TENTH EDITION

ROBBINS & COTRAN

PATHOLOGIC BASIS OF DISEASE

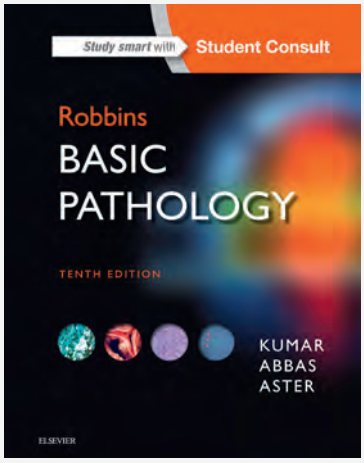
KUMAR
ABBAS
ASTER

ELSEVIER

Robbins and Cotran.

Where an effective understanding of disease begins.

A current and accurate understanding of pathophysiology is integral to effective practice in every specialty. No source delivers that understanding better than the Robbins and Cotran family! *Robbins Basic Pathology*, *Robbins Essential Pathology*, and their many companion resources provide a clear, up-to-date explanation of disease mechanisms you need to succeed.

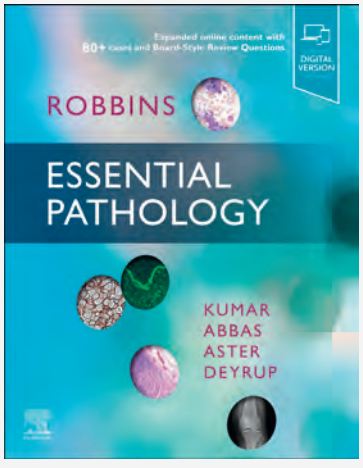


Robbins Basic Pathology, 10th Edition

ISBN: 9780323353175

Key Features:

- Excellent art program boasts high-quality photomicrographs, gross photos, and radiologic images to supplement the world-class illustrations.
- Bulleted summary boxes provide quick access to key information and easy review of key concepts.
- Highlights pathogenesis, morphology, and pathophysiologic content throughout.



Robbins Essential Pathology, 1st Edition

ISBN: 9780323640251

Key Features:

- The most concise Robbins text available, providing high-quality content and a case-based approach in a focused, multimedia learning package for coursework and exam preparation.
- Focuses on the core knowledge of disease mechanisms and essential clinical aspects that medical students need to know.
- Features more than 500 images and tables that illustrate key disorders and concepts.

Explore the entire Robbins and Cotran Family at elsevierhealth.com/pathology

ROBBINS&COTRAN PATHOLOGIC BASIS OF DISEASE

ROBBINS & COTRAN PATHOLOGIC BASIS OF DISEASE

TENTH EDITION

Vinay Kumar, MBBS, MD, FRCPath
Alice Hogge and Arthur A. Baer Distinguished Service Professor of Pathology
Biologic Sciences Division and the Pritzker Medical School
The University of Chicago
Chicago, Illinois

Abul K. Abbas, MBBS
Emeritus Professor and Chair
Department of Pathology
University of California San Francisco
San Francisco, California

Jon C. Aster, MD, PhD
Michael Gimbrone Professor of Pathology
Brigham and Women's Hospital and Harvard Medical School
Boston, Massachusetts

Associate Editor
Jerrold R. Turner, MD, PhD
Professor of Pathology and Medicine
Brigham and Women's Hospital
Harvard Medical School
Boston, Massachusetts

WITH ILLUSTRATIONS BY:
James A. Perkins, MS, MFA
Distinguished Professor of Medical Illustration
Rochester Institute of Technology
Rochester, New York

ELSEVIER

Elsevier

1600 John F. Kennedy Blvd.
Ste 1800
Philadelphia, PA 19103-2899

ROBBINS & COTRAN PATHOLOGIC BASIS OF DISEASE, TENTH EDITION ISBN: 978-0-323-53113-9
INTERNATIONAL EDITION ISBN: 978-0-323-60992-0

International Standard Book Number: 978-0-323-53113-9

Publisher: Jeremy Bowes
Content Development Director: Rebecca Gruliow
Publishing Services Manager: Catherine Jackson
Health Content Management Specialist: Kristine Feeherty
Design Direction: Brian Salisbury

Printed in Canada

Last digit is the print number: 9 8 7 6 5 4 3 2 1

DEDICATION

To
Our teachers
For inspiring us

To
Our students
For constantly challenging us

To our spouses
Raminder Kumar
Ann Abbas
Erin Malone
For their unconditional support

Contributors

Mahul B. Amin, MD
Chairman
Department of Pathology
College of Medicine
University of Tennessee Health Science Center
Memphis, Tennessee
The Lower Urinary Tract and Male Genital System

Douglas C. Anthony, MD, PhD
Professor
Pathology, Laboratory Medicine, and Neurology
Warren Alpert Medical School of Brown University;
Chief of Pathology
Lifespan Academic Medical Center
Providence, Rhode Island
Peripheral Nerves and Skeletal Muscles

Anthony Chang, MD
Professor
Department of Pathology
The University of Chicago
Chicago, Illinois
The Kidney

Nicole A. Cipriani, MD
Associate Professor of Pathology
The University of Chicago
Chicago, Illinois
Head and Neck

Andrew J. Connolly, MD, PhD
Professor of Pathology
University of California San Francisco
San Francisco, California
The Heart

Lora Hedrick Ellenson, MD
Attending Physician and Director of Gynecologic Pathology
Department of Pathology
Memorial Sloan Kettering Cancer Center
New York, New York
The Female Genital Tract

Robert Folberg, MD
Departments of Ophthalmology and Pathology
Beaumont Health - Royal Oak
Royal Oak, Michigan
The Eye

Karen M. Frank, MD, PhD, D(ABMM)
Chief
Department of Laboratory Medicine
Clinical Center
National Institutes of Health
Bethesda, Maryland
Infectious Diseases

Ryan M. Gill, MD, PhD
Professor of Pathology
University of California San Francisco School of Medicine
San Francisco, California
Liver and Gallbladder

Marc K. Halushka, MD, PhD
Deputy Director of Education
Department of Pathology
Professor of Pathology
The Johns Hopkins Hospital
Baltimore, Maryland
Blood Vessels

Andrew Horvai, MD, PhD
Professor
Department of Pathology
University of California San Francisco
San Francisco, California
Bones, Joints, and Soft Tissue Tumors

Aliya N. Husain, MBBS
Professor
Department of Pathology
The University of Chicago
Chicago, Illinois
Diseases of Infancy and Childhood; The Lung

Sanjay Kakar, MD
Professor of Pathology
University of California San Francisco School of Medicine
San Francisco, California
Liver and Gallbladder

Selene C. Koo, MD, PhD
Assistant Professor
Department of Pathology
The Ohio State University;
Pathologist
Department of Pathology and Laboratory Medicine
Nationwide Children's Hospital
Columbus, Ohio
Diseases of Infancy and Childhood

Zoltan G. Laszik, MD, PhD
Professor of Pathology
University of California San Francisco
San Francisco, California
The Kidney

Alexander J. Lazar, MD, PhD
Professor
Departments of Pathology, Genomic Medicine, Dermatology, and Translational Molecular Pathology
The University of Texas MD Anderson Cancer Center
Houston, Texas
The Skin

Susan C. Lester, MD, PhD
Assistant Professor
Breast Pathology Services
Department of Pathology
Harvard Medical School
Brigham and Women's Hospital
Boston, Massachusetts
The Breast

Mark W. Lingen, DDS, PhD, FRCPath
Professor of Pathology
The University of Chicago
Chicago, Illinois
Head and Neck

Anirban Maitra, MBBS
Professor
Pathology and Translational Molecular Pathology
The University of Texas MD Anderson Cancer Center
Houston, Texas
The Pancreas; The Endocrine System

Marta Margeta, MD, PhD
Associate Professor of Pathology
University of California San Francisco School of Medicine
San Francisco, California
The Central Nervous System

Alexander J. McAdam, MD, PhD
Associate Professor of Pathology
Department of Pathology
Harvard Medical School;
Medical Director
Clinical Microbiology Laboratory
Boston Children's Hospital
Boston, Massachusetts
Infectious Diseases

Richard N. Mitchell, MD, PhD
Lawrence J. Henderson Professor of Pathology
Member of the Harvard/MIT Health Sciences and Technology Faculty
Department of Pathology
Brigham and Women's Hospital
Harvard Medical School
Boston, Massachusetts
The Cell as a Unit of Health and Disease; Blood Vessels; The Heart

George Jabboure Netto, MD
Chair, Department of Pathology
University of Alabama at Birmingham School of Medicine
Birmingham, Alabama
The Lower Urinary Tract and Male Genital System

Scott A. Oakes, MD
Professor and Vice Chair of Research and Academic Affairs
Department of Pathology
University of Chicago Pritzker School of Medicine
Chicago, Illinois
Cell Injury, Cell Death, and Adaptations

Arie Perry, MD
Professor of Pathology and Neurological Surgery
Director of Neuropathology
University of California San Francisco
San Francisco, California
The Central Nervous System

Edyta C. Pirog, MD
Associate Professor
Department of Pathology and Laboratory Medicine
Weill Cornell Medicine-New York Presbyterian Hospital
New York, New York
The Female Genital Tract

Peter Pytel, MD
Professor
Department of Pathology
University of Chicago
Chicago, Illinois
Peripheral Nerves and Skeletal Muscles

Preface

As we launch the tenth edition of *Pathologic Basis of Disease* we look to the future of pathology as a discipline and how this textbook can remain most useful to readers in the twenty-first century. It is obvious that an understanding of disease mechanisms is based more than ever on a strong foundation of basic science. We have always woven the relevant basic cell and molecular biology into the sections on pathophysiology in various chapters. In the previous edition we went one step further and introduced a new chapter at the very beginning of the book titled "The Cell as a Unit of Health and Disease." We are delighted that the chapter was found useful by many students and faculty. Because progress in basic cell biology is moving at a very brisk pace, the chapter has been updated significantly.

In the preface of the very first edition (1957), Stanley Robbins wrote:

"The pathologist is interested not only in the recognition of structural alterations, but also in their significance, i.e., *the effects of these changes on cellular and tissue function* and ultimately the effect of these changes on the patient. It is not a discipline isolated from the living patient, but rather *a basic approach to a better understanding of disease and therefore a foundation of sound clinical medicine*."

We hope we continue to illustrate the principles of pathology that Dr. Robbins enunciated with such elegance and clarity over half a century ago.

This edition, like all previous ones, has been extensively revised, and some areas have been completely rewritten. A few examples of significant changes are as follows:

- Chapter 2 has been updated to include novel pathways of cell death beyond the long-established pathways of necrosis and apoptosis. Indeed, the distinction between these two is being blurred. Autophagy, which has begun to take center stage in diseases ranging from aging to cancer and neurodegeneration, has been revised, as have the possible molecular mechanisms of aging.
- In Chapter 3, we have married morphology with molecular mechanisms. Thus, different patterns of inflammation can be related to distinct molecular pathways.
- Chapter 5 includes a discussion of gene editing technology and a revised and updated section on molecular diagnosis.
- Chapter 7 has been extensively revised to incorporate new insights into cancer pathogenesis, the interplay between cancer cells and host immunity, and cancer precision diagnostics.
- Chapter 10, covering pediatric diseases, includes discussion of newly approved therapies targeting specific mutated forms of the CFTR transporter.
- Chapter 11, covering vascular diseases, includes discussion of clonal hematopoiesis, a newly emerging risk factor for atherosclerosis and other inflammatory disorders.
- Chapters 18 and 28, covering diseases of the liver and the central nervous system, have a fresh look brought in by new contributors.
- In addition to the revision and reorganization of the text, many new photographs and schematics have been added and a large number of the older "gems" have been enhanced by digital technology.

We made the changes highlighted above while remaining focused on the same longstanding overarching goals, which serve as our guiding principles:

- To integrate into the discussion of pathologic processes and disorders the newest established information available—morphologic as well as molecular.
- To organize information into logical and uniform presentations, facilitating readability, comprehension, and learning.
- To maintain the book at a reasonable size and yet provide adequate discussion of significant lesions, processes, and disorders. Indeed, despite the addition of new information, we are happy to state that the overall length of the book is unchanged. One of our most challenging tasks is to decide what to eliminate to make room for key new findings.
- To place great emphasis on clarity of writing and proper use of language in the recognition that struggling to comprehend is time-consuming and wearisome and gets in the way of the learning process.
- To make this text useful to students throughout all of their years in medical school and into their residencies—but, at the same time, to provide sufficient detail and depth to meet the needs of more advanced readers.

We have repeatedly been told by readers that up-to-datedness is a special feature that makes this book very valuable. We have strived to remain current by providing new information from the recent literature, up to the current year, and by adding coverage of the COVID-19 epidemic.

We are now into the digital age, and so the text will be available online to those who own the print version. Such access gives the reader the ability to search across the entire text, bookmark passages, add personal notes, use PubMed to view references, and exploit many other exciting features. In addition, also available online are case studies developed by one of us (VK), and revised by Dr. Alex Gallan from the University of Chicago. The cases are designed to enhance and reinforce learning by challenging students to apply their knowledge to solve clinical cases. To assist in the classroom, we have also made the images available for instructors on the Evolve website. Instructors may register

at https://evolve.elsevier.com/ to gain access to the images for teaching purposes.

All three of us have reviewed, critiqued, and edited each chapter to ensure the uniformity of style and flow that have been the hallmarks of the book. Together, we hope that we have succeeded in equipping the readers with the scientific basis for the practice of medicine and in whetting their appetite for learning beyond what can be offered in any textbook.

Vinay Kumar
Abul K. Abbas
Jon C. Aster

Acknowledgments

First and foremost, all four of us offer thanks to our contributing authors for their commitment to this textbook. Many are veterans of previous editions; others are new to the tenth edition. All are acknowledged in the table of contents. Their names lend authority to this book, for which we are grateful. As in previous editions, the four of us have chosen not to add our own names to the chapters we have been responsible for writing, in part or whole. We welcome to this edition Dr. Jerry Turner in the capacity of Associate Editor. Jerry is a veteran of Robbins texts having written the chapter on diseases of the gastrointestinal tract in previous editions. His editing has strengthened several chapters.

Many colleagues have enhanced the text by reading various chapters and providing helpful critiques in their area of expertise. They include Dr. Celeste Thomas, University of Chicago; Dr. Meenakshi Jolly, Rush University, Chicago; Dr. Richard Aster, Blood Research Institute, Milwaukee; and Dr. Suneil Koliwad, UCSF. Many colleagues provided photographic gems from their collections. They are individually acknowledged in the text.

All of the graphic art in this book was created by Mr. James Perkins, Distinguished Professor of Medical Illustration at Rochester Institute of Technology. His ability to convert complex ideas into simple and aesthetically pleasing sketches has considerably enhanced this book.

Many individuals associated with our publisher, Elsevier, need our special thanks. Outstanding among them is Kristine Feeherty, Health Content Management Specialist, and our partner in the production of this book. Her understanding of the needs of the authors, promptness in responding to requests (both reasonable and unreasonable), and cheerful demeanor went a long way in reducing our stress and making our lives less complicated. Jim Merritt handed over the charge to Jeremy Bowes, Publisher. Our thanks also go to Director of Content Development, Rebecca Gruliow, and Designer Brian Salisbury. Undoubtedly there are many others who may have been left out unwittingly—to them we say "thank you" and tender apologies for not acknowledging you individually. We also want to acknowledge many readers—students, residents, and faculty members—scattered around the globe whose comments improve the book. We are impressed by their careful reading of the text.

Efforts of this magnitude take a toll on the families of the authors. We thank our spouses, Raminder Kumar, Ann Abbas, Erin Malone, and Judy Turner, for their patience, love, and support of this venture, and for their tolerance of our absences.

Vinay Kumar
Abul K. Abbas
Jon C. Aster
Jerrold R. Turner

Contents

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 1

The Cell as a Unit of Health and Disease

Richard N. Mitchell

CHAPTER CONTENTS

Pathology literally translates as the *study of suffering* (Greek *pathos* = suffering, *logos* = study); more prosaically, and as applied to modern medicine, it is the *study of disease.* Virchow was prescient in asserting that disease originates at the cellular level, but we now appreciate that cellular pathologies arise from perturbations in molecules (genes, proteins, and metabolites) that influence cell survival and behaviors. Thus the foundation of modern pathology is understanding the *cellular and molecular* aberrations that give rise to diseases. It is illuminating to consider these abnormalities in the context of normal cellular structure and function, which is the subject of this introductory chapter.

It is unrealistic (and even undesirable) to condense the vast and fascinating field of cell biology into a single chapter. Consequently, rather than attempting a comprehensive review, the goal here is to survey basic principles and highlight recent advances that are relevant to the mechanisms of disease that are emphasized throughout the rest of the book.

THE GENOME

The sequencing of the human genome at the beginning of the 21st century represented a landmark achievement of biomedical science. Since then the rapidly declining cost of sequencing, the burgeoning computational capacity to mine the ensuing data, and the expanding toolkits to analyze functional outputs (genomics, proteomics, and metabolomics) promise to revolutionize our understanding of health and disease. The emerging information has also revealed a breathtaking level of complexity far beyond the linear sequence of the genome. The potential of these powerful innovations to explain disease pathogenesis and drive therapeutic discovery excites and inspires scientists and the lay public alike.

Noncoding DNA

The human genome contains some 3.2 billion DNA base pairs. Yet, within the genome there are only about 20,000 protein-encoding genes, constituting just 1.5% of the genome. These are the blueprints that instruct the assembly of the enzymes, structural elements, and signaling molecules within the 50 trillion cells that make up the human body. Although 20,000 underestimates the actual number of encoded proteins (many genes produce multiple RNA transcripts that translate to different protein isoforms), it is nevertheless startling to realize that worms, which are composed of fewer than 1000 cells and have 30-fold smaller genomes also have about 20,000 protein-encoding genes. Many of these proteins are recognizable homologs of molecules expressed in humans. What then separates humans from worms?

The answer is not completely known, but evidence suggests that much of the difference lies in the 98.5% of the human genome that does not encode proteins. The function of such long stretches of DNA (so-called genome "dark matter") was mysterious for many years. However, over 85% of the human genome is ultimately transcribed; nearly 80% is devoted to regulation of gene expression. It follows that while proteins provide the building blocks and

machinery required for assembling cells, tissues, and organisms, it is the noncoding regions of the genome that provide the critical "architectural planning." Practically stated, the difference between worms and humans apparently lies more in the genomic "blueprints" than in the construction materials.

There are five major classes of functional non–protein-coding sequences in the human genome (Fig. 1.1):

- *Promoter* and *enhancer* regions that provide binding sites for transcription factors.
- Binding sites for factors that organize and maintain higher order *chromatin structures.*
- *Noncoding regulatory RNAs.* Over 60% of the genome is transcribed into RNAs that are never translated but regulate gene expression through a variety of mechanisms. The two best-studied varieties—micro-RNAs (miRNAs) and long noncoding RNAs (lncRNAs)—are described later.
- *Mobile genetic elements* (e.g., *transposons*) make up more than a third of the human genome. These "jumping genes" can move around the genome during evolution, resulting in variable copy number and positioning even among closely related species (e.g., humans and other primates). Although implicated in gene regulation and chromatin organization, the function of mobile genetic elements is not well established.
- Special structural regions of DNA, in particular, *telomeres* (chromosome ends) and *centromeres* (chromosome "tethers"). A major component of centromeres is so-called *satellite DNA,* consisting of large arrays—up to megabases in length—of repeating sequences (from 5 bp up to 5 kb). Although classically associated with spindle apparatus attachment, satellite DNA is also important in maintaining the dense, tightly packed organization of heterochromatin (discussed later).

Many genetic variations (polymorphisms) associated with diseases are located in non–protein-coding regions of the genome. Thus variation in gene regulation may prove to be more important in disease causation than structural changes in specific proteins. Another surprise that emerged from genome sequencing is that any two humans are typically more than 99.5% DNA-identical (and are 99% sequence-identical with chimpanzees)! Thus individual variation, including differential susceptibility to diseases and environmental stimuli, is encoded in less than 0.5% of our DNA (representing about 15 million bp).

The two most common forms of DNA variation in the human genome are single nucleotide polymorphisms (SNPs) and copy number variations (CNVs).

- *SNPs* are variants at single nucleotide positions and are almost always biallelic (only two choices exist at a given

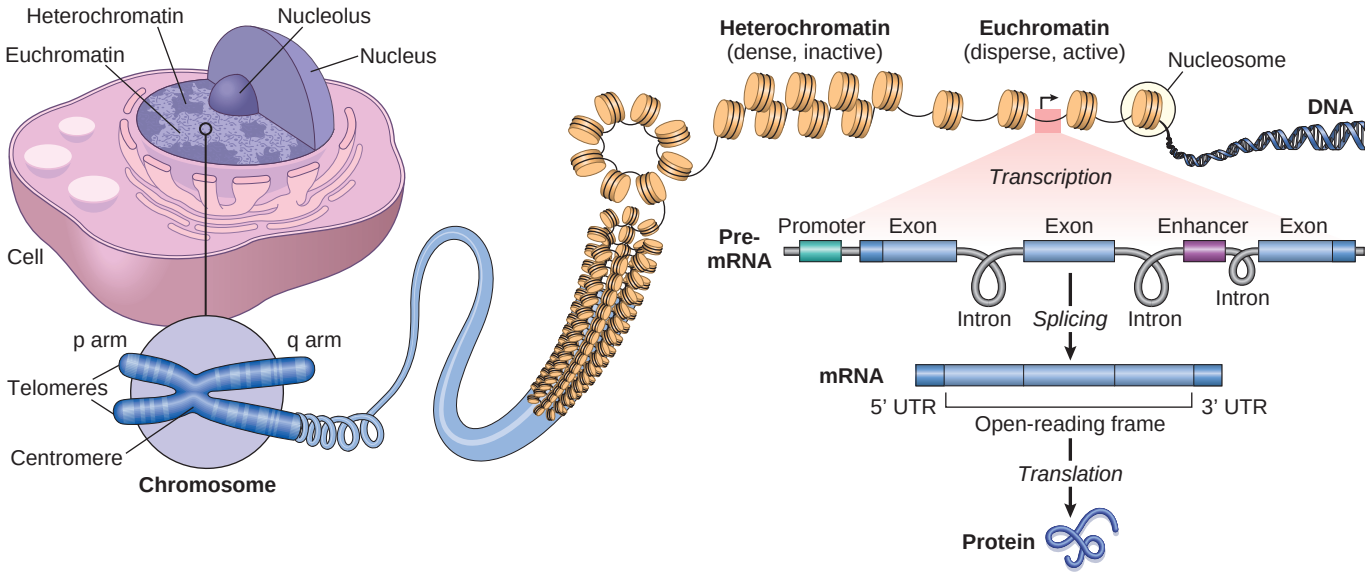


Figure 1.1 The organization of nuclear DNA. At the light microscopic level, the nuclear genetic material is organized into dispersed, transcriptionally active euchromatin and densely packed, transcriptionally inactive heterochromatin; chromatin can also be mechanically connected with the nuclear membrane, and membrane perturbation can thus influence transcription. Chromosomes (as shown) can be visualized only during mitosis. During mitosis, they are organized into paired chromatids connected at centromeres; the centromeres act as the locus for the formation of a kinetochore protein complex that regulates chromosome segregation at metaphase. The *telomeres* are repetitive nucleotide sequences that cap the termini of chromatids and permit repeated chromosomal replication without deterioration of genes near the ends. The chromatids are organized into short "P" ("petite") and long "Q" (next letter in the alphabet) arms. The characteristic banding pattern of chromatids has been attributed to relative GC content (less GC content in bands relative to interbands), with genes tending to localize to interband regions. Individual chromatin fibers are comprised of a string of nucleosomes—DNA wound around octameric histone cores—with the nucleosomes connected via DNA linkers. Promoters are noncoding regions of DNA that initiate gene transcription; they are on the same strand and upstream of their associated gene. Enhancers can modulate gene expression over distances of 100 kb or more by looping back onto promoters and recruiting additional factors that drive the expression of pre–messenger RNA (mRNA) species. Intronic sequences are spliced out of the pre-mRNA to produce the final message that is translated into protein—without the 3′–untranslated region (UTR) and 5′-UTR. In addition to the enhancer, promoter, and UTR sequences, noncoding elements, including short repeats, regulatory factor binding regions, noncoding regulatory RNAs, and transposons, are distributed throughout the genome.

site within the population, such as A or T). Over 6 million human SNPs have been identified, with many showing wide variation in frequency in different populations.
  - SNPs occur across the genome—within exons, introns, intergenic regions, and coding regions.
  - Roughly 1% of SNPs occur in coding regions, which is about what would be expected by chance, since coding regions comprise about 1.5% of the genome.
  - SNPs located in noncoding regions can occur within genomic regulatory elements, thereby altering gene expression; in such instances, SNPs influence disease susceptibility directly.
  - Some SNPs, termed "neutral" variants, are thought to have no effect on gene function or individual phenotype.
  - Even "neutral" SNPs may be useful markers if they happen to be coinherited with a disease-associated polymorphism as a result of physical proximity. In other words, the SNP and the causative genetic factor are in *linkage disequilibrium*.
  - The effect of most SNPs on disease susceptibility is weak, and it remains to be seen if identification of such variants, alone or in combination, can be used to develop effective strategies to identify those at risk and, ultimately, prevent disease.
- *CNVs* are a form of genetic variation consisting of different numbers of large contiguous stretches of DNA; these can range from 1000 base pairs to millions of base pairs. CNVs can be biallelic and simply duplicated or, alternatively, deleted in some individuals. At other sites there are complex rearrangements of genomic material, with multiple variants in the human population.
  - CNVs are responsible for between 5 million and 24 million base pairs of sequence difference between any two individuals.
  - Approximately 50% of CNVs involve gene-coding sequences; thus CNVs may underlie a large portion of human phenotypic diversity.

It is important to note that alterations in DNA sequence cannot by themselves explain the diversity of phenotypes in human populations; moreover, classic genetic inheritance cannot explain differing phenotypes in monozygotic twins. The answers to these conundrums probably lie in *epigenetics*—heritable changes in gene expression that are not caused by variations in DNA sequence (see the following section).

Histone Organization

Even though virtually all cells in the body have the same genetic composition, differentiated cells have distinct structures and functions that arise as a result of lineage-specific gene expression programs. Such cell type–specific differences in transcription and translation depend on *epigenetic factors* (literally, factors that are "above genetics") that can be conceptualized as follows (Fig. 1.2):

- *Histones and histone-modifying factors. Nucleosomes* consist of DNA segments 147 bp long that are wrapped around a central core structure of highly conserved low molecular weight proteins called *histones*. The resulting DNA-histone complex resembles a series of beads joined by short DNA linkers. The naked DNA of a single human cell is about 1.8 m long. By winding around histones, like spools of thread, the entire genome can be packed into a nucleus as small as 7 to 8 μm in diameter. In most cases, this structured DNA, termed chromatin, is not wound uniformly. Thus at the light microscopic level, nuclear chromatin is recognizable as cytochemically dense and transcriptionally inactive heterochromatin and disperse, transcriptionally active euchromatin (see Fig. 1.1). In general, only the regions that are "unwound" are available for transcription. Chromatin structure can therefore regulate transcription independent of traditional promoters and DNA-binding elements and, due to variations between cell types, helps to define cellular identity and activity.

  Histones are not static, but rather are highly dynamic structures regulated by a host of nuclear proteins. Thus chromatin remodeling complexes can reposition nucleosomes on DNA, exposing (or obscuring) gene regulatory elements such as promoters. "Chromatin writer" complexes, on the other hand, carry out over 70 different histone modifications generically denoted as "marks." Such covalent alterations include methylation, acetylation, or phosphorylation of specific amino acids within histones.

  Actively transcribed genes in euchromatin are associated with histone marks that make the DNA accessible to RNA polymerases. In contrast, inactive genes have histone marks that enable DNA compaction into heterochromatin. Histone marks are reversible through the activity of "chromatin erasers." Still other proteins function as "chromatin readers," binding histones that bear particular marks and thereby regulating gene expression.
- *Histone methylation.* Both lysines and arginines can be methylated by specific writer enzymes; methylation of histone lysine residues can lead to transcriptional activation or repression, depending on which histone residue is marked.
- *Histone acetylation.* Lysine residues are acetylated by histone acetyltransferases (HATs), whose modifications tend to open the chromatin and increase transcription. In turn, these changes can be reversed by histone deacetylases (HDACs), leading to chromatin condensation.
- *Histone phosphorylation.* Serine residues can be modified by phosphorylation; depending on the specific residue, the DNA may be opened for transcription or condensed and inactive.
- *DNA methylation.* High levels of DNA methylation in gene regulatory elements typically result in transcriptional silencing. Like histone modifications, DNA methylation is tightly regulated by methyltransferases, demethylating enzymes, and methylated-DNA-binding proteins.
- *Chromatin organizing factors.* Much less is known about these proteins, which are believed to bind to noncoding regions and control long-range looping of DNA, thus regulating the spatial relationships between enhancers and promoters that control gene expression.

Deciphering the mechanisms that allow epigenetic factors to control genomic organization and gene expression in a cell-type-specific fashion is an extraordinarily complex proposition. Despite the intricacies, there is already ample evidence that dysregulation of the "epigenome" has a central role in malignancy (Chapter 7), and emerging data indicate

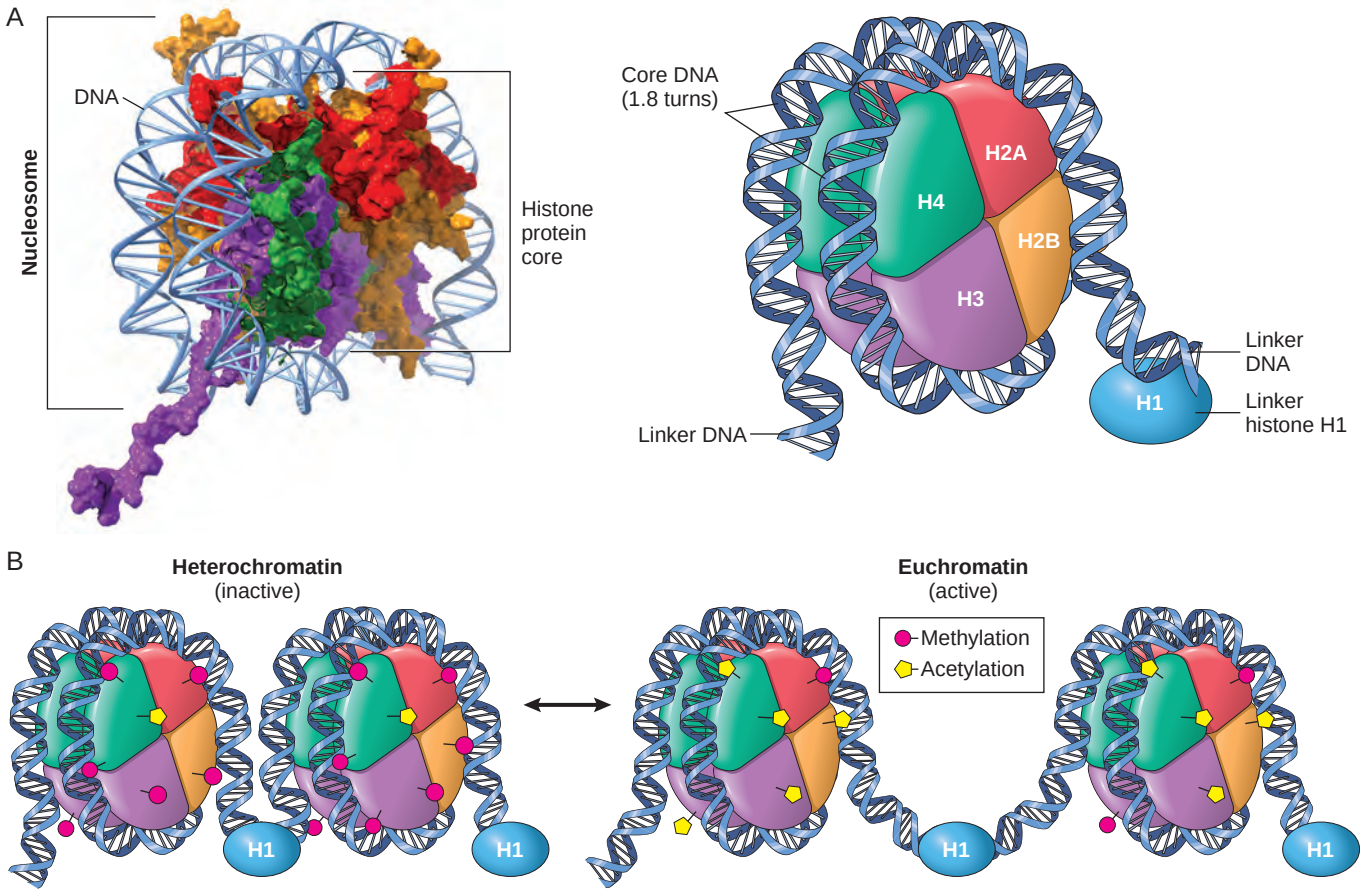


Figure 1.2 Histone organization. (A) Nucleosomes are comprised of octamers of histone proteins (two each of histone subunits H2A, H2B, H3, and H4) encircled by 1.8 147 bp DNA loops. Histones sit on 20- to 80-nucleotide stretches of linker DNA between nucleosomes. Histone subunits are positively charged, thus allowing compaction of negatively charged DNA. (B) The relative state of DNA unwinding (and thus access for transcription factors) is regulated by histone modification, including acetylation, methylation, and/or phosphorylation; these "marks" are dynamically written and erased. Certain marks such as histone acetylation "open up" the chromatin structure, whereas others such as methylation of particular histone residues condense DNA to silence genes. DNA can also be methylated, leading to transcriptional inactivation.

that many other diseases are associated with inherited or acquired epigenetic alterations. Unlike genetic changes, many epigenetic alterations (e.g., histone acetylation and DNA methylation) are reversible and amenable to therapeutic intervention; HDAC and DNA methylation inhibitors are already being tested in the treatment of various forms of cancer.

Micro-RNA and Long Noncoding RNA

Genes can also be regulated by noncoding RNAs. These genomic sequences are transcribed but not translated. Although many distinct families of noncoding RNAs exist, we will discuss only two examples here: small RNA molecules called *microRNAs (miRNAs)* and *long noncoding RNAs (lncRNAs)* (>200 nucleotides in length).

Micro-RNA

miRNAs do not encode proteins; they modulate translation of target messenger RNAs (mRNAs). Posttranscriptional silencing of gene expression by miRNA is a fundamental and well-conserved mechanism of gene regulation present in all eukaryotes (plants, animals, and fungi). Even bacteria have a primitive version of the same machinery that they use to protect themselves against foreign DNA (e.g., phage and virus DNA). The profound influence of miRNAs on protein expression allows these relatively short RNAs (22 nucleotides on average) to be critical regulators of developmental pathways as well as pathologic conditions (e.g., cancer).

The human genome encodes almost 6000 miRNA genes, about 30% of the total number of protein-coding genes. Individual miRNAs can regulate multiple protein-coding genes, allowing each miRNA to coregulate entire programs of gene expression. Transcription of miRNA genes produces a primary transcript (pri-miRNA) that is processed into progressively smaller segments, including trimming by the enzyme *Dicer*. This generates mature single-stranded miRNAs of 21 to 30 nucleotides that associate with a multiprotein aggregate called *RNA-induced silencing complex (RISC)* (Fig. 1.3). Subsequent base pairing between the miRNA strand and its target mRNA directs the RISC to either induce mRNA cleavage or repress its translation. In this way the target mRNA is *posttranscriptionally silenced*.

Small interfering RNAs (siRNAs) are short RNA sequences that can be introduced experimentally into cells where they

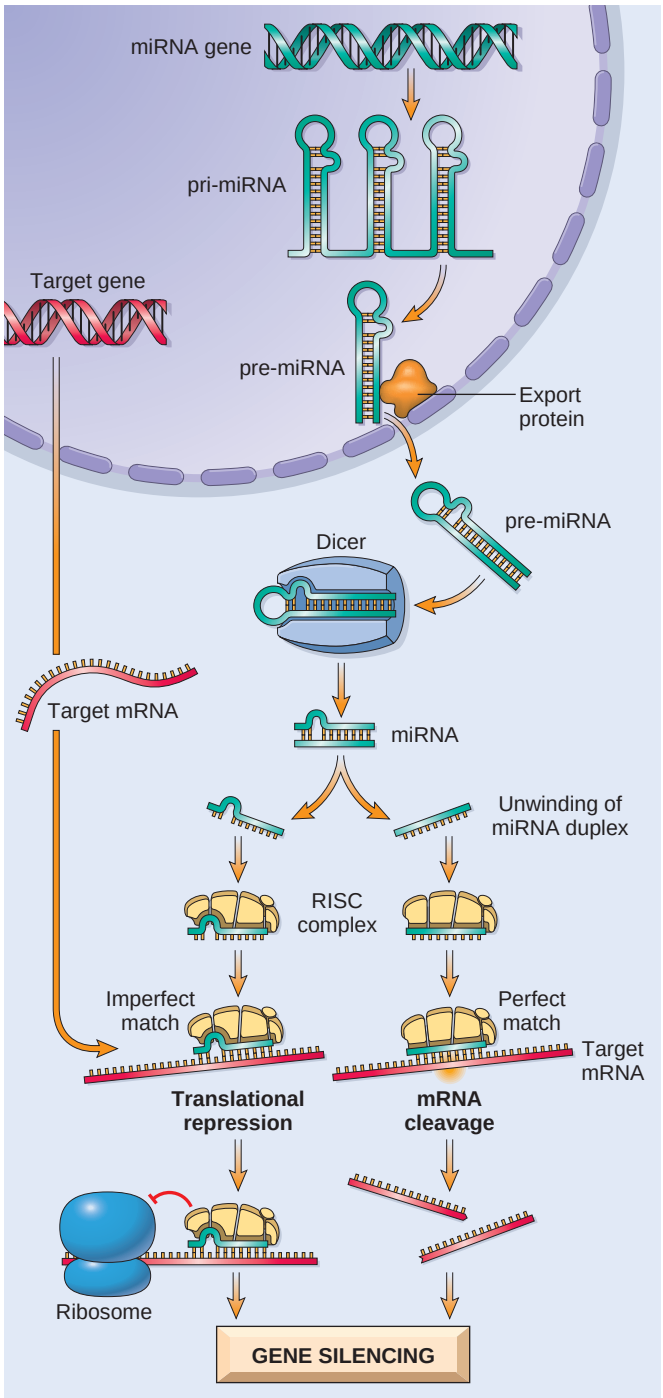


Figure 1.3 Generation of microRNAs (miRNAs) and their mode of action in regulating gene function. Transcription of a miRNA produces a primary miRNA (pri-miRNA), which is processed within the nucleus to form *pre-miRNA* composed of a single RNA strand with secondary hairpin loop structures and stretches of double-stranded RNA. After export out of the nucleus via specific transporter proteins, pre-miRNA is trimmed by the cytoplasmic *Dicer enzyme* to generate mature double-stranded miRNAs of 21 to 30 nucleotides. The miRNA subsequently unwinds and the single strands are incorporated into multiprotein RNA-induced silencing complexes (RISC). Base pairing between single-stranded miRNA and the targeted messenger RNA (mRNA) directs RISC to either cleave or repress translation of the mRNA, resulting in posttranscriptiontional silencing.

serve as substrates for Dicer and interact with RISC, thereby reproducing endogenous miRNAs function. Synthetic siRNAs that target specific mRNA species are powerful laboratory tools to study gene function (so-called knockdown technology) and are also being studied as potential therapeutic agents to silence pathogenic genes (e.g., oncogenes that drive neoplastic transformation).

Long Noncoding RNA

Recent studies have further identified an untapped universe of lncRNAs—by some calculations, the number of lncRNAs may exceed coding mRNAs by 10-fold to 20-fold. lncRNAs modulate gene expression by several mechanisms (Fig. 1.4). As one example, lncRNAs can bind to chromatin and restrict RNA polymerase from accessing coding genes within that region. The best-known example is XIST, which is transcribed from the X chromosome and plays an essential role in the physiologic X chromosome inactivation that occurs in

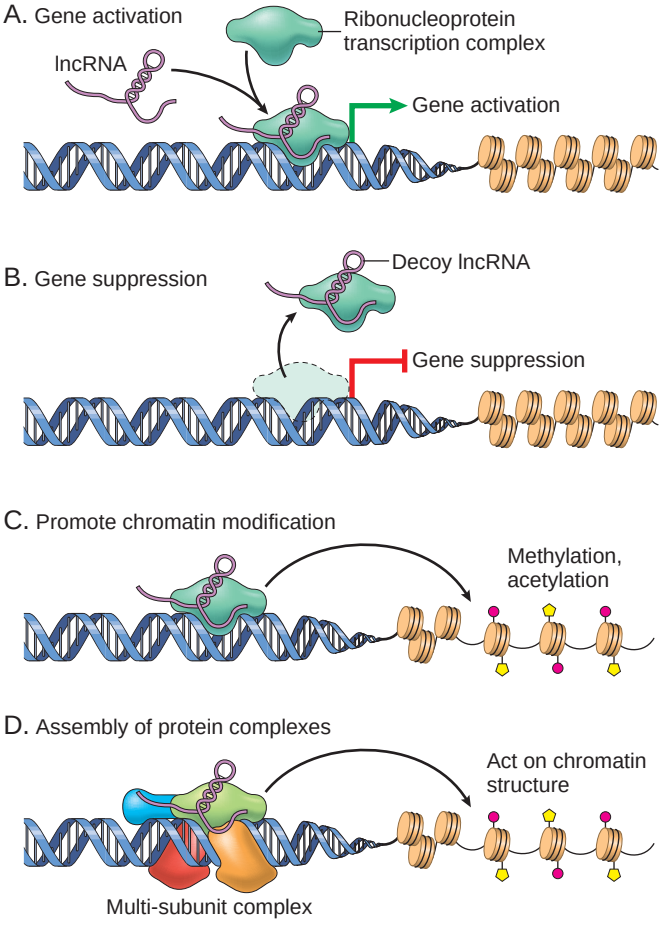


Figure 1.4 Roles of long noncoding RNAs (lncRNAs). (A) lncRNAs can facilitate transcription factor binding and thus promote gene activation. (B) Conversely, lncRNAs can preemptively bind transcription factors to inhibit transcription. (C) Histone and DNA modification by acetylases or methylases (or deacetylases and demethylases) may be directed by lncRNA binding. (D) In other instances, lncRNAs can act as scaffolds to stabilize secondary or tertiary structures and multisubunit complexes that influence chromatin architecture or gene activity. (Modified from Wang KC, Chang HY: Molecular mechanisms of long noncoding RNAs, *Mol Cell* 43:904, 2011.)

females. XIST itself escapes X inactivation but forms a repressive "cloak" on the X chromosome from which it is transcribed, resulting in gene silencing. Conversely, many enhancers are actually sites of lncRNA synthesis. In this case the lncRNAs expand transcription from gene promoters via a variety of mechanisms (see Fig. 1.4).

Gene Editing

An exciting new development that allows high-fidelity genome editing may usher in the next era of the molecular revolution. This advance comes from a wholly unexpected source: the discovery of *clustered regularly interspaced short palindromic repeats (CRISPRs)* and CRISPR-associated genes (Cas), such as the Cas9 nuclease. These are linked genetic elements that endow prokaryotes with a form of acquired immunity to phages and plasmids. Bacteria use the system to sample the DNA of infecting agents and integrate portions into their genomes as CRISPRs. These CRISPR segments are subsequently transcribed and processed into guide RNA sequences that bind and direct the Cas9 nuclease to specific sites (e.g., a phage sequence) so that it can be cleaved to disable the infecting agent.

Gene editing repurposes this process by using artificial 20-base guide RNAs (gRNAs) that bind Cas9 and are complementary to a targeted DNA sequence (Fig. 1.5). Cas9 then induces double-stranded DNA breaks at the site of gRNA binding. Repair of the highly specific cleavages can lead to random disruptive mutations (through nonhomologous end joining) or can introduce new genetic material with precision (by homologous recombination). Both the guide sequences and the Cas enzyme, either as a coding DNA (cDNA) or a protein, can be easily introduced into cells. The potential application to genetic engineering, due to the impressive specificity of the Cas9 system (up to 10,000-fold better than other previous editing systems), has led to great excitement. Applications include inserting specific mutations in cells and tissues to model cancers and other diseases and rapidly generating transgenic animal models from edited embryonic stem cells. CRISPR also makes it possible to selectively edit mutations that cause hereditable disease, or—perhaps more worrisome—to just eliminate less "desirable" traits. Predictably the technology has inspired a vigorous debate regarding the ethics of its use.

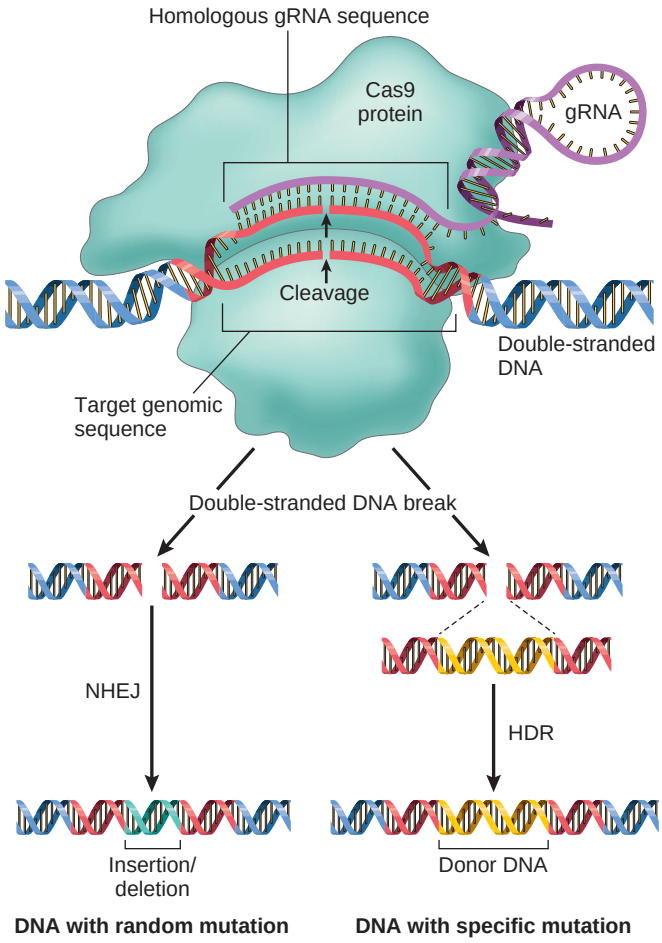


Figure 1.5 Gene editing with clustered regularly interspersed short palindromic repeats (CRISPRs) and the nuclease Cas9. In bacteria, DNA sequences consisting of CRISPRs are transcribed into guide RNAs *(gRNAs)* with a constant region and a variable sequence of approximately 20 bases. The gRNA constant regions bind to Cas9, while the variable regions form heteroduplexes with homologous DNA sequences of interest; the Cas9 nuclease then cleaves the bound DNA to produce a double-stranded DNA break. In nature, bacteria use the CRISPR/Cas9 system to protect against phages and plasmids; CRISPR sequences from previous assaults are transcribed into gRNA from the bacterial genome. These bind to pathogen nucleotide sequences and form a complex with the Cas9 nuclease that leads to cleavage and, ultimately, destruction of the invader's DNA.

To perform *gene editing*, gRNAs are designed with variable regions that are homologous to a specific DNA sequence of interest; coexpression of the gRNA and Cas9 then leads to efficient and highly specific cleavage of the target sequence. In the absence of homologous DNA, the double-stranded break is repaired by nonhomologous end-joining *(NHEJ)*, an error-prone mechanism that typically introduces disruptive insertions or deletions (indels). Conversely, in the presence of homologous "donor" DNA that spans the region target by the CRISPR/Cas9 complex, cells instead can use homologous DNA recombination *(HDR)* to repair the break. HDR is less efficient than NHEJ but has the capacity to introduce precise changes in DNA sequence. Potential applications of CRISPR/Cas9 coupled with HDR include repair of inherited genetic diseases and the creation of pathogenic mutations in inducible pluripotent stem cells.

CELLULAR HOUSEKEEPING

Normal functioning and intracellular homeostasis depend on a variety of fundamental cell housekeeping functions that all differentiated cells must perform to maintain viability and normal activity. These include protection from the environment, nutrient acquisition, metabolism, communication, movement, renewal of senescent molecules, molecular catabolism, and energy generation.

Many of the normal housekeeping functions of the cell are compartmentalized within membrane bound intracellular organelles (Fig. 1.6). By isolating certain cellular functions within distinct compartments, potentially injurious degradative enzymes or toxic metabolites can be kept at usefully high concentrations without risking damage to more delicate intracellular constituents. Moreover, compartmentalization also allows the creation of unique intracellular environments (e.g., low pH or high calcium) that permit more efficient functioning of certain enzymes or metabolic pathways.

New proteins destined for the plasma membrane or secretion are physically assembled in the *rough endoplasmic*

Relative volumes of intracellular organelles (hepatocyte)

Compartment	% total volume	number/cell	role in the cell
Cytosol	54%	1	metabolism, transport, protein translation
Mitochondria	22%	1700	energy generation, apoptosis
Rough ER	9%	1	synthesis of membrane and export protein
Smooth ER, Golgi	6%	1	protein modification, sorting, catabolism
Nucleus	6%	1	cell regulation, proliferation, DNA transcription
Endosomes	1%	200	intracellular transport and export
Lysosomes	1%	300	cellular catabolism
Peroxisomes	1%	400	very long-chain fatty acid metabolism

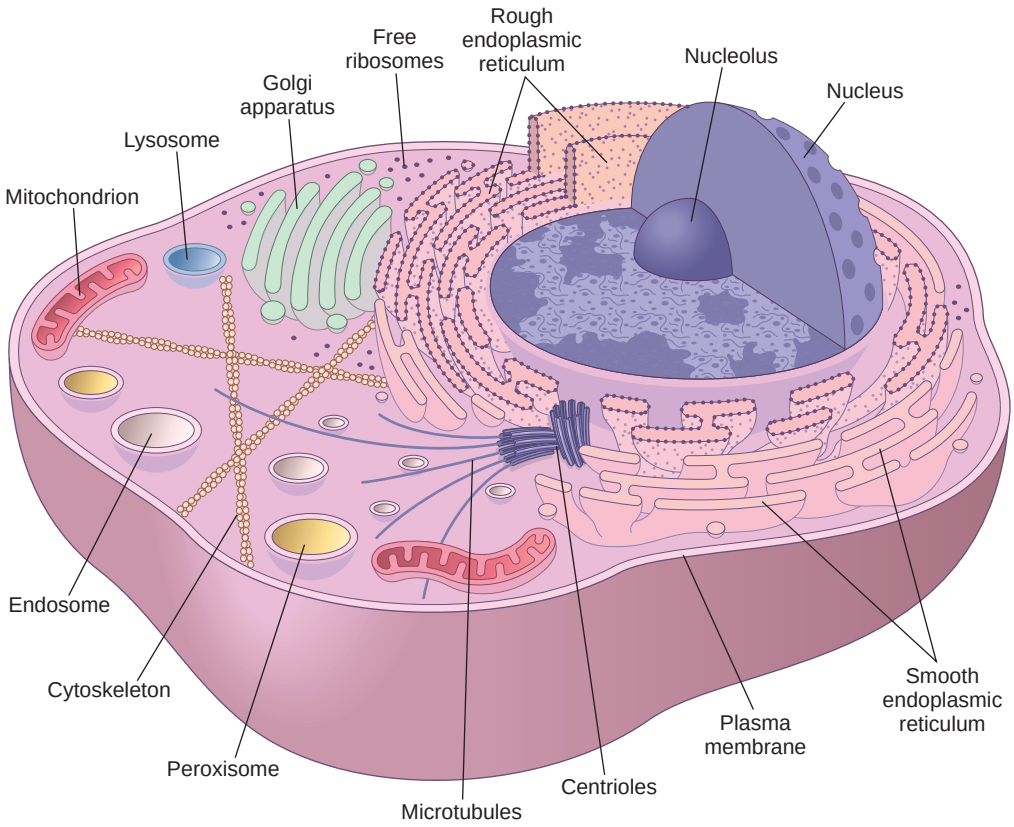


Figure 1.6 Basic subcellular constituents of cells. The table presents the number of the various organelles within a typical hepatocyte, as well as their volume within the cell. The figure shows geographic relationships but is not intended to be accurate to scale. *ER,* Endoplasmic reticulum. (Modified from Weibel ER, Stäubli W, Gnägi HR, et al: Correlated morphometric and biochemical studies on the liver cell. I. Morphometric model, stereologic methods, and normal morphometric data for rat liver. *J Cell Biol* 42:68, 1969.)

reticulum (RER) and *Golgi apparatus;* proteins intended for the cytosol are synthesized on free ribosomes. *Smooth endoplasmic reticulum (SER)* is used for steroid hormone and lipoprotein synthesis and modification of hydrophobic compounds into water-soluble molecules for export.

Cells catabolize the wide variety of molecules that they endocytose, as well as the entire repertoire of their own proteins and organelles—all of which are constantly being degraded and renewed. Breakdown of these constituents takes place at three different sites, ultimately serving different functions.

- *Proteasomes* are "disposal" complexes that degrade denatured or otherwise "tagged" cytosolic proteins. In antigen-presenting cells, the resulting short peptides are presented in the context of class I or class II major histocompatibility molecules to help drive the adaptive immune response (Chapter 6). In other cases, proteasomal degradation of regulatory proteins or transcription factors can trigger initiation or suppression of signaling pathways.
- *Lysosomes* are intracellular organelles containing degradative enzymes that permit digestion of a wide range of macromolecules, including proteins, polysaccharides, lipids, and nucleic acids. They are the site of senescent intracellular organelle breakdown (a process called *autophagy*) and where phagocytosed microbes are killed and catabolized.
- *Peroxisomes* contain catalase, peroxidase, and other oxidative enzymes; they play a specialized role in the breakdown of very long-chain fatty acids, generating hydrogen peroxide in the process.

The contents and location of cellular organelles are also highly regulated. *Endosomal vesicles* shuttle internalized material to the appropriate intracellular site(s), while other

membrane-bound vesicles direct newly synthesized materials to the cell surface or specific organelles. Movement—of both organelles and proteins within the cell, as well as the entire cell in its environment—is accomplished by the *cytoskeleton,* which is composed of filamentous actin (microfilaments), keratins (intermediate filaments), and microtubules. These structural proteins also maintain cellular shape and intracellular organization, which are essential to generation and maintenance of *cell polarity.* This is particularly important in epithelium where the top of the cell *(apical)* and the bottom and sides of the cell *(basolateral)* are exposed to different environments and have distinct functions. Loss of polarity could, for example, disrupt vectorial transcellular transport in the intestine or renal tubule.

Cell growth and maintenance require a constant supply of both energy and the building blocks that are needed for synthesis of macromolecules. Most of the adenosine triphosphate (ATP) that powers cells is generated via *mitochondrial oxidative phosphorylation.* Mitochondria also serve as an important source of metabolic intermediates needed for anabolic metabolism, are sites of synthesis of certain macromolecules (e.g., heme), and contain important sensors of cell damage that can initiate and regulate programmed cell death (e.g., apoptosis).

In growing and dividing cells, all of these organelles have to be replicated *(organellar biogenesis)* and correctly apportioned in daughter cells following mitosis. Moreover, because the macromolecules and organelles have finite lifespans (mitochondria, for example, last only about 10 days), mechanisms must also exist that allow for the recognition and degradation of "worn-out" cellular components.

With this as a primer, we will now move on to discuss cellular components and their function in greater detail.

Plasma Membrane: Protection and Nutrient Acquisition

Plasma membranes (and all other organellar membranes for that matter) are more than just static lipid sheaths. Rather, they are fluid bilayers of amphipathic phospholipids—hydrophilic head groups that face the aqueous environment and hydrophobic lipid tails that interact with each other to form a barrier to passive diffusion of large or charged molecules (Fig. 1.7). The bilayer has a remarkably heterogeneous composition of different phospholipids that vary by location and are also asymmetric—that is, membrane lipids preferentially associate with extracellular or cytosolic faces. Proper localization of these molecules is important for cell health. For example, specific phospholipids interact with particular membrane proteins and modify their distributions and functions.

- *Phosphatidylinositol* on the inner membrane leaflet can be phosphorylated, serving as an electrostatic scaffold for intracellular proteins; alternatively, polyphosphoinositides can be hydrolyzed by phospholipase C to generate intracellular second signals like diacylglycerol and inositol trisphosphate.

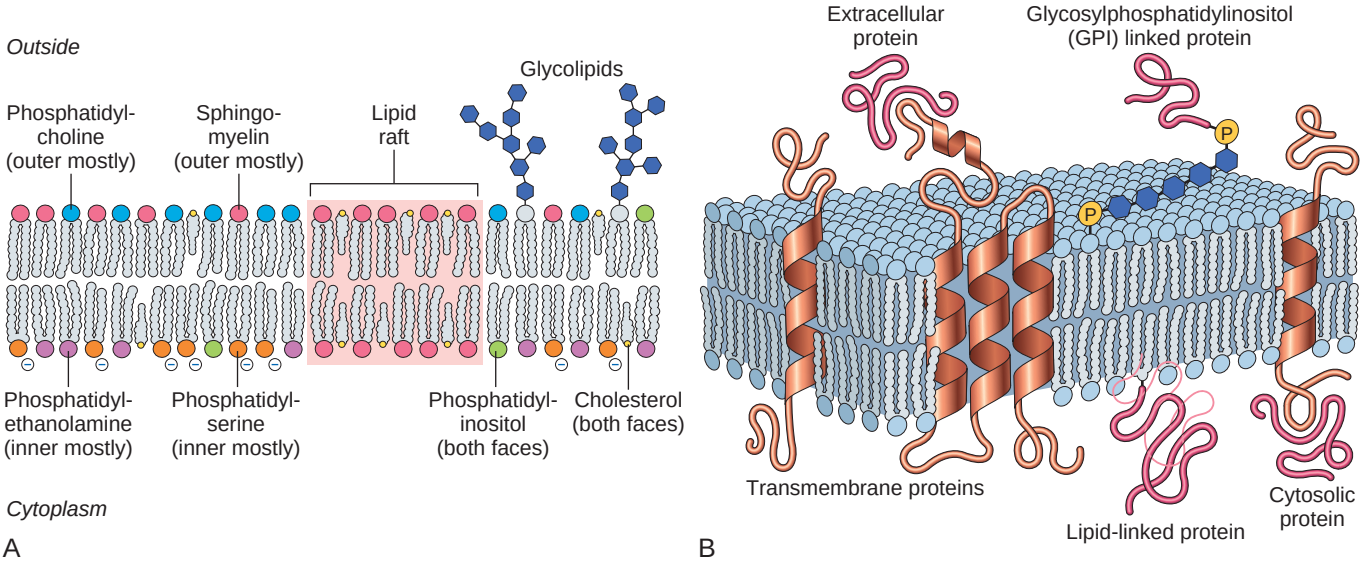


Figure 1.7 Plasma membrane organization and asymmetry. (A) The plasma membrane is a bilayer of phospholipids, cholesterol, and associated proteins. The phospholipid distribution within the membrane is asymmetric due to the activity of flippases; phosphatidylcholine and sphingomyelin are overrepresented in the outer leaflet, and phosphatidylserine (negative charge) and phosphatidylethanolamine are predominantly found on the inner leaflet; glycolipids occur only on the outer face where they contribute to the extracellular glycocalyx. Although the membrane is laterally fluid and the various constituents can diffuse randomly, specific domains, for example cholesterol and glycosphingolipid-rich lipid rafts, can also form. (B) Membrane-associated proteins may traverse the membrane (singly or multiply) via α-helical hydrophobic amino acid sequences; depending on the membrane lipid content and relative hydrophobicity of protein domains, such proteins may have nonrandom distributions within the membrane. Proteins on the cytosolic face can be associated with the plasma membrane through posttranslational modifications (e.g., farnesylation) or addition of palmitic acid. Proteins on the extracytoplasmic face can associate with the membrane via glycosylphosphatidylinositol *(GPI)* linkages. Besides protein-protein interactions within the membrane, membrane proteins can also associate with extracellular and/or intracytoplasmic proteins to generate distinct domains (e.g., the *focal adhesion complex*). Transmembrane proteins can translate mechanical forces (e.g., from the cytoskeleton or extracellular matrix), as well as chemical signals across the membrane.

- *Phosphatidylserine* is normally restricted to the inner face where it confers a negative charge involved in electrostatic protein interactions; however, when flipped to the extracellular leaflet, it becomes a potent "eat me" signal during programmed cell death (e.g., apoptosis). In platelets, phosphatidylserine is also a cofactor in blood clotting.
- *Glycolipids* and *sphingomyelin* are preferentially located on the extracellular face; glycolipids, including gangliosides with complex sugar linkages and terminal sialic acids that confer negative charges, support charge-based interactions that contribute to including inflammatory cell recruitment and sperm-egg fusion.

Despite substantial lateral fluidity, some membrane constituents concentrate into specialized domains (e.g., *lipid rafts*) that are enriched in glycosphingolipids and cholesterol. Since inserted membrane proteins have different intrinsic solubilities in domains with distinct lipid compositions, this membrane organization also impacts protein distribution. This geographic organization of plasma membrane components impacts cell-cell and cell-matrix interactions, intracellular signaling, and the specialized sites of vesicle budding or fusion.

The plasma membrane is liberally studded with a variety of proteins and glycoproteins involved in (1) ion and metabolite transport; (2) fluid-phase and receptor-mediated uptake of macromolecules; and (3) cell-ligand, cell-matrix, and cell-cell interactions. The means by which these proteins associate with membranes frequently reflects function. For example, multiple transmembrane-spanning proteins are often pores or molecular transporters, while proteins that are superficially attached to the membrane via labile linkages are more likely to participate in signaling. In general, proteins associate with the lipid bilayer by one of four mechanisms.

- Most proteins are integral or transmembrane proteins, having one or more relatively hydrophobic α-helical segments that traverse the lipid bilayer.
- Proteins synthesized on free ribosomes in the cytosol may be modified posttranslationally by addition of prenyl groups (e.g., farnesyl, related to cholesterol) or fatty acids (e.g., palmitic or myristic acid) that insert into the cytosolic side of the plasma membrane.
- Proteins on the extracellular face of the membrane may be anchored by glycosylphosphatidylinositol (GPI) tails that are added posttranslationally.
- Peripheral membrane proteins may noncovalently associate with true transmembrane proteins.

Many plasma membrane proteins function as large complexes; these may be aggregated either under the control of chaperone molecules in the RER or by lateral diffusion in the plasma membrane, followed by complex formation in situ. For example, many protein receptors (e.g., cytokine receptors) dimerize or trimerize in the presence of ligand to form functional signaling units. Although lipid bilayers are fluid within the plane of the membrane, components can be confined to discrete domains. This can occur by localization to lipid rafts (discussed earlier) or through intercellular protein-protein interactions (e.g., *tight junctions*) that establish discrete boundaries and also have unique lipid composition. The latter strategy is used to maintain *cell polarity* (e.g., top/apical/free vs. bottom/basolateral/bound to extracellular matrix [ECM]) in epithelial cells. Interactions of other membrane and cytosolic proteins with one another and the cytoskeleton also contributes to cell polarity.

The extracellular face of the plasma membrane is diffusely decorated by carbohydrates, not only as complex oligosaccharides on glycoproteins and glycolipids, but also as polysaccharide chains attached to integral membrane proteoglycans. This *glycocalyx* can form a chemical and mechanical barrier.

Membrane Transport

Although the barrier provided by plasma membranes is critical, transport of selected molecules across the lipid bilayer or to intracellular sites via vesicular transport is essential. Several mechanisms contribute to this transport.

Passive Diffusion. Small, nonpolar molecules like O_2 and CO_2 readily dissolve in lipid bilayers and therefore rapidly diffuse across them. Larger hydrophobic molecules, (e.g., steroid-based molecules like estradiol or vitamin D) can also cross lipid bilayers with relative impunity. While small polar molecules such as water (18 Da) can also diffuse across membranes at low rates, in tissues responsible for significant water movement (e.g., renal tubular epithelium), special integral membrane proteins called aquaporins form transmembrane channels for water, H_2O_2, and other small molecules. In contrast, the lipid bilayer is an effective barrier to the passage of larger polar molecules (>75 Da); at 180 Da, for example, glucose is effectively excluded. Lipid bilayers are also impermeant to ions due to their charge and hydration.

Carriers and Channels (Fig. 1.8)**.** Plasma membrane transport proteins are required for uptake and secretion of ions and larger molecules that are required for cellular function (e.g., nutrient uptake and waste disposal). Ions and small molecules can be transported by channel proteins and carrier proteins. Similar pores and channels also mediate transport across organellar membranes. These transporters that move ions, sugars, nucleotides, etc., frequently have exquisite specificities, and can be either active or passive (see below). For example, some transporters accommodate glucose but reject galactose.

- *Channel proteins* create hydrophilic pores, which, when open, permit rapid movement of solutes (usually restricted by size and charge).
- *Carrier proteins* bind their specific solute and undergo a series of conformational changes to transfer the ligand across the membrane; their transport is relatively slow.

Solute transport across the plasma membrane is frequently driven by a concentration and/or electrical gradient between the inside and outside of the cell via *passive transport* (virtually all plasma membranes have an electrical potential difference across them, with the inside negative relative to the outside). In other cases, *active transport* of certain solutes (against a concentration gradient) is accomplished by carrier molecules (never channels) at the expense of ATP hydrolysis or a coupled ion gradient. For example, most apical nutrient

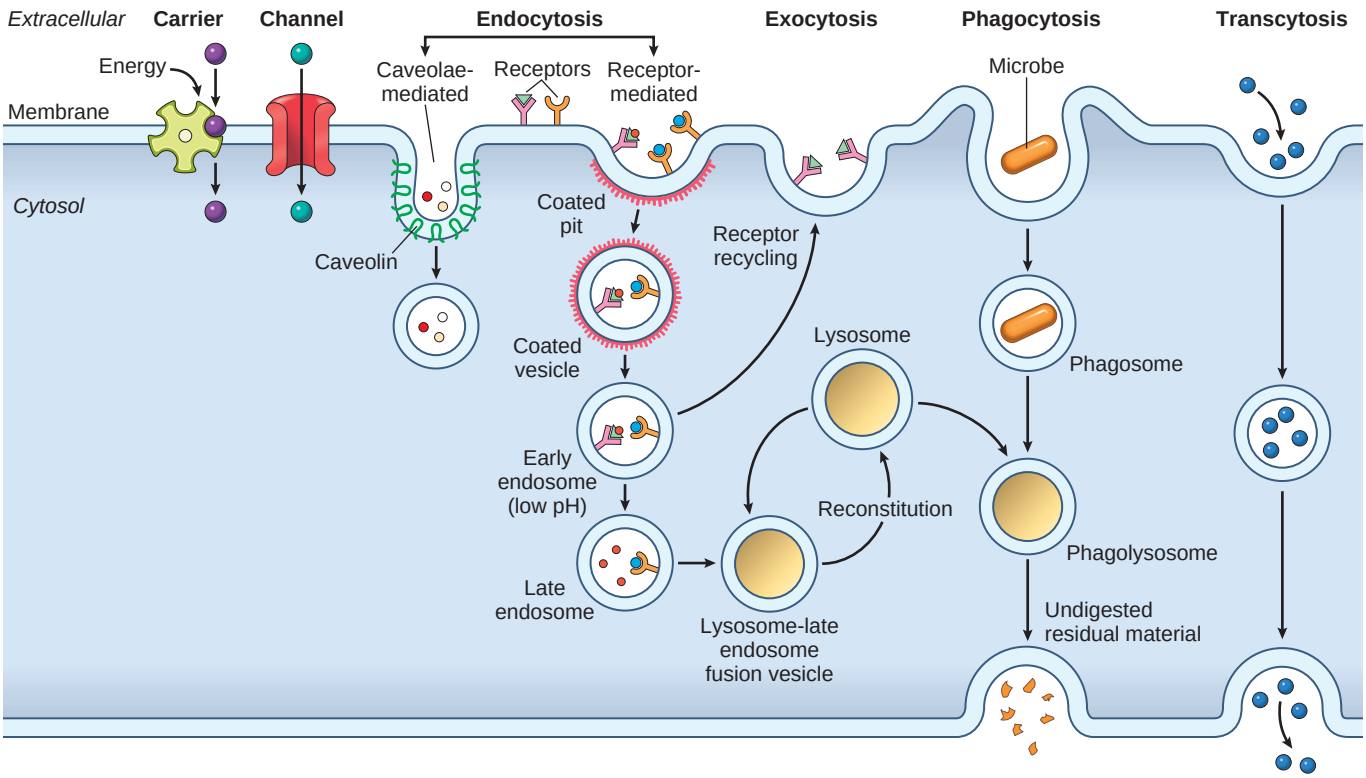


Figure 1.8 Movement of small molecules and larger structures across membranes. The lipid bilayer is relatively impermeable to all but the smallest and/or most hydrophobic molecules. Thus the import or export of charged species requires specific transmembrane transporter proteins, vesicular traffic, or membrane deformations.

From left to right in the figure: Small charged solutes can move across the membrane using either channels or carriers; in general, each molecule requires a unique transporter. Channels are used when concentration gradients can drive the solute movement; activation of the channel opens a hydrophilic pore that allows size-restricted and charge-restricted flow. Carriers are required when solute is moved against a concentration gradient; this typically requires energy expenditure to drive a conformational change in the carrier that facilitates the transmembrane delivery of specific molecules.

Receptor-mediated and fluid-phase uptake of material involves membrane bound vesicles. Caveolae endocytose extracellular fluid, membrane proteins, and some receptor bound molecules (e.g., folate) in a process driven by caveolin proteins concentrated within lipid rafts. They can subsequently fuse with endosomes or recycle to the membrane. Endocytosis of receptor-ligand pairs often involves clathrin-coated pits and vesicles. After internalization the clathrin disassembles and individual components can be re-used. The resulting vesicle becomes part of the endocytic pathway, in which compartments are progressively more acidic. After ligand is released, the receptor can be recycled to the plasma membrane to repeat the process (e.g., iron dissociates from transferin at pH ~5.5; apotransferrin and the transferrin receptor then return to the surface). Alternatively, receptor and ligand complexes can eventually be degraded within lysosomes (e.g., epidermal growth factor and its receptor are both degraded, which prevents excessive signaling). Exocytosis is the process by which membrane-bound vesicles fuse with the plasma membrane and discharge their contents to the extracellular space. This includes endosome recycling (shown), release of undigested residual material from lysosomes, transcytotic delivery of vesicles, and export of secretory vacuole contents (not shown). Phagocytosis involves membrane invagination to engulf large particles and is most common in specialized phagocytes (e.g., macrophages and neutrophils). The resulting phagosomes eventually fuse with lysosomes to facilitate the degradation of the internalized material. Transcytosis can mediate transcellular transport in either apical-to-basal or basal-to-apical directions, depending on the receptor and ligand.

transporters in the intestines and renal tubules exploit the extracellular to intracellular Na^+ gradient to allow absorption even when intracellular nutrient concentrations exceed extracellular concentrations. This form of active transport does not use ATP directly, but depends on the Na^+ gradient generated by Na^+-Ka^+ ATPase. Other transporters are ATPases. One example is the *multidrug resistance (MDR) protein*, which pumps polar compounds (e.g., chemotherapeutic drugs) out of cells and may render cancer cells resistant to treatment.

Water movement into or out of cells is passive and directed by solute concentrations. Thus extracellular salt in excess of that in the cytoplasm *(hypertonicity)* causes net movement of water out of cells, while *hypotonicity* causes net movement of water into cells. Conversely, the charged metabolites and proteins within the cytoplasm attract charged counterions that increase intracellular osmolarity. Thus to prevent overhydration, cells must constantly pump out small inorganic ions (e.g., Na^+)—typically through the activity of the membrane ion-exchanging ATPase. Loss of the ability to generate energy (e.g., in a cell injured by toxins or ischemia) therefore results in osmotic swelling and eventual cell rupture. Similar transport mechanisms also regulate concentrations of other ions (e.g., Ca^{2+} and H^+). This is critical to many processes. For example, cytosolic enzymes are most active at pH 7.4 and are often regulated by Ca^{2+}, whereas lysosomal enzymes function best at pH 5 or less.

Uptake of fluids or macromolecules by the cell is called *endocytosis*. Depending on the size of the vesicle, endocytosis may be denoted pinocytosis ("cellular drinking") or phagocytosis ("cellular eating"). Generally, phagocytosis is restricted to certain cell types (e.g., macrophages and

neutrophils) whose role is to specifically ingest invading organisms or dead cell fragments.

Receptor-Mediated and Fluid-Phase Uptake (see Fig. 1.8)

Certain small molecules—including some vitamins—bind to cell-surface receptors and are taken up through invaginations of the plasma membrane called caveolae. Uptake can also occur through membrane invaginations coated by an intracellular matrix of clathrin proteins that spontaneously assemble into a basket-like lattice which helps drive endocytosis (discussed more later). In both cases, activity of the "pinchase" dynamin is required for vesicle release.

Macromolecules can also be exported from cells by *exocytosis*. In this process, proteins synthesized and packaged within the RER and Golgi apparatus are concentrated in secretory vesicles, which then fuse with the plasma membrane to expel their contents. Common examples include peptide hormones (e.g., insulin) and cytokines.

Transcytosis is the movement of endocytosed vesicles between the apical and basolateral compartments of cells. This is a mechanism for transferring large amounts of intact proteins across epithelial barriers (e.g., ingested antibodies in maternal milk) or for rapid movement of large solute volumes.

We now return to the specifics of endocytosis (see Fig. 1.8).

- *Caveolae-mediated endocytosis. Caveolae* ("little caves") are noncoated plasma membrane invaginations associated with GPI-linked molecules, cyclic adenosine monophosphate (cAMP) binding proteins, src-family kinases, and the folate receptor; *caveolin* is the major structural protein of caveolae, which, like membrane rafts (see above), are enriched in glycosphingolipids and cholesterol. Internalization of caveolae along with bound molecules and associated extracellular fluid is called *potocytosis*—literally "cellular sipping." In addition to supporting transmembrane delivery of some molecules (e.g., folate), caveolae regulate transmembrane signaling and cellular adhesion via internalization of receptors and integrins.
- *Receptor-mediated endocytosis.* Macromolecules bound to membrane receptors (such as transferrin or low-density lipoprotein [LDL] receptors) are taken up at specialized regions of the plasma membrane called *clathrin-coated pits*. The receptors are efficiently internalized by membrane invaginations driven by the associated clathrin matrix, eventually pinching off to form *clathrin-coated vesicles*. Trapped within these vesicles is also a gulp of the extracellular milieu *(fluid-phase pinocytosis)*. The vesicles then rapidly lose their clathrin coating and fuse with an acidic intracellular structure called the *early endosome;* the endosomal vesicles undergo progressive maturation to late endosomes, ultimately fusing with lysosomes. In the acidic environment of the endosomes, LDL and transferrin receptors release their cargo (cholesterol and iron, respectively), which is then transported into the cytosol.

After release of bound ligand, some receptors recycle to the plasma membrane and are reused (e.g., transferrin and LDL receptors), while others are degraded within lysosomes (e.g., epidermal growth factor receptor). In the latter case, degradation after internalization results in receptor downregulation that limits receptor-mediated signaling. Defects in receptor-mediated transport of LDL underlie familial hypercholesterolemia, as described in Chapter 5.

Endocytosis requires recycling of internalized vesicles back to the plasma membrane (exocytosis) for another round of ingestion. This is critical, as a cell will typically ingest from the extracellular space the equivalent of 10% to 20% of its own cell volume each hour—amounting to 1% to 2% of its plasma membrane each minute! Without recycling, the plasma membrane would be rapidly depleted. Endocytosis and exocytosis must therefore be tightly coupled to avoid large changes in plasma membrane area.

Cytoskeleton

The ability of cells to adopt a particular shape, maintain polarity, organize intracellular organelles, and migrate depends on an intracellular scaffold of structural proteins that form the cytoskeleton (Fig. 1.9). Although the cytoskeleton can have an appearance similar to the beams and supports of a building, cytoskeletal structures are constantly elongating and shrinking; microfilaments and microtubules are most active, and their assembly and disassembly can drive cell migration.

In eukaryotic cells, there are three major classes of cytoskeletal proteins.

- *Actin microfilaments* are 5- to 9-nm diameter fibrils formed from the globular protein actin (G-actin), the most

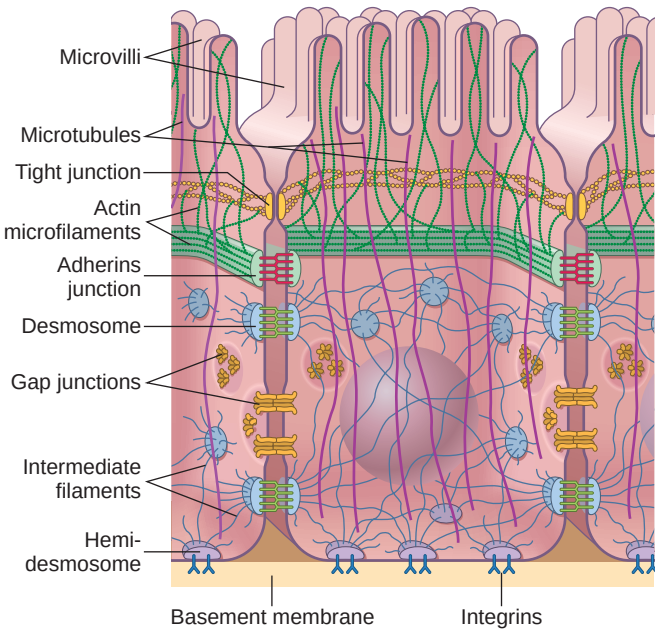


Figure 1.9 Cytoskeletal elements and cell-cell interactions. Interepithelial adhesion involves several different surface protein interactions at tight junctions, adherens junctions, and desmosomes; adhesion to the extracellular matrix involves cellular integrins (and associated proteins) within hemidesmosomes. The various adhesion proteins within the plasma membrane associate with actin microfilaments and intermediate filaments to provide a mechanical matrix for cell structure and signaling. Gap junctions do not impart structural integrity but allow cell-cell communication by the movement of small molecular weight and/or charged species. See text for details.

abundant cytosolic protein in cells. G-actin monomers noncovalently polymerize into long filaments (F-actin) that intertwine to form double-stranded helices with a defined polarity. Although the details are (as always) more nuanced, new subunits are typically added at the "positive" end of the strand and removed from the "negative" end—a process referred to as *actin treadmilling*. Actin nucleating, binding, and regulatory proteins organize polymerization, bundling, and branching to form networks that control cell shape and movement. This complex and its association with motor proteins (e.g., myosin) is so precisely arrayed in skeletal and cardiac muscle that a banding pattern is apparent by light microscopy. ATP hydrolysis by myosin slides the actin filaments relative to one another to cause muscle contraction. Although less coordinated, myosins, of which there are many, are responsible for other functions that depend on actin contraction including vesicular transport, epithelial barrier regulation, and cell migration.

- *Intermediate filaments* are 10-nm diameter fibrils that comprise a large and heterogeneous family that includes keratin proteins and nuclear lamins. Intermediate filaments predominantly form ropelike polymers and do not usually actively reorganize like actin and microtubules. This allows intermediate filaments to provide tensile strength so that cells can bear mechanical stress, e.g., in epithelia where intermediate filaments link *desmosomes* and *hemidesmosomes* (see Fig. 1.9). Individual intermediate filament proteins have characteristic tissue-specific patterns of expression that can be useful for assigning a cell of origin for poorly differentiated tumors. Examples include:
  - *Vimentin*, in mesenchymal cells (fibroblasts, endothelium).
  - *Desmin* in muscle cells forms the scaffold on which actin and myosin contract.
  - *Neurofilaments* are critical for neuronal axon structure and confer both strength and rigidity.
  - *Glial fibrillary acidic protein* is expressed in glial cells.
  - *Cytokeratins* are expressed in epithelial cells. There are at least 30 distinct different cytokeratins that are expressed in different cell lineages (e.g., lung vs. gastrointestinal epithelium).
  - *Lamins* are intermediate filament proteins that form the nuclear lamina, define nuclear shape, and can regulate transcription.
- *Microtubules* are 25-nm-thick fibrils composed of noncovalently polymerized α- and β-tubulin dimers organized into hollow tubes. These fibrils are extremely dynamic and polarized, with "+" and "–" ends. The "–" end is typically embedded in a *microtubule organizing center (MTOC* or *centrosome)* near the nucleus, where it is associated with paired centrioles; the "+" end elongates or recedes in response to various stimuli by the addition or subtraction of tubulin dimers. Microtubules:
  - Serve as mooring lines for molecular motor proteins that use ATP to translocate vesicles, organelles, or other molecules around cells. There are two varieties of these motor proteins, kinesins and dyneins, that typically (but not exclusively) transport cargo in anterograde (– to +) or retrograde (+ to –) directions, respectively.
  - Mediate sister chromatid segregation during mitosis.
  - Form the core of *primary cilia,* single nonmotile projections on many nucleated cells that contribute to the regulation of cellular proliferation and differentiation (mutations in the proteins of the primary cilia complex lead to forms of polycystic kidney disease; see Chapter 20).
  - Can be adapted to form the core of motile cilia (e.g., in bronchial epithelium) or flagella (in sperm).

Cell-Cell Interactions

Cells connect and communicate with each other via junctional complexes that form mechanical links and facilitate receptor-ligand interactions. Similar complexes also mediate interaction with the ECM. Cell-cell junctions are organized into three basic types (see Fig. 1.9):

- *Occluding junctions (tight junctions)* seal adjacent epithelial cells together to create a continuous barrier that restricts the paracellular (between cells) movement of ions and other molecules. Occluding junctions form a tight meshlike network (when viewed en face by freeze-fracture electron microscopy) of macromolecular contacts between neighboring cells; the complexes that mediate the cell-cell interactions are composed of transmembrane proteins including the tetraspanning *claudin* and *tight junction–associated MARVEL protein (TAMP)* families. These connect to a host of intracellular adaptor and scaffolding proteins, including the three members of the *zonula occludens protein family* (ZO-1, ZO-2, ZO-3) and cingulin. Besides forming a selectively permeable barrier that seals the space between cells (i.e., the paracellular space), this zone also represents the boundary that separates apical and basolateral membrane domains and helps to maintain cellular polarity. Nevertheless, these junctions are dynamic structures that can be modified to facilitate epithelial healing and inflammatory cell migration across epithelial-lined mucosal surfaces.
- *Anchoring junctions (adherens junctions* and *desmosomes)* mechanically attach cells—and their cytoskeletons—to other cells or the ECM. Adherens junctions are often closely associated with and beneath tight junctions. Desmosomes are more basal and form several types of junctions. When desmosomes attach the cell to the extracellular matrix (ECM) they are referred to as hemidesmosomes (half a desmosome), since the other half of the desmosome is not present within the ECM. Both adherens junctions and desmosomes are formed by homotypic extracellular interactions between transmembrane glycoproteins called cadherins, on adjacent cells.
  - In adherens junctions the transmembrane adhesion molecules are associated with intracellular actin microfilaments through which they can also influence cell shape and/or motility. Loss of the epithelial adherens junction protein E-cadherin explains the discohesive invasion pattern of some gastric cancers and lobular carcinomas of the breast (Chapters 17 and 23).
  - In desmosomes the cadherins are linked to intracellular intermediate filaments, allowing extracellular forces to be mechanically communicated (and dissipated) over multiple cells.

- In hemidesmosomes the transmembrane connector proteins are called *integrins;* like desmosomal cadherins, these attach to intermediate filaments and link the cytoskeleton to the ECM. *Focal adhesion complexes* are composed of >100 proteins and localize at hemidesmosomes. Their component proteins can generate intracellular signals when cells are subjected to shear stress (e.g., endothelium in the bloodstream or cardiac myocytes in a failing heart).

- *Communicating junctions (gap junctions)* permit the diffusion of chemical or electrical signals from one cell to another. The junction consists of a dense planar array of 1.5- to 2-nm pores (called *connexons*) formed by a pair of hexamers (one on each cell) of transmembrane connexin proteins. These form pores that permit passage of ions, nucleotides, sugars, amino acids, vitamins, and other small molecules; permeability of the junction is rapidly reduced by lowered intracellular pH or increased intracellular calcium. Gap junctions play a critical role in cell-cell communication. For example, gap junctions in cardiac myocytes allow cell-to-cell calcium fluxes that allow the many cells of the myocardium to behave as a functional syncytium with coordinated waves of contraction.

Biosynthetic Machinery: Endoplasmic Reticulum and Golgi Apparatus

All cellular constituents—including structural proteins, enzymes, transcription factors, and even the phospholipid membranes—are constantly renewed in an ongoing process balancing synthesis and degradation. The endoplasmic reticulum (ER) is the site for synthesis of all transmembrane proteins and lipids for plasma membrane and cellular organelles, including the ER itself. It is also the initial site of synthesis for secreted proteins. The ER is organized into a meshlike interconnected maze of branching tubes and flattened lamellae, forming a continuous sheet around a single lumen that is topologically equivalent to the extracellular environment. ER is composed of contiguous but distinct domains that are distinguished by the presence (RER) or absence (SER) of ribosomes (see Fig. 1.6).

- *Rough endoplasmic reticulum (RER):* Membrane-bound ribosomes on the cytosolic face of RER translate mRNA into proteins that are extruded into the ER lumen or become integrated into the ER membrane. This process is directed by specific signal sequences on the N-termini of nascent proteins; synthesis of proteins with signal peptides is initiated on free ribosomes, but the complex then becomes attached to the ER membrane, and the protein is inserted into or passed across the ER membrane as it is translated. For proteins lacking a signal sequence, translation remains on free ribosomes in the cytosol, forming polyribosomes as multiple ribosomes attach to the mRNA; such transcribed proteins remain within the cytoplasm.

  Proteins inserted into the ER fold into their active conformation and can form polypeptide complexes (oligomerize); in addition, disulfide bonds are formed, and N-linked oligosaccharides (sugar moieties attached to asparagine residues) are added. Chaperone molecules assist in folding and retaining proteins in the ER until the modifications are complete and the proper conformation is achieved. If a protein fails to appropriately fold or oligomerize, it is retained and degraded within the ER. A good example of this is the most common mutation of the CFTR protein in cystic fibrosis. In mutant CFTR, a codon deletion leads to the absence of a single amino acid (Phe508) which results in its misfolding, ER retention and catabolism and therefore reduced surface expression. Moreover, excess accumulation of misfolded proteins—exceeding the capacity of the ER to edit and degrade them—leads to the *ER stress response* (also called the *unfolded protein response [UPR]*) (Fig. 1.10B). Detection of excess abnormally folded proteins leads to a reduction in protein synthesis overall with a concurrent increase in chaperone proteins; failure to correct the overload can trigger cell death through *apoptosis* (Chapter 2).

- *Golgi apparatus:* From the RER, proteins and lipids destined for other organelles or extracellular export are shuttled into the Golgi apparatus. This consists of stacked cisternae that progressively modify proteins in an orderly fashion from *cis* (near the ER) to *trans* (near the plasma membrane). Cisternal progression, i.e., movement of *cis*-face cisternae to the *trans* aspect of the Golgi, akin to steps on an escalator, allows sequential processing of newly synthesized proteins and can be facilitated by small membrane-bound vesicles. Similar vesicles shuttle Golgi-resident enzymes from *trans* to *cis* in order to maintain the different cisternal contents along this assembly line. As cisternae mature, the N-linked oligosaccharides originally added in the ER are pruned and extended in a stepwise fashion; O-linked oligosaccharides (sugar moieties linked to serine or threonine) are also appended. Some of this glycosylation is important in sorting molecules to lysosomes (via the mannose-6-phosphate receptor); other glycosylation adducts are important for cell-cell or cell-matrix interactions or for clearing senescent cells (e.g., platelets and erythrocytes). In the *trans Golgi network,* proteins and lipids are sorted and dispatched to other organelles, plasma membrane, or secretory vesicles. The Golgi complex is especially prominent in cells specialized for secretion, including goblet cells of the intestinal or bronchial epithelium, which secrete large amounts of polysaccharide-rich mucus. In plasma cells that secrete antibodies, the Golgi can be recognized as a perinuclear hoff on simple hematoxylin and eosin stains (Chapter 6).

- *Smooth endoplasmic reticulum (SER):* In most cells, the SER is relatively sparse and primarily exists as the transition zone extending from RER to generate transport vesicles that carry newly synthesized proteins to the Golgi apparatus. The SER may, however, be particularly conspicuous in cells that synthesize steroid hormones (e.g., within the gonads or adrenals) or that catabolize lipid-soluble molecules (e.g., hepatocytes). Indeed, repeated exposure to compounds that are metabolized by the SER (e.g., phenobarbital, which is catabolized by the cytochrome P-450 system) can lead to SER hyperplasia. The SER is also responsible for sequestering intracellular calcium, which, when released into the cytosol, can mediate a number of responses to extracellular signals (including apoptotic cell death). In muscle cells, specialized SER called sarcoplasmic reticulum is responsible for the cyclic release and sequestration of calcium ions that regulate muscle contraction and relaxation, respectively.

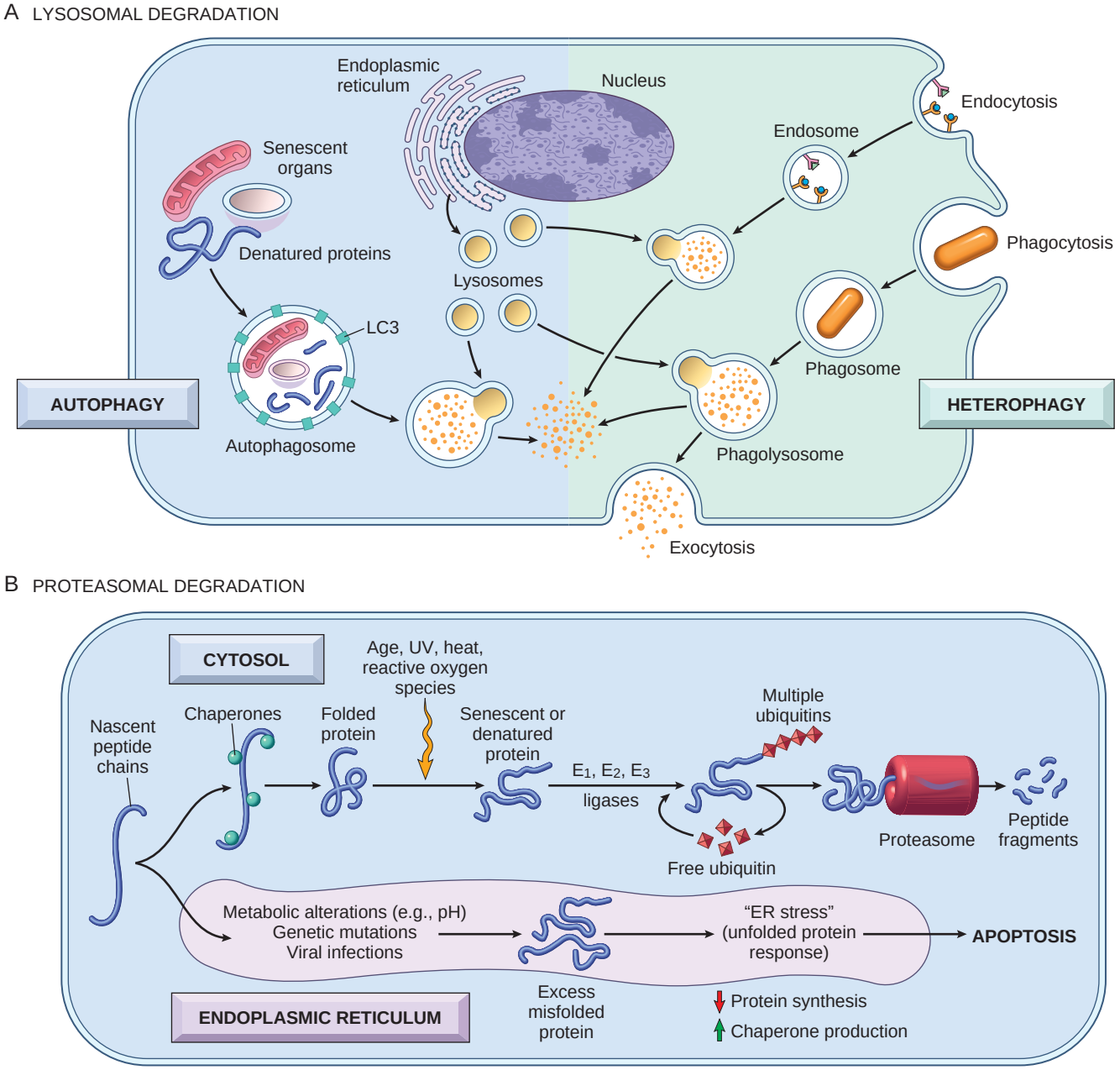


Figure 1.10 Intracellular catabolism. (A) Lysosomal degradation. In heterophagy (right side of panel A), lysosomes fuse with endosomes or phagosomes to facilitate the degradation of their internalized contents (see Fig. 1.8). The end products may be released into the cytosol for nutrition or discharged into the extracellular space (exocytosis). In autophagy (left side of panel A), senescent organelles or denatured proteins are targeted for lysosome-driven degradation as they are encircled within a double membrane vacuole derived from the endoplasmic reticulum and marked by LC3 protein (microtubule-associated protein 1A/1B-light chain 3). Cell stress, such as nutrient depletion or some intracellular infections, can also activate the autophagocytic pathway. (B) Proteosomal degradation. Cytosolic proteins destined for turnover (e.g., transcription factors or regulatory proteins), senescent proteins, or proteins that have become denatured due to extrinsic mechanical or chemical stresses can be tagged by multiple ubiquitin molecules (through the activity of E_1, E_2, and E_3 ubiquitin ligases). This marks the proteins for degradation by proteasomes, cytosolic multi-subunit complexes that degrade proteins to small peptide fragments. High levels of misfolded proteins within the endoplasmic reticulum (ER) trigger a protective unfolded protein response—engendering a broad reduction in protein synthesis, but specific increases in chaperone proteins that can facilitate protein refolding. If this is inadequate to cope with the levels of misfolded proteins it can lead to apoptosis.

Waste Disposal: Lysosomes and Proteasomes

Although cells rely primarily on lysosomes to digest internalized material and accumulated internal waste, there are multiple other routes to degrade intracellular macromolecules (see Fig. 1.10). These include proteosomes and peroxisomes. The latter are responsible for catabolism of long-chained fatty acids.

- *Lysosomes* are membrane-bound organelles containing roughly 40 different acid hydrolases (i.e., that function best at pH ≤5); these include proteases, nucleases, lipases, glycosidases, phosphatases, and sulfatases. Lysosomal

enzymes are initially synthesized in the ER lumen and then tagged with mannose-6-phosphate (M6P) within the Golgi apparatus. These M6P-modified proteins are subsequently delivered to lysosomes through *trans* Golgi vesicles that express M6P receptors. The other macromolecules destined for catabolism in the lysosomes arrive by one of three pathways (see Fig. 1.10).

- Material internalized by *fluid-phase* or *receptor-mediated endocytosis* passes from plasma membrane to early and then late endosomes and ultimately arrives at the lysosome. These compartments are progressively acidified such that proteolytic enzymes become active in late endosome and lysosomes.
- Senescent organelles and/or large, denatured protein complexes can be ferried into lysosomes by a process called *autophagy* (Chapter 2). Through a mechanism involving the products of a number of autophagy-related *(Atg)* genes, obsolete organelles are corralled by a double membrane derived from the ER. The membrane progressively expands to encircle a collection of organelles and cytosolic constituents, forming the definitive *autophagosome;* these structures are targeted for eventual destruction by fusion with lysosomes. In addition to facilitating turnover of aged and/or defunct structures, autophagy can be used to preserve viability during nutrient depletion; is involved in protective responses to intracellular infections; participates in intracellular repair; and, under some circumstances, triggers programmed cell death *(apoptosis)*. Autophagy is discussed in more detail in Chapter 2.
- *Phagocytosis* of microorganisms or large fragments of matrix or debris occurs primarily in professional phagocytes (macrophages or neutrophils). The material is engulfed to form a *phagosome* that subsequently fuses with lysosomes.

- *Proteasomes* play an important role in degrading cytosolic proteins (see Fig. 1.10); these include denatured or misfolded proteins as well as any other macromolecule whose lifespan needs to be regulated (e.g., signaling molecules). Many proteins destined for destruction are identified by covalently binding a small protein called *ubiquitin*. Polyubiquitinated molecules are unfolded and funneled into the polymeric proteasome complex—a cylinder containing multiple protease activities, each with its active site pointed at the hollow core. Proteasomes digest proteins into small (6 to 12 amino acids) fragments that can subsequently be degraded to their constituent amino acids and recycled.

CELLULAR METABOLISM AND MITOCHONDRIAL FUNCTION

Mitochondria evolved from ancestral prokaryotes that were engulfed by primitive eukaryotes about 1.5 billion years ago. Their origin explains why mitochondria contain their own DNA that encodes about 1% of total cellular protein and approximately 20% of the proteins involved in oxidative phosphorylation. Although their genomes are small, mitochondria can carry out all the steps of DNA replication, transcription, and translation using machinery similar to present-day bacteria; for example, mitochondria initiate protein synthesis with N-formylmethionine and are sensitive to some antibacterial antibiotics. Since mitochondria biogenesis requires a genetic contribution from preexisting mitochondria and the ovum contributes the vast majority of cytoplasmic organelles in the fertilized zygote, mitochondrial DNA is almost entirely maternally inherited. Nevertheless, because the protein constituents of mitochondria are encoded by both nuclear and mitochondrial DNA, mitochondrial disorders may be X-linked, autosomal, or maternally inherited.

Mitochondria are impressively dynamic, constantly undergoing fission and fusion with other newly synthesized mitochondria; this supports their renewal and defends against degenerative changes that occur through ongoing oxygen free radical damage. Even so, mitochondria are short-lived—being degraded through autophagy (a process called mitophagy)—with estimated half-lives of 1 to 10 days, depending on the tissue, nutritional status, metabolic demands, and intercurrent injury.

Besides providing the enzymatic machinery for oxidative phosphorylation (and thus the relatively efficient generation of energy from glucose and fatty acid substrates), mitochondria play a fundamental role in regulating apoptosis (Fig. 1.11). Details of mitochondrial functions follow:

- *Energy generation.* Each mitochondrion has two separate membranes with distinct functions; the inner membrane contains the enzymes of the respiratory chain folded into cristae. This encloses a core matrix space that harbors the bulk of the enzymes of the glycolytic and tricarboxylic acid cycles, and these constitute the major working parts of the organelle. Outside the inner membrane is the intermembrane space, which is the site of nucleotide phosphorylation and is, in turn, enclosed by the outer membrane; the latter is studded with porin proteins that form voltage-dependent anion channels that are permeable to small (<5000 Da) molecules. As with other cellular membranes, larger molecules (and even some smaller polar species) require specific transporters.

  The major source of the energy that facilitates all cellular functions derives from oxidative metabolism. Mitochondria oxidize substrates to CO_2, transferring the high-energy electrons from the original molecule (e.g., sugar) to molecular oxygen. Oxidation of various metabolites drives proton pumps that transfer H^+ from the core matrix into the intermembrane space. As the H^+ ions flow down their electrochemical gradient and out of the intermembrane space, the energy released is used to generate ATP.

  Notably, the electron transport chain need not necessarily be coupled to ATP generation. An inner membrane protein enriched in brown fat called *thermogenin* (also called uncoupling protein-1 [UCP-1]) is a hydrogen ion transporter that can dissipate the proton gradient, uncoupling it from oxidative phosphorylation. By this means, there is rapid substrate oxidation without ATP synthesis that allows tissues with high levels of UCP-1 to generate heat (nonshivering thermogenesis). As a natural (albeit usually low-level) by-product of substrate oxidation and electron transport, mitochondria are also an important source of reactive oxygen species (oxygen free radicals, hydrogen peroxide, etc.); importantly, hypoxia, toxic injury, or even mitochondrial aging can

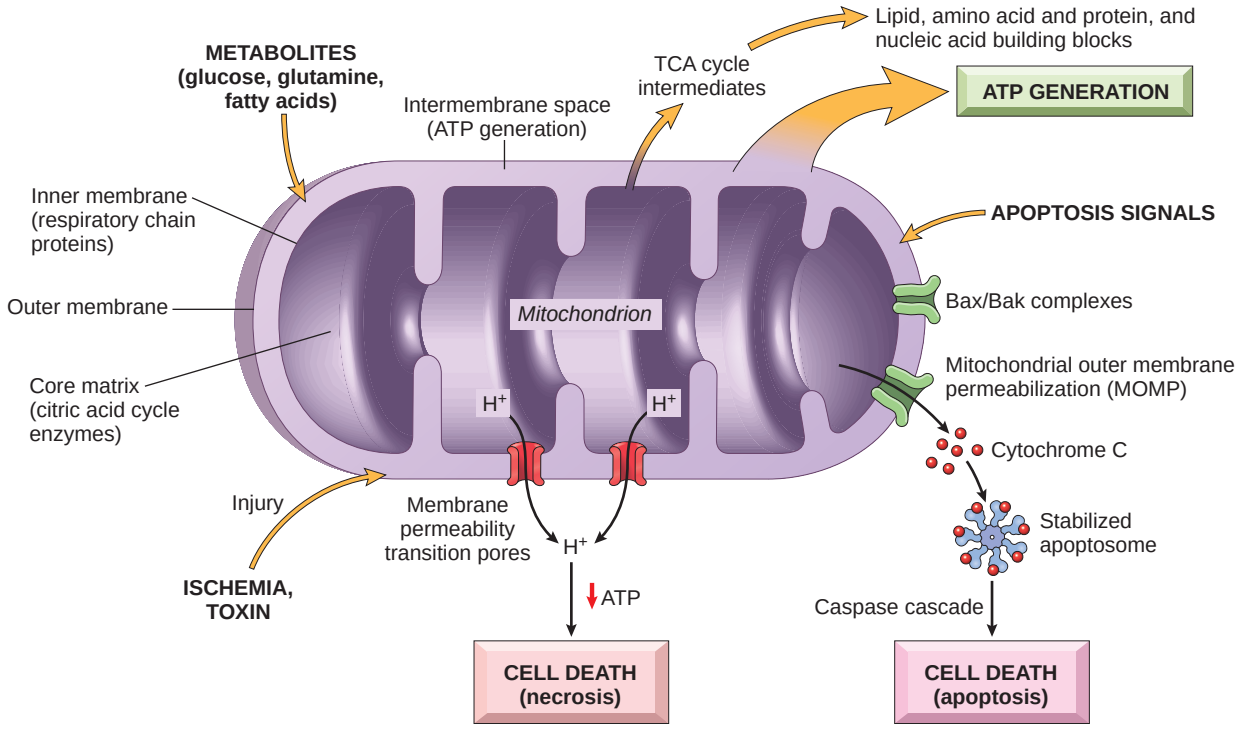


Figure 1.11 Roles of the mitochondria. Besides the efficient generation of adenosine triphosphate (ATP) from carbohydrate and fatty acid substrates, mitochondria also play an important role in intermediary metabolism—serving as the source of substrates used to synthesize lipids, proteins, and nucleotides. Significantly, they are centrally involved in cell life-and-death decisions: (1) toxic or ischemic injury will induce a membrane permeability transition that dissipates the intermembrane proton gradient, leading to cell death through loss of ATP generation; (2) intracellular signaling from both intrinsic and extrinsic sources can also result in the formation of oligomerized Bax and Bak protein pores on mitochondrial outer membrane permeabilization (MOMP) that facilitate release of cytochrome C from the electron transport chain proteins. Cytosolic cytochrome C stabilizes the multisubunit apoptosome to promote caspase activation and ultimately apoptotic cell death. TCA, Tricarboxylic acid.

lead to oxidative stress, characterized by increases in intracellular reactive oxygen species.

- *Intermediate metabolism.* Oxidative phosphorylation efficiently generates 36 to 38 ATP molecules per glucose molecule, but also "burns" substrates to their core CO_2 and H_2O, leaving no carbon moieties to use for building lipids and proteins. Consequently, to ensure the necessary building blocks for growth, rapidly dividing cells (both benign and malignant) increase their uptake of glucose and glutamine and switch to aerobic glycolysis, a phenomenon called the *Warburg effect*. In that setting, each glucose molecule is catabolized to lactic acid (even in the presence of adequate oxygen), generating only two net ATP molecules but "spinning off" intermediates that can be converted into new lipids, amino acids and proteins, and nucleic acids. Thus mitochondrial metabolism can be modulated to support either cellular energy maintenance or cellular proliferation.
- *Cell death.* In addition to providing the ATP that enables the bulk of cellular activity, mitochondria are fundamental to cell survival. The role of mitochondria in cell death is detailed in Chapter 2 and briefly mentioned here.
  - *Necrosis:* External cellular injury (toxin, ischemia, trauma) can damage mitochondria, inducing the formation of mitochondrial permeability transition pores in the outer membrane. These channels allow the dissipation of the proton gradient so that subsequent mitochondrial ATP generation becomes impossible and the cell dies.
  - *Apoptosis:* Programmed cell death is a central feature of normal tissue development and turnover and can be triggered by extrinsic signals (including cytotoxic T cells or inflammatory cytokines) or intrinsic pathways (including DNA damage or intracellular stress). Mitochondria integrate intracellular proapoptotic and antiapoptotic effector signals to generate a final "go" or "no go" signal for apoptosis. A net proapoptotic signal results in mitochondrial outer membrane permeabilization (MOMP)—distinct from the mitochondrial permeability transition—that releases cytochrome C (and other proteins) into the cytoplasm. In turn, these proteins activate intracellular programmed cell death pathways. Notably, a failure of normal proapoptotic signaling (or too many antiapoptotic effectors) can underlie malignancy—even in the face of mutations that would otherwise trigger cellular suicide. Conversely, too strong an apoptotic signal (or lack of antiapoptotic effectors) may lead to premature cell death, as in neurodegenerative disorders (Chapter 28).

CELLULAR ACTIVATION

Intercellular communication is essential to multicellular organisms. At the most basic level, extracellular signals may determine whether a cell lives or dies, remains quiescent,

or becomes active to perform its specific functions. Intercellular signaling is critical in the developing embryo and in maintaining tissue organization; it is also important in adult organisms, where intercellular signaling assures that all tissues act in appropriate concert (e.g., in response to local tissue trauma or a systemic infection). Loss of intercellular communication and "social controls" can also be reflected in unregulated growth (e.g., cancer) or in detrimental responses to extrinsic stress (e.g., shock).

Cell Signaling

Individual cells are chronically exposed to a remarkable cacophony of signals that must be integrated into rational output; some signals may induce a given cell type to differentiate, while others signal proliferation or direct the cell to perform specialized functions. Multiple signals at once, in a certain ratio, may trigger yet another totally unique response. Many cells require certain input just to continue living; in the absence of appropriate exogenous signals, they die by apoptosis.

The signals that most cells respond to can be classified into several groups.

- *Danger and pathogens.* Many cells have an innate capacity to sense and respond to damaged cells (danger signals), as well as foreign invaders such as microbes. The involved receptors are discussed in Chapters 3 and 6.
- *Cell-cell contacts,* mediated through adhesion molecules and/or gap junctions. As mentioned previously, gap junction signaling is accomplished between adjacent cells via hydrophilic connexon channels that permit the movement of small ions (e.g., calcium), metabolites, and second messenger molecules (e.g., cAMP).
- *Cell-ECM contacts,* mediated through integrins. We will return to a consideration of integrins in the context of leukocyte attachment to other cells during inflammation in Chapter 3.
- *Secreted molecules.* The most important secreted molecules include growth factors (discussed later); cytokines, a term reserved for mediators of inflammation and immune responses (discussed also in Chapters 3 and 6); and hormones, which are secreted by endocrine organs (Chapter 24).

Signaling pathways can also be classified based on the spatial relationships between the sending and receiving cells.

- *Paracrine signaling.* Only cells in the immediate vicinity are affected. To accomplish this, there can be only minimal diffusion, after which the secreted signal is rapidly degraded, taken up by other cells, or trapped in the ECM.
- *Autocrine signaling* occurs when molecules secreted by a cell affect that same cell. This can be a means to entrain groups of cells undergoing synchronous differentiation during development, or it can be used to amplify (positive feedback) or dampen (negative feedback) a particular response.
- *Synaptic signaling.* Activated neurons secrete neurotransmitters at specialized cell junctions (i.e., synapses) onto target cells.
- *Endocrine signaling.* A mediator is released into the bloodstream and acts on target cells at a distance.

Regardless of the nature of an extracellular stimulus (paracrine, autocrine, synaptic, or endocrine), the signal that is conveyed is transmitted to the cell via specific receptor proteins. Signaling molecules (ligands) bind their respective receptors and initiate a cascade of intracellular events culminating in the desired cellular response. Ligands usually have high affinities for their receptors ($\leq 10^{-8}$ M) and, at physiologic concentrations, bind receptors with exquisite specificity. Receptors may be present on the cell surface or within the cell (Fig. 1.12).

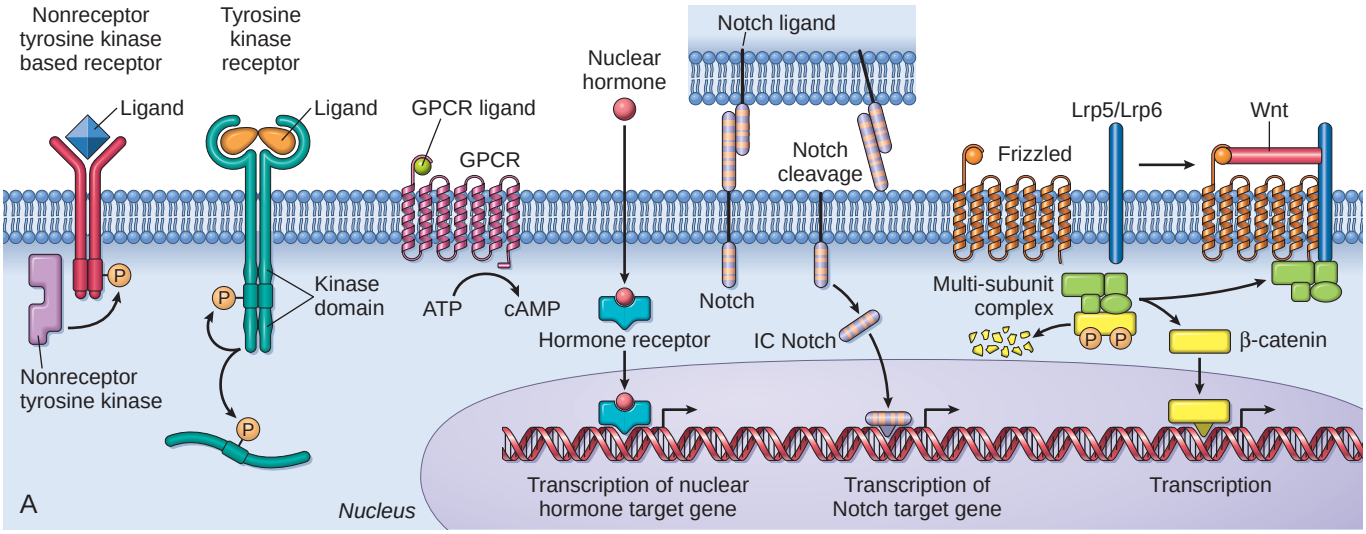
- *Intracellular receptors* include transcription factors that are activated by lipid-soluble ligands that can easily transit plasma membranes; vitamin D and steroid hormones that activate nuclear hormone receptors are good examples. In other cases a small and/or nonpolar signaling ligand produced by one cell type can influence the activity of adjacent cells. Thus nitric oxide produced by endothelial cells diffuses into underlying smooth muscle cells, activating the enzyme guanylyl cyclase to generate cyclic guanosine monophosphate (cGMP) that causes smooth muscle relaxation; endothelium can thereby regulate vasomotor tone.
- *Cell-surface receptors* are generally transmembrane proteins with extracellular domains that bind activating ligands. Depending on the receptor, ligand binding can:
  1. Open ion channels, typically at the synapse between electrically excitable cells.
  2. Activate an associated GTP-binding regulatory protein (G protein).
  3. Activate an endogenous or associated enzyme (often a tyrosine kinase).
  4. Trigger a proteolytic event or change protein binding or stability to activate a latent transcription factor.

G protein–coupled receptors and tyrosine kinase–associated receptors are typically involved in signaling that drives cellular proliferation; proteolytic or conformation changes are common features of multiple pathways (e.g., Notch, Wnt, and Hedgehog) that influence normal development. Secondary intracellular downstream events frequently involve phosphorylation or dephosphorylation of target molecules, with subsequent conformational changes that impact nuclear access or enzymatic activity (see later). Understandably, signals transduced by cell-surface receptors are frequently perturbed in developmental disorders and malignancy.

Signal Transduction Pathways

The interaction of a cell-surface receptor and its ligand can activate signaling through ligand-induced clustering of the receptor (receptor cross-linking) or by inducing a physical change in receptor structure (see Fig. 1.12). Either mechanism results in a conformational change in the cytosolic tail of the receptor that mediates additional intracellular biochemical events.

Cellular receptors are grouped into several types based on the signaling mechanisms they use and the intracellular biochemical pathways they activate (see Fig. 1.12). Receptor signaling most commonly leads to the formation or modification of biochemical intermediates and/or activation of



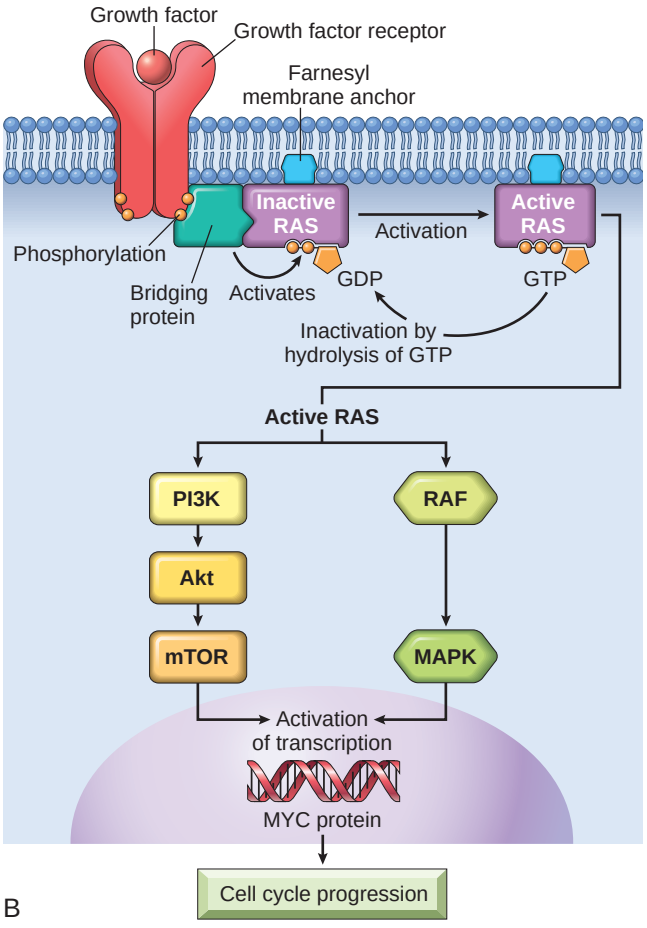


Figure 1.12 Receptor-mediated signaling. (A) Categories of signaling receptors including (from left to right) receptors that utilize a nonreceptor tyrosine kinase; a receptor tyrosine kinase; a G-protein coupled receptor (GPCR) seven-transmembrane protein linked to heterotrimeric G proteins; a nuclear receptor that binds its ligand and can then influence transcription; Notch, which recognizes a ligand on a distinct cell and is cleaved yielding an intracellular Notch (IC Notch) fragment that can enter the nucleus and influence transcription of specific target genes; and the Wnt/Frizzled pathway, where activation releases intracellular β-catenin from a protein complex that normally drives its constitutive degradation. The released β-catenin can then migrate to the nucleus and act as a transcription factor. (B) Signaling from a tyrosine kinase–based growth factor receptor. Binding of the growth factor (ligand) causes receptor dimerization and autophosphorylation of tyrosine residues. Attachment of adaptor (or bridging) proteins couples the receptor to inactive guanosine diphosphate (GDP)–bound RAS, allowing the GDP to be displaced in favor of guanosine triphosphate (GTP) and yielding activated RAS. Activated RAS interacts with and activates RAF (also known as MAP kinase kinase kinase). This kinase then phosphorylates mitogen-activated protein kinases *(MAPK)*, which, in turn, phosphorylates other cytoplasmic proteins and nuclear transcription factors that culminate in cellular responses. The phosphorylated tyrosine kinase receptor can also bind other proteins such as phosphatidylinositol 3-kinase (PI3K), which activates additional signaling pathways. The cascade is inactivated when the activated RAS eventually hydrolyzes GTP to GDP—converting RAS to its inactive form. Mutations in RAS that cause delayed GTP hydrolysis can thus lead to augmented proliferative signaling. *ATP,* Adenosine triphosphate; cAMP, cyclic adenosine monophosphate; Lrp5/Lrp6, low-density lipoprotein receptor–related proteins 5 and 6; mTOR, mammalian target of rapamycin.

enzymes that ultimately lead to generation of active transcription factors that enter the nucleus and alter the pattern of gene expression.

- *Receptors associated with kinase activity.* Downstream phosphorylation is a common pathway of signal transduction. Changes in receptor geometry can stimulate intrinsic receptor protein kinase activity or promote the enzymatic activity of recruited intracellular kinases; these kinases add charged phosphate residues to target molecules. Tyrosine kinases phosphorylate specific tyrosine residues, whereas serine/threonine kinases add phosphates to serine or threonine residues, and lipid kinases phosphorylate lipid substrates. For every phosphorylation event, there is also a phosphatase that removes phosphate residues to modulate signaling; phosphatases usually inhibit signal transduction.

- *Receptor tyrosine kinases (RTKs)* are integral membrane proteins (e.g., receptors for insulin, epidermal growth factor, and platelet-derived growth factor); ligand-induced cross-linking activates intrinsic tyrosine kinase domains located in their cytoplasmic tails.
- Several kinds of receptors have no intrinsic catalytic activity (e.g., immune receptors, some cytokine receptors, and integrins). For these, a separate intracellular protein—known as a *nonreceptor tyrosine* kinase—interacts with receptors after ligand binding and phosphorylates specific motifs on the receptor or other proteins. The cellular homologue of the transforming protein of the Rous sarcoma virus, called Src, is the prototype for an important family of such nonreceptor tyrosine kinases *(Src-family kinases).* Src contains unique functional regions called Src homology (SH) domains; SH2 domains typically bind to receptors phosphorylated by another kinase, allowing the aggregation of multiple enzymes, whereas SH3 domains mediate other protein-protein interactions, often involving proline-rich sequences.

- *G-protein coupled receptors (GPCRs)* characteristically traverse the plasma membrane seven times (hence their designation as seven-transmembrane or serpentine receptors); more than 1500 such receptors have been identified. After ligand binding, the GPCR associates with an intracellular guanosine triphosphate (GTP)–binding protein (G protein) that contains guanosine diphosphate (GDP). G-protein interaction with a GPCR-ligand complex results in activation through the exchange of GDP for GTP. Subsequent GPCR-mediated signaling pathways include the generation of cAMP and inositol-1,4,5-triphosphate (IP_3), with the latter triggering release of calcium from the ER.
- *Nuclear receptors*. Lipid-soluble ligands can diffuse into cells where they interact with intracellular proteins to form a receptor-ligand complex that directly binds to nuclear DNA; the results can be either activation or repression of gene transcription.
- *Other classes of receptors*. Other receptors—originally recognized as important for embryonic development and cell fate determination—are now recognized to participate in the functioning of mature cells, particularly within the immune system. Rather than enzymatic activity, these pathways rely on protein-protein interactions to transduce signals.
  - Receptor proteins of the *Notch* family fall in this category; ligand binding to Notch receptors leads to proteolytic cleavage of the receptor and subsequent nuclear translocation of the cytoplasmic piece (intracellular Notch) to form a transcription complex.
  - *Wnt* protein ligands can also influence cell development through a canonical pathway involving transmembrane *Frizzled* family receptors, a distinct set of GPCRs that regulate intracellular levels of *β-catenin*. Normally, β-catenin is continuously targeted for ubiquitin-directed proteasomal degradation. However, Wnt binding to frizzled (and other coreceptors) recruits yet another intracellular protein *(Dishevelled)* that leads to disruption of the degradation-targeting complex. This stabilizes β-catenin, allowing it to translocate to the nucleus and form a transcription complex.

Modular Signaling Proteins, Hubs, and Nodes

The traditional linear view of signaling—that receptor activation triggers an orderly sequence of biochemical intermediates that ultimately leads to changes in gene expression and the desired biological response—is oversimplified. Instead, it is increasingly clear that any initial signal impacts multiple processes, each of which contributes to the final outcome. This is particularly true of signaling pathways that rely on enzymatic activity. For example, specific phosphorylation of any given protein can allow it to associate with a host of other molecules, resulting in (among other effects):

- Enzyme activation (or inactivation).
- Nuclear (or cytoplasmic) localization of transcription factors (see later).
- Transcription factor activation (or inactivation).
- Actin polymerization (or depolymerization).
- Protein degradation (or stabilization).
- Activation of feedback inhibitory (or stimulatory) loops.

Adaptor proteins play a key role in organizing intracellular signaling pathways. These proteins function as molecular connectors that physically link different enzymes and promote the assembly of complexes; adaptors can be integral membrane proteins or cytosolic proteins. A typical adaptor contains specific domains (e.g., SH2 or SH3) that mediate protein-protein interactions. By influencing which proteins are recruited to signaling complexes, adaptors can determine downstream signaling events.

By analogy with computer networks, the protein-protein complexes can be considered nodes, and the biochemical events feeding into or emanating from these nodes can be thought of as hubs. Signal transduction can therefore be visualized as a kind of networking phenomenon; understanding this higher-order complexity is the province of *systems biology,* involving a melding of biology and computer science, i.e., computational biology.

Transcription Factors

Most signal transduction pathways ultimately induce durable effects on cellular function by modulating gene transcription; this occurs through the activation and/or nuclear localization of transcription factors. Some transcription factors drive expression of a relatively limited set of genes or a specific genetic program, while others have widespread effects. Among the transcription factors that regulate cell division are products of several growth-promoting genes, such as *MYC* and *JUN*, and of cell cycle–inhibiting genes, such as *TP53*. Transcription factors often contain modular domains that bind to DNA, small molecules such as steroid hormones, and intracellular regulatory proteins. Interactions mediated by these domains can be controlled by posttranslational modifications such as phosphorylation. These changes can result in translocation from the cytoplasm into the nucleus, modify transcription factor protein half-life, expose specific DNA binding motifs, or promote binding to components of the RNA polymerase complex to augment transcription factor activity.

- *DNA-binding domains* permit specific binding to short DNA sequences. Whereas some transcription factor

binding sites are found in promoters near the location of transcription initiation, other transcription factor binding sites can be found throughout the genome; in the latter case, transcription factor activation may lead to the simultaneous transcription of a cassette of genes (presumably interrelated and interacting). Transcription factors may also bind to long-range regulatory elements such as enhancers that function by bringing gene promoters into geographic proximity to the genes they regulate. The fact that these sites may be distant from one another based on the linear genetic sequence emphasizes the importance of chromatin organization in regulating gene expression.

- *Protein-protein interaction domains* within transcription factors directly or indirectly recruit additional proteins including coactivators, histone-modifying enzymes, and chromatin-remodeling complexes that unwind and/or otherwise expose initiation sites. Most importantly, they recruit RNA polymerase—the large multiprotein enzymatic complex that is responsible for RNA synthesis.

GROWTH FACTORS AND RECEPTORS

Growth factors stimulate the activity of signaling pathways and genes that augment cell survival, growth, and division. Growth factors bind to specific receptors and, ultimately, influence expression of genes that:

- Promote entry into the cell cycle.
- Relieve blocks on cell cycle progression (thus promoting replication).
- Prevent apoptosis.
- Enhance synthesis of components (nucleic acids, proteins, lipids, carbohydrates) required for cell division.

Although growth factors are characteristically thought of as proteins that "just" stimulate cell proliferation and/or survival, it is important to remember that they can also regulate a host of nongrowth activities including migration, differentiation, and synthetic capacity. Some growth factors relevant to tissue repair and regeneration are listed in Table 1.1 and described further in Chapter 3.

Growth factors regulate cellular proliferation at steady state and in response to injury, when irreversibly damaged cells must be replaced. Uncontrolled proliferation can result when the growth factor activity is dysregulated or when growth factor signaling pathways are altered to become constitutively active. Thus many growth factor pathway genes are *proto-oncogenes;* by virtue of their proliferative effects, gain-of-function mutations convert them into *oncogenes* that lead to unfettered cell division and can be precursors to malignancy. Although the growth factors described here all involve receptors with intrinsic kinase activity, growth factors can signal through each of the pathways shown in Fig. 1.12.

Epidermal Growth Factor (EGF) and Transforming Growth Factor-α (TGF-α). Both factors belong to the EGF family, bind to overlapping sets of receptors, and share many biologic activities. EGF and TGF-α, which are produced by macrophages and some epithelial cells, are mitogenic for hepatocytes, fibroblasts, and a host of epithelial cell types. The "EGF receptor family" includes four membrane receptors with intrinsic tyrosine kinase activity; the best-characterized receptor is EGFR1, also known as ERB-B1, or simply EGFR. EGFR1 mutations and/or amplification frequently occur in a number of cancers including lung, head and neck, breast, and brain. The ERB-B2 receptor (also known as *HER-2*) is overexpressed in a subset of breast cancers. Antibodies and small molecule antagonists that target many of these receptors have proven effective in some cancers.

Hepatocyte Growth Factor (HGF). HGF (also known as scatter factor) has mitogenic effects on hepatocytes and most epithelium including biliary, lung, kidney, breast, and skin.

Table 1.1 Growth Factors Involved in Regeneration and Repair

Growth Factor	Sources	Functions
Epidermal growth factor (EGF)	Activated macrophages, salivary glands, keratinocytes, many other cells	Mitogenic for many cell types; stimulates epithelial cell migration; stimulates formation of granulation tissue
Transforming growth factor-α (TGF-α)	Activated macrophages, keratinocytes, many other cells	Stimulates proliferation of hepatocytes and many other epithelial cells
Hepatocyte growth factor (HGF) (scatter factor)	Fibroblasts, stromal cells in the liver, endothelial cells	Enhances proliferation of hepatocytes and other epithelial cells; increases cell motility
Vascular endothelial growth factor (VEGF)	Mesenchymal cells	Stimulates proliferation of endothelial cells; increases vascular permeability
Platelet-derived growth factor (PDGF)	Platelets, macrophages, endothelial cells, smooth muscle cells, keratinocytes	Chemotactic for neutrophils, macrophages, fibroblasts, and smooth muscle cells; activates and stimulates proliferation of fibroblasts, endothelial cells, and other cells; stimulates ECM protein synthesis
Fibroblast growth factors (FGFs) including acidic (FGF-1) and basic (FGF-2)	Macrophages, mast cells, endothelial cells, many other cell types	Chemotactic and mitogenic for fibroblasts; stimulates angiogenesis and ECM protein synthesis
Transforming growth factor-β (TGF-β)	Platelets, T lymphocytes, macrophages, endothelial cells, epithelial cells, smooth muscle cells, fibroblasts	Chemotactic for leukocytes and fibroblasts; stimulates ECM protein synthesis; suppresses acute inflammation
Keratinocyte growth factor (KGF) (i.e., FGF-7)	Fibroblasts	Stimulates keratinocyte migration, proliferation, and differentiation

ECM, Extracellular matrix.

HGF acts as a *morphogen* in embryonic development (i.e., influences the pattern of tissue differentiation), promotes cell migration (hence the designation scatter factor), and enhances hepatocyte survival. HGF is produced by fibroblasts and most mesenchymal cells, endothelial cells, and non-hepatocyte liver cells. It is synthesized as an inactive precursor (pro-HGF) that is proteolytically activated by serine proteases released at sites of injury. The receptor for HGF is MET, which has intrinsic tyrosine kinase activity. MET is frequently overexpressed or mutated in tumors, particularly renal and thyroid papillary carcinomas. Consequently, MET inhibitors are being evaluated for cancer therapy.

Platelet-derived Growth Factor (PDGF). PDGF is a family of several closely related proteins, each consisting of two chains (designated by pairs of letters). Three isoforms of PDGF (AA, AB, and BB) are directly biologically active; PDGF-CC and PDGF-DD must be activated by proteolytic cleavage. PDGF proteins are stored in cytoplasmic granules and released by activated platelets. Although originally isolated from platelets (hence the name), PDGFs are produced by many cells including activated macrophages, endothelium, smooth muscle cells, and tumors. All PDGF isoforms exert their effects by binding to two cell-surface receptors (PDGFR α and β), both of which have intrinsic tyrosine kinase activity. PDGF induces fibroblast, endothelial, and smooth muscle cell proliferation and is also chemotactic for these cells (and inflammatory cells), thereby promoting their recruitment to sites of inflammation and tissue injury.

Vascular Endothelial Growth Factor (VEGF). VEGFs are a family of homodimeric proteins: VEGF-A, VEGF-B, VEGF-C, VEGF-D, and placental growth factor (PlGF). VEGF-A is generally referred to simply as VEGF; it is the major *angiogenic* factor (inducing blood vessel development) after injury and in tumors. In comparison, VEGF-B and PlGF are involved in embryonic vessel development, and VEGF-C and VEGF-D stimulate both angiogenesis and lymphatic development *(lymphangiogenesis)*. Separate from their roles in angiogenesis, VEGFs are also involved in the maintenance of normal endothelium, with highest expression in epithelial cells adjacent to fenestrated epithelium (e.g., kidney podocytes, retinal pigment epithelium, and choroid plexus). VEGF induces all the activities necessary for angiogenesis, including endothelial cell migration and proliferation (capillary sprouting), and promotes formation of vascular lumina. VEGF also affects vascular dilation and increases vascular permeability (VEGF was originally called vascular permeability factor to reflect that activity). As might be anticipated, hypoxia is the most important inducer of VEGF production, through pathways that involve activation of the transcription factor *hypoxia-inducible factor 1* (HIF-1). Other VEGF inducers—produced at sites of inflammation or wound healing—include PDGF and TGF-α.

VEGFs bind to a family of tyrosine kinase receptors (VEGFR-1, VEGFR-2, and VEGFR-3); VEGFR-2 is highly expressed in endothelium and is the most important for angiogenesis. Antibodies against VEGF are approved for the treatment of tumors such as renal and colon cancers that require angiogenesis for their spread and growth. Anti-VEGF therapies have had success in ophthalmic disorders including "wet" age-related macular degeneration (a disorder of inappropriate angiogenesis and vascular permeability that causes adult-onset blindness), angiogenesis associated with retinopathy of prematurity, and vascular leakage that leads to diabetic macular edema. Finally, increased levels of soluble VEGFR-1 (also known as s-FLT-1) in pregnant women may cause preeclampsia (hypertension and proteinuria) by sequestering the free VEGF required for maintenance of normal endothelium.

Fibroblast Growth Factor (FGF). FGF refers to a family of growth factors with more than 20 members. Acidic FGF (aFGF) (also known as FGF-1) and basic FGF (bFGF) (also known as FGF-2) are the best characterized; FGF-7 is also referred to as keratinocyte growth factor (KGF). Released FGFs associate with heparan sulfate in the ECM, which serves as a reservoir for inactive factors that can be subsequently released by proteolysis (e.g., at sites of wound healing). FGFs signal through four tyrosine kinase receptors (FGFR1 through FGFR4) to promote wound healing, hematopoiesis, and development; bFGF has all the activities necessary for angiogenesis.

Transforming Growth Factor β (TGF-β). TGF-β has three isoforms (TGF-β1, TGF-β2, TGF-β3) that belong to a larger family of about 30 members including bone morphogenetic proteins (BMPs), activins, inhibins, and Müllerian inhibiting substance. TGF-β1 has the most widespread distribution and is commonly referred to simply as TGF-β; it is a homodimeric protein produced by multiple cell types including platelets, endothelium, epithelial cells, and inflammatory cells. TGF-β is secreted as a precursor that requires proteolysis to yield the biologically active protein. There are two TGF-β receptors (types I and II) both with serine/threonine kinase activity that induce the phosphorylation of a variety of downstream transcription factors called *Smads*. Phosphorylated Smads form heterodimers, allowing nuclear translocation and association with other DNA-binding proteins to activate or inhibit gene transcription. TGF-β signaling has multiple—and often opposing—effects, depending on the tissue type and any concurrent signals. Agents with such multiplicity of effects are called *pleiotropic,* and TGF-β is "pleiotropic with a vengeance." Primarily, however, TGF-β can be conceptualized as driving scar formation and putting a brake on the inflammation that accompanies wound healing.

- TGF-β stimulates the production of collagen, fibronectin, and proteoglycans and inhibits collagen degradation by both decreasing matrix metalloproteinase (MMP) activity and increasing the activity of tissue inhibitors of proteinases (TIMPs) (discussed later). TGF-β is involved not only in scar formation after injury but also drives fibrosis in lung, liver, intestines, and kidneys in the setting of chronic inflammation.
- TGF-β is also an antiinflammatory cytokine that serves to limit and terminate inflammatory responses. It does this by inhibiting lymphocyte proliferation and activity of other leukocytes. Animals lacking TGF-β have widespread and persistent inflammation.

EXTRACELLULAR MATRIX

The ECM is a protein network that constitutes a significant proportion of any tissue. Cell interactions with the ECM are critical for development, healing, and maintenance of

normal tissue architecture (Fig. 1.13). Much more than a simple "space filler" around cells, the ECM functions as a:

- *Mechanical support* for cell anchorage, cell migration, and maintenance of cell polarity.
- *Regulator of cell proliferation* by binding and displaying growth factors and by signaling via cellular integrin family receptors. The ECM provides a depot for latent growth factors that can be activated within foci of injury or inflammation.
- *Scaffolding for tissue renewal*. Because maintenance of normal tissue structure requires a basement membrane or stromal scaffolds, integrity of the basement membrane or the stroma of parenchymal cells is critical for organized tissue regeneration. Thus ECM disruption prevents effective tissue regeneration and repair.
- *Foundation for establishment of tissue microenvironments*. Basement membrane acts as a boundary between epithelium and underlying connective tissue but often does more than just provide structural support; for example, in the kidney, it forms part of the filtration apparatus.

The ECM is constantly being remodeled; its synthesis and degradation accompany morphogenesis, tissue regeneration and repair, chronic fibrosis, and tumor invasion and metastasis.

The ECM occurs in two basic forms: interstitial matrix and basement membrane (Fig. 1.14).

- *Interstitial matrix.* Interstitial matrix occupies the spaces between stromal cells within connective tissue and between parenchymal epithelium and the underlying supportive vascular and smooth muscle structures in some organs. Interstitial matrix is synthesized by mesenchymal cells (e.g., fibroblasts), forming a three-dimensional, relatively amorphous, semi-fluid gel. In some tissues, such as the gastrointestinal tract, urinary bladder, and periarterial soft tissues, fluid within matrix cushions tissue compression associated with peristalsis, urination, and pulsatile arterial blood flow. The major nonfluid constituents of the interstitial matrix are fibrillar and nonfibrillar collagens, as well as *fibronectin, elastin, proteoglycans, hyaluronate,* and other constituents (see later).

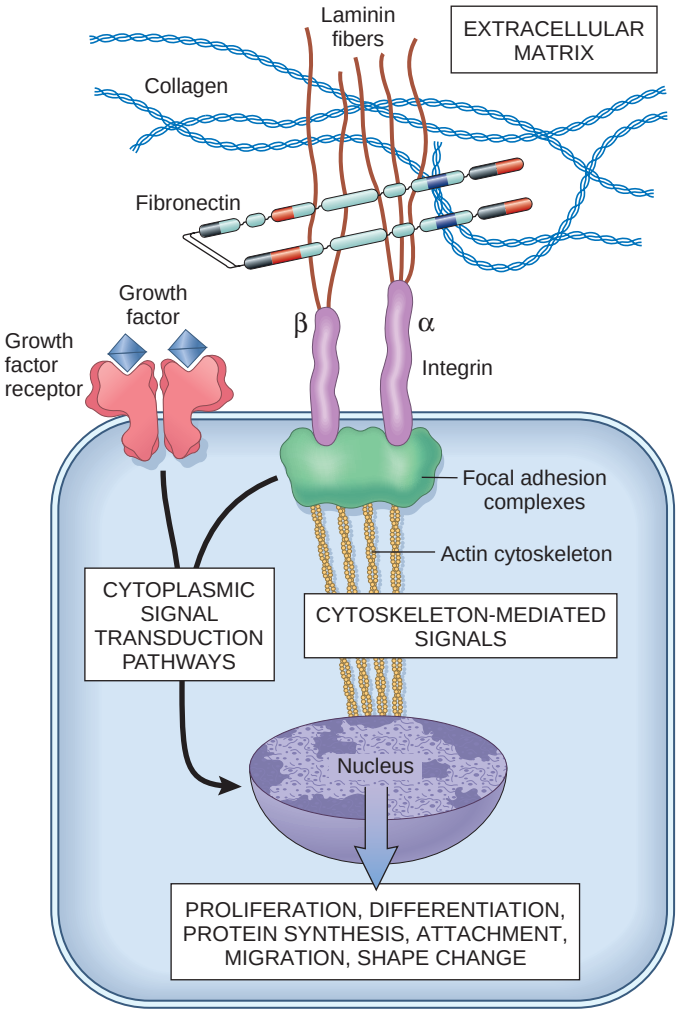


Figure 1.13 Interactions of extracellular matrix (ECM) and growth factors to mediate cell signaling. Cell surface integrins interact with the cytoskeleton at focal adhesion complexes (protein aggregates that include vinculin, α-actinin, and talin; see Fig. 1.17C). This can initiate the production of intracellular messengers or can directly transduce signals to the nucleus. Cell-surface receptors for growth factors can activate signal transduction pathways that overlap with those mediated through integrins. Signals from both ECM interactions and growth factors can be integrated by the cells to produce specific responses, including changes in proliferation, locomotion, or differentiation.

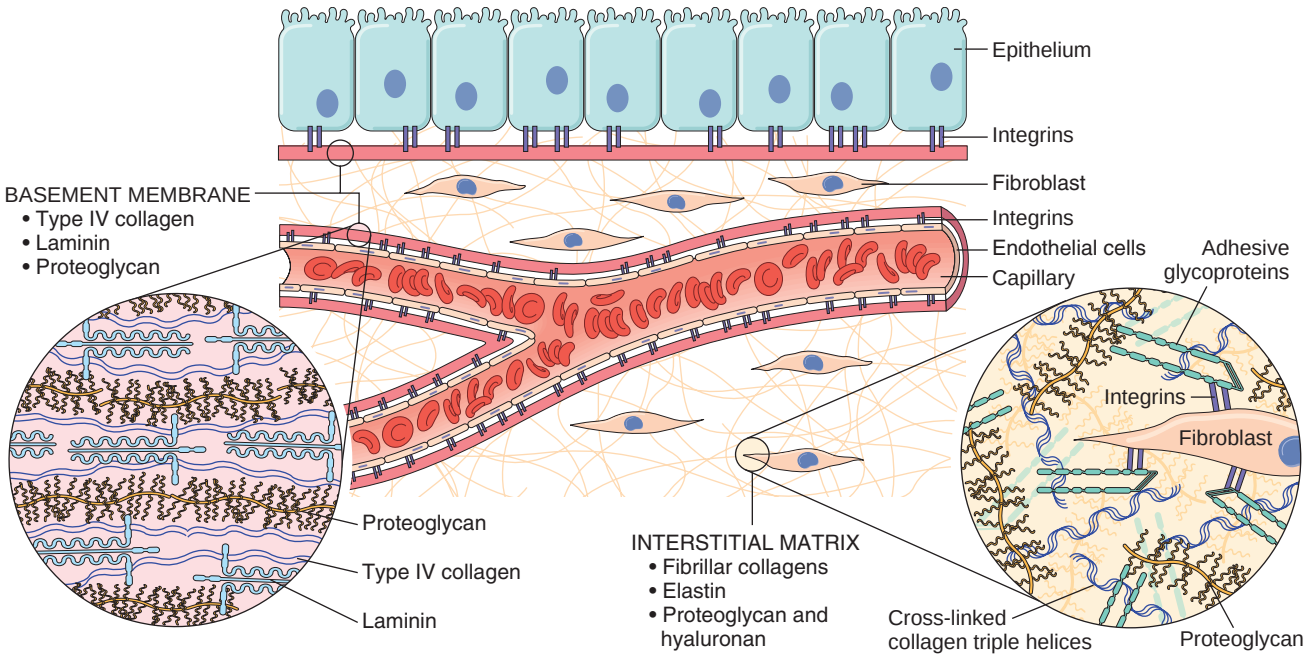


Figure 1.14 Main components of the extracellular matrix (ECM) including collagens, proteoglycans, and adhesive glycoproteins. Both epithelial and mesenchymal cells (e.g., fibroblasts) interact with ECM via integrins. Basement membranes and interstitial ECM have different architecture and general composition, although certain components are present in both. For the sake of clarity, many ECM components (e.g., elastin, fibrillin, hyaluronan, and syndecan) are not shown.

- *Basement membrane*. Interstitial matrix within connective tissues becomes highly organized around epithelial cells, endothelial cells, and smooth muscle cells, where it forms basement membranes, specialized surfaces for cell growth. Basement membrane components, which are synthesized by the overlying epithelium and underlying mesenchymal cells, form a flat lamellar mesh (although labeled as a membrane, it is quite porous). The major constituents are nonfibrillar *type IV collagen* and *laminin*.

Components of the Extracellular Matrix (Fig. 1.15)

ECM components fall into three families:

- *Fibrous structural proteins* such as collagens and elastin that confer tensile strength and recoil
- *Water-hydrated gels* such as proteoglycans and hyaluronan that permit compressive resistance and lubrication
- *Adhesive glycoproteins* that connect ECM elements to one another and to cells

Collagens. Collagens are composed of three separate polypeptide chains braided into a ropelike triple helix (Fig. 1.16). About 30 collagen types have been identified, some of which are unique to specific cells and tissues.

- *Fibrillar collagens:* Some collagen types (e.g., types I, II, III, and V) form linear fibrils stabilized by interchain hydrogen bonding; such fibrillar collagens comprise a major proportion of the connective tissue in bone, tendon, cartilage, blood vessels, and skin, as well as in healing wounds and scars. The tensile strength of the fibrillar collagens derives from lateral cross-linking of the triple helices via covalent bonds that follow lysine hydroxylation. The responsible enzyme, *lysyl hydroxylase,* is dependent on vitamin C, explaining why children with ascorbate deficiency have skeletal deformities and why individuals of any age with vitamin C deficiency heal poorly and bleed easily. Genetic defects, including collagen and lysyl hydroxylase mutations, cause diseases such as *osteogenesis imperfecta* and certain forms of *Ehlers-Danlos syndrome* (Chapter 5).
- *Nonfibrillar collagens* (e.g., type IV collagen) contribute to structures of planar basement membranes; help regulate collagen fibril diameters and collagen-collagen interactions via so-called fibril-associated collagen with interrupted triple helices (FACITs) (e.g., type IX collagen in cartilage); or provide anchoring fibrils that maintain structure of stratified squamous epithelium (e.g., type VII collagen; mutations lead to blistering skin diseases).

Elastin. The ability of tissues to elastically recoil and return to a baseline structure after physical stress is conferred by elastin (see Fig. 1.15A). Elasticity is especially important in cardiac valves and large blood vessels, which need to accommodate recurrent pulsatile flow, as well as in the uterus, skin, and ligaments. Morphologically, elastic fibers consist of a central core of elastin with an associated mesh-like network of *fibrillin* glycoprotein. The latter relationship partially explains why fibrillin synthetic defects lead to skeletal abnormalities and weakened aortic walls, as in patients with Marfan syndrome; fibrillin also controls the availability of free TGF-β, and this function plays a role in pathogenesis of Marfan syndrome (Chapter 5).

Proteoglycans and Hyaluronan (see Fig. 1.15B)**.** Proteoglycans form highly hydrated compressible gels that confer resistance to compressive forces; in joint cartilage, proteoglycans also provide a layer of lubrication between adjacent bony surfaces. Proteoglycans consist of long polysaccharides,

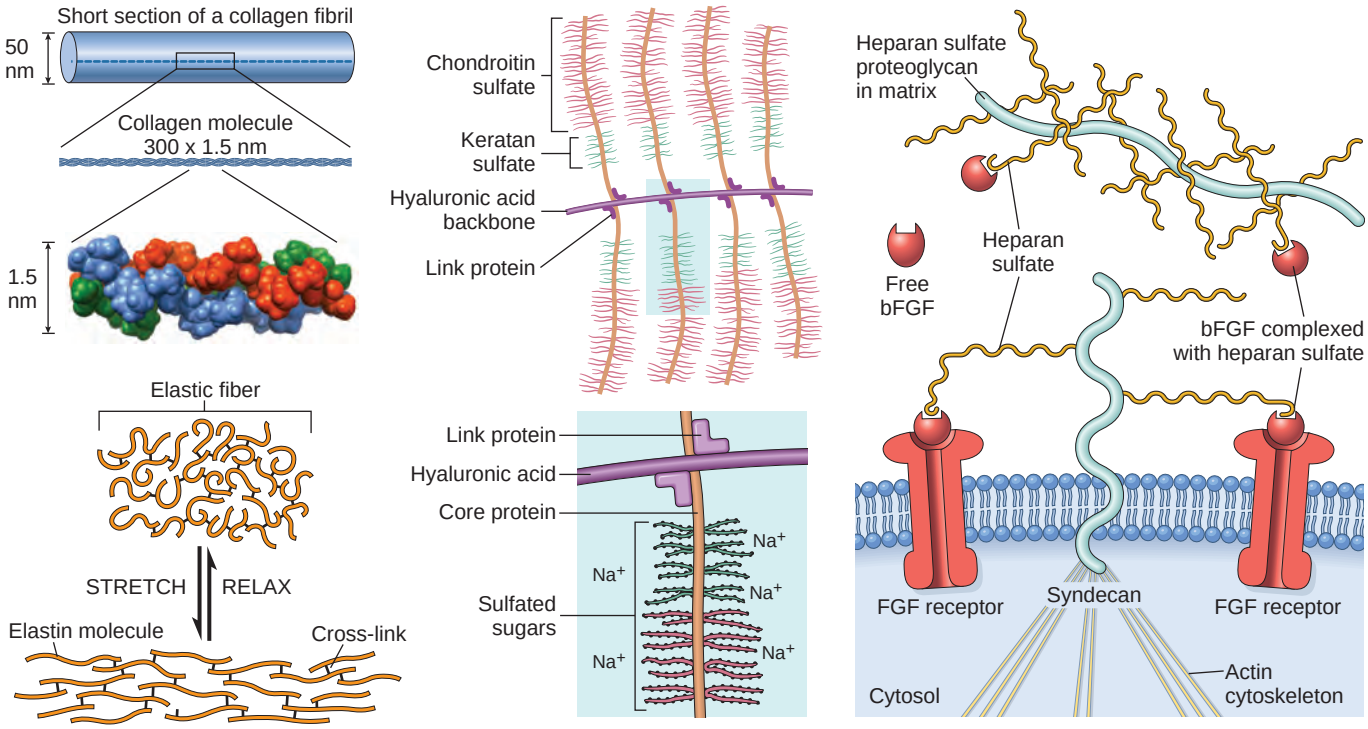


Figure 1.15 Extracellular matrix (ECM) components. (A) Fibrillar collagen and elastic tissue structures. Due to rodlike fibril stacking and extensive lateral cross-linking, collagen fibers have marked tensile strength but do not have much elasticity. Elastin is also cross-linked but differs in having large hydrophobic segments that form a dense globular configuration at rest. As stretch is exerted, the hydrophobic domains are pulled open, but the cross-links keep the tissue intact; release of the stretch tension allows the hydrophobic domains of the proteins to refold. (B) Proteoglycan structure. The highly negatively charged sulfated sugars on the proteoglycan "bristles" attract sodium and water to generate a viscous, but compressible matrix. (C) Regulation of basic fibroblast growth factor (bFGF, also known as FGF-2) activity by ECM and cellular proteoglycans. Heparan sulfate binds bFGF secreted in the ECM. Syndecan is a cell surface proteoglycan with a transmembrane core protein and extracellular glycosaminoglycan side chains that can bind bFGF and a cytoplasmic tail that interacts with the intracellular actin cytoskeleton. Syndecan side chains bind bFGF released from damaged ECM, thus facilitating bFGF interaction with cell-surface receptors. FGF, Fibroblast growth factor.

called *glycosaminoglycans* (examples are keratan sulfate and chondroitin sulfate) attached to a core protein; these are then linked to a long hyaluronic acid polymer called *hyaluronan,* in a manner reminiscent of the bristles on a round brush. The highly negatively charged nature of the densely packed sulfated sugars attracts cations (mostly sodium) and with them, abundant osmotically attracted water—producing a viscous, gel-like matrix. Besides providing compressibility to tissues, proteoglycans also serve as reservoirs for growth factors secreted into the ECM (e.g., FGF and HGF). Some proteoglycans are integral cell membrane proteins that have roles in cell proliferation, migration, and adhesion (e.g., by binding and concentrating growth factors and chemokines) (see Fig. 1.15C).

Adhesive Glycoproteins and Adhesion Receptors. Adhesive glycoproteins and adhesion receptors are structurally diverse molecules variously involved in cell-cell, cell-ECM, and ECM-ECM interactions (Fig. 1.17). Prototypical adhesive glycoproteins include *fibronectin* (a major component of the interstitial ECM) and *laminin* (a major constituent of basement membrane). *Integrins* are representative of the adhesion receptors, also known as cell adhesion molecules (CAMs); the CAMs also include immunoglobulin family members, cadherins, and selectins.

- *Fibronectin* is a large (450-kDa), disulfide-linked heterodimer that exists in tissue and plasma forms; it is synthesized by a variety of cells including fibroblasts, monocytes, and endothelium. Fibronectin has specific domains that can bind to distinct ECM components (e.g., collagen, fibrin, heparin, and proteoglycans), as well as attach to cell integrins (see Fig. 1.17A). In healing wounds, tissue and plasma fibronectin provide the scaffolding for subsequent ECM deposition, angiogenesis, and reepithelialization.
- *Laminin* is the most abundant glycoprotein in basement membranes. It is an 820-kDa cross-shaped heterotrimer that connects cells to underlying ECM components such as type IV collagen and heparan sulfate (see Fig. 1.17B). Besides mediating cell attachment to basement membrane, laminin can also modulate cell proliferation, differentiation, and motility.
- *Integrins* are a large family of transmembrane heterodimeric glycoproteins composed of α- and β-subunits; these allow cells to attach to ECM constituents such as laminin and fibronectin, thus functionally and structurally linking the intracellular cytoskeleton with the outside world. Integrins also facilitate cell-cell adhesive interactions; on leukocytes, they mediate the firm adhesion to and migration across endothelium and epithelium at sites of

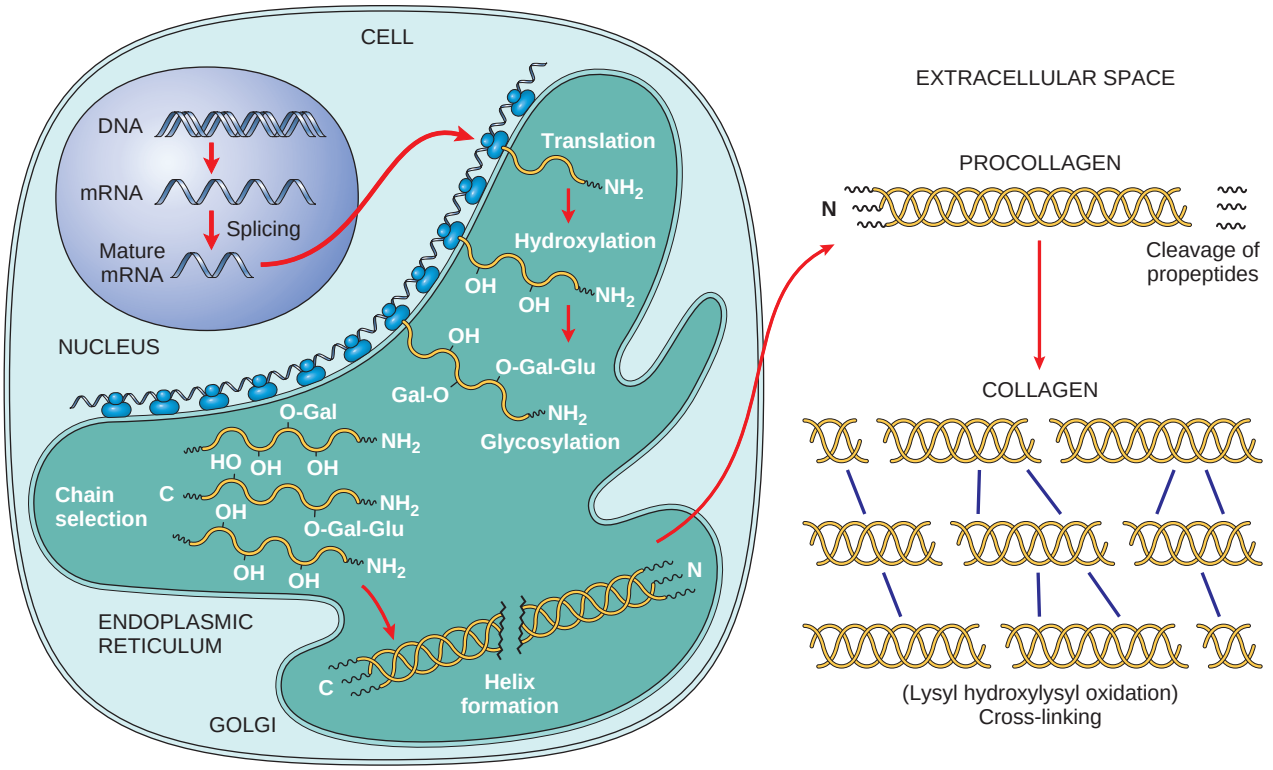


Figure 1.16 Collagen biosynthetic pathway. Some collagen types are heterotrimers (types I, V, and XI), while others are homotrimers (types II and III). The α-chains that make up fibrillar collagen molecules are synthesized as precursor pro-α-chains, with large globular polypeptide regions flanking the central triple-helical domain. After proline and lysine hydroxylation and lysine glycosylation within the endoplasmic reticulum, three procollagen chains align to form a triple helix. For all the fibrillary collagens, the carboxyl end of the propeptide is completely removed by endoproteinase activity after secretion, and the resulting triple helical rodlike domains polymerize in a staggered fashion to form fibrils. The N-terminus propeptide is variably processed depending on the collagen chain. For collagen types I and II, the N-propeptide processing is complete, while for collagens V and XI, a large portion of the N-propeptides remain attached; such "incomplete" processing can regulate fibril size. After secretion, collagen achieves lateral stability through cross-linking involving lysyl oxidase and the previously hydroxylated residues. mRNA, Messenger RNA. Defects in primary sequence, procollagen endopeptidase processing, hydroxylation, or cross-linking can all lead to connective tissue laxity. The specific structures (e.g., blood vessels, skin, bone, ligaments) affected by such disorders are predictable based on the collagen that predominates in that tissue.

inflammation (Chapter 3); they also play a critical role in platelet aggregation (Chapter 4). Integrins attach to ECM components via a tripeptide arginine-glycine-aspartic acid motif (abbreviated RGD). Besides providing focal attachment to underlying substrates, binding through the integrin receptors can trigger signaling cascades that regulate cell locomotion, proliferation, shape, and differentiation (see Fig. 1.17C).

MAINTAINING CELL POPULATIONS

Proliferation and the Cell Cycle

Cell proliferation is fundamental to organism development, to maintenance of steady-state tissue homeostasis, and to replacement of dead or damaged cells. The key elements of cellular proliferation are accurate DNA replication, coordinated synthesis of other cellular components, and equal apportionment of DNA and organelles to daughter cells through the processes of mitosis and cytokinesis.

The sequence of events that results in cell proliferation is called the *cell cycle;* it consists of G_1 (gap 1), S (DNA synthesis), G_2 (gap 2), and M (mitotic) phases; quiescent cells that are not actively cycling are in the G_0 (gap 0) state (Fig. 1.18). Cells can enter G_1 either from the G_0 quiescent cell pool or after completing a round of mitosis. Each stage requires completion of the previous step, as well as activation of necessary factors (see later); nonfidelity of DNA replication or cofactor deficiency results in arrest at various transition points.

The cell cycle is regulated by activators and inhibitors. Cell cycle progression is chaperoned by proteins called *cyclins*—named for the cyclic nature of their production and degradation—and cyclin-associated enzymes called *cyclin-dependent kinases* (CDKs) (Fig. 1.19). Constitutively synthesized CDKs acquire kinase activity—that is, the ability to phosphorylate protein substrates—by forming complexes with the relevant cyclins. Transiently increased synthesis of a particular cyclin thus leads to increased kinase activity of the appropriate CDK binding partner; as the CDK completes its round of phosphorylation, the associated cyclin is degraded and CDK activity abates. Consequently, as cyclin levels rise and fall, activity of associated CDKs will likewise wax and wane.

More than 15 cyclins have been identified; cyclins D, E, A, and B appear sequentially during the cell cycle and bind to one or more CDKs. The cell cycle thus resembles a relay

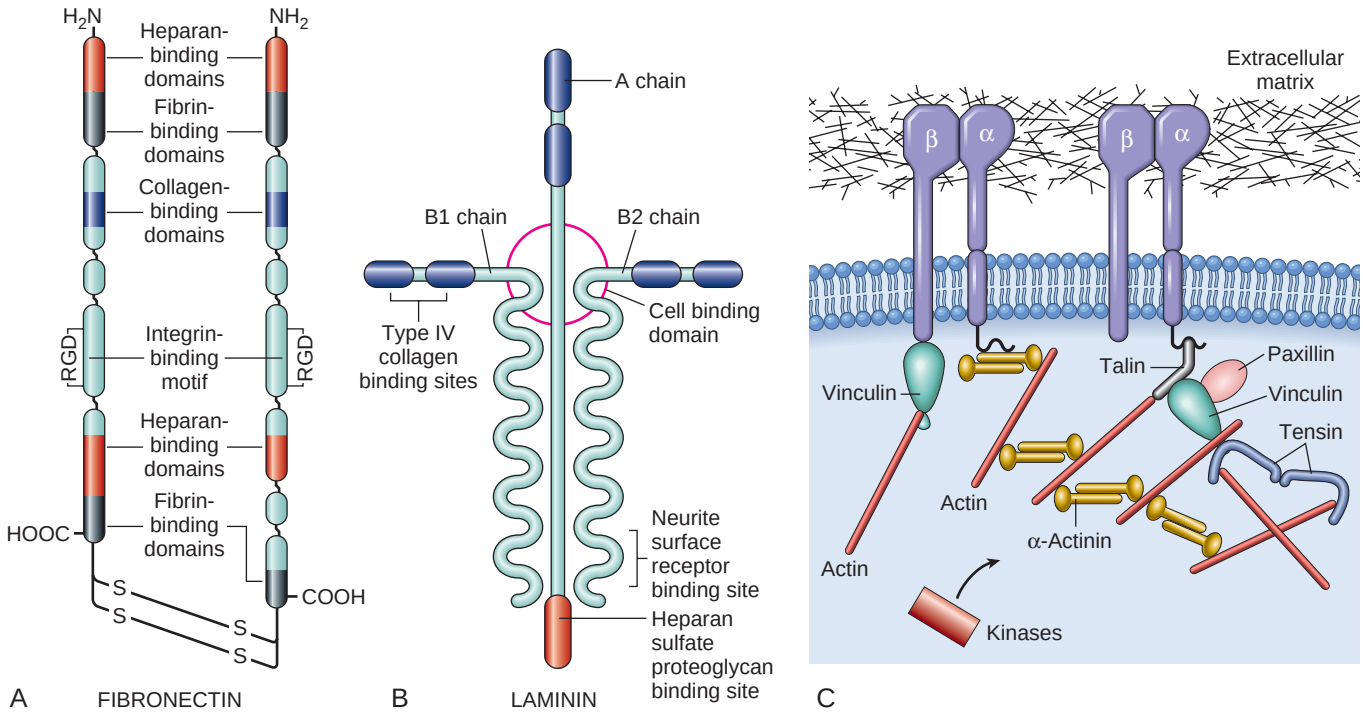


Figure 1.17 Cell and extracellular matrix (ECM) interactions: adhesive glycoproteins and integrin signaling. (A) Fibronectin consists of a disulfide-linked dimer, with several distinct domains that allow binding to ECM and to integrins, the latter through arginine-glycine-aspartic acid (RGD) motifs. (B) The cross-shaped laminin molecule is one of the major components of basement membranes; its multidomain structure allows interactions between type IV collagen, other ECM components, and cell-surface receptors. (C) Integrins and integrin-mediated signaling events at focal adhesion complexes. Each α-β heterodimeric integrin receptor is a transmembrane dimer that links ECM and the intracellular cytoskeleton. Focal adhesion complexes include linking molecules (e.g., vinculin and talin) that can recruit and activate kinases and, ultimately, trigger downstream signaling cascades.

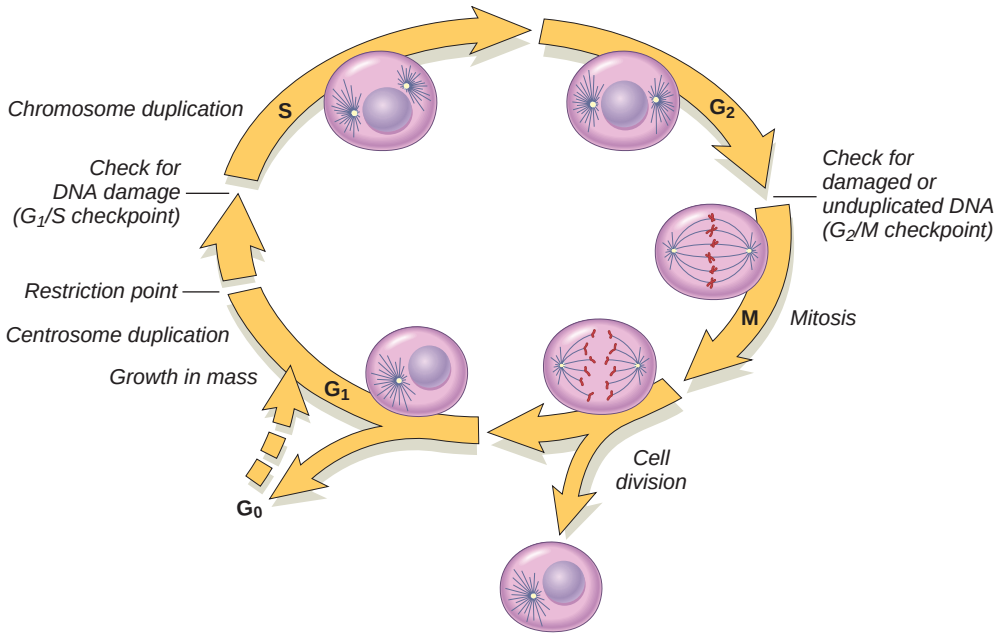


Figure 1.18 Cell cycle landmarks. The figure shows the cell cycle phases (G_0, G_1, G_2, S, and M), the location of the G_1 restriction point, and the G_1/S and G_2/M cell cycle checkpoints. *G_1 restriction point* refers to the stage in G_1 where the cell is committed to advance further into the cell cycle without requiring any more of the growth signal that initiated cell division. Cells from labile tissues such as the epidermis and the gastrointestinal tract may cycle continuously; stable cells such as hepatocytes are quiescent but can enter the cell cycle; permanent cells such as neurons and cardiac myocytes have lost the capacity to proliferate. (Modified from Pollard TD, Earnshaw WC: *Cell Biology,* Philadelphia, 2002, Saunders.)

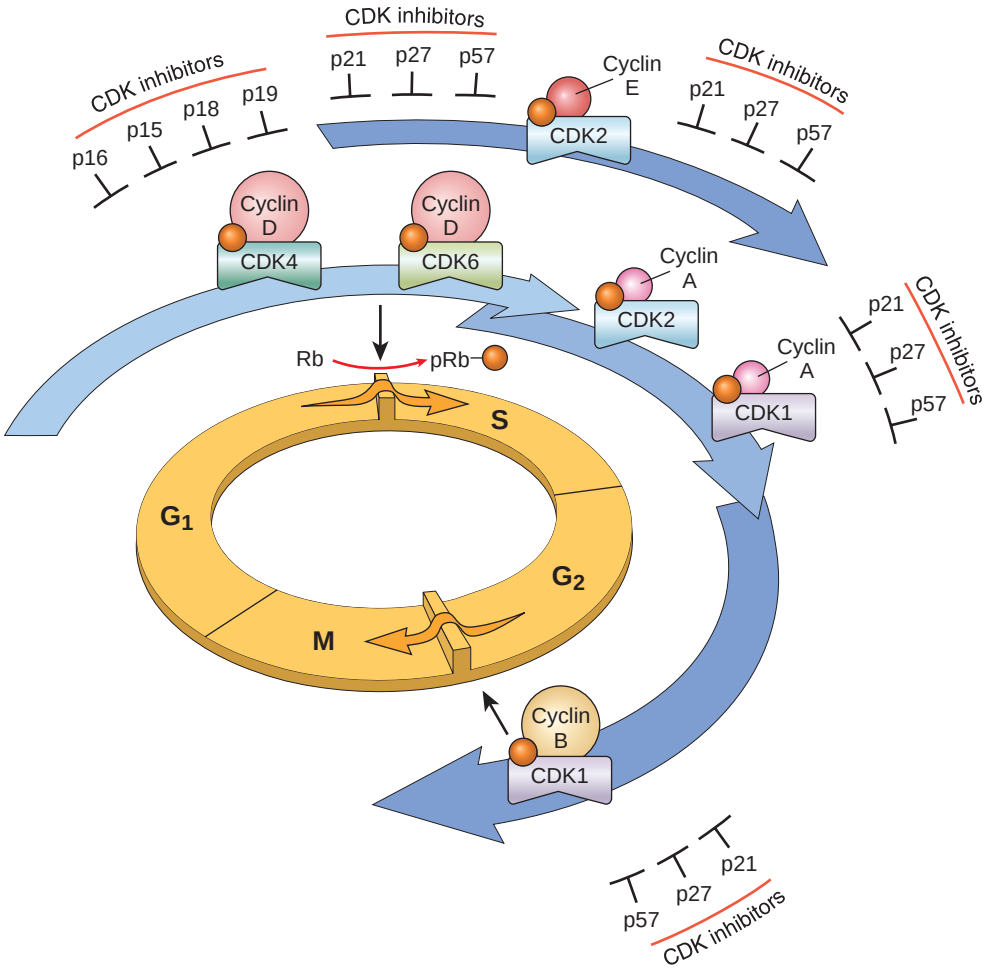


Figure 1.19 Role of cyclins, cyclin-dependent kinases (CDKs) and CDK inhibitors (CDKIs) in regulating the cell cycle. Shaded arrows represent the phases of the cell cycle during which specific cyclin-CDK complexes are active. As illustrated, cyclin D-CDK4, cyclin D-CDK6, and cyclin E-CDK2 regulate the G_1-to-S transition by phosphorylating the Rb protein (pRb). Cyclin A-CDK2 and cyclin A-CDK1 are active in the S phase. Cyclin B-CDK1 is essential for the G_2-to-M transition. Two families of CDK inhibitors can block activity of CDKs and progression through the cell cycle. The so-called INK4 inhibitors, composed of p15, p16, p18, and p19, act on cyclin D-CDK4 and cyclin D-CDK6. The other family of three inhibitors, p21, p27, and p57, can inhibit all CDKs.

race in which each leg is regulated by a distinct set of cyclins: as one collection of cyclins leaves the track, the next set takes over.

Surveillance mechanisms primed to sense DNA or chromosomal damage are embedded within the cell cycle. These quality control *checkpoints* ensure that cells with genetic imperfections do not complete replication. Thus the G_1-S checkpoint monitors DNA integrity before irreversibly committing cellular resources to DNA replication. Later in the cell cycle, the G_2-M restriction point insures that there has been accurate genetic replication before the cell actually divides. When cells do detect DNA irregularities, checkpoint activation delays cell cycle progression and triggers DNA repair mechanisms. If the genetic derangement is too severe to be repaired, cells either undergo apoptosis or enter a nonreplicative state called *senescence*—primarily through p53-dependent mechanisms (see later).

Enforcing the cell cycle checkpoints is the job of *CDK inhibitors* (CDKIs); they accomplish this by modulating CDK-cyclin complex activity. There are several different CDKIs.

- One family of CDKIs—composed of three proteins called p21 (CDKN1A), p27 (CDKN1B), and p57 (CDKN1C)—broadly inhibits multiple CDKs.
- Another family of CDKIs has selective effects on cyclin CDK4 and cyclin CDK6; these proteins are called p15 (CDKN2B), p16 (CDKN2A), p18 (CDKN2C), and p19 (CDKN2D).
- Defective CDKI checkpoint proteins allow cells with damaged DNA to divide, resulting in mutated daughter cells at risk for malignant transformation.

An equally important aspect of cell growth and division is the biosynthesis of the membranes, cytosolic proteins, and organelles necessary to make two daughter cells. Thus as growth factor receptor signaling stimulates cell cycle progression, it also activates events that promote the metabolic changes that support growth. Chief among these is the switch to aerobic glycolysis (with the counter-intuitive reduction in oxidative phosphorylation), also called the *Warburg effect*. These alterations in cell metabolism are an

important element in cancer cell growth and are discussed in greater detail in Chapter 7.

Stem Cells

Stem cells have the dual property of being able to self renew and to give rise to differentiated cells and tissues. During development, *totipotent stem cells* can give rise to the full range of differentiated tissues; in the mature organism, *adult stem cells* only have the capacity to replace damaged cells and maintain cell populations within the tissues where they reside. There are also populations of stem cells between these extremes with varying capacities to differentiate into multiple (but limited) cell lineages. Thus depending on the source and stage of development, there are limits on the cell types that a "stem cell" can generate.

In normal tissues—without healing, degeneration, or neoplasia—there is a homeostatic equilibrium between replication, self-renewal, and differentiation of stem cells and death of mature, fully differentiated cells (Fig. 1.20). The dynamic relationship between stem cells and terminally differentiated parenchyma is particularly evident in the continuously dividing epithelium of the skin; stem cells at the basal layer of the epithelium divide and their daughter cells progressively differentiate as they migrate to the upper layers of the epithelium before dying and being shed.

Under conditions of homeostasis, stem cells are maintained by self-renewal, which can involve two types of cell division:

- *Asymmetric division* refers to cell replication in which one daughter cell enters a differentiation pathway and gives rise to mature cells, while the other remains undifferentiated and retains its self-renewal capacity.
- *Symmetric division* occurs when both daughter cells retain self-renewal capacity. Such replication occurs early in embryogenesis—when stem cell populations are expanding—and under stress conditions, as in bone marrow repopulation after ablative chemotherapy.

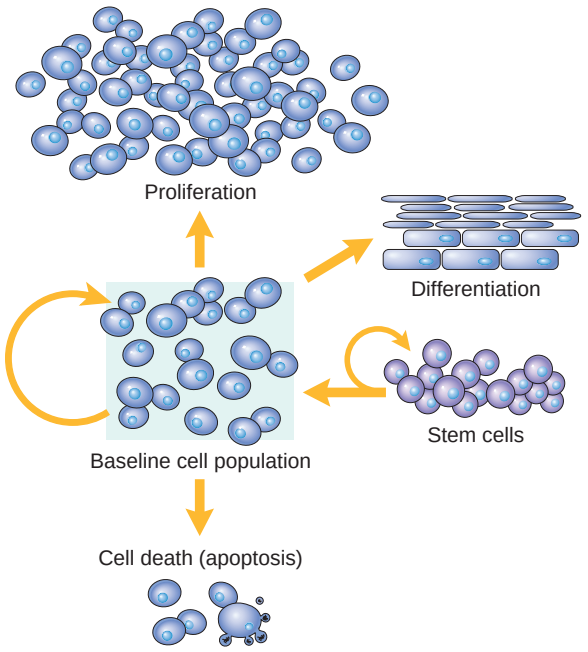


Figure 1.20 Mechanisms regulating cell populations. Cell numbers can be altered by increased or decreased rates of stem cell input, cell death resulting from apoptosis, or changes in the rates of proliferation or differentiation. (Modified from McCarthy NJ, et al: Apoptosis in the development of the immune system: growth factors, clonal selection and bcl-2, *Cancer Metastasis Rev* 11:157, 1992.)

Although some researchers separate stem cells into multiple different subsets, there are fundamentally only two varieties.

- *Embryonic stem (ES) cells* are the most undifferentiated. They are present in the inner cell mass of the blastocyst, have virtually limitless cell renewal capacity, and can give rise to every cell in the body; they are thus said to be *totipotent* (Fig. 1.21). While ES cells can be maintained for extended periods without differentiating, they can be induced under appropriate culture conditions to form specialized cells of all three germ cell layers.
- *Tissue stem cells* (also called *adult stem cells*) are found in intimate association with the differentiated cells of a given tissue. They are normally protected within specialized tissue microenvironments called *stem cell niches*. Such niches have been demonstrated in many organs, most notably the bone marrow, where hematopoietic stem cells characteristically congregate in perivascular niches, and the intestines, where epithelial stem cells are confined to the crypts. Other stem cell niches include the bulge region of hair follicles, the limbus of the cornea, and the subventricular zone in the brain. Soluble factors and other cells within the niches regulate the balance between stem cell quiescence and expansion and differentiation (Fig. 1.22).

Adult stem cells have a limited repertoire *(lineage potential)* of differentiated cells that they can generate. Thus although adult stem cells can maintain tissues with high (e.g., skin, and gastrointestinal tract) or low (e.g., endothelial) cell turnover, the adult stem cells in any given tissue can usually produce only those cells normally found within that tissue.

Hematopoietic stem cells continuously replenish all the cellular elements of the blood as they exit the circulation, senesce, or are otherwise consumed. They can be isolated directly from bone marrow, as well as from the peripheral blood after administration of certain colony-stimulating factors that induce their release from bone marrow niches. Although they are overall rare, hematopoietic stem cells can be purified to virtual purity based on cell surface markers. Clinically, these stem cells can be used to repopulate marrows depleted after chemotherapy (e.g., for leukemia) or to provide normal precursors to correct various blood cell defects (e.g., sickle cell disease; see Chapter 14).

Besides hematopoietic stem cells, the bone marrow (and, notably, other tissues such as fat) also contains a population of *mesenchymal stem cells*. These are multipotent cells that can differentiate into a variety of stromal cells including chondrocytes (cartilage), osteocytes (bone), adipocytes (fat), and myocytes (muscle). Because these cells can be massively expanded and can generate a locally immunosuppressive microenvironment (thus potentially evading rejection), they represent a potential means of manufacturing stromal cellular scaffoldings for tissue regeneration.

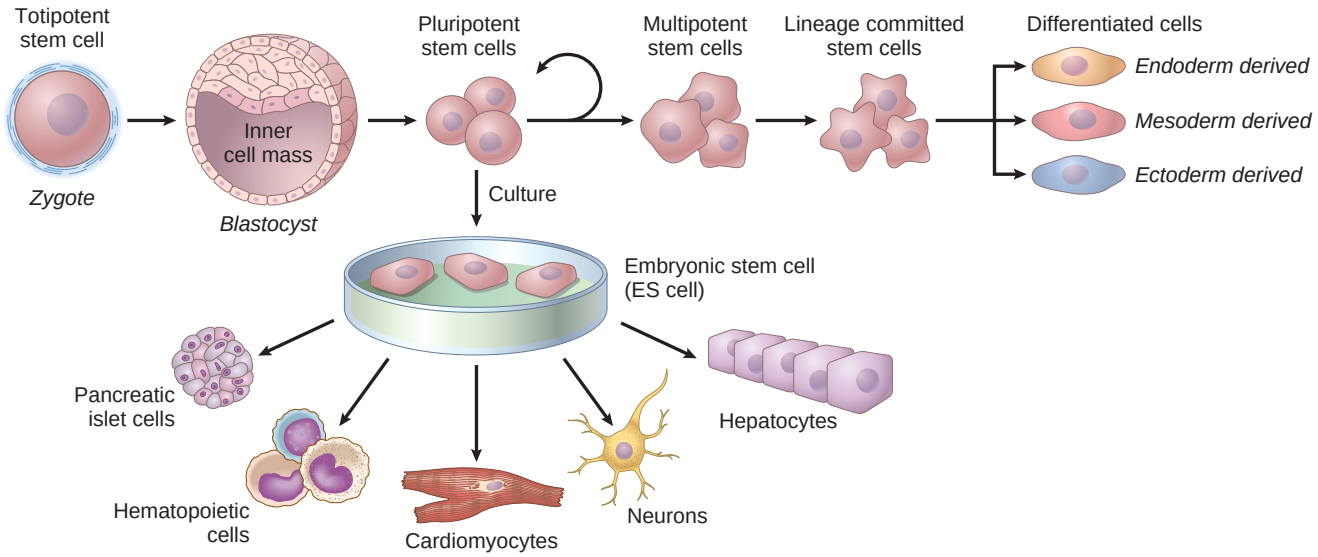


Figure 1.21 Embryonal stem *(ES)* cells. The zygote, formed by the union of sperm and egg, divides to form blastocysts, and the inner cell mass of the blastocyst generates the embryo. The pluripotent cells of the inner cell mass, known as ES cells, can be induced to differentiate into cells of multiple lineages. In the embryo, pluripotent stem cells can asymmetrically divide to yield a residual stable pool of ES cells in addition to generating populations that have progressively more restricted developmental capacity, eventually generating stem cells that are committed to just specific lineages. ES cells can be cultured in vitro and induced to differentiate into cells characteristic of all three germ layers.

Regenerative Medicine

The burgeoning field of regenerative medicine has been made possible by the ability to identify, isolate, expand, and transplant stem cells. Theoretically the differentiated progeny of ES or adult stem cells can be used to repopulate damaged tissues or even to construct entire replacement organs. There is, therefore, considerable interest in therapeutic opportunities for restoring tissues that have low intrinsic regenerative capacity, such as myocardium or neurons, to promote healing after myocardial infarction or stroke, respectively. Despite improved ability to purify and expand stem cells, success has, thus far, been limited by difficulties in introducing and functionally integrating replacement cells at sites of damage.

Another issue arises from the immunologic reactivity of most stem cells. Although mesenchymal stem cells may be relatively immunologically privileged, most other adult stem cells, as well as ES cells (from fertilized blastocysts), express histocompatibility molecules—human leukocyte antigen (HLA) in humans (see Chapters 3 and 6)—that provoke immunologic rejection when transplanted. Hence considerable effort has been expended to generate cells with the totipotential characteristics of ES cells from cells that can be harvested from an individual patient. This would, in principle, allow new tissues to be generated and transplanted

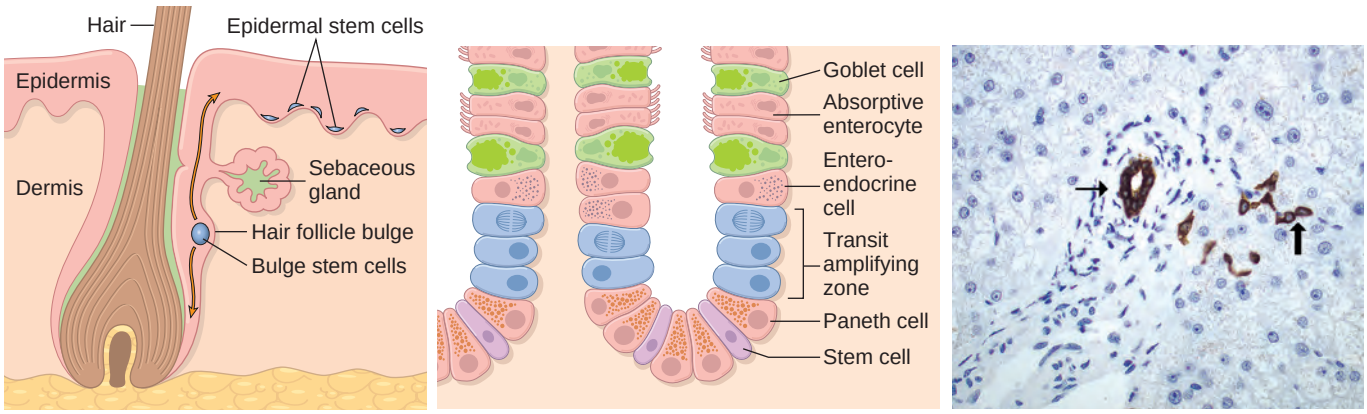


Figure 1.22 Stem cell niches in various tissues. (A) Skin stem cells are located in the bulge area of the hair follicle, in sebaceous glands, and in the lower layer of the epidermis. (B) Small intestinal crypt base columnar (CBC) stem cells are located at the base of the crypt interspersed between Paneth cells. (C) Liver stem cells (oval cells) are located in the canals of Hering (thick arrow), structures that connect bile ductules (thin arrow) to parenchymal hepatocytes. Bile duct cells and canals of Hering are highlighted here by an immunohistochemical stain for cytokeratin 7. (C, Courtesy Tania Roskams, MD, University of Leuven, Belgium.)

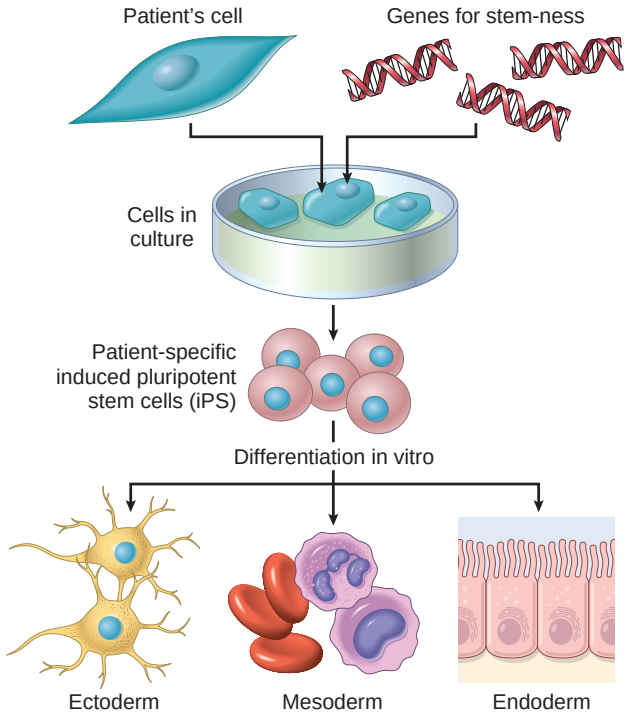


Figure 1.23 Induced pluripotent stem *(iPS)* cells. Genes that confer stem cell properties are introduced into a patient's differentiated cells, giving rise to stem cells that can be induced to differentiate into various lineages. (Modified from Hochedlinger K, Jaenisch R: Nuclear transplantation, embryonic stem cells, and the potential for cell therapy, *N Engl J Med* 349:275, 2003.)

without fear of immunologic graft rejection. To accomplish this, a handful of genes have been identified whose products can—remarkably—reprogram somatic cells to achieve the "stem-ness" of ES cells. When such genes are introduced into fully differentiated cells (e.g., fibroblasts), *induced pluripotent stem (iPS) cells* are generated (Fig. 1.23). While not yet in practice, their differentiated progeny could be remarkable therapeutic agents, e.g., by generating insulin-secreting β cells in a patient with diabetes.

Concluding Remarks. This survey of selected topics in cell biology will serve as a basis for our later discussions of pathology, and we will refer back to it throughout the book. Students should, however, remember that this summary is intentionally brief, and more information about some of the fascinating topics reviewed here can be readily found in textbooks devoted to cell and molecular biology.

SUGGESTED READINGS

Genetics and Epigenetics

Batista PJ, Chang HY: Long noncoding RNAs: cellular address codes in development and disease, *Cell* 152:1298, 2013. [*Good review regarding long noncoding RNA biology*].

Cech TR, Steitz JA: The noncoding RNA revolution—trashing old rules to forge new ones, *Cell* 157:77, 2014. [*An excellent review of the roles played by noncoding RNAs*].

Meller VH, Joshi SS, Deshpande N: Modulation of chromatin by noncoding RNA, *Annu Rev Genet* 49:673, 2015. [*Excellent overview of the roles played by noncoding RNAs in nuclear organization*].

Minarovits J, Banati F, Szenthe K et al: Epigenetic regulation, *Adv Exp Med Biol* 879:1, 2016. [*A brief primer on the pathways that regulate chromatin structure and accessibility*].

Rowley MJ, Corces VG: The three-dimensional genome: principles and roles of long-distance interactions, *Curr Opin Cell Biol* 40:8, 2016. [*An interesting discussion regarding the mechanisms by which three-dimensional conformations can influence nuclear transcription*].

Sun Q, Hao Q, Prasanth KV: Nuclear long noncoding RNAs: key regulators of gene expression, *Trends Genet* 2017. [*An up-to-date review on the role of long noncoding RNA on chromatin organization, as well as transcriptional and posttranscriptional gene expression*].

Wright AV, Nuñez JK, Doudna JA: Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering, *Cell* 164:29, 2016. [*A superb review on CRISPR/Cas from some of the researchers who discovered and developed its possible applications*].

Cellular Housekeeping

Andersson ER: The role of endocytosis in activating and regulating signal transduction, *Cell Mol Life Sci* 69:1755, 2011. [*Overview of endocytosis with specific emphasis on its role in modulating intracellular signaling*].

Choi AM, Ryter SW, Levine B: Autophagy in human health and disease, *N Engl J Med* 368:651, 2013. [*Superb review concerning the physiologic and pathophysiologic aspects of autophagy*].

English AR, Zurek N, Voeltz GK: Peripheral ER structure and function, *Curr Opin Cell Biol* 21:596, 2009. [*Overview of the structural and functional organization of the endoplasmic reticulum and its relationship to other cellular organelles*].

Guillot C, Lecuit T: Mechanics of epithelial tissue homeostasis and morphogenesis, *Science* 340:1185, 2013. [*Topical discussion about cellular interactions and the mechanical basis of tissue maintenance*].

Hetz C, Chevet E, Oakes SA: Proteostasis control by the unfolded protein response, *Nat Cell Biol* 17:829, 2015. [*Mechanisms underlying endoplasmic reticulum editing and cellular homeostasis*].

Kaur J, Debnath J: Autophagy at the crossroads of catabolism and anabolism, *Nat Rev Mol Cell Biol* 16:461, 2015. [*Excellent review of the mechanisms and consequences of cellular autophagy*].

Johnson DS: Establishing and transducing cell polarity: common themes and variations, *Curr Opin Cell Biol* 51:33, 2017. [*A well-written overview of the intracellular architecture that directs and maintains polarity*].

Simons K, Sampaio JL: Membrane organization and lipid rafts, *Cold Spring Harb Perspect Biol* 3:1, 2013. [*Nice review of the general principles of membrane architecture emphasizing domain organization*].

Wong E, Cuervo AM: Integration of clearance mechanisms: the proteasome and autophagy, *Cold Spring Harb Perspect Biol* 2:1, 2010. [*Overview of intracellular degradation pathways, specifically focusing on the elimination of aberrant or abnormal constituents*].

Cellular Metabolism and Mitochondrial Function

Andersen JL, Kornbluth S: The tangled circuitry of metabolism and apoptosis, *Mol Cell* 49:399, 2013. [*Solid review of the interplay between cell metabolism, cell proliferation, and cell death*].

Burke PJ: Mitochondria, bioenergetics and apoptosis in cancer, *Trends Cancer* 3:857, 2017. [*A well-written discussion linking mitochondrial energy generation and cell death functions, with a focus on possible cancer therapeutics*].

Friedman JR, Nunnari J: Mitochondrial form and function, *Nature* 505:335, 2014. [*Good overview of mitochondrial replication and response to cellular injury*].

Tait SW, Green DR: Mitochondria and cell death: outer membrane permeabilization and beyond, *Nat Rev Mol Cell Biol* 11:621, 2010. [*Review of the role of mitochondria in cell death pathways*].

Cellular Activation

Duronio RJ, Xiong Y: Signaling pathways that control cell proliferation, *Cold Spring Harb Perspect Biol* 5:1, 2013. [*Excellent overall review of cell signaling and proliferation*].

Morrison DK: MAP kinase pathways, *Cold Spring Harb Perspect Biol* 4:1, 2012. [*Review of mitogen-activated kinase signaling pathways*].

Nusse R, Clevers H: Wnt/β-catenin signaling, disease, and emerging therapeutic modalities, *Cell* 169:985, 2017. [*Nice overview of the Wnt/β-catenin signaling pathways, particularly focusing on their role in stem cell biology and malignancy*].

Perona R: Cell signalling: growth factors and tyrosine kinase receptors, *Clin Transl Oncol* 8:77, 2011. [*Overview of signaling pathways with an emphasis on how these become dysregulated in malignancy*].

Maintaining Cell Populations

Alvarado AS, Yamanaka S: Rethinking differentiation: stem cells, regeneration, and plasticity, *Cell* 157:110, 2014. [*Very good review of the fundamental concepts in stem cell biology*].

Blau HM, Daley GQ: Stem cells in the treatment of disease, *N Engl J Med* 380:1748, 2019. *[Excellent review on stem cell biology and its therapeutic potential].*

De Los Angeles A, Ferrari F, Xi R et al: Hallmarks of pluripotency, *Nature* 525:469, 2015. [*Excellent overview of pluripotent stem cells including the molecular pathways of their renewal and differentiation capacities*].

Fuchs E, Chen T: A matter of life and death: self-renewal in stem cells, *EMBO Rep* 14:39, 2013. [*Scholarly review on the conceptual framework and experimental underpinnings of our understanding regarding stem cell renewal, using cutaneous stem cells as a paradigm*].

Jang S, Collin del Hortet A, Soto-Gutierrezy A: Induced pluripotent stem cell-derived endothelial cells: overview, current advances, applications, and future directions, *Am J Pathol* 189:502, 2019. *[A focused review of applications of induced stem cell–derived endothelial cells].*

Martello G, Smith A: The nature of embryonic stem cells, *Annu Rev Cell Dev Biol* 30:647, 2014. [*Good comprehensive overview of cell plasticity and stemness*].

Scadden DT: Nice neighborhood: emerging concepts of the stem cell niche, *Cell* 157:41, 2014. [*Well-written paper describing the role of niche in stem cell biology*].

Wu J, Izpisua Belmonte JC: Stem cells: a renaissance in human biology research, *Cell* 165:1572, 2016. [*High-level review of the opportunities offered by new technologies and insights in stem cell biology in understanding human development and disease*].

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 2

Cell Injury, Cell Death, and Adaptations

Scott A. Oakes

CHAPTER CONTENTS

INTRODUCTION TO PATHOLOGY

Pathology is the study of the structural, biochemical, and functional changes in cells, tissues, and organs that underlie disease. By the use of morphologic, microbiologic, immunologic, and molecular techniques, pathology attempts to explain the whys and wherefores of the signs and symptoms manifested by patients while providing a rational basis for clinical care and therapy. It thus serves as the bridge between the basic sciences and clinical medicine, and is the scientific foundation for all of medicine. In Chapter 1 we examined the cellular and molecular properties of healthy cells. In this chapter, we will build on that knowledge to discuss the fundamental mechanisms that underlie various forms of cell injury and death.

Traditionally, the study of pathology is divided into general pathology and systemic pathology. General pathology is concerned with the common reactions of cells and tissues to injurious stimuli. Such reactions are often not tissue specific: thus, acute inflammation in response to bacterial infections produces a very similar reaction in most tissues. On the other hand, systemic pathology examines the alterations and underlying mechanisms in diseases of particular organ systems. In this book, we first cover the principles of general pathology and then proceed to specific disease processes as they affect different organs.

The four aspects of a disease process that form the core of pathology are causation *(etiology)*, biochemical and molecular mechanisms *(pathogenesis)*, the associated structural *(morphologic changes)* and functional alterations in cells and organs, and the resulting clinical consequences *(clinical manifestations).*

- *Etiology is the initiating cause of a disease.* Although there are myriad factors that cause disease, all can be grouped into two broad classes: genetic (e.g., inherited or acquired mutations, and disease-associated gene variants, or polymorphisms) and environmental (e.g., infectious, nutritional, chemical, physical). The idea that one etiologic agent is the cause of one disease arose from the study of infections and inherited disorders caused by single gene anomalies, but the majority of diseases are not this simple. In fact, most common afflictions, such as atherosclerosis and cancer, arise from the effects of various environmental insults on a genetically susceptible individual and hence are referred to as being multifactorial. The relative contribution of inherited susceptibility and environmental influences varies in different diseases, and it is challenging to precisely define their roles in most multifactorial diseases.
- *Pathogenesis refers to the sequence of molecular, biochemical, and cellular events that lead to the development of disease.* Thus, pathogenesis explains how the underlying etiologies produce the morphologic and clinical manifestations of the disease. The study of pathogenesis is a central focus of pathology. Even when the initial cause is known (e.g., infection or mutation), it is many steps removed from the expression of the disease. For example, to truly understand the disorder cystic fibrosis it is essential to know not only the defective gene and gene product, but also the biochemical and morphologic events that lead to clinically significant disease in the lungs, pancreas, and other organs. New technological advances, particularly the use of so-called "omics" technologies (genomics, proteomics, metabolomics) to interrogate diseases, hold great promise for elucidating pathogenic mechanisms. Hopefully, the application of these methods and the analysis of mounds of "big data" so generated will lead not only to better understanding of pathogenesis but also to the identification of *biomarkers* that predict disease progression and therapeutic responses. This, of course, is the goal of *precision medicine*.
- *Morphologic changes refer to the structural alterations in cells or tissues that are characteristic of a disease and hence diagnostic of an etiologic process.* Traditionally, diagnostic pathology has used morphology to determine the nature of disease and to follow its progression. Although morphology remains a cornerstone of diagnosis, it is now routinely supplemented by analysis of protein expression and genetic alterations. Nowhere is this more striking than in the study of neoplasms; breast cancers that are indistinguishable morphologically may result from different genetic abnormalities that result in widely different courses, therapeutic responses, and prognoses. Molecular analysis by techniques such as next-generation sequencing (Chapter 7) has revealed genetic differences that predict the behavior of tumors as well as their response to therapies, an increasing number of which are now chosen on the basis of the presence (or absence) of specific molecular alterations.
- *Clinical manifestations.* The end results of genetic, biochemical, and structural changes in cells and tissues are functional abnormalities that lead to the clinical manifestations (symptoms and signs) of disease, as well as its progression (clinical course and outcome). Hence, clinicopathologic correlations are very important in the study of disease.

Virtually all diseases start with molecular or structural alterations in cells. This concept of the cellular basis of disease was first put forth in the nineteenth century by Rudolf Virchow, known as the father of modern pathology. Virchow emphasized the idea that individuals are sick because their cells are sick. We therefore begin our consideration of pathology with the study of the causes, mechanisms, and morphologic and biochemical correlates of *cell injury*. Injury to cells and to the extracellular matrix ultimately leads to *tissue and organ injury,* which determines the morphologic and clinical patterns of disease.

OVERVIEW OF CELLULAR RESPONSES TO STRESS AND NOXIOUS STIMULI

The normal cell is confined to a fairly narrow range of function and structure dictated by its state of metabolism, differentiation, and specialization; by constraints imposed by neighboring cells; and by the availability of metabolic substrates. It is nevertheless able to handle physiologic demands, maintaining a healthy steady state called *homeostasis*. *Adaptations* are reversible functional and structural responses to changes in physiologic states (e.g., pregnancy) and some pathologic stimuli, during which new but altered

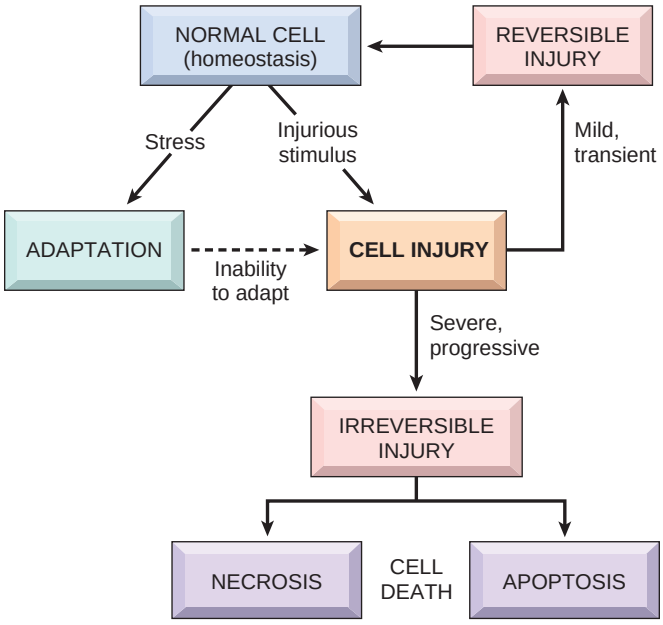


Figure 2.1 Stages of the cellular response to stress and injurious stimuli.

steady states are achieved, allowing the cell to survive and continue to function (Fig. 2.1). The adaptive response may consist of an increase in the size (hypertrophy) and functional activity of cells, an increase in cell number (hyperplasia), a decrease in the size and metabolic activity of cells (atrophy), or a change in the phenotype of cells (metaplasia). If the stress is eliminated, the cell can return to its original state without having suffered any harmful consequences.

If the limits of adaptive responses are exceeded or if cells are exposed to damaging insults, deprived of critical nutrients, or compromised by mutations that affect essential cellular functions, a sequence of events follows that is termed *cell injury* (see Fig. 2.1). Cell injury is *reversible* up to a point, but if the injurious stimulus is persistent or severe, the cell suffers *irreversible injury* and ultimately undergoes *cell death*. *Adaptation, reversible injury,* and *cell death* may be stages of progressive impairment following different types of insults. For instance, in response to increased hemodynamic loads, the heart muscle becomes enlarged, a form of adaptation, which because of increased metabolic demands is more susceptible to injury. If the blood supply to the myocardium is compromised or inadequate, the muscle first suffers reversible injury, manifested by certain cytoplasmic changes (described later). Unless the blood supply is quickly restored, the cells suffer irreversible injury and die (Fig. 2.2).

The removal of damaged, unneeded, and aged cells through *cell death* is a normal and essential process in embryogenesis, the development of organs, and the maintenance of homeostasis into adulthood. Conversely, excessive cell death as a result of progressive injury is one of the most crucial events in the evolution of disease in any tissue or organ. It results from diverse causes, including ischemia (reduced blood flow), infection, and toxins. There are two principal pathways of cell death, *necrosis* and *apoptosis*. Nutrient deprivation triggers an adaptive cellular response called *autophagy* that may also culminate in cell death. A detailed discussion of these and some other, less common pathways of cell death follows later in the chapter.

Stresses of different types may induce changes in cells and tissues other than typical adaptations, cell injury, and death. Metabolic derangements and chronic injury may be associated with *intracellular accumulations* of a number of substances, including proteins, lipids, and carbohydrates. Calcium may be deposited at sites of cell death, resulting in *pathologic calcification*. Finally, the normal process of *aging* is accompanied by characteristic morphologic and functional changes in cells.

This chapter discusses first the causes, mechanisms, and consequences of the various forms of cell damage, including reversible cell injury and cell death. We conclude with cellular adaptations to stress, and three other processes that affect

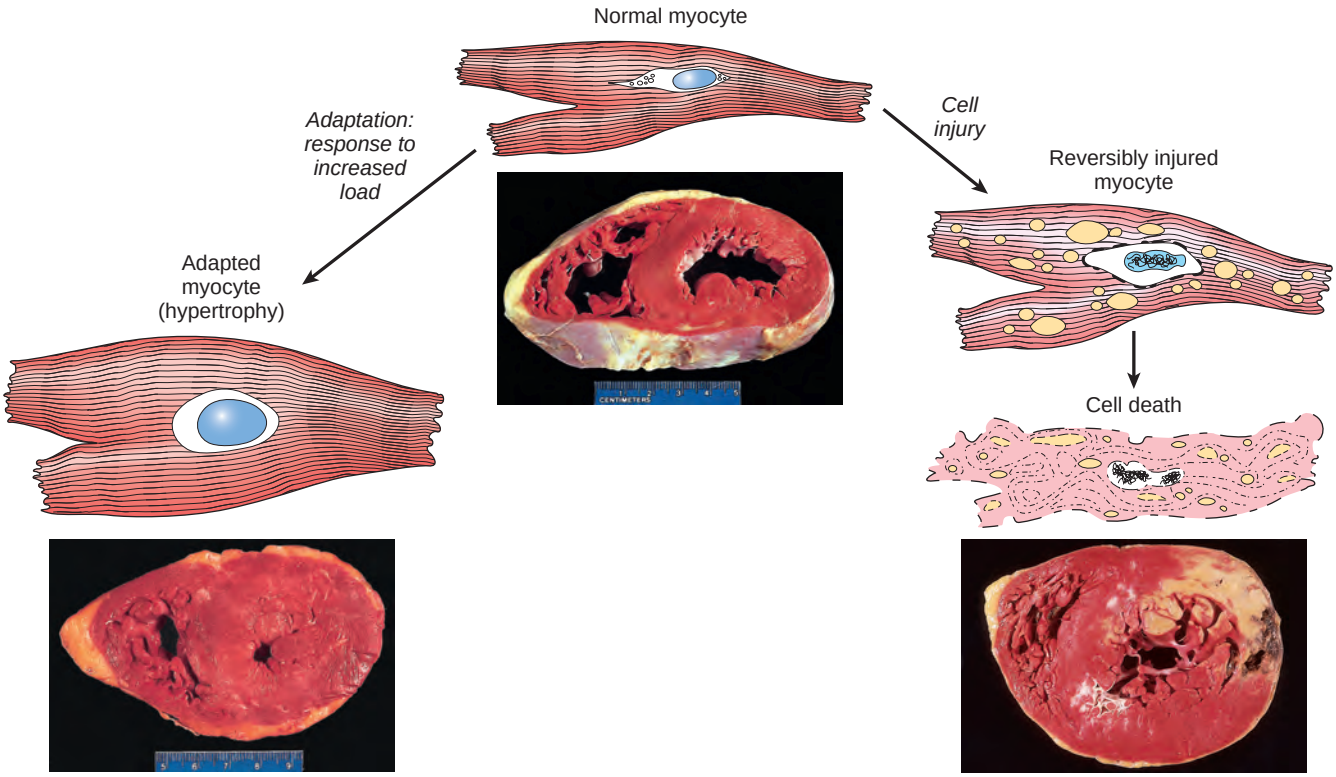


Figure 2.2 The relationship among normal, adapted, reversibly injured, and dead myocardial cells. All three transverse sections of the heart have been stained with triphenyltetrazolium chloride, an enzyme substrate that colors viable myocardium magenta. The cellular adaptation shown here is myocardial hypertrophy *(lower left)*, caused by increased blood pressure requiring greater mechanical effort by myocardial cells. This adaptation leads to thickening of the left ventricular wall (compare with the normal heart). In reversibly injured myocardium (illustrated schematically, *right*), there are functional alterations, usually without any gross or microscopic changes but sometimes with cytoplasmic changes such as cellular swelling and fat accumulation. In the specimen showing necrosis, a form of cell death *(lower right)*, the light area in the posterolateral left ventricle represents an acute myocardial infarction caused by reduced blood flow (ischemia).

cells and tissues: intracellular accumulations, pathologic calcification, and cell aging.

Causes of Cell Injury

The causes of cell injury range from the mechanical trauma of an automobile accident to subtle cellular abnormalities, such as a mutation causing lack of a vital enzyme that impairs normal metabolic function. Most injurious stimuli can be grouped into the following broad categories.

Oxygen Deprivation

Hypoxia is a deficiency of oxygen, which causes cell injury by reducing aerobic oxidative respiration. Hypoxia is an extremely important and common cause of cell injury and cell death. Causes of hypoxia include reduced blood flow *(ischemia);* inadequate oxygenation of the blood due to cardiorespiratory failure; and decreased oxygen-carrying capacity of the blood, as in anemia or carbon monoxide poisoning and severe blood loss. Depending on the severity of the hypoxic state, cells may adapt, undergo injury, or die. For example, if an artery is narrowed, the tissue supplied by that vessel may initially shrink in size (atrophy), whereas more severe or sudden hypoxia induces cell injury and cell death.

Physical Agents

Physical agents capable of causing cell injury include mechanical trauma, extremes of temperature (burns and deep cold), sudden changes in atmospheric pressure, radiation, and electric shock (Chapter 9).

Chemical Agents and Drugs

The list of chemicals that may produce cell injury defies compilation. Simple chemicals such as glucose or salt in hypertonic concentrations may cause cell injury directly or by deranging electrolyte and fluid balance in cells. Even oxygen at high concentrations is toxic. Trace amounts of *poisons,* such as arsenic, cyanide, or mercury, may damage sufficient numbers of cells within minutes or hours to cause death. Other potentially injurious substances are our daily companions: environmental pollutants, insecticides, and herbicides; industrial and occupational hazards, such as carbon monoxide and asbestos; recreational drugs such as alcohol; and the ever increasing variety of therapeutic drugs, many of which have toxic side effects. These are discussed further in Chapter 9.

Infectious Agents

These agents range from submicroscopic viruses to tapeworms several feet in length. In between are rickettsiae, bacteria, fungi, and higher forms of parasites. The ways by which these biologic agents cause injury are diverse (Chapter 8).

Immunologic Reactions

The immune system serves an essential function in defense against infectious pathogens, but immune reactions may also cause cell injury. Injurious reactions to endogenous self antigens are responsible for autoimmune diseases (Chapter 6). Immune reactions to many external agents, such as viruses and environmental substances, are also important causes of cell and tissue injury (Chapters 3 and 6).

Genetic Abnormalities

As described in Chapter 5, genetic aberrations as extreme as an extra chromosome, as in Down syndrome, or as subtle as a single base pair substitution leading to an amino acid substitution, as in sickle cell anemia, may produce highly characteristic clinical phenotypes ranging from congenital malformations to anemias. Genetic defects may cause cell injury because of deficient protein function, such as enzyme defects in inborn errors of metabolism, or accumulation of damaged DNA or misfolded proteins, both of which trigger cell death when they are beyond repair. DNA sequence variants that are common in human populations (polymorphisms) can also influence the susceptibility of cells to injury by chemicals and other environmental insults.

Nutritional Imbalances

Nutritional imbalances continue to be major causes of cell injury. Protein-calorie deficiencies cause an appalling number of deaths, chiefly among lower-income populations. Deficiencies of specific vitamins are found throughout the world (Chapter 9). Nutritional shortages can be self-imposed, as in anorexia nervosa (a psychological disorder of inadequate food consumption) or stem from food shortages or poor diet. Ironically, nutritional excess also is an important cause of cell injury. Obesity is rampant in the United States and is associated with an increased incidence of several important diseases, such as diabetes and cancer. In addition to the problems of undernutrition and overnutrition, the composition of the diet makes a significant contribution to a number of diseases. For example, diets high in certain lipids lead to elevated serum cholesterol and predispose to atherosclerosis, a leading risk factor for cardiovascular disease, the number one killer of adults in the United States.

The Progression of Cell Injury and Death

It is useful to describe the basic alterations that occur in damaged cells before discussing the mechanisms that bring about these changes. All stresses and noxious influences exert their effects first at the molecular or biochemical level. There is a time lag between the stress and the morphologic changes of cell injury or death; understandably, the early changes are subtle and are only detected with highly sensitive methods of examination (Fig. 2.3). With histochemical, ultrastructural, or biochemical techniques, changes may be seen in minutes to hours after injury, whereas changes visible by light microscopy or the naked eye may take considerably longer (hours to days) to appear. As would be expected, the morphologic manifestations of necrosis take more time to develop than those of reversible damage. For example, in ischemia of the myocardium, cell swelling is a reversible morphologic change that may occur in a matter of minutes, and is an indicator of ongoing cellular damage that may progress to irreversibility within 1 or 2 hours. Unmistakable light microscopic evidence of cell death, however, may not be seen until 4 to 12 hours after onset of ischemia.

The sequential structural changes in cell injury progressing to cell death are illustrated in Fig. 2.4 and described later.

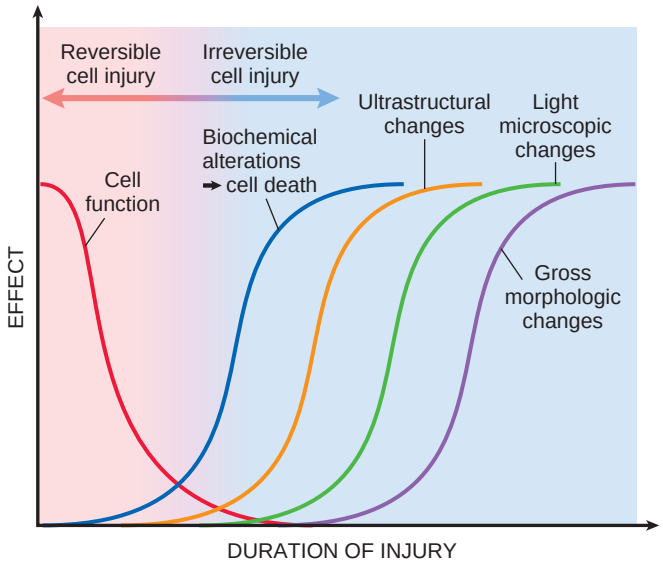


Figure 2.3 Sequential development of biochemical and morphologic changes in cell injury. Cells may become rapidly nonfunctional after the onset of injury, although they may still be viable, with potentially reversible damage; a longer duration of injury may lead to irreversible damage and cell death. Note that irreversible biochemical alterations may cause cell death, and typically this precedes ultrastructural, light microscopic, and grossly visible morphologic changes.

Within limits, the cell can repair the alterations seen in reversible injury and if the injurious stimulus abates, may return to normalcy. Persistent or excessive injury, however, causes cells to pass the rather nebulous "point of no return" into irreversible injury and cell death. Different injurious stimuli induce death mainly by necrosis and/or apoptosis (see Fig. 2.4 and Table 2.1).

REVERSIBLE CELL INJURY

Reversible cell injury is characterized by functional and structural alterations in early stages or mild forms of injury, which are correctable if the damaging stimulus is removed. Two features are consistently seen in reversibly injured cells.

- Early alterations in reversible injury include *generalized swelling of the cell* and its organelles, blebbing of the plasma membrane, detachment of ribosomes from the endoplasmic reticulum (ER), and clumping of nuclear chromatin. Swelling of cells results from influx of water. This is usually caused by failure of the adenosine triphosphate (ATP)-dependent Na^+-K^+ plasma membrane pump due to depletion of ATP resulting from oxygen deficiency, which interferes with mitochondrial oxidative phosphorylation, or mitochondrial damage by radiation or toxins (discussed later).
- *Fatty change* occurs in organs that are actively involved in lipid metabolism (e.g., liver). It results when toxic injury disrupts metabolic pathways and leads to rapid accumulation of triglyceride-filled lipid vacuoles. This is discussed in Chapter 18.
- Other alterations are described in the following sections.

Table 2.1 Features of Necrosis and Apoptosis

Feature	Necrosis	Apoptosis
Cell size	Enlarged (swelling)	Reduced (shrinkage)
Nucleus	Pyknosis, karyorrhexis, karyolysis	Fragmentation into nucleosome-size fragments
Plasma membrane	Disrupted	Intact; altered structure, especially orientation of lipids
Cellular contents	Enzymatic digestion; may leak out of cell	Intact; may be released in apoptotic bodies
Adjacent inflammation	Frequent	No
Physiologic or pathologic role	Usually pathologic (culmination of irreversible cell injury)	Often physiologic, means of eliminating unwanted cells; may be pathologic after some forms of cell injury, especially DNA damage

MORPHOLOGY

Cellular swelling is the earliest manifestation of almost all forms of injury to cells (Fig. 2.5B). When it affects many cells, it causes pallor, increased turgor, and increased weight of the affected organ. On microscopic examination, small clear vacuoles may be seen within the cytoplasm; these represent distended and pinched-off segments of the ER. This pattern of nonlethal injury is sometimes called **hydropic change** or **vacuolar degeneration.** The cytoplasm of injured cells appears red (eosinophilc) when stained with hematoxylin and eosin (H&E) due to loss of RNA, which binds the blue hematoxylin dye. The eosinophilia becomes more pronounced with progression toward necrosis.

The ultrastructural changes of reversible cell injury, visible by electron microscopy (Fig. 2.6B), include the following:

1. Plasma membrane alterations, such as blebbing, blunting, and loss of microvilli
2. Mitochondrial changes, including swelling and the appearance of small amorphous densities
3. Accumulation of "myelin figures" in the cytosol composed of phospholipids derived from damaged cellular membranes
4. Dilation of the ER, with detachment of polysomes
5. Nuclear alterations, with disaggregation of granular and fibrillar elements

CELL DEATH

There are two principal types of cell death, necrosis and apoptosis, which differ in their mechanisms, morphology, and roles in physiology and disease (see Table 2.1). Severe mitochondrial damage with depletion of ATP and rupture of lysosomal and plasma membranes are typically associated with necrosis. Necrosis occurs in many commonly encountered injuries, such as those following ischemia, exposure to toxins, various infections, and trauma. Apoptosis has many unique features (see later).

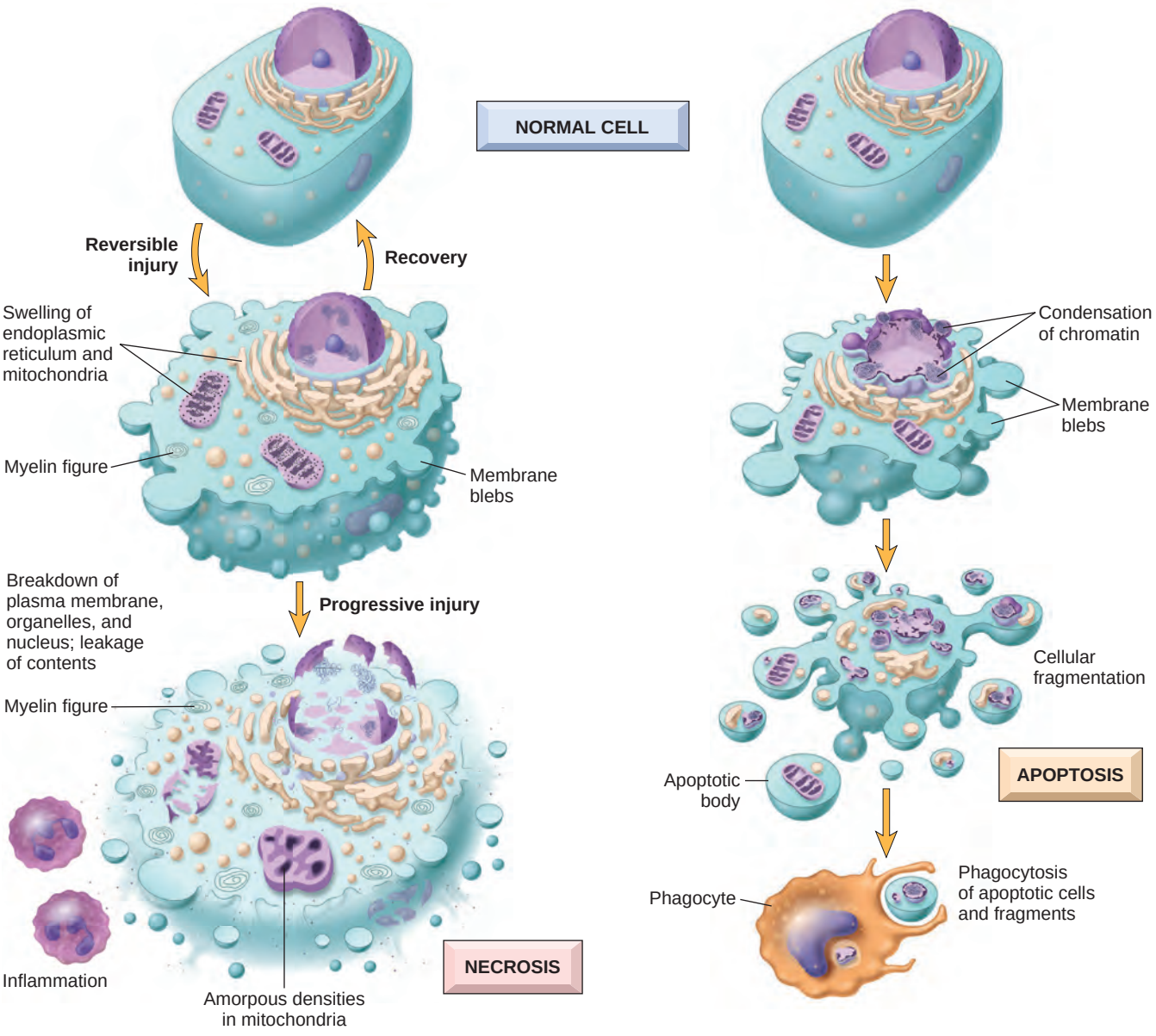


Figure 2.4 Schematic illustration of the morphologic changes in cell injury culminating in necrosis or apoptosis.

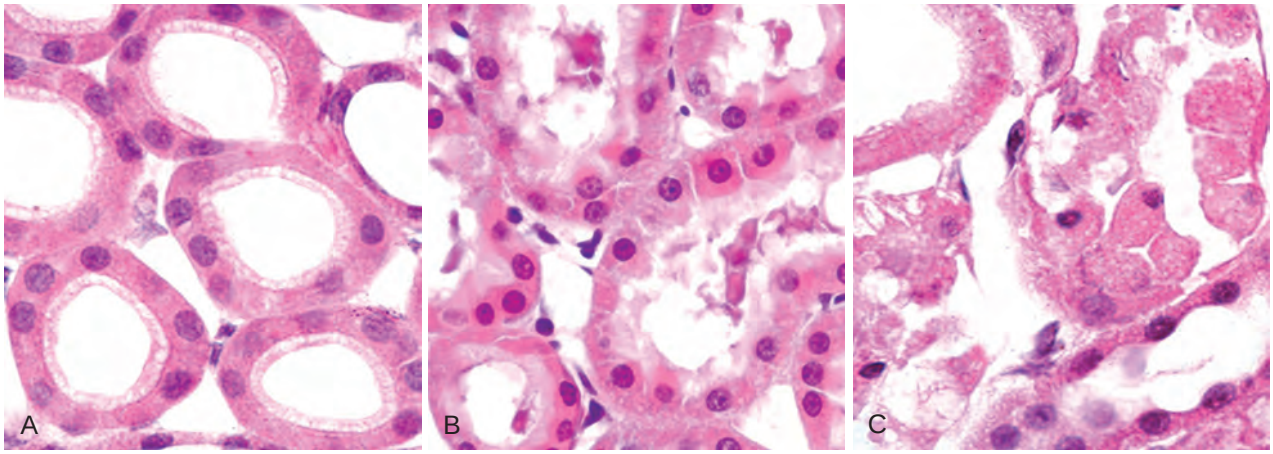


Figure 2.5 Morphologic changes in reversible cell injury and necrosis. (A) Normal kidney tubules with viable epithelial cells. (B) Early (reversible) ischemic injury showing surface blebs, increased eosinophilia of cytoplasm, and swelling of occasional cells. (C) Necrosis (irreversible injury) of epithelial cells, with loss of nuclei, fragmentation of cells, and leakage of contents. The ultrastructural features of these stages of cell injury are shown in Fig. 2.6. (Courtesy Drs. Neal Pinckard and M.A. Venkatachalam, University of Texas Health Sciences Center, San Antonio, Tex.)

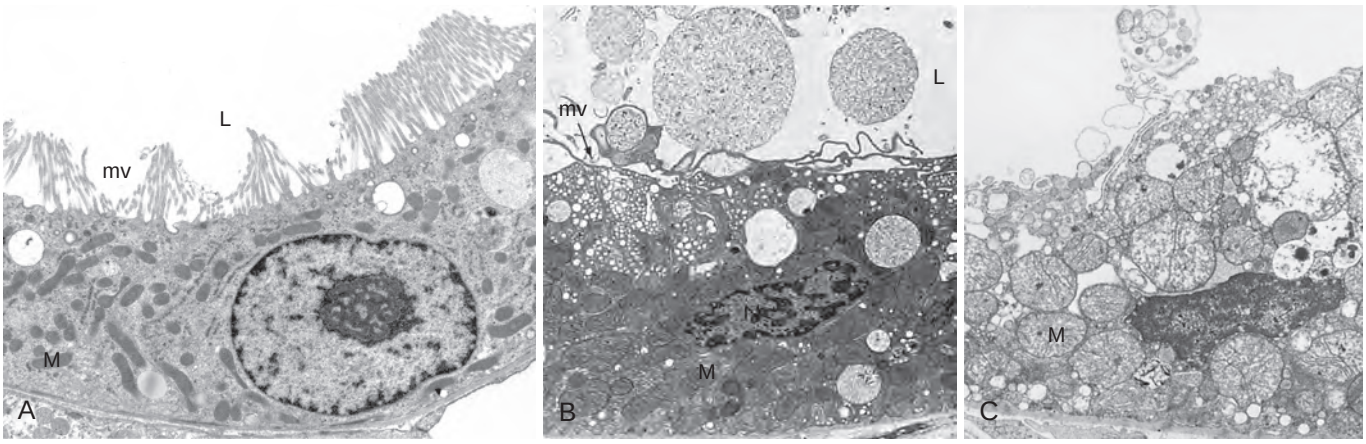


Figure 2.6 Ultrastructural features of reversible and irreversible cell injury (necrosis) in a rabbit kidney. (A) Electron micrograph of a normal epithelial cell of the proximal kidney tubule. Note abundant microvilli *(mv)* lining the luminal surface *(L)*. (B) Epithelial cell of the proximal tubule showing early cell injury resulting from reperfusion following ischemia. The microvilli are lost and have been incorporated in apical cytoplasm; blebs have formed and are extruded in the lumen. Mitochondria *(M)* would have been swollen during ischemia; with reperfusion, they rapidly undergo condensation and become electron-dense. (C) Proximal tubular cell showing late injury, expected to be irreversible. Note the markedly swollen mitochondria containing electron-dense deposits, expected to contain precipitated calcium and proteins. Higher magnification micrographs of the cell would show disrupted plasma membrane and swelling and fragmentation of organelles. (A, Courtesy Dr. Brigitte Kaisslin, Institute of Anatomy, University of Zurich, Switzerland. B, C, Courtesy Dr. M.A. Venkatachalam, University of Texas Health Sciences Center, San Antonio, Tex.)

Necrosis has historically been thought of as "accidental" cell death, reflecting severe injury that irreparably damages so many cellular components that the cell simply "falls apart." If early (reversible) injury progresses because the injurious stimulus persists, the end result is death by necrosis. When cells die by necrosis, there is a local inflammatory response that clears the "scene of the accident." In contrast, apoptosis is "regulated" cell death, mediated by defined molecular pathways that are activated under specific circumstances and kill cells with surgical precision, without inflammation or the associated collateral damage. The separation of necrosis and apoptosis is not always so clear, however, and some forms of necrosis are genetically controlled through a defined molecular pathway, called "necroptosis" (discussed later). Moreover, in some situations, cell death may show morphologic features of both apoptosis and necrosis, or progress from one to the other, so the distinctions may not be absolute. Nevertheless, it is useful to consider these as largely non-overlapping pathways of cell death because their principal mechanisms, morphological features, and functional consequences are usually different.

Necrosis

Necrosis is a pathologic process that is the consequence of severe injury. The main causes of necrosis include loss of oxygen supply (ischemia), exposure to microbial toxins, burns and other forms of chemical and physical injury, and unusual situations in which active proteases leak out of cells and damage surrounding tissues (as in pancreatitis). All of these initiating triggers lead to irreparable damage of numerous cellular components.

Necrosis is characterized by denaturation of cellular proteins, leakage of cellular contents through damaged membranes, local inflammation, and enzymatic digestion of the lethally injured cell. When damage to membranes is severe, lysosomal enzymes enter the cytoplasm and digest the cell. Cellular contents also leak through the damaged plasma membrane into the extracellular space, where they elicit a host reaction (inflammation). Some specific substances released from injured cells have been called *damage-associated molecular patterns (DAMPs)*. These include ATP (released from damaged mitochondria), uric acid (a breakdown product of DNA), and numerous other molecules that are normally confined within healthy cells and whose release is an indicator of severe cell injury. These molecules are recognized by receptors present in macrophages and most other cell types, and trigger phagocytosis of the debris as well as the production of cytokines that induce inflammation (Chapter 3). Inflammatory cells produce more proteolytic enzymes, and the combination of phagocytosis and enzymatic digestion usually leads to clearance of the necrotic cells.

Necrosis-associated leakage of intracellular proteins through damaged plasma membranes and ultimately into the circulation is the basis for blood tests that detect tissue-specific cellular injury. Cardiac muscle cells, for example, express cardiac-specific variants of the contractile protein troponin, while bile duct epithelium expresses a specific isoform of the enzyme alkaline phosphatase and hepatocytes express transaminases. Necrosis of these cell types and associated loss of membrane integrity is reflected in increased serum levels of these proteins, which serve as biomarkers that are used clinically to assess and quantify tissue damage. Cardiac-specific troponins can be detected in the blood as early as 2 hours after myocardial cell necrosis, well before histologic evidence of myocardial infarction becomes apparent. Because of their sensitivity and specificity, serial measurement of serum cardiac troponins has a central role in the diagnosis and management of patients with myocardial infarction (Chapter 12).

MORPHOLOGY

Necrotic cells show **increased eosinophilia** in H&E stains, attributable in part to the loss of cytoplasmic RNA and in part to accumulation of denatured cytoplasmic proteins (which bind the red dye eosin). The necrotic cell may have a glassy homogeneous appearance relative to normal cells, mainly as a result of the loss of glycogen particles (see Fig. 2.5C). When enzymes have digested the cell's organelles, the cytoplasm becomes vacuolated and appears moth-eaten. Dead cells may be replaced by large whorled phospholipid precipitates called **myelin figures**, which are either phagocytosed by other cells or further degraded into fatty acids; calcification of such fatty acid residues results in deposition of calcium-rich precipitates. By electron microscopy, necrotic cells are characterized by discontinuities in plasma and organelle membranes, marked dilation of mitochondria with the appearance of large amorphous densities, intracytoplasmic myelin figures, amorphous debris, and aggregates of fluffy material representing denatured protein (see Fig. 2.6C).

Nuclear changes appear in one of three patterns, all due to breakdown of DNA. The basophilia of the chromatin may fade **(karyolysis),** a change that presumably reflects loss of DNA because of enzymatic degradation by endonucleases. A second pattern (also seen in apoptotic cell death) is **pyknosis,** characterized by nuclear shrinkage and increased basophilia. Here the chromatin condenses into a dense, shrunken basophilic mass. In the third pattern, known as **karyorrhexis,** the pyknotic nucleus undergoes fragmentation. With the passage of time (1 or 2 days), the nucleus in the necrotic cell totally disappears.

It is useful to consider the possible events that determine when reversible injury becomes irreversible and progresses to necrosis. The clinical relevance of this question is obvious—if we can answer it, we may be able to devise strategies for preventing cell injury from having permanent deleterious consequences. Although the "point of no return," at which the damage becomes irreversible, is still largely undefined, *two phenomena consistently characterize irreversibility—the inability to reverse mitochondrial dysfunction* (lack of oxidative phosphorylation and ATP generation) even after resolution of the original injury, and *profound disturbances in membrane function*. As mentioned earlier, injury to lysosomal membranes results in the enzymatic dissolution of the injured cell that is characteristic of necrosis.

Patterns of Tissue Necrosis

The discussion of necrosis has focused so far on changes in individual cells. When large numbers of cells die, the tissue or organ is said to be necrotic; thus, a myocardial infarct is necrosis of a portion of the heart caused by death of many myocardial cells. Necrosis of tissues has several morphologically distinct patterns, which are important to recognize because they provide clues about the underlying cause. Although the terms that describe these patterns are somewhat dated, they are often used and their implications are understood by pathologists and clinicians.

MORPHOLOGY

Coagulative necrosis is a form of necrosis in which the architecture of dead tissue is preserved for a span of at least some days (Fig. 2.7). The affected tissue has a firm texture. Presumably, the injury denatures not only structural proteins but also enzymes and so blocks the proteolysis of the dead cells; as a result, intensely eosinophilic cells with indistinct or reddish nuclei may persist for days or weeks. Ultimately, the necrotic cells are broken down by the action of lysosomal enzymes derived from infiltrating leukocytes, which also remove the debris of the dead cells by phagocytosis. Ischemia caused by obstruction in a vessel may lead to coagulative necrosis of the supplied tissue in all organs except the brain (see next paragraph for explanation). A localized area of coagulative necrosis is called an **infarct**.

Liquefactive necrosis, in contrast to coagulative necrosis, is characterized by digestion of the dead cells, resulting in transformation of the tissue into a viscous liquid. It is seen in focal bacterial or, occasionally, fungal infections, because microbes stimulate the accumulation of leukocytes and the liberation of enzymes from these cells. The necrotic material is frequently creamy yellow because of the presence of leukocytes and is called **pus.** For

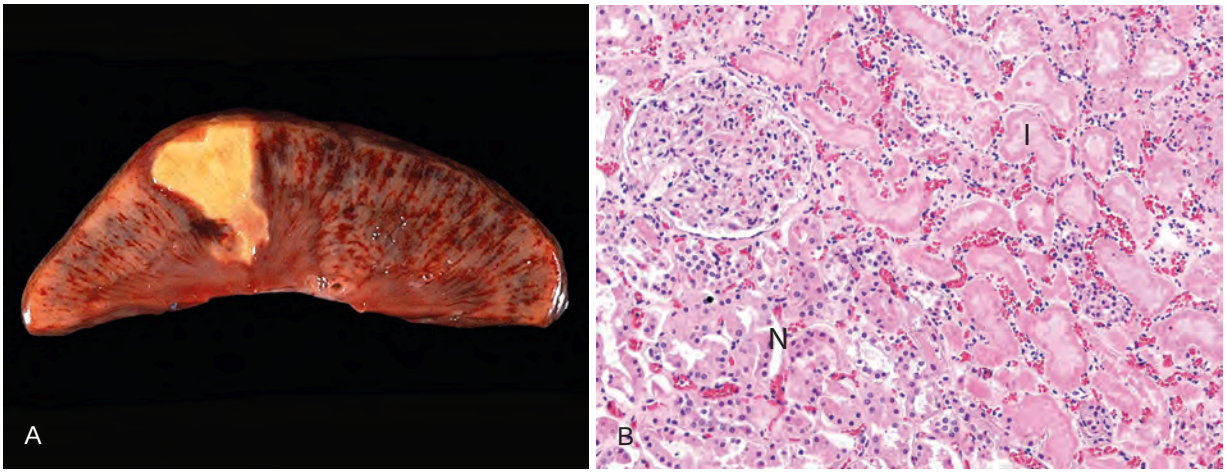


Figure 2.7 Coagulative necrosis. (A) A wedge-shaped kidney infarct *(yellow)*. (B) Microscopic view of the edge of the infarct, with normal kidney *(N)* and necrotic cells in the infarct *(I)* showing preserved cellular outlines with loss of nuclei and an inflammatory infiltrate (seen as nuclei of inflammatory cells in between necrotic tubules).

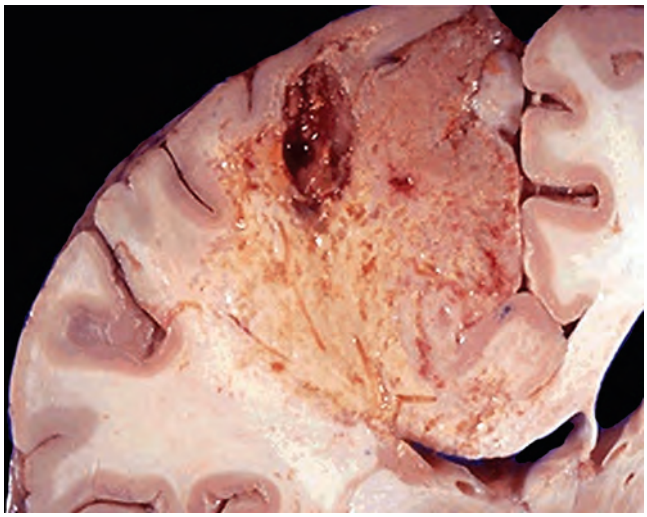


Figure 2.8 Liquefactive necrosis. An infarct in the brain, showing dissolution of the tissue.

unknown reasons, hypoxic death of cells within the central nervous system often manifests as liquefactive necrosis (Fig. 2.8).

Gangrenous necrosis is not a specific pattern of cell death, but the term is commonly used in clinical practice. It is usually applied to a limb, generally the lower leg, that has lost its blood supply and has undergone necrosis (typically coagulative necrosis) involving multiple tissue planes. When bacterial infection is superimposed, there is more liquefactive necrosis because of the actions of degradative enzymes in the bacteria and the attracted leukocytes (giving rise to so-called **wet gangrene**).

Caseous necrosis is encountered most often in foci of tuberculous infection (Chapter 8). The term *caseous* (cheeselike) is derived from the friable white appearance of the area of necrosis (Fig. 2.9). On microscopic examination, the necrotic area appears as a structureless collection of fragmented or lysed cells and amorphous granular debris enclosed within a distinctive inflammatory border; this appearance is characteristic of a focus of inflammation known as a **granuloma** (Chapter 3).

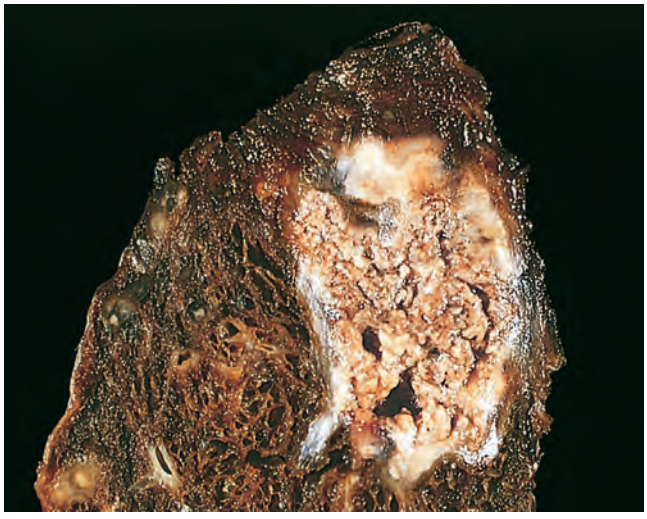


Figure 2.9 Caseous necrosis. Tuberculosis of the lung, with a large area of caseous necrosis containing yellow-white and "cheesy" appearing debris.

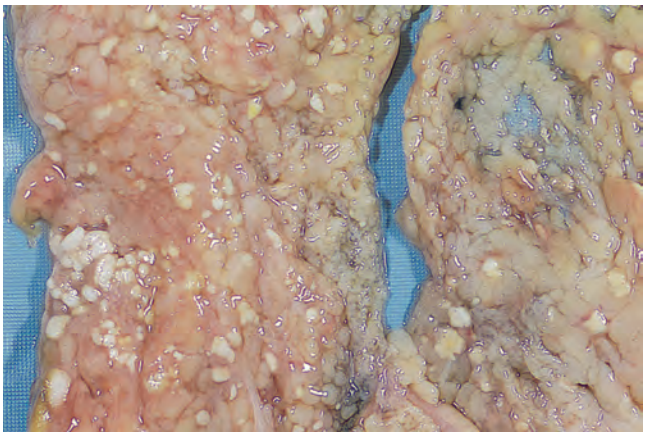


Figure 2.10 Fat necrosis. The areas of white chalky deposits represent foci of fat necrosis with calcium soap formation (saponification) at sites of lipid breakdown in the mesentery.

Fat necrosis refers to focal areas of fat destruction, typically resulting from release of activated pancreatic lipases into the substance of the pancreas and the peritoneal cavity. This occurs in the calamitous abdominal emergency known as acute pancreatitis (Chapter 19). In this disorder, pancreatic enzymes leak out of damaged acinar cells and liquefy the membranes of fat cells in the peritoneum, releasing triglyceride esters that are split by pancreatic lipases. Fatty acids are generated that combine with calcium to produce grossly visible chalky-white areas (fat saponification), which enable the surgeon and the pathologist to identify the underlying disorder (Fig. 2.10). On histologic examination, the necrotic areas contain the shadowy outlines of necrotic fat cells, basophilic calcium deposits, and an inflammatory reaction.

Fibrinoid necrosis is a special form of vascular damage usually seen in immune reactions involving blood vessels. It typically occurs when complexes of antigens and antibodies are deposited in the walls of arteries. Deposits of these immune complexes, together with plasma proteins that has leaked out of vessels, result in a bright pink and amorphous appearance in H&E stains called "fibrinoid" (fibrin-like) by pathologists (Fig. 2.11). The immunologically mediated vasculitis syndromes in which this type of vascular injury is seen are described in Chapter 11.

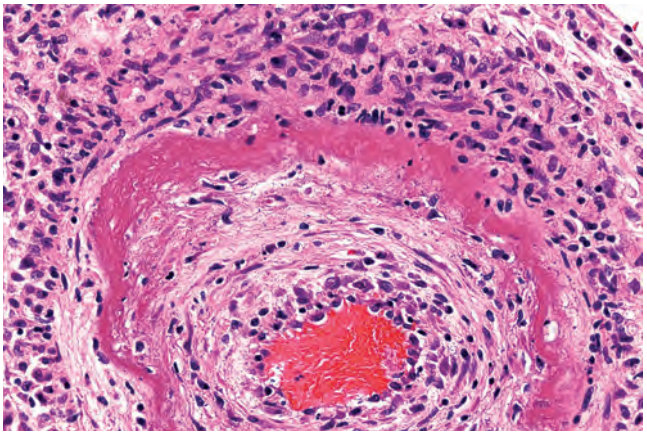


Figure 2.11 Fibrinoid necrosis in an artery. The wall of the artery shows a circumferential bright pink area of necrosis with inflammation (neutrophils with dark nuclei).

Ultimately, in the living patient most necrotic cells and their contents disappear due to enzymatic digestion and phagocytosis of the debris by leukocytes. If necrotic cells and cellular debris are not promptly destroyed and reabsorbed, they provide a nidus for the deposition of calcium salts and other minerals and thus tend to become calcified. This phenomenon, called *dystrophic calcification,* is considered later in the chapter.

KEY CONCEPTS

CELL INJURY AND NECROSIS

- Exposure of cells to stress or noxious agents causes cell injury that is reversible to a point but may progress to death of the cells, principally by necrosis.
- Reversible cell injury: Characterized by cellular swelling, fatty change, plasma membrane blebbing and loss of microvilli, mitochondrial swelling, dilation of the ER, and eosinophilia (due to decreased cytoplasmic RNA)
- Necrosis: A pathologic process in which cellular membranes are destroyed, enzymes and other constituents leak out, and local inflammation is induced to clear the damaged cells. Morphologic features are: eosinophilia; nuclear shrinkage, fragmentation, and dissolution; breakdown of plasma membrane and organellar membranes; abundant myelin figures; and leakage and enzymatic digestion of cellular contents
- Patterns of tissue necrosis: Under different conditions, necrosis in tissues may assume specific patterns: coagulative, liquefactive, gangrenous, caseous, fat, and fibrinoid

Apoptosis

Apoptosis is a type of cell death that is induced by a tightly regulated suicide program in which cells destined to die activate intrinsic enzymes that degrade the cells' genomic DNA and nuclear and cytoplasmic proteins. Apoptotic cells break up into plasma membrane–bound fragments, called *apoptotic bodies,* which contain portions of the cytoplasm and nucleus. While the plasma membrane remains intact, its surface components are altered so as to produce "find me" and "eat me" signals for phagocytes, discussed later. As a result, the dead cell and its fragments are rapidly devoured, before the contents leak out, and therefore apoptosis does not elicit an inflammatory reaction. Apoptosis was first recognized in 1972 by the distinctive morphologic appearance of membrane-bound fragments derived from cells, and named after the Greek designation for "falling off." It was subsequently discovered in model organisms such as worms that certain cells undergo apoptosis at precise times during development. This phenomenon, termed *programmed cell death*, is controlled by the action of a small number of genes and is required for normal embryogenesis. Thus, apoptosis is a unique mechanism of cell death, distinct from necrosis in many respects (see Fig. 2.4 and Table 2.1).

Causes of Apoptosis

Apoptosis occurs in two broad contexts: as part of normal physiologic processes, and as a pathophysiologic mechanism of cell loss in many different diseases.

Apoptosis in Physiologic Situations

Death by apoptosis is a normal phenomenon that serves to eliminate cells that are no longer needed, or as a mechanism to maintain a constant number of various cell populations in tissues. It is estimated that humans turn over almost 1 million cells per second! Central to this process is death of cells by apoptosis and their removal by phagocytes. Apoptosis is important in the following physiologic situations:

- *The removal of supernumerary cells (in excess of the required number) during development*. Cell death is critical for involution of primordial structures and remodeling of maturing tissues. Apoptosis is a generic term for this pattern of cell death, regardless of the context, while programmed cell death refers only to apoptosis during development.
- *Involution of hormone-dependent tissues on hormone withdrawal,* such as endometrial cell breakdown during the menstrual cycle, ovarian follicular atresia in menopause, and regression of the lactating breast after weaning.
- *Cell turnover in proliferating cell populations,* such as immature lymphocytes in the bone marrow and thymus, B lymphocytes in germinal centers that fail to express useful antigen receptors (Chapter 6), and epithelial cells in intestinal crypts, to maintain a constant cell number *(homeostasis)*.
- *Elimination of potentially harmful self-reactive lymphocytes* to prevent immune reactions against one's own tissues (Chapter 6).
- Death of host cells that have served their useful purpose, such as neutrophils in an *acute inflammatory response,* and lymphocytes at the end of an *immune response*.

In all of these situations, cells undergo apoptosis because they are deprived of necessary survival signals, such as growth factors and interactions with the extracellular matrix, or they receive pro-apoptotic signals from other cells or the surrounding environment.

Apoptosis in Pathologic Conditions

Apoptosis eliminates cells that are injured beyond repair without eliciting a host reaction, thus limiting collateral tissue damage. Death by apoptosis is responsible for loss of cells in a variety of pathologic states:

- *DNA damage*. Radiation and cytotoxic anticancer drugs can damage DNA, either directly or via production of free radicals. If repair mechanisms cannot correct the damage, the cell triggers intrinsic mechanisms that induce apoptosis. In these situations, apoptosis has a protective effect by preventing the survival of cells with DNA mutations that can lead to malignant transformation.
- *Accumulation of misfolded proteins*. Cell death triggered by improperly folded intracellular proteins and the subsequent endoplasmic reticulum (ER) stress response is discussed later.
- Apoptosis can be induced during certain *infections,* particularly viral infections, as a result of the virus itself (as in adenovirus and HIV infections) or the host immune response (as in viral hepatitis). An important host response to viruses consists of cytotoxic T lymphocytes

(CTLs) specific for viral proteins, which induce apoptosis of infected cells in an attempt to eliminate reservoirs of infection. During this process, there can be significant tissue damage. The same CTL-mediated mechanism is responsible for killing of *tumor cells,* cellular rejection of *transplants,* and tissue damage in graft-versus-host disease.

- Apoptosis may also contribute to *pathologic atrophy in parenchymal organs after duct obstruction,* such as occurs in the pancreas, parotid gland, and kidney.

Morphologic and Biochemical Changes in Apoptosis

Before discussing underlying mechanisms, the morphologic and biochemical characteristics of apoptosis are described.

MORPHOLOGY

The following morphologic features, some best seen with the electron microscope, characterize cells undergoing apoptosis (Fig. 2.12; see Fig. 2.4).

Cell shrinkage. Cell size is reduced, the cytoplasm is dense and eosinophilic (see Fig. 2.12A), and the organelles, although relatively normal, are more tightly packed. This contrasts with necrosis, in which an early feature is cell swelling, not shrinkage.

Chromatin condensation. This is the most characteristic feature of apoptosis. The chromatin aggregates peripherally, under the nuclear membrane, into dense masses of various shapes and sizes (see Fig. 2.12B). The nucleus itself may break up into two or more fragments.

Formation of cytoplasmic blebs and apoptotic bodies. The apoptotic cell first shows extensive surface membrane blebbing, which is followed by fragmentation of the dead cells into membrane-bound apoptotic bodies composed of cytoplasm and tightly packed organelles, with or without nuclear fragments (see Fig. 2.12C).

Phagocytosis of apoptotic cells or cell bodies, usually by macrophages. The apoptotic bodies are rapidly ingested by phagocytes and degraded by the phagocyte's lysosomal enzymes.

In H&E-stained tissue, the apoptotic cell appears as a round or oval mass of intensely eosinophilic cytoplasm with fragments of dense nuclear chromatin (see Fig. 2.12A). Because the cell shrinkage and formation of apoptotic bodies are rapid and the pieces are quickly cleared by phagocytes, considerable apoptosis may occur in tissues before it is apparent in histologic sections. The absence of an inflammatory response can also make it difficult to detect apoptosis by light microscopy.

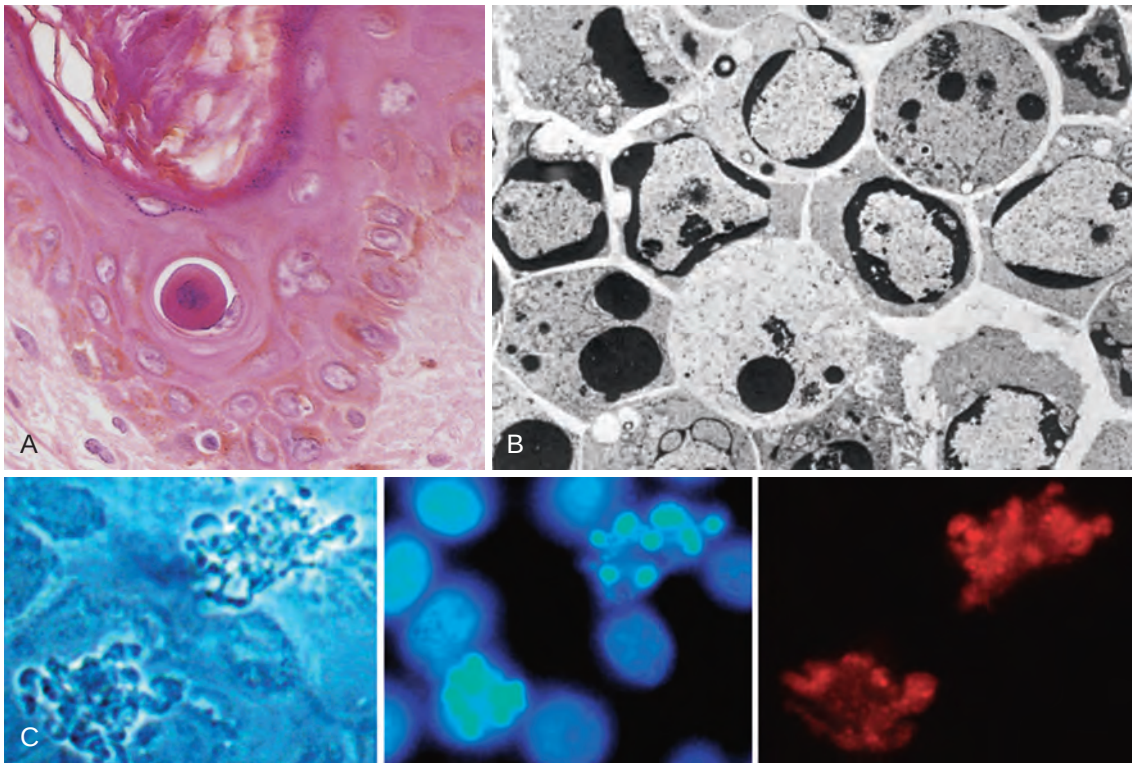


Figure 2.12 Morphologic features of apoptosis. (A) Apoptosis of an epidermal cell in an immune reaction. The cell is reduced in size and contains brightly eosinophilic cytoplasm and a condensed nucleus. (B) This electron micrograph of cultured cells undergoing apoptosis shows some nuclei with peripheral crescents of compacted chromatin, and others that are uniformly dense or fragmented. (C) These images of cultured cells undergoing apoptosis show blebbing and formation of apoptotic bodies (*left panel,* phase contrast micrograph), a stain for DNA showing nuclear fragmentation *(middle panel),* and activation of caspase-3 (*right panel,* immunofluorescence stain with an antibody specific for the active form of caspase-3, revealed as red color). (B, From Kerr JFR, Harmon BV: Definition and incidence of apoptosis: a historical perspective. In Tomei LD, Cope FO, editors: *Apoptosis: The Molecular Basis of Cell Death.* Cold Spring Harbor, NY, 1991, Cold Spring Harbor Laboratory Press, pp 5–29; C, Courtesy Dr. Zheng Dong, Medical College of Georgia, Augusta, Ga.)

Mechanisms of Apoptosis

Apoptosis results from the activation of enzymes called *caspases* (so named because they are proteases containing a *c*ysteine in their active site and cleave proteins after *asp*artic residues). Like many proteases, caspases exist as inactive proenzymes and must undergo enzymatic cleavage to become active. The presence of active caspases is therefore a marker for cells undergoing apoptosis (see Fig. 2.12C). The process of apoptosis may be divided into an *initiation phase,* during which some caspases become catalytically active and unleash a cascade of other caspases, and an *execution phase,* during which the terminal caspases trigger cellular fragmentation. Regulation of these enzymes depends on a finely tuned balance between the abundance and activity of pro-apoptotic and anti-apoptotic proteins.

Two distinct pathways converge on caspase activation: the mitochondrial pathway and the death receptor pathway (Fig. 2.13). Although these pathways intersect, they are generally induced under different conditions, involve different initiating molecules, and serve distinct roles in physiology and disease.

The Mitochondrial (Intrinsic) Pathway of Apoptosis

The mitochondrial pathway is responsible for apoptosis in most physiologic and pathologic situations. It results from increased permeability of the mitochondrial outer membrane with consequent release of death-inducing (pro-apoptotic) molecules from the mitochondrial intermembrane space into the cytoplasm (Fig. 2.14). Mitochondria are organelles that contain remarkable proteins such as cytochrome c, a double-edged sword that is essential for producing the energy (e.g., ATP) that sustains cell viability, but that when released into the cytoplasm (an indication that the cell is not healthy) initiates the suicide program of apoptosis. The release of pro-apoptotic proteins such as cytochrome c is determined by the integrity of the outer mitochondrial membrane, which is tightly controlled by the BCL2 family of proteins. This family is named after *BCL2,* a gene that is frequently overexpressed due to chromosomal translocations and other aberrations in certain B cell lymphomas (Chapter 13). There are more than 20 members of the BCL family, which can be divided into three groups based on their pro-apoptotic or anti-apoptotic function and the BCL2 homology (BH) domains they possess.

- *Anti-apoptotic.* BCL2, BCL-X_L, and MCL1 are the principal members of this group; they possess four BH domains (called BH1-4). These proteins reside in the outer mitochondrial membrane as well as in the cytosol and ER membranes. By keeping the mitochondrial outer membrane impermeable, they prevent leakage of cytochrome c and other death-inducing proteins into the cytosol (see Fig. 2.14A).

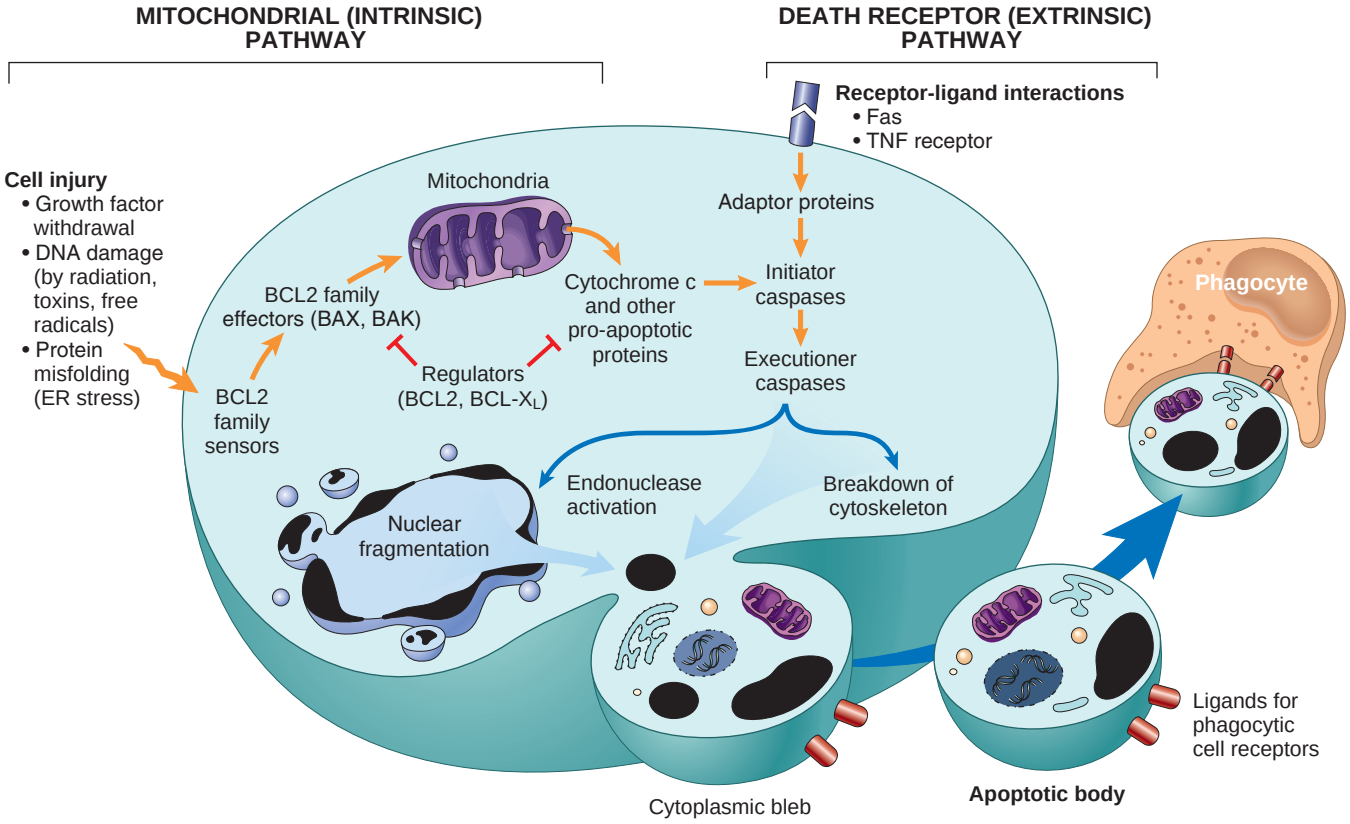


Figure 2.13 Mechanisms of apoptosis. Although the two pathways of apoptosis differ in their induction and regulation, they both culminate in the activation of caspases. In the mitochondrial pathway, proteins of the BCL2 family, which regulate mitochondrial permeability, become imbalanced such that the ratio of pro-apoptotic versus anti-apoptotic proteins results in the leakage of various substances from mitochondria that lead to caspase activation. In the death receptor pathway, signals from plasma membrane receptors lead to the assembly of adaptor proteins into a "death-inducing signaling complex," which activates caspases, and the end result is the same.

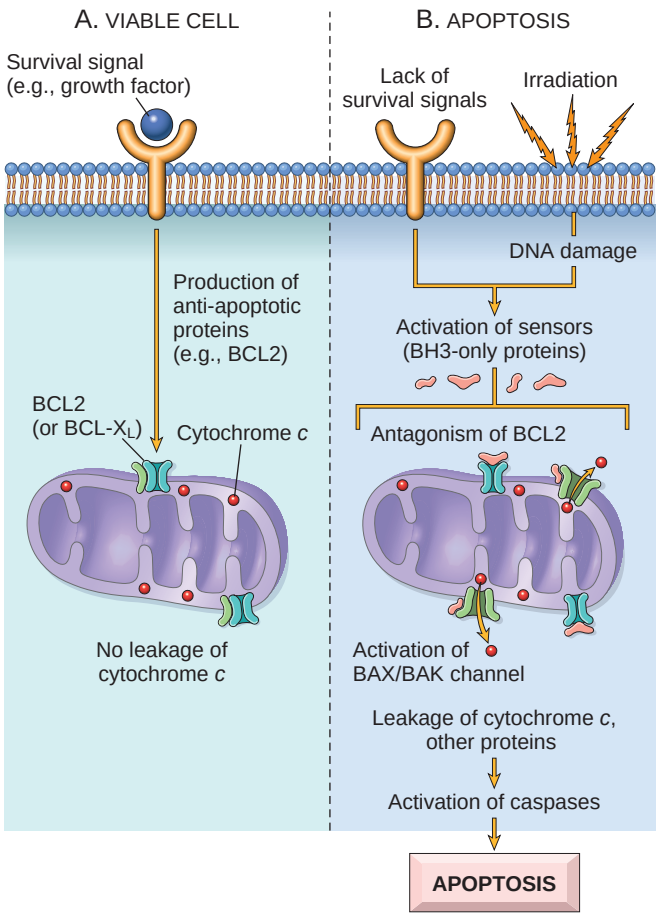


Figure 2.14 The intrinsic (mitochondrial) pathway of apoptosis. (A) Cell viability is maintained by the induction of anti-apoptotic proteins such as BCL2 by survival signals. These proteins maintain the integrity of mitochondrial membranes and prevent leakage of mitochondrial proteins. (B) Loss of survival signals, DNA damage, and other insults activate sensors that antagonize the anti-apoptotic proteins and activate the pro-apoptotic proteins BAX and BAK, which form channels in the mitochondrial membrane. The subsequent leakage of cytochrome c (and other proteins, *not shown*) leads to caspase activation and apoptosis.

- *Pro-apoptotic.* BAX and BAK are the two prototypic members of this group; they contain the first three BH domains (BH1-3). On activation, BAX and/or BAK oligomerize within the outer mitochondrial membrane and enhance its permeability. The precise mechanism by which BAX-BAK oligomers permeabilize membranes is not settled. According to one model illustrated in Fig. 2.14B, they form a channel in the outer mitochondrial membrane that allows cytochrome c leakage from the intermembranous space.
- *Regulated apoptosis initiators.* Members of this group, including BAD, BIM, BID, Puma, and Noxa, contain only one BH domain, the third of the four BH domains, and hence are sometimes called BH3-only proteins. The activity of BH3-only proteins is modulated by sensors of cellular stress and damage; when upregulated and activated, they can initiate apoptosis.

Growth factors and other survival signals stimulate the production of anti-apoptotic proteins such as BCL2, thus protecting cells from apoptosis. When cells are deprived of survival signals, suffer DNA damage, or develop ER stress due to the accumulation of misfolded proteins, BH3-only proteins are upregulated through increased transcription and/or post-translational modifications (e.g., phosphorylation). These BH3-only proteins in turn directly activate the two critical pro-apoptotic family members, BAX and BAK, which form oligomers that insert into the mitochondrial membrane and allow proteins from the inner mitochondrial membrane to leak out into the cytoplasm. BH3-only proteins may also bind to and block the function of BCL2 and BCL-X_L. At the same time, synthesis of BCL2 and BCL-X_L may decline because their transcription relies on survival signals. The net result of BAX-BAK activation coupled with loss of the protective functions of the anti-apoptotic BCL2 family members is the release into the cytoplasm of several mitochondrial proteins such as cytochrome c that can activate the caspase cascade (see Fig. 2.14).

Once released into the cytosol, cytochrome c binds to a protein called APAF-1 (apoptosis-activating factor-1), forming a multimeric structure called the *apoptosome*. This complex binds to caspase-9, the critical initiator caspase of the mitochondrial pathway, and promotes its autocatalytic cleavage, generating catalytically active forms of the enzyme. Active caspase-9 then triggers a cascade of caspase activation by cleaving and thereby activating other pro-caspases (such as caspase-3), which mediate the execution phase of apoptosis (discussed later). Other mitochondrial proteins with arcane names like Smac/DIABLO enter the cytoplasm, where they bind to and neutralize cytoplasmic proteins that function as physiologic inhibitors of apoptosis (IAPs). The normal function of the IAPs is to block the inappropriate activation of caspases, including executioners like caspase-3, and keep cells alive. Thus, IAP inhibition permits initiation of the caspase cascade.

The Extrinsic (Death Receptor–Initiated) Pathway of Apoptosis

This pathway is initiated by engagement of plasma membrane death receptors. Death receptors are members of the tumor necrosis factor (TNF) receptor family that contain a cytoplasmic domain involved in protein-protein interactions. This *death domain* is essential for delivering apoptotic signals. (Some TNF receptor family members do not contain cytoplasmic death domains; their function is to activate inflammatory cascades [Chapter 3], and their role in triggering apoptosis is much less established.) The best-known death receptors are the type 1 TNF receptor (TNFR1) and a related protein called Fas (CD95), but several others have been described. The mechanism of apoptosis induced by these death receptors is well illustrated by Fas, a death receptor expressed on many cell types (Fig. 2.15). The ligand for Fas is called Fas ligand (FasL). FasL is expressed on T cells that recognize self antigens (and functions to eliminate self-reactive lymphocytes that also express the receptor Fas upon recognition of self antigens) and on some CTLs that kill virus-infected and tumor cells. When FasL binds to Fas, three or more molecules of Fas are brought together, and their cytoplasmic death domains form a binding site for an adaptor protein called FADD (*F*as-*a*ssociated *d*eath *d*omain). Once attached to this complex, FADD binds inactive

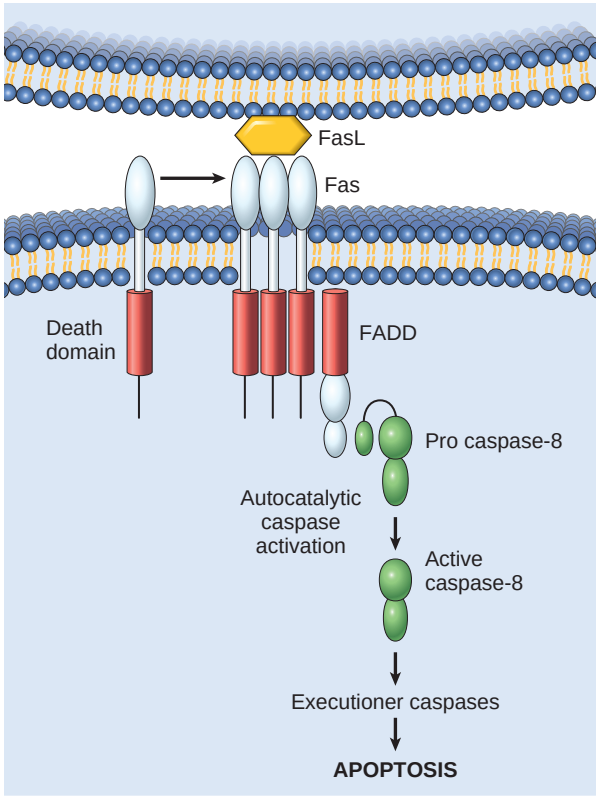


Figure 2.15 The extrinsic (death receptor–initiated) pathway of apoptosis, illustrated by the events following Fas engagement. *FADD,* Fas-associated death domain; *FasL,* Fas ligand.

caspase-8 (or caspase-10), bringing together multiple caspase molecules and leading to autocatalytic cleavage and generation of active caspase-8. In turn, active caspase-8 initiates the same executioner caspase sequence as in the mitochondrial pathway. This extrinsic apoptosis pathway can be inhibited by a protein called FLIP, which binds to procaspase-8, thereby blocking FADD binding, but cannot activate the caspase. Some viruses and normal cells produce FLIP as a mechanism to protect themselves from Fas-mediated apoptosis.

The extrinsic and intrinsic pathways of apoptosis are initiated in fundamentally different ways by distinct molecules, but there may be interconnections between them. For instance, in hepatocytes and pancreatic β cells, caspase-8 produced by Fas signaling cleaves and activates the BH3-only protein BID, which then feeds into the mitochondrial pathway. The combined activation of both pathways delivers a fatal blow to the cells.

The Execution Phase of Apoptosis

The intrinsic and extrinsic pathways converge to activate a caspase cascade that mediates the final phase of apoptosis. The intrinsic mitochondrial pathway activates the initiator caspase-9, whereas the extrinsic death receptor pathway activates caspase-8 and caspase-10. The active forms of these caspases trigger the rapid and sequential activation of the executioner caspases, such as caspase-3 and caspase-6, which then act on many cellular components. For instance, once activated these caspases cleave an inhibitor of a DNase, making the DNase enzymatically active and allowing DNA degradation to commence. Caspases also proteolyze structural components of the nuclear matrix and thus promote fragmentation of nuclei. Other steps in apoptosis are less well-defined. For instance, we do not know how membrane blebs and apoptotic bodies are formed.

Removal of Dead Cells

The formation of apoptotic bodies breaks cells up into "bite-sized" fragments that are edible for phagocytes. Apoptotic cells and their fragments also undergo several changes in their membranes that actively promote their phagocytosis so they are most often cleared before they lose membrane integrity and release their cellular contents. In healthy cells, phosphatidylserine is present on the inner leaflet of the plasma membrane, but in apoptotic cells this phospholipid "flips" out and is expressed on the outer layer of the membrane, where it is recognized by several macrophage receptors. Cells that are dying by apoptosis also secrete soluble factors that recruit phagocytes, and macrophages themselves may produce proteins that bind to apoptotic cells (but not live cells), leading to their engulfment. Apoptotic bodies may also become coated with natural antibodies and proteins of the complement system, notably C1q, which are recognized by phagocytes. Thus, numerous ligands induced on apoptotic cells serve as "eat me" signals and are recognized by receptors on phagocytes that bind and engulf these cells. This process of apoptotic cell phagocytosis is called *efferocytosis*; it is so efficient that dead cells disappear, often within minutes, without leaving a trace. In addition, production of pro-inflammatory cytokines is reduced in macrophages that have ingested apoptotic cells. Together with rapid clearance, this limits inflammatory reactions, even in the face of extensive apoptosis.

KEY CONCEPTS

APOPTOSIS

- Regulated mechanism of cell death that serves to eliminate unwanted and irreparably damaged cells, with the least possible host reaction
- Characterized by enzymatic degradation of proteins and DNA, initiated by caspases, and by recognition and removal of dead cells by phagocytes
- Initiated by two major pathways:
  - Mitochondrial (intrinsic) pathway is triggered by loss of survival signals, DNA damage, and accumulation of misfolded proteins (ER stress), which leads to leakage of pro-apoptotic proteins from mitochondrial membrane into the cytoplasm and subsequent caspase activation; can be inhibited by anti-apoptotic members of the BCL2 family, which are induced by survival signals including growth factors
  - Death receptor (extrinsic) pathway eliminates self-reactive lymphocytes and is a mechanism of cell killing by cytotoxic T lymphocytes; is initiated by engagement of death receptors (members of the TNF receptor family). The responsible ligands can be soluble or expressed on the surface of adjacent cells

Other Mechanisms of Cell Death

Although necrosis and apoptosis are the best-defined mechanisms of cell death, several other ways by which cells die have been described. Their importance in human diseases remains a topic of investigation, but students should be aware of their names and unique features.

- *Necroptosis.* As the name indicates, this form of cell death is a hybrid that shares aspects of both necrosis and apoptosis. Morphologically, and to some extent biochemically, it resembles necrosis, as both are characterized by loss of ATP, swelling of the cell and organelles, generation of reactive oxygen species (ROS), release of lysosomal enzymes, and ultimately rupture of the plasma membrane. Mechanistically, it is triggered by signal transduction pathways that culminate in cell death, a feature similar to apoptosis. Because of these overlapping features, necroptosis is sometimes called *programmed necrosis* to distinguish it from forms of necrosis driven passively by toxic or ischemic injury to the cell. In sharp contrast to apoptosis, the signals leading to necroptosis do not result in caspase activation, and hence it is also sometimes referred to as "caspase-independent" programmed cell death. The process of necroptosis starts in a manner similar to that of the extrinsic form of apoptosis, that is, by ligation of a receptor by its ligand. Ligation of TNFR1 is the most widely studied model of necroptosis, but many other signals, including ligation of Fas and yet to be identified sensors of viral DNA and RNA, can also trigger necroptosis. Since TNF can cause both apoptosis and necroptosis, the mechanisms underlying these effects of TNF are especially illustrative (Fig. 2.16).

  Although the entire set of signaling molecules and their interactions are not known, necroptosis involves two kinases called *receptor-interacting protein kinase 1* and *3* (*RIPK1* and *RIPK3*). As indicated in Fig. 2.16, ligation of TNFR1 recruits these kinases into a multiprotein complex, and RIPK3 phosphorylates a cytoplasmic protein called MLKL. In response to its phosphorylation, MLKL monomers assemble into oligomers, translocate from the cytosol to the plasma membrane, and cause the plasma membrane disruption that is characteristic of necrosis. This explains the morphologic similarity of necroptosis with necrosis initiated by other injuries.

  Necroptosis is postulated to be an important death pathway both in physiologic and pathologic conditions. For example, physiologic necroptosis occurs during the formation of the mammalian bone growth plate. In pathologic states, it is associated with cell death in steatohepatitis, acute pancreatitis, ischemia-reperfusion injury, and neurodegenerative diseases such as Parkinson disease. Necroptosis also acts as a backup mechanism in host defense against certain viruses that encode caspase inhibitors (e.g., cytomegalovirus).
- *Pyroptosis* is a form of apoptosis that is accompanied by the release of the fever-inducing cytokine IL-1 (*pyro* refers to fever). Microbial products that enter infected cells are recognized by cytoplasmic innate immune receptors and can activate the multiprotein complex called the *inflammasome* (Chapter 6). The function of the inflammasome is to activate caspase-1 (also known as interleukin-1β–converting enzyme), which cleaves a precursor form of interleukin-1 (IL-1) and releases its biologically active form. IL-1 is a mediator of many aspects of inflammation, including leukocyte recruitment and fever (Chapter 3). Caspase-1 and the closely related caspases-4 and -5 also induce death of the cells. Unlike classical apoptosis, this pathway of cell death is characterized by release of inflammatory mediators. Pyroptosis is thought to be the mechanism by which some microbes cause the death of infected cells and at the same time trigger local inflammation.
- *Ferroptosis.* Only discovered in 2012, ferroptosis is a distinct form of cell death that is triggered when excessive intracellular levels of iron or reactive oxygen species overwhelm the glutathione-dependent antioxidant defenses (discussed later) to cause unchecked membrane lipid peroxidation. The widespread peroxidation of lipids disrupts many aspects of membrane function, including fluidity, lipid-protein interactions, ion and nutrient transport, and signaling pathways. The overall effect is the loss of plasma membrane permeability, which ultimately leads to cell death resembling necrosis. The process is, however, regulated by specific signals (unlike necrosis) and can be prevented by reducing iron levels (hence its name). Ultrastructurally, the most prominent features are the loss

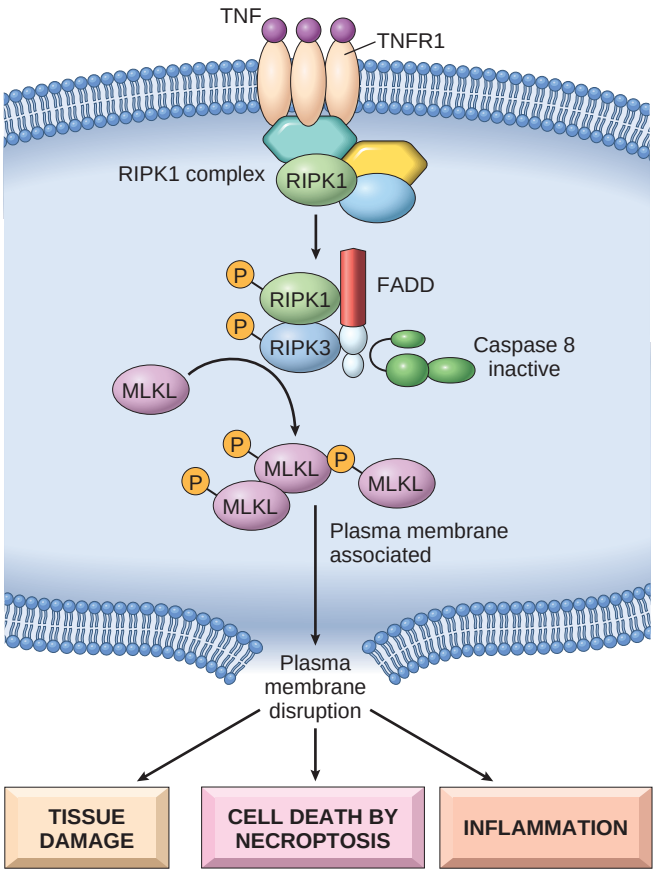


Figure 2.16 Molecular mechanism of TNF-mediated necroptosis. Cross-linking of TNFR1 by TNF initiates the illustrated series of downstream events, which ultimately lead to plasma membrane disruption, cell death, and inflammation. See text for details. (Modified from Galluzi L, et al: Programmed necrosis from molecules to health and disease, *Int Rev Cell Molec Biol* 289:1, 2011.)

of mitochondrial cristae and ruptured outer mitochondrial membrane. While its role in normal development and physiology remain controversial, ferroptosis has been linked to cell death in a variety of human pathologies, including cancer, neurodegenerative diseases, and stroke.

KEY CONCEPTS

NECROPTOSIS AND PYROPTOSIS

- Necroptosis resembles necrosis morphologically, but like apoptosis is a genetically controlled form of cell death.
- Necroptosis is triggered by ligation of TNFR1 and by proteins found in RNA and DNA viruses.
- Necroptosis is caspase-independent and depends on the RIPK1 and RIPK3 complex. RIPK1–RIPK3 signaling leads to the phosphorylation of MLKL, which then forms pores in the plasma membrane.
- Release of cellular contents evokes an inflammatory reaction as in necrosis.
- Pyroptosis occurs in cells infected by microbes. It involves activation of caspase-1, which cleaves the precursor form of IL-1 to generate biologically active IL-1. Caspase-1 along with other closely related caspases also cause death of the infected cell.
- Ferroptosis is an iron-dependent pathway of cell death induced by lipid peroxidation.

Autophagy

Autophagy is a process in which a cell eats its own contents (*Greek: auto,* self; *phagy,* eating). It involves the delivery of cytoplasmic materials to the lysosome for degradation. Autophagy is an evolutionarily conserved survival mechanism whereby, in states of nutrient deprivation, starved cells live by cannibalizing themselves and recycling the digested contents. Autophagy is implicated in many physiologic states (e.g., aging and exercise) and pathologic processes. It proceeds through several steps (Fig. 2.17):

- Nucleation and formation of an isolation membrane, also called a *phagophore;* the isolation membrane is believed to be derived from the ER, though other membrane sources such as the plasma membrane and mitochondria may contribute
- Formation of a vesicle, called the *autophagosome,* from the isolation membrane, inside which intracellular organelles and cytosolic structures are sequestered
- Maturation of the autophagosome by fusion with lysosomes, to deliver digestive enzymes that degrade the contents of the autophagosome

In recent years, more than a dozen "autophagy-related genes" called *Atgs* have been identified whose products are required for the creation of the autophagosome. Environmental cues like nutrient deprivation or depletion of growth factors activate an initiation complex of four proteins that promotes the hierarchical recruitment of *Atgs* to nucleate the initiation membrane. The initiation membrane elongates further, surrounds and captures its cytosolic cargo, and closes to form the autophagosome. The elongation and closure of the initiation membrane require the coordinated action of two ubiquitin-like conjugation systems that result in the covalent linkage of the lipid phosphatidylethanolamine (PE) to microtubule-associated protein light chain 3 (LC3). PE-lipidated LC3 is increased during autophagy, and it is therefore a useful marker for identifying cells in which autophagy is occurring. The newly formed autophagosome fuses with lysosomes to form an autophagolysosome. In the terminal step, the inner membrane and enclosed cytosolic cargoes are degraded by lysosomal enzymes. There is increasing evidence that autophagy is not a random process that engulfs cytosolic contents indiscriminately. Rather, the loading of cargo into the autophagosome is selective, and one of the functions of the lipidated LC3 is to target protein aggregates and effete organelles.

Autophagy functions as a survival mechanism under various stress conditions, maintaining the integrity of cells by recycling essential metabolites and clearing intracellular debris. It is therefore prominent in atrophic cells exposed to severe nutrient deprivation. Autophagy is also involved in the turnover of organelles like the ER, mitochondria, and lysosomes and the clearance of intracellular aggregates that accumulate during aging, stress, and various disease states. Autophagy can trigger cell death if it is inadequate to cope

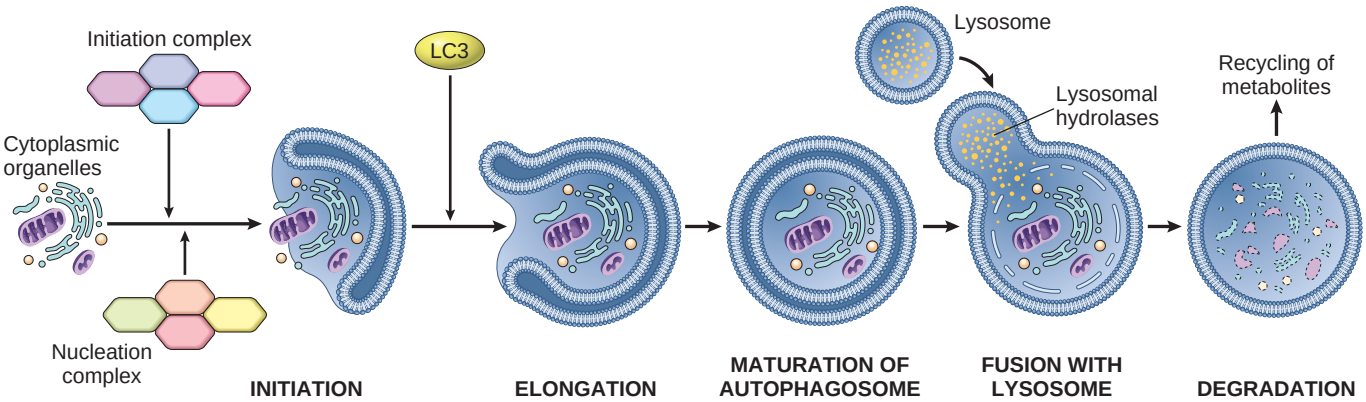


Figure 2.17 Autophagy. Cellular stresses, such as nutrient deprivation, activate an autophagy pathway that proceeds through several phases (initiation, nucleation, and elongation of isolation membrane) and eventually creates double-membrane-bound vacuoles (autophagosome) in which cytoplasmic materials, including organelles, are sequestered and then degraded after fusion of the vesicles with lysosomes. In the final stage, the digested materials are released for recycling of metabolites. See text for details. *LC3,* Light chain 3. (Modified from Choi, AMK, Ryter S, Levine B: Autophagy in human health and disease, *N Engl J Med* 368:651, 2013.)

with the stressor. This pathway of cell death is distinct from necrosis and apoptosis, but the mechanism is unknown. Furthermore, it is not clear whether cell death is caused by autophagy or by the stress that triggered autophagy. Nevertheless, autophagic vacuolization often precedes or accompanies cell death.

There is accumulating evidence that autophagy plays a role in human diseases, including the following:

- *Cancer:* Autophagy can both promote cancer growth and act as a defense against cancers. This is an area of active investigation, as discussed in Chapter 7.
- *Neurodegenerative disorders:* Many neurodegenerative disorders are associated with dysregulation of autophagy. Alzheimer disease is characterized by impaired autophagosome maturation, and in mouse models of the disease genetic defects in autophagy accelerate neurodegeneration. In Huntington disease, mutant huntingtin impairs autophagy.
- *Infectious diseases:* Many pathogens are degraded by autophagy; these include mycobacteria, *Shigella* spp., and HSV-1. This is one way by which microbial proteins are digested and delivered to antigen presentation pathways. Macrophage-specific deletion of Atg5 increases susceptibility to tuberculosis.
- *Inflammatory bowel diseases:* Genome-wide association studies have linked both Crohn disease and ulcerative colitis to single-nucleotide polymorphisms (SNPs) in the autophagy-related gene *ATG16L1*. How these polymorphisms promote intestinal inflammation is not known.

KEY CONCEPTS

AUTOPHAGY

- Autophagy involves sequestration of cellular organelles into cytoplasmic autophagic vacuoles (autophagosomes) that fuse with lysosomes and digest enclosed material.
- Autophagy is an adaptive response that is enhanced during nutrient deprivation, allowing the cell to cannibalize itself to survive.
- Autophagosome formation is regulated by more than a dozen proteins that act in a coordinated and sequential manner.
- Dysregulation of autophagy occurs in many disease states, including cancer, inflammatory bowel diseases, and neurodegenerative disorders. Autophagy plays a role in host defense against certain microbes.

MECHANISMS OF CELL INJURY

The discussion of the pathways of cell injury and death sets the stage for a consideration of the underlying biochemical mechanisms of cell injury. The molecular alterations that lead to cell injury are complex, but several principles are relevant to most forms of cell injury:

- **The cellular response to injurious stimuli depends on the nature of the injury, its duration, and its severity.** Small doses of a chemical toxin or brief periods of ischemia may induce reversible injury, whereas large doses of the same toxin or more prolonged ischemia might result either in rapid cell death or in slowly progressive injury that with time becomes irreversible and leads to cell death.
- **The consequences of cell injury depend on the type, state, and adaptability of the injured cell.** The cell's nutritional and hormonal status, metabolic demands, and functions dictate the response to injury. How vulnerable is a cell, for example, to loss of blood supply and hypoxia? When a skeletal muscle cell in the leg is deprived of its blood supply, it can be rested and preserved; cardiac muscle cells have no such option. Exposure of two individuals to identical concentrations of a toxin, such as carbon tetrachloride, may produce no effect in one and cell death in the other. This may be due to polymorphisms in genes encoding hepatic enzymes that metabolize carbon tetrachloride (CCl_4) to toxic by-products (Chapter 9). With the complete mapping of the human genome, there is great interest in identifying genetic polymorphisms that affect the responses of different individuals to various injurious agents.
- **Any injurious stimulus may simultaneously trigger multiple interconnected mechanisms that damage cells.** This is one reason why it can be difficult to ascribe cell injury in a particular situation to a single or even dominant biochemical derangement.

We start this section with a discussion of general mechanisms that are involved in reversible injury and necrosis caused by diverse stimuli and conclude with a discussion of the pathways of injury in selected clinical situations that illustrate general principles.

General Mechanisms of Cell Injury and Intracellular Targets of Injurious Stimuli

Cell injury results from abnormalities in one or more essential cellular components (Fig. 2.18). The principal targets of injurious stimuli are mitochondria, cell membranes, the machinery of protein synthesis and secretion, and DNA. The consequences of injury of each of these cellular components are distinct but overlapping.

Mitochondrial Damage

Mitochondria are critical players in all pathways leading to cell injury and death. This should be expected because mitochondria supply life-sustaining energy by producing ATP but are also targets of many injurious stimuli. Thus, in many ways, they are the arbiters of life and death of cells. Mitochondria can be damaged by increases of cytosolic Ca^{2+}, ROS (discussed later), and oxygen deprivation, which makes them sensitive to virtually all types of injurious stimuli, including hypoxia and toxins. In addition, mutations in mitochondrial genes are the cause of some inherited diseases (Chapter 5).

There are three major consequences of mitochondrial damage.

- *ATP depletion.* Decreased ATP synthesis and ATP depletion are frequently associated with both hypoxic and chemical (toxic) injury (Fig. 2.19). ATP is produced in two ways. The major pathway in mammalian cells, particularly those that are nondividing (e.g., brain and liver), is oxidative phosphorylation of adenosine diphosphate in a reaction that results in reduction of oxygen by the mitochondrial electron transport system. The second is the glycolytic pathway, which can generate ATP, albeit in far smaller

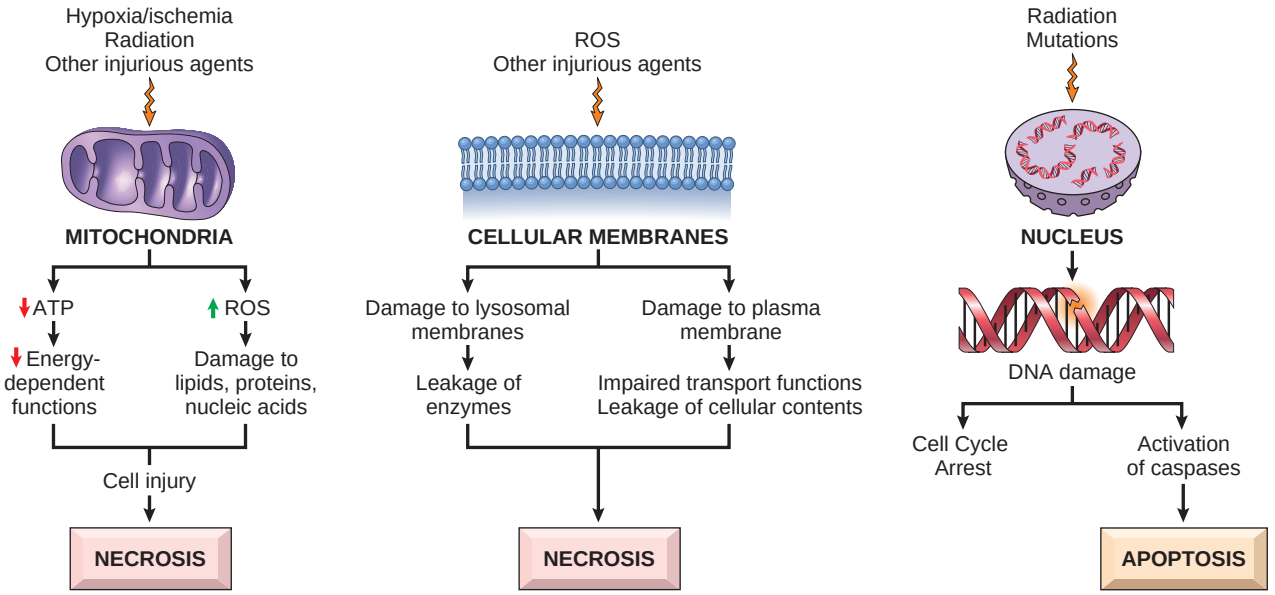


Figure 2.18 The principal forms and sites of damage in cell injury. *ATP*, Adenosine triphosphate; *ROS*, reactive oxygen species.

amounts, in the absence of oxygen using glucose derived either from body fluids or from the hydrolysis of glycogen. In addition to mitochondrial damage, *the major causes of ATP depletion are reduced supply of oxygen and nutrients (because of ischemia and hypoxia), and the actions of some toxins (e.g., cyanide)*. Mitochondrial damage often results in the formation of a high-conductance channel in the mitochondrial membrane, called the *mitochondrial permeability transition pore* (see Fig. 2.19). The opening of this conductance channel leads to the loss of mitochondrial membrane potential, resulting in failure of oxidative phosphorylation and progressive ATP depletion that culminates in necrosis.

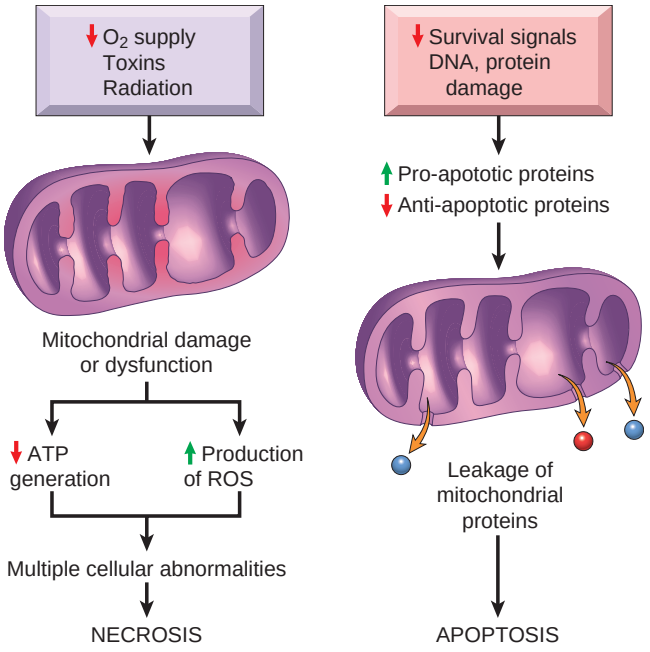


Figure 2.19 Role of mitochondria in cell injury and death. Mitochondria are affected by a variety of injurious stimuli, and their abnormalities lead to necrosis or apoptosis. *ATP*, Adenosine triphosphate; *ROS*, reactive oxygen species.

High-energy phosphate in the form of ATP is required for virtually all synthetic and degradative processes within the cell. These include membrane transport, protein synthesis, lipogenesis, and the deacylation-reacylation reactions necessary for phospholipid turnover. *Hence, depletion of ATP to 5% to 10% of normal levels has widespread effects on many critical cellular systems*.

- The activity of the *plasma membrane energy-dependent sodium pump* (Na^+,K^+-ATPase) is reduced (Chapter 1). Failure of this active transport system causes sodium to enter and accumulate inside cells and potassium concentrations to fall. The net solute gain results in osmotically driven water accumulation that leads to *cell swelling* and ER dilation.
- *Cellular energy metabolism is altered*. If the supply of oxygen to cells is reduced, as in ischemia, oxidative phosphorylation ceases, resulting in a decrease in cellular ATP and associated increase in adenosine monophosphate. These changes stimulate phosphorylase and phosphofructokinase activities, leading to increased rates of glycogenolysis and glycolysis, respectively, in an effort to maintain energy supplies by generating ATP through metabolism of glucose derived from glycogen. As a consequence, *glycogen stores are rapidly depleted*. Glycolysis under anaerobic conditions results in the accumulation of *lactic acid* and inorganic phosphates from the hydrolysis of phosphate esters. This reduces the intracellular pH, resulting in decreased activity of many cytosolic enzymes.
- With prolonged or worsening depletion of ATP, structural disruption of the protein synthetic apparatus occurs, manifest as detachment of ribosomes from the rough ER and dissociation of polysomes, with a consequent *reduction in protein synthesis*. There may

also be increased protein misfolding, with injurious effects that are discussed later.
    - Ultimately, there is irreversible damage to mitochondrial and lysosomal membranes, and the cell undergoes *necrosis*.
- Incomplete oxidative phosphorylation also leads to the formation of ROS, which have many deleterious effects, described later.
- As discussed earlier, leakage of mitochondrial proteins due to channel formation by pro-apoptotic BAX and BAK is the initial step in apoptosis by the intrinsic pathway. This action of BAX and BAK is specific to mitochondrial membranes only and leads to damage of other organelles indirectly.

Membrane Damage

Early loss of selective membrane permeability, leading ultimately to overt membrane damage, is a consistent feature of most forms of cell injury (except apoptosis). Membrane damage may affect the integrity and functions of all cellular membranes. In ischemic cells, membrane defects may be the result of ATP depletion and calcium-mediated activation of phospholipases. The plasma membrane can also be damaged directly by bacterial toxins, viral proteins, lytic complement components, and a variety of physical and chemical agents. Several biochemical mechanisms may contribute to membrane damage (Fig. 2.20):

- *ROS*. Oxygen free radicals cause injury to cell membranes by lipid peroxidation, discussed later.
- *Decreased phospholipid synthesis*. The production of phospholipids in cells may be reduced as a consequence of defective mitochondrial function or hypoxia, both of which decrease ATP production and thus affect energy-dependent biosynthetic pathways. Decreased phospholipid synthesis may affect all cellular membranes, including those of mitochondria.

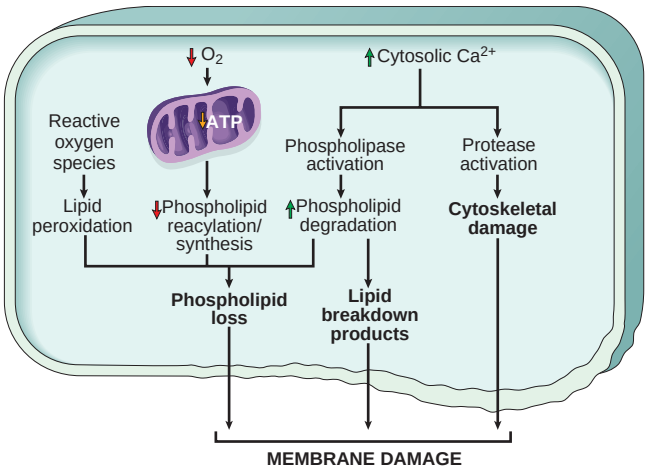


Figure 2.20 Mechanisms of membrane damage in cell injury. Decreased O_2 and increased cytosolic Ca^{2+} are typically seen in ischemia, but may accompany other forms of cell injury. Reactive oxygen species, which are often produced on reperfusion of ischemic tissues, also cause membrane damage *(not shown)*.

- *Increased phospholipid breakdown*. Severe cell injury is associated with increased degradation of membrane phospholipids, probably due to activation of calcium-dependent phospholipases by increased cytosolic and mitochondrial $Ca2^{+}$. Phospholipid breakdown leads to the accumulation of *lipid breakdown products*, including unesterified free fatty acids, acyl carnitine, and lysophospholipids, which have a detergent effect on membranes. They may also either insert into the lipid bilayer of the membrane or exchange with membrane phospholipids, potentially causing changes in permeability and electrophysiologic alterations.
- *Cytoskeletal abnormalities*. Cytoskeletal filaments serve as anchors connecting the plasma membrane to the cell interior, and proteases activated by cytosolic Ca^{2+} may damage these tethers. When exacerbated by cell swelling, particularly in myocardial cells, this damage leads to detachment of the cell membrane from the cytoskeleton, rendering it susceptible to stretching and rupture.

Damage to different cellular membranes has diverse effects on cells.

- *Mitochondrial membrane damage*. As discussed earlier, damage to mitochondrial membranes results in opening of the mitochondrial permeability transition pore, leading to decreased ATP generation and release of proteins that trigger apoptotic death.
- *Plasma membrane damage*. Plasma membrane damage results in loss of osmotic balance and influx of fluids and ions, as well as loss of cellular contents. The cells may also leak metabolites (e.g., glycolytic intermediates) that are vital for the reconstitution of ATP, thus further depleting energy stores.
- *Injury to lysosomal membranes* results in leakage of their enzymes into the cytoplasm and activation of acid hydrolases in the acidic intracellular pH of the injured cell. These lysosomal hydrolases include RNases, DNases, proteases, phosphatases, and glucosidases, which degrade RNA, DNA, proteins, phosphoproteins, and glycogen, respectively, and push cells into necrosis.

Damage to DNA

Damage to nuclear DNA activates sensors that trigger p53-dependent pathways (Chapter 7). DNA damage may be caused by exposure to radiation, chemotherapeutic (anticancer) drugs, and ROS, or may occur spontaneously as a part of aging, due largely to deamination of cytosine residues to uracil residues. DNA damage activates p53, which arrests cells in the G1 phase of the cell cycle and activates DNA repair mechanisms. If these mechanisms fail to correct the DNA damage, p53 triggers apoptosis by the mitochondrial pathway. Thus, the cell chooses to die rather than survive with abnormal DNA that has the potential to induce malignant transformation. Predictably, mutations in p53 that interfere with its ability to arrest cell cycling or to induce apoptosis are associated with numerous cancers (Chapter 7).

In addition to damage to various organelles, some biochemical alterations are involved in many situations that lead to cell injury. Two of these general pathways are discussed next.

Oxidative Stress: Accumulation of Oxygen-Derived Free Radicals

Cell injury induced by free radicals, particularly ROS, is an important mechanism of cell damage in many pathologic conditions, such as chemical and radiation injury, ischemia-reperfusion injury (induced by restoration of blood flow in ischemic tissue), cellular aging, and microbial killing by phagocytes. *Free radicals* are chemical species that have a single unpaired electron in an outer orbit. Unpaired electrons are highly reactive and "attack" and modify adjacent molecules, such as inorganic or organic chemicals—proteins, lipids, carbohydrates, nucleic acids—many of which are key components of cell membranes and nuclei. Some of these reactions are autocatalytic, whereby molecules that react with free radicals are themselves converted into free radicals, thus propagating the chain of damage.

ROS are a type of oxygen-derived free radical whose role in cell injury is well established. ROS are produced normally in cells during mitochondrial respiration and energy generation, but they are degraded and removed by intracellular ROS scavengers. These defense systems allow cells to maintain a steady state in which free radicals may be present at low concentrations but do not cause damage. Increased production or decreased scavenging of ROS may lead to an excess of free radicals, a condition called *oxidative stress*. Oxidative stress has been implicated in a wide variety of pathologic processes, including cell injury, cancer, aging, and some degenerative diseases, such as Alzheimer disease. ROS are also produced in large amounts by activated leukocytes, particularly neutrophils and macrophages, during inflammatory reactions aimed at destroying microbes and cleaning up dead cells and other unwanted substances (Chapter 3).

The following section discusses the generation and removal of ROS, and how they contribute to cell injury. The properties of some of the most important free radicals are summarized in Table 2.2.

Generation of Free Radicals

Free radicals may be generated within cells in several ways (Fig. 2.21):

- *The reduction-oxidation reactions that occur during normal metabolic processes.* As a part of normal respiration, molecular O_2 is reduced by the transfer of four electrons to H_2 to generate two water molecules. This conversion is catalyzed by oxidative enzymes in the ER, cytosol, mitochondria, peroxisomes, and lysosomes. During this process, small amounts of partially reduced intermediates are produced in which different numbers of electrons have been transferred from O_2; these include superoxide anion ($O_2^{\cdot-}$, one electron), hydrogen peroxide (H_2O_2, two electrons), and hydroxyl radicals ($^{\cdot}OH$, three electrons).
- *Absorption of radiant energy* (e.g., ultraviolet light, x-rays). For example, ionizing radiation can hydrolyze water into $^{\cdot}OH$ and hydrogen (H) free radicals.
- Rapid bursts of ROS are produced in activated leukocytes during *inflammation*. This occurs in a precisely controlled reaction carried out by a plasma membrane multiprotein complex that uses NADPH oxidase for the redox reaction (Chapter 3). In addition, some intracellular oxidases (e.g., xanthine oxidase) generate $O_2^{\cdot-}$. Defects in leukocytic superoxide production lead to chronic granulomatous disease (Chapter 6).
- *Enzymatic metabolism of exogenous chemicals or drugs* can generate free radicals that are not ROS but have similar effects (e.g., CCl_4 can generate $^{\cdot}CCl_3$, described later in the chapter).
- *Transition metals* such as iron and copper donate or accept free electrons during intracellular reactions and catalyze free radical formation, as in the Fenton reaction ($H_2O_2 + Fe^{2+} \rightarrow Fe^{3+} + {}^{\cdot}OH + OH^-$). Because most of the intracellular free iron is in the ferric (Fe^{3+}) state, it must be reduced to the ferrous (Fe^{2+}) form to participate in the Fenton reaction. This reduction can be enhanced by $O_2^{\cdot-}$, and

Table 2.2 Properties of the Principal Free Radicals Involved in Cell Injury

Properties	$O_2^{\cdot-}$	H_2O_2	$^{\cdot}OH$	$ONOO^-$
Mechanisms of production	Incomplete reduction of O_2 during oxidative phosphorylation; by phagocyte oxidase in leukocytes	Generated by SOD from $O_2^{\cdot-}$ and by oxidases in peroxisomes	Generated from H_2O by hydrolysis (e.g., by radiation); from H_2O_2 by Fenton reaction; from $O_2^{\cdot-}$	Produced by interaction of $O_2^{\cdot-}$ and NO generated by NO synthase in many cell types (endothelial cells, leukocytes, neurons, others)
Mechanisms of inactivation	Conversion to H_2O_2 and O_2 by SOD	Conversion to H_2O and O_2 by catalase (peroxisomes), glutathione peroxidase (cytosol, mitochondria)	Conversion to H_2O by glutathione peroxidase	Conversion to HNO_2 by peroxiredoxins (cytosol, mitochondria)
Pathologic effects	Stimulates production of degradative enzymes in leukocytes and other cells; may directly damage lipids, proteins, DNA; acts close to site of production	Can be converted to OH and OCl^-, which destroy microbes and cells; can act distant from site of production	Most reactive oxygen-derived free radical; principal ROS responsible for damaging lipids, proteins, and DNA	Damages lipids, proteins, DNA

HNO_2, Nitrite; *H_2O_2,* hydrogen peroxide; *NO,* nitric oxide; *$O_2^{\cdot-}$,* superoxide anion; *OCl^-,* hypochlorite; *OH,* hydroxyl radical; *$ONOO^-$,* peroxynitrite; *ROS,* reactive oxygen species; *SOD,* superoxide dismutase.

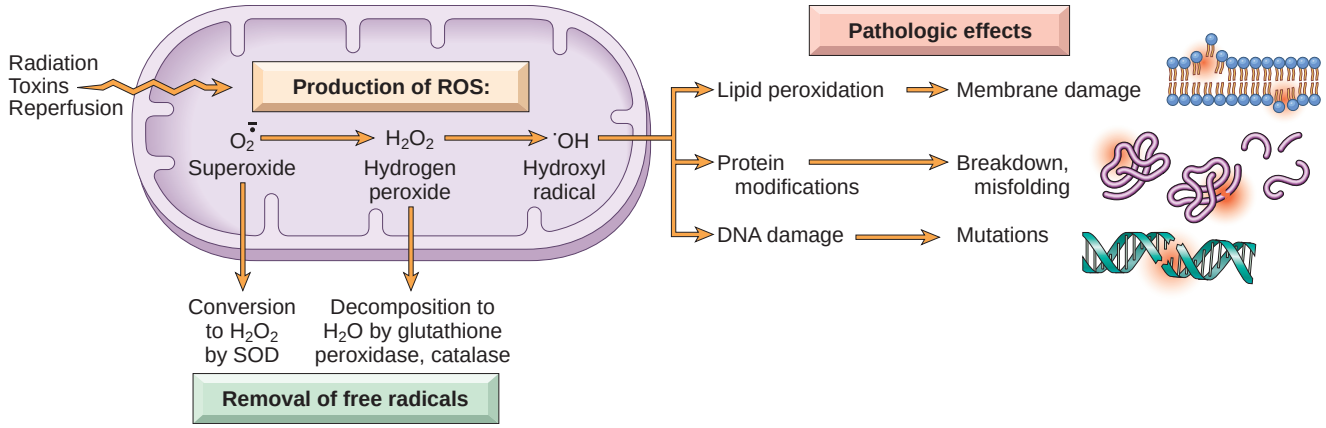


Figure 2.21 The generation, removal, and role of reactive oxygen species *(ROS)* in cell injury. The production of ROS is increased by many injurious stimuli. These free radicals are removed by spontaneous decay and by specialized enzymatic systems. Excessive production or inadequate removal leads to accumulation of free radicals in cells, which may damage lipids (by peroxidation), proteins, and deoxyribonucleic acid *(DNA)*, resulting in cell injury.

thus sources of iron and $O_2^{\cdot-}$ may cooperate in oxidative cell damage.

- *Nitric oxide (NO),* an important chemical mediator generated by endothelial cells, macrophages, neurons, and other cell types (Chapter 3), can act as a free radical and may also be converted to highly reactive peroxynitrite anion ($ONOO^-$) as well as NO_2 and NO_3^-.

Removal of Free Radicals

Free radicals are inherently unstable and generally decay spontaneously. $O_2^{\cdot-}$, for example, is unstable and decays (dismutates) spontaneously to O_2 and H_2O_2 in the presence of water. In addition, cells have developed multiple nonenzymatic and enzymatic mechanisms to remove free radicals and thereby minimize injury (see Fig. 2.21). These include the following:

- *Antioxidants* either block free radical formation or inactivate (e.g., scavenge) free radicals. Examples are the lipid-soluble vitamins E and A as well as ascorbic acid and glutathione in the cytosol.
- As we have seen, free *iron* and *copper* can catalyze the formation of ROS. Under normal circumstances, the reactivity of these metals is minimized by their binding to storage and transport proteins (e.g., transferrin, ferritin, lactoferrin, and ceruloplasmin), which prevents these metals from participating in reactions that generate ROS.
- Several *enzymes* act as free radical–scavenging systems and break down H_2O_2 and $O_2^{\cdot-}$. These enzymes are located near the sites of generation of the oxidants and include the following:
  1. *Catalase,* present in peroxisomes, decomposes H_2O_2 ($2H_2O_2 \rightarrow O_2 + 2H_2O$).
  2. *Superoxidase dismutases* (SODs) are found in many cell types and convert $O_2^{\cdot-}$ to H_2O_2 ($2O_2^{\cdot-} + 2H \rightarrow H_2O_2 + O_2$). This group of enzymes includes both manganese-SOD, which is localized in mitochondria, and copper-zinc-SOD, which is found in the cytoplasm.
  3. *Glutathione peroxidase* also protects against injury by catalyzing free radical breakdown ($H_2O_2 + 2GSH \rightarrow GSSG$ [glutathione homodimer] $+ 2H_2O$, or $2^{\cdot}OH + 2GSH \rightarrow GSSG + 2H_2O$). The intracellular ratio of oxidized glutathione (GSSG) to reduced glutathione (GSH) reflects the oxidative state of the cell and is an important indicator of the cell's ability to detoxify ROS.

Pathologic Effects of Free Radicals

The effects of ROS and other free radicals are wide-ranging, but three reactions are particularly relevant to cell injury (see Fig. 2.21):

- *Lipid peroxidation in membranes.* In the presence of O_2, free radicals may cause peroxidation of lipids within plasma and organellar membranes. Oxidative damage is initiated when the double bonds in unsaturated fatty acids of membrane lipids are attacked by O_2-derived free radicals, particularly by OH. The lipid–free radical interactions yield peroxides, which are themselves unstable and reactive, and an autocatalytic chain reaction ensues (called *propagation*) that can result in extensive membrane damage.
- *Oxidative modification of proteins.* Free radicals promote oxidation of amino acid side chains, formation of covalent protein-protein cross-links (e.g., disulfide bonds), and oxidation of the protein backbone. Oxidative modifications may also damage the active sites of enzymes, disrupt the conformation of structural proteins, and enhance proteasomal degradation of unfolded or misfolded proteins, thereby raising havoc throughout the cell.
- *Lesions in DNA.* Free radicals are capable of causing single- and double-strand breaks in DNA, cross-linking DNA strands, and forming adducts. Oxidative DNA damage has been implicated in cell aging (discussed later in the chapter) and in malignant transformation of cells (Chapter 7).

The traditional thinking about free radicals was that they cause cell injury and death by necrosis, and, in fact, the production of ROS is often a prelude to necrosis. However, it is now clear that free radicals can also trigger apoptosis. It is also possible that these potentially deadly molecules, when produced under controlled conditions in the "right" dose, serve important physiologic functions in signaling by cellular receptors and other pathways.

Disturbance in Calcium Homeostasis

Calcium ions normally serve as second messengers in several signaling pathways, but if released into the cytoplasm of cells in excessive amounts, are also an important source of cell injury. In keeping with this, calcium depletion protects cultured cells from injury induced by a variety of harmful stimuli. Cytosolic free Ca^{2+} is normally maintained at very low concentrations (~0.1 µmol) compared with extracellular levels of 1.3 mmol, and most intracellular Ca^{2+} is sequestered in mitochondria and the ER. Ischemia and certain toxins cause an excessive increase in cytosolic Ca^{2+}, initially because of release from intracellular stores, and later due to increased influx across the plasma membrane (Fig. 2.22). Excessive intracellular Ca^{2+} may cause cell injury by several mechanisms, although the significance of these mechanisms in cell injury in vivo is not established.

- The accumulation of Ca^{2+} in mitochondria results in opening of the mitochondrial permeability transition pore and, as described earlier, failure of ATP generation.
- Increased cytosolic Ca^{2+} activates a number of enzymes with potentially deleterious effects on cells. These enzymes include *phospholipases* (which cause membrane damage), *proteases* (which break down both membrane and cytoskeletal proteins), *endonucleases* (which are responsible for DNA and chromatin fragmentation), and *ATPases* (thereby hastening ATP depletion).

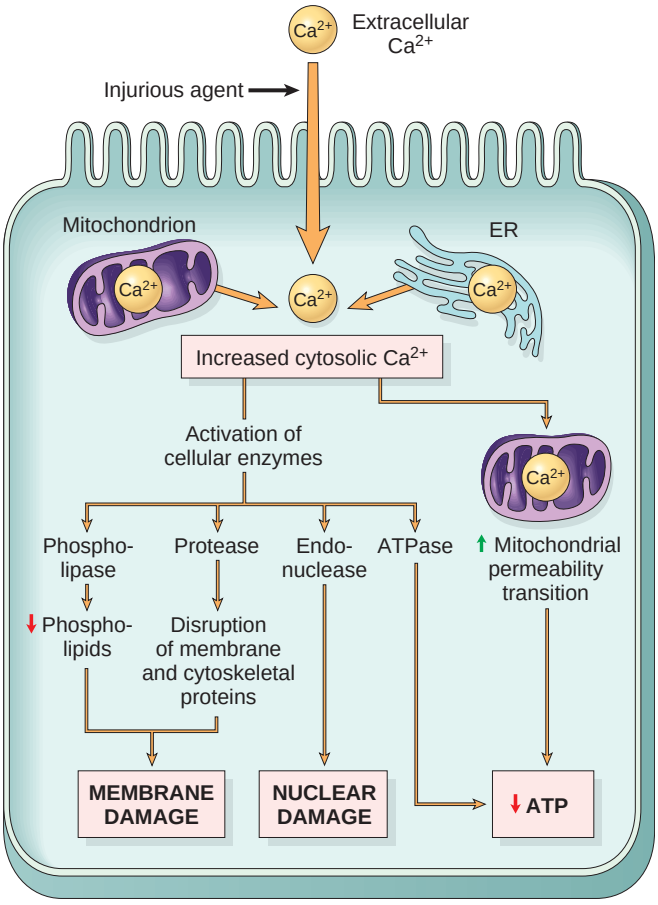


Figure 2.22 The role of increased cytosolic calcium in cell injury. *ER*, Endoplasmic reticulum.

Endoplasmic Reticulum Stress: the Unfolded Protein Response

The accumulation of misfolded proteins in the ER can stress adaptive mechanisms and trigger apoptosis. Chaperones in the ER control the proper folding of newly synthesized proteins, and misfolded polypeptides are shuttled into the cytoplasm where they are ubiquitinated and targeted for proteolysis in proteasomes (Chapter 1). If, however, unfolded or misfolded proteins accumulate in the ER, they trigger a number of alterations that are collectively called the *unfolded protein response*. The unfolded protein response activates signaling pathways that increase the production of chaperones, enhance proteasomal degradation of abnormal proteins, and slow protein translation, thus reducing the load of misfolded proteins in the cell (Fig. 2.23). However, if this cytoprotective response is unable to cope with the accumulation of misfolded proteins, the cell activates caspases and induces apoptosis. This process is called *ER stress.* Intracellular accumulation of misfolded proteins may be caused by an increased rate of misfolding or a reduction in the cell's ability to repair or eliminate them. Increased misfolding may be a consequence of deleterious mutations or decreased capacity to correct misfolded proteins, as occurs in aging. Protein misfolding may also be increased in viral infections when proteins encoded by the viral genome are synthesized in such large quantities that they overwhelm the quality control system that normally ensures proper protein folding. Increased demand for secretory proteins such as insulin in insulin-resistant states, and changes in intracellular pH and redox state are other stressors that result in misfolded protein accumulation. Protein misfolding is thought to be the causative cellular abnormality in several neurodegenerative diseases (Chapter 28). Given that many "foldases" require ATP to function, deprivation of glucose and oxygen, as in ischemia and hypoxia, also may increase the burden of misfolded proteins. Diseases caused by misfolded proteins are listed in Table 2.3.

KEY CONCEPTS

MECHANISMS OF CELL INJURY

- ATP depletion: failure of energy-dependent functions → reversible injury → necrosis
- Mitochondrial damage: ATP depletion → failure of energy-dependent cellular functions → ultimately, necrosis; under some conditions, leakage of mitochondrial proteins that cause apoptosis
- Increased permeability of cellular membranes: may affect plasma membrane, lysosomal membranes, mitochondrial membranes; typically culminates in necrosis
- Accumulation of damaged DNA and misfolded proteins: triggers apoptosis
- Accumulation of ROS: covalent modification of cellular proteins, lipids, nucleic acids
- Influx of calcium: activation of enzymes that damage cellular components and may also trigger apoptosis
- Unfolded protein response and ER stress: Accumulation of misfolded proteins in the ER activates adaptive mechanisms that help the cell to survive, but if their repair capacity is exceeded they trigger apoptosis.

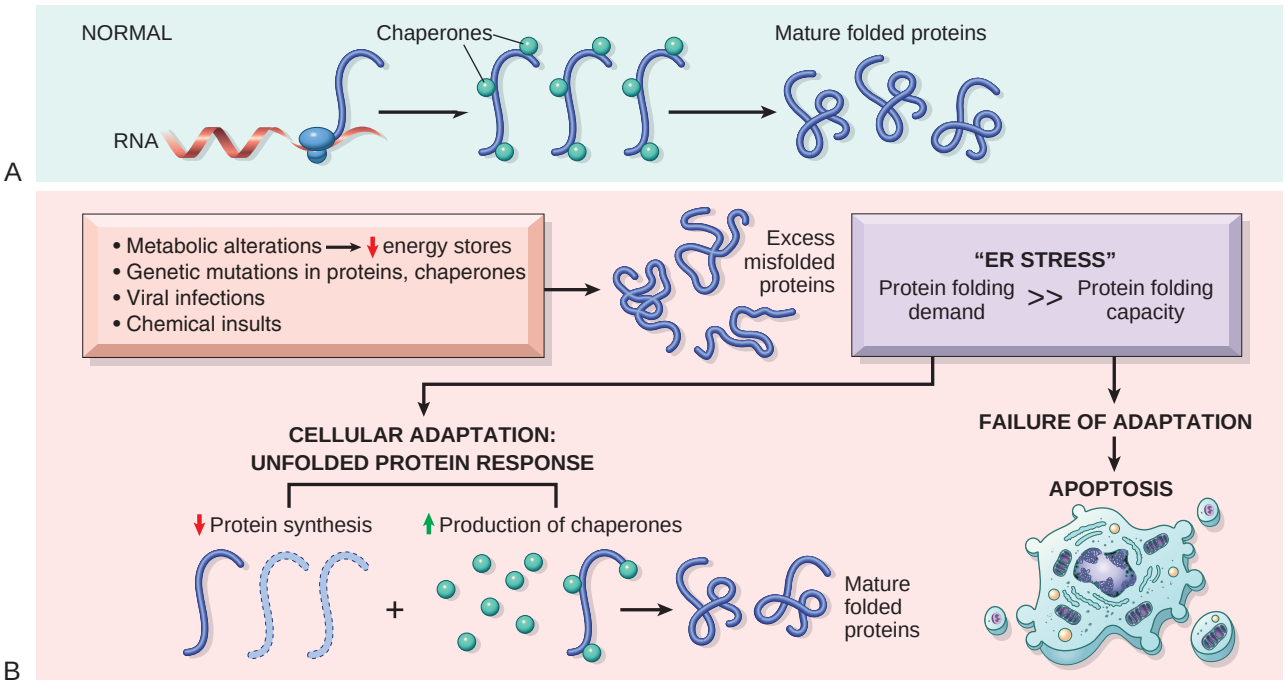


Figure 2.23 The unfolded protein response and endoplasmic reticulum *(ER)* stress. (A) In healthy cells, newly synthesized proteins are folded in the ER with the help of chaperones and are then incorporated into the cell or secreted. (B) Various external stresses or mutations induce a state called *ER stress*, in which the cell is unable to cope with the load of misfolded proteins. Accumulation of these proteins in the ER triggers the unfolded protein response, which tries to restore protein homeostasis; if this response is inadequate, the unfolded protein response actively signals apoptosis.

Table 2.3 Selected Examples of Diseases Caused by Misfolding of Proteins

Disease	Affected Protein	Pathogenesis
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator (CFTR)	Loss of CFTR leads to defects in chloride transport
Familial hypercholesterolemia	LDL receptor	Loss of LDL receptor leads to hypercholesterolemia
Tay-Sachs disease	Hexosaminidase β subunit	Lack of the lysosomal enzyme leads to storage of GM_2 gangliosides in neurons
α_1-antitrypsin deficiency	α_1-antitrypsin	Storage of nonfunctional protein in hepatocytes causes apoptosis; absence of enzymatic activity in lungs causes destruction of elastic tissue giving rise to emphysema
Creutzfeldt-Jacob disease	Prions	Abnormal folding of PrPsc causes neuronal cell death
Alzheimer disease	Aβ peptide	Abnormal folding of Aβ peptides causes aggregation within neurons and apoptosis

CLINICOPATHOLOGIC CORRELATIONS: SELECTED EXAMPLES OF CELL INJURY AND DEATH

Having briefly reviewed the causes, morphology, and general mechanisms of cell injury and death, we now describe some common and clinically significant forms of cell injury. These examples illustrate many of the mechanisms and sequence of events in cell injury described earlier.

Hypoxia and Ischemia

Ischemia, the most common cause of cell injury in clinical medicine, results from hypoxia induced by reduced blood flow, most often due to a mechanical arterial obstruction. It can also occur due to reduced venous drainage. In contrast to hypoxia, in which blood flow is maintained and during which energy production by anaerobic glycolysis can continue, ischemia compromises the delivery of substrates for glycolysis. Thus, in ischemic tissues, not only does aerobic metabolism cease but anaerobic energy generation also fails after glycolytic substrates are exhausted or glycolysis is inhibited by the accumulation of metabolites, which otherwise would be washed out by flowing blood. For this reason, *ischemia causes more rapid and severe cell and tissue injury than hypoxia*.

Mechanisms of Ischemic Cell Injury

The sequence of events following hypoxia or ischemia reflects many of the biochemical alterations in cell injury, summarized here in Fig. 2.24 and shown earlier in Figs. 2.5 and 2.6. As intracellular oxygen tension falls, oxidative phosphorylation fails and ATP generation decreases. The consequences of ATP depletion were described earlier in the Mitochondrial Damage section. In brief, loss of ATP results

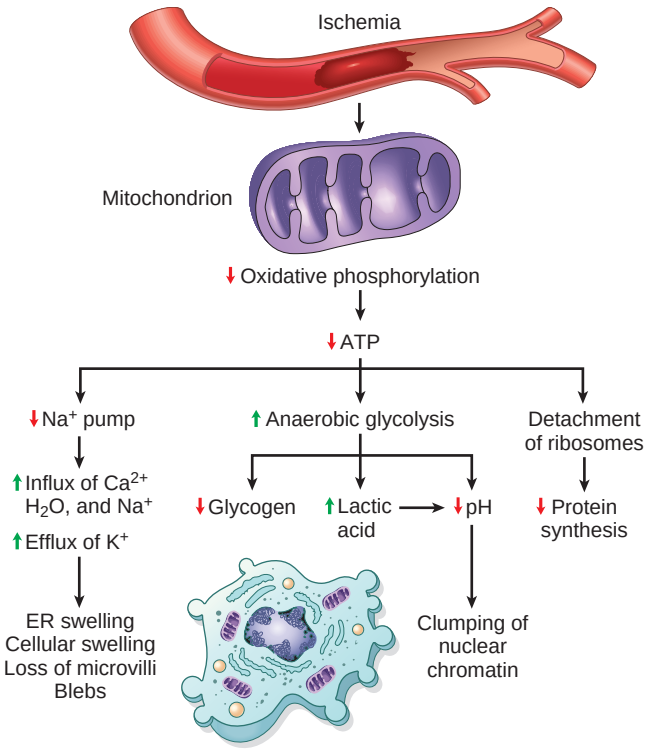


Figure 2.24 Functional and morphologic consequences of decreased intracellular adenosine triphosphate *(ATP)* in ischemic cell injury. The morphologic changes shown here are indicative of reversible cell injury. Further depletion of ATP results in cell death, typically by necrosis. *ER*, Endoplasmic reticulum.

initially in reversible cell injury (cell and organelle swelling) and later in cell death by necrosis.

Mammalian cells have developed protective responses to deal with hypoxic stress. The best defined of these is induction of a transcription factor called *hypoxia-inducible factor-1* (HIF-1), which promotes new blood vessel formation, stimulates cell survival pathways, and enhances glycolysis. Several promising investigational compounds that promote HIF-1 signaling are being developed. Nevertheless, there are still no reliable therapeutic approaches for reducing the injurious consequences of ischemia in clinical situations. The strategy that is perhaps the most useful in ischemic (and traumatic) brain and spinal cord injury is the transient induction of hypothermia (lowering the core body temperature to 92°F). This treatment reduces the metabolic demands of the stressed cells, decreases cell swelling, suppresses the formation of free radicals, and inhibits the host inflammatory response. All of these may contribute to decreased cell and tissue injury.

Ischemia-Reperfusion Injury

Restoration of blood flow to ischemic tissues can promote recovery of cells if they are reversibly injured but can also paradoxically exacerbate cell injury and cause cell death. As a consequence, reperfused tissues may sustain loss of viable cells in addition to those that are irreversibly damaged by the ischemia. This process, called *ischemia-reperfusion injury,* is clinically important because it contributes to tissue damage during myocardial and cerebral infarction following therapies that restore blood flow (Chapters 12 and 28).

How does reperfusion injury occur? The likely answer is that new damaging processes are set in motion during reperfusion, causing the death of cells that might have recovered otherwise. Several mechanisms have been proposed:

- *Oxidative stress.* New damage may be initiated during reoxygenation by increased generation of *reactive oxygen and nitrogen species.* These free radicals may be produced in reperfused tissue as a result of incomplete reduction of oxygen in leukocytes, and in damaged endothelial cells and parenchymal cells. Compromise of cellular antioxidant defense mechanisms during ischemia may sensitize cells to free radical damage.
- *Intracellular calcium overload.* As mentioned earlier, intracellular and mitochondrial calcium overload begins during acute ischemia; it is exacerbated during reperfusion due to influx of calcium resulting from cell membrane damage and ROS-mediated injury to sarcoplasmic reticulum. Calcium overload favors opening of the mitochondrial permeability transition pore with resultant ATP depletion. This, in turn, causes further cell injury.
- *Inflammation*. Ischemic injury is associated with inflammation as a result of "danger signals" released from dead cells, cytokines secreted by resident immune cells such as macrophages, and increased expression of adhesion molecules by hypoxic parenchymal and endothelial cells, all of which act to recruit circulating neutrophils to reperfused tissue. The inflammation causes additional tissue injury (Chapter 3). The importance of neutrophil influx in reperfusion injury has been demonstrated experimentally by the salutary effects of treatment with antibodies that block cytokines or adhesion molecules and thereby reduce neutrophil extravasation.
- Activation of the *complement system* may contribute to ischemia-reperfusion injury. For unknown reasons, some IgM antibodies have a propensity to deposit in ischemic tissues. When blood flow is resumed, complement proteins bind to the deposited antibodies, are activated, and exacerbate cell injury and inflammation.

Chemical (Toxic) Injury

Chemical injury remains a frequent problem in clinical medicine and is a major limitation to drug therapy. Because many drugs are metabolized in the liver, this organ is a major target of drug toxicity. In fact, toxic liver injury is often the reason for terminating the therapeutic use or development of a drug. The mechanisms by which chemicals, certain drugs, and toxins produce injury are described in greater detail in Chapter 9 in the discussion of environmental diseases. Here the major pathways of chemically induced injury with selected examples are described.

Chemicals induce cell injury by one of two general mechanisms:

- *Direct toxicity.* Some chemicals injure cells directly by combining with critical molecular components. For example, in mercuric chloride poisoning, *mercury* binds to the sulfhydryl groups of cell membrane proteins, causing increased membrane permeability and inhibition of ion transport. In such instances, the greatest damage is usually to the cells that use, absorb, excrete, or concentrate the chemicals—in the case of mercuric chloride, the cells of

the gastrointestinal tract and kidney (Chapter 9). *Cyanide* poisons mitochondrial cytochrome oxidase and thus inhibits oxidative phosphorylation. Many antineoplastic chemotherapeutic agents and antibiotics also induce cell damage by direct cytotoxic effects.

- *Conversion to toxic metabolites.* Most toxic chemicals are not biologically active in their native form but must be converted to reactive toxic metabolites, which then act on target molecules. This modification is usually accomplished by the cytochrome P-450 mixed-function oxidases in the smooth ER of the liver and other organs. The toxic metabolites cause membrane damage and cell injury mainly by formation of free radicals and subsequent lipid peroxidation; direct covalent binding to membrane proteins and lipids may also contribute. For instance, CCl_4, which was once widely used in the dry cleaning industry, is converted by cytochrome P-450 to the highly reactive free radical $^{\cdot}CCl_3$, which causes lipid peroxidation and damages many cellular structures. The analgesic drug acetaminophen is also converted to a toxic product during detoxification in the liver, leading to cell injury. These and other examples of chemical injury are described in Chapter 9.

KEY CONCEPTS

EXAMPLES OF CELL INJURY

- Mild ischemia: Reduced oxidative phosphorylation → ATP depletion → failure of Na pump → influx of sodium and water → organelle and cellular swelling (reversible)
- Severe/prolonged ischemia: Severe swelling of mitochondria, calcium influx into mitochondria and into the cell with rupture of lysosomes and plasma membrane. Death by necrosis and apoptosis due to release of cytochrome c from mitochondria.
- Reperfusion injury follows restoration of blood flow to ischemic tissues and is caused by oxidative stress due to release of free radicals from leukocytes and endothelial cells. Blood brings calcium that overloads reversibly injured cells with consequent mitochondrial injury, as well as oxygen and leukocytes, which generate free radicals and cytokines. Complement may be activated locally by IgM antibodies deposited in ischemic tissues.
- Chemicals may cause injury directly or by conversion into toxic metabolites. The organs chiefly affected are those involved in absorption or excretion of chemicals or others such as liver, where the chemicals are converted to toxic metabolites. Direct injury to critical organelles such as mitochondria or indirect injury from free radicals generated from the chemicals/toxins is involved.

ADAPTATIONS OF CELLULAR GROWTH AND DIFFERENTIATION

Adaptations are reversible changes in the size, number, phenotype, metabolic activity, or functions of cells in response to changes in their environment. Such adaptations may take several distinct forms.

Hypertrophy

Hypertrophy is an increase in the size of cells that results in an increase in the size of the affected organ. The hypertrophied organ has no new cells, just larger cells. The increased size of the cells is due to the synthesis and assembly of additional intracellular structural components. Cells capable of division may respond to stress by undergoing both hyperplasia (described later) and hypertrophy, whereas nondividing cells (e.g., myocardial fibers) increase tissue mass due to hypertrophy. In many sites, hypertrophy and hyperplasia may coexist, with both contributing to increased organ size.

Hypertrophy can be *physiologic* or *pathologic;* the former is caused by increased functional demand or stimulation by hormones and growth factors.

- *Pathologic hypertrophy.* The striated muscle cells in the heart and skeletal muscles have only a limited capacity for division, and respond to increased metabolic demands mainly by undergoing hypertrophy. *The most common stimulus for hypertrophy of skeletal and cardiac muscle is increased workload.* In both tissue types, muscle cells respond by synthesizing more protein and increasing the number of myofilaments per cell. This in turn increases the amount of force each myocyte can generate and thus the strength and work capacity of the muscle as a whole. A classic example of pathologic hypertrophy is enlargement of the heart in response to pressure overload, usually resulting from either hypertension or valvular disease (see Fig. 2.2). Initially, cardiac hypertrophy improves function, but over time this adaptation often fails, setting the stage for heart failure and other significant forms of heart disease (Chapter 12).
- *Physiologic hypertrophy.* The massive physiologic growth of the uterus during pregnancy is a good example of hormone-induced enlargement of an organ that results mainly from hypertrophy of smooth muscle fibers (Fig. 2.25). Uterine hypertrophy during pregnancy is stimulated by estrogenic hormone signaling through estrogen receptors that eventually result in increased synthesis of smooth muscle proteins and an increased cell size. The bulging muscles of bodybuilders engaged in "pumping iron" result from enlargement of individual skeletal muscle fibers in response to increased demand.

Mechanisms of Hypertrophy

Hypertrophy is a result of increased cellular protein production. Much of our understanding of hypertrophy is based on studies of the heart. There is great interest in defining the molecular basis of myocardial hypertrophy because beyond a certain point, it becomes maladaptive. Hypertrophy results from the action of growth factors and direct effects on cellular proteins (Fig. 2.26):

- Mechanical sensors in the cell detect the increased load.
- These sensors activate a complex downstream web of signaling pathways, including the phosphoinositide 3-kinase (PI3K)/AKT pathway (postulated to be most important in physiologic, e.g., exercise-induced, hypertrophy) and G-protein–coupled receptor–initiated pathways (activated by many growth factors and vasoactive agents, and thought to be more important in pathologic hypertrophy).
- Some of the signaling pathways stimulate increased production of growth factors (e.g. TGF-β, insulin-like growth factor 1 [IGF1], fibroblast growth factor) and vasoactive agents (e.g., α-adrenergic agonists, endothelin-1, and angiotensin II).

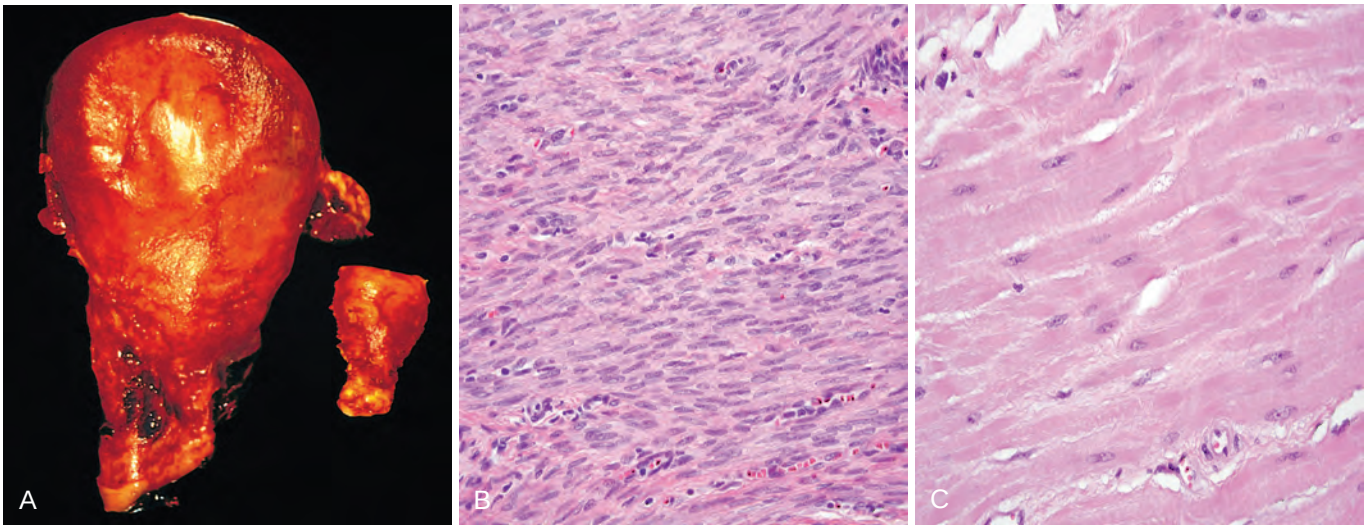


Figure 2.25 Physiologic hypertrophy of the uterus during pregnancy. (A) Gross appearance of a normal uterus *(right)* and a gravid uterus (removed for postpartum bleeding) *(left)*. (B) Small spindle-shaped uterine smooth muscle cells from a normal uterus, compared with (C) large plump cells from the gravid uterus, at the same magnification.

- These and other pathways activate transcription factors, including GATA4, nuclear factor of activated T cells (NFAT), and myocyte enhancer factor 2 (MEF2), which increase the expression of genes that encode muscle proteins.

Cardiac hypertrophy is also associated with a switch in gene expression from genes that encode adult-type contractile proteins to genes that encode functionally distinct fetal isoforms of the same proteins. For example, the α isoform of myosin heavy chain is replaced by the β isoform, which has a slower, more energetically economical contraction. Other proteins that are altered in hypertrophic myocardial cells are the products of genes that participate in the cellular response to stress. For example, cardiac hypertrophy is associated with increased atrial natriuretic factor gene expression. Atrial natriuretic factor is a peptide hormone that causes salt secretion by the kidney, decreases

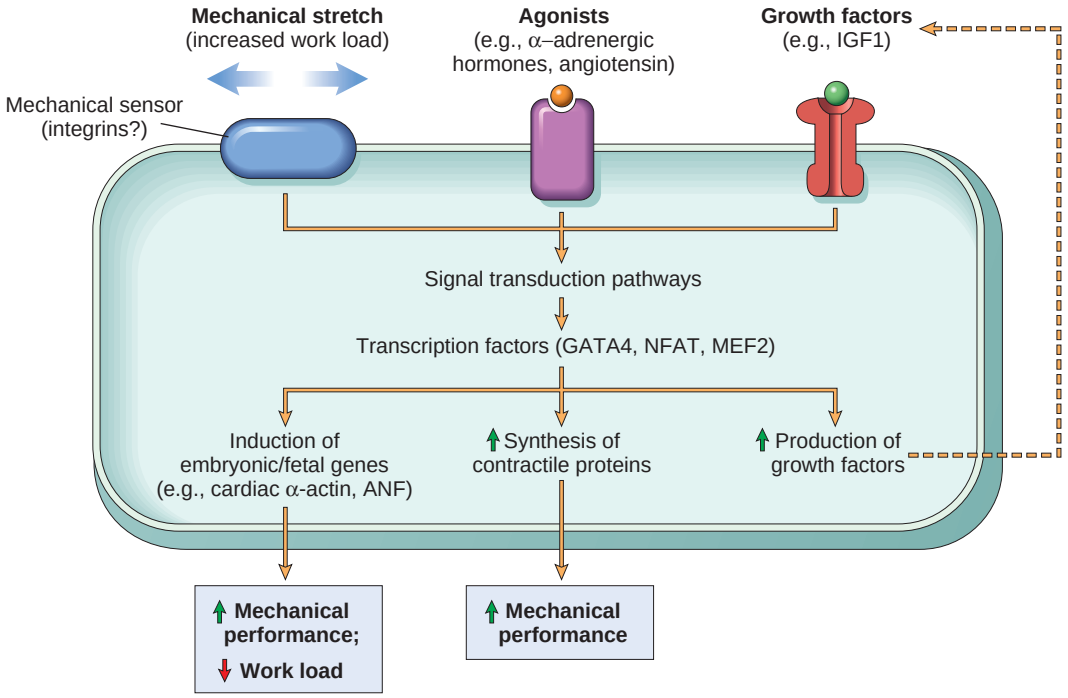


Figure 2.26 Biochemical mechanisms of myocardial hypertrophy. The major known signaling pathways and their functional effects are shown. Mechanical sensors appear to be the major triggers for physiologic hypertrophy, and agonists and growth factors may be more important in pathologic states. *ANF,* Atrial natriuretic factor; *GATA4,* transcription factor that binds to DNA sequence GATA; *IGF1,* insulin-like growth factor; *NFAT,* nuclear factor–activated T cells; *MEF2,* myocardial enhancing factor 2.

blood volume and pressure, and therefore serves to reduce hemodynamic load.

Whatever the exact cause and mechanism of cardiac hypertrophy, it eventually reaches a limit beyond which enlargement of muscle mass is no longer able to cope with the increased burden. At this stage, several regressive changes occur in the myocardial fibers, of which the most important are degradation and loss of myofibrillar contractile elements. In extreme cases, myocyte death can occur. The net result of these changes is cardiac failure, a sequence of events that illustrates how an adaptation to stress can progress to functionally significant cell injury if the stress is not relieved.

Hyperplasia

Hyperplasia is an increase in the number of cells in an organ or tissue in response to a stimulus. Although hyperplasia and hypertrophy are distinct processes, they frequently occur together, and may be triggered by the same external stimuli. Hyperplasia can only take place if the tissue contains cells capable of dividing, thus increasing the number of cells. It can be physiologic or pathologic.

- *Physiologic hyperplasia.* **Physiologic hyperplasia due to the action of hormones or growth factors occurs when there is a need to increase functional capacity of hormone sensitive organs, or when there is need for compensatory increase after damage or resection.** Hormonal hyperplasia is well illustrated by the proliferation of the glandular epithelium of the female breast at puberty and during pregnancy, usually accompanied by enlargement (hypertrophy) of the glandular epithelial cells. The classic illustration of compensatory hyperplasia comes from the study of liver regeneration. In individuals who donate one lobe of the liver for transplantation, the remaining cells proliferate so that the organ soon grows back to its original size. Experimental models of partial hepatectomy have been very useful for defining the mechanisms that stimulate regeneration of the liver (Chapter 3). The bone marrow is also remarkable in its capacity to undergo rapid hyperplasia in response to a deficiency of mature blood cells. For example, in the setting of an acute bleeding or premature breakdown of red blood cells (hemolysis), feedback loops involving the growth factor erythropoietin are activated that stimulate the growth of red blood cell progenitors, allowing red blood cell production to increase as much as eightfold. The regulation of hematopoiesis is discussed further in Chapter 13.
- *Pathologic hyperplasia.* **Most forms of pathologic hyperplasia are caused by excessive or inappropriate actions of hormones or growth factors acting on target cells.** Endometrial hyperplasia is an example of abnormal hormone-induced hyperplasia. Normally, after a menstrual period there is a rapid burst of proliferative activity in the endometrium that is stimulated by pituitary hormones and ovarian estrogen. It is brought to a halt by the rising levels of progesterone, usually about 10 to 14 days before the end of the menstrual period. In some instances, however, the balance between estrogen and progesterone is disturbed, resulting in absolute or relative increases in the amount of estrogen, with consequent hyperplasia of the endometrial glands. This form of pathologic hyperplasia is a common cause of abnormal uterine bleeding. Benign prostatic hyperplasia is another common example of pathologic hyperplasia, in this case as a response to hormonal stimulation by androgens. Although these forms of pathologic hyperplasias are abnormal, the process remains controlled and the hyperplasia can either regress or stabilize if the hormonal stimulation is eliminated. As is discussed in Chapter 7, the increased cell division associated with hyperplasia elevates the risk of acquiring genetic aberrations that drive unrestrained proliferation and cancer. Thus, *although hyperplasia is distinct from cancer, pathologic hyperplasia constitutes a fertile soil in which cancerous proliferations may eventually arise.* For instance, patients with hyperplasia of the endometrium are at increased risk for developing endometrial cancer (Chapter 22).

  Hyperplasia is a characteristic response to certain *viral infections,* such as papillomaviruses, which cause skin warts and several mucosal lesions composed of masses of hyperplastic epithelium. Here, the viruses make factors that interfere with host proteins that regulate cell proliferation. Like other forms of hyperplasia, some of these virally induced proliferations are also precursors to cancer (Chapter 7).

Mechanisms of Hyperplasia

Hyperplasia is the result of growth factor–driven proliferation of mature cells and, in some cases, by increased output of new cells from tissue stem cells. For instance, after partial hepatectomy, growth factors are produced in the liver that engage receptors on the surviving cells and activate signaling pathways that stimulate cell proliferation. But if the proliferative capacity of the liver cells is compromised, as in some forms of hepatitis causing cell injury, hepatocytes can instead regenerate from intrahepatic stem cells. The roles of growth factors and stem cells in cellular replication and tissue regeneration are discussed in more detail in Chapter 3.

Atrophy

Atrophy is a reduction in the size of an organ or tissue due to a decrease in cell size and number. Atrophy can be physiologic or pathologic. *Physiologic atrophy* is common during normal development. Some embryonic structures, such as the notochord and thyroglossal duct, undergo atrophy during fetal development. The decrease in the size of the uterus that occurs shortly after parturition is another form of physiologic atrophy.

Pathologic atrophy has several causes, and it can be local or generalized. Common causes of atrophy include the following:

- *Decreased workload (disuse atrophy).* When a fractured bone is immobilized in a plaster cast or when a patient is restricted to complete bed rest, skeletal muscle atrophy rapidly ensues. The initial decrease in cell size is reversible once activity is resumed. With more prolonged disuse, skeletal muscle fibers decrease in number (due to apoptosis) as well as in size; muscle atrophy can be accompanied by increased bone resorption, leading to osteoporosis of disuse.
- *Loss of innervation (denervation atrophy).* The normal metabolism and function of skeletal muscle are dependent

on its nerve supply. Damage to the nerves leads to atrophy of the muscle fibers supplied by those nerves (Chapter 27).

- *Diminished blood supply.* A gradual decrease in blood supply (chronic ischemia) to a tissue as a result of slowly developing arterial occlusive disease results in tissue atrophy. In late adult life, the brain may undergo progressive atrophy, mainly because of reduced blood supply as a result of atherosclerosis (Fig. 2.27). This is called *senile atrophy*.
- *Inadequate nutrition.* Profound protein-calorie malnutrition (marasmus) is associated with the utilization of skeletal muscle proteins as a source of energy after other reserves such as adipose stores have been depleted. This results in marked muscle wasting (*cachexia;* Chapter 9). Cachexia is also seen in patients with chronic inflammatory diseases and cancer. In some cachectic states, chronic overproduction of the inflammatory cytokine TNF is thought to be responsible for appetite suppression and lipid depletion, culminating in muscle atrophy.
- *Loss of endocrine stimulation.* Many hormone-responsive tissues, such as the breast and reproductive organs, are dependent on endocrine stimulation for normal metabolism and function. The loss of estrogen stimulation after menopause results in atrophy of the endometrium, vaginal epithelium, and breast. Similarly, the prostrate atrophies following chemical or surgical castration (e.g., for treatment of prostate cancer).
- *Pressure.* Tissue compression for any length of time can cause atrophy. An enlarging benign tumor can cause atrophy in the surrounding uninvolved tissues. Atrophy in this setting is probably the result of ischemic changes caused by compromise of the blood supply by the pressure exerted by the expanding mass.

The fundamental cellular changes associated with atrophy are similar in all of these settings. The initial response is a decrease in cell size and organelles, which may reduce the metabolic needs of the cell sufficiently to permit its survival. In atrophic muscle, the cells contain fewer mitochondria and myofilaments and a reduced amount of rough ER. By bringing into balance the cell's metabolic demands and the lower levels of blood supply, nutrition, or trophic stimulation, a new equilibrium is achieved. *Early in the process, atrophic cells and tissues have diminished function, but cell death is minimal*. However, atrophy caused by gradually reduced blood supply may progress to the point at which cells are irreversibly injured and die, often by apoptosis. Cell death by apoptosis also contributes to the atrophy of endocrine organs after hormone withdrawal.

Mechanisms of Atrophy

Atrophy results from decreased protein synthesis and increased protein degradation in cells. Protein synthesis decreases because of reduced trophic signals (e.g., those produced by growth receptors), which enhance uptake of nutrients and increase mRNA translation.

The degradation of cellular proteins occurs mainly by the *ubiquitin-proteasome pathway*. Nutrient deficiency and disuse may activate ubiquitin ligases, which attach the small peptide ubiquitin to cellular proteins and target these proteins for degradation in proteasomes. This pathway is also thought to be responsible for the accelerated proteolysis seen in a variety of catabolic conditions, including cancer cachexia. In many situations, atrophy is also accompanied by increased *autophagy,* marked by the appearance of increased numbers of autophagic vacuoles. Some of the cell debris within the autophagic vacuoles may resist digestion and persist in the cytoplasm as membrane-bound *residual bodies*. An example of residual bodies is *lipofuscin granules,* discussed later in the chapter. When present in sufficient amounts, they impart a brown discoloration to the tissue *(brown atrophy)*. Autophagy is associated with various types of cell injury, as discussed earlier.

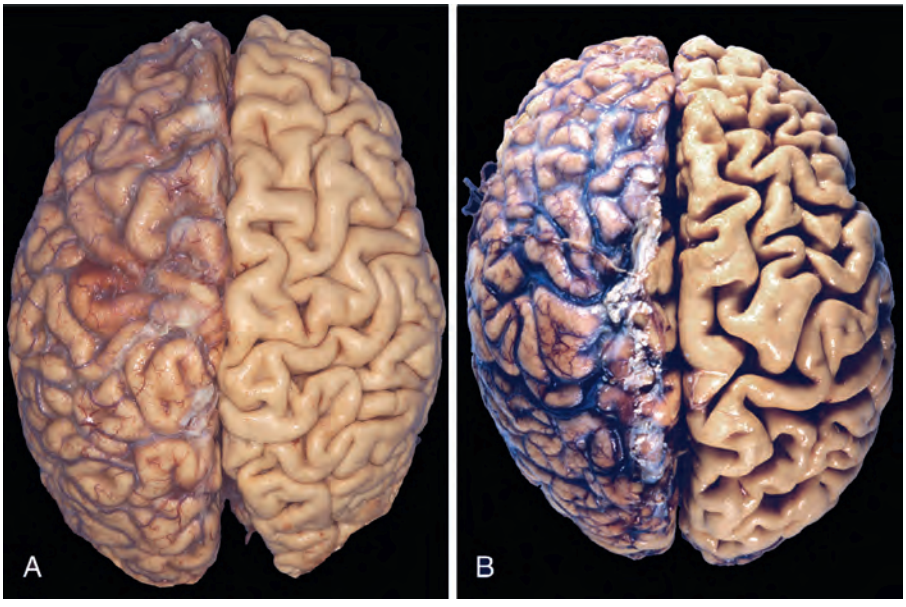


Figure 2.27 Atrophy. (A) Normal brain of a young adult. (B) Atrophy of the brain in an 82-year-old man with atherosclerotic cerebrovascular disease, resulting in reduced blood supply. Note that loss of brain substance narrows the gyri and widens the sulci. The meninges have been stripped from the right half of each specimen to reveal the surface of the brain.

Metaplasia

Metaplasia is a reversible change in which one differentiated cell type (epithelial or mesenchymal) is replaced by another cell type. It often represents an adaptive response in which one cell type that is sensitive to a particular stress is replaced by another cell type that is better able to withstand the adverse environment.

The most common epithelial metaplasia is columnar to squamous (Fig. 2.28), as occurs in the respiratory tract in response to chronic irritation. In the habitual cigarette smoker, the normal ciliated columnar epithelial cells of the trachea and bronchi are often replaced by stratified squamous epithelial cells. Vitamin A (retinoic acid) deficiency can also induce squamous metaplasia in the respiratory epithelium and in the cornea, the latter with highly deleterious effects on vision (Chapter 9). Stones in the excretory ducts of the salivary glands, pancreas, or bile ducts, which are normally lined by secretory columnar epithelium, may also lead to squamous metaplasia. In all these instances, the more rugged stratified squamous epithelium is able to survive under circumstances in which the more fragile specialized columnar epithelium might have succumbed. However, the change to metaplastic squamous cells comes with a price. In the respiratory tract, for example, although the epithelial lining becomes more durable, important mechanisms of protection against infection—mucus secretion and the ciliary action of the columnar epithelium—are lost. Thus, epithelial metaplasia, in most circumstances, represents an undesirable change. Moreover, *the influences that predispose to metaplasia, if persistent, can initiate malignant transformation in metaplastic epithelium*. The development of squamous cell carcinoma in areas of the lungs where the normal columnar epithelium has been replaced by squamous epithelium is one example.

Metaplasia from squamous to columnar type may also occur, as in *Barrett esophagus,* in which the esophageal squamous epithelium is replaced by intestinal-like columnar cells under the influence of refluxed gastric acid. As might be expected, the cancers that arise in these areas are typically glandular (adenocarcinomas) (Chapter 17).

Connective tissue metaplasia is the formation of cartilage, bone, or adipose cells (mesenchymal tissues) in tissues that normally do not contain these elements. For example, bone formation in muscle, designated *myositis ossificans,* occasionally occurs after intramuscular hemorrhage. This type of metaplasia is less clearly seen as an adaptive response, and may be a result of cell or tissue injury. Unlike epithelial metaplasia, this type of metaplasia is not associated with increased cancer risk.

Mechanisms of Metaplasia

Metaplasia does not result from a change in the phenotype of an already differentiated cell type; rather, it results from either reprogramming of local tissue stem cells or, alternatively, colonization by differentiated cell populations from adjacent sites. In either case, the metaplastic change is stimulated by signals generated by cytokines, growth factors, and extracellular matrix components in the cells' environment. In the case of stem cell reprogramming, these external stimuli promote the expression of genes that drive cells toward a specific differentiation pathway. A direct link between transcription factor dysregulation and metaplasia is seen with vitamin A (retinoic acid) deficiency or excess, both of which may cause metaplasia. Retinoic acid regulates gene transcription directly through nuclear retinoid receptors (Chapter 9), which can influence the differentiation of progenitors derived from tissue stem cells.

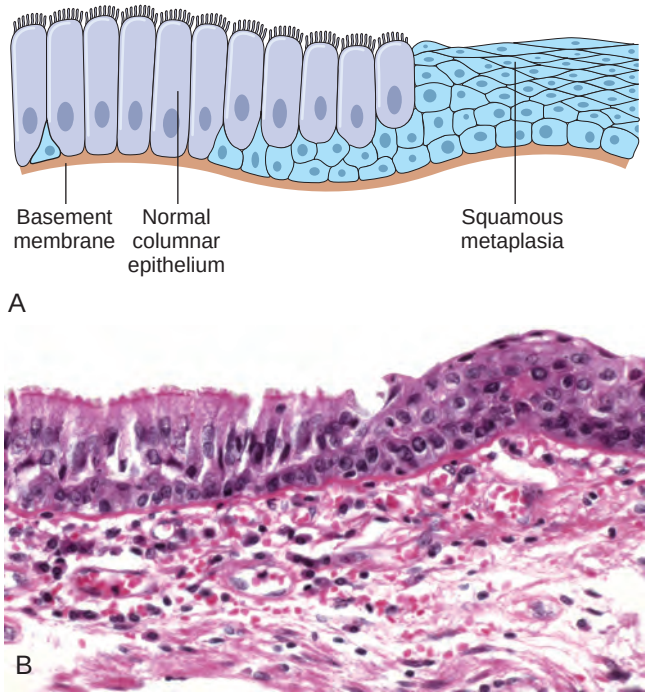


Figure 2.28 Metaplasia of columnar to squamous epithelium. (A) Schematic diagram. (B) Metaplasia of columnar epithelium *(left)* to squamous epithelium *(right)* in a bronchus (as often occurs with smoking).

KEY CONCEPTS

CELLULAR ADAPTATIONS TO STRESS

- Hypertrophy: increased cell and organ size, often in response to increased workload; induced by growth factors produced in response to mechanical stress or other stimuli; occurs in tissues incapable of cell division. It may be physiologic (e.g., enlargement of the uterus in pregnancy) or pathologic (e.g., enlargement of the heart in hypertension of valvular disease).
- Hyperplasia: increased cell numbers in response to hormones and other growth factors; occurs in tissues whose cells are able to divide or contain abundant tissue stem cells. It may be physiologic in response to increased need (e.g., breast acini during lactation) or pathologic in response to inappropriate secretion of hormones (e.g., endometrial hyperplasia due to excessive estrogenic stimulation).
- Atrophy: decreased cell and organ size, as a result of decreased nutrient supply or disuse; associated with decreased synthesis of cellular building blocks and increased breakdown of cellular organelles by increased autophagy
- Metaplasia: change in phenotype of differentiated cells, often in response to chronic irritation, that makes cells better able to withstand the stress; usually induced by altered differentiation pathway of tissue stem cells; may result in reduced functions or increased propensity for malignant transformation

INTRACELLULAR ACCUMULATIONS

One of the manifestations of metabolic derangements in cells is the intracellular accumulation of substances that may be harmless or cause further injury. These accumulations may be located in the cytoplasm, within organelles (typically lysosomes), or in the nucleus, and they may be composed of substances that are synthesized by the affected cells or are produced elsewhere.

There are four main mechanisms leading to abnormal intracellular accumulations (Fig. 2.29):

- *Inadequate removal* of a normal substance secondary to defects in packaging and transport, as in fatty change (steatosis) in the liver (Chapter 18)
- Accumulation of an endogenous substance as a result of genetic or acquired *defects in its folding, packaging, transport, or secretion,* as with certain mutated forms of α_1-antitrypsin (Chapter 15)
- *Failure to degrade* a metabolite due to inherited enzyme deficiencies, typically lysosomal enzymes. The resulting disorders are called *lysosomal storage diseases* (Chapter 5).
- Deposition and accumulation of an *abnormal exogenous substance* when the cell has neither the enzymatic machinery to degrade the substance nor the ability to transport it to other sites. Accumulation of carbon or silica particles is an example of this type of alteration (Chapter 15).

In many cases, if the overload can be controlled or stopped, the accumulation is reversible. In inherited storage diseases, accumulation is progressive and may cause cellular injury, leading in some instances to death of the tissue and the patient.

Lipids

All major classes of lipids can accumulate in cells: triglycerides, cholesterol/cholesterol esters, and phospholipids. Phospholipids are components of the myelin figures found in necrotic cells. In addition, abnormal complexes of lipids and carbohydrates accumulate in the lysosomal storage diseases (Chapter 5). Triglyceride and cholesterol accumulations are discussed here.

Steatosis (Fatty Change)

The terms *steatosis* and *fatty change* describe abnormal accumulations of triglycerides within parenchymal cells. Fatty change is often seen in the liver because it is the major organ involved in fat metabolism (Fig. 2.30), but it also occurs in the heart, muscle, and kidney. The causes of steatosis include toxins, protein malnutrition, diabetes mellitus, obesity, and anoxia. In higher-income nations, the most common causes of significant fatty change in the liver (fatty liver) are alcohol abuse and nonalcoholic fatty liver disease, which is often associated with diabetes and obesity. Fatty liver is discussed in more detail in Chapter 18.

Cholesterol and Cholesterol Esters

The cellular metabolism of cholesterol (Chapter 5) is tightly regulated such that most cells use cholesterol for the synthesis of cell membranes without intracellular accumulation of

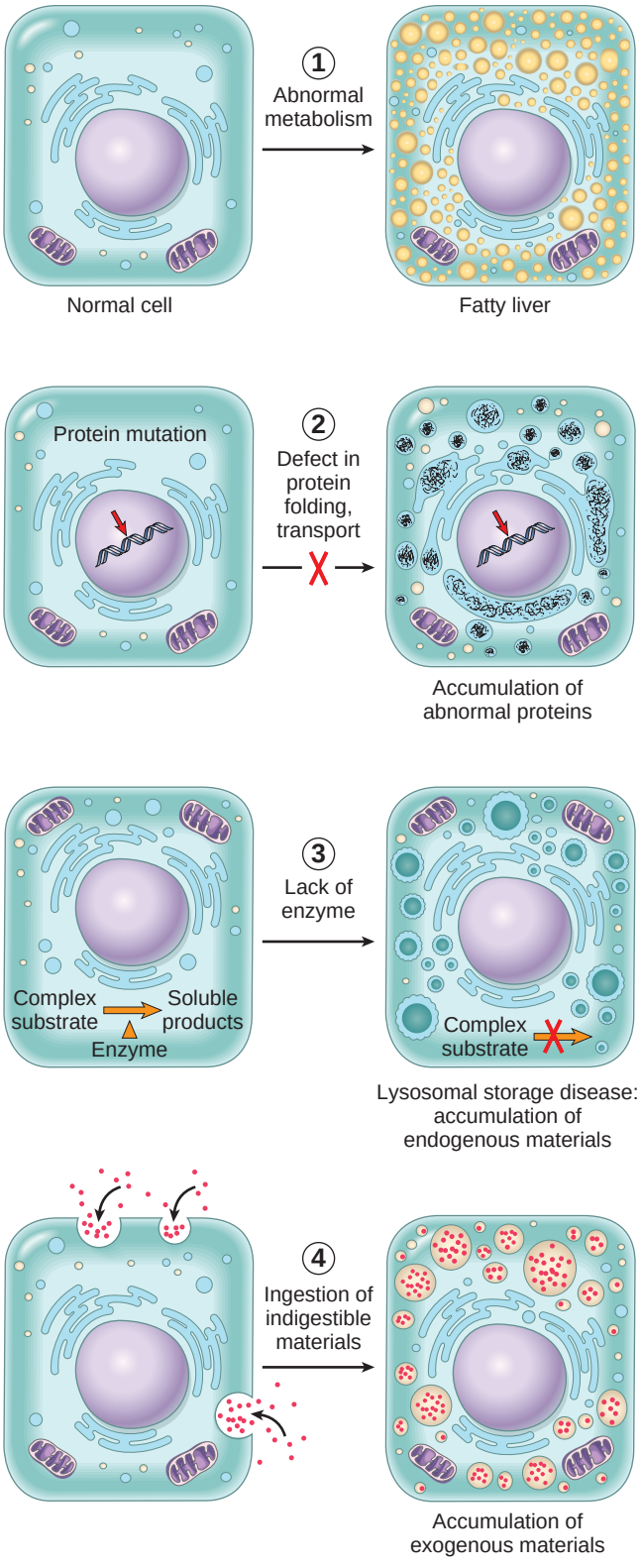


Figure 2.29 Mechanisms of intracellular accumulations discussed in the text.

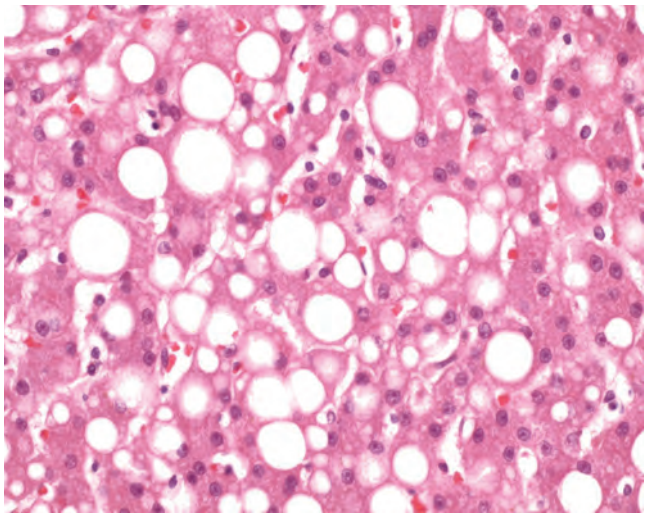


Figure 2.30 Fatty liver. High-power detail of fatty change of the liver. In most cells, the well-preserved nucleus is squeezed into the displaced rim of cytoplasm about the fat vacuole. (Courtesy Dr. James Crawford, Department of Pathology, Hofstra Northwell School of Medicine, NY.)

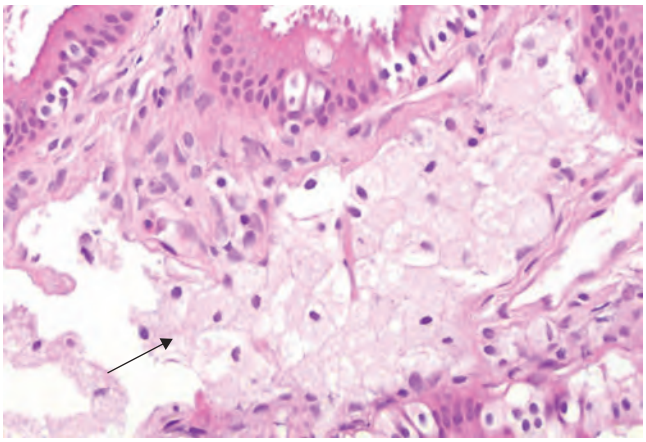


Figure 2.31 Cholesterolosis. Cholesterol-laden macrophages (foam cells, *arrow*) in a focus of gallbladder cholesterolosis. (Courtesy Dr. Matthew Yeh, Department of Pathology, University of Washington, Seattle, Wash.)

cholesterol or cholesterol esters. Accumulations manifested histologically by intracellular vacuoles are seen in several pathologic processes.

- *Atherosclerosis.* In atherosclerotic plaques, smooth muscle cells and macrophages within the intimal layer of the aorta and large arteries are filled with lipid vacuoles, most of which contain cholesterol and cholesterol esters. Such cells have a foamy appearance (foam cells), and aggregates of them in the intima produce the yellow cholesterol-laden atheromas characteristic of this serious disorder. Some of these fat-laden cells may rupture, releasing cholesterol and cholesterol esters into the extracellular space, where they may form crystals. Some form long needles that produce distinct clefts in tissue sections, while other small crystals are phagocytosed by macrophages and activate the inflammasome, contributing to local inflammation. The mechanisms of cholesterol accumulation and its pathogenic consequences in atherosclerosis are discussed in detail in Chapter 11.
- *Xanthomas.* Intracellular accumulation of cholesterol within macrophages is also characteristic of acquired and hereditary hyperlipidemic states. Clusters of foamy cells are found in the subepithelial connective tissue of the skin and in tendons, producing tumorous masses known as *xanthomas*.
- *Cholesterolosis.* This refers to the focal accumulations of cholesterol-laden macrophages in the lamina propria of the gallbladder (Fig. 2.31). The mechanism of accumulation is unknown.
- *Niemann-Pick disease, type C.* This lysosomal storage disease is caused by mutations affecting an enzyme involved in cholesterol trafficking, resulting in cholesterol accumulation in multiple organs (Chapter 5).

Proteins

Intracellular accumulations of proteins usually appear as rounded, eosinophilic droplets, vacuoles, or aggregates in the cytoplasm. By electron microscopy, they can be amorphous, fibrillar, or crystalline in appearance. In some disorders, such as certain forms of amyloidosis, abnormal proteins deposit primarily in extracellular spaces (Chapter 6).

Excesses of proteins within the cells sufficient to cause morphologically visible accumulation have diverse causes.

- *Reabsorption droplets in proximal renal tubules* are seen in renal diseases associated with protein loss in the urine (proteinuria). In the kidney, small amounts of protein filtered through the glomerulus are normally reabsorbed by pinocytosis in the proximal tubule. In disorders with heavy protein leakage across the glomerular filter, there is increased reabsorption of the protein into vesicles, and the protein appears as pink hyaline droplets within the cytoplasm of the tubular cell (Fig. 2.32). The process is reversible; if the proteinuria diminishes, the protein droplets are metabolized and disappear.
- The proteins that accumulate may be normal secreted proteins that are produced in excessive amounts and accumulate within the ER, as occurs in certain

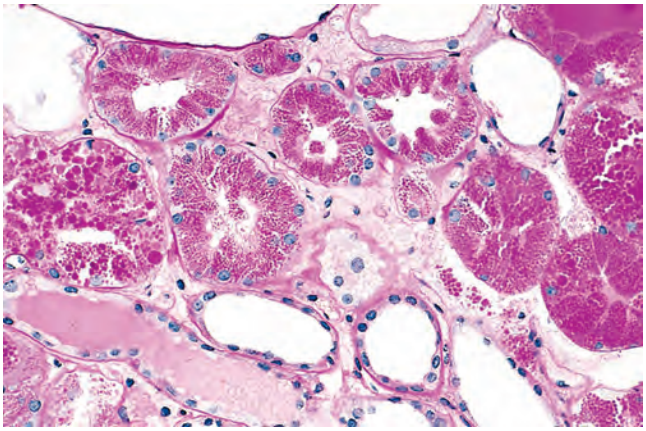


Figure 2.32 Protein reabsorption droplets in the renal tubular epithelium. (Courtesy Dr. Helmut Rennke, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

plasma cells engaged in active synthesis of immunoglobulins. The ER becomes hugely distended, producing large, homogeneous eosinophilic inclusions called *Russell bodies*.

- *Defective intracellular transport and secretion of critical proteins.* In α_1-antitrypsin deficiency, mutations in the protein significantly slow folding, resulting in the buildup of partially folded intermediates, which aggregate in the ER of the hepatocyte and are not secreted. The resultant deficiency of the circulating enzyme where it is needed in the lung causes emphysema (Chapter 15). In many of these protein-folding diseases, the pathology results not only from loss of protein function but also ER stress caused by the misfolded proteins, which initiates the unfolded protein response and culminates in cell death by apoptosis (discussed earlier).
- *Accumulation of cytoskeletal proteins.* There are several types of cytoskeletal proteins, including microtubules (20 to 25 nm in diameter), thin actin filaments (6 to 8 nm), thick myosin filaments (15 nm), and intermediate filaments (10 nm). Intermediate filaments, which provide a flexible intracellular scaffold that organizes the cytoplasm and resists forces applied to the cell, are divided into five classes: keratin filaments (characteristic of epithelial cells), neurofilaments (neurons), desmin filaments (muscle cells), vimentin filaments (connective tissue cells), and glial filaments (astrocytes). Accumulations of keratin filaments and neurofilaments are associated with certain types of cell injury. *Alcoholic hyaline* is an eosinophilic cytoplasmic inclusion in liver cells that is characteristic of alcoholic liver disease and is composed predominantly of keratin intermediate filaments (Chapter 18). The *neurofibrillary tangle* found in the brain in Alzheimer disease contains neurofilaments and other proteins (Chapter 28).
- *Aggregation of abnormal proteins.* Abnormal or misfolded proteins may deposit in tissues and interfere with normal functions. The deposits can be intracellular, extracellular, or both, and the aggregates may either directly or indirectly cause the pathologic changes. Certain forms of *amyloidosis* (Chapter 6) fall in this category of diseases. These disorders are sometimes called *proteinopathies* or *protein-aggregation diseases*.

Hyaline Change

The term *hyaline* usually refers to an alteration within cells or in the extracellular space that gives a homogeneous, glassy, pink appearance in routine histologic sections stained with H&E. It is widely used as a descriptive histologic term rather than a specific marker for cell injury. This morphologic change is produced by a variety of alterations and does not represent a specific pattern of accumulation.

Intracellular hyaline accumulations of protein iclude reabsorption droplets, Russell bodies, and alcoholic hyaline (described earlier). *Extracellular hyaline* has been more difficult to analyze. Collagenous fibers in old scars may appear hyalinized, but the biochemical basis of this change is not clear. In long-standing hypertension and diabetes mellitus, the walls of arterioles, especially in the kidney, become hyalinized, resulting from extravasated plasma protein and deposition of basement membrane material.

Glycogen

Excessive intracellular deposits of glycogen are seen in patients with an abnormality in either glucose or glycogen metabolism. Glycogen is a readily available source of glucose stored in the cytoplasm of healthy cells. Whatever the clinical setting, the glycogen masses appear as clear vacuoles within the cytoplasm because glycogen dissolves in aqueous fixatives; thus, it is most readily identified when tissues are fixed in absolute alcohol. Staining with Best carmine or the PAS reaction imparts a rose-to-violet color to the glycogen, but can also stain protein-bound carbohydrates. Diastase digestion of a parallel section that demonstrates loss of staining due to glycogen hydrolysis is therefore an important validation.

Diabetes mellitus is the prime example of a disorder of glucose metabolism. In this disease, glycogen is found in renal tubular epithelial cells, as well as within liver cells, β cells of the islets of Langerhans within the pancreas, and heart muscle cells.

Glycogen accumulates within select cells in a group of related genetic disorders that are collectively referred to as the *glycogen storage diseases*, or *glycogenoses* (Chapter 5). In these diseases, enzymatic defects in the synthesis or breakdown of glycogen result in massive accumulation, causing cell injury and cell death.

Pigments

Pigments are colored substances, some of which are normal constituents of cells (e.g., melanin), whereas others are abnormal and accumulate in cells under special circumstances. Pigments can be exogenous, coming from outside the body, or endogenous, synthesized within the body itself.

Exogenous Pigments

The most common exogenous pigment is carbon (coal dust), a ubiquitous air pollutant in urban areas. When inhaled, it is picked up by macrophages within the alveoli and then transported through lymphatic channels to lymph nodes in the tracheobronchial region. Accumulations of this pigment blacken the tissues of the lungs *(anthracosis)* and the involved lymph nodes. In coal miners, the aggregates of carbon dust may induce a fibroblastic reaction or even emphysema, and thus cause a serious lung disease known as *coal worker's pneumoconiosis* (Chapter 15). *Tattooing* is a form of localized, exogenous pigmentation of the skin. The pigments inoculated are phagocytosed by dermal macrophages, in which they reside for the remainder of the life of the embellished. The pigments do not usually evoke any inflammatory response.

Endogenous Pigments

Lipofuscin is an insoluble pigment, also known as lipochrome or wear-and-tear pigment. Lipofuscin is composed of polymers of lipids and phospholipids in complex with protein, suggesting that it is derived through lipid peroxidation of polyunsaturated lipids of intracellular membranes. Lipofuscin is not injurious to the cell or its functions. Its importance lies in its being a telltale sign of free radical injury and lipid peroxidation. The term is

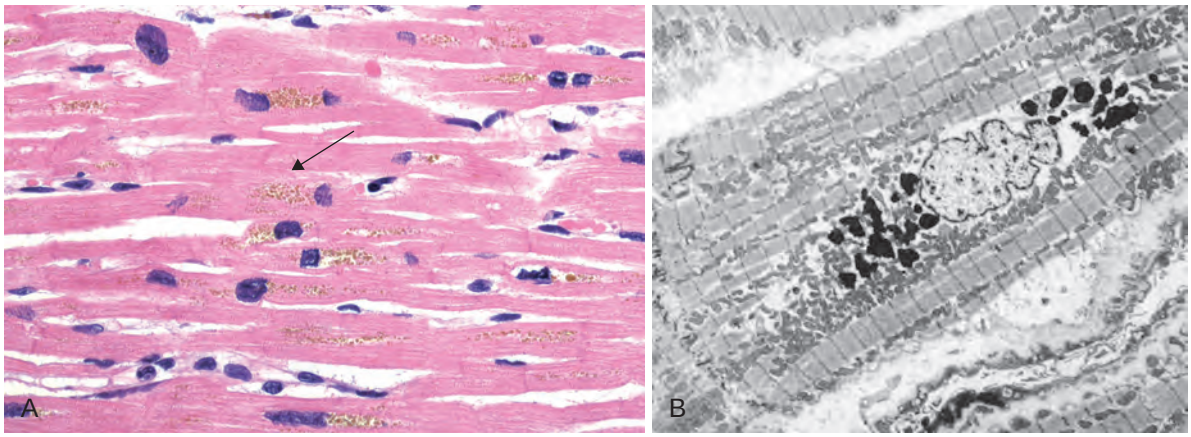


Figure 2.33 Lipofuscin granules in cardiac myocytes shown by (A) light microscopy (deposits indicated by *arrow*), and (B) electron microscopy (note the perinuclear, intralysosomal location).

derived from the Latin (*fuscus,* brown), referring to brown lipid. In tissue sections, it appears as a yellow-brown, finely granular cytoplasmic, often perinuclear, pigment (Fig. 2.33). It is seen in cells undergoing slow, regressive changes and is particularly prominent in the liver and heart of aging patients or patients with severe malnutrition and cancer cachexia.

Melanin, derived from the Greek (*melas*, black), is an endogenous, brown-black, pigment formed when the enzyme tyrosinase catalyzes the oxidation of tyrosine to dihydroxyphenylalanine in melanocytes. It is discussed further in Chapter 25. For practical purposes, melanin is the *only endogenous brown-black pigment*. The only other that could be considered in this category is homogentisic acid, a black pigment that occurs in patients with *alkaptonuria,* a rare metabolic disease. Here the pigment is deposited in the skin, connective tissue, and cartilage, and the pigmentation is known as *ochronosis.*

Hemosiderin, a hemoglobin-derived, golden yellow-to-brown, granular, or crystalline pigment is one of the major storage forms of iron. Iron metabolism and hemosiderin are considered in detail in Chapters 14 and 18. Iron is normally carried by a specific transport protein called transferrin. In cells, it is stored in association with a protein, apoferritin, to form ferritin micelles. Ferritin is a constituent of most cell types. When there is a local or systemic excess of iron, ferritin forms *hemosiderin granules,* which are easily seen with the light microscope. Hemosiderin pigment represents aggregates of ferritin micelles. Under normal conditions, small amounts of hemosiderin can be seen in the mononuclear phagocytes of the bone marrow, spleen, and liver, which are responsible for recycling of iron derived from hemoglobin during the breakdown of effete red blood cells.

Local or systemic excesses of iron cause hemosiderin to accumulate within cells. *Local excesses* result from hemorrhages in tissues. The best example of localized hemosiderosis is the common bruise. Extravasated red blood cells at the site of injury are phagocytosed over several days by macrophages, which break down the hemoglobin and recover the iron. After removal of iron, the heme moiety is converted first to biliverdin ("green bile") and then to bilirubin ("red bile"). In parallel, the iron released from heme is incorporated into ferritin and eventually hemosiderin. These conversions account for the often dramatic play of colors seen in a healing bruise, which typically changes from red-blue to green-blue to golden-yellow before it is resolved.

When there is *systemic iron overload*, hemosiderin may be deposited in many organs and tissues, a condition called *hemosiderosis*. The main causes of hemosiderosis are (1) increased absorption of dietary iron due to an inborn error of metabolism called *hemochromatosis* (Chapter 18), (2) hemolytic anemias, in which excessive lysis of red blood cells leads to release of abnormal quantities of iron (Chapter 14), and (3) repeated blood transfusions, because transfused red blood cells constitute an exogenous iron load.

PATHOLOGIC CALCIFICATION

Pathologic calcification is the abnormal tissue deposition of calcium salts, together with smaller amounts of iron, magnesium, and other mineral salts. There are two forms of pathologic calcification. When the deposition occurs locally in dying tissues, it is known as *dystrophic calcification;* it occurs despite normal serum levels of calcium and in the absence of derangements in calcium metabolism. In contrast, the deposition of calcium salts in otherwise normal tissues is known as *metastatic calcification,* and it almost always results from hypercalcemia secondary to some disturbance in calcium metabolism.

Dystrophic Calcification

Dystrophic calcification is encountered in areas of necrosis, whether they are of coagulative, caseous, or liquefactive type, and in foci of enzymatic necrosis of fat. Calcification is almost always present in the atheromas of advanced atherosclerosis. It also commonly develops in aging or damaged heart valves, further hampering their function (Fig. 2.34). Whatever the site of deposition, the calcium salts appear macroscopically as fine, white granules or clumps, often felt as gritty deposits. Sometimes a tuberculous lymph node is virtually converted to stone.

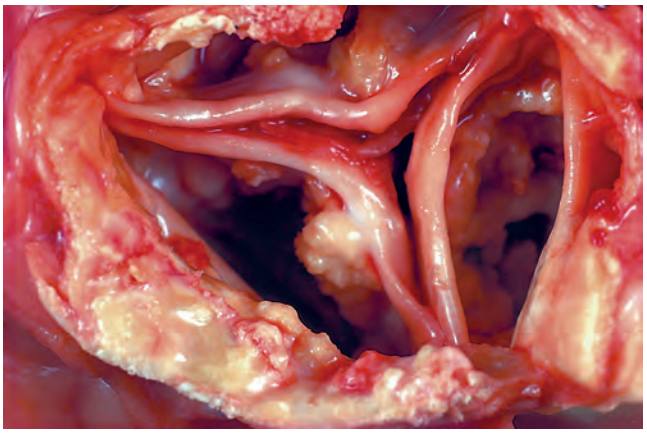


Figure 2.34 Dystrophic calcification of the aortic valve. View looking down onto the unopened aortic valve in a heart with calcific aortic stenosis. It is markedly narrowed (stenosis). The semilunar cusps are thickened and fibrotic, and behind each cusp are irregular masses of piled-up dystrophic calcification.

MORPHOLOGY

Histologically, with the usual H&E stain, calcium salts have a basophilic, amorphous granular, sometimes clumped appearance. They can be intracellular, extracellular, or in both locations. In the course of time, **heterotopic bone** may form in the focus of calcification. On occasion, single necrotic cells may constitute seed crystals that become encrusted by the mineral deposits. The progressive acquisition of outer layers may create lamellated configurations, called **psammoma bodies** because of their resemblance to grains of sand. Some types of papillary cancers (e.g., thyroid) are apt to develop psammoma bodies. In asbestosis, calcium and iron salts gather about long slender spicules of asbestos in the lung, creating exotic, beaded dumbbell forms known as **asbestos bodies** (Chapter 15).

Although dystrophic calcification may simply be a telltale sign of previous cell injury, it is often a cause of organ dysfunction. Such is the case in calcific valvular disease and atherosclerosis, as will become clear in further discussion of these diseases (Chapters 11 and 12). Serum calcium is normal in dystrophic calcification.

Metastatic Calcification

Metastatic calcification may occur in normal tissues whenever there is hypercalcemia. Hypercalcemia also accentuates dystrophic calcification. There are four principal causes of hypercalcemia: (1) *increased secretion of parathyroid hormone (PTH)* with subsequent bone resorption, as in hyperparathyroidism due to parathyroid tumors, and ectopic secretion of PTH-related protein by malignant tumors (Chapter 7); (2) *resorption of bone tissue,* secondary to primary tumors of bone marrow (e.g., multiple myeloma, leukemia) or diffuse skeletal metastasis (e.g., breast cancer), accelerated bone turnover (e.g., Paget disease), or immobilization; (3) *vitamin D–related disorders*, including vitamin D intoxication, sarcoidosis (in which macrophages activate a vitamin D precursor), and idiopathic hypercalcemia of infancy (Williams syndrome), characterized by abnormal sensitivity to vitamin D; and (4) *renal failure,* which causes retention of phosphate, leading to secondary hyperparathyroidism. Less common causes include aluminum intoxication, which occurs in patients on chronic renal dialysis, and milk-alkali syndrome, which is due to excessive ingestion of calcium and absorbable antacids such as milk or calcium carbonate.

Metastatic calcification may occur widely throughout the body but principally affects the interstitial tissues of the gastric mucosa, kidneys, lungs, systemic arteries, and pulmonary veins. Although quite different in location, all of these tissues excrete acid and therefore have an internal alkaline compartment that predisposes them to metastatic calcification. In all of these sites, the calcium salts morphologically resemble those described in dystrophic calcification. Thus, they may occur as noncrystalline amorphous deposits or, at other times, as hydroxyapatite crystals.

Usually the mineral salts cause no clinical dysfunction, but on occasion massive involvement of the lungs produces remarkable x-ray images and respiratory compromise. Massive deposits in the kidney (nephrocalcinosis) may in time cause renal damage (Chapter 20).

KEY CONCEPTS

ABNORMAL INTRACELLULAR DEPOSITIONS AND CALCIFICATIONS

Abnormal deposits of materials in cells and tissues are the result of excessive intake or defective transport or catabolism.

- Deposition of lipids
  - Fatty change: Accumulation of free triglycerides in cells, resulting from excessive intake or defective transport (often because of defects in synthesis of transport proteins); manifestation of reversible cell injury
  - Cholesterol deposition: Result of defective catabolism and excessive intake; in macrophages and smooth muscle cells of vessel walls in atherosclerosis
- Deposition of proteins: Reabsorbed proteins in kidney tubules; immunoglobulins in plasma cells
- Deposition of glycogen: In macrophages of patients with defects in lysosomal enzymes that break down glycogen (glycogen storage diseases) and in certain disorders of glycogen metabolism
- Deposition of pigments: Typically indigestible pigments, such as carbon, lipofuscin (breakdown product of lipid peroxidation), or iron (usually due to overload, as in hemosiderosis)
- Pathologic calcifications
  - Dystrophic calcification: Deposition of calcium at sites of cell injury and necrosis
  - Metastatic calcification: Deposition of calcium in normal tissues, caused by hypercalcemia (usually a consequence of parathyroid hormone excess)

CELLULAR AGING

Mankind has pursued immortality from time immemorial. Toth and Hermes, Egyptian and Greek deities, are said to have discovered the elixir of youth and become immortal. Sadly, Toth and Hermes are nowhere to be found, hence

the elixir remains a secret. Shakespeare probably characterized aging best in his elegant description of the seven ages of man. It begins at the moment of conception, involves the differentiation and maturation of the organism and its cells, at some variable point in time leads to the progressive loss of functional capacity characteristic of senescence, and ends in death.

Individuals age because their cells age. Although public attention on the aging process has traditionally focused on its cosmetic manifestations, aging has important health consequences, because age is one of the strongest independent risk factors for many chronic diseases, such as cancer, Alzheimer disease, and ischemic heart disease. Perhaps one of the most striking discoveries about cellular aging is that it is not simply a consequence of cells "running out of steam," but in fact is regulated by genes that are evolutionarily conserved from yeast to worms to mammals.

Cellular aging is the result of a progressive decline in cellular function and viability caused by genetic abnormalities and the accumulation of cellular and molecular damage due to the effects of exposure to exogenous influences (Fig. 2.35). Studies in model systems have clearly established that aging is influenced by a limited number of genes, and genetic anomalies underlie syndromes resembling premature aging in humans as well. Such findings suggest that aging is associated with definable mechanistic alterations. Several mechanisms, some cell intrinsic and others environmentally induced, are believed to play a role in aging.

DNA Damage. A variety of exogenous (physical, chemical, and biologic) agents and endogenous factors such as ROS threaten the integrity of nuclear and mitochondrial DNA. Although most DNA damage is repaired by DNA repair enzymes, some persists and accumulates as cells age. Several lines of evidence point to the importance of DNA repair in the aging process. Next-generation DNA sequencing studies have shown that the average hematopoietic stem cell suffers 14 new mutations per year, and it is likely that this accumulating damage explains why, like most cancers, the most common hematologic malignancies are diseases of the aged. Patients with *Werner syndrome* show premature aging, and the defective gene product is a DNA helicase, a protein involved in DNA replication and repair and other functions requiring DNA unwinding. A defect in this enzyme causes rapid accumulation of chromosomal damage that may mimic some aspects of the injury that normally accumulates during cellular aging. Genetic instability in somatic cells is also characteristic of other disorders in which patients display some of the manifestations of aging at an increased rate, such as *Bloom syndrome* and *ataxia-telangiectasia,* in which the mutated genes encode proteins involved in repairing double-strand breaks in DNA (Chapter 7).

Cellular Senescence. All normal cells have a limited capacity for replication, and after a fixed number of divisions cells become arrested in a terminally nondividing state, known as *replicative senescence.* Aging is associated with progressive replicative senescence of cells. Cells from children have the capacity to undergo more rounds of replication than do cells from older people. Two mechanisms are believed to underlie cellular senescence:

- *Telomere attrition*. One mechanism of replicative senescence involves progressive shortening of telomeres, which ultimately results in cell cycle arrest. *Telomeres* are short repeated sequences of DNA present at the ends of linear chromosomes that are important for ensuring the complete replication of chromosome ends and for protecting the ends from fusion and degradation. When somatic cells replicate, a small section of the telomere is not duplicated and telomeres become progressively shortened. As the telomeres become shorter, the ends of chromosomes cannot be protected and are seen as broken DNA, which signals cell cycle arrest. Telomere length is maintained by nucleotide addition mediated by an enzyme called *telomerase.* Telomerase is a specialized RNA-protein complex that uses its own RNA as a template for adding nucleotides to the ends of chromosomes. Telomerase is expressed in germ cells and is present at low levels in

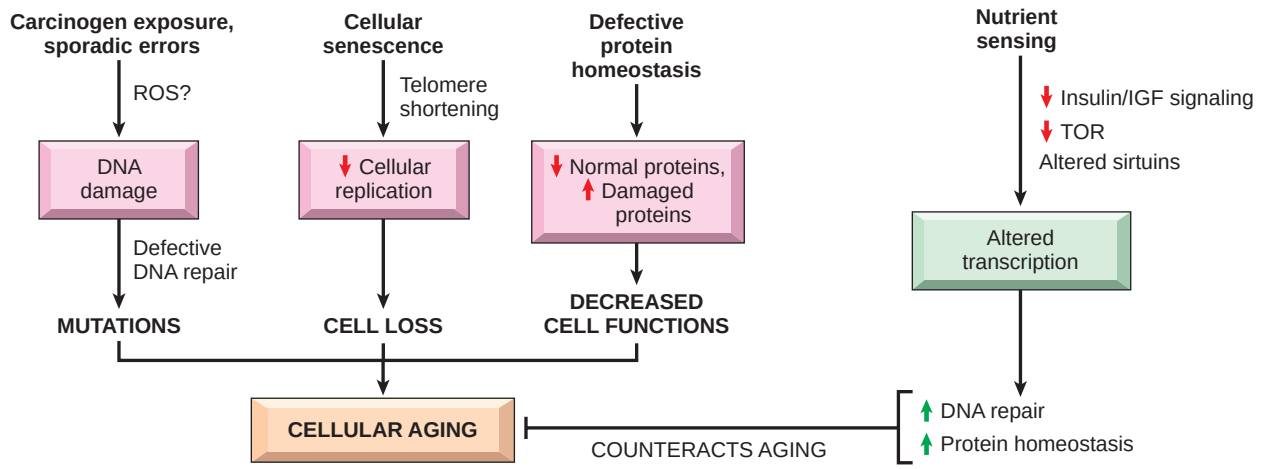


Figure 2.35 Mechanisms that cause and counteract cellular aging. DNA damage, replicative senescence, and decreased and misfolded proteins are among the best described mechanisms of cellular aging. Nutrient sensing, exemplified by caloric restriction, counteracts aging by activating various signaling pathways and transcription factors. *IGF,* Insulin-like growth factor; *ROS,* reactive oxygen species; *TOR,* target of rapamycin.

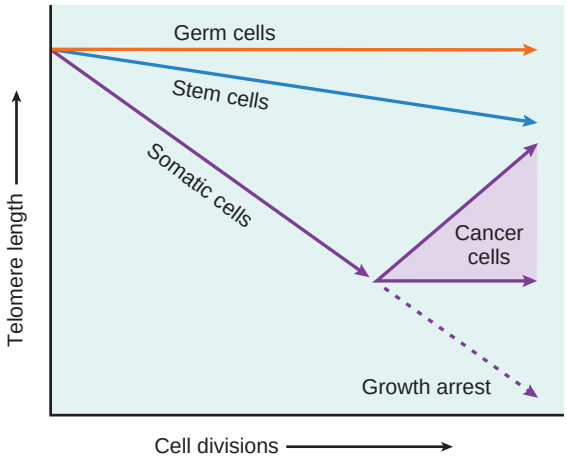


Figure 2.36 The role of telomeres and telomerase in replicative senescence of cells. Telomere length is plotted against the number of cell divisions. In most somatic cells there is no telomerase activity, and telomeres progressively shorten with increasing cell divisions until growth arrest or until senescence occurs. Germ cells and stem cells both contain telomerase, but only germ cells have sufficient levels of the enzyme to stabilize telomere length completely. In cancer cells, telomerase is often reactivated. (Data from Holt SE, Shay JW, Wright WE: Refining the telomere-telomerase hypothesis of aging and cancer, *Nat Biotechnol* 14:836, 1996.)

stem cells, but it is absent in most somatic tissues (Fig. 2.36) Therefore, as most somatic cells age, their telomeres become shorter and they exit the cell cycle, resulting in an inability to generate new cells to replace damaged ones. Conversely, in immortalized cancer cells, telomerase is usually reactivated and telomere length is stabilized, allowing the cells to proliferate indefinitely. This is discussed more fully in Chapter 7. The causal links between telomere length and cellular senescence have been established in mouse models. Genetically engineered mice with shortened telomeres exhibit reduced life spans that can be restored to normal by telomerase activation. As discussed in other chapters, telomere shortening is also associated with premature development of diseases, such as pulmonary fibrosis (Chapter 15) and aplastic anemia (Chapter 14).

- *Activation of tumor suppressor genes.* In addition to telomere attrition, activation of certain tumor suppressor genes, particularly those encoded by the *CDKN2A* locus, also seems to be involved in controlling replicative senescence. The *CDKN2A* locus encodes two tumor suppressor proteins, expression of one of which, known as p16 or INK4a, is correlated with chronologic age in virtually all human and mouse tissues examined. By controlling G1- to S-phase progression during the cell cycle (Chapter 1), p16 protects cells from uncontrolled mitogenic signals and pushes cells along the senescence pathway. This is discussed further in Chapter 7.

Defective Protein Homeostasis. Protein homeostasis involves two mechanisms: maintenance of proteins in their correctly folded conformations (mediated by chaperones) and degradation of misfolded, damaged, or unneeded proteins by the autophagy-lysosome system and ubiquitin-proteasome system. There is evidence that both normal folding and degradation of misfolded proteins are impaired with aging. Mutant mice deficient in chaperones of the heat shock protein family age rapidly, and conversely, those that overexpress such chaperones are long-lived. Similar data exist for the role of autophagy and proteasomal degradation of proteins. Of interest, administration of rapamycin, which inhibits the mTOR (molecular target of rapamycin) pathway, increases the life span of middle-aged mice. Rapamycin has multiple effects, including promotion of autophagy. Abnormal protein homeostasis can have many effects on cell survival, replication, and functions. In addition, it may lead to accumulation of misfolded proteins, which can trigger apoptosis.

Dysregulated Nutrient Sensing. Paradoxical though it may seem, eating less increases longevity. Caloric restriction increases life span in all eukaryotic species in which it has been tested, with encouraging results even in nonhuman primates and in a few unusually disciplined people who are the envy of others! Because of these observations, there has been much interest in deciphering the role of nutrient sensing in aging. Although incompletely understood, there are two major neurohormonal circuits that regulate metabolism.

- *Insulin and insulin-like growth factor 1 (IGF-1) signaling pathway.* IGF-1 is produced in many cell types in response to growth hormone secretion by the pituitary gland. IGF-1, as indicated by its name, mimics intracellular signaling by insulin and thereby informs cells of the availability of glucose, promoting an anabolic state as well as cell growth and replication. IGF-1 signaling has multiple downstream targets; relevant to this discussion are two kinases: AKT and its downstream target, mTOR, which, as the name implies, is inhibited by rapamycin.
- *Sirtuins.* Sirtuins are a family of NAD-dependent protein deacetylases. There are at least seven types of sirtuins in mammals that are distributed in different cellular compartments and have nonredundant functions designed to adapt bodily functions to various environmental stresses, including food deprivation and DNA damage. Sirtuins are thought to promote the expression of several genes whose products increase longevity. These include proteins that inhibit metabolic activity, reduce apoptosis, stimulate protein folding, and counteract the harmful effects of oxygen free radicals. Sirtuins also increase insulin sensitivity and glucose metabolism, and may be targets for the treatment of diabetes.

It is thought that caloric restriction increases longevity both by reducing the signaling intensity of the IGF-1 pathway and by increasing sirtuins. Attenuation of IGF-1 signaling leads to lower rates of cell growth and metabolism and possibly reduced cellular damage. This effect can be mimicked by rapamycin. An increase in sirtuins, particularly sirtuin-6, serves dual functions: the sirtuins (1) contribute to metabolic adaptations of caloric restriction and (2) promote genomic integrity by activating DNA repair enzymes through deacylation. Although the anti-aging effects of sirtuins have been widely publicized, much remains to be known before sirtuin-activating pills will be available to increase longevity. Nevertheless, optimistic wine lovers have been delighted to hear that a constituent of red wine may activate sirtuins and thus increase life span!

KEY CONCEPTS

CELLULAR AGING

- Cellular aging results from a combination of accumulating cellular damage (e.g., by free radicals), reduced capacity to divide (replicative senescence), reduced ability to repair damaged DNA, and defective protein homeostasis.
  - Accumulation of DNA damage: Defective DNA repair mechanisms; conversely, caloric restriction may activate DNA repair and slow aging in model organisms
  - Replicative senescence: Reduced capacity of cells to divide secondary to progressive shortening of chromosomal ends (telomeres)
  - Defective protein homeostasis: Resulting from impaired chaperone and proteasome functions.
  - Nutrient sensing system: Caloric restriction increases longevity. Mediators may be reduced IGF-1 signaling and increased sirtuins.

The various forms of cellular derangements and adaptations described in this chapter cover a wide spectrum, which includes adaptations in cell size, growth, and function; reversible and irreversible forms of acute cell injury; regulated cell death (e.g., in apoptosis); pathologic alterations in cell organelles; and less ominous forms of intracellular accumulations, including pigmentations. Reference is made to all of these alterations throughout this book, because all organ injury and ultimately all clinical disease arise from derangements in cell structure and function.

SUGGESTED READINGS

Mechanisms of Cell Injury

Galluzzi L, Kepp O, Kroemer G: Mitochondrial control of cell death: a phylogenetically conserved control, *Microb Cell* 3:101–108, 2016. *[The role of mitochondria in cellular response to stress].*

Hausenloy DJ, Yellon DM: Myocardial ischemia-reperfusion injury: a neglected therapeutic target, *J Clin Invest* 123:92–100, 2013. *[The molecular basis of reperfusion injury and possible therapeutic targets].*

Lambeth JD, Neish AS: Nox enzymes and new thinking on reactive oxygen: a double-edged sword revisited, *Annu Rev Pathol* 9:119–145, 2014. *[An interesting review that discusses the roles of reactive oxygen species in normal physiology and disease].*

Oakes SA, Papa FR: The role of endoplasmic reticulum stress in human pathology, *Annu Rev Pathol* 10:173–194, 2015. *[An up-to-date review of the unfolded protein response and the pathogenic importance of cell injury caused by misfolded proteins].*

Cell Death

Green DR: The coming decade of cell death research: five riddles, *Cell* 177:1094–1107, 2019. *[A forward-looking Perspectives article highlighting the major unanswered questions about different pathways of cell death].*

Hotchkiss RS, Strasser A, McDunn JE, et al: Cell death, *N Engl J Med* 361:1570–1583, 2009. *[An excellent review of the major pathways of cell death (necrosis, apoptosis, and autophagy-associated death) and their clinical implications and therapeutic targeting].*

Apoptosis

Schenk RL, Strasser A, Dewson G: BCL-2. Long and winding road from discovery to therapeutic target, *Biochem Biophys Res Commun* 482:459–469, 2017. *[An up-to-date review of the biochemistry and functions of the members of the BCL-2 family of apoptosis regulators].*

Nagata S: Apoptosis and clearance of apoptotic cells, *Annu Rev Immunol* 36:489–517, 2018. *[An excellent review of the mechanisms by which apoptotic cells and their fragments are cleared].*

Van Opdenbosch N, Lamkanfi M: Caspases in cell death, inflammation, and disease, *Immunity* 50:1352–1364, 2019. *[A summary of the functions of this family of enzymes involved in apoptosis and other pathologic processes].*

Necroptosis and Pyroptosis

Galluzzi L, Kepp O, Chan FK, et al: Necroptosis. Mechanisms and relevance to disease, *Annu Rev Pathol* 12:103–130, 2017. *[A current review of necroptosis and its pathophysiologic significance].*

Shi J, Gao W, Shao F: Pyroptosis: gasdermin-mediated programmed necrotic cell death, *Trends Biochem Sci* 42:245–254, 2017. *[A review of the mechanisms and consequences of pyroptosis].*

Tang D, Kang R, Berghe TV, et al: The molecular machinery of regulated cell death, *Cell Res* 29:347–364, 2019. *[A review of newly discovered pathways of cell death and possible therapeutic interventions].*

Tonnus W, Meyer C, Paliege A, et al: The pathologic features of regulated necrosis, *J Pathol* 247:697–707, 2019. *[A description of the pathology of various pathways of cell death].*

Weinlich R, Oberst A, Beere HM, et al: Necroptosis in development, inflammation, and disease, *Nat Rev Mol Cell Biol* 18:127–136, 2017. *[An excellent overview of the biochemistry and significance of necroptosis].*

Autophagy

Choi AMK, Ryter S, Levine B: Autophagy in human health and disease, *N Engl J Med* 368:651–662, 2013. *[An excellent discussion of the mechanisms and significance of autophagy].*

Doherty J, Baehrecke EH: Life, death and autophagy, *Nat Cell Biol* 20:1110–1117, 2018. *[A review of the link between autophagy and cell death].*

Levine B, Kroemer G: Biological functions of autophagy genes: a disease perspective, *Cell* 176:11–42, 2019. *[An excellent review of the cell biology and genetics of autophagy].*

Adaptations

Bonaldo P, Sandri M: Cellular and molecular mechanisms of muscle atrophy, *Dis Model Mech* 6:25–39, 2013. *[A review of the control of protein turnover and its role in atrophy of muscle].*

Giroux V, Rustgi AK: Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence, *Nat Rev Cancer* 17:594–604, 2017. *[An excellent review of the mechanisms of metaplasia, with a focus on the gastrointestinal tract and why it predisposes to cancer].*

Nakamura M, Sadoshima J: Mechanisms of physiological and pathological cardiac hypertrophy, *Nat Rev Cardiol* 15:387–407, 2018. *[A review of the mechanisms of hypertrophy, with an emphasis on the heart].*

Aging

Lopez-Otin C, Blasco MA, Partridge L, et al: The hallmarks of aging, *Cell* 153:1194–1217, 2013. *[A landmark review that suggests nine hallmarks of aging and directions for future research].*

Marques FC, Volovik Y, Cohen E: The roles of cellular and organismal aging in the development of late-onset maladies, *Annu Rev Pathol* 10:1–23, 2015. *[A review of the many ways by which cellular aging contributes to the development of chronic diseases].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 3

Inflammation and Repair

CHAPTER CONTENTS

OVERVIEW OF INFLAMMATION: DEFINITIONS AND GENERAL FEATURES

Inflammation is a response of vascularized tissues that delivers leukocytes and molecules of host defense from the circulation to the sites of infection and cell damage in order to eliminate the offending agents. Although in common medical and lay parlance, inflammation suggests a harmful reaction, it is actually a protective response that is essential for survival. It serves to rid the host of both the initial cause of cell injury (e.g., microbes, toxins) and the consequences of such injury (e.g., necrotic cells and tissues). The mediators of defense include phagocytic leukocytes, antibodies, and complement proteins. Most of these normally circulate in a resting state in the blood, from where they can be rapidly recruited to any site in the body. Some of the cells involved in inflammatory responses also reside in tissues, where they function as sentinels on the lookout for threats. The process of inflammation delivers circulating cells and proteins to tissues and activates the recruited and resident cells as well as the soluble molecules, which then function to get rid of the harmful or unwanted substances. Without inflammation, infections would go unchecked, wounds would never heal, and injured tissues might remain permanent festering sores. The suffix *-itis* after an organ denotes inflammation in that site, such as appendicitis, conjunctivitis, or meningitis.

The typical inflammatory reaction develops through a series of sequential steps (Fig. 3.1):

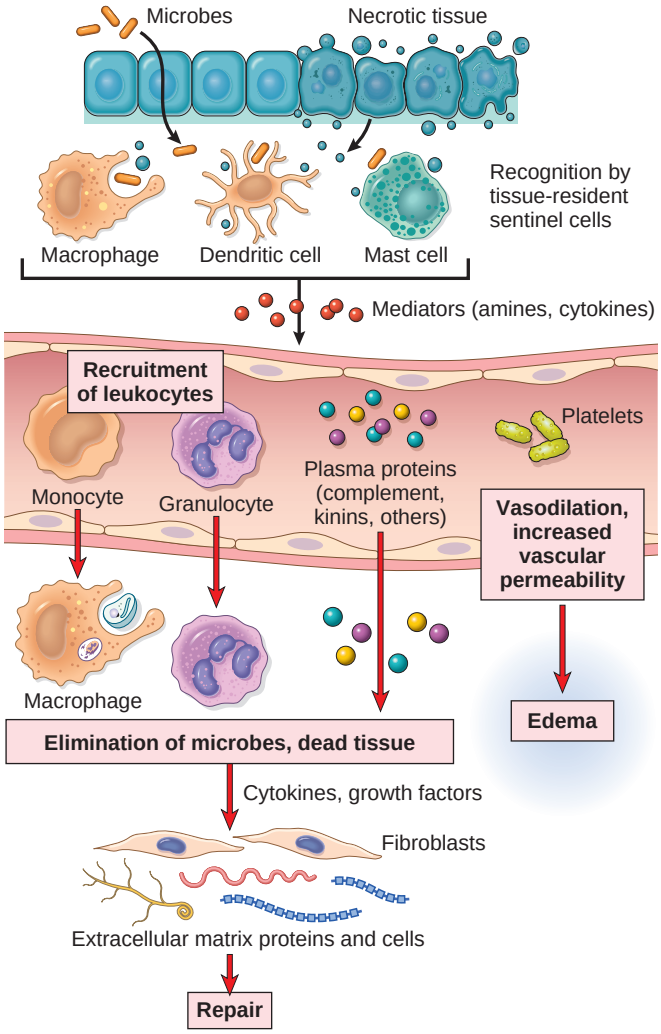


Figure 3.1 Sequence of events in an inflammatory reaction. Sentinel cells in tissues (macrophages, dendritic cells, and other cell types) recognize microbes and damaged cells and liberate mediators, which trigger the vascular and cellular reactions of inflammation.

- *Recognition* of the noxious agent that is the initiating stimulus for inflammation. The cells involved in inflammation (tissue-resident sentinel cells, phagocytes, and others) are equipped with receptors that recognize microbial products and substances released from damaged cells. These receptors are described in more detail later. Engagement of the receptors leads to the production of mediators of inflammation, which then trigger the subsequent steps in the inflammatory response.
- *Recruitment* of leukocytes and plasma proteins into the tissues. Since blood perfuses every tissue, leukocytes and proteins such as complement can be delivered to any site of microbial invasion or tissue injury. When pathogenic microbes invade the tissues, or tissue cells die, leukocytes (first mainly neutrophils, later monocytes and lymphocytes) and plasma proteins are rapidly recruited from the circulation to the extravascular site where the offending agent is located. The exodus of cells and plasma proteins from blood requires coordinated changes in blood vessels and secretion of mediators, described in detail later.
- *Removal* of the stimulus for inflammation is accomplished mainly by phagocytic cells, which ingest and destroy microbes and dead cells.
- *Regulation* of the response is important for terminating the reaction when it has accomplished its purpose.
- *Repair* consists of a series of events that heal damaged tissue. In this process the injured tissue is replaced through regeneration of surviving cells and filling of residual defects with connective tissue (scarring).

Before discussing the mechanisms, functions, and pathology of the inflammatory response, it is useful to review some of its fundamental properties.

- *Components of the inflammatory response.* The major participants in the inflammatory reaction in tissues are blood vessels and leukocytes (see Fig. 3.1). As will be discussed in more detail later, blood vessels respond to inflammatory stimuli by dilating and by increasing their permeability, enabling selected circulating proteins to enter the site of infection or tissue damage. In addition, the endothelium lining blood vessels also changes, such that circulating leukocytes adhere and then migrate into the tissues. Leukocytes, once recruited, are activated and acquire the ability to ingest and destroy microbes and dead cells, as well as foreign bodies and other unwanted materials in the tissues.
- *Harmful consequences of inflammation.* Protective inflammatory reactions to infections are often accompanied by local tissue damage and its associated signs and symptoms (e.g., pain and functional impairment). Typically, however, these harmful consequences are self-limited and resolve as the inflammation abates, leaving little or no permanent damage. In contrast, there are many diseases in which the inflammatory reaction is misdirected (e.g., against self tissues in autoimmune diseases), occurs against normally harmless environmental substances (e.g., in allergies), or is inadequately controlled. In these cases the normally protective inflammatory reaction becomes the cause of the disease, and the damage it causes is the dominant feature. In clinical medicine, great attention is given to the injurious consequences of inflammation (Table 3.1). Inflammatory reactions underlie common chronic diseases such as rheumatoid arthritis, atherosclerosis, and lung fibrosis, as well as life-threatening hypersensitivity reactions to insect bites, foods, drugs, and toxins. For this reason our pharmacies abound with antiinflammatory drugs, which ideally would control the harmful sequelae of inflammation yet not interfere with its beneficial effects. Inflammation also may contribute to a variety of diseases that are thought to be primarily metabolic, degenerative, or genetic, such as type 2 diabetes, Alzheimer disease, and cancer. In recognition of the wide-ranging harmful consequences of inflammation, the lay press has rather melodramatically referred to it as "the silent killer."
- *Local and systemic inflammation.* Much of this discussion focuses on the inflammatory response to a localized infection or tissue damage. Although even local reactions may have systemic manifestations (e.g., fever in the setting

Table 3.1 Diseases Caused by Inflammatory Reactions

Disorders	Cells and Molecules Involved in Injury
Acute	
Acute respiratory distress syndrome	Neutrophils
Asthma	Eosinophils; IgE antibodies
Glomerulonephritis	Antibodies and complement; neutrophils, monocytes
Septic shock	Cytokines
Chronic	
Arthritis	Lymphocytes, macrophages; antibodies?
Asthma	Eosinophils; IgE antibodies
Atherosclerosis	Macrophages; lymphocytes
Pulmonary fibrosis	Macrophages; fibroblasts

IgE, Immunoglobulin E.

Listed are selected examples of diseases in which the inflammatory response plays a significant role in tissue injury. Some, such as asthma, can present with acute inflammation or a chronic illness with repeated bouts of acute exacerbation. These diseases and their pathogenesis are discussed in relevant chapters.

of bacterial or viral pharyngitis), the inflammation is largely confined to the site of infection or damage. In rare situations, such as some disseminated bacterial infections, the inflammatory reaction is systemic and causes widespread pathologic abnormalities. This reaction has been called sepsis, which is one form of the *systemic inflammatory response syndrome*. This serious disorder is discussed in Chapter 4.

- *Mediators of inflammation.* The vascular and cellular reactions of inflammation are triggered by soluble factors that are produced by various cells or derived from plasma proteins and are generated or activated in response to the inflammatory stimulus. Microbes, necrotic cells (whatever the cause of cell death), and even hypoxia can trigger the elaboration of inflammatory mediators and thus elicit inflammation. Such mediators initiate and amplify the inflammatory response and determine its pattern, severity, and clinical and pathologic manifestations.
- *Acute and chronic inflammation.* The distinction between acute and chronic inflammation was originally based on the duration of the reaction, but we now know that they differ in several ways (Table 3.2). Acute inflammation is a rapid, often self-limited, response to offending agents that are readily eliminated, such as many bacteria and fungi, and dead cells. It typically develops within minutes or hours and is of short duration (several hours to a few days). It is characterized by the exudation of fluid and plasma proteins (edema) and the emigration of leukocytes, predominantly neutrophils. If the offending stimulus is eliminated, the reaction subsides, and residual injury is repaired.

  Chronic inflammation may follow acute inflammation or arise de novo. It is a response to agents that are difficult to eradicate, such as some bacteria (e.g., tubercle bacilli) and other pathogens (such as viruses and fungi), as well as self antigens and environmental antigens. Chronic inflammation is of longer duration and is associated with more tissue destruction and scarring (fibrosis). Sometimes, chronic inflammation may coexist with unresolved acute inflammation, as may occur in peptic ulcers.

Table 3.2 Features of Acute and Chronic Inflammation

Feature	Acute	Chronic
Onset	Fast: minutes or hours	Slow: days
Cellular infiltrate	Mainly neutrophils	Monocytes/macrophages and lymphocytes
Tissue injury, fibrosis	Usually mild and self-limited	Often severe and progressive
Local and systemic signs	Prominent	Less

Historical Highlights

Although clinical features of inflammation were described in an Egyptian papyrus dated around 3000 BC, Celsus, a Roman writer of the first century AD, first listed the four cardinal signs of inflammation: *rubor* (redness), *tumor* (swelling), *calor* (heat), and *dolor* (pain). These signs are hallmarks of acute inflammation. A fifth clinical sign, loss of function *(functio laesa),* was added by Rudolf Virchow in the 19th century. In 1793 the Scottish surgeon John Hunter noted what is now considered an obvious fact: inflammation is not a disease but a stereotypic response that has a salutary effect on its host. In the 1880s Russian biologist Elie Metchnikoff discovered the process of *phagocytosis* by observing the ingestion of rose thorns by amebocytes of starfish larvae and of bacteria by mammalian leukocytes. He concluded that the purpose of inflammation was to bring phagocytic cells to the injured area to engulf invading bacteria. This concept was satirized by George Bernard Shaw in his play *The Doctor's Dilemma,* in which one physician's cure-all is to "stimulate the phagocytes!" Sir Thomas Lewis, studying the inflammatory response in skin, established the concept that chemical substances, such as histamine (produced locally in response to injury), mediate the vascular changes of inflammation. This fundamental concept underlies the important discoveries of chemical mediators of inflammation and the use of antiinflammatory drugs in clinical medicine.

Causes of Inflammation

Inflammatory reactions may be triggered by a variety of stimuli:

- *Infections* (bacterial, viral, fungal, parasitic) and microbial toxins are among the most common and medically important causes of inflammation. Different infectious pathogens elicit varied inflammatory responses, from mild acute inflammation that causes little or no lasting damage and successfully eradicates the infection, to severe systemic reactions that can be fatal, to prolonged chronic reactions that cause extensive tissue injury. The outcomes are determined largely by the type of pathogen and the host response and, to some extent, by other, poorly defined characteristics of the host.

- *Tissue necrosis* elicits inflammation regardless of the cause of cell death. Cells may die because of ischemia (reduced blood flow, the cause of myocardial infarction), trauma, and physical and chemical injury (e.g., thermal injury, as in burns or frostbite; irradiation; exposure to some environmental chemicals). Several molecules released from necrotic cells are known to trigger inflammation; some of these are described later.
- *Foreign bodies* (splinters, dirt, sutures) may elicit inflammation by themselves or because they cause traumatic tissue injury or carry microbes. Even endogenous substances can be harmful if they deposit in tissues; such substances include urate crystals (in gout), cholesterol crystals (in atherosclerosis), and lipids (in obesity-associated metabolic syndrome).
- *Immune reactions* (also called *hypersensitivity*) are reactions in which the normally protective immune system damages the individual's own tissues. The injurious immune responses may be inappropriately directed against self antigens, causing autoimmune diseases, or may be reactions against environmental substances, as in allergies, or against microbes. Inflammation is a major cause of tissue injury in these diseases (Chapter 6). Because the stimuli for the inflammatory responses (e.g., self antigens and environmental antigens) cannot be eliminated, autoimmune and allergic reactions tend to be persistent and difficult to cure, are often associated with chronic inflammation, and are important causes of morbidity and mortality. The inflammation is induced largely by cytokines produced by T lymphocytes and other cells of the immune system (Chapter 6).

Recognition of Microbes and Damaged Cells

Recognition of microbial components or substances released from damaged cells is the initiating step in inflammatory reactions. The cells and receptors that perform this function evolved to protect multicellular organisms from microbes in the environment, and the responses they trigger are critical for the survival of the organisms. Several cellular receptors and circulating proteins are capable of recognizing microbes and products of cell damage and triggering inflammation.

- *Cellular receptors for microbes.* Cells express receptors in the plasma membrane (for extracellular microbes), the endosomes (for ingested microbes), and the cytosol (for intracellular microbes) that enable the cells to sense the presence of foreign invaders in any cellular compartment. The best defined of these receptors belong to the family of *Toll-like receptors (TLRs);* these and other cellular receptors of innate immunity are described in Chapter 6. The receptors are expressed on many cell types including epithelial cells (through which microbes enter from the external environment), dendritic cells, macrophages, and other leukocytes (which may encounter microbes in various tissues). Engagement of these receptors triggers production of molecules involved in inflammation including adhesion molecules on endothelial cells, cytokines, and other mediators.
- *Sensors of cell damage.* All cells have cytosolic receptors, such as *NOD-like receptors (NLRs),* that recognize diverse molecules that are liberated or altered as a consequence of cell damage. These molecules include uric acid (a product of DNA breakdown), adenosine triphosphate (ATP) (released from damaged mitochondria), reduced intracellular K^+ concentrations (reflecting loss of ions because of plasma membrane injury), even DNA when it is released into the cytoplasm and not sequestered in nuclei, as it should be normally, and many others. These receptors activate a multiprotein cytosolic complex called the *inflammasome* (Chapter 6), which induces the production of the cytokine interleukin-1 (IL-1). IL-1 recruits leukocytes and thus induces inflammation (see later). Gain-of-function mutations in the genes encoding some of the receptors are the cause of rare diseases grouped under autoinflammatory syndromes that are characterized by spontaneous IL-1 production and inflammation; IL-1 antagonists are effective treatments for these disorders. The inflammasome has also been implicated in inflammatory reactions to urate crystals (the cause of gout), lipids (in metabolic syndrome and obesity-associated type 2 diabetes), cholesterol crystals (in atherosclerosis), and even amyloid deposits in the brain (in Alzheimer disease). These disorders are discussed later in this and other chapters.
- *Other cellular receptors involved in inflammation.* In addition to directly recognizing microbes, many leukocytes express receptors for the Fc tails of antibodies and for complement proteins. These receptors recognize microbes coated with antibodies and complement (the coating process is called opsonization) and promote ingestion and destruction of the microbes as well as inflammation.
- *Circulating proteins.* The *complement system* reacts against microbes and produces mediators of inflammation (discussed later). A circulating protein called mannose-binding lectin recognizes microbial sugars and promotes ingestion of the microbes and the activation of the complement system. Other proteins called collectins also bind to and combat microbes.

KEY CONCEPTS

GENERAL FEATURES AND CAUSES OF INFLAMMATION

- Inflammation is a beneficial host response to foreign invaders and necrotic tissue, but it may also cause tissue damage.
- The main components of inflammation are a vascular reaction and a cellular response, both activated by mediators that are derived from plasma proteins and various cells.
- The steps of the inflammatory response can be remembered as the five R's: (1) recognition of the injurious agent, (2) recruitment of leukocytes, (3) removal of the agent, (4) regulation (control) of the response, and (5) repair (resolution).
- Acute and chronic inflammation differ in the kinetics of the reaction, the principal cells involved, and the degree of injury. The outcome of acute inflammation is either elimination of the noxious stimulus followed by decline of the reaction and repair of the damaged tissue or persistent injury resulting in chronic inflammation.
- Causes of inflammation include infections, tissue necrosis, foreign bodies, trauma, and immune responses.
- Epithelial cells, tissue macrophages and dendritic cells, leukocytes, and other cell types express receptors that sense the presence of microbes and substances released from damaged cells. Circulating proteins recognize microbes that have entered the blood.

ACUTE INFLAMMATION

Acute inflammation has three major components: (1) dilation of small vessels leading to an increase in blood flow; (2) increased permeability of the microvasculature enabling plasma proteins and leukocytes to leave the circulation; and (3) emigration of leukocytes from the microcirculation, their accumulation in the focus of injury, and their activation to eliminate the offending agent (see Fig. 3.1). When an individual encounters an injurious agent, such as a microbe or dead cells, phagocytes that reside in tissues try to eliminate these agents. At the same time, phagocytes and other sentinel cells in the tissues recognize the presence of the foreign or abnormal substance and react by liberating cytokines, lipid messengers, and other mediators of inflammation. Some of these mediators act on small blood vessels in the vicinity and promote the efflux of plasma and the recruitment of circulating leukocytes to the site where the offending agent is located.

Reactions of Blood Vessels in Acute Inflammation

The vascular reactions of acute inflammation consist of changes in the flow of blood and the permeability of vessels, both designed to maximize the movement of plasma proteins and leukocytes out of the circulation and into the site of infection or injury. The escape of fluid, proteins, and blood cells from the vascular system into the interstitial tissue or body cavities is known as exudation (Fig. 3.2). An *exudate* is an extravascular fluid that has a high protein concentration and contains cellular debris. Its presence implies the existence of an inflammatory process that has increased the permeability of small blood vessels. In contrast, a *transudate* is a fluid with low protein content (most of which is albumin), little or no cellular material, and low specific gravity. It is essentially an ultrafiltrate of blood plasma that is produced as a result of osmotic or hydrostatic imbalance across the vessel wall without an increase in vascular permeability (Chapter 4). Edema denotes an excess of fluid in the interstitial tissue or serous cavities; it can be either an exudate or a transudate. Pus, a purulent exudate, is an inflammatory exudate rich in leukocytes (mostly neutrophils), the debris of dead cells, and, in many cases, microbes.

Changes in Vascular Flow and Caliber

Changes in vascular flow and caliber begin early after injury and consist of the following.

- *Vasodilation* is induced by the action of several mediators, notably histamine, on vascular smooth muscle. It is one of the earliest manifestations of acute inflammation. Vasodilation first involves the arterioles and then leads to opening of new capillary beds in the area. The result is increased blood flow, which is the cause of heat and redness (erythema) at the site of inflammation.
- Vasodilation is quickly followed by *increased permeability of the microvasculature,* with the outpouring of protein-rich fluid into the extravascular tissues; this process is described in detail later.
- The loss of fluid and increased vessel diameter lead to slower blood flow, concentration of red cells in small vessels, and increased viscosity of the blood. These changes result in engorgement of small vessels with slowly moving red cells, a condition termed *stasis,* which is seen

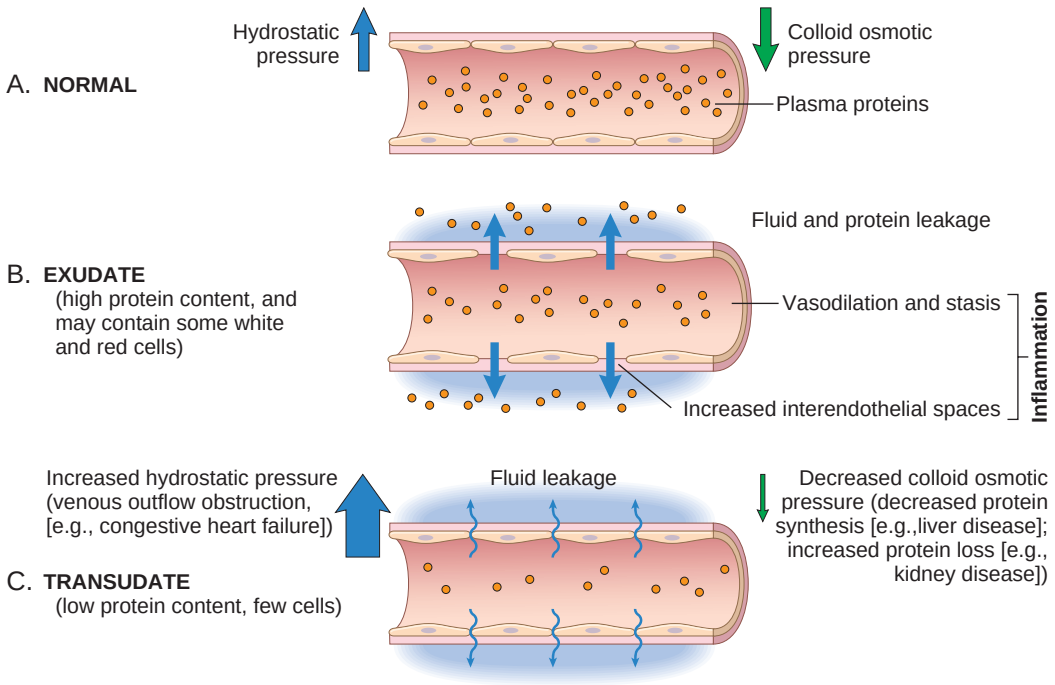


Figure 3.2 Formation of exudates and transudates. (A) Normal hydrostatic pressure *(blue arrow)* is about 32 mm Hg at the arterial end of a capillary bed and 12 mm Hg at the venous end; the mean colloid osmotic pressure of tissues is approximately 25 mm Hg *(green arrow)*, which is equal to the mean capillary pressure. Therefore the net flow of fluid across the vascular bed is almost nil. (B) An exudate is formed in inflammation because vascular permeability increases as a result of increased interendothelial spaces. (C) A transudate is formed when fluid leaks out because of increased hydrostatic pressure or decreased osmotic pressure.

as vascular congestion and localized redness of the involved tissue.

- As stasis develops, blood leukocytes, principally neutrophils, accumulate along the vascular endothelium. At the same time, endothelial cells are activated by mediators produced at sites of infection and tissue damage and express increased levels of adhesion molecules. Leukocytes then adhere to the endothelium, and soon afterward they migrate through the vascular wall into the interstitial tissue in a sequence that is described later.

Increased Vascular Permeability (Vascular Leakage)

Several mechanisms are responsible for the increased permeability of postcapillary venules, a hallmark of acute inflammation (Fig. 3.3).

- *Contraction of endothelial cells* resulting in opening of interendothelial gaps is the most common mechanism of vascular leakage. It is elicited by histamine, bradykinin, leukotrienes, and other chemical mediators. It is called the immediate transient response because it occurs rapidly after exposure to the mediator and is usually short-lived (15 to 30 minutes). In some forms of mild injury (e.g., after burns, irradiation or ultraviolet radiation, and exposure to certain bacterial toxins), vascular leakage begins after a delay of 2 to 12 hours and lasts for several hours or even days; this delayed prolonged leakage may be caused by contraction of endothelial cells or mild endothelial damage. Sunburn is a classic example of damage that results in late-appearing vascular leakage. Often the immediate and delayed responses occur along a continuum.

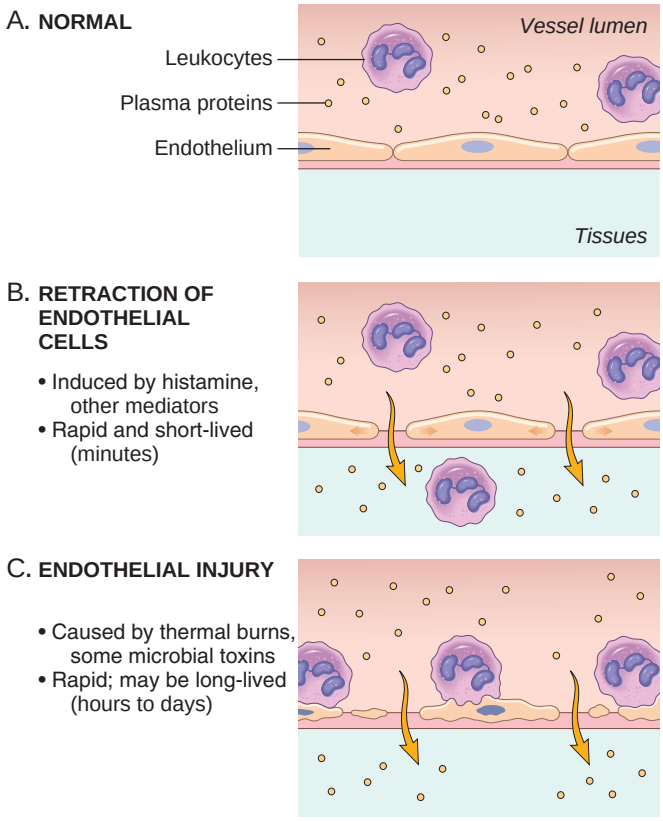


Figure 3.3 Principal mechanisms of increased vascular permeability in inflammation and their features and underlying causes.

- *Endothelial injury* resulting in endothelial cell necrosis and detachment. Direct damage to the endothelium is encountered in severe physical injuries, for example, in thermal burns, or is induced by the actions of microbes and microbial toxins that damage endothelial cells. Neutrophils that adhere to the endothelium during inflammation may also injure endothelial cells and thus amplify the reaction. In most instances leakage starts immediately after injury and is sustained for several hours until the damaged vessels are thrombosed or repaired.

Although these mechanisms of increased vascular permeability are described separately, all probably contribute in varying degrees in responses to most stimuli. For example, at different stages of a thermal burn, leakage results from chemically mediated endothelial contraction and direct and leukocyte-dependent endothelial injury. The vascular leakage induced by these mechanisms can cause life-threatening loss of fluid in severely burned patients.

Responses of Lymphatic Vessels and Lymph Nodes

In addition to blood vessels, lymphatic vessels also participate in acute inflammation. The system of lymphatics and lymph nodes filters and polices the extravascular fluids. Lymphatics drain the small amount of extravascular fluid that seeps out of capillaries in the healthy state. In inflammation, lymph flow is increased and helps drain edema fluid that accumulates because of increased vascular permeability. In addition to fluid, leukocytes and cell debris, as well as microbes, may find their way into lymph. Lymphatic vessels, like blood vessels, proliferate during inflammatory reactions to handle the increased load. The lymphatics may become secondarily inflamed (lymphangitis), as may the draining lymph nodes (lymphadenitis). Inflamed lymph nodes are often enlarged because of hyperplasia of the lymphoid follicles and increased numbers of lymphocytes and macrophages. This constellation of pathologic changes is termed reactive, or inflammatory, lymphadenitis (Chapter 13). The presence of red streaks near a skin wound is a telltale sign of bacterial infection. The streaks represent inflamed lymphatic channels and are diagnostic of lymphangitis; it may be accompanied by painful enlargement of the draining lymph nodes, indicating lymphadenitis.

KEY CONCEPTS

VASCULAR REACTIONS IN ACUTE INFLAMMATION

- Vasodilation is induced by chemical mediators such as histamine (described later) and is the cause of erythema and increased blood flow.
- Increased vascular permeability is induced by histamine, kinins, and other mediators that produce gaps between endothelial cells and by direct or leukocyte-induced endothelial injury.
- Increased vascular permeability allows plasma proteins and leukocytes, the mediators of host defense, to enter sites of infection or tissue damage. Fluid leak from blood vessels results in edema.
- Lymphatic vessels and lymph nodes are also involved in inflammation and often show redness and swelling.

Leukocyte Recruitment to Sites of Inflammation

The changes in blood flow and vascular permeability are quickly followed by an influx of leukocytes into the tissue. These leukocytes perform the key function of eliminating the offending agents. The most important leukocytes in typical inflammatory reactions are the ones capable of phagocytosis, namely neutrophils and macrophages. They ingest and destroy bacteria and other microbes, as well as necrotic tissue and foreign substances. Macrophages also produce growth factors that aid in repair. A price that is paid for the defensive potency of leukocytes is that, when activated, they may induce tissue damage and prolong the inflammatory reaction because the leukocyte products that destroy microbes and help "clean up" necrotic tissues can also injure normal bystander host tissues.

The journey of leukocytes from the vessel lumen to the tissue is a multistep process that is mediated and controlled by adhesion molecules and cytokines called chemokines. This process can be divided into sequential phases (Fig. 3.4).

1. In the lumen: *margination, rolling, and adhesion to endothelium*. Vascular endothelium in its normal state does not bind circulating cells or allow their passage. In inflammation the endothelium is activated and can bind leukocytes as a prelude to their exit from blood vessels.
2. *Migration across the endothelium* and vessel wall.
3. *Migration in the tissues* toward a chemotactic stimulus.

Leukocyte Adhesion to Endothelium

In normally flowing blood in venules, red cells are confined to a central axial column, displacing the leukocytes toward the wall of the vessel. Because of dilation of inflamed postcapillary venules, blood flow slows (stasis), and more white cells assume a peripheral position along the endothelial surface. This process of leukocyte redistribution is called margination. The slowed leukocytes sense signals from the endothelium, resulting first in the cells rolling on the vessel

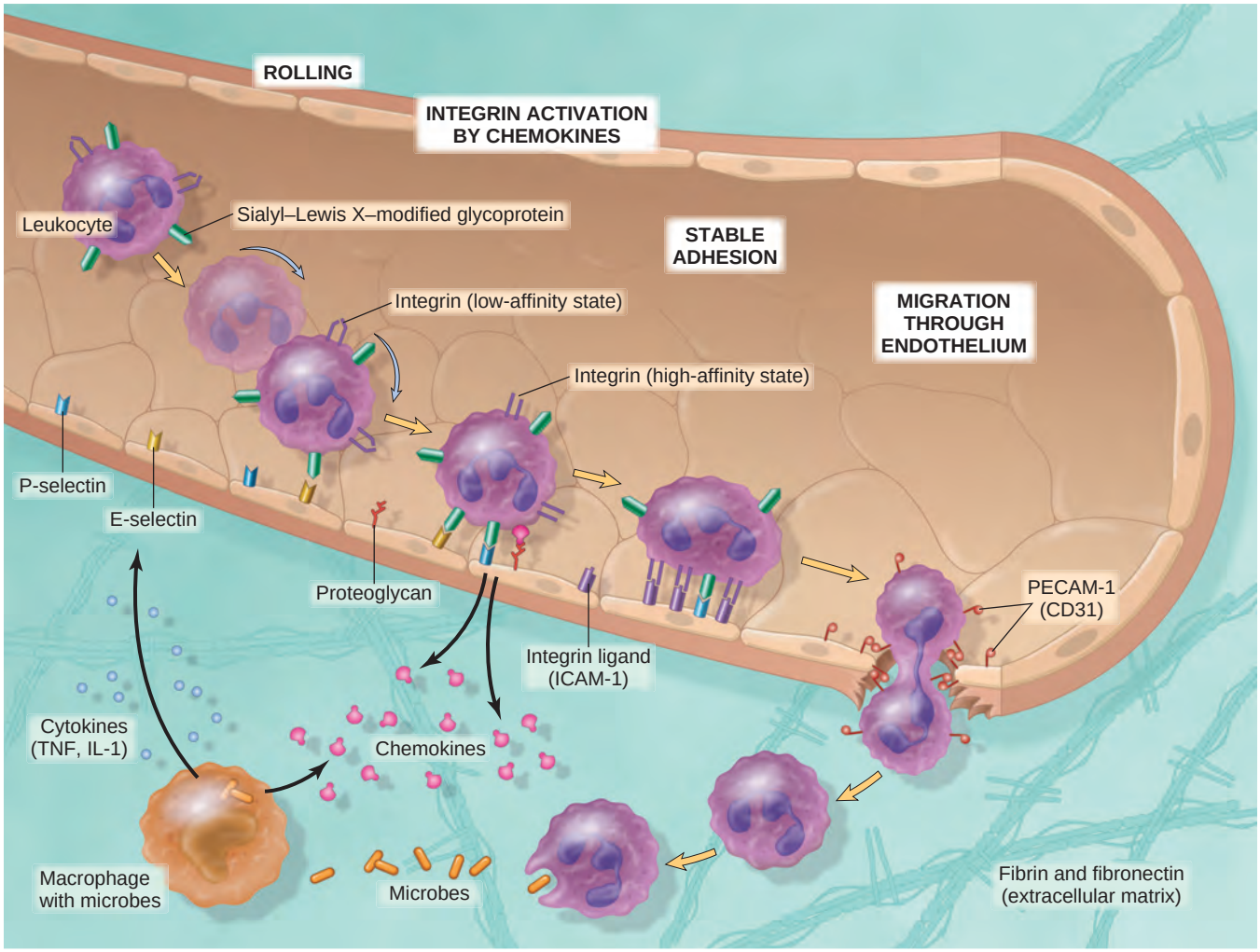


Figure 3.4 The multistep process of leukocyte migration through blood vessels, shown here for neutrophils. The leukocytes first roll, then become activated and adhere to endothelium, then transmigrate across the endothelium, pierce the basement membrane, and migrate toward chemoattractants emanating from the source of injury. Different molecules play predominant roles in different steps of this process: selectins, in rolling; chemokines (usually displayed bound to proteoglycans), in activating the neutrophils to increase avidity of integrins; integrins, in firm adhesion; and CD31 (PECAM-1), in transmigration. *ICAM-1*, Intercellular adhesion molecule 1; *IL-1*, interleukin-1; *PECAM-1*, platelet endothelial cell adhesion molecule (also known as CD31); *TNF*, tumor necrosis factor.

wall and then recognizing adhesion molecules expressed on the endothelium that lead to the cells adhering firmly (resembling pebbles over which a stream runs without disturbing them).

The attachment of leukocytes to endothelial cells is mediated by adhesion molecules whose expression is enhanced by cytokines, which are secreted by sentinel cells in tissues in response to microbes and other injurious agents, thus ensuring that leukocytes are recruited to the tissues where these stimuli are present. The two major families of proteins involved in leukocyte adhesion and migration are the selectins and integrins and their ligands (Table 3.3). They are expressed on leukocytes and endothelial cells.

- *Selectins.* The initial rolling interactions are mediated by selectins, of which there are three types: one expressed on leukocytes (L-selectin), one on endothelium (E-selectin), and one in platelets and on endothelium (P-selectin) (see Table 3.3). The ligands for selectins are sialylated oligosaccharides bound to mucin-like glycoproteins. The expression of selectins and their ligands is regulated by cytokines produced in response to infection and injury. Tissue macrophages, mast cells, and endothelial cells that encounter microbes and dead tissues respond by secreting several cytokines including tumor necrosis factor (TNF), IL-1, and chemokines (*chemo*attractant cyto*kines*). (Cytokines are described in more detail later and in Chapter 6.) TNF and IL-1 act on the endothelial cells of postcapillary venules adjacent to the infection and induce the coordinate expression of numerous adhesion molecules. Within 1 to 2 hours the endothelial cells begin to express E-selectin and the ligands for L-selectin. Other mediators such as histamine and thrombin, described later, stimulate the redistribution of P-selectin from its normal intracellular stores in endothelial cell granules (called *Weibel-Palade bodies*) to the cell surface. Leukocytes express L-selectin at the tips of their microvilli and also express ligands for E-selectin and P-selectin, all of which bind to the complementary molecules on the endothelial cells. These are low-affinity interactions with a fast off-rate, so they are easily disrupted by the flowing blood. As a result, the bound leukocytes bind, detach, and bind again and thus begin to roll along the endothelial surface.
- *Integrins.* The weak rolling interactions slow down the leukocytes and give them the opportunity to bind more firmly to the endothelium. Firm adhesion is mediated by a family of heterodimeric leukocyte surface proteins called integrins (see Table 3.3). TNF and IL-1 induce endothelial expression of ligands for integrins, mainly vascular cell adhesion molecule 1 (VCAM-1), the ligand for the β1 integrin VLA-4, and intercellular adhesion molecule-1 (ICAM-1), the ligand for the β2 integrins LFA-1 and MAC-1. Leukocytes normally express integrins in a low-affinity state. Chemokines that were produced at the site of injury bind to endothelial cell proteoglycans and are displayed at high concentrations on the endothelial surface. These chemokines bind to and activate the rolling leukocytes. One of the consequences of activation is the conversion of VLA-4 and LFA-1 integrins on the leukocytes to a high-affinity state. The combination of cytokine-induced expression of integrin ligands on the endothelium and increased integrin affinity on the leukocytes results in firm integrin-mediated adhesion of the leukocytes to the endothelium at the site of inflammation. The leukocytes stop rolling, their cytoskeleton is reorganized, and they spread out on the endothelial surface.

Leukocyte Migration Through Endothelium

The next step in the process of leukocyte recruitment is migration of the leukocytes through intact endothelium, called *transmigration* or *diapedesis*. Transmigration of leukocytes occurs mainly in postcapillary venules. Chemokines act on the adherent leukocytes and stimulate the cells to migrate through interendothelial gaps toward the chemical

Table 3.3 Endothelial and Leukocyte Adhesion Molecules

Family	Molecule	Distribution	Ligand
Selectin	L-selectin (CD62L)	Neutrophils, monocytes T cells (naïve and central memory) B cells (naïve)	Sialyl-Lewis X/PNAd on GlyCAM-1, CD34, MAdCAM-1, others; expressed on endothelium (HEV)
	E-selectin (CD62E)	Endothelium activated by cytokines (TNF, IL-1)	Sialyl-Lewis X (e.g., CLA) on glycoproteins; expressed on neutrophils, monocytes, T cells (effector, memory)
	P-selectin (CD62P)	Endothelium activated by cytokines (TNF, IL-1), histamine, or thrombin	Sialyl-Lewis X on PSGL-1 and other glycoproteins; expressed on neutrophils, monocytes, T cells (effector, memory)
Integrin	LFA-1 (CD11aCD18)	Neutrophils, monocytes, T cells (naïve, effector, memory)	ICAM-1 (CD54), ICAM-2 (CD102); expressed on endothelium (upregulated on activated endothelium)
	MAC-1 (CD11bCD18)	Monocytes, DCs	ICAM-1 (CD54), ICAM-2 (CD102); expressed on endothelium (upregulated on activated endothelium)
	VLA-4 (CD49aCD29)	Monocytes T cells (naïve, effector, memory)	VCAM-1 (CD106); expressed on endothelium (upregulated on activated endothelium)
	α4β7 (CD49dCD29)	Monocytes T cells (gut homing naïve effector, memory)	VCAM-1 (CD106), MAdCAM-1; expressed on endothelium in gut and gut-associated lymphoid tissues
Ig	CD31	Endothelial cells, leukocytes	CD31 (homotypic interaction)

CLA, Cutaneous lymphocyte antigen-1; *GlyCAM-1,* glycan-bearing cell adhesion molecule-1; *HEV,* high endothelial venule; *Ig,* immunoglobulin; *IL-1,* interleukin-1; *ICAM,* intercellular adhesion molecule; *MAdCAM-1,* mucosal adhesion cell adhesion molecule-1; *PSGL-1,* P-selectin glycoprotein ligand-1; *TNF,* tumor necrosis factor; *VCAM,* vascular cell adhesion molecule.

concentration gradient, that is, toward the site of injury or infection where the chemokines are being produced. Several adhesion molecules present in the intercellular junctions between endothelial cells are involved in the migration of leukocytes. These molecules include a member of the immunoglobulin superfamily called *CD31* or *PECAM-1* (platelet endothelial cell adhesion molecule). After traversing the endothelium, leukocytes pierce the basement membrane, probably by secreting collagenases, and enter the extravascular space. After leukocytes pass through, the basement membranes become continuous again. The cells that have exited the vessel then migrate toward the chemotactic gradient created by chemokines and other chemoattractants and accumulate in the extravascular site.

The most telling proof of the importance of leukocyte adhesion molecules in the host inflammatory response are genetic deficiencies in these molecules, which result in increased susceptibility to bacterial infections. These leukocyte adhesion deficiencies are described in Chapter 6.

Chemotaxis of Leukocytes

After exiting the circulation, leukocytes move in the tissues toward the site of injury by a process called *chemotaxis*, which is defined as locomotion along a chemical gradient. Both exogenous and endogenous substances act as chemoattractants. The most common exogenous factors are bacterial products, including peptides with *N*-formylmethionine terminal amino acids and some lipids. Endogenous chemoattractants include several chemical mediators (described later): (1) cytokines, particularly those of the chemokine family (e.g., IL-8); (2) components of the complement system, particularly C5a; and (3) arachidonic acid (AA) metabolites, mainly leukotriene B_4 (LTB_4). All these chemotactic agents bind to specific seven-transmembrane G protein–coupled receptors on the surface of leukocytes. Signals initiated from these receptors result in activation of second messengers that induce polymerization of actin at the leading edge of the cell and localization of myosin filaments at the back. This reorganization of the cytoskeleton allows the leading edge of the leukocyte to extend filopodia that pull the back of the cell in the direction of extension, much as an automobile with front-wheel drive is pulled by the wheels in front (Fig. 3.5). The net result is that leukocytes migrate in the direction of locally produced chemoattractants emanating from the site of the inflammatory stimulus.

The nature of the leukocyte infiltrate varies with the age of the inflammatory response and the type of stimulus. In most forms of acute inflammation, neutrophils predominate in the inflammatory infiltrate during the first 6 to 24 hours and are replaced by monocytes in 24 to 48 hours (Fig. 3.6). There are several reasons for the early preponderance of neutrophils: they are more numerous than are other leukocytes, respond more rapidly to chemokines, and may attach more firmly to the adhesion molecules that are rapidly induced on endothelial cells such as P-selectin and E-selectin. After entering tissues, neutrophils are short-lived; most neutrophils in extravascular tissues undergo apoptosis within a few days. Monocytes not only survive longer but may also proliferate in the tissues, and thus they become the dominant population in prolonged inflammatory reactions. There are, however, exceptions to this stereotypic pattern of cellular infiltration. In certain infections—for example, those produced by *Pseudomonas* bacteria—the cellular infiltrate is dominated by continuously recruited neutrophils for several days; in viral infections, lymphocytes may be the first cells to arrive; some hypersensitivity reactions are dominated by activated lymphocytes, macrophages, and plasma cells (reflecting the immune response); and in helminthic infections and allergic reactions, eosinophils may be the main cell type.

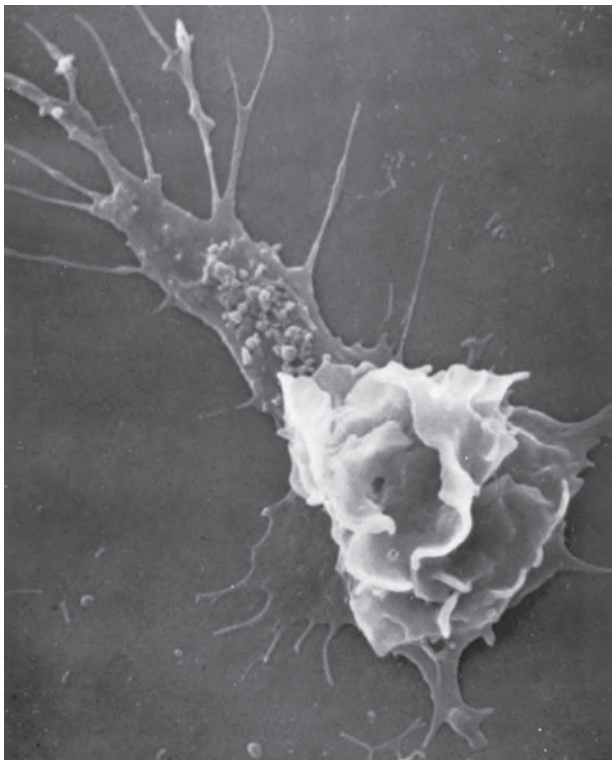


Figure 3.5 Scanning electron micrograph of a moving leukocyte in culture showing a filopodium *(upper left)* and a trailing tail. (Courtesy Dr. Morris J. Karnovsky, Harvard Medical School, Boston, Mass.)

The molecular understanding of leukocyte recruitment and migration has led to development of a large number of drugs for controlling harmful inflammation, including agents that block TNF (discussed later), and antagonists of leukocyte integrins that are approved for inflammatory diseases or are being tested in clinical trials. Predictably, these antagonists not only have the desired effect of controlling the inflammation but can also compromise the ability of treated patients to defend themselves against microbes, which, of course, is the physiologic function of the inflammatory response.

KEY CONCEPTS

LEUKOCYTE RECRUITMENT TO SITES OF INFLAMMATION

- Leukocytes are recruited from the blood into the extravascular tissue, where infectious pathogens or damaged tissues may be located, migrate to the site of infection or tissue injury, and are activated to perform their functions.

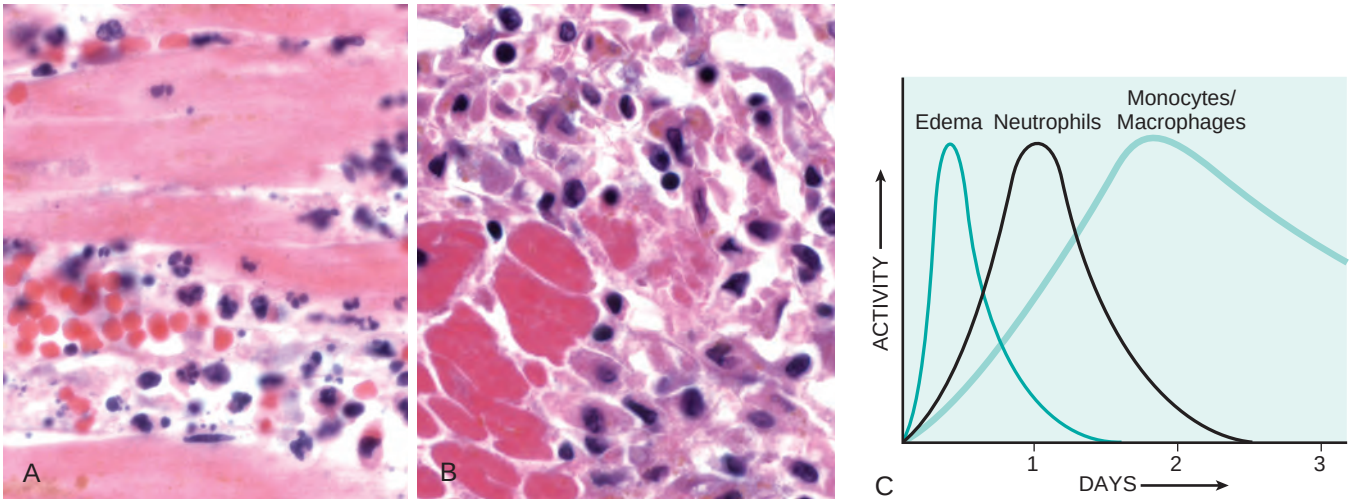


Figure 3.6 Nature of leukocyte infiltrates in inflammatory reactions. The photomicrographs show an inflammatory reaction in the myocardium after ischemic necrosis (infarction). (A) Early (neutrophilic) infiltrates and congested blood vessels. (B) Later (mononuclear) cellular infiltrates. (C) The approximate kinetics of edema and cellular infiltration. For simplicity, edema is shown as an acute transient response, although secondary waves of delayed edema and neutrophil infiltration can also occur.

- Leukocyte recruitment is a multistep process consisting of loose attachment to and rolling on endothelium (mediated by selectins), firm attachment to endothelium (mediated by integrins), and migration through interendothelial spaces.
- Various cytokines promote expression of selectins and integrin ligands on endothelium (e.g., TNF, IL-1), increase the avidity of integrins for their ligands (e.g., chemokines), and promote directional migration of leukocytes (also chemokines); many of these cytokines are produced by tissue macrophages and other cells responding to the pathogens or damaged tissues.
- Neutrophils predominate in the early inflammatory infiltrate and are later replaced by monocytes and macrophages.

Once leukocytes (particularly neutrophils and monocytes) are recruited to a site of infection or cell death, they must be activated to perform their functions. The responses of these leukocytes consist of recognition of the offending agents by TLRs and other receptors, described earlier, which deliver signals that activate the leukocytes to phagocytose and destroy the offending agents.

Phagocytosis and Clearance of the Offending Agent

The two major phagocytes are neutrophils and macrophages. Although these cell types share many functional properties, they also differ in significant ways (Table 3.4).

Recognition of microbes or dead cells induces several responses in leukocytes that are collectively called *leukocyte activation* (Fig. 3.7). Activation results from signaling pathways that are triggered in leukocytes, resulting in increases in cytosolic Ca^{2+} and activation of enzymes such as protein kinase C and phospholipase A_2. The functional responses that are most important for destruction of microbes and other offenders are phagocytosis and intracellular killing. Several other responses aid in the defensive functions of inflammation and may contribute to its injurious consequences.

Phagocytosis

Phagocytosis involves sequential steps (Fig. 3.8):

- Recognition and attachment of the particle to be ingested by the leukocyte;
- Engulfment, with subsequent formation of a phagocytic vacuole; and
- Killing of the microbe and degradation of the ingested material.

Phagocytic Receptors. Mannose receptors, scavenger receptors, and receptors for various opsonins enable phagocytes to bind and ingest microbes. The macrophage mannose receptor is a lectin that binds terminal mannose and fucose residues of glycoproteins and glycolipids. These sugars are typically part of molecules found on microbial cell walls, whereas mammalian glycoproteins and glycolipids contain terminal sialic acid or *N*-acetylgalactosamine. Therefore the mannose receptor recognizes microbes and not host cells. Scavenger receptors were originally defined as molecules that bind and mediate endocytosis of oxidized or acetylated low-density lipoprotein (LDL) particles that do not interact with the conventional LDL receptor. Macrophage scavenger receptors bind a variety of microbes in addition to modified LDL particles. Macrophage integrins, notably MAC-1 (CD11b/CD18), may also bind microbes for phagocytosis. The efficiency of phagocytosis is greatly enhanced when microbes are coated with opsonins for which the phagocytes express high-affinity receptors. The major opsonins are immunoglobulin G (IgG) antibodies, the C3b breakdown product of complement, and certain plasma lectins, notably mannose-binding lectin and collectins, all of which are recognized by specific receptors on leukocytes.

Engulfment. After a particle is bound to phagocyte receptors, extensions of the cytoplasm flow around it, and the plasma membrane pinches off to form an intracellular vesicle (phagosome) that encloses the particle. The phagosome then fuses with a lysosomal granule, which discharges its contents

Table 3.4 Properties of Neutrophils and Macrophages

	Neutrophils	Macrophages
Origin	HSCs in bone marrow	HSCs in bone marrow (in inflammatory reactions) Many tissue-resident macrophages: stem cells in yolk sac or fetal liver (early in development)
Lifespan in tissues	Several days	Inflammatory macrophages: days or weeks Tissue-resident macrophages: years
Responses to activating stimuli	Rapid, short-lived, mostly degranulation and enzymatic activity	More prolonged, slower, often dependent on new gene transcription
Reactive oxygen species	Rapidly induced by assembly of phagocyte oxidase (respiratory burst)	Less prominent
Nitric oxide	Low levels or none	Induced following transcriptional activation of iNOS
Degranulation	Major response; induced by cytoskeletal rearrangement	Not prominent
Cytokine production	Low levels or none	Major functional activity; requires transcriptional activation of cytokine genes
NET formation	Rapidly induced, by extrusion of nuclear contents	No
Secretion of lysosomal enzymes	Prominent	Less

HSC, Hematopoietic stem cells; *iNOS,* inducible nitric oxide synthase; *NET,* neutrophil extracellular trap.

This table lists the major differences between neutrophils and macrophages. The reactions summarized above are described in the text. Note that the two cell types share many features such as phagocytosis, ability to migrate through blood vessels into tissues, and chemotaxis.

into the phagolysosome (see Fig. 3.8). During this process the phagocyte may also release lysosome contents into the extracellular space.

The process of phagocytosis is complex and involves the integration of many receptor-initiated signals that lead to membrane remodeling and cytoskeletal changes. Phagocytosis is dependent on polymerization of actin filaments; it is therefore not surprising that the signals that trigger phagocytosis are many of the same that are involved in chemotaxis.

Intracellular Destruction of Microbes and Debris

Killing of microbes is accomplished by reactive oxygen species (ROS), also called reactive oxygen intermediates, and reactive nitrogen species, mainly derived from nitric

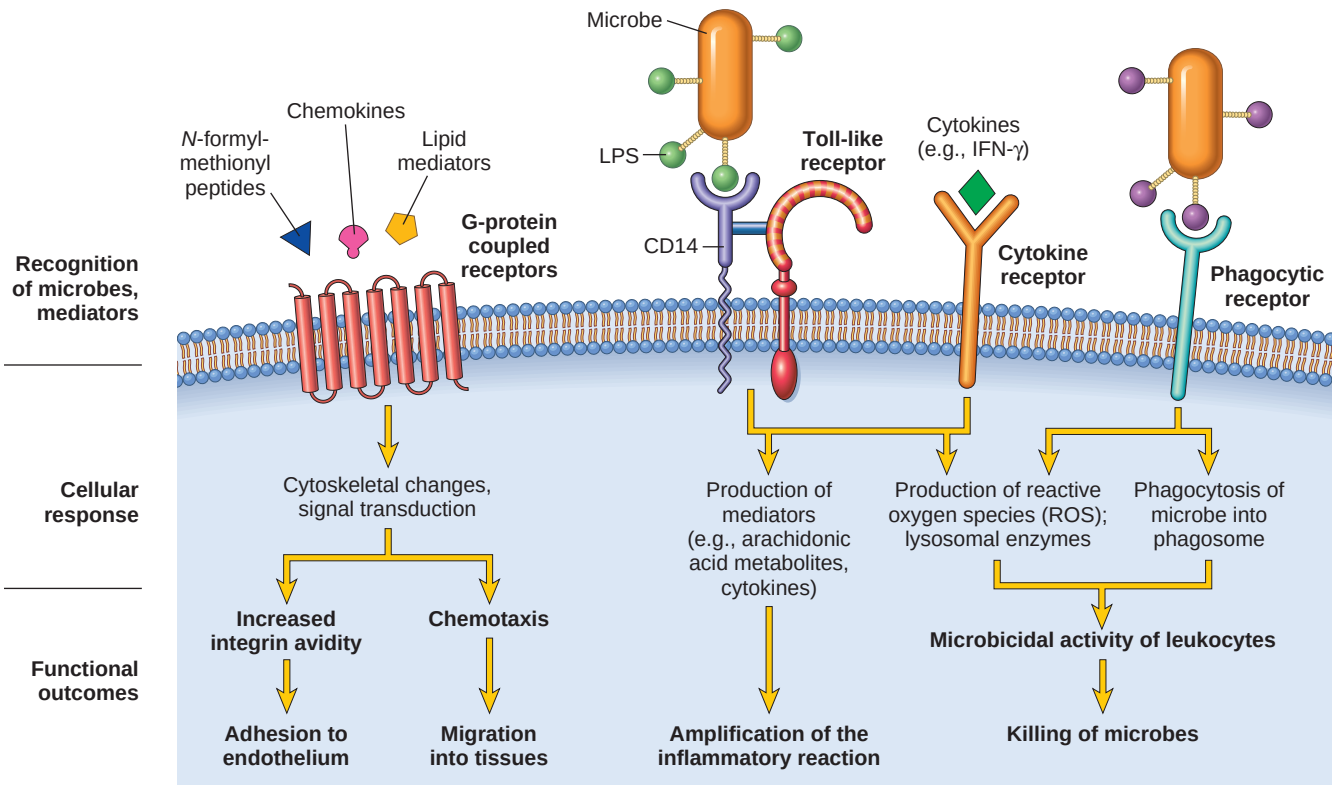


Figure 3.7 Leukocyte activation. Different classes of cell surface receptors of leukocytes recognize different stimuli. The receptors initiate responses that mediate the functions of the leukocytes. Only some receptors are depicted (see text for details). Lipopolysaccharide (*LPS*) first binds to a circulating LPS-binding protein (not shown). *IFN-γ,* Interferon-γ.

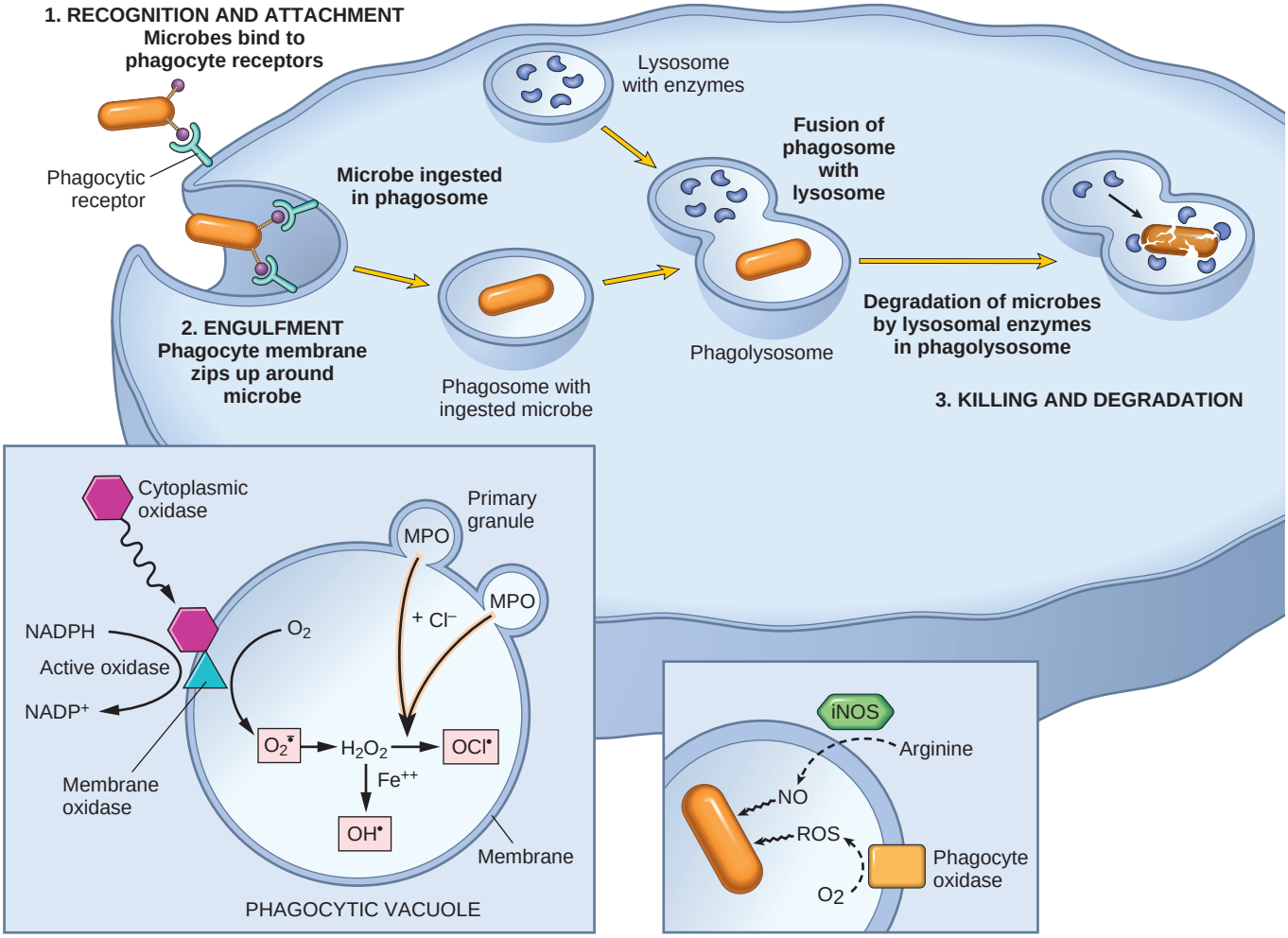


Figure 3.8 Phagocytosis and intracellular destruction of microbes. Phagocytosis of a particle (e.g., a bacterium) involves binding to receptors on the leukocyte membrane, engulfment, and fusion of the phagocytic vacuoles with lysosomes. This is followed by destruction of ingested particles within the phagolysosomes by lysosomal enzymes and by reactive oxygen and nitrogen species. Hypochlorite ($HOCl^{\cdot}$) and hydroxyl radical ($^{\cdot}OH$) are microbicidal products generated from superoxide ($O_2^{\cdot-}$), and peroxynitrite ($OONO^{\cdot}$) is generated from nitric oxide *(NO)*. During phagocytosis, granule contents may be released into extracellular tissues (not shown). *iNOS*, Inducible nitric oxide synthase; *MPO*, myeloperoxidase; *ROS*, reactive oxygen species.

oxide (NO), and these as well as lysosomal enzymes destroy phagocytosed materials (see Fig. 3.8). This is the final step in the elimination of infectious agents and necrotic cells. The killing and degradation of microbes and dead cell debris within neutrophils and macrophages occur most efficiently after activation of the phagocytes. All these killing mechanisms are normally sequestered in lysosomes, to which phagocytosed materials are brought. Thus, potentially harmful substances are segregated from the cell's cytoplasm and nucleus to avoid damage to the phagocyte while it is performing its normal function.

Reactive Oxygen Species. ROS are produced by the rapid assembly and activation of a multicomponent oxidase, NADPH oxidase (also called phagocyte oxidase), which oxidizes reduced nicotinamide-adenine dinucleotide phosphate (NADPH) and, in the process, reduces oxygen to superoxide anion ($O_2^{\cdot-}$). In neutrophils, this oxidative reaction is triggered by activating signals accompanying phagocytosis and is called the *respiratory burst*. Phagocyte oxidase is an enzyme complex consisting of at least seven proteins. In resting neutrophils, different components of the enzyme are located in the plasma membrane and the cytoplasm. In response to activating stimuli, the cytosolic protein components translocate to the phagosomal membrane, where they assemble and form the functional enzyme complex. Thus, the ROS are produced within the phagolysosome, where they can act on ingested particles without damaging the host cell. $O_2^{\cdot-}$ is converted into hydrogen peroxide (H_2O_2), mostly by spontaneous dismutation. H_2O_2 is not able to efficiently kill microbes by itself. However, the azurophilic granules of neutrophils contain the enzyme myeloperoxidase (MPO), which, in the presence of a halide such as Cl^-, converts H_2O_2 to hypochlorite ($HOCl^{\cdot}$), the active ingredient in household bleach. The latter is a potent antimicrobial agent that destroys microbes by halogenation (in which the halide is bound covalently to cellular constituents) or by oxidation of proteins and lipids (lipid peroxidation). The H_2O_2-MPO-halide system is the most potent bactericidal system of neutrophils. Nevertheless, inherited deficiency of MPO by

itself leads to minimal increase in susceptibility to infection, emphasizing the redundancy of microbicidal mechanisms in leukocytes. H_2O_2 is also converted to hydroxyl radical ($^{\cdot}OH$), another powerful destructive agent. As discussed in Chapter 2, these oxygen-derived free radicals bind to and modify cellular lipids, proteins, and nucleic acids and thus destroy cells such as microbes.

Oxygen-derived radicals may be released extracellularly from leukocytes after exposure to microbes, chemokines, and antigen-antibody complexes or following a phagocytic challenge. These ROS are implicated in tissue damage accompanying inflammation.

Plasma, tissue fluids, and host cells possess antioxidant mechanisms that protect healthy cells from these potentially harmful oxygen-derived radicals. These antioxidants are discussed in Chapter 2 and include (1) the enzyme superoxide dismutase, which is found in, or can be activated in, a variety of cell types; (2) the enzyme catalase, which detoxifies H_2O_2; (3) glutathione peroxidase, another powerful H_2O_2 detoxifier; (4) the copper-containing plasma protein ceruloplasmin; and (5) the iron-free fraction of plasma transferrin.

Inherited deficiencies of components of phagocyte oxidase cause an immunodeficiency disease called chronic granulomatous disease (CGD), which is discussed in Chapter 6.

Nitric Oxide. NO, a soluble gas produced from arginine by the action of nitric oxide synthase (NOS), also participates in microbial killing. There are three different types of NOS: endothelial (eNOS), neuronal (nNOS), and inducible (iNOS). eNOS and nNOS are constitutively expressed at low levels, and the NO they generate functions to maintain vascular tone and as a neurotransmitter, respectively. iNOS, the type that is involved in microbial killing, is induced when macrophages (and, to a lesser extent, neutrophils) are activated by cytokines (e.g., interferon-γ [IFN-γ]) or microbial products. In macrophages, NO reacts with superoxide ($O_2^{\cdot-}$) to generate the highly reactive free radical peroxynitrite ($ONOO^-$). These nitrogen-derived free radicals, similar to ROS, attack and damage the lipids, proteins, and nucleic acids of microbes (Chapter 2). Reactive oxygen and nitrogen species have overlapping actions, as shown by the observation that knockout mice lacking either phagocyte oxidase or iNOS are only mildly susceptible to infections, but mice lacking both succumb rapidly to disseminated infections by normally harmless commensal bacteria.

In addition to its role as a microbicidal substance, NO relaxes vascular smooth muscle and promotes vasodilation. It is not clear if this action of NO plays an important role in the vascular reactions of acute inflammation.

Lysosomal Enzymes and Other Lysosomal Proteins. Neutrophils and macrophages contain lysosomal granules that contribute to microbial killing and, when released, may cause tissue damage. Neutrophils have two main types of granules. The smaller specific (or secondary) granules contain lysozyme, collagenase, gelatinase, lactoferrin, plasminogen activator, histaminase, and alkaline phosphatase. The larger azurophil (or primary) granules contain MPO, bactericidal proteins (lysozyme, defensins), acid hydrolases, and a variety of neutral proteases (elastase, cathepsin G, nonspecific collagenases, proteinase 3). Both types of granules can fuse with phagocytic vacuoles containing engulfed material, or the granule contents can be released into the extracellular space during "frustrated phagocytosis" (discussed later).

Different granule enzymes serve different functions. Acid proteases degrade bacteria and debris within the phagolysosomes, which are acidified by membrane-bound proton pumps. Neutral proteases are capable of degrading various extracellular components such as collagen, basement membrane, fibrin, elastin, and cartilage, resulting in the tissue destruction that accompanies inflammatory processes. Neutral proteases can also cleave C3 and C5 complement proteins and release a kinin-like peptide from kininogen. The released components of complement and kinins act as mediators of acute inflammation (discussed later). Neutrophil elastase has been shown to degrade virulence factors of bacteria and thus combat bacterial infections. Macrophages also contain acid hydrolases, collagenase, elastase, phospholipase, and plasminogen activator.

Because of the destructive effects of lysosomal enzymes, the initial leukocytic infiltration, if unchecked, can potentiate further inflammation by damaging tissues. These harmful proteases, however, are normally controlled by a system of antiproteases in the serum and tissue fluids. Foremost among these is α_1-antitrypsin, which is the major inhibitor of neutrophil elastase. A deficiency of these inhibitors may lead to sustained action of leukocyte proteases, as is the case in patients with α_1-antitrypsin deficiency, who are at risk for emphysema due to destruction of elastic support fibers in the lung because of uncontrolled elastase activity (Chapter 15). α_2-Macroglobulin is another antiprotease found in serum and various secretions.

Other microbicidal granule contents include defensins, cationic arginine-rich granule peptides that are toxic to microbes; cathelicidins, antimicrobial proteins found in neutrophils and other cells; lysozyme, which hydrolyzes the muramic acid-*N*-acetylglucosamine bond found in the glycopeptide coat of all bacteria; lactoferrin, an iron-binding protein present in specific granules; and major basic protein, a cationic protein of eosinophils, which has limited bactericidal activity but is cytotoxic to many helminthic parasites.

Neutrophil Extracellular Traps

Neutrophil extracellular traps (NETs) are extracellular fibrillar networks that concentrate antimicrobial substances at sites of infection and trap microbes, helping to prevent their spread. They are produced by neutrophils in response to infectious pathogens (mainly bacteria and fungi) and inflammatory mediators (e.g., chemokines, cytokines [mainly interferons], complement proteins, and ROS). The extracellular traps consist of a viscous meshwork of nuclear chromatin that binds and concentrates granule proteins such as antimicrobial peptides and enzymes (Fig. 3.9). NET formation starts with ROS-dependent activation of an arginine deaminase that converts arginines to citrulline, leading to chromatin decondensation. Other enzymes that are produced in activated neutrophils, such as MPO and elastase, enter the nucleus and cause further chromatin decondensation, culminating in rupture of the nuclear envelope and release of chromatin. In this process, the nuclei of the neutrophils are lost, leading to death of the cells. NETs have also been detected in the blood during sepsis. The nuclear chromatin

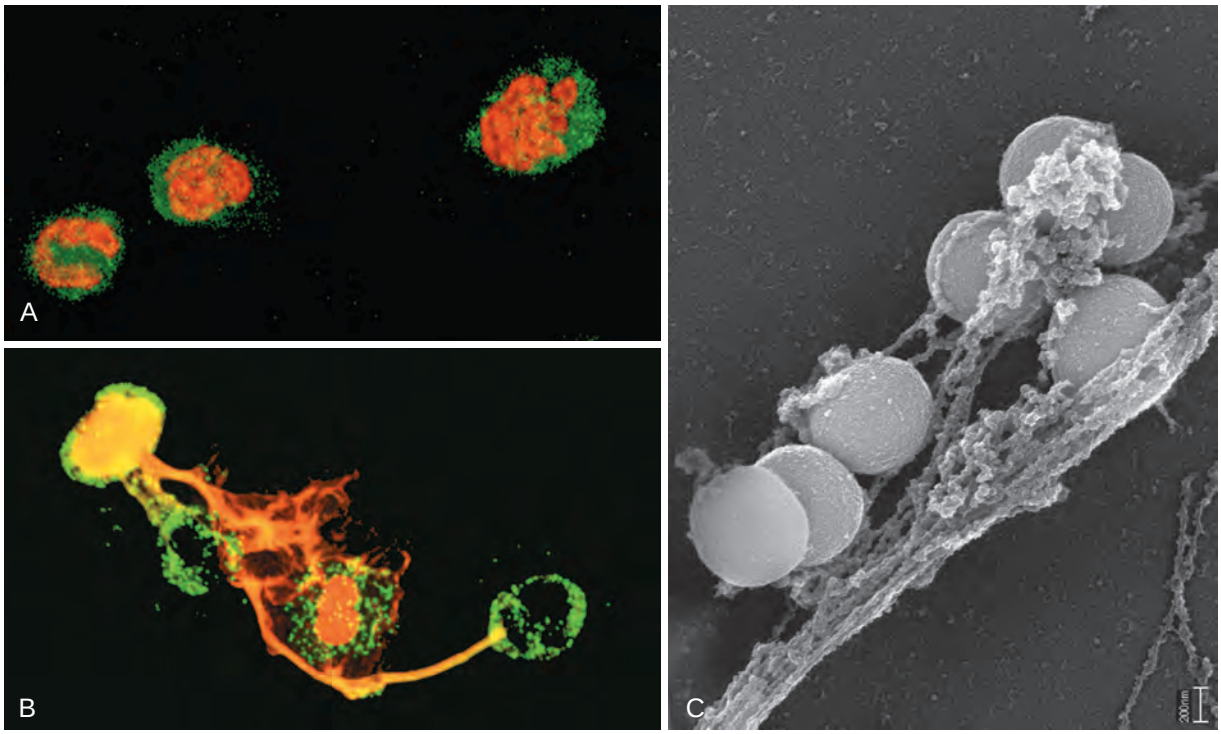


Figure 3.9 Neutrophil extracellular traps (NETs). (A) Healthy neutrophils with nuclei stained red and cytoplasm stained green. (B) Release of nuclear material from neutrophils (note that two have lost their nuclei), forming extracellular traps. (C) Electron micrograph of bacteria (staphylococci) trapped in NETs. (From Brinkmann V, Zychlinsky A: Beneficial suicide: why neutrophils die to make NETs, *Nat Rev Microbiol* 5:577, 2007, with permission.)

in the NETs, which includes histones and associated DNA, has been postulated to be a source of nuclear antigens in systemic autoimmune diseases, particularly lupus, in which individuals react against their own DNA and nucleoproteins (Chapter 6).

Leukocyte-Mediated Tissue Injury

Leukocytes are important causes of injury to normal cells and tissues under several circumstances.

- As part of a normal defense reaction against infectious microbes, when adjacent tissues suffer collateral damage. In some infections that are difficult to eradicate, such as tuberculosis and certain viral diseases, the prolonged host response contributes more to the pathology than does the microbe itself.
- When the inflammatory response is inappropriately directed against host tissues, as in certain autoimmune diseases.
- When the host reacts excessively against usually harmless environmental substances, as in allergic diseases, including asthma.

In all these situations, **the mechanisms by which leukocytes damage normal tissues are the same as the mechanisms involved in antimicrobial defense** because once the leukocytes are activated, their effector mechanisms do not distinguish between offender and host. During activation and phagocytosis, neutrophils and macrophages produce microbicidal substances (ROS, NO, and lysosomal enzymes) within the phagolysosome; under some circumstances, these substances are also released into the extracellular space. These released substances are capable of damaging host cells such as vascular endothelium and may thus amplify the effects of the initial injurious agent. If unchecked or inappropriately directed against host tissues, the leukocyte infiltrate itself becomes the offender, and indeed leukocyte-dependent inflammation and tissue injury underlie many acute and chronic human diseases (see Table 3.1). This fact becomes evident in the discussion of specific disorders throughout this book.

The contents of lysosomal granules are secreted by leukocytes into the extracellular milieu by several mechanisms. Controlled secretion of granule contents is a normal response of activated leukocytes. If phagocytes encounter materials that cannot be easily ingested, such as immune complexes deposited on large surfaces (e.g., glomerular basement membrane), the inability of the leukocytes to surround and ingest these substances (frustrated phagocytosis) triggers strong activation and the release of lysosomal enzymes into the extracellular environment. Some phagocytosed substances, such as urate crystals, may damage the membrane of the phagolysosome, also leading to the release of lysosomal granule contents.

Other Functional Responses of Activated Leukocytes

In addition to eliminating microbes and dead cells, activated leukocytes play several other roles in host defense. Importantly, these cells, especially macrophages, produce cytokines that can either amplify or limit inflammatory reactions, growth factors that stimulate the proliferation of endothelial cells and fibroblasts and the synthesis of collagen, and enzymes that remodel connective tissues. Because of these

activities, macrophages are also critical cells of chronic inflammation and tissue repair after inflammation has subsided. These functions of macrophages are discussed later in the chapter.

In this discussion of acute inflammation, we emphasize the importance of neutrophils and macrophages. However, it has recently become clear that some T lymphocytes, which are cells of adaptive immunity, also contribute to acute inflammation. The most important of these cells are those that produce the cytokine IL-17 (so-called Th17 cells), discussed in more detail in Chapter 6. IL-17 induces the secretion of chemokines that recruit other leukocytes. In the absence of effective Th17 responses, individuals are susceptible to fungal and bacterial infections and tend to develop "cold abscesses," particularly in the skin, that lack the classic features of acute inflammation, such as warmth and redness.

Termination of the Acute Inflammatory Response

Such a powerful system of host defense, with its inherent capacity to cause tissue injury, needs tight controls to minimize damage. In part, inflammation declines after the offending agents are removed simply because the mediators of inflammation are produced for only as long as the stimulus persists, have short half-lives, and are degraded after their release. Neutrophils also have short half-lives in tissues and die by apoptosis within several hours after leaving the blood. In addition, as inflammation develops, the process itself triggers a variety of stop signals that actively terminate the reaction. These active termination mechanisms include a switch in the type of arachidonic acid metabolite produced, from proinflammatory leukotrienes to antiinflammatory lipoxins (described later), and the liberation of antiinflammatory cytokines, including transforming growth factor-β (TGF-β) and IL-10, from macrophages and other cells. Other control mechanisms that have been demonstrated experimentally include neural impulses (cholinergic discharge) that inhibit the production of TNF in macrophages.

KEY CONCEPTS

LEUKOCYTE ACTIVATION AND REMOVAL OF OFFENDING AGENTS

- Leukocytes can eliminate microbes and dead cells by phagocytosis, followed by their destruction in phagolysosomes.
- Destruction is caused by free radicals (ROS, NO) generated in activated leukocytes and lysosomal enzymes.
- Neutrophils can extrude their nuclear contents to form extracellular nets that trap and destroy microbes.
- Enzymes and ROS may be released into the extracellular environment.
- The mechanisms that function to eliminate microbes and dead cells (the physiologic role of inflammation) are also capable of damaging normal tissues (the pathologic consequences of inflammation).
- Antiinflammatory mediators terminate the acute inflammatory reaction when it is no longer needed.

Mediators of Inflammation

Inflammatory mediators are the substances that initiate and regulate inflammatory reactions. Many mediators have been identified and targeted therapeutically to limit inflammation. In this discussion, we review their shared properties and the general principles that govern their production and actions.

- The most important mediators of acute inflammation are vasoactive amines, lipid products (prostaglandins and leukotrienes), cytokines (including chemokines), and products of complement activation (Table 3.5). These mediators induce various components of the inflammatory response, typically by distinct mechanisms, which is why inhibiting each has been therapeutically beneficial. However, there is also some overlap (redundancy) in the actions of the mediators.
- Mediators are either secreted by cells or generated from plasma proteins. *Cell-derived mediators* are normally sequestered in intracellular granules and can be rapidly secreted by granule exocytosis (e.g., histamine in mast cell granules) or are synthesized de novo (e.g., prostaglandins and leukotrienes, cytokines) in response to a stimulus. The major cell types that produce mediators of acute inflammation are the sentinels that detect invaders and tissue damage, that is, macrophages, dendritic cells, and mast cells, but platelets, neutrophils, endothelial cells, and most epithelia can also be induced to elaborate some

Table 3.5 Principal Mediators of Inflammation

Mediator	Source	Action
Histamine	Mast cells, basophils, platelets	Vasodilation, increased vascular permeability, endothelial activation
Prostaglandins	Mast cells, leukocytes	Vasodilation, pain, fever
Leukotrienes	Mast cells, leukocytes	Increased vascular permeability, chemotaxis, leukocyte adhesion and activation
Cytokines (TNF, IL-1, IL-6)	Macrophages, endothelial cells, mast cells	Local: endothelial activation (expression of adhesion molecules) Systemic: fever, metabolic abnormalities, hypotension (shock)
Chemokines	Leukocytes, activated macrophages	Chemotaxis, leukocyte activation
Platelet-activating factor	Leukocytes, mast cells	Vasodilation, increased vascular permeability, leukocyte adhesion, chemotaxis, degranulation, oxidative burst
Complement	Plasma (produced in liver)	Leukocyte chemotaxis and activation, direct target killing (membrane attack complex), vasodilation (mast cell stimulation)
Kinins	Plasma (produced in liver)	Increased vascular permeability, smooth muscle contraction, vasodilation, pain

IL, Interleukin; *TNF,* tumor necrosis factor.

of the mediators. *Plasma-derived mediators* (e.g., complement proteins) are produced mainly in the liver and are present in the circulation as inactive precursors that must be activated, usually by a series of proteolytic cleavages, to acquire their functional properties.

- Active mediators are produced only in response to offending stimuli. These stimuli include microbial products and substances released from necrotic cells. Some of the stimuli trigger well-defined receptors and signaling pathways, described earlier, but we still do not know how other stimuli induce the secretion of mediators (e.g., from mast cells in response to cell injury or mechanical irritation). The usual requirement for microbes or dead tissues as the initiating stimulus ensures that inflammation is normally triggered only when and where it is needed.
- Most mediators are short-lived. They quickly decay, or are inactivated by enzymes, or they are otherwise scavenged or inhibited. There is thus a system of checks and balances that regulates mediator actions. These built-in control mechanisms are discussed with each class of mediator.
- One mediator can stimulate the release of other mediators. For instance, products of complement activation stimulate the release of histamine, and the cytokine TNF acts on endothelial cells to stimulate the production of another cytokine, IL-1, and many chemokines. The secondary mediators may have the same actions as the initial mediators but may also have different and even opposing activities. Such cascades provide mechanisms for amplifying—or, in certain instances, counteracting—the initial action of a mediator.

We next discuss the more important mediators of acute inflammation, focusing on their mechanisms of action and roles in acute inflammation.

Vasoactive Amines: Histamine and Serotonin

The two major vasoactive amines, so named because they have important actions on blood vessels, are *histamine* and *serotonin*. They are stored as preformed molecules in cells and are therefore among the first mediators to be released during inflammation.

The richest source of histamine are mast cells that are normally present in the connective tissue adjacent to blood vessels. It is also found in basophils and platelets. Histamine is stored in mast cell granules and is released by mast cell degranulation in response to a variety of stimuli, including (1) physical injury (such as trauma), cold, and heat, all by unknown mechanisms; (2) binding of antigen to IgE antibodies displayed on the surfaces of mast cells, which underlies immediate hypersensitivity (allergic) reactions (Chapter 6); and (3) products of complement called anaphylatoxins (C3a and C5a), described later. Antibodies and complement products bind to specific receptors on mast cells and trigger signaling pathways that induce rapid degranulation. Neuropeptides (e.g., substance P) and cytokines (IL-1, IL-8) may also trigger release of histamine.

Histamine causes dilation of arterioles and increases the permeability of venules. Histamine is considered to be the principal mediator of the immediate transient phase of increased vascular permeability, producing interendothelial gaps in venules, as discussed earlier. Its vasoactive effects are mediated mainly via binding to receptors on microvascular endothelial cells. The antihistamine drugs that are commonly used to treat some inflammatory reactions, such as allergies, are histamine receptor antagonists that bind to and block the receptor. Histamine also causes contraction of some smooth muscles.

Serotonin (5-hydroxytryptamine) is a preformed vasoactive mediator present in platelets and certain neuroendocrine cells, such as in the gastrointestinal tract. Its primary function is as a neurotransmitter in the gastrointestinal tract and the central nervous system. It is also a vasoconstrictor, but the importance of this action in inflammation is unclear.

Arachidonic Acid Metabolites

The lipid mediators *prostaglandins* and *leukotrienes* are produced from arachidonic acid (AA) present in membrane phospholipids and stimulate vascular and cellular reactions in acute inflammation. AA is a 20-carbon polyunsaturated fatty acid (5,8,11,14-eicosatetraenoic acid) that is derived from dietary sources or by synthesis from a precursor molecule, the essential fatty acid linoleic acid. Active AAs are derived from an esterified precursor found in membrane phospholipids. Mechanical, chemical, and physical stimuli or other mediators (e.g., C5a) release AA from membrane phospholipids through the action of cellular phospholipases, mainly phospholipase A_2. AA-derived mediators, also called eicosanoids (because they are derived from 20-carbon fatty acids; Greek *eicosa* = 20), are synthesized by two major classes of enzymes: cyclooxygenases (which generate prostaglandins) and lipoxygenases (which produce leukotrienes and lipoxins) (Fig. 3.10). Eicosanoids bind to G protein–coupled receptors on many cell types and can mediate virtually every step of inflammation (Table 3.6).

Prostaglandins

Prostaglandins (PGs) are produced by mast cells, macrophages, endothelial cells, and many other cell types, and are involved in the vascular and systemic reactions of inflammation. They are generated by the actions of two cyclooxygenases, called COX-1 and COX-2. COX-1 is constitutively expressed in most tissues, where it may serve various homeostatic functions (e.g., fluid and electrolyte balance in the kidneys, cytoprotection in the gastrointestinal tract), and is also induced by inflammatory stimuli. By contrast, COX-2 expression is mainly confined to cells that are participating in inflammatory reactions.

Prostaglandins are named based on common structural features coded by a letter (PGD, PGE, PGF, PGG, and PGH)

Table 3.6 Principal Actions of Arachidonic Acid Metabolites in Inflammation

Action	Eicosanoid
Vasodilation	Prostaglandins PGI_2 (prostacyclin), PGE_1, PGE_2, PGD_2
Vasoconstriction	Thromboxane A_2, leukotrienes C_4, D_4, E_4
Increased vascular permeability	Leukotrienes C_4, D_4, E_4
Chemotaxis, leukocyte adhesion	Leukotrienes B_4, HETE

HETE, Hydroxyeicosatetraenoic acid.

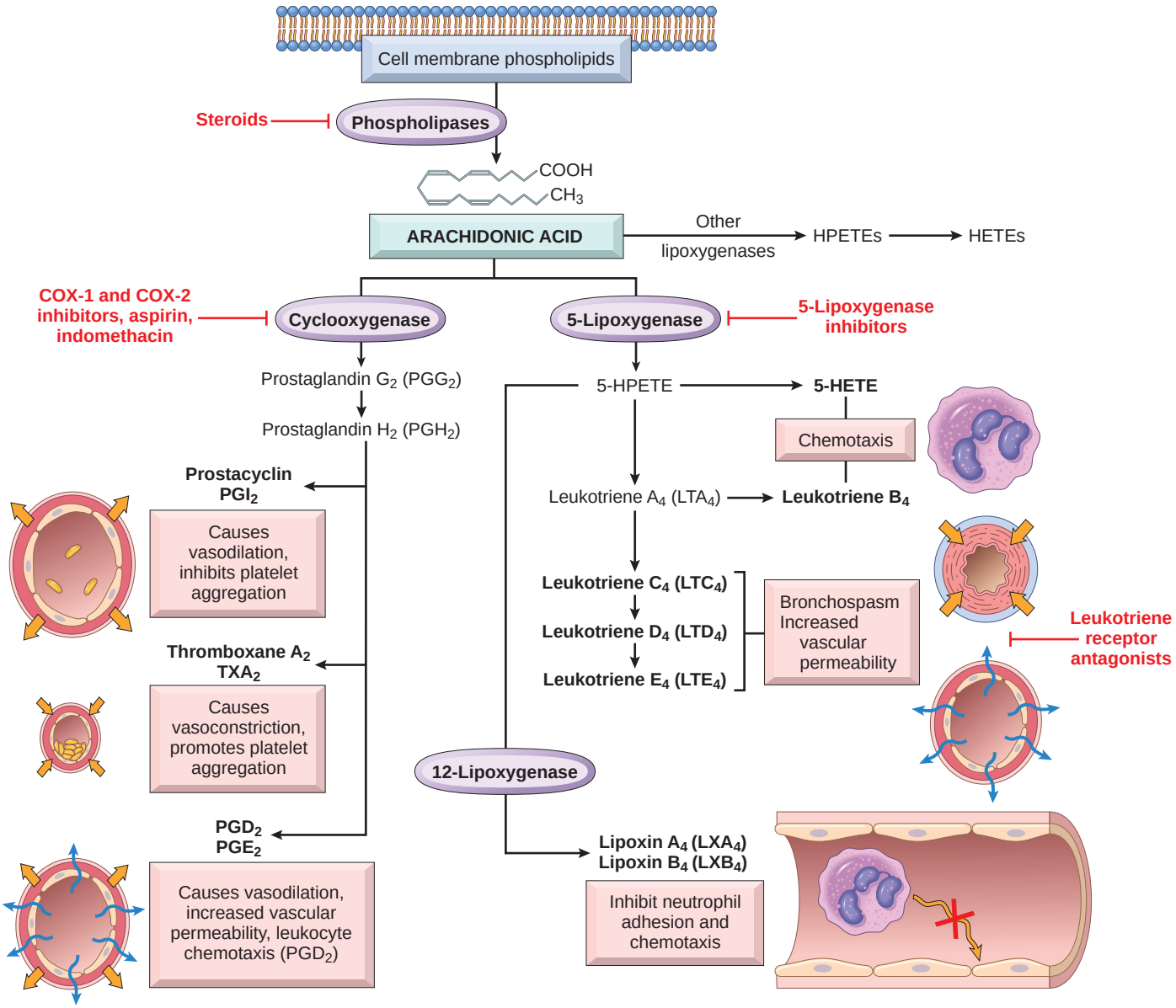


Figure 3.10 Production of arachidonic acid metabolites and their roles in inflammation. Note the enzymatic activities whose inhibition through pharmacologic intervention blocks major pathways (denoted with a red *X*). *COX-1*, *COX-2*, Cyclooxygenase 1 and 2; *HETE*, hydroxyeicosatetraenoic acid; *HPETE*, hydroperoxyeicosatetraenoic acid.

and a subscript numeral (e.g., 1, 2) that indicates the number of double bonds in the compound. The most important ones in inflammation are PGE_2, PGD_2, PGF_{2a}, PGI_2 (prostacyclin), and thromboxane A_2 (TxA_2), each of which is synthesized by a specific enzyme acting on an intermediate in the pathway. Some of these enzymes have restricted tissue distributions. For example, platelets contain the enzyme thromboxane synthase, and hence TxA_2 is the major product in these cells. TxA_2, a potent platelet-aggregating agent and vasoconstrictor, is itself unstable and rapidly converted to its inactive form. Vascular endothelium lacks thromboxane synthase but contains prostacyclin synthase, which is responsible for the formation of prostacyclin (PGI_2) and its stable end product PGF_{1a}. Prostacyclin is a vasodilator and a potent inhibitor of platelet aggregation, and also markedly potentiates the permeability-increasing and chemotactic effects of other mediators. A thromboxane-prostacyclin imbalance has been implicated as an early event in thrombus formation in coronary and cerebral blood vessels. PGD_2 is the major prostaglandin made by mast cells; along with PGE_2 (which is more widely distributed), it causes vasodilation and increases the permeability of postcapillary venules, thus potentiating edema formation. PGD_2 also is a chemoattractant for neutrophils.

In addition to their local effects, prostaglandins are involved in the pathogenesis of pain and fever in inflammation. PGE_2 is hyperalgesic, making the skin hypersensitive to painful stimuli such as intradermal injection of suboptimal concentrations of histamine and bradykinin. It is involved in cytokine-induced fever during infections (described later).

Leukotrienes

Leukotrienes are produced in leukocytes and mast cells by the action of lipoxygenase and are involved in vascular and smooth muscle reactions and leukocyte recruitment. There are three different lipoxygenases, 5-lipoxygenase being the predominant one in neutrophils. This enzyme converts AA to 5-hydroxyeicosatetraenoic acid, which is chemotactic for neutrophils and is the precursor of the leukotrienes. LTB_4 is a potent chemotactic agent and activator of neutrophils, causing aggregation and adhesion of the cells to venular endothelium, generation of ROS, and release of lysosomal enzymes. The cysteinyl-containing leukotrienes LTC_4, LTD_4, and LTE_4 cause intense vasoconstriction, bronchospasm (important in asthma), and increased permeability of venules. Leukotrienes are more potent than histamine in increasing vascular permeability and causing bronchospasm.

Lipoxins

Lipoxins are also generated from AA by the lipoxygenase pathway, but unlike prostaglandins and leukotrienes, the lipoxins suppress inflammation by inhibiting neutrophil chemotaxis and adhesion to endothelium. They are also unusual in that two cell populations are required for the transcellular biosynthesis of these mediators. Neutrophils synthesize precursors of active lipoxins and pass these to platelets, where they are converted to mature lipoxins.

Pharmacologic Inhibitors of Prostaglandins and Leukotrienes

The importance of eicosanoids in inflammation has driven attempts to develop drugs that inhibit their production or actions and thus suppress inflammation. These antiinflammatory drugs include the following.

- *Cyclooxygenase inhibitors* include aspirin and other nonsteroidal antiinflammatory drugs (NSAIDs), such as ibuprofen. They inactivate both COX-1 and COX-2 and thus inhibit prostaglandin synthesis (hence their efficacy in treating pain and fever); aspirin does this by irreversibly acetylating and inactivating cyclooxygenases. Selective COX-2 inhibitors are 200- to 300-fold more potent in blocking COX-2 than COX-1. There has been great interest in COX-2 as a therapeutic target because of the possibility that COX-1 is responsible for the production of prostaglandins that are involved in both inflammation and physiologic protective functions, whereas COX-2 generates prostaglandins that are involved only in inflammatory reactions. If this idea is correct, selective COX-2 inhibitors should be antiinflammatory without having the toxicities of the nonselective inhibitors, such as gastric ulceration. However, these distinctions are not absolute, as COX-2 also seems to play a role in normal homeostasis. Furthermore, selective COX-2 inhibitors may increase the risk of cardiovascular and cerebrovascular events, possibly because they impair endothelial cell production of PGI_2 (prostacyclin), a vasodilator and inhibitor of platelet aggregation, but leave intact the COX-1–mediated production by platelets of TxA_2, an important mediator of platelet aggregation and vasoconstriction. Thus, selective COX-2 inhibition may tilt the balance toward thromboxane and promote vascular thrombosis, especially in individuals with other factors that increase the risk of thrombosis. Nevertheless, these drugs are still used in individuals who do not have risk factors for cardiovascular disease when their benefits outweigh their risks.
- *Lipoxygenase inhibitors.* 5-Lipoxygenase is not affected by NSAIDs, and many new inhibitors of this enzyme pathway have been developed. Pharmacologic agents that inhibit leukotriene production are useful in the treatment of asthma.
- *Corticosteroids* are broad-spectrum antiinflammatory agents that reduce the transcription of genes encoding many proteins involved in inflammation, including COX-2, phospholipase A_2, proinflammatory cytokines (e.g., IL-1 and TNF), and iNOS.
- *Leukotriene receptor antagonists* block leukotriene receptors and prevent the actions of the leukotrienes. These drugs are useful in the treatment of asthma.
- Another approach to manipulating inflammatory responses has been to modify the intake and content of dietary lipids by increasing the consumption of fish oil. The proposed explanation for the effectiveness of this approach is that the polyunsaturated fatty acids in fish oil are poor substrates for conversion to active metabolites by the cyclooxygenase and lipoxygenase pathways but are better substrates for the production of antiinflammatory lipid products, including lipids called resolvins.

Cytokines and Chemokines

Cytokines are proteins produced by many cell types (principally activated lymphocytes, macrophages, and dendritic cells, but also endothelial, epithelial, and connective tissue cells) that mediate and regulate immune and inflammatory reactions. By convention, growth factors that act on epithelial and mesenchymal cells are not grouped under cytokines. The general properties and functions of cytokines are discussed in Chapter 6. The cytokines involved in acute inflammation are reviewed here (Table 3.7).

Tumor Necrosis Factor and Interleukin-1

TNF and IL-1 serve critical roles in leukocyte recruitment by promoting adhesion of leukocytes to endothelium and their migration through blood vessels. These cytokines are produced mainly by activated macrophages and dendritic cells; TNF is also produced by T lymphocytes and mast cells, and IL-1 is produced by some epithelial cells as well. The secretion of TNF and IL-1 can be stimulated by microbial products, dead cells, immune complexes, foreign bodies, physical injury, and a variety of other inflammatory stimuli. The production of TNF is induced by signals through TLRs and other microbial sensors. The synthesis of IL-1 is stimulated by the same signals, but the generation of the biologically active form of this cytokine is dependent on the inflammasome (described earlier).

The actions of TNF and IL-1 contribute to the local and systemic reactions of inflammation (Fig. 3.11). The most important roles of these cytokines in inflammation are the following.

- *Endothelial activation.* Both TNF and IL-1 act on endothelium to induce a spectrum of changes referred to as endothelial activation. These changes include increased expression of endothelial adhesion molecules, mostly

Table 3.7 Cytokines in Inflammation

Cytokine	Principal Sources	Principal Actions in Inflammation
In Acute Inflammation		
TNF	Macrophages, mast cells, T lymphocytes	Stimulates expression of endothelial adhesion molecules and secretion of other cytokines; systemic effects
IL-1	Macrophages, endothelial cells, some epithelial cells	Similar to TNF; greater role in fever
IL-6	Macrophages, other cells	Systemic effects (acute phase response)
Chemokines	Macrophages, endothelial cells, T lymphocytes, mast cells, other cell types	Recruitment of leukocytes to sites of inflammation; migration of cells in normal tissues
IL-17	T lymphocytes	Recruitment of neutrophils and monocytes
In Chronic Inflammation		
IL-12	Dendritic cells, macrophages	Increased production of IFN-γ
IFN-γ	T lymphocytes, NK cells	Activation of macrophages (increased ability to kill microbes and tumor cells)
IL-17	T lymphocytes	Recruitment of neutrophils and monocytes

IFN-γ, Interferon-γ; *IL,* interleukin; *NK,* natural killer; *TNF,* tumor necrosis factor.

The most important cytokines involved in inflammatory reactions are listed. Many other cytokines may play lesser roles in inflammation. There is also considerable overlap between the cytokines involved in acute and chronic inflammation. Specifically, all the cytokines listed under acute inflammation may also contribute to chronic inflammatory reactions.

E-selectin and P-selectin and ligands for leukocyte integrins; increased production of various mediators, including other cytokines and chemokines, growth factors, and eicosanoids; and increased procoagulant activity of the endothelium.

- *Activation of leukocytes and other cells.* TNF augments responses of neutrophils to other stimuli such as bacterial endotoxin and stimulates the microbicidal activity of macrophages, in part by inducing production of NO. IL-1 activates fibroblasts to synthesize collagen and stimulates proliferation of synovial and other mesenchymal cells. IL-1 also stimulates Th17 responses, which in turn induce acute inflammation.
- *Systemic acute-phase response.* IL-1 and TNF (as well as IL-6) induce the systemic acute-phase responses associated with infection or injury, discussed later. TNF regulates energy balance by promoting lipid and protein mobilization and by suppressing appetite. Therefore,

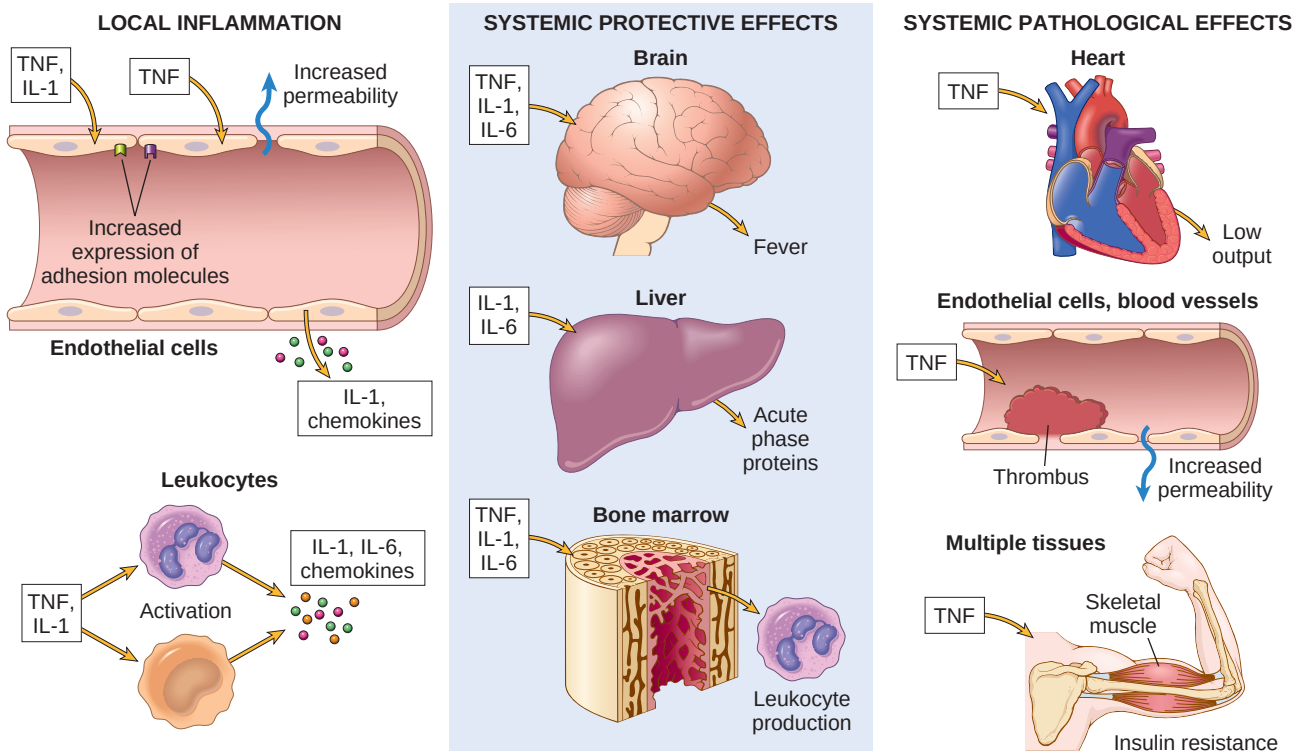


Figure 3.11 Major roles of cytokines in acute inflammation. *IL,* Interleukin; *TNF,* tumor necrosis factor.

sustained production of TNF contributes to cachexia, a pathologic state characterized by weight loss and anorexia that is seen in some chronic infections and neoplastic diseases.

TNF antagonists have been remarkably effective in the treatment of chronic inflammatory diseases, particularly rheumatoid arthritis, psoriasis, and some types of inflammatory bowel disease. One of the complications of this therapy is that patients become susceptible to mycobacterial infection, reflecting the reduced ability of macrophages to kill intracellular microbes. Although many of the actions of TNF and IL-1 seem overlapping, IL-1 antagonists are not as effective for reasons that remain unclear. Also, blocking either cytokine has no effect on the outcome of sepsis, perhaps because other cytokines contribute to this systemic inflammatory reaction.

Chemokines

Chemokines are a family of small (8 to 10 kDa) proteins that act primarily as chemoattractants for specific types of leukocytes. About 40 different chemokines and 20 different receptors for chemokines have been identified. They are classified into four major groups, according to the arrangement of cysteine (C) residues in the proteins.

- *C-X-C chemokines* have one amino acid residue separating the first two of the four conserved cysteine residues. A subset of these chemokines acts primarily on neutrophils. IL-8 (now called CXCL8) is typical of this group. It is secreted by activated macrophages, endothelial cells, and other cell types and causes activation and chemotaxis of neutrophils, with limited activity on monocytes and eosinophils. Its most important inducers are microbial products and other cytokines, mainly IL-1 and TNF.
- *C-C chemokines* have the first two conserved cysteine residues adjacent. The C-C chemokines, which include monocyte chemoattractant protein (MCP-1, CCL2), eotaxin (CCL11), macrophage inflammatory protein-1α (MIP-1α, CCL3), and others, generally attract monocytes, eosinophils, basophils, and lymphocytes, but are less potent chemoattractants for neutrophils. Although most of the chemokines in this class have overlapping actions, eotaxin selectively recruits eosinophils.
- *C chemokines* lack the first and third of the four conserved cysteines. The C chemokines (e.g., lymphotactin, XCL1) are relatively specific for lymphocytes.
- *CX_3C chemokines* contain three amino acids between the two cysteines. The only known member of this class is called fractalkine (CX3CL1). This chemokine exists in two forms: a cell surface–bound protein induced on endothelial cells by inflammatory cytokines that promotes strong adhesion of monocytes and T cells, and a soluble form derived by proteolysis of the membrane-bound protein that has potent chemoattractant activity for the same cells.

Chemokines mediate their activities by binding to seven-transmembrane G protein–coupled receptors. These receptors usually exhibit overlapping ligand specificities, and leukocytes generally express more than one receptor type. As discussed in Chapter 6, certain chemokine receptors (CXCR4, CCR5) act as coreceptors for a viral envelope glycoprotein of human immunodeficiency virus (HIV), the cause of AIDS, and are thus involved in binding and entry of the virus into cells.

Chemokines may be displayed at high concentrations attached to proteoglycans on the surface of endothelial cells and in the extracellular matrix (ECM). They have two main functions.

- *In acute inflammation.* Inflammatory chemokines are the ones whose production is induced by microbes and other stimuli. These chemokines stimulate leukocyte attachment to endothelium by acting on leukocytes to increase the affinity of integrins, and they stimulate migration (chemotaxis) of leukocytes in tissues to the site of infection or tissue damage.
- *Maintenance of tissue architecture.* Some chemokines are produced constitutively in tissues and are sometimes called homeostatic chemokines. These organize various cell types in different anatomic regions of the tissues, such as T and B lymphocytes in discrete areas of the spleen and lymph nodes (Chapter 6).

Although the role of chemokines in inflammation is well established, it has proved difficult to develop antagonists that are effective therapeutic agents.

Other Cytokines in Acute Inflammation

The list of cytokines implicated in inflammation is huge and constantly growing. In addition to the ones described earlier, two that have received considerable recent interest are IL-6, made by macrophages and other cells, which is involved in local and systemic reactions, and IL-17, produced mainly by T lymphocytes, which promotes neutrophil recruitment. Antagonists against both are effective for the treatment of inflammatory diseases, such as juvenile arthritis (anti-IL-6 receptor) and psoriasis (anti-IL-17). Cytokines also play key roles in chronic inflammation; these are described later.

Complement System

The complement system is a collection of plasma proteins that function mainly in host defense against microbes and in pathologic inflammatory reactions. The system consists of more than 20 proteins, some of which are numbered C1 through C9. The complement proteins function in both innate and adaptive immunity for defense against microbial pathogens. In the process of complement activation, several cleavage products of complement proteins are elaborated that cause increased vascular permeability, chemotaxis, and opsonization. The activation and functions of complement are outlined in Fig. 3.12.

Complement proteins are present in inactive forms in the plasma, and many of them are activated to become proteolytic enzymes that degrade other complement proteins, thus forming an enzymatic cascade capable of tremendous amplification. The critical step in complement activation is the proteolysis of the third (and most abundant) component, C3. **Cleavage of C3 can occur by one of three pathways:**

- The *classical pathway,* which is triggered by binding of C1 to antibody (IgM or IgG) that has combined with antigen.

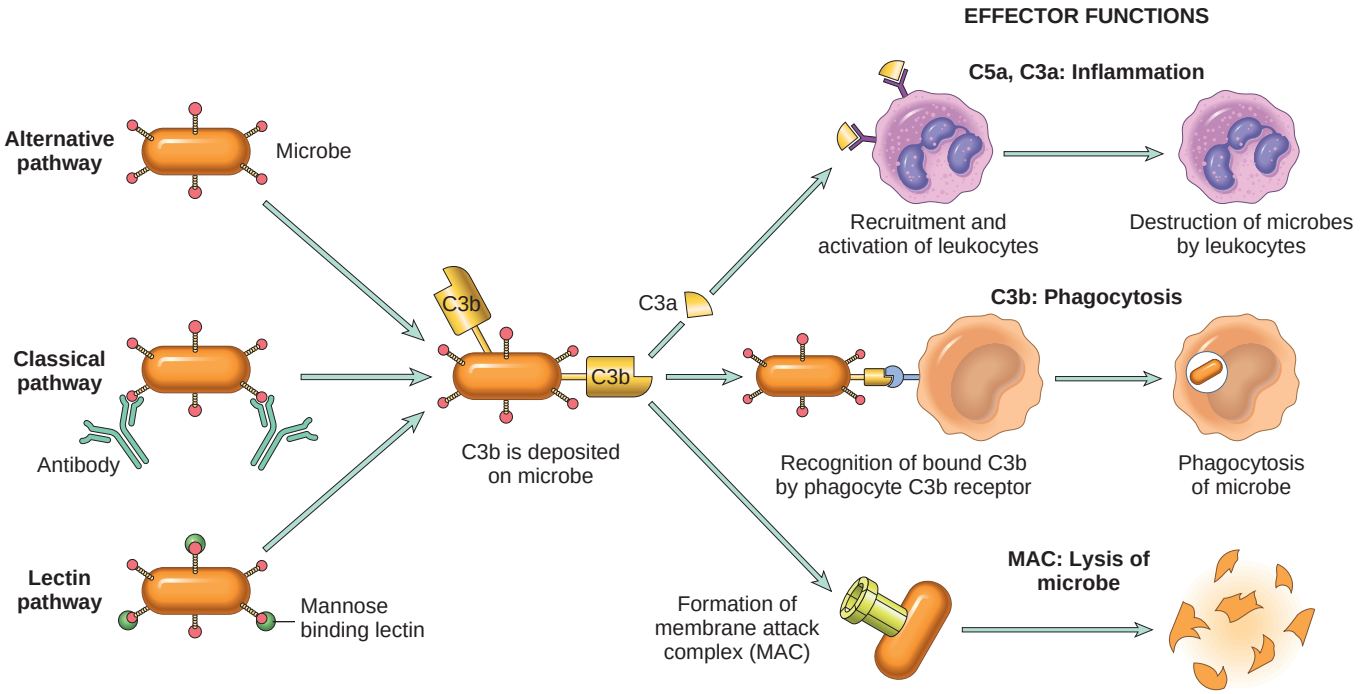


Figure 3.12 Activation and functions of the complement system. Activation of complement by different pathways leads to cleavage of C3. The functions of the complement system are mediated by breakdown products of C3 and other complement proteins and by the membrane attack complex *(MAC)*.

- The *alternative pathway,* which can be triggered by microbial surface molecules (e.g., endotoxin, or lipopolysaccharide [LPS]), complex polysaccharides, cobra venom, and other substances, in the absence of antibody.
- The *lectin pathway,* in which plasma mannose-binding lectin binds to carbohydrates on microbes and directly activates C1.

All three pathways of complement activation lead to the formation of an active enzyme called *C3 convertase,* which splits C3 into two functionally distinct fragments, C3a and C3b. C3a is released, and C3b becomes covalently attached to the cell or molecule where complement is being activated. More C3b then binds to the previously generated fragments to form C5 convertase, which cleaves C5 to release C5a and leave C5b attached to the cell surface. C5b binds the late components (C6–C9), culminating in the formation of the membrane attack complex (composed of multiple C9 molecules).

The complement system has three main functions (see Fig. 3.12).

- *Inflammation.* C5a, C3a, and, to a lesser extent, C4a are cleavage products of the corresponding complement components that stimulate histamine release from mast cells and thereby increase vascular permeability and cause vasodilation. They are called anaphylatoxins because they have effects similar to those of mast cell mediators that are involved in the reaction called anaphylaxis (Chapter 6). C5a is also a chemotactic agent for neutrophils, monocytes, eosinophils, and basophils. In addition, C5a activates the lipoxygenase pathway of AA metabolism in neutrophils and monocytes, causing further release of inflammatory mediators.
- *Opsonization and phagocytosis. C3b* and its cleavage product inactive C3b (iC3b), when fixed to a microbial cell wall, act as opsonins and promote phagocytosis by neutrophils and macrophages, which bear cell surface receptors for the complement fragments.
- *Cell lysis.* The deposition of the membrane attack complex on cells makes these cells permeable to water and ions and results in osmotic lysis of the cells. This role of complement is particularly important for killing microbes with thin cell walls, such as *Neisseria* bacteria, and deficiency of the terminal components of complement predisposes to *Neisseria* infections.

The activation of complement is tightly controlled by cell-associated and circulating regulatory proteins. Different regulatory proteins inhibit the production of active complement fragments or remove fragments that deposit on cells. These regulators are expressed on normal host cells and are thus designed to prevent healthy tissues from being injured at sites of complement activation. Regulatory proteins can be overwhelmed when large amounts of complement are deposited on host cells and in tissues, as happens in autoimmune diseases, in which individuals produce complement-fixing antibodies against their own cell and tissue antigens (Chapter 6). The most important of these regulatory proteins are the following:

- *C1 inhibitor (C1 INH)* blocks the activation of C1, the first protein of the classical complement pathway. Inherited deficiency of this inhibitor is the cause of hereditary angioedema.
- *Decay accelerating factor (DAF)* and *CD59* are two proteins that are linked to plasma membranes by a glycophosphatidylinositol (GPI) anchor. DAF prevents formation of C3 convertases, and CD59 inhibits formation of the

membrane attack complex. An acquired deficiency of the enzyme that creates GPI anchors leads to deficiency of these regulators and excessive complement activation and lysis of red cells (which are sensitive to complement-mediated cell lysis) in the disease called paroxysmal nocturnal hemoglobinuria (PNH) (Chapter 14).

- *Complement Factor H* is a circulating glycoprotein that inhibits the alternative pathway of complement activation by promoting the cleavage and destruction of C3b and the turnover of the C3 convertases. Inherited defects in Factor H and several other regulatory proteins that interact with Factor H cause an atypical form of hemolytic uremic syndrome (Chapter 20), in which complement deposits in glomerular vessels, leading to endothelial damage and formation of platelet-rich thrombi. Polymorphisms in the *Factor H* gene have also been linked to age-related macular degeneration (Chapter 29), an important cause of vision loss in older adults.

The complement system contributes to disease in several ways. The activation of complement by antibodies or antigen-antibody complexes deposited on host cells and tissues is an important mechanism of cell and tissue injury (Chapter 6). Inherited deficiencies of complement proteins cause increased susceptibility to infections (Chapter 6), and, as mentioned earlier, deficiencies of regulatory proteins cause a variety of disorders resulting from excessive complement activation.

Other Mediators of Inflammation

Platelet-Activating Factor

Platelet-activating factor (PAF) is a phospholipid-derived mediator that was discovered as a factor that caused platelet aggregation, but it is now known to have multiple inflammatory effects. A variety of cell types, including platelets, basophils, mast cells, neutrophils, macrophages, and endothelial cells, can elaborate PAF in both secreted and cell-bound forms. In addition to platelet aggregation, PAF causes vasoconstriction and bronchoconstriction, and at low concentrations it induces vasodilation and increased venular permeability. Despite these actions, the use of PAF antagonists in various inflammatory diseases has not been found to be useful.

Products of Coagulation

A link between the coagulation pathway and inflammation is supported by the presence of protease-activated receptors (PARs) on leukocytes that are activated by thrombin (the protease that cleaves fibrinogen to produce fibrin, which forms the clot). It is, however, likely that the major role of the PARs is in platelet activation during clotting (Chapter 4). There is also some evidence that cleavage products of fibrin (fibrinopeptides) can stimulate inflammation. However, many forms of tissue injury are associated with both clotting and inflammation, and it is difficult to establish a cause-and-effect relationship.

Kinins

Kinins are vasoactive peptides derived from plasma proteins, called kininogens, by the action of specific proteases called kallikreins. The enzyme kallikrein cleaves a plasma glycoprotein precursor, high-molecular-weight kininogen, to produce bradykinin. **Bradykinin increases vascular permeability and causes contraction of smooth muscle, dilation of blood vessels, and pain when injected into the skin.** These effects are similar to those of histamine. The action of bradykinin is short-lived because it is quickly inactivated by an enzyme called kininase. Bradykinin has been implicated as a mediator in some forms of allergic reaction, such as anaphylaxis (Chapter 6).

Neuropeptides

Neuropeptides are secreted by sensory nerves and various leukocytes and may play a role in the initiation and regulation of inflammatory responses. These small peptides, such as substance P and neurokinin A, are produced in the central and peripheral nervous systems. Nerve fibers containing substance P are prominent in the lung and gastrointestinal tract. Substance P has many activities that may be important in inflammation, including the transmission of pain signals and increasing vascular permeability. Leukocytes express receptors for many neuropeptides, so these neural products could provide a mechanism for "cross-talk" between the nervous system and immune and inflammatory reactions. For instance, activation of the efferent vagus nerve inhibits the production of proinflammatory cytokines such as TNF, providing a mechanism for suppressing inflammation. This observation has led to clinical trials of vagus nerve stimulation in patients with rheumatoid arthritis.

When Lewis discovered the role of histamine in inflammation, one mediator was thought to be enough. Now, we are wallowing in them! Yet, from this large compendium, it is likely that a few mediators are most important for the reactions of acute inflammation in vivo, and these are summarized in Table 3.8. The redundancy of the mediators and their actions ensures that this protective response is robust and not readily subverted.

Table 3.8 Role of Mediators in Different Reactions of Inflammation

Reaction of Inflammation	Principal Mediators
Vasodilation	Histamine Prostaglandins
Increased vascular permeability	Histamine and serotonin C3a and C5a (by liberating vasoactive amines from mast cells, other cells) Leukotrienes C_4, D_4, E_4
Chemotaxis, leukocyte recruitment and activation	TNF, IL-1 Chemokines C3a, C5a Leukotriene B_4
Fever	IL-1, TNF Prostaglandins
Pain	Prostaglandins Bradykinin Substance P
Tissue damage	Lysosomal enzymes of leukocytes Reactive oxygen species

IL-1, Interleukin-1; *TNF*, tumor necrosis factor.

KEY CONCEPTS

ACTIONS OF THE PRINCIPAL MEDIATORS OF INFLAMMATION

- Vasoactive amines, mainly histamine: Vasodilation and increased vascular permeability.
- Arachidonic acid metabolites (prostaglandins and leukotrienes): Several forms exist and are involved in vascular reactions, leukocyte chemotaxis, and other reactions of inflammation; antagonized by lipoxins.
- Cytokines: Proteins produced by many cell types; usually act at short range; mediate multiple effects, mainly leukocyte recruitment and migration; principal ones in acute inflammation are TNF, IL-1, and chemokines.
- Complement proteins: Activation of the complement system by microbes or antibodies leads to the generation of multiple breakdown products, which are responsible for leukocyte chemotaxis, opsonization, phagocytosis of microbes and other particles, and cell killing.
- Kinins: Produced by proteolytic cleavage of precursors; mediate vascular reaction, pain.

Morphologic Patterns of Acute Inflammation

The morphologic hallmarks of acute inflammatory reactions are dilation of small blood vessels and accumulation of leukocytes and fluid in the extravascular tissue. However, special morphologic patterns are often superimposed on these general features, depending on the severity of the reaction, its specific cause, and the particular tissue and site involved. The importance of recognizing the gross and microscopic patterns is that they often provide valuable clues about the underlying cause.

Serous Inflammation

Serous inflammation is marked by the exudation of cell-poor fluid into spaces created by cell injury or into body cavities lined by the peritoneum, pleura, or pericardium. Typically, the fluid in serous inflammation does not contain microbes or large numbers of leukocytes (which tend to produce purulent inflammation, described later). In body cavities the fluid may be derived from the plasma (as a result of increased vascular permeability) or from the secretions of mesothelial cells (as a result of local irritation); accumulation of fluid in these cavities is called an effusion. (Effusions also occur in noninflammatory conditions, such as reduced blood outflow in heart failure or reduced plasma protein levels in some kidney and liver diseases.) The skin blister resulting from a burn or viral infection represents accumulation of serous fluid within or immediately beneath the damaged epidermis of the skin (Fig. 3.13).

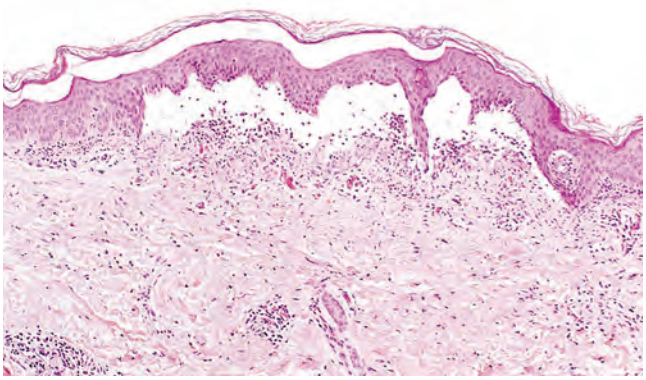


Figure 3.13 Serous inflammation. Low-power view of a cross section of a skin blister showing the epidermis separated from the dermis by a focal collection of serous effusion.

Fibrinous Inflammation

With greater increase in vascular permeability, large molecules such as fibrinogen pass out of the blood, and fibrin is formed and deposited in the extracellular space. **A fibrinous exudate develops when the vascular leaks are large or there is a local procoagulant stimulus** (e.g., caused by cancer cells). A fibrinous exudate is characteristic of inflammation in the lining of body cavities, such as the meninges, pericardium (Fig. 3.14A), and pleura. Histologically, fibrin appears as an eosinophilic meshwork of threads or sometimes as an amorphous coagulum (Fig. 3.14B). Fibrinous exudates may be dissolved by fibrinolysis and cleared by macrophages. If the fibrin is not removed, over time it may stimulate the ingrowth of fibroblasts and blood vessels and thus lead to

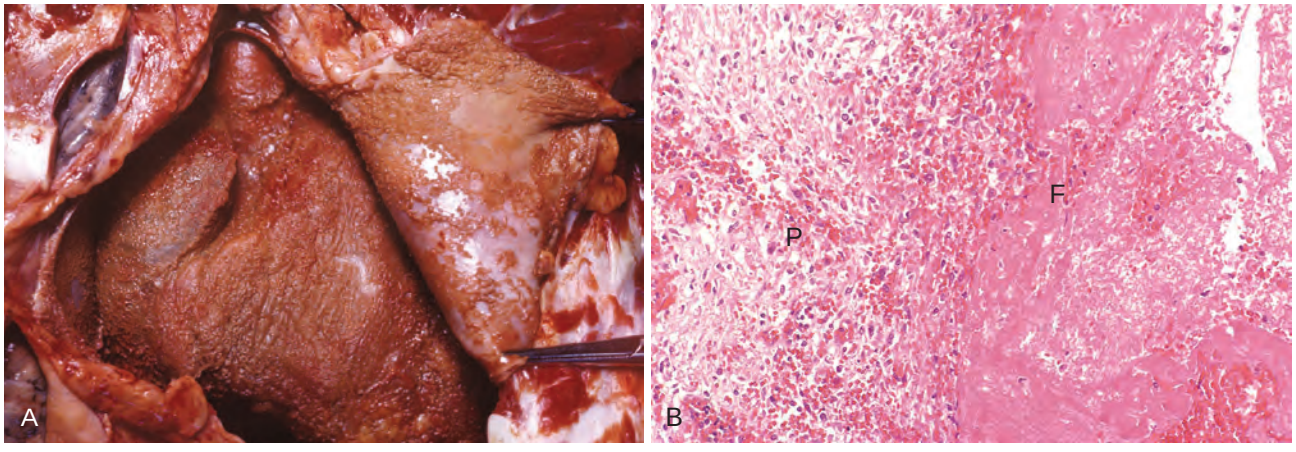


Figure 3.14 Fibrinous pericarditis. (A) Deposits of fibrin on the pericardium. (B) A pink meshwork of fibrin exudate *(F)* overlies the pericardial surface *(P)*.

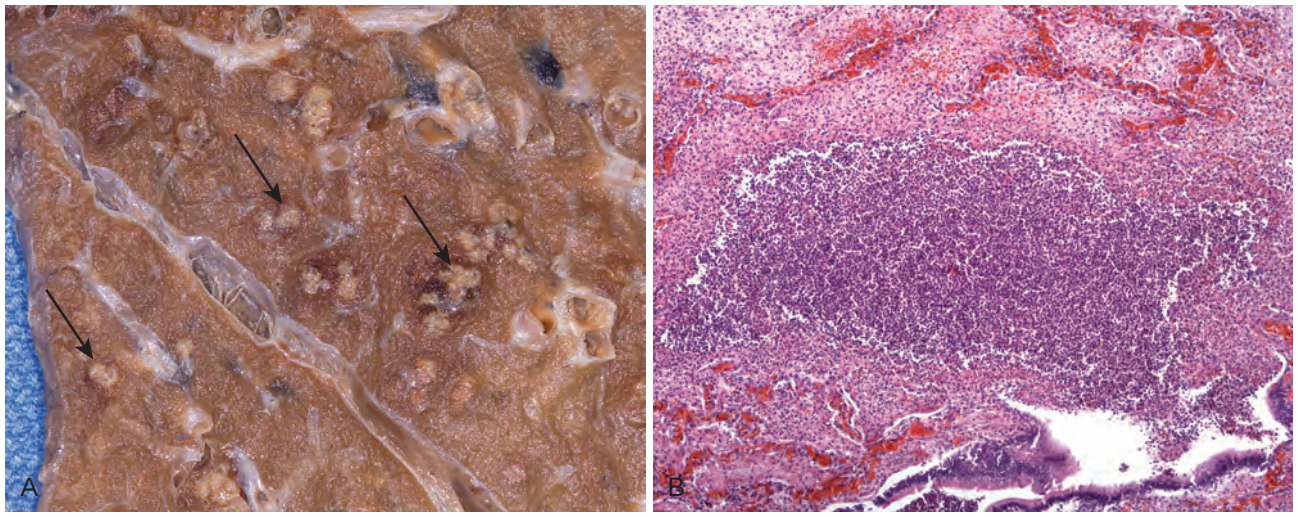


Figure 3.15 Purulent inflammation. (A) Multiple bacterial abscesses *(arrows)* in the lung in a case of bronchopneumonia. (B) The abscess contains neutrophils and cellular debris and is surrounded by congested blood vessels.

scarring. Conversion of the fibrinous exudate to scar tissue (organization) within the pericardial sac leads to opaque fibrous thickening of the pericardium and epicardium in the area of exudation and, if the fibrosis is extensive, obliteration of the pericardial space.

Purulent (Suppurative) Inflammation and Abscess

Purulent inflammation is characterized by the production of pus, an exudate consisting of neutrophils, the liquefied debris of necrotic cells, and edema fluid. The most frequent cause of purulent (also called suppurative) inflammation is infection with bacteria that cause liquefactive tissue necrosis, such as staphylococci; these pathogens are referred to as pyogenic (pus-producing) bacteria. A common example of an acute suppurative inflammation is acute appendicitis.

Abscesses **are localized collections of pus** caused by suppuration buried in a tissue, an organ, or a confined space. They are produced by seeding of pyogenic bacteria into a tissue (Fig. 3.15). Abscesses have a central liquefied region composed of necrotic leukocytes and tissue cells. There is usually a zone of preserved neutrophils around this necrotic focus, and outside this region there may be vascular dilation and parenchymal and fibroblastic proliferation, indicating chronic inflammation and repair. In time the abscess may become walled off and ultimately replaced by connective tissue.

Ulcers

An ulcer is a local defect, or excavation, of the surface of an organ or tissue that is produced by the sloughing (shedding) of inflamed necrotic tissue (Fig. 3.16). Ulceration can occur only when tissue necrosis and resultant inflammation exist on or near a surface. It is most common in (1) the mucosa of the mouth, stomach, intestines, or genitourinary tract, and (2) the skin and subcutaneous tissue of the lower extremities in individuals with disorders that predispose to vascular insufficiency, such as diabetes, sickle cell anemia, and peripheral vascular disease.

Ulcerations are best exemplified by peptic ulcer of the stomach or duodenum, in which acute and chronic inflammation coexist. During the acute stage there is intense polymorphonuclear infiltration and vascular dilation in the margins of the defect. With chronicity, the margins and base of the ulcer develop fibroblastic proliferation, scarring, as well as the accumulation of lymphocytes, macrophages, and plasma cells.

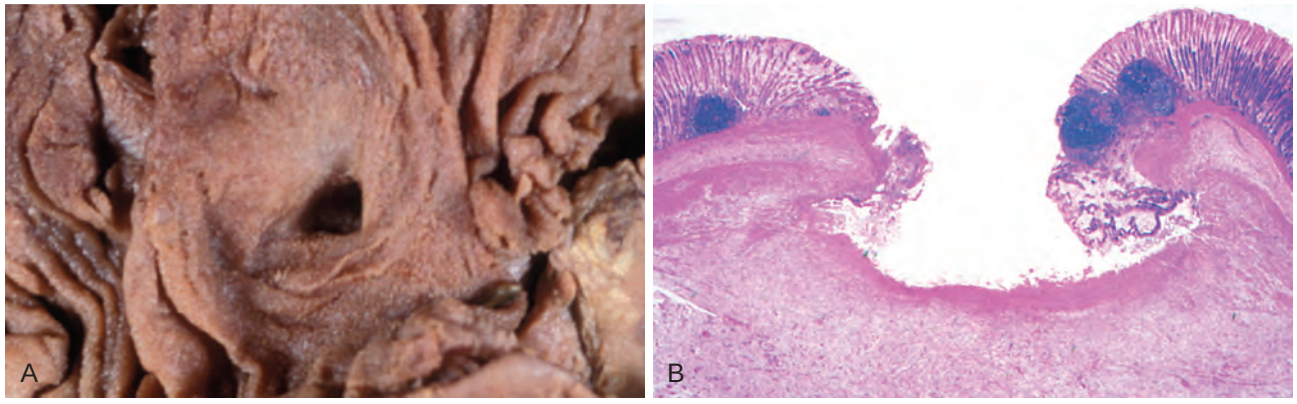


Figure 3.16 The morphology of an ulcer. (A) A chronic duodenal ulcer. (B) Low-power cross-section view of a duodenal ulcer crater with an acute inflammatory exudate in the base.

Outcomes of Acute Inflammation

Although, as might be expected, many variables may modify the basic process of inflammation, including the nature and intensity of the injury, the site and tissue affected, and the responsiveness of the host, **acute inflammatory reactions typically have one of three outcomes** (Fig. 3.17).

- *Complete resolution.* In a perfect world, all inflammatory reactions, once they have succeeded in eliminating the offending agent, would end with restoration of the site of acute inflammation to normal. This is called resolution and is the usual outcome when the injury is limited or short-lived or when there has been little tissue destruction and the damaged parenchymal cells can regenerate. Resolution involves removal of cellular debris and microbes by macrophages and resorption of edema fluid by lymphatics, followed by regeneration of the damaged tissue.
- *Healing by connective tissue replacement (scarring, or fibrosis).* This occurs after substantial tissue destruction, when the inflammatory injury involves tissues that are incapable of regeneration, or when there is abundant fibrin exudation in tissue or in serous cavities (pleura, peritoneum) that cannot be adequately cleared. In all these situations, connective tissue grows into the area of damage or exudate, converting it into a mass of fibrous tissue, a process also called organization.
- Progression of the response to *chronic inflammation* (discussed later). Acute to chronic transition occurs when the acute inflammatory response cannot be resolved, as a result of either the persistence of the injurious agent or some interference with the normal process of healing.

Summary of Acute Inflammation

Now that we have described the components, mediators, and pathologic manifestations of acute inflammatory responses, it is useful to summarize the main features of a typical response of this type. When a host encounters an injurious agent, such as a microbe or dead cells, resident phagocytes and phagocytes recruited from the blood try to eliminate these agents. At the same time, phagocytes and other host cells react to the presence of the foreign or abnormal substance by liberating cytokines, lipid messengers, and other mediators of inflammation. Some of these mediators act on small blood vessels in the vicinity and promote the efflux of plasma proteins and the recruitment of circulating leukocytes to the site where the offending agent is located. The recruited leukocytes are activated by molecules derived from microbes and injured cells and by locally produced mediators, and the activated leukocytes try to remove the offending agent by phagocytosis. As the injurious agent is eliminated and antiinflammatory mechanisms become active, the process subsides and the host returns to a normal state of health. If the injurious agent cannot be quickly eliminated, the result may be chronic inflammation.

The vascular and cellular reactions account for the cardinal signs of inflammation: rubor, calor, tumor, dolor, and functio laesa. The increased blood flow to the injured area and increased vascular permeability lead to the accumulation of extravascular fluid rich in plasma proteins, known as *edema*. The redness *(rubor),* warmth *(calor),* and swelling *(tumor)* are caused by the increased blood flow and edema. Circulating leukocytes, initially predominantly neutrophils, adhere to the endothelium via adhesion

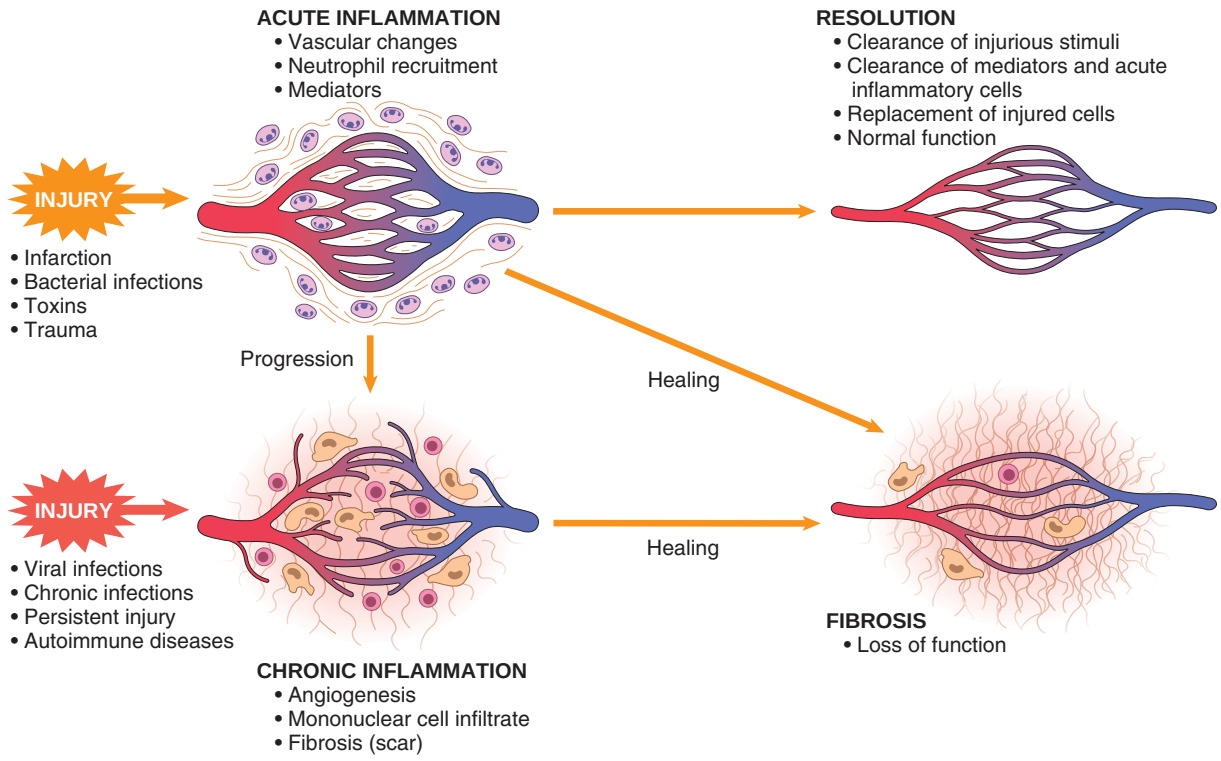


Figure 3.17 Outcomes of acute inflammation: resolution, healing by fibrosis, or chronic inflammation. The components of the various reactions and their functional outcomes are listed.

molecules, traverse the endothelium, and migrate to the site of injury under the influence of chemotactic agents. Leukocytes that are activated by the offending agent and by endogenous mediators may release toxic metabolites and proteases extracellularly, causing tissue damage. Due to the damage and the liberation of prostaglandins, neuropeptides, and cytokines, one of the local symptoms is pain *(dolor)*. Loss of function *(functio laesa)* results from pain and injury to the tissues.

CHRONIC INFLAMMATION

Chronic inflammation is a response of prolonged duration (weeks or months) in which inflammation, tissue injury, and attempts at repair coexist in varying combinations. It may follow acute inflammation, as described earlier, or chronic inflammation may begin insidiously, as a low-grade, smoldering response without any manifestations of a preceding acute reaction.

Causes of Chronic Inflammation

Chronic inflammation arises in the following settings.

- *Persistent infections* by microorganisms that are difficult to eradicate, such as mycobacteria and certain viruses, fungi, and parasites. These organisms often evoke an immune reaction called delayed-type hypersensitivity (Chapter 6). Chronic inflammatory responses sometimes develop a specific pattern called a granulomatous reaction (discussed later). In other cases an unresolved acute inflammation may evolve into chronic inflammation, as may occur in acute bacterial infection of the lung that progresses to a chronic lung abscess. Acute and chronic inflammation may coexist, as in a peptic ulcer.
- *Hypersensitivity diseases.* Chronic inflammation plays an important role in a group of diseases that are caused by excessive and inappropriate activation of the immune system. Under certain conditions, immune reactions develop against the individual's own tissues, leading to autoimmune diseases (Chapter 6). In these diseases, autoantigens evoke a self-perpetuating immune reaction that results in chronic tissue damage and inflammation; examples of such diseases include rheumatoid arthritis and multiple sclerosis. In other cases, chronic inflammation is the result of unregulated immune responses against microbes, as in inflammatory bowel disease. Immune responses against common environmental substances are the cause of allergic diseases, such as bronchial asthma (Chapter 6). Because these autoimmune and allergic reactions are triggered against antigens that are normally harmless, the reactions serve no useful purpose and only cause disease. Such diseases may show morphologic patterns of mixed acute and chronic inflammation because they are characterized by repeated bouts of inflammation. Fibrosis may dominate the late stages.
- Prolonged exposure to potentially *toxic agents*, either exogenous or endogenous. An example of an exogenous agent is particulate silica, a nondegradable inanimate material that, when inhaled for prolonged periods, results in an inflammatory lung disease called silicosis (Chapter 15). Atherosclerosis (Chapter 11) is a chronic inflammatory process of the arterial wall induced, at least in part, by excessive production and tissue deposition of endogenous cholesterol and other lipids.

Morphologic Features

In contrast to acute inflammation, which is manifested by vascular changes, edema, and predominantly neutrophilic infiltration, **chronic inflammation is characterized by** the following:

- *Infiltration with mononuclear cells,* which include macrophages, lymphocytes, and plasma cells (Fig. 3.18).
- *Tissue destruction,* induced by the persistent offending agent or by the inflammatory cells.
- *Attempts at healing* by connective tissue replacement of damaged tissue, accomplished by *angiogenesis* (proliferation of small blood vessels) and, in particular, *fibrosis.*

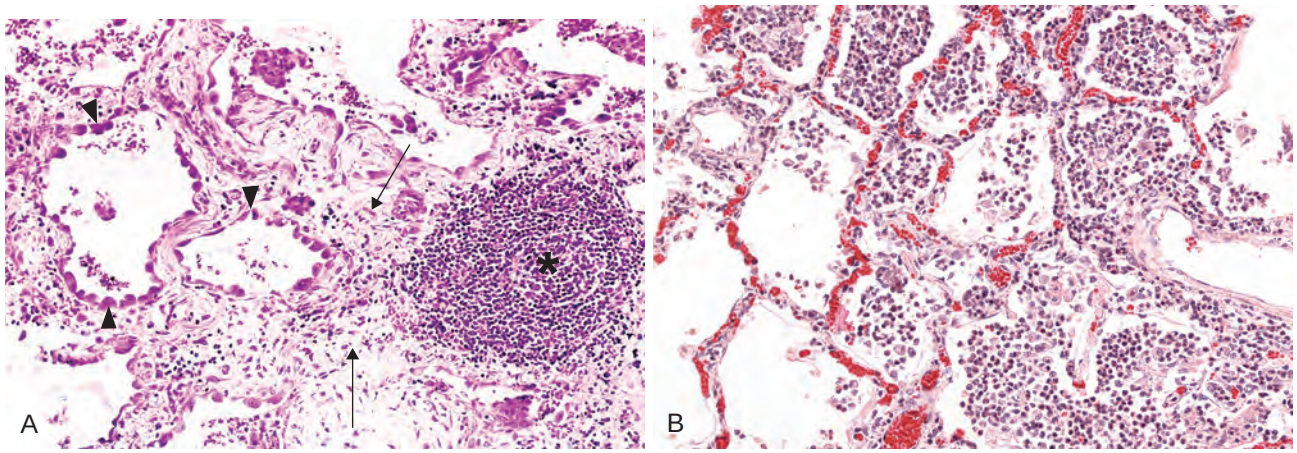


Figure 3.18 (A) Chronic inflammation in the lung, showing all three characteristic histologic features: (1) collection of chronic inflammatory cells *(asterisk)*, (2) destruction of parenchyma (normal alveoli are replaced by spaces lined by cuboidal epithelium) *(arrowheads)*, and (3) replacement by connective tissue (fibrosis) *(arrows)*. (B) In contrast, in acute inflammation of the lung (acute bronchopneumonia), neutrophils fill the alveolar spaces, and blood vessels are congested.

Because angiogenesis and fibrosis are components of wound healing and repair, they are discussed later, in the context of tissue repair.

Cells and Mediators of Chronic Inflammation

The combination of leukocyte infiltration, tissue damage, and fibrosis that characterize chronic inflammation is the result of the local activation of several cell types and the production of mediators.

Role of Macrophages

The dominant cells in most chronic inflammatory reactions are macrophages, which contribute to the reaction by secreting cytokines and growth factors that act on various cells, destroying foreign invaders and tissues, and activating other cells, notably T lymphocytes. Macrophages are professional phagocytes that eliminate microbes and damaged tissues. They also serve important roles in the repair of injured tissues. Here we review the development and functions of macrophages.

Macrophages are tissue cells derived from hematopoietic stem cells in the bone marrow in postnatal life and from progenitors in the embryonic yolk sac and fetal liver during early development (Fig. 3.19). Circulating cells of this lineage are known as monocytes. Macrophages are normally diffusely scattered in most connective tissues. In addition, they are present in specific locations in organs such as the liver (where they are called Kupffer cells), spleen and lymph nodes (called sinus histiocytes), central nervous system (microglial cells), and lungs (alveolar macrophages). Together, these cells comprise the mononuclear phagocyte system, also known by the older (and inaccurate) name reticuloendothelial system.

Committed progenitors in the bone marrow give rise to monocytes, which enter the blood, migrate into various tissues, and differentiate into macrophages. This is typical of macrophages at sites of inflammation and in some tissues such as the skin and intestinal tract. The half-life of blood monocytes is about 1 day, whereas the lifespan of tissue macrophages may be several months or years. Other specialized types of macrophages, such as microglia, Kupffer cells, and alveolar macrophages, arise from progenitors in the yolk sac or fetal liver very early in embryogenesis and migrate to the developing brain, liver, and lung, where they persist throughout life as a stable population of resident cells. As discussed earlier, in inflammatory reactions, monocytes begin to emigrate into extravascular tissues quite early, and within 48 hours they may constitute the predominant cell type. Extravasation of monocytes is governed by the same factors that are involved in neutrophil emigration, that is, adhesion molecules and chemotactic factors.

The products of activated macrophages eliminate injurious agents such as microbes and initiate the process of repair, but are also responsible for much of the tissue injury in chronic inflammation. Several functions of macrophages are central to the development and persistence of chronic inflammation and the accompanying tissue injury.

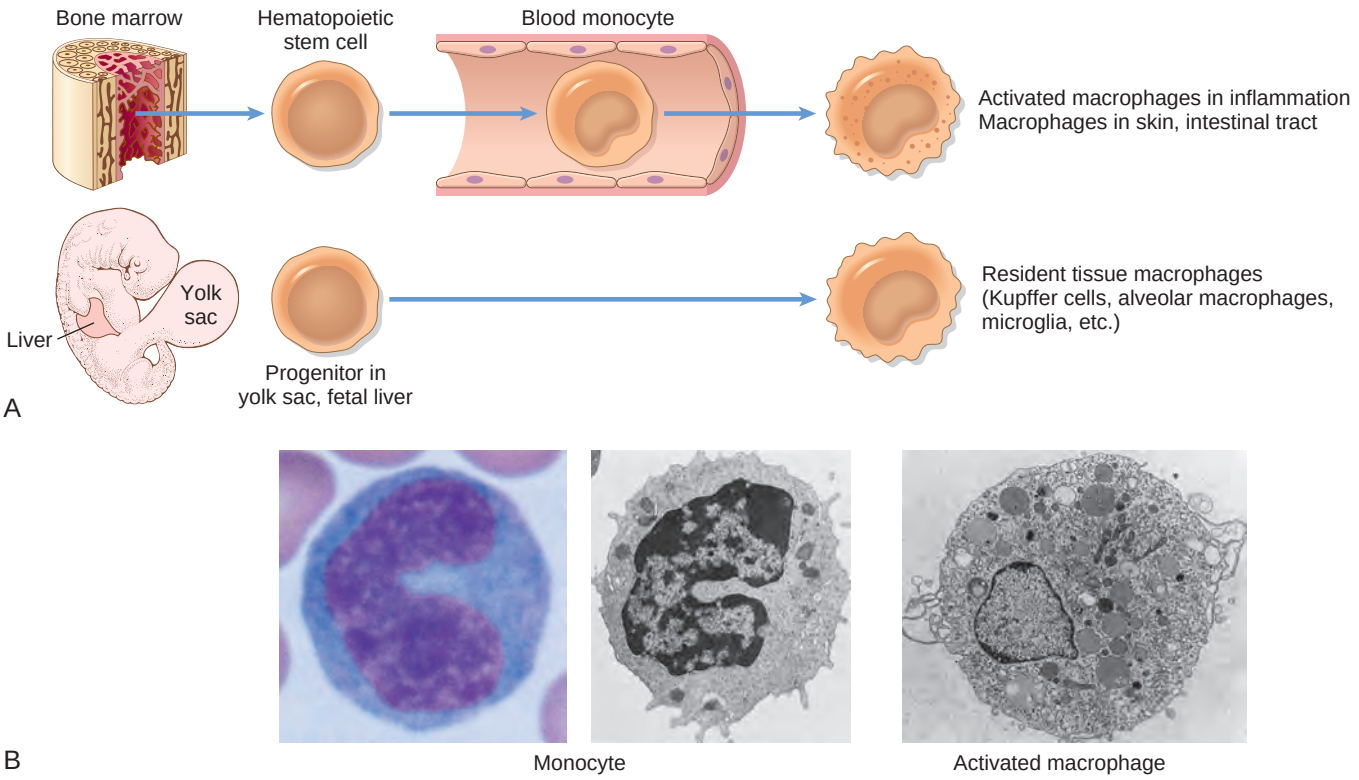


Figure 3.19 Maturation of mononuclear phagocytes. (A) In postnatal life, macrophages arise mainly from bone marrow progenitors and blood monocytes. These cells make up the majority of resident macrophages in some tissues and become more prominent after injury and during inflammation. Some tissue macrophages, including microglia and alveolar macrophages, arise from embryonic precursors and migrate into tissue, where they persist throughout life. (B) The morphology of a monocyte and activated macrophage.

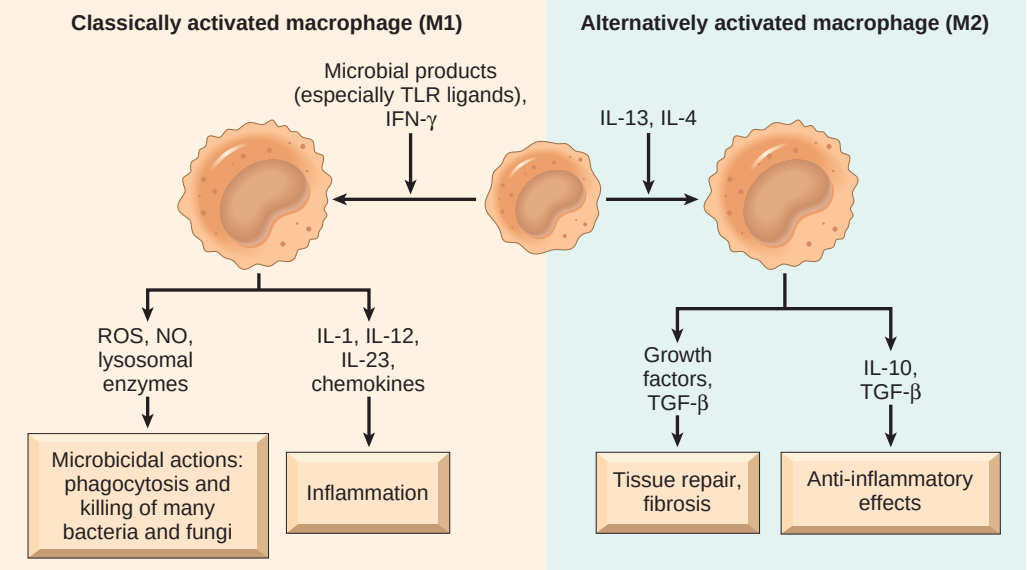


Figure 3.20 Classical and alternative macrophage activation. Different stimuli activate monocytes/macrophages to develop into functionally distinct populations. Classically activated macrophages are induced by microbial products and cytokines, particularly interferon-γ (IFN-γ). They phagocytose and destroy microbes and dead tissues and can potentiate inflammatory reactions. Alternatively activated macrophages are induced by other cytokines and are important in tissue repair and resolution of inflammation. *IL*, Interleukin; *NO*, nitric oxide; *ROS*, reactive oxygen species; *TGF-β*, transforming growth factor-β; *TLR*, Toll-like receptor.

- Macrophages, like the other type of phagocytes, the neutrophils, ingest and eliminate microbes and dead tissues.
- Macrophages initiate the process of tissue repair and are involved in scar formation and fibrosis. These processes are discussed later in the chapter.
- Macrophages secrete mediators of inflammation, such as cytokines (TNF, IL-1, chemokines, and others) and eicosanoids. Thus, macrophages contribute to the initiation and propagation of inflammatory reactions.
- Macrophages display antigens to T lymphocytes and respond to signals from T cells, thus setting up a feedback loop that is essential for defense against many microbes by cell-mediated immune responses. These interactions are described further in the subsequent discussion of the role of lymphocytes in chronic inflammation and in more detail in Chapter 6 where cell-mediated immunity is considered.

There are two major pathways of macrophage activation, called *classical* and *alternative*, that endow macrophages with different functional activities (Fig. 3.20).

- *Classical macrophage activation* may be induced by microbial products such as endotoxin, which engage TLRs and other sensors; by T cell–derived signals, importantly the cytokine IFN-γ, in immune responses; or by foreign substances, including crystals and particulate matter. Classically activated (also called M1) macrophages produce NO and lysosomal enzymes, which enhance their ability to kill ingested organisms, and secrete cytokines that stimulate inflammation. The main role of these macrophages in host defense is to destroy microbes and promote the inflammatory response.
- *Alternative macrophage activation* is induced by cytokines other than IFN-γ, such as IL-4 and IL-13, produced by T lymphocytes and other cells. These macrophages are not actively microbicidal; instead, their principal functions are to terminate inflammation and promote tissue repair.

It seems plausible that in response to most injurious stimuli the first activation pathway is the classical one, designed to destroy the offending agents, and this is followed by alternative activation, which initiates tissue repair. However, such a precise sequence is not documented in most inflammatory reactions, and most reactions contain varying numbers of both types. Also, the M1 and M2 phenotypes are the extreme forms, and there may be many intermediate types that are difficult to characterize as typical M1 or M2.

Their impressive arsenal of mediators makes macrophages powerful allies in the body's defense against unwanted invaders, but the same weaponry can also induce considerable tissue destruction when macrophages are inappropriately or excessively activated. It is largely because of these activities of macrophages that tissue destruction is one of the hallmarks of chronic inflammation.

In some instances, if the irritant is eliminated, macrophages eventually disappear (either dying off or making their way into the lymphatics and lymph nodes). In others, macrophage accumulation persists, as a result of continuous recruitment from the circulation and local proliferation at the site of inflammation.

Role of Lymphocytes

Microbes and other environmental antigens activate T and B lymphocytes, which amplify and propagate chronic inflammation. Activation of these cells is often prominent in chronic inflammatory reactions, and when they are involved, the inflammation tends to be persistent and severe, in part because lymphocyte activation leads to the generation

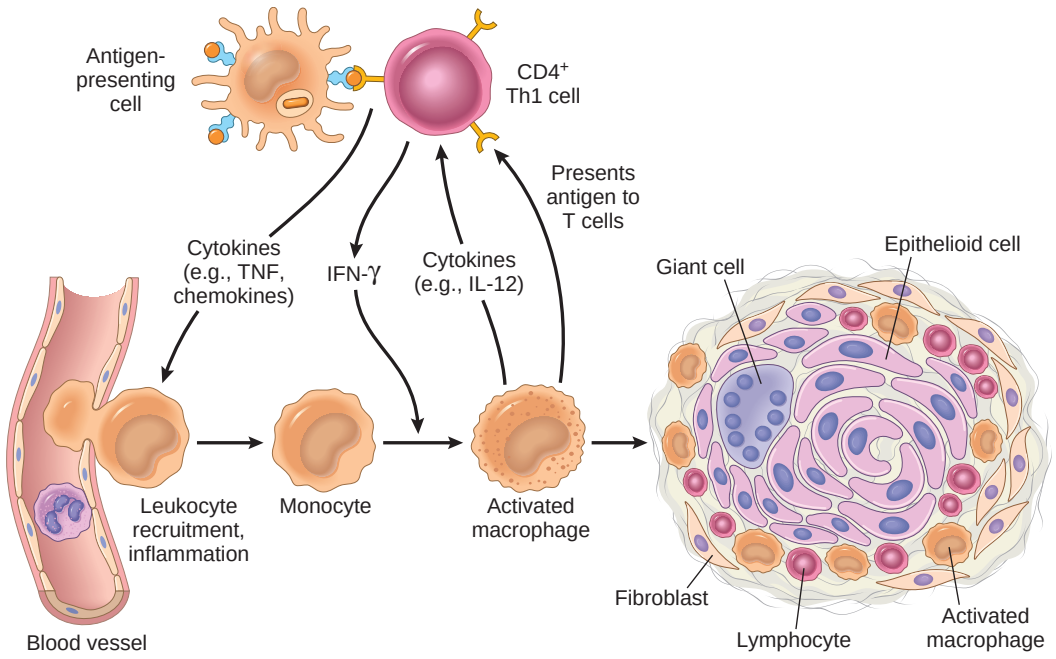


Figure 3.21 Macrophage–lymphocyte interactions in chronic inflammation. Activated T cells produce cytokines that recruit macrophages (tumor necrosis factor *[TNF]*, interleukin-17 *[IL-17]*, chemokines) and others that activate macrophages (interferon-γ *[IFN-γ]*). Activated macrophages in turn stimulate T cells by presenting antigens and via cytokines such as IL-12. Prolonged reactions involving T cells and macrophages may result in granuloma formation.

of long-lived memory cells. Some of the persistent chronic inflammatory reactions, such as granulomatous inflammation, described later, are dependent on generation of memory lymphocyte responses. Lymphocytes may be the dominant population in the chronic inflammation seen in various autoimmune diseases.

By virtue of their ability to secrete cytokines, CD4+ T lymphocytes promote inflammation and influence the nature of the inflammatory reaction. There are three subsets of CD4+ T cells that secrete different types of cytokines and elicit different types of inflammation.

- Th1 cells produce the cytokine IFN-γ, which activates macrophages by the classical pathway.
- Th2 cells secrete IL-4, IL-5, and IL-13, which recruit and activate eosinophils and are responsible for the alternative pathway of macrophage activation.
- Th17 cells secrete IL-17 and other cytokines, which induce the secretion of chemokines responsible for recruiting neutrophils (and monocytes) into the reaction.

Both Th1 and Th17 cells are involved in defense against many types of bacteria and viruses and in autoimmune diseases in which tissue injury is caused by chronic inflammation. Th2 cells are important in defense against helminthic parasites and in allergic inflammation. These T-cell subsets and their functions are described in more detail in Chapter 6.

Lymphocytes and macrophages interact in a bidirectional way, and these interactions play an important role in propagating chronic inflammation (Fig. 3.21). Macrophages display antigens to T cells, express membrane molecules called costimulators, and produce cytokines (IL-12 and others) that stimulate T-cell responses (Chapter 6). Activated T lymphocytes, in turn, produce cytokines, described earlier, which recruit and activate macrophages, promoting more antigen presentation and cytokine secretion. The result is a cycle of cellular reactions that fuel and sustain chronic inflammation. A typical consequence of such a chronic T cell–macrophage reaction is the formation of granulomas, described later.

Activated B lymphocytes and antibody-producing plasma cells are also often present at sites of chronic inflammation. The antibodies produced may be specific for persistent foreign antigens or self antigens in the inflammatory site or against altered tissue components. However, the contribution of antibodies to most chronic inflammatory disorders is unclear.

In some chronic inflammatory reactions, the accumulated lymphocytes, antigen-presenting cells, and plasma cells cluster together to form organized lymphoid follicles resembling those seen in lymph nodes. These are called tertiary lymphoid organs; this type of lymphoid organogenesis is often seen in the synovium of patients with long-standing rheumatoid arthritis and in the thyroid in Hashimoto thyroiditis. It has been postulated that the local formation of lymphoid follicles may perpetuate the immune reaction, but the significance of these structures is not established.

Other Cells in Chronic Inflammation

Other cell types may be prominent in chronic inflammation induced by particular stimuli.

- *Eosinophils* are abundant in immune reactions mediated by IgE and in parasitic infections (Fig. 3.22). Their recruitment is driven by adhesion molecules similar to those used by neutrophils and by specific chemokines (e.g., eotaxin) derived from leukocytes and epithelial cells. Eosinophils have granules that contain major basic protein, a highly cationic protein that is toxic to helminths but

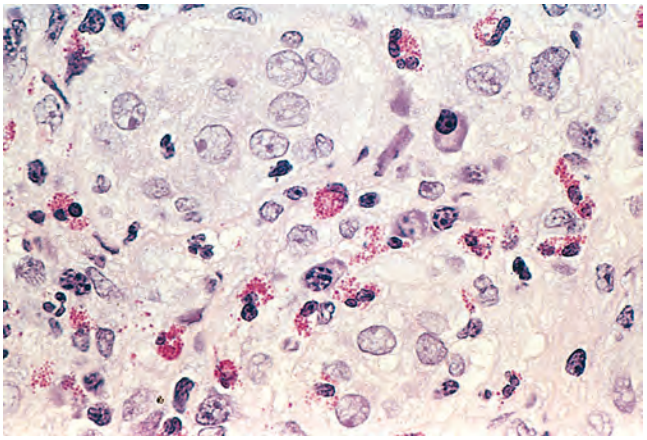


Figure 3.22 Focus of inflammation containing numerous eosinophils.

also may injure host epithelial cells. This is why eosinophils are of benefit in controlling helminth infections, yet they also contribute to tissue damage in immune reactions such as allergies (Chapter 6).

- *Mast cells* are widely distributed in connective tissues and participate in both acute and chronic inflammatory reactions. Mast cells express on their surface the receptor (FcεRI) that binds the Fc portion of IgE antibody. In immediate hypersensitivity reactions, IgE antibodies bound to the cells' Fc receptors specifically recognize antigen, and the cells degranulate and release mediators such as histamine and prostaglandins (Chapter 6). This type of response occurs during allergic reactions to foods, insect venom, or drugs, sometimes with catastrophic results (e.g., anaphylactic shock). Mast cells are also present in chronic inflammatory reactions, and because they secrete a plethora of cytokines, they may promote inflammatory reactions in different situations.
- Although *neutrophils* are characteristic of acute inflammation, many forms of chronic inflammation continue to show large numbers of neutrophils, induced either by persistent microbes or by mediators produced by activated macrophages and T lymphocytes. In chronic bacterial infection of bone (osteomyelitis), a neutrophilic exudate can persist for many months. Neutrophils are also important in the chronic damage induced in lungs by smoking and other irritant stimuli (Chapter 15). This pattern of inflammation has been called acute on chronic.

Granulomatous Inflammation

Granulomatous inflammation is a form of chronic inflammation characterized by collections of activated macrophages, often with T lymphocytes, and sometimes associated with necrosis. Granuloma formation is a cellular attempt to contain an offending agent that is difficult to eradicate. In this attempt there is often strong activation of T lymphocytes leading to macrophage activation, which can cause injury to normal tissues. The activated macrophages may develop abundant cytoplasm and begin to resemble epithelial cells and are called epithelioid cells. Some activated macrophages may fuse, forming multinucleate giant cells.

There are two types of granulomas, which differ in their pathogenesis.

- *Foreign body granulomas* are incited by inert foreign bodies, which induce inflammation in the absence of T cell–mediated immune responses. Typically, foreign body granulomas form around materials such as talc (associated with intravenous drug abuse) (Chapter 9), sutures, or other fibers that are large enough to preclude phagocytosis by a macrophage and are not immunogenic, so they do not incite any specific immune response. Epithelioid cells and giant cells are apposed to the surface of the foreign body. The foreign material can usually be identified in the center of the granuloma, sometimes within the giant cells, particularly if viewed with polarized light, in which it may appear refractile.
- *Immune granulomas* are caused by a variety of agents that are capable of inducing a persistent T cell–mediated immune response. This type of immune response usually produces granulomas when the inciting agent is difficult to eradicate, such as a persistent microbe. In such responses, activated Th1 cells produce cytokines such as IFN-γ, which activates macrophages (see Fig. 3.21). In some parasitic infections, such as schistosomiasis, the granulomas are associated with strong Th2 responses and eosinophils.

MORPHOLOGY

In the usual hematoxylin and eosin preparations, the activated macrophages in granulomas have pink granular cytoplasm with indistinct cell boundaries and are called **epithelioid cells** because of their resemblance to epithelia (Fig. 3.23). The aggregates of epithelioid macrophages are surrounded by a collar of lymphocytes. Older granulomas may have a rim of fibroblasts and connective tissue. Frequently, but not invariably, multinucleated **giant cells** 40 to 50 μm in diameter are found in granulomas; these are called Langhans giant cells. They consist of a large mass of cytoplasm and many nuclei, and they derive from the fusion of multiple activated macrophages. In granulomas associated with certain infectious organisms (most classically *Mycobacterium tuberculosis*), a combination of hypoxia and free radical–mediated injury leads

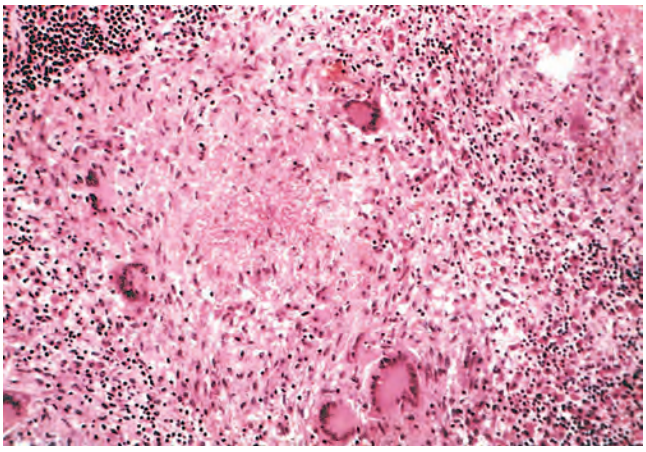


Figure 3.23 Typical tuberculous granuloma showing an area of central necrosis surrounded by multiple Langhans-type giant cells, epithelioid cells, and lymphocytes.

to a central zone of necrosis. Grossly, this has a granular, cheesy appearance and is therefore called **caseous necrosis.** Microscopically, this necrotic material appears as amorphous, structureless, eosinophilic, granular debris, with complete loss of cellular details (as opposed to coagulative necrosis, in which cell outlines are preserved). The granulomas in Crohn disease, sarcoidosis, and foreign body reactions tend to not have necrotic centers and are said to be noncaseating. Healing of granulomas is accompanied by fibrosis that may be extensive in involved organs.

Recognition of granulomatous inflammation is important because of the limited number of conditions (some life-threatening) that cause it (Table 3.9). **Tuberculosis is the prototype of a granulomatous disease caused by infection and should always be excluded as the cause when granulomas are identified.** In this disease the granuloma is referred to as a tubercle. Granulomas may also develop in some immune-mediated inflammatory diseases. Notable among these are Crohn disease, one type of inflammatory bowel disease and an important cause of granulomatous inflammation in the United States, and sarcoidosis, a disorder of unknown etiology. The morphologic patterns in the various granulomatous diseases may be sufficiently different to allow reasonably accurate diagnosis by an experienced pathologist (see Table 3.9); however, there are so many atypical presentations that it is always necessary to identify the specific etiologic agent by special stains for organisms (e.g., acid-fast stains for tubercle bacilli), by culture methods (e.g., in tuberculosis and fungal diseases), by molecular techniques (e.g., the polymerase chain reaction in tuberculosis), and by serologic studies (e.g., in syphilis).

Table 3.9 Examples of Diseases With Granulomatous Inflammation

Disease	Cause	Tissue Reaction
Tuberculosis	*Mycobacterium tuberculosis*	Caseating granuloma (tubercle): focus of activated macrophages (epithelioid cells), rimmed by fibroblasts, lymphocytes, histiocytes, occasional Langhans giant cells; central necrosis with amorphous granular debris; acid-fast bacilli
Leprosy	*Mycobacterium leprae*	Acid-fast bacilli in macrophages; noncaseating granulomas
Syphilis	*Treponema pallidum*	Gumma: microscopic to grossly visible lesion, enclosing wall of histiocytes; plasma cell infiltrate; central cells are necrotic without loss of cellular outline
Cat-scratch disease	Gram-negative bacillus	Rounded or stellate granuloma containing central granular debris and recognizable neutrophils; giant cells uncommon
Sarcoidosis	Unknown etiology	Noncaseating granulomas with abundant activated macrophages
Crohn disease (inflammatory bowel disease)	Immune reaction against intestinal bacteria, possibly self antigens	Occasional noncaseating granulomas in the wall of the intestine, with dense chronic inflammatory infiltrate

KEY CONCEPTS

CHRONIC INFLAMMATION

- Chronic inflammation is a prolonged host response to persistent stimuli.
- It is caused by microbes that resist elimination, immune responses against self antigens and environmental antigens, and some toxic substances (e.g., silica), and it underlies many medically important diseases.
- It is characterized by coexisting inflammation, tissue injury, and attempted repair by scarring.
- The cellular infiltrate consists of macrophages, lymphocytes, plasma cells, and other leukocytes.
- It is mediated by cytokines produced by macrophages and lymphocytes (notably T lymphocytes); bidirectional interactions between these cells tend to amplify and prolong the inflammatory reaction.
- Granulomatous inflammation is a pattern of chronic inflammation induced by T-cell and macrophage activation in response to an agent that is resistant to eradication.

SYSTEMIC EFFECTS OF INFLAMMATION

Inflammation, even if localized, is associated with cytokine-induced systemic reactions that are collectively called the *acute-phase response.* Anyone who has suffered through a severe bout of a viral illness (e.g., influenza) has experienced the systemic manifestations of acute inflammation. These changes are reactions to cytokines whose production is stimulated by bacterial products and by other inflammatory stimuli. The cytokines TNF, IL-1, and IL-6 are important mediators of the acute-phase reaction; other cytokines, notably interferons, also contribute.

The acute-phase response consists of several clinical and pathologic changes.

- *Fever,* characterized by an elevation of body temperature, usually by 1°C to 4°C, is one of the most prominent manifestations of the acute-phase response, especially when inflammation is associated with infection. Substances that induce fever are called pyrogens and include bacterial products (exogenous pyrogens, e.g., LPS) and cytokines, principally IL-1 and TNF (called endogenous pyrogens). Exogenous pyrogens act by stimulating immune cells to release IL-1 and TNF, which upregulate cyclooxygenases, the enzymes that synthesize prostaglandins. Among the cells that respond to IL-1 and TNF are vascular and perivascular cells of the hypothalamus. Prostaglandins released by these cells, especially PGE_2, stimulate the production of neurotransmitters by the hypothalamus that reset the body's steady-state temperature to a higher level by reducing heat loss (via vasoconstriction) and

increasing heat generation (through effects on brown fat and skeletal muscle). NSAIDs, including aspirin, reduce fever by inhibiting prostaglandin synthesis. An elevated body temperature has been shown to help amphibians ward off microbial infections, and although the details are not understood, it is assumed that fever is a protective host response in mammals as well.

- Elevated levels of *acute-phase proteins,* which are plasma proteins, mostly synthesized in the liver, whose plasma concentrations may increase several hundred–fold as part of the response to inflammatory stimuli. Three of the best known of these proteins are C-reactive protein (CRP), fibrinogen, and serum amyloid A (SAA) protein. Synthesis of these molecules in hepatocytes is stimulated by cytokines, especially IL-6 (for CRP and fibrinogen) and IL-1 or TNF (for SAA). Many acute-phase proteins, such as CRP and SAA, bind to microbial cell walls, and they may act as opsonins and fix complement. They also bind chromatin, possibly aiding in clearing necrotic cell nuclei. Fibrinogen binds to red cells and causes them to form stacks (rouleaux) that sediment more rapidly at unit gravity than do individual red cells. This is the basis for measuring the *erythrocyte sedimentation rate* as a simple test for an inflammatory response caused by any stimulus. Acute-phase proteins have beneficial effects during acute inflammation, but prolonged production of these proteins (especially SAA) in states of chronic inflammation causes secondary amyloidosis (Chapter 6). Elevated serum levels of CRP have been proposed as a marker for increased risk of myocardial infarction in patients with coronary artery disease. It is postulated that inflammation involving atherosclerotic plaques in the coronary arteries may predispose to thrombosis and subsequent infarction. Two other liver proteins that are released in increased amounts as part of the acute phase response often lead to altered blood counts. *Hepcidin* is a small protein that reduces the availability of iron to erythroid progenitors in the bone marrow; over time, this effect may lead to the anemia of chronic inflammation (Chapter 14). *Thrombopoietin*, the major growth factor for megakaryocytes (platelet precursors) in the bone marrow, is also upregulated and as a result systemic inflammation may be associated with an elevated platelet count (thrombocytosis).
- *Leukocytosis* is a common feature of inflammatory reactions, especially those induced by bacterial infections. The leukocyte count usually climbs to 15,000 or 20,000 cells/mL, but sometimes it may reach extraordinarily high levels of 40,000 to 100,000 cells/mL. These extreme elevations are sometimes called leukemoid reactions because they are similar to the white cell counts observed in leukemia, from which they must be distinguished. The leukocytosis occurs initially because of accelerated release of granulocytes from the bone marrow (caused by cytokines, including TNF and IL-1) and is therefore associated with a rise in the number of both mature and immature neutrophils in the blood, referred to as a shift to the left (for arcane reasons having to do with the way these cells were counted manually by technicians in days past). Prolonged infection also induces proliferation of precursors in the bone marrow, caused by increased production of colony-stimulating factors (CSFs) mainly from activated macrophages and marrow stromal cells. The increase in the bone marrow output of leukocytes compensates for the loss of these cells in the inflammatory reaction. (See also the discussion of leukocytosis in Chapter 13.) Most bacterial infections induce an increase in the blood neutrophil count, called neutrophilia. Viral infections, such as infectious mononucleosis, mumps, and German measles, cause an absolute increase in the number of lymphocytes (lymphocytosis). In some allergies and helminth infestations, there is an increase in the absolute number of eosinophils, creating eosinophilia. Certain infections (typhoid fever and infections caused by some viruses, rickettsiae, and certain protozoa) are associated with a decreased number of circulating white cells (leukopenia), in part because of sequestration of activated leukocytes in vascular spaces and tissues.
- Other manifestations of the acute-phase response include increased pulse and blood pressure; decreased sweating, mainly because of redirection of blood flow from cutaneous to deep vascular beds, to minimize heat loss through the skin; rigors (shivering), chills (search for warmth), anorexia, somnolence, and malaise, probably because of the actions of cytokines on brain cells.
- In severe bacterial infections (sepsis), the large amounts of bacteria and their products in the blood stimulate the production of enormous quantities of several cytokines, notably TNF and IL-1. High blood levels of cytokines cause various clinical manifestations such as disseminated intravascular coagulation, hypotensive shock, and metabolic disturbances, including insulin resistance and hyperglycemia. This clinical triad is known as *septic shock* and is one form of a severe, often fatal disorder referred to as systemic inflammatory response syndrome; it is discussed in more detail in Chapter 4.

KEY CONCEPTS

SYSTEMIC EFFECTS OF INFLAMMATION

- Fever: cytokines (TNF, IL-1) stimulate production of prostaglandins in hypothalamus.
- Production of acute-phase proteins: C-reactive protein, others; synthesis stimulated by cytokines (IL-6, others) acting on liver cells.
- Leukocytosis: cytokines (CSFs) stimulate production of leukocytes from precursors in the bone marrow.
- In some severe infections, septic shock: fall in blood pressure, disseminated intravascular coagulation, metabolic abnormalities; induced by high levels of TNF and other cytokines.

Whereas excessive inflammation is the underlying cause of many human diseases, defective inflammation results mainly in increased susceptibility to infections. The most common cause of defective inflammation is leukocyte deficiency due to replacement of the bone marrow by leukemias and metastatic tumors and suppression of the marrow by therapies for cancer and graft rejection. Inherited genetic abnormalities of leukocyte adhesion and microbicidal function are rare but informative; these are described in Chapter 6, in the context of immunodeficiency diseases. Deficiencies of the complement system are mentioned earlier and are described further in Chapter 6.

TISSUE REPAIR

Overview of Tissue Repair

Repair, also called healing, refers to the restoration of tissue architecture and function after an injury. By convention, the term *repair* is used for parenchymal and connective tissues and *healing* for surface epithelia, but these distinctions are not based on biology, and we use the terms interchangeably. Critical to the survival of an organism is the ability to repair the damage caused by toxic insults and inflammation. Hence the inflammatory response to microbes and injured tissues not only serves to eliminate these dangers but also sets into motion the process of repair.

Repair of damaged tissues occurs by two processes: regeneration, which restores normal cells, and scarring, the deposition of connective tissue (Fig. 3.24).

- *Regeneration.* Some tissues are able to replace the damaged components and essentially return to a normal state; this process is called regeneration. Regeneration may occur by proliferation of differentiated cells that survive the injury and retain the capacity to proliferate, notably hepatocytes in the liver. In other tissues, particularly the epithelia of the skin and intestines, tissue stem cells and their progenitors contribute to the restoration of damaged tissues. However, mammals have a limited capacity to regenerate most damaged tissues and organs, and only some components of these tissues are able to fully restore themselves.

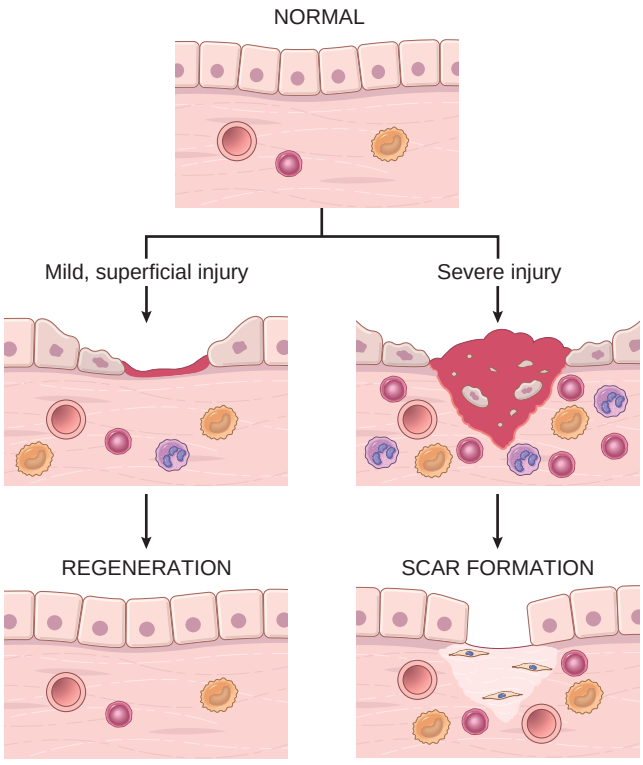


Figure 3.24 Mechanisms of tissue repair: regeneration and scar formation. Following mild injury, which damages the epithelium but not the underlying tissue, resolution occurs by regeneration, but after more severe injury with damage to the connective tissue, repair is by scar formation.

- *Connective tissue deposition (scar formation).* If the injured tissues are incapable of regeneration, or if the supporting structures of the tissue are too severely damaged to support regeneration of the tissue cells, repair occurs by the laying down of connective (fibrous) tissue, a process that may result in scar formation. Although the fibrous scar is not normal, it usually provides enough structural stability that the injured tissue is able to function. The term *fibrosis* is often used to describe the deposition of collagen that occurs in the lungs, liver, kidney, and other organs as a consequence of chronic inflammation or in the myocardium after extensive ischemic necrosis (infarction). If fibrosis develops in a tissue space occupied by an inflammatory exudate, it is called *organization* (as in organizing pneumonia affecting the lung).

After many common types of injury, both regeneration and scar formation contribute in varying degrees to the ultimate repair. Both processes involve the proliferation of cells and close interactions between cells and the ECM. We first discuss the general mechanisms of cellular proliferation and regeneration, then the salient features of healing by scar formation, and we conclude with a description of cutaneous wound healing and fibrosis (scarring) in parenchymal organs as illustrations of the repair process.

Cell and Tissue Regeneration

The regeneration of injured cells and tissues involves cell proliferation, which is driven by growth factors and is critically dependent on the integrity of the ECM, and by the development of mature cells from tissue stem cells. Before describing examples of repair by regeneration, the general principles of cell proliferation are discussed.

Cell Proliferation: Signals and Control Mechanisms

Several cell types proliferate during tissue repair. These include the remnants of the injured tissue (which attempt to restore normal structure), vascular endothelial cells (to create new vessels that provide the nutrients needed for the repair process), and fibroblasts (the source of the fibrous tissue that forms the scar to fill defects that cannot be corrected by regeneration).

The ability of tissues to repair themselves is determined, in part, by their intrinsic proliferative capacity and the presence of tissue stem cells. Based on these criteria, the tissues of the body are divided into three groups.

- *Labile (continuously dividing) tissues.* Cells of these tissues are continuously being lost and replaced by maturation from tissue stem cells and by proliferation of mature cells. Labile cells include hematopoietic cells in the bone marrow and the majority of surface epithelia, such as the stratified squamous epithelia of the skin, oral cavity, vagina, and cervix; the cuboidal epithelia of the ducts draining exocrine organs (e.g., salivary glands, pancreas, biliary tract); the columnar epithelium of the gastrointestinal tract, uterus, and fallopian tubes; and the transitional epithelium of the urinary tract. These tissues can readily regenerate after injury as long as the pool of stem cells is preserved.
- *Stable tissues.* Cells of these tissues are quiescent (in the G_0 stage of the cell cycle) and have only minimal

proliferative activity in their normal state. However, these cells are capable of dividing in response to injury or loss of tissue mass. Stable cells constitute the parenchyma of most solid tissues, such as liver, kidney, and pancreas. They also include endothelial cells, fibroblasts, and smooth muscle cells; the proliferation of these cells is particularly important in wound healing. With the exception of liver, stable tissues have a limited capacity to regenerate after injury.

- *Permanent tissues.* The cells of these tissues are considered to be terminally differentiated and nonproliferative in postnatal life. The majority of neurons and cardiac muscle cells belong to this category. Thus, injury to the brain or heart is irreversible and results in a scar because neurons and cardiac myocytes cannot regenerate. Limited stem cell replication and differentiation occur in some areas of the adult brain, and there is some evidence that heart muscle cells may proliferate after myocardial necrosis. Nevertheless, whatever proliferative capacity may exist in these tissues, it is insufficient to produce tissue regeneration after injury. Skeletal muscle is usually classified as a permanent tissue, but satellite cells attached to the endomysial sheath provide some regenerative capacity for muscle. In permanent tissues, repair is typically dominated by scar formation.

Cell proliferation is driven by signals provided by growth factors and from the ECM. Many different growth factors have been described; some act on multiple cell types, and others are cell-type specific (see Table 1.1 in Chapter 1). Growth factors are typically produced by cells near the site of damage. The most important sources of these growth factors are macrophages that are activated by the tissue injury, but epithelial and stromal cells also produce some of these factors. Several growth factors are displayed at high concentrations bound to ECM proteins. All growth factors activate signaling pathways that stimulate DNA replication (Chapter 1), while also fostering changes in metabolism that promote the biosynthesis of other cellular components (membranes, organelles, proteins) that are needed for a "mother" cell to produce two daughter cells. In addition to responding to growth factors, cells use integrins to bind to ECM proteins, and signals from the integrins can also stimulate cell proliferation.

Mechanisms of Tissue Regeneration

We will consider liver regeneration as a model of tissue regeneration, because it has been studied extensively and illustrates the mechanisms that underlie this process.

Liver Regeneration

The human liver has a remarkable capacity to regenerate, as demonstrated by its growth after partial hepatectomy, which may be performed for tumor resection or for living-donor hepatic transplantation. The mythologic image of liver regeneration is the regrowth of the liver of Prometheus, which was eaten every day by an eagle sent by Zeus as punishment for stealing the secret of fire and grew back overnight. The reality, although less dramatic, is still quite impressive.

Regeneration of the liver occurs by two major mechanisms: proliferation of remaining hepatocytes and repopulation from progenitor cells. Which mechanism plays the dominant role depends on the nature of the injury.

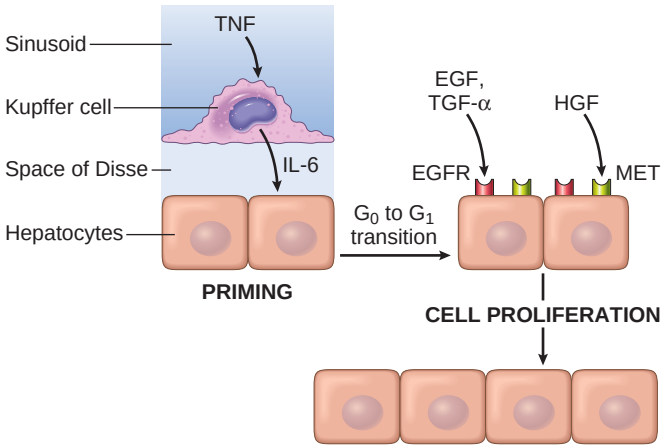


Figure 3.25 Liver regeneration by proliferation of hepatocytes. Following partial hepatectomy, the liver regenerates by proliferation of surviving cells. The process occurs in stages, including priming, followed by growth factor–induced proliferation. The main signals involved in these steps are shown. Once the mass of the liver is restored, the proliferation is terminated (not shown). *EGF*, Epidermal growth factor; *EGFR*, epidermal growth factor receptor; *HGF*, hepatocyte growth factor; *IL-6*, interleukin-6; *TGF-α*, transforming growth factor-α; *TNF*, tumor necrosis factor.

- *Proliferation of hepatocytes following partial hepatectomy.* In humans, resection of up to 90% of the liver can be corrected by proliferation of the residual hepatocytes. This classic model of tissue regeneration has been used experimentally to study the initiation and control of the process.

  Hepatocyte proliferation in the regenerating liver is triggered by the combined actions of cytokines and polypeptide growth factors. The process occurs in distinct stages (Fig. 3.25). In the first, or priming, phase, cytokines such as IL-6 produced mainly by Kupffer cells act on hepatocytes to make the parenchymal cells competent to receive and respond to growth factor signals. In the second, or growth factor, phase, growth factors such as hepatocyte growth factor (HGF) and TGF-α, produced by many cell types, act on primed hepatocytes to stimulate cell metabolism and entry of the cells into the cell cycle. Because hepatocytes are quiescent cells, it takes them several hours to enter the cell cycle, progress from G_0 to G_1, and reach the S phase of DNA replication. Almost all hepatocytes replicate during liver regeneration after partial hepatectomy. During the phase of hepatocyte replication, numerous genes are activated; these include genes encoding transcription factors, cell cycle regulators, regulators of energy metabolism, and others. The wave of hepatocyte proliferation is followed by replication of nonparenchymal cells (Kupffer cells, endothelial cells, and stellate cells). In the final, termination, phase, hepatocytes return to quiescence. The nature of the stop signals is poorly understood; antiproliferative cytokines of the TGF-β family are likely involved.
- *Liver regeneration from progenitor cells.* In situations where the proliferative capacity of hepatocytes is impaired, such as after chronic liver injury or inflammation, progenitor cells in the liver contribute to repopulation. In rodents, these progenitor cells have been called oval cells because

of the shape of their nuclei. Some of these progenitor cells reside in specialized niches called canals of Hering, where bile canaliculi connect with larger bile ducts. The signals that drive proliferation of progenitor cells and their differentiation into mature hepatocytes are topics of active investigation.

Restoration of normal tissue structure can occur only if the residual tissue is structurally intact, as after partial surgical resection. By contrast, if the tissue is damaged by infection or inflammation, regeneration is incomplete and is accompanied by scarring. For example, extensive destruction of the liver with collapse of the reticulin framework, as occurs in a liver abscess, leads to scar formation even though the remaining liver cells have the capacity to regenerate.

KEY CONCEPTS

REPAIR BY REGENERATION

- Tissues are classified as labile, stable, and permanent, according to the proliferative capacity of their cells.
- Continuously dividing tissues (labile tissues) contain stem cells that differentiate to replenish lost cells and maintain tissue homeostasis.
- Cell proliferation is controlled by the cell cycle and is stimulated by growth factors and interactions of cells with the ECM.
- Regeneration of the liver is a classic example of repair by regeneration. It is triggered by cytokines and growth factors produced in response to loss of liver mass and inflammation. In different situations, regeneration may occur by proliferation of surviving hepatocytes or repopulation from progenitor cells.

Repair by Connective Tissue Deposition

If repair cannot be accomplished by regeneration alone, it occurs by replacement of the injured cells with connective tissue, leading to the formation of a scar, or by a combination of regeneration of some residual cells and scar formation. In contrast to regeneration, which involves the restitution of tissue components, scar formation is a response that "patches" rather than restores the tissue. The term scar is most often used in connection to wound healing in the skin, but may also be used to describe the replacement of parenchymal cells in any tissue by collagen, as in the heart after myocardial infarction.

Steps in Scar Formation

Repair by connective tissue deposition consists of sequential processes that follow tissue injury. Within minutes after injury, a hemostatic plug composed of platelets (Chapter 4) is formed, which stops bleeding and provides a scaffold for the deposition of fibrin. The subsequent steps are summarized here (Fig. 3.26):

- *Inflammation.* Breakdown products of complement activation, chemokines released from activated platelets, and other mediators produced at the site of injury function as chemotactic agents to recruit neutrophils and then monocytes over the next 6 to 48 hours. These inflammatory cells eliminate the offending agents, such as microbes that may have entered through the wound, and clear the debris. As the injurious agents and necrotic cells are cleared, the inflammation resolves.
- *Cell proliferation.* In the next stage, which takes up to 10 days, several cell types, including epithelial cells, endothelial and other vascular cells, and fibroblasts, proliferate and migrate to close the now clean wound. Each cell type serves unique functions.
  - *Epithelial cells* respond to locally produced growth factors and migrate over the wound to cover it up.
  - *Endothelial cells and pericytes* proliferate to form new blood vessels, a process known as *angiogenesis*. Because of the importance of this process in physiologic host responses and in many pathologic conditions, it is described in more detail subsequently.
  - *Fibroblasts* proliferate and migrate into the site of injury and lay down collagen fibers that form the scar.
- *Formation of granulation tissue.* Migration and proliferation of fibroblasts and deposition of loose connective tissue,

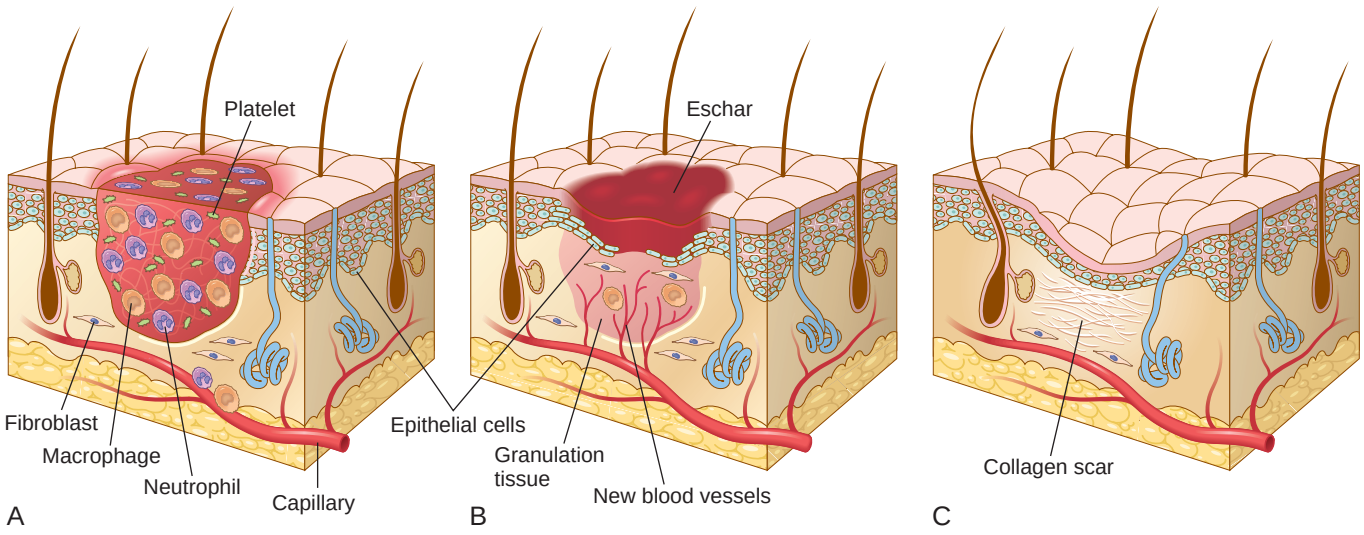


Figure 3.26 Steps in repair by scar formation: wound healing in the skin. (A) Inflammation. (B) Proliferation of epithelial cells; formation of granulation tissue by vessel growth and proliferating fibroblasts. (C) Remodeling to produce the fibrous scar.

together with the vessels and interspersed mononuclear leukocytes, form granulation tissue. The term granulation tissue derives from the pink, soft, granular gross appearance, such as that seen beneath the scab of a skin wound. Its histologic appearance is characterized by proliferation of fibroblasts and new thin-walled, delicate capillaries (angiogenesis) in a loose ECM, often with admixed inflammatory cells, mainly macrophages (Fig. 3.27A). Granulation tissue progressively fills the site of injury; the amount of granulation tissue that is formed depends on the size of the tissue defect created by the wound and the intensity of inflammation.

- *Deposition of connective tissue.* Granulation tissue is progressively replaced by deposition of collagen. The amount of connective tissue increases in the granulation tissue, eventually resulting in the formation of a stable fibrous *scar* (Fig. 3.27B).

Macrophages play a central role in repair by clearing offending agents and dead tissue, providing growth factors for the proliferation of various cells, and secreting cytokines that stimulate fibroblast proliferation and connective tissue synthesis and deposition. The macrophages that are involved in repair are mostly of the alternatively activated (M2) type. It is not clear how the classically activated macrophages that dominate during inflammation, and are involved in getting rid of microbes and dead tissues, are gradually replaced by alternatively activated macrophages that serve to terminate inflammation and induce repair.

We next describe the steps in the formation of granulation tissue and the scar.

Angiogenesis

Angiogenesis is the process of new blood vessel development from existing vessels. It is critical in healing at sites of injury, in the development of collateral circulations at sites of ischemia, and in allowing tumors to increase in size beyond the constraints of their original blood supply. Much work has been done to understand the mechanisms underlying angiogenesis, and therapies to either augment the process (e.g., to improve blood flow to a heart ravaged by coronary atherosclerosis) or inhibit it (to frustrate tumor growth or block pathologic vessel growth such as in diabetic retinopathy) have been developed.

Angiogenesis involves sprouting of new vessels from existing ones and consists of the following steps (Fig. 3.28):

- *Vasodilation* in response to nitric oxide and increased permeability induced by vascular endothelial growth factor (VEGF).
- *Separation of pericytes* from the abluminal surface and breakdown of the basement membrane to allow formation of a vessel sprout.
- *Migration of endothelial cells* toward the area of tissue injury.
- *Proliferation of endothelial cells* just behind the leading front ("tip") of migrating cells.
- Remodeling into capillary tubes.
- *Recruitment of periendothelial cells* (pericytes for small capillaries and smooth muscle cells for larger vessels) to form the mature vessel.
- *Suppression of endothelial proliferation* and migration and deposition of the basement membrane.

Mechanisms of Angiogenesis. The process of angiogenesis involves several signaling pathways, cell–cell interactions, ECM proteins, and tissue enzymes. *VEGFs,* mainly VEGF-A (Chapter 1), stimulate both migration and proliferation of endothelial cells, thus initiating the process of capillary sprouting in angiogenesis. VEGF promotes vasodilation by stimulating the production of NO and contributes to the formation of the vascular lumen. Fibroblast growth factors (FGFs), mainly FGF-2, stimulate the proliferation of endothelial cells. They also promote the migration of macrophages and fibroblasts to the damaged area and stimulate epithelial cell migration to cover epidermal wounds. Angiopoietins 1 and 2 (Ang 1 and Ang 2) are growth factors that play a

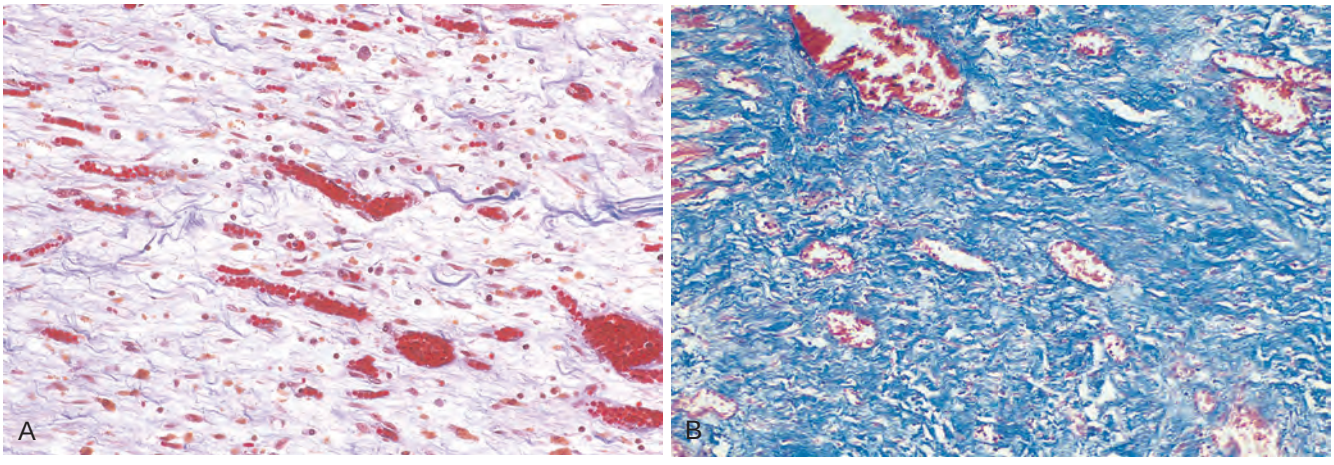


Figure 3.27 (A) Granulation tissue showing numerous blood vessels, edema, and a loose extracellular matrix containing occasional inflammatory cells. Collagen is stained blue by the trichrome stain; minimal mature collagen can be seen at this point. (B) Trichrome stain of mature scar, showing dense collagen, with only scattered vascular channels.

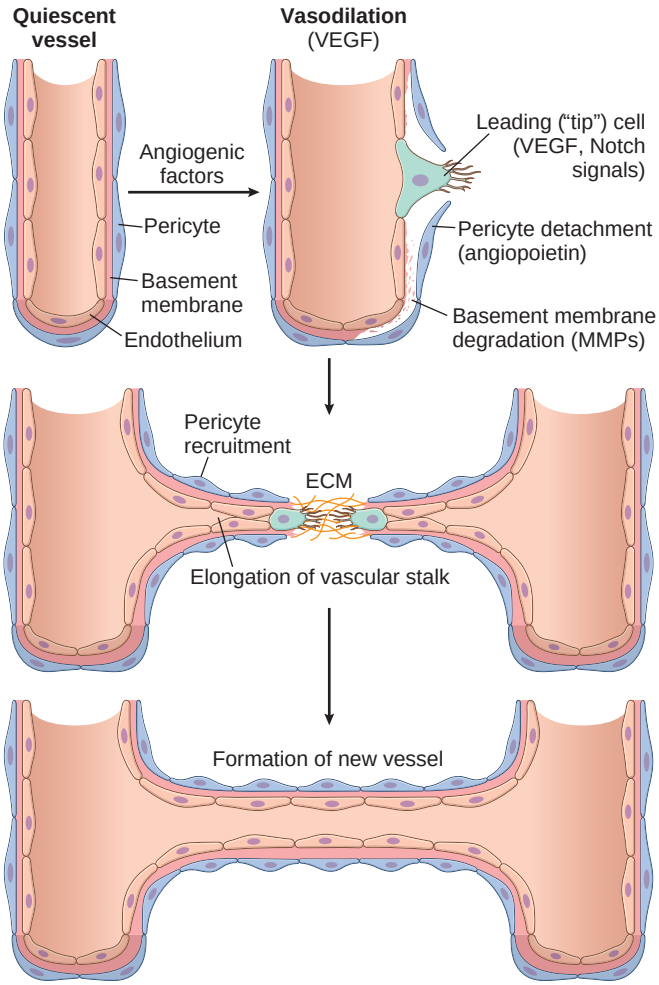


Figure 3.28 Angiogenesis. In tissue repair, angiogenesis occurs mainly by sprouting of new vessels. The steps in the process and the major signals involved are illustrated. The newly formed vessel joins up with other vessels (not shown) to form the new vascular bed. *ECM*, Extracellular matrix; *MMPs*, matrix metalloproteinases; *VEGF*, vascular endothelial growth factor.

role in angiogenesis and the structural maturation of new vessels. Newly formed vessels need to be stabilized by pericytes and smooth muscle cells and by the deposition of connective tissue. The growth factors platelet-derived growth factor (PDGF) and TGF-β also participate in the stabilization process: PDGF recruits smooth muscle cells and TGF-β suppresses endothelial proliferation and migration and enhances the production of ECM proteins.

The Notch signaling pathway regulates the sprouting and branching of new vessels and thus ensures that the new vessels that are formed have the proper spacing to effectively supply the healing tissue with blood. VEGF stimulates the expression of Notch ligands, which bind to the Notch receptor on endothelial cells and regulate the pattern of vessel branching.

ECM proteins participate in the process of vessel sprouting in angiogenesis, largely through interactions with integrin receptors in endothelial cells and by providing the scaffold for vessel growth. Enzymes in the ECM, notably the matrix metalloproteinases (MMPs), degrade the ECM to permit remodeling and extension of the vascular tube.

Deposition of Connective Tissue

The laying down of connective tissue occurs in two steps: migration and proliferation of fibroblasts into the site of injury, and deposition of ECM proteins produced by these cells. These processes are orchestrated by locally produced cytokines and growth factors including PDGF, FGF-2, and TGF-β. The major sources of these factors are inflammatory cells, particularly alternatively activated (M2) macrophages, which are present at sites of injury and in granulation tissue. Sites of inflammation are also rich in mast cells, and in the appropriate chemotactic milieu, lymphocytes may also be present. Each of these can secrete cytokines and growth factors that contribute to fibroblast proliferation and activation.

***TGF-β* is the most important cytokine for the synthesis and deposition of connective tissue proteins.** It is produced by most of the cells in granulation tissue, including alternatively activated macrophages. The levels of TGF-β in tissues are primarily regulated not by the transcription of the gene but by the posttranscriptional activation of latent TGF-β, the rate of secretion of the active molecule, and molecules that make up or that interact with the ECM, notably integrins and microfibrils, that enhance or diminish TGF-β activity. TGF-β stimulates fibroblast migration and proliferation, increases synthesis of collagen and fibronectin, and decreases degradation of ECM by inhibiting metalloproteinases. TGF-β is involved not only in scar formation after injury but also in the development of fibrosis in lung, liver, and kidneys in response to chronic inflammation. TGF-β is also an antiinflammatory cytokine that serves to limit and terminate inflammatory responses. It does this by inhibiting lymphocyte proliferation and the activity of other leukocytes.

As healing progresses, the number of proliferating fibroblasts and new vessels decreases; however, the fibroblasts progressively assume a more synthetic phenotype, and hence there is increased deposition of ECM. Collagen deposition is critical for the development of strength in a healing wound site. As the scar matures, there is progressive vascular regression, which eventually transforms the highly vascularized granulation tissue into a pale, largely avascular scar. Some of the fibroblasts also acquire features of smooth muscle cells, including the presence of actin filaments, and are called *myofibroblasts*. These cells contribute to the contraction of the scar over time.

Remodeling of Connective Tissue

The outcome of the repair process is influenced by a balance between synthesis and degradation of ECM proteins. After its deposition, the connective tissue in the scar continues to be modified and remodeled. The degradation of collagens and other ECM components is accomplished by a family of matrix metalloproteinases (MMPs), so called because they are dependent on metal ions (e.g., zinc) for their activity. MMPs include interstitial collagenases (MMP-1, MMP-2, and MMP-3), which cleave fibrillar collagen; gelatinases,

which degrade amorphous collagen and fibronectin; and stromelysins, which degrade a variety of ECM constituents, including proteoglycans, laminin, fibronectin, and amorphous collagen.

MMPs are produced by a variety of cell types (fibroblasts, macrophages, neutrophils, synovial cells, and some epithelial cells), and their synthesis and secretion are regulated by growth factors, cytokines, and other agents. The activity of the MMPs is tightly controlled. They are produced as inactive precursors (zymogens) that must be first activated; this is accomplished by proteases (e.g., plasmin) likely to be present only at sites of injury. In addition, activated collagenases can be rapidly inhibited by specific tissue inhibitors of metalloproteinases (TIMPs), produced by most mesenchymal cells. Thus, during scar formation, MMPs are activated to remodel the deposited ECM, and then their activity is shut down by the TIMPs.

KEY CONCEPTS

REPAIR BY SCAR FORMATION

- Tissues are repaired by replacement with connective tissue and scar formation if the injured tissue is not capable of proliferation or if the structural framework is damaged and cannot support regeneration.
- The main components of connective tissue repair are angiogenesis, migration and proliferation of fibroblasts, collagen synthesis, and connective tissue remodeling.
- Repair by connective tissue starts with the formation of granulation tissue and culminates in the laying down of fibrous tissue.
- Multiple growth factors stimulate the proliferation of the cell types involved in repair.
- TGF-β is a potent fibrogenic agent; ECM deposition depends on the balance between fibrogenic agents, metalloproteinases (MMPs) that digest ECM, and TIMPs.

Factors That Influence Tissue Repair

Tissue repair may be altered by several factors, which impact the quality or adequacy of the reparative process. Variables that modify healing may be extrinsic (e.g., infection) or intrinsic to the injured tissue and systemic or local:

- *Infection* is clinically one of the most important causes of delayed healing; it prolongs inflammation and potentially increases the local tissue injury.
- *Diabetes* is a metabolic disease that compromises tissue repair for many reasons (Chapter 24) and is one of the most important systemic causes of abnormal wound healing.
- *Nutritional status* has profound effects on repair; protein deficiency and vitamin C deficiency inhibit collagen synthesis and retard healing.
- *Glucocorticoids (steroids)* have well-documented antiinflammatory effects, and their administration may result in weakness of the scar due to inhibition of TGF-β production and diminished fibrosis. In some instances, however, these effects of glucocorticoids are desirable. For example, in corneal infections, glucocorticoids are sometimes prescribed (along with antibiotics) to reduce the likelihood of opacity that may result from scarring.
- *Mechanical factors* such as increased local pressure or torsion may cause wounds to pull apart, or dehisce.
- *Poor perfusion,* due to peripheral vascular disease, arteriosclerosis, and diabetes or due to obstructed venous drainage (e.g., in varicose veins), also impairs healing.
- *Foreign bodies* such as fragments of steel, glass, or even bone impede healing by perpetuating chronic inflammation.
- *The type and extent of tissue injury* and the character of the tissue in which the injury occurs affect the subsequent repair. Complete restoration can occur only in tissues composed of stable and labile cells. Injury to tissues composed of permanent cells inevitably results in scarring and some loss of function.
- The *location of the injury* is also important. For example, inflammation arising in tissue spaces (e.g., pleural, peritoneal, synovial cavities) develops extensive exudates. Subsequent repair may occur by digestion of the exudate, initiated by the proteolytic enzymes of leukocytes, and resorption of the liquefied exudate. This is called resolution, and in the absence of cellular necrosis, normal tissue architecture is generally restored. However, in the setting of larger accumulations, granulation tissue grows into the exudate, and a fibrous scar ultimately forms. This is called organization.

Examples of Tissue Repair and Fibrosis

So far, we have discussed the general principles and mechanisms of repair by regeneration and scar formation. In this section we describe two clinically significant types of repair—the healing of skin wounds (cutaneous wound healing) and fibrosis in injured parenchymal organs.

Healing of Skin Wounds

Based on the nature and size of the wound, the healing of skin wounds is said to occur by first or second intention.

Healing by First Intention

When the injury involves only the epithelial layer, the principal mechanism of repair is epithelial regeneration, also called primary union or healing by first intention. One of the simplest examples of this type of wound repair is the healing of a clean, uninfected surgical incision approximated by surgical sutures (Fig. 3.29). The incision causes only focal disruption of epithelial basement membrane continuity and death of relatively few epithelial and connective tissue cells. The repair consists of the same three connected processes that we have described previously: inflammation, proliferation of epithelial and other cells, and maturation of the connective tissue scar.

- Wounding causes the rapid activation of coagulation pathways, which results in the formation of a blood clot on the wound surface (Chapter 4). The clot serves to stop bleeding and supports migrating cells, which are attracted by growth factors, cytokines, and chemokines released into the area.
- Within 24 hours, neutrophils are seen at the incision margin, migrating toward the fibrin clot. Basal cells at

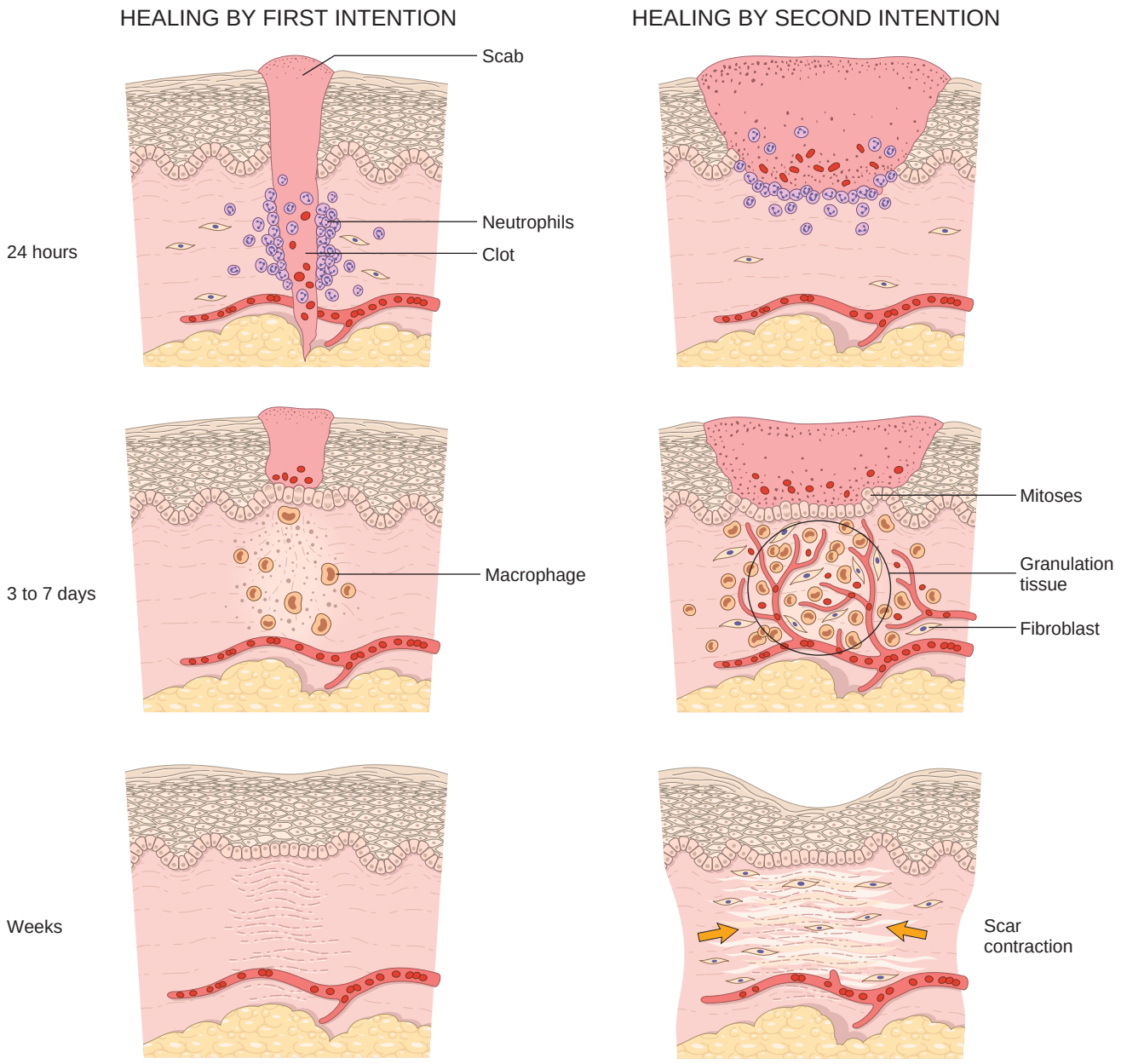


Figure 3.29 Steps in wound healing by first intention (left) and second intention (right). In the latter, note the large amount of granulation tissue and wound contraction.

the edge of the incision begin to show increased mitotic activity. Within 24 to 48 hours, epithelial cells from both edges have begun to migrate and proliferate along the dermis yielding a thin but continuous epithelial layer that closes the wound.

- By day 3, neutrophils have been largely replaced by macrophages, and granulation tissue progressively invades the incision space. The macrophages clear extracellular debris, fibrin, and other foreign material and promote angiogenesis and ECM deposition.
- By day 5, neovascularization reaches its peak as granulation tissue fills the incisional space. The new vessels are leaky, allowing the passage of plasma proteins and fluid into the extravascular space. Thus, new granulation tissue is often edematous. Fibroblasts progressively migrate into the granulation tissue, where they proliferate and lay down collagen and ECM.
- During the second week, there is continued collagen accumulation and fibroblast proliferation. The leukocyte infiltrate, edema, and increased vascularity are substantially diminished.
- By the end of the first month, the scar comprises a cellular connective tissue largely devoid of inflammatory cells and covered by an essentially normal epidermis. The tensile strength of the wound increases with time, as described later.

Healing by Second Intention

Healing by second intention, also called secondary union, differs from primary healing in several respects.

- In wounds causing large tissue deficits, the fibrin clot is larger, and there is more exudate and necrotic debris in the wounded area. Inflammation is more intense because large tissue defects have a greater volume of necrotic debris, exudate, and fibrin that must be removed. Consequently, large defects have a greater potential for secondary, inflammation-mediated injury.
- Much larger amounts of granulation tissue are formed to fill a bigger gap caused by the larger area of deficit. A greater volume of granulation tissue generally results in a greater mass of scar tissue.
- At first a provisional matrix containing fibrin, plasma fibronectin, and type III collagen is formed, but in about 2 weeks this is replaced by a matrix composed primarily of type I collagen. Ultimately the original granulation tissue scaffold is converted into a pale, avascular scar. The dermal appendages that have been destroyed in the line of the incision are permanently lost. By the end of the first month, the scar is made up of acellular connective tissue devoid of inflammatory infiltrate, covered by intact epidermis.
- Wound contraction helps to close the wound by decreasing the gap between its dermal edges and by reducing the wound surface area. Hence it is an important feature in healing by secondary union. Wound contraction involves the formation of a network of *myofibroblasts,* which are modified fibroblasts which have contractile properties. Within 6 weeks, large skin defects may be reduced to 5% to 10% of their original size, largely by contraction.

Wound Strength

Carefully sutured wounds have approximately 70% of the strength of normal skin, largely because of the placement of sutures. The recovery of tensile strength results from the excess of collagen synthesis over collagen degradation during the first 2 months of healing by cross-linking of collagen fibers and increased fiber size. Wound strength reaches approximately 70% to 80% of normal by 3 months but usually does not substantially improve beyond that point.

Fibrosis in Parenchymal Organs

The term fibrosis is used to denote the excessive deposition of collagen and other ECM components in a tissue. As already mentioned, the terms scar and fibrosis are used interchangeably. The basic mechanisms of fibrosis are the same as those of scar formation in the skin during tissue repair. Fibrosis may be responsible for substantial organ dysfunction and even organ failure.

Fibrotic disorders include diverse chronic and debilitating diseases such as liver cirrhosis, systemic sclerosis (scleroderma), fibrosing diseases of the lung (idiopathic pulmonary fibrosis, pneumoconioses, and drug- and radiation-induced pulmonary fibrosis), end-stage kidney disease, and constrictive pericarditis. These conditions are discussed in the appropriate chapters throughout the book. Because of the tremendous functional impairment caused by fibrosis in these conditions, there is great interest in the development of antifibrotic drugs.

Abnormalities in Tissue Repair

Complications in tissue repair can arise from abnormalities in any of the basic components of the process, including deficient scar formation, excessive formation of the repair components, and formation of contractures.

Defects in Healing: Chronic Wounds

These are seen in numerous clinical situations, as a result of local and systemic factors. The following are some common examples.

- *Venous leg ulcers* (Fig. 3.30A) develop most often in elderly people as a result of chronic venous hypertension, which may be caused by severe varicose veins or congestive heart failure. Deposits of iron pigment (hemosiderin) are common, resulting from red cell breakdown, and there may be accompanying chronic inflammation. These ulcers fail to heal because of poor delivery of oxygen to the site of the ulcer.
- *Arterial ulcers* (Fig. 3.30B) develop in individuals with atherosclerosis of peripheral arteries, especially associated with diabetes. The ischemia results in atrophy and then necrosis of the skin and underlying tissues. These lesions can be quite painful.
- *Diabetic ulcers* (Fig. 3.30C) affect the lower extremities, particularly the feet. There is tissue necrosis and failure to heal as a result of vascular disease causing ischemia, neuropathy, systemic metabolic abnormalities, and secondary infections. Histologically, these lesions are characterized by epithelial ulceration (Fig. 3.30E) and extensive granulation tissue in the underlying dermis (Fig. 3.30F).
- *Pressure sores* (Fig. 3.30D) are areas of skin ulceration and necrosis of underlying tissues caused by prolonged compression of tissues against a bone, e.g., in elderly patients with numerous morbidities lying in bed without moving. The lesions are caused by mechanical pressure and local ischemia.

When a surgical incision reopens internally or externally it is called wound dehiscence. The risk factors for such an occurrence are obesity, malnutrition, infections, and vascular insufficiency. In abdominal wounds it can be precipitated by vomiting and coughing.

Excessive Scarring

Excessive formation of the components of the repair process can give rise to hypertrophic scars and keloids. The accumulation of excessive amounts of collagen may give rise to a raised scar known as a hypertrophic scar. These often grow rapidly and contain abundant myofibroblasts, but they tend to regress over several months (Fig. 3.31A). If the scar tissue grows beyond the boundaries of the original wound and does not regress, it is called a *keloid* (Fig. 3.31B, C). Keloid formation seems to be an individual predisposition, and for unknown reasons it is somewhat more common in African Americans. Hypertrophic scars generally develop after thermal or traumatic injury that involves the deep layers of the dermis.

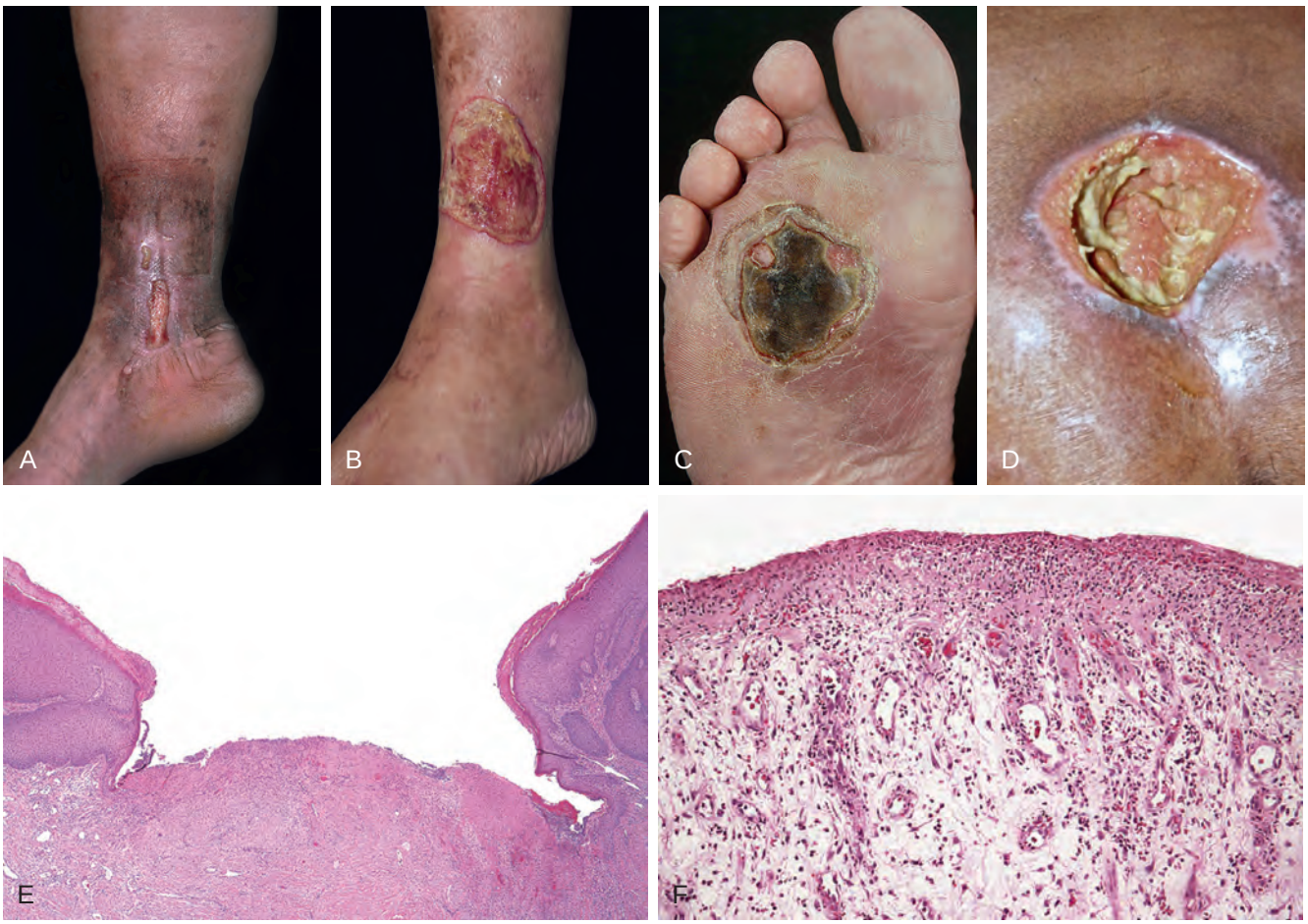


Figure 3.30 Chronic wounds illustrating defects in wound healing. (A–D) External appearance of skin ulcers. (A) Venous leg ulcer; (B) arterial ulcer, with more extensive tissue necrosis; (C) diabetic ulcer; (D) pressure sore. (E, F) Histologic appearance of a diabetic ulcer. (E) Ulcer crater; (F) chronic inflammation and granulation tissue. (A–D, From Eming SA, Margin P, Tomic-Canic M: Wound repair and regeneration: mechanisms, signaling, and translation, *Sci Transl Med* 6:265, 2014.)

Exuberant granulation is another deviation in wound healing consisting of the formation of excessive amounts of granulation tissue, which protrudes above the level of the surrounding skin and blocks reepithelialization (this process has been called, with more literary fervor, proud flesh). Excessive granulation must be removed by cautery or surgical excision to permit restoration of the continuity of the epithelium. Rarely, incisional scars or traumatic injuries may be followed by exuberant proliferation of fibroblasts and other connective tissue elements that may recur after excision. Called desmoids, or aggressive fibromatoses, these neoplasms lie in the interface between benign and malignant (though low-grade) tumors.

Contraction in the size of a wound is an important part of the normal healing process. An exaggeration of this process gives rise to *contracture* and results in deformities of the wound and the surrounding tissues. Contractures are particularly prone to develop on the palms, the soles, and the anterior aspect of the thorax. Contractures are commonly seen after serious burns and can compromise the movement of joints.

KEY CONCEPTS

CUTANEOUS WOUND HEALING AND PATHOLOGIC ASPECTS OF REPAIR

- The main phases of cutaneous wound healing are inflammation, formation of granulation tissue, and ECM remodeling.
- Cutaneous wounds can heal by primary union (first intention) or secondary union (secondary intention); secondary healing involves more extensive scarring and wound contraction.
- Wound healing can be altered by many conditions, particularly infection and diabetes; the type, volume, and location of the injury are important factors that influence the healing process.
- Excessive production of ECM can cause keloids in the skin.
- Persistent stimulation of collagen synthesis in chronic inflammatory diseases leads to fibrosis of the tissue, often with extensive loss of the tissue and functional impairment.

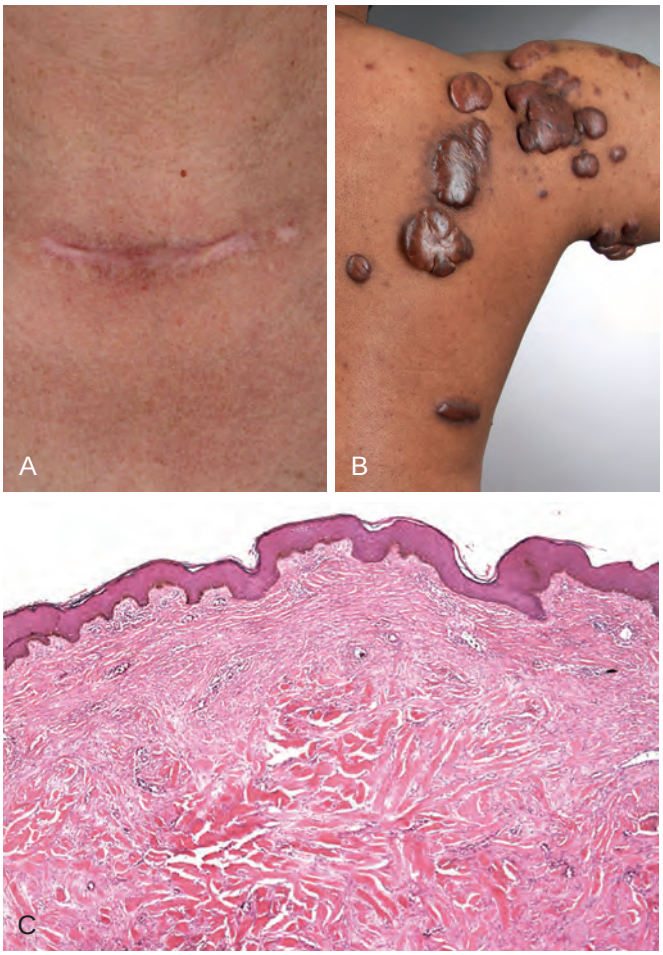


Figure 3.31 Clinical examples of excessive scarring and collagen deposition. (A) Hypertrophic scar. (B) Keloid. (C) Microscopic appearance of a keloid. Note the thick connective tissue deposition in the dermis. (A–B, From Eming SA, Margin P, Tomic-Canic M: Wound repair and regeneration: mechanisms, signaling, and translation, *Sci Transl Med* 6:265, 2014; C, Courtesy Z. Argenyi, MD, University of Washington, Seattle, Wash.)

SUGGESTED READINGS

General Mechanisms of Inflammation

Kotas ME, Medzhitov R: Homeostasis, inflammation, and disease susceptibility, *Cell* 160:816–827, 2015. *[A conceptual discussion of physiologic homeostasis and how it may be disturbed, accounting for the association between inflammation and many of the diseases of the modern world].*

Rock KL, Latz E, Ontiveros F, et al: The sterile inflammatory response, *Annu Rev Immunol* 28:321–342, 2010. *[An excellent discussion of how the immune system recognizes and responds to necrotic cells and other noninfectious harmful agents].*

Sensors of Inflammatory Stimuli

Manthiram K, Zhou Q, Aksentijevich I, et al: The monogenic autoinflammatory diseases define new pathways in human innate immunity and inflammation, *Nat Immunol* 18:832–842, 2017. *[A comprehensive review of diseases caused by mutations in inflammasome-related pathways].*

Rathinam VA, Fitzgerald KA: Inflammasome complexes: emerging mechanisms and effector functions, *Cell* 165:792–800, 2016. *[An up-to-date review of inflammasomes and their roles in inflammation].*

Takeuchi O, Akira S: Pattern recognition receptors and inflammation, *Cell* 140:805, 2010. *[An excellent overview of Toll-like receptors and other pattern recognition receptor families, as well as their roles in host defense and inflammation].*

Acute Inflammation: Vascular Reactions

Alitalo K: The lymphatic vasculature in disease, *Nat Med* 17:1371–1380, 2011. *[An excellent review of the cell biology of lymphatic vessels; their functions in immune and inflammatory reactions; and their roles in inflammatory, neoplastic, and other diseases].*

Vestweber D: Relevance of endothelial junctions in leukocyte extravasation and vascular permeability, *Ann N Y Acad Sci* 1257:184–192, 2012. *[A good review of the basic processes of vascular permeability and how interendothelial junctions are regulated].*

Acute Inflammation: Role of Neutrophils

Castanheira FVS, Kubes P: Neutrophils and NETs in modulating acute and chronic inflammation, *Blood* 133:2178–2185, 2019. *[A discussion of the role of neutrophil extracellular traps in tissue injury and resolution].*

Flannagan RS, Jaumouillé V, Grinstein S: The cell biology of phagocytosis, *Annu Rev Pathol* 7:61–98, 2012. *[A modern discussion of the receptors involved in phagocytosis, the molecular control of the process, and the biology and functions of phagosomes].*

Kolaczkowska E, Kubes P: Neutrophil recruitment and function in health and inflammation, *Nat Rev Immunol* 13:159–175, 2013. *[An excellent review of neutrophil generation, recruitment, functions and fates, and their roles in different types of inflammatory reactions].*

Kourtzelis I, Mitroulis I, von Renesse J, et al: From leukocyte recruitment to resolution of inflammation: the cardinal role of integrins, *J Leukoc Biol* 102:677–683, 2017. *[A discussion of the diverse functions of integrins].*

Mayadas TN, Cullere X, Lowell CA: The multifaceted functions of neutrophils, *Annu Rev Pathol* 9:181, 2014. *[An excellent review of neutrophil biology].*

McEver RP: Selectins: initiators of leukocyte adhesion and signaling at the vascular wall, *Cardiovasc Res* 107:331–339, 2015. *[An overview of the roles of selectins in leukocyte recruitment].*

Muller WA: Mechanisms of leukocyte transendothelial migration, *Annu Rev Pathol* 6:323, 2011. *[A thoughtful review of the mechanisms by which leukocytes traverse the endothelium].*

Schmidt S, Moser M, Sperandio M: The molecular basis of leukocyte recruitment and its deficiencies, *Mol Immunol* 55:49–58, 2013. *[A review of the mechanisms of leukocyte recruitment and leukocyte adhesion deficiencies].*

Sollberger G, Tilley DO, Zychlinsky A: Neutrophil extracellular traps: the biology of chromatin externalization, *Dev Cell* 44:542–553, 2018. *[A review of the mechanisms of neutrophil extracellular trap formation].*

Mediators of Inflammation

Chavan SS, Pavlov VA, Tracey KJ: Mechanisms and therapeutic relevance of neuro-immune communication, *Immunity* 46:927–942, 2017. *[A thoughtful review that explores the mechanisms by which neural reflexes may influence inflammatory reactions].*

Dennis EA, Norris PC: Eicosanoid storm in infection and inflammation, *Nat Rev Immunol* 15:511, 2015. *[A review of the proinflammatory and antiinflammatory activities of eicosanoids].*

Di Gennaro A, Haeggström JZ: The leukotrienes: immune-modulating lipid mediators of disease, *Adv Immunol* 116:51–92, 2012. *[A comprehensive review of the biochemistry of leukotrienes and their receptors, as well as the roles of these mediators in various diseases].*

Griffith JW, Sokol CL, Luster AD: Chemokines and chemokine receptors: positioning cells for host defense and immunity, *Annu Rev Immunol* 32:659–702, 2014. *[An overview of the functions of chemokines in host defense and inflammation].*

Holers VM: Complement and its receptors: new insights into human disease, *Annu Rev Immunol* 32:433–459, 2015.

Nathan C, Cunningham-Bussel A: Beyond oxidative stress: an immunologist's guide to reactive oxygen species, *Nat Rev Immunol* 13:349–361,

2013. *[An excellent modern review of the production, catabolism, targets, and actions of reactive oxygen species, as well as their roles in inflammation].*

Ricklin D, Lambris JD: Complement in immune and inflammatory disorders, *J Immunol* 190:3831–3839, 2013. *[Two companion articles on the biochemistry and biology of the complement system and the development of therapeutic agents to alter complement activity in disease].*

Zlotnik A, Yoshie O: The chemokine superfamily revisited, *Immunity* 36:705–716, 2012. *[An excellent update on the classification, functions, and clinical relevance of chemokines and their receptors].*

Chronic Inflammation: Role of Macrophages and Other Cells

Nagy JA, Dvorak AM, Dvorak HF: VEGF-A and the induction of pathological angiogenesis, *Annu Rev Pathol* 2:251–275, 2007. *[A review of the vascular endothelial growth factor family of growth factors and their role in angiogenesis in cancer, inflammation, and various disease states].*

Nathan C, Ding A: Nonresolving inflammation, *Cell* 140:871–882, 2010. *[A discussion of the abnormalities that lead to chronic inflammation].*

Sica A, Mantovani A: Macrophage plasticity and polarization: in vivo veritas, *J Clin Invest* 122:787–795, 2012. *[An excellent review of macrophage subpopulations; their generation; and their roles in inflammation, infections, cancer, and metabolic disorders].*

Sepsis

Aziz M, Jacob A, Yang WL, et al: Current trends in inflammatory and immunomodulatory mediators in sepsis, *J Leukoc Biol* 93:329–342, 2013. *[A comprehensive review of the roles of cytokines and other mediators in the development and resolution of the systemic inflammatory response syndrome].*

Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, et al: The pathogenesis of sepsis, *Annu Rev Pathol* 6:19–48, 2011. *[An excellent review of the pathogenesis of sepsis, focusing on the cells involved, the importance of cell death, the links between inflammation and coagulation, the utility of biomarkers, and therapeutic approaches].*

Tissue Repair: Regeneration and Fibrosis

Duffield JS, Lupher M, Thannickal VJ, et al: Host responses in tissue repair and fibrosis, *Annu Rev Pathol* 8:241–276, 2013. *[An overview of the cellular mechanisms of fibrosis, with emphasis on the role of the immune system in fibrotic reactions to chronic infections].*

Eming SA, Martin P, Tomic-Canic M: Wound repair and regeneration: mechanisms, signaling, and translation, *Sci Transl Med* 6:265r6, 2014. *[A modern review of host responses that contribute to tissue repair].*

Friedman SL, Sheppard D, Duffield JS, et al: Therapy for fibrotic diseases: nearing the starting line, *Sci Transl Med* 5:167sr1, 2013. *[An excellent review of the current concepts of the pathogenesis of fibrosis, emphasizing the roles of different cell populations and the extracellular matrix and the potential for translating basic knowledge to the development of new therapies].*

Klingberg F, Hinz B, White ES: The myofibroblast matrix: implications for tissue repair and fibrosis, *J Pathol* 229:298–309, 2013. *[An excellent review of the properties of myofibroblasts and extracellular matrix proteins and their roles in tissue repair and fibrosis].*

Mantovani A, Biswas SK, Galdiero MR, et al: Macrophage plasticity and polarization in tissue repair and remodelling, *J Pathol* 229:176–185, 2013. *[An excellent review comparing macrophage subsets and the role of these cells in the resolution of inflammation and tissue repair].*

Preziosi ME, Monga SP: Update on the mechanisms of liver regeneration, *Semin Liver Dis* 37:141–151, 2017. *[A modern discussion of the biochemical mechanisms of liver regeneration].*

Murawala P, Tanaka EM, Currie JD: Regeneration: the ultimate example of wound healing, *Semin Cell Dev Biol* 23:954–962, 2012. *[A thoughtful discussion of the process of regeneration in amphibians and how it is different in mammals].*

Novak ML, Koh TJ: Macrophage phenotypes during tissue repair, *J Leukoc Biol* 93:875–881, 2013. *[A review of macrophage phenotypes and how they change during inflammation and tissue repair, emphasizing the concept that these phenotypes are plastic and dynamic and may not represent committed lineages].*

Wick G, Grundtman C, Mayerl C, et al: The immunology of fibrosis, *Annu Rev Immunol* 31:107–135, 2013. *[An excellent review of the role of the immune system in fibrosis during tissue repair and in chronic diseases including autoimmune disorders, atherosclerosis, fibrotic disorders of different organs, and cancer].*

Wynn TA, Ramalingam TR: Mechanisms of fibrosis: therapeutic translation for fibrotic disease, *Nat Med* 18:1028–1040, 2012. *[An excellent review of the mechanisms of tissue fibrosis and the antifibrotic therapeutic strategies being developed].*

Wynn TA, Vannella KM: Macrophages in tissue repair, regeneration, and fibrosis, *Immunity* 44:450–462, 2016. *[An excellent review of the multifaceted roles of macrophage subsets in inflammation and repair].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 4

Hemodynamic Disorders, Thromboembolic Disease, and Shock

CHAPTER CONTENTS

The health of cells and tissues depends on the circulation of blood, which delivers oxygen and nutrients and removes wastes generated by cellular metabolism. Under normal conditions, as blood passes through capillary beds, proteins in the plasma are retained within the vasculature, and there is little net movement of water and electrolytes into the tissues. This balance is often disturbed by pathologic conditions that alter endothelial function, increase vascular hydrostatic pressure, or decrease plasma protein content, all of which promote edema—the accumulation of fluid in tissues resulting from a net movement of water into extravascular spaces. Depending on its severity and location, edema may have minimal or profound effects. In the lower extremities, it may only make one's shoes feel snugger after a long sedentary day; in the lungs, however, edema fluid can fill alveoli, causing life-threatening hypoxia.

The structural integrity of blood vessels is frequently compromised by trauma. *Hemostasis* is the process of blood clotting that prevents excessive bleeding after blood-vessel damage. Inadequate hemostasis may result in hemorrhage, which can compromise regional tissue perfusion and, if massive and rapid, may lead to hypotension, shock, and death. Conversely, inappropriate clotting *(thrombosis)* or migration of clots *(embolism)* can obstruct blood vessels, potentially causing ischemic cell death *(infarction)*. Indeed, thromboembolism lies at the heart of three major causes of morbidity and death in high income countries: myocardial infarction, pulmonary embolism (PE), and cerebrovascular accident (stroke).

Herein, we focus on disorders of hemodynamics (edema, effusions, congestion, and shock), provide an overview of disorders of abnormal bleeding and clotting (thrombosis), and discuss the various forms of embolism.

EDEMA AND EFFUSIONS

Disorders that perturb cardiovascular, renal, or hepatic function are often marked by the accumulation of fluid in tissues (edema) or body cavities (effusions). Under normal circumstances, the tendency of vascular hydrostatic pressure to push water and salts out of capillaries into the interstitial space is nearly balanced by the tendency of plasma colloid osmotic pressure to pull water and salts back into vessels. There is usually a small net movement of fluid into the interstitium, but this drains into lymphatic vessels and ultimately returns to the bloodstream via the thoracic duct, keeping the tissues "dry" (Fig. 4.1). **Elevated hydrostatic pressure or diminished colloid osmotic pressure disrupts this balance and results in increased movement of fluid out of vessels.** If the net rate of fluid movement exceeds the rate of lymphatic drainage, fluid accumulates. Within tissues the result is edema, and if a serosal surface is involved, fluid may accumulate within the adjacent body cavity as an effusion.

Edema fluids and effusions may be *inflammatory* or *noninflammatory* (Table 4.1). Inflammation-related edema and effusions are discussed in detail in Chapter 3. These

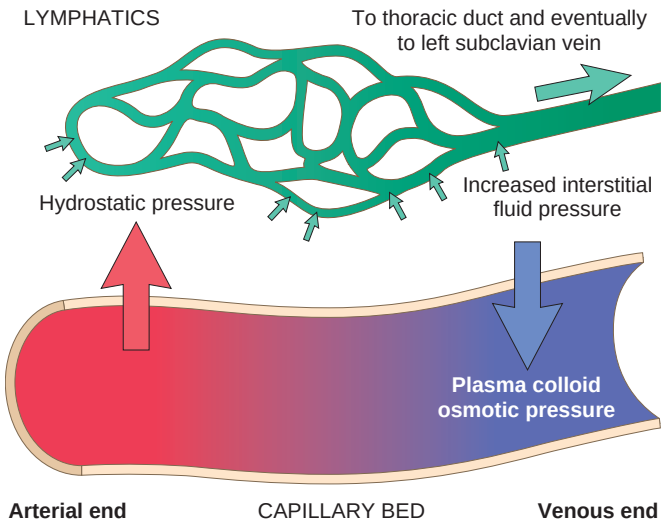


Figure 4.1 Factors influencing fluid movement across capillary walls. Normally, hydrostatic and osmotic forces are nearly balanced so that there is little net movement of fluid out of vessels. Many different pathologic disorders (see Table 4.1) are associated with increases in capillary hydrostatic pressure or decreases in plasma osmotic pressure that lead to the extravasation of fluid into tissues. Lymphatic vessels remove much of the excess fluid, but if the capacity for lymphatic drainage is exceeded, tissue edema results.

protein-rich *exudates* accumulate due to increases in vascular permeability caused by inflammatory mediators. In contrast, noninflammatory edema and effusions are protein-poor fluids called *transudates*. Noninflammatory edema and effusions are common in many disorders, including heart failure, liver failure, renal disease, and malnutrition (Fig. 4.2). We will now discuss the various causes of edema.

Increased Hydrostatic Pressure

Increases in hydrostatic pressure are mainly caused by disorders that impair venous return. If the impairment is localized (e.g., a deep venous thrombosis [DVT] in a lower extremity), then the resulting edema is confined to the affected part. Conditions leading to systemic increases in venous pressure (e.g., congestive heart failure, Chapter 12) are understandably associated with more widespread edema.

Reduced Plasma Osmotic Pressure

Under normal circumstances albumin accounts for almost half of the total plasma protein; it follows that conditions leading to inadequate synthesis or increased loss of albumin from the circulation are common causes of reduced plasma oncotic pressure. Reduced albumin synthesis occurs mainly in severe liver diseases (e.g., end-stage cirrhosis, Chapter 18) and protein malnutrition (Chapter 9). An important cause of albumin loss is the *nephrotic syndrome* (Chapter 20), in which albumin leaks into the urine through abnormally permeable glomerular capillaries. Regardless of cause, reduced plasma osmotic pressure leads in a stepwise fashion to edema, reduced intravascular volume, renal hypoperfusion, and secondary hyperaldosteronism. Not only does the ensuing salt and water retention by the kidney fail to correct the plasma volume deficit, but it also exacerbates the edema, because the primary defect—a low plasma protein level—persists.

Sodium and Water Retention

Increased salt retention—with obligate retention of associated water—causes both increased hydrostatic pressure (due to intravascular fluid volume expansion) and diminished vascular colloid osmotic pressure (due to dilution). Salt retention occurs whenever renal function is compromised, such as in primary kidney disorders and in cardiovascular disorders that decrease renal perfusion. One of the most important causes of renal hypoperfusion is congestive heart failure, which (like hypoproteinemia) results in the activation of the renin-angiotensin-aldosterone axis. In early heart failure, this response is beneficial, as the retention of sodium and water and other adaptations, including increased vascular tone and elevated levels of antidiuretic hormone, improve cardiac output and restore normal renal perfusion. However, as heart failure worsens and cardiac output diminishes, the retained fluid merely increases the hydrostatic pressure, leading to edema and effusions.

Lymphatic Obstruction

Trauma, fibrosis, invasive tumors, and infectious agents can all disrupt lymphatic vessels and impair the clearance of interstitial fluid, resulting in lymphedema in the affected part of the body. A dramatic example is seen in parasitic *filariasis,* in which the organism induces obstructive fibrosis

Table 4.1 Pathophysiologic Categories of Edema

Increased Hydrostatic Pressure
Impaired Venous Return
Congestive heart failure
Constrictive pericarditis
Ascites (liver cirrhosis)
Venous obstruction or compression
Thrombosis
External pressure (e.g., mass)
Lower extremity inactivity with prolonged dependency
Arteriolar Dilation
Heat
Neurohumoral dysregulation
Reduced Plasma Osmotic Pressure (Hypoproteinemia)
Protein-losing glomerulopathies (nephrotic syndrome)
Liver cirrhosis
Malnutrition
Protein-losing gastroenteropathy
Lymphatic Obstruction
Inflammatory
Neoplastic
Postsurgical
Postirradiation
Sodium Retention
Excessive salt intake with renal insufficiency
Increased tubular reabsorption of sodium
Renal hypoperfusion
Increased renin-angiotensin-aldosterone secretion
Inflammation
Acute inflammation
Chronic inflammation
Angiogenesis

Modified from Leaf A, Cotran RS: *Renal Pathophysiology,* ed 3, New York, 1985, Oxford University Press, p 146.

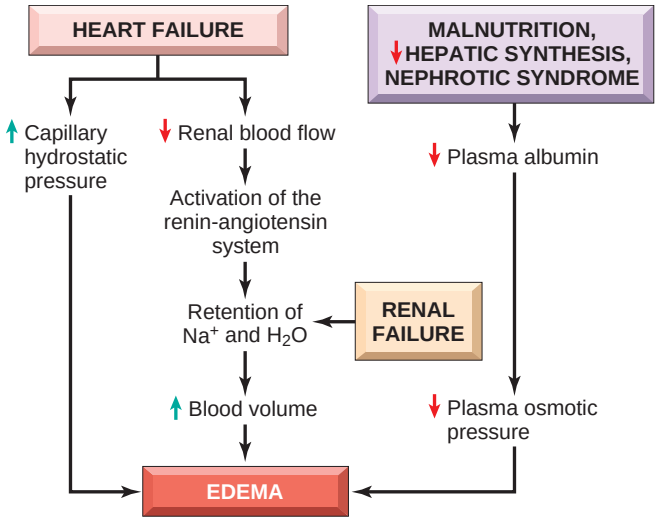


Figure 4.2 Mechanisms of systemic edema in heart failure, renal failure, malnutrition, hepatic failure, and nephrotic syndrome.

of lymphatic channels and lymph nodes. This may result in edema of the external genitalia and lower limbs that is so massive as to earn the appellation *elephantiasis*. Severe edema of the upper extremity may also complicate surgical removal and/or irradiation of the breast and associated axillary lymph nodes in patients with breast cancer.

MORPHOLOGY

Edema is easily recognized grossly; microscopically, it is appreciated as clearing and separation of the extracellular matrix (ECM) and subtle cell swelling. Edema is most commonly seen in subcutaneous tissues, the lungs, and the brain. **Subcutaneous edema** can be diffuse or more conspicuous in regions with high hydrostatic pressures. Its distribution is often influenced by gravity (e.g., it appears in the legs when standing and the sacrum when recumbent), a feature termed **dependent edema.** Finger pressure over markedly edematous subcutaneous tissue displaces the interstitial fluid and leaves a depression, a sign called **pitting edema.**

Edema resulting from **renal dysfunction** often appears initially in parts of the body containing loose connective tissue, such as the eyelids; **periorbital edema** is thus a characteristic finding in severe renal disease. With **pulmonary edema,** the lungs are often two to three times their normal weight, and sectioning yields frothy, blood-tinged fluid—a mixture of air, edema, and extravasated red cells. **Brain edema** can be localized or generalized depending on the nature and extent of the pathologic process or injury. The swollen brain exhibits narrowed sulci and distended gyri, which are compressed by the unyielding skull (Chapter 28).

Effusions involving the pleural cavity **(hydrothorax),** the pericardial cavity **(hydropericardium),** or the peritoneal cavity (**hydroperitoneum** or **ascites**) are common in a wide range of clinical settings. Transudative effusions are typically protein-poor, translucent, and straw colored; an exception are peritoneal effusions caused by lymphatic blockage (chylous effusion), which may be milky due to the presence of lipids absorbed from the gut. In contrast, exudative effusions are protein-rich and often cloudy due to the presence of white cells.

Clinical Features

The consequences of edema range from merely annoying to rapidly fatal. Subcutaneous edema is important primarily because it signals potential underlying cardiac or renal disease; however, when significant, it can also impair wound healing and the clearance of infections. Pulmonary edema is a common clinical problem that is most frequently seen in the setting of left ventricular failure; it can also occur with renal failure, acute respiratory distress syndrome (Chapter 15), and pulmonary inflammation or infection. Edema in the pulmonary interstitium and the alveolar spaces impedes gas exchange (leading to hypoxemia) and also creates a favorable environment for bacterial infection. Pulmonary edema is often exacerbated by pleural effusions, which may further compromise gas exchange by compressing the underlying pulmonary parenchyma. Peritoneal effusions (ascites) resulting most commonly from portal hypertension are prone to seeding by bacteria, leading to serious and sometimes fatal infections. Brain edema is life threatening; if severe, brain substance can herniate (extrude) through the foramen magnum, or the brain stem vascular supply can be compressed. Either condition can injure the medullary centers and cause death (Chapter 28).

KEY CONCEPTS

EDEMA

Edema is the result of the movement of fluid from the vasculature into the interstitial spaces; the fluid may be protein-poor (transudate) or protein-rich (exudate).

Edema may be caused by:

- Increased hydrostatic pressure (e.g., heart failure)
- Decreased colloid osmotic pressure caused by reduced plasma albumin, either due to decreased synthesis (e.g., liver disease, protein malnutrition) or to increased loss (e.g., nephrotic syndrome)
- Increased vascular permeability (e.g., inflammation)
- Lymphatic obstruction (e.g., infection or neoplasia)
- Sodium and water retention (e.g., renal failure)

HYPEREMIA AND CONGESTION

Hyperemia and congestion both stem from increased blood volumes within tissues, but have different underlying mechanisms and consequences. Hyperemia is an active process in which arteriolar dilation (e.g., at sites of inflammation or in skeletal muscle during exercise) leads to increased blood flow. Affected tissues turn red *(erythema)* because of increased delivery of oxygenated blood. Congestion is a passive process resulting from reduced venous outflow of blood from a tissue. It can be systemic, as in cardiac failure, or localized, as in isolated venous obstruction. Congested tissues have an abnormal blue-red color *(cyanosis)* that stems from the accumulation of deoxygenated hemoglobin in the affected area. In long-standing *chronic passive congestion,* the associated chronic hypoxia may result in ischemic tissue injury and scarring. In chronically congested tissues, capillary rupture can also produce small hemorrhagic foci; subsequent catabolism of extravasated red cells can

leave residual telltale clusters of hemosiderin-laden macrophages. As a result of increased hydrostatic pressures, congestion commonly leads to edema.

MORPHOLOGY

Congested tissues take on a dusky reddish-blue color *(cyanosis)* due to red cell stasis and the presence of deoxygenated hemoglobin. Microscopically, **acute pulmonary congestion** is marked by engorged alveolar capillaries, alveolar septal edema, and focal intra-alveolar hemorrhage. In **chronic pulmonary congestion,** which is often caused by congestive heart failure, the septa are thickened and fibrotic, and the alveoli often contain numerous macrophages laden with hemosiderin **(heart failure cells)** derived from phagocytosed red cells. In **acute hepatic congestion,** the central vein and sinusoids are distended. Because the centrilobular area is at the distal end of the hepatic blood supply, centrilobular hepatocytes may undergo ischemic necrosis, and the periportal hepatocytes—better oxygenated because of proximity to hepatic arterioles—may only develop fatty change. In **chronic passive hepatic congestion,** the centrilobular regions are grossly red-brown and slightly depressed (because of cell death) and are accentuated against the surrounding zones of uncongested tan liver **(nutmeg liver)** (Fig. 4.3A). Microscopically, there is centrilobular congestion and hemorrhage, hemosiderin-laden macrophages, and variable degrees of hepatocyte dropout and necrosis (see Fig. 4.3*B*).

HEMOSTASIS, HEMORRHAGIC DISORDERS, AND THROMBOSIS

Hemostasis can be defined simply as the process by which blood clots form at sites of vascular injury. Hemostasis is essential for life and is deranged to varying degrees in a broad range of disorders, which can be divided into two groups. In hemorrhagic disorders, characterized by excessive bleeding, hemostatic mechanisms are either blunted or insufficient to prevent blood loss. By contrast, in thrombotic disorders blood clots (often referred to as thrombi) form within intact blood vessels or within the chambers of the heart. As is discussed in Chapters 11 and 12, thrombosis has a central role in the most common and clinically important forms of cardiovascular disease.

Although useful, it must be recognized that this division between bleeding and thrombotic disorders sometimes breaks down, in that generalized activation of clotting sometimes paradoxically produces bleeding due to the consumption of coagulation factors, as in *disseminated intravascular coagulation (DIC)*. To provide context for understanding disorders of bleeding and clotting, this discussion begins with normal hemostasis, focusing on the contribution of platelets, coagulation factors, and endothelium.

Normal Hemostasis

Hemostasis is a precisely orchestrated process involving platelets, clotting factors, and endothelium that occurs at the site of vascular injury and culminates in the formation of a blood clot, which serves to prevent or limit the extent of bleeding. The general sequence of events leading to hemostasis at a site of vascular injury is shown in Fig. 4.4.

- *Arteriolar vasoconstriction* occurs immediately and markedly reduces blood flow to the injured area (see Fig. 4.4A). It is mediated by reflex neurogenic mechanisms and may be augmented by the local secretion of factors such as *endothelin,* a potent endothelium-derived vasoconstrictor. This effect is transient, however, and bleeding would resume if not for activation of platelets and coagulation factors.
- *Primary hemostasis: the formation of the platelet plug*. Disruption of the endothelium exposes subendothelial von Willebrand factor (vWF) and collagen, which promote platelet adherence and activation. Activation of platelets results in a dramatic shape change (from small rounded discs to flat plates with spiky protrusions that markedly increase surface area), as well as the release of secretory granules. Within a few minutes, the secreted products recruit additional platelets that undergo aggregation to form a primary hemostatic plug (see Fig. 4.4B).
- *Secondary hemostasis: deposition of fibrin.* Vascular injury exposes *tissue factor* at the site of injury. Tissue factor is

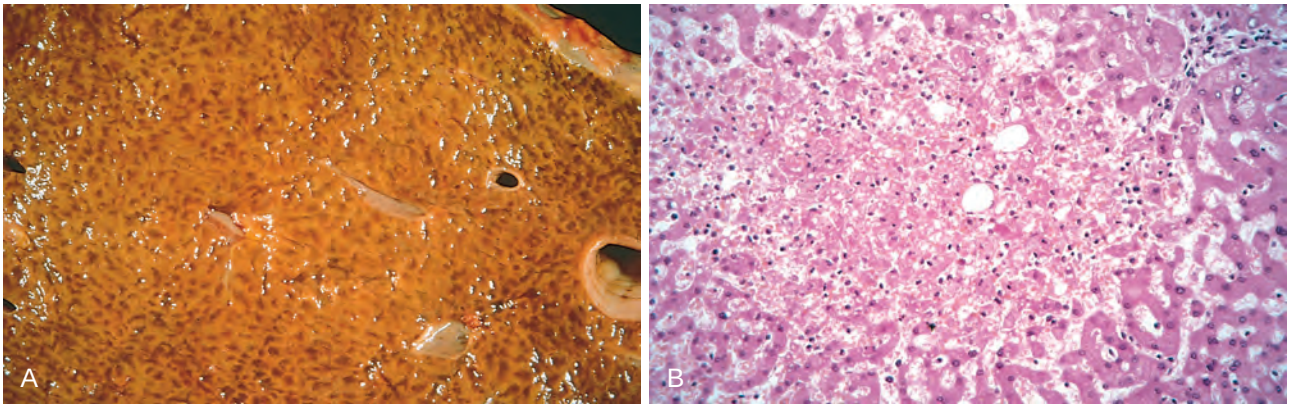
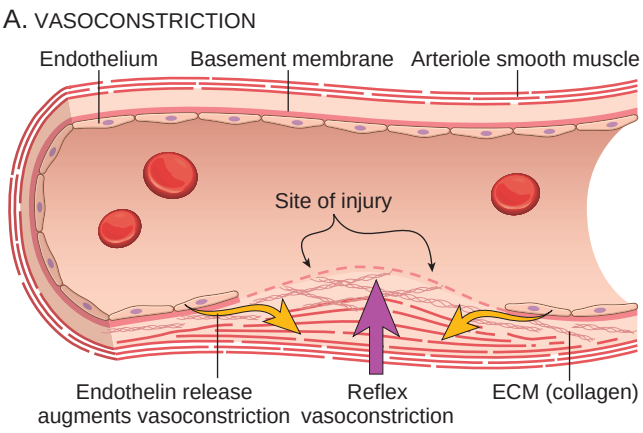
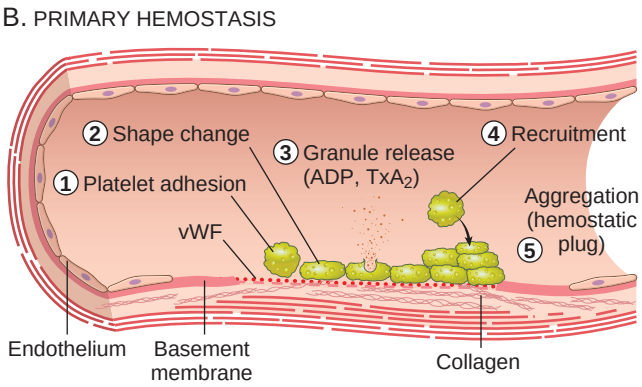
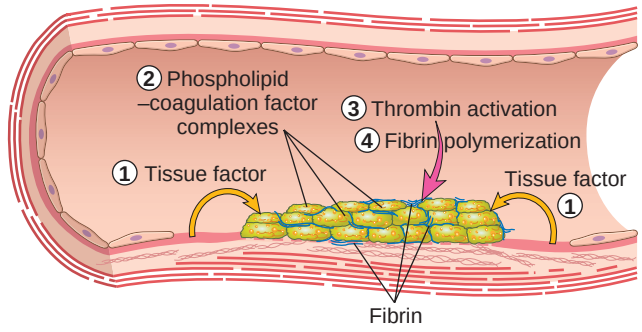


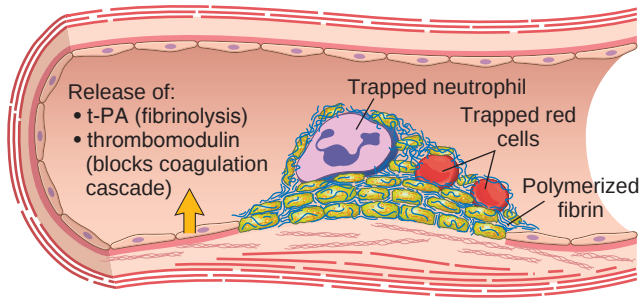
Figure 4.3 Liver with chronic passive congestion and hemorrhagic necrosis. (A) Central areas are red and slightly depressed compared with the surrounding tan viable parenchyma, forming a "nutmeg liver" pattern (so-called because it resembles the cut surface of a nutmeg). (B) Centrilobular necrosis with degenerating hepatocytes and hemorrhage. (Courtesy Dr. James Crawford, Department of Pathology, University of Florida, Gainesville, Fla.)

Figure 4.4 Normal hemostasis. (A) After vascular injury, neurohumoral factors induce transient vasoconstriction. (B) Platelets bind via glycoprotein Ib *(GpIb)* receptors to von Willebrand factor *(vWF)* on exposed extracellular matrix *(ECM)* and are activated, undergoing a shape change and granule release. Released adenosine diphosphate *(ADP)* and thromboxane A_2 *(TxA_2)* induce additional platelet aggregation through platelet GpIIb-IIIa receptor binding to fibrinogen, and form the primary hemostatic plug. (C) Local activation of the coagulation cascade (involving tissue factor and platelet phospholipids) results in fibrin polymerization, "cementing" the platelets into a definitive secondary hemostatic plug. (D) Counterregulatory mechanisms, mediated by tissue plasminogen activator (*t-PA*, a fibrinolytic product) and thrombomodulin, confine the hemostatic process to the site of injury.









a membrane-bound procoagulant glycoprotein that is normally expressed by subendothelial cells in the vessel wall, such as smooth muscle cells and fibroblasts. Tissue factor binds and activates factor VII (see later), setting in motion a cascade of reactions that culminates in *thrombin* generation. Thrombin cleaves circulating fibrinogen into insoluble *fibrin,* creating a fibrin meshwork, and also is a potent activator of platelets, leading to additional platelet aggregation at the site of injury. This sequence, referred to as *secondary hemostasis,* consolidates the initial platelet plug (see Fig. 4.4C).

- *Clot stabilization and resorption*. Polymerized fibrin and platelet aggregates undergo contraction to form a solid, permanent plug that prevents further hemorrhage. At this stage, counterregulatory mechanisms (e.g., tissue plasminogen activator [t-PA] made by endothelial cells) are set into motion that limit clotting to the site of injury (see Fig. 4.4D) and eventually lead to clot resorption and tissue repair.

It should be emphasized that endothelial cells are central regulators of hemostasis; the balance between the antithrombic and prothrombotic activities of endothelium determines whether thrombus formation, propagation, or dissolution occur. Normal endothelial cells express a variety of *anticoagulant* factors that inhibit platelet aggregation and coagulation and promote fibrinolysis; after injury or activation, however, this balance shifts, and endothelial cells acquire numerous *procoagulant* activities (activation of platelets and clotting factor, described earlier; see also Fig. 4.10). Besides trauma, endothelium can be activated by microbial pathogens, hemodynamic forces, and pro-inflammatory mediators.

We now describe roles of platelets, coagulation factors, and endothelium in hemostasis in greater detail, following the scheme illustrated in Fig. 4.4.

Platelets

Platelets play a critical role in hemostasis by forming the primary plug that initially seals vascular defects and by providing a surface that binds and concentrates activated coagulation factors. Platelets are disc-shaped anucleate cell fragments that are shed from megakaryocytes in the bone marrow into the bloodstream. Their function depends on several glycoprotein receptors, a contractile cytoskeleton, and two types of cytoplasmic granules. α-*Granules* have the adhesion molecule P-selectin on their membranes (Chapter 3)

and contain proteins involved in coagulation, such as fibrinogen, coagulation factor V, and vWF, as well as protein factors that may be involved in wound healing, such as fibronectin, platelet factor 4 (PF4, a heparin-binding chemokine), platelet-derived growth factor (PDGF), and transforming growth factor-β. *Dense* (or δ) *granules* contain adenosine diphosphate (ADP), adenosine triphosphate, ionized calcium, serotonin, and epinephrine.

After a traumatic vascular injury, platelets encounter constituents of the subendothelial connective tissue, such as vWF and collagen. On contact with these proteins, platelets undergo a sequence of reactions that culminate in the formation of a platelet plug (see Fig. 4.4B).

- *Platelet adhesion* is mediated largely via interactions between the platelet surface receptor glycoprotein Ib (GpIb) and vWF in the subendothelial matrix (Fig. 4.5). Platelets also adhere to exposed collagen via the platelet collagen receptor Gp1a/IIa. Notably, genetic deficiencies of vWF (von Willebrand disease, Chapter 14) or GpIb (Bernard-Soulier syndrome) result in bleeding disorders, attesting to the importance of these factors.
- *Platelets rapidly change shape* following adhesion, being converted from smooth discs to spiky "sea urchins" with greatly increased surface area. This change is accompanied by conformational changes in cell surface *glycoprotein IIb/IIIa* that increase its affinity for fibrinogen and by the translocation of *negatively charged phospholipids* (particularly phosphatidylserine) to the platelet surface. These phospholipids bind calcium and serve as nucleation sites for the assembly of coagulation factor complexes.
- *Secretion (release reaction) of granule contents* occurs along with changes in shape; these two events are often referred to together as platelet activation. Platelet activation is triggered by a number of factors, including the coagulation factor thrombin and ADP. Thrombin activates platelets through a special G-protein–coupled receptor referred to as *protease-activated receptor-1* (PAR-1), which is switched on by a proteolytic cleavage carried out by thrombin. ADP is a component of dense-body granules; thus, platelet activation and ADP release begets additional rounds of platelet activation, a phenomenon referred to as *recruitment*. ADP acts by binding two G-protein–coupled receptors. $P2Y_1$ and $P2Y_{12}$. Activated platelets also produce the prostaglandin *thromboxane* A_2 (TxA_2), a potent inducer of platelet aggregation. An understanding of the biochemical pathways involved in platelet activation has led to the development of drugs with antiplatelet activities. Aspirin, the oldest of all, inhibits platelet aggregation and produces a mild bleeding defect by inhibiting cyclooxygenase, a platelet enzyme that is required for TxA_2 synthesis. More recently, drugs that inhibit platelet function by antagonizing PAR-1 or $P2Y_{12}$ have been developed. All of these antiplatelet drugs are used in the treatment of coronary artery disease.
- *Platelet aggregation* follows their activation. The conformational change in glycoprotein IIb/IIIa that occurs with platelet activation allows binding of fibrinogen, a large bivalent plasma polypeptide that forms bridges between adjacent platelets, leading to their aggregation. Predictably, inherited deficiency of GpIIb-IIIa results in a bleeding disorder called *Glanzmann thrombasthenia*. The initial wave of aggregation is reversible, but concurrent activation of thrombin stabilizes the platelet plug by causing further platelet activation and aggregation, and by promoting irreversible platelet contraction. Platelet contraction is dependent on the cytoskeleton and consolidates the aggregated platelets. In parallel, thrombin also converts fibrinogen into insoluble *fibrin*, cementing the platelets in place and creating the definitive *secondary hemostatic plug*. Entrapped red cells and leukocytes are also found in hemostatic plugs, in part due to adherence of leukocytes to P-selectin expressed on activated platelets.

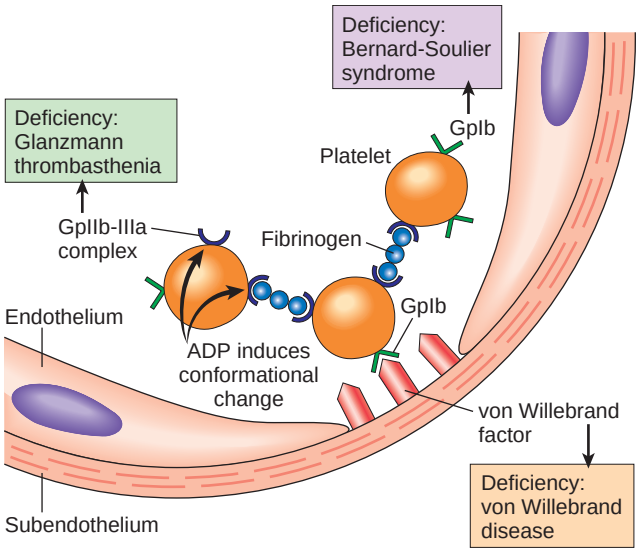


Figure 4.5 Platelet adhesion and aggregation. Von Willebrand factor functions as an adhesion bridge between subendothelial collagen and the glycoprotein Ib *(GpIb)* platelet receptor. Aggregation is accomplished by fibrinogen bridging GpIIb-IIIa receptors on different platelets. Congenital deficiencies in the various receptors or bridging molecules lead to the diseases indicated in the colored boxes. *ADP*, Adenosine diphosphate.

KEY CONCEPTS

PLATELET ADHESION, ACTIVATION, AND AGGREGATION

- Endothelial injury exposes the underlying basement membrane ECM; platelets adhere to the ECM primarily through the binding of platelet GpIb receptors to VWF.
- Adhesion leads to platelet activation, an event associated with secretion of platelet granule contents, including calcium (a cofactor for several coagulation proteins) and ADP (a mediator of further platelet activation); dramatic changes in shape and membrane composition; and activation of GpIIb/IIIa receptors.
- The GpIIb/IIIa receptors on activated platelets form bridging cross-links with fibrinogen, leading to platelet aggregation.
- Concomitant activation of thrombin promotes fibrin deposition, cementing the platelet plug in place.

Coagulation Cascade

The coagulation cascade is a series of amplifying enzymatic reactions that lead to the deposition of an insoluble fibrin clot. As discussed later, the dependency of clot formation on various factors differs in the laboratory test tube and in

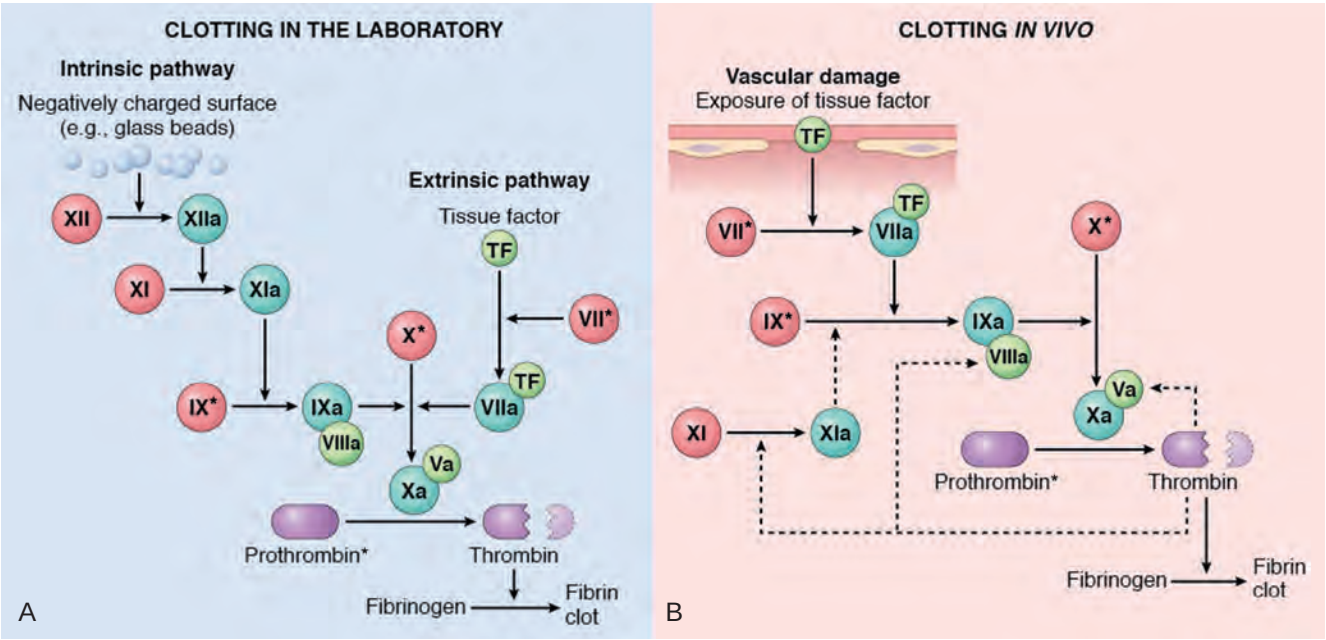


Figure 4.6 The coagulation cascade in the laboratory and in vivo. (A) Clotting is initiated in the laboratory by adding phospholipids, calcium, and either a negatively charged substance such as glass beads (intrinsic pathway) or a source of tissue factor (extrinsic pathway). (B) In vivo, tissue factor is the major initiator of coagulation, which is amplified by feedback loops involving thrombin (dotted lines). The red polypeptides are inactive factors, the dark green polypeptides are active factors, and the light green polypeptides correspond to cofactors (reaction accelerators). Factors marked with an asterisk (*) are vitamin K dependent as are protein C and S (not depicted). Warfarin acts as an anticoagulant by inhibiting the γ-carboxylation of the vitamin K–dependent coagulation factors. Vitamin K is an essential cofactor for the synthesis of all of these vitamin K–dependent clotting factors.

blood vessels in vivo (Fig. 4.6). However, clotting in vitro and in vivo both follow the same general principles, as follows.

The cascade of reactions in the pathway can be likened to a "dance" in which coagulation factors are passed from one partner to the next (Fig. 4.7). Each reaction step involves an enzyme (an activated coagulation factor), a substrate (an inactive proenzyme form of a coagulation factor), and a cofactor (a reaction accelerator). These components are assembled on a negatively charged phospholipid surface, which is provided by activated platelets. Assembly of

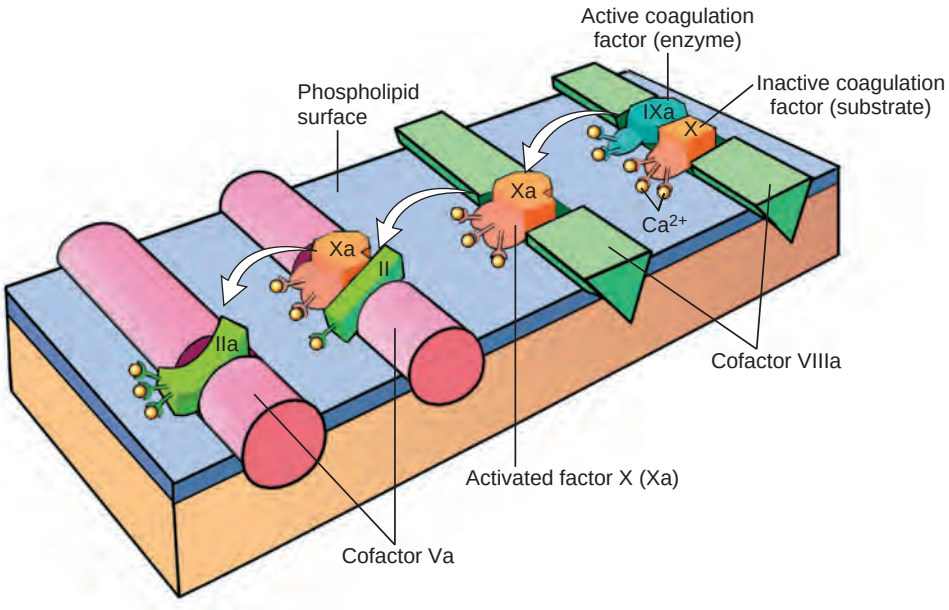


Figure 4.7 Schematic illustration of the conversion of factor X to factor Xa, which in turn converts factor II (prothrombin) to factor IIa (thrombin) (see Fig. 4.6B). The initial reaction complex consists of a proteolytic enzyme (factor IXa), a substrate (factor X), and a reaction accelerator (factor VIIIa), all assembled on a platelet phospholipid surface. Calcium ions hold the assembled components together and are essential for the reaction. Activated factor Xa becomes the protease for the second adjacent complex in the coagulation cascade, converting prothrombin substrate (II) to thrombin (IIa) using factor Va as the reaction accelerator.

reaction complexes also depends on calcium, which binds to γ-carboxylated glutamic acid residues that are present in factors II, VII, IX, and X. The enzymatic reactions that produce γ-carboxylated glutamic acid use vitamin K as a cofactor and are antagonized by drugs such as coumadin, used as an anticoagulant.

Based on assays performed in clinical laboratories, the coagulation cascade can be divided into the *extrinsic* and *intrinsic* pathways (see Fig. 4.6A).

- The *prothrombin time* (PT) assay assesses the function of the proteins in the extrinsic pathway (factors VII, X, V, II [prothrombin], and fibrinogen). In brief, tissue factor, phospholipids, and calcium are added to plasma, and the time for a fibrin clot to form is recorded.
- The *partial thromboplastin time* (PTT) assay screens the function of the proteins in the intrinsic pathway (factors XII, XI, IX, VIII, X, V, II, and fibrinogen). In this assay, clotting of plasma is initiated by the addition of negatively charged particles (e.g., ground glass) that activate factor XII (Hageman factor) together with phospholipids and calcium, and the time to fibrin clot formation is recorded.

Although the PT and PTT assays are of great utility in evaluating coagulation factor function in patients, they do not recapitulate the events that lead to coagulation in vivo. This point is most clearly made by considering the clinical effects of deficiencies of various coagulation factors. Deficiencies of factors V, VII, VIII, IX, and X are associated with moderate to severe bleeding disorders, and prothrombin deficiency is likely incompatible with life. In contrast, factor XI deficiency is only associated with mild bleeding, and individuals with factor XII deficiency do not bleed and in fact may be susceptible to thrombosis. By contrast, there is evidence from experimental models suggesting that in some circumstances factor XII may contribute to thrombosis. These paradoxical findings may stem from involvement of factor XII in several pathways, including the pro-inflammatory bradykinin pathway as well as fibrinolysis (discussed later).

Based on the effects of various factor deficiencies in humans, it is believed that, in vivo, factor VIIa/tissue factor complex is the most important activator of factor IX and that factor IXa/factor VIIIa complex is the most important activator of factor X (see Fig. 4.6B). The mild bleeding tendency seen in patients with factor XI deficiency is likely explained by the ability of thrombin to activate factor XI (as well as factors V and VIII), a feedback mechanism that amplifies the coagulation cascade.

Among the coagulation factors, thrombin is the most important, because its various enzymatic activities control diverse aspects of hemostasis and link clotting to inflammation and repair. Among thrombin's most important activities are the following:

- *Conversion of fibrinogen into cross-linked fibrin.* Thrombin directly converts soluble fibrinogen into fibrin monomers that polymerize into an insoluble fibril, and also amplifies the coagulation process, not only by activating factor XI, but also by activating two critical cofactors: factors V and VIII. It also stabilizes the secondary hemostatic plug by activating factor XIII, which covalently cross-links fibrin.
- *Platelet activation.* Thrombin is a potent inducer of platelet activation and aggregation through its ability to activate PAR-1, thereby linking platelet function to coagulation.

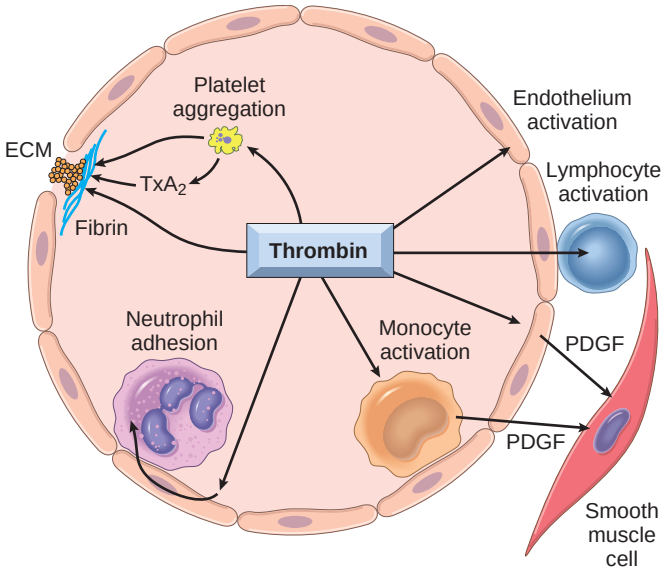


Figure 4.8 Role of thrombin in hemostasis and cellular activation. Thrombin plays a critical role in generating cross-linked fibrin (by cleaving fibrinogen to fibrin and by activating factor XIII), as well as activating several other coagulation factors (see Fig. 4.6B). Through protease-activated receptors (PARs, see text), thrombin also modulates several cellular activities. It directly induces platelet aggregation and TxA_2 production, and activates endothelial cells, which respond by expressing adhesion molecules and cytokine mediators (e.g., *PDGF*). Thrombin also directly activates leukocytes. *ECM*, Extracellular matrix; *PDGF*, platelet-derived growth factor. See Fig. 4.10 for additional anticoagulant activities mediated by thrombin. (Courtesy Shaun Coughlin, MD, PhD, Cardiovascular Research Institute, University of California at San Francisco; modified with permission.)

- *Pro-inflammatory effects.* PARs also are expressed on inflammatory cells, endothelium, and other cell types (Fig. 4.8), and activation of these receptors by thrombin is believed to mediate pro-inflammatory effects that contribute to tissue repair and angiogenesis.
- *Anticoagulant effects.* Remarkably, through mechanisms described later, on encountering normal endothelium, thrombin changes from a procoagulant to an anticoagulant; this reversal in function prevents clots from extending beyond the site of the vascular injury.

Factors That Limit Coagulation

Once initiated, coagulation must be restricted to the site of vascular injury to prevent deleterious consequences. One limiting factor is simple dilution; blood flowing past the site of injury washes out activated coagulation factors, which are rapidly removed by the liver. A second is the requirement for negatively charged phospholipids, which, as mentioned, are mainly provided by platelets that have been activated by contact with subendothelial matrix at sites of vascular injury. However, the most important counter-regulatory mechanisms involve factors that are expressed by intact endothelium adjacent to the site of injury (described later).

Activation of the coagulation cascade also sets into motion a fibrinolytic cascade that limits the size of the clot and contributes to its later dissolution (Fig. 4.9). Fibrinolysis is largely accomplished through the enzymatic activity of

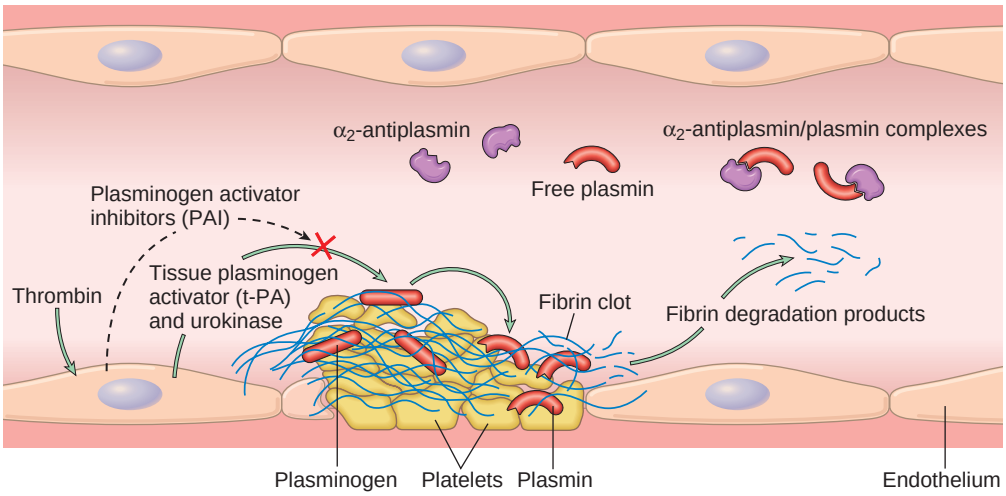


Figure 4.9 The fibrinolytic system, illustrating various plasminogen activators and inhibitors (see text).

plasmin, which breaks down fibrin and interferes with its polymerization. An elevated level of breakdown products of fibrinogen (often called fibrin split products), most notably fibrin-derived *D-dimers,* are a useful clinical marker of several thrombotic states (described later). Plasmin is generated by enzymatic catabolism of the inactive circulating precursor *plasminogen*, either by a factor XII–dependent pathway (possibly explaining the association of factor XII deficiency and thrombosis) or by plasminogen activators. The most important plasminogen activator is t-PA; it is synthesized principally by endothelium and is most active when bound to fibrin. This characteristic makes t-PA a useful therapeutic agent, since its fibrinolytic activity is largely confined to sites of recent thrombosis. Once activated, plasmin is in turn tightly controlled by counterregulatory factors such as α_2-plasmin inhibitor, a plasma protein that binds and rapidly inhibits free plasmin.

KEY CONCEPTS

COAGULATION FACTORS

- Coagulation occurs via the sequential enzymatic conversion of a cascade of circulating and locally synthesized proteins.
- Tissue factor elaborated at sites of injury is the most important initiator of the coagulation cascade in vivo.
- At the final stage of coagulation, thrombin converts fibrinogen into insoluble fibrin that contributes to formation of the definitive hemostatic plug.
- Coagulation normally is restricted to sites of vascular injury by:
  - Limiting enzymatic activation to phospholipid surfaces provided by activated platelets or endothelium
  - Circulating inhibitors of coagulation factors, such as antithrombin III, whose activity is augmented by heparin-like molecules expressed on endothelial cells
  - Expression of thrombomodulin on normal endothelial cells, which binds thrombin and converts it to an anticoagulant
  - Activation of fibrinolytic pathways (e.g., by association of t-PA with fibrin)

Endothelium

The balance between the anticoagulant and procoagulant activities of endothelium often determines whether clot formation, propagation, or dissolution occurs (Fig. 4.10). Normal endothelial cells express a multitude of factors that inhibit the procoagulant activities of platelets and coagulation factors and that augment fibrinolysis. These factors act in concert to prevent thrombosis and to limit clotting to sites of vascular damage. However, if injured or exposed to proinflammatory factors, endothelial cells lose many of their antithrombotic properties. Here, we complete the discussion of hemostasis by focusing on the antithrombotic activities of normal endothelium; we return to the "dark side" of endothelial cells later when discussing thrombosis.

The antithrombotic properties of endothelium can be divided into activities directed at platelets, coagulation factors, and fibrinolysis.

- *Platelet inhibitory effects.* An obvious effect of intact endothelium is to serve as a barrier that shields platelets from subendothelial vWF and collagen. However, normal endothelium also releases a number of factors that inhibit platelet activation and aggregation. Among the most important are *prostacyclin (PGI_2), nitric oxide (NO),* and *adenosine diphosphatase*; the latter degrades ADP, already discussed as a potent activator of platelet aggregation. The major regulator of NO and PGI_2 production appears to be flow; precisely what senses flow is uncertain, though changes in cell shape and cytoskeleton are correlated. PGI_2 is produced by COX-1, which is expressed constitutively by "healthy" endothelium under normal flow conditions. NO is the product of endothelial nitric oxide synthase eNOS.
- *Anticoagulant effects.* Normal endothelium shields coagulation factors from tissue factor in vessel walls and expresses multiple factors that actively oppose coagulation, most notably thrombomodulin, endothelial protein C receptor, heparin-like molecules, and tissue factor pathway inhibitor. Thrombomodulin and endothelial protein C receptor bind thrombin and protein C, respectively, in a complex on the endothelial cell surface. When bound in this complex, thrombin loses its ability to activate coagulation

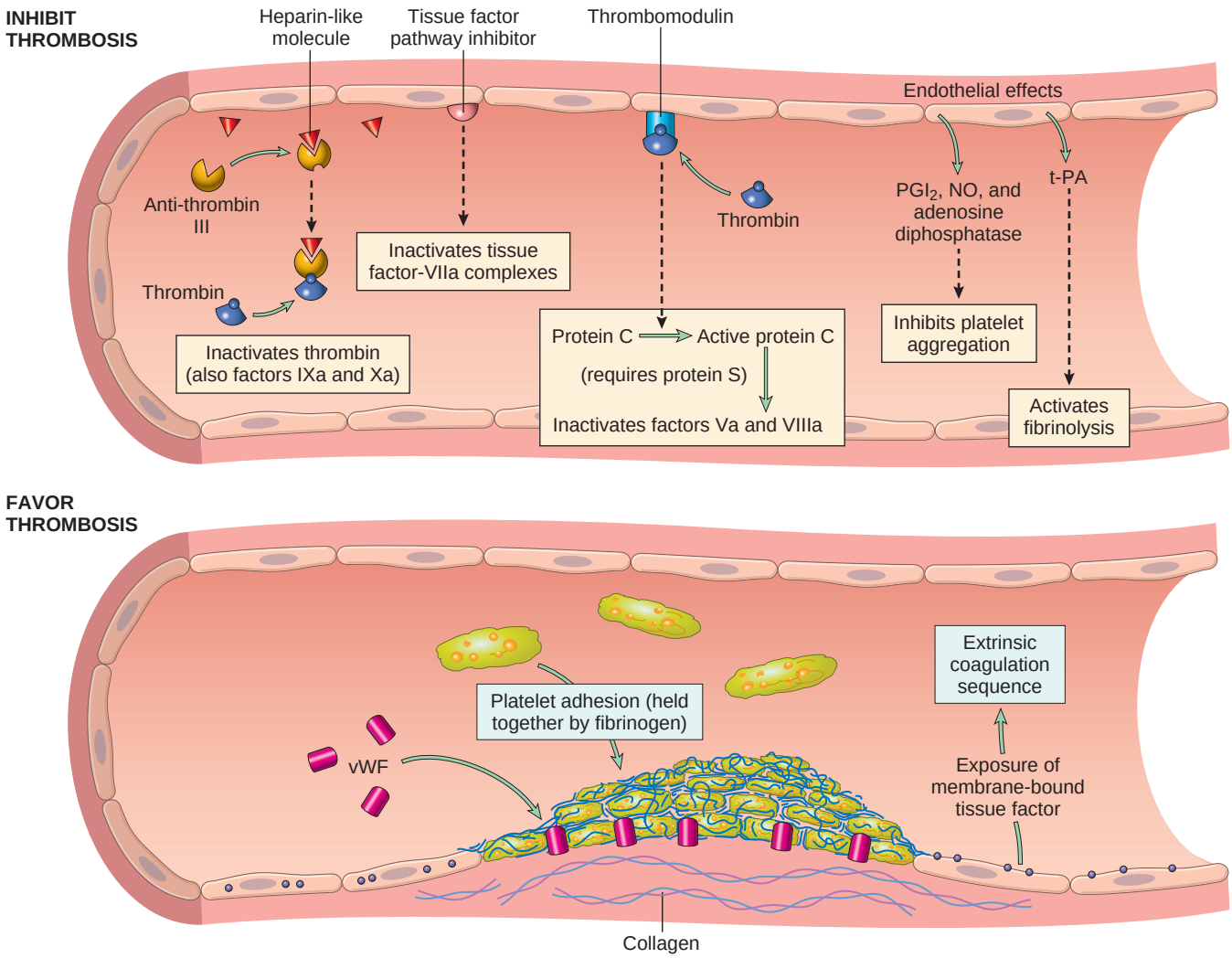


Figure 4.10 Anticoagulant activities of normal endothelium *(top)* and procoagulant properties of injured or activated endothelium *(bottom)*. *NO*, Nitric oxide; *PGI_2*, prostacyclin; *t-PA*, tissue plasminogen activator; *vWF*, von Willebrand factor. The thrombin receptor is one of the protease-activated receptors (PAR).

factors and platelets, and instead cleaves and activates protein C, a vitamin K–dependent protease that requires a cofactor, protein S. Activated protein C/protein S complex is a potent inhibitor of coagulation cofactors Va and VIIIa. *Heparin-like molecules* on the surface of endothelium bind and activate antithrombin III, which then inhibits thrombin and factors IXa, Xa, XIa, and XIIa. The clinical utility of heparin and related drugs is based on their ability to stimulate antithrombin III activity. Tissue factor pathway inhibitor (TFPI), like protein C, requires protein S as a cofactor and, as the name implies, binds and inhibits tissue factor/factor VIIa complexes.

- *Fibrinolytic effects.* Normal endothelial cells synthesize t-PA, as already discussed, a key component of the fibrinolytic pathway.

Hemorrhagic Disorders

Disorders associated with abnormal bleeding inevitably stem from primary or secondary defects in vessel walls, platelets, or coagulation factors, all of which must function properly to ensure hemostasis. The presentation of abnormal bleeding varies widely. At one end of the spectrum are massive bleeds associated with ruptures of large vessels such as the aorta or of the heart; these catastrophic events simply overwhelm hemostatic mechanisms and are often fatal. Diseases associated with sudden, massive hemorrhage include aortic dissection and aortic abdominal aneurysm (Chapter 11) and myocardial infarction (Chapter 12) complicated by rupture of the aorta or the heart. At the other end of the spectrum are subtle defects in clotting that only become evident under conditions of hemostatic stress, such as surgery, childbirth, dental procedures, menstruation, or trauma. Among the most common causes of mild bleeding tendencies are inherited defects in vWF (Chapter 14), aspirin consumption, and uremia (renal failure); the latter alters platelet function through uncertain mechanisms. Between these extremes lie deficiencies of coagulation factors (the hemophilias, Chapter 14), which are usually inherited and lead to severe bleeding disorders if untreated.

Additional specific examples of disorders associated with abnormal bleeding are discussed throughout the book. The

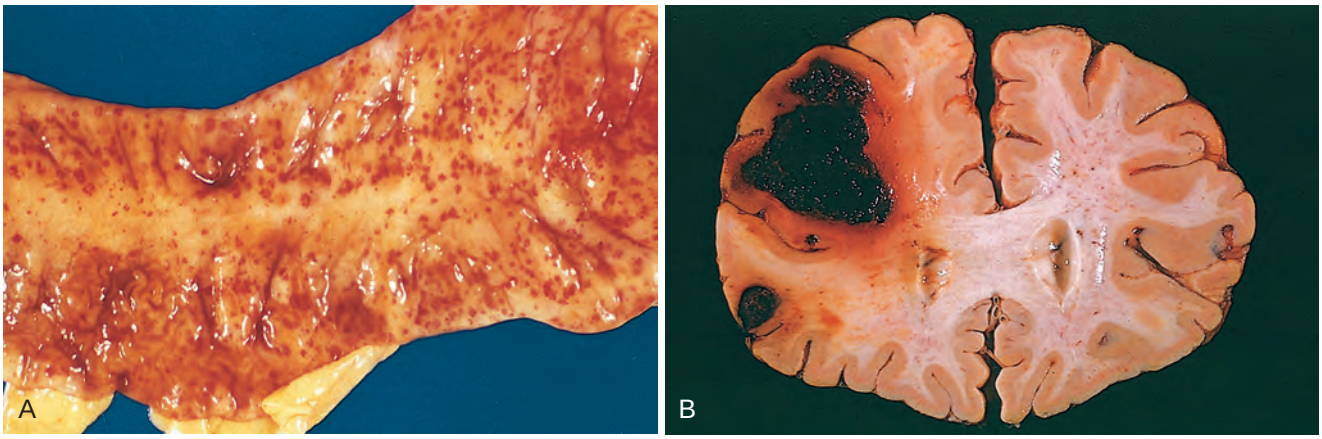


Figure 4.11 (A) Punctate petechial hemorrhages of the colonic mucosa, a consequence of thrombocytopenia. (B) Fatal intracerebral bleed.

following are general principles related to abnormal bleeding and its consequences.

- *Defects of primary hemostasis (platelet defects or von Willebrand disease)* often present with small bleeds in skin or mucosal membranes. These bleeds typically take the form of petechiae, minute 1- to 2-mm hemorrhages (Fig. 4.11A), or *purpura,* which are slightly larger (≥3 mm) than petechiae. It is believed that the capillaries of the mucosa and skin are particularly prone to rupture following minor trauma and that under normal circumstances platelets seal these defects virtually immediately. Mucosal bleeding associated with defects in primary hemostasis may also take the form of epistaxis (nosebleeds), gastrointestinal bleeding, or excessive menstruation (menorrhagia). A feared complication of very low platelet counts *(thrombocytopenia)* is intracerebral hemorrhage, which may be fatal.
- *Defects of secondary hemostasis (coagulation factor defects)* often present with bleeds into soft tissues (e.g., muscle) or joints. Bleeding into joints *(hemarthrosis)* following minor trauma is particularly characteristic of hemophilia (Chapter 14). It is unknown why severe defects in secondary hemostasis present with this peculiar pattern of bleeding; as with severe platelet defects, intracranial hemorrhage, sometimes fatal, may also occur.
- *Generalized defects involving small vessels* often present with "palpable purpura" and ecchymoses. *Ecchymoses* (sometimes simply called *bruises*) are hemorrhages of 1 to 2 cm in size. In both purpura and ecchymoses, the volume of extravasated blood may be large enough to create a palpable mass of blood known as a *hematoma*. Purpura and ecchymoses are particularly characteristic of systemic disorders that disrupt small blood vessels (e.g., vasculitis, Chapter 11) or that lead to blood vessel fragility (e.g., amyloidosis, Chapter 6; scurvy, Chapter 9).

The clinical significance of hemorrhage depends on the volume of the bleed, the rate at which it occurs, and its location. Rapid loss of up to 20% of the blood volume may have little impact in healthy adults; greater losses, however, can cause *hemorrhagic (hypovolemic) shock* (discussed later). Bleeding that is relatively trivial in the subcutaneous tissues can cause death if located in the brain (see Fig. 4.11B); because the skull is unyielding, intracranial hemorrhage may increase intracranial pressure to a level that compromises the blood supply or causes herniation of the brain stem (Chapter 28). Finally, chronic or recurrent external blood loss (e.g., peptic ulcer or menstrual bleeding) causes iron loss and can lead to iron deficiency anemia. In contrast, when red cells are retained (e.g., hemorrhage into body cavities or tissues), iron is recovered and recycled for use in the synthesis of hemoglobin.

Thrombosis

The primary abnormalities that lead to thrombosis are (1) endothelial injury, (2) stasis or turbulent blood flow, and (3) hypercoagulability of the blood (the so-called *Virchow triad*) (Fig. 4.12). Thrombosis is one of the scourges of modern man, because it underlies the most serious and common forms of cardiovascular disease. Here, the focus is on its causes and consequences; its role in cardiovascular disorders is discussed in detail in Chapters 11 and 12.

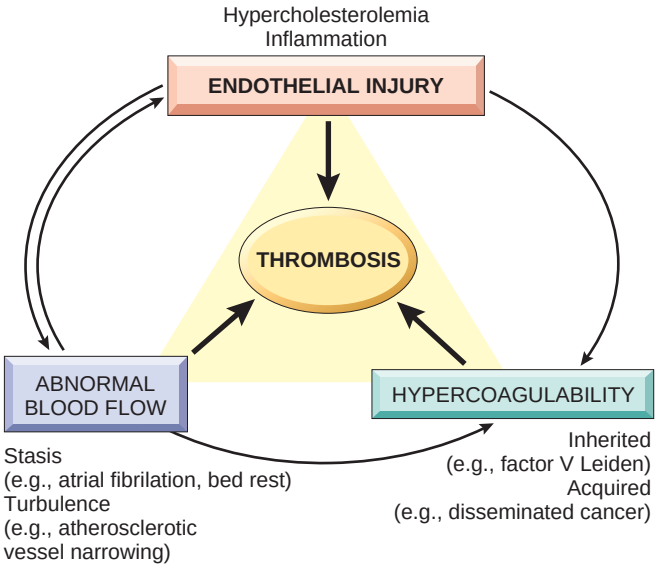


Figure 4.12 The Virchow triad in thrombosis. Endothelial integrity is the most important factor. Injury to endothelial cells can alter local blood flow and affect coagulability. Abnormal blood flow (stasis or turbulence), in turn, can cause endothelial injury. These factors may promote thrombosis independently or in combination.

Endothelial Injury

Endothelial injury leading to platelet activation almost inevitably underlies thrombus formation in the heart and the arterial circulation, where the high rates of blood flow impede clot formation. Notably, cardiac and arterial clots are typically rich in platelets, and it is believed that platelet adherence and activation is a necessary prerequisite for thrombus formation under high shear stress, such as exists in arteries. This insight provides part of the reasoning behind the use of aspirin and other platelet inhibitors in coronary artery disease and acute myocardial infarction.

Obviously, severe endothelial injury may trigger thrombosis by exposing vWF and tissue factor. However, inflammation and other noxious stimuli also promote thrombosis by shifting the pattern of gene expression in endothelium to one that is "prothrombotic." This change is sometimes referred to as *endothelial activation* or *dysfunction* and can be produced by diverse exposures, including physical injury, infectious agents, abnormal blood flow, inflammatory mediators, metabolic abnormalities, such as hypercholesterolemia or homocystinemia, and toxins absorbed from cigarette smoke. Endothelial activation is believed to have an important role in triggering arterial thrombotic events.

The role of endothelial cell activation and dysfunction in arterial thrombosis is discussed in detail in Chapters 11 and 12. Here it suffices to mention several of the major prothrombotic alterations:

- *Procoagulant changes.* Endothelial cells activated by inflammatory cytokines downregulate the expression of *thrombomodulin,* already described as a key modulator of thrombin activity, enhancing the procoagulant and pro-inflammatory actions of thrombin. In addition, inflamed endothelium also downregulates the expression of other anticoagulants, such as protein C and tissue factor protein inhibitor, changes that further promote a procoagulant state.
- *Antifibrinolytic effects.* Activated endothelial cells secrete *plasminogen activator inhibitors* (PAIs), which limit fibrinolysis, and downregulate the expression of t-PA, alterations that also favor the development of thrombi.

Alterations in Normal Blood Flow

Turbulence contributes to arterial and cardiac thrombosis by causing endothelial injury or dysfunction as well as by forming countercurrents that contribute to local pockets of stasis, whereas stasis is a major contributor to the development of venous thrombi Normal blood flow is laminar such that the platelets (and other blood cellular elements) flow centrally in the vessel lumen, separated from the endothelium by a slower moving layer of plasma. Turbulence and stasis therefore:

- *Promote endothelial activation,* enhancing procoagulant activity and leukocyte adhesion, in part through flow-induced changes in the expression of adhesion molecules and pro-inflammatory factors
- *Disrupt laminar flow* and bring platelets into contact with the endothelium
- *Prevent washout and dilution of activated clotting factors* by fresh flowing blood and the inflow of clotting factor inhibitors

Altered blood flow contributes to thrombosis in several clinical settings. Ulcerated atherosclerotic plaques not only expose subendothelial vWF and tissue factor but also cause turbulence. Aortic and arterial dilations called *aneurysms* result in local stasis and are therefore fertile sites for thrombosis (Chapter 11). Acute myocardial infarction results in areas of noncontractile myocardium and sometimes in cardiac aneurysm; both are associated with stasis and flow abnormalities that promote the formation of cardiac mural thrombi (Chapter 12). Rheumatic mitral valve stenosis results in left atrial dilation; in conjunction with atrial fibrillation, a dilated atrium is a site of profound stasis and a prime location for thrombosis (Chapter 12). Hyperviscosity (such as is seen with polycythemia vera; Chapter 13) increases resistance to flow and causes small vessel stasis, and the deformed red cells in *sickle cell anemia* (Chapter 14) impede blood flow through small vessels, with the resulting stasis also predisposing to thrombosis.

Hypercoagulability

Hypercoagulability refers to an abnormally high tendency of the blood to clot, and is typically caused by alterations in coagulation factors. Hypercoagulability has a particularly important role in venous thrombosis and can be divided into primary (genetic) and secondary (acquired) disorders (Table 4.2). Of the inherited causes of hypercoagulability,

Table 4.2 Hypercoagulable States

Primary (Genetic)
Common
Factor V mutation: factor V Leiden (Arg to Gln substitution in amino acid residue 506 leading to resistance to activated protein C)
Prothrombin mutation (G20210A noncoding sequence variant leading to increased prothrombin levels)
Increased levels of factors VIII, IX, XI, or fibrinogen (genetics unknown)
Rare
Antithrombin III deficiency
Protein C deficiency
Protein S deficiency
Very Rare
Fibrinolysis defects
Homozygous homocystinuria (deficiency of cystathione β-synthetase)
Secondary (Acquired)
Strong Risk Factors for Thrombosis
Prolonged bed rest or immobilization
Myocardial infarction
Atrial fibrillation
Tissue injury (surgery, fracture, burn)
Cancer
Prosthetic cardiac valves
Disseminated intravascular coagulation
Heparin-induced thrombocytopenia
Antiphospholipid antibody syndrome
Other Risk Factors for Thrombosis
Cardiomyopathy
Nephrotic syndrome
Hyperestrogenic states (pregnancy and postpartum)
Oral contraceptive use
Sickle cell anemia
Smoking

point mutations in the factor V gene and prothrombin gene are the most common. These are listed below:

- *Factor V Leiden.* Approximately 2% to 15% of Caucasians carry a single-nucleotide mutation in factor V that is called factor V Leiden, after the city in the Netherlands where it was discovered. Among individuals with recurrent DVT, the frequency of this mutation is considerably higher, approaching 60%. This mutation renders factor V resistant to cleavage and inactivation by protein C. As a result, an important antithrombotic counterregulatory pathway is lost (see Fig. 4.10). The inheritance pattern for factor V Leiden is autosomal dominant. Heterozygotes have a fivefold increased relative risk of venous thrombosis, and homozygotes have a 50-fold increase.
- *Prothrombin gene mutation.* A single nucleotide change (G20210A) in the 3′-untranslated region of the prothrombin gene is another common mutation (1% to 2% of the population) associated with hypercoagulability. It leads to elevated prothrombin levels and an almost threefold increased risk of venous thrombosis.
- *Other inherited causes.* Rare inherited causes of primary hypercoagulability include deficiencies of anticoagulants such as antithrombin III, protein C, or protein S; affected individuals typically present with venous thrombosis and recurrent thromboembolism beginning in adolescence or early adulthood.
- *Homocysteinemia.* Elevated levels of homocysteine may be inherited or acquired. Marked elevations of homocysteine may be caused by an inherited deficiency of cystathione β-synthetase. Acquired causes include deficiency of vitamin B_6, B_{12}, and folic acid. Prothrombotic effects of homocysteine may be due to thioester linkages formed between homocysteine metabolites and a variety of proteins, including fibrinogen.

The most common thrombophilic genotypes found in various populations (heterozygosity for factor V Leiden and heterozygosity for the prothrombin G20210A variant) impart only a moderately increased risk of thrombosis; most individuals with these genotypes, when otherwise healthy, are free from thrombotic complications. However, factor V and prothrombin mutations are frequent enough that homozygosity and compound heterozygosity are not rare, and such genotypes are associated with greater risk. Moreover, individuals with such mutations have a significantly increased frequency of venous thrombosis in the setting of other acquired risk factors (e.g., pregnancy or prolonged bed rest). Thus, factor V Leiden heterozygosity may trigger deep vein thrombosis (DVT) when combined with enforced inactivity, such as during prolonged airplane travel. Consequently, **inherited causes of hypercoagulability must be considered in patients younger than 50 years of age who present with thrombosis—even when acquired risk factors are present.**

Unlike hereditary disorders, the pathogenesis of *acquired thrombophilia* is frequently multifactorial (see Table 4.2). In some cases (e.g., cardiac failure or trauma), stasis or vascular injury may be most important. Hypercoagulability due to oral contraceptive use or the hyperestrogenic state of pregnancy is probably caused by increased hepatic synthesis of coagulation factors and reduced anticoagulant synthesis. In disseminated cancers, release of various procoagulants from tumors predisposes to thrombosis. The hypercoagulability seen with advancing age may be due to reduced endothelial PGI_2 production. Smoking and obesity promote hypercoagulability by unknown mechanisms.

Among the acquired thrombophilic states, heparin-induced thrombocytopenia and antiphospholipid antibody syndrome are particularly important clinical problems that deserve special mention.

Heparin-Induced Thrombocytopenia (HIT) Syndrome

HIT syndrome is a serious, potentially life-threatening disorder that occurs following the administration of *unfractionated heparin*. It results from the formation of antibodies that recognize complexes of heparin and PF4 on the surface of platelets (Chapter 14), as well as complexes of heparin-like molecules and PF4-like proteins on endothelial cells. PF4 protein is normally found in platelet alpha granules and is released on activation of platelets. Released PF4 binds to heparin and undergoes a conformational change that results in the formation of a neoantigen against which IgG antibodies are formed. PF4-IgG immune complex (Fig. 4.13) attaches to and cross-links the Fc receptors on the platelet surface, which leads to platelet activation and aggregation. Platelet activation results in the release of more PF4, creating more target antigen for HIT antibodies. The prothrombotic state may also be augmented by activation of endothelium by binding of HIT antibodies to PF4-like proteins on their surface. Binding of HIT antibodies to platelets results in their removal by macrophages (hence the thrombocytopenia in the syndrome name). Although thrombocytopenia is the most common manifestation, thrombosis is the most serious complication. It occurs in approximately 50% of cases and affects both veins and arteries. Necrosis of the skin, gangrene of the limbs, stroke, and myocardial infarction are some of the sequelae. Diagnosis requires the demonstration of anti–PF4-heparin antibodies. Low-molecular-weight heparin preparations induce HIT less frequently, and other classes of anticoagulants such as direct inhibitors of factor X and thrombin may also obviate the risk.

Antiphospholipid Antibody Syndrome (APS)

APS is an autoimmune disorder characterized by:

- *Presence of one or more antiphospholipid (aPL) autoantibodies*
- *Venous or arterial thromboses, or pregnancy complications such as recurrent miscarriages, unexplained fetal death, and premature birth.*

APS may be primary or secondary. Individuals with a well-defined autoimmune disease, such as systemic lupus erythematosus (Chapter 6), are designated as having secondary antiphospholipid syndrome (hence the earlier term *lupus anticoagulant syndrome*). In primary antiphospholipid syndrome, patients exhibit only the manifestations of a hypercoagulable state and lack evidence of other well-defined autoimmune disorders. Approximately 50% of the patients with APS have the primary form, and the rest occur in association with well-defined autoimmune disease, most commonly SLE. Our focus here is on the primary form.

The clinical manifestations of APS are varied; they include recurrent thromboses, repeated miscarriages, cardiac valve

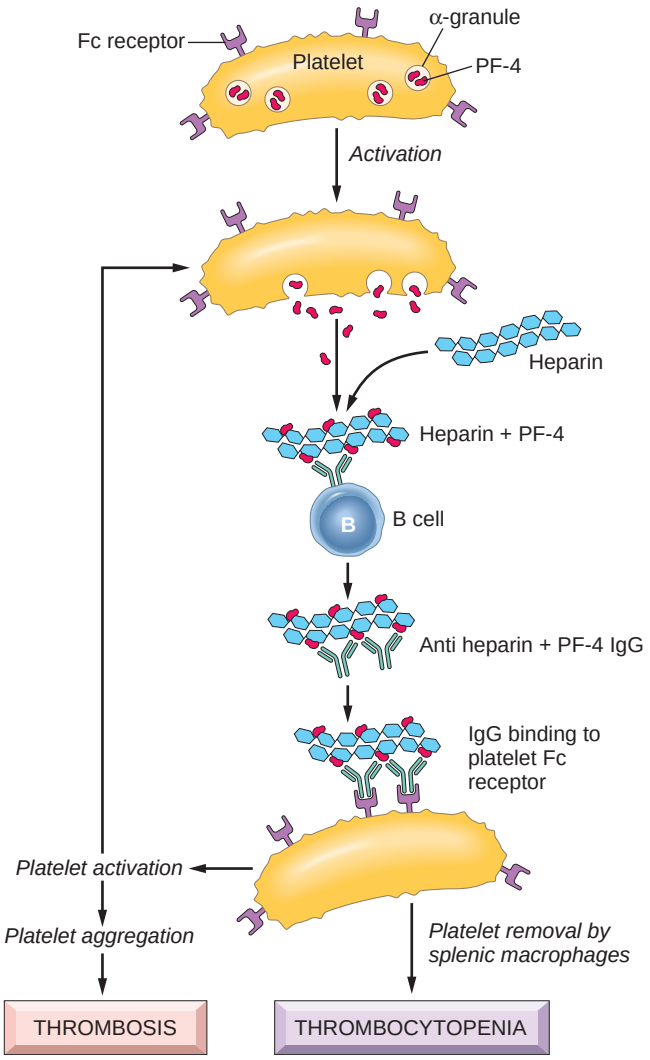


Figure 4.13 Mechanism of heparin-induced thrombocytopenia.

vegetations, and thrombocytopenia. Depending on the vascular bed involved, the clinical presentations can include pulmonary embolism (following lower extremity venous thrombosis), pulmonary hypertension (from recurrent subclinical pulmonary emboli), valvular heart disease, stroke, bowel infarction, or renovascular hypertension.

The pathogenesis of antiphospholipid syndrome is complex and not fully understood. The aPL antibodies are directed against anionic membrane phospholipids or proteins associated with phospholipids. Proteins that are recognized by these antibodies include cardiolipin and β_2-glycoprotein I. This glycoprotein is found in plasma, but it has strong avidity for phospholipids expressed on the surfaces of endothelial cells, monocytes, platelets, thrombin, and trophoblasts. Anti–β_2-glycoprotein antibodies are suspected to have a major role in APS by activating endothelial cells, monocytes, and platelets. Their pathogenicity is supported by the observation that transfer of these antibodies into rodents can induce thrombosis. Patients with APS also show evidence of complement activation and inhibition of fibrinolytic processes, both of which favor the prothrombotic state. As mentioned earlier, pregnancy morbidity is a defining manifestation of APS. However, unlike most other clinical features, fetal loss does not appear to be caused by thrombosis, but rather seems to stem from antibody-mediated interference with the growth and differentiation of trophoblasts, leading to a failure of placentation.

Although antiphospholipid antibodies are clearly associated with thrombotic diatheses, they have also been identified in 5% to 15% of apparently normal individuals, implying that they are not sufficient to cause the full-blown syndrome. It is postulated that a "second hit" is required that may be provided by infection, smoking, or pregnancy, among others. Although these antibodies induce a hypercoagulable state in vivo, they interfere with phospholipids and thus inhibit coagulation in vitro, thereby prolonging the PTT. The antibodies also frequently result in a false-positive serologic test for syphilis because the antigen in the standard assay is embedded in cardiolipin, which cross reacts with phospholipids of *Treponema pallidum*. Diagnosis of APS is based on clinical features and demonstration of aPL antibodies in the serum. Therapy of APS involves various forms of anticoagulation.

MORPHOLOGY

Thrombi can develop anywhere in the cardiovascular system and vary in size and shape depending on the involved site and the underlying cause. Arterial or cardiac thrombi usually begin at sites of turbulence or endothelial injury, whereas venous thrombi characteristically occur at sites of stasis. Thrombi are focally attached to the underlying vascular surface, particularly at the point of initiation. From here, arterial thrombi tend to grow retrograde, and venous thrombi extend in the direction of blood flow; thus both propagate toward the heart. The propagating portion of a thrombus is often poorly attached and therefore prone to fragmentation and embolization.

Thrombi often have grossly and microscopically apparent laminations called **lines of Zahn,** which are pale platelet and fibrin deposits alternating with darker red cell–rich layers. Such laminations signify that a thrombus has formed in flowing blood; their presence can therefore distinguish antemortem clots from the bland nonlaminated clots that occur postmortem (see later).

Thrombi occurring in heart chambers or in the aortic lumen are designated **mural thrombi.** Abnormal myocardial contraction (arrhythmias, dilated cardiomyopathy, or myocardial infarction) or endomyocardial injury (myocarditis or catheter trauma) promotes cardiac mural thrombi (Fig. 4.14A), and ulcerated atherosclerotic plaque and aneurysmal dilation underlie aortic thrombi (see Fig. 4.14B).

Arterial thrombi are frequently **occlusive**; the most common sites in decreasing order of frequency are the coronary, cerebral, and femoral arteries. They typically consist of a friable meshwork of platelets, fibrin, red cells, and degenerating leukocytes. Although these are usually superimposed on a ruptured atherosclerotic plaque, other vascular injuries (vasculitis, trauma) may be the underlying cause.

Venous thrombosis (phlebothrombosis) is almost invariably occlusive, with the thrombus forming a long luminal cast. Because these thrombi form in the sluggish venous circulation, they tend to contain more enmeshed red cells (and relatively few platelets) and are therefore known as **red thrombi** or **stasis thrombi.** Venous thrombi are firm, are focally attached to the

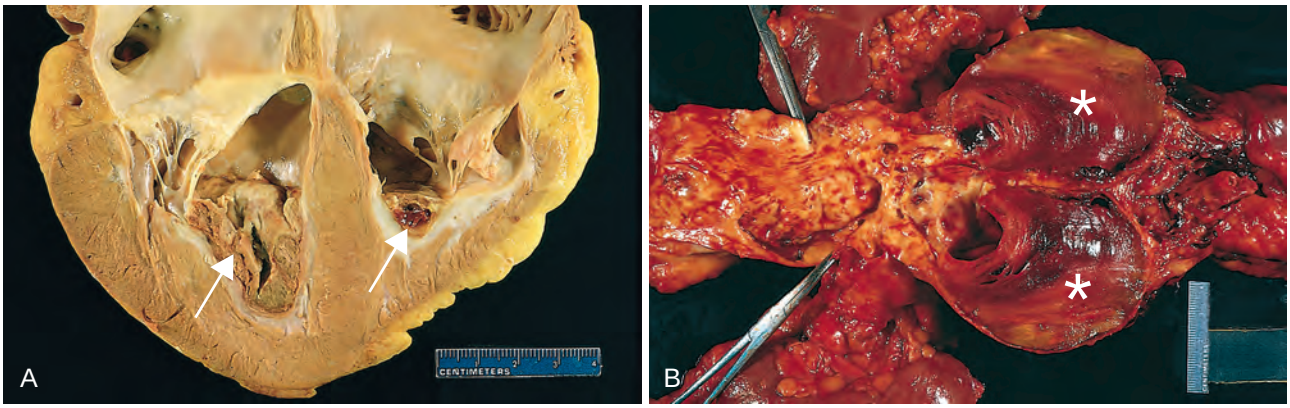


Figure 4.14 Mural thrombi. (A) Thrombus in the left and right ventricular apices *(arrows)*, overlying a white fibrous scar. (B) Laminated thrombus in a dilated abdominal aortic aneurysm *(asterisks)*. Numerous friable mural thrombi are also superimposed on advanced atherosclerotic lesions of the more proximal aorta *(left side of picture)*.

vessel wall, and contain lines of Zahn, features that help distinguish them from postmortem clots (see later). The veins of the lower extremities are most commonly involved (90% of cases); however, upper extremities, periprostatic plexus, or the ovarian and periuterine veins can also develop venous thrombi. Under special circumstances, they can also occur in the dural sinuses, portal vein, or hepatic vein.

Postmortem clots can sometimes be mistaken for antemortem venous thrombi. However, clots that form after death are gelatinous and have a dark-red dependent portion where red cells have settled by gravity and a yellow "chicken fat" upper portion, and are usually not attached to the underlying vessel wall.

Thrombi on heart valves are called **vegetations**, which may be infected or sterile. Blood-borne bacteria or fungi can adhere to previously damaged valves (e.g., due to rheumatic heart disease) or may cause valve damage directly; in either case, endothelial injury and disturbed blood flow can induce the formation of large thrombotic masses (**infective endocarditis**; Chapter 12). Sterile vegetations can also develop on noninfected valves in persons with hypercoagulable states, so-called **nonbacterial thrombotic endocarditis** (Chapter 12). Less commonly, sterile verrucous endocarditis **(Libman-Sacks endocarditis)** can occur in the setting of systemic lupus erythematosus (Chapter 6).

Fate of the Thrombus

If a patient survives the initial thrombosis, in the ensuing days to weeks thrombi undergo some combination of the following four events:

- *Propagation.* Thrombi accumulate additional platelets and fibrin (discussed earlier).
- *Embolization.* Thrombi dislodge and travel to other sites in the vasculature (discussed later).
- *Dissolution.* Dissolution is the result of fibrinolysis, which can lead to the rapid shrinkage and total disappearance of recent thrombi. In contrast, the extensive fibrin deposition and cross-linking in older thrombi render them more resistant to lysis. This distinction explains why therapeutic administration of fibrinolytic agents such as t-PA (e.g., in the setting of acute coronary thrombosis) is generally effective only when given during the first few hours of a thrombotic event.
- *Organization and recanalization.* Older thrombi become organized by the ingrowth of endothelial cells, smooth muscle cells, and fibroblasts (Fig. 4.15). Capillary channels eventually form that reestablish the continuity of the original lumen, albeit to a variable degree. Continued recanalization may convert a thrombus into a smaller mass of connective tissue that becomes incorporated into the vessel wall. Eventually, with remodeling and contraction of the mesenchymal elements, only a fibrous lump may remain to mark the original thrombus (Chapter 11).

Occasionally the centers of thrombi undergo enzymatic digestion, presumably as a result of the release of lysosomal enzymes from trapped leukocytes and platelets. In the setting of bacteremia, such thrombi may become infected, producing an inflammatory mass that erodes and weakens the vessel

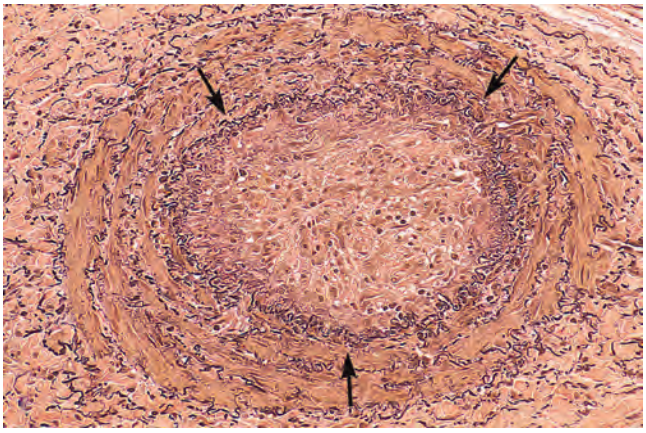


Figure 4.15 Low-power view of a thrombosed artery stained for elastic tissue. The original lumen is delineated by the internal elastic lamina *(arrows)* and is totally filled with organized thrombus, now punctuated by several recanalized endothelium-lined channels *(white spaces)*.

wall. If unchecked, this may result in a mycotic aneurysm (Chapter 11).

Clinical Features

Thrombi come to clinical attention when they obstruct arteries or veins, or give rise to emboli. The clinical presentation depends on the involved site. Venous thrombi can cause painful congestion and edema distal to an obstruction, but are mainly of concern due to their tendency to embolize to the lungs (see later). Conversely, although arterial thrombi can also embolize and cause downstream infarctions, the chief clinical problem is more often related to occlusion of a critical vessel (e.g., a coronary or cerebral artery), which can have serious or fatal consequences.

Venous Thrombosis (Phlebothrombosis). Most venous thrombi occur in the superficial or deep veins of the leg. Superficial venous thrombi typically occur in the saphenous veins in the setting of varicosities. Such thrombi can cause local congestion, swelling, pain, and tenderness, but rarely embolize. Nevertheless, the associated edema and impaired venous drainage predispose the overlying skin to the development of infection and ulcers *(varicose ulcers)*. DVT involving one of the large leg veins—at or above the knee (e.g., the popliteal, femoral, and iliac veins)—is considered serious because such thrombi more often embolize to the lungs and give rise to pulmonary infarction (see later and Chapter 15). Although DVTs may cause local pain and edema due to venous obstruction, these symptoms are often absent due the opening of venous collateral channels. Consequently, DVTs are asymptomatic in approximately 50% of affected individuals and are recognized only in retrospect after embolization.

Lower extremity DVTs are often associated with hypercoagulable states, as described earlier (see Table 4.2). Common predisposing factors include bed rest and immobilization (because they reduce the milking action of the leg muscles, resulting in stasis), and congestive heart failure (also a cause of impaired venous return). Trauma, surgery, and burns not only immobilize a person but are also associated with vascular insults, procoagulant release from injured tissues, increased hepatic synthesis of coagulation factors, and decreased t-PA production. Many elements contribute to the thrombotic diathesis of pregnancy, including decreased venous return from leg veins and systemic hypercoagulability associated with the hormonal changes of late pregnancy and the postpartum period. Tumor-associated inflammation and coagulation factors (tissue factor, factor VIII), as well as procoagulants (e.g., mucin) released from tumor cells, all contribute to the increased risk of thromboembolism in disseminated cancers, so-called *migratory thrombophlebitis* or *Trousseau syndrome*. Regardless of the specific clinical setting, advanced age also increases the risk of DVT.

Arterial and Cardiac Thrombosis. Atherosclerosis is a major cause of arterial thromboses because it is associated with loss of endothelial integrity and with abnormal blood flow (see Fig. 4.14B). Myocardial infarction can predispose to cardiac mural thrombi by causing dyskinetic myocardial contraction and endocardial injury (see Fig. 4.14A), and rheumatic heart disease may engender atrial mural thrombi by causing atrial dilation and fibrillation. Both cardiac and aortic mural thrombi are prone to embolization. Although any tissue can be affected, the brain, kidneys, and spleen are particularly likely targets because of their rich blood supply.

KEY CONCEPTS

THROMBOSIS

- Thrombus development usually is related to one or more components of the Virchow triad:
  - Endothelial injury (e.g., by toxins, hypertension, inflammation, or metabolic products) associated with endothelial activation and changes in endothelial gene expression that favor coagulation
  - Abnormal blood flow—stasis or turbulence (e.g., due to aneurysms, atherosclerotic plaque)
  - Hypercoagulability, either primary (e.g., factor V Leiden, increased prothrombin synthesis, antithrombin III deficiency) or secondary (e.g., bed rest, tissue damage, malignancy, or development of aPL antibodies [antiphospholipid antibody syndrome]) or antibodies against PF4/heparin complexes [heparin-induced thrombocytopenia])
- Thrombi may propagate, resolve, become organized, or embolize.
- Thrombosis causes tissue injury by local vascular occlusion or by distal embolization.

Disseminated Intravascular Coagulation (DIC)

DIC is widespread thrombosis within the microcirculation that may be of sudden or insidious onset. It is not a specific disease but rather a complication of a large number of conditions associated with systemic activation of thrombin. Disorders ranging from obstetric complications to advanced malignancy can be complicated by DIC, which leads to widespread formation of thrombi in the microcirculation. These microvascular thrombi can cause diffuse circulatory insufficiency and organ dysfunction, particularly of the brain, lungs, heart, and kidneys. To complicate matters, the runaway thrombosis "uses up" platelets and coagulation factors (hence the synonym *consumptive coagulopathy*) and often activates fibrinolytic mechanisms. Thus, symptoms initially related to thrombosis can evolve into a bleeding catastrophe, such as hemorrhagic stroke or hypovolemic shock. DIC is discussed in greater detail along with other bleeding diatheses in Chapter 14.

EMBOLISM

An embolus is a detached intravascular solid, liquid, or gaseous mass that is carried by the blood from its point of origin to a distant site, where it often causes tissue dysfunction or infarction. The vast majority of emboli are dislodged thrombi, hence the term *thromboembolism*. Other rare emboli are composed of fat droplets, nitrogen bubbles, atherosclerotic debris *(cholesterol emboli)*, tumor fragments, bone marrow, or even foreign bodies. Emboli travel through

the blood until they encounter vessels too small to permit further passage, causing partial or complete vascular occlusion. Depending on where they originate, emboli can lodge anywhere in the vascular tree; as discussed later, the clinical consequences vary widely depending on the size and the position of the lodged embolus, as well as the vascular bed that is impacted.

Pulmonary Embolism (PE)

Pulmonary emboli originate from DVT and are the most common form of thromboembolic disease. PE is a common and serious disorder with an estimated incidence of 100 to 200 cases per 100,000 people in the United States. It is somewhat more common in males than in females. PE causes about 100,000 deaths per year in the United States. An estimated 20% of individuals with PE die before or shortly after a diagnosis is made. In more than 95% of cases, PE originates from leg DVT. Hence the risk factors for PE are the same as for DVT (see Table 4.2).

Fragmented thrombi from DVT are carried through progressively larger veins and the right side of the heart before slamming into the pulmonary arterial vasculature. Depending on the size of the embolus, it can occlude the main pulmonary artery, straddle the pulmonary artery bifurcation *(saddle embolus),* or pass out into the smaller, branching arteries (Fig. 4.16). Frequently there are multiple emboli, occurring either sequentially or simultaneously as a shower of smaller emboli from a single large mass; in general, **the patient who has had one PE is at high risk for more.** Rarely, a venous embolus passes through an interatrial or interventricular defect and gains access to the systemic arterial circulation *(paradoxical embolism).* A more complete discussion of PEs is presented in Chapter 15; the following is an overview of the major functional consequences of pulmonary emboli.

- *Most pulmonary emboli (60% to 80%) are clinically silent* because they are small. With time they become organized and are incorporated into the vascular wall; in some cases organization of the thromboembolus leaves behind a delicate, bridging fibrous web.

Figure 4.16 Embolus from a lower extremity deep venous thrombosis, lodged at a pulmonary artery branchpoint.

- *Sudden death, acute right heart failure (cor pulmonale),* or cardiovascular collapse occurs when emboli obstruct 60% or more of the pulmonary circulation.
- *Embolic obstruction of medium-sized arteries with subsequent vascular rupture can result in pulmonary hemorrhage* but usually does not cause pulmonary infarction. This is because the lung is supplied by both the pulmonary arteries and the bronchial arteries, and the intact bronchial circulation is usually sufficient to perfuse the affected area. Understandably, if the bronchial arterial flow is compromised (e.g., by left-sided cardiac failure), infarction may occur.
- *Embolic obstruction of small end-arteriolar pulmonary branches often does produce hemorrhage or infarction.*
- *Multiple emboli over time may cause pulmonary hypertension* and right ventricular failure.

Systemic Thromboembolism

Most systemic emboli (80%) arise from intracardiac mural thrombi, two-thirds of which are associated with left ventricular wall infarcts and another one-fourth with left atrial dilation and fibrillation. The remainder originates from aortic aneurysms, atherosclerotic plaques, valvular vegetations, or venous thrombi (paradoxical emboli); 10% to 15% are of unknown origin. In contrast to venous emboli, the vast majority of which lodge in the lung, arterial emboli can travel to a wide variety of sites; the point of arrest depends on the source and the relative amount of blood flow that downstream tissues receive. Most come to rest in the lower extremities (75%) or the brain (10%), but other tissues, including the intestines, kidneys, spleen, and upper extremities, may be involved. The consequences of systemic emboli depend on the vulnerability of the affected tissues to ischemia, the caliber of the occluded vessel, and whether a collateral blood supply exists; in general, however, the outcome is tissue infarction.

Fat Embolism

Fat embolism refers to the presence of microscopic fat globules—sometimes with associated hematopoietic bone marrow—in the vasculature after fractures of long bones or, rarely, in the setting of soft-tissue trauma and burns. It is fairly common, occurring in some 90% of individuals with severe skeletal injuries (Fig. 4.17). Presumably these injuries rupture vascular sinusoids in the marrow or small venules, allowing marrow or adipose tissue to herniate into the vascular space and travel to the lung. Fat embolism syndrome is the term applied to the minority of patients who become symptomatic.

Fat embolism syndrome is characterized by pulmonary insufficiency, neurologic symptoms, anemia, and thrombocytopenia, and is fatal in 5% to 15% of cases. Typically, 1 to 3 days after injury there is a sudden onset of tachypnea, dyspnea, and tachycardia; irritability; and restlessness that can progress to delirium or coma. Thrombocytopenia is attributed to platelet adhesion to fat globules and subsequent aggregation or splenic sequestration; anemia can result from similar red cell aggregation and/or hemolysis. A diffuse petechial rash (seen in 20% to 50% of cases) is related to rapid onset of thrombocytopenia and can be a useful diagnostic clue.

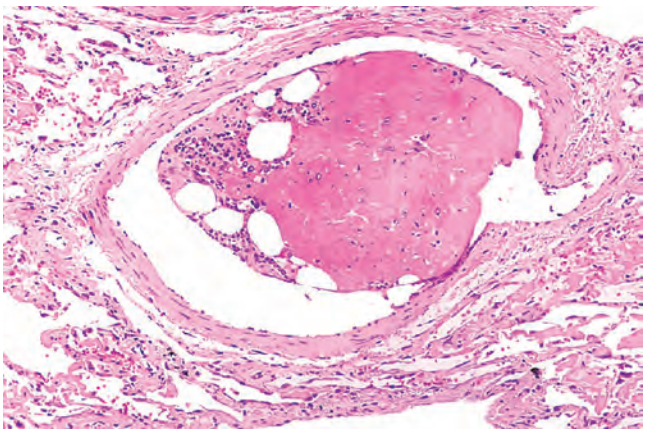


Figure 4.17 Bone marrow embolus in the pulmonary circulation. The cellular elements on the left side of the embolus are hematopoietic cells, and the cleared vacuoles represent marrow fat. The relatively uniform red area on the right of the embolus is an early organizing thrombus.

The pathogenesis of fat embolism syndrome involves both mechanical obstruction and biochemical injury. Fat microemboli and associated red cell and platelet aggregates can occlude the pulmonary and cerebral microvasculature. Release of free fatty acids from the fat globules exacerbates the situation by causing local toxic injury to endothelium, and platelet activation and granulocyte recruitment (with free radical, protease, and eicosanoid release) complete the vascular assault. Because lipids are dissolved out of tissue preparations by solvents routinely used in paraffin embedding, the microscopic demonstration of fat microglobules typically requires specialized techniques, including frozen sections and stains for fat.

Air Embolism

Gas bubbles within the circulation can coalesce to form frothy masses that obstruct vascular flow and cause distal ischemic injury. Air embolism occurs when there is communication between the vasculature and outside air, and a negative pressure gradient that "sucks" in the air. For example, air may be introduced into the cerebral circulation by neurosurgery in the "sitting position," creating a gravitational gradient. Air may also be introduced during endovascular and interventional procedures as well as during mechanical ventilation. A large volume of air, generally more than 100 mL, is necessary to produce a clinical effect in the pulmonary circulation; unless care is taken, this volume of air can be inadvertently introduced during obstetric or laparoscopic procedures, or as a consequence of chest wall injury. Introduction of 300 to 500 mL of air at 100 mL/sec may be fatal. Entry of air in the pulmonary vasculature does not merely block perfusion of downstream region. Microemboli of air trapped in pulmonary capillaries induce an intense inflammatory response with release of cytokines that may injure the alveoli. Bubbles in the central nervous system can cause mental impairment and even sudden onset of coma.

A particular form of gas embolism, called *decompression sickness,* occurs when individuals experience sudden decreases in atmospheric pressure. Scuba and deep sea divers, and underwater construction workers are all at risk. When air is breathed at high pressure (e.g., during a deep sea dive), increased amounts of gas (particularly nitrogen) are dissolved in the blood and tissues. If the diver then ascends (depressurizes) too rapidly, the nitrogen comes out of solution in the tissues and the blood.

The rapid formation of gas bubbles within skeletal muscles and supporting tissues in and about joints is responsible for the painful condition called *the bends* (so named in the 1880s because it was noted that those afflicted characteristically arched their backs in a manner reminiscent of a then-popular women's fashion pose called the *Grecian bend*). In the lungs, gas bubbles in the vasculature cause edema, hemorrhage, and focal atelectasis or emphysema, leading to a form of respiratory distress called the *chokes*. A more chronic form of decompression sickness is called *caisson disease* (named for the pressurized vessels used in bridge construction; workers in these vessels suffered both acute and chronic forms of decompression sickness). In caisson disease, persistence of gas emboli in the skeletal system leads to multiple foci of ischemic necrosis; the more common sites are the femoral heads, tibia, and humeri.

Individuals affected by acute decompression sickness are treated by being placed in a chamber under sufficiently high pressure to force the gas bubbles back into solution. Subsequent slow decompression permits gradual resorption and exhalation of the gases, which prevents the obstructive bubbles from reforming.

Amniotic Fluid Embolism

Amniotic fluid embolism is the fifth most common cause of maternal mortality worldwide; it accounts for roughly 10% of maternal deaths in the United States and results in permanent neurologic deficit in as many as 85% of survivors. Amniotic fluid embolism is an ominous complication of labor and the immediate postpartum period. Although the incidence is only approximately 2 to 6 in 100,000 deliveries, the mortality rate is up to 80%. The onset is characterized by sudden severe dyspnea, cyanosis, and shock, followed by neurologic impairment ranging from headache to seizures and coma, and by DIC. Note that these features differ from those observed with pulmonary embolism from DVT; in fact, much of the morbidity and mortality in amniotic fluid embolism stems from the biochemical activation of coagulation factors, components of the innate immune system, and release of vasoactive substances, rather than the mechanical obstruction of pulmonary vessels by amniotic debris. The vasoactive substances cause acute pulmonary hypertension and right heart failure, which causes hypoxia, left heart failure, pulmonary edema, and diffuse alveolar damage.

The underlying cause is the infusion of amniotic fluid or fetal tissue into the maternal circulation via a tear in the placental membranes or rupture of uterine veins. Classic findings at autopsy include the presence of squamous cells shed from fetal skin, lanugo hair, fat from vernix caseosa, and mucin derived from the fetal respiratory or gastrointestinal tract in the maternal pulmonary microvasculature (Fig. 4.18).

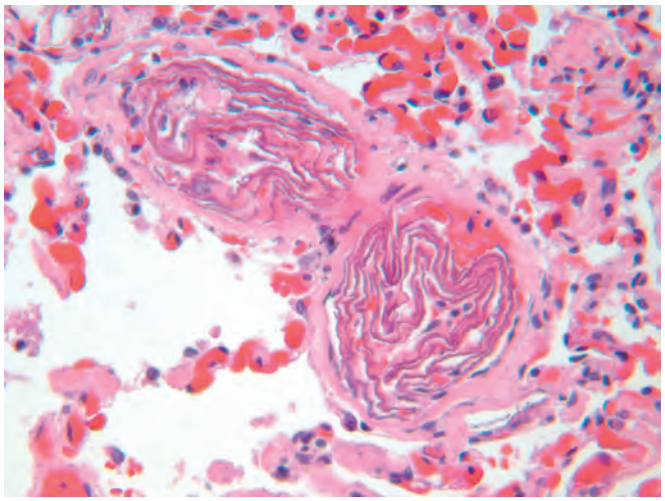


Figure 4.18 Amniotic fluid embolism. Two small pulmonary arterioles are packed with laminated swirls of fetal squamous cells. There is marked edema and congestion. Elsewhere the lung contained small organizing thrombi consistent with disseminated intravascular coagulation. (Courtesy Dr. Beth Schwartz, Baltimore, Md.)

KEY CONCEPTS

EMBOLISM

- An embolus is a solid, liquid, or gaseous mass carried by the blood to a site distant from its origin; most are dislodged thrombi.
- Pulmonary emboli derive primarily from lower extremity DVT; their effects depend on the size of the embolus and the location in which it lodges. Consequences may include right-sided heart failure, pulmonary hemorrhage, pulmonary infarction, or sudden death.
- Systemic emboli derive from cardiac mural or valvular thrombi, aortic aneurysms, or atherosclerotic plaques; whether an embolus causes tissue infarction depends on the site of embolization and the presence or absence of collateral circulation.

INFARCTION

An infarct is an area of ischemic necrosis caused by occlusion of either the arterial supply or the venous drainage. Tissue infarction is a common and extremely important cause of clinical illness. Roughly 40% of all deaths in the United States are caused by cardiovascular disease, and most of these are attributable to myocardial or cerebral infarction. Pulmonary infarction is also a common complication in many clinical settings, bowel infarction is frequently fatal, and ischemic necrosis of the extremities *(gangrene)* is a serious problem in the diabetic population.

Arterial thrombosis or arterial embolism underlies the vast majority of infarctions. Less common causes of arterial obstruction leading to infarction include local vasospasm, hemorrhage into an atheromatous plaque, or extrinsic vessel compression (e.g., by tumor). Other uncommon causes of tissue infarction include torsion of a vessel (e.g., in testicular torsion or bowel volvulus), traumatic vascular rupture, or vascular compromise by edema (e.g., *anterior compartment syndrome* of the leg) or by entrapment in a hernia sac. Although venous thrombosis can cause infarction, the more common outcome is just congestion; in this setting, bypass channels rapidly open and permit vascular outflow, which then improves arterial inflow. Infarcts caused by venous thrombosis are thus more likely in organs with a single efferent vein (e.g., testis and ovary).

MORPHOLOGY

Infarcts are classified according to color and the presence or absence of infection; they are either red (hemorrhagic) or white (anemic) and may be septic or bland.

- **Red infarcts** (Fig. 4.19A) occur (1) with venous occlusions (e.g., testicular torsion, Chapter 19), (2) in loose, spongy tissues (e.g., lung) where blood can collect in the infarcted zone, (3) in tissues with dual circulations (e.g., lung and small intestine) that allow blood to flow from an unobstructed parallel supply into a necrotic zone, (4) in tissues previously congested by sluggish venous outflow, and (5) when flow is reestablished to a site of previous arterial occlusion and necrosis (e.g., following angioplasty of an arterial obstruction).
- **White infarcts** (see Fig. 4.19B) occur with arterial occlusions in solid organs with end-arterial circulation (e.g., heart, spleen, and kidney), and where tissue density limits the seepage of blood from adjoining capillary beds into the necrotic area.

Infarcts tend to be wedge-shaped, with the occluded vessel at the apex and the periphery of the organ forming the base (see Fig. 4.19); when the base is a serosal surface, there may be an overlying fibrinous exudate resulting from an acute inflammatory response to mediator release from injured and necrotic cells. Fresh infarcts are poorly defined and slightly hemorrhagic, but over a few days the margins tend to become better defined by a narrow rim of congestion attributable to inflammation. With further passage of time, infarcts resulting from arterial occlusions in organs without a dual blood supply typically become progressively paler and more sharply defined (see Fig. 4.19B). In comparison, in the lung hemorrhagic infarcts are the rule (see Fig. 4.19A). Extravasated red cells in hemorrhagic infarcts are phagocytosed by macrophages, which convert heme iron into hemosiderin; small

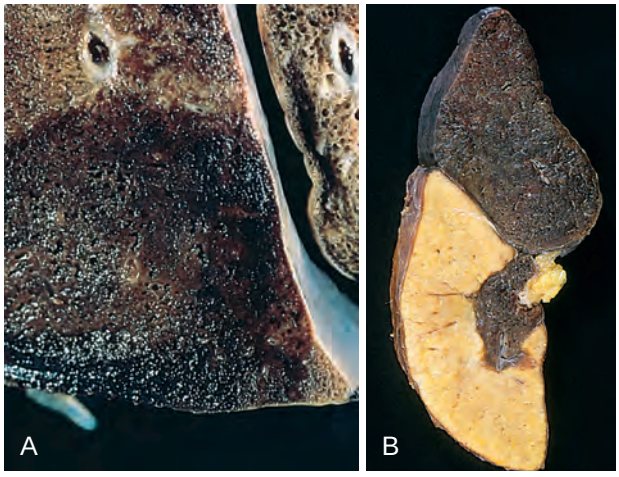


Figure 4.19 Red and white infarcts. (A) Hemorrhagic, roughly wedge-shaped pulmonary red infarct. (B) Sharply demarcated white infarct in the spleen.

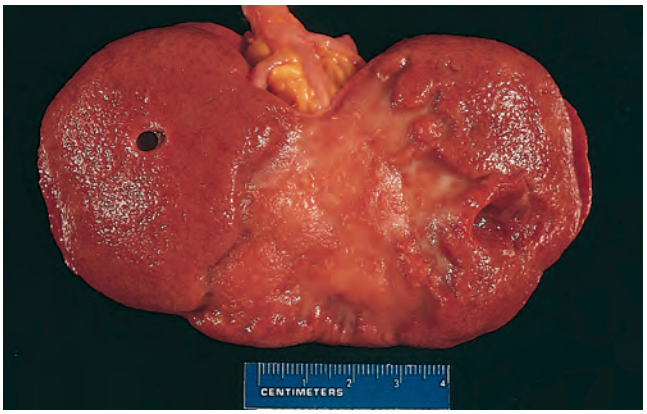


Figure 4.20 Remote kidney infarct replaced by a large fibrotic scar.

amounts do not grossly impart any appreciable color to the tissue, but extensive hemorrhage can leave a firm, brown hemosiderin-rich residuum.

The dominant histologic characteristic of infarction is **ischemic coagulative necrosis** (Chapter 2). Importantly, if the vascular occlusion has occurred shortly (minutes to hours) before the death of the person, histologic changes may be absent; it takes 4 to 12 hours for the dead tissue to show microscopic evidence of necrosis. Acute inflammation is present along the margins of infarcts within a few hours and is usually well defined within 1 to 2 days. Eventually a reparative response begins in the preserved margins (Chapter 3). If the tissue harbors tissue stem cells, parenchymal regeneration can occur at the periphery where underlying stromal architecture is preserved. However, most infarcts are ultimately replaced by scar (Fig. 4.20). The brain is an exception to these generalizations, in that central nervous system infarction results in **liquefactive necrosis** (Chapter 2).

Septic infarctions occur when infected cardiac valve vegetations embolize or when microbes seed necrotic tissue. In these cases the infarct is converted to an abscess, with a correspondingly greater inflammatory response (Chapter 3). The eventual sequence of organization, however, follows the pattern already described.

Factors That Influence Development of an Infarct. A vascular occlusion can cause effects ranging from virtually nothing to tissue dysfunction and necrosis sufficient to result in death. The variables that influence the outcome of vascular occlusion are as follows:

- *Anatomy of the vascular supply.* The availability of an alternative blood supply is the most important determinant of whether vessel occlusion will cause tissue damage. As mentioned earlier, the lungs have a dual pulmonary and bronchial artery blood supply that protects against thromboembolism-induced infarction. Similarly, the liver, with its dual hepatic artery and portal vein circulation, and the hand and forearm, with their dual radial and ulnar arterial supply, are all relatively resistant to infarction. In contrast, renal and splenic circulations are end-arterial, and vascular obstruction generally causes tissue death.
- *Rate of occlusion.* Slowly developing occlusions are less likely to cause infarction, because they provide time for development of collateral pathways of perfusion. For example, small interarteriolar anastomoses—normally with minimal functional flow—interconnect the three major coronary arteries in the heart. If one of the coronaries is occluded slowly (i.e., by an encroaching atherosclerotic plaque), flow within this collateral circulation may increase sufficiently to prevent infarction, even though the larger coronary artery is eventually occluded.
- *Tissue vulnerability to hypoxia.* Neurons undergo irreversible damage when deprived of their blood supply for only 3 to 4 minutes. Myocardial cells, although hardier than neurons, are also quite sensitive and die after only 20 to 30 minutes of ischemia (although, as mentioned earlier, changes in the appearance of the dead cells take 4 to 12 hours to develop). In contrast, fibroblasts within myocardium remain viable even after many hours of ischemia (Chapter 12).
- *Hypoxemia.* Understandably, abnormally low blood oxygen content (regardless of cause) increases both the likelihood and extent of infarction.

KEY CONCEPTS
INFARCTION

- Infarcts are areas of ischemic necrosis most commonly caused by arterial occlusion (typically due to thrombosis or embolization); venous outflow obstruction is a less frequent cause.
- Infarcts caused by venous occlusion or occurring in spongy tissues with dual blood supply and where blood can collect typically are hemorrhagic (red); those caused by arterial occlusion in compact tissues typically are pale (white).
- Whether or not vascular occlusion causes tissue infarction is influenced by collateral blood supply, the rate at which an obstruction develops, intrinsic tissue susceptibility to ischemic injury, and blood oxygenation.

SHOCK

Shock is a state of circulatory failure that impairs tissue perfusion and leads to cellular hypoxia. At the outset, the cellular injury is reversible; however, prolonged shock eventually leads to irreversible tissue injury and can be fatal. Shock may complicate severe hemorrhage, extensive trauma or burns, myocardial infarction, pulmonary embolism, and microbial sepsis. Its causes fall into three general categories (Table 4.3):

- *Cardiogenic shock* results from low cardiac output due to myocardial pump failure. This can be due to intrinsic myocardial damage (infarction), ventricular arrhythmias, extrinsic compression (cardiac tamponade; Chapter 11), or outflow obstruction (e.g., pulmonary embolism).
- *Hypovolemic shock* results from low cardiac output due to low blood volume, such as can occur with massive hemorrhage or fluid loss from severe burns.
- *Sepsis, septic shock, and the systemic inflammatory response syndrome* are interrelated and somewhat overlapping conditions. The definitions that follow are based on The Third International Consensus Definitions for Sepsis and Septic Shock (2016):
  - *Sepsis* is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection.

Table 4.3 Three Major Types of Shock

Type of Shock	Clinical Example	Principle Mechanisms
Cardiogenic	Myocardial infarction Ventricular rupture Arrhythmia Cardiac tamponade Pulmonary embolism	Failure of myocardial pump resulting from intrinsic myocardial damage, extrinsic compression, or obstruction to outflow
Hypovolemic	Fluid loss (e.g., hemorrhage, vomiting, diarrhea, burns, or trauma)	Inadequate blood or plasma volume
Shock associated with systemic inflammation	Overwhelming microbial infections (bacterial and fungal) Superantigens (e.g., toxic shock syndrome) Trauma, burns, pancreatitis	Activation of cytokine cascades; peripheral vasodilation and pooling of blood; endothelial activation/injury; leukocyte-induced damage, disseminated intravascular coagulation

- *Septic shock* is defined as a subset of sepsis in which particularly profound circulatory, cellular, and metabolic abnormalities are associated with a greater risk of mortality than with sepsis alone.
- *Systemic inflammatory response syndrome (SIRS)* is a sepsis-like condition associated with systemic inflammation that may be triggered by a variety of nonmicrobial insults, such as burns, trauma, and/or pancreatitis. Pathogenic feature common to SIRS and septic shock is a massive outpouring of inflammatory mediators from innate and adaptive immune cells that produce arterial vasodilation, vascular leakage, and venous blood pooling. With the advent of chimeric antigen receptor T-cell (CAR-T) therapy, a similar iatrogenic syndrome called *cytokine release syndrome* has been observed in cancer patients. The cardiovascular abnormalities associated with SIRS result in tissue hypoperfusion, cellular hypoxia, and metabolic derangements that lead to organ dysfunction and, if severe and persistent, organ failure and death.

- Less commonly, shock can occur in the setting of a spinal cord injury *(neurogenic shock),* or an IgE-mediated hypersensitivity reaction (*anaphylactic shock,* Chapter 6). In both of these forms of shock, acute vasodilation leads to hypotension and tissue hypoperfusion.

Pathogenesis of Septic Shock

Septic shock is responsible for 2% of all hospital admissions in the United States. Of these, 50% require treatment in intensive care units. The number of cases in the United States exceeds 750,000 per year, and the incidence is rising, which is ironically due to improvements in life support for critically ill patients, as well as the growing ranks of immunocompromised hosts (because of chemotherapy, immunosuppression, advanced age, or human immunodeficiency virus infection) and the increasing prevalence of multidrug-resistant organisms in the hospital setting. Despite improvements in care, the mortality rate remains around 50%. Septic shock is most frequently triggered by gram-positive bacterial infections, followed by gram-negative bacteria and fungi.

The ability of diverse microorganisms to cause septic shock is consistent with the idea that a variety of microbial constituents can trigger the process. As mentioned in Chapter 3, macrophages, neutrophils, dendritic cells, endothelial cells, and soluble components of the innate immune system (e.g., complement) recognize and are activated by diverse substances derived from microorganisms. After activation, these cells and factors initiate a number of inflammatory and counter-inflammatory responses that interact in a complex, incompletely understood fashion to produce septic shock and organ failure (Fig. 4.21).

Factors believed to play major roles in the pathophysiology of septic shock include the following:

- *Inflammatory and counter-inflammatory responses*. In sepsis, various microbial cell wall constituents engage receptors on cells of the innate immune system, triggering pro-inflammatory responses. Likely initiators of inflammation in sepsis are signaling pathways that lie downstream of Toll-like receptors (TLRs) (Chapter 5), which recognize a host of microbe-derived substances containing *pathogen-associated molecular patterns* (PAMPs) and *damage-associated molecular patterns* (DAMPs), as well as G-protein–coupled receptors that detect bacterial peptides and C-type lectin receptors such as Dectins. Ligation of these receptors leads to increased expression of the genes encoding inflammatory mediators via activation and nuclear translocation of the transcription factor nuclear factor-κB (NF-κB). The upregulated mediators include numerous cytokines such as tumor necrosis factor (TNF), interleukin 1 (IL-1), IL-12, IL-18, and interferon-γ (IFN-γ), as well as other inflammatory mediators such as high-mobility group box 1 protein (HMGB1). Markers of acute inflammation such as C-reactive protein and *procalcitonin* are also elevated. Reactive oxygen species and lipid mediators such as prostaglandins and platelet-activating factor (PAF) are also elaborated. These effector molecules induce endothelial cells (and other cell types) to upregulate adhesion molecule expression and further stimulate cytokine and chemokine production. The complement cascade is also activated by microbial components, both directly and through the proteolytic activity of plasmin (Chapter 3), resulting in the production of anaphylotoxins (C3a, C5a), chemotactic fragments (C5a), and opsonins (C3b), all of which contribute to the pro-inflammatory state. In addition, microbial components can activate coagulation directly through factor XII and indirectly through altered endothelial function (discussed later). The accompanying widespread activation of thrombin may further augment inflammation by triggering protease-activated receptors on inflammatory cells.

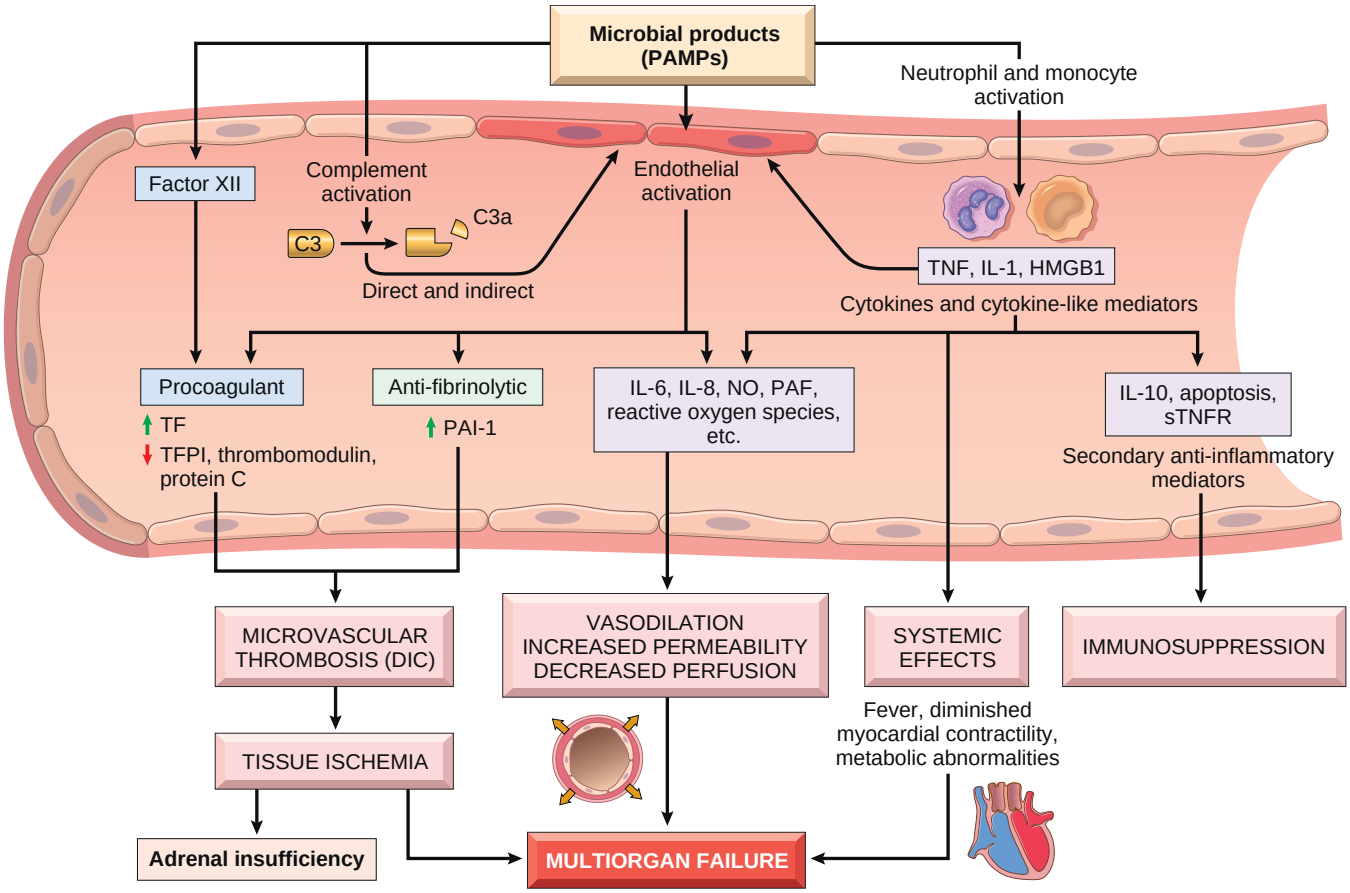


Figure 4.21 Major pathogenic pathways in septic shock. Microbial products (*PAMPs*, or pathogen-associated molecular patterns) activate endothelial cells and cellular and humoral elements of the innate immune system, initiating a cascade of events that lead to end-stage multiorgan failure. Additional details are given in the text. *C3*, Complement component 3; *C3a*, complement component 3a; *DIC*, disseminated intravascular coagulation; *HMGB1*, high-mobility group box 1 protein; *IL*, interleukin; *NO*, nitric oxide; *PAF*, platelet-activating factor; *PAI-1*, plasminogen activator inhibitor 1; *sTNFR*, soluble tumor necrosis factor receptor; *TF*, tissue factor; *TFPI*, tissue factor pathway inhibitor.

The hyperinflammatory state, initiated by sepsis, triggers counter-regulatory immunosuppressive mechanisms, which may involve both innate and adaptive immune cells. As a result, septic patients may oscillate between hyperinflammatory and immunosuppressed states during their clinical course. Proposed mechanisms for the immune suppression include a shift from pro-inflammatory (Th1) to anti-inflammatory (Th2) cytokines (Chapter 5), production of anti-inflammatory mediators (e.g., soluble TNF receptor, IL-1 receptor antagonist, and IL-10), lymphocyte apoptosis, the immunosuppressive effects of apoptotic cells, and the induction of cellular anergy.

Current evidence suggests that both are triggered simultaneously. The intensity of each of these reactions depends on multiple factors both intrinsic to the host (e.g., genetics and underlying diseases) and pathogen (e.g., virulence and burden).

- *Endothelial activation and injury*. The pro-inflammatory state and endothelial cell activation associated with sepsis lead to widespread vascular leakage and tissue edema, which have deleterious effects on both nutrient delivery and waste removal. One effect of inflammatory cytokines is to loosen endothelial cell tight junctions, making vessels leaky and resulting in the accumulation of protein-rich edema fluid throughout the body. This alteration impedes tissue perfusion and may be exacerbated by attempts to support the patient with intravenous fluids. Activated endothelium also upregulates production of NO and other vasoactive inflammatory mediators (e.g., C3a, C5a, and PAF), which may contribute to vascular smooth muscle relaxation and systemic hypotension. Another feature of sepsis is microvascular dysfunction. There is an increase in capillaries with intermittent flow, and heterogeneity of flow in various capillary beds, and the normal autoregulation of flow based on tissue metabolic environment is lost. These changes cause a mismatch in oxygen needs and oxygen delivery.
- *Induction of a procoagulant state*. The derangement in coagulation is sufficient to produce the formidable complication of *DIC* in up to one-half of septic patients. Sepsis alters the expression of many factors so as to favor coagulation. Pro-inflammatory cytokines increase tissue factor production by monocytes and possibly endothelial

cells as well, and decrease the production of endothelial anticoagulant factors, such as tissue factor pathway inhibitor, thrombomodulin, and protein C (see Fig. 4.10). They also dampen fibrinolysis by increasing PAI-1 expression (see Fig. 4.10). Some studies suggest a role for neutrophil extracellular traps (NETs, Chapter 3) in promoting the procoagulant state by stimulating both intrinsic and extrinsic pathways of coagulation. The vascular leak and tissue edema decrease blood flow at the level of small vessels, producing stasis and diminishing the washout of activated coagulation factors. Acting in concert, these effects lead to systemic activation of thrombin and the deposition of fibrin-rich thrombi in small vessels, often throughout the body, further compromising tissue perfusion. In full-blown DIC, the consumption of coagulation factors and platelets is so great that deficiencies of these factors appear, leading to concomitant bleeding and hemorrhage (Chapter 14).

- *Metabolic abnormalities*. Septic patients exhibit insulin resistance and hyperglycemia. Cytokines such as TNF and IL-1, stress-induced hormones (such as glucagon, growth hormone, and glucocorticoids), and catecholamines all drive gluconeogenesis. At the same time, the pro-inflammatory cytokines suppress insulin release while simultaneously promoting insulin resistance in the liver and other tissues, likely by impairing the surface expression of glucose transporter-4 (GLUT-4). Hyperglycemia decreases neutrophil function—thereby suppressing bactericidal activity—and causes increased adhesion molecule expression on endothelial cells. Although sepsis is initially associated with an acute surge in glucocorticoid production, this phase may be followed by adrenal insufficiency and a functional deficit of glucocorticoids. This may stem from depression of the synthetic capacity of intact adrenal glands or frank adrenal necrosis resulting from DIC *(Waterhouse-Friderichsen syndrome)* (Chapter 20). Finally, cellular hypoxia and diminished oxidative phosphorylation lead to increased lactate production and lactic acidosis.
- *Organ dysfunction*. Systemic hypotension, interstitial edema, microvascular dysfunction, and small vessel thrombosis all decrease the delivery of oxygen and nutrients to the tissues that, because of cellular hypoxia, fail to properly use those nutrients that are delivered. Mitochondrial damage resulting from oxidative stress impairs oxygen use. High levels of cytokines and secondary mediators diminish myocardial contractility and cardiac output; increased vascular permeability and endothelial injury can lead to the *acute respiratory distress syndrome* (Chapter 13). Ultimately, these factors may conspire to cause failure of multiple organs, particularly the kidneys, liver, lungs, and heart, culminating in death.

The severity and outcome of septic shock are likely dependent on the extent and virulence of the infection; the immune status of the host; the presence of other comorbid conditions; and the pattern and level of mediator production. The multiplicity of factors and the complexity of the interactions that underlie sepsis explain why most attempts to intervene therapeutically with antagonists of specific mediators have not been effective and even had deleterious effects in some cases. As mentioned earlier, another factor in the failure of neutralization of pro-inflammatory cytokines is the concurrent activation of pro-inflammatory and anti-inflammatory mediators. The standard of care remains antibiotics to treat the underlying infection and intravenous fluids, pressors, and supplemental oxygen to maintain blood pressure and limit tissue hypoxia. Suffice it to say that even in the best of clinical centers, septic shock remains an obstinate clinical challenge.

An additional group of secreted bacterial proteins called *superantigens* also cause a syndrome similar to septic shock (e.g., *toxic shock syndrome*). Superantigens are polyclonal T-lymphocyte activators that induce the release of high levels of cytokines that result in a variety of clinical manifestations, ranging from a diffuse rash to vasodilation, hypotension, shock, and death.

Stages of Shock

Shock is a progressive disorder that leads to death if the underlying problems are not corrected. The exact mechanisms of sepsis-related death are still unclear; aside from increased lymphocyte and enterocyte apoptosis, cellular necrosis is minimal. Death typically follows the failure of multiple organs, which usually offer no morphologic clues to explain their dysfunction. For hypovolemic and cardiogenic shock, however, the pathways leading to a patient's demise are reasonably well understood. Unless the insult is massive and rapidly lethal (e.g., exsanguination from a ruptured aortic aneurysm), shock tends to evolve through three general (albeit somewhat artificial) stages. These stages have been documented most clearly in hypovolemic shock but are common to other forms as well:

- *An initial nonprogressive stage* during which reflex compensatory mechanisms are activated and vital organ perfusion is maintained
- *A progressive stage* characterized by tissue hypoperfusion and onset of worsening circulatory and metabolic derangement, including acidosis
- *An irreversible stage* in which cellular and tissue injury is so severe that even if the hemodynamic defects are corrected, survival is not possible

In the early nonprogressive phase of shock, various neurohumoral mechanisms help maintain cardiac output and blood pressure. These mechanisms include baroreceptor reflexes, release of catecholamines and antidiuretic hormone, activation of the renin-angiotensin-aldosterone axis, and generalized sympathetic stimulation. The net effect is tachycardia, peripheral vasoconstriction, and renal fluid conservation; cutaneous vasoconstriction causes the characteristic "shocky" skin coolness and pallor (notably, septic shock can initially cause cutaneous vasodilation, so the patient may present with warm, flushed skin). Coronary and cerebral vessels are less sensitive to sympathetic signals and maintain relatively normal caliber, blood flow, and oxygen delivery. Thus, blood is shunted away from the skin to the vital organs such as the heart and the brain.

If the underlying causes are not corrected, shock passes imperceptibly to the progressive phase, which as noted is characterized by widespread tissue hypoxia. In the setting

of persistent oxygen deficit, intracellular aerobic respiration is replaced by anaerobic glycolysis with excessive production of lactic acid. The resultant metabolic lactic acidosis lowers the tissue pH, which blunts the vasomotor response; arterioles dilate, and blood begins to pool in the microcirculation. Peripheral pooling not only worsens the cardiac output but also puts endothelial cells at risk for the development of anoxic injury with subsequent DIC. With widespread tissue hypoxia, vital organs are affected and begin to fail.

In the absence of appropriate intervention, or in severe cases, the process eventually enters an irreversible stage. Widespread cell injury is reflected in lysosomal enzyme leakage, further aggravating the shock state. Myocardial contractile function worsens, in part because of increased NO synthesis. The ischemic bowel may allow intestinal flora to enter the circulation, and thus bacteremic shock may be superimposed. Commonly, further progression to renal failure occurs as a consequence of ischemic injury of the kidney (Chapter 14), and despite the best therapeutic interventions, the downward spiral culminates in death.

MORPHOLOGY

The cellular and tissue effects of shock are essentially those of hypoxic injury (Chapter 2) and are caused by a combination of **hypoperfusion** and **microvascular thrombosis.** Although any organ can be affected, the brain, heart, kidneys, adrenals, and gastrointestinal tract are most commonly involved. **Fibrin thrombi** can form in any tissue but typically are most readily visualized in kidney glomeruli. **Adrenal cortical cell lipid depletion** is akin to that seen in all forms of stress and reflects increased use of stored lipids for steroid synthesis. Whereas the lungs are resistant to hypoxic injury in hypovolemic shock occurring after hemorrhage, sepsis or trauma can precipitate diffuse alveolar damage (Chapter 13), leading to so-called **"shock lung."** Except for neuronal and cardiomyocyte loss, affected tissues can recover completely if the patient survives.

Clinical Features

The clinical manifestations of shock depend on the precipitating insult. In hypovolemic and cardiogenic shock, patients exhibit hypotension, a weak rapid pulse, tachypnea, and cool, clammy, cyanotic skin. As already noted, in septic shock, the skin may be warm and flushed owing to peripheral vasodilation. The primary threat to life is the underlying initiating event (e.g., myocardial infarction, severe hemorrhage, bacterial infection). However, the cardiac, cerebral, and pulmonary changes rapidly aggravate the situation. If patients survive the initial insult, worsening renal function can provoke a phase dominated by progressive oliguria, acidosis, and electrolyte imbalances.

Prognosis varies with the origin of shock and its duration. Thus, more than 90% of young, otherwise healthy patients with hypovolemic shock survive with appropriate management; by comparison, septic or cardiogenic shock is associated with substantially poorer outcomes, even with state-of-the-art care.

KEY CONCEPTS

SHOCK

- Shock is defined as a state of systemic tissue hypoperfusion resulting from reduced cardiac output and/or reduced effective circulating blood volume.
- The major types of shock are cardiogenic (e.g., myocardial infarction), hypovolemic (e.g., blood loss), and septic (e.g., infections).
- Shock of any form can lead to hypoxic tissue injury if not corrected.
- Septic shock is caused by a dysregulated host response to bacterial or fungal infections; it is characterized by endothelial cell activation, vasodilation, edema, DIC, and metabolic derangements.

SUGGESTED READINGS

Fluid Dynamics

Chen H, Schrier R: Pathophysiology of volume overload in acute heart failure syndromes, *Am J Med* 119:S11, 2006. *[An older but still useful review of heart failure and fluid overload].*

Hemostasis and Bleeding

Chapman JC, Hajjar KA: Fibrinolysis and the control of blood coagulation, *Blood Rev* 29:17, 2015. *[An updated discussion of fibrinolysis and its role in the regulation of coagulation].*

Ellery PE, Adams MJ: Tissue factor pathway inhibitor: then and now, *Semin Thromb Hemost* 40:881, 2014. *[Advances in understanding the role of tissue factor in coagulation].*

Rao LVM, Pendurthi UR, Rao LVM: Endothelial cell protein C receptor dependent signaling, *Curr Opin Hematol* 25:219, 2018. *[An excellent review of the multiple functions of protein C].*

Rudinga GR et al: Protease-activated receptor 4 (PAR4): a promising target for antiplatelet therapy, *Int J Mol Sci* 19:572, 2018. *[A discussion of the role of protease-activated receptors in hemostasis and for development of drugs that target PARs].*

Versteeg HH, Heemskerk JWM, Levi M et al: New fundamentals in hemostasis, *Physiol Rev* 93:327, 2013. *[An update of several issues in hemostasis].*

Vojacek JF: Should we replace the terms intrinsic and extrinsic coagulation pathways with tissue factor pathway? *Clin Appl Thromb Hemost* 23:922, 2017. *[A short review on the central role of tissue factor in coagulation].*

Thrombosis and Thromboembolism

Coleman DM, Obi A, Henke PK: Update in venous thromboembolism: pathophysiology, diagnosis, and treatment for surgical patients, *Curr Probl Surg* 52:233, 2015. *[An exhaustive review of this common clinical condition].*

Hotoleanu C: Genetic risk factors in venous thromboembolism, *Adv Exp Med Biol – Adv In Internal Med* 906(253):2017. *[An excellent clinical and molecular discussion of genetic thrombophilic states].*

Montagnana M, Lippi G, Danese: An overview of thrombophilia and associated laboratory testing. In Favaloro EJ, Lippi G, editors: *Hemostasis and thrombosis: methods and protocols, methods in molecular biology, 1646*, 2017, p 113. *[A laboratory-oriented review of genetic and acquired factors underlying thrombosis].*

Prince M, Wenham T: Heparin-induced thrombocytopenia, *Postgrad Med J* 94:453, 2018. *[A concise and up-to-date review of this serious clinical disorder].*

Nisio MD, van Es N, Buller H: Deep Vein thrombosis and pulmonary embolism, *Lancet* 388, 2016. *[An excellent review of this important clinical condition].*

Sciascia S, Amigo M-C, Roccatello M et al: Diagnosing antiphospholipid syndrome: 'extra-criteria' manifestations and technical advances, *Nat Rev Rheumatol* 13:549, 2017. *[An excellent discussion of the pathogenesis and clinical features of APS]*.

Unusual Forms of Embolic Disease

Brull SJ, Prielipp RC: Vascular air embolism: a silent hazard to patient safety, *J Crit Care* 42:255, 2017. *[An exhaustive review of an uncommon but serious disorder]*.

Fukumoto LE, Fukomoto KD: Fat embolism syndrome, *Nurs Clin North Am* 335, 2018. *[An excellent discussion of the pathophysiology and clinical features of this disorder]*.

Tamura N, Farhana M, Oda T et al: Amniotic fluid embolism: pathophysiology from the perspective of pathology, *J Obstet Gynaecol Res* 43:627, 2017. *[A discussion of an uncommon syndrome with high mortality]*.

Septic Shock

Cecconi M, Evans L, Levy M et al: Sepsis and septic shock, *Lancet* 392:75, 2018. *[A lucid and comprehensive review of septic shock]*.

Levi M, Poll TVD: Coagulation and sepsis, *Thromb Res* 149:38, 2017. *[A discussion of coagulopathy in sepsis]*.

Moskovitz JB, Levy ZD, Todd LS: Cardiogenic shock, *Emerg Med Clin North Am* 645, 2015. *[A discussion of the pathophysiology of cardiogenic shock]*.

Pfeiler S, Stark K, Massberg F et al: Propagation of thrombosis by neutrophils and extracellular nucleosome networks, *Hematologica* 102:206, 2017. *[The emerging role of NETs in septic shock]*.

Pool R, Gomez H, Kellum JA: Mechanism of organ dysfunction in sepsis, *Crit Care Clin* 34:63, 2018. *[An in-depth discussion of the multiple mechanisms that contribute to organ dysfunction in sepsis]*.

Russell JA, Rush B, Boyd J: Pathophysiology of septic shock, *Crit Care Clin* 34:43, 2017. *[An excellent discussion of the pathogenesis of septic shock]*.

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 5

Genetic Disorders

CHAPTER CONTENTS

GENES AND HUMAN DISEASES

In Chapter 1, we discussed the architecture of the normal human genome. Here we build on that knowledge to discuss the genetic basis of human diseases.

Genetic disorders are far more common than is widely appreciated. The lifetime frequency of genetic diseases is estimated to be 670 per 1000. Furthermore, the genetic diseases encountered in medical practice represent only the tip of the iceberg, that is, those with less extreme genotypic errors that permit full embryonic development and live birth. It is estimated that 50% of spontaneous abortuses during the early months of gestation have a demonstrable chromosomal abnormality; there are, in addition, numerous smaller detectable errors and many other genetic lesions that are only now being recognized thanks to advances in DNA sequencing. About 1% of all newborn infants possess a gross chromosomal abnormality, and serious disease with a significant genetic component develops in approximately 5% of individuals younger than age 25 years. How many more mutations remain hidden?

Before discussing specific aberrations that may cause genetic diseases, it is useful to note that **human genetic disorders can be broadly classified into three categories**.

- *Disorders related to mutations in single genes with large effects.* These mutations cause the disease or predispose to the disease and, with some exceptions like hemoglobinopathies, are typically not present in the normal population. Such mutations and their associated disorders are highly

penetrant, meaning that the presence of the mutation is associated with the disease in a large proportion of individuals. Because these diseases are caused by single gene mutations, they usually follow the classic Mendelian pattern of inheritance and are also referred to as Mendelian disorders. A few important exceptions to this rule are noted later.

Study of single genes and mutations with large effects has been extremely informative in medicine, since a great deal of what is known about several physiologic pathways (e.g., cholesterol transport, chloride secretion) has been learned from analysis of single-gene disorders. Although informative, these disorders are generally rare unless they are maintained in a population by strong selective forces (e.g., sickle cell anemia in areas where malaria is endemic; Chapter 14).

- *Chromosomal disorders*. These arise from structural or numerical alteration in the autosomes and sex chromosomes. Like monogenic disease, they are uncommon but associated with high penetrance.
- *Complex multigenic disorders*. These are far more common than diseases in the previous two categories. They are caused by interactions between multiple variant forms of genes and environmental factors. Such variations in genes are common within the population and are also called *polymorphisms*. Each such variant gene confers a small increase in disease risk, and no single susceptibility gene is necessary or sufficient to produce the disease. It is only when several such polymorphisms are present in an individual that disease occurs—hence the term *multigenic* or *polygenic*. Thus, unlike mutant genes with large effects that are highly penetrant and give rise to Mendelian disorders, each polymorphism has a small effect and is of low penetrance. Since environmental interactions are important in the pathogenesis of these diseases, they are also called multifactorial disorders. In this category are some of the most common diseases that afflict humans, including atherosclerosis, diabetes mellitus, hypertension, and autoimmune diseases. Even normal traits such as height and weight are governed by polymorphisms in several genes.

The following discussion describes mutations that affect single genes, which underlie Mendelian disorders, followed by transmission patterns and selected samples of single-gene disorders.

Mutations

A *mutation* is defined as a permanent change in the DNA. Mutations that affect germ cells are transmitted to the progeny and can give rise to inherited diseases. Mutations that arise in somatic cells understandably do not cause hereditary diseases but are important in the genesis of cancers and some congenital malformations.

General principles relating to the effects of gene mutations follow.

- *Point mutations within coding sequences.* A point mutation is a change in which a single base is substituted with a different base. It may alter the code in a triplet of bases and lead to the replacement of one amino acid by another in the gene product. Because these mutations alter the meaning of the sequence of the encoded protein, they are often termed *missense mutations.* If the substituted amino acid is biochemically similar to the original, typically it causes little change in the function of the protein, and the mutation is called a "conservative" missense mutation. On the other hand, a "nonconservative" missense mutation replaces the normal amino acid with a biochemically different one. An excellent example of this type is the sickle mutation affecting the β-globin chain of hemoglobin (Chapter 14). Here the nucleotide triplet CTC (or GAG in mRNA), which encodes glutamic acid, is changed to CAC (or GUG in mRNA), which encodes valine. This single amino acid substitution alters the physicochemical properties of hemoglobin, giving rise to sickle cell anemia. Besides producing an amino acid substitution, a point mutation may change an amino acid codon to a chain terminator, or stop codon *(nonsense mutation).* Taking again the example of β-globin, a point mutation affecting the codon for glutamine (CAG) creates a stop codon (UAG) if U is substituted for C (Fig. 5.1). This change leads to premature termination of β-globin gene translation, and the short peptide that is produced is rapidly degraded. The resulting deficiency of β-globin chains can give rise to a severe form of anemia called β^0-thalassemia (Chapter 14).
- *Mutations within noncoding sequences.* Deleterious effects may also result from mutations that do not involve the exons. Recall that transcription of DNA is initiated and regulated by promoter and enhancer sequences (Chapter 1). Point mutations or deletions involving these regulatory sequences may interfere with binding of transcription factors and thus lead to a marked reduction in or total lack of transcription. Such is the case in certain forms of hereditary anemias called thalassemias (Chapter 14). In addition, point mutations within introns may lead to defective splicing of intervening sequences. This, in turn, interferes with normal processing of the initial mRNA transcripts and results in a failure to form mature mRNA. Therefore translation cannot take place, and the gene product is not synthesized.
- *Deletions and insertions.* Small deletions or insertions involving the coding sequence can have two possible effects on the encoded protein. If the number of base pairs involved is three or a multiple of three, the reading frame will remain intact, and an abnormal protein lacking

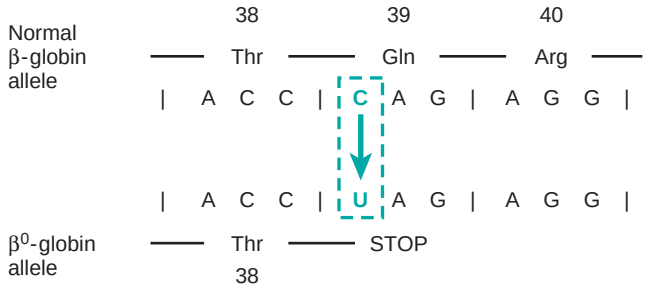


Figure 5.1 Nonsense mutation leading to premature chain termination. Partial mRNA sequence of the β-globin chain of hemoglobin showing codons for amino acids 38 to 40. A point mutation (C → U) in codon 39 changes a glutamine (Gln) codon to a stop codon, and hence protein synthesis stops at amino acid 38.

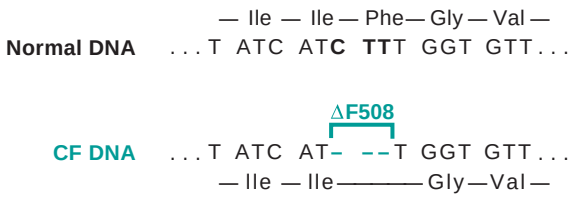


Figure 5.2 Three-base deletion in the common cystic fibrosis *(CF)* allele results in synthesis of a protein that lacks amino acid 508 (phenylalanine). Because the deletion is a multiple of three, this is not a frameshift mutation. (From Thompson MW, et al: *Thompson and Thompson Genetics in Medicine*, ed 5, Philadelphia, 1991, WB Saunders, p 135.)

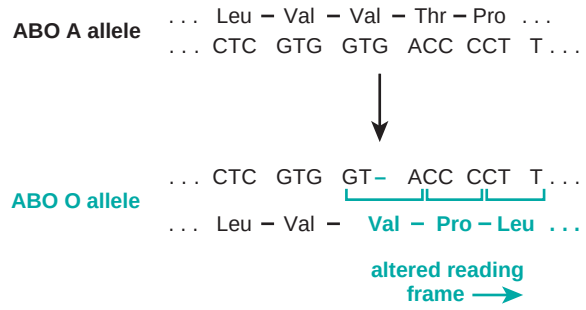


Figure 5.3 Single-base deletion at the ABO (glycosyltransferase) locus, leading to a frameshift mutation responsible for the O allele. (From Thompson MW, et al: *Thompson and Thompson Genetics in Medicine*, ed 5, Philadelphia, 1991, WB Saunders, p 134.)

or gaining one or more amino acids will be synthesized (Fig. 5.2). If the number of affected coding bases is not a multiple of three, this will result in an alteration of the reading frame of the DNA strand, producing what is referred to as a *frameshift mutation* (Figs. 5.3 and 5.4). Typically the result is the incorporation of a variable number of incorrect amino acids followed by truncation resulting from a premature stop codon.

- *Alterations in protein-coding genes other than mutations.* In addition to alterations in DNA sequence, coding genes also can undergo structural variations, such as copy number changes—*amplifications* or *deletions*—or *translocations* that result in aberrant gain or loss of protein function. As with mutations, structural changes may occur in the germline or be acquired in somatic tissues. In many instances, pathogenic germline alterations involve a contiguous portion of a chromosome rather than a single gene, such as in the 22q microdeletion syndrome, discussed later. With the widespread availability of next-generation sequencing (NGS) technology for assessing genome-wide DNA copy number variation at very high resolution, potentially pathogenic structural alterations have now been discovered in common disorders such as autism. Cancers often contain somatically acquired structural alterations, including amplifications, deletions, and translocations. The so-called Philadelphia chromosome—translocation t(9;22) between the *BCR* and *ABL* genes in chronic myeloid leukemia (Chapter 13)—is a classic example.
- *Alterations in noncoding RNAs.* It is worth noting that in the past, the major focus of gene hunting was discovery of genes that encode proteins. We now know that a very large number of genes do not encode proteins but produce transcripts—so-called noncoding RNAs (ncRNAs)—that serve important regulatory functions. Although many distinct families of ncRNAs exist, the two most important examples—small RNA molecules called microRNAs (miRNAs) and long noncoding RNAs (lncRNAs)—are discussed in Chapter 1.
- *Trinucleotide-repeat mutations.* Trinucleotide-repeat mutations belong to a special category of genetic anomaly. These mutations are characterized by amplification of a sequence of three nucleotides. Although the specific nucleotide sequence that undergoes amplification differs in various disorders, almost all affected sequences share the nucleotides guanine (G) and cytosine (C). For example, in fragile X syndrome (FXS), prototypical of this category of disorders, there are 250 to 4000 tandem repeats of the sequence CGG within the regulatory region of a gene called *familial mental retardation 1 (FMR1).* In normal populations the number of repeats is small, averaging 29. Such expansions of the trinucleotide sequences prevent normal expression of the *FMR1* gene, thus giving rise to intellectual disability. Another *distinguishing feature of trinucleotide-repeat mutations is that they are dynamic* (i.e., the degree of amplification increases during gametogenesis). These features, discussed in greater detail later, influence the pattern of inheritance and the phenotypic manifestations of the diseases caused by this class of mutation.

To summarize, mutations can interfere with gene expression at various levels. Transcription may be suppressed by gene deletions and point mutations involving promoter sequences. Abnormal mRNA processing may result from mutations affecting introns or splice junctions or both. Translation is affected if a nonsense mutation creates a stop

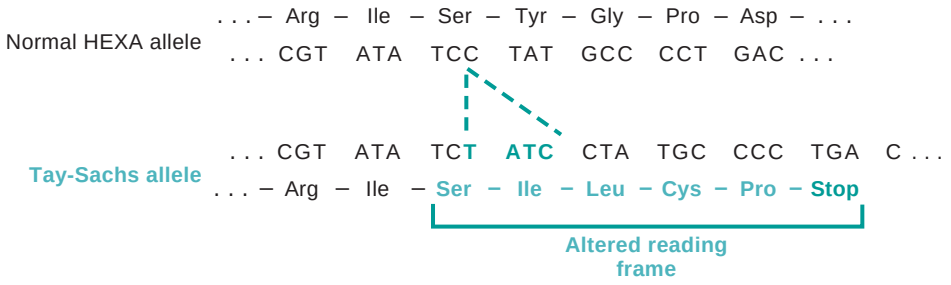


Figure 5.4 Four-base insertion in the hexosaminidase A gene, leading to a frameshift mutation. This mutation is the major cause of Tay-Sachs disease in Ashkenazi Jews. (From Nussbaum RL, et al: *Thompson and Thompson Genetics in Medicine*, ed 6, Philadelphia, 2001, WB Saunders, p 212.)

codon (chain termination mutation) within an exon. Finally, some pathogenic point mutations may lead to expression of normal amounts of a dysfunctional protein.

Against this background, we now turn our attention to the three major categories of genetic disorders: (1) disorders related to mutant genes of large effect, (2) diseases with multifactorial inheritance, and (3) chromosomal disorders. To these three well-known categories must be added a heterogeneous group of *single-gene disorders with nonclassic patterns of inheritance.* This group includes disorders resulting from triplet-repeat mutations, those arising from mutations in mitochondrial DNA (mtDNA), and those in which the transmission is influenced by genomic imprinting or gonadal mosaicism. Diseases within this group are caused by mutations in single genes, but they do not follow the Mendelian pattern of inheritance. These are discussed later in this chapter.

It is beyond the scope of this book to review normal human genetics. Some fundamentals of DNA structure and regulation of gene expressions were described in Chapter 1. It is important here to clarify several commonly used terms—*hereditary, familial,* and *congenital.* Hereditary disorders, by definition, are derived from one's parents and are transmitted in the germline through the generations and therefore are familial. The term *congenital* simply implies "born with." Some congenital diseases are not genetic (e.g., congenital syphilis). Not all genetic diseases are congenital; individuals with Huntington disease, for example, begin to manifest their condition only after their 20s or 30s.

MENDELIAN DISORDERS

Virtually all Mendelian disorders are the result of mutations in single genes that have large effects. It is not necessary to detail Mendel's laws here, since every student in biology, and possibly every garden pea, has learned about them at an early age. Only some comments of medical relevance are made.

It is estimated that every individual is a carrier of several deleterious genes; most of these genes are recessive and therefore do not have serious phenotypic effects. About 80% to 85% of these mutations are familial. The remainder represent new mutations acquired de novo by an affected individual.

Most mutations in autosomal genes produce partial expression in the heterozygote and full expression in the homozygote. Sickle cell anemia is caused by substitution of normal hemoglobin (HbA) by hemoglobin S (HbS). When an individual is homozygous for the mutant gene, all the hemoglobin is of the abnormal, HbS, type, and even with normal saturation of oxygen the disorder is fully expressed (i.e., sickling deformity of all red cells and hemolytic anemia). In the heterozygote, only a proportion of the hemoglobin is HbS (the remainder being HbA), and therefore red cell sickling occurs only under unusual circumstances, such as exposure to lowered oxygen tension. This is referred to as the *sickle cell trait* to differentiate it from full-blown sickle cell anemia.

Although Mendelian traits are usually described as dominant or recessive, in some cases both of the alleles of a gene pair contribute to the phenotype—a condition called *codominance.* Histocompatibility and blood group antigens are good examples of codominant inheritance.

A single mutant gene may lead to many end effects, termed *pleiotropism;* conversely, mutations at several genetic loci may produce the same trait *(genetic heterogeneity).* Sickle cell anemia is an example of pleiotropism. In this hereditary disorder, not only does the point mutation in the gene give rise to HbS, which predisposes the red cells to hemolysis, but also the abnormal red cells tend to cause a logjam in small vessels, inducing, for example, splenic fibrosis, organ infarcts, and bone changes. The numerous differing end-organ derangements all are related to the primary defect in hemoglobin synthesis. On the other hand, profound childhood deafness, an apparently homogeneous clinical entity, results from many different types of autosomal recessive mutations. Recognition of genetic heterogeneity not only is important in genetic counseling but also is relevant in the understanding of the pathogenesis of some common disorders, such as diabetes mellitus.

Transmission Patterns of Single-Gene Disorders

Mutations involving single genes typically follow one of three patterns of inheritance: autosomal dominant, autosomal recessive, and X-linked. The general rules that govern the transmission of single-gene disorders are well known; only a few salient features are summarized. Single-gene disorders with nonclassic patterns of inheritance are described later.

Autosomal Dominant Disorders

Autosomal dominant disorders are manifested in the heterozygous state, so at least one parent of an index case is usually affected; both males and females are affected, and both can transmit the condition. When an affected person marries an unaffected one, every child has one chance in two of having the disease. In addition to these basic rules, autosomal dominant conditions are characterized by the following:

- *With every autosomal dominant disorder, some proportion of patients do not have affected parents.* Such patients owe their disorder to new mutations involving either the egg or the sperm from which they were derived. Their siblings are neither affected nor at increased risk for disease development. The proportion of patients who develop the disease as a result of a new mutation is related to the effect of the disease on reproductive capability. If a disease markedly reduces reproductive fitness, most cases would be expected to result from new mutations. Many new mutations seem to occur in germ cells of relatively older fathers.
- *Clinical features can be modified by variations in penetrance and expressivity.* Some individuals inherit the mutant gene but are phenotypically normal. This is referred to as *incomplete penetrance.* Penetrance is expressed in mathematical terms. Thus 50% penetrance indicates that 50% of those who carry the gene express the trait. In contrast to penetrance, if a trait is seen in all individuals carrying the mutant gene but is expressed differently among individuals, the phenomenon is called *variable expressivity.* For example, manifestations of neurofibromatosis type 1 range from brownish spots on the skin to multiple skin tumors and skeletal deformities. The mechanisms underlying

incomplete penetrance and variable expressivity are not fully understood, but they most likely result from effects of other genes or environmental factors that modify the phenotypic expression of the mutant allele. For example, the phenotype of a patient with sickle cell anemia (resulting from mutation at the β-globin locus) is influenced by the genotype at the α-globin locus because the latter influences the total amount of hemoglobin made (Chapter 14). The influence of environmental factors is exemplified by individuals heterozygous for familial hypercholesterolemia (FH). The expression of the disease in the form of atherosclerosis is conditioned by the dietary intake of lipids.

- In many conditions the age at onset is delayed; symptoms and signs may not appear until adulthood (as in Huntington disease).

The molecular mechanisms of autosomal dominant disorders depend on the nature of the mutation and the type of protein affected. Most mutations lead to the reduced production of a gene product or give rise to a dysfunctional or inactive protein. Whether such a mutation gives rise to dominant or recessive disease depends on whether the remaining copy of the gene is capable of compensating for the loss. Thus, understanding the reasons why particular loss-of-function mutations give rise to dominant versus recessive disease patterns requires an understanding of the biology. Many autosomal dominant diseases arising from deleterious mutations fall into one of a few familiar patterns:

1. *Diseases involved in regulation of complex metabolic pathways that are subject to feedback inhibition.* Membrane receptors such as the low-density lipoprotein (LDL) receptor provide one such example; in FH, discussed later, a 50% loss of LDL receptors results in a secondary elevation of cholesterol that, in turn, predisposes to atherosclerosis in affected heterozygotes.
2. *Key structural proteins, such as collagen and cytoskeletal elements of the red cell membrane* (e.g., spectrin). The biochemical mechanisms by which a 50% reduction in the amounts of such proteins results in an abnormal phenotype are not fully understood. In some cases, especially when the gene encodes one subunit of a multimeric protein, the product of the mutant allele can interfere with the assembly of a functionally normal multimer. For example, the collagen molecule is a trimer in which the three collagen chains are arranged in a helical configuration. Each of the three collagen chains in the helix must be normal for the assembly and stability of the collagen molecule. Even with a single mutant collagen chain, normal collagen trimers cannot be formed, and hence there is a marked deficiency of collagen. In this instance the mutant allele is called *dominant negative* because it impairs the function of a normal allele. This effect is illustrated by some forms of osteogenesis imperfecta, characterized by marked deficiency of collagen and severe skeletal abnormalities (Chapter 26).

Less common than loss-of-function mutations are *gain-of-function* mutations, which can take two forms. Some mutations result in an increase in normal function of a protein, for example, excessive enzymatic activity. In other cases, mutations impart a wholly new activity completely unrelated to the affected protein's normal function. The transmission of disorders produced by gain-of-function mutations is almost always autosomal dominant, as illustrated by Huntington disease (Chapter 28). In this disease the trinucleotide-repeat mutation affecting the Huntington gene (see later) gives rise to an abnormal protein, called *huntingtin,* that is toxic to neurons, and hence even heterozygotes develop a neurologic deficit.

Table 5.1 lists common autosomal dominant disorders. Many are discussed more logically in other chapters. A few conditions not considered elsewhere are discussed later in this chapter to illustrate important principles.

Autosomal Recessive Disorders

Autosomal recessive traits make up the largest category of Mendelian disorders. They occur when both alleles at a given gene locus are mutated. These disorders are characterized by the following features: (1) The trait does not usually affect the parents of the affected individual, but siblings may show the disease; (2) siblings have one chance in four of having the trait (i.e., the recurrence risk is 25% for each birth); and (3) if the mutant gene occurs with a low frequency in the population, there is a strong likelihood that the affected individual (proband) is the product of a consanguineous marriage. The following features generally apply to most autosomal recessive disorders and distinguish them from autosomal dominant diseases:

- The expression of the defect tends to be more uniform than in autosomal dominant disorders.
- Complete penetrance is common.
- Onset is frequently early in life.
- Although new mutations associated with recessive disorders do occur, they are rarely detected clinically. Since the individual with a new mutation is an asymptomatic heterozygote, several generations may pass before the descendants of such a person mate with other heterozygotes and produce affected offspring.
- Many of the mutated genes encode enzymes. In heterozygotes, equal amounts of normal and defective enzyme are synthesized. Usually the natural "margin of safety"

Table 5.1 Autosomal Dominant Disorders

System	Disorder
Nervous	Huntington disease Neurofibromatosis Myotonic dystrophy Tuberous sclerosis
Urinary	Polycystic kidney disease
Gastrointestinal	Familial polyposis coli
Hematopoietic	Hereditary spherocytosis von Willebrand disease
Skeletal	Marfan syndrome[a] Ehlers-Danlos syndrome (some variants)[a] Osteogenesis imperfecta Achondroplasia
Metabolic	Familial hypercholesterolemia[a] Acute intermittent porphyria

[a]Discussed in this chapter. Other disorders listed are discussed in appropriate chapters in the book.

Table 5.2 Autosomal Recessive Disorders

System	Disorder
Metabolic	Cystic fibrosis Phenylketonuria Galactosemia Homocystinuria Lysosomal storage diseases[a] α_1-Antitrypsin deficiency Wilson disease Hemochromatosis Glycogen storage diseases[a]
Hematopoietic	Sickle cell anemia Thalassemias
Endocrine	Congenital adrenal hyperplasia
Skeletal	Ehlers-Danlos syndrome (some variants)[a] Alkaptonuria
Nervous	Neurogenic muscular atrophies Friedreich ataxia Spinal muscular atrophy

[a]Discussed in this chapter. Many others are discussed elsewhere in the text.

ensures that cells with half the usual complement of the enzyme function normally.

Autosomal recessive disorders include almost all inborn errors of metabolism. The various consequences of enzyme deficiencies are discussed later. The more common of these conditions are listed in Table 5.2. Most are presented elsewhere; a few prototypes are discussed later in this chapter.

X-Linked Disorders

All sex-linked disorders are X-linked, and almost all are recessive. Several genes are located in the male-specific region of Y; all of these are related to spermatogenesis. *Males with mutations affecting the Y-linked genes are usually infertile, and hence there is no Y-linked inheritance.* As discussed later, a few additional genes with homologs on the X chromosome have been mapped to the Y chromosome, but only a few rare disorders resulting from mutations in such genes have been described.

X-linked recessive inheritance accounts for a small number of well-defined clinical conditions. The Y chromosome, for the most part, is not homologous to the X chromosome, and so mutant genes on the X chromosome do not have corresponding alleles on the Y chromosome. Thus, males are said to be *hemizygous* for X-linked mutant genes, so these disorders are expressed in males. Other features that characterize these disorders are as follows:

- An affected male does not transmit the disorder to his sons, but all daughters are carriers. Sons of heterozygous women have, of course, one chance in two of receiving the mutant gene.
- A heterozygous female usually does not express the full phenotypic change because of the paired normal allele. Because of the random inactivation of one of the X chromosomes in the female, however, females have a variable proportion of cells in which the mutant X chromosome is active. Thus it is remotely possible for the normal allele to be inactivated in most cells, permitting full expression of heterozygous X-linked conditions in females. Much more commonly, the normal allele is inactivated in only some of the cells, and thus a heterozygous female expresses the disorder partially. An illustrative condition is *glucose-6-phosphate dehydrogenase (G6PD) deficiency*. Transmitted on the X chromosome, this enzyme deficiency, which predisposes to red cell hemolysis in patients receiving certain types of drugs (Chapter 14), is expressed principally in males. In females, a proportion of the red cells may be derived from precursors with inactivation of the normal allele. Such red cells are at the same risk for undergoing hemolysis as the red cells in hemizygous males. Thus, females are not only carriers of this trait but also are susceptible to drug-induced hemolytic reactions. Because the proportion of defective red cells in heterozygous females depends on the random inactivation of one of the X chromosomes, however, the severity of the hemolytic reaction is almost always less in heterozygous women than in hemizygous men. Most of the X-linked conditions listed in Table 5.3 are covered elsewhere in the text.

There are only a few *X-linked dominant* conditions. They are caused by dominant disease-associated alleles on the X chromosome. These disorders are transmitted by an affected heterozygous female to half her sons and half her daughters and by an affected male parent to all his daughters but none of his sons, if the female parent is unaffected. Vitamin D–resistant rickets and Alport syndrome are examples of this type of inheritance.

KEY CONCEPTS

TRANSMISSION PATTERNS OF SINGLE-GENE DISORDERS

- Autosomal dominant disorders are characterized by expression in heterozygous state; they affect males and females equally, and both sexes can transmit the disorder.
- Enzyme proteins are not affected in autosomal dominant disorders; instead, receptors and structural proteins are involved.
- Autosomal recessive diseases occur when both copies of a gene are mutated; enzyme proteins are frequently involved. Males and females are affected equally.
- X-linked disorders are transmitted by heterozygous females to their sons, who manifest the disease. Female carriers usually are protected because of random inactivation of one X chromosome.

Table 5.3 X-Linked Recessive Disorders

System	Disease
Musculoskeletal	Duchenne muscular dystrophy
Blood	Hemophilia A and B Chronic granulomatous disease Glucose-6-phosphate dehydrogenase deficiency
Immune	Agammaglobulinemia Wiskott-Aldrich syndrome
Metabolic	Diabetes insipidus Lesch-Nyhan syndrome
Nervous	Fragile X syndrome[a]

[a]Discussed in this chapter. Others are discussed in appropriate chapters in the text.

Biochemical and Molecular Basis of Single-Gene (Mendelian) Disorders

Mendelian disorders result from alterations involving single genes. The genetic defect may lead to the formation of an abnormal protein or a reduction in the output of the gene product. Virtually any type of protein may be affected in single-gene disorders and by a variety of mechanisms (Table 5.4). To some extent the pattern of inheritance of the disease is related to the kind of protein affected by the mutation. For this discussion, the mechanisms involved in single-gene disorders can be classified into four categories: (1) enzyme defects and their consequences; (2) defects in membrane receptors and transport systems; (3) alterations in the structure, function, or quantity of nonenzyme proteins; and (4) mutations resulting in unusual reactions to drugs.

Enzyme Defects and Their Consequences

Mutations may result in the synthesis of an enzyme with reduced activity or a reduced amount of a normal enzyme. In either case, the consequence is a metabolic block. Fig. 5.5 provides an example of an enzyme reaction in which the substrate is converted by intracellular enzymes, denoted as 1, 2, and 3, into an end product through intermediates 1 and 2. In this model the final product exerts feedback control on enzyme 1. A minor pathway producing small quantities of M1 and M2 also exists. The biochemical consequences of an enzyme defect in such a reaction may lead to three major consequences:

- *Accumulation of the substrate,* depending on the site of block, may be accompanied by accumulation of one or both intermediates. Moreover, an increased concentration of intermediate 2 may stimulate the minor pathway and

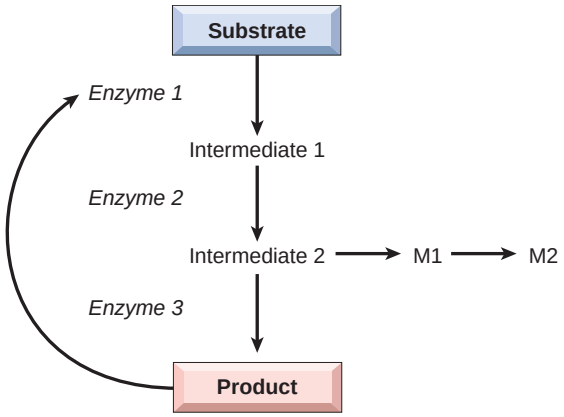


Figure 5.5 A possible metabolic pathway in which a substrate is converted to an end product by a series of enzyme reactions. *M1, M2,* Products of a minor pathway.

Table 5.4 Biochemical and Molecular Basis of Some Mendelian Disorders

Protein Type/Function	Example	Molecular Lesion	Disease
Enzyme	Phenylalanine hydroxylase	Splice-site mutation: reduced amount	Phenylketonuria
	Hexosaminidase A	Splice-site mutation or frameshift mutation with stop codon: reduced amount	Tay-Sachs disease
	Adenosine deaminase	Point mutations: abnormal protein with reduced activity	Severe combined immunodeficiency
Enzyme inhibitor	α_1-Antitrypsin	Missense mutations: impaired secretion from liver to serum	Emphysema and liver disease
Receptor	Low-density lipoprotein receptor	Deletions, point mutations: reduction of synthesis, transport to cell surface, or binding to low-density lipoprotein	Familial hypercholesterolemia
	Vitamin D receptor	Point mutations: failure of normal signaling	Vitamin D–resistant rickets
Transport			
Oxygen	Hemoglobin	Deletions: reduced amount	α-Thalassemia
		Defective mRNA processing: reduced amount	β-Thalassemia
		Point mutations: abnormal structure	Sickle cell anemia
Ion channels	Cystic fibrosis transmembrane conductance regulator	Deletions and other mutations: nonfunctional or misfolded proteins	Cystic fibrosis
Structural			
Extracellular	Collagen	Deletions or point mutations cause reduced amount of normal collagen or normal amounts of defective collagen	Osteogenesis imperfecta Ehlers-Danlos syndromes
	Fibrillin	Missense mutations	Marfan syndrome
Cell membrane	Dystrophin	Deletion with reduced synthesis	Duchenne/Becker muscular dystrophy
	Spectrin, ankyrin, or protein 4.1	Heterogeneous	Hereditary spherocytosis
Hemostasis	Factor VIII	Deletions, insertions, nonsense mutations, and others: reduced synthesis or abnormal factor VIII	Hemophilia A
Growth regulation	Rb protein	Deletions	Hereditary retinoblastoma
	Neurofibromin	Heterogeneous	Neurofibromatosis type I

thus lead to an excess of M1 and M2. Under these conditions tissue injury may result if the precursor, the intermediates, or the products of alternative minor pathways are toxic in high concentrations. For example, in galactosemia, the deficiency of galactose-1-phosphate uridyltransferase (Chapter 10) leads to the accumulation of galactose and consequent tissue damage. Excessive accumulation of complex substrates within the lysosomes as a result of deficiency of degradative enzymes is responsible for a group of diseases generally referred to as *lysosomal storage diseases.*

- *An enzyme defect can lead to a metabolic block and a decreased amount of end product* that may be necessary for normal function. For example, a deficiency of melanin may result from lack of tyrosinase, which is necessary for the biosynthesis of melanin from its precursor, tyrosine, resulting in the clinical condition called *albinism.* If the end product is a feedback inhibitor of the enzymes involved in the early reactions (in Fig. 5.5 it is shown that the product inhibits enzyme 1), the deficiency of the end product may permit overproduction of intermediates and their catabolic products, some of which may be injurious at high concentrations. A prime example of a disease with such an underlying mechanism is Lesch-Nyhan syndrome (Chapter 26).
- *Failure to inactivate a tissue-damaging substrate* is best exemplified by α_1-antitrypsin deficiency. Individuals who have an inherited deficiency of serum α_1-antitrypsin are unable to inactivate neutrophil elastase in their lungs. Unchecked activity of this protease leads to destruction of elastin in the walls of lung alveoli, leading eventually to pulmonary emphysema (Chapter 15).

Defects in Receptors and Transport Systems

As we discussed in Chapter 1, biologically active substances have to be actively transported across the cell membrane. In some cases, transport is achieved by receptor-mediated endocytosis. A genetic defect in a receptor-mediated transport system is exemplified by familial hypercholesterolemia (FH), in which reduced synthesis or function of LDL receptors leads to defective transport of LDL into the cells and secondarily to excessive cholesterol synthesis by complex intermediary mechanisms. In cystic fibrosis the transport system for chloride and bicarbonate ions in exocrine glands, sweat ducts, lungs, and pancreas is defective. By complex mechanisms not fully understood, impaired anion transport leads to serious injury to the lungs and pancreas (Chapter 10).

Alterations in Structure, Function, or Quantity of Nonenzyme Proteins

Genetic defects resulting in alterations of nonenzyme proteins often have widespread secondary effects, as exemplified by sickle cell disease. The hemoglobinopathies, sickle cell disease being one, all of which are characterized by defects in the structure of the globin molecule, best exemplify this category. In contrast to the hemoglobinopathies, the thalassemias result from mutations in globin genes that affect the amount of globin chains synthesized. Thalassemias are associated with reduced amounts of structurally normal α-globin or β-globin chains (Chapter 14). Other examples of genetic disorders involving defective structural proteins include osteogenesis imperfecta (defect in collagen, Chapter 26), hereditary spherocytosis (spectrin, Chapter 14), and muscular dystrophies (dystrophin, Chapter 27).

Genetically Determined Adverse Reactions to Drugs

Certain genetically determined enzyme deficiencies are unmasked only after exposure of the affected individual to certain drugs. This special area of genetics, called *pharmacogenetics,* is of considerable clinical importance. The classic example of drug-induced injury in the genetically susceptible individual is associated with a deficiency of the enzyme G6PD. Under normal conditions, G6PD deficiency does not result in disease, but on administration, for example, of the antimalarial drug primaquine, a severe hemolytic anemia results (Chapter 14). In recent years an increasing number of polymorphisms of genes encoding drug-metabolizing enzymes, transporters, and receptors have been identified. In some cases these genetic factors have a major impact on drug sensitivity and adverse reactions. It is hoped that advances in pharmacogenetics will lead to patient-tailored therapy, an example of personalized medicine.

With this overview of the biochemical basis of single-gene disorders, we now consider selected examples grouped according to the underlying defect.

Disorders Associated With Defects in Structural Proteins

Several diseases caused by mutations in genes that encode structural proteins are listed in Table 5.4. Many are discussed elsewhere in the text. Only Marfan syndrome and Ehlers-Danlos syndromes (EDSs) are discussed here because they affect connective tissue and hence involve multiple organ systems.

Marfan Syndrome

Marfan syndrome is a disorder of connective tissues, manifested principally by changes in the skeleton, eyes, and cardiovascular system. Its prevalence is estimated to be 1 in 5000. Approximately 70% to 85% of cases are familial and transmitted by autosomal dominant inheritance. The remainder are sporadic and arise from new mutations.

Pathogenesis

Marfan syndrome results from an inherited defect in an extracellular glycoprotein called *fibrillin-1*. There are two fundamental mechanisms by which loss of fibrillin leads to the clinical manifestations of Marfan syndrome: loss of structural support in microfibril-rich connective tissue and excessive activation of transforming growth factor (TGF)-β signaling. Each of these is discussed below.

- *Fibrillin is the major component of microfibrils found in the extracellular matrix* (Chapter 1). These fibrils provide a scaffold on which tropoelastin is deposited to form elastic fibers. Although microfibrils are widely distributed in the body, they are particularly abundant in the aorta, ligaments, and the ciliary zonules that support the lens; these tissues are prominently affected in Marfan syndrome. Fibrillin occurs in two homologous forms, fibrillin-1 and fibrillin-2, encoded by two separate genes, *FBN1* and *FBN2,* mapped on chromosomes 15q21.1 and 5q23.31, respectively. Mutations of *FBN1* underlie Marfan syndrome; mutations of the related *FBN2* gene are less

common, and they give rise to *congenital contractural arachnodactyly,* an autosomal dominant disorder characterized by skeletal abnormalities. Mutational analysis has revealed close to 1000 distinct mutations of the *FBN1* gene in individuals with Marfan syndrome. Most of these are missense mutations that give rise to abnormal fibrillin-1. These can inhibit polymerization of fibrillin fibers (dominant negative effect). Alternatively, the reduction of fibrillin content below a certain threshold weakens the connective tissue (haploinsufficiency).

- *TGF-β bioavailability*. While many clinical manifestations of Marfan syndrome can be explained by changes in the mechanical properties of the extracellular matrix resulting from abnormalities of fibrillin, several others such as bone overgrowth and myxoid changes in mitral valves cannot be attributed to changes in tissue elasticity. It is now established that fibrillin-1 controls the bioavailability of TGF-β. Reduced or altered forms of fibrillin-1 give rise to abnormal and excessive activation of TGF-β, since normal microfibrils sequester TGF-β. Excessive TGF-β signaling, in turn, leads to inflammation, has deleterious effects on vascular smooth muscle development, and increases the activity of metalloproteases, causing loss of extracellular matrix. This schema is supported by two sets of observations.
  - First, gain-of-function mutations in the TGF-β type II receptor give rise to a related syndrome, called Marfan syndrome type 2 (MFS2). Furthermore, patients with germline mutations in one isoform of TGF-β, called TGF-β3, present with an inherited predisposition to aortic aneurysm and other cardiovascular manifestations similar to those found in patients with classic Marfan syndrome.
  - Second, angiotensin receptor II blockers, which inhibit TGF-β activity, markedly reduce the aortic root diameter in mouse models of Marfan syndrome. Clinical trials are under way to evaluate the efficacy of angiotensin receptor blockade in patients with Marfan syndrome.

MORPHOLOGY

Skeletal abnormalities are the most striking feature of Marfan syndrome. Typically the patient is unusually tall with exceptionally long extremities and long, tapering fingers and toes. The joint ligaments in the hands and feet are lax, suggesting that the patient is double-jointed; typically the thumb can be hyperextended back to the wrist. The head is commonly dolichocephalic (long-headed) with bossing of the frontal eminences and prominent supraorbital ridges. A variety of spinal deformities may appear, including kyphosis, scoliosis, or rotation or slipping of the dorsal or lumbar vertebrae. The chest is classically deformed, presenting either pectus excavatum (deeply depressed sternum) or a pigeon-breast deformity.

Ocular changes take many forms. Most characteristic is bilateral subluxation or dislocation (usually outward and upward) of the lens, referred to as ectopia lentis. This abnormality, resulting from weakening of ciliary zonules, is so uncommon in persons who do not have this disease that the finding of bilateral ectopia lentis should raise the suspicion of Marfan syndrome.

Cardiovascular lesions are the most life-threatening features of this disorder. The two most common lesions are mitral valve prolapse, which occurs in 40% to 50% of cases and, of greater importance, dilation of the ascending aorta due to cystic medionecrosis. Histologically the changes in the media are virtually identical to those found in cystic medionecrosis not related to Marfan syndrome (Chapter 12). Loss of medial support results in progressive dilation of the aortic valve ring and the root of the aorta, giving rise to severe aortic incompetence. Weakening of the media predisposes to an intimal tear, which may initiate an intramural hematoma that cleaves the layers of the media to produce **aortic dissection.** After cleaving the aortic layers for considerable distances, sometimes back to the root of the aorta or down to the iliac arteries, the hemorrhage often ruptures through the aortic wall.

Clinical Features

Although mitral valve lesions are more frequent, they are clinically less important than aortic lesions. Loss of connective tissue support in the mitral valve leaflets makes them soft and billowy, creating a so-called floppy valve (Chapter 12). Valvular lesions, along with lengthening of the chordae tendineae, frequently give rise to mitral regurgitation. Similar changes may affect the tricuspid and, rarely, the aortic valves. Echocardiography greatly enhances the ability to detect the cardiovascular abnormalities and is therefore extremely valuable in the diagnosis of Marfan syndrome. The great majority of deaths are caused by rupture of aortic dissections, followed in importance by cardiac failure.

While the lesions just described typify Marfan syndrome, it must be emphasized that there is great variation in the clinical expression of this genetic disorder. Patients with prominent eye or cardiovascular changes may have few skeletal abnormalities, whereas others with striking changes in body habitus have no eye changes. Although variability in clinical expression may be seen within a family, interfamilial variability is much more common and extensive. Because of such variations, the clinical diagnosis of Marfan syndrome is currently based on the so-called revised Ghent criteria. These take into account family history, cardinal clinical signs in the absence of family history, and presence or absence of fibrillin mutation. In general, major involvement of two of the four organ systems (skeletal, cardiovascular, ocular, and skin) and minor involvement of another organ is required for diagnosis.

The variable expression of the Marfan defect is best explained on the basis of the many different mutations that affect the fibrillin locus, which number around 1000. This genetic heterogeneity also poses formidable challenges in the diagnosis of Marfan syndrome. The evolving high throughput sequencing technologies discussed later in this chapter may overcome this problem in the future.

Ehlers-Danlos Syndromes (EDSs)

EDSs comprise a clinically and genetically heterogeneous group of disorders that result from some mutations in genes that encode collagen, enzymes that modify collagen, and less commonly other proteins present in the extracellular matrix. Other disorders resulting from mutations affecting collagen synthesis include osteogenesis imperfecta (Chapter 26), Alport syndrome (Chapter 20), and epidermolysis bullosa (Chapter 25).

Biosynthesis of collagen is a complex process (Chapter 1) that can be disturbed by genetic errors that may affect any one of the numerous structural collagen genes or enzymes necessary for posttranscriptional modifications of collagen. Hence the mode of inheritance of EDSs encompasses all three Mendelian patterns. On the basis of clinical and molecular characteristics, six variants of EDS have been recognized. These are listed in Table 5.5. More recently, next-generation sequencing has revealed other subgroups bringing the total to 11 molecular types. Although individually rare, the collective frequency of EDSs is 1 in 5000 births. It is beyond the scope of this book to discuss each variant individually; instead, the important clinical features common to most variants are summarized, and clinical manifestations are correlated with the underlying molecular defects in collagen synthesis or structure.

As might be expected, tissues rich in collagen, such as skin, ligaments, and joints, are frequently involved in most variants of EDS. Because the abnormal collagen fibers lack adequate tensile strength, *skin is hyperextensible, and the joints are hypermobile.* These features permit grotesque contortions, such as bending the thumb backward to touch the forearm and bending the knee forward to create almost a right angle. It is believed that most contortionists have one of the EDSs. A predisposition to joint dislocation, however, is one of the prices paid for this virtuosity. *The skin is extraordinarily stretchable, extremely fragile, and easily bruised.* Minor injuries produce gaping defects, and surgical repair or intervention is accomplished with great difficulty because of the lack of normal tensile strength. *The basic defect in connective tissue may lead to serious internal complications.* These include rupture of the colon and large arteries (vascular EDS), ocular fragility with rupture of cornea and retinal detachment (kyphoscoliosis EDS), and diaphragmatic hernia (classic EDS).

The biochemical and molecular bases of these abnormalities are known in all but one form of EDS—the so-called hypermobility type. Some of the EDS types are described briefly because they offer some insights into the perplexing clinical heterogeneity of these disorders. Perhaps the best characterized is the *kyphoscoliosis type*, the most common autosomal recessive form of EDS. It results from mutations in the *PLOD1* gene encoding lysyl hydroxylase, an enzyme necessary for hydroxylation of lysine residues during collagen synthesis. Affected patients have markedly reduced levels of this enzyme. Because hydroxylysine is essential for intermolecular and intramolecular cross-linking of collagen fibers, a deficiency of lysyl hydroxylase results in the synthesis of collagen that lacks normal structural stability.

The *vascular type of EDS* results from abnormalities of type III collagen. This form is genetically heterogeneous because at least three distinct types of mutations affecting the *COL3A1* gene encoding collagen type III can give rise to this variant. Some affect the rate of synthesis of pro-α1 (III) chains, others affect the secretion of type III procollagen, and still others lead to the synthesis of structurally abnormal type III collagen. Some mutant alleles behave as dominant negatives (see discussion under "Autosomal Dominant Disorders") and thus produce severe phenotypic effects. These molecular studies provide a rational basis for the pattern of transmission and clinical features that are characteristic of this variant. First, because vascular-type EDS results from mutations involving a structural protein (rather than an enzyme protein), an autosomal dominant pattern of inheritance would be expected. Second, because blood vessels and intestines are known to be rich in collagen type III, an abnormality of this collagen is consistent with severe structural defects (e.g., vulnerability to spontaneous rupture) in these organs. Unlike many other EDS subtypes, the skin is not usually hyperextensible.

In two forms of EDS—arthrochalasia type and dermatosparaxis type—the fundamental defect is in the conversion of type I procollagen to collagen. This step in collagen synthesis involves cleavage of noncollagen peptides at the N terminus and C terminus of the procollagen molecule. This is accomplished by N-terminal–specific and C-terminal–specific peptidases. The defect in the conversion of procollagen to collagen in the arthrochalasia type has been traced to mutations that affect one of the two type I collagen genes, *COL1A1* and *COL1A2*. As a result, structurally abnormal pro-α1 (I) or pro-α2 (I) chains that resist cleavage of N-terminal peptides are formed. In patients with a single mutant allele, only 50% of the type I collagen chains are abnormal, but because these chains interfere with the formation of normal collagen helices, heterozygotes manifest the disease. In contrast, the related dermatosparaxis type is caused by mutations in the *ADAMTS2* gene that encodes procollagen-N-peptidase, essential for the cleavage of procollagens. Because in this case the disease is caused by an enzyme deficiency, it follows an autosomal recessive form of inheritance.

Finally, in the *classic type of EDS,* molecular analysis suggests that genes other than those that encode collagen may also be involved. In close to 90% of cases, mutations in the genes for type V collagen (*COL5A1* and *COL5A2*) have been detected. Surprisingly, in the remaining cases, no other collagen gene abnormalities have been found despite clinical similarity with the classic type EDS. It is suspected

Table 5.5 Classification of Ehlers-Danlos Syndromes

EDS Type[a]	Clinical Findings	Inheritance	Gene Defects
Classic (I/II)	Skin and joint hypermobility, atrophic scars, easy bruising	Autosomal dominant	*COL5A1*, *COL5A2*
Hypermobility (III)	Joint hypermobility, pain, dislocations	Autosomal dominant	Unknown
Vascular (IV)	Thin skin, arterial or uterine rupture, bruising, small joint hyperextensibility	Autosomal dominant	*COL3A1*
Kyphoscoliosis (VI)	Hypotonia, joint laxity, congenital scoliosis, ocular fragility	Autosomal recessive	Lysyl hydroxylase
Arthrochalasia (VIIa, VIIb)	Severe joint hypermobility, skin changes (mild), scoliosis, bruising	Autosomal dominant	*COL1A1*, *COL1A2*
Dermatosparaxis (VIIc)	Severe skin fragility, cutis laxa, bruising	Autosomal recessive	Procollagen *N*-peptidase

[a]EDS types were previously classified by Roman numerals. Parentheses show previous numerical equivalents.

that in some cases, genetic defects that affect the biosynthesis of other extracellular matrix molecules that influence collagen synthesis indirectly may be involved. One example is a classic EDS-like condition caused by mutation in the *TNXB* gene encoding tenascin-X, a large multimeric protein that interacts with fibrillar types I, III, and V collagens.

To summarize, the common thread in EDSs is some abnormality of collagen. These disorders, however, are extremely heterogeneous. At the molecular level, a variety of defects, varying from mutations involving structural genes for collagen to those involving enzymes that are responsible for posttranscriptional modifications of mRNA, have been detected. Such molecular heterogeneity results in the expression of EDSs as clinically variable disorders with several patterns of inheritance.

KEY CONCEPTS

MARFAN SYNDROME AND EHLERS-DANLOS SYNDROMES

Marfan Syndrome

- Marfan syndrome is caused by a mutation in the *FBN1* gene encoding fibrillin, which is required for structural integrity of connective tissues and regulation of TGF-β signaling.
- The major tissues affected are the skeleton, eyes, and cardiovascular system.
- Clinical features may include tall stature, long fingers, bilateral subluxation of lens, mitral valve prolapse, aortic aneurysm, and aortic dissection.

Ehlers-Danlos Syndromes

- There are several variants of EDS, all characterized by defects in collagen synthesis or assembly. Each of the variants is caused by a distinct mutation involving one of several collagen genes or genes that encode other ECM proteins like tenascin-X.
- Clinical features may include fragile, hyperextensible skin vulnerable to trauma; hypermobile joints; and ruptures involving the colon, cornea, or large arteries. Wound healing is poor.

Disorders Associated With Defects in Receptor Proteins

Familial Hypercholesterolemia

Familial hypercholesterolemia (FH) is caused most commonly by mutations in the gene encoding the receptor for LDL, resulting in inadequate removal of plasma LDL by the liver. Mutations in the LDL receptor gene *(LDLR)* account for 80% to 85% of cases of FH. Much less commonly, FH is caused by mutations in two other genes involved in clearance of plasma LDL. They encode (1) apolipoprotein B-100 (ApoB), the ligand for LDL receptor on the LDL particle (5% to 10% cases), and (2) proprotein convertase subtilisin/kexin type 9 (1% to 2% cases). This enzyme, better known by its abbreviation *PCSK9*, reduces expression of LDL receptors by downregulating their recycling and consequent degradation in lysosomes. Each of these three types of mutations impairs hepatic clearance of LDL and increases serum levels of cholesterol, giving rise to premature atherosclerosis and a greatly increased risk of myocardial infarction. The functions of these genes in cholesterol metabolism are discussed below.

FH caused by LDL receptor mutations is one of the most frequently occurring Mendelian disorders. Heterozygotes with one mutant gene, representing about 1 in 200 individuals, have from birth a twofold to threefold elevation of plasma cholesterol level, leading to tendinous xanthomas and premature atherosclerosis in adult life (Chapter 11). Homozygotes, having a double dose of the mutant gene, are much more severely affected and may have fivefold to sixfold elevations in plasma cholesterol levels. Skin xanthomas and coronary, cerebral, and peripheral vascular atherosclerosis may develop in these individuals at an early age. Myocardial infarction may occur before age 20 years. Large-scale studies have found that FH is present in 3% to 6% of survivors of myocardial infarction.

Normal Cholesterol Metabolism and Transport

Cholesterol may be derived from the diet or from endogenous synthesis. Dietary triglycerides and cholesterol are incorporated into chylomicrons in the intestinal mucosa and travel by way of the gut lymphatics to the blood. These chylomicrons are hydrolyzed by an endothelial lipoprotein lipase in the capillaries of muscle and fat. The chylomicron remnants, rich in cholesterol, are then delivered to the liver. Some of the cholesterol enters the metabolic pool (to be described), and some is excreted as free cholesterol or as bile acids into the biliary tract. The endogenous synthesis of cholesterol and LDL begins in the liver (Fig. 5.6). The first step in this process is the secretion of very-low-density

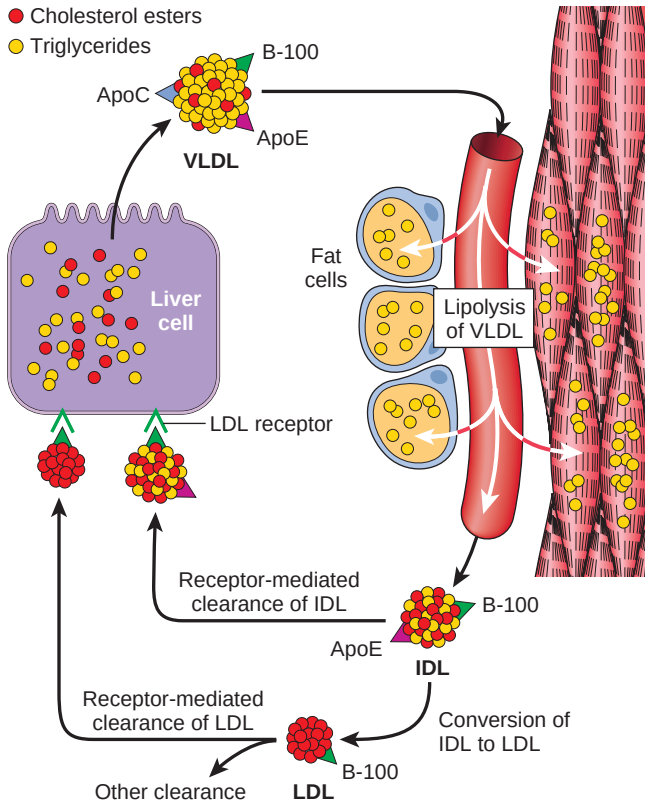


Figure 5.6 Low-density lipoprotein *(LDL)* metabolism and the role of the liver in its synthesis and clearance. Lipolysis of very-low-density lipoprotein *(VLDL)* by lipoprotein lipase in the capillaries releases triglycerides, which are then stored in fat cells and used as a source of energy in skeletal muscles. *ApoC*, Apolipoprotein C; *ApoE*, apolipoprotein E; *B-100*, apolipoprotein B-100 (ApoB); *IDL*, intermediate-density lipoprotein.

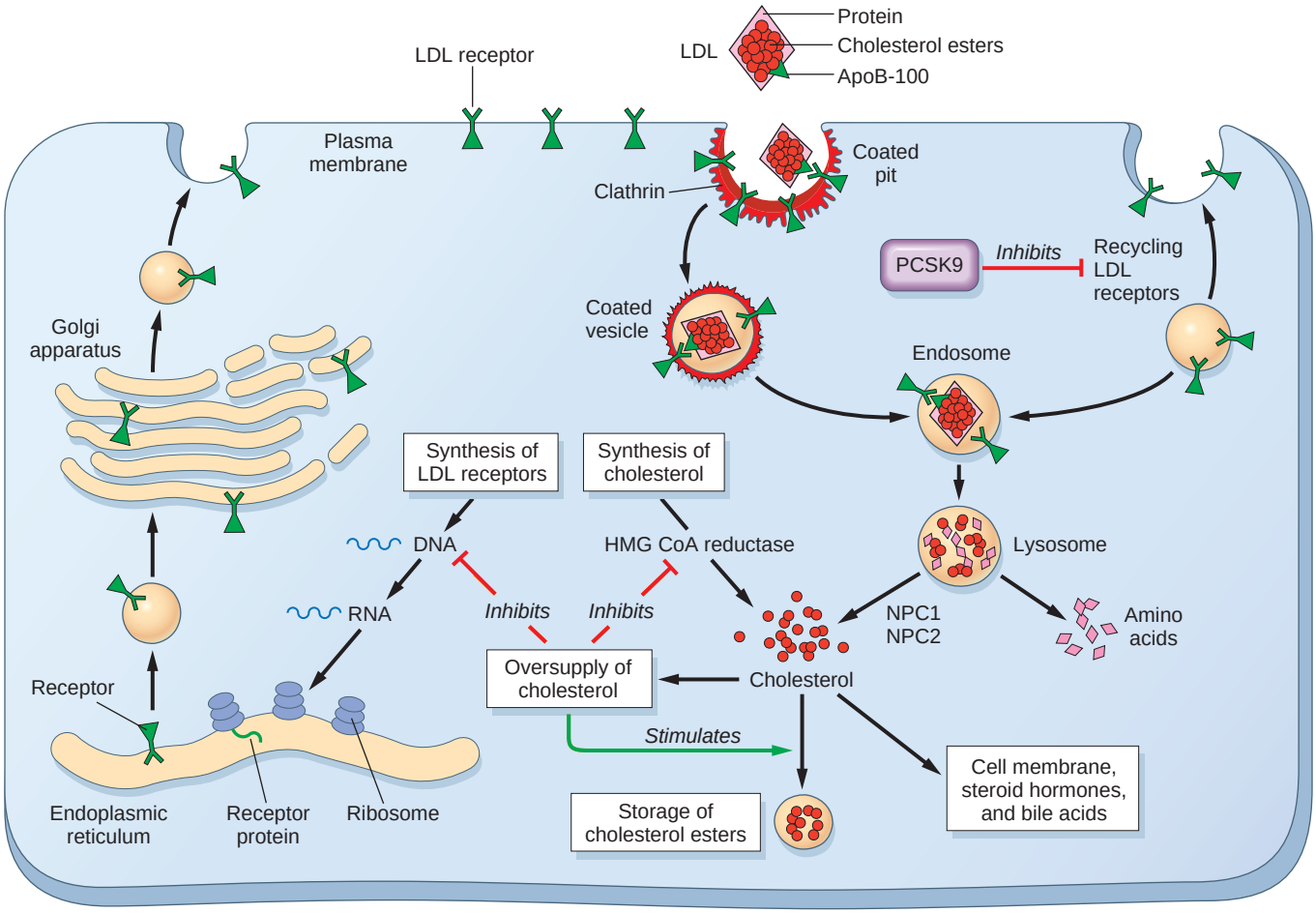


Figure 5.7 The low-density lipoprotein *(LDL)* receptor pathway and regulation of cholesterol metabolism. *ApoB-100*, Apolipoprotein B-100 (ApoB); *HMG CoA*, 3-hydroxy-3-methylglutaryl coenzyme A.

lipoprotein (VLDL) from the liver into the blood. VLDL particles are rich in triglycerides, but they contain lesser amounts of cholesteryl esters. In addition, they carry apolipoproteins ApoB, ApoC, and ApoE on their surface. In the capillaries of adipose tissue and muscle, the VLDL particle undergoes lipolysis and is converted to VLDL remnant, also called intermediate-density lipoprotein (IDL). Compared with VLDL, IDL particles have reduced content of triglycerides and an increase in cholesteryl esters. ApoC is lost, but ApoB and ApoE are retained. After release from the capillary endothelium, the IDL particles have one of two fates. Approximately 50% of newly formed IDL is rapidly taken up by the liver by receptor-mediated transport. The receptor responsible for the binding of IDL to the liver cell membrane recognizes both ApoB and ApoE. It is called ApoB/E, or more commonly the LDL receptor, because it is also involved in the hepatic clearance of LDL (described later). In the liver cells, IDL is recycled to generate VLDL. The IDL particles not taken up by the liver are subjected to further metabolic processing that removes most of the remaining triglycerides and ApoE, yielding ApoB carrying cholesterol-rich LDL particles.

Although many cell types, including fibroblasts, lymphocytes, smooth muscle cells, hepatocytes, and adrenocortical cells, possess high-affinity LDL receptors, approximately 70% of the plasma LDL is cleared by the liver, using a quite sophisticated transport process (Fig. 5.7). The first step involves binding of LDL to cell-surface receptors, which are clustered in specialized regions of the plasma membrane called *coated pits* (Chapter 1). After binding, the coated pits containing the receptor-bound LDL are internalized by invagination to form coated vesicles, after which they migrate within the cell to fuse with the lysosomes. Here the LDL dissociates from the receptor, which is recycled to the surface. The recycling of LDL receptors is regulated by PCSK9, which binds to LDL receptors on the surface of hepatocytes and causes their degradation after endocytosis. In the lysosomes the LDL molecule is enzymatically degraded; the apoprotein part is hydrolyzed to amino acids, whereas the cholesteryl esters are broken down to free cholesterol. This free cholesterol, in turn, crosses the lysosomal membrane to enter the cytoplasm, where it is used for membrane synthesis and as a regulator of cholesterol homeostasis. The exit of cholesterol from the lysosomes requires the action of two proteins, called NPC1 and NPC2 (see "Niemann-Pick Disease Type C"). Four separate processes are affected by the released intracellular cholesterol, as follows (see Fig. 5.7):

- Cholesterol *suppresses* cholesterol synthesis within the cell by inhibiting the activity of the enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase, which is the rate-limiting enzyme in the synthetic pathway.

- Cholesterol *activates* the enzyme acyl-coenzyme A:cholesterol acyltransferase, favoring esterification and storage of excess cholesterol.
- Cholesterol *suppresses* the synthesis of LDL receptors, thus protecting the cells from excessive accumulation of cholesterol.
- Cholesterol *upregulates* the expression of PCSK9, which reduces recycling of LDL receptors and causes degradation of endocytosed LDL receptors. This provides an additional mechanism of protecting the cells from excessive accumulation of cholesterol.

As mentioned earlier, FH results from mutations in the gene encoding the receptor for LDL or the two genes that compromise its function. The impact of mutations in these genes is as follows:

- *Mutations in* LDLR *gene*. Heterozygotes with FH due to mutation in the *LDLR* gene possess only 50% of the normal number of high-affinity LDL receptors because they have only one normal gene. As a result of this defect in transport, the catabolism of LDL by the receptor-dependent pathways is impaired, and the plasma level of LDL increases two- to three-fold. Homozygotes have virtually no normal LDL receptors in their cells and have much higher levels of circulating LDL. In addition to defective LDL clearance, both the homozygotes and the heterozygotes have increased synthesis of LDL. The increased synthesis that contributes to hypercholesterolemia also results from a lack of LDL receptors (see Fig. 5.6). As mentioned above, IDL, the immediate precursor of plasma LDL, also uses hepatic LDL receptors (apoB/E receptors) for its transport into the liver. In FH, impaired IDL transport into the liver secondarily diverts a greater proportion of plasma IDL into the precursor pool for plasma LDL.
- *Mutations in gene encoding ApoB*. Since ApoB on the surface of LDL particles is the ligand for LDL receptors, mutant ApoB reduces the binding of LDL molecules with LDL receptors. This compromise in the binding of LDL particles to its receptors increases serum LDL cholesterol.
- *Activating mutation in the PCSK9 gene*. This mutation greatly reduces the number of LDL receptors on the cell surface because of their increased degradation during the recycling process.

The transport of LDL via the scavenger receptor seems to occur at least in part into the cells of the mononuclear phagocyte system. Monocytes and macrophages have receptors for chemically altered (e.g., acetylated or oxidized) LDL. Normally the amount of LDL transported along this scavenger receptor pathway is less than that mediated by the LDL receptor–dependent mechanisms. In the face of hypercholesterolemia, however, there is a marked increase in the scavenger receptor–mediated traffic of LDL cholesterol into the cells of the mononuclear phagocyte system and possibly the vascular walls (Chapter 11). This increase is responsible for the appearance of xanthomas and contributes to the pathogenesis of premature atherosclerosis.

The molecular genetics of FH is extremely complex. More than 2000 mutations involving the LDL receptor gene, including DNA copy number variations, insertions, deletions, and missense and nonsense mutations, have been identified.

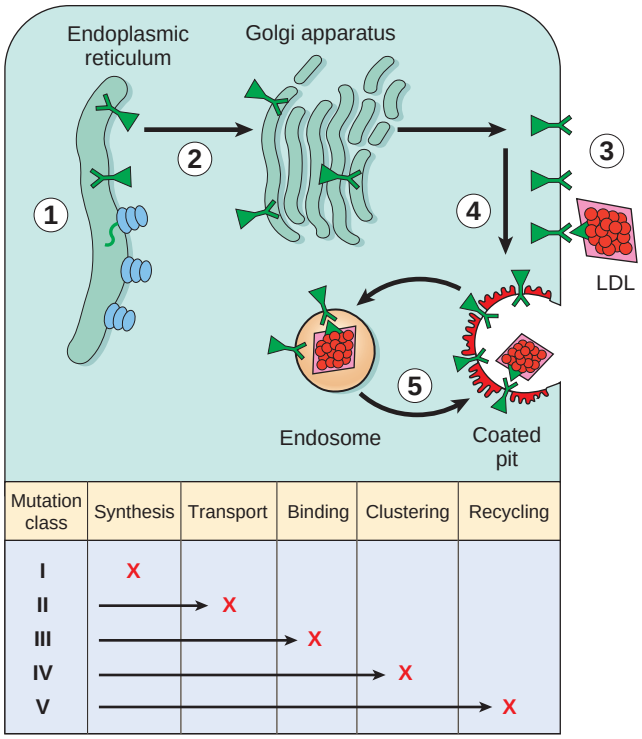


Figure 5.8 Classification of low-density lipoprotein *(LDL)* receptor mutations based on abnormal function of the mutant protein. These mutations disrupt the receptor's synthesis in the endoplasmic reticulum, transport to the Golgi complex, binding of apoprotein ligands, clustering in coated pits, and recycling in endosomes. Not shown is class VI mutation, in which initial targeting of the receptor to basolateral membrane fails to occur. Each class is heterogeneous at the DNA level. (Modified with permission from Hobbs HH, et al: The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein, *Annu Rev Genet* 24:133–170, 1990. © 1990 by Annual Reviews.)

These can be classified into six groups (Fig. 5.8). *Class I mutations* are relatively uncommon and lead to a complete failure of synthesis of the LDL receptor protein (null allele). *Class II mutations* are fairly common; they encode LDL receptor proteins that accumulate in the endoplasmic reticulum because their folding defects make it impossible for them to be transported to the Golgi complex. *Class III mutations* affect the ApoB binding site of the receptor; the mutant LDL receptors reach the cell surface but fail to bind LDL or do so poorly. *Class IV mutations* encode LDL receptors that are synthesized and transported to the cell surface efficiently. They bind LDL normally, but they fail to localize in coated pits, and hence the bound LDL is not internalized. *Class V mutations* encode LDL receptors that are expressed on the cell surface, can bind LDL, and can be internalized; however, the pH-dependent dissociation of the receptor and the bound LDL fails to occur. Such receptors are trapped in the endosome, where they are degraded, and hence they fail to recycle to the cell surface. *Class VI mutations* result in the failure of initial targeting of the LDL receptor to the basolateral membrane.

The discovery of the critical role of LDL receptors in cholesterol homeostasis has led to the rational design of drugs that lower plasma cholesterol by increasing the number of LDL receptors (see Fig. 5.8). One strategy is

based on the ability of certain drugs (statins) to suppress intracellular cholesterol synthesis by inhibiting the enzyme HMG CoA reductase. The reduction in intracellular cholesterol allows greater synthesis of LDL receptors by removing the braking action of cholesterol on LDL receptor synthesis. In another strategy, antibodies that inhibit PCSK9 function reduce the degradation of LDL receptors, thereby increasing their abundance on the cell membrane and consequent increased clearance of LDL cholesterol from the blood. These agents have profound cholesterol-lowering effects, and large clinical trials have shown the benefits of using this class of drugs in the treatment of patients who do not respond adequately to statins alone. Interestingly, PCSK9 was discovered accidentally in individuals with exceedingly low serum cholesterol who were found to have loss-of-function variants of the *PCSK9* gene.

KEY CONCEPTS

FAMILIAL HYPERCHOLESTEROLEMIA

- FH is an autosomal dominant disorder caused by mutations in the genes encoding the (1) LDL receptor (85% cases), (2) ApoB protein (5% to 10% cases), or (3) activating mutations of PCSK9 (1% to 2% cases).
- Patients develop hypercholesterolemia as a consequence of impaired transport of LDL into the cells.
- In heterozygotes for mutations in the *LDLR* gene, elevated serum cholesterol greatly increases the risk of atherosclerosis and resultant coronary artery disease; homozygotes have an even greater increase in serum cholesterol and a higher frequency of ischemic heart disease. Cholesterol also deposits along tendon sheaths to produce xanthomas.

Disorders Associated With Enzyme Defects

Lysosomal Storage Diseases

Lysosomes are key components of the intracellular digestive system. They contain a battery of hydrolytic enzymes, which have two special properties. First, they function in the acidic milieu of the lysosomes. Second, these enzymes constitute a special category of secretory proteins that are destined not for the extracellular fluids but for intracellular organelles. This latter characteristic requires special processing within the Golgi apparatus, which merits brief discussion.

Similar to all other secretory proteins, lysosomal enzymes (or *acid hydrolases,* as they are sometimes called) are synthesized in the endoplasmic reticulum and transported to the Golgi apparatus. Within the Golgi complex they undergo a variety of posttranslational modifications including the attachment of terminal mannose-6-phosphate groups to some of the oligosaccharide side chains. The phosphorylated mannose residues serve as an "address label" that is recognized by specific receptors found on the inner surface of the Golgi membrane. Lysosomal enzymes bind these receptors and are thereby segregated from the numerous other secretory proteins within the Golgi. Subsequently, small transport vesicles containing the receptor-bound enzymes are pinched off from the Golgi and proceed to fuse with the lysosomes. Thus the enzymes are targeted to their intracellular abode, and the vesicles are shuttled back to the Golgi (Fig. 5.9). As indicated later, genetically determined errors in this remarkable sorting mechanism may give rise to one form of lysosomal storage disease. Recent studies have established a close link between lysosomal storage diseases and several neurodegenerative disorders. The cellular and molecular mechanisms of this linkage will be discussed below.

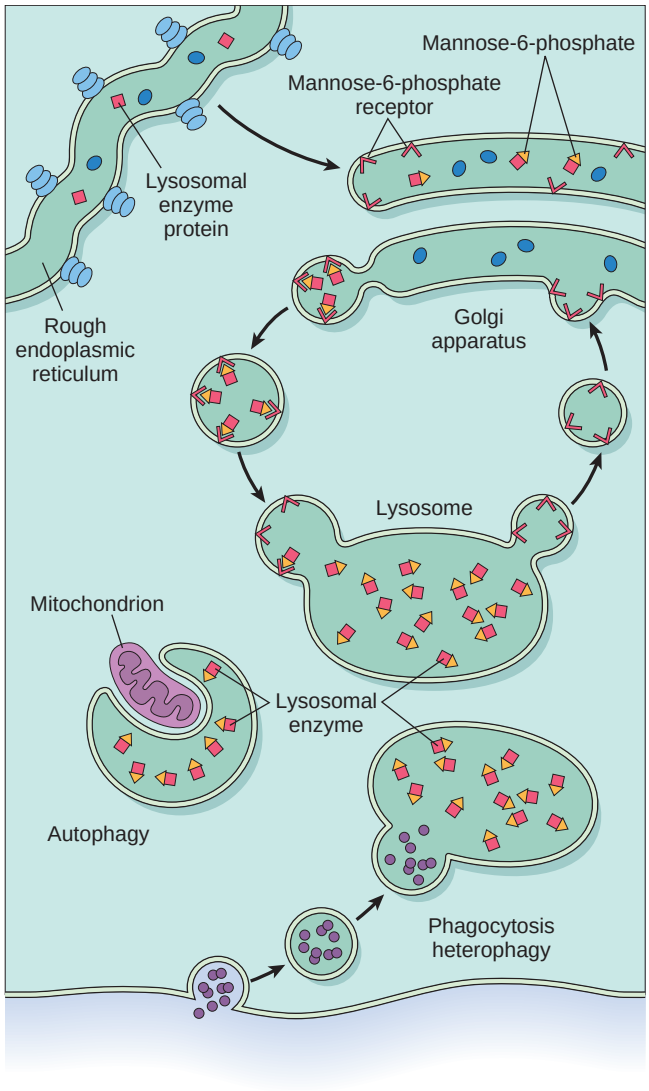


Figure 5.9 Synthesis and intracellular transport of lysosomal enzymes.

The lysosomal enzymes catalyze the breakdown of a variety of complex macromolecules. These large molecules may be derived from the metabolic turnover of intracellular organelles (autophagy), or they may be acquired from outside the cells by phagocytosis (heterophagy). An inherited deficiency of a functional lysosomal enzyme gives rise to two pathologic consequences (Fig. 5.10).

- *Catabolism of the substrate of the missing enzyme remains incomplete,* leading to the accumulation of the partially degraded insoluble metabolite within the lysosomes. This is called "primary accumulation." Stuffed with incompletely digested macromolecules, lysosomes become large and numerous enough to interfere with normal cell functions.

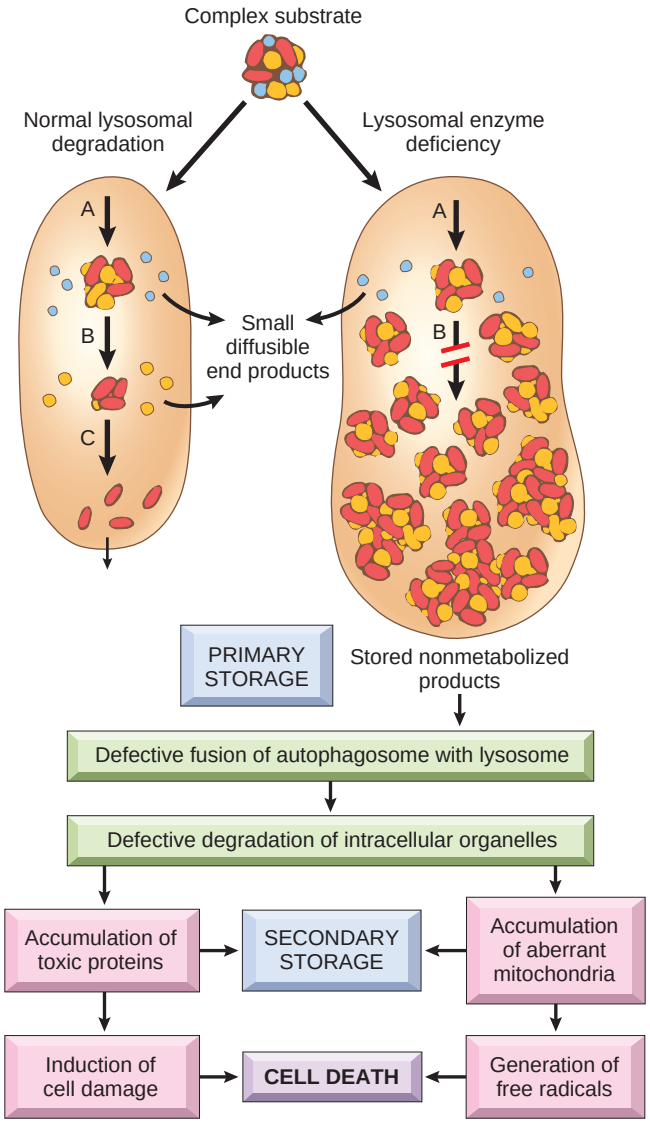


Figure 5.10 Pathogenesis of lysosomal storage diseases. In the example shown, a complex substrate is normally degraded by a series of lysosomal enzymes (*A*, *B*, and *C*) into soluble end products. If there is a deficiency or malfunction of one of the enzymes (e.g., *B*), catabolism is incomplete, and insoluble intermediates accumulate in the lysosomes. In addition to this primary storage, secondary storage and toxic effects result from defective autophagy.

- *There is a tight linkage between autophagy, mitochondrial functions, and lysosomes*. As discussed in Chapter 2, a diverse range of cellular organelles and molecules are degraded by autophagy including complex lipids, polyubiquitinated proteins, mitochondria, and fragments of the endoplasmic reticulum. In particular, autophagy is essential for turnover of mitochondria by a process termed *mitophagy*. This serves as a quality control system whereby dysfunctional mitochondria are degraded. Because of the accumulation of undigested macromolecules in the lysosomes, the rate at which lysosomes process organelles delivered by autophagocytic vacuoles is markedly reduced. This leads to persistence of dysfunctional and leaky mitochondria with poor calcium-buffering capacity and altered membrane potentials in the lysosomes. Damaged mitochondria generate free radicals and release molecules that trigger the intrinsic pathway of apoptosis. Impaired autophagy gives rise to secondary accumulation of autophagic substrates including ubiquitinated and aggregate-prone polypeptides such as α-synuclein and Huntingtin protein. This provides a molecular link between neurodegenerative disorders and lysosomal storage diseases such as Gaucher disease (discussed below).

There are three general approaches to the treatment of lysosomal storage diseases. The most obvious is enzyme replacement therapy, currently in use for several of these diseases. Another approach, substrate reduction therapy, is based on the premise that if the substrate to be degraded by the lysosomal enzyme can be reduced, the residual enzyme activity may be sufficient to catabolize it and prevent accumulation. A more recent strategy is based on the understanding of the molecular basis of enzyme deficiency. In many disorders, exemplified by Gaucher disease, the enzyme activity is low because the mutant proteins are unstable and prone to misfolding and hence degraded in the endoplasmic reticulum. In such diseases an exogenous competitive inhibitor of the enzyme can, paradoxically, bind to the mutant enzyme and act as the folding template that assists proper folding of the enzyme and thus prevents its degradation. Such *molecular chaperone therapy* is under active investigation. In addition to the aforementioned, hematopoietic stem cell transplants and gene therapy are also being evaluated in specific cases.

Although the combined frequency of lysosomal storage disorders is about 1 in 5000 live births, lysosomal dysfunction may be involved in the etiology of several more common diseases. For example, an important genetic risk factor for developing Parkinson disease is the carrier state for Gaucher disease, and virtually all patients with Gaucher disease develop Parkinson disease. Niemann-Pick type C disease is another lysosomal storage disorder that increases the risk for Alzheimer disease. Such interconnectedness stems from the multifunctionality of the lysosome. For example, lysosomes play critical roles in (1) autophagy, resulting from fusion with the autophagosome; (2) immunity, because they fuse with phagosomes; and (3) membrane repair, through fusion with the plasma membrane.

Approximately 70 lysosomal storage diseases have been identified. These may result from abnormalities of lysosomal enzymes or proteins involved in substrate degradation, endosomal sorting, or lysosomal membrane integrity. Lysosomal storage disorders are divided into categories based on the biochemical nature of the substrates and the accumulated metabolites (Table 5.6). Within each group are several entities, each resulting from the deficiency of a specific enzyme.

In general, the distribution of the stored material, and hence the organs affected, is determined by two interrelated factors: (1) the tissue where most of the material to be degraded is found and (2) the location where most of the degradation normally occurs. For example, the brain is rich in gangliosides, and hence defective hydrolysis of gangliosides, as occurs in G_{M1} and G_{M2} gangliosidoses, results primarily in accumulation within neurons and consequent neurologic symptoms. Defects in degradation of mucopolysaccharides affect virtually every organ because mucopolysaccharides are

Table 5.6 Lysosomal Storage Diseases

Disease	Enzyme Deficiency	Major Accumulating Metabolites
Glycogenosis	Type 2—Pompe disease α-1,4-Glucosidase (lysosomal glucosidase)	Glycogen
Sphingolipidoses		
G_{M1} gangliosidosis	G_{M1} ganglioside β-galactosidase	G_{M1} ganglioside, galactose-containing oligosaccharides
Type 1—infantile, generalized		
Type 2—juvenile		
G_{M2} gangliosidosis		
Tay-Sachs disease	Hexosaminidase A	G_{M2} ganglioside
Sandhoff disease	Hexosaminidase A and B	G_{M2} ganglioside, globoside
G_{M2} gangliosidosis variant AB	Ganglioside activator protein	G_{M2} ganglioside
Sulfatidoses		
Metachromatic leukodystrophy	Arylsulfatase A	Sulfatide
Multiple sulfatase deficiency	Arylsulfatase A, B, C; steroid sulfatase; iduronate sulfatase; heparan *N*-sulfatase	Sulfatide, steroid sulfate, heparan sulfate, dermatan sulfate
Krabbe disease	Galactosylceramidase	Galactocerebroside
Fabry disease	α-Galactosidase A	Ceramide trihexoside
Gaucher disease	Glucocerebrosidase	Glucocerebroside
Niemann-Pick disease: types A and B	Sphingomyelinase	Sphingomyelin
Mucopolysaccharidoses (MPSs)		
MPS I-H (Hurler)	α-L-Iduronidase	Dermatan sulfate, heparan sulfate
MPS II (Hunter)	Iduronate 2-sulphatase	
Mucolipidoses (MLs)		
I-cell disease (ML II) and pseudo-Hurler polydystrophy	Deficiency of phosphorylating enzymes essential for the formation of mannose-6-phosphate recognition marker; acid hydrolases lacking the recognition marker cannot be targeted to the lysosomes, but are secreted extracellularly	Mucopolysaccharide, glycolipid
Other diseases of complex carbohydrates		
Fucosidosis	α-Fucosidase	Fucose-containing sphingolipids and glycoprotein fragments
Mannosidosis	α-Mannosidase	Mannose-containing oligosaccharides
Aspartylglycosaminuria	Aspartylglycosamine amide hydrolase	Aspartyl-2-deoxy-2-acetamido-glycosylamine
Other lysosomal storage diseases		
Wolman disease	Acid lipase	Cholesterol esters, triglycerides

widely distributed in the body. Because cells of the mononuclear phagocyte system are especially rich in lysosomes and are involved in the degradation of a variety of substrates, organs rich in phagocytic cells, such as the spleen and liver, are frequently enlarged in several forms of lysosomal storage disorders. The ever-expanding number of lysosomal storage diseases can be divided into rational categories based on the biochemical nature of the accumulated metabolite, thus creating such subgroups as *glycogenoses, sphingolipidoses (lipidoses), mucopolysaccharidoses (MPSs),* and *mucolipidoses* (see Table 5.6).

Most of these conditions are very rare, and their detailed description is better relegated to specialized texts and reviews. Only a few of the more common conditions are considered here.

Tay-Sachs Disease (G_{M2} Gangliosidosis: Hexosaminidase α-Subunit Deficiency)

G_{M2} gangliosidoses are a group of three lysosomal storage diseases caused by deficiency of the enzyme β-hexosaminidase resulting in an inability to catabolize G_{M2} gangliosides. There are two isoenzymes of β-hexosaminidase: Hex A, consisting of two subunits, α and β, and Hex B, a homodimer of β-subunits. Degradation of G_{M2} gangliosides requires three polypeptides encoded by three distinct genes—*HEXA* (on chromosome 15), which encodes the α-subunit of Hex A; *HEXB* (on chromosome 5), which encodes the β-subunit of Hex A and Hex B; and *GM2A* (on chromosome 5), which encodes the activator of hexosaminidase. The phenotypic effects of mutations affecting these genes are fairly similar because they result from accumulation of G_{M2} gangliosides. The underlying enzyme defect, however, is different for each. Tay-Sachs disease, the most common form of G_{M2} gangliosidosis, results from mutations in the α-subunit locus on chromosome 15 that cause a severe deficiency of hexosaminidase A. This disease is especially prevalent among Jews, particularly among those of Eastern European (Ashkenazic) origin, in whom a carrier rate of 1 in 30 has been reported.

The molecular basis for neuronal injury resulting from hexosaminidase deficiency is not fully understood. More than 100 mutations have been described in the *HEXA*

α-subunit gene; most affect protein folding. Because the mutant protein is misfolded, it induces the so-called unfolded protein response (Chapter 2). If such misfolded enzymes are not stabilized by chaperones, they undergo proteasomal degradation, leading to accumulation of toxic substrates and intermediates within neurons. These findings have spurred clinical trials of molecular chaperone therapy for some variants of later-onset Tay-Sachs and other selected lysosomal storage diseases. Such therapy involves the use of synthetic chaperones that can cross the blood–brain barrier, bind to the mutated protein, and enable its proper folding. Sufficient functional enzyme can then be rescued to ameliorate the effects of the inborn error.

MORPHOLOGY

Hexosaminidase A is absent from virtually all the tissues, so G_{M2} ganglioside accumulates in many tissues (e.g., heart, liver, spleen, nervous system), but the **involvement of neurons in the central and autonomic nervous systems and retina dominates the clinical picture.** On histologic examination, the neurons are ballooned with cytoplasmic vacuoles, each representing a markedly distended lysosome filled with gangliosides (Fig. 5.11A). With the electron microscope, several types of **cytoplasmic inclusions** can be visualized, the most prominent being whorled configurations within lysosomes composed of onion-skin layers of membranes (Fig. 5.11B). In time there is progressive destruction of neurons, proliferation of microglia, and accumulation of complex lipids in phagocytes within the brain substance. The ganglion cells in the retina are similarly swollen with G_{M2} ganglioside, particularly at the margins of the macula. A **cherry-red spot thus appears in the macula,** representing accentuation of the normal color of the macular choroid contrasted with the pallor produced by the swollen ganglion cells in the remainder of the retina (Chapter 29). This finding is characteristic of Tay-Sachs disease and other storage disorders affecting the neurons.

Clinical Features

Affected infants appear normal at birth but begin to manifest signs and symptoms at about age 6 months. There is relentless motor and mental deterioration, resulting in motor incoordination and intellectual disability leading eventually to muscular flaccidity, blindness, and increasing dementia. Sometime during the early course of the disease, the characteristic, but not pathognomonic, cherry-red spot appears in the macula of the eye in almost all patients. Over the span of 1 or 2 years a complete vegetative state is reached, followed by death at age 2 to 3 years. Antenatal diagnosis and carrier detection are possible by enzyme assays and DNA-based analysis.

Niemann-Pick Disease Types A and B

Niemann-Pick disease types A and B are two related disorders that are characterized by lysosomal accumulation of sphingomyelin due to an inherited deficiency of sphingomyelinase. Type A is a severe infantile form with extensive neurologic involvement, marked visceral accumulations of sphingomyelin, and progressive wasting and early death within the first 3 years of life. In contrast, patients with type B disease have organomegaly but generally no central nervous system involvement. Patients usually survive into adulthood. As with Tay-Sachs disease, Niemann-Pick disease types A and B are common in Ashkenazi Jews. The gene for acid sphingomyelinase maps to chromosome 11p15.4 and is one of the imprinted genes that is preferentially expressed from the maternal chromosome as a result of epigenetic silencing of the paternal gene (discussed later). Although this disease is typically inherited as an autosomal recessive, heterozygotes who inherit the mutant allele from the mother can develop Niemann-Pick disease. More than 180 mutations have been found in the acid sphingomyelinase gene, and there seems to be a correlation between the type of mutation, the severity of enzyme deficiency, and the phenotype.

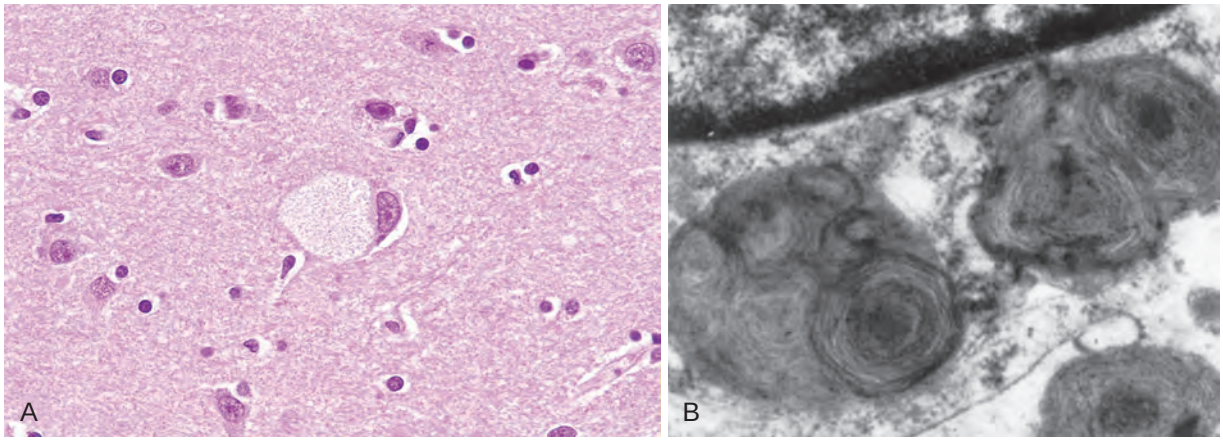


Figure 5.11 Ganglion cells in Tay-Sachs disease. (A) Under the light microscope, a large neuron has obvious lipid vacuolation. (B) A portion of a neuron under the electron microscope shows prominent lysosomes with whorled configurations. Part of the nucleus is shown above. (A, Courtesy Dr. Arthur Weinberg, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.; B, Courtesy Dr. Joe Rutledge, University of Texas Southwestern Medical Center, Dallas, Tex.)

MORPHOLOGY

In the classic infantile type A variant, a missense mutation causes almost complete deficiency of sphingomyelinase. Sphingomyelin is a ubiquitous component of cellular (including organellar) membranes, and so the enzyme deficiency blocks degradation of the lipid, resulting in its progressive accumulation within lysosomes, particularly within cells of the mononuclear phagocyte system. **Affected cells become enlarged, sometimes to 90 μm in diameter, due to the distention of lysosomes with sphingomyelin and cholesterol.** Innumerable small vacuoles of relatively uniform size are created, imparting foaminess to the cytoplasm (Fig. 5.12). Electron microscopy confirms that the vacuoles are engorged secondary lysosomes that often contain membranous cytoplasmic bodies resembling concentric lamellated myelin figures, sometimes called zebra bodies.

The lipid-laden phagocytic foam cells are widely distributed in the spleen, liver, lymph nodes, bone marrow, tonsils, gastrointestinal tract, and lungs. **The involvement of the spleen generally produces massive enlargement,** sometimes to 10 times its normal weight, but the hepatomegaly is usually not quite so striking. The lymph nodes are generally moderately to markedly enlarged throughout the body.

The neuronal involvement is diffuse, affecting all parts of the nervous system. **Vacuolation and ballooning of neurons** constitute the dominant histologic change, which in time leads to cell death and loss of brain substance. A **retinal cherry-red spot** similar to that seen in Tay-Sachs disease is present in about one-third to one-half of affected individuals.

Clinical manifestations in type A disease may be present at birth and almost invariably become evident by age 6 months. Infants typically have a protuberant abdomen because of hepatosplenomegaly. Once the manifestations appear, they are followed by progressive failure to thrive, vomiting, fever, and generalized lymphadenopathy as well as progressive deterioration of psychomotor function. Death occurs usually within the first or second year of life.

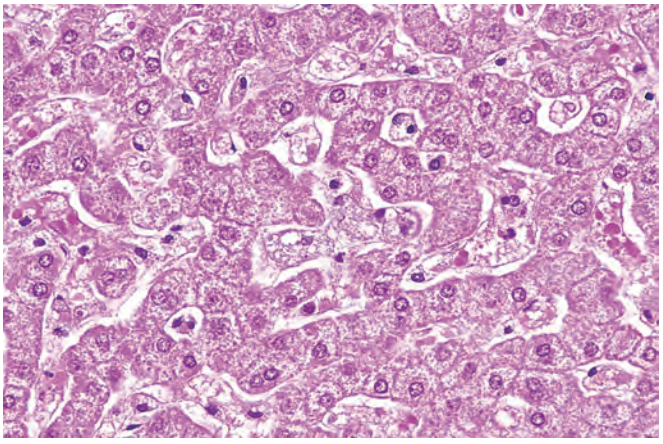


Figure 5.12 Niemann-Pick disease in liver. The hepatocytes and Kupffer cells have a foamy, vacuolated appearance due to deposition of lipids. (Courtesy Dr. Arthur Weinberg, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.)

The diagnosis is established by biochemical assays for sphingomyelinase activity in leukocytes or bone marrow biopsy. Individuals affected with types A and B as well as carriers can be detected by DNA analysis.

Niemann-Pick Disease Type C

Although previously considered to be related to types A and B, Niemann-Pick disease type C is distinct at the biochemical and genetic levels and is more common than types A and B combined. Mutations in two related genes, *NPC1* and *NPC2*, can give rise to Niemann-Pick disease type C, with *NPC1* being responsible for 95% of cases. Unlike most other storage diseases, Niemann-Pick disease type C is due to a primary defect in nonenzymatic lipid transport. NPC1 is membrane bound, whereas NPC2 is soluble. Both are involved in the transport of free cholesterol from the lysosomes to the cytoplasm. Niemann-Pick disease type C is clinically heterogeneous. It may present as hydrops fetalis and stillbirth, as neonatal hepatitis, or, most commonly, as a chronic form characterized by progressive neurologic damage. The clinical course is marked by ataxia, vertical supranuclear gaze palsy, dystonia, dysarthria, and psychomotor regression.

Gaucher Disease

Gaucher disease refers to a cluster of autosomal recessive disorders resulting from mutations in the gene encoding glucocerebrosidase. It is the most common lysosomal storage disorder. The affected gene encodes glucocerebrosidase, an enzyme that normally cleaves the glucose residue from ceramide. As a result of the enzyme defect, glucocerebrosides accumulate principally in phagocytes but in some subtypes also in the central nervous system. Glucocerebrosides are continually formed from the catabolism of glycolipids derived mainly from the cell membranes of senescent leukocytes and red cells. It is clear now that the pathologic changes in Gaucher disease are caused not just by the burden of storage material but also by activation of macrophages and the consequent secretion of cytokines such as interleukin (IL)-1, IL-6, and tumor necrosis factor (TNF).

Three clinical subtypes of Gaucher disease have been distinguished:

- The most common, accounting for 99% of cases, is type I, or the chronic nonneuronopathic form. In this type, storage of glucocerebrosides is limited to the mononuclear phagocytes throughout the body without involving the brain. Splenic and skeletal involvements dominate this pattern of the disease. It is found principally in Jews of European stock. Individuals with this disorder have reduced but detectable levels of glucocerebrosidase activity. Longevity is shortened, but not markedly.
- Type II, or acute neuronopathic Gaucher disease, is the infantile acute cerebral pattern. This form has no predilection for Jews. In these patients there is virtually no detectable glucocerebrosidase activity in the tissues. Hepatosplenomegaly is also seen in this form of Gaucher disease, but the clinical picture is dominated by progressive central nervous system involvement leading to death at an early age.
- A third pattern, type III, is intermediate between types I and II. These patients have the systemic involvement characteristic of type I but have progressive central

nervous system disease that usually begins in adolescence or early adulthood.

MORPHOLOGY

Glucocerebrosides accumulate in massive amounts within phagocytic cells throughout the body in all forms of Gaucher disease. The **distended phagocytic cells, known as Gaucher cells, are found in the spleen, liver, bone marrow, lymph nodes, tonsils, thymus, and Peyer patches.** Similar cells may be found in both the alveolar septa and the air spaces in the lung. In contrast to other lipid storage diseases, Gaucher cells rarely appear vacuolated, but instead have a fibrillary type of cytoplasm likened to crumpled tissue paper (Fig. 5.13). Gaucher cells are often enlarged, sometimes up to 100 μm in diameter, and have one or more dark, eccentrically placed nuclei. Periodic acid–Schiff staining is usually intensely positive. With the electron microscope **the fibrillary cytoplasm can be resolved as elongated, distended lysosomes,** containing the stored lipid in stacks of bilayers.

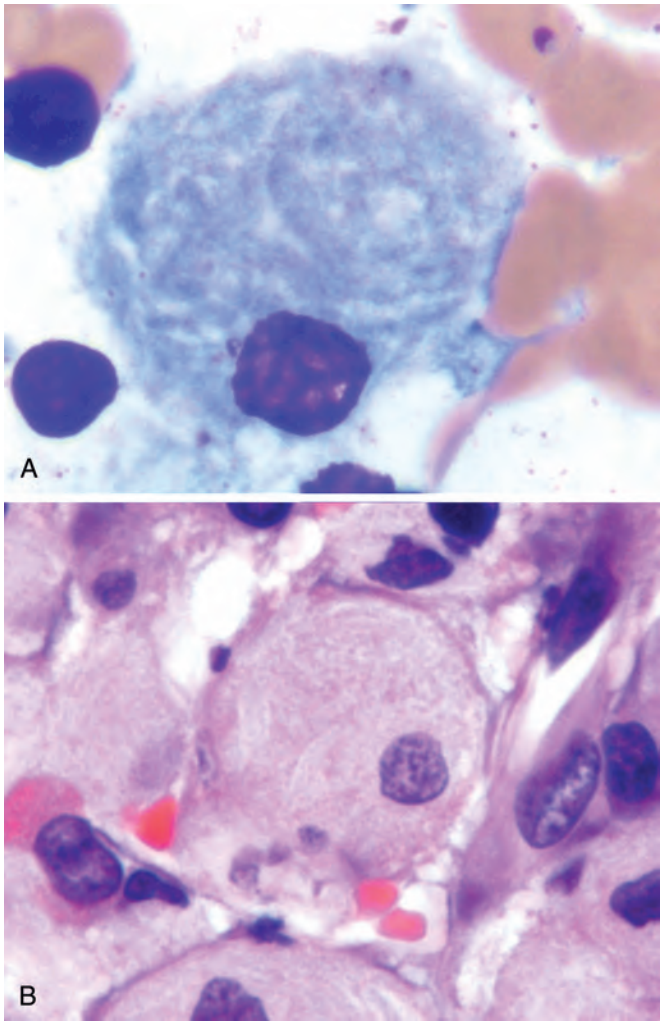


Figure 5.13 Gaucher disease involving the bone marrow. (A–B) Gaucher cells are plump macrophages that characteristically have the appearance in the cytoplasm of crumpled tissue paper due to accumulation of glucocerebroside. (A) Wright stain; (B) hematoxylin and eosin. (Courtesy Dr. John Anastasi, Department of Pathology, University of Chicago, Chicago, Ill.)

In type I disease, the **spleen is enlarged, sometimes up to 10 kg.** The lymphadenopathy is mild to moderate and is body-wide. Accumulation of Gaucher cells in the bone marrow occurs in 70% to 100% of cases of type I Gaucher disease. It produces areas of **bone erosion** that are sometimes small but in other cases sufficiently large to give rise to pathologic fractures. Bone destruction occurs due to the secretion of cytokines by activated macrophages. In patients with cerebral involvement, Gaucher cells are seen in the Virchow-Robin spaces, and arterioles are surrounded by swollen adventitial cells. There is no storage of lipids in the neurons, yet neurons appear shriveled and are progressively destroyed. It is suspected that the lipids that accumulate in the phagocytic cells around blood vessels secrete cytokines that damage nearby neurons.

Mutation of the glucocerebrosidase gene is the most common known genetic risk factor for development of Parkinson disease. Patients with Gaucher disease have a 20-fold higher risk of developing Parkinson disease (compared with controls), and 5% to 10% of patients with Parkinson disease have mutations in the gene encoding glucocerebrosidase. There is a reciprocal relationship between the level of this enzyme and aggregation of α-synuclein, the protein involved in the pathogenesis of Parkinson disease (Chapter 24). As discussed earlier, patients with lysosomal diseases have impairment in autophagy, which allows aggregation and persistence of proteins like α-synuclein.

The level of glucocerebrosides in leukocytes or cultured fibroblasts is helpful in diagnosis and in the detection of heterozygote carriers. DNA testing is also available in select populations.

Clinical Features

The course of Gaucher disease depends on the clinical subtype. In type I, symptoms and signs first appear in adult life and are related to splenomegaly or bone involvement. Most commonly there is pancytopenia or thrombocytopenia secondary to hypersplenism. Pathologic fractures and bone pain occur if there has been extensive expansion of the marrow space. Although the disease is progressive in adults, it is compatible with long life. In types II and III, central nervous system dysfunction, convulsions, and progressive mental deterioration dominate, although organs such as the liver, spleen, and lymph nodes are also affected. The diagnosis in homozygotes can be made by measurement of glucocerebrosidase activity in peripheral blood leukocytes or in extracts of cultured skin fibroblasts. The enzyme assay cannot identify heterozygotes because the levels of glucocerebrosidase are difficult to distinguish from those in normal cells. In principle, heterozygotes can be identified by detection of mutations. However, because more than 150 mutations in the glucocerebrosidase gene can cause Gaucher disease, currently it is not possible to use a single genetic test. However, when the causative mutation in a patient is known, a heterozygote can be identified with molecular tests. The landscape of genetic testing is rapidly changing with the application of next-generation sequencing.

Replacement therapy with recombinant enzymes is the mainstay for treatment of Gaucher disease; it is effective, and

those with type I disease can expect normal life expectancy. However, such therapy is extremely expensive. Because the fundamental defect resides in mononuclear phagocytic cells originating from marrow stem cells, allogeneic hematopoietic stem cell transplantation can be curative. Other work is directed toward correction of the enzyme defect by transfer of the normal glucocerebrosidase gene into the patient's hematopoietic stem cells. Substrate reduction therapy with inhibitors of glucosylceramide synthetase is also being evaluated.

Mucopolysaccharidoses

MPSs are a group of closely related syndromes that result from genetically determined deficiencies of enzymes involved in the degradation of mucopolysaccharides (glycosaminoglycans). Chemically, mucopolysaccharides are long-chain complex carbohydrates that are linked with proteins to form proteoglycans. They are abundant in extracellular matrix, joint fluid, and connective tissue. The glycosaminoglycans that accumulate in MPSs are dermatan sulfate, heparan sulfate, keratan sulfate, and chondroitin sulfate. Eleven enzymes involved in the degradation of these molecules cleave terminal sugars from the polysaccharide chains disposed along a polypeptide or core protein. In the absence of enzymes, these chains accumulate within lysosomes in various tissues and organs of the body.

There are 11 clinical variants of MPS, each resulting from the deficiency of one specific enzyme (some variants have subvariants). All the MPSs except one are inherited as autosomal recessive traits; the exception, *Hunter syndrome,* is an X-linked recessive trait. Within a given group (e.g., MPS I, characterized by a deficiency of α-L-iduronidase), subgroups exist that result from different mutant alleles at the same genetic locus. Thus the severity of enzyme deficiency and the clinical picture even within subgroups are often different.

In general, MPSs are progressive disorders, characterized by *coarse facial features, clouding of the cornea, joint stiffness, and intellectual disability.* Urinary excretion of the accumulated mucopolysaccharides is often increased and is used as a diagnostic tool.

MORPHOLOGY

The accumulated **mucopolysaccharides are generally found in mononuclear phagocytic cells, endothelial cells, intimal smooth muscle cells, and fibroblasts** throughout the body. Common sites of involvement are thus the spleen, liver, bone marrow, lymph nodes, blood vessels, and heart.

Microscopically, affected cells are distended and have apparent clearing of the cytoplasm to create so-called balloon cells. Under the electron microscope, the clear cytoplasm can be resolved as numerous minute vacuoles. These are swollen lysosomes containing a finely granular periodic acid–Schiff–positive material that can be identified biochemically as mucopolysaccharide. Similar lysosomal changes are found in the neurons of those syndromes characterized by central nervous system involvement. In addition, however, some of the lysosomes in neurons are replaced by lamellated zebra bodies similar to those seen in Niemann-Pick disease. **Hepatosplenomegaly, skeletal deformities, valvular lesions and subendothelial arterial deposits, particularly in the coronary arteries, and lesions in the brain are common threads that run through all the MPSs.** In many of the more protracted syndromes, coronary subendothelial lesions lead to myocardial ischemia. Thus myocardial infarction and cardiac decompensation are important causes of death.

Clinical Features

Of the 11 recognized variants, only 2 well-characterized syndromes are described briefly here. *Hurler syndrome,* also called MPS I-H, results from a deficiency of α-L-iduronidase. It is one of the most severe forms of MPS. Affected children appear normal at birth but develop hepatosplenomegaly by age 6 to 24 months. Their growth is retarded, and, as in other forms of MPS, they develop coarse facial features and skeletal deformities. Death occurs by age 6 to 10 years and is often due to cardiovascular complications. *Hunter syndrome,* also called MPS II, differs from Hurler syndrome in mode of inheritance (X-linked), absence of corneal clouding, and milder clinical course.

KEY CONCEPTS

LYSOSOMAL STORAGE DISEASES

Inherited mutations leading to defective lysosomal enzyme functions gives rise to accumulation and storage of complex substrates in the lysosomes and defects in autophagy resulting in cellular injury.

- Tay-Sachs disease is caused by an inability to metabolize G_{M2} gangliosides due to lack of the α-subunit of lysosomal hexosaminidase. G_{M2} gangliosides accumulate in the central nervous system and cause severe intellectual disability, blindness, motor weakness, and death by 2 to 3 years of age.
- Niemann-Pick disease types A and B are caused by a deficiency of sphingomyelinase. In the more severe type A variant, accumulation of sphingomyelin in the nervous system results in neuronal damage. Lipid also is stored in phagocytes within the liver, spleen, bone marrow, and lymph nodes, causing their enlargement. In type B, neuronal damage is not present.
- Niemann-Pick disease type C is caused by a defect in cholesterol transport and resultant accumulation of cholesterol and gangliosides in the nervous system. Affected children most commonly exhibit ataxia, dysarthria, and psychomotor regression.
- Gaucher disease results from lack of the lysosomal enzyme glucocerebrosidase and accumulation of glucocerebroside in mononuclear phagocytic cells. In the most common, type I variant, affected phagocytes become enlarged (Gaucher cells) and accumulate in liver, spleen, and bone marrow, causing hepatosplenomegaly and bone erosion. Types II and III are characterized by variable neuronal involvement. Gaucher disease has a strong association with Parkinson disease.
- MPSs result in accumulation of mucopolysaccharides in many tissues including liver, spleen, heart, blood vessels, brain, cornea, and joints. Affected patients in all forms have coarse facial features. Manifestations of Hurler syndrome include corneal clouding, coronary arterial and valvular deposits, and death in childhood. Hunter syndrome is associated with a milder clinical course.

Glycogen Storage Diseases (Glycogenoses)

The glycogen storage diseases result from a hereditary deficiency of one of the enzymes involved in the synthesis or sequential degradation of glycogen. Depending on the

normal tissue or organ distribution of the specific enzyme, glycogen storage in these disorders may be limited to a few tissues, may be more widespread while not affecting all tissues, or may be systemic.

The significance of a specific enzyme deficiency is best understood from the perspective of the normal metabolism of glycogen (Fig. 5.14). Glycogen is a storage form of glucose. Glycogen synthesis begins with the conversion of glucose to glucose-6-phosphate by the action of a hexokinase (glucokinase). A phosphoglucomutase then transforms the glucose-6-phosphate to glucose-1-phosphate, which, in turn, is converted to uridine diphosphoglucose. A highly branched, large polymer is then built (molecular weight as high as 100 million), containing as many as 10,000 glucose molecules linked together by α-1,4-glucoside bonds. The glycogen chain and branches continue to be elongated by the addition of glucose molecules mediated by glycogen synthetases. During degradation, distinct phosphorylases in the liver and muscle split glucose-1-phosphate from the glycogen until about four glucose residues remain on each branch, leaving a branched oligosaccharide called limit dextrin. This can be further degraded only by the debranching enzyme. In addition to these major pathways, glycogen is also degraded in the lysosomes by acid alpha-glucosidase. If the lysosomes are deficient in this enzyme, the glycogen contained within them is not accessible to degradation by cytoplasmic enzymes such as phosphorylases.

On the basis of specific enzyme deficiencies and the resultant clinical pictures, glycogenoses have traditionally been divided into a dozen or so syndromes designated by

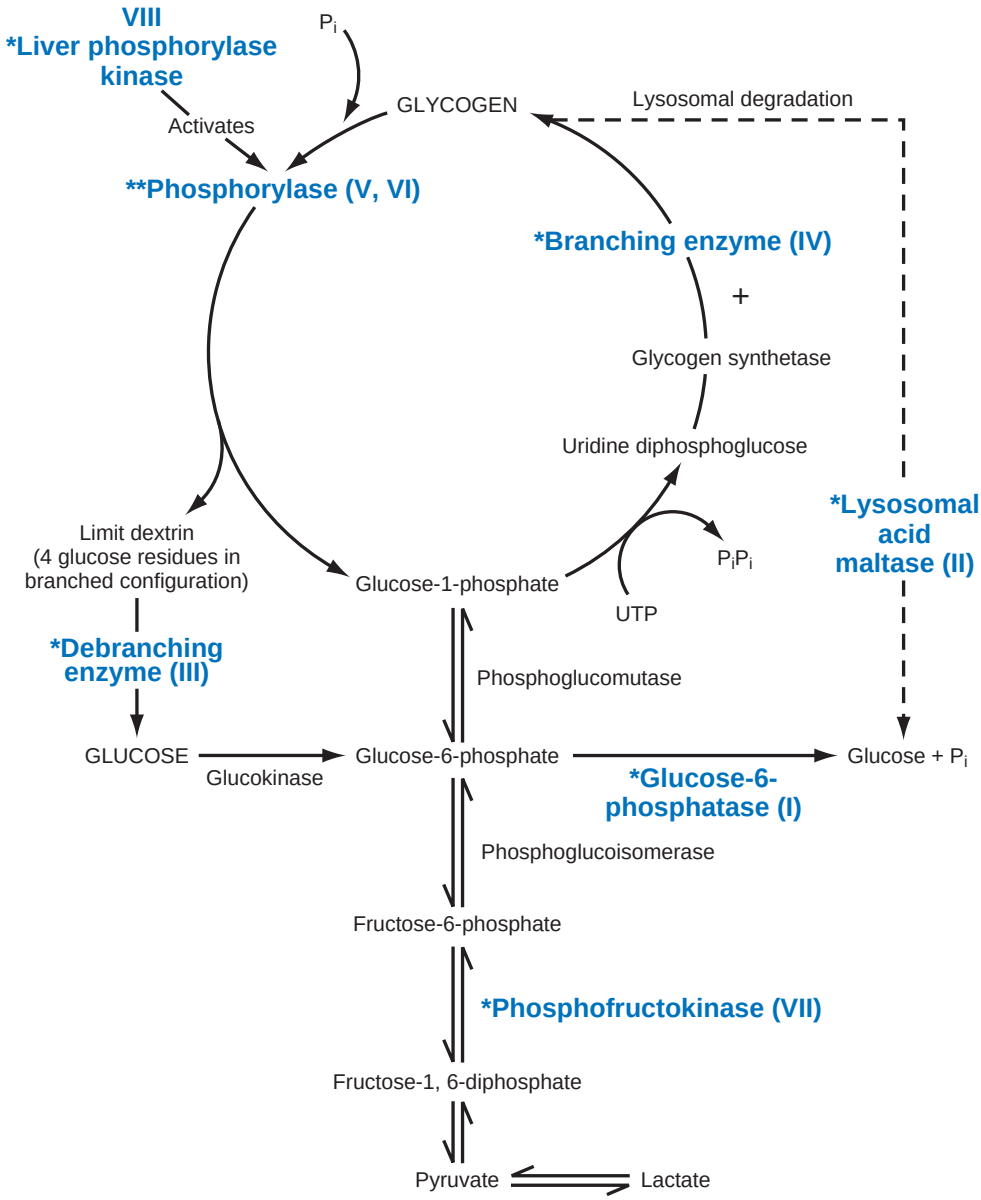


Figure 5.14 Pathways of glycogen metabolism. *Asterisks* mark the enzyme deficiencies associated with glycogen storage diseases. Roman numerals indicate the type of glycogen storage disease associated with the given enzyme deficiency. Types V and VI result from deficiencies of muscle and liver phosphorylases, respectively. (Modified from Hers H, et al: Glycogen storage diseases. In Scriver CR, et al, editors: *The Metabolic Basis of Inherited Disease*, ed 6, New York, 1989, McGraw-Hill, p 425.)

Table 5.7 Principal Subgroups of Glycogenoses

Clinicopathologic Category	Specific Type	Enzyme Deficiency	Morphologic Changes	Clinical Features
Hepatic type	Hepatorenal—von Gierke disease (type I)	Glucose-6-phosphatase	Hepatomegaly—intracytoplasmic accumulations of glycogen and small amounts of lipid; intranuclear glycogen Renomegaly—intracytoplasmic accumulations of glycogen in cortical tubular epithelial cells	In untreated patients: failure to thrive, stunted growth, hepatomegaly, and renomegaly Hypoglycemia due to failure of glucose mobilization, often leading to convulsions Hyperlipidemia and hyperuricemia resulting from deranged glucose metabolism; many patients develop gout and skin xanthomas Bleeding tendency due to platelet dysfunction With treatment: most survive and develop late complications (e.g., hepatic adenomas)
Myopathic type	McArdle disease (type V)	Muscle phosphorylase	Skeletal muscle only—accumulations of glycogen predominant in subsarcolemmal location	Painful cramps associated with strenuous exercise; myoglobinuria occurs in 50% of cases; onset in adulthood (>20 years); muscular exercise fails to raise lactate level in venous blood; serum creatine kinase always elevated; compatible with normal longevity
Miscellaneous types	Generalized glycogenosis—Pompe disease (type II)	Lysosomal acid alpha-glucosidase	Mild hepatomegaly—ballooning of lysosomes with glycogen, creating lacy cytoplasmic pattern Cardiomegaly—glycogen within sarcoplasm as well as membrane-bound Skeletal muscle—similar to changes in heart	Massive cardiomegaly, muscle hypotonia, and cardiorespiratory failure within 2 years; a milder adult form with only skeletal muscle involvement, presenting with chronic myopathy; enzyme replacement therapy available

Roman numerals, and the list continues to grow. On the basis of pathophysiology glycogenoses can be divided into three major subgroups (Table 5.7).

- *Hepatic forms.* The liver is a key player in glycogen metabolism. It contains enzymes that synthesize glycogen for storage and ultimately break it down into free glucose, which is then released into the blood. An inherited deficiency of hepatic enzymes that are involved in glycogen degradation therefore leads not only to the accumulation of glycogen in the liver but also to a reduction in blood glucose concentrations (hypoglycemia) (Fig. 5.15). Deficiency of the enzyme glucose-6-phosphatase (von Gierke disease, or type I glycogenosis) is a prime example of the hepatic-hypoglycemic form of glycogen storage disease (see Table 5.7). Other examples include deficiencies of liver phosphorylase and debranching enzyme, both involved in the breakdown of glycogen (see Fig. 5.15). In all these disorders, glycogen is stored in many organs, but hepatic enlargement and hypoglycemia dominate the clinical picture.
- *Myopathic forms.* In the skeletal muscles, as opposed to the liver, glycogen is used predominantly as a source of energy during physical activity. Adenosine triphosphate (ATP) is generated by glycolysis, which leads ultimately to the formation of lactate (Fig. 5.16). If the enzymes that fuel the glycolytic pathway are deficient, glycogen storage occurs in the muscles and is associated with muscular weakness due to impaired energy production. Examples in this category include deficiencies of muscle phosphorylase (McArdle disease, or type V glycogenosis), muscle phosphofructokinase (type VII glycogen storage disease), and several others. Typically, individuals with the myopathic forms present with muscle cramps after exercise, and lactate levels in the blood fail to rise after exercise due to a block in glycolysis.
- *Glycogen storage diseases associated with (1) deficiency of acid alpha-glucosidase (acid maltase) and (2) lack of branching enzyme* do not fit into the hepatic or myopathic categories. They are associated with glycogen storage in many organs and death early in life. Acid alpha-glucosidase is a lysosomal enzyme, and hence its deficiency leads to lysosomal storage of glycogen (type II glycogenosis, or *Pompe disease*) in all organs, but *cardiomegaly is the most prominent feature* (see Fig. 5.16).

KEY CONCEPTS

GLYCOGEN STORAGE DISEASES

- Inherited deficiency of enzymes involved in glycogen metabolism can result in storage of normal or abnormal forms of glycogen, predominantly in liver or muscles, but also in other tissues as well.
- In the most common hepatic form (von Gierke disease), liver cells store glycogen because of a lack of hepatic glucose-6-phosphatase. The liver is enlarged, and patients have hypoglycemia.
- There are several myopathic forms, including McArdle disease, in which lack of muscle phosphorylase gives rise to storage in skeletal muscles and cramps after exercise.
- In Pompe disease there is lack of lysosomal acid alpha-glucosidase, and all organs are affected, but heart involvement is predominant.

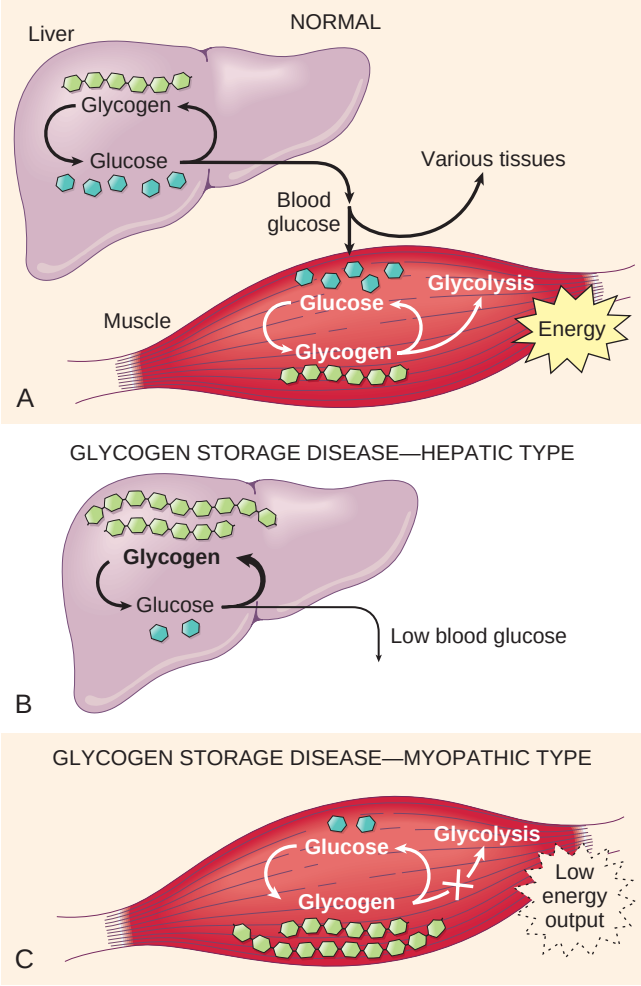


Figure 5.15 (A) Normal glycogen metabolism in the liver and skeletal muscles. (B) Effects of an inherited deficiency of hepatic enzymes involved in glycogen metabolism. (C) Consequences of a genetic deficiency in the enzymes that metabolize glycogen in skeletal muscles.

Disorders Associated With Defects in Proteins That Regulate Cell Growth

Normal growth and differentiation of cells are regulated by two classes of genes: proto-oncogenes and tumor suppressor genes, whose products promote or restrain cell growth (Chapter 7). It is well established that mutations in these two classes of genes are important in the pathogenesis of tumors. In the vast majority of cases, cancer-causing mutations affect somatic cells and hence are not passed in the germline. In approximately 5% of all cancers, however, mutations transmitted through the germline contribute to the development of cancer. Most familial cancers are inherited in an autosomal dominant fashion, but a few recessive disorders have also been described. This subject is discussed in Chapter 7. Specific forms of familial tumors are described in various chapters.

COMPLEX MULTIGENIC DISORDERS

As discussed previously, such disorders are caused by interactions between variant forms of genes and environmental factors. A gene that has at least two alleles, each of which occurs at a frequency of at least 1% in the population, is polymorphic, and each variant allele is referred to as a polymorphism. According to the common disease/common variant hypothesis, complex genetic disorders occur when many polymorphisms, each with a modest effect and low penetrance, are coinherited. Two additional facts have emerged from studies of common complex disorders, such as type 1 diabetes:

- While complex disorders result from the collective inheritance of many polymorphisms, different polymorphisms vary in significance. For example, of the 20 to 30 genes implicated in type 1 diabetes, 6 or 7 are most important, and a few HLA alleles contribute more than 50% of the risk (Chapter 24).

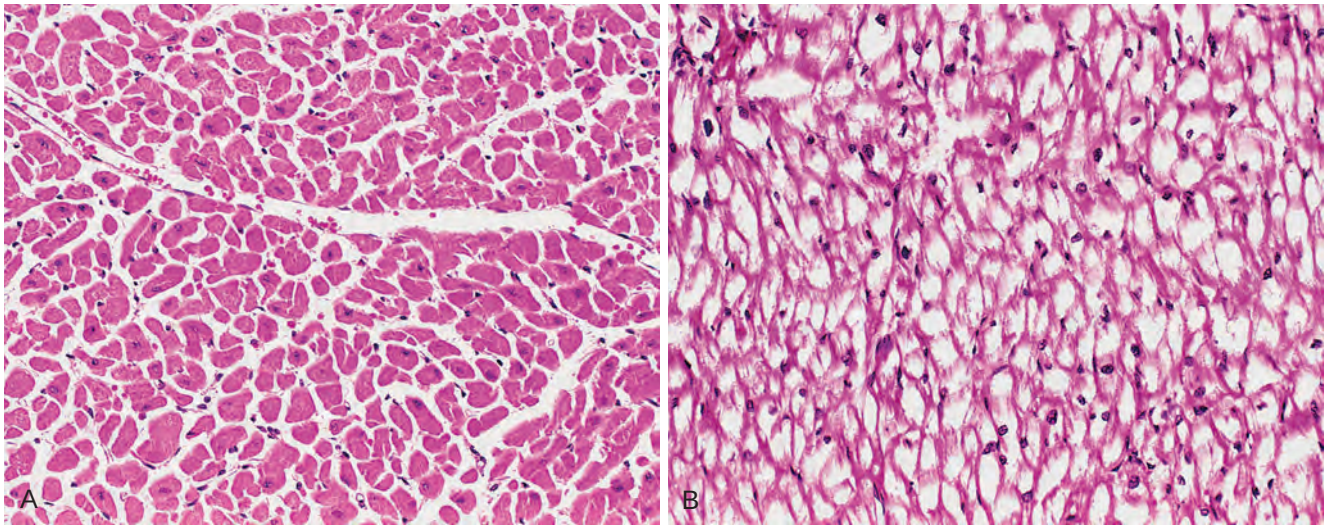


Figure 5.16 Pompe disease (glycogen storage disease type II). (A) Normal myocardium with abundant eosinophilic cytoplasm. (B) Patient with Pompe disease (same magnification) showing the myocardial fibers full of glycogen seen as clear spaces. (Courtesy Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.)

- Some polymorphisms are common to multiple diseases of the same type, while others are disease specific. This is best illustrated in autoimmune diseases (Chapter 6).

Several normal phenotypic characteristics are governed by multifactorial inheritance such as hair color, eye color, skin color, height, and intelligence. These characteristics show a continuous variation in population groups, producing the standard bell-shaped curve of distribution. Environmental influences, however, significantly modify the phenotypic expression of complex traits. For example, type 2 diabetes has many of the features of a multifactorial disorder. It is well recognized that individuals often first manifest this disease after weight gain. Thus, obesity as well as other environmental influences unmask the diabetic genetic trait. Nutritional influences may cause even monozygous twins to achieve different heights. A culturally deprived child cannot achieve his or her full intellectual capacity.

Assigning a disease to this mode of inheritance must be done with caution. It depends on many factors but first on familial clustering and the exclusion of Mendelian and chromosomal modes of transmission. A range of levels of severity of a disease is suggestive of a complex multigenic disorder, but, as pointed out earlier, variable expressivity and reduced penetrance of single mutant genes may also account for this phenomenon. Because of these problems, sometimes it is difficult to distinguish between Mendelian and multifactorial disease.

CHROMOSOMAL DISORDERS

Normal Karyotype

As you will remember, human somatic cells contain 46 chromosomes—22 homologous pairs of autosomes and two sex chromosomes, XX in the female and XY in the male. The study of chromosomes—*karyotyping*—is the basic tool of the cytogeneticist. The usual procedure to examine chromosomes is to arrest dividing cells in metaphase with mitotic spindle inhibitors (e.g., *N*-diacetyl-*N*-methylcolchicine [colcemid]) and then to stain the chromosomes. In a metaphase spread, the individual chromosomes take the form of two chromatids connected at the centromere. A karyotype is obtained by arranging each pair of autosomes according to length, followed by sex chromosomes.

A variety of staining methods have been developed that allow identification of individual chromosomes on the basis of distinctive and reliable patterns of alternating light and dark bands. The one most commonly used involves a Giemsa stain and is hence called *G banding*. A normal male karyotype with G banding is illustrated in Fig. 5.17. With standard G banding, approximately 400 to 800 bands per haploid set can be detected. The resolution obtained by banding can be markedly improved by obtaining the cells in prophase. The individual chromosomes appear markedly elongated, and as many as 1500 bands per karyotype can be recognized. The use of these banding techniques permits certain identification

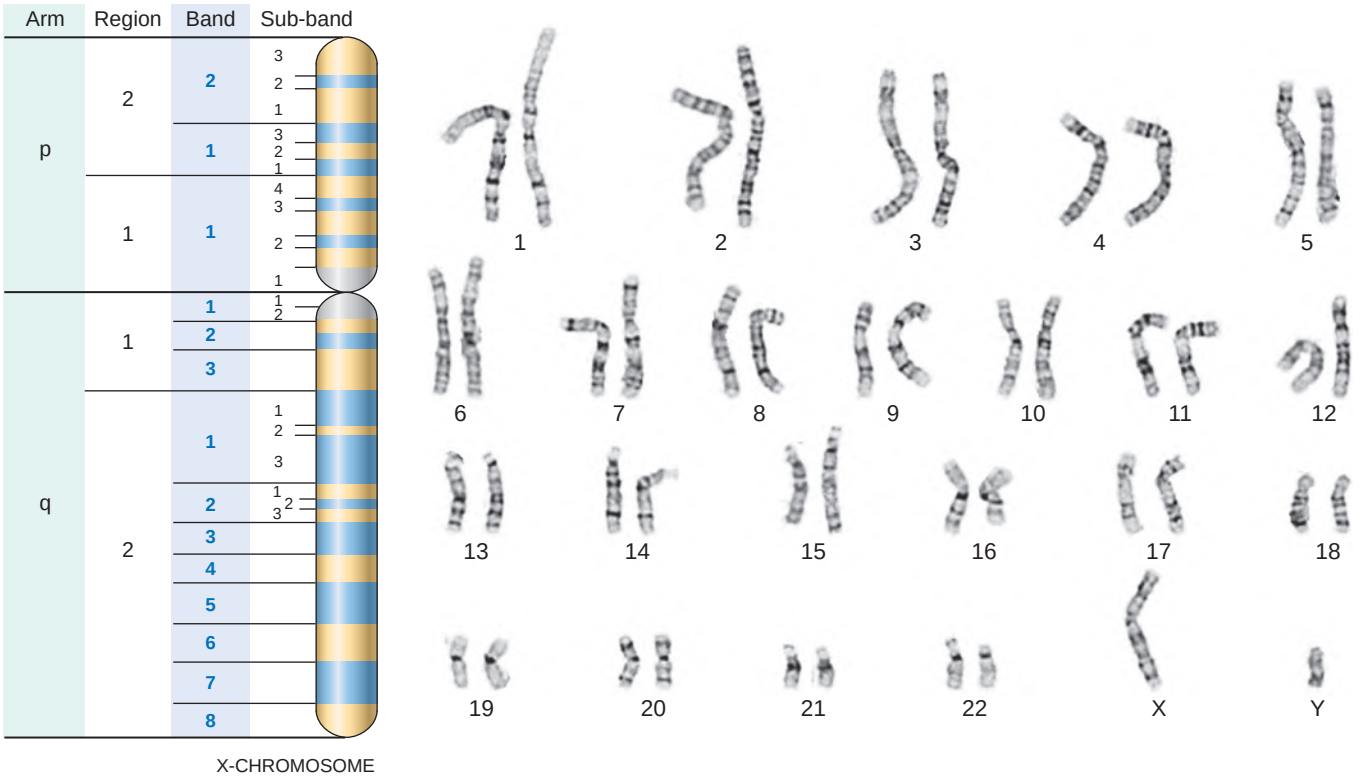


Figure 5.17 G-banded karyotype from a normal male (46,XY). Also shown is the banding pattern of the X chromosome with nomenclature of arms, regions, bands, and sub-bands. (Courtesy Dr. Stuart Schwartz, Department of Pathology, University of Chicago, Chicago, Ill.)

of each chromosome and roughly delineates breakpoints and other gross alterations (described later).

Commonly Used Cytogenetic Terminology

Karyotypes are usually described using a shorthand system of notations in the following order: total number of chromosomes is given first, followed by the sex chromosome complement, and finally the description of abnormalities in ascending numerical order. For example, a male with trisomy 21 is designated *47,XY,+21*. Notations denoting structural alterations of chromosomes and their corresponding abnormalities are described later.

The short arm of a chromosome is designated *p* (for petit), and the long arm is referred to as *q* (the next letter of the alphabet). In a banded karyotype, each arm of the chromosome is divided into two or more regions bordered by prominent bands. The regions are numbered (e.g., 1, 2, 3) from the centromere outward. Each region is further subdivided into bands and sub-bands, and these are ordered numerically as well (see Fig. 5.17). Thus the notation *Xp21.2* refers to a chromosomal segment located on the short arm of the X chromosome, in region 2, band 1, and sub-band 2.

Structural Abnormalities of Chromosomes

The aberrations underlying cytogenetic disorders may take the form of an abnormal number of chromosomes or alterations in the structure of one or more chromosomes. The normal chromosome complement is expressed as 46,XX for females and 46,XY for males. Any exact multiple of the haploid number of chromosomes (23) is called *euploid.* If an error occurs in meiosis or mitosis and a cell acquires a chromosome complement that is not an exact multiple of 23, it is referred to as *aneuploidy*. The usual causes for aneuploidy are *nondisjunction* and *anaphase lag*. When nondisjunction occurs during gametogenesis, the gametes formed have either an extra chromosome (n + 1) or one less chromosome (n – 1). Fertilization of such gametes by normal gametes results in two types of zygotes—trisomic (2n + 1) or monosomic (2n – 1). In anaphase lag, one homologous chromosome in meiosis or one chromatid in mitosis lags behind and is left out of the cell nucleus. The result is one normal cell and one cell with monosomy. As discussed later, monosomies or trisomies involving the sex chromosomes, or even more bizarre aberrations, are compatible with life and are usually associated with variable degrees of phenotypic abnormalities. Monosomy involving an autosome generally causes loss of too much genetic information to permit live birth or even embryogenesis, but several autosomal trisomies do permit survival. With the exception of trisomy 21, all exhibit severe malformations and almost invariably die at an early age.

Occasionally, mitotic errors in early development give rise to two or more populations of cells with different chromosomal complement in the same individual, a condition referred to as *mosaicism*. Mosaicism can result from mitotic errors during the cleavage of the fertilized ovum or in somatic cells. Mosaicism affecting the sex chromosomes is relatively common. In the division of the fertilized ovum, an error may lead to one of the daughter cells receiving three sex chromosomes, whereas the other receives only one, yielding, for example, a 45,X/47,XXX mosaic. All descendant cells derived from each of these precursors thus have either a 47,XXX complement or a 45,X complement. Such a patient is a mosaic variant of Turner syndrome, with the extent of phenotypic expression dependent on the number and distribution of the 45,X cells.

Autosomal mosaicism seems to be much less common than that involving the sex chromosomes. An error in an early mitotic division affecting the autosomes usually leads to a nonviable mosaic due to autosomal monosomy. Rarely, the nonviable cell population is lost during embryogenesis, yielding a viable mosaic (e.g., 46,XY/47,XY,+21). Such a patient is a trisomy 21 mosaic with variable expression of Down syndrome, depending on the proportion of cells containing the trisomy.

A second category of chromosomal aberrations is associated with changes in the structure of chromosomes. To be visible by routine banding techniques, a fairly large amount of DNA (approximately 2 to 4 million bp), containing many genes, must be involved. The resolution is much higher with fluorescence in situ hybridization (FISH), which can detect changes as small as kilobases. Structural changes in chromosomes usually result from chromosome breakage followed by loss or rearrangement of material. In the next section the more common forms of alterations in chromosome structure and the notations used to signify them are reviewed.

Deletion refers to loss of a portion of a chromosome (Fig. 5.18). Most deletions are interstitial, but rarely terminal deletions may occur. Interstitial deletions occur when there are two breaks within a chromosome arm, followed by loss of the chromosomal material between the breaks and fusion of the broken ends. One can specify in which regions and at what bands the breaks have occurred. For example, *46,XY,del(16)(p11.2p13.1)* describes breakpoints in the short arm of chromosome 16 at 16p11.2 and 16p13.1 with loss of material between breaks. Terminal deletions result from a single break in a chromosome arm, producing a fragment with no centromere, which is then lost at the next cell division. The deleted end of the retained chromosome is protected by acquiring telomeric sequences.

A *ring chromosome* is a special form of deletion. It is produced when a break occurs at both ends of a chromosome with fusion of the damaged ends (see Fig. 5.18). If significant genetic material is lost, phenotypic abnormalities result. This might be expressed as *46,XY,r(14)*. Ring chromosomes do not behave normally in meiosis or mitosis and usually result in serious consequences.

Inversion refers to a rearrangement that involves two breaks within a single chromosome with reincorporation of the inverted, intervening segment (see Fig. 5.18). An inversion involving only one arm of the chromosome is known as *paracentric.* If the breaks are on opposite sides of the centromere, it is known as *pericentric.* Inversions are often fully compatible with normal development.

Isochromosome formation results when one arm of a chromosome is lost and the remaining arm is duplicated, resulting in a chromosome consisting of two short arms only or of two long arms (see Fig. 5.18). An isochromosome has morphologically identical genetic information in both arms. The most common isochromosome present in live births involves the long arm of the X chromosome and is designated *i(X)(q10)*. The Xq isochromosome is associated

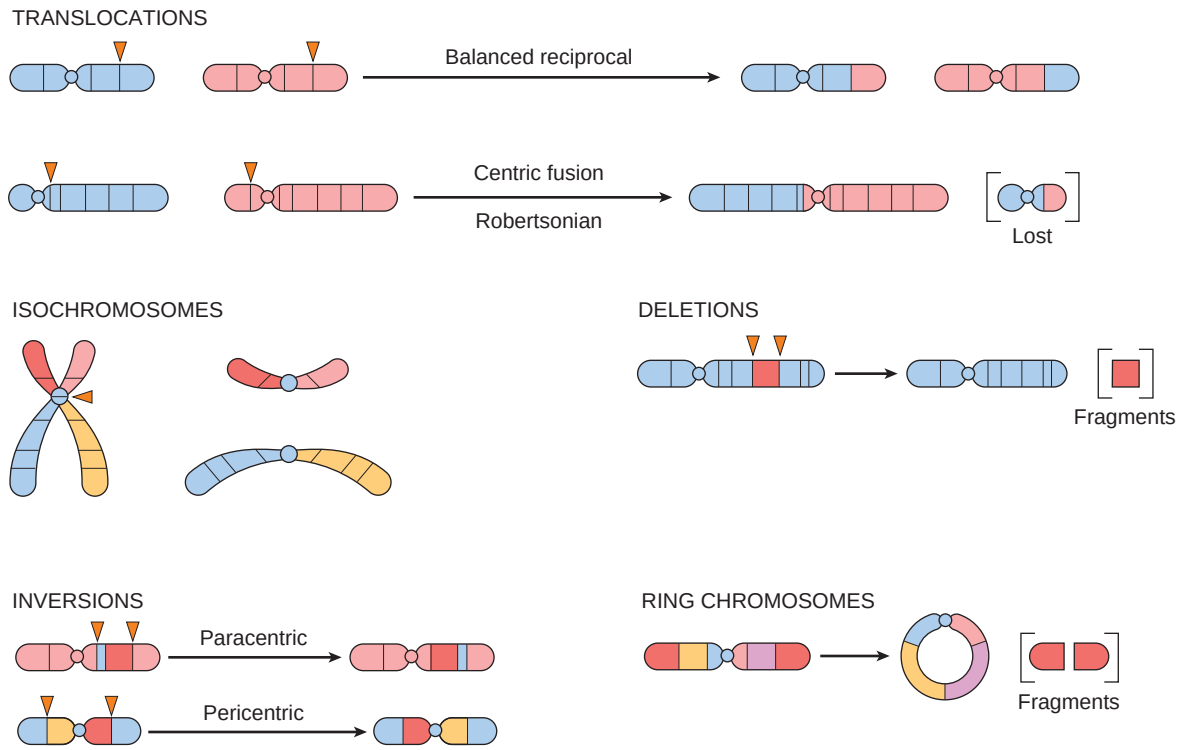


Figure 5.18 Types of chromosomal rearrangements.

with monosomy for genes on the short arm of X and with trisomy for genes on the long arm of X.

In a *translocation,* a segment of one chromosome is transferred to another (see Fig. 5.18). In one form, called *balanced reciprocal translocation,* there are single breaks in each of two chromosomes, with exchange of material. A balanced reciprocal translocation between the long arm of chromosome 2 and the short arm of chromosome 5 would be written *46,XX,t(2;5)(q31;p14)*. This individual has 46 chromosomes with altered morphology of one of the chromosomes 2 and one of the chromosomes 5. Because there has been no or very little loss of genetic material, the individual is likely to be phenotypically normal. A balanced translocation carrier, however, is at increased risk for producing abnormal gametes. For example, in the case cited earlier, a gamete containing one normal chromosome 2 and a translocated chromosome 5 may be formed. Such a gamete would be unbalanced because it would not contain the normal complement of genetic material. Subsequent fertilization by a normal gamete would lead to the formation of an abnormal (unbalanced) zygote, resulting in spontaneous abortion or birth of a malformed child. The other important pattern of translocation is called a *robertsonian translocation* (or centric fusion), a translocation between two acrocentric chromosomes. Typically the breaks occur close to the centromeres of each chromosome. Transfer of the segments then leads to one very large chromosome and one extremely small one. Usually the small product is lost (see Fig. 5.18); however, because it carries only highly redundant genes (e.g., ribosomal RNA genes), this loss is compatible with a normal phenotype. Robertsonian translocation between two chromosomes is encountered in 1 in 1000 apparently normal individuals. The significance of this form of translocation also lies in the production of abnormal progeny, as discussed later with Down syndrome.

As pointed out earlier, the clinically detected chromosome disorders represent only the "tip of the iceberg." It is estimated that approximately 7.5% of all conceptions have a chromosomal abnormality, most of which are not compatible with survival or live birth. Even in live-born infants the frequency is approximately 0.5% to 1.0%. It is beyond the scope of this book to discuss most of the clinically recognizable chromosomal disorders. Hence we focus attention on those few that are most common.

Cytogenetic Disorders Involving Autosomes

Trisomy 21 (Down Syndrome)

Down syndrome is the most common of the chromosomal disorders and is a major cause of intellectual disability. In the United States the incidence in newborns is about 1 in 700. Approximately 95% of affected individuals have trisomy 21, so their chromosome count is 47. FISH with chromosome 21–specific probes reveals the extra copy of chromosome 21 in such cases (Fig. 5.19). As mentioned earlier, the most common cause of trisomy and therefore of Down syndrome is meiotic nondisjunction. The parents of such children have a normal karyotype and are normal in all respects.

Maternal age has a strong influence on the incidence of trisomy 21. It occurs once in 1550 live births in women under age 20, in contrast to 1 in 25 live births for mothers

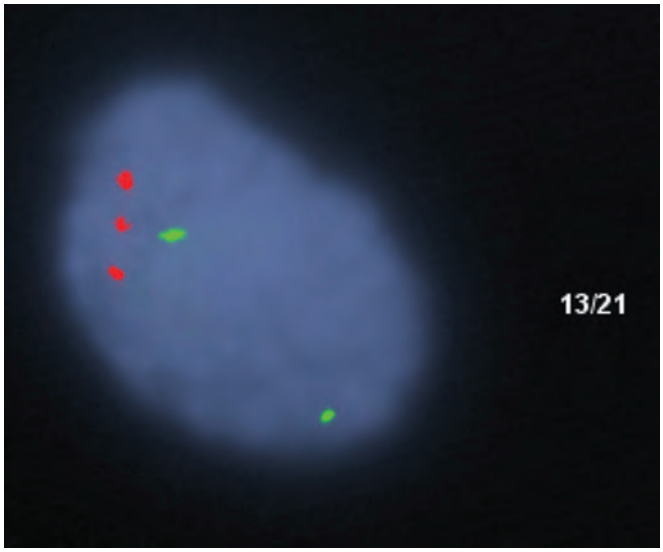


Figure 5.19 Fluorescence in situ hybridization analysis of an interphase nucleus using locus-specific probes to chromosome 13 (green) and chromosome 21 (red), revealing three red signals consistent with trisomy 21. (Courtesy Dr. Stuart Schwartz, Department of Pathology, University of Chicago, Chicago, IL.)

older than age 45. The correlation with maternal age suggests that in most cases the meiotic nondisjunction of chromosome 21 occurs in the ovum. Indeed, studies in which DNA polymorphisms were used to trace the parental origin of chromosome 21 have revealed that in 95% of the cases with trisomy 21 the extra chromosome is of maternal origin. Although many hypotheses have been advanced, the reason for the increased susceptibility of the ovum to nondisjunction remains unknown.

In about 4% of cases of Down syndrome the extra chromosomal material derives from the presence of a robertsonian translocation of the long arm of chromosome 21 to another acrocentric chromosome (e.g., 22 or 14). Because the fertilized ovum already possesses two normal autosomes 21, the translocated material provides the same triple gene dosage as in trisomy 21. Such cases are frequently (but not always) familial, and the translocated chromosome is inherited from one of the parents (usually the mother), who is a carrier of a robertsonian translocation, for example, a mother with karyotype 45,XX,der(14;21)(q10;q10). In cells with robertsonian translocations, the genetic material normally found on long arms of two pairs of chromosomes is distributed among only three chromosomes. This affects chromosome pairing during meiosis, and as a result the gametes have a high probability of being aneuploid.

Approximately 1% of patients with Down syndrome are mosaics, having a mixture of cells with 46 or 47 chromosomes. This mosaicism results from mitotic nondisjunction of chromosome 21 during an early stage of embryogenesis. Symptoms in such cases are variable and milder, depending on the proportion of abnormal cells. Clearly, in cases of translocation or mosaic Down syndrome, maternal age is of no importance.

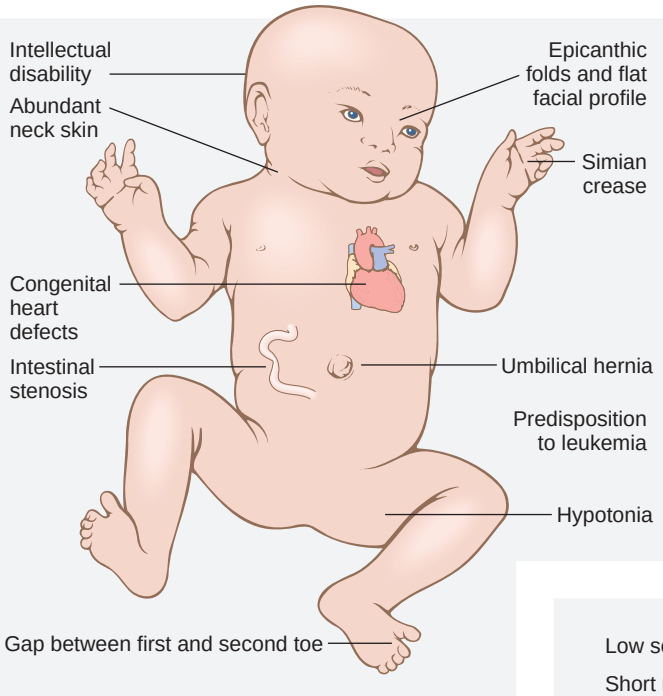
The diagnostic clinical features of this condition—flat facial profile, oblique palpebral fissures, and epicanthic folds (Fig. 5.20)**—are usually readily evident, even at birth.** Down syndrome is a leading cause of severe intellectual disability. It should be pointed out that some mosaics with Down syndrome have mild phenotypic changes and may even have normal or near-normal intelligence. In addition to the phenotypic abnormalities and the intellectual disability already noted, some other clinical features are worthy of note.

- *Approximately 40% of the patients have congenital heart disease.* The most frequent forms of congenital heart diseases in Down syndrome are atrioventricular septal defects constituting 43% of cases, whereas ventricular septal defects, atrial septal defects, and tetralogy of Fallot constitute 32%, 19%, and 6%, respectively. Cardiac problems are responsible for the majority of the deaths in infancy and early childhood. Several other congenital malformations, including atresias of the esophagus and small bowel, are also common.
- *Children with trisomy 21 have a high risk of developing leukemia;* there is 20-fold increased risk of developing acute B lymphoblastic leukemias and 500-fold increased risk of acute myeloid leukemias. The latter, most commonly, is acute megakaryoblastic leukemia.
- Virtually all patients with trisomy 21 older than age 40 develop *neuropathologic changes* characteristic of Alzheimer disease, a degenerative disorder of the brain.
- Patients with Down syndrome *have abnormal immune responses* that predispose them to serious infections, particularly of the lungs, and to thyroid autoimmunity. Although several abnormalities, affecting mainly T-cell functions, have been reported, the basis of immunologic disturbances is not clear.

Despite all these problems, improved medical care has increased the longevity of individuals with trisomy 21. Currently the median age at death is 47 years (up from 25 years in 1983).

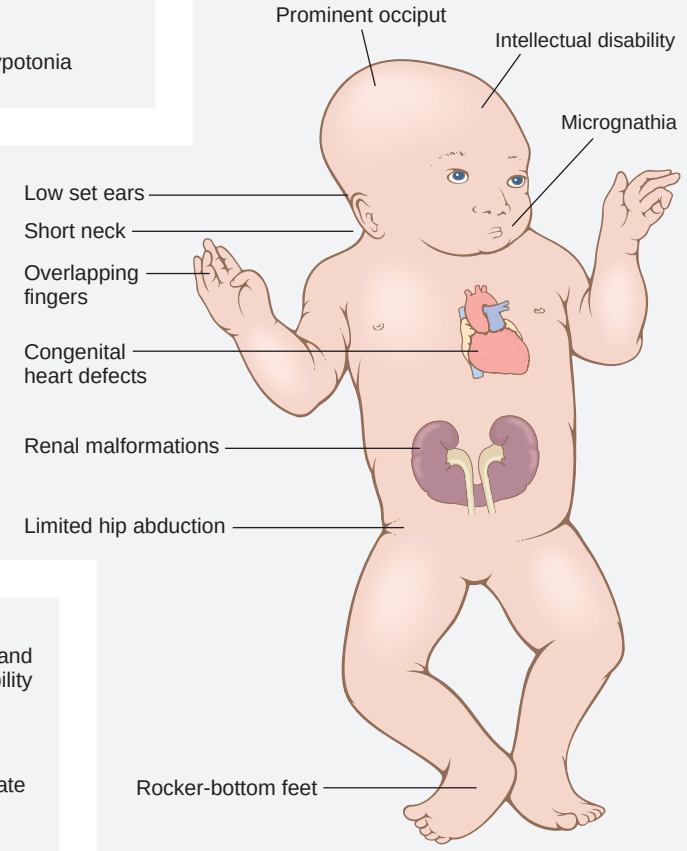
Although the long arm of chromosome 21 was fully sequenced in 2000, progress in unraveling the molecular basis of Down syndrome has been quite slow. This stems in part from the fact that Down syndrome results from gene dosage imbalance rather than the action of a few genes. A sampling of some of the observations follows.

- *Gene dosage*. The majority of the protein coding genes mapped to chromosome 21 are overexpressed. Included among these is the gene for amyloid-beta precursor protein (*APP*). As discussed in Chapter 28, aggregation of amyloid-beta proteins is the critical initiating event in the development of Alzheimer disease and could contribute to the early onset of Alzheimer disease that occurs in individuals with Down syndrome.
- *Mitochondrial dysfunction.* Approximately 10% of the genes overexpressed in Down syndrome are directly or indirectly involved in regulation of mitochondrial functions. In keeping with this, mitochondria are abnormal both morphologically and functionally in several tissues. For example, cristae are broken or swollen; there is evidence of mitochondrial stress with generation of reactive oxygen species and activation of apoptosis.
- *Noncoding RNAs*. Chromosome 21 has the highest density of lncRNAs whose target sequences are widely expressed



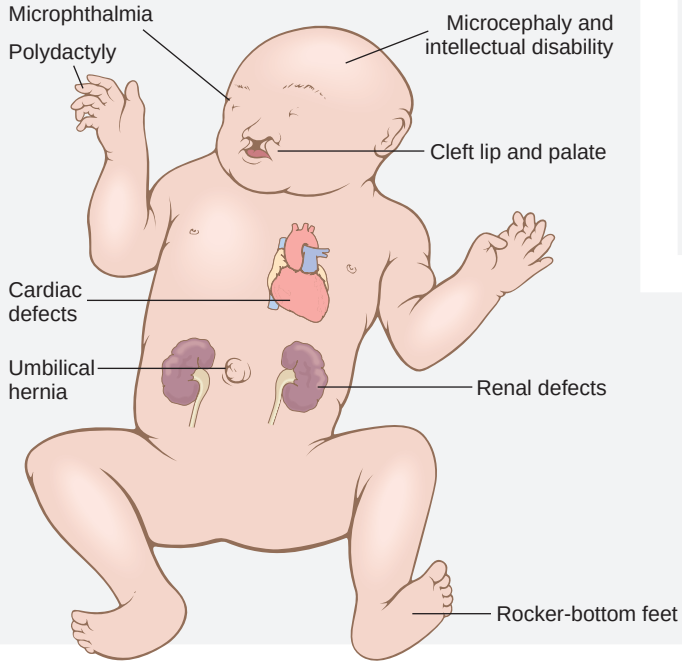
TRISOMY 21: DOWN SYNDROME

Incidence: 1 in 700 births
Karyotypes:
Trisomy 21 type: 47,XX, +21
Translocation type: 46,XX,der(14;21)(q10;q10),+21
Mosaic type: 46,XX/47,XX, +21



TRISOMY 18: EDWARDS SYNDROME

Incidence: 1 in 8000 births
Karyotypes:
Trisomy 18 type: 47,XX, +18
Mosaic type: 46,XX/47,XX, +18



TRISOMY 13: PATAU SYNDROME

Incidence: 1 in 15,000 births
Karyotypes:
Trisomy 13 type: 47,XX, +13
Translocation type: 46,XX,+13,der(13;14)(q10;q10)
Mosaic type: 46,XX/47,XX, +13

Figure 5.20 Clinical features and karyotypes of selected autosomal trisomies.

on other chromosomes, making the task of identification of genes whose products dictate the phenotype of trisomy 21 exceedingly difficult.

Much progress is being made in the molecular diagnosis of Down syndrome prenatally. Approximately 5% to 10% of the total cell free DNA in maternal blood is derived from the fetus and can be identified by polymorphic genetic markers. By using next-generation sequencing, the gene dosage of chromosome 21 linked genes in fetal DNA can be determined with great precision. This is a powerful noninvasive screening test for prenatal diagnosis of trisomy 21 as well as other trisomies. Most laboratories require confirmation of a positive screening test with conventional karyotyping.

Other Trisomies

A variety of other trisomies involving chromosomes 8, 9, 13, 18, and 22 have been described. Only trisomy 18 (Edwards syndrome) and trisomy 13 (Patau syndrome) are common enough to merit brief mention here. As noted in Fig. 5.20, they share several karyotypic and clinical features with trisomy 21. Thus most cases result from meiotic nondisjunction and therefore carry a complete extra copy of chromosome 13 or 18. As in Down syndrome, an association with increased maternal age is also noted. In contrast to trisomy 21, however, the malformations are much more severe and wide ranging. As a result, only rarely do infants survive beyond the first year of life. Most succumb within a few weeks to months.

Chromosome 22q11.2 Deletion Syndrome

Chromosome 22q11.2 deletion syndrome encompasses a spectrum of disorders that result from a small deletion of band q11.2 on the long arm of chromosome 22. The syndrome is fairly common, occurring in as many as 1 in 4000 births, but it is often missed because of variable clinical features. These include congenital heart defects, abnormalities of the palate, facial dysmorphism, developmental delay, and variable degrees of T-cell immunodeficiency and hypocalcemia. Previously, these clinical features were considered to represent two different disorders—*DiGeorge syndrome* and *velocardiofacial syndrome*. In a very small number of cases there is a deletion of 10p13-14.

Patients with DiGeorge syndrome have thymic hypoplasia, with resultant T-cell immunodeficiency (Chapter 6), parathyroid hypoplasia giving rise to hypocalcemia, a variety of cardiac malformations affecting the outflow tract, and mild facial anomalies. Atopic disorders (e.g., allergic rhinitis) and autoimmunity (e.g., thrombocytopenia) may also be seen. The clinical features of the so-called velocardiofacial syndrome include facial dysmorphism (prominent nose, retrognathia), cleft palate, cardiovascular anomalies, and learning disabilities. Less frequently, these patients also have immunodeficiency.

Despite the overlapping clinical features of these two conditions (e.g., cardiac malformations, facial dysmorphology) it was only after these two apparently unrelated syndromes were found to be associated with a similar cytogenetic abnormality that the clinical overlap came into focus. Recent studies indicate that, in addition to the numerous structural malformations, individuals with the 22q11.2 deletion syndrome are at a particularly high risk for psychotic illnesses, such as *schizophrenia and bipolar disorders.* In fact, it is estimated that schizophrenia develops in approximately 25% of adults with this syndrome. Conversely, deletions of the region can be found in 2% to 3% of individuals with childhood-onset schizophrenia. In addition, attention-deficit/hyperactivity disorder is seen in 30% to 35% of affected children.

The diagnosis of this condition may be suspected on clinical grounds but can be established only by detection of the deletion by FISH (Fig. 5.21). By this test, approximately 90% of individuals previously diagnosed as having DiGeorge syndrome and 80% of those with the velocardiofacial syndrome have a deletion of 22q11.2. Thirty percent of individuals with conotruncal cardiac defects but no other features of this syndrome also reveal deletions of the same chromosomal region.

The molecular basis of this syndrome is not fully understood. The deleted region is large (approximately 1.5 Mb) and includes 30 to 40 genes. The clinical heterogeneity, with predominant immunodeficiency in some cases (DiGeorge syndrome) and predominant dysmorphology and cardiac malformations in other cases, probably reflects the variable position and size of the deleted segment from this genetic region. Approximately 30 candidate genes have been mapped to the deleted region. Among these, *TBX1,* a T-box transcription factor, is most closely associated with the phenotypic features of this syndrome. This gene is expressed in the pharyngeal mesenchyme and endodermal pouch from which facial structures, thymus, and parathyroid are derived. The targets of *TBX1* include *PAX9,* a gene that controls the development of the palate, parathyroids, and thymus. Clearly there are other genes that contribute to the behavioral and psychiatric disorders that remain to be identified.

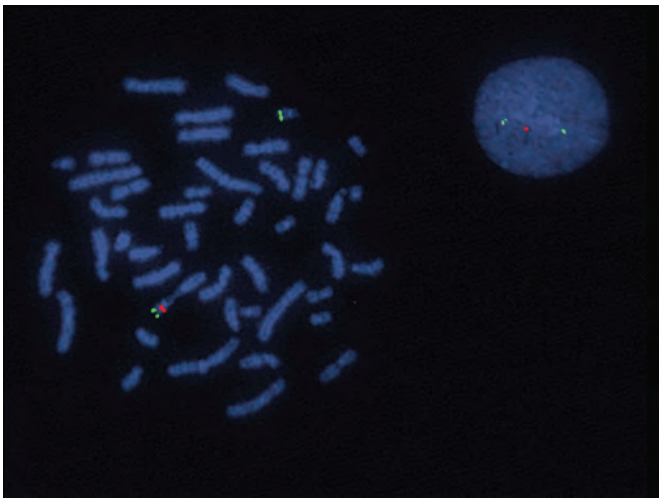


Figure 5.21 Fluorescence in situ hybridization of both metaphase chromosomes and an interphase cell from a patient with DiGeorge syndrome demonstrating the deletion of a probe that maps to chromosome 22q11.2. The 22q11.2 probe is in red, and the control probe, localized to 22q, is in green. The metaphase spread shows one chromosome 22 with both a green signal (control probe) and a red signal (from the 22q11.2 probe). The other chromosome 22 shows only hybridization with the control probe (green), but no red 22q11.2 signal since there is a deletion on this chromosome. The interphase cell also shows a hybridization pattern consistent with a deletion of chromosome 22q11.2. (Courtesy Dr. Stuart Schwartz, Department of Pathology, University of Chicago, Chicago, Ill.)

KEY CONCEPTS

CYTOGENETIC DISORDERS INVOLVING AUTOSOMES

- Down syndrome is associated with an extra copy of genes on chromosome 21, most commonly due to trisomy 21 and less frequently from translocation of extra chromosomal material from chromosome 21 to other chromosomes or from mosaicism.
- Patients with Down syndrome have severe intellectual disability, flat facial profile, epicanthic folds, cardiac malformations, higher risk of leukemia and infections, and premature development of Alzheimer disease.
- Deletion of genes at chromosomal locus 22q11.2 gives rise to malformations affecting the face, heart, thymus, and parathyroids. The resulting disorders are recognized as DiGeorge syndrome (thymic hypoplasia with diminished T-cell immunity and parathyroid hypoplasia with hypocalcemia) and velocardiofacial syndrome (congenital heart disease involving outflow tracts, facial dysmorphism, and developmental delay).

Cytogenetic Disorders Involving Sex Chromosomes

Genetic diseases associated with changes involving the sex chromosomes are far more common than those related to autosomal aberrations. Furthermore, imbalances (excess or loss) of sex chromosomes are much better tolerated than are similar imbalances of autosomes. In large part, this latitude relates to two factors that are peculiar to the sex chromosomes: (1) lyonization or inactivation of all but one X chromosome and (2) the modest amount of genetic material carried by the Y chromosome. These features are discussed briefly in relation to sex chromosomal disorders.

In 1961, Mary Lyon outlined the idea of X-inactivation, now commonly known as the Lyon hypothesis. It states that (1) *only one of the X chromosomes is genetically active,* (2) *the other X chromosome of either maternal or paternal origin undergoes heteropyknosis and is rendered inactive,* (3) *inactivation of either the maternal or the paternal X chromosome occurs at random among all the cells of the blastocyst on or about day 5.5 of embryonic life,* and (4) *inactivation of the same X chromosome persists in all the cells derived from each precursor cell.* Thus, the great preponderance of normal females are in reality mosaics and have two populations of cells, one with an inactivated maternal X chromosome and the other with an inactivated paternal X chromosome. Herein lies the explanation of why females have the same dosage of X-linked active genes as males. The inactive X chromosome can be seen in the interphase nucleus as a darkly staining small mass in contact with the nuclear membrane known as the *Barr body,* or X chromatin. The molecular basis of X inactivation involves a unique gene called *XIST,* whose product is a lncRNA (Chapter 1) that is retained in the nucleus, where it "coats" the X chromosome that it is transcribed from and initiates a gene-silencing process by chromatin modification and DNA methylation. The *XIST* allele is switched off in the active X chromosome.

Although it was initially thought that all the genes on the inactive X chromosome are "switched off," it is now established that many genes escape X inactivation. Molecular studies suggest that 30% of genes on Xp and a smaller number (3%) on Xq escape X inactivation. At least some of the genes that are expressed from both X chromosomes are important for normal growth and development. This notion is supported by the fact that patients with monosomy of the X chromosome (Turner syndrome: 45,X) have severe somatic and gonadal abnormalities. If a single dose of X-linked genes were sufficient, no detrimental effect would be expected in such cases. Furthermore, although one X chromosome is inactivated in all cells during embryogenesis, it is selectively reactivated in oogonia before the first meiotic division. Thus, it seems that both X chromosomes are required for normal growth as well as oogenesis. The tips of short and long arms of X and Y chromosomes have regions of homology that recombine during meiosis and are therefore inherited as autosomal loci. For this reason they are called pseudoautosomal regions. These genes also escape X inactivation. These mechanisms ensure that males and females have equivalent doses of genes that map on X and Y chromosomes.

With respect to the Y chromosome, it is well known that this chromosome is both necessary and sufficient for male development. **Regardless of the number of X chromosomes, the presence of a single Y chromosome determines the male sex.** The gene that dictates testicular development (*SRY* [sex-determining region Y gene]) is located on its distal short arm. For quite some time this was considered to be the only gene of significance on the Y chromosome. Recent studies, however, have yielded a rich harvest of several gene families in the so-called male-specific Y region, or MSY region, harboring 75 protein coding genes. All of these are believed to be testis-specific and are involved in spermatogenesis. In keeping with this, all Y chromosome deletions are associated with azoospermia. By comparison, the X chromosome has 840 coding genes. The following features are common to all sex chromosome disorders.

- In general, sex chromosome disorders cause subtle, chronic problems relating to sexual development and fertility.
- Sex chromosome disorders are often difficult to diagnose at birth, and many are first recognized at the time of puberty.
- In general, the greater the number of X chromosomes, in both males and females, the greater the likelihood of intellectual disability.

The two most important disorders arising in aberrations of sex chromosomes are described briefly here.

Klinefelter Syndrome

Klinefelter syndrome is best defined as male hypogonadism that occurs when there are two or more X chromosomes and one or more Y chromosomes. It is one of the most frequent forms of genetic diseases involving the sex chromosomes as well as one of the most common causes of hypogonadism in males. The incidence of this condition is reported to be approximately 1 in 660 live male births. This is an underestimate since Klinefelter syndrome has a range of phenotypic manifestations, and those with mild features are never seen by health care providers. The clinical features of Klinefelter syndrome can be attributed to two major factors: (1) aneuploidy and the impact of increased gene

dosage by the supernumerary X and (2) the presence of hypogonadism. Klinefelter syndrome is rarely diagnosed before puberty, particularly because manifestations of hypogonadism do not develop before early puberty.

Most patients have a distinctive body habitus with an increase in length between the soles and the pubic bone, which creates the appearance of an elongated body. Also characteristic are eunuchoid body habitus with abnormally long legs; small atrophic testes often associated with a small penis; and lack of such secondary male characteristics as deep voice, beard, and male distribution of pubic hair. Gynecomastia may be present. The cognitive abilities range from average to below average with modest deficit in verbal skills particularly those that are used in reading and language comprehension. Patients with Klinefelter syndrome develop several comorbid conditions. There is increased incidence of type 2 diabetes and the metabolic syndrome that gives rise to insulin resistance. Patients are at a higher risk for congenital heart disease, particularly mitral valve prolapse, which is seen in about 50% of adults. In addition there is a higher prevalence of atrial and ventricular septal defects. There is also an increased incidence of osteoporosis and fractures due to sex hormonal imbalance. Patients with Klinefelter syndrome have a 20- to 30-fold higher risk of developing extragonadal germ cell tumors, mostly mediastinal teratomas. In addition, breast cancer and autoimmune diseases such as systemic lupus erythematosus occur more frequently. It should be noted that the physical attributes described here are quite variable, the only consistent finding being hypogonadism.

Klinefelter syndrome is an important genetic cause of reduced spermatogenesis and male infertility. In some patients the testicular tubules are totally atrophied and replaced by pink, hyaline, collagenous ghosts. In others, apparently normal tubules are interspersed with atrophic tubules. In some patients, all tubules are primitive and appear embryonic, consisting of cords of cells that never developed a lumen or progressed to mature spermatogenesis. Leydig cells appear prominent, as a result of the atrophy and crowding of the tubules and elevation of gonadotropin concentrations. Plasma gonadotropin concentrations, particularly follicle-stimulating hormone (FSH), are consistently elevated, whereas testosterone levels are variably reduced. Mean plasma estradiol levels are elevated by an as yet unknown mechanism. The ratio of estrogens and testosterone determines the degree of feminization in individual cases.

Classic Klinefelter syndrome is associated with a 47,XXY karyotype (90% of cases). This complement of chromosomes results from nondisjunction during the meiotic divisions in the germ cells of one of the parents. Maternal and paternal nondisjunction at the first meiotic division is roughly equally involved. There is no phenotypic difference between those who receive the extra X chromosome from their father and those who receive it from their mother. Advanced maternal age (>40 years) is a risk factor. In addition to this classic karyotype, approximately 15% of patients with Klinefelter syndrome have a variety of mosaic patterns, most of them being 46,XY/47,XXY; in some there is a cell line with structurally abnormal X chromosome (e.g., 47,iXq,Y). As is the case with normal females, all but one X chromosome undergoes inactivation in patients with Klinefelter syndrome. Why then, do the patients with this disorder have hypogonadism and associated features? The explanation for this lies in genes on the X chromosome that escape lyonization and in the pattern of X inactivation.

- One pathogenic mechanism is related to uneven dosage compensation during X inactivation. It is now known that about 35% of the X-linked genes escape inactivation. Thus there is an extra dose of these genes compared with normal males in whom only one copy of the X chromosome is active; it appears that "overexpression" of one or more of these genes leads to hypogonadism. A similar mechanism may also dictate some somatic features. The *Short-stature HomeobOX (SHOX)* gene that maps on the pseudoautosomal region of Xp is one of the genes that is not subject to X inactivation. An extra copy of this growth-related gene is probably responsible for the tall stature and long legs typical of Klinefelter syndrome. It should be noted that most genes whose expression is upregulated in Klinefelter syndrome lie outside the X chromosome. This implies that the supernumerary X chromosome can regulate gene expression on autosomes.
- A second mechanism involves the gene encoding the androgen receptor, through which testosterone mediates its effects. The androgen receptor gene maps to the X chromosome and contains highly polymorphic CAG (trinucleotide) repeats. The functional response of the receptor to any particular dose of androgen is dictated, in part, by the number of CAG repeats, as receptors with shorter CAG repeats are more sensitive to androgens than those with long CAG repeats. In persons with Klinefelter syndrome, the X chromosome bearing the androgen receptor allele with the shortest CAG repeat is preferentially inactivated. In XXY males with low testosterone levels, expression of androgen receptors with long CAG repeats exacerbates the hypogonadism and appears to account for certain aspects of the phenotype, such as small penis size.

Turner Syndrome

Turner syndrome results from complete or partial monosomy of the X chromosome and is characterized by hypogonadism in phenotypic females. It is the most common sex chromosome abnormality in females, affecting about 1 in 2000 live-born females.

With routine cytogenetic methods, three types of karyotypic abnormalities are seen in individuals with Turner syndrome.

- *Approximately 57% are missing an entire X chromosome, resulting in a 45,X karyotype.* Of the remaining 43%, approximately one-third (14%) have structural abnormalities of the X chromosomes, and two-thirds (29%) are mosaics.
- *The common feature of the structural abnormalities is to produce partial monosomy of the X chromosome.* In order of frequency, the structural abnormalities of the X chromosome include (1) an isochromosome of the long arm, 46,X,i(X)(q10), resulting in the loss of the short arm; (2) deletion of portions of both long and short arms, resulting in the formation of a ring chromosome, 46,X,r(X); and (3) deletion of portions of the short or long arm, 46X,del(Xq) or 46X,del(Xp).
- *The mosaic patients have a 45,X cell population along with one or more karyotypically normal or abnormal cell types.*

Examples of karyotypes that mosaic Turner females may have are the following: (1) 45,X/46,XX, (2) 45,X/46,XY, (3) 45,X/47,XXX, or (4) 45,X/46,X,i(X)(q10). Studies suggest that the prevalence of mosaicism in Turner syndrome may be much higher than the 30% detected by conventional cytogenetic studies. With the use of more sensitive techniques, the prevalence of mosaic Turner syndrome increases to 75%. Because 99% of conceptuses with an apparent 45,X karyotype are nonviable, many authorities believe that there are no truly nonmosaic Turner syndrome patients. While this issue remains controversial, it is important to appreciate the karyotypic heterogeneity associated with Turner syndrome because it is responsible for significant variations in phenotype. In patients in whom the proportion of 45,X cells is high, the phenotypic changes are more severe than in those who have readily detectable mosaicism. The latter may have an almost normal appearance and may present only with primary amenorrhea. A very small number of patients are able to conceive.

Five percent to 10% of patients with Turner syndrome have Y chromosome sequences either as a complete Y chromosome (e.g., 45,X/46,XY karyotype) or as fragments of Y chromosomes translocated on other chromosomes. These patients are at a higher risk for development of a gonadal tumor (gonadoblastoma).

The most severely affected patients generally present during infancy with edema of the dorsum of the hand and foot due to lymph stasis and sometimes *swelling of the nape of the neck*. The latter is related to markedly distended lymphatic channels, producing a so-called cystic hygroma (Chapter 10). As these infants develop, the swellings subside but often leave bilateral *neck webbing* and persistent looseness of skin on the back of the neck. *Congenital heart disease* is also common, affecting 25% to 50% of patients. Left-sided cardiovascular abnormalities, particularly preductal coarctation of the aorta and bicuspid aortic valve, are seen most frequently. Approximately 5% of young women initially diagnosed with coarctation of aorta have Turner syndrome. Aortic root dilatation is present in 30% of cases, and there is a 100-fold higher risk of aortic dissection. Cardiovascular abnormalities are the most important cause of increased mortality in children with Turner syndrome.

The principal clinical features in adolescents and adults are illustrated in Fig. 5.22. At puberty there is failure to develop normal secondary sex characteristics. The genitalia remain infantile, breast development is inadequate, and there is little pubic hair. The mental status of patients is usually normal, but subtle defects in nonverbal, visual-spatial information processing have been noted. Of particular importance in establishing the diagnosis in an adult is the shortness of stature (rarely exceeding 150 cm in height) and amenorrhea. Turner syndrome is the single most important cause of primary

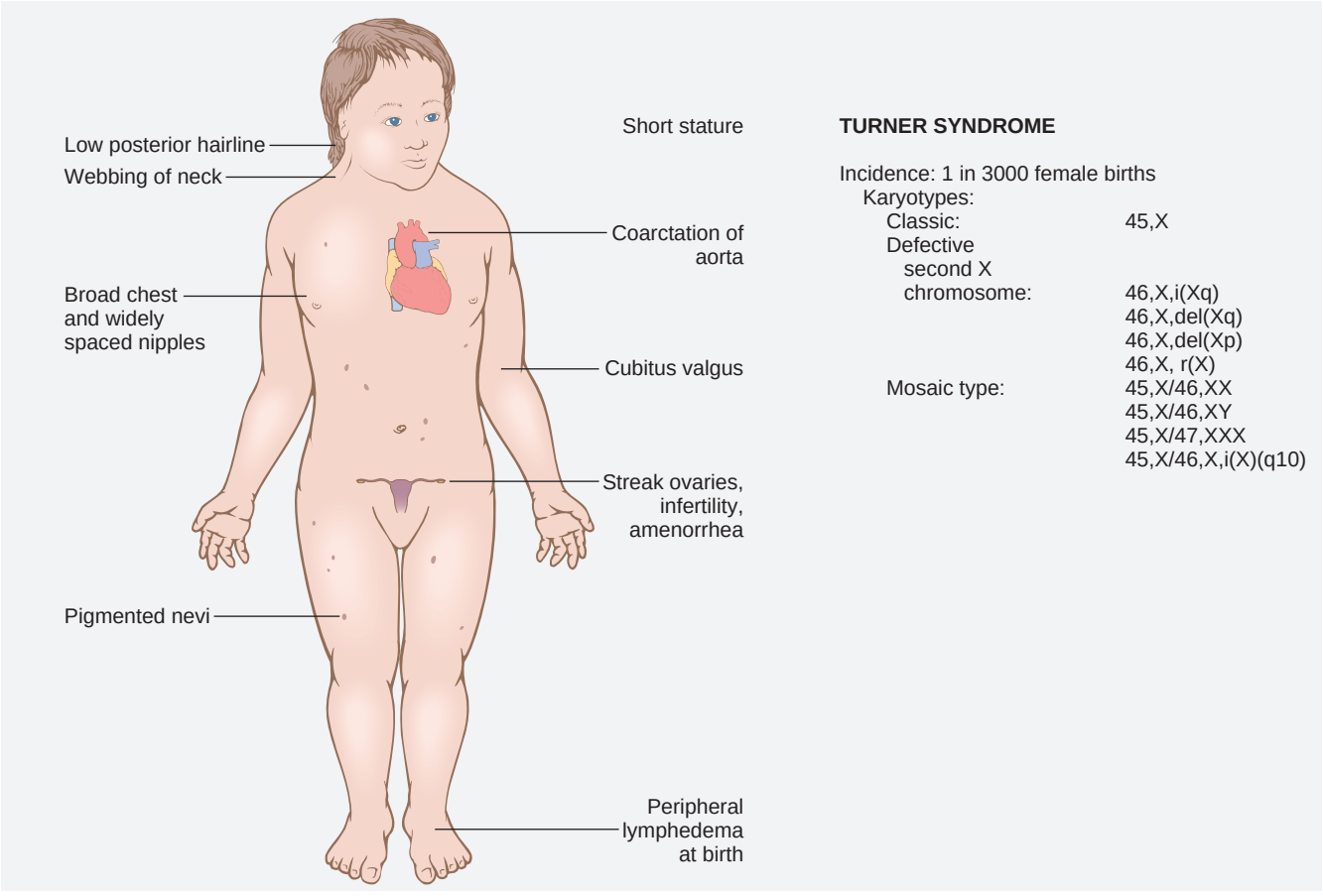


Figure 5.22 Clinical features and karyotypes of Turner syndrome.

amenorrhea, accounting for approximately one-third of the cases. For reasons that are not clear, approximately 50% of patients develop autoantibodies that react with the thyroid gland, and up to half of these develop clinically manifest hypothyroidism. Equally mysterious is the presence of glucose intolerance, obesity, nonalcoholic fatty liver disease, and insulin resistance in a subset of patients. Some have full-fledged metabolic syndrome. The occurrence of insulin resistance is significant because therapy with growth hormone, commonly used in these patients, worsens insulin resistance.

The molecular pathogenesis of Turner syndrome is not completely understood, but studies have begun to shed some light. In approximately 80% of cases the X chromosome is maternal in origin, suggesting that there is an abnormality in paternal gametogenesis. As mentioned earlier, both X chromosomes are active during oogenesis and are essential for normal development of the ovaries. During normal fetal development, ovaries contain as many as 7 million oocytes. The oocytes gradually disappear so that by menarche their numbers have dwindled to a mere 400,000, and when menopause occurs fewer than 10,000 remain. In Turner syndrome, fetal ovaries develop normally early during the first 18 weeks of gestation, but the absence of the second X chromosome leads to an accelerated loss of oocytes, which is complete by age 2 years. In a sense, therefore, "menopause occurs before menarche," and the ovaries are reduced to atrophic fibrous strands, devoid of ova and follicles *(streak ovaries)*. Studies of patients with Turner syndrome with deletions affecting the short or long arm have revealed that many of the somatic features are determined by genes on the short arm, whereas genes on the long arm affect fertility and menstruation. Among the genes involved in the Turner phenotype is the *SHOX* gene at Xp22.33. As mentioned earlier, *SHOX* maps at the pseudo autosomal region of the X and Y chromosomes and escapes X inactivation. Thus both normal males and normal females have two copies of this gene. Haploinsufficiency of *SHOX* in Turner syndrome is believed to give rise to short stature. Indeed, deletions of the *SHOX* gene are noted in 2% to 5% of otherwise normal children with short stature. In keeping with its role as a critical regulator of growth, *SHOX* is expressed during fetal life in the growth plates of several long bones including the radius, ulna, tibia, and fibula. It is also expressed in the first and second pharyngeal arches. Just as the loss of *SHOX* is always associated with short stature, excess copies of this gene (in Klinefelter syndrome) are associated with tall stature. Whereas haploinsufficiency of *SHOX* can explain growth deficit in Turner syndrome, it cannot explain other clinical features such as cardiac malformations and metabolic abnormalities. Clearly several other genes located on the short arm of X chromosome are involved. Growth hormone and estradiol are used to treat Turner syndrome with a reasonable degree of success.

Hermaphroditism and Pseudohermaphroditism

The problem of sexual ambiguity is exceedingly complex, and only limited observations are possible here; for more details, the reader should refer to specialized sources. It will be no surprise to medical students that the sex of an individual can be defined on several levels. *Genetic sex* is determined by the presence or absence of a Y chromosome. No matter how many X chromosomes are present, a single Y chromosome dictates testicular development and the genetic male sex. The initially indifferent gonads of both male and female embryos have an inherent tendency to feminize, unless influenced by Y chromosome–dependent masculinizing factors. *Gonadal sex* is based on the histologic characteristics of the gonads. *Ductal sex* depends on the presence of derivatives of the müllerian or wolffian ducts. *Phenotypic,* or *genital, sex* is based on the appearance of the external genitalia. Sexual ambiguity is present whenever there is disagreement among these various criteria for determining sex.

The term *true hermaphrodite* implies the presence of both ovarian and testicular tissue. In contrast, a pseudohermaphrodite represents a disagreement between the phenotypic and gonadal sex (i.e., a female pseudohermaphrodite has ovaries but male external genitalia; a male pseudohermaphrodite has testicular tissue but female-type genitalia). The genetic bases of these conditions are quite variable and beyond the scope of our discussion here.

KEY CONCEPTS

CYTOGENETIC DISORDERS INVOLVING SEX CHROMOSOMES

- In females, one X chromosome, maternal or paternal, is randomly inactivated during development (Lyon hypothesis).
- In Klinefelter syndrome, there are two or more X chromosomes with one Y chromosome as a result of nondisjunction of sex chromosomes. Patients have testicular atrophy, sterility, reduced body hair, gynecomastia, and eunuchoid body habitus. It is the most common cause of male sterility.
- In Turner syndrome, there is partial or complete monosomy of genes on the short arm of the X chromosome, most commonly due to absence of one X chromosome (45,X), mosaicism, or deletions involving the short arm of the X chromosome. Short stature, webbing of the neck, cubitus valgus, cardiovascular malformations, amenorrhea, lack of secondary sex characteristics, and fibrotic ovaries are typical clinical features.

SINGLE-GENE DISORDERS WITH NONCLASSIC INHERITANCE

It has become increasingly evident that transmission of certain single-gene disorders does not follow classic Mendelian principles. This group of disorders can be classified into four categories.

- Diseases caused by trinucleotide-repeat mutations
- Disorders caused by mutations in mitochondrial genes
- Disorders associated with genomic imprinting
- Disorders associated with gonadal mosaicism

Clinical and molecular features of some single-gene diseases that exemplify nonclassic patterns of inheritance are described next.

Diseases Caused by Trinucleotide-Repeat Mutations

Expansion of trinucleotide repeats is an important genetic cause of human disease, particularly neurodegenerative

Table 5.8 Examples of Trinucleotide-Repeat Disorders

Disease	Gene	Locus	Protein	Repeat	No. Repeats	
					Normal	Disease
Expansions Affecting Noncoding Regions						
Fragile X syndrome	*FMRI (FRAXA)*	Xq27.3	FMR-I protein (FMRP)	CGG	6–55	55–200 (pre); >230 (full)
Friedreich ataxia	*FXN*	9q21.1	Frataxin	GAA	7–34	34–80 (pre); >100 (full)
Myotonic dystrophy	*DMPK*	19q13.3	Myotonic dystrophy protein kinase (DMPK)	CTG	5–37	34–80 (pre); >100 (full)
Expansions Affecting Coding Regions						
Spinobulbar muscular atrophy (Kennedy disease)	*AR*	Xq12	Androgen receptor (AR)	CAG	5–34	37–70
Huntington disease	*HTT*	4p16.3	Huntingtin	CAG	6–35	39–250
Dentatorubral-pallidoluysian atrophy (Haw River syndrome)	*ATNL*	12p13.31	Atrophin-I	CAG	7–35	49–88
Spinocerebellar ataxia type I	*ATXN1*	6p23	Ataxin-I	CAG	6–44	>39
Spinocerebellar ataxia type 2	*ATXN2*	12q24.1	Ataxin-2	CAG	13–33	>31
Spinocerebellar ataxia type 3 (Machado-Joseph disease)	*ATXN3*	14q21	Ataxin-3	CAG	12–40	55–84
Spinocerebellar ataxia type 6	*Ataxin-6*	19p13.3	α_{1A}-Voltage-dependent calcium channel subunit	CAG	4–18	21–33
Spinocerebellar ataxia type 7	*Ataxin-7*	3p14.1	Ataxin-7	CAG	4–35	37–306

disorders. The discovery in 1991 of expanding trinucleotide repeats as a cause of fragile X syndrome (FXS) was a landmark in human genetics. Since then the origins of about 40 human diseases (Table 5.8) have been traced to unstable nucleotide repeats, and the number continues to grow. All the disorders discovered so far are associated with neurodegenerative changes. Some general principles apply to these diseases.

- The causative mutations are associated with the expansion of a stretch of trinucleotides that usually share the nucleotides G and C. In all cases the DNA is unstable, and an expansion of the repeats above a certain threshold impairs gene function in various ways, discussed later. In recent years, diseases associated with unstable tetranucleotides, pentanucleotides, and hexanucleotides have also been found, establishing this as a fundamental mechanism of neuromuscular diseases.
- The proclivity to expand depends strongly on the sex of the transmitting parent. In FXS, expansions occur during oogenesis, whereas in Huntington disease they occur during spermatogenesis.
- There are three key mechanisms by which unstable repeats cause diseases: (1) *Loss of function* of the affected gene, typically by transcription silencing, as in FXS; in such cases the repeats are generally in the noncoding part of the gene. (2) *A toxic gain of function* by alterations of protein structure, as in Huntington disease and spinocerebellar ataxias; in such cases the expansions occur in the coding regions of the genes. (3) *A toxic gain of function mediated by RNA* as is seen in fragile X–associated tremor/ataxia syndrome; as in FXS, the noncoding parts of the gene are affected (Fig. 5.23).

The pathogenetic mechanisms underlying disorders caused by mutations that affect coding regions seem to be distinct from those in which the expansions affect noncoding regions. The former usually involve CAG repeats coding for polyglutamine tracts in the corresponding proteins. Such polyglutamine diseases are characterized by progressive neurodegeneration, typically striking in midlife. Polyglutamine expansions lead to toxic gain of function, whereby the abnormal protein may interfere with the function of the normal protein (a dominant negative activity) or acquire a novel pathophysiologic toxic activity. The precise mechanisms by which expanded polyglutamine proteins cause disease are not fully understood. In most cases the proteins are misfolded and tend to aggregate; the aggregates may suppress transcription of other genes, cause mitochondrial dysfunction, or trigger the unfolded-protein stress response and apoptosis (Chapters 1 and 2). *A morphologic hallmark of these diseases is the accumulation of aggregated mutant proteins in large intranuclear inclusions.* While formation of aggregates is common to many polyglutamine diseases, evidence of a direct toxic role of aggregates is not universal. In fact, some observers believe that aggregation may be protective by sequestration of the misfolded protein. Other models of pathogenicity implicate downstream effects mediated by proteolytic fragments of the polyglutamine fragment. Much more needs to be learned before therapeutic strategies can be developed.

Fragile X Syndrome (FXS)

FXS is the most common genetic cause of intellectual disability in males and overall the second most common cause after Down syndrome. It results from a trinucleotide expansion mutation in the *familial mental retardation 1 (FMR1)* gene. Although discovered initially as the cause of FXS, expansion mutations affecting the *FMR1* gene are now known to be present in two other well-defined disorders—fragile X–associated tremor/ataxia syndrome and fragile X–associated primary ovarian insufficiency. We begin our discussion of these disorders with consideration of FXS.

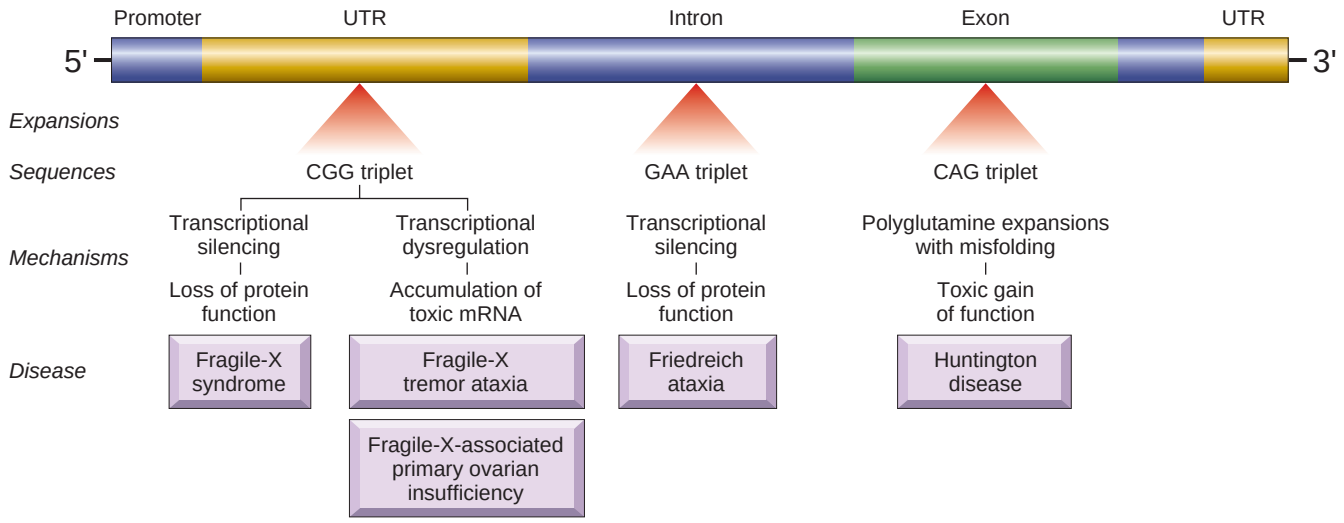


Figure 5.23 Sites of expansion and the affected sequence in selected diseases caused by nucleotide-repeat mutations. *UTR,* Untranslated region.

FXS has a frequency of 1 in 1550 for affected males and 1 in 8000 for affected females. Its name derives from an inducible cytogenetic abnormality in the X chromosome within which the *FMR1* gene maps. The cytogenetic alteration was discovered as a discontinuity of staining or as a constriction in the long arm of the X chromosome when cells are cultured in a folate-deficient medium. Because it appears that the chromosome is "broken" at this locale, it was named as a *fragile site* (Fig. 5.24). This method of detection has now been supplanted by DNA-based analysis of triplet repeat size, as discussed later.

Males with FXS have marked *intellectual disability.* They have a characteristic physical phenotype that includes a long face with a large mandible, large everted ears, and large testicles (macro-orchidism). Hyperextensible joints, a high arched palate, and mitral valve prolapse noted in some patients mimic a connective tissue disorder. These and other physical abnormalities described in this condition, however, are not always present and, in some cases, are quite subtle. The most distinctive feature is *macro-orchidism,* which is observed in at least 90% of affected postpubertal males.

In addition to intellectual disability, several neurologic and neuropsychiatric manifestations have been recognized in patients with FXS. These include epilepsy in 30% of cases, aggressive behavior in 90% of cases, autism spectrum disorder (includes several conditions such as autism and Asperger syndrome), and anxiety disorder/hyperactivity disorder. The latter two affect 50% to 75% of males with FXS. Two percent to 5% of patients diagnosed first with nonsyndromic autism have a mutation in the *FMR1* gene.

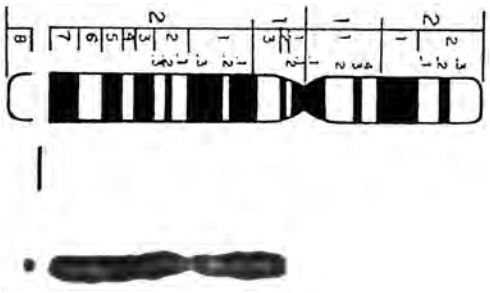


Figure 5.24 Fragile X seen as discontinuity of staining. (Courtesy Dr. Patricia Howard-Peebles, University of Texas Southwestern Medical Center, Dallas, Tex.)

As with other X-linked diseases, FXS affects males predominantly. Analysis of several pedigrees, however, reveals some patterns of transmission not typically associated with other X-linked recessive disorders (Fig. 5.25).

- *Carrier males:* Approximately 20% of males who, by pedigree analysis and by molecular tests, are known to carry a fragile X mutation are clinically normal. Because carrier males transmit the trait through all their phenotypically normal daughters to affected grandchildren, they are called *normal transmitting males.*
- *Affected females:* Thirty percent to 50% of carrier females are affected (i.e., have intellectual disability as well as other features described here), a number much higher than that in other X-linked recessive disorders.
- *Risk of phenotypic effects:* Risk depends on the position of the individual in the pedigree. For example, brothers of transmitting males are at a 9% risk of having intellectual disability, whereas grandsons of transmitting males incur a 40% risk.
- *Anticipation:* This refers to the observation that clinical features of FXS worsen with each successive generation, as if the mutation becomes increasingly deleterious as it is transmitted from a man to his grandsons and great-grandsons (through daughters).

The first breakthrough in resolving these perplexing observations came when linkage studies localized the mutation responsible for this disease to Xq27.3, within the cytogenetically abnormal region. At this locale lies the *FMR1* gene, characterized by multiple tandem repeats of the nucleotide sequence CGG in its 5′ untranslated region. In the normal population, the number of CGG repeats is small, ranging from 6 to 55 (average, 29). The presence and severity of clinical symptoms is related to the amplification of the CGG repeats. Thus, normal transmitting males and carrier females carry 55 to 200 CGG repeats. Expansions of this size are called *premutations.* In contrast, patients with FXS

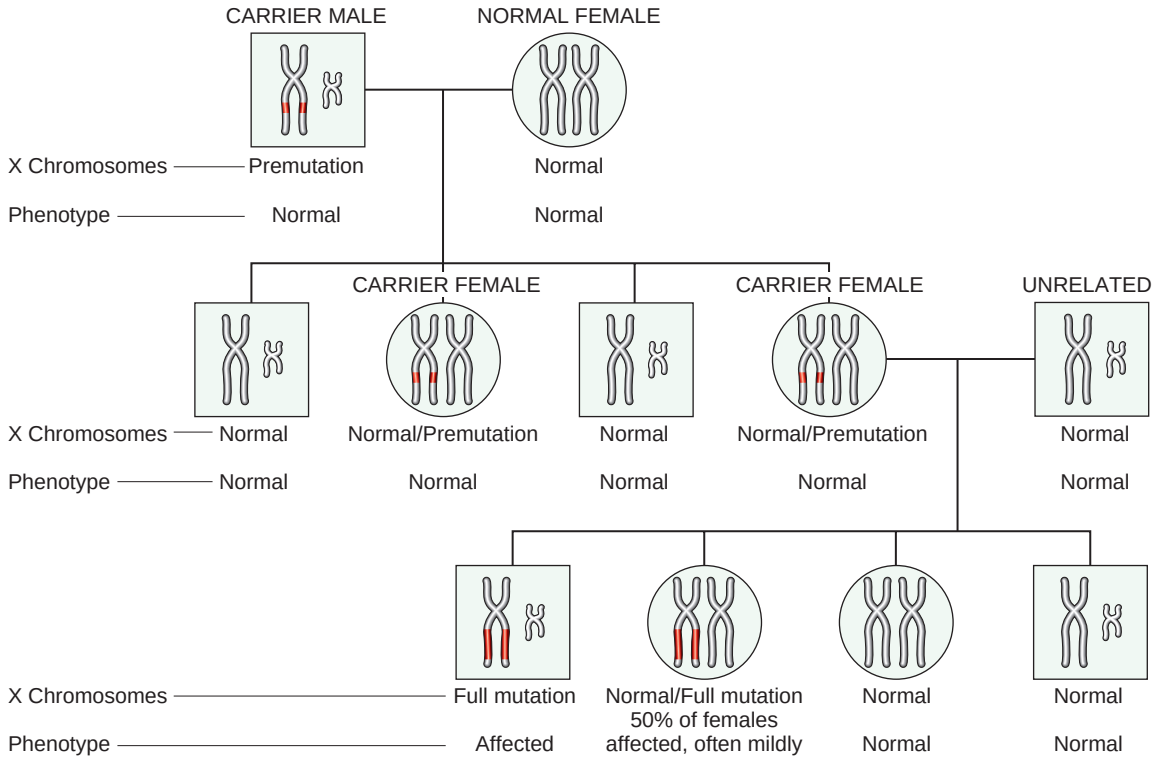


Figure 5.25 Fragile X pedigree. Note that in the first generation all sons are normal and all females are carriers. During oogenesis in the carrier female, premutation expands to full mutation; hence, in the next generation all males who inherit the X with full mutation are affected. However, only 50% of females who inherit the full mutation are affected and only mildly. Not shown are fragile X–associated ataxia/tremor and fragile X–associated primary ovarian failure that can occur in permutation careers. (Courtesy Dr. Nancy Schneider, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.)

have an extremely large expansion of the repeat region (200 to 4000 repeats, or *full mutations*). Full mutations are believed to arise by further amplification of the CGG repeats seen in premutations. How this process takes place is quite peculiar. Carrier males transmit the repeats to their progeny with small changes in repeat number. When the premutation is passed on by a carrier female, however, there is a high probability of a dramatic amplification of the CGG repeats, leading to intellectual disability in most male offspring and 50% of female offspring. Thus *it seems that during the process of oogenesis, but not spermatogenesis, premutations can be converted to mutations by triplet-repeat amplification*. This explains the unusual inheritance pattern; that is, the likelihood of intellectual disability is much higher in grandsons than in brothers of transmitting males because grandsons incur the risk of inheriting a premutation from their grandfather that is amplified to a full mutation in their mothers' ova. By comparison, brothers of transmitting males, being higher up in the pedigree, are less likely to have a full mutation. These molecular details also provide a satisfactory explanation of anticipation—a phenomenon that remained unexplained until triplet-repeat mutations were identified. Why only 50% of the females with the full mutation are clinically affected is not clear. Presumably in those who are clinically affected, there is unfavorable lyonization (i.e., there is a higher frequency of cells in which the X chromosome carrying the mutation is active).

The molecular basis of intellectual disability and other somatic changes is related to loss of function of the fragile X mental retardation protein (FMRP), the product of *FMR1* gene. As mentioned earlier, the normal *FMR1* gene contains up to 55 CGG repeats in its 5′ untranslated region. When the trinucleotide repeats in the *FMR1* gene exceed approximately 230, the DNA of the entire 5′ region of the gene becomes abnormally methylated. Methylation also extends upstream into the promoter region of the gene, resulting in transcriptional suppression of *FMR1*. The resulting absence of FMRP is believed to cause the phenotypic changes.

FMRP is a widely expressed cytoplasmic protein, most abundant in the brain and testis, the two organs most affected in this disease. Its proposed functions in the brain are the following:

- *FMRP selectively binds mRNAs associated with polysomes and regulates their intracellular transport to dendrites.* Unlike other cells, in neurons, protein synthesis occurs both in the perinuclear cytoplasm and in dendritic spines. Newly made FMRP translocates to the nucleus, where it assembles into a complex containing mRNA transcripts that encode presynaptic and postsynaptic proteins. The FRMP-mRNA complexes are then exported to the cytoplasm, from where they are trafficked to dendrites near neuronal synapses (Fig. 5.26).
- *FRMP is a translation regulator.* At synaptic junctions, FMRP suppresses protein synthesis from the bound mRNAs in response to signaling through group I metabotropic glutamate receptors (mGlu-R). Thus a reduction in FMRP in FXS results in increased translation of the bound mRNAs at synapses. This leads to an imbalance

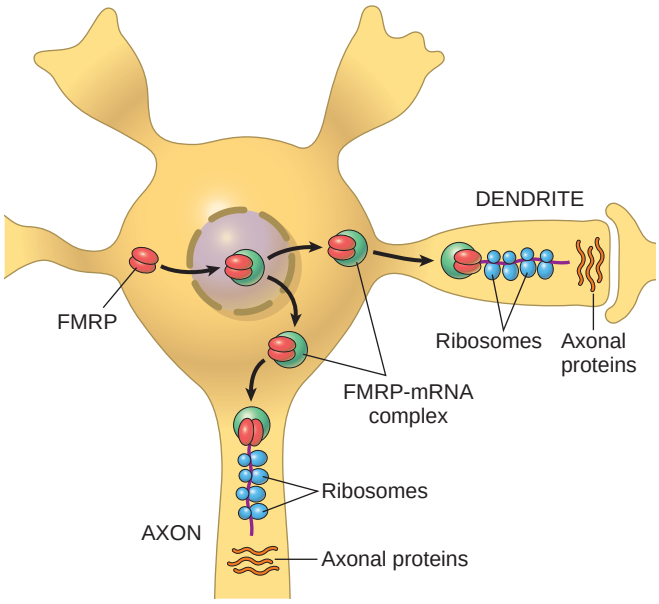


Figure 5.26 A model for the action of familial mental retardation protein *(FMRP)* in neurons. (Modified from Hin P, Warren ST: New insights into fragile X syndrome: from molecules to neurobehavior, *Trends Biochem Sci* 28:152, 2003.)

in the production of proteins at the synapses resulting in loss of synaptic plasticity—the ability of synapses to change and adapt in response to specific signals. Synaptic plasticity is essential for learning and memory.

Although demonstration of an abnormal karyotype led to the identification of this disorder, polymerase chain reaction (PCR)–based detection of the repeats is now the method of choice for diagnosis.

Fragile X–Associated Tremor/Ataxia Syndrome and Fragile X–Associated Primary Ovarian Failure

Although initially assumed to be innocuous, CGG premutations in the *FMR1* gene can cause two disorders that are phenotypically different from FXS and occur through a distinct mechanism involving a toxic gain of function. A decade after the discovery that CGG repeat expansions cause FXS, it became clear that approximately 20% of females carrying the premutation (carrier females) have premature ovarian failure (before the age of 40 years). This condition is called *fragile X–associated primary ovarian failure.* Affected women have menstrual irregularities and decreased fertility. FSH levels are elevated, and antimüllerian hormone levels are decreased—both markers of declining ovarian function. They develop menopause approximately 5 years earlier than controls. Approximately 50% of premutation-carrying males (transmitting males) exhibit a progressive neurodegenerative syndrome starting in their sixth decade. This syndrome, referred to as *fragile X–associated tremor/ataxia,* is characterized by intention tremors and cerebellar ataxia and may progress to parkinsonism.

How do premutations cause disease? In these patients, the *FMR1* gene instead of being methylated and silenced continues to be transcribed. CGG-containing *FMR1* mRNAs, so formed, are "toxic." They recruit RNA-binding proteins and impair their function by sequestration from their normal locales. The expanded *FMR1* mRNA and the sequestered RNA-binding proteins aggregate in the nucleus and form intranuclear inclusions in both the central and the peripheral nervous systems. As in FXS, males are affected much more frequently and more severely than female permutation carriers. The pathogenesis of fragile X–associated primary ovarian insufficiency is less well understood. Aggregates containing *FMR1* mRNA have been detected in granulosa cells and ovarian stromal cells. Perhaps these aggregates cause premature death of ovarian follicles.

KEY CONCEPTS

FRAGILE X SYNDROME (FXS)

- Pathologic amplification of trinucleotide repeats causes loss-of-function (FXS) or gain-of-function (Huntington disease) mutations. Most such mutations produce neurodegenerative disorders.
- FXS results from loss of *FMR1* gene function and is characterized by severe intellectual disability and a variety of neuropsychiatric conditions such as autism spectrum disorders.
- In the normal population, there are about 29 to 55 CGG repeats in the *FMR1* gene. The genomes of carrier males and females contain premutations with 55 to 200 CGG repeats that can expand to 4000 repeats (full mutations) during oogenesis. When full mutations are transmitted to progeny, FXS occurs.
- Carriers of premutations develop fragile X–associated tremor/ataxia and fragile X–associated primary ovarian failure due to toxic gain of function by the abnormal *FMR1* mRNA.

Mutations in Mitochondrial Genes—Leber Hereditary Optic Neuropathy

The vast majority of genes are located on chromosomes in the cell nucleus and are inherited in classic Mendelian fashion. Several mitochondrial genes exist, however, that are inherited in quite a different manner. *A feature unique to mtDNA is maternal inheritance.* This peculiarity exists because ova contain numerous mitochondria within their abundant cytoplasm, whereas spermatozoa contain few, if any. Hence the mtDNA complement of the zygote is derived entirely from the ovum. Thus, mothers transmit mtDNA to all their offspring, male and female; daughters, but not sons, transmit the DNA further to their progeny (Fig. 5.27). Several other features apply to mitochondrial inheritance.

- *Human mtDNA contains 37 genes,* of which 22 are transcribed into transfer RNAs and two into ribosomal RNAs. The remaining 13 genes encode subunits of the respiratory chain enzymes. Because mtDNA encodes enzymes involved in oxidative phosphorylation, mutations affecting these genes exert their deleterious effects primarily on the organs most dependent on oxidative phosphorylation such as the central nervous system, skeletal muscle, cardiac muscle, liver, and kidneys.
- *Each mitochondrion contains thousands of copies of mtDNAs,* and typically deleterious mutations of mtDNA affect some, but not all, of these copies. Thus tissues and, indeed, individuals may harbor both wild-type and mutant mtDNA, a situation called *heteroplasmy.* A minimum number of mutant mtDNAs must be present in a cell or

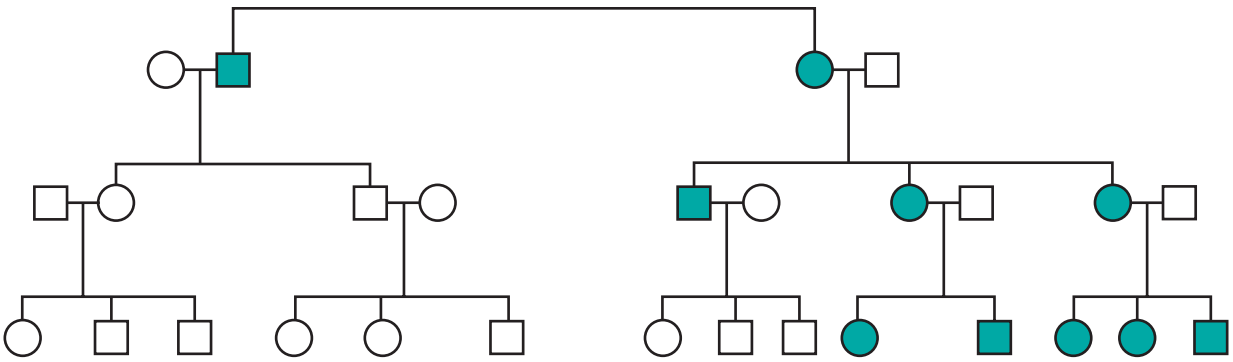


Figure 5.27 Pedigree of Leber hereditary optic neuropathy, a disorder caused by mutation in mitochondrial DNA. Note that all progeny of an affected male *(shaded squares)* are normal, but all children, male and female, of the affected female *(shaded circles)* manifest disease to a variable degree as discussed in the text.

tissue before oxidative dysfunction gives rise to disease. This is called the "threshold effect." Not surprisingly, the threshold is reached most easily in the metabolically active tissues listed earlier.

- *During cell division, mitochondria and their contained DNA are randomly distributed* to the daughter cells. Thus when a cell containing normal and mutant mtDNA divides, the proportion of the normal and mutant mtDNA in daughter cells is extremely variable. Therefore the expression of disorders resulting from mutations in mtDNA is quite variable.

Diseases associated with mitochondrial inheritance are rare, and, as mentioned earlier, many of them affect the neuromuscular system. *Leber hereditary optic neuropathy* is a prototype of this type of disorder. It is a neurodegenerative disease that manifests as progressive bilateral loss of central vision. Visual impairment is first noted between ages 15 and 35, leading eventually to blindness. Cardiac conduction defects and minor neurologic manifestations have also been observed in some families.

Genomic Imprinting

We all inherit two copies of each autosomal gene, carried on homologous maternal and paternal chromosomes. In the past, it had been assumed that there is no functional difference between the alleles derived from the mother or the father. Studies have now provided definite evidence that, at least with respect to some genes, important functional differences exist between the paternal allele and the maternal allele. These differences result from an epigenetic process called *imprinting.* In most cases, imprinting selectively inactivates either the maternal or the paternal allele. Thus, *maternal imprinting* refers to transcriptional silencing of the maternal allele, whereas *paternal imprinting* implies that the paternal allele is inactivated.

Imprinting occurs in the ovum or the sperm, before fertilization, and then is stably transmitted to all somatic cells through mitosis. As with other instances of epigenetic regulation, imprinting is associated with differential patterns of DNA methylation at CG nucleotides. Other mechanisms include histone H4 deacetylation and methylation (Chapter 1). Regardless of the mechanism, it is believed that such marking of paternal and maternal chromosomes occurs during gametogenesis, and thus it seems that from the moment of conception some chromosomes remember where they came from. The exact number of imprinted genes is not known; estimates range from 200 to 600. Although imprinted genes may occur in isolation, more commonly they are found in groups that are regulated by common *cis*-acting elements called imprinting control regions. Genomic imprinting is best illustrated by two uncommon genetic disorders: Prader-Willi syndrome and Angelman syndrome, which were originally believed to be unrelated until the genetic lesions responsible for them were mapped to the same location. They are described next.

Prader-Willi Syndrome and Angelman Syndrome

Prader-Willi syndrome is characterized by intellectual disability, short stature, hypotonia, profound hyperphagia, obesity, small hands and feet, and hypogonadism. In 65% to 70% of cases, an interstitial deletion of band q12 in the long arm of chromosome 15, del(15)(q11.2q13), can be detected. In most cases the breakpoints are the same, causing a 5-Mb deletion. *It is striking that in all cases the deletion affects the paternally derived chromosome 15.* In contrast to Prader-Willi syndrome, patients with the phenotypically distinct Angelman syndrome are *born with a deletion of the same chromosomal region derived from their mothers.* **Patients with Angelman syndrome also have intellectual disability, but in addition they present with microcephaly, ataxic gait, seizures, and inappropriate laughter.** Because of their laughter and ataxia, they have been referred to as "happy puppets." A comparison of these two syndromes clearly demonstrates the *parent-of-origin* effects on gene function.

The molecular basis of these two syndromes lies in the process of genomic imprinting (Fig. 5.28). Three mechanisms are involved.

- *Deletions.* It is known that a gene or set of genes on maternal chromosome 15q12 is imprinted (and hence silenced), and thus the only functional alleles are provided by the paternal chromosome. When these are lost as a result of a deletion, the person develops Prader-Willi syndrome. Conversely, a distinct gene that also maps to the same region of chromosome 15 is imprinted on the paternal chromosome. Only the maternally derived allele of this gene is normally active. Deletion of this maternal

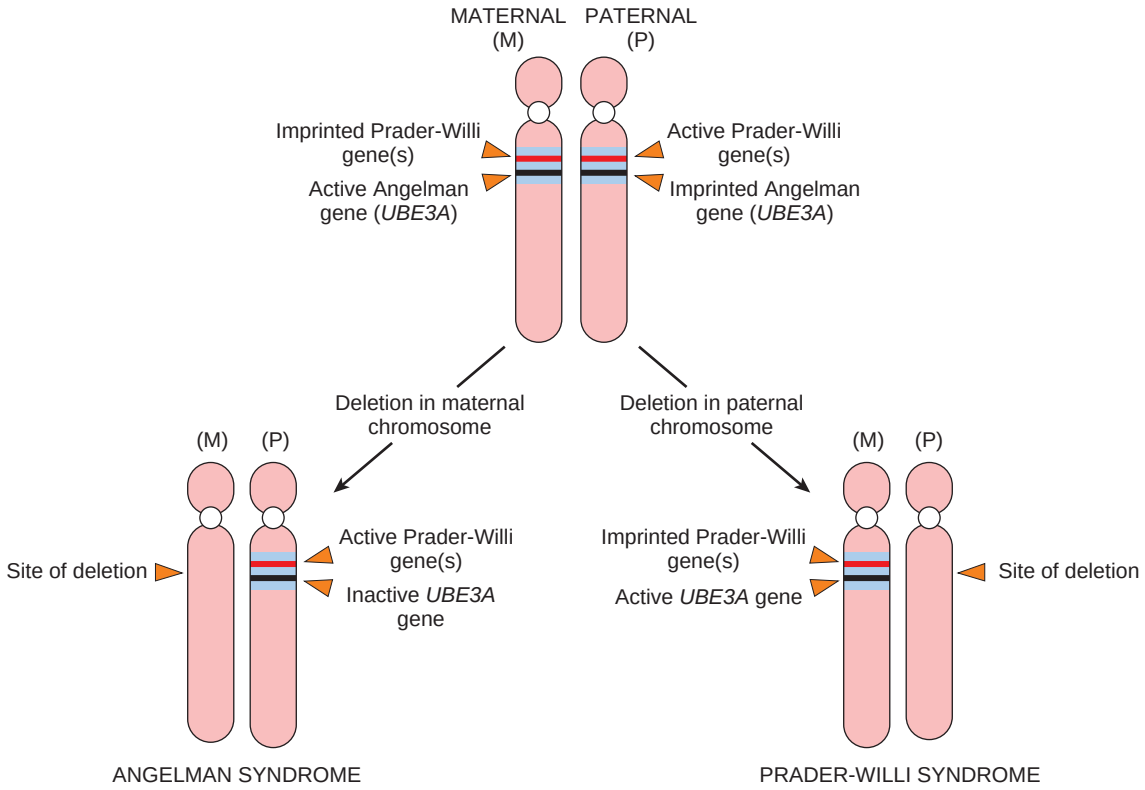


Figure 5.28 Diagrammatic representation of Prader-Willi and Angelman syndromes.

gene on chromosome 15 gives rise to the Angelman syndrome. Deletions account for about 70% cases.

- *Uniparental disomy.* Molecular studies of cytogenetically normal individuals with Prader-Willi syndrome (i.e., those without the deletion) have revealed that they have two maternal copies of chromosome 15. Inheritance of both chromosomes of a pair from one parent is called *uniparental disomy.* The net effect is the same (i.e., the person does not have a functional set of genes from the [nonimprinted] paternal chromosomes 15). Angelman syndrome, as might be expected, can also result from uniparental disomy of paternal chromosome 15. This is the second most common mechanism, responsible for 20% to 25% cases.
- *Defective imprinting.* In a small minority of patients (1% to 4%), there is an imprinting defect. In some patients with Prader-Willi syndrome, the paternal chromosome carries the maternal imprint, and conversely in Angelman syndrome, the maternal chromosome carries the paternal imprint (hence there are no functional alleles).

The genetic basis of these two imprinting disorders is now being unraveled.

- In Angelman syndrome, the affected gene is a ubiquitin ligase that is involved in catalyzing the transfer of activated ubiquitin to target protein substrates. The gene, called *UBE3A,* maps within the 15q12 region, is imprinted on the paternal chromosome, and is expressed from the maternal allele primarily in specific regions of the brain. Absence of *UBE3A* inhibits synapse formation and synaptic plasticity. The imprinting is tissue-specific in that *UBE3A* is expressed from both alleles in most tissues.
- In contrast to Angelman syndrome, no single gene has been implicated in Prader-Willi syndrome. Instead, a series of genes located in the 15q11.2-q13 interval (which are imprinted on the maternal chromosome and expressed from the paternal chromosome) are believed to be involved. These include the SNORP family of genes that encode small nucleolar RNAs. These RNAs are noncoding molecules that are involved in posttranscriptional modifications of ribosomal RNAs and other small nuclear RNAs. Loss of SNORP functions is believed to contribute to Prader-Willi syndrome, but the precise mechanisms are unclear.

Molecular diagnosis of these syndromes is based on assessment of methylation status of marker genes and FISH. The importance of imprinting is not restricted to rare chromosomal disorders. Parent-of-origin effects have been identified in a variety of inherited diseases such as Huntington disease and myotonic dystrophy and in tumorigenesis.

KEY CONCEPTS

GENOMIC IMPRINTING

- Imprinting involves transcriptional silencing of the paternal or maternal copies of certain genes during gametogenesis. For such genes, only one functional copy exists in the individual. Loss of the functional (not imprinted) allele by deletion gives rise to diseases.

- In Prader-Willi syndrome there is deletion of band q12 on the long arm of paternal chromosome 15. Genes in this region of maternal chromosome 15 are imprinted, so there is complete loss of their functions. Patients have intellectual disability, short stature, hypotonia, hyperphagia, small hands and feet, and hypogonadism.
- In Angelman syndrome there is deletion of the same region from the maternal chromosome, and genes on the corresponding region of paternal chromosome 15 are imprinted. These patients have intellectual disability, ataxia, seizures, and inappropriate laughter.

Gonadal Mosaicism

It was mentioned earlier that with every autosomal dominant disorder some patients do not have affected parents. In such patients the disorder results from a new mutation in the egg or the sperm from which they were derived; as such, their siblings are neither affected nor at increased risk for development of the disease. This is not always the case, however. *In some autosomal dominant disorders, exemplified by osteogenesis imperfecta, phenotypically normal parents have more than one affected child.* This clearly violates the laws of Mendelian inheritance. Studies indicate that gonadal mosaicism may be responsible for such unusual pedigrees. Such mosaicism results from a mutation that occurs postzygotically during early (embryonic) development. If the mutation affects only cells destined to form the gonads, the gametes carry the mutation, but the somatic cells of the individual are completely normal. A phenotypically normal parent who has gonadal mosaicism can transmit the disease-causing mutation to the offspring through their mutated gametes. Because the progenitor cells of the gametes carry the mutation, there is a possibility that more than one child of such a parent would be affected. Obviously the likelihood of such an occurrence depends on the proportion of germ cells carrying the mutation.

MOLECULAR GENETIC DIAGNOSIS

The field of molecular genetic diagnostics emerged in the latter half of the 20th century in the form of labor-intensive, low throughput methods such as conventional karyotyping for diagnosis of cytogenetic disorders (e.g., Down syndrome) and DNA-based assays such as Southern blotting for the diagnosis of Huntington disease. Over time, a steady stream of technologic breakthroughs has led to ever-increasing capabilities, particularly in the area of nucleic acid sequencing. These advances began with the development of Sanger DNA sequencing in 1977 and PCR in 1983 and accelerated rapidly with the advent of high throughput, massively parallel sequencing strategies (often lumped together under the rubric next-generation sequencing [NGS]) in the late 1990s. NGS has continued to improve in terms of speed and cost, and it is now possible to sequence an entire genome for less than $1000 in a few days. As a result, nucleic acid–based testing is taking a central role in the diagnosis and management of many diseases, and we currently are in a transition period during which many simple, single-gene tests using "old" technologies are being slowly but surely supplanted by more comprehensive NGS-based approaches. That said, certain types of genetic lesions are difficult to detect by NGS, and for the foreseeable future older technologies will continue to have an important role in testing for genetic diseases. An exhaustive discussion of molecular diagnostics is beyond the scope of this book; here, we highlight some of the most useful and widely deployed approaches.

When considering molecular genetic testing, it is important to remember that genetic markers can be either constitutional (i.e., present in every cell of the affected person, as with a *CFTR* mutation in a patient with cystic fibrosis) or somatic (i.e., restricted to specific tissue types or lesions, as with mutations in *RAS* in a variety of human cancers). Similarly, in suspected infections, nucleic acids that are specific to the infectious agent may be confined to particular cells or body sites. These considerations determine the nature of the sample used for the assay (e.g., peripheral blood cells, tumor tissue, nasopharyngeal swab).

Diagnostic Methods and Indications for Testing

There are a dizzying number of techniques and indications for performing molecular genetic diagnostic tests on patient specimens. The burden of choice is often problematic, both for molecular pathologists who design tests and for clinicians who need to choose the optimal test for their patients.

Laboratory Considerations

Pathologists developing tests focus on the sensitivity, specificity, accuracy, and reproducibility of different methods as well as practical factors like cost, labor, reliability, and turnaround time. To choose the most appropriate diagnostic technique, one must have a thorough knowledge of the spectrum of genetic anomalies that are responsible for the disease in the patient population under study. Disease-causing genetic anomalies range in size from single base substitutions to gains or losses of entire chromosomes and often vary widely in frequency among ethnic groups. Proper test design requires careful consideration of these factors. For example, standard cystic fibrosis testing for pathogenic mutations in *CFTR* has a sensitivity of 94% in Ashkenazi Jews, but identifies less than 50% of affected patients in Asian populations. In cases with negative standard test results and a high clinical suspicion, further testing is indicated. However, even with extensive sequencing of *CFTR,* no pathogenic mutations are identified in approximately 10% of patients with classic cystic fibrosis; this is likely because NGS assays miss certain genetic lesions (e.g., kilobase scale deletions and gene rearrangements) that are better detected with other methods. Issues like this arise frequently in genetic testing, and close communication between primary care clinicians, medical genetics specialists, and diagnosticians is required to select the optimal test strategy in difficult cases.

Indications for Analysis of Inherited Genetic Alterations

Although it is not unusual for inherited genetic disorders to present in adulthood, most testing is performed during the prenatal or postnatal/childhood periods. Mendelian disorders caused by mutations in specific genes number in the thousands, and definitive diagnosis of most of them is possible by DNA sequencing tests.

Other inherited disorders are the result of chromosomal aberrations that usually present prenatally or at birth. Prenatal testing should be offered for all fetuses at risk for a cytogenetic abnormality. Possible indications include the following:

- Advanced maternal age
- A parent known to carry a balanced chromosomal rearrangement, which greatly increases the frequency of abnormal chromosome segregation during meiosis and the risk of aneuploidy in the fertilized ovum
- Fetal anomalies observed on ultrasound
- Routine maternal blood screening indicating an increased risk of Down syndrome (trisomy 21) or another trisomy

Prenatal testing may also be considered for fetuses at known risk for Mendelian disorders (e.g., cystic fibrosis, spinal muscular atrophy) based on family history. At present it is usually performed on cells obtained by amniocentesis, chorionic villus biopsy, or umbilical cord blood. However, as much as 10% of the free DNA in a pregnant mother's blood is of fetal origin, and new sequencing technologies have opened the door to an era of noninvasive prenatal diagnostics utilizing this source of DNA.

Following birth, testing is ideally done as soon as the possibility of constitutional genetic disease arises. It is most commonly performed on peripheral blood DNA and is targeted based on clinical suspicion. In newborns or children, indications are as follows:

- Multiple congenital anomalies
- Suspicion of a metabolic syndrome
- Unexplained intellectual disability, and/or developmental delay
- Suspected aneuploidy (e.g., features of Down syndrome) or other syndromic chromosomal abnormality (e.g., deletions, inversions)
- Suspected monogenic disease, whether previously described in the family or new

In older patients, testing logically becomes more focused toward genetic diseases that manifest at later stages of life. Again, the possibilities are vast, but the more common indications include the following:

- Inherited cancer syndromes (triggered by either family history or an unusual cancer presentation, such as multiple cancer types or unusually young age at diagnosis)
- Atypically mild monogenic disease (e.g., attenuated cystic fibrosis)
- Neurodegenerative disorders (e.g., familial Alzheimer disease, Huntington disease)

Indications for Analysis of Acquired Genetic Alterations

In this era of molecularly targeted therapies it is increasingly important to identify nucleic acid sequences or aberrations that are specific for acquired diseases (e.g., cancer and infectious disease). The technical approaches are the same as those used for germline Mendelian disorders. Common indications include the following:

- *Diagnosis and management of cancer* (Chapter 7)
  - Detection of tumor-specific mutations and cytogenetic alterations that are the hallmarks of specific tumors (e.g., *BCR-ABL* fusion genes in chronic myeloid leukemia [CML])
  - Determination of clonality as an indicator of a neoplastic condition
  - Identification of specific genetic alterations that can direct therapeutic choices (e.g., amplification of *HER2* [official gene name *ERBB2*] in breast cancer or mutation of *EGFR* [official gene name *ERBB1*] in lung cancer)
  - Determination of treatment efficacy (e.g., minimal residual disease detection in patients with CML by quantitative PCR for BCR-ABL)
  - Detection of drug-resistant secondary mutations in malignancies treated with therapies that target specific proteins (e.g., mutated EGFR)
- *Diagnosis and management of infectious disease* (Chapter 8)
  - Detection of microorganism-specific genetic material for definitive diagnosis (e.g., human immunodeficiency virus [HIV], mycobacteria, human papillomavirus, herpesvirus in central nervous system)
  - Identification of genetic alterations in the genomes of microbes that are associated with drug resistance
  - Determination of treatment efficacy (e.g., assessment of viral loads in HIV, Epstein-Barr virus, and hepatitis C virus infection)

PCR and Detection of DNA Sequence Alterations

PCR analysis, which involves synthesis of relatively short DNA fragments from a DNA template, has been a mainstay of molecular diagnostics for the last few decades. By using appropriate heat-stable DNA polymerases and thermal cycling, the target DNA (usually <1000 bp) lying between designed primer sites is exponentially amplified from as little as one original copy, greatly simplifying secondary sequence analysis. Many options exist for subsequent analysis, each with different strengths and weaknesses. A few of the more common options are described below:

- *Sanger sequencing.* A single PCR product is mixed with a DNA polymerase, a specific primer, nucleotides, and four dead-end (di-deoxy terminator) nucleotides (A, T, G, and C) labeled with different fluorescent tags. The ensuing reaction produces a ladder of DNA molecules of all possible lengths, each labeled with a tag corresponding to the base at which the reaction stopped due to incorporation of a terminator nucleotide. After size separation by electrophoresis, the sequence is "read" and compared with the normal sequence to detect mutations.
- *Next-generation sequencing (NGS).* PCR using primers for many different genomic regions is performed simultaneously, and the resulting mixture of PCR products enriched for regions of interest are subjected to NGS (described in more detail later). NGS is more sensitive than Sanger sequencing, as it can reliably identify the presence of mutations in only a small percentage of the individual sequencing reads. This situation is encountered frequently in cancers, either because the mutation in question is only present in a small fraction of tumor cells or because the tumor cells are heavily contaminated with genetically normal stromal cells.
- *Single-base primer extension.* This is a useful approach for identifying mutations at a specific nucleotide position (e.g., an oncogenic mutation in codon 600 of the *BRAF* gene). A primer is added to the PCR product that

hybridizes one base upstream of the target, differently colored terminator fluorescent nucleotides are added (corresponding to normal and variant bases), and a single base polymerase extension is performed. The relative amounts of normal/variant fluorescence are then detected (Fig. 5.29). This technique is very sensitive, but has the obvious disadvantage of only producing 1 bp of sequence data.

- *Restriction fragment length analysis.* This simple approach takes advantage of the digestion of DNA with endonucleases known as restriction enzymes that recognize and cut DNA at specific sequences. If the specific mutation is known to affect a restriction site, the amplified PCR product may be digested, and the normal and mutant PCR products will yield fragments of different sizes that are easily distinguished. Needless to say, this approach is considerably less comprehensive than direct sequencing but remains useful for molecular diagnosis when the causal mutation always occurs at an invariant nucleotide position.

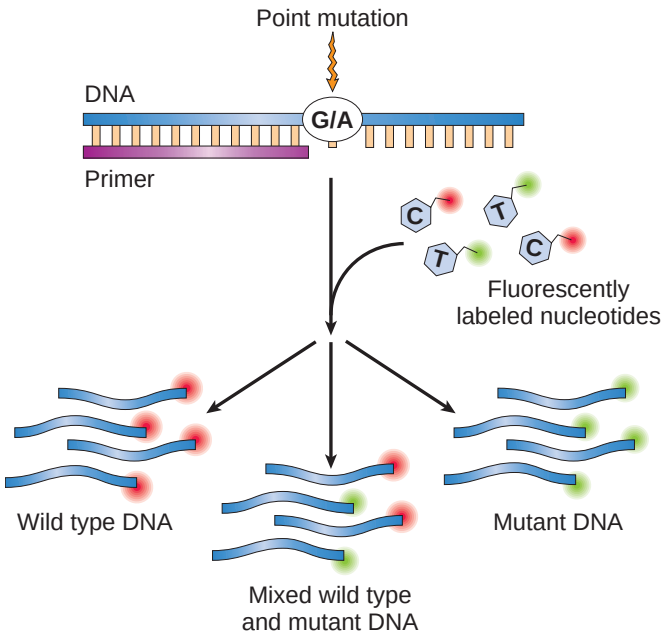


Figure 5.29 Single-base extension analysis of a polymerase chain reaction product, using a primer to interrogate a single base position. Nucleotides complementary to the mutant and wild-type bases at the queried position are labeled with different fluorophores, such that incorporation yields fluorescent signals of varying intensity based on the ratio of mutant to wild-type DNA present.

- *Amplicon length analysis.* Mutations that affect the length of DNA (e.g., deletions or expansions) can be easily detected by PCR. As discussed earlier, several diseases, such as FXS, are associated with alterations in trinucleotide repeats. Fig. 5.30 reveals how PCR analysis can be used to detect this mutation. Two primers that flank the region containing the trinucleotide repeats at the 5′ end of the *FMR1* gene are used to amplify the intervening sequences. Because there are large differences in the number of repeats, the size of the PCR products obtained from the DNA of normal individuals, or those with a premutation, is quite different and can be easily distinguished by gel electrophoresis. An important caveat is that if the trinucleotide expansion is so large that it cannot be amplified by convention PCR, Southern blot analysis of genomic DNA may have to be performed. As discussed earlier, large expansions are not uncommon in FXS.

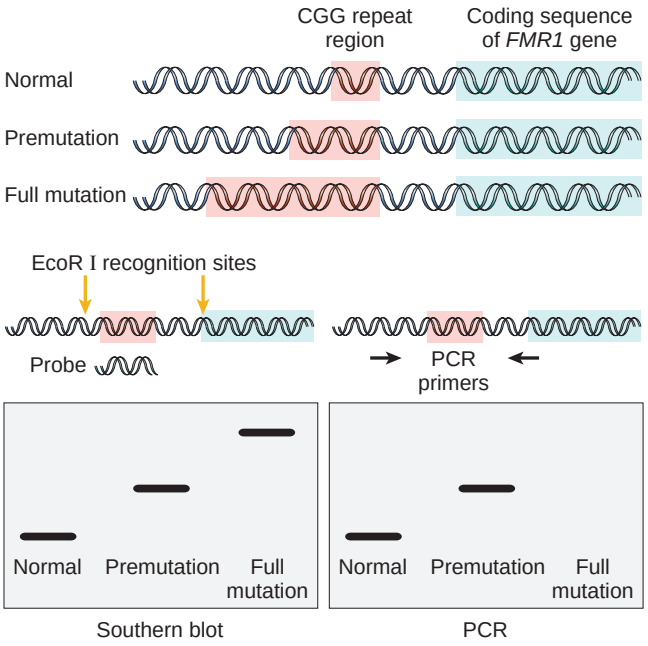


Figure 5.30 Diagnostic application of polymerase chain reaction *(PCR)* and Southern blot analysis in fragile X syndrome. With PCR, the differences in the size of CGG repeats between normal and premutation give rise to products of different sizes and mobility. With a full mutation, the region between the primers is too large to be amplified by conventional PCR. In Southern blot analysis the DNA is cut by enzymes that flank the CGG repeat region and is then probed with complementary DNA that binds to the affected part of the gene. A single small band is seen in normal males, a band of higher molecular weight is seen in males with premutation, and a very large (usually diffuse) band is seen in males with the full mutation.

- *Real-time PCR.* A variety of PCR-based technologies that use fluorophore indicators can detect and quantify the presence of particular nucleic acid sequences in real time (i.e., during the exponential phase of DNA amplification rather than post-PCR). It is most often used to monitor the frequency of cancer cells bearing characteristic genetic lesions in the blood or in tissues (e.g., the level of *BCR-ABL* fusion gene sequences in patients with CML), or the infectious load of certain viruses (e.g., HIV, Epstein-Barr virus).

Molecular Analysis of Genomic Alterations

A significant number of genetic lesions involve large deletions, duplications, or more complex rearrangements that are not readily amenable to detection using PCR or DNA sequencing approaches. Such genomic-scale alterations can be studied using a variety of hybridization-based techniques.

Fluorescence in Situ Hybridization (FISH)

FISH uses DNA probes that recognize sequences specific to particular chromosomal regions. To perform FISH, large fragments of cloned genomic DNA spanning up to 200 kb are labeled with fluorescent dyes and applied to

metaphase chromosome preparations or interphase nuclei that are pretreated so as to "melt" the genomic DNA. The probe hybridizes to its homologous genomic sequence and thus labels a specific chromosomal region that can be visualized under a fluorescent microscope. FISH can be performed on prenatal samples, peripheral blood cells, touch preparations from cancer biopsies, and even fixed archival tissue sections. It is used to detect numeric abnormalities of chromosomes (aneuploidy) (see Fig. 5.19), subtle microdeletions (see Fig. 5.21) and complex translocations that are not demonstrable by routine karyotyping, and gene amplification (e.g., *HER2* in breast cancer or *NMYC* amplification in neuroblastomas). It is also used in certain circumstances when rapid diagnosis is essential (e.g., when deciding to treat a patient with suspected acute promyelocytic leukemia with retinoic acid, which is effective only when a particular chromosomal translocation involving the retinoic acid receptor gene is present in the tumor cells; Chapter 13).

Cytogenomic Array Technology

FISH requires prior knowledge of the one or few specific chromosomal regions suspected of being altered in the test sample. However, **genomic abnormalities can also be detected without prior knowledge by using microarray technology to perform a global genomic survey.** First-generation platforms were designed for comparative genomic hybridization (CGH), while newer platforms incorporate single nucleotide polymorphism (SNP) genotyping approaches, offering multiple benefits.

- *Array-based comparative genomic hybridization.* In array CGH, the test DNA and a reference (normal) DNA are labeled with two different fluorescent dyes. The samples are then mixed and hybridized to an array spotted with DNA probes that span the human genome at regularly spaced intervals. At each chromosomal probe location, the binding of the labeled DNA from the two samples is compared. If the two samples are equal (i.e., the test sample is diploid), all spots on the array will fluoresce yellow due to equal admixture of green and red dyes. In contrast, if the test sample has a deletion or duplication, the probe spots corresponding to it will show skewing toward red or green (depending on gain or loss of material), allowing highly accurate determination of even focal copy number gain or loss anywhere across the genome.
- *Single nucleotide polymorphism genotyping arrays.* Newer types of genomic arrays are based on a similar concept, but some or all of the probes are designed to identify SNP sites genome-wide, which provides a number of advantages. As discussed in Chapter 1, SNPs are the most common type of DNA polymorphism, occurring on average approximately every 1000 nucleotides throughout the genome. They serve both as a physical landmark within the genome and as a genetic marker whose transmission can be followed from parent to child.

There are several testing platforms using different methodologies that allow SNPs to be analyzed genome-wide on arrays; details of these methods are beyond the scope of this discussion. Like CGH probes, these methods involving SNPs can be used to make copy number variation (CNV) calls, but by discriminating between SNP alleles at each particular location they also provide zygosity data (Fig. 5.31). The current generation of SNP arrays is quite comprehensive, with the largest containing greater than 4 million SNP probes.

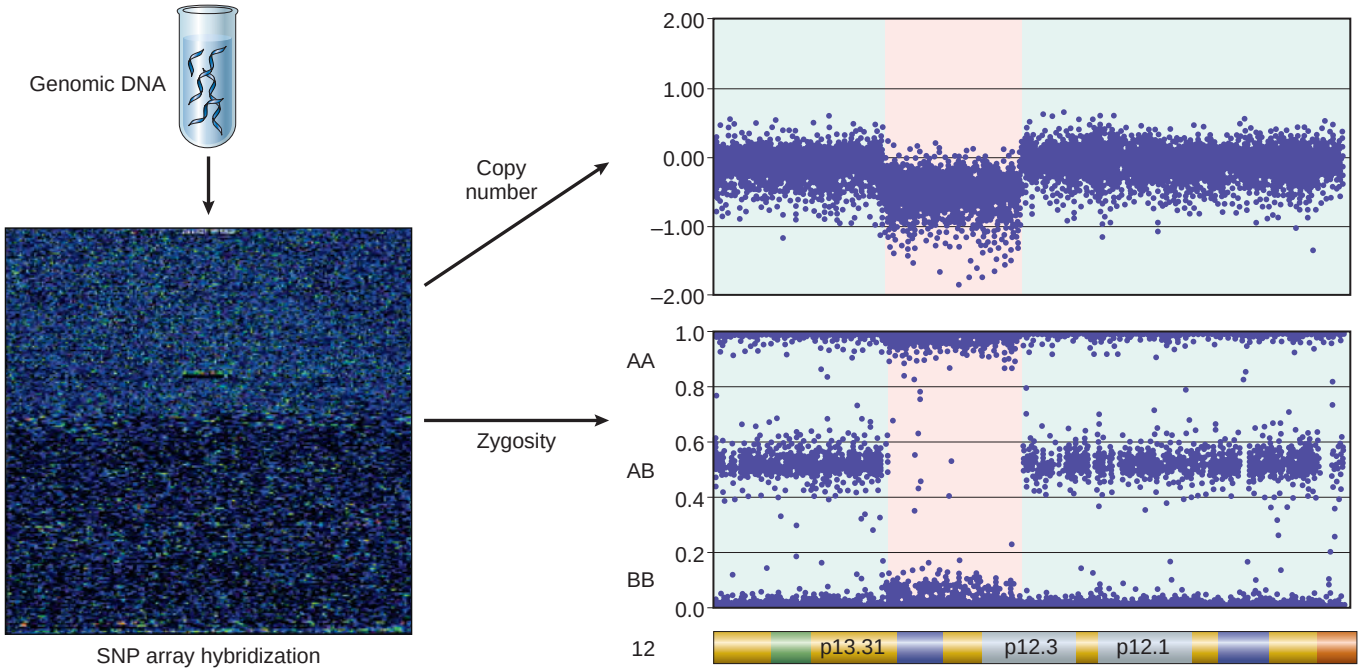


Figure 5.31 Analysis of copy number variation via single nucleotide polymorphism *(SNP)* cytogenomic array. Genomic DNA is labeled and hybridized to an array containing potentially millions of probe spots. Copy number is determined by overall intensity, and genotype is determined by allelic ratio. The example shown is the p arm of chromosome 12 in a patient with pediatric leukemia. The normal areas (green) show neutral (diploid) DNA content, and the zygosity plot shows the expected ratio of AA, AB, and BB SNP genotypes. The anomalous area (red) shows decreased overall intensity, and the zygosity plot shows absence of the mixed AB genotype, indicating a full heterozygous deletion. (Modified from Paulsson K, et al: Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia, *Proc Natl Acad Sci U S A* 107:21719–21724, 2010.)

In the clinical laboratory, SNP arrays are routinely used to uncover copy number abnormalities in pediatric patients when the karyotype is normal but a structural chromosomal abnormality is still suspected. Common indications include congenital abnormalities, dysmorphic features, developmental delay, and autism. Notably, unlike CGH arrays, SNP arrays can identify loss of heterozygosity. This capability is important in the diagnosis of disorders caused by uniparental disomy (e.g., in Prader-Willi syndrome and Angelman syndrome), where despite diploid genomic copy number, the SNP calls in the affected chromosomal regions showed reduction to homozygosity. SNP data can also help uncover other anomalies, such as mosaicism, which produces complex but distinctive skewing of zygosity plots.

Polymorphic Markers and Molecular Diagnosis

Clinical detection of disease-specific mutations is possible only if the gene responsible for the disorder is known and its sequence has been identified. If the exact nature of the genetic aberration is not known, or if testing for the primary defect is technically challenging or unfeasible, diagnostic laboratories can take advantage of the phenomenon of *linkage*. In humans, two DNA loci even 100,000 bp apart on the same chromosome are almost certain to cosegregate during meiosis due to the extremely low chance of a crossover event happening between them. Thus, the closer two loci are, the higher the probability that they will travel together in family pedigrees. In the event of a challenging or unknown pathogenic allele, a diagnostic laboratory can instead choose as a surrogate approach to examine nearby marker loci in the context of the family pedigree.

The two types of genetic polymorphisms most useful for linkage analysis are SNPs (described earlier) and repeat-length polymorphisms known as minisatellite and microsatellite repeats. Human DNA contains short repetitive sequences of DNA giving rise to what are called repeat-length polymorphisms. These polymorphisms are often subdivided on the basis of their length into microsatellite repeats and minisatellite repeats. Microsatellites are usually less than 1 kb and are characterized by a repeat size of 2 to 6 bp. Minisatellite repeats, by comparison, are larger (1 to 3 kb), and the repeat motif is usually 15 to 70 bp. The number of repeats, both in microsatellites and in minisatellites, is extremely variable within a given population, and hence these stretches of DNA can be used to establish genetic identity for linkage analysis. Different microsatellites and minisatellites are readily distinguishable on the basis of size, which can be measured by doing PCR amplification using primers that flank the repeat region. Fig. 5.32 depicts the application of microsatellite linkage analysis to the *PKD1* gene (historically very difficult to sequence) for the familial diagnosis of adult polycystic kidney disease. It can be seen that the longest microsatellite allele is linked in the family to the disease allele and can be used to track its transmission.

Assays to detect genetic polymorphisms are also important in many other areas of medicine, including in the determination of relatedness and identity in transplantation, cancer genetics, paternity testing, and forensic medicine. Since microsatellite markers are scattered throughout the human genome and have such a high level of polymorphism, they are ideal for differentiating between two individuals and to follow transmission of the marker from parent to child. Assays employing PCR amplification of informative microsatellite markers are now routinely used for determining paternity and for criminal investigations. PCR can be performed even with highly degraded biologic samples, allowing DNA testing to play a central role in forensic identifications. The same assays are used in patients following allogeneic hematopoietic stem cell transplant to detect and quantify chimerism, which is evaluated by assessing the relative amounts of donor-specific and host-specific microsatellite markers in the patient's blood cells.

Epigenetic Alterations

Epigenetics is defined as the study of heritable chemical modification of DNA or chromatin that does not alter the protein-encoding DNA sequence itself. Examples of such modifications include methylation of DNA and methylation and acetylation of histones (Chapter 1). Our understanding of the importance of these types of molecular alterations is rapidly growing, and it is clear that epigenetic modifications are critical for normal human development—including the regulation of tissue-specific gene expression, X chromosome inactivation, and imprinting—as well as for understanding the cellular perturbations in aging and cancer.

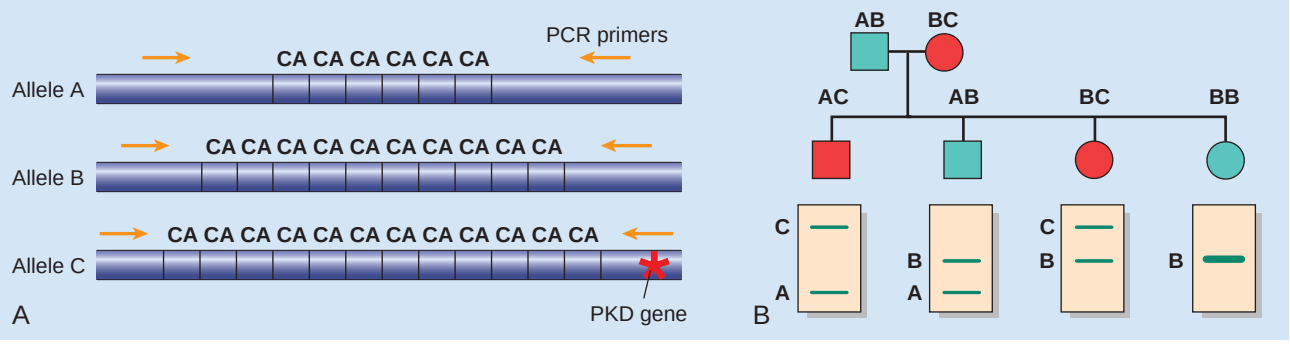


Figure 5.32 DNA polymorphisms resulting from a variable number of CA repeats. The three alleles (A) produce polymerase chain reaction *(PCR)* products of different sizes (B), thus identifying their origins from specific chromosomes. In the example depicted, allele C is linked to a mutation responsible for autosomal dominant polycystic kidney disease *(PKD)*. Application of this to detect progeny carrying the disease-related gene (red symbols) is illustrated in one hypothetical pedigree. Males (squares); females (circles).

Gene expression frequently correlates negatively with the level of methylation of DNA, particularly of cytosine residues in CG dinucleotide-rich promoter regions known as CpG islands. As discussed earlier in the section on genomic imprinting, increased methylation of CpG islands is associated with decreased gene expression and is accompanied by concomitant alterations of histone methylation and acetylation. Diagnosis of an increasing number of diseases involves the analysis of promoter methylation—for example, FXS, in which hypermethylation results in *FMR1* silencing. Methylation analysis is also essential in the diagnosis of Prader-Willi syndrome and Angelman syndrome.

Special techniques are needed to detect DNA methylation. One approach is to treat genomic DNA with sodium bisulfite, a chemical that converts unmethylated cytosine to uracil, which acts like thymine in downstream reactions. Methylated cytosines are protected from modification and remain unchanged. After treatment, it is then straightforward to discriminate the unmethylated (modified) DNA from the methylated (unmodified) DNA by DNA sequencing.

RNA Analysis

Because DNA exerts its effects on the cell through RNA expression and mature mRNA contains the coding sequences of all expressed genes, RNA can in principle substitute for DNA in a wide range of diagnostic applications. From a practical standpoint, DNA-based diagnosis is usually preferred, since DNA is much more stable. Nonetheless, RNA analysis is critical in several areas of molecular diagnostics. The most important applications are the detection and quantification of RNA viruses such as HIV and hepatitis C virus, but increasingly RNA analysis is also being used to evaluate cancer. mRNA expression profiling (described for breast cancer in Chapter 23) has become an important tool for molecular stratification of certain tumors. In some instances, cancer cells bearing particular chromosomal translocations are detected with greater sensitivity by analyzing mRNA (e.g., the *BCR-ABL* fusion transcript in CML). The principal reason for this is that most translocations occur in scattered locations within particular introns, which can be very large, complicating detection by PCR amplification of DNA. Since introns are removed by splicing during the formation of mRNA, PCR analysis is possible if RNA is first converted to complementary DNA (cDNA) by reverse transcriptase. Real-time PCR performed on cDNA is the method of choice for monitoring residual disease in patients with CML and certain other hematologic malignancies (Chapter 13).

Next-Generation Sequencing (NGS)

***Next-generation sequencing* is a term used to describe several newer DNA sequencing technologies that produce very large amounts of sequence data in a massively parallel manner.** These technologies have revolutionized biomedical research and are increasingly impacting molecular diagnostics. The factors propelling rapid adoption of NGS are price and performance: NGS allows us to perform previously impossible analyses at extremely low relative cost.

One feature that makes NGS far more clinically applicable than Sanger sequencing is its input sample requirement. Whereas Sanger sequencing requires a single homogeneous template DNA (usually a specific PCR product), NGS has no such requirement: any DNA from almost any source can be used. Because Sanger sequencing essentially provides an "average" result for a particular DNA region in a sample, samples with substantial sequence heterogeneity among input DNA molecules produce uninterpretable results. NGS, by contrast, is well suited to analysis of heterogeneous DNA samples due to the application of three common basic principles (Fig. 5.33).

- *Spatial separation.* At the beginning of the procedure, individual DNA molecules are physically isolated from each other in space. The specifics of this process are platform-dependent.
- *Local amplification.* After separation, the individual DNA molecules are amplified in place using a limited number of PCR cycles. Amplification is necessary so that sufficient signal can be generated to ensure detection and accuracy.
- *Parallel sequencing.* The amplified DNA molecules are then simultaneously sequenced by the addition of polymerases and other reagents, with each spatially separated and amplified original molecule yielding a "read" corresponding to its sequence. Sequence reads from NGS instruments are generally short, less than 500 bp.

Bioinformatics

Each NGS analysis generates a staggering amount of sequence data; for high throughput sequencing instruments, this amounts to 400 billion bp or more of sequence per day, enough to produce a high-quality sequence spanning an entire human genome. The downstream analysis necessary to make sense of these enormous data sets is sufficiently complex that specialized training in bioinformatics is needed to ensure its proper interpretation. Bioinformatics computational "pipelines" vary widely across applications and sample types, and a detailed discussion is beyond the scope of this text. However, it is worth describing the basic steps necessary to process this type of data.

- *Alignment.* Alignment is the process by which the sequencing reads from a sample are mapped onto a reference genome, allowing them to be viewed and interpreted in context.
- *Variant calling.* This process involves a systematic comparison of all of the sequence data from a sample with the reference sequence. The more reads that cover a particular location (sequencing depth), the more likely that a variant will be detected if present. If a locus shows sufficient evidence of a difference from the reference sequence, a variant call is made.
- *Variant annotation and interpretation.* Called variants can be annotated with various features (e.g., gene name, coding change and protein effect predictions, information from databases listing benign and pathogenic variants, clinical information). Once completed by the clinical laboratory, the annotated data are ready for reporting. Because the pace of advances in genetics and genomics has far outpaced the knowledge of the typical clinician, an important part of this report is a brief description of the biological and clinical significance of each pathogenic variant, which can be numerous in the case of some cancers.

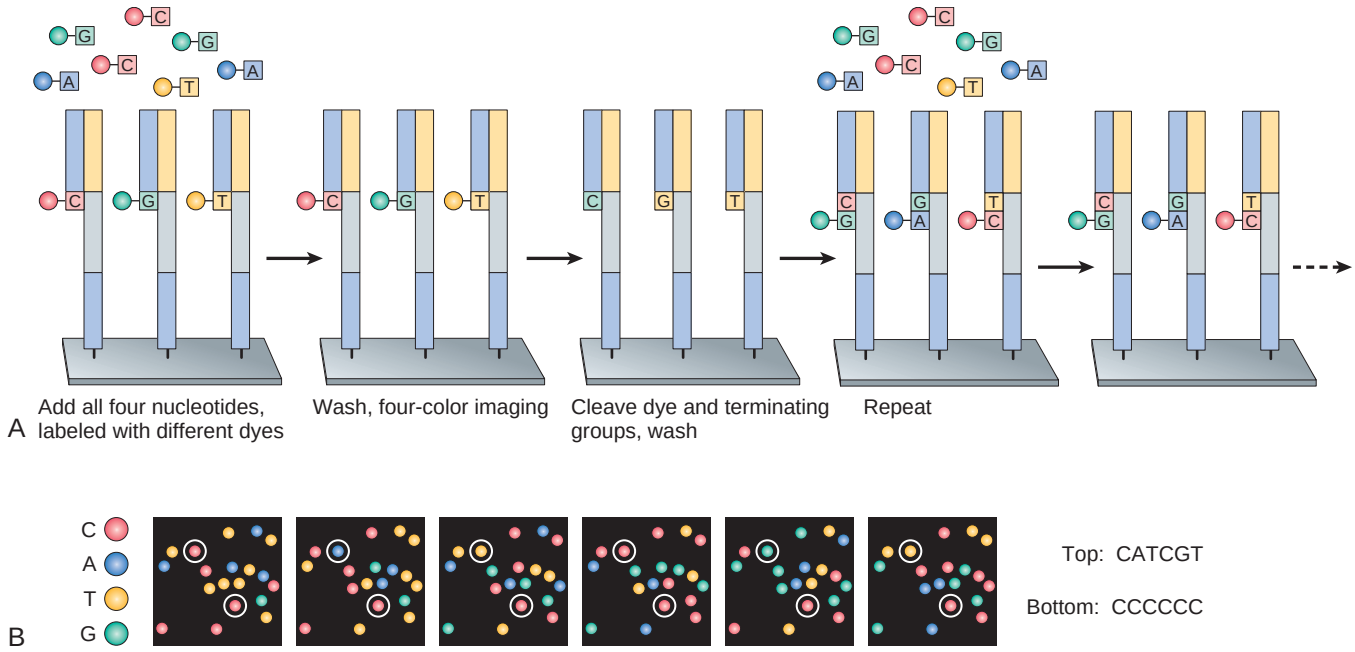


Figure 5.33 Principle of next-generation sequencing. Several alternative approaches currently are available for next-generation sequencing, and one of the more commonly used platforms is illustrated. (A) Short fragments of genomic DNA (template) 100 to 500 bp in length are immobilized on a solid-phase platform such as a glass slide, using universal capture primers that are complementary to adaptors that have previously been added to ends of the template fragments. The addition of fluorescently labeled complementary nucleotides, one per template DNA per cycle, occurs in a "massively parallel" fashion, at millions of templates immobilized on the solid phase at the same time. A four-color imaging camera captures the fluorescence emanating from each template location (corresponding to the specific incorporated nucleotide), following which the fluorescent dye is cleaved and washed away, and the entire cycle is repeated. (B) Powerful computational programs can decipher the images to generate sequences complementary to the template DNA at the end of one run, and these sequences are then mapped back to the reference genomic sequence to identify alterations. (Reproduced with permission from Metzker M: Sequencing technologies—the next generation, *Nat Rev Genet* 11:31–46, 2010, © Nature Publishing Group.)

- *Mutational signature calling*. In addition to totaling up individual mutations, informatics algorithms have been developed that detect patterns of mutations that point to particular environmental exposures (e.g., ultraviolet light) or underlying defects in DNA repair. The latter has become clinically relevant with the discovery that cancers with acquired defects in mismatch repair genes that lead to microsatellite instability (MSI) are highly responsive to immune checkpoint inhibitors (Chapter 7). Based on this observation, many clinical laboratories are now "calling" the MSI status of cancers based on NGS results.

Clinical Applications of Next-Generation DNA Sequencing

As discussed, any DNA sample can be analyzed by NGS. However, the DNA needs to first be prepared for sequencing by preparing a *library* of short DNA sequences that are enriched for genomic regions of interest (e.g., particular exons). There are several methods for NGS library preparation from genomic DNA, depending on the question and desired result. In the clinical laboratories, most applications of NGS are targeted toward constitutional genetic disease and cancer diagnostics, using a few different basic approaches:

- *Targeted sequencing*. Most current clinical laboratory NGS tests fall in this category. Targeted sequencing is focused on a carefully selected panel of genes, which maximizes sequencing depth while minimizing costs and the time and expense required for interpretation and clinical reporting. Sample preparation for targeted sequencing can be performed either by enriching for sequences of interest through a method called hybrid capture via custom complementary probes or by multiplex PCR. Single-gene NGS assays for inherited disorders are also becoming more common, and many laboratories that previously performed costly Sanger sequencing of multiple PCR products (i.e., to enable comprehensive *CFTR* sequencing) are now moving these assays to NGS platforms. Targeted NGS with gene panels is also now commonplace in evaluating individuals with genetic diseases such as cardiomyopathy and congenital deafness. In cancer testing, gene panels are widely used to perform detailed tumor profiling. Each tumor has a unique set of somatic mutations, and these assays aim to detect as many treatable or prognostic mutations as possible to offer individually tailored patient care. Increasingly, repeat testing at the time of disease recurrence is also being performed to understand mechanisms of drug resistance, which can serve to guide the selection of second-line therapies.
- *Whole-exome sequencing (WES)*. WES is a more expansive type of targeted sequencing, in which hundreds of thousands of custom probes are used to enrich for the roughly 1.5% of the genome that consists of protein-encoding exons prior to NGS. It is not routinely used in the evaluation of suspected germline disorders, but has

led to some wonderful success stories, allowing physicians to deliver answers and even therapies for children with orphan diseases who had suffered through prolonged and unsuccessful diagnostic odysseys. WES is also used in oncology to perform a very broad analysis, mostly in the research setting.

- *Whole-genome sequencing (WGS).* WGS is the most comprehensive type of DNA analysis that can be performed on an individual. Current indications for use in medical genetics are mostly limited to cases where exome sequencing has failed to provide an answer, but the clinical suspicion of genetic disease remains high. For cancer applications, WGS is the only form of NGS that can detect novel structural rearrangements (e.g., insertions, deletions, translocations) that may be clinically relevant, but relatively high costs, slow turnaround times, and significant informatics challenges still preclude its routine use in clinical practice.

Future Applications

Because NGS can be used to detect genetic anomalies of essentially any size scale, from SNPs to very large rearrangements and even aneuploidy, almost all of today's genetic diagnostic test modalities could in principle be supplanted by NGS. This includes RNA analysis because NGS-based analysis of the transcriptome (RNA-seq) is straightforward. As costs continue to drop, it is reasonable to expect NGS to occupy an increasingly prominent place in the diagnostics laboratory. Additionally, NGS has great promise for clinical applications in a number of other areas including microbiome analysis, blood screening for early markers of diseases (e.g., cancer), and vastly more sensitive methods of gauging the response of cancers to therapy (i.e., using circulating "free" DNA released from tumor cells that is harvested from the blood). Continuing technologic advances may even extend the applications further. For example, third-generation ("single molecule" or "next next-generation") technologies are emerging that can rapidly sequence single molecules in parallel without the need for focal amplification, and these could soon have an impact in the clinical laboratory.

ACKNOWLEDGMENT

The assistance of Jeremy Segal, MD, PhD, Director, Division of Genomic and Molecular Pathology, University of Chicago, in the revision of the section on molecular diagnosis is greatly appreciated.

SUGGESTED READINGS

Molecular Basis of Single-Gene Disorders—General

Dietz HC: New therapeutic approaches to mendelian disorders, *N Engl J Med* 363:852, 2010. *[An excellent discussion of treatment of genetic disorders based on molecular pathogenesis].*

Disorders Associated With Defects in Structural Proteins

Cortini F, Villa C, Marinelli B et al: Understanding the basis of Ehlers-Danlos syndrome in the era of the next-generation sequencing, *Arch Dermatol Res* 311:265, 2019. *[This review describes the power of next-generation sequencing in understanding the molecular basis of Ehlers-Danlos syndromes].*

Pyeritz RE: Etiology and pathogenesis of the Marfan syndrome: current understanding, *Ann Cardiothorac Surg* 6(6):595, 2017. *[A nice review of Marfan syndrome by a pioneer in the field].*

Salik I, Rawla P: Marfan syndrome. [Updated 2019 Feb 28]. *In StatPearls [internet]*, Treasure Island, FL, 2019, StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537339/?report=classic. *[An up-to-date and well-written article on the clinical and molecular features of Marfan syndrome].*

Tinkle B, Castori M, Berglund B et al: Hypermobile Ehlers-Danlos syndrome (a.k.a. Ehlers-Danlos syndrome Type III and Ehlers-Danlos syndrome hypermobility type): clinical description and natural history, *Am J Med Genet Part C Semin Med Genet* 175C:48, 2017. *[Clinically oriented review of the most common form of Ehlers-Danlos syndrome].*

Disorders Associated With Defects in Receptor Proteins

Benito-Vicente A, Uribe KB, Jebari S et al: Familial hypercholesterolemia: the most frequent cholesterol metabolism disorder caused disease, *Int J Mol Sci* 19(11):3426, 2018. *[A good accounting of low-density lipoprotein metabolism in familial hypercholesterolemia].*

Berberich AJ, Hegele RA: The complex molecular genetics of familial hypercholesterolaemia, *Nat Rev Cardiol* 16:9, 2019. *[The various less well-known genetic causes of familial hypercholesterolemia are discussed].*

Defesche JC, Gidding SS, Harada-Shiba M et al: Familial hypercholesterolaemia, *Nat Rev Dis Primers* 3(17093):1, 2017. *[A good and easy-to-follow overview of familial hypercholesterolemia].*

Sabatine MS: PCSK9 inhibitors: clinical evidence and implementation, *Nat Rev Cardiol* 16:155, 2019. *[The discovery of regulation of low-density lipoprotein receptor by PCSK9 and its clinical application].*

Lysosomal Storage Disorders

Aflaki E, Westbroek W, Sidransky E: The complicated relationship between Gaucher disease and Parkinsonism: insights from a rare disease, *Neuron* 93:737, 2017. *[A discussion of lysosomal storage disorders and neurodegenerative diseases].*

Breiden B, Sandhoff K: Ganglioside metabolism and its inherited diseases, *Methods Mol Biol* 1804:97, 2018. *[A review of ganglioside metabolism and pathogenesis of Tay-Sachs disease].*

Ferreiraa CR, Gahlc WA: Lysosomal storage diseases, *Transl Sci Rare Dis* 2(1):2017. *[An exhaustive description of all the lysosomal storage disorders].*

Filocamo M, Tomanin R, Bertola F et al: Biochemical and molecular analysis in mucopolysaccharidoses: what a paediatrician must know, *Ital J Pediatr* 44(2):129, 2018. *[A detailed discussion of these disorders in the pediatric population].*

Kohler L, Puertollano R, Raben N: Pompe disease: from basic science to therapy, *Neurother* 15:928, 2018. *[This article discusses details of Pompe disease].*

Ohashi T: Gene therapy for lysosomal storage diseases and peroxisomal diseases, *J Hum Genet* 64:139, 2019. *[Current approaches for the treatment of lysosomal storage disorders].*

Plotegher N, Duchen MR: Mitochondrial dysfunction and neurodegeneration in lysosomal storage disorders, *Trends Mol Med* 23(2):116, 2017. *[This review explains the role of mitochondrial injury and dysfunction in pathogenesis of lysosomal storage disorders].*

Schuchman EH, Desnick RJ: Types A and B Niemann-Pick disease, *Mol Genet Metab* 120(1–2):27, 2017. *[This article discusses details of types A and B Niemann-Pick disease].*

Seranova E, Connolly KJ, Zatyka M et al: Dysregulation of autophagy as a common mechanism in lysosomal storage diseases, *Essays Biochem* 61:733, 2017. *[A discussion of the critical role of dysregulation of autophagy in the development of lysosomal storage disorders].*

Stirnemann J, Belmatoug N, Camou F et al: A review of Gaucher disease pathophysiology, clinical presentation and treatments, *Int J Mol Sci* 18(441):1, 2017. *[This article discusses details of Gaucher disease].*

Weinstein DA, Steuerwald U, De Souza CFM et al: Inborn errors of metabolism with hypoglycemia: glycogen storage diseases and inherited disorders of gluconeogenesis, *Pediatr Clin North Am* 65(2):247, 2018. *[This article discusses details of glycogen storage diseases and inherited disorders of gluconeogenesis.].*

Cytogenetic Disorders Affecting Autosomes

Antonarakis SE: Down syndrome and the complexity of genome dosage imbalance, *Nat Rev Genet* 18:147, 2017. *[A discussion of the causes of Down syndrome due to imbalance in gene expression].*

Iwarsson E, Jacobsson B, Dagerhamn J et al: Analysis of cell-free fetal DNA in maternal blood for detection of trisomy 21, 18 and 13 in a general pregnant population and in a high risk population—a systematic review and meta-analysis, *Acta Obstet Gynecol Scand* 96:7, 2017. *[Utility of prenatal noninvasive testing of trisomies].*

Morsheimer M, Whitehorn TFB, Heimall J et al: The immune deficiency of chromosome 22q11.2 deletion syndrome, *Am J Med Genet* 173A:2366, 2017. *[A discussion of the emerging evidence that immunodeficiency in this syndrome is broader than typically assumed].*

Zamponi Z, Helguera PR: The shape of mitochondrial dysfunction in Down syndrome. *Dev Neurobiol Epub 2019. [A discussion of emerging evidence for the role of mitochondrial dysfunction in Down syndrome].*

Cytogenetic Disorders Affecting Sex Chromosomes

Kanakis GA, Nieschlag E: Klinefelter syndrome: more than hypogonadism, *Metab Clin Exp* 86:135, 2018. *[An exhaustive and updated review of pathogenesis of Klinefelter syndrome].*

Levitsky LL, O'Donnell Luria AH, Hayes FJ et al: Turner syndrome: update on biology and management across the life span, *Curr Opin Endocrinol Diabetes Obes* 22(1):65, 2015. *[A somewhat older but excellent succinct review of the clinical features and pathogenesis of Turner syndrome].*

Skuse D, Printzlau F, Wolstencroft J: Sex chromosomal aneuploidies. In Geschwind DH, Paulson HL, Klein C, editors: *Handbook of clinical neurology*, 2018, p 147. (3rd series) Neurogenetics, Part 1355, *[An excellent overview of sex chromosomal disorders].*

Diseases Caused by Trinucleotide Mutations

Bagni C, Zukin RS: A synaptic perspective of fragile X syndrome and autism spectrum disorders, *Neuron* 101:1070, 2019. *[A discussion of the loss of synaptic plasticity in the pathogenesis of fragile X syndrome].*

Hall DA, Berry-Kravis E: Fragile X syndrome and fragile X-associated tremor ataxia syndrome, *Handb Clin Neurol* 147:377, 2018. *[A well-balanced discussion of the genetic basis of fragile X syndrome and related disorders].*

Man L, Lekovich J, Rosenwaks Z et al: Fragile X-associated diminished ovarian reserve and primary ovarian insufficiency from molecular mechanisms to clinical manifestations, *Front Mol Neurosci* 12:1, 2017. *[A detailed discussion of the mechanism of ovarian failure in fragile X–associated primary ovarian insufficiency].*

Mila M, Alvarez-Mora MI, Madrigal I et al: Fragile X syndrome: an overview and update of the FMR1 gene, *Clin Genet* 93:197, 2018. *[A succinct update on the molecular basis of fragile X syndrome and related conditions].*

Salcedo-Arellano MJ, Dufour B, McLennan Y et al: Fragile X syndrome and associated disorders: clinical aspects and pathology, *Neurobiol Dis* 136:1, 2020. *[An excellent up-to-date review].*

Diseases Caused by Genomic Imprinting

Angulo MA, Butler MG: Cataletto ME: Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings, *J Endocrinol Invest* 38:1249, 2015. *[A well-written review of Prader-Willi syndrome].*

Buiting K, Williams C, Horsthemke B: Angelman syndrome—insights into a rare neurogenetic disorder, *Nat Rev Neurol* 12:584, 2016. *[A review focused on neurologic disturbances].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 6

Diseases of the Immune System

CHAPTER CONTENTS

The immune system is vital for survival because it protects us from infectious pathogens that abound in the environment and from the development of cancer. Predictably, immune deficiencies render individuals easy prey to infections and increase the incidence of certain cancers. But the immune system is itself capable of causing tissue injury and disease. Examples of disorders caused by immune responses include reactions to environmental substances that cause *allergies* and reactions against an individual's own tissues and cells *(autoimmunity)*.

This chapter is devoted to diseases caused by too little or too much immunologic reactivity. We also consider amyloidosis, a disease in which an abnormal protein, derived in many cases from immunoglobulins, is deposited in tissues.

First, we review some of the important features of normal immune responses, to provide a foundation for understanding the abnormalities that give rise to immunologic diseases.

THE NORMAL IMMUNE RESPONSE

The classic definition of immunity is protection from infectious pathogens, and the normal immune response is best understood in this context. However, immunity in its broader sense includes host reactions against cancers (tumor immunity), tissue transplants, and even self antigens (autoimmunity). The mechanisms of immunity fall into two broad categories (Fig. 6.1). *Innate immunity* (also called natural, or native, immunity) refers to intrinsic mechanisms that are poised to react immediately, and thus constitute the first line of defense. It is mediated by cells and molecules that recognize products of microbes and dead cells and induce rapid protective host reactions. *Adaptive immunity* (also called acquired, or specific, immunity) consists of mechanisms that are stimulated by ("adapt to") exposure to microbes and other foreign substances. It develops more slowly than innate immunity, but is even more powerful in combating infections. By convention, the term immune response usually refers to adaptive immunity.

Innate Immunity

Innate immunity is always present, ready to provide immediate defense against microbes and to eliminate damaged cells. The receptors and components of innate immunity have evolved to serve these purposes.

Components of Innate Immunity

The major components of innate immunity are epithelial barriers that block entry of microbes, phagocytic cells (mainly neutrophils and macrophages), dendritic cells, natural killer cells and other innate lymphoid cells, and several plasma proteins, including the proteins of the complement system.

- *Epithelia* of the skin and gastrointestinal and respiratory tracts act as mechanical barriers to the entry of microbes from the external environment. Epithelial cells also produce antimicrobial molecules such as defensins, and lymphocytes located in the epithelia combat microbes at these sites. If microbes breach epithelial boundaries, other defense mechanisms come into play.
- *Monocytes* and *neutrophils* are phagocytes in the blood that can be rapidly recruited to any site of infection; monocytes that enter the tissues and mature are called *macrophages*. Some tissue-resident macrophages (Kupffer cells in the liver, microglia in the brain, and alveolar macrophages in the lungs) develop from the yolk sac or fetal liver early in life and populate various tissues. Phagocytes sense the presence of microbes and other offending agents, ingest (phagocytose) these invaders, and destroy them. Because macrophages are the dominant cells of chronic inflammation, we described them in more detail in Chapter 3 in the discussion of chronic inflammation.
- *Dendritic cells (DCs)* are specialized cells present in epithelia, lymphoid organs, and most tissues. They capture protein antigens and display peptides for recognition by T lymphocytes. In addition to their antigen-presenting function, DCs are endowed with a rich collection of receptors that sense microbes and cell damage and stimulate the secretion of cytokines, mediators that play critical roles in inflammation and antiviral defense. Thus, DCs serve as sentinels that detect danger and initiate innate immune responses, but, unlike macrophages, they are not key participants in the destruction of microbes and other offending agents.
- *Innate lymphoid cells (ILCs)* are tissue-resident lymphocytes that lack T-cell antigen receptors and cannot respond to

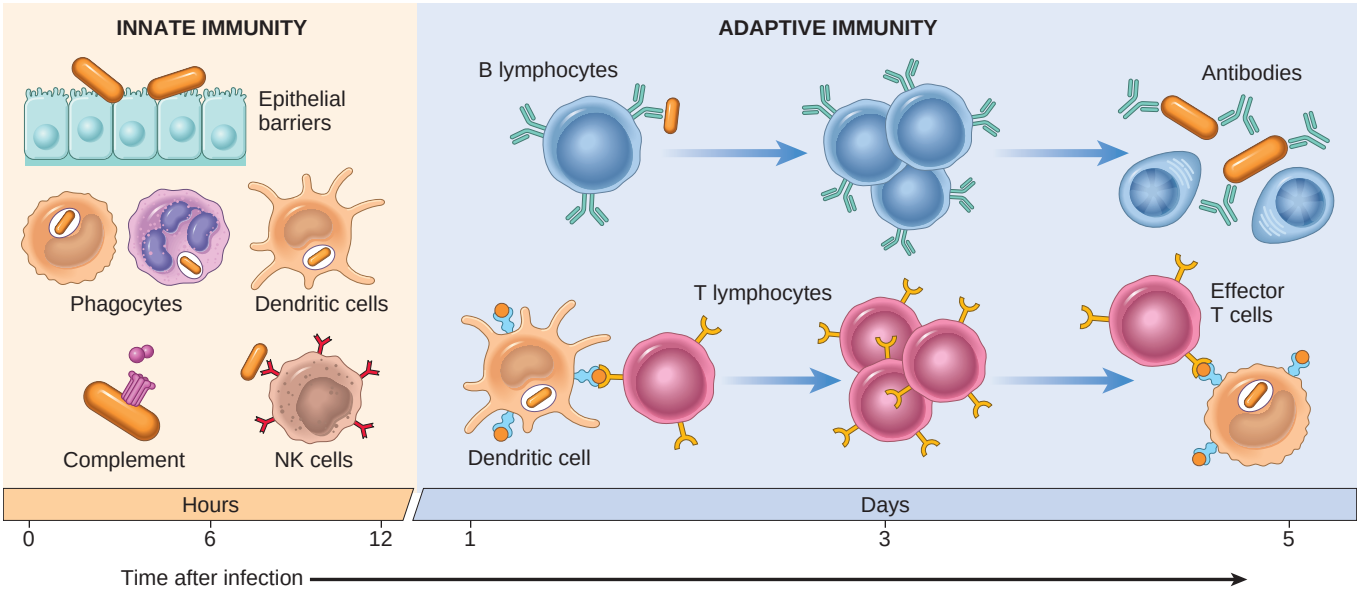


Figure 6.1 The principal components of innate and adaptive immunity. *NK*, Natural killer.

antigens, but instead are activated by cytokines and other mediators produced at sites of tissue damage. They are thought to be sources of inflammatory cytokines during early phases of immune reactions. ILCs are classified into groups based on the dominant cytokines they produce: groups 1, 2, and 3 ILCs produce many of the same cytokines as Th1, Th2, and Th17 subsets of CD4+ T cells, described later. *Natural killer (NK) cells* are one type of ILC that provide early protection against many viruses and intracellular bacteria; their properties and functions are described later.

- *Other cell types.* Several other cell types can sense and react to microbes. These include *mast cells*, which are capable of producing many mediators of inflammation (discussed later), and even epithelial and endothelial cells.
- *Plasma proteins.* In addition to these cells, several soluble proteins play important roles in innate immunity. The *complement system* (described in Chapter 3) consists of plasma proteins that are activated by microbes. Complement activation may occur through the alternative and lectin pathways as part of innate immune responses or through the classical pathway, which involves antibody-antigen complexes, as part of adaptive immune responses (Chapter 3). Other circulating proteins of innate immunity are mannose-binding lectin and C-reactive protein, both of which coat microbes and promote phagocytosis. Lung surfactant is also a component of innate immunity, providing protection against inhaled microbes.

Cellular Receptors for Microbes, Products of Damaged Cells, and Foreign Substances

Cells that participate in innate immunity are capable of recognizing certain components that are shared among related microbes and that are often essential for infectivity (and thus cannot be mutated to allow the microbes to evade the defense mechanisms). These microbial structures are called pathogen-associated molecular patterns. Leukocytes also recognize molecules released by injured and necrotic cells, which are called damage-associated molecular patterns. Collectively, the cellular receptors that recognize these molecules are called pattern recognition receptors.

Pattern recognition receptors are located in all cellular compartments where microbes may be present: plasma membrane receptors detect extracellular microbes, endosomal receptors detect ingested microbes, and cytosolic receptors detect microbes in the cytoplasm (Fig. 6.2). Several classes of these receptors have been identified.

Toll-like receptors. The best known of the pattern recognition receptors are the Toll-like receptors (TLRs), whose founding member, *Toll,* was discovered in *Drosophila* as a gene involved in development of the fly. A family of related proteins was later shown to be essential for host defense against microbes. Mammals have 10 TLRs, each recognizing a different set of microbial molecules. The TLRs are present in the plasma membrane and endosomal vesicles (see Fig. 6.2). All TLRs signal by a common pathway that culminates in the activation of two sets of transcription factors: (1) NF-κB, which stimulates the synthesis and secretion of cytokines and the expression of adhesion molecules, both of which are critical for the recruitment and activation of leukocytes (Chapter 3), and (2) interferon regulatory factors (IRFs), which stimulate the production of the antiviral cytokines, type I interferons. Inherited loss-of-function mutations affecting TLRs and their signaling pathways are associated with rare but serious immunodeficiency syndromes, described later in the chapter.

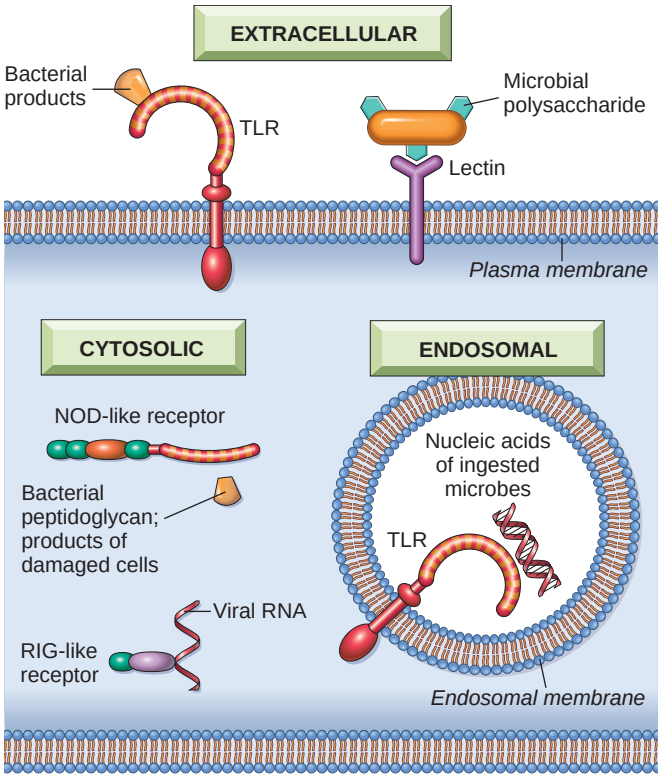


Figure 6.2 Cellular receptors for microbes and products of cell injury. Phagocytes, dendritic cells, and many types of epithelial cells express different classes of receptors that sense the presence of microbes and dead cells. Toll-like receptors *(TLRs)* located in different cellular compartments, as well as other cytoplasmic and plasma membrane receptors, recognize products of different classes of microbes. Major classes of innate immune receptors are TLRs, NOD-like receptors (NLRs) in the cytosol, C-type lectin receptors (CLRs), RIG-like receptors (RLRs) for viral nucleic acids, and cytosolic receptors for DNA (not shown). *RIG,* retinoic acid-inducible gene.

NOD-like receptors and the inflammasome. NOD-like receptors (NLRs) are cytosolic receptors named after the founding member NOD-2. They recognize a wide variety of substances, including products released from necrotic or damaged cells (e.g., uric acid and adenosine triphosphate [ATP]), loss of intracellular K^+ ions, and some microbial products. How this family of sensors is capable of detecting so many diverse signs of danger or damage is not known. Several of the NLRs signal via a cytosolic multiprotein complex called the *inflammasome,* which activates an enzyme (caspase-1) that cleaves a precursor form of the cytokine interleukin-1 (IL-1) to generate the biologically active form (Fig. 6.3). As discussed later, IL-1 is a mediator of inflammation that recruits leukocytes and induces fever. Gain-of-function mutations in NLRs and related proteins, and loss-of-function mutations in regulators of the inflammasome, result in periodic fever syndromes called *autoinflammatory syndromes* (to be distinguished from autoimmune diseases,

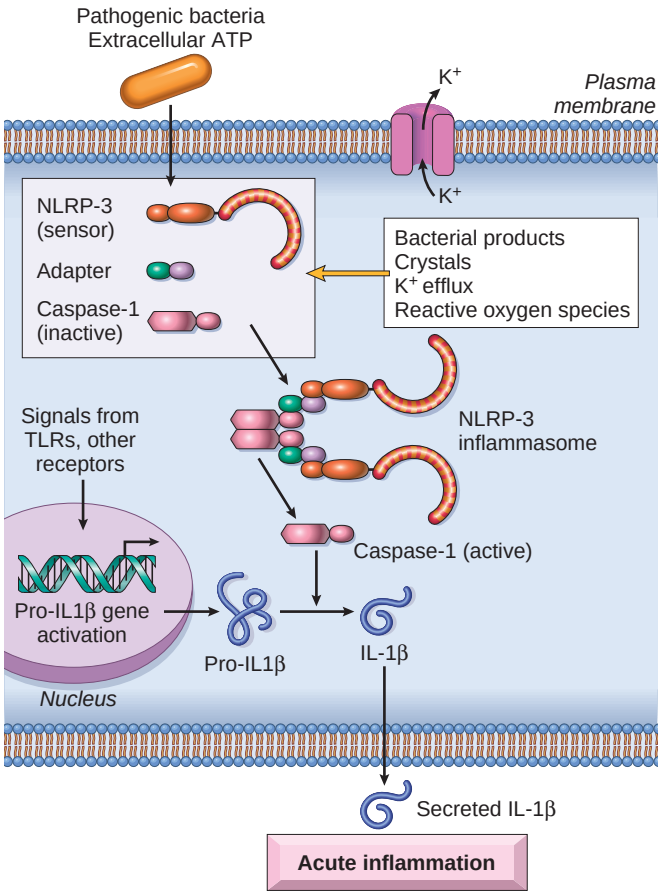


Figure 6.3 The inflammasome. The inflammasome is a protein complex that recognizes products of dead cells and some microbes and induces the secretion of biologically active interleukin 1. The inflammasome consists of a sensor protein (an example is the leucine-rich protein NLRP3), an adapter, and the enzyme caspase-1, which is converted from an inactive to an active form. *ATP*, Adenosine triphosphate; *IL*, interleukin; *TLR*, Toll-like receptor.

which result from T and B lymphocyte reactions against self antigens). The autoinflammatory syndromes respond very well to treatment with IL-1 antagonists. The inflammasome pathway may also play a role in many common disorders. For example, recognition of urate crystals by a class of NLRs underlies the inflammation associated with gout. These receptors are also capable of detecting lipids and cholesterol crystals that are deposited in abnormally large amounts in tissues, and the resulting inflammation appears to contribute to obesity-associated type 2 diabetes and atherosclerosis, respectively.

Other cellular receptors for microbial products. C-type lectin receptors (CLRs) expressed on the plasma membrane of macrophages and DCs detect fungal glycans and elicit inflammatory reactions to fungi. RIG-like receptors (RLRs), named after the founding member RIG-I (retinoic acid-inducible gene-I), are located in the cytosol of most cell types and detect nucleic acids of viruses that replicate in the cytoplasm of infected cells. These receptors stimulate the production of antiviral cytokines. Cytosolic receptors for microbial DNA, often derived from viruses in the cell, activate a pathway called STING (for stimulator of interferon genes), which leads to the production of the antiviral cytokine interferon-α. Excessive activation of the STING pathway causes systemic inflammatory disorders collectively called *interferonopathies*. Plasma membrane G protein–coupled receptors on neutrophils, macrophages, and most other types of leukocytes recognize short bacterial peptides containing *N*-formylmethionyl residues. Because all bacterial proteins and few mammalian proteins (only those synthesized within mitochondria) are initiated by *N*-formylmethionine, this receptor enables neutrophils to detect bacterial proteins and move toward their source (chemotaxis). Mannose receptors recognize microbial sugars (which often contain terminal mannose residues, unlike mammalian glycoproteins) and induce phagocytosis of the microbes.

Natural Killer Cells

The function of NK cells is to recognize and destroy severely stressed or abnormal cells, such as virus-infected cells and tumor cells. NK cells make up approximately 5% to 10% of peripheral blood lymphocytes. NK cells express CD16, a receptor for IgG Fc tails that confers on NK cells the ability to lyse IgG-coated target cells. This phenomenon is known as antibody-dependent cellular cytotoxicity (ADCC).

Killing of target cells by NK cells is regulated by signals from activating and inhibitory receptors (Fig. 6.4). There are many types of activating receptors, which recognize surface molecules that are induced by various kinds of stress, such as infection and DNA damage. Thus, these receptors enable NK cells to recognize damaged or infected cells. NK cell inhibitory receptors recognize self class I MHC molecules, which are expressed on all healthy cells. The inhibitory receptors prevent NK cells from killing normal cells. Virus infection or neoplastic transformation often enhances expression of ligands for activating receptors and at the same time reduces the expression of class I MHC molecules. As a result, when NK cells engage these abnormal cells the balance is tilted toward activation, and the infected or tumor cell is killed.

NK cells also secrete cytokines such as interferon-γ (IFN-γ), which activates macrophages to destroy ingested microbes, and thus NK cells provide an early defense against intracellular microbial infections. The activity of NK cells is regulated by many cytokines, including the interleukins IL-2, IL-15, and IL-12. IL-2 and IL-15 stimulate proliferation of NK cells, whereas IL-12 activates the killing of target cells and the secretion of IFN-γ.

Reactions of Innate Immunity

The innate immune system provides host defense by two main reactions.

- *Inflammation.* Cytokines and products of complement activation, as well as other mediators, are produced during innate immune reactions and trigger the vascular and cellular components of inflammation. The recruited leukocytes destroy microbes and ingest and eliminate damaged cells. The innate immune response also triggers the repair of damaged tissues. These processes are described in Chapter 3.
- *Antiviral defense.* Type I interferons produced in response to viruses act on infected and uninfected cells and activate

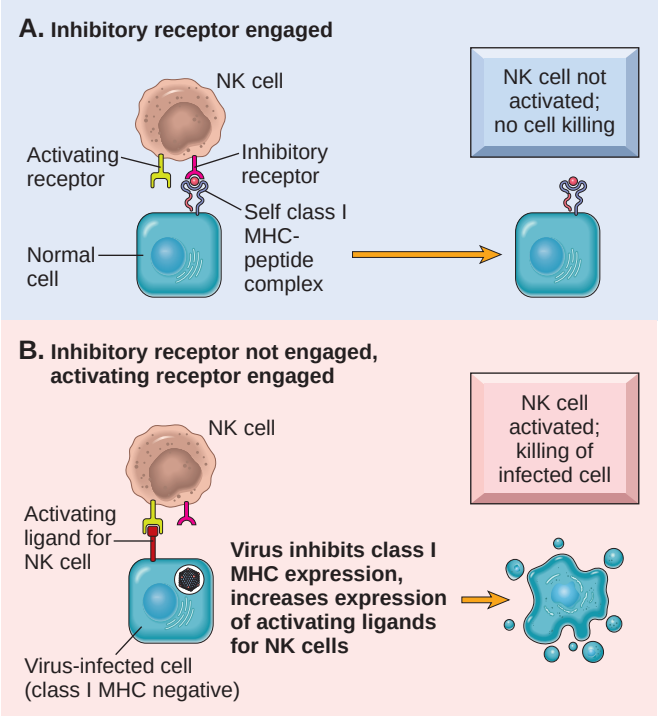


Figure 6.4 Activating and inhibitory receptors of natural killer *(NK)* cells. (A) Healthy cells express self class I major histocompatibility complex *(MHC)* molecules, which are recognized by inhibitory receptors, thus ensuring that NK cells do not attack normal cells. Note that healthy cells may express ligands for activating receptors *(not shown)* or may not express such ligands *(as shown)*, but they do not activate NK cells because they engage the inhibitory receptors. (B) In infected and stressed cells, class I MHC expression is reduced so that the inhibitory receptors are not engaged, and ligands for activating receptors are expressed. The result is that NK cells are activated and the infected cells are killed.

enzymes that degrade viral nucleic acids and inhibit viral replication, inducing what has been called an antiviral state. NK cells recognize virus-infected cells, as described above.

In addition to these direct defense functions, innate immunity produces the danger signals that stimulate the subsequent more powerful adaptive immune response. The nature of some of these signals is described later.

Innate immunity, unlike adaptive immunity, does not have memory or fine antigen specificity. It is estimated that innate immunity uses about 100 different receptors to recognize 1000 molecular patterns. In contrast, adaptive immunity uses two types of receptors (antibodies and T-cell receptors [TCRs], described later), each with millions of variations, to recognize millions of antigens.

Adaptive Immunity

The adaptive immune system consists of lymphocytes and their products, including antibodies. The lymphocytes of adaptive immunity use highly diverse receptors to recognize a vast array of foreign substances. In the remainder of this introductory section, we focus on lymphocytes and the reactions of the adaptive immune system.

There are two types of adaptive immunity: humoral immunity, which protects against extracellular microbes and their toxins, and cell-mediated (or cellular) immunity, which is responsible for defense against intracellular microbes and against cancers. Humoral immunity is mediated by B (bone marrow–derived) lymphocytes and their secreted products, *antibodies* (also called *immunoglobulins*, Ig), and cellular immunity is mediated by T (thymus-derived) lymphocytes. Both classes of lymphocytes express highly specific receptors for a wide variety of substances, which are called *antigens*.

Cells of the Adaptive Immune System

Although T and B lymphocytes and their subsets are morphologically unimpressive and appear quite similar to one another, they are actually remarkably heterogeneous and specialized in molecular properties and functions. The major classes of lymphocytes and their functions in adaptive immunity are illustrated in Fig. 6.5. Lymphocytes and other cells involved in immune responses are not fixed in particular tissues (as are cells in most of the organs of the body) but constantly circulate among lymphoid and other tissues via the blood and the lymphatic system. This feature promotes immune surveillance throughout the body and allows lymphocytes to home to any site of infection. In lymphoid organs, different classes of lymphocytes are anatomically segregated in such a way that they interact with one another only when stimulated to do so by encounters with antigens and other stimuli. Mature lymphocytes that have not encountered the antigen for which they are specific are said to be *naïve* (immunologically inexperienced). After they are activated by recognition of antigens and other signals described later, lymphocytes differentiate into *effector cells,* which perform the function of eliminating microbes, and *memory cells,* which live in a state of heightened awareness and are able to react rapidly and strongly to combat the microbe in case it returns. The process of lymphocyte differentiation into effector and memory cells is summarized later. We start with a consideration of the diversity of T and B lymphocytes.

Lymphocyte Diversity

Lymphocytes specific for a large number of antigens exist before exposure to antigen, and when an antigen appears it selectively activates the antigen-specific cells. This fundamental concept is called *clonal selection*. Lymphocytes of the same specificity are said to constitute a clone; all members of one clone express identical antigen receptors, which are different from the receptors in all other clones. There are about 10^{12} lymphocytes in a healthy adult, and it is estimated that there are 10^7 to 10^9 clones, each expressing receptors specific for a different antigen. It follows that the number of cells specific for any one antigen is very small, probably fewer than 1 in 100,000 lymphocytes. It is remarkable that so few cells with a particular specificity can accomplish the difficult task of combating various microbes; as discussed later, the immune system has developed many mechanisms for optimizing reactions to microbial antigens. It is also remarkable that the system is capable of producing so many receptors, far more than could be individually

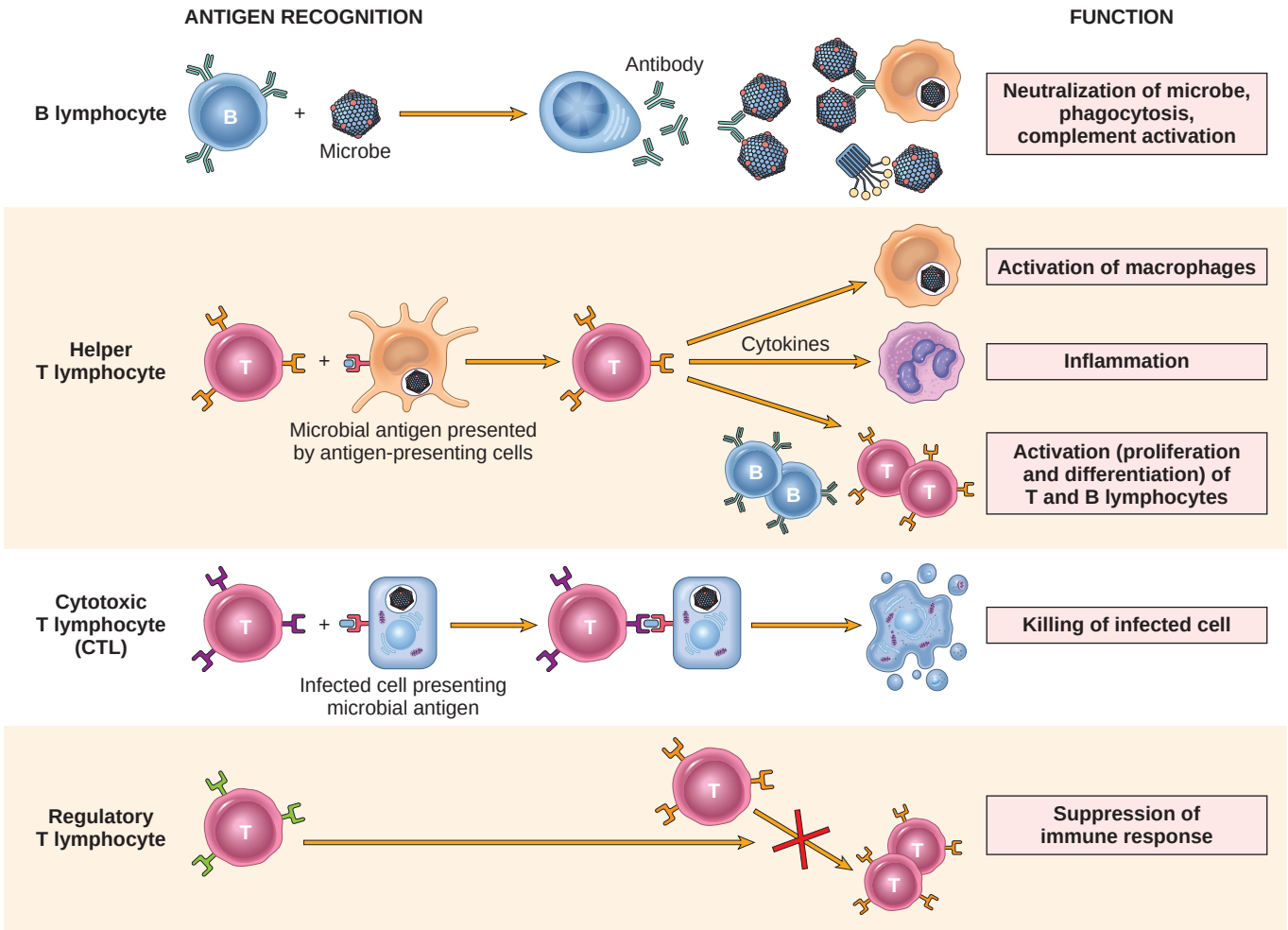


Figure 6.5 The principal classes of lymphocytes and their functions. B and T lymphocytes are the cells of adaptive immunity. Several other classes of lymphocytes have been identified, including NK-T cells and so-called innate lymphoid cells (ILCs); the functions of these cells are not as well established as those of B and T lymphocytes. NK cells are discussed earlier.

encoded in the genome. The mechanisms by which this happens are now well understood and have many interesting clinical implications.

Antigen receptor diversity is generated by somatic recombination of the genes that encode antigen receptors. All cells of the body, including lymphocyte progenitors, contain antigen receptor genes in the germline (inherited) configuration, in which the genes encoding these receptors consist of spatially separated segments that cannot be expressed as mRNAs. During lymphocyte maturation (in the thymus for T cells and the bone marrow for B cells), these gene segments are assembled by recombination, and DNA sequence variation is introduced at the sites where the gene segments are joined. This creates many different genes that can be transcribed and translated into antigen receptors with diverse amino acid sequences, particularly in the regions of the receptors that recognize and bind antigen. The enzyme in developing lymphocytes that mediates recombination of these gene segments is the product of *RAG-1* and *RAG-2* (recombination-activating genes); inherited defects in RAG proteins result in a failure to generate mature lymphocytes. It is important to note that germline antigen receptor genes are present in all cells in the body, but only T and B cells contain recombined (also called rearranged) antigen receptor genes (the TCR in T cells and Ig in B cells). Hence, the presence of recombined TCR or Ig genes, which can be demonstrated by molecular analysis, is a marker of T- or B-lineage cells. Furthermore, because each T or B cell and its clonal progeny have a unique DNA rearrangement (and hence a unique antigen receptor), it is possible to distinguish polyclonal (nonneoplastic) lymphocyte proliferations from monoclonal (neoplastic) lymphoid tumors by assessing the diversity of antigen receptor rearrangements within a population of lymphocytes. Thus, **assays that assess the clonality of antigen receptor gene rearrangements are useful in diagnosing lymphoid neoplasms** (Chapter 13).

T Lymphocytes

There are three major populations of T cells, which serve distinct functions. **Helper T lymphocytes stimulate B lymphocytes to make antibodies and activate other leukocytes (e.g., phagocytes) to destroy microbes; cytotoxic (killer) T lymphocytes (CTLs) kill infected cells; and**

regulatory T lymphocytes limit immune responses and prevent reactions against self antigens.

T lymphocytes develop in the thymus from precursors that arise from hematopoietic stem cells (HSCs). Mature T cells are found in the blood, where they constitute 60% to 70% of lymphocytes, and in T-cell zones of secondary lymphoid organs (described later). Each T cell recognizes a specific cell-bound antigen by means of an antigen-specific TCR. In approximately 95% of T cells, the TCR consists of a disulfide-linked heterodimer made up of an α and a β polypeptide chain (Fig. 6.6), each having a variable (antigen-binding) region and a constant region. **The αβ TCR recognizes peptide antigens that are bound to and presented by major histocompatibility complex (MHC) molecules on the surfaces of antigen-presenting cells (APCs).** By limiting the specificity of T cells for peptides displayed by cell surface MHC molecules, called MHC restriction, the immune system ensures that T cells see only cell-associated antigens (e.g., those derived from microbes in cells or from proteins ingested by cells).

Each TCR is noncovalently linked to six polypeptide chains, which form the CD3 complex and the ζ chain dimer (see Fig. 6.6). The CD3 and ζ proteins are invariant (i.e., identical) in all T cells. They are involved in the transduction of signals into the T cell that are triggered by binding of antigen to the TCR. Together with the TCR, these proteins form the TCR complex.

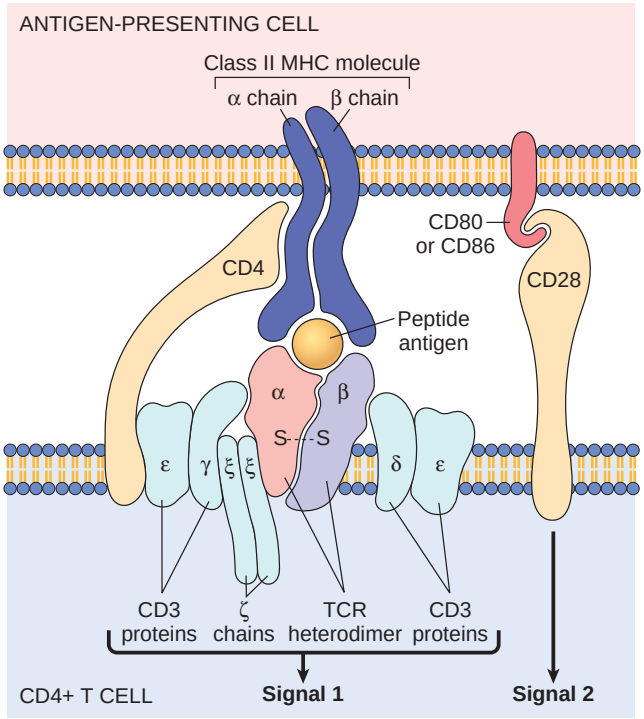


Figure 6.6 The T-cell receptor *(TCR)* complex and other molecules involved in T-cell activation. The TCR heterodimer, consisting of an α and a β chain, recognizes antigen (in the form of peptide-MHC complexes expressed on antigen-presenting cells, or APCs), and the linked CD3 complex and ζ chains initiate activating signals. CD4 and CD28 are also involved in T-cell activation. (Note that some T cells express CD8 and not CD4; these molecules serve analogous roles.) The sizes of the molecules are not drawn to scale. *MHC*, Major histocompatibility complex.

A small population of mature T cells expresses another type of TCR composed of γ and δ polypeptide chains. The γδ TCR recognizes peptides, lipids, and small molecules, without a requirement for display by MHC proteins. γδ T cells tend to aggregate at epithelial surfaces, such as the skin and mucosa of the gastrointestinal and urogenital tracts, suggesting that these cells are sentinels that protect against microbes that try to enter through epithelia. However, the functions of γδ T cells are not established. Another small subset of T cells expresses markers that are also found on NK cells; these cells are called NK-T cells. NK-T cells express a very limited diversity of TCRs, and they recognize glycolipids that are displayed by the MHC-like molecule CD1. The functions of NK-T cells are also not well defined.

In addition to CD3 and ζ proteins, T cells express several other proteins that assist the TCR complex in functional responses. These include CD4, CD8, CD28, and integrins. CD4 and CD8 are expressed on two mutually exclusive subsets of αβ T cells. Normally, approximately 60% of mature T cells are CD4+, and about 30% are CD8+. Most CD4+ T cells function as cytokine-secreting helper cells that assist macrophages and B lymphocytes to combat infections. Most CD8+ cells function as CTLs that destroy host cells harboring microbes. CD4 and CD8 serve as coreceptors in T-cell activation. During antigen recognition, CD4 molecules bind to class II MHC molecules that are displaying antigen (see Fig. 6.6); CD8 molecules bind to class I MHC molecules; and the CD4 or CD8 coreceptor initiates signals that are necessary for activation of the T cells. Because of this requirement for coreceptors, CD4+ helper T cells can recognize and respond to antigen displayed only by class II MHC molecules, whereas CD8+ cytotoxic T cells recognize cell-bound antigens only in association with class I MHC molecules; this segregation is described later. Integrins are adhesion molecules that promote the attachment of T-cells to APCs.

To respond, T cells have to not only recognize antigen-MHC complexes, but also have to receive additional signals provided by antigen-presenting cells. This process, in which CD28 plays an important role, is described later, when the steps in cell-mediated immune responses are summarized.

B Lymphocytes

B lymphocytes are the only cells in the body capable of producing antibodies, the mediators of humoral immunity. B lymphocytes develop from precursors in the bone marrow. Mature B cells constitute 10% to 20% of lymphocytes in the blood and are also present in peripheral lymphoid tissues such as lymph nodes, spleen, and mucosa-associated lymphoid tissues. B cells recognize antigen via the B-cell antigen receptor complex. Membrane-bound antibodies of the IgM and IgD isotypes, present on the surface of all mature, naïve B cells, are the antigen-binding component of the B-cell receptor (BCR) complex (Fig. 6.7). After stimulation by antigen and other signals (described later), B cells develop into *plasma cells,* veritable protein factories for producing antibodies, as well as long-lived memory cells. It is estimated that a single plasma cell can secrete hundreds to thousands of antibody molecules per second, a remarkable measure of the power of the immune response for combating pathogens.

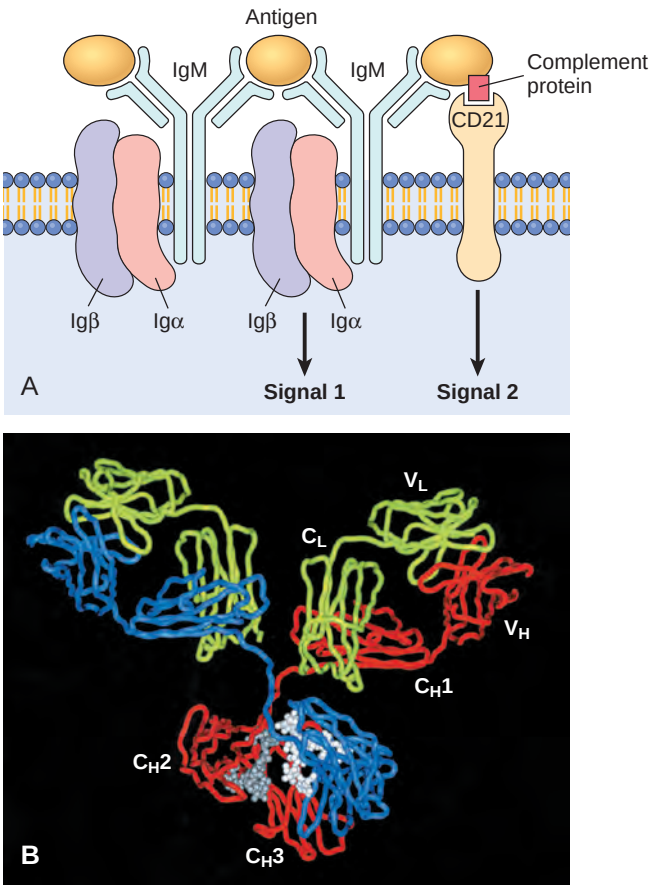


Figure 6.7 Structure of antibodies and the B-cell antigen receptor. (A) The B-cell antigen receptor complex is composed of membrane immunoglobulin M (*IgM;* or IgD, *not shown*), which recognizes antigens, and the associated signaling proteins Igα and Igβ. CD21 is a receptor for a complement component that also promotes B-cell activation. (B) Crystal structure of a secreted IgG molecule, showing the arrangement of the variable *(V)* and constant *(C)* regions of the heavy *(H)* and light *(L)* chains. (Courtesy Dr. Alex McPherson, University of California, Irvine, Calif.)

In addition to membrane Ig, the B-cell antigen receptor complex contains a heterodimer of two invariant proteins called Igα and Igβ. Similar to the CD3 and ζ proteins of the TCR complex, Igα (CD79a) and Igβ (CD79b) are essential for signal transduction in response to antigen recognition. B cells also express several other molecules that are essential for their responses. These include the type 2 complement receptor (CR2, or CD21), which recognizes complement products generated during innate immune responses to microbes, and CD40, which receives signals from helper T cells. CR2 is also used by Epstein-Barr virus (EBV) as a receptor to enter and infect B cells.

Dendritic Cells (DCs)

DCs (sometimes called interdigitating DCs) **are the most important antigen-presenting cells for initiating T-cell responses against protein antigens.** These cells have numerous fine cytoplasmic processes that resemble dendrites, from which they derive their name. Several features of DCs account for their key role in antigen presentation. First, these cells are located at the right place to capture antigens—under epithelia, the common site of entry of microbes and foreign antigens, and in the interstitia of all tissues, where antigens may be produced. Immature DCs within the epidermis are called *Langerhans cells*. Second, DCs express many receptors for capturing and responding to microbes (and other antigens), including TLRs and lectins. Third, in response to microbes, DCs are recruited to the T-cell zones of lymphoid organs, where they are ideally located to present antigens to naïve T cells. Fourth, DCs express high levels of MHC and other molecules needed for antigen presentation and T-cell activation.

A second type of cell with dendritic morphology is present in the germinal centers of lymphoid follicles in the spleen and lymph nodes and is called the *follicular dendritic cell (FDC)*. These cells bear Fc receptors for IgG and receptors for C3b and can trap antigen bound to antibodies or complement proteins. They play a role in humoral immune responses by presenting antigens to B cells in the germinal center, part of a process through which only B cells that express antibodies with high affinity for antigen survive and go on to mature into plasma cells or memory cells.

Macrophages

Macrophages are a part of the mononuclear phagocyte system; their origin, differentiation, and role in inflammation are discussed in Chapter 3. Here, their important functions in the induction and effector phases of adaptive immune responses are discussed.

- *Macrophages that have phagocytosed microbes and protein antigens process the antigens and present peptide fragments to T cells.* Thus, macrophages function as antigen-presenting cells in T-cell activation.
- *Macrophages are key effector cells in certain forms of cell-mediated immunity, the reaction that serves to eliminate intracellular microbes.* In this type of response, T cells activate macrophages and enhance their ability to kill ingested microbes (discussed later).
- *Macrophages also participate in the effector phase of humoral immunity.* Macrophages efficiently phagocytose and destroy microbes that are opsonized (coated) by IgG or C3b.

Tissues of the Immune System

The tissues of the immune system consist of the *primary* (also called *generative*, or *central*) lymphoid organs, in which T and B lymphocytes mature and become competent to respond to antigens, and the *secondary* (or *peripheral*) lymphoid organs, in which adaptive immune responses to microbes are initiated.

Primary Lymphoid Organs. **The principal primary lymphoid organs are the thymus, where T cells develop, and the bone marrow, the site of production of all other blood cells, including naïve B cells.** These organs are described in Chapter 13.

Secondary Lymphoid Organs. **The secondary lymphoid organs—lymph nodes, spleen, and the mucosal and cutaneous lymphoid tissues—are the tissues where adaptive immune responses occur**. Several features of these organs promote the generation of adaptive immunity—antigens are concentrated in these organs, naïve lymphocytes circulate through them searching for the antigens, and different

lymphocyte populations (such as T and B cells) are brought together when they need to interact.

- *Lymph nodes* are nodular aggregates of lymphoid tissues located along lymphatic channels throughout the body (Fig. 6.8). As lymph slowly suffuses through lymph nodes, antigen-presenting cells are positioned to recognize antigens (e.g., derived from microbes that may enter through epithelia into tissues and are carried in the lymph). In addition, DCs pick up and transport antigens of microbes from epithelia and tissues via lymphatic vessels to the lymph nodes. Thus, the antigens of microbes that enter through epithelia or colonize tissues become concentrated in draining lymph nodes. Because most foreign antigens enter through epithelia or are produced in tissues, lymph nodes are the site of generation of the majority of adaptive immune responses.
- The *spleen* is an abdominal organ that serves the same role in immune responses to blood-borne antigens as the lymph nodes do in responses to lymph-borne antigens. Blood entering the spleen flows through a network of sinusoids lined by macrophages and DCs. Blood-borne antigens are trapped in the spleen by these cells, which can then initiate adaptive immune responses to these antigens.
- The *cutaneous and mucosal lymphoid systems* are located under the epithelia of the skin and the gastrointestinal and respiratory tracts, respectively. They respond to antigens that enter through breaches in the epithelium. Pharyngeal tonsils and Peyer patches of the intestine are two anatomically defined mucosal lymphoid tissues. At any time, a large fraction of the body's lymphocytes are in the mucosal tissues (reflecting the large size of these tissues), and many of these are memory cells.

Within the secondary lymphoid organs, naïve T and B lymphocytes are segregated into different regions (see Fig. 6.8). In lymph nodes, the B cells are concentrated in discrete structures, called follicles, located around the periphery, or cortex, of each node. If the B cells in a follicle have recently responded to an antigen, this follicle may contain a central region called a germinal center. The T lymphocytes are concentrated in the paracortex, adjacent to the follicles. The follicles contain the FDCs that are involved in the activation of B cells, and the paracortex contains the DCs that present antigens to T lymphocytes. In the spleen, T lymphocytes are concentrated in periarteriolar lymphoid sheaths surrounding small arterioles, and B cells reside in follicles akin to those found in lymph nodes (the so-called splenic white pulp).

Lymphocyte Recirculation

Lymphocytes constantly recirculate between tissues and home to particular sites; naïve lymphocytes traverse the secondary lymphoid organs where immune responses are initiated, and effector lymphocytes migrate to sites of infection and inflammation. This process of lymphocyte recirculation is most important for T cells, because naïve T cells have to circulate through the secondary lymphoid organs where antigens are concentrated and effector T cells have to migrate to sites of infection to eliminate microbes. In contrast, plasma cells remain in lymphoid organs and the bone marrow and do not need to traffic to sites of infection because they secrete

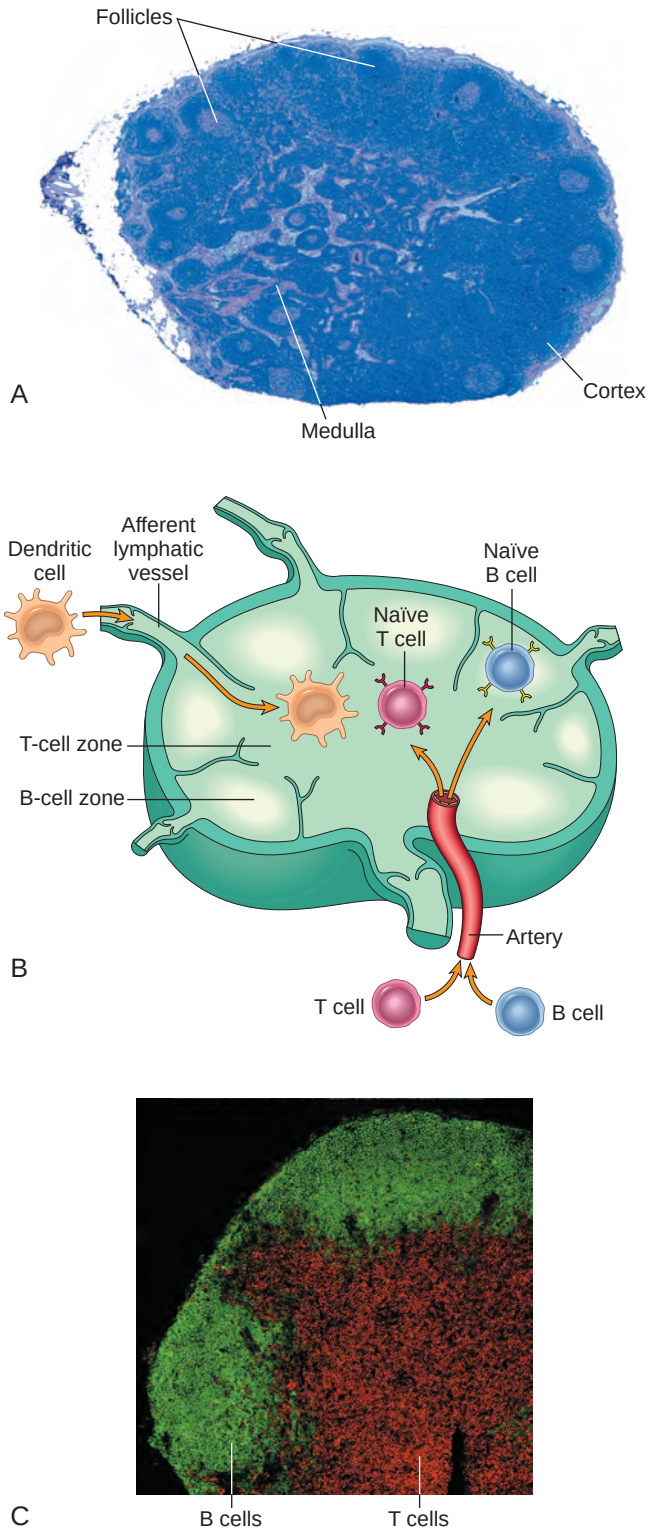


Figure 6.8 Morphology of a lymph node. (A) The histology of a lymph node, with an outer cortex containing follicles and an inner medulla. (B) The segregation of B cells and T cells in different regions of the lymph node, illustrated schematically. (C) The location of B cells (*stained green,* using the immunofluorescence technique) and T cells *(stained red)* in a lymph node. (Courtesy Drs. Kathryn Pape and Jennifer Walter, University of Minnesota School of Medicine, Minneapolis, Minn.)

antibodies that are carried through the blood to distant tissues.

Major Histocompatibility Complex Molecules: the Peptide Display System of Adaptive Immunity

The function of MHC molecules is to display peptide fragments of protein antigens for recognition by antigen-specific T cells. Because MHC molecules are fundamental to antigen recognition by T cells and are linked to many autoimmune diseases, it is important to briefly review their structure and function. MHC molecules were discovered as products of genes that evoke rejection of transplanted organs, and their name derives from their role in determining tissue compatibility between individuals. In humans, the MHC molecules are called *human leukocyte antigens* (HLA) because they were initially detected on leukocytes. The genes encoding HLA molecules are clustered on a small segment of chromosome 6 (Fig. 6.9). The HLA system is highly polymorphic; there are thousands of distinct MHC gene alleles in humans, and as a result each individual's HLA alleles differ from those inherited by most other individuals in the population. This, as we see subsequently, constitutes a formidable barrier in organ transplantation.

On the basis of their structure, cellular distribution, and function, MHC gene products are divided into two major classes.

- *Class I MHC molecules* are expressed on all nucleated cells and platelets. They are heterodimers consisting of a polymorphic α, or heavy, chain (44-kD) linked noncovalently to a smaller (12-kD) nonpolymorphic protein called β_2-microglobulin. The α chains are encoded by three genes, designated *HLA-A, HLA-B,* and *HLA-C,* that lie close to one another in the MHC locus (see Fig. 6.9). The extracellular region of the α chain is divided into three domains: α_1, α_2, and α_3. The α_1 and α_2 domains form a cleft, or groove, where peptides bind. The polymorphic amino acid residues line the sides and the base of the peptide-binding groove, explaining why different class I alleles bind different peptides.

  Class I MHC molecules display peptides that are derived from cytoplasmic proteins, including normal proteins and virus- and tumor-specific antigens, which are all recognized bound to class I MHC molecules by CD8+ T cells. Cytoplasmic proteins are degraded in proteasomes, and peptides are transported into the endoplasmic reticulum (ER), where they bind to newly synthesized class I molecules. Peptide-loaded MHC

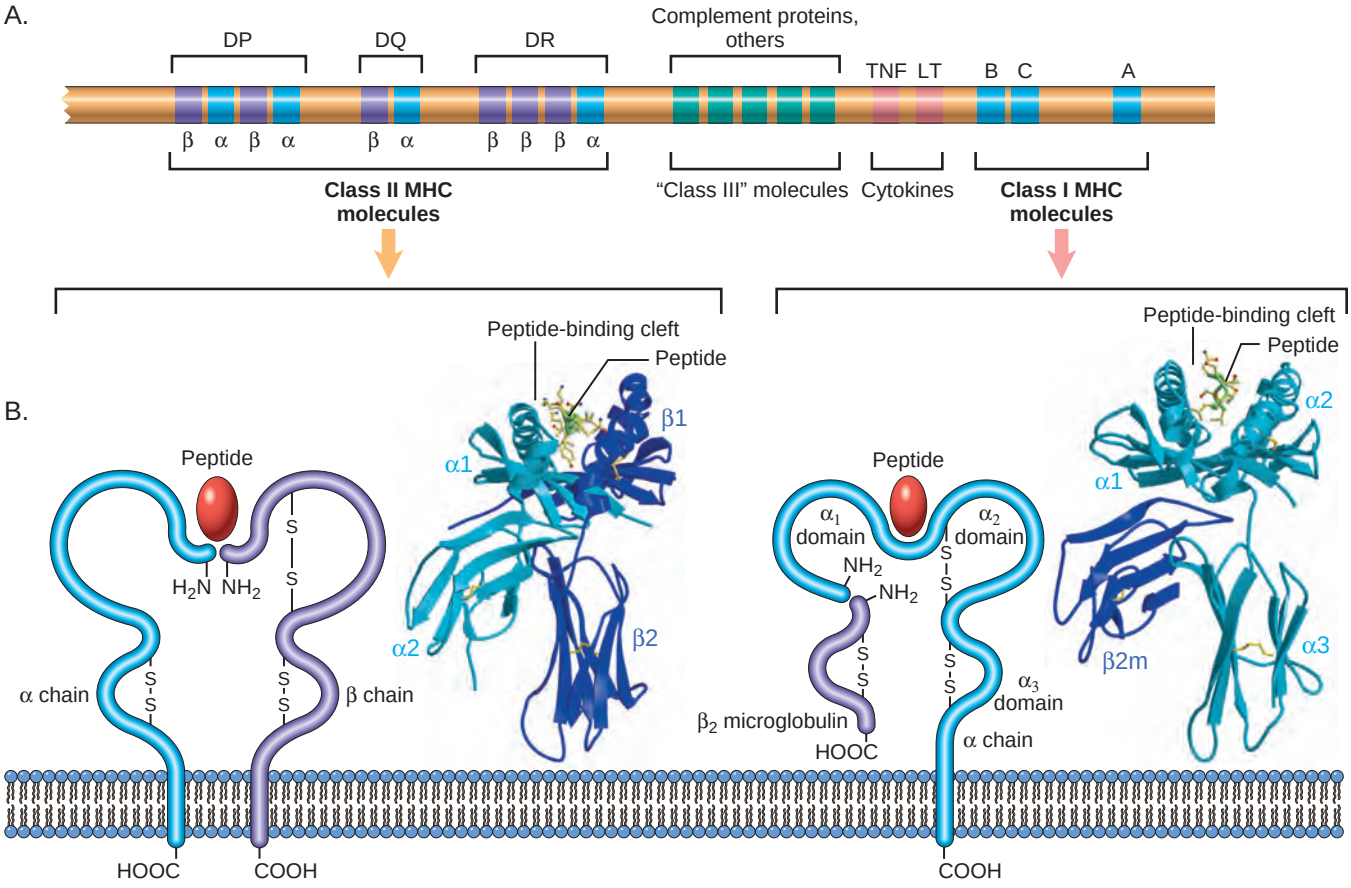


Figure 6.9 The human leukocyte antigen (HLA) complex and the structure of HLA molecules. (A) The location of genes in the HLA complex. The relative locations, sizes, and distances between genes are not to scale. Genes that encode several proteins involved in antigen processing (the TAP transporter, components of the proteasome, and HLA-DM) are located in the class II region *(not shown).* (B) Schematic diagrams and crystal structures of class I and class II HLA molecules. (Crystal structures are courtesy Dr. P. Bjorkman, California Institute of Technology, Pasadena, Calif.)

molecules then associate with β_2-microglobulin to form a stable complex, which is transported to the cell surface. The nonpolymorphic α_3 domain of class I MHC molecules has a binding site for CD8, and therefore the peptide-class I complexes are recognized by CD8+ T cells, which function as CTLs. In this interaction, the TCR recognizes the MHC-peptide complex, and the CD8 molecule, acting as a coreceptor, binds to the class I heavy chain. Since CD8+ T cells recognize peptides only if presented as a complex with class I MHC molecules, CD8+ T cells are said to be class I MHC-restricted. Because important functions of CD8+ CTLs include the elimination of viruses, which may infect any nucleated cell, and killing of tumor cells, which may arise from any nucleated cell, it makes good sense that all nucleated cells express class I MHC molecules and can be surveyed by CD8+ T cells.

- *Class II MHC molecule*s are encoded in a region called *HLA-D*, which has three subregions: *HLA-DP, HLA-DQ,* and *HLA-DR*. Each class II molecule is a heterodimer consisting of a noncovalently associated α chain and β chain, both of which are polymorphic. The extracellular portions of the α and β chains both have two domains designated α_1 and α_2, and β_1 and β_2. The crystal structure of class II molecules has revealed that, similar to class I molecules, they have peptide-binding clefts facing outward (see Fig. 6.9). This cleft is formed by an interaction of the α_1 and β_1 domains, and it is in this portion that most class II alleles differ. Thus, as with class I molecules, polymorphism of class II molecules is associated with differential binding of antigenic peptides.

  Class II MHC molecules present antigens derived from extracellular microbes and proteins following their internalization into endosomes or lysosomes. Here, the internalized proteins are proteolytically digested, producing peptides that then associate with class II heterodimers in the vesicles, from which they are transported to the cell surface as stable peptide-MHC complexes. The class II β_2 domain has a binding site for CD4, and therefore, the class II-peptide complex is recognized by CD4+ T cells, which function as helper cells. Because CD4+ T cells can recognize antigens only in the context of self class II molecules, they are referred to as class II MHC-restricted. In contrast to class I molecules, class II molecules are mainly expressed on cells that present ingested antigens and respond to T-cell help (macrophages, B lymphocytes, and DCs).

The combination of HLA alleles in each individual is called the HLA haplotype. Any given individual inherits one set of HLA genes from each parent and (assuming the parents are unrelated) typically expresses two different molecules for every locus. Because of the polymorphism of the HLA genes, virtually innumerable combinations of molecules exist in the population, and each individual expresses an MHC profile on his or her cell surface that is different from the haplotypes of most other individuals. It is believed that this degree of polymorphism evolved to ensure that at least some individuals in a species would be able to display any microbial peptide and thus provide protection against any infection. This polymorphism also means that no two individuals (other than identical twins) are likely to express the same MHC molecules, and therefore grafts exchanged between these individuals are recognized as foreign and attacked by the immune system. Because each haplotype is inherited as a block, and there are two sets of genes from each parent, the chance that siblings will have the same MHC is 1 in 4. This is why siblings are screened first as potential donors for patients in need of a kidney or hematopoietic stem cell transplant.

MHC molecules play several key roles in regulating T cell–mediated immune responses. First, because different antigenic peptides bind to different MHC molecules, it follows that an individual mounts an immune response against a protein antigen only if he or she inherits an MHC variant that can bind peptides derived from the antigen and present it to T cells. The consequences of inheriting a given MHC (e.g., class II) variant depend on the nature of the antigen bound by the class II molecule. For example, if the antigen is a peptide from ragweed pollen, the individual who expresses class II molecules capable of binding the antigen would be genetically prone to allergic reactions against ragweed. In contrast, an inherited capacity to bind a bacterial peptide may provide resistance to the infection by evoking a protective antibody response. Second, by segregating cytoplasmic and internalized antigens, MHC molecules ensure that the correct immune response is mounted against different microbes—CTL-mediated killing of cells harboring cytoplasmic microbes and tumor antigens, and helper T cell–mediated antibody production and macrophage activation to combat extracellular and phagocytosed microbes.

Interest in HLA molecules was spurred by the realization, in the 1960s and 1970s, that a number of autoimmune and other diseases are associated with the inheritance of particular HLA alleles. These associations are discussed when the pathogenesis of autoimmune diseases is considered later in the chapter.

Cytokines: Messenger Molecules of the Immune System

The induction and regulation of immune responses involve multiple interactions among lymphocytes, DCs, macrophages, other inflammatory cells (e.g., neutrophils), and endothelial cells. Some of these interactions depend on cell-to-cell contact; however, **many functions of leukocytes are stimulated and regulated by secreted proteins called cytokines.** Molecularly defined cytokines are called interleukins because they mediate communications between leukocytes (although many also act on cells other than leukocytes). Most cytokines have a wide spectrum of effects, and some are produced by several different cell types. The majority of these cytokines act on the cells that produce them (autocrine actions) or on neighboring cells (paracrine) and rarely at a distance (endocrine).

Cytokines contribute to different types of immune responses.

- In innate immune responses, cytokines are produced rapidly after encounter with microbes and other stimuli, and they function to induce inflammation and inhibit virus replication. These cytokines include TNF, IL-1, IL-12, type I IFNs, IFN-γ, and chemokines (Chapter 3). Their major sources are macrophages, DCs, ILCs, and NK cells, but endothelial and epithelial cells can also produce them.
- In adaptive immune responses, cytokines are produced principally by CD4+ T lymphocytes activated by antigen

and other signals, and they function to promote lymphocyte proliferation and differentiation and to activate effector cells. The main ones in this group are IL-2, IL-4, IL-5, IL-17, and IFN-γ; their roles in immune responses are described later. Some cytokines serve mainly to limit and terminate immune responses; these include TGF-β and IL-10.

- Some cytokines stimulate hematopoiesis and are called colony-stimulating factors (CSFs) because they are assayed by their ability to stimulate formation of blood cell colonies from bone marrow progenitors (Chapter 13). Their functions are to increase leukocyte production during immune and inflammatory responses, both to increase their numbers and to replace leukocytes that die during such responses. They are produced by marrow stromal cells, T lymphocytes, macrophages, and other cells. Examples include GM-CSF and other CSFs, and IL-3.

The knowledge gained about cytokines has numerous practical therapeutic applications. Inhibiting cytokine production or actions is an approach for controlling the harmful effects of inflammation and tissue-damaging immune reactions. Patients with rheumatoid arthritis often show dramatic responses to TNF antagonists, an elegant example of rationally designed and molecularly targeted therapy. Many other cytokine antagonists are now approved for the treatment of various inflammatory disorders. Conversely, administration of cytokines is used to boost reactions that are normally dependent on these proteins, such as hematopoiesis and defense against some viruses. An important therapeutic application of cytokines is to mobilize hematopoietic stem cells from bone marrow to peripheral blood, from which they can be collected for stem cell transplantation.

Overview of Lymphocyte Activation and Immune Responses

All adaptive immune responses develop in steps, consisting of: antigen recognition; activation of specific lymphocytes to proliferate and differentiate into effector and memory cells; elimination of the antigen; and decline of the response, with memory cells being the long-lived survivors. The major events in each step are summarized next; these general principles apply to protective responses against microbes as well as pathologic responses that injure the host.

Display and Recognition of Antigens

Microbes and other foreign antigens can enter the body anywhere. It is obviously impossible for lymphocytes to effectively patrol every possible portal of antigen entry, because there are not enough antigen-specific lymphocytes to constantly cover all of this "terrain." To overcome this problem, antigens are captured and concentrated in secondary lymphoid organs through which naïve lymphocytes circulate, thus increasing the likelihood of a lymphocyte finding antigens it can recognize. Microbes and their protein antigens are captured by DCs that are resident in epithelia and tissues. These cells carry their antigenic cargo to draining lymph nodes (Fig. 6.10). Here the antigens are processed and displayed complexed with MHC molecules on the cell surface, where the antigens are recognized by T cells.

B lymphocytes use their antigen receptors (membrane-bound antibody molecules) to recognize antigens of many different chemical types, including proteins, polysaccharides, and lipids.

Even before the antigens of a microbe are recognized by T and B lymphocytes, the microbe elicits an immune response through pattern recognition receptors expressed on innate immune cells; this is the first line of defense that also serves to activate adaptive immunity. In the case of immunization with a protein antigen, microbial mimics, called adjuvants, are given with the antigen, and these stimulate innate immune responses. During the innate response, the microbe or adjuvant activates antigen-presenting cells to express molecules called *costimulators* and to secrete cytokines that stimulate the proliferation and differentiation of T lymphocytes. The principal costimulators for T cells are the B7 proteins (CD80 and CD86) that are expressed on antigen-presenting cells and are recognized by the CD28 receptor on naïve T cells. Thus, antigen ("signal 1") and costimulatory molecules produced during innate immune responses to microbes ("signal 2") function cooperatively to activate antigen-specific lymphocytes (see Fig. 6.6). The requirement for microbe-triggered signal 2 ensures that the adaptive immune response is induced by microbes and not by harmless substances. In immune responses to tumors and transplants, "signal 2" may be provided by substances released from necrotic cells (the "damage-associated molecular patterns" mentioned earlier).

The reactions and functions of T and B lymphocytes differ in important ways and are best considered separately even though both may be activated concurrently in an immune response.

Cell-Mediated Immunity: Activation of T Lymphocytes and Elimination of Intracellular Microbes

Naïve T lymphocytes are activated by antigen and costimulators in peripheral lymphoid organs, and proliferate and differentiate into effector cells that migrate to any site where microbial antigens are present (see Fig. 6.10). One of the earliest responses of CD4+ helper T cells is secretion of the cytokine IL-2 and expression of high-affinity receptors for IL-2. This creates an autocrine loop wherein IL-2 acts as a growth factor that stimulates T-cell proliferation, leading to an increase in the number of antigen-specific lymphocytes. The functions of helper T cells are mediated by the combined actions of CD40-ligand (CD40L) and cytokines. When CD4+ helper T cells recognize antigens being displayed by macrophages or B lymphocytes, the T cells express CD40L, which engages CD40 on the macrophages or B cells and activates these cells.

Some of the activated CD4+ T cells differentiate into effector cells that secrete distinct sets of cytokines and perform different functions (Fig. 6.11). Cells of the Th1 subset secrete the cytokine IFN-γ, which is a potent macrophage activator. The combination of CD40- and IFN-γ-mediated activation results in "classical" macrophage activation (Chapter 3), leading to the production of microbicidal substances in macrophages and the destruction of ingested microbes. Th2 cells produce IL-4, which stimulates B cells to differentiate into IgE-secreting plasma cells, and IL-5, which stimulates the production of eosinophils in the marrow and activates eosinophils at sites of immune responses. Eosinophils and mast cells bind to IgE-coated microbes such as helminthic parasites, and function to eliminate helminths. Th2 cells also induce the "alternative"

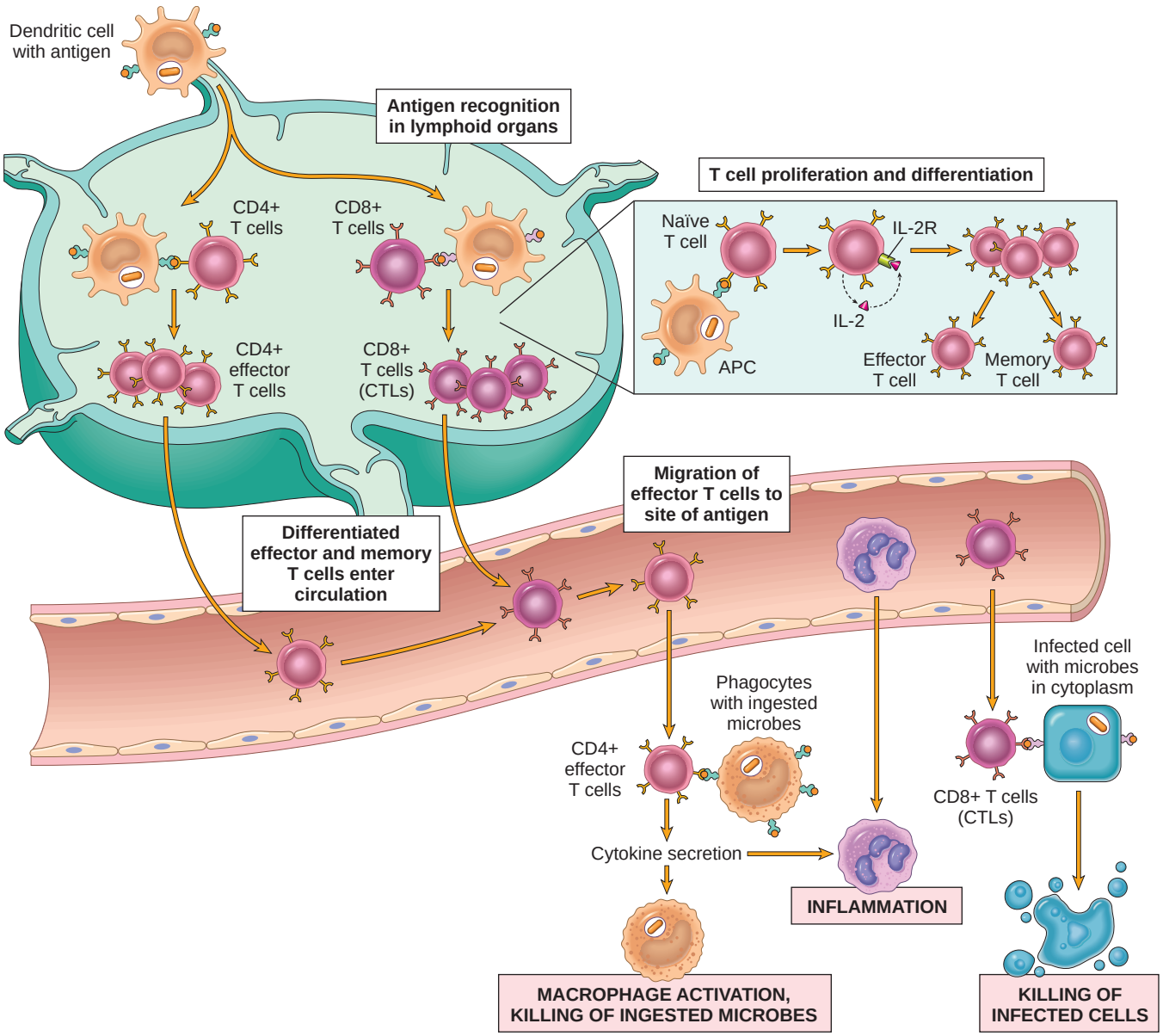


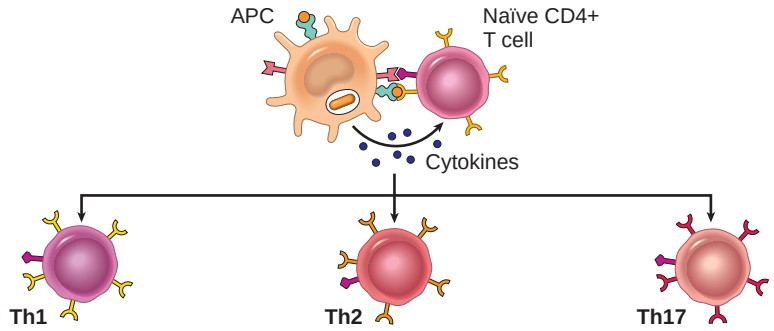
Figure 6.10 Cell-mediated immunity. Dendritic cells capture microbial antigens from epithelia and tissues and transport the antigens to lymph nodes. During this process, the dendritic cells mature, and express high levels of MHC molecules and costimulators. Naïve T cells recognize MHC-associated peptide antigens displayed on dendritic cells. The T cells are activated to proliferate and to differentiate into effector and memory cells, which migrate to sites of infection and serve various functions in cell-mediated immunity. CD4+ effector T cells of the Th1 subset recognize the antigens of microbes ingested by phagocytes and activate the phagocytes to kill the microbes; other subsets of effector cells enhance leukocyte recruitment and stimulate different types of immune responses. CD8+ cytotoxic T lymphocytes *(CTLs)* kill infected cells harboring microbes in the cytoplasm. Some activated T cells remain in the lymphoid organs and help B cells to produce antibodies, and some T cells differentiate into long-lived memory cells (not shown). *APC,* Antigen-presenting cell.

pathway of macrophage activation, which is associated with tissue repair and fibrosis (Chapter 3). Th17 cells, so called because the signature cytokine of these cells is IL-17, recruit neutrophils and monocytes, which destroy extracellular bacteria and fungi and are involved in some inflammatory diseases.

Activated CD8+ T lymphocytes differentiate into CTLs that kill cells harboring microbes in the cytoplasm. By destroying the infected cells, CTLs eliminate the reservoirs of infection. CTLs also kill tumor cells by recognizing tumor-specific antigens derived from mutated or abnormal cytoplasmic proteins.

Humoral Immunity: Activation of B Lymphocytes and Elimination of Extracellular Microbes

Upon activation, B lymphocytes proliferate and then differentiate into plasma cells that secrete different classes of antibodies with distinct functions (Fig. 6.12). Antibody responses to most protein antigens require T cell help and are said to be T-dependent. In these responses, B cells that recognize protein antigens by their Ig receptors endocytose these antigens into vesicles, degrade them, and display peptides bound to class II MHC molecules for recognition by helper T cells. The helper T cells are activated and express CD40L



	Th1	Th2	Th17
Cytokines produced	IFN-γ	IL-4, IL-5, IL-13	IL-17, IL-22
Cytokines that induce this subset	IFN-γ, IL-12	IL-4	TGF-β, IL-6, IL-1, IL-23
Immunological reactions triggered	Macrophage activation	Stimulation of IgE production, activation of mast cells and eosinophils	Recruitment of neutrophils, monocytes
Host defense against	Intracellular microbes	Helminthic parasites	Extracellular bacteria, fungi
Role in disease	Immune-mediated chronic inflammatory diseases (often autoimmune)	Allergies	Immune-mediated chronic inflammatory diseases (often autoimmune)

Figure 6.11 Subsets of helper T *(Th)* cells. In response to stimuli (mainly cytokines) present at the time of antigen recognition, naïve CD4+ T cells may differentiate into populations of effector cells that produce distinct sets of cytokines and perform different functions. The dominant immune reactions elicited by each subset, and its role in host defense and immunologic diseases, are summarized. These populations may be capable of converting from one to another. Some activated T cells produce multiple cytokines and do not fall into a distinct subset.

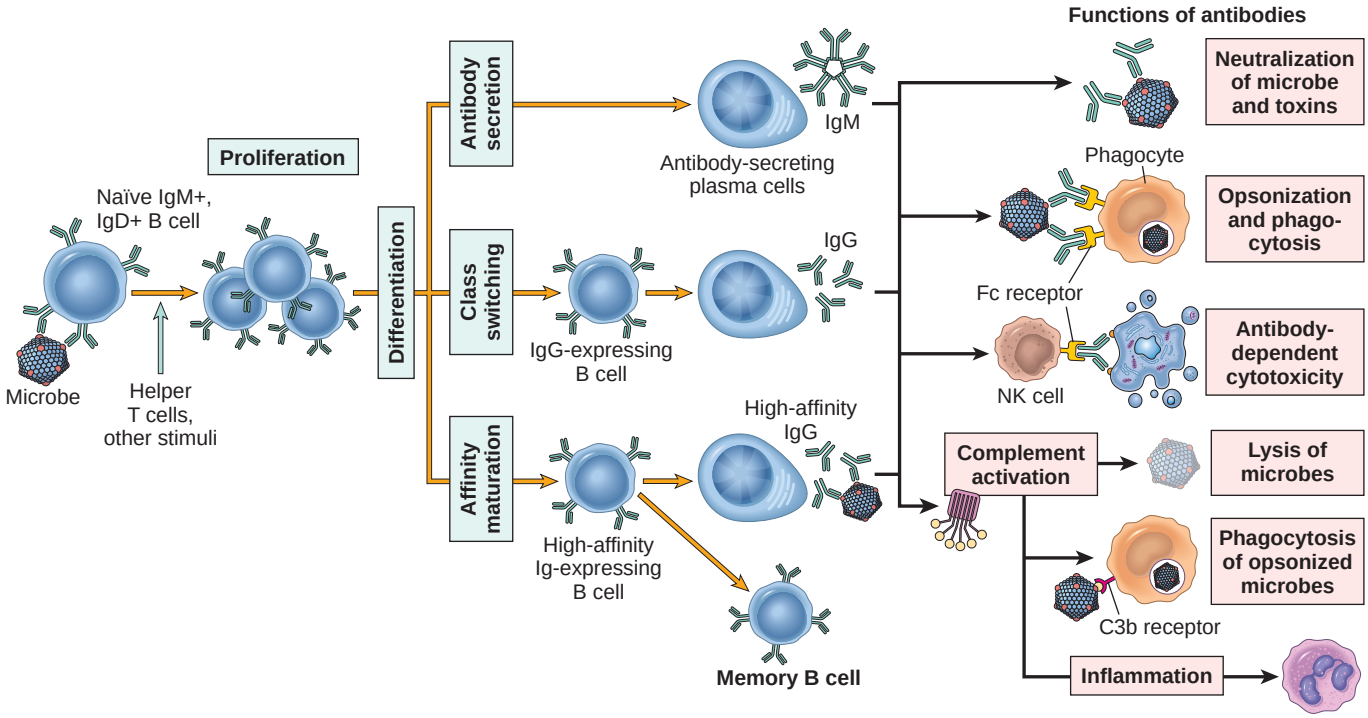


Figure 6.12 Humoral immunity. Naïve B lymphocytes recognize antigens, and under the influence of helper T cells and other stimuli *(not shown)*, the B cells are activated to proliferate and to differentiate into antibody-secreting plasma cells. Some of the activated B cells undergo heavy-chain class switching and affinity maturation, and some become long-lived memory cells. Antibodies of different heavy-chain classes (isotypes) perform different effector functions, shown on the right. Note that the antibodies shown are IgG; these and IgM activate complement; and the specialized functions of IgA (mucosal immunity) and IgE (mast cell and eosinophil activation) are not shown.

and secrete cytokines, which work together to stimulate the B cells. Many polysaccharide and lipid antigens cannot be recognized by T cells (because they cannot bind to MHC molecules) but have multiple identical antigenic determinants (epitopes) that are able to engage many antigen receptor molecules on each B cell and initiate the process of B-cell activation; these responses are said to be T-independent. T-independent responses are relatively simple, whereas T-dependent responses show features such as Ig isotype switching and affinity maturation (described later), which require T cell help and lead to responses that are more varied and effective.

Each plasma cell is derived from an antigen-stimulated B cell and secretes antibodies that recognize the same antigen that was bound by the BCR to initiate the response. Polysaccharides and lipids stimulate secretion mainly of IgM antibody. Protein antigens, by virtue of CD40L- and cytokine-mediated helper T-cell actions, induce the production of antibodies of different classes, or isotypes (IgG, IgA, IgE), a process called *isotype switching*. Helper T cells also stimulate the production of antibodies with high affinities for the antigen. This process, called *affinity maturation,* improves the quality of the humoral immune response. These two processes are initiated when activated B cells that receive signals from helper T cells during responses to protein antigens migrate into follicles and begin to proliferate to form germinal centers, which are the major sites of isotype switching and affinity maturation. The helper T-cells that stimulate these processes in B lymphocytes also migrate to and reside in the germinal centers and are called T follicular helper (TFH) cells.

The humoral immune response combats microbes in many ways (see Fig. 6.12). Antibodies bind to microbes and prevent them from infecting cells, thus neutralizing the microbes. IgG antibodies coat (opsonize) microbes and target them for phagocytosis, since phagocytes (neutrophils and macrophages) express receptors for the Fc tails of IgG. IgG and IgM activate the complement system by the classical pathway, and complement products promote phagocytosis and destruction of microbes. Some antibodies serve special roles at particular anatomic sites. IgA is secreted from mucosal epithelia and neutralizes microbes in the lumens of the respiratory and gastrointestinal tracts (and other mucosal tissues). IgG is actively transported across the placenta and protects the newborn until the immune system becomes mature. IgE and eosinophils cooperate to kill parasites, mainly by release of eosinophil granule contents that are toxic to the worms. As mentioned earlier, Th2 cytokines stimulate the production of IgE and activate eosinophils, and thus the response to helminths is orchestrated by Th2 cells.

Most circulating IgG antibodies have half-lives of about 3 weeks. Some antibody-secreting plasma cells, particularly those that are generated in germinal centers, migrate to the bone marrow and take up residence for months or even years, continuously producing antibodies during this time.

Decline of Immune Responses and Immunologic Memory

The majority of effector lymphocytes induced by an infectious pathogen die by apoptosis after the microbe is eliminated, thus returning the immune system to its resting state. The initial activation of lymphocytes also generates long-lived memory cells, which may survive for many years after the infection. Memory cells are an expanded pool of antigen-specific lymphocytes (more numerous than the naïve cells specific for any antigen that are present before encounter with that antigen), and they respond faster and more effectively when reexposed to the antigen than do naïve cells. Generation of memory cells underlies the effectiveness of vaccination.

KEY CONCEPTS

THE NORMAL IMMUNE RESPONSE: OVERVIEW OF CELLS, TISSUES, RECEPTORS, AND MEDIATORS

- The innate immune system uses several families of receptors (e.g., Toll-like receptors) to recognize molecules present in and shared by various types of microbes and produced by damaged cells.
- Lymphocytes are the mediators of adaptive immunity and the only cells that produce specific and diverse receptors for antigens.
- T (thymus-derived) lymphocytes express antigen receptors called TCRs that recognize peptide fragments of protein antigens that are displayed by MHC molecules on the surface of antigen-presenting cells.
- B (bone marrow–derived) lymphocytes express membrane-bound antibodies that recognize a wide variety of antigens. B cells are activated to become plasma cells, which secrete antibodies.
- NK cells destroy cells that are infected by some microbes, or are stressed and damaged beyond repair. NK cells express inhibitory receptors that recognize MHC molecules that are normally expressed on healthy cells, and are thus prevented from killing normal cells.
- Antigen-presenting cells (APCs) capture microbes and other antigens, transport them to secondary lymphoid organs, and display them for recognition by lymphocytes. The most efficient APCs are dendritic cells (DCs), which are present in epithelia and most other tissues.
- The cells of the immune system are organized in tissues, some of which are the sites of production of mature lymphocytes (the primary, or generative, lymphoid organs, the bone marrow and thymus), and others are the sites of immune responses (the secondary, or peripheral, lymphoid organs, including lymph nodes, spleen, and mucosal lymphoid tissues).
- The early reaction to microbes is mediated by the mechanisms of innate immunity, which are always ready to respond to microbes. These mechanisms include epithelial barriers, phagocytes, innate lymphoid cells (ILCs), NK cells, and certain plasma proteins (e.g., the complement system). Innate immune reactions often manifest as inflammation. Innate immunity, unlike adaptive immunity, does not have fine antigen specificity or memory.
- The defense reactions of adaptive immunity develop over several days, but are more potent and specialized.
- Microbes and other foreign antigens are captured by DCs and transported to lymph nodes, where the antigens are recognized by naïve lymphocytes. The lymphocytes are activated to proliferate and differentiate into effector and memory cells.
- Cell-mediated immunity is the reaction of T lymphocytes, designed to combat cell-associated microbes (e.g., phagocytosed microbes and microbes in the cytoplasm of infected cells). Humoral immunity is mediated by antibodies and is effective against extracellular microbes (in the circulation and mucosal lumens).
- CD4+ helper T cells help B cells to make antibodies, activate macrophages to destroy ingested microbes, stimulate recruitment

of leukocytes, and regulate all immune responses to protein antigens. The functions of CD4+ T cells are mediated by secreted proteins called cytokines. CD8+ cytotoxic T lymphocytes kill cells that express antigens in the cytoplasm that are seen as foreign (e.g., virus-infected and tumor cells) and can also produce cytokines.
- Antibodies secreted by plasma cells neutralize microbes and block their infectivity, and they promote the phagocytosis and destruction of pathogens. Antibodies also confer passive immunity to neonates.

The brief outline of basic immunology presented here provides a foundation for considering the diseases of the immune system. We first discuss the immune reactions that cause injury, called hypersensitivity reactions, and then disorders caused by the failure of tolerance to self antigens, called autoimmune disorders, and the rejection of transplants. This is followed by diseases caused by a defective immune system, called immunodeficiency diseases. We close with a consideration of amyloidosis, a disorder that is often associated with immune and inflammatory diseases.

HYPERSENSITIVITY: IMMUNOLOGICALLY MEDIATED TISSUE INJURY

Injurious immune reactions, called hypersensitivity, are responsible for the pathology associated with immunologic diseases. This term arose from the idea that individuals who have been previously exposed to an antigen manifest detectable reactions to that antigen and are therefore said to be sensitized. Hypersensitivity implies an excessive or harmful reaction to an antigen. There are several important general features of hypersensitivity disorders.

- Hypersensitivity reactions can be elicited by exogenous environmental antigens (microbial and nonmicrobial) or endogenous self antigens. Exogenous antigens include those in dust, pollen, food, drugs, microbes, and various chemicals. The immune responses against such exogenous antigens may take several forms, ranging from annoying but trivial discomforts, such as itching of the skin, to potentially fatal diseases, such as anaphylaxis. Some of the most common reactions to environmental antigens cause the group of diseases known as allergy. Immune responses against self, or autologous, antigens, cause autoimmune diseases.
- Hypersensitivity usually results from an imbalance between the effector mechanisms of immune responses and the control mechanisms that serve to limit such responses. In fact, in many hypersensitivity diseases, it is suspected that the underlying cause is a failure of normal regulation. We will return to this concept when we consider autoimmunity.
- The development of hypersensitivity diseases (both allergic and autoimmune) is often associated with the inheritance of particular susceptibility genes. HLA genes and many non-HLA genes have been implicated in different diseases; specific examples will be described in the context of the diseases.
- The mechanisms of tissue injury in hypersensitivity reactions are the same as the effector mechanisms of defense against infectious pathogens. The problem in hypersensitivity is that these reactions are poorly controlled, excessive, or misdirected (e.g., against normally harmless environmental and self antigens).

Classification of Hypersensitivity Reactions

Hypersensitivity reactions can be classified on the basis of the underlying immunologic mechanism (Table 6.1). This classification is of value in distinguishing the manner in which an immune response causes tissue injury and disease, and the accompanying pathologic and clinical manifestations. However, it is now increasingly recognized that multiple mechanisms may be operative in any one disease. The main types of hypersensitivity reactions are as follows:

- In *immediate hypersensitivity (type I hypersensitivity)*, the injury is caused by Th2 cells, IgE antibodies, and mast cells and other leukocytes. Mast cells release mediators

Table 6.1 Mechanisms of Hypersensitivity Reactions

Type	Immune Mechanisms	Histopathologic Lesions	Prototypical Disorders
Immediate (type I) hypersensitivity	Production of IgE antibody → immediate release of vasoactive amines and other mediators from mast cells; later recruitment of inflammatory cells	Vascular dilation, edema, smooth muscle contraction, mucus production, tissue injury, inflammation	Anaphylaxis; allergies; bronchial asthma (atopic forms)
Antibody-mediated (type II) hypersensitivity	Production of IgG, IgM → binds to antigen on target cell or tissue → phagocytosis or lysis of target cell by activated complement or Fc receptors; recruitment of leukocytes	Phagocytosis and lysis of cells; inflammation; in some diseases, functional derangements without cell or tissue injury	Autoimmune hemolytic anemia; Goodpasture syndrome
Immune complex–mediated (type III) hypersensitivity	Deposition of antigen-antibody complexes → complement activation → recruitment of leukocytes by complement products and Fc receptors → release of enzymes and other toxic molecules	Inflammation, necrotizing vasculitis (fibrinoid necrosis)	Systemic lupus erythematosus; some forms of glomerulonephritis; serum sickness; Arthus reaction
Cell-mediated (type IV) hypersensitivity	Activated T lymphocytes → (1) release of cytokines, inflammation and macrophage activation; (2) T cell–mediated cytotoxicity	Perivascular cellular infiltrates; edema; granuloma formation; cell destruction	Contact dermatitis; multiple sclerosis; type 1 diabetes; tuberculosis

Ig, Immunoglobulin.

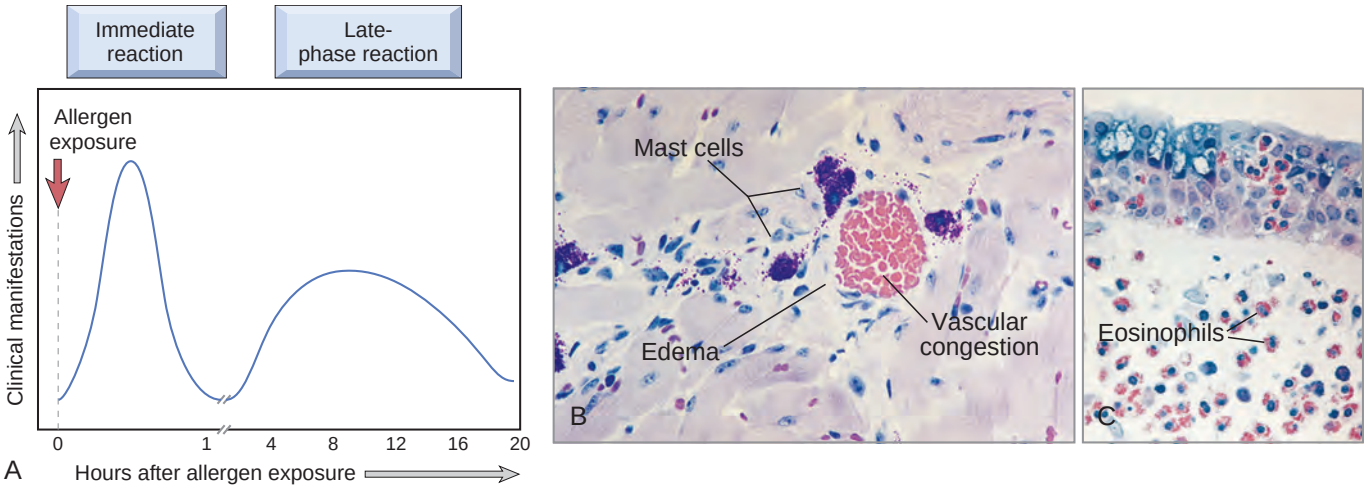


Figure 6.13 Phases of immediate hypersensitivity reactions. (A) Kinetics of the immediate and late-phase reactions. The immediate vascular and smooth muscle reaction to allergen develops within minutes after challenge (allergen exposure in a previously sensitized individual), and the late-phase reaction develops 2 to 24 hours later. The immediate reaction (B) is characterized by vasodilation, congestion, and edema, and the late-phase reaction (C) is characterized by an inflammatory infiltrate rich in eosinophils, neutrophils, and T cells. (Courtesy Dr. Daniel Friend, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

that act on vessels and smooth muscle and proinflammatory cytokines that recruit inflammatory cells.

- In *antibody-mediated disorders (type II hypersensitivity)*, secreted IgG and IgM antibodies injure cells by promoting their phagocytosis or lysis and injure tissues by inducing inflammation. Antibodies may also interfere with cellular functions and cause disease without tissue injury.
- In *immune complex–mediated disorders (type III hypersensitivity)*, IgG and IgM antibodies bind antigens usually in the circulation, and the antigen-antibody complexes deposit in tissues and induce inflammation. The leukocytes that are recruited (neutrophils and monocytes) produce tissue damage by release of lysosomal enzymes and generation of toxic free radicals.
- In *cell-mediated immune disorders (type IV hypersensitivity)*, T lymphocytes (Th1 and Th17 cells and CD8+ CTLs) are the cause of the tissue injury.

Immediate (Type I) Hypersensitivity

Immediate, or type I, hypersensitivity is a rapid immunologic reaction occurring in a previously sensitized individual that is triggered by the binding of an antigen to IgE antibody on the surface of mast cells. These reactions are often called *allergy,* and the antigens that elicit them are allergens. Immediate hypersensitivity may occur as a systemic disorder or as a local reaction. The systemic reaction most often follows injection of an antigen into a sensitized individual (e.g., by a bee sting), but can also follow antigen ingestion (e.g., peanut allergens). Sometimes, within minutes the patient goes into a state of shock, which may be fatal. Local reactions are diverse and vary depending on the portal of entry of the allergen. They may take the form of localized cutaneous rash or blisters (skin allergy, hives), nasal and conjunctival discharge (allergic rhinitis and conjunctivitis), hay fever, bronchial asthma, or allergic gastroenteritis (food allergy).

Many local type I hypersensitivity reactions have two well-defined phases (Fig. 6.13). The immediate reaction is characterized by vasodilation, vascular leakage, and, depending on the tissue, smooth muscle spasm or glandular secretions. These changes usually become evident within minutes after exposure to an allergen and tend to subside in a few hours. In many instances (e.g., allergic rhinitis and bronchial asthma), a second, late-phase reaction sets in 2 to 24 hours later without additional exposure to antigen and may last for several days. This late-phase reaction is characterized by infiltration of tissues with eosinophils, neutrophils, basophils, monocytes, and CD4+ T cells, as well as tissue destruction, typically in the form of mucosal epithelial cell damage.

Most immediate hypersensitivity disorders are caused by excessive Th2 responses, and these cells play a central role by stimulating IgE production and promoting inflammation. These Th2-mediated disorders show a characteristic sequence of events (Fig. 6.14), described next.

Activation of Th2 Cells and Production of IgE Antibody

The first step in the generation of Th2 cells is the presentation of the antigen to naïve CD4+ helper T cells, probably by DCs that capture the antigen from its site of entry. For reasons that are still not understood, only some environmental antigens elicit strong Th2 responses and thus serve as allergens. In response to antigen and other stimuli, including cytokines such as IL-4 produced at the local site, the T cells differentiate into Th2 cells. The newly minted Th2 cells produce a number of cytokines on subsequent encounter with the antigen; as mentioned earlier, the signature cytokines of this subset are IL-4, IL-5, and IL-13. IL-4 acts on B cells to stimulate class switching to IgE and promotes the development of additional Th2 cells. IL-5 is involved in the development and activation of eosinophils, which are important effectors of type I hypersensitivity (discussed later). IL-13 enhances IgE production and acts on epithelial cells to stimulate mucus secretion. In addition, Th2 cells (as well as mast cells and epithelial cells) produce chemokines that attract more Th2 cells, as well as other leukocytes, to the reaction site. Patients with chronic atopic diseases such as

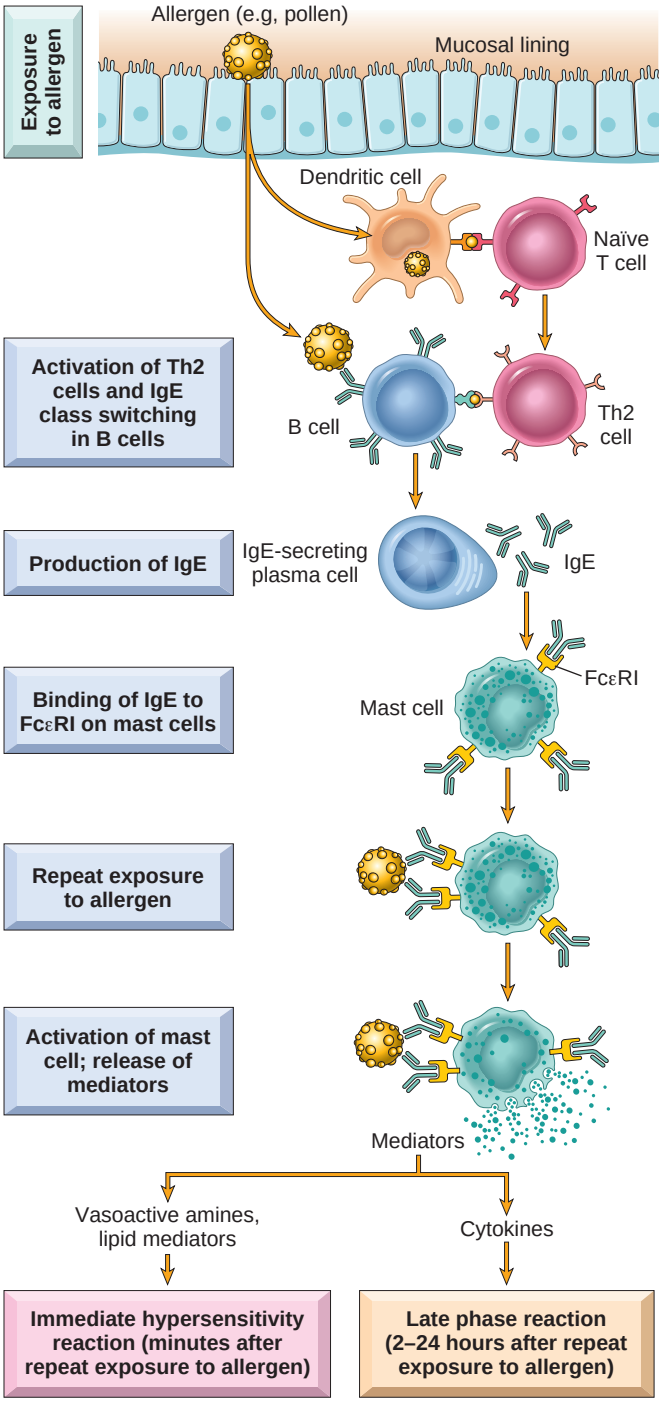


Figure 6.14 Sequence of events in immediate (type I) hypersensitivity. Immediate hypersensitivity reactions are initiated by the introduction of an allergen, which stimulates Th2 responses and IgE production in genetically susceptible individuals. IgE binds to Fc receptors (FcεRI) on mast cells, and subsequent exposure to the allergen activates the mast cells to secrete the mediators that are responsible for the pathologic manifestations of immediate hypersensitivity.

asthma and atopic dermatitis are sometimes classified into Th2-high and Th2-low based on biomarkers that reflect the intensity of the pathologic T-cell response in individual patients. This separation may serve as a guide to therapy, as antagonists of Th2 cytokines (IL-4, IL-5) are predictably most effective in the Th2-high group.

Before Th2 responses develop, type 2 ILCs in tissues may respond to cytokines produced by damaged epithelia. These ILCs secrete IL-5 and IL-13 and are thus able to induce the same tissue reactions as the classical Th2 cells. Over time, the Th2 cells become the dominant contributors to the local cytokine response.

Sensitization and Activation of Mast Cells

Because mast cells are central to the development of immediate hypersensitivity, we first review some of their salient characteristics. Mast cells are bone marrow–derived cells that are widely distributed in the tissues. They are abundant near small blood vessels and nerves and in subepithelial tissues, which explains why local immediate hypersensitivity reactions often occur at these sites. Mast cells have cytoplasmic membrane-bound granules that contain a variety of biologically active mediators, described later. The granules also contain acidic proteoglycans that bind basic dyes such as toluidine blue. (*Mast* in German refers to fattening of animals, and the name of these cells came from the erroneous belief that their granules fed the tissue where the cells were located.) As detailed next, mast cells (and their circulating counterpart, basophils) are activated by the cross-linking of high-affinity IgE Fc receptors; in addition, mast cells may also be triggered by several other stimuli, such as complement components C5a and C3a (called anaphylatoxins because they elicit reactions that mimic anaphylaxis), both of which act by binding to receptors on the mast cell membrane. Other mast cell secretagogues include some chemokines (e.g., IL-8), drugs such as codeine and morphine, adenosine, melittin (present in bee venom), and physical stimuli (e.g., heat, cold, sunlight). Basophils are similar to mast cells in many respects, including the presence of cell surface IgE Fc receptors as well as cytoplasmic granules. In contrast to mast cells, however, basophils are not normally present in tissues but rather circulate in the blood in small numbers. Similar to other granulocytes, basophils can be recruited to inflammatory sites.

When a mast cell armed with IgE antibodies previously produced in response to an antigen is exposed to the same antigen, the cell is activated, leading to the release of an arsenal of powerful mediators that are responsible for immediate hypersensitivity reactions. Mast cells and basophils express a high-affinity receptor called FcεRI that is specific for the Fc portion of IgE and avidly binds IgE antibodies. IgE-coated mast cells are said to be sensitized because they are activated by subsequent encounters with antigen. In the first step of activation, the antigen binds to the IgE antibodies on the mast cell surface. Multivalent antigens bind to and cross-link adjacent IgE antibodies, bringing the underlying Fcε receptors together. This triggers signal transduction pathways from the cytoplasmic portion of the receptors that lead to the release of preformed mediators and de novo production of mediators that are responsible for the initial, sometimes explosive, symptoms of immediate hypersensitivity, and they also set into motion the events that lead to the late-phase reaction.

Mediators of Immediate Hypersensitivity

Mast cell activation leads to degranulation, with the discharge of preformed mediators that are stored in the granules, and

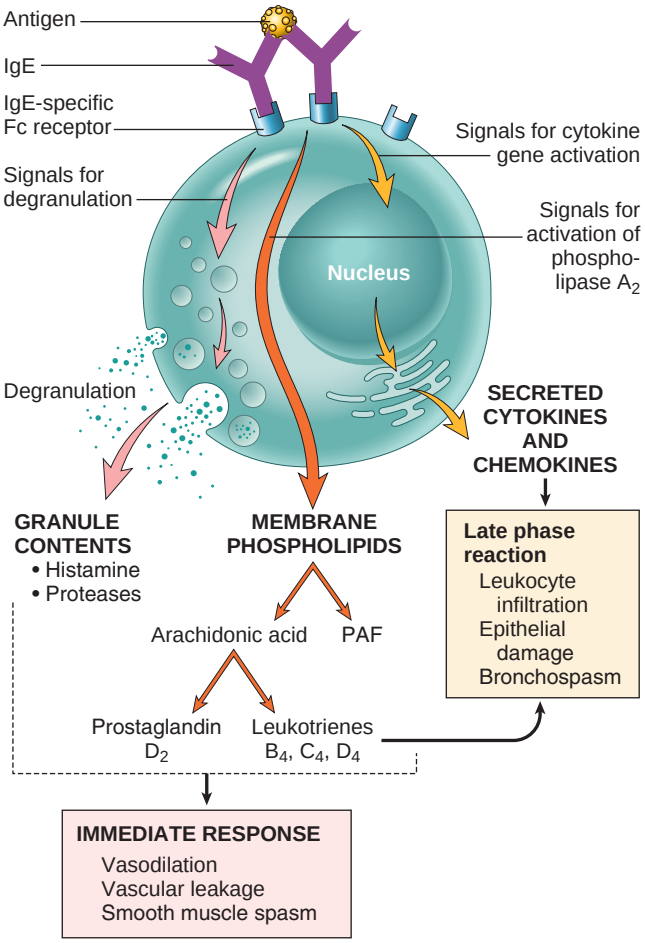


Figure 6.15 Mast cell mediators. On activation, mast cells release various classes of mediators that are responsible for the immediate and late-phase reactions. *PAF,* Platelet-activating factor.

de novo synthesis and release of additional mediators including lipid products and cytokines (Fig. 6.15).

Granule contents. Mediators contained within mast cell granules are the first to be released and can be divided into three categories:

- *Vasoactive amines*. The most important mast cell–derived amine is *histamine* (Chapter 3). Histamine causes intense smooth muscle contraction, increases vascular permeability, and stimulates mucus secretion by nasal, bronchial, and gastric glands.
- *Enzymes.* These are contained in the granule matrix and include neutral proteases (chymase, tryptase) and several acid hydrolases. The enzymes cause tissue damage and lead to the generation of kinins and activated components of complement (e.g., C3a) by acting on their precursor proteins.
- *Proteoglycans.* These include heparin, a well-known anticoagulant, and chondroitin sulfate. The proteoglycans serve to package and store the amines in the granules.

Lipid mediators. The major lipid mediators are arachidonic acid–derived products (Chapter 3). Mast cell activation is associated with activation of phospholipase A_2, an enzyme that converts membrane phospholipids to arachidonic acid. This is the parent compound from which leukotrienes and prostaglandins are produced by the 5-lipoxygenase and the cyclooxygenase pathways, respectively.

- *Leukotrienes.* Leukotrienes C_4 and D_4 are the most potent vasoactive and spasmogenic agents known. On a molar basis, they are several thousand times more active than histamine in increasing vascular permeability and causing bronchial smooth muscle contraction. Leukotriene B_4 is highly chemotactic for neutrophils, eosinophils, and monocytes.
- *Prostaglandin D_2.* This is the most abundant mediator produced in mast cells by the cyclooxygenase pathway. It causes intense bronchospasm and increases mucus secretion.
- *Platelet-activating factor (PAF).* PAF (Chapter 3) is a lipid mediator produced by some mast cell populations that is not derived from arachidonic acid. It causes platelet aggregation, histamine release, bronchospasm, increased vascular permeability, and vasodilation. Its role in immediate hypersensitivity reactions is not well established.

Cytokines. Mast cells are sources of many cytokines, which may play an important role at several stages of immediate hypersensitivity reactions. The cytokines include: TNF, IL-1, and chemokines, which promote leukocyte recruitment (typical of the late-phase reaction); IL-4, which amplifies the Th2 response; and numerous others. The inflammatory cells that are recruited by mast cell–derived TNF and chemokines are additional sources of cytokines.

The mediators produced by mast cells are responsible for most of the manifestations of immediate hypersensitivity reactions. Some, such as histamine and leukotrienes, are released rapidly from sensitized mast cells and trigger the intense immediate reactions characterized by edema, mucus secretion, and smooth muscle spasm; others, exemplified by cytokines, including chemokines, set the stage for the late-phase response by recruiting additional leukocytes. Not only do these inflammatory cells release additional waves of mediators (including cytokines), but they also cause epithelial cell damage. Epithelial cells themselves are not passive bystanders in this reaction; they can also produce soluble mediators, such as chemokines.

Late-Phase Reaction

In the late-phase reaction, leukocytes are recruited that amplify and sustain the inflammatory response without additional exposure to the triggering antigen. *Eosinophils* are often an abundant leukocyte population in these reactions (see Fig. 6.13C). They are recruited to sites of immediate hypersensitivity by chemokines, such as eotaxin, and others that may be produced by epithelial cells, Th2 cells, and mast cells. The Th2 cytokine IL-5 is the most potent eosinophil-activating cytokine known. Upon activation, eosinophils liberate proteolytic enzymes as well as two unique proteins called major basic protein and eosinophil cationic protein, which damage tissues. Eosniophils contain crystals called Charcot-Leyden crystals composed of the protein galectin-10, which are sometimes released into the extracellular space and can be detected in the sputum of patients with asthma. These crystals promote inflammation and enhance Th2 responses, so they may contribute to allergic reactions. It is now believed that the late-phase reaction is a major cause of symptoms in some type I hypersensitivity disorders, such as allergic asthma. Therefore, treatment of these diseases

requires the use of broad-spectrum antiinflammatory drugs, such as steroids, rather than antihistamine drugs, which are of benefit only in the immediate reaction as occurs in allergic rhinitis (hay fever).

Development of Allergies

Susceptibility to immediate hypersensitivity reactions is genetically determined. A propensity to develop immediate hypersensitivity reactions is called *atopy*. Atopic individuals tend to have higher serum IgE levels and more IL-4–producing Th2 cells than does the general population. A positive family history of allergy is found in 50% of atopic individuals. The basis of familial predisposition is not clear, but studies in patients with asthma reveal linkage to polymorphisms in several genes encoding cytokines with important roles in allergic reaction, including the genes for the cytokines IL-3, IL-4, IL-5, IL-9, IL-13, and GM-CSF. How the disease-associated polymorphisms influence the development of allergies is not known. Linkage has also been noted to polymorphisms lying within the HLA genes located on chromosome 6, suggesting that the inheritance of certain HLA alleles permits reactivity to certain allergens.

Environmental factors are also important in the development of allergic diseases. Exposure to environmental pollutants, which is common in industrialized societies, is an important predisposing factor for allergy. For example, dogs and cats diverged from humans about 95 million years ago and are genetically distant from humans compared with chimpanzees, which diverged only about 4 to 5 million years ago and are >95% identical to humans genetically; yet dogs and cats, who live in the same environment as humans, develop allergies, and chimps do not. This observation suggests that environmental factors may be more important in the development of allergic disease than genetics. Viral infections of the airways are triggers for bronchial asthma, an allergic disease affecting the lungs (Chapter 15).

It is estimated that 20% to 30% of immediate hypersensitivity reactions are triggered by non-antigenic stimuli such as temperature extremes and exercise, and do not involve Th2 cells or IgE; such reactions are sometimes called *nonatopic allergy*. It is believed that in these cases mast cells are abnormally sensitive to activation by various nonimmune stimuli.

The incidence of many allergic diseases has increased in high income countries, as populations have urbanized and exposure to the natural environment has diminished. These observations have led to an idea, sometimes called the *hygiene hypothesis,* that early childhood and even prenatal exposure to microbial antigens "educates" the immune system in such a way that subsequent pathologic responses against common environmental allergens are prevented. Thus, paradoxically, improved hygiene in early childhood may increase allergies later in life. This hypothesis, however, is difficult to prove, and the underlying mechanisms are not defined.

With this consideration of the basic mechanisms of type I hypersensitivity, we turn to some clinically important examples of IgE-mediated disease. These reactions can lead to a wide spectrum of injury and clinical manifestations (Table 6.2).

Table 6.2 Examples of Disorders Caused by Immediate Hypersensitivity

Clinical Syndrome	Clinical and Pathologic Manifestations
Anaphylaxis (may be caused by drugs, bee sting, food)	Fall in blood pressure (shock) caused by vascular dilation; airway obstruction due to laryngeal edema
Bronchial asthma	Airway obstruction caused by bronchial smooth muscle hyperactivity; inflammation and tissue injury caused by late-phase reaction
Allergic rhinitis, sinusitis (hay fever)	Increased mucus secretion; inflammation of upper airways, sinuses
Food allergies	Increased peristalsis due to contraction of intestinal muscles

Systemic Anaphylaxis

Systemic anaphylaxis is characterized by vascular shock, widespread edema, and difficulty in breathing. It may occur in sensitized individuals in hospital settings after administration of foreign proteins (e.g., antisera), hormones, enzymes, polysaccharides, and drugs (e.g., the antibiotic penicillin), or in the community setting following exposure to food allergens (e.g., peanuts, shellfish) or insect toxins (e.g., those in bee venom). Extremely small doses of antigen may trigger anaphylaxis, for example, the tiny amounts used in skin testing for allergies. Because of the risk of severe allergic reactions to minute quantities of peanuts in confined quarters, commercial airlines have largely stopped serving peanuts or foods containing peanuts. Within minutes after exposure to allergens, itching, hives, and skin erythema appear, followed shortly thereafter by a striking contraction of pulmonary bronchioles and respiratory distress. Laryngeal edema results in hoarseness and further compromises breathing. Vomiting, abdominal cramps, diarrhea, and laryngeal obstruction follow, and the patient may go into shock and even die within the hour. The risk of anaphylaxis must be borne in mind when certain therapeutic agents are administered. Although patients at risk often have a previous history of some form of allergy, the absence of such a history does not preclude the possibility of an anaphylactic reaction.

Local Immediate Hypersensitivity Reactions

About 10% to 20% of the population suffers from allergies involving localized reactions to common environmental allergens, such as pollen, animal dander, house dust, foods, and the like. Specific diseases include urticaria, allergic rhinitis (hay fever), bronchial asthma, atopic dermatitis, and food allergies; these are discussed elsewhere in this text.

KEY CONCEPTS

IMMEDIATE (TYPE I) HYPERSENSITIVITY

- These are also called allergic reactions, or allergies.
- They are induced by environmental antigens (allergens) that stimulate strong Th2 responses and IgE production in genetically susceptible individuals.
- IgE attaches to mast cells by binding to Fcε receptors; reexposure to the allergen leads to cross-linking of the IgE and FcεRI, activation of mast cells, and release of mediators.
- The principal mediators are histamine, proteases, and other granule contents; prostaglandins and leukotrienes; and cytokines.

- The mediators are responsible for the immediate vascular and smooth muscle reactions and the late-phase reaction (inflammation).
- The clinical manifestations may be local or systemic, and range from mildly annoying rhinitis to fatal anaphylaxis.

Antibody-Mediated (Type II) Hypersensitivity

Antibodies that react with antigens present on cell surfaces or in the extracellular matrix cause disease by destroying these cells, triggering inflammation, or interfering with normal functions. The antibodies may be specific for normal cell or tissue antigens (autoantibodies) or for exogenous antigens, such as chemical or microbial proteins, that bind to a cell surface or tissue matrix. The antibody-dependent mechanisms that cause tissue injury and disease are illustrated in Fig. 6.16 and described next. These reactions are the cause of several important diseases (Table 6.3).

Opsonization and Phagocytosis

Phagocytosis is largely responsible for depletion of cells coated with antibodies. Cells opsonized by IgG antibodies are recognized by phagocyte Fc receptors, which are specific for the Fc portions of some IgG subclasses. In addition, when IgM or IgG antibodies are deposited on the surfaces of cells, they may activate the complement system by the classical pathway. Complement activation generates cleavage products of C3, mainly C3b and C4b, which are deposited on the surfaces of the cells and recognized by phagocytes that express receptors for these proteins. The net result is phagocytosis of the opsonized cells and their destruction inside the phagocytes (see Fig. 6.16A). Complement activation on cells also leads to the formation of the membrane attack complex, which disrupts membrane integrity by "drilling holes" through the lipid bilayer, thereby causing osmotic lysis of the cells. This mechanism is probably effective in destroying only cells and microbes with thin cell walls.

Antibody-mediated destruction of cells also may occur by another process called antibody-dependent cellular

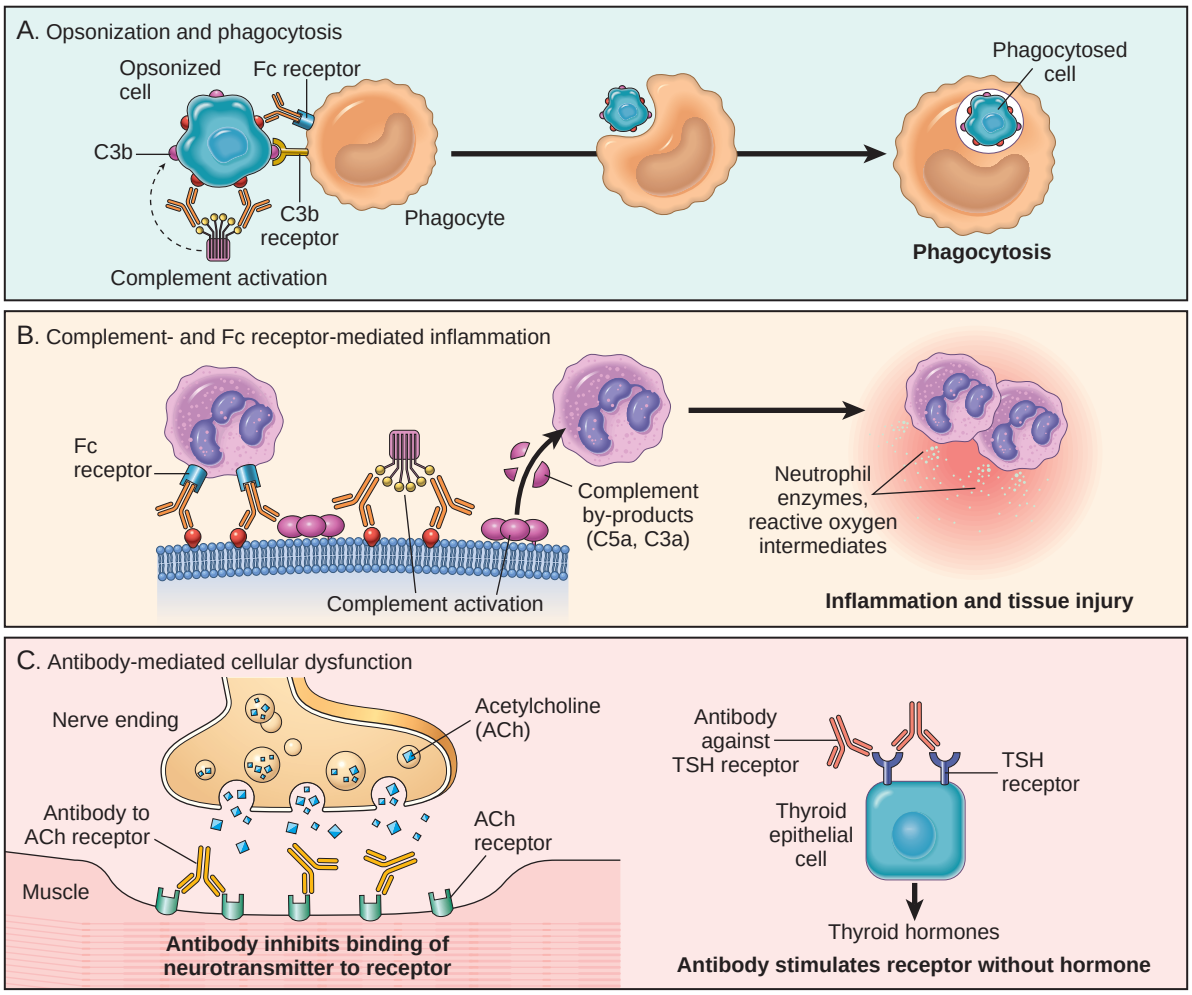


Figure 6.16 Mechanisms of antibody-mediated injury. (A) Opsonization of cells by antibodies and complement components and ingestion by phagocytes. (B) Inflammation induced by antibody binding to Fc receptors of leukocytes and by complement breakdown products. (C) Antireceptor antibodies disturb the normal function of receptors. In these examples, antibodies to the acetylcholine *(ACh)* receptor impair neuromuscular transmission in myasthenia gravis, and antibodies against the thyroid-stimulating hormone *(TSH)* receptor activate thyroid cells in Graves disease.

Table 6.3 Examples of Antibody-Mediated Diseases (Type II Hypersensitivity)

Disease	Target Antigen	Mechanisms of Disease	Clinicopathologic Manifestations
Autoimmune hemolytic anemia	Red cell membrane proteins (Rh blood group antigens, I antigen)	Opsonization and phagocytosis of red cells	Hemolysis, anemia
Autoimmune thrombocytopenic purpura	Platelet membrane proteins (GpIIb:IIIa integrin)	Opsonization and phagocytosis of platelets	Bleeding
Pemphigus vulgaris	Proteins in intercellular junctions of epidermal cells (desmogleins)	Antibody-mediated activation of proteases, disruption of intercellular adhesions	Skin vesicles (bullae)
Vasculitis caused by ANCA	Neutrophil granule proteins, presumably released from activated neutrophils	Neutrophil degranulation and inflammation	Vasculitis
Goodpasture syndrome	Noncollagenous protein in basement membranes of kidney glomeruli and lung alveoli	Complement- and Fc receptor–mediated inflammation	Nephritis, lung hemorrhage
Acute rheumatic fever	Streptococcal cell wall antigen; antibody cross-reacts with myocardial antigen	Inflammation, macrophage activation	Myocarditis, arthritis
Myasthenia gravis	Acetylcholine receptor	Antibody inhibits acetylcholine binding, down-modulates receptors	Muscle weakness, paralysis
Graves disease (hyperthyroidism)	TSH receptor	Antibody-mediated stimulation of TSH receptors	Hyperthyroidism
Pernicious anemia	Intrinsic factor of gastric parietal cells	Neutralization of intrinsic factor, decreased absorption of vitamin B_{12}	Abnormal erythropoiesis, anemia

ANCA, Antineutrophil cytoplasmic antibodies; *TSH*, thyroid-stimulating hormone.

cytotoxicity (ADCC). Cells that are coated with IgG antibody are killed by effector cells, mainly NK cells and macrophages, which bind to the target by their receptors for the Fc fragment of IgG, and cell lysis proceeds without phagocytosis. The contribution of ADCC to common hypersensitivity diseases is uncertain.

Clinically, antibody-mediated cell destruction and phagocytosis occur in the following situations: (1) transfusion reactions, in which cells from a blood group mismatched donor react with and are opsonized by preformed antibody in the host (Chapter 14); (2) hemolytic disease of the fetus and newborn (erythroblastosis fetalis), in which there is an antigenic difference between the mother and the fetus, and IgG anti-erythrocyte antibodies from the mother cross the placenta and cause destruction of fetal red cells (Chapter 10); (3) autoimmune hemolytic anemia, agranulocytosis, and thrombocytopenia, in which individuals produce antibodies to their own blood cells, which are then destroyed (Chapter 14); and (4) certain drug reactions. In some cases of drug-induced antibody-mediated destruction of blood cells, the drug binds to plasma membrane proteins on host cells, and antibodies are produced against the drug-protein complex. In other instances, the offending drug modifies the conformation of an antigen, generating new antigenic epitopes against which the individual reacts.

Inflammation

When antibodies deposit in fixed tissues, such as basement membranes and extracellular matrix, the resultant injury is due to inflammation. The deposited antibodies activate complement, generating cleavage products, including chemotactic agents (mainly C5a), which direct the migration of granulocytes and monocytes, and anaphylatoxins (C3a and C5a), which increase vascular permeability (see Fig. 6.16B). The leukocytes are activated by engagement of their C3b and Fc receptors. This results in the release of substances from the leukocytes that damage tissues, such as lysosomal enzymes, including proteases capable of digesting basement membrane, collagen, elastin, and cartilage, and by generation of reactive oxygen species.

Antibody-mediated inflammation is the mechanism responsible for tissue injury in some forms of glomerulonephritis, vascular rejection in organ grafts, and numerous other disorders (see Table 6.3).

Cellular Dysfunction

In some cases, antibodies directed against cell surface receptors impair or dysregulate function without causing cell injury or inflammation (see Fig. 6.16C). For example, in myasthenia gravis, antibodies reactive with acetylcholine receptors in the motor end plates of skeletal muscles block neuromuscular transmission and therefore cause muscle weakness. The converse (i.e., antibody-mediated stimulation of cell function) is the basis of Graves disease. In this disorder, antibodies against the thyroid-stimulating hormone receptor on thyroid epithelial cells stimulate the cells, resulting in hyperthyroidism.

Immune Complex–Mediated (Type III) Hypersensitivity

Antigen-antibody complexes produce tissue damage mainly by eliciting inflammation at the sites of deposition. The pathologic reaction is usually initiated when antigen combines

Table 6.4 Examples of Immune Complex–Mediated Diseases

Disease	Antigen Involved	Clinicopathologic Manifestations
Systemic lupus erythematosus	Nuclear antigens	Nephritis, skin lesions, arthritis, others
Poststreptococcal glomerulonephritis	Streptococcal cell wall antigen(s); may be "planted" in glomerular basement membrane	Nephritis
Polyarteritis nodosa	Hepatitis B virus antigens in some cases	Systemic vasculitis
Reactive arthritis	Bacterial antigens (e.g., *Yersinia*)	Acute arthritis
Serum sickness	Various proteins, e.g., foreign serum protein (horse antithymocyte globulin)	Arthritis, vasculitis, nephritis
Arthus reaction (experimental)	Various foreign proteins	Cutaneous vasculitis

with antibody in the circulation, creating immune complexes that typically deposit in vessel walls. Less frequently, the complexes may be formed at sites where antigen has been "planted" previously (called in situ immune complexes). The antigens that form immune complexes may be exogenous, such as a foreign protein that is injected or produced by an infectious microbe, or endogenous, if the individual produces antibody against self antigens (autoimmunity). Examples of immune complex disorders and the antigens involved are listed in Table 6.4. Immune complex–mediated diseases tend be systemic, but often preferentially involve the kidney (glomerulonephritis), joints (arthritis), and small blood vessels (vasculitis), all of which are common sites of immune complex deposition for reasons mentioned below.

Systemic Immune Complex Disease

Serum sickness is the prototype of a systemic immune complex disease; it was once a frequent sequela to the administration of large amounts of foreign serum (e.g., serum from immunized horses used for protection against diphtheria). In modern times, the disease is infrequent and usually seen in individuals who receive antibodies from other individuals or species. Nevertheless, it is an informative model that has taught us a great deal about systemic immune complex disorders.

The pathogenesis of systemic immune complex disease can be divided into three phases (Fig. 6.17).

1. *Formation of immune complexes.* The introduction of a protein antigen triggers an immune response that results in the formation of antibodies, typically about 1 week after the injection of the protein. These antibodies are secreted into the blood, where they react with the antigen still present in the circulation and form antigen-antibody complexes.
2. *Deposition of immune complexes.* In the next phase, the circulating antigen-antibody complexes are deposited in vessels. The factors that determine whether immune complex formation will lead to tissue deposition and disease are not fully understood, but the major influences seem to be the characteristics of the complexes and local vascular alterations. In general, complexes that are of medium size, formed under conditions of slight antigen excess, are the most pathogenic. Organs where blood is filtered at high pressure to form other fluids, like urine and synovial fluid, are sites where immune complexes become concentrated and tend to deposit; hence, immune complex disease often affects glomeruli and joints.

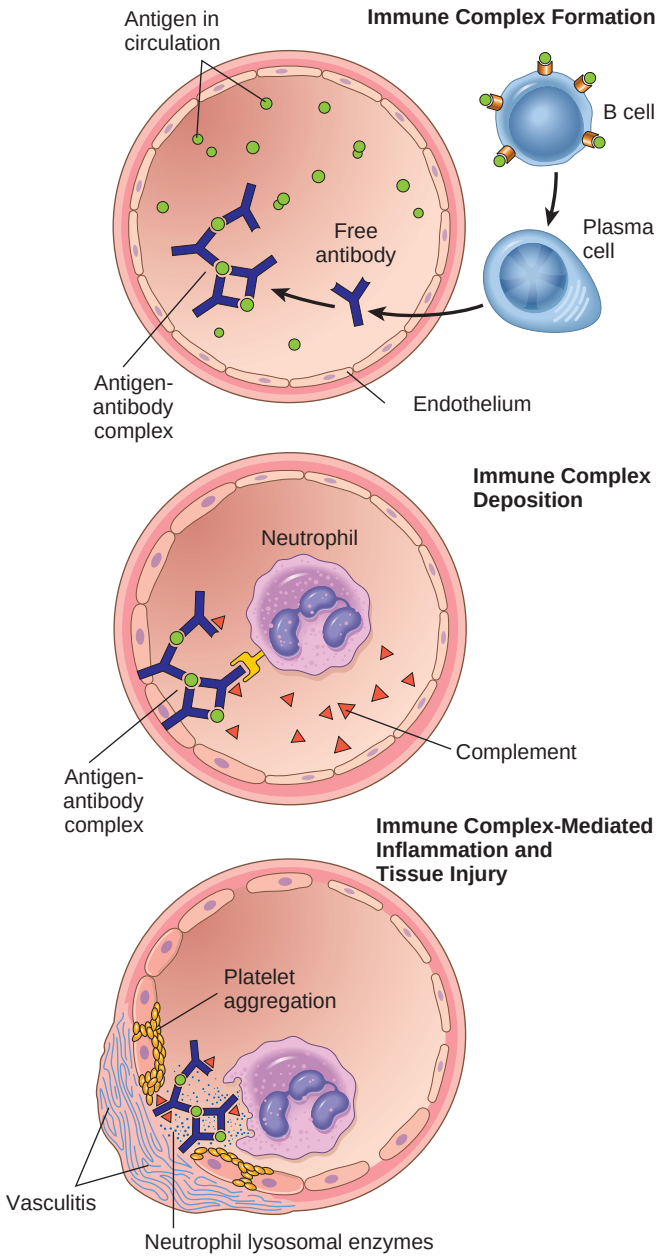


Figure 6.17 Immune complex disease. The sequential phases in the induction of systemic immune complex–mediated diseases (type III hypersensitivity).

3. *Inflammation and tissue injury.* Once immune complexes are deposited in the tissues, they initiate an acute inflammatory reaction. During this phase (approximately 10 days after antigen administration), clinical features such as fever, urticaria, joint pain, lymph node enlargement, and proteinuria appear. Wherever complexes deposit, inflammation and tissue injury occur through the antibody-mediated mechanisms that were discussed earlier. The important role of complement in the pathogenesis of the tissue injury is supported by the observations that complement proteins can be detected at the site of injury and, during the active phase of the disease, consumption of complement leads to a decrease in serum levels of C3. In fact, serum C3 levels can, in some cases, be used to monitor disease activity.

MORPHOLOGY

The principal morphologic manifestation of immune complex injury is acute vasculitis, associated with necrosis of the vessel wall and intense neutrophilic infiltration. The necrotic tissue and deposits of immune complexes, complement, and plasma protein appear as a smudgy eosinophilic area of tissue destruction, an appearance termed **fibrinoid necrosis** (see Fig. 2.15). When deposited in the kidney, the complexes can be seen on immunofluorescence microscopy as granular lumpy deposits of immunoglobulin and complement and on electron microscopy as electron-dense deposits along the glomerular basement membrane (see Figs. 6.31 and 6.32).

In *acute serum sickness*, caused by a single exposure to a large amount of antigen, the lesions tend to resolve as a result of catabolism of the immune complexes. A form of *chronic serum sickness* results from repeated or prolonged exposure to an antigen. This occurs in several diseases, such as systemic lupus erythematosus (SLE), which is associated with persistent antibody responses to autoantigens. In many diseases, the morphologic changes and other findings suggest immune complex deposition, but the inciting antigens are unknown. Included in this category are membranous glomerulonephritis and several vasculitides.

Local Immune Complex Disease

The *Arthus reaction* is a localized area of tissue necrosis resulting from acute immune complex vasculitis, usually elicited in the skin. The reaction can be produced experimentally by intracutaneous injection of antigen in a previously immunized animal that contains circulating antibodies against the antigen. As the antigen diffuses into the vascular wall, it binds the preformed antibody, and large immune complexes are formed locally. These complexes precipitate in the vessel walls and cause fibrinoid necrosis, and superimposed thrombosis worsens the ischemic injury.

KEY CONCEPTS

PATHOGENESIS OF DISEASES CAUSED BY ANTIBODIES AND IMMUNE COMPLEXES

- Antibodies can coat (opsonize) cells, with or without complement proteins, and target these cells for phagocytosis by phagocytes (macrophages), which express receptors for the Fc tails of IgG and for complement proteins. The result is depletion of the opsonized cells.
- Antibodies and immune complexes may deposit in tissues or blood vessels and elicit an acute inflammatory reaction by activating complement, with release of breakdown products, or by engaging Fc receptors of leukocytes. The inflammatory reaction causes tissue injury.
- Antibodies can bind to cell surface receptors or other essential molecules and cause functional derangements (either inhibition or unregulated activation) without cell injury.

T Cell–Mediated (Type IV) Hypersensitivity

Cell-mediated hypersensitivity is caused mainly by inflammation resulting from cytokines produced by CD4+ T cells (Fig. 6.18). CD4+ T cell–mediated hypersensitivity induced by environmental and self antigens is the cause of many autoimmune and other chronic inflammatory diseases (Table 6.5). Cell killing by CD8+ cells may also be involved in some autoimmune diseases and may be the dominant mechanism of tissue injury in certain reactions, especially those that follow viral infections.

CD4+ T Cell–Mediated Inflammation

In CD4+ T cell–mediated hypersensitivity reactions, cytokines produced by T cells induce inflammation that may be chronic and destructive. The prototype of T cell–mediated inflammation is *delayed-type hypersensitivity (DTH),* a tissue reaction to antigens given to immune individuals. In this reaction, an antigen administered into the skin of a previously immunized individual results in a detectable cutaneous reaction within 24 to 48 hours (hence the term delayed, in contrast to immediate hypersensitivity). Both Th1 and Th17 cells contribute to organ-specific diseases in which inflammation is a prominent aspect of the pathology. The inflammatory reaction associated with Th1 cells is dominated by activated macrophages, and that triggered by Th17 cells has a greater neutrophil component.

The inflammatory reactions stimulated by CD4+ T cells can be divided into sequential stages.

Activation of CD4+ T Cells

As described earlier, naïve CD4+ T cells recognize peptides displayed by DCs and secrete IL-2, which functions as an autocrine growth factor to stimulate proliferation of the antigen-responsive T cells. The subsequent differentiation of antigen-stimulated T cells to Th1 or Th17 cells is driven by the cytokines produced by APCs at the time of T-cell activation. In some situations, the APCs (DCs and macrophages) produce IL-12, which induces differentiation of CD4+ T cells to the Th1 subset. IFN-γ produced by these effector cells promotes further Th1 development, thus amplifying the reaction. If the APCs produce the inflammatory cytokines IL-1, IL-6, and a close relative of IL-12 called IL-23, the T cells are induced to differentiate to the Th17 subset. Some of the differentiated effector cells enter the circulation and join the pool of memory T cells, where they persist for long periods, sometimes years.

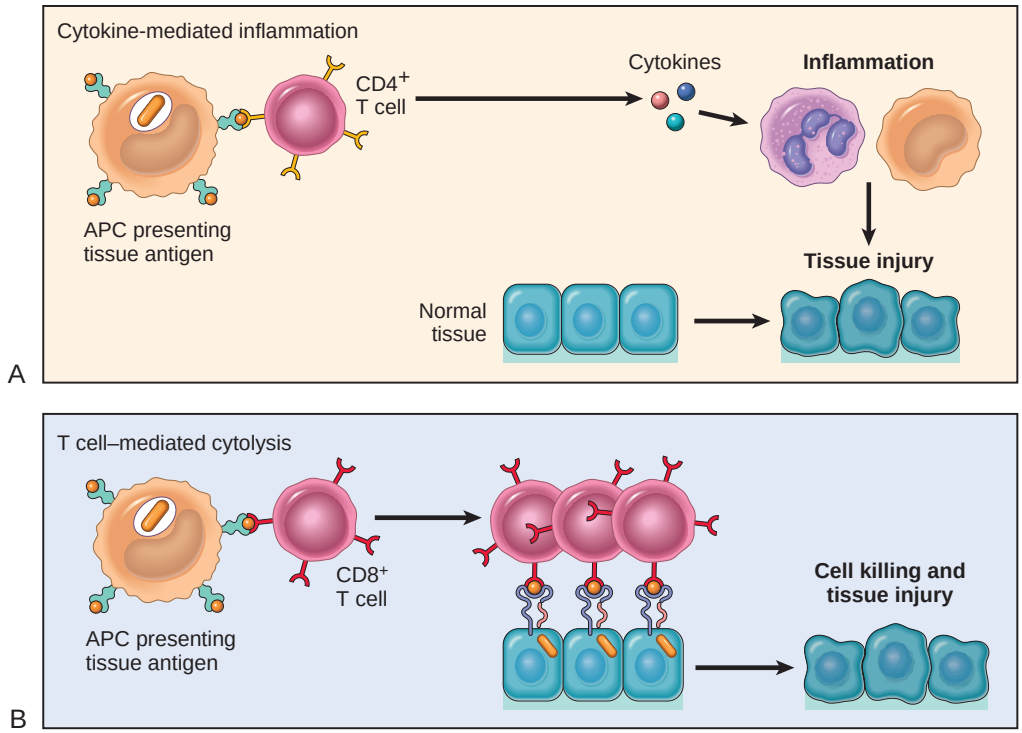


Figure 6.18 Mechanisms of T-cell–mediated (type IV) hypersensitivity reactions. (A) CD4+ Th1 cells (and sometimes CD8+ T cells, *not shown*) respond to tissue antigens by secreting cytokines that stimulate inflammation and activate phagocytes, leading to tissue injury. CD4+ Th17 cells contribute to inflammation by recruiting neutrophils (and, to a lesser extent, monocytes). (B) In some diseases, CD8+ cytotoxic T lymphocytes directly kill tissue cells. *APC*, Antigen-presenting cell.

Responses of Differentiated Effector T Cells

On repeat exposure to an antigen, Th1 cells secrete cytokines, mainly IFN-γ, which are responsible for many of the manifestations of delayed-type hypersensitivity. IFN-γ-activated ("classically activated") macrophages are altered in several ways: their ability to phagocytose and kill microorganisms is markedly augmented; they express more class II MHC molecules on the surface, enhancing antigen presentation; they secrete TNF, IL-1, and chemokines, which promote inflammation (Chapter 3); and they produce more IL-12, amplifying the Th1 response. Thus, activated macrophages serve to eliminate the offending antigen; if the activation is sustained, continued inflammation and tissue injury result.

Table 6.5 T Cell–Mediated Diseases

Disease	Specificity of Pathogenic T Cells	Principal Mechanisms of Tissue Injury	Clinicopathologic Manifestations
Rheumatoid arthritis	Collagen? Citrullinated self proteins?	Inflammation mediated by Th17 (and Th1?) cytokines; role of antibodies and immune complexes?	Chronic arthritis with inflammation, destruction of articular cartilage
Multiple sclerosis	Protein antigens in myelin (e.g., myelin basic protein)	Inflammation mediated by Th1 and Th17 cytokines, myelin destruction by activated macrophages	Demyelination in CNS with inflammation; paralysis, optic neuritis
Type 1 diabetes	Antigens of pancreatic islet β cells (insulin, glutamic acid decarboxylase, others)	T-cell–mediated inflammation, destruction of islet cells by CTLs	Insulitis (chronic inflammation in islets), destruction of β cells; diabetes
Inflammatory bowel disease	Enteric bacteria; self antigens?	Inflammation mediated by Th1 and Th17 cytokines	Chronic intestinal inflammation, obstruction
Psoriasis	Unknown	Inflammation mediated mainly by Th17 cytokines	Destructive plaques in the skin
Contact sensitivity	Various environmental chemicals (e.g., urushiol from poison ivy or poison oak)	Inflammation mediated by Th1 (and Th17?) cytokines	Epidermal necrosis, dermal inflammation, causing skin rash and blisters

Examples of human T cell–mediated diseases are listed. In many cases, the specificity of the T cells and the mechanisms of tissue injury are inferred based on the similarity with experimental animal models of the diseases. *CTLs*, Cytotoxic T lymphocytes.

Activated Th17 cells secrete IL-17, IL-22, chemokines, and several other cytokines. Collectively, these cytokines recruit neutrophils and monocytes to the reaction, thus promoting inflammation.

Clinical Examples of CD4+ T Cell–Mediated Inflammatory Reactions

The classic example of DTH is the tuberculin reaction, which is produced by the intracutaneous injection of purified protein derivative (PPD, also called tuberculin), a protein-containing antigen of the tubercle bacillus. In a previously sensitized individual, reddening and induration of the site appear in 8 to 12 hours, reach a peak in 24 to 72 hours, and thereafter slowly subside. Morphologically, delayed-type hypersensitivity is characterized by the accumulation of mononuclear cells, mainly CD4+ T cells and macrophages, around venules, producing perivascular "cuffing" (Fig. 6.19). In fully developed lesions, the venules show marked endothelial hypertrophy, reflecting cytokine-mediated endothelial activation.

With certain persistent or nondegradable antigens, such as tubercle bacilli colonizing the lungs or other tissues, the infiltrate is dominated by macrophages over a period of 2 or 3 weeks. With sustained activation, macrophages often undergo a morphologic transformation into epithelioid cells, large cells with abundant cytoplasm. Aggregates of epithelioid cells, usually surrounded by lymphocytes, form grossly visible small nodules called *granulomas* (Fig. 6.20). This pattern of chronic inflammation, called granulomatous inflammation (Chapter 3), is commonly associated with strong Th1-cell activation and production of cytokines such as IFN-γ. It can also be caused by indigestible foreign bodies, which activate macrophages without eliciting an adaptive immune response. In some helminthic infections, such as schistosomiasis, the worms lay eggs that elicit granulomatous reactions. These reactions are usually rich in eosinophils and are elicited by strong Th2 responses, which are typical of many helminthic infections.

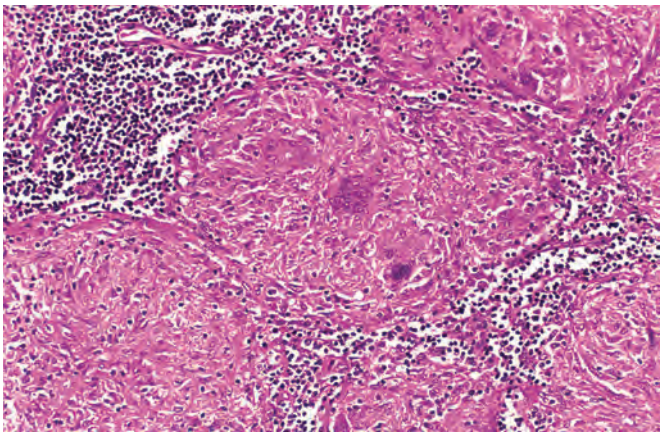


Figure 6.20 Granulomatous inflammation. A section of a lymph node shows several granulomas, each made up of an aggregate of epithelioid cells and surrounded by lymphocytes. The granuloma in the center shows several multinucleate giant cells. (Courtesy Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

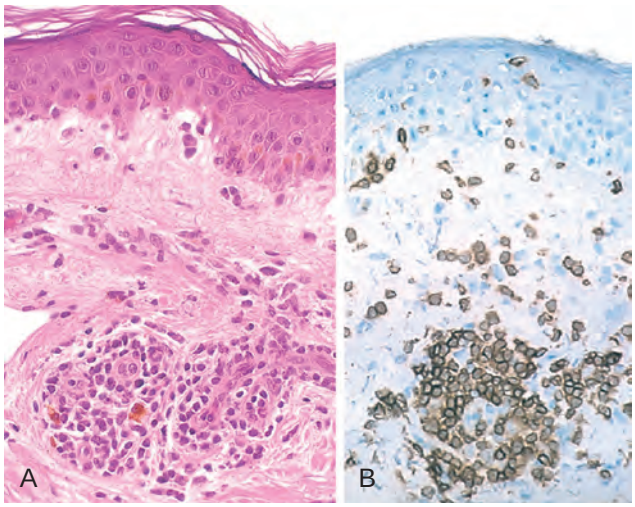


Figure 6.19 Delayed type hypersensitivity reaction in the skin. (A) Perivascular accumulation ("cuffing") of mononuclear inflammatory cells (lymphocytes and macrophages), with associated dermal edema and fibrin deposition. (B) Immunoperoxidase staining reveals a predominantly perivascular cellular infiltrate that marks positively with anti-CD4 antibodies. (Courtesy Dr. Louis Picker, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Contact dermatitis is a common example of tissue injury resulting from DTH reactions. It may be evoked by contact with urushiol, the antigenic component of poison ivy or poison oak, and presents as an itchy, vesicular (blistering) dermatitis. It is thought that in these reactions, the environmental chemical binds to and structurally modifies self proteins, and peptides derived from these modified proteins are recognized by T cells and elicit the reaction. Chemicals may also modify HLA molecules, making them appear foreign to T cells. The same mechanism is responsible for most *drug reactions,* among the most common immunologic reactions of humans. These often manifest as skin rashes.

CD4+ T cell–mediated inflammation is the basis of tissue injury in many organ-specific and systemic autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis, as well as diseases linked to uncontrolled reactions to bacterial commensals, such as inflammatory bowel disease (see Table 6.5).

CD8+ T Cell–Mediated Cytotoxicity

In this type of T cell–mediated reaction, CD8+ CTLs kill antigen-expressing target cells. Tissue destruction by CTLs may be a component of some T cell–mediated diseases, such as type 1 diabetes. CTLs directed against cell surface histocompatibility antigens play an important role in graft rejection, to be discussed later. They also play a role in reactions against viruses. In a virus-infected cell, viral peptides are displayed by class I MHC molecules, and the complex is recognized by the TCR of CD8+ T lymphocytes. The killing of infected cells leads to the elimination of the infection, but in some cases it is responsible for cell damage that accompanies the infection (e.g., in viral hepatitis). Tumor antigens are also presented on the surface of tumor cells, and CTLs are involved in the host response to transformed cells (Chapter 7).

The principal mechanism of T cell–mediated killing of targets involves perforins and granzymes, preformed mediators contained in the lysosome-like granules of CTLs. CTLs that recognize the target cells secrete a complex consisting of perforin, granzymes, and other proteins that enters target cells by endocytosis. In the target cell cytoplasm, perforin facilitates the release of the granzymes from the complex. Granzymes are proteases that cleave and activate caspases, which induce apoptosis of the target cells (Chapter 2). Activated CTLs also express Fas ligand, a molecule with homology to TNF, which also can trigger apoptosis by binding and activating Fas receptor expressed on target cells.

CD8+ T cells also produce cytokines, notably IFN-γ, and are involved in inflammatory reactions resembling DTH, especially following viral infections and exposure to some contact sensitizing agents.

KEY CONCEPTS

MECHANISMS OF T CELL–MEDIATED HYPERSENSITIVITY REACTIONS

- Cytokine-mediated inflammation: CD4+ T cells are activated by exposure to a protein antigen and differentiate into Th1 and Th17 effector cells. Subsequent exposure to the antigen results in the secretion of cytokines. IFN-γ activates macrophages to produce substances that cause tissue damage and promote fibrosis, and IL-17 and other cytokines recruit leukocytes, thus promoting inflammation. The classical T cell–mediated inflammatory reaction is DTH.
- T cell–mediated cytotoxicity: CD8+ cytotoxic T lymphocytes (CTLs) specific for an antigen recognize cells expressing the target antigen and kill these cells. CD8+ T cells also secrete IFN-γ.

Now that we have described how the immune system can cause tissue damage, we turn to diseases in which normal mechanisms of immune regulation fail. The prototypes of such diseases are autoimmune disorders, which are the result of failure of tolerance to self antigens.

AUTOIMMUNE DISEASES

Immune reactions against self antigens—autoimmunity—are an important cause of certain diseases in humans, estimated to affect 5% to 10% of the US population, and the incidence is increasing, especially in high income countries. A growing number of diseases have been attributed to autoimmunity (Table 6.6). It should be noted, however, that the mere presence of autoantibodies does not indicate that an autoimmune disease exists. Autoantibodies can be found in the serum of apparently normal individuals, particularly in older age groups. Furthermore, innocuous autoantibodies are sometimes produced after damage to tissues and may serve a physiologic role in the removal of tissue breakdown products. How, then, does one define pathologic autoimmunity? Ideally, at least three requirements should be met before a disorder is categorized as truly caused by autoimmunity: (1) the presence of an immune reaction specific for some self antigen or self tissue; (2) evidence that such a reaction is not secondary to tissue damage but is of primary pathogenic significance; and (3) the absence of another well-defined cause of the disease. Similarity with experimental models of proven autoimmunity is also often used to support this mechanism in human diseases. Disorders in which chronic inflammation is a prominent component are sometimes grouped under immune-mediated inflammatory diseases; these may be autoimmune, or the immune response may be directed against normally harmless microbes such as gut commensal bacteria.

Table 6.6 Autoimmune Diseases

Organ-Specific	Systemic
Diseases Mediated by Antibodies	
Autoimmune hemolytic anemia	Systemic lupus erythematosus
Autoimmune thrombocytopenia	
Autoimmune atrophic gastritis of pernicious anemia	
Myasthenia gravis	
Graves disease	
Goodpasture syndrome	
Diseases Mediated by T Cells[a]	
Type 1 diabetes mellitus	Rheumatoid arthritis
Multiple sclerosis	Systemic sclerosis (scleroderma)[b]
	Sjögren syndrome[b]
Diseases Postulated to Be Autoimmune	
Inflammatory bowel diseases (Crohn disease, ulcerative colitis)[c]	Polyarteritis nodosa[b]
Primary biliary cirrhosis[b]	Inflammatory myopathies[b]
Autoimmune (chronic active) hepatitis	

[a]A role for T cells has been demonstrated in these disorders, but antibodies may also be involved in tissue injury.

[b]An autoimmune basis of these disorders is suspected, but the supporting evidence is not strong.

[c]These disorders may result from excessive immune responses to commensal enteric microbes, autoimmunity, or a combination of the two.

The clinical manifestations of autoimmune disorders are extremely varied. On one end are conditions in which the immune responses are directed against a single organ or tissue, resulting in *organ-specific disease,* and on the other end are diseases in which the autoimmune reactions are against widespread antigens, resulting in *systemic disease*. Examples of organ-specific autoimmune diseases are type 1 diabetes, in which the autoreactive T cells and antibodies are specific for β cells of the pancreatic islets, and multiple sclerosis, in which autoreactive T cells react against central nervous system (CNS) myelin. The best example of systemic autoimmune disease is SLE, in which a diversity of antibodies directed against DNA, platelets, red cells, and protein-phospholipid complexes result in widespread lesions throughout the body. In the middle of the spectrum falls Goodpasture syndrome, in which antibodies to basement membranes of lung and kidney induce lesions in these organs.

It is obvious that autoimmunity results from the loss of self-tolerance, and the question arises as to how this happens. Before we look for answers to this question, we review the mechanisms of immunologic tolerance to self antigens.

Immunologic Tolerance

Immunologic tolerance is the phenomenon of unresponsiveness to an antigen induced by exposure of lymphocytes to that antigen. *Self-tolerance* refers to lack of responsiveness to an individual's own antigens, and it underlies our ability to live in harmony with our cells and tissues.

The mechanisms of self-tolerance can be broadly classified into two groups: central tolerance and peripheral tolerance (Fig. 6.21). Each of these is considered briefly.

Central Tolerance

In this process, immature self-reactive T and B lymphocyte clones that recognize self antigens during their maturation in the central (primary, or generative) lymphoid organs (the thymus for T cells and the bone marrow for B cells) are killed or rendered harmless. In developing lymphocytes, random somatic antigen receptor gene rearrangements generate diverse antigen receptors, many of which by chance may have high affinity for self antigens. The mechanisms of central tolerance eliminate these potentially dangerous lymphocytes.

- When immature T cells expressing TCRs specific for self antigens encounter these antigens in the thymus, signals are produced that result in killing of the cells by apoptosis. This process is called *negative selection* or *clonal deletion*. A wide variety of autologous protein antigens, including antigens thought to be restricted to peripheral tissues, are processed and presented by thymic antigen-presenting cells in association with self MHC molecules and can, therefore, be recognized by potentially self-reactive T cells. A protein called AIRE (autoimmune regulator) stimulates expression of some peripheral tissue–restricted self antigens in the thymus and is thus critical for deletion of immature T cells specific for these antigens. The importance of this mechanism is emphasized by rare patients with germline loss-of-function mutations in the *AIRE* gene, who develop an autoimmune disorder called *autoimmune polyendocrine syndrome* that leads to destruction of multiple endocrine organs (Chapter 24). In the CD4+ T-cell lineage, some of the cells that see self antigens in the thymus do not die but develop into regulatory T cells (described later). What determines the choice between deletion and development of regulatory T cells in the thymus is not established; it may be partly related to the affinity of the antigen receptor on immature T cells for antigens present in the thymus.
- When developing B cells strongly recognize self antigens in the bone marrow, many of the cells reactivate the machinery of antigen receptor gene rearrangement and begin to express new antigen receptors, not specific for self antigens. This process is called *receptor editing*; it is estimated that one-fourth to one-half of all B cells in the body may have undergone receptor editing during their maturation. If receptor editing does not occur, the self-reactive cells undergo apoptosis, thus purging potentially dangerous lymphocytes from the mature pool.

Central tolerance, however, is imperfect. Not all self antigens may be present in the thymus and bone marrow, and hence lymphocytes bearing receptors for such autoantigens escape into the periphery. Self-reactive lymphocytes that

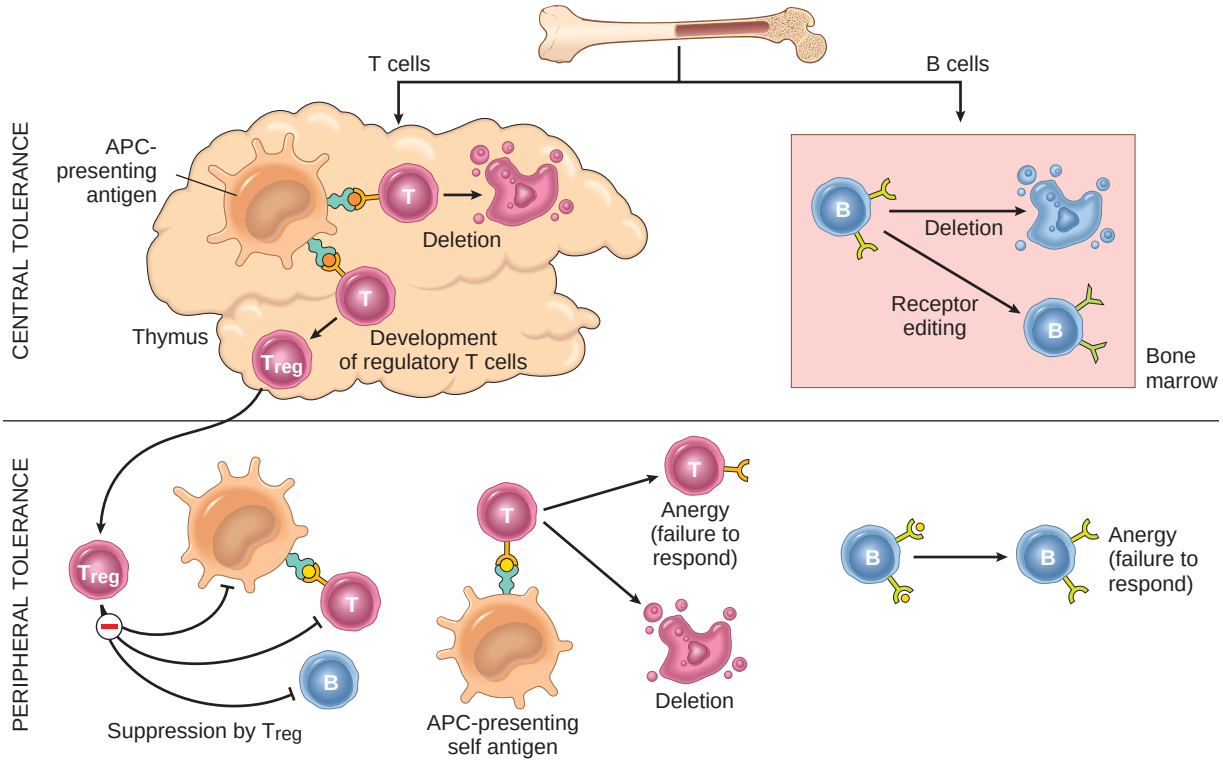


Figure 6.21 Mechanisms of immunologic tolerance to self antigens. The principal mechanisms of central and peripheral self-tolerance in T and B cells are illustrated. *APC*, Antigen-presenting cell.

escape negative selection can inflict tissue injury unless they are deleted or muzzled in the peripheral tissues.

Peripheral Tolerance

Several mechanisms silence potentially autoreactive T and B cells in peripheral tissues; these are best defined for T cells. These mechanisms include the following:

- *Anergy.* Lymphocytes that recognize self antigens may be rendered functionally unresponsive, a phenomenon called anergy. We discussed earlier that activation of antigen-specific T cells requires two signals: recognition of peptide antigen in association with self MHC molecules on the surface of APCs and a set of costimulatory signals ("second signals") from APCs. These second signals are provided by certain T cell–associated molecules, such as CD28, that bind to their ligands (the costimulators B7-1 and B7-2) on APCs. If the antigen is presented to T cells without adequate levels of costimulators, the cells become anergic. Because costimulatory molecules are not expressed or are weakly expressed on resting DCs in normal tissues, the encounter between autoreactive T cells and their specific self antigens displayed by these DCs may lead to anergy. Several mechanisms of T-cell anergy have been demonstrated in various experimental systems. One of these, which has clinical implications, is that T cells that recognize self antigens receive an inhibitory signal from receptors that are structurally homologous to CD28 but serve the opposite functions. Two of these inhibitory receptors, sometimes called coinhibitors (to contrast them with costimulators mentioned earlier), are CTLA-4, which (like CD28) binds to B7 molecules, and PD-1, which binds to two ligands, PD-L1 and PD-L2, that are expressed on a wide variety of cells. The importance of these inhibitory mechanisms has been established by study of CTLA-4 and PD-1 "knockout" mice and humans with inherited mutations in *CTLA4,* all of which lead to systemic inflammatory diseases. Interestingly, some tumors and viruses use the same pathways of immune regulation to evade immune attack. This realization has led to the development of antibodies that block CTLA-4, PD-1, and its ligand PD-L1 for tumor immunotherapy; by removing the brakes on the immune response, these antibodies promote responses against tumors.

  Anergy also affects mature B cells in peripheral tissues. It is believed that if B cells encounter self antigen in peripheral tissues, especially in the absence of specific helper T cells, the B cells become unable to respond to subsequent antigenic stimulation and may be excluded from lymphoid follicles, resulting in their death.
- *Suppression by regulatory T cells.* A population of T cells called regulatory T cells functions to prevent immune reactions against self antigens. Regulatory T cells develop mainly in the thymus, as a result of recognition of self antigens (see Fig. 6.21), but they may also be induced in peripheral lymphoid tissues. The best-defined regulatory T cells are CD4+ cells that express high levels of CD25, the α chain of the IL-2 receptor, and FOXP3, a transcription factor of the forkhead family. Both IL-2 and FOXP3 are required for the development and maintenance of functional CD4+ regulatory T cells. Mutations in *FOXP3* result in severe autoimmunity in humans and mice; in humans these mutations are the cause of a systemic autoimmune disease called *IPEX* (an acronym for *i*mmune dysregulation, *p*olyendocrinopathy, *e*nteropathy, *X*-linked). Mutation of the gene encoding IL-2 or the IL-2 receptor α or β chain also results in a rare multiorgan autoimmune disease because IL-2 is essential for the maintenance of regulatory T cells. Recent genome-wide association studies have revealed that polymorphisms in the promoter of the *CD25* gene are associated with type 1 diabetes, multiple sclerosis, and other autoimmune diseases, raising the possibility of a regulatory T-cell defect contributing to these diseases. Regulatory T cells may suppress immune responses by multiple mechanisms. Their inhibitory activity may be mediated in part by the secretion of immunosuppressive cytokines such as IL-10 and TGF-β, which inhibit lymphocyte activation and effector functions.

  Regulatory T cells prevent immune responses not only against self antigens but also against the fetus and commensal microbes. Placental mammals face a unique challenge because the developing fetus expresses paternal antigens that are foreign to the mother yet have to be tolerated. There is emerging evidence that regulatory T cells prevent immune reactions against fetal antigens that are inherited from the father and therefore foreign to the mother. In line with this idea, during evolution, placentation appeared simultaneously with the ability to stably express the FOXP3 transcription factor. Experiments in mice have shown that fetal antigens induce long-lived FOXP3+ regulatory T cells, and depletion of these cells results in fetal loss. There is great interest in determining the contribution of regulatory T cells in human pregnancy and possible defects in these cells as the basis for recurrent spontaneous abortions.
- *Deletion by apoptosis.* T cells that recognize self antigens may receive signals that promote their death by apoptosis. Depletion of T cells occurs not only in the thymus, discussed earlier, but also in the periphery. Two mechanisms of deletion of mature T cells in the periphery have been proposed, based mainly on studies in mice. It is postulated that if T cells recognize self antigens, they may express a pro-apoptotic member of the Bcl family, called Bim, without antiapoptotic members of the family like Bcl-2 and Bcl-x (whose induction requires the full set of signals for lymphocyte activation). Unopposed Bim triggers apoptosis by the mitochondrial pathway (Chapter 2). A second mechanism involves the Fas-Fas ligand system. Upon recognition of self antigens, lymphocytes express the death receptor Fas (CD95), a member of the TNF-receptor family. Fas ligand (FasL), a membrane protein that is structurally homologous to the cytokine TNF, is expressed mainly on activated T lymphocytes. The engagement of Fas by FasL induces apoptosis by the death receptor pathway (Chapter 2). If self antigens engage antigen receptors of self-reactive T cells, Fas and FasL are co-expressed, leading to elimination of the cells via Fas-mediated apoptosis. Self-reactive B cells may also be deleted by FasL on T cells engaging Fas on the B cells. Mutations in Fas or FasL in mice and humans result in an autoimmune disease reminiscent of human SLE called the *autoimmune lymphoproliferative syndrome (ALPS).*

  Some antigens are hidden (sequestered) from the immune system, because the tissues in which these

antigens are located do not communicate with the blood and lymph. As a result, self antigens in these tissues fail to elicit immune responses and are essentially ignored by the immune system. This is believed to be the case for the testis, eye, and brain, all of which are called *immune-privileged sites* because antigens introduced into these sites tend to elicit weak or no immune responses. If the antigens of these tissues are released, for example, as a consequence of trauma or infection, the result may be an immune response that leads to prolonged tissue inflammation and injury. This is the postulated mechanism for post-traumatic orchitis and uveitis.

Mechanisms of Autoimmunity: General Principles

The immune system normally exists in an equilibrium in which lymphocyte activation, which is required for defense against pathogens, is balanced by the mechanisms of tolerance, which prevent reactions against self antigens. The underlying cause of autoimmune diseases is the failure of tolerance, which allows responses to develop against self antigens. Understanding why tolerance fails in these diseases is an important goal of immunologists.

Autoimmunity arises from a combination of the inheritance of susceptibility genes, which may contribute to the breakdown of self-tolerance, and environmental triggers, such as infections and tissue damage, which promote the activation of self-reactive lymphocytes (Fig. 6.22). Although much remains unknown about the enigma of autoimmunity, the following abnormalities appear to contribute to its development:

- *Defective tolerance or regulation.* Some clues about how tolerance mechanisms are disrupted have come from the analysis of patients with rare inherited (Mendelian) autoimmune disorders and from gene knockout mice that develop autoimmune lesions, mentioned earlier and discussed in more detail later. However, despite these advances, it is still not known why self-tolerance fails in the majority of common autoimmune diseases.
- *Abnormal display of self antigens.* Abnormalities may include increased expression and persistence of self antigens that are normally cleared, or structural changes in these antigens resulting from posttranslational enzymatic modifications or from cellular stress or injury. If these changes lead to the display of "neoantigens" (new epitopes that are not expressed normally), the immune system may not be tolerant to these epitopes, and anti-self responses may develop.
- *Inflammation or an initial innate immune response.* As discussed earlier, the innate immune response is a strong stimulus for the subsequent activation of lymphocytes and the generation of adaptive immune responses. Microbes or cell injury may elicit local inflammatory reactions that may be triggers for subsequent autoimmunity.

Although these are appealing hypotheses, which of these abnormalities actually play a role in specific autoimmune diseases in humans remains largely a matter of speculation.

Role of Susceptibility Genes

Most autoimmune diseases are complex multigenic disorders. It has been known for decades that autoimmunity has a genetic component. The incidence of many autoimmune diseases is greater in twins of affected individuals than in the general population, and greater in monozygotic than in dizygotic twins, proof that genetics contributes to the development of these disorders.

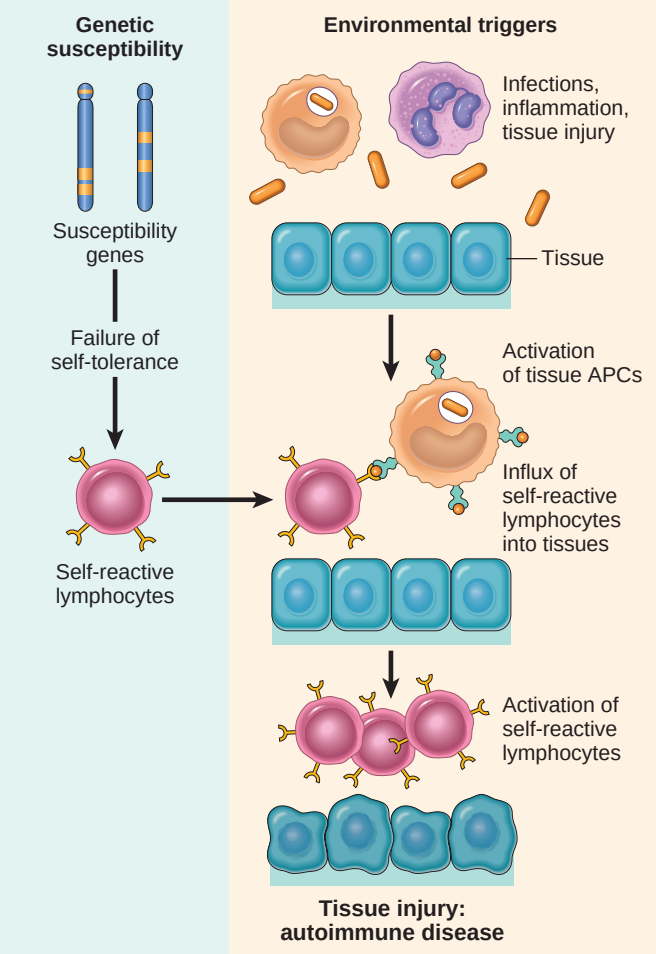


Figure 6.22 Pathogenesis of autoimmunity. Autoimmunity results from multiple factors, including susceptibility genes that may interfere with self-tolerance and environmental triggers (such as infections, tissue injury, and inflammation) that promote lymphocyte entry into tissues, activation of self-reactive lymphocytes, and tissue damage.

Association of HLA Alleles With Disease

Among the genes known to be associated with autoimmunity, the greatest contribution is that of HLA genes (Table 6.7). The most striking of these associations is between ankylosing spondylitis and *HLA-B27;* individuals who inherit this class I HLA allele have a 100- to 200-fold greater chance (odds ratio, or relative risk) of developing the disease compared with those who do not carry *HLA-B27.* Many autoimmune diseases are associated with different class II *HLA* alleles. Although it is reasonable to postulate that these associations reflect the ability of some HLA molecules to preferentially display natural or modified self peptides, it has been difficult to prove that this is the case. It is also important to understand that although different HLA alleles may contribute to a disease, their presence is not, by itself, the cause of any disease. Thus, in the example of HLA-B27,

Table 6.7 Association of HLA Alleles and Inflammatory Diseases

Disease	HLA Allele	Odds Ratio[a]
Rheumatoid arthritis (anti-CCP Ab positive)[b]	DRB1, 1 SE allele[c] DRB1, 2 SE alleles	4 12
Type 1 diabetes	DRB1*0301-DQA1*0501-DQB1*0201 haplotype DRB1*0401-DQA1*0301-DQB1*0302 haplotype DRB1*0301/0401 haplotype heterozygotes	4 8 35
Multiple sclerosis	DRB1*1501	3
Systemic lupus erythematosus	DRB1*0301 DRB1*1501	2 1.3
Ankylosing spondylitis	B*27 (mainly B*2705 and B*2702)	100–200
Celiac disease	DQA1*0501-DQB1*0201 haplotype	7

[a]The odds ratio reflects approximate values of increased risk of the disease associated with the inheritance of particular HLA alleles. The data are from European-derived populations.

[b]Anti-CCP Ab = antibodies directed against cyclic citrullinated peptides. Data are from patients who test positive for these antibodies in the serum.

[c]SE refers to *shared epitope*, so called because the susceptibility alleles map to one region of the DRB1 protein (positions 70–74).

Courtesy Dr. Michelle Fernando, Imperial College London.

the vast majority of individuals who inherit this allele never develop ankylosing spondylitis.

Association of Non-MHC Genes With Autoimmune Diseases

Genome-wide association studies and family studies have shown that multiple non-MHC genes are associated with various autoimmune diseases (Table 6.8). Some of these genes are disease-specific, but many of the associations are seen in multiple disorders, suggesting that the products of these genes affect general mechanisms of immune regulation and self-tolerance. Three recently described genetic associations are especially interesting.

- Polymorphisms in a gene called *PTPN22,* which encodes a protein tyrosine phosphatase, are associated with rheumatoid arthritis, type 1 diabetes, and several other autoimmune diseases. Because these disorders have a fairly high prevalence (especially rheumatoid arthritis), *PTPN22* is said to be the non-HLA gene that is most frequently implicated in autoimmunity. It is postulated that the disease-associated variants encode forms of the phosphatase that are functionally defective and thus unable to dampen the activity of tyrosine kinases that are involved in many responses of lymphocytes. The net result is excessive lymphocyte activation.
- Polymorphisms in the gene for NOD2 are associated with Crohn disease, a form of inflammatory bowel disease, especially in certain ethnic populations. NOD2, a member of the NOD-like receptor (NLR) family (discussed earlier), is a cytoplasmic sensor of microbes that is expressed in intestinal epithelial and other cells. According to one hypothesis, the disease-associated variant is ineffective at sensing gut microbes, including commensal bacteria, resulting in entry of and chronic inflammatory responses against these normally well-tolerated organisms.
- Polymorphisms in the gene encoding the IL-2 receptor (CD25) are associated with multiple sclerosis and other autoimmune diseases. These variants may influence the expression and activity of this key receptor and thereby affect the balance between regulatory and effector T cells.

Many other polymorphisms have been described in different autoimmune diseases, and we will mention some of these when we describe specific disorders. We have previously mentioned that in mice and humans, gene

Table 6.8 Selected Non–HLA Genes Associated With Autoimmune Diseases

Putative Gene Involved	Diseases	Postulated Function of Encoded Protein and Role of Mutation/Polymorphism in Disease
Genes Involved in Immune Regulation		
PTPN22	RA, T1D, IBD	Protein tyrosine phosphatase, may affect signaling in lymphocytes and may alter negative selection or activation of self-reactive T cells
IL23R	IBD, PS, AS	Receptor for the Th17-inducing cytokine IL-23; may alter differentiation of CD4+ T cells into pathogenic Th17 effector cells
CTLA4	T1D, RA	Inhibits T-cell responses by terminating activation and promoting activity of regulatory T cells; may interfere with self-tolerance
IL2RA	MS, T1D	α chain of the receptor for IL-2, which is a growth and survival factor for activated and regulatory T cells; may affect development of effector cells and/or regulation of immune responses
Genes Involved in Immune Responses to Microbes		
NOD2	IBD	Cytoplasmic sensor of bacteria expressed in Paneth and other intestinal epithelial cells; may control resistance to gut commensal bacteria
ATG16	IBD	Involved in autophagy; possible role in defense against microbes and maintenance of epithelial barrier function
IRF5, IFIH1	SLE	Role in type I interferon production; type I interferon is involved in the pathogenesis of SLE (see text)

AS, Ankylosing spondylitis; *IBD,* inflammatory bowel disease; *MS,* multiple sclerosis; *PS,* psoriasis; *RA,* rheumatoid arthritis; *SLE,* systemic lupus erythematosus.

The probable linkage of these genes with various autoimmune diseases has been defined by genome-wide association studies and other methods for studying disease-associated polymorphisms.

Modified from Zenewicz LA, Abraham C, Flavell RA, Cho JH: Unraveling the genetics of autoimmunity. *Cell* 140:791, 2010.

knockouts and sporadic mutations affecting several individual genes result in autoimmunity. These genes include *AIRE, CTLA4, PD1, FAS, FASL,* and *IL2* and its receptor *CD25*. In addition, B cells express an Fc receptor that recognizes IgG antibodies bound to antigens and switches off further antibody production (a normal negative-feedback mechanism). Knockout of this receptor results in autoimmunity, presumably because the B cells can no longer be controlled. These examples provide valuable information about pathways of self-tolerance and immune regulation, but the diseases caused by these single-gene mutations are rare, and mutations in these genes are not the cause of most common autoimmune disorders.

Role of Infections and Other Environmental Factors

Autoimmune reactions may be triggered by infections. Two mechanisms have been postulated to explain the link between infections and autoimmunity (Fig. 6.23). First, infections may upregulate the expression of costimulators on APCs. If these cells are presenting self antigens, the result may be a breakdown of anergy and activation of T cells specific for the self antigens. Second, some microbes may express antigens that share amino acid sequences with self antigens. Immune responses against the microbial antigens may result in the activation of self-reactive lymphocytes. This phenomenon is called *molecular mimicry*. A clear example of such mimicry is rheumatic heart disease, in which antibodies against streptococcal proteins cross-react with myocardial proteins and cause myocarditis (Chapter 12). More subtle molecular mimicry may be involved in classic autoimmune diseases as well.

Microbes may induce other abnormalities that promote autoimmune reactions. Some viruses, such as EBV and HIV, cause polyclonal B-cell activation, which may result in production of autoantibodies. The tissue injury that is common in infections may release and structurally modify self antigens, creating neoantigens that are able to activate T cells. Infections may induce the production of cytokines that recruit lymphocytes, including potentially self-reactive lymphocytes, to sites of self antigens.

Infections may protect against some autoimmune diseases. Although the role of infections in triggering autoimmunity has received a great deal of attention, epidemiologic studies suggest that the incidence of autoimmune diseases is increasing in high income countries in parallel with better infection control. In some animal models (e.g., of type 1 diabetes) infections greatly reduce the incidence of disease. The underlying mechanisms are unclear; one possibility is that infections promote low-level IL-2 production, and this is essential for maintaining regulatory T cells.

Recently, there has been great interest in the idea that the normal gut and skin microbiome influences the development of autoimmunity. It is possible that different nonpathogenic microbes affect the relative proportions of effector and regulatory T cells and shape the host response toward or away from aberrant activation. However, it is still not clear which microbes actually contribute to specific diseases in humans, or if the microbiome can be manipulated to prevent or treat these disorders.

In addition to infections, the display of tissue antigens also may be altered by a variety of environmental insults. As discussed later, ultraviolet (UV) radiation causes cell death and may lead to the exposure of nuclear antigens, which elicit pathologic immune responses in lupus; this mechanism is the proposed explanation for the association of lupus flares with exposure to sunlight. Smoking is a risk factor for rheumatoid arthritis, perhaps because it leads to chemical modification of self antigens. Local tissue injury for any reason may lead to the release of self antigens and autoimmune responses.

Finally, autoimmunity has a strong gender bias, with many of these diseases being more common in women than

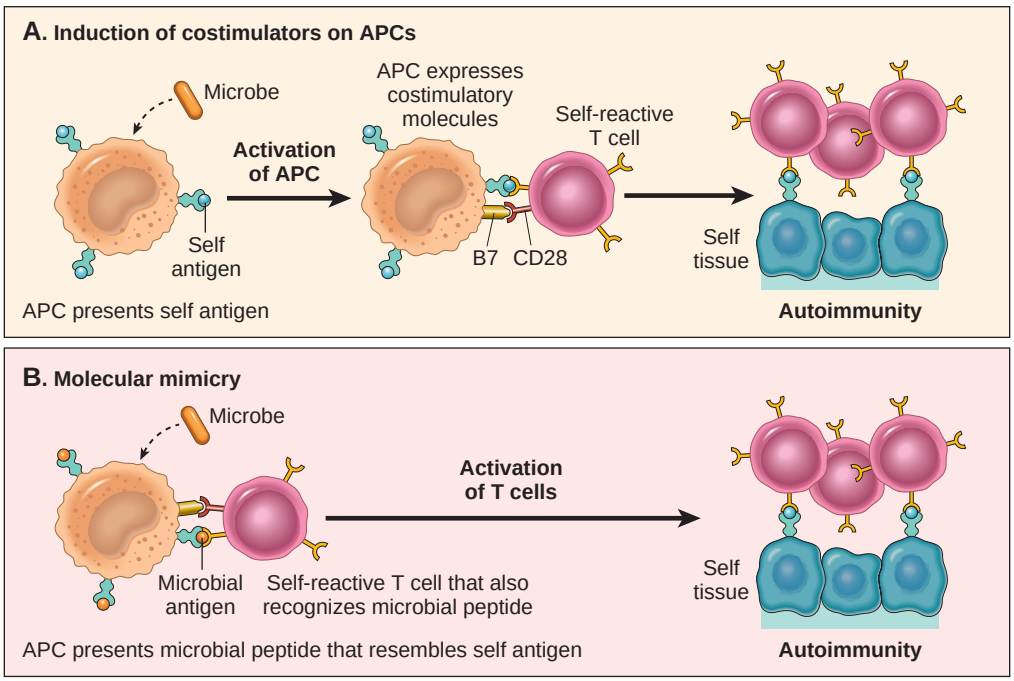


Figure 6.23 Postulated role of infections in autoimmunity. Infections may promote activation of self-reactive lymphocytes by inducing the expression of costimulators (A), or microbial antigens may mimic self antigens and activate self-reactive lymphocytes as a cross-reaction (B).

in men. The underlying mechanisms are not understood but may involve the effects of hormones and currently unknown genes on the X chromosome.

General Features of Autoimmune Diseases

Diseases caused by autoimmunity have some important general features.

- Autoimmune diseases tend to be chronic, sometimes with relapses and remissions, and the damage is often progressive. One reason for the chronicity is that the immune system contains many intrinsic amplification loops that allow small numbers of antigen-specific lymphocytes to accomplish the task of eradicating complex infections. When the response is inappropriately directed against self tissues, the same amplification mechanisms exacerbate and prolong the injury. Another reason for the persistence and progression of autoimmune disease is the phenomenon of *epitope spreading,* in which an immune response against one self antigen causes tissue damage, releasing other antigens, and resulting in the activation of lymphocytes that recognize these newly encountered epitopes.
- The clinical and pathologic manifestations of an autoimmune disease are determined by the nature of the underlying immune response. Some of these diseases are caused by autoantibodies, whose formation may be associated with dysregulated germinal center reactions. Most chronic inflammatory diseases are caused by abnormal and excessive Th1 and Th17 responses; examples of these diseases include psoriasis, multiple sclerosis, and some types of inflammatory bowel disease. In some autoimmune diseases, such as rheumatoid arthritis, both antibodies and T cell–mediated inflammation may be involved.

With this background, we can proceed to a discussion of specific autoimmune diseases. Table 6.6 lists both systemic and organ-specific autoimmune disorders. The systemic diseases tend to involve blood vessels and connective tissues, and therefore, they are often called *collagen vascular diseases* or *connective tissue diseases*. Our focus here is on selected systemic autoimmune diseases; organ-specific disorders are covered in relevant chapters.

KEY CONCEPTS

IMMUNOLOGIC TOLERANCE AND AUTOIMMUNITY

- Tolerance (unresponsiveness) to self antigens is a fundamental property of the immune system, and breakdown of tolerance is the basis of autoimmune diseases.
- Central tolerance: immature lymphocytes that recognize self antigens in the central (generative) lymphoid organs are killed by apoptosis; in the B-cell lineage, some of the self-reactive lymphocytes switch to new antigen receptors that are not self-reactive.
- Peripheral tolerance: mature lymphocytes that recognize self antigens in peripheral tissues become functionally inactive (anergic) or are suppressed by regulatory T lymphocytes, or they die by apoptosis.
- The factors that lead to a failure of self-tolerance and the development of autoimmunity include (1) inheritance of susceptibility genes that may disrupt different tolerance pathways and (2) infections and tissue injury that may expose self antigens and activate APCs and lymphocytes in the tissues.
- Autoimmune diseases are usually chronic and progressive, and the type of tissue injury is determined by the nature of the dominant immune response.

Systemic Lupus Erythematosus (SLE)

SLE is an autoimmune disease involving multiple organs, characterized by a vast array of autoantibodies, particularly antinuclear antibodies (ANAs), in which injury is caused mainly by deposition of immune complexes and binding of antibodies to various cells and tissues. The disease may be acute or insidious in its onset, and is typically a chronic, remitting and relapsing, often febrile, illness. Injury to the skin, joints, kidney, and serosal membranes is most prominent, but virtually any organ in the body may be affected. Because of this, the disease is very heterogeneous, and any patient may present with any number of clinical features. In recognition of this, the American College of Rheumatology has established a complex set of criteria for this disorder, which is helpful for clinicians and for the assessment of patients in clinical trials (Table 6.9).

SLE is a fairly common disease, with a prevalence that may be as high as 1 in 2500 in certain populations. Similar to many autoimmune diseases, SLE predominantly affects women, with a frequency of 1 in 700 among women of childbearing age and a female-to-male ratio of 9:1 in the reproductive age group of 17 to 55 years. By comparison, the female-to-male ratio is only 2:1 for disease developing during childhood or after 65 years of age. The prevalence of the disease is twofold to threefold higher in blacks and Hispanics than in whites. Although SLE most often presents in the twenties and thirties, it may manifest at any age, even in early childhood.

Spectrum of Autoantibodies in Systemic Lupus Erythematosus

The hallmark of SLE is the production of autoantibodies, several of which (antibodies to double-stranded DNA and the so-called Smith [Sm] antigen) are virtually diagnostic. These and other autoantibodies are pathogenic, either by forming immune complexes or by attacking their target cells. The levels of these autoantibodies in the blood are also helpful for the diagnosis and management of patients with SLE. Autoantibodies are found in many diseases in addition to SLE, and antibodies of different specificities tend to be associated with different autoimmune disorders (Table 6.10).

ANAs. These are directed against nuclear antigens and can be grouped into four categories based on their specificity for: (1) DNA, (2) histones, (3) nonhistone proteins bound to RNA, and (4) nucleolar antigens. Table 6.10 lists several ANAs and their association with SLE as well as with other autoimmune diseases to be discussed later. The most widely used method for detecting ANAs is indirect immunofluorescence, which can identify antibodies that bind to a variety of nuclear antigens, including DNA, RNA, and proteins (collectively called generic ANAs). The pattern of nuclear fluorescence suggests the type of antibody present in the patient's serum. Four basic patterns are recognized (Fig. 6.24):

Table 6.9 1997 Revised Criteria for Classification of Systemic Lupus Erythematosus[a]

Criterion	Definition
Clinical Criteria	
Acute cutaneous lupus	Malar rash (fixed erythema, flat or raised, over the malar eminences), photosensitivity
Chronic cutaneous lupus	Discoid rash: erythematous raised patches with adherent keratotic scaling and follicular plugging
Nonscarring alopecia	Diffuse thinning or hair fragility in the absence of other causes
Oral or nasal ulcers	Oral or nasopharyngeal ulceration, usually painless
Joint disease	Nonerosive synovitis involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
Serositis	Pleuritis (pleuritic pain or rub or evidence of pleural effusion), pericarditis
Renal disorder	Persistent proteinuria >0.5 g/24 hours, or red cell casts
Neurologic disorder	Seizures, psychosis, myelitis, or neuropathy, in the absence of offending drugs or other known causes
Hemolytic anemia	Hemolytic anemia
Leukopenia or lymphopenia	Leukopenia: $<4.0 \times 10^9$ cells/L (4000 cells/mm^3) total on two or more occasions, *or* Lymphopenia: $<1.5 \times 10^9$ cells/L (1500 cells/mm^3) on two or more occasions
Thrombocytopenia	Thrombocytopenia: $<100 \times 10^9$ cells/L (100×10^3 cells/mm^3) in the absence of offending drugs and other conditions
Immunologic Criteria	
Antinuclear antibody (ANA)	Abnormal titer of antinuclear antibody by immunofluorescence
Anti-dsDNA antibody	Abnormal titer
Anti-Sm antibody	Presence of antibody to Sm nuclear antigen
Antiphospholipid antibody	Positive finding of antiphospholipid antibodies based on (1) an abnormal serum level of IgG or IgM anti-cardiolipin antibodies, (2) a positive test for lupus anticoagulant using a standard test, or (3) a false-positive serologic test for syphilis known to be positive for at least 6 months and confirmed by negative *Treponema pallidum* immobilization or fluorescent treponemal antibody absorption test
Low complement	Low C3, C4, or CH50
Direct Coombs test	Assay for anti–red cell antibody, in the absence of clinically evident hemolytic anemia

[a]This classification was initially proposed in 1997 by the American College of Rheumatology for the purpose of identifying patients in clinical studies. It was updated in 2012 by the Systemic Lupus International Collaborating Clinics. A patient is classified as having SLE if four of the clinical and immunologic criteria are present at any time (not necessarily concurrently), including at least one clinical and one immunologic criterion. Some details have been omitted from the table.

Modified from Petri M, Orbai AM, Alarcón GS, et al: Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus, *Arthritis Rheum* 64:2677, 2012.

Table 6.10 Autoantibodies in Systemic Autoimmune Diseases

Disease	Specificity of Autoantibody	% Positive	Association With Specific Disease Features
Systemic lupus erythematosus (SLE)	Double-stranded DNA	40–60	Nephritis; specific for SLE
	U1-RNP	30–40	
	Smith (Sm) antigen (core protein of small RNP particles)	20–30	Specific for SLE
	Ro (SS-A)/La (SS-B) nucleoproteins	30–50	Congenital heart block; neonatal lupus
	Phospholipid-protein complexes (anti-PL)	30–40	Antiphospholipid syndrome (in ~10% of SLE patients)
	Multiple nuclear antigens ("generic ANAs")	95–100	Found in other autoimmune diseases, not specific
Systemic sclerosis	DNA topoisomerase I	30–70	Diffuse skin disease, lung disease; specific for systemic sclerosis
	Centromeric proteins (CENPs) A, B, C	20–40	Limited skin disease, ischemic digital loss, pulmonary hypertension
	RNA polymerase III	15–20	Acute onset, scleroderma renal crisis, cancer
Sjögren syndrome	Ro/SS-A La/SS-B	70–95	
Autoimmune myositis	Histidyl aminoacyl-tRNA synthetase, Jo1	25	Interstitial lung disease, Raynaud phenomenon
	Mi-2 nuclear antigen	5–10	Dermatomyositis, skin rash
	MDA5 (cytoplasmic receptor for viral RNA)	20–35 (Japanese)	Vascular skin lesions, interstitial lung disease
	TIF1γ nuclear protein	15–20	Dermatomyositis, cancer
Rheumatoid arthritis	CCP (cyclic citrullinated peptides); various citrullinated proteins	60–80	Specific for rheumatoid arthritis
	Rheumatoid factor (not specific)	60–70	

"Generic" antinuclear antibodies (ANAs), which may react against many nuclear antigens, are positive in a large fraction of patients with SLE but are also positive in other autoimmune diseases. % positive refers to the approximate % of patients who test positive for each antibody.

The table was compiled with the help of Dr. Antony Rosen, Johns Hopkins University, Baltimore, Md.

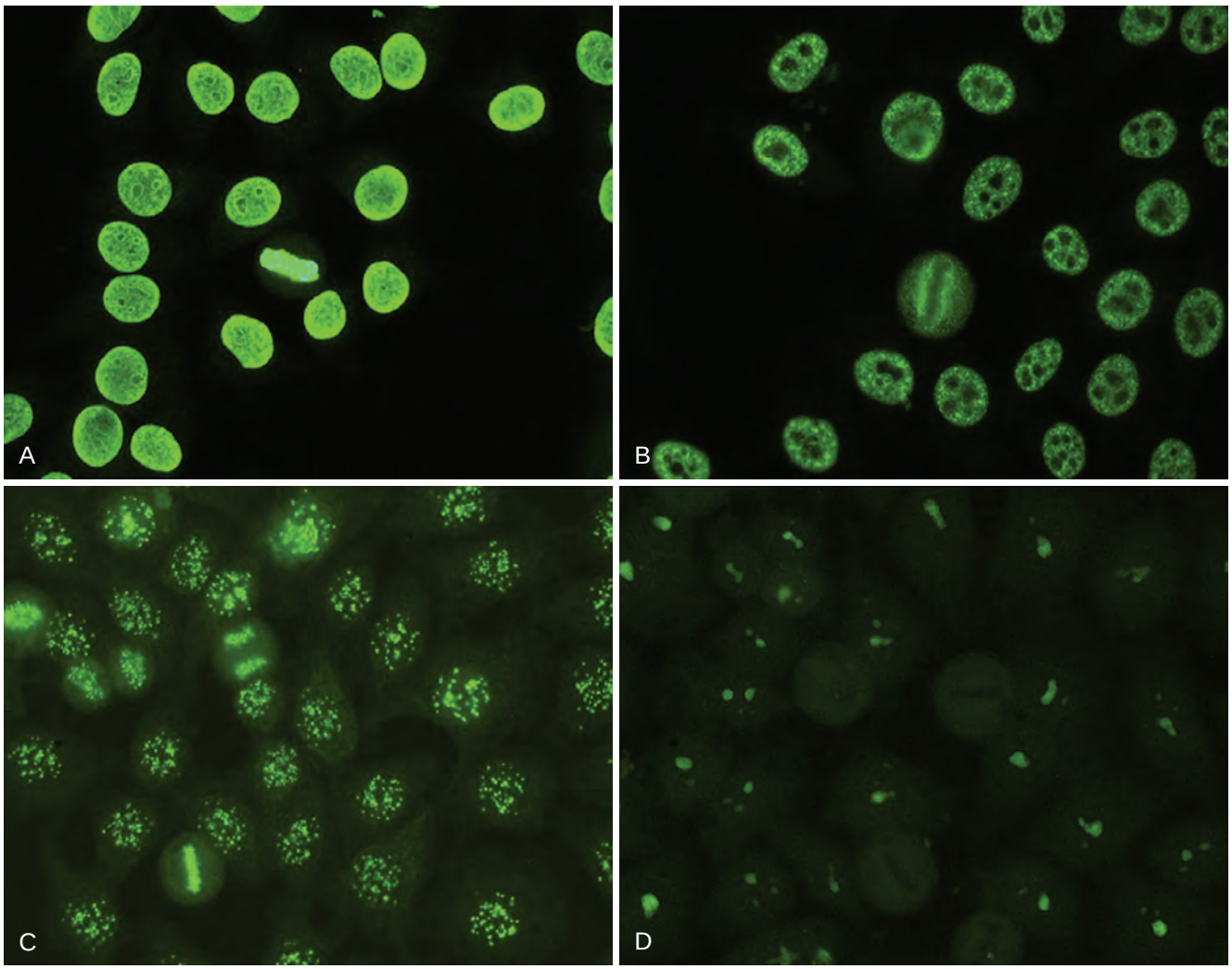


Figure 6.24 Staining patterns of antinuclear antibodies. (A) Homogeneous or diffuse staining of nuclei is typical of antibodies reactive with dsDNA, nucleosomes, and histones, and is common in systemic lupus erythematosus. (B) Speckled pattern is seen with antibodies against various nuclear antigens, including Sm and RNPs. (C) The pattern of staining of anti-centromere antibodies is seen in some cases of systemic sclerosis, Sjögren syndrome, and other diseases. (D) Nucleolar pattern is typical of antibodies against nucleolar proteins. (Images reproduced with permission from Wiik AS, Høier-Madsen M, Forslid J, et al: Antinuclear antibodies: a contemporary nomenclature using Hep-2 cells, *J Autoimm* 35:276–290, 2010.)

- *Homogeneous or diffuse nuclear staining* usually reflects antibodies to chromatin, histones, and, occasionally, double-stranded DNA.
- *Rim or peripheral staining* patterns are most often indicative of antibodies to double-stranded DNA and sometimes to nuclear envelope proteins.
- *Speckled pattern* refers to the presence of uniform or variable-sized speckles. This is one of the most commonly observed patterns of fluorescence and therefore the least specific. It reflects the presence of antibodies to non-DNA nuclear constituents such as Sm antigen, ribonucleoprotein, and SS-A and SS-B reactive antigens.
- *Nucleolar pattern* refers to the presence of a few discrete spots of fluorescence within the nucleus and represents antibodies to RNA. This pattern is reported most often in patients with systemic sclerosis.
- *Centromeric pattern.* Patients with systemic sclerosis often contain antibodies specific for centromeres, which give rise to this pattern.

The fluorescence patterns are not absolutely specific for the type of antibody, and because many autoantibodies may be present, combinations of patterns are frequent. Concerns have been raised about the sensitivity and subjective nature of this assay, and attempts are ongoing to replace it with ELISAs (enzyme linked immunosorbent assays) for specific nuclear and other antigens. Nevertheless, the staining pattern is considered of diagnostic value, and the test remains in use.

In addition to ANAs, lupus patients have a host of other autoantibodies. Some are directed against blood cells, such as red cells, platelets, and lymphocytes; others react with proteins in complex with phospholipids. *Antiphospholipid antibodies* are present in 30% to 40% of lupus patients. They are actually specific for epitopes of plasma proteins that are revealed when the proteins are in complex with phospholipids. Included among these proteins are prothrombin, annexin V, β_2-glycoprotein I, protein S, and protein C. Antibodies against the phospholipid-β_2-glycoprotein complex

also bind to cardiolipin antigen, used in syphilis serology, and therefore lupus patients may have a false-positive test result for syphilis. Some of these antibodies interfere with in vitro clotting tests, such as partial thromboplastin time. Therefore, these antibodies are sometimes referred to as *lupus anticoagulant*. Despite the observed clotting delays in vitro, however, patients with antiphospholipid antibodies have complications related to excessive clotting (a hypercoagulable state), such as thrombosis (Chapter 4).

Pathogenesis

The fundamental defect in SLE is a failure of the mechanisms that maintain self-tolerance. Although what causes this failure of self-tolerance remains unknown, as is true of most autoimmune diseases, both genetic and environmental factors play a role.

Genetic Factors. SLE is a genetically complex disease with contributions from MHC and multiple non-MHC genes. Many lines of evidence support a genetic predisposition.

- Family members of patients have an increased risk of developing SLE. In addition, laboratory testing has revealed that as many as 20% of clinically unaffected first-degree relatives of SLE patients have autoantibodies and other immune abnormalities.
- There is a higher rate of concordance (>20%) in monozygotic twins when compared with dizygotic twins (1% to 3%).
- Studies of HLA associations support the concept that MHC genes regulate production of particular autoantibodies. Specific alleles of the *HLA-DQ* locus have been linked to the production of anti–double-stranded DNA, anti-Sm, and antiphospholipid antibodies, although the relative risk is small.
- Some lupus patients have inherited deficiencies of early complement components, such as C2, C4, or C1q. Lack of complement may impair removal of circulating immune complexes by the mononuclear phagocyte system, thus favoring tissue deposition. Knockout mice lacking C4 or certain complement receptors are also prone to develop lupus-like autoimmunity. Various mechanisms have been invoked, including failure to clear immune complexes and loss of B-cell self-tolerance. It has also been proposed that deficiency of C1q results in defective phagocytic clearance of apoptotic cells. Many cells normally undergo apoptosis, and if they are not cleared their nuclear components may elicit immune responses.
- Genome-wide association studies have identified several genetic loci that may be associated with the disease. Many of these loci encode proteins involved in lymphocyte signaling and interferon responses, both of which may play a role in lupus pathogenesis, as discussed later. The relative risk for each locus is small, and even taken together these loci account for 20% or less of the disease predisposition, suggesting an important role for environmental factors, discussed later.

Immunologic Factors. Recent studies in animal models and patients have revealed several immunologic aberrations that collectively may result in the persistence and uncontrolled activation of self-reactive lymphocytes.

- Failure of self-tolerance in B cells results from defective elimination of self-reactive B cells in the bone marrow or defects in peripheral tolerance mechanisms.
- Activation of CD4+ helper T cells specific for nucleosomal antigens that escape tolerance and help B cells to produce high-affinity pathogenic autoantibodies.
- TLR engagement by nuclear DNA and RNA contained in immune complexes may activate B lymphocytes. These TLRs function normally to sense microbial products, including nucleic acids. Thus, B cells specific for nuclear antigens may get second signals from TLRs and may be activated, resulting in increased production of antinuclear autoantibodies.
- Type I interferons play a role in lymphocyte activation in SLE. High levels of circulating type I interferons and a molecular signature in blood cells suggesting exposure to these cytokines has been reported in SLE patients and correlates with disease severity. Type I interferons are antiviral cytokines that are normally produced during innate immune responses to viruses. It may be that nucleic acids engage TLRs on DCs and stimulate the production of interferons. In other words, self nucleic acids mimic their microbial counterparts. How interferons contribute to the development of SLE is unclear; these cytokines may activate DCs and B cells and promote Th1 responses, all of which may stimulate the production of pathogenic autoantibodies and induce inflammation.

Environmental Factors. There are many indications that environmental factors must also be involved in the pathogenesis of SLE.

- Exposure to ultraviolet (UV) light exacerbates the disease in many individuals. UV irradiation may induce apoptosis in cells and may alter the DNA in such a way that its recognition by TLRs is enhanced. In addition, UV light may modulate the immune response, for example, by stimulating keratinocytes to produce IL-1, a cytokine known to promote inflammation.
- The gender bias of SLE is partly attributable to actions of sex hormones and partly related to unknown genes on the X chromosome, independent of hormone effects.
- Drugs such as hydralazine, procainamide, and D-penicillamine can induce an SLE-like response in humans.

A Model for the Pathogenesis of SLE. It is clear from this discussion that the immunologic abnormalities in SLE—both documented and postulated—are varied and complex. Nevertheless, an attempt can be made to synthesize results from patients and animal studies into a model of the pathogenesis of SLE (Fig. 6.25). UV irradiation and other environmental insults lead to the apoptosis of cells. Inadequate clearance of the nuclei of these cells results in a large burden of nuclear antigens. Underlying abnormalities in B and T lymphocytes are responsible for defective tolerance, because of which self-reactive lymphocytes survive and remain functional. These lymphocytes are stimulated by nuclear self antigens, and antibodies are produced against the antigens. Complexes of the antigens and antibodies bind to Fc receptors on B cells and DCs, and may be internalized. The nucleic acid components engage TLRs and stimulate B cells to produce more autoantibodies. TLR stimuli also activate DCs to produce interferons and other cytokines,

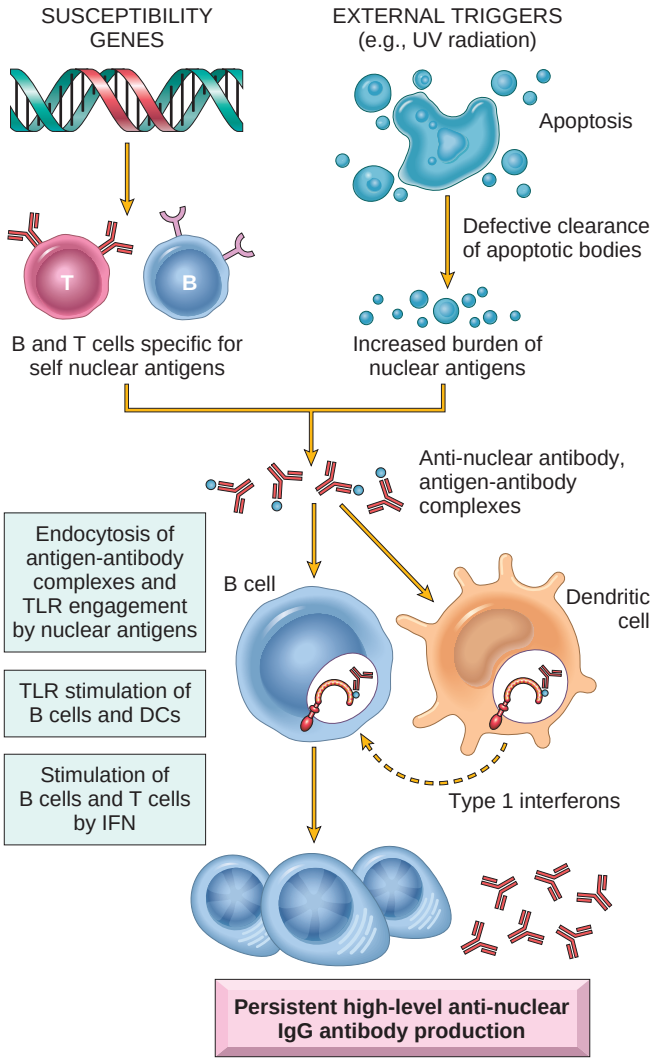


Figure 6.25 Model for the pathogenesis of systemic lupus erythematosus. In this hypothetical model, susceptibility genes interfere with the maintenance of self-tolerance, and external triggers lead to persistence of nuclear antigens. The result is an antibody response against self nuclear antigens, which is amplified by the action of nucleic acids on dendritic cells *(DCs)* and B cells, and the production of type I interferons. *IFN,* Interferon; *TLRs,* Toll-like receptors.

which further enhance the immune response and cause more apoptosis. The net result is a cycle of antigen release and immune activation resulting in the production of high-affinity autoantibodies.

Mechanisms of Tissue Injury. Different autoantibodies are the cause of most of the lesions of SLE.

- Most of the systemic lesions are caused by *immune complexes (type III hypersensitivity).* DNA–anti-DNA complexes can be detected in the glomeruli and small blood vessels in animal models. Low levels of serum complement (secondary to consumption of complement proteins) and granular deposits of complement and immunoglobulins in the glomeruli further support the role of immune complex deposition. T-cell infiltrates are also frequently seen in the kidneys and may be involved in tissue damage. There is no evidence that ANAs, which are present in immune complexes, can penetrate intact cells. If cell nuclei are exposed, however, the ANAs can bind to them. In tissues, nuclei of damaged cells react with ANAs, lose their chromatin pattern, and become homogeneous, to produce so-called LE bodies or hematoxylin bodies. Related to this phenomenon is the *LE cell,* which is readily seen when blood is agitated in vitro. The LE cell is any phagocytic leukocyte (blood neutrophil or macrophage) that has engulfed the denatured nucleus of an injured cell. The demonstration of LE cells in vitro was used in the past as a test for SLE. With new techniques for detection of ANAs, however, this test is now largely of historical interest. Sometimes, LE cells are found in pericardial or pleural effusions in patients.
- Autoantibodies specific for red blood cells, white blood cells, and platelets opsonize these cells and promote their phagocytosis and destruction, resulting in cytopenias. These are examples of antibody-mediated (type II) hypersensitivity. The most common disorder caused by this type of autoantibody is immune thromobocytopenic purpura (ITP; Chapter 14), which occurs in up to 10% of patients with SLE. ITP is most commonly due to autoantibodies that bind platelet membrane glycoproteins, leading to removal of platelets by macrophages, particularly in splenic sinusoids.
- *Antiphospholipid antibody syndrome.* Patients with antiphospholipid antibodies may develop venous and arterial thromboses, which may be associated with recurrent spontaneous miscarriages and focal cerebral or ocular ischemia. This constellation of clinical features, in association with lupus, is referred to as the secondary antiphospholipid antibody syndrome. The mechanisms of thrombosis are not defined, and antibodies against clotting factors, platelets, and endothelial cells have all been proposed as being responsible for thrombosis (Chapter 4). Some patients develop these autoantibodies and the clinical syndrome without associated SLE. They are said to have the primary antiphospholipid antibody syndrome (Chapter 4).
- The *neuropsychiatric manifestations* of SLE have been attributed to antibodies that cross the blood-brain barrier and react with neurons or receptors for various neurotransmitters. Other immune factors, such as cytokines, may also be involved in the cognitive dysfunction and CNS abnormalities that are associated with SLE.

MORPHOLOGY

The morphologic changes in SLE are extremely variable. The frequency of individual organ involvement is shown in Table 6.11. The most characteristic lesions result from immune complex deposition in blood vessels, kidneys, connective tissue, and skin.

Blood Vessels. An acute necrotizing vasculitis involving capillaries, small arteries, and arterioles may be present in any tissue. The arteritis is characterized by fibrinoid necrosis of the vessel walls. In chronic stages, vessels undergo fibrous thickening with luminal narrowing.

Kidney. Up to 50% of SLE patients have clinically significant renal involvement mainly in the form of glomerulonephritis and tubulointerstitial nephritis. The glomerular lesions are the result

Table 6.11 Clinical and Pathologic Manifestations of Systemic Lupus Erythematosus

Clinical Manifestation	Prevalence in Patients (%)[a]
Hematologic	100
Arthritis, arthralgia or myalgia	80–90
Skin	85
Fever	55–85
Fatigue	80–100
Weight loss	60
Renal	50–70
Neuropsychiatric	25–35
Pleuritis	45
Pericarditis	25
Gastrointestinal	20
Raynaud phenomenon	15–40
Ocular	5–15
Peripheral neuropathy	15

[a]Percentages are approximate and may vary with age, ethnicity, and other factors. Table compiled with the assistance of Dr. Meenakshi Jolly, Rush Medical Center, Chicago, Ill.

of deposition of immune complexes on the glomerular basement membrane, in the mesangium, and sometimes throughout the glomerulus. According to the currently accepted classification, six patterns of glomerular disease are seen in SLE. It should be noted that there is some overlap between these classes, and over time lesions may evolve from one class to another. Thus, the exact percentage of patients with each of the six classes of lesions is difficult to determine. Suffice it to say that class I is the least common and class IV is the most common pattern.

- *Minimal mesangial lupus nephritis* (class I) is very uncommon and is characterized by immune complex deposition in the mesangium, identified by immunofluorescence and by electron microscopy, but without structural changes by light microscopy.
- *Mesangial proliferative lupus nephritis* (class II) is characterized by mesangial cell proliferation, often accompanied by accumulation of mesangial matrix, and granular mesangial deposits of immunoglobulin and complement without involvement of glomerular capillaries.
- *Focal lupus nephritis* (class III) is defined by involvement of fewer than 50% of glomeruli. The lesions may be segmental (affecting only a portion of the glomerulus) or global (involving the entire glomerulus). Affected glomeruli may exhibit swelling and proliferation of endothelial and mesangial cells, leukocyte accumulation, capillary necrosis, and hyaline thrombi. There is also often extracapillary proliferation associated with focal necrosis and crescent formation (Fig. 6.26A). The clinical presentation ranges from mild hematuria and proteinuria to acute renal insufficiency. Red blood cell casts in the urine are common when the disease is active. Some patients progress to diffuse glomerulonephritis. The active (or proliferative) inflammatory lesions can heal completely or lead to chronic global or segmental glomerular scarring.
- *Diffuse lupus nephritis* (class IV) is the most common and severe form of lupus nephritis. The lesions are similar to those in class III, but differ in extent; typically, in class IV nephritis half or more of the glomeruli are affected. As in class III, the lesions may be segmental or global and on the basis of this, it can be subclassified as Class IV segmental (IV-S) or Class IV global (IV-G). Involved glomeruli show proliferation of endothelial, mesangial, and epithelial cells (Fig. 6.26B), with the latter producing cellular crescents that fill Bowman's space (Chapter 20). Subendothelial immune complex deposits may cause circumferential thickening of the capillary wall, forming "wire loop" structures on light microscopy (Fig. 6.26C). Immune complexes can be readily detected by electron microscopy (Fig. 6.26D) and immunofluorescence (Fig. 6.26E). Lesions may progress to scarring of glomeruli. Patients with diffuse glomerulonephritis are usually symptomatic, showing hematuria as well as proteinuria. Hypertension and mild to severe renal insufficiency are also common.
- *Membranous lupus nephritis* (class V) is characterized by diffuse thickening of the capillary walls due to deposition of subepithelial immune complexes, similar to idiopathic membranous nephropathy, described in Chapter 20. The immune complexes are usually accompanied by increased production of basement membrane–like material. This lesion typically causes severe proteinuria or nephrotic syndrome and may occur concurrently with focal or diffuse lupus nephritis.
- *Advanced sclerosing lupus nephritis* (class VI) is characterized by sclerosis of more than 90% of the glomeruli and represents end-stage renal disease.

Changes in the interstitium and tubules are frequently present in lupus nephritis. Rarely, **tubulointerstitial lesions** may be the dominant abnormality. Immune complexes similar to those in glomeruli are present in the tubular or peritubular capillary basement membranes in many lupus nephritis patients. Occasionally, there are well-organized B-cell follicles in the interstitium, with plasma cells that may be sources of autoantibodies.

Skin. Characteristic erythema affects the face along the bridge of the nose and cheeks (the "butterfly" rash) in approximately 50% of patients, and a similar rash may be seen on the extremities and trunk. Urticaria, bullae, maculopapular lesions, and ulcerations also occur. Exposure to sunlight incites or accentuates the erythema. Histologically the involved areas show vacuolar degeneration of the basal layer of the epidermis (Fig. 6.27A). In the dermis, there is variable edema and perivascular inflammation. Vasculitis with fibrinoid necrosis may be prominent. Immunofluorescence microscopy shows deposition of immunoglobulin and complement along the dermoepidermal junction (Fig. 6.27B), which may also be present in uninvolved skin. This finding is not specific for SLE and is sometimes seen in scleroderma or dermatomyositis.

Joints. Joint involvement is typically a nonerosive synovitis with little deformity, which contrasts with rheumatoid arthritis.

CNS. No clear morphologic abnormalities account for the neuropsychiatric symptoms of SLE. Noninflammatory occlusion of small vessels by intimal proliferation is sometimes noted, which may be due to endothelial damage by autoantibodies or immune complexes.

Pericarditis and Other Serosal Cavity. Inflammation of the serosal lining membranes may be acute, subacute, or chronic. During the acute phases, the mesothelial surfaces are sometimes covered with fibrinous exudate. Later they become thickened, opaque, and coated with a shaggy fibrous tissue that may lead to partial or total obliteration of the serosal cavity. Pleural and pericardial effusions may be present.

Cardiovascular System. Involvement may manifest as damage to any layer of the heart. Symptomatic or asymptomatic pericardial involvement is present in up to 50% of patients. Myocarditis, or

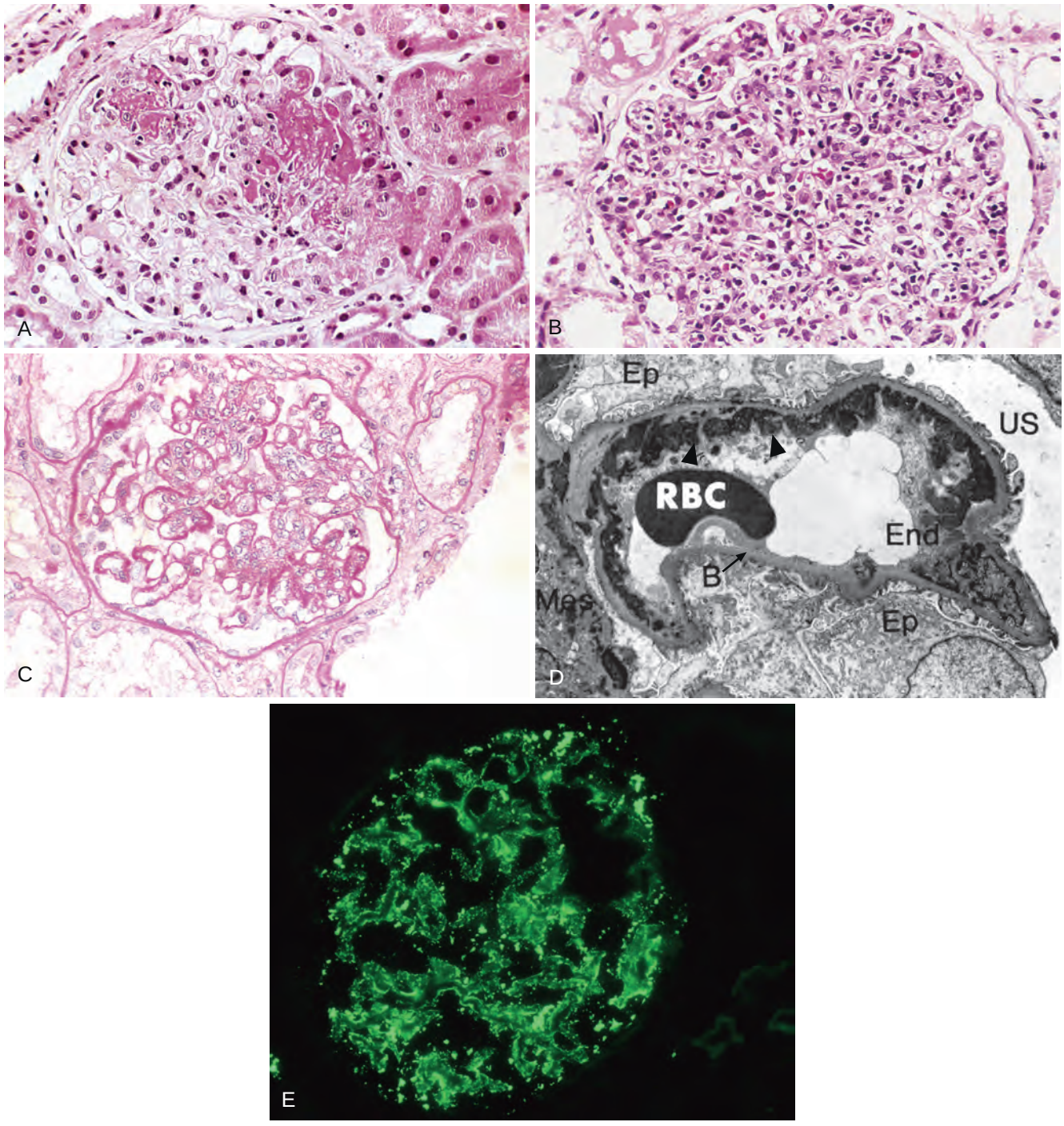


Figure 6.26 Lupus nephritis. (A) Focal proliferative glomerulonephritis, with two focal necrotizing lesions at the 11 o'clock and 2 o'clock positions (H&E stain). Extracapillary proliferation is not prominent in this case. (B) Diffuse proliferative glomerulonephritis. Note the marked increase in cellularity throughout the glomerulus (H&E stain). (C) Lupus nephritis showing a glomerulus with several "wire loop" lesions representing extensive subendothelial deposits of immune complexes (periodic acid-Schiff stain). (D) Electron micrograph of a renal glomerular capillary loop from a patient with lupus nephritis. Subendothelial dense deposits *(arrowheads)* correspond to "wire loops" seen by light microscopy. B *(with arrow)* refers to the basement membrane. (E) Deposition of IgG antibody in a granular pattern, detected by immunofluorescence. *B,* Basement membrane; *End,* endothelium; *Ep,* epithelial cell with foot processes; *Mes,* mesangium; *RBC,* red blood cell in capillary lumen; *US,* urinary space. (A–C, Courtesy Dr. Helmut Rennke, Department of Pathology, Brigham and Women's Hospital, Boston, Mass. D, Courtesy Dr. Edwin Eigenbrodt, Department of Pathology, University of Texas, Southwestern Medical School, Dallas, Tex. E, Courtesy Dr. Jean Olson, Department of Pathology, University of California, San Francisco, Calif.)

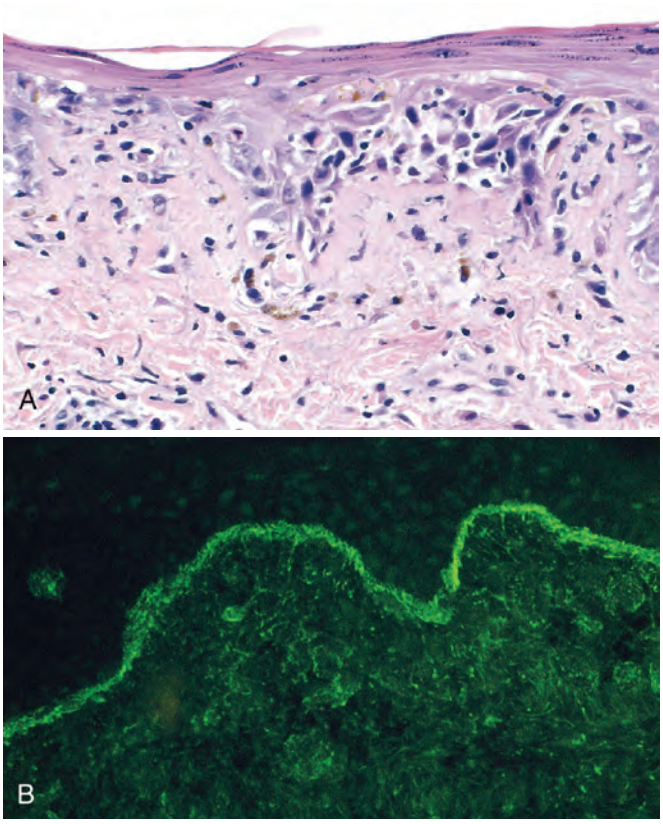


Figure 6.27 Systemic lupus erythematosus involving the skin. (A) An H&E-stained section shows liquefactive degeneration of the basal layer of the epidermis and edema at the dermoepidermal junction. (B) An immunofluorescence micrograph stained for IgG reveals deposits of Ig along the dermoepidermal junction. (A, Courtesy Dr. Jag Bhawan, Boston University School of Medicine, Boston, Mass. B, Courtesy Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Tex.)

mononuclear cell infiltration, is less common and may cause resting tachycardia and electrocardiographic abnormalities. Valvular abnormalities, primarily of the mitral and aortic valves, manifest as diffuse leaflet thickening that may be associated with dysfunction (stenosis and/or regurgitation). Valvular endocarditis (called Libman-Sacks endocarditis) was more common before the widespread use of steroids. This nonbacterial verrucous endocarditis takes the form of single or multiple 1- to 3-mm warty deposits on any heart valve, distinctively on either surface of the leaflets (Fig. 6.28). By comparison, the vegetations in infective endocarditis are considerably larger, and those in rheumatic heart disease (Chapter 12) are smaller and confined to the lines of closure of the valve leaflets.

An increasing number of SLE patients have evidence of coronary artery disease (angina, myocardial infarction) resulting from coronary atherosclerosis. This complication is particularly notable in young patients with long-standing disease, and especially prevalent in those who have been treated with corticosteroids. The pathogenesis of accelerated coronary atherosclerosis is unclear but is probably multifactorial. Risk factors for atherosclerosis, including hypertension, obesity, and hyperlipidemia, are more commonly present in SLE patients than in the population at large. In addition, immune complexes and antiphospholipid antibodies may cause endothelial damage and promote atherosclerosis.

Spleen. Splenomegaly, capsular thickening, and follicular hyperplasia are common features. Central penicilliary arteries may show concentric intimal and smooth muscle cell hyperplasia, producing so-called onion-skin lesions.

Lungs. In addition to pleuritis and pleural effusions, which are present in almost 50% of patients, in some cases, there is chronic interstitial fibrosis and secondary pulmonary hypertension. None of these changes is specific for SLE.

Other Organs and Tissues. LE, or hematoxylin, bodies in the bone marrow or other organs are strongly indicative of SLE. Lymph nodes may be enlarged due to the presence of hyperplastic germinal centers, or may even demonstrate necrotizing lymphadenitis, usually associated with the presence of activated CTLs and macrophages. The activated T cells may be so prominent in such cases as to mimic certain features of T-cell lymphoma but are polyclonal and reactive in nature.

Clinical Features

SLE is a multisystem disease that is highly variable in its presentation, and its diagnosis relies on a constellation of clinical, serologic, and morphologic changes (see Table 6.9). It may be acute or insidious in its onset. Often, the patient is a young woman with some, but not necessarily all, of the following features: a butterfly rash over the face, fever, pain but no deformity in one or more peripheral joints (feet, ankles, knees, hips, fingers, wrists, elbows, shoulders), pleuritic chest pain, and photosensitivity. In many patients, however, the presentation of SLE is subtle and puzzling, taking forms such as a febrile illness of unknown origin, abnormal urinary findings, or joint disease masquerading

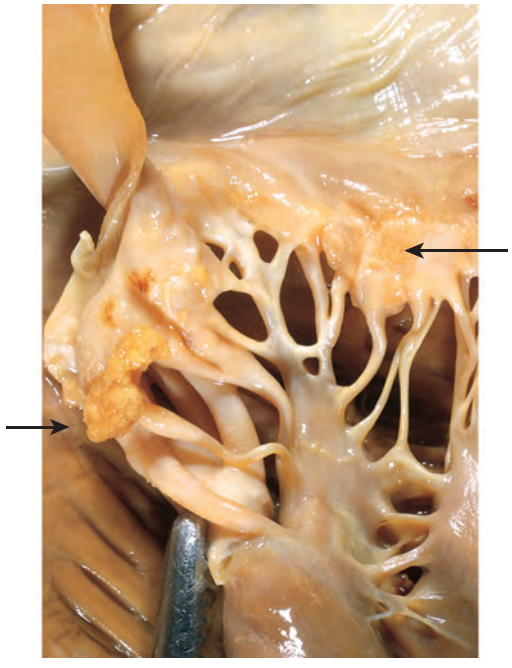


Figure 6.28 Libman-Sacks endocarditis of the mitral valve in lupus erythematosus. The vegetations attached to the margin of the thickened valve leaflet are indicated by *arrows*. (Courtesy Dr. Fred Schoen, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

as rheumatoid arthritis or rheumatic fever. "Generic" ANAs, detected by immunofluorescence assays, are found in virtually 100% of patients, but these are not specific for SLE. A variety of clinical findings may point toward renal involvement, including hematuria, red cell casts, proteinuria, and in some cases the classic nephrotic syndrome (Chapter 20). Laboratory evidence of some hematologic derangement is seen in virtually every case, but in some patients anemia or thrombocytopenia may be the presenting manifestation as well as the dominant clinical problem. In still others, mental aberrations, including psychosis or convulsions, or coronary artery disease may be prominent clinical problems. Patients with SLE are also prone to infections, presumably because of their underlying immune dysfunction and treatment with immunosuppressive drugs.

The course of the disease is variable and unpredictable. Rarely, death ensues within weeks to months. More often, with appropriate therapy, the disease follows a relapsing and remitting course spanning a period of years or even decades. During acute flareups, increased formation of immune complexes results in complement activation, often leading to hypocomplementemia. Disease flares are usually treated with corticosteroids or other immunosuppressive drugs. Even without therapy, in some patients the disease runs an indolent course with relatively mild manifestations, such as skin changes and mild hematuria, for years. The outcome has improved significantly, and an approximately 90% 5-year and 80% 10-year survival can be expected. The most common causes of death are renal failure and intercurrent infections. Coronary artery disease is also becoming an important cause of death. Patients treated with steroids and immunosuppressive drugs incur the usual risks associated with such therapy.

As mentioned earlier, involvement of skin along with multisystem disease is fairly common in SLE. The following sections describe two syndromes in which the cutaneous involvement is the exclusive or most prominent feature.

Chronic Discoid Lupus Erythematosus

Chronic discoid lupus erythematosus is a disease in which the skin manifestations may mimic SLE, but systemic manifestations are rare. It is characterized by the presence of skin plaques showing varying degrees of edema, erythema, scaliness, follicular plugging, and skin atrophy surrounded by an elevated erythematous border. The disease usually remains confined to the skin, but 5% to 10% of patients develop multisystem manifestations after many years. Approximately 35% of patients test positive for generic ANAs, but antibodies to double-stranded DNA are rarely present. Immunofluorescence studies of skin biopsy specimens show deposition of immunoglobulin and C3 at the dermoepidermal junction, similar to what is seen in SLE.

Subacute Cutaneous Lupus Erythematosus

Subacute cutaneous lupus erythematosus appears to define a group of patients with features intermediate between SLE and chronic discoid lupus erythematosus. The condition presents with predominant skin involvement and is distinguished from chronic discoid lupus erythematosus by several criteria. Firstly, the skin rash tends to be widespread, superficial, and nonscarring (though exceptions occur). Secondly, most patients have mild systemic symptoms consistent with SLE. There is a strong association with antibodies to the SS-A antigen and with the *HLA-DR3* genotype.

Drug-Induced Lupus Erythematosus

An SLE-like syndrome may develop in patients receiving a variety of drugs, including hydralazine, procainamide, isoniazid, and D-penicillamine, to name only a few. Somewhat surprisingly, anti-TNF therapy, which is effective in rheumatoid arthritis and other autoimmune diseases, can also cause drug-induced lupus. Many of these drugs are associated with the development of ANAs, but most patients do not have symptoms of SLE. For example, 80% of patients receiving procainamide test positive for ANAs, but only one-third of these manifest clinical symptoms, such as arthralgias, fever, and serositis. Although multiple organs are affected, renal and CNS involvement is distinctly uncommon. There are serologic and genetic differences from classic SLE, as well. Antibodies specific for double-stranded DNA are rare, but there is an extremely high frequency of antibodies specific for histones.

KEY CONCEPTS

SYSTEMIC LUPUS ERYTHEMATOSUS

- SLE is a systemic autoimmune disease caused by autoantibodies produced against numerous self antigens and the formation of immune complexes.
- The major autoantibodies, and the ones responsible for the formation of circulating immune complexes, are directed against nuclear antigens. Other autoantibodies react with red cells, platelets, and various complexes of phospholipids with proteins.
- Disease manifestations include nephritis, skin lesions, arthritis (all caused by the deposition of immune complexes), and hematologic and neurologic abnormalities.

The underlying cause of the breakdown in self-tolerance in SLE is unknown; it may include excess or persistence of nuclear antigens, multiple inherited susceptibility genes, and environmental triggers (e.g., UV irradiation, which results in cellular apoptosis and release of nuclear proteins).

Rheumatoid Arthritis

Rheumatoid arthritis is a chronic inflammatory disease that primarily affects the joints, but may also involve extraarticular tissues such as the skin, blood vessels, lungs, and heart. Abundant evidence supports the autoimmune nature of the disease. Because the principal manifestations of the disease are in the joints, it is discussed in Chapter 26.

Sjögren Syndrome

Sjögren syndrome is a chronic disease characterized by dry eyes (keratoconjunctivitis sicca) and dry mouth (xerostomia) resulting from immunologically mediated destruction of the lacrimal and salivary glands. It occurs as an isolated disorder (primary form), also known as the *sicca syndrome,* or more often in association with another autoimmune disease (secondary form). Among the associated disorders, rheumatoid arthritis is the most common, but some patients have SLE, polymyositis, scleroderma, vasculitis, mixed connective tissue disease, or thyroiditis.

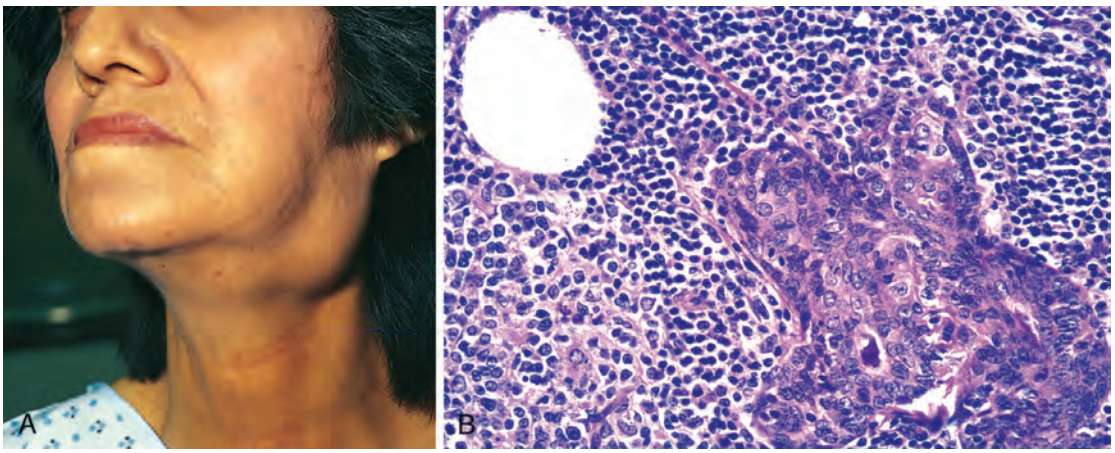


Figure 6.29 Sjögren syndrome. (A) Enlargement of the salivary gland. (B) Intense lymphocytic and plasma cell infiltration with ductal epithelial hyperplasia in a salivary gland. (A) Courtesy Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Tex. (B) Courtesy Dr. Dennis Burns, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Pathogenesis

The characteristic decrease in tears and saliva (sicca syndrome) is the result of lymphocytic infiltration and fibrosis of the lacrimal and salivary glands. The infiltrate contains predominantly activated CD4+ helper T cells and some B cells, including plasma cells. About 75% of patients have rheumatoid factor (an antibody reactive with self IgG) whether or not coexisting rheumatoid arthritis is present. ANAs are detected in 50% to 80% of patients by immunofluorescence assay. A host of other organ-specific and non–organ-specific antibodies have also been identified. Most important, however, are antibodies directed against two ribonucleoprotein antigens, SS-A (Ro) and SS-B (La) (see Table 6.10), which can be detected in as many as 90% of patients by sensitive techniques. These antibodies are thus considered serologic markers of the disease. These autoantibodies are also present in a smaller percentage of patients with SLE and hence are not entirely specific for Sjögren syndrome.

As with other autoimmune diseases, Sjögren syndrome shows some association, albeit weak, with certain HLA alleles. The significance of such associations in disease causation or manifestations is not clear.

Although the pathogenesis of Sjögren syndrome remains obscure, aberrant T-cell and B-cell activation are both implicated. The initiating trigger may be a viral infection of the salivary glands, which causes local cell death and release of tissue self antigens. In genetically susceptible individuals, CD4+ T cells and B cells specific for these self antigens may have escaped tolerance and are able to react. The result is inflammation, tissue damage, and, eventually, fibrosis. However, the role of particular cytokines or T cell subsets in the development of the lesions is not established. The nature of the autoantigens recognized by these lymphocytes is also mysterious. Sjögren syndrome–like disease is seen in some patients with human T-lymphotropic virus (HTLV), human immunodeficiency virus (HIV), and hepatitis C virus infections, but the link between these viruses and the autoimmune disorder is obscure.

MORPHOLOGY

Lacrimal and salivary glands are the major targets of the disease, but other exocrine glands, including those lining the respiratory and gastrointestinal tracts and the vagina, may also be involved. The earliest histologic finding in involved salivary glands is periductal and perivascular lymphocytic infiltration. Eventually the lymphocytic infiltrate becomes extensive (Fig. 6.29B), and lymphoid follicles with germinal centers may appear. The ductal lining epithelial cells may show hyperplasia, thus obstructing the ducts. Later there is atrophy of the acini, fibrosis, and hyalinization; still later in the course, atrophy and replacement of parenchyma with fat are seen. In some cases, the lymphoid infiltrate may be so intense as to give the appearance of a lymphoma. Indeed, these patients are at high risk for development of B-cell lymphomas.

The lack of tears leads to drying of the corneal epithelium, which becomes inflamed, eroded, and ulcerated; the oral mucosa may atrophy, with inflammatory fissuring and ulceration; and dryness and crusting of the nose may lead to ulcerations and even perforation of the nasal septum.

Clinical Features

Sjögren syndrome occurs most commonly in women between 50 and 60 years of age. As might be expected, symptoms result from inflammatory destruction of the exocrine glands. The keratoconjunctivitis produces blurring of vision, burning, and itching, and thick secretions accumulate in the conjunctival sac. The xerostomia results in difficulty in swallowing solid foods, a decrease in the ability to taste, cracks and fissures in the mouth, and dryness of the buccal mucosa. Parotid gland enlargement is present in one-half of patients; dryness of the nasal mucosa, epistaxis, recurrent bronchitis, and pneumonitis are other symptoms. Manifestations of extraglandular disease are seen in one-third of patients and include synovitis, diffuse pulmonary fibrosis, and peripheral neuropathy. These are more common in patients with high titers of antibodies specific for SS-A. In contrast to SLE, glomerular lesions are extremely rare in Sjögren syndrome. Defects of tubular function, however, including renal tubular

acidosis, uricosuria, and phosphaturia, are often seen and are associated with tubulointerstitial nephritis (Chapter 20). About 60% of patients have another accompanying autoimmune disorder, such as rheumatoid arthritis, and these patients also have the symptoms and signs of that disorder.

The combination of lacrimal and salivary gland inflammation was once called Mikulicz disease. The name has now been replaced by *Mikulicz syndrome,* broadened to include lacrimal and salivary gland enlargement from any cause, including sarcoidosis, IgG4-related disease (described later), lymphoma, and other tumors.

As mentioned earlier, the involved glands show intense inflammatory infiltrates. In early stages of the disease, this immune infiltrate consists of a mixture of polyclonal T and B cells. However, if the reaction continues unabated, there is a strong tendency over time for individual clones within the population of B cells to gain a growth advantage, presumably because of the acquisition of somatic mutations. Emergence of a dominant B-cell clone is usually indicative of the development of a marginal zone lymphoma, a specific type of B-cell malignancy that often arises in the setting of chronic lymphocytic inflammation. About 5% of Sjögren patients develop lymphoma, an incidence that is 40-fold greater than normal. Certain other autoimmune disorders (e.g., Hashimoto thyroiditis) are also associated with a high risk of marginal zone lymphoma (Chapter 13), which typically arises within the organ or tissue that is the target of the autoimmune inflammation.

KEY CONCEPTS

SJÖGREN SYNDROME

- Sjögren syndrome is an inflammatory disease that affects primarily the salivary and lacrimal glands, causing dryness of the mouth and eyes.
- The disease is believed to be caused by an autoimmune T-cell reaction against an unknown self antigen expressed in these glands, or immune reactions against the antigens of a virus that infects the tissues.

Systemic Sclerosis (Scleroderma)

Systemic sclerosis is characterized by: (1) chronic inflammation thought to be the result of autoimmunity, (2) widespread damage to small blood vessels, and (3) progressive interstitial and perivascular fibrosis in the skin and multiple organs. Although the term *scleroderma* is ingrained in clinical medicine, this disease is better named *systemic sclerosis* because it is characterized by excessive fibrosis throughout the body. The skin is most commonly affected, but the gastrointestinal tract, kidneys, heart, muscles, and lungs also are frequently involved. In some patients, the disease seems to remain confined to the skin for many years, but in the majority it progresses to visceral involvement with death from renal failure, cardiac failure, pulmonary insufficiency, or intestinal malabsorption. The clinical heterogeneity of systemic sclerosis has been recognized by classifying the disease into two major categories: *diffuse scleroderma,* characterized by widespread skin involvement at onset, with rapid progression and early visceral involvement; and *limited scleroderma,* in which the skin involvement is often confined to fingers, forearms, and face. Visceral involvement occurs late; hence, the clinical course is relatively benign. Some patents with the limited disease develop a combination of calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia, called the *CREST syndrome*. Several other variants and related conditions, such as eosinophilic fasciitis, occur far less frequently and are not described here.

Pathogenesis

The cause of systemic sclerosis is not known, but the **disease likely results from three interrelated processes—autoimmune responses, vascular damage, and collagen deposition** (Fig. 6.30).

- *Autoimmunity.* It is proposed that CD4+ T cells responding to an as yet unidentified antigen accumulate in the skin and release cytokines that activate inflammatory cells and fibroblasts. Although inflammatory infiltrates are typically sparse in the skin of patients with systemic sclerosis, activated CD4+ T cells can be found in many patients,

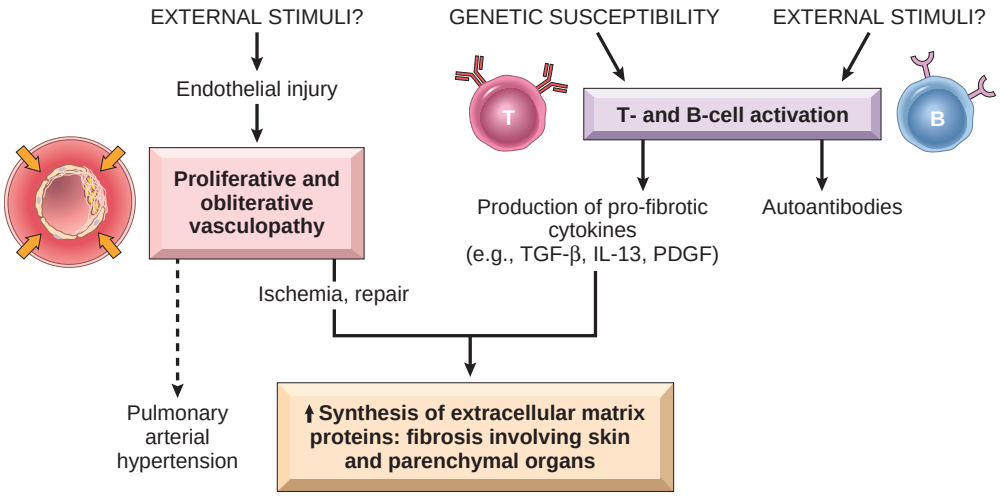


Figure 6.30 A model for the pathogenesis of systemic sclerosis. Unknown external stimuli cause vascular abnormalities and immune activation in genetically susceptible individuals, and both contribute to the excessive fibrosis.

and Th2 cells have been isolated from the skin. Several cytokines produced by these T cells, including TGF-β and IL-13, can stimulate the transcription of genes encoding collagen and other extracellular matrix proteins (e.g., fibronectin) in fibroblasts. Other cytokines recruit leukocytes and propagate the chronic inflammation. There is also evidence for inappropriate activation of humoral immunity, and the presence of various autoantibodies, notably ANAs, provides diagnostic and prognostic information. The role of these ANAs in the pathogenesis of the disease is unclear; it has been postulated that some of these antibodies may stimulate fibrosis, but the evidence in support of this idea is not convincing.

- *Vascular damage.* Microvascular disease is consistently present early in the course of systemic sclerosis and may be the initial lesion. Intimal proliferation is evident in the digital arteries of patients with systemic sclerosis. Capillary dilation with leaking, as well as destruction, is also common. Nailfold capillary loops are distorted early in the course of disease, and later they disappear. Telltale signs of endothelial activation and injury (e.g., increased levels of von Willebrand factor) and increased platelet activation (increased circulating platelet aggregates) have also been noted. However, what causes the vascular injury is not known; it could be the initiating event or the result of chronic inflammation, with mediators released by inflammatory cells inflicting damage on microvascular endothelium. Repeated cycles of endothelial injury followed by platelet aggregation lead to release of platelet and endothelial factors (e.g., PDGF, TGF-β) that trigger perivascular fibrosis. Vascular smooth muscle cells also show abnormalities, such as increased expression of adrenergic receptors. Eventually, widespread narrowing of the microvasculature leads to ischemic injury and scarring.
- *Fibrosis.* The progressive fibrosis characteristic of the disease may be the culmination of multiple abnormalities, including the accumulation of alternatively activated macrophages, actions of fibrogenic cytokines produced by infiltrating leukocytes, hyperresponsiveness of fibroblasts to these cytokines, and scarring following ischemic damage caused by the vascular lesions. There is some evidence that fibroblasts from patients with systemic sclerosis have an intrinsic abnormality that causes them to produce excessive amounts of collagen. This idea is based on studies with cultured fibroblasts, and whether or how this abnormality relates to pathogenesis in vivo is unknown.

MORPHOLOGY

Virtually all organs may be involved in systemic sclerosis. Prominent changes occur in the skin, alimentary tract, musculoskeletal system, and kidney, and lesions also are often present in the blood vessels, heart, lungs, and peripheral nerves.

Skin. The majority of patients have diffuse, sclerotic atrophy of the skin, which usually begins in the fingers and distal regions of the upper extremities and extends proximally to involve the upper arms, shoulders, neck, and face. Histologically, there are perivascular infiltrates containing CD4+ T cells and edema, together with degeneration of collagen fibers, which become eosinophilic. Capillaries and small arteries (150 to 500 μm in diameter) may show thickening of the basal lamina, endothelial cell damage, and partial occlusion. With progression of the disease, there is increasing fibrosis of the dermis, which becomes tightly bound to the subcutaneous structures. There is marked increase of compact collagen in the dermis, usually with thinning of the epidermis, loss of rete pegs, atrophy of the dermal appendages, and hyaline thickening of the walls of dermal arterioles and capillaries (Fig. 6.31B). Focal and sometimes diffuse subcutaneous calcifications may develop, especially in patients with the CREST syndrome. In advanced stages, the fingers take on a tapered clawlike appearance, joint motion is limited, and the face becomes a drawn mask. Loss of blood supply may lead to cutaneous ulcerations and to atrophic changes in the terminal phalanges (Fig. 6.31C). Sometimes the tips of the fingers undergo autoamputation.

Alimentary Tract. The alimentary tract is affected in approximately 90% of patients. Progressive atrophy and collagenous fibrous replacement of the muscularis may develop at any level of the gut but are most severe in the esophagus. The lower two-thirds of the esophagus often develops a rubber-hose–like inflexibility. The associated dysfunction of the lower esophageal sphincter gives rise to gastroesophageal reflux and its complications, including Barrett esophagus (Chapter 17) and strictures. The mucosa is thinned and may be ulcerated, and there is excessive collagenization of the lamina propria and submucosa. Loss of villi and microvilli in the small bowel sometimes produces a malabsorption syndrome.

Musculoskeletal System. Inflammation of the synovium, associated with synovial hypertrophy and hyperplasia, is common in the early stages; fibrosis later ensues. These changes are reminiscent of rheumatoid arthritis, but joint destruction is not common in systemic sclerosis. In a small subset of patients (approximately 10%), inflammatory myositis may develop.

Kidneys. Renal abnormalities occur in two-thirds of patients with systemic sclerosis, most prominently vascular lesions. Interlobular arteries show intimal thickening as a result of deposition of mucinous or finely collagenous material, which stains histochemically for glycoprotein and acid mucopolysaccharides. There is also concentric proliferation of intimal cells. These changes may resemble those seen in malignant hypertension, but in scleroderma the alterations are restricted to vessels 150 to 500 μm in diameter and are not always associated with hypertension. Hypertension, however, does occur in 30% of patients with scleroderma, and in 20% it takes an ominously rapid, downhill course (malignant hypertension). In hypertensive patients, vascular alterations are more pronounced and are often associated with fibrinoid necrosis of arterioles, thrombosis, and infarction. Such patients often die of renal failure, which overall accounts for about 50% of deaths from systemic sclerosis. There are no specific glomerular changes.

Lungs. Pulmonary involvement is seen in more than 50% of individuals with systemic sclerosis and may manifest as interstitial fibrosis and pulmonary hypertension. Pulmonary vasospasm, secondary to pulmonary vascular endothelial dysfunction, is considered important in the pathogenesis of pulmonary hypertension. Pulmonary fibrosis, when present, is indistinguishable from that seen in idiopathic pulmonary fibrosis (Chapter 15).

Heart. Pericarditis with effusion, myocardial fibrosis, and thickening of intramyocardial arterioles occur in one-third of patients. Clinical impairment by myocardial involvement, however, is less common.

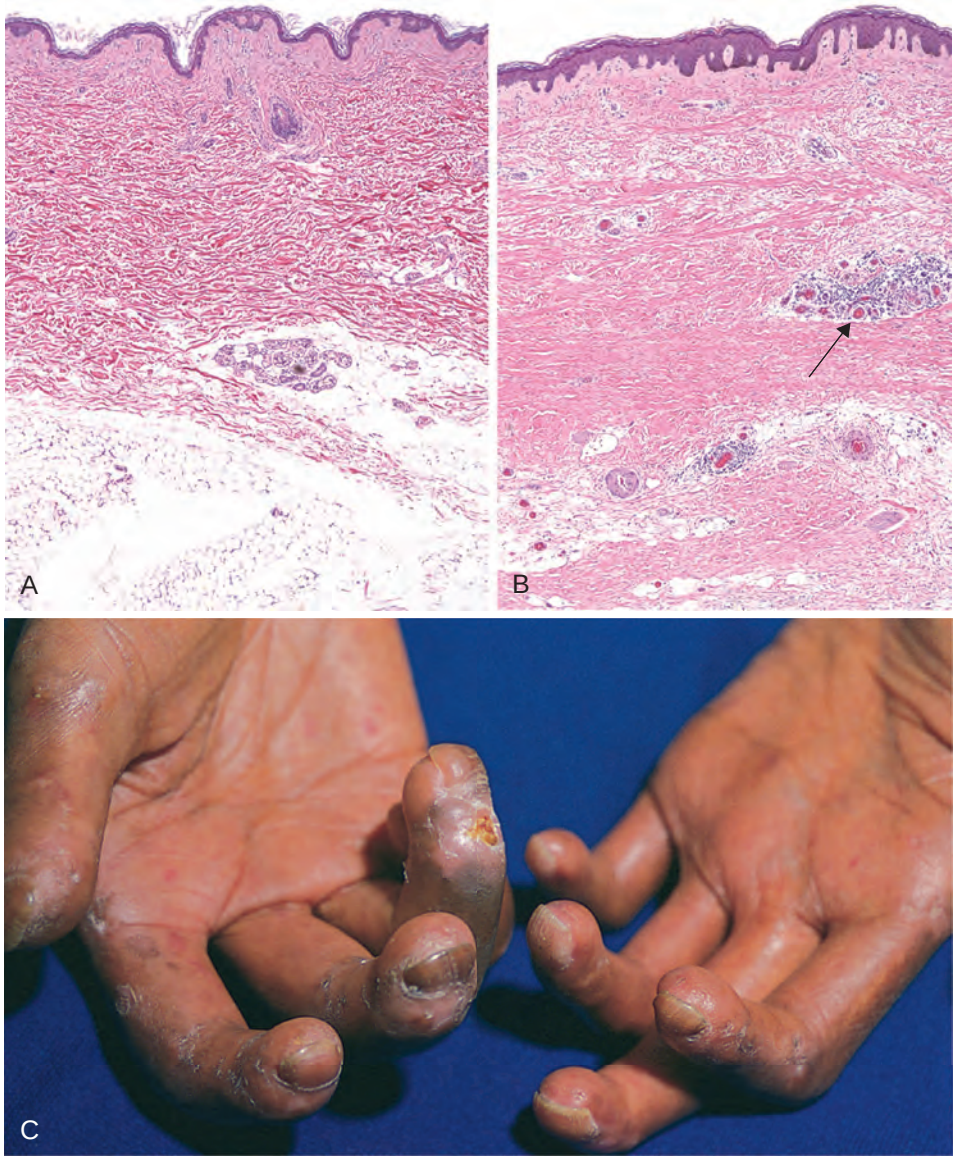


Figure 6.31 Systemic sclerosis. (A) Normal skin. (B) Skin biopsy from a patient with systemic sclerosis. Note the extensive deposition of dense collagen in the dermis with virtual absence of appendages (e.g., hair follicles) and foci of inflammation *(arrow)*. (C) The extensive subcutaneous fibrosis has virtually immobilized the fingers, creating a clawlike flexion deformity. Loss of blood supply has led to cutaneous ulcerations. (C, Courtesy Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Tex.)

Clinical Features

Systemic sclerosis has a female-to-male ratio of 3:1, with a peak incidence in the 50- to 60-year age group. Although systemic sclerosis shares many features with SLE, rheumatoid arthritis (Chapter 26), and polymyositis (Chapter 27), its distinctive features are the striking cutaneous changes, notably skin fibrosis (Fig. 6.31C). *Raynaud phenomenon,* manifested as numbness and tingling of the fingers and toes caused by episodic vasoconstriction of arteries and arterioles, is seen in virtually all patients and precedes other symptoms in 70% of cases. Dysphagia attributable to esophageal fibrosis and its resultant hypomotility are present in more than 50% of patients. Eventually, destruction of the esophageal wall leads to atony and dilation, especially at its lower end. Abdominal pain, intestinal obstruction, or malabsorption syndrome with weight loss and anemia due to nutritional deficiencies reflect involvement of the small intestine. Respiratory difficulties caused by the pulmonary fibrosis may result in right-sided cardiac dysfunction, and myocardial fibrosis may cause either arrhythmias or cardiac failure. Proteinuria occurs in as many as 30% of patients, but rarely is severe enough to cause nephrotic syndrome. The most ominous manifestation is malignant hypertension (Chapter 11), with the subsequent development of renal failure, but in its absence progression of the disease may be slow. The disease tends to be more severe in people of African descent, especially women. As treatment of the renal crises has improved, pulmonary disease has become the major cause of death in systemic sclerosis.

Virtually all patients have ANAs that react with a variety of nuclear antigens. Two ANAs strongly associated with systemic sclerosis have been described. One of these, directed

against DNA topoisomerase I (anti-Scl 70), is highly specific. Depending on the ethnic group and the assay, it is present in 10% to 20% of patients with diffuse systemic sclerosis. Patients who have this antibody are more likely to have pulmonary fibrosis and peripheral vascular disease. The other, an anticentromere antibody, is found in 20% to 30% of patients, who tend to have the CREST syndrome. Patients with this syndrome have relatively limited involvement of skin, often confined to fingers, forearms, and face, and calcification of the subcutaneous tissues. Involvement of the viscera, including esophageal lesions, pulmonary hypertension, and biliary cirrhosis, may not occur at all or occur late. In general, these patients live longer than those with diffuse visceral involvement at the outset.

KEY CONCEPTS

SYSTEMIC SCLEROSIS

- Systemic sclerosis (commonly called scleroderma) is characterized by progressive fibrosis involving the skin, gastrointestinal tract, and other tissues.
- Fibrosis may be the result of activation of fibroblasts by cytokines produced by T cells, but what triggers T-cell responses is unknown.
- Endothelial injury and microvascular disease are commonly present in the lesions of systemic sclerosis, perhaps causing chronic ischemia, but the pathogenesis of vascular injury is not known.

Inflammatory Myopathies

Inflammatory myopathies comprise an uncommon, heterogeneous group of disorders characterized by injury and inflammation of mainly the skeletal muscles that are probably immunologically mediated. These diseases are described in Chapter 27.

Mixed Connective Tissue Disease

The term mixed connective tissue disease is used to describe a disease with clinical features that overlap with those of SLE, systemic sclerosis, and polymyositis. The disease is characterized serologically by high titers of antibodies to U1 ribonucleoprotein. Typically, mixed connective tissue disease presents with synovitis of the fingers, Raynaud phenomenon, myositis, and renal involvement, which is modest and responds well to corticosteroids, at least in the short term. It is debated whether mixed connective tissue disease is a truly distinct entity, or if different patients represent subsets of SLE, systemic sclerosis, and myositis. The disease may, over time, evolve into classic SLE or systemic sclerosis. However, in other patients the mixture of features are maintained over time, and the salutary response to steroids is not universal, suggesting that there is a form of mixed connective tissue disease that is distinct from other autoimmune diseases.

Polyarteritis Nodosa and Other Vasculitides

Polyarteritis nodosa belongs to a group of diseases characterized by necrotizing inflammation of the walls of blood vessels and showing strong evidence of an immunologic pathogenic mechanism. The general term *noninfectious vasculitis* differentiates these conditions from those due to direct infection of the blood vessel wall (such as occurs in the wall of an abscess). Noninfectious vasculitis is encountered in many clinical settings. A detailed classification and description of vasculitides is presented in Chapter 11, where the immunologic mechanisms are also discussed.

IgG4-Related Disease

IgG4-related disease (IgG4-RD) is a constellation of disorders characterized by tissue infiltrates dominated by IgG4 antibody–producing plasma cells and lymphocytes (particularly T cells), fibrosis, obliterative phlebitis, and usually increased serum IgG4. Although recognized only recently when extrapancreatic manifestations were identified in patients with autoimmune pancreatitis, IgG4-related disease has now been described in virtually every organ system: the biliary tree, salivary glands, periorbital tissues, kidneys, lungs, lymph nodes, meninges, aorta, breast, prostate, thyroid, pericardium, and skin (Fig. 6.32). Many medical conditions long viewed as confined to single organs are part of the IgG4-RD spectrum. These include some forms of Mikulicz syndrome, Riedel thyroiditis, idiopathic retroperitoneal fibrosis, autoimmune pancreatitis, and inflammatory pseudotumors of the orbit, lungs, and kidneys, to name a few. The disease most often affects middle-age and older men.

The pathogenesis of this condition is not understood, and although IgG4 production in lesions is a hallmark of the disease it is not known if this antibody type contributes to the pathology. The key role of B cells is supported by initial clinical trials in which depletion of B cells with anti–B-cell reagents such as rituximab provided clinical benefit. It is unclear if the disease is truly autoimmune in nature, and no target autoantigens have been identified.

REJECTION OF TISSUE TRANSPLANTS

Transplant rejection is discussed here because it involves several of the immunologic reactions that underlie immune-mediated inflammatory diseases. A major barrier to transplantation is the process of rejection, in which the recipient's immune system recognizes the graft as being foreign and attacks it.

Mechanisms of Recognition and Rejection of Allografts

Rejection is a process in which T lymphocytes and antibodies produced against graft antigens react against and destroy tissue grafts. Grafts exchanged between individuals of the same species (the usual clinical situation) are called *allografts*, and grafts from one species to another (still an experimental procedure) are called *xenografts*. Since most clinical transplants are allografts, the discussion is focused on these.

Recognition of Graft Alloantigens by T and B Lymphocytes

The major antigenic differences between a donor and recipient that result in rejection of transplants are differences in HLA alleles. Because HLA genes are highly

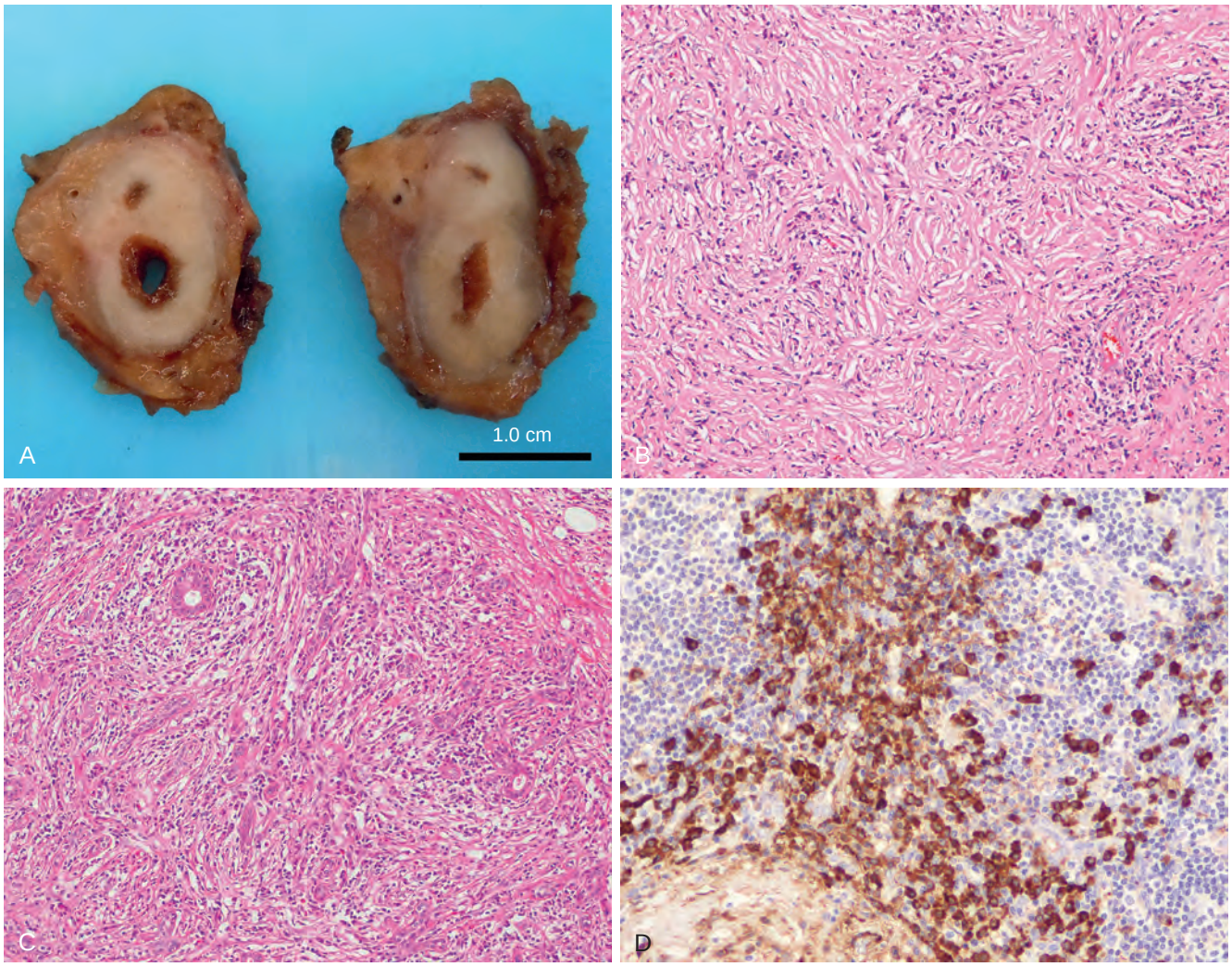


Figure 6.32 IgG4-related disease: representative lesions. (A) Bile duct showing sclerosing cholangitis. (B) Sclerotic area of the bile duct with storiform fibrosis. (C) Submandibular gland with infiltrates of lymphocytes and plasma cells and whorls of fibrosis. (D) Section of an involved lacrimal gland stained with an antibody against IgG_4, showing large numbers of IgG4-producing plasma cells. (From Kamisawa T, Zen Y, Pillai S, et al: IgG4-related disease, *Lancet* 385:1460, 2015.)

polymorphic, there are always some differences between individuals (except, of course, identical twins). Following transplantation, the recipient's T cells recognize donor HLA antigens from the graft (the allogeneic antigens, or alloantigens) by two pathways. The graft antigens are either presented directly to recipient T cells by graft APCs, or the graft antigens are picked up by host APCs, processed (like any other foreign antigen), and presented to host T cells. These are called the direct and indirect pathways of recognition of alloantigens. Both lead to the activation of CD8+ T cells, which develop into CTLs, and CD4+ T cells, which become cytokine-producing effector cells, mainly Th1 cells. The direct pathway may be most important for CTL-mediated acute rejection, and the indirect pathway may play a greater role in chronic rejection.

The frequency of T cells that can recognize the foreign antigens in a graft is much higher than the frequency of T cells specific for any microbe. For this reason, immune responses to allografts are stronger than responses to pathogens. Predictably, these strong reactions can destroy grafts rapidly, and their control requires powerful immunosuppressive agents.

B lymphocytes also recognize antigens in the graft, including HLA and other antigens that differ between donor and recipient. The activation of these B cells typically requires T cell help.

Patterns and Mechanisms of Graft Rejection

Graft rejection is classified into hyperacute, acute, and chronic, on the basis of clinical and pathologic features. This historical classification was devised by clinicians based on rejection of kidney allografts, and has stood the test of time remarkably well. Each type of rejection is mediated by a particular kind of immune response. In the following discussion, the description of the morphology of rejection is limited to kidney allografts, but similar changes are seen in other organ transplants.

- **Hyperacute rejection is mediated by preformed antibodies specific for antigens on graft endothelial cells.** The preformed antibodies may be natural IgM antibodies

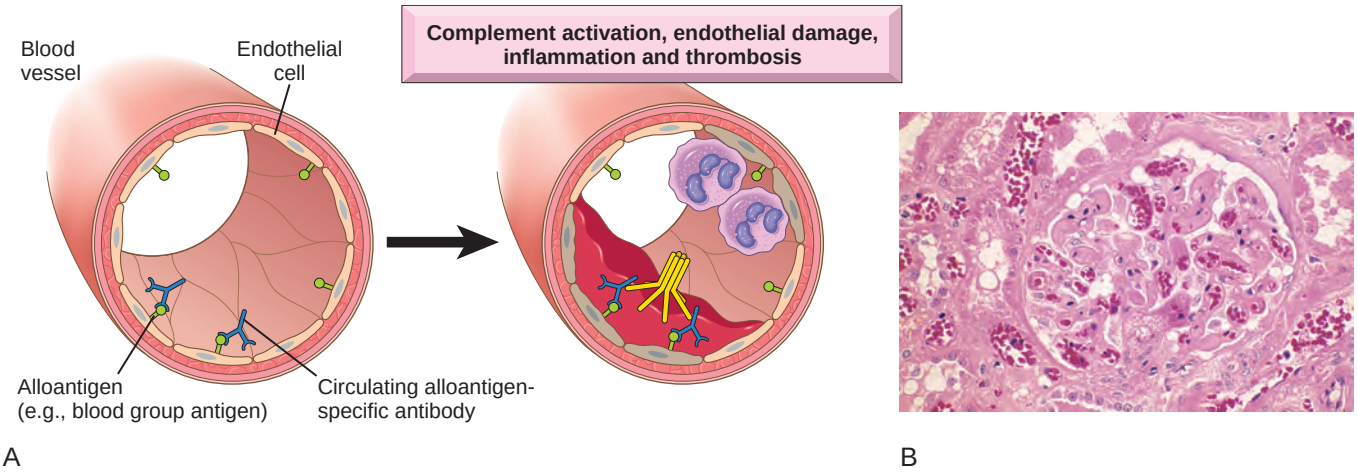


Figure 6.33 Hyperacute rejection. (A) Deposition of antibody on endothelium and activation of complement causes thrombosis. (B) Hyperacute rejection of a kidney allograft showing platelet fibrin thrombi and severe ischemic injury in a glomerulus.

specific for blood group antigens, or may be antibodies specific for allogeneic MHC molecules that were induced by prior exposure of the organ recipient to allogeneic cells through blood transfusions, pregnancy, or transplantation of another organ. Immediately after the graft is implanted and blood flow is restored, the antibodies bind to antigens on the graft vascular endothelium and activate the complement system, leading to endothelial injury, thrombosis, and ischemic necrosis of the graft (Fig. 6.33A). Hyperacute rejection is rare because every donor and recipient are matched for blood type, and potential recipients are tested for antibodies against the cells of the prospective donor, a test called a cross-match.

MORPHOLOGY

In hyperacute rejection, the affected kidney rapidly becomes cyanotic, mottled, and anuric. Virtually all arterioles and arteries exhibit acute fibrinoid necrosis of their walls and narrowing or complete occlusion of their lumens by thrombi (Fig. 6.33B). Neutrophils rapidly accumulate within arterioles, glomeruli, and peritubular capillaries. As these changes intensify and become diffuse, the glomerular capillaries also undergo thrombotic occlusion, and eventually the kidney cortex undergoes outright necrosis (infarction). Affected kidneys are nonfunctional and have to be removed.

- **Acute rejection is mediated by T cells and antibodies that are activated by alloantigens in the graft**. It occurs within days or weeks after transplantation and is the principal cause of early graft failure. It also may appear suddenly much later after transplantation if immunosuppression is tapered or terminated. Based on the role of T cells or antibodies, acute rejection is divided into two types, although in most rejecting grafts, both patterns are present.

  In *acute cellular rejection,* CD8+ CTLs may directly destroy graft cells, or CD4+ cells secrete cytokines and induce inflammation, which damages the graft (Fig. 6.34A). T cells may also react against graft vessels, leading to vascular damage. Current immunosuppressive therapy is designed mainly to prevent and reduce acute rejection by blocking the activation of alloreactive T cells.

MORPHOLOGY

Acute cellular (T cell–mediated) rejection may produce two different patterns of injury.

- In the tubulointerstitial pattern (sometimes called type I), there is extensive interstitial inflammation and tubular inflammation (tubulitis) associated with focal tubular injury (Fig. 6.34B). As might be expected, the inflammatory infiltrates contain activated CD4+ and CD8+ T lymphocytes.
- The vascular pattern shows inflammation of vessels (type II) (Fig. 6.34C) and sometimes necrosis of vessel walls (type III). The affected vessels have swollen endothelial cells, and lymphocytes are seen between the endothelium and the vessel wall, a finding termed endotheliitis or intimal arteritis. The recognition of cellular rejection is important because, in the absence of accompanying humoral rejection, most patients respond well to immunosuppressive therapy.

In acute *antibody-mediated (vascular, or humoral) rejection,* antibodies bind to vascular endothelium and activate complement via the classical pathway (Fig. 6.35A). The resultant inflammation and endothelial damage cause graft failure.

MORPHOLOGY

Acute antibody-mediated rejection is manifested mainly by damage to glomeruli and small blood vessels. Typically, there is inflammation of glomeruli and peritubular capillaries (Fig. 6.35B), associated with deposition of complement products due to activation of the complement system by the antibody-dependent classical pathway (Fig. 6.35C). Small vessels also may show focal thrombosis.

- **Chronic rejection is an indolent form of graft damage that occurs over months or years, leading to progressive**

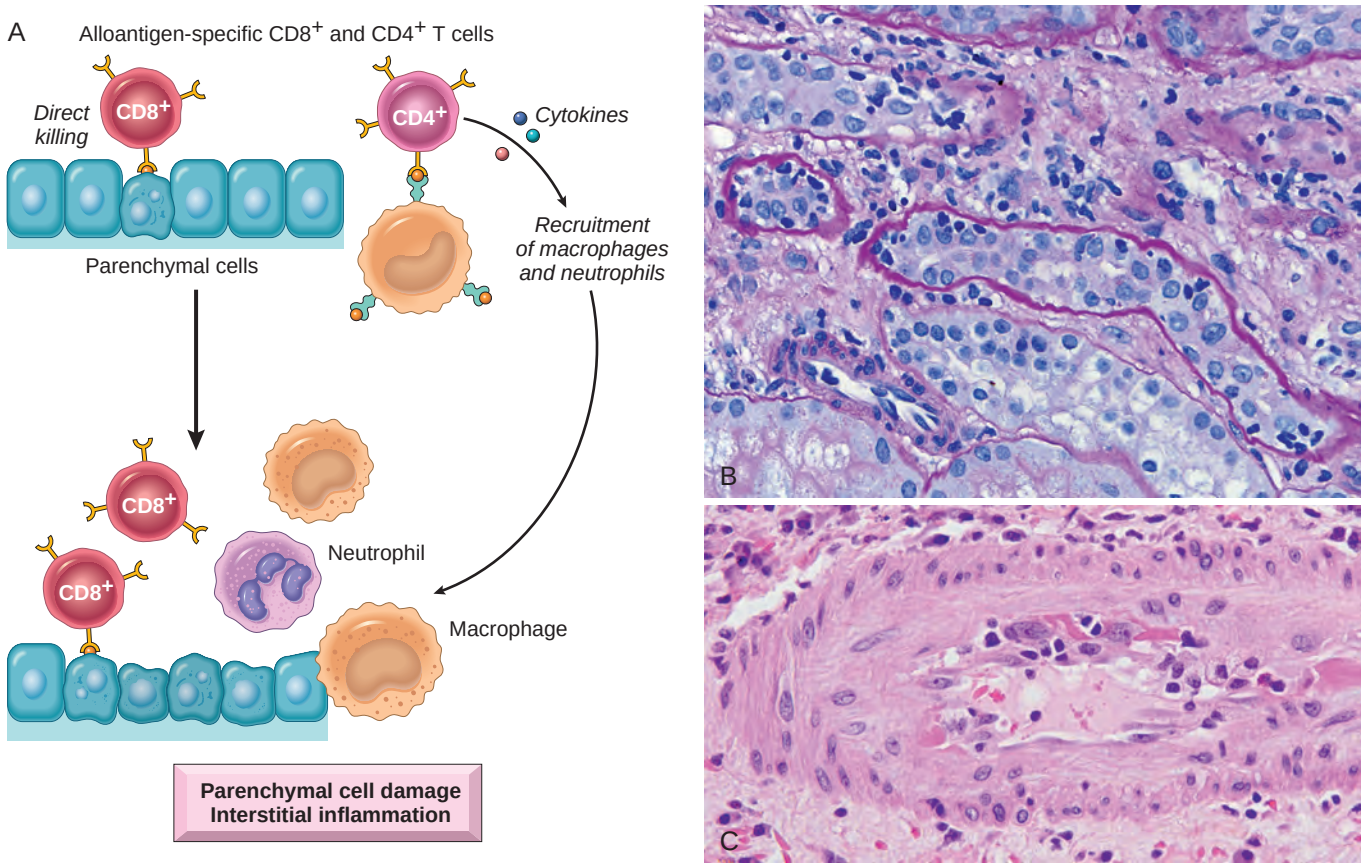


Figure 6.34 Acute cellular rejection. (A) Destruction of graft cells by T cells. Acute T-cell–mediated rejection involves direct killing of graft cells by CD8+ CTLs and inflammation caused by cytokines produced by CD4 T cells. (B) Acute cellular rejection of a kidney graft, manifested by inflammatory cells in the interstitium and between epithelial cells of the tubules (tubulitis). Collapsed tubules are outlined by wavy basement membranes. (C) Rejection vasculitis in a kidney graft. An arteriole is shown with inflammatory cells attacking and undermining the endothelium (endotheliitis) *(arrow)*. (Courtesy Drs. Zoltan Laszik and Kuang-Yu Jen, Department of Pathology, University of California, San Francisco, Calif.)

loss of graft function. Chronic rejection manifests as interstitial fibrosis and gradual narrowing of graft blood vessels *(graft arteriosclerosis)*. In both lesions, the culprits are believed to be T cells that react against graft alloantigens and secrete cytokines, which stimulate the proliferation and activities of fibroblasts and vascular smooth muscle cells in the graft (Fig. 6.36A). Alloantibodies also contribute to chronic rejection. Although treatments to prevent or curtail acute rejection have steadily improved, leading to better 1-year survival of transplants, chronic rejection is refractory to most therapies and is becoming the principal cause of graft failure.

MORPHOLOGY

Chronic rejection is dominated by vascular changes, often with intimal thickening and vascular occlusion (Fig. 6.36B). Chronically rejecting kidney grafts show glomerulopathy, with duplication of the basement membrane, likely secondary to chronic endothelial injury (Fig. 6.36C), and peritubular capillaritis with multilayering of peritubular capillary basement membranes. Interstitial fibrosis and tubular atrophy with loss of renal parenchyma may occur secondary to the vascular lesions (Fig. 6.36D). Interstitial mononuclear cell infiltrates are typically sparse.

In addition to the kidney, a variety of organs, such as the liver (Chapter 18), heart (Chapter 12), lungs, and pancreas, are also transplanted. The rejection of these grafts is described in the relevant chapters.

Methods of Increasing Graft Survival

The value of HLA matching between donor and recipient varies in different solid-organ transplants. In kidney transplants, there is substantial benefit if all the polymorphic HLA alleles are matched (both inherited alleles of *HLA-A*, *-B*, and *DR*). However, HLA matching is usually not done for transplants of liver, heart, and lungs, because other considerations, such as anatomic compatibility, severity of the underlying illness, and the need to minimize the time of organ storage, override the potential benefits of HLA matching.

Except for identical twins, immunosuppressive therapy is essential in all donor-recipient combinations. Immunosuppressive drugs in current use include steroids (which reduce inflammation), mycophenolate mofetil (which inhibits lymphocyte proliferation), and tacrolimus (FK506). Tacrolimus, like its predecessor cyclosporine, is an inhibitor of the phosphatase calcineurin, which is required for activation of a transcription factor called nuclear factor of activated T cells (NFAT). NFAT stimulates transcription of cytokine

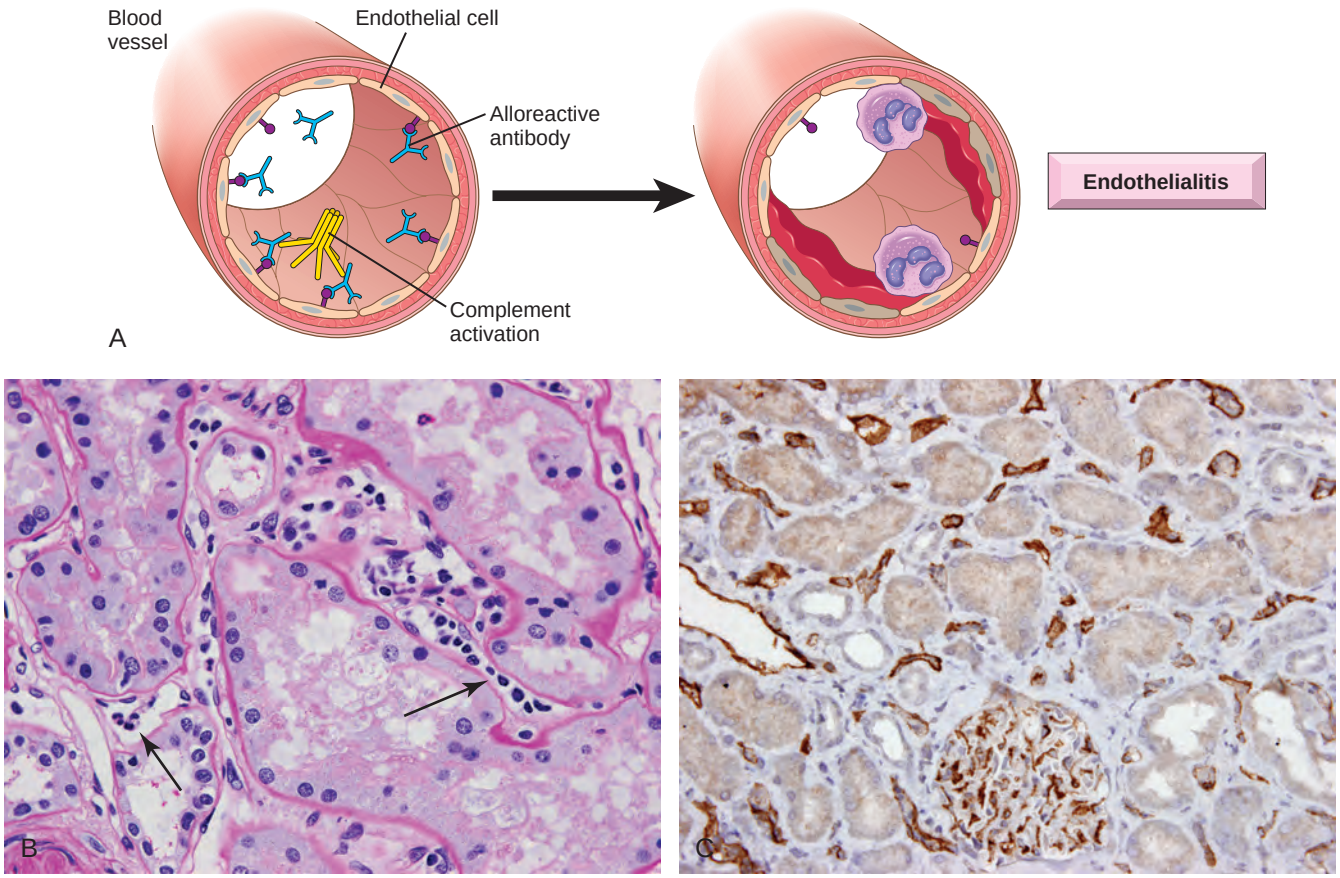


Figure 6.35 Acute antibody-mediated (humoral) rejection. (A) Graft damage caused by antibody deposition in vessels. (B) Light micrograph showing inflammation (capillaritis) in peritubular capillaries *(arrows)* in a kidney graft. (C) Immunoperoxidase stain shows C4d deposition in peritubular capillaries and a glomerulus. (Courtesy Dr. Zoltan Laszik, Department of Pathology, University of California, San Francisco, Calif.)

genes, in particular the gene that encodes the growth factor IL-2. Thus, tacrolimus inhibits T cell responses. Additional drugs that are used to treat rejection include T cell– and B cell–depleting antibodies, and pooled intravenous IgG (IVIG), which suppresses inflammation by unknown mechanisms. Plasmapheresis is used in cases of severe antibody-mediated rejection. Another more recent strategy for reducing antigraft immune responses is to prevent host T cells from receiving costimulatory signals from DCs during the initial phase of sensitization. This can be accomplished by interrupting the interaction between the B7 molecules on the DCs of the graft donor with the CD28 receptors on host T cells, for example, by administration of proteins that bind to B7 costimulators.

Although immunosuppression prolongs graft survival, it carries its own risks. The price paid in the form of increased susceptibility to opportunistic infections is not small. One of the most frequent infectious complications is reactivation of polyoma virus. The virus establishes latent infection of epithelial cells in the lower genitourinary tract of healthy individuals, and on immunosuppression, it is reactivated, infects renal tubules, and may even cause graft failure. These patients are also at increased risk for developing EBV-induced lymphomas, human papillomavirus–induced squamous cell carcinomas, and Kaposi sarcoma (KS; Chapter 11), all probably the result of reactivation of latent viral infections because of diminished host defenses. To circumvent the untoward effects of immunosuppression, much effort is being devoted to induce donor-specific tolerance in graft recipients. Strategies being tested include injecting regulatory T cells and blocking the costimulatory signals that are required for lymphocyte activation, as mentioned earlier.

Transplantation of Hematopoietic Stem Cells (HSCs)

Use of HSC transplants to treat hematologic malignancies, bone marrow failure syndromes (such as aplastic anemia), and inherited bone marrow disorders (such as sickle cell anemia, thalassemia, and immunodeficiency states) is increasing each year. Transplantation of genetically "reengineered" HSCs obtained from affected patients may also be useful for somatic cell gene therapy, and is being evaluated in some immunodeficiencies and hemoglobinopathies. Historically, HSCs were obtained from the bone marrow, but now they usually are harvested from peripheral blood after they are mobilized from the bone marrow by administration of hematopoietic growth factors, or from the umbilical cord blood of newborn infants, a rich source of HSCs. In most of the conditions in which HSC transplantation is indicated, the recipient is irradiated or treated with high doses of chemotherapy to destroy the immune system (and sometimes, cancer cells) and to "open up" niches in the microenvironment of the marrow that nurture HSCs, thus

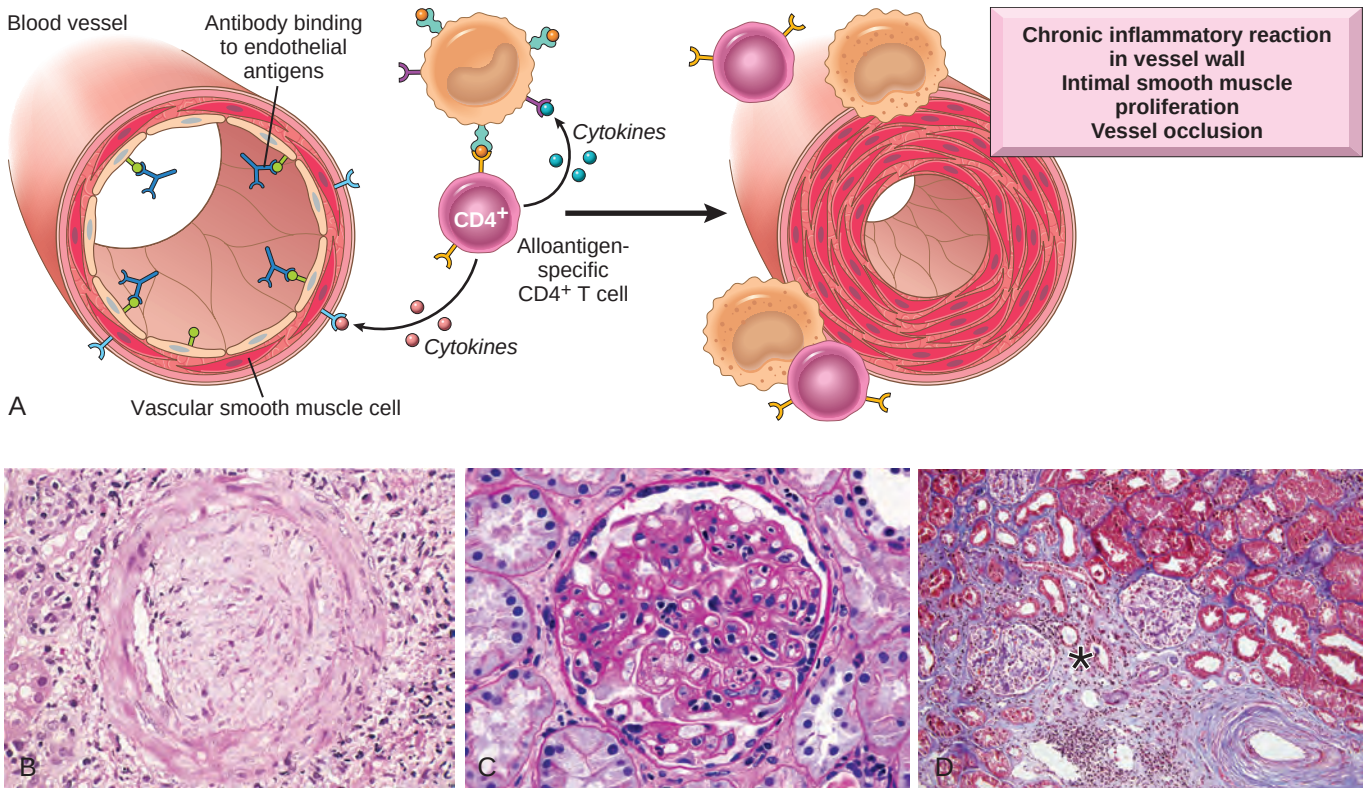


Figure 6.36 Chronic rejection. (A) Graft arteriosclerosis caused by T-cell cytokines and antibody deposition. (B) Graft arteriosclerosis in a cardiac transplant. (C) Transplant glomerulopathy, the characteristic manifestation of chronic antibody-mediated rejection in the kidney. The glomerulus shows inflammatory cells within the capillary loops (glomerulitis), accumulation of mesangial matrix, and duplication of the capillary basement membrane. (D) Interstitial fibrosis and tubular atrophy, resulting from arteriosclerosis of arteries and arterioles in a chronically rejecting kidney allograft. In this trichrome stain, the blue area *(asterisk)* shows fibrosis, contrasted with the normal kidney *(top right)*. An artery showing prominent arteriosclerosis is shown *(bottom right)*. (B, Courtesy of Dr. Richard Mitchell, Department of Pathology, Brigham and Women's Hospital, Boston, Mass. C and D, Courtesy of Dr. Zoltan Laszik, Department of Pathology, University of California, San Francisco, Calif.)

allowing the transplanted HSCs to engraft. Several features distinguish HSC transplants from solid-organ transplants. Two problems that are unique to HSC transplantation are graft-versus-host disease (GVHD) and immunodeficiency.

GVHD occurs when immunologically competent cells or their precursors are transplanted into immunologically compromised recipients, and the transferred cells recognize alloantigens in the host and attack host tissues. It is seen most commonly in the setting of HSC transplantation but, rarely, may occur following transplantation of solid organs rich in lymphoid cells (e.g., the liver) or transfusion of unirradiated blood. When immune-compromised recipients receive HSC preparations from allogeneic donors, the immunocompetent T cells present in the donor inoculum recognize the recipient's HLA antigens as foreign and react against them. To try to minimize GVHD, HSC transplants are done between donor and recipient who are HLA-matched using precise DNA sequencing–based methods for molecular typing of HLA alleles.

- *Acute GVHD* occurs within days to weeks after allogeneic HSC transplantation. Although any organ may be affected, the major clinical manifestations result from involvement of the immune system and epithelia of the skin, liver, and intestines. Involvement of skin in GVHD is manifested by a generalized rash that may lead to desquamation in severe cases (Fig. 6.37A). Destruction of small bile ducts gives rise to jaundice, and mucosal ulceration of the gut results in bloody diarrhea. Although tissue injury may be severe, the affected tissues are usually not heavily infiltrated by lymphocytes. It is believed that in addition to direct cytotoxicity by CD8+ T cells, considerable damage is inflicted by cytokines released by the sensitized donor T cells.
- *Chronic GVHD* may follow the acute syndrome or may occur insidiously. These patients have extensive cutaneous injury, with destruction of skin appendages and fibrosis of the dermis (Fig. 6.37B). The changes may resemble systemic sclerosis (discussed earlier). Chronic liver disease manifested by cholestatic jaundice is also frequent. Damage to the gastrointestinal tract may cause esophageal strictures. The immune system is devastated, with involution of the thymus and depletion of lymphocytes in the lymph nodes. Not surprisingly, patients experience recurrent and life-threatening infections. Other patients develop manifestations of autoimmunity, postulated to result from the grafted CD4+ helper T cells reacting with host B cells and stimulating these cells, some of which may be capable of producing autoantibodies.

Because GVHD is mediated by T lymphocytes contained in the transplanted donor cells, depletion of donor T cells before transfusion virtually eliminates the disease. This protocol, however, is a decidedly mixed blessing: GVHD is ameliorated, but the recurrence of tumor in leukemic

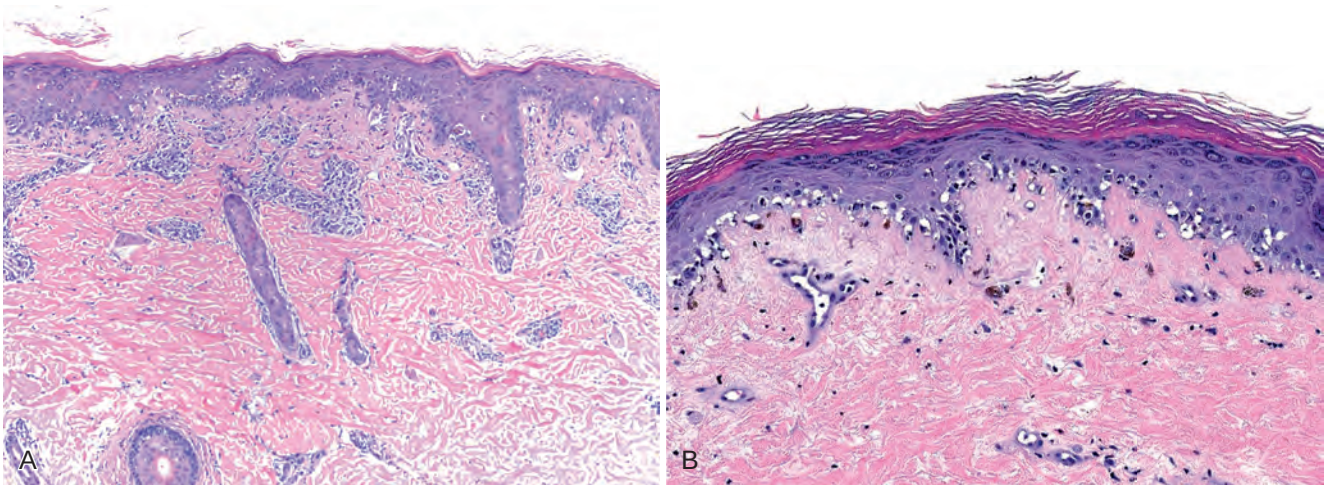


Figure 6.37 Graft-versus-host disease (GVHD) involving the skin. (A) Acute GVHD showing patchy epithelial cell death and dermal infiltrates of mononuclear cells (lymphocytes and macrophages). (B) Chronic GVHD showing a sparse lymphocytic infiltrate at the dermal-epidermal junction, which has resulted in a vacuolar interface reaction and occasional damaged keratinocytes. The epidermis is thinned, signifying atrophy. The underlying dermis shows thickening of collagen bundles, indicative of sclerosis. (B, Courtesy Dr. Jarish Cohen, Department of Pathology, University of California, San Francisco, Calif.)

patients as well as the incidence of graft failures and EBV-related B-cell lymphoma increase. It seems that the multifaceted T cells not only mediate GVHD but also are required for engraftment of the transplanted HSCs, suppression of EBV-infected B-cell clones, and control of leukemia cells. The latter graft-versus-leukemia effect can be quite dramatic. In fact, deliberate induction of graft-versus-leukemia effect by infusion of allogeneic T cells is used to treat chronic myeloid leukemia that has relapsed after HSC transplantation.

Immunodeficiency is a frequent complication of HSC transplantation. The immunodeficiency may be a result of prior treatment (e.g., for leukemia), myeloablation prior to HSC transplantation, a delay in repopulation of the recipient's immune system, and attack on the host's immune cells by grafted lymphocytes. Affected individuals are profoundly immunosuppressed and are easy prey to infections. Although many different types of organisms may infect patients, infection with cytomegalovirus is particularly important. This usually results from activation of latent infection. Cytomegalovirus-induced pneumonitis can be a fatal complication.

KEY CONCEPTS

RECOGNITION AND REJECTION OF TRANSPLANTS (ALLOGRAFTS)

- The rejection response against solid organ transplants is initiated mainly by host T cells that recognize the foreign HLA antigens of the graft, either directly (on APCs in the graft) or indirectly (after uptake and presentation by host APCs).
- Types and mechanisms of rejection of solid organ grafts:
  - Hyperacute rejection. Preformed antidonor antibodies bind to graft endothelium immediately after transplantation, leading to thrombosis, ischemic damage, and rapid graft failure.
  - Acute cellular rejection. T cells destroy graft parenchyma (and vessels) by cytotoxicity and inflammatory reactions.
  - Acute humoral rejection. Antibodies damage graft vasculature.
  - Chronic rejection. Dominated by arteriosclerosis, this type is caused by T-cell activation and antibodies. The T-cells may secrete cytokines that induce proliferation of vascular smooth muscle cells, and the antibodies cause endothelial injury. The vascular lesions and T-cell reactions cause parenchymal fibrosis.
- Treatment of graft rejection relies on immunosuppressive drugs, which inhibit immune responses against the graft.
- Transplantation of HSCs requires careful matching of donor and recipient and is often complicated by GVHD and immune deficiency.

IMMUNODEFICIENCY DISEASES

Immunodeficiency diseases can be divided into primary (or congenital) immunodeficiencies, which are genetically determined, and secondary (or acquired) immunodeficiencies, which may arise as complications of cancers, infections, malnutrition, or side effects of immunosuppression, irradiation, or chemotherapy for cancer and other diseases. Immunodeficiencies are manifested clinically by increased infections, which may be newly acquired or a reactivation of latent infection. The primary immunodeficiency syndromes are accidents of nature that provide valuable insights into some of the critical molecules of the human immune system. Here we briefly discuss the more important and best-defined primary immunodeficiencies, to be followed by a more detailed description of acquired immunodeficiency syndrome (AIDS), the most devastating example of secondary immunodeficiency.

Primary Immunodeficiencies

Primary immunodeficiency diseases are caused by genetic (inherited) defects that affect the defense mechanisms of innate immunity (phagocytes, NK cells, or complement) or the humoral and/or cellular arms of adaptive immunity (mediated by B and T lymphocytes, respectively). Most of

these diseases are detected in infancy, between 6 months and 2 years of life, the telltale signs being susceptibility to recurrent infections. Here we present selected examples of primary immunodeficiencies, beginning with defects in innate immunity and then defects in the maturation and activation of B and T lymphocytes. We conclude with immune defects associated with some systemic diseases.

Defects in Innate Immunity

Inherited defects in the early innate immune response typically affect leukocyte functions or the complement system, and all lead to increased vulnerability to infections (Table 6.12). Some of the defects whose molecular basis is defined are summarized next.

Defects in Leukocyte Function

- *Inherited defects in leukocyte adhesion.* Individuals with *leukocyte adhesion deficiency type 1* have a defect in the biosynthesis of the β_2 chain shared by the integrins LFA-1 and Mac-1. *Leukocyte adhesion deficiency type 2* is caused by the absence of sialyl-Lewis X, the fucose-containing ligand for E- and P-selectins, as a result of a defect in a fucosyl transferase, an enzyme that attaches fucose moieties to protein backbones. Both conditions result in a failure of leukocyte adhesion to endothelium, preventing the cells from migrating into tissues and making patients prone to bacterial infections, which are often recurrent and frequently life threatening.
- *Inherited defects in phagolysosome function.* One such disorder is *Chédiak-Higashi syndrome,* an autosomal recessive condition characterized by defective fusion of phagosomes and lysosomes, resulting in defective phagocyte function and susceptibility to infections. The main leukocyte abnormalities are neutropenia (decreased numbers of neutrophils), defective degranulation, and delayed microbial killing. Leukocytes contain giant granules, which can be readily seen in peripheral blood smears and are thought to result from aberrant phagolysosome fusion. In addition, there are abnormalities in melanocytes (leading to albinism), cells of the nervous system (associated with nerve defects), and platelets (causing bleeding disorders). The dysfunctional gene underlying this disorder encodes a cytosolic protein called LYST, which is believed to regulate lysosomal trafficking.
- *Inherited defects in microbicidal activity.* The importance of oxygen-dependent bactericidal mechanisms is shown by the existence of a group of congenital disorders called *chronic granulomatous disease* (CGD), which are characterized by defects in bacterial killing that render patients susceptible to recurrent bacterial infection. CGD results from inherited defects in the genes encoding components of phagocyte oxidase, the phagolysosomal enzyme that generates superoxide ($O_2^{\cdot-}$). The most common variants are an X-linked defect in one of the membrane-bound components (gp91phox) and autosomal recessive defects in the genes encoding two of the cytoplasmic components (p47phox and p67phox). The name of this disease comes from the macrophage-rich chronic inflammatory reaction that tries to control the infection when the initial neutrophil defense is inadequate. This often leads to collections of activated macrophages that wall off the microbes, forming granulomas.
- *Defects in TLR signaling.* Rare defects have been described in various TLRs and their signaling molecules. Defects in TLR3, a receptor for viral RNA, result in recurrent herpes simplex encephalitis, and defects in MyD88, the adaptor protein downstream of multiple TLRs, are associated with destructive bacterial pneumonias.

Deficiencies Affecting the Complement System

Hereditary deficiencies have been described for virtually all components of the complement system and several of the regulators. In addition, one disease, paroxysmal nocturnal hemoglobinuria, is caused by an acquired deficiency of complement regulatory factors.

- Deficiency of C2 is the most common complement protein deficiency. A deficiency of C2 or C4, early components of the classical pathway, is associated with increased bacterial or viral infections. However, many patients have no clinical manifestations, presumably because the alternative complement pathway is adequate for the control of most infections. Surprisingly, in some of these patients, as well as in patients with C1q deficiency, the dominant manifestation is SLE-like autoimmune disease, as discussed earlier.

Table 6.12 Defects in Innate Immunity

Disease	Defect
Defects in Leukocyte Function	
Leukocyte adhesion deficiency 1	Defective leukocyte adhesion because of mutations in β chain of CD11/CD18 integrins
Leukocyte adhesion deficiency 2	Defective leukocyte adhesion because of mutations in fucosyl transferase required for synthesis of sialylated oligosaccharide (receptor for selectins)
Chédiak-Higashi syndrome	Decreased leukocyte functions because of mutations affecting protein involved in lysosomal membrane traffic
Chronic granulomatous disease	Decreased oxidative burst
X-linked	Phagocyte oxidase (membrane component)
Autosomal recessive	Phagocyte oxidase (cytoplasmic components)
Myeloperoxidase deficiency	Decreased microbial killing because of defective MPO-H_2O_2 system
Defects in the Complement System	
C2, C4 deficiency	Defective classical pathway activation results in reduced resistance to infection and reduced clearance of immune complexes
C3 deficiency	Defects in all complement functions
Deficiency of complement regulatory proteins	Excessive complement activation; clinical syndromes include angioedema, paroxysmal hemoglobinuria, others

The table lists some of the more common inherited immune deficiencies affecting phagocytic leukocytes and the complement system.

Modified in part from Gallin JI: Disorders of phagocytic cells. In Gallin JI, Goldstein IM, Snyderman R, editors: *Inflammation: Basic Principles and Clinical Correlates,* ed 2, New York, 1992, Raven Press, pp 860–861.

- Deficiency of components of the alternative pathway (properdin and factor D) is rare. It is associated with recurrent pyogenic infections.
- The C3 component of complement is required for both the classical and alternative pathways, and hence a deficiency of this protein results in susceptibility to serious and recurrent pyogenic infections. There is also an increased incidence of immune complex–mediated glomerulonephritis. Complement is involved in the removal of immune complexes, and in its absence inflammation is presumably caused by Fc receptor–dependent leukocyte activation.
- The terminal components of complement (C5, 6, 7, 8, and 9) are required for the assembly of the membrane attack complex involved in the lysis of organisms. Deficiency of these late-acting components is associated with increased susceptibility to recurrent *Neisseria* (gonococcal and meningococcal) infections; *Neisseria* bacteria have thin cell walls that make them susceptible to the lytic actions of complement.
- Other patients inherit a defective form of mannose-binding lectin, the plasma protein that initiates the lectin pathway of complement. These individuals also show increased susceptibility to infections.
- A deficiency of C1 inhibitor (C1 INH) gives rise to *hereditary angioedema*. This autosomal dominant disorder is more common than complement deficiency states. The C1 inhibitor's targets are proteases, specifically C1r and C1s of the complement cascade, factor XII of the coagulation pathway, and the kallikrein system. With deficiency of C1 INH, unregulated activation of kallikrein may lead to increased production of vasoactive peptides such as bradykinin. Although the exact nature of the bioactive compound produced in hereditary angioedema is uncertain, these patients have episodes of edema affecting skin and mucosal surfaces such as the larynx and the gastrointestinal tract. This may result in life-threatening asphyxia or nausea, vomiting, and diarrhea after minor trauma or emotional stress. Acute attacks of hereditary angioedema can be treated with C1 inhibitor concentrates prepared from human plasma.
- Deficiencies of other complement regulatory proteins are causes of paroxysmal nocturnal hemoglobinuria (Chapter 14), chronic forms of hemolytic uremic syndrome (Chapter 20), and age-related macular degeneration.

Defects in Adaptive Immunity

Defects in adaptive immunity are often subclassified on the basis of the primary component involved (i.e., B cells or T cells, or both). However, these distinctions are not clear-cut; for instance, T-cell defects almost always lead to impaired antibody synthesis, and hence isolated deficiencies of T cells are often indistinguishable clinically from combined deficiencies of T and B cells. These immunodeficiencies result from abnormalities in lymphocyte maturation or activation. The mutations responsible for many of these diseases have now been identified (Fig. 6.38).

Severe Combined Immunodeficiency

Severe combined immunodeficiency (SCID) spans a constellation of genetically distinct syndromes, all having in common defects in both humoral and cell-mediated immune responses. Affected infants present with prominent thrush (oral candidiasis), extensive diaper rash, and failure to thrive. Some patients develop a morbilliform rash shortly after birth because maternal T cells are transferred across the placenta and attack the fetus, causing GVHD. Persons with SCID are extremely susceptible to recurrent, severe infections by a wide range of pathogens, including *Candida albicans, Pneumocystis jirovecii, Pseudomonas,* cytomegalovirus, varicella, and a host of bacteria. Without HSC transplantation, death occurs within the first year of life. Despite the common clinical manifestations, the underlying genetic defects are quite varied and in many cases are unknown. Often, the SCID defect resides in the T-cell compartment, with a secondary impairment of humoral immunity.

X-linked SCID. The most common form, accounting for 50% to 60% of cases, is X-linked, and hence SCID is more common in boys than in girls. The genetic defect in the X-linked form is a mutation in the common γ-chain (γc) subunit of cytokine receptors. This transmembrane protein is a signal-transducing component of the receptors for IL-2, IL-4, IL-7, IL-9, IL-11, IL-15, and IL-21. IL-7 is required for the survival and proliferation of lymphoid progenitors, particularly T-cell precursors. As a result of defective IL-7 receptor signaling, there is a profound defect in the earliest stages of lymphocyte development, especially T-cell development. T-cell numbers are greatly reduced, and although B cells may be normal in number, antibody synthesis is impaired because of lack of T-cell help. IL-15 is important for the maturation and proliferation of NK cells, and because the common γ chain is a component of the receptor for IL-15, these individuals often have a deficiency of NK cells as well.

Autosomal Recessive SCID. The remaining forms of SCID are autosomal recessive disorders. The most common cause of autosomal recessive SCID is a deficiency of the enzyme adenosine deaminase (ADA). Although the mechanisms by which ADA deficiency causes SCID are not entirely clear, it has been proposed that deficiency of the enzyme leads to accumulation of deoxyadenosine and its derivatives (e.g., deoxy-ATP), which are toxic to rapidly dividing immature lymphocytes, especially those of the T-cell lineage. Hence there may be a greater reduction in the number of T lymphocytes than of B lymphocytes.

Several other less common causes of autosomal recessive SCID have been identified:

- Mutations in recombinase-activating genes (RAG) or other components of the antigen receptor gene recombination machinery prevent the somatic gene rearrangements that are essential for the assembly of TCR and Ig genes. This blocks the development of T and B cells.
- An intracellular kinase called JAK3 is essential for signal transduction through cytokine receptors containing the common γ chain (which is mutated in X-linked SCID, as discussed earlier). Mutations of JAK3 therefore have the same effects as mutations in the γc chain.
- Several mutations have been described in signaling molecules, including kinases associated with the T-cell antigen receptor and components of calcium channels that are required for entry of calcium and activation of many signaling pathways.

The histologic findings in SCID depend on the underlying defect. In the two most common forms (γc mutation and ADA deficiency), the thymus is small and devoid of

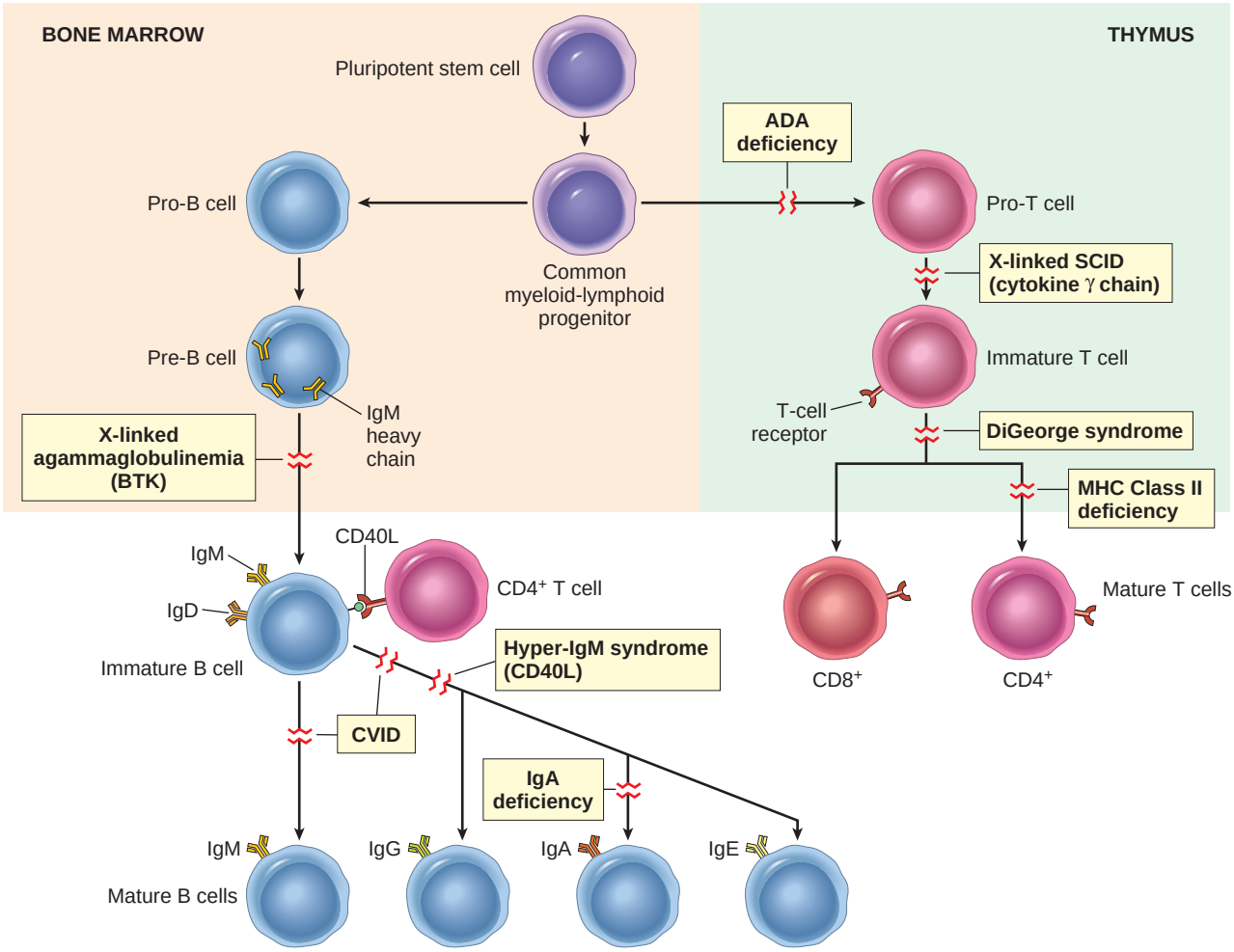


Figure 6.38 Primary immune deficiency diseases. Shown are the principal pathways of lymphocyte development and the blocks in these pathways in selected primary immune deficiency diseases. The affected genes are indicated in parentheses for some of the disorders. *ADA*, Adenosine deaminase; *CD40L*, CD40 ligand (also known as *CD154*); *CVID*, common variable immunodeficiency; *SCID*, severe combined immunodeficiency.

lymphoid cells. In X-linked SCID, the thymus contains lobules of undifferentiated epithelial cells resembling fetal thymus, whereas in SCID caused by ADA deficiency, remnants of Hassall's corpuscles can be found. In both diseases, other lymphoid tissues are hypoplastic as well, with marked depletion of T-cell areas and in some cases both T-cell and B-cell zones.

Currently, HSC transplantation is the mainstay of treatment, but X-linked SCID is the first human disease in which gene therapy has been successful. In this type of ex vivo gene therapy, a normal γc gene is expressed using a viral vector in HSCs taken from patients, and the cells are then transplanted back into the patients. The clinical experience is small, but some patients have shown reconstitution of their immune systems for over 10 years after therapy. Unfortunately, about 20% of patients in the initial clinical trials developed T-cell lymphoblastic leukemia, highlighting the dangers of this particular approach to gene therapy. The uncontrolled T-cell proliferation was triggered by the activation of oncogenes by the integrated virus, an effect that may have been augmented by a growth advantage conferred by the introduced γc gene. Current trials are using new vectors with safety features built in. Patients with ADA deficiency have also been treated with HSC transplantation and, more recently, with administration of the enzyme or gene therapy involving the introduction of a normal *ADA* gene into T-cell precursors.

X-Linked Agammaglobulinemia (Bruton Agammaglobulinemia)

X-linked agammaglobulinemia is characterized by the failure of B-cell precursors (pro–B cells and pre–B cells) to develop into mature B cells. It is one of the more common forms of primary immunodeficiency. During normal B-cell maturation in the bone marrow, the Ig heavy-chain genes are rearranged first, in pre–B cells. These form a complex with a "surrogate" light chain on the cell surface called the pre–B cell receptor (pre-BCR) that delivers signals that induce rearrangement of the Ig light-chain genes and further maturation. This need for Ig-initiated signals is a quality-control mechanism that ensures that maturation will proceed only if functional Ig proteins are expressed. X-linked agammaglobulinemia is caused by mutations in a cytoplasmic tyrosine kinase, called Bruton tyrosine kinase (BTK); the gene that encodes it is located on the long arm of the X chromosome at Xq21.22. BTK is a protein tyrosine kinase that is associated with the pre-BCR and with the B-cell receptor (BCR) complexes that are found on mature B

cells. When BTK is mutated, the pre-BCR cannot deliver the signals that are needed for light chain rearrangement, and maturation is arrested. BTK is also an important transducer of BCR signals that stimulate growth and increase cell survival in benign and malignant mature B cells, and inhibitors of BTK are effective therapies for several B cell malignancies (Chapter 13).

As an X-linked disease, this disorder is seen almost exclusively in males, but sporadic cases have been described in females, possibly caused by mutations in other genes that function in the same pathway. The disease usually does not become apparent until about 6 months of age, as maternal immunoglobulins are depleted. In most cases, recurrent bacterial infections of the respiratory tract, such as acute and chronic pharyngitis, sinusitis, otitis media, bronchitis, and pneumonia, call attention to the underlying immune defect. Almost always the causative organisms are *Haemophilus influenzae, Streptococcus pneumoniae,* or *Staphylococcus aureus*. These organisms are normally opsonized by antibodies and cleared by phagocytosis. Because antibodies are important for neutralizing infectious viruses, individuals with this disease are also susceptible to certain viral infections, especially those caused by enteroviruses, such as echovirus, poliovirus, and coxsackievirus. These viruses infect the gastrointestinal tract, and from here they can disseminate to the nervous system via the blood. Thus, immunization with live poliovirus carries the risk of paralytic poliomyelitis, and echovirus can cause fatal encephalitis. For similar reasons, *Giardia lamblia,* an intestinal protozoan that is normally resisted by secreted IgA, causes persistent infections in persons with this disorder. In general, however, most fungal, protozoal, and intracellular viral infections are handled quite well by the intact T cell–mediated immunity.

The classic form of this disease has the following characteristics:

- B cells are absent or markedly decreased in the circulation, and serum levels of all classes of immunoglobulins are depressed. Pre–B cells, which express the B-lineage marker CD19 but not membrane Ig, are found in normal numbers in the bone marrow.
- Germinal centers of lymph nodes, Peyer patches, the appendix, and tonsils are underdeveloped.
- Plasma cells are absent throughout the body.
- T cell–mediated reactions are normal.

Paradoxically, autoimmune diseases, such as arthritis and dermatomyositis, occur in as many as 30% of individuals with this disease. It is likely that these autoimmune disorders are caused by a breakdown of self-tolerance resulting in autoimmunity, but chronic infections associated with the immune deficiency may play a role in inducing the inflammatory reactions. The treatment of X-linked agammaglobulinemia is replacement therapy with immunoglobulins. In the past, most patients succumbed to infection in infancy or early childhood. Prophylactic IVIG therapy allows most individuals to reach adulthood.

DiGeorge Syndrome (Thymic Hypoplasia)

DiGeorge syndrome is a T-cell deficiency that results from failure of development of the thymus. The third and fourth pharyngeal pouches, which give rise to the thymus, the parathyroids, some of the C cells of the thyroid, and the ultimobranchial body, do not develop normally. Thus, individuals with this syndrome have a variable loss of T cell–mediated immunity (resulting from hypoplasia or lack of the thymus), tetany (resulting from lack of the parathyroids), and congenital defects of the heart and great vessels. In addition, the appearance of the mouth, ears, and facies may be abnormal. Absence of cell-mediated immunity is caused by low numbers of T lymphocytes in the blood and lymphoid tissues and poor defense against certain fungal and viral infections. The T-cell zones of lymphoid organs—paracortical areas of the lymph nodes and the periarteriolar sheaths of the spleen—are depleted. Ig levels may be normal or reduced, depending on the severity of the T-cell deficiency.

In the vast majority (90%) of cases, DiGeorge syndrome is caused by a small germline deletion that maps to chromosome 22q11, and DiGeorge syndrome is now considered a component of the 22q11 deletion syndrome, discussed in Chapter 5. One gene in the deleted region is *TBX1,* which is required for development of the branchial arch and the great vessels. Notably, *TBX1* is involved by loss-of-function mutations in a few cases of DiGeorge syndrome that lack 22q11 deletions, strongly suggesting that its loss contributes to the observed phenotype.

Other Defects in Lymphocyte Maturation

Many other rare causes of immunodeficiency resulting from defective lymphocyte maturation have been documented. One of these, *bare lymphocyte syndrome,* is usually caused by mutations in transcription factors that are required for class II MHC gene expression. The lack of class II MHC molecules prevents the development of CD4+ T cells, which are involved in cellular immunity and provide help to B cells; hence, class II MHC deficiency results in combined immunodeficiency. Other defects are caused by mutations in antigen receptor chains or signaling molecules involved in T- or B-cell maturation.

Hyper-IgM Syndrome

In this disorder, affected patients have IgM antibodies but are deficient in IgG, IgA, and IgE antibodies. Mature B cells are present, but they are incapable of Ig class switching and affinity maturation, because of a defect in CD4+ helper T cells or an intrinsic B-cell defect. As discussed earlier in the chapter, many of the functions of CD4+ helper T cells require the engagement of CD40 on B cells, macrophages, and DCs by CD40L (also called CD154) expressed on antigen-activated T cells. This interaction triggers Ig class switching and affinity maturation in B cells, and also stimulates the microbicidal functions of macrophages. Approximately 70% of individuals with hyper-IgM syndrome have the X-linked form of the disease, caused by mutations in the gene encoding CD40L located on Xq26 that interfere with CD4+ helper T cell function. In the remaining patients, the disease is inherited in an autosomal recessive pattern. Most of these patients have loss-of-function mutations involving either CD40 or activation-induced cytidine deaminase (AID), a DNA-editing enzyme expressed in B cells that is required for Ig class switching and affinity maturation.

The serum of persons with this syndrome contains normal or elevated levels of IgM but no IgA or IgE and extremely low levels of IgG. The numbers of B and T cells in the blood are normal. Clinically, patients present with recurrent pyogenic

infections, because the level of opsonizing IgG antibodies is low, and also because affinity maturation, a process necessary for production of high-affinity antibodies, is impaired. In addition, those with CD40L mutations are susceptible to pneumonia caused by the intracellular organism *Pneumocystis jirovecii* because CD40L-mediated macrophage activation, a key reaction of cell-mediated immunity, is also defective. Occasionally, the IgM antibodies react with blood cells, giving rise to autoimmune hemolytic anemia, thrombocytopenia, and neutropenia. In older patients, there may be a proliferation of IgM-producing plasma cells that infiltrates the mucosa of the gastrointestinal tract.

Common Variable Immunodeficiency

This relatively frequent entity encompasses a heterogeneous group of disorders in which the common feature is hypogammaglobulinemia, generally affecting all the antibody classes but sometimes only IgG. The diagnosis of common variable immunodeficiency is based on exclusion of other well-defined causes of decreased antibody production.

Both sporadic and inherited forms of the disease occur. In familial forms, there is no single pattern of inheritance. Relatives of such patients have a high incidence of selective IgA deficiency (see later), suggesting that at least in some cases, selective IgA deficiency and common variable immunodeficiency represent different expressions of a common genetic defect in antibody synthesis. In contrast to X-linked agammaglobulinemia, most individuals with common variable immunodeficiency have normal or near-normal numbers of B cells in the blood and lymphoid tissues. These B cells, however, are not able to differentiate into plasma cells.

Both intrinsic B-cell defects and abnormalities in helper T cell–mediated activation of B cells may account for the antibody deficiency in this disease. Families have been reported in which the underlying abnormality is in a receptor for a cytokine called *BAFF* that promotes the survival and differentiation of B cells, or in a molecule called *ICOS* (inducible costimulator) that is homologous to CD28 and is involved in T-cell activation and in interactions between T and B cells. However, the known mutations explain less than 10% of cases.

The clinical manifestations of common variable immunodeficiency are caused by antibody deficiency, and hence they resemble those of X-linked agammaglobulinemia. Patients typically present with recurrent sinopulmonary pyogenic infections. In addition, about 20% of patients have recurrent herpesvirus infections. Serious enterovirus infections causing meningoencephalitis may also occur. Individuals with this disorder are also prone to the development of persistent diarrhea caused by *G. lamblia.* Common variable immunodeficiency affects both sexes equally, and the onset of symptoms is later than in X-linked agammaglobinemina, in childhood or adolescence. Histologically, the germinal centers in lymphoid tissues (i.e., nodes, spleen, and gut) are hyperplastic due to the presence of cells that can proliferate in response to antigen but are defective in further maturation in some respect.

As in X-linked agammaglobulinemia, these patients have a high frequency of autoimmune diseases (approximately 20%), including rheumatoid arthritis. The risk of lymphoid malignancy is also increased, and an increase in gastric cancer has been reported.

Isolated IgA Deficiency

Isolated IgA deficiency is a common immunodeficiency caused by impaired differentiation of naïve B lymphocytes to IgA-producing plasma cells. In the United States, it occurs in about 1 in 600 individuals of European descent. It is far less common in people of African descent and Asians. Affected individuals have extremely low levels of both serum and secretory IgA. The molecular basis of this defect in most patients is unknown; defects in a receptor for a B cell–activating cytokine, BAFF, have been described in some patients.

Most individuals with IgA deficiency are asymptomatic. Because IgA is the major antibody in mucosal secretions, mucosal defenses are weakened, and infections occur in the respiratory, gastrointestinal, and urogenital tracts. Symptomatic patients commonly present with recurrent sinopulmonary infections and diarrhea. In addition, IgA-deficient patients have a high frequency of respiratory tract allergy and a variety of autoimmune diseases, particularly SLE and rheumatoid arthritis. The basis of the increased frequency of autoimmune and allergic diseases is not known. When transfused with blood containing normal IgA, some patients develop severe, even fatal, anaphylactic reactions, because the IgA behaves like a foreign antigen.

X-Linked Lymphoproliferative Disease

X-linked lymphoproliferative disease is characterized by an inability to eliminate EBV, eventually leading to fulminant infectious mononucleosis and the development of B-cell tumors. In about 80% of cases, the disease is due to mutations in the gene encoding an adapter molecule called SLAM-associated protein (SAP). SAP binds to a family of cell surface molecules involved in the activation of NK cells and T and B lymphocytes, including the signaling lymphocyte activation molecule (SLAM). Defects in SAP attenuate NK and T-cell activation and result in increased susceptibility to viral infections. SAP is also required for the development of T follicular helper cells, and because of this defect X-linked lymphoproliferative disease patients are unable to form germinal centers or produce high-affinity antibodies, additional abnormalities that also likely contribute to susceptibility to viral infection. This immunodeficiency is most commonly manifested by severe EBV infection, including severe and often fatal infectious mononucleosis (Chapter 8), but not other viral infections, for reasons that are not clear.

Other Defects in Lymphocyte Activation

Many rare cases of lymphocyte activation defects have been described, affecting antigen receptor signaling and various biochemical pathways. Mutations affecting Th1 responses are associated with atypical mycobacterial infections; the syndrome is called *Mendelian susceptibility to mycobacterial disease*. Inherited defects in Th17 responses lead to chronic mucocutaneous candidiasis and bacterial infections of the skin (a disorder called *Job syndrome*).

Immunodeficiencies Associated With Systemic Diseases

In some inherited systemic disorders, immune deficiency is a prominent clinical problem. Two representative examples of such diseases are described next.

Wiskott-Aldrich Syndrome

Wiskott-Aldrich syndrome is an X-linked disease characterized by thrombocytopenia, eczema, and a marked vulnerability to recurrent infection, resulting in early death. The thymus is morphologically normal, at least early in the course of the disease, but there is progressive loss of T lymphocytes in the peripheral blood and in the T-cell zones (paracortical areas) of the lymph nodes, with variable defects in cellular immunity. Patients do not make antibodies to polysaccharide antigens, and the response to protein antigens is poor. IgM levels in the serum are low, but levels of IgG are usually normal. Paradoxically the levels of IgA and IgE are often elevated. Patients are also prone to developing B-cell lymphomas. Wiskott-Aldrich syndrome is caused by mutations in the gene located at Xp11.23 that encodes Wiskott-Aldrich syndrome protein (WASP). WASP belongs to a family of proteins that are believed to link membrane receptors, such as antigen receptors, to cytoskeletal elements. WASP may be involved in cytoskeleton-dependent responses, including cell migration and signal transduction, but the essential functions of this protein in lymphocytes and platelets are unclear. The only treatment is HSC transplantation.

Ataxia Telangiectasia

Ataxia telangiectasia is an autosomal-recessive disorder characterized by abnormal gait (ataxia), vascular malformations (telangiectases), neurologic deficits, increased incidence of tumors, and immunodeficiency. The immunologic defects are of variable severity and may affect both B and T cells. The most prominent humoral immune abnormalities are defective production of isotype-switched antibodies, mainly IgA and IgG_2. The T-cell defects, which are usually less pronounced, are associated with thymic hypoplasia. Patients experience upper and lower respiratory tract bacterial infections, multiple autoimmune phenomena, and increasingly frequent cancers with advancing age. The gene responsible for this disorder is located on chromosome 11 and encodes a protein kinase called ATM (ataxia telangiectasia mutated). In response to DNA damage (double-strand breaks), ATM activates p53 by phosphorylation, which in turn can activate cell cycle checkpoints and apoptosis in cells with damaged DNA. ATM has also been shown to contribute to the stability of DNA double-strand break complexes during V(D)J recombination. The abnormalities in DNA repair resulting from ATM deficiency may impair the generation of antigen receptors. In addition, defective DNA repair may lead to abnormalities in the DNA recombination events that are involved in Ig isotype switching. Like several other immunodeficiency syndromes, patients with ataxia telangiectasia have a markedly increased incidence of lymphoma.

KEY CONCEPTS

PRIMARY (INHERITED) IMMUNODEFICIENCIES

- These diseases are caused by inherited mutations in genes involved in lymphocyte maturation or function, or in innate immunity.
- Deficiencies in innate immunity include defects of phagocyte function, complement, and innate immune receptors.
- Some of the common disorders affecting lymphocytes and the adaptive immune response are:
  - X-SCID: failure of T-cell and B-cell maturation; mutation in the common γ chain of a cytokine receptor, leading to failure of IL-7 signaling and defective lymphopoiesis
  - Autosomal recessive SCID: failure of T-cell development, secondary defect in antibody responses; approximately 50% of cases caused by mutation in the gene encoding ADA, leading to accumulation of toxic metabolites during lymphocyte maturation and proliferation
  - X-linked agammaglubulinemia: failure of B-cell maturation, absence of antibodies; caused by mutations in the *BTK* gene, which encodes B-cell tyrosine kinase, required for maturation signals from the pre-BCRs and BCRs
  - Common variable immunodeficiency: defects in antibody production; cause unknown in most cases
  - Selective IgA deficiency: failure of IgA production; cause unknown
  - X-linked hyper-IgM syndrome: failure to produce isotype-switched high-affinity antibodies (IgG, IgA, IgE); mutation in gene encoding CD40L
  - X-linked lymphoproliferative disease: defect in a signaling molecule causing defective responses against Epstein-Barr virus and lymphoproliferation
- These diseases present clinically with increased susceptibility to infections in early life.

Secondary Immunodeficiencies

Secondary (acquired) immune deficiencies may be encountered in individuals with cancer, diabetes and other metabolic diseases, malnutrition, chronic infection, and in persons receiving chemotherapy or radiation therapy for cancer, or immunosuppressive drugs to prevent graft rejection or to treat autoimmune diseases (Table 6.13). As a group, the secondary immune deficiencies are more common than the disorders of primary genetic origin. Some of these secondary immunodeficiency states can be caused by defective lymphocyte maturation (when the bone marrow is damaged by radiation or chemotherapy or involved by tumors, such as leukemias), inadequate Ig synthesis (as in malnutrition), or lymphocyte depletion (from drugs or severe infections). The most serious secondary immunodeficiency is AIDS, which is described next.

Table 6.13 Causes of Secondary (Acquired) Immunodeficiencies

Cause	Mechanism
Human immunodeficiency virus infection	Depletion of CD4+ helper T cells
Irradiation and chemotherapy treatments for cancer	Decreased bone marrow precursors for all leukocytes
Involvement of bone marrow by cancers (metastases, leukemias)	Reduced site of leukocyte development
Protein-calorie malnutrition	Metabolic derangements inhibit lymphocyte maturation and function
Removal of spleen	Decreased phagocytosis of microbes

Acquired Immunodeficiency Syndrome (AIDS)

AIDS is caused by the retrovirus human immunodeficiency virus (HIV) and is characterized by profound immunosuppression that leads to opportunistic infections, secondary neoplasms, and neurologic manifestations. The magnitude of this modern plague is truly staggering. In the United States, AIDS is the second leading cause of death in men between 25 and 44 years of age, and the third leading cause of death in women in this age group. Although initially recognized in the United States, AIDS is a global problem. It has now been reported from more than 190 countries around the world, and the pool of HIV-infected persons in Africa and Asia is large and expanding. By 2016, HIV had infected 60 million people worldwide, and nearly 30 million adults and children have died of the disease. There are about 37 million people living with HIV (1.1 million in the United States), of whom 70% are in Africa and more than 20% in Asia; the prevalence rate of infection in adults in sub-Saharan Africa is 7% overall and exceeds 25% in some countries. It is estimated that 1.7 million people were newly infected with HIV in 2018, and about 800,000 deaths were caused by AIDS. Only about 80% of HIV-infected people living worldwide know their status.

There is some good news. Because of public health measures, the infection rate seems to be decreasing, and some authorities believe it may have peaked in the late 1990s. Furthermore, improved antiviral therapies have resulted in fewer people dying of the disease. However, these newer treatments are not readily available in many low income countries, and toxic side effects remain a problem. The advent of these drugs raises its own tragic concern; because more people are living with HIV, the risk of spreading the infection will increase if vigilance is relaxed.

The enormous medical and social burden of AIDS has led to an explosion of research aimed at understanding HIV and its remarkable ability to cripple host defenses. The literature on HIV and AIDS is vast. Here we summarize the currently available data on the epidemiology, pathogenesis, and clinical features of HIV infection.

Epidemiology

Epidemiologic studies in the United States have identified five groups of adults at high risk for developing AIDS. The case distribution in these groups is as follows:

- *Men who have sex with men* account for more than 50% of the reported cases. This includes about 5% who are intravenous drug users as well. Transmission of HIV in this category appears to be on the decline, but in 2016 this remained the largest affected group, constituting about 70% of new cases.
- *Heterosexual transmission,* chiefly due to contact with members of other high-risk groups (e.g., intravenous drug users), is responsible for approximately 20% of cases in the United States. Heterosexual transmission has declined modestly since 2011 in the United States. Globally, however, heterosexual transmission is by far the most common mode by which HIV is spread, and as a result, women are much more frequently infected outside of the United States. In sub-Saharan Africa, where the infection rate is estimated to be about 10,000 new cases every day, more than one-half of infected individuals are women. Heterosexual spread of the virus is occurring most rapidly in female sex workers and in women in long-term marital or cohabitating relationships, particularly among adolescents.
- *Intravenous drug users* with no previous history of homosexuality are the next largest group, representing about 20% of infected individuals.
- *HIV infection of the newborn.* Children of HIV-positive women are at risk for infection in utero, at birth, or through breast milk (discussed later). Close to 2% of all AIDS cases occur in children, and it is estimated that worldwide there are 1.7 million HIV-infected individuals younger than 15 years of age.
- Patients with *hemophilia,* especially those who received large amounts of factor VIII or factor IX concentrates before 1985, make up about 0.5% of all cases.
- *Recipients of blood and blood components* who are not hemophiliacs but who received transfusions of HIV-infected whole blood or components (e.g., platelets, plasma) account for about 1% of patients. (Organs obtained from HIV-infected donors can also transmit the virus.)
- In approximately 5% of cases, the risk factors cannot be determined.

It should be apparent from the preceding discussion that transmission of HIV occurs under conditions that facilitate exchange of blood or body fluids containing the virus or virus-infected cells. **The three major routes of transmission are sexual contact, parenteral inoculation, and passage of the virus from infected mothers to their newborns.**

- *Sexual transmission* is the dominant mode of infection worldwide, accounting for more than 75% of all cases of HIV transmission. Because the majority of infected people in the United States are men who have sex with men, most sexual transmission has occurred among homosexual men. The virus is carried in the semen, and it enters the recipient's body through abrasions in rectal or oral mucosa or by direct contact with mucosal lining cells. Viral transmission occurs in two ways: (1) direct inoculation into the blood vessels breached by trauma and (2) infection of DCs or CD4+ cells within the mucosa. In addition to male-to-male and male-to-female transmission, female-to-male transmission also occurs.

  Sexual transmission of HIV is enhanced by coexisting sexually transmitted diseases, especially those associated with genital ulceration. In this regard, syphilis, chancroid, and herpes are particularly important. Other sexually transmitted diseases, including gonorrhea and chlamydia, are also cofactors for HIV transmission, perhaps because in these genital inflammatory states there is greater concentration of the virus and virus-containing cells in genital fluids, as a result of increased numbers of inflammatory cells in the semen.
- *Parenteral transmission* of HIV has occurred in three groups of individuals: intravenous drug users, hemophiliacs who received factor VIII and factor IX concentrates, and random recipients of blood transfusion. Of these three, intravenous drug users constitute by far the largest group. Transmission occurs by sharing of needles, syringes, and other paraphernalia contaminated with HIV-containing blood.

Transmission of HIV by transfusion of blood or blood products, such as lyophilized factor VIII and factor IX concentrates, has been virtually eliminated. This fortunate outcome resulted from increasing use of recombinant clotting factors and from three public health measures: screening of donated blood and plasma for antibody to HIV, stringent purity criteria for factor VIII and factor IX preparations, and screening of donors on the basis of history. However, an extremely small risk of acquiring AIDS through transfusion of seronegative blood persists, because a recently infected individual may be antibody-negative. Currently, this risk is estimated to be 1 in more than 2 million units of blood transfused.

- As alluded to earlier, *mother-to-infant transmission* is the major cause of pediatric AIDS. Infected mothers can transmit the infection to their offspring by three routes: (1) in utero by transplacental spread, (2) during delivery through an infected birth canal, and (3) after birth by ingestion of breast milk. Of these, transmission during birth (intrapartum) and in the immediate period thereafter (peripartum) is considered to be the most common mode in the United States. The reported transmission rates vary from 7% to 49% in different parts of the world. Higher risk of transmission is associated with high maternal viral load and low CD4+ T-cell counts as well as chorioamnionitis. Fortunately, antiretroviral therapy (ART) given to infected pregnant women in the United States has virtually eliminated mother-to-child transmission, but it remains a major source of infection in areas where these treatments are not readily available.

Much concern has arisen in the lay public and among health care workers about the spread of HIV infection outside the high-risk groups. Extensive studies indicate that HIV infection cannot be transmitted by casual personal contact in the household, workplace, or school. Spread by insect bites is virtually impossible. Regarding transmission of HIV infection to health care workers, an extremely small but definite risk is present. Seroconversion has been documented after accidental needle-stick injury or exposure of nonintact skin to infected blood in laboratory accidents. After needle-stick accidents, the risk of seroconversion is believed to be about 0.3%, and antiretroviral therapy given within 24 to 48 hours of a needle stick can reduce the risk of infection eightfold. By comparison, approximately 30% of those accidentally exposed to hepatitis B–infected blood become seropositive.

Etiology: the Properties of HIV

HIV is a nontransforming human retrovirus belonging to the lentivirus family. Included in this group are feline immunodeficiency virus, simian immunodeficiency virus, visna virus of sheep, bovine immunodeficiency virus, and the equine infectious anemia virus.

Two genetically different but related forms of HIV, called *HIV-1 and HIV-2,* have been isolated from patients with AIDS. HIV-1 is the most common type associated with AIDS in the United States, Europe, and Central Africa, whereas HIV-2 causes a similar disease principally in West Africa and India. Specific tests for HIV-2 are available, and blood collected for transfusion is routinely screened for both HIV-1 and HIV-2 seropositivity. The ensuing discussion relates primarily to HIV-1 and diseases caused by it, but the information is generally applicable to HIV-2 as well.

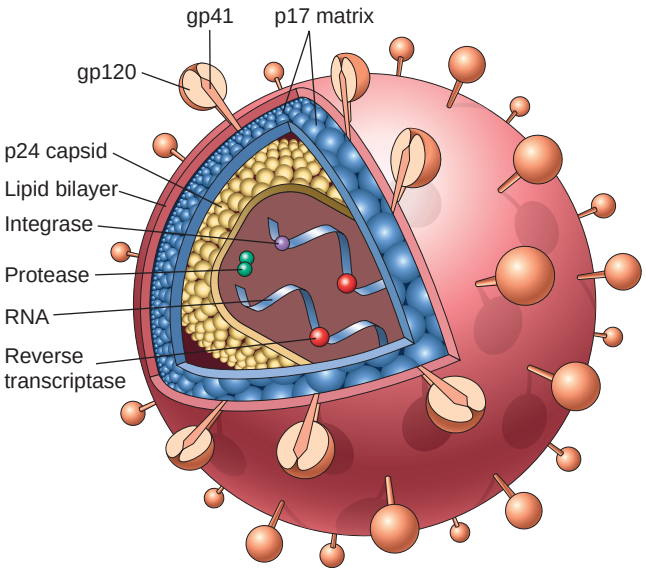


Figure 6.39 The structure of the HIV-1 virion. The viral particle is covered by a lipid bilayer derived from the host cell and studded with viral glycoproteins gp41 and gp120.

Structure of HIV

Similar to most retroviruses, the HIV-1 virion is spherical and contains an electron-dense, cone-shaped core surrounded by a lipid envelope derived from the host cell membrane (Fig. 6.39). **The virus core contains (1) the major capsid protein p24; (2) nucleocapsid protein p7/p9; (3) two copies of viral genomic RNA; and (4) the three viral enzymes (protease, reverse transcriptase, and integrase).** p24 is the most abundant viral antigen and is detected by an ELISA that is widely used to diagnose HIV infection. The viral core is surrounded by a matrix protein called *p17,* which lies underneath the virion envelope. Studding the viral envelope are two viral glycoproteins, gp120 and gp41, which are critical for HIV infection of cells.

The HIV-1 RNA genome contains the *gag, pol,* and *env* genes, which are typical of retroviruses (Fig. 6.40). The products of the *gag* and *pol* genes are large precursor proteins that are cleaved by the viral protease to yield the mature proteins. In addition to these three standard retroviral genes, HIV contains several other accessory genes, including *tat, rev, vif, nef, vpr,* and *vpu,* which regulate the synthesis and assembly of infectious viral particles and the pathogenicity of the virus. For example, the product of the *tat* (transactivator) gene causes a 1000-fold increase in the transcription of viral genes and is critical for virus replication. The functions of other accessory proteins are indicated in Fig. 6.40.

Pathogenesis of HIV Infection and AIDS

Although HIV can infect many tissues, **the major target of HIV infection is the immune system.** The central nervous system (CNS) is also affected.

Profound immune deficiency, primarily affecting cell-mediated immunity, is the hallmark of AIDS. This results chiefly from infection and subsequent death of CD4+ T cells

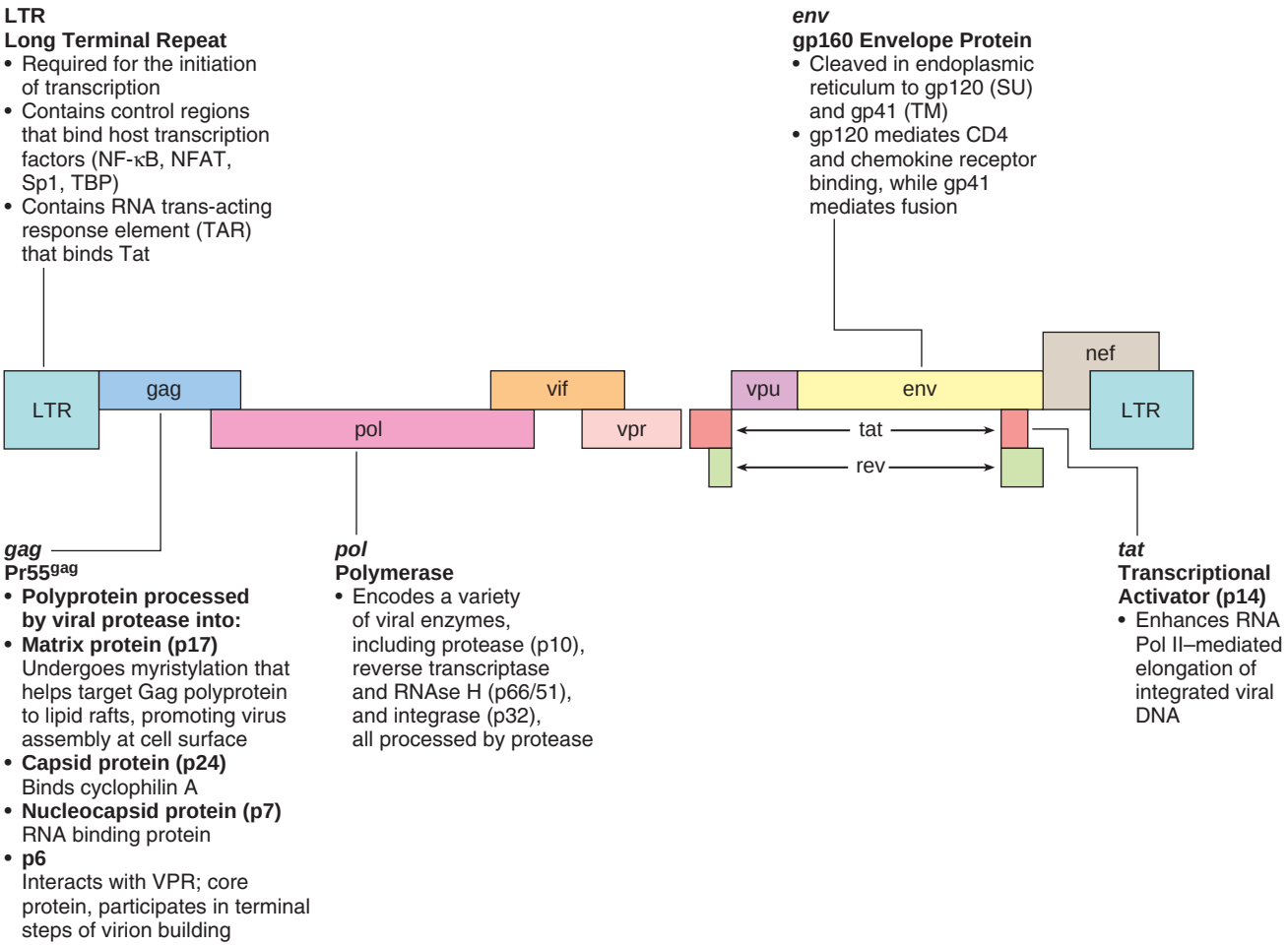


Figure 6.40 The HIV genome. The functions of selected genes are shown.

as well as impairment in the function of surviving helper T cells, but infection of macrophages and DCs also contributes (discussed later). HIV enters the body through mucosal tissues and blood and first infects T cells as well as DCs and macrophages. The infection becomes established in lymphoid tissues, where the virus may remain latent for long periods. Active viral replication is associated with more infection of cells and progression to AIDS. We first describe the mechanisms involved in viral entry into T cells and macrophages and the replicative cycle of the virus within cells. This is followed by a more detailed review of the interaction between HIV and its cellular targets.

Life Cycle of HIV

The life cycle of HIV consists of infection of cells, integration of the provirus into the host cell genome, activation of viral replication, and production and release of infectious virus (Fig. 6.41). The molecules and mechanisms of each of these steps are understood in considerable detail.

Infection of Cells by HIV. **HIV infects cells by using the CD4 molecule as a receptor and various chemokine receptors as coreceptors** (see Fig. 6.41). The requirement for CD4 binding explains the selective tropism of the virus for CD4+ T cells and other CD4+ cells, particularly monocytes/macrophages and DCs. Binding to CD4 is not sufficient for infection, however. HIV gp120 must also bind to other cell surface molecules (coreceptors) for entry into the cell. Chemokine receptors, particularly CCR5 and CXCR4, serve this role. HIV isolates can be distinguished by their use of these receptors: R5 strains use CCR5, X4 strains use CXCR4, and some strains (R5X4) use both. R5 strains preferentially infect cells of the monocyte/macrophage lineage and are thus referred to as M-tropic, whereas X4 strains are T-tropic, preferentially infecting T cells. In approximately 90% of cases, the R5 (M-tropic) type of HIV is the dominant virus found in the blood of acutely infected individuals and early in the course of infection. Over the course of infection, however, T-tropic viruses gradually accumulate; these are especially virulent because T-tropic viruses are capable of infecting many T cells and even thymic T-cell precursors and cause greater T-cell depletion and impairment.

Molecular details of the deadly handshake between HIV glycoproteins and their cell surface receptors have been elucidated and are important to understand because they may provide the basis for additional anti-HIV therapies. The HIV envelope contains two noncovalently associated glycoproteins, surface gp120 and the transmembrane protein gp41. The initial step in infection is the binding of the gp120 envelope glycoprotein to CD4, which leads to a conformational change that produces a new recognition site on gp120 for the coreceptors CCR5 or CXCR4. Binding

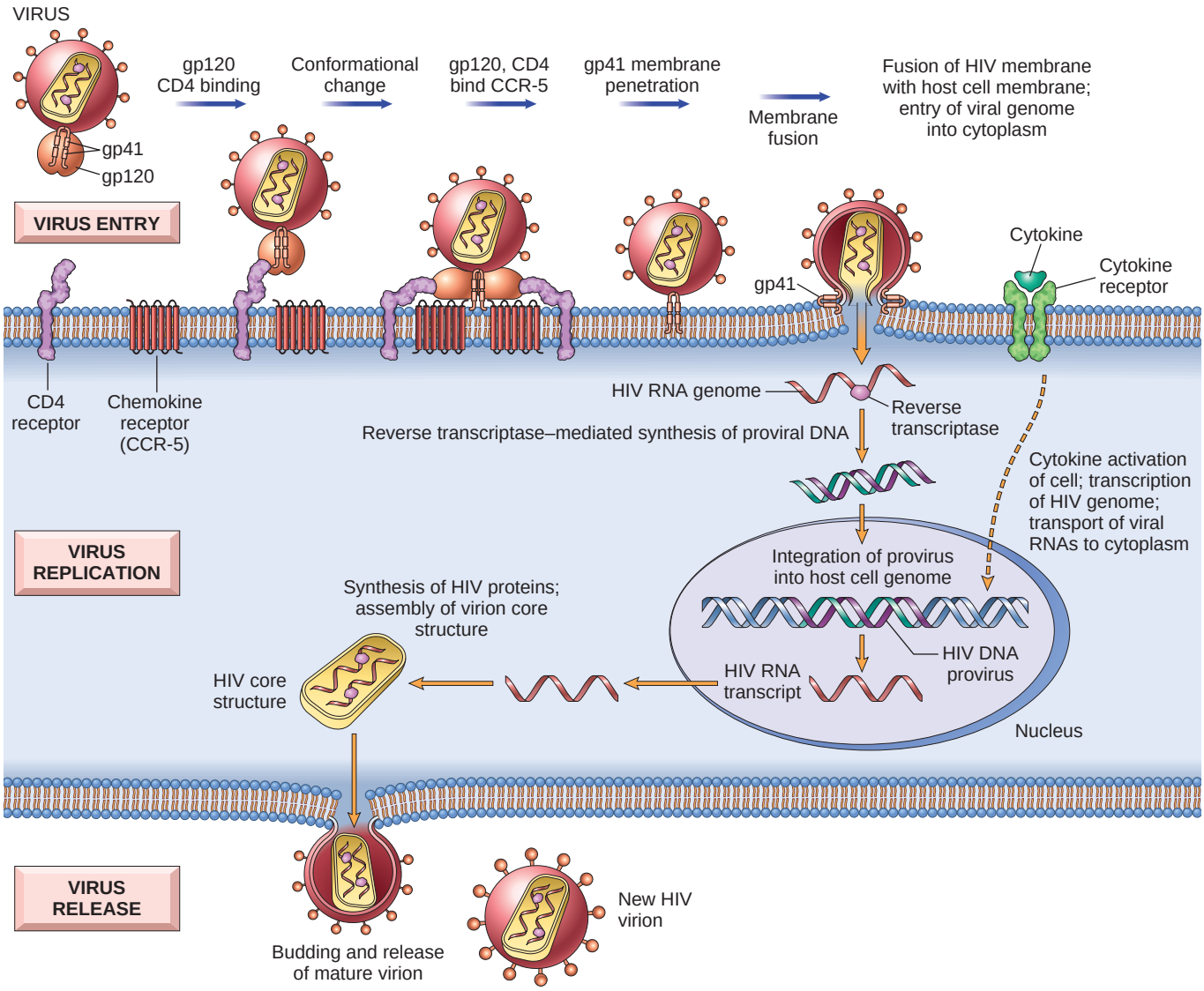


Figure 6.41 The life cycle of HIV showing the steps from viral entry to production of infectious virions. (Modified with permission from Wain-Hobson S: HIV. One on one meets two, *Nature* 384:117, 1996. Copyright 1996, Macmillan Magazines Limited.)

to the coreceptors induces conformational changes in gp41 that result in the exposure of a hydrophobic region called the fusion peptide at the tip of gp41. This peptide inserts into the cell membrane of the target cells (e.g., T cells or macrophages), leading to fusion of the virus and the host cell membrane, an event that allows the virus core containing the HIV genome to enter the cell. The requirement for HIV binding to coreceptors may have important implications for the pathogenesis of AIDS. Chemokines sterically hinder HIV infection of cells in culture by occupying their receptors, and therefore, the level of chemokines in the tissues may influence the efficiency of viral infection in vivo. Also, polymorphisms in the gene encoding CCR5 are associated with different susceptibility to HIV infection. About 1% of white Americans inherit two mutant copies of the *CCR5* gene and are resistant to infection and the development of AIDS associated with R5 HIV isolates. About 20% of individuals are heterozygous for this protective CCR5 allele; these persons are not protected from AIDS, but the onset of their disease after infection is somewhat delayed. Only rare homozygotes for the mutation have been found in African or East Asian populations.

Viral Replication. Once internalized, the RNA genome of the virus undergoes reverse transcription, leading to the synthesis of double-stranded complementary DNA (cDNA; proviral DNA) (see Fig. 6.41). In quiescent T cells, HIV cDNA may remain in the cytoplasm in a linear episomal form. In dividing T cells, the cDNA circularizes, enters the nucleus, and integrates into the host genome. After integration, the provirus may be silent for months or years, a form of latent infection. Alternatively, proviral DNA may be transcribed, with the formation of complete viral particles that bud from the cell membrane. Such productive infection, when associated with extensive viral budding, leads to death of infected cells.

HIV infects memory and activated T cells but is inefficient at productively infecting naïve (resting) T cells.

Naïve T cells contain an active form of an enzyme that introduces mutations in the HIV genome. This enzyme has the cumbersome name APOBEC3G (for apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3G). It is a cytidine deaminase that introduces cytosine-to-uracil mutations in the viral DNA that is produced by reverse transcription. These mutations inhibit further DNA replication by mechanisms that are not fully defined. Activation of T cells converts APOBEC3G into an inactive, high-molecular-mass complex, explaining why the virus can replicate in previously activated (e.g., memory) T cells. HIV has also evolved to counteract this cellular defense mechanism; the viral protein Vif binds to APOBEC3G and promotes its degradation by cellular proteases.

Completion of the viral life cycle in latently infected cells occurs only after cell activation, and in the case of most CD4+ T cells, viral replication and shedding result in cell lysis. Activation of T cells by antigens or cytokines upregulates several transcription factors, including NF-κB, that stimulate transcription of genes encoding cytokines such as IL-2 and its receptor. In resting T cells, NF-κB is held inactive in the cytoplasm in a complex with the IκB (inhibitor of κB) protein. Stimulation of cells by antigen or cytokines activates cytoplasmic kinases that phosphorylate IκB and target it for enzymatic degradation, thus releasing NF-κB and allowing it to translocate to the nucleus. In the nucleus, NF-κB binds to the regulatory sequences of several genes, including those of cytokines that are expressed in activated T cells. The long-terminal-repeat sequences that flank the HIV genome also contain NF-κB–binding sites that can drive the expression of viral RNA. Imagine now a latently infected CD4+ cell that encounters an environmental antigen. Induction of NF-κB in such a cell (a physiologic response) activates the expression of HIV genes (a pathologic outcome) and leads ultimately to the production of virions and to cell lysis. Furthermore, TNF and other cytokines produced by activated macrophages also stimulate NF-κB activity in T cells. Thus, it seems that HIV thrives when the host T cells and macrophages are physiologically activated, an act that can be best described as "subversion from within." Such activation in vivo may result from antigenic stimulation by HIV itself or by other infecting microorganisms. HIV-infected people are at increased risk for other infections, which lead to increased lymphocyte activation and production of proinflammatory cytokines. These, in turn, stimulate more HIV production, loss of additional CD4+ T cells, and more infection. Thus, it is easy to visualize how in individuals with AIDS a vicious cycle is set up that culminates in inexorable destruction of the immune system.

Mechanism of T-Cell Depletion in HIV Infection

Loss of infected CD4+ T cells is mainly because of the direct cytopathic effects of the replicating virus. In infected individuals, approximately 100 billion new viral particles are produced every day, and 1 to 2 billion CD4+ T cells die each day. Because the frequency of infected cells in the circulation is very low, for many years it was suspected that the immunodeficiency is out of proportion to the level of infection and cannot be attributed to death of infected cells. In fact, many infected cells are within tissues (e.g., secondary lymphoid organs and mucosal sites), and death of these cells is a major cause of the relentless, and eventually profound, cell loss. Also, up to a point the immune system can replace the dying T cells, and hence the rate of T cell loss may appear deceptively low, but as the disease progresses, renewal of CD4+ T cells cannot keep up with their loss.

In addition to direct killing of cells by the virus, other mechanisms may contribute to the loss of T cells. These include the following:

- Chronic activation of uninfected cells, responding to HIV itself or to infections that are common in individuals with AIDS, leads to apoptosis of these cells by the process of activation-induced cell death. Thus, the numbers of CD4+ T cells that die may be considerably more than the numbers of infected cells. The molecular mechanism of this type of cell death is not known.
- Noncytopathic (abortive) HIV infection activates the inflammasome pathway and leads to a form of cell death that has been called pyroptosis (Chapter 2). During this process, inflammatory cytokines and cellular contents are released, thus potentiating recruitment of new cells and increasing the numbers of cells that can be infected. This form of cell death may play a role in spread of the infection.
- HIV infects cells in secondary lymphoid organs (spleen, lymph nodes, tonsils) and may cause progressive destruction of the architecture and cellular composition of lymphoid tissues.
- Loss of immature precursors of CD4+ T cells can also occur, either by direct infection of thymic progenitor cells or by infection of accessory cells that secrete cytokines essential for CD4+ T-cell maturation.
- Fusion of infected and uninfected cells with formation of syncytia (giant cells) can occur. In tissue culture, the gp120 expressed on productively infected cells binds to CD4 molecules on uninfected T cells, followed by cell fusion. Fused cells usually die within a few hours. This property of syncytia formation is generally confined to the T-tropic X4 type of HIV-1. For this reason, this type is often referred to as syncytia-inducing (SI) virus, in contrast to the R5 virus.
- Although marked reduction in CD4+ T cells, a hallmark of AIDS, can account for most of the immunodeficiency late in the course of HIV infection, there is evidence of qualitative defects in T cells even in asymptomatic HIV-infected persons. These include reduced antigen-induced T-cell proliferation, decreased Th1-type responses, defects in intracellular signaling, and many more. The loss of Th1 responses results in a profound deficiency of cell-mediated immunity, leading to increased susceptibility to infections by viruses and other intracellular microbes. There is also a selective loss of the memory subset of CD4+ helper T cells early in the course of disease, which explains poor recall responses to previously encountered antigens.

Low-level chronic or latent infection of T cells is an important feature of HIV infection. It is widely believed that integrated provirus, without viral gene expression (latent infection), can remain in the cells for months to years. Even with potent antiviral therapy, which practically sterilizes the peripheral blood, latent virus lurks within the CD4+ cells (both T cells and macrophages) in the lymph nodes. According to some estimates, 0.05% of CD4+ T cells in the

Table 6.14 Major Abnormalities of Immune Function in AIDS

Lymphopenia
Predominantly caused by death of the CD4+ helper T-cell subset
Decreased T-Cell Function in Vivo
Preferential loss of activated and memory T cells Decreased delayed-type hypersensitivity Susceptibility to opportunistic infections Susceptibility to neoplasms
Altered T-Cell Function in Vitro
Decreased proliferative response to mitogens, alloantigens, and soluble antigens Decreased cytotoxicity Decreased helper function for B-cell antibody production Decreased IL-2 and IFN-γ production
Polyclonal B-Cell Activation
Hypergammaglobulinemia and circulating immune complexes Inability to mount de novo antibody response to new antigens Poor responses to normal B-cell activation signals in vitro
Altered Monocyte or Macrophage Functions
Decreased chemotaxis and phagocytosis Decreased class II HLA expression Diminished capacity to present antigen to T cells

HLA, Human leukocyte antigen; *IFN-γ*, interferon-γ; *IL-2*, interleukin-2; *TNF*, tumor necrosis factor.

lymph nodes are latently infected. Because most of these CD4+ T cells are memory cells, they are long-lived, with a life span of months to years, and thus provide a persistent reservoir of virus.

CD4+ T cells play a pivotal role in regulating both cellular and humoral immune responses. Therefore, loss of this "master regulator" has ripple effects on virtually every other component of the immune system, as summarized in Table 6.14.

HIV Infection of Non–T Cells

In addition to infection and loss of CD4+ T cells, **infection of macrophages and DCs is also important in the pathogenesis of HIV infection**. Similar to T cells, the number of macrophages infected by HIV in the tissues greatly exceeds the number of infected blood monocytes. In certain tissues, such as the lungs and brain, as many as 10% to 50% of macrophages are infected. Several aspects of HIV infection of macrophages should be emphasized:

- Although cell division is required for nuclear entry and replication of most retroviruses, HIV-1 can infect and multiply in terminally differentiated nondividing macrophages. This property of HIV-1 is dependent on the viral *vpr* gene. The Vpr protein allows nuclear targeting of the HIV pre-integration complex through the nuclear pore.
- Infected macrophages bud relatively small amounts of virus from the cell surface, but these cells contain large numbers of virus particles, often located in intracellular vesicles. Even though macrophages allow viral replication, they are quite resistant to the cytopathic effects of HIV, in contrast to CD4+ T cells. Thus, in late stages of HIV infection, when CD4+ T-cell numbers decline greatly, macrophages may be an important site of continued viral replication and a reservoir for viral persistence.
- Macrophages may act as portals of infection because in more than 90% of cases acute HIV infection is by M-tropic strains.
- Even uninfected monocytes are reported to have unexplained functional defects that may have important consequences for host defense. These defects include impaired microbicidal activity, decreased chemotaxis, decreased secretion of IL-1, inappropriate secretion of TNF, and poor capacity to present antigens to T cells. Also, infected monocytes may carry HIV from the blood to various parts of the body, including the nervous system, where infiltrating infected monocytes may be a source of the virus that infects resident microglial cells (described later).

Studies have documented that, in addition to macrophages, two types of DCs are also important targets for the initiation and maintenance of HIV infection: mucosal and follicular DCs. It is thought that mucosal DCs are infected by the virus and may transport it to regional lymph nodes, where the virus is transmitted to CD4+ T cells. DCs also express a lectin-like receptor that binds HIV and displays it in an intact, infectious form to T cells, thus promoting infection of the T cells.

FDCs in the germinal centers of lymph nodes are potential reservoirs of HIV. Although some FDCs may be susceptible to HIV infection, most virus particles are found on the surface of their dendritic processes. FDCs have receptors for the Fc portion of immunoglobulins, and hence they trap HIV virions coated with anti-HIV antibodies. The antibody-coated virions localized to FDCs retain the ability to infect CD4+ T cells as they traverse the intricate meshwork formed by the dendritic processes of the FDCs. Infected T follicular helper cells in the germinal centers are also reservoirs of HIV. Since CTLs are largely excluded from germinal centers, these viral reservoirs cannot be readily eliminated by the host immune response.

B-Cell Function in HIV Infection

Although HIV infects T cells, macrophages, and DCs, individuals with AIDS also display profound abnormalities of B-cell function. Paradoxically, early in the disease course there is polyclonal activation of B cells, resulting in germinal center B-cell hyperplasia, sometimes accompanied by autoimmune phenomena like immune thromoboctyopenic purpura. B-cell activation may result from multiple mechanisms: reactivation of or reinfection with EBV, which is a polyclonal B-cell activator; viral gp41, which can promote B-cell growth and differentiation; and increased production of IL-6, which stimulates proliferation of B cells, by HIV-infected macrophages. Plasma cells also increase in number, leading to hypergammaglobulinemia and bone marrow plasmacytosis.

Despite the increased activation of B cells and risk of autoimmunity, patients with AIDS are unable to mount effective antibody responses to newly encountered pathogens. This could be due, in part, to lack of T-cell help, but antibody responses against T-independent antigens are also suppressed, and hence there may be intrinsic defects in B cells as well. Impaired humoral immunity renders these patients

prey to disseminated infections caused by encapsulated bacteria, such as *S. pneumoniae* and *H. influenzae*, both of which require antibodies for effective opsonization and clearance.

Pathogenesis of Central Nervous System Involvement

The pathogenesis of neurologic manifestations deserves special mention because, in addition to the lymphoid system, the nervous system is a target of HIV infection. The clinical syndrome of CNS abnormalities is called HIV-associated neurocognitive disorder (HAND). Microglia, cells in the CNS that belong to the macrophage lineage, are the predominant cell types in the brain that are infected with HIV. It is believed that HIV is carried into the brain by infected T cells or monocytes. The mechanism of HIV-induced damage of the brain remains obscure. Because neurons are not infected by HIV and the extent of neuropathologic changes is less than might be expected from the severity of neurologic symptoms, most workers believe that the neurologic deficit is caused indirectly by viral products and by soluble factors produced by infected microglia. Included among the soluble factors are the usual culprits, such as IL-1, TNF, and IL-6. In addition, nitric oxide induced in neuronal cells by gp41 has been implicated. Direct damage of neurons by soluble HIV gp120 has also been postulated.

Natural History of HIV Infection

The virus typically enters through mucosal epithelia. The subsequent pathologic and clinical manifestations of the infection can be divided into several phases: (1) an acute retroviral syndrome; (2) a middle, chronic phase, in which most individuals are asymptomatic; and (3) clinical AIDS (Figs. 6.42 and 6.43).

Primary Infection, Virus Dissemination, and the Acute Retroviral Syndrome

Acute (early) infection is characterized by infection of memory CD4+ T cells (which express CCR5) in mucosal lymphoid tissues, and death of many infected cells. Because the mucosal tissues are the largest reservoir of memory T cells in the body, and memory T cells are susceptible to HIV infection, this local loss results in considerable depletion of lymphocytes. Few infected cells are detectable in the blood and other tissues. Mucosal infection is often associated with damage to the epithelium, defects in mucosal barrier functions, and translocation of other microbes across the epithelium.

Mucosal infection is followed by dissemination of the virus and the development of host immune responses. DCs in epithelia at sites of virus entry capture the virus and then migrate into the lymph nodes. Once in lymphoid tissues, DCs may pass HIV on to CD4+ T cells through direct cell-to-cell contact. Within days after the first exposure to HIV, viral replication can be detected in the lymph nodes. This replication leads to viremia, during which high numbers of HIV particles are present in the patient's blood. The virus disseminates throughout the body and infects helper T cells, macrophages, and DCs in peripheral lymphoid tissues.

As the HIV infection spreads, the individual mounts antiviral humoral and cell-mediated immune responses. These responses lead to seroconversion (usually within 3 to 7 weeks of presumed exposure) and the development of

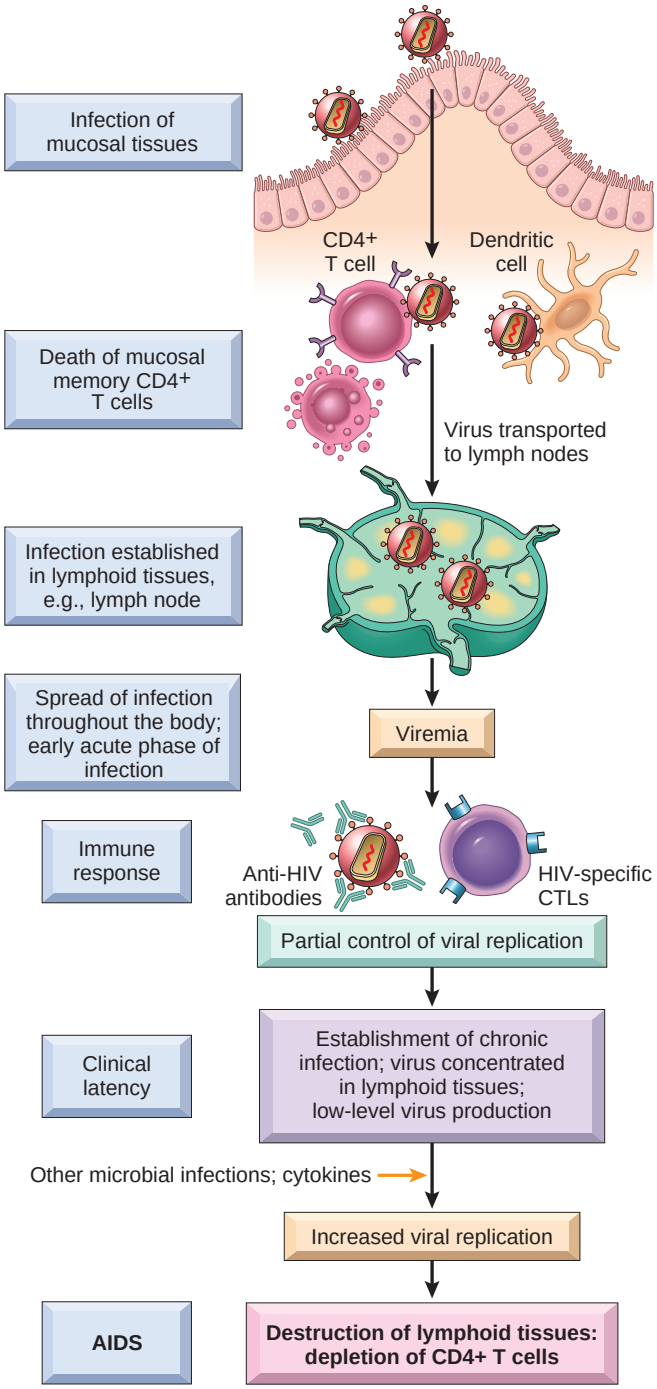


Figure 6.42 Pathogenesis of HIV-1 infection. The initial infection starts in mucosal tissues, involving mainly memory CD4+ T cells and dendritic cells, and spreads to lymph nodes. Viral replication leads to viremia and widespread seeding of lymphoid tissue. This corresponds to the early acute phase of HIV infection. The viremia is controlled by the host immune response, and the patient then enters a phase of clinical latency. During this phase, viral replication in both T cells and macrophages continues unabated, but there is some immune containment of virus *(not illustrated)*. There continues a gradual erosion of CD4+ cells, and ultimately CD4+ T-cell numbers decline, and the patient develops clinical symptoms of full-blown AIDS. *CTL,* Cytotoxic T lymphocyte.

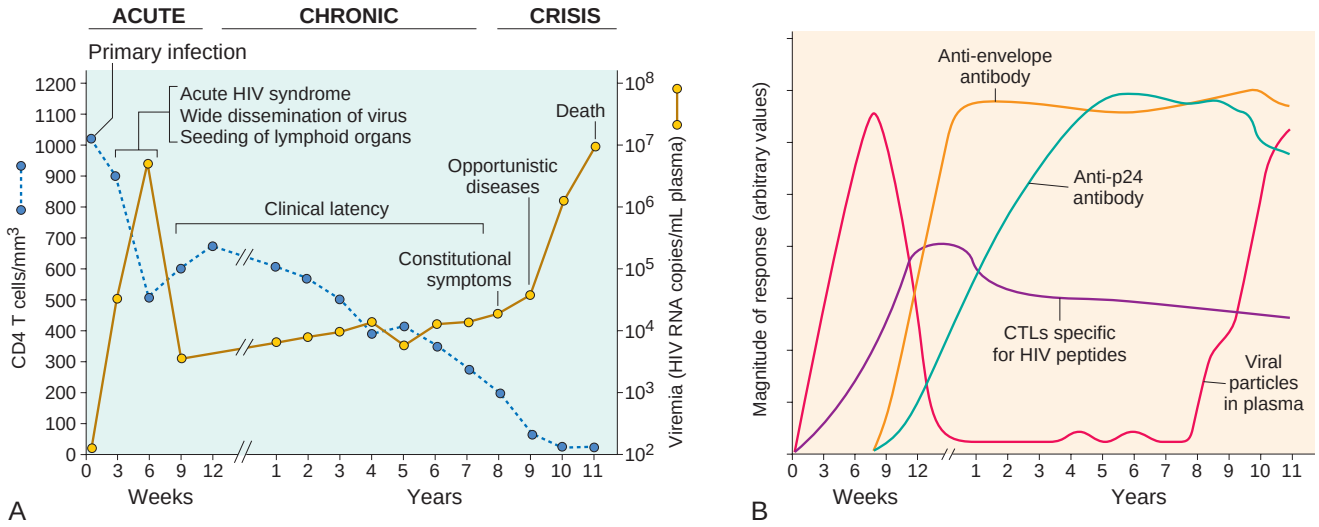


Figure 6.43 Clinical course of HIV infection. (A) During the early period after primary infection, there is dissemination of virus, development of an immune response to HIV, and often an acute viral syndrome. During the period of clinical latency, viral replication continues, and the CD4+ T-cell count gradually decreases until it reaches a critical level below which there is a substantial risk of AIDS-associated diseases. (B) Immune response to HIV infection. A cytotoxic T lymphocyte *(CTL)* response to HIV is detectable by 2 to 3 weeks after the initial infection, and it peaks by 9 to 12 weeks. Marked expansion of virus-specific CD8+ T-cell clones occurs during this time, and up to 10% of a patient's CTLs may be HIV specific at 12 weeks. The humoral immune response to HIV peaks at about 12 weeks. (A, Redrawn from Fauci AS, Lane HC: Human immunodeficiency virus disease: AIDS and related conditions. In Fauci AS, et al., editors: *Harrison's Principles of Internal Medicine*, ed 14, New York, 1997, McGraw-Hill, p 1791.)

virus-specific CD8+ cytotoxic T cells. HIV-specific CD8+ T cells are detected in the blood at about the time viral titers begin to fall and are most likely responsible for the initial containment of HIV infection. These immune responses partially control the infection and viral production, as reflected by a drop in viremia to low but detectable levels by about 12 weeks after the primary exposure.

The *acute retroviral syndrome* is the clinical presentation of the initial spread of the virus and the host response. It is estimated that 40% to 90% of individuals who acquire a primary infection develop this syndrome. This typically occurs 3 to 6 weeks after infection, and resolves spontaneously in 2 to 4 weeks. This phase is marked by a nonspecific self-limited acute illness with flulike symptoms, including sore throat, myalgias, fever, weight loss, and fatigue, sometimes accompanied by rash, cervical adenopathy, diarrhea, and vomiting.

The level of HIV-1 RNA in the blood (viral load) is a useful marker of HIV disease progression and is of clinical value in the management of people with HIV infection. The viral load at the end of the acute phase reflects the equilibrium reached between the virus and the host response, and in a given patient it may remain fairly stable for several years. This level of steady-state viremia, called the viral set point, is a predictor of the rate of decline of CD4+ T cells, and, therefore, progression of HIV disease. In one study, only 8% of patients with a viral load of less than 4350 copies of viral mRNA per microliter of blood progressed to clinical AIDS in 5 years, whereas 62% of those with a viral load of greater than 36,270 copies developed AIDS in the same period.

Because the loss of immune containment is associated with declining CD4+ T-cell counts, the Centers for Disease Control and Prevention (CDC) classification stratifies HIV infection into three categories on the basis of CD4+ cell counts: CD4+ cells greater than or equal to 500 cells/μL, 200 to 499 cells/μL, and fewer than 200 cells/μL (Table 6.15). For clinical management, the blood CD4+ T-cell count is perhaps the most reliable short-term indicator of disease progression. For this reason, the CD4+ cell count (rather than viral load) is the primary clinical measurement used to determine when to start antiretroviral therapy (ART).

Chronic Infection: Phase of Clinical Latency

In the next, chronic phase of the disease, lymph nodes and the spleen are sites of continuous HIV replication and cell destruction (see Fig. 6.42). During this period of the disease, few or no clinical manifestations of the HIV infection are present. Therefore, this phase of HIV disease is called the *clinical latency* period. Although the majority of peripheral blood T cells do not harbor the virus, destruction of CD4+ T cells within lymphoid tissues continues and the number of CD4+ T cells in the blood steadily declines. More than 90% of the body's approximately 10^{11} T cells are normally found in secondary lymphoid organs, and it is estimated that HIV destroys up to 2×10^9 CD4+ T cells every day. Early in the course of the disease, the T cells can be replaced almost as quickly as they are destroyed. At this stage, up to 10% of CD4+ T cells in lymphoid organs may be infected, but the frequency of infected CD4+ T cells in the blood may be less than 0.1% of the total circulating CD4+ T cells. Eventually, over a period of years, the slowly amplifying cycle of virus infection, T-cell death, and new infection leads to a steady decline in the number of CD4+ T cells in the lymphoid tissues and the circulation.

Concomitant with this loss of CD4+ T cells, host defenses wane, and the proportion of the surviving CD4+ cells infected with HIV increases, as does the viral burden per CD4+ cell. Not unexpectedly, HIV RNA levels increase as the host begins to lose the battle with the virus. How HIV escapes

Table 6.15 CDC Classification Categories of HIV Infection

Clinical Categories	CD4+ T-Cell Categories		
	1 ≥500 Cells/μL	2 200–499 Cells/μL	3 <200 Cells/μL
A. Asymptomatic, acute (primary) HIV, or persistent generalized lymphadenopathy	A1	A2	A3
B. Symptomatic, not A or C conditions	B1	B2	B3
C. AIDS indicator conditions: including opportunistic infections, neurologic disease, and tumors			

Data from Centers for Disease Control and Prevention: 1993 revised classification system and expanded surveillance definition for AIDS among adolescents and adults, *MMWR* 41(RR-17):1, 1992.

immune control is not entirely clear, but several mechanisms have been proposed. These include destruction of the CD4+ T cells that are critical for effective immunity, antigenic variation, and down-modulation of class I MHC molecules on infected cells so that viral antigens are not recognized by CD8+ CTLs. During this period, the virus may evolve and switch from relying solely on CCR5 to enter its target cells to relying on either CXCR4 or both CCR5 and CXCR4. This coreceptor switch is associated with more rapid decline in CD4+ T-cell counts, presumably because of greater infection of T cells.

In this chronic phase of infection, patients are either asymptomatic or develop minor opportunistic infections, such as oral candidiasis (thrush), vaginal candidiasis, herpes zoster, and perhaps tuberculosis (the latter being particularly common in resource-poor regions such as sub-Saharan Africa). Autoimmune thrombocytopenia may also be noted (Chapter 14).

AIDS

The final phase is progression to AIDS, characterized by a breakdown of host defense, a dramatic increase in viral load, and severe, life-threatening clinical disease. The typical patient presents with long-lasting fever (>1 month), fatigue, weight loss, diarrhea, and generalized lymph node enlargement. After a variable period, serious opportunistic infections, secondary neoplasms, or clinical neurologic disease (grouped under the rubric AIDS indicator diseases, discussed later) emerge, and the patient is said to have developed AIDS.

In the absence of treatment, most patients with HIV infection progress to AIDS after a chronic phase lasting from 7 to 10 years. Exceptions to this typical course are rapid progressors and long-term nonprogressors. In *rapid progressors,* the middle, chronic phase is telescoped to 2 to 3 years after primary infection. About 5% to 15% of infected individuals are *long-term nonprogressors,* defined as untreated HIV-1–infected individuals who remain asymptomatic for 10 years or more, with stable CD4+ T-cell counts and low viral loads (usually <500 viral RNA copies per milliliter). Remarkably, about 1% of infected individuals have undetectable plasma virus (<50 to 75 RNA copies/mL); these have been called *elite controllers*. Individuals with such an uncommon clinical course have attracted great attention in the hope that studying them may shed light on host and viral factors that influence disease progression. Studies thus far indicate that this group is heterogeneous with respect to the variables that influence the course of the disease. In most cases, the viral isolates do not show qualitative abnormalities, suggesting that the course of the disease cannot be attributed to a "wimpy" virus. In all cases, there is evidence of a vigorous anti-HIV immune response, but the immune correlates of protection are still unknown. Some of these individuals have high levels of HIV-specific CD4+ and CD8+ T-cell responses, and these levels are maintained over the course of infection. The inheritance of particular HLA alleles seems to correlate with resistance to disease progression, perhaps reflecting the ability to mount antiviral T-cell responses. Further studies, it is hoped, will provide the answers to this and other questions critical to understanding disease progression.

Clinical Features of AIDS

The salient clinical features of the acute early and chronic middle phases of HIV infection were described earlier; here we summarize the clinical manifestations of the terminal phase, AIDS. At the outset, it should be pointed out that the clinical manifestations and opportunistic infections associated with HIV infection vary in different parts of the world. Also, the course of the disease has been greatly modified by new antiretroviral therapies, and many devastating complications that were once common are now infrequent.

In the United States, the typical adult patient with AIDS presents with fever, weight loss, diarrhea, generalized lymphadenopathy, multiple opportunistic infections, neurologic disease, and, in many cases, secondary neoplasms. The infections and neoplasms listed in Table 6.16 are included in the surveillance definition of AIDS.

Opportunistic Infections

Opportunistic infections account for the majority of deaths in untreated patients with AIDS. Many of these infections represent reactivation of latent infections, which are normally kept in check by a robust immune system but are not completely eradicated because the infectious agents have evolved to coexist with their hosts. The frequency of infections varies in different regions of the world, and has been markedly reduced by ART, which relies on a combination of three or four drugs that block different steps of the HIV life cycle. A brief summary of selected opportunistic infections is provided here.

- Approximately 15% to 30% of untreated HIV-infected people develop pneumonia caused by the fungus *Pneumocystis jirovecii* (reactivation of a prior latent infection) during the course of the disease. Before the advent of ART, this infection was the presenting feature in about 20% of cases, but the incidence is much less in patients who respond to ART.

Table 6.16 AIDS-Defining Opportunistic Infections and Neoplasms Found in Patients With HIV Infection

Infections
Protozoal and Helminthic Infections
Cryptosporidium or *Cystoisospora* (enteritis)
Pneumocystis (pneumonia or disseminated infection)
Toxoplasma (pneumonia or CNS infection)
Fungal Infections
Candida (esophageal, tracheal, or pulmonary)
Cryptococcus (CNS infection)
Coccidioides (disseminated)
Histoplasma (disseminated)
Bacterial Infections
Mycobacterium ("atypical," e.g., *Mycobacterium avium-intracellulare*, disseminated or extrapulmonary; *Mycobacterium tuberculosis*, pulmonary or extrapulmonary)
Nocardia (pneumonia, meningitis, disseminated)
Salmonella infections, disseminated
Viral Infections
Cytomegalovirus (pulmonary, intestinal, retinitis, or CNS infections)
Herpes simplex virus (localized or disseminated infection)
Varicella-zoster virus (localized or disseminated infection)
Progressive multifocal leukoencephalopathy
Neoplasms
Kaposi sarcoma
Primary lymphoma of brain
Invasive cancer of the uterine cervix

CNS, Central nervous system.

- Many patients present with other opportunistic infections. Among the most common pathogens are *Candida,* cytomegalovirus, atypical and typical mycobacteria, *Cryptococcus neoformans, Toxoplasma gondii, Cryptosporidium,* herpes simplex virus, papovaviruses, and *Histoplasma capsulatum*.
- *Candidiasis* is the most common fungal infection in patients with AIDS, and infection of the oral cavity, vagina, and esophagus are its most common clinical manifestations. In asymptomatic HIV-infected individuals, oral candidiasis is a sign of immunologic decompensation and often heralds the transition to AIDS. Invasive candidiasis is infrequent in patients with AIDS, and it usually occurs when there is drug-induced neutropenia or use of indwelling catheters.
- *Cytomegalovirus* (CMV) may cause disseminated disease or may be localized to the eye and gastrointestinal tract. At one time, CMV chorioretinitis was seen in approximately 25% of patients, but its incidence has decreased dramatically due to ART. Cytomegalovirus retinitis occurs almost exclusively in patients with CD4+ T cell counts less than 50/μL. Gastrointestinal disease, seen in 5% to 10% of cases, manifests as esophagitis and colitis, the latter associated with multiple mucosal ulcerations.
- Disseminated bacterial infection with *atypical mycobacteria* (mainly *Mycobacterium avium-intracellulare*) also occurs late, in the setting of severe immunosuppression. Coincident with the AIDS epidemic, the incidence of tuberculosis has risen dramatically. Worldwide, almost one-third of all deaths in AIDS patients are attributable to tuberculosis, but this complication remains uncommon in the United States. Patients with AIDS are at risk for reactivation of latent pulmonary disease as well as primary infection. In contrast to infection with atypical mycobacteria, *M. tuberculosis* infection manifests early in the course of AIDS. As with tuberculosis in other settings, the infection may be confined to lungs or may involve multiple organs. The extent of infection depends on the degree of immunosuppression; dissemination is more common in patients with very low CD4+ T-cell counts. Most worrisome are reports indicating that a growing number of isolates are resistant to multiple antimycobacterial drugs.
- A variety of opportunistic infections target the CNS. *Cryptococcosis* occurs in about 10% of AIDS patients. As in other settings with immunosuppression, meningitis is the major clinical manifestation of cryptococcosis. *Toxoplasma gondii,* another frequent invader of the CNS in AIDS, causes encephalitis and is responsible for 50% of all mass lesions in the CNS. *JC virus,* a human papovavirus, is another important cause of CNS infections in HIV-infected patients. It causes progressive multifocal leukoencephalopathy (Chapter 28).
- Other infections preferentially involve the gastrointestinal tract and genitalia. *Herpes simplex virus infection* is manifested by mucocutaneous ulcerations involving the mouth, esophagus, external genitalia, and perianal region. *Persistent diarrhea,* which is common in untreated patients with advanced AIDS, is often caused by infections with protozoans such as *Cryptosporidium, Isospora belli,* or microsporidia. These patients have chronic, profuse, watery diarrhea with massive fluid loss. Diarrhea may also result from infection with enteric bacteria, such as *Salmonella* and *Shigella,* as well as *M. avium-intracellulare.*

Tumors

Patients with AIDS have a high incidence of certain tumors, especially Kaposi sarcoma (KS), B-cell lymphoma, cervical cancer in women, and anal cancer in men. It is estimated that 25% to 40% of untreated HIV-infected individuals will eventually develop a malignancy. A common feature of these tumors is that they are caused by oncogenic DNA viruses, specifically human herpesvirus-8 (KS), EBV (B-cell lymphoma), and human papillomavirus (cervical and anal carcinoma). These viruses may establish latent infections even in healthy people, but they are usually kept in check by a competent immune system. The risk of malignancy in AIDS patients is increased mainly because of failure to contain the infections and decreased immunity against the tumors.

Kaposi Sarcoma. KS, a vascular tumor that is otherwise rare in the United States, is the most common neoplasm in patients with AIDS. The morphology of KS and its occurrence in patients not infected with HIV are discussed in Chapter 11. At the onset of the AIDS epidemic, up to 30% of infected homosexual or bisexual men had KS, but in recent years, with use of ART there has been a dramatic decline in its incidence. In contrast, in areas of sub-Saharan Africa where HIV infection is both frequent and largely untreated, KS is one of the most common tumors.

The lesions of KS are characterized by the proliferation of spindle-shaped cells that express markers of both

endothelial cells (vascular or lymphatic) and smooth muscle cells. There is also a profusion of slitlike vascular spaces, suggesting that the lesions may arise from primitive mesenchymal precursors of vascular channels. In addition, KS lesions display chronic inflammatory cell infiltrates. Many of the features of KS suggest that it is not a malignant tumor (despite its ominous name). For instance, spindle cells in many KS lesions are polyclonal or oligoclonal, although more advanced lesions occasionally show monoclonality. The current model of KS pathogenesis is that the spindle cells produce proinflammatory and angiogenic factors, which recruit the inflammatory and neovascular components of the lesion, and the latter components supply signals that aid in spindle cell survival and growth.

There is compelling evidence that KS is caused by *human herpesvirus 8* (HHV8), also called KS herpesvirus (KSHV). Exactly how HHV8 infection leads to KS is still unclear. Like other herpesviruses, HHV8 establishes latent infection, during which several proteins are produced with potential roles in stimulating spindle cell proliferation and preventing apoptosis. These include a viral homologue of cyclin D and several inhibitors of p53. However, HHV8 infection, although necessary for KS development, is not sufficient, and additional cofactors are needed. In the AIDS-related form, that cofactor is clearly HIV (the relevant cofactors for HIV-negative KS remain unknown). HIV-mediated immune suppression may aid in widespread dissemination of HHV8 in the host.

HHV8 infection is not restricted to endothelial cells. The virus is related phylogenetically to the lymphotropic subfamily of herpesviruses (γ-herpesvirus); in keeping with this, its genome is found in B cells of infected subjects. In fact, HHV8 infection is also linked to rare B-cell lymphomas in AIDS patients (called primary effusion lymphoma) and to multicentric Castleman disease, a B-cell lymphoproliferative disorder.

Clinically, AIDS-associated KS is quite different from the sporadic form (Chapter 11). In HIV-infected individuals, the tumor is usually widespread, affecting the skin, mucous membranes, gastrointestinal tract, lymph nodes, and lungs. These tumors also tend to be more aggressive than classic KS.

Lymphomas. Lymphoma occurs at a markedly increased rate in individuals with AIDS, making it one of several AIDS-defining conditions. Roughly 5% of AIDS patients present with lymphoma, and approximately another 5% develop lymphoma during their subsequent course, virtually all of which originate from transformed B cells. With the advent of effective ART, the incidence of lymphoma has fallen substantially in some HIV-infected populations. However, lymphoma continues to occur in HIV-infected people at an incidence that is at least 10-fold greater than the population average. These epidemiologic findings suggest that the association of lymphoma and HIV infection is only partially explained by T-cell immunodeficiency. Indeed, based on molecular characterization of HIV-associated lymphomas and these epidemiologic considerations, at least two mechanisms appear to underlie the increased risk of B-cell tumors in HIV-infected individuals (Fig. 6.44).

- Unchecked proliferation of B cells infected with oncogenic herpesviruses in the setting of profound T cell depletion (AIDS). T-cell immunity is required to restrain the proliferation of B cells infected with oncogenic viruses such as EBV and HHV8. With the appearance of severe T-cell depletion late in the course of HIV infection, this control is lost. As a result, AIDS patients are at high risk of developing aggressive B cell lymphomas composed of tumor cells infected by oncogenic viruses, particularly EBV.

  By adulthood, most normal individuals are infected with EBV. Once immunity is established, EBV persists in such individuals as a latent infection in approximately 1 in 100,000 B cells, most of which have a memory B-cell phenotype. Activation of such cells, by antigen or by cytokines, reawakens an EBV-encoded program of gene expression that drives B-cell proliferation. Patients with AIDS have high levels of several cytokines, some of which, including IL-6, are growth factors for B cells. These patients are also chronically infected with pathogens that may lead to B-cell stimulation. In the absence of T-cell immunity, these activated, EBV-infected clones proliferate and eventually acquire additional somatic mutations, leading to their outgrowth as full-blown EBV-positive B-cell lymphomas. The tumors often occur in extranodal

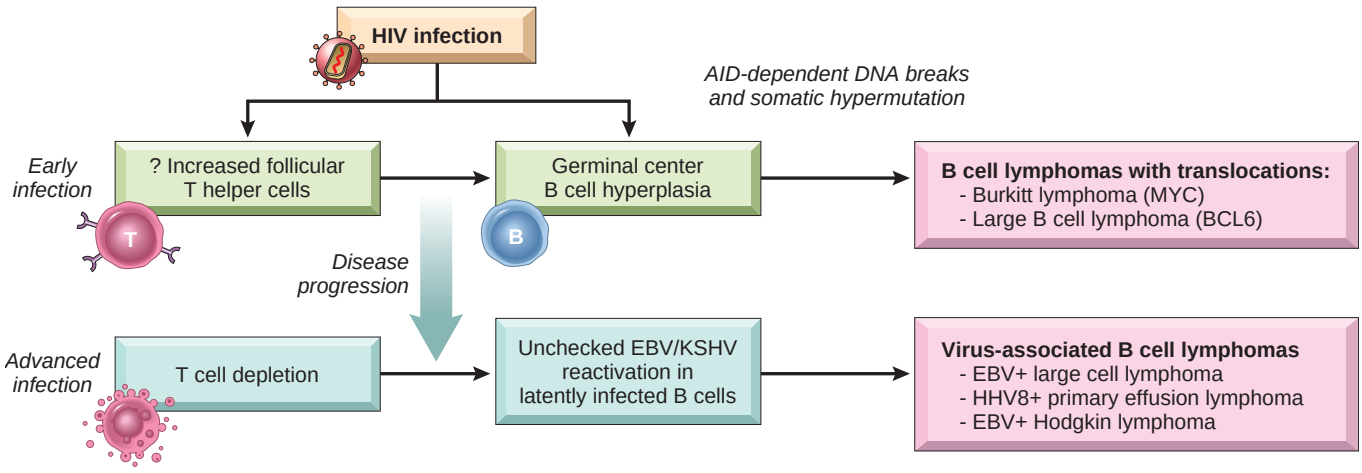


Figure 6.44 A model for the pathogenesis of B-cell lymphomas in HIV infection. HIV infection results in several changes that may cooperate to produce B-cell lymphomas. An increase in T follicular helper cells early in the course of the disease is postulated, but its basis is unknown.

sites, such as the CNS, but also the gut, the orbit, the lungs, and elsewhere. The primary effusion lymphomas mentioned earlier present as malignant effusions and are remarkable in that the tumor cells are typically coinfected by both EBV and HHV8, an unusual example of cooperativity between two oncogenic viruses.

- Germinal center B-cell hyperplasia in the setting of early HIV infection. As mentioned, even in the face of effective ART, the overall rate of lymphoma in the HIV-infected population remains elevated, even in those with normal CD4+ T-cell counts. The majority of the lymphomas that arise in patients with preserved CD4 T-cell counts are not associated with EBV or HHV8. What then explains the increased risk of lymphoma? The answer is not known, but it may be related to the profound germinal center B-cell hyperplasia that occurs early in HIV infection. Recall that in germinal centers, B cells undergo class switching and somatic hypermutation in their immunoglobulin genes. Both processes introduce breaks in DNA and are error-prone, sometimes leading to translocation of oncogenes. Of note, B cell tumors that arise outside of the setting of full-blown AIDS in HIV-infected individuals, such as Burkitt lymphoma and diffuse large B-cell lymphoma, are often associated with translocations of oncogenes into Ig gene loci. Thus, the striking germinal center B-cell hyperplasia that occurs early in HIV infection may contribute to lymphomagenesis in part by increasing the number of B cells that are at-risk for acquiring lymphoma-initiating events.

Several other EBV-related proliferations also merit mention. Hodgkin lymphoma, an unusual B cell tumor associated with a pronounced tissue inflammatory response (Chapter 13), occurs at increased frequency in HIV-infected individuals. In virtually all instances of HIV-associated Hodgkin lymphoma, the characteristic tumor cells (Reed-Sternberg cells) are infected with EBV. Many (but not all) HIV patients with Hodgkin lymphoma have low CD4 counts at the time of disease presentation. EBV infection also is responsible for oral hairy leukoplakia (white projections on the tongue), which results from EBV-driven squamous cell proliferation of the oral mucosa (Chapter 16).

Other Tumors. In addition to KS and lymphoma, patients with AIDS are at increased risk of human papilloma virus (HPV)-associated carcinomas of the uterine cervix and the anus. HPV is intimately associated with squamous cell carcinoma of the cervix and its precursor, squamous intraepithelial lesion (SIL; Chapters 7 and 22). HPV-associated SIL is 10 times more common in HIV-infected women as compared with uninfected women attending family planning clinics, and is confined to HIV-infected women with CD4 counts of less than 500 cells/μL, suggesting that the risk is attributable to diminished immune surveillance. The rate of progression from SIL to overt cervical carcinoma also appears to be accelerated in HIV-infected women. Hence, gynecologic examination is a particularly important part of a routine workup for HIV-infected women.

Central Nervous System Disease

Involvement of the CNS is a common and important manifestation of AIDS. Ninety percent of patients demonstrate some form of neurologic involvement at autopsy, and 40% to 60% have clinically apparent neurologic dysfunction. Importantly, in some patients, neurologic manifestations may be the sole or earliest presenting feature of HIV infection. In addition to opportunistic infections and neoplasms, several HIV-related neuropathologic changes occur. These include a self-limited meningoencephalitis occurring at the time of seroconversion, aseptic meningitis, vacuolar myelopathy, peripheral neuropathies, and, most commonly, a progressive encephalopathy designated clinically as HIV-associated neurocognitive disorder (Chapter 28). It is believed to result from a combination of HIV infection of microglia and an immune response in the CNS.

Effect of Antiretroviral Drug Therapy

Fortunately for those affected, ART with drugs that target the viral reverse transcriptase, protease, and integrase have altered the clinical course of HIV infection quite dramatically. These drugs are given in combination to reduce the emergence of drug-resistant mutants. Over 25 antiretroviral drugs from six distinct drug classes have been developed for the management of HIV infection. When a combination of at least three effective drugs is used in a motivated, compliant patient, HIV viral load is reduced to below the level of detection (<50 copies RNA/mL) and remains there indefinitely (as long as the patient adheres to therapy). Even when a drug-resistant virus breaks through, there are several second- and third-line options to combat the virus. Once the virus is suppressed, the progressive loss of CD4+ T cells is halted, and over a period of several years the peripheral blood CD4+ T-cell count slowly increases and often returns to normal. With the use of ART, in the United States the annual death rate from AIDS has decreased from its peak of 16 to 18 per 100,000 people in 1995 to 1996 to less than 4 per 100,000. Many AIDS-associated disorders, such as *P. jirovecii* infection and KS, are very uncommon now. ART also has reduced the transmission of the virus, especially from infected mothers to newborns. However, because of the reduced mortality, more people are living with HIV, and the risk of HIV spread in the population will inevitably increase if vigilance is relaxed. Indeed, there is compelling evidence that even patients who have virtually undetectable viral loads for years on ART develop active infection if they stop drug treatment.

Despite these dramatic improvements, several new complications associated with HIV infection and its treatment have emerged. Some patients with advanced disease who are given ART have a paradoxical deterioration in their clinical condition as the immune system recovers. This occurs despite increasing CD4+ T-cell counts and decreasing viral load. This disorder, called *immune reconstitution inflammatory syndrome,* is not understood but is postulated to be a poorly regulated host response to the high antigenic burden of persistent microbes. Perhaps a more important complication of long-term ART pertains to adverse side effects of the drugs. These include lipoatrophy (loss of facial fat), lipoaccumulation (excess fat deposition centrally), elevated lipids, insulin resistance, peripheral neuropathy, and potentially deleterious effects on cardiovascular, renal, and hepatic function. Because CD4+ T-cell counts are normalized, non-AIDS morbidity is far more common than classic AIDS-related morbidity in long-term ART-treated patients. Major causes of morbidity are cancer, and accelerated cardiovascular, kidney, and

liver disease. The mechanism for these non–AIDS-related complications is not known, but persistent inflammation and immune dysfunction may play a role.

MORPHOLOGY

The anatomic changes in the tissues are neither specific nor diagnostic. Common pathologic features of AIDS include opportunistic infections, KS, and B-cell lymphomas. Most of these lesions are discussed elsewhere, because they also occur in individuals who do not have HIV infection. Lesions in the CNS are described in Chapter 28.

Biopsy specimens from enlarged lymph nodes in the early stages of HIV infection reveal a marked hyperplasia of B-cell follicles. The follicles are enlarged and often take on unusual, serpiginous shapes. The mantle zones that surround the follicles are attenuated, and the germinal centers impinge on interfollicular T-cell areas. This hyperplasia of B cells is the morphologic reflection of the polyclonal B-cell activation that is seen in HIV-infected individuals.

With disease progression, the frenzy of B-cell proliferation subsides and gives way to a pattern of severe lymphoid involution. The lymph nodes are depleted of lymphocytes, and the organized network of FDCs is disrupted. The germinal centers may even become hyalinized. These small, atrophic, "burnt-out" lymph nodes may harbor numerous opportunistic pathogens, often within macrophages. Because of profound immunosuppression, inflammatory responses to infections, both in the lymph nodes and at extranodal sites, may be sparse or atypical. For example, mycobacteria may not evoke granuloma formation because CD4+ cells are deficient. In the empty-looking lymph nodes and in other organs, the presence of infectious agents may not be readily apparent without special stains. As might be expected, lymphoid involution is not confined to the nodes; in later stages of AIDS, the spleen and thymus also are converted to "wastelands" that are virtually devoid of lymphocytes.

Despite spectacular advances in our understanding of HIV infection, the long-term prognosis of HIV-positive patients remains guarded. With ART, the mortality rate has declined in the United States, but treated patients still carry viral DNA in their lymphoid tissues, and a cure remains elusive. Although a considerable effort has been mounted to develop a vaccine, many hurdles remain to be crossed before vaccine-based prophylaxis becomes a reality. Molecular analyses have revealed an alarming degree of variation in viral isolates from patients; this renders the task of producing a vaccine extremely difficult. Recent efforts have focused on producing broadly neutralizing antibodies against relatively invariant portions of HIV proteins. The task of developing an effective vaccine is complicated by the fact that the correlates of immune protection are not fully understood. At present, therefore, prevention, public health measures, and antiretroviral drugs remain the mainstays in the fight against AIDS.

KEY CONCEPTS

PATHOGENESIS AND COURSE OF HIV INFECTION AND AIDS

- Virus entry into cells: requires CD4 and co-receptors, which are receptors for chemokines; involves binding of viral gp120 and fusion with the cell mediated by viral gp41 protein; main cellular targets are CD4+ helper T cells, macrophages, and DCs
- Viral replication: provirus genome integrates into host cell DNA; viral gene expression is triggered by stimuli that activate infected cells (e.g., infectious microbes, cytokines produced during normal immune responses)
- Progression of infection: acute infection of mucosal T cells and DCs; viremia with dissemination of virus; latent infection of cells in lymphoid tissue; continuing viral replication and progressive loss of CD4+ T cells
- Mechanisms of immune deficiency
  - Loss of CD4+ T cells: T-cell death during viral replication and budding (similar to other cytopathic infections); apoptosis as a result of chronic stimulation; decreased thymic output; functional defects
  - Defective macrophage and DC functions
  - Destruction of architecture of lymphoid tissues (late)
- Clinical manifestations of AIDS include opportunistic infections, tumors such as B-cell lymphomas, and CNS abnormalities.

AMYLOIDOSIS

Amyloidosis is a condition associated with a number of inherited and inflammatory disorders in which extracellular deposits of fibrillar proteins are responsible for tissue damage and functional compromise. These abnormal fibrils are produced by the aggregation of misfolded proteins (which are soluble in their normal folded configuration). The fibrillar deposits bind a wide variety of proteoglycans and glycosaminoglycans, including heparan sulfate and dermatan sulfate, and plasma proteins, notably serum amyloid P. The presence of abundant charged sugar groups in these adsorbed proteins give the deposits staining characteristics that were thought to resemble starch (amylose). Therefore, the deposits were called *amyloid,* a name that is firmly entrenched despite the realization that the deposits are unrelated to starch.

Amyloid is deposited in the extracellular space in various tissues and organs in several clinical settings. With progressive accumulation, it encroaches on and produces pressure atrophy of adjacent cells. Because amyloid deposition appears insidiously and sometimes mysteriously, its clinical recognition ultimately depends on morphologic identification of this distinctive substance in appropriate biopsy specimens. With the light microscope and hematoxylin and eosin stains, amyloid appears as an amorphous, eosinophilic, hyaline, extracellular substance. To differentiate amyloid from other hyaline materials (e.g., collagen, fibrin), a variety of histochemical techniques, described later, are employed. Perhaps most widely used is the Congo red stain, which under ordinary light imparts a pink or red color to tissue deposits, but far more striking and specific is the green birefringence of the stained amyloid when observed by polarizing microscopy (see later).

Properties of Amyloid Proteins

Even though all amyloid deposits have a similar microscopic appearance, amyloid is not a single chemical entity. In fact, more than 20 different proteins can aggregate to form amyloid. There are three major and several minor biochemical

forms, which are deposited by different pathogenic mechanisms. Therefore, amyloidosis is not a single disease, but rather a group of diseases having in common the deposition of similar-appearing proteins. At the heart of the morphologic similarity is the remarkably uniform physical organization of amyloid, which we consider first.

Physical Nature of Amyloid. By electron microscopy, all types of amyloid, regardless of clinical setting or chemical composition, consist of continuous, nonbranching fibrils with a diameter of approximately 7.5 to 10 nm. X-ray crystallography and infrared spectroscopy demonstrate a characteristic cross-β-pleated sheet conformation (Fig. 6.45). This conformation is responsible for the distinctive Congo red staining and birefringence of amyloid.

Chemical Nature of Amyloid. Approximately 95% of the amyloid consists of fibril proteins, with the remaining 5% being the P component and other glycoproteins. The three most common forms of amyloid are as follows:

- *Amyloid light chain (AL) protein* is made up of complete immunoglobulin light chains, the amino-terminal fragments of light chains, or both. Most of the AL proteins analyzed are composed of λ light chains or their fragments, but κ chains are present in some cases. The amyloid fibril protein of the AL type is produced from free Ig light chains secreted by a monoclonal population of plasma cells, and its deposition is associated with certain plasma cell tumors (Chapter 13).
- *Amyloid-associated (AA) protein* is a form of amyloid derived from a unique non-Ig protein made by the liver. The AA protein found in the fibrils is created by proteolysis of a larger precursor called SAA (serum amyloid-associated) protein that is synthesized in the liver and circulates in the blood bound to high-density lipoproteins. The production of SAA protein is increased in inflammatory states as part of the acute phase response; therefore, this form of amyloidosis is associated with chronic inflammation, and is often called secondary amyloidosis.
- *β-amyloid (Aβ) protein* constitutes the core of cerebral plaques found in Alzheimer disease as well as the amyloid deposited in walls of cerebral blood vessels in individuals with this disease. The Aβ protein is derived by proteolysis from a much larger transmembrane glycoprotein called amyloid precursor protein. This form of amyloid is discussed in Chapter 28.

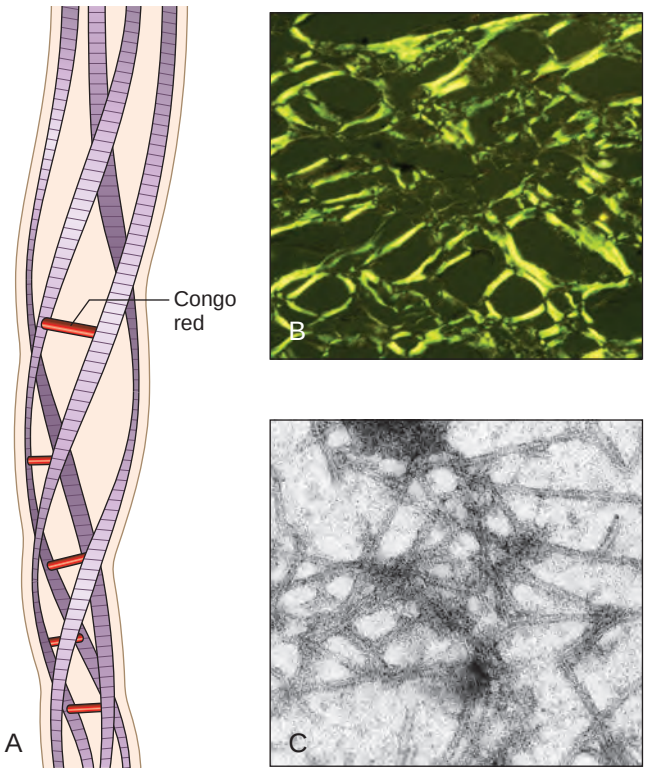


Figure 6.45 Structure of amyloid. (A) A schematic diagram of an amyloid fiber showing four fibrils (there can be as many as six in each fiber) wound around one another with regularly spaced binding of the Congo red dye. (B) Congo red staining shows apple-green birefringence under polarized light, a diagnostic feature of amyloid. (C) Electron micrograph of 7.5- to 10-nm amyloid fibrils. (From Merlini G, Bellotti V: Molecular mechanisms of amyloidosis, *N Engl J Med* 349:583–596, 2003, with permission of the Massachusetts Medical Society.)

As mentioned, multiple other biochemically distinct proteins can also deposit as amyloid in a variety of clinical settings. Among these rare forms of amyloid, the proteins most often involved are as follows:

- *Transthyretin* (TTR) is a normal serum protein that binds and transports thyroxine and retinol. Several distinct mutant forms of TTR (and its fragments) form amyloid in a group of genetically determined disorders referred to as *familial amyloid polyneuropathies.* Unmutated TTR may also deposit as amyloid in the heart of aged individuals (senile systemic amyloidosis).
- *β_2-microglobulin,* a component of MHC class I molecules and a normal serum protein, has been identified as the major component of a form of amyloid ($A\beta_2m$) that deposits in or around the joints or soft tissues of patients on long-term hemodialysis.
- In a minority of cases of prion disease in the CNS, the misfolded prion proteins aggregate in the extracellular space and acquire the structural and staining characteristics of amyloid.

Pathogenesis and Classification of Amyloidosis

Amyloidosis results from abnormal folding of proteins, which become insoluble, aggregate, and deposit as fibrils in extracellular tissues. Normally, misfolded proteins are degraded intracellularly in proteasomes, or extracellularly by macrophages. It appears that in amyloidosis, these quality-control mechanisms fail, leading to accumulation of a misfolded protein outside cells. The proteins that form amyloid fall into two general categories (Fig. 6.46): (1) normal proteins that have an inherent tendency to fold improperly, associate and form fibrils, and do so when they are produced in increased amounts and (2) mutant proteins that are prone to misfolding and subsequent aggregation. The mechanisms of deposition of different types of amyloid are discussed here along with classification.

Because a given biochemical form of amyloid (e.g., AA) may be associated with amyloid deposition in diverse clinical settings, we follow a combined biochemical-clinical classification for our discussion (Table 6.17). Amyloid may be *systemic* (generalized), involving several organ systems, or it may

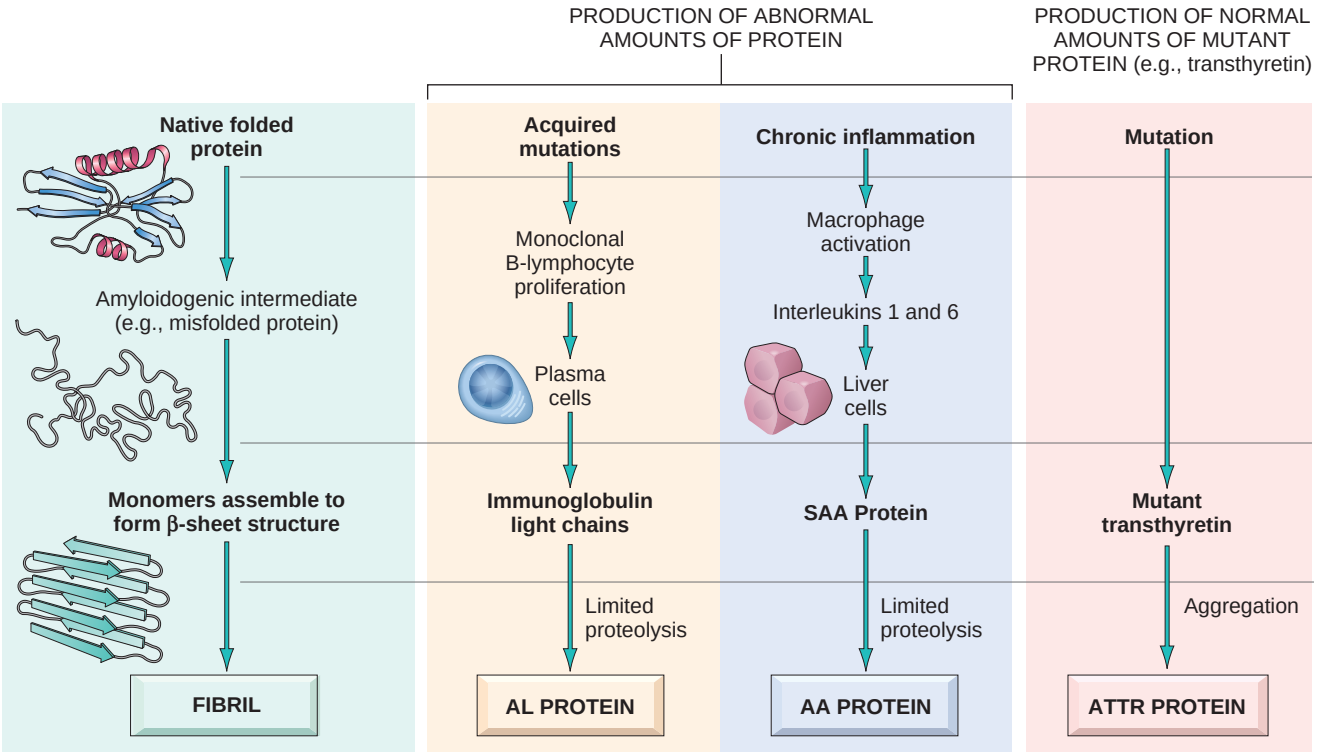


Figure 6.46 Pathogenesis of amyloidosis, showing the proposed mechanisms underlying deposition of the major forms of amyloid fibrils.

be *localized* to a single organ, such as the heart. On clinical grounds, the systemic, or generalized, pattern is subclassified into *primary amyloidosis,* when it is associated with clonal plasma cell proliferations, or *secondary amyloidosis,* when it occurs as a complication of an underlying chronic inflammatory process. *Hereditary* or *familial amyloidosis* constitutes a separate, albeit heterogeneous, group, with several distinctive patterns of organ involvement.

Primary Amyloidosis: Plasma Cell Disorders Associated With Amyloidosis

Amyloid in this category is usually systemic in distribution and is of the AL type. With approximately 2000 to 3000 new cases every year in the United States, this is the most common form of amyloidosis. In all cases, the disorder is caused by a clonal proliferation of plasma cells that

Table 6.17 Classification of Amyloidosis

Clinicopathologic Category	Associated Diseases	Major Fibril Protein	Chemically Related Precursor Protein
Systemic (Generalized) Amyloidosis			
Immunoglobulin light chain amyloidosis (primary amyloidosis)	Multiple myeloma and other monoclonal plasma cell proliferations	AL	Immunoglobulin light chains, chiefly λ type
Reactive systemic amyloidosis (secondary amyloidosis)	Chronic inflammatory conditions	AA	SAA
Hemodialysis-associated amyloidosis	Chronic renal failure	$A\beta_2m$	β_2-microglobulin
Hereditary Amyloidosis			
Familial Mediterranean fever		AA	SAA
Familial amyloidotic neuropathies (several types)		ATTR	Transthyretin
Systemic senile amyloidosis		ATTR	Transthyretin
Localized Amyloidosis			
Senile cerebral	Alzheimer disease	Aβ	APP
Endocrine Medullary carcinoma of thyroid Islets of Langerhans	Type 2 diabetes	 A Cal AIAPP	 Calcitonin Islet amyloid peptide
Isolated atrial amyloidosis		AANF	Atrial natriuretic factor

synthesize an Ig light chain that is prone to form amyloid due to its intrinsic physiochemical properties. Best defined is the occurrence of systemic amyloidosis in 5% to 15% of individuals with multiple myeloma, a plasma-cell tumor characterized by multiple osteolytic lesions throughout the skeletal system (Chapter 13). The malignant plasma cells synthesize abnormal amounts of a single Ig, producing an M (myeloma) protein "spike" on serum electrophoresis. In addition to the synthesis of whole Ig molecules, the malignant plasma cells often secrete free, unpaired κ or λ light chains (referred to as Bence-Jones protein). These are often found in the serum, and due to their small molecular size, Bence-Jones proteins are excreted and concentrated in the urine. In primary amyloidosis, the free light chains are not only present in serum and urine but are also deposited in tissues as amyloid. It should be noted, however, that the majority of myeloma patients who have free light chains in serum and urine do not develop amyloidosis. Clearly, not all free light chains are equally likely to produce amyloid, and it is believed that the amyloidogenic potential of any particular light chain is largely determined by its specific amino acid sequence.

Most persons with AL amyloid do not have classic multiple myeloma or any other overt B-cell neoplasm; such cases have been traditionally classified as primary amyloidosis, because their clinical features derive from the effects of amyloid deposition without any other associated disease. In virtually all such cases, however, monoclonal immunoglobulins or free light chains, or both, can be found in the serum or urine. Most of these patients also have a modest increase in the number of plasma cells in the bone marrow, which presumably secrete the precursors of AL protein. Thus, these patients have an underlying monoclonal proliferation of Ig-producing cells *(monoclonal gammopathy)* in which production of an abnormal protein, rather than presence of tumor masses, is the dominant manifestation.

Reactive Systemic Amyloidosis

The amyloid deposits in this pattern are systemic in distribution and are composed of AA protein. This category was previously referred to as *secondary amyloidosis* because it is secondary to an associated inflammatory condition. At one time, tuberculosis, bronchiectasis, and chronic osteomyelitis were the most important underlying conditions, but with the advent of effective antibiotic therapy, the importance of these conditions has diminished. More commonly now, reactive systemic amyloidosis complicates rheumatoid arthritis, other connective tissue disorders such as ankylosing spondylitis, and inflammatory bowel disease, both Crohn disease and ulcerative colitis. Among these, the most frequently associated condition is rheumatoid arthritis. Amyloidosis is reported to occur in approximately 3% of patients with rheumatoid arthritis and is clinically significant in one-half of those affected. Heroin users who inject the drug subcutaneously also have a high occurrence rate of generalized AA amyloidosis. The chronic skin infections associated with "skin-popping" of narcotics seem to be responsible for the amyloidosis. Reactive systemic amyloidosis may also occur in association with solid tumors, the most common being renal cell carcinoma and Hodgkin lymphoma.

In this form of amyloidosis, SAA synthesis by liver cells is stimulated by cytokines such as IL-6 and IL-1 that are produced during inflammation; thus, long-standing inflammation leads to a sustained elevation of SAA levels. However, increased production of SAA by itself is not sufficient for the deposition of amyloid. There are two possible explanations for this. According to one view, SAA is normally degraded to soluble end products by the action of monocyte-derived enzymes. Conceivably, individuals who develop amyloidosis have an enzyme defect that results in incomplete breakdown of SAA, thus generating insoluble AA molecules. Alternatively, a genetically determined structural abnormality in the SAA molecule may render it resistant to degradation by macrophages.

Heredofamilial Amyloidosis

A variety of familial forms of amyloidosis have been described. Most of them are rare and occur in limited geographic areas. The most common and best studied is an autosomal recessive condition called *familial Mediterranean fever*. This is an autoinflammatory syndrome associated with excessive production of the cytokine IL-1 in response to inflammatory stimuli. It is characterized clinically by attacks of fever accompanied by inflammation of serosal surfaces, including peritoneum, pleura, and synovial membrane. The gene for familial Mediterranean fever encodes a protein called pyrin (for its relation to fever), which is one of a complex of proteins that regulate inflammatory reactions via activation of the inflammasome (Chapter 3). This disorder is encountered largely in individuals of Armenian, Sephardic Jewish, and Arabic origins. It is sometimes associated with widespread amyloidosis. The amyloid fibril proteins are made up of AA proteins, reflecting the increased production of SAA in response to excessive IL-1 secretion.

In contrast to familial Mediterranean fever, a group of autosomal dominant familial disorders is characterized by deposition of amyloid predominantly in peripheral and autonomic nerves. These familial amyloidotic polyneuropathies have been described in different parts of the world. As mentioned earlier, in all of these genetic disorders, the fibrils are made up of mutant TTR. In these disorders, TTR is deposited as amyloid fibrils because genetically determined alterations of its structure render it prone to misfolding and aggregation, and resistant to proteolysis.

Hemodialysis-Associated Amyloidosis

Patients on long-term hemodialysis for renal failure can develop amyloidosis as a result of deposition of β_2-microglobulin. This protein is present in high concentrations in the serum of persons with renal disease, and in the past it was retained in the circulation because it could not be filtered through dialysis membranes. Patients usually present with symptoms related to β_2-microglobulin deposition in joints, muscle, tendons, or ligaments; one relatively common presentation is as carpal tunnel syndrome. With new dialysis filters, the incidence of this complication has decreased substantially.

Localized Amyloidosis

Sometimes, amyloid deposits are limited to a single organ or tissue without involvement of any other site in the body. The deposits may produce grossly detectable nodular masses or be evident only on microscopic examination. Nodular deposits of amyloid are most often encountered in the lung,

larynx, skin, urinary bladder, tongue, and the region around the eye. Frequently, there are infiltrates of lymphocytes and plasma cells in the periphery of these amyloid masses. At least in some cases, the amyloid consists of AL protein and may therefore represent a localized form of plasma cell–derived amyloid.

Endocrine Amyloid

Microscopic deposits of localized amyloid may be found in certain endocrine tumors, such as medullary thyroid carcinoma, pancreatic islet tumor, and pheochromocytoma, and in the islets of Langerhans in individuals with type 2 diabetes mellitus. In these settings, the amyloidogenic proteins seem to be derived either from polypeptide hormones (e.g., medullary carcinoma) or from unique proteins (e.g., islet amyloid polypeptide). In medullary carcinoma of the thyroid, the presence of amyloid is a helpful diagnostic feature.

Amyloid of Aging

Several well-documented forms of amyloid deposition occur with aging. Senile systemic amyloidosis refers to the systemic deposition of amyloid in elderly patients (usually in their seventies and eighties). Because of the dominant involvement and related dysfunction of the heart, this form was previously called senile cardiac amyloidosis. Those who are symptomatic present with a restrictive cardiomyopathy and arrhythmias (Chapter 12). The amyloid in this form is derived from normal TTR. In addition to the sporadic senile systemic amyloidosis, another form affecting predominantly the heart that results from the deposition of a mutant form of TTR has also been recognized. Approximately 4% of people of African descent in the United States express this mutant form of TTR, and cardiomyopathy has been identified in both homozygous and heterozygous patients. The precise prevalence of patients with this mutation who develop clinically manifest cardiac disease is not known.

MORPHOLOGY

There are no consistent or distinctive patterns of organ or tissue distribution of amyloid deposits in any of the categories cited. Nonetheless, a few generalizations can be made. In amyloidosis secondary to chronic inflammatory disorders, the kidneys, liver, spleen, lymph nodes, adrenals, and thyroid, as well as many other tissues, are typically affected. Although amyloidosis associated with plasma cell proliferations cannot reliably be distinguished from the secondary form by its organ distribution, it more often involves the heart, gastrointestinal tract, respiratory tract, peripheral nerves, skin, and tongue. The localization of amyloid deposits in the hereditary syndromes is varied. In familial Mediterranean fever, the amyloidosis may be widespread, involving the kidneys, blood vessels, spleen, respiratory tract, and (rarely) liver. The localization of amyloid in the remaining hereditary syndromes can be inferred from the designation of these entities.

Whatever the clinical disorder, the amyloid may or may not be apparent on macroscopic examination. When amyloid accumulates in larger amounts, the organ is frequently enlarged and the tissue appears gray with a waxy, firm consistency. Histologically, the amyloid deposition is always extracellular and begins between cells, often closely adjacent to basement membranes (Fig. 6.47A). As the amyloid accumulates, it encroaches on the cells, in time surrounding and destroying them. In the form associated with plasma cell proliferation, perivascular and vascular deposits are common.

The histologic diagnosis of amyloid is based almost entirely on its staining characteristics. The most common staining technique uses the dye Congo red, which under ordinary light imparts a pink or red color to amyloid deposits. Under polarized light, the Congo red–stained amyloid shows so-called apple-green birefringence (Fig. 6.47B). This reaction is shared by all forms of amyloid and is caused by the cross-β-pleated configuration of amyloid fibrils. Confirmation can be obtained by electron microscopy, which reveals amorphous nonoriented thin fibrils. AA, AL, and ATTR types of amyloid can also be distinguished by specific immunohistochemical staining.

The pattern of organ involvement in different clinical forms of amyloidosis is variable.

Kidney. Amyloidosis of the kidney is the most common and potentially the most serious form of organ involvement. Grossly, the kidneys may be of normal size and color, or, in advanced cases, they may be shrunken because of ischemia caused by vascular narrowing induced by the deposition of amyloid within arterial and arteriolar walls.

Histologically, the amyloid is deposited primarily in the glomeruli, but the interstitial peritubular tissue, arteries, and arterioles are also affected. The glomerular deposits first appear as subtle thickenings of the mesangial matrix, accompanied usually by uneven widening of the basement membranes of the glomerular capillaries.

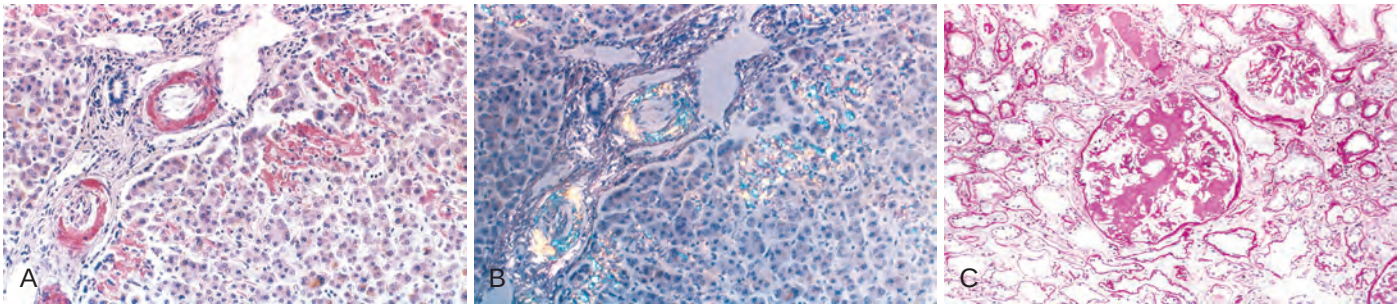


Figure 6.47 Amyloidosis. (A) A section of the liver stained with Congo red reveals pink-red deposits of amyloid in the walls of blood vessels and along sinusoids. (B) Note the yellow-green birefringence of the deposits when observed by polarizing microscope. (C) Amyloidosis of the kidney. The glomerular architecture is almost totally obliterated by the massive accumulation of amyloid. (A, B, Courtesy Dr. Trace Worrell and Sandy Hinton, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

In time, the mesangial depositions and the deposits along the basement membranes cause capillary narrowing and distortion of the glomerular vascular tuft. With progression of the glomerular amyloidosis, the capillary lumens are obliterated, and the obsolescent glomerulus is flooded by confluent masses or interlacing broad ribbons of amyloid (Fig. 6.47C).

Spleen. Amyloidosis of the spleen may be inapparent grossly or may cause moderate to marked splenomegaly (up to 800 g). For completely mysterious reasons, one of two patterns of deposition is seen. In one, the deposits are largely limited to the splenic follicles, producing tapioca-like granules on gross inspection, designated *sago spleen.* In the other pattern, the amyloid involves the walls of the splenic sinuses and connective tissue framework in the red pulp. Fusion of the early deposits gives rise to large, maplike areas of amyloidosis, creating what has been designated *lardaceous spleen.*

Liver. The deposits may be inapparent grossly or may cause moderate to marked hepatomegaly. Amyloid appears first in the space of Disse and then progressively encroaches on adjacent hepatic parenchymal cells and sinusoids (see Fig. 6.47). In time, deformity, pressure atrophy, and disappearance of hepatocytes occur, causing total replacement of large areas of liver parenchyma. Vascular involvement is frequent. Even with extensive involvement, liver function is usually preserved.

Heart. Amyloidosis of the heart (Chapter 12) may occur in any form of systemic amyloidosis. It is also the major organ involved in senile systemic amyloidosis. The heart may be enlarged and firm, but more often it shows no significant changes on gross inspection. The deposits begin as focal subendocardial accumulations and within the myocardium between the muscle fibers. Expansion of these myocardial deposits eventually causes pressure atrophy of myocardial fibers. When the amyloid deposits are subendocardial, the conduction system may be damaged, accounting for the electrocardiographic abnormalities noted in some patients.

Other Organs. Nodular deposits in the tongue may cause macroglossia, giving rise to the designation *tumor-forming amyloid of the tongue.* The respiratory tract may be involved focally or diffusely from the larynx down to the smallest bronchioles. As mentioned earlier, a distinct form of amyloid is found in the brains of patients with Alzheimer disease. It may be present in so-called *plaques* as well as blood vessels (Chapter 28). Amyloidosis of peripheral and autonomic nerves is a feature of several familial amyloidotic neuropathies. Depositions of amyloid in patients on long-term hemodialysis are most prominent in the carpal ligament of the wrist, resulting in compression of the median nerve (carpal tunnel syndrome). These patients may also have extensive amyloid deposition in the joints.

Clinical Features

Amyloidosis may be found as an unsuspected anatomic change, having produced no clinical manifestations, or it may cause serious clinical problems and even death. The symptoms depend on the magnitude of the deposits and on the sites or organs affected. Clinical manifestations at first are often nonspecific, such as weakness, weight loss, lightheadedness, or syncope. Somewhat more specific findings appear later and most often relate to renal, cardiac, and gastrointestinal involvement.

Renal involvement gives rise to proteinuria that may be severe enough to cause the nephrotic syndrome (Chapter 20). Progressive obliteration of glomeruli in advanced cases ultimately leads to renal failure and uremia. Renal failure is a common cause of death. *Cardiac amyloidosis* may present as an insidious congestive heart failure. The most serious aspects of cardiac amyloidosis are conduction disturbances and arrhythmias, which may prove fatal. Occasionally, cardiac amyloidosis produces a restrictive pattern of cardiomyopathy and masquerades as chronic constrictive pericarditis (Chapter 12). *Gastrointestinal amyloidosis* may be entirely asymptomatic, or it may present in a variety of ways. Amyloidosis of the tongue may cause sufficient enlargement and inelasticity to hamper speech and swallowing. Depositions in the stomach and intestine may lead to malabsorption, diarrhea, and disturbances in digestion. *Vascular amyloidosis* causes vascular fragility that may lead to bleeding, sometimes massive, that can occur spontaneously or following seemingly trivial trauma. Additionally, in some cases AL amyloid binds and inactivates factor X, a critical coagulation factor, leading to a life-threatening bleeding disorder.

The diagnosis of amyloidosis depends on the histologic demonstration of amyloid deposits in tissues. The most common sites biopsied are the kidney, when renal manifestations are present, or rectal or gingival tissues in patients suspected of having systemic amyloidosis. Examination of abdominal fat aspirates stained with Congo red can also be used for the diagnosis of systemic amyloidosis. The test is quite specific, but its sensitivity is low. In suspected cases of AL amyloidosis, serum and urine protein electrophoresis and immunoelectrophoresis should be performed. Bone marrow aspirates in such cases often show monoclonal plasmacytosis, even in the absence of overt multiple myeloma. Scintigraphy with radiolabeled serum amyloid P component is a rapid and specific test, since it binds to the amyloid deposits and reveals their presence. It also gives a measure of the extent of amyloidosis and can be used to follow patients undergoing treatment.

The prognosis for individuals with generalized amyloidosis is poor. Those with AL amyloidosis have an overall median survival of 2 years after diagnosis, and the prognosis is even poorer with myeloma-associated AL amyloidosis. Nevertheless, many new, more effective drugs have recently been deployed to treat plasma cell tumors and offer some hope. The outlook for individuals with reactive systemic amyloidosis is somewhat better and depends to some extent on the control of the underlying condition. Resorption of amyloid after treatment of the associated condition has been reported, but this is a rare occurrence. New therapeutic strategies aimed at correcting protein misfolding and inhibiting fibrillogenesis are being developed.

KEY CONCEPTS

AMYLOIDOSIS

- Amyloidosis is a disorder characterized by the extracellular deposits of misfolded proteins that aggregate to form insoluble fibrils.
- The deposition of these proteins may result from: excessive production of proteins that are prone to misfolding and aggregation; mutations that produce proteins that cannot fold properly and tend to aggregate; and defective or incomplete proteolytic degradation of extracellular proteins.

- Amyloidosis may be localized or systemic. It is seen in association with a variety of primary disorders, including monoclonal B-cell proliferations (in which the amyloid deposits consist of immunoglobulin light chains); chronic inflammatory diseases such as rheumatoid arthritis (deposits of amyloid A protein, derived from an acute-phase protein produced in inflammation); Alzheimer disease (amyloid β protein); familial conditions in which the amyloid deposits consist of mutants of normal proteins (e.g., transthyretin in familial amyloid polyneuropathies); and amyloidosis associated with dialysis (deposits of β_2-microglobulin, whose clearance is defective).
- Amyloid deposits cause tissue injury and impair normal function by causing pressure on cells and tissues. They do not evoke an inflammatory response.

SUGGESTED READINGS

Innate Immunity

Broz P, Dixit V: Inflammasomes: mechanisms of assembly, regulation and signaling, *Nat Rev Immunol* 16:407–420, 2016. *[An excellent discussion of the biochemistry of inflammasome activation].*

Brubaker SW, Bonham KS, Zanoni I et al: Innate immune pattern recognition: a cell biological perspective, *Ann Rev Immunol* 33:257–290, 2015. *[An excellent review of the structure and biology of innate immune receptors].*

Goubau D, Deddouche S, Reis E et al: Cytosolic sensing of viruses, *Immunity* 38:855–869, 2013. *[An excellent review on the numerous mechanisms used by cells to recognize viral DNA and RNA].*

Manthiram K, Zhou Q, Aksentijevich I et al: The monogenic autoinflammatory diseases define new pathways in human innate immunity and inflammation, *Nat Immunol* 18:832–842, 2017. *[A comprehensive review of diseases caused by mutations in inflammasome-related pathways, and what they teach us about these pathways].*

Pandey S, Kawai T, Akira S: Microbial sensing by Toll-like receptors and intracellular nucleic acid sensors. *Cold Spring Harb Persp Biol* 7:a016246, 2014. *[An excellent review of the receptors used by the innate immune system to sense microbes].*

Cell-Mediated Immunity

O'Shea JJ, Paul WE: Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells, *Science* 327:1098–1102, 2010. *[An excellent review of the development and functions of helper T cell subsets, and the uncertainties in the field].*

Raphael I, Nalawade S, Eagar TN et al: T cell subsets and their signature cytokines in autoimmune and inflammatory diseases, *Cytokine* 74:5–17, 2015. *[An overview of helper T cell subsets and the functions of their cytokines].*

Humoral Immunity

Crotty S: T follicular helper cell biology: a decade of discovery and diseases, *Immunity* 50:1132–1148, 2019. *[A discussion of the properties and generation of follicular helper T cells and their roles in antibody production and autoimmunity].*

Goodnow CC, Vinuesa CG, Randall KL et al: Control systems and decision making for antibody production, *Nat Immunol* 11:681–688, 2010. *[An excellent discussion of the major control points in the process of antibody production].*

Immediate Hypersensitivity, Allergy

Galli SJ, Tsai M: IgE and mast cells in allergic disease, *Nat Med* 18:693–704, 2012. *[An excellent review of the roles of IgE antibody and mast cells in chronic allergic diseases].*

Kauffmann F, Demenais F: Gene-environment interactions in asthma and allergic diseases: challenges and perspectives, *J Allergy Clin Immunol* 130:1229–1240, 2012. *[A thoughtful discussion of the complex interactions between susceptibility genes and environmental influences that may underlie allergic diseases].*

Lambrecht BN, Hammad H: The immunology of asthma, *Nat Immunol* 16:45–56, 2015. *[A comprehensive discussion of the roles of Th2 cells, cytokines, and other cells of the immune system in the development and resolution of asthma].*

Portelli MA, Hodge E, Sayers I: Genetic risk factors for the development of allergic disease identified by genome-wide association, *Clin Exp Allergy* 45:21–31, 2015. *[A summary of the genetic associations identified in asthma and other allergic diseases, and their potential significance*].

Other Hypersensitivity Reactions

Jancar S, Sanchez Crespo M: Immune complex–mediated tissue injury: a multistep paradigm, *Trends Immunol* 26:48–55, 2005. *[A summary of the mechanisms of immune complex–mediated tissue injury].*

Sturfelt G, Truedsson L: Complement in the immunopathogenesis of rheumatic disease, *Nat Rev Rheumatol* 8:458–468, 2012. *[A review of complement deficiencies and the role of the complement system in autoimmune diseases].*

Weaver CT, Elson CO, Fouser LA et al: The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin, *Annu Rev Pathol Mech Dis* 8:477–512, 2013. *[A detailed discussion of the biology of Th17 cells and their involvement in inflammatory diseases].*

Immunological Tolerance

Cheng M, Anderson MS: Thymic tolerance as a key brake on autoimmunity, *Nat Immunol* 19:659–664, 2018. *[An excellent review of thymic selection as a mechanism for self-tolerance].*

Dominguez-Villar M, Hafler DA: Regulatory T cells in autoimmune disease, *Nat Immunol* 19:665–673, 2018. *[A review discussing the abnormalities in regulatory T cells that may contribute to autoimmunity].*

Mueller DL: Mechanisms maintaining peripheral tolerance, *Nat Immunol* 11:21–27, 2010. *[A discussion of the mechanisms of peripheral tolerance, with an emphasis on T cells].*

Ohkura N, Kitagawa Y, Sakaguchi S: Development and maintenance of regulatory T cells, *Immunity* 38:414–423, 2013. *[An excellent review of the molecular mechanisms underlying the generation, maintenance, and stability of regulatory T cells].*

Schildberg FA, Klein SR, Freeman GJ et al: Coinhibitory pathways in the B7-CD28 ligand-receptor family, *Immunity* 44:955–972, 2016. *[An excellent review of coinhibitors and their roles in autoimmune diseases and cancer progression].*

Schwartz RH: Historical overview of immunological tolerance, *Cold Spring Harb Perspect Biol* 4:2012. a006908. *[A thoughtful summary of the mechanisms of tolerance, the experimental studies behind the elucidation of these mechanisms, and how they may be disrupted to give rise to autoimmunity].*

Mechanisms of Autoimmunity: General

Bach J-F: The hygiene hypothesis in autoimmunity: the role of pathogens and commensals, *Nat Rev Immunol* 18:105–120, 2018. *[An update on the contribution of microbes to the development of autoimmunity].*

Cheng MH, Anderson MS: Monogenic autoimmunity, *Annu Rev Immunol* 30:393–427, 2012. *[An excellent review of autoimmune syndromes caused by single-gene mutations, and what they teach us about pathways of immunological tolerance].*

Dendrou CA, Petersen J, Rossjohn J et al: HLA variation and disease, *Nat Rev Immunol* 18:325–339, 2018. *[An excellent review of how HLA polymorphisms may contribute to autoimmune diseases].*

Goodnow CC: Multistep pathogenesis of autoimmune disease, *Cell* 130:25, 2007. *[An excellent discussion of the checkpoints that prevent autoimmunity and why these might fail].*

Inshaw JRJ, Cutler AJ, Burren OS et al: Approaches and advances in the genetic causes of autoimmunity and their implications, *Nat Immunol* 19:674–684, 2018. *[A review of genetic associations with autoimmune diseases and how these may be analyzed].*

Lenardo M, Lo B, Lucas CL: Genomics of immune diseases and new therapies, *Ann Rev Immunol* 34:121–149, 2016. *[An excellent review of genetic causes of autoimmunity].*

Marson A, Housley WJ, Hafler DA: Genetic basis of autoimmunity, *J Clin Invest* 125:2234–2241, 2015. *[A review of genes associated with autoimmunity].*

Mathis D, Benoist C: Microbiota and autoimmune disease: the hosted self, *Cell Host Microbe* 10:297–301, 2011. *[A review of the evidence that the microbiome influences immune activation and autoimmunity, and the relevance of these findings to human autoimmune diseases].*

Theofilopoulos A, Kono DH, Baccala R: The multiple pathways to autoimmunity, *Nat Immunol* 18:716–724, 2017. *[A summary of immunologic aberrations that may cause autoimmunity].*

Systemic Lupus Erythematosus

Crow MK, Olferiev M, Kirou K: Type I interferons in autoimmune disease, *Ann Rev Pathol Mech Dis* 14:369–393, 2019. *[A review of the recently discovered role of interferons in SLE and other autoimmune diseases, and the potential for targeting this family of cytokines for therapy].*

Kaul A, Gordon C, Crow MK et al: Systemic lupus erythematosus, *Nat Rev Dis Primers* 2:16039, 2016. *[An excellent review of the clinical features and pathogenesis of lupus].*

Maria NI, Davidson A: Emerging areas for therapeutic discovery in SLE, *Curr Opin Immunol* 55:1–8, 2018. *[An excellent review of recent advances in understanding the genetics of lupus and the roles of innate and adaptive immune responses in the disease, and how these advances are shaping the development of novel therapies].*

Sjogren Syndrome, Systemic Sclerosis, and Other Systemic Autoimmune Diseases

Jennette JC, Falk RJ, Hu P et al: Pathogenesis of antineutrophil cytoplasmic autoantibody-associated small-vessel vasculitis, *Annu Rev Pathol* 8:139–160, 2013. *[A comprehensive review of the clinical and pathologic features and pathogenesis of small vessel vasculitis].*

Katsumoto TR, Whitfield ML, Connolly MK: The pathogenesis of systemic sclerosis, *Annu Rev Pathol Mech Dis* 6:509–537, 2011. *[An excellent review of the pathogenesis of systemic sclerosis, and the many unanswered questions].*

Mahajan VS, Mattoo H, Deshpande V et al: IgG4-related disease, *Annu Rev Pathol Mech Dis* 9:315–347, 2014. *[An excellent review of the clinical and pathologic features and likely autoimmune pathogenesis of a quite recently identified multisystem fibrotic disease].*

Mavragani CP: Mechanisms and new strategies for primary Sjogren's Syndrome, *Annu Rev Med* 68:331–343, 2017. *[A review of the pathogenesis and clinical features of Sjogren's syndrome].*

Pattanaik D, Brown M, Postlethwaite BC et al: Pathogenesis of systemic sclerosis, *Front Immunol* 6:272, 2015. *[A discussion of current concepts of the pathogenesis of systemic sclerosis].*

Rejection of Transplants

Mitchell RN: Graft vascular disease: immune response meets the vessel wall, *Annu Rev Pathol Mech Dis* 4:19–47, 2009. *[A review of the mechanisms that lead to vascular disease in chronic graft rejection].*

Nagy ZA: Alloreactivity: an old puzzle revisited, *Scand J Immunol* 75:463–470, 2012. *[A thoughtful discussion of the evolution of ideas about allorecognition, and the current understanding of the phenomenon].*

Nankivell BJ, Alexander SI: Rejection of the kidney allograft, *N Engl J Med* 363:1451, 2010. *[Good review of the mechanisms of recognition and rejection of allografts and the development of new strategies for treating rejection].*

Valenzuela NM, Reed EF: Antibody-mediated rejection across solid organ transplants: manifestations, mechanisms, and therapies, *J Clin Invest* 127:2492–2504, 2017. *[A review of the role of antibodies in the rejection of transplants].*

Wood KJ, Goto R: Mechanisms of rejection: current perspectives, *Transplantation* 93:1–10, 2012. *[An excellent review of the steps in the recognition of alloantigens and the activation of alloreactive lymphocytes, and the mechanisms of graft rejection].*

Primary Immunodeficiency Diseases

Casanova JL, Abel L: Human genetics of infectious diseases: unique insights into immunological redundancy, *Semin Immunol* 36:1–12, 2018. *[A thoughtful review of the genetic basis of primary immunodeficiency diseases, focusing on the restricted clinical manifestations of many of these disorders].*

Durandy A, Kracker S, Fischer A: Primary antibody deficiencies, *Nat Rev Immunol* 13:519–533, 2013. *[A summary of primary immunodeficiencies affecting B cells].*

Fischer A, Notarangelo LD, Neven B et al: Severe combined immunodeficiencies and related disorders, *Nat Rev Dis Primers* 1:15061, 2015. *[A review of the pathogenesis and clinical features of severe combined immunodeficiency].*

Parvaneh N, Casanova JL, Notarangelo LD et al: Primary immunodeficiencies: a rapidly evolving story, *J Allergy Clin Immunol* 131:314–323, 2013. *[An excellent review of newly described primary immunodeficiency syndromes].*

HIV and AIDS

Douek DC, Roederer M, Koup RA: Emerging concepts in the immunopathogenesis of AIDS, *Annu Rev Med* 60:471, 2009. *[A balanced discussion of the pathogenesis of AIDS, and the unresolved issues].*

Cohen MS, Shaw GM, McMichael AJ et al: 1 infection, *N Engl J Med* 364:1943–1954, 2011. *[An excellent summary of the early phase of HIV infection and host responses to the virus].*

Moir S, Chun TW, Fauci AS: Pathogenic mechanisms of HIV disease, *Annu Rev Pathol Mech Dis* 6:223–248, 2011. *[A discussion of the mechanisms by which HIV causes immunodeficiency].*

Amyloidosis

Buxbaum JN, Linke RP: A molecular history of the amyloidoses, *J Mol Biol* 421:142–159, 2012. *[A thoughtful review of how our understanding of amyloid proteins and their role in disease has evolved, and the molecular studies that have led to current concepts].*

Merlini G, Dispenzieri A, Sanchorawala V et al: Systemic immunoglobulin light chain amyloidosis, *Nat Rev Dis Primers* 4:38, 2018. *[A review of the pathogenesis and clinical features of the AL form of amyloidosis].*

Westermark GT, Fandrich M, Westermark P: AA amyloidosis: pathogenesis and targeted therapies, *Ann Rev Pathol Mech Dis* 10:321, 2015. *[An excellent review of the pathogenesis and clinical features of a major form of amyloidosis].*

Dogan A: Amyloidosis: insights from proteomics, *Annu Rev Pathol Mech Dis* 12:277–304, 2017. *[An overview of the biochemistry of amyloid and the identification of different forms of amyloidosis].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 7

Neoplasia

CHAPTER CONTENTS

Cancer is the second leading cause of death in the United States; only cardiovascular diseases exact a higher toll. Even more agonizing than the mortality rate is the emotional and physical suffering inflicted by cancers. Patients and the public often ask, "When will there be a cure for this scourge?" The answer to this simple question is difficult because cancer is not one disease but many disorders with widely different pathogeneses, natural histories, and responses to treatments. Some cancers, such as Hodgkin lymphoma, are curable, whereas others, such as pancreatic adenocarcinoma, are virtually always fatal. The only hope for controlling cancer lies in learning more about its causes and pathogenesis. Fortunately, great strides have been made in understanding its molecular basis, and some good news has emerged: cancer mortality for both men and women in the United States declined during the last decade of the 20th century and has continued its downward course in the 21st century.

In this chapter, we describe the vocabulary of tumor biology and pathology and then review the morphologic characteristics that define neoplasia and allow benign and malignant tumors to be identified and distinguished. Also reviewed is the epidemiology of cancer, which provides a measure of the impact of cancer on human populations as well as clues to its environmental causes, insights that have led to effective prevention campaigns against certain cancers. Building on this foundation, we then discuss the biologic properties of tumors and the molecular basis of carcinogenesis, emphasizing the critical role that genetic alterations play in the development of neoplasia. Finally, we turn to cancer diagnosis, focusing on new technologies that are helping to direct the use of cancer drugs that are targeted at particular molecular lesions. Throughout, we give examples of new analytic methods and therapies that are not only changing our approach to cancer treatment but also providing new insights into cancer pathophysiology.

NOMENCLATURE

Neoplasia means "new growth," and the collection of cells and stroma composing new growths are referred to as *neoplasms*. *Tumor* originally described swelling caused by inflammation, but is now equated with neoplasm. *Oncology* (Greek *oncos* = tumor) is the study of tumors or neoplasms. Although physicians know what they mean when they use the term *neoplasm,* it has been difficult to develop a precise definition. In the modern era, a neoplasm is defined as a genetic disorder of cell growth that is triggered by acquired or less commonly inherited mutations affecting a single cell and its clonal progeny. As discussed later, these causative mutations alter the function of particular genes and give the neoplastic cells a survival and growth advantage, resulting in excessive proliferation that is independent of physiologic growth signals and controls.

All tumors are composed of two components: (1) neoplastic cells that constitute the tumor parenchyma and (2) reactive stroma made up of connective tissue, blood vessels, and cells of the adaptive and innate immune system. The classification of tumors and their biologic behavior are based primarily on the parenchymal component, but their growth and spread are critically dependent on their stroma. In some tumors, connective tissue is scant, and the neoplasm is soft and fleshy. In others, parenchymal cells stimulate the formation of abundant collagenous stroma, referred to as *desmoplasia.* Some desmoplastic tumors—for example, some cancers of the female breast—are stony hard or *scirrhous.*

Benign Tumors

Benign tumors remain localized at their site of origin and are generally amenable to surgical removal. Predictably, the patient generally survives. Exceptions arise, however, when benign tumors occur in vulnerable locations such as the brain; here, even "benign" tumors may cause significant morbidity and are sometimes even fatal.

Naming of benign tumors of mesenchymal cells is relatively simple; in general, the suffix "-oma" is attached to the name of the cell type from which the tumor arises. Thus a benign tumor of fibroblast-like cells is called a *fibroma,* a benign cartilaginous tumor is a *chondroma,* and so on. The nomenclature of benign epithelial tumors is more complex; some are classified based on their cell of origin, others on their microscopic appearance, and still others on their macroscopic architecture. *Adenoma* is applied to benign epithelial neoplasms derived from glandular tissues even if the tumor cells fail to form glandular structures. Thus, a benign epithelial neoplasm that arises from renal tubular cells and forms tightly clustered glands and a mass of adrenal cortical cells growing as a solid sheet are both referred to as adenomas. Benign epithelial neoplasms producing fingerlike or warty projections from epithelial surfaces are called *papillomas,* whereas those that form large cystic masses, such as in the ovary, are referred to as *cystadenomas.* Some tumors produce papillary projections that protrude into cystic spaces and are called papillary cystadenomas. When a neoplasm—benign or malignant—produces a grossly visible projection above a mucosal surface, for example, into the gastric or colonic lumen, it is termed a *polyp*. If the polyp has glandular tissue, it is called an adenomatous polyp (Fig. 7.1).

Malignant Tumors

Malignant tumors can invade and destroy adjacent structures and spread to distant sites (metastasize). Malignant tumors are collectively referred to as *cancers,* derived from the Latin word for crab, because they tend to adhere to any part that they seize on in an obstinate manner. Not all cancers pursue a deadly course; some are discovered at early stages

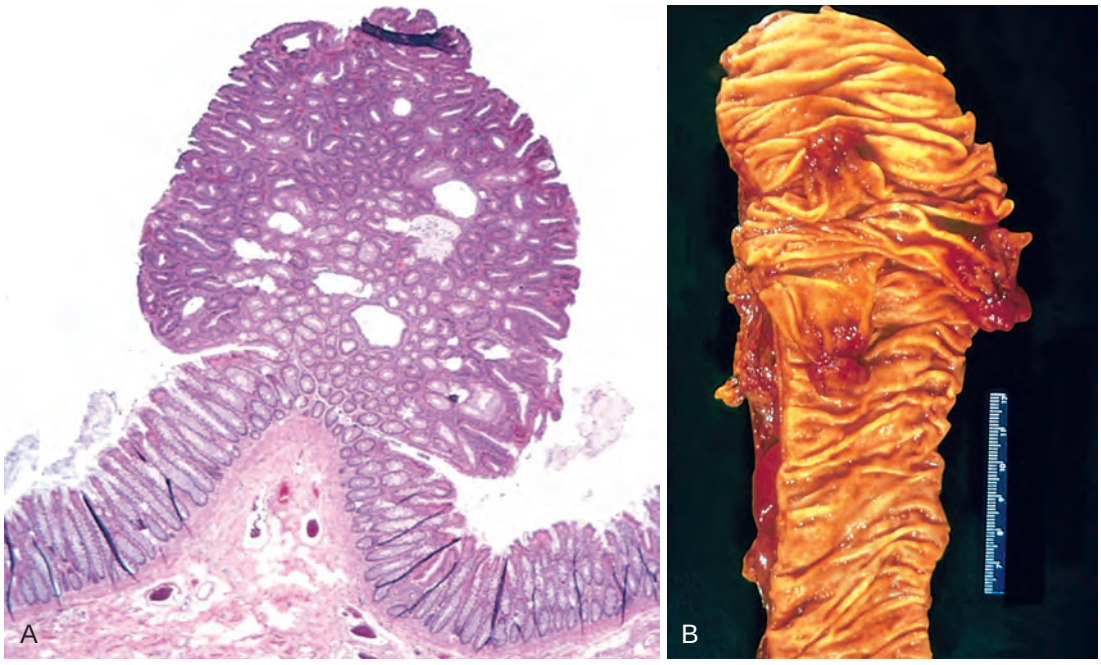


Figure 7.1 Colonic polyp. (A) An adenomatous (glandular) polyp is projecting into the colonic lumen and is attached to the mucosa by a distinct stalk. (B) Gross appearance of several colonic polyps.

that allow for surgical excision, and others are cured with systemically administered drugs or therapeutic antibodies. Nevertheless, the designation "malignant" always raises a red flag.

The nomenclature of malignant tumors follows essentially the same schema used for benign neoplasms, with certain additions. Malignant tumors arising in solid mesenchymal tissues are usually called *sarcomas* (Greek *sar* = fleshy; e.g., fibrosarcoma and chondrosarcoma), whereas those arising from blood-forming cells are designated *leukemias* (literally, white blood) or *lymphomas* (tumors of lymphocytes or their precursors). Malignant neoplasms of epithelial cell origin are called *carcinomas.* Carcinomas may be further qualified. In *squamous cell carcinoma* the tumor cells resemble stratified squamous epithelium, whereas in *adenocarcinoma* the neoplastic epithelial cells grow in a glandular pattern. Sometimes the tissue or organ of origin can be identified and is added as a descriptor, as in renal cell adenocarcinoma or bronchogenic squamous cell carcinoma. In approximately 2% of cases, cancers are composed of cells of unknown origin and must be designated merely as undifferentiated malignant tumors.

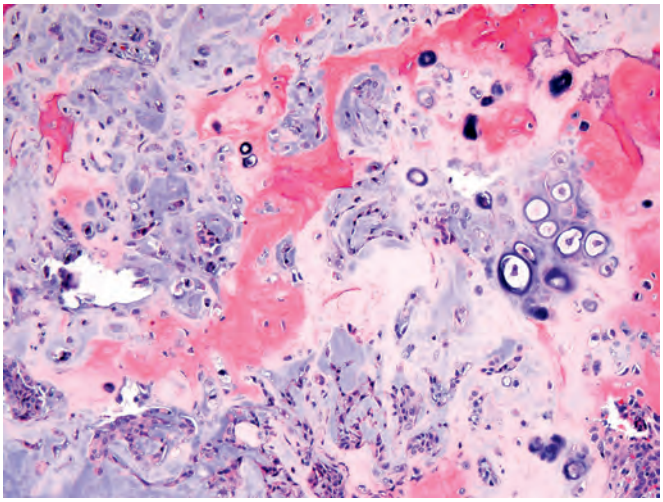


Figure 7.2 Mixed tumor of the parotid gland. Small nests of epithelial cells and myxoid stroma forming cartilage and bone (an unusual feature) are present in this field. (Courtesy Dr. Vicky Jo, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

Mixed Tumors

In most neoplasms, all parenchymal cells closely resemble one another, but in some types of tumors more than one line of differentiation is evident, creating distinct subpopulations of cells. A classic example is the mixed tumor of the salivary gland, which contains epithelial components scattered within a myxoid stroma that may contain islands of cartilage or bone (Fig. 7.2). All of these elements arise from a single neoplastic clone capable of producing both epithelial and mesenchymal cells; thus the preferred designation of this neoplasm is *pleomorphic adenoma.* The great majority of neoplasms, including mixed tumors, are composed of cells from a single germ layer (mesoderm, endoderm, or ectoderm). An exception is a tumor called a *teratoma,* which contains recognizable mature or immature cells or tissues belonging to more than one germ cell layer (and sometimes all three). Teratoma originates from totipotential germ cells that are normally present in the ovary and testis and sometimes also found in abnormal midline embryonic rests. Such cells can differentiate into any cell type found in the body and so, not surprisingly, may give rise to neoplasms that contain, in a helter-skelter fashion, bone, epithelium, muscle, fat, nerve, and other tissues. A particularly common pattern is seen in the *ovarian cystic teratoma (dermoid cyst),* which differentiates principally along ectodermal lines to create a cystic tumor lined by squamous epithelium that is replete with hair, sebaceous glands, and tooth structures (Fig. 7.3).

The nomenclature of the more common types of tumors is presented in Table 7.1. Included in this list are some inappropriate but deeply entrenched names. For instance, the benign-sounding designations lymphoma, melanoma, mesothelioma, and seminoma are used for malignant neoplasms. Alternatively, ominous-sounding terms are applied to some trivial lesions. *Hamartomas* are disorganized masses composed of cells indigenous to the involved tissue. Once thought to be a developmental malformation unworthy

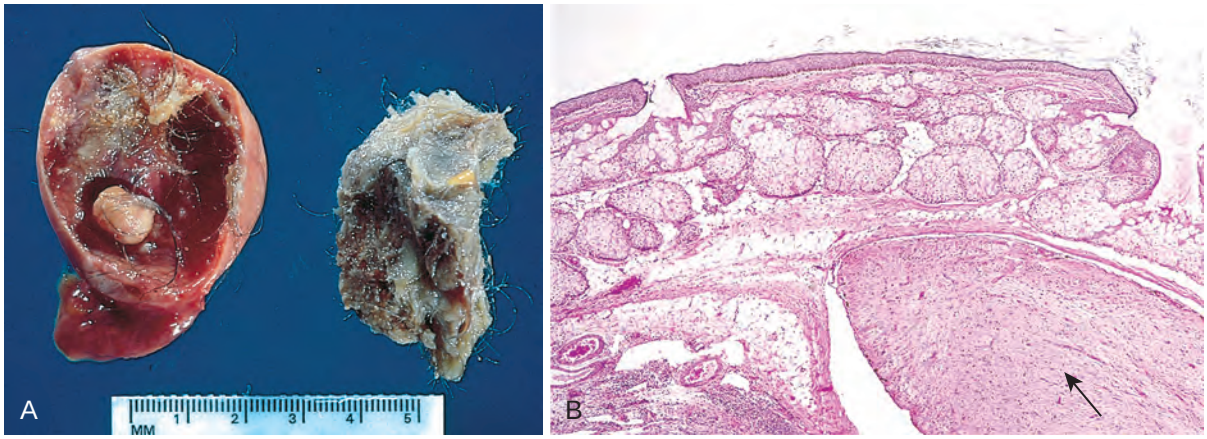


Figure 7.3 (A) Gross appearance of an opened cystic teratoma of the ovary. Note the presence of hair, sebaceous material, and a tooth. (B) Microscopic view of a similar tumor shows skin, sebaceous glands, fat cells, and a tract of neural tissue *(arrow).*

Table 7.1 Nomenclature of Tumors

Tissue of Origin	Benign	Malignant
Composed of One Parenchymal Cell Type		
Tumors of Mesenchymal Origin		
Connective tissue and derivatives	Fibroma Lipoma Chondroma Osteoma	Fibrosarcoma Liposarcoma Chondrosarcoma Osteogenic sarcoma
Vessels and Surface Coverings		
Blood vessels	Hemangioma	Angiosarcoma
Lymph vessels	Lymphangioma	Lymphangiosarcoma
Mesothelium	Benign fibrous tumor	Mesothelioma
Brain coverings	Meningioma	Invasive meningioma
Blood Cells and Related Cell Types		
Hematopoietic cells		Leukemias
Lymphoid tissue		Lymphomas
Muscle		
Smooth	Leiomyoma	Leiomyosarcoma
Striated	Rhabdomyoma	Rhabdomyosarcoma
Tumors of Epithelial Origin		
Stratified squamous	Squamous cell papilloma	Squamous cell carcinoma
Basal cells of skin or adnexa		Basal cell carcinoma
Melanocytes	Nevus	Malignant melanoma
Epithelial lining of glands or ducts	Adenoma Papilloma Cystadenoma	Adenocarcinoma Papillary carcinomas Cystadenocarcinoma
Respiratory passages	Bronchial adenoma	Bronchogenic carcinoma
Renal epithelium	Renal tubular adenoma	Renal cell carcinoma
Liver cells	Hepatic adenoma	Hepatocellular carcinoma
Urinary tract epithelium (transitional epithelium)	Transitional cell papilloma	Transitional cell carcinoma
Placenta epithelium	Hydatidiform mole	Choriocarcinoma
Testicular epithelium (germ cells)		Seminoma Embryonal carcinoma
More Than One Neoplastic Cell Type—Mixed Tumors, Usually Derived From One Germ Cell Layer		
Salivary glands	Pleomorphic adenoma (mixed tumor of salivary origin)	Malignant mixed tumor of salivary gland origin
Renal anlage		Wilms tumor
More Than One Neoplastic Cell Type Derived From More Than One Germ Cell Layer—Teratogenous		
Totipotential cells in gonads or in embryonic rests	Mature teratoma, dermoid cyst	Immature teratoma, teratocarcinoma

of the "-oma" designation, most hamartomas have clonal chromosomal aberrations that are acquired through somatic mutation and on this basis are now considered benign neoplasms. *Choristoma* is the term applied to a heterotopic (misplaced) rest of cells. For example, a small nodule of well-developed, normally organized pancreatic tissue may be found in the submucosa of the stomach or small intestine. The term *choristoma,* suggesting a neoplasm, imparts a gravity to these lesions that far exceeds their actual significance.

CHARACTERISTICS OF BENIGN AND MALIGNANT NEOPLASMS

The differentiation between benign and malignant tumors is one of the most important distinctions a pathologist can make, as nothing is more important to an individual with a tumor than being told, "It is benign." Although an innocent face may mask an ugly nature, benign and malignant tumors usually can be distinguished on the basis of various histologic and anatomic features (described later). Malignant tumors also tend to grow more rapidly than benign tumors, but there are so many exceptions that growth rate is not a reliable discriminator between benignity and malignancy. In fact, even cancers exhibit remarkably varied growth rates, from slow-growing tumors associated with survival for many years, often without treatment, to rapidly growing tumors that may be lethal within months or weeks.

Differentiation and Anaplasia

Differentiation refers to the extent to which neoplastic parenchymal cells resemble the corresponding normal parenchymal cells, both morphologically and functionally; lack of differentiation is called *anaplasia*. In general, benign tumors are well differentiated (Figs. 7.4 and 7.5). The neoplastic cells of a lipoma, a proliferation of benign adipocytes, may so closely resemble normal adipocytes as to be unrecognizable as a tumor by microscopic examination. Only the growth of these cells into a discrete mass discloses

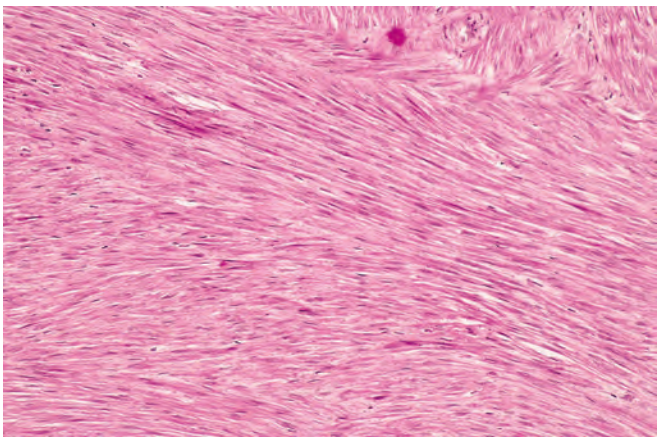


Figure 7.4 Leiomyoma of the uterus. This benign, well-differentiated tumor contains interlacing bundles of neoplastic smooth muscle cells that are virtually identical in appearance to normal smooth muscle cells in the myometrium.

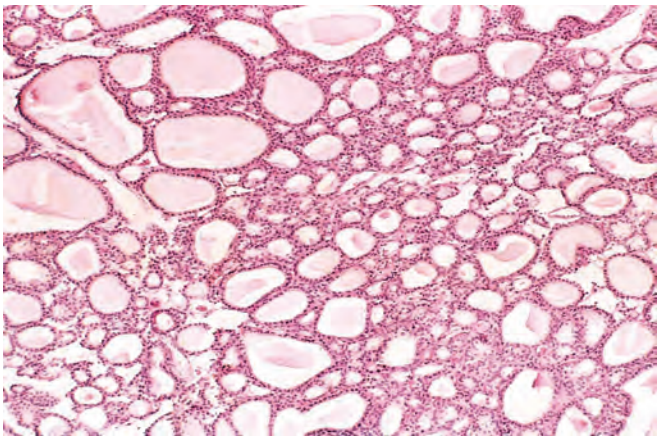


Figure 7.5 Benign tumor (adenoma) of the thyroid. Note the normal-looking (well-differentiated), colloid-filled thyroid follicles. (Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.)

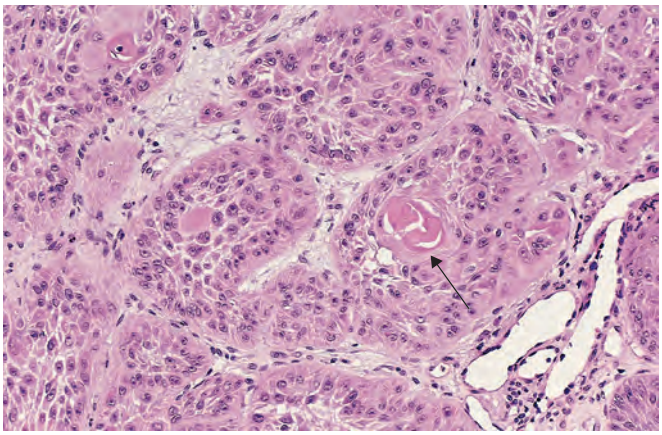


Figure 7.6 Well-differentiated squamous cell carcinoma of the skin. The tumor cells are strikingly similar to normal squamous epithelial cells, with intercellular bridges and nests of keratin pearls *(arrow)*. (Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.)

their neoplastic nature. One may get so close to the tree that one loses sight of the forest. In well-differentiated benign tumors, mitoses are usually rare and are of normal configuration.

By contrast, **most malignant neoplasms exhibit morphologic alterations that betray their potential for aggressive behavior**. In well-differentiated tumors, these features may be quite subtle. Well-differentiated adenocarcinomas of the thyroid, for example, form normal-appearing follicles, and some squamous cell carcinomas contain cells that appear identical to normal squamous epithelial cells (Fig. 7.6). The malignant nature of such tumors is revealed by invasion of adjacent tissues and their ability to metastasize. At the other end of the spectrum lie highly anaplastic, poorly differentiated tumors exhibiting little or no evidence of differentiation (Fig. 7.7), a morphologic appearance that is highly predictive of malignant behavior. In between these two extremes lie tumors that are loosely referred to as moderately well differentiated (Fig. 7.8).

In addition to anaplasia, cancer cells often exhibit other telltale morphologic changes:

- *Pleomorphism.* Pleomorphism refers to variation in cell size and shape. Thus, cells within the same tumor are not uniform, but range from small cells with an undifferentiated appearance to *tumor giant cells* many times larger than their neighbors. Some tumor giant cells possess only a single huge polymorphic nucleus, while others may have two or more large, hyperchromatic nuclei (Fig. 7.9). These giant cells are not to be confused with inflammatory Langhans or foreign body giant cells, which are derived from macrophages and contain many small, normal-appearing nuclei.
- *Abnormal nuclear morphology.* Characteristically, cancer cells have nuclei that are disproportionately large, with a nuclear-to-cytoplasm ratio that may approach 1:1 instead of the normal 1:4 to 1:6. The nuclear shape is variable and often irregular, and the chromatin is often coarsely clumped and distributed along the nuclear

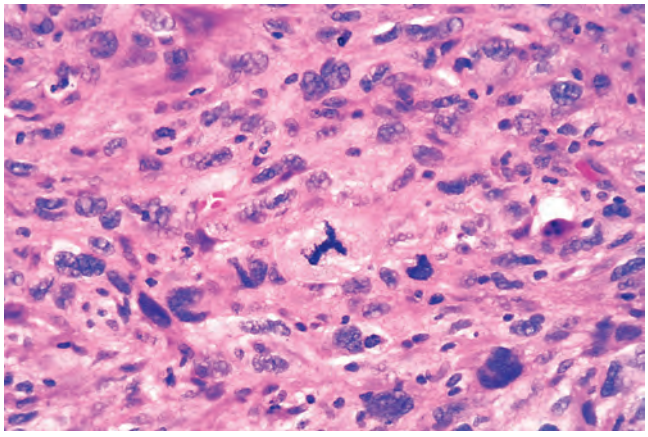


Figure 7.7 Anaplastic tumor showing cellular and nuclear variation in size and shape. The prominent cell in the center field has an abnormal tripolar spindle.

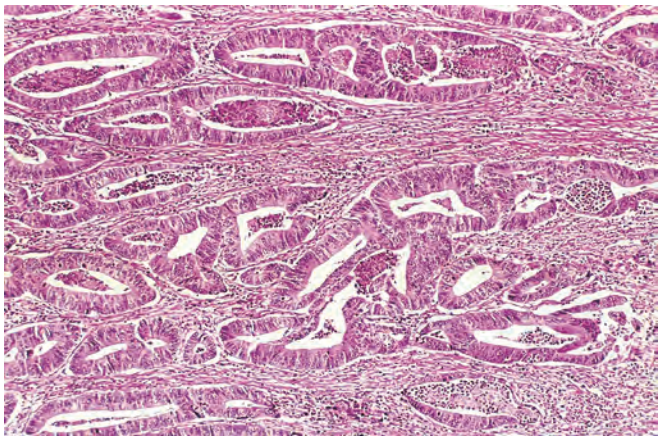


Figure 7.8 Malignant tumor (adenocarcinoma) of the colon. Note that compared with the well-formed and normal-looking glands characteristic of a benign tumor, the cancerous glands are irregular in shape and size and do not resemble the normal colonic glands. This tumor is considered moderately well differentiated because gland formation is seen. The malignant glands have invaded the muscular layer of the colon. (Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.)

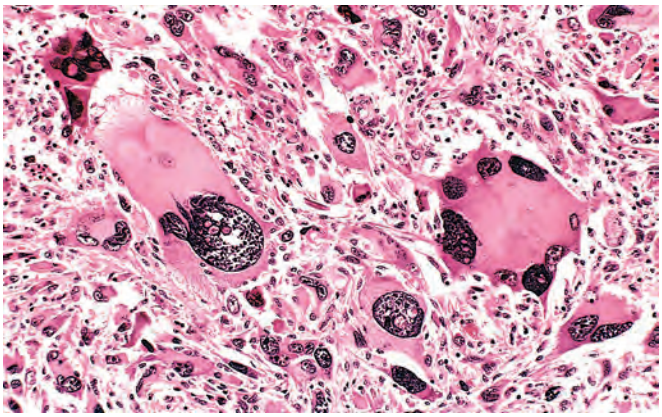


Figure 7.9 Pleomorphic tumor of the skeletal muscle (rhabdomyosarcoma). Note the marked cellular and nuclear pleomorphism, hyperchromatic nuclei, and tumor giant cells. (Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.)

membrane or more darkly stained than normal *(hyperchromatic)*. Abnormally large nucleoli are also commonly seen.

- *Mitoses.* Unlike benign tumors and some well-differentiated malignant neoplasms, undifferentiated cancers often contain many cells in mitosis, reflecting their high rate of proliferation. The presence of mitoses, however, does not equate with malignancy. For example, cells in mitosis are often seen in normal tissues exhibiting rapid turnover, such as the epithelial lining of the gut and nonneoplastic proliferations such as hyperplasias. More important as a morphologic feature of malignancy are *atypical, bizarre mitotic figures* (see Fig. 7.8).
- *Loss of polarity.* In addition to cytologic abnormalities, the orientation of anaplastic cells with respect to each other or to supporting structures like basement membranes is markedly disturbed. Sheets or large masses of tumor cells grow in a disorganized fashion.
- *Other changes.* While growing tumor cells must have a blood supply, the vascular stroma is often insufficient; as a result, many rapidly growing cancers develop areas of ischemic necrosis.

As one might surmise, well-differentiated transformed cells have a greater likelihood of retaining the functional capabilities of their normal counterparts. Benign tumors are almost always well differentiated and often retain normal functions, as do many well-differentiated cancers. Thus, well-differentiated tumors of endocrine glands frequently secrete hormones characteristic of their origin into the blood, where they can be detected and quantified to diagnose and follow the response of such tumors to treatment. Similarly, well-differentiated squamous cell carcinomas synthesize keratin, and well-differentiated hepatocellular carcinomas elaborate bile. By contrast, highly anaplastic undifferentiated tumors typically lose the specialized functional activities of their tissue of origin, but sometimes acquire new and unanticipated functions. Thus some malignant tumors express fetal proteins that are not produced by their normal adult counterparts, while others express proteins that are normally found only in other types of cells. For example, bronchogenic carcinomas may produce corticotropin, parathyroid-like hormone, insulin, glucagon, and other hormones, giving rise to paraneoplastic syndromes (described later).

Metaplasia, Dysplasia, and Carcinoma In Situ

These terms describe morphologically recognizable changes in differentiation that variously represent an adaptation to chronic injury (metaplasia), a premalignant change (dysplasia), or a cancer that has yet to invade (carcinoma in situ).

- *Metaplasia* is defined as the replacement of one type of cell with another type (Chapter 2). Metaplasia is nearly always found in association with tissue damage, repair, and regeneration. Often the replacing cell type is better suited to some alteration in the local environment. For example, in Barrett esophagus, gastroesophageal reflux damages the squamous epithelium of the esophagus, leading to its replacement by glandular (gastric or intestinal) epithelium better suited to an acidic environment. Unfortunately, the metaplastic epithelium is prone to malignant transformation. The same is true of squamous metaplasia of the bronchial epithelium in chronic smokers, often a prelude to the development of lung cancer.
- *Dysplasia* is a term that literally means "disordered growth." It is encountered principally in epithelial cells and is recognized on the basis of several morphologic changes. Dysplastic cells may exhibit considerable pleomorphism and often contain large hyperchromatic nuclei with a high nuclear-to-cytoplasmic ratio. Dysplastic epithelial surfaces also typically show architectural disarray and a loss of orderly differentiation. For example, in dysplastic squamous epithelium the normal progressive maturation of tall cells in the basal layer to flattened squames on the surface may fail in part or entirely, leading to replacement of the epithelium by basal-like cells with hyperchromatic nuclei. In addition, mitotic figures are more abundant than in the normal squamous epithelium and may be seen throughout dysplastic epithelium, rather than being confined to the basal layer, as is the normal case.
- *Carcinoma in situ.* When dysplasia is severe and involves the full thickness of the epithelium but the lesion does not penetrate the basement membrane, it is referred to as *carcinoma in situ* (Fig. 7.10). Carcinoma in situ is often seen in the skin, breast, bladder, and uterine cervix. In situ epithelial cancers display all of the cytologic features of malignancy and unless treated have high probability of progression to invasive cancers.

Dysplastic changes are often found adjacent to foci of invasive carcinoma, and in some situations, such as in the cervix, severe epithelial dysplasia or carcinoma in situ frequently antedates the appearance of cancer. Moreover, some mutations associated with full-blown cancer (described later) may be present in even "mild" dysplasias. Nevertheless, **although dysplasia may be a precursor to malignant transformation, it does not always progress to cancer.** With removal of inciting causes, even moderately severe dysplasias may be completely reversible. Even carcinoma in situ may persist for years before it becomes invasive. As discussed later, cancers arise by accumulation of mutations, and the time interval for evolution of full-blown cancers from in situ lesions relates most likely to the time required for

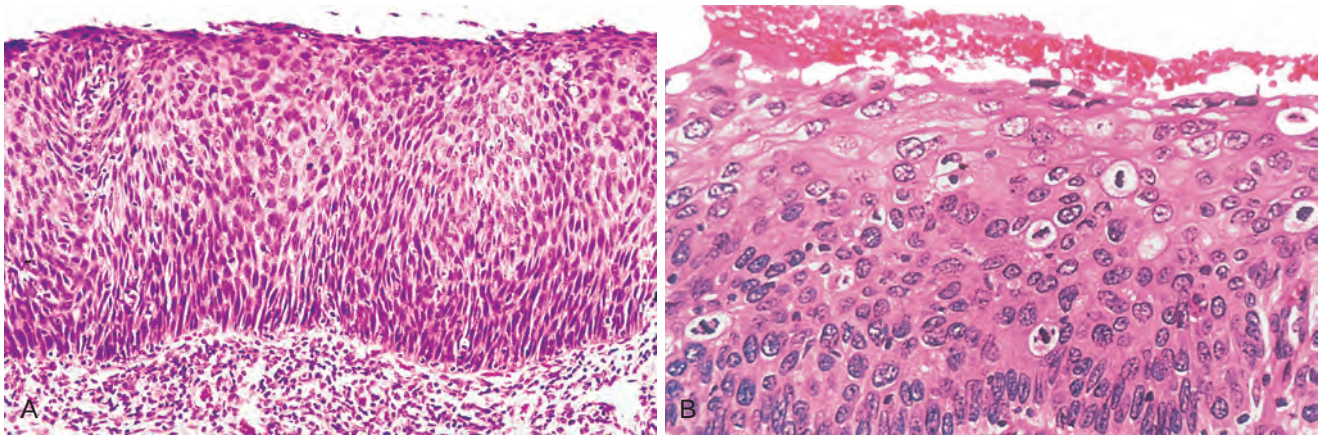


Figure 7.10 (A) Carcinoma in situ. Low-power view shows that the epithelium is entirely replaced by atypical dysplastic cells. There is no orderly differentiation of squamous cells. The basement membrane is intact, and there is no tumor in the subepithelial stroma. (B) High-power view of another region shows failure of normal differentiation, marked nuclear and cellular pleomorphism, and numerous mitotic figures extending toward the surface.

accumulation of all the mutations that are needed to induce a fully malignant phenotype. Finally, it should be noted that while dysplasia often occurs in metaplastic epithelium, not all metaplastic epithelium is dysplastic.

Local Invasion

The growth of cancers is accompanied by progressive invasion, destruction of surrounding tissue, and eventually systemic spread, whereas nearly all benign tumors grow as cohesive, expansile masses that remain localized to their site of origin and lack the capacity to invade or metastasize to distant sites. Because benign tumors grow and expand slowly, they usually develop a rim of compressed fibrous tissue called a capsule that separates them from the surrounding normal tissue. The tumor capsule consists of extracellular matrix (ECM) deposited by stromal cells such as fibroblasts, which are activated by hypoxic damage resulting from the pressure of the expanding tumor. Such encapsulation creates a tissue plane that makes the tumor discrete, readily palpable, movable (nonfixed), and easily excisable by surgical enucleation (Fig. 7.11). There are a few exceptions to this rule, however. For example, hemangiomas (benign neoplasms composed of tangled blood vessels) are often unencapsulated and permeate the site in which they arise (e.g., the dermis of the skin and the liver); when such lesions are extensive, they may be unresectable.

In contrast, malignant tumors are, in general, poorly demarcated from the surrounding normal tissue and lack well-defined cleavage planes (Fig. 7.12). Slowly expanding malignant tumors may develop an apparently enclosing fibrous capsule and push along a broad front into adjacent normal structures. However, histologic examination of such "pseudoencapsulated" masses almost always shows rows of tumor cells penetrating the margin and infiltrating adjacent structures, a crablike pattern of growth that fits the popular image of cancer.

Next to the development of metastases, invasiveness is the most reliable discriminator of malignant and benign tumors. Most malignant tumors do not recognize normal anatomic boundaries. Given time, they will penetrate the wall of the colon or uterus, for example, or fungate through the surface of the skin. This invasiveness makes their complete surgical resection difficult or impossible, and even if the tumor appears well circumscribed it is necessary to remove a large margin of adjacent, apparently normal tissue to ensure complete local excision.

Metastasis

Metastasis is defined as the spread of a tumor to sites that are physically discontinuous with the primary tumor, an

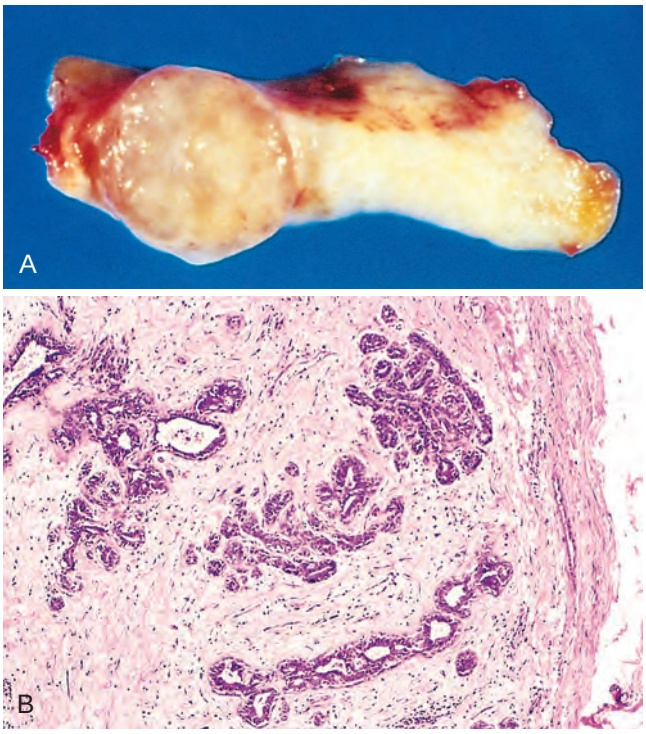


Figure 7.11 Fibroadenoma of the breast. (A) The tan-colored, encapsulated small tumor is sharply demarcated from the whiter breast tissue. (B) Microscopic view shows that the fibrous capsule *(right)* delimits the tumor from the surrounding tissue. (B, Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.)

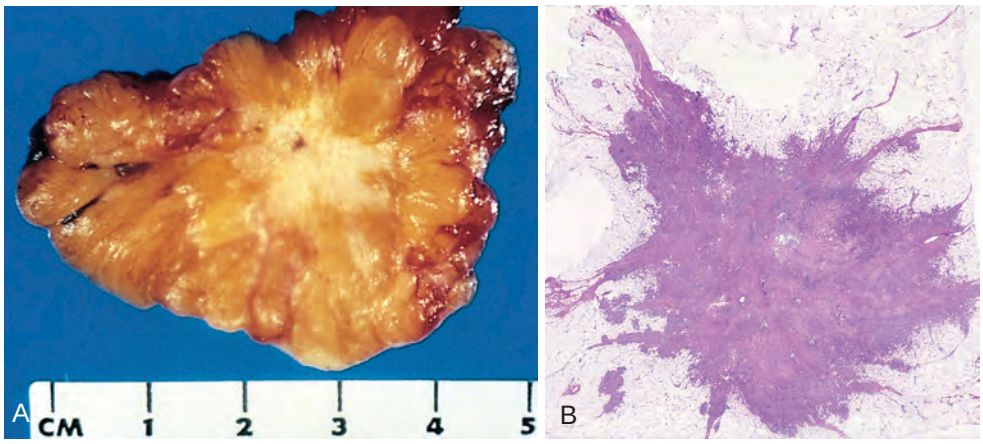


Figure 7.12 Invasive ductal carcinoma of the breast. (A) On cut section, the lesion is retracted and infiltrates the surrounding breast substance and would be stony hard on palpation. (B) Low-power microscopic view shows irregular infiltrative borders without a well-defined capsule and intense stromal reaction. (A, Courtesy Dr. Trace Worrell, University of Texas Southwestern Medical School, Dallas, Tex.; B, Courtesy Dr. Susan Lester, Brigham and Women's Hospital, Boston, Mass.)

event that unequivocally marks a tumor as malignant. The invasiveness of cancers permits them to penetrate blood vessels, lymphatics, and body cavities, providing the opportunity for spread. All malignant tumors can metastasize, but some do so very infrequently. Examples include malignant neoplasms of the glial cells in the central nervous system, called *gliomas,* and basal cell carcinomas of the skin, both of which invade early in their course but rarely metastasize. It is evident then that the properties of invasiveness and metastasis are separable. Blood cancers (leukemias and lymphomas) are a special case. These tumors are derived from cells that normally have the capacity to enter the bloodstream and travel to distant sites. As a result, leukemias and lymphomas (sometimes referred to as "liquid tumors") are often disseminated at diagnosis and are always taken to be malignant, unlike all other tumors (so-called "solid" tumors), which are derived from cells that do not normally circulate in the bloodstream.

Overall, approximately 30% of solid tumors (excluding skin cancers other than melanomas) present as metastatic disease. In general, the likelihood of metastasis of a solid tumor correlates with other features of malignancy, including lack of differentiation, aggressive local invasion, rapid growth, and large size. There are innumerable exceptions, however. Small, well-differentiated, slow-growing lesions sometimes metastasize widely; conversely, some rapidly growing, large lesions remain localized for years. Metastasis is thus a complex and unpredictable process that involves many factors relating to both invader and host (discussed later). Metastatic spread strongly reduces the possibility of cure; hence, short of prevention of cancer, no achievement would be of greater benefit to patients than an effective means to block metastasis, with the important caveat that many tumors that kill the patient have already spread by the time of initial diagnosis.

Pathways of Spread

Dissemination of cancers occurs through three pathways: (1) direct seeding of body cavities or surfaces, (2) lymphatic spread, and (3) hematogenous spread. Although iatrogenic spread of tumor cells on surgical instruments may occur—it is the reason, for example, why biopsies of testicular masses are never done—it is rare and not discussed further.

Seeding of Body Cavities and Surfaces. Seeding of body cavities and surfaces occurs when a malignant neoplasm penetrates into a natural "open field" lacking physical barriers. Most often involved is the peritoneal cavity (Fig. 7.13), but any body cavity—pleural, pericardial, subarachnoid, and joint spaces—may be affected. Such seeding is particularly characteristic of carcinomas arising in the ovaries, which often spread to peritoneal surfaces, producing a heavy cancerous coating. Remarkably, the tumor cells may remain confined to the surface of the abdominal viscera without penetrating into the substance. Sometimes, mucus-secreting

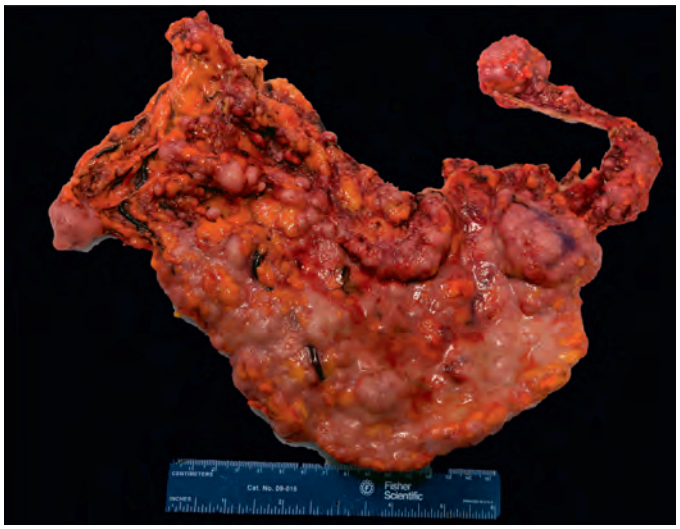


Figure 7.13 Involvement of omentum by metastatic ovarian carcinoma. Innumerable nodules and more subtle "glazing" are evident due to seeding by carcinoma cells via the peritoneal cavity. (Courtesy Dr. Sarah Hill, Brigham and Women's Hospital, Boston, Mass.)

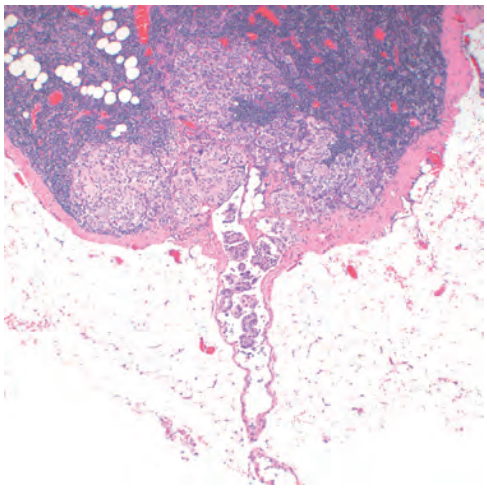


Figure 7.14 Axillary lymph node with metastatic breast carcinoma. Note the aggregates of tumor cells within the substance of the node and the dilated lymphatic channel. (Courtesy Dr. Susan Lester, Brigham and Women's Hospital, Boston, Mass.)

appendiceal carcinomas or ovarian carcinomas fill the peritoneal cavity with a gelatinous neoplastic mass referred to as *pseudomyxoma peritonei*.

Lymphatic Spread. **Transport through lymphatic vessels is the most common pathway for the initial dissemination of carcinomas** (Fig. 7.14). Sarcomas also sometimes use this route. Tumors do not contain functional lymphatic vessels, but lymphatic vessels located at the margins of invading cancers are apparently sufficient for the lymphatic spread of tumor cells. The pattern of spread follows the natural routes of lymphatic drainage. For example, because carcinomas of the breast usually arise in the upper outer quadrants, they generally disseminate first to the axillary lymph nodes and then to infraclavicular and supraclavicular lymph nodes. Carcinomas of the lung arising in the major respiratory passages metastasize first to perihilar tracheobronchial and mediastinal lymph nodes. Local lymph nodes, however, may be bypassed—so-called skip metastasis—possibly because microscopic metastases are missed or because of variation in normal patterns of lymphatic drainage.

In breast cancer, axillary lymph node examination is used to determine the prognosis and select the most suitable therapeutic options. To avoid the surgical morbidity associated with a full axillary lymph node dissection, biopsy of sentinel nodes is often used to assess the presence or absence of metastatic lesions. **A sentinel lymph node is defined as "the first node in a regional lymphatic basin that receives lymph flow from the primary tumor."** Sentinel node mapping can be done by injection of radiolabeled tracers or colored dyes, and examination of frozen sections of the sentinel lymph node performed during surgery can guide the surgeon to the appropriate therapy. Sentinel node examination has also been used to assess the spread of melanomas, colon cancers, and other tumors.

In many cases the regional nodes serve as effective barriers against further dissemination of the tumor, at least for a while. Conceivably, after arrest within the node the cells may be destroyed by a tumor-specific immune response. The immune response to tumor cells or antigens in draining lymph nodes may lead to enlargement (hyperplasia) of the nodes. Thus, enlarged lymph nodes do not always harbor metastases, which can be assessed definitively only by microscopic examination.

Hematogenous Spread. **Hematogenous spread is typical of sarcomas but is also seen with carcinomas**. In general, histologic evidence of penetration of small vessels at the site of the primary neoplasm is an ominous feature associated with hematogenous metastasis. The involved vessels are usually small veins, as arteries, with their thicker walls, are more resistant to penetration. Arterial spread may occur, however, when tumor cells pass through pulmonary capillary beds or pulmonary arteriovenous shunts or when cancers in the lung (primary or metastatic) give rise to tumor emboli.

Several factors influence the patterns of vascular metastasis. With venous invasion, the blood-borne tumor cells often come to rest in the first capillary bed they encounter. Understandably, the liver and the lungs (Fig. 7.15) are most frequently involved by hematogenous dissemination because all portal area drainage flows to the liver and all caval blood

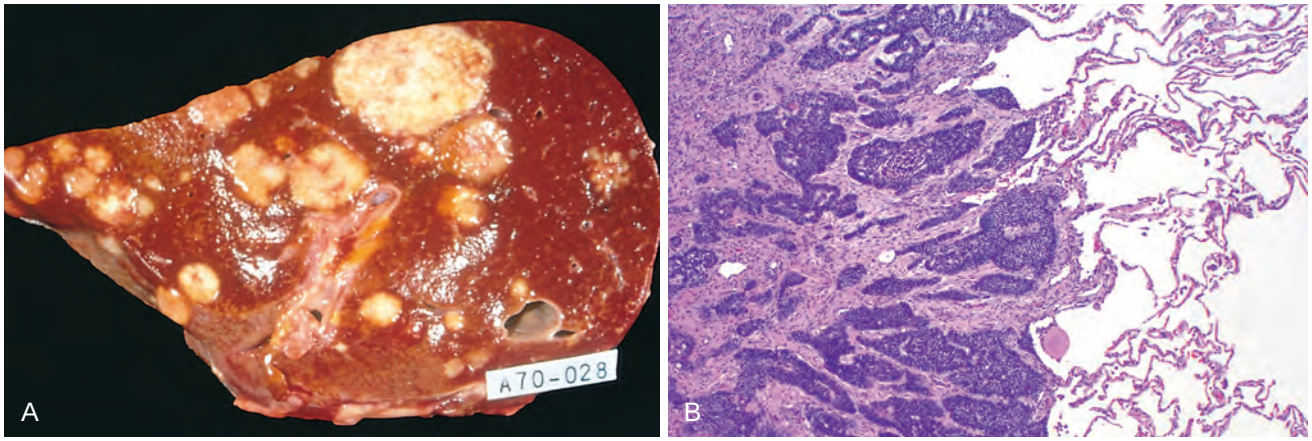


Figure 7.15 Cancer metastasis. (A) Liver studded with metastatic cancer. (B) Microscopic view of lung metastasis. A colonic adenocarcinoma has formed a metastatic nodule in the lung. (B, Courtesy Dr. Shuji Ogino, Dana Farber Cancer Institute, Boston, Mass.)

Table 7.2 Comparisons Between Benign and Malignant Tumors

Characteristics	Benign	Malignant
Differentiation/ anaplasia	Well differentiated; structure sometimes typical of tissue of origin	Some lack of differentiation (anaplasia); structure often atypical
Rate of growth	Usually progressive and slow; may come to a standstill or regress; mitotic figures rare and normal	Erratic, may be slow to rapid; mitotic figures may be numerous and abnormal
Local invasion	Usually cohesive, expansile, well-demarcated masses that do not invade or infiltrate surrounding normal tissues	Locally invasive, infiltrating surrounding tissue; sometimes may be misleadingly cohesive and expansile
Metastasis	Absent	Frequent; more likely with large undifferentiated primary tumors

flows to the lungs. Cancers arising in close proximity to the vertebral column often embolize through the paravertebral plexus; this pathway produces frequent vertebral metastases from carcinomas of the thyroid and prostate. Nonetheless, many observations suggest that the location of the primary tumor and its natural pathways of venous drainage do not wholly explain the observed patterns of metastatic spread, which are often cancer-specific. Unfortunately, most cancers have not read pathology textbooks! The basis of tissue-specific patterns of metastasis is discussed later.

Certain cancers have a curious propensity for growth within large veins. Renal cell carcinoma often invades the branches of the renal vein and then the renal vein itself, growing in a snakelike fashion up the inferior vena cava until it sometimes reaches the right side of the heart. Similarly, hepatocellular carcinomas often penetrate portal and hepatic radicles and then grow into the main venous channels. Remarkably, such intravenous growth may not be accompanied by widespread metastasis.

The distinguishing features of benign and malignant tumors are summarized in Table 7.2 and Fig. 7.16. Having completed our overview of the morphology and behavior of neoplasms, we now discuss the pathogenesis of neoplasia, starting with clues gleaned from studies of the epidemiology of cancer.

KEY CONCEPTS

CHARACTERISTICS OF BENIGN AND MALIGNANT NEOPLASMS

- Benign and malignant tumors can be distinguished from one another based on the degree of differentiation, local invasiveness, and distant spread.
- Benign tumors resemble the tissue of origin and are well differentiated; malignant tumors are less well differentiated or completely undifferentiated (anaplastic).

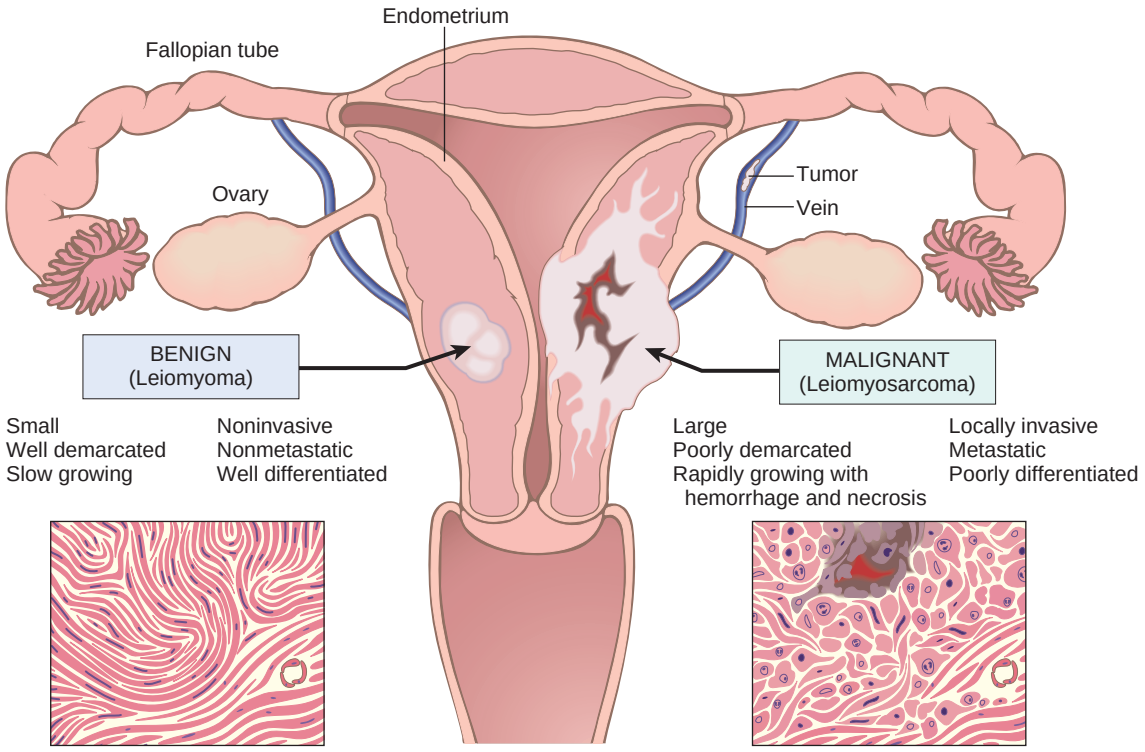


Figure 7.16 Comparison between a benign tumor of the myometrium (leiomyoma) and a malignant tumor of the same origin (leiomyosarcoma).

- Benign tumors are more likely to retain functions of their cells of origin, whereas malignant tumors sometimes acquire unexpected functions due to derangements in differentiation.
- Benign tumors are slow growing, while malignant tumors generally grow faster.
- Benign tumors are circumscribed and have a capsule; malignant tumors are poorly circumscribed and invade surrounding normal tissues.
- Benign tumors remain localized at the site of origin; malignant tumors metastasize to distant sites. Carcinomas tend to spread via lymphatics, whereas sarcomas prefer the hematogenous route.

EPIDEMIOLOGY OF CANCER

Study of cancer in defined populations has contributed substantially to knowledge about its origins. Epidemiologic studies have established the causative link between smoking and lung cancer, and comparison of diet and cancer rates in different regions of the world has linked diets high in fat and low in fiber to colon cancer. It is hoped that additional insights into the causes of cancer will be obtained by studies that relate particular environmental, racial (possibly hereditary), and cultural influences to specific neoplasms. The strong association of certain inflammatory and other diseases with cancer also provides clues to its pathogenesis. In the following sections, we discuss the overall incidence of cancer and then review environmental and host factors that influence the predisposition to cancer.

The Global Impact of Cancer

In 2018 it was estimated that there were over 9.5 million deaths caused by cancer worldwide, representing nearly 1 in 6 of all deaths. Moreover, due to increasing population size and age, cancer cases and cancer-related deaths worldwide are projected to increase to 21.4 million and 13.2 million, respectively by the year 2030. The major organ sites affected and the estimated frequency of cancer deaths in the United States are shown in Fig. 7.17. The most common tumors in men arise in the prostate, lung, and colon/rectum. In women, cancers of the breast, lung, and colon/rectum are the most frequent. Cancers of the lung, female breast, prostate, and colon/rectum constitute more than 50% of cancer diagnoses and cancer deaths in the United States.

Most longitudinal data pertaining to cancer incidence come from higher income countries, where age-adjusted death rates (deaths per 100,000 population) for many cancers have changed significantly over the years. In the last 50 years of the 20th century, the age-adjusted cancer death rate increased significantly in both men and women. However, since 1995 the cancer incidence rate in men has been stable, and since 1990 the cancer death rate has decreased by approximately 20%. Similarly, the cancer incidence rate also stabilized in women in 1995, and the cancer death rate has fallen by approximately 10% since 1991. Among men, nearly 80% of the decrease is accounted for by lower death rates from lung, prostate, and colorectal cancers; among women, nearly 60% of the decrease is due to reductions in death rates from breast and colorectal cancers. Decreased use of tobacco products is responsible for the reduction in lung cancer deaths, while improved detection and treatment are responsible for the decrease in death rates for colorectal, female breast, and prostate cancer.

The last half-century has also seen a sharp decline in deaths caused by cervical cancer in the United States. This decrease is largely attributable to the Papanicolaou (Pap) smear test, which enables detection of "precursor lesions" (discussed later) and early, curable cancers. By contrast,

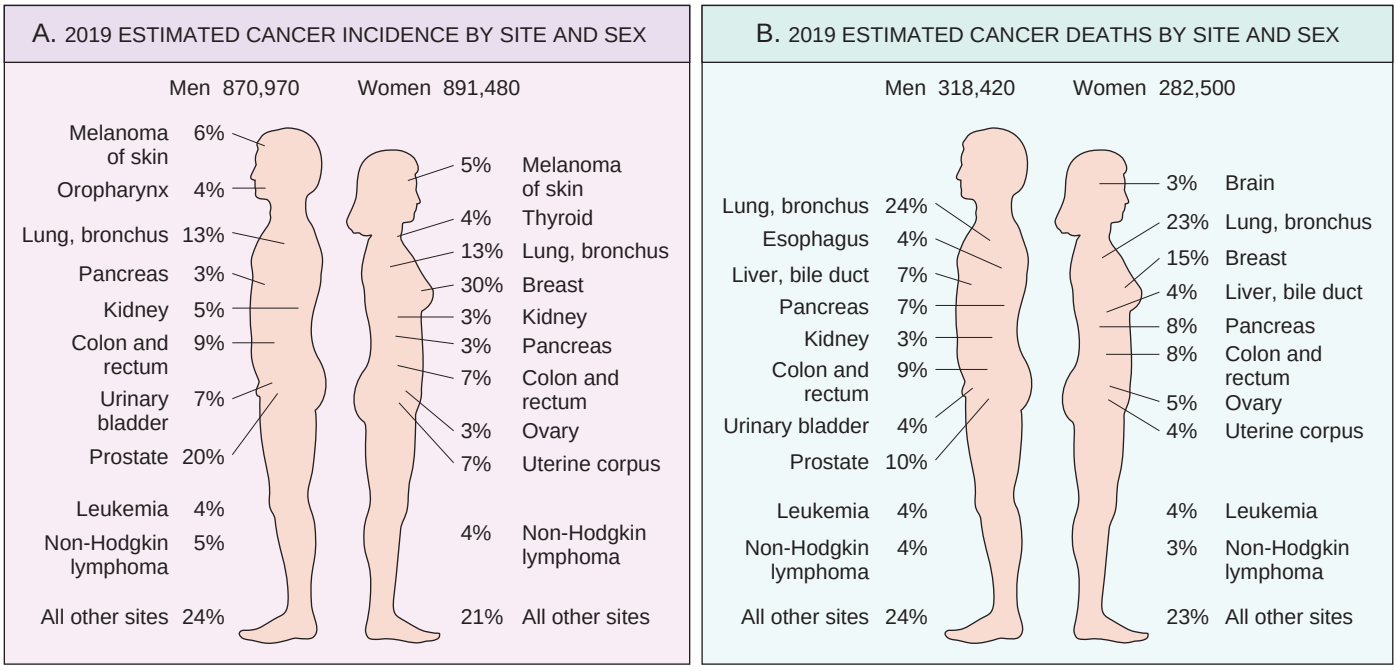


Figure 7.17 Cancer incidence (A) and mortality (B) by site and sex. Excludes basal cell and squamous cell skin cancers and in situ carcinomas except urinary bladder. (Modified from Siegel RL, Miller KD, Jemal A: Cancer statistics, 2017, *CA Cancer J Clin* 67:7–30, 2017.)

between 1990–1991 and 2004, lung cancer death rates in women and liver and intrahepatic bile duct cancer death rates in men increased substantially, offsetting some of the improvement in survival from other cancers. Indeed, although carcinomas of the breast occur about 2.5 times more frequently than those of the lung in women, lung cancer now causes more deaths in women.

Race is not a discrete biologic variable, but it can define groups at risk for certain cancers. The disparity in cancer mortality rates between Americans who are Caucasian or of African descent persists, but African Americans had the largest decline in cancer mortality during the past decade. People identifying as Hispanic living in the United States have a lower frequency of the most common cancers affecting the Caucasian non-Hispanic population and a higher incidence of cancers of the stomach, liver, uterine cervix, and gallbladder as well as certain leukemias.

Environmental Factors

Although both genetic and environmental factors contribute, environmental influences are the dominant risk factors for most cancers. One line of evidence supporting this idea comes from longitudinal changes in cancer incidence in the United States. Examples include the tight tracking of lung cancer incidence with changes in smoking habits over time; the sharp drop in stomach cancer incidence during the 20th century, believed to stem from decreased exposure to unknown environmental carcinogens; and a recent increase in the incidence of liver cancer, which is likely due to rising rates of chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infection and obesity. Other evidence is found in the wide geographic variation that exists in the incidence of specific cancers (Fig. 7.18). For example, the most common cancer of men in the United States and in most other higher income countries is prostate cancer, but in other countries or regions, cancers of the liver, stomach, esophagus, bladder, lung, oropharynx, and the immune system rise to the top of the list. Similarly, the incidence of breast cancer is generally much higher in women living in higher income countries than in lower income countries. Although racial predisposition may factor in, it is believed that environmental influences—some known, some not—underlie most of these differences in cancer incidence.

Among the best-established environmental factors affecting cancer risk are the following.

- *Infectious agents.* About 15% of all cancers worldwide are caused directly or indirectly by infectious agents, with the burden of cancers linked to infections being roughly three times higher in the developing world than in the developed world. For example, *human papillomavirus (HPV),* an agent spread through sexual contact, has a causative role in the majority of cervical carcinomas and an increasing fraction of head and neck cancers. Specific infectious agents and their associated cancers are discussed later in this chapter.
- *Smoking.* Cigarette smoking is the single most important environmental factor contributing to premature death in the United States. Smoking, particularly of cigarettes, is implicated in cancer of the mouth, pharynx, larynx, esophagus, pancreas, and bladder and, most significantly, in about 90% of lung cancers (Chapter 9).
- *Alcohol consumption.* Alcohol abuse increases the risk of carcinoma of the oropharynx (excluding lip), larynx, and esophagus and, by the development of alcoholic cirrhosis, hepatocellular carcinoma. Moreover, the risk of cancers in the upper airways and digestive tract imposed by alcohol is increased synergistically when combined with tobacco use.
- *Diet.* Although the precise dietary factors that affect cancer risk remain a matter of debate, wide geographic variation in the incidences of colorectal carcinoma, prostate carcinoma, and breast carcinoma has been ascribed to differences in diet.
- *Obesity.* Given that the obesity epidemic in the United States is spreading to other parts of the world (Chapter 9), it is concerning that obesity is associated with increased cancer risk. The most overweight individuals in the U.S. population have 52% (men) to 62% (women) higher death rates from cancer than do their slimmer counterparts; it follows that approximately 14% of cancer deaths in men and 20% in women are associated with obesity.
- *Reproductive history.* Lifelong cumulative exposure to estrogen stimulation, particularly if unopposed by progesterone, increases the risk of cancers of the breast and endometrium, tissues that are responsive to these hormones. It is likely that some of the geographic variation in breast cancer incidence is related to differing cultural mores that influence the timing and number of pregnancies a woman has during her lifetime.
- *Environmental carcinogens.* There is no paucity of well-characterized environmental carcinogens: they lurk in the ambient environment, in the workplace (Table 7.3), in food, and in personal practices. Individuals may be exposed to carcinogenic factors when they go outside (e.g., ultraviolet [UV] rays, smog), drink well water (e.g., arsenic, particularly in Bangladesh), take certain medications (e.g., methotrexate), go to work (e.g., asbestos), or lounge at home (e.g., grilled meat, high-fat diet, alcohol).

It appears that almost everything one does to earn a livelihood or for pleasure is fattening, immoral, illegal, or, even worse, carcinogenic!

Age

Age has an important influence on the risk of cancer. Most carcinomas occur in adults older than 55 years of age. Cancer is the leading cause of death among women aged 40 to 79 and among men aged 60 to 79; the decline in cancer deaths after age 80 is due to the lower number of individuals who reach this age. The rising incidence of cancer with age is likely explained by the accumulation of somatic mutations that accompanies the aging of cells (discussed later). A decline in immune competence in older individuals may also be a factor.

Tragically, children are not spared; cancer accounts for approximately 10% of all deaths in children younger than age 15 in the United States, second only to accidents. However, the types of cancers that predominate in children are different from those seen in adults; in part, this is because pediatric cancers are more likely to be caused by inherited mutations (particularly in tumor suppressor genes, described later) and much less likely to stem from exposure to

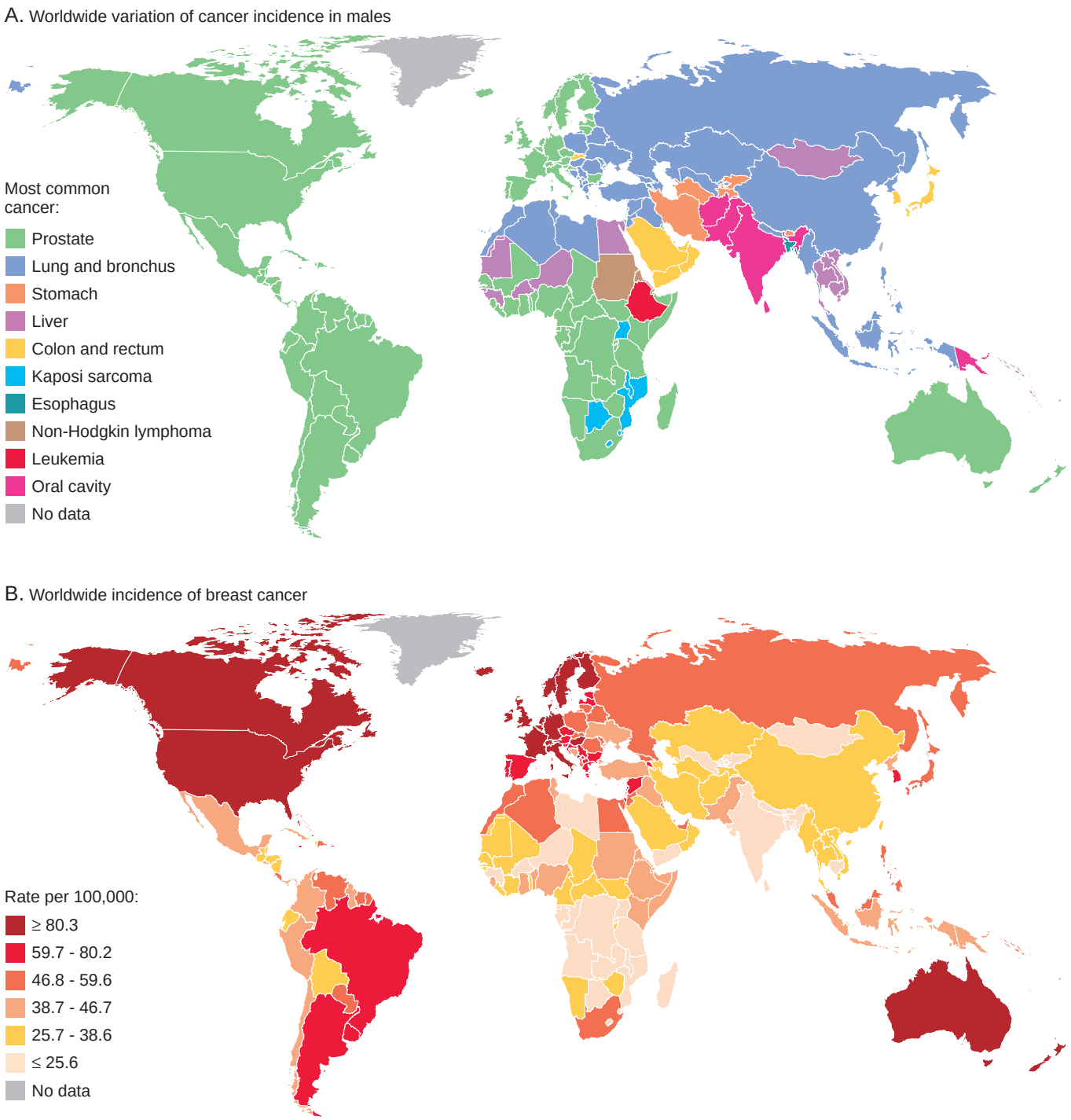


Figure 7.18 Geographic variation in cancer incidence. (A) Most common cancers in men by country. (B) Variation in breast cancer incidence in women by country. (Modified from American Cancer Society: *Global Cancer Facts & Figures*, ed 3, Atlanta, 2015, American Cancer Society.)

environmental carcinogens (e.g., cigarette smoking). This difference explains why carcinomas, which are frequently caused by carcinogens and are the most common general type of tumor in adults, are very rare in children. Instead, acute leukemia and distinctive neoplasms of the central nervous system cause approximately 60% of childhood cancer deaths. The common neoplasms of infancy and childhood include the so-called small round blue cell tumors such as neuroblastoma, Wilms tumor, retinoblastoma, acute lymphoblastic leukemia, and rhabdomyosarcoma. These are discussed in Chapter 10 and elsewhere in the text.

Acquired Predisposing Conditions

Acquired conditions that predispose to cancer can be divided into chronic inflammatory disorders, precursor

Table 7.3 Occupational Cancers

Agents or Groups of Agents	Human Cancers for Which Reasonable Evidence Is Available	Typical Use or Occurrence
Arsenic and arsenic compounds	Lung carcinoma, skin carcinoma	By-product of metal smelting; component of alloys, electrical and semiconductor devices, medications and herbicides, fungicides, and animal dips
Asbestos	Lung, esophageal, gastric, and colon carcinoma; mesothelioma	Formerly used for many applications because of fire, heat, and friction resistance; still found in existing construction as well as fire-resistant textiles, friction materials (i.e., brake linings), underlayment and roofing papers, and floor tiles
Benzene	Acute myeloid leukemia	Principal component of light oil; despite known risk, many applications exist in printing and lithography, paint, rubber, dry cleaning, adhesives and coatings, and detergents; formerly widely used as solvent and fumigant
Beryllium and beryllium compounds	Lung carcinoma	Missile fuel and space vehicles; hardener for lightweight metal alloys, particularly in aerospace applications and nuclear reactors
Cadmium and cadmium compounds	Prostate carcinoma	Uses include yellow pigments and phosphors; found in solders; used in batteries and as alloy and in metal platings and coatings
Chromium compounds	Lung carcinoma	Component of metal alloys, paints, pigments, and preservatives
Nickel compounds	Lung and oropharyngeal carcinoma	Nickel plating; component of ferrous alloys, ceramics, and batteries; by-product of stainless-steel arc welding
Radon and its decay products	Lung carcinoma	From decay of minerals containing uranium; potentially serious hazard in quarries and underground mines
Vinyl chloride	Hepatic angiosarcoma	Refrigerant; monomer for vinyl polymers; adhesive for plastics; formerly inert aerosol propellant in pressurized containers

Modified from Stellman JM, Stellman SD: Cancer and workplace, *CA Cancer J Clin* 46:70, 1996.

lesions, and immunodeficiency states. Chronic inflammatory disorders and precursor lesions span a diverse set of conditions that are all associated with increased cellular replication, which appears to create a "fertile" soil for the development of malignant tumors. Indeed, repeated rounds of cell division may be required for neoplastic transformation, as proliferating cells are most at risk for somatic mutations that lead to carcinogenesis. Immunodeficiency states predispose to virus-induced cancers. Each of these acquired predisposing conditions is described next.

- *Chronic inflammation.* Virchow first proposed a cause-and-effect relationship between chronic inflammation and cancer in 1863. The scope of this association is now clear; cancer risk is increased in individuals affected by a wide variety of chronic inflammatory diseases, both infectious and noninfectious (Table 7.4). Tumors arising in the context of chronic inflammation are mostly carcinomas, but also include mesothelioma and several kinds of lymphoma. As with any cause of tissue injury, these disorders are accompanied by a compensatory proliferation of cells that serves to repair the damage. In some cases, chronic inflammation may increase the pool of tissue stem cells, which may be particularly susceptible to transformation. Additionally, activated immune cells produce reactive oxygen species that may damage DNA and inflammatory mediators that may promote cell survival, even in the face of genomic damage. Whatever the precise mechanism, the link between chronic inflammation and cancer has practical implications. For instance, diagnosis and effective treatment of *Helicobacter pylori* gastritis with antibiotics can quell a chronic inflammatory condition that might otherwise lead to the development of a gastric cancer.
- *Precursor lesions.* Precursor lesions are defined by localized morphologic changes that identify a field of epithelium that is at increased risk for malignant transformation. These changes may take the form of hyperplasia, metaplasia, or dysplasia. The link between epithelial dysplasia and metaplasia with various forms of carcinoma has already been mentioned. Precursor lesions consisting of hyperplasias often stem from chronic exposure to trophic factors. One of the most common precursors of this type is *endometrial hyperplasia,* which is caused by sustained estrogenic stimulation of the endometrium. Other "at risk" lesions consist of benign neoplasms. A classic lesion of this type is the colonic *villous adenoma,* which progresses to cancer in about 50% of cases if left untreated. It should be emphasized, however, that most benign tumors transform rarely (e.g., uterine leiomyomas, pleomorphic adenoma) and others not at all (e.g., lipomas). Why most benign tumors have a negligible risk of malignant transformation is an unsettled question; one possibility is that benign tumors at high risk for malignant transformation possess the cancer-enabling property of genomic instability (discussed later), whereas other benign tumors do not.
- *Immunodeficiency and cancer.* Patients who are immunodeficient, particularly those with deficits in T-cell immunity, are at increased risk for cancer, especially types caused by oncogenic viruses, presumably because

Table 7.4 Chronic Inflammatory States and Cancer

Pathologic Condition	Associated Neoplasm(s)	Etiologic Agent(s)
Asbestosis, silicosis	Mesothelioma, lung carcinoma	Asbestos fibers, silica particles
Inflammatory bowel disease	Colorectal carcinoma	
Lichen sclerosis	Vulvar squamous cell carcinoma	
Pancreatitis	Pancreatic carcinoma	Alcoholism, germline mutations (e.g., in the trypsinogen gene)
Chronic cholecystitis	Gallbladder cancer	Bile acids, bacteria, gallbladder stones
Reflux esophagitis, Barrett esophagus	Esophageal carcinoma	Gastric acid
Sjögren syndrome, Hashimoto thyroiditis	MALT lymphoma	
Opisthorchis, cholangitis	Cholangiocarcinoma, colon carcinoma	Liver flukes (*Opisthorchis viverrini*)
Gastritis/ulcers	Gastric adenocarcinoma, MALT lymphoma	*Helicobacter pylori*
Hepatitis	Hepatocellular carcinoma	Hepatitis B and/or C virus
Osteomyelitis	Carcinoma in draining sinuses	Bacterial infection
Chronic cervicitis	Cervical carcinoma	Human papillomavirus
Chronic cystitis	Bladder carcinoma	Schistosomiasis

MALT, Mucosa-associated lymphoid tissue.

Modified from Tlsty TD, Coussens LM: Tumor stroma and regulation of cancer development, *Ann Rev Pathol Mech Dis* 1:119, 2006.

these individuals have a higher than normal incidence of chronic infection with viruses. These virus-associated tumors include lymphomas, certain carcinomas, and some sarcomas and sarcoma-like proliferations. The relationship between infections, immunity, and cancer is discussed later in this chapter.

Genetic Predisposition and Interactions Between Environmental and Inherited Factors

In some families, cancer is an inherited trait, usually due to germline mutations in a tumor suppressor gene (described later). What then can be said about the influence of heredity on sporadic malignant neoplasms, which constitute roughly 95% of cancers in the United States? While the evidence suggests that these cancers are largely attributable to environmental factors or acquired predisposing conditions, lack of family history does not preclude an inherited component. It is generally difficult to sort out hereditary and nonhereditary contributions because their interactions are often complex, particularly when tumor development depends on the action of multiple genes. Even in cancers with a well-defined inherited component, the risk of cancer development may be greatly influenced by nongenetic factors. For instance, breast cancer risk in females who inherit mutated copies of the *BRCA1* or *BRCA2* tumor suppressor genes (discussed later) has been observed to be almost threefold higher for women born after 1940 than for women born before that year, perhaps because of changes in reproductive history. Conversely, genetic factors can alter the likelihood of cancers that are primarily induced by environmental carcinogens. This is because genetic variation (polymorphisms) in certain enzymes, such as the cytochrome P-450 system, influences the conversion of procarcinogens to active carcinogens. A cardinal example, discussed later, is a polymorphism in one of the P-450 genes that confers susceptibility to smoking-induced lung cancer.

KEY CONCEPTS

EPIDEMIOLOGY OF CANCER

- The incidence of cancer varies with geography, age, race, and genetic background. Cancers are most common in adults older than 55 years of age, but occur in adults at all ages and in children and infants. The geographic variation is thought to mainly stem from different environmental exposures.
- Important environmental factors implicated in carcinogenesis include infectious agents, smoking, alcohol, diet, obesity, reproductive history, and exposure to environmental carcinogens.
- The risk of cancer is increased by reparative proliferations caused by chronic inflammation or tissue injury, certain forms of hyperplasia, and immunodeficiency.
- Interactions between environmental factors and genetic factors may be important determinants of cancer risk.

MOLECULAR BASIS OF CANCER: ROLE OF GENETIC AND EPIGENETIC ALTERATIONS

Evidence for the genetic origins of cancer has been building for decades. However, a full accounting of the extent of these genetic aberrations is only now nearing completion, made possible by technologic advances in DNA sequencing and other methods that permit genome-wide analysis of cancer cells. The complexity of these data is daunting, and the messages hidden within them have yet to be fully decoded, but certain "genomic themes" have emerged that are likely relevant to every cancer.

- *Nonlethal genetic damage lies at the heart of carcinogenesis.* The initial damage (or mutation) may be caused by environmental exposures, may be inherited in the germline, or may be spontaneous and random, falling into the category of "bad luck." The term environmental, used in this context, refers to any acquired mutation caused by exogenous agents such as viruses or environmental chemicals or by endogenous products of cellular metabolism that have the potential to damage DNA (such as reactive oxygen species) or alter gene expression through epigenetic mechanisms (e.g., so-called oncometabolites, described later).

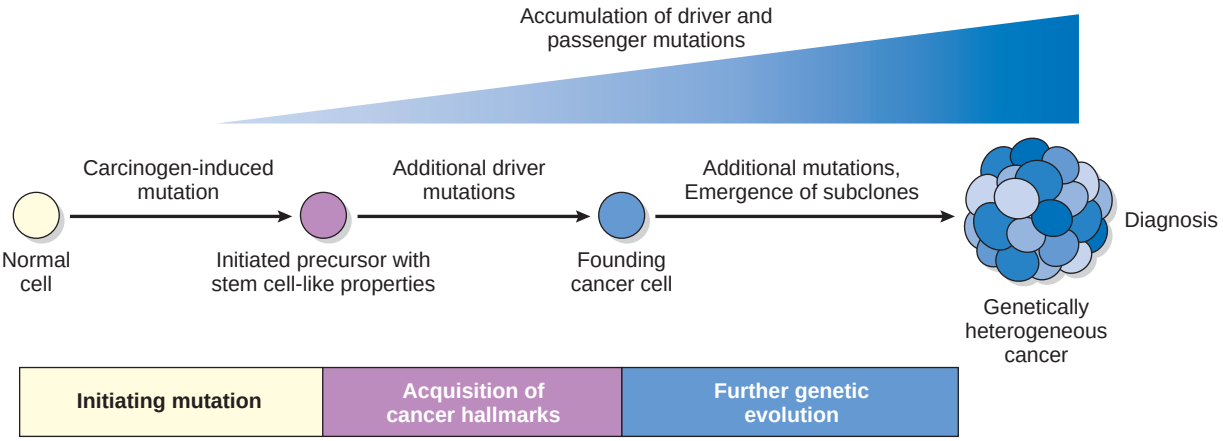


Figure 7.19 Development of a cancer through stepwise acquisition of complementary mutations. The order in which various driver mutations occur in initiated precursor cells is not known and may vary from tumor to tumor. See text for details.

- *A tumor is formed by the clonal expansion of a single precursor cell that has incurred genetic damage (i.e., tumors are clonal).* Alterations in DNA are heritable, being passed to daughter cells, and thus all cells within an individual tumor share the same set of mutations that were present at the moment of transformation. Such tumor-specific mutations are most often identified by DNA sequencing (e.g., point mutations) or by chromosomal analyses (e.g., chromosomal translocations and copy number changes, discussed later).
- *Four classes of genes—growth-promoting proto-oncogenes, growth-inhibiting tumor suppressor genes, genes that regulate programmed cell death (apoptosis), and genes that are responsible for DNA repair—are the principal targets of cancer-causing mutations.* Mutations that activate proto-oncogenes may either cause an increase in one or more normal functions of the encoded gene product that promote tumorigenesis or the appearance of a completely new function that is oncogenic. Because these mutations cause a "gain of function," they can transform cells despite the presence of a normal copy of the same gene. Thus, in genetic parlance, oncogenes are dominant over their normal counterparts. Mutations that affect tumor suppressor genes generally cause a "loss of function," and in most instances both alleles must be damaged before transformation can occur. As a result, mutated tumor suppressor genes usually behave in a recessive fashion. However, there are exceptions: sometimes loss of only a single tumor suppressor gene allele (a state termed haploinsufficiency) reduces the quantity of the encoded protein enough to release the brakes on cell proliferation and survival. Such a finding indicates that two "doses" of the gene are essential for normal function. Apoptosis-regulating genes may acquire abnormalities that result in less cell death and therefore enhanced survival. These abnormalities include gain-of-function mutations in genes whose products suppress apoptosis and loss-of-function mutations in genes whose products promote cell death. Loss-of-function mutations affecting DNA repair genes contribute to carcinogenesis indirectly by impairing the ability of the cell to recognize and repair nonlethal genetic damage in other genes. As a result, affected cells acquire mutations at an accelerated rate, a state referred to as a *mutator phenotype* that is marked by *genomic instability.*
- *Carcinogenesis results from the accumulation of complementary mutations in a stepwise fashion over time* (Fig. 7.19). Malignant neoplasms have several phenotypic attributes referred to as *cancer hallmarks* (discussed in detail later), such as excessive growth, local invasiveness, and the ability to form distant metastases, which stem from genomic alterations that change the expression and function of key genes and thereby impart a malignant phenotype.
  - *Mutations that contribute to the acquisition of cancer hallmarks are referred to as driver mutations.* The first driver mutation that starts a cell on the path to malignancy is the *initiating mutation,* which is typically maintained in all the cells of the subsequent cancer. However, because no single mutation appears to be fully transforming, development of a cancer requires that the "initiated" cell acquire a number of additional driver mutations, each of which also contributes to the development of the cancer. The time over which this occurs is unknown in most cancers, but appears to be lengthy; even in aggressive cancers that clinically seem to appear "out of the blue," such as childhood acute lymphoblastic leukemia, cells bearing initiating mutations may be found in blood samples taken as long as a decade before diagnosis. The persistence of initiated cells during this long preclinical prodrome is consistent with the idea that cancers arise from cells with stem cell–like properties, so-called *cancer stem cells,* that have a capacity for self-renewal and long-term persistence.
  - *Loss-of-function mutations in genes that maintain genomic integrity appear to be a common early step on the road to malignancy, particularly in solid tumors.* Mutations that lead to genomic instability not only increase the likelihood of acquiring driver mutations, but also greatly increase the frequency of mutations that have no phenotypic consequence, so-called *passenger mutations,* which are much more common than driver mutations. As a result, by the time a cell acquires all of the driver mutations that are needed for malignant behavior, it may bear hundreds or even thousands of passenger mutations.

- Mutations in many other genes contribute to tumorigenesis by interfering with host immune responses or altering interactions with the stroma, or by other mechanisms. By convention, these are not classified under driver and passenger mutations, since the terms are largely restricted to genes that influence the behavior of the cells in a cell-intrinsic manner.

Once established, tumors evolve genetically during their outgrowth and progression under the pressure of Darwinian selection (survival of the fittest). Early on, all the cells in a tumor are genetically identical, being the progeny of a single founding transformed cell. However, by the time a tumor comes to clinical attention (generally when it attains a mass of about 1 g, or about 10^9 cells), it has gone through a minimum of 30 cell doublings (this number is actually a substantial underestimation because a fraction of cells in all tumors dies by apoptosis during preclinical stages of tumor development). During the expansion process, individual tumor cells acquire additional mutations at random; this is particularly true in tumors with driver mutations conferring a mutator phenotype. As a result of this tumor evolution, even though cancers are clonal in origin, by the time they become clinically evident their constituent cells are often extremely heterogeneous genetically (see Fig. 7.19). These diverse tumor subclones compete for access to nutrients and microenvironmental niches, and those that are most fit "win" this Darwinian struggle and come to dominate the tumor mass. This pernicious tendency of tumors to become more aggressive over time is referred to as *tumor progression*.

A skeptical student might well ask, "How do we know that genetically distinct subclones really exist in any particular cancer?" Supportive data have emerged from studies of solid cancers such as renal cell carcinoma, in which multiple regions of the primary tumor and metastatic deposits from the same patient have been subjected to DNA sequencing (Fig. 7.20). As predicted, two types of mutations were identified in these studies: (1) mutations that are present in all tumor sites tested, which were presumably present in the founding cell at the moment of transformation, and (2) mutations that are unique to a subset of tumor sites, which were likely acquired after transformation during the outgrowth and spread of the tumor. This second type of mutation can be used to create tumor "family trees" showing the genetic relationships of various subclones. Remarkably, subclones within tumors appear to diverge genetically in a fashion that is very similar to the manner in which new species are thought to emerge in complex ecosystems; a cardinal example of the latter are the finches on the Galapagos Islands that inspired Darwin, in part, to propose evolution as the origin of the species. In the case of species, this genetic divergence occurs over a period of many millennia, whereas in tumors, subclones may arise and diverge on a timescale of years, months, or even weeks.

Selection of the fittest cells can explain not only the natural history of cancer, but also changes in tumor behavior following therapy. One of the most profound selective pressures that cancer cells face is effective therapy given by treating physicians. Tumors that recur after therapy are almost always found to be resistant if the same treatment is given again, presumably because therapy selects for subclones that, by chance, have a genotype that allows them to survive.

In addition to DNA mutations, epigenetic aberrations also contribute to the malignant properties of cancer cells. Epigenetic modifications include DNA methylation, which tends to silence gene expression, and modifications of histones, the proteins that package DNA into chromatin, which depending on their nature may either enhance or dampen gene expression. The epigenetic state of the cell dictates which genes are expressed, which in turn determines the lineage commitment and differentiation state of both normal and neoplastic cells. Epigenetic modifications are

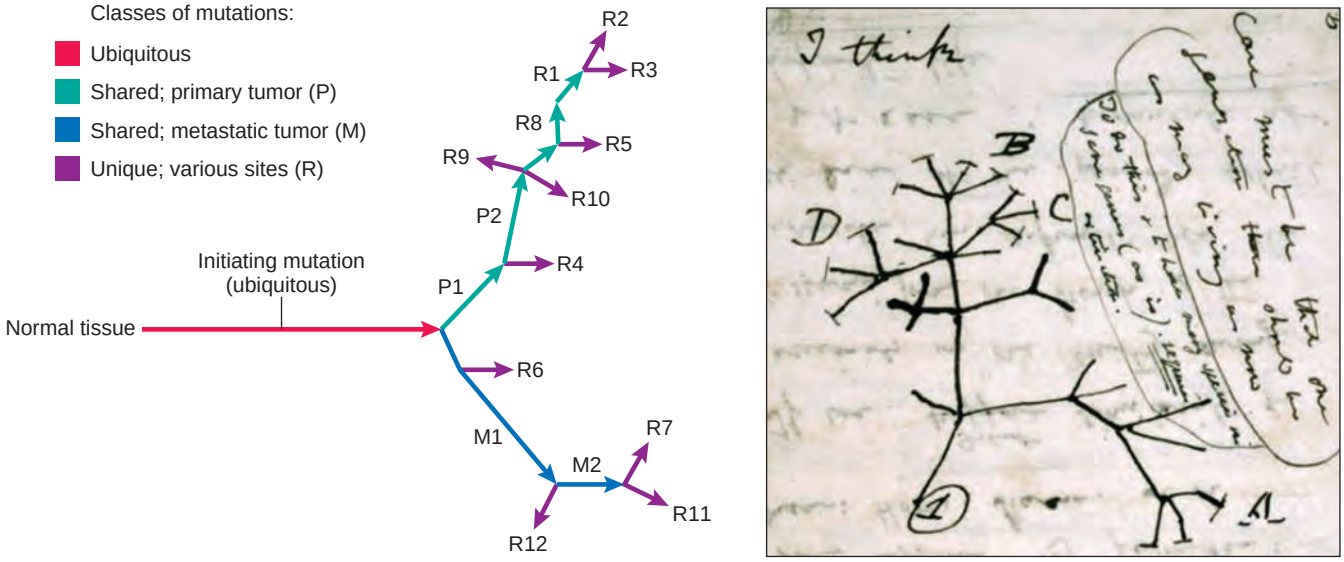


Figure 7.20 Tumor evolution. Evolution of a renal cell carcinoma *(left panel)* and Darwin's finches *(right panel)*. The renal cell carcinoma evolutionary tree is based on genetic comparisons drawn from sequencing of DNA obtained from different tumor sites; the finch evolutionary tree was surmised by Darwin based on morphologic comparisons of different species of finches on the Galapagos Islands. (*right panel*, from Darwin CR: Notebook B: Transmutation of species, 1837–1838, p 26.)

usually passed on faithfully to daughter cells, but on occasion (just as with DNA mutations) alterations may occur that result in changes in gene expression. Aberrant DNA methylation in cancer cells is responsible for the silencing of some tumor suppressor genes, while tumor-specific changes in histone modifications may have far-ranging effects on gene expression (see later). The increasing awareness of the role of epigenetic alterations in cancer has revealed a new path forward for cancer treatment; unlike DNA mutations, epigenetic changes are potentially reversible by drugs that inhibit DNA-modifying or histone-modifying factors. Thus, there is great interest in treating cancers with drugs that correct epigenetic abnormalities.

We will come back to these themes throughout the subsequent discussion, which next turns to the cellular and molecular properties that underlie the malignant behavior of cancer cells.

Cellular and Molecular Hallmarks of Cancer

Over the past several decades, hundreds of genes that are mutated in cancer have been discovered. Traditionally, the functional consequences of these alterations were described one gene at a time. However, the blizzard of mutated genes emerging from the sequencing of cancer genomes has blanketed the landscape and revealed the limitations of trying to grasp the fundamental properties of cancer gene by gene. For example, compilation of a partially complete catalog of recurrent genetic alterations in breast carcinoma required whole genomic sequencing of thousands of tumors and led to the identification of hundreds of distinct driver mutations—and this is just one of hundreds of different kinds of cancer, some of which are substantially more genetically complex than breast carcinoma.

A more tractable and conceptually satisfying way to think about the biology of cancer is to consider the common biologic properties that are imparted to cancer cells by their diverse genomic and epigenomic alterations. It appears that **all cancers display eight fundamental changes in cell physiology, which are considered the hallmarks of cancer.** These changes are illustrated in Fig. 7.21 and consist of the following:

- *Self-sufficiency in growth signals.* Tumors have the capacity to proliferate without external stimuli, usually as a consequence of oncogene activation.
- *Insensitivity to growth-inhibitory signals.* Tumors may not respond to molecules that inhibit the proliferation of normal cells, usually because of inactivation of tumor suppressor genes that encode components of growth inhibitory pathways.
- *Altered cellular metabolism.* Tumor cells undergo a metabolic switch to aerobic glycolysis (called the *Warburg effect*), which enables the synthesis of the macromolecules and organelles that are needed for rapid cell growth.
- *Evasion of apoptosis.* Tumors are resistant to programmed cell death.
- *Limitless replicative potential (immortality).* Tumors have unrestricted proliferative capacity, a stem cell–like property that permits tumor cells to avoid cellular senescence and mitotic catastrophe.
- *Sustained angiogenesis.* Tumor cells, like normal cells, are not able to grow without a vascular supply to bring nutrients and oxygen and remove waste products. Hence, tumors must induce angiogenesis.
- *Ability to invade and metastasize.* Tumor metastases are the cause of the vast majority of cancer deaths and arise from the interplay of processes that are intrinsic to tumor cells and signals that are initiated by the tissue environment.
- *Ability to evade the host immune response.* You will recall that the cells of the innate and adaptive immune system can recognize and eliminate cells displaying abnormal antigens (e.g., a mutated oncoprotein). Cancer cells exhibit a number of alterations that allow them to evade the host immune response.

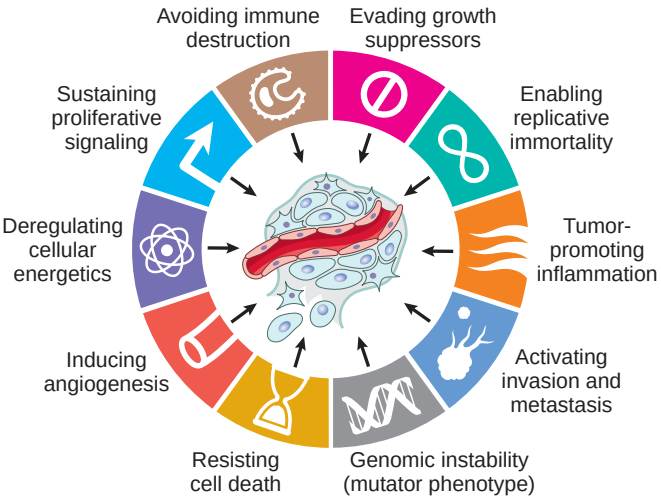


Figure 7.21 Hallmarks of cancer. (Modified from Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation, *Cell* 144:646, 2011.)

The acquisition of the genetic and epigenetic alterations that confer these hallmarks may be accelerated by *genomic instability* and by *cancer-promoting inflammation*. These are considered enabling characteristics because they promote cellular transformation and subsequent tumor progression.

In the following sections, each of the hallmarks and enabling characteristics of cancer cells is discussed, focusing on the most important contributing genes and cellular pathways. The discussion of cancer pathophysiology ends with a review of the roles that epigenetic changes and noncoding RNAs play in the disease.

Self-Sufficiency in Growth Signals: Oncogenes

Oncogenes are mutated genes that cause excessive cell growth, even in the absence of growth factors and other growth-promoting external cues. A major discovery in cancer was that oncogenes are mutated or overexpressed versions of normal cellular genes, which are called proto-oncogenes. Through a variety of mechanisms, discussed later, these mutations increase or alter the function of oncoproteins, which are constitutively active and resistant to control by external signals. Cells expressing oncoproteins are thus freed from normal checkpoints and proliferate excessively.

To aid in the understanding of the nature and functions of oncoproteins and their role in cancer, it is necessary to briefly describe how normal cells respond to growth factors.

Physiologic growth factor–induced signaling can be resolved into the following steps:

- The binding of a growth factor to its specific receptor
- Transient and limited activation of the growth factor receptor, which in turn activates several cytoplasmic signal-transducing proteins
- Transmission of the transduced signal to the nucleus via additional cytoplasmic effector proteins and second messengers or by a cascade of signal transduction molecules
- Induction and activation of transcription factors and epigenetic alterations that initiate and sustain DNA transcription
- Expression of genes and encoded factors that promote entry and progression of the cell into the cell cycle, ultimately resulting in cell division
- In parallel, changes in the expression of other genes that support cell survival and metabolic alterations that are needed for optimal growth

Aberrations in multiple signaling pathways have been identified in neoplasms, and many components of these pathways act as oncoproteins when mutated (Table 7.5). Conversely, many tumor suppressors act by inhibiting one or more components of these same pro-growth pathways (discussed later). In Chapter 1, the major signaling pathways

Table 7.5 Selected Oncogenes, Their Mode of Activation, and Associated Human Tumors

Category	Proto-Oncogene	Mode of Activation in Tumor	Associated Human Tumor
Growth Factors			
PDGF-β	*PDGFB*	Overexpression	Astrocytoma
Fibroblast growth factors	*HST1* *FGF3*	Overexpression Amplification	Osteosarcoma Stomach cancer Bladder cancer Breast cancer Melanoma
TGF-α	*TGFA*	Overexpression	Astrocytomas
HGF	*HGF*	Overexpression	Hepatocellular carcinomas Thyroid cancer
Growth Factor Receptors			
EGF-receptor family	*ERBB1 (EGFR)* *ERBB2 (HER)*	Mutation Amplification	Adenocarcinoma of lung Breast carcinoma
FMS-like tyrosine kinase 3	*FLT3*	Point mutation or small duplications	Leukemia
Receptor for neurotrophic factors	*RET*	Point mutation	Multiple endocrine neoplasia 2A and B, familial medullary thyroid carcinomas
PDGF receptor	*PDGFRB*	Amplification, translocation	Gliomas, leukemias
Receptor for KIT ligand	*KIT*	Point mutation	Gastrointestinal stromal tumors, seminomas, leukemias
ALK receptor	*ALK*	Translocation Point mutation	Adenocarcinoma of lung, certain lymphomas Neuroblastoma
Proteins Involved in Signal Transduction			
GTP-binding (G) proteins	*KRAS* *HRAS* *NRAS* *GNAQ* *GNAS*	Point mutation Point mutation Point mutation Point mutation Point mutation	Colon, lung, and pancreatic tumors Bladder and kidney tumors Melanomas, hematologic malignancies Uveal melanoma Pituitary adenoma, other endocrine tumors
Nonreceptor tyrosine kinase	*ABL*	Translocation	Chronic myelogenous leukemia Acute lymphoblastic leukemia
RAS signal transduction	*BRAF*	Point mutation	Melanomas, leukemias, colon carcinoma, others
Notch signal transduction	*NOTCH1*	Point mutation, translocation	Leukemias, lymphomas, breast carcinoma
JAK/STAT signal transduction	*JAK2*	Point mutation, translocation	Myeloproliferative disorders Acute lymphoblastic leukemia
Nuclear Regulatory Proteins			
Transcriptional activators	*MYC* *NMYC*	Translocation Amplification	Burkitt lymphoma Neuroblastoma
Cell Cycle Regulators			
Cyclins	*CCND1* (cyclin D1)	Translocation Amplification	Mantle cell lymphoma, multiple myeloma Breast and esophageal cancers
Cyclin-dependent kinase	*CDK4*	Amplification or point mutation	Glioblastoma, melanoma, sarcoma

that regulate cellular behavior are laid out, including the receptor tyrosine kinase pathway, the G protein–coupled receptor pathway, the JAK/STAT pathway, the WNT pathway, the Notch pathway, the Hedgehog pathway, the TGF-β/SMAD pathway, and the NF-κB pathway. Abnormalities in each of these pathways are implicated in the development and progression of various cancers.

Traditionally, discussion of oncoproteins and tumor suppressors has centered on their ability to accelerate or inhibit, respectively, DNA replication and cell cycle progression. This view has merit, and we will follow it in our initial description of their activities. However, the proliferation of cells requires not only DNA replication but also sufficient biosynthesis of membrane, protein, and various macromolecules and organelles to enable a "mother" cell to divide and produce two complete daughter cells. Cell growth pathways implicated in oncogenesis also initiate signals that promote and coordinate the biosynthesis of all essential cellular components (discussed later). This insight has generated interest in therapeutic targeting of many aspects of oncogenic pro-growth signaling including the altered cellular metabolism that is characteristic of cancer cells.

Building on this framework, we next discuss some of the most important oncoproteins and the mechanisms by which they contribute to the autonomous growth of cancer cells.

Oncoproteins and Cell Growth

Oncogenes have multiple roles, but virtually all encode constitutively active oncoproteins that participate in signaling pathways that drive the proliferation of cells. Thus proto-oncogenes, the normal regulated versions of oncogenes, may encode growth factors, growth factor receptors, signal transducers, transcription factors, or cell cycle components. In most instances, the corresponding oncogenes encode oncoproteins that serve functions similar to their normal counterparts, with the important difference that they are usually constitutively active and thereby relieve cells of their normal dependency on growth factors.

In the following sections, we "walk down" a prototypical growth factor signaling pathway from the membrane to the nucleus (Fig. 7.22), discussing at each step along the way some of the genes and factors that are most commonly dysregulated (and therefore most important) in cancer.

Growth Factors. Most growth factors are made by one cell type and act on a neighboring cell of a differing type expressing the appropriate growth factor receptor (paracrine action). Some cancer cells, however, synthesize the same growth factor to which they are responsive, creating an autocrine loop. For example, brain tumors called *glioblastomas* (Chapter 28) often express both platelet-derived growth factor (PDGF) and PDGF receptor (PDGFR), and many sarcomas overexpress transforming growth factor α (TGF-α) and its receptor, epidermal growth factor receptor (EGFR).

Growth Factor Receptors. A large number of oncogenes encode growth factor receptors, of which receptor tyrosine kinases are arguably the most important in cancer. Recall that receptor tyrosine kinases are transmembrane proteins with an extracellular growth factor–binding domain and a cytoplasmic tyrosine kinase domain (Chapter 1). Normally the receptor is activated transiently by binding of a specific growth factor, an event that induces a rapid change in receptor conformation to an active dimeric state. The activated receptor then autophosphorylates tyrosine residues in its own intracellular tail, and these modified residues serve as sites for recruitment of other signaling molecules including RAS and PI3K, key players in receptor tyrosine kinase signaling (described later). The oncogenic versions of these receptors are associated with mutations that lead to constitutive, growth factor–independent tyrosine kinase activity. Hence, the mutant receptors deliver mitogenic signals to the cell continuously, even in the absence of growth factor in the environment.

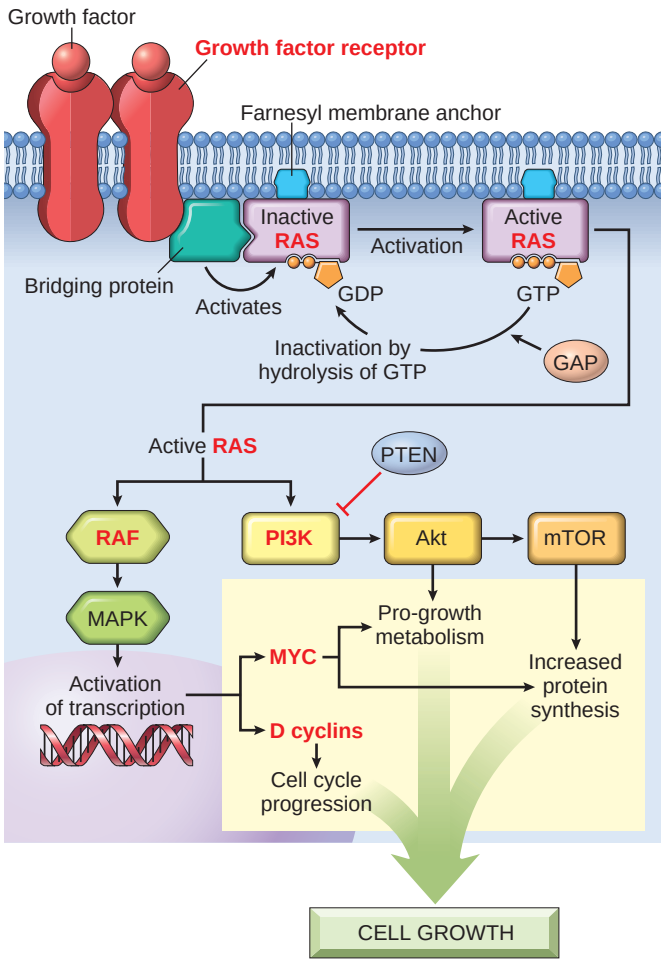


Figure 7.22 Growth factor signaling pathways in cancer. Growth factor receptors RAS, PI3K, MYC, and D cyclins are oncoproteins that are activated by mutations in various cancers. GTPase-activating proteins *(GAPs)* apply brakes to RAS activation, and phosphatase and tensin homologue *(PTEN)* serves the same function for PI3K. *GDP,* Guanosine diphosphate; *GTP,* guanosine triphosphate.

Receptor tyrosine kinases are constitutively activated in tumors by multiple mechanisms including point mutations, gene rearrangements, and gene amplifications. A few of the best-characterized oncogenic mutations involving growth factor receptors are listed in Table 7.5; the following are salient examples of particular clinical importance.

- *ERBB1* encodes EGFR. Several different *ERBB1* point mutations found in a subset of lung adenocarcinomas produce constitutive activation of the EGFR tyrosine kinase.
- *ERBB2* encodes a different member of the receptor tyrosine kinase family, HER2. Rather than being activated by point mutations, the *ERBB2* gene is amplified in certain breast carcinomas, leading to overexpression of the HER2 receptor and constitutive tyrosine kinase activity.
- *ALK* is a receptor tyrosine kinase that may be produced in a constitutively active form as a result of a gene rearrangement. For example, in a subset of lung adenocarcinomas, a deletion on chromosome 5 fuses part of the *ALK* gene with part of another gene called *EML4*. The resulting *EML4-ALK* fusion gene encodes a chimeric EML4-ALK protein with constitutive tyrosine kinase activity.

The importance of these mutated receptor tyrosine kinases has been proven in no small part by the therapeutic effectiveness of agents that inhibit their enzymatic activities. Breast cancers with *ERBB2* amplification and overexpression of HER2 respond well to antibodies or drugs that inhibit HER2 activity. These inhibitors not only block tumor growth but also induce apoptosis and tumor regression, reflecting the ability of receptor tyrosine kinase signaling to augment cell survival as well as proliferation. Inhibitors of EGFR and ALK produce similar therapeutic responses in patients with lung adenocarcinomas harboring *ERBB1* mutations or *EML4-ALK* fusion genes, respectively.

Unfortunately, these targeted therapies are usually not curative in advanced cancers. The tumor cells that withstand therapy typically are found to have other acquired mutations that sidestep the effects of the drug. For example, lung cancers that develop resistance to EGFR inhibitors often have mutations in EGFR that prevent inhibitors from binding, or amplifications in a gene called *MET,* which encodes yet another receptor tyrosine kinase. This experience highlights one of the most daunting clinical problems in the treatment of advanced cancers—the presence of subclones within the genetically heterogeneous tumor cell population that afford resistance to targeted therapies.

Downstream Components of the Receptor Tyrosine Kinase Signaling Pathway. As mentioned, receptor tyrosine kinase activation stimulates RAS and two major downstream signaling "arms," the MAPK cascade and the PI3K/AKT pathway. In line with the importance of these pathways in mediating cell growth, RAS, PI3K, and other components of these pathways are frequently involved by gain-of-function mutations in different types of cancer. Of interest, when *RAS* mutations are present in a tumor, activating mutations in receptor tyrosine kinases are almost always absent, at least within the dominant tumor clone, implying that in such tumors activated RAS can completely substitute for tyrosine kinase activity. Thus, lung adenocarcinomas fall into mutually exclusive molecular subtypes that are associated with mutations involving *RAS* or various tyrosine kinase genes, an insight that has important implications for targeted therapies in this type of cancer.

Point mutations of *RAS* family genes constitute the most common type of abnormality involving proto-oncogenes in human tumors. The RAS genes, of which there are three in humans (*HRAS*, *KRAS*, and *NRAS*), were discovered initially within the genomes of transforming retroviruses. Approximately 15% to 20% of all human tumors have RAS mutations, but in some types of cancers the frequency of RAS mutations is much higher. For example, 90% of pancreatic adenocarcinomas contain RAS mutations, as do about 50% of colon, endometrial, and thyroid cancers and 30% of lung adenocarcinomas and myeloid leukemias.

Recall that RAS proteins are members of a family of membrane-associated small G proteins that bind guanosine nucleotides (guanosine triphosphate [GTP] and guanosine diphosphate [GDP]), similar to the larger trimolecular G proteins. RAS normally flips back and forth between an excited signal-transmitting state in which it is bound to GTP and a quiescent state in which it is bound to GDP. Stimulation of receptor tyrosine kinases by growth factors leads to exchange of GDP for GTP and subsequent conformational changes that generate active RAS. Activation of RAS is transient because RAS has an intrinsic GTPase activity that is accelerated by GTPase-activating proteins (GAPs), which bind to active RAS and augment its GTPase activity by more than 1000-fold, thereby terminating signal transduction. Thus, GAPs prevent uncontrolled RAS activity.

Several distinct *RAS* point mutations have been identified in cancer cells that markedly reduce the GTPase activity of the RAS protein. These mutated forms of RAS are trapped in the activated GTP-bound form, and as a result the cell receives pro-growth signals continuously. It follows from this scenario that the consequences of gain-of-function mutations in RAS proteins should be mimicked by loss-of-function mutations in GAPs that normally restrain RAS activity. Indeed, disabling mutations of neurofibromin 1, a GAP encoded by the *NF1* gene, are associated with the inherited cancer syndrome *familial neurofibromatosis type 1* (Chapter 25). *NF1* is therefore an example of a tumor suppressor gene that acts through negative regulation of RAS signaling.

The MAPK and PI3K/AKT cascades both lie downstream of RAS and consist of a series of kinases, many of which are mutated in cancer cells. Components positioned close to the top of each cascade are frequently involved by oncogenic gain-of-function mutations in various cancers, as follows:

- *Mutations in BRAF,* a member of the RAF family of serine/threonine protein kinases, are found in close to 100% of hairy cell leukemias, 60% of melanomas, and a smaller percentage of a wide variety of other neoplasms including colon carcinomas. Like activating RAS mutations, activating mutations in BRAF stimulate downstream kinases and ultimately activate transcription factors. Mutations in other MAPK family members downstream of BRAF are less common in cancer, suggesting mutations affecting factors near the top of the cascade are most effective at producing pro-growth effects.
- *Mutations in kinases of the PI3K family* are also very common in certain cancers. For example, about 30% of breast carcinomas have PI3K gain-of-function mutations. In other instances, PI3K is "unbridled" by loss-of-function mutations in its negative regulator, called PTEN, a tumor suppressor that is commonly mutated in endometrial carcinoma. Under normal circumstances, following

receptor tyrosine kinase activation, PI3K is recruited to plasma membrane–associated protein complexes. Here, like BRAF, it activates a cascade of serine/threonine kinases, including AKT. AKT phosphorylates more than 150 proteins and constitutes a major signaling node. Its substrates include key regulators of protein synthesis (mTOR) and apoptosis (BAD, FOXO transcription factors, MDM2, and IAP, all described elsewhere).

Because RAS proteins are so frequently mutated in cancer, much effort has been spent trying to develop drugs that inhibit RAS. Unfortunately, none of these strategies has been successful, in large part because what is required is the restoration of a missing enzymatic activity (GTPase activity), an effect that is generally difficult to achieve with drugs. In contrast, treatment of patients with advanced melanomas with BRAF inhibitors has produced striking clinical responses. Such responses are strictly limited to tumors with *BRAF* mutations since these are dependent on BRAF signaling, whereas melanomas with wild-type *BRAF* genes do not respond. This phenomenon, termed *oncogene addiction* (described below), highlights the need for molecular analysis to guide appropriate therapy. Multiple drugs that inhibit various PI3K isoforms have also been developed, and some are now approved for treatment of particular cancers.

Nonreceptor Tyrosine Kinases. Oncogenic mutations also occur in several nonreceptor tyrosine kinases that normally localize to the cytoplasm or the nucleus. In many instances the mutations take the form of chromosomal translocations or rearrangements that create fusion genes encoding constitutively active tyrosine kinases. Despite their nonmembranous localization, these oncoproteins seem to activate the same signaling pathways as receptor tyrosine kinases. An important example of this oncogenic mechanism involves the ABL tyrosine kinase. In chronic myeloid leukemia (CML) and a subset of acute lymphoblastic leukemia, the *ABL* gene is translocated from its normal abode on chromosome 9 to chromosome 22 (Fig. 7.23), where it fuses with the *BCR* gene (see discussion of chromosomal translocations later in this chapter). The resultant fusion gene encodes a chimeric BCR-ABL protein with constitutive tyrosine kinase activity. The most important contribution of the BCR moiety is to promote self-association of BCR-ABL, which appears to be sufficient to unleash the tyrosine kinase activity of ABL. This represents a recurrent story in cancer, as many different oncogenic tyrosine kinases consist of chimeric proteins in which the non–tyrosine kinase partner drives self-association.

Treatment of CML has been revolutionized by the development of BCR-ABL kinase inhibitors, another example of rational drug design emerging from an understanding of the molecular basis of cancer. The remarkable therapeutic response of CML to BCR-ABL inhibitors is one of the first and best examples of *oncogene addiction,* in which tumor cells are highly dependent on the activity of one oncoprotein. Despite accumulation of mutations in other cancer-associated genes in CML cells, signaling through the BCR-ABL tyrosine kinase is required for most CML tumor cells to proliferate and survive; hence, inhibition of its activity is a highly effective therapy. The presence of a *BCR-ABL* fusion gene defines CML and must be the initiating event in this disease;

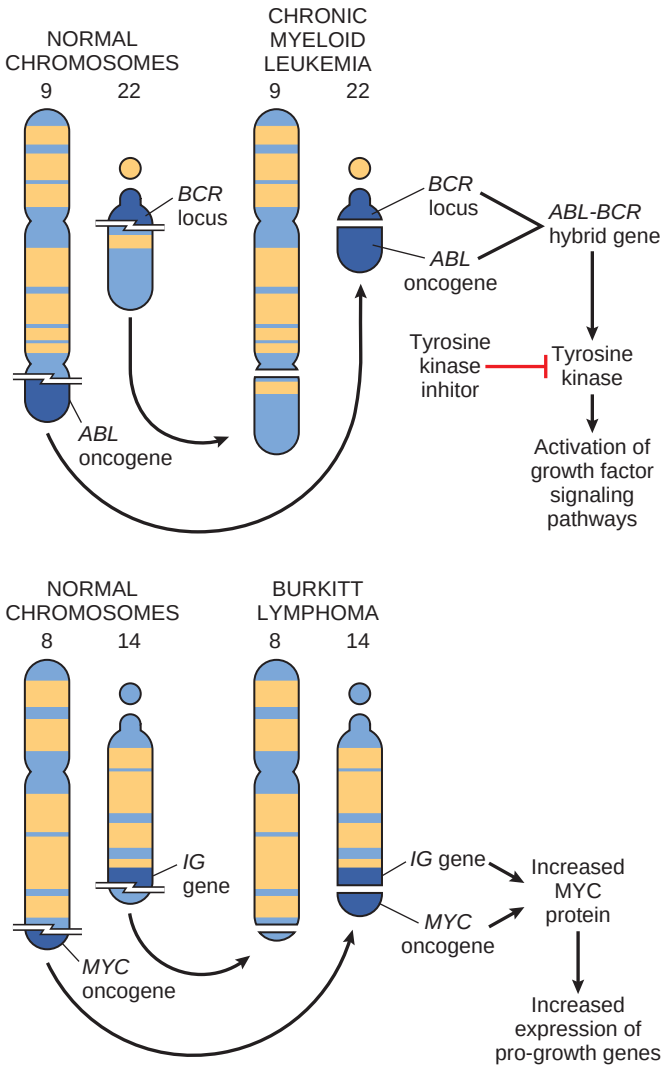


Figure 7.23 Chromosomal translocation and associated oncogenes in Burkitt lymphoma and chronic myeloid leukemia.

thus, additional mutations acquired by the founding clone are selected for their ability to complement the effects of incessant BCR-ABL signaling. BCR-ABL can be seen as the central lodgepole around which an oncogenic signaling "structure" is built. If the lodgepole is removed by treatment with BCR-ABL kinase inhibitors, the entire structure collapses. Unfortunately, treatment of this "addiction" to BCR-ABL does not lead to cure. Even though the proliferating component of the tumor is suppressed by BCR-ABL inhibitors and the patient seems completely well, rare CML "stem cells" harboring the *BCR-ABL* fusion gene persist, apparently because these cells do not require BCR-ABL signals for their survival. As a result, therapy with BCR-ABL inhibitors must be continued indefinitely; otherwise, the malignant stem cells spawn rapidly proliferating offspring, and the full-blown leukemia returns. This outcome highlights a second important concept that we will return to: the existence of "stem-like" cells in certain cancers that may be particularly resistant to therapeutic targeting.

In other instances, nonreceptor tyrosine kinases are activated by point mutations that abrogate the function of negative autoregulatory domains that normally hold enzyme

activity in check. An example of this type of mutation is found in the nonreceptor tyrosine kinase JAK2. JAK2 participates in the JAK/STAT signaling pathway, which transduces mitogenic signals from growth factor and cytokine receptors that lack tyrosine kinase activity (as described in Chapter 1). JAK/STAT activation alters the expression of target genes that bind STAT transcription factors. Several myeloid neoplasms are frequently associated with activating point mutations in JAK2 that relieve the tumor cells of their normal dependence on hematopoietic growth factors such as erythropoietin (Chapter 13). Recognition of this molecular lesion has led to the clinical development of JAK2 inhibitors and has stimulated searches for activating mutations in other nonreceptor tyrosine kinases.

Transcription Factors. Just as all roads lead to Rome, all signal transduction pathways converge on the nucleus, where the expression of target genes that orchestrate the cell's orderly advance through the cell cycle is activated. Indeed, the ultimate consequence of deregulated mitogenic signaling pathways is inappropriate and continuous stimulation of nuclear transcription factors that drive growth-promoting genes. Thus not surprisingly, growth autonomy may also occur as a consequence of mutations affecting transcription factors that regulate the expression of pro-growth genes and cyclins. Transcription factors of this class include the products of the *MYC, MYB, JUN, FOS,* and *REL* proto-oncogenes. Of these, *MYC* is most commonly affected in cancer, and hence a brief overview of its regulation and function follows.

MYC. The *MYC* proto-oncogene is expressed in virtually all eukaryotic cells and belongs to the immediate early response genes, which are rapidly and transiently induced by RAS/MAPK signaling following growth factor stimulation of quiescent cells. Under normal circumstances, MYC protein concentrations are tightly controlled at the level of transcription, translation, and protein stability, and virtually all pathways that regulate growth impinge on MYC through one or more of these mechanisms.

How MYC promotes normal and neoplastic cell growth is incompletely understood, but a multitude of studies have shown that **MYC has remarkably broad activities, several of which contribute not only to deregulated cell growth but also to several other hallmarks of cancer**.

- *MYC activates the expression of many genes that are involved in cell growth.*
  - Some MYC target genes, like D cyclins, are directly involved in cell cycle progression.
  - MYC also upregulates the expression of ribosomal RNA (rRNA) genes and rRNA processing, thereby enhancing the assembly of ribosomes needed for protein synthesis.
  - MYC upregulates a program of gene expression that leads to metabolic reprogramming and the Warburg effect, another cancer hallmark (discussed later). Among the genes involved in metabolism that are upregulated by MYC are multiple glycolytic enzymes and factors involved in glutamine metabolism, both of which contribute to the generation of metabolic intermediates that are needed for synthesis of macromolecules such as DNA, proteins, and lipids.
  - Based on these protean effects, MYC can be considered a master transcriptional regulator of cell growth. Indeed, the fastest growing human tumors, such as Burkitt lymphoma, which virtually always bears a chromosomal translocation involving *MYC* (see Fig. 7.23), are those with the highest levels of MYC.
- *In some contexts, MYC upregulates expression of telomerase.* As discussed later, telomerase is one of several factors that contribute to the endless replicative capacity (the immortalization) of cancer cells.
- *MYC is one of a handful of transcription factors that can act together to reprogram somatic cells into pluripotent stem cells* (Chapter 1). This capacity has led to suspicions that MYC also contributes to cancer cell "stemness," another important aspect of the immortality of cancers.

Given the importance of MYC in regulation of cell growth, it should come as no surprise that it is deregulated in cancer through a large variety of mechanisms. Sometimes deregulation involves genetic alterations of *MYC* itself. In Burkitt lymphoma and a subset of other B- and T-cell tumors, the *MYC* gene is translocated into an antigen receptor gene locus, which contains gene regulatory elements called enhancers that are highly active in lymphocytes. Rather than driving the expression of B- or T-cell receptors, these misplaced enhancers instead cause the deregulation and overexpression of MYC protein. Alternatively the *MYC* gene is amplified in cancers of the breast, colon, lung, and other tissues, again resulting in overexpression of MYC. The functionally identical *NMYC* and *LMYC* genes are also amplified in neuroblastomas (Fig. 7.24) and small cell cancers of the lung, respectively. In many other instances, oncogenic

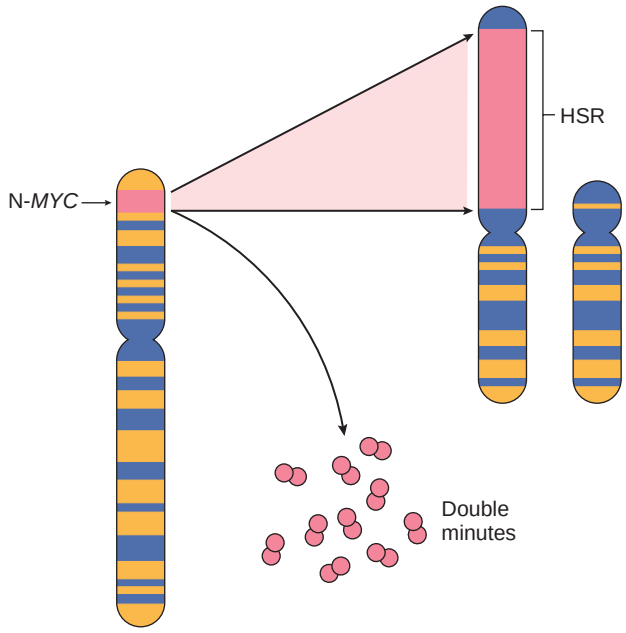


Figure 7.24 Amplification of the *NMYC* gene in human neuroblastomas. The *NMYC* gene, normally present on chromosome 2p, becomes amplified and is seen either as extra chromosomal double minutes or as a chromosomally integrated, homogeneous staining region *(HSR)*. The integration involves other autosomes such as 4, 9, or 13. (Modified from Brodeur GM: Molecular correlates of cytogenetic abnormalities in human cancer cells: implications for oncogene activation. In Brown EB, editor: *Progress in Hematology,* vol 14, Orlando, Fla, 1986, Grune & Stratton, pp 229–256.)

mutations involving components of upstream signaling pathways elevate MYC protein levels by increasing *MYC* transcription, enhancing *MYC* messenger RNA (mRNA) translation, and/or stabilizing MYC protein. Thus, constitutive RAS/MAPK signaling (many cancers), Notch signaling (several cancers), Wnt signaling (colon carcinoma), and Hedgehog signaling (medulloblastoma) all transform cells in part through upregulation of *MYC*. Finally, several single nucleotide polymorphisms (SNPs) associated with an inherited risk for cancers such as prostate and ovarian carcinoma and certain leukemias lie within enhancer elements that flank *MYC;* these variants appear to stimulate higher levels of *MYC* RNA expression in response to growth-promoting signals. Thus, there is seemingly no end to the ways in which MYC may be deregulated in cancer cells.

Cyclins and Cyclin-Dependent Kinases. As mentioned in Chapter 1, growth factors transduce signals that stimulate the orderly progression of cells through the various phases of the cell cycle, the process by which cells replicate their DNA in preparation for cell division. Progression of cells through the cell cycle is orchestrated by cyclin-dependent kinases (CDKs), which are activated by binding to cyclins, so called because of the cyclic nature of their production and degradation. The CDK-cyclin complexes phosphorylate crucial target proteins that drive cells forward through the cell cycle. While cyclins arouse the CDKs, CDK inhibitors, of which there are many, silence the CDKs and exert negative control over the cell cycle (Table 7.6). Expression of these inhibitors is downregulated by mitogenic signaling pathways, thus promoting the progression of the cell cycle.

There are two main cell cycle checkpoints, one at the G_1/S transition and the other at the G_2/M transition, both of which are tightly regulated by a balance of growth-promoting and growth-suppressing proteins, as well as by sensors of DNA damage (Chapter 1). If activated, these DNA-damage sensors transmit signals that arrest cell cycle progression and, if the damage cannot be repaired, initiate apoptosis. Understandably, defects in the G_1/S checkpoint are more important in cancer because these not only lead to dysregulated growth but may also impair DNA repair, creating a "mutator" phenotype that (as mentioned) enables cancer development and progression.

The major cancer-associated mutations that affect the G_1/S checkpoint can be broadly grouped into two classes.

- *Gain-of-function mutations in D cyclin genes and CDK4, which promote unregulated G_1/S progression and thus function as oncogenes.* There are three D cyclin genes, D1, D2, and D3, which are functionally interchangeable and often dysregulated by acquired mutations in cancer, including chromosomal translocations in lymphoid tumors and gene amplification in a variety of solid tumors. Amplification of the *CDK4* gene also occurs in melanomas, sarcomas, and glioblastomas. CDK4 inhibitors are effective in treatment of advanced breast cancers associated with excessive CDK4 activity. Mutations affecting other CDKs and cyclin E also occur in cancers, but they are infrequent, presumably because these factors are less important in control of the G_1/S transition, which has a preeminent role in regulating tumor growth rates.
- *Loss-of-function mutations in genes that inhibit G_1/S progression.* Examples of these tumor suppressor genes are those encoding CDK inhibitors, which inhibit cyclin D/CDK complexes and are frequently mutated or otherwise silenced in many human malignancies. For example, germline mutations of *p16 (CDKN2A)* are present in 25% of melanoma-prone kindreds, and somatically acquired deletion or inactivation of p16 is seen in 75% of pancreatic carcinomas, 40% to 70% of glioblastomas, 50% of esophageal cancers, 20% to 70% of acute lymphoblastic leukemias, and 20% of non–small cell lung carcinomas, soft tissue sarcomas, and bladder cancers. Furthermore, the two most important tumor suppressor genes, *RB* and *TP53,* both encode proteins that inhibit G_1/S progression.

Table 7.6 Cell Cycle Components and Inhibitors That Are Frequently Mutated in Cancer

Cell Cycle Component	Main Function
Cyclins and Cyclin-Dependent Kinases	
CDK4; D cyclins	Form a complex that phosphorylates RB, allowing the cell to progress through the G_1 restriction point
Cell Cycle Inhibitors	
CIP/KIP family: p21, p27 (CDKN1A–D)	Block the cell cycle by binding to cyclin-CDK complexes p21 is induced by tumor suppressor p53 p27 responds to growth suppressors such as TGF-β
INK4/ARF family (CDKN2A–C)	p16/INK4a binds to cyclin D–CDK4 and promotes the inhibitory effects of RB p14/ARF increases p53 levels by inhibiting MDM2 activity
Cell Cycle Checkpoint Components	
RB	Tumor suppressive "pocket" protein that binds E2F transcription factors in its hypophosphorylated state, preventing G_1/S transition Interacts with transcription factors that regulate differentiation
p53	Tumor suppressor altered in the majority of cancers Induced by DNA damage Causes cell cycle arrest by upregulating the CDK inhibitor p21 Induces apoptosis by upregulating BAX and other pro-apoptotic genes

KEY CONCEPTS

ONCOGENES, ONCOPROTEINS, AND UNREGULATED CELL PROLIFERATION

Proto-oncogenes: normal cellular genes whose products promote cell proliferation.

Oncogenes: mutated or overexpressed versions of proto-oncogenes that function autonomously, having lost dependence on normal growth-promoting signals.

Oncoprotein: a protein encoded by an oncogene that drives increased cancer cell proliferation, which may result from a variety of aberrations.

- Constitutive expression of growth factors and their cognate growth factor receptors, setting up an autocrine cell signaling loop.
- Mutations in growth factor receptors, nonreceptor tyrosine kinases, or downstream signaling molecules that lead to constitutive signaling, such as:
  - Activation of the EGF receptor tyrosine kinase by point mutations (lung cancer), activation of the HER2 receptor tyrosine kinase by gene amplification (breast cancer), and activation of the JAK2 tyrosine kinase by point mutations (myeloproliferative neoplasms).
  - Activation of the ABL nonreceptor tyrosine kinase by chromosomal translocation and creation of a *BCR-ABL* fusion gene (chronic myeloid leukemia, acute lymphoblastic leukemia).
  - Activation of RAS by point mutations (many cancers).
  - Activation of PI3K and BRAF serine/threonine kinases by point mutations (many cancers).
  - Increased expression of MYC, a master transcription factor that regulates genes needed for rapid cell growth by deregulation through chromosomal translocations (Burkitt lymphoma, other hematologic malignancies), gene amplication (neuroblastoma), and increased activity of upstream signaling pathways (many cancers).
  - Mutations that increase the activity of cyclin-dependent kinase 4 (CDK4)/D cyclin complexes, which promote cell cycle progression.

Insensitivity to Growth Inhibition: Tumor Suppressor Genes

Whereas oncogenes drive the proliferation of cells, the products of most tumor suppressor genes apply brakes to cell proliferation, and abnormalities in these genes lead to failure of growth inhibition, another fundamental hallmark of carcinogenesis. Tumor suppressor proteins control a series of checkpoints that prevent uncontrolled growth. Many tumor suppressors, such as RB and p53, are part of a regulatory network that recognizes genotoxic stress from any source and responds by shutting down proliferation. Indeed, expression of an oncogene in normal cells with intact tumor suppressor genes leads to quiescence or permanent cell cycle arrest (oncogene-induced senescence, discussed later), rather than uncontrolled proliferation. Ultimately, the growth inhibitory pathways may prod the cells into apoptosis. Another set of tumor suppressors seems to be involved in cell differentiation, causing cells to enter a postmitotic, differentiated pool without replicative potential. Similar to mitogenic signals, signals that induce growth inhibition and differentiation originate outside the cell and use receptors, signal transducers, and nuclear transcription regulators to accomplish their effects; tumor suppressors form a portion of these networks. Thus the protein products of tumor suppressor genes may function as transcription factors, cell cycle inhibitors, signal transduction molecules, cell surface receptors, and regulators of cellular responses to DNA damage.

In this section, we describe tumor suppressor genes, their products, and mechanisms by which loss of their function contributes to unregulated cell growth (Table 7.7). Many of our current concepts of tumor suppressors evolved from studies of the retinoblastoma (*RB*) gene, the first tumor suppressor gene discovered, which remains a prototype of genes of this type. Like many discoveries in medicine, *RB* was identified by studying a rare inherited disease, familial retinoblastoma. Approximately 40% of retinoblastomas are familial, with the predisposition to develop the tumor being transmitted as an autosomal dominant trait. Carriers of this trait have a 10,000-fold increased risk of developing retinoblastoma (often in both eyes) compared with the general population and are at greatly increased risk of developing osteosarcoma and other soft tissue sarcomas. The remaining 60% of retinoblastomas occur sporadically (virtually always in only one eye), and such patients are not at increased risk for other forms of cancer. To explain these two patterns of occurrence of retinoblastoma, Knudson proposed his now canonic "two-hit" hypothesis of oncogenesis. In molecular terms, Knudson's hypothesis can be stated as follows (Fig. 7.25):

- Two mutations (hits), involving both alleles of *RB* are required to produce retinoblastoma.
- In familial cases, children inherit one defective copy of *RB* (the first hit) and one normal copy of *RB* in the germline. Retinoblastoma develops when the normal *RB* allele is mutated in retinoblasts as a result of a spontaneous somatic mutation (the second hit). Because second hits seem to be virtually inevitable in a small fraction of retinoblasts, most individuals inheriting a germline defect in one *RB* allele develop unilateral or bilateral retinoblastoma, and the disease is inherited as an autosomal dominant trait.
- In sporadic cases both normal *RB* alleles must undergo somatic mutation in the same retinoblast (two hits). The probability of this event is low (explaining why retinoblastoma is uncommon in the general population), but the end result is the same: a retinal cell loses *RB* function and becomes cancerous.

A child carrying an inherited mutant *RB* allele in all somatic cells is perfectly normal (except for the increased risk of developing cancer); it follows that one defective *RB* gene does not have adverse effects on cell behavior. Thus, although the genetic trait (increased cancer risk) associated with germline *RB* mutations is inherited in an autosomal dominant fashion, at the level of individual cells the phenotype associated with *RB* loss of function behaves like a recessive trait.

Following the identification of *RB,* a large number of other tumor suppressor genes were discovered, often through study of other types of familial cancer. In general, the major themes that emerged from the study of familial retinoblastoma hold for other familial cancers: the risk of cancer is inherited as an autosomal dominant trait due to a germline mutation in a tumor suppressor gene; tumors have second "hits" in the sole normal tumor suppressor gene allele; and the same tumor suppressor gene is frequently mutated in sporadic tumors of the same type.

Some of the most important tumor suppressor genes, their associated familial syndromes, and their normal

Table 7.7 Selected Tumor Suppressor Genes and Associated Familial Syndromes and Cancers, Sorted by Cancer Hallmarks[a]

Gene	Protein	Function	Familial Syndromes	Sporadic Cancers
Inhibitors of Mitogenic Signaling Pathways				
APC	Adenomatous polyposis coli protein	Inhibitor of WNT signaling	Familial colonic polyps and carcinomas	Carcinomas of stomach, colon, pancreas; melanoma
NF1	Neurofibromin-1	Inhibitor of RAS/MAPK signaling	Neurofibromatosis type 1 (neurofibromas and malignant peripheral nerve sheath tumors)	Neuroblastoma, juvenile myeloid leukemia
NF2	Merlin	Cytoskeletal stability, Hippo pathway signaling	Neurofibromatosis type 2 (acoustic schwannoma and meningioma)	Schwannoma, meningioma
PTCH	Patched	Inhibitor of Hedgehog signaling	Gorlin syndrome (basal cell carcinoma, medulloblastoma, several benign tumors)	Basal cell carcinoma, medulloblastoma
PTEN	Phosphatase and tensin homologue	Inhibitor of PI3K/AKT signaling	Cowden syndrome (variety of benign skin, GI, and CNS growths; breast, endometrial, and thyroid carcinoma)	Diverse cancers, particularly carcinomas and lymphoid tumors
SMAD2, SMAD4	SMAD2, SMAD4	Component of the TGF-β signaling pathway, repressors of MYC and CDK4 expression, inducers of CDK inhibitor expression	Juvenile polyposis	Frequently mutated (along with other components of the TGF-β signaling pathway) in colonic and pancreatic carcinoma
Inhibitors of Cell Cycle Progression				
RB	Retinoblastoma (RB) protein	Inhibitor of G_1/S transition during cell cycle progression	Familial retinoblastoma syndrome (retinoblastoma, osteosarcoma, other sarcomas)	Retinoblastoma; osteosarcoma; carcinomas of breast, colon, lung
CDKN2A	p16/INK4a and p14/ARF	p16: Negative regulator of cyclin-dependent kinases; p14, indirect activator of p53	Familial melanoma	Pancreatic, breast, and esophageal carcinoma; melanoma; certain leukemias
Inhibitors of Pro-growth Programs of Metabolism and Angiogenesis				
VHL	von Hippel–Lindau (VHL) protein	Inhibitor of hypoxia-induced transcription factors (e.g., HIF1α)	von Hippel–Lindau syndrome (cerebellar hemangioblastoma, retinal angioma, renal cell carcinoma)	Renal cell carcinoma
STK11	Liver kinase B1 (LKB1) or STK11	Activator of AMPK family of kinases; suppresses cell growth when cell nutrient and energy levels are low	Peutz-Jeghers syndrome (GI polyps, GI cancers, pancreatic carcinoma, and other carcinomas)	Diverse carcinomas (5%–20% of cases, depending on type)
SDHB, SDHD	Succinate dehydrogenase complex subunits B and D	TCA cycle, oxidative phosphorylation	Familial paraganglioma, familial pheochromocytoma	Paraganglioma
Inhibitors of Invasion and Metastasis				
CDH1	E-cadherin	Cell adhesion, inhibition of cell motility	Familial gastric cancer	Gastric carcinoma, lobular breast carcinoma
Enablers of Genomic Stability				
TP53	p53 protein	Cell cycle arrest and apoptosis in response to DNA damage	Li-Fraumeni syndrome (diverse cancers)	Most human cancers
DNA Repair Factors				
BRCA1, BRCA2	Breast cancer-1 and breast cancer-2 (BRCA1 and BRCA2)	Repair of double-stranded breaks in DNA	Familial breast and ovarian carcinoma; carcinomas of male breast; chronic lymphocytic leukemia (BRCA2)	Rare
MSH2, MLH1, MSH6	MSH1, MLH1, MSH6	DNA mismatch repair	Hereditary nonpolyposis colon carcinoma	Colonic and endometrial carcinoma
Unknown Mechanisms				
WT1	Wilms tumor-1 (WT1)	Transcription factor	Familial Wilms tumor	Wilms tumor, certain leukemias
MEN1	Menin	Transcription factor	Multiple endocrine neoplasia-1 (MEN1) (pituitary, parathyroid, and pancreatic endocrine tumors)	Pituitary, parathyroid, and pancreatic endocrine tumors

[a]Some tumor suppressors impact multiple cancer phenotypes (e.g., p53 affects cell cycle progression, genomic stability, susceptibility to cell death, and cellular metabolism); only a subset of major effects are given for each tumor suppressor gene listed.

CNS, Central nervous system; *GI,* gastrointestinal; *TCA,* tricarboxylic acid.

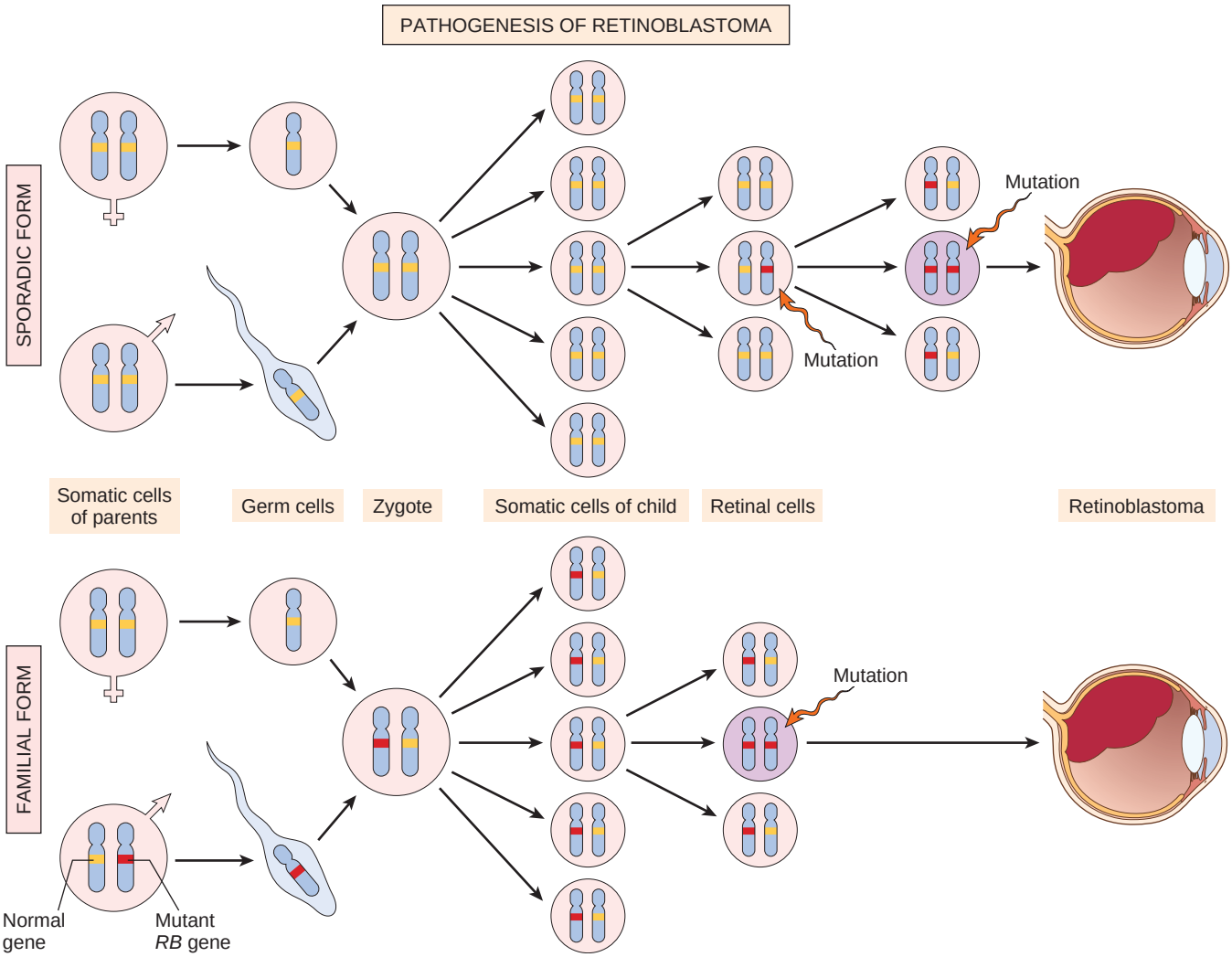


Figure 7.25 Pathogenesis of retinoblastoma. Two mutations of the *RB* locus on chromosome 13q14 lead to neoplastic proliferation of the retinal cells. In the sporadic form, both *RB* mutations in the tumor-founding retinal cell are acquired. In the familial form, all somatic cells inherit one mutated copy of *RB* gene from a carrier parent, and as a result only one additional *RB* mutation in a retinal cell is required for complete loss of RB function.

functions are listed in Table 7.7. Note that while tumor suppressors were initially thought of narrowly as proteins that put the brakes on cell cycle progression and DNA replication, it is now appreciated that some tumor suppressors prevent cellular transformation through other mechanisms such as by altering cell metabolism or by ensuring genomic stability. Thus, while most tumor suppressors have inhibitory effects on cell growth through one mechanism or another, a more inclusive definition of a tumor suppressor is simply a protein or gene that opposes any of the various hallmarks of cancer.

We next consider how specific tumor suppressors function, focusing on factors that are frequently mutated in cancer or that highlight pathogenically important molecular mechanisms.

RB: Governor of Proliferation. **RB, a key negative regulator of the G_1/S cell cycle transition, is directly or indirectly inactivated in most human cancers.** RB exists in an active hypophosphorylated state in quiescent cells and an inactive hyperphosphorylated state in cells passing through the G_1/S cell cycle transition (Chapter 1). Its function may be compromised in two different ways.

- Loss-of-function mutations involving both *RB* alleles.
- A shift from the active hypophosphorylated state to the inactive hyperphosphorylated state caused by gain-of-function mutations that upregulate CDK/cyclin D activity or by loss-of-function mutations that abrogate the activity of CDK inhibitors.

As discussed previously, the "decision" of a cell to progress from G_1 into S is of great importance, as once a cell enters S phase it is obligated to complete mitosis. High levels of CDK4/cyclin D, CDK6/cyclin D, and CDK2/cyclin E complexes lead to hyperphosphorylation and inhibition of RB, releasing E2F transcription factors that drive the expression of genes that are needed for progression to S phase (Fig. 7.26). Growth factor signaling pathways generally upregulate the activity of CDK/cyclin complexes and drive cells through the G_1/S transition, whereas growth inhibitors

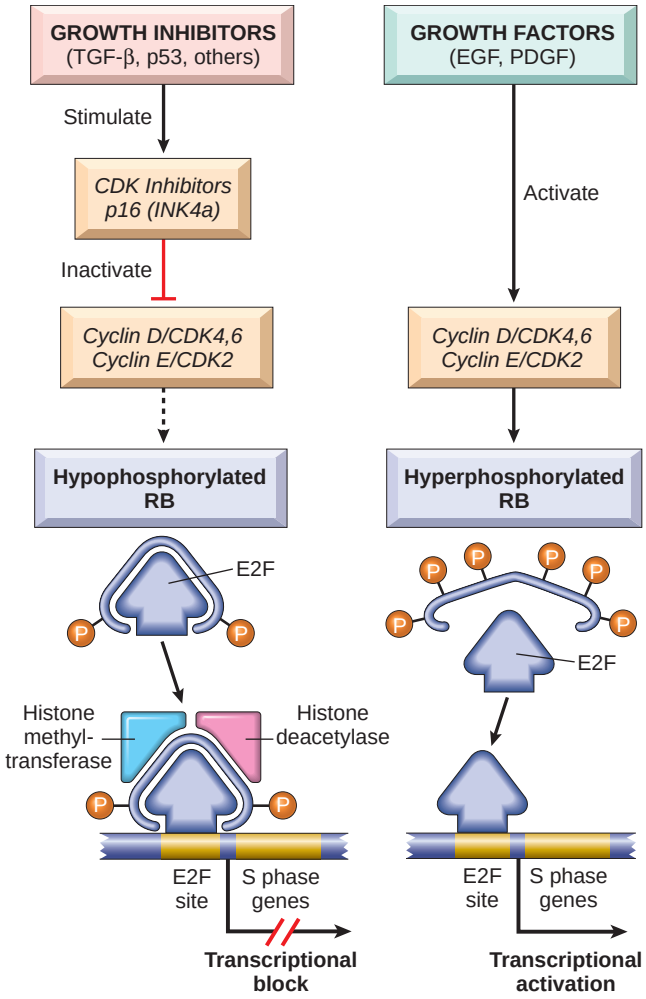


Figure 7.26 The role of RB in regulating the G_1-S checkpoint of the cell cycle. Hypophosphorylated RB in complex with the E2F transcription factors binds to DNA, recruits chromatin-modifying factors (histone deacetylases and histone methyltransferases), and inhibits transcription of genes whose products are required for the S phase of the cell cycle. When RB is phosphorylated by the cyclin D–CDK4, cyclin D–CDK6, and cyclin E–CDK2 complexes, it releases E2F. The latter then activates transcription of S-phase genes. Phosphorylation of RB is inhibited by cyclin-dependent kinase inhibitors because they inactivate cyclin–CDK complexes. In most cancers, the G_1-S checkpoint is defective as a result of mutation of one of four genes that regulate the phosphorylation of RB; these genes are *RB*, *CDK4*, the genes encoding cyclin D proteins, and *CDKN2A* (p16). *EGF,* Epidermal growth factor; PDGF, platelet-derived growth factor; *TGF-β,* transforming growth factor-β.

tip the balance the other way by upregulating CDK inhibitors. RB is the point of integration of these opposing signals, making it a key regulator of cell cycle progression.

It was mentioned previously that germline and somatic loss-of-function *RB* mutations are associated with retinoblastoma and osteosarcoma, and analyses of cancer cell genomes have identified similar somatic *RB* mutations in subsets of glioblastoma and lung, breast, and bladder carcinoma. However, given that RB is expressed in all cells, one may ask, why do patients with germline *RB* mutations preferentially develop retinoblastoma instead of other cancers? And, conversely, why are acquired mutations of *RB* not found in all kinds of cancer?

The reason why persons who inherit one defective allele of *RB* develop retinoblastoma and not other tumors is not understood, but a possible explanation is that other RB family members exist with RB-like activities; these proteins may fulfill the role of RB in cell types other than retinoblasts. With respect to the second question, the answer is much simpler: mutations in other genes that control RB phosphorylation can mimic the effect of *RB* loss, and such genes are mutated in many cancers that have normal *RB* genes. Thus, for example, mutational activation of cyclin D or CDK4 and mutational inactivation of CDK inhibitors favors cell proliferation by facilitating the hyperphosphorylation and inactivation of RB. The current paradigm is that **loss of normal cell cycle control is central to malignant transformation and that at least one of four key regulators of the cell cycle (p16/INK4a, cyclin D, CDK4, RB) is dysregulated in the vast majority of human cancers**. In cells that harbor mutations in any one of these genes or in upstream factors that regulate their expression and function (e.g., receptor tyrosine kinases, RAS), RB may be functionally inactivated even if the *RB* gene itself is not mutated.

The transforming proteins of several oncogenic animal and human DNA viruses also neutralize the growth inhibitory activities of RB. Of greatest relevance to human cancer, polyomavirus large T antigens and E7 proteins from high-risk types of HPV (such as HPV16) bind to hypophosphorylated RB through the same "pocket" that RB uses to bind and sequester E2F transcription factors. Binding of the viral proteins thus inactivates RB and releases E2F transcription factors, freeing them to cause cell cycle progression.

KEY CONCEPTS

RB, GOVERNOR OF THE CELL CYCLE

- When hypophosphorylated, RB exerts antiproliferative effects by binding and inhibiting E2F transcription factors that regulate genes required for cells to pass through the G_1/S phase cell cycle checkpoint. Normal growth factor signaling leads to RB hyperphosphorylation and inactivation, thus promoting cell cycle progression.
- The antiproliferative effect of RB is abrogated in cancers through a variety of mechanisms, including:
  - Loss-of-function *RB* mutations
  - Amplifications of the CDK4 and cyclin D genes
  - Loss-of-function mutations affecting cyclin-dependent kinase inhibitors (e.g., p16/INK4a)
  - Viral oncoproteins that bind and inhibit RB (E7 protein of HPV)

TP53: Guardian of the Genome. ***TP53,* a tumor suppressor gene that regulates cell cycle progression, DNA repair, cellular senescence, and apoptosis, is the most frequently mutated gene in human cancers.** Loss-of-function mutations in *TP53,* located on chromosome 17p13.1, are found in more than 50% of cancers. Moreover, *TP53* mutations occur in virtually every type of cancer, including carcinomas of the lung, colon, and breast—the three leading causes of cancer death. In most cases, mutations are present in both *TP53* alleles and are acquired in somatic cells (not inherited in

the germline). Less commonly, individuals inherit one mutated *TP53* allele. As in the case of the *RB* tumor suppressor and retinoblastoma, inheritance of a mutated copy of *TP53* predisposes individuals to malignant tumors because only one additional "hit" in the remaining normal allele is needed to abrogate *TP53* function. Individuals with these inherited mutations, said to have the *Li-Fraumeni syndrome,* have a 25-fold greater chance of developing a malignant tumor by age 50 than the general population. In contrast to individuals who inherit a mutant *RB* allele, the spectrum of tumors that develop in persons with Li-Fraumeni syndrome is broad; the most common types are sarcomas, breast cancers, leukemias, brain tumors, and carcinomas of the adrenal cortex. Persons with Li-Fraumeni syndrome often develop cancer at younger ages and are more likely to suffer from multiple primary tumors of varying types than are normal individuals.

These mutational data, while impressive, only begin to tell the tale of altered *TP53* function in cancer. *TP53* encodes the protein p53, which is tightly regulated at several levels. Analogous to RB, many tumors lacking *TP53* mutations have instead other mutations affecting proteins that regulate p53 function. For example, MDM2 and related proteins of the MDM2 family stimulate the degradation of p53; these proteins are frequently overexpressed in malignancies with normal *TP53* alleles. Indeed, the *MDM2* gene is amplified in 33% of human sarcomas, leading to a functional deficiency of p53 in these tumors. Like RB, the transforming proteins of several DNA viruses bind p53 and promote its degradation. Best known of these viral oncoproteins is the E6 protein of high-risk HPVs, which have causative roles in cervical carcinoma and a subset of squamous cell carcinomas of the head and neck.

The frequent loss of p53 function in human tumors reflects its critical role in preventing cancer development. p53 is the focal point of a large network of signals that sense cellular stress, primarily DNA damage, but also shortened telomeres, hypoxia, and stress caused by excessive pro-growth signaling, as may occur in cells bearing mutations in genes such as *RAS* and *MYC*. In nonstressed, healthy cells, p53 is held at bay through its aforementioned association with MDM2, an enzyme that ubiquitinates p53, leading to its degradation by the proteasome. As a result, p53 is virtually undetectable in normal cells. In stressed cells, however, p53 is released from the inhibitory effects of MDM2 via two major mechanisms, which vary depending on the nature of the stress.

- *DNA damage and hypoxia*. The key initiators of p53 activation following DNA damage or hypoxic stress are two related protein kinases, ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad3 related (ATR). As the name implies, *ATM* was originally identified in individuals with ataxia-telangiectasia, which is caused by germline mutations in *ATM*. Affected patients suffer from an inability to repair certain kinds of DNA damage and have an increased incidence of cancer. The types of damage sensed by ATM and ATR are different, but the downstream effects are similar. Once triggered, both ATM and ATR stimulate the phosphorylation of p53 and MDM2. These posttranslational modifications disrupt the binding and degradation of p53 by MDM2, allowing p53 to accumulate.
- *"Oncogenic" stress*. Activation of oncoproteins such as RAS leads to sustained, supraphysiologic signaling through pro-growth pathways such as the MAPK and PI3K/AKT cascades. Through incompletely understood mechanisms, these aberrant signals create cellular stress and lead to increased expression of p14/ARF, which is encoded by the *CDKN2A* tumor suppressor gene. p14/ARF binds MDM2 and displaces p53, again allowing p53 levels to rise in the cell.

Once activated, p53 thwarts neoplastic transformation by inducing transient cell cycle arrest, senescence (permanent cell cycle arrest), or programmed cell death (apoptosis) (Fig. 7.27). p53 is a transcription factor that binds DNA in a sequence-specific fashion and activates the transcription of hundreds of target genes with p53-binding regulatory elements. The target genes that execute the functions of p53 are not completely defined but appear to fall into three major categories: (1) genes that cause cell cycle arrest, (2) genes that cause apoptosis, and (3) genes that enhance catabolic metabolism or inhibit anabolic metabolism. The last group of genes makes intuitive sense; there is no need for a cell that has stopped its cell cycle progression to continue to synthesize macromolecules (e.g., lipids and proteins) that are needed for cell growth and division.

Once p53 accumulates in a cell to levels that are sufficient to activate the transcription of target genes, several different outcomes are possible, each more serious than the last with respect to the ultimate fate of the affected cell.

- *Transient p53-induced cell cycle arrest.* Rapid onset, p53-mediated cell cycle arrest may be considered a primordial response to DNA damage. It occurs late in G_1 phase and is caused in part by p53-dependent transcription of the *CDKN1A* gene, which encodes the CDK inhibitor p21. p21 (like p16) inhibits CDK4/D cyclin complexes, thereby maintaining RB in an active, hypophosphorylated state and blocking the progression of cells from G_1 phase to S phase. This pause in cell cycling is welcome, as it gives the cells "breathing time" to repair DNA damage. p53 also helps the process by inducing proteins such as GADD45 (growth arrest and DNA damage) that enhance DNA repair. If DNA damage is repaired successfully, the signals responsible for p53 stabilization cease and p53 levels fall, releasing the cell cycle block. The cells may then revert to a normal state.
- *p53-induced senescence.* Senescence is a state of permanent cell cycle arrest characterized by specific changes in morphology and gene expression that differentiate it from reversible cell cycle arrest. How cells become fixed in the senescence state is unclear. One plausible idea is that senescence is the product of epigenetic changes that result in the formation of heterochromatin at key loci, including genes that are required for progression of cells from G_1 phase to S phase. Like other p53 responses, senescence may be stimulated in response to a variety of stresses, such as unopposed oncogene signaling, hypoxia, and shortened telomeres. Senescent cells, while not normal, cannot divide and therefore cannot develop into tumors.
- *p53-induced apoptosis.* Apoptosis of cells with irreversible DNA damage is the ultimate protective mechanism against neoplastic transformation. p53 directs the transcription

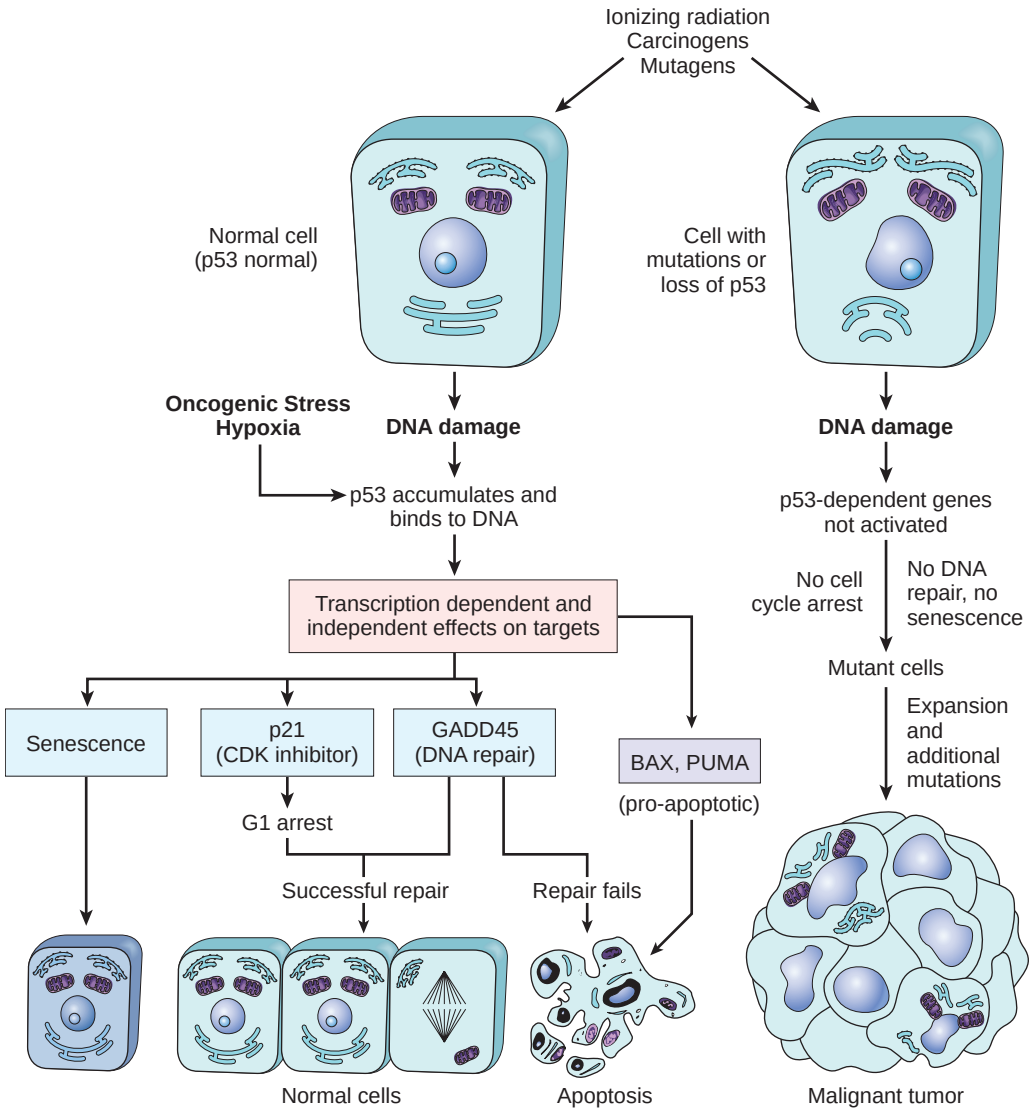


Figure 7.27 The role of p53 in maintaining the integrity of the genome. Activation of normal p53 by DNA-damaging agents or by hypoxia leads to cell cycle arrest in G_1 and induction of DNA repair by transcriptional upregulation of the cyclin-dependent kinase inhibitor *CDKN1A* (encoding the cyclin-dependent kinase inhibitor p21) and *GADD45* genes. Successful repair of DNA allows cells to proceed with the cell cycle; if DNA repair fails, p53 triggers either apoptosis or senescence. In cells with loss or mutations of the p53 gene, DNA damage does not induce cell cycle arrest or DNA repair, and genetically damaged cells proliferate, giving rise eventually to malignant neoplasms.

of several pro-apoptotic genes such as *BAX* and *PUMA* (described later), which are believed to tip the balance in favor of cell death via the intrinsic (mitochondrial) pathway.

What determines whether a cell repairs its DNA, becomes senescent, or undergoes apoptosis is uncertain, but both the duration and the level of p53 activation may be deciding factors. It appears that the affinity of p53 for its binding sites in the promoters and enhancers of DNA repair genes is higher than its affinity for binding sites in pro-apoptotic genes. Thus the DNA repair pathway is stimulated first as p53 begins to accumulate. If p53 is sustained at this level due to ineffective DNA repair or other chronic stresses (e.g., that induced by a potentially oncogenic RAS mutation), epigenetic silencing of genes that are needed for cell cycle progression occurs, leading to senescence. Alternatively, if enough p53 accumulates to stimulate the transcription of the pro-apoptotic genes, the cell dies. While this scheme seems to be generally correct, cell type–specific variations in response to p53 activation have been observed that are not easily explained, with some cell types succumbing rapidly to apoptosis and others opting mainly for senescence. Thus there is still much to be learned about the nuances of p53 function.

With loss of p53 function, DNA damage goes unrepaired, driver mutations accumulate in oncogenes and other cancer genes, and the cell marches along a dangerous path leading to malignant transformation. Moreover, once a cancer is established, its p53 status has important therapeutic implications. Irradiation and conventional chemotherapy, the two common modalities of cancer treatment, mediate their effects

by inducing DNA damage and subsequent apoptosis. Tumors with wild-type *TP53* alleles are more likely to be killed by such therapy than tumors with mutated *TP53* alleles. Such is the case with testicular germ cell tumors and childhood acute lymphoblastic leukemias, cancers with excellent clinical outcomes that usually have wild-type *TP53* alleles. In contrast, tumors such as lung cancers and colorectal cancers, which frequently carry *TP53* mutations, are relatively resistant to chemotherapy and irradiation. A second, less obvious but even more nefarious effect is that cells with defective p53 acquire a mutator phenotype, a tendency to acquire mutations at a high rate. Particularly in advanced stage tumors with mutator phenotypes, it is very likely (and perhaps inevitable) that genetically distinct subclones will arise by chance that are resistant to any single therapy, whether radiation, conventional chemotherapy, or molecularly targeted cancer drugs. This theme is discussed later when the enabling properties of genomic instability are discussed more broadly.

KEY CONCEPTS

P53, GUARDIAN OF THE GENOME

- The p53 protein is the central monitor of stress in the cell and can be activated by anoxia, inappropriate signaling by mutated oncoproteins, or DNA damage. p53 controls the expression and activity of proteins involved in cell cycle arrest, DNA repair, cellular senescence, and apoptosis.
- DNA damage is sensed by complexes containing kinases of the ATM/ATR family; these kinases phosphorylate p53, liberating it from inhibitors such as MDM2. Active p53 then upregulates the expression of proteins such as the cyclin-dependent kinase inhibitor p21, thereby causing cell-cycle arrest at the G_1/S checkpoint. This pause allows cells to repair DNA damage.
- If DNA damage cannot be repaired, p53 induces additional events that lead to cellular senescence or apoptosis.
- The majority of human cancers demonstrate biallelic loss-of-function mutations in *TP53*. Rare patients with Li-Fraumeni syndrome inherit one defective copy of *TP53* and have a very high incidence of a wide variety of cancers.
- Like RB, p53 is inactivated by viral oncoproteins, such as the E6 protein of HPV.

Other Tumor Suppressor Genes. The full panoply of tumor suppressor genes is still being defined. Often, they are disabled because they are the targets of recurrent chromosomal deletions, which are now being systematically identified and characterized by sequencing of cancer genomes. Tumor suppressor genes all appear to impact one or more of the hallmarks of cancer. Some that are associated with well-defined clinical syndromes (Table 7.7) or that serve to highlight mechanisms by which tumor suppressors function are described next; others that are organ- or tumor-specific are mentioned in the relevant chapters that follow.

APC: Gatekeeper of Colonic Neoplasia. **Adenomatous polyposis coli (APC) is a member of the class of tumor suppressors that function by downregulating growth-promoting signaling pathways.** Germline loss-of-function mutations involving the *APC* locus on chromosome 5q21 are associated with *familial adenomatous polyposis*, an autosomal dominant disorder in which individuals inheriting one mutant allele develop thousands of adenomatous polyps in the colon during their teens or 20s (Chapter 17). Almost invariably, one or more of these polyps undergoes malignant transformation, giving rise to colon cancer. As with other tumor suppressor genes, both copies of *APC* must be lost for an adenoma to arise. As discussed later, several additional mutations must then occur for adenomas to progress to cancers. In addition to these hereditary forms of colon cancer, 70% to 80% of nonfamilial colorectal carcinomas and sporadic adenomas also show acquired defects involving both APC genes, firmly implicating APC loss of function in the pathogenesis of colonic tumors.

APC is a component of the WNT signaling pathway, which has a major role in controlling cellular growth and differentiation during embryonic development (Fig. 7.28). WNT molecules signal by binding to cell surface receptors of the frizzled (FRZ) family. This stimulates several pathways, the central one involving APC and β-catenin. A major function of the APC protein is to hold β-catenin activity in check. In the absence of WNT signaling, APC participates in the formation a "destruction complex" that leads to the proteasomal degradation of β-catenin. WNT signaling blocks the formation of the destruction complex, stabilizing β-catenin and allowing it to translocate from the cytoplasm to the nucleus. Here, it forms a transcription activation complex with a DNA-binding factor called TCF that promotes the growth of colonic epithelial cells by increasing the expression of *MYC, cyclin D1,* and other genes. Because loss of APC function disrupts the destruction complex, cells that lose APC behave as if they are being continuously stimulated by WNT and show elevated expression of genes that are regulated by β-catenin. The importance of the β-catenin complex in tumorigenesis is attested to by the fact that many colon tumors with normal *APC* genes harbor β-catenin mutations that prevent its APC-dependent destruction, again leading to its accumulation and increased expression of β-catenin–dependent target genes. Thus β-catenin, the target of APC, is itself a proto-oncoprotein. Dysregulation of the APC/β-catenin pathway is not restricted to colon cancers; for example, gain-of-function mutations in β-catenin are present in approximately 20% of hepatocellular carcinomas.

E-Cadherin. β-catenin also binds to the cytoplasmic tail of E-cadherin, a cell surface protein that maintains intercellular adhesiveness. Loss of cell-cell contact, such as following epithelial wounding or injury, disrupts the interaction between E-cadherin and β-catenin. Like WNT signaling, this in turn allows β-catenin to translocate to the nucleus and stimulate genes that promote proliferation, an appropriate response to injury that can help repair a wound. Reestablishment of these E-cadherin contacts as the wound heals leads to sequestration of β-catenin at the membrane and reduces proliferative signaling; these cells are said to be "contact-inhibited." Loss of contact inhibition, by mutation of the E-cadherin/β-catenin axis or by other changes, is a characteristic of many carcinomas. Furthermore, loss of E-cadherin can contribute to the malignant phenotype by allowing easy disaggregation of cells, which can then invade locally or metastasize. Reduced cell surface expression of E-cadherin has been noted in many carcinomas including

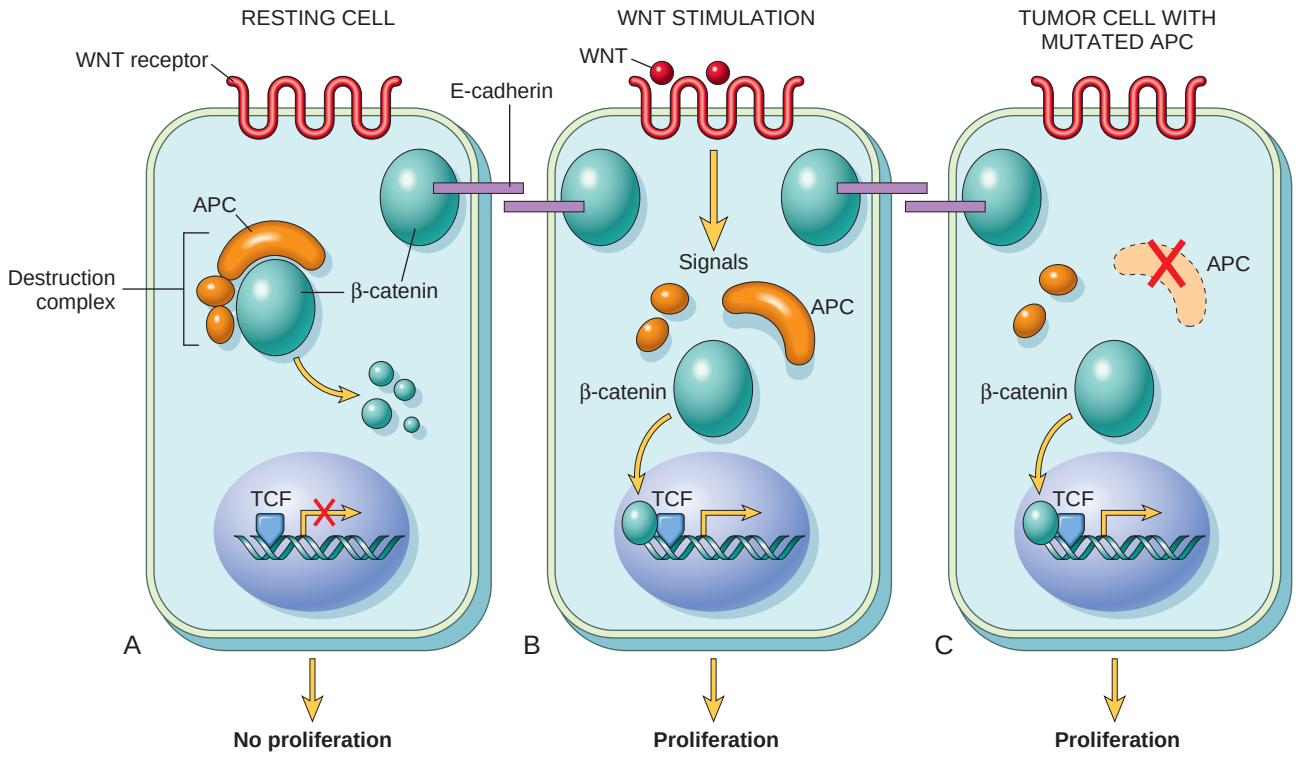


Figure 7.28 The role of adenomatous polyposis coli *(APC)* in regulating the stability and function of β-catenin. APC and β-catenin are components of the WNT signaling pathway. (A) In resting colonic epithelial cells (not exposed to WNT), β-catenin forms a macromolecular complex containing the APC protein. This complex leads to the destruction of β-catenin, and intracellular levels of β-catenin are low. (B) When normal colonic epithelial cells are stimulated by WNT molecules, the *destruction complex* is deactivated, β-catenin degradation does not occur, and cytoplasmic levels increase. β-Catenin translocates to the nucleus, where it binds to TCF, a transcription factor that activates genes involved in cell cycle progression. (C) When *APC* is mutated or absent, as frequently occurs in colonic polyps and cancers, the destruction of β-catenin cannot occur. β-Catenin translocates to the nucleus and coactivates genes that promote entry into the cell cycle, and cells behave as if they are under constant stimulation by the WNT pathway.

those that arise in the esophagus, colon, breast, ovary, and prostate. Germline loss-of-function mutations of the E-cadherin gene, known as *CDH1,* are associated with familial gastric carcinoma, and a proportion of sporadic gastric carcinomas are also associated with loss of E-cadherin expression, which can occur due to mutation of the *CDH1* gene or other indirect mechanisms.

CDKN2A. As mentioned earlier, the *CDKN2A* gene encodes two protein products: the p16/INK4a cyclin-dependent kinase inhibitor, which blocks CDK4/cyclin D–mediated phosphorylation of RB, thereby reinforcing the G_1/S checkpoint; and p14/ARF, which activates the p53 pathway by inhibiting MDM2 and preventing destruction of p53. p16 also appears to be important in induction of cellular senescence (described later). Germline mutations in *CDKN2A* are associated with familial forms of melanoma, and sporadic mutations of this locus have been detected in bladder cancer, head and neck tumors, acute lymphoblastic leukemia, and cholangiocarcinoma. In some tumors, such as cervical cancer, p16 is often silenced by hypermethylation of the gene rather than mutation (see "Epigenetic Changes"). Other cyclin-dependent kinase inhibitors also function as tumor suppressors and are frequently mutated or otherwise silenced in many human malignancies.

TGF-β Pathway. In most normal epithelial, endothelial, and hematopoietic cells, TGF-β is a potent inhibitor of proliferation. It regulates cellular processes by binding to TGF-β receptors, thereby stimulating intracellular signals that involve transcriptional regulators of the SMAD protein family. Under normal circumstances, these signals turn on antiproliferative genes (e.g., genes for cyclin-dependent kinase inhibitors) and turn off genes that drive cell growth (e.g., *MYC,* cyclins, and cyclin-dependent kinases). As can be inferred from our earlier discussion, these changes result in decreased phosphorylation of RB and cell cycle arrest.

In many forms of cancer these growth-inhibiting effects are impaired by loss-of-function mutations in the TGF-β signaling pathway. Mutations affecting TGF-β receptors are common in cancers of the colon, stomach, and endometrium, while mutational inactivation of SMAD4 is common in pancreatic cancers. In many other cancers, loss of TGF-β–mediated growth inhibition occurs at the level of key target genes; examples include mutations that lead to loss of p21 function and/or persistent expression of *MYC*. In such cases, other preserved elements of the TGF-β–induced program of gene expression may actually facilitate acquisition of cancer hallmarks, such as immune evasion or angiogenesis. Thus TGF-β signaling is a double-edged sword that can prevent or promote tumor growth, depending on the state of other genes in the cell.

PTEN. Phosphatase and tensin homologue (PTEN) is a membrane-associated phosphatase encoded by a gene on chromosome 10q23 that is mutated in Cowden syndrome,

an autosomal dominant disorder marked by frequent benign growths such as skin appendage tumors and an increased incidence of epithelial cancers, particularly of the breast (Chapter 21), endometrium, and thyroid. As already mentioned, PTEN acts as a tumor suppressor by serving as a brake on the PI3K/AKT signaling cascade. *PTEN* gene function is lost in many cancers through deletion, deleterious point mutations, or epigenetic silencing.

VHL. Germline loss-of-function mutations of the von Hippel–Lindau *(VHL)* gene on chromosome 3p are associated with hereditary renal cell carcinoma and several other tumors and proliferations. Somatic mutations of *VHL* also occur in a subset of sporadic renal cell carcinomas (Chapter 20). VHL is a component of a protein complex that covalently links ubiquitin chains to specific protein substrates, thereby promoting their degradation by the proteasome. A critical substrate for the VHL ubiquitin ligase is the transcription factor hypoxia-inducible transcription factor 1α (HIF1α). In the presence of oxygen, HIF1α is hydroxylated and binds to VHL, leading to its ubiquitination and degradation. In hypoxic environments the hydroxylation reaction does not occur, and HIF1α escapes recognition by VHL. As a result, HIF1α accumulates in the nuclei of hypoxic cells and turns on target genes, including genes encoding vascular endothelial growth factor (VEGF), a critical angiogenesis factor; PDGF, a potent mitogen; and the glucose transporter GLUT1 and several glycolytic enzymes, factors that contribute to Warburg metabolism (described later). Thus, VHL is part of the system that regulates cellular responses to oxygen levels. Loss-of-function mutations in *VHL* prevent the ubiquitination and degradation of HIF1α, even under normoxic conditions, and are accordingly associated with increased levels of angiogenic growth factors and alterations in cellular metabolism that favor growth.

STK11. The *STK11* gene, also known as *LKB1,* encodes a serine/threonine kinase that is an important regulator of cellular metabolism. Germline loss-of-function *STK11* mutations cause Peutz-Jeghers syndrome, an autosomal dominant disorder associated with benign polyps of the gastrointestinal tract and an increased risk of multiple epithelial cancers, particularly gastrointestinal and pancreatic carcinomas. STK11 has pleiotropic effects on multiple facets of cellular metabolism including glucose uptake, gluconeogenesis, protein synthesis, mitochondrial biogenesis, and lipid metabolism. Sporadic *STK11* loss-of-function mutations are found in diverse carcinomas, a finding pointing to the important role of altered cellular metabolism in the establishment and maintenance of the transformed state (discussed later).

KEY CONCEPTS

MECHANISM OF ACTION OF MAJOR TUMOR SUPPRESSOR GENES

APC: encodes a factor that negatively regulates the WNT pathway in colonic epithelium by promoting the formation of a complex that degrades β-catenin.
- Germline loss-of-function mutations cause familial adenomatous polyposis, an autosomal dominant disorder associated with development of thousands of colonic polyps and early-onset colon carcinoma.
- Acquired somatic *APC* mutations are found in about 70% of sporadic colon carcinomas.

E-cadherin: cell adhesion molecule that plays an important role in contact-mediated growth inhibition of epithelial cells; also binds and sequesters β-catenin, a signaling protein that functions in the WNT pathway.
- Germline loss-of-function mutations in the E-cadherin gene *(CDH1)* cause autosomal dominant familial gastric carcinoma.
- Loss of *CDH1* expression is seen in many sporadic carcinomas; these are associated with loss of contact inhibition, loss of cohesiveness, increased invasiveness, and increased WNT signaling.

CDKN2A: a complex locus that encodes two tumor suppressive proteins, p16/INK4a, a cyclin-dependent kinase inhibitor that augments RB function, and ARF, which stabilizes p53.
- Germline loss-of-function mutations cause autosomal dominant familial melanoma.
- Biallelic loss of function is seen in diverse cancers including leukemias, melanomas, and carcinomas.

TGF-β pathway: potent inhibitor of cellular proliferation in normal tissues.
- Frequent loss-of-function mutations involving TGF-β receptors (colon, stomach, endometrium) or downstream signal transducers (SMADs, pancreas) in diverse carcinomas.
- Complex role in carcinogenesis; may also have a pro-oncogenic role by enhancing the immune evasiveness of tumors.

PTEN: encodes a lipid phosphatase that is an important negative regulator of PI3K/AKT signaling.
- Germline loss-of-function mutations cause Cowden syndrome, an autosomal dominant disorder associated with a high risk of breast and endometrial carcinoma.
- Biallelic loss of function common in diverse cancers.

VHL: encodes a component of a ubiquitin ligase complex that is responsible for degradation of hypoxia-induced factors (HIFs), transcription factors that alter gene expression in response to hypoxia.
- Germline loss-of-function mutations cause von Hippel–Lindau syndrome, an autosomal dominant disorder associated with a high risk of renal cell carcinoma and pheochromocytoma.
- Acquired biallelic loss-of-function mutations are common in sporadic renal cell carcinoma.

Growth-Promoting Metabolic Alterations: The Warburg Effect

Even in the presence of ample oxygen, cancer cells demonstrate a distinctive form of cellular metabolism characterized by high levels of glucose uptake and increased conversion of glucose to lactose (fermentation) via the glycolytic pathway. This phenomenon, called the *Warburg effect* and also known as *aerobic glycolysis,* has been recognized for many years (Otto Warburg received the Nobel Prize in 1931 for discovery of the effect that bears his name). Clinically the "glucose hunger" of tumors is used to visualize them via positron emission tomography (PET) scanning, in which patients are injected with ^{18}F-fluorodeoxyglucose, a nonmetabolizable derivative of glucose that is preferentially taken up into tumor cells (as well as normal, actively dividing

tissues such as the bone marrow). Most tumors are PET-positive, and rapidly growing ones are markedly so.

At the heart of the Warburg effect lies a simple question: why is it advantageous for a cancer cell to rely on seemingly inefficient glycolysis (which generates two molecules of ATP per molecule of glucose) instead of oxidative phosphorylation (which generates 36 molecules of ATP per molecule of glucose)? While pondering this question, it is important to recognize that rapidly growing normal cells, such as in embryonic tissues, also rely on aerobic fermentation. Thus, "Warburg metabolism" is not cancer-specific, but instead is a general property of growing cells that is exploited by cancer cells.

The answer to this riddle is surprisingly simple: **aerobic glycolysis provides rapidly dividing tumor cells with metabolic intermediates that are needed for the synthesis of cellular components, whereas mitochondrial oxidative phosphorylation does not.** The reason growing cells rely on aerobic glycolysis becomes readily apparent when one considers that a growing cell has a strict biosynthetic requirement; it must duplicate all of its cellular components—DNA, RNA, proteins, lipid, and organelles—before it can divide and produce two daughter cells. Recall that the net effect of oxidative phosphorylation is to take a molecule of glucose, $C_6H_{12}O_6$, and combine it with six molecules of O_2 to produce six molecules of H_2O and six molecules of CO_2, which are lost through respiration. Thus, while "pure" oxidative phosphorylation yields abundant ATP, it fails to produce any carbon moieties that can be used to build cellular components that are needed for growth (proteins, lipids, and nucleic acids). Even cells that are not actively growing must shunt some metabolic intermediates away from oxidative phosphorylation to synthesize macromolecules that are needed for cellular maintenance.

By contrast, in actively growing cells, only a small fraction of the cellular glucose is shunted through the oxidative phosphorylation pathway, such that on average each molecule of glucose that is metabolized produces approximately four molecules of ATP (instead of the two molecules that would be produced by "pure" glycolysis). Presumably, this balance in glucose utilization (heavily biased toward aerobic fermentation, with a bit of oxidative phosphorylation) hits a metabolic "sweet spot" that is optimal for growth. It follows that growing cells do rely on mitochondrial metabolism. However, the main metabolic function of mitochondria in growing cells is not to generate ATP, but rather to carry out reactions that generate intermediates that can be diverted for use as precursors in the synthesis of cellular building blocks. For example, lipid biosynthesis requires acetyl coenzyme A (acetyl-CoA), and acetyl-CoA is largely synthesized in growing cells from intermediates such as citrate that are generated in mitochondria.

So how is this reprogramming of metabolism triggered in growing normal and malignant cells? As might be guessed, **metabolic reprogramming is produced by signaling cascades downstream of growth factor receptors, the very same pathways that are deregulated by mutations in oncogenes and tumor suppressor genes in cancers.** Thus, whereas in rapidly growing normal cells aerobic glycolysis ceases when the tissue is no longer growing, in cancer cells this reprogramming persists due to the action of oncogenes and the loss of tumor suppressor gene function. Some of the important points of crosstalk between pro-growth signaling factors and cellular metabolism are shown in Fig. 7.29 and include the following:

- *Receptor tyrosine kinase/PI3K/AKT signaling.* PI3K/AKT signaling upregulates the activity of glucose transporters and multiple glycolytic enzymes, thus increasing glycolysis; promotes shunting of mitochondrial intermediates to pathways leading to lipid biosynthesis; and stimulates factors that are required for protein synthesis. In addition, receptor tyrosine kinases phosphorylate and inhibit pyruvate kinase, which catalyzes the last step in the glycolytic pathway, the conversion of phosphoenolpyruvate to pyruvate. This creates a damming effect that leads to the buildup of upstream glycolytic intermediates, which are siphoned off for synthesis of DNA, RNA, and protein.
- *MYC.* As mentioned, pro-growth pathways upregulate expression of the transcription factor MYC, which drives changes in gene expression that support anabolic metabolism and cell growth. Among the most important metabolic factors that are upregulated by MYC are multiple glycolytic enzymes and glutaminase, which is required for mitochondrial utilization of glutamine, another important source of intermediates needed for biosynthesis of cellular components.

The flip side of the coin is that tumor suppressors often inhibit metabolic pathways that support growth. We have already discussed how the STK11 tumor suppressor antagonizes metabolic changes that produce Warburg metabolism. Indeed, it may be that many (and perhaps all) tumor suppressors that induce growth arrest suppress the Warburg effect. For example, p53, arguably the most important tumor suppressor, upregulates target genes that collectively inhibit glucose uptake, glycolysis, lipogenesis, and the generation of NADPH (a key cofactor needed for the biosynthesis of macromolecules). Thus, it is clear that the functions of many oncoproteins and tumor suppressors are inextricably intertwined with cellular metabolism.

Autophagy. Autophagy is a state of severe nutrient deficiency in which cells not only arrest their growth but also cannibalize their own organelles, proteins, and membranes as carbon sources for energy production (see Fig. 7.29 and Chapter 2). If this adaptation fails, the cells die of starvation. Tumor cells often seem to be able to grow under marginal environmental conditions without triggering autophagy, suggesting that the pathways that induce autophagy are disabled. In keeping with this, several genes that promote autophagy are tumor suppressors, meaning that the loss of autophagy enhances tumor growth. Whether autophagy is always bad from the vantage point of the tumor, however, is a matter of active investigation and debate. For example, under conditions of severe nutrient deprivation, tumor cells may use autophagy to become "dormant," a state of metabolic hibernation that allows cells to survive hard times for long periods. Such cells are believed to be resistant to therapies that kill actively dividing cells and could therefore be responsible for therapeutic failures. Thus, autophagy may be a tumor's friend or foe depending on how the signaling pathways that regulate it are "wired" in a given tumor.

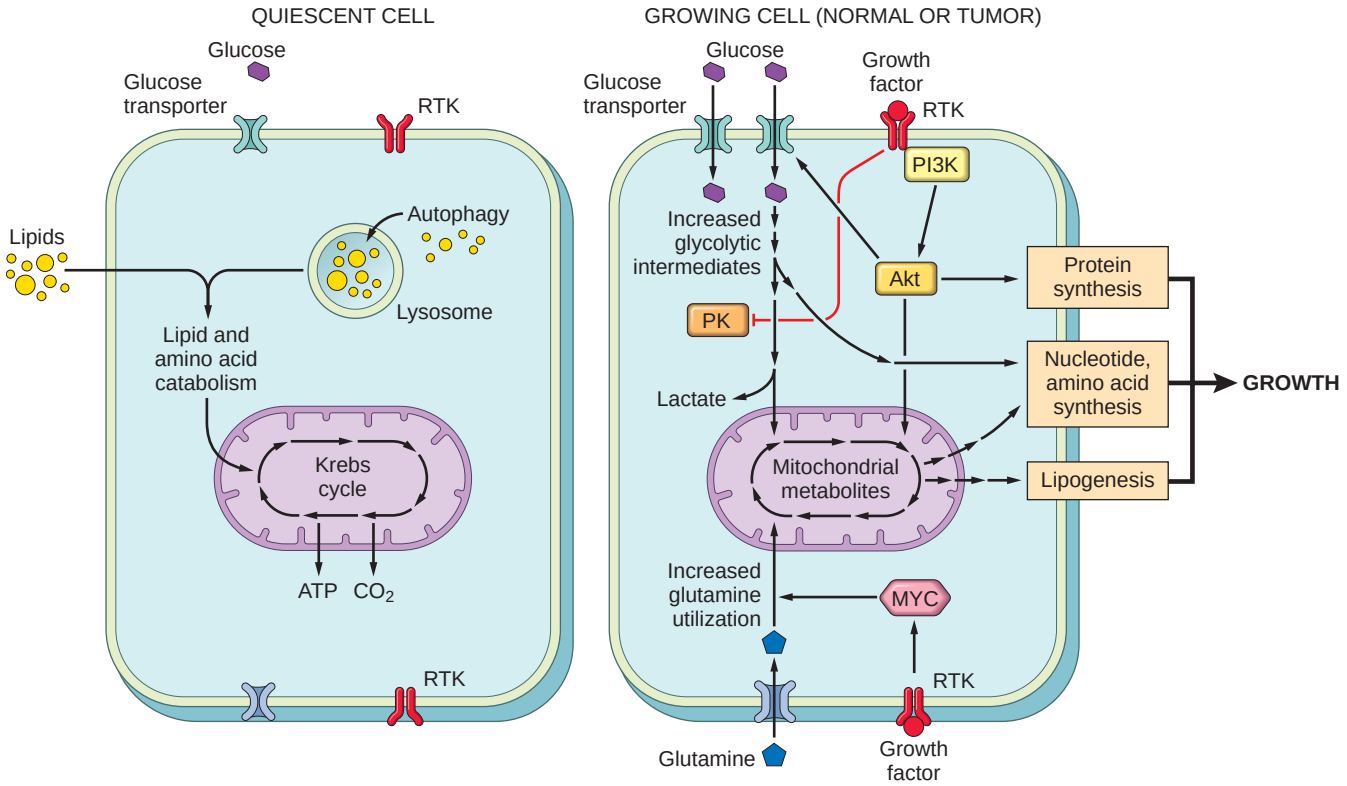


Figure 7.29 Metabolism and cell growth. Quiescent cells rely mainly on the Krebs cycle for adenosine triphosphate *(ATP)* production; if starved, autophagy (self-eating) is induced to provide a source of fuel. When stimulated by growth factors, normal cells markedly upregulate glucose and glutamine uptake, yielding glycolytic and Krebs cycle metabolic intermediates that provide carbon sources for synthesis of nucleotides, proteins, and lipids. In cancers, oncogenic mutations involving growth factor signaling pathways and other key factors such as MYC deregulate these metabolic pathways.

Oncometabolism. A surprising group of genetic alterations discovered through tumor genome sequencing studies consists of mutations in enzymes that participate in the Krebs cycle. Of these, mutations in isocitrate dehydrogenase (IDH) have garnered the most interest, as they have revealed a new mechanism of oncogenesis termed *oncometabolism* (Fig. 7.30).

The proposed steps in the oncogenic pathway involving IDH are as follows:

- IDH acquires a mutation that leads to a specific amino acid substitution involving residues in its active site. As a result, the mutated protein loses it IDH function and instead acquires a new enzymatic activity that catalyzes the production of 2-hydroxyglutarate (2-HG).
- 2-HG in turn acts as an inhibitor of several other enzymes that require a metabolite called α-ketoglutarate as a cofactor. Among the proteins that are inhibited by 2-HG are several members of the TET family, including TET2.
- TET2 is one of several factors that enhance DNA methylation, which you will recall is an epigenetic modification that controls normal gene expression and often goes awry in cancer. According to the model, loss of TET2 activity leads to abnormal patterns of DNA methylation.
- Abnormal DNA methylation in turn leads to misexpression of cancer genes, which drive cellular transformation and oncogenesis. Some data suggest that the net effect of TET2 loss in lineages in which TET2 is a tumor suppressor is the upregulation of RAS and receptor tyrosine kinase signaling.

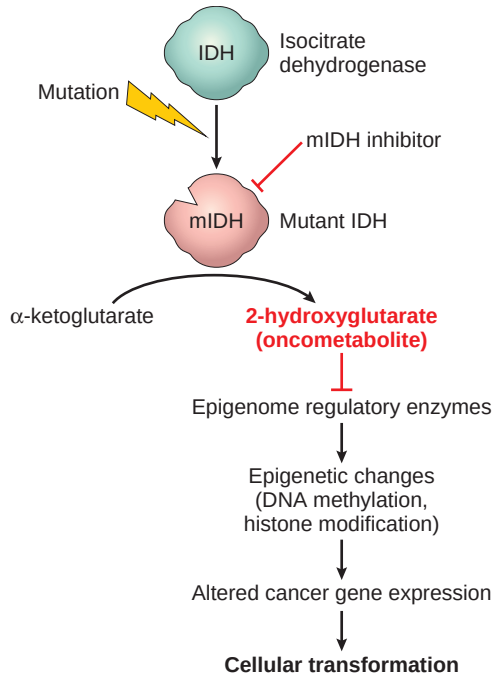


Figure 7.30 Proposed action of the oncometabolite 2-hydroxyglutarate in cancer cells with mutated isocitrate dehydrogenase *(mIDH)*. *IDH*, Isocitrate dehydrogenase.

Thus, according to this scenario, mutated IDH acts by producing 2-HG, which is considered a prototypical *oncometabolite*. Oncogenic IDH mutations occur in a diverse collection of cancers including a sizable fraction of cholangiocarcinomas, gliomas, acute myeloid leukemias, and sarcomas. Of clinical significance, because the mutated IDH proteins have an altered structure, it has been possible to develop drugs that inhibit mutated IDH and not the normal IDH enzyme. These drugs are now being used to treat leukemias with IDH mutations. This story exemplifies how detailed understanding of oncogenic mechanisms can yield entirely new kinds of anticancer drugs.

KEY CONCEPTS

ALTERED CELLULAR METABOLISM

- Warburg metabolism is a form of pro-growth metabolism favoring glycolysis over oxidative phosphorylation. It is induced in normal cells by exposure to growth factors and becomes fixed in cancer cells due to the action of certain driver mutations.
- Many oncoproteins (RAS, MYC, mutated growth factor receptors) induce or contribute to Warburg metabolism, and many tumor suppressors (*PTEN*, *NF1*, *p53*) oppose it.
- Stress may induce cells to consume their components in a process called autophagy. Cancer cells may accumulate mutations to avoid autophagy or may corrupt the process to provide nutrients for continued growth and survival.
- Some oncoproteins such as mutated IDH act by causing the formation of high levels of "oncometabolites" that alter the epigenome, thereby leading to changes in gene expression that are oncogenic.

Evasion of Cell Death

Tumor cells frequently contain mutations in genes that result in resistance to apoptotic cell death. As discussed in Chapter 2, apoptosis, or regulated cell death, refers to an orderly dismantling of cells into component pieces, which are then efficiently consumed by macrophages without stimulating inflammation. You will recall there are two pathways that lead to apoptosis: (1) the extrinsic (death receptor) pathway, triggered by death receptors of the tumor necrosis factor (TNF) receptor family, such as FAS, and their ligands; and (2) the intrinsic (mitochondrial) pathway, initiated by various stresses, such as the absence of growth factors and DNA damage. The intrinsic pathway appears to be the primary arbitrator of life and death in cancer cells, as cancer cells are subject to a number of intrinsic stresses that can initiate apoptosis, particularly DNA damage, but also metabolic disturbances stemming from dysregulated growth, hypoxia caused by insufficient blood supply, and in some cancers increased amounts of misfolded proteins. These stresses are enhanced manyfold when tumors are treated with chemotherapy or radiation (which kill tumor cells mainly by inducing apoptosis). Thus there is strong selective pressure, both before and after therapy, for cancer cells to develop mechanisms to evade apoptosis. This occurs mainly by way of acquired mutations and changes in gene expression that disable key components of the intrinsic pathway or that reset the balance of regulatory factors so as to favor cell survival in the face of intrinsic stresses (Fig. 7.31).

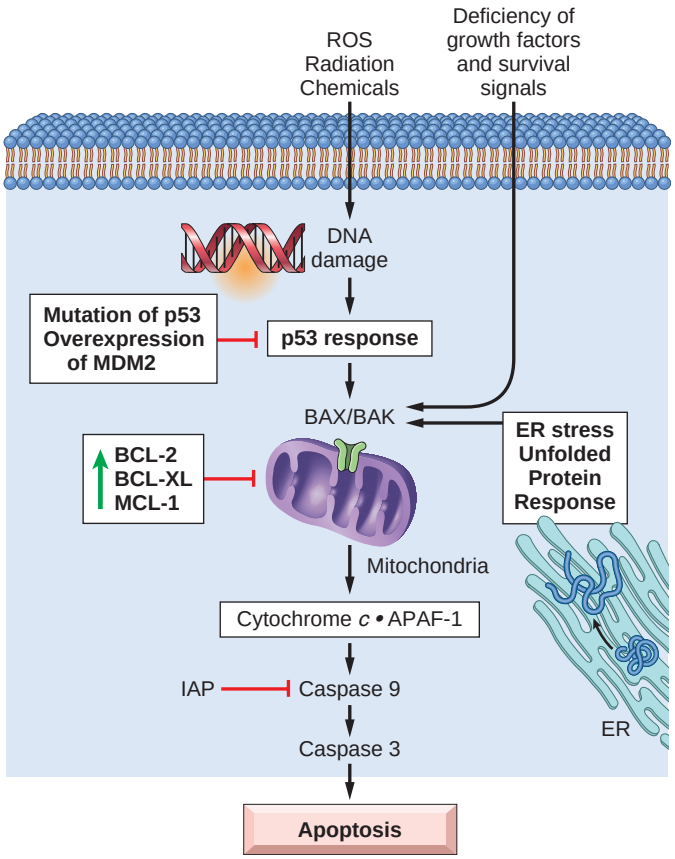


Figure 7.31 Intrinsic pathway of apoptosis and major mechanisms used by tumor cells to evade cell death. The most important mechanisms involve loss of p53 function, either through mutation or through antagonism by MDM2, and reduced egress of cytochrome *c* from mitochondria as a result of upregulation of anti-apoptotic factors that stabilize the mitochondrial membrane such as BCL2, BCL-XL, and MCL-1. Less commonly, tumors suppress apoptosis by unregulating members of the inhibitor of apoptosis *(IAP)* family.

Before delving into modes of resistance to apoptosis, a brief review of the intrinsic pathway is in order. Activation of this pathway leads to permeabilization of the mitochondrial outer membrane and release of molecules, such as cytochrome *c,* that initiate apoptosis. The integrity of the mitochondrial outer membrane is determined by a delicate balancing act between pro-apoptotic and anti-apoptotic members of the BCL2 protein family. The pro-apoptotic proteins BAX and BAK are required for apoptosis and directly promote mitochondrial permeabilization. Their action is inhibited by the anti-apoptotic members of this family, which are exemplified by BCL2 and BCL-XL. A third set of proteins, so-called *BH3-only proteins,* which include BIM, BAD, BID, and PUMA, shift the balance between the pro-apoptotic and anti-apoptotic family members by neutralizing the actions of anti-apoptotic proteins like BCL2 and BCL-XL, thereby promoting apoptosis. When the sum total of all BH3 proteins expressed "overwhelms" the anti-apoptotic BCL2/ BCL-XL protein barrier, BAX and BAK are activated and form pores in the mitochondrial membrane. This allows cytochrome *c* to leak into the cytosol, where it

binds to APAF-1 and activates caspase-9, which in turn cleaves and activates the executioner caspases. Another check on this system consists of a group of proteins called *inhibitor of apoptosis proteins (IAPs),* which block the action of caspase-9.

Within this framework, it is possible to illustrate the two major mechanisms by which apoptosis is avoided by cancer cells (see Fig. 7.31):

- *Loss of TP53 function*. While *TP53* is commonly mutated in cancers at diagnosis, the frequency of *TP53* mutations is even higher in tumors that relapse after therapy, probably because of their ability to convey resistance to genotoxic therapies. Other lesions in cancer impair p53 function indirectly, most notably amplification of *MDM2,* which encodes an inhibitor of p53. Loss of p53 function prevents the upregulation of PUMA, a pro-apoptotic BH3-only protein, in response to DNA damage and other stresses, allowing cells to survive that otherwise would be killed.
- *Overexpression of anti-apoptotic members of the BCL2 family*. Overexpression of BCL2 is a common event leading to the protection of tumor cells from apoptosis. One of the best-understood examples is found in follicular lymphoma (Chapter 12), a B-cell tumor carrying a characteristic (14;18)(q32;q21) translocation that fuses the *BCL2* gene to the transcriptionally active immunoglobulin heavy chain gene. The resulting overabundance of BCL2 protects the transformed lymphocytes from apoptosis. Because BCL2-overexpressing follicular lymphomas arise in large part through reduced cell death rather than explosive cell proliferation, they tend to be indolent (slow-growing). In other tumors such as chronic lymphocytic leukemia (Chapter 12), it appears that BCL2 is upregulated due to loss of specific micro-RNAs that normally restrain BCL2 levels. Many other mechanisms leading to overexpression of anti-apoptotic members of the BCL2 family have been proposed in a wide variety of cancers.

Recognition of the mechanisms by which cancers evade cell death has stimulated several lines of targeted drug development. Restoration of p53 function in *TP53*-mutated tumors is a daunting problem (because of the inherent difficulty of "fixing" defective genes) but is possible in tumors in which p53 is inactive because of overexpression of its inhibitor, MDM2. Indeed, inhibitors of MDM2 that reactivate p53 and induce apoptosis in tumors with *MDM2* gene amplification are being tested in clinical trials. Even more impressive results have been generated with drugs that inhibit the function of anti-apoptotic members of the BCL2 family, particularly BCL2 itself. These drugs have potent activity against certain tumors characterized by BCL2 overexpression (such as chronic lymphocytic leukemia) and are beoming a standard part of the treatment of particular cancers.

KEY CONCEPTS

EVASION OF CELL DEATH

- Apoptosis can be initiated through intrinsic or extrinsic pathways, both of which result in the activation of a proteolytic cascade of caspases that destroys the cell.
- Lesions that incapacitate the intrinsic (mitochondrial) pathway appear to be most common in cancers.
- In greater than 85% of follicular B-cell lymphomas, the anti-apoptotic gene *BCL2* is overexpressed due to a (14;18) translocation.
- Overexpression of other BCL2 family members is also linked to cancer cell survival and drug resistance.

Limitless Replicative Potential: The Stem Cell–Like Properties of Cancer Cells

All cancers contain cells that are immortal and have limitless replicative potential. Some cell lines established from cancers have now been proliferating ceaselessly in laboratories for more than 60 years, and it is reasonable to expect that they will continue to grow for as long as there are scientists to tend to them. How can it be that cancer cells have seemingly discovered the proverbial fountain of eternal youth? The answers are not completely known, but three interrelated factors appear to be critical to the immortality of cancer cells: (1) evasion of senescence, (2) evasion of mitotic crisis, and (3) the capacity for self-renewal.

- *Evasion of senescence.* As discussed in Chapter 2, most normal human cells have the capacity to divide 60 to 70 times. After this, the cells become senescent, permanently leaving the cell cycle and never dividing again. The mechanisms that produce senescence are not well understood, but the senescent state is associated with upregulation of p53 and INK4a/p16 (perhaps in response to the accumulation of DNA damage over time). These are believed to contribute to senescence in part by maintaining RB in a hypophosphorylated state, which favors cell cycle arrest. As already discussed, the RB-dependent G_1/S cell cycle checkpoint is disrupted in virtually all cancers by a wide variety of acquired genetic and epigenetic aberrations that may allow cells to bypass senescence.
- *Evasion of mitotic crisis.* While cells that are resistant to senescence have increased replicative capacity, they are not immortal; instead, they eventually enter a phase referred to as *mitotic crisis* and die. Mitotic crisis has been ascribed to progressive shortening of *telomeres,* special DNA sequences at the ends of chromosomes that bind several types of protective protein complexes (Chapter 2). Most somatic cells do not express *telomerase,* the enzyme that is responsible for the maintenance of telomeres, and with each cell division their telomeres shorten. When the telomeric DNA is completely eroded, the exposed chromosome ends are "sensed" as double-stranded DNA breaks. If the affected cells have functional p53, the cell arrests its growth and may undergo apoptosis, but if p53 is dysfunctional, the nonhomologous end-joining pathway is activated and may join the "naked" ends of two chromosomes. The resulting dicentric chromosomes are broken during attempted mitotic segregation during anaphase, producing new double-stranded DNA breaks. The snowballing genomic damage caused by repeated "bridge-fusion-breakage" cycles eventually produces mitotic catastrophe and cell death (Fig. 7.32). Telomerase is expressed at very low levels in most somatic cells, and thus proliferating cells that escape from senescence are

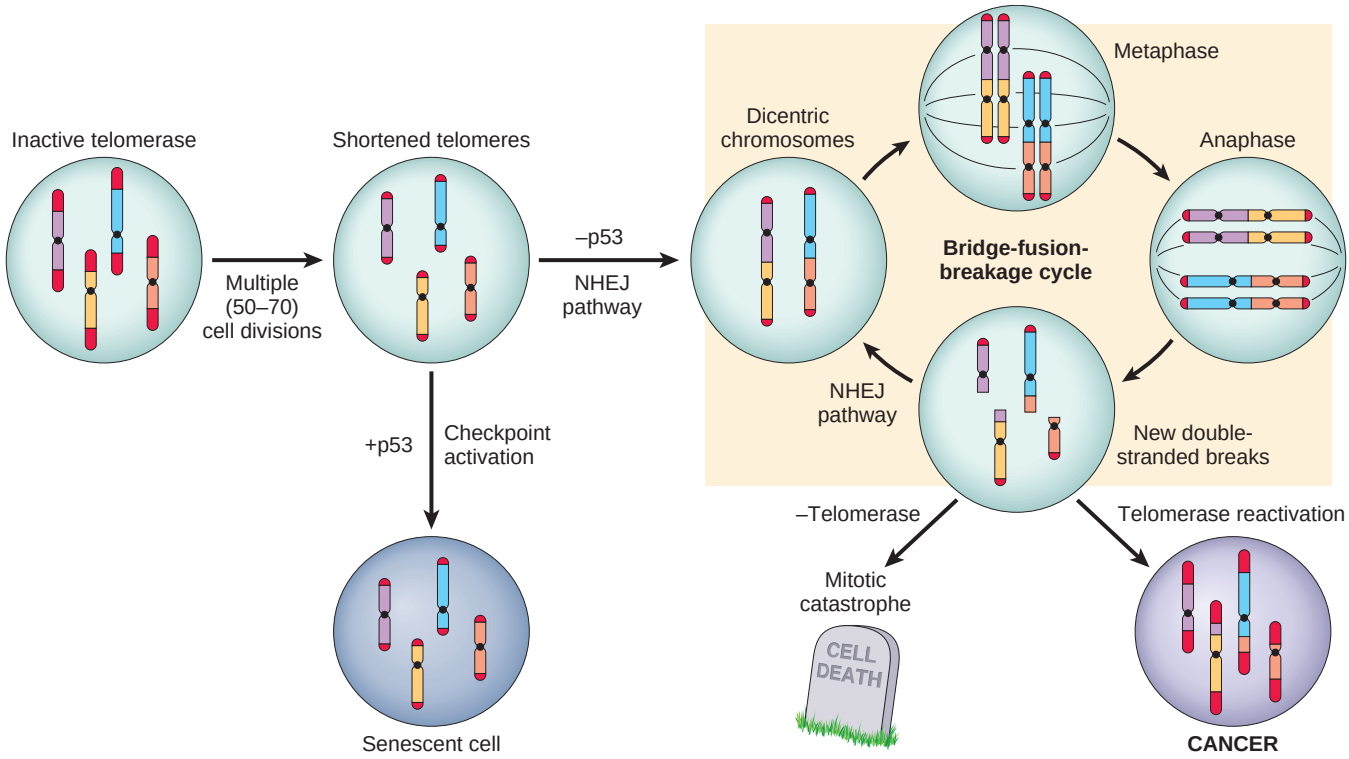


Figure 7.32 Escape of cells from senescence and mitotic catastrophe caused by telomere shortening. Replication of somatic cells, which do not express telomerase, leads to shortened telomeres. In the presence of competent checkpoints, cells undergo arrest and enter nonreplicative senescence. In the absence of checkpoints, DNA repair pathways such as the nonhomologous end joining *(NHEJ)* pathway are inappropriately activated, leading to the formation of dicentric chromosomes. At mitosis the dicentric chromosomes are pulled apart, generating random double-stranded breaks, which then activate DNA repair pathways, leading to the random association of double-stranded ends and the formation, again, of dicentric chromosomes. Cells undergo numerous rounds of this bridge-fusion-breakage cycle, which generates massive chromosomal instability and numerous mutations. If cells fail to reexpress telomerase, they eventually undergo mitotic catastrophe and death. Reexpression of telomerase allows the cells to escape the bridge-fusion-breakage cycle, thus promoting their survival and tumorigenesis.

likely to expire in this fashion. However, cells in crisis that reactivate telomerase can restore their telomeres and survive. Such cells may have suffered damage to oncogenes and tumor suppressor genes during crisis and are at high risk for malignant transformation. Alternatively, cancers may arise from stem cells (described later), which are long-lived in part because they express telomerase. Whatever the mechanism, telomere maintenance is seen in virtually all types of cancers. In 85% to 95% of tumors it is due to expression of telomerase. The remaining tumors use another mechanism to maintain their telomeres termed alternative lengthening of telomeres that depends on DNA recombination.

- *Self-renewal.* Unlike most cells, tissue stem cells and germ cells express telomerase, making them resistant to mitotic crisis, and also somehow avoid the genetic and epigenetic alterations that trigger senescence. In addition, long-lived stem cells possess another critical property, the capacity for self-renewal. In simple terms, self-renewal means that each time a stem cell divides at least one of the two daughter cells remains a stem cell. In a symmetric division, both daughter cells remain stem cells; such divisions may occur during embryogenesis, when stem cell pools are expanding, or during times of stress. In an asymmetric division, only one daughter cell remains a stem cell; in such circumstances, the non–stem cell daughter proceeds along some differentiation pathway, losing "stemness" but gaining one or more functions in the process. Cells in "transit" to a differentiated state are often highly proliferative, but they eventually differentiate, stop dividing, and may eventually become senescent or undergo apoptosis. The continued growth and maintenance of many tissues that contain short-lived cells, such as the formed elements of the bone marrow and blood and the epithelial cells of the gastrointestinal tract and skin, depend on a resident population of tissue stem cells that are capable of self-renewal. Following on this logic, because cancers are immortal and have limitless proliferative capacity, they too must contain cells that self-renew—so-called *cancer stem cells*. While there remains debate about the identity and number of stemlike cells in particular cancers, it is accepted that cells resembling stem cells must exist in all cancers.

Another open question is whether cancer stem cells arise from the transformation of tissue stem cells or from the conversion of conventional somatic cells to transformed cells with the acquired property of "stemness." The answer seems to be that both scenarios occur in different types of tumors, as exemplified by chronic myeloid leukemia and acute promyelocytic leukemia (Fig. 7.33). Recall also that expression of a small number of transcription factors can

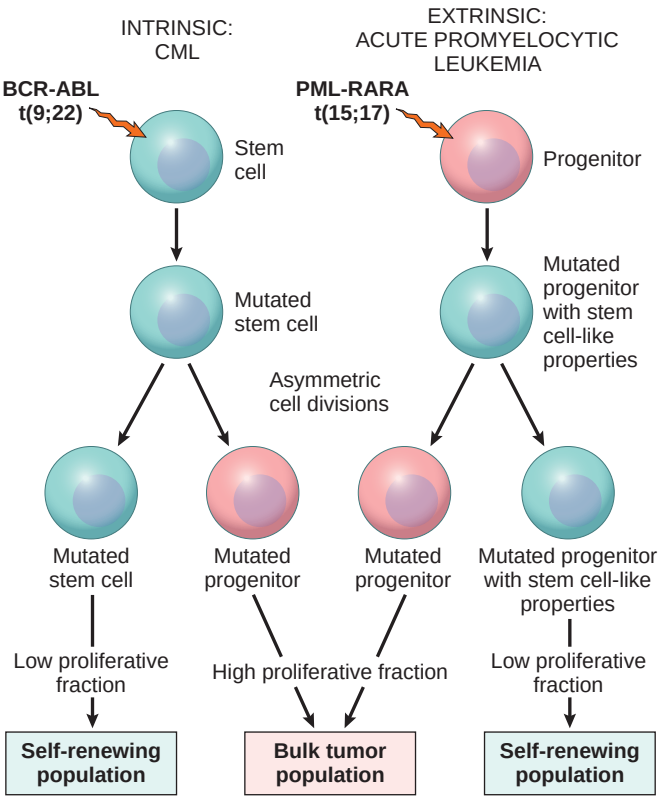


Figure 7.33 Origins of cells with self-renewing capacity in cancer. Cancer stem cells can arise from transformed tissue stem cells (e.g., hematopoietic stem cells in chronic myeloid leukemia *[CML]*) with intrinsic "stemness" or from proliferating cells that acquire a mutation that confers stemness (e.g., granulocyte progenitors in acute promyelocytic leukemia). In both instances, the cancer stem cells undergo asymmetric cell divisions that give rise to committed progenitors that proliferate more rapidly than the cancer stem cells; as a result, most of the malignant cells in both tumors lack self-renewing capacity.

result in the epigenetic reprogramming of a differentiated somatic cell such as a fibroblast into a pluripotent stem cell. Thus it is easy to imagine how mutations leading to misexpression of certain key transcription factors, such as MYC, might convert a somatic cell into a transformed cell with a capacity for self-renewal. A corollary of this idea is that, unlike normal stem cells and their more differentiated progeny, which have a fixed parent–offspring relationship, cancer cells within a tumor may be able to dedifferentiate to a stem cell–like state. Indeed, there is evidence that cancers can repopulate their stem cell pools from non–stem cell populations, further complicating efforts to precisely define and selectively target cancer stem cells.

Despite these uncertainties, the concept of cancer stem cells has important implications for cancer therapy. Most notably, if cancer stem cells are essential for tumor persistence, it follows that these cells must be eliminated to eradicate the tumor. It is hypothesized that like normal stem cells, cancer stem cells have a high intrinsic resistance to conventional therapies due to a low rate of cell division and the expression of factors such as multiple drug resistance-1 (MDR1) that counteract the effects of chemotherapeutic drugs. Thus the limited success of current therapies may in part be explained by their failure to kill the malignant stem cells that lie at the root of cancer.

KEY CONCEPTS

LIMITLESS REPLICATIVE POTENTIAL

- At least some cells in all cancers must be stem cell–like; these cells are sometimes referred to as cancer stem cells. These may arise through transformation of a normal stem cell or through acquired genetic lesions that impart a stem-like state on a more mature cell.
- Cancer cells acquire lesions that inactivate senescence signals and reactivate telomerase, which act together to convey limitless replicative potential.

Angiogenesis

Even if a solid tumor possesses all the genetic aberrations that are required for malignant transformation, it cannot enlarge beyond 1 to 2 mm in diameter unless it has the capacity to induce angiogenesis. Like normal tissues, the "health" of tumors requires delivery of oxygen and nutrients and removal of waste products; presumably the 1- to 2-mm size limit is the maximal distance across which oxygen, nutrients, and waste can diffuse from existing blood vessels. Growing cancers stimulate neoangiogenesis, during which new vessels sprout from capillaries (Chapter 3). Neovascularization supplies needed nutrients and oxygen, and proliferating endothelial cells also stimulate the growth of adjacent tumor cells by secreting growth factors such as insulin-like growth factors (IGFs) and PDGF. While the resulting tumor vasculature is effective at delivering nutrients and removing wastes, it is not entirely normal; the vessels are leaky and dilated and have a haphazard pattern of connection, features that can be appreciated on angiograms. By permitting tumor cells ready access to these abnormal vessels, angiogenesis also contributes to metastasis. Angiogenesis is thus an essential facet of malignancy.

How do growing tumors develop a blood supply? The current paradigm is that **angiogenesis is controlled by a balance between angiogenesis promoters and inhibitors; in angiogenic tumors this balance is skewed in favor of promoters.** Early in their development, most human tumors do not induce angiogenesis. Starved of nutrients, these tumors remain small or in situ, possibly for years, until an *angiogenic switch* terminates this stage of quiescence. The molecular basis of the angiogenic switch involves increased local production of angiogenic factors and/or loss of angiogenic inhibitors. The sources of these factors include tumor cells, infiltrating inflammatory cells (e.g., macrophages) or other tumor-associated stromal cells, and the ECM. In the case of tumor cells, several alterations enhance the production of pro-angiogenic factors:

- *Relative lack of oxygen due to hypoxia stabilizes HIF1α,* an oxygen-sensitive transcription factor mentioned earlier that activates the transcription of the pro-angiogenic cytokines VEGF and basic fibroblast growth factor (bFGF). These factors create an angiogenic gradient that stimulates the proliferation of endothelial cells and guides the growth of new vessels toward the tumor (Chapter 3).
- *Driver mutations in certain tumor suppressors and oncogenes favor angiogenesis.* For example, p53 stimulates the expression of anti-angiogenic molecules such as thrombospondin-1

and represses the expression of pro-angiogenic molecules such as VEGF. Thus, loss of p53 not only removes cell cycle checkpoints and alters tumor cell metabolism but also provides a more permissive environment for angiogenesis. Conversely, gain-of-function mutations in *RAS* or *MYC* upregulate the production of VEGF.

- *Proteases,* which may be elaborated by either tumor cells or by stromal cells, also influence the local balance of angiogenic and antiangiogenic factors. Many proteases release bFGF from the ECM, which constitutes a storage site for this factor, while other proteases release anti-angiogenic factors such as angiostatin and endostatin through proteolytic cleavage of plasminogen and collagen, respectively.

The idea that angiogenesis is essential for solid tumors to grow to clinically significant sizes provided a powerful impetus for the development of therapeutic agents that block angiogenesis. These agents are now a part of the armamentarium that oncologists use against cancers. A cardinal example is bevacizumab, a monoclonal antibody that neutralizes VEGF activity and is approved for treatment of multiple cancers. However, angiogenesis inhibitors have not been nearly as effective as was hoped; they can prolong life, but usually for only a few months and at very high financial cost. The mechanisms that underlie the persistence and ultimate progression of cancers in the face of angiogenesis inhibitors are not yet clear. Improvements will be possible only with greater understanding of the "escape routes" through which tumor cells sidestep the effects of the angiogenesis inhibitors that are now in use.

KEY CONCEPTS

ANGIOGENESIS

- Vascularization of tumors is essential for their growth and is controlled by the balance between angiogenic and anti-angiogenic factors that are produced by tumor and stromal cells.
- Hypoxia triggers angiogenesis through the actions of HIF1α on the transcription of the pro-angiogenic factor VEGF.
- Many other factors regulate angiogenesis; for example, p53 induces synthesis of the angiogenesis inhibitor thombospondin-1, while RAS, MYC, and MAPK signaling all upregulate VEGF expression and stimulate angiogenesis.
- VEGF inhibitors are used to treat a number of advanced cancers and prolong the clinical course, but are not curative.

Invasion and Metastasis

Invasion and metastasis are the major causes of cancer-related morbidity and mortality and hence are the subjects of intense investigation. Local invasion of tumor cells may damage or destroy vital structures and is a prerequisite for distant spread. Studies in mice and humans reveal that although many of these locally invasive cells enter the bloodstream each day, very few produce metastases. Why is the metastatic process so inefficient? The answer is uncertain but undoubtedly relates to the complexity of the process. For cancer cells to emerge from a primary mass, enter blood vessels or lymphatics, and produce a secondary growth at a distant site, they must go through a series of steps that involve an intricate interplay between tumor cells and many different types of host cells and factors (summarized in Fig. 7.34). At each point in this sequence, the breakaway cells must overcome the challenges of avoiding immune defenses (discussed later) and adapting to a microenvironment (e.g., lymph node, bone marrow, or brain) that is quite different from that of the site of origin of the tumor. The complexity of this series of events may explain why individual "metastasis genes" have not been found; it may be that the "metastatic phenotype" requires the

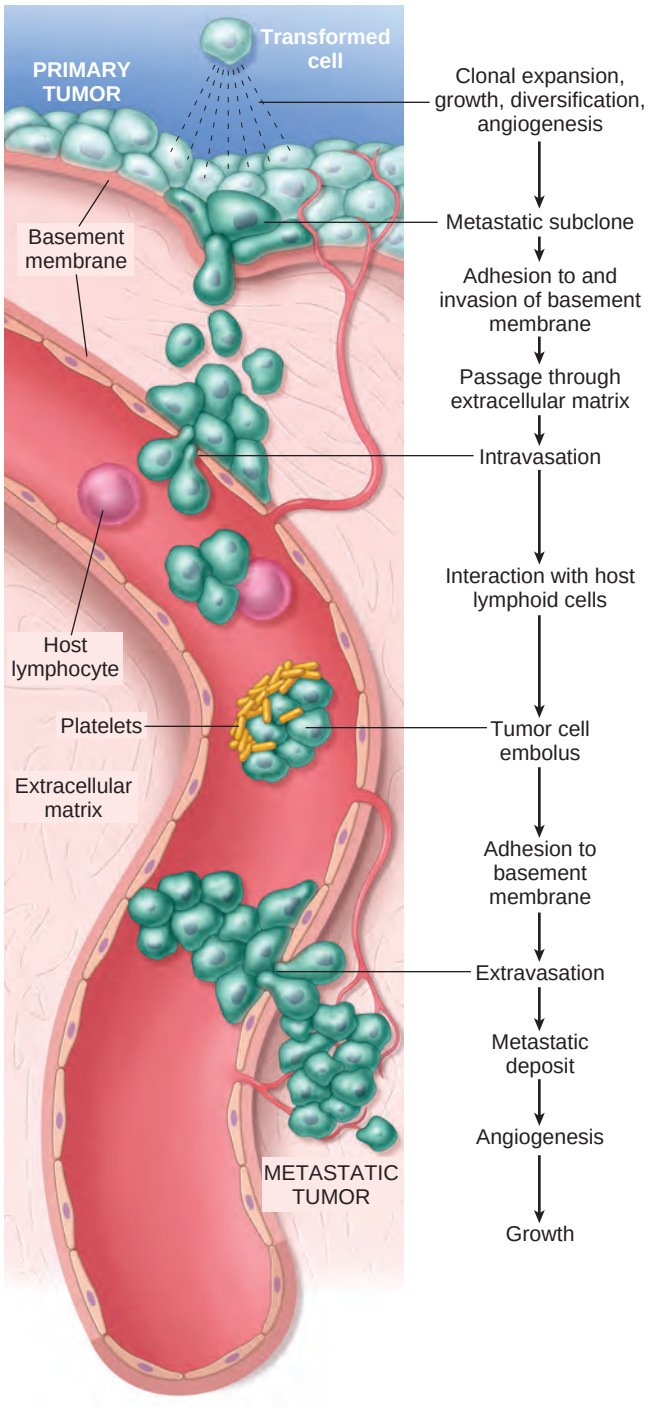


Figure 7.34 The metastatic cascade. Sequential steps involved in the hematogenous spread of a tumor.

accumulate complementary genetic and epigenetic alterations that collectively promote the metastatic cascade. This set of "skills" may be present only in rare cells, or might even require a collaboration between subclones, with each providing some needed function. This latter idea implies that successful metastases may arise from cells that migrate as cohesive groups, for which there is some evidence.

In the discussion that follows, the metastatic cascade is divided into two phases: (1) invasion of the ECM and (2) vascular dissemination, tissue homing, and colonization. Throughout, we touch on some of the proposed molecular mechanisms that underlie the process.

Invasion of Extracellular Matrix

The structural organization and function of normal tissues is determined by interactions between cells and the ECM. As discussed in Chapter 1, tissue compartments are separated from each other by two types of ECM, basement membrane and interstitial connective tissue, each made up of different combinations of collagens, glycoproteins, and proteoglycans. As shown in Fig. 7.34, tumor cells interact with the ECM at several stages in the metastatic cascade. To metastasize, carcinoma cells must breach the underlying basement membrane, traverse the interstitial connective tissue, and ultimately gain access to the circulation by penetrating the vascular basement membrane. This process is repeated in reverse when tumor cells extravasate at a distant site. Invasion of the ECM initiates the metastatic cascade and is an active process that can be resolved into several steps (Fig. 7.35):

- "Loosening up" of tumor cell–tumor cell interactions
- Degradation of ECM
- Attachment to "remodeled" ECM components
- Migration and invasion of tumor cells

Dissociation of cancer cells from one another is often the result of alterations in intercellular adhesion molecules and is the first step in the process of invasion. Normal epithelial cells are tightly glued to each other and to the ECM by a variety of adhesion molecules. As discussed earlier, *E-cadherins* are transmembrane glycoproteins that mediate the homotypic adhesion of epithelial cells, serving both to hold the cells together and to relay signals between cells. In several epithelial tumors including certain adenocarcinomas of the stomach and breast, E-cadherin function is lost due to pathogenic mutations, and in many other epithelial cancers it is hypothesized that E-cadherin expression is silenced, at least transiently, through a process called *epithelial-mesenchymal transition (EMT)*. It is postulated that EMT is integral to the metastasis of carcinomas, particularly breast and prostate cancers. EMT is controlled by the transcription factors SNAIL and TWIST and is defined not only by the downregulation of epithelial markers (e.g., E-cadherin) but also by the concomitant upregulation of mesenchymal markers (e.g., vimentin and smooth muscle actin), changes that are believed to favor the development of a promigratory phenotype that is essential for metastasis.

Degradation of the basement membrane and interstitial connective tissue is the second step in invasion. Tumor cells may accomplish this by secreting proteolytic enzymes or by inducing stromal cells (e.g., fibroblasts and inflammatory cells) to do so. Many different proteases such as matrix metalloproteinases (MMPs), cathepsin D, and urokinase plasminogen activator are overexpressed in tumors and have been implicated in tumor cell invasion. MMPs regulate tumor invasion not only by remodeling the basement membrane and interstitial connective tissue, but also by releasing factors that contribute to the malignant behavior

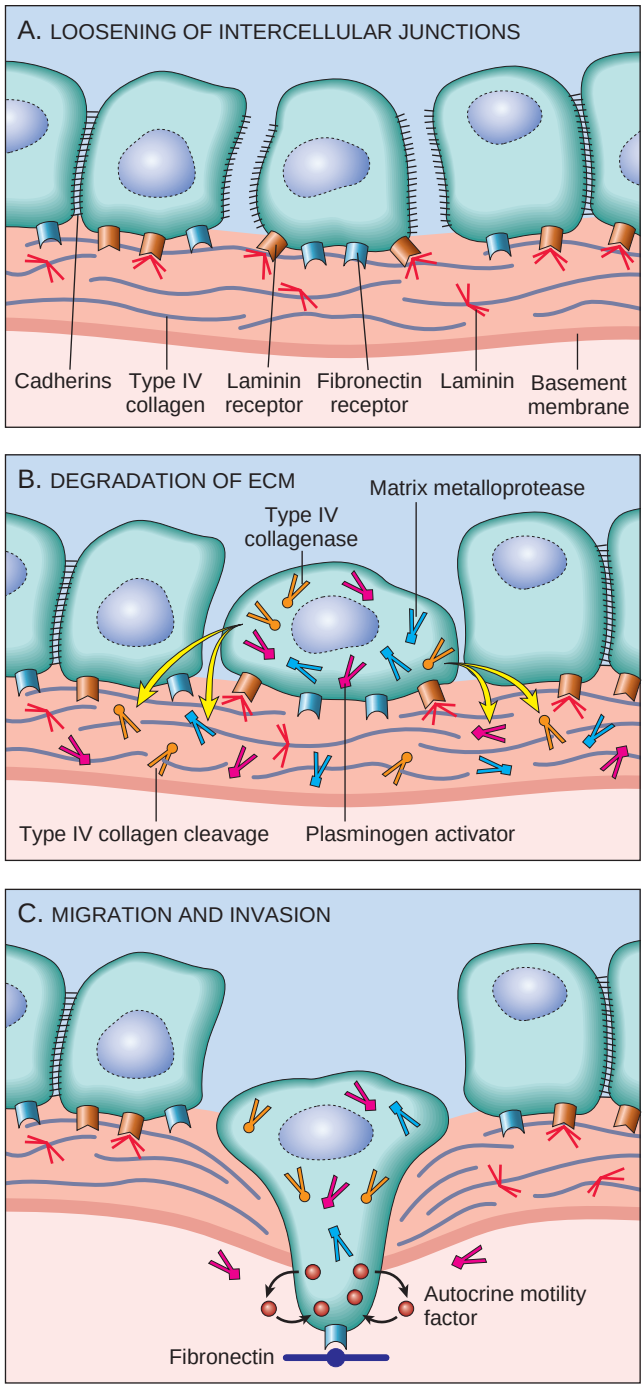


Figure 7.35 Sequence of events in the invasion of epithelial basement membranes by tumor cells. Tumor cells detach from each other because of reduced adhesiveness and attract inflammatory cells. Proteases secreted from tumor cells and inflammatory cells degrade the basement membrane. Binding of tumor cells to proteolytically generated binding sites and tumor cell migration follow. *ECM*, Extracellular matrix.

of cancers. For example, MMP-9, a gelatinase that cleaves type IV collagen found within the epithelial and vascular basement membrane, also stimulates the release of VEGF from ECM-sequestered pools and generates collagen and proteoglycan cleavage products with chemotactic, angiogenic, and growth-promoting effects. Benign tumors of the breast, colon, and stomach have little MMP-9 activity, whereas their malignant counterparts overexpress this enzyme. Concurrently the concentrations of metalloproteinase inhibitors are also reduced in many cancers, further tilting the balance toward tissue degradation.

The third step in invasion involves changes in how tumor cells attach to ECM proteins. Tumor cells demonstrate complex changes in the expression of integrins, which you will recall are transmembrane proteins that participate in adhesion of cells to other cells and to ECM (Chapter 3). In normal epithelial cells, integrins that bind basement membrane laminin and collagens are strictly restricted to the basal aspect of the cell; these receptors help to maintain the cells in a resting, polarized state. Loss of adhesion in normal cells leads to induction of apoptosis, but free tumor cells are resistant to this form of cell death (termed *anoikis,* meaning without a home), in part because of expression of other integrins that mitigate the loss of adhesion to ECM, apparently by transmitting signals that promote cell survival. Additionally, the matrix itself is modified in ways that promote invasion and metastasis. For example, cleavage of the basement membrane proteins collagen IV and laminin by MMP-2 or MMP-9 generates novel sites that bind to receptors on tumor cells and stimulate migration.

Locomotion is the final step of invasion, propelling tumor cells through the degraded basement membranes and zones of matrix proteolysis. Migration is a multistep process that involves many families of receptors and several signaling pathways that eventually impinge on the actin cytoskeleton. Cells must attach to the matrix at their leading edge, detach from the matrix at their trailing edge, and contract the actin cytoskeleton to ratchet forward. Such movement seems to be stimulated and directed by several types of factors, which likely vary among different types of tumors. These include:

- Tumor cell–derived cytokines including chemokines and growth factors (e.g., insulin-like growth factors), which act as autocrine motility factors
- Cleavage products of matrix components (e.g., collagen, laminin)
- Stromal cell–derived paracrine factors such as hepatocyte growth factor/scatter factor, which binds to the receptor tyrosine kinase MET on tumor cells and stimulates motility

This initial phase of metastasis culminates in penetration through the endothelial basement membrane and transmigration into the vascular space. Throughout this phase of the process, tumor cells interact not only with ECM but also with several types of stromal cells, including innate and adaptive immune cells, fibroblasts, and endothelial cells. Details remain to be worked out, but it is clear that "successful" invaders induce signals that modify stromal cells in ways that support their malignant behavior. For example, under the influence of invading cancer cells, so-called cancer-associated fibroblasts alter their expression of genes that encode ECM molecules, proteases, protease inhibitors, and various growth factors. It is easy to imagine that as tumors evolve over time, they come to be dominated by cancer cells that are most effective at co-opting the complex, ever-changing tumor microenvironment to serve their malignant purposes.

Vascular Dissemination, Homing, and Colonization

Once in the circulation, tumor cells are vulnerable to destruction by a variety of mechanisms including mechanical shear stress, apoptosis due to anoikis, and innate and adaptive immune defenses. Nevertheless, viable circulating tumor cells are not rare in patients with solid tumors such as carcinomas. What then separates those few cells that give rise to metastases from the large number of cells that fail? Also, what determines why metastases ultimately appear where they do in the body?

While definitive answers to these questions are not known, clues have emerged. One idea that has gained support proposes that the circulating cells that establish metastases are much more likely to migrate as multicellular aggregates than as single cells. Clumping of tumor cells in the blood is promoted by homotypic interactions as well as heterotypic interactions between tumor cells and blood elements, particularly platelets (see Fig. 7.34), which are believed to enhance tumor cell survival in the circulation. Tumor cells may also express anionic substances such as polyphosphate that activate factor XII (contact factor), resulting in fibrin deposition and further stabilization of tumor emboli, which may enhance the ability of the cells to arrest en masse within capillary beds. Another potential advantage possessed by tumor cells circulating as a group is that they may be far more likely than any single cell to possess all the properties that are needed to establish a metastasis. Among these essential attributes is the presence of cells with stem cell–like properties, which may contribute not only to the relentless growth of metastatic lesions but also to the "plasticity" that is needed for metastatic cells to adapt to growth in a new microenvironment.

Where circulating tumor cells arrest and eventually form clinically significant metastatic deposits appears to relate to three factors: (1) location and vascular drainage of the primary tumor; (2) tropism of particular kinds of tumor cells for specific tissues; and (3) escape from tumor dormancy. The first is a matter of simple anatomy; thus colon carcinomas are far more likely to give rise to metastases in the liver, the first organ downstream of the tumor, than to metastases elsewhere. However, many exceptions to this "rule" exist. For example, carcinomas of the prostate and breast preferentially spread to bone, bronchogenic carcinomas tend to involve the adrenals and the brain, and neuroblastomas spread to the liver and bones. Such organ tropism may be related to the following mechanisms:

- Tumor cells may express adhesion molecules whose ligands are found preferentially on the endothelial cells of the target organ. Of interest in this regard is the CD44 adhesion molecule, which is expressed on normal T lymphocytes and is used by these cells to migrate to selective sites in lymphoid tissues. Such migration is accomplished by the binding of CD44 to hyaluronate on high endothelial venules. Solid tumors also often express

CD44, which appears to enhance their spread to lymph nodes and other metastatic sites.
- Some cancer cells express chemokine receptors, which may guide tumor cells to tissues expressing chemokines, much in the way chemokines normally act as attractants for cells of the immune system.
- Some tissues may provide a favorable "soil" for the growth of tumor seedlings. According to this "seed-soil" hypothesis, originally proposed by Paget, the ability of tumor cells originating from a particular site to adapt to a foreign environment may be limited to certain tissue types. A corollary of this idea is that target tissues in which metastasis is rare despite a rich vascular supply (e.g., skeletal muscle, spleen) constitute nonpermissive environments—"unfavorable soil," so to speak.

Once arrested at distant sites, extravasation of tumor cells involves transmigration between endothelial cells followed by egress through the basement membrane. Little is known about how this process occurs, and the mechanisms may differ depending on whether the endothelium is fenestrated (as in tissues such as the liver and bone marrow) or is held together by tight junctions (as in the brain). Extravasation requires the action of adhesion molecules (integrins, laminin receptors), proteolytic enzymes, and chemokines, which may be derived from tumor cells or from innate immune cells such as monocytes and neutrophils.

Even when metastatic cells take root and survive within distant tissues, they may fail to grow. This phenomenon, called *tumor dormancy,* is well described in melanoma and in breast and prostate cancer. Although the molecular mechanisms of productive colonization are still being unraveled, a consistent theme seems to be that tumor cells secrete cytokines, growth factors, and ECM molecules that act on the resident stromal cells, which in turn make the metastatic site habitable for the cancer cell. For example, breast cancer cells that are metastatic to bone often secrete parathyroid hormone–related protein (PTHRP), which stimulates osteoblasts to make RANK ligand (RANKL). RANKL then activates osteoclasts, which degrade the bone matrix and release growth factors embedded within it, like IGF and TGF-β. These in turn bind to receptors on the cancer cells, activating signaling pathways that support the growth and survival of the cancer cells. It is likely that many similar feedback loops between metastatic tumor cells and stromal cells await discovery.

KEY CONCEPTS

INVASION AND METASTASIS

- Ability to invade tissues, a hallmark of malignancy, occurs in four steps: loosening of cell-cell contacts, degradation of ECM, attachment to novel ECM components, and migration of tumor cells.
- Cell-cell contacts are lost by the inactivation of E-cadherin through a variety of pathways.
- Basement membrane and interstitial matrix degradation is mediated by proteolytic enzymes secreted by tumor cells and stromal cells, such as MMPs and cathepsins.
- Proteolytic enzymes also release growth factors sequestered in the ECM and generate chemotactic and angiogenic fragments from cleavage of ECM glycoproteins.
- The metastatic site of many tumors can be predicted by the location of the primary tumor. Many tumors arrest in the first capillary bed they encounter (lung and liver, most commonly).
- Some tumors show organ tropism, probably due to expression of adhesion or chemokine receptors whose ligands are expressed by endothelial cells at the metastatic site.
- Genes that promote epithelial-mesenchymal transition (EMT), like *TWIST* and *SNAIL*, may be important metastasis genes in epithelial tumors.

Evasion of Immune Surveillance

Long one of the "holy grails" of oncology, the promise of therapies that enable the host immune system to recognize and destroy cancer cells is finally coming to fruition, largely due to a clearer understanding of the ways by which cancer cells evade the host response. Paul Ehrlich first conceived the idea that tumor cells can be recognized as "foreign" and eliminated by the immune system. Subsequently, Lewis Thomas and Macfarlane Burnet formalized this concept by coining the term *immune surveillance,* which implies that a normal function of the immune system is to constantly "scan" the body for emerging malignant cells and destroy them. This idea has been supported by many observations—the direct demonstration of tumor-specific T cells and antibodies in patients; data showing that the density and quality of immune infiltrates in cancers often correlate with outcome; the increased incidence of certain cancers in immunodeficient people and mice; and most recently and most directly, the response of advanced cancers to therapeutic agents that act by stimulating latent host T-cell responses (described later).

Assuming the immune system is capable of recognizing and eliminating nascent cancers, it follows that the tumors that grow out in immunocompetent individuals must be composed of cells that are either invisible to the host immune system or that activate mechanisms that suppress host immunity. In support of this concept, it is now evident that therapeutic agents that neutralize these mechanisms can lead to tumor regression, even in patients with advanced cancers. These encouraging clinical responses constitute strong evidence that evasion of host immunity is indeed a hallmark of many, if not all, human cancers.

The following section explores some of the important questions about tumor immunity: What is the nature of tumor antigens? What host effector systems recognize tumor cells? How do tumors evade these host mechanisms? And how can immune reactions against tumors be exploited therapeutically?

Tumor Antigens

Malignant tumors express various types of molecules that may be recognized by the immune system as foreign antigens (Fig. 7.36). Tumor antigens that elicit an immune response have been demonstrated in many experimentally induced tumors and in some human cancers. It appears that protein antigens that elicit CD8+ cytotoxic T-cell

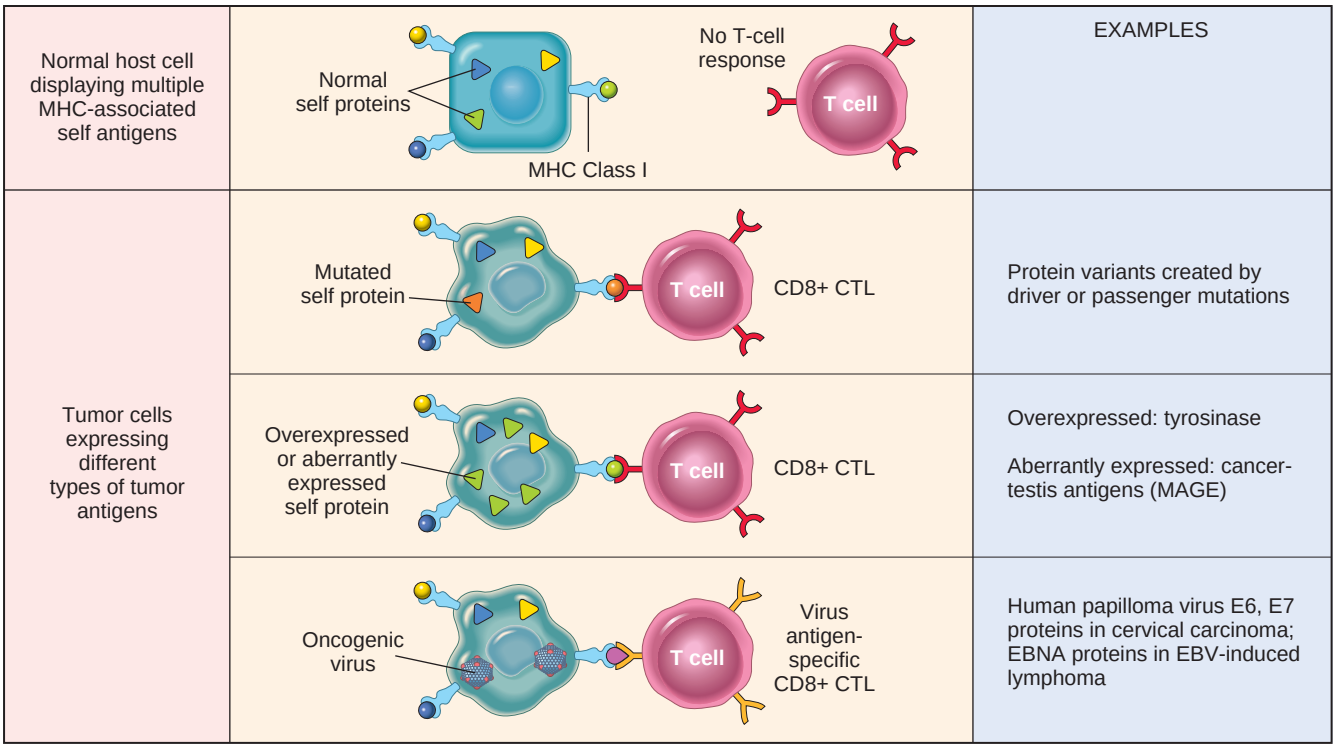


Figure 7.36 Tumor antigens recognized by CD8+ T cells. *EBV,* Epstein-Barr virus; *MHC,* major histocompatibility complex. (Modified from Abbas AK, Lichtman AH: *Cellular and Molecular Immunology,* ed 5, Philadelphia, 2003, WB Saunders.)

responses are most relevant for protective antitumor immunity. These protein antigens can be classified according to their source and molecular structure as follows:

- *Neoantigens produced from genes bearing passenger and driver mutations.* As discussed earlier, neoplastic transformation results from driver mutations in cancer genes; these mutated genes encode variant proteins that have never been seen by the immune system and may thus be recognized as nonself. Additionally, because of the enabling hallmark feature of genetic instability, cancers often have a high burden of passenger mutations throughout their genomes. Although these mutations are neutral in terms of cancer cell fitness and thus unrelated to the transformed phenotype, some by chance fall in the coding sequences of genes and give rise to protein variants that serve as tumor antigens (provided the mutation is in an epitope that binds to the MHC molecules of the individual). Notably, the size of the mutational load in a particular tumor, as deduced from DNA sequencing of tumor genomes, correlates well with the strength of the host CD8+ cytotoxic T-cell response and the effectiveness of immunomodulatory therapies (described later).
- *Overexpressed or aberrantly expressed normal cellular proteins.* Tumor antigens may also be normal cellular proteins that are overexpressed or aberrantly expressed in tumor cells. An example of an overexpressed protein that may be antigenic is tyrosinase, an enzyme involved in melanin biosynthesis that is expressed only in normal melanocytes and melanomas. It may be surprising that the immune system is able to respond to this normal self antigen. The probable explanation is that tyrosinase is normally produced in such small amounts and in so few healthy cells that it is not recognized by the immune system and fails to induce tolerance. Another group of tumor antigens are proteins that are aberrantly expressed in cancer cells at levels much greater than those seen in normal tissues. This may occur because of gene amplification or because of acquired epigenetic alterations that reactivate genes that are normally silenced in adult tissues. Some genes in this class encode *cancer-testis antigens,* proteins that are normally expressed only in testicular germ cells. Although the protein is present in the testis, it is not expressed on the cell surface in an antigenic form because sperm do not express major histocompatibility complex (MHC) class I antigens. Thus, for all practical purposes, these antigens are tumor-specific. Prototypic of this group is the melanoma antigen gene (MAGE) family. Although originally described in melanomas, MAGE antigens are expressed by a variety of tumor types.
- *Tumor antigens produced by oncogenic viruses.* In several cancers associated with ongoing active or latent viral infections (described later), the responsible viruses encode viral proteins that are recognized as foreign by the immune system. Cytotoxic T lymphocytes (CTLs) play a vital role in surveillance against virus-induced tumors by recognizing viral antigens and killing virus-infected cells. The importance of this immune mechanism is made evident by the high incidence of virally induced cancers that is observed in patients with inherited or acquired T-cell immunodeficiency. Many of these tumors are caused by Epstein-Barr virus (EBV) or HPV, DNA viruses that carry potent viral oncogenes.

Antitumor Effector Mechanisms

The principal immune mechanism of tumor eradication is killing of tumor cells by CTLs specific for tumor antigens. This mechanism predominates because the majority of tumor neoantigens consist of mutated gene products or viral proteins that are endogenously synthesized and presented in the context of MHC class I molecules, enabling their recognition by CTLs. Although sera from cancer patients may contain antibodies that recognize tumors, there is little evidence that they have any protective role under physiologic conditions. The cellular effectors that mediate immunity are described in Chapter 6. Here we focus on the CTL response to tumor antigens.

CTL responses against tumors are initiated by recognition of tumor antigens on host antigen presenting cells (APCs). Many cancers, particularly those that are rapidly growing, have a high fraction of cells that are undergoing cell death, either by apoptosis or necrosis. Dendritic cells and macrophages in the tumor microenvironment ingest tumor cells or released tumor antigens and migrate to draining lymph nodes. Here, they present the antigens in the context of MHC class II molecules and, through a mechanism called cross-presentation, in the context of MHC class I molecules (Fig. 7.37), allowing the antigens to be recognized by naïve CD8+ CTLs. Activation of antigen-specific CTLs also requires costimulatory molecules, which are upregulated on APCs presumably by "danger signals" released from damaged or necrotic tumor cells (DAMPs). Once activated by interaction with APCs, tumor-specific CTLs can migrate from lymph nodes to the tumor and kill tumor cells, directly and serially, without any assistance from other cell types.

The ability of CTLs to kill tumor cells independent of other cell types and factors underlies the ferocious antitumor activity of CTLs engineered to express chimeric antigen receptors (so-called CAR-T cells) against lineage-specific surface antigens found on certain tumors. For example, CAR-T cells specific for B-cell antigens are highly active against B-cell tumors but also annihilate normal B cells and release sufficient inflammatory cytokines to cause substantial morbidity and sometimes the death of the patient.

Although CTLs appear to have a preeminent role, other mechanisms also may play a role in tumor immunity. Antitumor CD4+ T-cell responses have been detected in patients, and increased numbers of CD4+ effector T cells, especially Th1 cells, in tumor infiltrates have been associated with a better prognosis in certain cancers, such as colorectal carcinoma. In experimental systems, natural killer (NK) cells and activated macrophages are capable of killing tumor cells. After activation with interleukin (IL)-2 and IL-15, NK cells can lyse a wide range of human tumors, including those that are nonimmunogenic for T cells due to loss of expression of MHC class I molecules. The ability of NK cells to kill tumor cells requires no prior sensitization, suggesting that they might constitute a first line of defense. While the importance of NK cells in host responses against spontaneous tumors is still not well established, cytokines that activate NK cells are being used for immunotherapy. Activated macrophages also exhibit cytotoxicity against tumor cells in vitro. Interferon-γ, a cytokine secreted by T cells and NK cells, is a potent activator of macrophages and may allow macrophages to kill tumors by mechanisms similar to those used to kill microbes (e.g., production of reactive oxygen species; see Chapter 3).

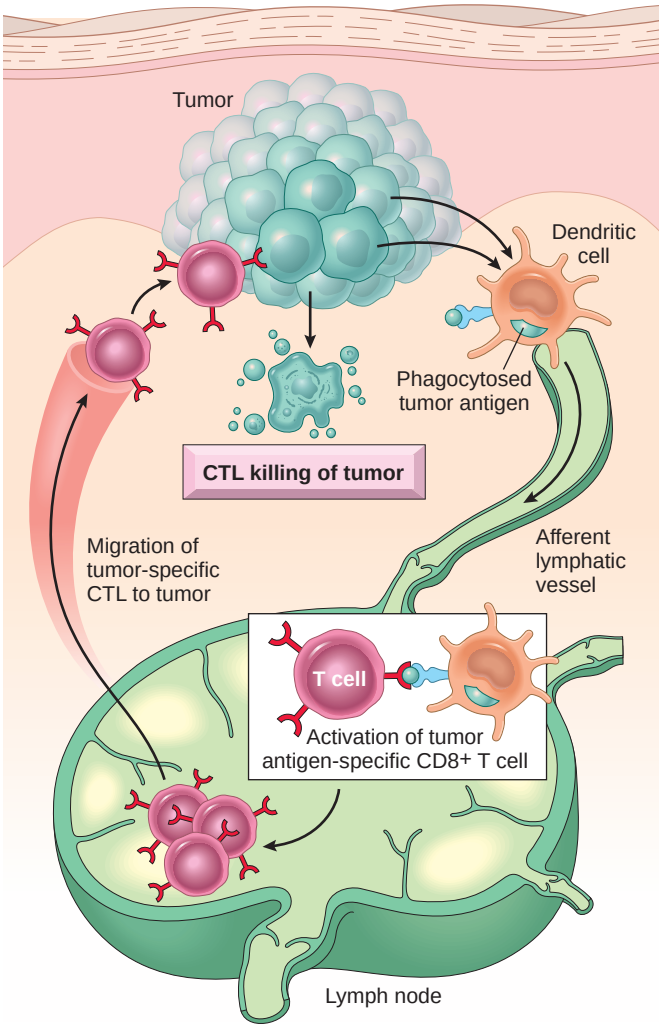


Figure 7.37 Cross-presentation of tumor antigens and induction of CD8+ cytotoxic T cell antitumor response. (Modified from Abbas AK, Lichtman AH: *Cellular and Molecular Immunology*, ed 8, Philadelphia, 2017, Elsevier.)

Mechanisms of Immune Evasion by Cancers

Immune responses often fail to check tumor growth because cancers evade immune recognition or resist immune effector mechanisms. Since cancer is all too common in persons who do not suffer from any overt immunodeficiency, it is evident that tumor cells must have ways to escape or evade the immune system in immunocompetent hosts. Several mechanisms appear to be operative (Fig. 7.38).

- *Selective outgrowth of antigen-negative variants.* During tumor progression, strongly immunogenic antigen-expressing subclones may be eliminated, and tumor cells that survive are those that have lost their antigens. If tumor cells express a large number of neoantigens, it is unlikely that all can be lost, but the same goal may be accomplished by other strategies.

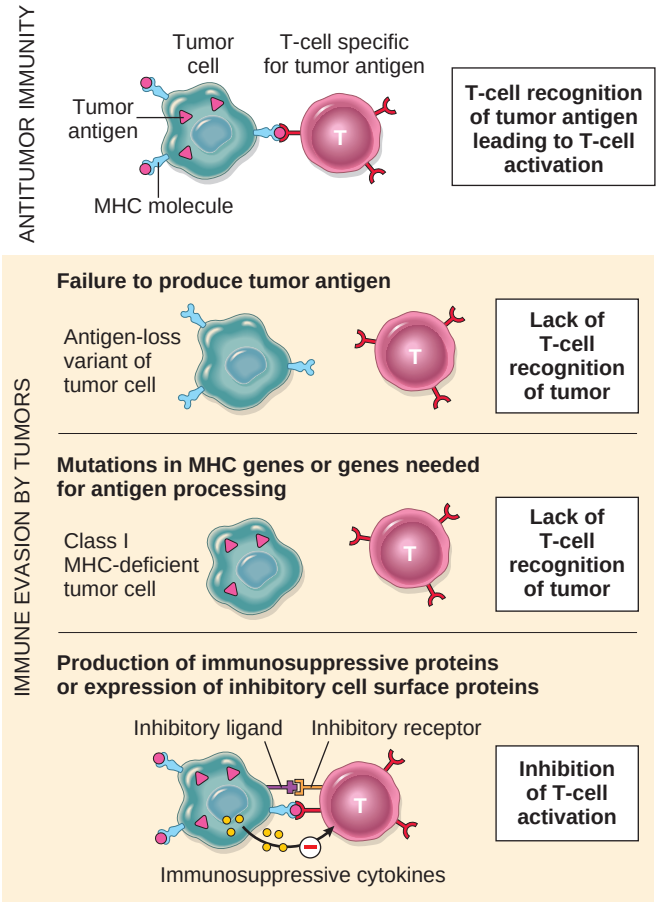


Figure 7.38 Mechanisms by which tumors evade the immune system. Tumors may evade immune responses by losing expression of antigens or major histocompatibility complex *(MHC)* molecules or by producing immunosuppressive cytokines or ligands such as PD-L1 for inhibitory receptors on T cells. (From Abbas AK, Lichtman AH, Pillai S: *Cellular and Molecular Immunology*, ed 7, Philadelphia, 2012, WB Saunders.)

- *Loss or reduced expression of MHC molecules.* Tumor cells may fail to express normal levels of HLA class I molecules, thereby losing the ability to display cytosolic antigens and escaping attack by cytotoxic T cells. Such cells, however, may trigger NK cells if the tumor cells express ligands for NK cell activating receptors.
- *Engagement of pathways that inhibit T-cell activation.* Tumor cells actively inhibit tumor immunity by upregulating negative regulatory "checkpoints" that suppress immune responses. Through a variety of mechanisms, tumor cells may promote the expression of the inhibitory receptor CTLA-4 on tumor-specific T cells. CTLA-4 binds to and removes its ligands, the B7 molecules, from APCs, thus reducing the engagement of the activating costimulatory receptor CD28. This not only prevents sensitization but also may induce long-lived unresponsiveness in tumor-specific T cells. Tumor cells also may upregulate the expression of PD-L1 and PD-L2, cell surface proteins that activate the programmed death-1 (PD-1) receptor on effector T cells. PD-1, like CTLA-4, inhibits T-cell activation. In some instances, overexpression of PD-L1 and PD-L2 is caused by amplification or translocation of the PD-L1 and PD-L2 genes, placing them in the pantheon of true oncogenes. Based on promising results in clinical trials, antibodies that restore T-cell function by blocking CTLA-4, PD-L1, or inhibitory PD-1 receptors are now approved for treatment of patients with advanced stage solid tumors and certain forms of lymphoma. The success of these agents has led to a new paradigm in cancer immunotherapy, sometimes called "checkpoint blockade," which is centered on the idea that agents that remove the "brakes" (checkpoints) imposed by tumors on host antitumor immune responses can be highly effective in treating cancer (Fig. 7.39).
- *Secretion of immunosuppressive factors.* Tumors may secrete several products that inhibit the host immune response. TGF-β is secreted in large quantities by many tumors and is a potent immunosuppressant. Many other soluble factors produced by tumors are also suspected of inhibiting the host immune response, including IL-10, prostaglandin E_2, certain metabolites derived from tryptophan, and VEGF, which can inhibit the movement of T cells from the vasculature into the tumor bed.
- *Induction of regulatory T cells (Tregs).* Some studies suggest that tumors produce factors that favor the development of immunosuppressive Tregs, which could also contribute to "immunoevasion."

Thus, it seems that there is no dearth of mechanisms by which tumor cells can outwit the host immune system. Nevertheless, the aforementioned response of tumors to immunomodulatory agents, such as antibodies that block CTLA-4 and PD-1, has generated tremendous excitement around the potential of rationally designed cancer immunotherapy. One of the remarkable features of this therapy is that it can achieve long-term remission and possibly even cures (which is not likely with any other treatment), perhaps because long-lived tumor-specific memory T cells are activated in the treated patients. The major challenges now are to determine which immune evasion mechanisms are most important in the cancers of individual patients and to develop a broader set of therapies that stymie various evasion mechanisms and thereby induce effective host immunity. In this regard, the treatment of human cancers with highly specific checkpoint inhibitor antibodies affords the opportunity to develop a mechanistic understanding of certain aspects of tumor response and resistance. Lessons learned to date from clinical trials conducted with these inhibitors include the following:

- *Tumor neoantigen burden is a good predictor of response.* Tumors that have deficiencies in mismatch repair enzymes, which normally correct errors in DNA replication that lead to point mutations, have the highest mutation burdens of all cancers, and these cancers are most likely to respond to checkpoint blockade therapy. Based on this observation, anti–PD-1 therapy is now approved for all metastatic tumors that have mismatch repair deficiency, which leads to a high mutational burden, regardless of histologic type.
- *Only a subset of tumors (25% to 40%) responds to checkpoint inhibitors,* probably because nonresponding tumors rely on evasion strategies other than engaging checkpoint pathways. For example, tumors that do not express PD-L1, or do not have exhausted PD-1–positive CD8+ T-cell infiltrates, are not likely to respond to anti–PD-1 or

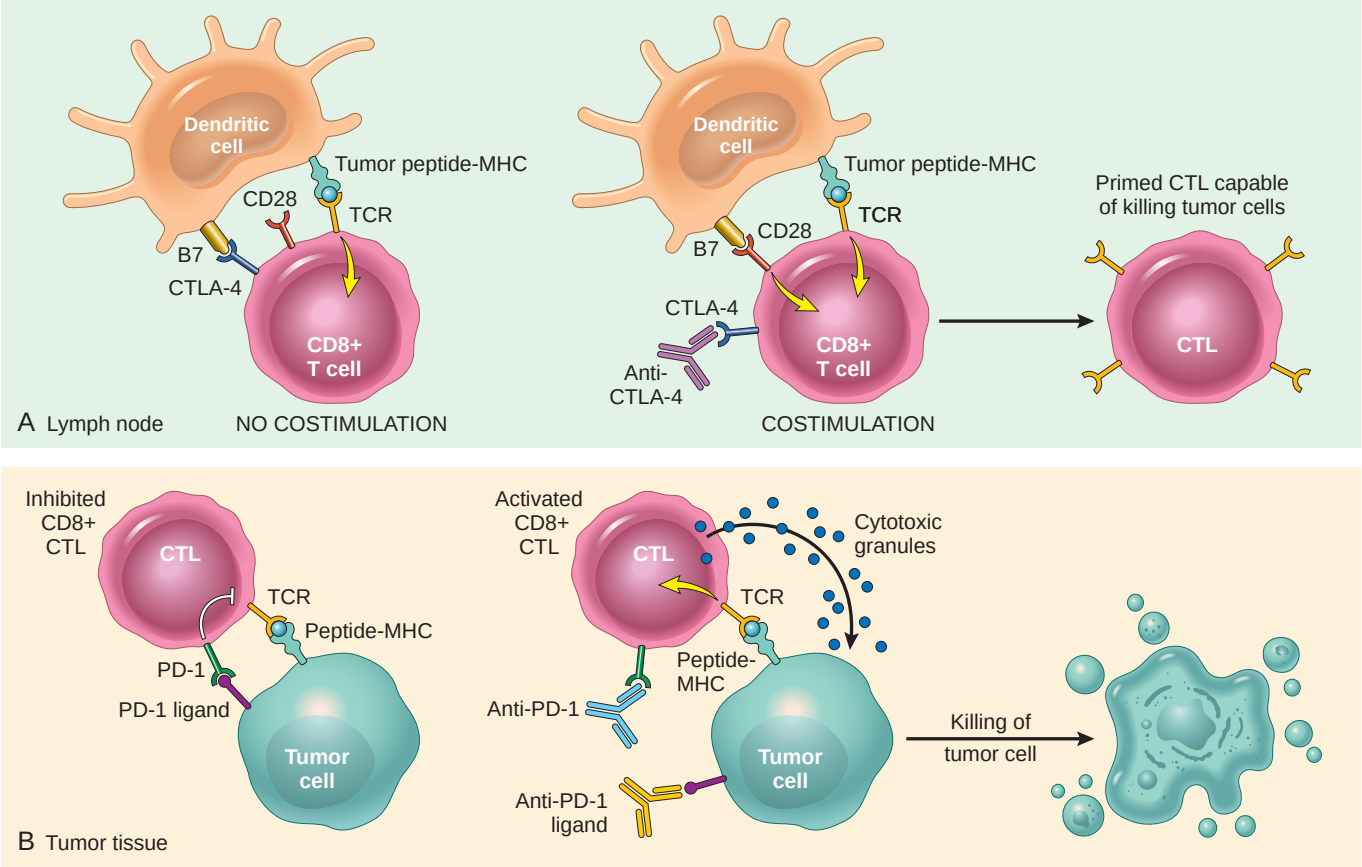


Figure 7.39 Activation of host antitumor immunity by checkpoint inhibitors. (A) Blockade of the CTLA4 surface molecule with an inhibitor antibody allows cytotoxic CD8+ T cells *(CTLs)* to engage B7 family coreceptors, leading to T-cell activation. (B) Blockade of PD-1 receptor or PD-1 ligand by inhibitory antibodies abrogates inhibitory signals transmitted by PD-1, again leading to activation of CTLs. *MHC,* Major histocompatibility complex. (From Abbas AK, Lichtman AH, Pillai S: *Cellular and Molecular Immunology*, ed 9, Philadelphia, 2017, Elsevier.)

anti–PD-L1 therapy. Investigators are currently studying which biomarkers will predict responsiveness to different checkpoint blockade approaches.

- *Combination of checkpoint inhibitors with other therapeutic agents.* The combined use of different checkpoint inhibitors, or of checkpoint inhibitors with other types of therapeutic agents (e.g., drugs that target oncoproteins like tyrosine kinases), will likely be necessary to achieve higher rates of therapeutic success. The first successful example of this is the combined use of anti–CTLA-4 and anti–PD-1 to treat melanoma, an approach that is more effective than anti–CTLA-4 alone. This reflects the fact that the mechanisms by which CTLA-4 and PD-1 inhibit T-cell activation are different (Fig. 7.39). There are numerous ongoing or planned clinical trials using combinations of checkpoint blockade together with other kinds of therapeutic agents.
- *The most common toxicities associated with checkpoint blockade are autoimmunity and/or inflammatory damage to organs.* This is predictable because the physiologic function of inhibitory receptors and ligands is to maintain tolerance to self antigens. A wide range of organs may be affected including colon, lungs, endocrine organs, heart, and skin, each requiring different clinical interventions, sometimes including cessation of the potentially life-saving checkpoint blockade therapy.

KEY CONCEPTS

EVASION OF IMMUNE SURVEILLANCE

- Tumor cells can be recognized by the immune system as nonself and destroyed.
- Antitumor activity is mediated by predominantly cell-mediated mechanisms. Tumor antigens are presented on the cell surface by MHC class I molecules and are recognized by CD8+ CTLs.
- The different classes of tumor antigens include products of mutated proto-oncogenes, tumor suppressor genes, overexpressed or aberrantly expressed proteins, and tumor antigens produced by oncogenic viruses.
- Immunosuppressed patients have an increased risk for development of cancer, particularly types caused by oncogenic DNA viruses.
- In immunocompetent patients, tumors may avoid the immune system by several mechanisms including selective outgrowth of antigen-negative variants, loss or reduced expression of histocompatibility antigens, and immunosuppression mediated by expression of certain factors (e.g., TGF-β, PD-1 ligand) by the tumor cells.
- Antibodies that overcome mechanisms of immune evasion involving "immune checkpoints" are approved for treatment of patients with advanced cancer and are most likely to be effective against cancers with a high mutational burden.

Genomic Instability

Genetic aberrations that increase mutation rates are very common in cancers and expedite the acquisition of driver mutations that are required for transformation and subsequent tumor progression. Although exposure to environmental agents that are mutagenic (e.g., chemicals, radiation, sunlight) is unavoidable, cancers are relatively rare outcomes of these encounters. This state of affairs results from the ability of normal cells to repair DNA damage, the death of cells with irreparable damage (see "Evasion of Cell Death" earlier), and other mechanisms such as oncogene-induced senescence and immune surveillance.

Several mechanisms contribute to genomic instability in cancer cells. We previously discussed how p53 protects the genome from potentially oncogenic damage by arresting cell division to provide time for repair of DNA damage and by initiating apoptosis in irreparably damaged cells. Cancers with loss of p53 function not only accumulate point mutations but also are strongly associated with aneuploidy, which may take the form of deletions, amplifications, and complex chromosomal rearrangements. These genomic aberrations may occur in cells with defective telomeres during break-fusion-breakage cycles (see Fig. 7.32) or may be created by other types of chromosomal "catastrophes" (such as chromothrypsis, described later) that lead to DNA breaks in multiple chromosomes. In the absence of p53 function, cells with severely damaged genomes that normally would be eliminated persist and stitch their chromosome back together in an error-prone way using the nonhomologous end-joining pathway.

TP53 is the most commonly mutated gene in cancer, and loss of p53 function is thus the preeminent source of genomic instability in cancers. In the following sections, we discuss two other classes of proteins that normally function to protect against genomic instability: DNA repair factors and DNA polymerase itself. As with p53, dysfunction of these factors leads to more rapid accumulation of genomic damage (a "mutator" phenotype), speeding cancer development and progression. Finally, we will describe a special type of regulated genomic instability specific to lymphoid cells that also is a source of oncogenic mutations.

DNA Mismatch Repair Factors. DNA mismatch repair proteins work together to act as "spell checkers" during the process of DNA replication. For example, if there is an erroneous pairing of G with T rather than the normal A with T, the mismatch-repair factors correct the defect. With "proofreading" function lost, errors accumulate throughout the genome. Some of these errors may by chance create driver mutations, and with time a cancer may result. One of the hallmarks of mismatch-repair defects is *microsatellite instability*. Microsatellites are tandem repeats of one to six nucleotides found throughout the genome. In normal people the length of these microsatellites remains constant. However, if mismatch repair is defective, these satellites are unstable and increase or decrease in length, creating mutated alleles.

Hereditary nonpolyposis colon cancer (HNPCC) syndrome (also known as Lynch syndrome) is an autosomal dominant disorder associated with carcinomas that arise predominantly in the cecum and proximal colon (Chapter 17). Individuals with HNPCC syndrome inherit one abnormal copy of a mismatch repair gene. Trouble arises when cells acquire somatic loss-of-function mutations, presumably at random, in their single normal alleles. Germline loss-of-function mutations in at least four different genes can produce HNPCC syndrome. The most commonly affected genes are *MSH2* and *MLH1*, each accounting for approximately 30% of HNPCC syndrome. Microsatellite instability also is observed in about 15% of sporadic colon cancers and less frequently in many other cancer types. In sporadic cancers, defects in mismatch repair usually stem from epigenetic silencing of the *MLH1* gene, rather than somatic mutations.

Nucleotide Excision Repair Factors. UV radiation causes cross-linking of pyrimidine residues, preventing normal DNA replication. Such DNA damage is repaired by the nucleotide excision repair system. Several genes are involved in nucleotide excision repair. Inherited loss-of-function mutations in any of these genes gives rise to a syndrome called *xeroderma pigmentosum* that is marked by an extraordinarily high risk of skin cancers, specifically squamous cell carcinoma and basal cell carcinoma.

Homologous Recombination Repair Factors. Other types of DNA damage, particularly covalent DNA cross-links and double-stranded DNA breaks, are repaired through a complex process called homologous recombination. Several disorders caused by defects in homologous recombination factors are associated with an increased risk of cancer, as follows:

- *Bloom syndrome* is an autosomal recessive disorder caused by loss-of-function mutations in a helicase that is required for homologous recombination repair. Affected individuals have developmental anomalies and an increased risk of developing many different types of cancer.
- *Ataxia telangiectasia* is an autosomal recessive disorder caused by defects in *ATM,* a gene encoding a kinase that acts upstream of p53. This syndrome is characterized by neurodegeneration (particularly of the cerebellum, hence the ataxia), immunodeficiency, hypersensitivity to radiation (due to an inability to repair double-stranded DNA breaks), and predisposition to cancer, particularly certain forms of leukemia and lymphoma. Somatic driver mutations in *ATM* also are common in certain types of lymphoid neoplasms.
- *Fanconi anemia* is an autosomal recessive disorder that may be caused by mutations in more than a dozen different genes, each encoding a protein that participates in a pathway that repairs DNA cross-links through homologous recombination. It is characterized by developmental abnormalities (short stature, skeletal abnormalities), hypersensitivity to chemotherapeutic agents that cross-link DNA, and increased risk of bone marrow failure (aplasia) and leukemia.
- *Familial breast cancer* often is associated with inherited defects in genes that are required for homologous recombination repair. Mutations in two genes, *BRCA1* and *BRCA2,* account for 25% of cases. Certain germline *BRCA2* mutations cause Fanconi anemia, and it appears that BRCA proteins and Fanconi proteins function cooperatively in a DNA damage response network linked to homologous recombination repair. Defects in this pathway lead to the activation of the salvage

nonhomologous end-joining pathway, formation of dicentric chromosomes, bridge-fusion-breakage cycles, and aneuploidy, just as occurs in p53-deficient cells that undergo telomere shortening (see Fig. 7.32). In addition to breast cancer, women with *BRCA1* mutations have a substantially higher risk of epithelial ovarian cancers, and men have a slightly higher risk of prostate cancer. Mutations in the *BRCA2* gene are associated with a broader spectrum of cancers including breast cancer in men and women as well as cancers of the ovary, prostate, pancreas, bile ducts, stomach, melanocytes, and B lymphocytes.

DNA Polymerase. Under normal circumstances, cellular DNA polymerases involved in DNA replication have a very low rate of error, defined as addition of nucleotide that does not match its partner on the template strand of DNA. This fidelity stems in part from an inherent exonuclease activity that allows DNA polymerase to pause, excise mismatched bases, and insert the proper nucleotide before proceeding down the template strand. Subsets of certain cancers, most often endometrial carcinomas and colon cancers, harbor mutations in DNA polymerase that result in a loss of this "proofreading" function and the accumulation of numerous point substitutions. Cancers with DNA polymerase mutations are the most heavily mutated of all human cancers and, presumably because of a high burden of neoantigens, appear to have excellent responses to immune checkpoint inhibitors.

Regulated Genomic Instability in Lymphoid Cells. A special type of DNA damage plays a central role in the pathogenesis of tumors of B and T lymphocytes. As described in Chapter 6, adaptive immunity relies on the ability of B and T cells to diversify their antigen receptor genes. Developing B and T cells both express a pair of gene products, RAG1 and RAG2, that carry out V(D)J segment recombination, permitting the assembly of functional antigen receptor genes. In addition, after encountering antigen, mature B cells express a specialized enzyme called activation-induced cytosine deaminase (AID), which is required for both immunoglobulin gene class switch recombination and somatic hypermutation. These processes are associated with AID-induced DNA breaks or nucleotide substitutions, both of which are prone to errors such as translocations and mutations that cause lymphoid neoplasms (Chapter 13).

KEY CONCEPTS

GENOMIC INSTABILITY AS ENABLER OF MALIGNANCY

- Persons with inherited mutations of genes involved in DNA repair systems are at greatly increased risk for the development of cancer.
- Patients with HNPCC syndrome have defects in the mismatch repair system, leading to development of carcinomas of the colon. These patients' genomes show microsatellite instability, characterized by changes in length of short repeats throughout the genome.
- Patients with xeroderma pigmentosum have a defect in the nucleotide excision repair pathway and are at increased risk for the development of cancers of the skin exposed to UV light because of an inability to repair pyrimidine dimers.
- Syndromes involving defects in the homologous recombination DNA repair system constitute a group of disorders—Bloom syndrome, ataxia-telangiectasia, and Fanconi anemia—that are characterized by developmental disorders and hypersensitivity to DNA-damaging agents, such as ionizing radiation. *BRCA1* and *BRCA2*, which are mutated in familial breast cancers, are involved in DNA repair.
- Mutations in DNA polymerase that abolish proofreading function leads to genomic instability in subsets of colonic and endometrial carcinomas.
- T and B cells undergo regulated genomic instability during somatic gene rearrangements. Errors in this process are an important cause of lymphoid neoplasms.

Cancer-Enabling Inflammation

Infiltrating cancers provoke a chronic inflammatory reaction, leading some to liken them to "wounds that do not heal." In patients with advanced cancers, this inflammatory reaction can be so extensive as to cause systemic signs and symptoms such as anemia (Chapter 14), fatigue, and cachexia (described later). However, studies carried out on cancers in animal models suggest that inflammatory cells also modify the tumor cells and the local microenvironment to enable many of the hallmarks of cancer. These effects may stem from direct interactions between inflammatory cells and tumor cells or through indirect effects of inflammatory cells on other resident stromal cells, particularly cancer-associated fibroblasts and endothelial cells. Proposed cancer-enabling effects of inflammatory cells and resident stromal cells include the following:

- *Release of factors that promote proliferation.* Infiltrating leukocytes and activated stromal cells secrete a wide variety of growth factors such as EGF, as well as proteases that can liberate growth factors from the ECM.
- *Removal of growth suppressors.* The growth of epithelial cells is normally suppressed by cell–cell and cell–ECM interactions. Proteases released by inflammatory cells can degrade the adhesion molecules that mediate these interactions, removing a barrier to growth.
- *Enhanced resistance to cell death.* Recall that detachment of epithelial cells from basement membranes and from cell–cell interactions leads to a form of cell death called *anoikis*. It is suspected that tumor-associated macrophages prevent anoikis by expressing adhesion molecules such as integrins that promote direct physical interactions with the tumor cells. There is also substantial evidence that stromal cell–cancer cell interactions increase the resistance of cancer cells to chemotherapy, presumably by activating signaling pathways that promote cell survival in the face of stresses such as DNA damage.
- *Inducing angiogenesis.* Inflammatory cells release numerous factors, including VEGF, which stimulate angiogenesis.
- *Activating invasion and metastasis.* Proteases released from macrophages foster tissue invasion by remodeling the ECM, while factors such as TNF and EGF may directly stimulate tumor cell motility. As mentioned, other factors released from stromal cells, such as TGF-β, may promote

EMT, which is considered to be a key event in the process of invasion and metastasis.

- *Evading immune destruction.* A variety of soluble factors released by macrophages and other stromal cells may contribute to the immunosuppressive microenvironment of tumors, including TGF-β and a number of other factors that either favor the recruitment of immunosuppressive Tregs or suppress the function of CD8+ cytotoxic T cells. Furthermore, there is abundant evidence in murine cancer models and emerging evidence in human disease that advanced cancers contain alternatively activated (M2) macrophages (Chapter 3), cells induced by cytokines such as IL-4 and IL-13. These macrophages produce cytokines that promote angiogenesis, fibroblast proliferation, and collagen deposition, all of which are commonly observed in invasive cancers. In addition, macrophages may suppress effective host immune responses to cancer cells by expressing the immune checkpoint factor PD-L1 and through other mechanisms that remain to be fully elucidated.

Although a thorough understanding of how cancers "manipulate" inflammatory cells to support their growth and survival remains elusive, there is substantial interest in the development of therapies directed at tumor-induced inflammation and its downstream consequences. Of note in this regard, antiinflammatory cyclooxygenase-2 (COX-2) inhibitors have been shown to decrease the incidence of colonic adenomas and are approved for treatment of patients with familial adenomatous polyposis.

Dysregulation of Cancer-Associated Genes

The genetic damage that activates oncogenes or inactivates tumor suppressor genes may be subtle (e.g., point mutations) or may involve segments of chromosomes large enough to be detected in a routine karyotype. Activation of oncogenes and loss of function of tumor suppressor genes by mutations were discussed earlier in this chapter. Here we discuss chromosomal abnormalities and epigenetic changes that contribute to carcinogenesis and then briefly touch on the role of noncoding RNAs.

Chromosomal Changes

Certain chromosomal abnormalities are highly associated with particular neoplasms and inevitably lead to the dysregulation of genes with an integral role in the pathogenesis of that tumor type. Specific recurrent chromosomal abnormalities have been identified in most leukemias and lymphomas, many sarcomas, and an increasing number of carcinomas. In addition, whole chromosomes may be gained or lost. Although changes in chromosome number (aneuploidy) and structure are generally considered to be late phenomena in cancer progression, in some cases (e.g., in cells that have lost their telomeres; see Fig. 7.32), it can be an early event that initiates the transformation process.

Historically, chromosomal changes in cancer were identified through karyotyping, the morphologic identification of metaphase chromosomes prepared from clinical specimens. Today, however, cancer cell karyotypes are being reconstructed in research laboratories from deep sequencing of cancer cell genomes, and it is possible that conventional karyotyping will be supplanted by other methods even in clinical laboratories in the years to come. Whatever technology is used, the study of chromosomal changes in tumor cells is important. First, genes in the vicinity of recurrent chromosomal breakpoints or deletions are very likely to be either oncogenes (e.g., *MYC, BCL2, ABL*) or tumor suppressor genes (e.g., *APC, RB*). Second, certain karyotypic abnormalities have diagnostic value or important prognostic or therapeutic implications. For example, tests that detect and quantify *BCR-ABL* fusion genes or their mRNA products are essential for the diagnosis of CML and are used to monitor the response to BCR-ABL kinase inhibitors. Many other chromosomal aberrations that are characteristic of specific tumor types are presented in later chapters.

Table 7.8 Selected Examples of Oncogenes Created by Translocations

Malignancy	Translocation	Affected Genes
Chronic myeloid leukemia (CML)	(9;22)(q34;q11)	*ABL* 9q34 *BCR* 22q11
Acute myeloid leukemia (AML)	(8;21)(q22;q22) (15;17)(q22;q21)	*AML* 8q22 *ETO* 21q22 *PML* 15q22 *RARA* 17q21
Burkitt lymphoma	(8;14)(q24;q32)	*MYC* 8q24 *IGH* 14q32
Mantle cell lymphoma	(11;14)(q13;q32)	*CCND1* 11q13 *IGH* 14q32
Follicular lymphoma	(14;18)(q32;q21)	*IGH* 14q32 *BCL2* 18q21
Ewing sarcoma	(11;22)(q24;q12)	*FLI1* 11q24 *EWSR1* 22q12
Prostatic adenocarcinoma	(7;21)(p22;q22) (17;21)(p21;q22)	*TMPRSS2* (21q22.3) *ETV1* (7p21.2) *ETV4* (17q21)

Chromosomal Translocations. Any type of chromosomal rearrangement—translocations, inversions, amplifications, and even small deletions—can activate proto-oncogenes, but chromosomal translocation is the most common mechanism. Notable examples of oncogenes activated by chromosomal translocations are listed in Table 7.8. Translocations can activate proto-oncogenes in two ways:

- By promoter or enhancer substitution, in which the translocation results in overexpression of a proto-oncogene by swapping its regulatory elements with those of another gene, typically one that is highly expressed.
- By formation of a fusion gene in which the coding sequences of two genes are fused in part or in whole, leading to the expression of a novel chimeric protein with oncogenic properties.

Overexpression of a proto-oncogene caused by translocation is exemplified by Burkitt lymphoma. Virtually all Burkitt lymphomas have a translocation involving chromosome 8q24, where the *MYC* gene resides, and one of the three chromosomes that carry an immunoglobulin gene. At its normal locus, *MYC* is tightly controlled and is most highly expressed in actively dividing cells. In Burkitt lymphoma the most common translocation moves the *MYC*-containing

segment of chromosome 8 to chromosome 14q32 (see Fig. 7.23), placing it close to the immunoglobulin heavy chain *(IGH)* gene. The genetic notation for the translocation is t(8;14)(q24;q32). The molecular mechanisms of the translocation-mediated overexpression of *MYC* are variable, as are the precise breakpoints within the *MYC* gene. In most cases the translocation removes regulatory sequences of the *MYC* gene and replaces them with the control regions of the *IGH* gene, which is highly expressed in B cells. The *MYC* coding sequences remain intact, and the MYC protein is constitutively expressed at high levels. The almost invariable presence of *MYC* translocations in Burkitt lymphomas attests to the importance of *MYC* overactivity in the pathogenesis of this tumor.

There are many other examples of translocations involving oncogenes and antigen receptor loci in lymphoid tumors. For these (or any other) translocations to occur, double-stranded DNA breaks must occur simultaneously in at least two places in the genome, and the free DNA ends must then be joined to create two new derivative chromosomes. In lymphoid cells, most of these molecular misadventures are believed to occur during attempts at normal antigen receptor gene recombination (which occurs in both B- and T-cell progenitors) or class-switch recombination (which is confined to antigen-stimulated mature B cells). Not unexpectedly, tumors with translocations involving immunoglobulin genes are always of B-cell origin, and tumors with translocations involving T-cell receptor genes are always of T-cell origin. The affected genes are diverse, but as with translocations involving *MYC,* the net effect is overexpression of some protein with oncogenic activity.

The *Philadelphia chromosome,* characteristic of CML and a subset of B-cell acute lymphoblastic leukemias (Chapter 13), provides the prototypic example of a chromosomal rearrangement that creates a fusion gene encoding a chimeric oncoprotein. In this instance the two chromosome breaks lie within the *ABL* gene on chromosome 9 and within the *BCR (b*reakpoint *c*luster *r*egion) gene on chromosome 22 (see Fig. 7.23). Nonhomologous end-joining then leads to a reciprocal translocation that creates an oncogenic *BCR-ABL* fusion gene on the derivative chromosome 22 (the so-called Philadelphia chromosome). *BCR-ABL* fusion genes encode chimeric BCR-ABL proteins with constitutive tyrosine kinase activity. Since the discovery of *BCR-ABL* in CML, many other fusion oncogenes encoding constitutively active tyrosine kinases have been described in a broad array of human cancers. Like BCR-ABL, these fusion proteins drive oncogenic signaling pathways and have sometimes proven to be targets of effective therapies.

Other oncogenic fusion genes encode nuclear factors that regulate transcription or chromatin structure. In contrast to overactive tyrosine kinases, less is known about how nuclear fusion oncoproteins contribute to cancer. An exception with important clinical consequences is found in *acute promyelocytic leukemia* (APML). APML is virtually always associated with a reciprocal translocation between chromosomes 15 and 17 that produces a *PML-RARA* fusion gene (Fig. 7.40). How this fusion gene functions is now reasonably well understood.

- The fusion gene encodes a chimeric protein consisting of part of a protein called PML and part of the retinoic acid receptor-α (RARα). Normal RARα binds to DNA and activates transcription in the presence of retinoids. Among the RARα responsive genes are a number that are needed for the differentiation of myeloid progenitors into neutrophils.
- The PML-RARα oncoprotein has diminished affinity for retinoids, such that at physiologic levels retinoids do not bind to PML-RARα to any significant degree. In this "unliganded" state, it retains the capacity to bind DNA, but instead of activating transcription, it inhibits transcription through recruitment of transcriptional repressors. This interferes with the expression of genes that are needed for differentiation, leading to a "pile-up" of proliferating myeloid progenitors that replace normal bone marrow elements.
- When given in pharmacologic doses, all-*trans* retinoic acid (ATRA) binds to PML-RARα and causes a conformational change that results in the displacement of repressor complexes and the recruitment of different complexes that activate transcription. There also is evidence that ATRA-bound PML-RARA complexes are degraded more rapidly. These changes overcome the block in gene expression, causing the neoplastic myeloid progenitors to differentiate into neutrophils and die (as normal mature neutrophils do), clearing the marrow over several days and allowing for recovery of normal hematopoiesis.

This highly effective therapy is the first example of *differentiation therapy,* in which immortal tumor cells are induced to differentiate into their mature progeny, which have limited lifespans. It has also spurred efforts to develop drugs that target other nuclear oncoproteins, despite the inherent difficulty of the problem.

Deletions. Chromosomal deletions are another very common structural abnormality in tumor cells. Deletion of specific regions of chromosomes is associated with the loss of particular tumor suppressor genes.

As we discussed earlier, deletions involving chromosome 13q14, the site of the *RB* gene, are associated with retinoblastoma, and deletion of the *VHL* tumor suppressor gene on chromosome 3p is a common event in renal cell carcinomas. Sequencing of cancer cell genomes has revealed many more examples of deletions involving tumor suppressor genes, as well as small insertions of DNA from one site into another. Not all deletions lead to loss of gene function, however; a subset activates oncogenes through the same mechanisms as chromosomal translocations. For example, about 25% of T-cell acute lymphoblastic leukemias have small deletions of chromosome 1 that juxtapose the *TAL1* proto-oncogene with a nearby active promoter, leading to overexpression of the TAL1 transcription factor. Many other examples of "addition by genomic subtraction" have now been discovered through sequencing of cancer genomes.

Gene Amplification. Overexpression of oncogenes may also result from reduplication and amplification of their DNA sequences. Such amplification may produce up to several hundred copies of the oncogene in the tumor cell. In some cases the amplified genes produce chromosomal changes that can be identified microscopically. Two mutually exclusive patterns are seen: (1) multiple small extrachromosomal structures called *double minutes* and (2) *homogeneous*

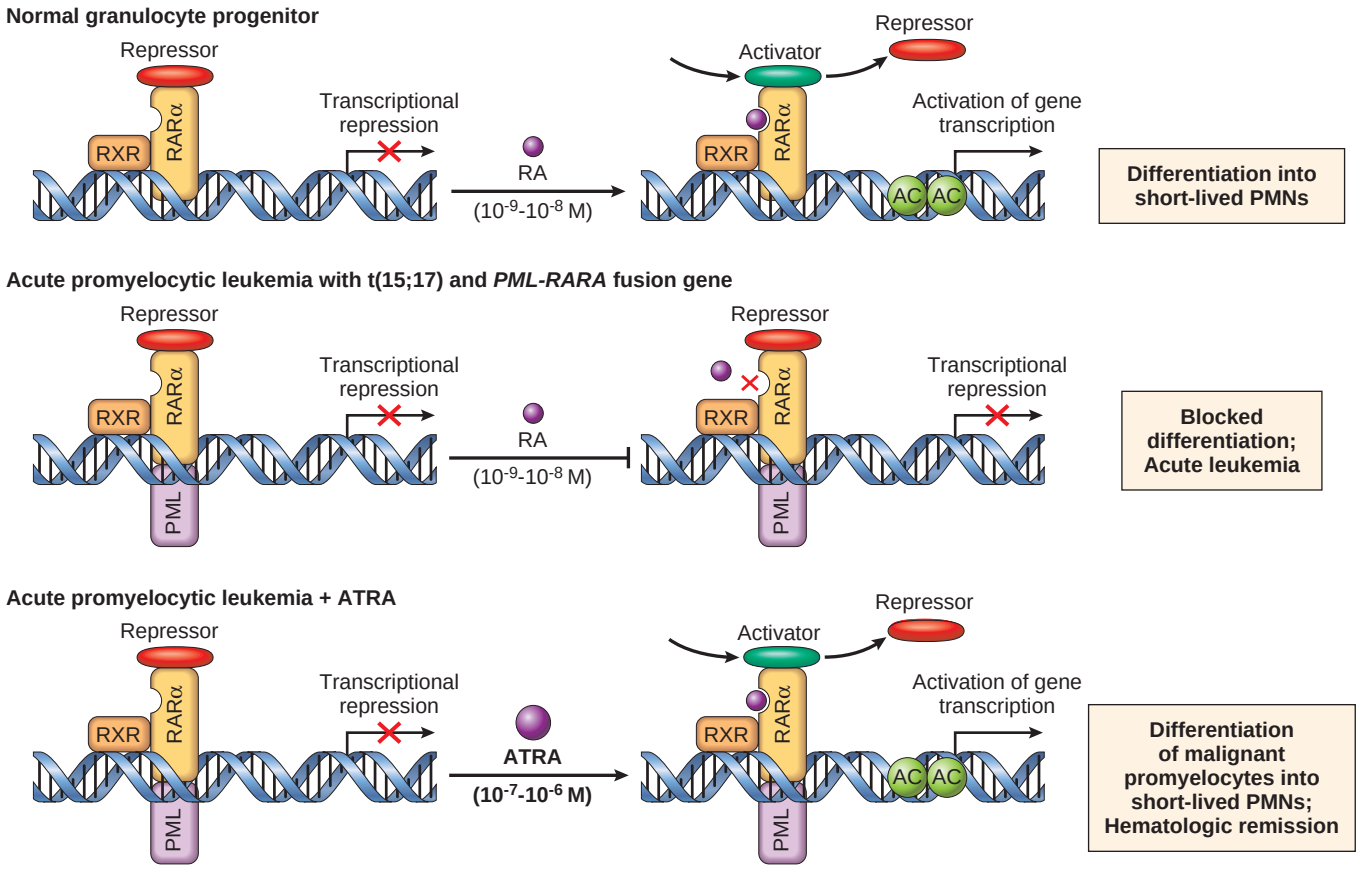


Figure 7.40 Molecular pathogenesis of acute promyelocytic leukemia *(PML)* and basis for response to all-*trans* retinoic acid *(ATRA)*. *PMN*, Polymorphonuclear neutrophil; *RA*, retinoic acid; *RXR*, binding partner for normal RARa and PML-RARa fusion protein encoded by a chimeric gene created by the (15;17) translocation in acute promyelocytic leukemia.

staining regions. The latter derive from the insertion of the amplified genes into new chromosomal locations, which may be distant from the normal location of the involved oncogene. The affected chromosomal regions lack a normal pattern of light- and dark-staining bands, appearing homogeneous in karyotypes (see Fig. 7.24). From a clinical perspective the most important amplifications are *NMYC* in neuroblastoma and *ERBB2* in breast cancers. *NMYC* is amplified in 25% to 30% of neuroblastomas, and its amplification is associated with poor prognosis. *ERBB2* amplification occurs in about 20% of breast cancers. As already mentioned, antibody therapy directed against the HER2 receptor encoded by *ERBB2* is an effective therapy for this molecular subset of breast cancers.

Complex Chromosomal Rearrangements. The true extent of chromosome rearrangements in cancer is only now coming into view thanks to sequencing of entire cancer cell genomes, which allows for comprehensive "reconstruction" of chromosomes from DNA sequences. This exercise has revealed not only a large number of simple rearrangements (e.g., small deletions, duplications, or inversions), but also much more dramatic chromosomal "catastrophes" that stem from chromothrypsis (literally, chromosome shattering). *Chromothrypsis* is observed in 1% to 2% of cancers as a whole and is particularly common in osteosarcomas and gliomas. It appears to result from a single event in which dozens to hundreds of chromosome breaks occur in a single chromosome or several chromosomes. The genesis of these breaks is uncertain, but DNA repair mechanisms are activated that stitch the pieces together in a haphazard way, creating many rearrangements, deletions, and even amplifications. It is hypothesized that such catastrophic events may by chance mutate multiple cancer genes simultaneously thereby expediting the process of carcinogenesis.

KEY CONCEPTS

GENETIC LESIONS IN CANCER

- Tumor cells may acquire several types of oncogenic mutations including point mutations and other nonrandom chromosomal abnormalities, such as translocations, deletions, and gene amplifications.
- Translocations contribute to carcinogenesis by overexpression of oncogenes or generation of novel fusion proteins with altered signaling capacity. Deletions frequently cause loss of tumor suppressor gene function and occasionally activate proto-oncogenes. Gene amplification generally increases the expression and function of oncogenes.
- Genomic sequencing has revealed numerous "cryptic" (subcytogenetic) rearrangements, mainly small deletions and insertions ("indels"), as well as chromothrypsis, in which a chromosome is "shattered" and then reassembled in a haphazard way.

Epigenetic Changes

Epigenetic changes have been implicated in many aspects of the malignant phenotype including the expression of cancer genes, the control of differentiation and self-renewal, and drug sensitivity and drug resistance. As discussed in Chapter 1, "epigenetics" refers to factors other than the sequence of DNA that regulate gene expression (and thereby cellular phenotype). Recall that epigenetic mechanisms include histone modifications catalyzed by enzymes associated with chromatin regulatory complexes; DNA methylation, a modification created by DNA methyltransferases; and other alterations that regulate the higher order organization of DNA (e.g., looping of enhancer elements onto gene promoters).

It has been recognized for more than a hundred years that the nuclei of cancer cells display abnormal morphologies, which (as discussed earlier) may take the form of hyperchromasia, chromatin clumping, or chromatin clearing (so-called vesicular nuclear chromatin). These altered appearances stem from disturbances of chromatin organization. One of the most notable findings emerging from the sequencing of cancer genomes has been the identification of numerous mutations involving genes that encode epigenetic regulatory proteins (Table 7.9). As a result, it is now suspected that the altered morphologic appearance of cancer cells reflects acquired genetic defects in factors that maintain the epigenome. Indeed, methods that allow genome-wide assessments have begun to reveal widespread alterations in cancer cell epigenomes, which can be broadly divided into the following categories:

- *Silencing of tumor suppressor genes by local hypermethylation of DNA.* Some cancer cells exhibit selective hypermethylation of the promoters of tumor suppressor genes that results in their transcriptional silencing. Typically, hypermethylation occurs on only one allele, and the function of the other copy of the affected tumor suppressor gene is lost through another mechanism, such as a disabling point mutation or a deletion. One of several examples of a tumor suppressor gene that is hypermethylated in several cancers is *CDKN2A,* which you will recall is a complex locus that encodes two tumor suppressors, p14/ARF and p16/INK4a, that enhance p53 and RB activity, respectively.
- *Global changes in DNA methylation.* In addition to local hypermethylation of tumor suppressor genes, many tumors exhibit abnormal patterns of DNA methylation throughout their genomes. Tumors commonly exhibiting abnormal DNA methylation, such as acute myeloid leukemia, sometimes have mutations in genes encoding DNA methyltransferases or other proteins that influence DNA methylation (see Table 7.9), suggesting that the observed alterations have a genetic basis. The most obvious potential consequence of global changes in methylation is altered expression of multiple genes, which may be overexpressed or underexpressed compared to normal depending on the nature of local changes.
- *Changes in histones.* You will recall that histones are responsible for "packaging" of DNA and that changes in histone positioning or posttranslational modifications (so-called histone marks) regulate gene transcription. Cancer cells often demonstrate changes in histones near genes that influence cellular behavior. As with DNA methylation, in many instances these alterations appear to have a genetic basis, being attributable to mutations in proteins that "write," "read," and "erase" histone marks or that position nucleosomes on DNA (see Table 7.9). Remarkably, in some cancers, driver mutations occur in the histone genes themselves. Details have yet to emerge, but it is certain that these lesions alter the expression of genes that contribute to the malignant phenotype.

Table 7.9 Examples of Epigenomic Regulatory Genes That Are Mutated in Cancer

Gene(s)	Function	Tumor (Approximate Frequency of Mutation)
DNMT3A	DNA methylation	Acute myeloid leukemia (20%)
MLL1	Histone methylation	Acute leukemia in infants (90%)
MLL2	Histone methylation	Follicular lymphoma (90%)
CREBBP/EP300	Histone acetylation	Diffuse large B-cell lymphoma (40%)
ARID1A	Nucleosome positioning/chromatin remodeling	Ovarian clear cell carcinoma (60%), endometrial carcinoma (30%–40%)
SNF5	Nucleosome positioning/chromatin remodeling	Malignant rhabdoid tumor (100%)
PBRM1	Nucleosome positioning/chromatin remodeling	Renal carcinoma (30%)
H3F3A, HIST1H3B	Histone H3 variants (nucleosome components)	Pediatric gliomas (30%–80%, depending on anatomic location)

Much remains to be deciphered about the state of the "epigenome" in various cancers and its contribution to the malignant state, but several aspects of the relationship merit emphasis.

- *The lineage-specificity of certain oncogenes and tumor suppressor genes has an epigenetic basis.* You may have noticed that tumor suppressors and oncoproteins can be broadly divided into two classes, those that are mutated or otherwise dysregulated in many cancers (e.g., RAS, MYC, p53) and those that are mutated in a restricted subset of tumors (e.g., RB in retinoblastoma, VHL in renal cell carcinoma, APC in colon carcinoma) and are thus lineage-restricted. A cancer cell's lineage or differentiation state, like that of normal cells, is dependent on epigenetic modifications that produce a pattern of gene expression that characterizes that particular cell type. It follows that lineage-restricted cancer genes act only within epigenetic contexts in which key oncogenic targets are controlled by these genes. At its extremes, this allows some genes, such as those encoding Notch receptors, to act as a tumor suppressor in one lineage and behave as an oncogene in another. Thus, the *NOTCH1* gene is one of the most commonly mutated tumor suppressor genes in squamous

cell carcinoma of the skin (in which the mutations result in loss of function and lead to impaired differentiation) and is also the most commonly mutated oncogene in T-cell acute lymphoblastic leukemia (in which mutations in different parts of the gene result in gain of function and drive the expression of pro-growth genes such as *MYC*).

- *The epigenome is a therapeutic target.* Because the epigenetic state of a cell depends on reversible modifications that are carried out by enzymes (which are generally good drug targets), there is intense interest in developing drugs that target epigenomic modifiers in cancer and other diseases. Inhibitors of histone deacetylases, chromatin erasers that remove acetyl groups from histones, are approved for use in certain lymphoid tumors, and DNA methylation inhibitors are used to treat myeloid tumors, based in part on the idea that these drugs may reactivate tumor suppressor genes. Other drugs that target specific chromatin writers and chromatin readers are now being tested in clinical trials.
- *Cancers likely exhibit considerable epigenetic heterogeneity.* Just as genomic instability gives rise to genetic heterogeneity in cancers, it is feared that cancers will also prove to have extensive epigenetic heterogeneity from cell to cell within individual tumors. One consequence of such heterogeneity may be drug resistance. For example, epigenetic alterations can lead to the resistance of lung cancer cells to inhibitors of EGF receptor signaling. When the inhibitors are removed, the lung cancer cells revert to their prior inhibitor-sensitive state. If widespread, such epigenetic plasticity may join genetic heterogeneity as yet another barrier to the development of curative cancer therapies.

Noncoding RNAs and Cancer

Noncoding RNAs participate in carcinogenesis by regulating the expression of protein-coding cancer-associated genes. The best characterized of these noncoding RNAs are microRNAs. As discussed in Chapter 1, microRNAs (miRs) are small noncoding, single-stranded RNAs, approximately 22 nucleotides in length, that mediate sequence-specific inhibition of mRNAs. Given that miRs control normal cell survival, growth, and differentiation, it is not surprising that they play a role in carcinogenesis. Altered miR expression, sometimes stemming from amplifications and deletions of miR loci, has been identified in many cancers. Decreased expression of certain miRs increases the translation of oncogenic mRNAs; such miRs have tumor suppressive activity. For example, deletions affecting miR-15 and miR-16 are among the most frequent genetic lesions in chronic lymphocytic leukemia, a common tumor of older adults (Chapter 13). In this tumor, loss of these mIRs leads to upregulation of the anti-apoptotic protein BCL-2, enhancing tumor cell survival. Conversely, overexpression of other miRs represses the expression of tumor suppressor genes; such miRs promote tumor development and are referred to as *onco-miRs.* One example of an onco-miR is miR-155, which is overexpressed in many human B-cell lymphomas and indirectly upregulates a large number of genes that promote proliferation, including *MYC*.

The involvement of miRs may be the proverbial tip of the iceberg with respect to the role of noncoding RNAs in cancer. Systematic genomic analyses have revealed that more than 60% of the genome is transcribed into RNAs, most of which are noncoding and believed to have regulatory functions (Chapter 1).

Molecular Basis of Multistep Carcinogenesis

Given that malignant tumors must acquire multiple "hallmarks" of cancer, it follows that cancers result from the stepwise accumulation of multiple mutations that act in complementary ways to produce a fully malignant tumor. The notion that malignant tumors arise from a sequential accumulation of cancer-promoting alterations is supported by epidemiologic, experimental, and molecular studies, and the study of oncogenes and tumor suppressor genes has provided a firm molecular footing for the concept of multistep carcinogenesis. Genome-wide sequencing of cancers has revealed as few as 10 or so mutations in certain leukemias to many thousands of mutations (most of which are passengers rather than drivers) in tumors that arise following chronic exposure to carcinogens, such as lung cancers associated with cigarette smoking. A more direct answer to the question "how many mutations does it take to establish a fully malignant tumor?" comes from experimental attempts to transform normal human cells with combinations of oncogenes, some derived from transforming viruses (described later). For example, normal human epithelial cells can be transformed by the following combination of events: (1) activation of RAS; (2) inactivation of RB; (3) inactivation of p53; (4) inactivation of PP2A, a tumor suppressive phosphatase that is a negative regulator of many signaling pathways; and (5) constitutive expression of telomerase. Cells bearing all of these alterations are immortal and produce invasive, fully malignant growths when injected into immunodeficient mice.

Unlike in the laboratory, these events presumably never occur simultaneously during the natural development of a human cancer, but instead occur in a stepwise fashion. What is the evidence that this is so? A classic example of incremental acquisition of the malignant phenotype is found in colon carcinoma. Many of these cancers evolve through a series of morphologically identifiable stages, most notably the formation of adenomas that progressively enlarge and ultimately undergo malignant transformation (Chapter 17). Molecular analyses of proliferations at each stage has indeed shown that precancerous lesions have fewer mutations than adenocarcinomas and that certain mutations tend to occur early (e.g., mutations in the tumor suppressor gene *APC*) or late (e.g., mutations in *TP53*) in the process (discussed in detail in Chapter 17). Similar evidence for stepwise progression exists for other recognizable precursor lesions to epithelial cancers, such as dysplasias of the cervix, epidermis, and oral mucosa, and hyperplasias of the endometrium. These are also described in subsequent chapters.

CARCINOGENIC AGENTS AND THEIR CELLULAR INTERACTIONS

More than 200 years ago the London surgeon Sir Percival Pott correctly attributed scrotal skin cancer in chimney sweeps to chronic exposure to soot. Based on this observation,

Table 7.10 Major Chemical Carcinogens

Direct-Acting Carcinogens
Alkylating Agents
β-Propiolactone Dimethyl sulfate Diepoxybutane Anticancer drugs (cyclophosphamide, chlorambucil, nitrosoureas, and others)
Acylating Agents
1-Acetyl-imidazole Dimethylcarbamyl chloride
Procarcinogens That Require Metabolic Activation
Polycyclic and Heterocyclic Aromatic Hydrocarbons
Benz[*a*]anthracene Benzo[*a*]pyrene Dibenz[*a,h*]anthracene 3-Methylcholanthrene 7,12-Dimethylbenz[*a*]anthracene
Aromatic Amines, Amides, Azo Dyes
2-Naphthylamine (β-naphthylamine) Benzidine 2-Acetylaminofluorene Dimethylaminoazobenzene (butter yellow)
Natural Plant and Microbial Products
Aflatoxin B_1 Griseofulvin Cycasin Safrole Betel nuts
Others
Nitrosamine and amides Vinyl chloride, nickel, chromium Insecticides, fungicides Polychlorinated biphenyls

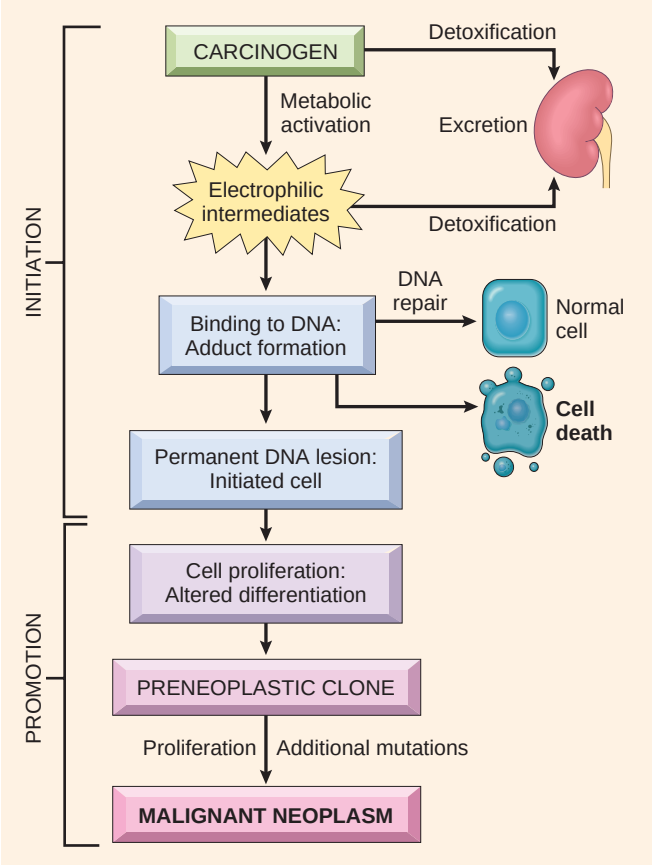


Figure 7.41 General schema of events in chemical carcinogenesis. Note that promoters cause clonal expansion of the initiated cell, thus producing a preneoplastic clone. Further proliferation induced by the promoter or other factors causes accumulation of additional mutations and emergence of a malignant tumor.

the Danish Chimney Sweeps Guild ruled that its members must bathe daily. No public health measure since that time has achieved as much in controlling a form of cancer! Subsequently, hundreds of chemicals have been shown to be carcinogenic in animals. Some of the major agents are listed in Table 7.10.

Chemical Carcinogenesis

As discussed earlier, carcinogenesis is a multistep process. This is readily demonstrated in experimental models of chemical carcinogenesis, in which the stages of initiation and promotion during cancer development were first described. The classic experiments that allowed the distinction between initiation and promotion were performed on mouse skin and revealed the following concepts relating to the initiation-promotion sequence: *Initiation* results from exposure of cells to a sufficient dose of a carcinogenic agent. It causes permanent DNA damage (mutations). *Promoters* can induce tumors to arise from initiated cells, but they are not tumorigenic by themselves. Application of promoters leads to proliferation and clonal expansion of initiated (mutated) cells. Driven to proliferate, subclones of the initiated cells suffer various additional mutations, and eventually a cancerous clone with all the hallmark characteristics emerges. It is likely that many factors contributing to oncogenesis in humans also act by stimulating proliferation and thus can be thought of conceptually as tumor promoters; examples include unopposed estrogenic stimulation of the endometrium and breast and chronic inflammatory processes associated with tissue repair (e.g., inflammatory bowel disease, chronic hepatitis, and Barrett esophagus).

Although the concepts of initiation and promotion have been derived largely from experiments in mice, they are useful concepts when considering the roles of certain factors that contribute to human cancers. With this brief overview, initiation and promotion can be examined in more detail (Fig. 7.41). All initiating chemical carcinogens are highly reactive electrophiles (have electron-deficient atoms) that can react with nucleophilic (electron-rich) atoms in the cell. Their targets are DNA, RNA, and proteins, and in some cases these interactions cause cell death. Initiation, obviously, inflicts nonlethal damage to the DNA that is repaired in some error-prone fashion. The mutated cell then passes on the DNA lesions to its daughter cells. Chemicals that can cause initiation of carcinogenesis fall into two categories: direct acting and indirect acting.

Direct-Acting Carcinogens

***Direct-acting carcinogens* do not require metabolic conversion to become carcinogenic.** Most are weak carcinogens, but some are important because they are cancer chemotherapeutic drugs (e.g., alkylating agents). Tragically, in some instances these agents have cured, controlled, or delayed recurrence of certain types of cancer (e.g., leukemia, lymphoma, or breast carcinoma), only to evoke a second form of cancer, usually acute myeloid leukemia. The risk of induced cancer is low, but its existence dictates judicious use of such agents.

Indirect-Acting Carcinogens

***Indirect-acting carcinogens* require metabolic conversion to become active carcinogens; the carcinogenic products are called *ultimate carcinogens*.** Most chemical carcinogens act indirectly and require metabolic activation for conversion into ultimate carcinogens (Fig. 7.41). Some of the most potent indirect chemical carcinogens—the polycyclic hydrocarbons—are present in fossil fuels. Others, for example, benzo[*a*]pyrene (the active component of soot, which Potts showed to be carcinogenic), are formed during the high-temperature combustion of tobacco in cigarettes and are implicated in the causation of lung cancer. Polycyclic hydrocarbons also are produced from animal fats during the process of broiling or grilling meats and are present in smoked meats and fish. The aromatic amines and azo dyes are another class of indirect-acting carcinogens that were widely used in the past in the aniline dye and rubber industries. Many other occupational carcinogens are listed in Table 7.10.

Most indirect carcinogens are metabolized by *cytochrome P-450–dependent monooxygenases*. The genes that encode these enzymes are polymorphic, and the activity and inducibility of these enzymes vary significantly among individuals (described further in Chapter 9). Because these enzymes are essential for the activation of procarcinogens, the susceptibility to carcinogenesis is related in part to the particular polymorphic variants that an individual inherits. Thus it may be possible to assess cancer risk in a given individual by genetic analysis of such enzyme polymorphisms.

The metabolism of polycyclic aromatic hydrocarbons, such as benzo[*a*]pyrene by the product of the P-450 gene, *CYP1A1,* provides an instructive example. Approximately 10% of the white population carry a highly inducible form of this gene. Light smokers with the susceptible *CYP1A1* genotype have a sevenfold higher risk of developing lung cancer compared with smokers without the permissive genotype. Metabolic pathways also are involved in the inactivation (detoxification) of certain procarcinogens or their derivatives, and variation in these pathways also may influence cancer risk.

Molecular Targets of Chemical Carcinogens. Because malignant transformation results from mutations, it comes as no surprise that most chemical initiating agents target DNA and are mutagenic. There is no single or unique alteration associated with cancer initiation. Nor is there any apparent predisposition for initiators to cause mutations in particular genes; presumably, mutations occur throughout the genome, and cells that by chance suffer damage to the "usual suspects"—oncogenes and tumor suppressors such as *RAS* and *TP53*—gain a potential selective advantage and are at risk for subsequent transformation.

This is not to say that mutations induced by carcinogens occur in an entirely random fashion. Because of their chemical structures, some carcinogens interact preferentially with particular DNA sequences or bases and thus produce mutations that are clustered at "hotspots" or that are enriched for particular base substitutions. This phenomenon is illustrated by a mutational "hotspot" associated with exposure to *aflatoxin* B_1, a naturally occurring agent produced by some strains of the mold *Aspergillus*. *Aspergillus* grows on improperly stored grains and nuts, and there is a strong correlation between the dietary level of this food contaminant and the incidence of hepatocellular carcinoma in parts of Africa and the Far East. Interestingly, aflatoxin B_1–associated hepatocellular carcinomas tend to have a particular mutation in *TP53*, a G:C→T:A transversion in codon 249 that produces an arginine-to-serine substitution in the p53 protein that interferes with its function. In contrast, *TP53* mutations are infrequent in liver tumors from areas where aflatoxin contamination of food does not occur, and few of these mutations involve codon 249. Similarly, lung cancers associated with smoking have a 10-fold higher mutational burden on average than lung cancers in nonsmokers, and these excess mutations are strongly skewed toward particular base substitutions known to be caused by carcinogens in cigarette smoke (the proverbial "smoking gun"). Sequencing of cancer genomes has revealed several dozen other mutational "signatures." One of these signatures reflects exposure to chemotherapeutic agents, but the rest are largely unexplained, suggesting that other carcinogenic agents lurk in the environment, awaiting discovery.

Additional potential carcinogens in the workplace and at home include vinyl chloride, arsenic, nickel, chromium, insecticides, fungicides, and polychlorinated biphenyls. Nitrites used as food preservatives also have caused concern, as they react with amines contained in the food to form nitrosoamines, which are suspected carcinogens.

KEY CONCEPTS

CHEMICAL CARCINOGENESIS

- Chemical carcinogens have highly reactive electrophile groups that directly damage DNA, leading to mutations and eventually cancer.
- Direct-acting agents do not require metabolic conversion to become carcinogenic, while indirect-acting agents are not active until converted to an ultimate carcinogen by endogenous metabolic pathways. Hence, polymorphisms of endogenous enzymes such as cytochrome P-450 may influence carcinogenesis.
- After exposure of a cell to a mutagen or an initiator, tumorigenesis can be enhanced by exposure to promoters, which stimulate proliferation of the mutated cells.
- Examples of human carcinogens are direct-acting agents (e.g., alkylating agents used for chemotherapy), indirect-acting agents (e.g., benzo[*a*]pyrene, azo dyes, aflatoxin), and promoters or agents that cause pathologic hyperplasias of the endometrium or regenerative activity in the liver.

Radiation Carcinogenesis

Radiant energy, in the form of the UV rays of sunlight or as ionizing electromagnetic and particulate radiation, is mutagenic and carcinogenic. UV light exposure causes skin cancers, and ionizing radiation exposure from medical or occupational exposure, nuclear plant accidents, and atomic bomb detonations is associated with a variety of cancers. Although the contribution of ionizing radiation to the total human burden of cancer is probably small, those cancers that do occur may arise decades later, and long periods of observation are necessary to ascertain its full effect. An increased incidence of breast cancer has become apparent decades after women were exposed during childhood to atomic bomb tests. The incidence peaked during 1988–1992 and then declined. Moreover, radiation may have additive or synergistic effects with other potentially carcinogenic factors.

Ultraviolet Rays

Exposure to UV rays derived from the sun, particularly in fair-skinned individuals, is associated with an increased incidence of squamous cell carcinoma, basal cell carcinoma, and melanoma of the skin. The degree of risk depends on the type of UV rays, the intensity of exposure, and skin pigmentation, the latter reflecting the quantity of light-absorbing melanin in the skin. Thus, persons of European origin with fair skin that sunburns easily and stalwartly refuses to tan and who live in locales receiving a great deal of sunlight (e.g., Queensland, Australia, close to the equator) have the highest incidence of skin cancers in the world. The UV portion of the solar spectrum can be divided into three wavelength ranges: UVA (320–400 nm), UVB (280–320 nm), and UVC (200–280 nm). Of these, UVB is believed to be responsible for the induction of cutaneous cancers. UVC, although a potent mutagen, is not considered significant because it is filtered out by the ozone layer surrounding the earth (hence concerns about ozone depletion).

UVB light is carcinogenic because of its ability to cause pyrimidine dimers to form in DNA. Absorption of the energy in a photon of UV light by DNA produces a chemical reaction that leads to covalent cross-linking of pyrimidine bases, particularly adjacent thymidine residues in the same strand of DNA. This distorts the DNA helix and prevents proper pairing of the dimer with bases in the opposite DNA strand. Pyrimidine dimers are repaired by the nucleotide excision repair pathway, a process that may involve 30 or more proteins. It is postulated that with excessive sun exposure, the capacity of the nucleotide excision repair pathway is overwhelmed, and error-prone nontemplated DNA repair mechanisms become operative. These allow the cell to survive but also introduce mutations that, in some instances, lead to cancer. The importance of the nucleotide excision repair pathway of DNA repair is most graphically illustrated by the high frequency of cancers in individuals with the hereditary disorder *xeroderma pigmentosum* (discussed previously). Controversy about the role of UV exposure in the etiology of melanoma was put to rest by sequencing of melanoma genomes. This revealed that melanomas arising in sun-exposed skin harbor enormous numbers of mutations bearing the signature of error-prone repair of pyrimidine dimers, confirming that UV light has an important causative role in this potentially lethal cancer.

Ionizing Radiation

Electromagnetic (x-rays, γ rays) and particulate (α particles, β particles, protons, neutrons) radiations are all carcinogenic. The evidence is voluminous, and a few examples suffice. Many individuals pioneering the use of x-rays developed skin cancers. Miners of radioactive elements in central Europe and the Rocky Mountain region of the United States have a 10-fold higher incidence of lung cancers than the rest of the population. Most telling is the follow-up of survivors of the atomic bombs dropped on Hiroshima and Nagasaki. Initially there was a marked increase in the incidence of certain forms of leukemia after an average latent period of about 7 years. Subsequently the incidence of many solid tumors with longer latent periods (e.g., carcinomas of the breast, colon, thyroid, and lung) increased. Of great concern in the current era of widespread use of computed tomography (CT) scans are studies that have shown that children who get two or three CT scans have a threefold higher risk of leukemia, and those who receive five to 10 such scans have a threefold higher risk of brain tumors. The overall risk in children is very low (roughly one excess leukemia and one excess brain tumor over 10 years per 10,000 CT scans), but nevertheless emphasizes the need to minimize radiation exposure whenever possible.

In humans, for reasons that are not clear, there is a hierarchy of tissue vulnerability to radiation-induced cancers. Most frequent are myeloid leukemias (tumors of granulocytes and their precursors; see Chapter 13). Cancer of the thyroid follows closely but only in young patients. In the intermediate category are cancers of the breast, lungs, and salivary glands. In contrast, skin, bone, and the gastrointestinal tract are relatively resistant to radiation-induced neoplasia, even though gastrointestinal epithelial cells are vulnerable to the acute cell-killing effects of radiation, and the skin is "first in line" for all external radiation. Nonetheless, the physician must not forget: practically *any* cell can be transformed into a cancer cell by sufficient exposure to radiant energy.

KEY CONCEPTS

RADIATION CARCINOGENESIS

- Ionizing radiation causes chromosome breakage, translocations, and, less frequently, point mutations, leading to genetic damage and carcinogenesis.
- UV rays induce the formation of pyrimidine dimers within DNA, leading to mutations. Therefore UV rays can give rise to skin cancers. Individuals with defects in the repair of pyrimidine dimers suffer from xeroderma pigmentosa and are at particularly high risk.
- Exposure to radiation during imaging procedures such as CT scans is linked to a very small, but measurable, increase in cancer risk in children.

Microbial Carcinogenesis

Many RNA and DNA viruses have proved to be oncogenic in animals as disparate as frogs and primates. Despite intense scrutiny, however, only a few viruses have been linked with human cancer. Our discussion focuses on human oncogenic viruses as well as the role of the bacterium *H. pylori* in gastric cancer. A common theme in the pathogenesis

of microbial carcinogenesis is that the infecton triggers cell proliferation, which is initially polyclonal but with time becomes monoclonal by acquisition of driver mutations in rapidly dividing cells.

Oncogenic RNA Viruses

Human T-Cell Leukemia Virus Type 1. Although the study of animal retroviruses has provided spectacular insights into the molecular basis of cancer, only one human retrovirus, human T-cell leukemia virus type 1 (HTLV-1), is firmly implicated in the pathogenesis of cancer in humans.

HTLV-1 causes adult T-cell leukemia/lymphoma (ATLL), a tumor that is endemic in certain parts of Japan, the Caribbean basin, South America, and Africa and found sporadically elsewhere, including the United States. Worldwide, it is estimated that 15 to 20 million people are infected with HTLV-1. Similar to human immunodeficiency virus (HIV), which causes AIDS, HTLV-1 has tropism for CD4+ T cells, and hence this subset of T cells is the major target for neoplastic transformation. Human infection requires transmission of infected T cells via sexual intercourse, blood products, or breastfeeding. Leukemia develops in only 3% to 5% of the infected individuals, typically after a long latent period of 40 to 60 years.

There is little doubt that HTLV-1 infection of T lymphocytes is necessary for leukemogenesis, but the molecular mechanisms of transformation are not defined. In contrast to several murine retroviruses, HTLV-1 does not contain an oncogene, and no consistent integration next to a proto-oncogene has been discovered. In leukemic cells, however, viral integration shows a clonal pattern. In other words, although the site of viral integration in host chromosomes is random (the viral DNA is found at different locations in different cancers), the site of integration is identical within all cells of a given cancer. This would not occur if HTLV-1 were merely a passenger that infects cells after transformation; rather, it means that HTLV-1 must have been present at the moment of transformation, placing it at the "scene of the crime."

The HTLV-1 genome contains the *gag, pol, env,* and long-terminal-repeat regions typical of all retroviruses, but, in contrast to other leukemia viruses, it contains two other genes referred to as *tax* and *HBZ*. Several aspects of HTLV-1's transforming activity may be attributable to the protein products of these genes. Tax is essential for viral replication because it stimulates transcription of viral RNA from the 5′ long terminal repeat. HBZ is a transcription factor, and Tax and HBZ alter the transcription of host cell genes and interact with certain host cell signaling proteins. In doing so, they appear to contribute to the acquisition of cancer hallmarks, though the mechanisms remain unclear; effects on intracellular signals that regulate growth and cell survival, induction of genomic instability, and inhibition of senescence all have been suggested. Whatever the actual mechanism, it is quite inefficient, given the typical latency period of many decades between infection and the development of leukemia, which appears in only a small subset of infected individuals.

Oncogenic DNA Viruses

As with RNA viruses, several oncogenic DNA viruses that cause tumors in animals have been identified. Of the various human DNA viruses, five—HPV, EBV, HBV, Merkel cell polyomavirus, and human herpesvirus 8 (HHV8, also called Kaposi sarcoma herpesvirus)—have been implicated in the causation of human cancer. Merkel cell polyomavirus has been identified in Merkel cell carcinomas and is described in Chapter 25. HHV8 is discussed in Chapters 6 and 11. Although not a DNA virus, HCV is also associated with cancer and is discussed here briefly.

Human Papillomavirus. At least 70 genetically distinct types of HPV have been identified. Some types (e.g., 1, 2, 4, and 7) cause benign squamous papillomas (warts) in humans. In contrast, high-risk HPVs (e.g., types 16 and 18) have been implicated in the genesis of squamous cell carcinomas of the cervix, anogenital region, and head and neck (particularly tumors arising in the tonsillar mucosa). These cancers are sexually transmitted diseases caused by chronic HPV infection. In contrast to cervical cancers, genital warts have low malignant potential and are associated with low-risk HPVs, predominantly HPV-6 and HPV-11.

What explains the variation in cancer risk among HPV strains? In benign warts, the HPV genome is maintained in a nonintegrated episomal form, while in cancers the HPV genome is integrated into the host genome, suggesting that integration of viral DNA is one factor. As with HTLV-1, the site of viral integration in host chromosomes is random, but the pattern of integration is clonal. Integration always occurs in a fashion that interrupts the viral DNA within the E1/E2 open reading frame, leading to loss of the E2 viral repressor and increased expression of the HPV E6 and E7 genes, which are responsible for the oncogenic potential of HPV (Fig. 7.42).

- *Oncogenic activities of E6.* The E6 protein binds to and mediates the degradation of p53 and stimulates the expression of telomerase reverse transcriptase (TERT), the catalytic subunit of telomerase, which you will recall contributes to the immortalization of cells. E6 from high-risk HPV types has a higher affinity for p53 than E6 from

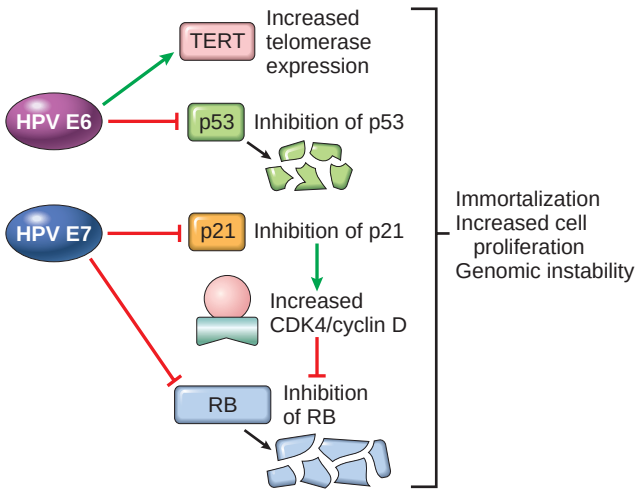


Figure 7.42 Transforming effects of human papillomavirus *(HPV)* E6 and E7 proteins. The net effect of HPV E6 and E7 proteins is to immortalize cells and remove the restraints on cell proliferation (see Fig. 7.26). *TERT,* Telomerase reverse transcriptase. (Modified from Münger K, Howley PM: Human papillomavirus immortalization and transformation functions, *Virus Res* 89:213–228, 2002.)

low-risk HPV types, a distinction that likely accounts for the difference in cancer risk.

- *Oncogenic activities of E7.* The E7 protein has effects that complement those of E6, all of which are centered on speeding cells through the G_1/S cell cycle checkpoint. It binds to the RB protein and displaces the E2F transcription factors that are normally sequestered by RB, promoting progression through the cell cycle. As with E6 proteins and p53, E7 proteins from high-risk HPV types have a higher affinity for RB than do E7 proteins from low-risk HPV types. E7 also inactivates the CDK inhibitors p21 and p27. Finally, E7 proteins from high-risk HPVs (types 16, 18, and 31) also bind and presumably activate cyclins A and E.

To summarize, **high-risk HPV types express oncogenic proteins that inactivate tumor suppressors, activate cyclins, inhibit apoptosis, and combat cellular senescence.** Thus, it is evident that HPV proteins promote many of the hallmarks of cancer. The primacy of HPV infection in the causation of cervical cancer is confirmed by the effectiveness of HPV vaccines in preventing cervical cancer. However, infection with HPV itself is not sufficient for carcinogenesis. For example, when human keratinocytes are transfected with DNA from HPV types 16, 18, or 31 in vitro, they are immortalized but do not form tumors. Cotransfection with a mutated *RAS* gene results in full malignant transformation. In addition to such genetic cofactors, HPV in all likelihood also acts in concert with environmental factors. These include cigarette smoking, coexisting microbial infections, dietary deficiencies, and hormonal changes, all of which have been implicated in the pathogenesis of cervical cancers. A high proportion of women infected with HPV clear the infection by immunologic mechanisms, but others do not because of acquired immune abnormalities, such as those that result from HIV infection, or for unknown reasons. As might be expected, women who are coinfected with high-risk HPV types and HIV have an elevated risk of cervical cancer.

Epstein-Barr Virus. EBV, a member of the herpesvirus family, was the first virus linked to a human tumor, Burkitt lymphoma. Since its initial discovery 50 years ago, EBV has been implicated in the pathogenesis of a diverse collection of human tumors including various lymphomas, several carcinomas, and even rare sarcomas. The most common EBV-associated tumors are lymphomas derived from B cells and nasopharyngeal carcinoma; other EBV-associated neoplasms are discussed elsewhere in this book.

The manner in which EBV causes B-cell tumors such as Burkitt lymphoma is complex and incompletely understood, but best appreciated by considering its effects on normal B cells. EBV has surface glycoproteins that recognize and bind the complement receptor CD21, allowing the virus to attach to and infect B cells. This probably occurs in the tonsils following exposure to the virus in saliva. Viral infection of B cells is latent; that is, there is no viral replication, and the cells are not killed. However, EBV proteins are expressed in latently infected B cells that allow the cells to grow indefinitely (immortalization). The molecular basis of B-cell growth and immortalization is complex, but as with other viruses it involves the "hijacking" of several normal signaling pathways. One EBV gene, latent membrane protein-1 *(LMP-1),* is an oncogene capable of inducing B-cell lymphomas in mice. LMP-1 behaves like a constitutively active CD40 receptor, a key recipient of helper T-cell signals that stimulate B-cell growth (Chapter 6). LMP-1 activates the NF-κB and JAK/STAT signaling pathways and promotes B-cell survival and proliferation, all of which occur autonomously (i.e., without T cells or other outside signals) in EBV-infected B cells. Concurrently, LMP-1 prevents apoptosis by activating BCL2. Thus, the virus "borrows" normal B-cell activation pathways to expand the pool of latently infected cells. Another EBV gene, *EBNA-2,* encodes a nuclear protein that mimics a constitutively active Notch receptor. EBNA-2 transactivates several host genes, including cyclin D and the *SRC* family of proto-oncogenes. In addition, the EBV genome contains a gene encoding a homologue of IL-10 (vIL-10) that was "borrowed" from the host genome. vIL-10 suppresses the activation of T cells by macrophages and contributes to EBV-dependent transformation of B cells.

The EBV proteins that are required for B-cell immortalization and proliferation are highly immunogenic, and in normal individuals the EBV-driven polyclonal B-cell proliferation is readily controlled by a cytotoxic T-cell response. Depending on the timing and intensity of this response, the individual either remains asymptomatic or develops a self-limited episode of infectious mononucleosis (Chapter 8). If T-cell immunity is defective, however, EBV transformed B cells can produce a rapidly progressive, fatal lymphoma.

Burkitt lymphoma is a neoplasm of B lymphocytes that is endemic in central Africa and New Guinea, areas where it is the most common tumor of childhood. A morphologically identical lymphoma occurs sporadically throughout the world. The association between endemic Burkitt lymphoma and EBV is strong:

- More than 90% of endemic tumors carry the EBV genome.
- All affected patients have elevated antibody titers against viral capsid antigens.
- Serum antibody titers against viral capsid antigens are correlated with the risk of developing the tumor.

Although EBV is intimately involved in the causation of Burkitt lymphoma, several observations suggest that additional factors are involved. (1) EBV infection is not limited to the geographic locales where Burkitt lymphoma is found; in fact, it is a ubiquitous virus that infects almost all humans worldwide. (2) The EBV genome is found in only 15% to 20% of Burkitt lymphomas outside of endemic regions. (3) There are significant differences in the patterns of viral gene expression in EBV-transformed (but not tumorigenic) B-cell lines and Burkitt lymphoma cells. Most notably, Burkitt lymphoma cells do not express LMP-1, EBNA2, and other EBV proteins that drive B-cell growth and immortalization.

Given these observations, how then does EBV contribute to the genesis of endemic Burkitt lymphoma? One possibility is shown in Fig. 7.43. In regions where Burkitt lymphoma is endemic, concomitant infections such as malaria impair immune competence, allowing sustained B-cell proliferation. Eventually, T-cell immunity directed against EBV antigens such as EBNA2 and LMP-1 eliminates most of the EBV-infected B cells, but a small number of cells downregulate expression of these immunogenic antigens. These cells persist indefinitely, even in the face of normal immunity. Lymphoma cells may

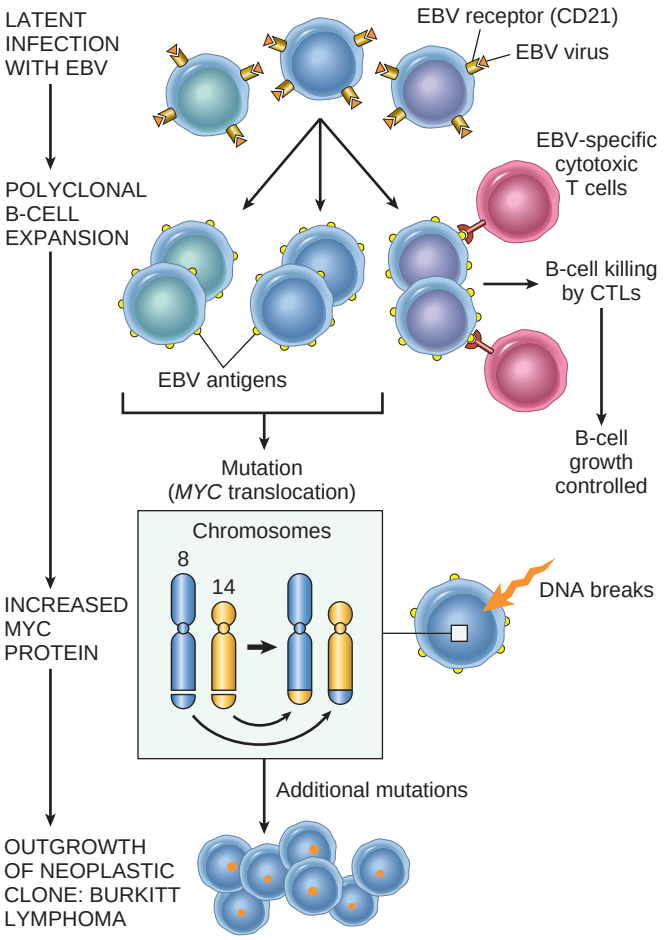


Figure 7.43 Pathogenesis of Epstein-Barr virus *(EBV)*–induced Burkitt lymphoma. *CTLs*, Cytotoxic T lymphocytes.

emerge from this population only with the acquisition of specific mutations, most notably translocations involving the *MYC* oncogene, as virtually all endemic and sporadic tumors possess the t(8;14) or other translocations that dysregulate *MYC*. Thus although sporadic Burkitt lymphomas are triggered by mechanisms other than EBV, they appear to develop through similar oncogenic pathways.

In summary, **in the case of Burkitt lymphoma, it seems that EBV is not directly oncogenic, but by acting as a polyclonal B-cell mitogen sets the stage for the acquisition of the (8;14) translocation and other mutations that ultimately produce a full-blown cancer.** In most individuals, EBV infection is readily controlled by effective immune responses, and lymphomagenesis is rare. By contrast, in regions where Burkitt lymphoma is endemic, cofactors such as chronic malaria may favor the acquisition of additional genetic events (e.g., t(8;14)) that lead to transformation.

The role played by EBV is more direct in B-cell lymphomas arising in immunosuppressed patients. Some persons with AIDS or who receive immunosuppressive therapy for preventing allograft rejection develop EBV-positive B-cell tumors, often at multiple sites and within extranodal tissues such as the gut or the central nervous system. These proliferations are polyclonal at the outset but can evolve into monoclonal neoplasms. In contrast to Burkitt lymphoma, tumors in immunosuppressed patients usually express LMP-1 and EBNA2, which are antigenic and would normally be recognized by cytotoxic T cells. Also, in contrast to Burkitt lymphoma, B-cell tumors in immunosuppressed individuals usually lack *MYC* translocations. These potentially lethal proliferations can be subdued if T-cell immunity can be restored, as may occur with withdrawal of immunosuppressive drugs in transplant recipients.

Nasopharyngeal carcinoma also is strongly associated with EBV. This tumor is endemic in southern China, parts of Africa, and the Inuit population of the Arctic. In contrast to Burkitt lymphoma, all nasopharyngeal carcinomas obtained from all parts of the world contain EBV, and antibody titers to viral capsid antigens are uniformly elevated in affected patients. The structure of the viral genome is identical (clonal) within individual tumors, indicating that EBV infection occurred before tumor development. EBV thus has a central role in the genesis of nasopharyngeal carcinoma, but (as with Burkitt lymphoma) its restricted geographic distribution indicates that genetic or environmental cofactors also contribute to its development. Unlike Burkitt lymphoma, LMP-1 is expressed in nasopharyngeal carcinoma cells and (as in B cells) activates the NF-κB pathway, which upregulates the expression of factors such as VEGF, FGF-2, MMP-9, and COX-2 that may contribute to oncogenesis. Nasopharyngeal carcinomas typically contain prominent infiltrates composed of T cells, which may be responding to viral antigens such as LMP-1, but this response is ineffective, suggesting that immune evasion mechanisms are likely to be important in this cancer. In line with this idea, nasopharyngeal carcinoma cells often express the immune checkpoint molecule PD-L1 and are responsive to PD-L1 inhibitors. Interestingly, EBV-positive carcinomas resembling nasopharyngeal carcinoma occasionally arise at other sites, such as the stomach and the thymus.

The relationship of EBV to the pathogenesis of Hodgkin lymphoma, yet another EBV-associated tumor, is discussed in Chapter 13.

***Hepatitis B and C Viruses.* Worldwide, 70% to 85% of hepatocellular carcinomas are associated with infection with HBV or HCV.** HBV is endemic in countries of the Far East and Africa; correspondingly, these areas have the highest incidence of hepatocellular carcinoma. Despite compelling evidence incriminating HBV and HCV, the mode of action of these viruses in liver tumorigenesis is not fully elucidated. Oncogenes have yet to be identified in HBV or HCV genomes, and although the HBV DNA is integrated within the human genome, there is no consistent pattern of integration in liver cells. Indeed, while the oncogenic effects of HBV and HCV are multifactorial, the dominant effect seems to be immunologically mediated chronic inflammation and hepatocyte death leading to hepatocyte proliferation during regeneration and, over time, genomic damage.

As with any cause of hepatocellular injury, chronic viral infection leads to the compensatory proliferation of hepatocytes. This regenerative process is aided and abetted by a plethora of growth factors, cytokines, chemokines, and other bioactive substances. These are produced by activated immune cells and promote cell survival, tissue remodeling, and angiogenesis (Chapter 3). The activated immune cells

also produce other mediators, such as reactive oxygen species, that are genotoxic and mutagenic. One key molecular step may be activation of the NF-κB pathway, which blocks apoptosis, allowing the dividing hepatocytes to incur genotoxic stress and to accumulate mutations. Although this seems to be a dominant mechanism in the pathogenesis of virus-induced hepatocellular carcinoma, the HBV genome also contains genes that may directly promote the development of cancer. For example, an HBV gene known as *HBx* can activate a variety of transcription factors and several signal transduction pathways. In addition, viral integration can cause structural changes in chromosomes that dysregulate oncogenes and tumor suppressor genes.

Although not a DNA virus, HCV is also strongly linked to the pathogenesis of liver cancer. The molecular mechanisms used by HCV are less well defined than are those of HBV. In addition to chronic liver cell injury and compensatory regeneration, components of the HCV genome, such as the HCV core protein, may have a direct effect on tumorigenesis, possibly by activating a variety of growth-promoting signal transduction pathways.

Helicobacter pylori

First incriminated as a cause of peptic ulcers, *H. pylori* now has acquired the dubious distinction of being the first bacterium classified as a carcinogen. Indeed, *H. pylori* infection is implicated in the genesis of both gastric adenocarcinomas and gastric lymphomas.

The proposed scenario for the development of gastric adenocarcinoma in the setting of *H. pylori* infection is similar to that of HBV- and HCV-induced liver cancer, as it involves increased epithelial cell proliferation in a background of chronic inflammation. As in viral hepatitis, the inflammatory milieu contains numerous genotoxic agents, such as reactive oxygen species. The *H. pylori* genome also contains genes directly implicated in oncogenesis. Strains associated with gastric adenocarcinoma have been shown to contain a "pathogenicity island" that contains cytotoxin-associated A *(CagA)* gene. Although *H. pylori* is noninvasive, *CagA* penetrates into gastric epithelial cells, where it has a variety of effects including the initiation of a signaling cascade that mimics unregulated growth factor stimulation. The infection initially leads to development of chronic gastritis, followed by gastric atrophy, intestinal metaplasia of the lining cells, dysplasia, and cancer. This sequence takes decades to complete and occurs in only 3% of infected patients.

H. pylori is specifically associated with the development of gastric lymphomas of B-cell origin (also discussed in Chapters 13 and 17). Their molecular pathogenesis is incompletely understood but seems to involve strain-specific *H. pylori* factors as well as host genetic factors, such as polymorphisms in the promoters of inflammatory cytokines such as IL-1β and TNF. It is thought that *H. pylori* infection leads to the appearance of *H. pylori*–reactive T cells, which in turn stimulate a polyclonal B-cell proliferation. In chronic infections, currently unknown mutations may be acquired that give individual cells a growth advantage. These cells grow out into a monoclonal MALToma that nevertheless remains dependent on T-cell stimulation of B-cell pathways that activate the transcription factor NF-κB. At this stage, eradication of *H. pylori* by antibiotic therapy "cures" the lymphoma by removing the antigenic stimulus for T cells. At later stages, however, additional mutations may be acquired that cause constitutive NF-κB activation. At this point, the MALToma no longer requires the antigenic stimulus of the bacterium for growth and survival and develops the capacity to spread beyond the stomach to other tissues.

KEY CONCEPTS

VIRAL AND BACTERIAL ONCOGENESIS``

HTLV-1: a retrovirus that is endemic in Japan, the Caribbean, and parts of South America and Africa that causes adult T-cell leukemia/lymphoma.

- HTLV-1 encodes two viral proteins, Tax and HBX, which are suspected to contribute to leukemogenesis through uncertain mechansisms.
- After a long latent period (decades), a small fraction of HTLV-1–infected individuals develop adult T-cell leukemia/lymphoma, a CD4+ tumor that arises from an HTLV-1–infected cell, presumably due to acquisition of additional mutations in the host cell genome.

HPV: an important cause of benign warts, cervical cancer, and oropharyngeal cancer.

- Oncogenic types of HPV encode the viral oncoproteins E6 and E7, which bind to p53 and Rb, respectively, with high affinity and neutralize their function.
- Development of cancer is associated with integration of HPV into the host genome and additional mutations needed for acquisition of cancer hallmarks.
- HPV cancers can be prevented by vaccination against high-risk HPV types.

EBV: ubiquitous herpesvirus implicated in the pathogenesis of Burkitt lymphomas, B-cell lymphomas in patients with T-cell immunosuppression (HIV infection, transplant recipients), and several other cancers.

- The EBV genome harbors several genes encoding proteins that trigger B-cell signaling pathways; in concert, these signals are potent inducers of B-cell growth and transformation.
- In the absence of T-cell immunity, EBV-infected B cells can rapidly "grow out" as aggressive B-cell tumors.
- In the presence of normal T-cell immunity, a small fraction of infected patients develop EBV-positive B-cell tumors (Burkitt lymphoma, Hodgkin lymphoma) or carcinomas (e.g., nasopharyngeal carcinoma).

HBV and HCV: cause of 70% to 85% of hepatocellular carcinomas worldwide.

- Oncogenic effects are multifactorial; dominant effect seems to be immunologically mediated chronic inflammation, hepatocellular injury, and reparative hepatocyte proliferation.
- HBx protein of HBV and the HCV core protein can activate signal transduction pathways that also may contribute to carcinogenesis.

H. pylori: implicated in gastric adenocarcinoma and MALToma.

- Pathogenesis of *H. pylori*–induced gastric cancers is multifactorial including chronic inflammation and reparative gastric cell proliferation.
- *H. pylori* pathogenicity genes, such as *CagA*, also may contribute by stimulating growth factor pathways.
- Chronic *H. pylori* infection leads to polyclonal B-cell proliferations that may give rise to a B-cell lymphoma (MALToma) of the stomach as a result of accumulation of mutations.

CLINICAL ASPECTS OF NEOPLASIA

Clinical Manifestations

Ultimately the importance of neoplasms lies in their effects on patients. Although malignant tumors are of course more threatening than benign tumors, any tumor, even a benign one, may cause morbidity and mortality.

Local and Hormonal Effects

Location is a critical determinant of the clinical effects of benign and malignant tumors. Tumors may impinge upon vital tissues and impair their function, cause death of involved tissues, and provide a nidus for infection. A small (1 cm) pituitary adenoma, although benign and possibly nonfunctional, can compress and destroy the surrounding normal gland and thus lead to serious hypopituitarism. Cancers arising within or metastatic to an endocrine gland may cause an endocrine insufficiency by destroying the gland. Neoplasms in the gut, both benign and malignant, may cause obstruction as they enlarge. Infrequently, peristaltic movement telescopes the neoplasm and its affected segment into the downstream segment, producing an obstructing intussusception (Chapter 17). Symptoms produced by a cancer due to its position can (ironically) be life-saving; for example, the few survivors of pancreatic cancer are those whose tumors "fortuitously" obstruct bile ducts early in their course, leading to the appearance of jaundice and other symptoms at a stage of the disease when surgical cure is possible.

Benign and malignant neoplasms arising in endocrine glands can cause clinical problems by producing hormones. Such functional activity is more typical of benign than of malignant tumors, which are more likely to be poorly differentiated and nonfunctional. A benign beta-cell adenoma of the pancreatic islets less than 1 cm in diameter may produce sufficient insulin to cause fatal hypoglycemia. In addition, nonendocrine tumors may elaborate hormones or hormone-like products and give rise to paraneoplastic syndromes (discussed later). The erosive and destructive growth of cancers or the expansile pressure of a benign tumor on any natural surface, such as the skin or mucosa of the gut, may cause ulcerations, secondary infections, and bleeding. Melena (blood in the stool) and hematuria, for example, are characteristic of neoplasms of the gut and urinary tract.

Cancer Cachexia

Cancer cachexia is a hypercatabolic state defined by a loss of muscle mass (with or without loss of fat) that cannot be explained by diminished food intake. It occurs in about 50% of cancer patients, most commonly in individuals with advanced gastrointestinal, pancreatic, and lung cancers, and is responsible for about 30% of cancer deaths. It is a highly debilitating condition characterized by extreme weight loss, fatigue, muscle atrophy, anemia, anorexia, and edema. Mortality is generally the consequence of atrophy of the diaphragm and other respiratory muscles.

The precise causes of cancer cachexia are not known, but inflammatory mediators, particularly TNF, IL-1, and IL-6, appear to have important roles. Administration of any of these cytokines to mice induces cachexia, whereas tumor-bearing mice are protected from cachexia by knockout of the TNF receptor. These observations are bolstered by clinical studies showing that cachexia in cancer patients is associated with higher levels of inflammatory cytokines. Evidence suggests that muscle loss occurs through a direct effect of inflammatory cytokines on skeletal muscle cells. Specifically, it appears that cytokines increase the degradation of major skeletal muscle structural proteins, such as myosin heavy chain, through signaling pathways that lead to ubiquitination of target proteins followed by proteolysis via the proteasome. It is also worth noting that similar cachetic states may be seen in patients without cancer, such as in those with chronic disseminated infections and AIDS, presumably at least in part due to the effects of inflammatory cytokines.

However, it must also be recognized that therapies directed against individual cytokines (e.g., TNF) in cancer patients have not been effective in reversing cachexia, suggesting that either a multiplicity of inflammatory cytokines, or other factors entirely, are also involved in its pathogenesis. In line with the latter possibility, while muscle wasting is the cardinal feature of cancer cachexia, many patients lose fat stores as well. One factor that may contribute to fat loss is a protein called lipid mobilizing factor, which has been detected in the sera and urine of patients with advanced cancer and which appears to sensitize adipocytes to lipolytic stimuli. It is likely that additional mechanisms underlying cancer cachexia await discovery.

Paraneoplastic Syndromes

Some cancer-bearing individuals develop signs and symptoms that cannot readily be explained by the anatomic distribution of the tumor or by the elaboration of hormones indigenous to the tissue from which the tumor arose; these are known as paraneoplastic syndromes. These occur in about 10% of persons with cancer. Paraneoplastic syndromes are important to recognize for several reasons:

- They may be the earliest manifestation of an occult neoplasm.
- In affected patients they can cause significant clinical problems and may even be lethal.
- They may mimic metastatic disease and therefore confound treatment.

A classification of paraneoplastic syndromes and their presumed origins is presented in Table 7.11. A few comments on some of the more common and interesting syndromes follow.

Endocrinopathies are frequently encountered paraneoplastic syndromes. By definition, the responsible cancers are not of endocrine origin and the secretory activity of such tumors is referred to as ectopic hormone production. *Cushing syndrome* is the most common endocrinopathy. Approximately 50% of affected individuals have carcinoma of the lung, chiefly the small-cell type. It is caused by excessive production of corticotropin (adrenocorticotropic hormone [ACTH]) or corticotropin-like peptides. The precursor of corticotropin is a large molecule known as pro-opiomelanocortin. Lung cancer patients with Cushing syndrome have elevated serum levels of both pro-opiomelanocortin and corticotropin. The former is not found in serum of patients with excess corticotropin produced by the pituitary.

Table 7.11 Paraneoplastic Syndromes

Clinical Syndromes	Major Forms of Underlying Cancer	Causal Mechanism
Endocrinopathies		
Cushing syndrome	Small cell carcinoma of lung Pancreatic carcinoma Neural tumors	ACTH or ACTH-like substance
Syndrome of inappropriate antidiuretic hormone secretion	Small cell carcinoma of lung Intracranial neoplasms	Antidiuretic hormone or atrial natriuretic hormones
Hypercalcemia	Squamous cell carcinoma of lung Breast carcinoma Renal carcinoma Adult T-cell leukemia/lymphoma	Parathyroid hormone-related protein (PTHRP), TGF-α, TNF, IL-1
Hypoglycemia	Ovarian carcinoma Fibrosarcoma Other mesenchymal sarcomas	Insulin or insulin-like substance
Polycythemia	Renal carcinoma Cerebellar hemangioma Hepatocellular carcinoma	Erythropoietin
Osteomalacia	Phosphaturic mesenchymal tumor	FGF-23
Nerve and Muscle Syndromes		
Myasthenia	Bronchogenic carcinoma Thymic neoplasms	Immunologic
Disorders of the central and peripheral nervous systems	Breast carcinoma	
Dermatologic Disorders		
Acanthosis nigricans	Gastric carcinoma Lung carcinoma Uterine carcinoma	Immunologic; secretion of epidermal growth factor
Dermatomyositis	Bronchogenic carcinoma Breast carcinoma	Immunologic
Osseous, Articular, and Soft Tissue Changes		
Hypertrophic osteoarthropathy and clubbing of the fingers	Bronchogenic carcinoma Thymic neoplasms	Unknown
Vascular and Hematologic Changes		
Venous thrombosis (Trousseau phenomenon)	Pancreatic carcinoma Bronchogenic carcinoma Other cancers	Tumor products (mucins that activate clotting)
Disseminated intravascular coagulation	Acute promyelocytic leukemia Prostatic carcinoma	Tumor products that activate clotting
Nonbacterial thrombotic endocarditis	Advanced cancers	Hypercoagulability
Red cell aplasia	Thymic neoplasms	Unknown
Others		
Nephrotic syndrome	Various cancers	Tumor antigens, immune complexes

ACTH, Adrenocorticotropic hormone; *IL*, interleukin; *TGF*, transforming growth factor; *TNF*, tumor necrosis factor; *FGF-23*, fibroblast growth factor-23.

Hypercalcemia is probably the most common paraneoplastic syndrome; in fact, symptomatic hypercalcemia is more often related to cancer than to hyperparathyroidism. Two general processes are involved in cancer-associated hypercalcemia: (1) *osteolysis* induced by cancer, whether primary in bone, such as multiple myeloma, or metastatic to bone from any primary lesion, and (2) the production of *calcemic humoral substances* by extraosseous neoplasms. Only the second mechanism is considered to be paraneoplastic.

The humoral factor that is most commonly associated with paraneoplastic hypercalcemia is *parathyroid hormone–related protein* (PTHRP). As its name implies, PTHRP has partial structural homology to parathyroid hormone (PTH). PTHRP and PTH bind to the same G protein–coupled receptor, known as the PTH/PTHRP receptor (often referred to as PTH-R or PTHRP-R), and share some, but not all, biologic activities. Like PTH, PTHRP increases bone resorption and renal calcium uptake, while inhibiting renal phosphate transport, effects that raise serum calcium levels. In contrast to PTH, PTHRP is produced in small amounts by many normal tissues, including epithelial cell types such as keratinocytes, which may explain the relatively frequent

association of squamous cell carcinomas with PTHRP-induced hypercalcemia. In addition to PTHRP, several other factors, such as IL-1, TGF-α, TNF, and dihydroxyvitamin D, have been causally implicated in the hypercalcemia of malignancy.

The tumors that are most often associated with paraneoplastic hypercalcemia are carcinomas of the breast, lung, kidney, and ovary. In breast cancers, paraneoplastic hypercalcemia is often exacerbated by osteolytic bone metastases. The most common lung neoplasm associated with hypercalcemia is squamous cell carcinoma, typically because of release of PTHRP.

The *neuromyopathic paraneoplastic syndromes* take diverse forms, such as peripheral neuropathies, encephalitis, cortical cerebellar degeneration, a polymyopathy resembling polymyositis, and a myasthenic syndrome similar to *myasthenia gravis* (Chapter 27). The cause of these syndromes is poorly understood, but appears to involve a cancer-induced immunologic attack on normal tissues. The initiating event may be the ectopic expression of antigens that normally are restricted to the neuromuscular system by tumor cells. For unknown reasons, the immune system recognizes these antigens as foreign and mounts a response that leads to tissue damage. This often takes the form of T-cell responses; in some cases, antibodies that cross-react with neuronal cell antigens are detected.

Acanthosis nigricans is a disorder characterized by gray-black patches of thickened, hyperkeratotic skin with a velvety appearance. It occurs rarely as a genetically determined disease in juveniles or adults (Chapter 25). In about 50% of the cases, particularly in adults older than age 40, the appearance of such lesions is associated with cancer, most commonly carcinoma of the stomach. Acanthosis nigricans is sometimes accompanied by the abrupt development of multiple seborrheic keratoses (Leser-Trélat sign). These skin changes may appear before the cancer is discovered.

Hypertrophic osteoarthropathy is encountered in 1% to 10% of patients with lung carcinoma. Rarely, other forms of cancer are involved. This disorder is characterized by (1) periosteal new bone formation, primarily at the distal ends of long bones, metatarsals, metacarpals, and proximal phalanges; (2) arthritis of the adjacent joints; and (3) clubbing of the digits. Although osteoarthropathy is seldom seen in non-cancer patients, clubbing of the fingertips may be encountered in patients with liver diseases, diffuse lung disease, congenital cyanotic heart disease, ulcerative colitis, and other disorders. The cause is unknown.

Several vascular and hematologic manifestations may appear in association with a variety of cancers. As mentioned in the discussion of thrombosis (Chapter 4), *migratory thrombophlebitis* (Trousseau syndrome) may be encountered in association with deep-seated cancers, most often carcinomas of the pancreas or lung. *Disseminated intravascular coagulation* may complicate a diversity of clinical disorders (Chapter 14); among cancers, it is most commonly associated with acute promyelocytic leukemia and prostatic adenocarcinoma. Bland, small, nonbacterial fibrinous vegetations sometimes form on the cardiac valve leaflets (more often on left-sided valves), particularly in individuals with advanced mucin-secreting adenocarcinomas. These lesions, called *nonbacterial thrombotic endocarditis,* are described further in Chapter 12. The vegetations are potential sources of emboli that can further complicate the course of the cancer.

Grading and Staging of Tumors

Methods to quantify the probable clinical aggressiveness of a given neoplasm and its apparent extent and spread are necessary for accurate prognostication and for comparing results of various treatments. For instance, the results of treating well-differentiated thyroid adenocarcinoma that is localized to the thyroid gland will be different from those obtained from treating highly anaplastic thyroid cancers that have invaded surrounding tissues. Systems have been developed that use the level of differentiation, or grade, and the extent of cancer spread, or stage, as parameters of the clinical gravity of the disease.

- *Grading.* Grading of a cancer is based on the degree of differentiation of the tumor cells and, in some cancers, the number of mitoses or architectural features. Grading schemes have evolved for each type of malignancy and generally range from two (low grade and high grade) to four categories. Criteria for the individual grades vary in different types of tumors and so are not detailed here, but all attempt, in essence, to judge the extent to which the tumor cells resemble or fail to resemble their normal counterparts. Although histologic grading is useful, the correlation between histologic appearance and biologic behavior is less than perfect. In recognition of this problem and to avoid spurious quantification, it is common practice to characterize a particular neoplasm in descriptive terms, for example, well-differentiated, mucin-secreting adenocarcinoma of the stomach, or poorly differentiated pancreatic adenocarcinoma.
- *Staging.* The staging of solid cancers is based on the size of the primary lesion, whether it has spread to regional lymph nodes, and the presence or absence of blood-borne metastases. The major staging system currently in use is the American Joint Committee on Cancer Staging. This system uses a classification called the *TNM system—T* for primary tumor, *N* for regional lymph node involvement, and *M* for metastases. TNM staging varies for specific forms of cancer, but there are general principles. The primary lesion is characterized as T1 to T4 based on increasing size. T0 is used to indicate an in situ lesion. N0 would mean no nodal involvement, whereas N1 to N3 would denote involvement of an increasing number and range of nodes. M0 signifies no distant metastases, whereas M1 or sometimes M2 indicates the presence of metastases and some judgment as to their number.

It is increasingly apparent, however, that the molecular features of solid tumors provide important complementary prognostic information that is independent of anatomic staging of cancers. In recognition of this, the American Joint Committee on Cancer (AJCC) has included molecular characterization as an essential part of the standard work-up of early breast cancers (T1–2, N0, M0) in the latest edition of the cancer staging manual. It can be anticipated that as molecular characterization of all cancers becomes widespread, many more prognostic schemes incorporating both anatomic and molecular information will become a routine part of standard-of-care clinical and pathologic practice.

KEY CONCEPTS

CLINICAL ASPECTS OF TUMORS

Cachexia: progressive loss of body fat and lean body mass, accompanied by profound weakness, anorexia, and anemia, that is caused by release of factors by the tumor or host immune cells.

Paraneoplastic syndromes: symptom complexes in individuals with cancer that cannot be explained by tumor spread or release of hormones that are indigenous to the tumor "cell of origin." For example:

- Endocrinopathies (Cushing syndrome, hypercalcemia)
- Neuropathic syndromes (polymyopathy, peripheral neuropathies, encephalopathy, neural degeneration, myasthenic syndromes)
- Skin disorders (acanthosis nigricans)
- Skeletal and joint abnormalities (hypertrophic osteoarthritis)
- Hypercoagulability (migratory thrombophlebitis, disseminated intravascular coagulation, nonbacterial thrombotic endocarditis)
- Grading: determined by cytologic appearance; based on the idea that behavior and differentiation are related, with poorly differentiated tumors having more aggressive behavior.
- Staging: determined by surgical exploration or imaging, is based on size, local and regional lymph node spread, and distant metastases; of greater clinical value than grading.

Laboratory Diagnosis of Cancer

Every year the laboratory diagnosis of cancer becomes more complex, more sophisticated, and more "personalized," increasingly allowing patients to receive therapy that is tailored to the molecular characteristics of their specific tumor. The following sections present the current state of this rapidly evolving art, starting with older standard methods and then moving to new molecular approaches, avoiding technologic details throughout.

Histologic and Cytologic Methods. The laboratory diagnosis of cancer is, in most instances, not difficult. The two ends of the benign-malignant spectrum pose no problem; however, in the middle is a gray zone where even experts tread cautiously. The focus here is on the roles of the clinician (often a surgeon) and the pathologist in facilitating the correct diagnosis.

Clinical data are invaluable for accurate pathologic diagnosis, but clinicians often underestimate its value. Radiation changes in the skin or mucosa can be similar to those associated with cancer. Sections taken from a healing fracture can mimic an osteosarcoma. Moreover, the laboratory evaluation of a lesion is only as good as the specimen made available for examination. It must be adequate, representative, and properly preserved. Several sampling approaches are available: (1) excision or biopsy, (2) needle aspiration, and (3) cytologic smears. When excision is not possible, selection of an appropriate site for biopsy of a mass requires awareness that the periphery may not be representative and the center may be largely necrotic. Appropriate preservation involves such actions as prompt immersion of at least a portion of the specimen in a fixative (usually a formalin solution) and (depending on the differential diagnosis) rapid allocation of tissue for other studies such as cytogenetics, flow cytometry, and molecular diagnostics (described later). Requesting "quick-frozen section" diagnosis is sometimes desirable, for example, in determining the nature of a mass lesion, in evaluating the margins of an excised cancer to ascertain that the entire neoplasm has been removed, or in making decisions about what additional studies beyond histology are needed. This method permits histologic evaluation within minutes. In experienced, competent hands, frozen-section diagnosis is highly accurate, but there are particular instances in which the superior morphologic detail provided by standard histology is needed—for example, when extremely radical surgery, such as the amputation of an extremity, may be indicated. Better to wait a day or two, despite the delay, than to perform inadequate or unnecessary surgery.

Fine-needle aspiration of tumors is another approach that is widely used. The procedure involves aspirating cells and attendant fluid with a small-bore needle, followed by cytologic examination of the stained smear. This method is used most commonly for assessment of readily palpable lesions in sites such as the breast, thyroid, and lymph nodes. With guidance from imaging, the method can also be used to evaluate lesions in deep-seated structures, such as pelvic lymph nodes and pancreas. Fine-needle aspiration is less invasive and more rapidly performed than needle biopsies, and it obviates surgery and its attendant risks. While it may be confounded by sampling errors, in experienced hands it is rapid and quite reliable.

Cytologic smears provide yet another method for cancer detection (Chapter 22). This approach is most widely used to screen for carcinoma of the cervix and its precursor lesions, but it is suitable for evaluation of any form of suspected malignancy in which tumor cells are shed into fluids or are easily accessible. Types of specimens that are commonly examined in cytologic smears for cancer cells include urine, cerebrospinal fluid, pleural effusions, and bronchial washes.

As pointed out earlier, cancer cells have lowered cohesiveness and exhibit a range of morphologic changes encompassed by the term *anaplasia.* Thus shed cells can be evaluated for the features of anaplasia indicative of their origin from a tumor (Fig. 7.44). In these cases, judgment must be rendered based on the features of individual cells or, at most, a clump of cells, without the supporting evidence of loss of orientation of one cell to another, and (most importantly) evidence of invasion. This method permits differentiation among normal, dysplastic, and malignant cells and, in addition, permits the recognition of cellular changes characteristic of carcinoma in situ. The gratifying control of cervical cancer through screening with Pap smears is the best testament to the value of cytology.

Although histology and exfoliative cytology remain the foundation of cancer diagnosis, they have inherent limits; for example, it can be difficult to determine the nature and tissue of origin of a poorly differentiated tumor, and some specific tumor types are notoriously difficult to distinguish based on their morphologic appearance alone (e.g., various acute leukemias and lymphomas). These limitations have spurred the widespread application of immunohistochemistry and flow cytometry, which can be used to make more accurate diagnoses. Another rapidly expanding modality is molecular diagnostics, which is being used increasingly to identify cancers that are amenable to treatment with

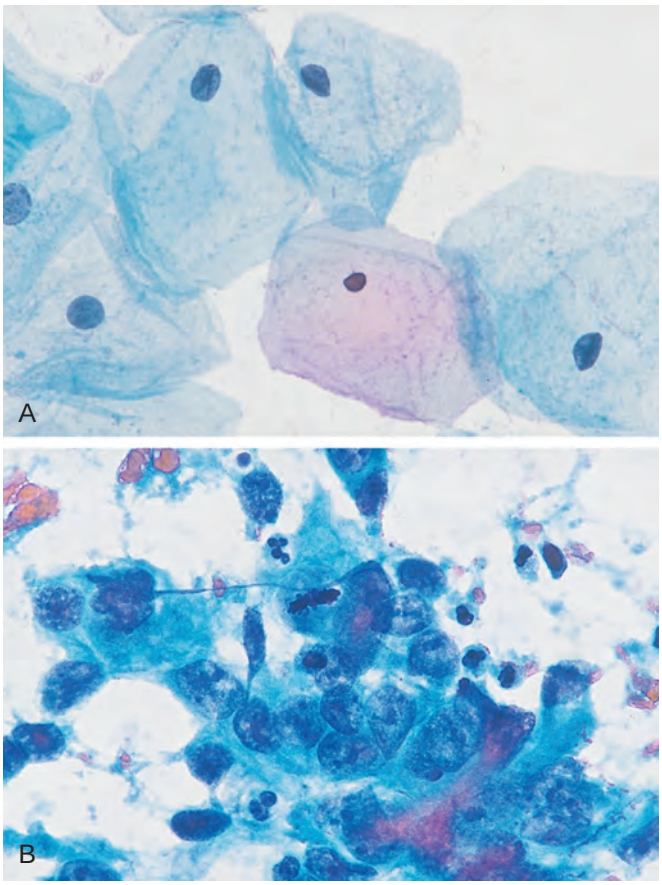


Figure 7.44 Cytology of cervical smears. (A) Normal cervicovaginal smear shows large, flattened squamous cells and groups of metaplastic cells; interspersed are neutrophils. There are no malignant cells. (B) Abnormal cervicovaginal smear shows numerous malignant cells that have pleomorphic, hyperchromatic nuclei; interspersed are normal polymorphonuclear leukocytes. (Courtesy Dr. P. K. Gupta, University of Pennsylvania, Philadelphia, Pa.)

so-called targeted therapies, drugs that are directed at mutated oncoproteins. Only some highlights of these diagnostic modalities are presented.

Immunohistochemistry. The availability of specific antibodies has greatly facilitated the identification of cell products or surface markers. Examples of the utility of immunohistochemistry in the diagnosis or management of malignant neoplasms follow.

- *Categorization of undifferentiated malignant tumors.* In many cases malignant tumors of diverse types resemble each other morphologically because of limited differentiation. Such tumors may be impossible to distinguish in routine hematoxylin and eosin (H&E)–stained tissue sections. For example, certain anaplastic carcinomas, lymphomas, melanomas, and sarcomas may look quite similar, but they must be accurately identified because their treatment and prognosis are different. Antibodies specific to intermediate filaments have proved to be of particular value in such cases because solid tumor cells often contain intermediate filaments characteristic of their cell of origin. For example, the presence of cytokeratins, detected by immunohistochemistry, points to an epithelial origin (carcinoma) (Fig. 7.45), whereas desmin is specific for neoplasms of muscle cell origin, and hematologic malignancies lack these cytoskeletal structures. Other useful immunohistochemical markers include lineage-specific membrane proteins (e.g., CD20, a marker of B-cell tumors) and transcription factors.
- *Determination of site of origin of metastatic tumors.* Many cancer patients present with metastases. In some the primary site is obvious or readily detected on the basis of clinical or radiologic features. In cases in which the origin of the tumor is obscure, immunohistochemical detection of tissue-specific or organ-specific antigens in a biopsy specimen of the metastatic deposit can lead to the identification of the tumor source. For example, prostate-specific antigen (PSA) and thyroglobulin are markers of carcinomas of the prostate and thyroid, respectively.
- *Detection of molecules that have prognostic or therapeutic significance.* Immunohistochemical detection of hormone (estrogen/progesterone) receptors in breast cancer cells is of prognostic and therapeutic value because these cancers are susceptible to antiestrogen therapy (Chapter 23). In general, receptor-positive breast cancers have a better prognosis than receptor-negative tumors. Protein products of oncogenes such as *ERBB2* in breast cancers can also be detected by immunostaining. Breast cancers with strong immunohistochemical staining for the protein product of the *ERBB2* gene product, HER2, generally have a poor prognosis but are amenable to treatment with antibodies that block the activity of the HER2 receptor. Because high-level expression of HER2 is caused by amplification of the *ERBB2* gene, fluorescent in situ hybridization (FISH) to confirm *ERBB2* gene amplification is sometimes used as an adjunct to immunohistochemical studies. Similarly, immunohistochemical stains for ALK protein can be used to identify lung cancers and lymphomas expressing constitutively active ALK fusion proteins, which is an indication for treatment with drugs that inhibit the ALK tyrosine kinase.

Flow Cytometry. Flow cytometry rapidly and quantitatively measures several cell characteristics but requires viable cells

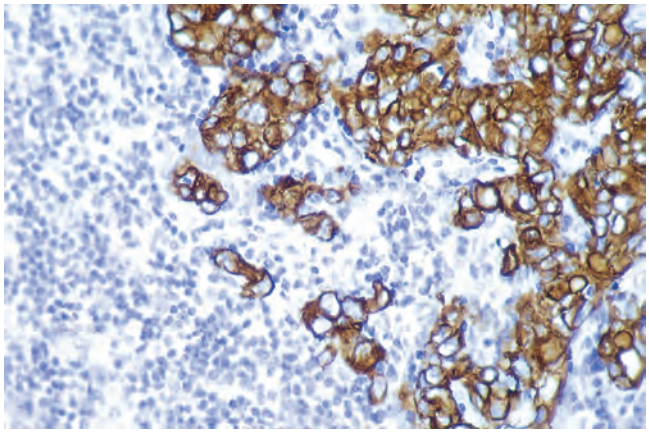


Figure 7.45 Anticytokeratin immunoperoxidase stain of a tumor of epithelial origin (carcinoma). (Courtesy Dr. Melissa Upton, University of Washington, Seattle, Wash.)

in suspension. It is mainly used to identify cellular antigens expressed by "liquid" tumors, those that arise from blood-forming tissues. These include B-cell and T-cell lymphomas and leukemias as well as myeloid neoplasms. An advantage of flow cytometry over immunohistochemistry is that multiple antigens are assessed simultaneously on individual cells using combinations of specific antibodies linked to different fluorescent dyes. Monoclonal antibodies directed against antigens found on blood cells and their progenitors that are frequently detected by flow cytometry are listed in Chapter 13.

Circulating Tumor Cells. Instrumentation that permits detection, quantification, and characterization of rare solid tumor cells (e.g., carcinoma, melanoma) circulating in the blood is being developed as a diagnostic modality. Some of the devices rely on three-dimensional flow cells coated with antibodies specific for tumor cells of interest (e.g., carcinoma cells) that efficiently capture rare tumor cells present in the blood. Such methods have the potential to permit earlier diagnosis, to gauge the risk of metastasis, and to provide a minimally invasive means of assessing the response of tumor cells to therapy, but are mainly being used currently in the realm of clinical research.

Molecular Diagnostics and Cytogenetics. Several molecular or cytogenetic techniques—some established, others emerging—have been used for diagnosis and, in some cases, for predicting behavior of tumors.

- *Diagnosis of malignant neoplasms.* Although molecular methods are not the primary modality of cancer diagnosis, they are of considerable value in selected cases. T- and B-cell tumors are derived from single cells with unique antigen receptor gene rearrangements, whereas reactive lymphoid proliferations contain many different lymphocyte clones, each with a different set of rearrangements antigen receptor genes. For this reason, polymerase chain reaction (PCR)–based evaluation of rearranged T-cell receptor or immunoglobulin genes allows distinction between monoclonal (neoplastic) and polyclonal (reactive) proliferations. Many hematopoietic neoplasms (leukemias and lymphomas) are associated with specific translocations that activate oncogenes. Detection of such translocations, usually by routine cytogenetic analysis or by FISH (Chapter 5), is often extremely helpful in diagnosis. Diagnosis of sarcomas (Chapter 26) with characteristic translocations is also aided by molecular techniques, in part because chromosome preparations are often difficult to obtain from solid tumors. For example, many sarcomas of childhood, so-called round blue cell tumors (Chapter 10), can be difficult to distinguish from each other on the basis of morphology. However, the presence of the characteristic (11;22)(q24;q12) translocation, established by PCR, in such tumors can help to confirm the diagnosis of Ewing sarcoma. Another diagnostic platform with some utility is *DNA microarrays,* either tiling arrays, which cover the entire human genome, or SNP arrays (SNP chips), which allow high-resolution mapping of copy number changes (either deletions or amplifications) genome-wide.
- *Prognosis of malignant neoplasms.* Certain genetic alterations are associated with poor prognosis, and hence their detection allows stratification of patients for therapy. For example, amplification of the *NMYC* gene and deletions of chromosome 1p bode poorly for patients with neuroblastoma, and oligodendrogliomas in which the only genomic abnormality is the loss of chromosomes 1p and 19q respond well to therapy and are associated with long-term survival when compared to tumors without 1p and 19q deletion and with EGF receptor amplification instead.
- *Detection of minimal residual disease.* After treatment of patients with leukemia or lymphoma, the presence of minimal disease or the onset of relapse can be monitored by PCR-based amplification of nucleic acid sequences unique to the malignant clone. For example, detection of *BCR-ABL* transcripts by PCR gives a measure of the residual leukemia cells in treated patients with CML. The prognostic importance of minimal residual disease has been established in acute leukemia and is being evaluated in other neoplasms.
- *Diagnosis of hereditary predisposition to cancer.* As discussed earlier, germline mutations in several tumor suppressor genes, including *BRCA1* and *BRCA2,* as well as the *RET* proto-oncogene, are associated with a high risk of developing specific cancers. Thus, detection of these mutated alleles may allow the patient and physician to devise an aggressive screening program, consider the option of prophylactic surgery, and counsel relatives, who may also be at risk. Such analysis usually requires detection of a specific mutation or sequencing of the entire gene. The latter is necessary when several different cancer-associated mutations are known to exist. Although the detection of mutations in such cases is relatively straightforward, the ethical issues surrounding presymptomatic diagnosis are complex.
- *Guiding therapy with oncoprotein-directed drugs.* An increasing number of chemotherapeutic agents target oncoproteins that are present only in a subset of cancers of a particular type. Thus the molecular identification of genetic lesions that produce these oncoproteins is essential for optimal treatment of patients. Current examples of genetic lesions that guide therapy and are frequently tested for in molecular diagnostic laboratories include the *PML-RARA* fusion gene in acute promyelocytic leukemia, the *BCR-ABL* fusion gene in chronic myeloid leukemia and acute lymphoblastic leukemia, *ERBB1* (EGFR) mutations and *ALK* gene rearrangements in lung cancer, and *BRAF* mutations in melanoma.
- *Identifying mechanisms of drug resistance: liquid biopsies.* These tests rely on circulating tumor cells or, increasingly, cell-free DNA that is shed from dying tumor cells into the blood. The most advanced tests of this type analyze cell-free tumor DNA for the presence of specific driver mutations that create targetable oncoproteins, for example, in lung cancers. These tests use extremely sensitive PCR-based approaches and can be performed on peripheral blood at multiple timepoints after treatment has begun, avoiding the need for repeat tissue biopsies. It is hoped that such approaches will make it possible to detect the early emergence of new genetic variants that convey drug resistance, thereby allowing oncologists to select alternative, more effective drugs before relapse becomes evident clinically.

Molecular Profiles of Tumors: The Future of Cancer Diagnostics

Until recently, molecular studies of tumors involved the analysis of individual genes. However, the past few years have seen the introduction of revolutionary technologies that can rapidly sequence an entire genome; assess epigenetic modifications genome-wide (the epigenome); quantify all of the RNAs expressed in a cell population (the transcriptome); measure many proteins simultaneously (the proteome); and take a snapshot of all of the cell's metabolites (the metabolome). Thus, we are living in the age of "omics"!

Among these various "omic" technologies, DNA sequencing has had the greatest impact on the evaluation of cancer. RNA sequencing is also being widely employed in research laboratories, but RNA is prone to degradation and is a more difficult analyte to work with than DNA in clinical practice. By contrast, DNA sequencing can be routinely performed using DNA retrieved from routinely fixed and processed tissues, the type that exists in most pathology departments.

These advances have enabled the systematic sequencing and cataloging of genomic alterations in various human cancers, much of it within a large consortium sponsored by the National Cancer Institute called The Cancer Genome Atlas (TCGA). The complexity of the genetic aberrations identified in these genome-wide studies has inspired informaticians to create new ways of displaying the data, such as circos plots (Fig. 7.46), which provide a snapshot of all of the genetic alterations that exist in a particular tumor.

The main impact of cancer genome sequencing to date has been in the area of research. However, next-generation DNA sequencing of cancer genomes is rapidly transitioning to standard clinical practice. "Targeted" sequencing is being performed routinely in many academic cancer centers and reference laboratories, at reasonable costs and with turnaround times of a few days to a few weeks. These tests typically cover the exons of several hundred key genes at sufficient "depth" (fold coverage of the sequence in question) to reliably detect any mutation that might be present in as few as 5% of tumor cells. One goal of these analyses is to identify therapeutically "actionable" genetic lesions. Such approaches are particularly applicable to tumors such as lung carcinomas, which are genetically diverse and require a "personalized" approach if targeted therapy is to succeed. Other clinically useful information is obtained, however. For example, the observed pattern of mutations (including passenger as well as driver mutations) can be used to identify the mutational signature (microsatellite instability) that is associated with mismatch repair defects, which as mentioned earlier is an indication for treatment with immune checkpoint inhibitors, regardless of cancer type. Other identified mutations may help guide therapy; for example, *TP53* mutations portend a very poor prognosis in acute leukemias, which otherwise might be expected to respond well to chemotherapy.

Fewer tests measuring RNA levels have entered into the clinical domain, but breast cancers are routinely subjected to RNA profiling to determine whether chemotherapy is likely to be beneficial. Newer test platforms that allow easier and more reliable measurement of RNAs retrieved from fixed tissue samples may expedite the development of additional tests of this type. Certain assays aimed at detecting driver mutations that use RNA have advantages over DNA-based tests, including assays designed to identify chimeric RNAs produced from fusion genes, which are now in various stages of development and clinical application.

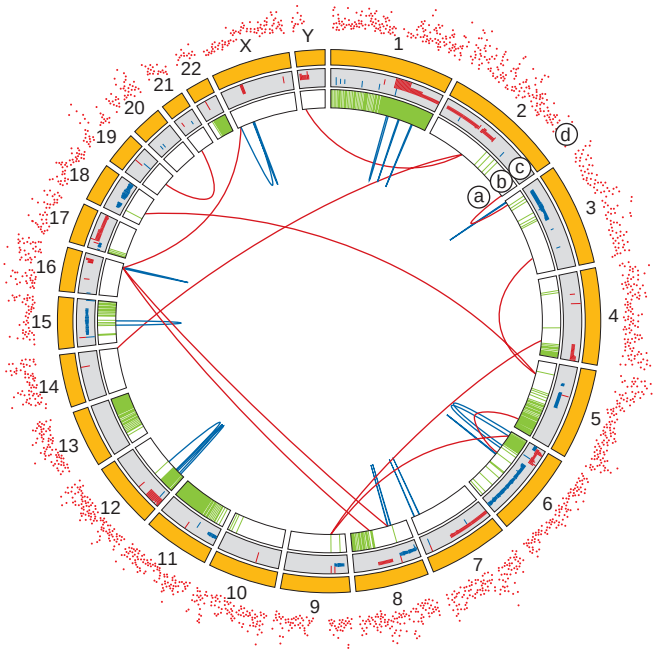


Figure 7.46 Circos plot showing genetic alterations in a single lung cancer in a male patient. Each of the 24 chromosomes in the cancer is displayed in a circle. The positions of various tumor-specific aberrations are mapped onto these chromosomes as follows: *a*, Structural rearrangements in chromosomes. The blue lines denote intrachromosomal rearrangements, while the red lines denote interchromosomal rearrangements. *b*, Regions of loss of heterozygosity and allelic imbalance (overrepresentation of one allele versus the second) are in green. *c*, Copy number profiles, showing copy number losses (in red) and copy gains (in blue). *d*, Point mutations, represented as red dots.

The excitement created by the development of new techniques for the global molecular analysis of tumors has led some scientists to predict that the end of histopathology is in sight. Indeed, with the advent of targeted therapies, it can be argued that we are in the midst of a paradigm shift in which the most important part of the work-up of a cancer sample is the identification of molecular targets, rather than histopathologic diagnosis (Fig. 7.47). For example, it is now appreciated that histopathologically distinct cancers may harbor the same gain-of-function mutation in the serine/threonine kinase BRAF, a component of the RAS signaling pathway (Fig. 7.48). In principle, all of these diverse "BRAFomas" are candidates for treatment with BRAF inhibitors. However, clinical studies have shown that the effectiveness of BRAF inhibitors (for reasons that remain to be determined) vary widely depending on histologic subtype: hairy cell leukemias with *BRAF* mutations typically show sustained responses, melanomas respond transiently, and colon carcinomas respond little, if at all, emphasizing the value of morphologic diagnosis. An example of a treatment based entirely on molecular features is the use of checkpoint inhibitors in patients with recurrent and metastatic tumors based on defects in mismatch repair genes and not on histologic features, which we mentioned earlier.

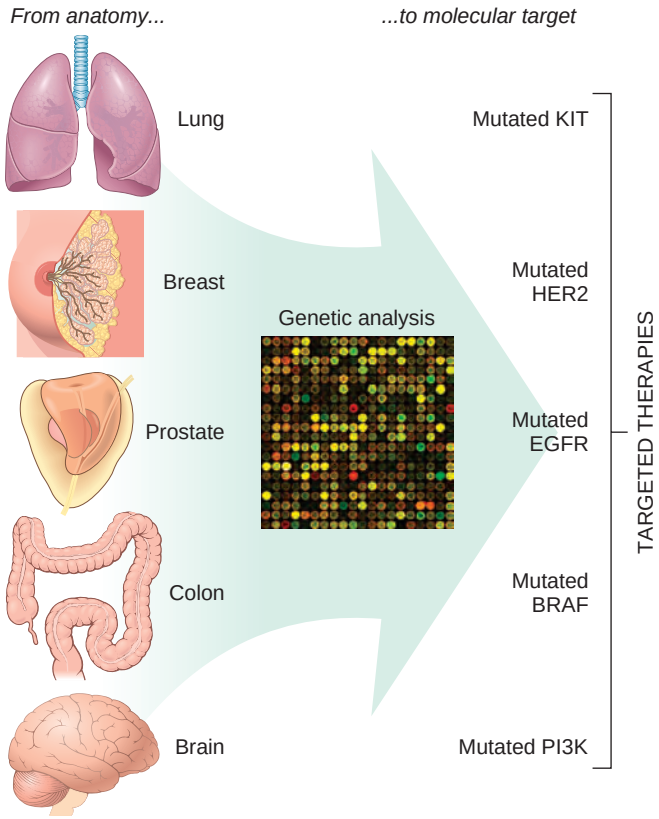


Figure 7.47 A paradigm shift: classification of cancer according to therapeutic targets rather than cell of origin and morphology. (Courtesy Dr. Levi Garraway, Dana Farber Cancer Institute, Boston, Mass.)

In most cases, however, histopathologic inspection of tumors provides information about other important cancer characteristics, such as anaplasia, invasiveness, and tumor heterogeneity, and when coupled with in situ biomarker tests performed on tissue sections, also remains the best way to assess tumor–stromal cell interactions such as angiogenesis and host immune responses. The latter may have an increasingly important role in guiding therapeutic interventions that are designed to counteract immune evasion by tumors. Thus, what lies ahead is not the replacement of one set of techniques by another. On the contrary, for the foreseeable future the most accurate diagnosis and assessment of prognosis in cancer patients will be arrived at by a combination of morphologic and molecular techniques.

Tumor Markers

Biochemical assays for tumor-associated enzymes, hormones, and other tumor markers in the blood lack the sensitivity and specificity necessary to diagnose cancer; however, in concert with other tests, they may contribute to the detection of cancer, and in many instances they are useful in following a tumor's response to therapy and in detecting tumor recurrence.

A host of tumor markers have been described, and new candidates are identified every year. Only a few have stood the test of time and proved to have clinical usefulness.

The application of several markers, listed in Table 7.12, is considered in the discussion of specific forms of neoplasia in other chapters, so only a few widely used examples suffice here. Blood tests for *prostate-specific antigen (PSA),* a marker for prostatic adenocarcinoma, are frequently used in clinical practice. Prostatic carcinoma can be suspected when elevated levels of PSA are found in the blood. However, PSA screening highlights problems encountered with virtually every tumor marker. Although PSA levels are often elevated by cancer, PSA levels also may be elevated by benign prostatic hyperplasia (Chapter 18). Furthermore, there is no PSA level that ensures that a person does not have prostate cancer. These limitations are discussed in detail in Chapter 18.

Other tumor markers occasionally used in clinical practice include *carcinoembryonic antigen* (CEA), which is elaborated by carcinomas of the colon, pancreas, stomach, and breast; and *alpha-fetoprotein* (AFP), which is produced

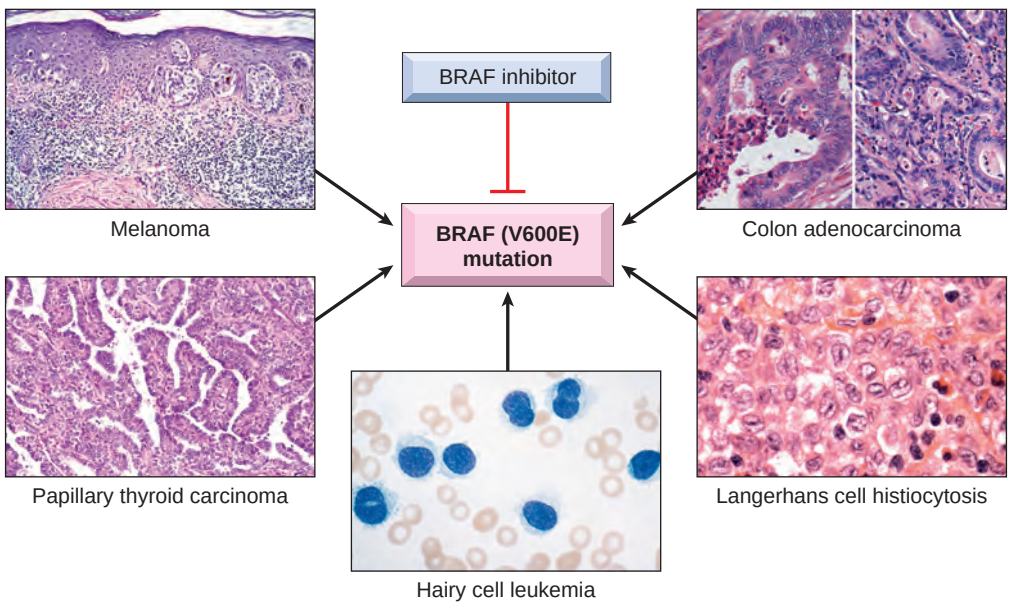


Figure 7.48 Diverse tumor types with a common molecular pathogenesis.

Table 7.12 Selected Tumor Markers

Tumor Markers	Tumor Types
Hormones	
Human chorionic gonadotropin	Trophoblastic tumors, nonseminomatous testicular tumors
Calcitonin	Medullary carcinoma of thyroid
Catecholamine and metabolites	Pheochromocytoma and related tumors
Ectopic hormones	See Table 7.11
Oncofetal Antigens	
α-Fetoprotein	Liver cell cancer, nonseminomatous germ cell tumors of testis
Carcinoembryonic antigen	Carcinomas of the colon, pancreas, lung, stomach, and heart
Lineage-Specific Proteins	
Immunoglobulins	Multiple myeloma and other gammopathies
Prostate-specific antigen and prostate-specific membrane antigen	Prostate cancer
Mucins and Other Glycoproteins	
CA-125	Ovarian cancer
CA-19-9	Colon cancer, pancreatic cancer
CA-15-3	Breast cancer
Cell-Free DNA Markers	
EGFR mutants in serum	Lung cancer
TP53, *APC*, *RAS* mutants in stool and serum	Colon cancer
TP53, *RAS* mutants in stool and serum	Pancreatic cancer
TP53, *RAS* mutants in sputum and serum	Lung cancer
TP53 mutants in urine	Bladder cancer

by hepatocellular carcinomas, yolk sac tumors, and occasionally teratocarcinomas and embryonal cell carcinomas. Unfortunately, serum levels of CEA and AFP can be elevated in a variety of nonneoplastic conditions, limiting their value as screening tests. Their utility lies in their ability to follow response to therapy. With successful resection of the tumor, these markers disappear from the serum; their persistence or reappearance almost always signifies tumor lurking within. Other widely used markers include *human chorionic gonadotropin* (HCG) for testicular tumors, *CA-125* for ovarian tumors, and monoclonal *immunoglobulin* in multiple myeloma and other secretory plasma cell tumors.

The development of accurate, sensitive, and specific screening tests to detect cancer markers in blood and body fluids remains an active area of research, as it is evident that early detection of cancer would greatly improve clinical outcomes. Sensitive tests that identify specific driver mutations in DNA isolated from urine, blood, or stool have been developed for this purpose. Nevertheless, it has been difficult to find a test focused on any single marker that has sufficient specificity to allow screening of populations, and it may be (given that cancer is the product of multiple genetic and epigenetic events rather than singular events) that such markers simply do not exist. An alternative approach is to measure many cancer-associated markers simultaneously. In principle, for example, it may be possible to detect multiple driver mutations in a single "liquid biopsy," and there is evidence that RNA shed from tumor cells into the blood is amenable to molecular profiling. These "circulating RNAs" are enriched in exosomes, membrane vesicles that can be found in the blood and other body fluids that are released from the cells of some tumors.

With all the advances in genomic analyses and targeted therapies and immunotherapies, one can confidently predict that we are on the cusp of the golden age of tumor diagnosis and treatment. Those of you who are in medical school now can safely assume that the expectations for rapid advances in cancer diagnosis and therapy will be realized while you are still in practice. Get ready!

KEY CONCEPTS

LABORATORY DIAGNOSIS OF CANCER

- Several sampling approaches exist for the diagnosis of tumors including excision, biopsy, fine-needle aspiration, and cytologic smears.
- Immunohistochemistry and flow cytometry studies help in the diagnosis and classification of tumors because distinct protein expression patterns define different entities.
- Molecular analyses are used to determine diagnosis and prognosis, to detect minimal residual disease, and to diagnose hereditary predisposition to cancer.
- Molecular profiling of tumors by DNA sequencing, RNA profiling, and DNA copy number arrays is useful in molecular stratification of otherwise identical tumors or those of distinct histogenesis that share a mutation or a mutational signature for purposes of targeted treatment and prognostication.
- Proteins released by tumors into the serum, such as PSA, can be used to monitor for recurrence after treatment but are problematic as screening tests because of low sensitivity and specificity.
- Assays of circulating tumor cells and tumor DNA are under development.

SUGGESTED READINGS

Cancer Epidemiology

Plummer M, de Martel C, Vignat J et al: Global burden of cancers attributable to infections in 2012: a synthetic analysis, *Lancet Glob Health* 4:e609–e616, 2016. *[A detailed analysis estimating that approximately 15% of cancers (2.2 million cases per year) is attributable to infectious agents].*

Liang J, Shang Y: Estrogen and cancer, *Annu Rev Physiol* 75:225–240, 2013. *[A summary of epidemiologic evidence connecting hyperestrogenism to cancer and understanding of the oncogenic mechanisms of estrogen signaling].*

Colditz G, Peterson LL: Obesity and cancer: evidence, impact, and future directions, *Clin Chem* 64:154–162, 2018. *[Review of evidence linking risk for various cancers to obesity].*

Alexandrov LB, Kim J, Haradhvala NJ et al: The repertoire of mutational signatures in human cancer, *bioRxiv* 2018. doi.org/10.1101/322859.

[An update on the various specific mutational signatures that have been identified through the sequencing of cancer genomes, work that falls in the realm of molecular epidemiology].

Cancer Evolution and Genomics

McGranahan N, Swanton C: Clonal heterogeneity and tumor evolution: past, present, and the future, *Cell* 168:613–628, 2018. *[A review discussing the relative contributions of stepwise and large scale changes in cancer genomes during clonal evolution].*

Greaves M, Maley CC: Clonal evolution in cancer, *Nature* 481:306–313, 2012. *[A discussion of the iterative processes of clonal expansion, genetic diversification, and clonal selection that promote cancer evolution and ultimately lead to therapeutic failure].*

Bailey MH, Tokheim C, Porta-Pardo E et al: Comprehensive characterization of cancer driver genes and mutations, *Cell* 173:371–385, 2018. *[An updated compendium of driver mutations and cancer genes across a broad range of cancer].*

Hallmarks of Cancer

Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation, *Cell* 144:646–674, 2011. *[An update of a classic paper describing the key features common to all cancers].*

Oncogenes

Cilloni D, Saglio G: Molecular pathways: BCR-ABL, *Clin Cancer Res* 18:930–937, 2012. *[A discussion of the functional consequences and clinical significance of aberrant tyrosine kinase activity in chronic myeloid leukemia mediated by the constitutive enzyme activity of BCR-ABL].*

Stine ZE, Walton ZE, Altman BJ et al: MYC, metabolism, and cancer, *Cancer Discov* 5:1024–1039, 2015. *[A review of the widespread oncogenic role of the transcription factor MYC in cancer].*

Pakkala S, Ramalingam SS: Personalized therapy for lung cancer: striking a moving target, *JCI Insight* 3:e120858, 2018. *[A review summarizing challenges in the effective treatment of lung cancers with "targetable" driver mutations in various kinases].*

Simanshu DK, Nissley DV, McCormick F: RAS proteins and their regulators in human disease, *Cell* 170:17–33, 2017. *[A discussion of the mechanisms, consequences, and unanswered questions pertaining to RAS signaling].*

Tumor Suppressor Genes

Joerger AC, Fersht AR: The p53 pathway: origins, inactivation in cancer, and emerging therapeutic opportunities, *Annu Rev Biochem* 85:375–404, 2016. *[Discussion of how p53 functions, its inactivation in cancer, and how one might target p53 mutated tumors].*

Dyson NJ: *RB1*: a prototype tumour suppressor and an enigma, *Genes Dev* 30:1492–1502, 2016. *[A discussion of what is known and unknown about RB function].*

Chen C-C, Feng W, Lim PX et al: Homology-directed repair and the role of BRCA1, BRCA2, and related proteins in genome integrity and cancer, *Annu Rev Cancer Biol* 2:313–336, 2018. *[A discussion of the roles of the BRCA1 and BRCA2 tumor suppressor proteins in protecting the genome from damage during DNA replication].*

Worby CA, Dixon JE: PTEN, *Annu Rev Biochem* 83:641–669, 2014. *[A detailed discussion of the function of the lipid phosphatase PTEN, an important tumor suppressor].*

Cancer Cell Metabolism

DeBerardnis RJ, Chandel NS: Fundamentals of cancer metabolism, *Sci Adv* 2:e1600200, 2016. *[A review discussing reprogramming of cancer cells to meet their bioenergetics, biosynthetic, and redox demands].*

Autophagy

Santana-Codina N, Mancias JD, Kimmelman AC: The role of autophagy in cancer, *Annu Rev Cancer Biol* 1:19–39, 2017. *[A discussion of the complex role of autophagy in cancer].*

Evasion of Apoptosis

Letai A: Apoptosis and cancer, *Annu Rev Cancer Biol* 1:275094, 2017. *[A review focusing on anti-apoptotic mechanisms in cancer cells and how they can be targeted].*

Cancer Stem Cells

Batlle E, Clevers H: Cancer stem cells revisited, *Nat Med* 23:1124–1134, 2017. *[A critical appraisal of the cancer stem cell model, with an eye toward therapeutic applications].*

Maciejowski J, de Lange T: Telomeres in cancer: tumour suppression and genome instability, *Nat Rev Mol Cell Biol* 18:175–186, 2017. *[A review discussing the double-edged effect of telomere shortening in cancer; growth suppression versus genomic instability due to telomere "crisis"].*

Angiogenesis

De Palma M, Biziato D, Petrova TV: Microenvironmental regulation of tumour angiogenesis, *Nat Rev Cancer* 17:457–474, 2017. *[A discussion of the pro-oncogenic role of the tumor microenvironment in cancer and its implications for development of new therapies].*

Invasion and Metastasis

Riggi N, Aguet M, Stamenkovic I: Cancer metastasis: a reappraisal of its underlying mechanisms and their relevance to treatment, *Annu Rev Pathol* 13:117–140, 2018. *[A review of the cellular and molecular mechanisms that underlie solid tumor metastasis].*

Brabletz T, Kalluri R, Nieto MA et al: EMT in cancer, *Nat Rev Cancer* 18:128–134, 2018. *[A far-ranging discussion of the possible roles of epithelial-mesenchymal transition in metastasis and other cancer hallmarks].*

Evasion of Host Defense

Spranger S, Gajewski TF: Mechanisms of tumor cell-intrinsic immune evasion, *Annu Rev Cancer Biol* 2:213–228, 2017. *[A broad discussion of how cancer cells evade the innate and adaptive immune system and what might be done to counter it].*

Brahmer JR, Tykodi SS, Chow LQ et al: Safety and activity of anti-PD-L1 antibody in patients with advanced cancer, *N Engl J Med* 366:2455–2465, 2012. Also, Topalian SL, Hodi FS, Brahmer JR, et al: Safety, activity, and immune correlates of anti-PD-1 antibody in cancer, *N Engl J Med* 366:2443–2454, 2012. *[Two back-to-back landmark papers describing the anticancer activity of antibodies that interfere with the activity of PD-1, a signaling receptor that inhibits T-cell function].*

Havel JJ, Chowell D, Chan TA: The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy, *Nat Rev Cancer* 19:133–150, 2019. [*A discussion of factors that determine response and resistance to checkpoint inhibitor therapy in patients with advance cancer*].

Cancer Enabling Inflammation

Pathria P, Louis TL, Varner JA: Targeting tumor-associated macrophages in cancer, *Trends Immunol* 40:310–327, 2019. *[A discussion of clinical and experimental studies describing protumorigenic roles for tumor-associated macrophages that elicit cancer-associated inflammation].*

Chromosomal Aberrations

Ly P, Cleveland D: Rebuilding chromosomes after catastrophe: emerging mechanisms of chromothripsis, *Trends Cell Biol* 27:917–930, 2017. *[A review describing mechanisms and consequences of chromothripsis in cancer].*

Gostissa M, Alt FW, Chiarle R: Mechanisms that promote and suppress chromosomal translocations in lymphocytes, *Annu Rev Immunol* 29:319–350, 2011. *[A discussion of pathogenic insights gleaned from experimental models of oncogenic chromosomal translocations in lymphocytes].*

Epigenetics and Cancer

Flavahan WA, Gaskell E, Bernstein BE: Epigenetic plasticity and the hallmarks of cancer, *Science* 357:2017. eaa12380. *[A discussion of how dysregulation of the cancer cell epigenome leads to phenotypic heterogeneity and the emergence of cancer hallmarks].*

Nebbioso A, Tambaro FP, Dell'Aversana C et al: Cancer epigenetics: moving forward, *PLoS Genet* 14:e1007362, 2018. [*A critical evaluation of causal epigenetic aberrations in cancer*].

Noncoding RNAs

Anastasiadou E, Jacob LS, Slack FJ: Non-coding RNA networks in cancer, *Nat Rev Genet* 18:5–18, 2018. *[A review that discusses the roles of noncoding RNAs as "drivers" in cancer].*

Lin C-P, He L: Noncoding RNAs in cancer development, *Annu Rev Cancer Biol* 1:163–184, 2017. *[A discussion of the role of micro-RNAs in regulation of gene networks involved in cancer development].*

Environmental Carcinogens

Berger MF, Hodis E, Heffernan TP et al: Melanoma genome sequencing reveals frequent PREX2 mutations, *Nature* 485:502–506, 2012. *[A study that describes and quantifies the mutational impact of sun exposure on melanoma genomes].*

Pleasance ED, Stephens PJ, O'Meara S et al: A small-cell lung cancer genome with complex signatures of tobacco exposure, *Nature* 463:184–190, 2010. *[A study that describes and quantifies the mutational impact of carcinogens in tobacco smoke on lung cancer genomes].*

Microbial Carcinogenesis

Ringehan M, McKeating JA, Protzer U: Viral hepatitis and liver cancer, *Philos Trans R Soc Lond B Biol Sci* 372:e20160274, 2017. *[A discussion of the roles of hepatitis viruses in hepatocellular carcinoma].*

Magrath I: Epidemiology: Clues to the pathogenesis of Burkitt lymphoma, *Br J Haematol* 156:744–756, 2012. *[A discussion of the roles of Epstein-Barr virus and malaria in genesis of Burkitt lymphoma, an aggressive B-cell malignancy].*

Moody CA, Laimins LA: Human papillomavirus oncoproteins: pathways to transformation, *Nat Rev Cancer* 10:550–560, 2010. *[A review focused on the oncogenic activities of human papillomavirus E6 and E7 proteins].*

Peleteiro B, La Vecchia C, Lunet N: The role of *Helicobacter pylori* infection in the web of gastric cancer causation, *Eur J Cancer Prev* 21:118–125, 2012. *[A review that proposes a conceptual framework for understanding the role of* Helicobacter pylori *infection in gastric cancer].*

Cancer Cachexia and Paraneoplastic Syndromes

Baracos VE, Martin L, Korc M et al: Cancer-associated cachexia, *Nat Rev Dis Primers* 4:e17105, 2018. *[An overview of the clinical features and possible mechanisms underlying cancer-related cachexia].*

Henry K: Paraneoplastic syndromes: definitions, classification, pathophysiology, and principles of treatment, *Semin Diagn Pathol* 36:204–210, 2019. *[A broad discussion of paraneoplastic syndromes].*

Cancer Diagnostics

Heitzer E, Haque IS, Roberts CES et al: Current and future perspectives of liquid biopsies in genomics-driven oncology, *Nat Rev Genet* 20:71–88, 2019. *[Potential and pitfalls of using circulating tumor DNA and circulating tumor cells to diagnose cancer and follow responses to therapy].*

Mardis ER: The impact of next-generation sequencing on cancer genomics: from discovery to clinic, *Cold Spring Harb Perspect Med* 9:a036269, 2019. *[A discussion of the current and potential future uses of NGS in oncology].*

Pao W, Iafrate AJ, Su Z: Genetically informed lung cancer medicine, *J Pathol* 223:230–240, 2011. *[A review of the adoption and application of genomic analysis to the diagnosis and treatment of lung cancer].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 8

Infectious Diseases

Karen M. Frank • Alexander J. McAdam

CHAPTER CONTENTS

GENERAL PRINCIPLES OF MICROBIAL PATHOGENESIS

Despite the availability of effective vaccines and antibiotics, infectious diseases remain a major health problem throughout the world. In the United States and other high-income countries, infectious diseases are particularly important causes of death among older adults and in people who are immunosuppressed or who suffer from debilitating chronic diseases. In lower-income nations inadequate access to medical care and malnutrition contribute to a heavy burden of infectious diseases. In these regions of the world, five of the ten leading causes of death are infectious diseases. Tragically, most of these deaths occur in children, with respiratory infections, infectious diarrhea, and malaria taking the greatest toll.

How Microorganisms Cause Disease

Most infectious diseases are caused by pathogenic organisms that exhibit a wide range of virulence. These organisms are acquired from a variety of sources, including people, animals, insect vectors, and the environment, but they are not found in the normal microbiota of healthy people. Thus, their presence is diagnostic of an infection. Over the past few years, it has become evident that humans and other animals harbor a complex ecosystem of microbes (the *microbiome*) that has important roles in health and disease. Most of these commensal organisms coexist peacefully with their human hosts, occupying microenvironmental niches that might otherwise be filled by potential pathogens, and in doing so help prevent infectious disease. However, under conditions in which normal host defenses are breached or attenuated (described later), even commensal microbiota may cause symptomatic infections and can even be fatal.

We will start our review of infectious disease at the beginning of the process, the establishment of a beachhead in the host, and then discuss dissemination and transmission of infection, before turning to specific infections.

Routes of Entry of Microbes

Microbes can enter the host by breaching epithelial surfaces, inhalation, ingestion, or sexual transmission (Table 8.1). In general, respiratory, gastrointestinal, and genitourinary tract infections in otherwise healthy persons are caused by virulent microorganisms with the ability to damage or penetrate the epidermis or mucosal epithelium. By contrast, skin infections in healthy persons are mainly caused by organisms that enter the skin through superficial injuries.

Skin

The intact keratinized epidermis protects against infection by serving as a mechanical barrier, having a low pH, and by producing antimicrobial fatty acids and defensins, small peptides that are toxic to bacteria. **Most skin infections are initiated by mechanical injury of the epidermis.** The injury may range from minor trauma to large wounds, burns, and pressure-related ulcers. In the hospital setting, infections may stem from intravenous catheters in patients or needle sticks in health care workers. Some pathogens penetrate the skin via an insect *(vector)* or animal bite; vectors include fleas, ticks, mosquitoes, and lice. The larvae of *Schistosoma* can traverse unbroken skin by releasing enzymes that dissolve the adhesive proteins that hold keratinocytes together. Certain fungi *(dermatophytes)* can cause superficial infections of the intact stratum corneum, hair, and nails.

Gastrointestinal Tract

Gastrointestinal tract infections may occur when local defenses are circumvented by a pathogen, or when they are so weakened that even normal flora produce disease. Most gastrointestinal pathogens are transmitted by food or drink contaminated with fecal material; when hygiene fails, diarrheal disease becomes rampant. The gastrointestinal tract has several local defenses. Of these, acidic gastric secretions are particularly important because they are highly effective at killing certain organisms. Neutralizing the stomach acid of healthy volunteers increased the infectivity of *Vibrio cholerae* by 10,000-fold. A layer of mucus covers the gut throughout its length, preventing access of luminal pathogens to the surface epithelium. Pancreatic enzymes and bile detergents can destroy organisms with lipid

Table 8.1 Routes of Microbial Infection

Site	Major Local Defense(s)	Basis for Failure of Local Defense	Pathogens (Examples)
Skin	Epidermal barrier	Mechanical defects (punctures, burns, ulcers)	*Staphylococcus aureus*, *Candida albicans*, *Pseudomonas aeruginosa*
		Needle sticks	Human immunodeficiency virus, hepatitis viruses
		Arthropod and animal bites	Yellow fever, plague, Lyme disease, malaria, rabies
		Direct penetration	*Schistosoma* spp.
Gastrointestinal tract	Epithelial barrier	Attachment and local proliferation of microbes	*Vibrio cholerae*, *Giardia duodenalis*
		Attachment and local invasion of microbes	*Shigella* spp., *Salmonella* spp., *Campylobacter* spp.
		Uptake through M cells	Poliovirus, *Shigella* spp., *Salmonella* spp.
	Acidic secretions	Acid-resistant cysts and eggs	Many protozoa and helminths
	Peristalsis	Obstruction, ileus, postsurgical adhesions	Mixed aerobic and anaerobic bacteria (*Escherichia coli*, *Bacteroides* spp.)
	Bile and pancreatic enzymes	Resistant microbial external coats	Hepatitis A, rotavirus, norovirus
	Normal protective microbiota	Broad-spectrum antibiotic use	*Clostridioides difficile*
Respiratory tract	Mucociliary clearance	Attachment and local proliferation of microbes	Influenza viruses
		Ciliary paralysis by toxins	*Haemophilus influenzae*, *Mycoplasma pneumoniae*, *Bordetella pertussis*
	Resident alveolar macrophages	Resistance to killing by phagocytes	*Mycobacterium tuberculosis*
Urogenital tract	Urination	Obstruction, microbial attachment, and local proliferation	*Escherichia coli*
	Normal vaginal microbiota	Antibiotic use	*Candida albicans*
	Intact epidermal/epithelial barrier	Microbial attachment and local proliferation	*Neisseria gonorrhoeae*
		Direct infection/local invasion	Herpes viruses, syphilis
		Local trauma	Various sexually transmitted infections (e.g., human papillomavirus)

envelopes. Antimicrobial defensins are produced by gut epithelial cells. IgA antibodies, produced in mucosal lymphoid tissues such as Peyer patches and secreted into the gut lumen (Chapter 17), can neutralize potential pathogens. Peristalsis can clear organisms, preventing their local overgrowth. Finally, the normal gut microbiota competitively inhibits colonization and overgrowth by potential pathogens, such as *Clostridioides difficile*. Many common gastrointestinal pathogens are resistant to local defenses. Norovirus (the scourge of the cruise ship industry) is a nonenveloped virus that is resistant to inactivation by acid, bile, and pancreatic enzymes and hence easily spreads in places where people are crowded together. Intestinal protozoa and helminths transmitted as cysts or eggs, respectively, have acid-resistant outer coats.

Pathogens may establish symptomatic gastrointestinal disease through several distinct mechanisms:

- *Toxin production.* Some organisms contaminating food can produce gastrointestinal disease without necessarily establishing an infection in the host. An example is *Staphylococcus aureus,* which elaborates a powerful exotoxin during its growth in contaminated food that is responsible for acute food poisoning.
- *Bacterial colonoization and toxin production.* Other bacteria establish an infection and produce damaging toxins. Examples include *V. cholerae* and enterotoxigenic *Escherichia coli,* which bind to the intestinal epithelium and multiply in the overlying mucous layer. These organisms elaborate potent exotoxins that are responsible for symptomatic disease.
- *Adhesion and mucosal invasion. Shigella spp., Salmonella enterica, Campylobacter jejuni,* and *Entamoeba histolytica* invade the intestinal mucosa and lamina propria and cause ulceration, inflammation, and hemorrhage that manifest clinically as dysentery. *Candida albicans* invades superficially into oral and esophageal squamous mucosa in immunocompromised patients to cause thrush.

Respiratory Tract

A plethora of microorganisms, including viruses, bacteria, and fungi, are inhaled daily, mainly in dust or aerosol particles. Microorganisms in large particles are trapped in the mucociliary blanket that lines the nose and the upper respiratory tract and transported by ciliary action to the back of the throat, where they are swallowed and cleared. Particles smaller than 5 microns are carried into the alveoli, where they are phagocytosed by leukocytes.

The microorganisms that infect the healthy respiratory tract evade local defenses through several different mechanisms. Some respiratory viruses attach to and enter epithelial cells in the lower respiratory tract and pharynx. For example, influenza viruses have envelope proteins called *hemagglutinins* that bind to sialic acid on the surface of epithelial cells. Attachment induces the host cell to endocytose the virus, leading to viral entry and replication. Certain bacterial respiratory pathogens, including *Mycoplasma pneumoniae* and *Bordetella pertussis,* release toxins that enhance their ability to establish an infection by impairing ciliary activity. Another important mechanism of establishing respiratory infection is resistance to killing following phagocytosis. *Mycobacterium tuberculosis* gains a foothold in alveoli by surviving within the phagolysosomes of macrophages.

Other organisms establish disease when local or systemic defenses are impaired. The damage to respiratory mucociliary clearance by influenza, mechanical ventilation, smoking or cystic fibrosis sets the stage for superinfection by bacteria. Many other infectious agents cause respiratory infections primarily in the setting of systemic immunodeficiency. Examples include fungal infections by *Pneumocystis jirovecii* in acquired immunodeficiency syndrome (AIDS) patients and by *Aspergillus* spp. in patients with neutropenia.

Urogenital Tract

Urine contains small numbers of low-virulence bacteria. The urinary tract is protected from infection by regular emptying during micturition. Urinary tract pathogens (e.g., *E. coli*) almost always gain access via the urethra and must be able to adhere to urothelium to avoid being washed away. Women have more than 10 times as many urinary tract infections as men because the length of the urethra is 5 cm in women versus 20 cm in men, making women more susceptible to entry of bacteria from the rectum. Expulsion of urine from the bladder eliminates microbes. Predictably, **obstruction of urinary flow or reflux of urine is a major factor in susceptibility to urinary tract infections.**

From puberty until menopause the vagina is protected from pathogens by lactobacilli, which ferment glucose to lactic acid, producing a low pH environment that suppresses the growth of pathogens. Antibiotics can kill the lactobacilli and allow overgrowth of yeast, causing vaginal candidiasis.

Vertical Transmission

Vertical transmission of infectious agents from mother to fetus or newborn child is a common mode of transmission of certain pathogens and may occur through several different routes.

- *Placental-fetal transmission* is most likely to occur when the mother becomes infected with a pathogen during pregnancy. Some resulting infections interfere with fetal development, and the degree and type of damage depend on the age of the fetus at the time of infection. Rubella virus infection during the first trimester can lead to heart malformations, intellectual disability, cataracts, or deafness, but rubella virus infection during the third trimester has little effect.
- *Transmission during birth* is caused by contact with infectious agents during passage through the birth canal. Examples include gonococcal and chlamydial conjunctivitis.
- *Postnatal transmission in maternal milk* can transmit cytomegalovirus (CMV), human immunodeficiency virus (HIV), and hepatitis B virus (HBV).

Spread and Dissemination of Microbes Within the Body

Although some disease-causing microorganisms remain localized to the initial site of infection, others have the capacity to invade tissues and spread to distant sites via the lymphatics, the blood, or the nerves (Fig. 8.1). Pathogens can spread within the body in several ways. Some pathogens secrete enzymes that break down tissues, allowing the organisms to spread contiguously in tissue. Organisms that disseminate often travel through the lymphatics to regional

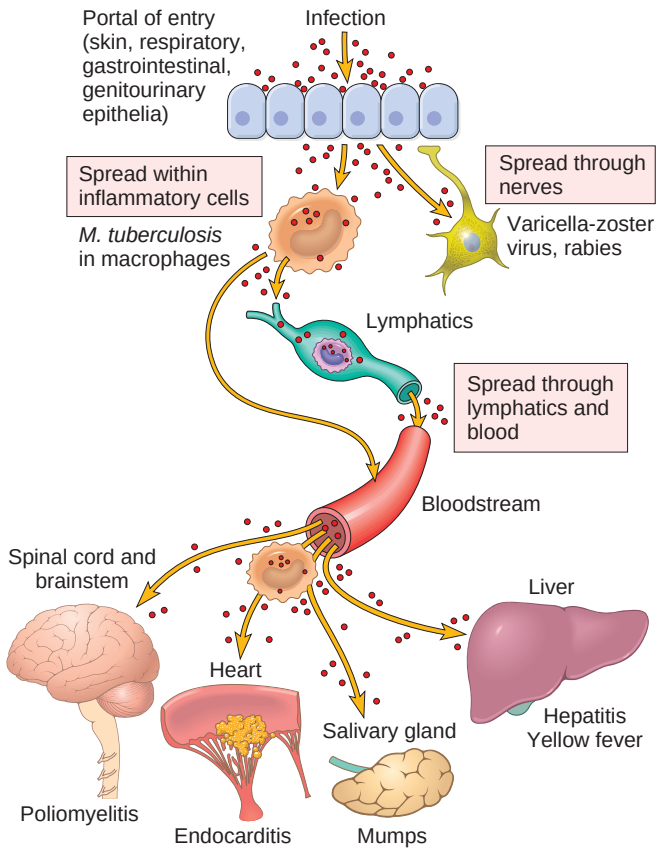


Figure 8.1 Routes of entry and dissemination of microbes. To enter the body, microbes penetrate the epithelial or mucosal barriers. Infection may remain localized at the site of entry or spread to other sites in the body. Most common microbes *(selected examples are shown)* spread through the lymphatics or bloodstream (either freely or within inflammatory cells). However, certain viruses and bacterial toxins may also travel through nerves. (Modified from Mims CA: *The Pathogenesis of Infectious Disease*, ed 4, San Diego, 1996, Academic Press.)

lymph nodes, from which they may reach the bloodstream. Certain viruses, such as rabies virus, poliovirus, and varicella-zoster virus, spread to the central nervous system (CNS) by infecting peripheral nerves and then traveling along axons. However, the most common and efficient mode of microbial dissemination is through the bloodstream, by which the organism can reach all organs. The consequences of blood-borne spread of pathogens vary widely depending on the virulence of the organism, the magnitude of the infection, the pattern of seeding, and host factors such as immune status. As discussed in Chapter 4, disseminated pathogens may produce a systemic inflammatory response syndrome called *sepsis* that manifests as fever, low blood pressure, and coagulopathies that may progress to organ failure and death if unchecked, even in previously healthy individuals. In other instances, the major signs of spread of the infection are related to seeding of tissues by pathogens. These may take the form of a single infectious nidus (an abscess or tuberculoma), multiple small sites of infection (e.g., miliary tuberculosis or *Candida* microabscesses), or infection of the heart and vessels (infectious endocarditis and mycotic aneurysm).

Release From the Body and Transmission of Microbes

Microbes use a variety of exit strategies to ensure their transmission from one host to the next. Release may be accomplished by skin shedding, coughing, sneezing, voiding of urine or feces, during sexual contact, or through insect vectors. Some pathogens are released for only brief periods of time or periodically during disease flares, but others may be shed for long periods by asymptomatic carrier hosts. Pathogens vary in hardiness in the environment. Fragile pathogens persist outside of the body for only short periods of time and must be passed quickly from person to person, often by direct contact.

Most pathogens are transmitted from person to person by respiratory, fecal-oral, or sexual routes.

- Viruses and bacteria spread by the respiratory route are aerosolized in droplets and released into the air. Some respiratory pathogens, including influenza viruses, are spread in large droplets that travel no more than 3 feet from the source, but others, including *M. tuberculosis* and varicella-zoster virus, spread in small particles that can travel longer distances.
- Most enteric pathogens are spread by the fecal-oral route, that is, by ingestion of stool-contaminated water or food. Water-borne pathogens involved in epidemic outbreaks that are spread in this fashion include hepatitis A and E viruses (HAV and HEV), poliovirus, rotavirus, *V. cholera, Shigella* spp., *C. jejuni,* and *S. enterica*. Some parasitic helminths (e.g., hookworms, schistosomes) shed eggs in stool that hatch as larvae that are capable of penetrating the skin of the next host.
- Sexual transmission often entails prolonged intimate or mucosal contact and is responsible for spread of a wide variety of pathogens, including viruses (e.g., herpes simplex virus [HSV], HIV, human papillomavirus [HPV]); bacteria (*Treponema pallidum, Neisseria gonorrhoeae*); protozoa *(Trichomonas vaginalis);* and arthropods (*Phthiris pubis,* lice).

There are several additional routes of transmission. Saliva is responsible for transmitting viruses that replicate in the salivary glands or the oropharynx, including Epstein-Barr virus (EBV). Some important human pathogens are protozoa that are spread through blood meals taken by arthropod vectors (mosquitoes, ticks, mites). Finally, *zoonotic infections* are those transmitted from animals to humans, either by direct contact (including animal bites), use of animal products, or via an invertebrate vector.

KEY CONCEPTS

HOW MICROORGANISMS CAUSE DISEASE

- Transmission of infections can occur by contact (direct and indirect), the respiratory route, the fecal-oral route, sexual transmission, vertical transmission, or insect/arthropod vectors.
- A pathogen can establish infection if it possesses virulence factors that overcome normal host defenses or if the host defenses are compromised.
- Host defenses against infection include the following:
  - Skin: tough keratinized barrier, low pH, fatty acids
  - Respiratory system: alveolar macrophages, mucociliary clearance by bronchial epithelium, IgA

- GI system: acidic gastric pH, viscous mucus, pancreatic enzymes and bile, defensins, IgA, and normal microbiota
- Urogenital tract: repeated flushing and acidic environment created by commensal microbiota

• Pathogens can proliferate locally, at the site of initial infection, or spread to other sites by direct extension (invasion) or by transport in the lymphatics, the blood, or nerves.

Host-Pathogen Interactions

The outcome of infection is determined by the virulence of the microbe and the nature of the host immune response, which may eliminate the infection or, in some cases, exacerbate or cause tissue damage. The host has a large and complex armamentarium of defenses against pathogens, including physical barriers and components of the innate and adaptive immune systems, which were discussed in Chapter 6. Microbes undergo continuous evolution to combat host defenses. Here we discuss some specific features of microbes and the host response that are important determinants of the outcome of infections.

Immune Evasion by Microbes

Most pathogenic microbes have developed one or more strategies to evade host defenses (Fig. 8.2).

- *Antigenic variation.* Antibodies against microbial antigens block microbial adhesion and uptake into cells, act as opsonins to facilitate phagocytosis, and fix complement. To escape recognition, microbes have strategies that allow them to "change their coats" by expressing different surface antigens (*Borrelia* spp. and trypanosomes). Influenza viruses have a segmented RNA genome that allows for frequent recombination events, permitting antigenic drift (changes in antibody-binding sites of major viral envelope glycoproteins) and antigenic shift (reassortment of genomes between two viral strains creating a new strain). Other microbes generate numerous genetic variants through mutation (over 95 capsular polysaccharides in different strains of *Streptococcus pneumoniae*) (Table 8.2).
- *Resistance to antimicrobial peptides.* Pathogens, such as *Shigella* spp., *S. aureus,* and *Candida* spp., use strategies to avoid killing by cationic antimicrobial peptides that include changes in net surface charge and membrane hydrophobicity, which prevent antimicrobial peptide insertion and pore formation, secretion of proteins that inactivate or degrade the peptides, and pumps that export the peptides.
- *Resistance to killing by phagocytes.* The carbohydrate capsule on the surface of many bacteria *(S. pneumoniae, Neisseria meningitidis, Haemophilus influenzae)* prevents phagocytosis of the organisms by neutrophils. *S. aureus* expresses protein A, which binds the Fc portion of antibodies and inhibits phagocytosis by competitively reducing binding of the antibodies to phagocyte Fc receptors. Some pathogens are resistant to intracellular killing in phagocytes, including mycobacteria (which inhibit phagosome-lysosome fusion), *Listeria monocytogenes* (which disrupts the phagosome membrane and escapes onto the cytosol), *Cryptococcus neoformans,* and certain protozoa (e.g., *Leishmania* spp., *Trypanosoma* spp., *Toxoplasma gondii*).
- *Evasion of apoptosis and manipulation of host cell metabolism.* Some viruses produce proteins that interfere with apoptosis of the host cell, buying time to replicate, enter latency, or even transform infected cells. Microbes that replicate intracellularly (viruses, some bacteria, fungi, and protozoa) also express factors that modulate autophagy, thus evading degradation.
- *Resistance to cytokine-, chemokine- and complement-mediated host defense.* Some viruses interfere with interferon (IFN) function by producing soluble homologues of IFN-α/β or IFN-γ receptors that function as "decoys" that inhibit the actions of secreted IFNs; by producing proteins that

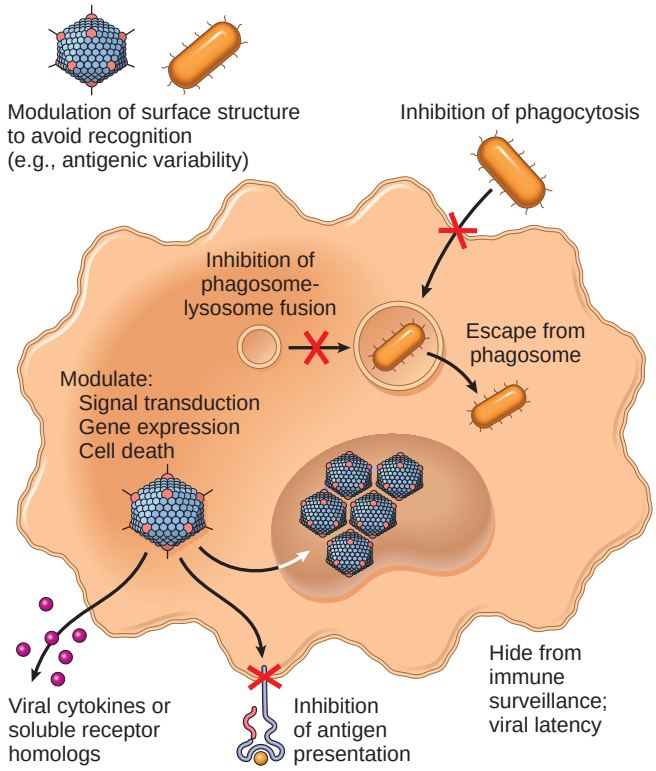


Figure 8.2 An overview of mechanisms used by viral and bacterial pathogens to evade innate and adaptive immunity. (Modified with permission from Finlay B, McFadden G: Anti-immunology: evasion of the host immune system by bacterial and viral pathogens, *Cell* 2006;124:767–782.)

Table 8.2 Mechanisms of Antigenic Variation

Type	Example	Disease
High mutation rate	Human immunodeficiency virus	AIDS
	Influenza virus	Influenza
Genetic reassortment	Influenza virus	Influenza
	Rotavirus	Diarrhea
Genetic rearrangement (e.g., gene recombination, gene conversion, site-specific inversion)	*Borrelia burgdorferi*	Lyme disease
	Neisseria gonorrhoeae	Gonorrhea
	Trypanosoma brucei	African sleeping sickness
	Plasmodium falciparum	Malaria
Large diversity of serotypes	Rhinoviruses	Colds
	Streptococcus pneumoniae	Pneumonia and meningitis

inhibit the JAK/STAT cytokine receptor signaling pathway; or by producing proteins that inactivate or inhibit double-stranded RNA–dependent protein kinase (protein kinase R [PKR]), through which IFNs inhibit viral replication.

- *Evasion of recognition by CD8+ cytotoxic T lymphocytes (CTLs) and CD4+ helper T cells.* T cells recognize microbial antigens presented by MHC molecules, class I for CTLs and class II for CD4+ cells (see Chapter 6). Several DNA viruses (e.g., HSV, CMV, EBV) bind to or alter localization of MHC class I proteins, impairing peptide presentation to CD8+ T cells. Herpesviruses also can target MHC class II molecules for degradation, impairing antigen presentation to CD4+ T-helper cells.
- Another strategy exploits *immunoregulatory mechanisms to downregulate antimicrobial T cell responses*. Loss of T cell potency over time, termed *T-cell exhaustion,* is a feature of chronic infections by HIV, hepatitis C virus (HCV), and HBV. Programmed cell death protein 1 (PD-1) is an immune checkpoint cell surface receptor, and the PD-1 pathway, which normally functions to maintain T-cell tolerance to self antigens, is an important mediator of T-cell exhaustion during chronic viral infection. Tumors exploit the same mechanisms to suppress destructive immune responses (Chapter 7). Anti–PD-1 immunotherapy is approved for the treatment of cancers and is being explored as a possible adjunctive therapy for chronic infections that are resistant to antimicrobial therapy.
- Another way of avoiding the immune system is to "lie low" by establishing a state of *latent infection* in which few, if any, viral genes are expressed, until later reactivation. Examples include latent infections of neurons by HSV and varicella-zoster virus, and of B lymphocytes by EBV.
- Pathogens can infect immune cells and interfere with their function (e.g., HIV, which infects and destroys CD4+ T cells).

KEY CONCEPTS

IMMUNE EVASION BY MICROBES

After bypassing host tissue barriers, infectious microorganisms must also evade host innate and adaptive immunity to successfully proliferate and be transmitted to the next host. Strategies include the following:

- Antigenic variation
- Inactivating antibodies or complement
- Resisting phagocytosis (e.g., by producing a capsule)
- Suppressing the host adaptive immune response (e.g., by interfering with cytokines or inhibiting MHC expression and antigen presentation)
- Establishing latency, during which viruses survive in a silent state in infected cells
- Infecting and disabling or killing immune cells

Injurious Effects of Host Immunity

The host immune response to microbes can be a major cause of tissue injury. The granulomatous inflammatory reaction to *M. tuberculosis* sequesters the bacilli and prevents their spread, but it also can produce tissue damage and fibrosis. Similarly, damage to hepatocytes following HBV and HCV infection is mainly due to the effects of the immune response on infected liver cells: in an attempt to clear the virus, host T cells and, possibly, natural killer (NK) cells kill the hepatocytes. Antibodies produced against the streptococcal M protein can cross-react with cardiac proteins and damage the heart, leading to rheumatic heart disease. Poststreptococcal glomerulonephritis is caused by immune complexes formed between antistreptococcal antibodies and streptococcal antigens that deposit in the renal glomeruli, producing inflammation. A cycle of inflammation and epithelial injury contributes to the pathogenesis of inflammatory bowel disease, with microbes playing a role (Chapter 17). Viruses (HBV, HCV) and bacteria *(Helicobacter pylori)* that are not known to carry or activate oncogenes are associated with cancers, presumably because these microbes trigger chronic inflammation, which provides fertile ground for the development of cancer (Chapter 7).

Infections in People With Immunodeficiencies

Inherited or acquired defects in innate and adaptive immunity often impair the immune system, rendering the affected individual susceptible to infections (Chapter 6). Organisms that cause disease in immunodeficient individuals but not in people with intact immune systems are called *opportunistic*. Worldwide, the most devastating immunodeficiency is that caused by infection with HIV, the cause of AIDS. Other causes of acquired immunodeficiencies include infiltrative processes that suppress bone marrow function (such as leukemia), immunosuppressive drugs used to treat patients with autoimmune diseases and organ transplant recipients, as well as drugs used to treat cancer, and hematopoietic stem cell transplantation. Opportunistic organisms (e.g., *Aspergillus* spp. and *Pseudomonas* spp.) cause significant disease in these patients. Decline of immune responses can result in reactivation of latent infection (e.g., herpesviruses and *M. tuberculosis*). Age-related decline in immune function may increase infections in the elderly. Nonimmune diseases or injuries also increase susceptibility to infection: *Pseudomonas aeruginosa* and *Burkholderia cepacia* in cystic fibrosis due to a defective transmembrane conductance regulator, *S. pneumoniae* in people with sickle cell disease due to loss of splenic macrophages, and *P. aeruginosa* in burns due to barrier disruption. Finally, malnutrition can impair immune defenses.

In addition to common immunodeficiencies, rare inherited (primary) immunodeficiency diseases illuminate important aspects of specific components of host defense, as well as the unique vulnerabilities of certain pathogens.

- *Antibody deficiencies,* as seen in patients with X-linked agammaglobulinemia, lead to increased susceptibility to infections by extracellular bacteria, including *S. pneumoniae, H. influenzae,* and *S. aureus,* as well as a few viruses (rotavirus and enteroviruses).
- *Complement defects* involving the early components of the complement cascade lead to susceptibility to infections by encapsulated bacteria, such as *S. pneumoniae,* whereas deficiencies of the late membrane attack complex components (C5 to C9) are associated with infections due to *Neisseria* spp.
- *Defects in neutrophil function,* as in chronic granulomatous disease, lead to increased susceptibility to infections with *S. aureus,* some gram-negative bacteria, and fungi.

- *Defects in Toll-like receptor (TLR) signaling pathways* have varied effects. Mutations in MyD88 or IRAK4, which are signaling proteins downstream of several TLRs, predispose to pyogenic bacterial infections *(S. pneumoniae)*, and impaired TLR3 responses are associated with childhood HSV encephalitis.
- *T-cell defects* lead to susceptibility to intracellular pathogens, particularly viruses and some parasites. Inherited mutations that impair the generation of T-helper 1 (Th1) cells (such as mutations in IL-12 or IFN-γ receptors, or the transcription factor STAT1) are associated with atypical mycobacterial infections, discussed later. By contrast, defects that impair the generation of Th17 cells (mutations in STAT3) are associated with chronic mucocutaneous candidiasis.

Host Damage by Microbes

Infectious agents establish infection and damage tissues by a few mechanisms:

- They can contact or enter host cells and cause *cell death* directly, or cause changes in cellular metabolism and proliferation that can eventually lead to *transformation*.
- They may release *toxins* that kill cells at a distance, release *enzymes* that degrade tissue components, or damage blood vessels and cause *ischemic necrosis*.
- They can induce host *immune responses* that, though directed against the invader, cause additional tissue damage.

Mechanisms of Viral Injury

Viruses can directly damage host cells by entering them and replicating at the cell's expense. The predilection for viruses to infect certain cells and not others is called *tropism*. **A major determinant of tissue tropism is the presence of viral receptors on host cells.** Viruses bind to proteins found on the surface of host cells that normally function as receptors for host factors. This is presumably one way in which viruses evolve to infect, survive within cells, and spread. For example, HIV glycoprotein gp120 binds to CD4 on T cells and to the chemokine receptors CXCR4 (mainly on T cells) and CCR5 (mainly on macrophages) (Chapter 6), and EBV binds to complement receptor 2 (also known as CR2 or CD21) on B cells. Other tropisms are explained by cell-lineage–specific factors. JC virus infection, which causes leukoencephalopathy (Chapter 28), is restricted to oligodendroglial cells in the CNS because JC viral genes require glial-specific host transcription factors for their expression.

Physical barriers can contribute to tissue tropism. For example, enteroviruses replicate in the intestine in part because they can resist inactivation by acids, bile, and digestive enzymes. Rhinoviruses infect host cells within the upper respiratory tract because they replicate optimally at the lower temperatures found in sites exposed to the ambient atmosphere.

Once viruses are inside host cells, they can damage or kill the cells by a number of mechanisms (Fig. 8.3):

- *Direct cytopathic effects.* Some viruses kill cells by preventing synthesis of critical host macromolecules (e.g., host cell DNA, RNA, or proteins), or by producing degradative enzymes and toxic proteins. Poliovirus inactivates cap-binding protein, which is essential for translation of host cell mRNAs but leaves translation of poliovirus mRNAs unaffected. Viruses can induce cell death by activating so-called *death receptors* (in the tumor necrosis factor [TNF] receptor family) on the plasma membrane and by triggering the intracellular apoptotic machinery. Large amounts of viral proteins are synthesized in infected cells, including unfolded or misfolded proteins that activate the ER stress response; this, too, activates pro-apoptotic pathways. Finally, some viruses encode proteins that are pro-apoptotic, such as the HIV viral protein R (Vpr).
- *Antiviral immune responses.* Host lymphocytes can recognize and destroy virus-infected cells, but, as mentioned earlier, CTLs also can be responsible for tissue injury.
- *Transformation of infected cells.* Oncogenic viruses can stimulate cell growth and survival by a variety of mechanisms, including expression of virus-encoded oncogenes, expression of viral proteins that inactivate key tumor suppressors, and insertional mutagenesis, in which expression of host genes is altered by the insertion of viral genes into host genes or flanking sequences (see Chapter 7).

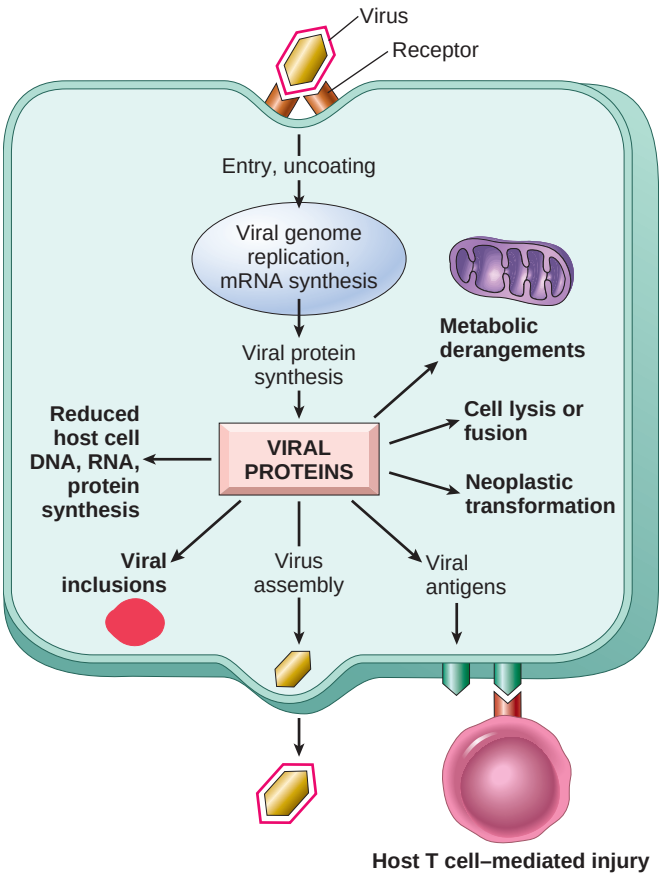


Figure 8.3 Mechanisms by which viruses cause injury to cells.

Mechanisms of Bacterial Injury

Bacterial Virulence. **Bacterial damage to host tissues depends on the ability of the bacteria to adhere to host cells, to invade cells and tissues, and to deliver toxins.** Pathogenic bacteria have *virulence genes* that encode proteins responsible for these properties. An example of the importance of such genes can be found in the various strains of *S. enterica*. All *S. enterica* strains that infect humans are so

closely related that they form a single species, meaning that they share many housekeeping genes. Differences in a relatively small number of virulence genes determine whether an isolate of *S. enterica* causes life-threatening typhoid fever or self-limited enteritis. Virulence genes are frequently found grouped together in clusters called *pathogenicity islands*.

Mobile genetic elements such as plasmids and bacteriophages can transmit functionally important genes to bacteria, including genes that influence pathogenicity and drug resistance. Genes for toxins are sometimes found in plasmids but are more often found in the genomes of bacteriophages, including the genes that encode the toxins responsible for the pathogenesis of the infections cholera, diphtheria, and botulism. Genes for acquired antibiotic resistance traits are more frequently found on plasmids, which can spread not only within bacterial species but also among more distantly related organisms. For example, a plasmid with genes for vancomycin resistance can spread not only among *Enterococcus* spp., but also to more distantly related (and virulent) *S. aureus*.

Many bacteria coordinately regulate gene expression within a large population by a process called *quorum sensing*. Quorum sensing allows bacteria to turn on gene expression and express specific traits only when the organism grows to reach a high concentration. To do this, bacteria secrete small autoinducer molecules which, when present at high levels, induce expression of genes for toxin production *(S. aureus),* competence for genetic transformation *(S. pneumoniae),* or generation of biofilms *(P. aeruginosa).* Autoinducers can be N-acyl-homoserine lactones in gram-negative bacteria, or peptides in gram-positive bacteria. Coordinated expression of virulence factors within bacterial populations may allow bacteria growing in discrete host sites, such as an abscess or consolidated pneumonia, to overcome host defenses. Interestingly, with quorum sensing, different bacterial colonies within the same population may express different genes. Thus, unicellular bacteria acquire some of the more complex properties of multicellular organisms, in which different cells perform different functions.

Communities of bacteria form *biofilms* in which the organisms live within a viscous layer of extracellular polysaccharides that adhere to host tissues or devices such as intravascular catheters and artificial joints. In addition to enhancing adherence to host tissues, biofilms increase the virulence of bacteria by protecting the microbes from immune effector mechanisms and increasing their resistance to antimicrobial drugs. Biofilm formation seems to be particularly important in the persistence and relapse of bacterial endocarditis, artificial joint infections, and respiratory infections in people with cystic fibrosis.

Bacterial Adherence to Host Cells. Bacteria use various surface structures to attach to host cells and tissues.

- *Adhesins* are bacterial surface proteins that bind the organisms to host cells or extracellular matrix. Adhesins have a broad range of host cell specificity. For example, *Streptococcus pyogenes* adheres to host tissues using the adhesins protein F and teichoic acid, which project from the bacterial cell wall and bind to fibronectin on the surface of host cells and in the extracellular matrix.
- *Pili* are filamentous structures on the surface of bacteria that act as adhesins. The stalks of pili are composed of conserved repeating protein subunits, and the variable tip fibrillum determines the tissue-binding specificity of the bacteria. The precise tissue-tropism of *E. coli* that cause urinary tract infections is determined by the tip fibrillum expressed by the bacteria. Adhesion to bladder epithelium is mediated by tip fibrillum that binds d-mannosylated receptors, and adhesion to the renal epithelium depends on binding of tip fibrillum to galabiose-containing glycosphingolipids. Pili can be targets of the host antibody response and, in turn, some bacteria such as *N. gonorrhoeae* vary their pili to escape from the host immune system.

Intracellular Bacteria. Bacteria have evolved a variety of mechanisms for entering host cells. Some bacteria use receptors that are important in the host immune response to gain entry into macrophages. *M. tuberculosis* uses host receptors for opsonins (antibodies and C3b) as well as poorly defined nonopsonic receptors on macrophages. Some gram-negative bacteria use a type III secretion system to enter epithelial cells. This consists of needlelike structures projecting from the bacterial surface that bind to host cells. These proteins then form pores in the host cell membrane and inject bacterial proteins that mediate the rearrangement of the host cell cytoskeleton in a fashion that facilitates bacterial entry.

After bacteria enter the host cell, their fate (and that of the infected cell) varies greatly depending on the organism. *Shigella* spp. and *E. coli* inhibit host protein synthesis, replicate rapidly, and lyse the host cell within hours. Most bacteria are killed within macrophages when the phagosome fuses with an acidic lysosome to form a phagolysosome, where ingested microbes are destroyed, but certain bacteria elude this host defense. *M. tuberculosis* blocks fusion of the lysosome with the phagosome, allowing it to proliferate unchecked within the macrophage. Other bacteria avoid destruction in macrophages by leaving the phagosome and entering the cytoplasm. *L. monocytogenes* produces a pore-forming protein called listeriolysin O and two phospholipases that degrade the phagosome membrane, allowing the bacterium to escape into the cytoplasm, protected from the killing mechanisms of macrophages. In the cytoplasm, *L. monocytogenes* modifies the host cell actin cytoskeleton to promote direct spreading of the organism to neighboring cells. The growth of bacteria inside cells can allow them to escape from certain effector mechanisms of the immune response (e.g., antibodies and complement), and can also facilitate the spread of the bacteria. An example of the latter is the migration of infected macrophages carrying *M. tuberculosis* from the lung to draining lymph nodes and other more distant sites.

Bacterial Toxins. Any bacterial substance that contributes to illness can be considered a toxin. Toxins are classified as endotoxin, which is a component of the gram-negative bacterial cell, and exotoxins, which are proteins that are secreted by many kinds of bacteria.

Bacterial endotoxin is a lipopolysaccharide (LPS) in the outer membrane of gram-negative bacteria that both stimulates host immune responses and injures the host. Lipid A, the part of LPS that anchors the molecule in the host cell membrane, has the endotoxin activity of LPS. LPS

binds to the cell-surface receptor CD14, and the complex of LPS/CD14 then binds to TLR4. Other molecules in the outer structures of gram-positive bacteria can have effects similar to LPS, including lipoteichoic acid, which binds to TLR2. The response to lipid A or lipotechoic acid is beneficial to the host in that it activates protective immunity in several ways. It induces the production of important cytokines and chemoattractants (chemokines) by immune cells and increases the expression of costimulatory molecules, which enhance T-lymphocyte activation. However, high levels of endotoxin play a pathogenic role in septic shock, disseminated intravascular coagulation (DIC), and adult respiratory distress syndrome, mainly through induction of excessive levels of cytokines such as TNF, IL-6, and IL-12.

Exotoxins are secreted bacterial proteins that cause cellular injury and disease. They can be classified into broad categories by their mechanism of action. These are briefly described next and discussed in more detail in the specific sections about each type of bacteria.

- *Enzymes.* Bacteria secrete a variety of enzymes (proteases, hyaluronidases, coagulases, fibrinolysins) that act on substrates in host tissues or on host cells. These enzymes have roles in tissue destruction and abscess formation. For example, exfoliative toxins produced by *S. aureus* cause staphylococcal scalded skin syndrome by degrading proteins that hold keratinocytes together, causing the epidermis to detach from the deeper skin.
- *Toxins that alter intracellular signaling or regulatory pathways.* Most of these toxins have an active (A) subunit with enzymatic activity and a binding (B) subunit that binds to receptors on the cell surface and delivers the A subunit into the cell cytoplasm. The effects of these toxins are diverse and depend on the binding specificity of the B domain and the cellular pathways affected by the A domain. A-B toxins are made by many bacteria including *Bacillus anthracis, V. cholerae,* and some strains of *E. coli*. *Neurotoxins* are A-B toxins produced by *Clostridium botulinum* and *Clostridium tetani* that inhibit release of neurotransmitters, resulting in paralysis. These toxins do not kill neurons; instead, the A domains interact specifically with proteins involved in secretion of neurotransmitters at the synaptic junction. Both tetanus and botulism can result in death from respiratory failure due to paralysis of the chest and diaphragm muscles.
- *Superantigens* are bacterial toxins that stimulate a large number of T lymphocytes by binding to conserved portions of the T-cell receptor, leading to massive T-lymphocyte proliferation and cytokine release. The high levels of cytokines can lead to systemic inflammatory response syndrome.

KEY CONCEPTS

HOST DAMAGE

- Diseases caused by microbes involve interplay between microbial virulence factors and host responses.
- Infectious agents cause death or dysfunction by directly interacting with the host cells.
- Injury may be due to local or systemic release of microbial products including endotoxin (LPS), exotoxins, or superantigens.
- Pathogens can induce immune responses that cause tissue damage. Absence of an immune response may reduce damage induced by some infections; conversely, immune compromise can allow uncontrolled expansion of microorganisms that can directly cause injury.

Spectrum of Inflammatory Responses to Infection

In contrast to the vast molecular diversity of microbes, the morphologic patterns of tissue responses to microbes are limited, as are the mechanisms directing these responses. Therefore, many pathogens produce similar reaction patterns, and few features are unique or pathognomonic for a particular microorganism. Moreover, sometimes the immune status of the host determines the histologic features of the inflammatory response to microbes. Thus, pyogenic bacteria, which normally evoke vigorous leukocyte responses, may cause rapid tissue necrosis with little leukocyte exudation in a profoundly neutropenic host. Similarly, in a patient who is not immunocompromised, *M. tuberculosis* causes well-formed granulomas with few mycobacteria present, whereas in an AIDS patient the same mycobacteria multiply profusely in macrophages, which fail to coalesce into granulomas.

There are five major histologic patterns of tissue reaction in infections (Table 8.3).

Suppurative (Purulent) Inflammation

This pattern is characterized by increased vascular permeability and leukocytic infiltration, predominantly of neutrophils (see Fig. 3.15). The neutrophils are attracted to the site of infection by release of chemoattractants from the "pyogenic" (pus-forming) bacteria that evoke this response, mostly extracellular gram-positive cocci and gram-negative rods. Masses of dying and dead neutrophils and liquefactive necrosis of the tissue form pus. The sizes of purulent lesions range from tiny microabscesses formed in multiple organs during bacterial sepsis secondary to a colonized heart valve, to diffuse involvement of entire lobes of the lung in pneumonia. How destructive the lesions are depends on their location and the organism involved. For example, *S. pneumoniae* usually spare alveolar walls and cause lobar pneumonia that resolves completely, whereas *S. aureus* and *Klebsiella pneumoniae* destroy alveolar walls and form abscesses that heal with scar formation. Bacterial pharyngitis *(S. pyogenes)* can resolve without sequelae, whereas untreated acute bacterial inflammation of a joint can destroy the joint in a few days.

Mononuclear and Granulomatous Inflammation

Diffuse, predominantly mononuclear, interstitial infiltrates are a common feature of all chronic inflammatory processes, but when they develop acutely, they often are a response to viruses, intracellular bacteria, or intracellular parasites. In addition, spirochetes and helminths provoke chronic inflammatory responses. Which mononuclear cell predominates within the inflammatory lesion depends on the host immune response to the organism. For example, plasma cells are abundant in the primary and secondary lesions of syphilis (Fig. 8.4), whereas lymphocytes predominate in HBV infection or viral infections of the brain. The presence

Table 8.3 Spectrum of Inflammatory Responses to Infection

Type of Response	Pathogenesis	Examples
Suppurative (Purulent) Infection	Increased vascular permeability Leukocyte infiltration (neutrophils) Chemoattractants from bacteria Formation of "pus"	Pneumonia *(Staphylococcus aureus)* Abscesses (*Staphylococcus* spp., anaerobic and other bacteria)
Mononuclear and granulomatous inflammation	Mononuclear cell infiltrates (monocytes, macrophages, plasma cells, lymphocytes) Cell-mediated immune response to pathogens ("persistent antigen") Formation of granulomata	Syphilis Tuberculosis
Cytopathic-cytoproliferative reactions	Viral transformation of cells Necrosis or proliferation (including multinucleation) Linked to neoplasia	Cervical cancer (human papillomavirus) Chicken pox, shingles Herpes
Tissue necrosis	Toxin- or lysis-mediated destruction Lack of inflammatory cells Rapidly progressive processes	Gangrene *(Clostridium perfringens)* Hepatitis (hepatitis B virus)
Chronic inflammation/ scarring	Repetitive injury leads to fibrosis Loss of normal parenchyma	Chronic hepatitis with cirrhosis (hepatitis B and C viruses)
No reaction	Severe immune compromise	*Mycobacterium avium* in untreated AIDS (T-cell deficiency) Mucormycosis in bone marrow transplant patients (neutropenia)

of these lymphocytes reflects cell-mediated immune responses against the pathogen or pathogen-infected cells. At the other extreme, macrophages may become filled with organisms, as occurs in *M. avium complex* infections in AIDS patients, who cannot mount an effective immune response to the organisms.

Granulomatous inflammation is a distinctive form of mononuclear inflammation usually evoked by infectious agents that resist eradication and are capable of stimulating strong T cell–mediated immunity (e.g., *M. tuberculosis, Histoplasma capsulatum*, schistosome eggs). Granulomatous inflammation is characterized by accumulation and aggregation of activated macrophages called "epithelioid" cells, some of which may fuse to form giant cells. Granulomas may contain a central area of caseous necrosis (see Chapter 3 and discussion of tuberculosis later in this chapter).

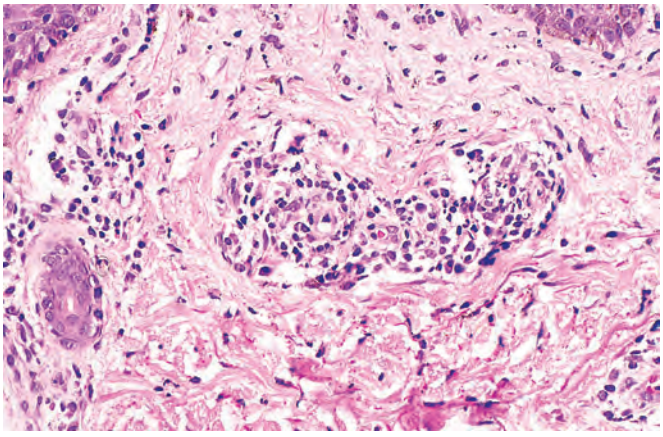


Figure 8.4 Secondary syphilis in the dermis with perivascular lymphoplasmacytic infiltrate and endothelial proliferation.

Cytopathic-Cytoproliferative Reaction

These reactions are usually produced by viruses. The lesions are characterized by cell necrosis or cellular proliferation, usually with sparse inflammatory cells. Some viruses replicate within cells and make viral aggregates that are visible as inclusion bodies (e.g., herpesviruses or adenovirus) or induce cells to fuse and form multinucleated cells called *polykaryons* (e.g., measles virus or herpesviruses). Focal cell damage in the skin may cause epithelial cells to become detached, forming blisters (Fig. 8.5). Some viruses can cause epithelial cells to proliferate (e.g., venereal warts caused by HPV or the umbilicated papules of molluscum contagiosum caused by poxviruses). Finally, viruses can contribute to the development of malignant neoplasms (Chapter 7).

Tissue Necrosis

Clostridium perfringens and other organisms such as *Corynebacterium diphtheriae* that secrete powerful toxins cause such rapid and severe necrosis (gangrenous necrosis) that tissue

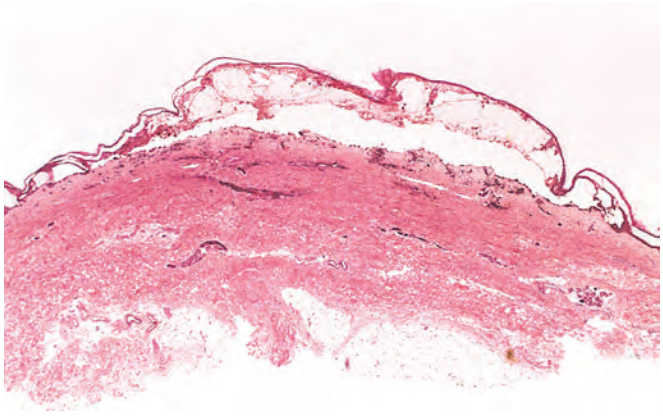


Figure 8.5 Herpesvirus blister in mucosa. See Fig. 8.9 for viral inclusions.

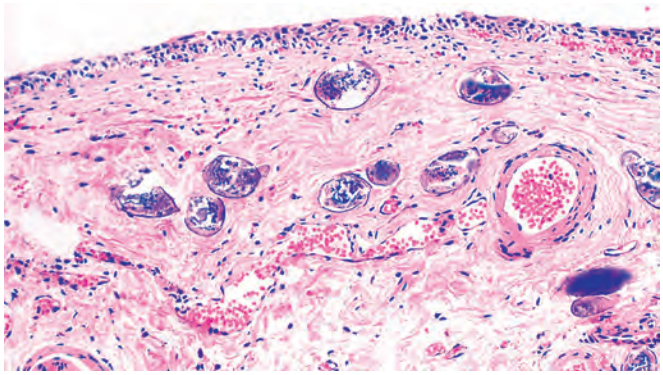


Figure 8.6 *Schistosoma haematobium* infection of the bladder with numerous calcified eggs and extensive scarring.

damage is the dominant feature. The parasite *E. histolytica* causes colonic ulcers and liver abscesses characterized by extensive tissue destruction with liquefactive necrosis and little inflammatory infiltrate. Some viruses can cause widespread and severe necrosis of host cells associated with inflammation, as exemplified by total destruction of the temporal lobes of the brain by HSV or the liver by HBV.

Chronic Inflammation and Scarring

Many infections elicit chronic inflammation, which can lead either to complete healing or to extensive scarring. For example, chronic HBV infection may cause cirrhosis of the liver, in which dense fibrous septae surround nodules of regenerating hepatocytes with complete loss of normal liver architecture and consequent changes in blood flow. Sometimes the exuberant scarring response is the major cause of dysfunction (e.g., the "pipestem" fibrosis of the liver or fibrosis of the bladder wall caused by schistosomal eggs [Fig. 8.6] or the constrictive fibrous pericarditis in tuberculosis).

These patterns of tissue reaction are useful guidelines for analyzing microscopic features of infectious processes, but they rarely appear in pure form because different types of host reactions often occur at the same time. For example, the lung of an AIDS patient may be infected with CMV, which causes cytolytic changes, and at the same time by *Pneumocystis* spp., which causes interstitial inflammation. Similar patterns of inflammation also can be seen in tissue responses to physical or chemical agents and in inflammatory diseases of unknown cause (Chapter 3).

This concludes our discussion of the general principles of the pathogenesis and pathology of infectious disease. We now turn to specific infections caused by viruses, bacteria, fungi, and parasites, and focus on their pathogenic mechanisms and pathologic effects rather than details of clinical features, which are available in clinical textbooks. Infections that typically involve a specific organ are discussed in other chapters.

VIRAL INFECTIONS

Viruses are the cause of many clinically important acute and chronic infections, which may affect virtually every organ system (Table 8.4).

Acute (Transient) Infections

The viruses that cause transient infections are structurally heterogeneous, but all elicit effective immune responses that eliminate the pathogens, limiting the durations of these infections. However, specific viruses exhibit widely differing degrees of genetic diversity, a variable that has an important impact on the susceptibility of the host to reinfection by viruses of the same type. The mumps virus, for example, has only one genetic subtype and infects people only once, whereas other viruses, such as influenza viruses, can repeatedly infect the same individual because new genetic variants arise periodically in nature. Immunity to some viruses,

Table 8.4 Selected Human Viruses and Viral Diseases

Organ System	Species	Disease
Respiratory	Adenovirus	Upper and lower respiratory tract infections, conjunctivitis, diarrhea
	Rhinovirus	Upper respiratory tract infection
	Influenza viruses A, B	Influenza
	Respiratory syncytial virus	Bronchiolitis, pneumonia
Digestive	Mumps virus	Mumps, pancreatitis, orchitis
	Rotavirus	Childhood gastroenteritis
	Norovirus	Gastroenteritis
	Hepatitis A virus	Acute viral hepatitis
	Hepatitis B virus	Acute or chronic hepatitis
	Hepatitis D virus	With hepatitis B virus, acute or chronic hepatitis
	Hepatitis C virus	Acute or chronic hepatitis
	Hepatitis E virus	Acute viral hepatitis
Systemic with skin eruptions	Measles virus	Measles (rubeola)
	Rubella virus	German measles (rubella)
	Varicella-zoster virus	Chickenpox, shingles
	Herpes simplex virus 1	Oral herpes ("cold sore")
	Herpes simplex virus 2	Genital herpes
Systemic with hematologic disorders	Cytomegalovirus	Cytomegalic inclusion disease
	Epstein-Barr virus	Infectious mononucleosis
	Human immunodeficiency viruses 1 and 2	Acquired immunodeficiency syndrome
Arboviral and hemorrhagic fevers	Dengue viruses 1 to 4	Dengue hemorrhagic fever
	Yellow fever virus	Yellow fever
Skin/genital warts	Human papillomavirus	Condyloma; cervical carcinoma
Central nervous system	Poliovirus	Poliomyelitis
	JC virus	Progressive multifocal leukoencephalopathy (opportunistic)

particularly respiratory and gastrointestinal viruses, wanes with time, and this too allows the same virus to infect the host repeatedly.

Measles

Measles is an acute viral infection that affects multiple organs and causes a wide range of disease, from mild, self-limited infections to severe systemic manifestations. A major initiative by several governments and international bodies, including the World Health Organization (WHO), to increase vaccination for measles reduced the number of measles-related deaths by 84% from 2000 to 2016, however there were still almost 90,000 deaths worldwide. Because of poor nutrition and lack of access to medical care, children in lower-income countries are 10 to 1000 times more likely to die of measles than are children in higher-income countries. Measles can produce severe disease in people with defects in cellular immunity (e.g., people infected with HIV or people with a hematologic malignancy). In higher-income nations, epidemics of measles occur when the virus is introduced by individual(s) traveling from an area of endemic disease, and then spreads, primarily to unvaccinated individuals. In recent years, such outbreaks have occurred several times each year in the United States. The diagnosis may be made clinically, by serology, or by detection of viral RNA in respiratory secretions or urine.

Pathogenesis

Measles virus is a single-stranded RNA virus of the Paramyxoviridae family, which includes mumps, respiratory syncytial virus, parainfluenza virus (a cause of croup), and human metapneumovirus. There is only one serotype of measles virus. Measles virus is very efficiently transmitted by the airborne route via aerosolized respiratory secretions. Three cell-surface receptors have been identified for measles hemagglutinin protein. Signaling lymphocytic activation molecule family member 1 (SLAMF1) is expressed on activated lymphocytes, dendritic cells, and monocytes, and it serves as the initial receptor for viral infection. Nectin-4 is found on the basal surface of epithelial cells and is thought to be important for replication of the virus within the respiratory tract, before spread of the virus in respiratory secretions. CD46 was the first cell-surface receptor identified for measles virus, but it has been found to be used only by culture-adapted virus (including the vaccine strain), and not wild-type virus.

Measles can replicate in a variety of cell types, including epithelial cells and leukocytes. The virus initially multiplies within the respiratory tract and then spreads to local lymphoid tissues. Replication of the virus in lymphatic tissue is followed by viremia and dissemination to many tissues, including the conjunctiva, skin, respiratory tract, urinary tract, small blood vessels, lymphatic system, and CNS. Most children develop T cell–mediated immunity to measles virus that helps control the viral infection and produces the measles rash. Hence, the rash is less frequent in people with deficiencies in cell-mediated immunity. In addition, in malnourished children with poor medical care, measles virus may cause croup, pneumonia, diarrhea and protein-losing enteropathy, keratitis leading to scarring and blindness, encephalitis, and hemorrhagic rashes. Subacute sclerosing panencephalitis (Chapter 28) and measles inclusion body encephalitis (in immunocompromised individuals) are rare late complications of measles. The pathogenesis of subacute sclerosing panencephalitis is not well understood, but a replication-defective variant of measles may be involved in this persistent viral infection.

Antibody-mediated immunity to measles virus protects against reinfection. Measles also can cause transient but profound immunosuppression, resulting in secondary bacterial and viral infections, which are responsible for much of measles-related morbidity and mortality. Delayed-type hypersensitivity responses are reduced following measles infection, indicating a reduction in lymphocyte responses. This may be associated with an inhibition of the ability of infected dendritic cells to stimulate lymphocytes.

MORPHOLOGY

The blotchy, reddish brown rash of measles virus infection on the face, trunk, and proximal extremities is produced by dilated skin vessels, edema, and a mononuclear perivascular infiltrate. Ulcerated mucosal lesions in the oral cavity near the opening of the Stensen ducts (the pathognomonic *Koplik spots*) are marked by necrosis, neutrophilic exudate, and neovascularization. The lymphoid organs typically have marked follicular hyperplasia, large germinal centers, and randomly distributed multinucleate giant cells, called *Warthin-Finkeldey cells*, which have eosinophilic nuclear and cytoplasmic inclusion bodies (Fig. 8.7). These are pathognomonic of measles and are also found in the lung and sputum. The milder forms of measles pneumonia show the same peribronchial and interstitial mononuclear cell infiltration that is seen in other nonlethal viral infections.

Mumps

Mumps is an acute systemic viral infection usually associated with pain and swelling of the salivary glands. Like measles virus, mumps virus is a member of the Paramyxoviridae family. Mumps virus has two types of surface glycoproteins, one with hemagglutinin and neuraminidase activities and the other with cell fusion and cytolytic activities. Mumps viruses enter the upper respiratory tract through inhalation of or contact with respiratory droplets, spread to draining lymph nodes where they replicate in lymphocytes

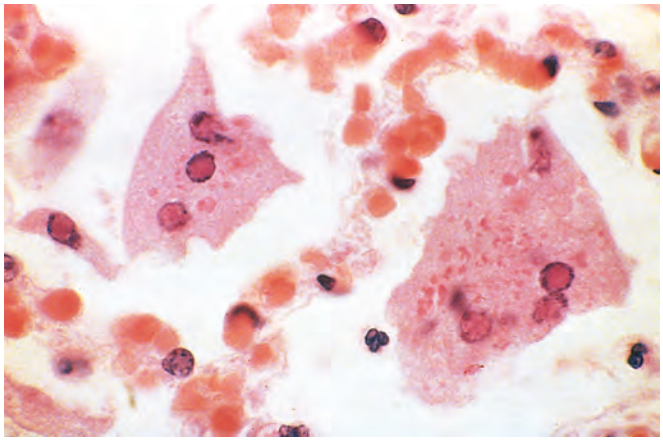


Figure 8.7 Measles giant cells in the lung. Note the glassy eosinophilic intranuclear inclusions.

(preferentially in activated T cells), and then spread through the blood to the salivary and other glands. Mumps virus infects salivary gland ductal epithelial cells, resulting in desquamation of involved cells, edema, and inflammation that leads to the classic salivary gland pain and swelling. Mumps virus also can spread to other sites, including the CNS, testis, ovary, and pancreas. Aseptic meningitis is the most common extrasalivary gland complication of mumps, occurring in up to 15% of cases. In the United States, outbreaks of mumps have occurred in populations with close contact (e.g., university or high-school settings) in recent years, resulting in approximately 6000 cases per year. It is important to note that this is still more than 99% fewer cases than occurred annually in the United States before use of the mumps vaccine. The diagnosis is usually made clinically, but serology or detection of viral RNA in saliva can be used for definitive diagnosis.

MORPHOLOGY

Mumps parotitis is bilateral in 70% of cases. The affected glands are enlarged, have a doughy consistency, and are moist, glistening, and reddish-brown on cross-section. On microscopic examination, the gland interstitium is edematous and diffusely infiltrated by macrophages, lymphocytes, and plasma cells, which compress acini and ducts. Neutrophils and necrotic debris may fill the duct lumen and cause focal damage to the lining epithelium.

In **mumps orchitis,** testicular swelling may be marked, caused by edema, mononuclear cell infiltration, and focal hemorrhages. Because the testis is tightly contained within the tunica albuginea, parenchymal swelling may compromise the blood supply and cause areas of infarction. The testicular damage can lead to scarring, atrophy, and, if severe, sterility.

Infection and damage of acinar cells in the pancreas may release digestive enzymes, causing parenchymal and fat necrosis and neutrophil-rich inflammation. **Mumps encephalitis** is associated with perivenous demyelination and perivascular mononuclear cuffing.

Poliomyelitis

Poliovirus causes an acute systemic viral infection, leading to a wide range of manifestations, from mild, self-limited infections to paralysis of limb muscles and respiratory muscles. Poliovirus is a spherical, unencapsulated RNA virus of the enterovirus genus. Other enteroviruses cause childhood diarrhea as well as rashes (coxsackievirus A), conjunctivitis (enterovirus 70), viral meningitis (coxsackieviruses and echovirus), and myopericarditis (coxsackievirus B). There are three serotypes of poliovirus, but it is likely that most infections are caused by type 1. The inactivated (injected) poliovirus vaccine protects against all three serotypes, and the attenuated (oral) poliovirus vaccine is available in various combinations of one, two, or all three serotypes, although only formulations containing one or two serotypes are currently in use. Use of these vaccines has nearly eradicated polio, because poliovirus infects only humans, shows limited genetic variation, and is effectively neutralized by antibodies generated by immunization. According to global polio surveillance data, in 2017, a total of only 22 polio cases were reported; however, additional cases were reported in 2018 and 2019 caused by the natural infection and the vaccine strain.

Poliovirus, like other enteroviruses, is transmitted by the fecal-oral route. The virus infects human cells by binding to CD155, a molecule expressed on a variety of cell types, including epithelial cells, lymphocytes, and neurons. The virus is ingested and replicates in the mucosa of the pharynx and gut, including tonsils and Peyer patches in the ileum. Poliovirus then spreads through lymphatics to lymph nodes and eventually the blood, producing transient viremia and fever. Although most poliovirus infections are asymptomatic, in about 1 in 100 infected persons poliovirus invades the CNS and replicates in motor neurons of the spinal cord (spinal poliomyelitis) or brainstem (bulbar poliomyelitis). Antiviral antibodies control the disease in most cases; it is not known why they fail to contain the virus in some individuals. Viral spread to the nervous system may be through the blood or by retrograde transport of the virus along axons of motor neurons. Rare cases of poliomyelitis that occur after vaccination are caused by mutations in the attenuated viruses to revert to wild-type, virulent, forms. The diagnosis can be made by viral culture or detection of viral RNA in throat secretions or, more often, stool, or by serology. The neurologic features and neuropathology of poliovirus infection are described in Chapter 28.

West Nile Virus Infections

West Nile Virus causes an acute systemic infection that has two very different presentations: a mild, self-limited infection or neuroinvasive disease associated with long-term neurologic sequelae. West Nile virus is an arthropod-borne virus (arbovirus) of the flavivirus group, which also includes viruses that cause dengue fever and yellow fever. West Nile virus has a broad geographic distribution in the Old World, including Africa, the Middle East, Europe, Southeast Asia, and Australia. It was first detected in the United States in 1999 during an outbreak in New York City and has since spread across the United States; in 2017, a least one case was reported in 47 states. West Nile virus is transmitted by mosquitoes to birds and to mammals. Infected birds develop prolonged viremia and are the major reservoir for the virus. Humans are incidental hosts. Most affected patients acquire the infection from a mosquito bite; less commonly, human-to-human transmission occurs by blood transfusion, organ transplantation, breastfeeding, or transplacental spread.

After inoculation by a mosquito, West Nile virus replicates in skin dendritic cells, which then migrate to lymph nodes. Here, the virus replicates further, enters the bloodstream, and, in some individuals, crosses the blood-brain barrier and infects neurons in the CNS.

West Nile virus infection is usually asymptomatic, but in 20% of infected individuals it gives rise to a fever, headache, myalgia, fatigue, anorexia, and nausea. A maculopapular rash is seen in approximately one-half of cases. CNS complications (meningitis, encephalitis, meningoencephalitis) occur in about 1 in 150 clinically apparent infections. Meningoencephalitis has a mortality rate of about 10% and results in long-term cognitive and neurologic impairment in many survivors. Immunosuppressed persons and older adults appear to be at the greatest risk for severe disease. Rare complications include hepatitis, myocarditis, and pancreatitis. The diagnosis is usually made by serology, but viral culture and polymerase chain reaction (PCR)-based tests are also used.

MORPHOLOGY

Perivascular and leptomeningeal chronic inflammation, microglial nodules (Chapter 28), and neuronophagia predominantly involving the temporal lobes and brainstem have been observed in patients who died of West Nile virus infection.

Viral Hemorrhagic Fever

Viral hemorrhagic fever is a severe life-threatening multisystem syndrome in which there is vascular damage, leading to widespread hemorrhage and shock. Viral hemorrhagic fever is caused by enveloped RNA viruses belonging to four different genera: *Arenaviridae, Filoviridae, Bunyaviridae,* and *Flaviviridae*. These viruses can produce a spectrum of illnesses, ranging from a mild acute disease characterized by fever, headache, myalgia, rash, neutropenia, and thrombocytopenia to severe, life-threatening disease in which there is sudden hemodynamic deterioration and shock. These viruses pass through an animal or insect host during their life cycles, and therefore their ranges are restricted to areas in which their hosts reside. Humans are incidental hosts who are infected when they come into contact with infected hosts (typically rodents) or insect vectors (mosquitoes and ticks). Some viruses that cause hemorrhagic fever (Ebola, Marburg, Lassa) also can spread from person to person.

The pathogenesis of the infection and its complications vary among the different viruses, but there are some common features. Damage to blood vessels is often prominent. It may be caused by direct infection of and damage to endothelial cells, or infection of macrophages and dendritic cells leading to production of inflammatory cytokines.

There was a major outbreak of Ebola virus in Sierra Leone, Liberia and Guinea in 2014 to 2016, with over 28,000 cases and 11,000 deaths reported to the WHO. A second outbreak in the Democratic Republic of Congo occurred in 2019. The outbreaks were characterized by explosive spread of the virus by person-to-person transmission through exposure to mucosal secretions. Ebola carries a very high mortality rate, about 40% during the 2019 outbreak. Effective vaccines for Ebola have been developed, and use of these will reduce the burden of disease. Ebola virus has specific mechanisms by which it can evade the immune response, allowing the virus to rapidly replicate to reach high levels. Two viral proteins, VP24 and VP35, inhibit the action of type I IFN. VP24 blocks type I IFN signaling by preventing tyrosine-kinase dimerization and nuclear translocation of STAT-1. VP35 binds to double-stranded viral RNA in infected host cells, preventing detection of the RNA by cytoplasmic receptors that stimulate type I IFN production. In addition, the viral surface protein, GP, is found in non–virus-associated soluble form, which might act as a decoy for binding of host antibodies, blocking the antibodies from neutralizing intact virus.

MORPHOLOGY

With viral hemorrhagic fever, there may be hemorrhagic manifestations, including petechiae, caused by a combination of thrombocytopenia or platelet dysfunction, endothelial injury, cytokine-induced DIC, and deficiency of clotting factors because of hepatic injury. Hemorrhages may be prominent in some infections (e.g., Congo-Crimean fever) but are rarely life-threatening. Necrosis of tissues secondary to the vascular lesions and hemorrhages may be seen and varies from mild and focal to massive, but the attendant inflammatory response is usually minimal. In Ebola virus infection, there are widespread hemorrhage and viral replication in many organs including the liver with hepatocellular necrosis and scant inflammation, and the spleen with lymphocyte apoptosis and viral replication in dendritic cells, fibroblasts, and monocytes.

Zika Virus Infections

Zika virus is a Flavivirus that was discovered in 1947 and was subsequently found to be widespread in Africa, Asia, and the Middle East. Large outbreaks occurred in the State of Yap in 2007 and French Polynesia in 2013 to 2014. In 2015, an outbreak of Zika virus from March until June in Brazil caused up to 1.3 million suspected cases. Subsequently, a marked increase in the occurrence of microcephaly in infants born in Brazil was noted with 4300 cases reported, and Zika virus has now been clearly linked to this unfortunate outcome. This Zika virus outbreak spread to at least 33 countries and territories in the Americas, with one of the largest outbreaks occurring in Colombia, which had over 50,000 suspected cases. In the United States, most of the approximately 5600 cases occurred in travelers returning from affected areas, however over 200 cases were acquired by local transmission. Zika virus is transmitted by *Aedes* mosquitos, primarily *Aedes aegypti*. It is likely that there is an animal reservoir for the virus, although none has been demonstrated to be a source of human infection. In addition to mosquito-borne and perinatal transmission, the virus can also infect through blood transfusion and sexual contact. The virus is detectable in semen, vaginal secretions, breast milk, and urine for up to several months after infection, although the duration of infectivity from these secretions is not yet clear.

The manifestations of Zika virus infection in adults are usually mild and nonspecific, with fever, myalgia, arthralgia, conjunctivitis, and maculopapular rash lasting a few days to 1 week. The infection is associated with a small number of cases of neurologic complications in adults, primarily Guillain-Barré syndrome. Perinatal transmission of the virus can result in fetal death or moderate to severe brain defects in the fetus and newborn child. In Brazil, the rate of adverse effects on newborns following maternal Zika virus infection was 46%, with the highest rates occurring following infection during the first and second trimesters. The rate of adverse effects on newborns has been higher in Brazil than in other nations. In the US territories, the rates of Zika virus–associated birth defects for infections during the first, second, and third trimesters of pregnancy were 8%, 5%, and 4%, respectively. The reasons for the differences in these rates are not known but might include differences in data-collection methods or effects of genetics, co-infections, immunity from past infections, or as-yet undefined environmental factors.

Our developing understanding of the pathogenesis of Zika virus is from autopsies of infected newborns, in-vitro studies of infected human tissues, and limited animal studies. Viral proteins are present in degenerating neurons and glial cells, as well as in chorionic villi of the placenta in some cases, indicating replication of virus with associated necrosis

in these cells. Viral RNA can be detected in brain tissues in all cases, but virus is absent or present only variably in other tissues. Studies in mice with immune defects suggest that Zika virus replicates initially in the placenta followed by infection of the fetus, and that viral replication occurs in neural precursor cells, leading to reduced brain size and thinning of the cortex. In vitro studies using cell lines and brain organoids, composed of differentiated pluripotent stem cells in a tissue-like structure, confirm that Zika virus replicates in and kills human neurons.

MORPHOLOGY

Cerebral calcifications, cerebral atrophy, ventricular enlargements, and hypoplastic cerebral structures were the most common adverse outcomes associated with Zika virus infection (Fig. 8.8). Microcephaly occurred in a small number of infants, 3.4% of those born to infected women, with one-half of cases being proportionate microcephaly in children who were small for gestational age, and the other half being disproportionate microcephaly, meaning that the head size was small relative to the size of the child. Ocular abnormalities, including pigment mottling, chorioretinal atrophy, and optic nerve abnormalities, are also common.

In small autopsy series, common findings in newborn children include microcephaly, ventriculomegaly, and congenital joint contractures (arthrogryposis) and pulmonary hypoplasia. Severe neuronal depletion and associated thinning of the brain parenchyma occurred, with microcalcifications and microglial nodules.

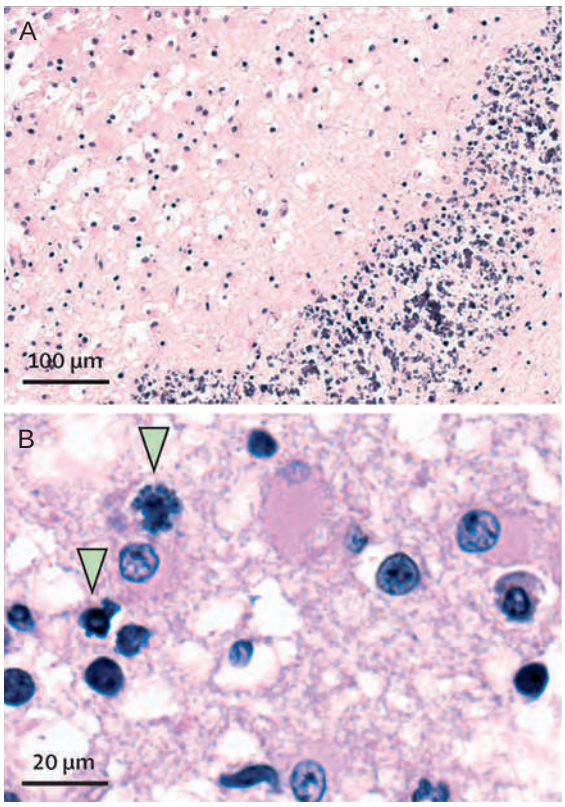


Figure 8.8 The morphology of brain from a two month old child with congenital Zika virus infection. (A) Subcortical band of degenerating cells with prominent calcifications. (B) Cortex with degenerating neurons *(arrows)*. (From Martinez, Pathology of congenital Zika syndrome in Brazil: a case series, *Lancet* 388(10047):898–904, 2016.)

Dengue

Dengue virus is a flavivirus transmitted by *Aedes* mosquitoes in tropical and subtropical regions. It is estimated to infect 390 million people each year, 960,000 of whom have severe infection that requires hospital admission. Clinical manifestations of dengue vary from fever with headache, macular rash and severe myalgias (breakbone fever) to severe dengue *(dengue hemorrhagic fever)* with bleeding, liver failure, reduced consciousness, organ failure, and plasma leakage leading to shock and respiratory distress. In severe dengue, there are widespread hemorrhages throughout the body with hepatic necrosis and mononuclear infiltrates, septal thickening, and hyaline membrane formation in the lung.

The immune response against dengue virus is thought to determine, in part, the severity of the infection. There are four serotypes of dengue virus. Infection with each serotype stimulates protective immunity against that serotype, but also stimulates a cross-reactive antibody response that is weak and nonprotective for other serotypes of the virus. Severe dengue usually occurs in people who have had previous infection with a different serotype than the one associated with their severe illness. Antibody-dependent enhancement, in which the cross-reactive antibodies enhance uptake of virus into macrophages via Fc receptors, is thought to increase infectivity of the virus and contribute to severe dengue. Severe dengue also occurs in infants who have maternal antibodies against dengue virus, consistent with this model.

Novel Coronavirus SARS-CoV-2 (COVID-19)

A novel coronavirus SARS-CoV-2-mediated disease called COVID-19 was first detected in Wuhan, China, and was reported to the WHO in December 2019. By the end of March 2020, the infection had become a worldwide pandemic, with more than 800,000 cases and about 40,000 deaths. This devastating pandemic has caused major social and economic disruptions, in addition to its severe health consequences. Epidemiologic studies suggest that the origin was a seafood and animal market in Wuhan, consistent with initial animal-to-human transmission followed by person-to-person spread soon thereafter. Sequencing of the full viral genome led to development of a rapid molecular diagnostic assay during the evolving outbreak. The clinical manifestations ranged from mild to severe respiratory illness. The vast majority of individuals who contract the virus recover after a flu-like disease. Severe and sometimes fatal illness with respiratory compromise, often associated with bilateral ground-glass opacities on chest imaging, occurs mainly in older individuals and those with comorbidities such as diabetes, COPD, and heart failure. The genomic sequence shows COVID-19 is related to bat coronaviruses and the SARS coronavirus. Histologic analysis of lung tissue done in a few patients revealed diffuse alveolar damage and inflammation with mainly mononuclear cells.

Latent Infections (Herpesvirus Infections)

Latency is defined as the persistence of viral genomes in cells that do not produce infectious virus. Dissemination of the infection and tissue injury stem from reactivation of the latent virus. The viruses that most frequently establish latent infections in humans are herpesviruses. These are large encapsulated viruses with double-stranded DNA

genomes that encode approximately 70 proteins. Herpesviruses cause acute infection followed by latent infection in which the viruses persist in a noninfectious form with periodic reactivation and shedding of infectious virus.

There are eight types of human herpesviruses, belonging to three subgroups that are defined by the type of cell most frequently infected and the site of latency.

- The *α-group viruses,* including HSV-1, HSV-2, and VZV, infect epithelial cells and produce latent infection in postmitotic neurons.
- The *β-group viruses* include CMV and human herpesviruses 6 and 7 (HHV-6 and HHV-7), which infect and produce latent infection in a variety of cell types. HHV-6 and HHV-7 cause exanthem subitum, also known as *roseola infantum,* and sixth disease, a benign rash of infants, and also have been associated with encephalitis, pneumonitis, hepatitis, and myelitis on reactivation.
- The *γ-group viruses,* including EBV and Kaposi sarcoma–associated virus (KSHV/HHV-8), produce latent infection mainly in lymphoid cells.
- *Herpesvirus simiae* (monkey B virus) is an Old World monkey virus that resembles HSV-1 and can cause fatal neurologic disease in animal handlers, usually resulting from an animal bite.

Herpes Simplex Virus (HSV) Infections

HSV-1 and HSV-2 differ serologically but are closely related genetically and cause a similar set of primary and recurrent infections. Herpes simplex viruses are common throughout the world, with an estimated 4 billion infections globally. Both viruses replicate in the skin and the mucous membranes at the site of entry of the virus (usually oropharynx or genitals), where they produce infectious virions and cause vesicular lesions of the epidermis. The viruses spread to sensory neurons that innervate these primary sites of viral replication and release. Viral nucleocapsids are transported along axons to the neuronal cell bodies, where the viruses establish latent infection. In immunocompetent hosts, primary HSV infection resolves in a few weeks, although the virus remains latent in nerve cells. During latency the viral DNA remains within the nucleus of the neuron, and only latency-associated viral RNA transcripts (LATs) are synthesized. No viral proteins appear to be produced during latency. LATs may contribute to latency by conferring resistance to apoptosis, silencing lytic gene expression through heterochromatin formation, and serving as precursors for micro-RNAs that downregulate expression of critical HSV lytic genes. Reactivation of HSV-1 and HSV-2 may occur repeatedly with or without symptoms, and results in the spread of virus from the neurons to the skin or to mucous membranes. It is proposed that during reactivation, release from epigenetic silencing occurs through a methylation/phosphorylation process that is linked to the cellular stress response pathways in neurons. Reactivation can occur in the presence of host immunity, because HSVs have developed ways to avoid immune recognition. For example, HSVs can evade antiviral CTLs by inhibiting the class I MHC recognition pathway, and elude humoral immune defenses by producing "decoy" receptors that bind the Fc domain of immunoglobulin and inhibitors of complement. HSVs can infect multiple cell types, including dendritic cells that are important for the antiviral immune response.

In addition to causing cutaneous and genital lesions, HSV-1 is the major infectious cause of corneal blindness in the United States. Corneal epithelial disease is thought to be due to direct viral damage, and corneal stromal disease appears to be immune-mediated. HSV-1 is also the major cause of fatal sporadic encephalitis in the United States. When the infection spreads to the brain, it usually involves the temporal lobes and orbital gyri of the frontal lobes. Inherited mutations in TLR3 or components of its signaling pathway increase the risk of HSV encephalitis. In addition, neonates and individuals with compromised cellular immunity (e.g., secondary to HIV infection or chemotherapy) may suffer disseminated herpesvirus infections. Due to ulceration and dampening of the immune response, HSV-2 infection increases the risk of HIV transmission fourfold and increases the risk of HIV acquisition twofold to threefold.

MORPHOLOGY

HSV-infected cells contain large, pink to purple *intranuclear inclusions* (Cowdry type A) that consist of viral replication proteins and virions at various stages of assembly that push the host cell chromatin out to the edges of the nucleus (Fig. 8.9). Due to cell fusion, HSV also produces inclusion-bearing multinucleated syncytia.

HSV-1 and HSV-2 cause lesions ranging from self-limited cold sores and gingivostomatitis to life-threatening disseminated visceral infections and encephalitis. **Fever blisters or cold sores** favor the facial skin around mucosal orifices (lips, nose), where their distribution is frequently bilateral and independent of skin dermatomes. Intraepithelial vesicles (blisters), which are formed by intracellular edema and ballooning degeneration of epidermal cells, frequently burst and crust over, but some may result in superficial ulcerations.

Gingivostomatitis, which is usually encountered in children, is caused primarily by HSV-1. It is a vesicular eruption extending from the tongue to the retropharynx and causing cervical lymphadenopathy. Swollen, painful, erythematous HSV lesions of the fingers or palm (herpetic whitlow) occur in infants and, occasionally, in health care workers due to occupational exposure.

Genital herpes is more often caused by HSV-2 than by HSV-1. It is characterized by vesicles on the genital mucous membranes as well as on the external genitalia that are rapidly converted into superficial ulcerations that are rimmed by an

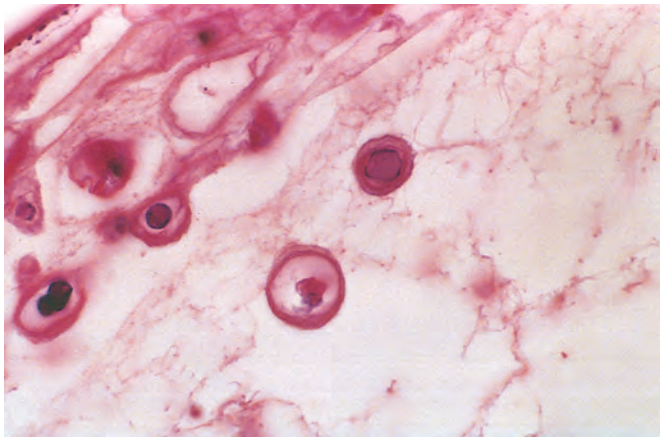


Figure 8.9 A herpesvirus blister showing glassy intranuclear viral inclusion bodies.

inflammatory infiltrate (Chapter 22). Herpesvirus (usually HSV-2) can be transmitted to neonates during passage through the birth canal of infected mothers. Although HSV-2 infection in the neonate may be mild, more commonly it is fulminating with generalized lymphadenopathy, splenomegaly, and necrotic foci throughout the lungs, liver, adrenals, and CNS.

Two forms of corneal lesions are caused by HSV (Chapter 29). **Herpes epithelial keratitis** shows typical virus-induced cytolysis of the superficial epithelium. In contrast, **herpes stromal keratitis** is characterized by infiltrates of mononuclear cells around keratinocytes and endothelial cells, leading to neovascularization, scarring, opacification of the cornea, and eventual blindness. Here, the damage is caused by an immunologic reaction to the HSV infection, rather than the cytopathic effects of the virus itself.

Herpes simplex encephalitis is described in Chapter 28.

Disseminated skin and visceral herpes infections are usually encountered in hospitalized patients with underlying cancer or immunosuppression. **Herpes esophagitis** is frequently complicated by superinfection with bacteria or fungi. **Herpes bronchopneumonia,** sometimes stemming from intubation of a patient with active oral lesions, is often necrotizing, and **herpes hepatitis** may cause liver failure.

Varicella-Zoster Virus (VZV) Infections

Acute infection with VZV causes chickenpox, and reactivation of latent VZV causes shingles (also called herpes zoster). Chickenpox is mild in children but more severe in adults and in immunocompromised people. Shingles is a source of morbidity in older and immunosuppressed persons. Like HSV, VZV infects mucous membranes, skin, and neurons and causes a self-limited primary infection in immunocompetent individuals. Also like HSV, VZV evades immune responses and establishes a latent infection in sensory ganglia. In contrast to HSV, VZV is transmitted in epidemic fashion by respiratory aerosols, disseminates hematogenously, and causes widespread vesicular skin lesions. Latent VZV infection is seen in neurons and/or satellite cells around neurons in the dorsal root ganglia. Reactivation and clinical recurrences causing shingles are uncommon but may occur many years after the primary infection. Localized recurrence of VZV is most frequent and painful in dermatomes innervated by the trigeminal ganglia, where the virus is most likely to be latent. Shingles rarely recurs in immunocompetent individuals (in only 1% to 4% of infected individuals), but immunosuppressed or older persons can have multiple recurrences. For this reason, vaccination to prevent shingles is now recommended in all patients older than 50 years of age, and in younger adults with chronic disorders that may impair immunity. VZV infection is diagnosed by viral culture, PCR, or detection of viral antigens in cells scraped from superficial lesions.

MORPHOLOGY

The **chickenpox** rash occurs approximately 2 weeks after respiratory infection. Lesions appear in multiple waves centrifugally from the torso to the head and extremities. Each lesion progresses rapidly from a macule to a vesicle, which resembles a dewdrop on a rose petal. On histologic examination, chickenpox lesions show intraepithelial vesicles (Fig. 8.10) with intranuclear inclusions in epithelial cells at the base of the vesicles. After a few days, most chickenpox vesicles rupture, crust over, and heal by regeneration, leaving no scars. However, bacterial superinfection of vesicles that are ruptured by trauma may lead to destruction of the basal epidermal layer and residual scarring.

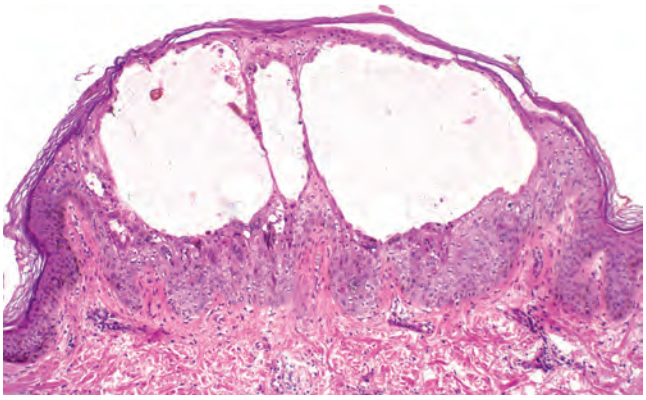


Figure 8.10 Skin lesion of chickenpox (varicella-zoster virus) with intraepithelial vesicle.

Shingles occurs when VZV that has long remained latent in the dorsal root ganglia after a previous chickenpox infection is reactivated and infects sensory nerves that carry it to one or more dermatomes. There, the virus infects keratinocytes and causes vesicular lesions, which, unlike chickenpox, are often associated with intense itching, burning, or sharp pain because of concomitant radiculoneuritis. This pain is especially severe when the trigeminal nerves are involved; rarely, the geniculate nucleus is involved, causing facial paralysis (Ramsay Hunt syndrome). The sensory ganglia contain a dense, predominantly mononuclear infiltrate, with herpetic intranuclear inclusions within neurons and their supporting cells (Fig. 8.11). VZV can also cause interstitial pneumonia, encephalitis, transverse myelitis, vasculopathy, and necrotizing visceral lesions, particularly in immunosuppressed people. Similar to HSV, patients with mutations in TLR3 have an increased risk of VZV encephalitis.

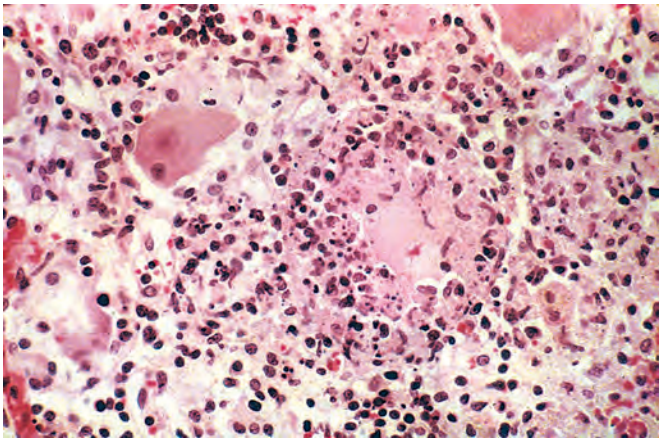


Figure 8.11 Dorsal root ganglion with varicella-zoster virus infection. Note the ganglion cell necrosis and associated inflammation. (Courtesy Dr. James Morris, Radcliffe Infirmary, Oxford, England.)

Cytomegalovirus (CMV) Infections

CMV, a β-group herpesvirus, can produce a variety of disease manifestations, depending on the age of the host, and, more importantly, on the host's immune status. CMV latently infects monocytes and their bone marrow progenitors and can be reactivated when cellular immunity is depressed. CMV causes an asymptomatic or mononucleosis-like infection in healthy individuals but devastating systemic infections in neonates and in immunocompromised people, in whom the virus may infect many different cell types and tissues. As its name implies, CMV-infected cells exhibit gigantism of both the entire cell and its nucleus, which typically contains a large inclusion surrounded by a clear halo ("owl's eye").

Transmission of CMV can occur by several routes, depending on the age group affected. These include the following:

- *Transplacental transmission,* from a newly acquired or primary infection in a mother who does not have protective antibodies (congenital CMV)
- *Neonatal transmission,* through cervical or vaginal secretions at birth, or later through breast milk from a mother who has active infection (perinatal CMV)
- *Transmission through saliva* may occur during preschool years, especially in day care centers. Toddlers so infected readily transmit the virus to their parents
- *Transmission by the genital route* is the dominant mode after about 15 years of age. Spread may also occur via respiratory secretions and the fecal-oral route
- *Iatrogenic transmission* can occur at any age through organ transplant or blood transfusion

Acute CMV infection induces transient but severe immunosuppression. CMV can infect dendritic cells and impair antigen processing and the ability of dendritic cells to stimulate T lymphocytes. Similar to other herpesviruses, CMV can evade immune defenses by down-modulating class I and II MHC molecules and by producing homologues of TNF receptor superfamily members, IL-10, and class I MHC molecules. Interestingly, CMV can evade NK cells by producing ligands that block activating receptors and class I–like proteins that engage inhibitory receptors. Adoptive transfer of CMV-specific T cells has been successfully used to prevent CMV disease after bone marrow transplantation, and the CD8+ T cell response has been thought to be the most important effector response. CD4+ T cells, γδ T cells, and NK cells are known to play a role in immune control of the infection. CMV encodes several Fcγ-binding glycoproteins that effectively operate as adversaries of host Fcγ receptors, inhibiting IgG-mediated immunity. Thus, CMV both hides from and actively suppresses immune responses.

MORPHOLOGY

Infected cells are strikingly enlarged, often to a diameter of 40 μm, and show cellular and nuclear pleomorphism. Prominent *intranuclear basophilic inclusions* spanning one-half of the nuclear diameter are usually set off from the nuclear membrane by a clear halo (Fig. 8.12). Within the cytoplasm of infected cells, smaller basophilic inclusions can also be seen. Parenchymal epithelial cells are infected in glandular organs, as well as neurons in the brain, alveolar macrophages, epithelial cells, and endothelial cells in the lungs, and tubular epithelial and glomerular endothelial cells in the kidneys. CMV causes focal necrosis with minimal inflammation in virtually any organ.

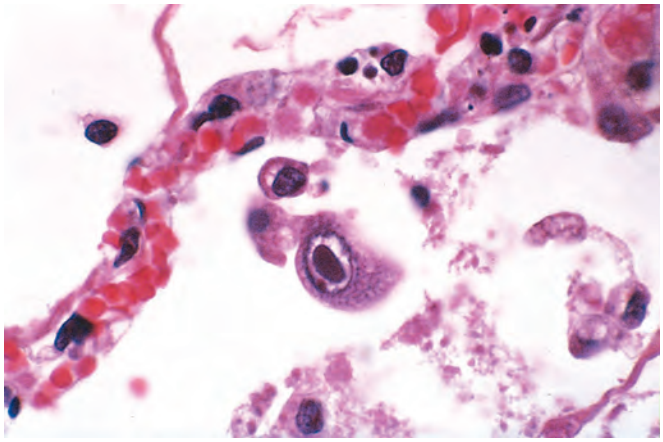


Figure 8.12 Cytomegalovirus: distinct nuclear and ill-defined cytoplasmic inclusions in the lung.

Congenital Infections. Infection acquired in utero may take many forms. In approximately 95% of cases, it is asymptomatic. However, sometimes when the virus is acquired from a mother with primary infection (who does not have protective antibodies), classic cytomegalic inclusion disease develops. Cytomegalic inclusion disease resembles erythroblastosis fetalis. Affected infants may suffer intrauterine growth restriction and present with jaundice, hepatosplenomegaly, anemia, bleeding due to thrombocytopenia, and encephalitis. In fatal cases, the brain is often smaller than normal (microcephaly) and may show foci of calcification. Diagnosis of neonatal CMV is made by viral culture or PCR amplification of viral DNA in urine or saliva.

Infants who survive usually have permanent deficits, including intellectual disability, hearing loss, and other neurologic impairments. The congenital infection is not always devastating, however, and may take the form of interstitial pneumonitis, hepatitis, or a hematologic disorder. Most infants with this milder form of cytomegalic inclusion disease recover, although a few develop intellectual disability later. Uncommonly, a totally asymptomatic infection may be followed months to years later by neurologic sequelae, including delayed-onset intellectual disability and deafness.

Perinatal Infections. Infection acquired during passage through the birth canal or from breast milk is usually asymptomatic due to protective maternal anti-CMV antibodies, which are transmitted to the fetus. Despite the lack of symptoms, many of these infants continue to excrete CMV in their urine or saliva for months to years. Subtle effects on hearing and intelligence later in life have been reported in some studies. Much less commonly, infected infants develop interstitial pneumonitis, failure to thrive, rash, or hepatitis.

CMV Mononucleosis. In healthy young children and adults, the disease is nearly always asymptomatic. In surveys

around the world, 50% to 100% of adults have antibodies to CMV, indicating previous exposure. The most common clinical manifestation of CMV infection in immunocompetent hosts beyond the neonatal period is an infectious mononucleosis-like illness, with fever, atypical lymphocytosis, lymphadenopathy, and hepatitis, marked by hepatomegaly and abnormal liver function tests. The diagnosis is made by serology. Most people recover without any sequelae, but the virus may continue to be excreted in body fluids for months to years. Irrespective of the presence or absence of symptoms, infected individuals remain seropositive for life, and the virus is never cleared, persisting in latently infected leukocytes.

CMV in Immunosuppressed Individuals. Immunocompromised individuals (e.g., transplant recipients and HIV-infected individuals) are susceptible to severe CMV infection, which may be either new infections or reactivation of latent CMV. In the past, CMV was the most common opportunistic viral pathogen in AIDS. Although CMV is still a significant cofactor in HIV disease, especially in Africa, the frequency of serious CMV infection in HIV-positive people has been greatly reduced by antiretroviral treatment. Recipients of solid-organ transplants (heart, liver, kidney) also may contract CMV from the donor organ.

In all of these settings, serious, even life-threatening, disseminated CMV infections in immunosuppressed people primarily affect the lungs (pneumonitis) and gastrointestinal tract (colitis). With pulmonary infection, an interstitial mononuclear infiltrate with foci of necrosis develops, accompanied by the typical enlarged cells with inclusions. The pneumonitis can progress to full-blown acute respiratory distress syndrome. Intestinal necrosis and ulceration can develop and be extensive, leading to the formation of pseudomembranes and debilitating diarrhea. CMV has been associated with vasculopathy and neurocognitive impairment with associated immune markers including the accumulation of multifunctional terminally differentiated αβ T cells, γδ T cells, and NK cells. Diagnosis of CMV infection is made by demonstration of characteristic morphologic alterations in tissue sections, viral culture, rising antiviral antibody titer, and PCR-based detection of CMV DNA. Quantitative PCR-based assays have revolutionized the approach to monitoring CMV infection in people after transplantation.

Chronic Productive Infections

In some infections the immune system is unable to eliminate the virus, and continued viral replication leads to persistent viremia. The high mutation rate of viruses such as HIV and HBV may contribute to their escape from control by the immune system. HIV and HBV infection are described in Chapters 6 and 18, respectively.

Transforming Viral Infections

Some viruses can transform infected cells into benign or malignant tumor cells. Oncogenic viruses can stimulate cell growth and survival by a variety of mechanisms. Several viruses have been implicated in the causation of human cancer, including EBV, HPV, HBV, and HTLV-1 (discussed in Chapters 7, 13, 18, and 22).

Epstein-Barr Virus (EBV) Infections

EBV causes infectious mononucleosis, a benign, self-limited lymphoproliferative disorder, and is associated with the pathogenesis of several human tumors, most commonly certain lymphomas and nasopharyngeal carcinoma. Only infectious mononucleosis is discussed here.

Infectious mononucleosis is characterized by fever, sore throat, generalized lymphadenopathy, splenomegaly, and the appearance in the blood of atypical activated T lymphocytes (mononucleosis cells). Some people develop hepatitis, meningoencephalitis, and pneumonitis. Infectious mononucleosis occurs principally in late adolescents or young adults among upper socioeconomic classes in higher-income nations. In the rest of the world, primary infection with EBV occurs in childhood and is usually asymptomatic.

Pathogenesis

EBV is transmitted by close human contact, frequently through the saliva during kissing. It is not known whether the source of the virus is B cells, oropharyngeal epithelial cells, or both. EBV infects B cells and possibly epithelial cells of the oropharynx. It has been hypothesized that EBV initially infects oropharyngeal epithelial cells and then spreads to underlying lymphoid tissue (tonsils and adenoids), where mature B cells are infected. Of note, people with X-linked agammaglobulinemia, who lack B cells, do not become latently infected with EBV or shed virus, suggesting that B cells are the main reservoir of infection. An EBV envelope glycoprotein binds CD21 (CR2), the receptor for the C3d component of complement, which is present on B cells. Infection of B cells may take one of two forms. In a minority of B cells, infection is lytic, leading to viral replication and eventual cell lysis accompanied by release of virions, which may infect other B cells. In most B cells, however, EBV establishes latent infection, during which the virus persists as an extrachromosomal episome.

B cells that are latently infected with EBV are activated and begin to proliferate and to disseminate. This uncontrolled, expanding polyclonal population of EBV-infected B cells secretes antibodies with many specificities, including antibodies that recognize sheep or horse red cells. These so-called *heterophile antibodies* are detected in diagnostic tests for mononucleosis. EBV-infected B cells may also produce autoantibodies, for example against platelets, leading to transient immune-mediated thrombocytopenia in a small subset of patients with mononucleosis.

The symptoms of infectious mononucleosis appear on initiation of the host immune response. Cellular immunity mediated by CD8+ CTLs and NK cells is the most important component of this response. The *atypical lymphocytes* seen in the blood, characteristic of this disease, are mainly EBV-specific CD8+ CTLs, but also include NK cells. The reactive proliferation of T cells is largely centered in lymphoid tissues, which accounts for the lymphadenopathy and splenomegaly. Early in the course of the infection, IgM antibodies are formed against viral capsid antigens; later, IgG antibodies are formed that persist for life. In otherwise healthy persons, the fully developed humoral and cellular responses to EBV act as brakes on viral shedding, resulting in the elimination of B cells expressing the full complement of EBV latency-associated genes. In hosts with acquired defects

in cellular immunity (e.g., AIDS, organ transplantation), reactivation of EBV can lead to B-cell proliferation, which can progress through a multistep process to EBV-associated B-cell lymphomas. EBV also contributes to the development of some cases of Burkitt lymphoma (Chapters 7 and 13), in which a chromosomal translocation (most commonly an 8:14 translocation) involving the *MYC* oncogene is the critical oncogenic event.

MORPHOLOGY

The major alterations involve the blood, lymph nodes, spleen, liver, CNS, and, occasionally, other organs. The **peripheral blood** shows absolute lymphocytosis; more than 60% of white blood cells are lymphocytes. Between 5% and 80% of these are large, **atypical lymphocytes,** 12 to 16 μm in diameter, characterized by an abundant cytoplasm containing multiple clear vacuolations; an oval, indented, or folded nucleus; and scattered cytoplasmic azurophilic granules (Fig. 8.13). These atypical lymphocytes, most of which express CD8, are sufficiently distinctive to strongly suggest the diagnosis.

The **lymph nodes** are typically discrete and enlarged throughout the body, particularly in the posterior cervical, axillary, and inguinal regions. On histologic examination, the most striking feature is the expansion of paracortical areas due to activation of T cells (immunoblasts). A minor population of EBV-infected B cells expressing Epstein-Barr nuclear antigen 2 *(EBNA2)*, latent membrane protein 1 *(LMP1)*, and other latency-specific genes can also be detected in the paracortex using specific antibodies. EBV-infected B cells resembling Reed-Sternberg cells (the malignant cells of Hodgkin lymphoma, Chapter 13) may be found. B-cell areas (follicles) may also show mild hyperplasia. The T-cell proliferation is sometimes so exuberant that it is difficult to distinguish the nodal morphology from that seen in malignant lymphomas. Similar changes commonly occur in the tonsils and lymphoid tissue of the oropharynx.

The **spleen** is enlarged in most cases, weighing between 300 and 500 g. It is usually soft and fleshy, with a hyperemic cut surface. The histologic changes are analogous to those of the lymph nodes, showing an expansion of white pulp follicles and red pulp sinusoids due to the presence of numerous activated T cells. These spleens are especially vulnerable to rupture, possibly in part because the rapid increase in size produces a tense, fragile splenic capsule.

The **liver** is usually involved to some degree, although hepatomegaly is at most moderate. On histologic examination, atypical lymphocytes are seen in the portal areas and sinusoids, and scattered, isolated cells or foci of parenchymal necrosis may be present. This histologic picture is similar to that of other forms of viral hepatitis.

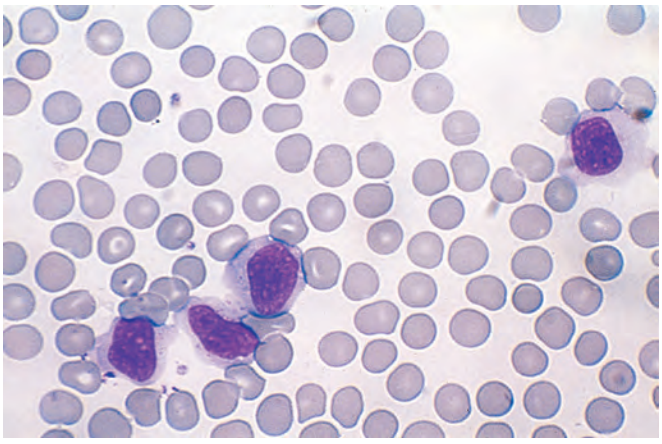


Figure 8.13 Atypical lymphocytes in infectious mononucleosis.

Clinical Features

EBV infection in young children classically presents with fever, sore throat, lymphadenitis, and the other features mentioned earlier. However, malaise, fatigue, and lymphadenopathy are the common presentation in young adults with infectious mononucleosis and can raise the specter of leukemia or lymphoma; EBV also can cause fever of unknown origin without significant lymphadenopathy or other localized findings, hepatitis resembling one of the hepatotropic viral syndromes, or a febrile rash resembling rubella. The diagnosis depends on the following findings (in increasing order of specificity): (1) lymphocytosis with the characteristic atypical lymphocytes in the peripheral blood, (2) a positive heterophile antibody reaction (Monospot test), and (3) a rising titer of specific antibodies for EBV antigens (viral capsid antigens, early antigens, or EBNA). In most patients, infectious mononucleosis resolves within 4 to 6 weeks, but sometimes the fatigue lasts longer. One or more complications occasionally supervene. Perhaps most common is marked hepatic dysfunction with jaundice, elevated hepatic enzyme levels, disturbed appetite, and, rarely, even liver failure. Other complications involve the nervous system, kidneys, bone marrow, lungs, eyes, heart, and spleen. Splenic rupture can occur even with minor trauma, leading to hemorrhage that may be fatal. Recent studies have found a correlation between the severity of infectious mononucleosis and the T-cell receptor repertoire of the host, leading to a hypothesis that heterologous immunity might explain variability in disease outcome.

A more serious complication in those lacking T-cell immunity, such as HIV-infected individuals and individuals receiving immunosuppressive therapy (e.g., bone marrow or solid-organ transplant recipients), is unimpeded B-cell proliferation. This process can be initiated by an acute infection or reactivation of latent B-cell infection and usually begins as polyclonal B-cell proliferation that transforms to monoclonal B-cell lymphoma.

Serious consequences of EBV infection occur in individuals suffering from the X-linked lymphoproliferation disease (also known as *Duncan disease*), a disorder caused by mutations in the *SH2D1A* gene, which encodes a signaling protein that participates in T-cell and NK-cell activation and antibody production. This rare inherited immunodeficiency is characterized by an ineffective immune response to EBV. Patients are usually normal until they are acutely infected with EBV, often during adolescence. In more than half the cases, EBV causes an acute overwhelming infection that may be fatal. Others succumb to EBV-positive B-cell lymphoma or infections related to hypogammaglobulinemia.

BACTERIAL INFECTIONS

Different classes of bacteria are responsible for diverse infections (Table 8.5). Selected examples of the most common or clinically significant infections are discussed next.

Gram-Positive Bacterial Infections

Common gram-positive cocci include *Staphylococcus* spp., *Streptococcus* spp., and *Enterococcus* spp., each of which causes many types of infections. Diphtheria, listeriosis, anthrax, and nocardiosis are less common infections caused by gram-positive rods.

Table 8.5 Selected Human Bacterial Pathogens and Associated Diseases

Organ System	Species	Frequent Disease Presentations
Respiratory	*Streptococcus pyogenes*	Pharyngitis
	Corynebacterium diphtheriae	Diphtheria
	Bordetella pertussis	Pertussis
	Streptococcus pneumonia	Lobar pneumonia
	Mycobacterium tuberculosis	Tuberculosis
	Legionella pneumophila	Legionnaires' disease
Gastrointestinal	*Helicobacter pylori*	Peptic ulcers
	Vibrio cholerae, enterotoxigenic *Escherichia coli*	Noninflammatory gastroenteritis
	Shigella spp., *Salmonella* sp. *Campylobacter jejuni*, enterohemorrhagic *Escherichia coli*	Inflammatory gastroenteritis
	Salmonella serotype *typhi*	Enteric (typhoid) fever
	Clostridioides difficile	Pseudomembranous colitis
Nervous system	*Neisseria meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Listeria monocytogenes*	Acute meningitis
	Clostridium tetani, *Clostridium botulinum*	Paralytic intoxications, tetanus, and botulism
Urogenital	*Escherichia coli*, *Pseudomonas aeruginosa*, *Enterococcus* spp.	Urinary tract infections
	Neisseria gonorrhoeae	Gonorrhea
	Chlamydia trachomatis	Chlamydia
	Treponema pallidum	Syphilis
Skin and adjacent soft tissue	*Staphylococcus aureus*	Abscess, cellulitis
	Streptococcus pyogenes	Impetigo, erysipelas, necrotizing fasciitis
	Clostridium perfringens	Gas gangrene
	Bacillus anthracis	Cutaneous anthrax
	Pseudomonas aeruginosa	Burn infections
	Mycobacterium leprae	Leprosy
Disseminated infections	*Yersinia pestis*	Plague
	Borrelia burgdorferi	Lyme disease
	Brucella spp.	Brucellosis (undulant fever)
Disseminated neonatal infection	*Streptococcus agalactiae*, *Listeria monocytogenes*	Neonatal bacteremia, meningitis
	Treponema pallidum	Congenital syphilis

Staphylococcal Infections

***S. aureus* causes a myriad of skin lesions (boils, carbuncles, impetigo, and scalded-skin syndrome) as well as abscesses, sepsis, osteomyelitis, pneumonia, endocarditis, food poisoning, and toxic shock syndrome** (Fig. 8.14). There are more than 1 million infections in the United States alone, with 20,000 attributable deaths per year. *S. aureus* is a pyogenic gram-positive coccus that forms clusters resembling bunches of grapes. The general characteristics of *S. aureus* infection are reviewed in the following sections. Infections of specific organs are described in other chapters. Coagulase-negative staphylococci, such as *S. epidermidis*, characteristically cause opportunistic infections in catheterized patients, patients with prosthetic cardiac valves, and intravenous drug users. *S. saprophyticus* is a common cause of urinary tract infection in young women.

Pathogenesis

S. aureus produces a multitude of virulence factors, which include surface proteins involved in adherence and evasion of the host immune response, secreted enzymes that degrade host structures, secreted toxins that damage host cells, and proteins that cause antibiotic resistance. *S. aureus* expresses surface receptors for fibrinogen (called clumping factor), fibronectin, and vitronectin and uses these molecules to bind to host endothelial cells. *S. aureus* produces a polysaccharide capsule, enabling attachment to artificial materials and resulting in significant prosthetic valve and catheter-associated infection and a resistance to host cell phagocytosis.

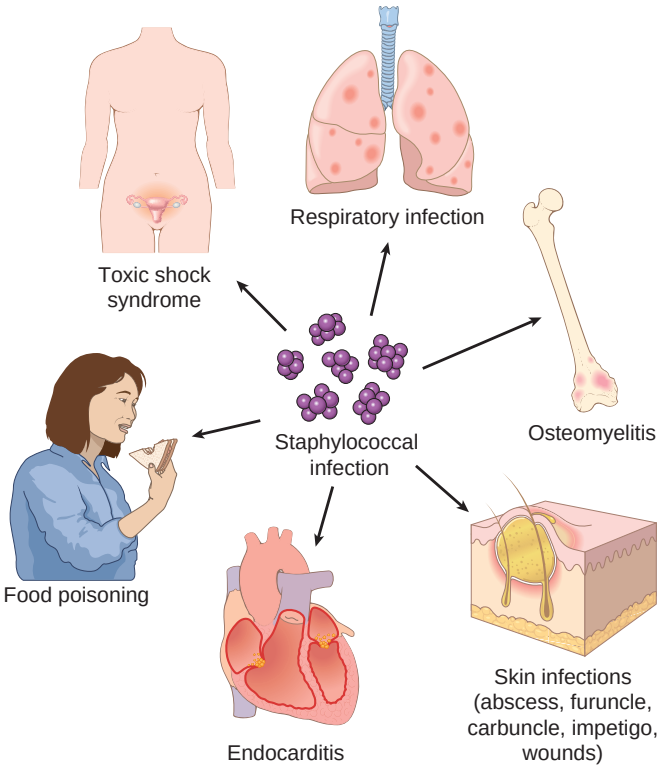


Figure 8.14 The many consequences of staphylococcal infection.

S. aureus also expresses surface protein A, which binds the Fc portion of immunoglobulins, allowing the organism to escape antibody-mediated killing.

***S. aureus* Toxins.** *S. aureus* produces multiple *membrane-damaging (hemolytic) toxins*. These include α-toxin, a protein that intercalates into the plasma membrane of host cells, forming pores that allow toxic levels of calcium to leak into cells; β-toxin, a sphingomyelinase; and δ-toxin, which is a detergent-like peptide. *S. aureus* γ-toxin and leukocidin lyse red cells and phagocytes, respectively.

The *exfoliative A and B toxins* produced by *S. aureus* are serine proteases that cleave the desmosomal protein desmoglein 1, which holds epidermal cells tightly together. This causes keratinocytes to detach from one another and from the underlying basement membrane, resulting in a loss of barrier function that often leads to secondary skin infections. Exfoliation may occur locally at the site of infection (bullous impetigo) or may result in widespread loss of the superficial epidermis (staphylococcal scalded-skin syndrome).

Superantigens produced by *S. aureus* cause food poisoning and toxic shock syndrome (TSS). TSS came to public attention because of its association with the use of hyperabsorbent tampons, which became colonized with *S. aureus* during use. It is now clear that TSS can be caused by growth of *S. aureus* at many sites, most commonly the vagina and infected surgical sites. This syndrome is characterized by hypotension (shock), renal failure, coagulopathy, liver disease, respiratory distress, a generalized erythematous rash, and soft-tissue necrosis at the site of infection. If not promptly treated, it can be fatal. TSS can also be caused by *S. pyogenes*. Bacterial superantigens cause polyclonal T-cell proliferation by binding to conserved portions of MHC molecules and to relatively conserved portions of T-cell receptor β chains. In this manner, superantigens may stimulate up to 20% of T lymphocytes, leading to release of cytokines such as TNF and IL-1, in such large amounts that they may trigger the systemic inflammatory response syndrome (Chapter 4). Superantigens produced by *S. aureus* also cause vomiting when ingested in food, presumably by affecting the CNS or the enteric nervous system.

Antibiotic resistance is a growing problem in treatment of *S. aureus* infections. Methicillin-resistant *S. aureus* (MRSA) are resistant to nearly all penicillin and cephalosporin antibiotics. Previously, MRSA was mainly found in health care facilities, but community-acquired MRSA infections are now common. As a result, empirical treatment of *S. aureus* infections with cephalosporin antibiotics is not recommended.

MORPHOLOGY

Whether the lesion is located in the skin, lungs, bones, or heart valves, ***S. aureus* causes pyogenic inflammation that is distinctive for its local destruction of host tissue.**

Excluding impetigo, which is restricted to the superficial epidermis, staphylococcal skin infections are centered around the hair follicles where they begin. A **furuncle,** or **boil,** is a focal suppurative inflammation of the skin and subcutaneous tissue. They may be solitary, or multiple, or recur in successive crops. Furuncles are most frequent in moist, hairy areas, such as the face, axillae, groin, legs, and submammary folds. Beginning in a single hair follicle, a boil develops into a growing and deepening abscess that eventually "comes to a head" by thinning and rupturing the overlying skin. A **carbuncle** is a deeper suppurative infection that spreads laterally beneath the deep subcutaneous fascia and then burrows superficially to erupt in multiple adjacent skin sinuses. Carbuncles typically appear beneath the skin of the upper back and posterior neck, where fascial planes favor their spread. **Hidradenitis** is chronic suppurative infection of apocrine glands, most often in the axilla. Infections of the nail bed **(paronychia)** or on the palmar side of the fingertips **(felons)** are exquisitely painful. They may follow trauma or embedded splinters and, if deep enough, destroy the bone of the terminal phalanx or detach the fingernail.

S. aureus lung infections (Fig. 8.15) have a polymorphonuclear infiltrate similar to that of *S. pneumoniae* infections (see Fig. 3.18B), but they cause much more tissue destruction. Lung infections usually arise from a hematogenous source, such as an infected thrombus, or in the setting of a predisposing condition such as influenza.

Staphylococcal scalded-skin syndrome, also called Ritter disease, most frequently occurs in children with *S. aureus* infection of the nasopharynx or skin. There is a sunburn-like rash that spreads over the entire body and evolves into fragile bullae that lead to partial or total skin loss. The desquamation of the epidermis in staphylococcal scalded-skin syndrome occurs at the level of the granulosa layer, distinguishing it from toxic epidermal necrolysis, or Lyell disease, which is secondary to drug hypersensitivity and causes desquamation at the level of the epidermal-dermal junction (Chapter 25).

Streptococcal and Enterococcal Infections

Streptococci cause suppurative infections of the skin, oropharynx, lungs, and heart valves. They are also responsible for a number of postinfectious syndromes, including rheumatic fever (Chapter 12), poststreptococcal glomerulonephritis (Chapter 20), and erythema nodosum (Chapter 25). These bacteria are gram-positive cocci that grow in pairs or chains. β-hemolytic streptococci are typed according to their

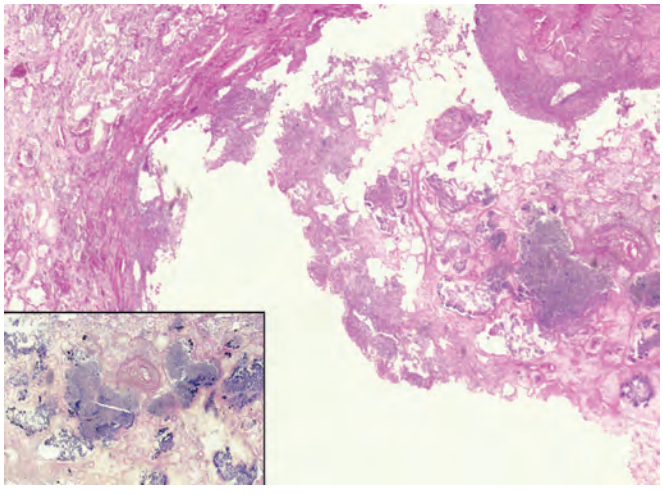


Figure 8.15 Staphylococcal abscess of the lung with extensive neutrophilic infiltrate and destruction of the alveoli (contrast with Fig. 3.18B). The *inset* shows the same area on Gram stain highlighting clusters of bacteria.

surface carbohydrate antigens. *S. pyogenes* (group A) causes pharyngitis, scarlet fever, erysipelas, impetigo, rheumatic fever, TSS, and glomerulonephritis. *Streptococcus agalactiae* (group B) colonizes the female genital tract and causes sepsis and meningitis in neonates and chorioamnionitis in pregnancy. *S. pneumoniae,* the most important α-hemolytic streptococcus, is a common cause of community-acquired pneumonia in older adults and meningitis in children and adults. The viridans-group streptococci include α-hemolytic and nonhemolytic streptococci found in normal oral microbiota that are a common cause of endocarditis. *Streptococcus mutans* is the major cause of dental caries. Streptococcal infections are diagnosed by culture, and, in those with pharyngitis, by the rapid streptococcal antigen test.

Enterococci are gram-positive cocci that also grow in pairs and chains and are hence difficult to distinguish from streptococci by morphology alone. They are often resistant to commonly used antibiotics and are a significant cause of endocarditis and urinary tract infection.

Pathogenesis

The different species of streptococci produce many virulence factors and toxins. *S. pyogenes, S. agalactiae,* and *S. pneumoniae* have capsules that resist phagocytosis. *S. pyogenes* also expresses M protein, a surface protein that prevents bacteria from being phagocytosed, and a complement C5a peptidase that degrades this chemotactic peptide. *S. pyogenes* secretes a phage-encoded pyrogenic exotoxin that causes fever and rash in scarlet fever. Rheumatic fever is probably caused by antistreptococcal M protein antibodies and T cells that cross-react with cardiac proteins. Virulent *S. pyogenes* have been called flesh-eating bacteria because they cause a rapidly progressive necrotizing fasciitis. Although the antiphagocytic capsule is the most important virulence factor of *S. pneumoniae,* this organism also produces pneumolysin, a toxin that inserts into host cell membranes and lyses cells, greatly increasing tissue damage. *S. mutans* produces caries by metabolizing sucrose to lactic acid (which causes demineralization of tooth enamel) and by secreting high-molecular-weight glucans that promote aggregation of bacteria and plaque formation.

Enterococci are low-virulence bacteria, although they do have an antiphagocytic capsule and produce enzymes that injure host tissues. The emergence of enterococci as pathogens is primarily due to their resistance to antibiotics, and incidence is higher in immunosuppressed patient populations, such as those undergoing organ or stem cell transplantation, who receive frequent antimicrobial agents and can have altered commensal microbiota.

MORPHOLOGY

Streptococcal infections are characterized by diffuse interstitial neutrophilic infiltrates with minimal destruction of host tissues. The skin lesions caused by streptococci (furuncles, carbuncles, and impetigo) resemble those of staphylococci.

Erysipelas is caused by exotoxins from superficial infection with *S. pyogenes*. It is characterized by rapidly spreading erythematous cutaneous swelling that may begin on the face or, less frequently, on the body or an extremity. The rash has a sharp, well-demarcated, serpiginous border and may form a "butterfly" distribution on the face (Fig. 8.16). On histologic examination, there is a diffuse, edematous, neutrophilic inflammatory reaction in the dermis and epidermis extending into the subcutaneous tissues. Microabscesses may be formed, but tissue necrosis is usually minor.

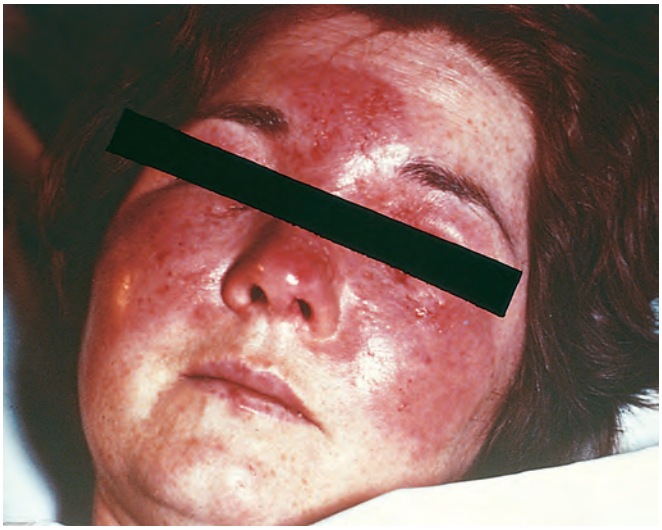


Figure 8.16 Streptococcal erysipelas.

Streptococcal pharyngitis, which is the major antecedent of poststreptococcal glomerulonephritis (Chapter 20), is marked by edema, epiglottic swelling, and punctate abscesses of the tonsillar crypts, sometimes accompanied by cervical lymphadenopathy. Swelling associated with severe pharyngeal infection may encroach on the airways, especially if there is peritonsillar or retropharyngeal abscess formation.

Scarlet fever, associated with pharyngitis caused by *S. pyogenes,* is most common between 3 and 15 years of age. It is manifested by a punctate erythematous rash that is most prominent over the trunk and inner aspects of the arms and legs. The face is also involved, but usually a small area about the mouth remains relatively unaffected, producing circumoral pallor. The skin usually becomes hyperkeratotic and scaly during defervescence.

S. pneumoniae is an important cause of lobar pneumonia (Chapter 15).

Diphtheria

Diphtheria is caused by *Corynebacterium diphtheriae,* a slender gram-positive rod with clubbed ends that spreads from person to person in respiratory droplets or skin exudate. Fewer than five cases of diphtheria were reported to the Centers for Disease Control and Prevention (CDC) in the United States in the past decade, but annually more than 7000 cases worldwide have been reported to the WHO. Respiratory diphtheria causes pharyngeal or, less often, nasal or laryngeal infection. Damage to the heart, nerves, and other organs may be present. Cutaneous diphtheria causes chronic ulcers with a dirty gray membrane, but does not cause systemic damage. *C. diphtheriae* produces a phage-encoded A-B toxin that blocks host cell protein synthesis. The A fragment does this by catalyzing the covalent transfer of adenosine diphosphate (ADP)-ribose to elongation factor-2 (EF-2). This inhibits EF-2 function, which is required for the translation of mRNA into protein. A single molecule of

diphtheria toxin can kill a cell by ADP-ribosylating, and thereby inactivating, more than a million EF-2 molecules. Immunization with diphtheria toxoid (formalin-fixed toxin) stimulates production of toxin-neutralizing antibodies that protect people from the lethal effects of the toxin.

MORPHOLOGY

Inhaled *C. diphtheriae* carried in respiratory droplets proliferate at the site of attachment on the mucosa of the nasopharynx, oropharynx, larynx, or trachea. The bacteria also form satellite lesions in the esophagus or lower airways. Release of exotoxin causes necrosis of the epithelium, accompanied by an outpouring of a dense fibrinosuppurative exudate. The coagulation of this exudate on the ulcerated necrotic surface creates a tough, dirty, gray to black superficial membrane, sometimes called **pseudomembrane** because it is not formed by viable tissue (Fig. 8.17). There is an intense neutrophilic infiltrate in the underlying tissues with marked vascular congestion, interstitial edema, and fibrin exudation. When the membrane sloughs off its inflamed and vascularized bed, bleeding and asphyxiation may occur. With control of the infection, the membrane is coughed up or removed by enzymatic digestion, and the inflammatory reaction subsides.

Although the bacterial invasion remains localized, with entry of exotoxin into the blood and its systemic distribution, there may be fatty change in the myocardium with isolated myofiber necrosis, polyneuritis with degeneration of the myelin sheaths and axis cylinders, and (less commonly) fatty change and focal necroses of parenchymal cells in the liver, kidneys, and adrenals.

Listeriosis

***Listeria monocytogenes* is a gram-positive bacillus that causes gastroenteritis in most individuals who ingest it in sufficient quantity, and severe food-borne infections in vulnerable hosts.** Outbreaks of *L. monocytogenes* infection have been linked to many foods, but most are associated with contaminated dairy products or processed fruits and vegetables. In the United States, all clinical isolates of *L. monocytogenes* have been typed for epidemiologic surveillance since 1998, resulting in a fivefold increase in the number of outbreaks detected. Pregnant women, neonates, older adults, and immunosuppressed persons are particularly susceptible to severe *L. monocytogenes* infection. In pregnant women, *L. monocytogenes* causes an amnionitis that may result in abortion, stillbirth, or neonatal sepsis. In neonates and immunosuppressed adults, it can cause disseminated disease (granulomatosis infantiseptica of the newborn) and an exudative meningitis.

L. monocytogenes is a facultative intracellular pathogen. The bacteria bind to receptors on host epithelial cells and macrophages and are phagocytosed. The bacteria escape from the phagolysosome using a pore-forming protein, listeriolysin O, and two phospholipases. Listeriolysin O also affects membrane bound organelles, including rounding and shrinking of mitochondria, inhibition of the endoplasmic reticulum, and damage to the lysosomal membrane. In the host cell cytoplasm, Act A, a bacterial surface protein, binds to the Arp2/3 complex, an actin nucleating complex, which induces actin polymerization. This generates force to propel the bacteria into adjacent, uninfected host cells. Resting macrophages fail to kill the intracellular bacteria, whereas macrophages that are activated by IFN-γ can. Accordingly, an effective host response to *L. monocytogenes* depends on IFN-γ produced by NK cells early in the course of the infection and Th1 and CD8+ T cells in chronic infection. Patients with defects in cell-mediated immunity, such as those with reduced levels of CD4+ lymphocytes, are at increased risk for listeriosis.

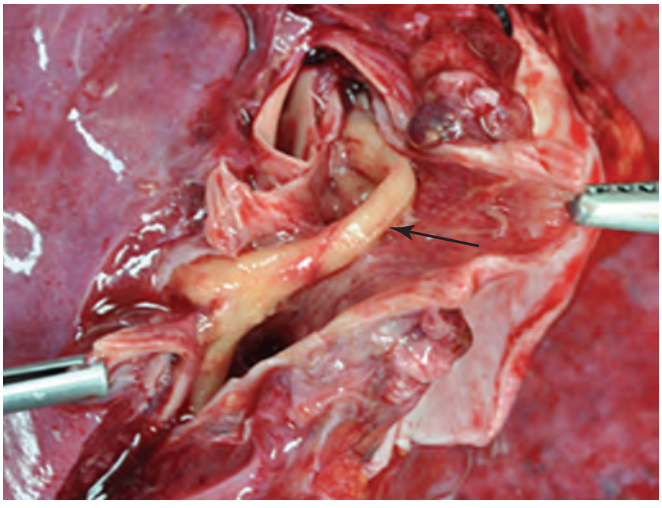


Figure 8.17 Membrane of diphtheria *(arrow)* lying within a transverse bronchus. (Courtesy Dr. Robin A. Cooke, Department of Anatomical Pathology, Princess Alexandria Hospital, Brisbane, Australia.)

MORPHOLOGY

In acute infections, *L. monocytogenes* evokes an exudative pattern of inflammation with numerous neutrophils. The meningitis due to *L. monocytogenes* is macroscopically and microscopically indistinguishable from meningitis due to other pyogenic bacteria (Chapter 28). **The finding of gram-positive bacilli in the CSF is virtually diagnostic.** More varied lesions may be encountered in neonates and immunosuppressed adults. Focal abscesses alternating with grayish or yellow nodules representing necrotic amorphous tissue debris can occur in any organ, including the lung, liver, spleen, and lymph nodes. In infections of longer duration, macrophages appear in large numbers, but granulomas are rare. Infants born with *L. monocytogenes* sepsis often have a papular red rash over the extremities, and listerial abscesses can be seen in the placenta. A smear of the meconium will disclose the gram-positive bacilli.

Anthrax

Anthrax is characterized by necrotizing inflammatory lesions in the skin or gastrointestinal tract or systemically. It is caused by *Bacillus anthracis,* a large, spore-forming gram-positive rod-shaped bacterium found in environmental sources. Livestock become infected by spores in their environment or feed. Humans usually become infected by eating or handling meat or products from infected animals (e.g., wool or hides). There are a small number of cases of anthrax each year, most of which occur in lower-income nations. Anthrax spores can be made into a fine powder, creating a potent biologic weapon that is a potential bioterrorism threat. In 1979, accidental release of *B. anthracis* spores

at a military research institute in Russia killed 66 people. In 2001, 22 people in the United States were infected with *B. anthracis*, mostly through domestic bioterrorism with spores delivered in the mail.

There are three major forms of anthrax:

- *Cutaneous anthrax,* which makes up 95% of naturally occurring infections, begins as a painless, pruritic papule that develops into a vesicle within 2 days. As the vesicle enlarges, striking edema may occur around it, with development of regional lymphadenopathy. After the vesicle ruptures, the remaining ulcer becomes covered with a characteristic black eschar, which dries and falls off as the person recovers. Bacteremia is rare.
- *Inhalational anthrax* occurs when airborne spores are inhaled. The spores are carried by phagocytes to lymph nodes where they germinate, producing bacilli that release toxins that cause hemorrhagic mediastinitis. After a prodromal illness of 1 to 6 days characterized by fever, cough, and chest or abdominal pain, there is abrupt onset of increased fever, hypoxia, and sweating. Frequently, meningitis develops from bacteremia. Inhalational anthrax rapidly leads to shock and frequently death within 1 to 2 days.
- *Gastrointestinal anthrax* is usually contracted by eating undercooked meat contaminated with *B. anthracis*. Initially, the person has nausea, abdominal pain, and vomiting, followed by severe bloody diarrhea and, sometimes, bacteremia. Mortality is approximately 40%.

Pathogenesis

B. anthracis produces potent toxins and an antiphagocytic polyglutamyl capsule. The mechanisms of action of anthrax toxins are well understood (Fig. 8.18). There are two A subunits and one B subunit. The two A subunits, *edema factor* (EF) and *lethal factor* (LF), are named for their effects in experimental animals. The B subunit is called the *protective antigen* (PA) because antibodies against it protect against the effects of the toxins. PA is not toxic, but it serves to deliver the toxic EF and LF into cells. The bacterium releases each subunit as a separate protein. PA binds to a cell surface receptor that is highly expressed on endothelial cells. Then a host protease removes a fragment of the PA, and the remaining fragment self-associates to form a heptamer. One to three molecules of the EF or LF bind to a PA heptamer, and this complex is endocytosed into the host cell. The low pH of the endosome causes a conformational change in the PA heptamer, which then forms a channel in the endosome membrane through which the EF or LF moves into the cytoplasm. In the cytoplasm, EF binds to calcium and calmodulin to form an adenylate cyclase. The active enzyme converts ATP to intracellular cyclic adenosine monophosphate (cAMP), altering cell function. LF has a different mechanism of action. LF is a protease that destroys mitogen-activated protein kinase kinases (MAPKKs). These kinases regulate the activity of MAPKs, which are important regulators of cell growth and differentiation (Chapter 1). The mechanism of cell death caused by dysregulation of MAPKs is not understood.

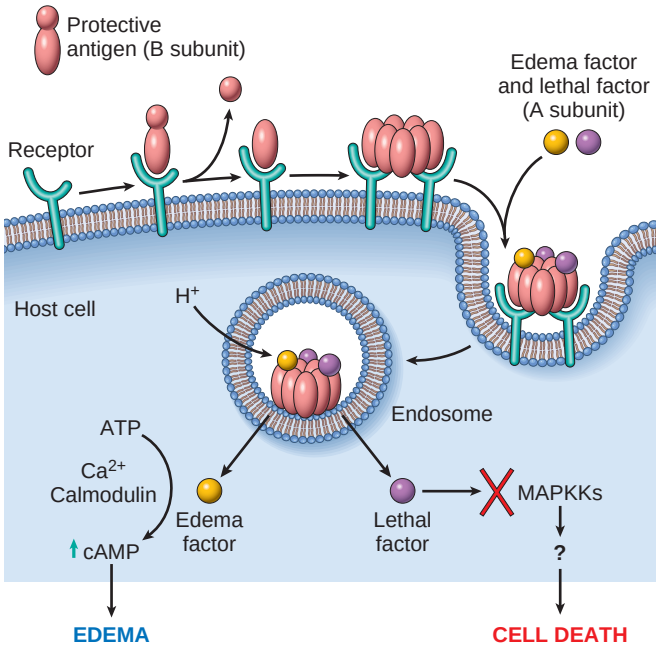


Figure 8.18 Mechanism of action of anthrax toxins. Note that each cluster of B subunits binds either the edema factor or the lethal factor, but not both (as shown for simplicity). (Modified from Mourez M, Lacy DB, Cunningham K, et al: 2001: A year of major advances in anthrax toxin research, *Trends Microbiol* 10:287–293, 2002.)

MORPHOLOGY

Anthrax lesions at any site are typified by necrosis and exudative inflammation rich in neutrophils and macrophages. The presence of large, boxcar-shaped gram-positive extracellular bacteria in chains, seen histopathologically using the Brown and Brenn stain or grown in culture, suggests the diagnosis.

Inhalational anthrax causes numerous foci of hemorrhage in the mediastinum and hemorrhagic lymphadenitis of hilar and peribronchial lymph nodes. The lungs typically show a perihilar interstitial pneumonia with infiltration of macrophages and neutrophils and pulmonary vasculitis. Hemorrhagic lung lesions associated with vasculitis are also present in about one-half of cases. Mediastinal lymph nodes are expanded by edema and by macrophages containing phagocytosed apoptotic lymphocytes. *B. anthracis* is most likely to be seen in the alveolar capillaries and venules and, to a lesser degree, within the alveolar space and draining hilar lymph nodes (Fig. 8.19). In fatal cases, however, the organism may be found in multiple organs (spleen, liver, intestines, kidneys, adrenal glands, and meninges).

Nocardial Infections

***Nocardia* spp. are aerobic gram-positive bacteria found in soil that cause opportunistic infections.** The organism grows in distinctive branched chains. In culture, *Nocardia* form thin aerial filaments resembling hyphae. Despite this morphologic similarity to molds, *Nocardia* are true bacteria.

Nocardia spp. cause respiratory infections, most often in patients with defects in immunity due to prolonged steroid use, HIV infection, or diabetes. Respiratory infection with *Nocardia* spp. manifests as an indolent illness with fever, weight loss, and cough, which may be mistaken for tuberculosis or malignancy. In some cases, *Nocardia* spp. infections disseminate from the lungs to the CNS. Infections of the

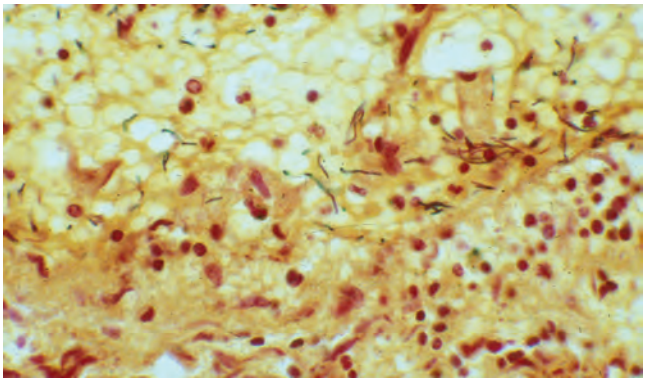


Figure 8.19 Bacillus anthracis in the subcapsular sinus of a hilar lymph node of a patient who died of inhalational anthrax. (Courtesy Dr. Lev Grinberg, Department of Pathology, Hospital 40, Ekaterinburg, Russia, and Dr. David Walker, UTMB Center for Biodefense and Emerging Infectious Diseases, Galveston, Tex.)

CNS are also indolent and cause varying neurologic deficits depending on the site of the lesions. *N. brasiliensis* causes skin infections following injuries contaminated with soil. Manifestations include cellulitis, lymphocutaneous disease, and actinomycetoma with formation of nodules that progress to form chronic draining fistulae.

MORPHOLOGY

Nocardia spp. appear in tissue as slender gram-positive organisms arranged in branching filaments (Fig. 8.20). Irregular staining gives the filaments a beaded appearance. *Nocardia* spp. stain with modified acid-fast stains (Fite-Faraco stain), unlike *Actinomyces* spp., which may appear similar on Gram stain of tissue. *Nocardia* spp. elicit a suppurative response with central liquefaction and surrounding granulation and fibrosis. Granulomas do not form.

Gram-Negative Bacterial Infections

There are a large number of gram-negative bacterial pathogens. Many gram-negative bacteria are increasingly resistant to antibiotics, including carbapenem-resistant *K. pneumoniae* and cephalosporin-resistant *N. gonorrhoeae*. Only a few gram-negative bacteria are discussed in this section. A number of important gram-negative pathogens are discussed in the appropriate chapters of organ systems, including bacterial causes of gastrointestinal infections and urinary tract infections. Anaerobic gram-negative organisms are considered later in this chapter. Gram-negative bacterial infections are usually diagnosed by culture.

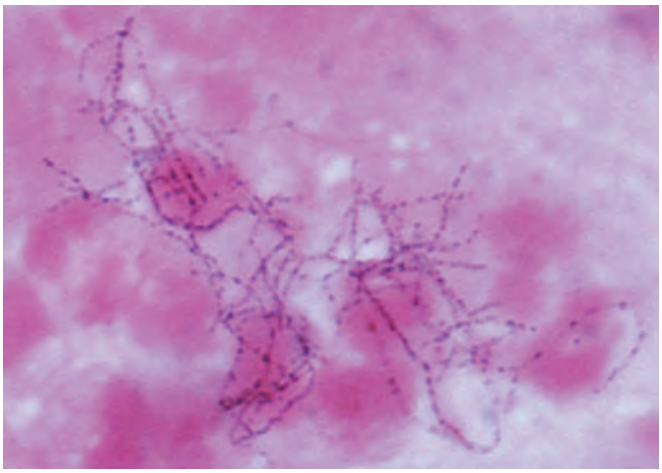


Figure 8.20 *Nocardia asteroides* in a Gram-stained sputum sample. Note the beaded, branched gram-positive organisms and leukocytes. (Courtesy Dr. Ellen Jo Baron, Stanford University Medical Center, Stanford, Calif.)

Neisserial Infections

Neisseria spp. are gram-negative diplococci that are flattened on the adjoining sides, giving the pair the shape of a coffee bean. These aerobic bacteria have stringent nutritional requirements and grow best on enriched media such as lysed sheep's blood (chocolate) agar. Pathogenic *Neisseria* spp. often have the ability to secrete single-stranded DNA for transformation of other *Neisseria* spp., but commensal *Neisseria* spp. usually lack this ability. The two clinically significant *Neisseria* spp. are *N. meningitidis* and *N. gonorrhoeae*.

***N. meningitidis* is a significant cause of bacterial meningitis, particularly among adolescents and young adults.** The organism is a common colonizer of the oropharynx and is spread by the respiratory route. An immune response leads to elimination of the colonizing organism in most people, and this response is protective against subsequent infection with the same serotype of bacteria. There are several capsular serotypes of *N. meningitidis,* however five of them cause most cases of meningitis. Invasive disease mainly occurs when people encounter new strains to which they are not immune, as may happen to young children or to young adults living in crowded quarters such as military barracks or college dormitories. *N. meningitidis* is endemic in the United States, but epidemics occur periodically in sub-Saharan Africa and cause thousands of deaths. Highly effective conjugate vaccines for *N. meningitidis* composed of capsular polysaccharides conjugated to antigenic proteins are available for four serogroups of *N. meningitidis* (A, C, W, and Y), and recombinant protein vaccines are available for serogroup B.

Even in the absence of preexisting immunity, only a small fraction of people infected with *N. meningitidis* develop meningitis. The bacteria must invade respiratory epithelial cells and enter the blood. In the blood, the bacterial capsule inhibits opsonization and destruction of the bacteria by complement proteins. The importance of complement as a first-line defense against *N. meningitidis* is shown by the increased rates of serious infection among people who have inherited defects in the complement proteins (C5 to C9) that form the membrane attack complex, or patients with paroxysmal nocturnal hemoglobinuria (Chapter 14) who are being treated with an antibody inhibitor of the membrane attack complex. If *N. meningitidis* escapes the host response, the consequences can be severe. Although antibiotic treatment greatly reduces the mortality of *N. meningitidis* infection, about 10% of infected patients still die. The pathology of pyogenic meningitis is discussed in Chapter 28.

***N. gonorrhoeae* is an important cause of sexually transmitted infection (STI),** with more than 450,000 cases reported each year in the United States. It is second only to *C. trachomatis* among bacterial STIs. Infection in men causes urethritis. In women, *N. gonorrhoeae* infection is often

asymptomatic and so may go unnoticed. Untreated gonorrhea may lead to pelvic inflammatory disease, which can cause infertility or ectopic pregnancy (Chapter 22). Infection is diagnosed by culture and PCR tests. Increased resistance to oral cephalosporins has led to the recommendation that gonorrhea be treated with intramuscular ceftriaxone and oral azithromycin. Ceftriaxone-resistant *N. gonorrhoeae* is rare but has been detected in Canada and Japan.

Although *N. gonorrhoeae* infection usually manifests locally in the genital or cervical mucosa, pharynx, or anorectum, disseminated infections may occur. Like *N. meningitidis, N. gonorrhoeae* is much more likely to become disseminated in people who lack the complement proteins that form the membrane attack complex. Disseminated infection of adults and adolescents usually causes septic arthritis accompanied by a rash of hemorrhagic papules and pustules. Neonatal *N. gonorrhoeae* infection causes conjunctivitis that may lead to blindness and, rarely, sepsis. The eye infection, which is preventable by instillation of silver nitrate or antibiotics in the newborn's eyes, remains an important cause of blindness in some lower-income nations.

Pathogenesis

Neisseria spp. adhere to and invade nonciliated epithelial cells at the site of entry (nasopharynx, urethra, or cervix). Adherence of *N. gonorrhoeae* to epithelial cells is initially mediated by long pili, which bind to CD46, a protein expressed on all human nucleated cells. OPA proteins (so named because they make bacterial colonies opaque), located in the outer membrane of the bacteria, increase binding of *Neisseria* spp. to epithelial cells and promote entry of bacteria into cells.

***Neisseria* spp. use antigenic variation as a strategy to escape the immune response.** The existence of multiple capsular serotypes of *N. meningitidis* results in meningitis in some people on exposure to a new strain, as discussed earlier. In addition, *Neisseria* spp. also can generate new antigens by genetic mechanisms, which permit a single bacterial clone to change its expressed antigens and escape immune defenses. Such mechanisms involve both pili and OPA proteins:

- *Recombination of genes encoding pili proteins.* The pili are composed of polypeptides encoded by the pilin gene, which consists of a promoter and coding sequences for 10 to 15 pili protein variants. At any point in time, only one of these coding sequences is adjacent to the promoter, allowing it to be expressed. Homologous recombination periodically shuttles one of the other pilin coding sequences next to the promoter, resulting in expression of a different pilin variant.
- *Expression of different OPA proteins.* Each *OPA* gene has several repeats of a five-nucleotide sequence, which are frequently deleted or duplicated. These changes shift the reading frame of the gene so that it encodes new sequences. Stop codons are also introduced by the additions and deletions, which determine whether each *OPA* gene is expressed or silent. Thus, *Neisseria* spp. can express one, none, or multiple OPA proteins at any time.

Pertussis

Pertussis, or whooping cough, caused by the gram-negative coccobacillus *Bordetella pertussis*, is an acute, highly communicable illness characterized by paroxysms of violent coughing followed by a loud inspiratory "whoop" as the patient gasps for air. Infants younger than 1 year of age are at highest risk of death. Children with pertussis can have coughing spells for up to 10 weeks. Worldwide, it is estimated that there are more than 160,000 deaths in children younger than 5 years of age due to pertussis. For diagnosis, PCR is more sensitive than culture, but specificity of the assays can vary, so some level of culture confirmation is critical in an outbreak setting. Widespread use of a vaccine resulted in a dramatic decrease in cases; however, concern about side effects of the whole-cell vaccine increased after the 1980s and led to a decrease in use. An acellular vaccine became available in the 1990s, but protection wanes over time, so there has been an increased incidence in recent years, even in areas where vaccination rates appear satisfactory. Outbreaks observed every 3 to 5 years affect primarily not the infants, but instead older children, adolescents, and adults. It is hypothesized that modifications of *B. pertussis* may also be playing a role, but this is unproven.

Pathogenesis

B. pertussis colonizes the brush border of the bronchial epithelium and also invades macrophages. It contains a filamentous hemagglutinin that binds to carbohydrates on the surface of respiratory epithelial cells, as well as to CR3 (Mac-1) integrins on macrophages. Virulence factors of *B. pertussis* include pertussis toxin, adenylate cyclase toxin, dermonecrotic toxin, and tracheal cytotoxin. Pertussis toxin is a typical A-B toxin that is composed of five subunits. The A unit, like cholera toxin, ADP-ribosylates and inactivates guanine nucleotide-binding proteins, so these G proteins can no longer transduce signals, interrupting the effect of chemokines that use G protein-coupled receptors. The B component contains four subunits that bind to extracellular molecules and allow the A subunit to enter cells. The B subunit can also bind to cell surface molecules such as TLR4, and through these it can initiate signaling events in cells. Collectively, pertussis toxin subunits impair host defenses by inhibiting neutrophil and macrophage recruitment and activation and paralyzing cilia, among other effects.

B. pertussis also produces adenylate cyclase toxin that enters host cells and converts ATP to supraphysiologic levels of cAMP. The rise in cAMP inhibits phagocytosis, the oxidative burst, and nitric oxide–mediated killing in neutrophils and macrophages, and the formation of neutrophil extracellular traps. The interactions of the multiple toxins and virulence factors that determine the course of disease require further study.

MORPHOLOGY

Bordetella spp. cause a laryngotracheobronchitis that in severe cases features bronchial mucosal erosion, hyperemia, and copious mucopurulent exudate (Fig. 8.21). Unless superinfected, the lung alveoli remain open and intact. In parallel with a striking peripheral lymphocytosis (up to 90%), there is hypercellularity and enlargement of the mucosal lymph follicles and peribronchial lymph nodes.

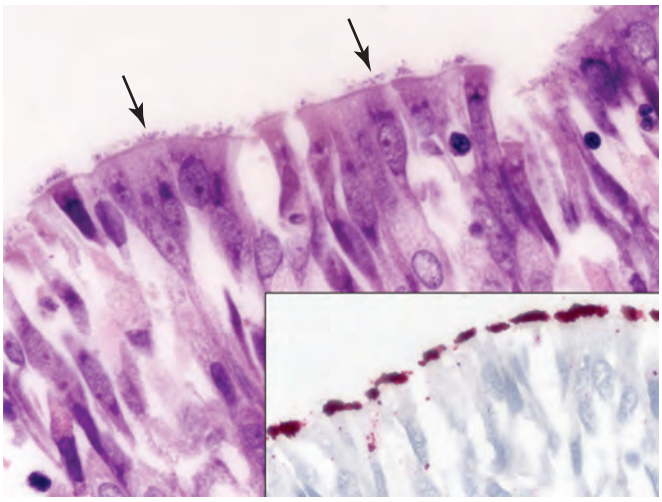


Figure 8.21 Whooping cough showing a haze of bacilli *(arrows)* entangled with the cilia of bronchial epithelial cells. The *inset* highlights the haze of bacilli by immunohistochemistry using a monoclonal antibody reactive to the lipo-oligosaccharide A of *Bordetella* pertussis. (Images courtesy Dr. Christopher Paddock of the Centers for Disease Control, Atlanta, Ga.)

Pseudomonal Infections

***Pseudomonas aeruginosa* is an opportunistic aerobic gram-negative bacillus that is a frequent, deadly pathogen in people with cystic fibrosis, severe burns, or neutropenia.** Many people with cystic fibrosis die of pulmonary failure secondary to chronic infection with *P. aeruginosa*. This bacterium can be very resistant to antibiotics, making these infections difficult to treat. It often infects extensive skin burns, which can lead to sepsis. *P. aeruginosa* is a common cause of hospital-acquired infections; it has been cultured from washbasins, respirator tubing, nursery cribs, and even antiseptic-containing bottles. It also causes corneal keratitis in wearers of contact lenses, endocarditis and osteomyelitis in intravenous drug abusers, external otitis (swimmer's ear) in healthy individuals, and severe external otitis in people with diabetes.

Pathogenesis

P. aeruginosa produces several toxins that contribute to local tissue damage. Early during infection of the lungs of people with cystic fibrosis, the organism secretes an A-B exotoxin called exotoxin A that, like diphtheria toxin, inhibits protein synthesis by ADP-ribosylating the ribosomal protein EF-2, leading to the death of host cells. At this early stage, the organism uses a type III secretion system to transport effector proteins into host cells; these reduce the ability of host cells to make antibacterial reactive oxygen species and also induce apoptosis of the cells. Later in chronic infection in the lungs of people with cystic fibrosis, the bacteria become organized into biofilms composed, in part, of alginate they secrete. Within the biofilm, the bacteria are protected from antibodies, complement, phagocytes, and antibiotics. In addition, the release of exotoxin A is reduced, as is the expression of the type III secretion system, and the organism evolves to become somewhat lower in virulence, although it does continue to damage the host by stimulating inflammation and by releasing enzymes (proteases and elastases) that damage tissue. During chronic infection, the organism develops antibiotic resistance both through biofilm production and genetic changes, making treatment difficult.

MORPHOLOGY

Pseudomonas spp. cause **necrotizing pneumonia** that is distributed through the terminal airways in a fleur-de-lis pattern, with striking pale necrotic centers and red, hemorrhagic peripheral areas. On microscopic examination, masses of organisms are seen, often concentrated in the walls of blood vessels, where host cells undergo coagulative necrosis (Fig. 8.22). This picture of gram-negative **bacterial vasculitis** accompanied by thrombosis and hemorrhage, although not pathognomonic, is highly suggestive of *P. aeruginosa* infection.

Bronchial obstruction caused by mucus plugging and subsequent *P. aeruginosa* infection are frequent complications of cystic fibrosis. Despite antibiotic treatment and the host immune response, chronic *P. aeruginosa* infection may result in bronchiectasis and pulmonary fibrosis (Chapter 15).

In skin burns, *P. aeruginosa* proliferates widely, penetrating deeply into the veins and spreading hematogenously. Well-demarcated necrotic and hemorrhagic oval skin lesions, called **ecthyma gangrenosum,** often appear. DIC is a frequent complication of *P. aeruginosa* bacteremia.

Plague

***Yersinia pestis* is a gram-negative facultative intracellular bacterium that causes an invasive, frequently fatal, infection called plague.** It is transmitted from rodents to humans by flea bites or, less often, from one human to another by aerosols. Plague caused three great pandemics that killed many millions of people; the devastation wreaked by the epidemic in Europe in the 1300s gave the infection the dramatic name *Black Death*. Most cases now occur in Africa, but the organism is endemic in many parts of the world, including nations in the former Soviet Union, the Americas, and Asia. Wild rodents in the rural western United States are infected with *Y. pestis* and are the source of about 10 to 15 human cases every year. In 2017, an outbreak in Madagascar resulted in over 2400 cases and more than 200 deaths. *Y. enterocolitica* and *Y. pseudotuberculosis* are genetically similar to *Y. pestis*; these bacteria cause fecal-orally transmitted ileitis and mesenteric lymphadenitis.

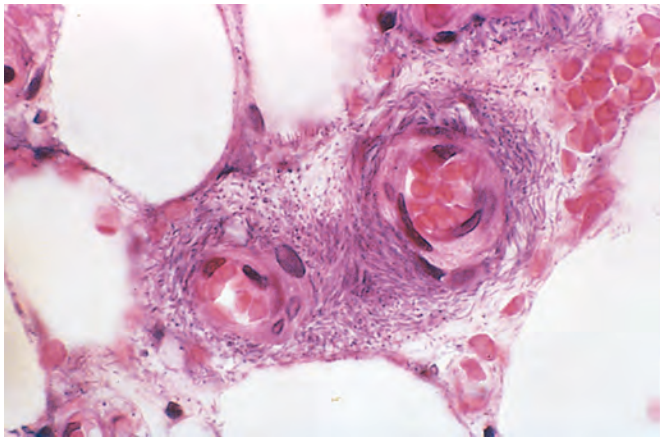


Figure 8.22 *Pseudomonas* vasculitis in which masses of organisms form a perivascular purple-blue haze.

Y. pestis has two growth phases in the flea: an early phase during which bacteria aggregate in the proventriculus but do not block blood flow completely, and a late phase during which a biofilm is formed that obstructs the gut of the infected flea. The starving flea bites and regurgitates before it feeds, and thus infects the rodent or human that it is biting. The bacterial gene regulation shifts on sensing relocation from the insect gut to the human host. The bacteria spread from the site of inoculation to lymphoid tissues, where they proliferate and inhibit the host from mounting an effective response. *Y. pestis* has a plasmid-borne complex of genes, the Yop virulon, which encodes a type III secretion system, a hollow syringe-like structure that projects from the bacterial surface, binds to host cells, and injects bacterial proteins called *Yops* (*Yersinia* outercoat proteins) into the cell. YopE, YopH, and YopT block phagocytosis by inactivating molecules that regulate actin polymerization. YopJ inhibits the signaling pathways that are activated by LPS, blocking the production of inflammatory cytokines.

MORPHOLOGY

Y. pestis causes lymph node enlargement (buboes), pneumonia, or sepsis with a striking neutrophilia. The distinctive histologic features include (1) massive proliferation of the organisms, (2) early appearance of protein-rich and polysaccharide-rich effusions with few inflammatory cells, (3) necrosis of tissues and blood vessels with hemorrhage, thrombosis, and marked tissue swelling, and (4) neutrophilic infiltrates that accumulate adjacent to necrotic areas as healing begins.

In **bubonic plague,** the infected flea bite is usually on the legs, where it forms a small pustule or ulcer. The draining lymph nodes enlarge dramatically within a few days and become soft, pulpy, and plum colored, and may infarct or rupture through the skin. These lymph nodes were called *buboes*, from the Greek word for "groin," giving rise to the name for this form of plague. In **pneumonic plague,** there is a severe, confluent, hemorrhagic, and necrotizing bronchopneumonia, often with fibrinous pleuritis. In **septicemic plague,** lymph nodes throughout the body as well as organs rich in mononuclear phagocytes develop foci of necrosis. Fulminant bacteremia also induces DIC with widespread hemorrhages and thrombi.

Chancroid (Soft Chancre)

Chancroid is an acute, ulcerative STI caused by *Haemophilus ducreyi*. The disease is most common in tropical and subtropical areas among lower socioeconomic groups and men who have frequent sex with prostitutes. Chancroid is one of the most common causes of genital ulcers in Africa, Southeast Asia, and the Caribbean, where it serves as an important cofactor in the transmission of HIV infection. Chancroid is uncommon in the United States, with fewer than 20 cases per year reported to the CDC in the past several years. The organism grows poorly in culture, requiring uncommon and highly-enriched media, and PCR-based tests are not widely available, so chancroid may be underdiagnosed.

MORPHOLOGY

Four to seven days after inoculation, a tender erythematous papule involving the external genitalia develops. In males, the primary lesion is usually on the penis; in females, most lesions occur in the vagina or the periurethral area. Over several days, the surface of the primary lesion erodes to produce an irregular, painful ulcer. In contrast to the primary chancre of syphilis, the ulcer of chancroid is not indurated, and multiple lesions may be present. The base of the ulcer is covered by shaggy, yellow-gray exudate. The regional lymph nodes become enlarged and tender in about 50% of cases within 1 to 2 weeks after primary infection. If the infection is not treated, the enlarged nodes (buboes) may erode the overlying skin to produce chronic, draining ulcers.

Microscopically, the ulcer of chancroid contains a superficial zone of neutrophilic debris and fibrin, and an underlying zone of granulation tissue containing areas of necrosis and thrombosed vessels. A dense, lymphoplasmacytic inflammatory infiltrate is present beneath the layer of granulation tissue. Coccobacilli are sometimes demonstrable in Gram or silver stains, but they are often obscured by other bacteria that colonize the ulcer base.

Granuloma Inguinale

Granuloma inguinale, or donovanosis, is a chronic inflammatory STI caused by *Klebsiella granulomatis* (formerly called *Calymmatobacterium donovani*), a minute, encapsulated, coccobacillus. Granuloma inguinale is an STI that is uncommon in the United States and Western Europe but is endemic in some rural tropical areas and some lower-income countries. Untreated cases are characterized by the development of extensive scarring, often associated with lymphatic obstruction and lymphedema (elephantiasis) of the external genitalia. Culture of the organism is difficult, and PCR assays are not widely available, so the diagnosis is made by microscopic examination of smears or biopsy samples of the ulcer.

MORPHOLOGY

Granuloma inguinale begins as a raised papular lesion on the moist stratified squamous epithelium of the genitalia or, rarely, the oral mucosa or pharynx. The lesion eventually ulcerates and develops abundant granulation tissue, which manifests grossly as a protuberant, soft, painless mass. As the lesion enlarges, its borders become raised and indurated. Disfiguring scars may develop in untreated cases and are sometimes associated with urethral, vulvar, or anal **strictures.** Regional lymph nodes typically are spared or show only nonspecific reactive changes, in contrast to chancroid.

Microscopic examination of active lesions reveals marked epithelial hyperplasia at the borders of the ulcer, sometimes mimicking carcinoma (pseudoepitheliomatous hyperplasia). A mixture of neutrophils and mononuclear inflammatory cells is present at the base of the ulcer and beneath the surrounding epithelium. The organisms are demonstrable in Giemsa-stained smears of the exudate as minute, encapsulated coccobacilli (Donovan bodies) in macrophages. Silver stains (e.g., the Warthin-Starry stain) may also demonstrate the organism.

Mycobacterial Infections

Bacteria in the genus *Mycobacterium* are slender, aerobic rods that grow in straight or branching chains. Mycobacteria have a unique waxy cell wall composed of unusual glycolipids and lipids including mycolic acid, which makes them acid-fast, meaning they will retain stains, even on treatment with a mixture of acid and alcohol. They are weakly gram-positive.

Tuberculosis

Tuberculosis is a chronic pulmonary and systemic disease caused most often by *M. tuberculosis*, and the leading infectious cause of death worldwide. The source of transmission is humans with active tuberculosis who release mycobacteria into the sputum. Oropharyngeal and intestinal tuberculosis contracted by drinking milk contaminated with *Mycobacterium bovis* is rare in countries where milk is routinely pasteurized, but it is still seen in countries that have tuberculous dairy cows and unpasteurized milk.

Epidemiology

According to the WHO, about 10 million people developed tuberculosis worldwide in 2018, with an estimated 1.3 million deaths in people without HIV infection and 300,000 in people with HIV infection. The mortality due to tuberculosis is falling by 3% per year. In 2018 there were about 9,000 cases of tuberculosis reported in the United States, 70% of which occurred in foreign-born people.

Tuberculosis flourishes wherever there is poverty, crowding, and chronic debilitating illness. In the United States, tuberculosis is mainly a disease of older adults, immigrants from high-burden countries, those living in crowded settings (prisons, homeless shelters, long-term care), and people with AIDS. Certain disease states also increase the risk: diabetes, Hodgkin lymphoma, chronic lung disease (particularly silicosis), chronic renal failure, malnutrition, alcoholism, and immunosuppression.

It is important that infection with *M. tuberculosis* be differentiated from active disease. Infection refers to the presence of bacteria in the body, which may be symptomatic (active disease) or not (latent infection). Most infections are acquired by person-to-person transmission of airborne organisms from an active case to a susceptible host. In most healthy people primary tuberculosis is asymptomatic, although it may cause fever and pleural effusion. Generally, the only evidence of infection, if any remains, is a tiny, fibrocalcific pulmonary nodule at the site of the infection. Viable organisms may remain dormant in such lesions for decades. If immune defenses are lowered, the infection may be reactivated, producing communicable and potentially life-threatening disease.

Pathogenesis

The outcome of infection in a previously unexposed, immunocompetent person depends on the development of antimycobacterial T cell–mediated immunity. These T cells control the host response to the bacteria and also result in development of pathologic lesions, such as caseating granulomas and cavitation.

Infection by *M. tuberculosis* proceeds in steps, from initial infection of macrophages to a subsequent Th1 response that both contains the bacteria and causes tissue damage (Fig. 8.23). Early in infection, *M. tuberculosis* replicates essentially unchecked within macrophages, and later in infection, the cell response stimulates macrophages to contain the proliferation of the bacteria. The steps in infection are as follows:

- *Entry into macrophages. M. tuberculosis* enters macrophages by phagocytosis mediated by several receptors expressed on the phagocyte, including mannose-binding lectin and the type 3 complement receptor (CR3).
- *Replication in macrophages. M. tuberculosis* inhibits maturation of the phagosome and blocks formation of the phagolysosome, allowing the bacterium to replicate unchecked within the vesicle, protected from the microbicidal mechanisms of lysosomes. The bacterium blocks phagolysosome formation by recruiting a host protein called coronin to the membrane of the phagosome. Coronin activates the phosphatase calcineurin, leading to inhibition of phagosome-lysosome fusion. Thus, during the earliest stage of primary tuberculosis (<3 weeks) in the nonsensitized individual, bacteria proliferate in the pulmonary alveolar macrophages and air spaces, resulting in bacteremia and seeding of multiple sites. Despite the bacteremia, most people at this stage are asymptomatic or have a mild flulike illness.
- *Innate immunity.* Multiple pathogen associated molecular patterns (Chapter 6) made by *M. tuberculosis* are recognized by innate immune receptors. Mycobacterial lipoarabinomannan binds TLR2, and unmethylated CpG nucleotides bind TLR9. These interactions initiate and enhance the innate and adaptive immune responses to *M. tuberculosis*, as described below.
- *The Th1 response.* About 3 weeks after infection, a Th1 response is mounted that activates macrophages, enabling them to become bactericidal. The response is initiated by mycobacterial antigens that enter draining lymph nodes and are displayed to T cells. Differentiation of Th1 cells depends on IL-12 and IL-18, which are produced by antigen-presenting cells that have encountered the mycobacteria. Stimulation of TLR2 by mycobacterial ligands promotes production of IL-12 by dendritic cells.
- *Th1-mediated macrophage activation and killing of bacteria.* Th1 cells, both in lymph nodes and in the lung, produce IFN-γ. IFN-γ is the critical mediator that activates macrophages and enables them to contain the *M. tuberculosis* infection. First, IFN-γ stimulates maturation of the phagolysosome in infected macrophages, exposing the bacteria to a lethal acidic, oxidizing environment. Second, IFN-γ stimulates expression of inducible nitric oxide (NO) synthase, which produces NO. NO combines with other oxidants to create reactive nitrogen intermediates, which are important for killing of mycobacteria. Third, IFN-γ mobilizes antimicrobial peptides (defensins) against the bacteria. Finally, IFN-γ stimulates autophagy, a process that sequesters and then destroys damaged organelles and intracellular bacteria such as *M. tuberculosis*.
- *Granulomatous inflammation and tissue damage.* In addition to stimulating macrophages to kill mycobacteria, the Th1 response orchestrates the formation of granulomas and caseous necrosis. Macrophages activated by IFN-γ differentiate into the "epithelioid histiocytes" that aggregate to form granulomas; some epithelioid cells may fuse to form giant cells. In many people this response halts the infection before significant tissue destruction or illness occur. In other people the infection progresses due to advanced age or immunosuppression, and the ongoing immune response results in caseous necrosis. Activated macrophages also secrete TNF and chemokines, which promote recruitment of more monocytes. The importance of TNF is underscored by the fact that

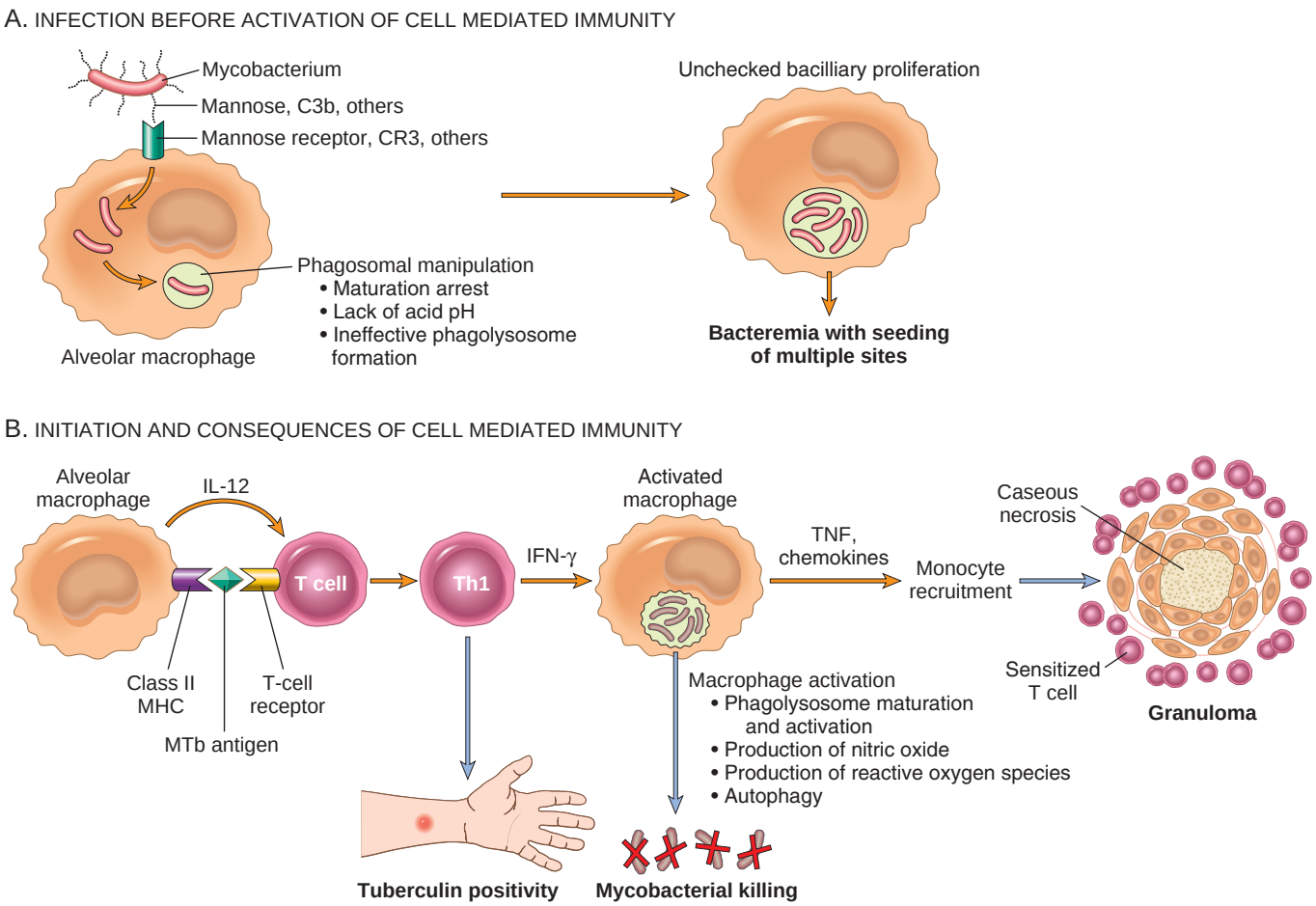


Figure 8.23 The sequence of events in primary pulmonary tuberculosis, commencing with inhalation of virulent *Mycobacterium* tuberculosis organisms and culminating with the development of cell-mediated immunity to the organism. (A) Events occurring during early infection, before activation of T-cell–mediated immunity. (B) The initiation and consequences of T-cell–mediated immunity. The development of resistance to the organism is accompanied by the appearance of a positive tuberculin test. *IFN-γ,* interferon-gamma; *MHC,* major histocompatibility complex; *MTb, M. tuberculosis; TNF,* tumor necrosis factor.

patients with rheumatoid arthritis who are treated with a TNF antagonist have an increased risk of tuberculosis reactivation.

- *Host susceptibility to disease.* AIDS is the greatest risk factor for progression to active disease, due to the loss of immunologic control of the organism. Other forms of immunosuppression, including glucocorticoids, TNF inhibitors, and transplants (solid organ and stem cell) also carry increased risk, as do renal failure and malnutrition. Rare inherited mutations that interfere with the Th1 response, such as loss of the IL-12 receptor β1 protein, result in increased susceptibility to severe tuberculosis and even symptomatic infection with normally avirulent (so-called "atypical") mycobacteria, such as the *Mycobacterium avium complex* (MAC), discussed later, or with the attenuated BCG vaccine strain. As previously mentioned, this group of genetic disorders is called *Mendelian susceptibility to mycobacterial disease.*

In summary, immunity to *M. tuberculosis* is primarily mediated by Th1 cells, which stimulate macrophages to kill the bacteria. This immune response, although largely effective, comes at the cost of accompanying tissue destruction. Reactivation of the infection or re-exposure to the bacilli in a previously sensitized host results in rapid mobilization of a defensive reaction but also increased tissue necrosis. Just as T-cell immunity and resistance are correlated, so, too, the loss of T-cell immunity (indicated by tuberculin negativity in a previously tuberculin-positive individual) may be an ominous sign that resistance to the organism has faded.

Clinical Features

Clinical tuberculosis is separated into two important types that differ in pathophysiology: primary tuberculosis, which occurs with the first infection, and secondary tuberculosis, which occurs in an individual who has been previously infected by *M. tuberculosis* (Fig. 8.24).

Primary tuberculosis is the form of disease that develops in a previously unexposed and therefore unsensitized person. Clinically significant disease develops in about 5% of newly infected people. With primary tuberculosis, the source of the organism is exogenous. In most people, the primary infection is contained, but in others, primary tuberculosis is progressive. The diagnosis of progressive primary tuberculosis in adults can be difficult. In contrast to secondary tuberculosis (apical disease with cavitation; see later), progressive primary tuberculosis more often

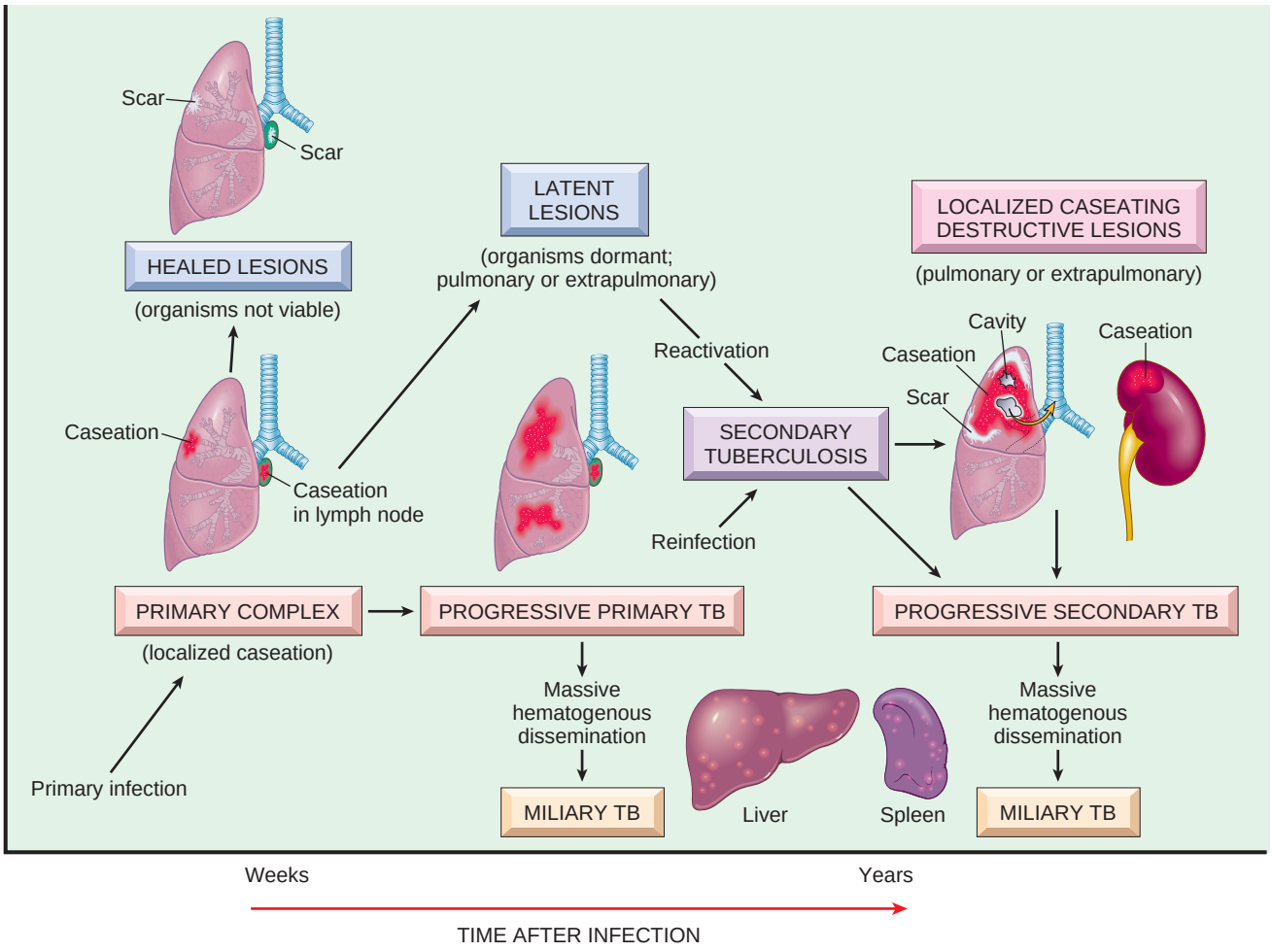


Figure 8.24 The natural history and spectrum of tuberculosis. (Modified from a sketch provided by Professor R. K. Kumar, The University of New South Wales, School of Pathology, Sydney, Australia.)

resembles an acute bacterial pneumonia with consolidation of the lobe, hilar adenopathy, and pleural effusion. Lymphatic and hematogenous dissemination following primary infection may result in the development of tuberculous meningitis and miliary tuberculosis (discussed later).

Secondary tuberculosis is the pattern of disease that arises in a previously sensitized host. It may follow shortly after primary tuberculosis, but more commonly it appears months to years after the initial infection, usually when host resistance is weakened. It most commonly stems from reactivation of a latent infection, but may also result from exogenous reinfection in the face of waning host immunity or when a large inoculum of virulent bacilli overwhelms the host immune system. Reactivation is more common in low-prevalence areas, and reinfection plays an important role in regions of high contagion.

Secondary pulmonary tuberculosis classically involves the apex of the upper lobes of one or both lungs. Because of the preexistence of hypersensitivity, the bacilli elicit a prompt and marked tissue response that tends to wall off the focus of infection. As a result, the regional lymph nodes are less prominently involved early in secondary disease than they are in primary tuberculosis. On the other hand, cavitation occurs readily in the secondary form. Indeed, cavitation is almost inevitable in neglected secondary tuberculosis, and erosion of the cavities into an airway is an important source of infection because the person now coughs sputum that contains bacteria.

Localized secondary tuberculosis may be asymptomatic. When manifestations appear, they are usually insidious in onset. Systemic symptoms, probably related to cytokines released by activated macrophages (e.g., TNF), often appear early in the course and include malaise, anorexia, weight loss, and fever. Commonly, the fever is low grade and remittent (appearing late each afternoon and then subsiding), and night sweats occur. With progressive pulmonary involvement, increasing amounts of sputum, at first mucoid and later purulent, appear. Some degree of hemoptysis is present in about one-half of all cases of pulmonary tuberculosis. Pleuritic pain may result from extension of the infection to the pleural surfaces. Extrapulmonary manifestations of tuberculosis are legion and depend on the organ system involved.

Multidrug resistance is now seen more commonly than it was in past years; hence, all newly diagnosed cases in the United States are treated with at least four drugs, unless the susceptibility of the bacterium from the source case is known. The prognosis is generally good if infections are localized to the lungs, except when they are caused by drug-resistant strains or occur in debilitated individuals.

Tests for tuberculosis detect either the bacteria or the T lymphocyte–mediated host response to the bacteria. There are several tests available for detection of *M. tuberculosis* in patients with active disease. Acid-fast smears and cultures of the sputum of patients suspected of having tuberculosis should be performed. Culture on solid agar media shows growth at 3 to 6 weeks, but culture in liquid media can provide an answer within 2 weeks. Comprehensive antibiotic susceptibility testing can only be performed by culture. PCR amplification of *M. tuberculosis* DNA allows for even more rapid diagnosis. An FDA-approved PCR test is available that identifies both the presence of *M. tuberculosis* and, if the organism is detected, whether it is resistant to rifampin. PCR assays are as sensitive as culture in acid-fast smear-positive samples, but slightly less sensitive in smear-negative tuberculosis, and substantially less sensitive in children. Although culture remains the gold standard, PCR should also be performed if active tuberculosis is suspected because results are available quickly.

Testing for latent tuberculosis is done by detecting T-cells specific for or delayed hypersensitivity to *M. tuberculosis* antigens. This can be detected by either IFN-γ release assays (IGRAs) or the tuberculin (purified protein derivative [PPD], or Mantoux) skin test. IGRAs are in vitro tests in which lymphocytes from the patient are stimulated with protein antigens from *M. tuberculosis*. Production of IFN-γ by the T cells is measured to assess the level of T-cell immunity to the organism. The tuberculin skin test is performed by intracutaneous injection of purified protein derivative of *M. tuberculosis,* which induces a visible and palpable induration that peaks in 48 to 72 hours. A positive IGRA or tuberculin test signifies T cell–mediated immunity to mycobacterial antigens but does not differentiate between infection and active disease. False-negative reactions may occur in the setting of certain viral infections, sarcoidosis, malnutrition, Hodgkin lymphoma, immunosuppression, and (notably) active tuberculous disease. A false-positive tuberculin skin test may result from infection by atypical mycobacteria or prior vaccination with BCG *(Bacillus Calmette-Guerin),* an attenuated strain of *M. bovis* that is used as a vaccine in some countries. False-positive results are uncommon with IGRAs.

Tuberculosis is a leading cause of death in people with AIDS. **All stages of HIV infection are associated with an increased risk of tuberculosis**. The use of antiretroviral therapy (ART) reduces the risk of tuberculosis in people with HIV infection, but even with ART, people infected with HIV are more likely to develop symptomatic tuberculosis. A low CD4 count before starting ART is an important risk factor for development of tuberculosis, which underscores the role of the immune response in keeping reactivation of *M. tuberculosis* in check. Pulmonary manifestations of tuberculosis in HIV-infected individuals are variable, ranging from focal lesions to multifocal infiltrates to localized apical disease with cavitation. The extent of immunodeficiency also determines the frequency of extrapulmonary involvement, rising from 10% to 15% in mildly immunosuppressed people to greater than 50% in those with severe immunodeficiency. Other atypical features of tuberculosis in HIV-positive people include an increased frequency of false-negative sputum smears and tuberculin tests (the latter sometimes referred to as "anergy"), and the absence of characteristic granulomas in tissues, particularly in the late stages of HIV. The increased frequency of sputum smear-negativity is paradoxical because these immunosuppressed patients typically have higher bacterial loads. The likely explanation is that cavitation and bronchial damage are more severe in immunocompetent individuals, resulting in more bacilli in expelled sputum. In contrast, the absence of bronchial wall destruction due to reduced T cell–mediated hypersensitivity results in the excretion of fewer bacilli in the sputum.

MORPHOLOGY

Primary Tuberculosis. In countries where consumption of infected milk has been eliminated, primary tuberculosis almost always begins in the lungs. Typically, the inhaled bacilli implant in the distal airspaces of the lower part of the upper lobe or the upper part of the lower lobe, usually close to the pleura. As sensitization develops, a 1- to 1.5-cm area of gray-white inflammation with consolidation emerges, known as the *Ghon focus*. In most cases, the center of this focus undergoes caseous necrosis. Tubercle bacilli, either free or within phagocytes, drain to the regional nodes, which also often caseate. **This combination of parenchymal lung lesion and nodal involvement is referred to as the *Ghon complex*** (Fig. 8.25). During the first few weeks, there is also lymphatic and hematogenous dissemination to other parts of the body. In approximately 95% of cases, development of cell-mediated immunity controls the infection. Hence, the Ghon complex undergoes progressive fibrosis, often followed by radiologically detectable calcification, and despite seeding of other organs, no lesions develop.

Histologically, sites of active involvement are marked by a characteristic granulomatous inflammatory reaction that forms both caseating and noncaseating tubercles (Fig. 8.26A to C). Individual tubercles are microscopic; it is only when multiple

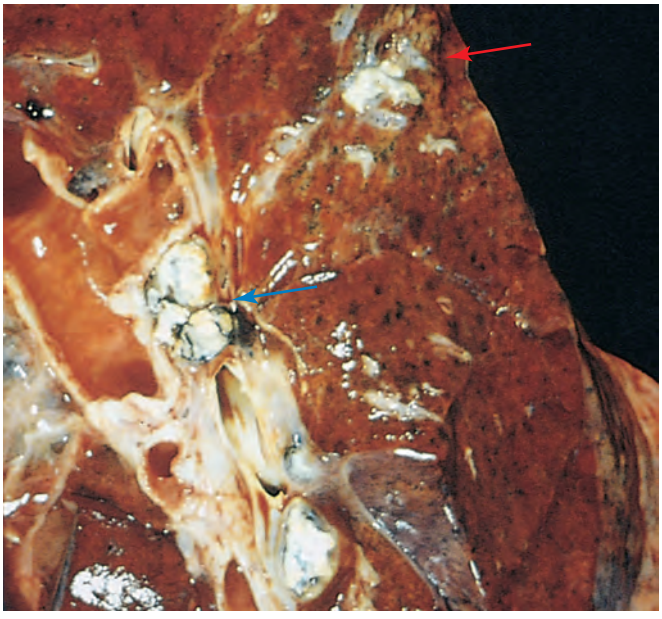


Figure 8.25 Primary pulmonary tuberculosis, Ghon complex. The gray-white parenchymal focus is under the pleura in the lower part of the upper lobe *(red arrow)*. Hilar lymph nodes with caseation are seen on the left *(blue arrow)*.

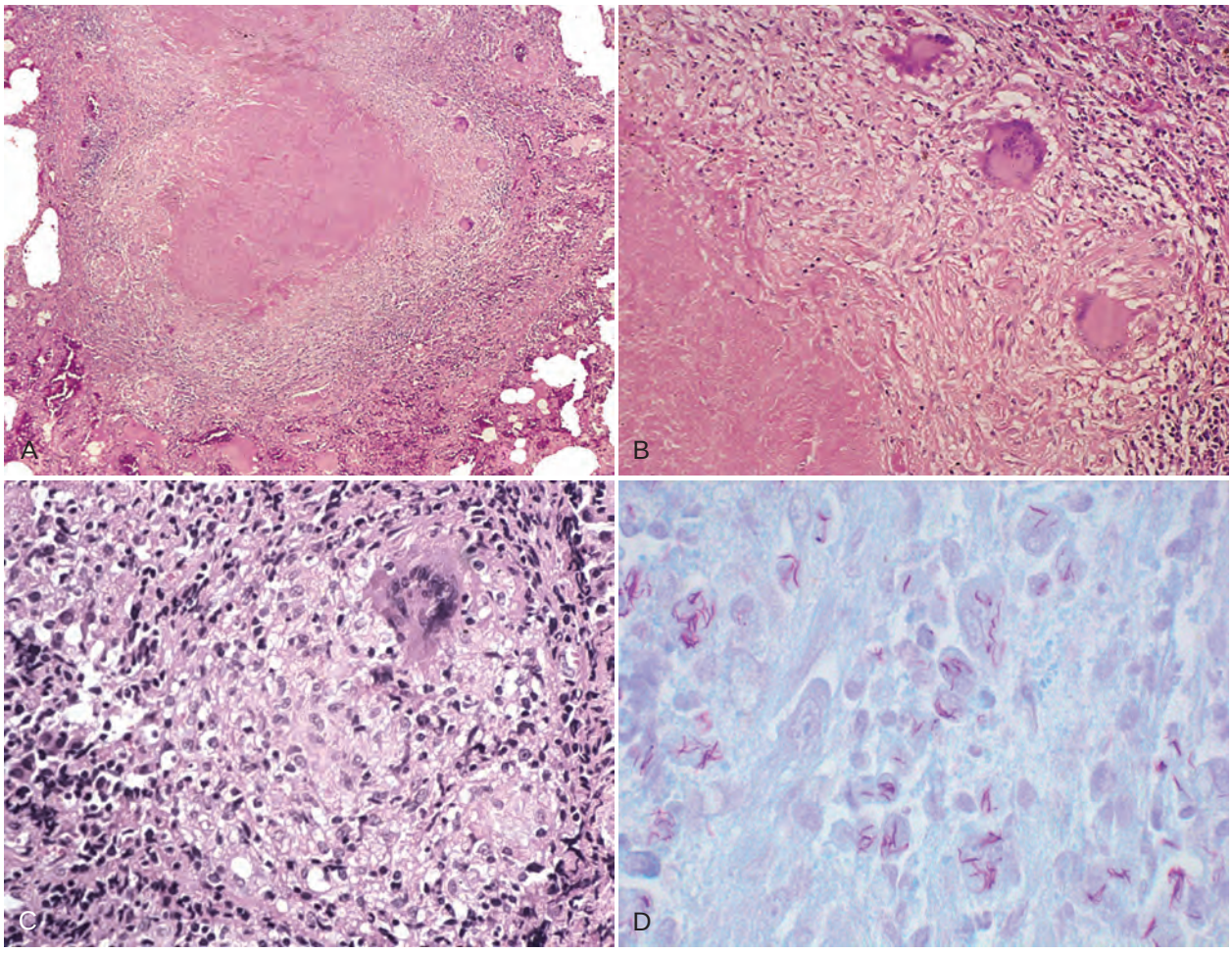


Figure 8.26 The morphologic spectrum of tuberculosis. Characteristic tubercle at low magnification (A) and high magnification (B) shows central granular caseation surrounded by epithelioid and multinucleate giant cells. This is the usual response in people who have developed cell-mediated immunity to the organism. (C) Occasionally, even in immunocompetent patients, tubercular granulomas may not show central caseation; hence regardless of the presence or absence of caseous necrosis, use of special stain for acid-fast organisms is indicated when granulomas are present. (D) In this specimen from an immunocompromised patient, macrophages with large numbers of mycobacteria are seen (acid-fast stain).

granulomas coalesce that they become macroscopically visible. The granulomas are usually enclosed within a fibroblastic rim punctuated by lymphocytes. Multinucleated giant cells are present in the granulomas. Immunocompromised people do not form the characteristic granulomas, and their macrophages contain many bacilli (Fig. 8.26D).

Secondary Tuberculosis. The initial lesion is usually a small focus of consolidation, less than 2 cm in diameter, within 1 to 2 cm of the apical pleura. Such foci are sharply circumscribed, firm and gray-white to yellow in color, and have variable degrees of central caseation and peripheral fibrosis (Fig. 8.27). In immunocompetent individuals, the initial parenchymal focus undergoes progressive fibrous encapsulation, leaving only fibrocalcific scars. Histologically, the active lesions show characteristic coalescent tubercles with central caseation. Tubercle bacilli can often be identified with acid-fast stains in early exudative and caseous phases of granuloma formation but are usually too few to be found in the late, fibrocalcific stages. Localized, apical, secondary pulmonary tuberculosis may heal with fibrosis either spontaneously or after therapy, or the disease may progress and extend along several different pathways.

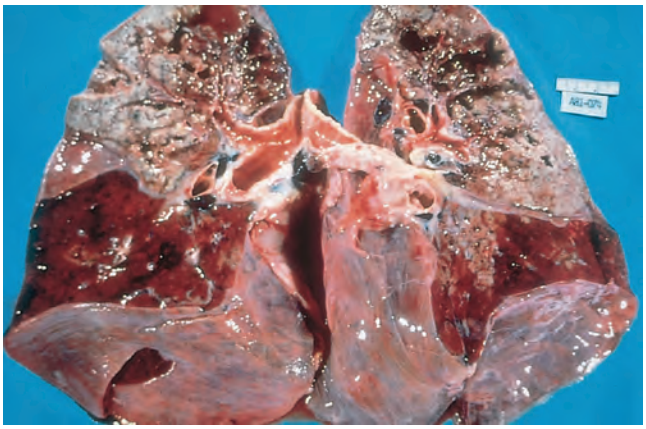


Figure 8.27 Secondary pulmonary tuberculosis. The upper parts of both lungs are riddled with gray-white areas of caseation and multiple areas of softening and cavitation.

Progressive pulmonary tuberculosis may ensue in older adults and immunosuppressed people. The apical lesion expands into adjacent lung and eventually erodes into bronchi and vessels. This evacuates the caseous center, creating a ragged, irregular cavity that is poorly walled off by fibrous tissue. Erosion of blood vessels results in hemoptysis. With adequate treatment the process may be arrested, although healing by fibrosis often distorts the pulmonary architecture. The cavities, now free of inflammation, may persist or become fibrotic. If the treatment is inadequate or if host defenses are impaired, the infection may spread via airways, lymphatic channels, or the vascular system. **Miliary pulmonary disease** occurs when organisms draining through lymphatics enter the venous blood and circulate back to the lung. Individual lesions are either microscopic or small, visible (2-mm) foci of yellow-white consolidation scattered through the lung parenchyma (the adjective "miliary" is derived from the resemblance of these foci to millet seeds). Miliary lesions may expand and coalesce, resulting in consolidation of large regions or even whole lobes of the lung. With progressive pulmonary tuberculosis, the pleural cavity is invariably involved, and serous **pleural effusions**, **tuberculous empyema**, or **obliterative fibrous pleuritis** may develop. Progressive primary tuberculosis that occurs in immunosuppressed individuals spreads in a similar manner.

Endobronchial, **endotracheal**, and **laryngeal tuberculosis** may develop by spread through lymphatic channels or from expectorated infectious material. The mucosal lining may be studded with minute granulomatous lesions that may only be apparent microscopically.

Systemic miliary tuberculosis occurs when bacteria disseminate through the systemic arterial system. Miliary tuberculosis is most prominent in the liver, bone marrow, spleen, adrenals, meninges, kidneys, fallopian tubes, and epididymis, but can involve any organ (Fig. 8.28).

Isolated tuberculosis may appear in any of the organs or tissues seeded hematogenously and may be the presenting manifestation. Organs that are commonly involved include the meninges (tuberculous meningitis), kidneys (renal tuberculosis), adrenals (formerly an important cause of Addison disease), bones (osteomyelitis), and fallopian tubes (salpingitis). When the vertebrae are affected, the disease is referred to as **Pott disease.** Paraspinal "cold" abscesses in these patients may track along tissue planes and present as an abdominal or pelvic mass.

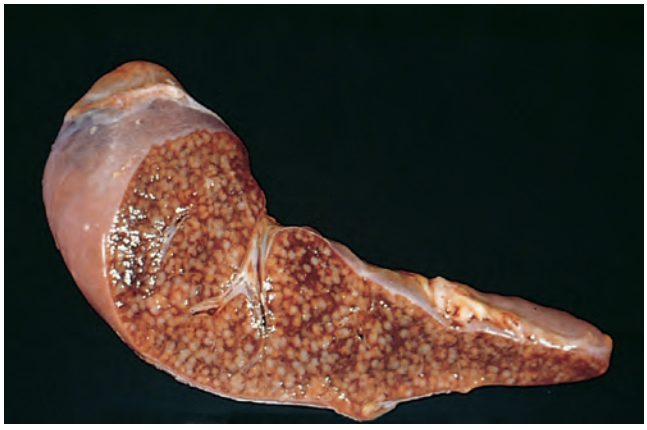


Figure 8.28 Miliary tuberculosis of the spleen. The cut surface shows numerous gray-white tubercles.

Lymphadenitis is the most frequent presentation of extrapulmonary tuberculosis, usually occurring in the cervical region ("scrofula"). In HIV-negative individuals, lymphadenitis tends to be unifocal and localized. HIV-positive people, on the other hand, almost always have multifocal disease, systemic symptoms, and either pulmonary or other organ involvement by active tuberculosis.

As previously mentioned, **intestinal tuberculosis** contracted by the drinking of contaminated milk is common in countries where bovine tuberculosis is present and milk is not pasteurized. In countries where milk is pasteurized, intestinal tuberculosis is more often caused by the swallowing of coughed-up infective material in patients with advanced pulmonary disease. Typically the organisms are seeded to mucosal lymphoid aggregates of the small and large bowel, which then undergo granulomatous inflammation that can lead to ulceration of the overlying mucosa, particularly in the ileum. Healing creates strictures.

Nontuberculous Mycobacterial Infections

Nontuberculous mycobacteria (NTM) refer to mycobacteria other than *M. tuberculosis* and very fastidious *M. leprae*. The prevalence of NTM disease has increased worldwide. Of the 150 NTM species, the most frequent human pathogens are *Mycobacterium avium* complex (MAC), *M. abscessus* complex, and *M. kansasii*. The prevalence of specific species of NTM varies in different geographic regions of the world. Treatment differs for these pathogens, so identifying the specific organism is important. It has been difficult to distinguish *M. avium* and *M. intracellulare* using traditional physical and biochemical tests, and the clinical features of these infections are similar, so they are grouped into a complex. Newer molecular methods are better able to distinguish these two species, as well as *M. chimaera* in the same complex. NTM are ubiquitous in the environment, water, and soil, and there is little human-to-human transmission, except for a small number of cases in cystic fibrosis patients.

Clinical presentations of NTM infections include chronic pulmonary disease (most common), lymphadenitis, cutaneous disease, and disseminated disease. Normal host defense mechanisms usually prevent infection, so vulnerable individuals include those with structural lung damage, cystic fibrosis, bronchiectasis, primary ciliary dyskinesia, chronic obstructive pulmonary disease, or pneumoconiosis. Additional predisposing factors include immunosuppression from HIV, transplantation, treatment with TNF inhibitors, and rare cases of anti-IFN-γ autoantibodies or inherited defects in the IL-12/IFN-γ pathways. Rapidly growing mycobacterial infections are often associated with postsurgical or posttraumatic infections. In patients with marked T-cell immunodeficiency, MAC causes widely disseminated infections, and organisms proliferate abundantly in many organs, including the lungs and gastrointestinal system. Patients are feverish, with drenching night sweats and weight loss. In cases of MAC infection in a person without HIV or another severe immunodeficiency, the organisms primarily infect the lung, causing a productive cough and sometimes fever and weight loss. Radiographic characteristics of disease may be fibrocavitary lesions primarily in the upper lobes or nodular bronchiectasis with multifocal clusters of small nodules. NTM biofilm development on medical devices is a well-documented source of a number of recent cases.

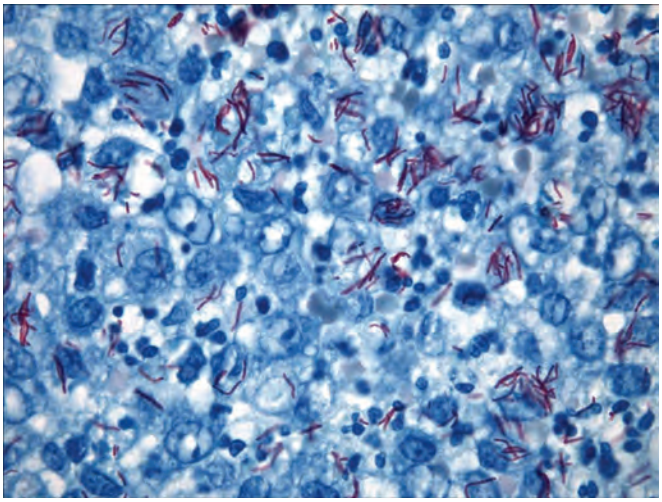


Figure 8.29 *Mycobacterium avium* infection in a patient with AIDS, showing massive infection with acid-fast organisms. This pattern is more common in patients with acquired immunodeficiencies.

MORPHOLOGY

The hallmark of MAC infections in patients with HIV is abundant acid-fast bacilli within macrophages (Fig. 8.29). Depending on the severity of immunodeficiency, MAC infections can be widely disseminated throughout the mononuclear phagocyte system, causing enlargement of involved lymph nodes, liver, and spleen, or localized to the lungs. There may be a yellowish pigmentation to these organs secondary to the large number of organisms present in swollen macrophages. Granulomas, lymphocytes, and tissue destruction are rare.

Leprosy

Leprosy, or *Hansen disease*, is a slowly progressive infection caused by *M. leprae* that mainly affects the skin and peripheral nerves. Despite its low communicability, leprosy remains endemic among people living in several lower-income tropical nations.

Pathogenesis

The source of infection and route of transmission are not known, however human respiratory secretions or soil are likely origins. *M. leprae* is taken up by macrophages and disseminates in the blood, but it replicates primarily in relatively cool tissues of the skin and extremities. It proliferates best at 32° to 34°C, the temperature of the human skin. Like *M. tuberculosis*, *M. leprae* secretes no toxins, and its virulence is based on properties of its cell wall, which is similar enough to that of *M. tuberculosis* that immunization with BCG confers some protection against *M. leprae* infection. Cell-mediated immunity is manifested by delayed-type hypersensitivity reactions to dermal injections of a bacterial extract called *lepromin*.

***M. leprae* causes two strikingly different patterns of disease, called tuberculoid and lepromatous, that are determined by the helper T-lymphocyte response to *M. leprae*.** People with the less severe tuberculoid leprosy have dry, scaly skin lesions that lack sensation. They often have asymmetric involvement of large peripheral nerves. The more severe form, lepromatous leprosy, includes symmetric skin thickening and nodules. In lepromatous leprosy, widespread invasion of the mycobacteria into Schwann cells and into endoneural and perineural macrophages damages the peripheral nervous system. In advanced cases of lepromatous leprosy, *M. leprae* is present in sputum and blood. People can also have intermediate forms of disease, called *borderline leprosy*.

As mentioned earlier, tuberculoid and lepromatous leprosy are associated with different T-cell responses. People with tuberculoid leprosy have a Th1 response associated with production of IL-2 and IFN-γ as well as a Th17 response. As with *M. tuberculosis*, IFN-γ functions to mobilize an effective host macrophage response, and hence the microbial burden is low. Also, antibody production is low. Lepromatous leprosy is associated with a weak Th1 response and, in some cases, a relative increase in the Th2 response. The net result is weak cell-mediated immunity and an inability to control the bacteria, which can be readily visualized in tissue sections. Occasionally, most often in the lepromatous form, antibodies are produced against *M. leprae* antigens. Paradoxically, these antibodies are usually not protective, but they may form immune complexes with free antigens that can lead to erythema nodosum, vasculitis, and glomerulonephritis. Some leprosy patients have a mixed Th1/Th2 cytokine pattern.

MORPHOLOGY

Tuberculoid leprosy begins with localized flat, red skin lesions that enlarge and develop irregular shapes with indurated, elevated, hyperpigmented margins and depressed pale centers (central healing). Neuronal involvement dominates tuberculoid leprosy. Nerves become enclosed within granulomatous inflammatory reactions and, if small (e.g., the peripheral twigs), are destroyed (Fig. 8.30). Nerve degeneration causes skin anesthesias and skin and muscle atrophy that render the person liable to trauma of the affected parts, leading to the development of chronic skin ulcers. Contractures, paralyses, and autoamputation of fingers or toes may ensue. Facial nerve involvement can lead to paralysis of the eyelids, with keratitis and corneal ulcerations. On microscopic examination, all sites of involvement have granulomatous lesions closely resembling those found in tuberculosis. Because of the strong host defense, bacilli are almost never found, hence the name **paucibacillary** leprosy. The presence of granulomas and absence of bacteria reflect strong T-cell immunity. Because leprosy pursues an extremely slow course, spanning decades, most patients die with leprosy rather than of it.

Lepromatous leprosy involves the skin, peripheral nerves, anterior eye chamber, upper airways (down to the larynx), testes, hands, and feet. The vital organs and CNS are rarely affected, presumably because the core temperature is too high for growth of *M. leprae*. Lepromatous lesions contain large aggregates of lipid-laden macrophages (lepra cells), often filled with masses ("globi") of acid-fast bacilli (Fig. 8.31). Because of the abundant bacteria, lepromatous leprosy is referred to as **multibacillary.** Macular, papular, or nodular lesions form on the face, ears, wrists, elbows, and knees. With progression, the nodular lesions coalesce to yield a distinctive leonine facies. Most skin lesions are hypoesthetic or anesthetic. Lesions in the nose may cause persistent inflammation and bacilli-laden discharge. The peripheral nerves,

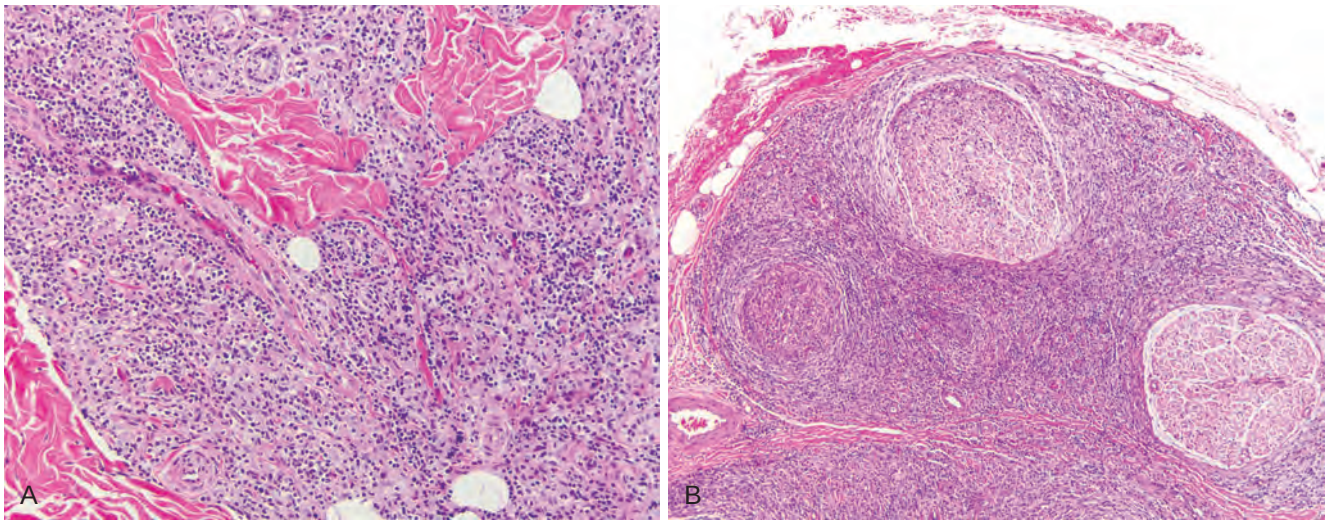


Figure 8.30 Two types of inflammatory infiltrates common in leprosy are: (A) dense dermal macrophage infiltration surrounding adnexa, vessels, and nerves (resulting in subcutaneous nodules) and (B) dense chronic lymphocytic and histiocytic infiltration into large nerve bundles (resulting in mononeuropathy).

particularly the ulnar and peroneal nerves where they approach the skin surface, are symmetrically invaded with mycobacteria, with minimal inflammation. Loss of sensation and trophic changes in the hands and feet follow the nerve lesions. Lymph nodes contain aggregates of bacteria-filled foamy macrophages in the paracortical (T-cell) areas and reactive germinal centers. In advanced disease, aggregates of macrophages are also present in the splenic red pulp and the liver. The testes are usually extensively involved, leading to destruction of the seminiferous tubules and consequent sterility.

Spirochete Infections

Spirochetes are gram-negative, slender, corkscrew-shaped bacteria with axial periplasmic flagella wound around a helical protoplasm. The bacteria are covered in a membrane called an outer sheath, which is thought to mask bacterial antigens from the host immune response. *Treponema pallidum* subsp. *pallidum* is the microaerophilic spirochete that causes syphilis, a chronic STI with multiple clinical presentations. Other closely related treponemes cause yaws *(T. pallidum* subsp. *pertenue)* and pinta *(T. pallidum* subsp. *carateum)*.

Syphilis

Syphilis is a chronic sexually transmitted infection (STI) with varied clinical and pathologic manifestations. The causative spirochete, *T. pallidum* subsp. *pallidum,* hereafter referred to simply as *T. pallidum,* is too slender to be seen in Gram stain, but it can be visualized by silver stains and immunofluorescence techniques (Fig. 8.32). Transplacental transmission of *T. pallidum* occurs readily, and active disease during pregnancy results in congenital syphilis. *T. pallidum* cannot be easily grown in culture. Public health programs and penicillin treatment reduced the number of cases of syphilis in the United States from the late 1940s until the late 1970s. In North America and western Europe, the incidence of syphilis has fluctuated since the 1970s and increased dramatically in the past decade among men who have sex with men, particularly those with HIV infection. Worldwide more than 5 million new cases of syphilis are diagnosed every year, and congenital infections are not uncommon.

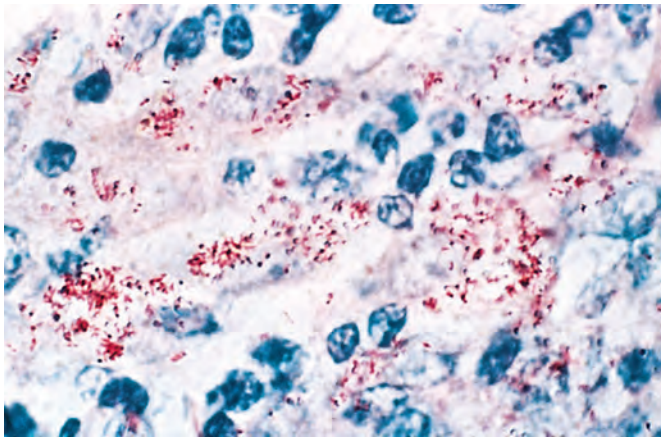


Figure 8.31 Lepromatous leprosy. Acid-fast bacilli ("red snappers") within macrophages.

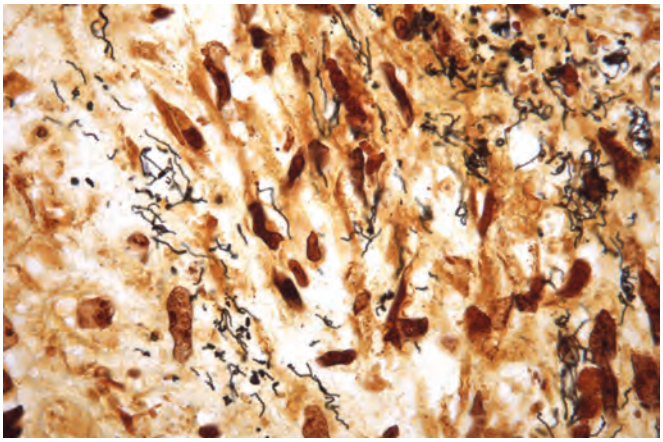


Figure 8.32 *Treponema pallidum* (Steiner silver stain) showing several spirochetes in histologic sections of placental syphilis.

Pathogenesis

Proliferative endarteritis affecting small vessels with a surrounding plasma cell–rich infiltrate is characteristic of all stages of syphilis. Much of the pathology of syphilis can be ascribed to the ischemia produced by the vascular lesions. The pathogenesis of endarteritis is unknown.

The immune response to *T. pallidum* reduces the burden of bacteria and can lead to resolution of local lesions but does not reliably eliminate the systemic infection. Superficial sites of infection (chancres and rashes) have an intense inflammatory infiltrate that includes T cells, plasma cells, and macrophages that surround the bacteria. The infiltrating CD4+ T cells are Th1 cells that may activate macrophages to kill the bacteria. Treponeme-specific antibodies are detectable, and these activate complement in the lesion and opsonize the bacteria for phagocytosis by macrophages. In many patients, the organism persists despite these host responses. A protein in the outer membrane of *T. pallidum*, TprK, accumulates structural diversity during the course of infection through gene conversion (recombination) between silent donor sites and the *tprK* gene, and this might contribute to antigenic diversity that allows the organism to persist. The difficulty of developing a continuous *in vitro* culture system for this organism has limited investigation of pathogenesis.

Syphilis is divided into three stages, with distinct clinical and pathologic manifestations (Fig. 8.33).

Primary Syphilis. This stage, occurring approximately 3 weeks after infection, features a single firm, nontender, raised, red lesion (chancre) located at the site of treponemal invasion on the penis, cervix, vaginal wall, or anus. The chancre heals with or without therapy. Spirochetes are plentiful within the chancre and spread from there throughout the body by hematologic and lymphatic dissemination.

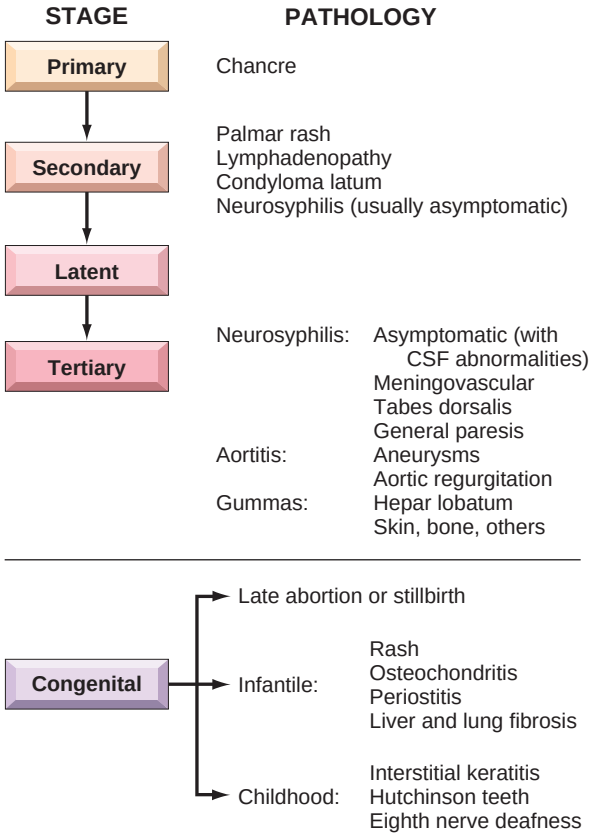


Figure 8.33 Protean manifestations of syphilis.

Secondary Syphilis. This stage is marked by painless, superficial lesions of the skin and mucosal surfaces. It occurs 2 to 10 weeks after the primary chancre in approximately 75% of untreated people. Skin lesions frequently occur on the palms or soles of the feet and may be maculopapular, scaly, or pustular. Moist areas of the skin, such as the anogenital region, inner thighs, and axillae, may have broad-based elevated plaques called *condylomata lata*. Silvery-gray superficial erosions may form on the oral, pharyngeal, and genital mucous membranes. Lymphadenopathy, mild fever, malaise, and weight loss are also common in secondary syphilis. Asymptomatic neurosyphilis (discussed later) occurs in 8% to 40% of patients, and symptomatic neurosyphilis with meningitis, visual changes, or hearing changes occurs in 1% to 2%. Secondary syphilis lasts several weeks, and then the person enters the latent stage of the disease.

Tertiary Syphilis. Tertiary syphilis has three main manifestations: cardiovascular syphilis, neurosyphilis, and so-called benign tertiary syphilis. These may occur alone or in combination. Tertiary syphilis occurs in one-third of untreated patients, usually after a latent period of 5 years or more.

- *Cardiovascular syphilis,* in the form of syphilitic aortitis, accounts for more than 80% of cases of tertiary disease. The pathogenesis of this vascular lesion is not known, but the scarcity of treponemes and the intense inflammatory infiltrate suggest that the immune response plays a role. The aortitis leads to slowly progressive dilation of the aortic root and arch, which causes aortic valve insufficiency and aneurysms of the proximal aorta (Chapter 11).
- *Neurosyphilis* may be symptomatic or asymptomatic. Symptomatic neurosyphilis is discussed in Chapter 28. Asymptomatic neurosyphilis, which accounts for about one-third of neurosyphilis cases, is initially suspected on detection of CSF abnormalities such as pleocytosis (increased numbers of inflammatory cells), elevated protein levels, or decreased glucose, and is confirmed by detection of antibodies stimulated by the spirochetes (discussed later) in the CSF. Antibiotics are given for a longer time if the spirochetes have spread to the CNS, so patients with tertiary syphilis should be tested for neurosyphilis even if they do not have neurologic symptoms.
- *Benign tertiary syphilis* is characterized by the formation of *gummas* in bone, skin, and the mucous membranes of the upper airway and mouth. Gummas are nodular lesions probably related to the development of delayed hypersensitivity to the bacteria. Skeletal involvement characteristically causes pain, tenderness, swelling, and pathologic fractures. Gummas in the skin and mucous membranes may produce nodular lesions or, rarely, destructive, ulcerative lesions. Gummas are rare because of the use of effective antibiotics.

Congenital Syphilis. Congenital syphilis occurs most frequently during maternal primary or secondary syphilis, when the spirochetes are most numerous. Intrauterine death

and perinatal death each occurs in approximately 25% of cases of untreated congenital syphilis.

Manifestations of congenital syphilis are divided into those that occur in the first 2 years of life (infantile syphilis) and those that occur later (tardive syphilis). Infantile syphilis is often manifested by nasal discharge and congestion (snuffles) in the first few months of life. A desquamating or bullous rash can lead to sloughing of the skin, particularly of the hands and feet and around the mouth and anus. Hepatomegaly and skeletal abnormalities are common. Late manifestations develop in nearly one-half of untreated children with neonatal syphilis.

Serologic Tests for Syphilis. Serology remains the mainstay of diagnosis of syphilis. Serologic tests include nontreponemal antibody tests and antitreponemal antibody tests. Nontreponemal tests measure antibody to a cardiolipin-cholesterol-lecithin antigen, present in both host tissues and *T. pallidum*. These antibodies are detected in the rapid plasma reagin (RPR) and Venereal Disease Research Laboratory (VDRL) tests. The nontreponemal assays are nonspecific but have been widely used due to low cost, ease of testing, and quantifiable results that can be used to follow response to treatment. Treponemal antibody tests measure antibodies that specifically react with *T. pallidum*. These include the fluorescent treponemal antibody absorption test and the *T. pallidum* enzyme immunoassay test. Automation of these more complex tests has led to their use as an initial screening test in some centers. A third option, rapid point-of-care tests, are designed for use with fingerstick blood (although serum shows higher sensitivity), are inexpensive, and are simple so that highly trained technologists are not required. There is only one FDA-approved point-of-care test in the United States. Rapid tests and PCR assays that have been developed are not commonly used at this time due to lower sensitivities and specificities.

The interpretation of these tests is complex because of differences in the antibody responses they measure and imperfections in the tests.

- Both treponemal and nontreponemal antibody tests are only moderately sensitive (~70% to 85%) for primary syphilis.
- Both types of test are very sensitive (>95%) for secondary syphilis.
- Treponemal tests are very sensitive for tertiary and latent syphilis. In contrast, nontreponemal antibody titers fall with time, and so nontreponemal tests are somewhat less sensitive for tertiary or latent syphilis.
- Nontreponemal antibody levels fall with successful treatment of syphilis, and so changes in the titers detected in these tests can be used to monitor therapy.
- Treponemal tests, which are nonquantitative, remain positive, even after successful therapy.
- Either nontreponemal or treponemal tests can be used for initial screening for syphilis, but positive results should be confirmed using a test of the other type (e.g., confirm nontreponemal positive test results with a treponemal test and *vice versa*). Confirmatory testing is needed because false-positive results can occur in both nontreponemal and treponemal tests. Causes of false-positive results in these tests include pregnancy, autoimmune diseases, and infections other than syphilis.

MORPHOLOGY

In primary syphilis, a chancre occurs on the penis or scrotum of 70% of men and on the vulva or cervix of 50% of women. The chancre is a slightly elevated, firm, reddened papule, up to several centimeters in diameter, that erodes to create a clean-based shallow ulcer. The contiguous induration creates a buttonlike mass directly adjacent to the eroded skin, providing the basis for the designation hard chancre (Fig. 8.34). On histologic examination, the chancre contains an intense infiltrate of plasma cells, with scattered macrophages and lymphocytes and a proliferative endarteritis. The endarteritis starts with endothelial cell activation and proliferation and progresses to intimal fibrosis (see Fig. 8.4). The regional nodes are usually enlarged due to nonspecific acute or chronic lymphadenitis, plasma cell–rich infiltrates, or granulomas.

In secondary syphilis, widespread mucocutaneous lesions involve the oral cavity, palms of the hands, and soles of the feet. The rash frequently consists of discrete red-brown macules less than 5 mm in diameter, but it may be follicular, pustular, annular, or scaling. Red lesions in the mouth or vagina contain the most organisms and are the most infectious. Histologically, the mucocutaneous lesions of secondary syphilis show the same plasma cell infiltrate and obliterative endarteritis as the primary chancre, although the inflammation is often less intense.

Tertiary syphilis most frequently involves the aorta, the CNS, and the liver, bones, and testes. The aortitis is caused by endarteritis of the vasa vasorum of the proximal aorta. Occlusion of the vasa vasorum results in scarring of the media of the proximal aortic wall, causing a loss of elasticity. There may be narrowing of the coronary artery ostia caused by subintimal scarring with resulting myocardial ischemia. The morphologic and

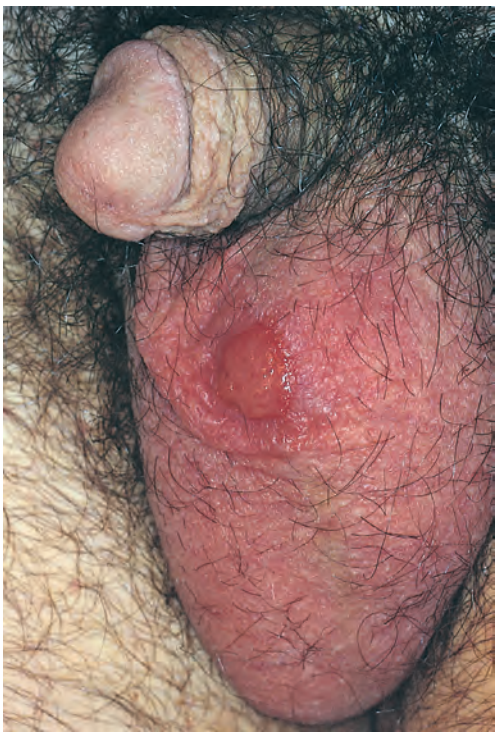


Figure 8.34 Syphilitic chancre in the scrotum (see Fig. 8.35 for the histopathology of syphilis). (Courtesy Dr. Richard Johnson, Beth Israel-Deaconess Hospital, Boston, Mass.)

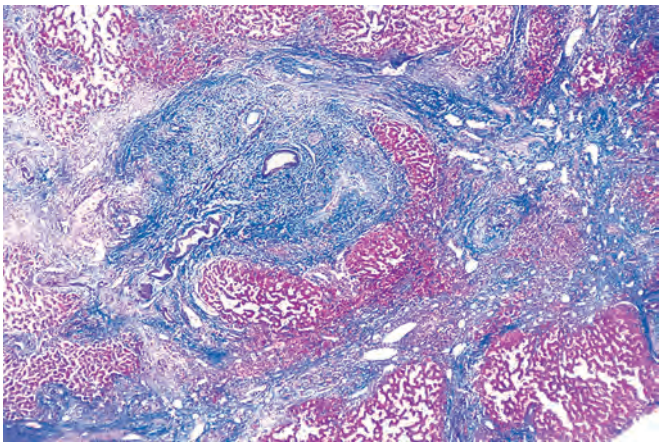


Figure 8.35 Trichrome stain of liver shows a gumma (scar), stained blue, caused by tertiary syphilis (the hepatic lesion is also known as *hepar lobatum*).

clinical features of syphilitic aortitis are discussed in greater detail with diseases of the blood vessels (Chapter 11).

Neurosyphilis takes one of several forms, designated meningovascular syphilis, tabes dorsalis, and general paresis (Chapter 28).

Syphilitic gummas are white-gray and rubbery, occur singly or multiply, and vary in size from microscopic lesions resembling tubercles to large tumorlike masses. They occur in most organs but particularly in skin, subcutaneous tissue, bone, and joints. In the liver, scarring as a result of gummas may cause a distinctive hepatic lesion known as *hepar lobatum* (Fig. 8.35). On histologic examination, the gummas have centers of coagulated, necrotic material and margins composed of plump, palisading macrophages and fibroblasts surrounded by large numbers of mononuclear leukocytes, chiefly plasma cells. Treponemes are scant in gummas and are difficult to demonstrate.

The rash of **congenital syphilis** is more severe than that of adult secondary syphilis. It is a bullous eruption of the palms and soles of the feet associated with epidermal sloughing. **Syphilitic osteochondritis and periostitis** affect all bones, but lesions of the nose and lower legs are most distinctive. Destruction of the vomer causes collapse of the bridge of the nose and, later on, the characteristic saddle nose deformity. Periostitis of the tibia leads to excessive new bone growth on the anterior surfaces and anterior bowing, or saber shin. There is also widespread disturbance in endochondral bone formation. The epiphyses become widened as the cartilage overgrows, and cartilage is found in displaced islands within the metaphysis.

The **liver** is often severely affected in congenital syphilis. Diffuse fibrosis permeates lobules to isolate hepatic cells into small nests, accompanied by the characteristic lymphoplasmacytic infiltrate and vascular changes. Gummas are occasionally found in the liver, even in early cases. The **lungs** may be affected by a diffuse interstitial fibrosis. In the syphilitic stillborn, the lungs appear pale and airless (pneumonia alba). The generalized spirochetemia may lead to diffuse interstitial inflammatory reactions in virtually any other organ (e.g., the pancreas, kidneys, heart, spleen, thymus, endocrine organs, and CNS).

The late manifestations of congenital syphilis include a distinctive **triad of interstitial keratitis, Hutchinson teeth, and eighth-nerve deafness**. In addition to interstitial keratitis, the ocular changes include choroiditis and abnormal retinal pigmentation. Hutchinson teeth are small incisors shaped like a screwdriver or a peg, often with notches in the enamel. Eighth-nerve deafness and optic nerve atrophy develop secondary to meningovascular syphilis.

Lyme Disease

Lyme disease is a common arthropod-borne illness caused by spirochetes in the genus *Borrelia* that can be localized or disseminated with a tendency to cause persistent chronic arthritis. It is named for the Connecticut town, Lyme, where there was an epidemic of arthritis associated with skin erythema in the mid-1970s. The common species of *Borrelia* that cause Lyme diseases are *Borrelia burgdorferi* in the United States, and additionally *B. afzelii* and *B. garinii* in Europe and Asia. These are each transmitted from rodents or, for *B. garinii* from birds, to people by *Ixodes* deer ticks. Lyme disease is endemic in the United States, Europe, and Asia. In the United States, there were between 30,000 and 40,000 cases reported each year for several years. Most cases occur in the northeastern states and the upper Midwest. In endemic areas, *B. burgdorferi* infects up to 50% of ticks, which may also be infected with *Ehrlichia* spp. and *Babesia* spp. Serology is the main method of diagnosis, but PCR can be done on infected tissue.

Lyme disease involves multiple organ systems and is divided into three stages (Fig. 8.36).

- In *early localized disease,* spirochetes multiply and spread in the dermis at the site of a tick bite, causing an expanding area of redness, often with a pale center. This lesion, called *erythema migrans,* may be accompanied by fever and lymphadenopathy. The rash spontaneously disappears in 4 to 12 weeks.
- In *early disseminated disease,* spirochetes spread hematogenously throughout the body and cause secondary skin lesions, lymphadenopathy, migratory joint and muscle pain, cardiac arrhythmias, and meningitis often associated with cranial nerve involvement. In Europe, *B. alfzelii* is associated with borrelial lymphocytoma, in which there is blue to red swelling of the earlobe or nipple, with lymphocytic infiltration.
- The third stage, *late disseminated disease,* manifests many months after the tick bite. *B. burgdorferi* usually causes a chronic arthritis sometimes with severe damage to large joints. Less often, patients will have polyneuropathy and encephalitis that vary from mild to debilitating. Lyme encephalopathy is less common in Europe than in the United States.

Pathogenesis

B. burgdorferi does not produce endotoxin or exotoxins that damage the host. Much of the pathology associated with the infection is thought to be secondary to the immune response against the bacteria and the inflammation that accompanies it. The initial immune response is stimulated by binding of bacterial lipoproteins to TLR2 on macrophages. In response, these cells release pro-inflammatory cytokines (IL-6 and TNF) and generate bactericidal reactive nitrogen intermediates, reducing but usually not eliminating the infection.

The inflammatory lesions are likely triggered by T cells and cytokines. *Borrelia*-specific antibodies, made 2 to 4 weeks

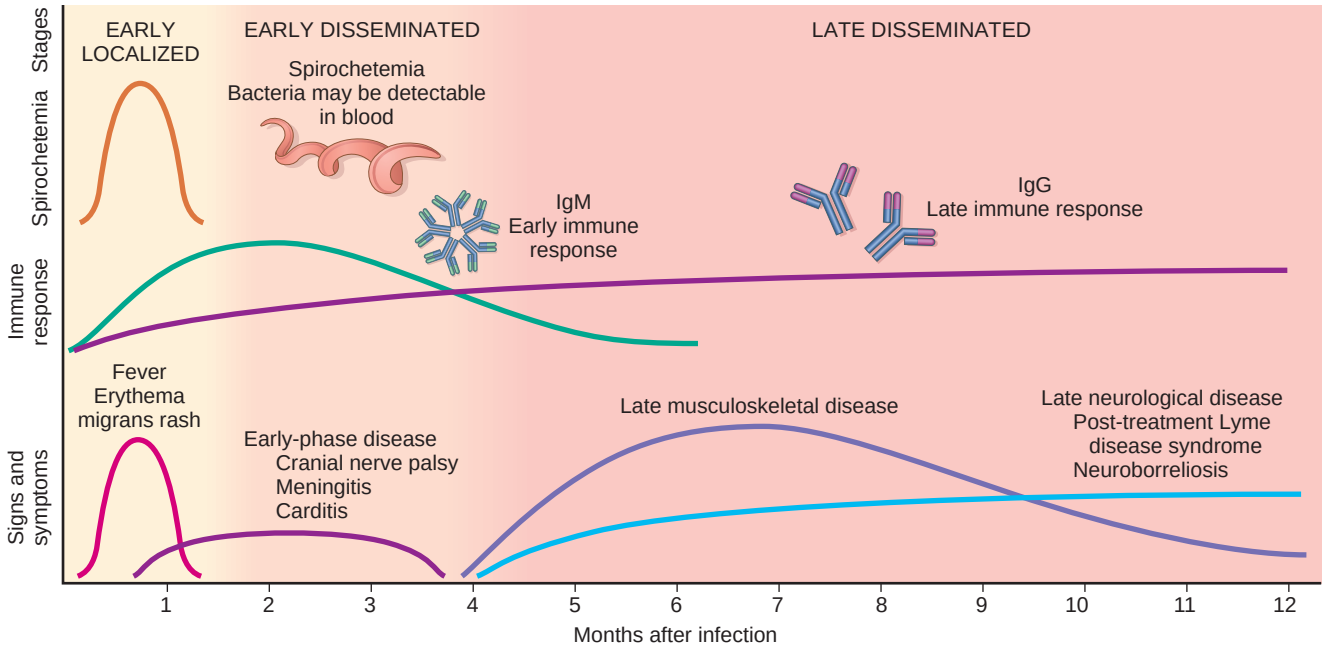


Figure 8.36 Lyme disease progresses through three clinically recognizable phases: early localized, early disseminated, and late disseminated. Although initial manifestations result directly from spirochete infection, later signs and symptoms are likely immune-mediated. (Figure modified from Dr. Charles Chiu, University of California, San Francisco, Calif. Used with permission.)

after infection, direct complement-mediated phagocytosis and killing of the bacteria; however, *B. burgdorferi* escapes the antibody response through antigenic variation. *B. burgdorferi* has a plasmid with a single promoter sequence and multiple coding sequences for an antigenic surface protein, VlsE, each of which can shuttle into position next to the promoter and be expressed. Thus, as the antibody response to one VlsE protein is mounted, bacteria expressing an alternate VlsE protein emerge and can escape immune recognition. Chronic manifestations of Lyme disease, such as late arthritis, may be caused by the immune response against some, as yet unknown, bacterial antigen that cross-reacts with a self antigen.

MORPHOLOGY

Skin lesions caused by *B. burgdorferi* are characterized by edema and a lymphocytic-plasma cell infiltrate. In early Lyme arthritis, the synovium resembles early rheumatoid arthritis, with villous hypertrophy, lining-cell hyperplasia, and abundant lymphocytes and plasma cells in the subsynovium. A distinctive feature of Lyme arthritis is an arteritis, which produces onionskin-like lesions resembling those seen in lupus (Chapter 6). In late Lyme disease, there may be extensive erosion of the cartilage in large joints. In Lyme meningitis, the CSF is hypercellular due to a marked lymphoplasmacytic infiltrate, and it contains antispirochete IgGs.

Anaerobic Bacterial Infections

Many anaerobic bacteria are normal microbiota in sites of the body that have low oxygen levels. The anaerobic microbiota cause disease (such as abscesses or peritonitis) when they are introduced into normally sterile sites or when the balance of organisms is upset and pathogenic anaerobes overgrow (e.g., *C. difficile* colitis with antibiotic treatment). Environmental anaerobes also cause disease (tetanus, botulism, and gas gangrene).

Abscesses

Commensal bacteria from adjacent sites (oropharynx, intestine, and female genital tract) are the usual cause of abscesses, so the species found in an abscess often reflect the normal microbiota in that site. Abscesses are usually caused by mixed anaerobic and facultative aerobic bacterial infections. Because most anaerobes that cause abscesses are part of the normal microbiota it is not surprising that these organisms do not produce significant toxins.

The bacteria found in head and neck abscesses reflect oral and pharyngeal microbiota. Common anaerobes at this site include the gram-negative bacilli *Prevotella* spp. and *Porphyromonas* spp., often mixed with the facultative *S. aureus* and *S. pyogenes*. *Fusobacterium necrophorum,* an oral commensal, causes Lemierre syndrome, characterized by infection of the lateral pharyngeal space and septic jugular vein thrombosis. Abdominal abscesses are caused by the anaerobes of the gastrointestinal tract, including gram-positive *Peptostreptococcus* spp. and *Clostridium* spp., as well as gram-negative *Bacteroides fragilis* and *E. coli*. Genital tract infections in women (e.g., Bartholin cyst abscesses and tubo-ovarian abscesses) are caused by anaerobic gram-negative bacilli, such as *Prevotella* spp., often mixed with *E. coli* or *S. agalactiae*.

MORPHOLOGY

Abscesses caused by anaerobes contain discolored and foul-smelling pus that is often poorly walled off. Otherwise, these lesions pathologically resemble those of the common pyogenic infections. Gram stain reveals mixed infection with gram-positive and gram-negative rods and gram-positive cocci mixed with neutrophils.

Clostridial Infections

Clostridium spp. are gram-positive bacilli that grow under anaerobic conditions and produce spores that are present in the soil. Four types of disease are caused by *Clostridium* spp.:

- *C. perfringens, C. septicum,* and other species cause cellulitis and myonecrosis of traumatic and surgical wounds *(gas gangrene),* uterine myonecrosis often associated with illegal abortions, mild food poisoning, and infection of the small bowel associated with ischemia or neutropenia that often leads to severe sepsis. Spontaneous gas gangrene due to *C. septicum* is highly associated with an underlying malignancy.
- *C. tetani,* the cause of *tetanus,* proliferates in puncture wounds and in the umbilical stump of newborn infants and releases a potent neurotoxin that causes increased muscle tone and generalized spasms of skeletal muscles (lockjaw). Tetanus toxoid (formalin-fixed tetanus toxin) is part of the DPT (diphtheria, pertussis, and tetanus) immunization, and this has greatly decreased the incidence of tetanus worldwide.
- *C. botulinum*, the cause of *botulism,* grows in inadequately cooked foods and releases a potent neurotoxin that blocks synaptic release of acetylcholine and causes flaccid paralysis of respiratory and skeletal muscles.
- *C. difficile* proliferates in the intestines when competition by normal commensal microbiota is reduced by antibiotic treatment. The bacteria release toxin and cause *pseudomembranous colitis* (Chapter 17).

Clostridial infections can be diagnosed by culture (cellulitis, myonecrosis), PCR assays that detect the toxin (pseudomembranous colitis), or both (botulism).

Pathogenesis

C. perfringens does not grow in the presence of oxygen, so tissue death is essential for growth of the bacteria in the host. These bacteria release collagenase and hyaluronidase that degrade extracellular matrix proteins and contribute to bacterial invasiveness, but their most powerful virulence factors are the many toxins they produce. *C. perfringens* secretes 14 toxins, the most important of which is α-toxin. This toxin has multiple actions. It is a phospholipase that degrades lecithin, a major component of cell membranes, and so destroys red cells, platelets, and muscle cells, causing myonecrosis. It also has a sphingomyelinase activity that contributes to nerve sheath damage.

Ingestion of food contaminated with *C. perfringens* causes a brief diarrhea. Spores, usually in contaminated meat, survive cooking, and the organism proliferates in cooling food. *C. perfringens* enterotoxin forms pores in the epithelial cell membranes, lysing the cells and disrupting tight junctions between epithelial cells.

The neurotoxins produced by *C. botulinum* and *C. tetani* both inhibit release of neurotransmitters, resulting in paralysis. Botulism toxin, eaten in contaminated foods or absorbed from wounds infected with *C. botulinum,* binds gangliosides on motor neurons and is transported into the cell. In the cytoplasm, the A fragment of botulism toxin cleaves a protein called synaptobrevin that mediates fusion of neurotransmitter-containing vesicles with the neuron membrane. By blocking vesicle fusion, botulism toxin prevents the release of acetylcholine at the neuromuscular junction, resulting in flaccid paralysis. If the respiratory muscles are affected, botulism can lead to death. The widespread use of botulism toxin (Botox) in cosmetic surgery is based on its ability to cause paralysis of strategically chosen muscles on the face. Tetanus toxin causes a violent spastic paralysis by blocking release of γ-aminobutyric acid, a neurotransmitter that inhibits motor neurons.

C. difficile produces toxin A, an enterotoxin that stimulates chemokine production and thus attracts leukocytes, and toxin B, a cytotoxin that causes distinctive cytopathic effects in cultured cells. Both toxins are glucosyl transferases and are part of a pathogenicity island that is absent from the chromosomes of nonpathogenic strains of *C. difficile*. There has been significant success in treatment of *C. difficile* with fecal transplantation, emphasizing the role of normal commensal microbiota in defense against pathogens.

MORPHOLOGY

The most significant lesions are caused by *C. perfringens;* these are described next. **Clostridial cellulitis,** which originates in wounds, can be differentiated from infection caused by pyogenic cocci by its foul odor, its thin, discolored exudate, and the relatively quick and wide tissue destruction. On microscopic examination, the amount of tissue necrosis is disproportionate to the number of neutrophils and gram-positive bacteria present (Fig. 8.37). Clostridial cellulitis, which often has granulation tissue at its borders, is treatable by debridement and antibiotics.

In contrast, **clostridial gas gangrene** is life-threatening and is characterized by marked edema and enzymatic necrosis of involved muscle cells 1 to 3 days after injury. An extensive fluid exudate, which is lacking in inflammatory cells, causes swelling of the affected region and the overlying skin, which develops large bullous vesicles that rupture. Gas bubbles caused by bacterial fermentation appear within the gangrenous tissues. As the infection progresses, the inflamed muscles become soft, blue-black, friable, and semifluid as a result of the massive proteolytic action of the released bacterial enzymes. On microscopic examination, there is severe **myonecrosis,** extensive hemolysis, and marked vascular injury with thrombosis. *C. perfringens* is also associated with

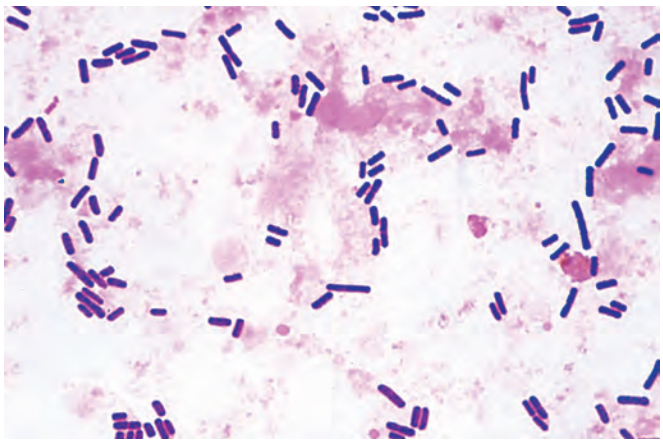


Figure 8.37 Boxcar-shaped gram-positive *Clostridium perfringens* intermingled with necrotic debris in gangrenous tissue.

dusk-colored, wedge-shaped infarcts in the small bowel, particularly in neutropenic people. Regardless of the site of entry, when *C. perfringens* disseminates hematogenously, there is widespread formation of gas bubbles.

Despite the severe neurologic damage caused by botulinum and tetanus toxins, the neuropathologic changes are subtle and nonspecific.

Obligate Intracellular Bacterial Infections

Obligate intracellular bacteria proliferate only within host cells, although some may survive outside of cells. These organisms are well adapted to the intracellular environment, with membrane pumps to capture amino acids and ATP for energy. Some are unable to synthesize ATP at all (e.g., *Chlamydia* spp.), and others synthesize at least some of their own ATP (e.g., the rickettsiae).

Chlamydial Infections

C. trachomatis is a small gram-negative bacterium that is an obligate intracellular pathogen. *C. trachomatis* exists in two forms during its unique life cycle. The infectious form, called the elementary body, is metabolically inactive. Elementary bodies use a type III secretion system to inject a protein, TARP, into host cells, leading to actin remodeling at the site of bacterial entry into the host cell by phagocytosis. The elementary body differentiates into a metabolically active form, called the reticulate body, within a membrane-bound inclusion derived from the endosome. The bacteria modify the inclusion to prevent it from trafficking as an endosome and maturing into an inhospitable phagolysosome. Host lipids, typically found in the Golgi apparatus, and GTPases typical of recycling endosomes and the Golgi apparatus are recruited to the inclusion, marking the inclusion as a nonphagocytic compartment. The reticulate body uses ATP and amino acids from the host cell to replicate and forms new infectious elementary bodies that lyse or are extruded from the infected cell.

The diseases caused by *C. trachomatis* infection are associated with different serotypes of the bacteria: urogenital infections and inclusion conjunctivitis are caused by serotypes D to K, lymphogranuloma venereum by serotypes L1, L2, and L3, and an ocular infection of children, trachoma, by serotypes A, B, and C. The venereal infections caused by *C. trachomatis* are discussed here.

Genital infection by *C. trachomatis* is the most common bacterial STI in the world. In 2016, approximately 1.6 million cases of genital chlamydia were reported to the CDC. The rate of genital chlamydia in the United States has risen fairly steadily over the past 10 years. Genital *C. trachomatis* infections (other than lymphogranuloma venereum, discussed later) are associated with clinical features that are similar to those caused by *N. gonorrhoeae*. Patients may develop epididymitis, prostatitis, pelvic inflammatory disease, pharyngitis, conjunctivitis, perihepatic inflammation, and proctitis. Unlike *N. gonorrhoeae* urethritis, *C. trachomatis* urethritis in men may be asymptomatic and so may go untreated. Both *N. gonorrhoeae* and *C. trachomatis* frequently cause asymptomatic infections in women. *C. trachomatis* urethritis can be diagnosed using amplified nucleic acid tests performed on genital swabs or urine specimens.

Genital infection with the L serotypes of *C. trachomatis* causes *lymphogranuloma venereum*, a chronic, ulcerative disease. Lymphogranuloma venereum is a sporadic disease in the United States and Western Europe, but it is endemic in parts of Asia, Africa, the Caribbean region, and South America. The infection initially manifests as a small, often unnoticed, papule on the genital mucosa or nearby skin. Two to 6 weeks later, growth of the organism and the host response in draining lymph nodes produce swollen, tender lymph nodes, which may coalesce and rupture. If not treated, the infection can subsequently cause fibrosis and strictures in the anogenital tract. Rectal strictures are particularly common in women.

MORPHOLOGY

The features of *C. trachomatis* **urethritis** are virtually identical to those of gonorrhea. The primary infection is characterized by a mucopurulent discharge containing a predominance of neutrophils. Organisms are not visible in Gram-stained smears or sections.

The lesions of **lymphogranuloma venereum** contain a mixed granulomatous and neutrophilic inflammatory response. Variable numbers of chlamydial inclusions are seen in the cytoplasm of epithelial cells or inflammatory cells. Regional lymphadenopathy is common, usually occurring within 30 days of infection. Lymph node involvement is characterized by a granulomatous inflammatory reaction associated with irregularly shaped foci of necrosis containing neutrophils (stellate abscesses). With time, the inflammatory reaction is dominated by nonspecific chronic inflammatory infiltrates and extensive fibrosis. The latter, in turn, may cause local lymphatic obstruction, lymphedema, and strictures. In active lesions, the diagnosis of lymphogranuloma venereum may be made by demonstration of the organism in biopsy sections or smears of exudate. In more chronic cases, the diagnosis rests with the demonstration of antibodies to the appropriate chlamydial serotypes in the patient's serum.

Infections Caused by Other Intracellular Bacteria

Rickettsia spp., *Orientia* spp., *Ehrlichia* spp., and *Anaplasma* spp. are vector-borne obligate intracellular bacteria that cause epidemic and scrub typhus, spotted fevers (*Rickettsia rickettsii* and others), ehrlichiosis, and anaplasmosis. These organisms have the structure of gram-negative, rod-shaped bacteria, although they stain poorly with Gram stain.

- *Epidemic typhus* (caused by *Rickettsia prowazekii*) is transmitted from person to person by body lice. It is associated with wars and poverty, when individuals live in close contact with poor hygiene. Manifestations include a rash that is initially macular, progressing to a petechial, maculopapular rash on the entire body except the face, palms, and soles.
- *Scrub typhus* (caused by *Orientia tsutsugamushi*) is transmitted by chiggers (mites). It is endemic in areas of Asia and Australia, and it has recently emerged in South Korea, Chile, and areas of China. Fever, headache, myalgia, and cough are usual symptoms, sometimes accompanied by a characteristic eschar and associated lymphadenopathy from the chigger bite.
- *Rocky Mountain spotted fever* (caused by *Rickettsia rickettsii*) is transmitted to humans by dog ticks. It is most common

in the southeastern and south-central United States. It begins as a nonspecific severe illness with fever, myalgias, and gastrointestinal distress, and then progresses to a widespread macular then petechial rash that can involve the palms and soles.

- *Ehrlichiosis* (caused mainly by *Ehrlichia chaffeensis*) and *anaplasmosis* (*Anaplasma phagocytophilum*) are transmitted by the lone star tick and deer tick, respectively. These bacteria predominantly infect monocytes *(E. chaffeensis)* or neutrophils *(A. phagocytophilum)*. Ehrlichiosis and anaplasmosis are characterized by abrupt onset of fever, headache, and malaise, and may progress to respiratory insufficiency, renal failure, and shock. Rash occurs in approximately 40% of people with *E. chaffeensis* infections.

Rickettsial diseases are usually diagnosed clinically and confirmed by serology or immunostaining of the organisms.

Pathogenesis

The severe manifestations of rickettsial infections are primarily due to infection of endothelial cells and the consequent endothelial dysfunction and injury. The rickettsiae that cause typhus and spotted fevers predominantly infect vascular endothelial cells, especially those in the lungs and brain. The bacteria enter cells by endocytosis, but escape from the endosome into the cytoplasm using hemolysins to disrupt phagosomal membranes. The organisms proliferate in the endothelial cell cytoplasm and then either lyse the cell (typhus group) or spread from cell to cell using actin-based motility (spotted fever group). The widespread endothelial dysfunction can cause shock, peripheral and pulmonary edema, and DIC, as well as renal failure and a variety of CNS manifestations that can include coma.

The innate immune response to rickettsial infection is complex, but includes NK cells, which produce IFN-γ. Subsequent CTL responses are critical for elimination of rickettsial infections. IFN-γ and TNF produced by activated NK cells and T cells stimulate the production of bactericidal nitric oxide derivatives. CTLs lyse infected cells, reducing bacterial proliferation.

MORPHOLOGY

Typhus Fever. In mild cases, the gross changes are limited to a rash and small hemorrhages due to the vascular lesions. In more severe cases, there may be areas of necrosis of the skin and gangrene of the tips of the fingers, nose, earlobes, scrotum, penis, and vulva. In such cases, irregular ecchymotic hemorrhages may be found internally, principally in the brain, heart muscle, testes, serosal membrane, lungs, and kidneys.

The most prominent microscopic changes are small vessel lesions and focal areas of hemorrhage and inflammation in various organs and tissues. Endothelial swelling in the capillaries, arterioles, and venules may narrow the lumens of these vessels. A cuff of mononuclear inflammatory cells usually surrounds the affected vessel. The vascular lumens are sometimes thrombosed. Necrosis of the vessel wall is unusual in typhus (compared with Rocky Mountain spotted fever). Vascular thromboses lead to gangrenous necrosis of the skin and other structures in a minority of cases. In the brain, characteristic typhus nodules are composed of focal microglial proliferations with an infiltrate of mixed T lymphocytes and macrophages (Fig. 8.38).

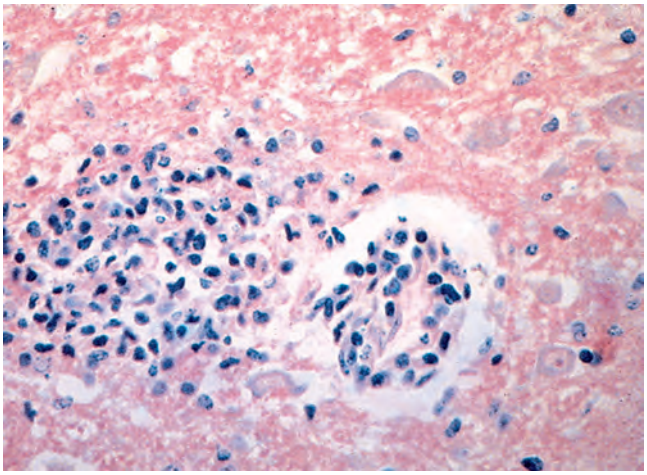


Figure 8.38 Typhus nodule in the brain.

Scrub typhus, or mite-borne infection, is usually a milder version of typhus fever. The rash is usually transitory or might not appear. Vascular necrosis or thrombosis is rare, but there may be a prominent inflammatory lymphadenopathy.

Rocky Mountain Spotted Fever. A hemorrhagic rash that extends over the entire body, including the palms of the hands and soles of the feet, is the hallmark of Rocky Mountain spotted fever. An eschar at the site of the tick bite is uncommon with Rocky Mountain spotted fever but is often seen with *R. akari, R. africae,* and *R. conorii* infection. The vascular lesions that underlie the rash often lead to acute necrosis, fibrin extravasation, and occasionally thrombosis of the small blood vessels, including arterioles (Fig. 8.39). In severe Rocky Mountain spotted fever, foci of necrotic skin appear, particularly on the fingers, toes, elbows, ears, and scrotum. The perivascular inflammatory response, similar to that of typhus, is seen in the brain, skeletal muscle, lungs, kidneys, testes, and heart muscle. The vascular lesions in the brain may involve larger vessels and produce microinfarcts. A noncardiogenic pulmonary edema causing adult respiratory distress

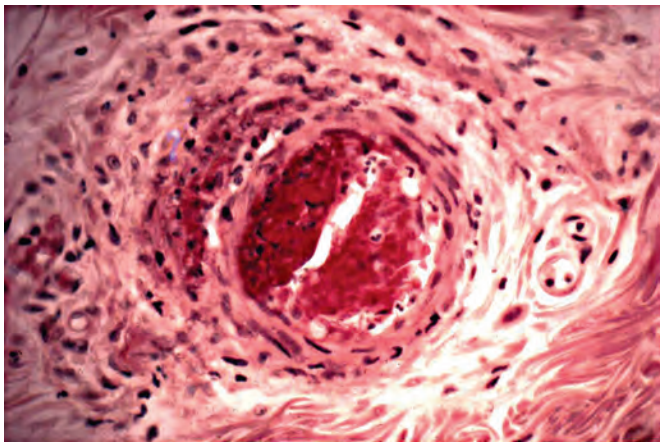


Figure 8.39 Rocky Mountain spotted fever with a thrombosed vessel and vasculitis.

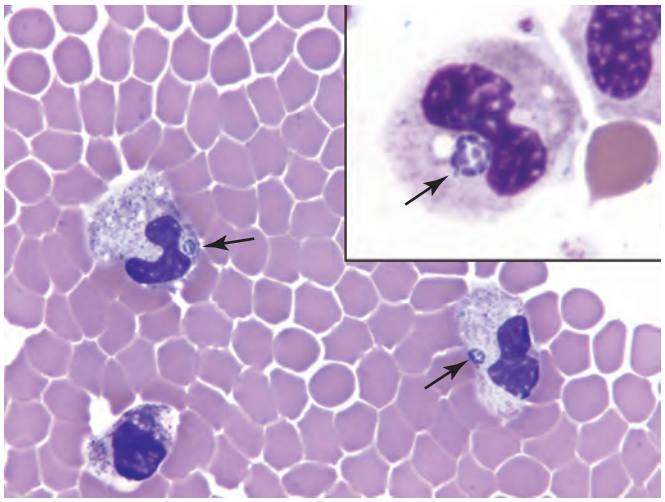


Figure 8.40 Peripheral blood granulocytes (band neutrophils) containing *Anaplasma* inclusions *(arrows).* (Courtesy Dr. Tad Weiczorek, Faulkner Hospital, Boston, Mass.)

syndrome is the major cause of death in patients with Rocky Mountain spotted fever.

Ehrlichiosis and **anaplasmosis** have similar presentations. The rash is nonspecific and can be macular, maculopapular, or petechial. Characteristic cytoplasmic inclusions (morulae), composed of masses of bacteria that occasionally take the shape of a mulberry, are present in leukocytes (Fig. 8.40).

FUNGAL INFECTIONS

Fungi are eukaryotes that grow as multicellular filaments (mold) or individual cells alone or in chains (yeast). Cell walls give fungi their shape. Yeasts are round to oval and mainly reproduce by budding. Some yeasts, such as *C. albicans,* can produce buds that fail to detach and become elongated, producing a chain of elongated yeast cells called *pseudohyphae*. Molds consist of threadlike filaments (hyphae) that grow and divide at their tips. They can produce round cells called *conidia* that easily become airborne, disseminating the fungus. Many medically important fungi are *dimorphic,* existing as yeast or molds, depending on environmental conditions (yeast forms at human body temperature and mold forms at room temperature). Fungal infections can be diagnosed by histologic examination, although definitive identification of some species requires culture.

Fungal infections, also called *mycoses,* are of four major types:

- *Superficial and cutaneous mycoses* are common and limited to the very superficial or keratinized layers of skin, hair, and nails.
- *Subcutaneous mycoses* involve the skin, subcutaneous tissues, and lymphatics and rarely disseminate systemically.
- *Endemic mycoses* are caused by dimorphic fungi that can produce serious systemic illness in healthy individuals.
- *Opportunistic mycoses* can cause life-threatening systemic diseases in individuals who are immunosuppressed or who carry implanted prosthetic devices or vascular catheters. Some of the fungi that cause opportunistic mycoses are discussed in the following sections; those involving specific organs are discussed in other chapters.

Yeast Infections

Candidiasis

***C. albicans* is the most prevalent fungal pathogen of humans.** Although there are more than 200 species of *Candida,* there are about 15 to 20 species that are frequently seen in human infections, with *C. glabrata, C. tropicalis, C. parapsilosis,* and *C. krusei* being more frequent. This section will focus on *C. albicans.*

Most *C. albicans* infections originate when these normal commensal microbiota breach the skin or mucosal barriers. Residing normally in the skin, mouth, gastrointestinal tract, and vagina, *Candida* spp. usually live as benign commensals and seldom produce disease in healthy people. These infections may be confined to the skin or mucous membranes or may disseminate widely. In otherwise healthy people, *C. albicans* causes vaginitis and diaper rash. Individuals with diabetes and burn patients are particularly susceptible to superficial candidiasis. In individuals with indwelling intravenous lines or catheters, or undergoing peritoneal dialysis, *C. albicans* can spread to the bloodstream. Severe disseminated candidiasis most commonly occurs in patients who are neutropenic due to leukemia, chemotherapy, or hematopoietic stem cell transplantation, and may cause shock and DIC.

Pathogenesis

A single strain of *C. albicans* can be successful as a commensal or a pathogen. There is no known environmental reservoir for *C. albicans,* unlike other *Candida* spp., and *C. albicans* has developed multiple adaptive mechanisms to complete its life cycle fully within the human host. *C. albicans* can shift between nine distinct cell shapes. Phenotypic switching involves coordinated transcriptional regulation of phase-specific genes and provides a way for *C. albicans* to adapt to changes in the environment such as temperature, nutrient availability, antibiotic therapy, or the immune response. These variants can exhibit altered colony morphology, cell shape, antigenicity, and virulence.

C. albicans produces a large number of functionally distinct adhesins involved in binding to fibrinogen, fibronectin, laminin, epithelial cells, and endothelial cells. *C. albicans* also produces a number of enzymes that contribute to invasiveness, including at least nine secreted aspartyl proteinases, which may promote tissue invasion by degrading extracellular matrix proteins, and catalases, which may enable the organism to resist oxidative killing by phagocytic cells.

The ability of *C. albicans* to grow as biofilms also contributes to its capacity to cause disease, with at least 30 factors identified to play a role in adhesion, maturation, and dispersion that affect biofilm formation. The biofilms are microbial communities consisting of mixtures of yeast, filamentous forms, and fungal-derived extracellular matrix. *C. albicans* can form biofilms on implanted medical devices that reduce the organism's susceptibility to immune responses and antifungal drug therapy.

Neutrophils, macrophages, and Th17 cells are important for protection against *Candida* infection.

- Neutrophils and macrophages phagocytose *C. albicans*, and oxidative killing by these phagocytes is a first line of host defense. The important role of neutrophils and macrophages is illustrated by the increased risk of *C. albicans* infection in individuals with neutropenia or defects in NADPH oxidase or myeloperoxidase. Filamentous forms, but not yeast, can escape from phagosomes and enter the cytoplasm and proliferate.
- *C. albicans* yeast forms activate dendritic cells through multiple pathways, more so than do the filamentous forms of the fungi. For example, β-1,3-glucan expressed by the yeast engages dectin on dendritic cells and induces IL-6 and IL-23 production, which promotes Th17 responses. The Th17 responses elicited by *C. albicans* promote the recruitment of neutrophils and monocytes (Chapter 6). These responses are critical for protection against *C. albicans* infection, as shown by recurrent mucocutaneous candidiasis in individuals with either low T-cell counts due to HIV infection or inherited defects in Th17 cell development.

MORPHOLOGY

In tissue sections, *C. albicans* can appear as yeast, pseudohyphae, and, less commonly, true hyphae, defined by the presence of septae, such as under reduced oxygen tension (Fig. 8.41). Pseudohyphae, an important diagnostic clue, are a chain of budding yeast cells joined end to end at constrictions. All forms may be present together in the same tissue. The organisms may be seen in routine hematoxylin and eosin stains, but a variety of special fungal stains (Gomori methenamine-silver, periodic acid-Schiff) are commonly used to better visualize them.

Most commonly, candidiasis takes the form of a superficial infection on mucosal surfaces of the oral cavity **(thrush).** Florid proliferation of the fungi creates gray-white, dirty-looking pseudomembranes composed of matted organisms and inflammatory debris. Deep under the surface, there is mucosal hyperemia and inflammation. This form of candidiasis is seen in newborns, debilitated people, children receiving oral steroids for asthma, and following a course of broad-spectrum antibiotics that destroys competing normal bacterial microbiota. The other major risk group includes HIV-positive patients; people with oral thrush for no obvious reason should be evaluated for HIV infection.

C. albicans **esophagitis** is commonly seen in AIDS patients and in those with hematologic malignancies. These patients present with dysphagia (painful swallowing) and retrosternal pain; endoscopy demonstrates white plaques and pseudomembranes resembling oral thrush on the esophageal mucosa (see Fig. 8.41).

C. albicans **vaginitis** is common, especially in women who are diabetic, pregnant, or on oral contraceptive pills. It is usually associated with intense itching and a thick, curdlike discharge.

Cutaneous candidiasis can present in many different forms, including infection of the nail proper (onychomycosis); nail folds (paronychia); hair follicles (folliculitis); moist, intertriginous skin,

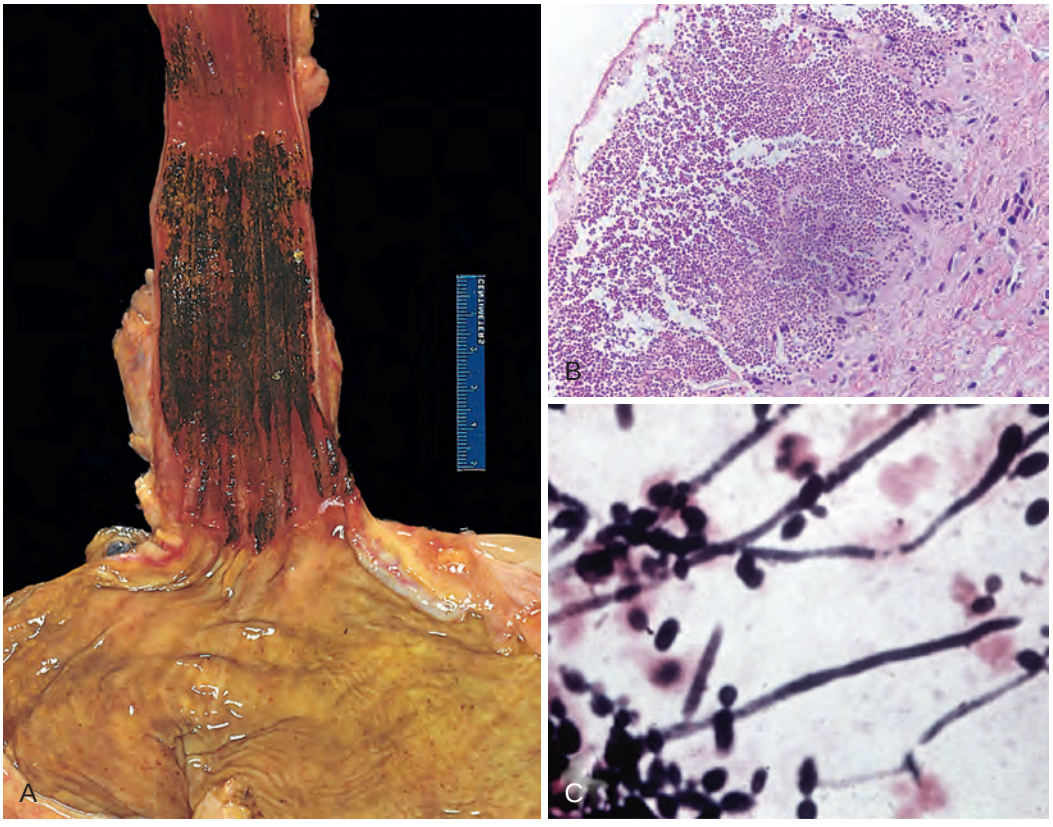


Figure 8.41 The morphology of Candida infections. (A) Severe candidiasis of the distal esophagus. (B) Hematoxylin and eosin stain of esophageal candidiasis reveals the dense mat of *Candida* spp. (C) Characteristic pseudohyphae and budding yeast of *Candida* spp. (C, Courtesy Dr. Dominick Cuvuoti, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

such as armpits or webs of the fingers and toes (intertrigo); and penile skin (balanitis). Diaper rash is a cutaneous candidal infection seen in the perineum of infants, the region in contact with wet diapers.

Invasive candidiasis is caused by blood-borne dissemination of organisms to various tissues or organs. Common patterns include (1) renal abscesses, (2) myocardial abscesses and endocarditis, (3) brain microabscesses and meningitis, (4) endophthalmitis (virtually any eye structure can be involved), and (5) hepatic abscesses. In any of these locations, depending on the immune status of the infected person, the fungus may evoke little inflammation, cause the usual suppurative response, or occasionally produce granulomas. People with acute leukemias who are profoundly neutropenic after chemotherapy are particularly prone to developing systemic disease. *Candida* spp. endocarditis is the most common fungal endocarditis, usually occurring in the setting of prosthetic heart valves or in intravenous drug abusers. In the latter group, the tricuspid valve is involved.

Candida auris Infections

C. auris is an emerging pathogen associated with multiple nosocomial infections on five continents. Genomic analysis indicates that separate clades of this organism have emerged simultaneously in different geographic regions. Antifungal resistance of this organism and difficulty in identification of the species with traditional laboratory diagnostics have heightened concern. Increased colonization and infection with non-*albicans Candida* spp. are attributed in part to increased use of prophylactic antifungal agents. There is an association of *C. auris* with intensive care units and central venous or urinary catheters, so biofilm formation is a suggested virulence mechanism of this organism. There has been variable resistance reported, but resistance to triazole antifungals and variable susceptibility to amphotericin B has been demonstrated in a number of studies. Colonization with *C. auris* has been reported in nares, groin, axilla, and rectum. Mortality in patients with invasive *C. auris* infection varies, but with some rates reported as high as 50%; attributable mortality has been difficult to determine in a vulnerable patient population with complex comorbidities. *C. auris* has been demonstrated to survive on dry and moist surfaces up to 14 days, so environmental cleaning to eliminate a source of nosocomial infections is a challenge.

Cryptococcosis

Two species of cryptococcus are known to cause disease in humans, *C. neoformans* and *C. gattii,* both of which grow as encapsulated yeasts. It has long been recognized that although *C. neoformans* may cause meningoencephalitis in otherwise healthy individuals, it more frequently presents as an opportunistic infection in people with AIDS, leukemia, lymphoma, systemic lupus erythematosus, or sarcoidosis, as well as in immunosuppressed transplant recipients. Many of these patients receive high-dose corticosteroids, a major risk factor for *C. neoformans* infection. It is estimated that there are more than 220,000 cases of cryptococcal meningitis occurring worldwide each year, with more than 180,000 associated deaths. *C. neoformans* is present in the soil and in bird (particularly pigeon) droppings and infects people when it is inhaled.

C. gattii is an obscure infectious agent that was classically viewed as a tropical or subtropical fungus until 1999, when it was identified as the cause of an outbreak of cryptococcal disease in the American Pacific Northwest and contiguous areas of British Columbia. It has subsequently been linked to cryptococcal infections in other regions of the world. Because most current tests used to diagnose cryptococcal infections do not distinguish between *C. gatti* and *C. neoformans,* the true incidence of infections caused by these two agents is currently uncertain. Based on findings from areas where *C. gattii* is now specifically monitored, it appears that *C. gattii* is more likely than *C. neoformans* to cause disease in immunologically normal individuals and to present with large lesions that produce mass effects or that mimic the radiologic appearance of a neoplasm. *C. gattii* is associated with certain species of trees, is found in soil, and, like *C. neoformans,* is acquired by inhalation.

Pathogenesis

Cryptococcus spp. have several virulence factors that enable it to evade host defenses, as follows:

- *Polysaccharide capsule.* Glucuronoxylomannan inhibits phagocytosis by alveolar macrophages, leukocyte migration, and recruitment of inflammatory cells. *Cryptococcus* spp. can block dendritic cell maturation by reducing MHC class II–dependent antigen presentation and inhibiting the production of IL-12 and IL-23. *Cryptococcus* spp. can make large cells, called *Titan cells,* that are greater than 12 µm and have a thickened cell wall. *Cryptococcus* spp. also produce small (micro) cells of 2 to 4 µm that may be adapted for growth in macrophages.
- *Melanin production.* Laccase in the yeast catalyzes the formation of melanin which (1) has antioxidant properties, (2) decreases antibody-mediated phagocytosis, (3) counteracts the effects of antifungal agents, (4) binds iron, (5) and provides cell wall integrity.
- *Enzymes.* Phospholipases degrade cell wall components and may aid tissue invasion. Urease helps neutralize the reactive oxygen species and pH of the phagocytic cell.
- *Differential cellular response to phagocytes.* A mechanism has been hypothesized to explain the success of the *C. gattii* strain in the Northwestern U.S. outbreak: In response to reactive oxygen species in the phagocyte, some cells stop growing and acquire an unusual morphology with tubularization of mitochondria, and other cells rapidly divide. Further investigation of these pathogenic pathways is needed for complete understanding.

MORPHOLOGY

Cryptococcus spp. have yeast forms, but not pseudohyphal or hyphal forms, in human hosts. **The typically 5- to 10-µm cryptococcal yeast form has a highly characteristic thick gelatinous capsule containing a polysaccharide that stains intense red with periodic acid-Schiff and mucicarmine in tissues** (Fig. 8.42) and can be detected in blood or CSF with various immunoassays. Although the lung is the primary site of infection, pulmonary involvement is usually mild and asymptomatic, even while the fungus is spreading to the CNS. *C. gattii* appears to be particularly likely to form a solitary pulmonary granuloma similar to the circumscribed (coin) lesions caused by *Histoplasma* spp.

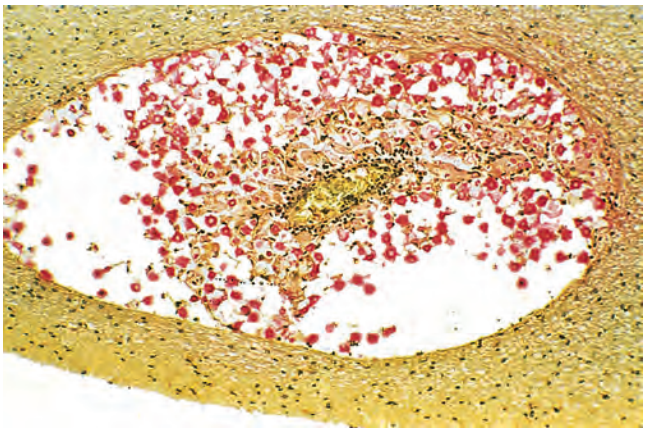


Figure 8.42 Mucicarmine stain of cryptococci *(staining red)* in a Virchow-Robin perivascular space of the brain (soap-bubble lesion).

The major lesions caused by *Cryptococcus* spp. are in the CNS, involving the meninges, cortical gray matter, and basal nuclei. The host response to cryptococci is extremely variable. In immunosuppressed people, organisms may evoke virtually no inflammatory reaction, so gelatinous masses of fungi grow in the meninges or expand the perivascular Virchow-Robin spaces within the gray matter, producing the so-called soap-bubble lesions (see Fig. 8.42). In severely immunosuppressed persons, *C. neoformans* may disseminate widely to the skin, liver, spleen, adrenals, and bones. In immunocompetent people or in those with protracted disease, the fungi induce a chronic granulomatous reaction composed of macrophages, lymphocytes, and foreign body-type giant cells. Suppuration also may occur, as well as a rare granulomatous arteritis of the circle of Willis.

Pneumocystis Infections

***P. jirovecii* is a yeastlike fungus that primarily causes lung infections and is a significant opportunistic infection in AIDS patients, despite the overall improvements seen with antiretroviral therapy.** The organism can cause a rapidly progressive, bilateral pneumonia. The organism was originally classified as a protozoal parasite, and descriptions of developmental forms reflect this historical classification. Three forms of the organism include trophozoites of 1 to 4 µm, sporocytes of 5 to 6 µm, and cysts of 5 to 8 µm. The cysts have a characteristic cup-shaped appearance, or they are oval with a central dot. Clusters of organisms in bronchoalveolar lavage fluid may be 200 µm in diameter. The nucleus and mitochondria are visible by Wright-Giemsa stain. The trophic forms are not visible with a cell wall stain such as methenamine silver, however, the sporocyst and ascus forms will be visible (Fig. 8.43). Histopathologic findings include alveolar interstitial thickening and eosinophilic honeycombed exudate in the lumen of the lung. Fluorescein-conjugated antibody stains are commonly used to diagnose these infections. Beta-D-glucan will be elevated with infection, although it is not specific for *Pneumocystis* spp., and sensitive and specific PCR tests are available for diagnosis.

Many *Pneumocystis* spp. exist, and each species is host-specific. No environmental source or external reservoir outside of humans has been identified for *P. jirovecii,* and the lack of a continuous in vitro culture method has hindered research of the pathogenesis. Transmission is believed to be via an airborne route, and there is evidence that healthy individuals might contribute to the reservoir. Most people are infected transiently early in life, with subsequent effective clearance. The average mortality rate of symptomatic infection is 10% to 14%, but it is higher in lower-income countries and urban communities. Extrapulmonary findings with *P. jirovecii* infection are not common, but can include involvement of lymph nodes, spleen, bone marrow, and liver, as well as other sites. Infection with *P. jirovecii* induces both a humoral and cellular immune response.

Mold Infections

The medically important dimorphic fungi are discussed in Chapter 15 and include *Blastomyces dermatitidis, H. capsulatum,* and *Coccidioides immitis*. This section discusses some important hyaline (transparent) molds.

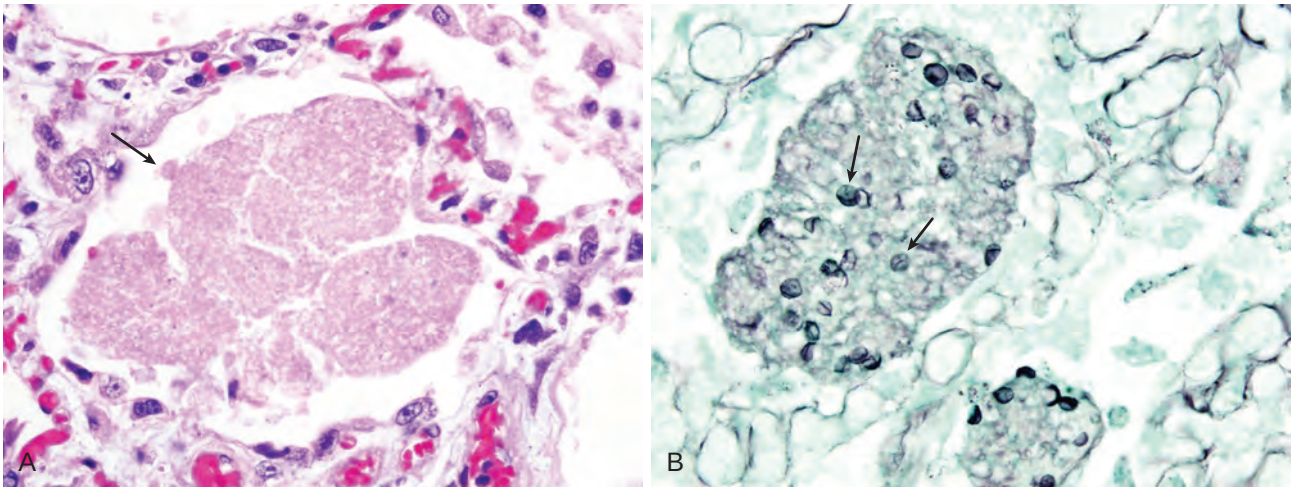


Figure 8.43 Pneumonia caused by *Pneumocystis jirovecii.* (A) Alveoli of lung filled with foamy exudate *(arrow)* and interstitial inflammation. (B) Silver stain of alveoli, showing cysts of *Pneumocystis jirovecii* stained black *(arrows).* (From Procop GW, Pritt BS: *Pathology of Infectious Diseases,* Philadelphia, 2015, Elsevier; Courtesy Dr. Ann M. Nelson, Joint Pathology Center, Silver Spring, Md.)

Aspergillosis

***Aspergillus* is a ubiquitous mold that causes allergies (allergic bronchopulmonary aspergillosis) in otherwise healthy people and serious sinusitis, pneumonia, and invasive disease in immunocompromised individuals.** The major conditions that predispose to *Aspergillus* infection are neutropenia and use of corticosteroids. *Aspergillus fumigatus* is the most common pathogenic species of the fungus.

Pathogenesis

Aspergillus spp. are transmitted as airborne conidia, and the lung is the major portal of entry. The small size of *A. fumigatus* spores, approximately 2 to 3 μm, enables them to reach alveoli. Conidia are coated in hydrophobic proteins that mask the microbial molecules from innate immune recognition. As conidia grow and form hyphae, these molecules are exposed. Alveolar macrophages recognize *Aspergillus* through TLR2 and the lectin dectin-1, which recognizes β-1,3-glucan in the fungal cell wall. Both receptors activate phagocytes to ingest and kill the conidia. In immunosuppressed states, conidia can germinate into hyphae, which then invade tissues. TLRs can recognize products of the fungal hyphae and trigger the release of pro-inflammatory mediators, including TNF, IL-1, and chemokines. Neutrophils produce reactive oxygen intermediates that kill hyphae. Invasive aspergillosis is highly associated with neutropenia and impaired neutrophil defenses.

***Aspergillus* produces several virulence factors, including adhesins, antioxidants, enzymes, and toxins.** The antioxidant defenses include melanin pigment, mannitol, catalases, and superoxide dismutases. This fungus also produces phospholipases, proteases, and toxins, but their roles in pathogenicity are not clear. *Aflatoxin* is made by *Aspergillus* spp. that grow on the surface of some crops, including corn and peanuts, particularly in warm regions if the crops are not stored or inspected appropriately. Aflatoxin causes acute and chronic hepatotoxicity and is associated with increased risk of liver cancer. Sensitization to *Aspergillus* spores produces an allergic alveolitis (Chapter 15). Allergic bronchopulmonary aspergillosis, associated with hypersensitivity arising from superficial colonization of the bronchial mucosa, often occurs in asthmatic people.

MORPHOLOGY

Colonizing aspergillosis (aspergilloma) refers to growth of the fungus in the respiratory tract with minimal or no invasion of the tissues. Colonized lung cavities are usually the result of prior tuberculosis, bronchiectasis, old infarcts, or abscesses. Proliferating masses of hyphae within proteinaceous debris form brownish "fungus balls" within the cavities. The surrounding inflammatory reaction may be sparse, or there may be chronic inflammation and fibrosis. People with aspergillomas usually have recurrent hemoptysis.

Invasive aspergillosis is an opportunistic infection that is confined to immunosuppressed hosts. The primary lesions are usually in the lung, but widespread hematogenous dissemination with involvement of the heart valves and brain is common. The pulmonary lesions take the form of necrotizing pneumonia with sharply delineated, rounded, gray foci and hemorrhagic borders; they are often referred to as target lesions (Fig. 8.44A). ***Aspergillus* forms fruiting bodies (usually in lung cavities) and septate filaments, 5 to 10 μm thick, branching at acute angles (40 degrees)** (see Fig. 8.44B). *Aspergillus* hyphae without the distinct fruiting body cannot be distinguished from *Pseudallescheria boydii* and *Fusarium* spp. by morphology alone. *Aspergillus* has a tendency to invade blood vessels, therefore areas of hemorrhage and infarction are usually superimposed on the necrotizing, inflammatory tissue reactions. Rhinocerebral *Aspergillus* infection in immunosuppressed individuals resembles that caused by Mucormycoses (e.g., *Mucor* spp., *Rhizopus* spp.).

Mucormycosis (Zygomycosis)

Mucormycotina are widely distributed in nature and cause no harm to immunocompetent individuals, but they infect immunosuppressed people, causing mucormycosis. Mucormycosis (formerly *zygomycosis*) is an opportunistic infection caused by environmental fungi, including *Mucor* spp., *Rhizopus* spp., and *Cunninghamella* spp., which belong to the subphylum *Mucormycotina*. Major predisposing factors are neutropenia, corticosteroid use, diabetes mellitus, iron overload, and breakdown of the cutaneous barrier (e.g., as a result of burns, surgical wounds, or trauma).

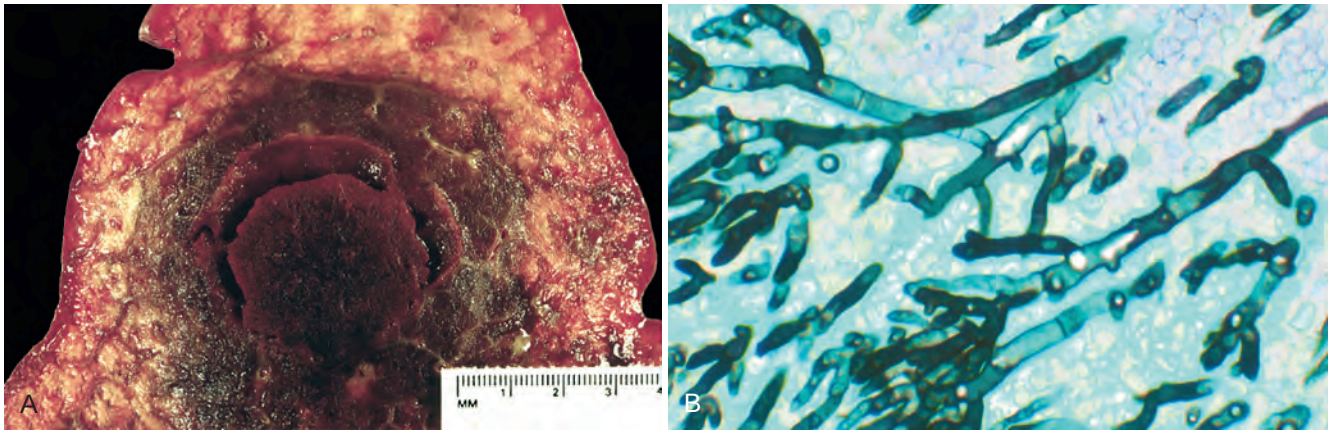


Figure 8.44 *Aspergillus* infection. (A) Invasive aspergillosis of the lung in a bone marrow transplant patient. (B) Gomori methenamine-silver stain shows septate hyphae with acute-angle branching, consistent with *Aspergillus*.

Pathogenesis

Mucormycotina are transmitted by airborne asexual spores. Inhalation of spores is the most common route of entry into the body, but percutaneous exposure or ingestion can also lead to infection. Macrophages provide the initial defense by phagocytosis and nonoxidative killing of germinating sporangiospores. *Mucormycotina* hyphal components are recognized by TLR2, which results in a pro-inflammatory cascade of cytokines including IL-6 and TNF. Neutrophils have a key role in killing hyphae after germination by directly damaging hyphae walls. If the macrophages or neutrophils are compromised in number or function, the probability of an established and then invasive infection is greatly increased. For unknown reasons, different *Mucor* species vary in resistance to phagocytosis of spores and in neutrophil damage to hyphae; thus, some infections can appear more aggressive than others despite a similar host milieu. The availability of free iron (a promoter of *Mucormycotina* growth) increases the probability of infection, as seen in people with diabetes (increased free iron due to ketoacidosis and/or glycosylation-induced poor iron affinity) and patients on chronic iron chelation treatment (where deferoxamine acts as a siderophore for the fungi).

MORPHOLOGY

Mucormycetes form nonseptate hyphae of variable width (6 to 50 μm) with frequent right-angle branching, distinct from *Aspergillus* hyphae, which are readily demonstrated by hematoxylin and eosin or special fungal stains (Fig. 8.45). The three primary sites of invasion are the nasal sinuses, lungs, and gastrointestinal tract, depending on whether the spores (which are widespread in dust and air) are inhaled or ingested. Most commonly in individuals with diabetes, the fungus may spread from nasal sinuses to the orbit and brain, giving rise to **rhinocerebral mucormycosis.** The *Mucormycotina* cause local tissue necrosis, invade arterial walls, and penetrate the periorbital tissues and cranial vault. Meningoencephalitis follows, sometimes complicated by cerebral infarctions when fungi invade arteries and induce thrombosis.

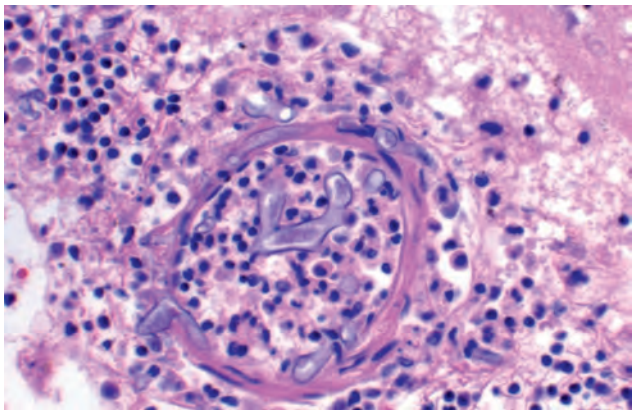


Figure 8.45 Meningeal blood vessels with angioinvasive *Mucor* spp. Note the irregular width and near right-angle branching of the hyphae. Compare with *Aspergillus*, Fig. 8.44.

Lung involvement with *Mucormycotina* may be secondary to rhinocerebral disease, or it may be primary in people with severe immunodeficiency. The lung lesions combine areas of hemorrhagic pneumonia with vascular thrombi and distal infarctions.

PARASITIC INFECTIONS

Protozoal Infections

Protozoa are unicellular eukaryotic organisms. The parasitic protozoa are transmitted by insects or by the fecal-oral route and, in humans, mainly reside in the blood or intestine (Table 8.6). Most protozoal infections are diagnosed by microscopic examination of blood smears or lesions.

Malaria

Malaria, caused by the intracellular parasite *Plasmodium,* affected an estimated 219 million people worldwide in 2017 and killed more than 435,000 people. According to the WHO, 90% of deaths from malaria occur in sub-Saharan Africa, where malaria is a leading cause of death in children younger than 5 years of age. *Plasmodium falciparum* (the cause of severe cerebral malaria) and the four other malarial parasites that infect humans (*P. vivax, P. ovale, P. knowlesi,* and *P. malariae*) are all transmitted by female *Anopheles* mosquitoes, which are widely distributed throughout Africa, Asia, and Latin America. Nearly all of the approximately 1700 new cases of malaria each year in the United States occur in travelers or immigrants. Mass spraying to eliminate the mosquito vectors was successful initially but ultimately failed when DDT was removed from the market due to environmental concerns. Worldwide public health efforts to control malaria today face the challenges of insecticide-resistant mosquitoes and drug-resistant *Plasmodium* spp. Currently, a combination of mosquito control and antimalarial drugs is viewed as the means to decrease the incidence.

Table 8.6 Selected Human Protozoal Diseases

Location	Species	Disease
Luminal or epithelial	*Entamoeba histolytica*	Amebic dysentery; liver abscess
	Balantidium coli	Colitis
	Giardia duodenalis	Diarrheal disease, malabsorption
	Cystoisospora belli	Chronic enterocolitis or malabsorption or both
	Cryptosporidium spp.	
	Trichomonas vaginalis	Urethritis, vaginitis
Central nervous system	*Naegleria fowleri*	Meningoencephalitis
	Acanthamoeba spp.	Meningoencephalitis or ophthalmitis
Bloodstream	*Plasmodium* spp.	Malaria
	Babesia spp.	Babesiosis
	Trypanosoma spp.	African sleeping sickness
Intracellular	*Trypanosoma cruzi*	Chagas disease
	Leishmania donovani	Kala-azar
	Leishmania spp.	Cutaneous and mucocutaneous leishmaniasis
	Toxoplasma gondii	Toxoplasmosis

Pathogenesis

The life cycles of the *Plasmodium* spp. are similar, although *P. falciparum* differs in ways that contribute to its greater virulence. *P. vivax*, *P. ovale*, *P. knowlesi,* and *P. malariae* cause low levels of parasitemia, mild anemia, and, in very rare instances, splenic rupture and nephrotic syndrome. ***P. falciparum* infection is associated with high levels of parasitemia that may lead to severe anemia, cerebral symptoms, renal failure, pulmonary edema, and death, depending on the susceptibility of the host.**

The life cycle of *Plasmodium* spp. involves only humans and mosquitoes, but the development of the parasite is complex, as it passes through several morphologically distinct forms. The infectious stage of *Plasmodium,* the *sporozoite,* is found in the salivary glands of female mosquitoes. When the mosquito takes a blood meal, sporozoites are released into the human's blood and, within minutes, attach to and invade liver cells by binding to the hepatocyte receptor for the serum proteins thrombospondin and properdin (Fig. 8.46). Within liver cells, malarial parasites multiply, releasing as many as 30,000 *merozoites* (asexual, haploid forms) when each infected hepatocyte ruptures. During *P. falciparum* infection, rupture usually occurs within 8 to 12 weeks. In contrast, *P. vivax* and *P. ovale* form latent *hypnozoites* in hepatocytes, which cause relapses of malaria weeks to months after initial infection. The infection of the liver and development of merozoites is called the *exoerythrocytic stage*. This stage is asymptomatic. Once released from the liver, *Plasmodium* merozoites use a lectin-like molecule to bind to sialic acid residues on glycophorin molecules on the surface of red cells and invade by active membrane penetration. Within the red cells (*erythrocytic stage*), the parasites grow in a membrane-bound digestive vacuole, hydrolyzing hemoglobin through secreted enzymes. The *trophozoite* is the first stage of the parasite in the red cell and is defined by the presence of a single chromatin mass. The next stage, the *schizont,* has multiple chromatin masses, each of which develops into a merozoite. Upon lysis of the red cell, the new merozoites infect additional red cells. Paroxysmal fever, chills, and rigors characteristic of malaria occur when the merozoites are released into the blood. As discussed later, release of merozoites induces the host cells to produce cytokines such as TNF that cause fever. The periodicity of such paroxysms (every 24-72 hours) varies with the species of the malaria parasite. Although most malaria parasites within the red cells develop into merozoites, some parasites develop, under specific conditions, into sexual forms called *gametocytes* that infect the mosquito when it takes its blood meal.

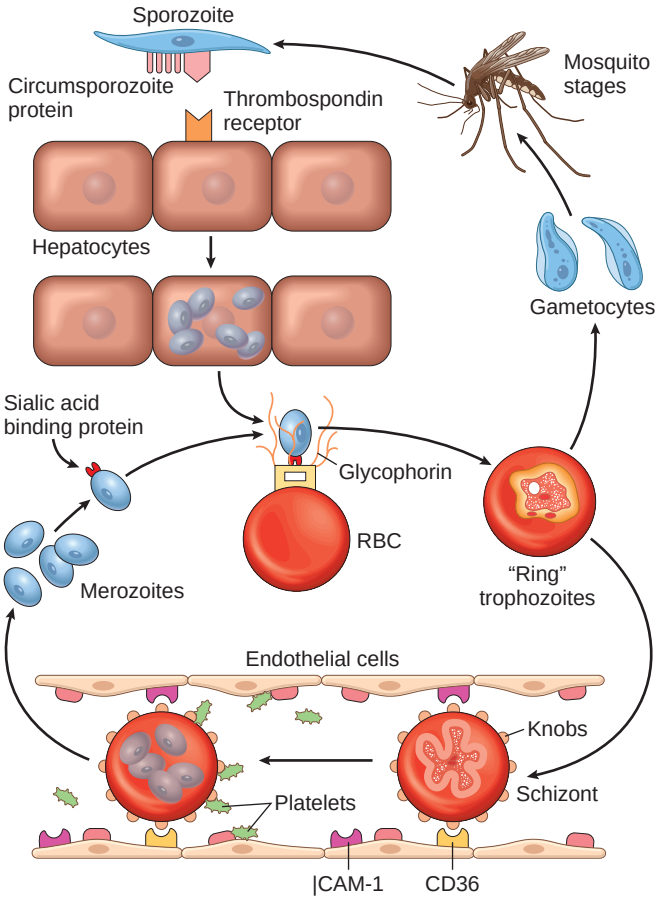


Figure 8.46 Life cycle of *Plasmodium falciparum.* Both exoerythrocytic and erythrocytic stages are depicted. *ICAM-1*, Intercellular adhesion molecule 1; *RBC*, red blood cell. (Drawn by Dr. Jeffrey Joseph, Beth Israel-Deaconess Hospital, Boston, Mass.)

Several features of *P. falciparum* account for its greater pathogenicity:

- *P. falciparum* is able to infect red blood cells of any age, whereas other species infect only young or old red cells, which are a smaller fraction of the red cell pool.
- *P. falciparum* causes infected red cells to clump together (rosette) and to stick to endothelial cells lining small blood vessels (sequestration), which blocks blood flow. Adhesive polypeptides, including *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), associate and form knobs on the surface of red cells (see Fig. 8.46). PfEMP1 binds to ligands on endothelial cells, including CD36, thrombospondin, VCAM-1, ICAM-1, and E-selectin. Red cell sequestration decreases tissue perfusion and leads to ischemia, which is responsible for the manifestations of cerebral malaria, the major cause of death in children with malaria.
- In *P. falciparum* infection, GPI-linked proteins, including merozoite surface antigens, are released from infected red cells and induce cytokine production by host cells. These cytokines increase fever, stimulate the production of reactive nitrogen species (leading to tissue damage), and induce expression of endothelial receptors for PfEMP1 (increasing sequestration).

Host resistance to *Plasmodium* can be intrinsic or acquired. Intrinsic resistance stems from inherited alterations that reduce the susceptibility of red cells to productive *Plasmodium* infections. Resistance may also be acquired following repeated or prolonged exposure to *Plasmodium* spp., which stimulates a partially protective immune response.

Several types of mutations affecting red cells are highly prevalent in parts of the world where malaria is endemic and are absent in other parts of the world. Most of these mutations are pathogenic in homozygous form, suggesting that they are maintained in populations due to a selective advantage for heterozygous carriers against malaria. The mutations fall into four broad classes.

- Point mutations in globin genes—sickle cell disease (HbS), HbC disease (hemoglobinopathies)
- Mutations leading to globin deficiencies—α- and β-thalassemia
- Mutations affecting red cell enzymes—glucose-6-phosphate dehydrogenase (G6PD) deficiency
- Mutations causing red cell membrane defects—absence of DARC (Duffy surface blood group), band 3, spectrin

P. vivax enters red cells by binding to the Duffy blood group antigen, and most of the population of West Africa is not susceptible to infection by *P. vivax* because they do not have the Duffy antigen. The mechanisms of the protective effects of the other three types of mutations are less well understood.

Individuals living where *Plasmodium* is endemic often gain partial immune-mediated resistance to malaria, evidenced by reduced illness despite infection. Antibodies and T lymphocytes specific for *Plasmodium* reduce disease manifestations, although the parasite has developed strategies to evade the host immune response. *P. falciparum* uses antigenic variation to escape from antibody responses to PfEMP1 and other surface proteins. Each haploid *P. falciparum* genome has multiple genes encoding variants of these parasite proteins. At least a percentage of the parasites switch genes each generation, producing antigenically new surface proteins. CTLs may also be important in resistance to *P. falciparum*. Multiple approaches to potential vaccines are in development; current vaccine trials have demonstrated decreases in severe disease but only modest efficacy against clinical infection.

MORPHOLOGY

The diagnostic test for malaria infection is examination of a Giemsa-stained peripheral blood smear, which permits the asexual stages of the parasite to be identified within infected red cells. PCR assays are more sensitive than the smear, but are not yet accepted as the gold standard; given the potential severity of disease, it is likely that both assays should be required to rule out infection. Insertion of parasite proteins into the red cell membrane leads to recognition by macrophages, particularly in the spleen. *Plasmodium falciparum* infection leads to splenomegaly, due to both congestion and hyperplasia of the red pulp, and the spleen may eventually exceed 1000 g in weight. In chronic infections, the spleen becomes increasingly fibrotic and brittle, with a thick capsule and fibrous trabeculae. The parenchyma is gray or black because the phagocytes contain granular, brown-black, faintly birefringent hemozoin pigment. Macrophages with engulfed parasitized red cells are also numerous.

With progression of malaria, the liver becomes enlarged and pigmented. Kupffer cells are heavily laden with malarial pigment, parasites, and cellular debris, and some pigment is also present in the parenchymal cells. Pigmented phagocytic cells may be found dispersed throughout the bone marrow, lymph nodes, subcutaneous tissues, and lungs. The kidneys are often enlarged and congested with a dusting of pigment in the glomeruli and hemoglobin casts in the tubules.

In **cerebral malaria** caused by *P. falciparum,* brain vessels are plugged with parasitized red cells (Fig. 8.47). Around the vessels, there are ring hemorrhages that are probably related to local hypoxia incident to the vascular stasis and small focal inflammatory reactions (called *malarial* or *Dürck granulomata*). With more severe hypoxia, there is degeneration of neurons, focal ischemic softening, and occasionally scant inflammatory infiltrates in the meninges.

Nonspecific focal hypoxic lesions in the heart may be induced by the progressive anemia and circulatory stasis in chronically infected people. In some, the myocardium shows focal interstitial infiltrates. Finally, in the nonimmune patient, pulmonary edema or shock with DIC may cause death, sometimes in the absence of other characteristic lesions.

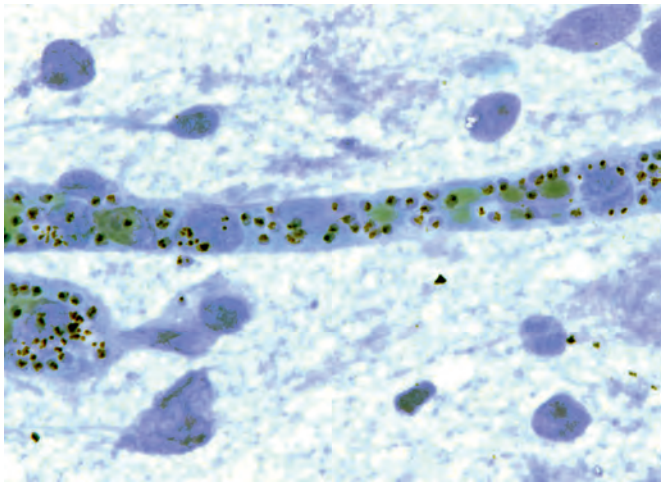


Figure 8.47 Field stain of *Plasmodium falciparum*–infected red cells marginating within a capillary in cerebral malaria.

Babesiosis

Babesia microti and *Babesia divergens,* the primary causes of babesiosis, are malaria-like protozoans transmitted in a manner similar to Lyme disease and granulocytic ehrlichiosis, by ticks, *Ixodes scapularis* (deer tick) and *Ixodes ricinus* (sheep tick), with additional cases due to *B. duncani* and *B. venatorum.* The white-footed mouse is the reservoir for *B. microti,* and in some areas, nearly all mice have a persistent low-level parasitemia. *B. microti* survives well in refrigerated blood, and a number of transfusion-acquired babesiosis cases have been reported. *Babesia* spp. parasitize red cells and cause fever and hemolytic anemia. Although most infections are asymptomatic, infection in debilitated or splenectomized individuals can cause severe, potentially fatal parasitemias.

MORPHOLOGY

In blood smears, *Babesia* spp. superficially resemble *P. falciparum* ring stages, but lack hemozoin pigment, exhibit greater pleomorphism, and form characteristic tetrads (Maltese cross), which are diagnostic, if found (Fig. 8.48). The level of *B. microti* parasitemia is a good indication of the severity of infection (about 1% in mild cases and up to 30% in splenectomized persons). In fatal cases, the anatomic findings are related to shock and hypoxia, and include jaundice, hepatic necrosis, acute renal tubular necrosis, adult respiratory distress syndrome, erythrophagocytosis, and visceral hemorrhage.

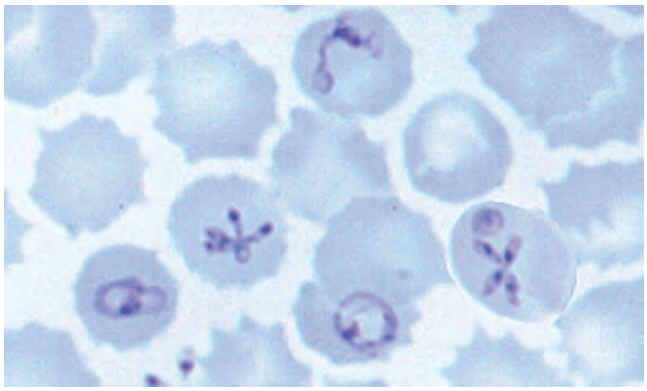


Figure 8.48 Erythrocytes with *Babesia*, including the distinctive Maltese cross form. (Courtesy Lynne Garcia, LSG and Associates, Santa Monica, Calif.)

Leishmaniasis

Leishmaniasis is a chronic inflammatory disease of the skin, mucous membranes, or viscera caused by obligate intracellular, kinetoplast-containing (kinetoplastid) protozoan parasites transmitted through the bite of infected sandflies. Leishmaniasis is endemic throughout the Middle East, South Asia, Africa, and Latin America. It may also be epidemic, as is tragically the case in Sudan, India, Bangladesh, and Brazil. There are an estimated 0.7 million to 1.2 million cases per year of cutaneous leishmaniasis, and 200,000 to 400,000 cases of visceral leishmaniasis with more than 20,000 deaths. Leishmanial infection, like infections by other intracellular organisms (mycobacteria, *Histoplasma* spp., *Toxoplasma* spp., and trypanosomes), is exacerbated by conditions that interfere with T-cell function, such as AIDS. Culture, PCR, or histologic examination is used to diagnose the infection.

Pathogenesis

The life cycle of *Leishmania* spp. involves two forms: the promastigote, which develops and lives extracellularly in the sandfly vector, and the amastigote, which multiplies intracellularly in host macrophages. There are about 21 species of *Leishmania* that infect humans and about 30 species of sandflies that serve as vectors. Mammals, including rodents, dogs, and foxes, are reservoirs of *Leishmania* spp. When sandflies bite infected humans or animals, macrophages harboring amastigotes are ingested. The amastigotes differentiate into promastigotes, multiply within the digestive tract of the sandfly, and migrate to the salivary gland, where they are poised for transmission by the fly bite. When the infected fly bites a person, the slender, flagellated infectious promastigotes are released into the host dermis along with the sandfly saliva, which potentiates parasite infectivity. The promastigotes are phagocytosed by macrophages, and the acidity within the phagolysosome induces them to transform into round amastigotes that lack flagella but contain a single mitochondrion with its DNA massed into a unique suborganelle, the kinetoplast. Amastigotes proliferate within macrophages, and dying macrophages release progeny amastigotes that can infect additional macrophages.

How far the amastigotes spread throughout the body depends on the specific *Leishmania* spp. and host, and determines the extent of disease. There are several forms of disease that are caused by different *Leishmania* spp. Old World (*L. major, L. tropica, L. donovani,* and *L. infantum*) refers to the Eastern Hemisphere (parts of Asia, the Middle East, Africa (particularly the tropical region and North Africa), and southern Europe. New World (*L. mexicana, L. braziliensis,* and *L. chagasi*) refers to the Western Hemisphere (parts of Mexico, Central America, and South America). A few cases have been acquired in Texas and Oklahoma.

***Leishmania* spp. manipulate innate host defenses to facilitate their entry and survival in phagocytes.** Promastigotes produce two abundant surface glycoconjugates that contribute to their virulence.

- Lipophosphoglycan forms a dense glycocalyx that both activates complement (leading to C3b deposition on the parasite surface) and inhibits complement action (by preventing membrane attack complex insertion into the parasite membrane). Thus, the parasite becomes coated with C3b but avoids destruction by the membrane attack complex. Instead, the C3b on the surface of the parasite binds to Mac-1 and CR1 on macrophages, targeting the promastigote for phagocytosis.
- Gp63 is a zinc-dependent proteinase that cleaves complement and some lysosomal antimicrobial enzymes. Gp63 also binds to fibronectin receptors on macrophages and promotes promastigote adhesion to macrophages.

To escape killing by neutrophils, *Leishmania* spp. use the following mechanisms: (1) interfere with the formation of phagolysosomes and fusion with granules, (2) localize to nonlytic compartments, (3) resist toxicity by reactive oxygen species, and (4) resist neutrophil extracellular trap (NET) formation by producing endonucleases that digest the NETs and expressing protease-resistant molecules.

Leishmania spp. amastigotes also produce molecules that facilitate their survival and replication within macrophages. Amastigotes reproduce in macrophage phagolysosomes, which normally have a pH of 4.5. However, the amastigotes protect themselves from this hostile environment by expressing a proton-transporting ATPase, which maintains the phagolysosome pH at 6.5.

Leishmania spp. require iron for survival, and they have an iron transporter from the ZRT-, IRT-like (ZIP) family of membrane proteins that is expressed on amastigotes and stimulates iron entry into the normally low iron environment of macrophage phagolysosomes. The host combats this iron acquisition by using pro-inflammatory cytokines to repress iron absorption (in part by increasing production of hepcidin, the principal iron exporter, and by activating synthesis of ferritin, which binds free iron. In addition, macrophages downregulate the transferrin receptor and remove iron from the phagosome.

The primary mechanisms of resistance and susceptibility to *Leishmania* spp. are determined by Th1 and Th2 responses. Parasite-specific CD4+ Th1 cells are needed to control *Leishmania* spp. in mice and humans. *Leishmania* spp. evade host immunity by impairing the development of the Th1 response. In animal models, mice that are resistant to *Leishmania* infection produce high levels of Th1-derived IFN-γ, which activates macrophages to kill the parasites. By contrast, mouse strains that are susceptible to leishmaniasis mount a dominant Th2 response. Th2 cytokines such as IL-4, IL-13,

and IL-10 prevent effective killing of *Leishmania* spp. by inhibiting the microbicidal activity of macrophages.

MORPHOLOGY

Leishmania spp. produce four different types of lesions in humans: visceral, cutaneous, mucocutaneous, and diffuse cutaneous. In **visceral leishmaniasis,** parasites invade and activate macrophages throughout the mononuclear phagocyte system (Fig. 8.49) and cause a systemic inflammatory disease marked by hepatosplenomegaly, lymphadenopathy, pancytopenia, fever, and weight loss. The spleen may weigh as much as 3 kg. Phagocytic cells are enlarged and filled with *Leishmania* spp., many plasma cells are present, and the normal architecture of the spleen is obscured. In the late stages, the liver becomes increasingly fibrotic. Phagocytic cells crowd the bone marrow and also may be found in the lungs, gastrointestinal tract, kidneys, pancreas, and testes. Often there is hyperpigmentation of the skin in individuals of South Asian ancestry, which is why the disease is called *kala-azar* (*black fever* in Hindi). In the kidneys there may be an immune complex–mediated mesangioproliferative glomerulonephritis, and in advanced cases there may be amyloid deposition. People with advanced leishmaniasis can develop life-threatening secondary bacterial infections, such as pneumonia, sepsis, or tuberculosis. Hemorrhages related to thrombocytopenia may also be fatal.

Cutaneous leishmaniasis is a relatively mild, localized disease consisting of ulcers on exposed skin. The lesion begins as a papule surrounded by induration, changes into a shallow and slowly expanding ulcer, often with heaped-up borders, and usually heals by involution within 6 to 18 months without treatment. On microscopic examination, the lesion shows granulomatous inflammation, usually with many giant cells and few parasites.

Mucocutaneous leishmaniasis is found only in the New World. Moist, ulcerating, or nonulcerating lesions develop in the nasopharyngeal areas and, with progression, may be highly destructive and disfiguring. Microscopic examination reveals a mixed inflammatory infiltrate composed of parasite-containing macrophages with lymphocytes and plasma cells. Later, the tissue inflammatory response becomes granulomatous, and the number of parasites declines. Eventually, the lesions remit and scar, although reactivation may occur after long intervals by mechanisms that are not currently understood.

Diffuse cutaneous leishmaniasis is a rare form of dermal infection found in Ethiopia and adjacent East Africa and in Central and South America. Diffuse cutaneous leishmaniasis begins as a single skin nodule, which continues spreading until the entire body is covered by nodular lesions. Microscopically, they contain aggregates of foamy macrophages stuffed with *Leishmania* organisms.

African Trypanosomiasis

African trypanosomes are kinetoplastid parasites (containing a large mass of DNA called a *kinetoplast*) that proliferate as extracellular forms in the blood and cause sustained or intermittent fevers, lymphadenopathy, splenomegaly, progressive brain dysfunction (sleeping sickness), cachexia, and death. *Trypanosoma brucei rhodesiense* infection, which occurs in East Africa and is often acute and virulent, is a zoonotic infection that is best combated by reducing infected fly (vector) populations. A few hundred cases are reported per year. *Trypanosoma brucei gambiense* infection occurring in West Africa spreads from human to human via fly bites and requires active case detection and treatment, with about 3000 cases reported annually. Tsetse flies (genus *Glossina*) transmit African *Trypanosoma* to humans either from wild and domestic animals *(T. brucei rhodesiense)* or from other humans *(T. brucei gambiense)*. Within the fly, the parasites multiply in the stomach and then in the salivary glands before developing into nondividing trypomastigotes, which are transmitted to humans and animals with the next blood meal. Diagnosis is made by microscopic examination of blood smears, lymph node, or chancre. Cerebrospinal fluid must be examined to determine whether there is infection of the CNS, as this determines the course of treatment.

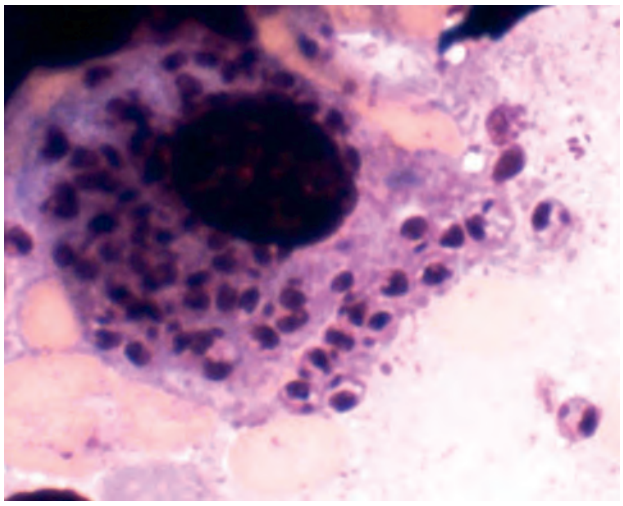


Figure 8.49 Giemsa stain of a tissue macrophage with *Leishmania donovani* parasites.

Pathogenesis

African trypanosomes are covered by a single, abundant, glycolipid-anchored protein called the *variant surface glycoprotein* (VSG). As parasites proliferate in the bloodstream, the host produces antibodies to the VSG, which, in association with phagocytes, kill most of the organisms, causing a spike of fever. A small number of parasites, however, undergo a genetic rearrangement and produce a different VSG on their surface and so escape the host immune response. These successor trypanosomes multiply until the host mounts a new anti-VSG response and kills most of them, but then another clone with a distinct VSG takes over. In this way, African trypanosomes cause waves of fever before they finally invade the CNS. Trypanosomes have many VSG genes, only one of which is expressed at a time. The parasite uses an elegant mechanism to turn VSG genes on and off. Although VSG genes are scattered throughout the trypanosome genome, only VSG genes found within bloodstream expression sites near the ends of chromosomes are transcribed. New VSG genes are periodically moved into these sites, mainly by homologous recombination, generating a new VSG. A specialized RNA polymerase that transcribes VSG genes associates with only a single bloodstream expression site, limiting expression to one VSG at a time.

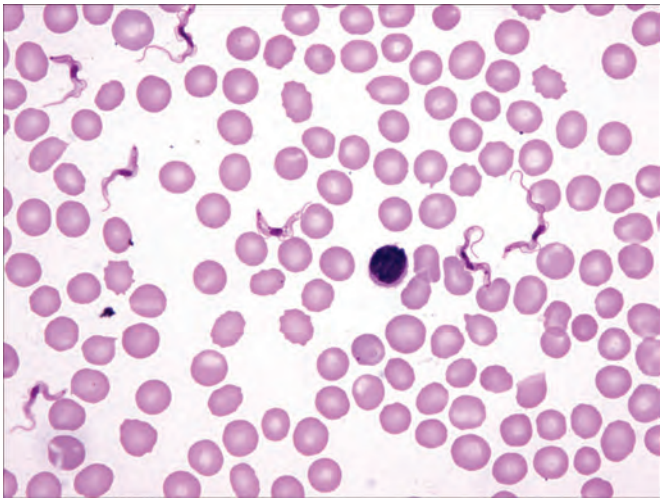


Figure 8.50 Slender bloodstream parasites of African trypanosomiasis.

MORPHOLOGY

A large, red, rubbery chancre forms at the site of the insect bite and contains numerous parasites surrounded by a dense, predominantly mononuclear, inflammatory infiltrate. With chronicity, the lymph nodes and spleen enlarge due to infiltration by lymphocytes, plasma cells, and macrophages, which are filled with dead parasites. Trypanosomes (Fig. 8.50) concentrate in capillary loops, such as the choroid plexus and glomeruli. When parasites breach the blood-brain barrier and invade the CNS, a leptomeningitis develops that extends into the perivascular Virchow-Robin spaces, and eventually a demyelinating panencephalitis occurs. Plasma cells containing cytoplasmic globules filled with immunoglobulins are frequent and are referred to as **Mott cells.** Chronic disease leads to progressive cachexia; patients, devoid of energy and normal mentation, literally waste away.

Chagas Disease

***Trypanosoma cruzi* is a kinetoplastid, intracellular protozoan parasite that causes American trypanosomiasis (Chagas disease).** Chagas disease occurs rarely in the United States and Mexico but is more common in South America, particularly Brazil, with an estimated 8 million people infected globally. *T. cruzi* parasites infect many animals, including cats, dogs, and rodents. The parasites are transmitted between animals and to humans by triatomine bugs (also known as kissing bugs or reduviids), which hide in the cracks of loosely constructed houses, feed on the sleeping inhabitants, and pass the parasites in feces; the infectious parasites enter the host through damaged skin or through mucous membranes. At the site of skin entry, there may be a transient, erythematous nodule. Another important route of infection is oral ingestion of the parasites due to contamination of food products with triatomine bugs and/or their feces. Other modes of infection include receipt of infected blood products, organ transplantation, or congenital transmission. Diagnosis can be made by a blood smear in the acute case, but is made more commonly by serology.

Pathogenesis

T. cruzi has a superfamily of glycosylphosphatidylinositol-anchored surface glycoproteins that interact with multiple ligands, such as laminin, fibronectin, collagen, cytokeratin, and other extracellular proteins, and are involved in cell adhesion and invasion. Although most intracellular pathogens avoid the toxic contents of lysosomes, after ingestion into macrophages, *T. cruzi* actually requires brief exposure to the acidic phagolysosome for development of amastigotes, the intracellular stage of the parasite. To ensure exposure to lysosomes, *T. cruzi* trypomastigotes elevate the concentration of cytoplasmic calcium in host cells, which promotes fusion of the phagosome and lysosome. In addition to enhancing amastigote development, the low pH of the lysosome activates pore-forming proteins that disrupt the lysosomal membrane, releasing the parasite into the cell cytoplasm. Parasites reproduce as rounded amastigotes in the cytoplasm of host cells and then develop flagella, lyse host cells, enter the bloodstream, and penetrate smooth, skeletal, and heart muscles. *T. cruzii* evades the complement system activation by expressing complement regulatory proteins.

Chagas disease primarily affects the heart, and in endemic areas it is a major cause of sudden death due to cardiac arrhythmia. In acute Chagas disease, which is mild in most individuals, cardiac damage results from direct invasion of myocardial cells by the organisms and the subsequent inflammation. Rarely, acute Chagas disease presents with high parasitemia, fever, or progressive cardiac dilation and failure, often with generalized lymphadenopathy or splenomegaly. In chronic Chagas disease, which occurs in 20% of people 5 to 15 years after initial infection, the mechanism of cardiac damage has two components.

- The presence of persistent *T. cruzi* parasites leads to a continued immune response with inflammatory infiltration of the myocardium, even though only scant organisms may be present.
- The parasite also may induce autoimmune responses, such that antibodies and T cells that recognize parasite proteins cross-react with host myocardial cells, nerve cells, and extracellular proteins such as laminin. For example, cross-reactive antibodies may induce electrophysiologic dysfunction of the heart.

Damage to myocardial cells and to conductance pathways causes a dilated cardiomyopathy and cardiac arrhythmias. In addition, damage to the myenteric plexus causes dilation of the colon (megacolon) and esophagus. This is particularly common in Brazilian endemic areas, where as many as 50% of the patients with lethal carditis have colonic and esophageal disease.

MORPHOLOGY

In lethal **acute myocarditis,** the changes are diffusely distributed throughout the heart. Clusters of amastigotes cause swelling of individual myocardial fibers and create intracellular pseudocysts. There is focal myocardial cell necrosis accompanied by extensive, dense, acute interstitial inflammatory infiltration throughout the

myocardium, often associated with four-chamber cardiac dilation (Chapter 12).

In **chronic Chagas disease,** the heart is typically dilated, rounded, and increased in size and weight. Often, there are mural thrombi that, in about one-half of autopsy cases, have given rise to pulmonary or systemic emboli or infarctions. On histologic examination, there are interstitial and perivascular inflammatory infiltrates composed of lymphocytes, plasma cells, and monocytes. There are scattered foci of myocardial cell necrosis and interstitial fibrosis, especially toward the apex of the left ventricle, which may undergo aneurysmal dilation and thinning. Even with dilation of the esophagus and colon, parasites cannot be found within ganglia of the myenteric plexus. Chronic Chagas cardiomyopathy often has to be treated by cardiac transplantation.

Toxoplasmosis

Although millions of people carry the ubiquitous parasitic protozoa *T. gondii* without symptoms due to control by their immune systems, *T. gondii* causes significant disease in the immunocompromised and in pregnant women and their offspring. It is estimated that 11% of the US population 6 years of age and older has been infected. In some regions of the world, 95% of the population has been infected. *T. gondii* infection occurs by eating undercooked meat (pork, lamb, venison) contaminated with the tissue cysts or by acquiring oocysts through drinking contaminated water, or by accidental ingestion following cleaning of a cat's litter box or handling contaminated soil. Another route of infection is vertical transmission during pregnancy. Rarely, infection occurs via blood transfusion or organ transplantation. Some infected individuals (10% to 20%) may experience swollen lymph nodes and muscle aches, with a benign, self-limited course of weeks to months, followed by a latent infection. The organism can be reactivated following immunosuppression, such as after organ transplantation or in HIV infection. Congenital toxoplasmosis symptoms may not become evident for months or many years after birth, and prompt treatment at birth may reduce the ultimate sequelae. Although congenital chorioretinitis is bilateral, ocular toxoplasmosis may be acquired and then is often unilateral. *Toxoplasma* encephalitis, due to reactivation, is the most common opportunistic pathogen of the CNS in AIDS patients. Myocarditis and pneumonitis are two other common manifestations in AIDS patients. Investigators have hypothesized a connection between toxoplasmosis and schizophrenia and other neurologic conditions, but studies are inconclusive.

Pathogenesis

The cat is the only definitive host, and humans are a dead-end host. Unsporulated oocysts are shed in cat feces, which sporulate in the environment. Intermediate hosts (birds, rodents) become infected from ingestion of contaminated soil, water, or plants. Ingested oocysts release sporozoites that invade the human intestinal epithelium, disseminating throughout the body. The sporozoites transform into tachyzoites that localize to tissues and develop into bradyzoites. Cats become infected after ingestion of intermediate hosts. *T. gondii* is an intracellular pathogen that can invade any nucleated cell by a unique form of actin-myosin moving junction and forms a parasitophorous vacuole (PV) protected from lysosomal fusion. To avoid clearance of the tachyzoites, *T. gondii* also inhibits autophagy of the parasite-containing vacuole.

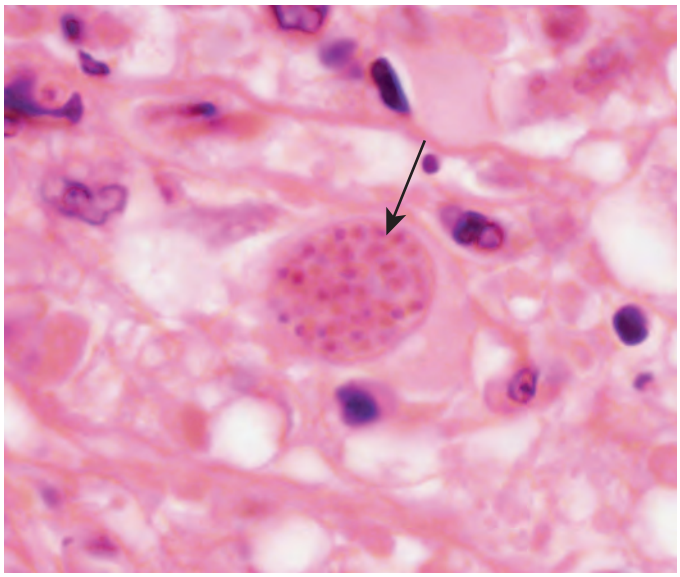


Figure 8.51 A brain biopsy showing a cyst of *Toxoplasma gondii (arrow).* (From Procop GW, Pritt BS: *Pathology of Infectious Diseases,* Philadelphia, Elsevier, 2015; Courtesy Dr. Bobbi S. Pritt, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minn.)

MORPHOLOGY

In humans, the parasites form cysts in skeletal muscle, myocardium, visceral organs, brain, and eyes. Although diagnosis is usually through serologic testing, cysts may be observed in biopsies (Fig. 8.51), or **tachyzoites** may be observed in stained fluid specimens, such as a bronchoalveolar lavage fluid from an immunocompromised host. Tachyzoites may be observed with periodic acid-Schiff, Giemsa, or hematoxylin and eosin stains. Tissue cysts can vary in size from 5 to 100 μm. Intact cysts can remain for the life of the host without causing inflammation. The cyst wall is thin and contains crescent-shaped bradyzoites of approximately 1.5 by 7 μm. PCR of blood or cerebrospinal fluid is a sensitive assay and permits less invasive testing in many cases. **Congenital toxoplasmosis** can result in fetal death and abortion, or hydrocephalus, microcephaly, cerebral calcifications, neurocognitive deficits, and chorioretinitis.

Metazoal Infections

Metazoa are multicellular, eukaryotic organisms with organ systems. Parasitic metazoal infections are contracted by consuming the parasite, often in undercooked meat, or by direct invasion of the host through the skin or via insect bites. Metazoa dwell in many sites of the body, including the intestine, skin, lung, liver, muscle, blood vessels, and lymphatics. The infections are diagnosed by microscopic identification of larvae or ova in excretions or tissues, and by serology.

Strongyloidiasis

Strongyloides stercoralis infects tens of millions of people worldwide and is endemic in tropical and subtropical regions of South America, sub-Saharan Africa, and Southeast Asia. It occurs sporadically in the United States, including in Appalachia. **The worms live in the soil and infect humans when larvae penetrate the skin, travel in the circulation to the lungs, and then travel up the trachea to be swallowed.** Female worms reside in the mucosa of the small intestine, where they produce eggs by asexual reproduction (parthenogenesis). Most of the larvae are passed in the stool and then may contaminate soil to continue the cycle of infection, however some develop and become infectious within the host, leading to autoinfection.

In immunocompetent hosts, *S. stercoralis* is often asymptomatic, but may cause diarrhea, bloating, and occasionally malabsorption. Immunocompromised hosts, particularly people on prolonged corticosteroid therapy, can have very high worm burdens (hyperinfection) due to uncontrolled autoinfection, leading to fatal disease. Corticosteroids inhibit the functions of eosinophils and other host immune cells that accumulate in tissues in response to infection, and stimulate female *Strongyloides* spp. to increase infective larvae production. In addition, other disease states that perturb immune control mechanisms (e.g., organ transplantation, lymphoma, HTLV-1) are associated with increased risk. Hyperinfection can be complicated by sepsis caused by intestinal bacteria, which enter the blood following damage to the intestinal wall by the invading larvae.

MORPHOLOGY

In mild strongyloidiasis, worms, mainly larvae, are present in the duodenal crypts but are not seen in the underlying tissue. There is an eosinophil-rich infiltrate in the lamina propria with mucosal edema. Hyperinfection with *S. stercoralis* results in invasion of larvae into the colonic submucosa, lymphatics, and blood vessels, with an associated mononuclear infiltrate. There are many adult worms, larvae, and eggs in the crypts of the duodenum and ileum (Fig. 8.52). Worms of all stages may be found in other organs, including skin and lungs, and may even be found in large numbers in sputum.

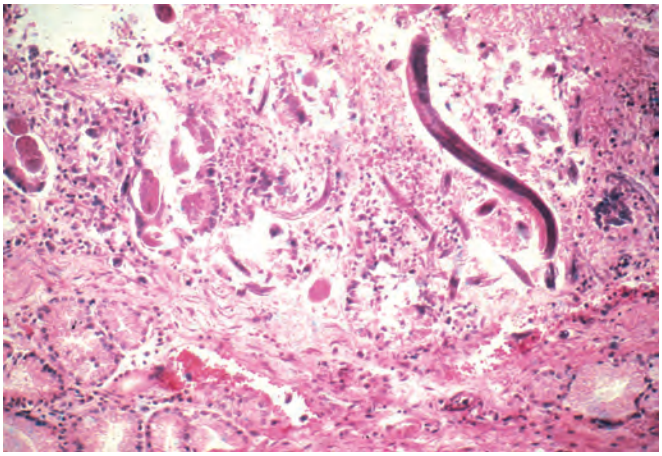


Figure 8.52 *Strongyloides* hyperinfection in a patient treated with high-dose cortisone. A female, her eggs, and rhabditoid larvae are in the duodenal crypts; filariform larvae are entering the blood vessels and muscularis mucosa. (Courtesy Dr. Franz C. Von Lichtenberg, Brigham and Women's Hospital, Boston, Mass.)

Cysticercosis and Hydatid Disease (Tapeworm Infections)

Taenia solium and *Echinococcus granulosus* are cestode parasites (tapeworms) that cause cysticercosis and hydatid infections, respectively. Both diseases are caused by larvae that develop after ingestion of tapeworm eggs. These tapeworms have a complex life cycle requiring two mammalian hosts: a definitive host, in which the worm reaches sexual maturity, and an intermediate host, in which the worm does not reach sexual maturity.

***T. solium* causes disease because of deposition of cysts (cysticerci) in organs and the resulting inflammatory response.** These tapeworms consist of a head (scolex) that has suckers and hooklets that attach to the intestinal wall, a neck, and many flat segments called proglottids that contain both male and female reproductive organs. New proglottids develop behind the scolex. The most distal proglottids are mature and contain many eggs, and they can detach and be shed in feces. *T. solium* can be transmitted to humans in two ways, ingestion of larval cysts or eggs, with distinct outcomes.

- Larval cysts, called *cysticerci,* are ingested in undercooked pork and attach to the intestinal wall, where they develop into mature adult tapeworms. These can grow to many meters in length and produce mild abdominal symptoms. The parasite life cycle is completed with this mode of infection, and disseminated cysticercosis does not develop.
- When intermediate hosts (pigs or humans) ingest eggs in food or water contaminated with human feces, the egg becomes an oncosphere that hatches and penetrates the gut wall, disseminating hematogenously, followed by transition to a cysticercus that can encyst in many organs, giving rise to clinical symptoms of cysticercosis. The most serious manifestations result from encystment in the brain (neurocysticercosis). Convulsions, increased intracranial pressure, and other neurologic disturbances may occur. Adult tapeworms are not produced with this mode of infection because larval cysts lodged in various tissues other than the intestine cannot develop into mature worms. Viable *T. solium* cysts often do not produce symptoms and can evade host immune defenses by producing taeniaestatin and paramyosin, which seem to inhibit complement activation. When the cysticerci die and degenerate, an inflammatory response develops.

Taenia saginata, the beef tapeworm, and *Diphyllobothrium latum,* the fish tapeworm, are acquired by eating undercooked meat or fish. In humans, these parasites live only in the gut and do not form cysticerci.

Hydatid disease is caused by ingestion of eggs of *Echinococcus* spp and formation of cysts in organs where the parasite larvae are deposited. For *E. granulosus* the definitive host is the dog, and the usual intermediate hosts are sheep. For *E. multilocularis* the fox is the most important definitive host, and rodents are intermediate hosts. Humans are accidental intermediate hosts, infected by ingestion of food contaminated with eggs shed by dogs or foxes. Eggs hatch in the duodenum and larvae invade the liver, lungs, or bones. Infection in humans is usually asymptomatic, but

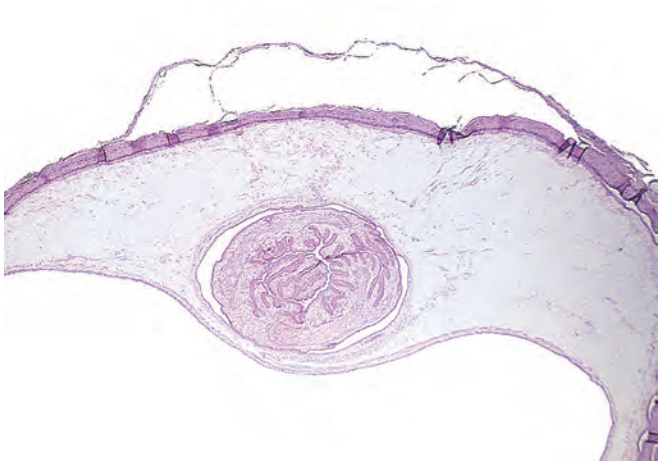


Figure 8.53 Portion of a cysticercus cyst in the skin.

large cysts in the liver can cause abdominal pain or obstruction, and pulmonary cysts can cause pain, cough, and hemoptysis. Caution is warranted if surgical removal of the cyst is considered, as anaphylaxis and/or dissemination of organisms can result from spillage of the cyst contents.

MORPHOLOGY

Cysticerci may be found in any organ, but the more common locations include the brain, muscles, skin, and heart. Cerebral symptoms depend on the precise location of the cysts, which may be intraparenchymal, attached to the arachnoid, or freely floating in the ventricular system. The cysts are ovoid and white to opalescent, often grape-sized, and contain an invaginated scolex with hooklets that are bathed in clear cyst fluid (Fig. 8.53). The cyst wall is more than 100 μm thick, is rich in glycoproteins, and evokes little host inflammatory response when it is intact. When cysts degenerate, however, there is inflammation, followed by focal scarring, and calcifications, which may be visible by radiography.

***E. granulosus* cysts** are found in the liver in about two thirds of cases, with 5% to 15% in the lung, and the rest in bones and brain, or other organs. In the various organs, the larvae lodge within the capillaries and first incite an inflammatory reaction composed principally of mononuclear leukocytes and eosinophils. Many such larvae are destroyed, but others encyst. The cysts begin at microscopic levels and progressively increase in size, so that in 5 years or more they may have achieved dimensions of more than 10 cm in diameter. Enclosing an opalescent fluid is an inner, nucleated, germinative layer and an outer, opaque, non-nucleated layer. The outer non-nucleated layer is distinctive and has innumerable delicate laminations. Outside this opaque layer, there is a host inflammatory reaction that produces a zone of fibroblasts, giant cells, and mononuclear and eosinophilic cells. In time a dense fibrous capsule forms. Daughter cysts often develop within the large mother cyst. These appear first as minute projections of the germinative layer that develop central vesicles and thus form tiny brood capsules. Degenerating scolices of the worm produce a fine, sandlike sediment within the hydatid fluid (hydatid sand).

Trichinosis

***Trichinella* spp. are nematode parasites that are acquired by ingestion of larvae in undercooked meat from infected animals (usually pigs, boars, or horses) that have themselves been infected by eating rats or meat products containing *T. spiralis*, *T. nativa*, or *T. britovi*.** There are an estimated 10,000 cases in the world each year. In the United States the number of *T. spiralis*–infected pigs has been greatly reduced by laws requiring proper cooking of food fed to hogs; the number of reported human infections in the United States is usually about 20 cases each year. Still, trichinosis remains widespread in other parts of the world, where undercooked meat, including noncommercial livestock and game (e.g., bear), is commonly eaten.

Pathogenesis

The life cycle of *T. spiralis* begins in the human intestine but ends within muscle as humans are dead-end hosts. In the human gut, *T. spiralis* larvae develop into adults that mate and release new larvae, which penetrate into the tissues. Larvae disseminate hematogenously and penetrate muscle cells, causing fever, myalgias, marked eosinophilia, and periorbital edema. Less commonly, the larvae lodge in the heart, lungs, and brain, and patients can develop dyspnea, encephalitis, and cardiac failure. In striated skeletal muscle, *T. spiralis* larvae become intracellular parasites, increase dramatically in size, and modify the host muscle cell so that it loses its striations, gains a collagenous capsule, and develops a plexus of new blood vessels around itself. The cell-parasite complex is largely asymptomatic, and the worm may persist for years before it dies and calcifies. Antibodies to larval antigens, which include an immunodominant carbohydrate epitope called **tyvelose,** may reduce reinfection and are useful for serodiagnosis of the disease.

T. spiralis and other invasive nematodes stimulate a Th2 response, with production of IL-4, IL-5, IL-10, and IL-13. The cytokines produced by Th2 cells activate eosinophils and mast cells, both of which are associated with the inflammatory response to these parasites. In animal models of *T. spiralis* infection, the Th2 response is associated with increased contractility of the intestine, which expels adult worms from the gut and subsequently reduces the number of larvae in the muscles. Although the Th2 response indirectly reduces the number of larvae in muscle by eliminating adults from the intestine, it is not clear whether the intramuscular inflammatory response, which is composed of mononuclear cells and eosinophils, is effective against the larvae.

MORPHOLOGY

During the invasive phase of trichinosis, cell destruction can be widespread with heavy infections and may be lethal. In the heart, there is a patchy interstitial myocarditis characterized by many eosinophils and scattered giant cells. The myocarditis can lead to scarring. Larvae in the heart do not encyst and are difficult to identify, because they die and disappear. In the lungs, trapped larvae cause focal edema and hemorrhages, sometimes with an allergic eosinophilic infiltrate. In the CNS, larvae cause a diffuse lymphocytic and eosinophilic infiltrate, with focal gliosis in and about small capillaries of the brain.

T. spiralis preferentially encysts in striated skeletal muscles with the richest blood supply, including the diaphragm and the extraocular, laryngeal, deltoid, gastrocnemius, and intercostal muscles (Fig. 8.54). Coiled larvae are approximately 1 mm long and are

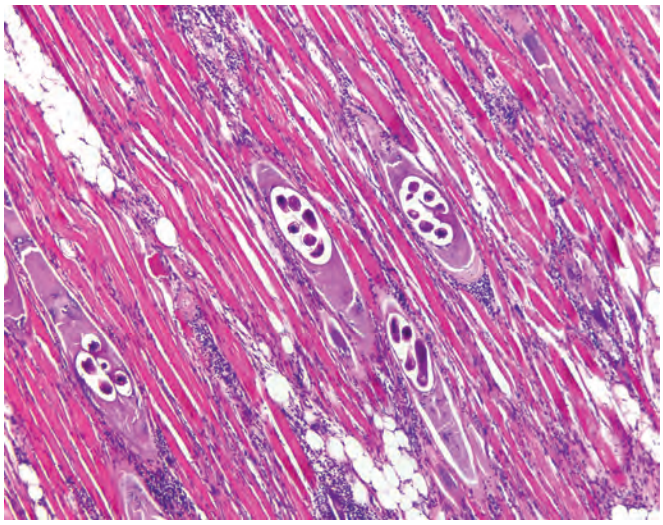


Figure 8.54 Multiple coiled *Trichinella spiralis* larvae within skeletal muscle cells.

surrounded by intracellular membrane-bound vacuoles, which in turn are surrounded by new blood vessels and an eosinophil-rich mononuclear cell infiltrate. This infiltrate is greatest around dying parasites, which eventually calcify and leave behind scars that are sufficiently characteristic to suggest the diagnosis of trichinosis.

Schistosomiasis

Schistosomiasis is estimated to infect approximately 230 million persons and kill more than 200,000 individuals annually. The affected organs and hence the site of major disease vary with the species. *Shistosoma mansoni* and *S. japonicum* affect the liver and the gut predominantly. Most deaths are due to hepatic cirrhosis, which is caused by *S. mansoni* in Latin America, Africa, and the Middle East and by *S. japonicum* and *S. mekongi* in East Asia. By contrast, *S. haematobium,* found in Africa, causes chronic granulomatous bladder inflammation that may lead to hematuria, obstructive uropathy, and carcinoma. *Schistosoma* flukes require passage through freshwater snails that live in the slow-moving water of tropical rivers, lakes, and irrigation ditches, ironically linking agricultural development with spread of the disease. Acute schistosomiasis in humans can be a severe febrile illness that peaks about 2 months after infection. Severe hepatic fibrosis is a serious manifestation of chronic schistosomiasis.

Pathogenesis

Much of the pathology of schistosomiasis is caused by host inflammatory reactions to different stages of the parasite. The life cycle of *Schistosoma* spp. involves stepwise infection of several human tissues, each associated with host inflammatory responses. After release from snails, ciliated miracidium larvae mature into infectious larvae (cercariae) that swim through fresh water and penetrate human skin with the aid of powerful proteolytic enzymes that degrade the keratinized layer. There is minimal skin reaction. Schistosomes migrate through the skin into the peripheral vasculature and lymphatics, travel to the lungs and heart, from where they are disseminated widely, including the mesenteric, splanchnic, and portal circulation, ultimately reaching the hepatic vessels, where they mature (*S. mansoni* and *S. japonicum*). Mature male-female worm pairs then migrate once again and settle in the venous system (commonly the portal or pelvic veins). Females produce hundreds to thousands of eggs per day that secrete proteases and elicit localized inflammatory reactions. This inflammatory response to egg migration is necessary for passive transfer across the intestine and, in the case of *S. haematobium,* bladder walls, allowing the eggs to be shed in stool or urine, respectively. Infection of freshwater snails completes the life cycle.

Eggs that are carried by the portal circulation into the hepatic parenchyma can cause severe chronic inflammation in the liver. This immune response to *S. mansoni* and *S. japonicum* eggs is responsible for the most serious complication of schistosomiasis, liver fibrosis. The helper T-cell response in the early stage is dominated by Th1 cells that produce IFN-γ, which stimulates macrophages to secrete high levels of the cytokines TNF-α, IL-1, and IL-6 that cause fever. Chronic schistosomiasis is associated with a dominant Th2 response, associated with the presence of activated macrophages. Both types of helper T cells contribute to the formation of granulomas surrounding eggs in the liver. Hepatic fibrosis is a serious manifestation of chronic schistosomiasis in which Th2 cells and activated macrophages may play the major role.

MORPHOLOGY

In early *S. mansoni* or *S. japonicum* infections, white, pinhead-sized granulomas are scattered throughout the gut and liver. At the center of the granuloma is the schistosome egg, which contains a miracidium; this degenerates over time and calcifies. The granulomas are composed of macrophages, lymphocytes, neutrophils, and eosinophils, which are distinctive for helminth infections (Fig. 8.55). The liver is darkened by regurgitated heme-derived

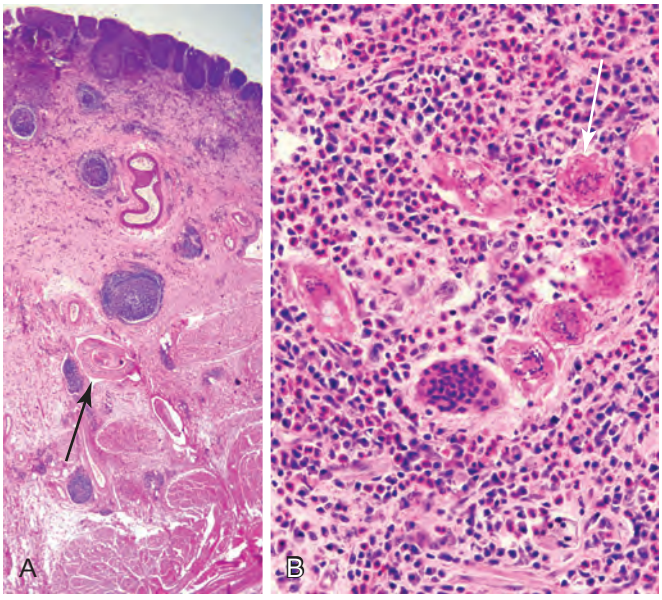


Figure 8.55 *Schistosoma haematobium* infection of the bladder (A) showing dense fibrosis, scattered granulomas, and a cross-section of adult worms in a vessel *(arrow)*. High magnification (B) demonstrates miracidium-containing eggs *(arrow)*, prominent eosinophils, histiocytes, and giant cells.

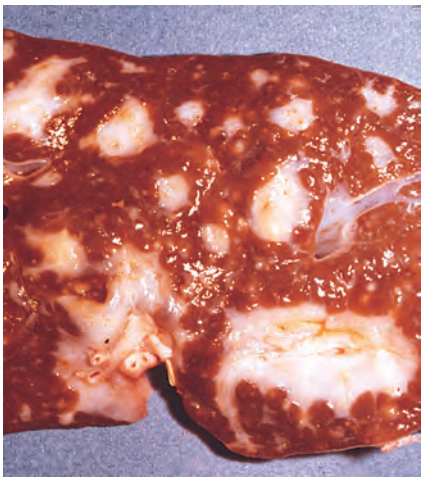


Figure 8.56 Pipe-stem fibrosis of the liver due to chronic *Schistosoma japonicum* infection.

pigments from the schistosome gut, which, like malaria pigments, are iron-free and accumulate in Kupffer cells and splenic macrophages.

In late *S. mansoni* or *S. japonicum* infections, inflammatory patches or pseudopolyps may form in the colon. The surface of the liver is bumpy, and cut surfaces reveal granulomas and widespread fibrosis and portal enlargement without intervening regenerative nodules. Because these fibrous tracts resemble the stem of a clay pipe, the lesion is named **pipe-stem fibrosis** (Fig. 8.56). The fibrosis often obliterates the portal veins, leading to portal hypertension, severe congestive splenomegaly, esophageal varices, and ascites. Schistosome eggs, diverted to the lung through portal collaterals, may produce granulomatous pulmonary arteritis with intimal hyperplasia, progressive arterial obstruction, and ultimately heart failure (cor pulmonale). On histologic examination, arteries in the lungs show disruption of the elastic layer by granulomas and scars, luminal organizing thrombi, and angiomatoid lesions similar to those of idiopathic pulmonary hypertension (Chapter 15). Patients with hepatosplenic schistosomiasis also have an increased frequency of mesangioproliferative or membranous glomerulopathy (Chapter 20) in which glomeruli contain deposits of immunoglobulin and complement but rarely schistosome antigen.

In *S. haematobium* infection, inflammatory cystitis due to massive egg deposition and granulomas appears early, leading to mucosal erosions and hematuria. Later, the granulomas calcify and develop a sandy appearance, which, if severe, may line the wall of the bladder and cause a dense concentric rim (calcified bladder) on radiographic films. The most frequent complication of *S. haematobium* infection is inflammation and fibrosis of the ureteral walls, leading to obstruction, hydronephrosis, and chronic pyelonephritis. There is also an association between urinary schistosomiasis and squamous cell carcinoma of the bladder (Chapter 21).

Lymphatic Filariasis

Lymphatic filariasis is transmitted by mosquitoes and is caused by closely related nematodes, *Wuchereria bancrofti* and *Brugia* spp. (*B. malayi* [90%] or *B. timori* [10%]), which are responsible for 120 million infections worldwide. In endemic areas, which include parts of Latin America, sub-Saharan Africa, and Southeast Asia, filariasis causes a spectrum of diseases:

- Asymptomatic microfilaremia
- Recurrent lymphadenitis
- Chronic lymphadenitis with swelling of the dependent limb or scrotum (elephantiasis)
- Tropical pulmonary eosinophilia

Pathogenesis

Infective larvae released by mosquitoes into the tissues during a blood meal develop within lymphatic channels into adult males and females, which mate and release microfilariae that enter into the bloodstream. Mosquitoes that bite infected individuals take up the microfilariae and can transmit the disease. The genomes of *W. bancrofi* and *B. malayi* have been sequenced, leading to the discovery of filarial molecules that may contribute to host invasion and enable these organisms to evade or inhibit immune defenses:

- *Elastases and trypsin-like proteases,* which facilitate invasion of host tissues
- *Several surface glycoproteins* with antioxidant function, which may protect the parasite from reactive oxygen species
- *Homologs of cystatins* (cysteine protease inhibitors), which can impair the MHC class II antigen-processing pathway
- *Serpins* (serine protease inhibitors), which can inhibit neutrophil proteases, critical inflammatory mediators
- *Homologs of host molecules,* such as TGF-β and macrophage migration inhibition factor, which could dampen the immune response.

In addition, symbiotic *Wolbachia* bacteria infect filarial nematodes and contribute to pathogenesis of disease. *Wolbachia* spp. are required for nematode development and reproduction, and antibiotics that eradicate *Wolbachia* spp. impair nematode survival and fertility.

Immune responses to the filarial worms produce damage to the human host. As is the case with leprosy and leishmanial infections, some of the different disease manifestations caused by lymphatic filariae are likely related to variations in host T-cell responses to the parasites. In chronic lymphatic filariasis, damage to the lymphatics is caused directly by the adult parasites and by a Th1-mediated immune response, which stimulates the formation of granulomas around the adult parasites. Hypersensitivity to microfilaria in the lungs is thought to be associated with tropical pulmonary eosinophilia. IgE and eosinophils may be stimulated by IL-4 and IL-5, respectively, which are secreted by filaria-specific Th2 helper T cells.

MORPHOLOGY

Chronic filariasis is characterized by **persistent lymphedema** of the extremities, scrotum, penis, or vulva (Fig. 8.57). Frequently there is hydrocele and lymph node enlargement. In severe and long-lasting infections, chylous weeping of the enlarged scrotum may ensue, or a chronically swollen leg may develop tough subcutaneous fibrosis and epithelial hyperkeratosis, termed **elephantiasis.** Elephantoid skin shows dilation of the dermal lymphatics, widespread lymphocytic infiltrates, and focal cholesterol

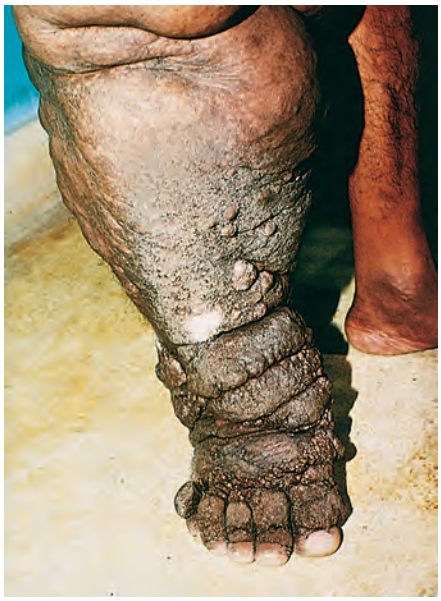


Figure 8.57 Massive edema and elephantiasis caused by filariasis of the leg. (Courtesy Dr. Willy Piessens, Harvard School of Public Health, Boston, Mass.)

deposits; the epidermis is thickened and hyperkeratotic. Adult filarial worms—live, dead, or calcified—are present in the draining lymphatics or nodes, surrounded by (1) mild or no inflammation, (2) an intense eosinophilia with hemorrhage and fibrin (recurrent filarial funiculoepididymitis), or (3) granulomas. Over time, the dilated lymphatics develop polypoid infoldings. In the testis, hydrocele fluid, which often contains cholesterol crystals, red cells, and hemosiderin, induces thickening and calcification of the tunica vaginalis.

Lung involvement by microfilariae is marked by eosinophilia caused by Th2 responses and cytokine production (tropical eosinophilia) or by dead microfilariae surrounded by stellate, hyaline, eosinophilic precipitates embedded in small epithelioid granulomas (Meyers-Kouwenaar bodies). Typically, these patients lack other manifestations of filarial disease.

Onchocerciasis

***Onchocerca volvulus* is a filarial nematode that is the leading cause of preventable blindness in sub-Saharan Africa**. It is transmitted by black flies and affects 17 million people in Africa, South America, and Yemen. An aggressive campaign of ivermectin treatment has dramatically reduced the incidence of *Onchocerca* spp. infections in Africa and South America. The WHO has declared that onchocerciasis has been eliminated in Colombia, Ecuador, Mexico, and Guatemala. Because the vector's habitat is near fast-moving water, there is higher incidence of human disease near rivers, accounting for the name *river blindness* given to this disease.

The disease attributable to onchocerciasis is primarily due to inflammation induced by microfilaria. Adult *O. volvulus* parasites mate in the dermis, where they are surrounded by a mixed infiltrate of host cells that produces a characteristic subcutaneous nodule *(onchocercoma)*. Inseminated females produce microfilariae, which accumulate in the skin and disseminate to the eye chambers. Ivermectin kills only immature worms, not adult worms, so parasites can repopulate the host, requiring retreatment every 3 to 6 months until symptoms are gone.

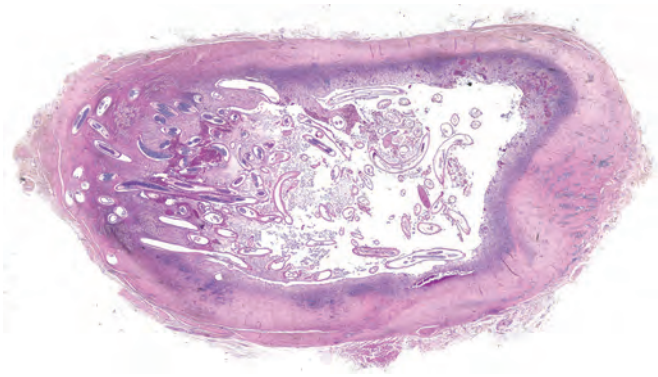


Figure 8.58 Microfilaria-laden gravid female of *Onchocerca volvulus* in a subcutaneous fibrous nodule.

MORPHOLOGY

O. volvulus causes chronic, itchy dermatitis with focal darkening or loss of pigment and scaling, referred to as leopard, lizard, or elephant skin. Foci of epidermal atrophy and elastic fiber breakdown may alternate with areas of hyperkeratosis, hyperpigmentation with pigment incontinence, dermal atrophy, and fibrosis. The subcutaneous onchocercoma is composed of a fibrous capsule surrounding adult worms and a mixed chronic inflammatory infiltrate that includes fibrin, neutrophils, eosinophils, lymphocytes, and giant cells (Fig. 8.58). The progressive eye lesions begin with punctate keratitis along with small, fluffy opacities of the cornea caused by degenerating microfilariae, which evoke an eosinophilic infiltrate. This is followed by a sclerosing keratitis that opacifies the cornea, beginning at the scleral limbus. Keratitis is sometimes accentuated by treatment with antifilarial drugs (Mazzotti reaction). Microfilariae in the anterior chamber cause inflammation of the anterior chamber of the eye and ciliary body (iridocyclitis) and glaucoma, whereas involvement of the choroid and retina results in atrophy and loss of vision.

SEXUALLY TRANSMITTED INFECTIONS

A variety of organisms can be transmitted through sexual contact (Table 8.7). Groups at greater risk for STIs include adolescents, men who have sex with men, and people who use illegal drugs parenterally. Although the increased risk among these groups is partially due to unsafe sexual practices, limited access to health care is often a contributing factor. The presence of an STI in a child, unless acquired during birth, strongly suggests sexual abuse.

Some pathogens, such as *Chlamydia trachomatis* and *N. gonorrhoeae,* are almost always spread by sexual intercourse, whereas others, such as *Shigella* spp. and *E. histolytica,* are typically spread by other means, but are also occasionally spread by oral-anal sex. To reduce the spread of STIs, these

Table 8.7 Classification of Important Sexually Transmitted Infections

Pathogens	Disease or Syndrome and Population Principally Affected		
	Males	**Females**	**Both**
Viruses			
Herpes simplex virus			Primary and recurrent herpes, neonatal herpes
Hepatitis B virus			Hepatitis
Human papillomavirus	Cancer of penis	Cervical dysplasia and cancer, vulvar cancer	Condyloma acuminatum
Human immunodeficiency virus			Acquired immunodeficiency syndrome
Chlamydiae			
Chlamydia trachomatis	Urethritis, epididymitis, proctitis	Urethral syndrome, cervicitis, bartholinitis, salpingitis, and sequelae	Lymphogranuloma venereum
Mycoplasmas			
Ureaplasma urealyticum	Urethritis		
Bacteria			
Neisseria gonorrhoeae	Epididymitis, prostatitis, urethral stricture	Cervicitis, endometritis, bartholinitis, salpingitis, and sequelae (infertility, ectopic pregnancy, recurrent salpingitis)	Urethritis, proctitis, pharyngitis, disseminated gonococcal infection
Treponema pallidum			Syphilis
Haemophilus ducreyi			Chancroid
Klebsiella granulomatis			Granuloma inguinale (donovanosis)
Protozoa			
Trichomonas vaginalis	Urethritis, balanitis	Vaginitis	

infections are often reported to public health authorities so that people who have had intimate contact with the person may be tested and treated.

Although the various pathogens that cause STIs differ in many ways, some general features should be noted.

- *STIs may become established locally and then spread from the urethra, vagina, cervix, rectum, or oral pharynx*. Organisms that cause STIs depend on direct contact for person-to-person spread because these pathogens do not survive in the environment. Transmission of STIs often occurs from asymptomatic people who do not realize that they have an infection.
- *Infection with one STI-associated organism increases the risk for additional STIs*. This is mainly because the risk factors are the same for all STIs. In addition, the epithelial injury caused by *N. gonorrhoeae* or *C. trachomatis* can increase the chance of co-infection with the other, as well as the risk of HIV infection if there is concomitant exposure.
- *The microbes that cause STIs can be spread from a pregnant woman to the fetus and cause severe damage to the fetus or child*. Perinatally acquired *C. trachomatis* causes conjunctivitis, and neonatal HSV infection is much more likely to cause visceral and CNS disease than is infection acquired later in life. Syphilis frequently causes miscarriage. Untreated HIV infection may be fatal to children infected with the virus prenatally or perinatally. Diagnosis of STIs in pregnant women is critical, because intrauterine or neonatal transmission of STIs can often be prevented by treatment of the mother or newborn. Antiretroviral treatment of pregnant women with HIV infection and their newborn infants can reduce the transmission of HIV to offspring from 25% to less than 1%.

Syphilis is discussed earlier in this chapter, and other STIs are described in Chapters 21 and 22.

EMERGING INFECTIOUS DISEASES

The rapidly expanding human population juxtaposed with environmental infractions allow the emergence of new pathogens and the re-emergence of old infectious agents. Table 8.8 lists the history of diseases that have emerged over the past half century. The infectious causes of some diseases were not new, but previously unrecognized because some of the infectious agents are difficult to culture; examples include *H. pylori* gastritis, HBV and HCV, and *Legionella pneumophila*. Some infectious agents are new to humans, such as HIV causing AIDS, and *B. burgdorferi* causing Lyme disease. Other infections have become much more common because of immunosuppression caused by AIDS or immunosuppressive therapy (e.g., CMV, KSHV, *M. avium-intracellulare, P. jirovecii*, and *Cryptosporidium parvum*). Finally, infectious diseases that are common in one area may be introduced into a new area. For example, West Nile virus was common in Europe, Asia, and Africa for years before it was described in the United States.

Human demographics and behavior are important contributors to the pattern of emergence of infectious diseases, such as AIDS in Africa versus AIDS in the United States. Annual outbreaks of Nipah virus infection have occurred in Bangladesh due to consumption of date palm sap. Bats, which shed the virus in their saliva and urine, are thought to contaminate the sap by feeding on it while it is being collected from the tree. Nipah virus can also

Table 8.8 Diseases That Have Emerged or Reemerged Over the Past 45 Years

Date Recognized	Infectious Agent	Manifestations
1977	Ebola virus	Epidemic Ebola hemorrhagic fever
	Hantaan virus	Hemorrhagic fever with renal syndrome
	Legionella pneumophila	Legionnaire disease
	Campylobacter jejuni	Enteritis
1980	Human T-lymphotropic virus I (HTLV-I)	T-cell lymphoma or leukemia, HTLV-associated myelopathy
1981	*Staphylococcus aureus*	Toxic shock syndrome
1982	*Escherichia coli* O157:H7	Hemorrhagic colitis, hemolytic-uremic syndrome
	Borrelia burgdorferi	Lyme disease
1983	Human immunodeficiency virus (HIV)	AIDS
	Helicobacter pylori	Gastric ulcers
1988	Hepatitis E	Enterically transmitted hepatitis
1989	Hepatitis C	Hepatitis C
1992	*Vibrio cholerae* O139	New epidemic cholera strain
	Bartonella henselae	Cat-scratch disease
1995	Kaposi sarcoma–associated virus (HHV-8)	Kaposi sarcoma in AIDS
1999	West Nile virus	West Nile fever, neuroinvasive disease
2003	Severe acute respiratory syndrome (SARS) coronavirus	SARS
2007	Zika virus	Fever, rash, Guillain-Barré syndrome Fetal loss, microcephaly, neurologic complications in newborns
2014	Ebola virus	Fever, rash, hemorrhage, multisystem organ failure
2016	*Candida auris*	Candidemia, wound and ear infections resistant to multiple antifungal agents
2019	COVID-19	Fever, cough, dyspnea, pneumonia

spread by person-to-person contact, and by transmission from pigs to people. Changes in the environment occasionally drive rates of infectious diseases. Reforestation of the eastern United States has led to massive increases in the populations of deer and mice, which carry the ticks that transmit Lyme disease, babesiosis, and ehrlichiosis. Failure of DDT to control the mosquitoes that transmit malaria and the development of drug-resistant parasites have dramatically increased the morbidity and mortality of *Plasmodium falciparum* infection in Asia, Africa, and Latin America. Microbial adaptation to widespread antibiotic use contributed to the emergence of drug resistance in many species of bacteria, including *M. tuberculosis, N. gonorrhoeae, S. aureus,* and *K. pneumoniae*. Human commercial use of dense populations of domestic animals (e.g., pigs, chickens) juxtaposed to habitat destruction of other disease reservoirs (e.g., bats and wild birds) can lead to acquisition of either unique traits in common pathogens such as influenza or emergence of unique viruses such as severe acute respiratory syndrome (SARS) virus, West Nile Virus, and Ebola virus. Because these pathogens are novel, humans lack immunity and so these infections can quickly spread through the population as pandemics, as was seen with influenza A H1N1 in 2009. Causes of more recent emerging infectious diseases discussed in this chapter include Ebola virus, Zika virus, COVID-19, *C. auris*, and some of the *Babesia* spp.

Agents of Bioterrorism. Bioterrorism is the use of biologic or chemical agents as weapons, and microorganisms are classified based on an assessment of which pose the greatest danger. The CDC classifies potential agents of bioterrorism based on the risk they present and whether they can be easily disseminated.

- Category A agents pose the highest risk and can be readily disseminated or transmitted from person to person, can cause high mortality, might cause public panic, and might require public health preparedness. For example, smallpox is a category A agent because of its high transmissibility, case mortality rate of 30% or greater, and lack of effective antiviral therapy. Because vaccination ended in the United States in 1972 and immunity has waned, the population is highly susceptible to smallpox. Concern that smallpox could be used for bioterrorism has led to a return of vaccination for selected groups. Other category A agents include *B. anthracis, Yersinia pestis,* and Ebola virus.
- Category B agents are relatively easy to disseminate, produce moderate morbidity but low mortality, and require specific diagnostic and disease surveillance. Many of these agents are food-borne or water-borne. Examples include *Brucella* spp. and *V. cholerae*.
- Category C agents include emerging pathogens that could be engineered for mass dissemination because of availability, ease of production and dissemination, potential for high morbidity and mortality, and great impact on health. Examples include Hantavirus and Nipah virus.

SPECIAL TECHNIQUES FOR DIAGNOSING INFECTIOUS AGENTS

The gold standards for diagnosis of infections are identification from culture, serology, and molecular techniques, depending on the organism in question. Some infectious agents or their products can be directly observed in hematoxylin and eosin–stained sections (e.g., the inclusion bodies formed by CMV and HSV; bacterial clumps, which usually stain blue). Many infectious agents, however, are best visualized by special stains that identify organisms on the basis of particular characteristics of their cell wall or coat or by staining with specific antibodies (Table 8.9). Organisms are typically easiest to identify at the advancing edge of a

Table 8.9 Special Techniques for Diagnosing Infectious Agents

Techniques	Infectious Agents
Gram stain	Most bacteria
Acid-fast stain	*Mycobacteria* spp., *Nocardia* spp. (modified)
Silver stains	Fungi, *Legionella* spp., *Pneumocystis jirovecii*
Periodic acid-Schiff	Fungi, amebae
Mucicarmine	*Cryptococcus* spp.
Giemsa	*Campylobacter* spp., *Leishmania* spp., malarial parasites
Antibody stains	All classes
Culture	All classes
DNA probes	All classes

lesion rather than at its center, particularly if there is necrosis. Acute infections can be diagnosed serologically by detecting pathogen-specific antibodies in the serum. The presence of specific IgM antibody shortly after the onset of symptoms is diagnostic. Alternatively, specific antibody titers can be measured during the early infection and again 4 to 6 weeks later; a fourfold rise in titer is considered diagnostic.

Nucleic acid amplification tests, such as PCR and transcription-mediated amplification, are common for rapid identification of microbes. These molecular diagnostic assays have become routine for diagnosis of gonorrhea, chlamydial infection, tuberculosis, and herpes encephalitis. In some cases, molecular assays are much more sensitive than conventional testing, for example for detection of HSV in CSF or for Chlamydia in urine or genital swabs. Multiplex PCR panels for detection of over 20 pathogens are now in clinical use for detection of pathogens in stool, respiratory samples, and cerebrospinal fluid, as a well as positive blood culture broths. In people infected with HIV, HBC, or HCV, quantification of viral RNA is an important guide to management of ART. High-throughput or next-generation sequencing methods, most often following PCR of the 16S rDNA gene for bacteria or the ribosomal internal transcribed spacer (ITS) region for fungi, allow rapid sequencing of large numbers of DNA molecules. Next-generation sequencing can be used to detect bacteria, viruses, parasites, or fungi directly in patient specimens. This method currently has limited clinical application, but is likely to become increasingly important and common in the future.

SUGGESTED READINGS

General Principles of Microbial Pathogenesis

Centers for Disease Control and Prevention: https://www.cdc.gov/. *[An extremely useful website resource maintained by the key US public health center.]*

Dutta D, Clevers H: Organoid culture systems to study host-pathogen interactions, *Curr Opin Immunol* 48:15, 2017. *[A good introduction to the utility of organoids for in vitro studies of pathogenesis.]*

Lee PP, Lau YL: Cellular and molecular defects underlying invasive fungal infections-revelations from endemic mycoses, *Front Immunol* 8:735, 2017. *[Discusses how specific immune defects provide the basis for an understanding of the host-fungus interaction during infection.]*

Santi-Rocca J, Blanchard N: Membrane trafficking and remodeling at the host-parasite interface, *Curr Opin Microbiol* 40:145, 2017. *[A review of how protist parasites alter endosomal membranes so they can grow within host cells.]*

Wheeler ML, Limon JJ, Underhill DM: Immunity to commensal fungi: detente and disease, *Annu Rev Pathol* 12:359, 2017. *[A very useful introduction to the fungal microbiota, and how these organisms persist within the host.]*

Virology

Baseler L, Chertow DS, Johnson KM et al: The pathogenesis of Ebola virus disease, *Annu Rev Pathol* 12:387, 2017. *[A comprehensive introduction to Ebola, with in-depth discussion of pathogenesis of Ebola virus infection.]*

Brown JC: Herpes simplex virus latency: the DNA repair-centered pathway, *Adv Virol* 2017:7028194, 2017. *[A concise summary of the mechanism of latency used by HSV.]*

Graham BS, Sullivan NJ: Emerging viral diseases from a vaccinology perspective: preparing for the next pandemic, *Nat Immunol* 19:20, 2018. *[Provides an overview of the challenges to vaccine development and the emerging technologies to address these challenges.]*

Grose C, Buckingham EM, Carpenter JE et al: Varicella-zoster virus infectious cycle: ER stress, autophagic flux, and amphisome-mediated trafficking, *Pathogens* 5:2016. *[A nice summary of VZV induced cellular stress and autophagic flux during infection.]*

Musso D, Gubler DJ: Zika virus, *Clin Microbiol Rev* 29:487, 2016. *[An excellent, comprehensive discussion of Zika virus infection.]*

Taylor GS, Long HM, Brooks JM et al: The immunology of Epstein-Barr virus-induced disease, *Annu Rev Immunol* 33:787, 2015. *[A good discussion of EBV and the associated host immune response.]*

Bacteriology

Abel L, Fellay J, Haas DW et al: Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives, *Lancet Infect Dis* 18:e64, 2018. *[Introduces our relatively limited understanding of the host genetics underlying vulnerability to tuberculosis.]*

Awuh JA, Flo TH: Molecular basis of mycobacterial survival in macrophages, *Cell Mol Life Sci* 74:1625–1648, 2017. *[Describes the mechanisms by which this intracellular pathogen avoids destruction within macrophages.]*

Dal Peraro M, van der Goot FG: Pore-forming toxins: ancient, but never really out of fashion, *Nat Rev Microbiol* 14:77, 2016. *[An interesting summary of the structure of pore-forming toxins, and how these assemble on host membranes.]*

Hinnebusch BJ, Jarrett CO, Bland DM: "Fleaing" the plague: adaptations of *Yersinia pestis* to its insect vector that lead to transmission, *Annu Rev Microbiol* 71:215, 2017. *[An excellent review of the biology of Yersinia and its vector.]*

Hryckowian AJ, Pruss KM, Sonnenburg JL: The emerging metabolic view of *Clostridium difficile* pathogenesis, *Curr Opin Microbiol* 35:42, 2017. *[A concise presentation of a model of* C. difficile *pathogenesis focused on the metabolic state of the gut.]*

Kilgore PE, Salim AM, Zervos MJ et al: Pertussis: microbiology, disease, treatment, and prevention, *Clin Microbiol Rev* 29:449, 2016. *[A comprehensive review of* Bordetella pertussis *and associated disease.]*

Koh WJ: Nontuberculous Mycobacteria-Overview, *Microbiol Spectr* 5:2017. *[Provides an introduction to the epidemiology, manifestations, and diagnostic testing for these bacteria.]*

Radolf JD, Deka RK, Anand A et al: *Treponema pallidum*, the syphilis spirochete: making a living as a stealth pathogen, *Nat Rev Microbiol* 14:744, 2016. *[A good review of the pathogenesis of* Treponema pallidum.*]*

Radoshevich L, Cossart P: *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis, *Nat Rev Microbiol* 16:32, 2018. *[An excellent and comprehensive review of the pathogenesis of* Listeria.*]*

de Sousa JR, Sotto MN, Simoes Quaresma JA: Leprosy as a complex infection: breakdown of the Th1 and Th2 immune paradigm in the immunopathogenesis of the disease, *Front Immunol* 8:1635, 2017. *[A short discussion of Th1 and Th2 immune responses to* M. leprae.*]*

Thomer L, Schneewind O, Missiakas D: Pathogenesis of *Staphylococcus aureus* bloodstream infections, *Annu Rev Pathol* 11:343, 2016.

[An excellent review of the pathogenesis of Staphylococcus aureus *infection.]*

Fungi

Alanio A, Desnos-Ollivier M, Garcia-Hermoso D et al: Investigating clinical issues by genotyping of medically important fungi: why and how?, *Clin Microbiol Rev* 30:671, 2017. *[A discussion of the changing methods of classifying medically important fungi.]*

Hoving JC, Kolls JK: New advances in understanding the host immune response to *Pneumocystis*, *Curr Opin Microbiol* 40:65, 2017. *[An excellent review of how the understanding of the genetics of* Pneumocystis *and of the host illuminates the understanding of the host response.]*

Jeffery-Smith A, Taori SK, Schelenz S et al: *Candida auris*: a Review of the Literature, *Clin Microbiol Rev* 31:2018. *[An outstanding and comprehensive discussion of our current understanding of this emerging pathogen.]*

May RC, Stone NR, Wiesner DL et al: *Cryptococcus*: from environmental saprophyte to global pathogen, *Nat Rev Microbiol* 14:106, 2016. *[A comprehensive review of* Cryptococcus *biology disease.]*

Noble SM, Gianetti BA, Witchley JN: *Candida albicans* cell-type switching and functional plasticity in the mammalian host, *Nat Rev Microbiol* 15:96, 2017. *[An excellent review of mechanisms of adaptability that* Candida *uses for virulence.]*

Perlin DS, Rautemaa-Richardson R, Alastruey-Izquierdo A: The global problem of antifungal resistance: prevalence, mechanisms, and management, *Lancet Infect Dis* 17:e383, 2017. *[A valuable introduction to the epidemiology and mechanisms of resistance in drug-resistant fungi.]*

van de Veerdonk FL, Gresnigt MS, Romani L et al: *Aspergillus fumigatus* morphology and dynamic host interactions, *Nat Rev Microbiol* 15:661, 2017. *[Introduces how* Aspergillus fumigatus *adapts to different environments and resists killing by the host immune system.]*

Parasitology

Buscher P, Cecchi G, Jamonneau V et al: Human African trypanosomiasis, *Lancet* 390:2397, 2017. *[A concise overview of disease caused by trypanosomes.]*

Coppens I: How Toxoplasma and malaria parasites defy first, then exploit host autophagic and endocytic pathways for growth, *Curr Opin Microbiol* 40:32, 2017. *[A concise discussion of the pathogenesis of intracellular pathogens.]*

Duncan SM, Jones NG, Mottram JC: Recent advances in *Leishmania* reverse genetics: manipulating a manipulative parasite, *Mol Biochem Parasitol* 216:30, 2017.

Grote A, Lustigman S, Ghedin E: Lessons from the genomes and transcriptomes of filarial nematodes, *Mol Biochem Parasitol* 215:23, 2017.

Ramirez-Toloza G, Ferreira A: *Trypanosoma cruzi* evades the complement system as an efficient strategy to survive in the mammalian host: the specific roles of host/parasite molecules and *Trypanosoma cruzi* calreticulin, *Front Microbiol* 8:1667, 2017. *[An excellent and comprehensive review of the pathogenesis of* Listeria.*]*

Wahlgren M, Goel S, Akhouri RR: Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria, *Nat Rev Microbiol* 15:479, 2017. *[A good review of the mechanism of antigenic variation of* Plasmodium *spp.]*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 9

Environmental and Nutritional Diseases

CHAPTER CONTENTS

Many diseases are caused or influenced by environmental factors. Broadly defined, the term *environment* encompasses the various indoor, outdoor, and occupational settings in which human beings live and work. In each of these settings, the air people breathe, the food and water they consume, and the toxic agents they are exposed to are major determinants of health. The environmental factors that influence our health pertain to individual behavior ("personal environment") and include tobacco use, alcohol ingestion, recreational drug consumption, diet, and the like, or the external (ambient and workplace) environment. In general, in higher income countries personal behavior has a larger effect on health than the ambient environment, but new threats related to global warming (described later) may change this equation.

The term *environmental disease* refers to conditions caused by exposure to chemical or physical agents in the ambient, workplace, and personal environment, including diseases of nutritional origin. The International Labour Organization estimates that work-related injuries and illnesses kill approximately 2.3 million people per year globally (more deaths than are caused by road accidents and wars combined). In the United States in 2018, there were nearly 3 million occupational injuries and illnesses. Disease related to malnutrition is even more pervasive. In 2019, it was estimated that 795 million people were malnourished—one in every nine persons worldwide. Children are disproportionately affected by undernutrition, which accounts for approximately 50% of childhood mortality worldwide. Estimating the burden of disease in the general population caused by nonoccupational exposures to toxic agents is complicated by the diversity of agents and difficulties in determining the extent and duration of exposures. But whatever the precise numbers, it is clear that environmental diseases are major causes of disability and suffering and constitute a heavy financial burden, particularly in low income countries.

In this chapter, we first consider the emerging problem of the health effects of climate change. We then discuss the mechanisms of toxicity of chemical and physical agents and address specific environmental disorders, including those of nutritional origin.

HEALTH EFFECTS OF CLIMATE CHANGE

Without immediate action, climate change stands to become the preeminent global cause of environmental disease in the 21st century and beyond. Global temperature measurements show that the earth has warmed significantly since the early 20th century and especially since the mid-1960s. Record-breaking global temperatures have become common, and the 5 years from 2014 to 2018 have been the warmest since 1880. During 2018, the global land temperature was 1.1°C warmer than the 20th century average. Mean global ocean temperatures also continue to warm, with the annual average temperature in 2018 being 0.66°C warmer than the 20th century average.

The rising atmospheric and oceanic temperatures have led to a large number of effects that include changes in storm frequency, drought, and flood, as well as large-scale ice losses in Greenland, Antarctica, and the vast majority of the other glaciated regions on earth, as well as dramatic thinning or disappearance of Arctic Ocean sea ice. The melting of land-based glacial ice and the thermal expansion of the warming oceans has produced approximately 13–20 cm of global average sea level rise since 1900, and the sea level currently is rising at a global average rate of 3.5 ± 0.4 mm/year.

Although politicians quibble, among scientists there is a general acceptance that climate change is in large part man-made. The culprit is the rising atmospheric level of greenhouse gases, particularly carbon dioxide (CO_2) released through the burning of fossil fuels (Fig. 9.1A), ozone (an important air pollutant, discussed later), and methane. These gases, along with water vapor, produce the so-called greenhouse effect by absorbing energy radiated from Earth's surface that otherwise would be lost into space. The annual average level of atmospheric CO_2 (about 410 ppm) in 2019 was higher than at any point in approximately 650,000 years and, without changes in human behavior, is expected to increase to 500 to 1200 ppm by the end of this century—levels not experienced for tens of millions of years. This increase stems not only from increased CO_2 production but also from deforestation and the attendant decrease in carbon fixation by plants. Depending on the computer model used, increased levels of greenhouse gases are projected to cause the global temperature to rise by 2°C to 5°C by the year 2100 (Fig. 9.1B).

The health consequences of climate change will depend on its extent and rapidity, the severity of the ensuing consequences, and humankind's ability to mitigate the damaging effects. The World Health Organization (WHO) estimates that approximately 250,000 deaths would occur annually between 2030 and 2050 as a consequence of climate change. This number does not include morbidity and disruption of health services from extreme changes in weather. Even in the best-case scenario, however, climate change is expected to have a serious negative impact on human health by increasing the incidence of a number of diseases including the following:

- *Cardiovascular, cerebrovascular, and respiratory diseases,* all of which will be exacerbated by heat waves and air pollution.

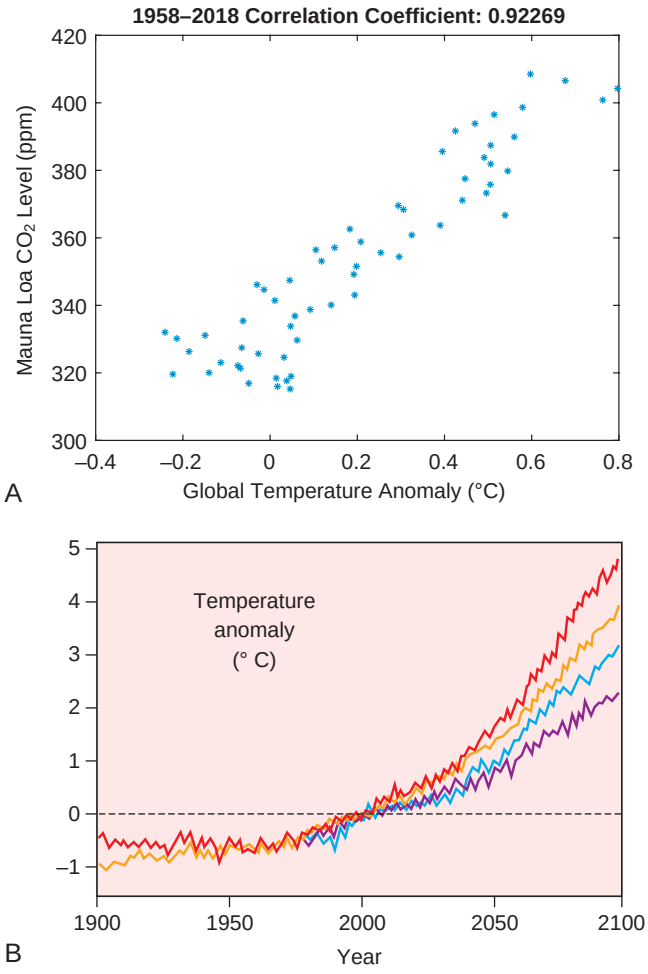


Figure 9.1 Climate change, past and future. (A) Correlation of carbon dioxide *(CO_2)* levels measured at the Mauna Loa Observatory in Hawaii with average global temperature trends over the past 60 years. Global temperature in any given year was deduced at the Hadley Center (United Kingdom) from measurements taken at more than 3000 weather stations located around the globe. (B) Predicted temperature increases during the 21st century. Different computer models plot anticipated rises in global temperatures of 2°C to 5°C by the year 2100. (A, Courtesy Dr. Richard Aster, Department of Geophysics, Colorado State University, Fort Collins, Colo.)

- *Gastroenteritis,* cholera, and other foodborne and waterborne infectious diseases, caused by contamination as a consequence of floods and disruption of clean water supplies and sewage treatment after heavy rains and other environmental disasters.
- *Vector-borne infectious diseases* such as malaria and dengue fever resulting from changes in vector number and geographic distribution related to increased temperatures, crop failures, and more extreme weather variation (e.g., more frequent and severe El Niño events).
- *Malnutrition,* caused by changes in local climate that disrupt crop production. Such changes are anticipated to be most severe in tropical locations, in which average temperatures may already be near or above crop tolerance levels; it is estimated that by 2080, agricultural productivity may decline by 10% to 25% in some low income countries as a consequence of climate change.

Beyond these disease-specific effects, it is estimated that the melting of glacial ice, particularly in Greenland and other parts of the Northern Hemisphere, combined with the thermal expansion of warming oceans will raise sea levels by 2 to 6 feet by 2100. Approximately 10% of the world's population—roughly 600 million people—live in low-lying areas that are at risk for flooding even if the rise in ocean levels is at the low end of these estimates. For example, a rise in sea level by 1.5 feet will submerge 70% of the land mass of the Maldive islands, while a 3-foot rise will inundate 100% of all of the islands, changes in sea level that are projected to occur by 2100. The resulting displacement of people will disrupt lives and commerce, creating conditions ripe for political unrest, war, and poverty—the "vectors" of malnutrition, sickness, and death.

Worldwide recognition of the potentially catastrophic effects of climate change led in late 2015 to a historic meeting of 196 countries in Paris, France, at which the participating countries agreed to the following objective: To hold the increase in the global average temperature to well below 2°C above preindustrial levels and to pursue efforts to limit the temperature increase to 1.5°C above preindustrial levels, recognizing that this would significantly reduce the risks and impacts of climate change. However, in 2017 the United States decided to withdraw from this agreement as of 2020, leading to uncertainty about the world's ability to meet the goals established in the Paris accord.

TOXICITY OF CHEMICAL AND PHYSICAL AGENTS

***Toxicology* is defined as the science of poisons. It studies the distribution, effects, and mechanisms of action of toxic agents.** More broadly, it also includes the study of the effects of physical agents such as radiation and heat. Of the approximately 100,000 chemicals in commercial use in the United States, only a small proportion has been tested experimentally for health effects. Several agencies in the United States set permissible levels of exposure to known environmental hazards (e.g., the maximum level of carbon monoxide [CO] in air that is noninjurious or the tolerable levels of radiation that are "safe"). Factors such as the complex interaction between various pollutants and the age, genetic predisposition, and different tissue sensitivities of exposed persons create wide variations in individual sensitivity to toxic agents, limiting the value of establishing "safe levels" for entire populations. Nevertheless, such cutoffs are useful for comparative studies of the effects of agents between specific populations and for estimating risk of disease in heavily exposed individuals.

We now consider some basic principles relevant to the effects of toxic chemicals and drugs.

- The *definition of a poison* is not straightforward. It is basically a quantitative concept that depends on dosage. The quote from Paracelsus in the 16th century that "all substances are poisons; the right dosage differentiates a poison from a remedy" is still valid today, given the number of pharmaceutical drugs with potentially harmful effects.
- *Xenobiotics* are exogenous chemicals in the environment in air, water, food, and soil that may be absorbed into the body through inhalation, ingestion, and skin contact (Fig. 9.2).
- Chemicals may be excreted in urine or feces; eliminated in expired air; or accumulate in bone, fat, brain, or other tissues.
- Chemicals may act at the site of entry or at other sites following transport through the blood.
- *Most solvents and drugs are lipophilic,* which facilitates their transport in the blood by lipoproteins and their penetration through the plasma membrane into cells.
- Most solvents, drugs, and xenobiotics are metabolized to form inactive water-soluble products (detoxification) or are activated to form toxic metabolites. The reactions

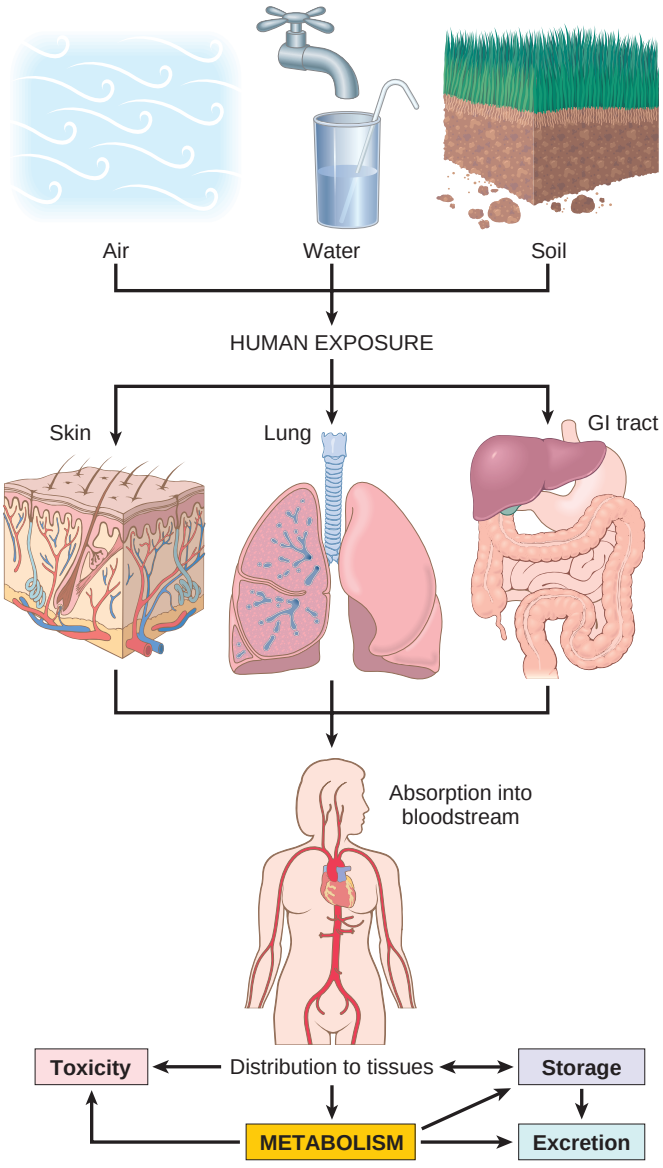


Figure 9.2 Human exposure to pollutants. Pollutants in the air, water, and soil are absorbed through the lungs, gastrointestinal *(GI)* tract, and skin. In the body they may act at the site of absorption, but are generally transported through the bloodstream to various organs where they may be stored or metabolized. Xenobiotics may be metabolized to water-soluble compounds that are excreted or to toxic metabolites, a process referred to as activation.

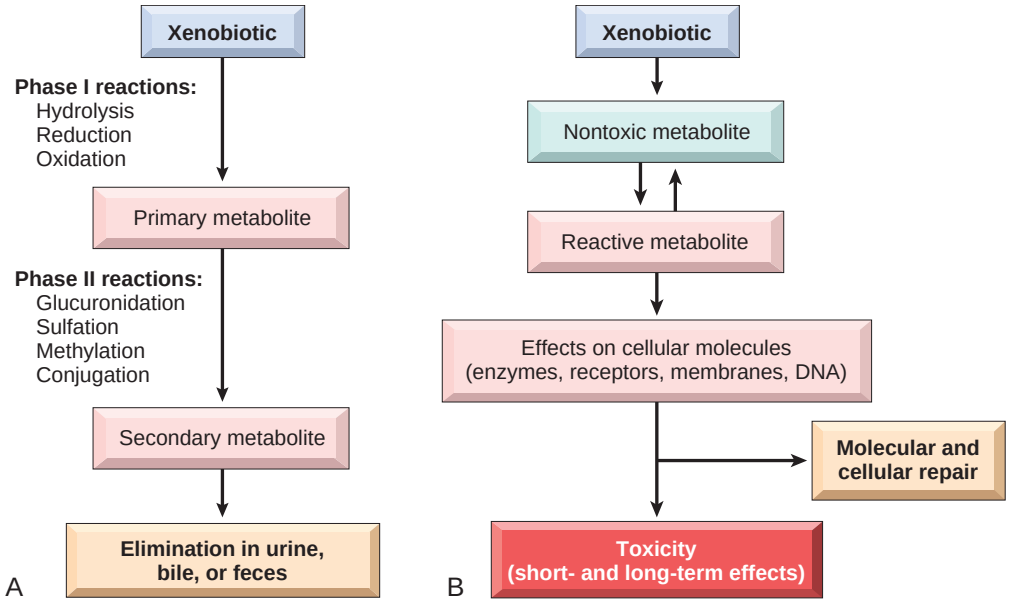


Figure 9.3 Xenobiotic metabolism. (A) Xenobiotics can be metabolized to nontoxic metabolites and eliminated from the body (detoxification). (B) Xenobiotic metabolism may also result in the formation of a reactive metabolite that is toxic to cellular components. If repair is not effective, short-term and long-term effects develop. (Based on Hodgson E: *A textbook of modern toxicology,* ed 3, Hoboken, NJ, 2004, Wiley.)

that metabolize xenobiotics into nontoxic products or that activate xenobiotics to generate toxic compounds (Figs. 9.2 and 9.3) occur in two phases. In phase I reactions, chemicals undergo hydrolysis, oxidation, or reduction. Products of phase I reactions are often metabolized into water-soluble compounds through phase II reactions, which include glucuronidation, sulfation, methylation, and conjugation with glutathione (GSH). Water-soluble compounds are readily excreted.

- The most important catalyst of phase I reactions is the cytochrome P-450 enzyme system. Cytochrome P-450 enzymes (CYPs) are a large family of heme-containing enzymes, each with preferred substrate specificities. These enzymes are primarily expressed in hepatocytes, in which they localize to the endoplasmic reticulum, but can also be found in skin, lungs, gastrointestinal mucosa, and other organs. **The P-450 system catalyzes reactions that either detoxify xenobiotics or, less commonly, convert xenobiotics into active compounds that cause cellular injury.** Both types of reactions may produce, as a by-product, reactive oxygen species (ROS), which can cause cellular damage (Chapter 2). Examples of metabolic activation of chemicals through CYPs are the production of the toxic trichloromethyl free radical from carbon tetrachloride in the liver and the generation of a DNA-binding metabolite from benzo[*a*]pyrene, a carcinogen present in cigarette smoke, in the lung. CYPs participate in the metabolism of alcohol (discussed later) and a large number of common therapeutic drugs, such as acetaminophen, barbiturates, warfarin, and anticonvulsants.

There is great variation in the activity of CYPs among individuals. The variation may be a consequence of genetic polymorphisms in specific CYPs, but more commonly it is due to exposure to drugs or chemicals that induce or diminish CYP activity. Known CYP inducers include environmental chemicals, drugs, smoking, alcohol, and hormones. In contrast, fasting or starvation can decrease CYP activity. Inducers of CYP act by binding to specific nuclear receptors, which then heterodimerize with the *retinoic X receptor (RXR)* to form a transcriptional activation complex that associates with promoter elements located in the 5′-flanking region of CYP genes. Nuclear receptors participating in CYP induction include the aryl hydrocarbon receptor, the peroxisome proliferator-activated receptors *(PPARs),* and two orphan nuclear receptors, constitutive androstane receptor (CAR) and pregnane X receptor (PXR).

This brief overview of the general mechanisms of toxicity provides the background for the discussion of environmental diseases presented in this chapter.

ENVIRONMENTAL POLLUTION

Air Pollution

Air pollution is a significant cause of morbidity and mortality worldwide, particularly among at-risk individuals with preexisting pulmonary or cardiac disease. Air is precious to life, but can also carry many potential causes of disease. Airborne microorganisms have long been major causes of morbidity and mortality, especially in low income countries. More widespread are airborne chemical and particulate pollutants, especially in high income nations. Here we consider these hazards in outdoor and indoor air.

Outdoor Air Pollution

The ambient air in industrialized nations is contaminated with an unsavory mixture of gaseous and particulate pollutants, more heavily in cities and in proximity to heavy industry. In the United States, the Environmental Protection Agency monitors and sets allowable upper limits for six

pollutants: sulfur dioxide, CO, ozone, nitrogen dioxide, lead, and particulate matter. Collectively, these agents produce the well-known *smog* (smoke and fog) that sometimes stifles large cities such as Beijing, Los Angeles, Houston, Cairo, New Delhi, Mexico City, and São Paulo. It may seem that air pollution is a modern phenomenon, but this is hardly the case. John Evelyn wrote in 1661 that inhabitants of London suffered from "Catharrs, Phthisicks and Consumptions" (bronchitis, pneumonia, and tuberculosis) and breathed "nothing but an impure and thick mist, accompanied by a fuliginous and filthy vapour, which renders them obnoxious to a thousand inconveniences, corrupting the lungs, and disordering the entire habit of their bodies." The first environmental control law, proclaimed by Edward I in 1306, was straightforward in its simplicity: "whoever should be found guilty of burning coal shall suffer the loss of his head." What has changed in modern times is the nature and sources of air pollutants and the types of regulations that control their emission.

Although the lungs bear the brunt of the adverse consequences, air pollutants can affect many organ systems. Except for some comments on smoking, pollutant-caused lung diseases are discussed in Chapter 15. Major health effects of outdoor pollutants are summarized in Table 9.1. Ozone, sulfur dioxide, particulates, and CO are discussed here.

Ozone (O_3) is produced by interaction of ultraviolet (UV) radiation and oxygen (O_2) in the stratosphere and naturally accumulates in the so-called ozone layer 10 to 30 miles above the earth's surface. This layer protects life on earth by absorbing the most dangerous UV radiation emitted by the sun. In 1985, it was discovered that ozone was nearly completely depleted over Antarctica and was thinned elsewhere, a dire effect stemming from the widespread use of chlorofluorocarbon gases in air conditioners and refrigerators and as aerosol propellants. When released into the atmosphere, these gases drift up into the stratosphere and participate in chemical reactions that destroy ozone. Due to prevailing stratospheric air currents, the resulting depletion is most profound in polar regions, particularly over Antarctica during the winter months. Recognition of the problem led in 1987 to the Montreal Protocol, international agreements that call for a complete phase-out of chlorofluorocarbon use by 2020 in higher income countries and by 2040 in lower income countries. Decreased use of chlorofluorocarbons over the past 30 years has reduced the size of the yearly ozone "hole" over Antarctica, indicating that this global environmental challenge is being met successfully.

Table 9.1 Health Effects of Outdoor Air Pollutants

Pollutant	Populations at Risk	Effects
Ozone	Healthy adults and children	Decreased lung function Increased airway reactivity Lung inflammation
	Athletes, outdoor workers	Decreased exercise capacity
	Asthmatics	Increased hospitalizations
Nitrogen dioxide	Healthy adults	Increased airway reactivity
	Asthmatics	Decreased lung function
	Children	Increased respiratory infections
Sulfur dioxide	Healthy adults	Increased respiratory symptoms
	Individuals with chronic lung disease	Increased mortality
	Asthmatics	Increased hospitalization Decreased lung function
Acid aerosols	Healthy adults	Altered mucociliary clearance
	Children	Increased respiratory infections
	Asthmatics	Decreased lung function Increased hospitalizations
Particulates	Children	Increased respiratory infections
	Individuals with chronic lung or heart disease	Decreased lung function
	Asthmatics	Excess mortality Increased attacks

From Bascom R, et al: Health effects of outdoor air pollution, *Am J Respir Crit Care Med* 153:477, 1996.

In contrast to the "good" ozone in the stratosphere, ozone that accumulates in the lower atmosphere (ground-level ozone) is one of the most pernicious air pollutants. Ground-level ozone is a gas formed by the reaction of nitrogen oxides and volatile organic compounds in the presence of sunlight. These chemicals are released by industrial emissions and motor vehicle exhaust. Ozone toxicity is in large part mediated by the production of free radicals, which injure epithelial cells along the respiratory tract and type I alveolar cells and cause the release of inflammatory mediators. Healthy individuals exposed to ozone experience upper respiratory tract inflammation and mild symptoms (decreased lung function and chest discomfort), but exposure is much more dangerous for people with asthma or emphysema.

To make matters worse, ozone often combines with other agents such as sulfur dioxide and particulates to create a veritable "witches' brew" of pollutants. Sulfur dioxide is produced by power plants burning coal and oil, from copper smelting, and as a by-product of paper mills. Released into the air, it may be converted into sulfuric acid and sulfuric trioxide, which cause a burning sensation in the nose and throat, difficulty in breathing, and asthma attacks in susceptible individuals.

Particulate matter (known as "soot") is a particularly important cause of morbidity and mortality related to pulmonary inflammation and secondary cardiovascular effects. Based on studies of large cities in the United States, it is estimated that there is a 0.5% increase in overall daily mortality for every 10 mg/m^3 increase in 10-µm particles in outdoor air, mainly due to exacerbations of pulmonary and cardiac disease. Particulates are emitted by coal- and oil-fired power plants, by industrial processes burning these fuels, and by diesel exhaust. Although the particles have not been well characterized chemically or physically, fine or ultrafine particles less than 10 µm in diameter are the most harmful. They are readily inhaled into the alveoli, where they are phagocytosed by macrophages and neutrophils, which respond by releasing a number of inflammatory mediators. In contrast, particles that are greater than 10 µm in diameter are of lesser consequence because they are generally removed in the nose or trapped by the mucociliary epithelium of the airways.

CO is a systemic asphyxiant that is an important cause of accidental and suicidal death. CO is a nonirritating, colorless, tasteless, odorless gas that is produced during any process that results in the incomplete oxidation of hydrocarbons. From the standpoint of human health, the most important environmental source of CO is the burning of carbonaceous materials, as occurs in automotive engines, furnaces, and cigarettes. CO is short-lived in the atmosphere, being rapidly oxidized to CO_2; thus elevated levels in ambient air are transient and occur only in close proximity to sources of CO. Chronic poisoning may occur in individuals working in environments such as tunnels, underground garages, and highway toll booths with high exposures to automobile fumes. Of greater concern is acute toxicity. In a small, closed garage, the average running car can produce sufficient CO to induce coma or death within 5 minutes, and CO concentrations can also rapidly rise to toxic levels with improper use of gasoline-powered generators (e.g., during power outages) or following mine fires. CO kills in part by inducing central nervous system (CNS) depression, which appears so insidiously that victims are often unaware of their plight. Hemoglobin has 200-fold greater affinity for CO than for oxygen, and the resultant carboxyhemoglobin cannot carry O_2. Systemic hypoxia develops when the hemoglobin is 20% to 30% saturated with CO; unconsciousness and death are likely with 60% to 70% saturation.

MORPHOLOGY

Chronic poisoning by CO develops because carboxyhemoglobin, once formed, is remarkably stable. Even with low-level, but persistent, exposure to CO, carboxyhemoglobin may rise to life-threatening levels in the blood. The slowly developing hypoxia can insidiously evoke widespread ischemic changes in the CNS; these are particularly marked in the basal ganglia and lenticular nuclei. With cessation of exposure to CO, the patient usually recovers, but there may be permanent neurologic sequelae, such as impairment of memory, vision, hearing, and speech. The diagnosis is made by measuring carboxyhemoglobin levels in the blood.

Acute poisoning by CO is generally a consequence of accidental exposure or suicide attempt. In light-skinned individuals, **acute poisoning is marked by a characteristic generalized cherry-red color of the skin and mucous membranes,** which results from high levels of carboxyhemoglobin. This effect of CO on coloration may result in a failure to recognize the oxygen-starved state of the victim (and is used by the meat industry in the United States to keep meat appearing fresh—caveat emptor!). If death occurs rapidly, morphologic changes may not be present; with longer survival, the brain may be slightly edematous, with punctate hemorrhages and hypoxia-induced neuronal changes. The morphologic changes are not specific and stem from systemic hypoxia.

Indoor Air Pollution

As we increasingly "button up" our homes to exclude the environment, the potential for pollution of the indoor air increases. The most common pollutant is tobacco smoke (discussed later), but additional offenders are CO, nitrogen dioxide (both already mentioned as outdoor pollutants), and asbestos (Chapter 15). Volatile substances containing polycyclic aromatic hydrocarbons generated by cooking oils and coal burning are important indoor pollutants in lower income parts of the world, particularly parts of Asia. Only a few comments about other agents are made here.

- *Smoke from burning of organic materials,* containing various oxides of nitrogen and carbon particulates, is an irritant that predisposes exposed persons to lung infections and may contain carcinogenic polycyclic hydrocarbons. It is estimated that one-third of the world's people burn carbon-containing material such as wood, dung, or charcoal in their homes for cooking, heating, and light.
- *Bioaerosols* range from microbiologic agents capable of causing infectious diseases such as legionnaires' disease, viral pneumonia, and the common cold to less threatening but nonetheless distressing allergens derived from pet dander, dust mites, fungi, and molds that are variously responsible for rhinitis, eye irritation, and asthma.
- *Radon,* a radioactive gas derived from uranium widely present in soil and in homes, can cause lung cancer in uranium miners. It is also suspected that low-level chronic exposures in the home increase lung cancer risk, particularly in those who smoke tobacco. Radon is the number one cause of lung cancer among nonsmokers, according to EPA estimates. Overall, radon is the second leading cause of lung cancer. Radon is responsible for about 21,000 lung cancer deaths every year. About 2,900 of these deaths occur among people who have never smoked.
- *Formaldehyde* is used in the manufacture of building materials (e.g., cabinetry, furniture, adhesives) and may accumulate in the air in poorly ventilated housing. At concentrations of 0.1 ppm or higher, it causes breathing difficulties and a burning sensation in the eyes and throat and can trigger asthma attacks. Formaldehyde is classified as a carcinogen for humans and animals.
- The so-called *sick building syndrome* remains an elusive problem; it may be a consequence of exposure to one or more indoor pollutants, possibly due to poor ventilation.

KEY CONCEPTS

ENVIRONMENTAL DISEASES AND ENVIRONMENTAL POLLUTION

- Environmental diseases are conditions caused by exposure to chemical or physical agents in the ambient, workplace, and personal environments.
- Exogenous chemicals known as xenobiotics enter the body through inhalation, ingestion, and skin contact and can either be eliminated or accumulate in fat, bone, brain, and other tissues.
- Xenobiotics can be converted into nontoxic products or activated to generate toxic compounds through a two-phase reaction process that involves the cytochrome P-450 system.
- The most common and important air pollutants are ozone (which in combination with oxides and particulate matter forms smog), sulfur dioxide, acid aerosols, and particles less than 10 μm in diameter.
- CO poisoning is an important cause of death from accidents and suicide; it binds hemoglobin with high affinity, leading to systemic asphyxiation associated with CNS depression.
- A variety of pollutants including smoke, bioaerosols, radon, and formaldehyde may accumulate in indoor air and cause disease.

Metals as Environmental Pollutants

Lead, mercury, arsenic, and cadmium are the heavy metals most commonly associated with harmful effects in humans.

Lead

Lead is a readily absorbed metal that binds to sulfhydryl groups in proteins and interferes with calcium metabolism, effects that lead to hematologic, skeletal, neurologic, gastrointestinal, and renal toxicities. Lead exposure may occur through contaminated air, food, and water. For most of the 20th century the major sources of lead in the environment were lead-containing house paints and gasoline. Although limits have been set for the amounts of lead contained in residential paints and use of leaded gasoline in road vehicles was banned in the United States in 1996, lead contamination remains an important health hazard, particularly for children. There are many sources of lead in the environment such as from mining, foundries, batteries, and spray painting that constitute occupational hazards. However, flaking lead paint in older houses and soil contamination pose major hazards to youngsters. Blood levels of lead in children living in older homes containing lead-based paint or lead-contaminated dust often exceed 5 µg/dL, the level at which the Centers for Disease Control and Prevention (CDC) recommends that measures be taken to limit further exposure. A dramatic case of lead contamination of drinking water occurred in Flint, Michigan, in 2014–2016. The Flint water crisis occurred when the source of water supply to the city was changed from Lake Huron to the Flint River. Because water from the Flint River had a higher chloride concentration than the lake waters, it leached lead from century-old lead pipes. This caused an increase in lead levels in tap water above the acceptable limit of 15 parts per billion (ppb) in about 25% of the homes and in some cases as high as 13,200 ppb. As a result, 6000 to 12,000 residents developed very high lead levels in their blood. Ingested lead is particularly harmful to children because they absorb more than 50% of lead from food, whereas adults absorb approximately 15%. A more permeable blood–brain barrier in children creates a high susceptibility to brain damage. The main clinical features of lead poisoning in children and adults are shown in Figs. 9.4 and 9.5.

Most absorbed lead (80% to 85%) is taken up into developing teeth and into bone, where it competes with calcium, binds phosphates, and has a half-life of 20 to 30 years. About 5% to 10% of the absorbed lead remains in the blood, and the remainder is distributed throughout soft tissues. Excess lead is toxic to nervous tissues in adults and children; peripheral neuropathies predominate in adults, whereas central effects are more common in children. The effects of chronic lead exposure in children may be subtle, producing mild dysfunction, including reduced IQ, learning disabilities, and delayed psychomotor development. At higher doses, however, the results can be devastating, taking the form of blindness, psychoses, seizures, coma, and even death. Lead-induced peripheral neuropathies in adults generally remit with the elimination of exposure, but both peripheral nervous system and CNS abnormalities in children usually are irreversible.

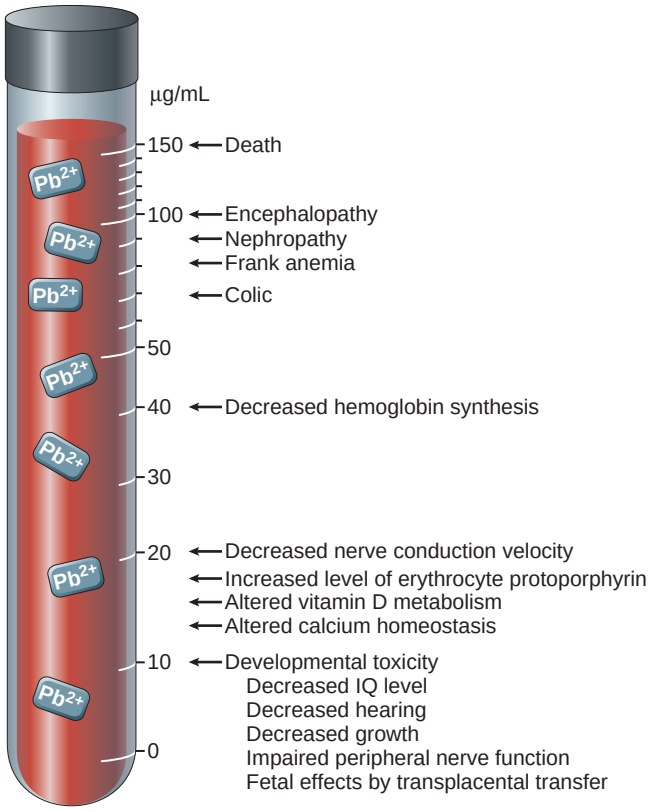


Figure 9.4 Effects of lead poisoning in children related to blood levels. (Modified from Bellinger DC, Bellinger AM: Childhood lead poisoning: the tortuous path from science to policy, *J Clin Invest* 116:853; 2006.)

Lead also affects many other tissues. Excess lead interferes with the normal remodeling of calcified cartilage and primary bone trabeculae in the epiphyses in children, causing increased bone density detected as radiodense lead lines (Fig. 9.6). Lead lines of a different sort also may occur in the gums, where excess lead stimulates hyperpigmentation. Lead inhibits the healing of fractures by increasing chondrogenesis and delaying cartilage mineralization. Excretion of lead occurs by way of the kidneys, and acute exposures may cause damage to proximal tubules.

Lead has a high affinity for sulfhydryl groups and interferes with two enzymes involved in heme synthesis: delta-aminolevulinic acid dehydratase and ferrochelatase. Iron incorporation into heme is impaired, leading to *microcytosis* (small red cells) and anemia. Lead also inhibits sodium- and potassium-dependent adenosine triphosphatases (ATPases) in cell membranes, an effect that may increase the fragility of red cells, causing hemolysis. The diagnosis of lead poisoning requires constant vigilance. It may be suspected on the basis of neurologic changes in children or unexplained microcytic anemia with basophilic stippling in red cells in adults and children. Elevated blood lead and red cell free protoporphyrin levels (greater than 50 µg/dL) or, alternatively, zinc protoporphyrin levels, are required for definitive diagnosis. In milder cases of lead exposure, anemia may be the only abnormality.

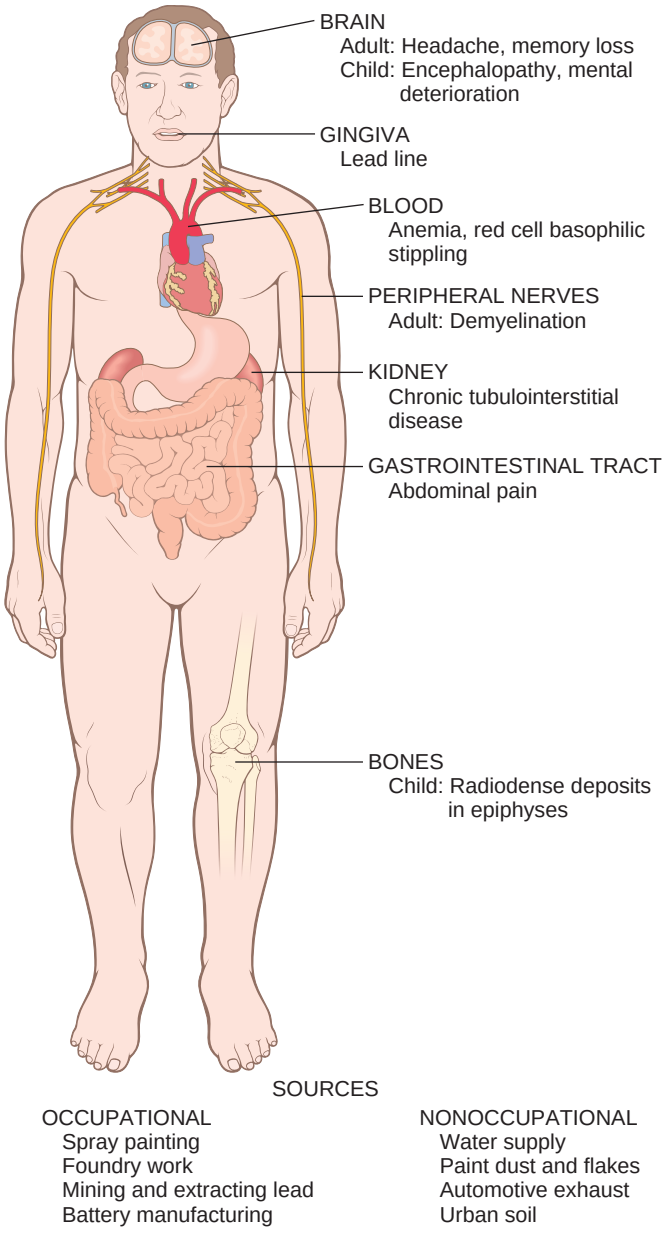


Figure 9.5 Pathologic features of lead poisoning in adults.

MORPHOLOGY

The major anatomic targets of lead toxicity are the bone marrow and blood, nervous system, gastrointestinal tract, and kidneys (see Fig. 9.5).

Blood and marrow changes occur fairly rapidly and are characteristic. The inhibition of ferrochelatase by lead may result in the appearance of a few **ring sideroblasts,** red cell precursors with iron-laden mitochondria that are detected with a Prussian blue stain. In the peripheral blood the defect in hemoglobin synthesis appears as a **microcytic, hypochromic anemia** that is often accompanied by mild **hemolysis.** Even more distinctive is a **punctate basophilic stippling of the red cells.**

Brain damage is prone to occur in children. In young children, sensory, motor, intellectual, and psychologic have been described. Lead toxicity in a pregnant woman may impair brain development in the fetus. The anatomic changes underlying the subtle functional deficits are ill-defined, but there is concern that these defects may be permanent. At the more severe end of the spectrum there is marked brain edema, demyelination of the cerebral and cerebellar white matter, and necrosis of cortical neurons accompanied by diffuse astrocytic proliferation. In adults the CNS is less often affected, but frequently **a peripheral demyelinating neuropathy** appears, typically involving the motor nerves of the most commonly used muscles. Thus the extensor muscles of the wrist and fingers are often the first to be affected (causing wristdrop), followed by paralysis of the peroneal muscles (causing footdrop).

The **gastrointestinal tract** is also a major source of clinical manifestations. Lead colic is characterized by severe, poorly localized abdominal pain.

Kidneys may develop proximal tubular damage associated with intranuclear inclusions consisting of protein aggregates. Chronic renal damage leads eventually to interstitial fibrosis and renal failure. Decreases in uric acid excretion can lead to gout (saturnine gout).

Mercury

Like lead, mercury binds to sulfhydryl groups in certain proteins with high affinity, leading to damage in the CNS and the kidney. Mercury has had many uses throughout history, for example, as a pigment in cave paintings, a cosmetic, a remedy for syphilis, and a component of diuretics. Alchemists tried (without much success) to produce gold from mercury.

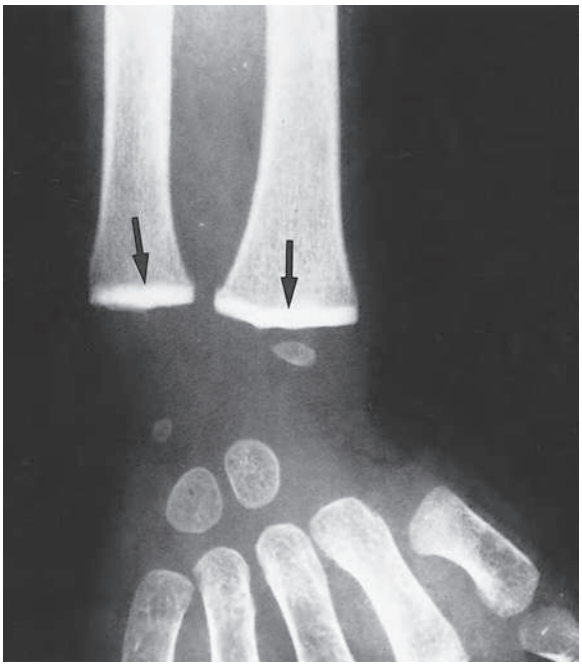


Figure 9.6 Lead poisoning. Impaired remodeling of calcified cartilage in the epiphyses *(arrows)* of the wrist has caused a marked increase in their radiodensity so that they are as radiopaque as the cortical bone. (Courtesy Dr. G. W. Dietz, Department of Radiology, University of Texas Southwestern Medical School, Dallas, Tex.)

Poisoning from inhalation of mercury vapors has long been recognized and is associated with tremor, gingivitis, and bizarre behavior, such as that displayed by the Mad Hatter in *Alice in Wonderland*. There are three forms of mercury: metallic mercury (also referred to as elemental mercury), inorganic mercury compounds (mostly mercuric chloride), and organic mercury (mostly methyl mercury). Today, the main sources of exposure to mercury are contaminated fish (methyl mercury) and mercury vapors released from metallic mercury in dental amalgams, a possible occupational hazard for dental workers. In some areas of the world, mercury used in gold mining has contaminated rivers and streams.

Inorganic mercury from the natural degassing of the earth's crust or from industrial contamination is converted to organic compounds such as methyl mercury by bacteria. Methyl mercury enters the food chain, and in carnivorous fish such as swordfish, shark, and bluefish, mercury levels may be 1 million times higher than in the surrounding water. Almost 90% of ingested methyl mercury is absorbed in the gastrointestinal tract. The consumption of fish contaminated by the release of methyl mercury from industrial sources in Minamata Bay and the Agano River in Japan caused widespread mortality and morbidity. Acute exposure through consumption of bread made from grain treated with a methyl mercury–based fungicide in Iraq in 1971 resulted in hundreds of deaths and thousands of hospitalizations. The medical disorders associated with the Minamata episode became known as *Minamata disease* and include cerebral palsy, deafness, blindness, intellectual disability, and major CNS defects in children exposed in utero. The lipid solubility of methyl mercury and metallic mercury facilitates their accumulation in the brain, disturbing neuromotor, cognitive, and behavioral functions. Intracellular GSH, by acting as a sulfhydryl donor, is the main protective mechanism against mercury-induced CNS and kidney damage.

Mercury continues to be released into the environment by power plants and other industrial sources, and there are serious concerns about the effects of chronic low-level exposure to methyl mercury in the food supply. To protect against potential fetal brain damage, the CDC has recommended that pregnant women avoid consumption of fish known to contain high levels of mercury. Ingested mercury can injure the gut and cause ulcerations and bloody diarrhea. In the kidneys, mercury can cause acute tubular necrosis and renal failure. Chronic exposure can cause nephrotic syndrome.

Arsenic

Arsenic salts interfere with several aspects of cellular metabolism, leading to toxicities that are most prominent in the gastrointestinal tract, nervous system, skin, and heart. Arsenic was the poison of choice in Renaissance Italy, with members of the Borgia and Medici families being highly skilled practitioners of the art of its use. Because of its favored use as an instrument of assassination among royal families, arsenic has been called "the poison of kings and the king of poisons." Deliberate poisoning is rare today, but unintentional exposure to arsenic is an important health problem in many areas of the world. Arsenic is found naturally in soils and water and is used in products such as wood preservers and herbicides and other agricultural products. It may be released into the environment from mines and smelting industries. Arsenic is present in Chinese and Indian herbal medicine, and high concentrations of inorganic arsenic are present in groundwater in countries such as Bangladesh, Chile, and China. It is estimated that 40 million people in Bangladesh drink water contaminated with arsenic, constituting one of the greatest environmental cancer risks yet uncovered.

The most toxic forms of arsenic are the trivalent compounds arsenic trioxide, sodium arsenite, and arsenic trichloride. If ingested in large quantities, arsenic causes acute gastrointestinal, cardiovascular, and CNS toxicities that are often fatal. These effects may be attributed in part to interference with mitochondrial oxidative phosphorylation, since trivalent arsenic can replace the phosphates in ATP. However, arsenic also has pleiotropic effects on the activity of a number of other enzymes and ion channels, and these too may contribute to certain toxicities.

- *Neurologic effects* usually occur 2 to 8 weeks after exposure and consist of a sensorimotor neuropathy that causes paresthesias, numbness, and pain.
- *Cardiovascular effects* include hypertension and prolonged Q-Tc interval with ventricular arrhythmias.
- *Skin changes* consisting of hyperpigmentation and hyperkeratosis occur with chronic exposure.
- *Increased risk for the development of cancers is* the most serious consequence of chronic exposure, particularly of the lung, bladder, and skin. Arsenic-induced skin tumors differ from those induced by sunlight; they are often multiple and usually appear on the palms and soles. The mechanisms of arsenic carcinogenesis in skin and lung have not been elucidated but may involve defects in nucleotide excision repair mechanisms that protect against DNA damage.

Cadmium

Cadmium is preferentially toxic to the kidneys and the lungs through uncertain mechanisms that may involve increased production of ROS. In contrast to the other metals discussed in this section, cadmium toxicity is a relatively modern problem. It is an occupational and environmental pollutant generated by mining, electroplating, and production of nickel-cadmium batteries, which are usually disposed of as household waste. Cadmium can contaminate the soil and plants directly or through fertilizers and irrigation water. Food is the most important source of cadmium exposure for the general population. Its toxic effects require its uptake into cells via transporters such as ZIP8, which normally serves as a transporter for zinc.

The principal toxic effects of excess cadmium are a form of obstructive lung disease caused by necrosis of alveolar epithelial cells and renal tubular damage that may progress to end-stage renal disease. Cadmium exposure can also cause skeletal abnormalities associated with calcium loss. Cadmium-containing water used to irrigate rice fields in Japan caused a disease in postmenopausal women known as *itai-itai* ("ouch-ouch"), a combination of osteoporosis and osteomalacia associated with renal disease. Cadmium exposure is also associated with an elevated risk of lung cancer, which has been demonstrated in workers exposed occupationally and in populations living near zinc

smelters. Cadmium is not directly genotoxic and most likely produces DNA damage through the generation of ROS (Chapter 2).

KEY CONCEPTS

TOXIC EFFECTS OF HEAVY METALS

- Lead, mercury, arsenic, and cadmium are the heavy metals most commonly associated with toxic effects in humans.
- Children absorb more ingested lead than adults; the main source of exposure for children is lead-containing paint in older housing and lead-containing drinking water.
- Excess lead causes CNS defects in children and peripheral neuropathy in adults. It also interferes with the remodeling of cartilage and causes anemia by interfering with hemoglobin synthesis.
- The major source of exposure to mercury is contaminated fish. The developing brain is highly sensitive to methyl mercury, which accumulates in the CNS.
- Exposure of the fetus to high levels of mercury in utero may lead to Minamata disease, characterized by cerebral palsy, deafness, and blindness.
- Arsenic is naturally found in soil and water and is a component of some wood preservatives and herbicides. Excess arsenic interferes with mitochondrial oxidative phosphorylation and the function of a variety of proteins. It causes toxic effects in the gastrointestinal tract, CNS, and cardiovascular system; long-term exposure causes skin lesions and carcinomas.
- Cadmium from nickel-cadmium batteries and chemical fertilizers can contaminate soil. Excess cadmium causes obstructive lung disease and kidney damage.

OCCUPATIONAL HEALTH RISKS: INDUSTRIAL AND AGRICULTURAL EXPOSURES

More than 10 million occupational injuries occur annually in the United States, and approximately 65,000 people die as a consequence of occupational injuries and illnesses. About 10% of these deaths occur due to falls and other traumatic injuries, while the remainder are secondary to toxic exposures that lead to cancer, lung disease, and other potentially fatal conditions. Industrial exposures to toxic agents are as varied as the industries themselves. They range from merely annoying irritations of respiratory airways by formaldehyde or ammonia fumes to lung cancers arising from exposure to asbestos, arsenic, or uranium. Human diseases associated with occupational exposures are listed in Table 9.2. In addition to toxic metals (already discussed), other important agents that contribute to environmental diseases include the following:

- *Organic solvents* are widely used in huge quantities worldwide. Some, such as *chloroform* and *carbon tetrachloride,* are found in degreasing and dry cleaning agents and paint removers. Acute exposure to high levels of these agents can cause dizziness and confusion, CNS depression, and even coma. Lower levels may cause liver and kidney toxicity. Occupational exposure to *benzene* and *1,3-butadiene* increases the risk of leukemia. Benzene is oxidized to an epoxide through hepatic CYP2E1, a component of the P-450 enzyme system already mentioned. The epoxide and other metabolites disrupt progenitor cell differentiation

Table 9.2 Human Diseases Associated With Occupational Exposures

Organ/System	Effect	Toxicant
Cardiovascular system	Heart disease	Carbon monoxide, lead, solvents, cobalt, cadmium
Respiratory system	Nasal cancer	Isopropyl alcohol, wood dust
	Lung cancer	Radon, asbestos, silica, bis(chloromethyl) ether, nickel, arsenic, chromium, mustard gas, uranium
	Chronic obstructive pulmonary disease	Grain dust, coal dust, cadmium
	Hypersensitivity	Beryllium, isocyanates
	Irritation	Ammonia, sulfur oxides, formaldehyde
	Fibrosis	Silica, asbestos, cobalt
Nervous system	Peripheral neuropathies	Solvents, acrylamide, methyl chloride, mercury, lead, arsenic, DDT
	Ataxic gait	Chlordane, toluene, acrylamide, mercury
	CNS depression	Alcohols, ketones, aldehydes, solvents
	Cataracts	Ultraviolet radiation
Urinary system	Renal toxicity	Mercury, lead, glycol ethers, solvents
	Bladder cancer	Naphthylamines, 4-aminobiphenyl, benzidine, rubber products
Reproductive system	Male infertility	Lead, phthalate plasticizers, cadmium
	Female infertility	Lead, mercury
	Stillbirths	Lead, mercury
	Teratogenesis	Mercury, polychlorinated biphenyls
Hematopoietic system	Leukemia	Benzene
Skin	Folliculitis and acneiform dermatosis	Polychlorinated biphenyls, dioxins, herbicides
	Cancer	Ultraviolet radiation
Gastrointestinal tract	Liver angiosarcoma	Vinyl chloride

CNS, Central nervous system; *DDT,* dichlorodiphenyltrichloroethane.

Data from Leigh JP, et al: Occupational injury and illness in the United States. Estimates of costs, morbidity, and mortality, *Arch Intern Med* 157:1557, 1997; Mitchell FL: Hazardous waste. In Rom WN, editor: *Environmental and Occupational Medicine,* ed 2, Boston, 1992, Little, Brown, p 1275; and Levi PE: Classes of toxic chemicals. In Hodgson E, Levi PE, editors: *A Textbook of Modern Toxicology,* Stamford, Conn, 1997, Appleton & Lange, p 229.

in the bone marrow and may lead to marrow aplasia and acute myeloid leukemia.

- *Polycyclic hydrocarbons* are released during the combustion of coal and gas, particularly at the high temperatures used in steel foundries, and also are present in tar and soot. (Pott identified soot as the cause of scrotal cancers in chimney sweeps in 1775, as mentioned in Chapter 7). Polycyclic hydrocarbons are among the most potent carcinogens, and industrial exposures have been implicated in the causation of lung and bladder cancer.
- *Organochlorines* (and halogenated organic compounds in general) are synthetic products that resist degradation and are lipophilic. Organochlorines used as pesticides are *dichlorodiphenyltrichloroethane (DDT)* and its metabolites and agents such as lindane, aldrin, and dieldrin. Nonpesticide organochlorines include *polychlorinated biphenyls (PCBs)* and *dioxin* (2,3,7,8-tetrachlorodibenzo-*p*-dioxin [TCDD]). DDT was banned in the United States in 1973. Acute DDT poisoning in humans causes neurologic toxicity. Most organochlorines are endocrine disruptors and have antiestrogenic or antiandrogenic activity in laboratory animals, but long-term health effects in humans have not been firmly established.
- *Nonpesticide organochlorines* include *PCBs* and *dioxin* (TCDD). Dioxins and PCBs can cause skin disorders such as folliculitis and acneiform dermatosis known as chloracne, which consists of acne, cyst formation, hyperpigmentation, and hyperkeratosis, generally around the face and behind the ears. It can be accompanied by abnormalities in the liver and CNS. Because PCBs induce the P-450 enzyme system, workers exposed to these substances may show altered drug metabolism. Environmental disasters in Japan and China in the late 1960s caused by the consumption of rice oil contaminated by PCBs poisoned about 2000 people in each episode. The primary manifestations of the disease (yusho in Japan, yu-cheng in China) were chloracne and hyperpigmentation of the skin and nails.
- *Bisphenol A* (BPA) is used in the synthesis of polycarbonate food and water containers and of epoxy resins that line almost all food bottles and cans; as a result, exposure to BPA is virtually ubiquitous in humans. BPA has long been known as a potential endocrine disruptor. Several large retrospective studies have linked elevated urinary BPA levels to heart disease in adult populations. In addition, infants who drink from BPA-containing containers may be particularly susceptible to the endocrine effects of BPA. In 2010, Canada was the first country to list BPA as a toxic substance, and the largest makers of baby bottles and "sippy" cups have stopped using BPA in the manufacturing process. The extent of the human health risks associated with BPA remains uncertain.
- *Vinyl chloride,* used in the synthesis of polyvinyl resins, can cause angiosarcoma of the liver, a rare type of liver tumor.
- Inhalation of *mineral dusts* causes chronic, nonneoplastic lung diseases called *pneumoconioses.* This group of disorders includes diseases induced by organic and inorganic particulates as well as chemical fume- and vapor-induced nonneoplastic lung diseases. The most common pneumoconioses are caused by exposures to coal dust (in mining of hard coal), silica (in sandblasting and stone cutting), asbestos (in mining, fabrication, and insulation work), and beryllium (in mining and fabrication). Exposure to these agents nearly always occurs in the workplace. The increased risk of cancer as a result of asbestos exposure, however, extends to family members of asbestos workers and to other persons exposed outside the workplace. Pneumoconioses and their pathogenesis are discussed in Chapter 13.

Effects of Tobacco

Smoking is the most readily preventable cause of death in humans. The main culprit is cigarette smoking, but smokeless tobacco (e.g., snuff, chewing tobacco) is also harmful to health and an important cause of oral cancer. The use of tobacco products not only creates personal risks, but passive tobacco inhalation from the environment *(second-hand smoke)* can cause lung cancer in nonsmokers. More than 20 million US residents have died of smoking-related diseases since the 1964 Surgeon General's report on the adverse effects of smoking. Of these, almost 2.5 million died as a result of inhalation of second-hand smoke. More than 10 times as many people in the United States have died as a result of cigarette smoking than have died in all the wars fought by the United States in its entire history. Annually, tobacco is responsible for more than 400,000 deaths in the United States, one-third of these due to lung cancer. Indeed, tobacco is the leading exogenous cause of human cancers, including 90% of lung cancers.

Worldwide, two-thirds of smokers live in 10 countries, led by China, which accounts for nearly 30%, and India with about 10%, followed by Indonesia, Russia, the United States, Japan, Brazil, Bangladesh, Germany, and Turkey.

From 1998 to 2007 in the United States, the incidence of smoking declined modestly, but this trend failed to continue, and approximately 20% of adults remain smokers. More disturbing, the world's most populous country, China, has become the world's largest producer and consumer of cigarettes. China has approximately 350 million smokers who in aggregate consume about 33% of all cigarettes smoked worldwide. It is estimated that more than 1 million people in China die each year of smoking-related diseases; this rate is projected to rise to 8 million deaths each year by 2050. Worldwide, cigarette smoking causes more than 4 million deaths annually, mostly from cardiovascular disease, various types of cancers, and chronic respiratory problems. These figures are expected to rise to 8 million tobacco-related deaths by 2020, the major increase occurring in lower income countries. Of people alive today, an estimated 500 million will die of tobacco-related illnesses.

Tobacco reduces overall survival through dose-dependent effects that are often expressed as pack-years, the average number of cigarette packs smoked each day multiplied by the number of years of smoking. The cumulative effects of smoking over time are striking. For instance, while about 75% of nonsmokers are alive at age 70, only about 50% of smokers survive to that age (Fig. 9.7). The only good news is that cessation of smoking greatly reduces, within 5 years, overall mortality and the risk of death from cardiovascular diseases. Lung cancer mortality decreases by 21% within 5 years, but the excess risk persists for 30 years.

The number of potentially noxious chemicals in tobacco smoke is extraordinary. Tobacco smoke contains a complex

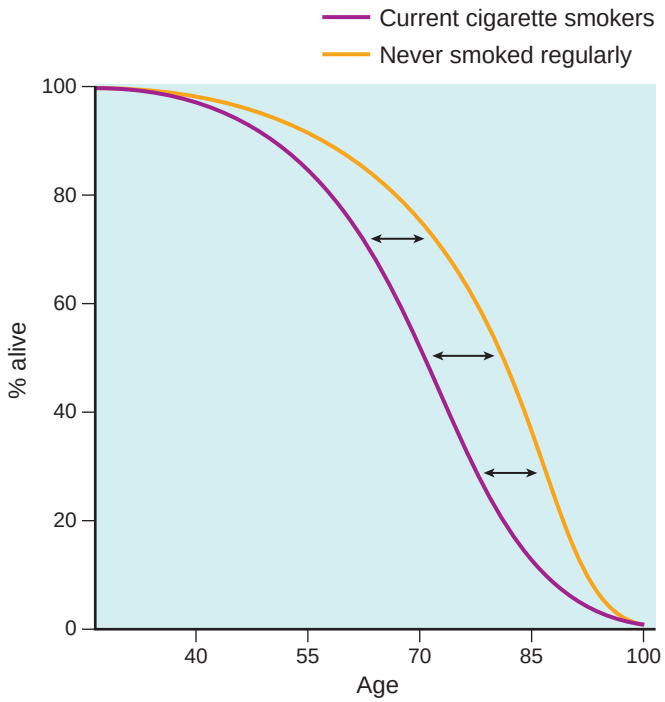


Figure 9.7 Effects of smoking on survival. The study compared age-specific death rates for current cigarette smokers with death rates of individuals who never smoked regularly (British Doctors Study). Measured at age 75, the difference in survival between smokers and nonsmokers is 7.5 years. (Modified from Stewart BW, Kleihues P, editors: *World Cancer Report,* Lyon, 2003, IARC Press.)

mixture of 7000 chemicals, more than 60 of which have been identified as carcinogens. Table 9.3 provides only a partial list and includes various types of injuries produced by these agents.

Nicotine, an alkaloid present in tobacco leaves, is strongly addictive. Nicotine binds to nicotinic acetylcholine receptors in the brain, and stimulates the release of catecholamines from sympathetic neurons. This activity is responsible for the acute effects of smoking, such as the increase in heart rate and blood pressure and the elevation in cardiac contractility and output. Recent studies indicate that in addition to being addictive, nicotine has other untoward effects. These include effects on the fetus, as nicotine, exposure affects fetal brain development and contributes to preterm birth and still birth.

Table 9.3 Effects of Selected Tobacco Smoke Constituents

Substance	Effect
Tar	Carcinogenesis
Polycyclic aromatic hydrocarbons	Carcinogenesis
Nicotine	Ganglionic stimulation and depression; tumor promotion
Phenol	Tumor promotion; mucosal irritation
Benzo[*a*]pyrene	Carcinogenesis
Carbon monoxide	Impaired oxygen transport and utilization
Formaldehyde	Toxicity to cilia; mucosal irritation
Nitrogen oxides	Toxicity to cilia; mucosal irritation
Nitrosamine	Carcinogenesis

Table 9.4 Suspected Organ-Specific Carcinogens in Tobacco Smoke

Organ	Carcinogen
Lung, larynx	Polycyclic aromatic hydrocarbons 4-(Methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK) Polonium 210
Esophagus	*N*′-Nitrosonornicotine (NNN)
Pancreas	NNK
Bladder	4-Aminobiphenyl, 2-naphthylamine
Oral cavity (smoking)	Polycyclic aromatic hydrocarbons, NNK, NNN
Oral cavity (snuff)	NNK, NNN, polonium 210

Data from Szczesny LB, Holbrook JH: Cigarette smoking. In Rom WH, editor: *Environmental and Occupational Medicine,* ed 2, Boston, 1992, Little, Brown, p 1211.

Smoking and Lung Cancer. Agents in smoke have a direct irritant effect on the tracheobronchial mucosa, producing increased mucus production (bronchitis). Components of cigarette smoke, particularly polycyclic hydrocarbons and nitrosamines (Table 9.4), are potent carcinogens in animals and are directly involved in the development of lung cancer in humans. CYPs (cytochrome P-450 phase I enzymes) and phase II enzymes increase the water solubility of the carcinogens, facilitating their excretion. However, some intermediates produced by CYPs are electrophilic and form DNA adducts that are repaired by error-prone mechanisms, leading to potentially oncogenic mutations (Chapter 7). Most tellingly, deep sequencing of the genomes of lung cancers that occur in smokers has revealed the presence of thousands of mutations of a type that is produced by carcinogens in tobacco smoke in experimental settings. The risk of developing lung cancer is related to the number of pack-years or cigarettes smoked per day (Fig. 9.8). Moreover, smoking increases the risk of other carcinogenic influences. Witness the 10-fold higher incidence of lung carcinomas in asbestos workers and uranium miners who smoke over those who do not smoke and the interaction between tobacco consumption and alcohol in the development of oral and laryngeal cancers (Fig. 9.9).

Smoking and Other Diseases. In addition to lung cancer, smoking is linked to many other malignant and nonmalignant disorders that affect numerous organ systems (Fig. 9.10).

- Cigarette smoking is associated with *cancers of the esophagus, pancreas, bladder, kidney, cervix,* and *bone marrow*. To this list, the latest report of the US Surgeon General has added *carcinomas of the liver and colon;* the latter is the second most common cause of cancer deaths.
- *The toll taken by nonmalignant conditions* associated with smoking is even more terrible. Agents in smoke have a direct irritant effect on the tracheobronchial mucosa, producing inflammation and increased mucus production (bronchitis). Cigarette smoke also causes the recruitment of leukocytes to the lung, with increased local elastase production and subsequent injury to lung tissue, leading

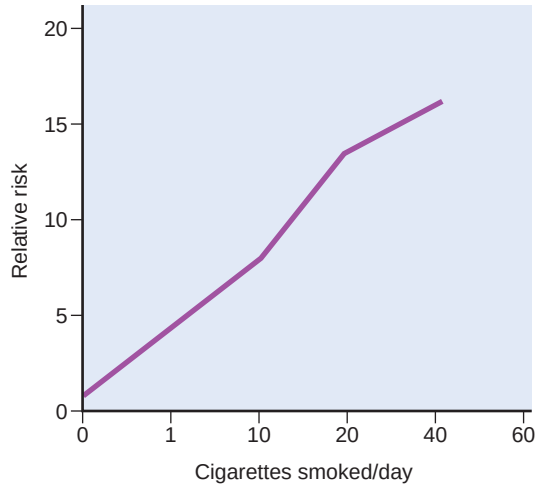


Figure 9.8 The risk of lung cancer is determined by the number of cigarettes smoked. (Modified from Stewart BW, Kleihues P, editors: *World Cancer Report,* Lyon, 2003, IARC Press.)

to emphysema, chronic bronchitis, and chronic obstructive pulmonary disease, conditions that are discussed in Chapter 15. Smoking exacerbates asthma and increases the risk for pulmonary tuberculosis.

- *Cigarette smoking is strongly linked to the development of atherosclerosis* and its major complications, myocardial infarction and stroke. The causal mechanisms probably relate to several factors including increased platelet aggregation, decreased myocardial oxygen supply

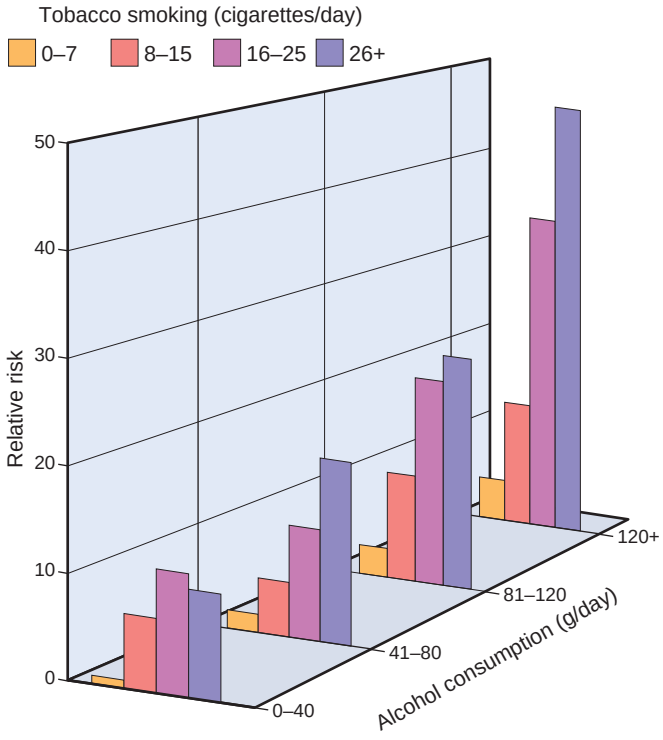


Figure 9.9 Multiplicative increase in the risk of laryngeal cancer from the interaction between cigarette smoking and alcohol consumption. (Modified from Stewart BW, Kleihues P, editors: *World Cancer Report,* Lyon, 2003, IARC Press.)

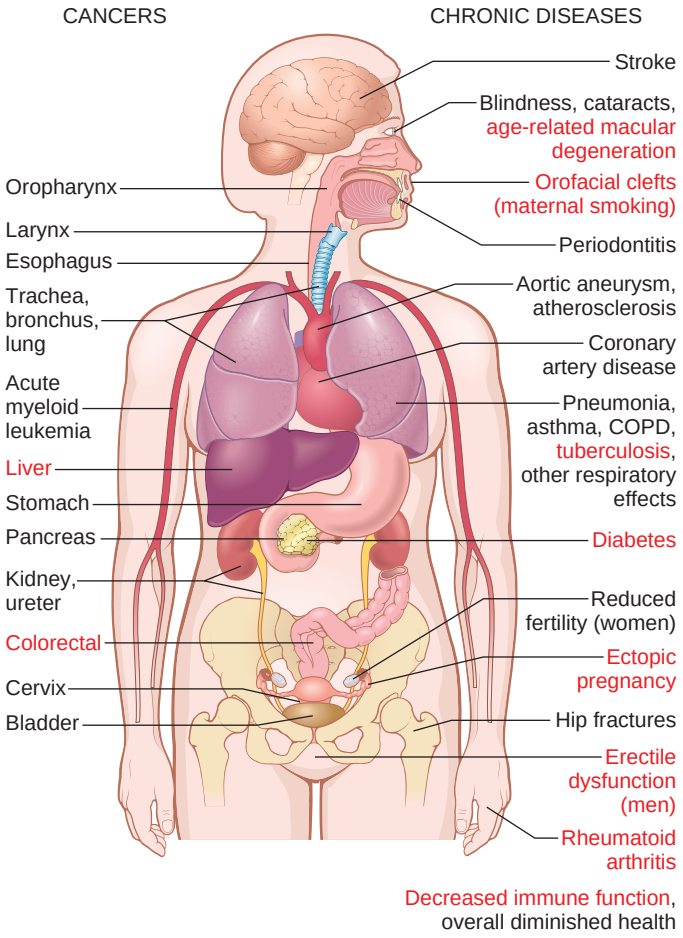


Figure 9.10 Health consequences causally linked to smoking. Items in red are new diseases added in the 2016 report from the Surgeon General (US Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, Ga, 2016, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health). *COPD,* Chronic obstructive pulmonary disease.

(because of significant lung disease coupled with the hypoxia related to the CO content of cigarette smoke) accompanied by increased myocardial oxygen demand and a decreased threshold for ventricular fibrillation. Smoking has a multiplicative effect on the incidence of myocardial infarction when combined with hypertension and hypercholesterolemia.

- *Smoking also harms the developing fetus.* Maternal smoking increases the risk of spontaneous abortions and preterm births and results in *intrauterine growth retardation* (Chapter 10). Birth weights of infants born to mothers who stopped smoking before pregnancy are, however, normal.
- The latest US Surgeon General's report has added several new diseases to the previously known list of smoking-associated diseases (Fig. 9.10). *Smoking increases the risk of type 2 diabetes, rheumatoid arthritis, age-related macular degeneration, ectopic pregnancy, and erectile dysfunction.*
- Exposure to environmental tobacco smoke *(passive smoke inhalation)* is associated with some of the same detrimental

effects that result from active smoking. It is estimated that the relative risk of lung cancer in nonsmokers exposed to environmental smoke is about 1.3 times higher than that of nonsmokers who are not exposed to smoke. In the United States, approximately 3000 lung cancer deaths in nonsmokers older than 35 years can be attributed each year to environmental tobacco smoke. Even more striking is the increased risk of coronary atherosclerosis and fatal myocardial infarction. Studies report that every year 30,000 to 60,000 cardiac deaths in the United States are associated with exposure to passive smoke. Children living in a household with an adult who smokes have an increased frequency of respiratory illnesses such as asthma. Passive smoke inhalation in nonsmokers can be estimated by measuring the blood levels of cotinine, a metabolite of nicotine. In the United States, median cotinine levels in nonsmokers have decreased by more than 60% since around 2000 because of the adoption of nonsmoking policies in public places. However, passive exposure to tobacco smoke in the home remains a major public health concern, particularly for children. It is clear that the transient pleasure of smoking comes with a heavy long-term price.

- *Electronic cigarettes* (e-cigarettes), devices that simulate cigarette smoking by delivering vaporized nicotine and flavorings, are rising in popularity. The use of flavored e-cigarettes, called "vaping," has been on the increase in recent years, especially among young adults. While for several years after the introduction of e-cigarettes no significant untoward effects were recorded, starting in the summer of 2019 an outbreak of vaping-associated acute lung injury occurred in the United States. By the end of 2019 close to 2000 cases had been reported to the CDC, with 42 fatalities. The pathogenesis of this outbreak is under intense investigations.

KEY CONCEPTS

HEALTH EFFECTS OF TOBACCO

- Smoking is the most prevalent preventable cause of human death.
- Tobacco smoke contains more than 7000 compounds. Among these are nicotine, which is responsible for tobacco addiction, and potent carcinogens—mainly, polycyclic aromatic hydrocarbons, nitrosamines, and aromatic amines. Nicotine also has other adverse effects, particularly on fetal development, and is associated with preterm birth and stillbirth.
- Approximately 90% of lung cancers occur in smokers. Smoking is also associated with an increased risk of cancers of the oral cavity, larynx, esophagus, stomach, bladder, and kidney, some forms of leukemia, as well as liver and colorectal cancer. Cessation of smoking reduces the risk of lung cancer.
- Smokeless tobacco use is an important cause of oral cancers. Tobacco consumption interacts with alcohol in multiplying the risk of oral, laryngeal, and esophageal cancer and increases the risk of lung cancers from occupational exposures to asbestos, uranium, and other agents.
- Tobacco use is an important risk factor for development of atherosclerosis and myocardial infarction, peripheral vascular disease, and cerebrovascular disease. In the lungs, in addition to cancer, it predisposes to emphysema, chronic bronchitis, and chronic obstructive pulmonary disease.
- Maternal smoking increases the risk of spontaneous abortion, premature birth, and intrauterine growth retardation.

EFFECTS OF ALCOHOL

Ethanol consumption in moderate amounts is generally not injurious (and may even protect against some disorders), but in excessive amounts alcohol causes serious physical and psychologic damage. In this section, we describe alcohol metabolism and the major health consequences associated with alcohol abuse.

Despite all the attention given to illicit drugs such as cocaine and opiates, alcohol abuse is a far more widespread hazard and claims many more lives. According to a 2017 survey conducted by the National Institute on Alcohol Abuse and Alcoholism, 60% of the people reported using alcohol in the previous month. Even more worrisome is the fact that 14 million of adults (over 18 years of age) suffer from alcohol abuse disorder (AUD) in the United States (5.7% of this age group). AUD is a chronic relapsing brain disease characterized by an impaired ability to stop or control alcohol use despite adverse social, occupational, or health consequences. It is estimated that alcohol consumption is responsible for more than 80,000 deaths annually. More than 50% of these deaths result from accidents caused by drunken driving and alcohol-related homicides and suicides, and about 15,000 annual deaths are a consequence of cirrhosis of the liver. Worldwide, alcohol accounts for approximately 3.3 million deaths per year (5.9% of all deaths).

After consumption, ethanol is absorbed unaltered in the stomach and small intestine and then distributes throughout the body in direct proportion to the blood level. Less than 10% is excreted unchanged in the urine, sweat, and breath. The amount exhaled is proportional to the blood level and forms the basis for the breath test used by law enforcement agencies. A concentration of 80 mg/dL in the blood constitutes the legal definition of drunk driving in most states. For an average individual, this alcohol concentration is reached after consumption of three standard drinks, about three (12 ounce) bottles of beer, 15 oz of wine, or 4 to 5 oz of 80-proof distilled spirits. Drowsiness occurs at 200 mg/dL, stupor at 300 mg/dL, and coma, with possible respiratory arrest, at higher levels. The rate of metabolism affects the blood alcohol level. Chronic alcoholics develop tolerance to alcohol. They metabolize alcohol at a higher rate than normal and hence show lower peak levels of alcohol than average for the same about of alcohol consumed. Most of the alcohol in the blood is metabolized to acetaldehyde in the liver by three enzyme systems: alcohol dehydrogenase, cytochrome P-450 isoenzymes, and catalase (Fig. 9.11). Of these, the main enzyme involved in alcohol metabolism is alcohol dehydrogenase, located in the cytosol of hepatocytes. At high blood alcohol levels, however, the microsomal ethanol-oxidizing system also plays an important role. This

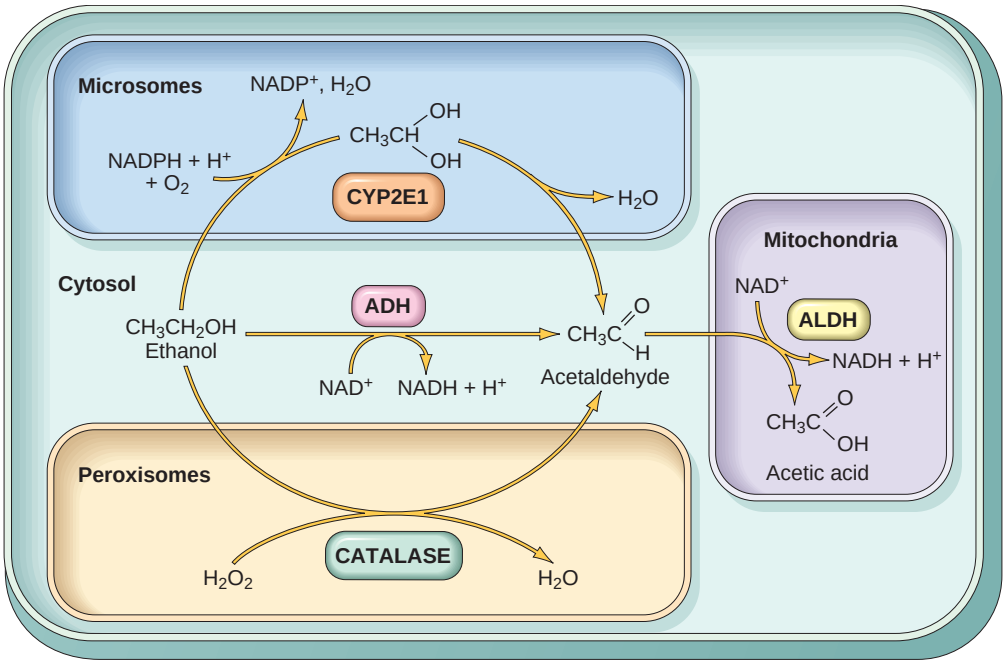


Figure 9.11 Metabolism of ethanol: oxidation of ethanol to acetaldehyde by three different routes and the generation of acetic acid. Note that oxidation by alcohol dehydrogenase *(ADH)* takes place in the cytosol; the cytochrome P-450 system and its CYP2E1 isoform are located in the endoplasmic reticulum (microsomes), and catalase is located in peroxisomes. Oxidation of acetaldehyde by aldehyde dehydrogenase *(ALDH)* occurs in mitochondria. ADH oxidation is the most important route; catalase is involved in only 5% of ethanol metabolism. Oxidation through CYPs may also generate reactive oxygen species (not shown). (From Parkinson A: Biotransformation of xenobiotics. In Klassen CD, editor: *Casarett and Doull's Toxicology: The Basic Science of Poisons*, ed 6, New York, 2001, McGraw-Hill, p 133.)

system involves cytochrome P-450 enzymes, particularly the CYP2E1 isoform, located in the smooth endoplasmic reticulum. Induction of P-450 enzymes by alcohol explains the increased susceptibility of alcoholics to other compounds metabolized by the same enzyme system, which include drugs (acetaminophen, cocaine), anesthetics, carcinogens, and industrial solvents. Of note, however, when alcohol is present in the blood at high concentrations, it competes with other CYP2E1 substrates and may delay the catabolism of other drugs, thereby potentiating their effects. Catalase is of minor importance, being responsible for only about 5% of alcohol metabolism. Acetaldehyde produced by these systems is in turn converted by acetaldehyde dehydrogenase to acetate, which is used in the mitochondrial respiratory chain or in lipid synthesis.

Several toxic effects result from ethanol metabolism. Listed here are only the most important of these.

- *Acetaldehyde,* the direct product of alcohol oxidation, has many toxic effects and is responsible for some of the acute effects of alcohol. The efficiency of alcohol metabolism varies among populations, depending on the expression levels of alcohol dehydrogenase and aldehyde dehydrogenase, as well as the presence of genetic variants that alter enzyme activity. About 50% of Asians have very low aldehyde dehydrogenase activity due to the substitution of lysine for glutamine at residue 487 (the normal allele is termed ALDH2*1, and the inactive variant is designated as ALDH2*2). The ALDH2*2 protein has dominant-negative activity, such that even one copy of the ALDH2*2 allele reduces enzyme activity significantly. Individuals homozygous for the ALDH2*2 allele are completely unable to oxidize acetaldehyde and cannot tolerate alcohol, experiencing nausea, flushing, tachycardia, and hyperventilation after its ingestion.
- *Alcohol oxidation* by alcohol dehydrogenase causes the reduction of nicotinamide adenine dinucleotide (NAD) to NADH, with a consequent decrease in NAD and increase in NADH. NAD is required for fatty acid oxidation in the liver and for the conversion of lactate into pyruvate. Its deficiency is a main cause of the accumulation of fat in the liver of alcoholics. The increase in the NADH/NAD ratio in alcoholics also causes lactic acidosis.
- *ROS generation*. Metabolism of ethanol in the liver by CYP2E1 produces ROS, which cause lipid peroxidation of hepatocyte membranes. Alcohol also provokes the release of endotoxin (lipopolysaccharide) from gram-negative bacteria in the intestinal flora, which stimulates the production of tumor necrosis factor (TNF) and other cytokines from macrophages and Kupffer cells, leading to hepatic injury.

The adverse effects of ethanol can be classified as acute or chronic.

***Acute alcoholism* exerts its effects mainly on the CNS, but it may induce hepatic and gastric changes that are reversible if alcohol consumption is discontinued.** Even with moderate intake of alcohol, multiple fat droplets accumulate in the cytoplasm of hepatocytes (fatty change or hepatic steatosis). The gastric changes are acute gastritis and ulceration. In the CNS, alcohol is a depressant, first affecting subcortical structures (probably the high brain stem

reticular formation) that modulate cerebral cortical activity. Consequently, there is stimulation and disordered cortical, motor, and intellectual behavior. At progressively higher blood levels, cortical neurons and then lower medullary centers are depressed, including those that regulate respiration. Respiratory arrest may follow.

***Chronic alcoholism* affects not only the liver and stomach, but virtually all other organs and tissues as well.** Chronic alcoholics suffer significant morbidity and have a shortened lifespan, related principally to damage to the liver, gastrointestinal tract, CNS, cardiovascular system, and pancreas.

- The *liver* is the main site of chronic injury. In addition to fatty change mentioned above, chronic alcoholism causes alcoholic hepatitis and cirrhosis, as described in Chapter 18. Cirrhosis is associated with portal hypertension and an increased risk for the development of hepatocellular carcinoma.
- In the *gastrointestinal tract,* chronic alcoholism can cause massive bleeding from gastritis, gastric ulcer, or esophageal varices (associated with cirrhosis), which may be fatal.
- *Neurologic effects.* Thiamine (vitamin B_1) deficiency is common in chronic alcoholics; the principal lesions resulting from this deficiency are peripheral neuropathies and Wernicke-Korsakoff syndrome (see Table 9.9 later and Chapter 28); cerebral atrophy, cerebellar degeneration, and optic neuropathy may also occur.
- *Cardiovascular effects*. Alcohol has diverse effects on the cardiovascular system. Injury to the myocardium may produce dilated congestive cardiomyopathy (alcoholic cardiomyopathy, discussed in Chapter 12). Chronic alcoholism is also associated with an increased incidence of hypertension, and heavy alcohol consumption, with attendant liver injury, results in decreased levels of high-density lipoprotein (HDL), increasing the likelihood of coronary heart disease.
- *Pancreatitis*. Excessive alcohol intake increases the risk of acute and chronic pancreatitis (Chapter 19).
- *Fetus.* The use of ethanol during pregnancy can cause fetal alcohol syndrome, which is marked by microcephaly, growth retardation, and facial abnormalities in the newborn and reduction in mental functions as the child grows older. It is difficult to establish the minimal amount of alcohol consumption that can cause fetal alcohol syndrome, but consumption during the first trimester of pregnancy is particularly harmful.
- *Carcinogenesis*. Chronic alcohol consumption is associated with an increased incidence of cancer of the oral cavity, esophagus, and liver. In women, low to moderate intake (12 oz beer or 5 oz of wine) incurs a slightly higher risk of breast cancer. Acetaldehyde is considered to be the main agent associated with alcohol-induced laryngeal and esophageal cancer, and acetaldehyde-DNA adducts have been detected in some tumors from these tissues. Individuals with one copy of the ALDH2*2 aldehyde dehydrogenase allele who drink are at a higher risk for developing cancer of the esophagus. As mentioned earlier, alcohol and cigarette smoke synergize in the causation of various cancers.
- *Malnutrition.* Ethanol is a substantial source of energy (empty calories). Chronic alcoholism leads to malnutrition and nutritional deficiencies, particularly of the B vitamins.

Not all is gloom and doom, however. Moderate amounts of alcohol (about 20 to 30 g/day, corresponding to approximately 250 mL of wine) appear to be protective against coronary heart disease. Possible mechanisms include increase in HDL levels, inhibition of platelet aggregation, and lowering of fibrinogen levels. It seems that the old saying is true, at least with respect to alcohol—all things in moderation!

KEY CONCEPTS

ALCOHOL—METABOLISM AND HEALTH EFFECTS

- Acute alcohol abuse causes drowsiness at blood levels of approximately 200 mg/dL. Stupor and coma develop at higher levels.
- Alcohol is oxidized to acetaldehyde in the liver by alcohol dehydrogenase, by the cytochrome P-450 system, and by catalase, which is of minor importance. Diminished ability to metabolize acetaldehyde is associated with acute toxicity and an increased risk of certain cancers.
- Alcohol oxidation by alcohol dehydrogenase depletes NAD, leading to accumulation of fat in the liver and metabolic acidosis.
- The main effects of chronic alcoholism are fatty liver, alcoholic hepatitis, and cirrhosis, which leads to portal hypertension and increases the risk for development of hepatocellular carcinoma.
- Chronic alcoholism can cause bleeding from gastritis and gastric ulcers, peripheral neuropathy associated with thiamine deficiency, alcoholic cardiomyopathy, and acute and chronic pancreatitis.
- Chronic alcoholism is a major risk factor for cancers of the oral cavity, larynx, and esophagus. The risk is greatly increased by concurrent smoking or use of smokeless tobacco.

INJURY BY THERAPEUTIC DRUGS AND DRUGS OF ABUSE

Injury by Therapeutic Drugs (Adverse Drug Reactions)

Adverse drug reactions refer to untoward effects of drugs that are given in conventional therapeutic settings. These reactions are extremely common in the practice of medicine, affecting almost 7% of patients admitted to the hospitals, with a 0.32% fatality rate (accounting for 106,000 deaths annually). An exotic, but easily seen example is discoloration of the skin caused by accumulation of an oxidized metabolite of the antibiotic minocycline (Fig. 9.12). Much more common are drug reactions that are due to direct actions of the drug or to immunologically based hypersensitivity reactions. Drug-induced hypersensitivity reactions most commonly manifest as skin rashes, but they may also mimic autoimmune disorders such as systemic lupus erythematosus (Chapter 6) or take the form of hemolytic anemia or immune thrombocytopenia (Chapter 13).

Older adults (above 65 years) are much more likely to suffer from adverse drug reactions. Table 9.5 lists common pathologic findings in adverse drug reactions and the drugs

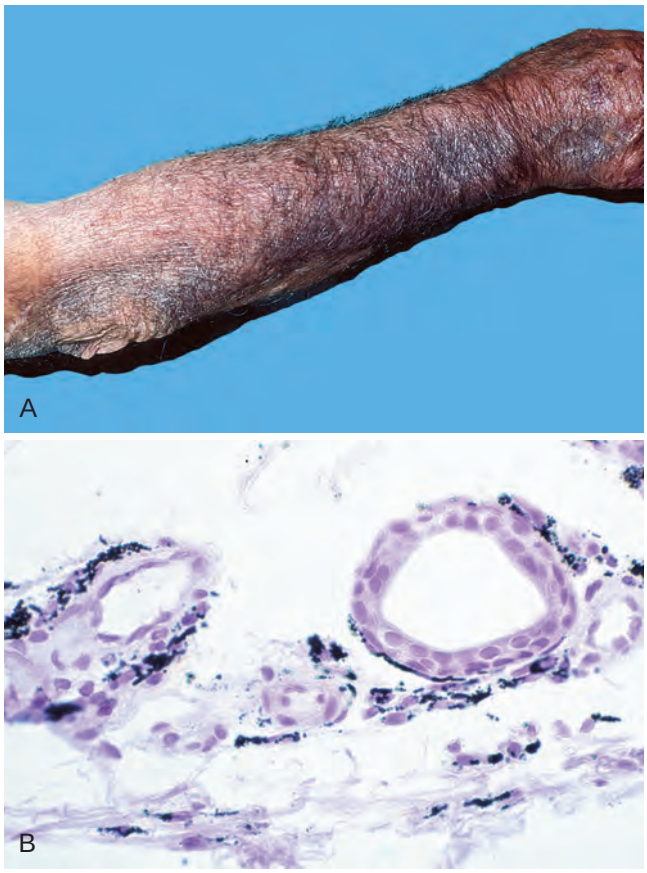


Figure 9.12 Adverse drug reaction. Skin pigmentation caused by minocycline, a long-acting tetracycline derivative. (A) Diffuse blue-gray pigmentation of the forearm. (B) Deposition of drug metabolite/iron/melanin pigment particles in the dermis. (Courtesy Dr. Zsolt Argenyi, Department of Pathology, University of Washington, Seattle, Wash.)

most frequently involved. Many of the drugs that produce adverse reactions, such as antineoplastic agents, are highly potent, and the adverse reactions are accepted risks of treatment. In this section, adverse reactions to commonly used drugs are examined, first discussing the unwelcome effects of anticoagulants, menopausal hormone therapy (MHT), and oral contraceptives (OCs) and then discussing the effects of acetaminophen and aspirin.

Anticoagulants

The two drugs that most frequently cause adverse reactions reported to the US Food and Drug Administration are the oral anticoagulants warfarin and dabigatran. Warfarin is an antagonist of vitamin K, and dabigatran is a direct inhibitor of thrombin. The principal complications associated with both of these medications are bleeding, which can be fatal, and thrombotic complications such as embolic stroke stemming from undertreatment. Warfarin is inexpensive, and its effects are easy to monitor, but many drugs and foods rich in vitamin K either interfere with its metabolism or abrogate its function. As a result, maintaining anticoagulation in a relatively safe therapeutic range can be problematic. Pharmacologic interactions of drugs with dabigatran metabolism have not been described, but many bleeding complications nevertheless occur. It is primarily used to prevent thromboembolism in patients with atrial fibrillation who are at high risk for thrombotic stroke.

Menopausal Hormone Therapy

The most common type of MHT (previously referred to as hormone replacement therapy) consists of the administration of estrogens together with a progestagen. Because of the risk of uterine cancer, therapy with estrogen alone is used only in hysterectomized women. Initially used to counteract

Table 9.5 Common Adverse Drug Reactions and Their Agents

Reaction	Major Offenders
Bone Marrow and Blood Cells[a]	
Granulocytopenia, aplastic anemia, pancytopenia	Antineoplastic agents, immunosuppressives, chloramphenicol
Hemolytic anemia, thrombocytopenia	Penicillin, methyldopa, quinidine, heparin
Cutaneous	
Urticaria, macules, papules, vesicles, petechiae, exfoliative dermatitis, fixed drug eruptions, abnormal pigmentation	Antineoplastic agents, sulfonamides, hydantoins, some antibiotics, and many other agents
Cardiac	
Arrhythmias	Theophylline, hydantoins, digoxin
Cardiomyopathy	Doxorubicin, daunorubicin
Renal	
Glomerulonephritis	Penicillamine
Acute tubular necrosis	Aminoglycoside antibiotics, cyclosporin, amphotericin B
Tubulointerstitial disease with papillary necrosis	Phenacetin, salicylates
Pulmonary	
Asthma	Salicylates
Acute pneumonitis	Nitrofurantoin
Interstitial fibrosis	Busulfan, nitrofurantoin, bleomycin
Hepatic	
Fatty change	Tetracycline
Diffuse hepatocellular damage	Halothane, isoniazid, acetaminophen
Cholestasis	Chlorpromazine, estrogens, contraceptive agents
Systemic	
Anaphylaxis	Penicillin
Lupus erythematosus syndrome (drug-induced lupus)	Hydralazine, procainamide
Bleeding	Warfarin, dabigatran
Central Nervous System	
Tinnitus and dizziness	Salicylates
Acute dystonic reactions and parkinsonian syndrome	Phenothiazine antipsychotics
Respiratory depression	Sedatives

[a]Affected in almost half of all drug-related deaths.

hot flashes and other symptoms of menopause, early clinical studies suggested that MHT use in postmenopausal women could prevent or slow the progression of osteoporosis (Chapter 26) and reduce the likelihood of myocardial infarction. However, subsequent randomized clinical trials have produced decidedly mixed results. In 2002, the Women's Health Initiative stunned the medical community by reporting that a large prospective placebo-controlled trial failed to find support for some of the presumed beneficial effects of the therapy. This study involved approximately 17,000 women who were taking a combination of estrogen (conjugated equine estrogens) and a synthetic progestin (medroxyprogesterone acetate). Although MHT did reduce the number of fractures in women on treatment, researchers also reported that after 5 years of treatment, combination MHT increased the risk of breast cancer (Chapter 23), stroke, and venous thromboembolism (VTE) and had no effect on the incidence of coronary heart disease. The shock waves produced by these findings led to a drastic decrease in the use of MHT, from 22% to 4.7% in 2010, which was accompanied by an apparent drop in the incidence of newly diagnosed breast cancers. But during the past few years there has been a reappraisal of the risks and benefits of MHT. These newer analyses showed that **MHT effects depend on the type of hormone therapy regimen used (combination estrogen-progestin versus estrogen alone); the age and risk factor status of the woman at the start of treatment; the duration of the treatment; and possibly the hormone dose, formulation, and route of administration.** The current risk/benefit consensus can be summarized as follows:

- *Combination estrogen-progestin increases the risk of breast cancer* after a median time of 5 to 6 years. By contrast, estrogen alone in women with hysterectomy is associated with a borderline reduction in risk of breast cancer. There is no increase in the risk for ovarian cancer.
- *MHT may have a protective effect on the development of atherosclerosis and coronary disease in women younger than age 60 years,* but there is no protection in women who started MHT at an older age. These data support the notion that there may be a critical therapeutic window for MHT effects on the cardiovascular system. Protective effects in younger women depend in part on the response of estrogen receptors in healthy vascular endothelium. However, MHT should not be used for prevention of cardiovascular disease or other chronic diseases.
- *MHT increases the risk of stroke and VTE including deep vein thrombosis and pulmonary embolism.* The increase in VTE is more pronounced during the first 2 years of treatment and in women who have other risk factors, such as immobilization and hypercoagulable states caused by prothrombin or factor V Leiden mutations (Chapter 4). The risks of VTE and stroke appear to be lower with transdermal than oral routes of estrogen. The effect of route of administration continues to be studied.

As can be appreciated from these associations, assessment of risks and benefits when considering the use of MHT in women is complex. The current feeling is that these agents have a role in the management of menopausal symptoms in early menopause but should not be used long term for chronic disease prevention.

Oral Contraceptives

Worldwide, millions of women use hormonal contraception. OCs nearly always contain a synthetic estradiol and a variable amount of a progestin, but some preparations contain only progestins. They act by inhibiting ovulation or preventing implantation. Currently prescribed OCs contain a much smaller amount of estrogens (as little as 20 μg of ethinyl estradiol) than the earliest formulations and are associated with fewer side effects. Transdermal and implantable formulations have also become available. Hence the results of epidemiologic studies should be interpreted in the context of the dosage and the delivery system. Nevertheless, there is good evidence to support the following conclusions.

- *Breast carcinoma.* The prevailing opinion is that OCs do not increase breast cancer risk.
- *Endometrial cancer and ovarian cancers.* OCs have a protective effect against these tumors that may last for decades following cessation of OC use.
- *Cervical cancer.* OCs may increase risk of cervical carcinomas in women infected with human papillomavirus.
- *Thromboembolism.* Most studies indicate that OCs including the newer low-dose (less than 50 μg of estrogen) preparations are associated with a twofold to fourfold increased risk of venous thrombosis and pulmonary thromboembolism due to a hypercoagulable state induced by elevated hepatic synthesis of coagulation factors. This risk may be even higher with newer third-generation OCs that contain synthetic progestins, particularly in women who are carriers of the factor V Leiden mutation. To put this complication into context, however, the risk of thromboembolism associated with OC use is two to six times lower than the risk of thromboembolism associated with pregnancy.
- *Cardiovascular disease.* There is considerable uncertainty about the risk of atherosclerosis and myocardial infarction in users of OCs. It seems that OCs do not increase the risk of coronary artery disease in women younger than 30 years or in older women who are nonsmokers, but the risk does approximately double in women older than 35 years who smoke.
- *Hepatic adenoma.* There is a well-defined association between the use of OCs and this rare benign hepatic tumor, especially in older women who have used OCs for prolonged periods.

Ultimately, the pros and cons of OCs must be viewed in the context of their wide applicability and acceptance as a form of contraception.

Acetaminophen

Acetaminophen is the most commonly used analgesic in the United States. It is present in more than 300 products, alone or in combination with other agents. In the United States, it is the cause of about 50% of cases of acute liver failure, with 30% mortality. Intentional overdose (attempted suicide) is the most common cause of acetaminophen toxicity in Great Britain, but unintentional overdose is the most frequent cause in the United States, representing almost 50% of the total intoxication cases.

At therapeutic doses, about 95% of acetaminophen undergoes detoxification in the liver by phase II enzymes and is excreted in the urine as glucuronate or sulfate conjugates

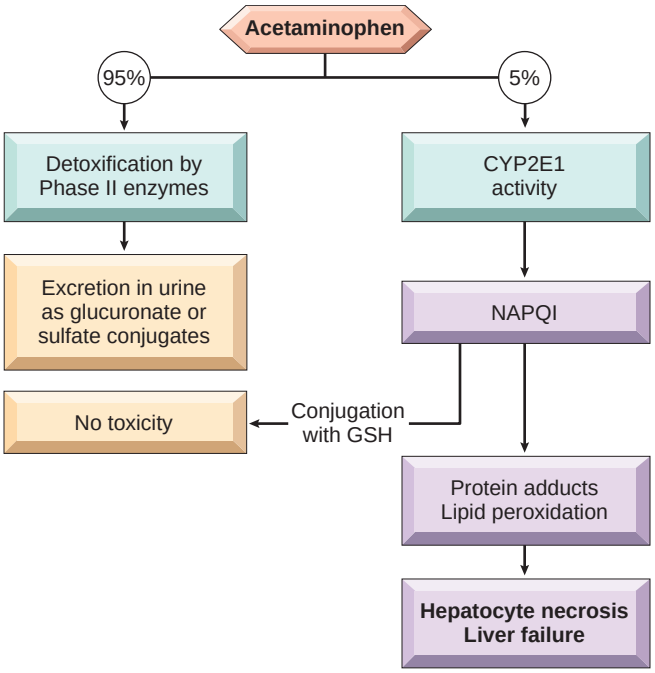


Figure 9.13 Acetaminophen metabolism and toxicity. See text for details. *GSH*, Glutathione; *NAPQI*, NAPQI, *N*-acetyl-*p*-benzoquinoneimine. (Courtesy Dr. Xavier Vaquero, Department of Pathology, University of Washington, Seattle, Wash.)

(Fig. 9.13). About 5% or less is metabolized through the activity of CYPs (primarily CYP2E) to NAPQI (*N*-acetyl-*p*-benzoquinoneimine), a highly reactive metabolite. NAPQI is normally conjugated with GSH, but when acetaminophen is taken in large doses, unconjugated NAPQI accumulates and causes hepatocellular injury, leading to centrilobular necrosis that may progress to liver failure. The injury produced by NAPQI involves two mechanisms: (1) covalent binding to hepatic proteins, which causes damage to cellular membranes and mitochondrial dysfunction, and (2) depletion of GSH, making hepatocytes more susceptible to ROS-induced injury. Because alcohol induces CYP2E in the liver, toxicity can occur at lower doses in chronic alcoholics.

The window between the usual dose (0.5 g) and the toxic dose (15 to 25 g) is large, and the drug is ordinarily very safe. Toxicity begins with nausea, vomiting, diarrhea, and sometimes shock, followed in a few days by evidence of jaundice. Overdose of acetaminophen can be treated in its early stages (within 12 hours) by administration of *N*-acetylcysteine, which restores GSH levels. In serious overdose, liver failure ensues, starting with centrilobular necrosis that may extend to entire lobules; in such circumstances liver transplantation is the only hope for survival. Some patients also show evidence of concurrent renal damage.

Aspirin (Acetylsalicylic Acid)

Aspirin overdose may result from accidental ingestion of a large number of tablets by young children; in adults, overdose is frequently suicidal. Much less commonly, salicylate poisoning is caused by the excessive use of ointments containing oil of wintergreen (methyl salicylate). Acute salicylate overdose causes alkalosis as a consequence of the stimulation of the respiratory center in the medulla. This is followed by metabolic acidosis and accumulation of pyruvate and lactate, caused by uncoupling of oxidative phosphorylation and inhibition of the Krebs cycle. Metabolic acidosis enhances the formation of nonionized forms of salicylates, which diffuse into the brain and produce effects ranging from nausea to coma. Ingestion of 2 to 4 g by children or 10 to 30 g by adults may be fatal, but survival has been reported after ingestion of doses five times larger.

Chronic aspirin toxicity (salicylism) may develop in persons who take 3 g or more daily for long periods of time for treatment of chronic pain or inflammatory conditions. Chronic salicylism is manifested by headaches, dizziness, ringing in the ears (tinnitus), hearing impairment, mental confusion, drowsiness, nausea, vomiting, and diarrhea. The CNS changes may progress to convulsions and coma. The morphologic consequences of chronic salicylism are varied. Most often there is an acute erosive gastritis (Chapter 17), which may produce gastrointestinal bleeding and lead to gastric ulceration. A bleeding tendency may appear concurrently with chronic toxicity because aspirin acetylates platelet cyclooxygenase and irreversibly blocks the production of thromboxane A_2, an activator of platelet aggregation. Petechial hemorrhages may appear in the skin and internal viscera, and bleeding from gastric ulcerations may be exacerbated by the clotting defect. With the recognition of gastric ulceration and bleeding as an important complication of ingestion of large doses of aspirin, chronic toxicity is now quite uncommon.

Proprietary analgesic mixtures of aspirin and phenacetin or its active metabolite, acetaminophen, when taken over several years, can cause tubulointerstitial nephritis with renal papillary necrosis, referred to as analgesic nephropathy (Chapter 20).

Injury by Nontherapeutic Agents (Drug Abuse)

Drug abuse generally involves the repeated or chronic use of mind-altering substances, beyond therapeutic or social norms, and may lead to drug addiction and overdose, both serious public health problems. According to the 2019 United Nations Office on Drugs and Crime report, it is estimated that about 275 million people worldwide, which is roughly 5.6% of the global population aged 15 to 64 years, used drugs at least once during 2016. Approximately, 450,000 people died in 2015 as a result of drug use. Of those deaths, 167,750 were a direct result of drug use disorders, in most cases involving opioids. Common drugs of abuse are listed in Table 9.6. Considered here are cocaine, opioids, amphetamines, and marijuana, among others.

Cocaine

According to the United Nations World Drug Report 2016, globally, there were 18 million users of cocaine in 2016. Of these, 5.1 million users were in North America, representing 1.4% of the population over 14 years of age.

Cocaine is extracted from the leaves of the coca plant and is usually prepared as a water-soluble powder, cocaine hydrochloride. Cocaine sold on the street is liberally diluted with talcum powder, lactose, or other look-alikes. It can be snorted or dissolved in water and injected subcutaneously or intravenously. Crystallization of the pure alkaloid yields nuggets of crack, so called because of the cracking or popping

Table 9.6 Common Drugs of Abuse

Class	Molecular Target	Example
Opioid narcotics	Mu opioid receptor (agonist)	Heroin, hydromorphone (Dilaudid) Oxycodone (OxyContin) Methadone (Dolophine) Meperidine (Demerol)
Sedative-hypnotics	$GABA_A$ receptor (agonist)	Barbiturates Ethanol Methaqualone (Quaalude) Glutethimide (Doriden) Ethchlorvynol (Placidyl)
Psychomotor stimulants	Dopamine transporter (antagonist) Serotonin receptors (toxicity)	Cocaine Amphetamines 3,4-Methylenedioxymethamphetamine (MDMA, Ecstasy)
Phencyclidine-like drugs	NMDA glutamate receptor channel (antagonist)	Phencyclidine (PCP, angel dust) Ketamine
Cannabinoids	CB1 cannabinoid receptors (agonist)	Marijuana Hashish
Hallucinogens	Serotonin 5-HT_2 receptors (agonist)	Lysergic acid diethylamide (LSD) Mescaline Psilocybin

GABA, γ-Aminobutyric acid; *5-HT_2*, 5-hydroxytryptamine; *NMDA*, *N*-methyl D-aspartate.

From Hyman SE: A 28-year-old man addicted to cocaine, *JAMA* 286:2586, 2001.

sound it makes when heated to produce vapors that are inhaled. The pharmacologic actions of cocaine and crack are identical, but crack is far more potent.

Cocaine produces intense euphoria and neurologic stimulation, making it one of the most addictive drugs. Experimental animals will press a lever more than 1000 times and forgo food and drink to obtain it. In the cocaine user, although physical dependence generally does not occur, the psychologic withdrawal is profound and can be extremely difficult to treat. Intense cravings are particularly severe in the first several months after abstinence and can recur for years. **Cocaine facilitates neurotransmission both in the CNS, where it blocks the reuptake of dopamine, and at adrenergic nerve endings, where it blocks the reuptake of both epinephrine and norepinephrine while stimulating the presynaptic release of norepinephrine.** The euphoria is due to its enhancement of brain dopamine activity, especially in the so-called mesolimbic dopamine reward pathway.

The acute and chronic effects of cocaine on various organ systems are as follows.

- *Cardiovascular effects*. The most serious cardiac effects of cocaine relate to its sympathomimetic activity resulting from blockade of reuptake of both epinephrine and norepinephrine at adrenergic nerve endings (Fig. 9.14). The net effect is the accumulation of these two neurotransmitters in synapses, resulting in excess stimulation, manifested by tachycardia, hypertension, and peripheral vasoconstriction. Cocaine may also induce myocardial ischemia by causing coronary artery vasoconstriction and by enhancing platelet aggregation and thrombus formation. These dual effects of cocaine increase myocardial oxygen demand and decrease coronary blood flow, setting the stage for myocardial ischemia and possible myocardial infarction. Cocaine can also precipitate lethal arrhythmias by enhanced sympathetic activity as well as by disrupting normal ion (K^+, Ca^{2+}, Na^+) transport in the myocardium. These toxic effects are not necessarily dose related, and a fatal event may occur in a first-time user with what is a typical mood-altering dose. They may also be potentiated by cigarette smoking, which increases the likelihood of cocaine-induced coronary vasospasm.
- *CNS*. The most common acute effects on the CNS are hyperpyrexia (thought to be caused by aberrations of the dopaminergic pathways that control body temperature) and seizures.
- *Effects on pregnancy*. In pregnant women, cocaine may cause acute decreases in blood flow to the placenta, resulting in fetal hypoxia and spontaneous abortion. Neurologic development may be impaired in the fetus of a pregnant woman who is a chronic drug user.
- *Other effects*. Chronic cocaine use may cause (1) perforation of the nasal septum in snorters, (2) decreased lung diffusing capacity in those who inhale the smoke, and (3) dilated cardiomyopathy.

Opioids and Opiates

According to the 2019 United Nations Office on Drugs and Crime report, in 2016 there were 19 million users of opiates (e.g., morphine and codeine, derived from poppy plants) and 34 million users of opioids (natural and synthetic). Opioid drugs of abuse include prescription drugs such as oxycodone, hydrocodone, fentanyl, tramadol, and methadone. With a death toll of 49,000 from opioid overdoses in 2016, the problem is of epidemic proportions in North America. Of these deaths, 40% were from prescription opioids.

Heroin is a street drug derived from the poppy plant that is closely related to morphine. Use of heroin is even more harmful than cocaine use. As sold on the street, it is cut (diluted) with an agent (often talc or quinine); thus the size of the dose is not only variable but also usually unknown to the buyer. Heroin, along with any contaminating substances, is usually self-administered intravenously or subcutaneously. The effects on the CNS are varied and include euphoria, hallucinations, somnolence, and sedation. Heroin has a wide range of other adverse physical effects related to (1) the pharmacologic action of the agent, (2) reactions to the

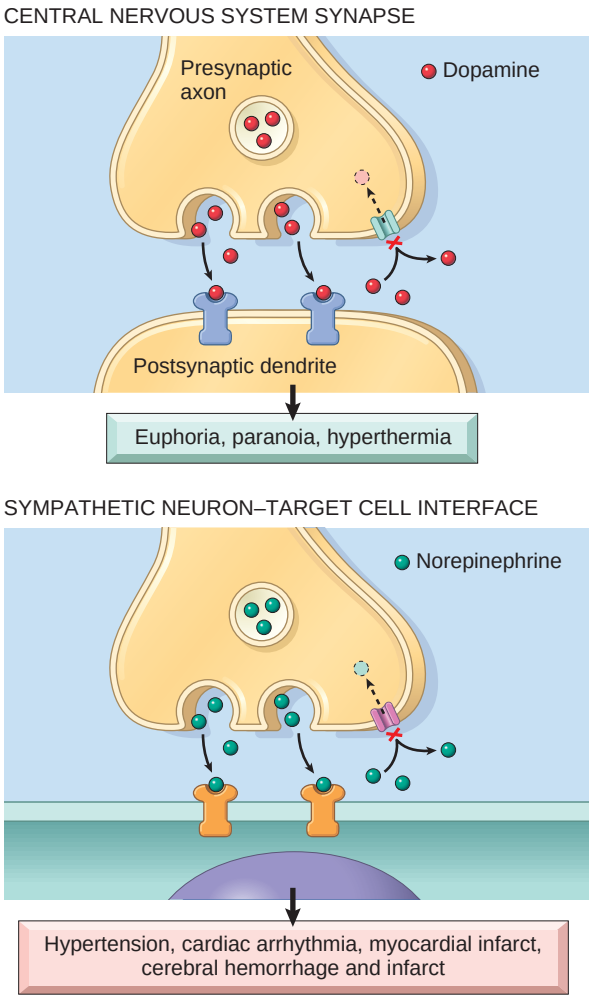


Figure 9.14 Effect of cocaine on neurotransmission. The drug inhibits reuptake of the neurotransmitters dopamine and norepinephrine in the central and peripheral nervous systems.

cutting agents or contaminants, (3) hypersensitivity reactions to the drug or its adulterants (quinine itself has neurologic, renal, and auditory toxicity), and (4) diseases contracted incident to the use of contaminated needles. Some of the most important adverse effects of heroin follow.

- *Sudden death.* Sudden death, usually related to overdose, is an ever-present risk because drug purity is generally unknown (ranging from 2% to 90%). The yearly mortality among heroin users in the United States is estimated to be between 1% and 3%. Addition of very potent synthetic opioids such as fentanyl and carfentanil to heroin, a common practice among dealers, greatly increases the chances of fatal overdose. Sudden death can also occur if heroin is taken after tolerance for the drug, built up over time, is lost (as during a period of incarceration). The mechanisms of death include profound respiratory depression, arrhythmia and cardiac arrest, and severe pulmonary edema.
- *Pulmonary injury.* Pulmonary complications include moderate to severe edema, septic embolism from endocarditis, lung abscess, opportunistic infections, and foreign-body granulomas from talc and other adulterants. Although granulomas occur principally in the lung, they are sometimes found in the mononuclear phagocyte system, particularly in the spleen, liver, and lymph nodes that drain the upper extremities. Examination under polarized light often highlights trapped talc crystals, sometimes enclosed within foreign-body giant cells.
- *Infections.* Infectious complications are common. The four sites most commonly affected are the skin and subcutaneous tissue, heart valves, liver, and lungs. In a series of addicted patients admitted to the hospital, more than 10% had endocarditis, which often takes a distinctive form involving right-sided heart valves, particularly the tricuspid. Most cases are caused by *Staphylococcus aureus,* but fungi and a multitude of other organisms have also been implicated. Viral hepatitis is the most common infection among addicted persons and is acquired by the sharing of dirty needles. In the United States, this practice has also led to a very high incidence of human immunodeficiency virus (HIV) infection in intravenous drug users.
- *Skin.* Cutaneous lesions are probably the most frequent telltale sign of heroin addiction. Acute changes include abscesses, cellulitis, and ulcerations due to subcutaneous injections. Scarring at injection sites, hyperpigmentation over commonly used veins, and thrombosed veins are the usual sequelae of repeated intravenous inoculations.
- *Kidneys.* Kidney disease is a relatively common hazard. The two forms most frequently encountered are amyloidosis (generally secondary to skin infections) and focal and segmental glomerulosclerosis; both induce proteinuria and nephrotic syndrome.

Use of oxycodone, an oral opioid available by prescription for treatment of pain, has increased sharply in recent years in the United States. According to the National Institute on Drug Abuse, approximately 5% of high school seniors took oxycodone in 2010, sometimes with tragic results due to the potent respiratory suppressant effect of the drug. The overall number of yearly fatalities attributed to abuse of prescription opiates in the United States rose from approximately 3000 in 1999 to approximately 49,000 in 2017. Most of this increase is attributable to abuse of oxycodone, which has surpassed heroin as the leading cause of opiate-related death in the United States.

Amphetamines and Related Drugs

Methamphetamine. This addictive drug, known as "speed" or "meth," is closely related to amphetamine but has stronger effects in the CNS. Methamphetamine use rose rapidly in the United States in the early 2000s, peaking in the year 2005, but has fallen steadily since that time. According to national surveys, use of methamphetamine fell to approximately 560,000 users in 2017, a decrease of more than 50% since 2006. Methamphetamine acts by releasing dopamine in the brain, which inhibits presynaptic neurotransmission at corticostriatal synapses, slowing glutamate release. Methamphetamine produces a feeling of euphoria, which is followed by a "crash." Long-term use leads to violent behaviors, confusion, and psychosis marked by paranoia and hallucinations.

Marijuana

It is estimated that worldwide 192 million people used marijuana (or "pot") in 2017, making it far and away the most widely used illicit drug globally. Several states in the United States have legalized the recreational use of marijuana

in 2018, and more appear poised to follow; thus, its status as an illicit drug is undergoing reevaluation.

Marijuana is made from the leaves of the *Cannabis sativa* plant, which contain the psychoactive substance Δ^9-tetrahydrocannabinol (THC). About 5% to 10% of THC is absorbed when it is smoked in a hand-rolled cigarette ("joint"). Marijuana use causes euphoria and a sense of relaxation. In many people it causes a heightened sensory perception (e.g., brighter colors), laughter, altered perception of time, and increased appetite. Among the beneficial effects of marijuana is its potential use to treat nausea secondary to cancer chemotherapy and as an agent capable of decreasing pain in some chronic conditions that are otherwise difficult to treat. The functional and organic CNS consequences of marijuana smoking have received the most scrutiny. Its use distorts sensory perception and impairs motor coordination, but these acute effects generally clear in 4 to 5 hours. With continued use these changes may progress to cognitive and psychomotor impairments, such as inability to judge time, speed, and distance, a potential cause of automobile accidents. Marijuana increases the heart rate and sometimes blood pressure, and it may cause angina in a person with coronary artery disease.

The respiratory system is also affected by chronic marijuana smoking; laryngitis, pharyngitis, bronchitis, cough and hoarseness, and asthma-like symptoms all have been described, along with mild but significant airway obstruction. Marijuana cigarettes contain a large number of carcinogens that are also present in tobacco. Smoking a marijuana cigarette, compared with a tobacco cigarette, is associated with a threefold increase in the amount of tar inhaled and retained in the lungs, presumably because of the larger puff volume, deeper inhalation, and longer breath holding. With the advent of e-cigarettes and vaping, an increasing number of users are inhaling aerosolized THC, a practice that has been associated with severe lung injury in some users, possibly due to the presence of adulterants. Paradoxically, some heavy users develop cannabis hyperemesis syndrome, marked by intractable nausea and vomiting that only remits with cessation of use.

Whether marijuana causes physical dependence and addiction remains controversial. According to the 2018 report from the US National Institute on Drug Abuse, chronic use of marijuana, especially if the exposure started during adolescence, gives rise to marijuana use disorder. Dependence in such individuals is manifested by withdrawal symptoms such as mood and sleep difficulties that may last up to 2 weeks when not taking the drug. A small subset of those who develop dependence may become addicted—individuals who cannot stop using the drug even though it interferes with many aspects of their life. The exact number of people who become addicted is uncertain since epidemiologic studies have not used uniform criteria to define addiction to marijuana.

Other Drugs

The variety of drugs that have been tried by individuals seeking "new experiences" (highs, lows, "out-of-body experiences") are legion. These drugs include various stimulants, depressants, analgesics, and hallucinogens (Table 9.6). Among these are 1-(1-phenylcyclohexyl) piperidine (PCP), or phencyclidine, and ketamine (related anesthetic agents); lysergic acid diethylamide (LSD), the most potent hallucinogen known; Ecstasy (3,4-methylenedioxymethamphetamine [MDMA]); and "bath salts," synthetic cathinones that are chemically related to khat, a widely used stimulant in East Africa. Not much is known about the long-term deleterious effects of any of these agents. Acutely, LSD has unpredictable effects on mood, affect, and thought, sometimes leading to bizarre and dangerous behaviors. Chronic use of Ecstasy may deplete the CNS of serotonin, potentially leading to sleep disorders, depression, anxiety, and aggressive behavior.

On the other hand, therapeutic uses for psychoactive drugs are emerging. Ketamine derivatives taken in low doses have recently been approved for treatment of severe depression, and clinical trials exploring the use of hallucinogens as treatment for psychiatric conditions such as post-traumatic stress disorder are ongoing.

KEY CONCEPTS

DRUG INJURY

- Drug injury may be caused by therapeutic drugs (adverse drug reactions) or nontherapeutic agents (drug abuse).
- Antineoplastic agents, anticoagulants, MHT preparations and OCs, acetaminophen, and aspirin are among the therapeutic drugs involved most frequently.
- MHT increases the risk of endometrial and breast cancers and thromboembolism and does not appear to protect against ischemic heart disease in women over 60 years of age. OCs have a protective effect against endometrial and ovarian cancer but increase the risk of thromboembolism and hepatic adenoma.
- Overdose of acetaminophen may cause centrilobular liver necrosis, leading to liver failure. Early treatment with agents that restore GSH levels may limit toxicity. Aspirin produces gastric ulceration and blocks the production of thromboxane A_2 in platelets, which may produce bleeding.
- The common drugs of abuse include sedative-hypnotics (barbiturates, ethanol), psychomotor stimulants (cocaine, methamphetamine, Ecstasy), opioid narcotics (heroin, oxycodone), hallucinogens, and cannabinoids (marijuana).

INJURY BY PHYSICAL AGENTS

Injury induced by physical agents is divided into the following categories: mechanical trauma, thermal injury, electrical injury, and injury produced by ionizing radiation. Each type is considered separately.

Mechanical Trauma

Mechanical forces may inflict a variety of forms of damage. The type of injury depends on the shape of the colliding object, the amount of energy discharged at impact, and the tissues or organs that bear the impact. Bone and head injuries result in unique damage and are discussed elsewhere (Chapters 26 and 28). All soft tissues react similarly to mechanical forces, and the patterns of injury can be divided into abrasions, contusions, lacerations, incised wounds, and puncture wounds. This is just a small sampling of the various forms of trauma encountered by forensic pathologists, who deal with wounds produced by shooting, stabbing, blunt

force, traffic accidents, and other causes. In addition to morphologic analyses, forensic pathology now includes molecular methods for identity testing and sophisticated methods to detect the presence of foreign substances. Details about the practice of forensic pathology can be found in specialized textbooks.

Thermal Injury

Both excessive heat and excessive cold are important causes of injury. Burns are the most common cause of thermal injury and are discussed first; a brief discussion of hyperthermia and hypothermia follows.

Thermal Burns

In the United States, approximately 450,000 persons per year receive medical treatment for burn injuries. Eighty percent of burns are caused by fire or by scalding, the latter being a major cause of injury in children. It is estimated that approximately 3500 persons die each year as a consequence of injuries caused by fire and smoke inhalation, mostly originating in homes. Since the 1970s, marked decreases have been seen in both mortality rates and the length of hospitalizations of burn patients. In recent years there were approximately 45,000 hospitalizations per year for burns; among patients treated in specialized burn centers (about 55% of those hospitalized), the survival rate was more than 95%, a remarkable testimony to improvements in the care of patients with severe burns. These improvements have been achieved by a better understanding of the systemic effects of extensive burns, the prevention of wound infection, and the use of treatments that promote the healing of skin surfaces.

The clinical significance of a burn injury depends on the following factors:

- *Depth* of the burns
- *Percentage* of body surface involved
- *Internal injuries* caused by the inhalation of hot and toxic fumes
- *Promptness and efficacy of therapy,* especially fluid and electrolyte management and prevention or control of wound infections

Burns used to be classified as first degree to fourth degree, according to the depth of the injury (first-degree burns being the most superficial), but are now classified as superficial, partial-thickness, and full-thickness.

- *Superficial burns* (formerly known as *first-degree burns*) are confined to the epidermis.
- *Partial-thickness burns* (formerly known as *second-degree burns*) involve injury to the dermis.
- *Full-thickness burns* (formerly known as *third-degree burns*) extend to the subcutaneous tissue. Full-thickness burns may also involve damage to underlying muscle tissue underneath the subcutaneous tissue (these were known formerly as fourth-degree burns).

Shock, sepsis, and respiratory insufficiency are the greatest threats to life in burn patients. Particularly in burns of more than 20% of the body surface, there is a rapid (within hours) shift of body fluids into the interstitial compartments throughout the body due to the *systemic inflammatory response syndrome,* leading to shock (Chapter 4). Because of widespread vascular leakiness, generalized edema, including pulmonary edema, can be severe. An important pathophysiologic effect of burns is the development of a hypermetabolic state associated with excess heat loss and an increased need for nutritional support. It is estimated that when more than 40% of the body surface is burned, the resting metabolic rate doubles.

The burn site is ideal for the growth of microorganisms; the serum and debris provide nutrients and the burn injury compromises blood flow, blocking effective inflammatory responses. As a result, virtually all burns become colonized with bacteria. Infections are defined by the presence of greater than 10^5 bacteria per gram of tissue, and invasive local infection is defined by the presence of greater than 10^5 bacteria per gram in unburned adjacent tissue. The most common offender is the opportunist *Pseudomonas aeruginosa,* but antibiotic-resistant strains of other common hospital-acquired bacteria, such as methicillin-resistant *S. aureus,* and fungi, particularly *Candida* species, may also be involved. Furthermore, systemic inflammatory response syndrome may impair or dysregulate both innate and adaptive immune responses (Chapter 4). Direct bacteremic spread and release of toxic substances such as endotoxin from the local site may also have dire consequences. Pneumonia or septic shock with renal failure and/or acute respiratory distress syndrome (Chapter 15) are the most common serious sequelae.

Organ system failure resulting from burn sepsis has greatly diminished during the past 30 years because of the introduction of techniques for early excision and grafting of the burn wound. Removal of the burn wound decreases infection and reduces the need for reconstructive surgery. Grafting is done with split-thickness skin grafts; dermal substitutes, which serve as a bed for cell repopulation, may be used in large full-thickness burns.

Injury to the airways and lungs may develop within 24 to 48 hours after the burn and may result from the direct effect of heat on the mouth, nose, and upper airways or from the inhalation of heated air and noxious gases in the smoke. Water-soluble gases, such as chlorine, sulfur oxides, and ammonia, may react with water to form acids or alkalis, particularly in the upper airways, producing inflammation and swelling, which may lead to partial or complete airway obstruction. Lipid-soluble gases, such as nitrous oxide and products of burning plastics, are more likely to reach deeper airways, producing pneumonitis.

In burn survivors the development of hypertrophic scars, both at the site of the original burn and at donor graft sites, and itching may become long-term, difficult-to-treat problems. Hypertrophic scarring is a common complication of burn injury marked by excessive deposition of collagen in the healing wound bed; its etiology is not well understood.

MORPHOLOGY

Grossly, full-thickness burns are white or charred, dry, and painless (because of destruction of nerve endings), whereas, depending on the depth, partial-thickness burns are pink or mottled with blisters and are painful. Histologically, devitalized tissue reveals coagulative necrosis, adjacent to vital tissue that quickly accumulates inflammatory cells and marked exudation.

Hyperthermia

Prolonged exposure to elevated ambient temperatures can result in heat cramps, heat exhaustion, and heat stroke.

- *Heat cramps* result from loss of electrolytes via sweating. Cramping of voluntary muscles, usually in association with vigorous exercise, is the hallmark. Heat-dissipating mechanisms are able to maintain normal core body temperature.
- *Heat exhaustion* is probably the most common hyperthermic syndrome. Its onset is sudden, with prostration and collapse, and it results from a failure of the cardiovascular system to compensate for hypovolemia caused by dehydration. After a period of collapse, which is usually brief, equilibrium is spontaneously reestablished if the victim is able to rehydrate.
- *Heat stroke* is associated with high ambient temperatures, high humidity, and exertion. Older adults, individuals undergoing intense physical stress (including young athletes and military recruits), and persons with cardiovascular disease are at particularly high risk for heat stroke. Thermoregulatory mechanisms fail, sweating ceases, and the core body temperature rises to more than 40°C, leading to multiorgan dysfunction that can be rapidly fatal. The hyperthermia is accompanied by marked generalized vasodilation, with peripheral pooling of blood and a decreased effective circulating blood volume. Hyperkalemia, tachycardia, arrhythmias, and other systemic effects are common. Particularly important, however, are sustained contractions of skeletal muscle that can exacerbate the hyperthermia and lead to muscle necrosis (rhabdomyolysis). These phenomena appear to stem from nitrosylation of ryanodine receptor 1 (RYR1), which is located in the sarcoplasmic reticulum of skeletal muscle. RYR1 regulates the release of calcium from the sarcoplasm. Heat stroke deranges RYR1 function and allows calcium to leak into the cytoplasm, where it stimulates muscle contraction and heat production.
- *Malignant hyperthermia,* despite the name, is not caused by exposure to high temperatures. It is a genetic condition resulting from mutations in genes such as *RYR1* that control calcium levels in skeletal muscle cells. In affected individuals, exposure to certain anesthetics during surgery triggers a rapid rise in calcium levels in skeletal muscle, leading to muscle rigidity and increased heat production. The resulting hyperthermia has a mortality rate of approximately 80% if untreated, but this falls to less than 5% if the condition is recognized and muscle relaxants are administered promptly.

Hypothermia

Prolonged exposure to low ambient temperature leads to hypothermia, a condition seen all too frequently in homeless persons. High humidity and wet clothing, sometimes exacerbated by dilation of superficial blood vessels due to ingestion of alcohol, hasten the lowering of body temperature. At a body temperature of about 90°F, loss of consciousness occurs, followed by bradycardia and atrial fibrillation at lower core temperatures.

Hypothermia causes injury by two mechanisms:

- *Direct effects* are probably mediated by physical disruptions within cells by high salt concentrations caused by the crystallization of intracellular and extracellular water (frostbite).
- *Indirect effects* result from circulatory changes, which vary depending on the rate and duration of the temperature drop. Slow chilling may induce vasoconstriction and increase vascular permeability, leading to edema and hypoxia. Such changes are typical of trench foot. This condition developed in soldiers who spent long periods of time in waterlogged trenches during World War I (1914–1918), frequently causing gangrene that necessitated amputation. With sudden, persistent chilling, the vasoconstriction and increased viscosity of the blood in the local area may cause ischemic injury and degenerative changes in peripheral nerves. In this situation, vascular injury and edema become evident only after the temperature begins to return to normal. However, if the period of ischemia is prolonged, hypoxic changes and infarction of the affected tissues (e.g., gangrene of toes or feet) may result.

Electrical Injury

Electrical injuries, which are often fatal, can arise from contact with low-voltage currents (i.e., in the home and workplace) or high-voltage currents carried by high-power lines or produced by lightning. Injuries are of two types: (1) burns and (2) ventricular fibrillation or cardiac and respiratory center failure, resulting from disruption of nerve impulse conduction. The type of injury and the severity and extent of burns depend on the strength (amperage), duration, and path of the electric current within the body.

Voltage in the household and workplace (120 or 220 V) is high enough that with low resistance at the site of contact (as when the skin is wet), sufficient current can pass through the body to cause serious injury, including ventricular fibrillation. If the current flow is sustained, it may generate enough heat to produce burns at the site of entry and exit as well as in internal organs. An important characteristic of alternating current, the type supplied to most homes, is that it induces tetanic muscle spasm, so that when a live wire or switch is grasped, irreversible clutching is likely to occur, prolonging the period of current flow. This results in a greater likelihood of developing extensive electrical burns and, in some cases, spasm of the chest wall muscles, producing death from asphyxia. Currents generated from high-voltage sources cause similar damage; however, because of the large current flows generated, these are more likely to produce paralysis of medullary centers and extensive burns. Lightning is a classic cause of high-voltage electrical injury.

Injury Produced by Ionizing Radiation

Radiation is energy that travels in the form of waves or high-speed particles. Radiation has a wide range of energies that span the electromagnetic spectrum; it can be divided into nonionizing and ionizing radiation. The energy of nonionizing radiation such as UV and infrared light, microwave, and sound waves, can move atoms in a molecule or cause them to vibrate but is not sufficient to displace bound electrons from atoms. By contrast, ionizing radiation has sufficient energy to remove tightly bound electrons.

Collision of electrons with other molecules releases electrons in a reaction cascade, referred to as ionization. The main sources of ionizing radiation are (1) x-rays and gamma rays (electromagnetic waves of very high frequencies), (2) high-energy neutrons, alpha particles (composed of two protons and two neutrons), and (3) beta particles, which are essentially electrons. At equivalent amounts of energy, alpha particles induce heavy damage in a restricted area, whereas x-rays and gamma rays dissipate energy over a longer, deeper course and produce considerably less damage per unit of tissue. About 50% of the total dose of ionizing radiation received by the US population is human-made, mostly originating from medical devices and radioisotopes. In fact, the exposure of patients to ionizing radiation during radiologic imaging tests roughly doubled between the early 1980s and 2006, mainly because of much more widespread use of computed tomography (CT) scans.

Ionizing radiation is a double-edged sword. It is indispensable in medical practice, being used in the treatment of cancer, in diagnostic imaging, and in therapeutic or diagnostic radioisotopes, but it also produces adverse short-term and long-term effects such as fibrosis, mutagenesis, carcinogenesis, and teratogenesis.

Radiation Units. Several somewhat confusing terms are used to describe radiation dose, which can be quantified according to the amount of radiation emitted by a source, the amount of radiation that is absorbed by a person, and the biologic effect of the radiation. Commonly used terms are as follows:

- *Curie* (Ci) represents the disintegrations per second of a radionuclide (radioisotope); 1 Ci is equal to 3.7×10^{10} disintegrations per second. This is an expression of the amount of radiation emitted by a source.
- *Gray* (Gy) is a unit that expresses the energy absorbed by the target tissue per unit mass; 1 Gy corresponds to absorption of 10^4 erg/g of tissue. A centigray (cGy), which is the absorption of 100 erg/g of tissue, is equivalent to 100 Rad (radiation absorbed dose), abbreviated as R. The cGy terminology has now replaced the Rad in medical practice.
- *Sievert* (Sv) is a unit of equivalent dose that depends on the biologic rather than the physical effects of radiation (it replaced a unit called Rem). For the same absorbed dose, various types of radiation produce different amounts of damage. The equivalent dose controls for this variation and thereby provides a uniform measure of biologic dose. The equivalent dose (expressed in *Sieverts*) corresponds to the absorbed dose (expressed in Grays) multiplied by the relative biologic effectiveness of the radiation. The relative biologic effectiveness depends on the type of radiation, the type and volume of the exposed tissue, the duration of the exposure, and some other biologic factors (discussed next). The effective dose of x-rays in radiographs and CT is commonly expressed in milliSieverts (mSv). For x-radiation, 1 mSv = 1 mGy.

Main Determinants of Biologic Effects of Ionizing Radiation. In addition to the physical properties of radiation, its biologic effects depend heavily on the following factors.

- *Rate of delivery* significantly modifies the biologic effect. Although the effect of radiant energy is cumulative, divided doses may allow cells to repair some of the damage between exposures. Thus, fractionated doses of radiant energy have a cumulative effect only to the extent that repair during the "recovery" intervals is incomplete. Radiation therapy of tumors exploits the general capability of normal cells to repair themselves and recover more rapidly than tumor cells and thus not sustain as much cumulative radiation damage.
- *Field size* has a great influence on the consequences of irradiation. The body can sustain relatively high doses of radiation when delivered to small, carefully shielded fields, whereas smaller doses delivered to larger fields may be lethal.
- *Cell proliferation.* Because ionizing radiation damages DNA, rapidly dividing cells are more vulnerable to injury than quiescent cells (Fig. 9.15). Except at extremely high doses that impair DNA transcription, irradiation does not kill nondividing cells, such as neurons and muscle cells. However, as discussed in Chapter 7, in dividing cells DNA damage is detected by sensors that produce signals leading to the upregulation of p53, the "guardian of the genome." p53 in turn upregulates the expression of genes that initially lead to cell cycle arrest and, if the DNA damage is too great to be repaired, genes that cause cell death through apoptosis. Understandably, therefore, tissues with a high rate of cell division, such as gonads, bone marrow, lymphoid tissue, and the mucosa of the gastrointestinal tract, are extremely vulnerable to radiation, and the injury is manifested early after exposure.
- *Oxygen effects and hypoxia*. The production of ROS from reactions with free radicals generated by radiolysis of

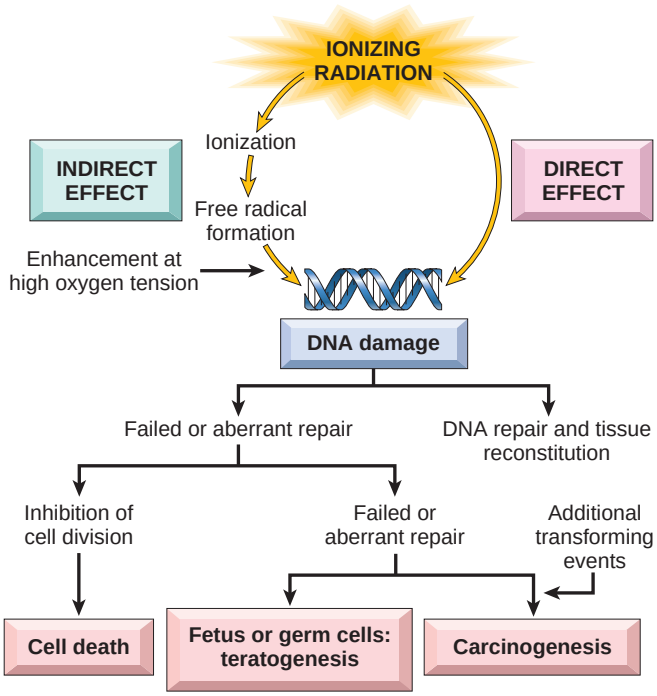


Figure 9.15 Effects of ionizing radiation on DNA and its consequences. The effects on DNA can be direct, or, more importantly, indirect, through free radical formation.

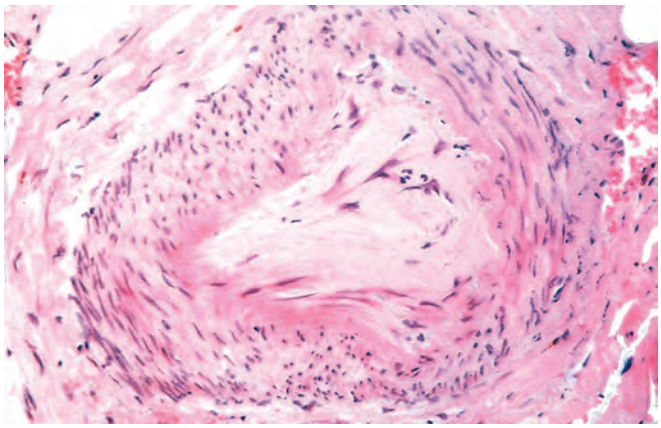


Figure 9.16 Radiation-induced chronic vascular injury with subintimal fibrosis occluding the lumen. (American Registry of Pathology © 1990.)

water is the major mechanism by which DNA is damaged by ionizing radiation. Poorly vascularized tissues with low oxygenation, such as the center of rapidly growing tumors, are generally less sensitive to radiation therapy than nonhypoxic tissues.

- *Vascular damage*. Damage to endothelial cells, which are moderately sensitive to radiation, may cause narrowing or occlusion of blood vessels leading to impaired healing, fibrosis, and chronic ischemic atrophy. These changes may appear months or years after exposure (Fig. 9.16).

Fig. 9.17 shows the overall consequences of radiation exposure. These consequences vary according to the dose of radiation and type of exposure. Table 9.7 lists the estimated threshold doses for acute effects of radiation aimed at specific organs; Table 9.8 lists the syndromes caused by exposure to various doses of total-body radiation.

MORPHOLOGY

Cells surviving radiant energy damage show a wide range of structural **changes in chromosomes** that are related to double-stranded DNA breaks including deletions, translocations, and fragmentation. The mitotic spindle often becomes disorderly, and polyploidy and aneuploidy may be encountered. **Nuclear swelling** and condensation and clumping of chromatin may appear; disruption of the nuclear membrane may also be noted. Apoptosis may occur. Several **abnormal nuclear morphologies** may be seen. Giant cells with pleomorphic nuclei or more than one nucleus may appear and persist for years after exposure. At extremely high doses of radiant energy, indicators of cell death, such as nuclear pyknosis and lysis, appear quickly.

In addition to affecting DNA and nuclei, radiant energy may induce a variety of **cytoplasmic changes** including swelling, mitochondrial distortion, and degeneration of the endoplasmic reticulum. Plasma membrane breaks and focal defects may be seen. The histologic constellation of cellular pleomorphism, giant-cell formation, conformational changes in nuclei, and abnormal mitotic figures creates a more than passing similarity between radiation-injured cells and cancer cells, a problem that plagues the pathologist

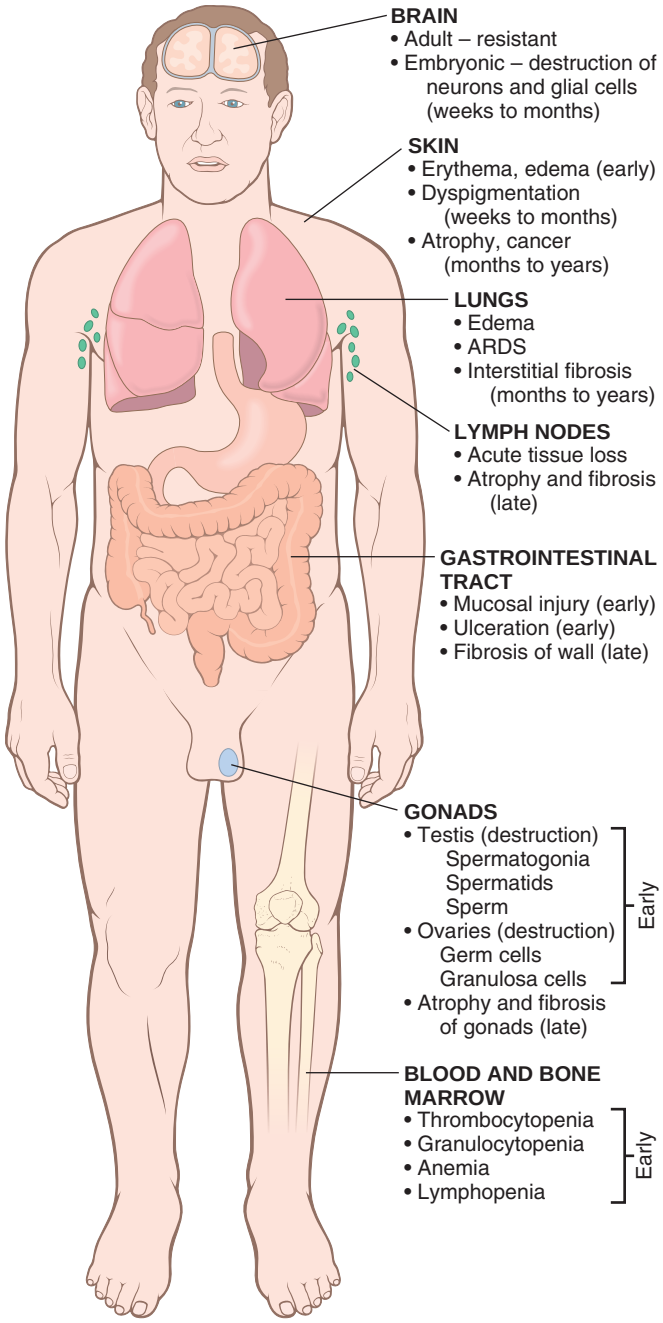


Figure 9.17 Overview of the major morphologic consequences of radiation injury. Early changes occur in hours to weeks; late changes occur in months to years. *ARDS*, Acute respiratory distress syndrome.

Table 9.7 Estimated Threshold Doses for Acute Radiation Effects on Specific Organs

Health Effect	Organ	Dose (Sv)
Temporary sterility	Testes	0.15
Depression of hematopoiesis	Bone marrow	0.50
Reversible skin effects (e.g., erythema)	Skin	1–2
Permanent sterility	Ovaries	2.5–6
Temporary hair loss	Skin	3–5
Permanent sterility	Testis	3.5
Cataract	Lens of eye	5

Table 9.8 Effects of Total-Body Ionizing Radiation

	0–1 Sv	1–2 Sv	2–10 Sv	10–20 Sv	>50 Sv
Main site of injury	None	Lymphocytes	Bone marrow	Small bowel	Brain
Main signs and symptoms	None	Moderate granulocytopenia, lymphopenia	Leukopenia, hemorrhage, hair loss, vomiting	Diarrhea, fever, electrolyte imbalance, vomiting	Ataxia, coma, convulsions, vomiting
Time of development	—	1 day to 1 week	2 to 6 weeks	5 to 14 days	1 to 4 hours
Lethality	None	None	Variable (0%–80%)	100%	100%

when evaluating irradiated tissues for the possible persistence of tumor cells.

Vascular changes and interstitial fibrosis are also prominent in irradiated tissues (Fig. 9.18). During the immediate postirradiation period, vessels may show only dilation. With time or with higher doses, a variety of degenerative changes appear, including endothelial cell swelling and vacuolation, or even necrosis and dissolution of the walls of small vessels such as capillaries and venules. Affected vessels may rupture or thrombose. Still later, endothelial cell proliferation and collagenous hyalinization and thickening of the intima are seen in irradiated vessels, resulting in marked narrowing or even obliteration of the vascular lumens. At this time, an increase in interstitial collagen in the irradiated field usually becomes evident, leading to scarring and contractions.

Total-Body Irradiation. Exposure of large areas of the body to even very small doses of radiation may have devastating effects. Doses below 1 Sv produce minimal symptoms, if any. However, higher levels of exposure cause health effects known as acute radiation syndromes, which at progressively higher doses involve the hematopoietic, gastrointestinal, and central nervous systems. The syndromes associated with total-body exposure to ionizing radiation are presented in Table 9.8.

Acute Effects on Hematopoietic and Lymphoid Systems. The hematopoietic and lymphoid systems are extremely susceptible to radiation injury and deserve special mention. With high dose levels and large exposure fields, severe lymphopenia appears within hours along with shrinkage of the lymph nodes and spleen. Radiation kills lymphocytes directly, both in the circulation and in tissues (nodes, spleen, thymus, gut). With sublethal doses of radiation, regeneration from viable precursors is prompt, leading to restoration of a normal blood lymphocyte count within weeks to months. Hematopoietic precursors in the bone marrow are also quite sensitive to radiant energy, which produces a dose-dependent marrow aplasia. The acute effects of marrow irradiation on peripheral blood counts reflect the kinetics of turnover of the formed elements, the granulocytes, platelets, and red cells, which have half-lives of less than a day (granulocytes), 10 days (platelets), and 120 days (red cells). After a brief rise in the circulating neutrophil count, neutropenia appears within several days. Neutrophil counts reach their nadir, often at counts near zero, during the second week. If the patient survives, recovery of a normal granulocyte count may require 2 to 3 months.

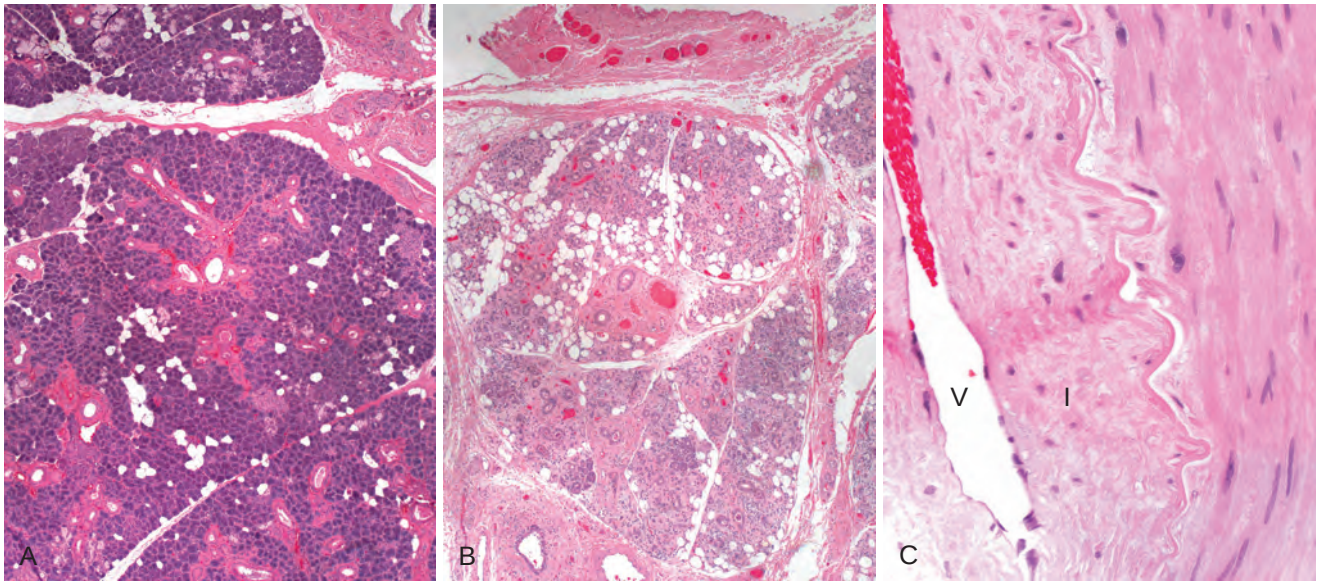


Figure 9.18 Fibrosis and vascular changes in salivary glands produced by radiation therapy of the neck region. (A) Normal salivary gland. (B) Fibrosis caused by radiation. (C) Fibrosis and vascular changes consisting of fibrointimal thickening and arteriolar sclerosis. *I*, Thickened intima; *V*, vessel lumen. (Courtesy Dr. Melissa Upton, Department of Pathology, University of Washington, Seattle, Wash.)

Thrombocytopenia appears by the end of the first week, with the platelet count nadir occurring somewhat later than that of granulocytes; recovery is similarly delayed. Anemia appears after 2 to 3 weeks and may persist for months. Understandably, higher doses of radiation produce more severe cytopenias and more prolonged periods of recovery. Very high doses kill hematopoietic stem cells and induce permanent aplasia (aplastic anemia) marked by a failure of blood count recovery, whereas with lower doses the aplasia is transient.

Fibrosis. A common consequence of radiation therapy for cancer is the development of fibrosis in the tissues included in the irradiated field (see Fig. 9.18). Fibrosis may occur weeks or months after irradiation as a consequence of the replacement of dead parenchymal cells by connective tissue, leading to the formation of scars and adhesions. Vascular damage, the death of tissue stem cells, and the release of cytokines and chemokines that promote inflammation and fibroblast activation are the main contributors to the development of radiation-induced fibrosis (Figs. 9.19 and 9.20). Common sites of fibrosis after radiation treatment are the lungs, the salivary glands after radiation therapy for head and neck cancers, and colorectal and pelvic areas after treatment for cancer of the prostate, rectum, or cervix.

DNA Damage and Carcinogenesis. Ionizing radiation can cause multiple types of DNA damage including single-base damage, single- and double-stranded breaks (DSBs), and DNA-protein cross-links. In surviving cells, simple defects may be repaired by various enzyme systems present in most mammalian cells. The most serious damage to DNA consists of DSBs. Two types of mechanisms can repair DSBs in mammalian cells: homologous recombination and nonhomologous end joining (NHEJ), with NHEJ being the most common repair pathway. DNA repair through NHEJ often produces mutations including short deletions or duplications or gross chromosomal aberrations such as translocations and inversions. If the replication of cells containing DSBs is not stopped by cell cycle checkpoint controls (Chapters 1 and 7), cells with chromosomal damage persist and may initiate carcinogenesis many years later.

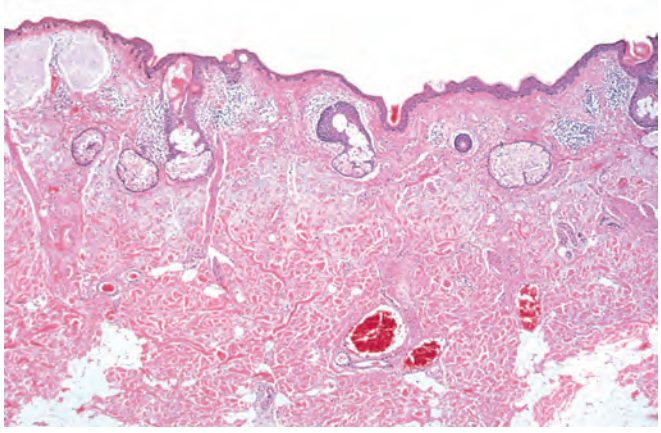


Figure 9.19 Chronic radiation dermatitis with atrophy of epidermis, dermal fibrosis, and telangiectasia of the subcutaneous blood vessels. (American Registry of Pathology © 1990.)

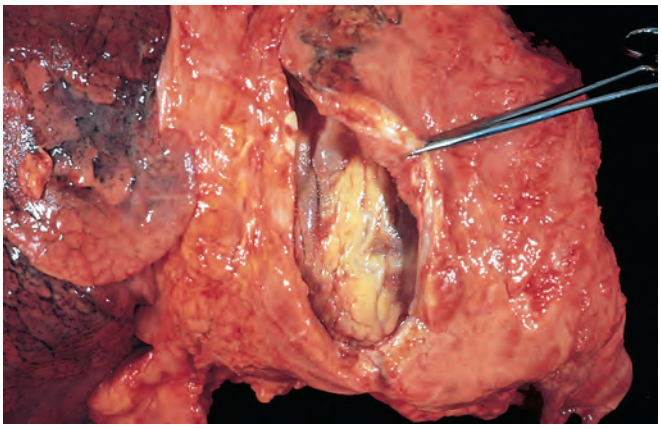


Figure 9.20 Extensive mediastinal fibrosis after radiotherapy for carcinoma of the lung. Note the markedly thickened pericardium. (From the Collection of the Department Pathology, Southwestern Medical School, Dallas, Tex.)

Cancer Risks From Exposures to Radiation. Any cell capable of division that has sustained a mutation has the potential to become cancerous. Thus, an increased incidence of neoplasms may occur in any organ after exposure to ionizing radiation. The level of radiation required to increase the risk of cancer development is difficult to determine, but there is little doubt that acute or prolonged exposures that result in doses of greater than 100 mSv cause serious consequences, including cancer. Proof of this risk is found in the increased incidence of leukemias and solid tumors in several organs (e.g., thyroid, breast, and lungs) in survivors of the atomic bombings of Hiroshima and Nagasaki; the high number of thyroid cancers in survivors of the Chernobyl accident; the high incidence of thyroid tumors and the elevated frequency of leukemias and birth defects in inhabitants of the Marshall Islands exposed to nuclear fallout; and the development of second cancers, such as acute myeloid leukemia, myelodysplastic syndrome, and solid tumors, in individuals who received radiation therapy for cancers such as Hodgkin lymphoma. The long-term cancer risks caused by radiation exposures in the range of 5 to 100 mSv are much more difficult to establish because accurate measurements of risks require large population groups ranging from 50,000 to 5 million people. Nevertheless, for x-rays and gamma rays there is good evidence for a statistically significant increase in the risk of cancer at acute doses of greater than 50 mSv and reasonable evidence for acute doses of greater than 5 mSv; as a point of reference, a single posteroanterior chest radiograph, a lateral chest film chest radiograph, and a CT scan of the chest deliver effective doses to the lungs of 0.01 mSv, 0.15 mSv, and 10 mSv. It is believed that the risk of secondary cancers following irradiation is greatest in children. This is based in part on a recent large-scale epidemiologic study showing that children who receive at least two CT scans have very small but measurable increased risks for leukemia and malignant brain tumors and on older studies showing that radiation therapy to the chest is particularly likely to produce breast cancers when administered to adolescent females.

Increased risk of cancer development may also be associated with occupational exposures. Radon gas is a ubiquitous product of the spontaneous decay of uranium. Its carcinogenic effects are largely attributable to two decay products, polonium 214 and polonium 218 (or "radon daughters"), which emit alpha particles. Polonium 214 and polonium 218 produced from inhaled radon tend to deposit in the lung, and chronic exposure in uranium miners may give rise to lung carcinomas. Risks are also present in homes in which the levels of radon are very high, comparable to those found in mines. It is suspected that lower levels of household radon may also contribute to lung cancer development, particularly in those who also smoke tobacco.

KEY CONCEPTS

RADIATION INJURY

- Ionizing radiation may injure cells directly or indirectly by generating free radicals from water or molecular oxygen.
- Ionizing radiation damages DNA; therefore rapidly dividing cells such as germ cells and those in the bone marrow and gastrointestinal tract are very sensitive to radiation injury.
- DNA damage that is not adequately repaired may result in mutations that predispose affected cells to neoplastic transformation.
- Ionizing radiation may cause vascular damage and sclerosis, resulting in ischemic necrosis of parenchymal cells and their replacement by fibrous tissue.

NUTRITIONAL DISEASES

Malnutrition is a consequence of inadequate intake of proteins and calories or deficiencies in the digestion or absorption of proteins resulting in the loss of fat and muscle mass, weight loss, and generalized weakness. Millions of people in lower income nations are malnourished or living on the cruel edge of starvation. In the higher income world and, more recently, also in some lower income countries, obesity has become a major public health problem due to its association with diseases such as diabetes, atherosclerosis, and cancer.

The sections that follow barely skim the surface of nutritional disorders. Particular attention is devoted to childhood malnutrition, anorexia nervosa and bulimia, deficiencies of vitamins and trace minerals, obesity, and the relationships of diet to cancer and atherosclerosis. Other nutrients and nutritional issues are discussed in the context of specific diseases.

Dietary Insufficiency

An appropriate diet should provide (1) sufficient energy, in the form of carbohydrates, fats, and proteins, for the body's daily metabolic needs; (2) amino acids and fatty acids to be used as building blocks for synthesis of proteins and lipids; and (3) vitamins and minerals, which function as coenzymes or hormones in vital metabolic pathways or, as in the case of calcium and phosphate, as important structural components. In *primary malnutrition,* one or all of these components are missing from the diet. By contrast, *secondary malnutrition* results from malabsorption, impaired utilization or storage, excess loss, or increased need for nutrients.

There are several conditions that may lead to primary or secondary malnutrition.

- *Poverty.* Homeless persons, aged individuals, and children of the poor often suffer from severe malnutrition as well as trace nutrient deficiencies. In lower income countries, poverty, crop failures, livestock deaths, and drought, often in times of war and political upheaval, create the setting for the malnourishment of children and adults.
- *Acute and chronic illnesses*. The basal metabolic rate rises in many illnesses, resulting in increased daily requirements for all nutrients. Failure to recognize these nutritional needs may delay recovery. Malnutrition is often present in patients with wasting diseases, such as advanced cancers and AIDS that are complicated by cachexia (Chapter 7).
- *Chronic alcoholism*. Alcoholic persons may sometimes suffer from malnutrition but more frequently have deficiencies of vitamins, especially thiamine, pyridoxine, folate, and vitamin A, as a result of poor diet, defective gastrointestinal absorption, abnormal nutrient utilization and storage, increased metabolic needs, and an increased rate of loss. A failure to recognize the likelihood of thiamine deficiency in persons with chronic alcoholism may result in irreversible brain damage (e.g., *Wernicke encephalopathy* and *Korsakoff psychosis,* discussed in Chapter 28).
- *Ignorance and failure of diet supplementation.* Even the affluent may fail to recognize that infants, adolescents, and pregnant women have increased nutritional needs. Ignorance about the nutritional content of foods is also a contributing factor. Some examples are iron deficiency in infants fed exclusively artificial milk diets; thiamine deficiency in diets in which polished rice is the mainstay; and lack of iodine from food and water in regions removed from the oceans, unless supplementation is provided.
- *Self-imposed dietary restriction.* Anorexia nervosa, bulimia, and less overt eating disorders affect individuals who are concerned about body image and are obsessed with body weight (anorexia and bulimia are discussed later).
- *Other causes.* Additional causes of malnutrition include gastrointestinal diseases and malabsorption syndromes, genetic diseases, specific drug therapies (which block uptake or utilization of particular nutrients), and inadequate total parenteral nutrition.

Severe Acute Malnutrition

The WHO defines severe acute malnutrition (SAM) as a state characterized by a weight for height ratio that is 3 standard deviations below the normal range. Worldwide about 50 million children are affected by SAM. It is common in low income countries, where as many as 25% of children may be affected and where it is a major contributor to the high death rates among the very young. In addition to loss of life, wars also exact a heavy toll on refugees who live in abject poverty. In recent years, camps set up for refugees from Syria, as many as 20% of the children are severely or moderately malnourished.

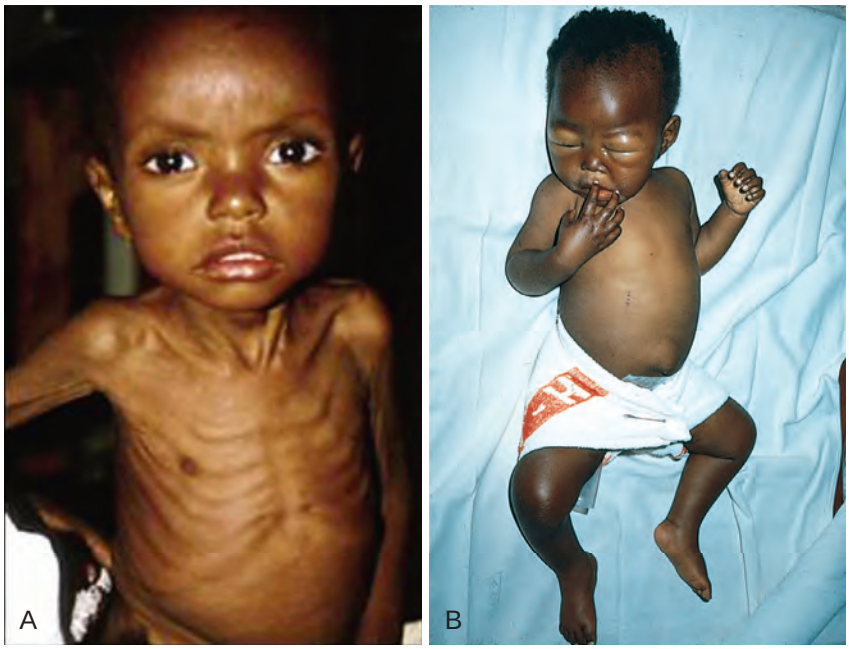


Figure 9.21 Childhood malnutrition. (A) Marasmus. Note the loss of muscle mass and subcutaneous fat; the head appears to be too large for the emaciated body. (B) Kwashiorkor. The infant shows generalized edema, seen as ascites and puffiness of the face, hands, and legs. (A, From Clinic Barak, Reisebericht, Kenya.)

SAM, previously called protein energy malnutrition (PEM), manifests as a spectrum of clinical syndromes, all resulting from a dietary intake of protein and calories that is inadequate to meet the body's needs. At two ends of SAM are *marasmus* and *kwashiorkor*. It should be noted that from a functional standpoint, there are two protein compartments in the body: the somatic compartment, represented by proteins in skeletal muscles, and the visceral compartment, represented by protein stores in the visceral organs, primarily the liver. These two compartments are regulated differently, as detailed subsequently. The somatic compartment is affected more severely in marasmus, and the visceral compartment is depleted more severely in kwashiorkor. Clinical assessment of undernutrition is discussed next, followed by descriptions of marasmus and kwashiorkor.

The diagnosis of SAM is obvious in its most severe forms. In mild to moderate forms, the usual approach is to compare the body weight for a given height against standard tables; other helpful parameters are fat stores, muscle mass, and levels of certain serum proteins. With a loss of fat, measured skinfold thickness (which includes skin and subcutaneous tissue) is reduced. If the somatic protein compartment is catabolized, the resultant reduction in muscle mass is reflected by reduced circumference of the midarm. Measurement of serum proteins (albumin, transferrin, and others) provides an estimate of the adequacy of the visceral protein compartment. Recent studies suggest a role for the gut microbiome in the pathogenesis of SAM. There is a substantial difference in the microbial flora of children with SAM when compared with the gut microbiome of normal children. It seems that the alterations in the microbiome are not merely the consequences of SAM but play a role in their causation. The most compelling evidence for this notion comes from fecal transplants from children with SAM into germ-free mice. Malnutrition was induced in host mice by fecal transplants from affected but not well-nourished children.

Marasmus. **Marasmus develops when the diet is severely lacking in calories** (Fig. 9.21A). A child is considered to have marasmus when weight falls to 60% of normal for sex, height, and age. A marasmic child suffers growth retardation and loss of muscle mass as a result of catabolism and depletion of the somatic protein compartment. This seems to be an adaptive response that provides the body with amino acids as a source of energy. Of interest, the visceral protein compartment, which presumably is more critical for survival, is depleted only marginally, so serum albumin levels are either normal or only slightly reduced. In addition to muscle proteins, subcutaneous fat is also mobilized and used as fuel. Leptin (discussed later under "Obesity") production is low, which may stimulate the hypothalamic-pituitary-adrenal axis to produce high levels of cortisol that contribute to lipolysis. Due to losses of muscle and subcutaneous fat, the extremities are emaciated; by comparison, the head appears too large for the body. Anemia and manifestations of multivitamin deficiencies are present, and there is evidence of immune deficiency, particularly of T-cell–mediated immunity. Hence, concurrent infections are usually present, which impose additional stress on a weakened body.

Kwashiorkor. **Kwashiorkor occurs when protein deprivation is relatively greater than the reduction in total calories** (Fig. 9.21B). This is the most common form of SAM seen in African children, who have been weaned too early and subsequently fed, almost exclusively, a carbohydrate diet (the name kwashiorkor, from the Ga language in Ghana, describes the illness in a young child that appears after the

arrival of another baby). The prevalence of kwashiorkor also is high in lower income countries of Southeast Asia. Less severe forms occur worldwide in persons with chronic diarrheal states, in which protein is not absorbed, or in persons with chronic protein loss (e.g., protein-losing enteropathies, the nephrotic syndrome, or the aftermath of extensive burns). Rare cases of kwashiorkor resulting from fad diets or replacement of milk by rice-based beverages have been reported in the United States.

In kwashiorkor, unlike in marasmus, marked protein deprivation is associated with severe loss of the visceral protein compartment, and the resultant hypoalbuminemia gives rise to generalized or dependent edema (see Fig. 9.21). The weight of children with severe kwashiorkor typically is 60% to 80% of normal. However, the true loss of weight is masked by the increased fluid retention (edema). In further contrast with marasmus, there is relative sparing of subcutaneous fat and muscle mass. The modest loss of these compartments may also be masked by edema.

Children with kwashiorkor have characteristic skin lesions with alternating zones of hyperpigmentation, desquamation, and hypopigmentation, giving a "flaky paint" appearance. Hair changes include loss of color or alternating bands of pale and darker color, straightening, fine texture, and loss of firm attachment to the scalp. Other features that distinguish kwashiorkor from marasmus include an enlarged, fatty liver (resulting from reduced synthesis of the carrier protein component of lipoproteins) and the development of apathy, listlessness, and loss of appetite. Like marasmus, vitamin deficiencies are likely to be present, as are defects in immunity and secondary infections, which produce inflammation and a catabolic state that aggravates the malnutrition. As already mentioned, marasmus and kwashiorkor represent two ends of a spectrum, and considerable overlap exists.

Malnutrition in the Upper Income World. In the United States, secondary malnutrition often develops in chronically ill, older, and bedridden patients. It is estimated that more than 50% of older residents in nursing homes in the United States are malnourished. The most obvious signs of secondary malnutrition include (1) depletion of subcutaneous fat in the arms, chest wall, shoulders, or metacarpal regions; (2) wasting of the quadriceps and deltoid muscles; and (3) ankle or sacral edema. Bedridden or hospitalized malnourished patients have an increased risk of infection, sepsis, impaired wound healing, and death after surgery.

MORPHOLOGY

The main anatomic changes in SAM are (1) growth failure; (2) peripheral edema in kwashiorkor; and (3) loss of body fat and atrophy of muscle, more marked in marasmus.

The **liver** in kwashiorkor, but not in marasmus, is enlarged and fatty; superimposed cirrhosis is rare. In kwashiorkor (rarely in marasmus) the **small bowel** shows a decrease in mitotic cells in the crypts of the glands, associated with mucosal atrophy and loss of villi and microvilli. In such cases concurrent loss of small intestinal enzymes occurs, most often manifested as disaccharidase deficiency. Hence infants with kwashiorkor initially may not respond well to full-strength, milk-based diets. With treatment, the mucosal changes are reversible.

The **bone marrow** in both kwashiorkor and marasmus may be hypoplastic, mainly as a result of decreased numbers of red cell precursors. The peripheral blood commonly reveals mild to moderate anemia, which is often multifactorial in origin; nutritional deficiencies of iron, folate, and protein, as well as the suppressive effects of infection (anemia of chronic inflammation) may all contribute. Depending on the predominant factor, the red cells may be microcytic, normocytic, or macrocytic.

The **brain** in infants who are born to malnourished mothers and who suffer SAM during the first 1 or 2 years of life has been reported by some to show cerebral atrophy, a reduced number of neurons, and impaired myelinization of white matter.

Many other changes may be present including (1) thymic and lymphoid atrophy (more marked in kwashiorkor than in marasmus); (2) anatomic alterations induced by intercurrent infections, particularly with all manner of endemic worms and other parasites; and (3) deficiencies of other required nutrients such as iodine and vitamins.

Cachexia. Secondary malnutrition is a common complication in patients with acquired immunodeficiency syndrome (AIDS) or advanced cancers, and in these settings it is known as cachexia. Because of its common association with cancer, cachexia is discussed in Chapter 7.

Anorexia Nervosa and Bulimia

Anorexia nervosa is self-induced starvation, resulting in marked weight loss; bulimia is a condition in which the patient binges on food and then induces vomiting. Anorexia nervosa has the highest death rate of any psychiatric disorder. Bulimia is more common than anorexia nervosa and generally has a better prognosis; it is estimated to occur in 1% to 2% of women and 0.1% of men, with an average onset at 20 years of age. These eating disorders occur primarily in previously healthy young women who have developed an obsession with body image and thinness. The neurobiologic underpinnings of these diseases are unknown, but it has been suggested that altered serotonin metabolism may be an important component.

The clinical findings in anorexia nervosa are generally similar to those in SAM. In addition, effects on the endocrine system are prominent. *Amenorrhea,* resulting from decreased secretion of gonadotropin-releasing hormone and (as a result) luteinizing hormone and follicle-stimulating hormone, is so common that its presence is considered a diagnostic feature. Other common findings related to *decreased thyroid hormone release* include cold intolerance, bradycardia, constipation, and changes in the skin and hair. In addition, dehydration and electrolyte abnormalities are frequently present. The skin becomes dry and scaly. Increased fat in the marrow (paradoxically, since fat is decreased elsewhere) associated with a peculiar deposition of mucinous matrix material that is referred to as *gelatinous transformation* is virtually pathognomonic for anorexia nervosa (Fig. 9.22). The bone density is decreased, most likely because of low estrogen levels, mimicking the postmenopausal acceleration of osteoporosis. Anemia, lymphopenia, and hypoalbuminemia may

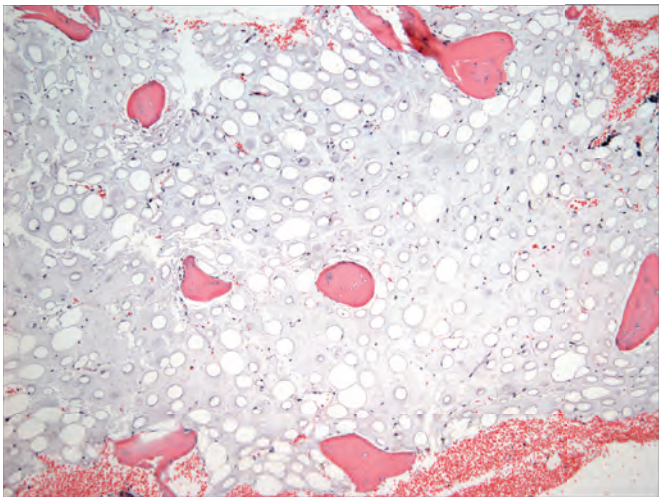


Figure 9.22 Anorexia nervosa. Increased fat in the marrow associated with deposition of mucinous matrix material (gelatinous transformation) is pathognomonic of this disease.

be present. A major complication of anorexia nervosa (and bulimia) is an increased susceptibility to *cardiac arrhythmia* and *sudden death,* resulting from hypokalemia.

In bulimia, *binge eating* is the norm. Large amounts of food, principally carbohydrates, are ingested, followed by induced vomiting. Although menstrual irregularities are common, amenorrhea occurs in less than 50% of bulimic patients because weight and gonadotropin levels remain near normal. The major medical complications relate to frequent vomiting and the chronic use of laxatives and diuretics. They include (1) electrolyte imbalances (hypokalemia), which predispose the patient to cardiac arrhythmias; (2) pulmonary aspiration of gastric contents; and (3) esophageal and gastric rupture. Nevertheless, there are no specific signs or symptoms; thus the diagnosis of bulimia relies on a comprehensive psychologic assessment.

Vitamin Deficiencies

Thirteen vitamins are necessary for health; vitamins A, D, E, and K are fat-soluble, and all others are water-soluble. The distinction between fat-soluble and water-soluble vitamins is important. Fat-soluble vitamins are more readily stored in the body, but they may be poorly absorbed in fat malabsorption disorders, caused by disturbances of digestive functions (Chapter 17). Certain vitamins can be synthesized endogenously—vitamin D from precursor steroids; vitamin K and biotin by the intestinal microflora; and niacin from tryptophan, an essential amino acid. Notwithstanding this endogenous synthesis, a dietary supply of all vitamins is essential for health.

A deficiency of vitamins may be primary (dietary in origin) or secondary to disturbances in intestinal absorption, transport in the blood, tissue storage, or metabolic conversion. In the following sections, vitamins A, D, and C are presented in some detail because of their wide-ranging activities and the morphologic changes of deficient states. This is followed by presentation in tabular form of the main consequences of deficiencies of the remaining vitamins (E, K, and the B complex) and some essential minerals. However, it should be emphasized that deficiency of a single vitamin is uncommon, and that single or multiple vitamin deficiencies may be associated with SAM.

Vitamin A

The major functions of vitamin A are maintenance of normal vision, regulation of cell growth and differentiation, and regulation of lipid metabolism. Vitamin A is the name given to a group of related compounds that include *retinol*, *retinal*, and *retinoic acid*, which have similar biologic activities. The generic term *retinoids* encompasses vitamin A in its various forms and both natural and synthetic chemicals that are structurally related to vitamin A, but may not necessarily have vitamin A–like biologic activity. Animal-derived foods such as liver, fish, eggs, milk, and butter are important dietary sources of preformed vitamin A. Yellow and leafy green vegetables such as carrots, squash, and spinach supply large amounts of carotenoids, provitamins that can be metabolized to active vitamin A in the body. Carotenoids contribute approximately 30% of the vitamin A in human diets; the most important of these is β-carotene, which is efficiently converted to vitamin A. The Recommended Dietary Allowance for vitamin A is expressed in retinol equivalents, to take into account both preformed vitamin A and β-carotene.

Vitamin A is a fat-soluble vitamin, and its absorption requires bile, pancreatic enzymes, and some level of antioxidant activity in the food. Retinol (generally ingested as retinol ester) and β-carotene are absorbed in the intestine, where β-carotene is also converted to retinol (Fig. 9.23). Retinol is then transported in chylomicrons to the liver for esterification and storage. Uptake in liver cells takes place through the apolipoprotein E receptor. More than 90% of the body's vitamin A reserves are stored in the liver, predominantly in the perisinusoidal stellate (Ito) cells. In healthy persons who consume an adequate diet, these reserves are sufficient to meet the body's demands for at least 6 months. Retinol esters stored in the liver can be mobilized; before release, retinol binds to a specific retinol-binding protein (RBP), synthesized in the liver. The uptake of retinol/RBP in peripheral tissues is dependent on cell surface receptors specific for RBP. Once taken up into peripheral tissues, retinol may also be stored as retinol ester or may be oxidized to form retinoic acid, which has important effects on epithelial differentiation and growth.

Function. In humans, the main functions of vitamin A are the following:

- *Maintenance of normal vision.* The visual process involves four forms of vitamin A–containing pigments: rhodopsin in the rods, the most light-sensitive pigment and therefore important in reduced light, and three iodopsins in cone cells, each responsive to specific colors in bright light. The synthesis of rhodopsin from retinol involves (1) oxidation to all-*trans*-retinal, (2) isomerization to 11-*cis*-retinal, and (3) covalent association with the 7-transmembrane rod protein opsin to form rhodopsin. A photon of light causes the isomerization of 11-*cis*-retinal to all-*trans*-retinal, which dissociates from rhodopsin. This induces a conformational change in opsin, triggering a series of downstream events that generate a nerve impulse, which is transmitted

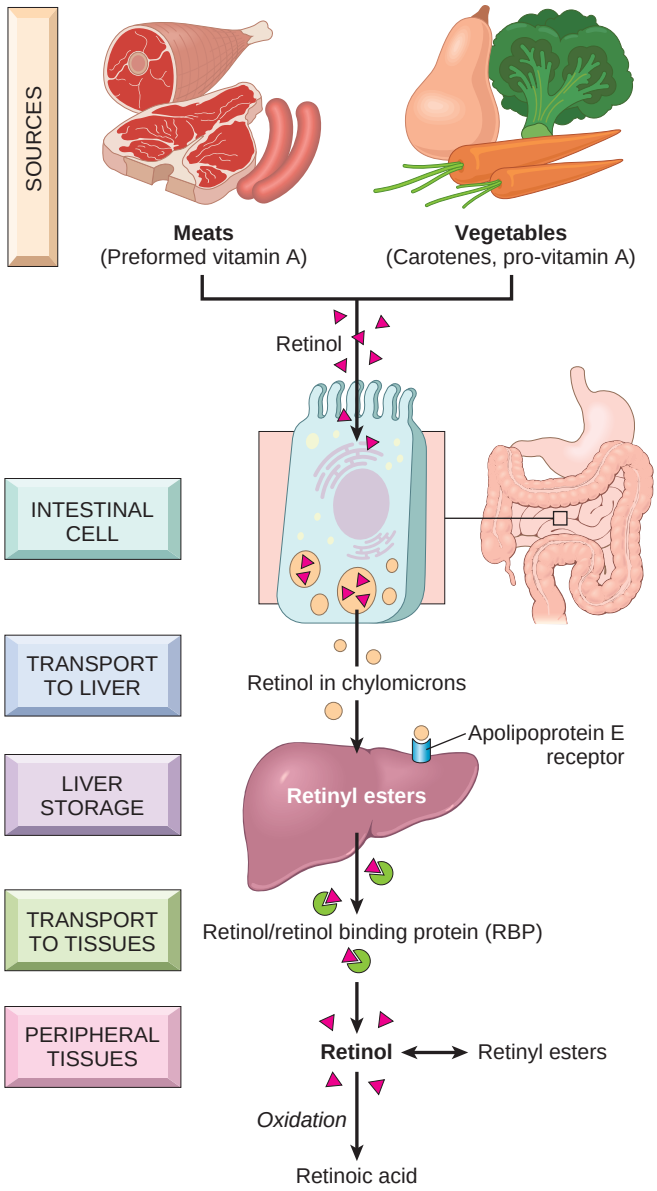


Figure 9.23 Vitamin A metabolism.

via neurons from the retina to the brain. During dark adaptation, some of the all-*trans*-retinal is reconverted to 11-*cis*-retinal, but most is reduced to retinol and lost to the retina, dictating the need for continuous supply.

- *Cell growth and differentiation.* Vitamin A and retinoids play an important role in the orderly differentiation of mucus-secreting epithelium; when a deficiency state exists, the epithelium undergoes squamous metaplasia, differentiating into a keratinizing epithelium. Activation of retinoic acid receptors (RARs) by their ligands causes the release of corepressors and the obligatory formation of heterodimers with another retinoid receptor, *RXR*. Both RAR and RXR have three isoforms, α, β, and γ. The RAR/RXR heterodimers bind to retinoic acid response elements located in the regulatory regions of genes that encode receptors for growth factors, tumor suppressor genes, and secreted proteins. Through these effects, retinoids regulate cell growth and differentiation, cell cycle control, and other biologic responses. All-*trans*-retinoic acid, a potent acid derivative of vitamin A, has the highest affinity for RARs compared with other retinoids.
- *Metabolic effects of retinoids.* The RXR, believed to be activated by 9-*cis*-retinoic acid, can form heterodimers with other nuclear receptors, such as nuclear receptors involved in drug metabolism, peroxisome proliferator-activated receptors (PPARs), and vitamin D receptors. PPARs are key regulators of lipid metabolism including fatty acid oxidation in fat cells and muscle, adipogenesis, and lipoprotein metabolism. The association of RXR and PPARγ thus explains the metabolic effects of retinoids on adipogenesis.
- *Enhancing immunity to infections.* Vitamin A supplementation can reduce morbidity and mortality rates from diarrhea by 15% and 28%, respectively. The effects of vitamin A on infections probably derive in part from its ability to stimulate the immune system through unclear mechanisms. Infections may reduce the bioavailability of vitamin A, possibly by inducing the acute phase response, which appears to inhibit RBP synthesis in the liver. The drop in hepatic RBP causes a decrease in circulating retinol, which reduces the tissue availability of vitamin A. In addition, retinoids, β-carotene, and some related carotenoids function as photoprotective and antioxidant agents.

Retinoids are used clinically in the treatment of skin disorders such as severe acne and certain forms of psoriasis, as well as in the treatment of acute promyelocytic leukemia. As discussed in Chapter 7, all-*trans*-retinoic acid induces the differentiation and subsequent apoptosis of acute promyelocytic leukemia cells through its ability to bind to a PML-RARα fusion protein that characterizes this form of cancer. A different isomer, 13-*cis*-retinoic acid, has been used with some success in the treatment of childhood neuroblastoma.

Vitamin A Deficiency. Vitamin A deficiency occurs worldwide either as a consequence of primary malnutrition or secondary to conditions that cause malabsorption of fats. In children, stores of vitamin A are depleted by infections, and the absorption of the vitamin is poor in newborn infants. Adult patients with malabsorption syndromes such as celiac disease, Crohn disease and colitis may develop vitamin A deficiency in conjunction with depletion of other fat-soluble vitamins. Bariatric surgery and, in older persons, continuous use of mineral oil as a laxative may lead to deficiency. The pathologic effects of vitamin A deficiency are summarized in Fig. 9.24.

As already discussed, vitamin A is a component of rhodopsin and other visual pigments. Not surprisingly, one of the earliest manifestations of vitamin A deficiency is impaired vision, particularly in reduced light (night blindness). Other effects of deficiency are related to the role of vitamin A in regulating the differentiation of epithelial cells. Persistent deficiency gives rise to epithelial metaplasia and keratinization. The most devastating changes occur in the eyes and are referred to as xerophthalmia (dry eye). First, there is dryness of the conjunctiva (xerosis conjunctivae) as the normal lacrimal and mucus-secreting epithelium is replaced by keratinized epithelium. This is followed by a

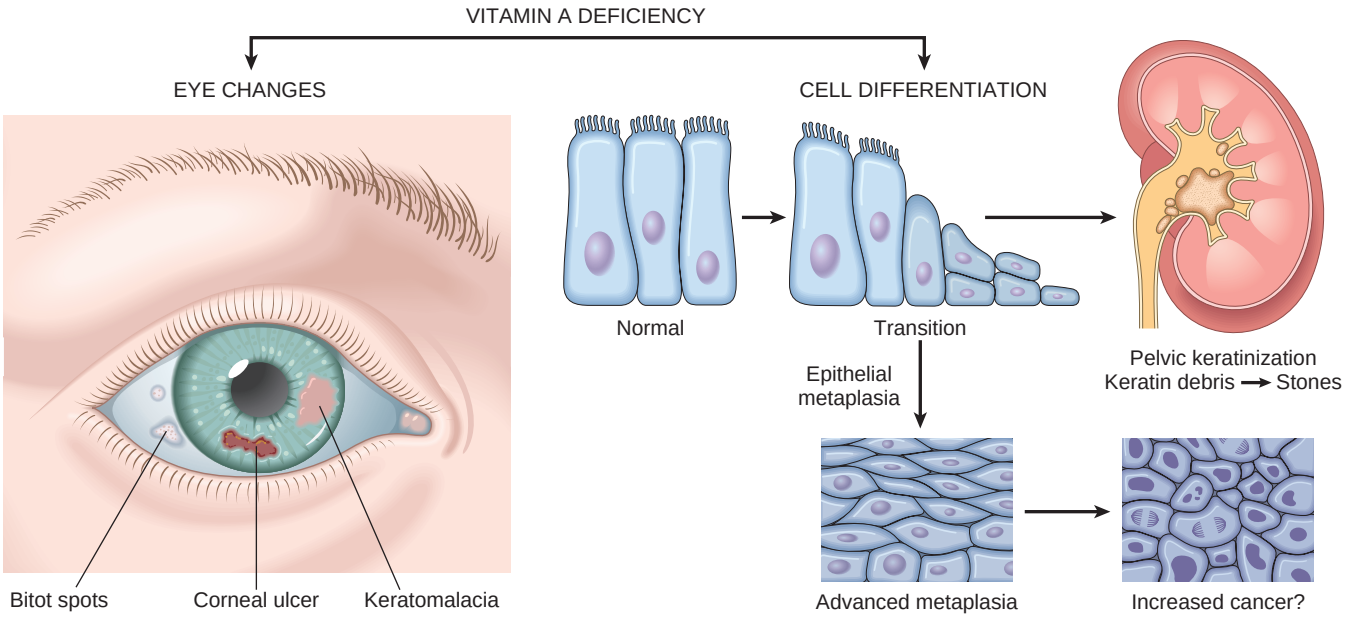


Figure 9.24 Vitamin A deficiency, its major consequences in the eye and in the production of keratinizing metaplasia of specialized epithelial surfaces, and its possible role in epithelial metaplasia. Not depicted are night blindness and immune deficiency.

buildup of keratin debris in small opaque plaques *(Bitot spots)* that progresses to erosion of the roughened corneal surface, softening and destruction of the cornea (keratomalacia), and blindness.

In addition to the ocular epithelium, the epithelium lining the upper respiratory passage and urinary tract also undergoes squamous metaplasia. Loss of the mucociliary epithelium of the airways predisposes to secondary pulmonary infections, and desquamation of keratin debris in the urinary tract predisposes to renal and urinary bladder stones. Hyperplasia and hyperkeratinization of the epidermis and plugging of the ducts of the adnexal glands may produce follicular or papular dermatosis. Another serious consequence is immune deficiency, which is responsible for higher mortality rates from common infections such as measles, pneumonia, and infectious diarrhea. In parts of the world where deficiency of vitamin A is prevalent, dietary supplements reduce mortality by 20% to 30%.

Vitamin A Toxicity. Both short-term and long-term excesses of vitamin A may produce toxic manifestations. The consequences of acute hypervitaminosis A were first described by Gerrit de Veer in 1597, a ship's carpenter stranded in the Arctic, who recounted in his diary the serious symptoms that he and other members of the crew developed after eating polar bear liver. With this cautionary tale in mind, the adventurous eater should be aware that acute vitamin A toxicity has also been described in individuals who ingested the livers of whales, sharks, and even tuna.

The symptoms of acute vitamin A toxicity include headache, dizziness, vomiting, stupor, and blurred vision, symptoms that may be confused with those of a brain tumor *(pseudotumor cerebri)*. Chronic toxicity is associated with weight loss, anorexia, nausea, vomiting, and bone and joint pain. Retinoic acid stimulates osteoclast production and activity, leading to increased bone resorption and high risk of fractures. Although synthetic retinoids used for the treatment of acne are not associated with these types of conditions, their use in pregnancy should be avoided because of the well-established teratogenic effects of retinoids.

Vitamin D

The major function of vitamin D is the maintenance of adequate plasma levels of calcium and phosphorus to support metabolic functions, bone mineralization, and neuromuscular transmission. Vitamin D is a fat-soluble vitamin required for the prevention of bone diseases known as *rickets* (in children whose epiphyses have not already closed) and *osteomalacia* (in adults), as well as *hypocalcemic tetany*. With respect to tetany, vitamin D maintains the correct concentration of ionized calcium in the extracellular fluid compartment. When deficiency develops, the drop in ionized calcium in the extracellular fluid results in continuous excitation of muscle (tetany). It should be noted, however, that any reduction in the level of serum calcium is usually corrected by increased secretion of parathyroid hormone followed by bone resorption; hence, tetany is quite uncommon. Our attention here is focused on the function of vitamin D in the regulation of serum calcium levels.

Metabolism of Vitamin D. **The major source of vitamin D for humans is its endogenous synthesis from a precursor, 7-dehydrocholesterol, in a photochemical reaction that requires solar or artificial UV light in the range of 290 to 315 nm (UVB radiation).** This reaction results in the synthesis of *cholecalciferol,* known as *vitamin* D_3. Herein, the term vitamin D is used to refer to this compound. Under usual conditions of sun exposure, about 90% of the required vitamin D is endogenously synthesized by the skin. However, individuals with dark skin generally have a lower level of vitamin D production because of melanin pigmentation. Dietary sources, such as deep-sea fish, plants, and grains, contribute the remaining required vitamin D and depend

on adequate intestinal fat absorption. In plants, vitamin D is present in a precursor form (ergosterol), which is converted to vitamin D in the body.

The main steps of vitamin D metabolism are summarized as follows:

1. Photochemical synthesis of vitamin D from 7-dehydrocholesterol in the skin and absorption of vitamin D from foods and supplements in the gut
2. Binding of vitamin D from both of these sources to plasma α_1-globulin (D-binding protein [DBP]) and transport into the liver
3. Conversion of vitamin D into 25-hydroxycholecalciferol (25-OH-D) in the liver, through the action of 25-hydroxylases, including CYP27A1 and other CYPs
4. Conversion of 25-OH-D into 1,25-dihydroxyvitamin D [$1\alpha,25(OH)_2D_3$], the most active form of vitamin D, by the enzyme 1α-hydroxylase in the kidney

The production of 1,25-dihydroxyvitamin D in the kidney is regulated by three main mechanisms (Fig. 9.25).

- *Hypocalcemia stimulates secretion of parathyroid hormone (PTH),* which in turn augments the conversion of

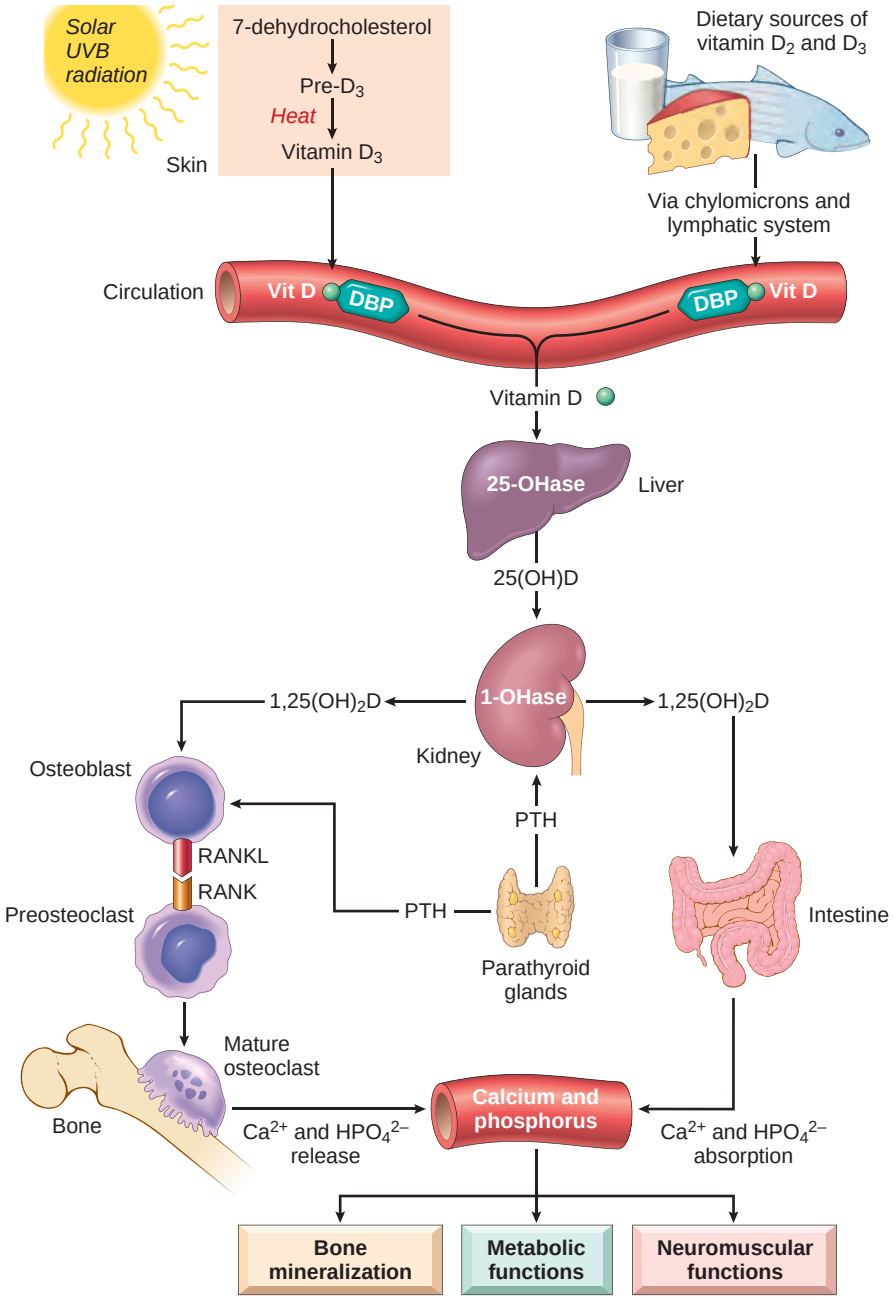


Figure 9.25 Vitamin D metabolism. Vitamin D is produced from 7-dehydrocholesterol in the skin or is ingested in the diet. It is converted in the liver into 25(OH)D and in the kidney into 1,25-dihydroxyvitamin D (*1,25(OH)$_2$D*), the active form of the vitamin. 1,25(OH)$_2$D stimulates the expression of RANKL, an important regulator of osteoclast maturation and function, on osteoblasts and enhances the intestinal absorption of calcium and phosphorus. *DBP,* Vitamin D–binding protein (α_1-globulin).

25-OH-D into 1,25-dihydroxyvitamin D by upregulating the expression of 1α-hydroxylase.
- *Hypophosphatemia also upregulates 1α-hydroxylase expression,* increasing the production of 1,25-dihydroxyvitamin D.
- *Through a feedback mechanism,* increased levels of 1,25-dihydroxyvitamin D downregulate its own synthesis through inhibition of 1α-hydroxylase activity.

Functions. Like retinoids and steroid hormones, 1,25-dihydroxyvitamin D acts by binding to a high-affinity nuclear receptor (vitamin D receptor), which associates with the already mentioned RXR. This heterodimeric complex binds to vitamin D response elements located in the regulatory sequences of vitamin D target genes. The receptors for 1,25-dihydroxyvitamin D are present in most cells of the body. In the small intestine, bones, and kidneys, signals transduced via these receptors regulate plasma levels of calcium and phosphorus. Beyond its role on skeletal homeostasis, vitamin D has immunomodulatory and antiproliferative effects. 1,25-Dihydroxyvitamin D also appears to act through mechanisms that do not require the transcription of target genes. These alternative mechanisms involve the binding of 1,25-dihydroxyvitamin D to a membrane-associated vitamin D receptor (mVDR), leading to the activation of protein kinase C and opening of calcium channels.

Effects of Vitamin D on Calcium and Phosphorus Homeostasis. The main functions of 1,25-dihydroxyvitamin D on calcium and phosphorus homeostasis are the following:
- *Stimulation of intestinal calcium absorption.* 1,25-Dihydroxyvitamin D stimulates intestinal absorption of calcium in the duodenum through the interaction of 1,25-dihydroxyvitamin D with nuclear vitamin D receptor and the formation of a complex with RXR. The complex binds to vitamin D response elements and activates the transcription of TRPV6 (a member of the transient receptor potential vanilloid family), which encodes a critical calcium transport channel.
- *Stimulation of calcium reabsorption in the kidney.* 1,25-Dihydroxyvitamin D increases calcium influx in distal tubules of the kidney through the increased expression of TRPV5, another member of the transient receptor potential vanilloid family. TRPV5 expression is also regulated by PTH in response to hypocalcemia.
- *Interaction with PTH in the regulation of blood calcium.* Vitamin D maintains calcium and phosphorus at supersaturated levels in the plasma. The parathyroid glands have a key role in the regulation of extracellular calcium concentrations. These glands have a calcium receptor that senses even small changes in blood calcium concentrations. In addition to their effects on calcium absorption in the intestine and kidneys already described, both 1,25-dihydroxyvitamin D and PTH enhance the expression of RANKL (receptor activator of NF-κB ligand) on osteoblasts. RANKL binds to its receptor (RANK) located in preosteoclasts, thereby inducing the differentiation of these cells into mature osteoclasts (Chapter 26). Through the secretion of hydrochloric acid and activation of proteases such as cathepsin K, osteoclasts dissolve bone and release calcium and phosphorus into the circulation.
- *Mineralization of bone.* Vitamin D contributes to the mineralization of osteoid matrix and epiphyseal cartilage in both flat and long bones. It stimulates osteoblasts to synthesize the calcium-binding protein osteocalcin, which is involved in the deposition of calcium during bone development. Flat bones develop by intramembranous bone formation, in which mesenchymal cells differentiate directly into osteoblasts, which synthesize the collagenous osteoid matrix on which calcium is deposited. Long bones develop by endochondral ossification, through which growing cartilage at the epiphyseal plates is provisionally mineralized and then progressively resorbed and replaced by osteoid matrix that is mineralized to create bone (Fig. 9.26A).

When *hypocalcemia* occurs due to vitamin D deficiency (Fig. 9.27), PTH production is elevated, causing (1) activation of renal 1α-hydroxylase, increasing the amount of active vitamin D and calcium absorption; (2) increased resorption of calcium from bone by osteoclasts; (3) decreased renal calcium excretion; and (4) increased renal excretion of phosphate. Although a normal serum level of calcium may be restored, hypophosphatemia persists, impairing the mineralization of bone. Increased production of FGF-23 may be responsible for tumor-induced osteomalacia and some forms of hypophosphatemic rickets.

Deficiency States. The normal reference range for circulating 25-(OH)-D is 20 to 100 ng/mL; concentrations of less than 20 ng/mL constitute vitamin D deficiency.

Rickets in growing children and osteomalacia in adults are skeletal diseases with worldwide distributions. They may result from diets deficient in calcium and vitamin D, but an equally important cause of vitamin D deficiency is limited exposure to sunlight. This most often affects inhabitants of northern latitudes, but can even be a problem in tropical countries, in heavily veiled women, and in children born to mothers who have frequent pregnancies followed by lactation. In all of these situations, vitamin D deficiency can be prevented by a diet high in fish oils. Other, less common causes of rickets and osteomalacia include renal disorders causing decreased synthesis of 1,25-dihydroxyvitamin D, phosphate depletion, malabsorption disorders, and some rare inherited disorders. Although rickets and osteomalacia rarely occur outside high-risk groups, milder forms of vitamin D deficiency leading to an increased risk of bone loss and hip fractures are quite common in older adults in the United States and Europe. Some genetically determined variants of the vitamin D receptors are also associated with accelerated loss of bone minerals during aging and certain familial forms of osteoporosis (Chapter 26).

MORPHOLOGY

Vitamin D deficiency in both rickets and osteomalacia results in an **excess of unmineralized matrix.** The following sequence ensues in rickets:
- Overgrowth of epiphyseal cartilage due to inadequate provisional calcification and failure of the cartilage cells to mature and disintegrate
- Persistence of distorted, irregular masses of cartilage that project into the marrow cavity

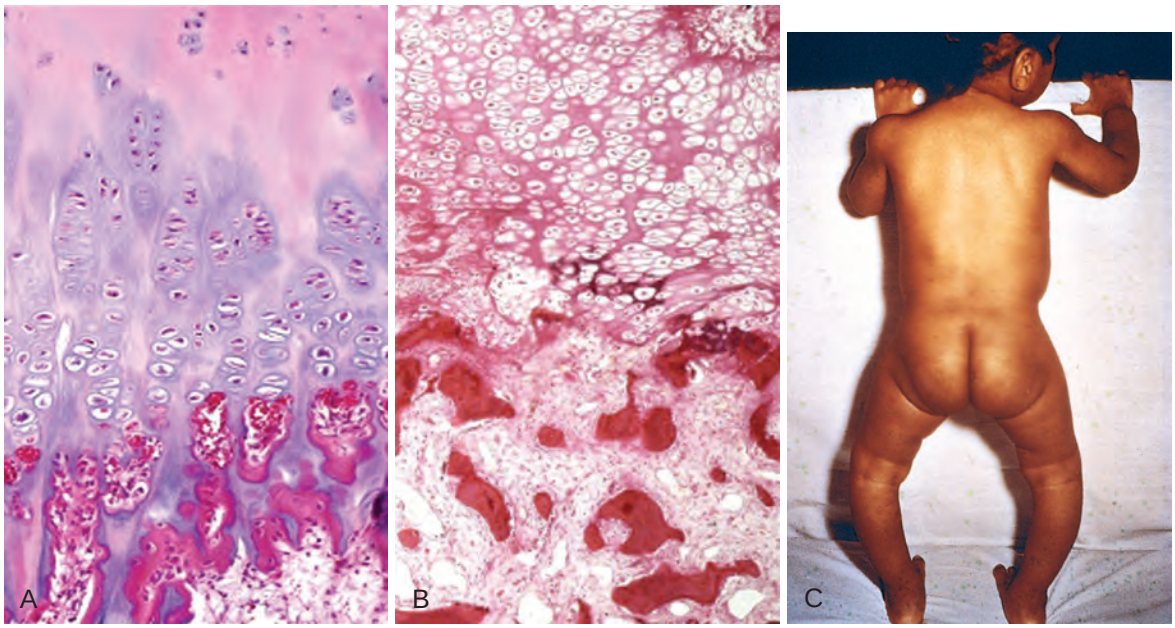


Figure 9.26 Rickets. (A) Normal costochondral junction of a young child illustrating formation of cartilage palisades and orderly transition from cartilage to new bone. (B) Detail of a rachitic costochondral junction in which the palisades of cartilage is lost. Darker trabeculae are well-formed bone; paler trabeculae consist of uncalcified osteoid. (C) Rickets: note bowing of legs due to formation of poorly mineralized bones. (B, Courtesy Dr. Andrew E. Rosenberg, Massachusetts General Hospital, Boston, Mass.)

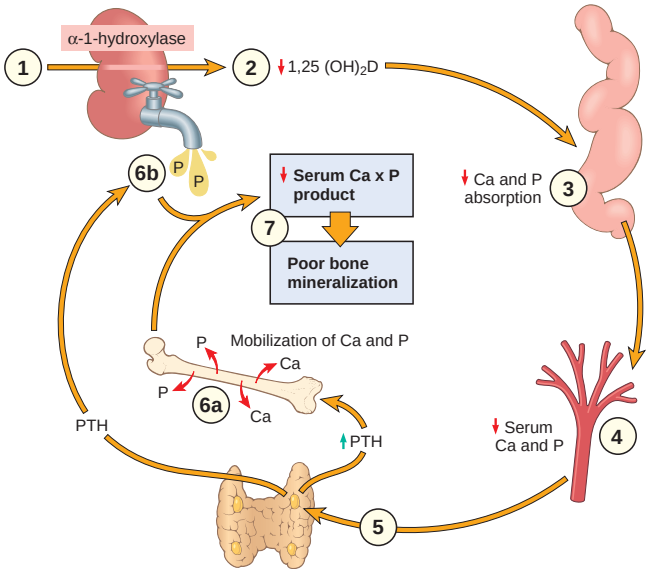


Figure 9.27 Vitamin D deficiency. There is inadequate substrate for renal 1α-hydroxylase *(1)*, yielding a deficiency of $1,25(OH)_2D$ *(2)*, and deficient absorption of calcium *(Ca)* and phosphorus *(P)* from the gut *(3)*, with consequently depressed serum levels of both *(4)*. The hypocalcemia activates the parathyroid glands *(5)*, causing mobilization of calcium and phosphorus from bone *(6a)*. Simultaneously, the parathyroid hormone *(PTH)* induces wasting of phosphate in the urine *(6b)* and calcium retention. As a result, the serum levels of calcium are normal or nearly normal, but phosphate levels are low; hence mineralization is impaired *(7)*.

- Deposition of osteoid matrix on inadequately mineralized cartilaginous remnants
- Disruption of the orderly replacement of cartilage by osteoid matrix, with enlargement and lateral expansion of the osteochondral junction (see Fig. 9.26B)
- Abnormal overgrowth of capillaries and fibroblasts in the disorganized zone resulting from microfractures and stresses on the inadequately mineralized, weak, poorly formed bone
- Deformation of the skeleton due to the loss of structural rigidity of the developing bones

The gross skeletal changes in rickets depend on the severity and duration of the process and, in particular, the stresses to which individual bones are subjected. During the nonambulatory stage of infancy, the head and chest sustain the greatest stresses. The softened occipital bones may become flattened, and the parietal bones can be buckled inward by pressure; with the release of the pressure, elastic recoil snaps the bones back into their original positions **(craniotabes).** An excess of osteoid produces **frontal bossing** and a **squared appearance to the head**. Deformation of the chest results from overgrowth of cartilage or osteoid tissue at the costochondral junction, producing the **rachitic rosary.** The weakened metaphyseal areas of the ribs are subject to the pull of the respiratory muscles and thus bend inward, creating anterior protrusion of the sternum **(pigeon breast deformity).** When an ambulating child develops rickets, deformities are likely to affect the spine, pelvis, and tibia, causing **lumbar lordosis** and **bowing of the legs** (see Fig. 9.26C).

In adults with **osteomalacia,** the lack of vitamin D deranges the normal bone remodeling that occurs throughout life. The newly formed osteoid matrix laid down by osteoblasts is inadequately mineralized, thus producing the excess of persistent osteoid that is characteristic of osteomalacia. Although the contours of the bone are not affected, the bone is weak and vulnerable to

gross fractures or microfractures, which are most likely to affect vertebral bodies and femoral necks. The unmineralized osteoid appears as a thickened layer of matrix (which stains pink in hematoxylin and eosin preparations) arranged about the more basophilic, normally mineralized trabeculae.

Nonskeletal Effects of Vitamin D. As mentioned earlier, the vitamin D receptor is present in various cells and tissues that do not participate in calcium and phosphorus homeostasis. In addition, macrophages, keratinocytes, and tissues such as breast, prostate, and colon can produce 1,25-dihydroxyvitamin D. Within macrophages, synthesis of 1,25-dihydroxyvitamin D occurs through the activity of CYP27B located in the mitochondria. It appears that pathogen-induced activation of Toll-like receptors in macrophages causes increased expression of vitamin D receptor and CYP27B, leading to local synthesis of 1,25-dihydroxyvitamin D and activation of vitamin D–dependent gene expression in macrophages and other neighboring immune cells. The net effect of this altered gene expression on the immune response remains to be determined, however, and clinical trials have failed to demonstrate beneficial effects of vitamin D supplements on the course of respiratory infections, including tuberculosis.

Vitamin D Toxicity. Prolonged exposure to normal sunlight does not produce an excess of vitamin D, but megadoses of orally administered vitamin can lead to hypervitaminosis. In children, hypervitaminosis D may take the form of metastatic calcifications of soft tissues such as the kidney; in adults, it causes bone pain and hypercalcemia. The toxic potential of this vitamin is so great that in sufficiently large doses it is a potent rodenticide.

Vitamin C (Ascorbic Acid)

A deficiency of water-soluble vitamin C leads to the development of *scurvy*, characterized principally by bone disease in growing children and by hemorrhages and healing defects in both children and adults. Sailors of the British Royal Navy were nicknamed "limeys" because at the end of the 18th century the Navy began to provide lime and lemon juice (rich sources of vitamin C) to sailors to prevent scurvy during their long sojourn at sea. It was not until 1932 that ascorbic acid was identified and synthesized. Ascorbic acid is not synthesized endogenously in humans; therefore we are entirely dependent on the diet for this nutrient. Vitamin C is present in milk and some animal products (liver, fish) and is abundant in a variety of fruits and vegetables. All but the most restricted diets provide adequate amounts of vitamin C.

Function. Ascorbic acid has many functions affecting a variety of processes:

- *Collagen synthesis*. The best-established function of vitamin C is the *activation of prolyl and lysyl hydroxylases* from inactive precursors, providing for *hydroxylation of procollagen*. Inadequately hydroxylated procollagen cannot acquire a stable helical configuration, so it is poorly secreted from the fibroblast. Those molecules that are secreted are inadequately cross-linked, lack tensile strength, and are more soluble and vulnerable to enzymatic degradation. Collagen, which normally has the highest content of hydroxyproline of any polypeptide, is most affected, particularly in blood vessels, accounting for the predisposition to hemorrhages in scurvy.
- *Neurotransmitter synthesis*. Synthesis of norepinephrine requires hydroxylation of dopamine, a step that requires vitamin C.
- *Antioxidant functions.*
- *Modulating the immune response.*

The effect on the latter two has formed the basis of clinical trials based on supplementation of vitamin C in sepsis.

Deficiency States. Consequences of vitamin C deficiency (scurvy) are illustrated in Fig. 9.28. Because of the abundance

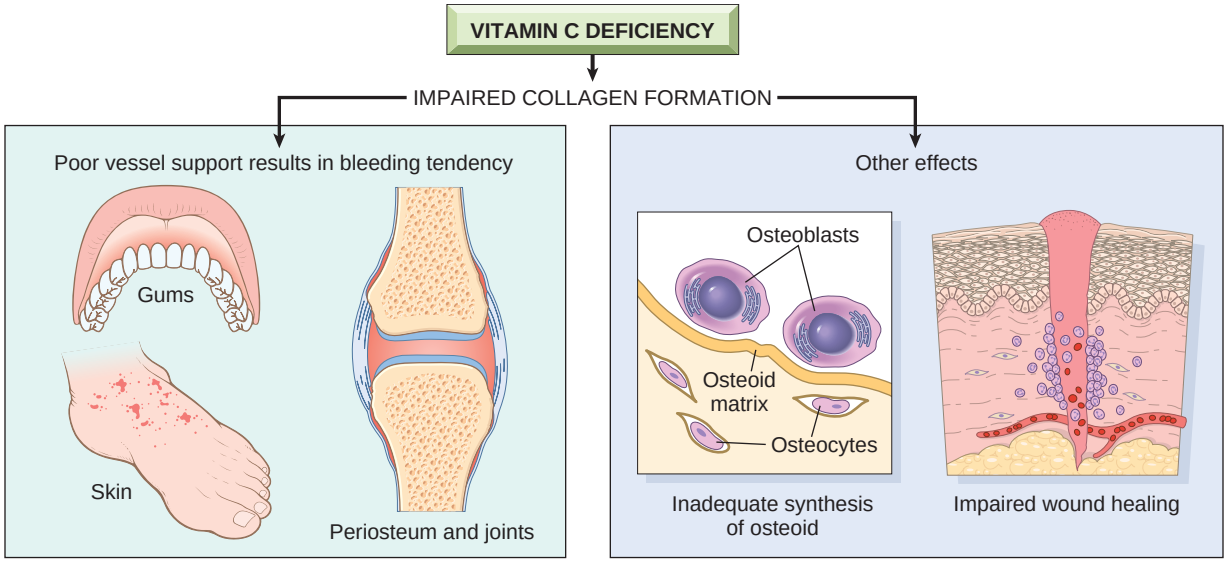


Figure 9.28 Major consequences of vitamin C deficiency caused by impaired formation of collagen.

of ascorbic acid in many foods, scurvy has ceased to be a global problem. It is sometimes encountered even in affluent populations as a secondary deficiency, particularly among older individuals, persons who live alone, and chronic alcoholics, groups that often have erratic and inadequate eating patterns. Occasionally, scurvy occurs in patients undergoing peritoneal dialysis and hemodialysis and in food faddists. The condition also sometimes appears in infants who are maintained on formulas of evaporated milk without supplementation of vitamin C.

Vitamin C Excess. The popular notion that megadoses of vitamin C protect against the common cold, or at least allay the symptoms, has not been borne out by controlled clinical studies. Such slight relief as may be experienced is probably due to the mild antihistamine action of ascorbic acid. Similarly, there is no evidence that large doses of vitamin C protect against cancer development. The physiologic availability of excess vitamin C is limited due to its inherent instability, poor intestinal absorption, and rapid urinary excretion. Fortunately, toxicities related to high doses of vitamin C are rare, consisting of possible iron overload (due to increased absorption), hemolytic anemia in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency (Chapter 14), and calcium oxalate kidney stones.

Other vitamins and some essential minerals are listed and briefly described in Tables 9.9 and 9.10 and are discussed in other chapters.

KEY CONCEPTS

NUTRITIONAL DISEASES

- SAM is a common cause of childhood deaths in low income countries. The two main primary forms of SAM syndromes are marasmus and kwashiorkor. Secondary malnutrition occurs in the chronically ill and in patients with advanced cancer (as a result of cachexia).
- Kwashiorkor is characterized by hypoalbuminemia, generalized edema, fatty liver, skin changes, and defects in immunity. It is caused by diets low in protein but normal in calories.
- Marasmus is characterized by emaciation resulting from loss of muscle mass and fat with relative preservation of serum albumin. It is caused by diets severely lacking in both protein and nonprotein calories.
- Anorexia nervosa is self-induced starvation; it is characterized by amenorrhea and multiple manifestations of low thyroid hormone levels. Bulimia is a condition in which food binges alternate with induced vomiting.
- Vitamins A and D are fat-soluble vitamins with a wide range of activities. Vitamin A is required for vision, epithelial differentiation, and immune function. Vitamin D is a key regulator of calcium and phosphate homeostasis.
- Vitamin C and members of the vitamin B family are water-soluble. Vitamin C is needed for collagen synthesis and collagen cross-linking and tensile strength. B vitamins have diverse roles in cellular metabolism.

Table 9.9 Vitamins: Major Functions and Deficiency Syndromes

Vitamin	Functions	Deficiency Syndromes
Fat-Soluble		
Vitamin A	A component of visual pigment Maintenance of specialized epithelia Maintenance of resistance to infection	Night blindness, xerophthalmia, blindness Squamous metaplasia Vulnerability to infection, particularly measles
Vitamin D	Facilitates intestinal absorption of calcium and phosphorus and mineralization of bone	Rickets in children Osteomalacia in adults
Vitamin E	Major antioxidant; scavenges free radicals	Spinocerebellar degeneration, hemolytic anemia
Vitamin K	Cofactor in hepatic carboxylation of procoagulants—factors II (prothrombin), VII, IX, and X and protein C and protein S	Bleeding diathesis (Chapter 14)
Water-Soluble		
Vitamin B_1 (thiamine)	As pyrophosphate, is coenzyme in decarboxylation reactions	Dry and wet beriberi, Wernicke syndrome, Korsakoff syndrome (Chapter 28)
Vitamin B_2 (riboflavin)	Converted to coenzymes flavin mononucleotide and flavin adenine dinucleotide, cofactors for many enzymes in intermediary metabolism	Ariboflavinosis, cheilosis, stomatitis, glossitis, dermatitis, corneal vascularization
Niacin	Incorporated into nicotinamide adenine dinucleotide (NAD) and NAD phosphate, involved in a variety of redox reactions	Pellagra—"three Ds": dementia, dermatitis, diarrhea
Vitamin B_6 (pyridoxine)	Derivatives serve as coenzymes in many intermediary reactions	Cheilosis, glossitis, dermatitis, peripheral neuropathy (Chapter 28) Maintenance of myelinization of spinal cord tracts
Vitamin B_{12}	Required for normal folate metabolism and DNA synthesis	Megaloblastic pernicious anemia and degeneration of posterolateral spinal cord tracts (Chapter 14)
Vitamin C	Serves in many oxidation-reduction (redox) reactions and hydroxylation of collagen	Scurvy
Folate	Essential for transfer and use of one-carbon units in DNA synthesis	Megaloblastic anemia, neural tube defects (Chapter 14)
Pantothenic acid	Incorporated in coenzyme A	No nonexperimental syndrome recognized
Biotin	Cofactor in carboxylation reactions	No clearly defined clinical syndrome

Table 9.10 Selected Trace Elements and Deficiency Syndromes

Element	Function	Basis of Deficiency	Clinical Features
Zinc	Component of enzymes, principally oxidases	Inadequate supplementation in artificial diets Interference with absorption by other dietary constituents Inborn error of metabolism	Rash around eyes, mouth, nose, and anus called acrodermatitis enteropathica Anorexia and diarrhea Growth retardation in children Depressed mental function Depressed wound healing and immune response Impaired night vision Infertility
Iron	Essential component of hemoglobin as well as several iron-containing metalloenzymes	Inadequate diet Chronic blood loss	Hypochromic microcytic anemia (Chapter 14)
Iodine	Component of thyroid hormone	Inadequate supply in food and water	Goiter and hypothyroidism (Chapter 24)
Copper	Component of cytochrome *c* oxidase, dopamine β-hydroxylase, tyrosinase, lysyl oxidase, and unknown enzymes involved in cross-linking collagen	Inadequate supplementation in artificial diet Interference with absorption	Muscle weakness Neurologic defects Abnormal collagen cross-linking
Fluoride	Mechanism unknown	Inadequate supply in soil and water Inadequate supplementation	Dental caries (Chapter 16)
Selenium	Component of glutathione peroxidase Antioxidant with vitamin E	Inadequate amounts in soil and water	Myopathy Cardiomyopathy (Keshan disease)

Obesity

Obesity is defined as an accumulation of adipose tissue that is of sufficient magnitude to impair health. How does one measure fat accumulation? Several high-tech methods have been devised, but for practical purposes the *body mass index* (BMI) is most commonly used. BMI is calculated as (weight in kilograms)/(height in meters)^2, or kg/m^2. The normal BMI range is 18.5 to 25 kg/m^2, although the range differs for different countries due to differences in ethnicity and genetic backgrounds. Individuals with BMI greater than 30 kg/m^2 are classified as obese; individuals with BMI between 25 kg/m^2 and 30 kg/m^2 are considered overweight. Unless otherwise noted, the term obesity herein applies to both truly obese and overweight individuals.

Obesity is associated with several of the most important diseases of humans, including type 2 diabetes, dyslipidemia, cardiovascular disease, hypertension, and cancer. The strength of this association is affected not only by the quantity of excess fat but also by its distribution. Central, or visceral, obesity, in which excess fat accumulates preferentially in the trunk and in the abdominal cavity (in the mesentery and around viscera), is associated with a much higher risk for several diseases than is excess accumulation of subcutaneous fat.

Obesity is a major public health problem in higher income countries and an emerging health problem in lower income nations, such as India. According to the WHO, >1.9 billion adults were estimated to be overweight or obese globally in 2015, of which 650 million were obese. In the United States, obesity has reached epidemic proportions. In 2016 the prevalence of obesity in men was 37.9%, and the prevalence of obesity in women was 41.1%. Equally troubling is the prevalence of obesity in children and adolescents, approximately one-third of whom are obese. The increase in obesity in the United States has been associated with the higher caloric content of the diet, mostly caused by increased consumption of refined sugars, sweetened beverages, and vegetable oils.

The etiology of obesity is complex and incompletely understood. Genetic, environmental, and psychologic factors are involved. However, simply put, obesity is a disorder of energy homeostasis. **The two sides of the energy equation, intake and expenditure, are finely regulated by neural and hormonal mechanisms so that body weight is maintained within a narrow range for many years.** Apparently, this fine balance is controlled by an internal set-point, or "lipostat," that senses the quantity of energy stores (adipose tissue) and appropriately regulates food intake as well as energy expenditure. The hypothalamus is the master regulator of energy homeostasis. It receives inputs from the periphery about the state of energy stores; if they are inadequate, it triggers anabolic circuits, and if they are adequate, catabolic circuits are activated. The effect of the anabolic circuits is to increase intake of food and reduce energy expenditure, whereas catabolic circuits reduce food intake and increase energy expenditure. The neurohumoral mechanisms that regulate energy balance can be subdivided into three components (Figs. 9.29 and 9.30).

- The *peripheral or afferent system* generates signals from various sites. Its main components are leptin produced by fat cells, ghrelin from the stomach, peptide YY (PYY) and glucagon-like peptide 1 (GLP-1) from the ileum and colon, and insulin from the pancreas. The afferent systems provide signals to the central processing system in the brain.
- *The central processing system resides in the arcuate nucleus of the hypothalamus* where neurohumoral peripheral signals are integrated to generate efferent signals. Two sets of neurons participate in central processing (see Fig. 9.30).
  - A pair of first-order neurons: (1) pro-opiomelanocortin *(POMC)* and cocaine and amphetamine-regulated transcript *(CART)* neurons and (2) neurons containing neuropeptide Y *(NPY)* and agouti-related peptide

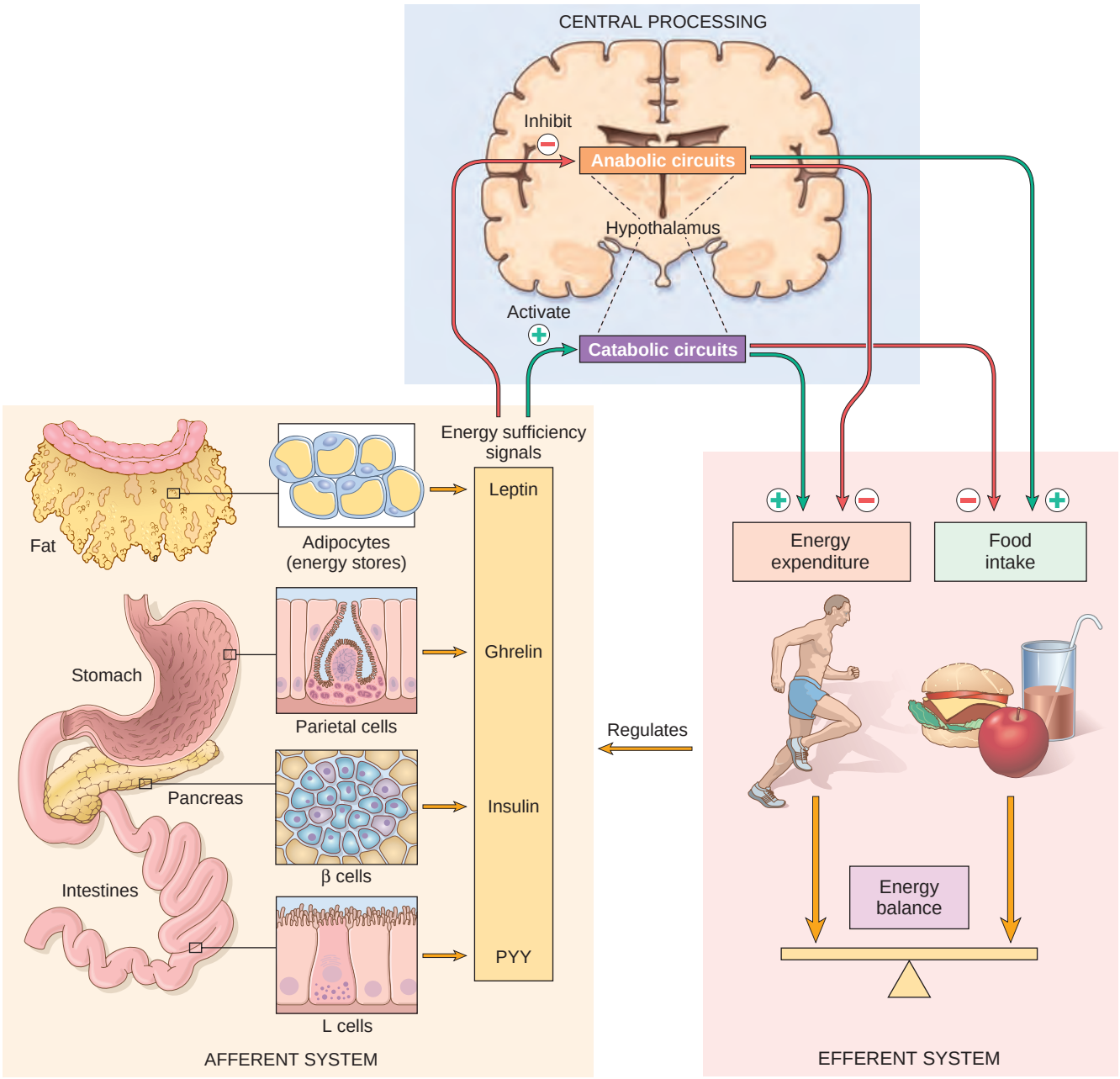


Figure 9.29 Energy balance regulatory circuitry. When sufficient energy is stored in adipose tissue and the individual is well fed, afferent signals (insulin, leptin, ghrelin, peptide YY) are delivered to the central neuronal processing units in the hypothalamus. Here these signals inhibit anabolic circuits and activate catabolic circuits. The effector arms of these central circuits then influence energy balance by inhibiting food intake and promoting energy expenditure. This in turn reduces the energy stores, and energy sufficiency signals are blunted. Conversely, when energy stores are low, the anabolic circuits take over, at the expense of catabolic circuits, to generate energy stores in the form of adipose tissue.

(AgRP). These first-order neurons communicate with second-order neurons.

- A pair of second-order neurons: (1) neurons that bear melanocortin receptors 3 and 4 (MC3/4R) and receive signals from first-order *POMC/CART* neurons and (2) neurons that bear Y1 and Y5 receptors and receive signals from first-order *NPY/AgRP* neurons.

- The *efferent system* consists of signals generated by second-order neurons and is organized along two pathways, catabolic (downstream of MC3/4R) and anabolic (downstream of Y1 and Y5 receptors) that control food intake and energy expenditure. In addition to these circuits (within the hypothalamus), the hypothalamic nuclei also communicate with forebrain and midbrain centers that control the autonomic nervous system.

With this background on the organization of the hypothalamic centers that regulate energy balance, we can now discuss how they function. Upon nutrient intake, POMC is cleaved from *POMC/CART* neurons and gives rise to α-melanocyte–stimulating hormone (MSH), which activates MC3/4R receptors in second-order neurons. These

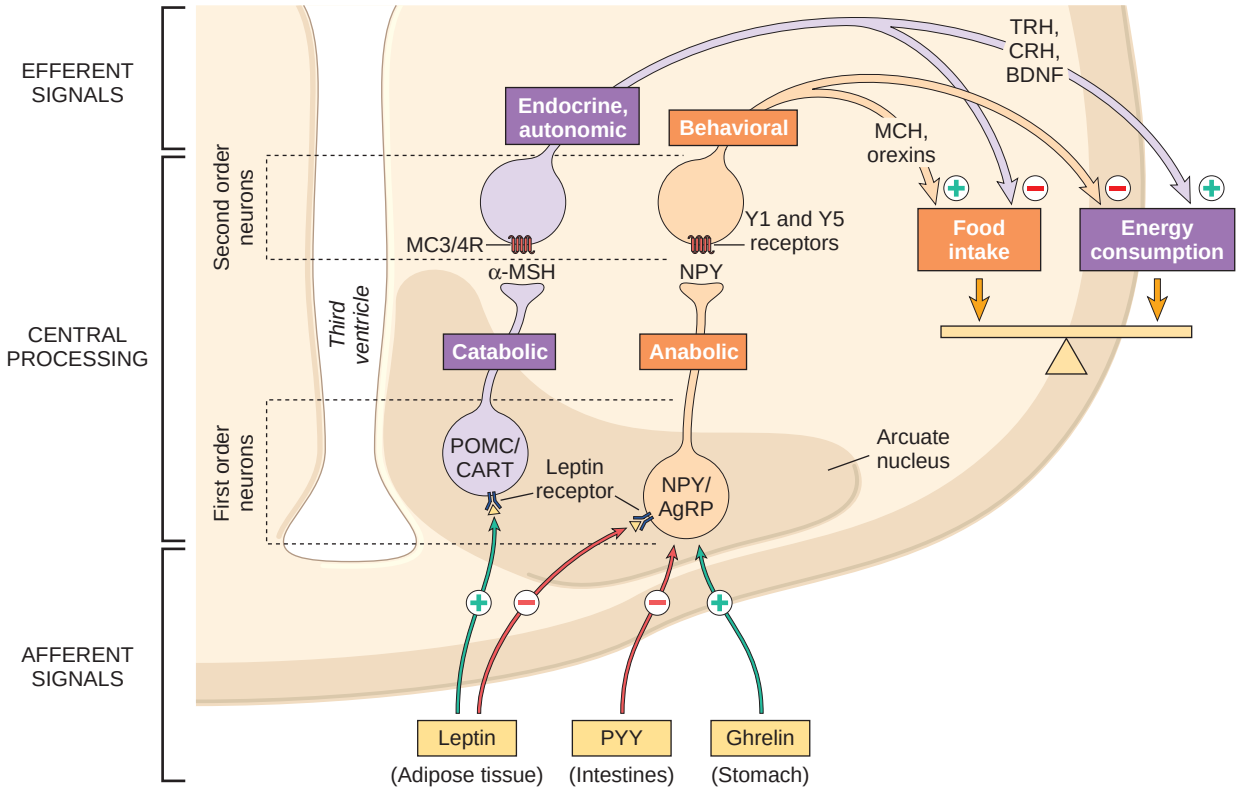


Figure 9.30 Neurohumoral circuits in the hypothalamus that regulate energy balance. Shown are POMC/CART anorexigenic neurons and NPY/AgRP orexigenic neurons in the arcuate nucleus of the hypothalamus and their pathways. See text for details. *BDNF,* Brain-derived neurotrophic factor; *CRH,* corticotropin-releasing hormone; *TRH,* thyroid-releasing hormone; *MCH,* melanin-concentrating hormone; *PYY,* peptide YY.

second-order neurons are responsible for reducing food intake and increasing energy expenditure by producing brain-derived neurotrophic factor (BDNF), thyroid-releasing hormone (TSH), and corticotropin-releasing hormone (CRH). By contrast, fasting activates the NPY/AgRP neurons to release NPY, which activates Y1and Y5 receptors in second-order neurons. These second-order neurons are responsible for increasing food intake by producing melanin-concentrating hormone (MCH) and orexin, and for reducing energy expenditure by downregulation of sympathetic output. NPY/AgRP neurons also directly inhibit POMC/CART neurons, thus blunting their anorexigenic effect.

To summarize, NPY/AgRP neurons may be thought of as the gas pedals for appetite, whereas POMC/CART neurons represent the brake pedal. The orderly functioning of these two pedals maintains energy homeostasis.

Discussed next are three important components of the afferent system, which regulates appetite and satiety: leptin, gut hormones, and adiponectin, which regulates energy consumption.

Leptin. **Leptin is secreted by fat cells, and its output is regulated by the adequacy of fat stores.** BMI and body fat stores are directly related to leptin secretion. With abundant adipose tissue, leptin secretion is stimulated and it crosses the blood–brain barrier and travels to the hypothalamus, where it reduces food intake by stimulating POMC/CART neurons and inhibiting NPY/AgRP neurons. The opposite sequence of events occurs when there are inadequate stores of body fat: leptin secretion is diminished, and food intake is increased. In persons of stable weight, the activities of these pathways are balanced.

Leptin regulates not only food intake but also energy expenditure through a distinct set of pathways. Thus an abundance of leptin stimulates physical activity, heat production, and energy expenditure. The neurohumoral mediators of leptin-induced energy expenditure are less well defined. *Thermogenesis,* an important catabolic effect mediated by leptin, is controlled in part by hypothalamic signals that increase the release of norepinephrine from sympathetic nerve endings in adipose tissue.

The central role of leptin in energy homeostasis is underscored by the study of mutations affecting components of the leptin pathway in mice and in humans. Mice with mutations that disable the leptin gene or its receptor fail to sense the adequacy of fat stores, so they behave as if they are undernourished, eating ravenously and becoming massively obese. As in mice, mutations of the leptin gene or receptor in humans, although rare, cause massive obesity. More common are mutations in the melanocortin receptor 4 gene *(MC4R),* found in 4% to 5% of patients with massive obesity. These monogenic traits confirm the importance of leptin signaling in the control of body weight, and it is possible that other genetic or acquired defects in the leptin pathway play a role in the more common forms of obesity. In keeping with this notion, the anorexigenic response of leptin is blunted in states of obesity despite high levels of circulating leptin (leptin resistance). Furthermore, injections of leptin in obese humans fail to affect food intake and

energy expenditure, dashing initial enthusiasm of leptin therapy for obesity.

In closing it should be noted that although our discussion has centered on the actions of leptin, there is increasing evidence that insulin, like leptin, exerts anorexigenic responses. Both POMC/CART and NPY/AgRP neurons express insulin receptors. However, while insulin can mimic the actions of leptin, most of the evidence suggests the primacy of leptin in the regulation of energy homeostasis.

Adiponectin. **Adiponectin, produced in the adipose tissue, has been called a "fat-burning molecule,"** as it stimulates fatty acid oxidation in skeletal muscle, thereby reducing fatty acid levels. Evidence suggests that excess fatty acids can cross the blood–brain barrier and enter the hypothalamus, where they are sensed by microglial cells. These cells respond by releasing inflammatory factors that appear to act on hypothalamic neurons to cause leptin resistance, thereby blunting its antiadiposity signals. Because of its actions on reducing fatty acids by promoting their oxidation, adiponectin is called the "guardian angel against obesity." Adiponectin also decreases glucose production in the liver and increases insulin sensitivity, protecting against the metabolic syndrome. In addition to its metabolic effects, adiponectin has antiinflammatory, antiatherogenic, antiproliferative, and cardioprotective effects. Its serum levels are lower in obese than in lean individuals. These effects contribute to obesity-associated insulin resistance, type 2 diabetes, and nonalcoholic fatty liver disease (Chapter 18).

Adiponectin binds to two receptors, AdipoR1 and AdipoR2. These receptors are found in many tissues, including the brain, but AdipoR1 and AdipoR2 are most highly expressed in skeletal muscle and liver, respectively. Binding of adiponectin to its receptors triggers signals that activate cyclic adenosine monophosphate (cAMP)–dependent protein kinase (protein kinase A), which in turn phosphorylates and inactivates acetyl coenzyme A carboxylase, a key enzyme required for fatty acid synthesis.

Gut Hormones. Gut peptides act as short-term meal initiators and terminators. They include ghrelin, PYY, and GLP-1 (glucagon-like peptide-1), among others. *Ghrelin* is produced in the stomach and is the only known gut hormone that increases food intake (orexigenic effect). Its injection in rodents elicits voracious feeding, even after repeated administration. Long-term injections cause weight gain, by increasing caloric intake and reducing energy utilization. Ghrelin acts centrally by activating orexigenic NPY/AgRP neurons.

Ghrelin levels normally rise before meals and fall 1 to 2 hours afterward, but this drop is attenuated in obese persons. Ghrelin levels are lower in obese individuals compared with normal-weight individuals, and levels increase with a reduction in obesity. Interestingly, the rise in ghrelin levels is much reduced in individuals in whom gastric bypass surgery is performed for the treatment of obesity, suggesting that the beneficial effects of such surgery may be in part due to a reduced surface of gastric mucosa that is exposed to food.

PYY and GLP-1 are secreted from endocrine cells in the ileum and colon. Plasma levels of PYY and GLP-1 are low during fasting and increase shortly after food intake. Both PYY and GLP-1 act centrally through NPY/AgRP neurons in the hypothalamus, causing a decrease in food intake. Because of the anorexigenic effect of GLP-1, agonists of GLP-1 receptor have recently been approved for treatment of obesity and type 2 diabetes since in addition to reducing food intake, GLP-1 enhances glucose-dependent insulin secretion.

Adipose Tissue. It has been known for quite some time that there are two types of adipose tissues: white adipose tissue (WAT) and brown adipose tissue (BAT). BAT has the unique property of expending energy by nonshivering thermogenesis. BAT functions by uncoupling energy production from energy storage and converting the energy produced into heat. BAT is abundant in newborns and is located primarily in interscapular and supraclavicular areas. Lesser amounts are present around kidneys, aorta, heart, pancreas, and trachea. Until recently, it was thought that BAT is lost in adults. However, recent imaging studies have revealed that some BAT is preserved in adolescents and adults. Much effort is now focused on how BAT can be preserved in adulthood and activated to burn energy.

In addition to leptin and adiponectin, white adipose tissue produces cytokines such as TNF, IL-6, IL-1, and IL-18; chemokines; and steroid hormones. The increased production of cytokines and chemokines by adipose tissue in obese patients creates a chronic proinflammatory state marked by high levels of circulating C-reactive protein (CRP). This relationship may be more than a one-way street, as emerging evidence suggests that immune cells, particularly tissue macrophages, have important roles in regulating adipocyte function.

It is now evident that through this panoply of mediators, adipose tissue participates in the control of energy balance and energy metabolism, functioning as a link between lipid metabolism, nutrition, and inflammatory responses. Thus the adipocyte, which was relegated to an obscure and passive role as the "Cinderella of cells of metabolism," is now "the belle of the ball" at the forefront of metabolic research.

Role of the Gut Microbiome. An interesting series of observations in mice suggest that the gut microbiome may be involved in the development of obesity. In support of this notion is the finding that the profiles of gut microbiota differ between genetically obese mice and their lean littermates. The microbiome of genetically obese mice can harvest much more energy from food compared with that of lean mice. Colonization of the gut of germ-free mice by microbiota from obese mice (but not microbiota from lean mice) is associated with increased body weight. The relevance of the mouse models to human obesity is tantalizing but remains to be proven. Differences between the gut microbiome of obese and lean humans have also been reported. Whether this difference is causal or just an association is unclear.

Clinical Consequences of Obesity

Obesity, particularly central obesity, is associated with an increase in all-cause mortality and is a known risk factor for a number of conditions including type 2 diabetes, cardiovascular disease, and cancer (Fig. 9.31). Central obesity also stands at the center of a cluster of alterations known

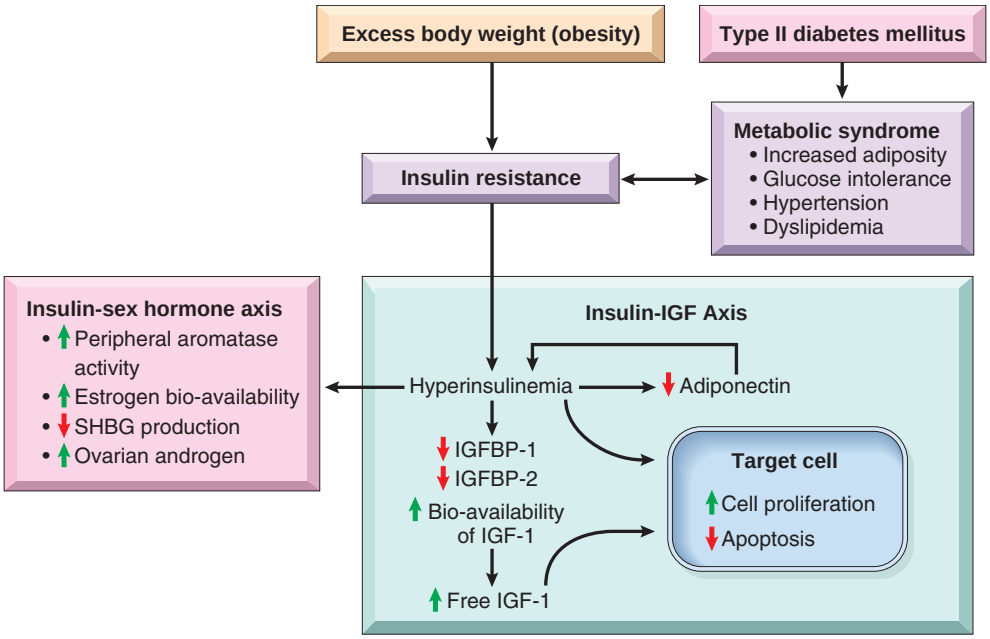


Figure 9.31 Obesity, metabolic syndrome, and cancer. Excessive weight and obesity are precursors of the metabolic syndrome, which is associated with insulin resistance, type 2 diabetes, and hormonal changes. Increases in insulin and insulin-like growth factor-1 stimulate cell proliferation and inhibit apoptosis and may contribute to tumor development. *IGF,* Insulin-like growth factor; *IGFBP,* insulin-like growth factor–binding protein; *SHBG,* sex hormone–binding globulin. (Modified from Renehan AG, et al: and cancer. Obesity and excessive weight: the role of the insulin-IGF axis, *Trends Endocrinol Metab* 17:328, 2006.)

as the *metabolic syndrome,* characterized by abnormalities of glucose and lipid metabolism coupled with hypertension and evidence of a systemic proinflammatory state. This seems to be caused by activation of the inflammasome by free fatty acids and excess levels of lipids in cells and tissue. This in turn stimulates secretion of IL-1, which induces systemic inflammation. The following associations are worthy of note.

- *Obesity is associated with insulin resistance and hyperinsulinemia,* important features of type 2 diabetes (Chapter 24). Inflammation induced by IL-1 and other factors likely contributes to insulin resistance. Excess insulin, in turn, may play a role in the retention of sodium, expansion of blood volume, production of excess norepinephrine, and smooth muscle proliferation that are the hallmarks of hypertension. Whatever the mechanism, the risk of developing hypertension among previously normotensive persons increases proportionately with weight.
- *Obese persons generally have hypertriglyceridemia and low HDL cholesterol levels,* factors that increase the risk of coronary artery disease. The risk of coronary heart disease is compounded by the existence of comorbid conditions including diabetes, hypertension, and dyslipidemia.
- *Nonalcoholic fatty liver disease* is commonly associated with obesity and type 2 diabetes. It can progress to fibrosis and cirrhosis (Chapter 18).
- *Cholelithiasis (gallstones)* is six times more common in obese than in lean subjects. The obesity-associated increase in total body cholesterol and cholesterol turnover leads to elevated biliary excretion of cholesterol in the bile, which in turn predisposes affected persons to the formation of cholesterol-rich gallstones (Chapter 18).
- *Obstructive sleep apnea* and consequent right-sided heart failure is strongly associated with obesity. Hypoventilation syndrome is a constellation of respiratory abnormalities in very obese persons. It has been called the pickwickian syndrome, after the fat lad who was constantly falling asleep in Charles Dickens' novel *The Pickwick Papers.*
- *Marked adiposity is a predisposing factor for the development of degenerative joint disease (osteoarthritis).* This form of arthritis, which typically appears in older persons, is attributed in large part to the cumulative effects of wear and tear on joints. The greater the body burdens of fat, the greater the trauma to joints with the passage of time.
- *Markers of inflammation, such as CRP and proinflammatory cytokines like TNF, are often elevated in obese persons, in particular people with central obesity.* The basis for the inflammation is uncertain; both a direct proinflammatory effect of excess circulating lipids and increased release of cytokines from fat-laden adipocytes have been proposed. Whatever the cause, it is thought that chronic inflammation may contribute to many of the complications of obesity including insulin resistance, metabolic abnormalities, thrombosis, cardiovascular disease, and cancer.

Obesity and Cancer

There is an increased incidence of certain cancers in overweight people, including cancers of the esophagus, thyroid, colon, and kidney in men and cancers of the esophagus, endometrium, gallbladder, and kidney in women. Though the "risk" associated with obesity is modest, because of the prevalence of obesity in the population, it is associated with approximately 40% of all cancers in the United States, somewhat more in women than in men. The underlying mechanisms are unknown and are likely to be multiple.

- *Elevated insulin levels.* Insulin resistance leads to hyperinsulinemia, which includes multiple effects that may

directly or indirectly contribute to cancer (see Fig. 9.31). For example, hyperinsulinemia causes a rise in levels of free insulin-like growth factor 1 (IGF-1). IGF-1 is a mitogen, and its receptor, IGFR-1, is highly expressed in many human cancers. IGFR-1 activates the RAS and PI3K/AKT pathways, which promote the growth of both normal and neoplastic cells (Chapter 7).

- Obesity has effects on *steroid hormones* that regulate cell growth and differentiation in the breast, uterus, and other tissues. Specifically, obesity increases the synthesis of estrogen from androgen precursors, increases androgen synthesis in ovaries and adrenals, and enhances estrogen availability in obese persons by inhibiting the production of sex hormone–binding globulin (SHBG) in the liver.
- As discussed earlier, *adiponectin* secretion from adipose tissue is reduced in obese individuals. Adiponectin suppresses cell proliferation and promotes apoptosis. It does so in part by promoting the actions of p53 and p21. In obese individuals these antineoplastic actions of adiponectin may be compromised.
- The *proinflammatory state* that is associated with obesity may itself be carcinogenic through mechanisms discussed in Chapter 7.

KEY CONCEPTS

OBESITY

- Obesity is a disorder of energy regulation. It increases the risk for a number of important conditions such as insulin resistance, type 2 diabetes, hypertension, and hypertriglyceridemia, which are associated with coronary artery disease, certain cancers, nonalcoholic fatty liver disease, and gallstones.
- The regulation of energy balance has three main components: (1) afferent signals provided mostly by insulin, leptin, ghrelin, GLP-1, and peptide YY; (2) the central hypothalamic system, which integrates afferent signals and triggers the efferent signals; and (3) efferent signals, which control energy balance.
- Leptin plays a key role in energy balance. Its output from adipose tissues is increased by the abundance of fat stores. Leptin binding to its receptors in the hypothalamus decreases appetite and increases energy consumption by stimulating POMC/CART neurons and inhibiting NPY/AgRP neurons.

Diet and Cancer

As you will recall from Chapter 7, the incidence of specific cancers varies as much as 100-fold in different geographic areas. It is well known that differences in incidence of various cancers are not fixed and can be modified by environmental factors including changes in diet. For instance, the incidence of colon cancer in Japanese men and women 55 to 60 years of age was negligible about 50 years ago, but it is now higher than that in men of the same age in the United Kingdom. Studies have also shown a progressive increase in colon cancers in Japanese populations as they moved from Japan to Hawaii and from there to the continental United States. Nevertheless, despite extensive experimental and epidemiologic research, relatively few mechanisms that link diets and specific types of cancer have been established.

With respect to carcinogenesis, three aspects of the diet are of major concern: (1) the content of exogenous carcinogens, (2) the endogenous synthesis of carcinogens from dietary components, and (3) the lack of protective factors.

- Regarding exogenous substances, *aflatoxin* is involved in the development of hepatocellular carcinomas in parts of Asia and Africa, generally in cooperation with hepatitis B virus. Exposure to aflatoxin causes a specific mutation in codon 249 of the *TP53* gene; when found in hepatocellular carcinomas, this mutation serves as a molecular signature for aflatoxin exposure. Debate continues about the carcinogenicity of food additives, artificial sweeteners, and contaminating pesticides.
- The concern about endogenous synthesis of carcinogens or enhancers of carcinogenicity from components of the diet relates principally to gastric carcinomas. *Nitrosamines and nitrosamides* are implicated in the generation of these tumors, as they have been clearly shown to induce gastric cancer in animals. These compounds can be formed in the body from nitrites and amines or amides derived from digested proteins. Sources of nitrites include sodium nitrite added to foods as a preservative, and nitrates, present in common vegetables, which are reduced in the gut by bacterial flora. There is, then, the potential for endogenous production of carcinogenic agents from dietary components, which might well have an effect on the stomach.
- *High animal fat intake combined with low fiber intake has been implicated in the causation of colon cancer.* It has been estimated that doubling the average level of total fiber consumption to about 40 g/day per person in most populations decreases the risk of colon cancer by 50%. The most plausible explanation for this association is that high fat intake increases the level of bile acids in the gut, which in turn modifies intestinal flora, favoring the growth of microaerophilic bacteria. Bile acid metabolites produced by these bacteria may function as carcinogens. The *protective effect of a high-fiber diet* might relate to (1) increased stool bulk and decreased transit time, which decreases the exposure of mucosa to putative offenders, and (2) the capacity of certain fibers to bind carcinogens and thereby protect the mucosa. However, attempts to document these theories in clinical and experimental studies have not generated consistent results.
- *Vitamins C and E, β-carotene, and selenium* have been assumed to have anticarcinogenic effects because of their antioxidant properties. However, thus far there is no convincing evidence that these antioxidants act as chemopreventive agents. As discussed earlier in this chapter, retinoids are effective agents in the therapy of acute promyelocytic leukemia, and associations between low levels of vitamin D and cancer of the colon, prostate, and breast have been reported.

On one hand, despite many tantalizing trends and proclamations by diet gurus, thus far there is no definitive proof that a particular diet can cause or prevent cancer. On the other hand, given the relationships between obesity and cancer development, prevention of obesity through the consumption of a healthy diet is a commonsense measure that goes a long way in preserving good health.

Diet and Systemic Diseases

The problems of undernutrition and overnutrition, as well as specific nutrient deficiencies, have been discussed; however, the composition of the diet, even in the absence of any of these problems, may make a significant contribution to the causation and progression of a number of diseases. A few examples suffice here.

Currently, one of the most important and controversial issues is the contribution of diet to atherogenesis. The central question is whether dietary modification—specifically, reduction in the consumption of foods high in cholesterol and saturated animal fats (e.g., eggs, butter, beef)—can reduce serum cholesterol levels and prevent or retard the development of atherosclerosis (of most importance, coronary heart disease) in people with no previous episode of cardiovascular disease. This is called primary prevention. We know some, but not all, the answers. The average adult in the United States consumes a large amount of fat and cholesterol daily, with a ratio of saturated fatty acids to polyunsaturated fatty acids of about 3:1. Lowering the level of saturates to the level of the polyunsaturates causes a 10% to 15% reduction in serum cholesterol within a few weeks. Vegetable oils (e.g., corn and safflower oils) and fish oils contain polyunsaturated fatty acids and are good sources of such cholesterol-lowering lipids. Fish oil fatty acids belonging to the omega-3, or *n-3,* family have more double bonds than the omega-6, or *n-6,* fatty acids found in vegetable oils. A corollary of this idea is that supplementation of diet with fish oils might protect against atherosclerosis. However, a recent large meta-analysis of 79 randomized controlled trials showed that dietary supplements of omega-3 fatty acids or consumption of oily fish had little or no effect on cardiovascular disease (ischemic heart disease, stroke).

Other specific effects of diet on disease have been recognized:

- Restricting sodium intake reduces hypertension.
- Dietary fiber, or roughage, resulting in increased fecal bulk, is thought by some investigators to provide a preventive effect against diverticulosis of the colon.
- Caloric restriction has been convincingly demonstrated to increase lifespan in experimental animals, including monkeys. The basis for this striking observation is not clear (Chapter 2), and the degree of calorie restriction that is needed to prolong life is so great that one can question if such a life is worth living. Perhaps by understanding the underlying mechanism, we may some day be able to have our cake and eat it, too.
- Even lowly garlic has been touted to protect against heart disease (and also against devils, werewolves, vampires, and, alas, kisses), although research has yet to prove this effect unequivocally. Of these, the effect on kisses is the best established!

SUGGESTED READINGS

Climate Change

Haines A, Ebi K: The imperative for climate action to protect health, *N Engl J Med* 380:263, 2019. *[An excellent review of the health effects of climate change].*

Air Pollution

Beelen R, Raaschau-Nielsen O, Massimo S et al: Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicenter ESCAPE project, *Lancet* 383:785, 2014. *[A large European study relating all-cause mortality and particulate inhalation].*

Heavy Metal Poisoning

Hanna-Attisha M, LaChance J, Sadler RC et al: Elevated blood lead levels in children associated with the Flint drinking water crisis: a spatial analysis of risk and public health response, *Am J Public Health* 106:283, 2016. *[An article that discusses public health issues resulting from water pollution with lead in Flint, Michigan].*

Hon KL, Fung CK, Leung AKC: Childhood lead poisoning: an overview, *Hong Kong Med J* 23:616, 2017. *[A discussion of lead poisoning in childhood].*

Rehman K, Fatima F, Waheed I et al: Prevalence of exposure of heavy metals and their impact on health consequences, *J Cell Biochem* 119:157, 2018. *[A comprehensive paper on heavy metal poisoning].*

Effects of Alcohol and Tobacco

Cao Y, Willet WC, Rimm EB et al: Light to moderate intake of alcohol, drinking patterns, and risk of cancer: results from two prospective US cohort studies, *BMJ* 351:h4238, 2015. *[A discussion of alcohol usage and cancer at different dose levels].*

Centers for Disease Control and Prevention: *Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products.* https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html. *[CDC report on severe lung injury associated with use of flavored e-cigarettes].*

Maddatu J, Anderson-Baucum E, Evans-Molina C: Smoking and the risk of type 2 diabetes, *Trans Res* 184:101, 2017. *[A discussion of accumulating evidence for association of type 2 diabetes and tobacco smoking].*

Rigotti NA: Balancing the benefits and harms of e-cigarettes: a national academies of science, engineering, and medicine report, *Ann Intern Med* 168:666, 2018. *[A summation of the evidence linking e-cigarette use and disease].*

US Department of Health and Human Services: *The health consequences of smoking – 50 years of progress: a report of the surgeon general*, Atlanta, GA, 2014, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. *[A comprehensive report by the US Surgeon General on the health effects of smoking].*

Menopausal Hormone Therapy and Oral Contraceptives

Gialeraki A, Valsami S, Pittaras T et al: Oral contraceptives and HRT risk of thrombosis, *Clin Appl Thromb Hemost* 24:217, 2018. *[A succinct review of the effects of hormone replacement therapy and oral contraceptives on thromboembolism].*

Drugs of Abuse and Radiation Injury

Guy GP Jr, Zhang K, Schieber LZ et al: County-level opioid prescribing in the United States, 2015 and 2017. JAMA Internal Medicine published online February 11, 2019. E1 *[A short article on the magnitude of the US opioid crisis].*

Matthew JD, Forsythe AV, Brady Z et al: Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians, *BMJ* 346:f2360, 2013. *[A paper showing that children who have undergone CT scans have a 24% increased risk of cancer, adding to accruing evidence that CT scans increase the risk of secondary cancers in children and adolescents].*

National Institute on Drug Abuse (NIDA). https://www.drugabuse.gov updated in June 2019. *[An extensive discussion of the effects of marijuana use and abuse including possible addiction].*

Malnutrition

Hariz A, Hamdi MS, Boukhris I et al: Gelatinous transformation of bone marrow in a patient with anorexia nervosa: an uncommon but reversible etiology, *Am J Case Rep* 19:1449, 2018. *[A discussion of*

gelatinous transformation of bone marrow that is pathognomonic of anorexia nervosa].

Million M, Diallo A, Raoult R: Review. Gut microbiota and malnutrition, *Microb Pathog* 106:127, 2017. *[An excellent discussion of the potential role of microbiome in pathogenesis of malnutrition].*

Vitamins

Abdelhamid AS, Brown TJ, Brainard JS et al: Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease, *Cochrane Database Syst Rev* (7):CD003177, 2018. *[A critical appraisal of the rate of omega-3 fatty acid supplementation and prevention of heart disease].*

Padayatty SJ, Levine M: Vitamin C: the known and unknown and Goldilocks, *Oral Dis* 22:463, 2016. *[A detailed discussion of vitamin C physiology and diseases associated with deficiency].*

Rautiainen S, Manson J, Lichtenstein AH: Dietary supplements and disease prevention—a global review, *Nat Rev Endocrinol* 12:407, 2016. *[A broad overview of the role of dietary supplements including vitamins in disease prevention].*

Sarri JC: Vitamin A metabolism in rod and cone visual cycles, *Annu Rev Nutr* 32:125, 2012. *[A discussion of the biochemical mechanism of vitamin A in vision].*

Teng J, Pourmand A, Mazer-Amirshahi M: Vitamin C: the next step in sepsis management?, *J Crit Care* 43:230, 2018. *[An appraisal of the possible role of vitamin C in sepsis management].*

Obesity

Aune D, Sen A, Prasad M et al: BMI and all-cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74million deaths among 30.3 million participants, *BMJ* 353:i2156, 2016. *[A review of several large studies linking obesity with all-cause mortality].*

Berthoud H-R, Munzberg H, Morrison CD: Blaming the brain for obesity: integration of hedonic and homeostatic mechanisms, *Gastroenterology* 152:1728, 2017. *[A discussion of neurohumoral mechanisms and obesity].*

Bouter KE, van Raalte DH, Groen AK et al: Role of gut microbiome in the pathogenesis of obesity and obesity-related metabolic dysfunction, *Gastroenterology* 152:1671, 2017. *[A balanced review of the role of the microbiome in obesity].*

Chechi K, Nedergaard J, Richard D: Brown adipose tissue as an anti-obesity tissue in humans, *Obes Rev* 15:92, 2014. *[A discussion of the role of brown adipose tissue in energy homeostasis].*

Chen Y, Copeland WK, Vedanthan R et al: Association between body mass index and cardiovascular disease mortality in east Asians and south Asians: pooled analysis of prospective data from the Asia Cohort Consortium, *BMJ* 347:f5446, 2013. *[A discussion of the high risk of cardiovascular disease with obesity in a South Asian population].*

Gonzalez-Muniesa P, Martinez-Gonzalez M-A, Hu FB et al: Obesity, *Nat Rev Dis Primers* 3:1, 2017. *[A general review of obesity with focus on visceral and truncal obesity].*

Steele CB, Thomas CC, Henley SJ et al: Vital signs: trends in incidence of cancers associated with overweight and obesity—United States, 2005-2014, *MMWR Morb Mortal Wkly Rep* 66(39):1052, 2017. *[A discussion of the association of cancers with obesity].*

Timper K, Bruning JC: Hypothalamic circuits regulating appetite and energy homeostasis: pathways to obesity, *Dis Model Mech* 10:679, 2017. *[An excellent discussion of energy homeostasis including neurohumoral mechanisms].*

Zhao VW, Scherer PE: Adiponectin, the past two decades, *J Mol Cell Biol* 8:93, 2016. *[A review of the role of adiponectin in regulation of fat metabolism].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 10

Diseases of Infancy and Childhood[a]

Aliya N. Husain • Selene C. Koo

CHAPTER CONTENTS

Children are not merely little adults, and their diseases are not merely variants of adult diseases. Many childhood conditions are unique to, or at least take distinctive forms in, this stage of life and so are discussed separately in this chapter. Diseases originating in the perinatal period are important in that they account for significant morbidity and mortality. The chances for survival of infants improve with each passing week. The infant mortality rate in the United States has shown a decline from a level of 20 deaths per 1000 live births in 1970 to 5.8 in 2014. Although the death rate has continued to decline for all infants, African Americans continue to have an infant mortality rate more than twice (11 deaths per 1000 live births) that of American whites (4.9 deaths). Worldwide, infant mortality rates vary widely, from as low as 1.8 deaths per 1000 live births in Slovenia, to as high as 110.6 deaths in Afghanistan. Rather dismayingly, the United States ranks thirtieth in infant mortality rate among high income nations in the Western hemisphere.

Each stage of development of the infant and child is prey to a somewhat different group of disorders: (1) the neonatal period (the first 4 weeks of life), (2) infancy (the first year of life), (3) 1 to 4 years of age, and (4) 5 to 14 years of age.

Congenital anomalies, prematurity and low birth weight, sudden infant death syndrome, and maternal complications and injuries represent the leading causes of death in the first 12 months of life. Once the infant survives the first year of life, the outlook brightens measurably. In the next two age groups—1 to 4 years and 5 to 9 years—unintentional injuries resulting from accidents are the leading cause of death. Among the natural diseases, in order of importance, congenital anomalies and malignant neoplasms assume major significance. In the 10- to 14-year age group, accidents, malignancies, suicide, homicide, and congenital malformations are the leading causes of death.

The following discussion looks at specific conditions encountered during the various stages of infant and child development.

[a]The prior contributions of Dr. Anirban Maitra to this chapter are gratefully acknowledged.

CONGENITAL ANOMALIES

Congenital anomalies are anatomic defects that are present at birth, but some, such as cardiac defects and renal anomalies, may not become clinically apparent until years later. The term *congenital* means "born with," but it does not imply or exclude a genetic basis for the birth defect. It is estimated that about 120,000 (1 in 33) babies are born with a birth defect each year in the United States. In a sense, anomalies found in live-born infants represent the less serious developmental failures in embryogenesis as they are compatible with live birth. Perhaps 20% of fertilized ova are so anomalous that they are blighted at early stages. Others may be compatible with early fetal development, only to

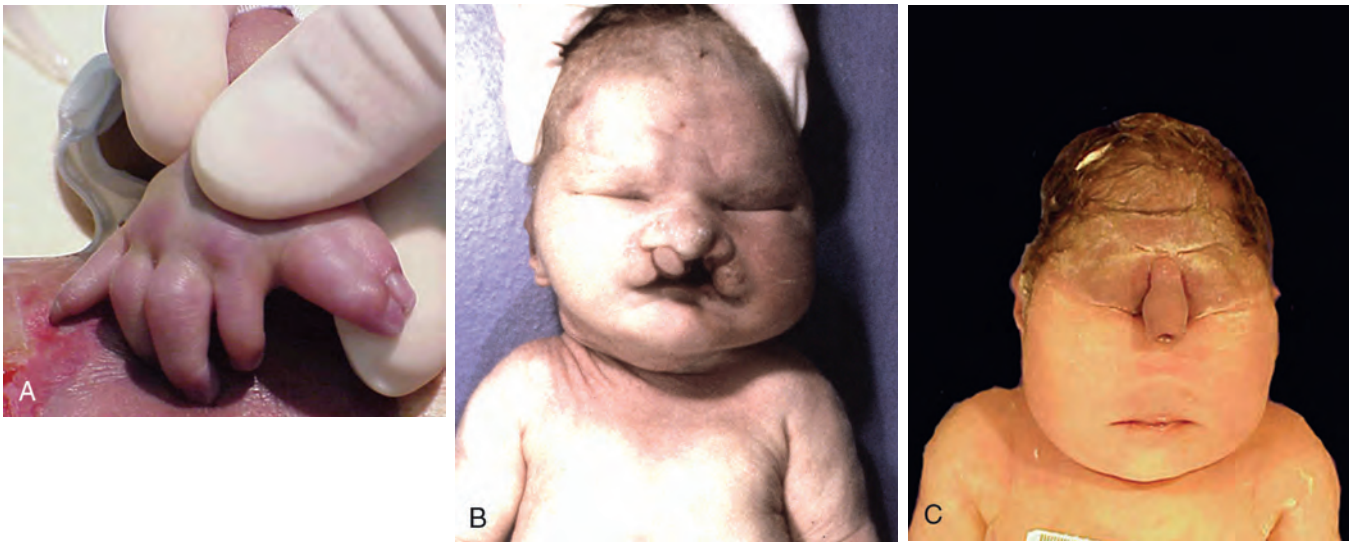


Figure 10.1 Examples of malformations. (A) *Polydactyly* (one or more extra digits) and *syndactyly* (fusion of digits) have little functional consequence when they occur in isolation. Similarly, *cleft lip* (B), with or without associated *cleft palate*, is compatible with life when it occurs as an isolated anomaly; in the present case, however, this neonate had an underlying *malformation syndrome* (trisomy 13) and died of severe cardiac defects. (C) The stillbirth illustrated represents a severe and essentially lethal malformation, wherein the midface structures are fused or ill-formed; in almost all cases, this degree of external dysmorphogenesis is associated with severe internal anomalies such as maldevelopment of the brain and cardiac defects. (A and C, Courtesy Dr. Reade Quinton; B, Courtesy Dr. Beverly Rogers, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.)

lead to spontaneous abortion. Less severe anomalies allow more prolonged intrauterine survival, with some disorders terminating in stillbirth and those still less significant permitting live birth despite the handicaps imposed.

Definitions

The process of morphogenesis (organ and tissue development) can be impaired by a variety of different errors.

- ***Malformations*** **represent primary errors of morphogenesis, in which there is an intrinsically abnormal developmental process** (Fig. 10.1). Malformations can be the result of a single gene or chromosomal defect, but are more commonly multifactorial in origin. Developmental anomalies may present in several patterns. Some, such as congenital heart defects and anencephaly (absence of part or all of the brain), involve single body systems, whereas in other cases multiple malformations involving many organs may coexist.
- ***Disruptions*** **result from secondary destruction of an organ or body region that was previously normal in development**; thus, in contrast with malformations, disruptions arise from an extrinsic disturbance in morphogenesis. Amniotic bands, denoting rupture of amnion with resultant formation of "bands" that encircle, compress, or attach to parts of the developing fetus, are the classic example of a disruption (Fig. 10.2). A variety of environmental agents may cause disruptions (see later). Disruptions are not heritable, hence they are not associated with increased risk of recurrence in subsequent pregnancies.
- ***Deformations*****, like disruptions, represent an extrinsic disturbance of development rather than an intrinsic error of morphogenesis.** Deformations are common problems, affecting approximately 2% of newborn infants to varying degrees. Fundamental to the pathogenesis of deformations is localized or generalized compression of the growing fetus by abnormal biomechanical forces, leading eventually to a variety of structural abnormalities. The most common underlying factor responsible for deformations is uterine constraint. Between 35 and 38 weeks of gestation, rapid increase in the size of the fetus outpaces the growth of the uterus, and the relative amount of amniotic fluid (which normally acts as a cushion) decreases. Thus, even the normal fetus is subjected to some degree of uterine constraint. Several factors increase

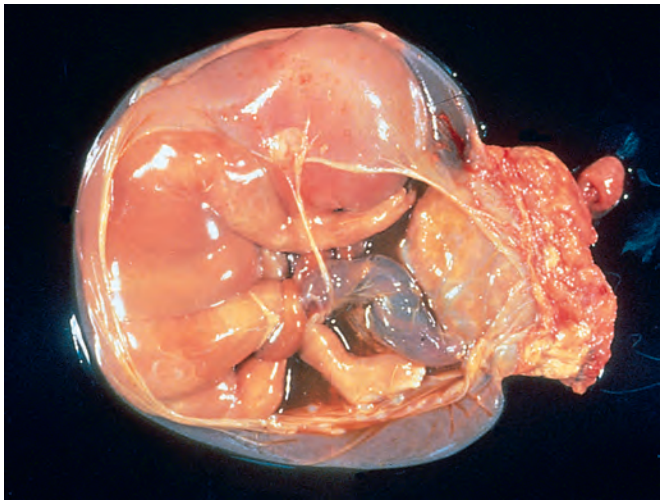


Figure 10.2 Disruption of morphogenesis by an amniotic band. Note the placenta at the right of the diagram and the band of amnion extending from the top portion of the amniotic sac to encircle the leg of the fetus. (Courtesy Dr. Theonia Boyd, Boston Children's Hospital, Boston, Mass.)

the likelihood of excessive compression of the fetus resulting in deformations. Maternal factors include first pregnancy, small uterus, malformed (e.g., bicornuate) uterus, and leiomyomas. Fetal or placental factors include oligohydramnios, multiple fetuses, and abnormal fetal presentation. For example, clubfeet can occur as a component of Potter sequence, described later.

- **A *sequence* is a cascade of anomalies triggered by one initiating aberration.** Approximately one-half the time, congenital anomalies occur singly; in the remaining cases, multiple congenital anomalies are recognized. In some instances, the constellation of anomalies may be explained by a single localized aberration in organogenesis (malformation, disruption, or deformation) that sets into motion secondary effects in other organs. A good example is the *oligohydramnios* (or *Potter*) *sequence* (Fig. 10.3). Oligohydramnios (decreased amniotic fluid) may be caused by a variety of unrelated maternal, placental, or fetal abnormalities. The most common cause of oligohydramnios is chronic leakage of amniotic fluid due to rupture of fetal membranes. Other causes include renal agenesis and urinary tract obstruction in the fetus (because fetal urine is a major constituent of amniotic fluid), and uteroplacental insufficiency resulting from maternal hypertension or severe preeclampsia. The fetal compression associated with significant oligohydramnios, in turn, results in a classic phenotype in the newborn infant, including flattened facies, positional abnormalities of the hands, and clubfeet (Fig. 10.4). The hips may be dislocated. Growth of the chest wall and the contained lungs is also compromised so that the lungs are frequently hypoplastic, and may cause fetal demise. Nodules in the amnion *(amnion nodosum)* are frequently present.
- **A *malformation syndrome* is a constellation of congenital anomalies**, believed to be pathologically related, that, in contrast to a sequence, cannot be explained on the basis of a single, localized, initiating defect. Syndromes are most often caused by a single etiologic agent, such as a viral infection or specific chromosomal abnormality, which simultaneously affects several tissues.

In addition to the aforementioned general definitions, a few organ-specific terms should be defined. *Agenesis* refers to the complete absence of an organ and its associated primordium. A closely related term, *aplasia,* also refers to the absence of an organ but one that occurs due to failure of growth of the existing primordium. *Atresia* describes the absence of an opening, usually of a hollow visceral organ, such as the trachea or intestine. *Hypoplasia* refers to incomplete development or decreased size of an organ with decreased numbers of cells, whereas *hyperplasia* refers to the converse, that is, the enlargement of an organ due to increased numbers of cells. An abnormality in an organ or a tissue as a result of an increase or a decrease in the size (rather than the number) of individual cells defines hypertrophy or hypotrophy, respectively. Finally, *dysplasia* in the context of malformations (versus *neoplasia*) describes an abnormal organization of cells.

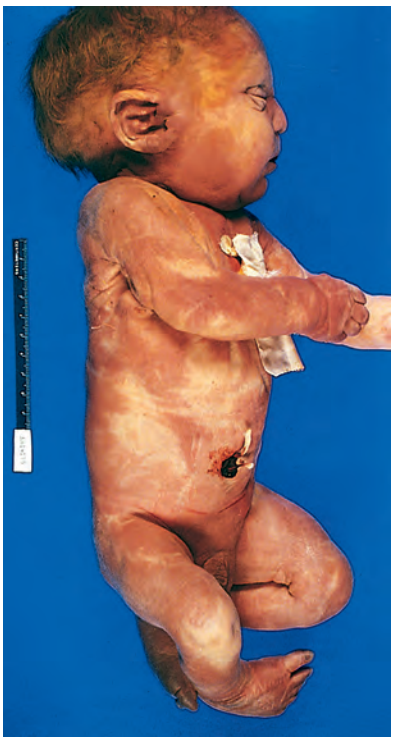


Figure 10.4 Infant with oligohydramnios sequence. Note the flattened facial features and deformed right foot (talipes equinovarus).

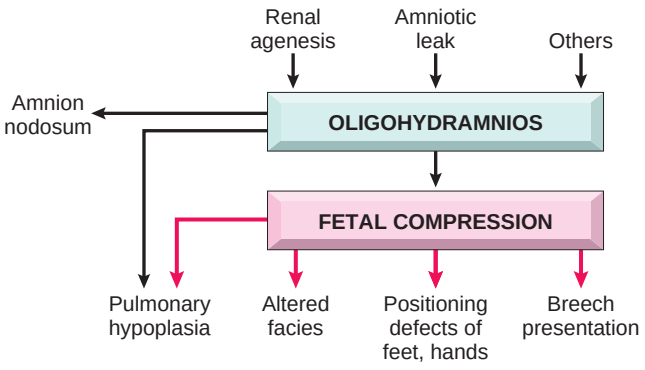


Figure 10.3 Schematic diagram of the pathogenesis of the oligohydramnios sequence.

Causes of Anomalies

Although we are learning a great deal about the molecular bases of some congenital anomalies, the exact cause remains unknown in 40% to 60% of cases. The era of molecular medicine promises to bring additional insights into the mechanisms by which malformations occur. The common known causes of congenital anomalies can be grouped into three major categories: genetic, environmental, and multifactorial (Table 10.1).

Genetic causes of malformations include all of the previously discussed mechanisms of genetic disease (Chapter 5). Virtually all chromosomal syndromes are associated with congenital malformations. Examples include Down syndrome and other trisomies, Turner syndrome, and Klinefelter syndrome. Most chromosomal disorders arise during gametogenesis and hence are not familial. Single-gene

Table 10.1 Causes of Congenital Anomalies in Live-Born Infants

Cause	Frequency (%)
Genetic	
Chromosomal aberrations	10–15
Mendelian inheritance	2–10
Environmental	
Maternal/placental infections Rubella Toxoplasmosis Syphilis Cytomegalovirus Human immunodeficiency virus	2–3
Maternal disease states Diabetes Phenylketonuria Endocrinopathies, including severe obesity	6–8
Drugs and chemicals Alcohol, smoking Folic acid antagonists Androgens Phenytoin Thalidomide Warfarin 13-*cis*-retinoic acid Others	1
Irradiation	1
Multifactorial	20–25
Unknown	40–60

Modified from Stevenson RE, Hall JG, Everman DB, Solomon B, editors: *Human Malformations and Related Anomalies*, ed 3, New York, 2016, Oxford University Press, p 12.

mutations, characterized by Mendelian inheritance, may underlie some major malformations. For example, holoprosencephaly is the most common developmental defect of the forebrain and midface in humans; the Hedgehog signaling pathway plays a critical role in the morphogenesis of these structures, and loss-of-function mutations of individual components within this pathway are reported in families with a history of recurrent holoprosencephaly.

Environmental influences, such as viral infections, drugs, and maternal irradiation, may cause fetal anomalies. Among the viral infections listed in Table 10.1, rubella was a major scourge of the nineteenth and early twentieth centuries. Fortunately, maternal rubella and the resultant *rubella embryopathy* have been virtually eliminated in high income countries as a result of maternal rubella vaccination. A variety of drugs and chemicals have been suspected to be teratogenic, but perhaps less than 1% of congenital malformations are caused by these agents. The list includes thalidomide, alcohol, anticonvulsants, warfarin (oral anticoagulant), and 13-*cis*-retinoic acid, which is used in the treatment of severe acne. For example, thalidomide, once used as a tranquilizer in Europe, causes an extremely high incidence (50% to 80%) of limb malformations. Alcohol, when consumed even in modest amounts during pregnancy, is an important environmental teratogen. Affected infants show prenatal and postnatal growth restriction, facial anomalies (microcephaly, short palpebral fissures, maxillary hypoplasia), and psychomotor disturbances. These in combination are labeled the *fetal alcohol syndrome* (also discussed in Chapter 9). Although cigarette smoke–derived nicotine has not been convincingly demonstrated to be a teratogen, there is a high incidence of spontaneous abortion, premature labor, and placental abnormalities in pregnant women who smoke, babies born to mothers who smoke often have a low birth weight and may be prone to sudden infant death syndrome (discussed later). In light of these findings, it is best to avoid nicotine exposure altogether during pregnancy. Among maternal conditions listed in Table 10.1, diabetes mellitus is a common entity, and despite advances in antenatal obstetric monitoring and glucose control, the incidence of major malformations in infants of diabetic mothers stands between 6% and 10% in most series. Maternal hyperglycemia-induced fetal hyperinsulinemia results in fetal macrosomia (organomegaly and increased body fat and muscle mass); cardiac anomalies, neural tube defects, and other central nervous system (CNS) malformations are some of the major anomalies seen in *diabetic embryopathy*.

Multifactorial inheritance, which implies the interaction of environmental influences with two or more genes of small effect, is the most common genetic etiology of congenital malformations. Examples include relatively common malformations such as cleft lip, cleft palate, and neural tube defects. The dramatic reduction in incidence of neural tube defects by periconceptional intake of folic acid is one case where understanding the environmental stimuli has prevented development of multifactorial malformations even though contributing genes have not been eliminated.

Pathogenesis

The pathogenesis of congenital anomalies is complex and still poorly understood, but two general principles of developmental pathology are relevant regardless of the etiologic agent.

1. The timing of the prenatal teratogenic insult has an important impact on the occurrence and the type of anomaly produced (Fig. 10.5). The intrauterine development of humans can be divided into two phases: (1) the early embryonic period occupying the first 9 weeks of pregnancy and (2) the fetal period terminating at birth.
   - In the *early embryonic period* (first 3 weeks after fertilization), an injurious agent damages either enough cells to cause death and abortion or only a few cells, presumably allowing the embryo to recover without developing defects. **Between the third and the ninth weeks, the embryo is extremely susceptible to teratogenesis;** peak sensitivity occurs between the fourth and the fifth weeks. During this period, organs are being crafted out of the germ cell layers.
   - The *fetal period* that follows organogenesis is marked chiefly by further growth and maturation of the organs, with greatly reduced susceptibility to teratogenic agents. Instead, the fetus is susceptible to growth restriction or injury to already formed organs. Thus a given agent may produce different anomalies if exposure occurs at different times of gestation.
2. The interplay between environmental teratogens and intrinsic genetic defects is exemplified by the fact that features of dysmorphogenesis caused by environmental insults can often be recapitulated by genetic defects in

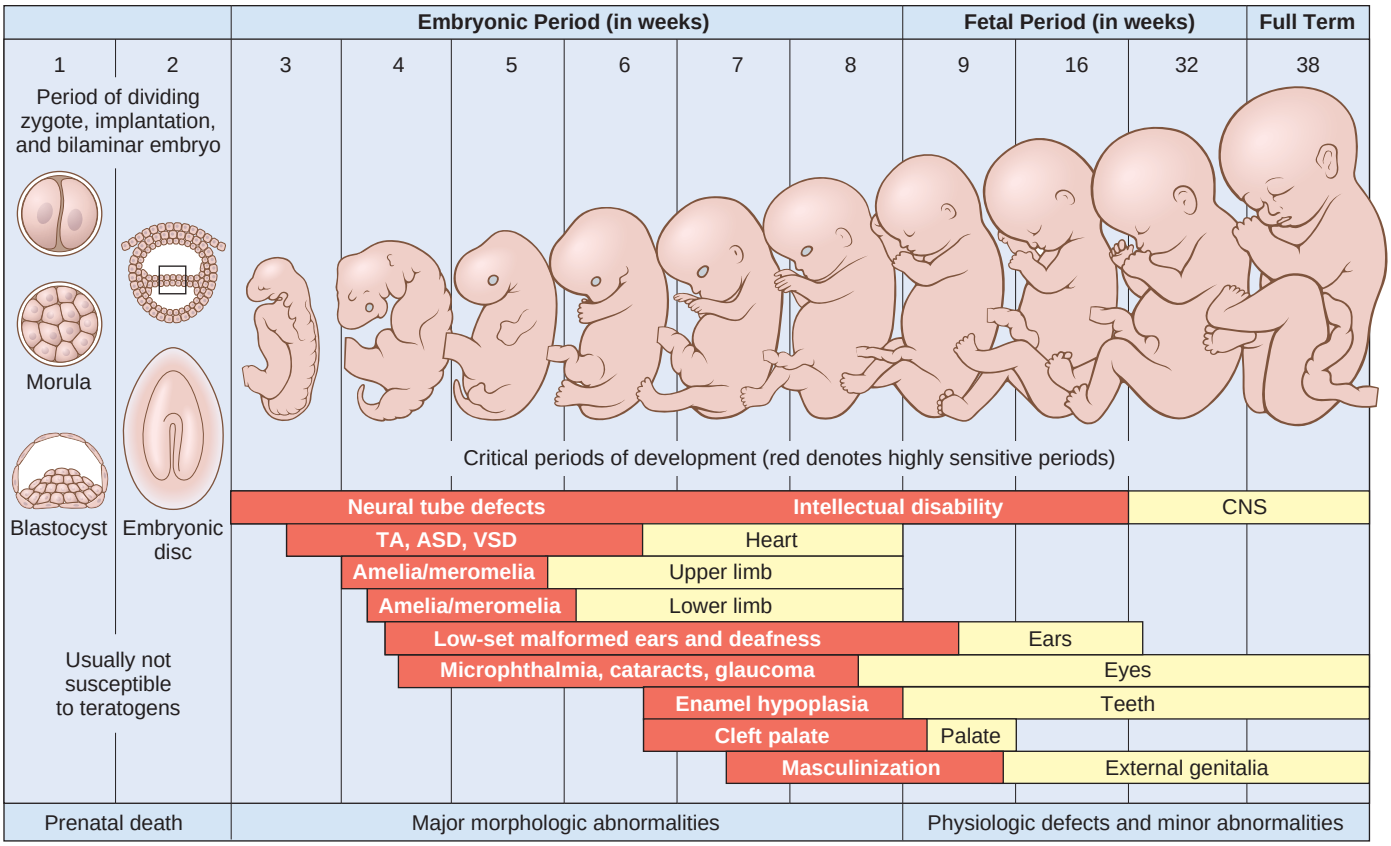


Figure 10.5 Critical periods of development for various organ systems and the resultant malformations. (Modified and redrawn from Moore KL: *The Developing Human,* ed 10, Philadelphia, 2016, WB Saunders, p 474.)

the pathways targeted by these teratogens. This is illustrated by the following representative examples.

- *Cyclopamine* is a plant teratogen contained in the corn lily; pregnant sheep who feed on this plant give birth to lambs with severe craniofacial abnormalities including holoprosencephaly and "cyclopia" (single fused eye, hence the origin of the moniker cyclopamine). This compound is an inhibitor of Hedgehog signaling in the embryo, and, as stated earlier, mutations of Hedgehog genes are present in subsets of patients with holoprosencephaly.
- *Valproic acid* is an antiepileptic and a recognized teratogen during pregnancy. Valproic acid disrupts expression of a family of highly conserved developmentally critical transcription factors known as *homeobox* (HOX) proteins. In vertebrates, HOX proteins have been implicated in the patterning of limbs, vertebrae, and craniofacial structures. Not surprisingly, mutations in the *HOX* family of genes are responsible for congenital anomalies that mimic features observed in valproic acid embryopathy.
- The vitamin A (retinol) derivative *all-trans-retinoic acid* is essential for normal development and differentiation, and its absence during embryogenesis results in a constellation of malformations affecting multiple organ systems, including the eyes, genitourinary system, cardiovascular system, diaphragm, and lungs (see Chapter 9 for effects of vitamin A deficiency in the postnatal period). Conversely, excessive exposure to retinoic acid is also teratogenic. Infants born to mothers treated with retinoic acid for severe acne have a predictable phenotype (retinoic acid embryopathy) including CNS, cardiac, and craniofacial defects, such as cleft lip and cleft palate. The latter may stem from retinoic acid–mediated deregulation of components of the transforming growth factor-β (TGF-β) signaling pathway, which is involved in palatogenesis. Mice with knockout of the *Tgfb3* gene uniformly develop cleft palate, once again illustrating the functional relationship between teratogenic exposure and signaling pathways in the causation of congenital anomalies.

PREMATURITY AND FETAL GROWTH RESTRICTION

Prematurity, defined by a gestational age less than 37 weeks, is the second most common cause of neonatal mortality, behind only congenital anomalies. The Centers for Disease Control and Prevention (CDC; cdc.gov/reproductivehealth) report that in 2016, preterm birth affected about 1 in 10 infants born in the United States. Preterm birth rates decreased from 2007 to 2014 due, in part, to declines in the number of births to teens and young mothers; however, the rate among African-American women (14%) remains greater than that in white women (9%). The major risk factors for prematurity include the following:

- *Preterm premature rupture of membranes* (PPROM): PPROM complicates about 3% of all pregnancies and is responsible for as many as one-third of all preterm deliveries. Rupture of membranes (ROM) before the onset of labor can be spontaneous or induced. PPROM refers to spontaneous ROM occurring *before* 37 weeks of gestation (hence the annotation "preterm"). In contrast, PROM refers to spontaneous ROM occurring *after* 37 weeks of gestation. This distinction is important because after 37 weeks the associated risk to the fetus is considerably decreased. Several clinical risk factors have been identified for PPROM, including a prior history of preterm delivery, preterm labor and/or vaginal bleeding during the current pregnancy, maternal smoking, low socioeconomic status, and poor maternal nutrition. The fetal and maternal outcome after PPROM depends on the gestational age of the fetus (second-trimester PPROM has a dismal prognosis) and the effective prophylaxis of infections in the exposed amniotic cavity.
- *Intrauterine infection:* This is a major cause of preterm labor with and without intact membranes. Intrauterine infection is present in approximately 25% of all preterm births, and the earlier the gestational age at delivery, the higher the frequency of intra-amniotic infection. The histologic correlates of intrauterine infection are inflammation of the placental membranes (chorioamnionitis) and inflammation of the fetal umbilical cord (funisitis). The most common microorganisms implicated in intrauterine infections leading to preterm labor are *Ureaplasma urealyticum, Mycoplasma hominis, Gardnerella vaginalis* (the dominant organism found in "bacterial vaginosis," a polymicrobial infection), *Trichomonas,* gonorrhea, and *Chlamydia.* In low income countries, malaria and HIV are significant contributors to the burden of preterm labor and prematurity. Recent studies have begun to elucidate the molecular mechanisms of inflammation-induced preterm labor. Endogenous Toll-like receptors (TLRs), which bind bacterial components as natural ligands (Chapter 6), have emerged as key players in this process. It is postulated that signals produced by TLR engagement deregulate prostaglandin expression, which, in turn, induces uterine smooth muscle contractions.
- *Uterine, cervical, and placental structural abnormalities:* Uterine distortion (e.g., uterine fibroids), compromised structural support of the cervix ("cervical incompetence"), placenta previa, and abruptio placentae (Chapter 22) are associated with an increased risk of prematurity.
- *Multiple gestation* (twin pregnancy)

The hazards of prematurity are manifold for the newborn and may give rise to one or more of the following:

- *Neonatal respiratory distress syndrome,* also known as *hyaline membrane disease*
- *Necrotizing enterocolitis (NEC)*
- *Sepsis*
- *Intraventricular and germinal matrix hemorrhage*

Fetal Growth Restriction

Although preterm infants have low birth weights, it is usually appropriate once adjusted for their gestational age. In contrast, as many as one-third of infants who weigh less than 2500 g are born at term and are therefore undergrown rather than immature. These small-for-gestational-age (SGA) infants suffer from fetal growth restriction (FGR),which may result from maternal, fetal, or placental abnormalities, although in many cases the specific cause is unknown.

Maternal Abnormalities

By far the most common factors associated with SGA infants are maternal conditions that result in decreased placental blood flow. Vascular diseases such as preeclampsia (toxemia of pregnancy) and chronic hypertension are often the underlying cause. The list of other maternal conditions associated with SGA infants is long, but some of the avoidable factors worth mentioning are maternal narcotic abuse, alcohol intake, and heavy cigarette smoking. Drugs causing FGR include both classic teratogens, such as chemotherapeutic agents, and some commonly administered therapeutic agents, such as phenytoin (Dilantin). Maternal malnutrition (in particular, prolonged hypoglycemia) may also affect fetal growth.

Fetal Abnormalities

Fetal influences are those that intrinsically reduce growth potential of the fetus despite an adequate supply of nutrients from the mother. Prominent among such fetal conditions are chromosomal disorders, congenital anomalies, and congenital infections. Chromosomal abnormalities may be detected in up to 17% of fetuses sampled for FGR and in up to 66% of fetuses with documented ultrasonographic malformations. Among the first group, the abnormalities include triploidy (7%), trisomy 18 (6%), trisomy 21 (1%), trisomy 13 (1%), and a variety of deletions and translocations (2%). Fetal infection should be considered in all infants with FGR. Those most commonly responsible for FGR are the TORCH group of infections (*t*oxoplasmosis, *o*ther viruses and bacteria such as syphilis, *r*ubella, *c*ytomegalovirus, and *h*erpesvirus). Infants who are SGA because of fetal factors usually have symmetric growth restriction (also referred to as proportionate FGR), meaning that all organ systems are similarly affected.

Placental Abnormalities

During the third trimester of pregnancy, vigorous fetal growth places particularly heavy demands on the uteroplacental blood supply. Therefore, the adequacy of placental growth in the preceding midtrimester is extremely important, and uteroplacental insufficiency is an important cause of growth restriction. This insufficiency may result from umbilical-placental vascular anomalies (such as single umbilical artery and abnormal cord insertion), placental abruption, placenta previa, placental thrombosis and infarction, chronic villitis of unknown etiology, placental infection, or multiple gestations (Chapter 22). In some cases the placenta (and the baby) may be small without any detectable underlying cause. Placental causes of FGR tend to result in asymmetric (or disproportionate) growth restriction of the fetus with relative sparing of the brain. Physiologically, this general type of FGR is viewed as a down-regulation of growth in the latter half of gestation because of limited availability of nutrients or oxygen.

The SGA infant faces a difficult course, not only during the struggle for survival in the perinatal period, but also in

childhood and adult life. Depending on the underlying cause of FGR and, to a lesser extent, the degree of prematurity, there is a significant risk of morbidity in the form of a major handicap such as cerebral dysfunction, learning disability, or hearing and visual impairment.

Neonatal Respiratory Distress Syndrome

There are many causes of respiratory distress in the newborn. The most common cause is neonatal respiratory distress syndrome (RDS), also known as hyaline membrane disease because of the deposition of a layer of hyaline proteinaceous material in the peripheral airspaces of infants who succumb to this condition. Others include excessive sedation of the mother, fetal head injury during delivery, aspiration of blood or amniotic fluid, and intrauterine hypoxia brought about by coiling of the umbilical cord about the neck. The incidence of RDS increases with decreasing gestational age, being 1% at 37 weeks, 10.5% at 34 weeks, and 93% in extremely preterm infants (28 weeks or below).

Pathogenesis

The fundamental defect in RDS is pulmonary immaturity and deficiency of surfactant. As described in Chapter 15, surfactant consists predominantly of dipalmitoyl phosphatidylcholine (lecithin), smaller amounts of phosphatidylglycerol, and two groups of surfactant-associated proteins. The first group is composed of hydrophilic glycoproteins SP-A and SP-D, which play a role in pulmonary host defense (innate immunity). The second group consists of hydrophobic surfactant proteins SP-B and SP-C, which, in concert with the surfactant lipids, are involved in the reduction of surface tension at the air-liquid barrier in the alveoli of the lung. With reduced surface tension in the alveoli, less pressure is required to keep them patent and hence aerated. The importance of surfactant proteins in normal lung function can be gauged by the occurrence of severe respiratory failure in neonates with congenital deficiency of surfactant caused by mutations in the *SFTPB* or *SFTPC* genes.

Surfactant production by type II alveolar cells is accelerated after the 35th week of gestation in the fetus. At birth, the first breath of life requires high inspiratory pressures to expand the lungs. With normal levels of surfactant, the lungs retain up to 40% of the residual air volume after the first breath; thus, subsequent breaths require far lower inspiratory pressures. With a deficiency of surfactant, the lungs collapse with each successive breath, so infants must work as hard with each successive breath as they did with the first. The problem of stiff atelectatic lungs is compounded by the soft thoracic wall that is pulled in as the diaphragm descends. Progressive atelectasis and reduced lung compliance then lead to a chain of events as depicted in Fig. 10.6, resulting in protein-rich, fibrin-rich exudation into the alveolar spaces with the formation of hyaline membranes. The fibrin-hyaline membranes are barriers to gas exchange, leading to carbon dioxide retention and hypoxemia. The hypoxemia itself further impairs surfactant synthesis, and a vicious cycle ensues.

Surfactant synthesis is modulated by a variety of hormones and growth factors, including cortisol, insulin, prolactin, thyroxine, and TGF-β. The role of glucocorticoids is particularly important. Conditions associated with intrauterine stress and FGR that increase corticosteroid release lower the risk of developing RDS. Surfactant synthesis can be suppressed by the compensatory high blood levels of insulin in infants of diabetic mothers, which counteracts the effects of steroids. This may explain, in part, why infants of diabetic mothers have a higher risk of developing RDS. Labor is known to increase surfactant synthesis; hence, cesarean delivery before the onset of labor may increase the risk of RDS.

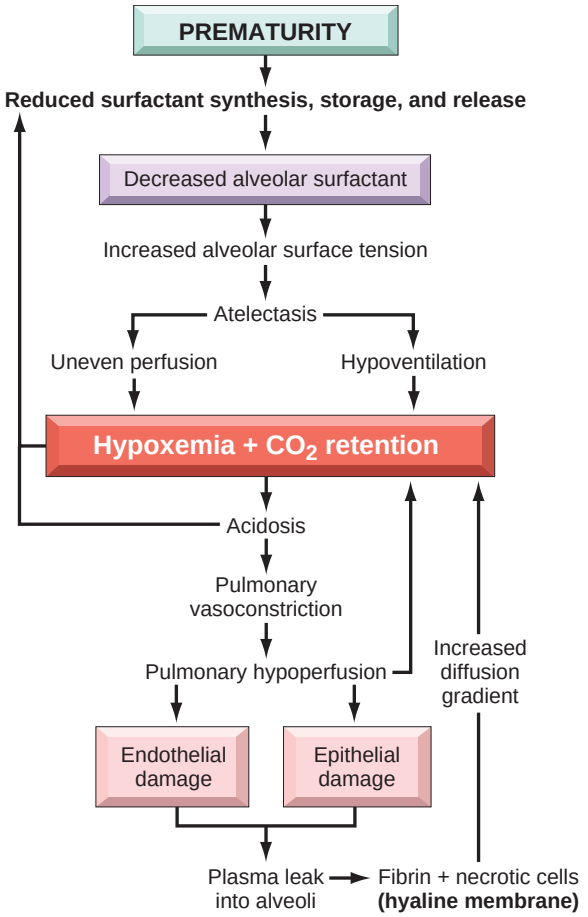


Figure 10.6 Schematic outline of the pathophysiology of respiratory distress syndrome (see text).

MORPHOLOGY

The lungs are distinctive on gross examination. Though of normal size, they are solid, airless, and reddish purple, similar to the color of the liver, and they usually sink in water, indicating the relative absence of entrapped air. Microscopically, alveoli are poorly developed, and those that are present are collapsed (Fig. 10.7). When the infant dies early in the course of the disease, necrotic cellular debris can be seen in the terminal bronchioles and alveolar ducts. The necrotic material becomes incorporated within **eosinophilic hyaline membranes** lining the respiratory bronchioles, alveolar ducts, and alveoli. The membranes are largely made up of fibrin admixed with cell debris derived chiefly from necrotic type II pneumocytes. The sequence of events that leads to the formation of hyaline membranes is depicted in Fig. 10.6.

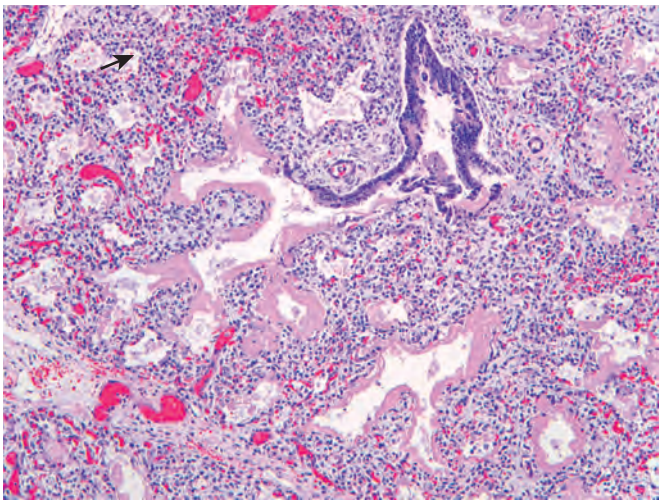


Figure 10.7 Hyaline membrane disease. There are eosinophilic thick hyaline membranes lining the dilated alveoli. Note that the remaining airspaces are lined by cuboidal epithelium *(arrow)*, indicating immaturity of the lung.

There is a remarkable paucity of neutrophilic inflammatory reaction associated with these membranes. The lesions of hyaline membrane disease are never seen in stillborn infants.

In infants who survive more than 48 hours, progressive reparative changes occur in the lungs. The alveolar epithelium proliferates under the surface of the membrane, which may detach into the airspace where it undergoes partial digestion or phagocytosis by macrophages. If it does not detach, fibroblasts grow into the membrane, and it becomes incorporated into the alveolar wall.

Clinical Features

The infant is almost always preterm but has weight appropriate for gestational age, and there are strong but not invariable associations with male gender, maternal diabetes, and cesarean delivery. Resuscitation may be necessary at birth, but usually within a few minutes rhythmic breathing and normal color are reestablished. Soon afterward, often within 30 minutes, breathing becomes more difficult, and within a few hours cyanosis becomes evident in the untreated infant. Fine rales can then be heard over both lung fields. A chest radiograph at this time usually reveals uniform minute reticulogranular densities, producing a so-called ground-glass picture. In the full-blown condition, the respiratory distress persists, cyanosis increases, and even the administration of 80% oxygen by a variety of ventilatory methods fails to improve the situation. If therapy staves off death for the first 3 or 4 days, however, the infant has an excellent chance of recovery.

Currently, the actual clinical course and prognosis for neonatal RDS vary, depending on the maturity and birth weight of the infant and the promptness of institution of therapy. A major thrust in the control of RDS focuses on prevention, either by delaying labor until the fetal lung reaches maturity or by inducing maturation of the lung in the at-risk fetus by use of antenatal steroids. Critical to these objectives is the ability to assess fetal lung maturity accurately. Because pulmonary secretions are discharged into the amniotic fluid, analysis of amniotic fluid phospholipids provides a good estimate of the level of surfactant in the alveolar lining. Prophylactic administration of exogenous surfactant at birth to extremely premature infants (gestational age <28 weeks) has been shown to be very beneficial. Together with newer assisted ventilation techniques, this has resulted in dramatic improvement in pulmonary function, symptom resolution, a shortened course, and markedly reduced mortality. In uncomplicated cases, recovery begins to occur within 3 or 4 days.

Complications of RDS, retrolental fibroplasia (also called retinopathy of prematurity) in the eyes, and bronchopulmonary dysplasia are now less frequent and less severe when they do occur.

- *Retinopathy of prematurity* has a two-phase pathogenesis. During the hyperoxic phase of RDS therapy (phase I), expression of the proangiogenic vascular endothelial growth factor (VEGF) is markedly decreased, causing endothelial cell apoptosis; VEGF levels rebound after return to relatively hypoxic room air ventilation (phase II), inducing retinal vessel proliferation (neovascularization) characteristic of the lesions in the retina (Chapter 29).
- *Bronchopulmonary dysplasia is* characterized by striking decrease in alveolar septation (manifested as large, simplified alveolar structures) and a dysmorphic capillary configuration. Thus, the current view is that bronchopulmonary dysplasia is caused by a potentially reversible impairment in the development of alveolar septation at the so-called "saccular" stage. Multiple factors—hyperoxemia, hyperventilation, prematurity, inflammatory cytokines, and vascular maldevelopment—contribute to bronchopulmonary dysplasia and probably act additively or synergistically to promote injury. The levels of a variety of proinflammatory cytokines (TNF, interleukin-1β [IL-1β], IL-6, and IL-8) are increased in the alveoli of infants who develop bronchopulmonary dysplasia, suggesting a role for these cytokines in arresting pulmonary development. Treatment with mesenchymal stem cells has potential but remains experimental. They act by secreting soluble factors that suppress inflammation and favor repair of the air spaces.

Infants who recover from RDS are also at increased risk for developing a variety of other complications associated with preterm birth; most important among these are patent ductus arteriosus, intraventricular hemorrhage, and necrotizing enterocolitis.

Necrotizing Enterocolitis

Necrotizing enterocolitis (NEC) is most common in premature infants, with the incidence of the disease being inversely proportional to the gestational age. It occurs in approximately 1 in 10 very-low-birth-weight infants (<1500 g). Approximately 2500 cases occur annually in the United States.

The pathogenesis of NEC is uncertain but multifactorial. In addition to prematurity, most cases are associated with enteral feeding, suggesting that some postnatal insult (such as introduction of bacteria) sets in motion a cascade of events culminating in tissue destruction. Although infectious agents may play a role in the pathogenesis of NEC, no single bacterial pathogen has been linked to the disease. Perhaps alteration in the microbiome on enteral feeding

plays a role. A large number of inflammatory mediators have been associated with NEC. One particular mediator, platelet activating factor (PAF), has been implicated in increasing mucosal permeability by promoting enterocyte apoptosis and compromising intercellular tight junctions, thus adding "fuel to the fire." Stool and serum samples of infants with NEC demonstrate higher PAF levels than age-matched controls. Ultimately, breakdown of mucosal barrier functions permits transluminal migration of gut bacteria, leading to a vicious cycle of inflammation, mucosal necrosis, and further bacterial entry, eventually culminating in sepsis and shock (Chapter 4).

The clinical course is fairly typical, with the onset of bloody stools, abdominal distention, and development of circulatory collapse. Abdominal radiographs often demonstrate gas within the intestinal wall *(pneumatosis intestinalis)*. When detected early, NEC often can be managed conservatively, but many cases (20% to 60%) require resection of the necrotic segments of bowel. NEC is associated with high perinatal mortality; those who survive often develop post-NEC strictures from fibrosis caused by the healing process. Probiotic therapies are being evaluated for prevention of NEC.

MORPHOLOGY

NEC typically involves the terminal ileum, cecum, and right colon, although any part of the small or large intestine may be involved. The involved segment is distended, friable, and congested, or it can be frankly gangrenous; intestinal perforation with accompanying peritonitis may occur. Microscopically, mucosal or transmural coagulative necrosis, ulceration, bacterial colonization, and submucosal gas bubbles may be seen (Fig. 10.8). Reparative changes, such as the formation of granulation tissue and fibrosis, may begin shortly after the acute episode.

PERINATAL INFECTIONS

In general, fetal and perinatal infections are acquired through one of two primary routes—transcervically (also referred to as ascending) or transplacentally (hematologic). Occasionally, infections occur by a combination of the two routes in that an ascending microorganism infects the endometrium and then invades the fetal bloodstream via the chorionic villi.

Transcervical (Ascending) Infections

Most bacterial and a few viral (e.g., herpes simplex II) infections are acquired by the cervicovaginal route. Such infections may be acquired in utero or around the time of birth. In general, the fetus acquires the infection either by inhaling infected amniotic fluid into the lungs shortly before birth or by passing through an infected birth canal during delivery. As stated earlier, preterm birth is a common and unfortunate consequence of infection. Preterm birth due to infection may be related either to damage and rupture of the amniotic sac as a direct consequence of the inflammation, or to the induction of labor by prostaglandins released from infiltrating neutrophils. Inflammation of the placental membranes and cord are usually seen, but the presence or absence and severity of chorioamnionitis do not necessarily correlate with the severity of the fetal infection. In the fetus infected by inhalation of amniotic fluid, pneumonia, sepsis, and meningitis are the most common sequelae.

Transplacental (Hematologic) Infections

Most parasitic (e.g., *Toxoplasma,* malaria) and viral infections and a few bacterial infections (i.e., *Listeria*, *Treponema*) gain access to the fetal bloodstream transplacentally via the

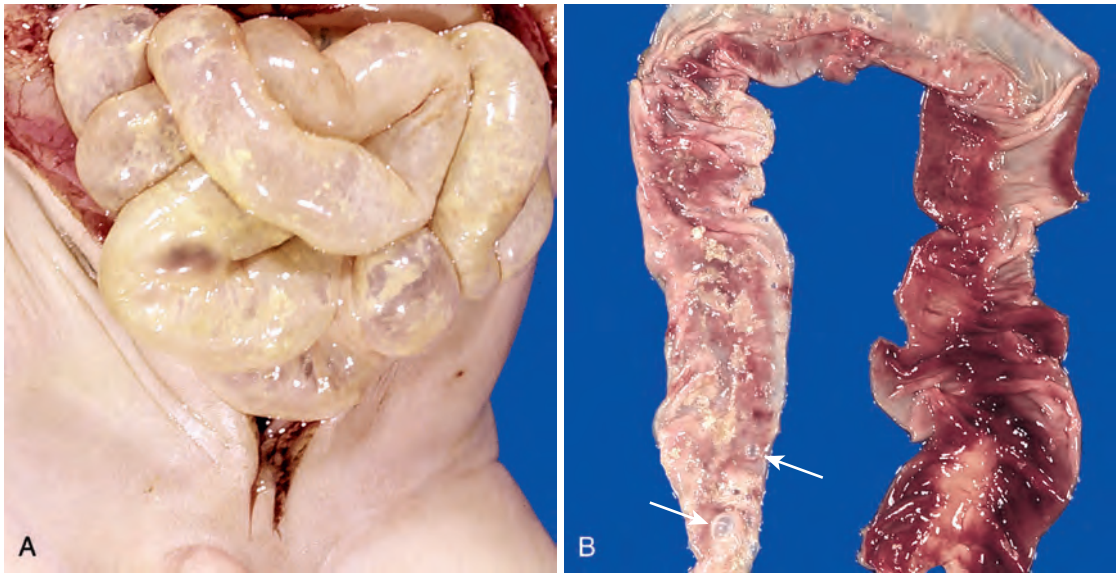


Figure 10.8 Necrotizing enterocolitis (NEC). (A) Postmortem examination in a severe case of NEC shows the entire small bowel is markedly distended with a perilously thin wall (usually this implies impending perforation). (B) The congested portion of the ileum corresponds to areas of hemorrhagic infarction and transmural necrosis microscopically. Submucosal gas bubbles (pneumatosis intestinalis) can be seen in several areas *(arrows)*.

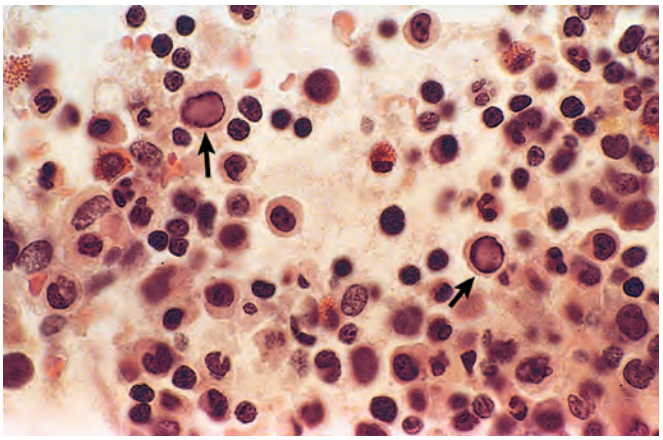


Figure 10.9 Bone marrow from an infant infected with parvovirus B19. The *arrows* indicate two erythroid precursors with large homogeneous intranuclear inclusions and a surrounding peripheral rim of residual chromatin.

chorionic villi. This hematogenous transmission may occur at any time during gestation or occasionally, as may be the case with hepatitis B and HIV, at the time of delivery via maternal-to-fetal transfusion. The clinical manifestations of these infections are highly variable, depending largely on the gestational timing and microorganism involved.

Parvovirus B19, which causes erythema infectiosum or "fifth disease of childhood" in immunocompetent older children, can infect 1% to 5% of seronegative (nonimmune) pregnant women, and the vast majority have a normal pregnancy outcome. Adverse pregnancy outcomes in a minority of intrauterine infections include spontaneous abortion (particularly in the second trimester), stillbirth, hydrops fetalis (see later), and congenital anemia. Parvovirus B19 has a particular tropism for erythroid cells, and diagnostic viral inclusions can be seen in early erythroid progenitors in infected infants (Fig. 10.9).

The *TORCH group of infections* (see earlier) are grouped together because they may evoke similar clinical and pathologic manifestations, including fever, encephalitis, chorioretinitis, hepatosplenomegaly, pneumonitis, myocarditis, hemolytic anemia, and vesicular or hemorrhagic skin lesions. Such infections occurring early in gestation may also cause chronic sequelae in the child, including growth retardation and intellectual disability, cataracts, congenital cardiac anomalies, and bone defects.

Sepsis

Perinatal sepsis can be grouped clinically based on early onset (within the first 7 days of life) versus late onset (from 7 days to 3 months). Most cases of early-onset sepsis are acquired at or shortly before birth and tend to result in clinical signs and symptoms of pneumonia, sepsis, and occasionally meningitis within 4 or 5 days of life. Group B *Streptococcus* is the most common cause of early-onset sepsis as well as early-onset bacterial meningitis. Infections with *Listeria* and *Candida* have longer latent periods between the time of microorganism inoculation and the appearance of clinical symptoms and present as late-onset sepsis.

FETAL HYDROPS

Fetal hydrops refers to the accumulation of edema fluid in the fetus during intrauterine growth. In the past, hemolytic anemia caused by Rh blood group incompatibility between mother and fetus (immune hydrops) was the most common cause, but with the successful prophylaxis of this disorder during pregnancy, nonimmune hydrops is more common. The intrauterine fluid accumulation can be quite variable, from progressive, generalized edema of the fetus (hydrops fetalis), a usually lethal condition, to more localized edema, such as isolated pleural and peritoneal effusions, or postnuchal fluid accumulation (cystic hygroma, see later) that are compatible with life.

Immune Hydrops

Immune hydrops is a hemolytic disease caused by blood group antigen incompatibility between mother and fetus. When the fetus inherits red cell antigenic determinants from the father that are foreign to the mother, a maternal immune reaction may occur. The major antigens known to induce clinically significant immunologic reactions are certain Rh antigens and the ABO blood groups. The reaction occurs in second and subsequent pregnancies in an Rh-negative mother with an Rh-positive father.

Etiology and Pathogenesis

The underlying basis of immune hydrops is the immunization of the mother by blood group antigens on fetal red cells and the free passage of antibodies from the mother through the placenta to the fetus (Fig. 10.10). Fetal red cells may reach the maternal circulation during the last trimester of pregnancy, when the cytotrophoblast is no longer present as a barrier, or during childbirth itself. The mother thus becomes sensitized to the foreign antigen. The initial exposure to Rh antigen evokes the formation of IgM antibodies that unlike IgG antibodies, do not cross the placenta. Thus, Rh disease is uncommon with the first pregnancy. Exposure during a subsequent pregnancy generally leads to a brisk IgG antibody response and the risk of immune hydrops.

Of the numerous antigens included in the Rh system, only the D antigen is a major cause of Rh incompatibility. Several factors influence the immune response to RhD-positive fetal red cells that reach the maternal circulation.

- *Concurrent ABO incompatibility protects the mother against Rh immunization* because the fetal red cells are promptly coated and removed from the maternal circulation by anti-A or anti-B IgM antibodies that do not cross the placenta.
- *The antibody response depends on the dose of immunizing antigen;* hence, hemolytic disease develops only when the mother has experienced a significant transplacental bleed (>1 mL of Rh-positive fetal red cells).

The incidence of maternal Rh isoimmunization has decreased significantly since the use of Rhesus immune globulin (RhIg) containing anti-D antibodies. Administration of RhIg at 28 weeks of gestation and within 72 hours of delivery to Rh-negative mothers significantly decreases the risk for hemolytic disease in Rh-positive neonates and

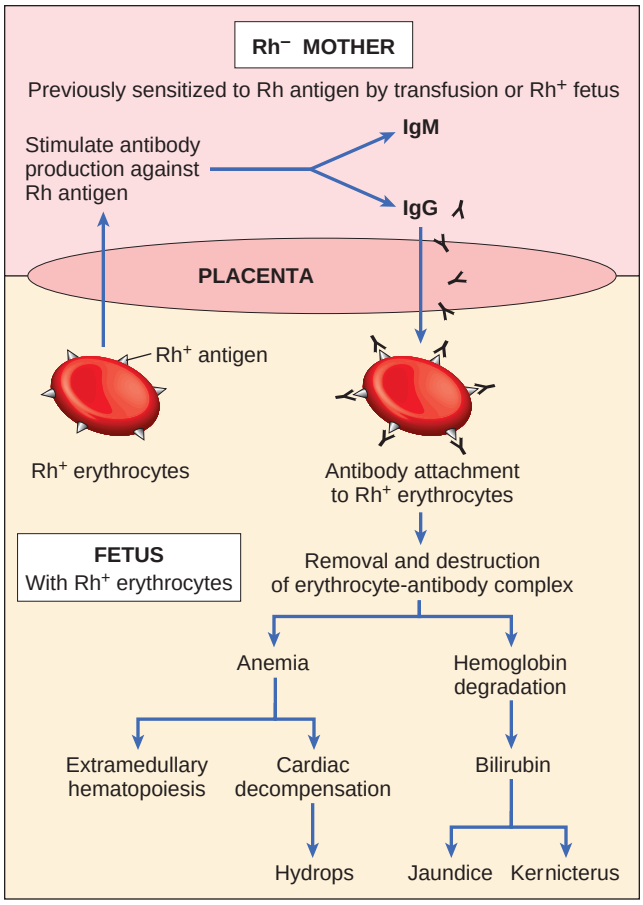


Figure 10.10 Pathogenesis of immune hydrops fetalis (see text).

in subsequent pregnancies; RhIg is also administered following abortions, because these too can lead to immunization. Antenatal identification and management of the at-risk fetus have been greatly facilitated by amniocentesis and chorionic villus and fetal blood sampling. In addition, cloning of the *RHD* gene has resulted in efforts to determine fetal Rh status using maternal serum because it contains fetal DNA. When identified, cases of severe intrauterine hemolysis may be treated by fetal intravascular transfusions via the umbilical cord and early delivery.

The pathogenesis of fetal hemolysis caused by maternal-fetal ABO incompatibility is slightly different from that caused by differences in the Rh antigens. ABO incompatibility occurs in approximately 20% to 25% of pregnancies, but laboratory evidence of hemolytic disease occurs in only 1 in 10 of such infants, and the hemolytic disease is severe enough to require treatment in only 1 in 200 cases. Several factors account for this. First, most anti-A and anti-B antibodies are of the IgM type and hence do not cross the placenta. Second, neonatal red cells express blood group antigens A and B poorly. Third, many cells other than red cells express A and B antigens and thus absorb some of the transferred antibody. ABO hemolytic disease occurs almost exclusively in infants of group A or B who are born of group O mothers. For unknown reasons, certain group O women possess IgG antibodies directed against group A or B antigens (or both) even without prior sensitization. Therefore, the firstborn may be affected. Fortunately, even with transplacentally acquired antibodies, lysis of the infant's red cells is minimal. There is no effective protection against ABO reactions.

There are two consequences of excessive destruction of red cells in the neonate (see Fig. 10.10). The severity of these changes varies considerably, depending on the degree of hemolysis and the maturity of the infant.

- *Anemia* is a direct result of red cell loss. If hemolysis is mild, increased red cell production may suffice to maintain near-normal levels of red cells. However, with more severe hemolysis, progressive anemia develops and may result in hypoxic injury to the heart and liver. Because of liver injury, plasma protein synthesis decreases, and levels of these proteins may drop to as low as 2 to 2.5 mg/dL. Cardiac hypoxia may lead to cardiac decompensation and failure. The combination of reduced plasma oncotic pressure and increased hydrostatic pressure in the circulation (secondary to cardiac failure) results in generalized edema and anasarca, culminating in hydrops fetalis.
- *Jaundice* develops because hemolysis produces unconjugated bilirubin (Chapter 18). Bilirubin also passes through the infant's poorly developed blood-brain barrier. Being water-insoluble, bilirubin blinds to lipids in the brain and can damage the CNS, causing kernicterus (see Fig. 10.13).

Nonimmune Hydrops

The three major causes of nonimmune hydrops include cardiovascular defects, chromosomal anomalies, and fetal anemia. Both structural and functional cardiovascular defects, such as congenital malformations and arrhythmias, may result in intrauterine cardiac failure and hydrops. Among the chromosomal anomalies, 45,X karyotype (Turner syndrome) and trisomies 21 and 18 are associated with fetal hydrops because of the accompanying structural cardiac anomalies. In Turner syndrome, abnormalities of lymphatic drainage from the neck may lead to postnuchal fluid accumulation (cystic hygromas) as well. Fetal anemia, not caused by Rh- or ABO-associated antibodies, can also result in hydrops. In fact, in some parts of the world (e.g., Southeast Asia), severe fetal anemia due to homozygous α-thalassemia, resulting from deletion of all four α-globin genes, is probably the most common cause of nonimmune hydrops (Chapter 14). Transplacental infection by parvovirus B19 is rapidly emerging as an important cause of hydrops (see earlier). The virus gains preferential entry into erythroid precursors (normoblasts), where it replicates, leading to apoptosis of red cell progenitors and isolated red cell aplasia. Parvoviral intranuclear inclusions can be seen within circulating and marrow erythroid precursors (see Fig. 10.9). Approximately 10% of cases of nonimmune hydrops are related to monozygous twin pregnancies and twin-to-twin transfusion occurring through anastomoses between the two circulations. In up to 20% of cases, the cause remains unknown.

MORPHOLOGY

The anatomic findings in fetuses with intrauterine fluid accumulation vary with both the severity of the disease and the underlying etiology. As previously noted, hydrops fetalis represents the most severe and generalized manifestation (Fig. 10.11), and lesser degrees

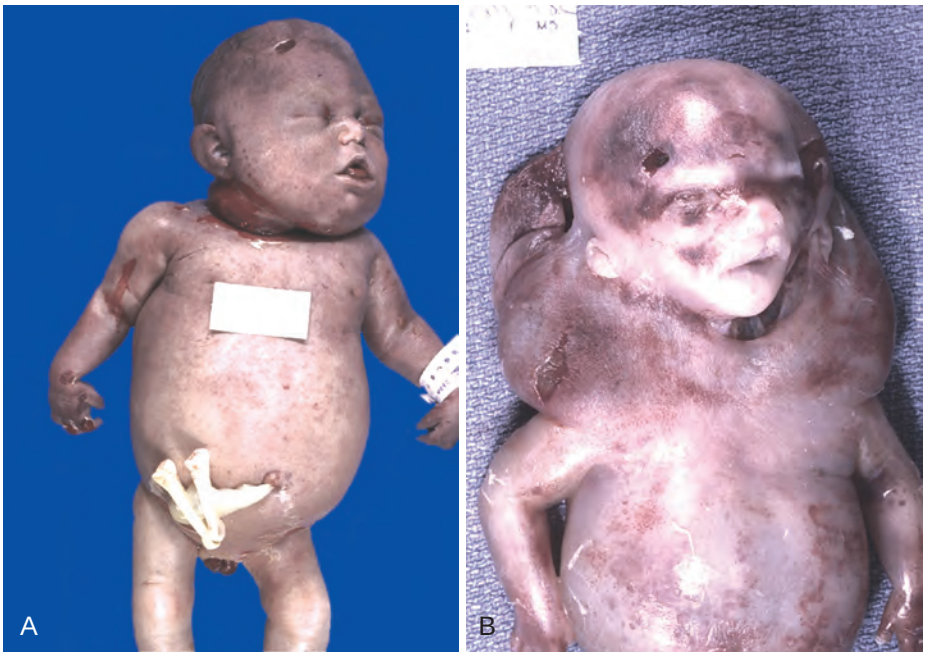


Figure 10.11 Hydrops fetalis. (A) There is generalized accumulation of fluid in the fetus. (B) Fluid accumulation is particularly prominent in the soft tissues of the neck, and this condition has been termed *cystic hygroma.* Cystic hygromas are characteristically seen, but not limited to, constitutional chromosomal anomalies such as 45,X karyotypes. (Courtesy Dr. Beverly Rogers, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Tex.)

of edema such as isolated pleural, peritoneal, or postnuchal fluid collections can occur. Accordingly, infants may be stillborn, die within the first few days, or recover completely. The presence of dysmorphic features suggests a chromosomal abnormality; postmortem examination may reveal an underlying cardiac anomaly.

In hydrops associated with fetal anemia, both the fetus and placenta are characteristically pale; in most cases the liver and spleen are enlarged from cardiac failure and congestion. Additionally, the bone marrow demonstrates compensatory hyperplasia of erythroid precursors (parvovirus-associated red cell aplasia being a notable exception), and extramedullary hematopoiesis is present in the liver, spleen, lymph nodes, and possibly other tissues such as the kidneys, lungs, and even the heart (Fig. 10.12). The increased hematopoietic activity accounts for the presence in the peripheral circulation of large numbers of immature red cells, including reticulocytes and erythroblasts **(erythroblastosis fetalis).**

The most serious threat in fetal hydrops is CNS damage, known as **kernicterus** (Fig. 10.13). The affected brain is enlarged and edematous and, when sectioned, has a bright yellow color, particularly the basal ganglia, thalamus, cerebellum, cerebral gray matter, and spinal cord. The precise level of bilirubin that induces kernicterus is unpredictable, but neural damage usually requires a blood bilirubin level greater than 20 mg/dL in term infants; in premature infants this threshold may be considerably lower.

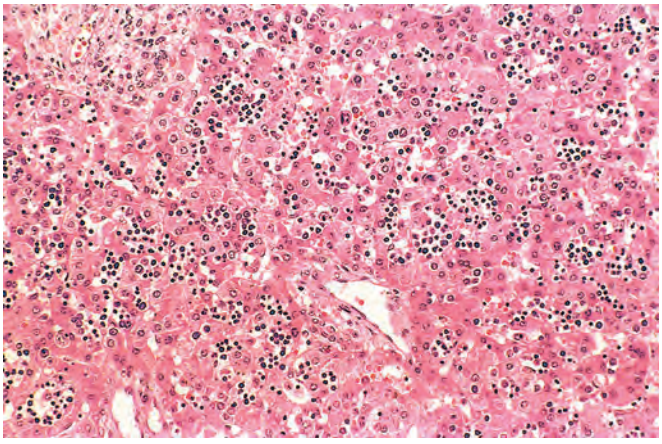


Figure 10.12 Numerous islands of extramedullary hematopoiesis (small blue cells) are scattered among mature hepatocytes in the liver of this infant with nonimmune hydrops fetalis.

Clinical Features

The clinical manifestations of fetal hydrops vary with the severity of the disease and can be inferred from the preceding discussion. Minimally affected infants display pallor, possibly accompanied by hepatosplenomegaly (to which may be added jaundice with more severe hemolytic reactions), whereas the most gravely ill neonates present with intense jaundice, generalized edema, and signs of neurologic injury. These infants may be supported by a variety of measures, including phototherapy (visual light oxidizes toxic unconjugated bilirubin to harmless, readily excreted, water-soluble dipyrroles) and, in severe cases, total exchange transfusion of the infant.

INBORN ERRORS OF METABOLISM AND OTHER GENETIC DISORDERS

Inborn errors of metabolism are genetic disorders that give rise to defects in metabolism. Since Sir Archibald Garrod

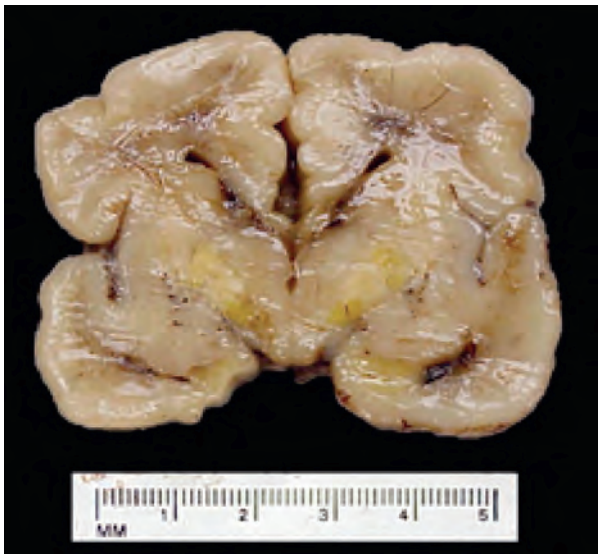


Figure 10.13 Kernicterus. Note the yellow discoloration of the brain parenchyma due to bilirubin accumulation, which is most prominent in the basal ganglia deep to the ventricles.

coined the term in 1908, the number of recognized metabolic diseases has increased exponentially. Most inborn errors of metabolism are rare diseases that are generally inherited as autosomal recessive or X-linked traits (Chapter 5). Mitochondrial disorders (Chapter 5) form a distinct entity by themselves. Some of the clinical features that suggest an underlying metabolic disorder in a neonate are listed in Table 10.2. The clinical manifestations of these diseases are generally the result of either abnormal metabolite accumulation or deficiency of the desired product. Three genetic disorders, phenylketonuria (PKU), galactosemia, and cystic fibrosis, are selected for discussion here. PKU and galactosemia are reviewed because their early diagnosis (via neonatal screening programs) is particularly important, and early death or intellectual disability can be prevented with appropriate dietary regimens. Cystic fibrosis is included because it is one of the most common, potentially lethal diseases occurring in individuals of Caucasian descent.

Phenylketonuria

There are several variants of this inborn error of metabolism, which affects 1 in 10,000 live-born Caucasian infants. The most common form, referred to as classic PKU, is quite common in persons of Scandinavian descent and is distinctly uncommon in African-American and Jewish populations.

PKU is an autosomal recessive disorder caused by a severe deficiency of the enzyme phenylalanine hydroxylase (PAH) leading to hyperphenylalaninemia. Affected infants are normal at birth but within a few weeks develop an elevated plasma phenylalanine level, which impairs brain development. Usually by 6 months of life, severe intellectual disability becomes evident; fewer than 4% of untreated phenylketonuric children have IQs greater than 50 or 60. About one-third of these children are never able to walk, and two-thirds cannot talk. Seizures, other neurologic abnormalities, decreased pigmentation of hair and skin, a characteristic musty odor, and eczema often accompany the intellectual disability in untreated children. Hyperphenylalaninemia and the resultant intellectual disability can be avoided by restricting phenylalanine intake early in life. Hence, all states in the United States have adopted screening procedures to detect PKU in the immediate postnatal period.

Many female PKU patients, if treated with dietary restriction early in life, reach childbearing age and are clinically asymptomatic. Most of them have marked hyperphenylalaninemia, because dietary treatment is discontinued after they reach adulthood. Between 75% and 90% of children born to such women have intellectual disability and are microcephalic, and 15% have congenital heart disease, even though the infants themselves are heterozygotes. This syndrome, termed maternal PKU, results from the teratogenic effects of phenylalanine or its metabolites that cross the placenta and affect specific fetal organs during development. The presence and severity of the fetal anomalies directly correlate with the maternal phenylalanine level, so it is imperative that maternal dietary restriction of phenylalanine be initiated before conception and continued throughout pregnancy.

The biochemical abnormality in PKU is an inability to convert phenylalanine into tyrosine. In normal children, less than 50% of the dietary intake of phenylalanine is necessary for protein synthesis. The remainder is converted to tyrosine by the phenylalanine hydroxylase system (Fig. 10.14). When phenylalanine metabolism is blocked because of a lack of functional PAH, minor shunt pathways come into play, yielding several off-pathway metabolites that are

Table 10.2 Abnormalities Suggesting Inborn Errors of Metabolism

General
Dysmorphic features
Deafness
Self-mutilation
Abnormal hair
Abnormal body or urine odor ("sweaty feet"; "mousy or musty"; "maple syrup")
Hepatosplenomegaly; cardiomegaly
Hydrops
Neurologic
Hypotonia or hypertonia
Coma
Persistent lethargy
Seizures
Gastrointestinal
Poor feeding
Recurrent vomiting
Jaundice
Eyes
Cataracts
Cherry red macula
Dislocated lens
Glaucoma
Muscle, Joints
Myopathy
Abnormal mobility

Modified from Barness LA, Gilbert-Barness E: Metabolic diseases. In Gilbert-Barness E, Barness LA, editors: *Potter's Pathology of the Fetus, Infant, and Child,* St, Louis, 2007, Mosby.

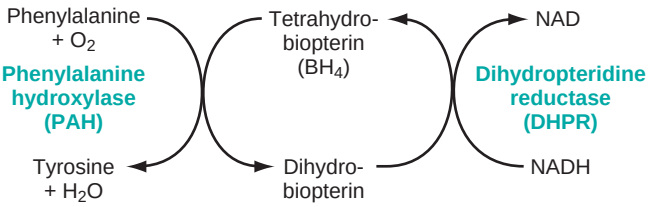


Figure 10.14 The phenylalanine hydroxylase *(PAH)* system. Deficiency of PAH or dihydropteridine reductase *(DHPR)* can give rise to phenylketonuria. *NAD,* Nicotinamide adenine dinucleotide (oxidized); *NADH,* nicotinamide adenine dinucleotide (reduced).

excreted in large amounts in the urine and sweat. These impart a strong musty or mousy odor to affected infants. It is believed that excess phenylalanine or its metabolites contribute to the brain damage in PKU. Concomitant lack of tyrosine (see Fig. 10.14), a precursor of melanin, is responsible for the light color of hair and skin.

At the molecular level, over 500 mutant alleles of the *PAH* gene have been identified. Mutations in both *PAH* alleles are required to develop the disease. Infants with mutations resulting in severely reduced PAH activity present with markedly elevated blood phenylalanine levels and the classic features of PKU, and those with up to 6% residual PAH activity present with milder disease. Some PAH mutations result in only modest elevations of blood phenylalanine levels without associated neurologic damage. This latter condition, referred to as benign hyperphenylalaninemia, is important to recognize because these individuals may have positive screening tests but do not develop the stigmata of classic PKU. Because of the numerous disease-causing alleles of the *PAH* gene, molecular diagnosis is not feasible, and measurement of serum phenylalanine levels is necessary to differentiate benign hyperphenylalaninemia from PKU; the levels in the latter are typically fivefold or more above normal. Once a biochemical diagnosis is established, the specific mutation causing PKU can be determined. With the identification of the mutation, carrier testing of at-risk family members can be performed.

Although 98% of PKU is attributable to mutations in *PAH*, approximately 2% occur due to abnormalities in synthesis or recycling of the cofactor tetrahydrobiopterin (BH_4; see Fig. 10.14). It is clinically important to recognize these variant forms of PKU because they cannot be treated by dietary restriction of phenylalanine.

Galactosemia

Galactosemia is an autosomal recessive disorder of galactose metabolism resulting from accumulation of galactose-1-phosphate in tissues. It affects 1 in 60,000 live-born infants. Normally, lactose, the major carbohydrate of mammalian milk, is split into glucose and galactose in the intestinal microvilli by lactase. Galactose is then converted to glucose in several steps catalyzed by distinct enzymes. Two variants of galactosemia have been identified. In the more common variant, there is a total lack of galactose-1-phosphate uridyl transferase (GALT). The rare variant arises from a deficiency of galactokinase. Because galactokinase deficiency leads to a milder form of the disease that is not associated with intellectual disability, it is not considered in this discussion.

As a result of GALT deficiency, galactose-1-phosphate accumulates in many locations, including the liver, spleen, lens of the eye, kidneys, heart muscle, cerebral cortex, and erythrocytes. Alternative metabolic pathways are activated, leading to the production of galactitol (a polyol metabolite of galactose) and galactonate, an oxidized by-product of excess galactose, both of which also accumulate in the tissues. Long-term toxicity in galactosemia has been variously imputed to these metabolic intermediates. Heterozygotes may have a mild enzyme deficiency and are spared the clinical and pathologic consequences of the homozygous state.

The clinical picture is variable, probably reflecting the heterogeneity of mutations in the *GALT* gene. The liver, eyes, and brain bear the brunt of the damage. The early-to-develop hepatomegaly is due largely to fatty change, but in time widespread scarring that closely resembles the cirrhosis of alcohol abuse may supervene. Opacification of the lens (cataract) develops, probably because the lens absorbs water and swells as galactitol, produced by alternative metabolic pathways, accumulates and increases osmotic pressure. Nonspecific alterations appear in the CNS, including loss of nerve cells, gliosis, and edema, particularly in the dentate nuclei of the cerebellum and the olivary nuclei of the medulla. Similar changes may occur in the cerebral cortex and white matter.

These infants fail to thrive almost from birth. Vomiting and diarrhea appear within a few days of milk ingestion. Jaundice and hepatomegaly usually become evident during the first week of life and may seem to be a continuation of the physiologic jaundice of the newborn. Cataracts develop within a few weeks, and within the first 6 to 12 months of life intellectual disability may be detected. Even in untreated infants, however, the disability is usually not as severe as that seen in PKU. Accumulation of galactose and galactose-1-phosphate in the kidney impairs amino acid transport, resulting in aminoaciduria. There is an increased frequency of fulminant *Escherichia coli* septicemia, possibly arising from depressed neutrophil bactericidal activity. Hemolysis and coagulopathy in the newborn period can occur as well. Newborn screening tests are widely utilized in the United States. They depend on fluorometric assay of GALT enzyme activity on a dried blood spot. A positive screening test must be confirmed by assay of GALT levels in red cells. Antenatal diagnosis is possible by assay of GALT activity in cultured amniotic fluid cells or determination of galactitol level in amniotic fluid supernatant.

Many of the clinical and morphologic changes of galactosemia can be prevented or ameliorated by early removal of galactose from the diet for at least the first 2 years of life. Control instituted soon after birth prevents the cataracts and liver damage and permits almost normal development. Even with dietary restrictions, however, it is now established that older patients are frequently affected by a speech disorder and gonadal failure (especially premature ovarian failure) and, less commonly, ataxia.

Cystic Fibrosis (Mucoviscidosis)

Cystic fibrosis is an inherited disorder of ion transport that affects fluid secretion in exocrine glands and in the epithelial lining of the respiratory, gastrointestinal, and

reproductive tracts. In many individuals this disorder leads to abnormally viscous secretions that obstruct organ passages, resulting in most of the clinical features of this disorder, such as chronic lung disease secondary to recurrent infections, pancreatic insufficiency, steatorrhea, malnutrition, hepatic cirrhosis, intestinal obstruction, and male infertility. These manifestations may appear at any point in life from before birth to much later in childhood or even in adolescence.

Cystic fibrosis is the most common lethal genetic disease that affects Caucasian populations, with an incidence of approximately 1 in 2500 live births. The carrier frequency in the United States is 1 in 20 among Caucasians but significantly lower in African Americans, Asians, and Hispanics. Although cystic fibrosis follows an autosomal recessive transmission pattern, recent data suggest that even heterozygote carriers have a higher incidence of respiratory and pancreatic disease compared with the general population. In addition, despite the classification of cystic fibrosis as a "Mendelian" disorder, there is a wide degree of phenotypic variation that results from three additional factors: diverse mutations in the gene associated with cystic fibrosis, the tissue-specific effects of the encoded gene product, and the influence of modifier genes.

Cystic Fibrosis Gene: Normal Structure and Function

The primary defect in cystic fibrosis is abnormal transport of chloride and bicarbonate ions mediated by an anion channel encoded by the cystic fibrosis transmembrane conductance regulator *(CFTR)* gene on chromosome 7q31.2. The 1480–amino acid polypeptide encoded by *CFTR* has two transmembrane domains (each containing six α-helices), two cytoplasmic nucleotide-binding domains (NBDs), and a regulatory domain (R domain) that contains protein kinase A and C phosphorylation sites (Fig. 10.15). The two transmembrane domains form a channel through which chloride passes. Activation of the CFTR channel is mediated by agonist-induced increases in cyclic adenosine monophosphate (cAMP), which lead to activation of a protein kinase A that phosphorylates the R domain. Adenosine triphosphate (ATP) binding and hydrolysis occurs at the NBD and is essential for opening and closing of the channel pore, which is enhanced by the presence of phosphorylated R domain. Several important facets of CFTR function have emerged in recent years:

- *CFTR regulates multiple ion channels and cellular processes.* Although initially characterized as a chloride-conductance channel, it is now recognized that CFTR can regulate additional ion channels and cellular processes. These include so-called outwardly rectifying chloride channels, inwardly rectifying potassium channels (Kir6.1), the epithelial sodium channel (ENaC), gap junction channels, and cellular processes involved in ATP transport and mucus secretion. Of these, the interaction of CFTR with the ENaC has possibly the most pathophysiologic relevance in cystic fibrosis. The ENaC is situated on the apical surface of epithelial cells and is responsible for sodium uptake from the luminal fluid, rendering the luminal fluid hypotonic. ENaC is inhibited by normally functioning CFTR; hence, in cystic fibrosis, ENaC activity increases, markedly augmenting sodium uptake across the apical membrane. The importance of this phenomenon is discussed later in the context of pulmonary and gastrointestinal pathology in cystic fibrosis. The one exception to this is the human sweat ducts, where ENaC activity decreases as a result of *CFTR* mutations; therefore, hypertonic fluid with high sodium chloride (the *sine qua non* of classic cystic fibrosis) is formed. This is the basis for the "salty" sweat that mothers can often detect in their affected infants.
- *The functions of* CFTR *are tissue-specific; therefore, the impact of a mutation in* CFTR *is also tissue-specific.* The major function of CFTR in the sweat gland ducts is to reabsorb luminal chloride ions and augment sodium reabsorption via the ENaC (see earlier). Therefore, in the sweat ducts, loss of CFTR function leads to decreased reabsorption of sodium chloride and the production of hypertonic sweat (Fig. 10.16). However, in the respiratory and intestinal epithelium, CFTR is one of the most important avenues for active luminal secretion of chloride. At these sites, *CFTR* mutations result in loss or reduction of chloride secretion into the lumen (Fig. 10.16), and active luminal sodium absorption is increased due to loss of inhibition of ENaC activity. These changes in ion distribution result in increased passive water reabsorption from the lumen, lowering the water content of the surface fluid

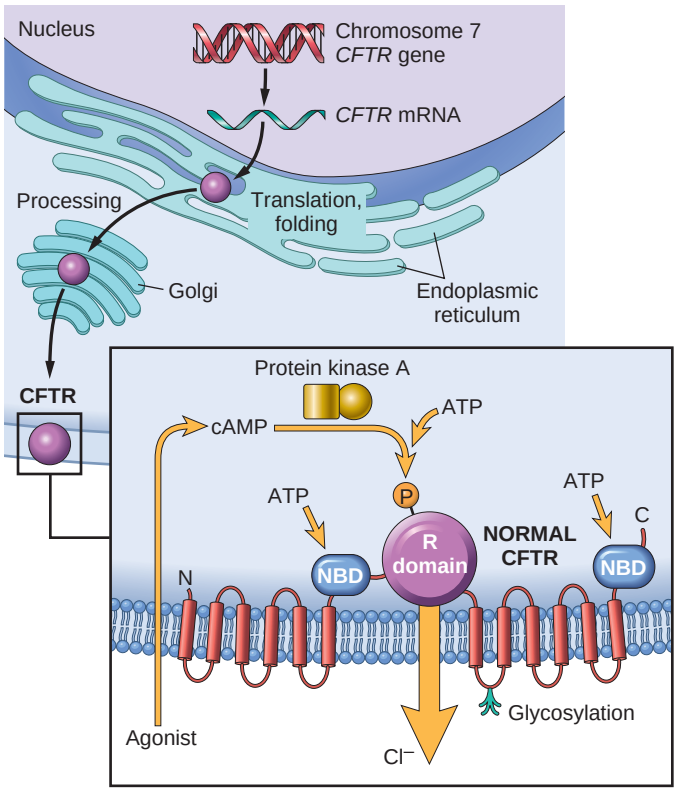


Figure 10.15 *Top*, CFTR from gene to protein. *Bottom*, Normal cystic fibrosis transmembrane conductance regulator (CFTR) structure and activation. CFTR consists of two transmembrane domains, two nucleotide-binding domains (NBDs), and a regulatory R domain. Agonists (e.g., acetylcholine) bind to epithelial cells and increase levels of cyclic adenosine monophosphate (cAMP), which activates protein kinase A; the latter phosphorylates the CFTR R domain using ATP. This results in opening of the chloride ion channel.

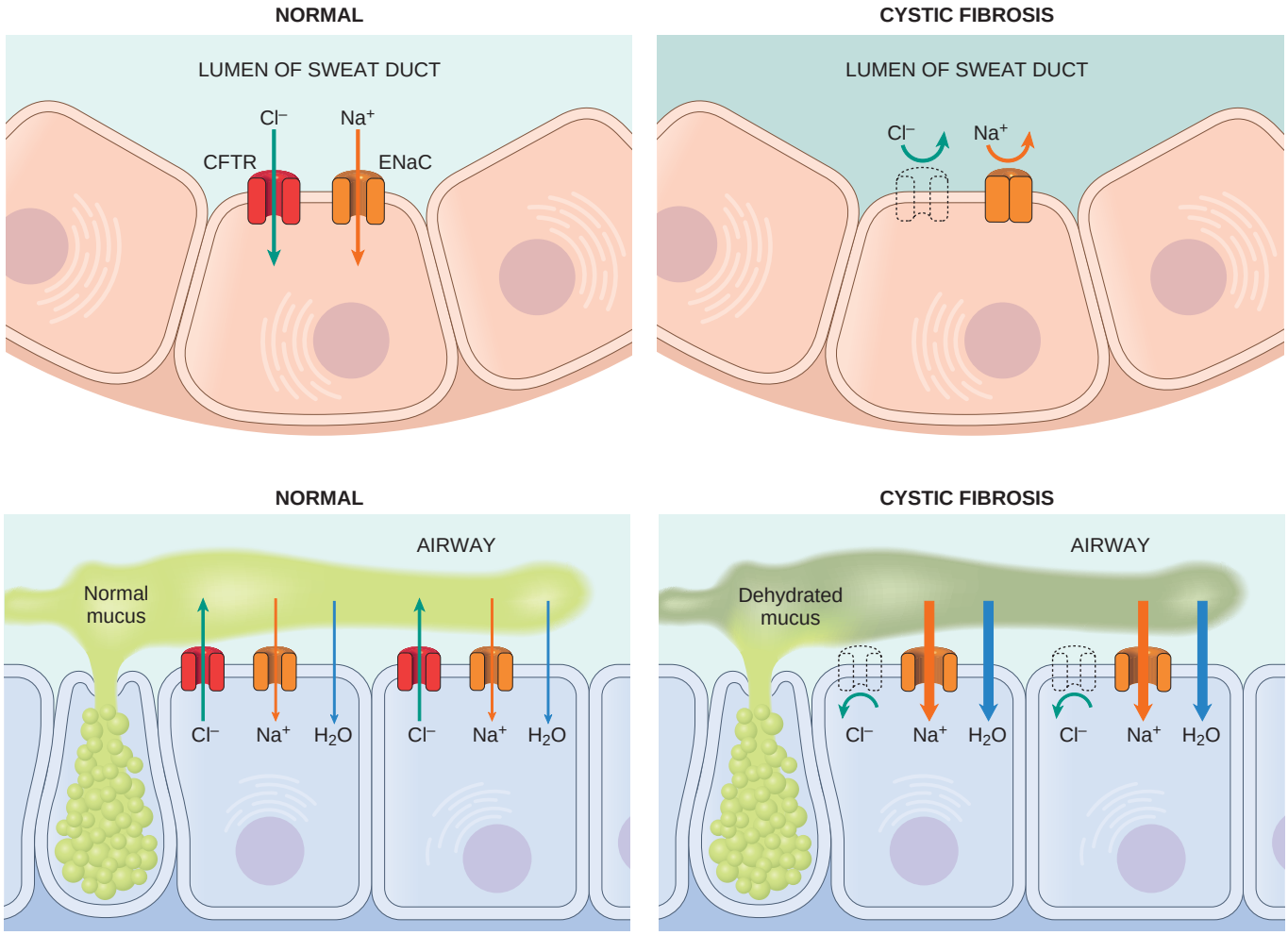


Figure 10.16 CFTR chloride channel defect in the sweat ducts *(top)* causes increased sodium chloride concentration in sweat. In the airway *(bottom)*, patients with cystic fibrosis have decreased chloride secretion and increased sodium and water reabsorption, leading to dehydration of the mucus layer coating epithelial cells, defective mucociliary action, and mucus plugging of airways. *CFTR*, Cystic fibrosis transmembrane conductance regulator; *ENaC*, epithelial sodium channel.

layer coating mucosal cells. Thus, unlike in the sweat ducts, there is no difference in the salt concentration of the surface fluid layer coating respiratory and intestinal mucosal cells in those with cystic fibrosis when compared with normal individuals. Instead, the pathogenesis of respiratory and intestinal complications in cystic fibrosis seems to stem from an isotonic but low-volume surface fluid layer. In the lungs, this dehydration leads to defective mucociliary action and the accumulation of hyperconcentrated, viscid secretions that obstruct the air passages and predispose to recurrent pulmonary infections.

- CFTR *regulates transport of bicarbonate ions.* CFTR plays both a direct and indirect role in regulating bicarbonate transport across the apical epithelial membrane. The anion channel of CFTR is not entirely specific for chloride and has been found to be permeable to bicarbonate as well. In addition, CFTR has reciprocal interactions with the SLC26 family of anion exchangers, which are co-expressed on the apical surface with CFTR. In some *CFTR* mutants, chloride transport is completely or substantially preserved, and bicarbonate transport is markedly abnormal. Normal tissue secretes alkaline fluids; in contrast, fluids that are acidic (due to absence of bicarbonate ions) are secreted by epithelia harboring these mutant *CFTR* alleles. The acidity of secretions results in decreased luminal pH that can lead to a variety of adverse effects such as increased mucin precipitation and plugging of ducts and increased binding of bacteria to plugged mucins. Pancreatic insufficiency, a feature of classic cystic fibrosis, is virtually always present when there are *CFTR* mutations with abnormal bicarbonate conductance.

Cystic Fibrosis Gene: Mutational Spectrum and Genotype-Phenotype Correlation

Since the *CFTR* gene was cloned in 1989, more than 2000 disease-associated mutations have been identified; only five mutations have a relative frequency of 1% or more of disease-causing genotypes. Most of the mutations are missense alterations but frameshift, splicing, and nonsense mutations are also present. The mutations can be grouped into six classes based on their effect on the CFTR protein (Fig. 10.17). Class I to III mutations result in less than 10% residual CFTR activity and are termed "severe" mutations,

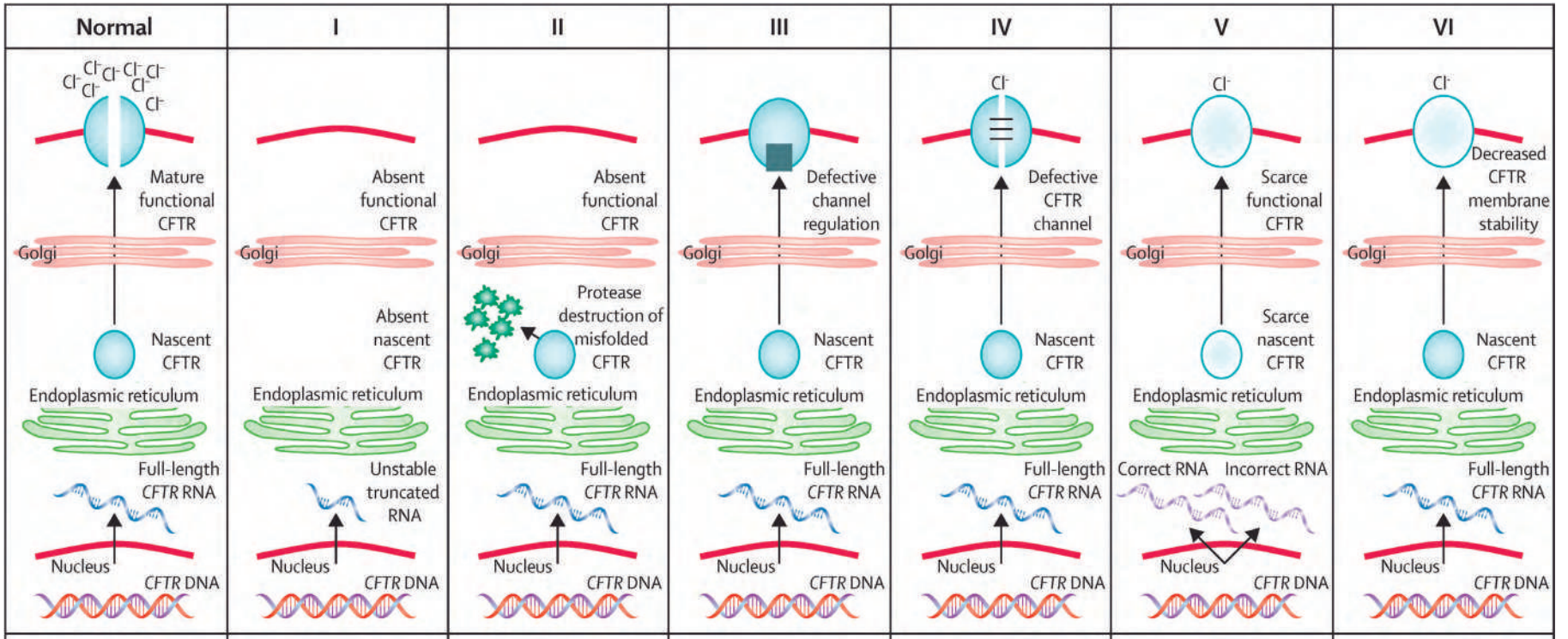


Figure 10.17 Classes of *CFTR* mutations. Mutations in the cystic fibrosis transmembrane conductance regulator *(CFTR)* gene can be divided into six classes. Class I mutations result in no protein production. Class II mutations (including the most prevalent, Phe508del or ΔF508) cause retention of a misfolded protein at the endoplasmic reticulum, and subsequent degradation in the proteasome. Class III mutations affect channel regulation, impairing channel opening (e.g., Gly551Asp). Class IV mutants show reduced conduction—i.e., decreased flow of ions (e.g., Arg117His). Class V mutations cause substantial reduction in mRNA or protein, or both. Class VI mutations cause substantial plasma membrane instability. (Reproduced from Elborn JS: Cystic fibrosis, *Lancet* 388:2519–2531, 2016, from Boyle MP, De Boeck K: A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect, *Lancet Respir Med* 1:158–163, 2013.)

and class IV to VI mutations result in less than 20% residual CFTR activity and are termed "mild" mutations.

- Class I: *Defective protein synthesis (null mutations).* These mutations result in premature termination of protein translation and are associated with complete lack of CFTR protein at the apical surface of epithelial cells.
- Class II: *Abnormal protein folding, processing, and trafficking (processing mutations).* These mutations result in defective processing of the protein from the endoplasmic reticulum to the Golgi apparatus; the protein does not become fully folded and glycosylated and is instead degraded before it reaches the cell surface. The most common *CFTR* mutation is a class II caused by deletion of three nucleotides coding for phenylalanine at amino acid position 508 (ΔF508). Worldwide, this mutation can be found in approximately 70% of Caucasian cystic fibrosis patients. Class II mutations are associated with near-complete lack of CFTR protein at the apical surface of epithelial cells.
- Class III: *Defective regulation (gating mutations).* Mutations in this class prevent activation of CFTR by abrogating ATP binding and hydrolysis, an essential prerequisite for ion transport (see earlier). Thus, there is a normal amount of CFTR on the apical surface, but it is nonfunctional.
- Class IV: *Decreased conductance (conduction mutations).* These mutations typically occur in the transmembrane domain of *CFTR*, which forms the ionic pore for chloride transport. There is a normal amount of CFTR at the apical membrane, but with reduced chloride conduction.
- Class V: *Reduced abundance (production mutations).* These mutations typically affect intronic splice sites or the *CFTR* promoter, such that there is a reduced amount of normal protein.
- Class VI: *Decreased membrane CFTR stability (instability mutations).* These mutations give rise to fully processed and functional proteins but with greatly reduced membrane stability.

Because cystic fibrosis is an autosomal recessive disease, affected individuals harbor mutations on both alleles. However, the nature of mutations on each of the two alleles can have a remarkable effect on the overall phenotype, as well as on organ-specific manifestations (Fig. 10.18). Thus, two "severe" mutations that produce virtual absence of membrane CFTR function are associated with the classic cystic fibrosis phenotype (pancreatic insufficiency, sinopulmonary infections, and gastrointestinal symptoms), and the presence of a "mild" mutation on one or both alleles results in a less severe phenotype. This general dictum of genotype-phenotype correlation is most consistent for pancreatic disease, wherein the presence of one allele with a "mild" mutation associated with some CFTR activity can prevent the pancreatic insufficiency that is virtually always seen with homozygosity for "severe" mutations. In contrast, genotype-phenotype correlations are far less consistent in pulmonary disease, due to the effect of secondary modifiers (see later). As genetic testing for *CFTR* mutations has expanded, it has become evident that some patients who present with clinical features apparently unrelated to cystic fibrosis may also harbor *CFTR* mutations. These include individuals with idiopathic chronic pancreatitis, late-onset chronic pulmonary disease, idiopathic bronchiectasis, and obstructive azoospermia caused by bilateral absence of the vas deferens (see detailed discussion of individual phenotypes later). Most of these patients do not demonstrate other features of cystic fibrosis, despite the presence of bi-allelic *CFTR* mutations; these patients are classified as having nonclassic or atypical cystic fibrosis. Identifying these individuals is important not only for subsequent management, but also for genetic counseling.

Genetic and Environmental Modifiers

Although cystic fibrosis remains one of the best-known examples of the "one gene, one disease" axiom, there is considerable evidence that genes other than *CFTR* modify the frequency and severity of certain organ-specific manifestations, especially pulmonary manifestations and neonatal meconium ileus. Not surprisingly, polymorphisms in genes whose products modulate neutrophil function in response to bacterial infections act as modifier loci for the severity of pulmonary disease in cystic fibrosis. Examples of such modifier genes include mannose binding lectin 2 *(MBL2)*, transforming growth factor β1 *(TGF-β1)*, and interferon-related developmental regulator 1 *(IFRD1)*. It is postulated that polymorphisms in these genes regulate the

Level of CFTR function	<5%	<10%	<20%	50%
Sinuses	Severe chronic sinusitis	Moderate chronic sinusitis		Increased rate of chronic sinusitis
Lungs	Severe lung disease	Variable lung disease		Increased rate of lung disease
Sweat glands	High sweat chloride	Intermediate sweat chloride		
Pancreas	Pancreatic insufficiency	Pancreatic sufficiency		
Intestines	Meconium ileus	Distal intestinal obstruction syndrome		
Vas deferens	Absent vas deferens			Increased risk of absent vas deferens

Figure 10.18 The many clinical manifestations of mutations in the cystic fibrosis gene, from most severe to asymptomatic. (Modified from Chang EH and Zabner J: Precision genomic medicine in cystic fibrosis, *Clin Transl Sci* 8(5):606–610, 2015.)

resistance of the lungs to exogenous infections with virulent microbes (see later), thus modifying the natural history of cystic fibrosis.

Environmental modifiers may also explain some of the significant phenotypic differences between individuals who share the same *CFTR* genotype. This is best exemplified in pulmonary disease, where *CFTR* genotype-phenotype correlations can be perplexing. As stated earlier, defective mucociliary action because of deficient hydration of the mucus results in an inability to clear bacteria from the airways. *Pseudomonas aeruginosa* species, in particular, colonize the lower respiratory tract, first intermittently and then chronically. Concurrent viral infections predispose to such colonization. The static mucus creates a hypoxic microenvironment in the airway surface fluid, which in turn favors the production of alginate, a mucoid polysaccharide capsule, by the colonizing bacteria. Alginate production permits the formation of a biofilm that protects the bacteria from antibodies and antibiotics, allowing them to evade host defenses and produce a chronic destructive lung disease. Antibody- and cell-mediated immune reactions induced by the organisms result in further pulmonary destruction, but are ineffective against the organism. Approximately 80% of cystic fibrosis patients are colonized by *Pseudomonas aeruginosa* by 20 years of age. Chronic colonization by these bacteria is a significant contributor to the morbidity and mortality of cystic fibrosis.

It is evident, therefore, that in addition to genetic factors (e.g., class of mutation), a plethora of environmental modifiers (e.g., intercurrent and concurrent infections by microorganisms) can influence the severity and progression of lung disease in cystic fibrosis.

MORPHOLOGY

The anatomic changes are highly variable in distribution and severity. In individuals with nonclassic cystic fibrosis, the disease is quite mild and does not seriously disturb their growth and development. In others, the pancreatic involvement is severe and impairs intestinal absorption because of pancreatic insufficiency (Chapter 19), so malabsorption stunts development and postnatal growth. In others, the mucus secretion defect leads to defective mucociliary action, obstruction of bronchi and bronchioles, and crippling fatal pulmonary infections. In all variants, the **sweat glands are morphologically unaffected.**

Pancreatic abnormalities are present in approximately 85% to 90% of patients with cystic fibrosis. In milder cases, there may be only accumulation of mucus in the small ducts with some dilation of the exocrine glands. In more severe cases, usually seen in older children or adolescents, the ducts are completely plugged, causing atrophy of the exocrine glands and progressive fibrosis (Fig. 10.19). Atrophy of the exocrine portion of the pancreas may occur, leaving only islets within a fibrofatty stroma. The loss of pancreatic exocrine secretion impairs fat absorption, and the associated avitaminosis A may contribute to squamous metaplasia of the pancreatic duct lining epithelium, which is already injured by the inspissated mucus secretions. Thick viscid plugs of mucus may also be found in the small intestine of infants. Sometimes these cause small-bowel obstruction, known as **meconium ileus.**

Liver involvement follows the same basic pattern. Bile canaliculi are plugged by mucus material, accompanied by ductular proliferation and portal inflammation. Hepatic **steatosis** is not an uncommon finding in liver biopsies. Over time, **focal biliary cirrhosis** develops in approximately one-third of patients (Chapter 18), which can eventually involve the entire liver, resulting in diffuse hepatic nodularity. Such severe hepatic involvement is encountered in less than 10% of patients.

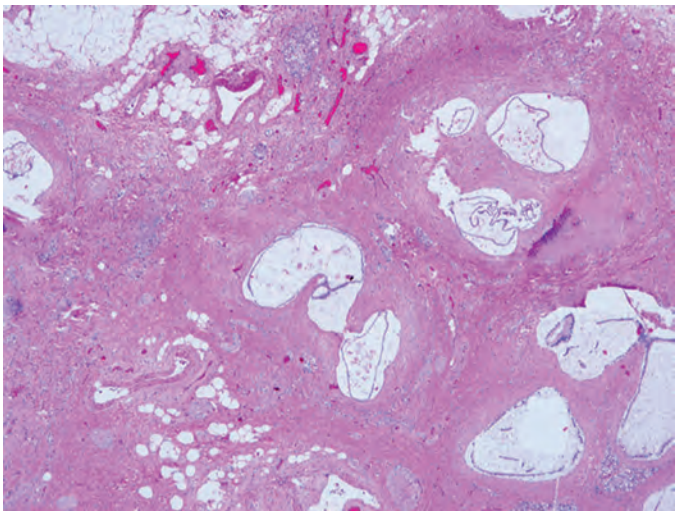


Figure 10.19 End-stage pancreatic disease in cystic fibrosis. The ducts are markedly dilated, and exocrine glands are destroyed and replaced by fibrous tissue.

The **salivary glands** frequently show histologic changes similar to those described in the pancreas: progressive dilation of ducts, squamous metaplasia of the lining epithelium, and glandular atrophy followed by fibrosis.

The **pulmonary changes** are the most serious complications of this disease (Fig. 10.20). These stem from the viscous mucus secretions of the submucosal glands of the respiratory tree, leading to secondary obstruction and infection of the air passages. The bronchioles are often distended with thick mucus associated with marked hyperplasia and hypertrophy of the mucus-secreting cells. Superimposed infections give rise to severe chronic bronchitis and bronchiectasis (Chapter 15). In many instances, lung abscesses develop. *Staphylococcus aureus, Haemophilus influenzae,* and *Pseudomonas aeruginosa* are the three most common organisms responsible for lung infections. As mentioned earlier, a mucoid form of *P. aeruginosa* (alginate-producing) is particularly frequent and causes chronic inflammation. Even more sinister is the increasing frequency of infection with another group of pseudomonads, the *Burkholderia cepacia* complex, which includes at least nine different species; of these, infections with *B. cenocepacia* are the most common in cystic fibrosis patients. This opportunistic bacterium is particularly hardy, and infection with this organism has been associated with fulminant illness ("cepacia syndrome"), longer hospital stays, and increased mortality. Other opportunistic bacterial pathogens include *Stenotrophomonas maltophilia* and nontuberculous mycobacteria; allergic bronchopulmonary aspergillosis also occurs with increased frequency in cystic fibrosis.

Azoospermia and infertility are found in 95% of males who survive to adulthood; **congenital bilateral absence of the vas deferens** is a frequent finding in these patients. In some males, bilateral absence of the vas deferens may be the only feature suggesting an underlying *CFTR* mutation.

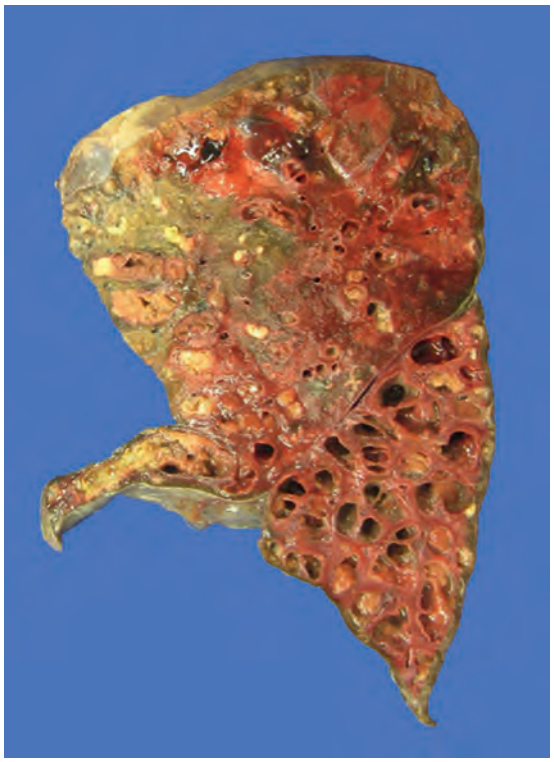


Figure 10.20 Lungs of a patient who died of cystic fibrosis. There is extensive mucus plugging and dilation of the tracheobronchial tree. The pulmonary parenchyma is consolidated by a combination of both secretions and pneumonia—the green color associated with *Pseudomonas* infection.

Clinical Features

Few childhood diseases are as protean as cystic fibrosis in clinical manifestations (Table 10.3; Fig. 10.18). The symptoms are extremely varied and may appear at birth to years later, and involve one organ system or many. Approximately 5% to 10% of cases come to clinical attention at birth or soon after because of meconium ileus. Distal intestinal obstruction can also occur in older individuals, manifesting as recurrent episodes of right lower quadrant pain sometimes associated with a palpable mass of meconium, with or without associated intussusception, in the right iliac fossa.

Exocrine pancreatic insufficiency occurs in the majority (85% to 90%) of patients with cystic fibrosis and is associated with "severe" *CFTR* mutations on *both* alleles (e.g., *ΔF508/ΔF508*), whereas 10% to 15% of patients with one "severe" and one "mild" *CFTR* mutation (e.g., *ΔF508/R117H*) or two "mild" *CFTR* mutations retain enough pancreatic exocrine function so as not to require enzyme supplementation (pancreas-sufficient phenotype). Pancreatic insufficiency is associated with protein and fat malabsorption and increased fecal loss. Manifestations of malabsorption (e.g., large foul-smelling stools, abdominal distention, and poor weight gain) may appear during the first year of life. The faulty fat absorption may induce deficiency of the fat-soluble vitamins, resulting in manifestations of avitaminosis A, D, E, or K. Hypoproteinemia may be severe enough to cause generalized edema. Persistent diarrhea may result in rectal prolapse in up to 10% of children with this disease. The pancreas-sufficient phenotype is usually not associated with other gastrointestinal complications, and, in general, these individuals demonstrate excellent growth and development. A subset of patients with pancreas-sufficient cystic fibrosis have recurrent bouts of pancreatitis associated with acute abdominal pain and occasionally, life-threatening complications. These patients have other features of classic cystic fibrosis, such as pulmonary disease. "Idiopathic" chronic pancreatitis can also occur as an isolated late-onset finding in the absence of other stigmata of cystic fibrosis (Chapter 19); bi-allelic *CFTR* mutations (usually one "mild," one "severe") are demonstrable in the majority of these individuals who have nonclassic or atypical cystic fibrosis. Endocrine pancreatic insufficiency (i.e., diabetes mellitus) occurs in up to 50% of adults with cystic fibrosis and is thought to be caused by severe destruction of pancreatic parenchyma including the islets.

Cardiorespiratory complications, such as persistent lung infections, obstructive pulmonary disease, and cor pulmonale, are the most common cause of death (~80%) in cystic fibrosis

Table 10.3 Clinical Features and Diagnostic Criteria for Cystic Fibrosis

Clinical Features of Cystic Fibrosis
1. Chronic sinopulmonary disease manifested by
a. Persistent colonization/infection with typical cystic fibrosis pathogens, including *Staphylococcus aureus*, nontypeable *Haemophilus influenzae*, mucoid and nonmucoid *Pseudomonas aeruginosa, Burkholderia cepacia*
b. Chronic cough and sputum production
c. Persistent chest radiograph abnormalities (e.g., bronchiectasis, atelectasis, infiltrates, hyperinflation)
d. Airway obstruction manifested by wheezing and air trapping
e. Nasal polyps; radiographic or computed tomographic abnormalities of paranasal sinuses
f. Digital clubbing
2. Gastrointestinal and nutritional abnormalities, including
a. Intestinal: meconium ileus, distal intestinal obstruction syndrome, rectal prolapse
b. Pancreatic: pancreatic insufficiency, recurrent acute pancreatitis, chronic pancreatitis
c. Hepatic: chronic hepatic disease manifested by clinical or histologic evidence of focal biliary cirrhosis or multilobular cirrhosis, prolonged neonatal jaundice
d. Nutritional: failure to thrive (severe acute malnutrition), hypoproteinemia, edema, complications secondary to fat-soluble vitamin deficiency
3. Salt-loss syndromes: acute salt depletion, chronic metabolic alkalosis
4. Male urogenital abnormalities resulting in obstructive azoospermia (congenital bilateral absence of vas deferens)
Criteria for Diagnosis of Cystic Fibrosis
One or more characteristic phenotypic features, OR a history of cystic fibrosis in a sibling, parent, or child, OR a positive newborn screening test result AND An increased sweat chloride concentration, OR a sweat chloride concentration in the intermediate range AND identification of two disease-causing *CFTR* mutations, OR a sweat chloride concentration in the intermediate range AND demonstration of abnormal epithelial nasal ion transport (nasal potential difference or intestinal current measurement)

Modified with permission from Farrell PM, White TB, Ren CL, et al: Diagnosis of cystic fibrosis: consensus guidelines from the Cystic Fibrosis Foundation, *J Pediatr* 181S:S4–S15, 2017.

patients in the United States. By 18 years of age, 80% of patients with classic cystic fibrosis harbor *P. aeruginosa,* and 3.5% harbor *B. cepacia*. With the indiscriminate use of antibiotic prophylaxis against *Staphylococcus,* there has been an unfortunate resurgence of resistant strains of *Pseudomonas* in many patients. Individuals who carry one "severe" and one "mild" *CFTR* mutation may develop late-onset mild pulmonary disease, another example of nonclassic or atypical cystic fibrosis. Patients with mild pulmonary disease usually have little or no pancreatic disease. Adult-onset "idiopathic" bronchiectasis has been linked to *CFTR* mutations in a subset of cases. Recurrent sinonasal polyps can occur in 10% to 25% of individuals with cystic fibrosis; hence, children who present with this finding should be tested for cystic fibrosis.

Significant *liver disease* occurs late in the natural history of cystic fibrosis and is gaining in clinical importance as life expectancies increase. In fact, after cardiopulmonary and transplantation-related complications, liver disease is the most common cause of death in cystic fibrosis. Most studies suggest that symptomatic or biochemical liver disease has its onset at or around puberty, with a prevalence of approximately 13% to 17%. However, asymptomatic hepatomegaly may be present in up to one-third of individuals. Obstruction of the common bile duct may occur due to stones or sludge; it presents with abdominal pain and the acute onset of jaundice. As previously noted, diffuse biliary cirrhosis develops in less than 10% of individuals with cystic fibrosis.

Approximately 95% of males with cystic fibrosis are infertile as a result of *obstructive azoospermia.* As mentioned earlier, this is most commonly due to congenital bilateral absence of the vas deferens, which is caused in 80% of cases by bi-allelic *CFTR* mutations.

In most cases, the diagnosis of cystic fibrosis is based on persistently elevated sweat electrolyte concentrations (often the mother makes the diagnosis by recognizing her infant's abnormally salty sweat), characteristic clinical findings (sinopulmonary disease and gastrointestinal manifestations), an abnormal newborn screening test, or a family history. The most common newborn screening test is based on measuring the blood level of immunoreactive trypsinogen, which is produced by the pancreas, and elevated levels result from pancreatic injury. A minority of patients with cystic fibrosis, especially those with at least one "mild" *CFTR* mutation, may have a normal or near-normal sweat test (<60 mM). Measurement of nasal transepithelial potential difference in vivo can be a useful adjunct under these circumstances; individuals with cystic fibrosis demonstrate a significantly more negative baseline nasal potential difference than controls. In patients with a positive screening test, suggestive clinical findings, or family history (or more than one of these), genetic analysis is warranted. Sequencing the *CFTR* gene is the gold standard for diagnosis of cystic fibrosis.

There have been major improvements in the management of acute and chronic complications for cystic fibrosis, including more potent antimicrobial therapies, pancreatic enzyme replacement, and bilateral lung transplantation. New treatment modalities for restoring mutant CFTR function are being tested in clinical trials. Based on the molecular defect, three classes of agents are being developed:

- *Potentiators.* These agents keep the "gate" of the CFTR channel open. Hence they are most useful in gating (class III) and conduction (class IV) mutations. They are also helpful for patients with production mutations (class V), by improving the function of reduced amounts of CFTR.
- *Correctors.* These agents assist in proper folding of the CFTR protein, thereby increasing its trafficking to the cell surface. Hence they have the potential to help patients with processing (class II) mutations. These include patients with ΔF508, the most common *CFTR* mutation.
- *Amplifiers.* These agents increase the amount of CFTR protein the cell makes and thus can be useful for patients with production (class V) mutations.

A combination therapy consisting of a "potentiator" agent and a small molecule that is a CFTR "corrector" has been approved for the treatment of patients with two copies of ΔF508, the most common *CFTR* mutation. For treatment of patients with class I nonsense mutations, molecules that allow "ribosomal read-through" of premature stop codons are in the pipeline. Such therapy is also being tested in muscular dystrophies in which stop codons prevent synthesis of dystrophin (Chapter 27). Overall, improvements in the management of cystic fibrosis have extended the median life expectancy to 50 years.

SUDDEN INFANT DEATH SYNDROME

According to the National Institute of Child Health and Human Development, **SIDS is defined as "the sudden death of an infant under 1 year of age which remains unexplained after a thorough case investigation, including performance of a complete autopsy, examination of the death scene, and review of the clinical history."** It is important to emphasize that many cases of sudden death in infancy may have an unexpected anatomic or biochemical basis discernible at autopsy (Table 10.4), and these should not be labeled as SIDS, but rather as sudden unexpected infant death (SUID). The Centers for Disease Control and Prevention estimates that SIDS accounts for approximately one-half of SUID cases in the United States. An aspect of SIDS that is not stressed in the definition is that the infant usually dies while asleep, mostly in the prone or side position, hence the pseudonyms of crib death or cot death.

Epidemiology

As infant deaths due to nutritional problems and infections have come under control in high income countries, SIDS has assumed greater importance, including in the United States. SIDS is the leading cause of death between 1 month and 1 year of age in this country and the fourth leading cause of death overall in infancy, after congenital anomalies, diseases of prematurity and low birth weight, and maternal complications. Mostly because of nationwide "Back to Sleep" (now "Safe to Sleep") campaigns, there has been a significant drop in SIDS-related mortality, from an estimated 120 deaths per 100,000 live births in 1992 to 35 per 100,000 in 2017. This number translates to about 1360 deaths due to SIDS in the United States. Despite the decline in SIDS and SUID in all races and ethnicities, the rate of SUID among

Table 10.4 Risk Factors and Postmortem Findings Associated With Sudden Infant Death Syndrome

Parental
Young maternal age (younger than 20 years of age)
Maternal smoking during pregnancy
Drug abuse in *either* parent, specifically paternal marijuana and maternal opiate, cocaine use
Short intergestational intervals
Late or no prenatal care
Low socioeconomic group
African-American and American Indian ethnicity (? socioeconomic factors)
Infant
Brainstem abnormalities, associated with delayed development of arousal and cardiorespiratory control
Prematurity and/or low birth weight
Male sex
Product of a multiple birth
SIDS in a prior sibling
Antecedent respiratory infections
Germline polymorphisms in autonomic nervous system genes
Environment
Prone or side sleep position
Sleeping on a soft surface
Hyperthermia
Co-sleeping in first 3 months of life
Postmortem Abnormalities Detected in Cases of Sudden Unexpected Infant Death (SUID)[a]
Infections
Viral myocarditis
Bronchopneumonia
Unsuspected congenital anomaly
Congenital aortic stenosis
Anomalous origin of the left coronary artery from the pulmonary artery
Traumatic child abuse
Intentional suffocation (filicide)
Genetic and metabolic defects
Long QT syndrome (*SCN5A* and *KCNQ1* mutations)
Fatty acid oxidation disorders (*MCAD, LCHAD, SCHAD* mutations)
Histiocytoid cardiomyopathy (*MTCYB* mutations)
Abnormal inflammatory responsiveness (partial deletions in *C4a* and *C4b*)

[a]SIDS is not the only cause of SUIDs, but rather is a diagnosis of exclusion. Therefore, performance of an autopsy may often reveal findings that would explain the cause of an SUID. These cases should *not*, strictly speaking, be labeled as "SIDS."

C4, Complement component 4; *KCNQ1*, potassium voltage-gated channel, KQT-like subfamily, member 1; *LCHAD*, long-chain 3-hydroxyacyl coenzyme A dehydrogenase; *MCAD*, medium-chain acyl coenzyme A dehydrogenase; *MTCYB*, mitochondrial cytochrome b; *SCHAD*, short-chain 3-hydroxyacyl coenzyme A dehydrogenase; *SCN5A*, sodium channel, voltage-gated, type V, alpha subunit.

non-Hispanic black (70 per 100,000 live births) and American Indian/Alaskan Native (77 per 100,000 live births) infants was more than double that of non-Hispanic white infants (35 per 100,000 live births) in 2017. SIDS rates for Asian/Pacific Islander and Hispanic infants were much lower than the rate for non-Hispanic white infants. Differences in the prevalence of supine positioning and other sleep environment conditions between racial and ethnic populations may contribute to these disparities. Worldwide, in countries where unexpected infant deaths are diagnosed as SIDS only after postmortem examination, the death rates from SIDS range from 10 per 100,000 live births in the Netherlands to 80 per 100,000 in New Zealand.

Approximately 90% of all SIDS deaths occur during the first 6 months of life, most between 2 and 4 months of age. This narrow window of peak susceptibility is a unique characteristic that is independent of other risk factors (to be described) and the geographic locale. Most infants who die of SIDS die at home, usually during the night after a period of sleep.

MORPHOLOGY

In infants who have died of suspected SIDS, a variety of findings have been reported at postmortem examination. They are usually subtle and of uncertain significance and are not present in all cases. Multiple petechiae are the most common finding (~80% of cases); these are usually present on the thymus, visceral and parietal pleura, and epicardium. Grossly, the lungs are usually congested, and vascular engorgement with or without pulmonary edema is demonstrable microscopically in the majority of cases. These changes possibly represent agonal events, because they are found with comparable frequencies in explained sudden deaths in infancy. Within the upper respiratory system (larynx and trachea), there may be some histologic evidence of recent infection (correlating with the clinical symptoms), although the changes are not sufficiently severe to account for death and should not detract from the diagnosis of SIDS. The CNS demonstrates astrogliosis of the brainstem and cerebellum. Sophisticated morphometric studies have revealed quantitative brainstem abnormalities such as hypoplasia of the arcuate nucleus or a decrease in brainstem neuronal populations in several cases; these observations are not uniform, however. Nonspecific findings include frequent persistence of hepatic extramedullary hematopoiesis and periadrenal brown fat; it is tempting to speculate that these latter findings relate to chronic hypoxemia, retardation of normal development, and chronic stress. Thus, autopsy usually fails to provide a clear cause of death, and this may well be related to the etiologic heterogeneity of SIDS. The importance of a postmortem examination rests in identifying other causes of SUID, such as unsuspected infection, congenital anomaly, or a genetic disorder (see Table 10.4), the presence of any of which would exclude a diagnosis of SIDS; and in ruling out the unfortunate possibility of traumatic child abuse.

Pathogenesis

The circumstances surrounding SIDS have been explored in great detail, and it is generally accepted that it is a multifactorial condition. **A "triple-risk" model of SIDS has been proposed, which postulates the intersection of three overlapping factors: (1) a vulnerable infant, (2) a critical developmental period in homeostatic control, and (3) an exogenous stressor.** According to this model, several factors make the infant vulnerable to sudden death during the critical developmental period (i.e., the first 6 months of life). These vulnerability factors may relate to the parents, infant, or exogenous environmental stressors (see Table 10.4).

Although numerous factors have been proposed to account for a vulnerable infant, the most compelling hypothesis is that SIDS reflects a delayed development of "arousal" and cardiorespiratory control. The brainstem, and in particular the medulla oblongata, plays a critical role in

the body's "arousal" response to noxious stimuli such as episodic hypercarbia, hypoxia, and thermal stress encountered during sleep. The serotonergic (5-HT) system of the medulla is implicated in these "arousal" responses, as well as regulation of other critical homeostatic functions such as respiratory drive, blood pressure, and upper airway reflexes. Abnormalities in serotonin-dependent signaling in the brainstem may be the underlying basis for SIDS in some infants.

Epidemiologic and genetic studies have identified additional vulnerability factors for SIDS in the "triple-risk" model. Infants who are born before term or who are low birth weight are at increased risk, and risk increases with decreasing gestational age or birth weight. Male sex is associated with a slightly greater incidence of SIDS. SIDS in a prior sibling is associated with a fivefold relative risk of recurrence, highlighting the importance of a genetic predisposition; traumatic child abuse must be carefully excluded under these circumstances. Most SIDS babies have an immediate prior history of a mild respiratory tract infection, but no single causative organism has been isolated. These infections may predispose an already vulnerable infant to even greater impairment of cardiorespiratory control and delayed arousal. In this context, laryngeal chemoreceptors have emerged as a putative "missing link" between upper respiratory tract infections, the prone position, and SIDS. When stimulated, these laryngeal chemoreceptors typically elicit an inhibitory cardiorespiratory reflex. Stimulation of the chemoreceptors is augmented by respiratory tract infections, which increase the volume of secretions, and by the prone position, which impairs swallowing and clearing of the airways, even in healthy infants. In a previously vulnerable infant with impaired arousal, the resulting inhibitory cardiorespiratory reflex may prove fatal. Genetic vulnerability factors in the infant include polymorphisms of genes related to serotonergic signaling and autonomic innervation, pointing to the importance of these processes in the pathophysiology of SIDS.

Maternal smoking during pregnancy has consistently emerged as a risk factor in epidemiologic studies of SIDS, with children exposed to in utero nicotine having more than double the risk of SIDS as compared with children born to nonsmokers. Young maternal age, frequent childbirths, and inadequate prenatal care are all risk factors associated with increased incidence of SIDS in the offspring.

Among the potential "environmental stressors," prone or side sleeping positions, sleeping with parents in the first 3 months, sleeping on soft surfaces, and thermal stress are probably the most important modifiable risk factors for SIDS. The prone or side positions predispose an infant to one or more recognized noxious stimuli (hypoxia, hypercarbia, and thermal stress) during sleep. The side position was considered a reliable alternative to the prone sleeping position, but the American Academy of Pediatrics now recognizes the supine sleeping position as the only safe position that reduces the risk of SIDS. This "Back to Sleep" campaign has resulted in substantial reductions in SIDS-related deaths since its inception in 1994.

SIDS is a diagnosis of exclusion, requiring careful examination of the death scene and a complete postmortem examination. The latter can reveal an unsuspected cause of sudden death in as many as 20% or more of "SIDS" babies. Infections (e.g., viral myocarditis or bronchopneumonia) are the most common causes of sudden "unexpected" death, followed by unsuspected congenital anomalies. In part as a result of advancements in molecular diagnostics and knowledge of the human genome, several genetic causes of sudden "unexpected" infant death have emerged (see Table 10.4). For example, fatty acid oxidation disorders, characterized by defects in mitochondrial fatty acid oxidative enzymes, may be responsible for as many as 5% of SUIDs.

TUMORS AND TUMORLIKE LESIONS OF INFANCY AND CHILDHOOD

Only 2% of all malignant tumors occur in infancy and childhood; nonetheless, cancer (including leukemia) accounts for about 16% of deaths in the United States in children between 5 and 14 years of age, and only accidents cause significantly more deaths. Benign tumors are even more common than cancers. Most benign tumors are of little concern, but on occasion they cause serious complications by virtue of their location or size.

It is sometimes difficult to separate, on morphologic grounds, true tumors or neoplasms from tumorlike lesions in the infant and child. In this context, two special categories of tumorlike lesions should be distinguished from true tumors.

- The term *heterotopia* (or *choristoma*) is applied to microscopically normal cells or tissues that are present in abnormal locations. Examples of heterotopias include a rest of pancreatic tissue found in the wall of the stomach or small intestine, or a small mass of adrenal cells found in the kidney, lungs, or ovaries. These heterotopic rests are usually of little significance, but they can be confused clinically with neoplasms. Rarely, they are sites of origin of true neoplasms, producing paradoxes such as an adrenal carcinoma arising in the ovary.
- The term *hamartoma* refers to an excessive, focal overgrowth of cells and tissues native to the organ in which it occurs. Although the cellular elements are mature and identical to those found in the remainder of the organ, they do not reproduce the normal architecture of the surrounding tissue. The line of demarcation between a hamartoma and a benign neoplasm is often unclear because both lesions can be clonal. Hemangiomas, lymphangiomas, rhabdomyomas of the heart, adenomas of the liver, and developmental cysts within the kidneys, lungs, or pancreas are interpreted by some as hamartomas and by others as true neoplasms. Their unequivocally benign histology, however, does not preclude bothersome and rarely life-threatening clinical problems in some cases.

Benign Tumors and Tumorlike Lesions

Virtually any tumor may be encountered in children, but within this wide array hemangiomas, fibrous lesions, and teratomas deserve special mention. You will notice that the most common neoplasms of childhood are soft-tissue tumors of mesenchymal derivation. This contrasts with adults, in whom the most common tumors, benign or malignant, have an epithelial origin. Benign tumors of various tissues are described in greater detail in appropriate chapters; here a

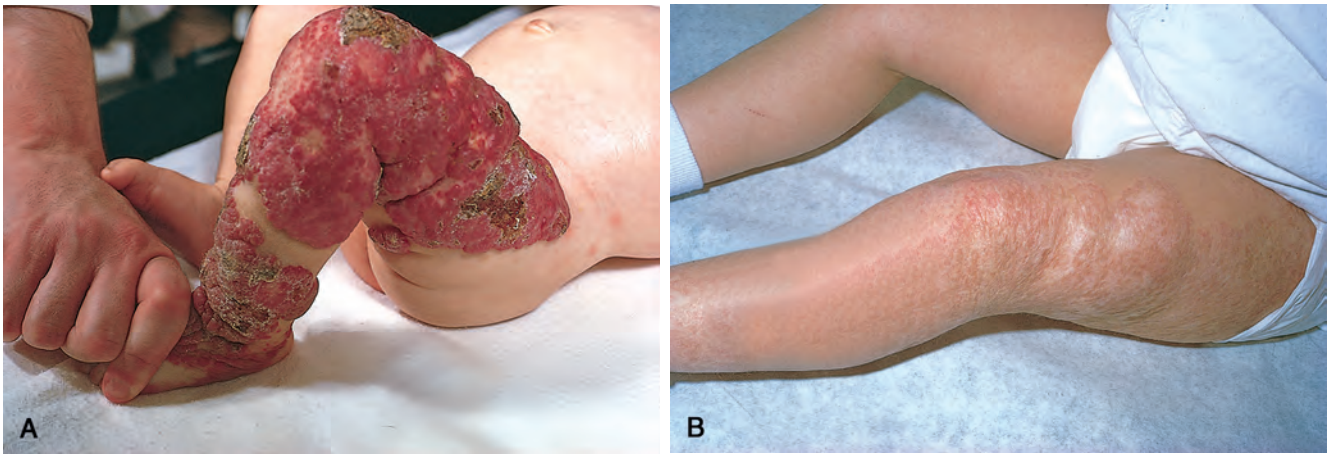


Figure 10.21 Congenital hemangioma at birth (A) and at 2 years of age (B) after spontaneous regression. (Courtesy Dr. Eduardo Yunis, Children's Hospital of Pittsburgh, Pittsburgh, Pa.)

few comments are made about their special features in childhood.

Hemangioma

Hemangiomas are the most common tumors of infancy (Chapter 11). Architecturally, they do not differ from those occurring in adults. Both cavernous and capillary hemangiomas may be encountered, although the latter are often more cellular than in adults, a feature that is deceptively worrisome. In children, most are located in the skin, particularly on the face and scalp, where they produce flat to elevated, irregular, red-blue masses; some of the flat, larger lesions (considered by some to represent vascular ectasias) are referred to as *port-wine stains*. Hemangiomas may enlarge along with the growth of the child, but in many instances they spontaneously regress (Fig. 10.21). In addition to their cosmetic significance, they can represent one facet of the hereditary disorder von Hippel-Lindau disease (Chapter 28). A subset of CNS cavernous hemangiomas can occur in the familial setting; these families harbor mutations in one of three genes (*KRIT1, CCM2,* or *PDCD10*).

Fibrous Tumors

Fibrous tumors occurring in infants and children include sparsely cellular proliferations of spindle-shaped cells (designated as fibromatosis) and richly cellular lesions indistinguishable from fibrosarcomas occurring in adults (designated as congenital-infantile fibrosarcomas). Biologic behavior cannot be predicted based on histology alone; however, despite their histologic similarities with adult fibrosarcomas, the congenital-infantile variants have an excellent prognosis. A characteristic chromosomal translocation, t(12;15)(p13;q25), has been described in congenital-infantile fibrosarcomas, which results in generation of an ETV6-NTRK3 fusion transcript. The normal *ETV6* gene product is a transcription factor, and the *NTRK3* gene product (also known as TRKC) is a tyrosine kinase. Like other tyrosine kinase fusion proteins found in human neoplasms, ETV6-NTRK3 is constitutively active and stimulates signaling through the oncogenic RAS-MAPK and PI3K/AKT pathways (Chapter 7). Among soft-tissue tumors, the *ETV6-NTRK3* fusion transcript is most commonly found in congenital-infantile fibrosarcomas, making it a useful diagnostic marker.

Teratomas

Teratomas illustrate the relationship of histologic maturity to biologic behavior. Teratomas may occur as benign, well-differentiated cystic lesions (mature teratomas), as lesions of indeterminate potential (immature teratomas), or as unequivocally malignant teratomas (usually admixed with another germ cell tumor component such as yolk sac tumor) (Chapter 21). They exhibit two peaks in incidence: the first at approximately 2 years of age and the second in late adolescence or early adulthood. The first peak is congenital neoplasms; the later occurring lesions may also be of prenatal origin but are more slowly growing. Sacrococcygeal teratomas are the most common teratomas of childhood, accounting for 40% or more of cases (Fig. 10.22). They occur with a frequency of 1 in 20,000 to 40,000 live

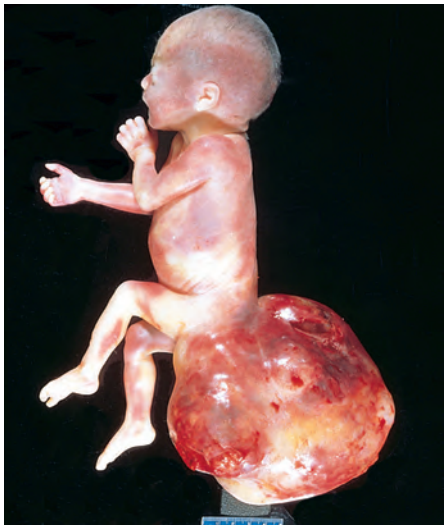


Figure 10.22 Sacrococcygeal teratoma. Note the size of the lesion compared with that of the stillbirth.

births, and they are four times more common in girls than in boys. Approximately 10% of sacrococcygeal teratomas are associated with congenital anomalies, primarily defects of the hindgut and cloacal region and other midline defects (e.g., meningocele, spina bifida) not believed to result from local effects of the tumor. Approximately 75% of these tumors are mature teratomas, and about 12% are unequivocally malignant and lethal. The remainder are immature teratomas; their malignant potential correlates with the amount of immature tissue, usually immature neuroepithelial elements, that are present. Most benign teratomas are encountered in younger infants (<4 months), whereas children with malignant lesions tend to be somewhat older. Other sites for teratomas in childhood include the testis (Chapter 21), ovary (Chapter 22), and various midline locations, such as the mediastinum, retroperitoneum, and head and neck.

Malignant Tumors

Cancers of infancy and childhood differ biologically and histologically from their counterparts occurring later in life. The main differences, some of which have already been alluded to, include the following:

- Incidence and type of tumor
- Relatively frequent demonstration of a close relationship between abnormal development (teratogenesis) and tumor induction (oncogenesis)
- Prevalence of underlying familial genetic aberrations
- Tendency of fetal and neonatal malignancies to regress or differentiate spontaneously
- Improved survival or cure of many childhood tumors, so that more attention is now being devoted to minimizing the adverse delayed effects of chemotherapy and radiation therapy in survivors, including the development of second malignancies

Incidence and Types

The most frequent childhood cancers arise in the hematopoietic system, nervous tissue (including the central and sympathetic nervous system, adrenal medulla, and retina), soft tissue, bone, and kidney. This is in sharp contrast to adults, in whom the skin, lung, breast, prostate, and colon are the most common sites of tumors.

Neoplasms that exhibit sharp peaks in incidence in children younger than 15 years of age and their approximate age distribution are indicated in Table 10.5. Leukemia alone accounts for more deaths in children under 15 years of age than all of the other tumors combined.

Histologically, many of the malignant nonhematopoietic pediatric neoplasms are unique. In general, they tend to have a more primitive (embryonal) undifferentiated appearance, often characterized by sheets of cells with small round nuclei, and frequently show features of organogenesis specific to the site of tumor origin. Because of this latter characteristic, these tumors are frequently designated by the suffix *-blastoma,* for example, nephroblastoma (Wilms tumor), hepatoblastoma, and neuroblastoma. Because of their primitive histologic appearance, many childhood tumors have been collectively referred to as *small round blue cell tumors.* The differential diagnosis of such tumors includes neuroblastoma, Wilms tumor, lymphoma (Chapter 13), rhabdomyosarcoma (Chapter 26), Ewing sarcoma/primitive neuroectodermal tumor (Chapter 26), medulloblastoma (Chapter 28), and retinoblastoma (Chapter 29). If the anatomic site of origin is known, diagnosis is usually possible on histologic grounds alone. Occasionally, a combination of chromosome analysis, immunoperoxidase stains, or electron microscopy is required. Two of these tumors are discussed here: the neuroblastic tumors, specifically neuroblastoma, and Wilms tumor. The remaining tumors are discussed in their respective organ-specific chapters.

Table 10.5 Common Malignant Neoplasms of Infancy and Childhood

0–4 Years	5–9 Years	10–14 Years
Leukemia	Leukemia	
Retinoblastoma		
Neuroblastoma	Neuroblastoma	
Wilms tumor		Renal cell carcinoma
Hepatoblastoma	Hepatocellular carcinoma	Hepatocellular carcinoma
Soft tissue sarcoma (especially rhabdomyosarcoma)	Soft tissue sarcoma	Soft tissue sarcoma
	Ewing sarcoma	Ewing sarcoma
		Osteogenic sarcoma
Teratoma		
Central nervous system tumors	Central nervous system tumors	
	Lymphoma	Lymphoma, including Hodgkin lymphoma
		Thyroid carcinoma

Neuroblastic Tumors

The term neuroblastic tumor includes tumors of the sympathetic ganglia and adrenal medulla that are derived from primordial neural crest cells populating these sites. Neuroblastoma is the most important member of this family. It is the most common extracranial solid tumor of childhood, and the most frequently diagnosed infant malignancy. The prevalence is about 1 in 7000 live births, and there are approximately 700 cases diagnosed each year in the United States. The median age at diagnosis is 18 months; approximately 40% of cases are diagnosed in infancy. Most neuroblastomas occur sporadically, but 1% to 2% are familial; in such cases the neoplasms may involve both of the adrenals or multiple primary autonomic sites. Germline mutations in the *anaplastic lymphoma kinase (ALK)* gene (Chapter 13) are a major cause of familial predisposition to neuroblastoma. Somatic gain-of-function *ALK* mutations are also observed in approximately 10% of sporadic neuroblastomas.

Despite the remarkable progress made in the therapy of this disease, long-term prognosis for high-risk subsets remains guarded, with 5-year survival in the range of 40%. As will be evident later, age and stage have a remarkable effect on prognosis, and, in general, children younger than 18 months of age tend to have a significantly more favorable prognosis than older individuals with similar disease burdens.

MORPHOLOGY

In childhood, about 40% of neuroblastomas arise in the adrenal medulla. The remainder occur anywhere along the sympathetic chain, with the most common locations being the paravertebral region of the abdomen (25%) and posterior mediastinum (15%). Tumors may arise in numerous other sites, including the pelvis, the neck, and within the brain (cerebral neuroblastomas).

Neuroblastomas range in size from minute nodules (so-called **in situ lesions**) to large masses more than 1 kg in weight. In situ neuroblastomas are reported to occur 40 times more frequently than clinically overt tumors. The great majority of these silent lesions spontaneously regress, leaving only a focus of fibrosis or calcification in the adult; this has led some to question the neoplastic connotation for the in situ lesions.

Some neuroblastomas are sharply demarcated by a fibrous pseudocapsule, but others are far more infiltrative and invade surrounding structures, including the kidneys, renal vein, and vena cava, and envelop the aorta. On transection, they are composed of soft, gray-tan tissue. Larger tumors have areas of necrosis, cystic softening, and hemorrhage. Occasionally, foci of punctate intratumoral calcification can be palpated.

Histologically, classic neuroblastomas are composed of small, primitive-appearing cells with dark nuclei, scant cytoplasm, and poorly defined cell borders growing in solid sheets. Such tumors may be difficult to differentiate morphologically from other small round blue cell tumors. Mitotic activity, nuclear breakdown ("karyorrhexis"), and pleomorphism may be prominent. The background often demonstrates a faintly eosinophilic fibrillary material **(neuropil)** that corresponds to neuritic processes of the primitive neuroblasts. Typically, **Homer-Wright pseudorosettes** can be found in which tumor cells are concentrically arranged about a central space filled with neuropil (Fig. 10.23). Other helpful features include positive immunohistochemical reactions for neuron-specific enolase and the transcription factor PHOX2B. Some neoplasms show signs of maturation that can be spontaneous or therapy-induced. Larger cells having more abundant cytoplasm, large vesicular nuclei, and a prominent nucleolus, representing ganglion cells in various stages of maturation, may be found in tumors admixed with primitive neuroblasts **(ganglioneuroblastoma).** Even better differentiated lesions contain many more large cells resembling mature ganglion cells with few if any residual neuroblasts; such neoplasms merit the designation **ganglioneuroma** (Fig. 10.24). Maturation of neuroblasts into ganglion cells is usually accompanied by the appearance of **Schwann cells.** The origin of Schwann cells in neuroblastoma remains an issue of contention; some investigators believe they represent a reactive population recruited by the tumor cells. However, studies using microdissection techniques have demonstrated that the Schwann cells harbor at least a subset of the same genetic alterations found in neuroblasts, and therefore are a component of the malignant clone. Irrespective of histogenesis, documenting the presence of schwannian stroma is essential, because its presence is associated with a **favorable outcome** (see later).

Metastases, when they develop, appear early and widely. In addition to local infiltration and lymph node spread, there is a pronounced tendency to spread through the bloodstream to the liver, lungs, bone marrow, and bones.

Staging. The International Neuroblastoma Staging System (INSS), which is the most widely used staging scheme worldwide, is detailed in Table 10.6.

Unfortunately, most (60% to 80%) children present with stage 3 or 4 tumors, and only 20% to 40% present with stage 1, 2A, 2B, or 4S neuroblastomas. The staging system is of paramount importance in determining prognosis.

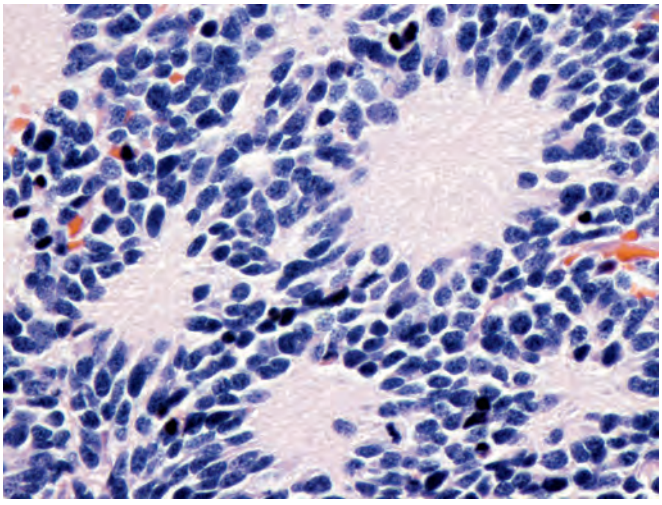


Figure 10.23 Adrenal neuroblastoma. This tumor is composed of small cells embedded in a finely fibrillar matrix (neuropil). Several Homer-Wright pseudorosettes are seen in this image.

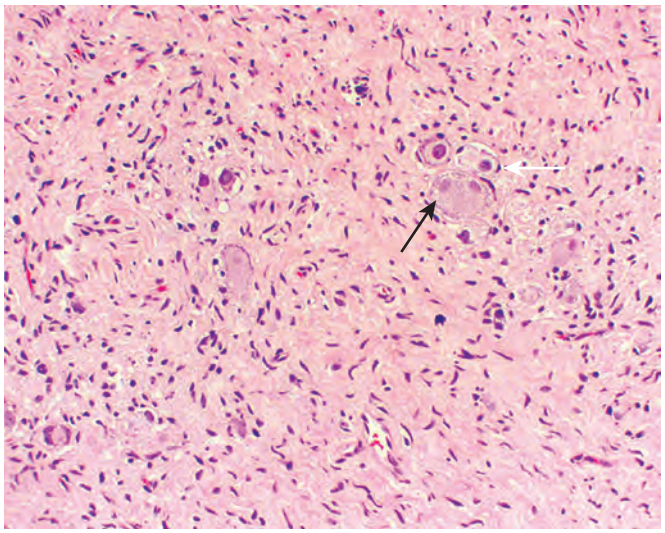


Figure 10.24 Ganglioneuromas are characterized by clusters of large cells with eccentric nuclei, prominent nucleoli, and abundant eosinophilic cytoplasm, representing neoplastic ganglion cells *(arrows)*. Spindle-shaped Schwann cells are present in the background stroma.

Clinical Course and Prognostic Features (Table 10.7)

In young children (younger than 2 years of age), neuroblastomas generally present with large abdominal masses, fever, and possibly weight loss. In older children, they may not come to attention until metastases produce manifestations, such as bone pain, respiratory symptoms, or gastrointestinal

Table 10.6 International Neuroblastoma Staging System

Stage 1	Localized tumor completely excised, with or without microscopic residual disease; representative ipsilateral nonadherent lymph nodes negative for tumor (nodes adherent to the primary tumor may be positive for tumor)
Stage 2A	Localized tumor resected incompletely grossly; representative ipsilateral nonadherent lymph nodes negative for tumor microscopically
Stage 2B	Localized tumor with or without complete gross excision, ipsilateral nonadherent lymph nodes positive for tumor; enlarged contralateral lymph nodes, which are negative for tumor microscopically
Stage 3	Unresectable unilateral tumor infiltrating across the midline with or without regional lymph node involvement; or localized unilateral tumor with contralateral regional lymph node involvement
Stage 4	Any primary tumor with dissemination to distant lymph nodes, bone, bone marrow, liver, skin, and/or other organs (except as defined for stage 4S)
Stage 4S	Localized primary tumor (as defined for stage 1, 2A, or 2B) with dissemination limited to skin, liver, and/or bone marrow (<10% of nucleated cells are constituted by neoplastic cells; >10% involvement of bone marrow is considered as stage 4); stage 4S limited to infants younger than 1 year of age

S, Special.

Modified from Brodeur GM, Pritchard J, Berthold F, et al: Revisions of the international neuroblastoma diagnosis, staging, and response to treatment, *J Clin Oncol* 11:1466, 1993.

Table 10.7 Prognostic Factors in Neuroblastomas

Variable	Favorable	Unfavorable
Stage[a]	Stage 1, 2A, 2B, 4S	Stage 3, 4
Age[a]	<18 months	>18 months
Histology[a]		
Evidence of schwannian stroma and gangliocytic differentiation[b]	Present	Absent
Mitosis-karyorrhexis index[c]	<200/5000 cells	>200/5000 cells
DNA ploidy[a]	Hyperdiploid (whole-chromosome gains)	Near-diploid (segmental chromosome losses; chromothripsis)
MYCN[a]	Not amplified	Amplified
Chromosome 1p loss	Absent	Present
Chromosome 11q loss	Absent	Present
TRKA expression	Present	Absent
TRKB expression	Absent	Present
Mutations of neuritogenesis genes	Absent	Present

[a]Corresponds to the most commonly used parameters in clinical practice for assessment of prognosis and risk stratification.

[b]It is not only the presence but also the amount of schwannian stroma that confer the designation of a favorable histology. At least 50% or more schwannian stroma is required before a neoplasm can be classified as ganglioneuroblastoma or ganglioneuroma.

[c]Mitosis-karyorrhexis index (MKI) is defined as the number of mitotic or karyorrhectic cells per 5000 tumor cells in random foci.

complaints. Neuroblastomas may metastasize widely through the hematogenous and lymphatic systems, particularly to liver, lungs, bones, and bone marrow. Proptosis and ecchymosis may also be present due to spread to the periorbital region, a common metastatic site. In neonates, disseminated neuroblastomas may present with multiple cutaneous metastases that cause deep blue discoloration of the skin (earning the unfortunate designation of "blueberry muffin baby"). About 90% of neuroblastomas, regardless of location, produce catecholamines (similar to the catecholamines associated with pheochromocytomas), which are an important diagnostic feature (i.e., elevated blood levels of catecholamines and elevated urine levels of the metabolites vanillylmandelic acid [VMA] and homovanillic acid [HVA]). Despite the elaboration of catecholamines, hypertension is much less frequent with these neoplasms than with pheochromocytomas (Chapter 24). Ganglioneuromas, unlike their malignant counterparts, tend to produce either asymptomatic mass lesions or symptoms related to compression.

The clinical course of neuroblastomas is extremely variable. Several clinical, histopathologic, molecular, and biochemical factors have been identified that have a bearing on prognosis (see Table 10.7); based on the collection of prognostic factors present in a given patient, they are classified as "low," "intermediate," or "high" risk. With improvements in therapy, long-term survival exceeds 90% of patients in the first two groups, but less than 50% of patients in the high-risk category are long-term survivors. The most pertinent prognostic factors include the following:

- *Tumor stage and patient age at time of diagnosis.* Neuroblastomas at stage 1, 2A, or 2B tend to have an excellent prognosis, irrespective of age ("low" or "intermediate" risk); tumors exhibiting amplification of the *MYCN* oncogene are a notable exception. Infants with localized primary tumors and metastases to the liver, bone marrow, and skin (stage 4S) represent a special subtype, wherein it is not uncommon for the disease to regress spontaneously. The biologic basis of this welcome behavior is not clear. The age of 18 months has emerged as a critical point of dichotomy in terms of prognosis. Children younger than 18 months of age, and especially those in the first year of life, have an excellent prognosis regardless of the stage of the neoplasm. Children older than 18 months of age generally fall into at least the "intermediate" risk category, and those with higher-stage tumors or with unfavorable prognostic features like *MYCN* amplification in the neoplastic cells are considered "high" risk.
- *Tumor morphology.* An age-linked morphologic classification of neuroblastic tumors has been proposed that divides them into favorable and unfavorable histologic subtypes. The specific morphologic features that bear on prognosis are listed in Table 10.7.
- *MYCN amplification status.* Amplification of the *MYCN* proto-oncogene in neuroblastomas markedly impacts the prognosis, particularly in tumors that would otherwise be classified as having a good prognosis. The presence of *MYCN* amplification "bumps" the tumor to the "high"-risk category, irrespective of age, stage, or histology. *MYCN* is located on the distal short arm of chromosome 2 (2p23-p24). Amplification of *MYCN* does not occur at the 2p23-p24 site, but as extrachromosomal double minute chromosomes or homogeneously staining regions on

other chromosomes (Fig. 10.25). *MYCN* amplification is present in about 20% to 30% of primary tumors, most presenting as advanced-stage disease, and the degree of amplification correlates with poorer prognosis. *MYCN* amplification is currently the most important genetic abnormality used in risk stratification of neuroblastic tumors (see later).

- *Ploidy* of the tumor cells correlates with outcome in children younger than 2 years of age but loses its independent prognostic significance in older children. Broadly, neuroblastomas can be divided into two categories: near-diploid and hyperdiploid (whole-chromosome gains), with the latter being associated with a more favorable prognosis.

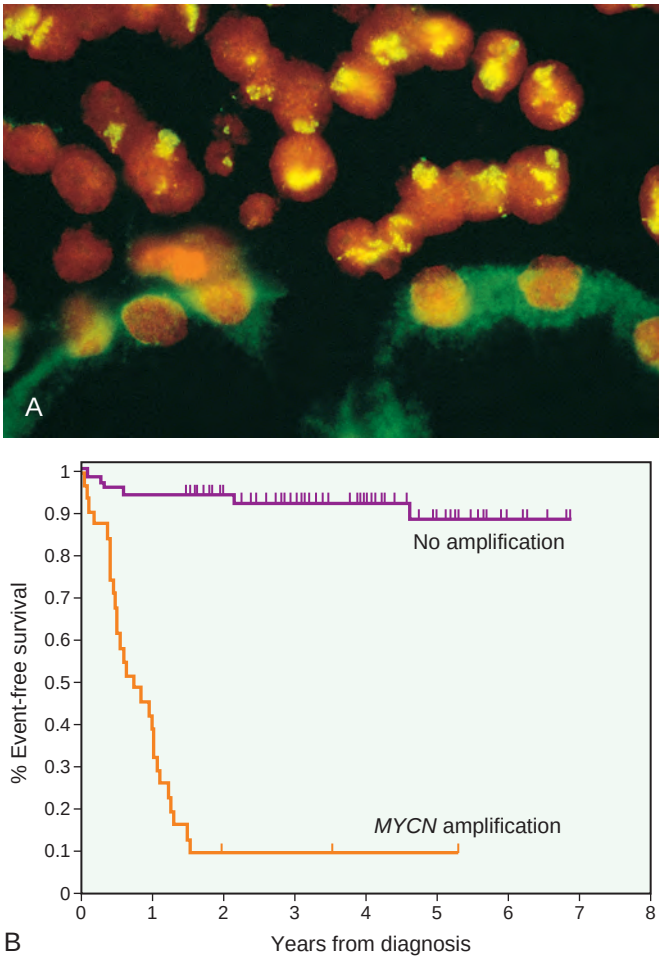


Figure 10.25 (A) Fluorescence in situ hybridization using a fluorescein-labeled cosmid probe for *MYCN* on a tissue section. Note the neuroblastoma cells on the upper half of the photo with large areas of staining *(yellow-green)*; this corresponds to amplified *MYCN* in the form of homogeneously staining regions. Renal tubular epithelial cells in the lower half of the photograph show no nuclear staining and background *(green)* cytoplasmic staining. (B) A Kaplan-Meier survival curve of infants younger than 1 year of age with metastatic neuroblastoma. The 3-year event-free survival of infants whose tumors lacked *MYCN* amplification was 93%, whereas those with tumors that had *MYCN* amplification had only a 10% event-free survival. (A, Courtesy Dr. Timothy Triche, Children's Hospital, Los Angeles, Calif.; B, Reproduced with permission from Brodeur GM: Neuroblastoma: biological insights into a clinical enigma, *Nat Rev Cancer* 3(3):203–216, 2003.)

It is postulated that neuroblastomas with hyperdiploidy have an underlying defect in the mitotic machinery leading to nondisjunction and whole-chromosome gains, but otherwise relatively banal karyotypes. The more aggressive near-diploid tumors harbor generalized genomic instability, with multiple segmental chromosomal aberrations that result in a complex karyotype with adverse prognostic implications. One peculiar form of segmental aberration in aggressive neuroblastomas is called chromothripsis (Chapter 7), which involves localized fragmentation of a chromosome segment followed by random assembly of the fragments. In a subset of neuroblastomas, chromothripsis can result in amplification of *MYCN* or other oncogenes, or losses in tumor suppressor loci.

Although age, stage, histology, *MYCN* status, and DNA ploidy are the "core" criteria used for formal risk stratification and therapeutic decision-making, various segmental chromosomal losses can also have prognostic significance. In particular, hemizygous deletion of the distal short arm of chromosome 1 in the region of band p36, which occurs in 25% to 35% of primary tumors, is strongly correlated with *MYCN* amplification, advanced disease stage, and an increased risk of disease relapse. Hemizygous loss of chromosome 11q is another common deletion event with adverse prognosis but is not correlated with *MYCN* amplification.

Whole-genome sequencing has uncovered alterations in a variety of genes. Some of these alterations have prognostic implications. These include those whose products are involved in neuritogenesis (a process in neuronal differentiation that includes the sprouting of neurites, which later become dendrites and axons). Selected examples of mutated genes in this category include *alpha thalassemia/mental retardation, X linked (ATRX);* genes involving chromatin remodeling, such as *AT-rich interaction domains 1A* and *1B* (*ARID1A* and *ARID1B*); and neurotrophin receptors (*NTRK1, NTRK2,* and *NTRK3*). The impact of these changes on prognosis is under investigation.

Although discussion of the treatment modalities for neuroblastoma is beyond the scope of this book, we mention in passing two promising experimental approaches. The first involves the use of retinoids as an adjunct therapy for inducing the differentiation of neuroblastoma. Recall that the retinoic acid pathway plays a critical role in cellular differentiation during embryogenesis. Secondly, clinical trials are evaluating the utility of small-molecule ALK inhibitors in treating tumors harboring activating *ALK* mutations.

Wilms Tumor

Wilms tumor (nephroblastoma) afflicts approximately 1 in 10,000 children in the United States, making it the most common primary renal tumor of childhood and the fourth most common pediatric malignancy in the United States. The peak incidence for Wilms tumor is between 2 and 5 years of age, and 95% of tumors occur before 10 years of age. Approximately 5% to 10% of Wilms tumors involve both kidneys, either simultaneously (synchronous) or one after the other (metachronous). Bilateral Wilms tumors have

a median age of onset approximately 10 months earlier than tumors restricted to one kidney, and these patients are presumed to harbor a germline mutation in one of the Wilms tumor–predisposing genes (see later). The biology of this tumor illustrates several important aspects of childhood neoplasms, such as the relationship between malformations and neoplasia, the histologic similarities between organogenesis and oncogenesis, the two-hit theory of recessive tumor suppressor genes (Chapter 7), the role of premalignant lesions, and perhaps most importantly, the potential for treatment to affect prognosis and outcome. Improvements in cure rates for Wilms tumor (from as low as 30% a few decades ago to approximately 90% currently) represent one of the greatest successes of pediatric oncology.

Pathogenesis and Genetics

The risk of Wilms tumor is increased with at least three recognizable groups of congenital malformations associated with distinct chromosomal loci. Although Wilms tumors arising in this setting account for no more than 10% of cases, these syndromic tumors have provided important insight into the biology of this neoplasm.

- The first group of patients has the *WAGR syndrome,* characterized by *W*ilms tumor, *a*niridia, *g*enitourinary anomalies, and intellectual disability (formerly called mental retardation). Their lifetime risk of developing Wilms tumor is approximately 33%. Individuals with WAGR syndrome carry constitutional (germline) deletions of 11p13. Studies on these patients led to the identification of the first Wilms tumor–associated gene, *WT1,* and a contiguously deleted autosomal dominant gene for aniridia, *PAX6,* both located on chromosome 11p13. Patients with deletions restricted to *PAX6,* with normal *WT1* function, develop aniridia, but they are not at increased risk for Wilms tumors. The presence of germline *WT1* deletions in WAGR syndrome represents the "first hit"; the development of Wilms tumor in these patients frequently correlates with the occurrence of a nonsense or frameshift mutation in the second *WT1* allele ("second hit").
- A second group of patients at much higher risk for Wilms tumor (~90%) has *Denys-Drash syndrome,* which is characterized by gonadal dysgenesis (male pseudohermaphroditism) and early-onset nephropathy leading to renal failure. The characteristic glomerular lesion in these patients is diffuse mesangial sclerosis (Chapter 20). As in patients with WAGR, these patients also demonstrate germline abnormalities in *WT1.* In patients with Denys-Drash syndrome, however, the genetic abnormality is a dominant-negative missense mutation in the zinc-finger region of the WT1 protein that affects its DNA-binding properties. This mutation interferes with the function of the remaining wild-type allele, yet strangely, it is sufficient only in causing genitourinary abnormalities but not tumorigenesis; Wilms tumors arising in Denys-Drash syndrome demonstrate bi-allelic inactivation of *WT1.* In addition to Wilms tumors, these individuals are also at increased risk for developing germ cell tumors called gonadoblastomas (Chapter 21), almost certainly a consequence of disruption in normal gonadal development.

  WT1 encodes a DNA-binding transcription factor that is expressed within several tissues, including the kidney and gonads, during embryogenesis. The WT1 protein is critical for normal renal and gonadal development. WT1 has multiple binding partners, and the choice of this partner can affect whether WT1 functions as a transcriptional activator or repressor in a given cellular context. Numerous transcriptional targets of WT1 have been identified, including genes encoding glomerular podocyte–specific proteins and proteins involved in induction of renal differentiation. Despite the importance of *WT1* in nephrogenesis and its unequivocal role as a tumor suppressor gene, only about 10% of patients with sporadic (nonsyndromic) Wilms tumors demonstrate *WT1* mutations, suggesting that the majority of these tumors are caused by mutations in other genes.
- A third group with an increased risk of developing Wilms tumor consists of children with *Beckwith-Wiedemann syndrome* (BWS), characterized by enlargement of body organs (organomegaly), macroglossia, hemihypertrophy, omphalocele, and abnormal large cells in the adrenal cortex (adrenal cytomegaly). BWS has served as a model for tumorigenesis associated with genomic imprinting (Chapter 5). The chromosomal region implicated in BWS has been localized to band 11p15.5 (so-called WT2 locus), distal to the *WT1* locus. This region contains multiple genes that are normally expressed from only one of the two parental alleles, with transcriptional silencing (i.e., imprinting) of the other parental homologue by methylation of the promoter region. Unlike WAGR and Denys-Drash syndromes, the genetic basis for BWS is considerably more heterogeneous in that no single 11p15.5 gene is involved in all cases. Moreover, the phenotype of BWS, including the predisposition to tumorigenesis, is influenced by the specific imprinting abnormalities present. One of the genes in the 11p15.5 region—insulin-like growth factor-2 (*IGF2*)—is normally expressed solely from the paternal allele, and the maternal allele is silenced by imprinting. In some Wilms tumors, loss of imprinting (i.e., re-expression of the maternal *IGF2* allele) can be demonstrated, leading to overexpression of the IGF2 protein. In other instances, there is selective deletion of the imprinted maternal allele combined with duplication of the transcriptionally active paternal allele in the tumor (uniparental paternal disomy), which has the identical functional effect of IGF2 overexpression. Because the IGF2 protein is an embryonal growth factor, it could conceivably explain the features of overgrowth associated with BWS as well as the increased risk for Wilms tumors in these patients.

In contrast with syndromic Wilms tumors, the molecular abnormalities underlying sporadic (i.e., nonsyndromic) tumors, which account for 90% of cases overall in children, are only beginning to be elucidated. Some of them are associated with specific histologic features described later. For example, gain-of-function mutations of the gene encoding β-catenin (Chapter 7) have been demonstrated in approximately 10% of sporadic Wilms tumors. Other recurrent mutations occur in genes encoding proteins involved in micro-RNA (miRNA) processing (*DROSHA,*

DGCR8, and *DICER1*); these are seen in 15% to 20% of Wilms tumors with predominantly blastemal histology (see later). It is postulated that aberrations in miRNA processing lead to reduced levels of many mature miRNAs, in particular in the miR-200 family, which is involved in "mesenchymal to epithelial transformation" during renal morphogenesis. The lack of mesenchymal to epithelial transformation likely leads to persistent blastemal "rests" in the kidney (see the following), which evolve into Wilms tumors. Finally, tumors with *TP53* mutations are associated with an especially poor prognosis and often have a distinctive anaplastic histologic appearance, described later.

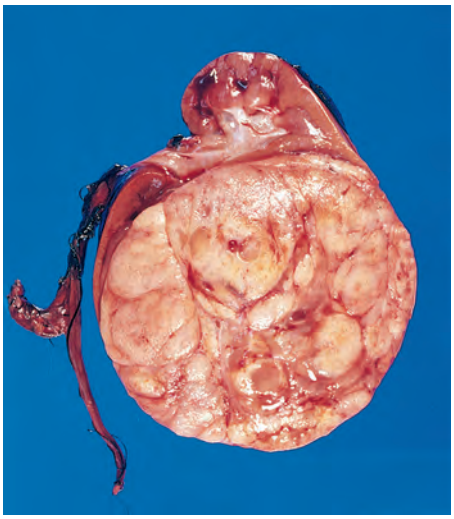


Figure 10.26 Wilms tumor in the lower pole of the kidney with the characteristic tan-to-gray color and well-circumscribed margins.

MORPHOLOGY

Grossly, Wilms tumor tends to present as a large, solitary, well-circumscribed mass. Approximately 10% are either bilateral or multicentric at the time of diagnosis. On cut section, the tumor is soft, homogeneous, and tan to gray with occasional foci of hemorrhage, cyst formation, and necrosis (Fig. 10.26).

Microscopically, Wilms tumors are characterized by recognizable attempts to recapitulate different stages of nephrogenesis. **The classic triphasic combination of blastemal, stromal, and epithelial cell types is observed in the vast majority of lesions,** although the percentage of each component is variable (Fig. 10.27A). Sheets of small blue cells with few distinctive features characterize the blastemal component. Epithelial differentiation is usually in the form of abortive tubules or glomeruli. Stromal cells are usually fibroblastic or myxoid in nature, although skeletal muscle differentiation is not uncommon. Rarely, other heterologous elements are identified, including squamous or mucinous epithelium, smooth muscle, adipose tissue, cartilage, and osteoid and neurogenic tissue. Approximately 5% of tumors have **anaplasia,** defined as the presence of cells with large, hyperchromatic, pleomorphic nuclei and abnormal mitoses (see Fig. 10.27B). The presence of anaplasia correlates with the presence of *TP53* mutations and the emergence of resistance to chemotherapy. Recall that p53 elicits pro-apoptotic signals in response to DNA damage (Chapter 7). The loss of p53 function might explain the relative unresponsiveness of anaplastic cells to cytotoxic chemotherapy.

Nephrogenic rests are putative precursor lesions of Wilms tumors and are seen in the renal parenchyma adjacent to approximately 25% to 40% of unilateral tumors; this frequency rises to nearly 100% in cases of bilateral Wilms tumors. In many instances the nephrogenic rests share genetic alterations with the adjacent Wilms tumor, pointing to their preneoplastic status. The appearance of nephrogenic rests varies from expansile masses that resemble Wilms tumors (hyperplastic rests) to sclerotic rests consisting predominantly of fibrous tissue and occasional admixed immature tubules or glomeruli. It is important to document the presence of nephrogenic rests in the resected specimen, because these patients are at an increased risk of developing Wilms tumors in the contralateral kidney and require frequent and regular surveillance for many years.

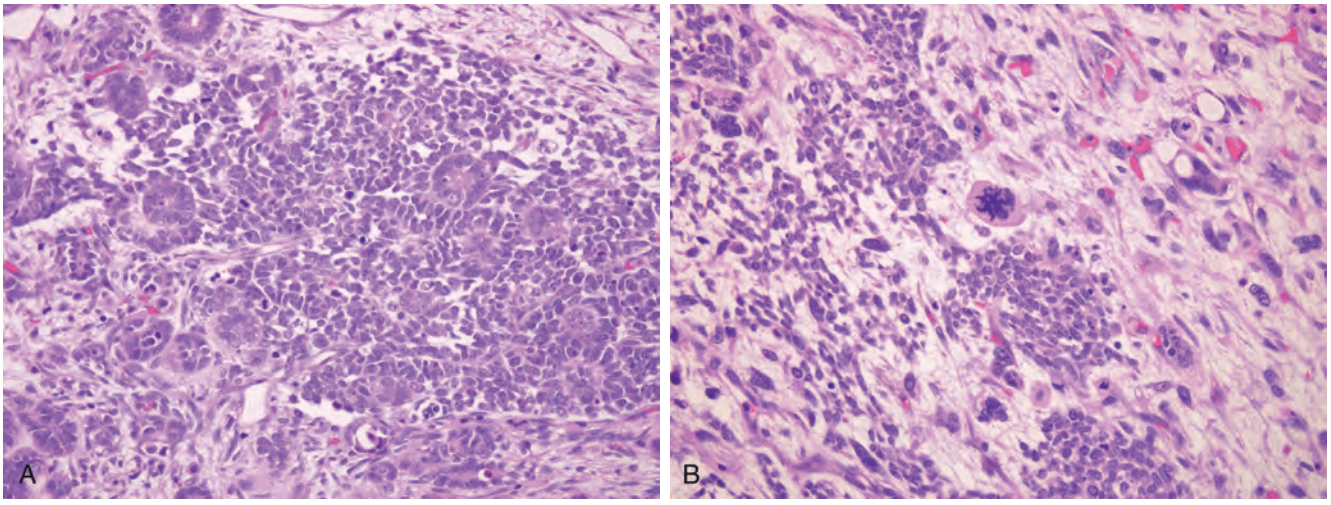


Figure 10.27 (A) Wilms tumor with tightly packed blue cells (blastemal component) and interspersed primitive tubules (epithelial component). Although multiple mitotic figures are seen, none are atypical in this field. (B) Focal anaplasia was present in this Wilms tumor, characterized by cells with hyperchromatic, pleomorphic nuclei, and abnormal mitoses (center).

Clinical Features

Most children with Wilms tumors present with a large abdominal mass that may be unilateral or, when very large, extend across the midline and down into the pelvis. Hematuria, pain in the abdomen after a traumatic incident, intestinal obstruction, and hypertension are other patterns of presentation. In a considerable number of these patients, pulmonary metastases are present at the time of primary diagnosis.

As stated, most patients with Wilms tumor can expect to be cured of their malignancy. Anaplastic histology is perhaps the most critical determinant of adverse prognosis. Even anaplasia restricted to the kidney (i.e., without extrarenal spread) confers an increased risk of recurrence and death, emphasizing the need for accurate identification of this histologic feature. Molecular parameters that correlate with adverse prognosis include loss of heterozygosity of chromosomes 1p and 16q, and gain of chromosome 1q in tumor cells. Along with the increased survival of individuals with Wilms tumor have come reports of an increased risk of developing second primary tumors, including bone and soft tissue sarcomas, leukemia and lymphomas, and breast cancers. Although some of these neoplasms result from the presence of a germline mutation in a cancer predisposition gene, others are a consequence of therapy, most commonly radiation administered to the cancer field. This tragic, albeit uncommon, outcome has mandated that radiation therapy be used judiciously in the treatment of this and other childhood cancers.

SUGGESTED READINGS

Congenital Anomalies, Hydrops, and Infections

Bellini C, Hennekam RC: Non-immune hydrops fetalis: a short review of etiology and pathophysiology, *Am J Med Genet A* 158A:597–605, 2012. *[A well written review on non-immune hydrops, which accounts for the overwhelming majority of fetal hydrops in the Western world].*

McCloskey B, Endericks T: The rise of Zika infection and microcephaly: what can we learn from a public health emergency?, *Public Health* 150:87–92, 2017. *[This article considers why Zika was declared a Public Health Emergency of International Concern (PHEIC), why it stopped being one, and what we can learn from this for the future].*

Mone F, Quinlan-Jones E, Ewer AK et al: Exome sequencing in the assessment of congenital malformations in the fetus and neonate, *Arch Dis Child Fetal Neonatal Ed* 104:F452–F456, 2019. *[A review of the utility of next-generation sequencing methods on finding genetic causes for congenital anomalies].*

Murphy SL, Mathews TJ, Martin JA et al: Annual summary of vital statistics: 2013-2014, *Pediatrics* 139(6):2017. PubMed PMID: 28814547. *[A periodically updated release from the United States Centers for Disease Control and Prevention, summarizing statistical data related to birth and death rates, congenital anomalies, causes of death, and so on].*

Cystic Fibrosis and Inborn Errors of Metabolism

Coelho AI, Rubio-Gozalbo ME, Vicente JB et al: Sweet and sour: an update on classic galactosemia, *J Inherit Metab Dis* 40:325–342, 2017. *[A summary of the current knowledge on this classic inborn error of metabolism as well as a discussion of ongoing avenues of study].*

Cutting GR: Cystic fibrosis genetics: from molecular understanding to clinical application, *Nat Rev Genet* 16:45–56, 2015. *[An excellent review of the genetics of cystic fibrosis, including a discussion of disease modifiers and targeted therapies].*

Elborn JS: Cystic fibrosis, *Lancet* 388:2519–2531, 2016. *[An excellent overview including clinical and molecular features of cystic fibrosis].*

Donlon J, Sarkissian C, Levy H et al: Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In Valle D, Beaudet AL, et al, editors: *The online metabolic and molecular bases of inherited disease*, New York, NY, 2014, McGraw-Hill. http://ommbid.mhmedical.com/content.aspx?bookid=971&sectionid=62673211. (Accessed December 11, 2017). *[An in-depth online resource that reviews this prototypical Mendelian disorder].*

Farrell PM, White TB, Ren CL et al: Diagnosis of cystic fibrosis: consensus guidelines from the Cystic Fibrosis Foundation, *J Pediatr* 181S:S4–S15, 2017. *[Updated consensus guidelines for the diagnosis of cystic fibrosis].*

Diseases of Prematurity and SIDS

Ahn SY, Chang YS, Kim JH et al: Two-year follow-up outcomes of premature infants enrolled in the phase I trial of mesenchymal stem cell transplantation for bronchopulmonary dysplasia, *J Pediatr* 2017. *[Although still in trial phase, it is encouraging to see that mesenchymal stem cells can be safely used in humans].*

Moon RY, AAP Task Force on Sudden Infant Death Syndrome: SIDS and other sleep-related infant deaths: evidence base for 2016 updated recommendations for a safe infant sleeping environment, *Pediatrics* 138(5):e20162940, 2016. *[A comprehensive guideline for preventing sudden infant death syndrome that takes into account the known predisposing conditions for this still unexplained syndrome].*

Pierro M, Ciarmoli E, Thébaud B: Bronchopulmonary dysplasia and chronic lung disease: stem cell therapy, *Clin Perinatol* 42(4):889–910, 2015. *[An exciting paper describing mesenchymal stromal (stem) cells ameliorating many critical aspects of bronchopulmonary dysplasia pathogenesis in preclinical models].*

Thakkar HS, Lakhoo K: The surgical management of necrotising enterocolitis (NEC), *Early Hum Dev* 97:25–28, 2016. *[A concise review of pathogenesis, diagnosis, and treatment of this condition].*

Solebo AL, Teoh L, Rahi J: Epidemiology of blindness in children, *Arch Dis Child* 102(9):853–857, 2017. *[A review on childhood visual impairment summarizing new evidence on the global epidemiology of, and the emerging therapies for, severe visual impairment and blindness including retinopathy of prematurity].*

Tumors of Infancy and Childhood

Cohn SL, Pearson ADJ, London WB et al: The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report, *J Clin Oncol* 27:289–297, 2009. *[The original paper that introduces the International Neuroblastoma Risk Group risk stratification system for neuroblastoma, which relies on pretreatment imaging].*

Dome JS, Perlman EJ, Graf N: Risk stratification for Wilms tumor: current approach and future directions, *Am Soc Clin Oncol Educ Book* 215–223, 2014. *[A review of clinical, histologic, and molecular determinants of prognosis for Wilms tumor].*

Gadd S, Huff V, Walz AL et al: A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor, *Nat Genet* 49:1487–1494, 2017. *[A large study on the genetic changes in Wilms tumor highlights the molecular complexity and heterogeneity that may be seen in this childhood tumor].*

Matthay KK, Maris JM, Schleiermacher G et al: Neuroblastoma, *Nat Rev Dis Primers* 2:16078, 2016. *[An excellent recent review of the clinical, pathologic, and molecular features of neuroblastoma].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 11

Blood Vessels

Richard N. Mitchell • Marc K. Halushka

CHAPTER CONTENTS

Vascular pathology is responsible for more morbidity and mortality than any other category of human disease. Although the most clinically significant lesions involve arteries, venous disorders are not inconsequential. Two principal mechanisms underlie vascular disease:

- *Narrowing (stenosis)* or *complete obstruction* of vessel lumens, either progressively (e.g., by atherosclerosis) or precipitously (e.g., by thrombosis or embolism)
- *Weakening* of vessel walls, leading to dilation or rupture

To better appreciate the pathogenesis of vascular disorders, it is important to first understand the normal vessel structure and function.

VASCULAR STRUCTURE AND FUNCTION

The general architecture and cellular composition of blood vessels are comparable throughout the cardiovascular system. However, structural specializations that reflect distinct functional roles characterize specific kinds of vessels (Fig. 11.1). For example, arterial walls are thicker than corresponding veins at the same level of branching to accommodate pulsatile flow and higher blood pressures. Arterial wall thickness gradually diminishes as the vessels become smaller, but the ratio of wall thickness to lumen diameter increases, allowing these muscular vessels to exert control over blood flow and pressure. Many disorders of the vasculature affect only particular types of vessels and thus have characteristic anatomic distributions. Thus atherosclerosis affects mainly elastic and muscular arteries, hypertension affects small muscular arteries and arterioles, and different varieties of vasculitis characteristically involve only vessels of a certain caliber.

The basic constituents of the walls of blood vessels are endothelial cells (ECs) and smooth muscle cells (SMCs), admixed with a variety of extracellular matrix (ECM) including elastin, collagen, and glycosaminoglycans. The relative amount and configuration of the basic constituents differ along the vasculature owing to local adaptations to mechanical or metabolic needs. In arteries and veins, these constituents are organized into three concentric layers—*intima, media,* and *adventitia,* which are anatomically more distinct in the arteries.

- The *intima* normally consists of a single layer of ECs attached to a basement membrane with a thin underlying

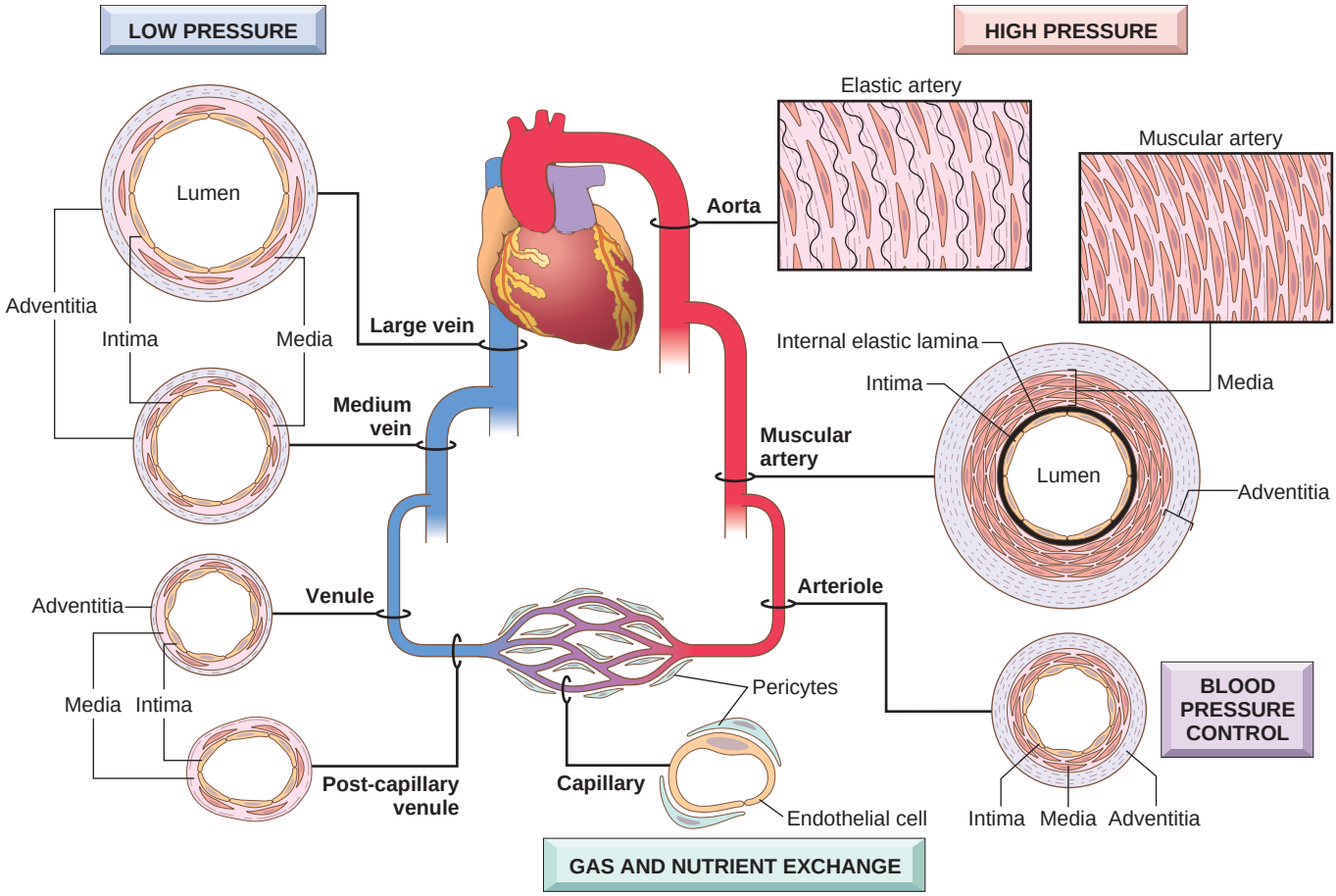


Figure 11.1 Regional specializations of the vasculature. Although the basic organization of the vasculature is constant, the thickness and composition of the various layers differs according to hemodynamic forces and tissue requirements. Thus, the aorta and other elastic arteries have substantial elastic tissue to accommodate high pulsatile forces, with the capacity to recoil and transmit energy into forward blood flow. These vessels have lamellar units that are comprised of repetitions of a layer of elastic fibers, a smooth muscle cell, and intervening extracellular matrix. Purely muscular arteries have elastic fibers only at the intersection of the intima and media or media and adventitia. In comparison, the venous system has relatively poorly developed thinner medial layers that permit greater capacitance, and the capillary wall permits ready diffusion of oxygen and nutrients because it is comprised only of an endothelial cell and sparse encircling pericytes. The different structure and functional attributes also influence the disorders that can affect the various parts of the vascular tree. Thus, loss of aortic elastic tissue will result in aneurysm, while stasis in a dilated venous bed can result in thrombosis.

layer of ECM; the intima is demarcated from the media by the internal elastic lamina.

- The *media* of *elastic arteries* (e.g., the aorta) are arranged in layers of lamellar units of elastin fibers and SMCs akin to tree rings. This high elastin content allows these vessels to expand during systole and recoil during diastole—functionally propelling blood toward the tissues. With aging and the loss of elasticity, the aorta and larger arteries become less compliant; besides transmitting higher pressures into distal tissues, the arteries of older individuals often become progressively tortuous and dilated *(ectatic)*.
  - In *muscular arteries,* the media is composed predominantly of circumferentially oriented SMCs. Arteriolar SMC contraction (vasoconstriction) or relaxation (vasodilation) is regulated by inputs from the autonomic nervous system and local metabolic factors.
  - *Arterioles are the principal points of physiologic resistance to blood flow*. Since the resistance to fluid flow is inversely proportional to the fourth power of the diameter (i.e., halving the diameter increases resistance 16-fold), small decreases in the lumen size of arterioles caused by structural changes or vasoconstriction can have profound effects on blood pressure.
- The *adventitia* lies external to the media and in many arteries is separated from the media by a well-defined external elastic lamina. The adventitia consists of loose connective tissue and can also contain nerve fibers. Diffusion of oxygen and nutrients from the lumen is adequate to sustain thin-walled vessels and the innermost SMCs of all vessels. In large- and medium-sized vessels, however, small arterioles within the adventitia (called *vasa vasorum*—literally, "vessels of the vessels") perfuse the outer half to two-thirds of the media.

As already alluded to, arteries are divided into three types based on their size and structural features: (1) large or elastic arteries including the aorta, the major branches of the aorta (innominate, subclavian, common carotid, and iliac), and pulmonary arteries; (2) medium-sized or muscular arteries comprising smaller branches of the aorta (e.g., coronary and renal arteries); and (3) small arteries (≤2 mm in diameter) and *arterioles* (20 to 100 μm in diameter) within tissues and organs.

Capillaries are slightly smaller (5 μm) than the diameter of a red blood cell (7 to 8 μm); they have an EC lining but no media, although variable numbers of pericytes, cells that resemble SMCs, typically lie just deep to the endothelium. Collectively, capillaries have a huge cross-sectional area and also have a relatively low flow rate. The combination of thin walls and slow flow makes capillaries ideally suited for the exchange of diffusible substances between blood and tissues. Because functionally useful oxygen diffusion is limited to a distance of only approximately 100 μm, the capillary network of most tissues is very rich. Tissues with high metabolic rates such as myocardium and brain have the highest density of capillaries.

Blood from capillary beds flows into *postcapillary venules* and then sequentially through collecting venules and small, medium, and large veins. In most forms of inflammation, vascular leakage and leukocyte exudation occur preferentially from postcapillary venules (Chapter 3).

Relative to arteries at the same level of branching, veins have larger diameters, larger lumens, and thinner and less well-organized walls (see Fig. 11.1). These structural features augment the capacitance of the venous side of the circulation, which on average contains about two-thirds of total blood volume. Reverse flow (due to gravity) is prevented in the extremities by venous valves.

Lymphatics are thin-walled, endothelium-lined channels that drain lymph (water, electrolytes, glucose, fat, proteins, and inflammatory cells) from the interstitium of tissues, eventually reconnecting with the blood stream via the thoracic duct. Lymphatics transport interstitial fluid and inflammatory cells from the periphery to lymph nodes, thereby facilitating antigen presentation and cell activation in the nodal tissues—and enabling continuous monitoring of peripheral tissues for infection. This can be a double-edged sword, however, as these channels can also disseminate disease by transporting microbes or tumor cells to distant sites.

KEY CONCEPTS

VASCULAR STRUCTURE AND FUNCTION

- All vessels except capillaries share a general three-layered architecture consisting of an endothelial-lined intima, surrounding smooth muscle media, and supportive adventitia.
- The SMCs and ECM of arteries, veins, and capillaries vary according to hemodynamic demands (e.g., pressure, pulsatility) and functional requirements.
- The specific composition of the vessel wall at any given site within the vascular tree influences the nature and consequences of pathologic injuries.

VASCULAR ANOMALIES

Although rarely symptomatic, anatomic variants of the usual vascular supply are important to recognize, as the failure to do so may lead to surgical complications and impede attempted therapeutic interventions (e.g., placement of coronary artery stents). Among the other congenital vascular anomalies, four are particularly significant, although not necessarily common.

- *Developmental* or *berry aneurysms* occur in cerebral vessels; when ruptured, these can cause fatal intracerebral hemorrhage (Chapter 28).
- *Arteriovenous fistulas* are direct connections (usually small) between arteries and veins that bypass capillaries. They occur most commonly as developmental defects but can also result from rupture of an arterial aneurysm into the adjacent vein, from penetrating injuries that pierce arteries and veins, or from inflammatory necrosis of adjacent vessels. Surgically generated arteriovenous fistulas provide vascular access for chronic hemodialysis. Like berry aneurysms, arteriovenous fistulas can rupture. Large or multiple arteriovenous fistulas become clinically significant by shunting blood from the arterial to the venous circulations, forcing the heart to pump additional volume and leading to high-output cardiac failure.
- *Fibromuscular dysplasia* is a focal irregular thickening in medium and large muscular arteries including renal, carotid, splanchnic, and vertebral vessels. The cause is unknown. Segments of the vessel wall are focally thickened by a combination of medial and intimal hyperplasia and fibrosis, resulting in luminal stenosis. In the renal arteries, it can be a cause of renovascular hypertension (Chapter 20). Immediately adjacent vessel segments can have markedly attenuated media (on angiography the vessels are said to have a "string of beads" appearance) leading to vascular outpouchings *(aneurysms)* that can rupture.
- *Anomalous coronary artery origin* occurs from a developmental anomaly in which both coronary arteries arise over the same coronary cusp of the aortic valve. Although many of these anomalies are benign, when a coronary vessel passes between the aorta and pulmonary artery it can be squeezed, for example, during exercise, limiting blood flow and resulting in sudden death.

VASCULAR WALL RESPONSE TO INJURY

The integrated functioning of ECs and SMCs impacts physiologic and pathophysiologic responses to hemodynamic and biochemical stimuli. Their function (and dysfunction) are described briefly, followed by discussion of specific vascular disorders.

ECs form a specialized simple squamous epithelium lining for blood vessels. Although ECs throughout the vascular tree share many attributes, populations that line different portions of the vascular tree (large vessels vs. capillaries, arteries vs. veins) have distinct gene expression profiles, behaviors, and morphologic appearances. Thus ECs in liver sinusoids or in renal glomeruli are fenestrated (they have holes, presumably to facilitate filtration), while central nervous system ECs (along with the associated perivascular cells) create an impermeable blood-brain barrier.

ECs are versatile multifunctional cells with a wealth of synthetic and metabolic properties. In the normal state they have several constitutive activities that are critical for vessel homeostasis and circulatory function (Table 11.1). Unless injured or otherwise activated, ECs have a nonthrombogenic surface that maintains blood in a fluid state (Chapter 4). They also modulate medial SMC tone (thereby influencing

Table 11.1 Endothelial Cell Properties and Functions

Property/Function	Mediators/Products
Maintenance of permeability barrier	
Elaboration of anticoagulant, antithrombotic, and fibrinolytic regulators	Prostacyclin Thrombomodulin Heparin-like molecules Plasminogen activator
Elaboration of prothrombotic molecules	Von Willebrand factor Tissue factor Plasminogen activator inhibitor
Production of extracellular matrix	Collagen, proteoglycans
Modulation of blood flow and vascular reactivity	*Vasoconstrictors:* endothelin, ACE *Vasodilators:* NO, prostacyclin
Regulation of inflammation and immunity	IL-1, IL-6, chemokines Adhesion molecules: VCAM-1, ICAM, E-selectin, P-selectin Histocompatibility antigens
Regulation of cell growth	Growth stimulators: PDGF, CSF, FGF Growth inhibitors: heparin, TGF-β

ACE, Angiotensin-converting enzyme; *CSF*, colony-stimulating factor; *FGF*, fibroblast growth factor; *ICAM*, intercellular adhesion molecule; *IL*, interleukin; *LDL*, low-density lipoprotein; *NO*, nitric oxide; *PDGF*, platelet-derived growth factor; *TGF-β*, transforming growth factor-β; *VCAM*, vascular cell adhesion molecule.

vascular resistance); metabolize hormones such as angiotensin; regulate inflammation; and affect the growth of other cell types, particularly SMCs. Although interendothelial junctions are largely impermeable in normal vessels, vasoactive agents (e.g., histamine) cause contraction of ECs and allow the rapid egress of fluids, electrolytes, and protein; in inflammation, even leukocytes can slip between adjacent ECs (Chapter 3).

ECs can respond to various stimuli by adjusting their steady-state (constitutive) functions and by expressing newly acquired (inducible) properties—a process termed *endothelial activation* (Fig. 11.2). There are numerous inducers of endothelial activation including cytokines and bacterial products, which elicit inflammation and, in severe cases, can promote septic shock (Chapter 4). Additional activators include hemodynamic stresses and lipid products that are critical to the pathogenesis of atherosclerosis (see later) and advanced glycation end-products that are important in the pathogenic sequelae of diabetes (Chapter 24). Viruses, complement components, and hypoxia also activate ECs. Activated ECs, in turn, express adhesion molecules (Chapter 3) and produce cytokines, chemokines, growth factors, vasoactive molecules that result either in vasoconstriction or in vasodilation, induction of major histocompatibility complex molecules, procoagulant and anticoagulant factors, and a variety of other biologically active products. ECs influence the vasoreactivity of the underlying SMCs through the production of both vasodilating (relaxing) factors (e.g., nitric oxide [NO]) and vasoconstrictive factors (e.g., endothelin). Normal EC function is characterized by a balance of these responses.

Endothelial dysfunction refers to an alteration in endothelial phenotype—seen in many different conditions—that is often both proinflammatory and prothrombogenic. It is responsible, at least in part, for the initiation of thrombus formation, atherosclerosis, and the vascular lesions of hypertension and other disorders. Certain forms of endothelial dysfunction are rapid in onset (within minutes), reversible, and independent of new protein synthesis (e.g., EC contraction induced by histamine and other vasoactive mediators that cause gaps in venular endothelium (Chapter 3). Other changes, such as the upregulation of adhesion molecules, involve alterations in gene expression and protein synthesis and may require hours or even days to develop.

Vascular SMCs are the predominant cellular element of the vascular media, playing important roles in normal vascular repair and pathologic processes such as atherosclerosis. SMCs have the capacity to proliferate when appropriately stimulated; they can also synthesize collagen, elastin, and proteoglycans and elaborate growth factors and cytokines. SMCs are also responsible for the vasoconstriction or dilation that occurs in response to physiologic or pharmacologic stimuli.

Intimal Thickening: A Stereotyped Response to Vascular Injury

Vascular injury associated with EC dysfunction or loss stimulates SMC recruitment and proliferation and associated ECM synthesis; the result is intimal thickening that can compromise vascular flow. Healing of injured vessels is analogous to the healing process that occurs in any other damaged tissue (Chapter 3) but has a somewhat unique outcome because the process impacts downstream blood flow. ECs involved in repair can migrate from adjacent uninjured areas into denuded areas or can derive from

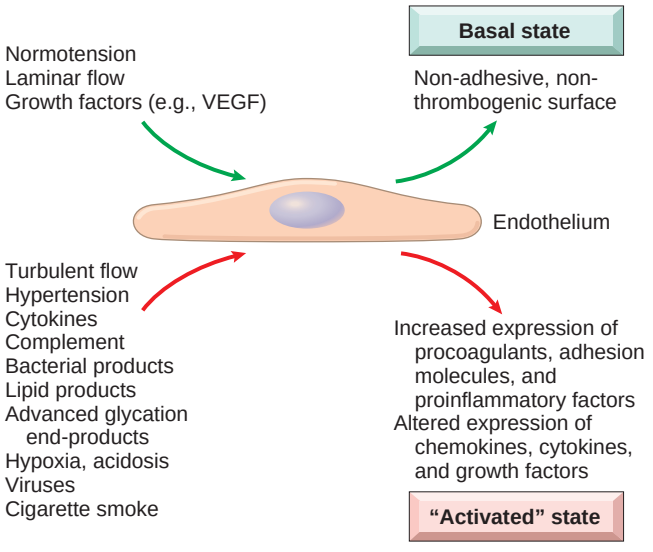


Figure 11.2 Basal and activated endothelial cell states. Normal blood pressure, laminar flow, and stable growth factor levels promote a basal endothelial cell state that maintains a nonthrombotic, nonadhesive surface with appropriate vascular wall smooth muscle tone. Injury or exposure to certain mediators results in endothelial activation, a state where endothelial cells develop a procoagulant surface that can be adhesive for inflammatory cells and express factors that cause smooth muscle contraction and/or proliferation and matrix synthesis. *VEGF*, Vascular endothelial growth factor.

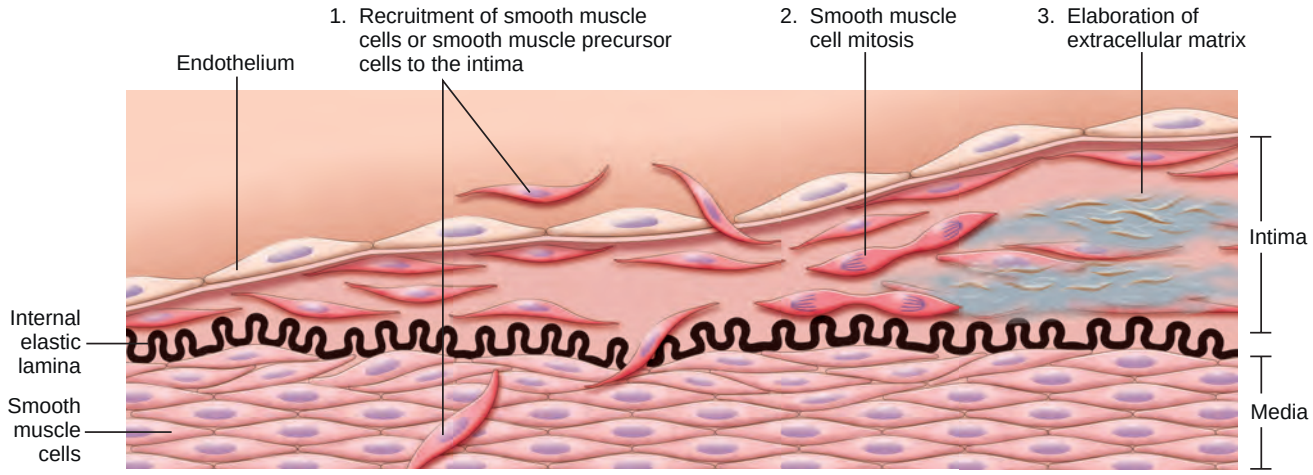


Figure 11.3 Stereotypical response to vascular injury. Schematic diagram of intimal thickening, emphasizing intimal smooth muscle cell migration and proliferation associated with extracellular matrix synthesis. Intimal smooth muscle cells can derive from the underlying media or can be recruited from circulating precursors; the intimal cells are shown in a darker shade to emphasize that they have a proliferative, synthetic, and noncontractile phenotype distinct from medial smooth muscle cells.

circulating precursors. Medial SMCs or circulating smooth muscle precursor cells also migrate into the intima, proliferate, and synthesize ECM in much the same way that fibroblasts fill in a wound (Fig. 11.3). The resulting neointima is typically completely covered by ECs. This neointimal response occurs with any form of vascular damage or dysfunction, regardless of cause. Thus intimal thickening is the stereotypical response of the vessel wall to any insult.

Neointimal SMCs have a phenotype that is distinct from that of medial SMCs. Specifically, neointimal SMCs are more proliferative, with increased biosynthetic capabilities and reduced contractile function. This neointimal SMC behavior is regulated by cytokines and growth factors derived from platelets, ECs, and macrophages, as well as thrombin and activated complement factors. With time and restoration and/or normalization of the EC layer, the neointimal SMCs can return to a nonproliferative state.

KEY CONCEPTS

RESPONSE OF VASCULAR WALL CELLS TO INJURY

- All vessels are lined by endothelium; although all ECs share certain homeostatic properties, ECs in specific vascular beds have special features that allow for tissue-specific functions (e.g., fenestrated ECs in renal glomeruli).
- EC function is tightly regulated in both the basal and activated states. Various physiologic and pathophysiologic stimuli induce endothelial activation and dysfunction that alter the EC phenotype (e.g., procoagulative vs. anticoagulative, proinflammatory vs. antiinflammatory, and nonadhesive vs. adhesive).
- Injury (of almost any type) to the vessel wall results in a stereotyped healing response involving SMC proliferation, ECM deposition, and intimal expansion.
- The recruitment and activation of the SMCs involves signals from cells (e.g., ECs, platelets, and macrophages) as well as mediators derived from coagulation and complement cascades.
- Excessive thickening of the intima can result in luminal stenosis and vascular obstruction.

HYPERTENSIVE VASCULAR DISEASE

Systemic and local tissue blood pressures must be maintained within a narrow range to prevent untoward consequences. Low pressures (hypotension) result in inadequate organ perfusion and can lead to dysfunction or tissue death. Conversely, high pressure (hypertension) can cause end-organ damage and is one of the major risk factors for atherosclerosis (see later).

Like height and weight, blood pressure is a continuously distributed variable. Detrimental effects of blood pressure increase continuously as the pressure rises—no rigidly defined threshold level of blood pressure identifies those who have an increased risk for cardiovascular disease. Both the systolic and the diastolic blood pressure are important in determining risk; specifically, according to the newest guidelines, individuals with diastolic pressures above 80 mm Hg or systolic pressures above 120 mm Hg are considered to have clinically significant hypertension. Approximately 46% of individuals in the general population are therefore hypertensive based on these newer criteria. However, such cutoffs do not reliably assess risk in all patients; for example, when other risk factors such as diabetes are present, lower thresholds are applicable.

Table 11.2 lists the major causes of hypertension. Although the molecular pathways that regulate normal blood pressure are reasonably well understood, the causes of hypertension in most individuals remain largely unknown. A small number of patients (approximately 10%) are said to have *secondary hypertension* resulting from an underlying renal or adrenal disease (e.g., primary aldosteronism, Cushing syndrome, or pheochromocytoma), renal artery stenosis, or other identifiable cause. However, approximately 90% of hypertension is idiopathic—so-called essential hypertension. It seems likely that hypertension is a multifactorial disorder resulting from the cumulative effects of multiple genetic polymorphisms and interacting environmental factors.

The prevalence and vulnerability to complications of hypertension increase with age and are higher among African

Table 11.2 Types and Causes of Hypertension (Systolic and Diastolic)

Essential Hypertension
Accounts for 90%–95% of all cases
Secondary Hypertension
Renal
Acute glomerulonephritis Chronic renal disease Polycystic disease Renal artery stenosis Renal vasculitis Renin-producing tumors
Endocrine
Adrenocortical hyperfunction (Cushing syndrome, primary aldosteronism, congenital adrenal hyperplasia, licorice ingestion) Exogenous hormones (glucocorticoids, estrogen [including pregnancy-induced and oral contraceptives], sympathomimetics and tyramine-containing foods, monoamine oxidase inhibitors) Pheochromocytoma Acromegaly Hyperthyroidism (thyrotoxicosis) Pregnancy-induced (preeclampsia)
Cardiovascular
Coarctation of the aorta Polyarteritis nodosa Increased intravascular volume Increased cardiac output Rigidity of the aorta
Neurologic
Psychogenic Increased intracranial pressure Sleep apnea Acute stress, including surgery

Americans. Besides increasing risk of atherosclerosis, hypertension can cause cardiac hypertrophy and heart failure *(hypertensive heart disease)* (Chapter 12), multi-infarct dementia (Chapter 28), aortic dissection (discussed later in this chapter), and renal failure (Chapter 20). Unfortunately, hypertension typically remains asymptomatic until late in its course, and even severely elevated pressures can be clinically silent for years. Left untreated, roughly half of hypertensive patients die of ischemic heart disease or congestive heart failure, and another third die of stroke. Treatment with blood pressure–lowering drugs dramatically reduces the incidence and death rates attributable to all forms of hypertension-related pathology.

A small percentage of hypertensive persons (as much as 5%) show a rapidly rising blood pressure that, if untreated, leads to death within 1 to 2 years. This form of hypertension, called *malignant hypertension,* is characterized by severe pressure elevations (i.e., systolic pressure more than 200 mm Hg, diastolic pressure more than 120 mm Hg), renal failure, and retinal hemorrhages and exudates, with or without papilledema (swelling of the optic nerve that reflects increased intracranial pressures). It can develop in previously normotensive persons but more often is superimposed on preexisting "benign" hypertension.

In this section, we will first briefly outline normal blood pressure homeostasis, followed by a discussion of pathogenic mechanisms that underlie hypertension and a description of hypertension-associated pathologic changes in vessels.

Blood Pressure Regulation

Blood pressure is a function of cardiac output and peripheral vascular resistance, both of which are influenced by multiple genetic and environmental factors (Fig. 11.4). The integration of the various inputs ensures adequate systemic perfusion, despite regional demand differences.

- *Cardiac output* is a function of stroke volume and heart rate. The most important determinant of stroke volume is the filling pressure, which is regulated through sodium homeostasis and its effect on blood volume. Heart rate and myocardial contractility (a second factor affecting stroke volume) are both regulated by the α- and β-adrenergic systems, which also have important effects on vascular tone.
- *Peripheral resistance* is regulated predominantly at the level of the arterioles by neural and hormonal inputs.

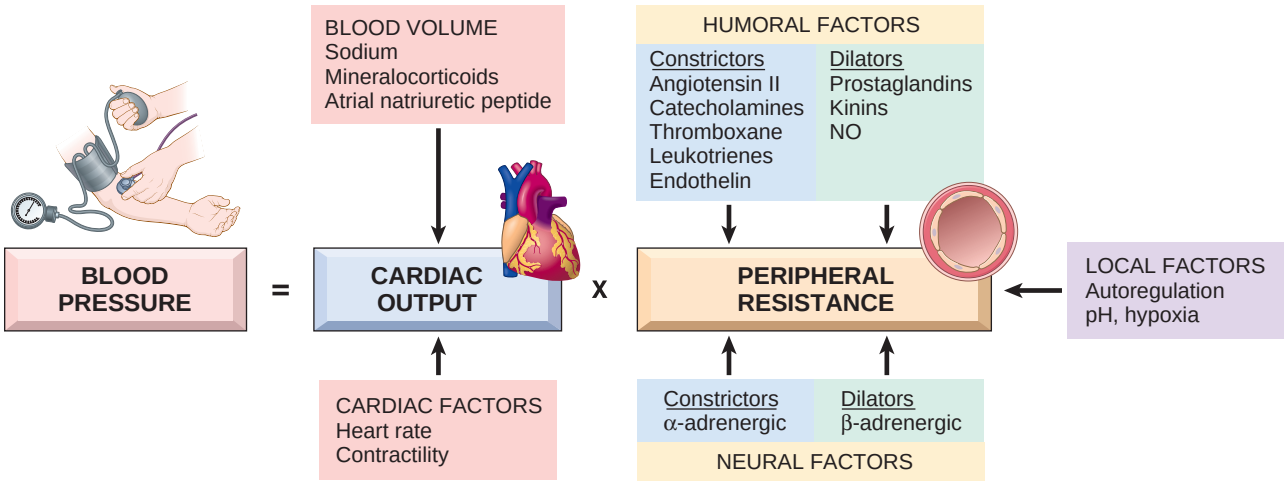


Figure 11.4 Blood pressure regulation. Diverse influences on cardiac output (e.g., blood volume and myocardial contractility) and peripheral resistance (neural, humoral, and local effectors) impact the output blood pressure.

Vascular tone reflects a balance between vasoconstrictors (including angiotensin II, catecholamines, and endothelin) and vasodilators (including kinins, prostaglandins, and NO). Resistance vessels also exhibit autoregulation, whereby increased blood flow induces vasoconstriction to protect tissues against hyperperfusion. Finally, blood pressure is fine-tuned by tissue pH and hypoxia to accommodate local metabolic demands.

Factors released from the kidneys, adrenals, and myocardium interact to influence vascular tone and to regulate blood volume by adjusting sodium balance (Fig. 11.5). The kidneys filter 170 liters of plasma containing 23 moles of salt daily. Thus with a typical diet containing 100 mEq of sodium, 99.5% of the filtered salt must be reabsorbed to maintain total body sodium levels. About 98% of the filtered sodium is reabsorbed by several constitutively active transporters. The small amount of remaining sodium is subject to resorption by the epithelial sodium channel (EnaC), which is tightly regulated by the renin-angiotensin-aldosterone system; it is this pathway that determines net sodium balance.

The kidneys and heart contain cells that sense changes in blood pressure or volume. In response, these cells release circulating effectors that act in concert to maintain normal blood pressure. Kidneys influence peripheral resistance and sodium excretion/retention primarily through the renin-angiotensin-aldosterone system.

- *Renin* is a proteolytic enzyme produced by renal juxtaglomerular cells, which are myoepithelial cells adjacent to the glomerular afferent arterioles. Renin is released in response to low blood pressure in afferent arterioles, elevated levels of circulating catecholamines, or low sodium levels in the distal convoluted renal tubules. The latter occurs when the glomerular filtration rate falls (e.g., when the cardiac output is low), leading to increased sodium resorption by the proximal tubules.
- *Renin cleaves plasma angiotensinogen to angiotensin I, which in turn is converted to angiotensin II by angiotensin-converting enzyme (ACE),* mainly a product of vascular endothelium. Angiotensin II raises blood pressure by (1) inducing vascular contraction, (2) stimulating aldosterone secretion by the adrenal gland, and (3) increasing tubular sodium resorption. Adrenal aldosterone increases blood pressure by its effect on blood volume; aldosterone increases sodium resorption (and thus water) in the distal convoluted tubules, which increases blood volume.
- *The kidney also produces a variety of vascular relaxing substances* (including prostaglandins and NO) that can counterbalance the vasopressor effects of angiotensin.

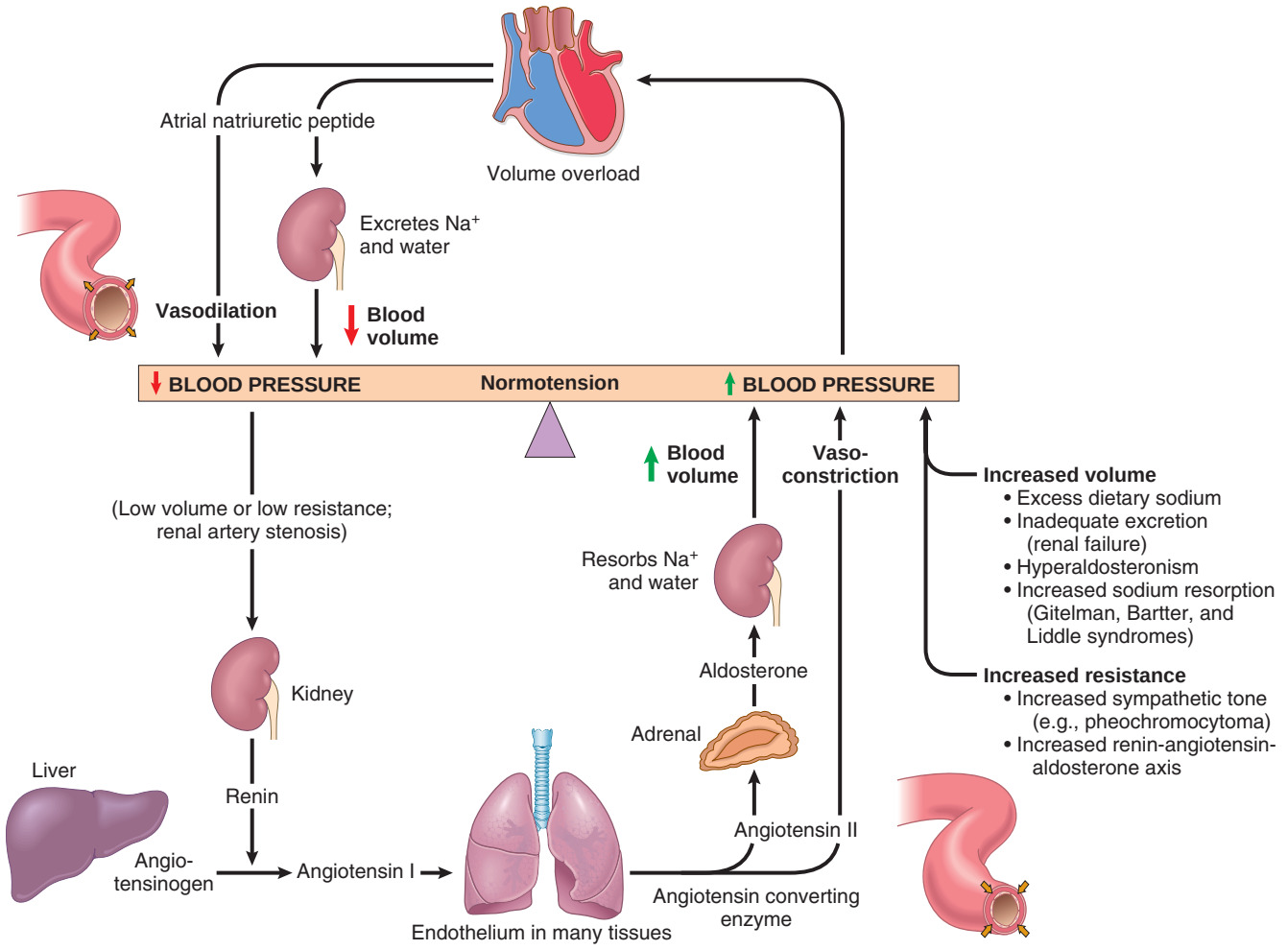


Figure 11.5 Interplay of renin-angiotensin-aldosterone and atrial natriuretic peptide in maintaining blood pressure homeostasis.

Myocardial natriuretic peptides are released from atrial (dominant contributor) and ventricular (minor contributor) myocardium in response to volume expansion; these inhibit sodium resorption in the distal renal tubules, thus leading to sodium excretion and diuresis. They also induce systemic vasodilation.

KEY CONCEPTS

BLOOD PRESSURE REGULATION

- Blood pressure is determined by vascular resistance and cardiac output.
- Vascular resistance is regulated at the level of the arterioles, influenced by neural and hormonal inputs.
- Cardiac output is determined by heart rate and stroke volume, which is strongly influenced by blood volume. Blood volume in turn is regulated mainly by renal sodium excretion or resorption.
- Renin, a major regulator of blood pressure, is secreted by the kidneys in response to decreased blood pressure in afferent arterioles. In turn, renin cleaves angiotensinogen to angiotensin I; subsequent endothelial catabolism produces angiotensin II, which regulates blood pressure by increasing vascular SMC tone and by increasing adrenal aldosterone secretion, thereby increasing renal sodium resorption.

Pathogenesis of Hypertension

The vast majority (90% to 95%) of hypertension is idiopathic, the result of interacting genetic and environmental factors. Even without knowing the specific lesions, it is reasonable to suppose that small changes in renal sodium homeostasis and/or vessel wall tone or structure act in combination to cause essential hypertension (see Fig. 11.5). Most other causes fall under the general rubric of renal disease, including renovascular hypertension (due to renal artery occlusion). Infrequently, hypertension has an underlying endocrine basis.

Mechanisms of Secondary Hypertension. For many of the secondary forms of hypertension, the underlying pathways are reasonably well understood.

- In *renovascular hypertension,* renal artery stenosis causes decreased glomerular flow and pressure in the afferent arteriole of the glomerulus. As already discussed, this induces renin secretion leading to increased blood volume and vascular tone via angiotensin and aldosterone pathways (see Fig. 11.5).
- *Primary hyperaldosteronism* is one of the most common causes of secondary hypertension (Chapter 24). It may be idiopathic or less commonly caused by aldosterone-secreting adrenal adenomas.
- *Single-gene disorders* cause severe but rare forms of hypertension.
  - *Gene defects affecting enzymes involved in aldosterone metabolism (e.g., aldosterone synthase, 11β-hydroxylase, 17α-hydroxylase)* can lead to increased aldosterone secretion with downstream increases in salt and water resorption, plasma volume expansion, and, ultimately, hypertension.
  - *Mutations affecting proteins that influence sodium reabsorption.* For example, the moderately severe form of salt-sensitive hypertension, called Liddle syndrome, is caused by mutations in an epithelial Na^+ channel protein that increase distal tubular reabsorption of sodium in response to aldosterone.

Mechanisms of Essential Hypertension. As mentioned earlier, in the vast majority of cases hypertension results from complex interactions between multiple genetic and environmental influences.

- *Genetic factors* definitely contribute to blood pressure regulation, as shown by comparisons of monozygotic and dizygotic twins and genetically related versus adopted children. Moreover, as noted earlier, several single-gene disorders cause relatively rare forms of hypertension (and hypotension) by altering net sodium reabsorption in the kidney. Large genome-wide association studies point to more than 60 genetic loci in which variants individually contribute minimally to blood pressure levels but in sum have larger effects.
- *Insufficient renal sodium excretion* in the presence of normal arterial pressure may be a key initiating event in essential hypertension and, indeed, a final common pathway for the pathogenesis of hypertension. Insufficient sodium excretion may lead sequentially to an increase in fluid volume, increased cardiac output, and peripheral vasoconstriction, thereby elevating blood pressure. At the higher blood pressure, enough additional sodium is excreted by the kidneys to equal intake and prevent further fluid retention. Thus a new steady state of sodium balance is achieved ("resetting of pressure natriuresis"), but at the expense of an increase in blood pressure.
- *Vasoconstrictive influences,* such as factors that induce vasoconstriction or stimuli that cause structural changes in the vessel wall, can lead to an increase in peripheral resistance and may also play a role in essential hypertension.
- *Environmental factors* such as stress, obesity, smoking, physical inactivity, and heavy salt consumption all are implicated in hypertension. Indeed, the evidence linking dietary sodium intake with the prevalence of hypertension in different populations is particularly impressive.

Vascular Pathology in Hypertension

Hypertension not only accelerates atherogenesis (see later) but also causes degenerative changes in the walls of large and medium arteries that can lead to both aortic dissection and cerebrovascular hemorrhage. Three forms of small vessel disease are hypertension-related (Fig. 11.6).

MORPHOLOGY

Hyaline arteriolosclerosis. Arterioles show homogeneous, pink hyaline thickening with associated luminal narrowing (Fig. 11.6A). These changes reflect both plasma protein leakage across injured ECs and increased SMC matrix synthesis in response to the chronic hemodynamic pressures of hypertension. Although the vessels of older patients (either normotensive or hypertensive) also frequently exhibit hyaline arteriosclerosis, it is more generalized and severe in patients with hypertension and diabetes (Chapter 24). In **nephrosclerosis** due to chronic hypertension, the arteriolar

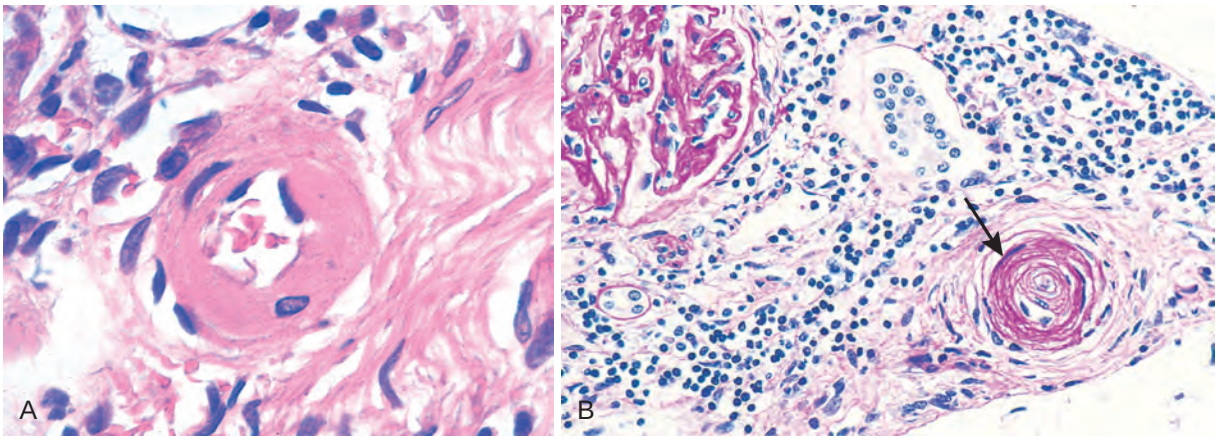


Figure 11.6 Vascular pathology in hypertension. (A) Hyaline arteriolosclerosis. The arteriolar wall is thickened with increased protein deposition (hyalinized), and the lumen is markedly narrowed. (B) Hyperplastic arteriolosclerosis (onion-skinning) causing luminal obliteration *(arrow)* (periodic acid-Schiff stain). (B, Courtesy Helmut Rennke, MD, Brigham and Women's Hospital, Boston, Mass.)

narrowing causes diffuse impairment of renal blood supply and glomerular scarring (Chapter 20).

Hyperplastic arteriolosclerosis. This lesion occurs in severe hypertension; vessels exhibit concentric, **laminated ("onion-skin")** thickening of the walls with luminal narrowing (Fig. 11.6B). The laminations consist of SMCs with thickened, reduplicated basement membrane; in malignant hypertension, they are accompanied by fibrinoid deposits and vessel wall necrosis **(necrotizing arteriolitis)**, particularly in the kidney (Chapter 20).

Pulmonary hypertension can be caused by several entities including left heart failure, congenital heart disease, valve disorders, obstructive or interstitial lung disease, and recurrent thromboemboli. The arterioles in such affected lungs typically show histologic changes ranging from fibrotic intimal thickening to medial hyperplasia. These are described in greater detail in Chapter 15.

KEY CONCEPTS

HYPERTENSION

- Hypertension is a common disorder affecting roughly half of adults in the United States; it is a major risk factor for atherosclerosis, congestive heart failure, and renal failure.
- Essential hypertension represents 90% to 95% of cases and is a complex, multifactorial disorder involving both environmental influences and genetic polymorphisms that influence sodium resorption, aldosterone pathways, and the renin-angiotensin-aldosterone system.
- Hypertension is occasionally caused by single-gene disorders or is secondary to diseases of the kidney, adrenal, or other endocrine organs.
- Sustained hypertension requires participation of the kidney, which normally responds to hypertension by eliminating salt and water. In established hypertension, both increased blood volume and increased peripheral resistance contribute to the increased pressure.
- Histologically, hypertension is associated with thickening of arterial walls caused by hyaline deposits and, in severe cases, by proliferation of ECs or SMCs and replication of the basement membrane.

ARTERIOSCLEROSIS

Arteriosclerosis literally means "hardening of the arteries"; it is a generic term for arterial wall thickening and loss of elasticity. There are four general patterns, with different clinical and pathologic consequences.

- *Arteriolosclerosis* affects small arteries and arterioles and may cause downstream ischemic injury. The two anatomic variants, hyaline and hyperplastic, were discussed earlier in relation to hypertension.
- *Mönckeberg medial sclerosis* is characterized by calcifications of the medial walls of muscular arteries, typically starting along the internal elastic membrane. Adults older than age 50 are most commonly affected. The calcifications do not encroach on the vessel lumen and are usually not clinically significant.
- *Fibromuscular intimal hyperplasia* occurs in muscular arteries larger than arterioles. It is driven by inflammation (as in a healed arteritis or transplant-associated arteriopathy; see Chapter 12) or by mechanical injury (e.g., associated with stents or balloon angioplasty; see later) and can be considered as a healing response. The affected vessels can become quite stenotic; indeed, such intimal hyperplasia underlies in-stent restenosis and is the major long-term limitation of solid-organ transplants.
- *Atherosclerosis,* from Greek root words for "gruel" and "hardening," is the most frequent and clinically important pattern and is discussed here.

ATHEROSCLEROSIS

Atherosclerosis underlies the pathogenesis of coronary, cerebral, and peripheral vascular disease and causes more morbidity and mortality (roughly half of all deaths) in the Western world than any other disorder. Because coronary artery disease is an important manifestation of the disease, epidemiologic data related to atherosclerosis mortality typically reflect deaths caused by ischemic heart disease (Chapter 12); indeed, myocardial infarction is responsible for almost a quarter of all deaths in the United

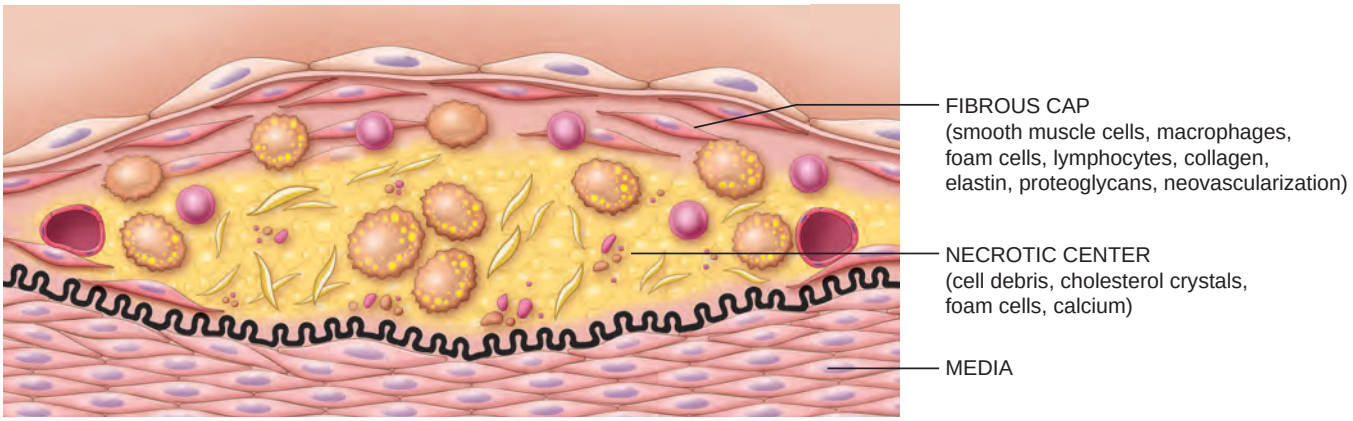


Figure 11.7 Basic structure of an atherosclerotic plaque. Note that atherosclerosis is an intimal-based process with a complex interplay of cells and extracellular materials. Plaques can have secondary effects on the underlying media including a reduction in smooth muscle cells.

States. Significant morbidity and mortality are also caused by aortic and carotid atherosclerotic disease and stroke.

The likelihood of atherosclerosis is determined by the combination of acquired (e.g., cholesterol levels, smoking, hypertension), inherited (e.g., low-density lipoprotein [LDL] receptor gene mutations), and gender- and age-associated risk factors. Acting in concert, they cause intimal lesions called *atheromas* (also called *atheromatous* or *atherosclerotic plaques*) that protrude into vessel lumens. An atheromatous plaque typically consists of a raised lesion with a soft grumous core of lipid (mainly cholesterol and cholesterol esters) covered by a fibrous cap (Fig. 11.7). Besides mechanically obstructing blood flow, atherosclerotic plaques can rupture leading to catastrophic obstructive vascular thrombosis. Atherosclerotic plaques can also increase the diffusion distance from the lumen to the media, leading to ischemic injury and weakening of the vessel wall, changes that can result in aneurysm formation.

Epidemiology. Although atherosclerosis-associated ischemic heart disease is ubiquitous among most developed nations, risk reduction and improved therapies have combined to moderate the associated mortality. At the same time, reduced mortality from infectious diseases and the adoption of Western lifestyles has led to the increased prevalence of ischemic heart disease in low income nations. As a result, the death rate for coronary artery disease in Africa, India, and Southeast Asia now exceeds that in the United States; eastern European countries have rates 3 to 5 times higher than the United States and 7 to 12 times higher than Japan.

Risk Factors

The prevalence and severity of atherosclerosis and ischemic heart disease among individuals and groups are related to a number of risk factors identified through several prospective analyses (e.g., the Framingham Heart Study); some of these are constitutional (and therefore less controllable), but others are acquired or related to specific behaviors and potentially amenable to intervention (Table 11.3). These risk factors typically have greater than additive effects, but treatment (even less than optimal) can mitigate some of the risk (Fig. 11.8).

Constitutional Risk Factors

- *Genetics.* Family history is the most important independent risk factor for atherosclerosis. Certain Mendelian disorders are strongly associated with atherosclerosis (e.g., familial hypercholesterolemia; see Chapter 5), but they account for only a small percentage of cases. The well-established familial predisposition to atherosclerosis and ischemic heart disease is usually polygenic, relating to small effects of many shared alleles common to a family or population.
- *Age is a dominant influence.* The development of atherosclerotic plaque is a progressive process that usually becomes clinically manifest in middle age or later (see later). Thus between ages 40 and 60, the incidence of myocardial infarction increases fivefold. Death rates from ischemic heart disease rise with each decade even into advanced age. Increasingly, however, this age association is being recognized as perhaps more than just the accumulated slings and arrows of vascular injury over the years. Indeed, with aging, there is a tendency for the outgrowth of hematopoietic clones (so-called *clonal hematopoiesis of indeterminate potential [CHIP]*) carrying mutations that confer a proliferative advantage. Many of these affect DNA modifications and transcriptional regulation (e.g., *TET2* encoding an enzyme that converts methylcytosine to 5-hydroxymethylcytosine); as expected, these can ultimately influence the risk of developing hematologic malignancies. Perhaps more remarkable, however, is that such clonal hematopoiesis is even more

Table 11.3 Major Risk Factors for Atherosclerosis

Nonmodifiable (Constitutional)
Genetic abnormalities
Family history
Increasing age
Male gender
Modifiable
Hyperlipidemia
Hypertension
Cigarette smoking
Diabetes
Inflammation

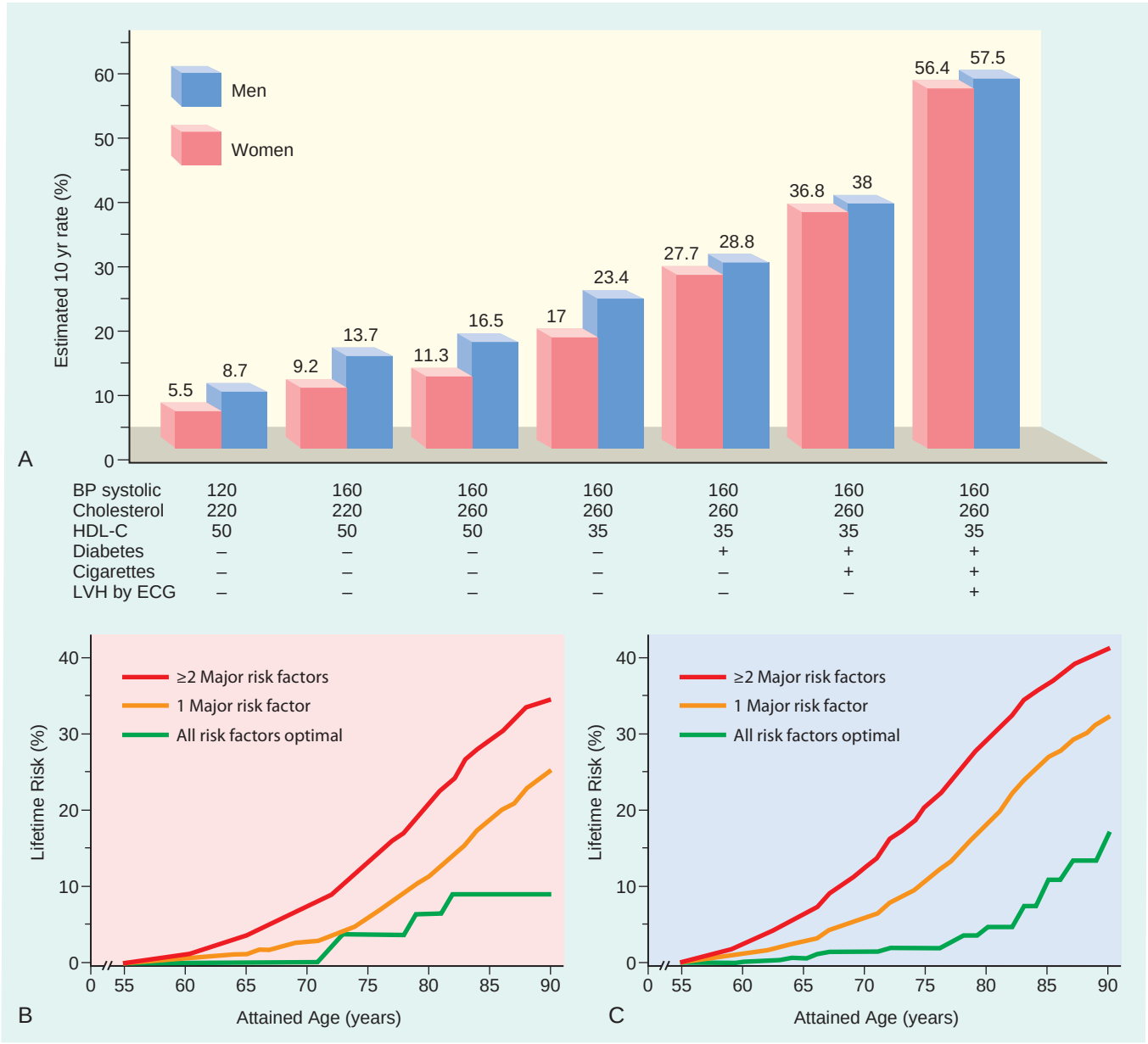


Figure 11.8 Lifetime risk of death from cardiovascular disease. (A) Estimated 10-year risk of coronary artery disease in hypothetical 55-year-old men and women as a function of traditional risk factors (hyperlipidemia, hypertension, smoking, and diabetes). *BP,* Blood pressure; *ECG,* electrocardiogram; *HDL-C,* high-density lipoprotein cholesterol; *LVH,* left ventricular hypertrophy. In women (B) and men (C), one or more risk factors of blood pressure, cholesterol, diabetes, and cigarettes significantly increases the lifetime risk of a cardiovascular event. (A, From O'Donnell CJ, Kannel WB: Cardiovascular risks of hypertension: lessons from observational studies, *J Hypertens Suppl* 16(6):S3–S7, 1998, with permission from Lippincott Williams & Wilkins; B and C, Modified from Berry JD, Dyer A, Cai X, et al: Lifetime risks of cardiovascular disease, *N Eng J Med* 366:321–329, 2012.)

significantly associated with increased all-cause cardiovascular mortality. An explanation is beginning to emerge that the same CHIP mutations that affect cellular proliferation (e.g., *TET2*) can also impact the inflammatory response of mononuclear cells and thereby influence atherogenesis (see also later).

- *Gender.* Premenopausal women are relatively protected against atherosclerosis and its consequences compared with age-matched men. Thus myocardial infarction and other complications of atherosclerosis are uncommon in premenopausal women unless they are otherwise predisposed by diabetes, hyperlipidemia, or severe hypertension. After menopause, however, the incidence of atherosclerosis-related diseases increases and at older ages exceeds that of men. Although a favorable influence of estrogen has been proposed to explain this effect, clinical trials of estrogen replacement did not show any benefit; indeed, postmenopausal estrogen therapy actually increased cardiovascular risk in some older women.

Modifiable Major Risk Factors

- *Hyperlipidemia*—and more specifically hypercholesterolemia—is a major risk factor for atherosclerosis; even in the absence of other risk factors, hypercholesterolemia is sufficient to initiate lesion development. The major component of serum cholesterol associated with increased risk is LDL cholesterol

("bad cholesterol"). LDL is the lipid-cholesterol-protein complex that delivers cholesterol to peripheral tissues; in contrast, high-density lipoprotein (HDL) is the complex that mobilizes cholesterol from the periphery (including atheromas) and transports it to the liver for catabolism and biliary excretion. Higher levels of HDL ("good cholesterol") correlate with reduced risk.

Understandably, dietary and pharmacologic interventions that lower LDL or total serum cholesterol are of considerable interest. Interestingly, approaches that exclusively raise HDL are not effective. Although previously considered important, the contribution of most dietary fats to atherosclerosis is now viewed as minimal. Nevertheless, omega-3 fatty acids (abundant in fish oils) are considered beneficial, whereas trans unsaturated fats produced by artificial hydrogenation of polyunsaturated oils (used in baked goods and margarine) can adversely affect cholesterol profiles. *Statins* are a class of drugs that lower circulating cholesterol levels by inhibiting hydroxymethylglutaryl coenzyme A (HMG CoA) reductase, the rate-limiting enzyme in hepatic cholesterol biosynthesis (Chapter 5). Statins are widely used to lower serum cholesterol levels, lowering rates of myocardial infarctions, arguably one of the most significant success stories of translational research. Interestingly, some of the benefit of the statins may be due to "off-target" effects on reducing inflammation (see also later).

- *Hypertension* can increase the risk of ischemic heart disease by approximately 60% versus normotensive populations. Chronic hypertension is the most common cause of left ventricular hypertrophy, and hence the latter is also a surrogate marker for cardiovascular risk.
- *Cigarette smoking*, and in particular prolonged (years) use, doubles the death rate from ischemic heart disease. Smoking cessation reduces that risk substantially.
- *Diabetes mellitus* induces hypercholesterolemia (Chapter 24) and markedly increases the risk of atherosclerosis. Other factors being equal, the incidence of myocardial infarction is twice as high in diabetics relative to normoglycemic individuals. There is also an increased risk of strokes and a 100-fold increased risk of atherosclerosis-induced gangrene of the lower extremities.

Additional Risk Factors

As many as 20% of all cardiovascular events occur in the absence of major risk factors (e.g., hypertension, hyperlipidemia, smoking, or diabetes). Indeed, more than 75% of cardiovascular events in previously healthy women occur with LDL cholesterol levels below 160 mg/dL (levels generally considered to connote low risk). Clearly, other factors are contributory. Among those that are proven or suspected are the following:

- *Inflammation*. Inflammation is present during all stages of atherogenesis and is intimately linked with atherosclerotic plaque formation and rupture (see later). With the increasing recognition that inflammation plays a significant causal role in ischemic heart disease, assessment of systemic inflammation has become important in overall risk stratification. A number of circulating markers of inflammation correlate with ischemic heart disease, and *C-reactive protein (CRP)* has emerged as one of the most stable and simplest to measure.

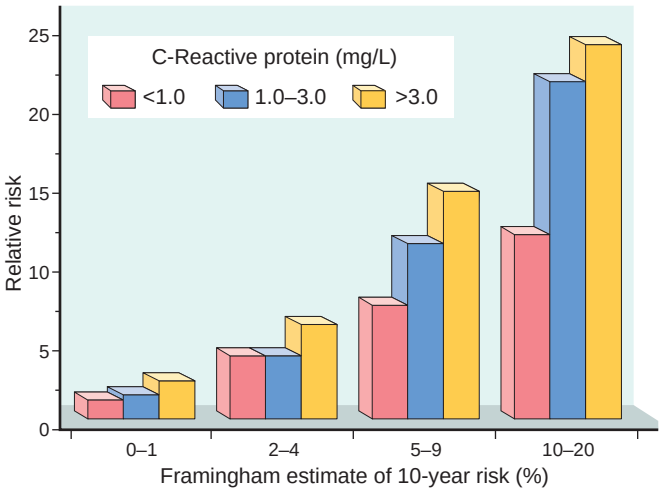


Figure 11.9 C-reactive protein adds prognostic information at all levels of traditional risk identified from the Framingham Heart Study. Relative risk (y-axis) refers to the risk of a cardiovascular event (e.g., myocardial infarction). The x-axis is the 10-year risk of a cardiovascular event derived from the traditional risk factors identified in the Framingham study. In each risk group, C-reactive protein values further stratify the patients. (Data from Ridker PM, et al: Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events, *N Engl J Med* 347:1557, 2002.)

CRP is an acute phase reactant synthesized primarily by the liver. Although CRP does not appear to be causally related to the development of atherosclerosis or its sequelae, it is well established that plasma CRP is a strong, independent marker of risk for myocardial infarction, stroke, peripheral arterial disease, and sudden cardiac death, even among apparently healthy individuals (Fig. 11.9). Accordingly, CRP levels are now incorporated into risk stratification algorithms. Interestingly, CRP levels are also typically lowered in concert with other risk reduction measures including smoking cessation, weight loss, exercise, and statin administration.

- *Hyperhomocysteinemia.* Serum homocysteine levels correlate with coronary atherosclerosis, peripheral vascular disease, stroke, and venous thrombosis. *Homocystinuria,* due to rare inborn errors of metabolism, results in elevated circulating homocysteine (>100 μmol/L) and is associated with premature vascular disease.
- *Metabolic syndrome.* Associated with central obesity (Chapter 9), this entity is characterized by insulin resistance, hypertension, dyslipidemia (elevated LDL and depressed HDL), hypercoagulability, and a proinflammatory state. Dyslipidemia, hyperglycemia, and hypertension all are cardiac risk factors, while the systemic hypercoagulable and proinflammatory state may contribute to endothelial dysfunction and/or thrombosis.
- *Lipoprotein a [Lp(a)]* is an altered form of LDL that contains the apolipoprotein B-100 portion of LDL linked to apolipoprotein A (apo A); Lp(a) levels are associated with coronary and cerebrovascular disease risk, independent of total cholesterol or LDL levels.
- *Factors affecting hemostasis*. Several markers of hemostatic and/or fibrinolytic function (e.g., elevated plasminogen activator inhibitor 1) are potent predictors of risk for

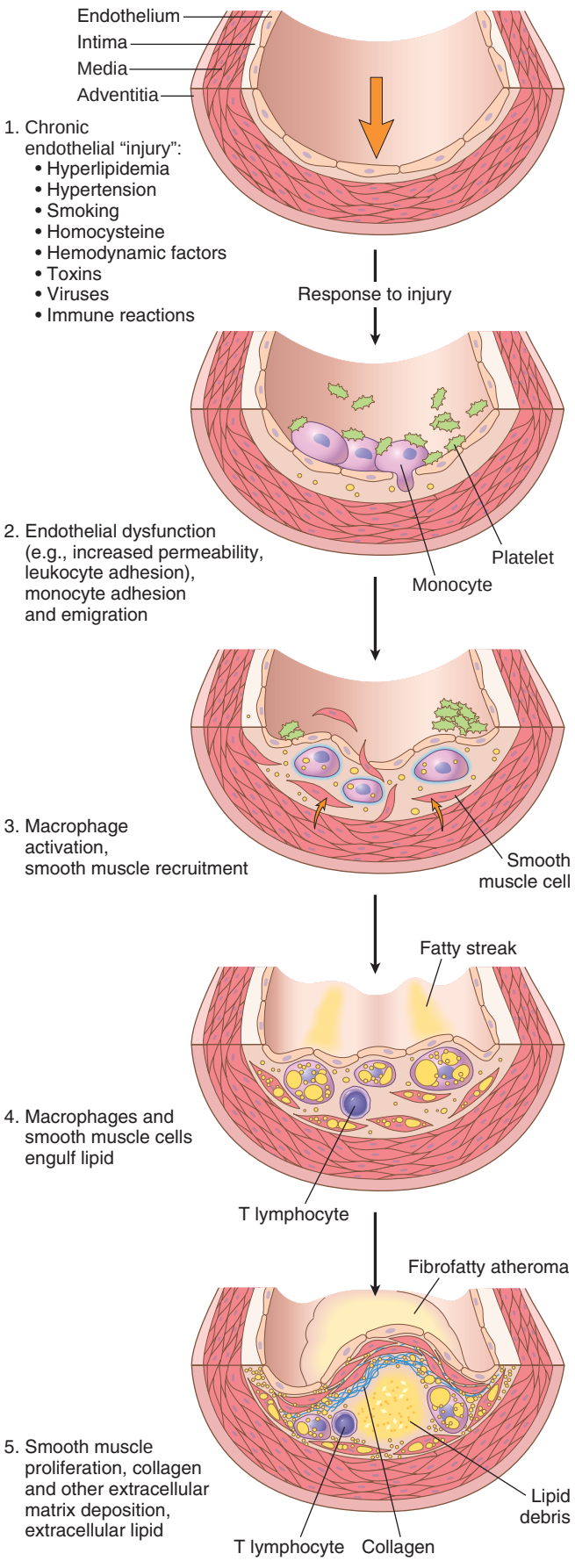


Figure 11.10 Evolution of arterial wall changes in the response to injury hypothesis. *1*, Normal. *2*, Endothelial injury with monocyte and platelet adhesion. *3*, Monocyte and smooth muscle cell migration into the intima, with macrophage activation. *4*, Macrophage and smooth muscle cell uptake of modified lipids, with further activation. *5*, Intimal smooth muscle cell proliferation with extracellular matrix production, forming a well-developed plaque.

major atherosclerotic events including myocardial infarction and stroke. Platelet-derived factors as well as thrombin—through both procoagulant and proinflammatory effects—are increasingly recognized as major contributors to local vascular pathology.

- *Other factors*. Factors associated with a less pronounced and/or difficult to quantitate risk include lack of exercise; competitive, stressful lifestyle (type A personality); and obesity (the last-mentioned also being complicated by hypertension, diabetes, hypertriglyceridemia, and decreased HDL).

Pathogenesis of Atherosclerosis

The clinical importance of atherosclerosis has stimulated enormous interest in understanding the mechanisms that underlie its evolution and complications. The contemporary view of atherogenesis integrates the risk factors previously discussed and is called the "response to injury" hypothesis. **This model views atherosclerosis as a chronic inflammatory and healing response of the arterial wall to endothelial injury. Lesion progression occurs through interaction of modified lipoproteins, macrophages, and T lymphocytes with ECs and SMCs of the arterial wall** (Fig. 11.10). According to this schema, atherosclerosis progresses in the following sequence:

- *Endothelial injury and dysfunction,* causing (among other things) increased vascular permeability, leukocyte adhesion, and thrombosis.
- *Accumulation of lipoproteins* (mainly LDL and its oxidized forms) in the vessel wall.
- *Monocyte adhesion to the endothelium,* followed by migration into the intima and transformation into *macrophages* and *foam cells.*
- *Platelet adhesion.*
- *Factor release* from activated platelets, macrophages, and vascular wall cells, inducing *SMC* recruitment, either from the media or from circulating precursors.
- *SMC proliferation, ECM production, and recruitment of T cells.*
- *Lipid accumulation* both extracellularly and within cells (macrophages and SMCs).
- *Calcification* of ECM and necrotic debris late in the pathogenesis.

We will now discuss the role of these factors in the pathogenesis of atherosclerosis in some detail, starting with endothelial injury.

Endothelial Injury. EC injury is the cornerstone of the response-to-injury hypothesis. Early lesions begin at sites of morphologically intact endothelium that exhibit

features of endothelial dysfunction—increased permeability, enhanced leukocyte adhesion, and altered gene expression. Causes of EC dysfunction include toxins from cigarette smoke, homocysteine, and the local production of inflammatory cytokines. However, the **three most important causes of endothelial dysfunction are hemodynamic disturbances, hypercholesterolemia, and inflammation.**

Hemodynamic Disturbances. The importance of hemodynamic turbulence in atherogenesis is illustrated by the observation that plaques do not occur randomly, but rather tend to locate at the ostia of exiting vessels, branch points, and along the posterior abdominal aorta where flow patterns are disturbed and nonlaminar. This is explained by the fact that laminar nonturbulent flow increases the production of transcription factors and, in particular, Krüppel-like factor-2 (KLF2) that turn on atheroprotective genes and turn off inflammatory gene transcription. Conversely, turbulent, nonlaminar flow drives a repertoire of genetic transcription that makes those sites atheroprone. It is noteworthy that some of the atheroprotective effects of statins also occur through KLF2 upregulation.

Hypercholesterolemia. Lipids are transported in the bloodstream bound to specific apoproteins (forming lipoprotein complexes). *Dyslipoproteinemias* are lipoprotein abnormalities that may be present in the general population (and, indeed, are found in many myocardial infarction survivors). These include (1) increased LDL cholesterol levels, (2) decreased HDL cholesterol levels, and (3) increased levels of the abnormal Lp(a). These may result from mutations in apoproteins or lipoprotein receptors or arise from other underlying disorders that affect circulating lipid levels such as nephrotic syndrome, alcoholism, hypothyroidism, or diabetes mellitus. All of these abnormalities are associated with an increased risk of atherosclerosis.

The evidence implicating hypercholesterolemia in atherogenesis includes the following:

- *The dominant lipids in atheromatous plaques are cholesterol and cholesterol esters.*
- *Genetic defects in lipoprotein uptake and metabolism* that cause hyperlipoproteinemia are associated with accelerated atherosclerosis. Thus familial hypercholesterolemia, caused by mutations affecting the LDL receptors and consequent inadequate hepatic LDL uptake and catabolism (Chapter 5), can precipitate myocardial infarctions before age 20 in those homozygous for the mutation. Similarly, accelerated atherosclerosis occurs in animal models with engineered deficiencies in apolipoproteins or LDL receptors.
- *Epidemiologic analyses* demonstrate a significant correlation between the severity of atherosclerosis and the levels of total plasma cholesterol or LDL.
- *Lowering serum cholesterol* by diet or drugs slows the rate of progression of atherosclerosis, causes regression of some plaques, and reduces the risk of cardiovascular events.

The mechanisms by which hyperlipidemia contributes to atherogenesis include the following:

- *Impaired EC function.* Chronic hyperlipidemia, particularly hypercholesterolemia, can directly impair EC function by increasing local reactive oxygen species production; besides causing membrane and mitochondrial damage, oxygen free radicals accelerate NO decay, dampening its vasodilator activity.
- *Modified LDL.* With chronic hyperlipidemia, lipoproteins accumulate within the intima, where they may aggregate or become oxidized by free radicals produced by inflammatory cells. Such modified LDL is then accumulated by macrophages through a variety of scavenger receptors (distinct from the LDL receptor). Because the modified lipoproteins cannot be completely degraded, chronic ingestion leads to the formation of lipid-filled macrophages called *foam cells;* SMCs can similarly transform into lipid-laden foam cells by ingesting modified lipids through LDL receptor–related proteins. Not only are the modified lipoproteins toxic to ECs, SMCs, and macrophages, but their binding and uptake also stimulate the release of growth factors, cytokines, and chemokines that create a vicious inflammatory cycle of monocyte recruitment and activation. The early lesions containing lipid-filled macrophages are called *fatty streaks.*

Inflammation. **Chronic inflammation contributes to the initiation and progression of atherosclerotic lesions.** It is believed that inflammation is triggered by the accumulation of cholesterol crystals and free fatty acids in macrophages and other cells (Fig. 11.11). These cells sense the presence of abnormal materials via cytosolic innate immune receptors that are components of the inflammasome (Chapter 6). The resulting inflammasome activation leads to the production of the proinflammatory cytokine interleukin (IL)-1, which recruits and activates mononuclear cells including macrophages and T lymphocytes. Such mononuclear cell activation, in turn, leads to local production of cytokines and chemokines that recruit and activate more inflammatory cells. Activated macrophages produce reactive oxygen species that enhance LDL oxidation and elaborate growth factors that drive SMC proliferation. Activated T lymphocytes in the growing intimal lesions elaborate inflammatory cytokines, (e.g., interferon-γ), which, in turn, can activate macrophages as well as ECs and SMCs. Thus many of the lesions of atherosclerosis are attributable to the chronic inflammatory reaction in the vessel wall.

Infection. Although circumstantial evidence has been presented linking atherosclerosis to herpesvirus, cytomegalovirus, and *Chlamydophila pneumoniae,* there is not an established causal role for infection.

Smooth Muscle Proliferation and Matrix Synthesis. Intimal SMC proliferation and ECM deposition convert a fatty streak into a mature atheroma and contribute to the progressive growth of atherosclerotic lesions (see Fig. 11.10). Several growth factors are implicated in SMC proliferation and ECM synthesis, including platelet-derived growth factor (PDGF), released by locally adherent platelets, as well as macrophages, ECs and SMCs; fibroblast growth factor; and transforming growth factor-α (Chapter 1). These factors also stimulate SMCs to synthesize ECM (notably collagen), which stabilizes atherosclerotic plaques. In contrast, activated inflammatory cells in atheromas may increase the breakdown of ECM components resulting in unstable plaques (see later).

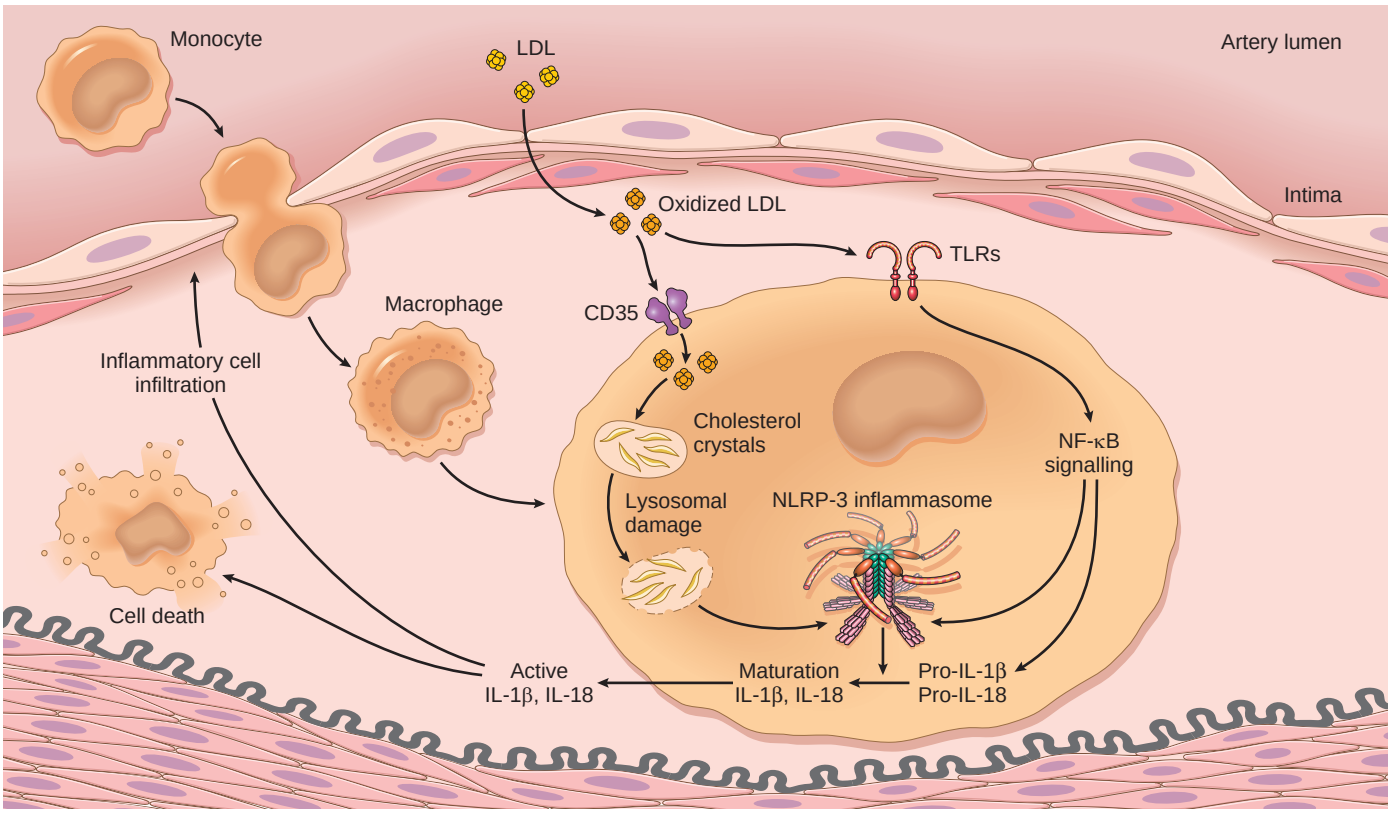


Figure 11.11 The role of cholesterol crystals in inflammasome activation and interleukin-1 production in atherosclerotic plaque. Oxidized low-density lipoprotein *(LDL)* leads to the activation of subendothelial macrophages through toll-like receptor *(TLR)* mediated initiation of downstream *NF-κB* transcription factor activation. The NF-κB complex then induces the production of a variety of proinflammatory mediators, including interleukin precursors and the components of the *NLRP3* inflammasome. Uptake of oxidized LDL and subsequent cholesterol crystallization results in assembly of the precursor components into an active inflammasome complex that proteolytically activates the pro-interleukin-1 beta *(IL-1β)* and interleukin-18 *(IL-18)* molecules. (Modified from Li X, Deroide N, Mallat Z: The role of the inflammasome in cardiovascular disease, *J Mol Med (Berl)* 92:307–319, 2014.)

Overview. Fig. 11.12 summarizes the major pathogenic pathways in atherogenesis, emphasizing the multifactorial nature of the disease. This schematic highlights the concept of atherosclerosis as a chronic inflammatory response—and ultimately an attempt at vascular "healing"—driven by a variety of insults including EC injury, lipid oxidation, lipid accumulation, and inflammation. Atheromas are dynamic lesions consisting of dysfunctional ECs, proliferating SMCs, and admixed T lymphocytes and macrophages. All four cell types are capable of liberating mediators that can influence atherogenesis. Thus at early stages, intimal plaques are little more than aggregates of SMCs, macrophages, and foam cells; death of these cells releases lipids and necrotic debris. With progression, the atheroma is modified by ECM synthesized by SMCs; connective tissue is particularly prominent on the intimal aspect forming a fibrous cap. Lesions typically retain a central core of lipid-laden cells and fatty debris that can become calcified. Although the intimal plaque can initially cause medial remodeling and outward expansion, the growing atherosclerotic plaque—with varying degrees of calcification depending on the nature of the surrounding matrix—eventually encroaches on the vessel lumen and compromises blood flow. The plaque can also compress the underlying media, leading to its degeneration, or can erode or rupture to expose thrombogenic factors resulting in thrombus formation and acute vascular occlusion.

MORPHOLOGY

Fatty streaks. Fatty streaks are composed primarily of lipid-filled foamy macrophages. Beginning as small flat yellow macules, these can eventually coalesce into elongated streaks 1 cm long or longer. The lesions are not particularly raised and do not cause any significant flow disturbance (Fig. 11.13). Although fatty streaks can evolve into plaques, not all are destined to become advanced lesions. Aortas of infants can exhibit fatty streaks, and such lesions are present in virtually all adolescents, even those without known risk factors. That coronary fatty streaks begin to form in adolescence at the same anatomic sites that later tend to develop plaques suggests a temporal evolution of these lesions.

Atherosclerotic plaque. The key processes in atherosclerosis are intimal thickening and lipid accumulation (see Figs. 11.7, 11.10, and 11.12). Atheromatous plaques are yellow-tan and are raised above the surrounding vessel wall; any superimposed thrombus over ulcerated plaques will be red-brown. Plaques vary in size but can coalesce to form larger masses (Fig. 11.14).

Atherosclerotic lesions are patchy and rarely circumferential, usually involving only a portion of any given arterial wall; on cross-section, the lesions therefore appear eccentric (Fig. 11.15A). The focality of atherosclerotic lesions—despite the uniform exposure of vessel walls to such factors as cigarette smoke toxins, elevated LDL, hyperglycemia, etc.—is attributable to the vagaries

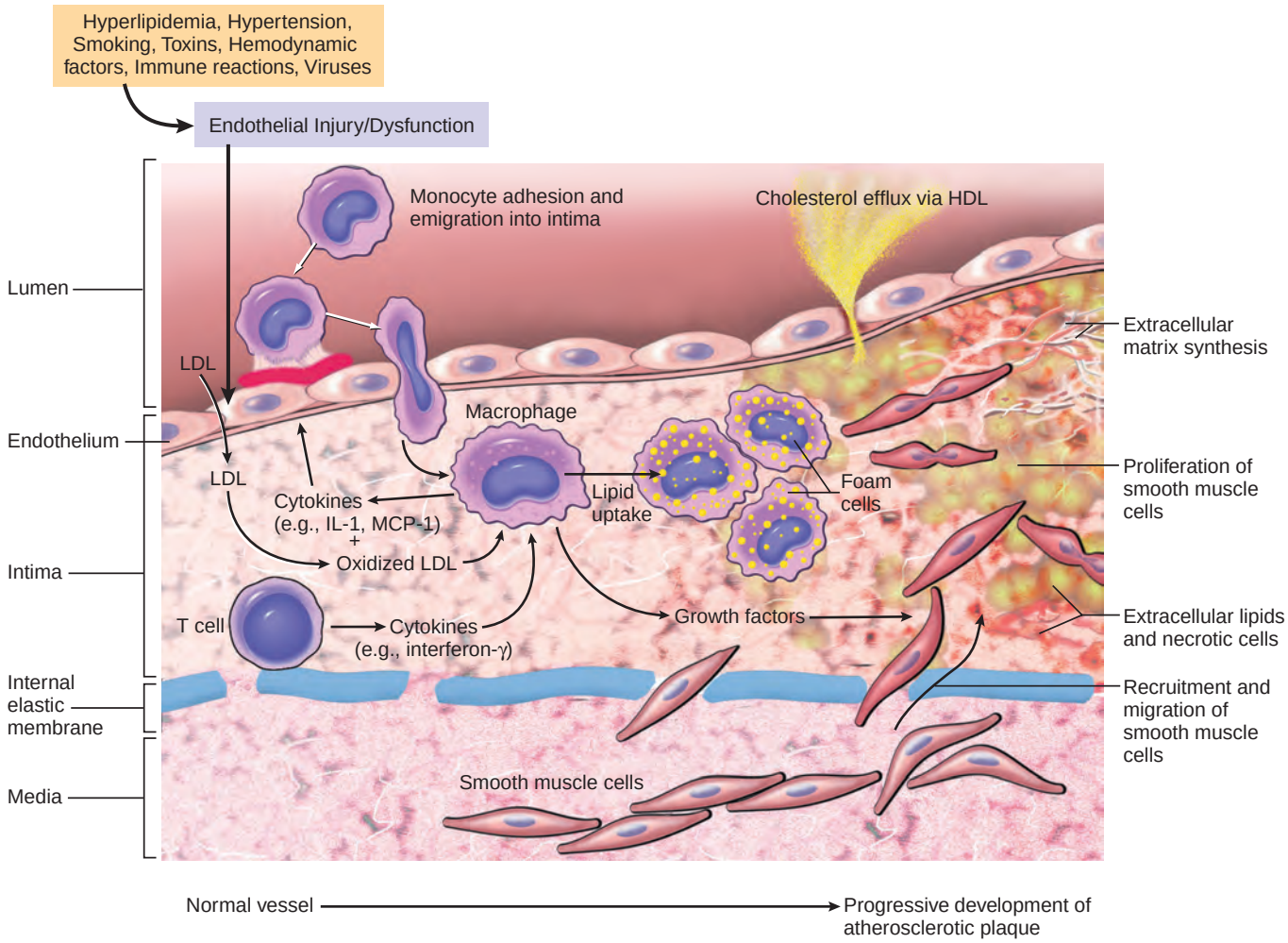


Figure 11.12 Sequence of cellular interactions in atherosclerosis. Hyperlipidemia, hyperglycemia, hypertension, and other influences cause endothelial dysfunction. This results in platelet and monocyte adhesion, with subsequent cytokine and growth factor release leading to smooth muscle cell migration and proliferation. Foam cells in atheromatous plaques derive from macrophages and smooth muscle cells that have accumulated modified lipids (e.g., oxidized and aggregated low-density lipoprotein *[LDL]*) via scavenger and LDL-receptor–related proteins. Extracellular lipid is derived from insudation from the vessel lumen, particularly in the presence of hypercholesterolemia, as well as from degenerating foam cells. Cholesterol accumulation in the plaque reflects an imbalance between influx and efflux; high-density lipoprotein *(HDL)* likely helps clear cholesterol from these accumulations. In response to the elaborated cytokines and chemokines, smooth muscle cells migrate to the intima, proliferate, and produce extracellular matrix including collagen and proteoglycans. *IL-1,* Interleukin-1; *MCP-1,* monocyte chemoattractant protein-1.

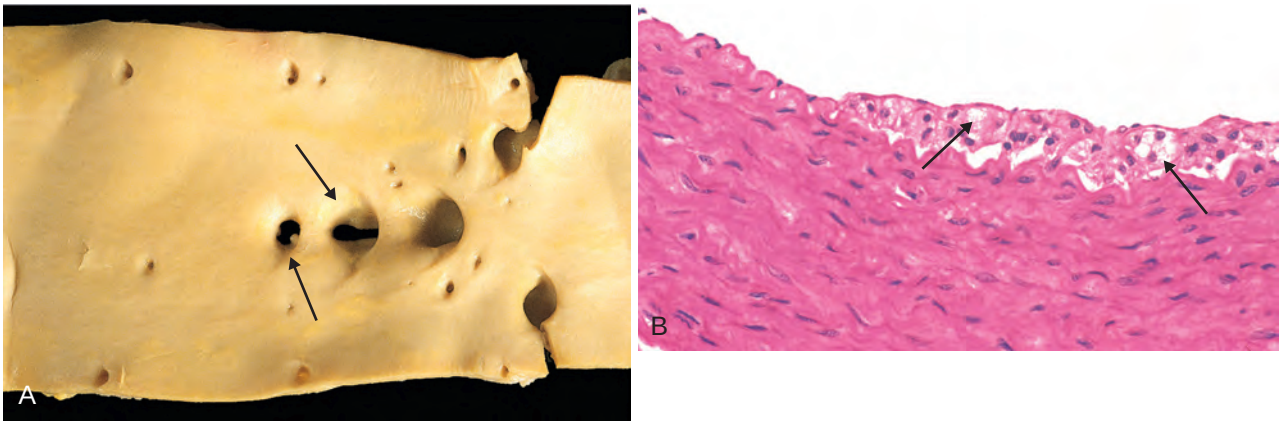


Figure 11.13 Fatty streak, a collection of foamy macrophages in the intima. (A) Aorta with fatty streaks *(arrows)*, associated largely with the ostia of branch vessels. (B) Photomicrograph of fatty streak in an experimental hypercholesterolemic rabbit, demonstrating intimal, macrophage-derived foam cells *(arrows)*. (B, Courtesy Myron I. Cybulsky, MD, University of Toronto, Toronto, Ontario, Canada).

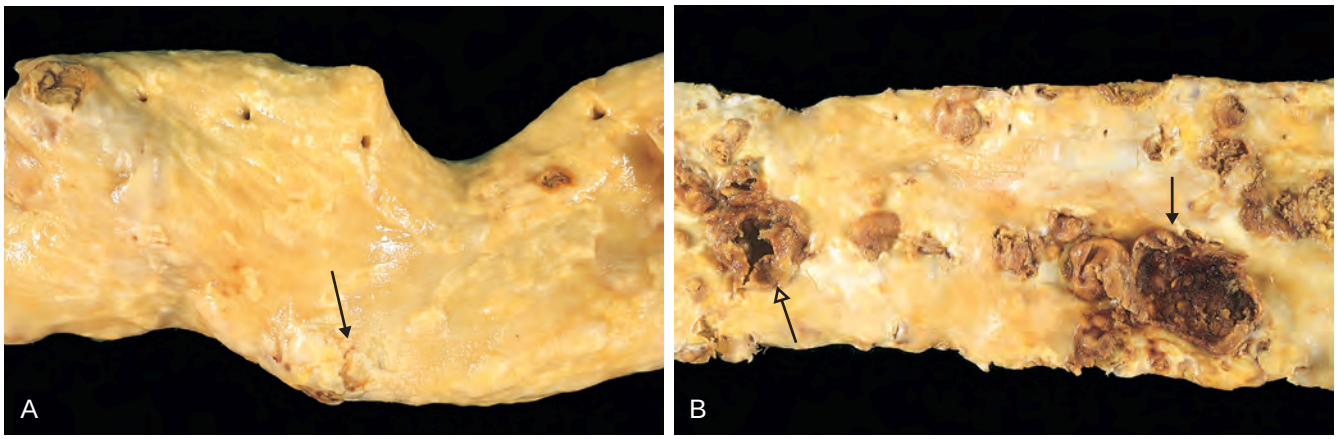


Figure 11.14 Gross views of atherosclerosis in the aorta. (A) Mild atherosclerosis composed of fibrous plaques, one of which is denoted by the *arrow.* (B) Severe disease with diffuse, complicated lesions including an ulcerated plaque *(open arrow)* and a lesion with overlying thrombus *(closed arrow)*.

of vascular hemodynamics. Local flow disturbances, such as nonlaminar flow (turbulence) at branch points, make certain portions of a vessel wall more susceptible to plaque formation. Although initially focal and sparsely distributed, atherosclerotic lesions tend to enlarge over time and become more numerous and broadly distributed. Moreover, in any given vessel, lesions at various stages often coexist.

In descending order, **the most extensively involved vessels are the lower abdominal aorta and iliac arteries, the coronary arteries, the popliteal arteries, the internal carotid arteries, and the vessels of the circle of Willis.** Vessels of the upper extremities are usually relatively spared, as are the mesenteric and renal arteries except at their ostia. Although most individuals tend to have a consistent degree of atherosclerotic burden in the affected vasculature, the severity of disease in one arterial distribution does not necessarily predict severity in another.

Atherosclerotic plaques have four principal components: (1) cells including variable numbers of SMCs, macrophages, and T lymphocytes; (2) ECM including collagen, elastic fibers, and proteoglycans; (3) intracellular and extracellular lipids; and (4) calcifications in later stage plaques (Fig. 11.15B and C). These components occur in varying proportions and configurations in different lesions. Typically, there is a superficial fibrous cap composed of SMCs and relatively dense collagen. Beneath and to the side of the cap (the "shoulder") is a more cellular area containing macrophages, T lymphocytes, and SMCs. Deep to the fibrous cap is a necrotic core containing lipid (primarily cholesterol and cholesterol esters), debris from dead cells, foam cells (lipid-laden macrophages and lipid-laden SMCs), fibrin, thrombus in varying degrees of organization, and other plasma proteins; the cholesterol content is frequently present as crystalline aggregates that are washed out during routine tissue processing

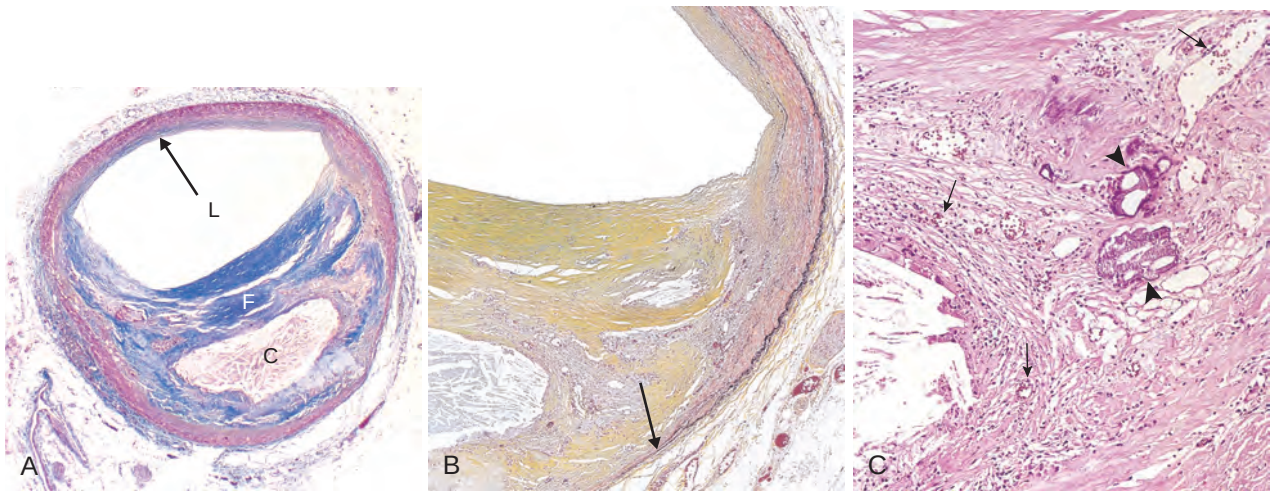


Figure 11.15 Histologic features of atheromatous plaque in the coronary artery. (A) Overall architecture demonstrating fibrous cap *(F)* and a central necrotic core *(C)* containing cholesterol and other lipids. The lumen *(L)* has been moderately compromised. Note that a segment of the wall is plaque-free *(arrow)*; the lesion is therefore "eccentric." In this section, collagen has been stained blue (Masson trichrome stain). (B) Higher power photograph of a section of the plaque shown in (A), stained for elastin (black), demonstrating that the internal and external elastic membranes are attenuated and the media of the artery is thinned under the most advanced plaque *(arrow)*. (C) Higher magnification photomicrograph at the junction of the fibrous cap and core showing scattered inflammatory cells, calcification *(arrowhead)*, and neovascularization *(small arrows)*.

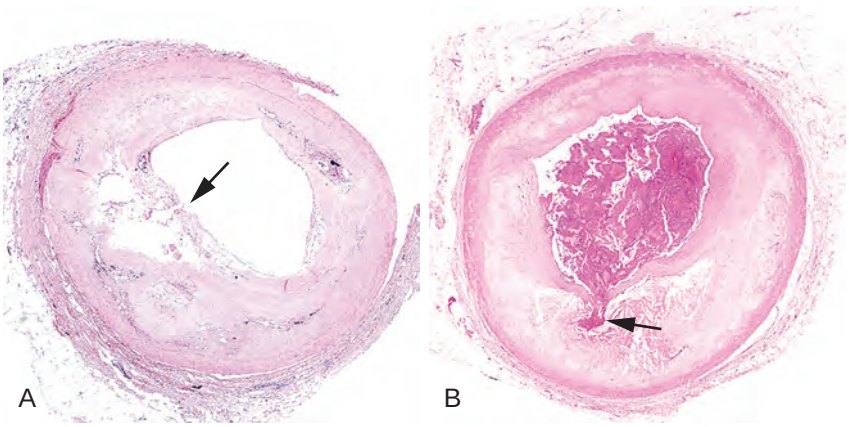


Figure 11.16 Atherosclerotic plaque rupture. (A) Plaque rupture without superimposed thrombus in a patient who died suddenly. (B) Acute coronary thrombosis superimposed on an atherosclerotic plaque with focal disruption of the fibrous cap, triggering fatal myocardial infarction. In both (A) and (B), an *arrow* points to the site of plaque rupture. (B, Reproduced from Schoen FJ: *Interventional and Surgical Cardiovascular Pathology: Clinical Correlations and Basic Principles*, Philadelphia, 1989, WB Saunders, p 61.)

and leave behind only empty "clefts." The periphery of the lesions demonstrate **neovascularization** (proliferating small blood vessels) (Fig. 11.15C). Most atheromas contain abundant lipid, but some plaques ("fibrous plaques") are composed almost exclusively of SMCs and fibrous tissue, which can be heavily calcified.

Plaques generally continue to change and progressively enlarge as a consequence of cell death and degeneration, synthesis and degradation (remodeling) of ECM, organization of any superimposed thrombus, and secondary calcification of phospholipid or necrotic debris (see Fig. 11.15C).

Atherosclerotic plaques are susceptible to clinically important pathologic changes. As discussed further later, the thickness and ECM content of the overlying fibrous cap will impact the stability or fragility of the plaque and its tendency to undergo secondary changes.

- **Rupture, ulceration, or erosion** of the surface of atheromatous plaques exposes the blood stream to highly thrombogenic substances and leads to **thrombosis,** which can partially or completely occlude the vessel lumen (Fig. 11.16). If the patient survives, the thrombus may organize and become incorporated into the growing plaque.
- **Hemorrhage into a plaque.** Rupture of the overlying fibrous cap or of the thin-walled vessels in the areas of neovascularization can cause intraplaque hemorrhage; a contained hematoma may expand the plaque or induce plaque rupture.
- **Atheroembolism.** Plaque rupture can discharge atherosclerotic debris into the blood stream, producing microemboli.
- **Aneurysm formation.** Atherosclerosis-induced pressure or ischemic atrophy of the underlying media, with loss of elastic tissue, causes weakness and potential rupture.

Consequences of Atherosclerotic Disease

Myocardial infarction (heart attack), cerebral infarction (stroke), aortic aneurysms, and peripheral vascular disease (gangrene of the legs) are the major consequences of atherosclerosis. Large elastic arteries (e.g., aorta, carotid, and iliac arteries) and large- and medium-sized muscular arteries (e.g., coronary and popliteal arteries) are the major targets of atherosclerosis. Symptomatic atherosclerotic disease most often involves the arteries supplying the heart, brain, kidneys, and lower extremities. The natural history, principal morphologic features, and main pathogenic events are schematized in Fig. 11.17.

We next describe the features of atherosclerotic lesions that are typically responsible for the clinicopathologic manifestations.

Atherosclerotic Stenosis. In small arteries, atherosclerotic plaques can gradually occlude vessel lumens, compromising blood flow and causing ischemic injury. At early stages of stenosis, outward remodeling of the vessel media tends to preserve the size of the lumen. However, there are limits on the extent of remodeling, and eventually the expanding atheroma impinges on the lumen to such a degree that blood flow is compromised (see Fig. 11.17). *Critical stenosis* is the stage at which the occlusion is sufficiently severe to cause tissue ischemia. In the coronary (and other) circulations, this typically occurs when the occlusion produces a 70% to 75% decrease in luminal cross-sectional area; with this degree of stenosis, chest pain may develop with exertion (so-called *stable angina*; see Chapter 12). Although acute plaque rupture (see later) is the most dangerous consequence, atherosclerosis also takes a toll through chronically diminished arterial perfusion. Mesenteric occlusion and bowel ischemia, sudden cardiac death, chronic ischemic heart disease, ischemic encephalopathy, and intermittent claudication (diminished perfusion of the extremities) all are consequences of flow-limiting stenoses. The effects of vascular occlusion ultimately depend on arterial supply and the metabolic demand of the affected tissue. If stenosis occurs slowly, smaller adjacent vessels can partially compensate by enlarging and creating collateral circulation to perfuse the organ.

Acute Plaque Change. Plaque erosion or rupture is typically promptly followed by partial or complete vascular thrombosis (see Fig. 11.16), resulting in acute tissue infarction (e.g., myocardial or cerebral infarction) (see Fig. 11.17). Plaque changes fall into three general categories:

- *Rupture/fissuring,* exposing highly thrombogenic plaque constituents that activate coagulation and induce thrombosis that is often completely occlusive
- *Erosion/ulceration,* exposing the thrombogenic subendothelial basement membrane to blood, less-frequently inducing fully occlusive thrombosis
- *Hemorrhage into the atheroma,* expanding its volume

It is now recognized that the plaques that are responsible for myocardial infarction and other acute coronary

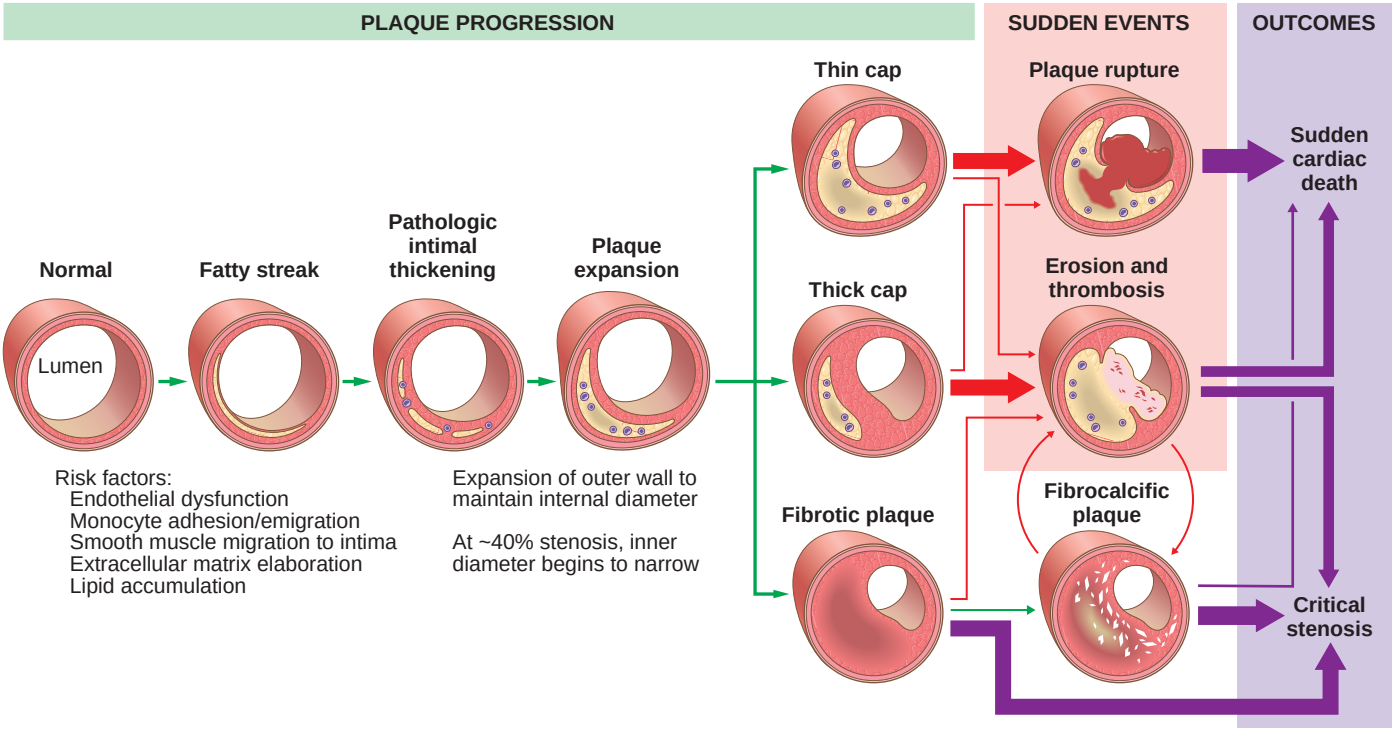


Figure 11.17 Atherosclerotic plaque formation, activities, and outcomes. In a lesion-prone area along with other risk factors, a reversible fatty streak is formed, which can progress to atherosclerosis. Compensatory enlargement initially prevents a reduction of blood flow through the vessel. A plaque can become a thin cap plaque (unstable, vulnerable), a thick cap plaque (stable), or a fibrotic plaque (stable). Thin cap plaques are the most prone to plaque rupture, generally leading to sudden cardiac death. Stable plaques can undergo surface erosion and thrombosis, rapidly expanding the plaque size and leading to more prominent calcifications. This event can lead to sudden cardiac death. Extensive narrowing of the luminal diameter from large plaques generally results in critical stenosis, reducing blood supply to the heart and resulting in angina.

syndromes are often asymptomatic before undergoing a sudden, typically unpredictable change. Thus pathologic and clinical studies show that the majority of plaques that undergo abrupt disruption and coronary occlusion previously showed only mild to moderate noncritical luminal stenosis. Thus a large number of asymptomatic adults may be at risk for a catastrophic coronary event. Although imaging modalities are being developed that could preemptively identify such lesions, it is clear that standard angiographic studies are woefully inadequate to visualize them until after the fact.

Plaques rupture when they are unable to withstand mechanical stresses generated by vascular shear forces. The events that trigger abrupt changes in plaques and subsequent thrombosis are complex and include both intrinsic factors (e.g., plaque structure and composition) and extrinsic elements (e.g., blood pressure, platelet reactivity, and vessel spasm).

The fibrous cap undergoes continuous remodeling that can stabilize the plaque or, conversely, render it more susceptible to rupture. Collagen is the major structural component of the fibrous cap and accounts for its mechanical strength and stability; the balance of collagen synthesis versus degradation affects cap integrity. Thus plaques with thin fibrous caps and active inflammatory cells over a necrotic core are more likely to rupture; these are referred to as "vulnerable plaques" (Fig. 11.18).

Collagen in atherosclerotic plaque is produced primarily by SMCs so that loss of these cells results in a less sturdy cap. Moreover, collagen turnover is controlled by metalloproteinases (MMPs), enzymes elaborated largely by macrophages and SMCs within the atheromatous plaque. Conversely, tissue inhibitors of metalloproteinases (TIMPs) produced by ECs, SMCs, and macrophages modulate MMP activity. In general, plaque inflammation results in a net increase in collagen degradation and reduced collagen synthesis, thereby destabilizing the mechanical integrity of

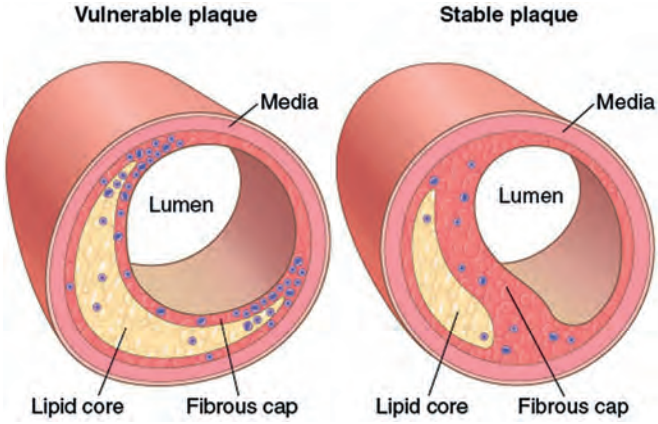


Figure 11.18 Vulnerable and stable atherosclerotic plaque. *Stable plaques* have thickened and densely collagenous fibrous caps with minimal inflammation and underlying atheromatous core. *Vulnerable plaques* have thin fibrous caps, large lipid cores, and increased inflammation. (Modified from Libby P: Molecular bases of the acute coronary syndromes, *Circulation* 91:2844, 1995.)

the fibrous cap (see later). The inflammation induced by cholesterol deposits themselves may contribute to plaque destabilization. Conversely, statins may have a beneficial therapeutic effect not only by reducing circulating cholesterol levels but also by stabilizing plaques through a reduction in plaque inflammation.

Influences extrinsic to plaques also contribute to acute plaque changes. Adrenergic stimulation can increase systemic blood pressure or induce local vasoconstriction, thereby increasing the physical stresses on a given plaque. Indeed, the adrenergic stimulation associated with awakening and rising can cause blood pressure spikes (followed by heightened platelet reactivity) that have been causally linked to the pronounced circadian periodicity for onset of acute myocardial infarction (peaking between 6 a.m. and 12 noon). Intense emotional stress can also contribute to plaque disruption; this is most dramatically illustrated by the uptick in the incidence of sudden death associated with natural or other disasters, such as earthquakes and the September 11, 2001, attack on the World Trade Center.

It is also important to note that not all plaque ruptures result in occlusive thromboses with catastrophic consequences. Indeed, plaque disruption and an ensuing superficial platelet aggregation and thrombosis are probably common, repetitive, and often clinically silent complications of atheroma. Healing of these subclinical plaque disruptions—and organization of their overlying thrombi—is an important mechanism in the growth of atherosclerotic lesions.

Thrombosis, partial or total, associated with a disrupted plaque is a central factor in acute coronary syndromes. In its most serious form, thrombus leads to total occlusion of the affected vessel. In contrast, in other coronary syndromes (Chapter 12), luminal obstruction by the thrombus is incomplete and may even wax and wane with time.

Mural thrombi in a coronary artery can also embolize. Small embolic fragments of thrombus can often be found in the distal intramyocardial circulation or in association with microinfarcts in patients with atherosclerosis who die suddenly. Thrombin and other factors associated with thrombosis are potent activators of SMCs and can thereby contribute to the growth of atherosclerotic lesions.

Vasoconstriction compromises lumen size and, by increasing the local mechanical forces, can potentiate plaque disruption. Vasoconstriction at sites of atheroma may be stimulated by (1) circulating adrenergic agonists, (2) locally released platelet contents, (3) EC dysfunction with impaired secretion of endothelial-derived relaxing factors (NO) relative to contracting factors (endothelin), and (4) mediators released from perivascular inflammatory cells.

KEY CONCEPTS

ATHEROSCLEROSIS

- Atherosclerosis is an intimal-based lesion composed of a fibrous cap and an atheromatous core; the constituents of the plaque include SMCs, ECM, inflammatory cells, calcifications, lipids, and necrotic debris.
- Atherogenesis is driven by an interplay of vessel wall injury and inflammation. The multiple risk factors for atherosclerosis all cause EC dysfunction and influence inflammatory cell and SMC recruitment and stimulation.
- Atherosclerotic plaques develop and generally grow slowly over decades. Stable plaques can produce symptoms related to chronic ischemia by narrowing vessel lumens, whereas unstable plaques can cause dramatic and potentially fatal ischemic complications related to acute plaque rupture, thrombosis, or embolization.
- Stable plaques tend to have a dense fibrous cap, minimal lipid accumulation and little inflammation, whereas "vulnerable" unstable plaques have thin caps, large lipid cores, and relatively dense inflammatory infiltrates.

ANEURYSMS AND DISSECTION

An aneurysm is a localized abnormal dilation of a blood vessel or the heart that may be congenital or acquired (Fig. 11.19). When an aneurysm involves all the layers of an intact (but attenuated) arterial wall or the thinned ventricular wall of the heart, it is called a true aneurysm. Atherosclerotic and congenital vascular aneurysms and ventricular aneurysms that follow transmural myocardial infarctions

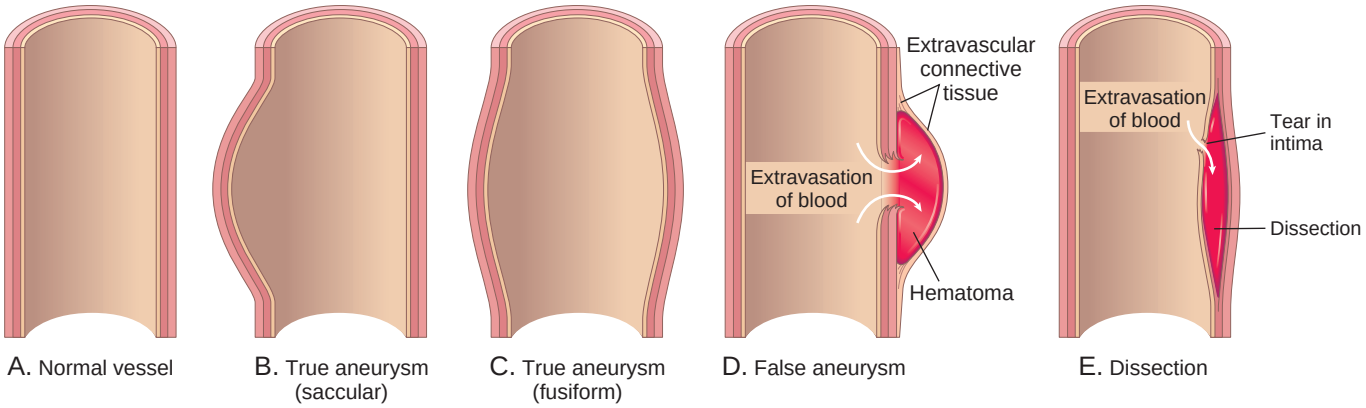


Figure 11.19 Aneurysms. (A) Normal vessel. (B) True aneurysm, saccular type. The wall focally bulges outward and may be attenuated but is otherwise intact. (C) True aneurysm, fusiform type. There is circumferential dilation of the vessel, without rupture. (D) False aneurysm. The wall is ruptured, and there is a collection of blood (hematoma) that is bounded externally by adherent extravascular tissues. (E) Dissection. Blood has entered *(dissected)* the wall of the vessel and separated the layers. Although this is shown as occurring through a tear in the lumen, dissections can also occur by rupture of the vessels of the *vasa vasorum* within the media.

are of this type. In contrast, a false aneurysm (also called a pseudoaneurysm) is a defect in the vascular wall leading to an extravascular hematoma that freely communicates with the intravascular space ("pulsating hematoma"). Examples include a ventricular rupture after myocardial infarction that is contained by a pericardial adhesion or a leak at the sutured junction of a vascular graft with a natural artery. An arterial dissection arises when blood enters a defect in the arterial wall and tunnels through medial or medial-adventitial planes. Dissections are often, but not always, aneurysmal (see later). Both true and false aneurysms as well as dissections can rupture, often with catastrophic consequences.

Descriptively, aneurysms are classified by macroscopic shape and size (see Fig. 11.19). *Saccular aneurysms* are essentially spherical outpouchings (involving only a portion of the vessel wall); in intracranial vessels they generally measure 2 to 20 mm; however, in the aorta they range from 5 to 10 cm in diameter and often contain thrombus. *Fusiform aneurysms* involve diffuse, circumferential dilation of a long vascular segment; they vary in diameter (in the aorta generally from 5 to 10 cm) and can involve extensive portions of the aortic arch, abdominal aorta, or even the iliacs. These types are not specific for any disease or clinical manifestations.

Pathogenesis

To maintain their structural and functional integrity, arterial walls constantly remodel by synthesizing, degrading, and repairing damage to their ECM constituents. Aneurysms can occur when the structure or function of the connective tissue within the vascular wall is compromised. Defects in synthesis or breakdown of connective tissue contribute to the pathobiology of both heritable aneurysmal diseases, as well as the common, sporadic forms of aneurysms. These defects include the following:

- *The intrinsic quality of the vascular wall connective tissue is poor.* Weak vessel walls due to defective type III collagen synthesis are a hallmark of the vascular form of *Ehlers-Danlos syndrome* (Chapter 5).
- *Abnormal transforming growth factor-β (TGF-β) signaling.* Excessive TGF-β activity alters vascular wall remodeling, primarily of the ascending aorta, leading ultimately to diminished ECM content and integrity with aneurysmal dilation. In *Marfan syndrome* (Chapter 5), defective synthesis of the scaffolding protein fibrillin leads to the inability to appropriately sequester endogenously produced TGF-β. In *Loeys-Dietz syndrome,* increased activity in the TGF-β signaling pathway can result from mutations in TGF-β receptors, an intracellular downstream signaling molecule in the TGF-β pathway (SMAD3), and even TGF-β3 itself. Aneurysms in patients with Loeys-Dietz syndrome can rupture at small sizes and are thus considered to follow an "aggressive" course.
- *The balance of collagen degradation and synthesis is altered by inflammation and associated proteases.* Inflammatory cells in the setting of an aortitis or associated with atherosclerosis can be found throughout the aortic wall. Increased MMP production, especially by macrophages, can contribute to aneurysm development by degrading ECM (elastin, collagens, proteoglycans, laminin, fibronectin) in all layers of the wall; decreased expression of TIMPs can also impact ECM degradation. Both processes result in a loss of elastic fibers necessary for recoil in diastole.

Genetic predisposition to aneurysm formation in the setting of inflammatory lesions (e.g., atherosclerosis) may be related to MMP and/or TIMP polymorphisms or to the nature of the local inflammatory response. Indeed, abdominal aortic aneurysms (AAAs) (see later) are associated with shifts in the local cytokine environments toward the production of T helper 2 cytokines (e.g., IL-4 and IL-10) that can drive macrophages to produce increased amounts of elastolytic MMP.

- *The vascular wall is weakened through loss of SMCs or the inappropriate synthesis of noncollagenous or nonelastic ECM.* Ischemia of the inner media occurs when there is atherosclerotic thickening of the intima, increasing the distance that oxygen and nutrients must diffuse. Systemic hypertension can also cause significant narrowing of arterioles of the vasa vasorum (e.g., in the aorta), which causes outer medial ischemia. Tertiary syphilis is another rare cause of aortic aneurysms in which obliterative endarteritis occurs in the vasa vasorum of the thoracic aorta; this engenders medial ischemia that leads to SMC loss, elastic fiber loss, and inadequate or inappropriate ECM synthesis.

All of the above-described processes lead to histopathologic changes in the arterial wall (referred to as medial degeneration) (Fig. 11.20). Although adventitial lymphoplasmacytic

Figure 11.20 Cystic medial degeneration. (A) Cross-section of aortic media from a patient with Marfan syndrome showing marked elastin fragmentation and formation of areas devoid of elastin that resemble cystic spaces but are actually filled with increased amounts of proteoglycans *(asterisks).* (B) Normal media for comparison showing the regular layered pattern of elastic tissue. In both (A) and (B), elastin is stained black.

inflammation can suggest syphilis, and clinical features (e.g., skeletal changes in Marfan syndrome) can suggest a potential genetic underpinning, medial degeneration is a nonspecific finding common to all forms of aortic disease.

The two most important causes of aortic aneurysms are atherosclerosis and hypertension; atherosclerosis is a greater factor in AAAs, while hypertension is the most common etiology associated with ascending aortic aneurysms. Other pathologies and risk factors that weaken vessel walls and lead to aneurysms include advanced age, smoking, trauma, vasculitis (see later), congenital defects (e.g., fibromuscular dysplasia and berry aneurysms typically in the circle of Willis; see Chapter 28), and infections (mycotic aneurysms). Mycotic aneurysms can originate from (1) embolization of a septic embolus, usually as a complication of infective endocarditis; (2) extension of an adjacent suppurative process; or (3) circulating organisms directly infecting the arterial wall.

Abdominal Aortic Aneurysm

Aneurysms occurring as a consequence of atherosclerosis form most commonly in the abdominal aorta and common iliac arteries. A variety of factors discussed earlier collaborate to weaken the media and predispose to aneurysm formation.

AAAs occur more frequently in men and in smokers, rarely developing before age 50. Atherosclerosis is a major cause of AAAs, but other factors clearly contribute, since the incidence is <5% in men older than age 60, despite almost universal abdominal aortic atherosclerosis in that population.

MORPHOLOGY

Usually located between the renal arteries and the bifurcation of the aorta, AAAs can be saccular or fusiform: greater than 3 cm, but often >5.5 cm, in diameter, and up to 25 cm in length (Fig. 11.21). There is severe complicated atherosclerosis with destruction and thinning of the underlying aortic media; the aneurysm frequently contains a bland, poorly organized mural thrombus. AAAs can occasionally affect the renal and superior or inferior mesenteric arteries, either by direct extension or by occluding vessel ostia with mural thrombi. Not infrequently, AAAs will be accompanied by smaller aneurysms of the iliac arteries.

Three AAA variants merit special mention:

- **Inflammatory AAAs** account for 5% to 10% of all AAAs; these typically occur in younger patients who often present with back pain and elevated inflammatory markers (e.g., erythrocyte sedimentation rate). Inflammatory aneurysms are characterized by abundant lymphoplasmacytic inflammation with many macrophages (and even giant cells) associated with dense periaortic scarring that can extend into the anterior retroperitoneum. The cause is a presumed localized immune response to the abdominal aortic wall; remarkably, most cases are not associated with inflammation of other arteries.
- A subset of inflammatory AAAs may be a vascular manifestation of **immunoglobulin G4–related disease (IgG4-RD).** This is a disorder marked by (in most cases) high circulating levels of IgG4, with storiform fibrosis and IgG4-positive infiltrating plasma cells in affected tissues. It may affect a host of organs including the pancreas and thyroid, as well as other vascular sites such as the ascending aorta and heart vessels. There may be retroperitoneal fibrosis with bilateral hydronephrosis. Recognition of this entity is important since it responds well to steroid and anti–B-cell therapies.
- **Mycotic AAAs** are lesions that have become infected by the lodging of circulating microorganisms in the wall. In such cases, suppuration further destroys the media, potentiating rapid dilation and rupture.

Clinical Features

Most cases of AAAs are completely asymptomatic and are discovered incidentally on physical examination as an abdominal mass (often palpably pulsating) that simulates a tumor. The other clinical manifestations of AAAs include:

- *Rupture* into the peritoneal cavity or retroperitoneal tissues with massive, potentially fatal hemorrhage
- *Obstruction* of a branch vessel resulting in downstream tissue ischemic injury; for example, iliac (leg), renal (kidney), mesenteric (gastrointestinal tract), or vertebral arteries (spinal cord)
- *Embolism* from atheroma or mural thrombus
- *Impingement* on an adjacent structure, for example, compression of a ureter or erosion of vertebrae

The risk of rupture is directly related to the size of the aneurysm, varying from nil for AAAs ≤4 cm in diameter to 1% per year for AAAs between 4 and 5 cm, 11% per year for AAAs between 5 and 6 cm, and 25% per year for aneurysms >6 cm. Most aneurysms expand at a rate of 0.2 to 0.3 cm/year, but 20% expand more rapidly. In general, aneurysms ≥5 cm are managed aggressively, either through

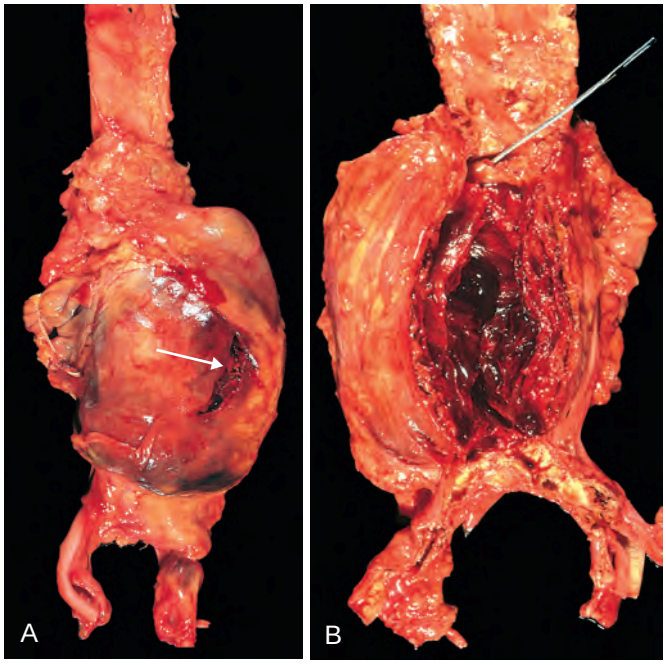


Figure 11.21 Abdominal aortic aneurysm. (A) External view, gross photograph of a large aortic aneurysm that ruptured (rupture site is indicated by the *arrow*). (B) Opened view, with the location of the rupture tract indicated by a probe. The wall of the aneurysm is exceedingly thin, and the lumen is filled by a large quantity of layered but largely unorganized thrombus.

open surgeries involving prosthetic grafts or endovascular approaches through the femoral artery deploying a stent graft (expandable wire frame covered by a cloth sleeve). Timely surgery is critical; operative mortality for unruptured aneurysms is approximately 5%, whereas emergency surgery after rupture carries a mortality rate of more than 50%. It is worth reiterating that because atherosclerosis is a systemic disease, a patient with an AAA is also very likely to have atherosclerosis in other vascular beds and is at a significantly increased risk of ischemic heart disease and stroke.

Thoracic Aortic Aneurysm

Thoracic aortic aneurysms are most commonly associated with hypertension, although other causes such as Marfan syndrome, Loeys-Dietz syndrome, and inflammatory conditions (aortitis) are increasingly recognized. Before dissection or rupture, symptoms can include (1) chest pain from encroachment or erosion into bone, (2) myocardial ischemia from compression of a coronary artery, (3) difficulty swallowing due to compression of the esophagus, (4) hoarseness from irritation of or pressure on the recurrent laryngeal nerves, or (5) respiratory complications from compression of the bronchi; the vast majority cause no symptoms at all until a catastrophic event occurs (see later).

Aortic Dissection

Aortic dissection occurs when blood separates the laminar planes of the media to form a blood-filled channel within the aortic wall (Fig. 11.22). Aortic dissection is generally associated with aortic dilation; it can be disastrous if the dissection then ruptures through the adventitia and hemorrhages into adjacent spaces.

Aortic dissection occurs principally in two groups of patients: (1) men aged 40 to 60 years, with antecedent hypertension (more than 90% of cases) and (2) younger patients with syndromic diseases affecting the aorta (e.g., Marfan syndrome). Dissections can be iatrogenic (e.g., following arterial cannulations during diagnostic catheterization or cardiopulmonary bypass). Pregnancy is also rarely associated with aortic (or other vessel) dissection (roughly 10 to 20 cases per 1 million births), typically occurring during or after the third trimester. It may be related to hormonally induced vascular remodeling and the hemodynamic stresses of the perinatal period. Dissection is unusual in the presence of substantial atherosclerosis or other cause of medial scarring, such as syphilis, presumably because the medial fibrosis inhibits propagation of the dissecting hematoma.

Pathogenesis

Hypertension is the major risk factor for aortic dissection. Aortas of hypertensive patients have medial degenerative changes with SMC loss and altered ECM content. Other dissections occur in the setting of inherited or acquired connective tissue disorders with defective TGF-β signaling or defective ECM synthesis or degradation. However, specific patterns of medial damage are neither a prerequisite for dissection nor a reliable predictor of its occurrence. Regardless of the underlying etiology, the ultimate trigger for the intimal tear and initial intramural aortic hemorrhage is not known in most cases. Nevertheless, once the tear has occurred, blood flow under systemic pressure dissects through the media leading to progression of the hematoma. Accordingly, aggressive pressure-reducing therapy may be effective in limiting an evolving dissection. In some cases, disruption of penetrating vessels of the vasa vasorum can give rise to an intramural hematoma without an intimal tear.

MORPHOLOGY

In most cases of dissection, medial degeneration is present at the site of tear but may not be seen in the area of propagation. These typically feature elastic fiber fragmentation and loss, mucoid ECM accumulation, and SMC attrition (see Fig. 11.20); inflammation is characteristically absent. Dissections can occur in the setting of rather trivial medial degeneration; conversely, marked degenerative changes are frequently seen at autopsies of patients who are completely free from dissection. Thus the relationship of the structural changes to the pathogenesis of dissection is uncertain.

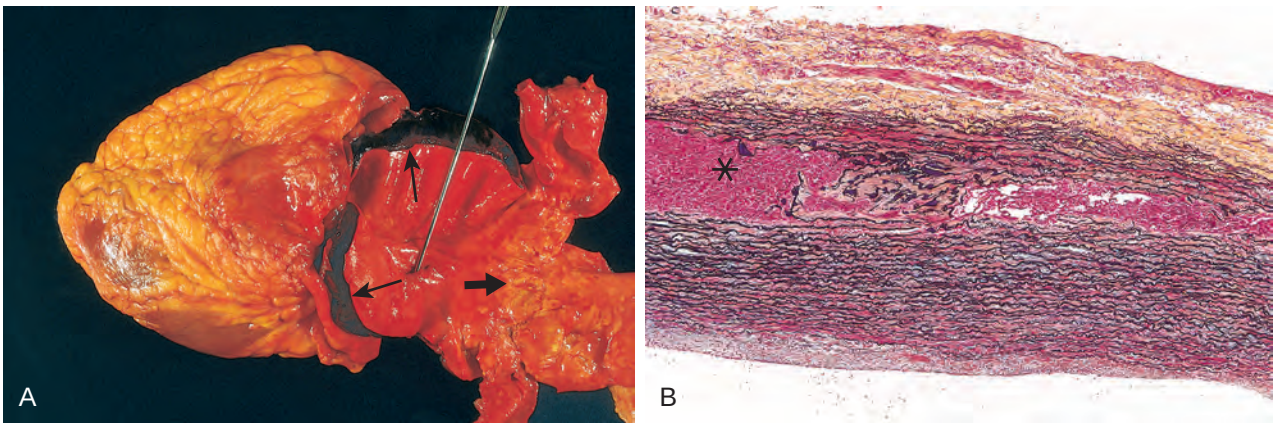


Figure 11.22 Aortic dissection. (A) Gross photograph of an opened aorta with proximal dissection originating from a small, oblique intimal tear *(probe)*, allowing blood to enter the media and creating a retrograde intramural hematoma *(thin arrows)*. Note that the intimal tear has occurred in a region largely free of atherosclerotic plaque and that propagation of the intramural hematoma distally is arrested where atherosclerosis begins *(thick arrow)*. (B) Histologic view of dissection demonstrating an aortic intramural hematoma *(asterisk)*. Aortic elastic layers are black and blood is red (Movat stain).

An aortic dissection usually initiates with an intimal tear. In the vast majority of spontaneous dissections, the tear occurs in the ascending aorta, usually within 10 cm of the aortic valve (see Fig. 11.22A). Such tears are typically transverse with sharp, jagged edges up to 1 to 5 cm in length. The dissection can extend retrograde toward the heart as well as distally, sometimes as far as the iliac and femoral arteries. The dissecting hematoma spreads characteristically between lamellar units of the outer third of the media or between media and adventitial layers (see Fig. 11.22B). It can rupture through the adventitia causing massive hemorrhage (e.g., into the thoracic or abdominal cavities) or cardiac tamponade (hemorrhage into the pericardial sac). In some (fortuitous) instances, the dissecting hematoma reenters the lumen of the aorta through a second distal intimal tear, creating a new false vascular channel ("double-barreled aorta"). This averts a fatal extraaortic hemorrhage, and over time, such false channels can be endothelialized to become recognizable **chronic dissections.**

Traumatic chest injury (e.g., motor vehicle accident, or even extremely vigorous cardiopulmonary resuscitation) can cause intimal tears that originate after the origin of arch vessels at the *ligamentum arteriosum.* In such cases the connective tissue that tethers the aorta and pulmonary vessels in that location can lead to aortic tearing when there is sudden extreme movement of the heart toward the anterior chest wall.

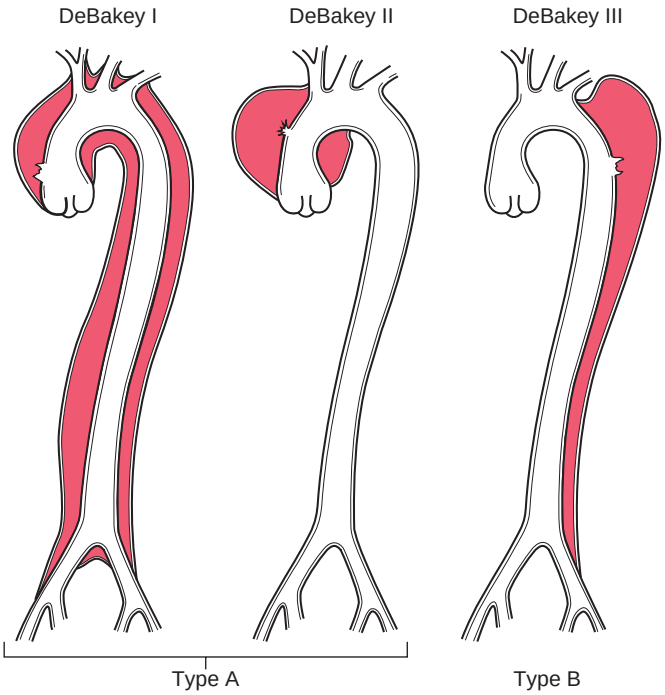


Figure 11.23 Classification of dissections. *Type A* (proximal) involves the ascending aorta, either as part of a more extensive dissection (DeBakey I) or in isolation (Debakey II). *Type B* (distal or Debakey III) dissections arise after the take-off of the great vessels. Type A dissections typically have the most serious complications and greatest mortality.

Clinical Features

The morbidity and mortality associated with dissections depends on which part of the aorta is involved; the most serious complications occur with dissections between the aortic valve and the distal arch. Accordingly, aortic dissections are generally classified into two types (Fig. 11.23):

- *Type A dissections.* These are the more common (and dangerous) *proximal* lesions involving either both the ascending and the descending aorta or the ascending aorta only (types I and II of the DeBakey classification).
- *Type B dissections.* Distal lesions not involving the ascending aorta and usually beginning distal to the subclavian artery (called DeBakey type III).

The classic clinical symptoms of aortic dissection are the sudden onset of excruciating pain, usually beginning in the anterior chest, radiating to the back between the scapulae, and moving downward as the dissection progresses; the pain can be confused with that of myocardial infarction.

The most common cause of death is rupture of the dissection and bleeding into the pericardial, pleural, or peritoneal cavities. Retrograde dissection into the aortic root can also disrupt the aortic valve annulus. Common clinical manifestations include cardiac tamponade and aortic insufficiency. Dissections can also extend into the great arteries of the neck or into the coronary, renal, mesenteric, or iliac arteries, causing vascular obstruction and ischemic consequences such as myocardial infarction; involvement of spinal arteries can cause transverse myelitis.

In type A dissections, rapid diagnosis and institution of intensive antihypertensive therapy coupled with surgical plication of the aortic tear can save 65% to 85% of patients. However, mortality approaches 70% in those who present with hemorrhage or symptoms related to distal ischemia, and the overall 10-year survival is 40% to 88% depending on whether a patient is a surgical candidate and other prognostic factors. Most type B dissections can be managed conservatively; such patients have a short-term 75% survival rate whether they are treated with surgery or antihypertensive medication only.

KEY CONCEPTS

ANEURYSMS AND DISSECTIONS

- Vascular aneurysms are congenital or acquired dilations of the blood vessels that involve the entire thickness of the wall. Complications are related to rupture, thrombosis, and embolization.
- Dissections occur when blood enters the wall of a vessel and dissects through the layers generally between lamellar units. Complications arise due to rupture or obstruction of vessels branching off the aorta.
- Aneurysms and dissections result from structural weakness of the vessel wall caused by aberrant TGF-β signaling, loss of SMCs, and changes in the ECM, which can result from ischemia, genetic defects, or defective matrix remodeling.

VASCULITIS

***Vasculitis* is a general term for vessel wall inflammation with protean manifestations depending on the vascular bed affected** (e.g., central nervous system vs. heart vs. small bowel). Besides the findings referable to the specific tissues involved, the clinical manifestations typically include constitutional signs and symptoms associated with systemic inflammation, such as fever, myalgias, arthralgias, and malaise.

Table 11.4 Primary Forms of Vasculitis

	Giant Cell Arteritis	Granulomatosis With Polyangiitis	Churg-Strauss Syndrome	Polyarteritis Nodosa	Leukocytoclastic Vasculitis	Buerger Disease	Behçet Disease
Sites of Involvement							
Aorta	+	–	–	–	–	–	–
Medium-sized arteries	+	+	+	+	–	+	+
Small-sized arteries	–	+	+	+	+	+	+
Capillaries	–	–	–	–	+	–	+
Veins	–	–	–	–	+	+	+
Inflammatory Cells Present							
Lymphocytes	+	+	+	±	±	±	±
Macrophages	+	+	+	±	±	±	±
Neutrophils	Rare	+	+	±	±	±	Required
Eosinophils	Very rare	±	Required	±	±	±	±
Other Features							
Granulomas	+[a]	Required[a]	±	–	–	–	–
Giant cells	Often; not required	±	–	–	–	–	–
Thrombosis	±	±	±	±	±	Required	±
Serum ANCA positivity	–	+	+	±	–	–	–
Clinical history	>40 years old, ± polymyalgia rheumatica	Any	Asthma, atopy	Any	Any	Young male smoker	Orogenital ulcers

[a]The granulomas of giant cell arteritis are found within the vessel wall as part of the inflammation comprising the vasculitis but need not be present to render the diagnosis. The granulomas of granulomatosis with polyangiitis are larger, spanning between vessels, and associated with areas of tissue necrosis.

ANCA, Antineutrophil cytoplasmic antibody.

From Seidman MA, Mitchell RN: Surgical pathology of small- and medium-sized vessels, *Surg Pathol Clin* 5:435, 2012.

Vessels of any type in virtually any organ can be affected; most vasculitides affect small vessels ranging from arterioles to capillaries to venules. Nevertheless, several vasculitides tend to affect only vessels of a particular size or location. Thus there are entities that primarily affect the aorta and medium-sized arteries, while others principally affect only smaller arterioles. Some 20 primary forms of vasculitis are recognized, and classification schemes attempt (with variable success) to group them according to vessel diameter, role of immune complexes, presence of specific autoantibodies, granuloma formation, organ specificity, and even population demographics (Table 11.4 and Fig. 11.24). As we will see, there is considerable clinical and pathologic overlap among many of them.

The two pathogenic mechanisms of vasculitis are immune-mediated inflammation and direct invasion of vascular walls by infectious pathogens. Infections can also indirectly induce a noninfectious vasculitis, for example, by generating immune complexes or triggering vascular cross-reactivity. In any given patient, it is critical to distinguish between infectious and immunologic mechanisms because immunosuppressive therapy is appropriate for immune-mediated vasculitis but could very well be counterproductive for infectious vasculitides. Physical and chemical injury (e.g., from irradiation, mechanical trauma, and toxins) can also cause vasculitis.

Noninfectious Vasculitis

The main immunologic mechanisms underlying noninfectious vasculitis are:

- *Immune complex deposition*
- *Antineutrophil cytoplasmic antibodies (ANCAs)*
- *Anti-EC antibodies*
- *Autoreactive T cells*

Immune Complex–Associated Vasculitis

This form of vasculitis can be seen in immunologic disorders such as systemic lupus erythematosus (SLE) (Chapter 6) that are associated with autoantibody production. The vascular lesions resemble those found in experimental immune complex–mediated disorders such as the Arthus phenomenon and serum sickness and in many cases contain readily identifiable antibody and complement. Often, however, this type of vasculitis presents a number of diagnostic challenges. Only rarely is the specific antigen responsible for immune complex formation identified. While immune complexes can occasionally be detected in the blood, in most cases it is not clear whether the pathogenic antigen-antibody complexes are deposited from the circulation or form in situ. Indeed, the sensitivity and specificity of circulating immune complex assays in such diseases are extremely low. In many suspected cases, even the antigen-antibody

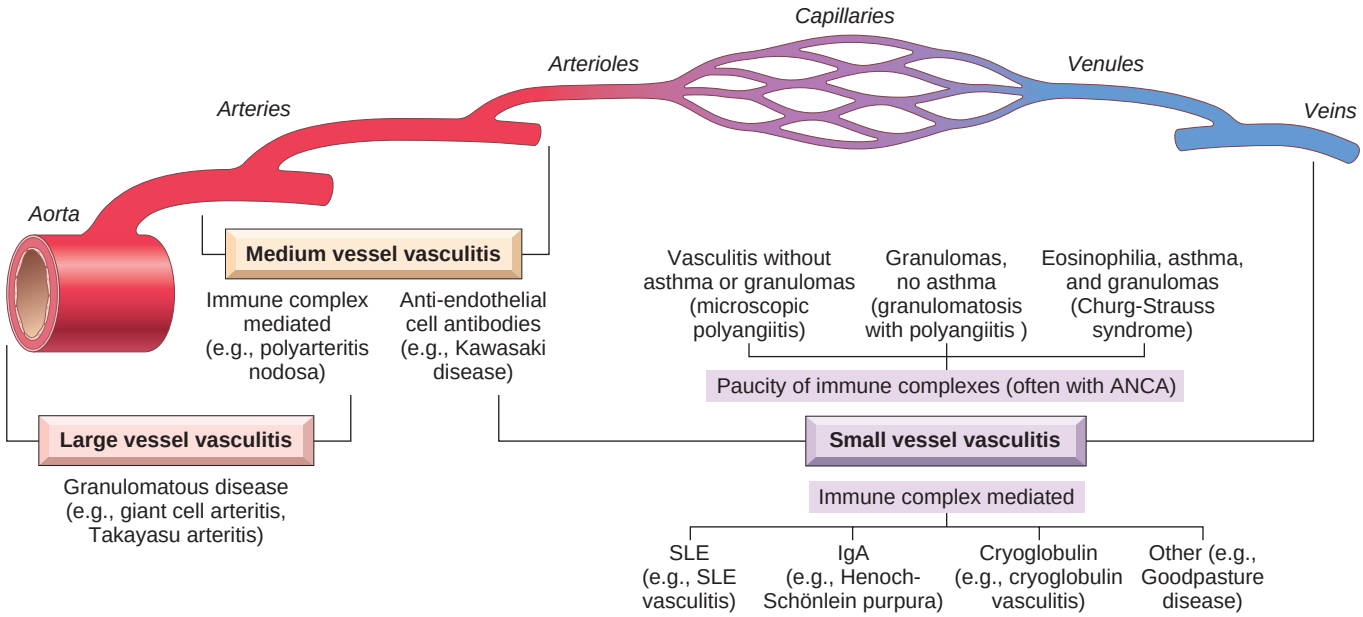


Figure 11.24 Vascular sites typically involved with the more common forms of vasculitis as well as their presumptive etiologies. Note that there is a substantial overlap in distributions. *ANCA*, Antineutrophil cytoplasmic antibody; *SLE*, systemic lupus erythematosus.

deposits are scarce. In such instances the immune complexes have putatively been degraded by the time of biopsy; alternatively, other underlying mechanisms need to be considered for such "pauci-immune" vasculitides.

Immune complex deposition is implicated in the following vasculitides:

- *Drug hypersensitivity vasculitis*. In some cases (e.g., penicillin), drugs act as haptens by binding to serum proteins; other agents are themselves foreign proteins (e.g., streptokinase). Regardless, antibodies directed against the drug-modified proteins or foreign molecules result in immune complex formation. The clinical manifestations can be mild and self-limiting or severe and even fatal; skin lesions are most common. It is always important to consider drug hypersensitivity as a cause of vasculitis since discontinuation of the offending agent usually leads to resolution.
- *Vasculitis secondary to infections*. Antibodies to microbial constituents can form immune complexes that circulate and deposit in vascular lesions. In up to 30% of patients with polyarteritis nodosa (see later), the vasculitis is attributable to immune complexes composed of hepatitis B surface antigen (HBsAg) and anti-HBsAg antibody.

Antineutrophil Cytoplasmic Antibodies

Many patients with vasculitis have circulating antibodies that react with neutrophil cytoplasmic antigens—*ANCAs*. ANCAs are a heterogeneous group of autoantibodies directed against constituents (mainly enzymes) of neutrophil primary granules, monocyte lysosomes, and ECs. ANCAs are very useful diagnostic markers; their titers generally mirror clinical severity, and a rise in titers after periods of quiescence is predictive of disease recurrence. Although a number of ANCAs have been described, two are most important. These were previously grouped according to the intracellular distribution of the target antigens (cytoplasmic [c-ANCA] or perinuclear [p-ANCA]), but are now classified according to their antigen specificity.

- *Anti-proteinase-3 (PR3-ANCA, previously c-ANCA)*. Proteinase-3 (PR3) is a neutrophil azurophilic granule constituent that shares homology with numerous microbial peptides, possibly explaining the generation of PR3-ANCAs. PR3-ANCAs are associated with granulomatosis with polyangiitis (see later).
- *Anti-myeloperoxidase (MPO-ANCA, previously p-ANCA)*. Myeloperoxidase (MPO) is a lysosomal granule constituent involved in oxygen free radical generation (Chapter 3). MPO-ANCAs are induced by several therapeutic agents, particularly propylthiouracil (used to treat hyperthyroidism). MPO-ANCAs are associated with microscopic polyangiitis and Churg-Strauss syndrome (see later).

The close association between ANCA titers and disease activity also suggests a pathogenic role for these antibodies. Of note, ANCAs can directly activate neutrophils, stimulating the release of reactive oxygen species and proteolytic enzymes; in vascular beds, such activation also leads to destructive interactions between inflammatory cells and ECs. While the antigenic targets of ANCA are primarily intracellular (and therefore not usually accessible to circulating antibodies), it is now clear that ANCA antigens (especially PR3) are either constitutively expressed at low levels on the plasma membrane or are translocated to the cell surface in activated and apoptotic leukocytes.

A plausible mechanism for ANCA vasculitis is the following.

- Drugs or cross-reactive microbial antigens induce ANCA formation; alternatively, leukocyte surface expression or release of PR3 and MPO (in the setting of infections) incites ANCA development in a susceptible host.
- Subsequent infection, endotoxin exposure, or inflammatory stimulus elicits cytokines such as tumor necrosis

factor [TNF] that upregulate the surface expression of PR3 and MPO on neutrophils and other cell types.
- ANCAs react with these cytokine-activated cells, causing either direct injury (e.g., to ECs) or further activation (e.g., of neutrophils).
- ANCA-activated neutrophils cause tissue injury by releasing granule contents and reactive oxygen species.

Since the ANCA autoantibodies are directed against cellular constituents and do not form circulating immune complexes, the vascular lesions do not typically contain demonstrable antibody and complement. Thus ANCA-associated vasculitides are often described as pauci-immune.

Anti–Endothelial Cell Antibodies

Antibodies to ECs, perhaps induced by defects in immune regulation, may predispose to certain vasculitides, for example, Kawasaki disease (see later).

We will now briefly present several of the best-characterized and generally recognized vasculitides, reemphasizing that there is substantial overlap among the different entities. Moreover, it should be kept in mind that some patients with vasculitis do not have a classic constellation of findings that allows them to be neatly pigeon-holed into one specific diagnosis.

Giant Cell (Temporal) Arteritis

Giant cell (temporal) arteritis is a chronic, classically granulomatous inflammation of large- to small-sized arteries that principally affects arteries in the head. It is the most common form of vasculitis among elderly adults in the United States and Europe. The temporal arteries are not particularly vulnerable, but their name is attached to the disorder since they are the most readily biopsied in making the diagnosis. Vertebral and ophthalmic arteries, as well as the aorta *(giant cell aortitis)*, also can be involved. Because ophthalmic artery involvement can lead to sudden and permanent blindness, affected persons must be diagnosed and treated promptly.

Pathogenesis

Giant cell arteritis likely occurs as a result of a T cell–mediated immune response to an as yet uncharacterized vessel wall antigen. Proinflammatory cytokines (especially TNF) and anti-EC antibodies also contribute. The characteristic granulomatous inflammation, an association with certain major histocompatibility complex (MHC) class II haplotypes, and the excellent therapeutic response to steroids all strongly support an immune etiology. The predilection for vessels of the head remains unexplained, although one hypothesis is that vessels in various parts of the body develop from distinct anlagen and may therefore express unique antigens.

MORPHOLOGY

Involved arterial segments develop **intimal thickening** (with occasional thromboses) **that reduces the luminal diameter.** Classic lesions exhibit medial **granulomatous inflammation** centered on the internal elastic membrane with **elastic lamina fragmentation;** there is an infiltrate of T cells (CD4+ > CD8+) and macrophages. Although multinucleated giant cells are seen in approximately 75% of adequately biopsied specimens (Fig. 11.25), granulomas and giant cells can be rare or absent. Inflammatory

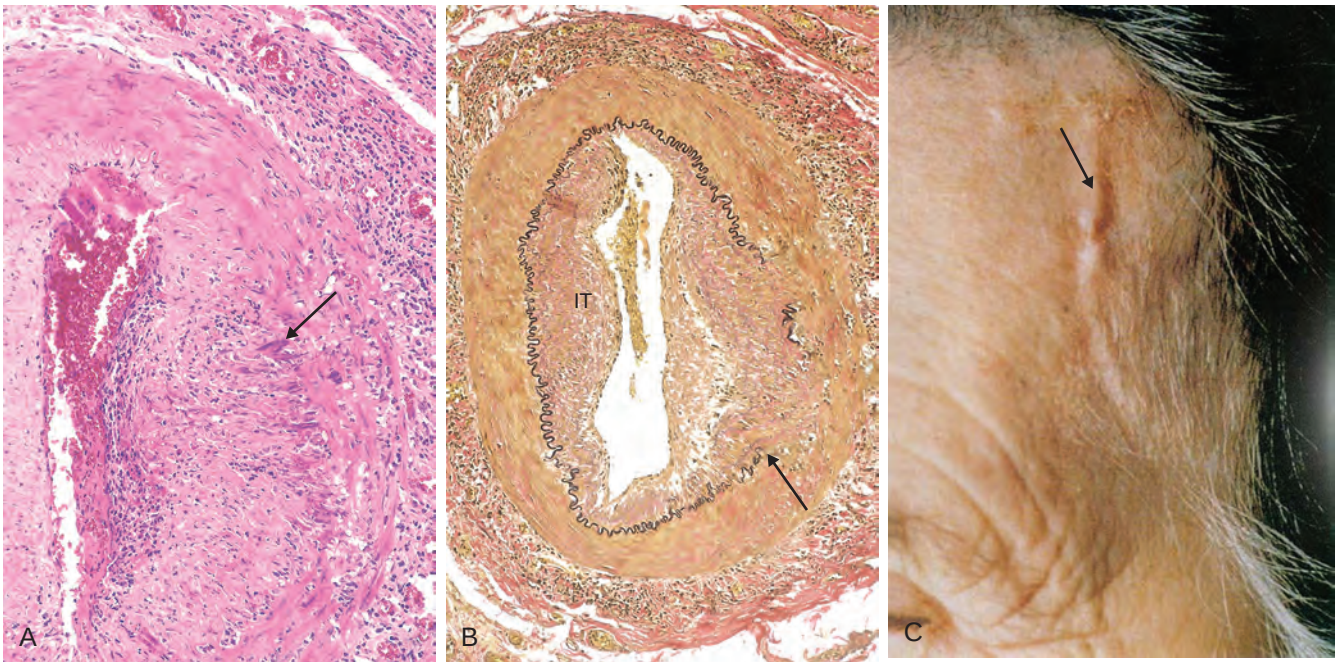


Figure 11.25 Giant cell (temporal) arteritis. (A) Hematoxylin-and-eosin stain of section of temporal artery showing giant cells at the degenerated internal elastic membrane in active arteritis *(arrow)*. (B) Elastic tissue stain demonstrating focal destruction of internal elastic membrane *(arrow)* and intimal thickening *(IT)* characteristic of long-standing or healed arteritis. (C) The temporal artery of a patient with classic giant cell arteritis shows a thickened, nodular, and tender segment of a vessel on the surface of head *(arrow)*. (C, From Salvarani C, et al. Polymyalgia rheumatica and giant-cell arteritis, *N Engl J Med* 347:261, 2002.)

lesions are only focally distributed along the vessel, and segments of relatively normal artery may be interposed. The healed stage is marked by medial attenuation and scarring with intimal thickening, fragmentation of greater than 30% of the circumference of the internal elastic lamina, and adventitial fibrosis.

Clinical Features

Temporal arteritis is rare before age 50. Symptoms may be only vague and constitutional—fever, fatigue, weight loss—or may involve facial pain or headache, most intense along the course of the superficial temporal artery, which can be painful to palpation. Ocular symptoms (associated with involvement of the ophthalmic artery) abruptly appear in about 50% of patients; these range from diplopia to complete vision loss. Diagnosis depends on biopsy and histologic confirmation. However, because giant cell arteritis can be extremely focal within an artery, adequate biopsy requires at least a 1-cm segment; even then, a negative biopsy result does not exclude the diagnosis. Corticosteroids or anti-TNF therapies are typically effective.

Takayasu Arteritis

Takayasu arteritis is a granulomatous vasculitis of medium and larger arteries characterized principally by ocular disturbances and marked weakening of the pulses in the upper extremities (hence the name *pulseless disease*). Takayasu arteritis manifests with transmural fibrous thickening of the aorta—particularly the aortic arch and great vessels—with severe luminal narrowing of the major branch vessels (Fig. 11.26). Takayasu aortitis shares many attributes with giant cell aortitis, including clinical features and histology; indeed, the distinction is typically made based on the age of the patient. Those over 50 years of age are designated as giant cell aortitis, while those under 50 are designated as Takayasu aortitis. Although traditionally associated with the Japanese population and a subset of human leukocyte antigen haplotypes, Takayasu aortitis has a global distribution. An autoimmune etiology is likely.

MORPHOLOGY

Takayasu arteritis classically involves the aortic arch. In a third of patients, it also affects the remainder of the aorta and its branches, with **pulmonary artery** involvement in half of cases; **coronary and renal arteries** may be similarly affected. There is irregular thickening of the vessel wall with intimal hyperplasia; when the aortic arch is involved, the great vessel lumens can be markedly narrowed or even obliterated (Fig. 11.26A and B). Such narrowing explains the weakness of the peripheral pulses. Histologically the changes range from adventitial mononuclear infiltrates with perivascular cuffing of the vasa vasorum to intense mononuclear inflammation in the media. Granulomatous inflammation, replete with giant cells and patchy medial necrosis, is also seen. The histology (Fig. 11.26C) is essentially indistinguishable from giant cell (temporal) arteritis. As the disease progresses, collagenous scarring, with admixed chronic inflammatory infiltrates, occurs in all three layers of the vessel wall. Occasionally, aortic root involvement causes dilation and aortic valve insufficiency.

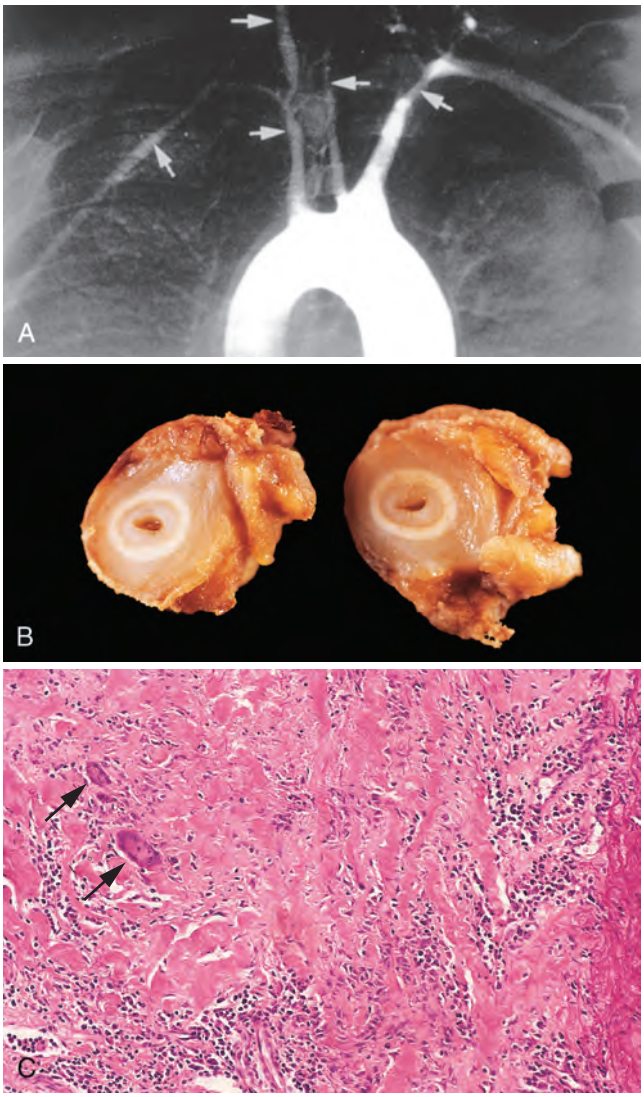


Figure 11.26 Takayasu arteritis. (A) Aortic arch angiogram showing narrowing of brachiocephalic, carotid, and subclavian arteries *(arrows)*. (B) Gross photograph of two cross-sections of the right carotid artery taken at autopsy of the patient shown in (A) demonstrating marked intimal thickening and adventitial fibrosis with minimal residual lumen. (C) Histologic appearance in active Takayasu aortitis illustrating destruction and fibrosis of the arterial media associated with mononuclear infiltrates and inflammatory giant cells *(arrows)*.

Clinical Features

Initial symptoms are usually nonspecific, including fatigue, weight loss, and fever. With progression, vascular symptoms appear and dominate the clinical picture, including reduced blood pressure and weaker pulses in the upper extremities; ocular disturbances, including visual defects, retinal hemorrhages, and total blindness; and neurologic deficits. Involvement of the more distal aorta may lead to claudication of the legs; pulmonary artery involvement can cause pulmonary hypertension. Narrowing of the coronary ostia may lead to myocardial infarction, and involvement of the renal arteries leads to systemic hypertension in roughly half of patients. The course of the disease is variable. In some patients there is rapid progression, while others enter a quiescent stage at 1 to 2 years, permitting long-term survival, albeit with visual or neurologic deficits.

Polyarteritis Nodosa (PAN)

PAN is a systemic vasculitis of small- or medium-sized muscular arteries that typically affects renal and visceral vessels but spares the pulmonary circulation. There is no association with ANCAs, but a third of patients with PAN have chronic hepatitis B, which leads to the formation of HBsAg-HbsAb complexes that deposit in affected vessels. The cause remains unknown in the majority of cases. Clinical manifestations result from ischemia and infarction of affected tissues and organs.

MORPHOLOGY

PAN is associated with segmental transmural necrotizing inflammation of small- to medium-sized arteries, often with superimposed aneurysms and/or thrombosis. Kidney, heart, liver, and gastrointestinal tract vessels are involved in descending order of frequency. Lesions usually involve only part of the vessel circumference with a predilection for branch points. Impaired perfusion of parenchyma distal to the lesions results in ulcerations, infarcts, ischemic atrophy, or hemorrhage.

During the acute phase, there is transmural inflammation of the arterial wall with a mixed infiltrate of neutrophils, eosinophils, and mononuclear cells, frequently accompanied by **fibrinoid necrosis and luminal thrombosis** (Fig. 11.27). Older lesions show fibrous thickening of the vessel wall extending into the adventitia. Characteristically, all stages of activity (from early to late) coexist in different vessels or even within the same vessel, suggesting ongoing injury.

Clinical Features

PAN is primarily a disease of young adults but can occur in all age groups. The course is frequently remitting and episodic, with long symptom-free intervals. Because the vascular involvement is widely scattered, the clinical signs and symptoms of PAN can be quite variable. Typical presentation involves some combination of rapidly accelerating hypertension due to renal artery involvement; abdominal pain and bloody stools caused by vascular gastrointestinal lesions; diffuse myalgias; and peripheral neuritis, predominantly affecting motor nerves. Renal involvement is often a major cause of mortality. Untreated, PAN is typically fatal; however, immunosuppression can yield remissions or cure in 90% of cases.

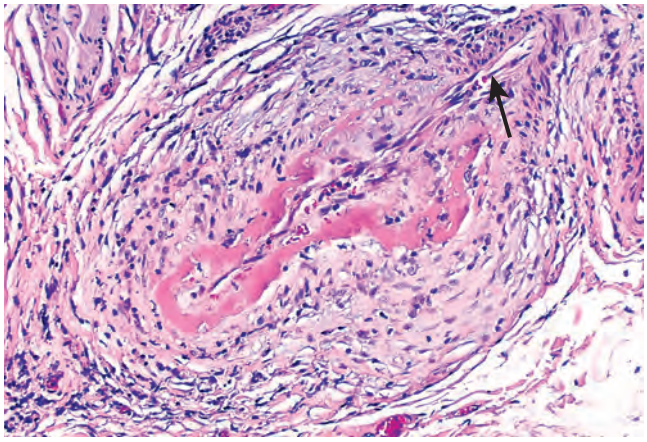


Figure 11.27 Polyarteritis nodosa. There is segmental fibrinoid necrosis and thrombotic occlusion of the lumen of this small artery. Note that part of the vessel wall at the upper right *(arrow)* is uninvolved. (Courtesy Sidney Murphree, MD, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Kawasaki Disease

Kawasaki disease is an acute, febrile, usually self-limited illness of infancy and childhood associated with large- to medium-sized vessel arteritis; 80% of the patients are younger than 4 years of age. Its clinical significance stems from the involvement of coronary arteries. Coronary arteritis can result in aneurysms that rupture or thrombose, causing myocardial infarction. Originally described in Japan, the disease is now recognized in the United States and elsewhere.

In genetically susceptible persons, a variety of infectious agents (mostly viral) have been posited to trigger the disease. The vasculitis may result from a delayed-type hypersensitivity response directed against cross-reactive or newly uncovered vascular antigens. Subsequent cytokine production and polyclonal B-cell activation result in autoantibodies to ECs and SMCs that precipitate the vasculitis.

MORPHOLOGY

Kawasaki disease vasculitis resembles that seen in PAN. There is a dense transmural inflammatory infiltrate, although the fibrinoid necrosis is usually less prominent than in PAN. The acute vasculitis typically subsides spontaneously or in response to treatment, but aneurysm formation due to wall damage can appear. As with other arteritides, healed lesions can also exhibit obstructive intimal thickening. Pathologic changes outside the cardiovascular system are rarely significant.

Clinical Features

Kawasaki disease typically presents with conjunctival and oral erythema and blistering, edema of the hands and feet, erythema of the palms and soles, a desquamative rash, and cervical lymph node enlargement (hence its other name, *mucocutaneous lymph node syndrome*). Approximately 20% of untreated patients develop cardiovascular sequelae, ranging from asymptomatic coronary arteritis, to coronary artery ectasia, to giant coronary artery aneurysms (7 to 8 mm); the last-mentioned are associated with rupture, thrombosis, myocardial infarction, or sudden death. With intravenous immunoglobulin therapy and aspirin, the rate of symptomatic coronary artery disease is less than 4%.

Microscopic Polyangiitis

Microscopic polyangiitis is a necrotizing vasculitis that generally affects capillaries as well as small arterioles and venules. Unlike PAN, all lesions of microscopic polyangiitis tend to be of the same age in any given patient, suggesting a single episode of antibody or immune complex deposition. It is also called hypersensitivity vasculitis or *leukocytoclastic vasculitis.* The skin, mucous membranes, lungs, brain, heart, gastrointestinal tract, kidneys, and muscle all can be

involved; necrotizing glomerulonephritis (90% of patients) and pulmonary capillaritis are particularly common. Microscopic angiitis can be a feature of a number of immune disorders (e.g., Henoch-Schönlein purpura, essential mixed cryoglobulinemia, and vasculitis associated with connective tissue disorders).

Pathogenesis

In some cases, antibody responses to antigens such as drugs (e.g., penicillin), microorganisms (e.g., streptococci), heterologous proteins, or tumor proteins have been implicated. These can either lead to immune complex deposition or trigger secondary immune responses (e.g., development of ANCAs) that are pathogenic. Indeed, most cases are associated with MPO-ANCA. Recruitment and activation of neutrophils within affected vascular beds is likely responsible for the disease manifestations.

MORPHOLOGY

Microscopic polyangiitis is characterized by **segmental fibrinoid necrosis of the media** with focal transmural necrotizing lesions; granulomatous inflammation is absent. These lesions morphologically resemble PAN but typically spare medium-sized and larger arteries; consequently, macroscopic infarcts are uncommon. In some areas (typically postcapillary venules), only infiltrating and fragmenting neutrophils are seen, giving rise to the term **leukocytoclastic vasculitis** (Fig. 11.28A). Although immunoglobulins and complement components can be demonstrated in early skin lesions, little or no immunoglobulin can be seen in most lesions (so-called pauci-immune injury).

Clinical Features

Depending on the vascular bed involved, major clinical features include hemoptysis, hematuria and proteinuria, bowel pain or bleeding, muscle pain or weakness, and palpable cutaneous purpura. With the exception of those who develop widespread renal or brain involvement, immunosuppression induces remission and markedly improves long-term survival.

Granulomatosis With Polyangiitis (GPA)

Previously called Wegener granulomatosis, GPA is a necrotizing vasculitis characterized by a triad of:

- *Acute necrotizing granulomas* of the upper respiratory tract (ear, nose, sinuses, throat) or the lower respiratory tract (lung) or both
- *Necrotizing or granulomatous vasculitis* affecting small- to medium-sized vessels (e.g., capillaries, venules, arterioles, and arteries), most prominent in the lungs and upper airways but affecting other sites as well
- *Focal necrotizing, often crescentic, glomerulonephritis*

"Limited" forms of GPA may be restricted to the respiratory tract. Conversely, a widespread form of the disease can affect eyes, skin, and other organs, notably the heart; clinically, widespread GPA resembles PAN except that there is also respiratory involvement.

Pathogenesis

GPA likely represents a form of T cell–mediated hypersensitivity response to normally "innocuous" inhaled microbial or other environmental agents. Such a pathogenesis is supported by the presence of granulomas and a dramatic response to immunosuppressive therapy. PR3-ANCAs are also present in up to 95% of cases; they are a useful marker of disease activity and may participate in disease pathogenesis. Following immunosuppression treatment, PR3-ANCA titers fall dramatically; subsequent rising titers are predictive of relapse.

MORPHOLOGY

Upper respiratory tract lesions range from inflammatory sinusitis with mucosal granulomas to ulcerative lesions of the nose, palate, or pharynx, rimmed by **granulomas with geographic patterns of central necrosis and accompanying vasculitis** (Fig. 11.28B). The necrotizing granulomas are surrounded by a zone of proliferating fibroblasts with giant cells and leukocyte infiltrate, reminiscent

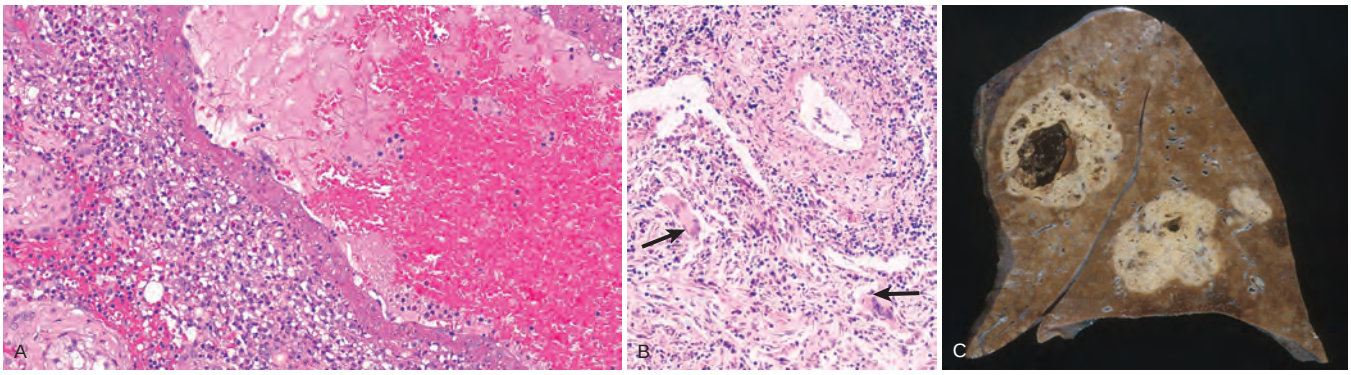


Figure 11.28 Antineutrophil cytoplasmic antibody–associated small vessel vasculitis. (A) Leukocytoclastic vasculitis (microscopic polyangiitis) with fragmentation of neutrophils in and around blood vessel walls. (B and C) Granulomatosis with polyangiitis. (B) Vasculitis of a small artery with adjacent granulomatous inflammation including epithelioid cells and giant cells *(arrows)*. (C) Gross photo from the lung of a patient with fatal granulomatosis with polyangiitis demonstrating large, nodular, centrally cavitating lesions. (A, Courtesy Michael A. Seidman, MD, PhD, Laboratory Medicine Program, University Health Network, Toronto, Canada; C, courtesy Sidney Murphree, MD, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

of mycobacterial or fungal infections. Multiple granulomata can coalesce to produce radiographically visible nodules that can also cavitate; late-stage disease may be marked by extensive necrotizing granulomatous involvement of the parenchyma (Fig. 11.28C), and alveolar hemorrhage may be prominent. Lesions may ultimately undergo progressive fibrosis and organization.

Renal lesions range from mild, focal glomerular necrosis with isolated capillary loop thrombosis **(focal and segmental necrotizing glomerulonephritis)** to more advanced glomerular lesions with diffuse necrosis with parietal cell proliferation resulting in crescent formation **(crescentic glomerulonephritis)**.

Clinical Features

Males are affected more often than females, at an average age of about 40 years. Classic features include bilateral pneumonitis with nodules and cavitary lesions (95%), chronic sinusitis (90%), mucosal ulcerations of the nasopharynx (75%), and renal disease (80%). Rashes, myalgias, articular involvement, neuritis, and fever can also occur. Left untreated, the disease is usually rapidly fatal with 80% mortality within 1 year. Treatment with steroids, cyclophosphamide, TNF antagonists, and anti–B-cell antibodies, have turned GPA into a chronic relapsing and remitting disease.

Churg-Strauss Syndrome

Churg-Strauss syndrome (also called *allergic granulomatosis and angiitis*) is a small-vessel necrotizing vasculitis classically associated with asthma, allergic rhinitis, lung infiltrates, peripheral eosinophilia, extravascular necrotizing granulomas, and a striking eosinophilic infiltration of vessels and tissues.

Cutaneous involvement (with palpable purpura), gastrointestinal bleeding, and renal disease (primarily as focal and segmental glomerulosclerosis) are the major associations. Cytotoxicity produced by the myocardial eosinophilic infiltrates often leads to cardiomyopathy; cardiac involvement is seen in 60% of patients and is a major cause of morbidity and death.

Churg-Strauss syndrome is likely a consequence of hyperresponsiveness to some normally innocuous allergic stimulus. MPO-ANCAs are present in a minority of cases, suggesting that the disorder is pathogenically heterogeneous. The vascular lesions differ from those of PAN or microscopic polyangiitis by virtue of the presence of both granulomas and eosinophils.

Thromboangiitis Obliterans (Buerger Disease)

Thromboangiitis obliterans is characterized by segmental, thrombosing, acute and chronic inflammation of medium- and small-sized arteries, especially the tibial and radial arteries, that often leads to vascular insufficiency, typically of the extremities. It occurs almost exclusively in heavy cigarette smokers, usually before the age of 35.

Pathogenesis

The strong relationship with cigarette use involves either direct idiosyncratic EC toxicity caused by some tobacco component or an immune response to the same agents that have modified host vascular wall proteins. Most patients with Buerger disease have hypersensitivity to intradermally injected tobacco extracts, and their vessels exhibit impaired endothelium-dependent vasodilation when challenged with acetylcholine. There is an increased prevalence in certain ethnic groups (Israeli, Indian subcontinent, Japanese) and an association with certain MHC haplotypes.

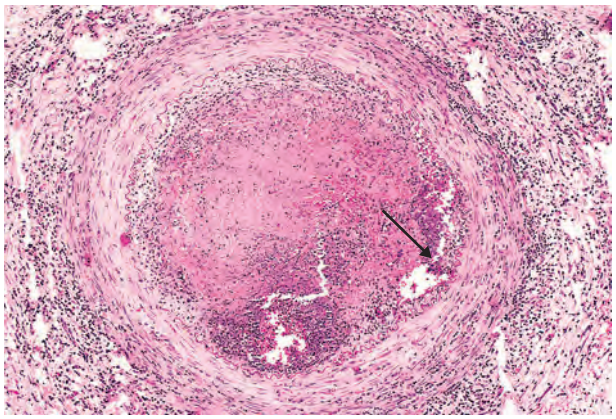


Figure 11.29 Thromboangiitis obliterans (Buerger disease). The lumen is occluded by a thrombus containing abscesses *(arrow)*, and the vessel wall is infiltrated with leukocytes.

MORPHOLOGY

Affected vessels of thombangitis obliterans show acute and chronic inflammation, accompanied by luminal thrombosis. The thrombus can contain small **microabscesses** composed of neutrophils surrounded by granulomatous inflammation (Fig. 11.29); the thrombus may eventually organize and recanalize. The inflammatory process extends into contiguous veins and nerves (rare with other forms of vasculitis), and with time all three structures can be encased in fibrous tissue.

Clinical Features

Early manifestations include Raynaud phenomenon (see later), instep foot pain induced by exercise (instep claudication), and a superficial nodular phlebitis (venous inflammation). The vascular insufficiency of Buerger disease tends to be accompanied by severe pain—even at rest—undoubtedly due to the neural involvement. Chronic extremity ulcerations can develop, progressing over time (occasionally precipitously) to frank gangrene. Smoking abstinence in the early stages of the disease can often ameliorate further attacks; however, once established, the vascular lesions do not respond to smoking abstinence.

Vasculitis Associated With Other Noninfectious Disorders

Vasculitis resembling hypersensitivity angiitis or classic PAN can sometimes be associated with other disorders, such as rheumatoid arthritis; systemic lupus erythematosus; malignancy; or systemic illnesses, such as mixed cryoglobulinemia, antiphospholipid antibody syndrome, and Henoch-Schönlein purpura. Rheumatoid vasculitis occurs predominantly in the setting of long-standing, severe rheumatoid arthritis and usually affects small- and medium-sized arteries,

leading to visceral infarction; it may also cause a clinically significant aortitis. Identifying the underlying pathology has therapeutic significance. For example, although classic immune complex lupus vasculitis and antiphospholipid antibody syndrome can be morphologically and clinically similar, antiinflammatory therapy is required in the former, and aggressive anticoagulant therapy is indicated in the latter.

Infectious Vasculitis

Arteritis can be caused by the direct invasion of infectious agents, usually bacteria or fungi, and in particular *Aspergillus* and *Mucor* species. Vascular invasion can be part of a localized tissue infection (e.g., bacterial pneumonia or adjacent to abscesses) or, less commonly, can arise from hematogenous spread of microorganisms during septicemia or embolization from infective endocarditis.

Vascular infections can weaken arterial walls and culminate in mycotic aneurysms (see earlier) or can induce thrombosis and downstream infarction. Thus inflammation-induced thrombosis of meningeal vessels in bacterial meningitis can eventually give rise to infarction of the underlying brain.

KEY CONCEPTS

VASCULITIS

- Vasculitis is defined as inflammation of vessel walls; it is frequently associated with systemic manifestations (including fever, malaise, myalgias, and arthralgias) and organ dysfunction that depends on the pattern of vascular involvement.
- Vasculitis most commonly has an immunologic basis resulting from immune complex deposition, ANCAs, or anti-EC antibodies. Less commonly it is caused by infections.
- Different forms of vasculitis tend to specifically affect vessels of a particular caliber and location (see Fig. 11.24 and Table 11.4).

DISORDERS OF BLOOD VESSEL HYPERREACTIVITY

Several disorders are characterized by inappropriate or excessive blood vessel vasoconstriction.

Raynaud Phenomenon

Raynaud phenomenon results from exaggerated vasoconstriction of arteries and arterioles in responses to cold or emotion. It most commonly affects the extremities, particularly the fingers and toes, but also occasionally the nose, earlobes, or lips. The restricted blood flow induces paroxysmal pallor and even cyanosis in severe cases; involved digits classically show "red, white, and blue" color changes from most proximal to most distal, correlating with proximal vasodilation, central vasoconstriction, and more distal cyanosis (Fig. 11.30). Raynaud phenomenon can be a primary entity or secondary to other disorders.

- *Primary Raynaud phenomenon* affects 3% to 5% of the general population and occurs most often in young women. It tends to symmetrically affect the extremities, and the severity and extent of involvement typically does not progress. It results from intrinsic hyperreactivity of medial SMCs. Structural changes in the arterial walls are absent except in long-standing disease, when intimal thickening can develop. The course is usually benign, but chronicity can lead to atrophy of the skin, subcutaneous tissues, and muscles. Ulceration and ischemic gangrene are rare.
- *Secondary Raynaud phenomenon* refers to vascular insufficiency due to arterial disease caused by other entities, including SLE, scleroderma, Buerger disease, or even atherosclerosis. Clinically, secondary Raynaud phenomenon tends to have asymmetric involvement of the extremities and progressively worsens (extent and severity) over time.

Since Raynaud phenomenon may be the first manifestation of immune-mediated vasculitides, any patient with new

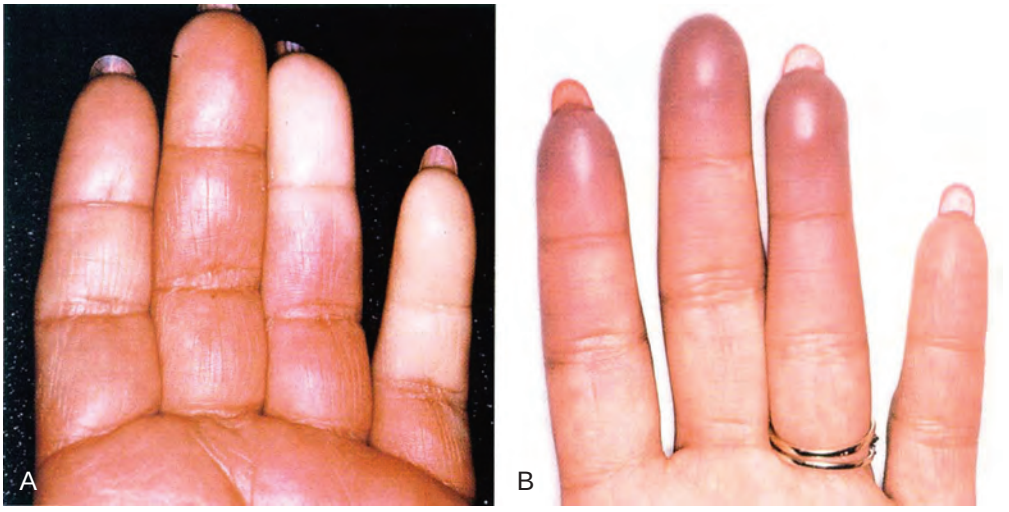


Figure 11.30 Raynaud phenomenon. (A) Sharply demarcated pallor of the distal fingers resulting from the closure of digital arteries. (B) Cyanosis of the fingertips. (Reproduced from Salvarani C, et al: Polymyalgia rheumatica and giant-cell arteritis, *N Engl J Med* 347:261, 2002.)

symptoms should be evaluated. Of these, some 10% will eventually manifest an underlying disorder.

Myocardial Vessel Vasospasm

Excessive constriction of arteries or arterioles may cause ischemia, and persistent vasospasm can even cause myocardial infarction. In addition to intrinsic hyperreactivity of medial SMCs, as described for primary Raynaud disease (see earlier), high levels of vasoactive mediators can precipitate prolonged vascular contraction. Such agents can be endogenous (e.g., epinephrine released by pheochromocytomas) or exogenous (cocaine or phenylephrine). Elevated thyroid hormone causes a similar effect by increasing the sensitivity of vessels to circulating catecholamines, while autoantibodies and T cells in scleroderma (Chapter 6) can cause vascular instability and vasospasm. Extreme psychological stress and the attendant release of catecholamines can lead to pathologic vasospasm.

When vasospasm of cardiac arterial or arteriolar beds (so-called cardiac Raynaud) is of sufficient duration (20 to 30 minutes), myocardial infarction or sudden cardiac death may occur.

VEINS AND LYMPHATICS

Varicose veins and phlebothrombosis/thrombophlebitis together account for at least 90% of clinical venous disease.

Varicose Veins

Varicose veins are abnormally dilated, tortuous veins produced by prolonged, increased intraluminal pressure with vessel dilation and incompetence of the venous valves. The superficial veins of the upper and lower leg are commonly involved because venous pressures in these sites can be markedly elevated (up to 10 times normal) by prolonged dependent posture. Up to 20% of men and a third of women develop lower extremity varicose veins. Obesity increases the risk, and the higher incidence in women probably reflects the prolonged elevation in venous pressure caused by compression of the inferior vena cava by the gravid uterus during pregnancy. A familial predilection to varicose veins reflects defective venous wall development.

Clinical Features

Incompetence of the venous valves leads to stasis, congestion, edema, pain, and thrombosis. Secondary tissue ischemia results from chronic venous congestion and poor vessel drainage leading to stasis dermatitis (also called "brawny induration"; the brawny color comes from the hemolysis of extravasated red cells) and ulcerations; poor wound healing and superimposed infections are common additional complications. Notably, embolism from these superficial veins is very rare, as opposed to the relatively frequent thromboembolism that arises from thrombosed deep veins (see later and Chapter 4).

Varicosities in two other sites also deserve mention:

- *Esophageal varices.* Liver cirrhosis (less frequently, portal vein obstruction or hepatic vein thrombosis) causes portal vein hypertension (Chapter 18). Portal hypertension leads to the opening of portosystemic shunts that increase blood flow into veins at the gastroesophageal junction (forming esophageal varices), rectum (forming hemorrhoids), and periumbilical veins of the abdominal wall (forming a caput medusa). Esophageal varices are the most important since their rupture can lead to massive (even fatal) upper gastrointestinal hemorrhage.
- *Hemorrhoids* can also result from primary varicose dilation of the venous plexus at the anorectal junction (e.g., through prolonged pelvic vascular congestion due to pregnancy or straining to defecate). Hemorrhoids are uncomfortable and may be a source of bleeding; they can also thrombose and are prone to painful ulceration.

Thrombophlebitis and Phlebothrombosis

Thrombophlebitis and phlebothrombosis are largely interchangeable designations for venous thrombosis and inflammation; deep leg veins account for more than 90% of cases. The periprostatic venous plexus in males and the pelvic venous plexus in females are additional sites, as are the large veins in the skull and the dural sinuses (especially in the setting of infection or inflammation). Portal vein thrombosis can occur in association with peritoneal infections (peritonitis, appendicitis, salpingitis, and pelvic abscesses), as well as certain thrombophilic conditions associated with platelet hyperactivity (e.g., polycythemia vera) (Chapter 14).

Decreased blood flow in the setting of prolonged immobilization is the most common cause of lower extremity *deep venous thrombosis (DVT).* This can occur with extended bed rest or sitting during long airplane or automobile excursions; the postoperative state can also precipitate DVT, and congestive heart failure, pregnancy, oral contraceptive use, malignancy, and obesity all constitute additional risk factors.

Systemic hypercoagulability, including genetic hypercoagulability syndromes (Chapter 4), can also contribute to thrombophlebitis. In patients with cancer, particularly adenocarcinomas, hypercoagulability occurs as a paraneoplastic syndrome related to tumor elaboration of procoagulant factors (Chapter 7). In this setting, venous thromboses classically appear in one location, disappear, and then occur in another site—so-called migratory thrombophlebitis *(Trousseau syndrome).*

Thrombi in the legs tend to produce few, if any, reliable signs or symptoms. Indeed, local manifestations including vein dilation, edema, cyanosis, heat, erythema, or pain may be entirely absent, especially in bedridden patients. In some cases, pain can be elicited by pressure over affected veins, squeezing the calf muscles, or forced dorsiflexion of the foot *(Homan sign)*; absence of these findings does not exclude a diagnosis of DVT.

Pulmonary embolism is a serious clinical complication of *DVT* (Chapter 4), resulting from fragmentation or detachment of the venous thrombus. In many cases the first manifestation of thrombophlebitis is a pulmonary embolus. Depending on the size and number of emboli, the outcome can range from no symptoms to death.

Superior and Inferior Vena Caval Syndromes

The superior vena caval syndrome is usually caused by neoplasms that compress or invade the superior vena cava (e.g., bronchogenic carcinoma or mediastinal lymphoma).

The resulting obstruction produces a characteristic clinical complex including marked dilation of the veins of the head, neck, and arms with cyanosis. Pulmonary vessels can also be compressed, inducing respiratory distress.

The inferior vena caval syndrome can be caused by neoplasms that compress or invade the inferior vena cava (IVC) or by thrombosis of the hepatic, renal, or lower extremity veins that propagates cephalad. Certain neoplasms—particularly hepatocellular carcinoma and renal cell carcinoma—show a striking tendency to grow within veins, and these can ultimately occlude the IVC. IVC obstruction induces marked lower extremity edema, distention of the superficial collateral veins of the lower abdomen, and—with renal vein involvement—massive proteinuria.

Lymphangitis and Lymphedema

Although primary disorders of lymphatic vessels are extremely uncommon, secondary processes frequently develop in association with inflammation or malignancies.

Lymphangitis represents acute inflammation caused by the spread of bacterial infections into lymphatics; group A β-hemolytic streptococcus is the most common agent. Affected lymphatics are dilated and filled with an exudate of neutrophils and monocytes; the infiltrates can extend through the vessel wall and, in severe cases, can produce cellulitis or focal abscesses. Lymphangitis is manifested by red, painful subcutaneous streaks (the inflamed lymphatics), with painful enlargement of the draining lymph nodes (lymphadenitis). If bacteria are not successfully contained within the lymph nodes, subsequent escape into the venous circulation can result in bacteremia or sepsis.

Primary lymphedema can occur as an isolated congenital defect (simple congenital lymphedema) or as the familial *Milroy disease (heredofamilial congenital lymphedema)*, which results in lymphatic agenesis or hypoplasia. Secondary or obstructive lymphedema stems from blockage of previously normal lymphatics, including:

- *Tumors* obstructing lymphatic channels or the regional lymph nodes
- *Surgical procedures* that sever lymphatic connections (e.g., axillary lymph node resection in a radical mastectomy)
- *Postradiation fibrosis*
- *Filariasis*
- *Postinflammatory thrombosis and scarring*

Regardless of the cause, lymphedema increases the hydrostatic pressure in the lymphatics distal to the obstruction and causes edema. Chronic edema, in turn, can lead to ECM deposition and fibrosis, producing brawny induration or a *peau d'orange* (orange peel) appearance of the overlying skin. Eventually inadequate tissue perfusion can lead to skin ulceration. Rupture of dilated lymphatics (e.g., secondary to obstruction from a tumor or following surgical injury) leads to milky accumulations of lymph designated as chylous ascites (abdomen), chylothorax, and chylopericardium.

VASCULAR TUMORS

Tumors of blood vessels and lymphatics constitute a spectrum from benign hemangiomas (extremely common) to locally aggressive neoplasms that rarely metastasize, to rare, highly malignant angiosarcomas (Table 11.5).

Table 11.5 Classification of Vascular Tumors and Tumor-Like Conditions

Benign Neoplasms, Developmental and Acquired Conditions
Hemangioma
Capillary hemangioma
Cavernous hemangioma
Pyogenic granuloma
Lymphangioma
Simple (capillary) lymphangioma
Cavernous lymphangioma (cystic hygroma)
Glomus tumor
Vascular ectasias
Nevus flammeus
Spider telangiectasia (arterial spider)
Hereditary hemorrhagic telangiectasis (Osler-Weber-Rendu disease)
Reactive vascular proliferations
Bacillary angiomatosis
Intermediate-Grade Neoplasms
Kaposi sarcoma
Hemangioendothelioma
Malignant Neoplasms
Angiosarcoma
Hemangiopericytoma

Vascular neoplasms arise either from endothelium (e.g., hemangioma, lymphangioma, angiosarcoma) or from cells that support or surround blood vessels (e.g., glomus tumor). Primary tumors of large vessels (aorta, pulmonary artery, and vena cava) are mostly sarcomas. Although a benign well-differentiated hemangioma is not usually confused with an anaplastic angiosarcoma, the distinction between benign and malignant can occasionally be difficult. Congenital or developmental malformations and nonneoplastic reactive vascular proliferations (e.g., *bacillary angiomatosis*) can also present as tumor-like lesions that can be diagnostically challenging. In general, benign and malignant vascular neoplasms can be distinguished by the following features.

- *Benign tumors* usually contain obvious vascular channels filled with blood cells (lymphatics will be filled with lymph), lined by a monolayer of normal-appearing ECs.
- *Malignant tumors* are more solidly cellular and more proliferative and exhibit cytologic atypia; they usually do not form well-organized vessels. The endothelial derivation of such neoplastic proliferations may require immunohistochemical detection of EC-specific markers such as CD31 or von Willebrand factor.

Benign Tumors and Tumor-Like Conditions

Vascular Ectasias. *Ectasia* is a generic term for any local dilation of a structure, while *telangiectasia* is used to describe a permanent dilation of preexisting small vessels (capillaries, venules, and arterioles, usually in the skin or mucous membranes) that forms a discrete red lesion. These can be congenital or acquired and are not true neoplasms.

- *Nevus flammeus* (a "birthmark"), the most common form of vascular ectasia, is a light pink to deep purple flat lesion on the head or neck composed of dilated vessels. Most ultimately regress spontaneously.
- The so-called *port wine stain* is a type of nevus flammeus that tends to grow during childhood, thickening the associated skin surface, and persisting over time (perhaps the most well-known example is that on the forehead of the former Soviet President Mikhail Gorbachev). Such lesions in the distribution of the trigeminal nerve can be associated with *Sturge-Weber syndrome* (also called encephalotrigeminal angiomatosis). This uncommon congenital disorder is associated with facial port wine nevi, ipsilateral venous angiomas in the cortical leptomeninges, intellectual disability, seizures, hemiplegia, and skull radiopacities. A large facial telangiectasia in a child with intellectual deficiency may indicate the presence of additional vascular malformations. The majority of cases of port wine stains (including those associated with Sturge-Weber syndrome) are caused by a somatic single-nucleotide missense mutation in GNAQ (the alpha subunit of heterotrimeric transmembrane signaling molecule).
- *Spider telangiectasias* are nonneoplastic vascular malformations grossly resembling a spider. These manifest as radial, often pulsatile lesions composed of dilated subcutaneous arteries or arterioles (resembling spider legs) about a central core (resembling a spider's body) that blanch with pressure. These commonly occur on the face, neck, or upper chest and are most frequently associated with hyperestrogenic states (e.g., in pregnant women or patients with cirrhosis).
- *Hereditary hemorrhagic telangiectasia (Osler-Weber-Rendu disease)* is an autosomal dominant disorder caused by mutations in genes that encode components of the TGF-β signaling pathway in ECs. The telangiectasias are malformations composed of dilated capillaries and veins that are present at birth. They are widely distributed over the skin and oral mucous membranes, as well as in the respiratory, gastrointestinal, and urinary tracts. The lesions can spontaneously rupture, causing serious epistaxis (nosebleed), gastrointestinal bleeding, or hematuria.

Hemangiomas. Hemangiomas are very common tumors composed of blood-filled vessels (Fig. 11.31); they constitute 7% of all benign tumors of infancy and childhood. Most are present from birth and initially increase in size, but many eventually regress spontaneously. While hemangiomas are typically confined to the head, neck, and thoracic skin, they can also arise internally and can occasionally be more extensive *(angiomatosis)*. Nearly one-third of internal lesions occur in the liver. Malignant transformation is rare. Several histologic and clinical variants have been described.

- *Capillary hemangiomas* are the most common type; these occur in the skin, subcutaneous tissues, and mucous

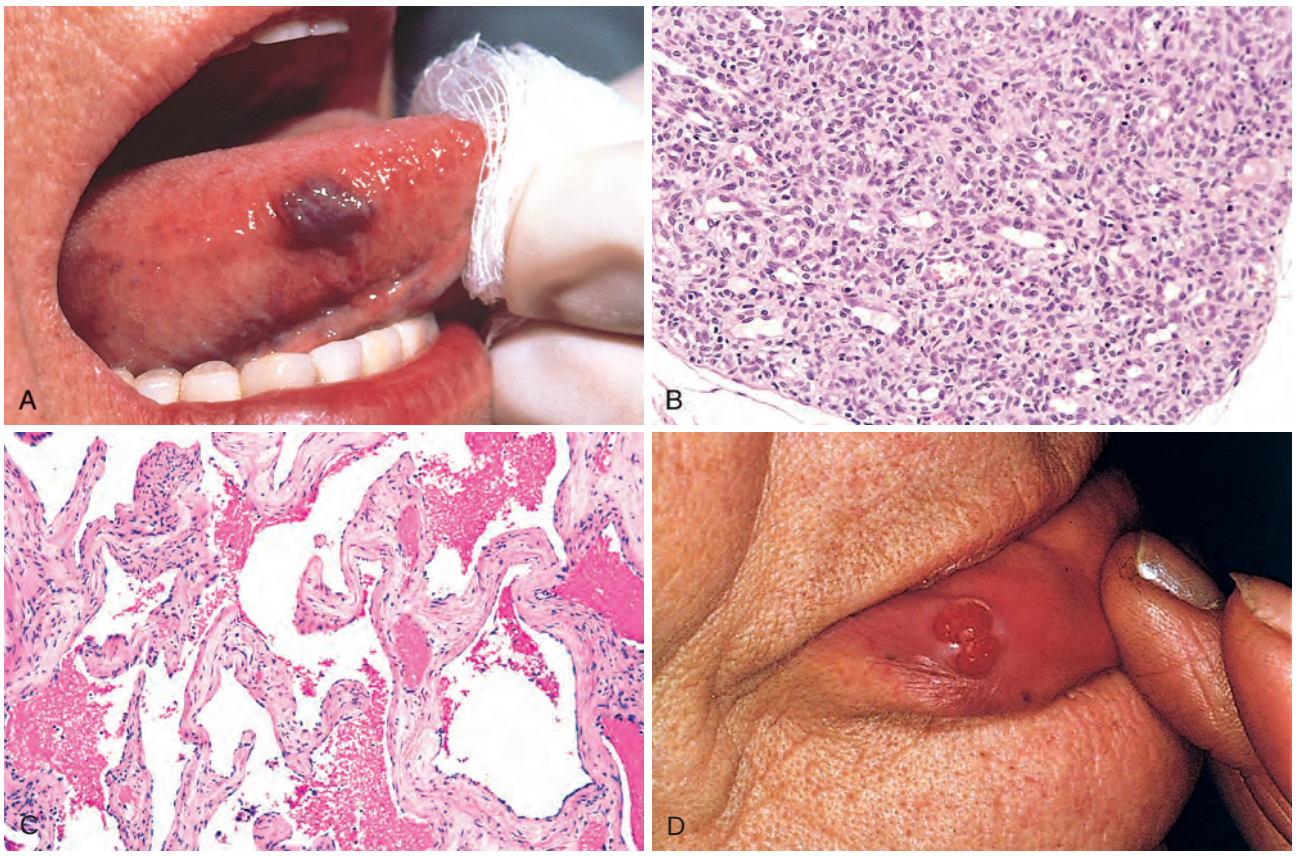


Figure 11.31 Hemangiomas. (A) Hemangioma of the tongue. (B) Histology of juvenile capillary hemangioma. (C) Histology of cavernous hemangioma. (D) Pyogenic granuloma of the lip. (A and D, Courtesy John Sexton, MD, Beth Israel Hospital, Boston, Mass.; B, courtesy Christopher D. M. Fletcher, MD, Brigham and Women's Hospital, Boston, Mass.; and C, courtesy Thomas Rogers, MD, University of Texas Southwestern Medical School, Dallas, Tex.)

membranes of the oral cavities and lips, as well as in the liver, spleen, and kidneys (Fig. 11.31A). Histologically, they are composed of thin-walled capillaries with scant stroma (Fig. 11.31B).

- *Juvenile hemangiomas* (so-called strawberry-type hemangiomas) of the newborn skin are extremely common (1 in 200 births) and can be multiple. These grow rapidly for a few months, but then fade by 1 to 3 years of age and completely regress by age 7 in most cases.
- *Cavernous hemangiomas* are composed of large, dilated vascular channels. In contradistinction to capillary hemangiomas, cavernous hemangiomas are more infiltrative, frequently involve deep structures, and do not spontaneously regress. Moreover, cavernous hemangiomas detected by imaging studies may be difficult to distinguish from their malignant counterparts. On histologic examination, the mass is sharply delineated but unencapsulated and composed of large, cavernous blood-filled vascular spaces, separated by connective tissue stroma (Fig. 11.31C). Intravascular thrombosis with associated dystrophic calcification is common. They can be locally destructive, but more often are of little clinical significance outside of a cosmetic disfigurement or their vulnerability to traumatic ulceration and bleeding. Brain hemangiomas are problematic, as they can cause symptoms related to compression of adjacent tissue or rupture. Cavernous hemangiomas are one component of *von Hippel-Lindau disease* (Chapter 28), in which vascular lesions are commonly found in the cerebellum, brain stem, retina, pancreas, and liver.
- *Pyogenic granulomas* are capillary hemangiomas that present as rapidly growing red pedunculated lesions on the skin, gingiva, or oral mucosa. They bleed easily and are often ulcerated (Fig. 11.31D). Roughly a quarter of lesions develop after trauma, reaching a size of 1 to 2 cm within a few weeks. Curettage and cautery is usually curative. Pregnancy tumor (granuloma gravidarum) is a pyogenic granuloma that occurs infrequently (1% of patients) in the gingiva of pregnant women. These lesions may spontaneously regress (especially after pregnancy) or undergo fibrosis, but occasionally require surgical excision.

Lymphangiomas. Lymphangiomas are the benign lymphatic counterparts of hemangiomas.

- *Simple (capillary) lymphangiomas* are slightly elevated or sometimes pedunculated lesions up to 1 to 2 cm in diameter that occur predominantly in the head, neck, and axillary subcutaneous tissues. Histologically, lymphangiomas exhibit networks of endothelium-lined spaces that can be distinguished from capillary channels by lymphatic endothelial markers (e.g., VEGFR-3, LYVE-1, and others) or by the absence of erythrocytes.
- *Cavernous lymphangiomas (cystic hygromas)* are typically found in the neck or axilla of children and more rarely in the retroperitoneum. Cavernous lymphangiomas can occasionally be enormous (up to 15 cm), filling the axilla or producing gross deformities about the neck; cavernous lymphangiomas of the neck are common in *Turner syndrome*. The lesions are composed of massively dilated lymphatic spaces lined by ECs and separated by intervening connective tissue stroma containing lymphoid aggregates. The tumor margins are indistinct and unencapsulated, making resection difficult.

Glomus Tumor (Glomangioma). Glomus tumors are benign but exquisitely painful tumors arising from modified SMCs of the glomus bodies, arteriovenous structures involved in thermoregulation. Although they may superficially resemble hemangiomas, glomangiomas arise from SMCs rather than ECs. They are most commonly found in the distal portion of the digits, especially under the fingernails. Excision is curative.

Bacillary Angiomatosis. *Bacillary angiomatosis* is a vascular proliferation in immunocompromised hosts (e.g., patients with acquired immunodeficiency syndrome [AIDS]) caused by opportunistic bacilli of the *Bartonella* family. Lesions can involve the skin, bone, brain, and other organs. Two species are implicated.

- *Bartonella henselae,* whose principal reservoir is the domestic cat, causes *cat-scratch disease* (a necrotizing granulomatous disorder of lymph nodes) in immunocompetent hosts.
- *Bartonella quintana* is transmitted by human body lice; this microbe was the cause of trench fever in World War I.

MORPHOLOGY

Skin lesions are red papules and nodules or rounded subcutaneous masses; histologically, there is capillary proliferation with prominent epithelioid ECs exhibiting nuclear atypia and mitoses, with scattered stromal neutrophils and causal bacteria (Fig. 11.32).

The bacteria induce host tissues to produce hypoxia-inducible factor-1α (HIF-1α), which, in turn, drives vascular endothelial growth factor (VEGF) production and ultimately vascular proliferation. The infections (and lesions) are cleared by antibiotics.

Intermediate-Grade (Borderline) Tumors

Kaposi Sarcoma. **Kaposi sarcoma (KS) is a vascular neoplasm caused by human herpesvirus 8 (HHV8, also known as Kaposi sarcoma herpesvirus).** Although it occurs in a number of contexts, it is by far most common in patients with AIDS; indeed, its presence is used as a criterion for diagnosis of AIDS. Based on population demographics and risk factors, KS is categorized into four forms:

- *Classic KS* is a disorder of older men of Mediterranean, Middle Eastern, or Eastern European descent (especially Ashkenazic Jews); it is uncommon in the United States. It can be associated with malignancy or altered immunity, but it is not associated with human immunodeficiency virus (HIV) infection. Classic KS manifests as multiple red-purple skin plaques or nodules, usually in the distal lower extremities; these progressively increase in size and number and spread proximally. Although persistent, the tumors are typically asymptomatic and remain localized to the skin and subcutaneous tissue.
- *Endemic African KS* typically occurs in HIV-seronegative individuals younger than age 40 years and can follow an indolent or aggressive course; it involves lymph nodes much more frequently than the classic variant. In combination with AIDS-associated KS (see later), KS is one of the

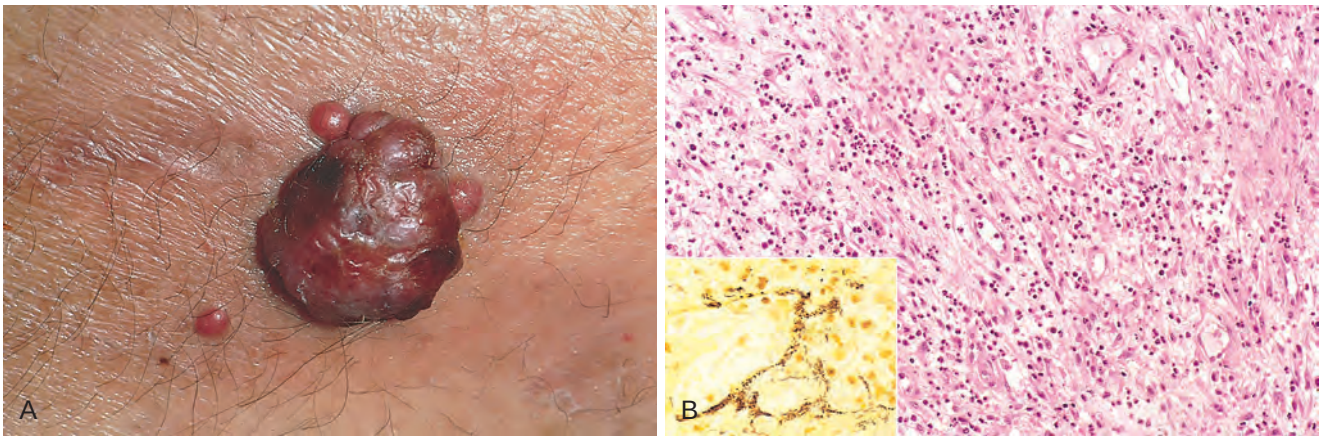


Figure 11.32 Bacillary angiomatosis. (A) Characteristic cutaneous lesion. (B) Histologic appearance with acute neutrophilic inflammation and vascular (capillary) proliferation. *Inset,* Modified silver (Warthin-Starry) stain demonstrates clusters of tangled bacilli (black). (A, Courtesy Richard Johnson, MD, Beth Israel Deaconess Medical Center, Boston, Mass.; B and *inset,* courtesy Scott Granter, MD, Brigham and Women's Hospital, Boston, Mass.)

more common tumors in central Africa. A particularly severe form, with prominent lymph node and visceral involvement, occurs in prepubertal children; the prognosis is poor, with almost 100% mortality within 3 years.

- *Transplant-associated KS* occurs in solid-organ transplant recipients receiving T-cell immunosuppression; the risk can be 100-fold greater than for immunocompetent patients. Transplant-associated KS typically follows an aggressive course involving lymph nodes, mucosa, and viscera; cutaneous lesions may be absent. Lesions can regress as immunosuppression is reduced, but at the risk of organ rejection.
- *AIDS-associated (epidemic) KS* is an AIDS-defining illness; worldwide, it represents the most common HIV-related malignancy. Although the incidence of KS has fallen more than 80% with the advent of aggressive antiretroviral therapies, it still occurs in HIV-infected individuals 300-fold more commonly than in transplant recipients and 1000-fold greater than in the general population. AIDS-associated KS often involves lymph nodes and disseminates widely to viscera early in its course; most patients eventually die of opportunistic infections rather than KS.

Pathogenesis

Virtually all KS lesions are infected by *HHV8.* Like Epstein-Barr virus, HHV8 is a γ-herpesvirus. It is transmitted sexually and potentially via oral secretions and cutaneous exposures (of note, the prevalence of endemic African KS is inversely related to the wearing of shoes). HHV8 and altered T-cell immunity are likely required for KS development; in elderly adults, diminished T-cell immunity may be related to aging. Acquired somatic mutations in the cells of origin also contribute to tumor development and progression.

HHV8 causes lytic and latent infections in ECs, both of which contribute to KS pathogenesis. A virally encoded G protein induces VEGF production, stimulating endothelial growth; in addition, cytokines produced by inflammatory cells recruited to sites of lytic infection create a local proliferative milieu. In latently infected cells, HHV8 encoded proteins disrupt normal cellular proliferation controls (e.g., through synthesis of a viral homologue of cyclin D) and prevent apoptosis by inhibiting p53. Thus the local inflammatory environment favors cellular proliferation, and latently infected cells have a growth advantage. In its early stages, only a few cells are HHV8-infected, but with time, virtually all of the proliferating cells carry the virus. Molecular pathogenesis is discussed in more detail in Chapter 6 along with HIV infections.

MORPHOLOGY

In **classic KS** (and sometimes in other variants), the cutaneous lesions progress through three stages:

- **Patches** are red-purple macules (Fig. 11.33A). Histology shows dilated irregular EC-lined vascular spaces with interspersed lymphocytes, plasma cells, and macrophages (sometimes containing hemosiderin). The lesions can be difficult to distinguish from granulation tissue.
- With time, lesions become larger, violaceous, **raised plaques** (see Fig. 11.33A) composed of dermal accumulations of dilated, jagged vascular channels lined and surrounded by plump spindle cells. Scattered between the vascular channels are extravasated erythrocytes, hemosiderin-laden macrophages, and other mononuclear inflammatory cells.
- Eventually, lesions become **nodular** and more distinctly neoplastic. These lesions are composed of sheets of plump, proliferating spindle cells, mostly in the dermis or subcutaneous tissues (Fig. 11.33B), containing small vessels and slitlike spaces with red cells. Marked hemorrhage, hemosiderin pigment, and mononuclear inflammation are present; mitotic figures are common, as are round, pink, cytoplasmic globules representing degenerating erythrocytes within phagolysosomes. The nodular stage often heralds lymph node and visceral involvement.

Clinical Features

The course of KS varies widely and is significantly influenced by the clinical setting. Most primary HHV8 infections are asymptomatic. Classic KS is—at least initially—largely restricted to the skin, and surgical resection is usually adequate for an excellent prognosis. Radiation can be used for multiple lesions in a restricted area, and chemotherapy yields satisfactory results for more disseminated disease

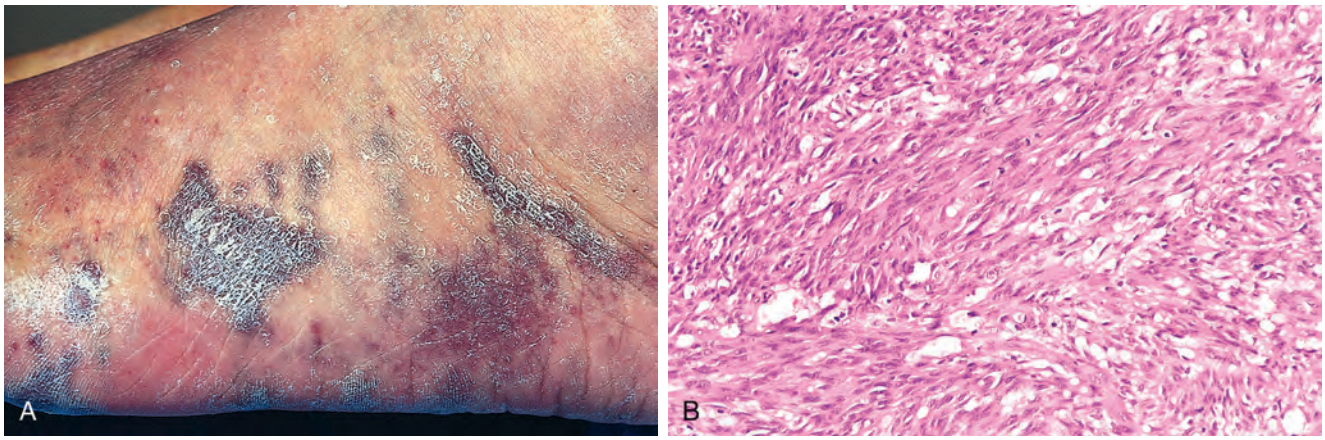


Figure 11.33 Kaposi sarcoma. (A) Gross photograph illustrating coalescent red-purple macules and plaques of the skin. (B) Histologic appearance of the nodular stage of Kaposi sarcoma, demonstrating sheets of plump, proliferating spindle cells. (B, Courtesy Christopher D. M. Fletcher, MD, Brigham and Women's Hospital, Boston, Mass.)

including nodal involvement. In immunosuppression-associated KS, withdrawal of immunosuppression (perhaps with adjunct chemotherapy or radiotherapy) is often effective. For AIDS-associated KS, HIV antiretroviral therapy is generally beneficial, with or without additional therapy. Interferon-γ and angiogenesis inhibitors also have proved variably effective, while newer strategies aimed at specific kinases downstream of VEGF receptors show promise.

Hemangioendothelioma. Hemangioendotheliomas comprise a spectrum of borderline vascular neoplasms with clinical behaviors intermediate between benign, well-differentiated hemangiomas and frankly anaplastic angiosarcomas.

Epithelioid hemangioendothelioma is a tumor of adults occurring around medium- and large-sized veins. Well-defined vascular channels are inconspicuous, and neoplastic cells are plump and often cuboidal (resembling epithelial cells). The clinical behavior is extremely variable; most are cured by excision, but up to 40% recur, 20% to 30% eventually metastasize, and 15% of patients will die of their tumor.

Malignant Tumors

Angiosarcoma. Angiosarcomas are malignant endothelial neoplasms (Fig. 11.34) with histology varying from highly differentiated tumors that resemble hemangiomas to profoundly anaplastic lesions. Older adults, males and females, are more commonly affected; angiosarcomas occur at any site, but most often involve skin, soft tissue, breast, and liver.

Angiosarcomas can also arise in the setting of lymphedema, classically in the ipsilateral upper extremity several years after radical mastectomy for breast cancer (i.e., after lymph node resection and/or radiation); in such cases the tumor presumably arises from lymphatic vessels *(lymphangiosarcoma).* Angiosarcomas have also been induced by radiation and are rarely associated with prolonged insertion of foreign material (e.g., prosthetic devices).

Hepatic angiosarcomas are associated with a variety of carcinogenic exposures including arsenic (e.g., in pesticides), Thorotrast (a radioactive contrast agent formerly used for radiologic imaging), and polyvinyl chloride (one of the best-known examples of human chemical carcinogenesis). All of these agents have long latencies between initial exposure and eventual tumor development.

Angiosarcomas are locally invasive and can readily metastasize, with 5-year survival rates of approximately 30%.

MORPHOLOGY

Cutaneous angiosarcomas can begin as multiple, deceptively small and asymptomatic red papules or nodules. More advanced lesions are large, fleshy masses of red-tan to gray-white tissue with margins blurring imperceptibly into surrounding structures (Fig. 11.34A). Necrosis and hemorrhage are common.

Microscopically, **all degrees of differentiation can be seen,** from plump, atypical ECs forming vascular channels (Fig. 11.34B) to wildly undifferentiated tumors with no discernible blood vessels. The endothelial origin of these tumors can be demonstrated by immunohistochemical staining for CD31 or von Willebrand factor (Fig. 11.34C).

KEY CONCEPTS

VASCULAR TUMORS

- Vascular ectasias are not neoplasms, but rather dilations of existing vessels.
- Vessel neoplasms can derive from either blood vessels or lymphatics and can be composed of ECs (hemangioma, lymphangioma, angiosarcoma) or other components of vascular wall cells (glomus tumor).
- Most vascular tumors are benign (e.g., hemangiomas); some have an intermediate, locally aggressive behavior (e.g., Kaposi sarcoma); and others are highly malignant (e.g., angiosarcoma).
- Benign tumors typically form obvious vascular channels lined by normal-appearing ECs. Malignant tumors are more often solid and cellular, exhibit cytologic atypia, and lack well-defined vessels.

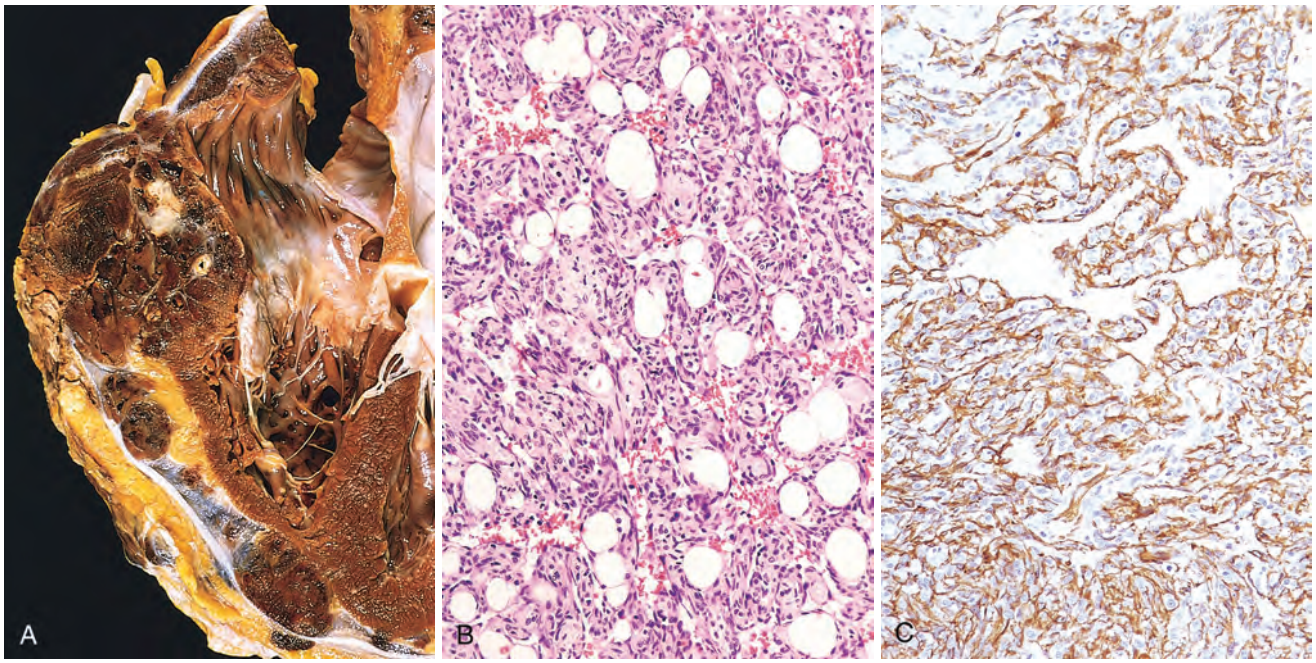


Figure 11.34 Angiosarcoma. (A) Angiosarcoma involving the right ventricle. (B) Moderately differentiated angiosarcoma with dense clumps of atypical cells lining distinct vascular lumens. (C) Immunohistochemical staining for the endothelial cell marker CD31 demonstrating the endothelial nature of the tumor cells.

PATHOLOGY OF VASCULAR INTERVENTION

The morphologic changes that occur in vessels following therapeutic intervention (e.g., stenting or bypass surgery) largely recapitulate the changes that occur in the setting of any vascular insult. Local trauma or thrombosis (e.g., due to a stent) or abnormal mechanical forces (e.g., a saphenous vein inserted into the arterial circulation as a coronary artery bypass graft) all induce the same stereotypical healing responses—fibromuscular intimal hyperplastic lesions composed of SMCs and ECM. Thus, just as with several risk factors for atherosclerosis, interventions that injure the endothelium also tend to induce intimal thickening by recruiting SMCs and promoting ECM deposition.

Endovascular Stenting

Arterial stenoses (especially those in coronary arteries) can be dilated by transiently inflating a balloon catheter to pressures sufficient to rupture the occluding plaque *(balloon angioplasty)*. Although most patients improve symptomatically following angioplasty alone, abrupt reclosure frequently occurs due to luminal compression caused by angioplasty-induced vascular dissection, by vessel wall spasm, or by thrombosis. Thus more than 95% of endovascular coronary procedures now involve both angioplasty and concurrent *coronary stent* placement.

Coronary stents are expandable tubes of metallic mesh. They provide a larger and more regular lumen, "tack down" the intimal flaps and dissections that occur during angioplasty, and mechanically limit vascular spasm. Nevertheless, due to endothelial injury, thrombosis is an important immediate post-stenting complication, and patients must receive potent antithrombotic agents (primarily platelet antagonists) to prevent acute catastrophic thrombotic occlusions. The long-term success of angioplasty is limited by the development of proliferative in-stent restenosis. This intimal thickening is due to SMC ingrowth, proliferation, and ECM synthesis, all driven by the initial vascular wall injury; it can cause clinically significant luminal occlusion in up to a third of patients within 6 to 12 months of stenting (Fig. 11.35).

The newest generation of *drug-eluting stents* is designed to avoid this complication by leaching antiproliferative drugs (e.g., paclitaxel or sirolimus) into the adjacent vessel wall to block SMC activation. Although the duration of drug elution is short (on the order of days to weeks), these drug-eluting stents nevertheless reduce the incidence of restenosis at 1 year by 50% to 80%. However, because of the antiproliferative effect of the drug-eluting stents, the time to reendothelialization is prolonged, and patients require extended courses of anticoagulation to prevent stent thrombosis.

Vascular Replacement

Synthetic or autologous vascular grafts can be used to replace damaged vessels or bypass diseased arteries. Large-bore (12- to 18-mm) synthetic conduits function well in high-flow locations such as the aorta; unfortunately, small-diameter artificial grafts (≤8 mm in diameter) generally fail as a result of early thrombosis or late intimal hyperplasia, the latter at the junction of the graft with the native vasculature (Fig. 11.36).

Consequently, when small-bore vessel replacement is required (e.g., in the >400,000 coronary bypass surgeries performed annually), the grafts are fashioned from saphenous

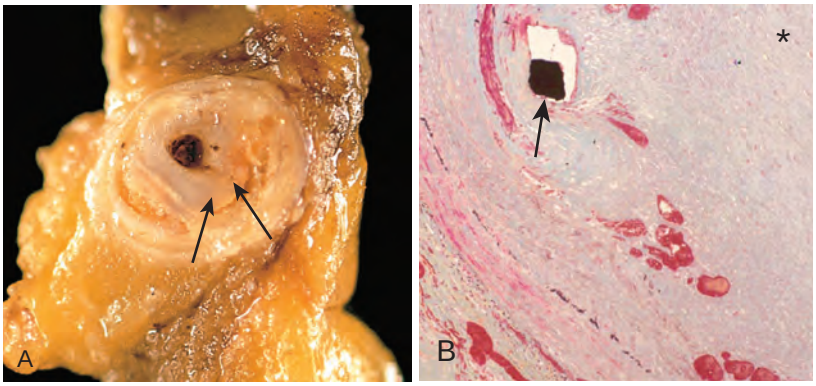


Figure 11.35 Restenosis after angioplasty and stenting. (A) Gross view demonstrating residual yellow atherosclerotic plaque *(arrows)* and a new, tan-white concentric intimal lesion inside of that plaque. (B) Histologic view shows a thickened neointima separating and overlying the stent wires (black diamond indicated by the *arrow*), which encroaches on the lumen (indicated by the *asterisk*); Movat stain with matrix staining gray-green. (B, Reproduced from Schoen FJ, Edwards WD: Pathology of cardiovascular interventions, including endovascular therapies, revascularization, vascular replacement, cardiac assist/replacement, arrhythmia control, and repaired congenital heart disease. In Silver MD, Gotlieb AI, Schoen FJ, editors: *Cardiovascular Pathology,* ed 3, Philadelphia, 2001, Churchill Livingstone.)

veins (taken from the patient's own leg) or left internal mammary arteries. The long-term patency of saphenous vein grafts averages only 50% at 10 years; grafts occlude due to thrombosis (typically early), intimal thickening (months to years postoperatively), and vein graft atherosclerosis—sometimes with superimposed plaque rupture, thrombosis, or aneurysms. By contrast, more than 90% of internal mammary artery grafts, which can be used only to bypass the left anterior descending artery, are patent at 10 years.

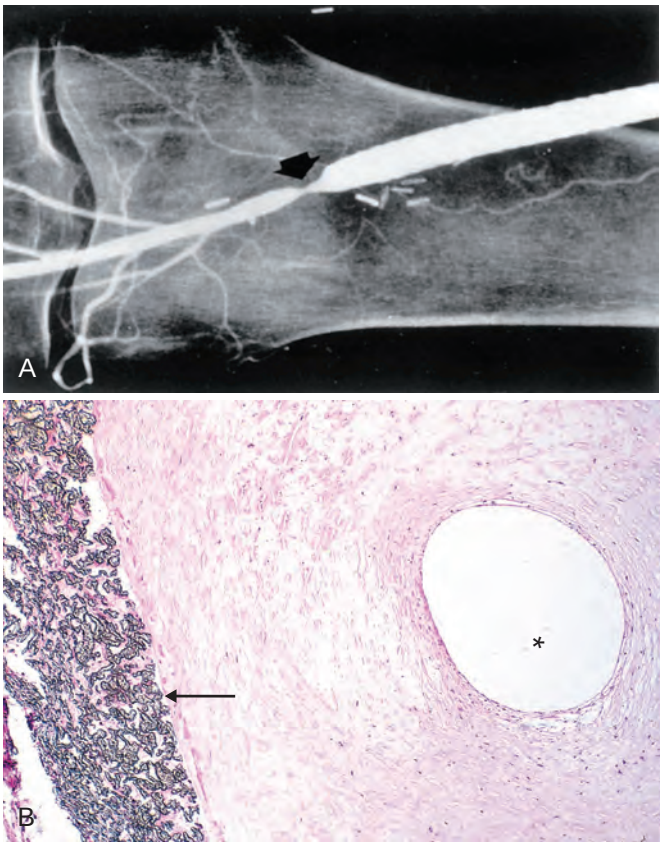


Figure 11.36 Intimal hyperplasia at the distal anastomosis of a synthetic femoral-popliteal graft. (A) Angiogram demonstrating constriction *(arrow)*. (B) Photomicrograph demonstrating Gore-Tex graft *(arrow)* with prominent intimal proliferation and very small residual lumen *(asterisk)*. (A, Courtesy Anthony D. Whittemore, MD, Brigham and Women's Hospital, Boston, Mass.)

SUGGESTED READINGS

Vascular Structure and Function

Axnick J, Lammert E: Vascular lumen formation, *Curr Opin Hematol* 19:192, 2012. *[Overview of physiologic and developmental blood vessel formation and remodeling].*

Bikfalvi A: The situation is more complex than anticipated. In *A Brief History of Blood and Lymphatic Vessels*, Cham, Switzerland, 2017, Springer. *[A nice review of the cellular and molecular underpinnings of vascular development including tumor vasculogenesis].*

Monahan-Earley R, Dvorak AM, Aird WC: Evolutionary origins of the blood vascular system and endothelium, *J Thromb Haemost* 11 Suppl 1:46, 2013. *[Interesting discussion of the evolutionary basis for vascular development including cogent explanations for endothelial heterogeneity].*

Vascular Wall Response to Injury

Gimbrone MA Jr, García-Cardeña G: Endothelial cell dysfunction and the pathobiology of atherosclerosis, *Circ Res* 11:620, 2016. *[Well-written review of endothelial responses to mechanical forces and other injury from one of the leading groups in the field].*

Pober JS, Min W, Bradley JR: Mechanisms of endothelial dysfunction, injury, and death, *Annu Rev Pathol* 4:71, 2009. *[Well-written scholarly review of the etiology and outcomes of endothelial injury].*

Hypertensive Vascular Disease

Coffman TM: Under pressure: the search for the essential mechanisms of hypertension, *Nat Med* 17:1402, 2011. *[Although not immediately current, this is a well-written review of the potential mechanisms underlying hypertension, the models that help elucidate pathways, and some translational advances].*

Singh M, Mensah GA, Bakris G: Pathogenesis and clinical physiology of hypertension, *Cardiol Clin* 28:545, 2010. *[Excellent overview of normal blood pressure regulation and the interaction of genetics and environment in the pathophysiology of hypertension].*

Atherosclerosis

Fuster JJ, Walsh K: Somatic mutations and clonal hematopoiesis: unexpected potential new drivers of age-related cardiovascular disease, *Circ Res* 122:523, 2018. *[Excellent paper elucidating the mechanistic links between aging, clonal hematopoiesis, and atherosclerosis].*

Hansson GK, Libby P, Tabas I: Inflammation and plaque vulnerability, *J Intern Med* 278:483, 2015. *[Nice review of the current concepts linking inflammation and plaque vulnerability].*

Jinnouchi H, Guo L, Sakamoto A et al: Diversity of macrophage phenotypes and responses in atherosclerosis, *Cell Mol Life Sci* 2019. [Epub ahead of print]. *[Solid review of the different macrophage activities in the atherosclerotic plaque].*

Jaiswal S et al: Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease, *N Engl J Med* 377:111, 2017. *[Remarkable new data showing the risk of atherosclerosis with increasing age may be related to clonal expansion of certain hematopoietic lineages.].*

Karasawa T, Takahashi M: Role of NLRP3 inflammasomes in atherosclerosis, *J Atheroscler Thromb* 24:443, 2017. *[Nice review describing the interaction of cholesterol microcrystals, phagocyte activation, inflammasome activation, cytokine production, and atherogenesis].*

Ridker P: From C-reactive protein to interleukin-6 to interleukin-1: moving upstream to identify novel targets for atheroprotection, *Circ Res* 118:145, 2016. *[Well-written opinion piece regarding how C-reactive protein links to other inflammatory mediators and potentially identifies therapeutic targets].*

Aneurysms and Dissection

Gillis E, Van Laer L, Loeys BL: Genetics of thoracic aortic aneurysm: at the crossroad of transforming growth factor-β signaling and vascular smooth muscle cell contractility, *Circ Res* 113:327, 2013. *[Good overview of the molecular and cellular basis for thoracic aneurysms].*

Michel JB, Jondeau G, Milewicz DM: From genetics to response to injury: vascular smooth muscle cells in aneurysms and dissections of the ascending aorta, *Cardiovasc Res* 11:578, 2018. *[Good discussion of the molecular pathways underlying human abdominal aortic aneurysm formation].*

Vasculitis

Jennette J, Falk R: Nosology of primary vasculitis, *Curr Opin Rheumatol* 19:10, 2007. *[Classification of vasculitis based on pathogenic mechanism and the vessels involved; provides a good organization to a complex and potentially confusing aspect of vascular pathology].*

Jennette JC, Falk RJ: Pathogenesis of antineutrophil cytoplasmic autoantibody-mediated disease, *Nat Rev Rheumatol* 10:463, 2014. *[Well-written and well-organized overview of the big picture in antineutrophil cytoplasmic autoantibody vasculitis].*

Disorders of Blood Vessel Hyperreactivity

MacAlpin RN: Some observations on and controversies about coronary arterial spasm, *Int J Cardiol* 181:389, 2015. *[Good clinical overview of coronary spasm and its sequelae].*

Veins and Lymphatics

Goldhaber SZ: Venous thromboembolism: epidemiology and magnitude of the problem, *Best Pract Res Clin Haematol* 25:235, 2012. *[Thorough clinically oriented overview of deep venous thrombosis and pulmonary embolism including epidemiology, mechanism of disease, and therapy].*

Vascular Tumors

Bhutani M, Polizzotto MN, Uldrick TS et al: Kaposi sarcoma-associated herpesvirus-associated malignancies: epidemiology, pathogenesis, and advances in treatment, *Semin Oncol* 42:247, 2015. *[Solid and thorough review].*

Goss JA, Greene AK: Congenital vascular tumors, *Otolaryngol Clin North Am* 51:89, 2018. *[Good overview of vascular tumors and malformations].*

Pathology of Vascular Intervention

Gallo G, Pierelli G, Forte M et al: Role of oxidative stress in the process of vascular remodeling following coronary revascularization, *Int J Cardiol* 268:27–33, 2018. *[A well-written review highlighting the important role of reactive oxygen species pathways in vascular remodeling].*

Inoue T et al: Vascular inflammation and repair: implications for re-endothelialization, restenosis, and stent thrombosis, *JACC Cardiovasc Interv* 4:1057, 2011. *[Excellent overview of the inflammatory pathways that influence the outcomes of percutaneous vessel interventions].*

Kim FY et al: Saphenous vein graft disease: review of pathophysiology, prevention, and treatment, *Cardiol Rev* 21:101, 2013. *[Good overview of the pathogenesis and clinical issues relating to venous bypass grafting, germane to all graft stenosis].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 12

The Heart

Richard N. Mitchell • Andrew J. Connolly

CHAPTER CONTENTS

The human heart is a remarkably efficient, durable, and reliable pump, propelling more than 7500 L of blood through the body each day, and beating more than 40 million times a year—the wellspring for tissue oxygenation, nutrition, and waste removal. In utero, the heart and vasculature are the first fully functional organ system (at roughly 8 weeks of gestation). Without a vascular supply and a beating heart, further development cannot occur and fetal demise is inevitable. When the heart fails postnatally, the results are equally catastrophic. Cardiovascular disease (including coronary artery disease [CAD], stroke, and peripheral vascular disease) is the number-one cause of worldwide mortality. In the United States alone, cardiovascular disease accounts for roughly 1 in 4 of all deaths, totaling about 610,000 individuals each year—greater than the number of deaths caused by all forms of cancer combined.

This chapter begins with a brief review of the normal heart because most cardiac diseases manifest as structural and/or functional changes in one or more cardiac components. General principles underlying cardiac hypertrophy and failure—common end points of several of the different forms of heart disease—are also discussed, before exploring the major categories of cardiac disease: congenital heart abnormalities, ischemic heart disease (IHD), hypertensive heart disease, diseases of the cardiac valves, and primary myocardial disorders. The chapter concludes with a few comments about pericardial diseases and cardiac neoplasms, as well as cardiac transplantation and devices.

CARDIAC STRUCTURE AND SPECIALIZATIONS

Heart weight varies with body habitus, averaging approximately 0.4% to 0.5% of body weight (250 to 320 g in the average adult female and 300 to 360 g in the average adult male). Increased heart weight or ventricular thickness indicates hypertrophy, and an enlarged chamber size implies dilation; both reflect compensatory changes in response to volume and/or pressure overloads (see later). Increased cardiac weight or size (or both)—resulting from hypertrophy and/or dilation—is called cardiomegaly.

Myocardium

The pumping function of the heart occurs through coordinated contraction (during systole) and relaxation (during diastole) of cardiac myocytes (the myocardium). Left ventricular myocytes are arranged in a spiral circumferential orientation to generate vigorous coordinated waves of contraction spreading from the cardiac apex to the base of the heart. In contrast, right ventricular myocytes have a less structured organization, generating overall less robust contractile forces. Contraction is achieved by shortening of serial contractile elements (sarcomeres) within parallel myofibrils. Although the heart is primarily a pump, it is worth remembering that it also has other functions (e.g., endocrine). For example, atrial cardiomyocytes have cytoplasmic storage granules that contain atrial natriuretic peptide, and ventricular myocytes contain B-type natriuretic peptide. Both of these are protein hormones that are released in response to increased stretch; they both promote arterial vasodilation and stimulate renal salt and water elimination (natriuresis and diuresis).

Valves

The four cardiac valves—tricuspid, pulmonary, mitral, and aortic—maintain unidirectional blood flow. Valve function depends on the mobility, pliability, and structural integrity of the leaflets of the atrioventricular valves (tricuspid and mitral) or cusps of the semilunar valves (aortic and pulmonary).

The function of the semilunar valves depends on the integrity and coordinated movements of the cuspal attachments. Thus, dilation of the aortic root can result in valvular regurgitation. In contrast, the competence of the atrioventricular valves depends on the proper function not only of the leaflets but also the tendinous cords and the attached papillary muscles of the ventricular wall. Left ventricular dilation, a ruptured cord, or papillary muscle dysfunction can all interfere with mitral valve closure, causing valvular insufficiency.

Cardiac valves are lined by endothelium and share a similar, trilayered architecture:

- *Fibrosa layer.* A dense collagenous layer at the outflow surface, connected to the valvular supporting structures and providing mechanical integrity
- *Spongiosa layer.* A central core of loose connective tissue
- *Ventricularis or atrialis layer* (depending on which chamber it faces). A layer rich in elastin on the inflow surface, providing leaflet recoil

Because they are thin enough to be nourished by diffusion from the blood, normal leaflets and cusps have only scant blood vessels limited to the proximal portion of the valve. Valve endothelium also does not express ABO or histocompatibility antigens, so cryopreserved valvular tissues can be transplanted with relative impunity.

Conduction System

Coordinated contraction of the cardiac muscle depends on the initiation and rapid propagation of electrical impulses—accomplished through specialized myocytes in the conduction system. The frequency of electrical impulses is sensitive to neural inputs (e.g., vagal stimulation), extrinsic adrenergic agents (e.g., circulating adrenaline), hypoxia, and potassium concentration (i.e., hyperkalemia can block signal transmission altogether).

The components of the conduction system include the following:

- *Sinoatrial (SA) node* pacemaker, at the junction of the right atrial appendage and superior vena cava
- *Atrioventricular (AV) node,* located in the right atrium along the interatrial septum
- *Bundle of His (AV bundle),* connecting the right atrium to the ventricular septum
- *Right and left bundle branch divisions* that stimulate their respective ventricles via further arborization into the *Purkinje network*

The cells of the cardiac conduction system depolarize spontaneously, potentially enabling all of them to function as cardiac pacemakers. Because the normal rate of spontaneous depolarization in the SA node (60 to 100 beats/min) is faster than the other components, it normally sets the pace. However, if nodal tissues become dysfunctional, other cells in the conduction system can take over, generating, for example, a junctional escape rhythm (usually at a much slower rhythm). The AV node has a gatekeeper function; by delaying the transmission of signals from the atria to the ventricles, it ensures that atrial contraction precedes ventricular systole.

Blood Supply

Cardiac myocytes rely almost exclusively on oxidative phosphorylation for their energy needs. Besides a high density of mitochondria (20% to 30% of myocyte volume), myocardial energy generation also requires a constant supply of oxygenated blood—rendering myocardium extremely vulnerable to ischemia. Nutrients and oxygen are delivered via the coronary arteries, with ostia immediately distal to the aortic valve. These initially course along the external surface of the heart (epicardial coronary arteries) and then penetrate the myocardium (intramural arteries), subsequently branching into arterioles, and forming a rich arborizing vascular network so that each myocyte contacts roughly three capillaries.

The right and left coronary arteries function as end arteries, although anatomically most hearts have numerous intercoronary anastomoses (connections called collateral circulation). Blood flow to the myocardium occurs during ventricular diastole, after closure of the aortic valve, and

when the microcirculation is not compressed by cardiac contraction. At rest, diastole comprises approximately two-thirds of the cardiac cycle; with tachycardia (increased heart rate), the relative duration of diastole also shortens, thus potentially compromising cardiac perfusion.

Cardiac Regeneration

There is considerable interest in exploring the possibility of replacing damaged myocardium by inducing cardiac regeneration in vivo or implanting stem cell–derived cardiac cells. Although cardiac regeneration in metazoans such as newts and zebrafish is well described, the myocardium of mammals has a very low replicative potential after fetal and neonatal life, averaging less than 1% cardiomyocyte turnover per year in adult humans. Increasing evidence, however, indicates that cardiomyocyte proliferation can be augmented in mice. The potential for stimulating cardiac regeneration in vivo in humans is tantalizing because it could facilitate recovery of myocardial function after a host of injurious stimuli. Another area of vigorous investigation is ex vivo expansion and subsequent administration of stem cell–derived myocardial cells into areas of myocardial injury. Unfortunately, results thus far have been less than exciting. Implanted cells may show some cardiomyocyte differentiation, but the durability of this benefit has been limited, and they do not contribute significantly to the restoration of contractile force; moreover, failure to successfully integrate these cells into the conduction pathways of the host heart carries the very real risk of autonomous arrhythmic foci.

Effects of Aging on the Heart

Most forms of heart disease become more prevalent with each advancing decade. Consequently, as the average populations in high income countries get progressively older, aging-associated changes in the cardiovascular system become ever more significant (Table 12.1).

The size of the left ventricular cavity, particularly in the base-to-apex dimension, is reduced in later life; this volume change is exacerbated by systemic hypertension as the basal ventricular septum protrudes into the left ventricular outflow tract (so-called sigmoid septum). Compared with younger myocardium, the "elderly" heart typically has fewer myocytes (due to degenerative attrition) and increased connective tissue; octogenarians (and older) also frequently have deposition of extracellular amyloid (most often poorly catabolized transthyretin; Chapter 6) that stiffens the heart and reduces diastolic filling. Valvular aging changes are major contributors to significant valvular disease (see later), and progressive atherosclerosis, with a strong aging component (Chapter 11), is the major cause of IHD.

Table 12.1 Changes in the Aging Heart

Chambers
Increased left atrial cavity size
Decreased left ventricular cavity size
Sigmoid-shaped ventricular septum
Valves
Aortic valve calcific deposits
Mitral valve annular calcific deposits
Fibrous thickening of leaflets
Buckling of mitral leaflets toward the left atrium
Lambl excrescences
Epicardial Coronary Arteries
Tortuosity
Diminished compliance
Calcific deposits
Atherosclerotic plaque
Myocardium
Decreased mass
Increased subepicardial fat
Brown atrophy
Lipofuscin deposition
Basophilic degeneration
Amyloid deposits
Aorta
Dilated ascending aorta with rightward shift
Elongated (tortuous) thoracic aorta
Sinotubular junction calcific deposits
Elastic fragmentation and collagen accumulation
Atherosclerotic plaque

OVERVIEW OF CARDIAC PATHOPHYSIOLOGY

Although a host of diseases can affect the cardiovascular system, the pathophysiologic pathways that result in a "broken" heart distill down to six principal mechanisms:

- *Failure of the pump.* In the most common situation, the cardiac muscle contracts weakly and the chambers cannot empty properly—so-called systolic dysfunction. In some cases, the myocardium cannot relax sufficiently to permit ventricular filling, resulting in diastolic dysfunction.
- *Obstruction to flow.* Lesions that prevent valve opening (e.g., calcific aortic valve stenosis) or cause increased ventricular chamber pressures (e.g., systemic hypertension or aortic coarctation) can overwork the myocardium, which has to pump against the obstruction.
- *Regurgitant flow.* Valve pathology that allows backward flow of blood results in increased volume workload and may overwhelm the pumping capacity of the affected chambers.
- *Shunted flow.* Defects (congenital or acquired) that divert blood inappropriately from one chamber to another, or from one vessel to another, lead to pressure and volume overloads.
- *Disorders of cardiac conduction.* Uncoordinated cardiac impulses or blocked conduction pathways can cause arrhythmias that slow contractions or prevent effective pumping altogether.
- *Rupture of the heart or major vessel.* Loss of circulatory continuity (e.g., a gunshot wound through the thoracic aorta) may lead to massive blood loss, hypotensive shock, and death.

Most cardiovascular disease results from a complex interplay of genetics and environmental factors; these can disrupt signaling pathways that control morphogenesis,

affect myocyte survival after injury, or affect contractility or electrical conduction in the face of biomechanical stressors. Indeed, the pathogenesis of many congenital heart defects involves an underlying genetic abnormality whose expression is modified by environmental factors (see later). Moreover, genes that control the development of the heart may also regulate the response to various forms of injury including aging. Subtle polymorphisms can significantly affect the risk of many forms of heart disease, and, as discussed later, a number of adult-onset heart disorders have a fundamentally genetic basis. Thus, cardiovascular genetics provides an important window on the pathogenesis of heart disease, and molecular diagnoses are increasingly a critical part of its classification.

HEART FAILURE

Heart failure, often called congestive heart failure (CHF), is a common, usually progressive condition with a poor prognosis. Each year in the United States, CHF affects more than 5 million individuals (approximately 2% of the population), necessitating more than 1 million hospitalizations, and contributing to the death of nearly 300,000 people. Roughly one-half of patients die within 5 years of receiving a diagnosis of CHF, and 1 in 9 deaths in the United States include heart failure as a contributory cause.

Heart failure is defined as the condition in which a heart cannot pump blood to adequately meet the metabolic demands of peripheral tissues, or can do so only at elevated filling pressures. It is the common end stage of many forms of chronic heart disease, often emerging insidiously from the cumulative effects of chronic work overload (e.g., in valve disease or hypertension) or IHD (e.g., after myocardial infarction [MI] with heart damage). However, acute hemodynamic stresses, such as fluid overload, abrupt valvular dysfunction, or myocardial infarction, can all precipitate sudden CHF.

When cardiac workload increases or cardiac function is compromised, several physiologic mechanisms swing into action, and can at least initially maintain arterial pressure and organ perfusion:

- *Frank-Starling mechanism:* Increased filling volumes dilate the heart, thereby increasing actin-myosin cross-bridge formation, and enhancing contractility and stroke volume.
- *Activation of neurohumoral systems:* These augment heart function and/or regulate filling volumes and pressures (and many of the therapies for CHF affect these systems when they become maladaptive).
  - *Release of norepinephrine* by adrenergic nerves of the autonomic nervous system, elevating heart rate, augmenting myocardial contractility and increasing vascular resistance
  - *Activation of the renin-angiotensin-aldosterone system,* promoting water and salt retention (augmenting circulatory volume) and increasing vascular tone
  - *Release of atrial natriuretic peptide,* counterbalancing the renin-angiotensin-aldosterone system through diuresis and vascular smooth muscle relaxation
- *Myocardial adaptations*: In many pathologic states, heart failure is preceded by cardiac hypertrophy, a compensatory response to increased mechanical work. Ventricular remodeling is the general term applied to the collective molecular, cellular, and structural changes that occur in response to injury or altered ventricular loading.

Although such adaptive mechanisms can potentially maintain adequate cardiac output in the face of acute perturbations, their capacity to do so may ultimately be overwhelmed. Heart failure can result from progressive deterioration of myocardial contractile function (systolic dysfunction)—reflected as a decrease in ejection fraction (EF, the percentage of blood volume ejected from the ventricle during systole; normal is approximately 45% to 65%). Reduction in EF can occur with ischemic injury, inadequate adaptation to pressure or volume overload due to hypertension or valvular disease, or ventricular dilation. Increasingly, heart failure is recognized as resulting from an inability of the heart chamber to expand and fill sufficiently during diastole (diastolic dysfunction), for example, due to left ventricular hypertrophy, myocardial fibrosis, constrictive pericarditis, or amyloid deposition.

Cardiac Hypertrophy: Pathophysiology and Progression to Heart Failure

Sustained increase in mechanical work of either ventricle due to pressure overload, volume overload, or trophic signals (e.g., those mediated through the activation of β-adrenergic receptors) causes myocytes to increase in size (cellular hypertrophy); cumulatively, this increases the size and weight of the heart (Fig. 12.1). Hypertrophy requires increased protein synthesis to form additional sarcomeres, as well as increasing the numbers of mitochondria. Hypertrophic myocytes also have multiple or enlarged nuclei, attributable to increased DNA ploidy resulting from DNA replication in the absence of cell division.

The pattern of hypertrophy reflects the nature of the stimulus.

- In *pressure-overload hypertrophy* (e.g., due to hypertension or aortic stenosis), new sarcomeres are predominantly assembled in parallel to the long axes of cells, expanding the cross-sectional area of myocytes in ventricles and causing a concentric increase in wall thickness.
- In contrast, *volume-overload hypertrophy* (e.g., due to valvular regurgitation) is characterized by new sarcomeres being assembled in series within existing sarcomeres, leading primarily to ventricular dilation. As a result, in dilation due to volume overload, or dilation that accompanies failure of a previously pressure overloaded heart, the wall thickness may be increased, normal, or less than normal. Consequently, heart weight, rather than wall thickness, is the best measure of hypertrophy in dilated hearts.

Heart disease can lead to dramatic levels of cardiac hypertrophy. Patients with systemic hypertension, IHD, aortic stenosis, mitral regurgitation, or dilated cardiomyopathy frequently have heart weights double or triple the average, and aortic regurgitation or hypertrophic cardiomyopathy can produce heart weights threefold to fourfold greater than normal.

Important changes at the tissue and cell level occur with cardiac hypertrophy. Significantly, myocyte hypertrophy

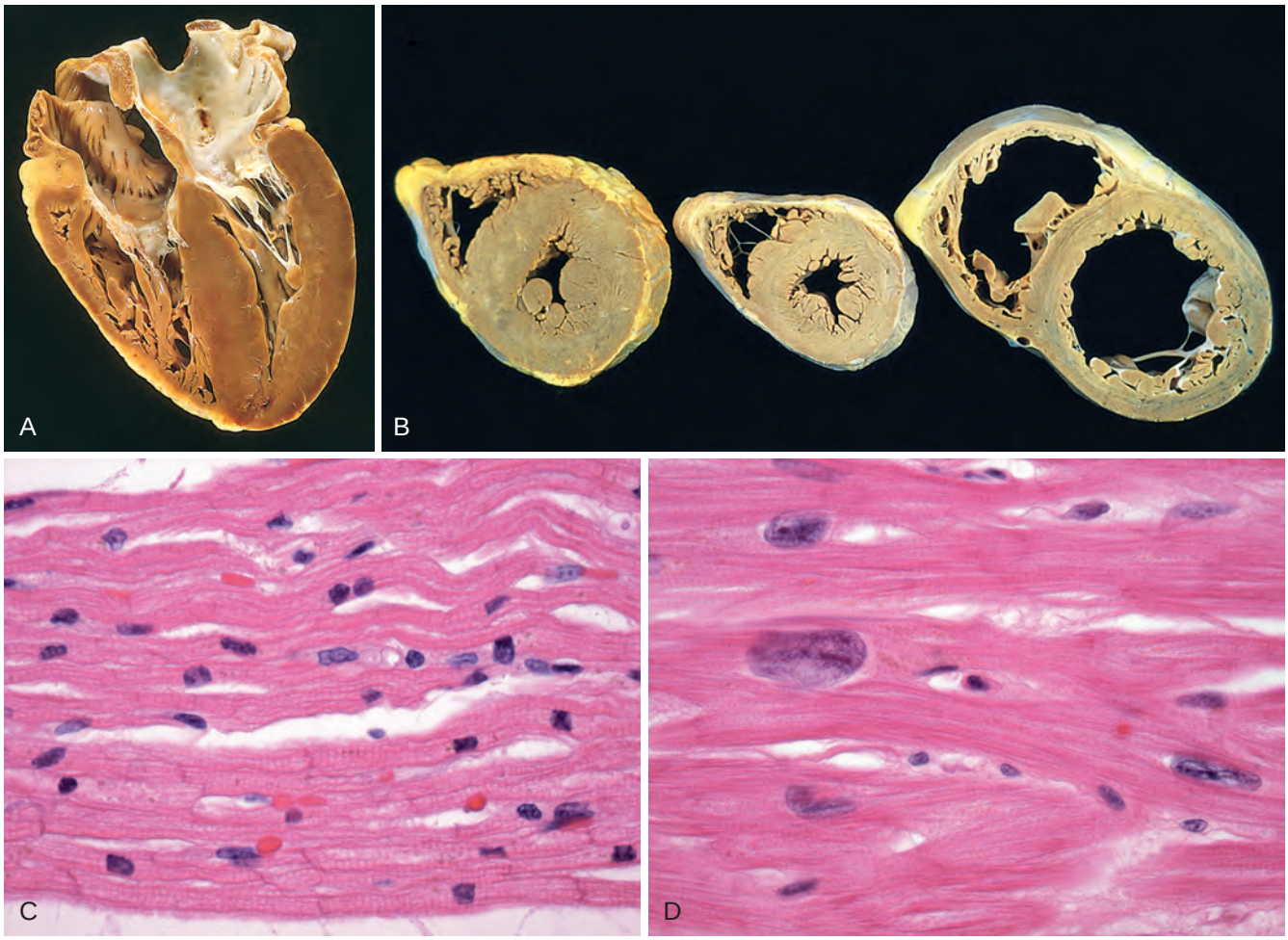


Figure 12.1 Left ventricular hypertrophy. (A) Pressure hypertrophy due to left ventricular outflow obstruction. The left ventricle is on the lower right in this apical four-chamber view of the heart. (B) Left ventricular hypertrophy with and without dilation, viewed in transverse heart sections. Compared with a normal heart *(center)*, the pressure-hypertrophied hearts *(left and in A)* have increased mass and a thick left ventricular wall, and the hypertrophied, dilated heart *(right)* has increased mass and an apparently normal wall thickness. (C) Normal myocardium. (D) Hypertrophied myocardium (C and D are photomicrographs at the same magnification). Note the increases in both cell size and nuclear size in the hypertrophied myocytes, and the interstitial cells remain small. (A and B, Reproduced with permission from Edwards WD: Cardiac anatomy and examination of cardiac specimens. In Emmanouilides GC, et al., editors: *Moss and Adams Heart Disease in Infants, Children, and Adolescents: Including the Fetus and Young Adults*, ed 5, Philadelphia, 1995, Williams & Wilkins, p 86.)

is not accompanied by a proportional increase in capillary numbers. As a result, the supply of oxygen and nutrients to the hypertrophied heart, particularly one undergoing pressure-overload hypertrophy, is more tenuous than in the normal heart. At the same time, cardiac hypertrophy is associated with heightened metabolic demands due to increases in mass, heart rate, and contractility (inotropic state, or force of contraction), all of which increase cardiac oxygen consumption. As a result of these changes, the hypertrophied heart is vulnerable to ischemia-related decompensation, which can evolve to cardiac failure. Adding insult to injury, hypertrophy is also typically accompanied by deposition of fibrous tissue (interstitial fibrosis), causing increased resistance to diastolic filling.

Molecular changes in hypertrophied cardiomyocytes include the expression of immediate-early genes (e.g., *FOS, JUN, MYC,* and *EGR1*) (Chapter 2) putatively driving cellular growth and altered protein expression; with prolonged hemodynamic overload, myocytes can even express genes usually only seen during fetal cardiac development (including fetal forms of myosin, natriuretic peptides, and collagen).

The proposed sequence of initially adaptive—and later harmful—events in response to increased cardiac work is summarized in Fig. 12.2. As illustrated, heart failure eventually supervenes. The degree of anatomic abnormality does not always reflect the severity of dysfunction; indeed, the gross appearance of the "failing heart" does not adequately convey the underlying structural, biochemical, and molecular basis for myocardial contractile failure. The hearts of patients with CHF are generally heavy and dilated, but can be relatively thin-walled, and histologically they exhibit variable degrees of myocyte hypertrophy. Loss of myocardial mass in the setting of infarction leads to work-related hypertrophy of the surrounding viable myocardium. In valvular heart disease, the increased pressure or volume overloads the myocardium globally. Increased heart mass owing to disease

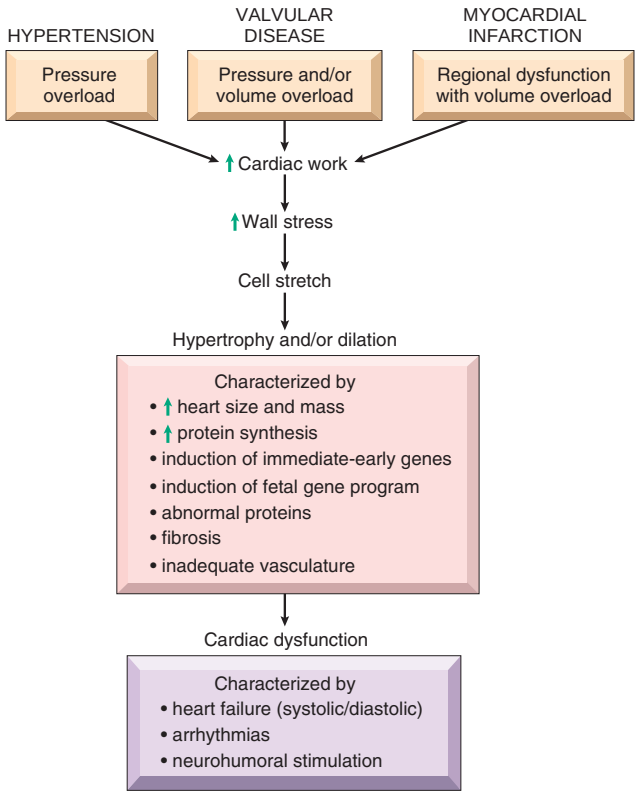


Figure 12.2 Schematic representation of the causes and consequences of cardiac hypertrophy.

is correlated with excess cardiac mortality and morbidity; indeed, cardiomegaly is an independent risk factor for sudden death.

In contrast to the pathologic hypertrophy associated with persistent mechanical stressors, regular exercise can promote potentially beneficial physiologic hypertrophy. Aerobic exercise (e.g., long-distance running) tends to be associated with volume-load hypertrophy accompanied by increases in capillary density (unlike other forms of hypertrophy); regular aerobic activity also decreases the resting heart rate and blood pressure—effects that are all beneficial. In comparison, purely static exercise (e.g., weight lifting) induces mild pressure hypertrophy (e.g., secondary to recurring Valsalva maneuvers) and less beneficial remodeling.

Whatever its basis, CHF is characterized by variable degrees of decreased cardiac output and tissue perfusion (forward failure), as well as pooling of blood in the venous capacitance system (backward failure); the latter may cause pulmonary edema, peripheral edema, or both. As a result, many of the significant clinical features and morphologic changes noted in CHF are actually secondary to disorders induced by hypoxia and congestion in noncardiac peripheral tissues.

The cardiovascular system is a closed circuit. Thus, although left-sided and right-sided failure can occur independently, failure of one side (particularly the left) often produces excessive strain on the other, terminating in global heart failure. Despite this interdependence, it is easiest to understand the pathology of heart failure by considering right- and left-sided heart failure separately.

Left-Sided Heart Failure

Left-sided heart failure is most often caused by the following:

- IHD
- Hypertension
- Aortic and mitral valvular diseases
- Primary myocardial diseases

The clinical and morphologic effects of left-sided CHF are a consequence of passive congestion (blood backing up in the pulmonary circulation), stasis of blood in the left-sided chambers, and inadequate perfusion of downstream tissues leading to organ dysfunction.

MORPHOLOGY

Heart. The findings depend on the disease process, ranging from myocardial infarcts, to stenotic or regurgitant valves, to intrinsic myocardial pathology. Except for failure caused by mitral valve stenosis or unusual restrictive cardiomyopathies (described later), the left ventricle is usually hypertrophied and often dilated, sometimes massively. Left ventricular diastolic dysfunction or dilation with mitral valve incompetence causes secondary dilation of the left atrium, increasing the risk of atrial fibrillation. This in turn results in stasis of blood, particularly in the atrial appendage, which is a common site of thrombus formation. The microscopic changes in the failing heart are nonspecific: variable degrees of myocyte hypertrophy and interstitial fibrosis.

Lungs. Pulmonary congestion and **edema** produce heavy, wet lungs, as described elsewhere (Chapters 4 and 15). Pulmonary changes—from mildest to most severe—include (1) perivascular and interstitial edema, particularly in the interlobular septa, (2) progressive edematous widening of alveolar septa, and (3) accumulation of edema fluid in the alveolar spaces. Extravasated red cells and plasma proteins in the alveoli are phagocytosed and digested by macrophages; the accumulated iron is stored as hemosiderin. These hemosiderin-laden macrophages (also known as **heart failure cells**) are telltale signs of previous episodes of pulmonary edema. Pleural effusions (typically serous) arise from elevated pleural capillary and lymphatic pressure and the resultant transudation of fluid into the pleural cavities.

Early left-sided heart failure symptoms are related to pulmonary congestion and edema. Initially, cough and dyspnea (breathlessness) may occur only with exertion. As CHF progresses, worsening pulmonary edema may cause orthopnea (dyspnea when supine, relieved by sitting or standing) or paroxysmal nocturnal dyspnea (dyspnea usually occurring at night that is so severe that it induces a feeling of suffocation). Dyspnea at rest may follow. The respiratory symptoms are characteristically associated with fine rales at the lung bases, caused when edematous pulmonary alveoli snap open during inspiration. Other manifestations of left ventricular failure include an enlarged heart (cardiomegaly, apparent on imaging), tachycardia, a third heart sound due to volume overload (S3), or a fourth heart sound (S4) due to increased myocardial stiffness. If heart failure is associated with progressive ventricular dilation, the papillary muscles are displaced outward, causing mitral regurgitation. Subsequent chronic dilation of the left atrium can cause atrial

fibrillation, and such uncoordinated, chaotic atrial contractions reduce the atrial contribution to ventricular filling, thus reducing the ventricular stroke volume.

In moderate CHF, a reduced ejection fraction leads to diminished renal perfusion, causing activation of the renin-angiotensin-aldosterone system as a compensatory mechanism to correct the "perceived" hypotension. This leads to salt and water retention, with expansion of the interstitial and intravascular fluid volumes (Chapters 4 and 11) exacerbating the ongoing pulmonary edema. If the hypoperfusion of the kidney becomes sufficiently severe, impaired excretion of nitrogenous products may cause azotemia (called prerenal azotemia; Chapter 20). In far-advanced CHF, cerebral hypoperfusion can give rise to hypoxic encephalopathy (Chapter 28), with irritability, loss of attention span, and restlessness that can progress to stupor and coma with ischemic cerebral injury.

Left-sided heart failure can be divided into systolic and diastolic failure:

- *Systolic failure* is defined by insufficient ejection fraction (pump failure) and can be caused by any of the many disorders that damage or derange the contractile function of the left ventricle.
- In *diastolic failure,* the left ventricle is abnormally stiff and cannot relax during diastole. Thus, although cardiac function is relatively preserved at rest, the heart is unable to increase its output in response to increases in the metabolic demands of peripheral tissues (e.g., during exercise). Moreover, because the left ventricle cannot expand normally, any increase in filling pressure is immediately transferred back into the pulmonary circulation, producing pulmonary edema. Hypertension is the most common underlying etiology; diabetes mellitus, obesity, and bilateral renal artery stenosis can also be causal. Reduced left ventricular relaxation may stem from myocardial fibrosis (e.g., in cardiomyopathies and IHD) or infiltrative disorders associated with restrictive cardiomyopathies (e.g., cardiac amyloidosis). Diastolic failure may appear in older patients without any known predisposing factors, possibly as an exaggeration of the normal stiffening of the heart with age. Constrictive pericarditis (discussed later) can also limit myocardial relaxation and therefore mimics primary diastolic dysfunction.

Right-Sided Heart Failure

Right-sided heart failure is most commonly caused by left-sided heart failure, as any increase in pressure in the pulmonary circulation from left-sided failure inevitably burdens the right side of the heart. Consequently, the causes of right-sided heart failure include all the etiologies for left-sided heart failure. Isolated right-sided heart failure is infrequent and typically occurs in patients with one of a variety of disorders affecting the lungs; hence it is often referred to as cor pulmonale. Besides parenchymal lung diseases, cor pulmonale can also arise secondary to disorders that affect the pulmonary vasculature, for example, primary pulmonary hypertension (Chapter 15), recurrent pulmonary thromboembolism (Chapter 4), or conditions that cause pulmonary vasoconstriction (obstructive sleep apnea, altitude sickness). The common feature of these disorders is pulmonary hypertension (discussed later), which results in hypertrophy and dilation of the right side of the heart. In extreme cases, leftward bulging of the interventricular septum can even cause left ventricular dysfunction. The major morphologic and clinical effects of primary right-sided heart failure differ from those of left-sided heart failure in that pulmonary congestion is minimal while engorgement of the systemic and portal venous systems is pronounced.

MORPHOLOGY

Heart. As in left-heart failure, the cardiac morphology varies with cause. Rarely, structural defects such as tricuspid or pulmonary valvular abnormalities or endocardial fibrosis (as in carcinoid heart disease) may be present. However, because isolated right heart failure is most often caused by lung disease, most cases exhibit only hypertrophy and dilation of the right atrium and ventricle.

Liver and Portal System. Congestion of the hepatic and portal vessels may produce pathologic changes in the liver, the spleen, and the gastrointestinal tract. The liver is usually increased in size and weight **(congestive hepatomegaly)** caused by **passive congestion,** greatest around the central veins (Chapter 4). Grossly, this is reflected as congested red-brown pericentral zones, with relatively normal-colored tan periportal regions, producing the characteristic "nutmeg liver" appearance (Chapter 4). In some instances, especially when left-sided heart failure with hypoperfusion is also present, severe centrilobular hypoxia produces **centrilobular necrosis.** With longstanding severe right-sided heart failure, the central areas can become fibrotic, eventually culminating in **cardiac cirrhosis** (Chapter 18). Portal venous hypertension also causes enlargement of the spleen with platelet sequestration **(congestive splenomegaly)** and can also contribute to chronic congestion and edema of the bowel wall. The latter may be sufficiently severe as to interfere with nutrient (and/or drug) absorption.

Pleural, Pericardial, and Peritoneal Spaces. Systemic venous congestion can lead to fluid accumulation **(effusions)** in the pleural, pericardial, or peritoneal spaces (a peritoneal effusion is also called **ascites**). Large pleural effusions can impact lung inflation, causing atelectasis, and substantial ascites can also limit diaphragmatic excursion, causing dyspnea on a purely mechanical basis.

Subcutaneous Tissues. Edema of the peripheral and dependent portions of the body, especially foot/ankle (pedal) and pretibial edema, is a hallmark of right-sided heart failure. In chronically bedridden patients, presacral edema may predominate. Generalized massive edema **(anasarca)** can also occur.

The kidney and the brain are also prominently affected in right-sided heart failure. Renal congestion is more marked with right-sided than with left-sided heart failure, leading to greater fluid retention and peripheral edema, and more pronounced azotemia. Venous congestion and hypoxia of the central nervous system can also produce deficits of mental function akin to those seen in left-sided heart failure with poor systemic perfusion.

Although we have discussed right and left heart failure separately, it is again worth emphasizing that in many cases of chronic cardiac decompensation, patients with biventricular CHF have symptoms reflecting both right-sided and left-sided heart failure. Besides a careful history and physical

examination, serum levels of B-type (or brain) natriuretic factor (BNP) have become a popular tool to quantitatively assess the extent of CHF. Recall that BNP is released by ventricular cardiomyocytes during increased wall stress; a low value has a high negative predictive value for CHF. Echocardiography is also an extremely valuable tool in following patients with CHF, providing a measure of ejection fraction, wall motion, valvular function, and possible mural thrombosis.

Treatment for CHF is initially focused on correcting any underlying cause, for example, a valvular defect or inadequate cardiac perfusion. Beyond that, the clinical approach includes salt restriction or pharmacologic agents that variously reduce volume overload (e.g., diuretics), increase myocardial contractility (so-called positive inotropes), or reduce afterload (via adrenergic blockade or inhibitors of angiotensin-converting enzymes [ACE]). Although many of these medications provide benefit through effects on neurohumoral pathways, ACE inhibitors also limit myocyte hypertrophy and cardiac remodeling. Although cardiac resynchronization therapy (exogenous pacing of both the right and left ventricles) and mechanical ventricular assist devices (VADs, discussed later) have also been added to the cardiologist's armamentarium, CHF remains a serious cause of human morbidity and mortality.

KEY CONCEPTS

HEART FAILURE

- Heart failure occurs when the heart is unable to provide adequate perfusion to meet the metabolic requirements of peripheral tissues; inadequate cardiac output is usually accompanied by increased congestion of the venous circulation.
- Left-sided heart failure is most commonly due to IHD, systemic hypertension, mitral or aortic valve disease, and primary diseases of the myocardium; symptoms are mainly a consequence of pulmonary congestion and edema, although systemic hypoperfusion can cause secondary renal and cerebral dysfunction.
- Right-sided heart failure is most often due to left heart failure, and less commonly to primary pulmonary disorders; symptoms are chiefly related to peripheral edema and visceral congestion.

CARDIAC DEVELOPMENT

The heart is a mechanical organ that generates pulsatile blood within just 3 weeks after fertilization. It is therefore likely that hemodynamic forces play an important role in cardiac development, just as they influence adaptations in the adult heart such as hypertrophy and dilation.

The diverse malformations seen in congenital heart disease (CHD) are caused by errors that occur during the complex migration and folding that constitutes cardiac morphogenesis. Derived from cells in the lateral mesoderm, the earliest cardiac precursors move to the midline in two migratory waves (called the first and second heart fields) within the first 15 days of fetal development. Although these are multipotent progenitor cells that can produce all of the major cell types of the heart (endocardium, myocardium, and smooth muscle cells), they rapidly assume distinct fates; cells in the first wave largely populate the developing left ventricle, whereas those in the second wave become the outflow tract, right ventricle, and most of the atria. Thus, defects in one or the other anlage can explain some of the CHDs involving discrete structures. By day 20 of development, the nascent heart has become a beating tube, which begins to form the basic heart chambers roughly 8 days later. At about the same time, (1) neural crest–derived cells migrate into the outflow tract, where they participate in the septation of the aortic and pulmonic outflow tracts and the formation of the aortic arch; and (2) interstitial connective tissue expands to become definitive endocardial cushions that will become the future atrioventricular canals and outflow tracts. By day 50, further septation of the ventricles, atria, and atrioventricular valves produces a four-chambered heart.

Proper orchestration of these remarkable transformations depends on a network of transcription factors that are regulated by a number of signaling pathways, particularly the Wnt, hedgehog, vascular endothelial growth factor (VEGF), bone morphogenetic protein, transforming growth factor-β (TGF-β), fibroblast growth factor, and Notch pathways (Chapter 1). It is not too surprising then that many of the inherited defects that affect heart development involve genes that encode transcription factors; these typically cause partial loss of function and are autosomal dominant (discussed later). In addition, specific micro-RNAs play critical roles in cardiac development by coordinating patterns and levels of transcription factor expression.

CONGENITAL HEART DISEASE

CHD refers to abnormalities of the heart or great vessels that are present at birth. Most CHD arises from faulty embryogenesis during gestational weeks 3 to 8, when major cardiovascular structures form and begin to function. The most severe anomalies preclude intrauterine survival, and significant heart malformations are common among stillborn infants. On the other hand, circumscribed defects affecting discrete regions of the heart or individual chambers can be compatible with live birth. In this latter category are the following:

- *Septal defects,* or "holes in the heart," including atrial septal defects (ASDs) or ventricular septal defects (VSDs)
- *Stenotic lesions,* either at the level of valves, or the entire cardiac chamber as in hypoplastic left heart syndrome
- *Outflow tract anomalies* including inappropriate routing of the great vessels from the ventricles, or anomalous coronary arteries

Such "tolerated" forms of CHD usually produce clinically important manifestations only after birth—uncovered by the transition from fetal to perinatal circulation; roughly one-half will be diagnosed in the first year of life, although some milder forms may not be discovered until adulthood (e.g., ASD).

Incidence

The incidence of CHD depends on what is counted as a defect. Thus, if echocardiography is performed routinely on neonates, small muscular VSDs or ASDs are detected in over 5% of live births. However, these typically close

Table 12.2 Frequencies of Congenital Cardiac Malformations[a]

Malformation	Incidence per Million Live Births	%
Ventricular septal defect	4482	42
Atrial septal defect	1043	10
Pulmonary stenosis	836	8
Patent ductus arteriosus	781	7
Tetralogy of Fallot	577	5
Coarctation of the aorta	492	5
Atrioventricular septal defect	396	4
Aortic stenosis	388	4
Transposition of the great arteries	388	4
Truncus arteriosus	136	1
Total anomalous pulmonary venous connection	120	1
Tricuspid atresia	118	1
Total	9757	

[a]Presented as upper quartile of 44 published studies. Percentages do not add up to 100% because of rounding. Does not include bicuspid aortic valves.

Data from: Hoffman JI, Kaplan S: The incidence of congenital heart disease, *J Am Coll Cardiol* 39(12):1890–1900, 2002.

spontaneously in the first year of life, and they probably should not be tallied with the burden of CHD. Similarly, bicuspid aortic valve—with an incidence of 1% to 2%—clearly persists beyond infancy, but often has modest manifestations and may not become evident until late adulthood. If the accounting is restricted to more serious defects, the worldwide incidence of congenital cardiovascular malformations is slightly less than 1%—still ranking CHD among the most prevalent birth defects. Twelve disorders account for about 85% of cases; their frequencies are listed in Table 12.2.

The number of individuals who survive into adulthood with CHD is increasing rapidly and is estimated at nearly 1.5 million people in the United States alone. Many have benefited from surgical advances that increasingly permit early postnatal repair of structural defects. In some cases, however, surgical intervention fails to restore complete normalcy; patients may have already sustained pulmonary or myocardial changes that are no longer reversible, or conversely, may suffer from arrhythmias due to surgical scarring. Other factors that impact the long-term outcome include complications associated with the use of prosthetic materials and devices (e.g., substitute valves or myocardial patches), and the cardiovascular stressors associated with childbearing that may tip a repaired heart into failure.

Etiology and Pathogenesis

Environmental exposures (e.g., congenital rubella infection, teratogens—including some therapeutic drugs, and gestational diabetes) and genetic factors are the best characterized causes but still account for a minority of CHD cases. Nutritional factors can also influence risk; folate supplementation during early pregnancy reduces CHD incidence.

Genetic factors include specific loci implicated in familial forms of CHD and certain chromosomal abnormalities (e.g., trisomies 13, 15, 18, and 21, and monosomy X/Turner syndrome). Indeed, the most common known genetic cause of CHD is trisomy 21 (Down syndrome); roughly 40% of patients with Down syndrome have one or more heart defects, most often affecting structures derived from the second migratory wave of cells (e.g., the atrioventricular septae). The mechanisms by which aneuploidy causes CHD likely involve the dysregulated expression of multiple genes.

A notable example of a small chromosomal lesion causing CHD is deletion of chromosome 22q11.2, occurring in patients with DiGeorge syndrome. In this syndrome, the fourth branchial arch and the derivatives of the third and fourth pharyngeal pouches (which contribute to the formation of the thymus, parathyroid glands, and heart) develop abnormally. Of the 30 or so genes present on this chromosome segment, deletion of the *TBX1* transcription factor gene is probably the culprit lesion. TBX1 regulates neural crest migration, as well as the expansion of cardiac progenitors in the second migratory wave. Interestingly, deletions in this region are also associated with mental illness, including schizophrenia.

In the case of single-gene mutations, the affected genes encode proteins belonging to several different functional classes (Table 12.3); as mentioned earlier, many of these involve transcription factors. Because affected patients are heterozygous for such mutations, it follows that a 50% reduction in the activity of these factors (or even less) may

Table 12.3 Selected Examples of Gene Defects Associated With Congenital Heart Disease[a]

Disorder	Gene(s)	Gene Product Function
Nonsyndromic		
ASD or conduction defects	*NKX2.5*	Transcription factor
ASD or VSD	*GATA4*	Transcription factor
Tetralogy of Fallot	*ZFPM2 or NKX2.5*	Transcription factors
Syndromic[b]		
Alagille syndrome—pulmonary artery stenosis or tetralogy of Fallot	*JAG1 or NOTCH2*	Signaling proteins or receptors
Char syndrome—PDA	*TFAP2B*	Transcription factor
CHARGE syndrome—ASD, VSD, PDA, or hypoplastic right side of the heart	*CHD7*	Helicase-binding protein
DiGeorge syndrome—ASD, VSD, or outflow tract obstruction	*TBX1*	Transcription factor
Holt-Oram syndrome—ASD, VSD, or conduction defect	*TBX5*	Transcription factor
Noonan syndrome—pulmonary valve stenosis, VSD, or hypertrophic cardiomyopathy	*PTPN11, KRAS, SOS1*	Signaling proteins

ASD, Atrial septal defect; *CHARGE*, posterior coloboma, heart defect, choanal atresia, retardation, genital and ear anomalies; *PDA*, patent ductus arteriosus; *VSD*, ventricular septal defect.

[a]Different mutations can cause the same phenotype, and mutations in some genes can cause multiple phenotypes (e.g., *NKX2.5*). Many of these congenital lesions also can occur sporadically, without specific genetic mutation.

[b]Only the cardiac manifestations of the syndrome are listed; the other skeletal, facial, neurologic, and visceral changes are not.

be sufficient to derange cardiac development. Even relatively minor decrements in activity of particular genes can result in significant defects. Thus, transient environmental stresses during the first trimester of pregnancy that alter the synthesis or activity of these same genes can conceivably lead to acquired defects that mimic those produced by heritable mutations. In addition, many of the transcription factors interact in large protein complexes, providing a rationale for why mutations in any one of several genes can produce similar defects. Thus, GATA4, TBX5, and NKX2-5, three transcription factors that are mutated in some patients with atrial and ventricular septal defects, all bind to one another and co-regulate the expression of target genes required for proper cardiac development.

Other single-gene mutations can alter structural proteins or affect signaling pathway molecules. Thus, mutations in genes encoding various components of the Notch pathway (Chapter 1) are associated with a variety of congenital heart defects, including bicuspid aortic valve (*NOTCH1*, discussed later) and tetralogy of Fallot (*JAG1* and *NOTCH2*). As described in Chapter 5 and 11, fibrillin mutations underlie Marfan syndrome—associated with valvular defects and aortic aneurysms. Although fibrillin is an important structural protein in the extracellular matrix (ECM), it is also an important negative regulator of TGF-β signaling, and hyperactive TGF-β signaling contributes to the cardiovascular abnormalities in Marfan syndrome and the related Loeys-Dietz syndrome.

Despite these advances, our understanding of the mechanisms underlying CHD remains rudimentary; the precise cause is unknown in almost 90% of cases. Most affected patients have no identifiable genetic risk, and even in those that do, the nature and severity of the defect are highly variable. It is likely that most forms of CHD arise by interactions of environmental factors with multiple genes. Guidelines for genetic testing are complex and evolving but suggest that genetic testing may be useful in patients with a broader noncardiac syndrome or a family history of CHD.

Clinical Features

Most of the various structural anomalies in CHD can be organized into three major categories according to the major functional abnormalities they cause:

- *Left-to-right shunt*
- *Right-to-left shunt*
- *Obstruction*

A shunt is an abnormal communication between chambers or blood vessels. Abnormal channels permit blood flow down pressure gradients from the left (systemic) side to the right (pulmonary) side of the circulation or vice versa. When blood from the right side of the circulation flows directly into the left side (right-to-left shunt), hypoxemia and *cyanosis* (a dusky blueness of the skin and mucous membranes) result because the pulmonary circulation is bypassed and poorly oxygenated venous blood shunts directly into the systemic arterial supply. In addition, right-to-left shunts can allow emboli from the peripheral veins to bypass the lungs and directly enter the systemic circulation (paradoxical embolism). Severe, long-standing hypoxia/cyanosis also causes increased numbers of circulating red blood cells (polycythemia), as well as a peculiar distal blunting and enlargement ("clubbing") of the tips of the fingers and toes that can include bony changes (called hypertrophic osteoarthropathy). The most important causes of right-to-left shunts are tetralogy of Fallot (TOF), transposition of the great arteries (TGA), persistent truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection.

In contrast, left-to-right shunts (e.g., ASD, VSD, and patent ductus arteriosus [PDA]) increase pulmonary blood flow but are not initially associated with cyanosis. However, left-to-right shunts chronically elevate both volume and pressure in the normally low-pressure, low-resistance pulmonary circulation. To maintain relatively normal distal pulmonary capillary and venous pressures, the muscular pulmonary arteries (<1 mm in diameter) initially respond by undergoing medial hypertrophy and vasoconstriction. However, prolonged pulmonary arterial vasoconstriction stimulates the development of irreversible obstructive intimal lesions analogous to the arteriolar changes seen in systemic hypertension; pulmonary arteries can even develop frank atherosclerotic lesions (Chapter 11). The right ventricle also responds to the pulmonary vascular changes by undergoing progressive hypertrophy. Eventually, pulmonary vascular resistance approaches systemic levels, and the original left-to-right shunt becomes a right-to-left shunt that introduces poorly oxygenated blood into the systemic circulation *(Eisenmenger syndrome)*.

Once irreversible pulmonary hypertension develops, the structural defects of CHD are considered irreparable; subsequent right heart failure can lead to the patient's death unless combined heart-lung transplantation can be performed. This provides the rationale for early intervention to close significant left-to-right shunts.

Obstructive CHD occurs when there is abnormal narrowing of chambers, valves, or blood vessels; these include coarctation of the aorta, aortic valvular stenosis, and pulmonary valvular stenosis. A complete obstruction is called an atresia. In some disorders (e.g., TOF), an obstruction (pulmonary stenosis) and a shunt (right-to-left through a VSD) are both present.

The altered hemodynamics of CHD usually cause cardiac dilation or hypertrophy (or both). However, some defects induce a decrease in the volume and muscle mass of a cardiac chamber; this is called hypoplasia if it occurs before birth and atrophy if it develops postnatally.

Left-to-Right Shunts

Left-to-right shunts are the most common CHD; these include ASD, VSD, and PDA (Fig. 12.3). ASDs typically increase only right ventricular and pulmonary outflow volumes, and VSDs and PDA cause both increased pulmonary blood flow and pressure. Depending on their size and location, manifestations of these shunts range in severity from no symptoms at all to fulminant heart failure.

Atrial Septal Defect

ASDs are abnormal, fixed openings in the atrial septum caused by incomplete tissue formation that allows communication of blood between the left and right atria; ASDs are usually asymptomatic until adulthood (see Table 12.2 and Fig. 12.3*A*). ASD should not be confused with patent foramen ovale (PFO; see later), which represents the postnatal

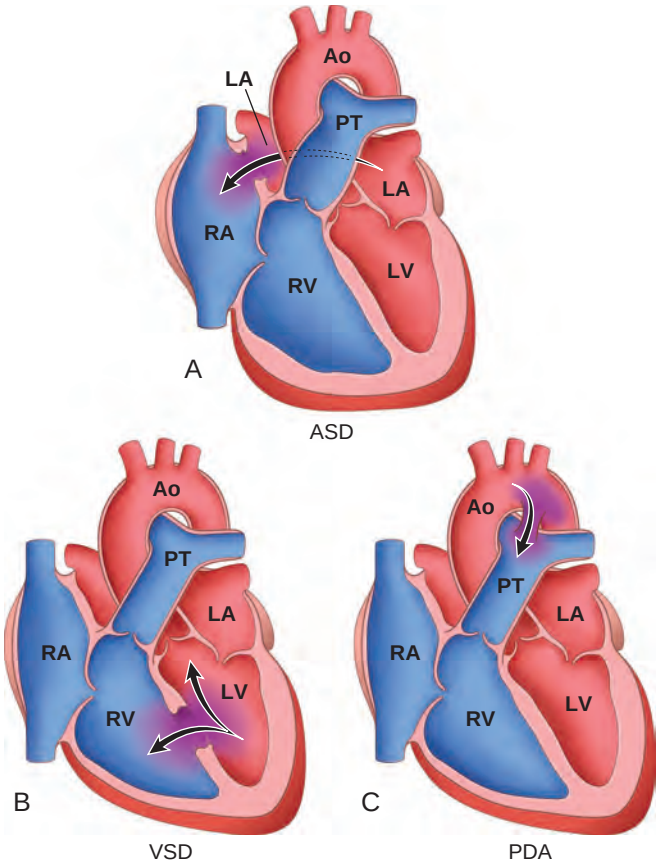


Figure 12.3 Common congenital left-to-right shunts (*arrows* indicate the direction of blood flow). (A) Atrial septal defect *(ASD).* (B) Ventricular septal defect *(VSD).* With VSD, the shunt is left-to-right, and the pressures are the same in both ventricles. Pressure hypertrophy of the right ventricle and volume hypertrophy of the left ventricle are generally present. (C) Patent ductus arteriosus *(PDA). Ao,* Aorta; *LA,* left atrium; *LV,* left ventricle; *PT,* pulmonary trunk; *RA,* right atrium; *RV,* right ventricle.

failure to close a foramen (channel) that is part of normal development. A brief summary of the developmental stages of the atrial septum follows:

- The septum primum is a crescent-shaped membranous ingrowth that sits posteriorly between the right and left atria and partially separates them; the remaining transient anterior opening, called the *ostium primum,* allows movement of blood from the right to left atrium during early fetal development.
- Before the growing septum primum completely obliterates the ostium primum, it develops a second posterior opening called the ostium secundum.
- The septum secundum is a subsequent membranous ingrowth located to the right and anterior of the septum primum.
- The septum secundum grows to cover the ostium secundum, leaving only a small channel called the foramen ovale that is continuous with the ostium secundum—the foramen ovale/ostium secundum permits continued right-to-left shunting of blood during intrauterine development.

This flap of septum secundum tissue opens and closes in response to pressure gradients between the left and right atria; the valve opens only when the pressure is greater in the right atrium. In fetal life, the lungs are nonfunctional, and the pressure in the pulmonary circulation is greater than that of the systemic circulation; thus, the right atrium is under higher pressures than the left atrium, and the valve of the foramen ovale is normally open. At birth, with lung expansion, the pulmonary vascular pressures drop, and the right atrial pressures fall below those in the left atrium. As a result, the valve of the foramen ovale closes—and is usually permanently sealed before adulthood (see later).

MORPHOLOGY

ASDs are classified according to their location. **Secundum ASD** (90% of all ASDs) result from a deficient septum secundum formation near the center of the atrial septum. These are usually not associated with other anomalies, may be of any size, and can be multiple or fenestrated. **Primum anomalies** (5% of ASD) occur adjacent to the AV valves and are often associated with AV valve abnormalities and/or a VSD. **Sinus venosus defects** (5%) are located near the entrance of the superior vena cava and can be associated with anomalous pulmonary venous return to the right atrium.

Clinical Features

ASDs are usually asymptomatic until adulthood. Although VSDs are more common, most close spontaneously. Consequently, ASDs—which are less likely to close spontaneously—are the most common defects to be diagnosed in adults. ASDs result in a left-to-right shunt, largely because pulmonary vascular resistance is considerably less than systemic vascular resistance and because the compliance (distensibility) of the right ventricle is much greater than that of the left. The resulting pulmonary flow volumes may be two to eight times normal. A murmur is often present as a result of excessive flow through the pulmonary valve and/or through the ASD. Despite the right-sided volume overload, ASDs are generally well tolerated and usually do not become symptomatic before 30 years of age; irreversible pulmonary hypertension is unusual.

Surgical or intravascular ASD closure is performed to preempt the development of heart failure, paradoxical embolization, and irreversible pulmonary vascular disease. Mortality is low, and postoperative survival is comparable to that for an unaffected population.

Patent Foramen Ovale

The foramen ovale closes permanently in approximately 80% of people by 2 years of age. However, in the remaining 20%, the unsealed flap can open if right-sided pressures become elevated. Thus, sustained pulmonary hypertension or even transient increases in right-sided pressures (e.g., during a bowel movement, coughing, or sneezing) can produce brief periods of right-to-left shunting, with the possibility of paradoxical embolism.

Ventricular Septal Defect

VSDs are incomplete closures of the ventricular septum, allowing free communication of blood between the left to right ventricles; they are the most common form of CHD (see Table 12.2 and Fig. 12.3*B*).

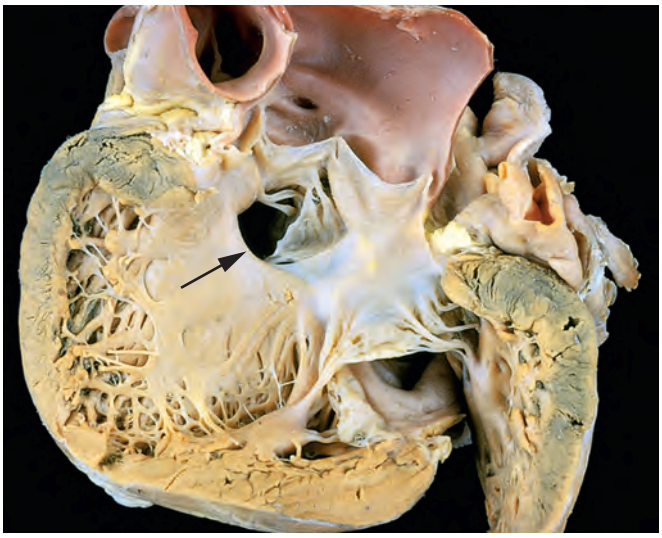


Figure 12.4 A membranous type ventricular septal defect *(arrow)* just proximal to the aortic valve. (Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, Minn.)

MORPHOLOGY

VSDs are classified according to their location and magnitude. About 90% occur in the region of the membranous interventricular septum (**membranous VSD**; Fig. 12.4), and the majority are 2 to 3 cm in diameter. The remaining 10% occur below the pulmonary valve **(infundibular VSD)** or within the muscular septum. Although most VSDs are single, those in the muscular septum may be multiple.

Clinical Features

Most VSDs that clinically manifest in the pediatric age group are associated with other congenital cardiac anomalies such as TOF; only 20% to 30% are isolated. Conversely, if a VSD is first detected only in an adult, it is usually an isolated defect. The functional consequences of a VSD depend on the size of the defect and whether there are associated right-sided malformations. Thus, large VSDs cause difficulties virtually from birth; smaller lesions are generally well tolerated for years and may not be recognized until much later in life. Moreover, approximately 50% of small muscular VSDs close spontaneously. Large defects are usually membranous or infundibular, and they generally cause significant left-to-right shunting, leading to early right ventricular hypertrophy and pulmonary hypertension. Over time, large unclosed VSDs almost universally lead to irreversible pulmonary vascular disease, ultimately resulting in shunt reversal and cyanosis. Surgical or catheter-based closure of asymptomatic VSD is generally delayed beyond infancy, in hope of spontaneous closure. Early correction, however, must be performed for large defects to prevent the development of irreversible obstructive pulmonary vascular disease.

Patent Ductus Arteriosus

The ductus arteriosus arises from the pulmonary artery and joins the aorta just distal to the origin of the left subclavian artery. During intrauterine life, it permits blood flow from the pulmonary artery to the aorta, thereby bypassing the unoxygenated lungs. In healthy term infants, the ductus constricts and is functionally closed within 1 to 2 days of birth; this occurs in response to increased arterial oxygenation, decreased pulmonary vascular resistance, and declining local levels of prostaglandin E_2. Complete structural obliteration occurs within the first few months of extrauterine life leaving behind the ligamentum arteriosum. Ductal closure is often delayed (or even absent) in infants with hypoxia (due to respiratory distress or heart disease) or when other congenital defects are present, particularly VSDs that increase pulmonary vascular pressures. PDAs account for about 7% of cases of CHD (see Table 12.2 and Fig. 12.3C), and 90% of these are isolated defects.

PDA produces a characteristic, continuous, harsh "machinery-like" murmur. The clinical impact of a PDA depends on its diameter and the cardiovascular status of the individual. PDA is usually asymptomatic at birth, and a narrow PDA may have no effect on the child's growth and development. Because the shunt is initially left-to-right, there is no cyanosis. However, with large shunts, the additional volume and pressure overloads eventually produce obstructive changes in small pulmonary arteries, leading to reversal of flow and its associated consequences. In general, isolated PDA should be closed as early in life as is feasible; therapy includes prostaglandin synthesis inhibitors and possibly percutaneous or surgical interventions. Conversely, preservation of ductal patency (by administering prostaglandin E1) may be life-saving for infants with various congenital malformations that obstruct the pulmonary or systemic outflow tracts. In CHD with aortic valve or pulmonary valve atresia, for example, a PDA may provide the entire systemic blood flow or pulmonary blood flow, respectively.

Right-to-Left Shunts

The diseases in this group cause cyanosis early in postnatal life (cyanotic CHD). TOF, the most common in this group, and TGA are illustrated schematically in Fig. 12.5. The others include persistent truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection. Note that the names of all of these conditions start with a *T*.

Tetralogy of Fallot

The four cardinal features of TOF are (1) VSD, (2) obstruction of the right ventricular outflow tract (subpulmonic stenosis), (3) an aorta that overrides the VSD, and (4) right ventricular hypertrophy (see Fig. 12.5A). The first three features result embryologically from anterosuperior displacement of the infundibular septum, and the right ventricular hypertrophy is a secondary consequence of the pressure overload.

MORPHOLOGY

The heart is typically enlarged and is classically "boot-shaped" due to marked right ventricular hypertrophy. The VSD is usually large with the aortic valve at the superior border, thereby overriding the defect and both ventricular chambers. The obstruction to right ventricular outflow is most often due to narrowing of the infundibulum (subpulmonic stenosis) but can be accompanied

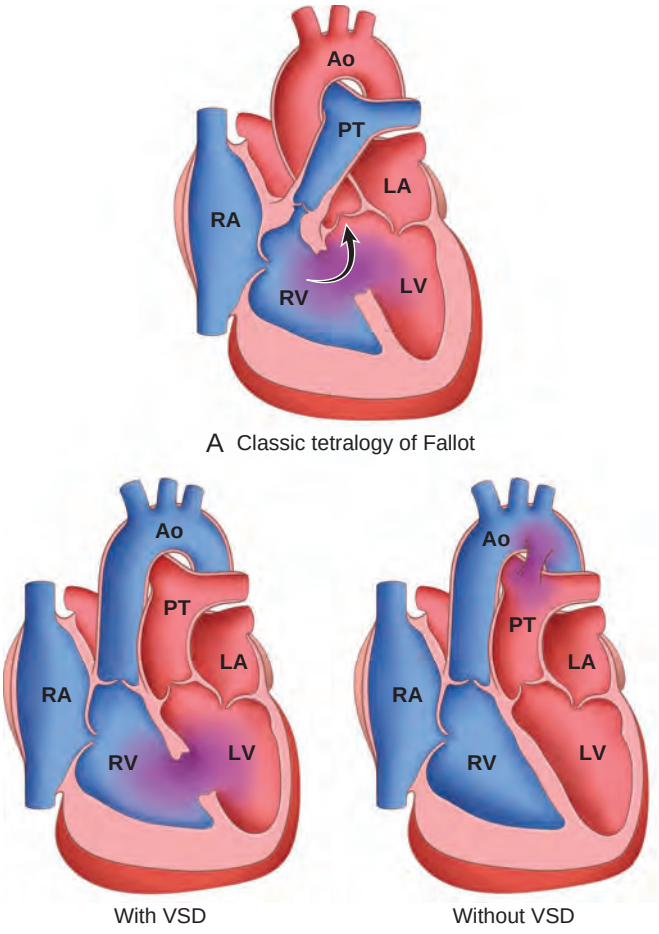


Figure 12.5 Common congenital right-to-left shunts (cyanotic congenital heart disease). (A) Tetralogy of Fallot. The direction of shunting across the ventricular septal defect *(VSD)* depends on the degree of the subpulmonic stenosis; when severe, a right-to-left shunt results *(arrow)*. (B) Dextro-transposition of the great arteries with and without VSD. *Ao,* Aorta; *LA,* left atrium; *LV,* left ventricle; *PT,* pulmonary trunk; *RA,* right atrium; *RV,* right ventricle.

by pulmonary valvular stenosis. Sometimes there is complete atresia of the pulmonary valve and variable portions of the pulmonary arteries, such that blood flow through a PDA, dilated bronchial arteries, or both, is necessary for survival. Aortic valve insufficiency or an ASD may also be present; a right aortic arch is present in about 25% of cases.

Clinical Features

Depending on the severity of the subpulmonic stenosis, untreated TOF can be tolerated into adulthood; 10% of unoperated individuals are alive at 20 years of age, and 3% survive for 40 years. If the subpulmonic stenosis is mild, the abnormality resembles an isolated VSD, and the shunt may be left-to-right, without cyanosis (so-called pink tetralogy). With more severe right ventricular outflow obstruction, right-sided pressures approach or exceed left-sided pressures, and right-to-left shunting develops, producing cyanosis (classic TOF). Most infants with TOF are cyanotic at birth or soon thereafter. The more severe the subpulmonic stenosis, the more hypoplastic are the pulmonary arteries (i.e., smaller and thinner-walled) and the larger is the overriding aorta. As the child grows and the heart increases in size, the pulmonic orifice does not expand proportionally, making the obstruction progressively worse. The subpulmonic stenosis, however, protects the pulmonary vasculature from pressure overload, and right ventricular failure is rare because the right ventricle is decompressed by the shunting of blood into the left ventricle and aorta. Complete surgical repair is possible but becomes complicated for individuals with pulmonary atresia and dilated bronchial arteries.

Transposition of the Great Arteries

TGA produces ventriculoarterial discordance (the ventricle outflow going to the wrong outflow vessel). In the more common variant (dextro-TGA or d-TGA), the aorta arises from the right ventricle, and the pulmonary artery emanates from the left ventricle (ventriculoarterial discordance; Fig. 12.5B and Fig. 12.6). The atrium-to-ventricle connections are normal (concordant), with the right atrium joining the right ventricle and the left atrium emptying into the left ventricle. The embryologic defect in complete TGA stems from abnormal formation of the spiraling truncal and aortopulmonary septae. The result is separation of the systemic and pulmonary circulations, a condition incompatible with postnatal life unless a shunt exists for adequate mixing of blood.

The outlook for infants with d-TGA depends on the degree of blood "mixing," the magnitude of tissue hypoxia, and the ability of the right ventricle to maintain the systemic circulation. Patients with d-TGA and a VSD (approximately 35%) often have a stable shunt. However, dependence on a patent foramen ovale or ductus arteriosus for blood mixing (approximately 65%) is problematic. These systemic-to-pulmonary connections tend to close early and thus require intervention to create a new shunt within the first few days of life (e.g., balloon atrial septostomy). With time, right

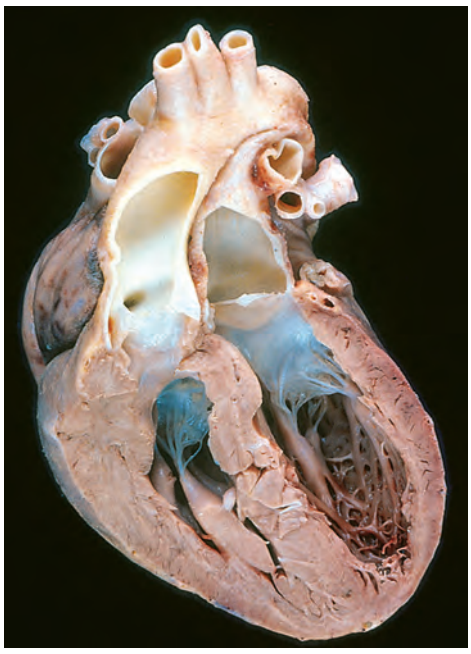


Figure 12.6 Dextro-transposition of the great arteries. (Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, Minn.)

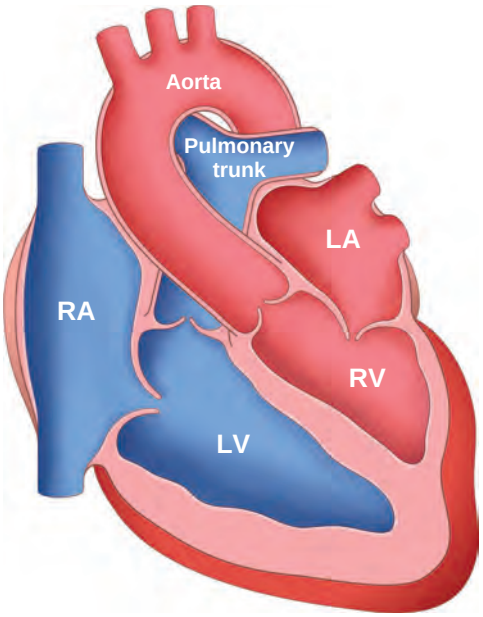


Figure 12.7 Schematic for "congenitally corrected" levo-transposition of the great arteries. The right atrium *(RA)* flows to the morphologic left ventricle *(LV)*, which connects to the pulmonary arteries; the left atrium *(LA)* connects to the morphologic right ventricle *(RV)*, which empties into the aorta, which is anomalously anterior and to the left. This congenital malformation may be well-tolerated, presenting only relatively late in adulthood with decompensation of the morphologic right ventricle.

ventricular hypertrophy becomes prominent because this chamber is functioning as the systemic ventricle. Concurrently, the left ventricle becomes thin-walled (atrophic) as it supports the low-resistance pulmonary circulation. Without surgery, most patients die within months. However, improved early surgical intervention, typically with an arterial switch operation, allows many patients with d-TGA to survive into adulthood.

In the less common levo-TGA (l-TGA, also commonly referred to as "congenitally corrected" TGA), the right atrium connects to a ventricle with the internal morphology of a left ventricle (atrioventricular discordance), which in turn empties into the pulmonary arteries (ventriculoarterial discordance); at the same time, the left atrium connects to a morphologic right ventricle, which empties into the aorta (Fig. 12.7). l-TGA does not lead to cyanosis and indeed can be entirely asymptomatic, being diagnosed in adulthood only during a workup for other cardiac issues. Nevertheless, l-TGA will result in hypertrophy of the morphologic right ventricle and eventually can cause heart failure; it is also often associated with other CHD such as VSD, ASD, and patent foramen ovale.

Tricuspid Atresia

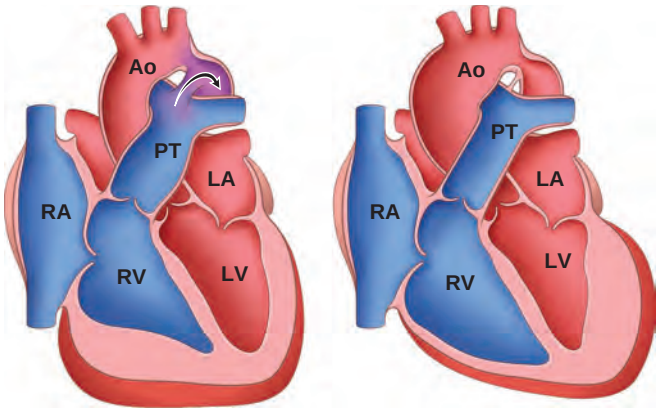
Tricuspid atresia represents complete occlusion of the tricuspid valve orifice. It results embryologically from unequal division of the AV canal; thus, the mitral valve is larger than normal, and there is right ventricular underdevelopment (hypoplasia). The circulation can be maintained by right-to-left shunting through an interatrial communication (ASD or patent foramen ovale), in addition to a VSD that connects the left ventricle and the pulmonary artery arising from the hypoplastic right ventricle. Cyanosis is present virtually from birth, and there is a high early mortality.

Obstructive Lesions

Congenital obstruction to blood flow can occur at the level of the heart valves or within a great vessel. Common examples include aortic or pulmonary valve stenosis or atresia, and coarctation of the aorta. Obstruction can also occur within a chamber, as with subpulmonic stenosis in TOF.

Coarctation of the Aorta

Coarctation (narrowing, constriction) of the aorta ranks high in frequency among the common structural anomalies. It is twice as common in males as in females; interestingly, females with Turner syndrome are also frequently affected (Chapter 5). There are two classic forms: (1) an "infantile" form—often symptomatic in early childhood—with tubular hypoplasia of the aortic arch proximal to a PDA, and (2) an "adult" form with a discrete ridgelike infolding of the aorta just opposite the closed ductus arteriosus (ligamentum arteriosum) distal to the arch vessels (Fig. 12.8). Encroachment



A Coarctation of aorta

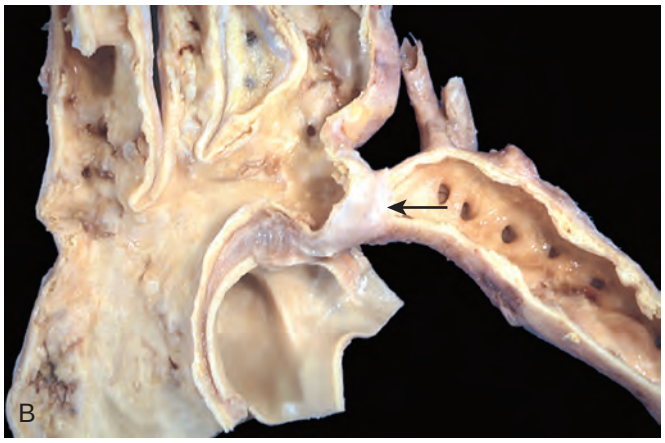


Figure 12.8 (A) Schematic of aortic coarctation with and without patent ductus arteriosus *(PDA)*. (B) Coarctation of the aorta, postductal type. The coarctation is a segmental narrowing of the aorta *(arrow)*. Such lesions typically manifest later in life than preductal coarctations. The dilated ascending aorta and major branch vessels are to the left of the coarctation. The lower extremities are perfused predominantly by way of dilated, tortuous collateral channels. *Ao*, Aorta; *LA*, left atrium; *LV*, left ventricle; *PT*, pulmonary trunk; *RA*, right atrium; *RV*, right ventricle; *PDA*, patent ductus arteriosus. (A, Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, Minn. B, Courtesy Sid Murphree, MD, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

on the aortic lumen is variable, sometimes leaving only a small channel and at other times producing only minimal narrowing. Although coarctation of the aorta may occur as a solitary defect, in 50% of cases it is accompanied by a bicuspid aortic valve and may also be associated with congenital aortic stenosis, ASD, VSD, mitral regurgitation, or berry aneurysms of the circle of Willis.

MORPHOLOGY

Preductal coarctation is characterized by circumferential narrowing of the aortic segment between the left subclavian artery and the ductus arteriosus; the ductus typically is patent and is the main source of (unoxygenated) blood delivered to the distal aorta. The pulmonary trunk is dilated to accommodate the increased blood flow; because the right side of the heart now perfuses the body distal to the narrowed segment ("coarct"), the right ventricle typically is hypertrophied.

In the more common **"adult" postductal coarctation,** the aorta is sharply constricted by a tissue ridge adjacent to the nonpatent ligamentum arteriosum (see Fig. 12.8B). The constricted segment is made up of smooth muscle and elastic fibers derived from the aortic media. Proximal to the coarctation, the aortic arch and its branch vessels are dilated and the left ventricle is hypertrophied.

Clinical Features

Clinical manifestations depend on the severity of the narrowing and the patency of the ductus arteriosus. Coarctation of the aorta with a PDA usually manifests early in life; indeed, it may cause signs and symptoms immediately after birth. In such cases, the delivery of unsaturated blood through the PDA produces cyanosis localized to the lower half of the body. Many such infants do not survive the neonatal period without surgical or catheter-based intervention to occlude the PDA.

The outlook is different with coarctation of the aorta without a PDA, unless the aortic constriction is severe. Most children are asymptomatic, and the disease may go unrecognized until well into adult life. Typically there is hypertension in the upper extremities with weak pulses and hypotension in the lower extremities, associated with manifestations of arterial insufficiency (i.e., claudication and coldness). Particularly characteristic is the development of collateral circulation between the pre-coarctation and post-coarctation arteries through enlarged intercostal and internal mammary arteries, often producing radiographically visible erosions ("notching") of the undersurfaces of the ribs.

With significant coarctations, murmurs are present throughout systole; sometimes a vibratory "thrill" is also present. The long-standing pressure overload leads to concentric left ventricular hypertrophy. With uncomplicated coarctation of the aorta, surgical resection and end-to-end anastomosis or replacement of the affected aortic segment by a prosthetic graft yields excellent results.

Pulmonary Stenosis and Atresia

Pulmonary stenosis or atresia is a relatively frequent malformation leading to obstruction at the level of the pulmonary valve. This can be mild to severe; the lesion can also be isolated or part of a more complex anomaly—either TOF or TGA. Right ventricular hypertrophy typically develops, and there is sometimes poststenotic dilation of the pulmonary artery due to injury of the wall by high-pressure jets through the stenotic valve. With coexistent subpulmonic stenosis (as in TOF), the pulmonary trunk is not dilated and may in fact be hypoplastic. When the valve is entirely atretic, there is no communication between the right ventricle and lungs. In such cases, the anomaly is associated with a hypoplastic right ventricle and an ASD; blood reaches the lungs through a PDA. Mild stenosis may be asymptomatic and compatible with long life, whereas symptomatic cases require surgical correction. When the RV is severely hypoplastic, blood can be surgically routed directly from the vena cavae to the pulmonary artery, bypassing the right side of the heart.

Aortic Stenosis and Atresia

Congenital narrowing and obstruction of the aortic valve can occur at three locations: valvular, subvalvular, and supravalvular. Congenital aortic valve stenosis is an isolated lesion in 80% of cases. With valvular aortic stenosis, the cusps may be hypoplastic (small), dysplastic (thickened, nodular), or abnormal in number (usually with one or no commissures). In severe congenital aortic stenosis or atresia, obstruction of the left ventricular outflow tract leads to hypoplasia of the left ventricle and ascending aorta, sometimes accompanied by dense, porcelain-like left ventricular endocardial fibroelastosis; the ductus arteriosus must be patent to allow blood flow to the aorta and coronary arteries. The constellation of findings is called hypoplastic left heart syndrome, and unless PDA patency is preserved, duct closure in the first week of life is generally lethal. Affected patients require surgeries that allow the right ventricle to become the pump for the systemic circulation, with the lungs perfused from either a parallel right ventricular conduit or passively from the venae cavae.

Subaortic stenosis is caused by a thickened ring or collar of dense endocardial fibrous tissue below the level of the cusps. Supravalvular aortic stenosis is a congenital aortic dysplasia with thickening and constriction of the ascending aortic wall. Elastin gene mutations can cause supravalvular stenosis by disrupting elastin–smooth muscle cell interactions during aortic morphogenesis. Subaortic stenosis is usually associated with a prominent systolic murmur and sometimes a thrill. Pressure hypertrophy of the left ventricle develops as a consequence of the obstruction to blood flow; unless very severe, the stenosis is generally well-tolerated, although left ventricular hypertrophy still carries a risk of sudden cardiac death.

KEY CONCEPTS

CONGENITAL HEART DISEASE

- CHD represents defects of cardiac chambers or the great vessels; these typically either result in shunting of blood between the right and left circulation or cause outflow obstruction. Lesions range from relatively asymptomatic to rapidly fatal. Environmental and genetic causes both contribute, and the manifestations depend on the timing of the environmental insult and which step in cardiac development is affected.

- Left-to-right shunts are most common and are typically associated with ASD, VSD, or PDA. These lesions result in chronic right-sided pressure and volume overloads that eventually cause pulmonary hypertension with reversal of flow and right-to-left shunts with cyanosis (Eisenmenger syndrome).
- Right-to-left shunts are most commonly caused by TOF or TGA. These are cyanotic lesions from the outset and are associated with polycythemia, peripheral cyanosis effects, and paradoxical emboli.
- Obstructive lesions include valve stenoses and aortic coarctation; the clinical severity of the lesion depends on the degree of stenosis and the patency of the ductus arteriosus.

ISCHEMIC HEART DISEASE

IHD represents a group of related entities resulting from myocardial ischemia—an imbalance between myocardial supply (perfusion) and cardiac demand for oxygenated blood. Ischemia not only limits tissue oxygenation (and thus ATP generation), but also reduces the availability of nutrients and the removal of metabolic wastes (Chapter 2). Thus, cardiac ischemia is generally less well tolerated than hypoxemia per se, such as may occur with severe anemia, cyanotic heart disease, or advanced lung disease.

In more than 90% of cases, myocardial ischemia results from reduced blood flow due to obstructive atherosclerotic lesions in the epicardial coronary arteries; consequently, IHD is frequently referred to as coronary artery disease (CAD). In most cases, there is a long period (up to decades) of silent, slow progression of coronary lesions before the sudden onset of symptoms. Thus, IHD is often the late manifestation of coronary atherosclerosis that began during childhood or adolescence (Chapter 11).

IHD can declare itself through one or more of the following clinical presentations:

- *Myocardial infarction (MI),* in which ischemia causes frank cardiac necrosis
- *Angina pectoris* (literally "chest pain"), in which ischemia is not severe enough to cause infarction, but the symptoms nevertheless portend infarction risk
- *Chronic IHD* with heart failure
- *Sudden cardiac death (SCD)*

In addition to coronary atherosclerosis, myocardial ischemia can be caused by coronary emboli, myocardial vessel inflammation, or vascular spasm. Moreover, otherwise modest coronary artery occlusions may become consequential in the setting of increased cardiac energy demand (e.g., myocardial hypertrophy or increased heart rate), hypoxemia, or systemic hypotension (e.g., shock). Some conditions can have multiple deleterious effects. Thus, tachycardia increases oxygen demand (because of more contractions per unit time) while decreasing functional supply (by decreasing the relative time spent in diastole, when cardiac perfusion occurs).

Epidemiology

IHD is the single largest cause of mortality worldwide, accounting for over 12% of global deaths; in the industrialized nations, this amounts to over 7.5 million casualties each year. Even in low income countries, advances in the treatment and prevention of infectious diseases, and the increasing adoption of westernized diets is accelerating IHD incidence; it is predicted soon to become the leading cause of mortality there as well.

At the same time, there is cause for optimism; since peaking in the mid-1960s, the overall death rate from IHD has fallen in the United States by over 50%. This remarkable improvement can be attributed to the following:

- *Prevention,* achieved by modifying important risk factors, such as smoking, level of blood cholesterol, and hypertension. Additional risk reduction can occur through weight loss, exercise, and maintaining good glycemic control in diabetic patients.
- *Diagnostic and therapeutic advances,* allowing earlier and more effective treatments. The latter include cholesterol- (and inflammation-) lowering drugs such as statins, thrombolysis for acute coronary occlusions, better medical management after MI, coronary angioplasty and stenting, coronary artery bypass graft (CABG) surgery, and improved therapies for heart failure and arrhythmias using left VADs, implantable defibrillators, and cardiac resynchronization approaches. Even a simple daily prophylactic aspirin can have therapeutic benefit.

Continuing this encouraging trend will be challenging, particularly in view of the increased longevity of "baby boomers" (which will lead to a doubling of individuals older than 65 years of age by 2050), and the "obesity epidemic." New therapeutic advances will depend on understanding the genetic determinants of coronary atherosclerosis and IHD. Thus, the observation that MIs occur in only a fraction of individuals with coronary disease suggests that simple control of atherosclerotic risk factors is only part of the story.

Pathogenesis

The dominant cause of IHD syndromes is insufficient coronary perfusion relative to myocardial demand; in the vast majority of cases, this is due to either of the following:

- *Chronic,* progressive atherosclerotic narrowing of the epicardial coronary arteries
- Variable degrees of superimposed *acute* plaque change, thrombosis, and vasospasm.

The individual elements and their interactions are discussed next.

Chronic Vascular Occlusion

More than 90% of patients with IHD have atherosclerosis involving one or more of the epicardial coronary arteries (Chapter 11). A fixed lesion obstructing greater than 70% of vascular cross-sectional area (so called "critical stenosis") is typically cited as the threshold for symptomatic ischemia precipitated by exercise (characteristically manifesting as exertional angina). With this degree of obstruction, compensatory coronary arterial vasodilation is no longer sufficient to meet even moderate increases in myocardial demand. Obstruction of 90% of the cross-sectional area of the lumen generally leads to inadequate coronary blood flow, even at rest. Slowly developing obstructions induce the formation

of collateral circulation that can mitigate the effects of even high-grade stenoses by allowing alternate channels to perfuse at-risk myocardium. Consequently, rather than just assessing the extent of local vascular stenosis, measurements of coronary flow across the stenosis (flow reserves) provide a better assessment of the consequences of fixed coronary artery occlusions, particularly when multifocal.

Although only a single major coronary epicardial vessel may be affected, two or all three—the left anterior descending (LAD), left circumflex (LCX), and right coronary artery (RCA)—are often involved simultaneously by obstructive atherosclerosis. Clinically significant plaques can be located anywhere along the course of the vessels, although they tend to predominate within the first several centimeters of the LAD and LCX. Sometimes the major epicardial branches are also involved (i.e., LAD diagonal branches, LCX obtuse marginal branches, or posterior descending branch of the RCA), but atherosclerosis of the intramyocardial (penetrating) branches is rare. Thus, most atherosclerotic stenoses can be accessed by coronary catheterization.

Acute Plaque Change

The risk of an individual developing clinically important IHD depends in part on the number, distribution, structure, and degree of obstruction by atheromatous plaques. However, the varied clinical manifestations of IHD cannot be explained by the anatomic disease burden and fixed stenoses alone. This is particularly true for the so-called acute coronary syndromes, namely when unstable angina, acute MI, and sudden death are caused abruptly by acute plaque changes. These acute coronary syndromes are typically initiated by an unpredictable and sudden conversion of a stable atherosclerotic plaque to an unstable and potentially life-threatening atherothrombotic lesion through rupture, superficial erosion, ulceration, fissuring, or deep hemorrhage (collectively called acute plaque change) (Chapter 11). In most instances, acute plaque changes—typically associated with intralesional inflammation—precipitate the formation of a superimposed thrombus that partially or completely occludes the artery. It remains to be seen whether aggressive anti-inflammatory regimens are a means to reduce such acute coronary events.

Consequences of Myocardial Ischemia

- *Stable angina* results from increases in myocardial oxygen demand that outstrip the ability of coronary arteries with fixed stenoses to increase oxygen delivery; it is usually not associated with plaque disruption.
- *Unstable angina* is caused by acute plaque change that results in thrombosis and/or vasoconstriction, and leads to incomplete or transient reductions in coronary blood flow. In some cases, microinfarcts can occur distal to disrupted plaques due to thromboemboli.
- *MI* is often the result of acute plaque change that induces an abrupt thrombotic occlusion, resulting in myocardial necrosis.
- *Sudden cardiac death* may be caused by regional myocardial ischemia that induces a fatal ventricular arrhythmia. This can result from a fixed stenosis or acute plaque change.

Each of these important syndromes is discussed in detail next, followed by an examination of the important myocardial consequences.

Angina Pectoris

Angina pectoris is characterized by paroxysmal and usually recurrent attacks of substernal or precordial chest discomfort caused by transient (15 seconds to 15 minutes) myocardial ischemia that is insufficient to induce myocyte necrosis. The anginal pain is a consequence of the ischemia-induced release of adenosine, bradykinin, and other molecules that stimulate sympathetic and vagal afferent nerves. Three overlapping patterns of angina pectoris are recognized, caused by varying combinations of decreased perfusion, increased demand, and coronary arterial pathology. Importantly, not all ischemic events are perceived by patients; silent ischemia is particularly common in the geriatric population and in the setting of diabetic neuropathy.

- *Stable (typical) angina* is the most common form of angina; it is caused by an imbalance in coronary perfusion (due to chronic stenosing coronary atherosclerosis) relative to myocardial demand. Stable angina does not occur at rest, but in a given patient can be reliably induced by activities that increase the energy requirements of the heart, including physical activity, emotional excitement, or psychological stress. Angina pectoris is classically described as a crushing or squeezing substernal sensation that can radiate down the left arm or to the left jaw *(referred pain)*. The pain is usually relieved by rest (decreasing demand) or by administering vasodilators, such as nitroglycerin and calcium channel blockers (thereby increasing perfusion).
- *Prinzmetal variant angina* is an uncommon form of episodic myocardial ischemia caused by coronary artery spasm. Although individuals with Prinzmetal variant angina can also have significant coronary atherosclerosis, the anginal attacks are unrelated to physical activity, heart rate, or blood pressure, and can occur at rest. Prinzmetal angina generally responds promptly to vasodilators.
- *Unstable* or *crescendo angina* refers to a pattern of increasingly frequent, prolonged (>20 min), or severe angina, precipitated by progressively lower levels of physical activity or even occurring at rest. Unstable angina is associated with plaque disruption and superimposed thrombosis, distal embolization of the thrombus, and/or vasospasm; it is an important harbinger of MI, potentially portending complete vascular occlusion.

Myocardial Infarction

MI, also commonly referred to as "heart attack," is the death of cardiac muscle due to prolonged ischemia. Roughly 1.5 million individuals in the United States suffer an MI each year, causing approximately 610,000 deaths annually. The major underlying cause of IHD is atherosclerosis; although MIs can occur at virtually any age, 10% of MIs occur in people younger than 40 years of age, and 45% occur in people younger than 65 years of age. Nevertheless, the frequency rises progressively with increasing age and with increasing atherosclerotic risk factors (Chapter 11). Through middle age, male gender increases the relative risk of MI; indeed, women are generally protected against MI during their reproductive years. However, postmenopausal decline in estrogen production is usually associated with accelerated CAD, and IHD is the most common cause of death in older

women. Unfortunately, postmenopausal hormonal replacement therapy has not been shown to be protective, and in fact, in some cases, may be detrimental.

Pathogenesis

Coronary Arterial Occlusion. The following sequence of events likely underlies most MIs (see Chapter 11 for additional details):

- An atheromatous plaque is eroded or suddenly disrupted by endothelial injury, intraplaque hemorrhage, or mechanical forces, exposing subendothelial collagen and necrotic plaque contents to the blood.
- Platelets adhere, aggregate, and are activated, releasing thromboxane A_2, adenosine diphosphate (ADP), and serotonin—causing further platelet aggregation and vasospasm (Chapter 4).
- Activation of coagulation by tissue factor and other mechanisms adds to the growing thrombus.
- Within minutes, the thrombus can evolve to completely occlude the coronary artery lumen.

The evidence for this scenario derives from autopsy studies of patients dying of acute MI, as well as imaging studies demonstrating a high frequency of thrombotic occlusion early after MI; interestingly, comparison to prior angiograms shows that these thrombi are usually at a site that did not previously have a critical (>70%) fixed stenosis. Typically, when angiography is performed within 4 hours of the onset of MI, it demonstrates coronary thrombosis in almost 90% of cases. However, when angiography is performed 12 to 24 hours after onset of symptoms, evidence of thrombosis is seen in only 60% of patients, even without intervention. Thus, at least some occlusions clear spontaneously through lysis of the thrombus or relaxation of spasm. This sequence of events in a typical MI also has therapeutic implications: early thrombolysis and/or angioplasty can be highly successful in limiting the extent of myocardial necrosis.

In approximately 10% of cases, MI occurs in the absence of the typical coronary atherothrombosis. In such situations, other mechanisms may be responsible for the reduced coronary blood flow:

- *Vasospasm* with or without coronary atherosclerosis, perhaps in association with platelet aggregation or due to drug ingestion (e.g., cocaine or ephedrine).
- *Emboli* from the left atrium in association with atrial fibrillation, a left-sided mural thrombus, vegetations of infective endocarditis (IE), intracardiac prosthetic material, or paradoxical emboli from the right side of the heart or the peripheral veins traversing a patent foramen ovale and into the coronary arteries
- *Uncommon causes* of MI without atherothrombosis include disorders of small intramural coronary vessels (e.g., vasculitis), hematologic abnormalities (e.g., sickle cell disease), amyloid deposition in vascular walls, vascular dissection, marked hypertrophy (e.g., due to aortic stenosis), lowered systemic blood pressure (e.g., shock), or inadequate myocardial "protection" during cardiac surgery.

Myocardial Response. Coronary arterial obstruction diminishes blood flow to a region of myocardium, causing ischemia, rapid myocardial dysfunction, and eventually—with prolonged vascular compromise—myocyte death. The anatomic region supplied by that artery is referred to as the area at risk. The outcome depends predominantly on the severity and duration of flow deprivation (Fig. 12.9).

The early biochemical consequence of myocardial ischemia is the cessation of aerobic metabolism within seconds, leading to inadequate production of high-energy phosphates (e.g., creatine phosphate and adenosine triphosphate) and accumulation of potentially noxious metabolites (e.g., lactic acid) (see Fig. 12.9A). Because of the exquisite dependence of myocardial function on oxygen and nutrients, myocardial contractility ceases within a minute or so of the onset of severe ischemia. Such loss of function contributes to decreased systolic function long before myocyte death occurs.

As detailed in Chapter 2, ultrastructural changes (including myofibrillar relaxation, glycogen depletion, cell and mitochondrial swelling) also develop within a few minutes

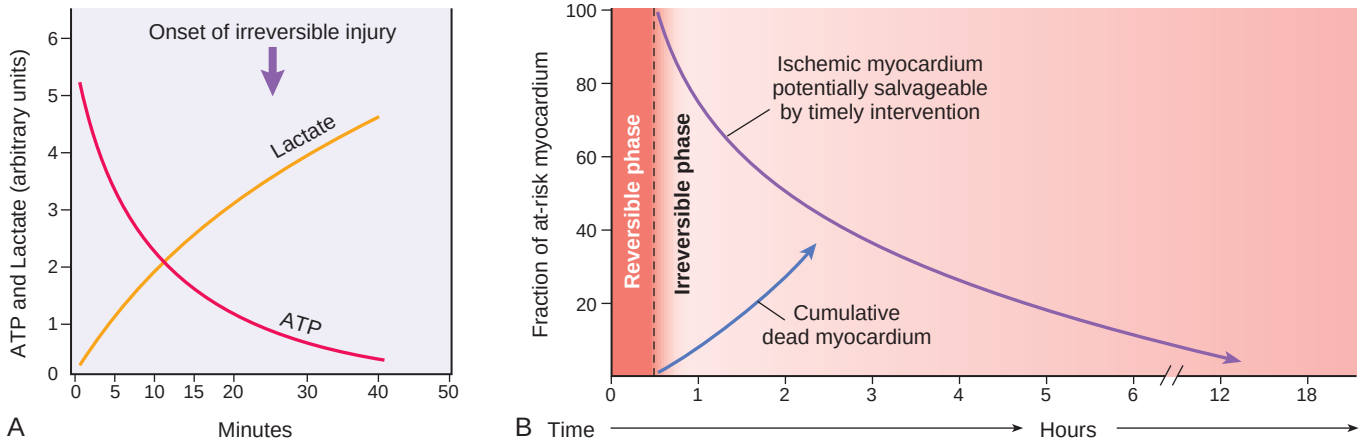


Figure 12.9 Temporal sequence of early biochemical findings and progression of cardiomyocyte necrosis after onset of severe myocardial ischemia. (A) Early changes include loss of adenosine triphosphate (ATP) and accumulation of lactate. (B) For approximately 30 minutes after the onset of even the most severe ischemia, myocardial injury is potentially reversible. Thereafter, progressive loss of viability occurs that becomes complete by 6 to 12 hours. The benefits of reperfusion are greatest when it is achieved early and are progressively lost when reperfusion is delayed. (Originally modified with permission from Antman E: Acute myocardial infarction. In Braunwald E, et al., editors: *Heart Disease: a Textbook of Cardiovascular Medicine*, ed 6, Philadelphia, 2001, WB Saunders, pp 1114–1231.)

of the onset of ischemia. Nevertheless, these early manifestations of ischemic injury are potentially reversible. Indeed, experimental and clinical evidence shows that only severe ischemia (blood flow 10% or less of normal) lasting 20 to 30 minutes or longer leads to irreversible damage (necrosis) of cardiac myocytes. This delay in the onset of permanent myocardial injury provides the rationale for rapid diagnosis in acute MI—to permit early coronary intervention to establish reperfusion and salvage as much "at risk" myocardium as possible.

The earliest detectable feature of myocyte necrosis is disruption of the integrity of the sarcolemmal membrane, allowing intracellular macromolecules to leak out of necrotic cells into the cardiac interstitium and ultimately into the microvasculature and lymphatics. This escape of intracellular myocardial proteins into the circulation forms the basis for blood tests that can sensitively detect irreversible myocyte damage, and are important for managing MI (see later). With prolonged severe ischemia, injury to the microvasculature follows injury to the cardiac myocytes. The temporal progression of these events is summarized in Table 12.4.

Table 12.4 Approximate Time of Onset of Key Events in Ischemic Cardiac Myocytes

Feature	Time
Onset of ATP depletion	Seconds
Loss of contractility	<2 minutes
ATP reduced to 50% of normal to 10% of normal	 10 minutes 40 minutes
Irreversible cell injury	20–40 minutes
Microvascular injury	>1 hour

ATP, Adenosine triphosphate.

The progression of ischemic necrosis in the myocardium is summarized in Fig. 12.10. Irreversible injury of ischemic myocytes first occurs in the subendocardial zone. This region is especially susceptible to ischemia because it is the last area to receive blood delivered by the epicardial vessels, and also because it is exposed to relatively high intramural

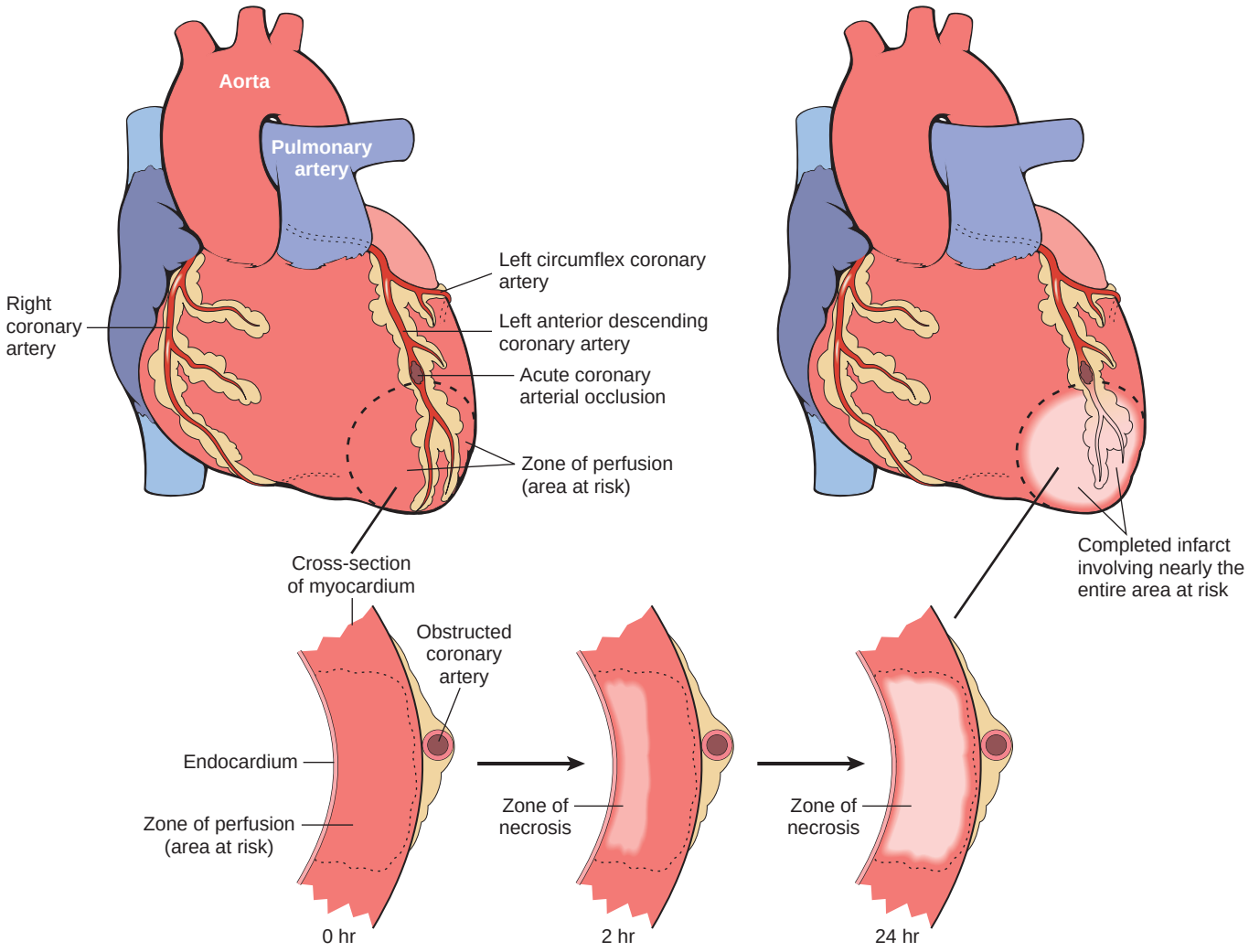


Figure 12.10 Progression of myocardial necrosis after coronary artery occlusion. Necrosis begins in a small zone of the myocardium beneath the endocardial surface in the center of the ischemic zone. The area that depends on the occluded vessel for perfusion is the "at risk" myocardium *(shaded)*. Note that a very narrow zone of myocardium immediately beneath the endocardium is spared from necrosis because oxygen and nutrition can be provided by diffusion from the ventricle.

pressures, which act to impede the inflow of blood. With more prolonged ischemia, a wavefront of cell death moves through other regions of the myocardium, driven by progressive tissue edema and myocardial-derived reactive oxygen species and inflammatory mediators.

The location, size, and specific morphologic features of an acute MI depend on the following:

- The *location, severity, and rate of development of coronary obstructions* due to atherosclerosis and thromboses
- The *size of the vascular bed* perfused by the obstructed vessels
- The *duration* of the occlusion
- The *metabolic and oxygen needs* of the myocardium at risk
- The extent of *vascular collateralization*
- The presence, site, and severity of coronary *arterial spasm*
- *Other factors,* such as heart rate, cardiac rhythm, and blood oxygenation

An infarct usually achieves its full extent within 3 to 6 hours; in the absence of intervention, an infarct caused by occlusion of an epicardial vessel can involve the entire wall thickness (transmural infarct). Clinical intervention within this critical window of time can lessen the size of the infarct within the territory at risk.

Patterns of Infarction. The distribution of myocardial necrosis correlates with the location and cause of the decreased perfusion (Fig. 12.11).

Knowledge of the areas of myocardium perfused by the major coronary arteries allows correlation of specific vascular obstructions with their corresponding areas of MI. Typically, the LAD branch of the left coronary artery supplies most of the apex of the heart, the anterior wall of the left ventricle, and the anterior two-thirds of the ventricular septum. By convention, the coronary artery—either RCA or LCX—that perfuses the posterior third of the septum is called "dominant" (even though the LAD and LCX collectively perfuse the majority of the left ventricular myocardium). In a right dominant circulation (present in approximately 80% of individuals), the RCA supplies the entire right ventricular free wall, the posterobasal wall of the left ventricle, and the posterior third of the ventricular septum, and the LCX generally perfuses only the lateral wall of the left ventricle. Thus, RCA occlusions can potentially lead to left ventricular damage.

Although most hearts have numerous intercoronary anastomoses (collateral circulation), relatively little blood normally courses through these. However, when a coronary artery is progressively narrowed over time, blood flows via the collaterals from the high- to the low-pressure circulation causing the channels to enlarge. Through such progressive dilation and growth of collaterals, stimulated by ischemia, blood flow is provided to areas of myocardium that would otherwise be deprived of adequate perfusion. Indeed, in the setting of extensive collateralization, the normal epicardial perfusion territories may be so expanded that subsequent occlusion leads to infarction in paradoxical distributions.

Transmural infarctions occur when there is occlusion of an epicardial vessel (in the absence of any therapeutic intervention)—the necrosis involves virtually the full thickness of the ventricular wall in the distribution of the affected coronary. This pattern of infarction is usually associated with a combination of chronic coronary atherosclerosis, acute

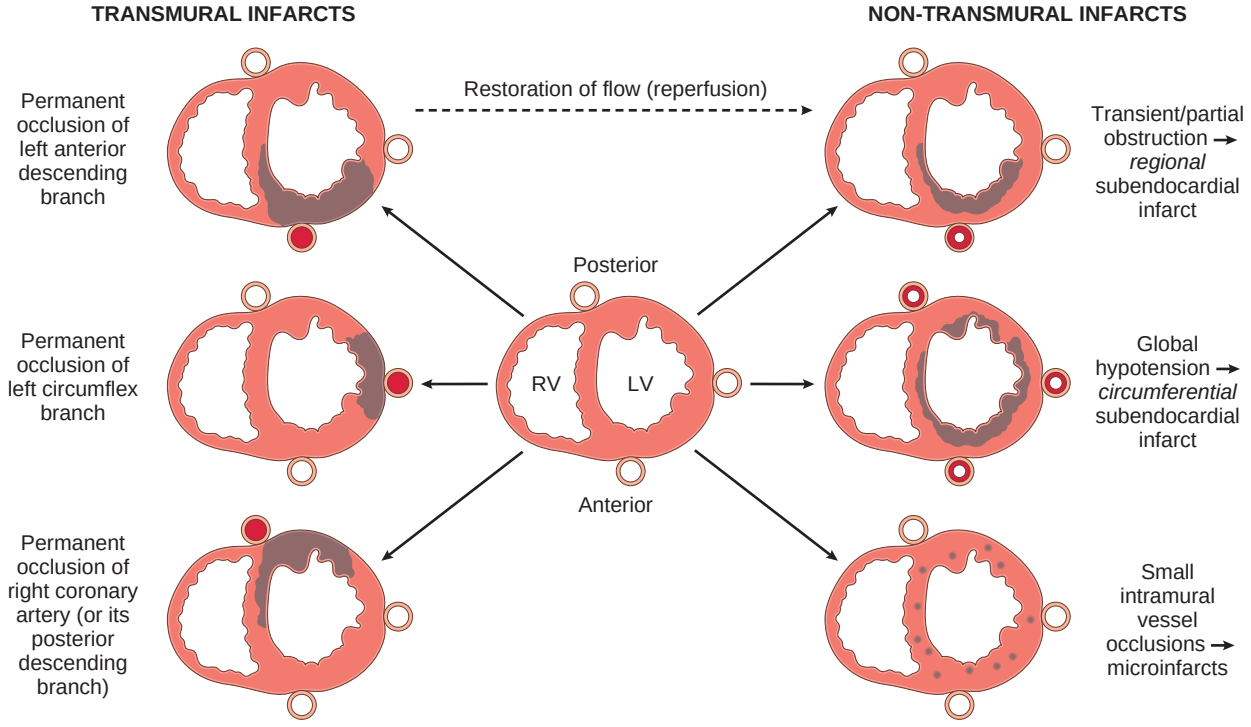


Figure 12.11 Distribution of myocardial ischemic necrosis correlates with the location and nature of decreased perfusion. *Left,* the positions of transmural acute infarcts resulting from occlusions of the major coronary arteries; *top to bottom,* left anterior descending, left circumflex, and right coronary arteries. *Right,* the types of infarcts that result from a partial or transient occlusion, global hypotension, or intramural small vessel occlusions.

plaque change, and superimposed thrombosis (discussed earlier).

Subendocardial (nontransmural) infarctions can occur as a result of a plaque disruption followed by a coronary thrombus that becomes lysed (therapeutically or spontaneously) before myocardial necrosis extends across the full thickness of the wall. Subendocardial infarcts can also result from prolonged, severe reduction in systemic blood pressure, as in shock superimposed on chronic, otherwise noncritical, coronary stenoses. In the subendocardial infarcts that occur as a result of global hypotension, myocardial damage is often circumferential, rather than being limited to the distribution of a single major coronary artery.

Multifocal microinfarction refers to a pattern that is seen when there is pathology involving only smaller intramural vessels. It can occur in the setting of microembolization, vasculitis, or vascular spasm, for example, due to endogenous catecholamines (epinephrine) or drugs (cocaine or ephedrine). Elevated levels of catecholamines also increase heart rate and myocardial contractility, exacerbating ischemia caused by the vasospasm. The outcome of such vasospasm can be sudden cardiac death (usually caused by a fatal arrhythmia) or an ischemic dilated cardiomyopathy (discussed later).

MORPHOLOGY

The temporal evolution of the morphologic changes in acute MI and subsequent healing are summarized in Table 12.5.

Nearly all transmural infarcts involve at least a portion of the left ventricle (comprising the free wall and ventricular septum); they encompass nearly the entire perfusion zone of the occluded coronary artery save for a narrow rim (approximately 0.1 mm) of viable subendocardial myocardium that is preserved by diffusion of oxygen and nutrients from the ventricular lumen.

The frequencies of involvement of each of the three main arterial trunks and the corresponding sites of myocardial lesions resulting in infarction (in the typical right dominant heart) are as follows (left side of Fig. 12.11):

- Left anterior descending coronary artery (40% to 50%): infarcts involving the anterior wall of left ventricle near the apex; the anterior portion of ventricular septum; and the apex circumferentially
- Right coronary artery (30% to 40%): infarcts involving the inferior/posterior wall of the left ventricle; posterior portion of ventricular septum; and the inferior/posterior right ventricular free wall in some cases
- Left circumflex coronary artery (15% to 20%): infarcts involving the lateral wall of the left ventricle except at the apex

Other locations of critical coronary arterial lesions causing infarcts are sometimes encountered, such as the left main coronary artery, the secondary (diagonal) branches of the left anterior descending coronary artery, or the marginal branches of the left circumflex coronary artery. Of MIs caused by a right coronary obstruction, 15% to 30% extend from the posterior free wall of the septal portion of the left ventricle into the adjacent right ventricular wall. Isolated infarction of the right ventricle is unusual (only 1% to 3% of cases), as is infarction of the atria.

The gross and microscopic appearance of an infarct depends on the interval of time between the MI and death. Damaged myocardium undergoes a progressive sequence of morphologic changes involving typical ischemic coagulative necrosis, the predominant mechanism of cell death in MI (although apoptosis

Table 12.5 Evolution of Morphologic Changes in Myocardial Infarction

Time	Gross Features	Light Microscope	Electron Microscope
Reversible Injury			
0–½ hour	None	None	Relaxation of myofibrils; glycogen loss; mitochondrial swelling
Irreversible Injury			
½–4 hours	None	Usually none; variable waviness of fibers at border	Sarcolemmal disruption; mitochondrial amorphous densities
4–12 hours	Dark mottling (occasional)	Early coagulative necrosis; edema; hemorrhage	
12–24 hours	Dark mottling	Ongoing coagulative necrosis; pyknosis of nuclei; myocyte hypereosinophilia; marginal contraction band necrosis; early neutrophilic infiltrate	
1–3 days	Mottling with yellow-tan infarct center	Coagulative necrosis, with loss of nuclei and striations; brisk interstitial infiltrate of neutrophils	
3–7 days	Hyperemic border; central yellow-tan softening	Beginning disintegration of dead myofibers, with dying neutrophils; early phagocytosis of dead cells by macrophages at infarct border	
7–10 days	Maximally yellow-tan and soft, with depressed red-tan margins	Well-developed phagocytosis of dead cells; granulation tissue at margins	
10–14 days	Red-gray depressed infarct borders	Well-established granulation tissue with new blood vessels and collagen deposition	
2–8 weeks	Gray-white scar, progressive from border toward core of infarct	Increased collagen deposition, with decreased cellularity	
>2 months	Scarring complete	Dense collagenous scar	

can also occur); this is followed by stereotypical inflammation and repair that parallels responses to injury in other tissues (Chapter 2).

Early morphologic recognition of acute MI can be difficult, particularly when death occurs within only a few hours of symptom onset. MIs less than 12 hours old are usually not apparent on gross examination alone; however, if the infarct preceded death by at least 2 to 3 hours, it is possible to highlight the area of necrosis by immersion of tissue slices in a solution of **triphenyltetrazolium chloride.** This gross histochemical stain imparts a brick-red color to intact, noninfarcted myocardium with preserved lactate dehydrogenase activity. Because dehydrogenases leak out through the damaged membranes of dead cells, an infarct appears as an unstained pale zone (Fig. 12.12). By 12 to 24 hours after infarction, an MI can usually be identified grossly as a reddish-blue area of discoloration caused by congestion and extravasated blood. By 3 to 7 days, it is rimmed by a hyperemic zone of highly vascularized early wound healing **(granulation tissue).** Thereafter, the infarct becomes progressively more sharply defined, yellow-tan, and soft. Over the succeeding weeks, the injured region evolves to a fibrous scar.

The histopathologic changes also proceed in a fairly predictable sequence (Fig. 12.13). The typical changes of coagulative necrosis become detectable in the first 6 to 12 hours. "Wavy fibers" may

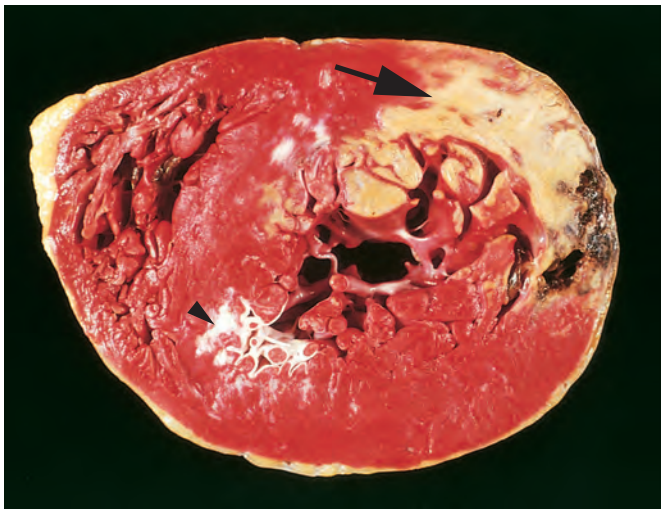


Figure 12.12 Acute myocardial infarct, predominantly of the posterolateral left ventricle, demonstrated histochemically by a lack of staining by triphenyltetrazolium chloride in areas of necrosis *(arrow)*. The staining defect is due to the lactate dehydrogenase leakage that follows cell death. Note the myocardial hemorrhage at one edge of the infarct that was associated with cardiac rupture, and the anterior scar *(arrowhead),* indicative of an old infarct. The specimen is oriented with the posterior wall at the top.

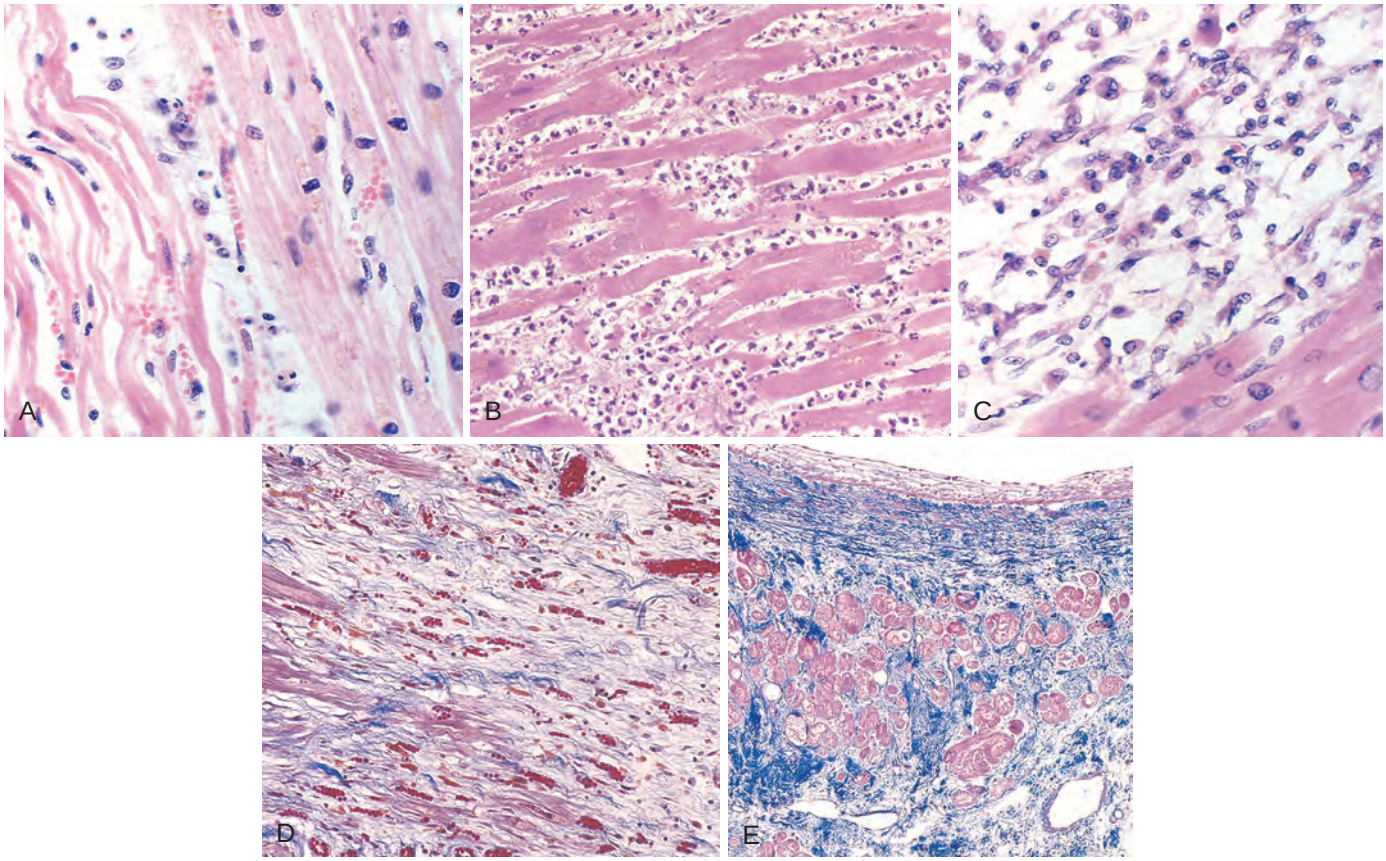


Figure 12.13 Microscopic features of myocardial infarction and its repair. (A) One-day-old infarct showing coagulative necrosis and wavy fibers (elongated and narrow, as compared with adjacent normal fibers at *right*). Widened spaces between the dead fibers contain edema fluid and scattered neutrophils. (B) Dense polymorphonuclear leukocytic infiltrate in an acute myocardial infarction that is 3 to 4 days old. (C) Removal of necrotic myocytes by phagocytosis (approximately 7 to 10 days). (D) Granulation tissue characterized by loose collagen and abundant capillaries. E, Healed myocardial infarct in which the necrotic tissue has been replaced by a dense collagenous scar. The residual cardiac muscle cells show evidence of compensatory hypertrophy. D and E are stained with Masson's trichrome, rendering collagenous connective tissue a deep blue color; note the accumulation of extracellular matrix (density of blue stain) between the early granulation tissue and the subsequent dense scar.

be present at the periphery of the infarct, resulting from the contractions of viable myocardium that cause the stretching and buckling of adjacent, noncontractile dead fibers. An additional sublethal ischemic change may be seen in the margins of infarcts: this so-called **myocyte vacuolization** reflects intracellular accumulations of salt and water within the sarcoplasmic reticulum, and marks myocytes that are viable but poorly contractile.

Necrotic myocardium elicits acute inflammation (typically most prominent 1 to 3 days after MI), followed by a wave of macrophages that remove necrotic myocytes and neutrophil fragments (most pronounced 3 to 7 days after MI). The infarcted zone is progressively replaced by granulation tissue (most prominent 7 to 10 days after MI), which in turn forms the provisional scaffolding on which dense collagenous scar forms; because cardiac myocytes are terminally differentiated cells, no cardiomyocyte proliferation is seen. In most instances, scarring is well advanced by the end of the sixth week, but the efficiency of repair depends on the size of the original lesion and the ability of the host tissues to heal. Healing requires the migration of inflammatory cells and ingrowth of new vessels from the infarct margins. Thus, an MI heals from its borders toward the center, and a large infarct may not heal as fast or as completely as a small one. Moreover, malnutrition, poor vasculature, or exogenous anti-inflammatory steroids can impede effective infarct scarring (Chapter 3). Once an MI is completely healed, it is impossible to distinguish its age: whether present for 8 weeks or 10 years, fibrous scars look the same.

The following discussion considers the changes that result from interventions that can limit infarct size by salvaging myocardium that is not yet necrotic.

Infarct Modification by Reperfusion. The therapeutic goal in acute MI is to salvage the maximal amount of ischemic myocardium; this is accomplished by restoration of tissue perfusion as quickly as possible (hence the adage "time is myocardium"). Such reperfusion is achieved by thrombolysis (dissolution of thrombus by tissue plasminogen activator), angioplasty, or coronary arterial bypass graft. Unfortunately, although preservation of viable (but at-risk) heart can improve both short- and long-term outcomes, reperfusion is not an unalloyed blessing. Indeed, late restoration of blood flow into ischemic tissues can be associated with arrhythmias, and can incite greater local damage than might otherwise have occurred—so-called reperfusion injury.

The effects of reperfusion on myocardial viability and function are summarized in Fig. 12.14. Although the clinical significance of myocardial reperfusion injury is debated, it has been estimated that up to 50% (or more) of the ultimate infarct size can be attributed to its effects. To date, clinical trials to prevent reperfusion injury have not been fruitful, but this remains an active area of investigation. Factors that contribute to reperfusion injury include the following:

- *Mitochondrial dysfunction:* Ischemia alters the mitochondrial membrane permeability, which allows proteins to move into the mitochondria. This leads to swelling and rupture of the outer membrane, releasing mitochondrial contents that promote apoptosis.
- *Myocyte hypercontracture:* During periods of ischemia, the intracellular levels of calcium are increased as a result of impaired calcium cycling and sarcolemmal damage. After reperfusion, the contraction of myofibrils is augmented and uncontrolled, causing cytoskeletal damage and cell death.
- *Free radicals* including superoxide anion (•O2), hydrogen peroxide (H_2O_2), hypochlorous acid (HOCl), nitric oxide–derived peroxynitrite, and hydroxyl radicals (•OH) are produced within minutes of reperfusion and cause damage to the myocytes by altering membrane proteins and phospholipids.
- *Leukocyte aggregation,* which may occlude the microvasculature and contribute to the "no-reflow" phenomenon. Further, leukocytes elaborate proteases and elastases that cause cell death.
- *Platelet and complement activation* also contribute to microvascular injury. Complement activation is thought to play a role in the no-reflow phenomenon by injuring the endothelium.

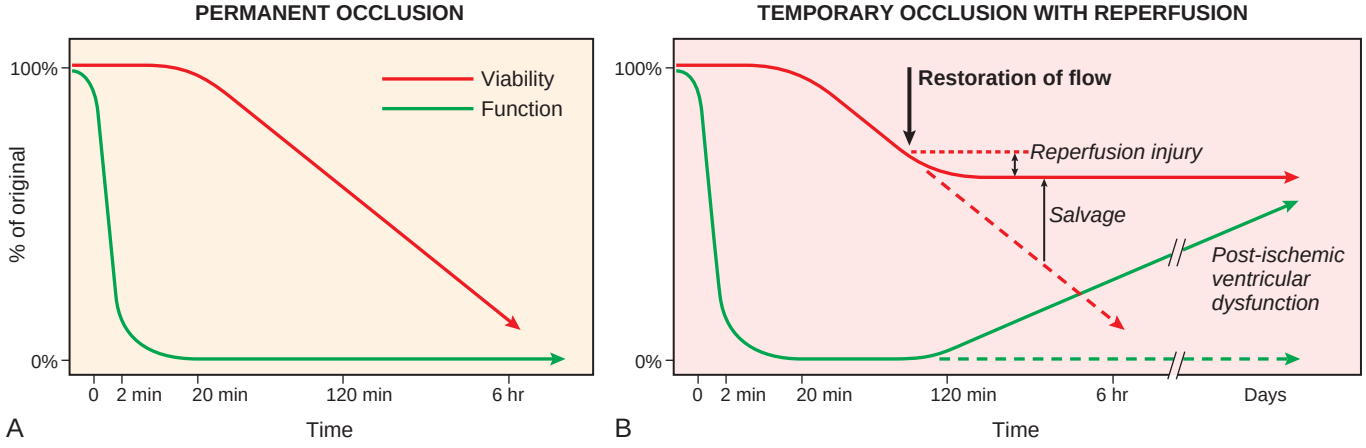


Figure 12.14 Effects of reperfusion on myocardial viability and function. After coronary occlusion, contractile function is lost within 2 minutes, and viability begins to diminish after approximately 20 minutes. If perfusion is not restored (A), then nearly all myocardium in the affected region suffers death. (B) If flow is restored, then some necrosis is prevented, myocardium is salvaged, and at least some function can return. The earlier reperfusion occurs, the greater the degree of salvage. However, the process of reperfusion itself may induce some damage (reperfusion injury), and return of function of salvaged myocardium may be delayed for hours to days (postischemic ventricular dysfunction or stunning).

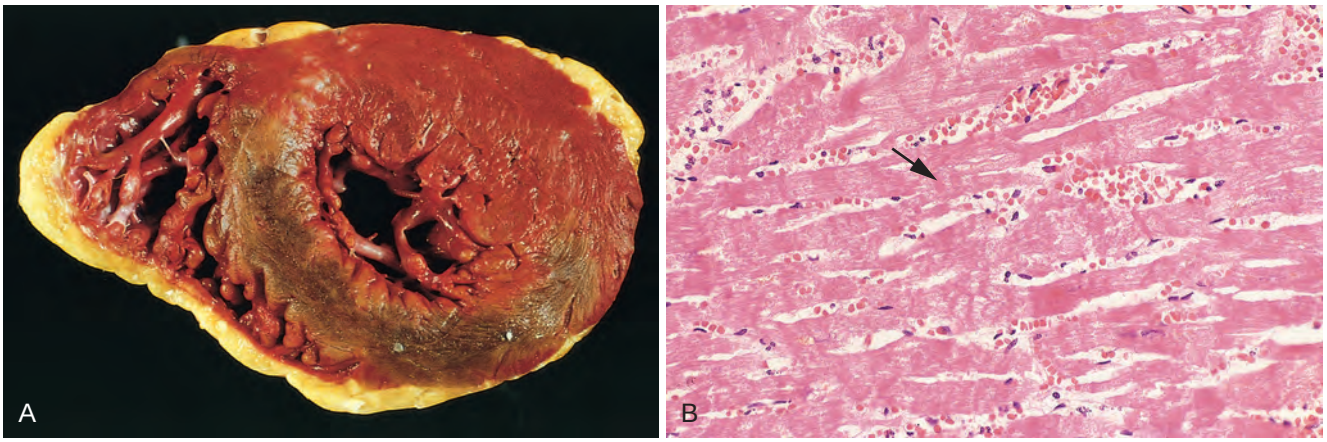


Figure 12.15 Consequences of myocardial ischemia followed by reperfusion. Gross (A) and microscopic (B) appearance of myocardium modified by reperfusion. (A) Large, densely hemorrhagic, anterior wall acute myocardial infarction in a patient with left anterior descending artery thrombus treated with streptokinase, a fibrinolytic agent (slice is stained with triphenyl tetrazolium chloride; see Fig. 12.12). Specimen oriented with posterior wall at top. (B) Myocardial necrosis with hemorrhage and contraction bands, visible as dark bands spanning some myofibers *(arrow)*.

The typical appearance of reperfused myocardium in the setting of an acute MI is shown in Fig. 12.15. Such infarcts typically are hemorrhagic as a consequence of vascular injury and leakiness. Microscopically, irreversibly damaged myocytes after reperfusion develop contraction band necrosis; in this pathologic process, intense eosinophilic bands of hypercontracted sarcomeres are created by an influx of calcium across plasma membranes that heightens actin-myosin interactions. In the absence of ATP, the sarcomeres cannot relax and get stuck in an agonal tetanic state. Thus, although reperfusion can salvage reversibly injured cells, it also alters the morphology of irreversibly injured cells.

The biochemical abnormalities (and their functional consequences) may also persist for days to weeks in reperfused myocytes. Such changes are thought to underlie a phenomenon referred to as stunned myocardium, a state of prolonged contractile dysfunction induced by short-term ischemia that usually recovers after several days. Myocardium that is subjected to chronic, sublethal ischemia can also enter into a state of lowered metabolism and function called hibernation. Subsequent revascularization (e.g., by CABG surgery, angioplasty, or stenting) often restores normal function to such hibernating myocardium.

Clinical Features

MI is diagnosed by clinical symptoms, laboratory tests for the presence of myocardial proteins in the plasma, and characteristic electrocardiographic changes. Patients with MI classically present with prolonged (more than 30 minutes) chest pain described as crushing, stabbing, or squeezing, associated with a rapid, weak pulse. Profuse sweating (diaphoresis), and nausea and vomiting are common and can suggest involvement of the posterior-inferior ventricle with secondary vagal stimulation. Dyspnea due to impaired contractility of the ischemic myocardium and the resultant pulmonary congestion and edema is a frequent symptom. However, in as many as 25% of patients the onset is entirely asymptomatic (e.g., in the setting of diabetic neuropathy), and the disease is discovered only by electrocardiographic changes or laboratory tests that show evidence of myocardial damage (see later).

Owing to the characteristic electrocardiographic changes resulting from myocardial ischemia or necrosis in various distributions, a transmural infarct is sometimes referred to as an ST-elevation myocardial infarct (STEMI) and a subendocardial infarct as a non–ST-elevation infarct (NSTEMI). Depending on the extent and location of the vascular involvement, microinfarctions show nonspecific changes or can even be electrocardiographically silent.

The laboratory evaluation of MI is based on measuring the blood levels of proteins that leak out of irreversibly damaged myocytes; previously, this involved measurement of the MB fraction of creatine kinase (CK-MB). Currently, the most clinically useful biomarkers of myocardial damage are cardiac-specific troponins T and I (cTnT and cTnI), proteins that normally regulate calcium-mediated contraction of cardiac muscle (Fig. 12.16). The diagnosis of myocardial injury is established when blood levels of troponin are elevated above threshold levels; the tempo and magnitude of appearance of these serum markers after MI depends on several factors, including the volume of damaged myocardium, blood flow and lymphatic drainage in the area of the infarct, and the rate of elimination of the marker from the blood. Cardiac troponins begin to rise in 2 to 4 hours and peak at 24 to 48 hours after an acute infarct. With reperfusion, troponin levels may be higher and peak earlier owing to more rapid washout of the marker from the necrotic tissue.

Significant acute elevation in serum troponin can also be seen in other conditions that cause death of cardiac myocytes, such as myocarditis and myocardial trauma. On the other hand, low-level serum troponin elevation ("troponin leak") can occur in a host of other conditions, including congestive heart failure, pulmonary embolus, renal failure, and sepsis. These elevations do not usually follow the same abrupt-injury time course, so serial measurements may be helpful in distinguishing different etiologies.

Recommended initial therapies for acute MI make sense based on the pathophysiologic features discussed earlier and include the following:

- *Oxygen supplementation* to improve blood oxygen saturation for patients with hypoxia or respiratory distress
- *Nitrates* to induce vasodilation and reverse vasospasm

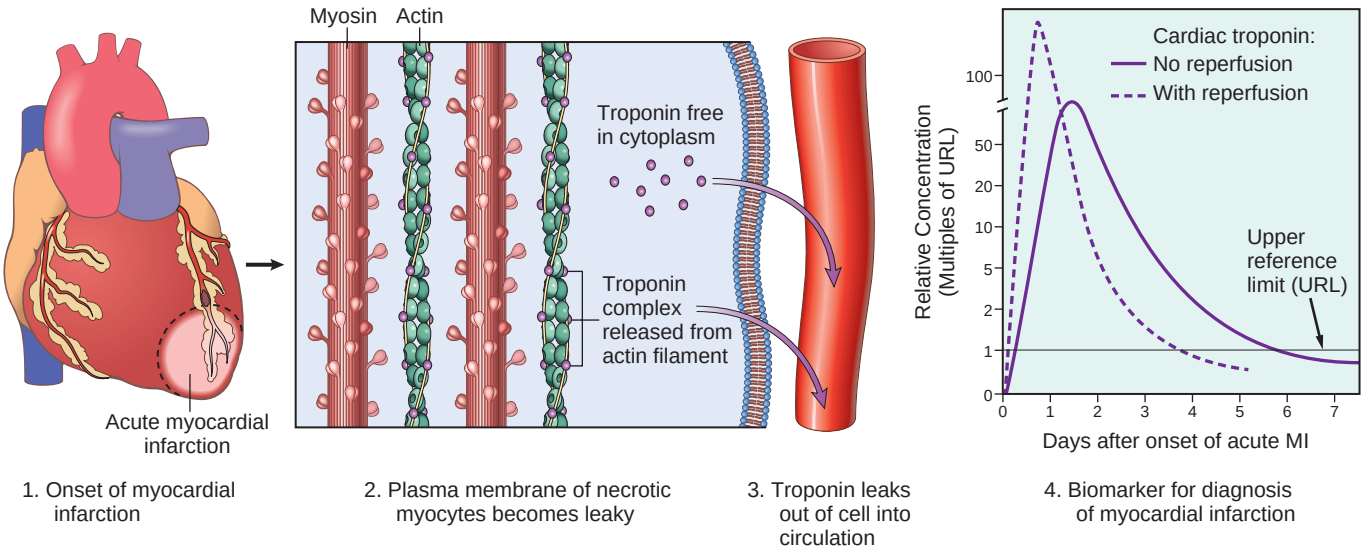


Figure 12.16 Release of myocyte proteins in myocardial infarction. Troponin I or troponin T are now most routinely used as diagnostic biomarkers of myocyte injury.

- *Antiplatelet agents* such as aspirin, ADP receptor inhibitors, and GPIIb/IIIa inhibitors
- *Anticoagulant therapy* with unfractionated heparin, low-molecular-weight heparin, direct thrombin inhibitors, and/or factor Xa inhibitors to prevent coronary artery thrombus propagation
- *Beta blockers* to decrease myocardial oxygen demand and to reduce the risk of arrhythmias, unless contraindicated such as in heart failure
- *Prompt reperfusion* to salvage myocardium, by either fibrinolytic medications or transcatheter interventions
- *Improve myocardial oxygen supply versus demand* by management of other factors, such as anxiety, ischemic pain, abnormal hemodynamics, anemia, and respiratory disorders
- *Early arrhythmia monitoring* and management

Consequences and Complications of Myocardial Infarction

Extraordinary progress has been made in improving patient outcomes after acute MI. The overall in-hospital death rate for MI is approximately 7% to 8%, with STEMI patients experiencing slightly higher mortality (10%) than NSTEMI patients (6%). Unfortunately, mortality for out-of-hospital MIs is substantially worse: one-third of patients who are unfortunate enough to suffer a STEMI at home will die, usually of an arrhythmia within 1 hour of symptom onset and before they can receive medical attention. Such statistics make the rising rate of CAD in countries with limited hospital facilities all the more worrisome.

Nearly three-fourths of patients experience one or more of the following complications after an acute MI (Fig. 12.17):

- *Contractile dysfunction.* In general, MIs affect left ventricular pump function in proportion to the volume of damage. In most cases, there is some degree of left ventricular failure manifested as hypotension, pulmonary congestion, and pulmonary edema. Severe "pump failure" *(cardiogenic shock)* occurs in roughly 10% of patients with transmural MIs and typically is associated with infarcts that damage 40% or more of the left ventricle.
- *Papillary muscle dysfunction.* Although papillary muscles rupture infrequently after MI, they often are dysfunctional and can be poorly contractile as a result of ischemia, leading to postinfarct mitral regurgitation. Much later, papillary muscle fibrosis and shortening, or global ventricular dilation also can cause mitral regurgitation.
- *Right ventricular infarction.* Although isolated right ventricular infarction occurs in only 1% to 3% of MIs, the right ventricle is affected by RCA occlusions leading to posterior septal or left ventricular infarction. In either case, right-sided heart failure is a common outcome, leading to pooling of blood in the venous circulation and systemic hypotension.
- *Myocardial rupture.* Rupture complicates only 1% to 5% of MIs, but is frequently fatal when it occurs. Left ventricular free wall rupture is most common, usually resulting in rapidly fatal hemopericardium and cardiac tamponade (see Fig. 12.17A). Ventricular septal rupture creates a VSD with left-to-right shunting (see Fig. 12.17B), and papillary muscle rupture leads to severe mitral regurgitation (see Fig. 12.17C). Rupture occurs most commonly within 3 to 7 days after infarction—the time in the healing process when lysis of necrotic myocardium is maximal and when much of the infarct has been converted to soft, friable granulation tissue. Risk factors for free wall rupture include age older than 60 years, anterior or lateral wall infarctions, female gender, lack of left ventricular hypertrophy, and first MI (because scarring associated with prior MIs tends to limit the risk for myocardial tearing).
- *Arrhythmias.* MIs lead to myocardial irritability and conduction disturbances that can cause sudden death. Approximately 90% of patients develop some form of rhythm disturbance, with the incidence being higher in STEMIs versus NSTEMIs. MI-associated arrhythmias include heart block of variable degree (including asystole),

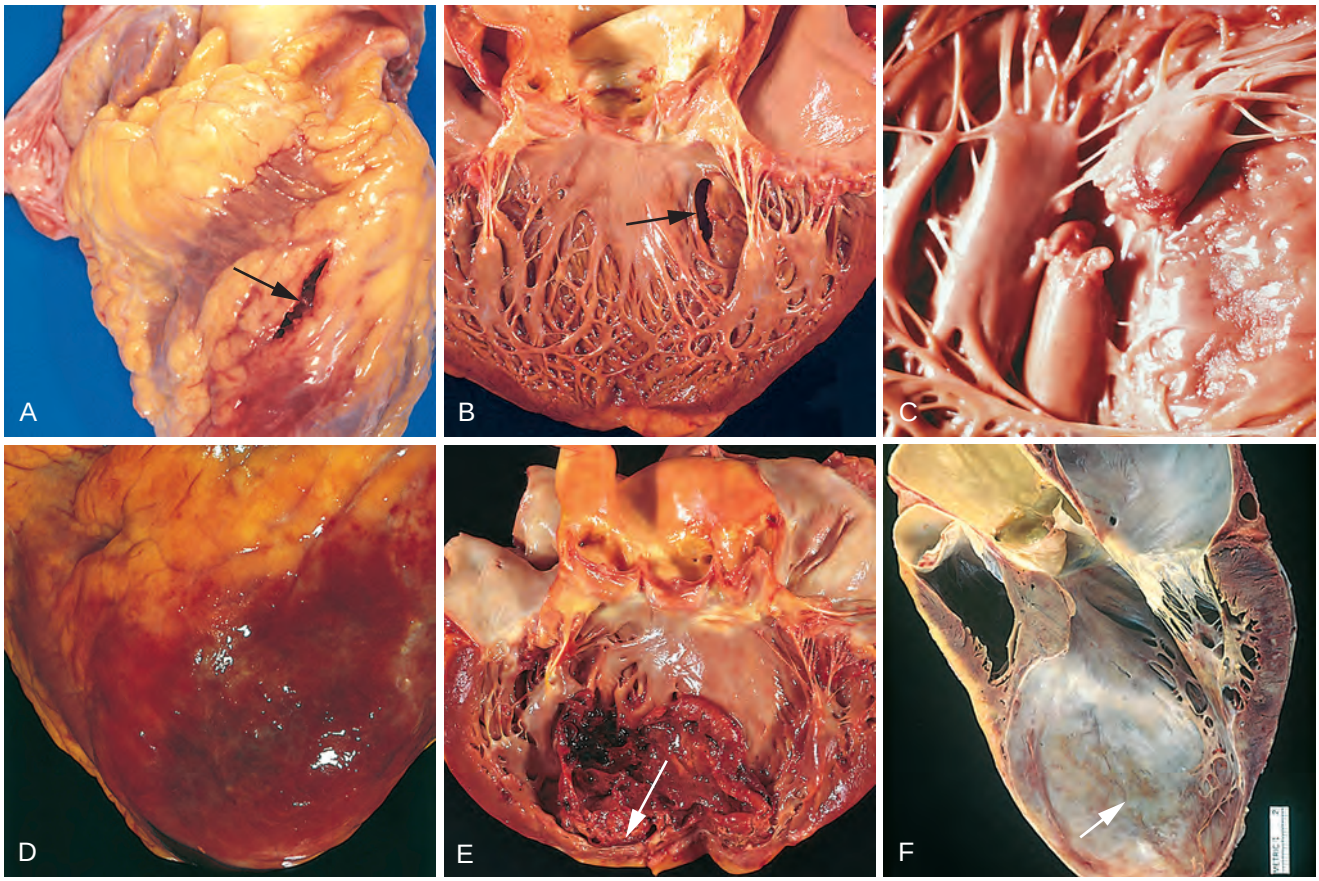


Figure 12.17 Complications of myocardial infarction. (A) Anterior myocardial rupture in an acute infarct *(arrow)*. (B) Rupture of the ventricular septum *(arrow)*. (C) Complete rupture of a necrotic papillary muscle. (D) Fibrinous pericarditis, showing a dark, roughened epicardial surface overlying an acute infarct. E, Early expansion of anteroapical infarct with wall thinning *(arrow)* and mural thrombus. F, Large apical left ventricular aneurysm *(arrow)*. The left ventricle is on the right in this apical four-chamber view of the heart. (A–E, Reproduced with permission from Schoen FJ: *Interventional and Surgical Cardiovascular Pathology: Clinical Correlations and Basic Principles*, Philadelphia, 1989, WB Saunders; F, Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, Minn.)

bradycardia, supraventricular tachyarrhythmias, ventricular premature contractions or ventricular tachycardia, and ventricular fibrillation. The risk for serious arrhythmias (e.g., ventricular fibrillation) is greatest in the first hour and declines thereafter.

- *Pericarditis.* Transmural MIs can elicit a fibrinohemorrhagic pericarditis; this is an epicardial manifestation of the underlying myocardial inflammation (see Fig. 12.17D). Heralded by anterior chest pain and a pericardial friction rub, pericarditis typically appears 2 to 3 days after infarction and then gradually resolves over the next few days. Extensive infarcts or severe pericardial inflammation occasionally can lead to large effusions or can organize to form dense adhesions that eventually manifest as a constrictive lesion. Rarely patients can develop an intense pericarditis weeks after MI *(Dressler syndrome)* due to formation of antibodies against damaged myocardium.
- *Chamber dilation.* Because of the weakening of necrotic muscle, there may be disproportionate stretching, thinning, and dilation of the infarcted region (especially with anteroseptal infarcts).
- *Mural thrombus.* With any infarct, the combination of attenuated myocardial contractility (causing stasis), chamber dilation, and endocardial damage (causing a thrombogenic surface) can foster mural thrombosis (see Fig. 12.17E), potentially leading to left-sided thromboembolism.
- *Ventricular aneurysm.* A late complication, aneurysms of the ventricle most commonly result from a large transmural anteroseptal infarct that heals with the formation of a thinned wall of scar tissue (see Fig. 12.17F). Although ventricular aneurysms frequently give rise to formation of mural thrombi, arrhythmias, and heart failure, they do not rupture.
- *Progressive heart failure.* This is discussed in the Chronic Ischemic Heart Disease section later in this chapter.

The risk of postinfarct complications and the prognosis of the patient depend primarily on the infarct size, location, and fraction of the wall thickness involved (subendocardial or transmural). Thus, large transmural infarcts yield a higher probability of cardiogenic shock, arrhythmias, and late CHF. Patients with anterior transmural infarcts are at greatest risk for free-wall rupture, expansion, mural thrombi, and aneurysm. In contrast, posterior transmural infarcts are more likely to be complicated by conduction blocks, right ventricular involvement, or both; when acute VSDs occur in this area they are more difficult to manage. Moreover, female gender, age older than 70 years, diabetes mellitus and previous MI are poor prognostic factors in patients with ST

elevation myocardial infarcts. With subendocardial infarcts, only rarely do pericarditis, rupture, and aneurysm occur.

In addition to the sequence of repair in the infarcted tissues described earlier, the noninfarcted segments of the ventricle undergo hypertrophy and dilation; collectively, these changes are termed ventricular remodeling. The compensatory hypertrophy of noninfarcted myocardium is initially hemodynamically beneficial. However, this adaptive effect may be overwhelmed by ventricular dilation (with or without ventricular aneurysm) and increased oxygen demand, which can exacerbate ischemia and depress cardiac function. There may also be changes in ventricular shape and stiffening of the ventricle due to scar formation and hypertrophy that further diminish cardiac output. Some of these deleterious effects appear to be reduced by ACE inhibitors, which lessen the ventricular remodeling that can occur after infarction.

Long-term prognosis after MI depends on many factors, the most important of which are the residual left ventricular function and the extent of any vascular obstructions in vessels that perfuse the remaining viable myocardium. The overall total mortality within the first year can be as high as 30%; thereafter, each passing year is associated with an additional 3% to 4% mortality among survivors. Infarct prevention (through control of risk factors) in individuals who have never experienced MI (primary prevention) and prevention of reinfarction in MI survivors (secondary prevention) are important strategies that have received much attention and achieved considerable success.

The relationship of the causes, pathophysiology, and consequences of MI are summarized in Fig. 12.18, including the possible outcomes of chronic IHD and sudden death, discussed later.

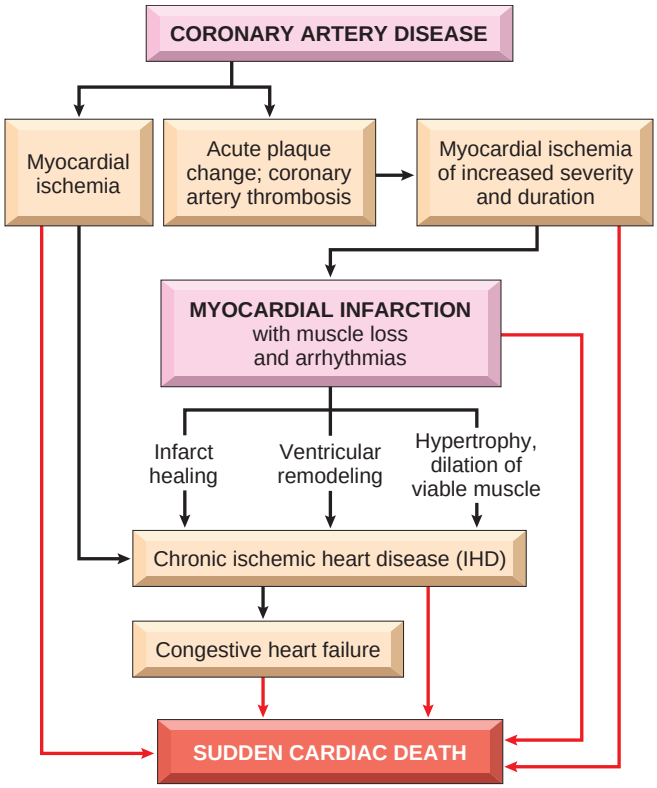


Figure 12.18 Schematic for the causes and outcomes of ischemic heart disease *(IHD)*, showing the interrelationships among coronary artery disease, acute plaque change, myocardial ischemia, myocardial infarction, chronic IHD, congestive heart failure, and sudden cardiac death.

Chronic Ischemic Heart Disease

The designation chronic IHD (often called ischemic cardiomyopathy by clinicians) is used here to describe progressive congestive heart failure as a consequence of accumulated ischemic myocardial damage and/or inadequate compensatory responses. In most instances, there has been a prior MI and sometimes previous coronary arterial interventions and/or bypass surgery. Chronic IHD usually appears postinfarction due to the functional decompensation of hypertrophied noninfarcted myocardium (see earlier discussion of cardiac hypertrophy). However, in other cases severe obstructive CAD may present as chronic congestive heart failure in the absence of prior infarction. Patients with chronic IHD account for almost 50% of cardiac transplant recipients.

MORPHOLOGY

Patients with chronic IHD have cardiomegaly, with left ventricular hypertrophy and dilation. Invariably there is some degree of stenotic coronary atherosclerosis. Discrete scars representing healed infarcts are usually present. The mural endocardium often has patchy fibrous thickenings (due to abnormal wall shear forces), and mural thrombi may be present. Microscopic findings include myocardial hypertrophy, diffuse subendocardial myocyte vacuolization, and interstitial fibrosis.

KEY CONCEPTS

ISCHEMIC HEART DISEASE

- The vast majority of IHD is due to coronary artery atherosclerosis. Vasospasm, vasculitis, or embolism are less common causes.
- Cardiac ischemia results from a mismatch in coronary supply and myocardial demand, and it presents as different, albeit overlapping, syndromes:
  - Angina pectoris is chest pain on exertion due to inadequate perfusion, and is typically due to atherosclerotic disease with greater than 70% fixed stenosis (so-called critical stenosis).
  - Unstable angina results from a small fissure or rupture of atherosclerotic plaque triggering platelet aggregation, vasoconstriction, and formation of a mural thrombus that needs not necessarily be occlusive. Chest pain may occur, even at rest or with minimal exertion.
  - Acute MI typically results from acute thromboses after plaque disruption; most occur in plaques that did not previously exhibit critical stenosis.
  - Chronic IHD is progressive heart failure due to ischemic injury, either from prior infarctions or chronic low-grade ischemia.
- Myocardial ischemia leads to loss of function within 1 to 2 minutes but causes necrosis only after 20 to 30 minutes. MI is diagnosed based on symptoms, electrocardiographic changes, and measurement of serum troponins. Gross and histologic changes of infarction require hours to days to develop.

- Infarction can be modified by therapeutic intervention (e.g., thrombolysis or stenting), which salvages myocardium at risk, but potentially induces reperfusion-related injury.
- Complications of infarction include: ventricular rupture, papillary muscle rupture, aneurysm formation, mural thrombus, arrhythmia, pericarditis, and CHF.

ARRHYTHMIAS

Abnormalities in myocardial conduction can be sustained or sporadic (paroxysmal). Aberrant rhythms can be initiated anywhere in the conduction system, from SA node down to the level of an individual myocyte; they are typically designated as originating from the atrium (supraventricular) or within the ventricular myocardium. Arrhythmias can manifest as tachycardia (fast heart rate), bradycardia (slow heart rate), an irregular rhythm with normal ventricular contraction, chaotic depolarization without functional ventricular contraction (ventricular fibrillation), or no electrical activity at all (asystole). Patients may be unaware of a rhythm disorder, or may note a "racing heart" or palpitations (irregular rhythm); loss of adequate cardiac output due to sustained arrhythmia can produce lightheadedness (near syncope), loss of consciousness (syncope), or sudden cardiac death (see later).

Ischemic injury is a common cause of rhythm disorders, either through direct damage or through the dilation of heart chambers that alters conduction system firing.

- *Sick sinus syndrome.* If the SA node is damaged, other fibers or even the AV node can take over pacemaker function, albeit at a much slower intrinsic rate (causing bradycardia).
- *Atrial fibrillation.* If the atrial myocytes become "irritable" and depolarize independently and sporadically (as occurs with atrial dilation), the signals are variably transmitted through the AV node, leading to the random "irregularly irregular" heart rate.
- *Heart block.* If the AV node is dysfunctional, varying degrees of heart block occur, ranging from simple prolongation of the P-R interval on electrocardiogram (first-degree heart block), to intermittent transmission of the signal (second-degree heart block), to complete failure (third-degree heart block).

As discussed earlier, coordinated cardiac contraction depends on the orderly transmission of electrical currents from myocyte to myocyte via gap junctions. Thus, abnormalities in the structure or spatial distribution of gap junctions, which are seen in a variety of disorders (e.g., IHD and dilated cardiomyopathies), can cause arrhythmias. Ischemia, myocyte hypertrophy, and inflammation (e.g., myocarditis or sarcoidosis) also promote increased "irritability" that leads to spontaneous aberrant myocyte depolarization; because of the electrical interconnection of myocytes, such random events can cause inappropriate firing of adjacent cells and create abnormal electrical circuits (so-called reentry circuits) that lead to ventricular tachycardia, which may progress to fatal ventricular fibrillation. Likewise, deposition of nonconducting material (e.g., amyloid) and even small areas of fibrosis can disrupt myocyte-to-myocyte signaling, again sowing the seeds for development of reentry circuits that can give rise to potentially fatal arrhythmias.

Heritable conditions associated with arrhythmias are important to recognize because they may alert physicians to the need for intervention to prevent sudden cardiac death (discussed later) in the proband and their family members. Some of these disorders are associated with recognizable anatomic abnormalities (e.g., congenital anomalies, hypertrophic cardiomyopathy, mitral valve prolapse). However, other heritable disorders precipitate arrhythmias and sudden death in the absence of structural cardiac pathology (so-called primary electrical disorders). These syndromes can only be diagnosed by genetic testing, which is performed in those with a positive family history or an unexplained nonlethal arrhythmia.

Channelopathies are the most important of the primary electrical abnormalities of the heart that predispose to arrhythmias (Table 12.6). Channelopathies are caused by mutations in genes that are required for normal ion channel function. These disorders (mostly with autosomal dominant inheritance) either involve genes that encode the structural components of ion channels (including Na^+, K^+, and Ca^+ channels) or accessory proteins that are essential for normal channel function. Ion channels are responsible for conducting the electrical currents that mediate contraction of the heart, and it is thus not surprising that defects in these channels may provoke arrhythmias. The prototype is the *long QT syndrome,* characterized by prolongation of the QT segment in ECGs with susceptibility to serious ventricular arrhythmias. Ion channels are needed for the normal function of

Table 12.6 Selected Examples of Causal Genes in Inherited Arrhythmogenic Diseases[a]

Disorder	Gene	Function
Long QT syndrome[b]	*KCNQ1*	K^+ channel (LOF)
	KCNH2	K^+ channel (LOF)
	SCN5A	Na^+ channel (GOF)
	CAV3	Caveolin, Na^+current (GOF)
Short QT syndrome[b]	*KCNQ1*	K^+ channel (GOF)
	KCNH2	K^+ channel (GOF)
Brugada syndrome[b]	*SCN5A*	Na^+ channel (LOF)
	CACNB2b	Ca^{++} channel (LOF)
	SCN1b	Na^+ channel (LOF)[a]
CPVT syndrome[b]	*RYR2*	Diastolic Ca^{++} release (GOF)
	CASQ2	Diastolic Ca^{++} release (LOF)

LOF, Loss-of-function mutations; *GOF,* gain-of-function mutations; *CPVT,* catecholaminergic polymorphic ventricular tachycardia.

[a]Different mutations can cause the same general syndrome, and mutations in some genes can cause multiple different phenotypes; thus, loss-of-function (LOF) mutations may cause long QT intervals, whereas gain-of-function (GOF) mutations result in short repolarization intervals.

[b]*Long QT syndrome* manifests as arrhythmias associated with excessive prolongation of the cardiac repolarization; patients often present with stress-induced syncope or sudden cardiac death (SCD), and some forms are associated with swimming. *Short QT syndrome* patients have arrhythmias associated with abbreviated repolarization intervals; they can present with palpitations, syncope, and SCD. *Brugada syndrome* manifests as ECG abnormalities (ST segment elevations and right bundle branch block) in the absence of structural heart disease; patients classically present with syncope or SCD during rest or sleep, or after large meals. *CPVT* does not have characteristic ECG changes; patients often present in childhood with life-threatening arrhythmias due to adrenergic stimulation (stress-related).

Modified from Cerrone M, Priori SG: Genetics of sudden death: focus on inherited channelopathies, *Eur Heart J* 32(17), 2109–2118, 2011.

many tissues, and certain channelopathies are also associated with skeletal muscle disorders and diabetes. Nevertheless, the most common channelopathies are isolated disorders of the heart, and their most feared consequence is sudden cardiac death (discussed in the next section).

Sudden Cardiac Death

SCD is most commonly defined as unexpected death from cardiac causes, either without symptoms or within 1 to 24 hours of symptom onset (different authors use different criteria); this happens in some 180,000 to 450,000 individuals each year in the United States alone. **The mechanism of SCD is most often a lethal arrhythmia (e.g., asystole or ventricular fibrillation).** Although ischemic injury (and other pathologies) can directly affect the major components of the conduction system, most cases of fatal arrhythmia are triggered by electrical irritability of myocardium distant from the major elements of the conduction system.

CAD is the leading cause of SCD, responsible for the majority of cases; unfortunately, SCD can be the first manifestation of IHD. In such cases, there is typically only chronic severe atherosclerotic disease with fixed critical stenoses; acute plaque disruption is found in only 10% to 20% of cases, and 80% to 90% of patients who suffer SCD but are successfully resuscitated do not show any enzymatic or ECG evidence of myocardial necrosis. Healed remote MIs are present in about 40%, and subendocardial myocyte vacuolization indicative of severe chronic ischemia is a common feature. With the decrease of prevalence of IHD in high income countries, SCD is increasingly being seen in individuals with hearts that are hypertrophic and fibrotic (from hypertension, obesity, substance abuse, etc.) without IHD.

With younger victims, nonatherosclerotic causes are more common etiologies for SCD:

- *Cardiac conduction abnormalities,* hereditary or acquired
- *Dilated* or *hypertrophic cardiomyopathy*
- *Congenital coronary arterial abnormalities*
- *Myocarditis*
- *Mitral valve prolapse*
- *Pulmonary hypertension*
- *Other miscellaneous causes,* such as pericardial tamponade, pulmonary embolism, systemic metabolic and hemodynamic alterations, catecholamines, and drugs of abuse, particularly cocaine and methamphetamine.

The prognosis of patients vulnerable to SCD is markedly improved by pharmaceutical intervention, and particularly by implantation of automatic cardioverter defibrillators that can sense and electrically counteract episodes of ventricular fibrillation.

KEY CONCEPTS

ARRHYTHMIAS

- Arrhythmias can be caused by ischemic or structural changes in the conduction system or by intrinsic myocyte electrical instability. In structurally normal hearts, arrhythmias are often due to mutations in ion channels that cause aberrant repolarization or depolarization.
- SCD typically results from ventricular fibrillation and is most frequently a consequence of CAD. Myocardial irritability typically results from nonlethal ischemia or from preexisting fibrosis from previous myocardial injury. SCD is less often due to acute plaque rupture with thrombosis that induces a rapidly fatal arrhythmia.

HYPERTENSIVE HEART DISEASE

Hypertensive heart disease (HHD) is a consequence of the increased demands placed on the heart by hypertension, causing pressure overload and ventricular hypertrophy. Although most commonly seen in the left heart as the result of systemic hypertension, pulmonary hypertension can cause right-sided HHD, or cor pulmonale.

Systemic (Left-Sided) Hypertensive Heart Disease

Hypertrophy of the heart is an adaptive response to the pressure overload of chronic hypertension. However, such compensatory changes may be ultimately maladaptive and can lead to myocardial dysfunction, myocardial "demand" ischemia, cardiac dilation, CHF, and in some cases sudden death.

The minimal pathologic criteria for the diagnosis of systemic HHD are the following: (1) left ventricular hypertrophy (usually concentric) in the absence of other cardiovascular pathology and (2) a clinical history or pathologic evidence of hypertension in other organs (e.g., kidney). The Framingham Study established unequivocally that even mild hypertension (levels only slightly above 140/90 mm Hg)—if sufficiently prolonged—induces left ventricular hypertrophy; some 30% of the population of the United States suffers from hypertension of at least this degree. More recently, the criteria for diagnosing hypertension were changed; according to the newest guidelines, diastolic pressures above 80 mm Hg, or systolic pressures above 120 mm Hg constitute clinically significant elevations in blood pressure. On that basis, almost one-half of individuals in the general population (!!) are hypertensive. The pathogenesis of hypertension is discussed in Chapter 11.

MORPHOLOGY

Hypertension induces left ventricular pressure overload hypertrophy, initially without ventricular dilation. As a result, the left ventricular wall thickening increases the weight of the heart disproportionately to the increase in overall cardiac size (Fig. 12.19A). The heart weight may exceed 500 g, and the left ventricular wall thickness may exceed 2.0 cm. Over time, the increased thickness of the left ventricular wall, associated with increased interstitial connective tissue, imparts a stiffness that impairs diastolic filling, often leading to left atrial enlargement.

Microscopically, the earliest change of systemic HHD is an increase in the transverse diameter of myocytes, which may be difficult to appreciate on routine microscopy. At a more advanced stage, variable degrees of cellular and nuclear enlargement become apparent, often accompanied by perivascular and interstitial fibrosis.

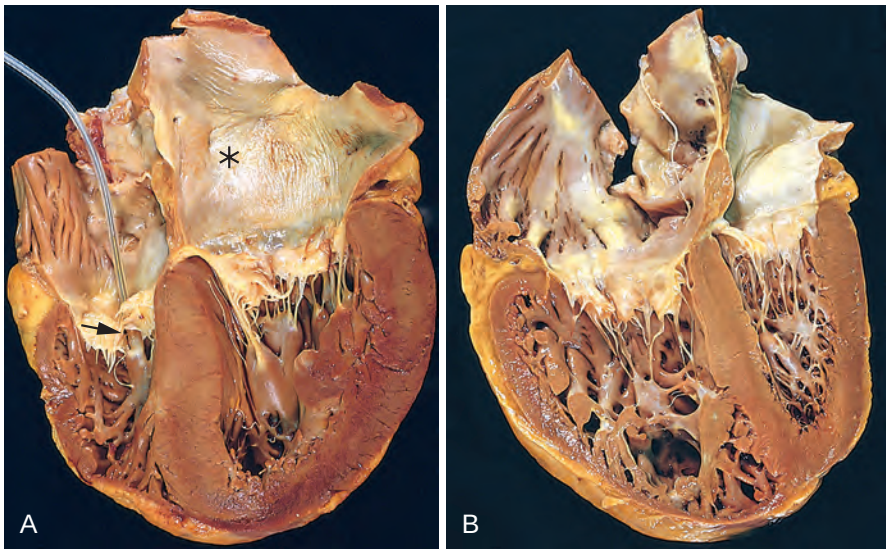


Figure 12.19 Hypertensive heart disease, systemic and pulmonary. (A) Systemic (left-sided) hypertensive heart disease. There is marked concentric thickening of the left ventricular wall causing reduction in lumen size. The left ventricle and left atrium *(asterisk)* are on the *right* in this apical four-chamber view of the heart. A pacemaker is present in the right ventricle *(arrow)*. (B) Pulmonary (right-sided) hypertensive heart disease (cor pulmonale). The right ventricle is markedly dilated and has a thickened free wall and hypertrophied trabeculae (apical four-chamber view of heart, right ventricle on left). The shape of the left ventricle *(to the right)* has been distorted by the enlarged right ventricle.

Compensated systemic HHD may be asymptomatic, producing only electrocardiographic or echocardiographic evidence of left ventricular enlargement. In many patients, systemic HHD comes to attention because of new atrial fibrillation induced by left atrial enlargement, or by progressive CHF. Depending on the severity, duration, and underlying basis of the hypertension, and on the adequacy of therapeutic control, the patient may (1) enjoy normal longevity and die of unrelated causes, (2) develop IHD due to both the potentiating effects of hypertension on coronary atherosclerosis and the ischemia induced by increased oxygen demand from the hypertrophic muscle, (3) suffer renal damage or cerebrovascular stroke as direct effects of hypertension, or (4) experience progressive heart failure or SCD. Effective control of hypertension can prevent cardiac hypertrophy and can even lead to its regression; with normalization of blood pressure, the associated risks of HHD are diminished.

Pulmonary (Right-Sided) Hypertensive Heart Disease (Cor Pulmonale)

Normally, because the pulmonary vasculature is the low-pressure side of the circulation, the right ventricle has a thinner and more compliant wall than the left ventricle. Isolated pulmonary HHD, or cor pulmonale, stems from right ventricular pressure overload. Chronic cor pulmonale is characterized by right ventricular hypertrophy, dilation, and potentially right-sided failure. Typical causes of chronic cor pulmonale are disorders of the lungs, especially chronic parenchymal diseases such as emphysema, and primary pulmonary hypertension (Table 12.7; see also Chapter 15). Acute cor pulmonale can follow massive pulmonary embolism. Nevertheless, it should also be remembered that pulmonary hypertension most commonly occurs as a complication of left-sided heart disease.

MORPHOLOGY

In acute cor pulmonale there is marked dilation of the right ventricle without hypertrophy. On cross-section, the normal crescent shape of the right ventricle chamber is transformed to a dilated ovoid. In chronic cor pulmonale, the right ventricular wall thickens, sometimes up to 1.0 cm or more (see Fig. 12.19B). More subtle right ventricular hypertrophy may take the form of thickening of the muscle bundles in the outflow tract, immediately

Table 12.7 Disorders Predisposing to Cor Pulmonale

Diseases of the Pulmonary Parenchyma
Chronic obstructive pulmonary disease
Diffuse pulmonary interstitial fibrosis
Pneumoconioses
Cystic fibrosis
Bronchiectasis
Diseases of the Pulmonary Vessels
Recurrent pulmonary thromboembolism
Primary pulmonary hypertension
Extensive pulmonary arteritis (e.g., granulomatosis with polyangiitis)
Drug-, toxin-, or radiation-induced vascular obstruction
Extensive pulmonary tumor microembolism
Disorders Affecting Chest Movement
Kyphoscoliosis
Marked obesity (sleep apnea, pickwickian syndrome)
Neuromuscular diseases
Disorders Inducing Pulmonary Arterial Constriction
Metabolic acidosis
Hypoxemia
Chronic altitude sickness
Obstruction of major airways
Idiopathic alveolar hypoventilation

below the pulmonary valve, or thickening of the moderator band, the muscle bundle that connects the ventricular septum to the anterior right ventricular papillary muscle. Sometimes, the hypertrophied right ventricle compresses the left ventricular chamber or leads to regurgitation and fibrous thickening of the tricuspid valve.

KEY CONCEPTS

HYPERTENSIVE HEART DISEASE

- HHD can affect either the left ventricle or the right ventricle; the latter is called cor pulmonale. Elevated pressures induce myocyte hypertrophy and interstitial fibrosis that increases wall thickness and myocardial stiffness.
- The chronic pressure overload of systemic hypertension causes left ventricular concentric hypertrophy, often associated with left atrial dilation due to impaired diastolic filling of the ventricle. Persistently elevated pressure overload can cause ventricular failure with dilation.
- Cor pulmonale results from pulmonary hypertension due to primary lung parenchymal or vascular disorders. There is commonly right ventricular and right atrial hypertrophy; right ventricular and atrial dilation can occur.

VALVULAR HEART DISEASE

Valvular disease can come to clinical attention due to stenosis, insufficiency (synonyms: regurgitation or incompetence), or both.

- *Stenosis* is the failure of a valve to open completely, obstructing forward flow. Acquired valvular stenosis is almost always due to a primary leaflet abnormality and is virtually always a chronic process (e.g., calcification or valve scarring).
- *Insufficiency* results from failure of a valve to close completely, thereby allowing regurgitation (backflow) of blood. Valvular insufficiency can result from either intrinsic disease of the valve leaflets (e.g., endocarditis) or disruption of the supporting structures (e.g., the aorta, mitral annulus, tendinous cords, papillary muscles, or ventricular free wall) without primary leaflet injury. It can appear abruptly, as with chordal rupture, or insidiously as a consequence of leaflet scarring and retraction.

Stenosis or insufficiency can occur alone or together in the same valve. Valvular disease can involve only one valve (the mitral valve being the most common target) or more than one valve. Abnormal flow through diseased valves typically produces abnormal heart sounds called murmurs; severe lesions can even be externally palpated as thrills. Depending on the valve involved, murmurs are best heard at different locations on the chest wall; moreover, the nature (regurgitation versus stenosis) and severity of the valvular disease determines the quality and timing of the murmur (e.g., harsh systolic or soft diastolic murmurs).

The clinical consequences of valve dysfunction vary depending on the valve involved, the degree of impairment, the tempo of disease onset, and the rate and quality of compensatory mechanisms. For example, sudden destruction of an aortic valve cusp by infection (infectious endocarditis; see later) can cause acute, massive, and rapidly fatal aortic regurgitation. In contrast, rheumatic mitral stenosis typically develops indolently over years, and its clinical effects can be well tolerated for extended periods. Certain conditions can complicate valvular heart disease by increasing the demands on the heart; for example, the increased output demands of pregnancy can exacerbate valve disease and lead to unfavorable maternal or fetal outcomes. Valvular stenosis or insufficiency often produces secondary changes, both proximal and distal to the affected valve, particularly in the myocardium. Generally, valvular stenosis leads to pressure overload cardiac hypertrophy, whereas valvular insufficiency leads to volume overload; both situations can culminate in heart failure. In addition, the ejection of blood through narrowed stenotic valves can produce high speed "jets" of blood that injure the endothelium or endocardium where they impact.

Valvular abnormalities can be congenital (discussed earlier) or acquired. Acquired valvular stenosis is almost always a consequence of a remote or chronic injury of the valve leaflets that declares itself clinically only after many years. In contrast, acquired valvular insufficiency can result from intrinsic disease of the valve leaflets or damage to or distortion of the supporting structures (e.g., the aorta, mitral annulus, tendinous cords, papillary muscles, ventricular free wall).

The causes of acquired heart valve diseases are summarized in Table 12.8. The most frequent causes of the major valvular lesions are as follows:

- *Aortic stenosis:* calcification and sclerosis of anatomically normal or congenitally bicuspid aortic valves
- *Aortic insufficiency:* dilation of the ascending aorta, often secondary to hypertension and/or aging
- *Mitral stenosis:* rheumatic heart disease (RHD)
- *Mitral insufficiency:* myxomatous degeneration (MVP), or left ventricular dilation due to ischemic or nonischemic heart failure

Calcific Valvular Degeneration

Heart valves are subjected to high levels of repetitive mechanical stress, particularly at the hinge points of the leaflets; this is a consequence of (1) millions of cardiac contractions per year, (2) substantial tissue deformations during each contraction, and (3) transvalvular pressure gradients in the closed phase of each contraction of approximately 120 mm Hg for the mitral and 80 mm Hg for the aortic valve. It is therefore not surprising that these delicate structures can suffer cumulative damage and calcification that lead to clinically important dysfunction.

Calcific Aortic Stenosis

The most common of all valvular abnormalities, calcific aortic stenosis is usually the consequence of age-associated "wear and tear" of either anatomically normal valves or congenitally bicuspid valves (found in approximately 1% of the population). The prevalence of aortic stenosis is estimated at 2% and is increasing as the general population ages. Aortic stenosis of previously normal valves (termed degenerative calcific aortic stenosis) usually comes to clinical attention in the seventh to ninth decades of life, whereas

Table 12.8 Etiology of Acquired Heart Valve Disease

Mitral Valve Disease	Aortic Valve Disease
Mitral Stenosis	**Aortic Stenosis**
Postinflammatory scarring (rheumatic heart disease)	Postinflammatory scarring (rheumatic heart disease) Senile calcific aortic stenosis Calcification of congenitally deformed valve
Mitral Regurgitation	**Aortic Regurgitation**
Abnormalities of leaflets and commissures Postinflammatory scarring Infective endocarditis Mitral valve prolapse "Fen-phen"–induced valvular fibrosis Abnormalities of tensor apparatus Rupture of papillary muscle Papillary muscle dysfunction (fibrosis) Rupture of chordae tendineae Abnormalities of left ventricular cavity and/or annulus Left ventricular enlargement (myocarditis, dilated cardiomyopathy) Calcification of mitral ring	Intrinsic valvular disease Postinflammatory scarring (rheumatic heart disease) Infective endocarditis Aortic disease Degenerative aortic dilation Syphilitic aortitis Ankylosing spondylitis Rheumatoid arthritis Marfan syndrome

Fen-phen, Fenfluramine-phentermine.

Data from Schoen FJ: Surgical pathology of removed natural and prosthetic valves, *Hum Pathol* 18(6):558–567, 1987.

stenotic bicuspid valves tend to become clinically significant one to two decades earlier.

Aortic valve calcification is likely a consequence of recurrent chronic injury due to hyperlipidemia, hypertension, inflammation, and other factors similar to those implicated in atherosclerosis. The chronic progressive injury leads to valvular degeneration and incites the deposition of hydroxyapatite (the same calcium salt found in bone). Although this model provides a good starting point for understanding calcific degeneration, it is increasingly clear that the valve injury of calcific aortic stenosis differs in some important respects from atherosclerosis. Most notably, the abnormal valves contain cells resembling osteoblasts that synthesize bone matrix proteins and promote the deposition of calcium salts. Moreover, interventions that improve atherosclerotic risk (e.g., statins) do not appear to significantly prevent valvular calcific degeneration. Bicuspid valves incur greater mechanical stress than normal tricuspid valves, which may explain their accelerated stenosis.

MORPHOLOGY

The gross morphologic hallmark of nonrheumatic, calcific aortic stenosis (involving either tricuspid or bicuspid valves) is mounded calcified masses on the outflow surfaces of the cusps that ultimately prevent cuspal opening. The free edges of the cusps are usually not involved (Fig. 12.20A). Microscopically, the layered architecture of the valve is largely preserved. The calcific process begins in the valvular fibrosa on the outflow surface of the valve, at the points of maximal cusp flexion (near the margins of attachment). Inflammation is variable, and metaplastic bone can be seen. In aortic stenosis, the functional valve area is decreased by large nodular calcific deposits that can eventually cause measurable outflow obstruction; this subjects the left ventricular myocardium to progressively increasing pressure overload. In contrast with rheumatic (and congenital) aortic stenosis (see Fig. 12.22E), commissural fusion is not usually seen.

Clinical Features

In calcific aortic stenosis (superimposed on a previously normal or bicuspid aortic valve), the obstruction to left ventricular outflow leads to gradual narrowing of the valve orifice (valve area approximately 0.5 to 1 cm^2 in severe aortic stenosis; normal is approximately 4 cm^2) and an increasing pressure gradient across the calcified valve. Left ventricular pressures rise to 200 mm Hg or more in such instances, producing concentric left ventricular (pressure overload) hypertrophy. The hypertrophied myocardium tends to be ischemic (as a result of diminished microcirculatory perfusion, often complicated by coronary atherosclerosis), and angina pectoris may occur. Both systolic and diastolic myocardial function may be impaired; eventually, cardiac decompensation and CHF can ensue. The onset of symptoms (angina, CHF, or syncope) in aortic stenosis heralds cardiac

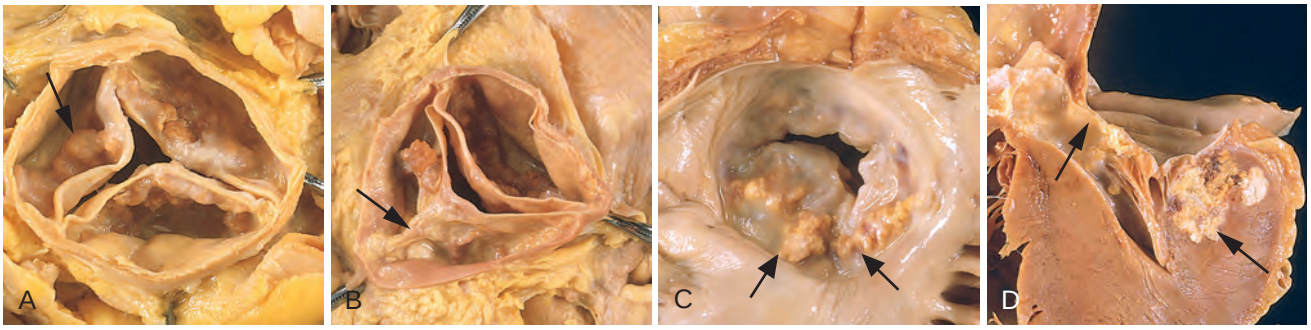


Figure 12.20 Calcific valvular degeneration. (A) Calcific aortic stenosis of a previously normal valve (viewed from aortic aspect). Nodular masses of calcium are heaped up within the sinuses of Valsalva *(arrow)*. Note that the commissures are not fused, as occurs with postrheumatic aortic valve stenosis (see Fig. 12.22E). (B) Calcific aortic stenosis of a congenitally bicuspid valve. One cusp has a partial fusion at its center, called a raphe *(arrow)*. (C and D) Mitral annular calcification, with calcific nodules at the base (attachment margin) of the anterior mitral leaflet *(arrows)*. (C) Left atrial view. (D) Cut section of myocardium showing the lateral wall with dense calcification that extends into the underlying myocardium *(arrow)*.

decompensation and carries an extremely poor prognosis. If untreated, most patients with aortic stenosis will die within 5 years of developing angina, within 3 years of developing syncope, and within 2 years of CHF onset. Treatment requires surgical valve replacement, as medical therapy is ineffective in severe symptomatic aortic stenosis.

Calcific Stenosis of Congenitally Bicuspid Aortic Valve

Bicuspid aortic valve (BAV) is a developmental abnormality with prevalence in the population of approximately 1%; calcified BAV comprises approximately 50% of cases of aortic stenosis in adults. Some cases of BAV show familial clustering, often with associated aortic or left ventricular outflow tract malformations. Loss-of-function mutations in *NOTCH1* (mapping to chromosome 9q34.3) have been specifically associated with BAV in a few families.

In a congenitally bicuspid aortic valve, there are only two functional cusps, usually of unequal size, with the larger cusp having a midline raphe, resulting from incomplete commissural separation during development; less frequently, the cusps are of equal size and the raphe is absent. The raphe is frequently a major site of calcific deposits (see Fig. 12.20B).

Although BAV is usually asymptomatic early in life, the onset of calcific degeneration heralds a clinical course similar to that described earlier for calcific aortic stenosis. Other bicuspid valve complications include aortic regurgitation and infective endocarditis. Interestingly, structural abnormalities of the aortic wall also commonly accompany BAV, even when the valve is hemodynamically normal, and this may potentiate aortic dilation or aortic dissection.

Mitral Annular Calcification

As opposed to the predominantly cuspal involvement in aortic valve calcification, degenerative calcific deposits in the mitral valve typically develop in the fibrous annulus. Grossly, these appear as irregular, stony hard, occasionally ulcerated nodules (2 to 5 mm in thickness) at the base of the leaflets (see Fig. 12.20C, D). Mitral annular calcification usually does not affect valvular function. However, in exceptional cases it can lead to the following:

- *Regurgitation* by interfering with physiologic contraction of the valve ring
- *Stenosis* by impairing opening of the mitral leaflets
- *Arrhythmias* and occasionally sudden death by penetration of calcium deposits to a depth sufficient to impinge on the atrioventricular conduction system

Because calcific nodules may also provide a site for thrombus formation, patients with mitral annular calcification have a mildly increased risk of embolic stroke, and the calcific nodules can become a nidus for infective endocarditis. Mitral annular calcification increases with age and is more common in women and individuals with mitral valve prolapse (see later).

Mitral Valve Prolapse (Myxomatous Degeneration of the Mitral Valve)

In mitral valve prolapse (MVP), one or both mitral valve leaflets are "floppy" and protrude into the left atrium during systole. MVP affects approximately 2% to 3% of adults in the United States and is more common in women; it is most often an incidental finding on physical examination, but may lead to serious complications in a small minority of affected individuals.

Pathogenesis

The etiologic basis for the changes that weaken the valve leaflets and associated structures is unknown in most cases. Uncommonly, MVP is associated with heritable disorders of connective tissue including Marfan syndrome, caused by fibrillin-1 *(FBN1)* mutations that alter cell-matrix interactions and dysregulate TGF-β signaling (Chapter 5). Interestingly, mice with mutated *FBN1* develop a form of MVP that is prevented by TGF-β inhibitors, indicating that excess TGF-β activity can cause the characteristic structural laxity and myxomatous changes.

MORPHOLOGY

The characteristic anatomic change in MVP is ballooning (hooding) of the mitral leaflets (Fig. 12.21A–C). The affected leaflets are often enlarged, redundant, thick, and rubbery. The associated tendinous cords may be elongated, thinned, or even ruptured, and the annulus may be dilated. The tricuspid, aortic, or pulmonary valves may also be affected. The key histologic change in the tissue is marked **myxomatous degeneration** of the spongiosa layer, reflected by increased deposition of a highly sulfated hydrophilic matrix (see Fig. 12.21E); the collagenous fibrosa layer of the valve is also attenuated, affecting the structural integrity of the leaflet. Secondary changes reflect the stresses and tissue injury incident to the billowing leaflets: (1) fibrous thickening of the valve leaflets, particularly where they rub against each other; (2) linear fibrous thickening of the left ventricular endocardial surface where the abnormally long cords snap or rub against it; (3) thickening of the mural endocardium of the left ventricle or atrium as a consequence of friction-induced injury induced by the prolapsing, hypermobile leaflets; (4) thrombi on the atrial surfaces of the leaflets or the atrial walls (Fig. 12.21B); and (5) focal calcifications at the base of the posterior mitral leaflet (Fig. 12.21C). Notably, mitral valve myxomatous degeneration can also occur as a secondary consequence of regurgitation of other etiologies (e.g., ischemic dysfunction).

Clinical Features

Most individuals diagnosed with MVP are asymptomatic; in such cases, the condition is discovered incidentally by auscultation of mid-systolic clicks, caused by abrupt tension on the redundant valve leaflets and chordae tendineae as the valve attempts to close; there may or may not be an associated regurgitant murmur. The diagnosis is confirmed by echocardiography. A minority of patients have chest pain mimicking angina (although not exertional in nature), and a subset has dyspnea, presumably related to valvular insufficiency. Although the great majority of persons with MVP have no untoward effects, approximately 3% develop one of four serious complications: (1) infective endocarditis; (2) mitral insufficiency, sometimes with chordal rupture; (3) stroke or other systemic infarct, resulting from embolism of leaflet thrombi; or (4) arrhythmias, both ventricular and atrial. Rarely, MVP is the only finding in sudden cardiac

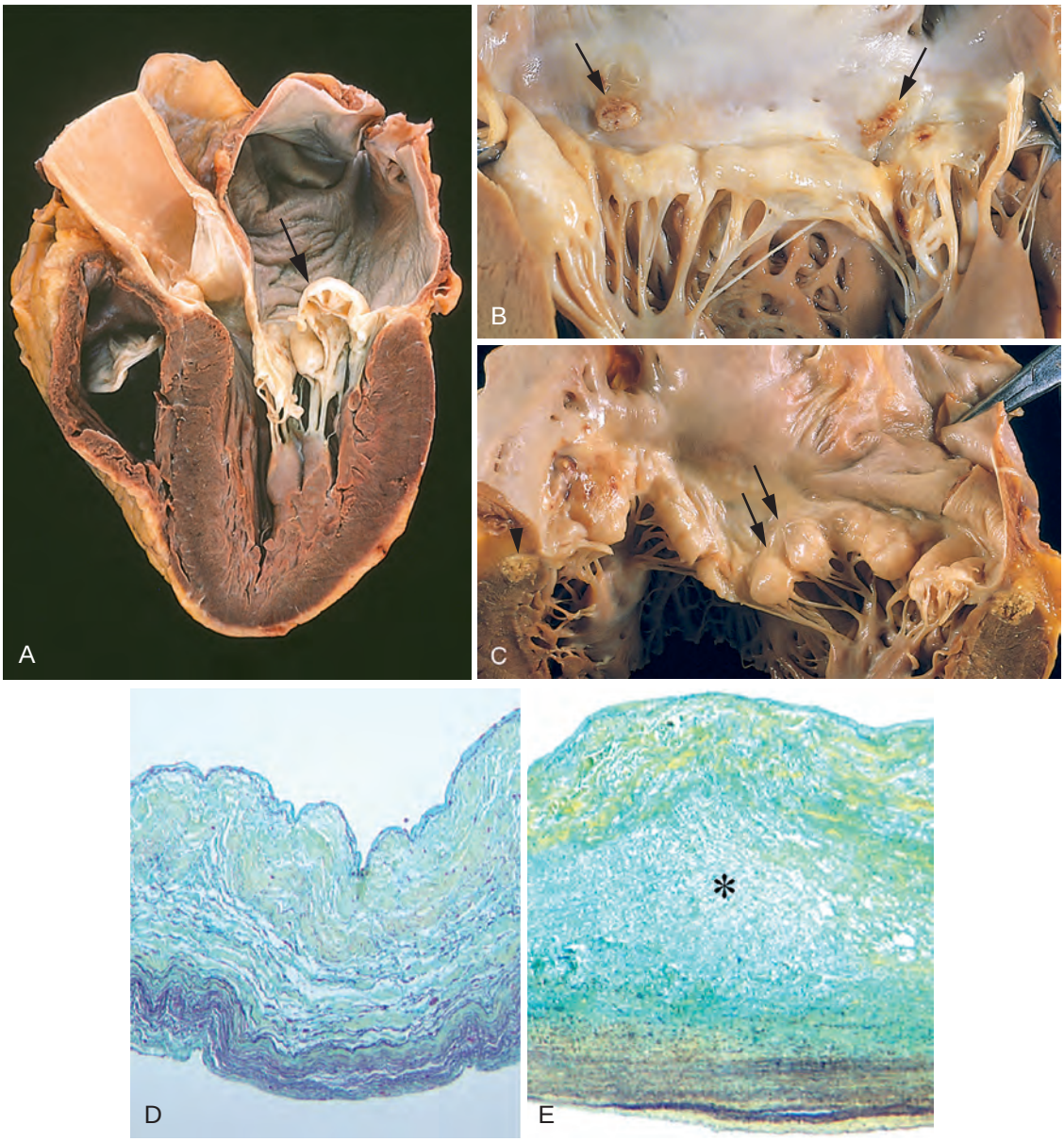


Figure 12.21 Myxomatous degeneration of the mitral valve. (A) Long axis view (left ventricle is on the *right*) demonstrating hooding with prolapse of the posterior mitral leaflet into the left atrium *(arrow)*. (B) Opened valve, showing pronounced hooding of the posterior mitral leaflet with thrombotic plaques at sites of leaflet-left atrium contact *(arrows)*. (C) Opened valve with pronounced hooding *(double arrows)* in a patient who died suddenly. Note also mitral annular calcification on the left side *(arrowhead)*. Normal heart valve (D) and myxomatous mitral valve (E). In myxomatous valves, collagen in the fibrosa layer is loose and disorganized, proteoglycan deposition *(asterisk)* in the central spongiosa layer is markedly expanded, and elastin in the atrialis layer is disorganized. (A, Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, Minn; D and E, Movat pentachrome stain, in which collagen is yellow, elastin is black, and proteoglycans are blue). From Rabkin E, et al: Activated interstitial myofibroblasts express catabolic enzymes and mediate matrix remodeling in myxomatous heart valves, *Circulation* 104:2525–2532, 2001.)

death. Valve repair or replacement surgery can be done for symptomatic patients or those with increased risk for significant complications.

Rheumatic Fever and Rheumatic Heart Disease

Rheumatic fever (RF) is an acute, immunologically mediated, multisystem inflammatory disease classically occurring a few weeks after group A streptococcal pharyngitis; occasionally, RF can follow streptococcal infections at other sites, such as the skin. Acute rheumatic carditis is a common manifestation of active RF and may progress over time to chronic rheumatic heart disease (RHD).

RHD is characterized principally by deforming fibrotic valvular disease, particularly involving the mitral valve; indeed, RHD is virtually the only cause of mitral stenosis. The incidence and mortality rate of RF and RHD have declined remarkably in many parts of the world over the past century, as a result of improved sanitation, and rapid diagnosis and treatment of streptococcal pharyngitis.

Nevertheless, in low income countries, and in many crowded, economically depressed urban areas, RHD remains an important public health problem.

Pathogenesis

Acute rheumatic fever results from host immune responses to group A streptococcal antigens that cross-react with host proteins. The characteristic 2- to 3-week delay in symptom onset after infection is explained by the time needed to generate an immune response; In particular, antibodies and CD4+ T cells directed against streptococcal M proteins can also in some cases recognize cardiac self antigens. Antibody binding can activate complement, as well as recruit Fc-receptor–bearing cells (neutrophils and macrophages); cytokine production by the stimulated T cells leads to macrophage activation (e.g., within Aschoff bodies). Damage to heart tissue may thus be caused by a combination of antibody- and T cell–mediated reactions (Chapter 6). In keeping with an immunologic basis of RHD, streptococci are completely absent from the lesions. Because only a small minority of infected patients develop rheumatic fever (estimated at 3%), a genetic susceptibility is likely to influence the development of the cross-reactive immune responses. The chronic fibrotic lesions are the predictable consequence of healing and scarring associated with the resolution of the acute inflammation.

MORPHOLOGY

Key pathologic features of acute RF and chronic RHD are shown in Fig. 12.22. During acute RF, focal inflammatory lesions are found in various tissues. Distinctive lesions in the heart—called **Aschoff bodies**—are composed of foci of T lymphocytes, occasional plasma cells, and plump activated macrophages called **Anitschkow cells.** These macrophages have abundant cytoplasm and central round-to-ovoid nuclei (occasionally binucleate) in which the chromatin condenses into a central, slender, wavy ribbon (hence the designation "caterpillar cells").

During acute RF, diffuse inflammation and Aschoff bodies may be found in any of the three layers of the heart, resulting in pericarditis, myocarditis, or endocarditis **(pancarditis).**

Inflammation of the endocardium and the left-sided valves typically results in fibrinoid necrosis within the cusps or tendinous

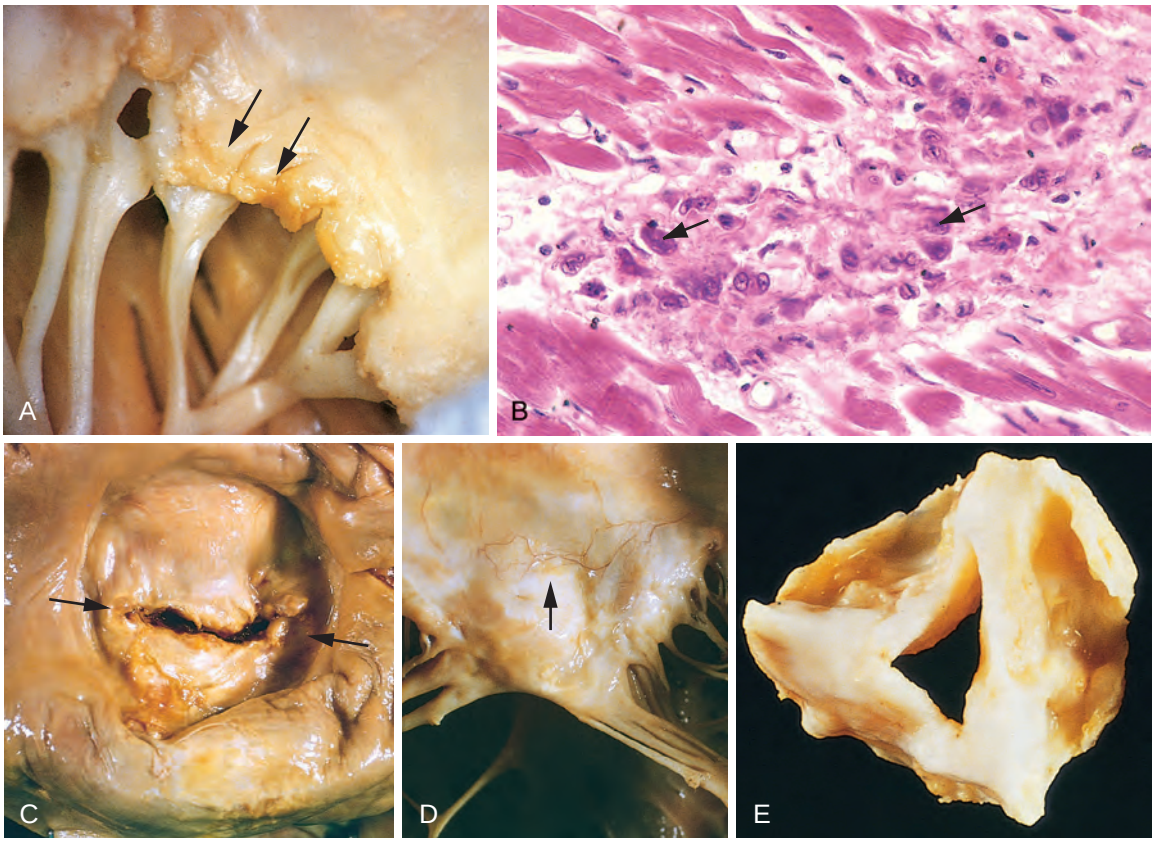


Figure 12.22 Acute and chronic rheumatic heart disease. (A) Acute rheumatic mitral valvulitis superimposed on chronic rheumatic heart disease. Small vegetations (verrucae) are visible along the line of closure of the mitral valve leaflet *(arrows)*. Previous episodes of rheumatic valvulitis have caused fibrous thickening and fusion of the chordae tendineae. (B) Microscopic appearance of an Aschoff body in a patient with acute rheumatic carditis. The myocardium exhibits a circumscribed nodule of mixed mononuclear inflammatory cells with associated necrosis; within the inflammation, large activated macrophages show prominent nucleoli, as well as chromatin condensed into long, wavy ribbons (caterpillar cells; *arrows*). (C and D) Mitral stenosis with diffuse fibrous thickening and distortion of the valve leaflets and commissural fusion (*arrows,* C), and thickening of the chordae tendineae (D). Note the neovascularization of the anterior mitral leaflet (*arrow,* D). (E) Surgically resected specimen of rheumatic aortic stenosis, demonstrating thickening and distortion of the cusps with commissural fusion. (E, Reproduced from Schoen FJ, St. John-Sutton M: Contemporary issues in the pathology of valvular heart disease, *Hum Pathol* 18:568, 1967.)

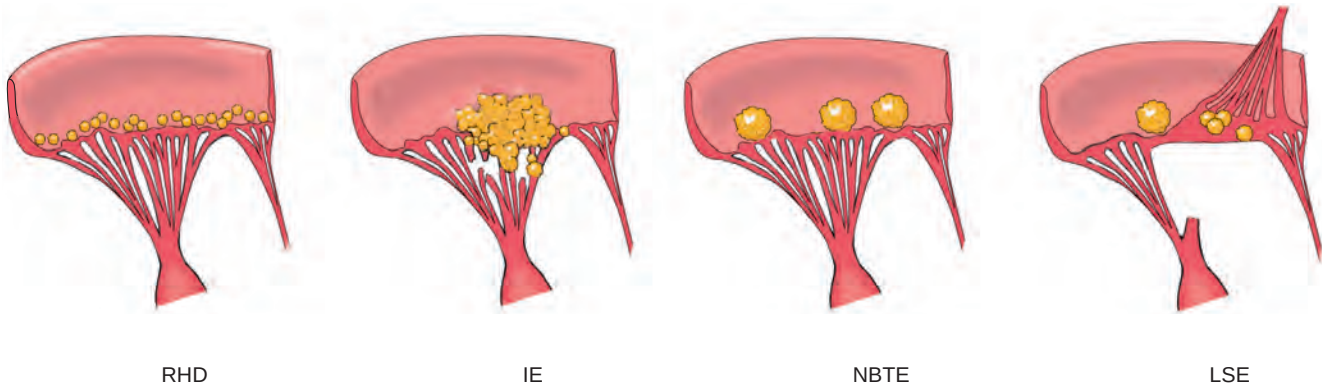


Figure 12.23 Comparison of the four major forms of vegetative endocarditis. The rheumatic fever phase of rheumatic heart disease *(RHD)* is marked by small, warty vegetations along the lines of closure of the valve leaflets. Infective endocarditis *(IE)* is characterized by large, irregular masses on the valve cusps that can extend onto the chordae (see Fig. 12.24A). Nonbacterial thrombotic endocarditis *(NBTE)* typically exhibits small, bland vegetations, usually attached at the line of closure. One or many may be present (see Fig. 12.25). Libman-Sacks endocarditis *(LSE)* has small- or medium-sized vegetations on either or both sides of the valve leaflets.

cords. Overlying these necrotic foci and along the lines of closure are small (1 to 2 mm) vegetations, called **verrucae.** Thus, RHD is one of the forms of vegetative valve disease, each of which exhibit their own characteristic morphologic features (Fig. 12.23). Subendocardial lesions, perhaps exacerbated by regurgitant jets, can induce irregular thickenings called **MacCallum plaques,** usually in the left atrium.

The cardinal anatomic changes of the mitral valve in chronic RHD are **leaflet thickening, commissural fusion and shortening, and thickening and fusion of the tendinous cords** (see Fig. 12.22D). The mitral valve is virtually always involved in chronic RHD; it is affected in isolation in roughly two-thirds of cases, and along with the aortic valve in another 25%. Tricuspid valve involvement is infrequent, and the pulmonary valve is only rarely affected.

In rheumatic mitral stenosis, calcification and fibrous bridging across the valvular commissures create "fish mouth" stenoses (see Fig. 12.22C). With tight mitral stenosis, the left atrium progressively dilates and may harbor mural thrombi that can embolize. Long-standing congestive changes in the lungs may induce pulmonary vascular and parenchymal changes; over time, these can lead to right ventricular hypertrophy. The left ventricle is largely unaffected by isolated pure mitral stenosis. Microscopically, valves show organization of the acute inflammation, with postinflammatory neovascularization and transmural fibrosis that obliterate the leaflet architecture. Aschoff bodies are rarely seen in surgical specimens or autopsy tissue from patients with chronic RHD, because of the long intervals between the initial insult and the development of the chronic deformity.

Clinical Features

RF is characterized by a constellation of major manifestations:

- *Migratory polyarthritis* of the large joints
- *Pancarditis* (myocarditis, pericarditis, or endocarditis)
- *Subcutaneous nodules* (typically on extensor surfaces of extremities)
- *Erythema marginatum,* an irregular circinate skin rash
- *Sydenham chorea,* a neurologic disorder with involuntary rapid movements

The diagnosis is established in accordance with the revised Jones criteria: evidence of a preceding group A streptococcal infection, and the presence of two major manifestations, or one major and two minor manifestations (nonspecific signs and symptoms that include fever, arthralgia, or elevated blood levels of acute-phase reactants); notably, these criteria are evolving and are applied differently in low- and high-risk settings.

Acute RF typically appears 10 days to 6 weeks after a group A streptococcal infection in about 3% of patients. It occurs most often in children between 5 and 15 years of age, but first attacks can occur in middle to later life. Although pharyngeal cultures for streptococci are negative by the time the illness begins, antibodies to one or more streptococcal enzymes, such as streptolysin O and DNase B, can be detected in the sera of most patients with RF. The predominant clinical manifestations are carditis and arthritis, the latter more common in adults than in children. Arthritis typically begins with migratory polyarthritis (accompanied by fever) in which one large joint after another becomes painful and swollen for a period of days and then subsides spontaneously, leaving no residual disability. Clinical features related to acute carditis include pericardial friction rubs, tachycardia, and arrhythmias. Myocarditis can cause cardiac dilation that may culminate in functional mitral valve insufficiency or even heart failure. Approximately 1% of affected individuals die of fulminant RF involvement of the heart.

After an initial attack, there is increased vulnerability to reactivation of the disease with subsequent pharyngeal infections, and the same manifestations are likely to appear with each recurrent attack. Damage to the valves is cumulative. Turbulence induced by ongoing valvular deformities leads to additional fibrosis. This is a prime example of a common theme in valvular heart disease in which the consequences of valve pathology may contribute to progression of that valve pathology in a positive feedback loop.

Clinical manifestations of RHD appear years or even decades after the initial episode of RF and depend on which valves are involved. In addition to cardiac murmurs, cardiac hypertrophy and dilation, and heart failure, individuals with chronic RHD may suffer from arrhythmias (particularly atrial fibrillation in the setting of mitral stenosis),

thromboembolic complications, and infective endocarditis (see later). The long-term prognosis is highly variable. Surgical repair or replacement of diseased valves has greatly improved the outlook for persons with RHD.

Infective Endocarditis (IE)

IE is a microbial infection of the heart valves or the mural endocardium that leads to the formation of vegetations composed of thrombotic debris and organisms, often associated with destruction of the underlying cardiac tissues. The aorta, aneurysms, other blood vessels, and prosthetic devices can also become infected. Although fungi and other classes of microorganisms can be responsible, most infections are bacterial (bacterial endocarditis). Prompt diagnosis, identification of the offending agent, and effective treatment of IE is important in limiting morbidity and mortality.

Traditionally, IE has been classified on clinical grounds into acute and subacute forms based on severity and tempo (reflecting microbial virulence), and whether there is underlying valvular pathology. Thus, acute IE is typically caused by infection of a previously normal heart valve by a highly virulent organism (e.g., *Staphylococcus aureus*) that rapidly produces destructive lesions. These infections may be difficult to cure with antibiotics alone and often require surgery; despite appropriate treatment, there can be substantial morbidity and even mortality. In contrast, subacute IE is characterized by organisms with lower virulence (e.g., viridans streptococci) that cause insidious infections of deformed valves with overall less destruction. In such cases the disease may pursue a protracted course of weeks to months, and cures can often be achieved with antibiotics alone. Of note, a clear delineation between acute and subacute endocarditis does not always exist, and many cases fall somewhere along the spectrum between the two forms.

Pathogenesis

Although highly virulent organisms can infect previously normal valves, a variety of cardiac abnormalities increase the risk of developing IE. RHD with valvular scarring has historically been the major antecedent disorder; as RHD becomes less common, it has been supplanted by mitral valve prolapse, degenerative calcific valvular stenosis, bicuspid aortic valve (whether calcified or not), artificial (prosthetic) valves, and congenital defects.

Endocarditis of native but previously damaged or otherwise abnormal valves is caused most commonly (50% to 60% of cases) by *Streptococcus viridans,* a normal component of the oral cavity flora. In contrast, more virulent *S. aureus* organisms commonly found on the skin can infect either healthy or deformed valves and are responsible for 20% to 30% of cases overall; notably, *S. aureus* is the major offender in IE among intravenous drug abusers. Other bacterial causes include enterococci and the so-called HACEK group (*Haemophilus, Actinobacillus, Cardiobacterium, Eikenella,* and *Kingella*), all commensals in the oral cavity. More rarely, Gram-negative bacilli and fungi can be involved. Prosthetic valve endocarditis occurring in the 1 to 2 months after surgical implantation is typically caused by skin flora (*S. aureus* and *S. epidermidis*); prosthetic valve infections 1 year or more after surgery tend to be streptococci and *S. aureus* (see later discussion of prosthetic valves). In about 10% of all cases of endocarditis, no organism can be isolated from the blood ("culture-negative" endocarditis); reasons include prior antibiotic therapy, difficulties in isolating the offending agent, or because deeply embedded organisms within the enlarging vegetation are not released into the blood.

Foremost among the factors predisposing to endocarditis are those that cause microorganism seeding into the bloodstream (bacteremia or fungemia). The source may be an obvious infection elsewhere, a dental or surgical procedure, a contaminated needle shared by intravenous drug users, or seemingly trivial breaks in the epithelial barriers of the gut, oral cavity, or skin. In patients with valve abnormalities, or with known bacteremia, IE risk can be lowered by antibiotic prophylaxis.

MORPHOLOGY

Vegetations on heart valves are the classic hallmark of IE; these are friable, bulky, potentially destructive lesions containing fibrin, inflammatory cells, and bacteria or other organisms (Figs. 12.23 and 12.24). The aortic and mitral valves are the most common

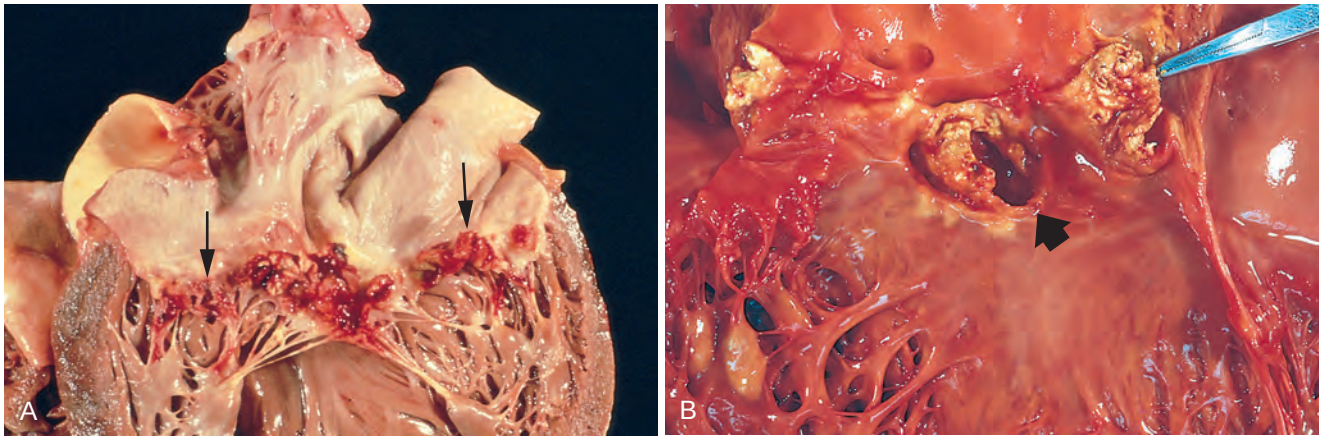


Figure 12.24 Infective (bacterial) endocarditis. (A) Endocarditis of mitral valve (subacute, caused by *Streptococcus viridans*). The large, friable vegetations are denoted by *arrows*. (B) Acute endocarditis of congenitally bicuspid aortic valve (caused by *Staphylococcus aureus*) with extensive cuspal destruction and ring abscess *(arrow)*.

sites of infection, although the valves of the right heart may also be involved, particularly in intravenous drug abusers. Vegetations can be single or multiple and may involve more than one valve; they can occasionally erode into the underlying myocardium and produce an abscess (ring abscess; see Fig. 12.24B). Vegetations are prone to embolization; because the embolic fragments often contain virulent organisms, abscesses frequently develop where they lodge, leading to sequelae such as septic infarcts or mycotic aneurysms.

The vegetations of subacute endocarditis are associated with less valvular destruction than those of acute endocarditis, although the distinction can be subtle. Microscopically, the vegetations of subacute IE typically exhibit granulation tissue at their bases indicative of healing. With time, fibrosis, calcification, and a chronic inflammatory infiltrate can develop.

Clinical Features

Acute endocarditis has a stormy onset with rapid onset of fever, chills, weakness, and lassitude. Although fever is the most consistent sign of IE, it can be slight or absent, particularly in older adults, and the only manifestations may be nonspecific fatigue, weight loss, and a flulike syndrome. Murmurs are present in the majority of patients with left-sided IE, either from a new valvular defect or from a preexisting abnormality. The modified Duke criteria (Table 12.9) facilitate diagnosis of individuals with suspected IE by taking into account predisposing factors, physical findings, blood culture results, echocardiographic findings, and laboratory information.

Complications of IE generally begin within the first few weeks of onset and can include glomerular antigen-antibody complex deposition causing glomerulonephritis (Chapter 20). Sepsis, arrhythmias (suggesting invasion into underlying myocardium and conduction system), and systemic embolization bode particularly ill for the patient. Left untreated, IE generally is fatal. However, with appropriate long-term (6 weeks or more) antibiotic therapy and/or valve replacement, mortality is reduced. For infections involving low-virulence organisms (e.g., *S. viridans*), the cure rate is 98%, and for enterococci and *S. aureus* infections, cure rates range from 60% to 90%; however, with infections due to gram-negative bacilli or fungi, one-half of the patients ultimately succumb. The cure rate for endocarditis arising on prosthetic valves is uniformly worse, and valve replacement is commonly required.

Earlier diagnosis and effective treatment has nearly eliminated some previously common clinical manifestations of long-standing IE—for example, microthromboemboli (manifest as splinter or subungual hemorrhages), erythematous or hemorrhagic nontender lesions on the palms or soles (Janeway lesions), subcutaneous nodules in the pulp of the digits (Osler nodes), and retinal hemorrhages in the eyes (Roth spots).

Noninfected Vegetations

Noninfected (sterile) vegetations occur in nonbacterial thrombotic endocarditis (NBTE) and the endocarditis of systemic lupus erythematosus (SLE).

Table 12.9 Diagnostic Criteria for Infective Endocarditis[a]

Pathologic Criteria
Microorganisms, demonstrated by culture or histologic examination, in a vegetation, embolus from a vegetation, or intracardiac abscess
Histologic confirmation of active endocarditis in a vegetation or intracardiac abscess
Clinical Criteria
Major
Blood culture(s) positive for a characteristic organism or persistently positive for an unusual organism
Echocardiographic identification of a valve-related or implant-related oscillating mass or abscess, or partial separation of artificial valve
New valvular regurgitation
Minor
Predisposing heart lesion or intravenous drug use
Fever
Vascular lesions, including major arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial hemorrhage, conjunctival hemorrhages, and Janeway lesions[b]
Immunological phenomena, including glomerulonephritis, Osler nodes,[c] Roth spots,[d] and rheumatoid factor
Microbiologic evidence, including a single culture positive for an unusual organism

[a]Diagnosis by these guidelines, often called the Modified Duke Criteria, requires either pathologic or clinical criteria; if clinical criteria are used, 2 major, 1 major + 3 minor, or 5 minor criteria are required for definitive diagnosis. "Possible" infective endocarditis diagnosis requires either 1 major +1 minor, or 3 minor.

[b]Janeway lesions are small erythematous or hemorrhagic, macular, nontender lesions on the palms and soles and are the consequence of septic embolic events.

[c]Osler nodes are small, tender subcutaneous nodules that develop in the pulp of the digits or occasionally more proximally in the fingers and persist for hours to several days.

[d]Roth spots are oval retinal hemorrhages with pale centers.

Modified from Li JS, Sexton DJ, Mick N, et al: Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis, *Clin Infect Dis* 30(4):633–638, 2000; Baddour LM: Cardiovascular infections. In Mann D, et al., editors: *Braunwald's Heart Disease. a Textbook of Cardiovascular Medicine*, ed 10, Philadelphia, 2015, WB Saunders, p 1524.

Nonbacterial Thrombotic Endocarditis (NBTE)

NBTE is characterized by the deposition of small (1 to 5 mm) sterile thrombi on the leaflets of the cardiac valves (Figs. 12.23 and 12.25). Histologically, these are bland thrombi, loosely attached to the underlying valve; the vegetations are nondestructive and do not elicit any inflammatory reaction. Although the local effect of the vegetations is usually trivial, they can be the source of systemic emboli that produce significant infarcts in the brain, heart, or elsewhere.

NBTE is often encountered in debilitated patients, such as those with cancer or sepsis—hence the previous term marantic endocarditis (root word *marasmus,* relating to malnutrition). Valvular damage is not a prerequisite for NBTE; indeed, the condition usually occurs on previously normal valves. Rather, hypercoagulable states are the usual precursor to NBTE; such conditions include chronic disseminated intravascular coagulation, hyperestrogenic states, and those associated with underlying malignancy, particularly mucinous adenocarcinomas. The latter association is likely related to the procoagulant effects of tumor-derived mucin or tissue factor that can also cause migratory thrombophlebitis (*Trousseau syndrome,* Chapter 4). Endocardial trauma, as from an indwelling catheter, is another well-recognized predisposing condition.

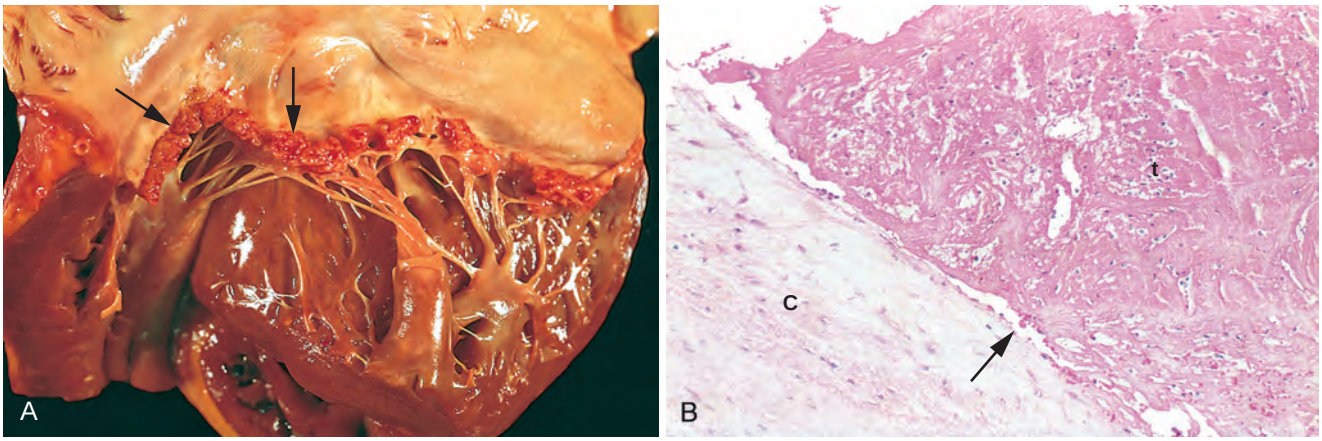


Figure 12.25 Nonbacterial thrombotic endocarditis (NBTE). (A) Nearly complete row of thrombotic vegetations along the line of closure of the mitral valve leaflets *(arrows)*. (B) Photomicrograph of NBTE, showing bland thrombus, with virtually no inflammation in the valve cusp *(c)* or the thrombotic deposit *(t)*. The thrombus is only loosely attached to the cusp *(arrow)*.

Endocarditis of Systemic Lupus Erythematosus (Libman-Sacks Disease)

Small (1 to 4 mm), sterile vegetations in the setting of systemic lupus erythematosus are termed Libman-Sacks endocarditis. The lesions develop as a consequence of immune complex deposition, with activation of complement and recruitment of Fc-receptor–bearing cells; histologically, there is an intense valvulitis and fibrinoid necrosis of the valve substance. The vegetations can occur anywhere on the valve surface, on the chordae, or even on the atrial or ventricular endocardium (see Fig. 12.23). Persistent injury can eventually result in valvular scarring and leaflet fusion, analogous to that seen in RHD. Similar lesions can occur in the antiphospholipid antibody syndrome (Chapter 4).

Carcinoid Heart Disease

The carcinoid syndrome refers to a systemic disorder marked by flushing, diarrhea, dermatitis, and bronchoconstriction that is caused by bioactive compounds such as serotonin released by carcinoid tumors (Chapter 17). Carcinoid heart disease refers to the cardiac manifestations caused by the bioactive compounds and occurs in roughly one-half of the patients in whom the systemic syndrome develops. Cardiac lesions do not typically occur until there is a massive hepatic metastatic burden, because the liver normally catabolizes circulating mediators before they can affect the heart. Classically, endocardium and valves of the right heart are primarily affected because they are the first cardiac tissues bathed by the mediators released by gastrointestinal carcinoid tumors. The left side of the heart is afforded some measure of protection because the pulmonary vascular bed degrades the mediators. However, left heart carcinoid lesions can occur in the setting of atrial or septal defects and right-to-left flow, or they can be induced by primary pulmonary carcinoid tumors.

Pathogenesis

The mediators elaborated by carcinoid tumors include serotonin (5-hydroxytryptamine), kallikrein, bradykinin, histamine, prostaglandins, and tachykinins. Although it is not clear which of these is causal, plasma levels of serotonin and urinary excretion of the serotonin metabolite 5-hydroxyindoleacetic acid correlate with the severity of the cardiac lesions. The valvular plaques in carcinoid syndrome are also similar to lesions that occured in patients taking fenfluramine (an appetite suppressant, no longer marketed) or ergot alkaloids (used previously for migraine headaches); interestingly, these agents affect systemic serotonin metabolism. Similarly, left-sided plaques have been reported after methysergide or ergotamine therapy for migraines; notably, these drugs are metabolized to serotonin as they pass through the pulmonary vasculature.

MORPHOLOGY

The cardiovascular lesions associated with the carcinoid syndrome are distinctive, glistening white intimal plaquelike thickenings of the endocardial surfaces of the cardiac chambers and valve leaflets (Fig. 12.26). The lesions are composed of smooth muscle cells and sparse collagen fibers embedded in an acid mucopolysaccharide-rich matrix material. Underlying structures are intact. With right-sided involvement, typical findings are tricuspid insufficiency and pulmonary stenosis.

Complications of Prosthetic Valves

Although prosthetic heart valves are less-than-perfect substitutes for the native tissues, their introduction has radically altered the prognosis for patients with valve disease. Two types of prosthetic valves are currently used, each with its own advantages and disadvantages:

- *Mechanical valves*. These consist of different configurations of rigid nonphysiologic material, such as caged balls, tilting disks, or hinged semicircular flaps (bileaflet tilting disk valves).
- *Tissue valves (bioprostheses)*. Porcine aortic valves or bovine pericardium are preserved in a dilute glutaraldehyde solution and then mounted on a prosthetic frame. Alternatively, frozen human valves from deceased donors

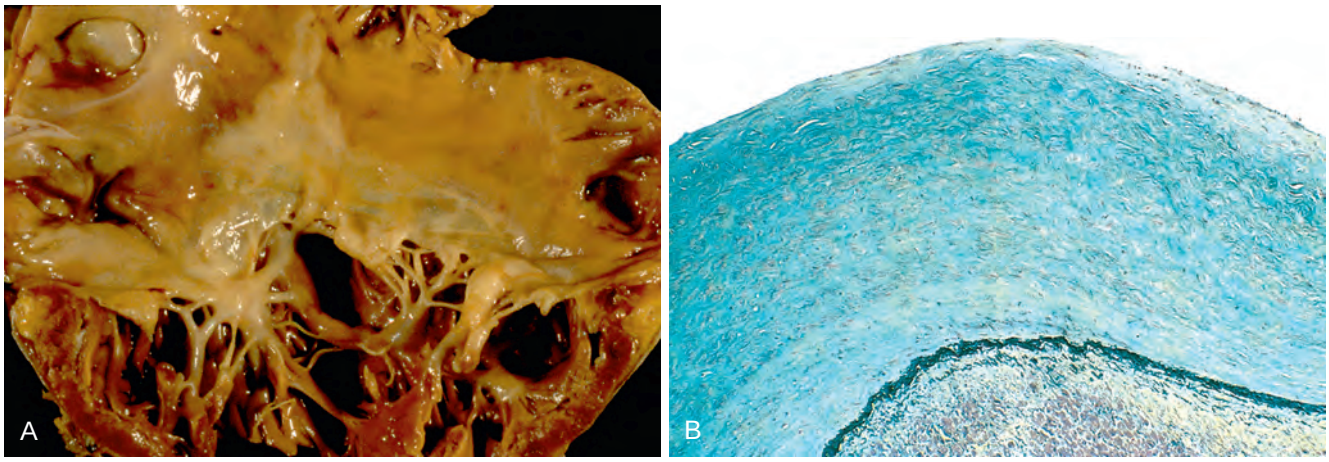


Figure 12.26 Carcinoid heart disease. (A) Characteristic endocardial fibrotic lesion involving the right ventricle and tricuspid valve. (B) Microscopic appearance of carcinoid heart disease with endocardial thickening. Movat stain shows myocardial elastic tissue *(black)* underlying the acid mucopolysaccharide-rich lesion *(blue-green)*. The underlying myocardium is unaffected.

(called cryopreserved "homografts") can also be used. Increasingly—especially in patients who are not good "open" surgical candidates—bioprosthetic valves are deployed by catheter-based approaches (transcatheter aortic valve replacement or TAVR) wedging the new valve into the outflow tract without resecting the original diseased valve. Tissue valves are flexible and function similarly to natural semilunar valves. However, the chemical treatment of the animal valves cross-links the valvular proteins, especially collagen, and renders the tissue nonviable. Similarly, the freezing and thawing of human homografts may also render them largely nonviable.

Approximately 60% of substitute valve recipients develop a serious prosthesis-related problem within 10 years after the surgery. The complications that occur depend on which type of valve has been implanted (Table 12.10 and Fig. 12.27).

- *Thromboembolism* is the major consideration with mechanical valves (see Fig. 12.27A); this may take the form of either thrombotic occlusion of the prosthesis or emboli released from thrombi formed on the valve. Because blood flow in all mechanical devices is nonlaminar, foci of turbulence and stasis are produced by prostheses that

Table 12.10 Complications of Cardiac Valve Prostheses

Thrombosis/thromboembolism
Anticoagulant-related hemorrhage
Prosthetic valve endocarditis
Structural deterioration (intrinsic)
Wear, fracture, poppet failure in ball valves, cuspal tear, calcification
Other forms of dysfunction
Inadequate healing (paravalvular leak), exuberant healing (obstruction), hemolysis

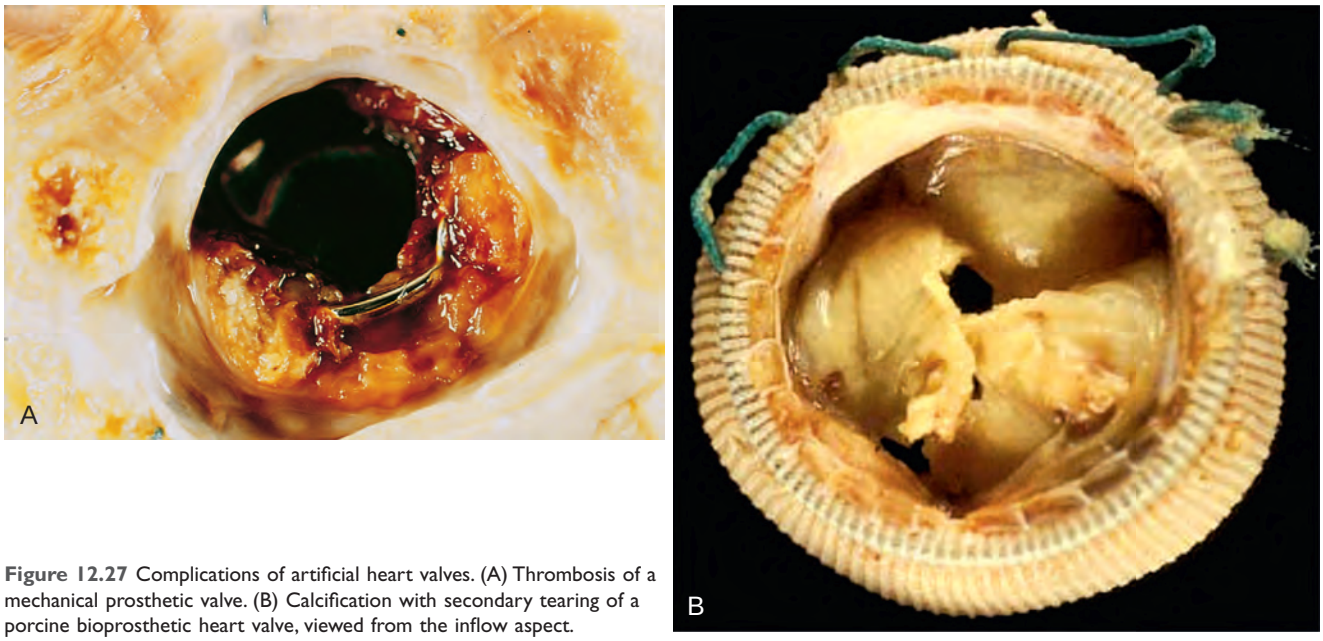


Figure 12.27 Complications of artificial heart valves. (A) Thrombosis of a mechanical prosthetic valve. (B) Calcification with secondary tearing of a porcine bioprosthetic heart valve, viewed from the inflow aspect.

predispose to thrombus formation. The risk of such complications necessitates long-term anticoagulation in all individuals with mechanical valves, with the attendant risk of hemorrhagic stroke or other forms of serious bleeding.

- *Structural deterioration* rarely causes failure of any of the mechanical valves in current use. However, virtually all bioprostheses eventually become incompetent due to calcification and/or tearing (see Fig. 12.27B).
- *Infective endocarditis* is a potentially serious complication of any valve replacement. The vegetations of prosthetic valve endocarditis are usually located at the prosthesis-tissue interface, and they often cause the formation of a ring abscess, which can eventually lead to a paravalvular regurgitant blood leak. In addition, vegetations may directly involve the tissue of bioprosthetic valvular cusps.
- *Other complications* include paravalvular leak due to inadequate healing, obstruction due to overgrowth of fibrous tissue during healing, valve-orifice disproportion—where the effective valve area is too small for the needs of the patient, leading to a relative stenosis—or intravascular hemolysis due to high shear forces.

KEY CONCEPTS

VALVULAR HEART DISEASE

- Valve pathology can lead to occlusion (stenosis) and/or regurgitation (insufficiency); acquired aortic and mitral valve stenoses account for approximately two-thirds of all valve disease; much of the remainder is mitral regurgitation.
- Valve calcification is a degenerative process that typically results in stenosis.
- Abnormal matrix synthesis and turnover result in myxomatous degeneration and insufficiency.
- Inflammatory valve diseases lead to postinflammatory scarring. RHD results from anti-streptococcal antibodies and T cells that cross-react with cardiac tissues; it most commonly affects the mitral valve and is responsible for 99% of acquired mitral stenoses.
- IE can be aggressive and rapidly destroy normal valves (acute IE), or can be indolent and minimally destructive of previously abnormal valves (subacute IE). Systemic embolization can produce septic infarcts.
- NBTE occurs on previously normal valves due to hypercoagulable states; embolization is an important complication.
- Mechanical prosthetic valves have thrombotic or hemorrhagic complications related to the nonlaminar flow of blood and the need for chronic anticoagulation. Bioprosthetic valves are nonviable and are therefore susceptible to long-term calcification and/or degeneration with tearing. Both types of valves have an increased risk of developing endocarditis relative to native valves.

CARDIOMYOPATHIES

The term cardiomyopathy (literally, heart muscle disease) has been historically applied to any cardiac dysfunction. Technically, that is not incorrect; the major heart disease categories already discussed—ischemic, valvular, hypertensive, or congenital—do cause "heart muscle disease." However, these major categories of cardiac diseases cause heart failure as predictable secondary consequences of other forms of cardiac dysfunctions—e.g., loss of pump function due to MI, or volume-pressure overloads due to an incompetent valve; in Table 12.11 these are referred to as secondary myocardial dysfunction. To distinguish these forms of "heart disease" from those associated with innate "primary" myocardial dysfunction, a more nuanced definition of cardiomyopathy has emerged. Thus, stimulated by the recognition of new phenotypes and the advent of more sophisticated molecular characterization—an expert panel has suggested: "Cardiomyopathies are a heterogeneous group of diseases of the myocardium associated with mechanical and/or electrical dysfunction that usually (but not invariably) exhibit inappropriate ventricular hypertrophy or dilatation and are due to a variety of causes that frequently are genetic. Cardiomyopathies either are confined to the heart or are part of generalized systemic disorders, often leading to cardiovascular death or progressive heart failure-related disability."

There are two major groups of cardiomyopathies:

- *Primary cardiomyopathies* involve predominantly the heart. They may be genetic or acquired (e.g., viral myocarditis, anthracycline cardiotoxic).
- *Secondary cardiomyopathies* have myocardial involvement as a component of a systemic or multiorgan disorder (e.g. hemochromatosis, amyloidosis).

Table 12.11 Cardiomyopathies: Functional Patterns and Causes

Functional Pattern	Left Ventricular Ejection Fraction[a]	Mechanisms of Heart Failure	Causes	Secondary Myocardial Dysfunction (Mimicking Cardiomyopathy)
Dilated	<40%	Impairment of contractility (systolic dysfunction)	Genetic; alcohol; peripartum; myocarditis; hemochromatosis; chronic anemia; anthracycline medications; sarcoidosis; idiopathic	Ischemic heart disease; valvular heart disease; hypertensive heart disease; congenital heart disease
Hypertrophic	50%–80%	Impairment of compliance (diastolic dysfunction)	Genetic; Friedreich ataxia; storage diseases; infants of diabetic mothers	Hypertensive heart disease; aortic stenosis
Restrictive	45%–90%	Impairment of compliance (diastolic dysfunction)	Amyloidosis; radiation-induced fibrosis; idiopathic	Pericardial constriction

[a]Range of normal values is approximately 50% to 65%.

A major advance in our understanding of the pathogenesis of cardiomyopathies stems from the accelerating identification of underlying genetic causes, including mutations in myocardial proteins involved in energy generation, contraction, cell-to-cell contacts, or connecting cytoskeleton to the extracellular matrix. These, in turn, lead to abnormal contraction or relaxation, or to dysregulated ion transport that can cause arrhythmias.

Cardiomyopathies can be classified according to a variety of criteria, including the underlying genetic basis of dysfunction; we have already discussed a number of the arrhythmia-inducing channelopathies. Here, we will confine our discussion to disorders that produce anatomic abnormalities falling into one of three distinct pathologic patterns (Fig. 12.28 and Table 12.11):

- *Dilated cardiomyopathy (DCM)* (including arrhythmogenic cardiomyopathy)
- *Hypertrophic cardiomyopathy*
- *Restrictive cardiomyopathy*

Among the three major patterns, DCM is most common (90% of cases), and restrictive cardiomyopathy is the least frequent. Within each pattern, there is a spectrum of clinical severity, and in some cases clinical features overlap among the groups. In addition, each of these patterns is associated with a specific identifiable cause or idiopathic; although many cases of DCM were previously labeled "idiopathic," the majority of these can now be assigned a genetic, toxic, or infectious etiology (Tables 12.11 and 12.12).

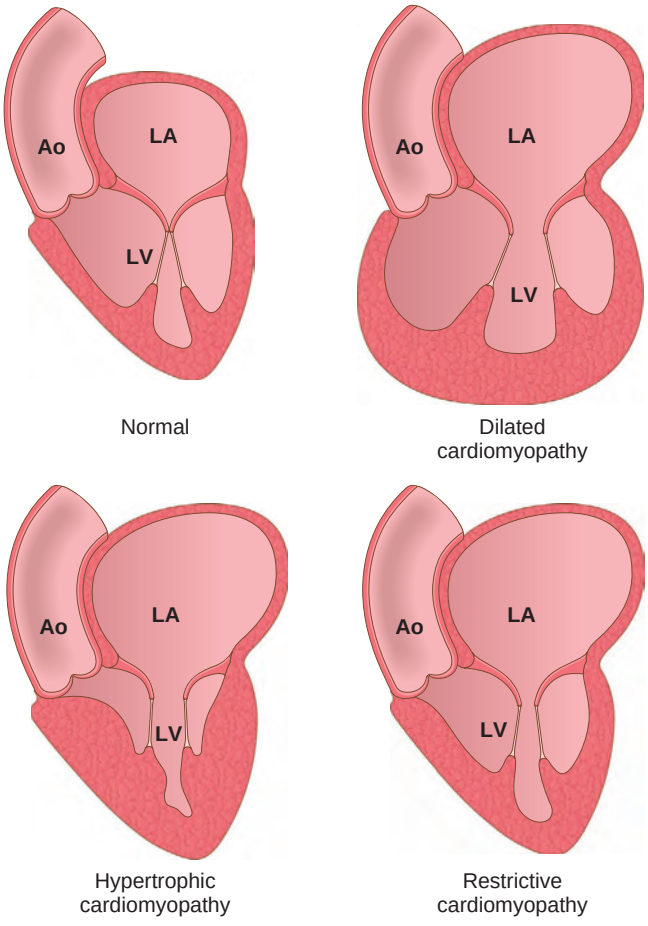


Figure 12.28 The three major morphologic patterns of cardiomyopathy. Dilated cardiomyopathy leads primarily to systolic dysfunction, whereas restrictive and hypertrophic cardiomyopathies result in diastolic dysfunction. Note the changes in atrial and/or ventricular wall thickness. *Ao*, Aorta; *LA*, left atrium; *LV*, left ventricle.

Table 12.12 Conditions Associated With Heart Muscle Diseases

Cardiac Infections
Viruses
Chlamydia
Rickettsia
Bacteria
Fungi
Protozoa
Toxins
Alcohol
Cobalt
Catecholamines
Carbon monoxide
Lithium
Hydrocarbons
Arsenic
Cyclophosphamide
Doxorubicin (Adriamycin) and daunorubicin
Metabolic
Hyperthyroidism
Hypothyroidism
Hyperkalemia
Hypokalemia
Nutritional deficiency (protein, thiamine, other avitaminoses)
Hemochromatosis
Neuromuscular Disease
Friedreich ataxia
Muscular dystrophy
Congenital atrophies
Storage Disorders and Other Depositions
Hunter-Hurler syndrome
Glycogen storage disease
Fabry disease
Amyloidosis
Infiltrative
Leukemia
Carcinomatosis
Sarcoidosis
Radiation-induced fibrosis
Immunologic
Myocarditis (several forms)
Posttransplant rejection

Dilated Cardiomyopathy (DCM)

DCM is characterized morphologically and functionally by progressive cardiac dilation and contractile (systolic) dysfunction, usually with concomitant hypertrophy. Many cases are familial, but the DCM phenotype can result from diverse causes, both primary and secondary.

Pathogenesis

Several different pathways can lead to DCM (Fig. 12.29). Identifying which may be causal in any given case is often

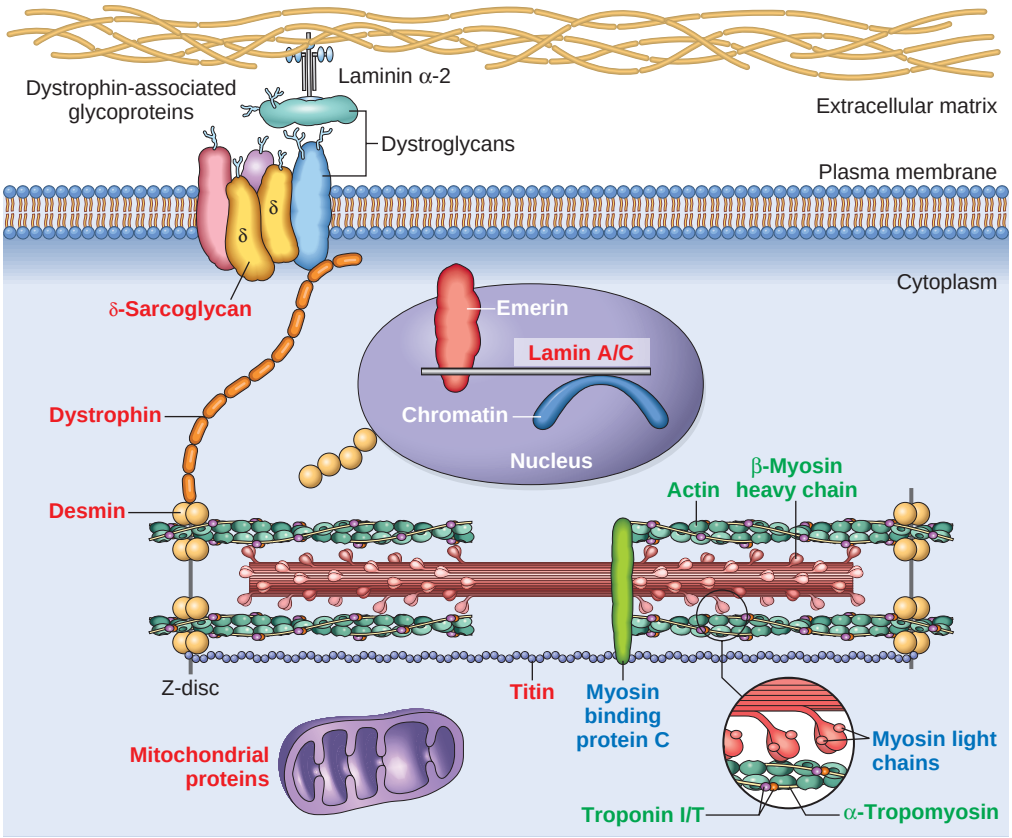


Figure 12.29 Schematic of a myocyte, showing key proteins mutated in dilated cardiomyopathy *(red labels)*, hypertrophic cardiomyopathy *(blue labels)*, or both *(green labels)*. Mutations in titin (the largest known human protein at approximately 30,000 amino acids) account for approximately 20% of all dilated cardiomyopathy. Titin spans the sarcomere and connects the Z and M bands, thereby limiting the passive range of motion of the sarcomere as it is stretched. Titin also functions like a molecular spring, with domains that unfold when the protein is stretched and refold when the tension is removed, thereby affecting the passive elasticity of striated muscle.

frustrating, because diagnosis is only made after the patient has developed end-stage heart failure; very simply, the heart is nonspecifically dilated and poorly contractile. The clinical evaluation at that point is done to exclude ischemic, valvular, hypertensive, or congenital causes; once they have been ruled out, a diagnosis of primary dilated cardiomyopathy can be suggested. Although an exhaustive evaluation may not be able to pinpoint a specific etiology ("idiopathic DCM"), familial (genetic) forms of DCM are increasingly recognized, and the final pathology can also result from a host of myocardial insults:

- *Genetic Influences.* DCM is probably familial in up to 50% of cases, caused by mutations in a diverse group of more than 20 genes encoding proteins involved in the cytoskeleton, sarcolemma, and nuclear envelope (e.g., lamin A/C). In particular, truncation mutations in *TTN,* a gene that encodes titin (so-called because it is the largest protein expressed in humans), may account for approximately 10% to 20% of all cases of DCM (Fig. 12.30). In the genetic forms of DCM, autosomal dominant inheritance is the predominant pattern, although there is often variable penetrance even within the same families, suggesting that the progression to DCM can be multifactorial.

  X-linked, autosomal recessive, and mitochondrial inheritance of DCM are less common. In some families, mitochondrial gene deletions affect oxidative phosphorylation; in others, there are mutations in genes encoding enzymes involved in fatty acid β-oxidation. Mitochondrial defects typically manifest in the pediatric population, whereas X-linked DCM typically presents after puberty and into early adulthood. X-linked cardiomyopathy can also be associated with mutations affecting the membrane-associated dystrophin protein that couples cytoskeleton to the extracellular matrix (recall that dystrophin is mutated in the most common skeletal myopathies, Duchenne and Becker muscular dystrophies; Chapter 27); some dystrophin gene mutations have DCM as the primary clinical feature. Interestingly, and probably resulting from the common developmental origin of contractile myocytes and conduction elements, congenital abnormalities of conduction may also be associated with DCM.
- *Myocarditis.* Sequential endomyocardial biopsies have documented progression from myocarditis to DCM. In other studies, the detection of the genetic fingerprints of coxsackie B and other viruses within myocardium of patients with DCM suggests that viral myocarditis can be causal (see later discussion).
- *Alcohol and other toxins.* Alcohol abuse is strongly associated with the development of DCM, raising the possibility that ethanol toxicity (Chapter 9) or a secondary nutritional disturbance can underlie myocardial injury. Alcohol or its metabolites (especially acetaldehyde) have a direct toxic effect on the myocardium. Moreover, chronic alcoholism may be associated with thiamine deficiency,

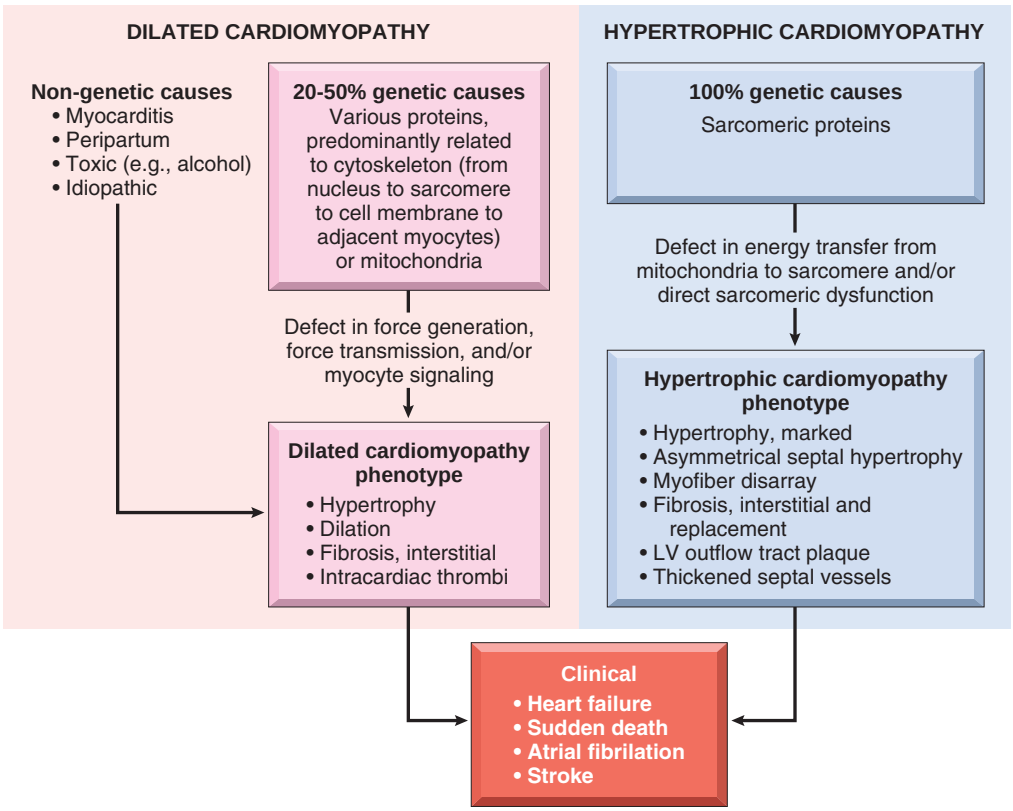


Figure 12.30 Causes and consequences of dilated and hypertrophic cardiomyopathy. Some dilated cardiomyopathies and virtually all hypertrophic cardiomyopathies are genetic in origin. The genetic causes of dilated cardiomyopathy involve mutations in any of a wide range of genes. They encode proteins predominantly of the cytoskeleton, but also the sarcomere, mitochondria, and nuclear envelope. In contrast, all of the mutated genes that cause hypertrophic cardiomyopathy encode proteins of the sarcomere. Although these two forms of cardiomyopathy differ greatly in subcellular basis and morphologic phenotypes, they share a common set of clinical complications. *LV,* left ventricle.

which can lead to beriberi heart disease (a form of DCM). Nevertheless, no morphologic features serve to distinguish alcoholic cardiomyopathy from DCM of other causes. Cobalt is an example of a heavy metal with cardiotoxicity and has caused DCM in the setting of inadvertent tainting (e.g., in beer production). Cardiotoxic drugs used for chemotherapy (discussed later) are also important causes of DCM.

- *Childbirth.* Peripartum cardiomyopathy can occur late in pregnancy or up to 5 months postpartum; the mechanism is probably multifactorial, including contributions from genetic susceptibility, pregnancy-associated hypertension, volume overload, nutritional deficiency, and/or other subtle metabolic derangements. In mouse models, DCM can also be induced by increased levels of circulating antiangiogenic mediators including vascular endothelial growth factor inhibitors (e.g., sFLT1, as occurs with preeclampsia) or antiangiogenic cleavage products of the hormone prolactin (which rises late in pregnancy). Thus, in pregnant women with a particular genetic predisposition, elevated antiangiogenic molecules can potentially lead to a microvascular angiogenic imbalance (microvascular loss in excess of angiogenesis), ultimately resulting in functional ischemic injury.
- *Iron overload* in the heart can result from either hereditary hemochromatosis (Chapter 18) or from multiple transfusions. DCM is the most common manifestation of such iron excess, and it may be caused by interference with metal-dependent enzyme systems or by injury from iron-mediated production of reactive oxygen species.
- *Supraphysiologic stress* can also result in DCM. This can happen with persistent tachycardia, hyperthyroidism, or even during development, as in the fetuses of insulin-dependent diabetic mothers. Excess catecholamines, in particular, cause multifocal myocardial contraction band necrosis that can eventually progress to DCM. This can happen in individuals with pheochromocytomas, tumors that elaborate epinephrine (Chapter 24); use of cocaine or vasopressor agents such as dopamine can have similar consequences. Such "catecholamine effect" also occurs in the setting of intense autonomic stimulation, for example, secondary to intracranial lesions or emotional duress. *Takotsubo cardiomyopathy* is an entity characterized by left ventricular contractile dysfunction after extreme psychological stress (thus also called broken heart syndrome); affected myocardium may be stunned or show contraction band necrosis. For unclear reasons, the left ventricular apex is most often affected, leading to "apical ballooning" that resembles a *takotsubo,* Japanese for "fishing pot for trapping octopus" (hence, the name). The mechanism of catecholamine cardiotoxicity is uncertain but likely relates either to direct myocyte toxicity due to calcium overload or to focal vasoconstriction in the coronary arterial macro- or microcirculation in the face of an increased heart rate.

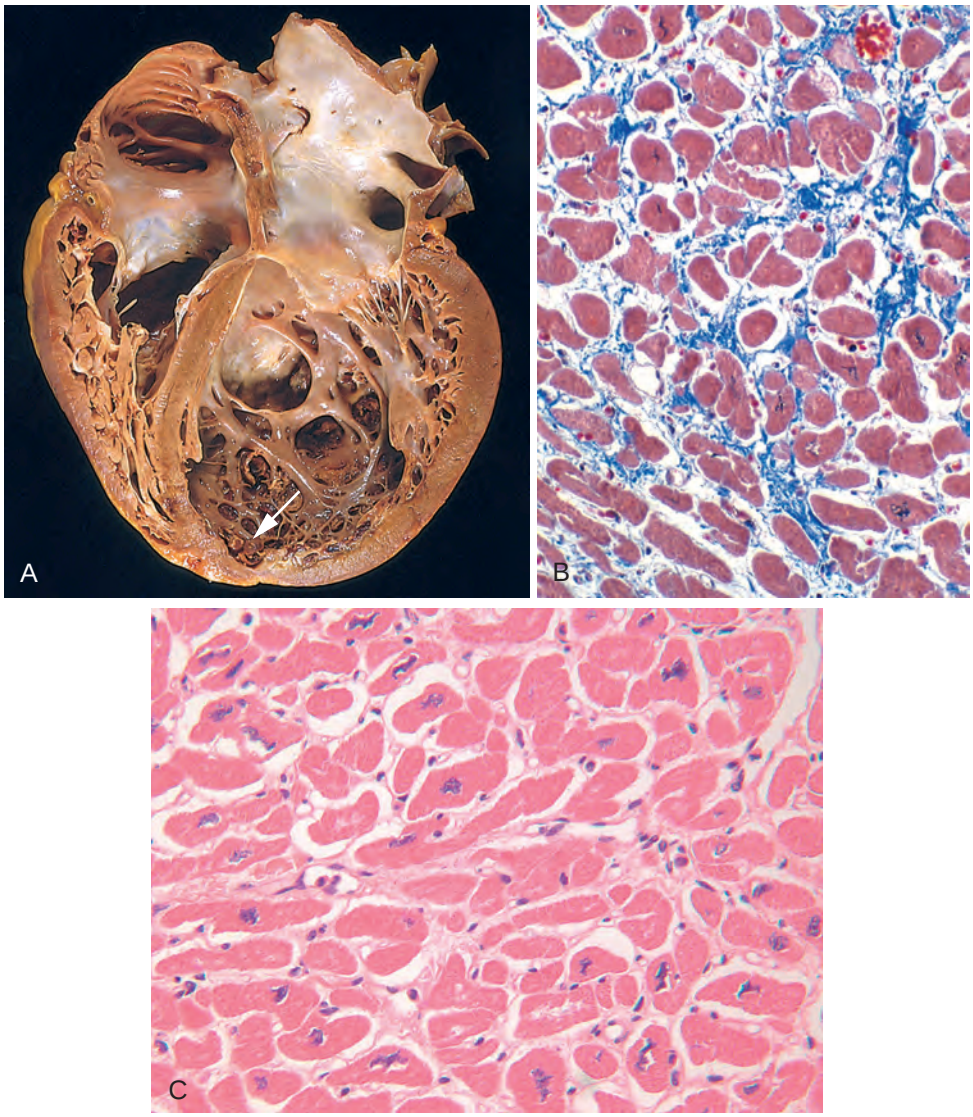


Figure 12.31 Dilated cardiomyopathy. (A) Four-chamber dilatation and hypertrophy are evident. There is a mural thrombus *(arrow)* at the apex of the left ventricle (seen on the right in this apical four-chamber view). The coronary arteries were patent. (B) Histologic section demonstrating variable myocyte hypertrophy and interstitial fibrosis (collagen is highlighted as blue in this Masson trichrome stain). (C) Histologic appearance of the enlarged, bizarre, hyperchromatic nuclei (likened to "Ninja stars") that are seen in greater numbers in patients with dilated cardiomyopathy caused by titin-truncating mutations.

MORPHOLOGY

In the DCM phenotype, the heart is typically enlarged, heavy (often weighing two to three times normal), and flabby, due to dilation of all chambers (Fig. 12.31). Mural thrombi can result from relative stasis of the blood in poorly contractile chambers and may be a source of thromboemboli. To be considered DCM, the heart should have no primary valvular alterations; if mitral (or tricuspid) regurgitation is present, it results from left (or right) ventricular chamber dilation (functional regurgitation). Either the coronary arteries are free from significant narrowing or the obstructions present are insufficient to explain the degree of cardiac dysfunction.

The histologic abnormalities in DCM are nonspecific and usually do not point to a specific etiology (see Fig. 12.31B). Moreover, the severity of morphologic changes may not reflect either the degree of dysfunction or the patient's prognosis.

Interstitial and endocardial fibrosis of variable degree is present, and small subendocardial scars may replace individual cells or groups of cells, probably reflecting healing of previous ischemic necrosis of myocytes caused by hypertrophy-induced imbalance between perfusion and demand. Most muscle cells are hypertrophied with enlarged nuclei, but some are attenuated, stretched, and irregular. In DCM caused by truncating mutations in the titin gene, myocytes may exhibit hyperchromatic, highly distorted "Ninja star"-like nuclei (see Fig. 12.31C); although these can be a nonspecific finding, identifying them in greater than or equal to 5% of myocytes is highly suggestive of a titin-truncation mutation.

Clinical Features

The fundamental defect in DCM is ineffective contraction. Thus, in end-stage DCM, the cardiac ejection fraction typically is less than 25% (normal is 50% to 65%). DCM can

occur at any age, including in childhood, but it most commonly affects individuals between 20 and 50 years of age. It presents with slowly progressive signs and symptoms of CHF including dyspnea, easy fatigability, and poor exertional capacity. Secondary mitral regurgitation and abnormal cardiac rhythms are common, and embolism from intracardiac thrombi can occur. Death usually results from progressive cardiac failure or arrhythmia, and it can occur suddenly. Although the annual mortality is high (10% to 50%), some severely affected patients respond well to pharmacologic or electrical resynchronization (biventricular pacing) therapies. Cardiac transplantation is also increasingly performed, and long-term ventricular assist can be beneficial. Interestingly, in some patients, relatively short-term mechanical cardiac support can induce durable improvement of cardiac function.

Arrhythmogenic Cardiomyopathy

Arrhythmogenic right ventricular cardiomyopathy is an autosomal dominant disorder that classically manifests with right-sided heart failure and rhythm disturbances, which can cause sudden cardiac death. Left-sided involvement with left-sided heart failure can also occur. Classically, the right ventricular wall is severely attenuated due to loss of myocytes, accompanied by massive fatty infiltration and focal fibrosis (Fig. 12.32). Although mononuclear inflammation may be present around degenerating cardiomyocytes, arrhythmogenic cardiomyopathy is not considered an inflammatory cardiomyopathy. Classical arrhythmogenic cardiomyopathy has an autosomal dominant inheritance with a variable penetrance; many of the causal mutations involve genes encoding desmosomal junctional proteins at the intercalated disk (e.g., plakoglobin) as well as proteins that interact with the desmosome (e.g., the intermediate filament desmin). Naxos syndrome is a disorder characterized by arrhythmogenic cardiomyopathy and hyperkeratosis of plantar and palmar skin surfaces, and is also associated with plakoglobin mutations.

Hypertrophic Cardiomyopathy

Hypertrophic cardiomyopathy (HCM) is a common (1 in 500 prevalence), clinically heterogeneous, genetic disorder characterized by myocardial hypertrophy, poorly compliant left ventricular myocardium leading to abnormal diastolic filling, and (in about one-third of cases) intermittent ventricular outflow obstruction. It is the leading cause of left ventricular hypertrophy unexplained by other clinical or pathologic causes. The heart is thick-walled, heavy, and hypercontracting, in striking contrast to the flabby, hypocontracting heart of DCM. HCM causes primarily diastolic dysfunction; systolic function is usually preserved. The two most common diseases that must be distinguished clinically from HCM are deposition diseases (e.g., amyloidosis, Fabry disease) and hypertensive heart disease coupled with age-related subaortic septal hypertrophy (see earlier discussion under Hypertensive Heart Disease in this chapter). Occasionally, valvular or congenital subvalvular aortic stenosis can also mimic HCM.

Pathogenesis

In most cases, the pattern of transmission is autosomal dominant with variable penetrance. **HCM is most commonly caused by mutations in any one of several genes that encode sarcomeric proteins;** there are more than 400 different known mutations in nine different genes, most being missense mutations. Mutations causing HCM are found most commonly in the genes encoding myosin-binding protein C *(MYBP-C)* or β-myosin heavy chain *(β-MHC/MYH7),* followed by the genes coding for cardiac TnI, TnT, and α-tropomyosin; overall, these account for 70% to 80% of all cases. Different affected families can have distinct mutations involving the same protein; thus, over 50 different *β-MHC* mutations can cause HCM. The prognosis of HCM varies widely and correlates strongly with specific mutations.

As mentioned earlier, HCM is a disease caused by mutations in proteins of the sarcomere. Although the precise mechanism by which these mutations cause HCM is unclear,

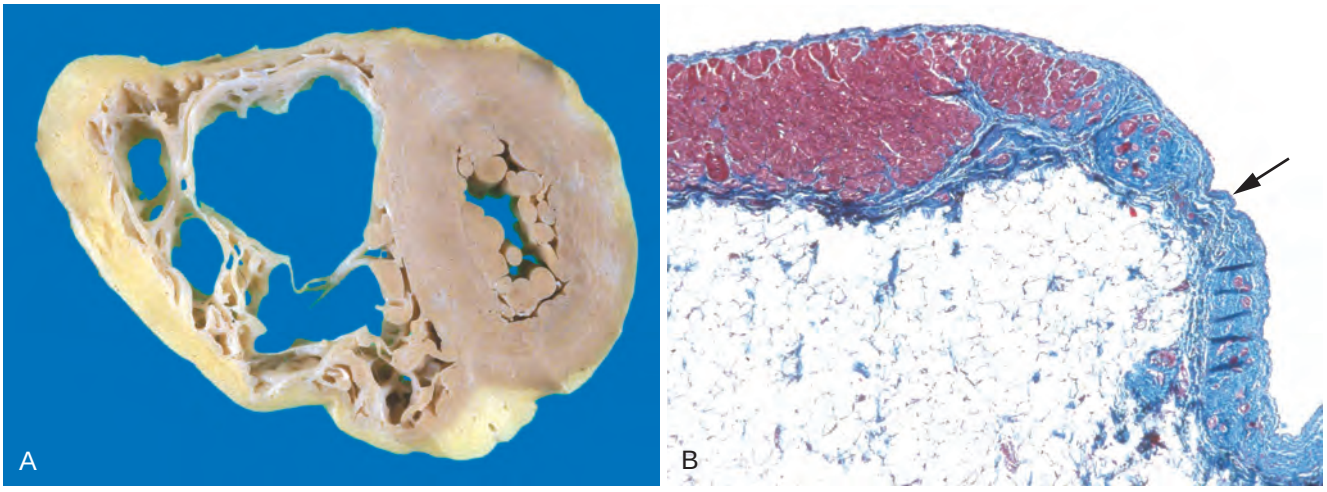


Figure 12.32 Arrhythmogenic cardiomyopathy. (A) Gross photograph, showing dilation of the right ventricle and near-transmural replacement of the right ventricular free-wall by fat and fibrosis. The left ventricle has a virtually normal configuration in this case, but can also be involved by the disease process. (B) Histologic section of the right ventricular free wall, demonstrating replacement of myocardium *(red)* by fibrosis *(blue, arrow)* and fat (Masson trichrome stain).

current evidence suggests that it arises from defective energy transfer from its source of generation (mitochondria) to its site of use (sarcomeres). In contrast, DCM is mostly associated with abnormalities of cytoskeletal proteins (see Fig. 12.30) and can be conceptualized as a disease of abnormal force generation, force transmission, or myocyte signaling. To complicate matters, mutations in certain genes (highlighted in Fig. 12.30) can give rise to either HCM or DCM, depending on the site and nature of the mutation. Genetic testing in cardiomyopathies is still hindered by the large number of genetic "variants of unknown significance," so genetic screening of family members of affected probands must be interpreted within the clinical context.

MORPHOLOGY

The essential feature of HCM is **massive myocardial hypertrophy, usually without ventricular dilation** (Fig. 12.33). The classic pattern involves disproportionate thickening of the ventricular septum relative to the left ventricle free wall, termed **asymmetric septal hypertrophy.** In about 10% of cases, the hypertrophy is concentric and symmetrical. On longitudinal sectioning, the normally round-to-ovoid left ventricular cavity may be compressed into a "banana-like" configuration by bulging of the ventricular septum into the lumen (see Fig. 12.33A). Although marked hypertrophy can involve the entire septum, it is usually most prominent in the subaortic region; the left ventricular outflow tract often exhibits a fibrous endocardial plaque and thickening of the anterior mitral leaflet. Both findings result from contact of the anterior mitral leaflet with the septum during ventricular systole; they correlate with the echocardiographic "systolic anterior motion" of the anterior leaflet, with functional left ventricular outflow tract obstruction during systole.

The most important histologic features of HCM myocardium are (1) massive myocyte hypertrophy, with transverse myocyte diameters frequently greater than 40 μm (normal is approximately 15 μm); (2) haphazard disarray of bundles of myocytes, individual myocytes, and contractile elements in sarcomeres within cells (termed ***myofiber disarray***); (3) fibrotic narrowing of small intramural arteries; and (4) interstitial and replacement fibrosis (see Fig. 12.33B).

Clinical Features

HCM is characterized by reduced stroke volume due to impaired diastolic filling—a consequence of a reduced chamber size, as well as the reduced compliance of the massively hypertrophied left ventricle. In addition, approximately 25% of patients with HCM have dynamic obstruction to the left ventricular outflow as the anterior mitral leaflet moves toward the ventricular septum during systole. The compromised cardiac output in conjunction with a secondary increase in pulmonary venous pressure explains the exertional dyspnea seen in these patients. Auscultation discloses a harsh systolic ejection murmur caused by the ventricular outflow obstruction. Because of the massive hypertrophy, high left ventricular chamber pressure, and frequently thick-walled intramural arteries, focal myocardial ischemia commonly results, even in the absence of concomitant epicardial CAD. Major clinical problems in HCM are atrial fibrillation, mural thrombus formation leading to embolization and possible stroke, intractable cardiac failure, ventricular arrhythmias, and, not infrequently, sudden death, especially with certain mutations. Indeed, HCM is one of the most common causes of sudden, otherwise unexplained death in young athletes.

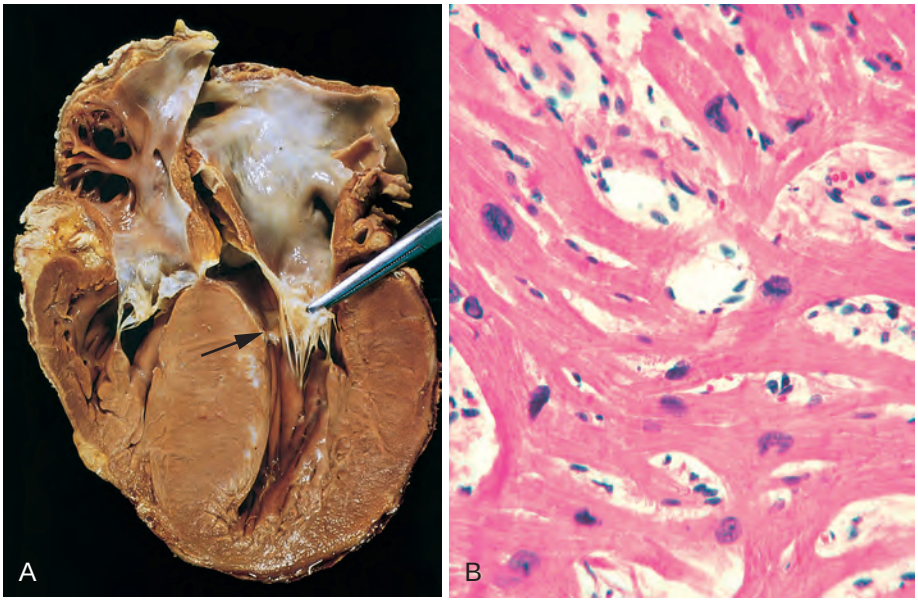


Figure 12.33 Hypertrophic cardiomyopathy with asymmetric septal hypertrophy. (A) The septal muscle bulges into the left ventricular outflow tract, and the left atrium is enlarged. The anterior mitral leaflet has been reflected away from the septum to reveal a fibrous endocardial plaque *(arrow)* (see text). (B) Histologic appearance demonstrating myocyte disarray, extreme hypertrophy, and exaggerated myocyte branching, as well as the characteristic interstitial fibrosis.

The natural history of HCM is highly variable. Most patients can be helped by pharmacologic intervention (e.g., β-adrenergic blockade) to decrease heart rate and contractility. Implantable cardioverter defibrillators are warranted for HCM patients at risk for ventricular arrhythmias. Some benefit can also be gained by reducing the septal myocardial mass, thus relieving the outflow tract obstruction. This can be achieved either by surgical excision of muscle or by carefully controlled septal infarction through a catheter-based infusion of alcohol.

Restrictive Cardiomyopathy

Restrictive cardiomyopathy is characterized by a primary decrease in ventricular compliance, resulting in impaired ventricular diastolic filling. Because the contractile (systolic) function of the left ventricle is usually unaffected, the functional abnormality can be confused with that of constrictive pericarditis or HCM. Restrictive cardiomyopathy can be idiopathic or associated with distinct disorders that affect the myocardium, principally amyloidosis (described later), sarcoidosis, radiation-induced fibrosis, metastatic tumors, or the accumulation of metabolites from inborn errors of metabolism.

The gross morphologic features are somewhat nonspecific; although bi-atrial dilation is commonly observed due to restricted ventricular filling and pressure overloads, the ventricles are of approximately normal size (or slightly enlarged), the cavities are not dilated, and the myocardium is largely unremarkable. Microscopically, there can be patchy or diffuse interstitial fibrosis, varying from minimal to extensive. Cardiac imaging is increasingly used to characterize myocardial "infiltrates," and endomyocardial biopsy can occasionally suggest a specific etiology.

Three other restrictive conditions merit brief mention.

- *Endomyocardial fibrosis* is principally a disease of children and young adults in Africa and other tropical areas, characterized by fibrosis of the ventricular endocardium and subendocardium that extends from the apex upward, often eventually involving the tricuspid and mitral valves. The fibrous tissue markedly diminishes the volume and compliance of affected chambers and so causes a restrictive functional defect. Ventricular mural thrombi sometimes develop, and indeed much of the endocardial fibrosis may result from thrombus organization. Endomyocardial fibrosis is linked to nutritional deficiencies and/or inflammation related to parasitic infections (e.g., hypereosinophilia); worldwide, it is the most common form of restrictive cardiomyopathy.
- *Loeffler endomyocarditis* also results in endomyocardial fibrosis, typically with large mural thrombi, and an overall morphology similar to the tropical disease. However, in addition to the cardiac changes, there is often a peripheral eosinophilia and eosinophilic infiltrates in multiple organs, including the heart. The release of toxic products of eosinophils, especially major basic protein, probably causes endomyocardial necrosis, followed by scarring of the necrotic area, layering of the endocardium by thrombus, and finally organization of the thrombus. Many patients with Loeffler endomyocarditis have a myeloproliferative disorder associated with chromosomal rearrangements involving either the platelet-derived growth factor receptor *(PDGFR)-α* or *-β* genes (Chapter 13). These rearrangements produce fusion genes that encode constitutively active PDGFR tyrosine kinases. Treatment of such patients with tyrosine kinase inhibitors has resulted in hematologic remissions and resolution of the endomyocardial lesions.
- *Endocardial fibroelastosis* is an uncommon heart disease characterized by fibroelastic thickening that typically involves the left ventricular endocardium. It is most common in the first 2 years of life; in one-third of cases, it is accompanied by aortic valve obstruction or other congenital cardiac anomalies. Endocardial fibroelastosis may actually represent a common morphologic end-point of several different insults including viral infections (e.g., intrauterine exposure to mumps) or mutations in the gene for tafazzin, which affects mitochondrial inner membrane integrity. Diffuse involvement may be responsible for rapid and progressive cardiac decompensation and death.

Amyloidosis

Amyloidosis is an important form of restrictive cardiomyopathy resulting from the extracellular accumulation of protein fibrils that form insoluble β-pleated sheets (Chapter 6). Cardiac amyloidosis can appear as a consequence of systemic amyloidosis (e.g., due to myeloma or inflammation-associated amyloid) or can be restricted to the heart, particularly in the aged (senile cardiac amyloidosis). Cardiac amyloidosis characteristically occurs in individuals 70 years of age and older, and it has a far better prognosis than systemic amyloidosis; the amyloid deposits are largely composed of transthyretin, a normal serum protein synthesized in the liver that transports thyroxine and retinol-binding protein. Mutant forms of transthyretin can accelerate the cardiac amyloid deposition; 4% of African Americans have a transthyretin mutation substituting isoleucine for valine at position 122 that produces a particularly amyloidogenic protein that is responsible for autosomal dominant familial transthyretin amyloidosis.

Cardiac amyloidosis most frequently produces a restrictive cardiomyopathy, but it can also be asymptomatic, manifest as dilation or arrhythmias, or mimic ischemic or valvular disease. The varied presentations depend on the predominant location of the deposits, for example, interstitium, conduction system, coronary vasculature, or valves.

MORPHOLOGY

In cardiac amyloidosis, the heart varies in consistency from normal to firm and rubbery. The chambers are usually of normal size but can be dilated and have thickened walls. Small, semitranslucent nodules resembling drips of wax may be seen on the atrial endocardial surface, particularly on the left. Histologically, hyaline eosinophilic deposits of amyloid may be found in the interstitium, conduction tissue, valves, endocardium, pericardium, and small intramural coronary arteries (Fig. 12.34). Amyloid accumulations can be distinguished from other deposits by special stains such as Congo red or sulfated Alcian blue; the former produces classic apple-green birefringence when viewed under polarized light (see Fig. 12.34B). Intramural arteries and arterioles may have sufficient amyloid in their walls to compress and occlude their lumens, inducing myocardial ischemia ("small-vessel disease").

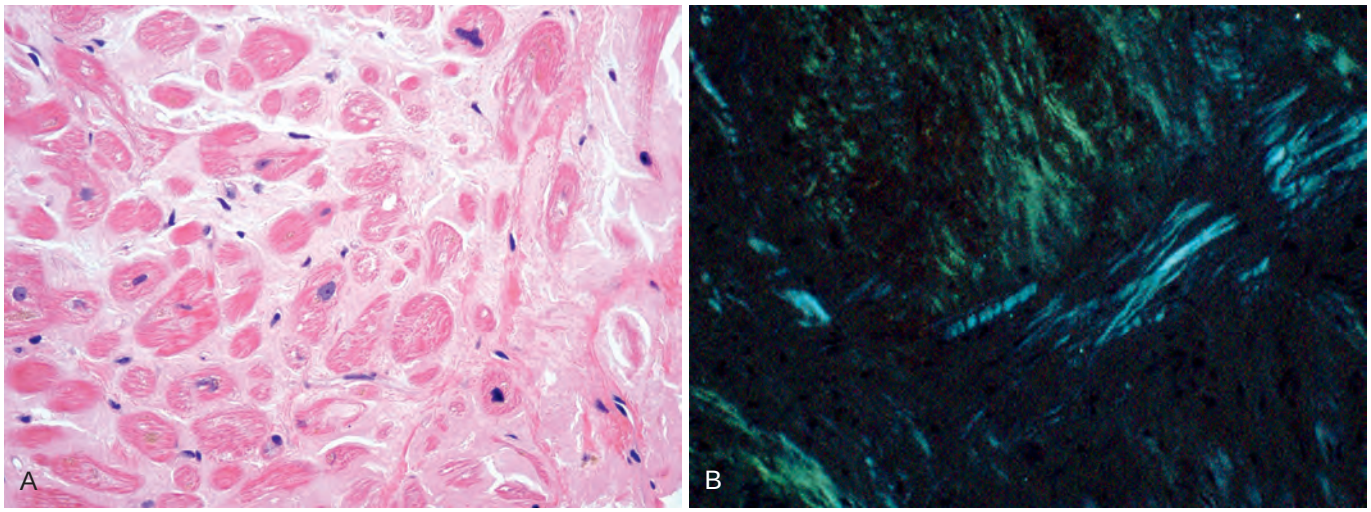


Figure 12.34 Cardiac amyloidosis. (A) Hematoxylin and eosin stain, showing amyloid appearing as amorphous pink material around myocytes. (B) Congo red stain viewed under polarized light, in which amyloid shows characteristic apple-green birefringence (compared with collagen, pale blue).

Myocarditis

Myocarditis is a diverse group of pathologic entities in which infectious microorganisms and/or a primary inflammatory process cause myocardial injury. Myocarditis should be distinguished from conditions such as IHD, in which myocardial inflammation is secondary to other causes.

Pathogenesis

In the United States, viral infections are the most common cause of myocarditis. Coxsackie viruses A and B and other enteroviruses probably account for most of the cases. Other less common etiologic agents include cytomegalovirus, HIV, and influenza (Table 12.13). In some (but not all) cases, the offending agent can be implicated by serologic studies or by identifying viral nucleic acid sequences in myocardial biopsies. Depending on the pathogen and the host, viruses can potentially cause myocardial injury either as a direct cytopathic effect, or by eliciting a destructive immune response. Inflammatory cytokines produced in response to myocardial injury can also cause myocardial dysfunction that is out of proportion to the degree of actual myocyte damage.

Nonviral agents are also important causes of infectious myocarditis, particularly the protozoan *Trypanosoma cruzi,* the agent of Chagas disease. Chagas disease is endemic in some regions of South America, with myocardial involvement in most infected individuals. About 10% of patients die during an acute attack; others develop a chronic immune-mediated myocarditis that may progress to cardiac insufficiency in 10 to 20 years. Trichinosis *(Trichinella spiralis)* is the most common helminthic disease associated with myocarditis. Parasitic diseases, including toxoplasmosis, and bacterial infections such as Lyme disease and diphtheria, can also cause myocarditis. In the case of diphtheritic myocarditis, the myocardial injury is a consequence of diphtheria toxin release by *Corynebacterium diphtheriae* (Chapter 8). Myocarditis occurs in approximately 5% of patients with Lyme disease, a systemic illness caused by the bacterial spirochete *Borrelia burgdorferi* (Chapter 8); it manifests primarily as a self-limited conduction system disorder that may require a temporary pacemaker. AIDS-associated myocarditis may reflect inflammation and myocyte damage without a clear etiologic agent, or a myocarditis attributable directly to HIV or to an opportunistic pathogen.

There are also noninfectious causes of myocarditis. Broadly speaking they are either immunologically mediated (hypersensitivity myocarditis) or idiopathic conditions with distinctive morphology (giant cell myocarditis) suspected to be of immunologic origin (see Table 12.13). More recently, immune checkpoint inhibitiors administered for treatment of cancer have occasionally led to an (often fatal) lymphocytic myocarditis.

Table 12.13 Major Causes of Myocarditis

Infections
Viruses (e.g., coxsackievirus, echovirus, influenza, HIV, cytomegalovirus)
Chlamydiae (e.g., *Chlamydophila psittaci*)
Rickettsiae (e.g., *Rickettsia typhi*, typhus fever)
Bacteria (e.g., *Corynebacterium diphtheriae, Neisseria meningococcus, Borrelia* (Lyme disease)
Fungi (e.g., *Candida*)
Protozoa (e.g., *Trypanosoma cruzi* [Chagas disease], toxoplasmosis)
Helminths (e.g., trichinosis)
Immune-Mediated Reactions
Postviral
Poststreptococcal (rheumatic fever)
Systemic lupus erythematosus
Drug hypersensitivity (e.g., methyldopa, sulfonamides)
Transplant rejection
Immune checkpoint inhibitor therapies
Unknown
Sarcoidosis
Giant cell myocarditis

HIV, Human immunodeficiency virus.

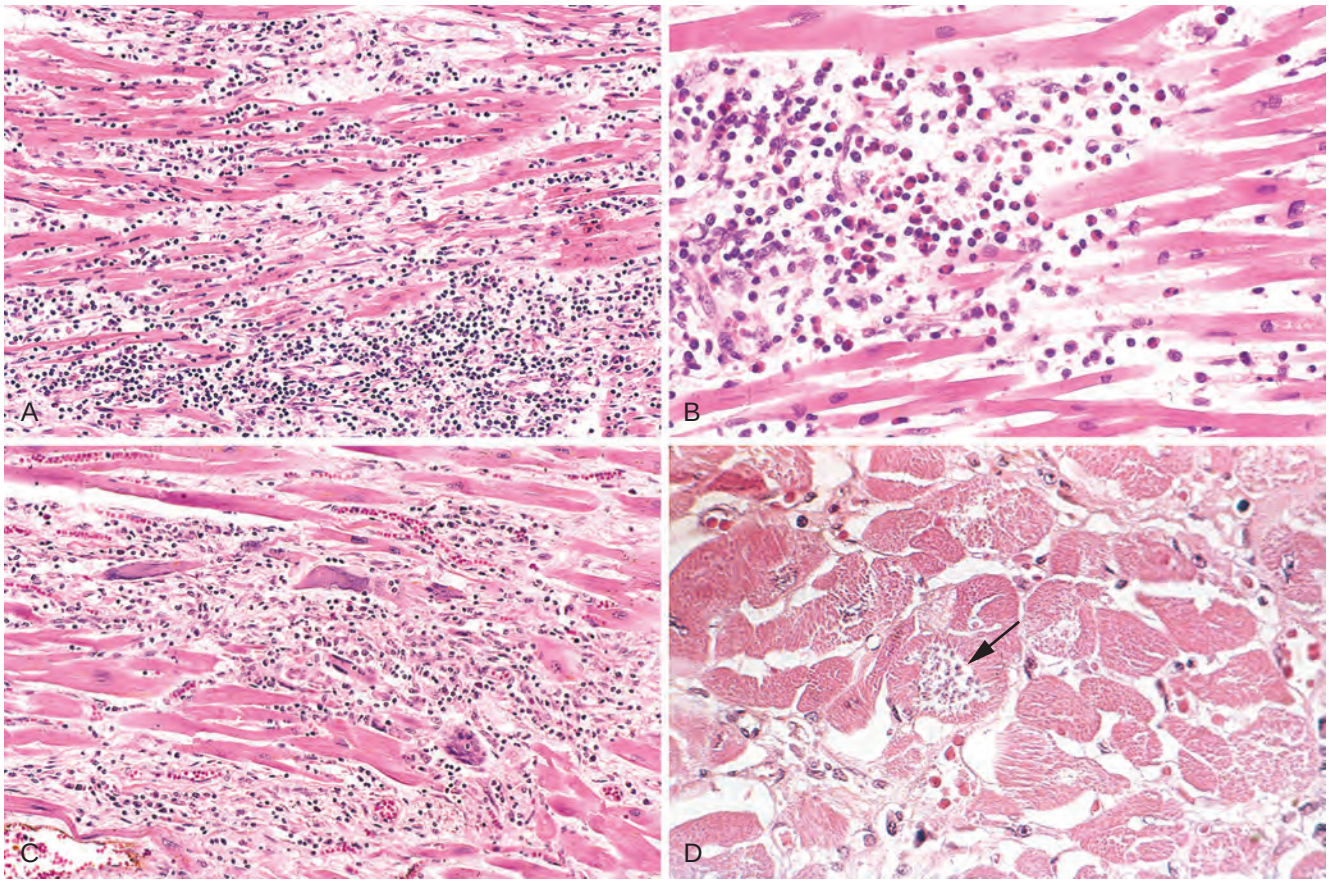


Figure 12.35 Myocarditis. (A) Lymphocytic myocarditis, associated with myocyte injury. (B) Hypersensitivity myocarditis, characterized by interstitial inflammatory infiltrate composed largely of eosinophils and mononuclear inflammatory cells, predominantly localized to perivascular and expanded interstitial spaces. (C) Giant-cell myocarditis, with mononuclear inflammatory infiltrate containing lymphocytes and macrophages, extensive loss of muscle, and multinucleated giant cells (fused macrophages). (D) The myocarditis of Chagas disease. A myofiber distended with trypanosomes *(arrow)* is present along with individual myofiber necrosis, and modest amounts of inflammation.

MORPHOLOGY

Grossly, the heart in myocarditis may appear normal or dilated; some hypertrophy may be present depending on disease duration. In advanced stages, the ventricular myocardium is flabby and often mottled by either pale foci or minute hemorrhagic lesions. Mural thrombi may be present.

Active myocarditis is characterized by an interstitial inflammatory infiltrate associated with degenerating or apoptotic myocytes (Fig. 12.35); larger zones of myocyte drop-out or necrosis with interstitial edema may reflect vascular injury with secondary ischemic damage. A diffuse, mononuclear, predominantly lymphocytic infiltrate is most common (see Fig. 12.35A). Although endomyocardial biopsies can be diagnostic, they can be spuriously negative because inflammation of the myocardium is patchy. If the patient survives the acute phase of myocarditis, the inflammatory lesions either resolve, leaving no residual changes, or heal by progressive fibrosis.

Hypersensitivity myocarditis is characterized by perivascular infiltrates composed of lymphocytes, macrophages, and a high proportion of eosinophils (see Fig. 12.35B). A morphologically distinctive form of myocarditis, called **giant-cell myocarditis,** is characterized by a widespread inflammatory cellular infiltrate containing multinucleate giant cells (fused macrophages) interspersed with lymphocytes, eosinophils, plasma cells, and macrophages. Focal to frequently extensive myocyte damage is present (see Fig. 12.35C). This variant likely represents the fulminant end of the myocarditis spectrum and carries a poor prognosis.

The myocarditis of **Chagas disease** is distinctive by virtue of the parasitization of scattered myofibers by trypanosomes accompanied by a mixed inflammatory infiltrate of neutrophils, lymphocytes, macrophages, and occasional eosinophils (see Fig. 12.35D).

Clinical Features

The clinical spectrum of myocarditis is broad. At one end, the disease is entirely asymptomatic, and patients can expect a complete recovery without sequelae; at the other extreme is the precipitous onset of heart failure or arrhythmias, occasionally followed by sudden death. Between these extremes are various levels of symptomatology, including fatigue, dyspnea, palpitations, precordial discomfort, and fever. The clinical features of myocarditis can mimic those of acute MI. As noted previously, patients can develop DCM as a late complication of myocarditis.

Other Causes of Myocardial Disease

Cardiotoxic Drugs

Cardiac complications of cytotoxic cancer therapies are an important clinical problem. Cardiotoxicity may be associated with many different conventional chemotherapeutic agents, as well as tyrosine kinase inhibitors. The anthracyclines doxorubicin and daunorubicin are the chemotherapeutic agents most often associated with toxic myocardial injury; they cause DCM with heart failure by direct damage to cardiomyocytes. Anthracycline toxicity is dose-dependent, with the cardiotoxicity risk increasing when cumulative lifetime doses exceed 250 mg/m^2.

Many other therapeutic agents, including lithium, phenothiazines, and chloroquine can idiosyncratically induce myocardial injury and sometimes sudden death. Common findings in affected myocardium include myofiber swelling, cytoplasmic vacuolization, and fatty change. Discontinuing the offending agent often leads to prompt resolution, without apparent sequelae. Occasionally, however, more extensive damage produces myocyte death that can evolve to a DCM.

Radiation

Radiation therapy to the thorax can cause free radical damage that in turn can lead to fibrosis of any of the cardiac structures; manifestations include accelerated atherosclerosis, valvular stenosis, pericardial constriction, and/or a restrictive cardiomyopathy due to interstitial fibrosis.

KEY CONCEPTS

CARDIOMYOPATHY

- Cardiomyopathies are intrinsic cardiac muscle diseases that may be genetic (idiopathic) or due to well-defined causes.
- There are three general pathophysiologic categories of cardiomyopathy: dilated (90%), hypertrophic, and restrictive (least common).
- DCM results in systolic (contractile) dysfunction. Causes include myocarditis, toxic exposures (e.g., alcohol), and pregnancy. In up to 50% of cases, genetic mutations are causal, with titin truncation mutations representing up to 20% of cases of DCM.
- HCM results in diastolic (relaxation) dysfunction. Virtually all cases are due to autosomal dominant mutations in the proteins comprising the contractile apparatus.
- Restrictive cardiomyopathy results in a stiff, noncompliant myocardium and can be due to deposition (e.g., amyloid), increased interstitial fibrosis (e.g., due to radiation), or endomyocardial scarring.
- Myocarditis is myocardial damage caused by inflammatory infiltrates secondary to infections or immune reactions. Viral infections are the most common causes in the United States. Clinically, myocarditis can be asymptomatic, give rise to acute heart failure, or evolve into DCM.

PERICARDIAL DISEASE

The most important pericardial disorders involve fluid accumulation, inflammation, fibrous constriction, or some combination of these processes, often in association with other cardiac pathology or a systemic disease.

Pericardial Effusion and Hemopericardium

Normally, the pericardial sac contains less than 50 mL of thin, clear, straw-colored fluid. Under various circumstances, the parietal pericardium may be distended by serous fluid (pericardial effusion), blood (hemopericardium), or pus (purulent pericarditis). With long-standing cardiac enlargement or with slowly accumulating fluid, the pericardium has time to remodel to accommodate the larger volume. This permits a slowly accumulating pericardial effusion to become quite substantial without restricting cardiac function. Thus, with chronic effusions of less than 500 mL in volume, the only clinical significance is a characteristic globular enlargement of the heart shadow on chest radiographs. In contrast, rapidly developing fluid collections of as little as 200 to 300 mL—e.g., due to hemopericardium caused by a ruptured MI or aortic dissection—can produce clinically devastating compression of the thin-walled atria and venae cavae, or the ventricles themselves; cardiac filling is thereby compromised, producing fatal cardiac tamponade.

Pericarditis

Pericardial inflammation can occur secondary to a variety of cardiac, thoracic, or systemic disorders, metastases from remote neoplasms, or cardiac surgical procedures. The major causes of pericarditis are listed in Table 12.14. Most evoke an acute pericarditis, but a few, such as tuberculosis and fungi, produce only chronic reactions.

Acute Pericarditis

Serous pericarditis is characteristically produced by noninfectious inflammatory diseases, including rheumatic fever, SLE, and scleroderma, as well as tumors and uremia. An infection in the tissues contiguous to the pericardium—for example, a bacterial pleuritis—may incite sufficient irritation of the parietal pericardial serosa to cause a sterile serous effusion that can progress to serofibrinous pericarditis and

Table 12.14 Causes of Pericarditis

Infectious Agents
Viruses
Pyogenic bacteria
Tuberculosis
Fungi
Other parasites
Presumably Immunologically Mediated
Rheumatic fever
Systemic lupus erythematosus
Scleroderma
Postcardiotomy
Postmyocardial infarction (Dressler) syndrome
Drug hypersensitivity reaction
Miscellaneous
Myocardial infarction
Uremia
After cardiac surgery
Neoplasia
Trauma
Radiation

ultimately to a frank suppurative reaction. In some cases, a well-defined viral infection elsewhere—upper respiratory tract, lung, parotid gland—antedates the pericarditis and serves as the primary focus of infection. Infrequently, usually in young adults, a viral pericarditis occurs as an apparent primary infection that may be accompanied by myocarditis (myopericarditis). Tumors can cause a serous pericarditis by lymphatic invasion or direct contiguous extension into the pericardium. Histologically, serous pericarditis elicits a mild inflammatory infiltrate in the epipericardial fat consisting predominantly of lymphocytes; tumor-associated pericarditis may also exhibit neoplastic cells. Organization into fibrous adhesions rarely occurs.

Fibrinous and serofibrinous pericarditis are the most frequent types of pericarditis; these are composed of serous fluid variably admixed with a fibrinous exudate. Common causes include acute MI (see Fig. 12.17D), postinfarction (Dressler) syndrome (an autoimmune response appearing weeks after an MI), uremia, rheumatic fever, SLE, and trauma. A fibrinous reaction also follows routine cardiac surgery. Radiation used to treat breast, lung, or mediastinal neoplasms can cause pericarditis, pericardial effusion, and chronic pericardial disorders.

Symptoms of fibrinous pericarditis characteristically include pain (sharp, pleuritic, and position dependent) and fever; congestive failure may also be present. A loud pericardial friction rub is the most striking clinical finding. However, the collection of serous fluid can actually prevent rubbing by separating the two layers of the pericardium.

Purulent or suppurative pericarditis reflects an active infection caused by microbial invasion of the pericardial space; this can occur through the following:

- *Direct extension* from neighboring infections, such as an empyema of the pleural cavity, lobar pneumonia, mediastinal infections, or extension of a ring abscess through the myocardium or aortic root
- *Seeding from the blood*
- *Lymphatic extension*
- *Direct introduction* during cardiotomy

The exudate ranges from a thin cloudy fluid to frank pus up to 500 mL in volume, and tuberculous pericarditis can exhibit foci of caseation. The serosal surfaces are reddened, granular, and coated with the exudate (Fig. 12.36). Microscopically there is an acute inflammatory reaction, which sometimes extends into surrounding structures (mediastinopericarditis). Complete resolution is infrequent, and organization by scarring is the usual outcome. The intense inflammatory response and the subsequent scarring frequently produce constrictive pericarditis (see later). Clinical findings in the active phase resemble those seen in fibrinous pericarditis, although the frank infection leads to more marked systemic symptoms including spiking fevers and rigors.

Hemorrhagic pericarditis has an exudate composed of blood mixed with a fibrinous or suppurative effusion. It is most commonly caused by the spread of a malignant neoplasm to the pericardial space; cytologic examination of fluid removed through a pericardial tap often reveals neoplastic cells. Hemorrhagic pericarditis can also be found in bacterial infections, in persons with an underlying bleeding diathesis, and in tuberculosis. Hemorrhagic pericarditis often follows cardiac surgery and is occasionally responsible for significant blood loss or even tamponade, requiring reoperation. The clinical significance is similar to that of fibrinous or suppurative pericarditis.

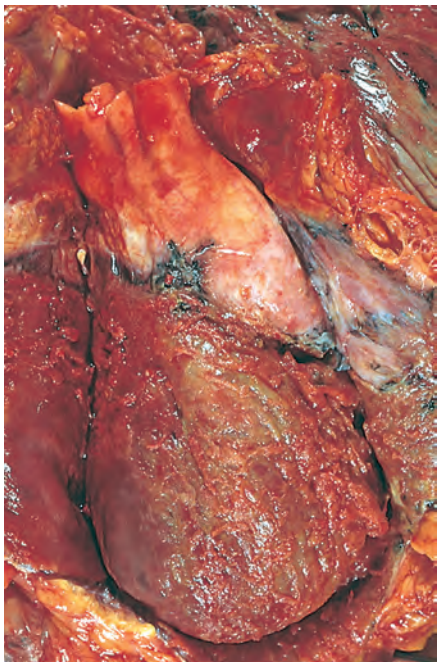


Figure 12.36 Acute suppurative pericarditis arising from direct extension of an adjacent pneumonia. Extensive purulent exudate is evident.

Chronic or Healed Pericarditis

Organization of pericardial inflammation may produce unobtrusive plaquelike fibrous thickenings of the serosal membranes ("soldiers' plaque") or thin, delicate adhesions that rarely cause impairment of cardiac function. In other cases, fibrosis in the form of meshlike stringy adhesions can largely obliterate the virtual space in the pericardial sac (adhesive pericarditis); in most instances, this has no effect on cardiac function. Two forms are worthy of some discussion.

- *Adhesive mediastinopericarditis* may follow infectious pericarditis, previous cardiac surgery, or mediastinal irradiation. The pericardial sac is obliterated, and adherence of the external aspect of the parietal layer to surrounding structures strains cardiac function. With each systolic contraction, the heart pulls not only against the parietal pericardium but also against the attached surrounding structures. Systolic retraction of the rib cage and diaphragm and pulsus paradoxus may be observed. The increased workload occasionally causes significant cardiac hypertrophy and dilation.
- In *constrictive pericarditis,* the heart is encased in a dense, fibrous, or fibrocalcific scar that limits diastolic expansion and cardiac output, features that mimic a restrictive cardiomyopathy. A prior history of pericarditis may or may not be present. The fibrous scar can be up to 1 centimeter in thickness, obliterating the pericardial space and sometimes calcifying; in extreme cases, it can resemble a plaster mold (concretio cordis). Because of the dense enclosing scar, cardiac hypertrophy and dilation

cannot occur. Cardiac output may be reduced at rest, but more importantly the heart has little if any capacity to increase its output in response to increased systemic demands. Signs of constrictive pericarditis include distant or muffled heart sounds, elevated jugular venous pressure, and peripheral edema. Treatment consists of surgical resection of the shell of constricting fibrous tissue (pericardiectomy).

TUMORS OF THE HEART

Primary Cardiac Tumors

Primary cardiac tumors are uncommon; moreover, most are (fortunately) benign. The five most common tumors have no malignant potential and account for almost 90% of all primary heart tumors. In descending order of frequency (combined pediatric and adult populations), these are myxomas, fibromas, lipomas, papillary fibroelastomas, and rhabdomyomas.

Primary malignant tumors of the heart are exceedingly rare, and usually are either an angiosarcoma or a sarcoma with poor differentiation and *MDM2* oncogene amplification. These and other sarcomas of the heart are not clinically or morphologically distinctive from sarcomas arising in other locations.

Fibromas and lipomas resemble their counterparts elsewhere; only the myxomas, papillary fibroelastomas, and rhabdomyomas merit further mention here.

Myxomas are the most common primary tumor of the adult heart (Fig. 12.37). These are benign neoplasms arising from primitive multipotent mesenchymal cells. Although sporadic myxomas do not show consistent genetic alterations, familial syndromes associated with myxomas have activating mutations in the *GNAS1* gene, encoding a subunit of G protein (Gsα) (in association with McCune-Albright syndrome) or null mutations in *PRKAR1A,* encoding a regulatory subunit of a cyclic-AMP–dependent protein kinase (Carney complex). About 90% of myxomas arise in the atria, with a left-to-right ratio of approximately 4:1.

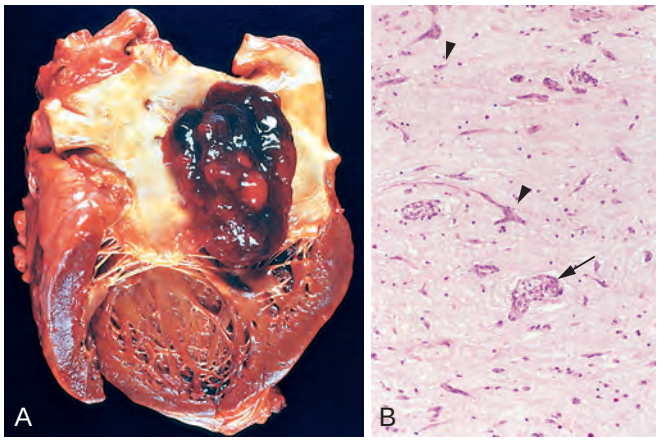


Figure 12.37 Atrial myxoma. (A) A large sessile lesion arises from the region of the fossa ovalis and extends into the mitral valve orifice. (B) Abundant amorphous extracellular matrix contains scattered multinucleate myxoma cells *(arrowheads)* in various groupings, including abnormal vessel-like formations *(arrow)*.

MORPHOLOGY

The tumors are usually single, but can rarely be multiple. The region of the fossa ovalis in the atrial septum is the favored site of origin. Myxomas range from small (<1 cm) to large (≥10 cm) and can be sessile or pedunculated lesions (see Fig. 12.37A). They vary from globular hard masses mottled with hemorrhage to soft, translucent, papillary, or villous lesions having a gelatinous appearance. The pedunculated form is often sufficiently mobile to move during systole into the atrioventricular valve opening, causing intermittent obstruction that may be position-dependent.

Histologically, myxomas are composed of stellate or globular myxoma cells embedded within an abundant acid mucopolysaccharide ground substance (see Fig. 12.37B). Peculiar vessel-like or glandlike structures are characteristic. Hemorrhage and mononuclear inflammation are usually present.

The major clinical manifestations are due to valvular "ball-valve" obstruction, embolization, or a syndrome of constitutional symptoms, such as fever and malaise. Sometimes fragmentation and systemic embolization calls attention to these lesions. Constitutional symptoms are probably due to the elaboration by some myxomas of the cytokine interleukin-6, a major mediator of the acute-phase response. Echocardiography provides the opportunity to identify these masses noninvasively. Surgical removal is usually curative; rarely, presumably with incomplete excision, the neoplasm can recur months to years later.

Papillary fibroelastomas are curious, sea-anemone–shaped endocardial tumors; they are usually incidental but can embolize and thereby become clinically important. Clonal cytogenetic abnormalities have been reported, suggesting that fibroelastomas are benign neoplasms. They resemble the much smaller, usually trivial, Lambl excrescences that project from sites of valve apposition.

MORPHOLOGY

Papillary fibroelastomas are usually (>80%) located on valves, particularly the ventricular surfaces of semilunar valves and the atrial surfaces of atrioventricular valves. Each lesion, typically 1 to 2 cm in diameter, consists of a distinctive cluster of hairlike projections up to 1 cm in length. Histologically, the projections are covered by a surface endothelium surrounding a core of myxoid connective tissue with abundant mucopolysaccharide matrix and elastic fibers.

Rhabdomyomas are the most frequent primary tumor of the pediatric heart; they are commonly discovered in the first years of life during an evaluation for a valve or other flow obstruction. Approximately one-half of cardiac rhabdomyomas are due to sporadic mutations; the other 50% of cases are associated with tuberous sclerosis (Chapter 28), with mutations in the tuberous sclerosis complex *TSC1* or *TSC2* tumor suppressor genes. The TSC1 and TSC2 proteins (hamartin and tuberin, respectively) function in a complex that inhibits the activity of the mammalian target of rapamycin (mTOR), a kinase that stimulates cell growth and regulates cell size. TSC1 or TSC2 expression is often absent in tuberous sclerosis–associated rhabdomyomas, providing a mechanism

for myocyte overgrowth. Because rhabdomyomas often regress spontaneously, they may be considered as hamartomas rather than true neoplasms.

MORPHOLOGY

Rhabdomyomas are gray-white myocardial masses that can be small or up to several centimeters in diameter. They are usually multiple and involve the ventricles preferentially, protruding into the lumen. Microscopically, they are composed of bizarre, markedly enlarged myocytes with large collections of glycogen. Routine histologic processing removes the glycogen and artifactually reduces the abundant cytoplasm to thin strands that stretch from the nucleus to the surface membrane, an appearance referred to as "spider" cells.

Metastatic Neoplasms

Metastatic tumors to the heart occur in roughly 5% of patients dying of cancer; the most frequent are lung and breast carcinomas, melanomas, leukemias, and lymphomas. Metastases can reach the heart and pericardium by retrograde lymphatic extension from the mediastinum (mostly carcinomas), by hematogenous seeding (many tumors), by direct contiguous extension (of intrathoracic tumors), or by venous extension (tumors of the kidney or liver). Clinical symptoms are most often associated with pericardial spread, which can cause symptomatic pericardial effusions or a mass-effect that is sufficient to restrict cardiac filling. Myocardial metastases are usually clinically silent or have nonspecific features, such as a generalized defect in ventricular contractility or compliance; occasionally they may cause arrhythmias. Bronchogenic carcinoma or malignant lymphoma can infiltrate the mediastinum extensively, causing encasement, compression, or invasion of the superior vena cava with resultant obstruction to blood coming from the head and upper extremities (superior vena cava syndrome). Renal cell carcinoma often invades the renal vein and can grow as a continuous column of tumor up the inferior vena cava lumen and into the right atrium, blocking venous return to the heart.

CARDIAC TRANSPLANTATION

Cardiac transplantation is performed for severe, intractable heart failure of diverse causes—most commonly DCM and IHD; over 3500 per year are performed worldwide. Three factors are primarily responsible for the dramatic improvement in allograft outcomes since the first transplant in 1967: (1) more effective immunosuppressive therapy (including the use of calcineurin inhibitors, mTOR inhibitors, glucocorticoids, and other agents), (2) careful selection of candidates, and (3) early histopathologic diagnosis of acute allograft rejection by endomyocardial biopsy.

Of the major complications, allograft rejection is the primary problem requiring surveillance; routine endomyocardial biopsy is the only reliable means of diagnosing acute cardiac rejection before substantial myocardial damage has occurred and at a stage that is reversible in the majority of instances.

Classic cellular rejection is characterized by interstitial lymphocytic inflammation with associated myocyte damage; the histology resembles myocarditis (Fig. 12.38A). There may also be interstitial edema due to vascular injury, and local cytokine elaboration can affect myocardial contractility without necessarily eliciting myocyte damage.

Increasingly, antibody-mediated rejection is also recognized as a mechanism of injury; donor-specific antibodies directed against major histocompatibility complex proteins lead to complement activation and the recruitment of Fc-receptor–bearing cells. These donor-specific antibodies cause allograft compromise by inducing endothelial cell injury leading to microvascular damage with thrombosis. Antibody-mediated rejection shows mild perivascular edema and scattered adherent intravascular inflammatory cells; it is often confirmed by immunohistochemical staining for complement fragment C4d, a long-lived catabolite that is released during complement cascade activation (see Fig. 12.38B and C).

Mild rejection may resolve spontaneously, and prompt recognition of more severe episodes allows successful treatment by augmenting baseline levels of immunosuppression; occasionally aggressive anti–T cell or anti–B cell immunotherapy, with or without plasmapheresis, may be necessary.

Allograft vasculopathy is the single most important long-term limitation for cardiac transplantation. It is a late, progressive, diffusely stenosing intimal proliferation in allograft blood vessels, predominantly in the coronary arteries (see Fig. 12.38D), leading to ischemic injury. Within 5 years of transplantation, 50% of patients develop significant allograft vasculopathy, and virtually all patients have lesions within 10 years. The pathogenesis of allograft vasculopathy involves immunologic responses that induce local production of growth factors which promote intimal smooth muscle cell recruitment and proliferation with ECM synthesis. Allograft vasculopathy is a particularly vexing problem because it can lead to silent MI (transplant patients have denervated hearts and do not typically experience angina), progressive CHF, or sudden cardiac death.

Other postoperative complications include infections and malignancies, particularly Epstein-Barr virus–associated B-cell lymphomas that arise in the setting of chronic T-cell immunosuppression. Despite these problems, the overall outlook is reasonably promising; the 1-year survival is 90%, and 5-year survival is greater than 70%.

CARDIAC DEVICES

Some of the most dramatic recent advances in cardiovascular medicine involve the use of mechanical devices. Thus, stents are now routinely deployed to maintain vascular patency (especially in the coronary arteries; Chapter 11), and endovascular conduit repairs are common approaches for treating abdominal aortic aneurysms. Implantable pacemakers and cardioverter defibrillators are indispensable for patients with cardiac rhythm disorders, and transcatheter aortic valve replacement (discussed earlier) has broadened the population that can benefit from valve replacement. Although impressive, these various devices are not without (somewhat predictable) risk. Mechanical failure can occur with any device, including strut fracture in stents, or lead fracture in pacing devices. Anything inserted into the cardiovascular system also has the potential to develop a thrombus that may become occlusive or embolize, and a foreign body is always a potential nidus for persistent (and difficult-to-treat)

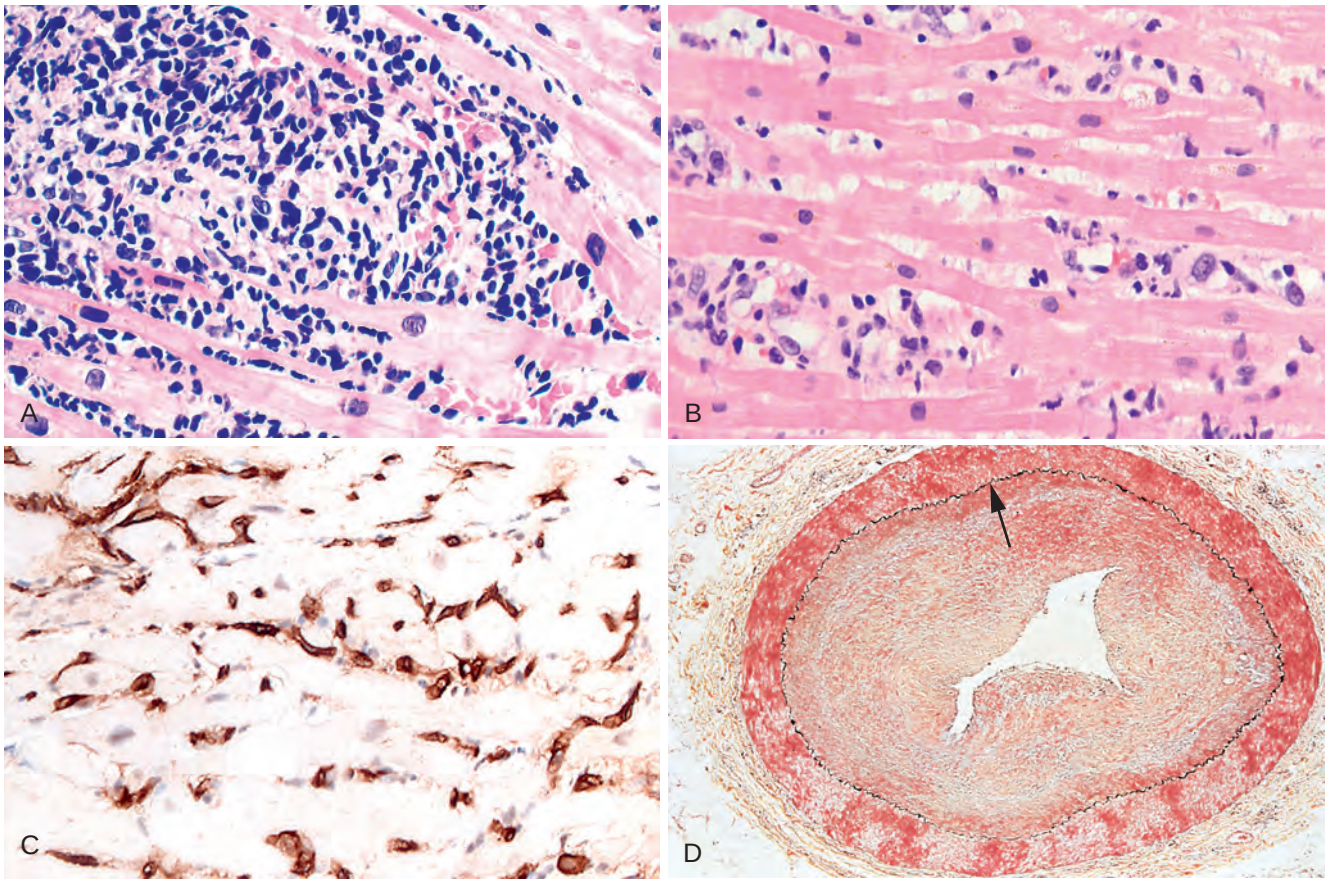


Figure 12.38 Complications of heart transplantation. (A) Cellular allograft rejection, typified by lymphocytic infiltrate associated with cardiac myocyte damage. B and C, Antibody-mediated rejection. (B) Hematoxylin and eosin–stained section shows plump, activated endothelium with adherent neutrophils and monocytes, but a sparse interstitial infiltrate; there is also perivascular edema. (C) immunohistochemistry for the complement C4d component shows sharp outlining of the capillaries *(brown stain)*, reflecting complement activation and deposition. (D) Allograft vasculopathy, with severe diffuse concentric intimal thickening producing critical stenosis. The internal elastic lamina *(arrow)* and media are intact (Movat pentachrome stain, elastin black). (C, Courtesy Dr. Robert Padera, Brigham and Women's Hospital, Boston, Mass.; D, Reproduced with permission from Salomon RN, Hughes CC, Schoen FJ, et al: Human coronary transplantation-associated arteriosclerosis. Evidence for chronic immune reaction to activated graft endothelial cells, *Am J Pathol* 138(4):791–798, 1991.)

infection. Fortunately, the manifest benefits significantly outweigh the potential risks.

Ventricular Assist Devices

Perhaps the most impressive engineering advances have occurred in the development of VADs. These allow the support of patients with cardiac failure caused by conditions ranging from transiently stunned myocardium after surgery, to permanent MI, to progressive cardiomyopathy. Given the limited availability of hearts suitable for transplantation, VADs can provide important life-sustaining support for end-stage failure patients, allowing them to survive (and indeed to improve their overall conditioning status) until a suitable donor can be identified. Increasingly, VADs are also being implanted as "destination therapy" for patients who are not otherwise candidates for transplantation (e.g., a patient with a history of malignancy and heart failure).

SUGGESTED READINGS

General

Buja L, Ottaviani G, Mitchell R: Pathobiology of cardiovascular diseases: an update, *Cardiovasc Pathol* 42:44, 2019. *[A review of cardiovascular pathobiology along with many useful references].*

Zipes DP et al, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 11, Philadelphia, 2018, Elsevier. *[An outstanding and authoritative text, with excellent sections on heart failure and atherosclerotic cardiovascular disease].*

Cardiac Structure and Specializations

Chien KR, Frisén J, Fritsche-Danielson R et al: Regenerating the field of cardiovascular cell therapy, *Nat Biotechnol* 37:232, 2019. *[A balanced review of research in cardiovascular regeneration].*

Cui M, Wang Z, Bassel-Duby R et al: Genetic and epigenetic regulation of cardiomyocytes in development, regeneration and disease, *Development* 145:24, 2018. *[A review of cardiomyocyte regenerative biology with a focus on basic science].*

Eschenhagen T et al: Cardiomyocyte regeneration: a consensus statement, *Circulation* 136:680, 2017. *[A brief outline consensus overview of the current evidence for cardiomyocyte regeneration].*

Lee RT, Walsh K: The future of cardiovascular regenerative medicine, *Circulation* 133:2618, 2016. *[An overview of trends in cardiovascular regenerative medicine].*

Heart Failure

Kemp CD, Conte JV: The pathophysiology of heart failure, *Cardiovasc Pathol* 21:365, 2012. *[An excellent review of the pathophysiology of heart failure].*

Nakamura M, Sadoshima J: Mechanisms of physiological and pathological cardiac hypertrophy, *Nat Rev Cardiol* 15:387, 2018. *[Good introduction to cardiac hypertrophy and then detailed review of possible molecular mechanisms involved].*

Congenital Heart Disease

Andersen TA, Troelsen KD, Larsen LL: Of mice and men: molecular genetics of congenital heart disease, *Cell Mol Life Sci* 71:1327, 2013. *[An overview of the relationships between cardiac development and congenital heart disease].*

Dinardo JA: Heart failure associated with adult congenital heart disease, *Semin Cardiothorac Vasc Anesth* 17:44, 2013. *[A well-written summary of the consequences of congenital heart disorders seen in the adult population].*

Huang JB, Liu YL, Sun PW et al: Molecular mechanisms of congenital heart disease, *Cardiovasc Pathol* 19:e183, 2010. *[A comprehensive review of the genes and pathways underlying congenital heart disease].*

Thiene G, Frescura C: Anatomical and pathophysiological classification of congenital heart disease, *Cardiovasc Pathol* 19:259, 2010. *[An excellent treatise on classification of congenital heart disease].*

Triedman JK, Newburger JW: Trends in congenital heart disease: the next decade, *Circulation* 133:2716, 2016. *[A review of trends in congenital heart disease medical science and clinical management].*

Ischemic Heart Disease

Crea F, Libby P: Acute coronary syndromes: the way forward from mechanisms to precision treatment, *Circulation* 136:1155, 2017. *[An excellent, well-written review on plaque erosion and rupture in acute coronary syndromes].*

Hausenloy DJ, Garcia-Dorado D, Bøtker HE et al: Novel targets and future strategies for acute cardioprotection: position paper of the European Society of Cardiology Working Group on Cellular Biology of the Heart, *Cardiovasc Res* 113:564, 2017. *[An update on the mechanisms and potential therapeutic interventions for ischemia-reperfusion injury and for cardioprotection and ischemic preconditioning in limiting infarct size].*

Hausenloy DJ, Yellon DM: Ischaemic conditioning and reperfusion injury, *Nat Rev Cardiol* 13:193, 2016. *[A great review of the mechanisms and potential therapeutic approaches to limiting reperfusion injury after MI].*

Nabel EG, Braunwald E: A tale of coronary artery disease and myocardial infarction, *N Engl J Med* 366:54, 2012. *[A terrific overview of the history of our understanding of the pathophysiology of coronary artery disease, and the successes of informed therapeutic interventions].*

Thygesen K et al: Fourth universal definition of myocardial infarction, *Circulation* 138:e618, 2018. *[An important consensus paper about definitions of myocardial infarction, with some discussion of pathobiology].*

Arrhythmias

Bagnall RD et al: A prospective study of sudden cardiac death among children and young adults, *N Engl J Med* 374:2441, 2016. *[Good study and discussion of sudden death in the young, an area of much recent investigation].*

Bezzina CR, Lahrouchi N, Priori SG: Genetics of sudden cardiac death, *Circ Res* 116:1919, 2015. *[An up-to-date, well-organized description of the known ion channel disorders and cardiomyopathies that cause sudden cardiac death].*

Hypertensive Heart Disease

Drazner MH: The progression of hypertensive heart disease, *Circulation* 123:327, 2011. *[A nice review of the cardiac consequences of systemic hypertension].*

Olschewski A et al: Pathobiology, pathology and genetics of pulmonary hypertension, *Int J Cardiol* 272S:4, 2018. *[An excellent review of right-sided hypertension-associated pathology and pathobiology].*

Valvular Heart Disease

Baddour LM et al: Infective endocarditis in adults: diagnosis, antimicrobial therapy, and management of complications, *Circulation* 132:1435, 2015. *[A good, clinically oriented overview of the developments in microorganisms, diagnosis, and therapies for infective endocarditis].*

Grozinsky-Glasberg S, Grossman AB, Gross DJ: Carcinoid heart disease: from pathophysiology to treatment, *Neuroendocrinology* 101:263, 2015. *[A good review of the current thinking on pathophysiology, diagnosis, and treatment of this entity].*

Guilherme L, Köhler KF, Kalil J: Rheumatic heart disease: mediation by complex immune events, *Adv Clin Chem* 53:31, 2011. *[A well-written and scholarly discussion of the pathogenic mechanisms regarding rheumatic heart disease].*

Li C, Xu S, Gotlieb AI et al: The response to valve injury. A paradigm to understand the pathogenesis of heart valve disease, *Cardiovasc Pathol* 20:183, 2011. *[A nice overview of pathologic concepts in valvular disease].*

New SE, Aikawa E: Molecular imaging insights into early inflammatory stages of arterial and aortic valve calcification, *Circ Res* 108:1381, 2011. *[A good overview of the mechanisms leading to degenerative calcification on valves and vessels].*

Otto CM, Prendergast B: Aortic-valve stenosis—From patients at risk to severe valve obstruction, *N Engl J Med* 371:744, 2014. *[A nice review of pathogenesis, clinical features, and clinical management].*

Schoen FJ, Gotlieb AI: Heart valve health, disease, replacement, and repair: a 25-year cardiovascular pathology perspective, *Cardiovasc Pathol* 25:341, 2016. *[An excellent review on mechanisms of valve disease and therapeutic approaches].*

Cardiomyopathies

Arany Z, Elkayam U: Peripartum cardiomyopathy, *Circulation* 133:1397, 2016. *[A review of pathophysiology and clinical management].*

Arbustini E, Narula N, Tavazzi L et al: The MOGE(S) classification of cardiomyopathy for clinicians, *J Am Coll Cardiol* 64:304, 2014. *[A consensus document regarding an updated classification of cardiomyopathies, weighted to genetic etiologies rather than pathophysiology].*

Braunwald E: Cardiomyopathies: an overview, *Circ Res* 121:711, 2017. *[The introductory article in an excellent series of nine articles covering many aspects of cardiomyopathy and myocarditis on pages 711-854 in this issue of Circ Res].*

Buggey J, ElAmm CA: Myocarditis and cardiomyopathy, *Curr Opin Cardiol* 33:341, 2018. *[A nice review of etiology, pathogenesis, and clinical features].*

Burke MA, Cook SA, Seidman JG et al: Clinical and mechanistic insights into the genetics of cardiomyopathy, *J Am Coll Cardiol* 68:2871, 2016. *[A well-written and authoritative overview of the genetics and pathophysiology of various cardiomyopathies from one of the leading groups in the world].*

Corrado D, Link MS, Calkins H: Arrhythmogenic right ventricular cardiomyopathy, *N Engl J Med* 376:61, 2017. *[A nice review of etiology, pathogenesis, and clinical features].*

Garfinkel AC, Seidman JG, Seidman CE: Genetic pathogenesis of hypertrophic and dilated cardiomyopathy, *Heart Fail Clin* 14(2):139, 2018. *[An excellent review of the genetic defects underlying many of the cardiomyopathies].*

Johnson DB et al: Fulminant myocarditis with combination immune checkpoint blockade, *N Engl J Med* 375:1749, 2016. *[An interesting report of a cardiovascular disease arising from a new therapy].*

Kwak-Glass C, Mitchell RN: Winning the battle, but losing the war: mechanisms and morphology of cancer-therapy-associated cardiovascular toxicity, *Cardiovasc Pathol* 30:55, 2017. *[A review of cardiovascular toxicity with current cancer therapies].*

Maleszewski JJ: Cardiac amyloidosis: pathology, nomenclature, and typing, *Cardiovasc Pathol* 24:343, 2015. *[A good review of cardiac amyloidosis].*

Tumors of the Heart

Burke A, Tavora F: The 2015 WHO Classification of tumors of the heart and pericardium, *J Thorac Oncol* 11:441, 2016. *[The most recent consensus classification scheme for tumors of the heart and pericardium].*

Sawyer DB: Anthracyclines and heart failure, *N Engl J Med* 38:1154, 2013. *[A succinct overview of chemotherapeutic cardiotoxicity].*

Cardiac Transplantation

Bruneval P, Angelini A, Miller D et al: Improving Antibody-Mediated Rejection Diagnostics: Strengths, Unmet Needs, and Future Directions, *Am J Transplant* 17(1):42, 2017. *[A consensus report on the state of antibody-mediated rejection diagnosis in cardiac transplantation].*

Mitchell RN: Graft vascular disease: immune response meets the vessel wall, *Annu Rev Pathol* 4:19, 2009. *[A comprehensive overview of allograft arteriopathy, including animal models, pathogenic mechanisms, clinical diagnosis, and therapy].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 13

Diseases of White Blood Cells, Lymph Nodes, Spleen, and Thymus

CHAPTER CONTENTS

The components of the hematopoietic system have been traditionally divided into the *myeloid tissues,* which include the bone marrow and the cells derived from it (e.g., red cells, platelets, granulocytes, and monocytes), and the *lymphoid tissues,* consisting of the thymus, lymph nodes, and spleen. It is important to recognize, however, that this subdivision is artificial with respect to both the normal physiology of hematopoietic cells and the diseases affecting them. For example, although bone marrow contains relatively few lymphocytes, it is the source of all lymphoid progenitors and the home of long-lived plasma cells and memory lymphocytes. Similarly, neoplastic disorders of myeloid progenitor cells (myeloid leukemias) originate in the bone marrow but secondarily involve the spleen and (to a lesser degree) the lymph nodes. Some red cell disorders (e.g., immunohemolytic anemia, discussed in Chapter 14) result from the formation of autoantibodies, indicating a primary disorder of lymphocytes. Thus it is not possible to draw neat lines between diseases involving the myeloid and lymphoid tissues.

Recognizing this difficulty, we somewhat arbitrarily divide diseases of the hematopoietic tissues into two chapters. In this chapter we discuss white cell diseases and disorders affecting the spleen and thymus. In Chapter 14 we consider diseases of red cells and those affecting hemostasis. Before delving into specific diseases, we will briefly discuss the origins of hematopoietic cells, since many disorders of white cells and red cells involve disturbances of their normal development and maturation.

Normal Hematopoiesis

Blood cell progenitors first appear during the third week of embryonic development in the yolk sac. Cells derived from the yolk sac are the source of long-lived tissue macrophages such as microglial cells in the brain and Kupffer cells in the liver (Chapter 3), but the contribution of the yolk sac to blood formation, mainly in the form of embryonic red cells, is only transient. Definitive *hematopoietic stem cells* (HSCs) arise several weeks later in the mesoderm of the intraembryonic aorta/gonad/mesonephros region. During the third month of embryogenesis, HSCs migrate to the liver, which becomes the chief site of blood cell formation until shortly before birth. HSCs also take up residence in the fetal placenta; this pool of HSCs is of uncertain physiologic relevance but has substantial clinical importance, as HSCs harvested at birth from umbilical cord blood are used in therapeutic HSC transplantation. By the fourth month of development, HSCs shift in location yet again to the bone marrow. By birth, marrow throughout the skeleton is hematopoietically active, and hepatic hematopoiesis dwindles to a trickle, persisting only in scattered foci that become inactive soon after birth. After puberty, hematopoiesis ceases in distal bones and becomes restricted to the axial skeleton. Thus in normal adults, only about half of the marrow space is hematopoietically active.

The formed elements of blood—red cells, granulocytes, monocytes, platelets, and lymphocytes—have a common origin from HSCs, pluripotent cells that sit at the apex of a hierarchy of bone marrow progenitors (Fig. 13.1). Most evidence supporting this scheme comes from studies in mice, but human hematopoiesis is believed to proceed in a similar way. The development of mature blood cells from HSCs involves progressive commitment to increasingly specialized cell populations. HSCs give rise to several kinds of early progenitor cells that have more restricted differentiation potential, such that they ultimately produce mainly myeloid cells or lymphoid cells. The origins of lymphoid cells are revisited when tumors derived from these cells are discussed. These early progenitors in turn give "birth" to progenitors that are further constrained in their differentiation potential. Some of these cells are referred to as *colony-forming units* (see Fig. 13.1) because they produce colonies composed of specific kinds of mature cells when grown in culture. From the various committed progenitors are derived the morphologically recognizable precursors, such as myeloblasts, proerythroblasts, and megakaryoblasts, which are the immediate progenitors of mature granulocytes, red cells, and platelets.

HSCs have two essential properties that are required for the maintenance of hematopoiesis: pluripotency and the capacity for self-renewal. Pluripotency refers to the ability of a single HSC to generate all mature blood cells. When an HSC divides, at least one daughter cell must self-renew to avoid stem cell depletion. Self-renewing divisions occur within a specialized marrow niche, in which stromal cells and secreted factors nurture and protect the HSCs. As one might surmise from their ability to migrate during embryonic development, HSCs are not sessile. Particularly under conditions of stress, such as severe anemia or acute inflammation, HSCs are mobilized from the bone marrow and appear in the peripheral blood. In fact, HSCs used in transplantation are now mainly collected from the peripheral blood of donors treated with granulocyte colony-stimulating factor (G-CSF), one of the factors that can mobilize marrow HSCs from their stem cell niches.

The marrow response to short-term physiologic needs is regulated by hematopoietic growth factors through their effects on the committed progenitors. These growth factors are called *colony-stimulating factors (CSFs)* because they were discovered by their ability to support the growth of colonies of blood cells in vitro. Because mature blood elements are terminally differentiated cells with finite lifespans, their numbers must be constantly replenished. In current models of hematopoiesis, some divisions of HSCs give rise to cells referred to as *multipotent progenitors,* which are more proliferative than HSCs but have a lesser capacity for self-renewal (see Fig. 13.1). Division of multipotent progenitors gives rise to at least one daughter cell that leaves the stem cell pool and begins to differentiate. Once past this threshold, these newly committed cells lose the capacity for self-renewal and commence an inexorable journey down a road that leads to terminal differentiation and death. However, as these progenitors differentiate, they also proliferate rapidly in response to growth factors, expanding their numbers. Some growth factors, such as stem cell factor (also called KIT ligand) and FLT3 ligand, act through receptors that are expressed on very early committed progenitors. Others, such as erythropoietin, granulocyte-macrophage colony-stimulating factor (GM-CSF), G-CSF, and thrombopoietin, act through receptors that are expressed only on committed progenitors with more restricted differentiation potentials. Feedback loops involving these lineage-specific growth factors tune the marrow output, allowing the numbers of formed blood elements (red cells, white cells, and platelets) to be maintained within appropriate ranges (Table 13.1).

Many diseases alter the production of blood cells. The marrow is the ultimate source of most cells of the innate and adaptive immune system and responds to infectious or inflammatory challenges by increasing its output of granulocytes under the direction of specific growth factors

Table 13.1 Adult Reference Ranges for Blood Cells[a]

Cell Type	Range
White cells ($\times 10^3/\mu L$)	4.8–10.8
Granulocytes (%)	40–70
Neutrophils ($\times 10^3/\mu L$)	1.4–6.5
Lymphocytes ($\times 10^3/\mu L$)	1.2–3.4
Monocytes ($\times 10^3/\mu L$)	0.1–0.6
Eosinophils ($\times 10^3/\mu L$)	0–0.5
Basophils ($\times 10^3/\mu L$)	0–0.2
Red cells ($\times 10^3/\mu L$)	
Men	4.3–5
Women	3.5–5
Platelets ($\times 10^3/\mu L$)	150–450

[a]Reference ranges vary among laboratories. The reference ranges for the laboratory providing the result should always be used.

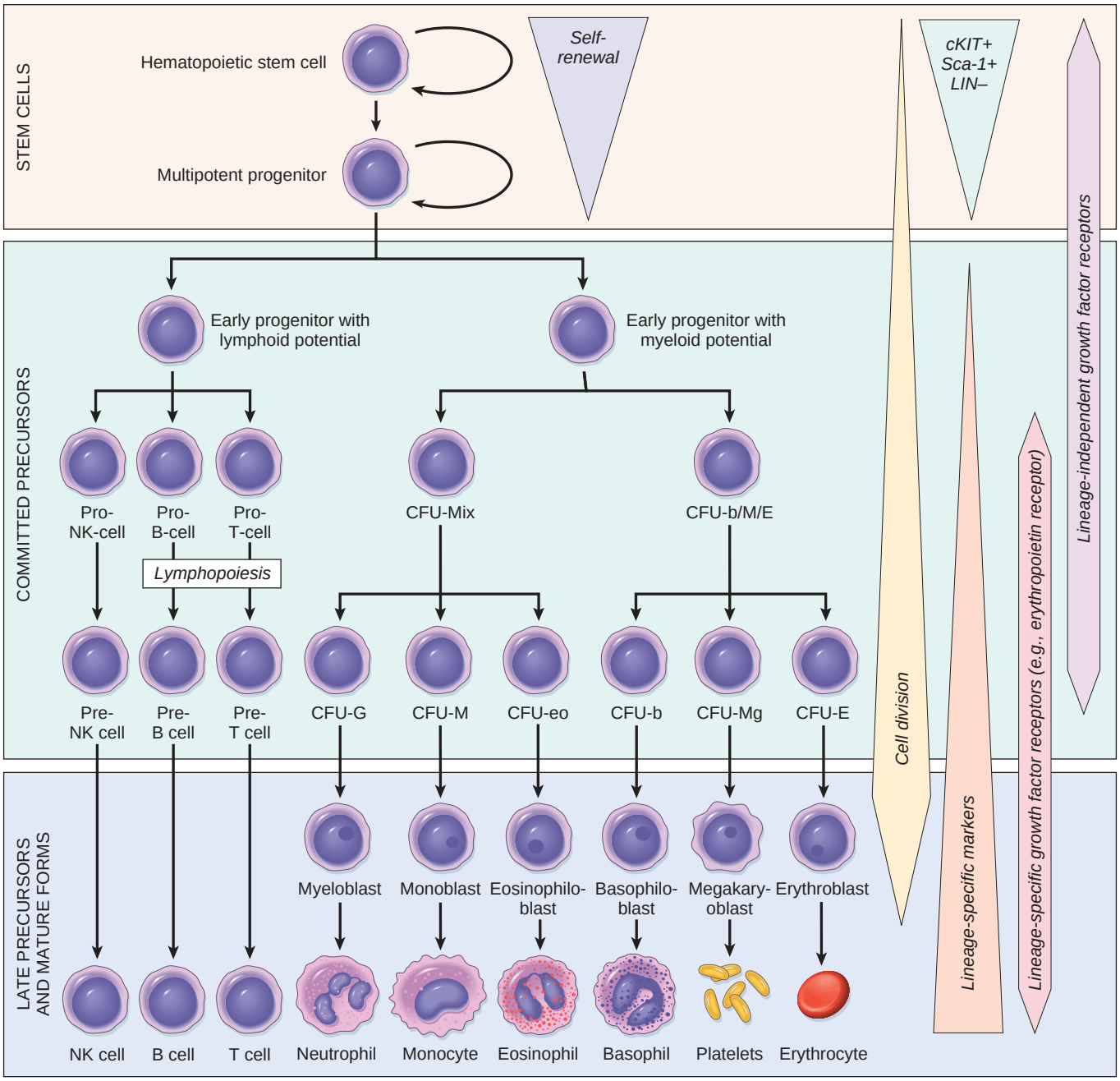


Figure 13.1 Differentiation of blood cells. *CFU*, Colony forming unit; *LIN–*, negative for lineage-specific markers; *NK*, natural killer.

and cytokines. By contrast, many other disorders are associated with defects in hematopoiesis that lead to deficiencies of one or more types of blood cells. Primary tumors of hematopoietic cells are among the most important diseases that interfere with marrow function, but certain genetic diseases, infections, toxins, and nutritional deficiencies, as well as chronic inflammation from any cause, may also decrease the production of blood cells by the marrow.

Tumors of hematopoietic origin are often associated with mutations that block progenitor cell maturation or abrogate their growth factor dependence. The net effect of such derangements is an unregulated clonal expansion of hematopoietic elements, which replace normal marrow progenitors and often spread to other hematopoietic tissues. In some instances these tumors originate from transformed HSCs that retain the ability to differentiate along multiple lineages, whereas in other instances the origin is a more differentiated progenitor that has acquired an abnormal capacity for self-renewal (Chapter 7).

MORPHOLOGY

The bone marrow is a unique microenvironment that supports the orderly proliferation, differentiation, and release of blood cells. It is filled with a network of thin-walled sinusoids lined by a single layer of endothelial cells, which are underlaid by a discontinuous basement membrane and adventitial cells. Within the

interstitium lie clusters of hematopoietic cells and fat cells. Differentiated blood cells enter the circulation by transcellular migration through the endothelial cells.

The normal marrow is organized in subtle, but important, ways. For example, normal megakaryocytes lie next to sinusoids and extend cytoplasmic processes that bud off into the bloodstream to produce platelets, while red cell precursors often surround macrophages that dispose of nuclear remnants produced when red cells disgorge their nucleus prior to release into the bloodstream. Processes that distort the marrow architecture, such as deposits of metastatic cancer or granulomatous disorders, can cause the abnormal release of immature precursors into the peripheral blood, a finding that is referred to as **leukoerythroblastosis.**

Marrow aspirate smears provide the best assessment of the morphology of hematopoietic cells. The most mature marrow precursors can be identified based on their morphology alone. Immature precursors ("blast" forms) of different types are morphologically similar and must be identified definitively using lineage-specific antibodies and histochemical markers (described later under white cell neoplasms). Bone marrow biopsies are a good means for estimating marrow activity. In normal adults the ratio of fat cells to hematopoietic elements is about 1:1. In hypoplastic states (e.g., aplastic anemia) the proportion of fat cells is greatly increased; conversely, fat cells often disappear when the marrow is involved by hematopoietic tumors and in diseases characterized by compensatory hyperplasias (e.g., hemolytic anemias) and neoplastic proliferations such as leukemias. Other disorders (e.g., metastatic cancers and granulomatous diseases) induce local marrow fibrosis. In such cases the marrow usually cannot be aspirated and the lesions are best seen in biopsies.

Disorders of White Cells

Disorders of white blood cells can be classified into two broad categories: proliferative disorders, in which there is an expansion of leukocytes, and leukopenias, which are defined as a deficiency of leukocytes. Proliferations of white cells can be *reactive* or *neoplastic.* Reactive proliferations in the setting of infections or inflammatory processes, when large numbers of leukocytes are needed for an effective host response, are fairly common. Neoplastic disorders, though less frequent, are much more important clinically. In the following discussion we will first describe the leukopenic states and summarize the common reactive disorders and then consider in some detail the malignant proliferations of white cells.

LEUKOPENIA

The number of circulating white cells may be decreased in a variety of disorders. An abnormally low white cell count *(leukopenia)* usually results from reduced numbers of neutrophils *(neutropenia, granulocytopenia)*. *Lymphopenia* is less common; in addition to congenital immunodeficiency diseases (Chapter 6), it is most commonly observed in advanced human immunodeficiency virus (HIV) infection, following therapy with glucocorticoids or cytotoxic drugs, autoimmune disorders, malnutrition, and certain acute viral infections. In the latter setting, lymphopenia actually stems from lymphocyte redistribution rather than a decrease in the number of lymphocytes in the body. Acute viral infections induce production of type I interferons, which activate T lymphocytes and change the expression of surface proteins that regulate T-cell migration. These changes result in the sequestration of activated T cells in lymph nodes and increased adherence to endothelial cells, both of which contribute to lymphopenia. Granulocytopenia is more common and is often associated with diminished granulocyte function and thus merits further discussion.

Neutropenia, Agranulocytosis

Neutropenia, a reduction in the number of neutrophils in the blood, occurs in a wide variety of circumstances. *Agranulocytosis,* a marked reduction in neutrophils, has the serious consequence of making individuals susceptible to bacterial and fungal infections.

Pathogenesis

Neutropenia can be caused by (1) inadequate or ineffective granulopoiesis or (2) increased destruction or sequestration of neutrophils in the periphery. Inadequate or ineffective granulopoiesis is observed in the setting of:

- *Suppression of HSCs,* as occurs in aplastic anemia (Chapter 14) and a variety of infiltrative marrow disorders (e.g., tumors, granulomatous disease); in these conditions, granulocytopenia is accompanied by anemia and thrombocytopenia.
- *Suppression of committed granulocytic precursors* by exposure to certain drugs (discussed later).
- Disease states associated with *ineffective hematopoiesis,* such as megaloblastic anemia (Chapter 14) and myelodysplastic syndrome, in which defective precursors die in the marrow.
- Rare *congenital conditions* (e.g., Kostmann syndrome), in which inherited defects in specific genes impair granulocytic differentiation.

Accelerated destruction or sequestration of neutrophils occurs with:

- *Immunologically mediated injury* to neutrophils, which can be idiopathic, associated with a well-defined immunologic disorder (e.g., systemic lupus erythematosus), or caused by exposure to drugs.
- *Splenomegaly,* in which splenic enlargement leads to sequestration and destruction of neutrophils in the spleen and modest neutropenia, sometimes associated with anemia and often with thrombocytopenia.
- *Increased peripheral utilization,* which can occur in overwhelming bacterial, fungal, or rickettsial infections.

The most common cause of agranulocytosis is drug toxicity. Certain drugs, such as alkylating agents and antimetabolites used in cancer treatment, produce agranulocytosis in a predictable, dose-related fashion. Because such drugs cause a generalized suppression of hematopoiesis,

production of red cells and platelets is also affected. Agranulocytosis may also occur as an idiosyncratic reaction to a large variety of agents including certain antibiotics, anticonvulsants, antiinflammatory drugs, antipsychotic drugs, and diuretics. The neutropenia induced by antipsychotic agents such as chlorpromazine and related phenothiazines results from a toxic effect on granulocytic precursors in the bone marrow. In contrast, agranulocytosis following administration of other drugs, such as sulfonamides, probably stems from antibody-mediated destruction of neutrophils through mechanisms similar to those involved in drug-induced immunohemolytic anemias (Chapter 14).

In some patients with acquired idiopathic neutropenia, autoantibodies directed against neutrophil-specific antigens are detected. Severe neutropenia may also occur in association with monoclonal proliferations of large granular lymphocytes (so-called *LGL leukemia*). The mechanism of this neutropenia is not clear; suppression of granulocytic progenitors by products of the neoplastic cell (usually a CD8+ cytotoxic T cell) is considered most likely.

MORPHOLOGY

The alterations in the **bone marrow** vary with cause. With excessive destruction of neutrophils in the periphery, the marrow is usually hypercellular due to a compensatory increase in granulocytic precursors. Hypercellularity is also the rule with neutropenias caused by ineffective granulopoiesis, as occurs in megaloblastic anemia and myelodysplastic syndrome. Agranulocytosis caused by agents that suppress or destroy granulocyte precursors is understandably associated with marrow hypocellularity.

Infections are a common consequence of agranulocytosis. Ulcerating necrotizing lesions of the gingiva, floor of the mouth, buccal mucosa, pharynx, or elsewhere in the oral cavity (agranulocytic angina) are quite characteristic. These are typically deep, undermined, and covered by gray to green-black necrotic membranes from which numerous bacteria or fungi can be isolated. Less frequently, similar ulcerative lesions occur in the skin, vagina, anus, or gastrointestinal tract. Severe life-threatening invasive bacterial or fungal infections may occur in the lungs, urinary tract, and kidneys. The neutropenic patient is at particularly high risk for deep fungal infections caused by *Candida* and *Aspergillus*. Sites of infection often show a massive growth of organisms with little leukocytic response. In the most dramatic instances, bacteria grow in colonies (botryomycosis) resembling those seen on agar plates.

Clinical Features

The symptoms and signs of neutropenia are related to infection and include malaise, chills, and fever, often followed by marked weakness and fatigability. With agranulocytosis, infections are often overwhelming and may cause death within hours to days.

Serious infections are most likely when the neutrophil count falls below 500/mm^3. Because infections are often fulminant, broad-spectrum antibiotics must be given expeditiously whenever signs or symptoms appear. In some instances, such as following myelosuppressive chemotherapy, neutropenia is treated with G-CSF, a growth factor that stimulates the production of granulocytes from marrow precursors.

REACTIVE PROLIFERATIONS OF WHITE CELLS AND LYMPH NODES

Leukocytosis

Leukocytosis refers to an increase in the number of white cells in the blood. It is a common reaction to a variety of inflammatory states.

Pathogenesis

The peripheral blood leukocyte count is influenced by several factors, including:

- The size of the myeloid and lymphoid precursor and storage cell pools in the bone marrow, thymus, circulation, and peripheral tissues.
- The rate of release of cells from the storage pools into the circulation.
- The proportion of cells that are adherent to blood vessel walls at any time (the marginal pool).
- The rate of extravasation of cells from the blood into tissues.

As discussed in Chapter 3, leukocyte homeostasis is maintained by cytokines, growth factors, and adhesion molecules through their effects on the proliferation, differentiation, and extravasation of leukocytes and their progenitors. Table 13.2 summarizes the major mechanisms of neutrophilic leukocytosis and its causes, the most important of which is infection. In acute infection there is a rapid increase in the egress of mature granulocytes from the bone marrow pool, an alteration that may be mediated through the effects of tumor necrosis factor (TNF) and interleukin-1 (IL-1). If the infection or an inflammatory process is prolonged, IL-1, TNF, and other inflammatory mediators stimulate macrophages, bone marrow stromal cells, and T cells to produce increased amounts of hematopoietic growth factors. These factors enhance the proliferation and differentiation of committed granulocytic progenitors and, over several days, cause a sustained increase in neutrophil production.

Some growth factors preferentially stimulate the production of a single type of leukocyte. For example, IL-5 mainly stimulates eosinophil production, while G-CSF induces

Table 13.2 Mechanisms and Causes of Leukocytosis

Increased Marrow Production
Chronic infection or inflammation (growth factor–dependent)
Paraneoplastic (e.g., Hodgkin lymphoma; growth factor–dependent)
Myeloproliferative neoplasms (e.g., chronic myeloid leukemia; growth factor–independent)
Increased Release From Marrow Stores
Acute inflammation (e.g., with infection)
Chronic inflammation (many causes)
Decreased Margination
Exercise
Catecholamines
Decreased Extravasation Into Tissues
Glucocorticoids

Table 13.3 Causes of Leukocytosis

Type of Leukocytosis	Causes
Neutrophilic leukocytosis	Acute bacterial infections, especially those caused by pyogenic organisms; sterile inflammation caused by, for example, tissue necrosis (myocardial infarction, burns)
Eosinophilic leukocytosis (eosinophilia)	Allergic disorders such as asthma, hay fever, parasitic infestations; drug reactions; certain malignancies (e.g., Hodgkin and some non-Hodgkin lymphomas); autoimmune disorders (e.g., pemphigus, dermatitis herpetiformis) and some vasculitides; atheroembolic disease (transient)
Basophilic leukocytosis (basophilia)	Rare, often indicative of a myeloproliferative neoplasm (e.g., chronic myeloid leukemia)
Monocytosis	Chronic infections (e.g., tuberculosis), bacterial endocarditis, rickettsiosis, and malaria; autoimmune disorders (e.g., systemic lupus erythematosus); inflammatory bowel diseases (e.g., ulcerative colitis)
Lymphocytosis	Accompanies monocytosis in many disorders associated with chronic immunologic stimulation (e.g., tuberculosis, brucellosis); viral infections (e.g., hepatitis A, cytomegalovirus, Epstein-Barr virus); *Bordetella pertussis* infection

neutrophilia. Such factors are differentially produced in response to various pathogenic stimuli, and, as a result, the five principal types of leukocytosis (neutrophilia, eosinophilia, basophilia, monocytosis, and lymphocytosis) tend to be observed in different clinical settings (Table 13.3).

In sepsis or severe inflammatory disorders (e.g., Kawasaki disease), leukocytosis is often accompanied by morphologic changes in neutrophils, such as toxic granulations, Döhle bodies, and cytoplasmic vacuoles (Fig. 13.2). *Toxic granules,* which are coarser and darker than normal neutrophilic granules, represent abnormal azurophilic (primary) granules. *Döhle bodies* are patches of dilated endoplasmic reticulum that appear as sky-blue cytoplasmic "puddles."

In most instances it is not difficult to distinguish reactive and neoplastic leukocytoses, but uncertainties may arise in two settings. Acute viral infections, particularly in children, can cause the appearance of large numbers of activated lymphocytes that resemble neoplastic lymphoid cells. At other times, particularly in severe infections, many immature granulocytes appear in the blood, mimicking a myeloid leukemia *(leukemoid reaction)*. Special laboratory studies (discussed later) are helpful in distinguishing reactive and neoplastic leukocytoses.

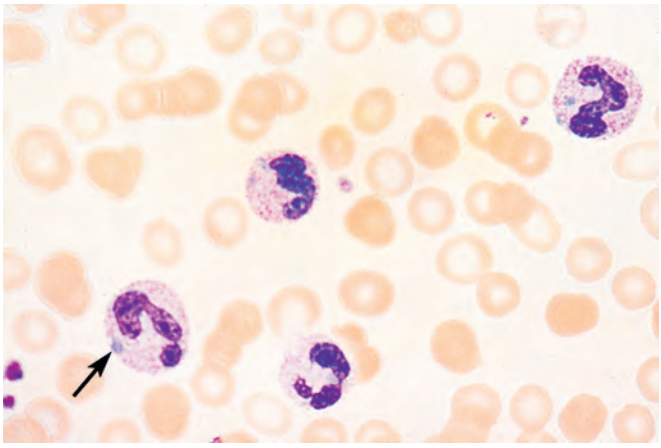


Figure 13.2 Reactive changes in neutrophils. Neutrophils containing coarse purple cytoplasmic granules (toxic granulations) and blue cytoplasmic patches of dilated endoplasmic reticulum (Döhle bodies) *(arrow)* are observed in this peripheral blood smear prepared from a patient with bacterial sepsis.

Lymphadenitis

Following their initial development from precursors in the central (also called primary) lymphoid organs—the bone marrow for B cells and the thymus for T cells—lymphocytes circulate through the blood and, under the influence of specific cytokines and chemokines, home to lymph nodes, spleen, tonsils, adenoids, and Peyer patches, which constitute the peripheral (secondary) lymphoid tissues. Lymph nodes, the most widely distributed and easily accessible lymphoid tissue, are frequently examined for diagnostic purposes. They are discrete encapsulated structures that contain separate B-cell and T-cell zones, each richly invested with phagocytes and antigen-presenting cells (see Fig. 6.8, Chapter 6).

The activation of resident immune cells leads to morphologic changes in lymph nodes. Within several days of antigenic stimulation, the primary follicles enlarge and develop pale-staining *germinal centers,* highly dynamic structures in which B cells acquire the capacity to make high-affinity antibodies against specific antigens. Paracortical T-cell zones may also undergo hyperplasia. The degree and pattern of morphologic change are dependent on the inciting stimulus and the intensity of the response. Trivial injuries and infections induce subtle changes, while more significant infections inevitably produce nodal enlargement and sometimes leave residual scarring. For this reason, lymph nodes in adults are almost never "normal" or "resting," and it is often necessary to distinguish morphologic changes secondary to past experience from those related to present disease. Infections and inflammatory stimuli often elicit regional or systemic immune reactions within lymph nodes. Some that produce distinctive morphologic patterns are described in other chapters. Most, however, cause stereotypical patterns of lymph node reaction designated acute and chronic nonspecific lymphadenitis.

Acute Nonspecific Lymphadenitis

Acute lymphadenitis in the cervical region is most often due to drainage of microbes or microbial products from infections of the teeth or tonsils, while in the axillary or inguinal regions it is most often caused by infections in the extremities. Acute lymphadenitis also occurs in mesenteric lymph nodes in the setting of acute appendicitis and other inflammatory conditions involving the gut (including self-limiting viral infections), a differential diagnosis that plagues the surgeon. Systemic viral infections (particularly in

children) and bacteremia often produce acute generalized lymphadenopathy.

MORPHOLOGY

The nodes are swollen, gray-red, and engorged. Microscopically, there is prominence of large reactive germinal centers containing numerous mitotic figures. Macrophages often contain particulate debris derived from dead bacteria or necrotic cells. When pyogenic organisms are the cause, neutrophils are prominent, and the centers of the follicles may undergo necrosis; sometimes the entire node is converted to pus. With less severe reactions, scattered neutrophils infiltrate about the follicles and accumulate within the lymphoid sinuses. The endothelial cells lining the sinuses become activated and enlarge in size.

Nodes involved by acute lymphadenitis are swollen and painful. When abscess formation is extensive the nodes become fluctuant. The overlying skin is red. Sometimes, suppurative infections penetrate through the capsule of the node and track to the skin to produce draining sinuses. Healing of such lesions is associated with scarring.

Chronic Nonspecific Lymphadenitis

A wide variety of chronic immunologic stimuli may produce nonspecific lymphadenitis. Several different patterns of morphologic change are seen, often within the same lymph node.

MORPHOLOGY

Follicular hyperplasia is caused by stimuli that activate humoral immune responses. It is defined by the presence of large oblong germinal centers (secondary follicles), which are surrounded by a collar of small resting naive B cells (the mantle zone) (Fig. 13.3). Germinal centers are polarized, consisting of two distinct regions: (1) a dark zone with proliferating blast-like B cells (centroblasts) and (2) a light zone composed of B cells with irregular or cleaved nuclear contours (centrocytes). Interspersed among the germinal center B cells is an inconspicuous network of antigen-presenting follicular dendritic cells and macrophages (often referred to as **tingible-body macrophages**) containing the nuclear debris of B cells, which undergo apoptosis if they fail to produce an antibody with a high affinity for antigen.

Causes of follicular hyperplasia include rheumatoid arthritis, toxoplasmosis, and early HIV infection. This form of hyperplasia is morphologically similar to follicular lymphoma (discussed later). Features favoring a reactive (nonneoplastic) hyperplasia include (1) preservation of the lymph node architecture, including the interfollicular T-cell zones and the sinusoids, (2) marked variation in the shape and size of the follicles, and (3) the presence of frequent mitotic figures, phagocytic macrophages, and recognizable light and dark zones, all of which tend to be absent from neoplastic follicles.

Paracortical hyperplasia is caused by stimuli that trigger T-cell–mediated immune responses, such as acute viral infections (e.g., infectious mononucleosis). The T-cell regions typically contain immunoblasts, activated T cells three to four times the size of resting lymphocytes that have round nuclei, open chromatin, several prominent nucleoli, and moderate amounts of pale cytoplasm.

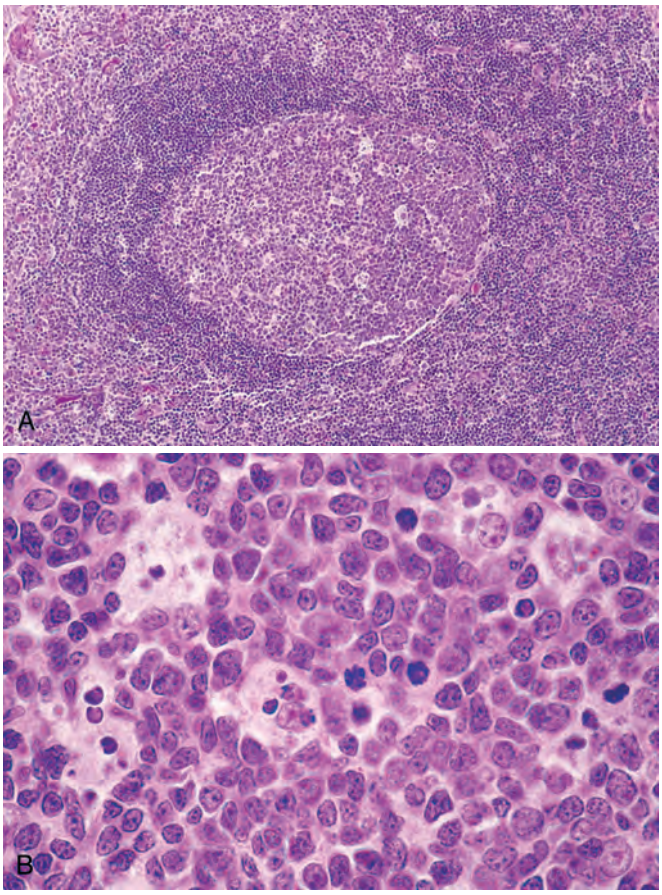


Figure 13.3 Follicular hyperplasia. (A) Low-power view showing a reactive follicle and surrounding mantle zone. The dark-staining mantle zone is more prominent adjacent to the germinal center light zone in the left half of the follicle. The right half of the follicle consists of the dark zone. (B) High-power view of the dark zone shows several mitotic figures and numerous macrophages containing phagocytosed apoptotic cells (tingible bodies).

The expanded T-cell zones encroach on and, in particularly exuberant reactions, may efface the B-cell follicles. In such cases, immunoblasts are so numerous that special studies may be needed to exclude a lymphoid neoplasm. In addition, there is often hypertrophy of sinusoidal and vascular endothelial cells, sometimes accompanied by infiltrating macrophages and eosinophils.

Sinus histiocytosis (also called *reticular hyperplasia*) is marked by an increase in the number and size of the endothelial cells that line lymphatic sinusoids and increased numbers of intrasinusoidal macrophages, which expand and distort the sinusoids. This form of hyperplasia may be particularly prominent in lymph nodes draining cancers such as carcinoma of the breast.

Characteristically, lymph nodes in chronic reactions are nontender, as enlargement occurs slowly over time and acute inflammation with associated tissue damage is absent. Chronic lymphadenitis is particularly common in inguinal and axillary nodes, which drain relatively large areas of the body and are frequently stimulated by immune reactions to trivial injuries and infections of the extremities.

Chronic immune reactions also can promote the appearance of organized collections of immune cells in nonlymphoid

tissues. These collections are sometimes called *tertiary lymphoid organs*. A classic example is that of chronic gastritis caused by *Helicobacter pylori,* in which aggregates of mucosal lymphocytes are seen that simulate the appearance of Peyer patches. A similar phenomenon occurs in rheumatoid arthritis, in which B-cell follicles often appear in the inflamed synovium. Lymphotoxin, a cytokine required for the formation of normal Peyer patches, is probably involved in the establishment of these "extranodal" inflammation-induced collections of lymphoid cells.

Hemophagocytic Lymphohistiocytosis

Hemophagocytic lymphohistiocytosis (HLH) is a reactive condition marked by cytopenias and signs and symptoms of systemic inflammation related to macrophage activation. For this reason, it is also sometimes referred to as *macrophage activation syndrome*. Some forms are familial and may appear early in life, even in infants, while other forms are sporadic and may affect people of any age.

Pathogenesis

The common feature of all forms of HLH is systemic activation of macrophages and CD8+ cytotoxic T cells. The activated macrophages phagocytose blood cell progenitors in the marrow and formed elements in the peripheral tissues, while the "stew" of mediators released from macrophages and lymphocytes suppress hematopoiesis and produce symptoms of systemic inflammation. These effects lead to cytopenias and a shock-like picture, sometimes referred to as "cytokine storm" or the systemic inflammatory response syndrome (Chapter 4).

Familial forms of HLH are associated with several different mutations, all of which impact the ability of cytotoxic T cells (CTLs) and natural killer (NK) cells to properly form or deploy cytotoxic granules. How these defects lead to HLH is not known. One common trigger for HLH is Epstein-Barr virus infection, suggesting that in some instances HLH stems from a defect in the ability of CD8+ CTLs to kill infected cells. As a result of the persistent infection, the CTLs continue to make cytokines, leading to excessive macrophage activation. HLH is also a common complication of peripheral T-cell lymphoma (discussed later), a tumor of mature T cells that is marked by immune dysregulation. Regardless of the trigger, HLH is uniformly associated with extremely high levels of inflammatory mediators such as interferon-γ, TNF-α, IL-6, and IL-12.

Clinical Features

Most patients present with an acute febrile illness associated with splenomegaly and hepatomegaly. Hemophagocytosis is usually seen on bone marrow examination, but is neither sufficient nor required to make the diagnosis. Laboratory studies typically reveal anemia, thrombocytopenia, and very high levels of plasma ferritin and soluble IL-2 receptor, both indicative of severe inflammation, as well as elevated liver function tests and triglyceride levels, both related to hepatitis. Coagulation studies may show evidence of disseminated intravascular coagulation. If untreated, this picture can progress rapidly to multiorgan failure, shock, and death.

Treatment involves the use of immunosuppressive drugs, "mild" chemotherapy, and administration of an antibody that neutralizes the activity of interferon-γ. Patients with germline mutations that cause HLH or who have persistent/resistant disease are candidates for HSC transplantation. Without treatment, the prognosis is grim, particularly in those with familial forms of the disease, who typically survive for less than 2 months. With prompt treatment, with or without subsequent HSC transplantation, roughly half of patients survive, though many do so with significant sequelae, such as renal damage in adults and growth retardation and intellectual disability in children.

NEOPLASTIC PROLIFERATIONS OF WHITE CELLS: OVERVIEW

Malignancies are clinically the most important disorders of white cells. These diseases fall into three broad categories:

- *Lymphoid neoplasms* include a diverse group of tumors of B-cell, T-cell, and NK-cell origin. In many instances the phenotype of the neoplastic cell closely resembles that of a particular lymphocyte class or stage of maturation, a feature that is used in the diagnosis and classification of these disorders.
- *Myeloid neoplasms* arise from early hematopoietic progenitors. Three categories of myeloid neoplasia are recognized: *acute myeloid leukemias (AMLs),* in which immature progenitor cells accumulate in the bone marrow; *myelodysplastic syndromes (MDSs),* which are associated with ineffective hematopoiesis and resultant peripheral blood cytopenias; and *myeloproliferative neoplasms,* in which increased production of one or more terminally differentiated myeloid elements (e.g., granulocytes) usually leads to elevated peripheral blood counts.
- The *histiocytoses* are uncommon proliferative lesions of macrophages and dendritic cells. Although "histiocyte" (literally, "tissue cell") is an archaic morphologic term, it is still often used. A special type of immature dendritic cell, the Langerhans cell, gives rise to a spectrum of neoplastic disorders referred to as the *Langerhans cell histiocytoses.*

Etiologic and Pathogenetic Factors in White Cell Neoplasia

As in other cancers, the development of white blood cell neoplasms involves genetic alterations, infections, and sometimes a background of chronic inflammation. Different types of tumors show different abnormalities and are, therefore, responsive to different therapies. Before delving into this complexity, we consider themes of general relevance to their etiology and pathogenesis.

Chromosomal Translocations and Other Acquired Mutations. Nonrandom chromosomal abnormalities, most commonly translocations, are present in the majority of white cell neoplasms. Many of these alterations are specifically associated with particular neoplasms and have a critical role in their genesis (Chapter 7).

- *Recurrently affected genes are often those that play crucial roles in the development, growth, or survival of the normal*

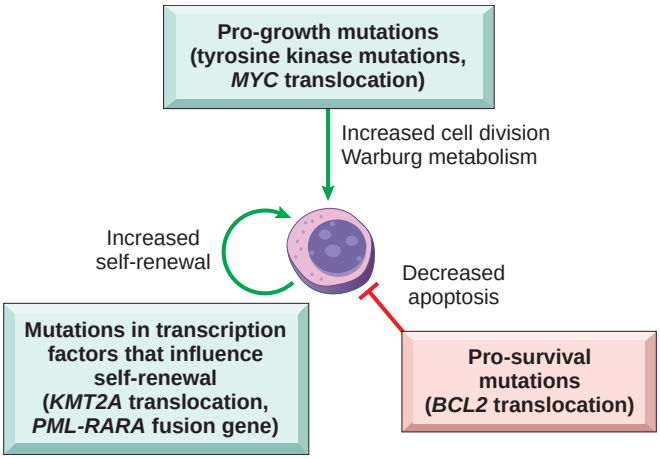


Figure 13.4 Pathogenesis of white cell malignancies. Various tumors harbor mutations that principally effect maturation or enhance self-renewal, drive growth, or prevent apoptosis. Examples of each type of mutation are listed; details are provided later under specific tumor types.

counterpart of the malignant cell. Mutations in certain genes are so strongly associated with specific tumor types that in some instances they are required for particular diagnoses. Some of these mutations produce a "dominant-negative" protein that interferes with a normal function (a loss of function); in others the result is an inappropriate increase in some normal activity (a gain of function).

- *Oncoproteins created by genomic aberrations often block normal maturation, turn on pro-growth signaling pathways, or protect cells from apoptotic cell death.* Fig. 13.4 highlights several well-characterized driver mutations and their pathogenic consequences in particular white cell neoplasms.
  - Many oncoproteins cause an arrest in differentiation, often at a stage when cells are proliferating rapidly. The importance of this mechanism is most evident in the acute leukemias, in which dominant-negative oncogenic mutations involving transcription factors interfere with early stages of lymphoid or myeloid cell differentiation.
  - Other mutations in transcriptional regulators directly enhance the self-renewal of tumors cells, giving them stem cell–like properties. These types of mutations often collaborate with mutations that constitutively activate tyrosine kinases, which in turn activate RAS and its two downstream signaling arms, the PI3K/AKT and MAPK pathways (Chapter 7), thereby driving cell growth.
  - Finally, mutations that inhibit apoptosis are prevalent in certain hematologic malignancies.
- *Proto-oncogenes are often activated in lymphoid cells by errors that occur during attempted antigen receptor gene diversification.* Among lymphoid cells, potentially oncogenic mutations occur most frequently in germinal center B cells. After antigen stimulation, B cells enter germinal centers and upregulate the expression of activation-induced cytosine deaminase (AID), a specialized DNA-modifying enzyme that is essential for two types of immunoglobulin (Ig) gene modifications: *class switching,* an intragenic recombination event in which the IgM heavy-chain constant gene segment is replaced with a different constant segment (e.g., IgG_3), leading to a switch in the class (isotype) of antibody produced, and *somatic hypermutation,* which creates point mutations within Ig genes that may increase antibody affinity for antigen (Chapter 6). Certain proto-oncogenes, such as *MYC,* are activated in germinal center B-cell lymphomas by translocations to the transcriptionally active Ig locus. Remarkably, AID expression induces *MYC/Ig* translocations in a small fraction of normal germinal center B cells, apparently because AID creates lesions in DNA that lead to chromosomal breaks. "Mistargeting" of AID also is implicated in point mutations that upregulate the expression and activity of BCL6, an oncogenic transcription factor with an important role in several B-cell malignancies. Another type of regulated genomic instability that is unique to precursor B and T cells is attributable to V(D)J recombinase, which cuts DNA at specific sites within the Ig and T-cell receptor loci, respectively. This process is essential for generating diversity in assembled antigen receptor genes but sometimes goes awry, leading to the joining of proto-oncogenes to antigen receptor gene regulatory elements. The resulting overexpression of the involved proto-oncogene converts it to an oncogene. This mechanism is particularly prevalent in tumors of precursor T cells but is observed in other types of lymphoid neoplasms as well.

Inherited Genetic Factors. As discussed in Chapter 7, individuals with genetic diseases that promote genomic instability, such as Bloom syndrome, Fanconi anemia, and ataxia telangiectasia, are at increased risk of acute leukemia. In addition, both Down syndrome (trisomy 21) and type I neurofibromatosis are associated with an increased incidence of childhood leukemia.

Viruses. Three lymphotropic viruses—human T-cell leukemia virus-1 (HTLV-1), EBV, and human herpesvirus-8 (HHV-8; also known as Kaposi sarcoma herpesvirus)—have been implicated as causative agents in particular lymphomas. The possible mechanisms of transformation by viruses are discussed in Chapter 7. HTLV-1 is associated with adult T-cell leukemia/lymphoma. EBV is found in a subset of Burkitt lymphoma, 30% to 40% of Hodgkin lymphoma (HL), many B-cell lymphomas arising in the setting of T-cell immunodeficiency, and rare NK-cell lymphomas. In addition to Kaposi sarcoma (Chapter 11), HHV-8 is associated with an unusual B-cell lymphoma that presents as a malignant effusion, often in the pleural cavity.

Chronic Inflammation. Several agents that cause localized chronic inflammation predispose to lymphoid neoplasia, which almost always arises within the inflamed tissue. Examples include the associations between *H. pylori* infection and gastric B-cell lymphomas (Chapter 17); gluten-sensitive enteropathy and intestinal T-cell lymphomas; and even breast implants, which are associated with an unusual subtype of T-cell lymphoma. This can be contrasted with HIV infection, which is associated with an increased risk of B-cell lymphomas that may arise within virtually any organ. Early in the course, T-cell dysregulation by HIV infection causes a systemic hyperplasia of germinal center B cells that is associated with an increased incidence of

germinal center B-cell lymphomas. In advanced infection (acquired immunodeficiency syndrome [AIDS]), severe T-cell immunodeficiency further elevates the risk for B-cell lymphomas, particularly those associated with EBV and KSHV/HHV-8. These relationships are discussed in more detail in Chapter 6.

Iatrogenic Factors. Ironically, radiation therapy and certain forms of chemotherapy used to treat cancer increase the risk of subsequent myeloid and lymphoid neoplasms. This association stems from the mutagenic effects of ionizing radiation and chemotherapeutic drugs on hematolymphoid progenitor cells.

Smoking. The incidence of AML is increased 1.3- to 2-fold in smokers, presumably because of exposure to carcinogens, such as benzene, in tobacco smoke.

LYMPHOID NEOPLASMS

Taken together, the diverse lymphoid neoplasms constitute a complex, clinically important group of cancers, with about 100,000 new cases being diagnosed each year in the United States.

Definitions and Classifications

Neoplasms that present with widespread involvement of the bone marrow and (usually, but not always) the peripheral blood are called *leukemias*. Proliferations of white cells, typically lymphocytes, that usually present as discrete tissue masses are called *lymphomas*. Originally these terms were attached to what were considered distinct entities, but with time and increased understanding these divisions have blurred. Many entities called "lymphoma" occasionally have leukemic presentations, and evolution to "leukemia" is not unusual during the progression of incurable "lymphomas." Conversely, tumors identical to "leukemias" sometimes arise as soft tissue masses without detectable bone marrow disease. Hence the terms leukemia and lymphoma merely reflect the usual tissue distribution of each disease at presentation.

Within the large group of lymphomas, *Hodgkin lymphoma* is segregated from other forms, which constitute the *non-Hodgkin lymphomas (NHLs)*. Hodgkin lymphoma has distinctive pathologic features and is treated in a unique fashion. Another special group of tumors includes *plasma cell neoplasms*. These most often arise in the bone marrow and only infrequently involve lymph nodes or the peripheral blood.

The clinical presentation of lymphoid neoplasms is most often determined by the anatomic distribution of disease. Two-thirds of NHLs and virtually all Hodgkin lymphomas present as enlarged nontender lymph nodes (often >2 cm). The remaining NHLs present with symptoms related to the involvement of extranodal sites (e.g., skin, stomach, or brain). Lymphocytic leukemias most often come to attention because of signs and symptoms related to the suppression of normal hematopoiesis by tumor cells in the bone marrow, whereas the most common plasma cell neoplasm, multiple myeloma, causes bony destruction of the skeleton and often presents with pain due to pathologic fractures. Other symptoms are frequently caused by proteins secreted from tumor cells or from immune cells responding to the tumor. Specific examples include the plasma cell tumors, in which much of the pathophysiology is related to the secretion of whole antibodies or Ig fragments; Hodgkin lymphoma, which is often associated with fever related to the release of cytokines from nonneoplastic inflammatory cells; and peripheral T-cell lymphomas, tumors of functional T cells that often release pro-inflammatory cytokines and chemokines.

Historically, few areas of pathology evoked as much controversy as the classification of lymphoid neoplasms, but consensus has been reached through use of objective molecular diagnostic tools. The current World Health Organization (WHO) classification scheme (Table 13.4) uses morphologic, immunophenotypic, genotypic, and clinical features to sort the lymphoid neoplasms into five broad categories, separated according to the cell of origin:

1. Precursor B-cell neoplasms (neoplasms of immature B cells)
2. Peripheral B-cell neoplasms (neoplasms of mature B cells)
3. Precursor T-cell neoplasms (neoplasms of immature T cells)

Table 13.4 World Health Organization Classification of Lymphoid Neoplasms

I. Precursor B-Cell Neoplasms
B-cell acute lymphoblastic leukemia/lymphoma (B-ALL)
II. Peripheral B-Cell Neoplasms
Chronic lymphocytic leukemia/small lymphocytic lymphoma
B-cell prolymphocytic leukemia
Lymphoplasmacytic lymphoma
Splenic and nodal marginal zone lymphomas
Extranodal marginal zone lymphoma
Mantle cell lymphoma
Follicular lymphoma
Marginal zone lymphoma
Hairy cell leukemia
Plasmacytoma/plasma cell myeloma
Diffuse large B-cell lymphoma
Burkitt lymphoma
III. Precursor T-Cell Neoplasms
T-cell acute lymphoblastic leukemia/lymphoma (T-ALL)
IV. Peripheral T-Cell and NK-Cell Neoplasms
T-cell prolymphocytic leukemia
Large granular lymphocytic leukemia
Mycosis fungoides/Sézary syndrome
Peripheral T-cell lymphoma, unspecified
Anaplastic large-cell lymphoma
Angioimmunoblastic T-cell lymphoma
Enteropathy-associated T-cell lymphoma
Panniculitis-like T-cell lymphoma
Hepatosplenic γδ T-cell lymphoma
Adult T-cell leukemia/lymphoma
Extranodal NK/T-cell lymphoma
NK-cell leukemia
V. Hodgkin Lymphoma
Classic subtypes
Nodular sclerosis
Mixed cellularity
Lymphocyte-rich
Lymphocyte depletion
Nodular lymphocyte predominant

NK, Natural killer.

4. Peripheral T-cell and NK-cell neoplasms (neoplasms of mature T cells and NK cells)
5. Hodgkin lymphomas (neoplasms of Reed-Sternberg cells and variants)

Before discussing the specific entities, some important principles relevant to lymphoid neoplasms should be emphasized.

- *Lymphoid neoplasia can be suspected based on clinical features, but histologic examination of lymph nodes or other involved tissues is required for diagnosis.* Analysis of lineage-specific protein (marker) expression and genetic alterations is an important complement to the morphologic studies. Markers recognized by antibodies that are helpful in the characterization of lymphomas and leukemias are listed in Table 13.5.
- *Antigen receptor gene rearrangement generally precedes transformation of lymphoid cells; hence all daughter cells derived from the malignant progenitor share the same antigen receptor gene configuration and sequence and synthesize identical antigen receptor proteins (either Igs or T-cell receptors).* In contrast, normal immune responses are comprised of polyclonal populations of lymphocytes that express many different antigen receptors. Thus, analyses of antigen receptor genes and their protein products can be used to distinguish reactive (polyclonal) and malignant (monoclonal) lymphoid proliferations. In addition, each antigen receptor gene rearrangement produces a unique DNA sequence that constitutes a highly specific clonal marker, which can be used to detect small numbers of residual malignant lymphoid cells after therapy.
- *Most lymphoid neoplasms resemble some recognizable stage of B- or T-cell differentiation* (Fig. 13.5), a feature that is used in their classification. The vast majority (85% to 90%) of lymphoid neoplasms are of B-cell origin, with most of the remainder being T-cell tumors; tumors of NK-cell origin are rare.
- *Lymphoid neoplasms are often associated with immune abnormalities.* Both a loss of protective immunity (susceptibility to infection) and a breakdown of tolerance (autoimmunity) may be seen, sometimes in the same patient. In a further ironic twist, individuals with inherited or acquired immunodeficiency are themselves at high risk of developing certain lymphoid neoplasms, particularly those caused by oncogenic viruses (e.g., EBV).
- *Neoplastic B and T cells tend to recapitulate the behavior of their normal counterparts.* Like normal lymphocytes, neoplastic B and T cells express adhesion molecules and chemokine receptors that govern their homing to certain tissue sites, leading to characteristic patterns of involvement. For example, follicular lymphomas home to germinal centers in lymph nodes, whereas cutaneous T-cell lymphomas home to the skin. Variable numbers of neoplastic B and T lymphoid cells also recirculate through the lymphatics and peripheral blood to distant sites; as a result, most lymphoid tumors are widely disseminated at the time of diagnosis. Notable exceptions to this rule include Hodgkin lymphomas, which are sometimes restricted to one group of lymph nodes, and marginal zone B-cell lymphomas, which are often restricted to sites of chronic inflammation.

Table 13.5 Some Immune Cell Antigens Detected by Monoclonal Antibodies

Antigen Designation	Normal Cellular Distribution
Primarily T-Cell Associated	
CD1	Thymocytes and Langerhans cells
CD3	Thymocytes, mature T cells
CD4	Helper T cells, subset of thymocytes
CD5	T cells and small subset of B cells
CD8	Cytotoxic T cells, subset of thymocytes, and some NK cells
Primarily B-Cell Associated	
CD10	Pre-B cells and germinal center B cells
CD19	Pre-B cells and mature B cells but not plasma cells
CD20	Pre-B cells after CD19 and mature B cells but not plasma cells
CD21	EBV receptor; mature B cells and follicular dendritic cells
CD23	Activated mature B cells
CD79a	Marrow pre-B cells and mature B cells
Primarily Monocyte or Macrophage Associated	
CD11c	Granulocytes, monocytes, and macrophages; also expressed by hairy cell leukemias
CD13	Immature and mature monocytes and granulocytes
CD14	Monocytes
CD15	Granulocytes; Reed-Sternberg cells and variants
CD33	Myeloid progenitors and monocytes
CD64	Mature myeloid cells
Primarily NK-Cell Associated	
CD16	NK cells and granulocytes
CD56	NK cells and a subset of T cells
Primarily Stem Cell and Progenitor Cell Associated	
CD34	Pluripotent hematopoietic stem cells and progenitor cells of many lineages
Activation Markers	
CD30	Activated B cells, T cells, and monocytes; Reed-Sternberg cells and variants
Present on All Leukocytes	
CD45	All leukocytes; also known as leukocyte common antigen (LCA)

CD, Cluster designation; *EBV,* Epstein-Barr virus; *NK,* natural killer.

- *Hodgkin lymphoma spreads in an orderly stepwise fashion, whereas most forms of NHL disseminate widely and somewhat unpredictably early in their course.* Hence, while lymphoma staging provides useful prognostic information, it is of most utility in guiding therapy in Hodgkin lymphoma.

We begin our discussion of specific entities with neoplasms of immature lymphoid cells and then move to mature B-cell tumors, plasma cell neoplasms, and tumors of T-cells and NK-cells. Some of the salient molecular and clinical features of these neoplasms are summarized in Table 13.6. We will finish by discussing Hodgkin lymphoma.

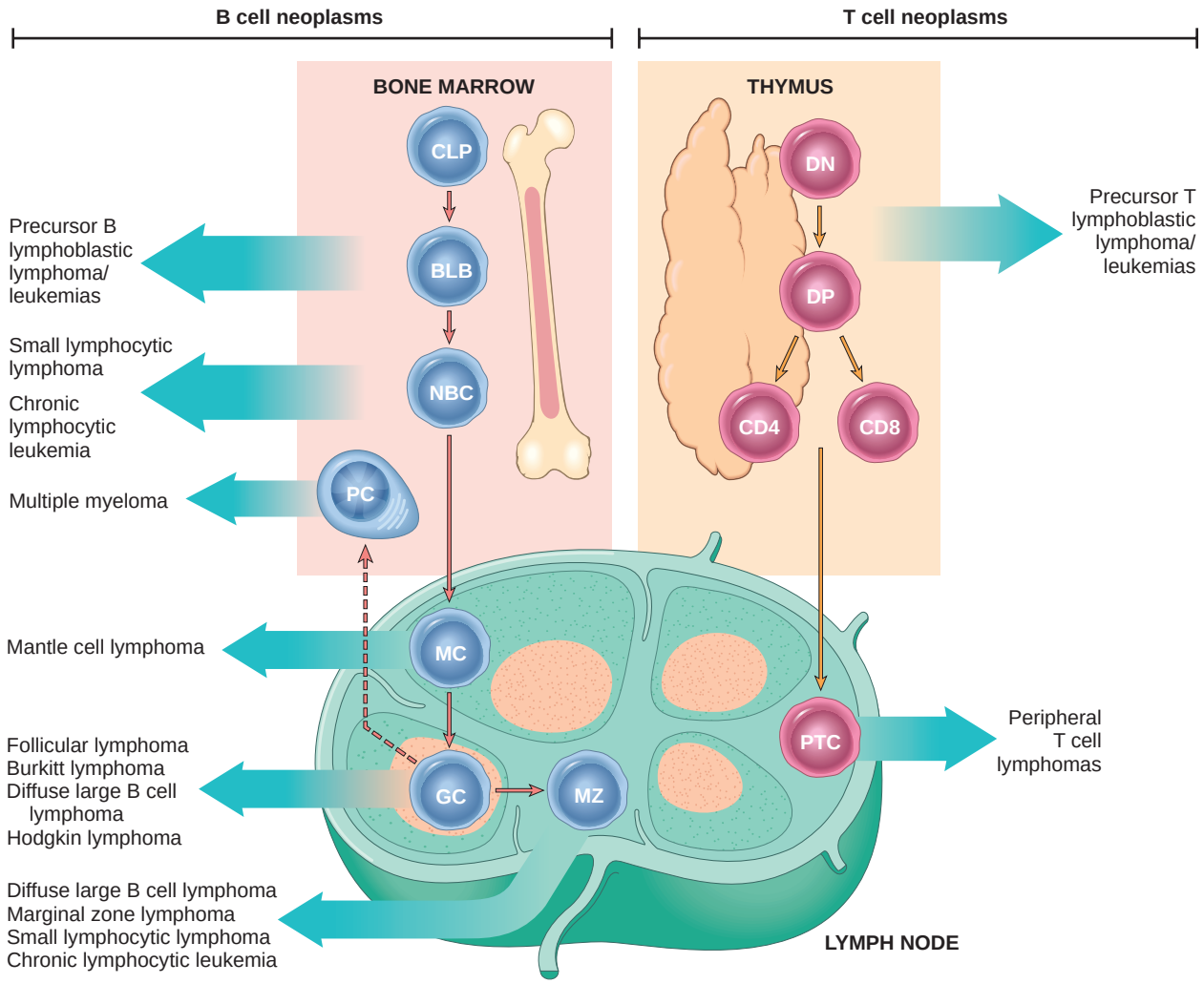


Figure 13.5 Origin of lymphoid neoplasms. Stages of B- and T-cell differentiation from which specific lymphoid tumors emerge are shown. *BLB*, Pre-B lymphoblast; *CLP*, common lymphoid precursor; *DN*, CD4/CD8 double-negative pro-T cell; *DP*, CD4/CD8 double-positive pre-T cell; *GC*, germinal center B cell; *MC*, mantle B cell; *MZ*, marginal zone B cell; *NBC*, naive B cell; *PTC*, peripheral T cell.

Precursor B- and T-Cell Neoplasms

Acute Lymphoblastic Leukemia/Lymphoma

Acute lymphoblastic leukemia/lymphomas (ALLs) are neoplasms composed of immature B (pre-B) or T (pre-T) cells, which are referred to as *lymphoblasts*. About 85% are B-ALLs, which typically manifest as childhood acute leukemias. The less common T-ALLs tend to present in adolescent males as thymic lymphomas. There is, however, considerable overlap in the clinical behavior of B-ALL and T-ALL; for example, B-ALL uncommonly presents as a mass in the skin or a bone, and many T-ALLs present with or evolve to a leukemic picture. Because of their morphologic and clinical similarities, the various forms of ALL are discussed here together.

ALL is the most common cancer of children. Approximately 2500 new cases are diagnosed each year in the United States, most occurring in individuals younger than 15 years of age. ALL is almost three times more common in Caucasians than in African-Americans and is slightly more frequent in boys than in girls. Hispanics have the highest incidence of any ethnic group. B-ALL peaks in incidence at about the age of 3, perhaps because the number of normal bone marrow pre-B cells (the cell of origin) is greatest very early in life. Similarly the peak incidence of T-ALL is in adolescence, the age when the thymus reaches maximum size. B-ALL and T-ALL also occur less frequently in adults of all ages.

Pathogenesis

Many of the chromosomal aberrations seen in ALL dysregulate the expression and function of transcription factors required for normal B- and T-cell development. Most T-ALLs have mutations in *NOTCH1*, a gene that is essential for T-cell development, while a high fraction of B-ALLs have mutations affecting genes such as *PAX5*, *TCF3*, *ETV6*, and *RUNX1*, all of which are required for the proper differentiation of early hematopoietic precursors. By disturbing the expression and function of "master" regulatory factors, these mutations promote maturation arrest and increased self-renewal, a stem cell–like phenotype. Similar themes are relevant in the genesis of AML (discussed later).

In keeping with the multistep origin of cancer (Chapter 7), mutations in transcription factor genes are not sufficient to produce ALL. The identity of other driver mutations is

Table 13.6 Summary of Major Types of Lymphoid Leukemias and Non-Hodgkin Lymphomas

Diagnosis	Cell of Origin	Genotype	Salient Clinical Features
Neoplasms of Immature B and T Cells			
B-cell acute lymphoblastic leukemia/lymphoma[a]	Bone marrow precursor B cell	Diverse chromosomal translocations; t(12;21) involving *RUNX1* and *ETV6* present in 25%	Predominantly children; symptoms relating to marrow replacement and pancytopenia; aggressive
T-cell acute lymphoblastic leukemia/lymphoma	Precursor T cell (often of thymic origin)	Diverse chromosomal translocations; *NOTCH1* mutations (50%–70%)	Predominantly adolescent males; thymic masses and variable bone marrow involvement; aggressive
Neoplasms of Mature B Cells			
Burkitt lymphoma[a]	Germinal center B cell	Translocations involving *MYC* and Ig loci, usually t(8;14); subset EBV-associated	Adolescents or young adults with extranodal masses; uncommonly presents as "leukemia"; aggressive
Diffuse large B-cell lymphoma[b]	Germinal center or post–germinal center B cell	Diverse chromosomal rearrangements, most often of *BCL6* (30%), *BCL2* (10%), or *MYC* (5%)	All ages, but most common in older adults; often appears as a rapidly growing mass; 30% extranodal; aggressive
Extranodal marginal zone lymphoma	Memory B cell	t(11;18), t(1;14), and t(14;18) creating *MALT1-IAP2*, *BCL10-IGH*, and *MALT1-IGH* fusion genes, respectively	Arises at extranodal sites in adults with chronic inflammatory diseases; may remain localized; indolent
Follicular lymphoma[b]	Germinal center B cell	t(14;18) creating *BCL2-IGH* fusion gene	Older adults with generalized lymphadenopathy and marrow involvement; indolent
Hairy cell leukemia	Memory B cell	Activating *BRAF* mutations	Older men with pancytopenia and splenomegaly; indolent
Mantle cell lymphoma	Naive B cell	t(11;14) creating cyclin D1–*IGH* fusion gene	Older men with disseminated disease; moderately aggressive
Multiple myeloma/solitary plasmacytoma[b]	Post–germinal center bone marrow homing plasma cell	Diverse rearrangements involving *IGH*; 13q deletions	Myeloma: older adults with lytic bone lesions, pathologic fractures, hypercalcemia, and renal failure; moderately aggressive Plasmacytoma: isolated plasma cell masses in bone or soft tissue; indolent
Small lymphocytic lymphoma/chronic lymphocytic leukemia	Naive B cell or memory B cell	Trisomy 12, deletions of 11q, 13q, and 17p; *NOTCH1* mutations; splicing factor mutations	Older adults with bone marrow, lymph node, spleen, and liver disease; autoimmune hemolysis and thrombocytopenia in a minority; indolent
Neoplasms of Mature T Cells or NK Cells			
Adult T-cell leukemia/lymphoma	Helper T cell	HTLV-1 provirus present in tumor cells	Adults with cutaneous lesions, marrow involvement, and hypercalcemia; occurs mainly in Japan, West Africa, and the Caribbean; aggressive
Peripheral T-cell lymphoma, unspecified	Helper or cytotoxic T cell	No specific chromosomal abnormality	Mainly older adults; usually presents with lymphadenopathy; aggressive
Anaplastic large-cell lymphoma	Cytotoxic T cell	Rearrangements of *ALK* (anaplastic large cell lymphoma kinase) in a subset	Children and young adults, usually with lymph node and soft tissue disease; aggressive
Extranodal NK/T-cell lymphoma	NK-cell (common) or cytotoxic T cell (rare)	EBV-associated; no specific chromosomal abnormality	Adults with destructive extranodal masses, most commonly sinonasal; aggressive
Mycosis fungoides/Sézary syndrome	Helper T cell	No specific chromosomal abnormality	Adult patients with cutaneous patches, plaques, nodules, or generalized erythema; indolent
Large granular lymphocytic leukemia	Two types: cytotoxic T cell and NK cell	Point mutations in *STAT3*	Adult patients with splenomegaly, neutropenia, and anemia, sometimes accompanied by autoimmune disease

[a]Most common tumors in children.

[b]Most common tumors in adults.

EBV, Epstein-Barr virus; *HTLV-1,* human T-cell leukemia virus-1; *Ig,* immunoglobulin; *NK,* natural killer.

incomplete, but aberrations that promote cell growth, such as mutations that increase tyrosine kinase activity and RAS signaling, are commonly present. Emerging data from deep sequencing of ALL genomes suggest that fewer than 10 mutations are sufficient to produce full-blown ALL; hence compared to solid tumors, ALL is genetically simple.

Approximately 90% of ALLs have numerical or structural chromosomal changes. Most common is hyperploidy (>50 chromosomes), but hypoploidy and a variety of balanced chromosomal translocations also are seen. Changes in chromosome numbers are of uncertain pathogenic significance but are important because they frequently correlate with immunophenotype and sometimes prognosis. For example, hyperdiploidy and hypodiploidy are seen only in B-ALL and are associated with better and worse prognoses, respectively. In addition, B-ALL and T-ALL are associated with completely

different sets of translocations; thus, while morphologically identical, they are genetically quite distinct.

MORPHOLOGY

In leukemic presentations, **the marrow is hypercellular and packed with lymphoblasts,** which replace normal marrow elements. Mediastinal thymic masses occur in 50% to 70% of T-ALLs, which are also more likely than B-ALL to be associated with lymphadenopathy and splenomegaly. In both B-ALL and T-ALL, the tumor cells have scant basophilic cytoplasm and nuclei somewhat larger than those of small lymphocytes (Fig. 13.6A). The nuclear chromatin is delicate and finely stippled, and nucleoli are usually small and often demarcated by a rim of condensed chromatin. The nuclear membrane is often deeply subdivided, imparting a convoluted appearance. In keeping with the aggressive clinical behavior, the mitotic rate is high. As with other rapidly growing lymphoid tumors, interspersed macrophages ingesting apoptotic tumor cells may impart a "starry sky" appearance (shown in Fig. 13.15).

Because of their different responses to chemotherapy, ALL must be distinguished from AML, a neoplasm of immature myeloid cells that can cause identical signs and symptoms. **Compared with myeloblasts, lymphoblasts have more condensed chromatin, less conspicuous nucleoli, and smaller amounts of cytoplasm that usually lacks granules.** However, these morphologic distinctions are not absolute, and definitive diagnosis relies on stains performed with antibodies specific for B- and T-cell antigens (Fig. 13.6B and C). Histochemical stains are also helpful, in that (in contrast to myeloblasts) lymphoblasts are myeloperoxidase-negative and often contain periodic acid–Schiff–positive cytoplasmic material.

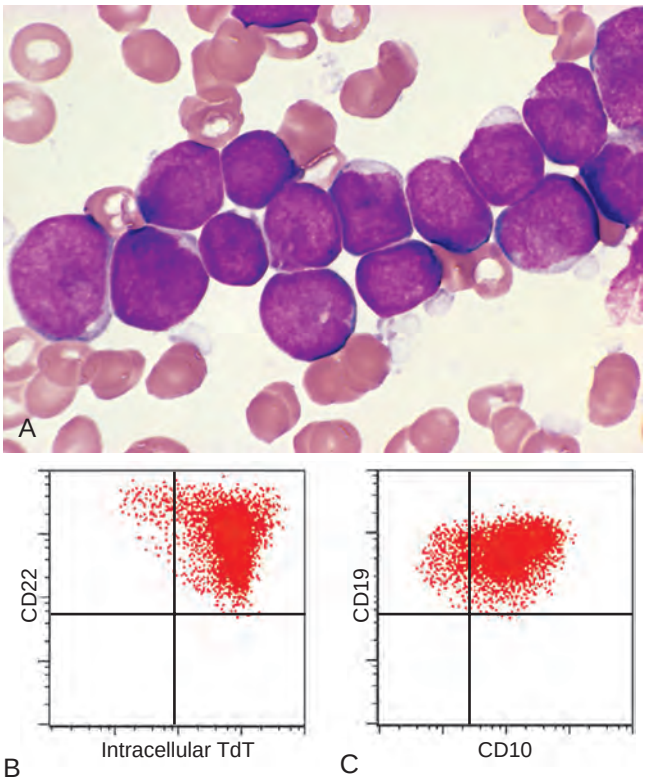


Figure 13.6 (A) Acute lymphoblastic leukemia/lymphoma (ALL). Lymphoblasts with condensed nuclear chromatin, small nucleoli, and scant agranular cytoplasm. (B and C) Phenotype of ALL shown in (A) analyzed by flow cytometry. (B) The lymphoblasts represented by the *red dots* express terminal deoxynucleotidyl transferase *(TdT)* and the B-cell marker CD22. (C) The same cells are positive for two other markers, CD10 and CD19, commonly expressed on pre-B lymphoblasts. Thus this is a B-ALL. (A, Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.; B and C, courtesy Dr. Louis Picker, Oregon Health Science Center, Portland, Ore.)

Immunophenotype. Immunostaining for terminal deoxynucleotidyl transferase (TdT), a specialized DNA polymerase that is expressed only in pre-B and pre-T lymphoblasts, is positive in more than 95% of cases (Fig. 13.6B). B-ALLs are arrested at various stages of pre–B-cell development, which correlate with the expression of certain proteins. The lymphoblasts usually express the pan B-cell marker CD19 and the transcription factor PAX5 as well as CD10. In very immature B-ALLs, CD10 is negative. Alternatively, more mature "late pre-B" ALLs express CD10, CD19, CD20, and cytoplasmic IgM heavy chain (μ chain). Similarly, T-ALLs are arrested at various stages of pre–T-cell development. In most cases the cells are positive for CD1, CD2, CD5, and CD7. The more immature tumors are usually negative for surface CD3, CD4, and CD8, whereas "late" pre–T-cell tumors are positive for these markers.

Clinical Features

Although ALLs and AMLs are genetically and immunophenotypically distinct, they are clinically very similar. In both, the accumulation of neoplastic "blasts" in the bone marrow suppresses normal hematopoiesis by physical crowding, competition for growth factors, and other poorly understood mechanisms. The common features and those more characteristic of ALL are the following:

- *Abrupt stormy onset* within days to a few weeks of the first symptoms.
- *Symptoms related to depression of marrow function,* including fatigue due to anemia; fever, reflecting infections secondary to neutropenia; and bleeding due to thrombocytopenia.
- *Mass effects caused by neoplastic infiltration* (which are more common in ALL), including bone pain resulting from marrow expansion and infiltration of the subperiosteum; generalized lymphadenopathy, splenomegaly, and hepatomegaly; testicular enlargement; and, in T-ALL, complications related to compression of large vessels and airways in the mediastinum.
- *Central nervous system manifestations* such as headache, vomiting, and nerve palsies resulting from meningeal spread, all of which are also more common in ALL.

Treatment of pediatric ALL is one of the great success stories of oncology. With aggressive chemotherapy about 95% of children with ALL obtain a complete remission, and 75% to 85% are cured. Despite these achievements, however, ALL remains a leading cause of cancer deaths in children, and only 35% to 40% of adults are cured. Several factors are associated with a worse prognosis: (1) age younger than 2 years, largely because of the strong association of infantile ALL with translocations involving the *MLL* gene; (2) presentation in adolescence or adulthood; and (3) peripheral

blood blast counts greater than 100,000, which probably reflects a high tumor burden. Favorable prognostic markers include (1) age between 2 and 10 years; (2) a low white cell count; (3) hyperdiploidy; (4) trisomy of chromosomes 4, 7, and 10; and (5) the presence of a t(12;21). Notably, the molecular detection of residual disease after therapy is predictive of a worse outcome in both B-ALL and T-ALL and is being used to guide clinical trials.

Although most chromosomal aberrations in ALL alter the function of transcription factors, the t(9;22) instead creates a fusion gene that encodes a constitutively active BCR-ABL tyrosine kinase (described in more detail under Chronic Myeloid Leukemia). In B-ALL, the BCR-ABL protein is usually 190 kDa in size and has stronger tyrosine kinase activity than the form of BCR-ABL that is found in chronic myeloid leukemia, in which a BCR-ABL protein of 210 kDa in size is usually seen. Treatment of t(9;22)-positive ALLs with BCR-ABL kinase inhibitors in combination with conventional chemotherapy is highly effective and has greatly improved the outcome for this molecular subtype of B-ALL in children and adults. Of interest, cryptic rearrangements involving genes encoding tyrosine kinases other than ABL also have been described in B-ALL, particularly in adults, and these too may be therapeutic targets. The outlook for adults with ALL lacking "targetable" molecular lesions remains more guarded, in part because of differences in the molecular pathogenesis of adult and childhood ALL, and also because older adults cannot tolerate the intensive chemotherapy regimens that are curative in children. Dramatic responses in B-ALL have been achieved with chimeric antigen receptor T cells directed against the B-cell antigen CD19 (Chapter 7), but at high cost and with associated toxicities that have sometimes proved fatal.

KEY CONCEPTS

ACUTE LYMPHOBLASTIC LEUKEMIA/ LYMPHOBLASTIC LYMPHOMA

- Most common type of cancer in children; may be derived from either precursor B or T cells.
- Highly aggressive tumors manifest with signs and symptoms of bone marrow failure or as rapidly growing masses.
- Tumor cells contain genetic lesions that block differentiation, leading to the accumulation of immature, nonfunctional blasts.
- A subset of tumors contains activating mutations in tyrosine kinases (e.g., BCR-ABL) that are important targets of therapy.

Peripheral B-Cell Neoplasms

Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma

Chronic lymphocytic leukemia (CLL) and small lymphocytic lymphoma (SLL) differ only in the degree of peripheral blood lymphocytosis. Most affected patients have sufficient lymphocytosis to fulfill the diagnostic requirement for CLL (absolute lymphocyte count >5000/mm^3). **CLL is the most common leukemia of adults in the Western world.** There are about 15,000 new cases of CLL each year in the United States. The median age at diagnosis is 60 years, and there is a 2:1 male predominance. In contrast, SLL constitutes only 4% of NHLs. CLL/SLL is much less common in Japan and other Asian countries than in the West.

Pathogenesis

Unlike most other lymphoid malignancies, chromosomal translocations are rare in CLL/SLL. The most common genetic anomalies are deletions of 13q14.3, 11q, and 17p and trisomy 12q. Molecular characterization of the region deleted on chromosome 13 has implicated two microRNAs, miR-15a and miR-16-1, tumor suppressor genes. Loss of these mIRs is believed to result in overexpression of the anti-apoptotic protein BCL2, which is uniformly observed in CLL/SLL. DNA sequencing has revealed that the Ig genes of some CLL/SLL are somatically hypermutated, whereas others are not, suggesting that the cell of origin may be either a post–germinal center memory B cell or a naive B cell. For unclear reasons, tumors with unmutated Ig segments (those putatively of naive B-cell origin) pursue a more aggressive course. Deep sequencing of CLL genomes has also revealed gain-of-function mutations involving the NOTCH1 receptor in 10% to 18% of tumors, as well as frequent mutations in genes that regulate RNA splicing.

The growth of CLL/SLL cells is largely confined to proliferation centers (described later), where tumor cells receive critical cues from the microenvironment. Stromal cells in proliferation centers express a variety of factors that stimulate the activity of the transcription factors nuclear factor kappa B (NF-κB), which promotes cell survival, and MYC, which promotes cell growth. Other critical signals are generated by the B-cell receptor (membrane bound Ig), which triggers a signaling pathway that includes Bruton tyrosine kinase (BTK). The importance of BTK in B cells is emphasized by congenital X-linked agammaglobulinemia (Chapter 6), a failure of B-cell development that is caused by loss-of-function mutations in the *BTK* gene. Notably, BTK inhibitors produce sustained therapeutic responses in a high fraction of CLL patients, indicating that CLL cells depend on B-cell receptor signaling and BTK activity for their growth and survival.

MORPHOLOGY

Lymph nodes are diffusely effaced by predominantly small lymphocytes 6 to 12 μm in diameter with round to slightly irregular nuclei, condensed chromatin, and scant cytoplasm (Fig. 13.7). Admixed are variable numbers of larger activated lymphocytes that often gather in loose aggregates referred to as **proliferation centers** that contain mitotically active cells. When present, **proliferation centers are pathognomonic for CLL/SLL.** The blood contains variable numbers of small round lymphocytes with scant cytoplasm (Fig. 13.8). Some of these cells are usually disrupted in the process of making smears, producing so-called **smudge cells.** In almost all cases, the bone marrow, spleen, and liver (Fig. 13.9) also are involved, albeit to widely varying degrees.

Immunophenotype. CLL/SLL has a distinctive immunophenotype. The tumor cells express the pan B-cell markers CD19 and CD20, as well as CD23 and CD5, the latter a marker that is found on a small subset of normal B cells. Low-level expression of surface Ig (usually IgM or IgM and IgD) is also typical, as is high-level expression of BCL2.

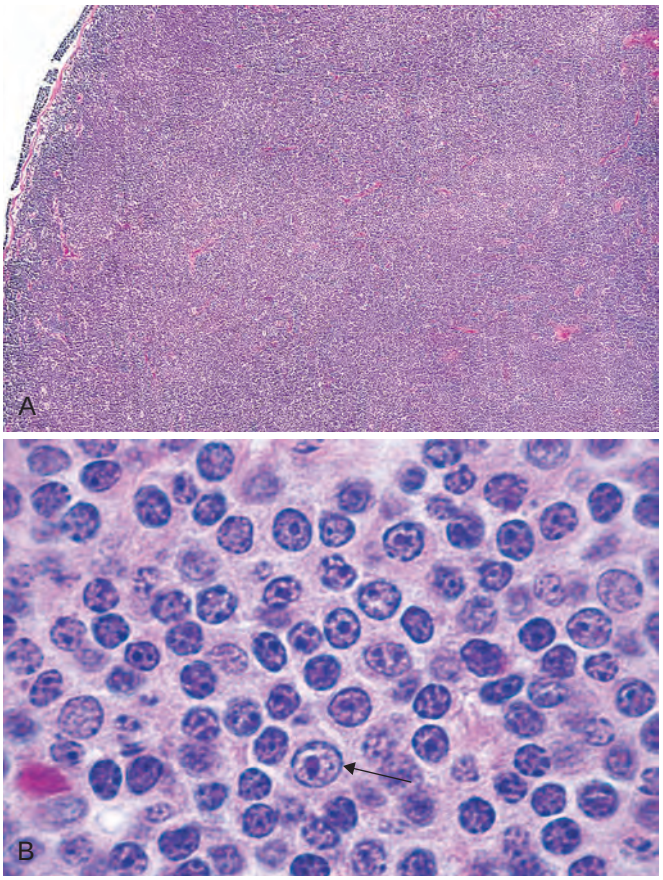


Figure 13.7 Small lymphocytic lymphoma/chronic lymphocytic leukemia (lymph node). (A) Low-power view shows diffuse effacement of nodal architecture. (B) At high power the majority of the tumor cells are small round lymphocytes. A prolymphocyte, a larger cell with a centrally placed nucleolus, is also present in this field *(arrow)*. (A, Courtesy Dr. José Hernandez, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

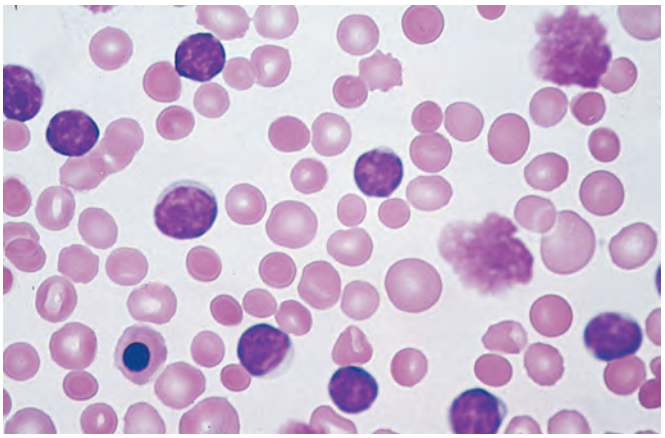


Figure 13.8 Chronic lymphocytic leukemia. This peripheral blood smear is flooded with small lymphocytes with condensed chromatin and scant cytoplasm. A characteristic finding is the presence of disrupted tumor cells (smudge cells), two of which are present in this smear. A coexistent autoimmune hemolytic anemia (Chapter 14) explains the presence of spherocytes (hyperchromatic, round erythrocytes). A nucleated erythroid cell is present in the lower left-hand corner of the field. In this setting, circulating nucleated red cells could stem from premature release of progenitors in the face of severe anemia, marrow infiltration by tumor (leukoerythroblastosis), or both.

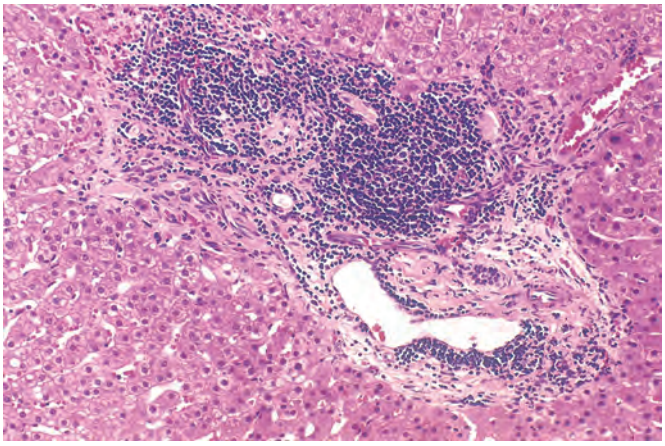


Figure 13.9 Small lymphocytic lymphoma/chronic lymphocytic leukemia involving the liver. Low-power view of a typical periportal lymphocytic infiltrate. (Courtesy Dr. Mark Fleming, Department of Pathology, Children's Hospital, Boston, Mass.)

Clinical Features

Patients are often asymptomatic at diagnosis. When symptoms appear, they are nonspecific and include easy fatigability, weight loss, and anorexia. Generalized lymphadenopathy and hepatosplenomegaly are present in 50% to 60% of symptomatic patients. The leukocyte count is highly variable; leukopenia can be seen in individuals with SLL and marrow involvement, while counts in excess of 200,000/mm^3 are sometimes seen in CLL patients with heavy tumor burdens. A small monoclonal Ig "spike" is present in the blood of some patients. At the other end of the spectrum are asymptomatic patients with monoclonal B cells in their peripheral blood that number too few to merit the diagnosis of CLL. This condition, referred to a monoclonal lymphocytosis of uncertain significance, is considered to be a precursor lesion and progresses to symptomatic CLL at a rate of 1% per year.

CLL/SLL disrupts normal immune function through uncertain mechanisms. Hypogammaglobulinemia is common and contributes to an increased susceptibility to infection, particularly those caused by bacteria. Conversely, 10% to 15% of patients develop hemolytic anemia or thrombocytopenia due to autoantibodies made by nonneoplastic B cells.

The course and prognosis are extremely variable and depend primarily on the clinical stage. Overall median survival is 4 to 6 years but is more than 10 years in individuals with minimal tumor burden at diagnosis. Other variables that correlate with a worse outcome include (1) the presence of deletions of 11q and 17p (the latter involving *TP53*), (2) a lack of somatic hypermutation, (3) the expression of ZAP-70, a protein that augments signals produced by the Ig receptor, and (4) the presence of *NOTCH1* mutations. Symptomatic patients are generally treated with "gentle" chemotherapy and immunotherapy with antibodies against proteins found on the surface of CLL/SLL cells, particularly CD20. Newly available, highly active targeted therapies include BTK inhibitors and BCL2 inhibitors; their impact on the disease course is still being determined.

Another factor that impacts patient survival is the tendency of CLL/SLL to transform to a more aggressive tumor. Most commonly this takes the form of a transformation to diffuse large B-cell lymphoma (DLBCL), so-called *Richter syndrome* (approximately 5% to 10% of patients). Richter syndrome is often heralded by the development of a rapidly enlarging mass within a lymph node or the spleen. It is often associated with acquisition of new abnormalities involving *TP53* or *MYC* and is an ominous event, with most patients surviving less than 1 year.

Follicular Lymphoma

Follicular lymphoma is the most common form of indolent NHL in the United States, trailing only diffuse large B-cell lymphoma (discussed later) in frequency among lymphomas. It affects 15,000 to 20,000 individuals per year. It usually presents in middle age and afflicts males and females equally. It is less common in Europe and rare in Asian populations.

Pathogenesis

Follicular lymphoma is strongly associated with chromosomal translocations involving *BCL2*. Its hallmark is a (14;18) translocation that juxtaposes the *IGH* locus on chromosome 14 and the *BCL2* locus on chromosome 18. The t(14;18) is seen in up to 90% of follicular lymphomas and leads to overexpression of BCL2 (see Fig. 13.12). BCL2 antagonizes apoptosis (Chapters 2 and 7) and promotes the survival of follicular lymphoma cells. Notably, while normal germinal centers contain numerous B cells undergoing apoptosis, follicular lymphoma is characteristically devoid of apoptotic cells. Deep sequencing of follicular lymphoma genomes have identified mutations in the *KMT2D* gene in about 90% of cases as well. *KMT2D* encodes a histone methyltransferase, suggesting that epigenetic abnormalities such as changes in the patterns of histone marks have an important role in this neoplasm.

Particularly early in the disease, follicular lymphoma cells growing in lymph nodes are found within a network of reactive follicular dendritic cells admixed with macrophages and T cells. Expression profiling studies have shown that differences in the genes expressed by these stromal cells are predictive of outcome, implying that the response of follicular lymphoma cells to therapy is influenced by the surrounding microenvironment.

MORPHOLOGY

In most cases, a nodular or nodular and diffuse growth pattern is observed in involved lymph nodes (Fig. 13.10A). Two principal cell types are present in varying proportions: (1) small cells with irregular or cleaved nuclear contours and scant cytoplasm, referred to as **centrocytes** (small cleaved cells), and (2) larger cells with open nuclear chromatin, several nucleoli, and modest amounts of cytoplasm, referred to as **centroblasts** (Fig. 13.10B). In most follicular lymphomas, centrocytes are in the majority. Peripheral blood involvement sufficient to produce lymphocytosis (usually <20,000 cells/mm^3) is seen in about 10% of cases. Bone marrow involvement occurs in 85% of cases and characteristically takes the form of paratrabecular lymphoid aggregates. The splenic white pulp (Fig. 13.11) and hepatic portal triads are also frequently involved.

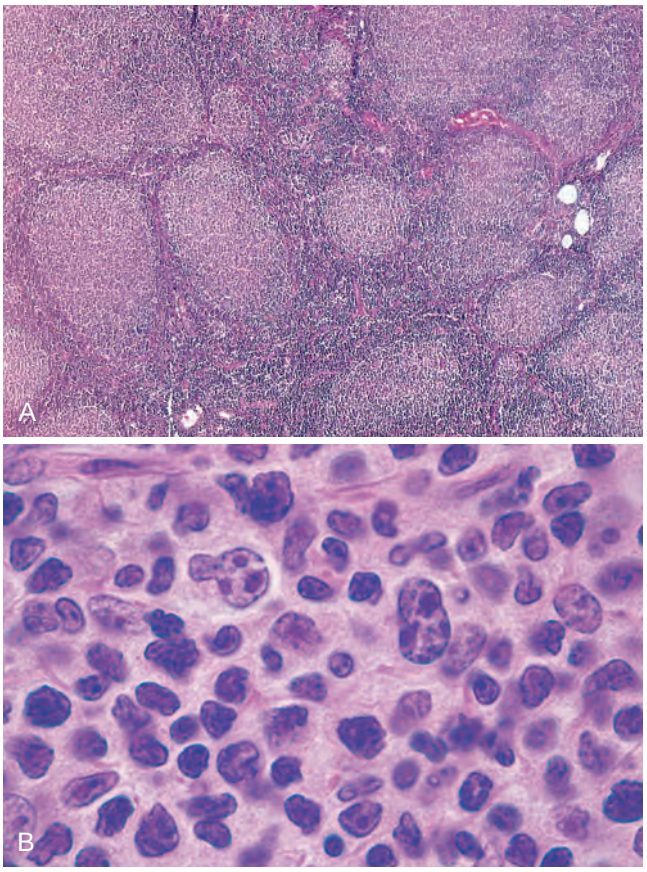


Figure 13.10 Follicular lymphoma (lymph node). (A) Nodular aggregates of lymphoma cells are present throughout lymph node. (B) At high magnification, small lymphoid cells with condensed chromatin and irregular or cleaved nuclear outlines (centrocytes) are mixed with a population of larger cells with nucleoli (centroblasts). (A, Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

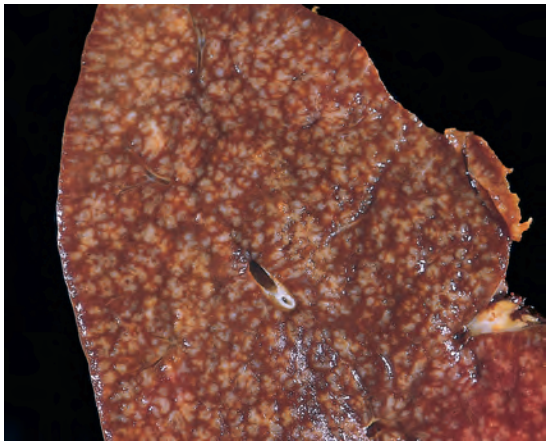


Figure 13.11 Follicular lymphoma (spleen). Prominent nodules represent white pulp follicles expanded by follicular lymphoma cells. Other indolent B-cell lymphomas (small lymphocytic lymphoma, mantle cell lymphoma, marginal zone lymphoma) can produce an identical pattern of involvement. (Courtesy Dr. Jeffrey Jorgenson, Department of Hematopathology, MD Anderson Cancer Center, Houston, Tex.)

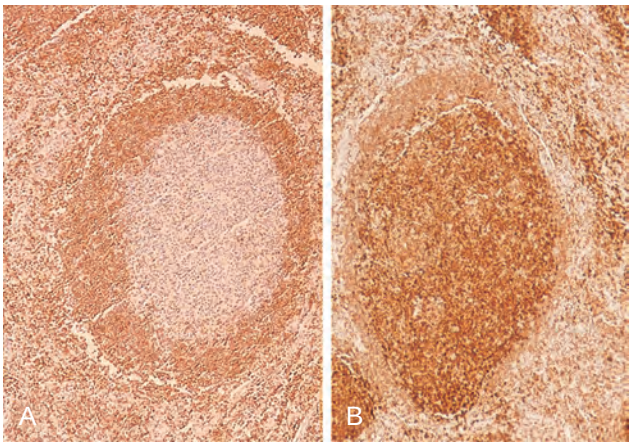


Figure 13.12 BCL2 expression in reactive and neoplastic follicles. BCL2 protein was detected by using an immunohistochemical technique that produces a brown stain. In reactive follicles (A), BCL2 is present in mantle zone cells but not follicular-center B cells, whereas follicular lymphoma cells (B) show strong BCL2 staining. (Courtesy Dr. Jeffrey Jorgenson, Department of Hematopathology, MD Anderson Cancer Center, Houston, Tex.)

Immunophenotype. The neoplastic cells closely resemble normal germinal center B cells, expressing CD19, CD20, CD10, surface Ig, and BCL6. Unlike CLL/SLL and mantle cell lymphoma, CD5 is not expressed. BCL2 is expressed in more than 90% of cases, in distinction to normal follicular center B cells, which are BCL2-negative (Fig. 13.12).

Clinical Features

Follicular lymphoma tends to present with painless, generalized lymphadenopathy. Involvement of extranodal sites, such as the gastrointestinal tract, central nervous system, or testis, is relatively uncommon. Although incurable, it usually follows an indolent waxing and waning course. Survival (median, 7 to 9 years) is not improved by aggressive therapy; hence the usual approach is to palliate patients with low-dose chemotherapy or immunotherapy (e.g., anti-CD20 antibody) when they become symptomatic. Like CLL, it also is responsive to inhibitors of B-cell receptor signaling (e.g., BTK inhibitors) and inhibitors of BCL2.

Histologic transformation occurs in 30% to 50% of follicular lymphomas, most commonly to DLBCL. These transformation events are frequently associated with aberrations that increase the expression of *MYC,* which you will recall drives Warburg metabolism and rapid cell growth. Follicular lymphomas show evidence of ongoing somatic hypermutation, which may promote transformation by causing point mutations or chromosomal aberrations. The median survival is less than 1 year after transformation.

Diffuse Large B-Cell Lymphoma

Diffuse large B-cell lymphoma (DLBCL) is the most common form of NHL. Each year in the United States there are about 25,000 new cases. There is a slight male predominance. The median patient age is about 60 years, but DLBCL also occurs in young adults and children.

Pathogenesis

Genetic, gene expression profiling, and immunohistochemical studies indicate that DLBCL is molecularly heterogeneous. One frequent pathogenic event is dysregulation of BCL6, a DNA-binding zinc-finger transcriptional repressor that is required for the formation of normal germinal centers. About 30% of DLBCLs contain various translocations that have in common a breakpoint in *BCL6* at chromosome 3q27. Acquired mutations in *BCL6* promoter sequences that abrogate *BCL6* autoregulation (an important negative-regulatory mechanism) are seen even more frequently. It is hypothesized that both types of lesions are inadvertent by-products of somatic hypermutation that result in overexpression of *BCL6,* which has several important consequences. BCL6 represses the expression of factors that normally serve to promote germinal center B-cell differentiation, growth arrest, and apoptosis, and each of these effects is believed to contribute to the development of DLBCL. Mutations similar to those found in *BCL6* are also seen in multiple other oncogenes, including *MYC,* suggesting that somatic hypermutation in DLBCL cells is "mistargeted" to a wide variety of loci.

Another 10% to 20% of tumors are associated with the t(14;18) (discussed earlier under Follicular Lymphoma), which leads to the overexpression of the antiapoptotic protein BCL2. Tumors with *BCL2* rearrangements usually lack *BCL6* rearrangements, suggesting that these rearrangements define two distinct molecular classes of DLBCL. Some tumors with *BCL2* rearrangements may arise from unrecognized underlying follicular lymphomas, which frequently transform to DLBCL. Roughly 5% of DLBCLs are associated with translocations involving *MYC;* these tumors may have a distinctive biology (discussed later under Burkitt lymphoma). Finally, sequencing of DLBCL genomes has identified frequent mutations in genes encoding histone acetyltransferases such as p300 and CREBP, proteins that regulate gene expression by modifying histones and altering chromatin structure.

MORPHOLOGY

The common features are a relatively **large cell size** (usually four to five times the diameter of a small lymphocyte) and a **diffuse pattern of growth** (Fig. 13.13). Other morphologic features show substantial variation. Most commonly, the tumor

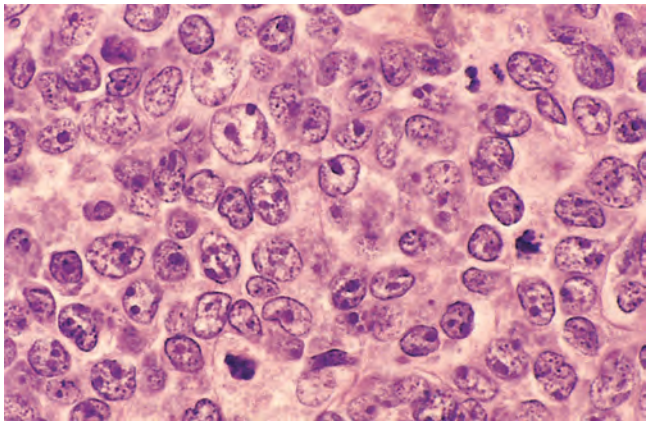


Figure 13.13 Diffuse large B-cell lymphoma. Tumor cells have large nuclei, open chromatin, and prominent nucleoli. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

cells have a round or oval nucleus that appears vesicular due to margination of chromatin to the nuclear membrane, but large multilobated or irregular nuclei are prominent in some cases. Nucleoli may be two to three in number and located adjacent to the nuclear membrane or single and centrally placed. The cytoplasm is usually moderately abundant and may be pale or basophilic. More anaplastic tumors may contain multinucleated cells with large inclusion-like nucleoli that resemble Reed-Sternberg cells (the malignant cell of Hodgkin lymphoma).

Immunophenotype. These mature B-cell tumors express CD19 and CD20 and show variable expression of germinal center B-cell markers such as CD10 and BCL6. Most have surface Ig. High-level expression of both MYC and BCL2 proteins is seen in some cases and may predict more aggressive behavior.

Special Subtypes. Several subtypes of DLBCL are sufficiently distinctive to merit brief discussion.

- *Immunodeficiency-associated large B-cell lymphoma* occurs in the setting of T-cell immunodeficiency (e.g., in advanced HIV infection and in recipients of organ or HSC transplants). The neoplastic B cells are usually infected with EBV, which plays a critical pathogenic role. Restoration of T-cell immunity may lead to regression of these proliferations.
- *Primary effusion lymphoma* presents as a malignant pleural or ascitic effusion, mostly in patients with advanced HIV infection or in older adults. The tumor cells are often anaplastic in appearance and typically fail to express surface B- or T-cell markers, but have clonal *IGH* gene rearrangements. In all cases the tumor cells are infected with HHV-8, which appears to have a causal role.

Clinical Features

DLBCL typically presents as a rapidly enlarging mass at a nodal or extranodal site. It can arise virtually anywhere in the body. Waldeyer ring, the oropharyngeal lymphoid tissue that includes the tonsils and adenoids, is involved commonly. Primary or secondary involvement of the liver and spleen may take the form of large destructive masses (Fig. 13.14).

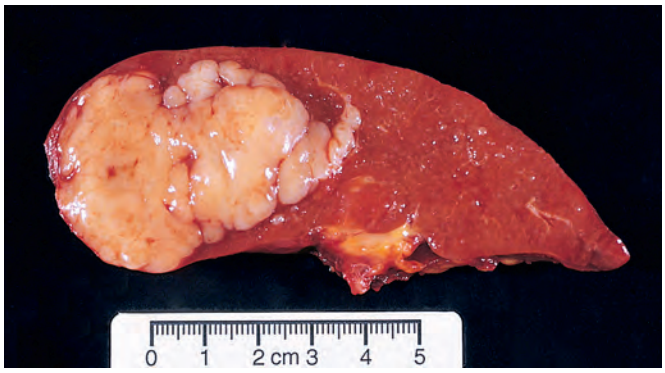


Figure 13.14 Diffuse large B-cell lymphoma involving the spleen. The isolated large mass is typical. In contrast, indolent B-cell lymphomas usually produce multifocal expansion of white pulp (see Fig. 13.11). (Courtesy Dr. Mark Fleming, Department of Pathology, Children's Hospital, Boston, Mass.)

Extranodal sites include the gastrointestinal tract, skin, bone, brain, and other tissues. Bone marrow involvement is relatively uncommon and usually occurs late in the course. Rarely, a leukemic picture emerges.

DLBCLs are aggressive tumors that are rapidly fatal without treatment. With intensive combination chemotherapy, 60% to 80% of patients achieve a complete remission, and 40% to 50% are cured. Adjuvant therapy with anti-CD20 antibody improves both the initial response and the overall outcome. Individuals with limited disease fare better than those with widespread disease or bulky tumor masses. Expression profiling has identified several distinct molecular subtypes, including one resembling germinal center B cells and a second resembling activated post–germinal center B cells, each with differing clinical outcomes. DLBCLs with *MYC* translocations have a worse prognosis than those without and may be better treated with chemotherapy regimens that are now standard for Burkitt lymphoma. CAR T cells directed against the B-cell antigen CD19 are now available for the treatment of patients with relapsed refractory DLBCL.

Burkitt Lymphoma

Within the category of Burkitt lymphoma fall (1) African (endemic) Burkitt lymphoma, (2) sporadic (nonendemic) Burkitt lymphoma, and (3) a subset of aggressive lymphomas occurring in individuals infected with HIV. Burkitt lymphomas occurring in these three settings are histologically identical but have distinct clinical, genotypic, and virologic characteristics.

Pathogenesis

All forms of Burkitt lymphoma are associated with translocations of the *MYC* gene on chromosome 8 that lead to increased MYC protein levels. MYC is a master transcriptional regulator that increases the expression of genes that are required for aerobic glycolysis, the so-called Warburg effect (Chapter 7). When nutrients such as glucose and glutamine are available, Warburg metabolism allows cells to biosynthesize all the building blocks—nucleotides, lipids, proteins—that are needed for growth and cell division. In line with the importance of MYC in regulating proliferation, Burkitt lymphoma is among the fastest-growing human tumors. The *MYC* translocation partner is usually the *IGH* locus [t(8;14)], but may also be the Ig κ [t(2;8)] or λ [t(8;22)] light chain loci. The breakpoints in the *IGH* locus in sporadic Burkitt lymphoma are usually found in the class switch regions, whereas the breakpoints in endemic Burkitt lymphoma tend to lie within more 5′ V(D)J sequences. The basis for this subtle molecular distinction is not known, but both types of translocations can be induced in germinal center B cells by AID, a specialized DNA-modifying enzyme required for both Ig class switching and somatic hypermutation (see earlier). The net effect of these translocations is similar; the *MYC* coding sequence is repositioned adjacent to strong Ig enhancer elements, which drive increased *MYC* expression. In addition, the translocated *MYC* allele often harbors point mutations that stabilize MYC protein and further increase its activity.

Sequencing of the genomes of Burkitt lymphoma cells has revealed that most tumors have mutations that increase the activity of the transcription factor TCF3 (also known as

E2A), an important regulator of gene expression in germinal center B cells. It is believed that TCF3 drives the expression of a set of genes, including cyclin D, that collaborate with MYC to enable the very rapid growth that characterizes Burkitt lymphoma.

Essentially all endemic Burkitt lymphomas are latently infected with EBV, which also is present in about 25% of HIV-associated tumors and 15% to 20% of sporadic cases. The configuration of the EBV DNA is identical in all tumor cells within individual cases, indicating that infection precedes transformation. Although this places EBV at the "scene of the crime," its precise role in the genesis of Burkitt lymphoma remains speculative (Chapter 7).

MORPHOLOGY

Involved tissues are effaced by a diffuse infiltrate of intermediate-sized lymphoid cells 10 to 25 μm in diameter with round or oval nuclei, coarse chromatin, several nucleoli, and a moderate amount of cytoplasm (Fig. 13.15). **The tumor exhibits a high mitotic index and contains numerous apoptotic cells,** the nuclear remnants of which are phagocytosed by interspersed benign macrophages. These phagocytes have abundant clear cytoplasm, creating a characteristic **"starry sky" pattern.** When the bone marrow is involved, aspirates reveal tumor cells with slightly clumped nuclear chromatin, two to five distinct nucleoli, and **royal blue cytoplasm containing clear cytoplasmic vacuoles.**

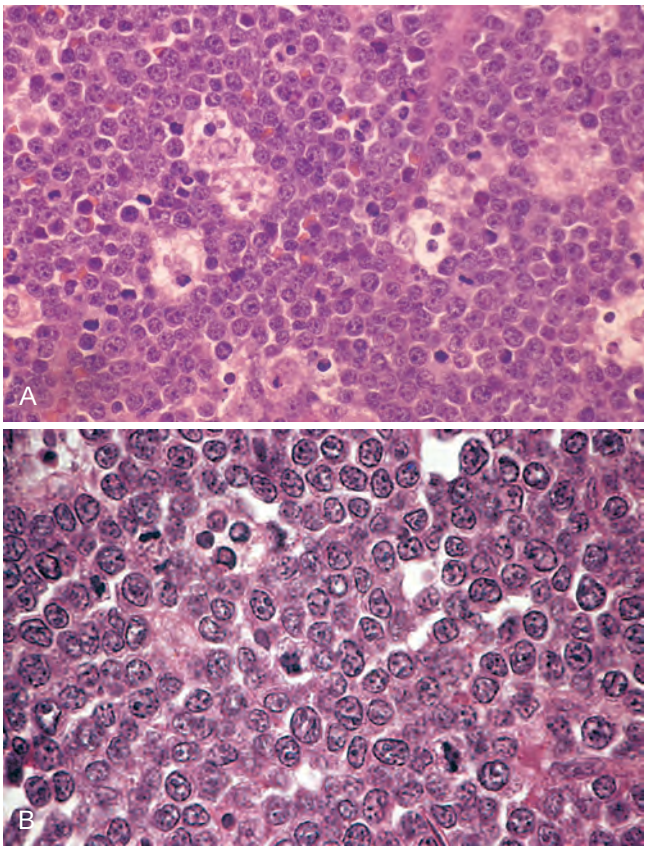


Figure 13.15 Burkitt lymphoma. (A) At low power, numerous pale tingible body macrophages are evident, producing a "starry sky" appearance. (B) At high power, tumor cells have multiple small nucleoli and high mitotic index. The lack of significant variation in nuclear shape and size lends a monotonous appearance. (B, Courtesy Dr. José Hernandez, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Immunophenotype. These are tumors of mature B cells that express surface IgM, CD19, CD20, CD10, and BCL6, a phenotype consistent with a germinal center B-cell origin. Unlike most other tumors of germinal center origin, Burkitt lymphoma almost always fails to express the antiapoptotic protein BCL2.

Clinical Features

Both endemic and sporadic Burkitt lymphomas are found mainly in children or young adults; overall, Burkitt lymphoma accounts for about 30% of childhood NHLs in the United States. Most tumors manifest at extranodal sites. Endemic Burkitt lymphoma often presents as a mass involving the mandible and shows an unusual predilection for involvement of abdominal viscera, particularly the kidneys, ovaries, and adrenal glands. In contrast, sporadic Burkitt lymphoma most often appears as a mass involving the ileocecum and peritoneum. Involvement of the bone marrow and peripheral blood is uncommon, especially in endemic cases.

Burkitt lymphoma is very aggressive but responds well to intensive chemotherapy. Most children and young adults can be cured. The outcome is more guarded in older adults.

Mantle Cell Lymphoma

Mantle cell lymphoma is an uncommon lymphoid neoplasm that makes up about 2.5% of NHL in the United States and 7% to 9% of NHL in Europe. It usually presents in the fifth to sixth decades of life and shows a male predominance. As the name implies, the tumor cells closely resemble the normal mantle zone B cells that surround germinal centers.

Pathogenesis

Virtually all mantle cell lymphomas have an (11;14) translocation involving the *IGH* locus on chromosome 14 and the cyclin D1 locus on chromosome 11 that leads to overexpression of cyclin D1. The resulting up-regulation of cyclin D1 promotes G_1- to S-phase progression during the cell cycle, as was described in Chapter 7.

MORPHOLOGY

At diagnosis the majority of patients have generalized lymphadenopathy, and 20% to 40% have peripheral blood involvement. Frequent sites of extranodal involvement include the bone marrow, spleen, liver, and gut. Occasionally, mucosal involvement of the small bowel or colon produces polyp-like lesions (lymphomatoid polyposis); of all forms of NHL, mantle cell lymphoma is most likely to spread in this fashion.

Nodal tumor cells may surround reactive germinal centers to produce a nodular appearance at low power or diffusely efface the node. **Typically, the proliferation consists of a homogeneous population of small lymphocytes with irregular**

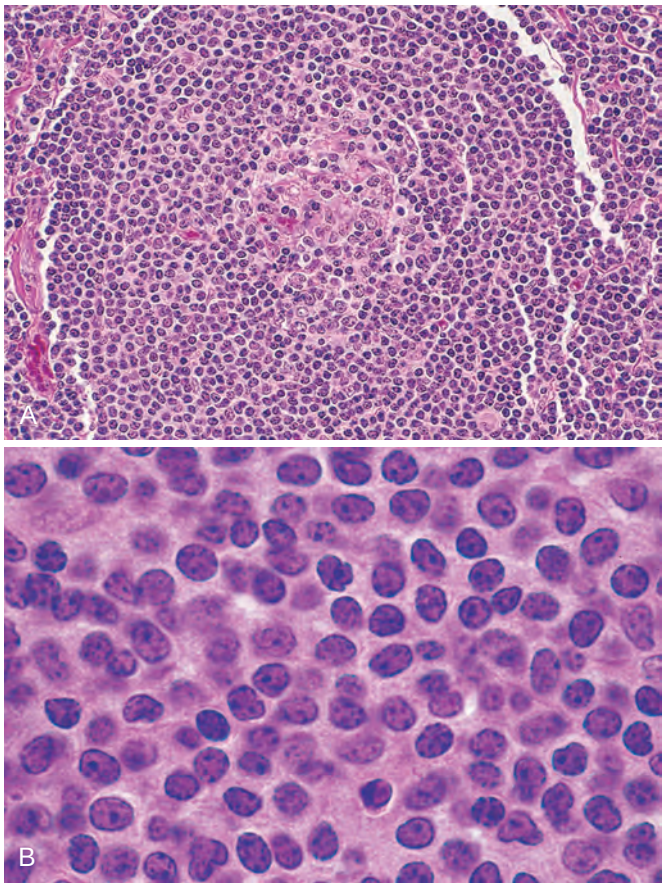


Figure 13.16 Mantle cell lymphoma. (A) At low power, neoplastic lymphoid cells surround a small, atrophic germinal center, producing a mantle zone pattern of growth. (B) High-power view shows a homogeneous population of small lymphoid cells with somewhat irregular nuclear outlines, condensed chromatin, and scant cytoplasm. Large cells resembling prolymphocytes (seen in chronic lymphocytic leukemia) and centroblasts (seen in follicular lymphoma) are absent.

to occasionally deeply clefted (cleaved) nuclear contours (Fig. 13.16). Large cells resembling centroblasts and proliferation centers are absent, distinguishing mantle cell lymphoma from follicular lymphoma and CLL/SLL, respectively. In most cases the nuclear chromatin is condensed, nucleoli are inconspicuous, and the cytoplasm is scant. Occasionally, tumors composed of intermediate-sized cells with more open chromatin and a brisk mitotic rate are observed; immunophenotyping is necessary to distinguish these "blastoid" variants from ALL.

Immunophenotype. Mantle cell lymphomas express high levels of cyclin D1, CD19, and CD20 and moderately high levels of surface Ig (usually IgM and IgD with κ or λ light chain). This tumor is usually CD5+ and CD23–, which help to distinguish it from CLL/SLL. The *IGH* genes lack somatic hypermutation, supporting an origin from a naive B cell.

Clinical Features

The most common presentation is painless lymphadenopathy. Symptoms related to involvement of the spleen (present in approximately 50% of cases) and gut also are common. Mantle cell lymphoma is moderately aggressive and incurable; the median survival is 8 to 10 years. The blastoid variant, a "proliferative" expression profiling signature, and *TP53* mutations are associated with shorter survivals.

Marginal Zone Lymphomas

The category of marginal zone lymphomas encompasses a heterogeneous group of B-cell tumors that arise within lymph nodes, spleen, or extranodal tissues. The extranodal tumors were initially recognized at mucosal sites and are often referred to as mucosa-associated lymphoid tumors (or MALTomas). In most cases, the tumor cells show evidence of somatic hypermutation and are considered to be of memory B-cell origin.

Although all marginal zone lymphomas share certain features, those occurring at extranodal sites deserve special attention because of their unusual pathogenesis and three exceptional characteristics.

- They often arise within tissues involved by chronic inflammatory disorders of autoimmune or infectious etiology; examples include the salivary gland in Sjögren disease, the thyroid gland in Hashimoto thyroiditis, and the stomach in *Helicobacter* gastritis.
- They remain localized for prolonged periods, spreading systemically only late in their course.
- They may regress if the inciting agent (e.g., *H. pylori*) is eradicated.

These characteristics suggest that **extranodal marginal zone lymphomas arising in chronically inflamed tissues lie on a continuum between reactive lymphoid hyperplasia and full-blown lymphoma.** The disease begins as a polyclonal immune reaction. With the acquisition of unknown initiating mutations, a B-cell clone emerges that still depends on antigen-stimulated T-helper cells for signals that drive growth and survival. At this stage, withdrawal of the responsible antigen causes tumor involution. A clinically relevant example is found in gastric MALToma, in which antibiotic therapy directed against *H. pylori* often leads to tumor regression (Chapter 17). With time, however, tumors may acquire additional mutations that render their growth and survival antigen-independent, such as (11;18), (14;18), or (1;14) chromosomal translocations, which are relatively specific for extranodal marginal zone lymphomas. All of these translocations up-regulate the expression and function of BCL10 or MALT1, protein components of a signaling complex that activates NF-κB and promotes the growth and survival of B cells. With further clonal evolution, spread to distant sites and transformation to DLBCL may occur. This theme of polyclonal-to-monoclonal transition during lymphomagenesis is also applicable to the pathogenesis of EBV-induced lymphoma and is discussed more fully in Chapter 7.

Hairy Cell Leukemia

This rare but distinctive B-cell neoplasm constitutes about 2% of all leukemias. It is predominantly a disease of middle-aged white males, with a median age of 55 and a male-to-female ratio of 5:1.

Pathogenesis

Hairy cell leukemia is associated in more than 90% of cases with activating point mutations in the serine/threonine

kinase BRAF, which lies immediately downstream of RAS in the MAPK signaling cascade (Chapter 7). The specific mutation, a valine to glutamate substitution at residue 600, is also found at high frequencies in many other neoplasms, including melanoma and Langerhans cell histiocytosis (discussed later).

MORPHOLOGY

Hairy cell leukemia derives its picturesque name from the appearance of the leukemic cells, which have **fine hair-like projections** that are best recognized under the phase-contrast microscope (Fig. 13.17). On routine peripheral blood smears, hairy cells have round, oblong, or reniform nuclei and moderate amounts of pale blue cytoplasm with thread-like or bleb-like extensions. The number of circulating cells is highly variable. The marrow is involved by a diffuse interstitial infiltrate of cells with oblong or reniform nuclei, condensed chromatin, and pale cytoplasm. Because these cells are enmeshed in an extracellular matrix composed of reticulin fibrils, they usually are inaspirable (a clinical difficulty referred to as a "dry tap") and are seen only in marrow biopsies. The splenic red pulp is usually heavily infiltrated, leading to obliteration of white pulp and a beefy red gross appearance. Hepatic portal triads are also involved frequently.

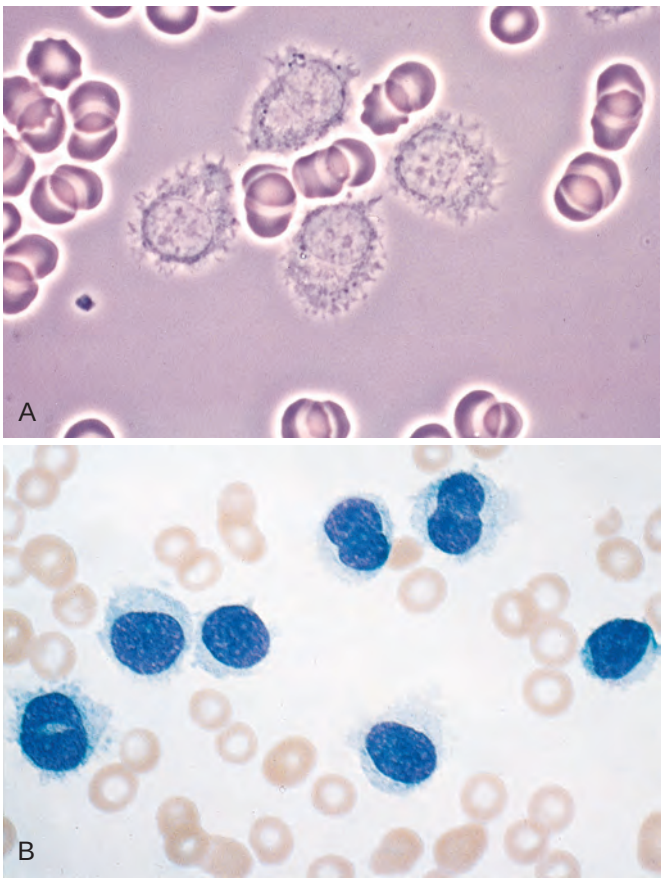


Figure 13.17 Hairy cell leukemia (peripheral blood smear). (A) Phase-contrast microscopy shows tumor cells with fine hairlike cytoplasmic projections. (B) In stained smears, these cells have round or folded nuclei and modest amounts of pale blue, agranular cytoplasm.

Immunophenotype. Hairy cell leukemias typically express the pan-B-cell markers CD19 and CD20; surface Ig (usually IgG); and certain relatively distinctive markers, such as CD11c, CD25, CD103, and annexin A1.

Clinical Features

Clinical manifestations result largely from infiltration of the bone marrow, liver, and spleen. Splenomegaly, often massive, is the most common and sometimes the only abnormal physical finding. Hepatomegaly is less common and not as marked; lymphadenopathy is rare. Pancytopenia resulting from marrow involvement and sequestration of cells in the enlarged spleen is seen in more than half the cases. About one-third of those affected present with infections. There is an increased incidence of atypical mycobacterial infections, possibly related to monocytopenia of uncertain origin, which is commonly seen in this disease.

Hairy cell leukemia follows an indolent course. For unclear reasons, it is exceptionally sensitive to "gentle" chemotherapeutic regimens, which produce long-lasting remissions. Tumors often relapse after 5 or more years, yet generally respond well when retreated with the same agents, a feature that is highly unusual among human cancers. BRAF inhibitors appear to produce excellent responses in tumors that have failed conventional chemotherapy. The overall prognosis is excellent.

KEY CONCEPTS

COMMON FORMS OF LYMPHOID LEUKEMIA AND LYMPHOMA

Small Lymphocytic Lymphoma/Chronic Lymphocytic Leukemia

- Most common leukemia of adults.
- Tumor of mature B cells that usually manifests with bone marrow and lymph node involvement.
- Indolent course, commonly associated with immune abnormalities, including an increased susceptibility to infection and autoimmune disorders.

Follicular Lymphoma

- Most common indolent lymphoma of adults.
- Tumor cells recapitulate the growth pattern of normal germinal center B cells; most cases are associated with a (14;18) translocation that results in overexpression of BCL2.

Diffuse Large B-Cell Lymphoma

- Most common lymphoma of adults.
- Heterogeneous group of mature B-cell tumors that shares a large cell morphology and aggressive clinical behavior.
- Rearrangements or mutations of *BCL6* gene are recognized associations; one-third carry a (14;18) translocation involving *BCL2* and may arise from follicular lymphomas; about 5% have translocations involving *MYC*.

Burkitt Lymphoma

- Very aggressive tumor of mature B cells that usually arises at extranodal sites.

- Strongly associated with translocations involving the *MYC* proto-oncogene.
- Tumor cells often are latently infected by EBV.

Mantle Cell Lymphoma

- Tumor of naive B cells that pursues a moderately aggressive course.
- Highly associated with translocations involving the cyclin D1 gene.

Marginal Zone Lymphoma

- Indolent tumors of antigen-primed B cells that arise at sites of chronic immune stimulation.
- Often remain localized for long periods of time.

Hairy Cell Leukemia

- Morphologically distinct, very indolent tumor of mature B cells that involves spleen and marrow.
- Highly associated with mutations in the BRAF serine/threonine kinase.

Peripheral T- and NK-Cell Neoplasms

These categories include a heterogeneous group of neoplasms having phenotypes resembling mature T cells or NK cells. Peripheral T-cell tumors make up about 5% to 10% of NHLs in the United States and Europe, while NK-cell tumors are rare. By contrast, for unknown reasons both T- and NK-cell tumors occur more frequently in the Far East. Only the most common diagnoses and those of particular pathogenetic interest will be discussed.

Peripheral T-Cell Lymphoma, Unspecified

Although the WHO classification includes a number of distinct peripheral T-cell neoplasms, many of these lymphomas are not easily categorized and are lumped into a "wastebasket" diagnosis, *peripheral T-cell lymphoma, unspecified*. As might be expected, no morphologic feature is pathognomonic, but certain findings are characteristic. These tumors efface lymph nodes diffusely and are typically composed of a pleomorphic mixture of variably sized malignant T cells (Fig. 13.18). There is often a prominent infiltrate of reactive cells, such as eosinophils and macrophages, probably attracted by tumor-derived cytokines. Brisk neoangiogenesis may also be seen.

By definition, all peripheral T-cell lymphomas are derived from mature T cells. They usually express CD2, CD3, CD5, and either αβ or γδ T-cell receptors. Some also express CD4 or CD8; such tumors are taken to be of helper or cytotoxic T-cell origin, respectively. However, many tumors have phenotypes that do not resemble any known normal T cell. In difficult cases where the differential diagnosis lies between lymphoma and a florid reactive process, DNA analysis is used to confirm the presence of clonal T-cell receptor gene rearrangements.

Most patients present with generalized lymphadenopathy, sometimes accompanied by eosinophilia, pruritus, fever, and weight loss. Although cures of peripheral T-cell lymphoma have been reported, these tumors have a significantly worse prognosis than comparably aggressive mature B-cell neoplasms (e.g., DLBCL).

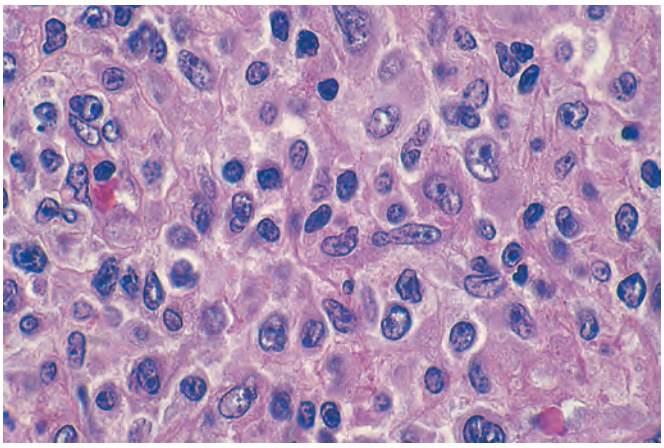


Figure 13.18 Peripheral T-cell lymphoma, unspecified (lymph node). A spectrum of small, intermediate, and large lymphoid cells, many with irregular nuclear contours, is visible.

Anaplastic Large-Cell Lymphoma (ALK Positive)

This uncommon entity is defined by the presence of rearrangements in the *ALK* gene on chromosome 2p23. These rearrangements break the *ALK* locus and lead to the formation of chimeric genes encoding ALK fusion proteins, constitutively active tyrosine kinases that trigger the RAS and JAK/STAT signaling pathways.

As the name implies, this tumor is typically composed of large anaplastic cells, some containing horseshoe-shaped nuclei and voluminous cytoplasm (so-called *hallmark cells*) (Fig. 13.19A). The tumor cells often cluster about venules and infiltrate lymphoid sinuses, mimicking the appearance of metastatic carcinoma. ALK is not expressed in normal lymphocytes; thus the detection of ALK protein in tumor cells (Fig. 13.19B) is a reliable indicator of an *ALK* gene rearrangement.

T-cell lymphomas with *ALK* rearrangements tend to occur in children or young adults, frequently involve soft tissues, and carry a very good prognosis (unlike other aggressive peripheral T-cell neoplasms). The cure rate with chemotherapy is 75% to 80%. Inhibitors of ALK have been developed and have produced excellent responses in some tumors that have failed to respond to conventional chemotherapy.

T-cell lymphomas that are morphologically similar but lack ALK rearrangements (termed ALK– anaplastic large cell lymphoma) tend to occur in older adults and have a substantially worse prognosis. Both ALK+ and ALK– tumors usually express CD30, a member of the TNF receptor family. Notably, recombinant antibodies linked to toxins that recognize CD30 have significant antitumor activity against CD30+ T-cell lymphomas and Hodgkin lymphoma, another CD30+ tumor (described later).

Adult T-Cell Leukemia/Lymphoma

This neoplasm of CD4+ T cells is observed only in adults infected by HTLV-1 (Chapter 7). It occurs mainly in regions where HTLV-1 is endemic, namely southern Japan, West Africa, and the Caribbean basin. Common findings include skin lesions, generalized lymphadenopathy, hepatosplenomegaly, peripheral blood lymphocytosis, and hypercalcemia.

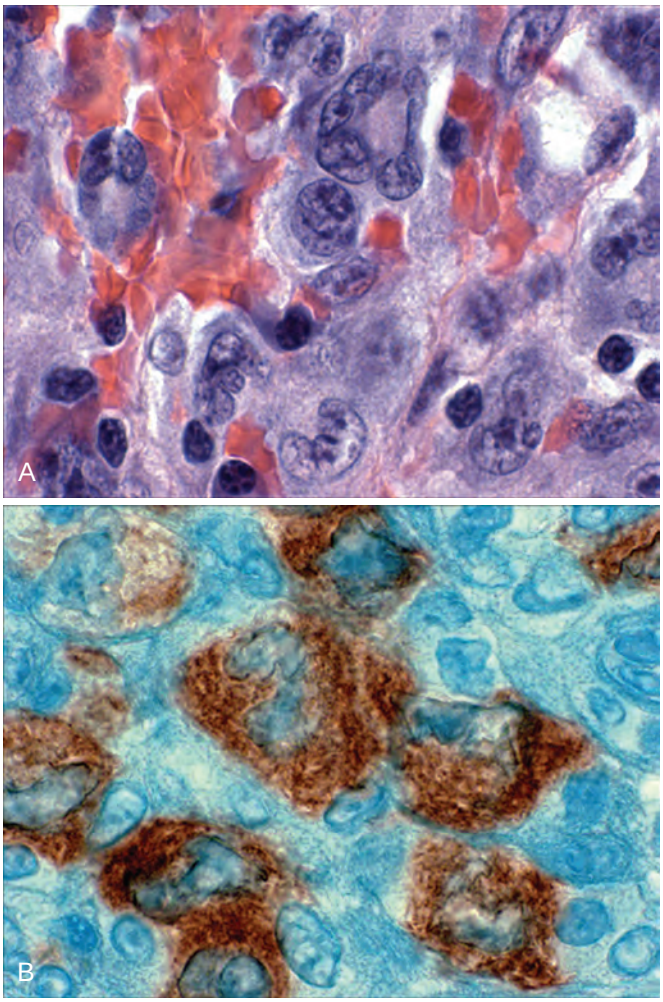


Figure 13.19 Anaplastic large-cell lymphoma. (A) Several hallmark cells with horseshoe-like or embryoid nuclei and abundant cytoplasm lie near the center of the field. (B) Immunohistochemical stain demonstrating the presence of ALK fusion protein. (Courtesy Dr. Jeffery Kutok, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

In addition to adult T-cell leukemia/lymphoma, HTLV-1 infection also may give rise to a progressive demyelinating disease of the central nervous system and spinal cord (Chapter 28).

The appearance of the tumor cells varies, but cells with multilobated nuclei ("cloverleaf" or "flower" cells) are frequently observed. The tumor cells always contain clonal HTLV-1 proviruses, suggesting a direct pathogenic role for the virus. However, while several proviral proteins have been implicated in the development of adult T-cell leukemia/lymphoma (Chapter 7), the tumor develops in only a small fraction of those who are infected, usually after a latent period of decades, and details of the transformation process remain poorly understood.

Most patients present with rapidly progressive disease that is fatal within months to 1 year despite aggressive chemotherapy. Less commonly, the tumor involves only the skin and follows a much more indolent course, like that of mycosis fungoides (described next).

Mycosis Fungoides/Sézary Syndrome

Mycosis fungoides and Sézary syndrome are different manifestations of a tumor of CD4+ helper T cells that home to the skin. Clinically, the cutaneous lesions of *mycosis fungoides* typically progress through three somewhat distinct stages, an inflammatory *premycotic phase,* a *plaque phase,* and a *tumor phase* (Chapter 25). Histologically, the epidermis and upper dermis are infiltrated by neoplastic T cells, which often have a cerebriform appearance due to marked infolding of the nuclear membrane. Late disease progression is characterized by extracutaneous spread, most commonly to lymph nodes and bone marrow.

Sézary syndrome is a variant in which skin involvement is manifested as a *generalized exfoliative erythroderma.* In contrast to mycosis fungoides, the skin lesions rarely proceed to tumefaction, and there is an associated leukemia of "Sézary" cells with characteristic cerebriform nuclei.

The tumor cells express the adhesion molecule cutaneous leukocyte antigen and the chemokine receptors CCR4 and CCR10, all of which contribute to the homing of normal CD4+ T cells to the skin. Although cutaneous disease dominates the clinical picture, sensitive molecular analyses have shown that the tumor cells circulate through the blood, marrow, and lymph nodes even early in the course. Nevertheless, these are indolent tumors, with a median survival of roughly 10 years. Transformation to aggressive T-cell lymphoma occurs occasionally as a terminal event.

Large Granular Lymphocytic Leukemia

T- and NK-cell variants of this rare neoplasm are recognized, both of which occur mainly in adults. Individuals with T-cell disease usually present with mild to moderate lymphocytosis and splenomegaly. Lymphadenopathy and hepatomegaly are usually absent. NK-cell disease often presents in an even more subtle fashion, with little or no lymphocytosis or splenomegaly.

Approximately 30% to 40% of large granular lymphocytic leukemias have acquired mutations in the transcription factor STAT3, which functions downstream of cytokine receptors. These mutations occur in both T- and NK-cell forms of the disease and result in cytokine-independent activation of STAT3, which appears to have a major role in the pathogenesis of these proliferations.

The tumor cells are large lymphocytes with abundant blue cytoplasm and a few coarse azurophilic granules, best seen in peripheral blood smears. The marrow usually contains a sparse interstitial lymphocytic infiltrate that is difficult to appreciate without immunohistochemical stains. Infiltrates also are usually present in the spleen and liver. T-cell variants express CD3 on their surfaces, whereas NK-cell variants lack surface CD3 and express NK markers such as CD56.

Despite the relative paucity of marrow infiltration, **neutropenia and anemia dominate the clinical picture.** Neutropenia may be accompanied by a striking decrease in late myeloid forms in the marrow. Rarely, *pure red cell aplasia* is seen. There is also an increased incidence of rheumatologic disorders. Some patients with *Felty syndrome,* a triad of rheumatoid arthritis, splenomegaly, and neutropenia, have this disorder as an underlying cause. The basis for these varied clinical abnormalities is unknown, but autoimmunity, provoked in some way by the tumor, seems likely.

The course is variable, being largely dependent on the severity of the cytopenias and their responsiveness to low-dose chemotherapy or steroids. In general, tumors of T-cell origin pursue an indolent course, whereas NK-cell tumors behave more aggressively.

Extranodal NK/T-Cell Lymphoma

This neoplasm is rare in the United States and Europe, but constitutes as many as 3% of NHLs in Asia. It most frequently presents as a destructive nasopharyngeal mass; less common sites of presentation include the testis and the skin. The tumor cell infiltrate typically surrounds and invades small vessels, leading to extensive *ischemic necrosis.* In touch preparations, *large azurophilic granules* are seen in the cytoplasm of the tumor cells that resemble those found in normal NK cells.

Extranodal NK/T-cell lymphoma is highly associated with Epstein-Barr virus (EBV). Within individual patients, all of the tumor cells contain identical EBV episomes, indicating that the tumor originates from a single EBV-infected cell. How EBV gains entry is uncertain, since the tumor cells do not express CD21, the surface protein that serves as the B-cell EBV receptor. Most tumors are CD3– and lack T-cell receptor rearrangements and express NK-cell markers, supporting an NK-cell origin. No consistent chromosome aberration has been described.

Most extranodal NK/T-cell lymphomas are highly aggressive neoplasms that respond well to radiation therapy but are resistant to chemotherapy. Thus, the prognosis is generally poor in patients with advanced disease. Recent work has shown, however, that like other virus-driven cancers, extranodal NK/T-cell lymphoma responds well to immune checkpoint inhibitors, albeit with all of the attendant risks (e.g., autoimmune disease) that are associated with these agents.

KEY CONCEPTS

PERIPHERAL T- AND NK-CELL NEOPLASMS

Peripheral NK/T-cell Lymphomas and Leukemias

- Anaplastic large cell lymphoma: Aggressive T cell tumor, associated in a subset with translocations that lead to constitutive activation of the ALK tyrosine kinase.
- Adult T cell leukemia/lymphoma: Aggressive tumor of CD4+ T cells that is uniformly associated with HTLV-1 infection.
- Large granular lymphocytic leukemia: Indolent tumor of cytotoxic T cells or NK cells that is associated with mutations in the transcription factor STAT3 and with autoimmune phenomena and cytopenias.
- Extranodal NK/T cell lymphoma: Aggressive tumor, usually derived from NK cells, that is strongly associated with EBV infection.

Plasma Cell Neoplasms and Related Disorders

These B-cell proliferations contain neoplastic plasma cells that virtually always secrete monoclonal Ig or Ig fragments, which serve as tumor markers and often have pathologic consequences. Collectively, the plasma cell neoplasms account for about 15% of deaths caused by lymphoid neoplasms. The most common and deadly of these neoplasms is multiple myeloma, of which there are about 15,000 new cases per year in the United States.

A monoclonal Ig identified in the blood is referred to as an *M component,* in reference to myeloma. Because complete M components have molecular weights of 160,000 or higher, they are restricted to the plasma and extracellular fluid and excluded from the urine in the absence of glomerular damage. However, **neoplastic plasma cells often synthesize excess light chains along with complete Igs.** Occasionally only light chains are produced, and rare tumors secrete only heavy chains. In most patients with plasma cell tumors, the level of free light chains is elevated and markedly skewed toward one light chain (e.g., κ) at the expense of the second (e.g., λ). Because free light chains are small in size, they are also excreted in the urine, where they are referred to as *Bence Jones proteins.*

Terms used to describe the abnormal Igs associated with plasma cell neoplasms include *monoclonal gammopathy* and *paraproteinemia.* These abnormal proteins are associated with the following clinicopathologic entities:

- *Multiple myeloma (plasma cell myeloma),* the most important plasma cell neoplasm, usually presents as tumor masses scattered throughout the skeletal system. *Solitary myeloma (plasmacytoma)* is an infrequent variant that presents as a single mass in bone or soft tissue. *Smoldering myeloma* refers to another uncommon variant defined by a lack of symptoms and a high plasma M component.
- *Waldenström macroglobulinemia* is a syndrome in which high levels of IgM lead to symptoms related to hyperviscosity of the blood. It occurs in older adults, most commonly in association with lymphoplasmacytic lymphoma (described later).
- *Heavy-chain disease* is a rare monoclonal gammopathy seen in association with a diverse group of disorders, including lymphoplasmacytic lymphoma and an unusual small bowel marginal zone lymphoma that occurs in malnourished populations (so-called *Mediterranean lymphoma*). The common feature is the synthesis and secretion of free heavy-chain fragments.
- *Primary or immunocyte-associated amyloidosis* results from a monoclonal proliferation of plasma cells secreting light chains (usually of λ isotype) that are deposited as amyloid. Some patients have overt multiple myeloma, but others have only a minor clonal population of plasma cells in the marrow.
- *Monoclonal gammopathy of undetermined significance (MGUS)* is applied to patients without signs or symptoms who have small to moderately large M components in their blood. MGUS is very common in older adults and may transform to multiple myeloma or other symptomatic plasma cell neoplasms.

With this background, we now turn to some of these specific entities; primary amyloidosis is discussed in Chapter 6.

Multiple Myeloma

Multiple myeloma is a plasma cell neoplasm commonly associated with lytic bone lesions, hypercalcemia, renal failure, and acquired immune abnormalities. Although bony disease dominates, it can spread late in its course to lymph nodes and extranodal sites. Multiple myeloma causes

1% of cancer deaths in Western countries. Its incidence is higher in men and people of African descent. It is chiefly a disease of older adults, with a peak age of incidence of 65 to 70 years.

Pathogenesis

Multiple myeloma is genetically heterogeneous and is associated with frequent rearrangements involving the *IGH* locus on chromosome 14q32 and various proto-oncogenes. Included among the rearranged genes are the cell cycle–regulatory genes cyclin D1 on chromosome 11q13 and cyclin D3 on chromosome 6p21. Deletions of chromosome 17p that involve the *TP53* tumor suppressor locus also occur and are associated with a poor outcome. Late-stage, highly aggressive forms of the disease such as plasma cell leukemia are associated with rearrangements involving *MYC*. Deep sequencing of myeloma genomes has identified frequent mutations involving components of the NF-κB pathway, which supports B-cell survival.

The proliferation and survival of myeloma cells are dependent on several cytokines, most notably IL-6, which is an important growth factor for plasma cells. It is produced by the tumor cells themselves and by resident marrow stromal cells. High serum levels of IL-6 are seen in patients with active disease and are associated with a poor prognosis. Myeloma cell growth and survival are also augmented by direct physical interactions with bone marrow stromal cells, a phenomenon that is an area of intense research interest.

Factors produced by neoplastic plasma cells mediate bone destruction, the major pathologic feature of multiple myeloma. One important factor appears to be myeloma-derived MIP1α (also known as CCL3), a chemokine that augments osteoclast formation through several different mechanisms. Other factors released from tumor cells, such as modulators of the Wnt pathway, are potent inhibitors of osteoblast function. The net effect is a marked increase in bone resorption, which leads to hypercalcemia and pathologic fractures.

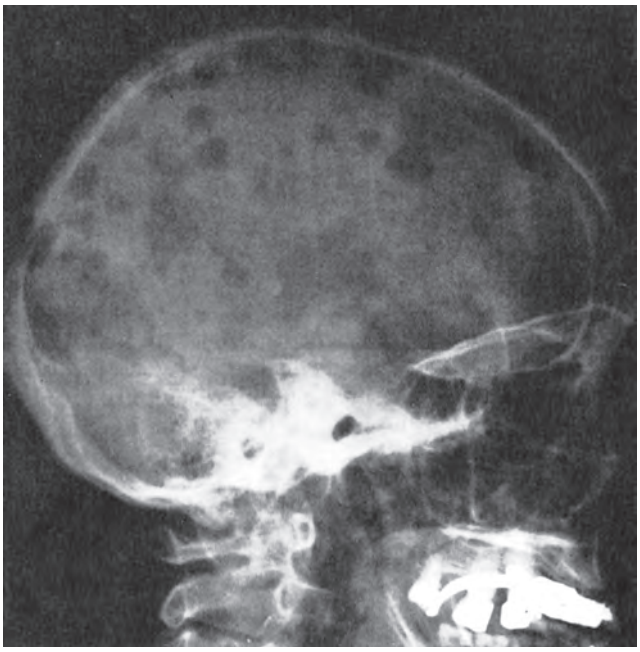


Figure 13.20 Multiple myeloma of the skull (radiograph, lateral view). The sharply punched-out bone lesions are most obvious in the calvaria.

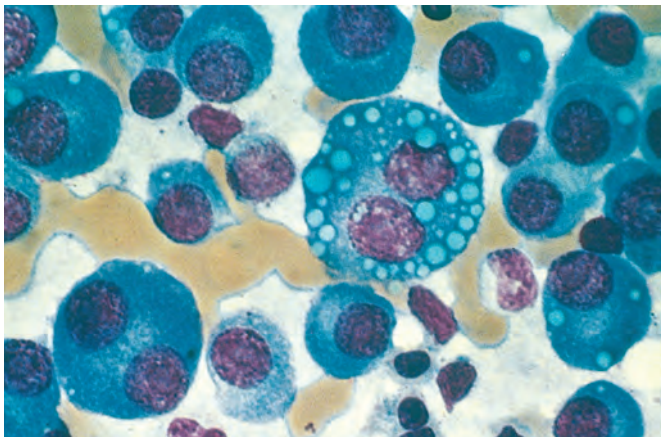


Figure 13.21 Multiple myeloma (bone marrow aspirate). Normal marrow cells are largely replaced by plasma cells including forms with multiple nuclei, prominent nucleoli, and cytoplasmic droplets containing immunoglobulin.

MORPHOLOGY

Multiple myeloma usually presents as destructive plasma cell tumors (plasmacytomas) involving the axial skeleton. The bones most commonly affected (in descending order of frequency) are the vertebral column, ribs, skull, pelvis, femur, clavicle, and scapula. Lesions begin in the medullary cavity, erode cancellous bone, and progressively destroy the bony cortex, often leading to pathologic fractures; these are most common in the vertebral column but may occur in any affected bone. **The bone lesions appear radiographically as punched-out defects, usually 1 to 4 cm in diameter** (Fig. 13.20), and consist of soft, gelatinous, red tumor masses. Less commonly, widespread myelomatous bone disease produces diffuse demineralization (osteopenia) rather than focal defects.

Even away from overt tumor masses, the marrow contains an increased number of plasma cells, which usually constitute more than 30% of the cellularity. The plasma cells may infiltrate the interstitium in a subtle fashion or completely replace normal elements. Like their benign counterparts, malignant plasma cells have a perinuclear clearing due to a prominent Golgi apparatus and an eccentrically placed nucleus (Fig. 13.21). Relatively normal-appearing plasma cells, **plasmablasts** with vesicular nuclear chromatin and a prominent single nucleolus, or **bizarre, multinucleated cells** may predominate. Other cytologic variants stem from the dysregulated synthesis and secretion of Ig, which often leads to intracellular accumulation of intact or partially degraded protein. Such variants include **flame cells** with fiery red cytoplasm, **Mott cells** with multiple grapelike cytoplasmic droplets, and cells containing a variety of other inclusions, including fibrils, crystalline rods, and globules. The globular inclusions are referred to as **Russell bodies** (if cytoplasmic) or **Dutcher bodies** (if nuclear). In advanced disease, plasma cell infiltrates may be present in the spleen, liver, kidneys, lungs, lymph nodes, and other soft tissues.

Commonly, high levels of M protein in the blood cause red cells to stick to one another in linear arrays in smears, a finding referred to as **rouleaux formation.** Rouleaux formation is characteristic but not specific, as it may be seen in other conditions in which Ig levels are elevated, such as lupus erythematosus and early HIV infection. Rarely, tumor cells flood the peripheral blood, giving rise to **plasma cell leukemia.**

Bence Jones proteins are excreted in the kidney and contribute to a form of renal disease called **myeloma kidney.** This important complication is discussed in detail in Chapter 20.

Immunophenotype. Plasma cell tumors are positive for CD138, an adhesion molecule also known as syndecan-1, and often express CD56, a feature that can be helpful in identifying small populations of neoplastic cells.

Clinical Features

The clinical features of multiple myeloma stem from (1) the effects of plasma cell growth in tissues, particularly the bones; (2) the production of excessive Igs, which often have abnormal physicochemical properties; and (3) the suppression of normal humoral immunity.

Bone resorption often leads to pathologic fractures and chronic pain. The attendant hypercalcemia can give rise to neurologic manifestations, such as confusion, weakness, lethargy, constipation, and polyuria, and contributes to renal dysfunction. Decreased production of normal Igs sets the stage for recurrent bacterial infections. Cellular immunity is relatively unaffected. Of great significance is renal insufficiency, which trails only infections as a cause of death. The pathogenesis of renal failure (Chapter 20), which occurs in up to 50% of patients, is multifactorial. However, the single most important factor seems to be Bence Jones proteinuria, as the excreted light chains are toxic to renal tubular epithelial cells. Certain light chains (particularly those of the λ6 and λ3 families) are prone to cause amyloidosis of the AL type (Chapter 6), which can exacerbate renal dysfunction and deposit in other tissues as well.

In 99% of patients, laboratory analyses reveal increased levels of Igs in the blood and/or light chains (Bence Jones proteins) in the urine. The monoclonal Igs are usually first detected as abnormal protein "spikes" in serum or urine electrophoresis and then further characterized by immunofixation (Fig. 13.22). Most myelomas are associated with more than 3 g/dL of serum Ig and/or more than 6 mg/dL of urine Bence Jones protein. The most common monoclonal Ig (M protein) is IgG (approximately 55% of patients), followed by IgA (approximately 25% of cases). Myelomas expressing IgM, IgD, or IgE occur but are rare. Excessive production and aggregation of M proteins, usually of the IgA or IgG_3 subtype, leads to symptoms related to hyperviscosity (described under Lymphoplasmacytic Lymphoma) in about 7% of patients. Both free light chains and a serum M protein are observed together in 60% to 70% of patients. However, in about 20% of patients only free light chains are present, and around 1% of myelomas are nonsecretory; hence the absence of M proteins does not completely exclude the diagnosis.

The clinicopathologic diagnosis of multiple myeloma relies on identification of clonal plasma cells in the marrow and the presence of CRAB criteria (hypercalcemia, renal dysfunction, anemia, and bone lesions). It can be strongly suspected when the distinctive radiographic changes are present, but definitive diagnosis requires a bone marrow examination and tests assessing calcium levels, renal function, blood counts, and serum and urine Igs.

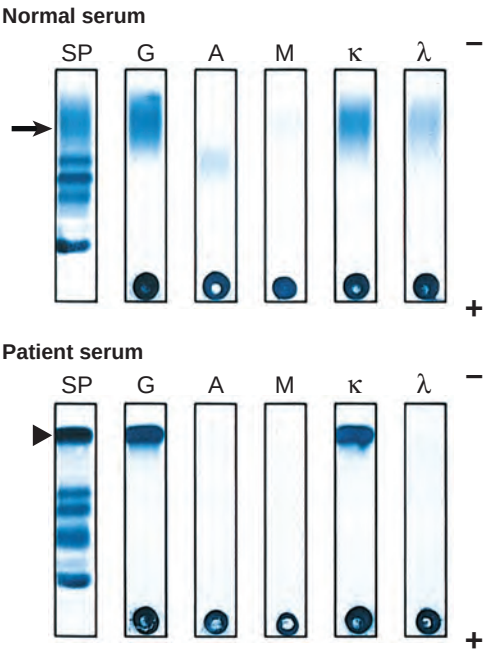


Figure 13.22 M protein detection in multiple myeloma. Serum protein electrophoresis *(SP)* is used to screen for a monoclonal immunoglobulin (Ig) (M protein). Polyclonal IgG in normal serum *(arrow)* appears as a broad band; in contrast, serum from a patient with multiple myeloma contains a single sharp protein band *(arrowhead)* in this region of the electropheretogram. The suspected monoclonal Ig is confirmed and characterized by immunofixation, in which proteins are trapped in the gel with antibodies specific for IgG *(G)*, IgA *(A)*, IgM *(M)*, or kappa *(κ)*, or lambda *(λ)* light chain and then visualized with a protein stain. Note the sharp band in the patient serum is cross-linked by antisera specific for IgG heavy chain and kappa light chain, indicating the presence of an IgGκ M protein. Levels of polyclonal IgG, IgA *(A)*, and lambda light chain *(λ)* are also decreased in the patient serum relative to normal, a finding typical of multiple myeloma. (Courtesy Dr. David Sacks, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

The prognosis is variable. The median survival is 4 to 7 years, and cures have yet to be achieved. Patients with multiple bony lesions, if untreated, rarely survive for more than 6 to 12 months, whereas patients with "smoldering myeloma" may be asymptomatic for many years. Translocations involving cyclin D1 are associated with a good outcome, whereas deletions of 13q, deletions of 17p, and the t(4;14) all portend a more aggressive course.

While still incurable, outcomes have improved over the past decade with advances in therapy. Myeloma cells are sensitive to inhibitors of the proteasome, a cellular organelle that degrades unwanted and misfolded proteins (Chapter 2). Because of the high levels of Ig synthesis, misfolding of Ig heavy and light chains may occur even in normal plasma cells, and this tendency may be exacerbated in myeloma cells in which heavy and light chain synthesis is unbalanced. If these misfolded proteins are not degraded in proteasomes, they trigger apoptosis. Proteasome inhibitors induce cell

death by exploiting this vulnerability and also seem to retard bone resorption through effects on stromal cells. Thalidomide and related compounds such as lenalidomide also have activity against myeloma. Interestingly, this also appears to involve changes in protein degradation, as lenalidomide redirects and activates certain ubiquitin ligases, thereby targeting proteins for proteolysis that are required for myeloma growth. Bisphosphonates, drugs that inhibit bone resorption, reduce pathologic fractures and limit hypercalcemia. HSC transplantation prolongs life but has not been proven to be curative.

Smoldering Myeloma

This entity defines a middle ground between multiple myeloma and MGUS. Plasma cells make up 10% to 30% of the marrow cellularity, and the serum M protein level is greater than 3 g/dL, but patients are asymptomatic. About 75% of patients progress to multiple myeloma over a 15-year period.

Solitary Osseous Plasmacytoma

About 3% to 5% of plasma cell neoplasms present as a solitary lesion of bone or soft tissue. The bone lesions tend to occur in the same locations as in multiple myeloma. Extraosseous lesions are often located in the lungs, oronasopharynx, or nasal sinuses. Modest elevations of M proteins in the blood or urine may be found in some patients. Solitary osseous plasmacytoma almost inevitably progresses to multiple myeloma, but this can take 10 to 20 years or longer. In contrast, extraosseous plasmacytomas, particularly those involving the upper respiratory tract, are frequently cured by local resection.

Monoclonal Gammopathy of Undetermined Significance

MGUS is the most common plasma cell disorder, occurring in about 3% of persons older than 50 years of age and in about 5% of individuals older than 70 years of age. By definition, patients are asymptomatic and the serum M protein level is less than 3 g/dL. **Approximately 1% of patients with MGUS develop a symptomatic plasma cell neoplasm, usually multiple myeloma, per year.** The clonal plasma cells in MGUS contain many of the same chromosomal translocations and deletions that are found in full-blown multiple myeloma, indicating that MGUS is an early stage of myeloma development. As with smoldering myeloma, progression to multiple myeloma is unpredictable; hence, periodic assessment of serum M component levels and Bence Jones proteinuria is warranted.

Lymphoplasmacytic Lymphoma

Lymphoplasmacytic lymphoma is a B-cell neoplasm of older adults that usually presents in the sixth or seventh decade of life. Although superficially resembling CLL/SLL, it differs in that a substantial fraction of the tumor cells undergo terminal differentiation to plasma cells. Most commonly, the plasma cell component secretes monoclonal IgM, often in amounts sufficient to cause a hyperviscosity syndrome known as *Waldenström macroglobulinemia*. Unlike multiple myeloma, complications stemming from the secretion of free light chains (e.g., renal failure and amyloidosis) are relatively rare, and bone destruction does not occur.

Pathogenesis

Virtually all cases of lymphoplasmacytic lymphoma are associated with acquired mutations in *MYD88*. *MYD88* encodes an adaptor protein that participates in signaling events that activate NF-κB, which may promote the growth and survival of the tumor cells.

MORPHOLOGY

Typically, the marrow contains an infiltrate of lymphocytes, plasma cells, and plasmacytoid lymphocytes in varying proportions, often accompanied by mast cell hyperplasia (Fig. 13.23). Some tumors also contain a population of larger lymphoid cells with more vesicular nuclear chromatin and prominent nucleoli. Periodic acid–Schiff–positive inclusions containing Ig are frequently seen in the cytoplasm **(Russell bodies)** or the nucleus **(Dutcher bodies)** of some of the plasma cells. At diagnosis the tumor has usually disseminated to the lymph nodes, spleen, and liver. Infiltration of the nerve roots, meninges, and more rarely the brain can also occur with disease progression.

Immunophenotype. The lymphoid component expresses B-cell markers such as CD20 and surface Ig, whereas the plasma cell component secretes the same Ig that is expressed on the surface of the lymphoid cells, usually IgM.

Clinical Features

The dominant presenting complaints are nonspecific and include weakness, fatigue, and weight loss. Approximately half the patients have lymphadenopathy, hepatomegaly, and splenomegaly. Anemia caused by marrow infiltration is common. About 10% of patients have autoimmune hemolysis caused by cold agglutinins, IgM antibodies that bind to red cells at temperatures of less than 37°C (Chapter 14).

Patients with IgM-secreting tumors have additional signs and symptoms stemming from the physicochemical

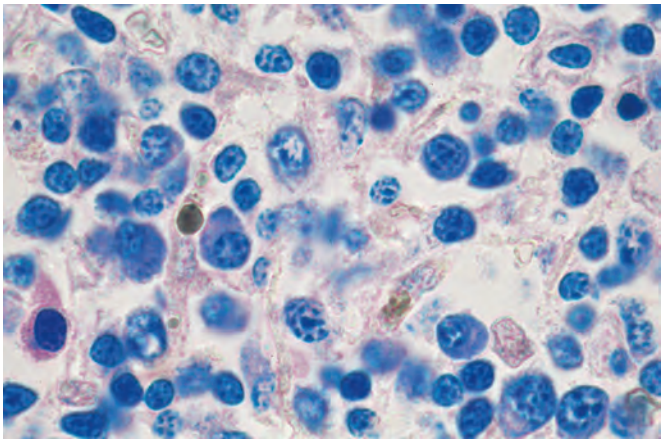


Figure 13.23 Lymphoplasmacytic lymphoma. Bone marrow biopsy shows a characteristic mixture of small lymphoid cells exhibiting various degrees of plasma cell differentiation. In addition, a mast cell with purplish red cytoplasmic granules is present at the left-hand side of the field.

properties of IgM. Because of its large size, at high concentrations IgM greatly increases the viscosity of the blood, giving rise to a *hyperviscosity syndrome* characterized by the following:

- *Visual impairment* associated with venous congestion, reflected by striking tortuosity and distention of retinal veins; retinal hemorrhages and exudates may also contribute to the visual problems.
- *Neurologic problems* such as headaches, dizziness, deafness, and stupor, all stemming from sluggish blood flow.
- *Bleeding* related to the formation of complexes between macroglobulins and clotting factors as well as interference with platelet function.
- *Cryoglobulinemia* resulting from the precipitation of macroglobulins at low temperatures, producing symptoms such as Raynaud phenomenon and cold urticaria.

Lymphoplasmacytic lymphoma is an indolent disorder. Because most IgM is intravascular, symptoms caused by the high IgM levels (e.g., hyperviscosity and hemolysis) can be alleviated by plasmapheresis. Tumor growth can be controlled with low doses of chemotherapeutic drugs and immunotherapy with anti-CD20 antibody. Like other indolent B-cell tumors, lymphoplasmacytic lymphoma is highly responsive to BTK inhibitors, indicating a dependence of the tumor cells on the B-cell receptor signaling pathway. Transformation to large-cell lymphoma occurs but is uncommon. Median survival is about 8 years and may improve further with the recent addition of BTK inhibitors to the therapeutic armamentarium.

KEY CONCEPTS

PLASMA CELL NEOPLASMS

Multiple Myeloma

- Plasma cell tumor that manifests with multiple lytic bone lesions associated with pathologic fractures and hypercalcemia.
- Neoplastic plasma cells suppress normal humoral immunity and secrete partial Igs that are nephrotoxic.
- Associated with diverse translocations involving the *IGH* locus; frequent dysregulation and overexpression of D cyclins.
- May be associated with AL amyloidosis (as may other neoplasms later).

Other Plasma Cell Neoplasms

- MGUS: common in older adults, progresses to myeloma at a rate of 1% per year.
- Smoldering myeloma: disseminated disease that pursues an unusually indolent course.
- Solitary osseous plasmacytoma: solitary bone lesion identical to disseminated myeloma; most progress to myeloma within 7 to 10 years.
- Extramedullary plasmacytoma: solitary mass, usually in the upper aerodigestive tract; rarely progresses to systemic disease.
- Lymphoplasmacytic lymphoma: B-cell lymphoma that exhibits plasmacytic differentiation; clinical symptoms dominated by hyperviscosity related to high levels of tumor-derived IgM; highly associated with mutations in the *MYD88* gene.

Table 13.7 Differences Between Hodgkin and Non-Hodgkin Lymphomas

Hodgkin Lymphoma	Non-Hodgkin Lymphoma
More often localized to a single axial group of nodes (cervical, mediastinal, para-aortic)	More frequent involvement of multiple peripheral nodes
Orderly spread by contiguity	Noncontiguous spread
Mesenteric nodes and Waldeyer ring rarely involved	Waldeyer ring and mesenteric nodes commonly involved
Extranodal presentation rare	Extranodal presentation common

Hodgkin Lymphoma

Hodgkin lymphoma encompasses a group of lymphoid neoplasms that differ from NHL in several respects (Table 13.7). While NHLs frequently occur at extranodal sites and spread in an unpredictable fashion, **Hodgkin lymphoma arises in a single node or chain of nodes and spreads first to anatomically contiguous lymphoid tissues. Morphologically, the distinctive feature of Hodgkin lymphoma is the presence of neoplastic giant cells called Reed-Sternberg cells.** These cells release factors that induce the accumulation of reactive lymphocytes, macrophages, and granulocytes, which typically make up greater than 90% of the tumor cellularity. Molecular studies have shown that the neoplastic Reed-Sternberg cells are derived from germinal center or post–germinal center B cells.

Hodgkin lymphoma accounts for 0.7% of all new cancers in the United States; there are about 8000 cases each year. The average age at diagnosis is 32 years. It is one of the most common cancers of young adults and adolescents, but also occurs in the aged. It was the first human cancer to be successfully treated with radiation therapy and chemotherapy and is curable in most cases.

Classification. The WHO classification recognizes five subtypes of Hodgkin lymphoma:

1. Nodular sclerosis
2. Mixed cellularity
3. Lymphocyte-rich
4. Lymphocyte depletion
5. Nodular lymphocyte predominance

In the first four subtypes—nodular sclerosis, mixed cellularity, lymphocyte-rich, and lymphocyte depletion—the Reed-Sternberg cells have a similar distinctive immunophenotype. These subtypes are often lumped together as *classic* forms of Hodgkin lymphoma. In the remaining subtype, lymphocyte predominance, the Reed-Sternberg cells have a B-cell immunophenotype that differs from that of the classic types.

Pathogenesis

The origin of the neoplastic Reed-Sternberg cells of classic Hodgkin lymphoma was solved through elegant molecular studies of single isolated Reed-Sternberg cells. These revealed clonal *IGH* gene rearrangements and the telltale signs of somatic hypermutation, establishing that Reed-Sternberg cells originate from a germinal center or post–germinal center B cell. Despite their B-cell origin, the Reed-Sternberg cells

of classic Hodgkin lymphoma fail to express most B-cell–specific genes, including the Ig genes. The cause of this wholesale reprogramming of gene expression has yet to be explained and presumably results from widespread epigenetic changes of uncertain etiology.

Activation of the transcription factor NF-κB is a common event in classic Hodgkin lymphoma and turns on genes that are believed to promote the growth and survival of Reed-Sternberg cells. This can occur by several mechanisms:

- EBV+ tumor cells express latent membrane protein-1 (LMP-1), a protein encoded by the EBV genome that transmits signals that up-regulate NF-κB.
- Activation of NF-κB may occur in EBV– tumors as a result of acquired loss-of-function mutations in IκB or TNF-α–induced protein 3, both of which are negative regulators of NF-κB.

It is hypothesized that activation of NF-κB rescues "crippled" germinal center B cells that cannot express Ig from apoptosis, setting the stage for the acquisition of other unknown mutations that collaborate to produce Reed-Sternberg cells. Little is known about the basis for the morphology of Reed-Sternberg cells and variants, but it is intriguing that EBV-infected B cells resembling Reed-Sternberg cells may be found in the lymph nodes of individuals with infectious mononucleosis, strongly suggesting that EBV-encoded proteins play a part in the remarkable metamorphosis of B cells into Reed-Sternberg cells.

Reed-Sternberg cells are aneuploid and possess diverse clonal chromosomal aberrations. Copy number gains in the *REL* proto-oncogene on chromosome 2p are particularly common and may also contribute to increases in NF-κB activity. Also frequent are copy number gains in genes encoding PD-L1 and PD-L2, located together on chromosome 9p, which you will recall are immune checkpoint proteins that inhibit antitumor T-cell responses (Chapter 7).

The florid accumulation of reactive cells in tissues involved by classic Hodgkin lymphoma occurs in response to a wide variety of cytokines (e.g., IL-5, IL-10, and M-CSF), chemokines (e.g., eotaxin), and other factors that are secreted by Reed-Sternberg cells. Once attracted, the reactive cells produce factors that support the growth and survival of the tumor cells and further modify the reactive cell response. For example, eosinophils and T cells express ligands that activate the CD30 and CD40 receptors found on Reed-Sternberg cells, producing signals that up-regulate NF-κB. Although Reed-Sternberg cells induce a host response, it is ineffective because of factors produced by the Reed-Sternberg cells. Most notably among these factors are PD-L1 and PD-L2, which antagonize cytotoxic T-cell responses. Other examples of "cross-talk" between Reed-Sternberg cells and surrounding reactive cells are provided in Fig. 13.28.

MORPHOLOGY

Identification of Reed-Sternberg cells and their variants is essential for the diagnosis. **Diagnostic Reed-Sternberg cells are large cells (45 μm in diameter) with multiple nuclei or a single nucleus with multiple nuclear lobes, each with a large inclusion-like nucleolus about the size of a small lymphocyte (5 to 7 μm in diameter)** (Fig. 13.24A). The cytoplasm is abundant. Several Reed-Sternberg cell variants are also recognized. **Mononuclear variants** contain a single nucleus with a large inclusion-like nucleolus (Fig. 13.24B). **Lacunar cells** (seen in the nodular sclerosis subtype) have more delicate, folded, or multilobate nuclei and abundant pale cytoplasm that is often disrupted during the cutting of sections, leaving the nucleus sitting in an empty space (a lacuna) (Fig. 13.24C). In classic forms of Hodgkin lymphoma, Reed-Sternberg cells undergo a peculiar form of cell death in which the cells shrink and become pyknotic, a process described as "mummification." **Lymphohistiocytic variants** (L&H cells) with polypoid nuclei, inconspicuous nucleoli, and moderately abundant cytoplasm are characteristic of the lymphocyte predominance subtype (Fig. 13.24D).

Hodgkin lymphoma must be distinguished from other conditions in which cells resembling Reed-Sternberg cells may be seen, such as infectious mononucleosis, solid tissue cancers, and large-cell NHLs. The diagnosis of Hodgkin lymphoma depends on the identification of Reed-Sternberg cells in a background of non-neoplastic inflammatory cells. Reed-Sternberg cells also have a characteristic immunohistochemical profile, a feature that is used to confirm the morphologic impression.

With this as background, we turn to the subclasses of Hodgkin lymphoma, pointing out some of the salient morphologic and immunophenotypic features of each (Table 13.8). The clinical manifestations common to all are presented later.

Nodular Sclerosis Type. This is the most common form of Hodgkin lymphoma, constituting 65% to 70% of cases. It is characterized by the presence of lacunar variant Reed-Sternberg cells and the **deposition of collagen in bands that divide involved lymph nodes into circumscribed nodules** (Fig. 13.25). The fibrosis may be scant or abundant. The Reed-Sternberg cells are found in a polymorphous background of T cells, eosinophils, plasma cells, and macrophages. Diagnostic Reed-Sternberg cells are often uncommon. The Reed-Sternberg cells in this and other classic Hodgkin lymphoma subtypes have a characteristic immunophenotype; they are positive for PAX5 (a B-cell transcription factor), CD15, and CD30 and negative for other B-cell markers, T-cell markers, and CD45 (leukocyte common antigen). As in other forms of Hodgkin lymphoma, involvement of the spleen, liver, bone marrow, and other organs and tissues may appear in due course in the form of irregular tumor nodules resembling those seen in lymph nodes. This subtype is uncommonly associated with EBV.

The nodular sclerosis type occurs with equal frequency in males and females. It has a propensity to involve the lower cervical, supraclavicular, and mediastinal lymph nodes of adolescents or young adults. The prognosis is excellent.

Mixed-Cellularity Type. This form of Hodgkin lymphoma constitutes about 20% to 25% of cases. Involved lymph nodes are diffusely effaced by a heterogeneous cellular infiltrate, which includes T cells, eosinophils, plasma cells, and benign macrophages admixed with Reed-Sternberg cells (Fig. 13.26). **Diagnostic Reed-Sternberg cells and mononuclear variants are usually plentiful. The Reed-Sternberg cells are infected with EBV in about 70% of cases.** The immunophenotype is identical to that observed in the nodular sclerosis type.

Mixed-cellularity Hodgkin lymphoma is more common in males. Compared with the lymphocyte predominance and nodular sclerosis subtypes, it is more likely to be associated with older age, systemic symptoms such as night sweats and weight loss, and advanced tumor stage. Nonetheless, the overall prognosis is very good.

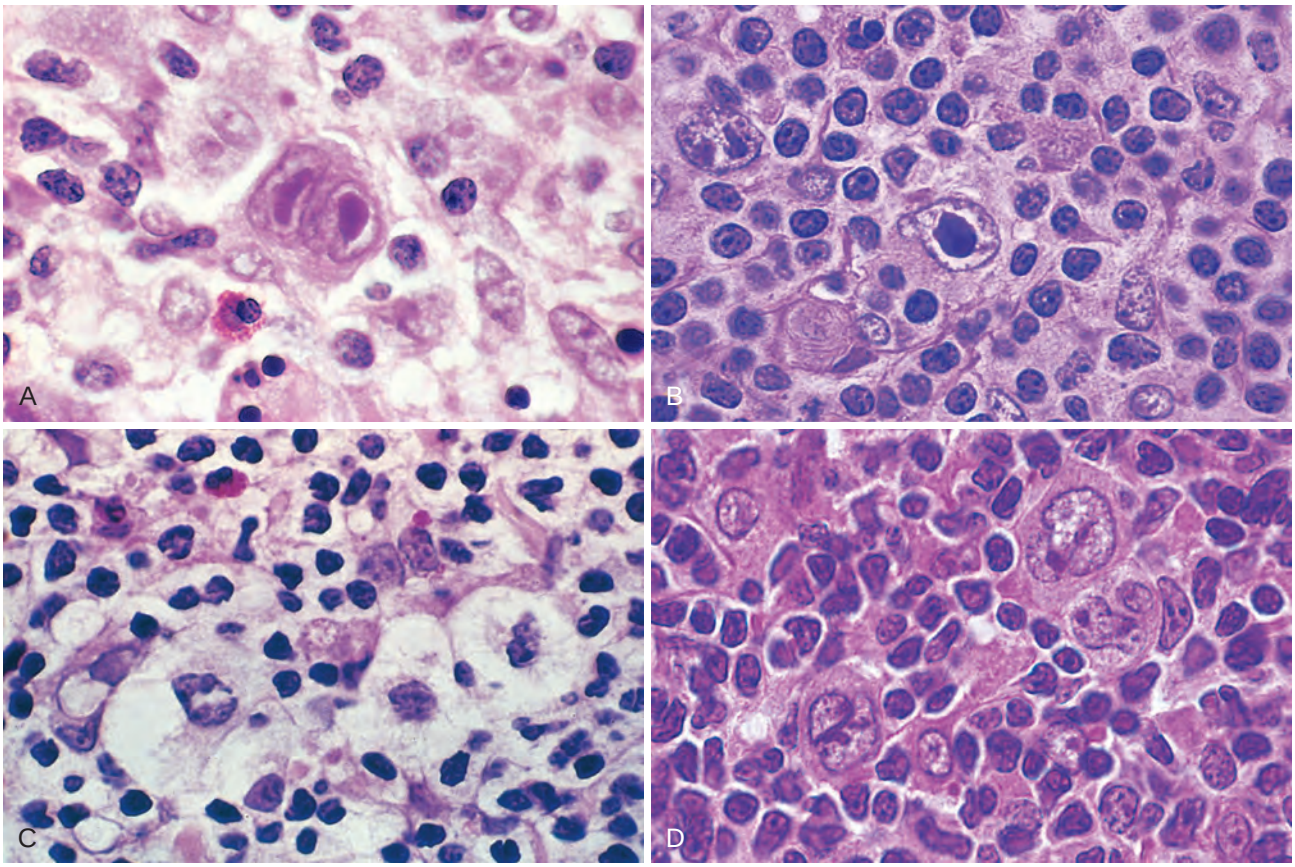


Figure 13.24 Reed-Sternberg cells and variants. (A) Diagnostic Reed-Sternberg cell, with two nuclear lobes, large inclusion-like nucleoli, and abundant cytoplasm, surrounded by lymphocytes, macrophages, and an eosinophil. (B) Reed-Sternberg cell, mononuclear variant. (C) Reed-Sternberg cell, lacunar variant. This variant has a folded or multilobated nucleus and lies within a open space, which is an artifact created by disruption of the cytoplasm during tissue sectioning. (D) Reed-Sternberg cell, lymphohistiocytic variant. Several such variants with multiply infolded nuclear membranes, small nucleoli, fine chromatin, and abundant pale cytoplasm are present. (A, Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Table 13.8 Subtypes of Hodgkin Lymphoma

Subtype	Morphology and Immunophenotype	Typical Clinical Features
Nodular sclerosis	Frequent lacunar cells and occasional diagnostic RS cells; background infiltrate composed of T lymphocytes, eosinophils, macrophages, and plasma cells; fibrous bands dividing cellular areas into nodules. RS cells CD15+, CD30+; usually EBV–	Most common subtype; usually stage I or II disease; frequent mediastinal involvement; equal occurrence in males and females, most patients young adults
Mixed cellularity	Frequent mononuclear and diagnostic RS cells; background infiltrate rich in T lymphocytes, eosinophils, macrophages, plasma cells; RS cells CD15+, CD30+; 70% EBV+	More than 50% present as stage III or IV disease; occurrence greater in males than females; biphasic incidence, peaking in young adults and again in adults older than 55
Lymphocyte-rich	Frequent mononuclear and diagnostic RS cells; background infiltrate rich in T lymphocytes; RS cells CD15+, CD30+; 40% EBV+	Uncommon; occurrence greater in males than females; tends to be seen in older adults
Lymphocyte depletion	Reticular variant: Frequent diagnostic RS cells and variants and a paucity of background reactive cells; RS cells CD15+, CD30+; most EBV+	Uncommon; more common in older men, HIV-infected individuals, and people in low income countries; often presents with advanced disease
Nodular lymphocyte predominant	Frequent L&H (popcorn cell) variants in a background of follicular dendritic cells and reactive B cells; RS cells CD20+, CD15–, CD30–; EB–	Uncommon; young males with cervical or axillary lymphadenopathy; mediastinal

HIV, human immunodeficiency virus; *L&H,* lymphohistiocytic; *RS,* Reed-Sternberg.

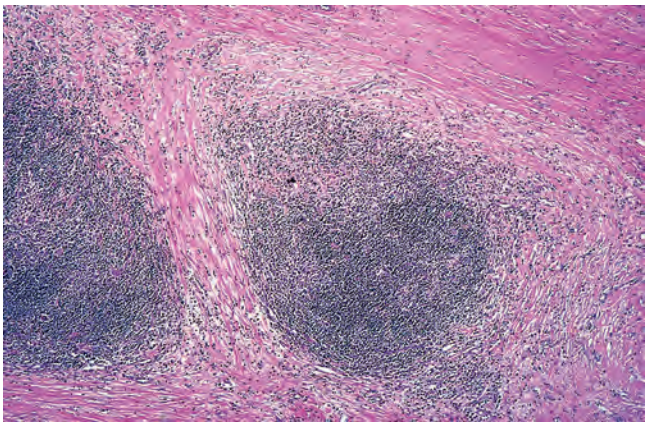


Figure 13.25 Hodgkin lymphoma, nodular sclerosis type. Low-power view shows well-defined bands of pink, acellular collagen that subdivide the tumor into nodules. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

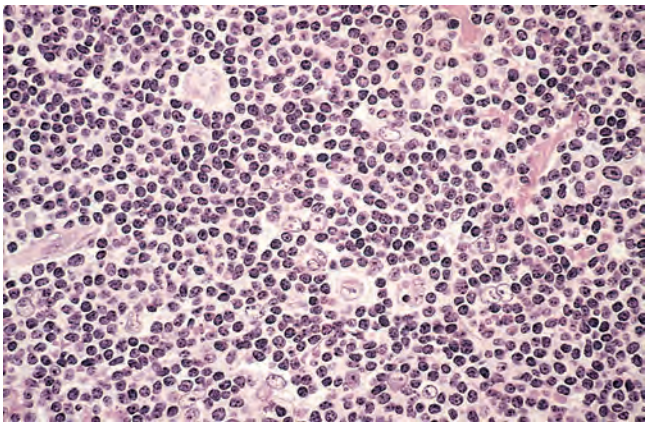


Figure 13.27 Hodgkin lymphoma, lymphocyte predominance type. Numerous mature-looking lymphocytes surround scattered, large, pale-staining lymphohistiocytic variants (popcorn cells). (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

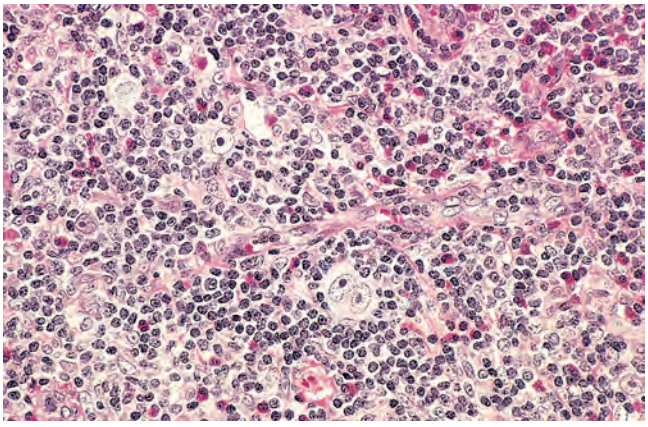


Figure 13.26 Hodgkin lymphoma, mixed-cellularity type. A diagnostic, binucleate Reed-Sternberg cell is surrounded by reactive cells including eosinophils (bright red cytoplasm), lymphocytes, and macrophages. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Lymphocyte-Rich Type. This is an uncommon form of classic Hodgkin lymphoma in which **reactive lymphocytes make up the vast majority of the cellular infiltrate.** In most cases, involved lymph nodes are diffusely effaced, but vague nodularity due to the presence of residual B-cell follicles is sometimes seen. This entity is distinguished from the lymphocyte predominance type by the presence of frequent mononuclear variants and diagnostic Reed-Sternberg cells with a "classic" immunophenotypic profile. It is associated with EBV in about 40% of cases and has a very good to excellent prognosis.

Lymphocyte Depletion Type. This is the least common form of Hodgkin lymphoma, amounting to less than 5% of cases. It is characterized by a paucity of lymphocytes and a relative abundance of Reed-Sternberg cells or their pleomorphic variants. The immunophenotype of the Reed-Sternberg cells is identical to that seen in other classic types of Hodgkin lymphoma. Immunophenotyping is essential, since most tumors suspected of being lymphocyte depletion Hodgkin lymphomas actually prove to be large-cell NHLs. **The Reed-Sternberg cells are infected with EBV in over 90% of cases.**

Lymphocyte depletion Hodgkin lymphoma occurs predominantly in older adults, in HIV-positive individuals of any age, and in individuals living in lower-income countries. Advanced stage and systemic symptoms are frequent, and the overall outcome is less favorable than in the other subtypes.

Nodular Lymphocyte Predominance Type. This uncommon "nonclassic" variant of Hodgkin lymphoma accounts for about 5% of cases. Involved nodes are effaced by nodules of small lymphocytes admixed with variable numbers of macrophages (Fig. 13.27). "Classic" Reed-Sternberg cells are usually difficult to find. Instead, this tumor contains so-called L&H variants with multilobed nuclei resembling popcorn kernels (popcorn cell). Eosinophils and plasma cells are usually scant or absent.

In contrast to the Reed-Sternberg cells found in classic forms of Hodgkin lymphoma, **L&H variants express B-cell markers typical of germinal center B cells,** such as CD20 and BCL6, and are usually negative for CD15 and CD30. The nodular pattern of growth is due to the presence of expanded B-cell follicles populated by L&H variants, numerous reactive B cells, and follicular dendritic cells. The *IGH* genes of the L&H variants show evidence of ongoing somatic hypermutation, further marking these cells as transformed germinal center B cells. In 3% to 5% of cases, this type transforms into a tumor resembling diffuse large B-cell lymphoma. EBV is rarely associated with this subtype.

A majority of patients are males, usually younger than 35 years of age, presenting with cervical or axillary lymphadenopathy. Mediastinal and bone marrow involvement is rare. In some series, this form of Hodgkin lymphoma is more likely to recur than the classic subtypes, but the prognosis is excellent.

Clinical Features

Hodgkin lymphoma most commonly presents as painless lymphadenopathy. Patients with the nodular sclerosis or lymphocyte predominance types tend to have stage I or II disease and are usually free of systemic manifestations. Patients with disseminated disease (stages III and IV) or the mixed-cellularity or lymphocyte depletion subtypes are

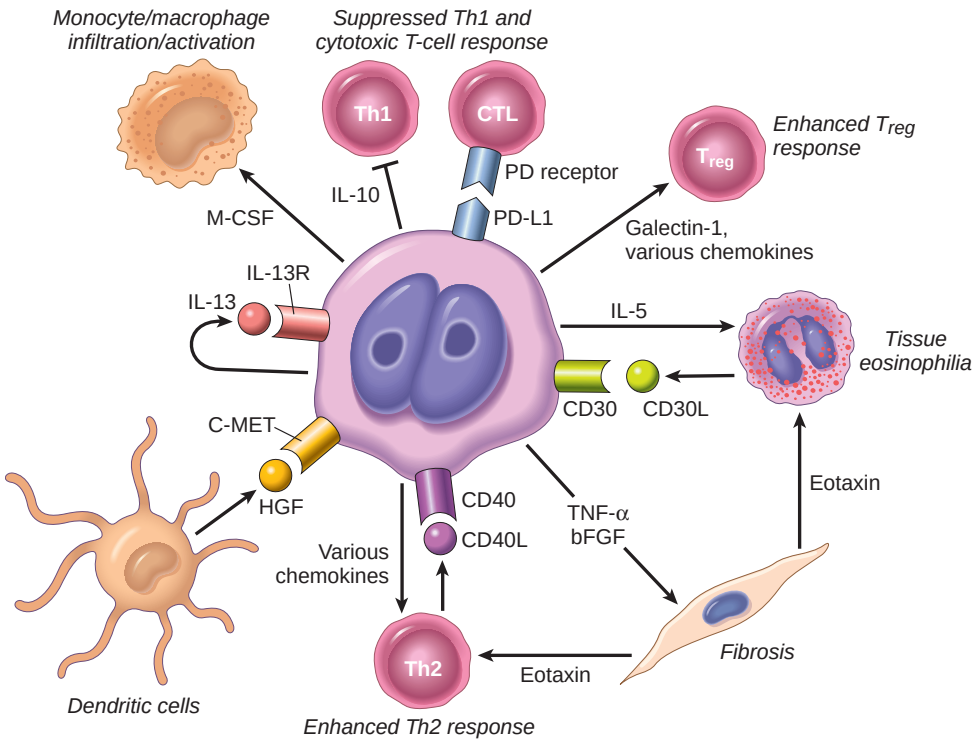


Figure 13.28 Proposed signals mediating crosstalk between Reed-Sternberg cells and surrounding normal cells in classic forms of Hodgkin lymphoma. *bFGF*, basic fibroblast growth factor; *CD30L*, CD30 ligand; *CTL*, CD8+ cytotoxic T cell; *HGF*, hepatocyte growth factor (binds to the c-MET receptor); *IL*, interleukin; *M-CSF*, monocyte colony-stimulating factor; *Th1*, *Th2*, CD4+ T helper cell subsets; *TNF-α*, tumor necrosis factor-α; T_{reg}, regulatory T cell.

more likely to have constitutional symptoms such as fever, night sweats, and weight loss. Cutaneous immune unresponsiveness (also called anergy) resulting from depressed cell-mediated immunity is seen in most cases of classic Hodgkin lymphoma and is attributable to the expression of factors such as IL-10 by Reed-Sternberg cells that suppress Th1 immune responses. This is but one of many examples of cross-talk between Reed-Sternberg cells, various stromal cells, and immune cells (Fig. 13.28).

The spread of Hodgkin lymphoma is remarkably stereotypic: nodal disease first, then splenic disease, hepatic disease, and finally involvement of the marrow and other tissues. Staging involves physical examination; radiologic imaging of the abdomen, pelvis, and chest; and biopsy of the bone marrow (Table 13.9). With current treatment protocols, tumor stage rather than histologic type is the most important prognostic variable. The cure rate of patients with stages I and IIA is close to 90%. Even with advanced disease (stages IVA and IVB), disease-free survival at 5 years is 60% to 70%.

Low-stage localized Hodgkin lymphoma can be cured with involved field radiotherapy, and indeed cure of such patients was one of the early success stories in oncology. However, it was subsequently recognized that long-term survivors treated with radiotherapy had a much higher incidence of certain other malignancies, including lung cancer, melanoma, and breast cancer. Patients treated with early chemotherapy regimens containing alkylating agents also had a high incidence of secondary tumors, particularly AML. These sobering results spurred the development of current treatment regimens, which minimize the use of radiotherapy and employ less genotoxic chemotherapeutic agents; as a result, the incidence of secondary tumors appears to have been reduced markedly, without any loss of therapeutic efficacy.

Table 13.9 Clinical Staging of Hodgkin and Non-Hodgkin Lymphomas (Ann Arbor Classification)

Stage	Distribution of Disease
I	Involvement of a single lymph node region (I) or a single extralymphatic organ or site (IE)
II	Involvement of two or more lymph node regions on the same side of the diaphragm alone (II) or localized involvement of an extralymphatic organ or site (IIE)
III	Involvement of lymph node regions on both sides of the diaphragm without (III) or with (IIIE) localized involvement of an extralymphatic organ or site
IV	Diffuse involvement of one or more extralymphatic organs or sites with or without lymphatic involvement
All stages are further divided on the basis of the absence (A) or presence (B) of the following symptoms: unexplained fever, drenching night sweats, and/or unexplained weight loss of greater than 10% of normal body weight.	

Data from Carbone PT, et al: Symposium (Ann Arbor): Staging in Hodgkin's disease, *Cancer Res* 31:1707, 1971.

For patients with classic Hodgkin lymphoma who fail conventional therapy, immune checkpoint inhibitors that block PD-1, the receptor for PD-L1 and PD-L2, have proven to be highly effective. These agents prevent the CD8+ cytotoxic T-cell "exhaustion" that is caused by PD-L1 and PD-L2 expressed on Reed-Sternberg cells (see Fig. 13.28) and lead to sustained responses in almost 90% of cases;

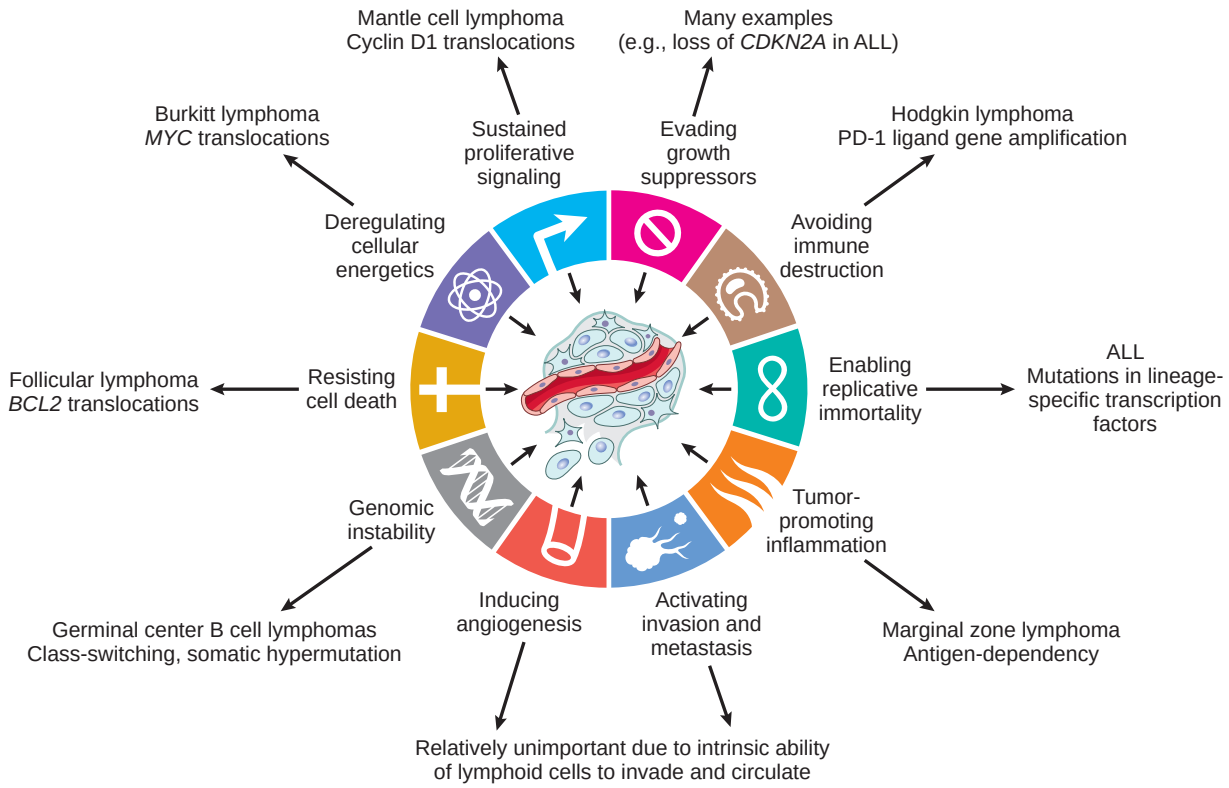


Figure 13.29 Cancer hallmarks exemplified by particular lymphoid neoplasms. See text for details.

indeed, classic Hodgkin lymphoma appears to be the human cancer that is most responsive to immune checkpoint inhibitor therapy.

KEY CONCEPTS

HODGKIN LYMPHOMA

- Unusual tumor consisting mostly of reactive lymphocytes, macrophages, eosinophils, plasma cells and stromal cells mixed with rare tumor giant cells called Reed-Sternberg cells and variants.
- Two broad types, classic (which has several subtypes) and nodular lymphocyte predominant, which are distinguished based on morphologic and immunophenotypic grounds.
- Reed-Sternberg cells of classic types express multiple cytokines, chemokines, and ligands that influence the host response, and the responding host cells in turn make factors that support the growth of the tumor cells.
- Classic forms are frequently associated with acquired mutations that activate the transcription factor NF-κB and with EBV infection.
- Lymphocyte predominance type expresses B-cell markers and is not associated with EBV.
- Classic Hodgkin lymphoma is highly responsive to immune checkpoint inhibitors, which antagonize the activity of PD-L1 and PD-L2 expressed on the surface of Reed-Sternberg cells.

Pathogenic Lessons From Lymphoid Neoplasms

Before ending our review of lymphoid neoplasms, it is worthwhile pausing to summarize the manner in which common driver mutations in specific entities produce changes in cellular behavior that exemplify particular hallmarks of cancer. Some of the best characterized pathogenic mechanisms are summarized in Fig. 13.29, including dysregulation of MYC in Burkitt lymphoma (leading to Warburg metabolisms and rapid cell growth); dysregulation of BCL2 in follicular lymphoma (leading to resistance to apoptosis); PD-1 ligand gene amplification in Hodgkin lymphoma (leading to evasion of host immunity); events leading to loss of cell cycle control (cyclin D1 rearrangements in mantle cell lymphoma and loss of the *CDKN2A* gene in acute lymphoblastic leukemia [ALL]); mutations in various transcription factors, particularly in ALL, that block differentiation and enhance "leukemia stem cell" self-renewal; and chronic immune stimulation, in marginal zone lymphoma. Such alterations not only highlight important pathogenic principles, but are increasingly the targets of effective therapies, such as antibodies that block PD-1 (Hodgkin lymphoma) and drugs that antagonize BCL2 (follicular lymphoma and other B-cell tumors). By contrast, because lymphoid cells normally circulate throughout the body, there is relatively little selective pressure in lymphoid malignancies for aberrations that increase angiogenesis or activate invasion and metastasis, points of distinction from other forms of cancer.

MYELOID NEOPLASMS

The common feature of this heterogeneous group of neoplasms is their origin from hematopoietic progenitor cells. These diseases primarily involve the marrow and to

a lesser degree the secondary hematopoietic organs (spleen, liver, and lymph nodes) and usually present with symptoms related to altered hematopoiesis. Three broad, major categories of myeloid neoplasms exist:

- *Acute myeloid leukemias,* in which an accumulation of immature myeloid forms (blasts) in the bone marrow suppresses normal hematopoiesis
- *Myelodysplastic syndromes,* in which defective maturation of myeloid progenitors gives rise to ineffective hematopoiesis, leading to cytopenias
- *Myeloproliferative neoplasms,* in which there is usually increased production of one or more types of blood cells

The pathogenesis of myeloid neoplasms is best understood in the context of normal hematopoiesis, which involves a hierarchy of HSCs, committed progenitors, and more differentiated elements (see Fig. 13.1). Normal hematopoiesis is finely tuned by homeostatic feedback mechanisms involving cytokines and growth factors that modulate the production of red cells, white cells, and platelets by the marrow. These mechanisms are deranged by myeloid neoplasms, which "escape" from normal homeostatic controls and suppress the function of residual normal HSCs and progenitors. The specific manifestations of the different myeloid neoplasms are influenced by

- *The position of the transformed cell within the hierarchy of progenitors* (i.e., a pluripotent HSC versus a more committed progenitor).
- *The effect of the transforming events on differentiation*, which may be inhibited, skewed, or deranged by particular oncogenic mutations.

Given that all myeloid neoplasms originate from transformed hematopoietic progenitors, it is not surprising that divisions between these neoplasms are sometimes blurred. Myeloid neoplasms, like other malignancies, tend to evolve over time to more aggressive forms of disease. In particular, both myelodysplastic syndromes and myeloproliferative neoplasms often "transform" to acute myeloid leukemia. In one of the most important myeloproliferative neoplasms, chronic myeloid leukemia, transformation to acute lymphoblastic leukemia is also seen, indicating that it originates from a transformed pluripotent HSC. As a final twist, it is now evident that many older adults with normal blood counts have clonal hematopoiesis, a state in which a substantial portion of the marrow output is the product of an expanded hematopoietic clone that bears one or more driver mutations associated with hematopoietic neoplasia, most commonly myeloid disease. These patients are not only at risk for hematopoietic neoplasms such as myelodysplastic syndromes, but also are at increased risk for cardiovascular disease (discussed in Chapter 12).

Acute Myeloid Leukemia

Acute myeloid leukemia (AML) is a tumor of hematopoietic progenitors caused by acquired oncogenic mutations that impede differentiation, leading to the accumulation of immature myeloid blasts in the marrow. The replacement of the marrow with blasts produces marrow failure and complications related to anemia, thrombocytopenia, and neutropenia. AML occurs at all ages, but the incidence rises throughout life, peaking after 60 years of age. There are about 13,000 new cases each year in the United States.

Classification. AML is quite heterogeneous, reflecting the complexities of myeloid cell differentiation. The current WHO classification subdivides AML into four categories (Table 13.10). The first includes forms of AML that are associated with particular genetic aberrations, which are important because they correlate with prognosis and guide therapy. Also included are categories of AML arising from myelodysplastic syndrome (MDS) or with MDS-like features and therapy-related AML. AMLs in these categories have distinct genetic features and respond poorly to therapy. A fourth "wastebasket" category includes AMLs lacking any of these features. These are classified based on the degree of differentiation and the lineage of the leukemic blasts. Given the increasing role of cytogenetic and molecular features in directing therapy, a further shift toward genetic classification of AML is both inevitable and desirable.

Pathogenesis

Driver mutations in AML tend to fall into four functional categories:

- *Transcription factor mutations that interfere with normal myeloid differentiation.* For example, the two most common chromosomal rearrangements, t(8;21) and inv(16), disrupt the *RUNX1* and *CBFB* genes, respectively. These two genes encode polypeptides that bind one another to form a RUNX1/CBFB transcription factor that is required for normal hematopoiesis. The t(8;21) and the inv(16) create chimeric genes encoding fusion proteins that interfere with the function of RUNX1/CBFB and block the maturation of myeloid cells. Another important example is found in acute promyelocytic leukemia, a distinctive subtype of AML associated with the t(15;17). The t(15;17) creates a fusion gene encoding a chimeric protein consisting of the retinoic acid receptor-α (RARα) and a portion of a protein called PML. As discussed in Chapter 7, this fusion protein interferes with the terminal differentiation of granulocytes, an effect that can be overcome by treatment with all-*trans*-retinoic acid and arsenic trioxide.
- *Mutation of signaling proteins that result in constitutive activation of pro-growth/survival pathways.* For example, AMLs with the t(15;17) also frequently have activating mutations in FLT3, a receptor tyrosine kinase that transmits signals that mimic normal growth factor signaling, thereby increasing cellular proliferation and survival. The combination of PML-RARα and activated FLT3 is a potent inducer of AML in mice, whereas expression of PML-RARα alone is only weakly leukemogenic. Mutations in a large number of other genes involved in pro-growth signaling, such as RAS, also occur in subsets of AML.
- *Mutation of genes that regulate or maintain the "epigenome."* Some of these mutations lead to abnormal DNA methylation patterns or involve members of the cohesin family, proteins that regulate the three-dimensional organization of chromatin in the nucleus. Another group of mutations involving the enzymes IDH1 or IDH2 result in the acquisition of a new enzymatic activity that produces the oncometabolite 2-hydroxyglutarate (described in Chapter 7). IDH inhibitors are effective in treating IDH-mutated forms of AML. The precise mechanism by which

Table 13.10 Major Subtypes of AML in the World Health Organization Classification

Class	Prognosis	Morphology/Comments
I. AML With Genetic Aberrations		
AML with t(8;21)(q22;q22); *RUNX1/RUNXT1* fusion gene	Favorable	Full range of myelocytic maturation; Auer rods easily found; abnormal cytoplasmic granules
AML with inv(16)(p13;q22); *CBFB/MYH11* fusion gene	Favorable	Myelocytic and monocytic differentiation; abnormal eosinophilic precursors with abnormal basophilic granules
AML with t(15;17)(q22;11-12); *RARA/PML* fusion gene	Very favorable	Numerous Auer rods, often in bundles within individual progranulocytes; primary granules usually very prominent, but inconspicuous in microgranular variant; high incidence of DIC
AML with t(11q23;v); diverse *KMT2A* fusion genes	Poor	Usually some degree of monocytic differentiation
AML with normal cytogenetics and mutated *NPM1*	Favorable	Detected by DNA sequencing
II. AML With MDS-Like Features		
With prior MDS	Poor	Diagnosis based on clinical history
AML with multilineage dysplasia	Poor	Maturing cells with dysplastic features typical of MDS
AML with MDS-like cytogenetic aberrations	Poor	Associated with 5q-, 7q-, 20q- aberrations
III. AML, Therapy-Related		
	Very poor	If following alkylator therapy or radiation therapy, 2- to 8-year latency period, MDS-like cytogenetic aberrations (e.g., 5q-, 7q-); if following topoisomerase II inhibitor (e.g., etoposide) therapy, 1- to 3-year latency, translocations involving *KMT2A* (11q23)
IV. AML, Not Otherwise Specified		
AML, minimally differentiated	Intermediate	Negative for myeloperoxidase; myeloid antigens detected on blasts by flow cytometry
AML without maturation	Intermediate	>3% of blasts positive for myeloperoxidase
AML with myelocytic maturation	Intermediate	Full range of myelocytic maturation
AML with myelomonocytic maturation	Intermediate	Myelocytic and monocytic differentiation
AML with monocytic maturation	Intermediate	Nonspecific esterase-positive monoblasts and promonocytes predominate in marrow; may see monoblasts or mature monocytes in the blood
AML with erythroid maturation	Intermediate	Erythroid/myeloid subtype defined by >50% dysplastic maturing erythroid precursors and >20% myeloblasts; pure erythroid subtype defined by >80% erythroid precursors without myeloblasts
AML with megakaryocytic maturation	Intermediate	Blasts of megakaryocytic lineage predominate; detected with antibodies against megakaryocyte-specific markers (GPIIb/IIIa or vWF); often associated with marrow fibrosis; most common AML in Down syndrome

AML, Acute myeloid leukemia; *DIC,* disseminated intravascular coagulation; *GPIIb/IIIa,* glycoprotein IIb-IIIa; *MDS,* myelodysplastic syndrome; *vWF,* von Willebrand factor.

disturbances of the epigenome contribute to the development of AML remains to be determined, but presumably they lead to alterations in gene expression that contribute to the acquisition of one or more cancer hallmarks.

- *Mutation of TP53 or genes that regulate p53.* With increasingly routine sequencing of AML genomes, it has become apparent that AMLs with mutations that impair p53 function have distinctive clinicopathologic features, including associations with complex karyotype, marked dysplasia, and particularly poor prognosis due to resistance to standard therapies.

MORPHOLOGY

The diagnosis of AML is based on the presence of at least 20% myeloid blasts in the bone marrow. Several types of myeloid blasts are recognized, and individual tumors may have more than one type of blast or blasts with hybrid features. **Myeloblasts** have delicate nuclear chromatin, two to four nucleoli, and more voluminous cytoplasm than lymphoblasts (Fig. 13.30A). The cytoplasm often contains fine, peroxidase-positive azurophilic granules. **Auer rods,** distinctive needle-like azurophilic granules, are present in many cases; they are particularly numerous in AML with the t(15;17) (acute promyelocytic leukemia) (Fig. 13.31A). **Monoblasts** (Fig. 13.31B) have folded or lobulated nuclei, lack Auer rods, and are nonspecific esterase-positive. In some AMLs, blasts show megakaryocytic differentiation, which is often accompanied by marrow fibrosis caused by the release of fibrogenic cytokines. Rarely, the blasts of AML show erythroid differentiation.

The number of leukemic cells in the blood is highly variable. Blasts may be more than 100,000/mm^3, but are under 10,000/mm^3 in about 50% of patients. **Occasionally, blasts are entirely absent from the blood (aleukemic leukemia).** For this reason, a bone marrow examination is essential to exclude acute leukemia in pancytopenic patients.

Immunophenotype. Because it can be difficult to distinguish myeloblasts and lymphoblasts morphologically, the diagnosis of AML is confirmed by performing stains for myeloid-specific antigens (Fig. 13.30B and C).

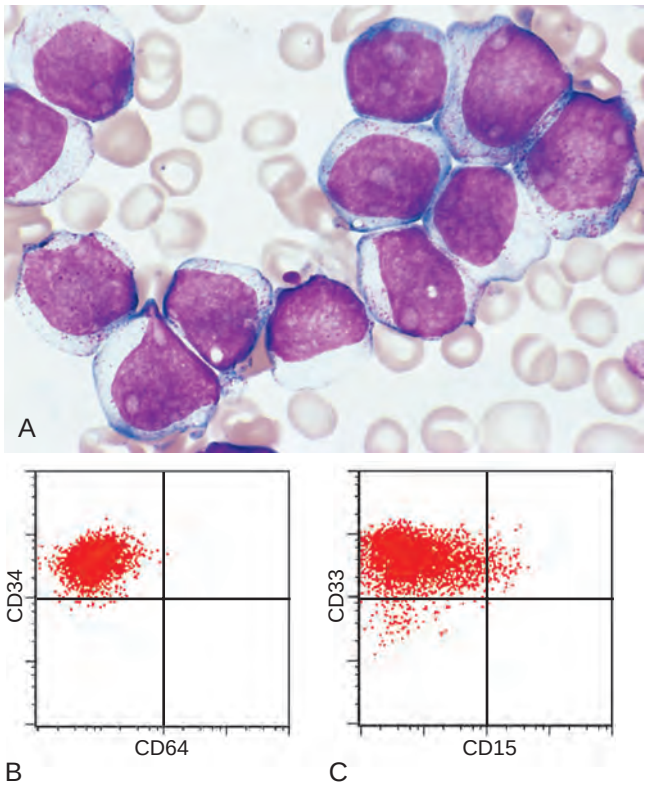


Figure 13.30 (A) Acute myeloid leukemia without maturation (FAB M1 subtype). Myeloblasts have delicate nuclear chromatin, prominent nucleoli, and fine azurophilic granules in the cytoplasm. (B) In the flow cytometric analysis shown, myeloid blasts, represented by the *red dots*, express CD34, a marker of multipotent stem cells, but do not express CD64, a marker of mature myeloid cells. (C) The same myeloid blasts express CD33, a marker of immature myeloid cells, and a subset express CD15, a marker of more mature myeloid cells. Thus these blasts are myeloid cells showing limited maturation. (A, Courtesy Dr. Robert W. McKenna Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.; B and C, courtesy Dr. Louis Picker, Oregon Health Science Center, Portland, Ore.)

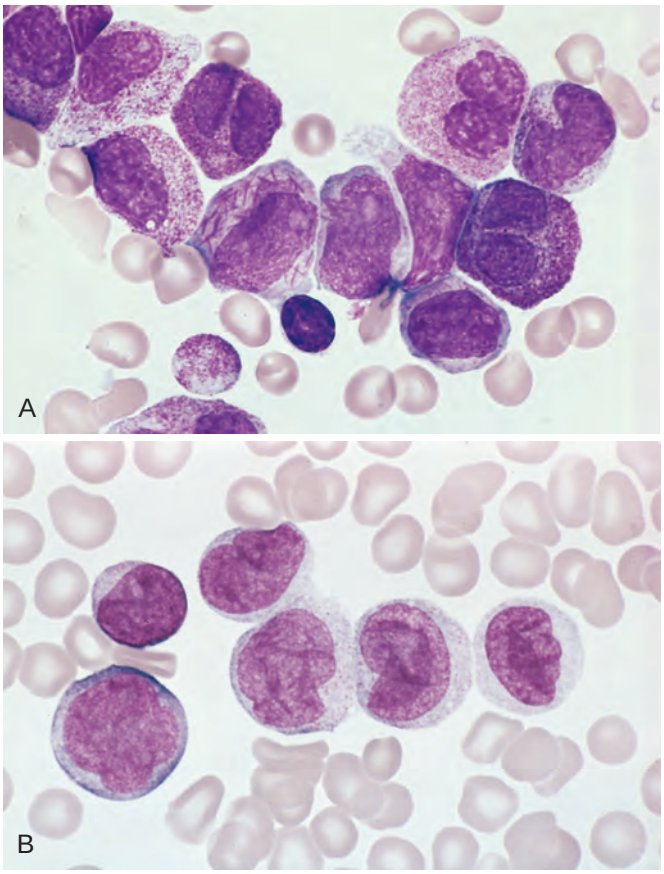


Figure 13.31 Acute myeloid leukemia subtypes. (A) Acute promyelocytic leukemia with the t(15;17) (FAB M3 subtype). Bone marrow aspirate shows neoplastic promyelocytes with abnormally coarse and numerous azurophilic granules. Other characteristic findings include the presence of several cells with bilobed nuclei and a cell in the center of the field that contains multiple needle-like Auer rods. (B) Acute myeloid leukemia with monocytic differentiation (FAB M5b subtype). Peripheral smear shows one monoblast and five promonocytes with folded nuclear membranes. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Cytogenetics. Cytogenetic analysis has a central role in the classification of AML. Karyotypic aberrations are detected in 50% to 70% of cases with standard techniques and in approximately 90% of cases using special high-resolution banding. Particular chromosomal abnormalities correlate with certain clinical features. AML arising de novo in younger adults is commonly associated with balanced chromosomal translocations, particularly t(8;21), inv(16), and t(15;17). In contrast, AML following MDS or exposure to DNA-damaging agents (such as chemotherapy or radiation therapy) often has deletions or monosomies involving chromosomes 5 and 7 and usually lack chromosomal translocations. The exception to this rule is AML occurring after treatment with topoisomerase II inhibitors, which is strongly associated with translocations involving the *KTM2A* gene on chromosome 11q23. AML in older adults also is more likely to be associated with "bad" aberrations, such as deletions of chromosomes 5q and 7q and complex karyotypic abnormalities, features that are associated with mutations that impair p53 function.

Clinical Features

Most patients present within weeks or a few months of the onset of symptoms with complaints related to anemia, neutropenia, and thrombocytopenia, most notably fatigue, fever, and spontaneous mucosal and cutaneous bleeding. You will remember that these findings are very similar to those produced by ALL. Thrombocytopenia results in abnormal bleeding, which often is prominent. Cutaneous petechiae and ecchymoses, serosal hemorrhages into the linings of the body cavities and viscera, and mucosal hemorrhages into the gingivae and urinary tract are common. Procoagulants and fibrinolytic factors released by leukemic cells, especially in AML with the t(15;17), exacerbate the bleeding tendency. Infections are frequent, particularly in the oral cavity, skin, lungs, kidneys, urinary bladder, and colon, and are often caused by opportunists such as fungi, *Pseudomonas,* and commensals.

Signs and symptoms related to involvement of tissues other than the marrow are usually less striking in AML than in ALL, but tumors with monocytic differentiation often infiltrate the skin (leukemia cutis) and the gingiva; this probably reflects the normal tendency of monocytes to extravasate into tissues. Central nervous system spread is less common than in ALL. AML occasionally presents as

a localized soft tissue mass known variously as a myeloblastoma, myeloid sarcoma, or chloroma. Without systemic treatment, such tumors inevitably progress to full-blown AML over time.

The overall prognosis is guarded, as AML is a difficult disease to treat; about 60% of patients achieve complete remission with chemotherapy, but only 15% to 30% remain free of disease for 5 years. However, the outcome varies markedly among different molecular subtypes. With targeted therapy using all-*trans*-retinoic acid and arsenic salts (described in Chapter 7), AMLs with the t(15;17) have the best prognosis of any type, being curable in more than 90% of patients. AMLs with t(8;21) or inv(16) also have a relatively good prognosis with conventional chemotherapy. By contrast, the prognosis is much worse for patients with AML who are over age 60 years, or who develop disease following MDS or genotoxic therapy, in large part because "good" genetic subtypes are rare in these settings and also because those that follow genotoxic therapy are often associated with *TP53* mutations. These "high-risk" forms of AML (as well as relapsed AML of all types) are treated with HSC transplantation when possible. Encourgingly, new therapies have emerged from insights gained from DNA sequencing; one success story is the deployment of inhibitors of mutated forms of IDH, which often produce excellent responses in patients with *IDH*-mutated AML.

Myelodysplastic Syndrome

The term "*myelodysplastic syndrome*" refers to a group of clonal stem cell disorders characterized by maturation defects that are associated with ineffective hematopoiesis and a high risk of transformation to AML. In MDS, the bone marrow is partly or wholly replaced by the clonal progeny of a neoplastic multipotent stem cell that retains the capacity to differentiate but does so in an ineffective and disordered fashion. Because the "birthing process" of blood cells is defective in MDS, patients have peripheral blood cytopenias of varying severity.

MDS may be either primary (idiopathic) or secondary to previous genotoxic drug or radiation therapy (t-MDS). t-MDS usually appears from 2 to 8 years after the genotoxic exposure. All forms of MDS can transform to AML, but transformation occurs with highest frequency and most rapidly in t-MDS. Although characteristic morphologic changes are typically seen in the marrow and the peripheral blood, the diagnosis frequently requires correlation with other laboratory tests. Cytogenetic analysis is often helpful, since certain chromosomal aberrations (discussed later) are often observed, and DNA sequencing also is now routinely used to help establish the diagnosis.

Pathogenesis

MDS is associated with driver mutations that partially overlap with those seen in AML, which is unsurprising given that MDS often evolves to AML. The affected proteins can be lumped into three major functional categories, as follows:

- *Epigenetic factors.* Frequent mutations are seen involving many of the same epigenetic factors that are mutated in AML, including factors that regulate DNA methylation, histone modifications, and chromatin looping; thus like AML, dysregulation of the epigenome appears to be important in the genesis of MDS.
- *RNA splicing factors.* A subset of tumors has mutations involving components of the 3′ end of the RNA splicing machinery. Precisely how these mutations contribute to the development of MDS has yet to be determined.
- *Transcription factors.* These mutations affect transcription factors that are required for normal myelopoiesis and may contribute to the deranged differentiation that characterizes MDS. Notably, classic chromosomal rearrangements that are seen in de novo AML, such as the t(8;21), inv(16), and t(15;17), are not seen in MDS; instead, most of the mutations are loss-of-function mutations in genes such as *RUNX1*.

In addition, roughly 10% of MDS cases have loss-of-function mutations in *TP53,* which as in AML correlate with the presence of a complex karyotype and particularly poor clinical outcomes. Both primary MDS and t-MDS are associated with similar recurrent chromosomal abnormalities, including monosomies 5 and 7; deletions of 5q, 7q, and 20q; and trisomy 8. As with aneuploidy in other cancers, how these aberrations contribute to MDS is incompletely understood. One idea is that the gain or loss of single copies of key genes is sufficient to give cells a growth advantage and that aneuploidy is one way to achieve this result. For example, subtle increases in the notorious oncoprotein MYC is sufficient to stimulate cell growth. Notably, the *MYC* gene is located on chromosome 8, and trisomy 8 is one of the most common forms of aneuploidy in a wide range of myeloid tumors. Similarly, the region that is commonly lost on chromosome 5q contains a gene encoding the ribosomal protein RPS14. In experimental systems, loss of one copy of RPS14 produces ineffective erythropoiesis, one of the hallmarks of MDS.

MDS appears to frequently arise from an asymptomatic state called clonal hematopoiesis of indeterminate potential (CHIP), defined by the presence of one or more pathogenic mutations associated with MDS in an individual with normal blood counts. Like CHIP (as discussed in Chapter 11), MDS patients show evidence of pro-inflammatory state and are at increased risk of dying from cardiovascular disease. Recent work suggests that MDS-associated inflammation may stem from activation of the inflammasome in neoplastic myeloid cells; precisely how this occurs remains to be determined.

MORPHOLOGY

Although the marrow is usually hypercellular at diagnosis, it is sometimes normocellular or, less commonly, hypocellular. The most characteristic finding is disordered (dysplastic) differentiation affecting the erythroid, granulocytic, monocytic, and megakaryocytic lineages to varying degrees (Fig. 13.32). Within the erythroid series, common abnormalities include **ring sideroblasts,** erythroblasts with iron-laden mitochondria visible as perinuclear granules in Prussian blue–stained aspirates or biopsies; **megaloblastoid maturation,** resembling that seen in vitamin B_{12} and folate deficiency (Chapter 14); and **nuclear budding abnormalities,** recognized as nuclei with misshapen, often polyploid, outlines. Neutrophils frequently contain decreased numbers of secondary granules, toxic granulations, and/or Döhle bodies. **Pseudo–Pelger-Hüet**

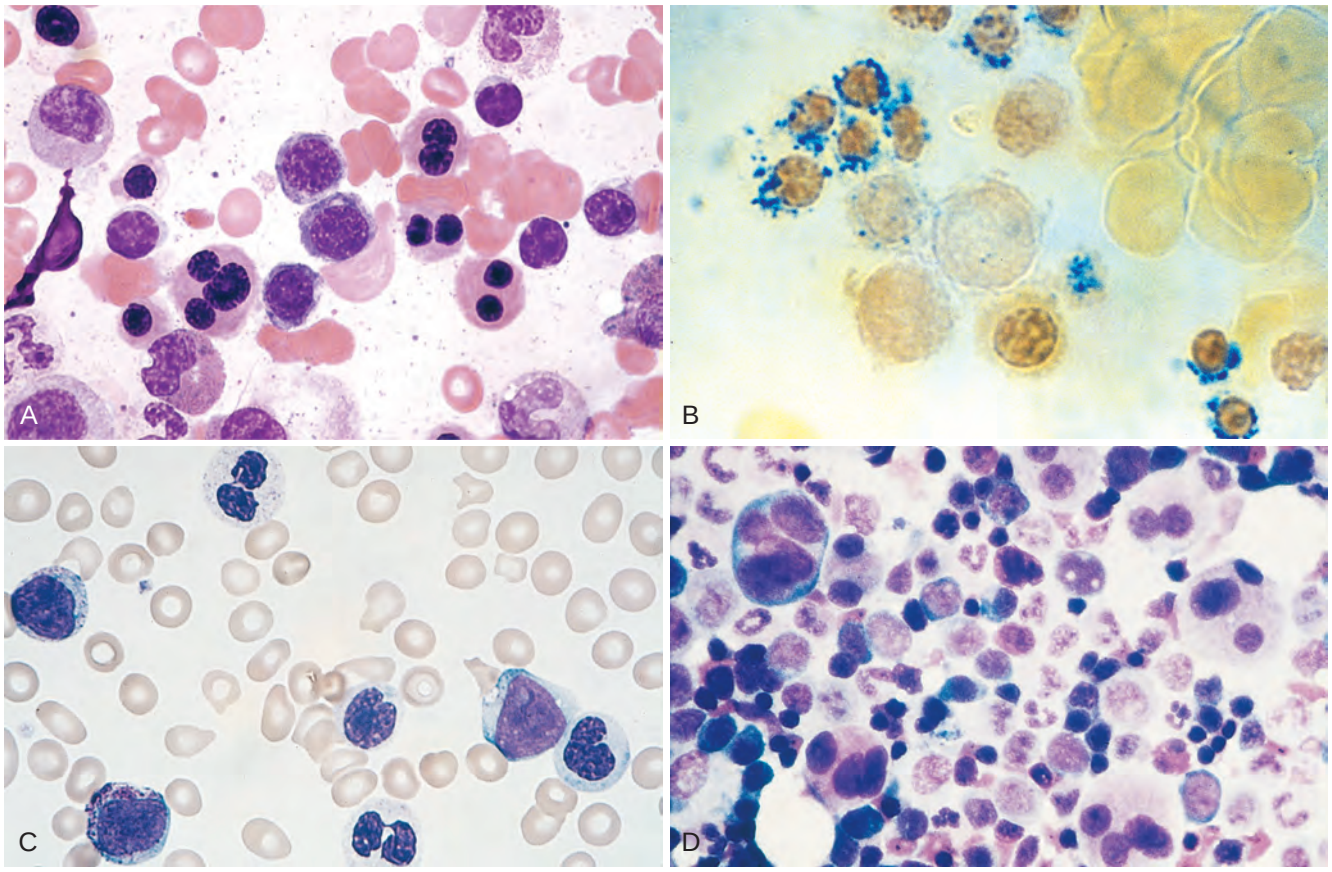


Figure 13.32 Myelodysplasia. Characteristic forms of dysplasia are shown. (A) Nucleated red cell progenitors with multilobated or multiple nuclei. (B) Ringed sideroblasts, erythroid progenitors with iron-laden mitochondria seen as blue perinuclear granules (Prussian blue stain). (C) Pseudo-Pelger-Hüet cells, neutrophils with only two nuclear lobes instead of the normal three to four, are observed at the top and bottom of this field. (D) Megakaryocytes with multiple nuclei instead of the normal single multilobated nucleus. (A, B, and D, Marrow aspirates; C, peripheral blood smear.)

cells, neutrophils with only two nuclear lobes, are commonly observed, and neutrophils are seen occasionally that completely lack nuclear segmentation. Megakaryocytes with single nuclear lobes or multiple separate nuclei **(pawn ball megakaryocytes)** are also characteristic. **Myeloid blasts** may be increased but make up less than 20% of the overall marrow cellularity. The blood often contains pseudo–Pelger-Hüet cells, giant platelets, macrocytes, and poikilocytes, accompanied by a relative or absolute monocytosis. Myeloid blasts usually make up less than 10% of the leukocytes in the blood.

Clinical Features

Primary MDS is predominantly a disease of older adults; the mean age of onset is 70 years. In up to half of the cases, it is discovered incidentally on routine blood testing. When symptomatic, it presents with weakness, infections, and hemorrhages, all due to pancytopenia.

Primary MDS is divided into six categories based on morphologic and cytogenetic features in the WHO classification, details of which are beyond our scope. Several prognostic scoring systems have been developed. In each, worse outcomes are (understandably) predicted by higher blast counts and more severe cytopenias, as well as the presence of multiple clonal chromosomal abnormalities.

Depending on the prognostic group, the median survival in primary MDS varies from less than 6 months to greater than 5 years. Progression to AML is more likely and occurs more rapidly in poor prognosis groups and is often accompanied by the appearance of additional cytogenetic abnormalities. Patients frequently succumb to the complications of thrombocytopenia (bleeding) and neutropenia (infection). The outlook is even grimmer in t-MDS, which has a median survival of only 4 to 8 months and often progresses to AML within 2 to 3 months of diagnosis.

Treatment options are fairly limited. In younger patients, allogeneic HSC transplantation offers hope for reconstitution of normal hematopoiesis and possible cure. Older patients with MDS are treated supportively with antibiotics and blood product transfusions. Thalidomide-like drugs and DNA methylation inhibitors improve the effectiveness of hematopoiesis and the peripheral blood counts in a subset of patients. The presence of isolated 5q deletion is correlated with a hematologic response to thalidomide-like drugs, but response to DNA methylation inhibitors remains unpredictable.

Myeloproliferative Neoplasms

The common pathogenic feature of myeloproliferative neoplasms is the presence of mutated, constitutively

activated tyrosine kinases or other acquired aberrations in signaling pathways that lead to growth factor independence. Hematopoietic growth factors act on normal progenitors by binding to surface receptors and activating tyrosine kinases, which turn on pathways that promote growth and survival (Chapter 7). The mutated tyrosine kinases found in myeloproliferative neoplasms circumvent normal controls and lead to growth factor–independent proliferation and survival of marrow progenitors. Because the tyrosine kinase mutations underlying the various myeloproliferative neoplasms do not impair differentiation, the most common consequence is an increase in the production of one or more mature blood elements. Most myeloproliferative neoplasms originate in multipotent myeloid progenitors, whereas others arise in pluripotent stem cells that give rise to both lymphoid and myeloid cells.

There is a considerable degree of clinical and morphologic overlap among myeloproliferative neoplasms. The common features include:

- *Increased proliferative drive* in the bone marrow
- Homing of the neoplastic stem cells to secondary hematopoietic organs, producing *extramedullary hematopoiesis*
- Variable transformation to a spent phase characterized by *marrow fibrosis* and peripheral blood *cytopenias*
- Variable transformation to *acute leukemia*

Certain myeloproliferative neoplasms are strongly associated with activating mutations in specific tyrosine kinases. This insight and the availability of kinase inhibitors have increased the importance of molecular tests for tyrosine kinase mutations, both for purposes of diagnosis and the selection of therapy. This discussion is confined to the more common myeloproliferative neoplasms, which are classified based on clinical, laboratory, and molecular criteria. Systemic mastocytosis, a distinctive neoplasm that is associated with mutations in the KIT tyrosine kinase, is discussed under disorders of the skin (Chapter 25). The association of various myeloproliferative neoplasms with specific tyrosine kinase mutations (including rare disorders not discussed here) is summarized in Table 13.11.

Chronic Myeloid Leukemia

CML is distinguished from other myeloproliferative neoplasms by the presence of a chimeric *BCR-ABL* gene derived from portions of the *BCR* gene on chromosome 22 and the *ABL* gene on chromosome 9. *BCR-ABL* directs the synthesis of a constitutively active BCR-ABL tyrosine kinase, which in CML is usually 210 kDa in size. In more than 90% of cases, *BCR-ABL* is created by a reciprocal (9;22)(q34;q11) translocation (the so-called *Philadelphia chromosome* [Ph]). In the remaining cases the *BCR-ABL* fusion gene is formed by cytogenetically complex or cryptic rearrangements and must be detected by other methods, such as fluorescence in situ hybridization or polymerase chain reaction (PCR)–based tests. The cell of origin is a pluripotent HSC.

Pathogenesis

Tyrosine kinases are normally regulated by ligand-mediated dimerization and autophosphorylation, which creates an activated kinase capable of phosphorylating other protein substrates (Chapters 3 and 7). The BCR moiety of BCR-ABL contains a dimerization domain that self-associates, leading to the activation of the ABL tyrosine kinase moiety (Fig. 13.33). The ABL kinase in turn phosphorylates proteins that induce signaling through the same pro-growth and pro-survival pathways that are turned on by hematopoietic growth factors, including the RAS and JAK/STAT pathways. For unknown reasons, BCR-ABL preferentially drives the proliferation of granulocytic and megakaryocytic progenitors and also causes the abnormal release of immature granulocytic forms from the marrow into the blood.

MORPHOLOGY

The marrow is markedly **hypercellular** because of massively increased numbers of maturing granulocytic precursors, which usually include an elevated proportion of eosinophils and basophils. Megakaryocytes are also increased and usually include small, dysplastic forms. Erythroid progenitors are present in normal or

Table 13.11 Mutations That Activate Tyrosine Kinase Signaling in Myeloproliferative Neoplasms

Disorder	Mutation	Frequency[a]	Consequences
Chronic myeloid leukemia	*BCR-ABL* fusion gene	100%	Constitutive ABL kinase activation
Polycythemia vera	*JAK2* mutations	>95%	Constitutive JAK2 kinase activation
Essential thrombocythemia	*JAK2* mutations *CALR* mutations *MPL* mutations	50%–60% 25%–35% 5%–10%	Constitutive JAK2 kinase activation Alternative MPL ligand Constitutive MPL kinase activation
Primary myelofibrosis	*JAK2* mutations *CALR* mutations *MPL* mutations	50%–60% 25%–35% 5%–10%	Constitutive JAK2 kinase activation Alternative MPL ligand Constitutive MPL kinase activation
Systemic mastocytosis	*KIT* mutations	>90%	Constitutive KIT kinase activation
Chronic eosinophilic leukemia	*FIP1L1-PDGFRA* fusion gene *PDE4DIP-PDGFRB* fusion gene	Common Rare	Constitutive PDGFRα kinase activation Constitutive PDGFRβ kinase activation
Myeloid/lymphoid neoplasms with eosinophilia and tyrosine kinase gene mutations[b]	Various fusion genes involving *FGFR1*, *PDGFRA*, *PDGFRB*, or *JAK2*	100%	Constitutive tyrosine kinase activation

[a]Refers to frequency within a diagnostic category.

[b]Rare disorder originating in pluripotent hematopoietic stem cells that presents with concomitant myeloproliferative neoplasm and lymphoblastic leukemia/lymphoma.

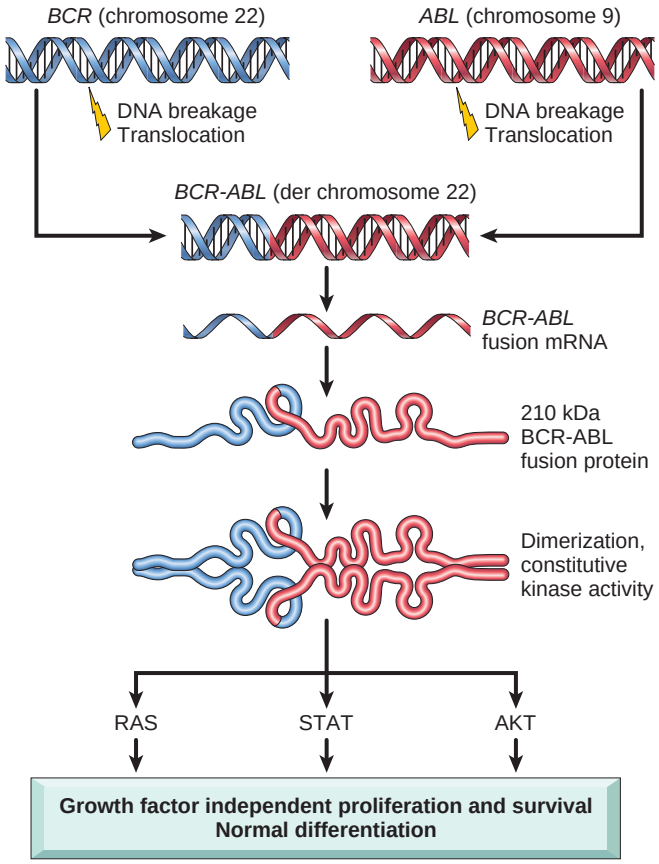


Figure 13.33 Pathogenesis of chronic myeloid leukemia. Breakage and joining of *BCR* and *ABL* creates a chimeric *BCR-ABL* fusion gene that encodes a constitutively active BCR-ABL tyrosine kinase. BCR-ABL activates multiple downstream pathways, which drive growth factor–independent proliferation and survival of bone marrow progenitors. Because BCR-ABL does not interfere with differentiation, the net result is an increase in mature elements in the peripheral blood, particularly granulocytes and platelets. *der,* Derivative chromosome.

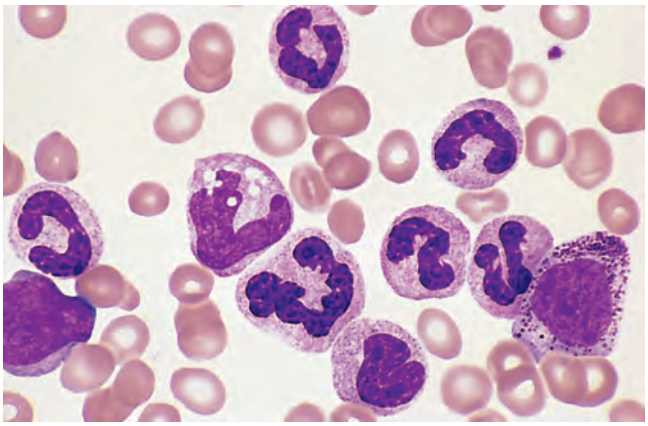


Figure 13.34 Chronic myeloid leukemia. Peripheral blood smear shows many mature neutrophils, some metamyelocytes, and a myelocyte. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Figure 13.35 Chronic myeloid leukemia (spleen). Enlarged spleen (2630 g; normal: 150 to 200 g) with greatly expanded red pulp stemming from neoplastic hematopoiesis. (Courtesy Dr. Daniel Jones, Department of Pathology, Ohio State University, Columbus, OH.)

mildly decreased numbers. A characteristic finding is the presence of scattered macrophages with abundant wrinkled, green-blue cytoplasm—so-called sea-blue histiocytes. Increased deposition of reticulin is typical, but overt marrow fibrosis is rare early in the course. The blood reveals a **leukocytosis, often exceeding 100,000 cells/mm^3** (Fig. 13.34), which consists predominantly of neutrophils, band forms, metamyelocytes, myelocytes, eosinophils, and basophils. Blasts usually make up less than 10% of the circulating cells. Platelets are also usually increased, sometimes markedly. The spleen is often greatly enlarged as a result of extensive extramedullary hematopoiesis (Fig. 13.35), and often contains infarcts of varying age. Extramedullary hematopoiesis can also produce mild hepatomegaly and lymphadenopathy.

Clinical Features

CML is primarily a disease of adults but also occurs in children and adolescents. The peak incidence is in the fifth to sixth decades of life. There are about 4500 new cases per year in the United States.

The onset is insidious. Mild-to-moderate anemia and hypermetabolism due to increased cell turnover lead to fatigability, weakness, weight loss, and anorexia. Sometimes the first symptom is a dragging sensation in the abdomen caused by splenomegaly or the acute onset of left upper quadrant pain due to splenic infarction. CML is best differentiated from other myeloproliferative neoplasms by detection of the *BCR-ABL* fusion gene through either chromosomal analysis or PCR-based tests.

The natural history of untreated disease is one of slow progression; without any treatment, the median survival is about 3 years. After a variable period averaging 3 years, about 50% of patients enter an "accelerated phase" marked by increasing anemia and thrombocytopenia, sometimes accompanied by a rise in the number of basophils in the blood. Additional clonal cytogenetic abnormalities, such as trisomy 8, isochromosome 17q, or duplication of the Ph chromosome, often appear. Within 6 to 12 months, the accelerated phase terminates in a picture resembling acute leukemia *(blast crisis)*. In the other 50% of patients, blast crises occur abruptly without an accelerated phase. In 70% of crises, the blasts are of myeloid origin (myeloid blast crisis), whereas in most of the remainder the blasts are of pre–B cell origin (lymphoid blast crisis). This is taken as evidence that CML originates from a pluripotent stem cell with both myeloid and lymphoid potential.

Given that acute leukemias often stem from complementary mutations involving a transcription factor and a tyrosine kinase, it might be predicted that blast crisis would be caused by an acquired mutation in a key transcriptional regulator. This prediction has been realized in lymphoid blast crisis, in which 85% of cases are associated with mutations that interfere with the activity of Ikaros, a transcription factor encoded by the gene *IKZF1* that regulates the differentiation of hematopoietic progenitors. The same types of Ikaros mutations are also seen in BCR-ABL–positive B-ALL, suggesting that these two varieties of aggressive leukemia have a similar molecular origin.

Understanding of the pathogenesis of CML has led to the use of drugs that target BCR-ABL. Treatment with BCR-ABL inhibitors results in sustained hematologic remissions in greater than 90% of patients, with generally tolerable side effects. These inhibitors markedly decrease the number of BCR-ABL–positive cells in the marrow and elsewhere, but do not extinguish the CML "stem cell," which persists at low levels. As a result, it is not clear if BCR-ABL inhibitors are ever truly curative. However, this form of targeted therapy controls blood counts and substantially decreases the risk of transformation to the accelerated phase and blast crisis, which is the greatest threat to the patient. It may be that by lowering the proliferative drive of the BCR-ABL–positive progenitors, BCR-ABL inhibitors decrease the rate at which these cells acquire mutations that lead to disease progression.

The other major threat to the patient is the emergence of resistance to first-generation BCR-ABL inhibitors, which in about 50% of cases stems from mutations in BCR-ABL and in the remaining cases mutations in other kinases. This problem has been overcome in part by development of second- and third-generation kinase inhibitors that are active against mutated forms of BCR-ABL. For relatively young patients, HSC transplantation performed in the stable phase is curative in about 75% of cases. The outlook is less favorable once the accelerated phase or blast crisis supervenes, as transplantation and treatment with BCR-ABL inhibitors are both less effective in these settings.

Polycythemia Vera

Polycythemia vera (PCV) is characterized by increased marrow production of red cells, granulocytes, and platelets (panmyelosis), but it is the increase in red cells (polycythemia) that is responsible for most of the clinical symptoms. PCV must be differentiated from relative polycythemia resulting from hemoconcentration and other causes of absolute polycythemia (Chapter 14).

Pathogenesis

In PCV, the transformed progenitor cells have markedly decreased requirements for erythropoietin and other hematopoietic growth factors due to activating mutations in the tyrosine kinase JAK2. JAK2 participates in the JAK/STAT pathway, which lies downstream of multiple hematopoietic growth factor receptors, including the erythropoietin receptor. Because the pathway is constitutively active and red cell numbers are abnormally high, serum erythropoietin levels in PCV are low. This is in contradistinction to secondary forms of polycythemia, in which erythropoietin levels are high. The elevated hematocrit leads to increased blood viscosity and sludging. These hemodynamic factors, together with thrombocytosis and abnormal platelet function, make patients with PCV prone to both thrombosis and bleeding.

More than 97% of cases are associated with a mutation in *JAK2* that results in a valine-to-phenylalanine substitution at residue 617; other *JAK2* mutations are found in most (and perhaps all) of the remaining cases. The mutated forms of *JAK2* found in PCV render hematopoietic cell lines growth factor–independent and when expressed in murine bone marrow progenitors cause a PCV-like disease that is associated with marrow fibrosis. In 25% to 30% of cases the tumor cells contain two mutated copies of *JAK2,* a genotype that is associated with higher white cell counts, more significant splenomegaly, symptomatic pruritus, and a greater rate of progression to the spent phase.

The proliferative drive in PCV (and other myeloproliferative neoplasms associated with *JAK2* mutations) is less than in CML, which is associated with more pronounced marrow hypercellularity, leukocytosis, and splenomegaly. Presumably, JAK2 signals are quantitatively weaker or qualitatively different from those produced by BCR-ABL (see Fig. 13.33).

MORPHOLOGY

The marrow is hypercellular, but some residual fat is usually present. **The increase in red cell progenitors is subtle and usually accompanied by an increase in granulocytic precursors and megakaryocytes as well.** At diagnosis, a moderate to marked increase in reticulin fibers is seen in about 10% of marrows. Mild organomegaly is common, being caused early in the course largely by congestion; at this stage extramedullary hematopoiesis is minimal. The peripheral blood often contains increased numbers of basophils and abnormally large platelets.

Late in the course, PCV often progresses to a **spent phase** characterized by extensive marrow fibrosis that displaces hematopoietic cells. This is accompanied by increased extramedullary hematopoiesis in the spleen and liver, often leading to prominent organomegaly (Fig. 13.36). Transformation to AML, with its typical features, occurs in about 1% of patients.

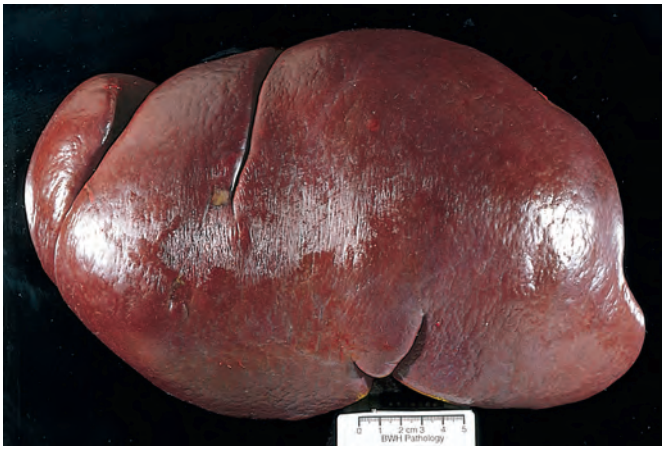


Figure 13.36 Polycythemia vera, spent phase. Massive splenomegaly (3020 g; normal: 150 to 200 g) largely due to extramedullary hematopoiesis occurring in the setting of advanced marrow myelofibrosis. (Courtesy Dr. Mark Fleming, Department of Pathology, Children's Hospital, Boston, Mass.)

Clinical Features

PCV is uncommon, having an incidence of 1 to 3 per 100,000 per year. It appears insidiously, usually in adults of late middle age. Most symptoms are related to the increased red cell mass and hematocrit. Usually, there is also an increased total blood volume. Together, these factors cause abnormal blood flow, particularly on the low-pressure venous side of the circulation, which becomes greatly distended. Patients are plethoric and cyanotic due to stagnation and deoxygenation of blood in peripheral vessels. Headache, dizziness, hypertension, and gastrointestinal symptoms are common. Intense pruritus and peptic ulceration may occur, both possibly resulting from the release of histamine from basophils. High cell turnover gives rise to hyperuricemia; symptomatic gout is seen in 5% to 10% of cases.

More ominously, the abnormal blood flow and platelet function that accompany PCV lead to an increased risk of both major bleeding and thrombotic episodes. About 25% of patients first come to attention due to deep venous thrombosis, myocardial infarction, or stroke. Thromboses sometimes also occur in the hepatic veins (producing Budd-Chiari syndrome) and the portal and mesenteric veins (leading to bowel infarction). It must be remembered that thrombotic complications may precede the appearance of the typical hematologic findings. Minor hemorrhages (epistaxis, bleeding gums) are common, and life-threatening hemorrhages occur in 5% to 10% of cases.

The hemoglobin concentration is typically greater than 16 g/dL, and the hematocrit is often 55% or more. Sometimes chronic bleeding leads to iron deficiency, which can suppress erythropoiesis sufficiently to lower the hematocrit into the normal range, an example of two defects counteracting one another to "correct" a laboratory abnormality. The white cell count ranges from 12,000 to 50,000 cells/mm^3, and the platelet count is often greater than 500,000 platelets/mm^3. The platelets usually exhibit morphologic abnormalities such as giant forms and are often defective in functional aggregation studies.

Without treatment, death from bleeding or thrombosis occurs within months of diagnosis. However, simply maintaining the red cell mass at nearly normal levels by phlebotomy extends the median survival to about 10 years. Although JAK2 inhibitors are an obvious targeted therapy for PCV, current JAK2 inhibitors are not as effective as was hoped (possibly because of the quality of the available drugs) and are used mainly in patients with significant splenomegaly. Extended survival with treatment has revealed that PCV tends to evolve to a "spent phase," during which clinical and anatomic features of primary myelofibrosis develop. The disease undergoes this transition in about 15% to 20% of patients after an average period of 10 years. It is marked by the appearance of obliterative fibrosis in the bone marrow (myelofibrosis) and extensive extramedullary hematopoiesis, principally in the spleen, which enlarges greatly. The mechanisms underlying the progression to the spent phase are not known.

In about 2% of patients, PCV transforms to AML. Surprisingly, the AML clone often lacks *JAK2* mutations, suggesting that the causative *JAK2* mutations occur in an abnormal stem cell that already harbors other oncogenic mutations and therefore is "at risk" for giving rise to several different myeloid tumors. Unlike CML, transformation to ALL is rarely observed, consistent with the cell of origin being a progenitor committed to myeloid differentiation.

Essential Thrombocytosis

Essential thrombocytosis (ET) is associated with diverse mutations that increase JAK-STAT signaling and mimic constitutive growth factor receptor signaling. Over 90% of cases have either activating mutations in JAK2 (approximately 50% to 60% of cases); MPL (5% to 10% of cases), a receptor tyrosine kinase that is normally activated by thrombopoietin and that signals through JAK2; or calreticulin (approximately 30% of cases). Remarkably, the mutated forms of CALR seen in ET are secreted and bind and activate the thrombopoietin receptor, an example of an oncogenic autocrine feedback loop.

ET manifests clinically with elevated platelet counts and is separated from PCV and primary myelofibrosis based on the absence of polycythemia and marrow fibrosis, respectively. In cases without tyrosine kinase mutations, causes of reactive thrombocytosis, such as inflammatory disorders and iron deficiency, must be excluded before the diagnosis can be established.

Pathogenesis

Constitutive JAK2 or MPL signaling renders the progenitors thrombopoietin-independent and leads to hyperproliferation. The *JAK2* mutation is the same as that found in almost all cases of PCV. Why some patients with *JAK2* mutations present with PCV and others with ET is not understood. Some cases thought to be ET may in fact be PCV disguised by iron deficiency (which is more common in individuals diagnosed with ET), but this is probably true of only a small fraction of patients. As mentioned, most cases without JAK2 or MPL mutations have calreticulin mutations instead.

MORPHOLOGY

Bone marrow cellularity is usually only mildly increased, but megakaryocytes are often markedly increased in number and include abnormally large forms. Delicate reticulin fibrils are often seen, but the overt fibrosis of primary myelofibrosis (see later) is absent. Peripheral smears usually reveal abnormally large platelets (Fig. 13.37), often accompanied by mild leukocytosis. Modest degrees of extramedullary hematopoiesis may occur, producing mild organomegaly in about 50% of patients. Uncommonly, a spent phase of marrow fibrosis or transformation to AML supervenes.

Clinical Features

The incidence of ET is 1 to 3 per 100,000 per year. It usually occurs past the age of 60 but may also be seen in young adults. Dysfunctions of platelets derived from the neoplastic clone can lead to thrombosis and hemorrhage, the major clinical manifestations. Platelets are not only increased in number but also frequently demonstrate qualitative abnormalities in functional tests. The types of thrombotic events resemble those observed in PCV; they include deep venous thrombosis, portal and hepatic vein thrombosis, and myocardial infarction. One characteristic symptom is *erythromelalgia,* a throbbing and burning of hands and feet caused by occlusion of small arterioles by platelet aggregates, which also may be seen in PCV.

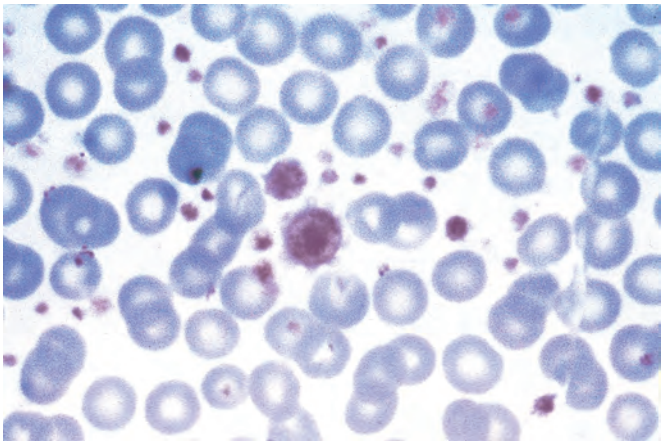


Figure 13.37 Essential thrombocytosis. Peripheral blood smear shows marked thrombocytosis including giant platelets approximating the size of surrounding red cells.

ET is an indolent disorder with long asymptomatic periods punctuated by occasional thrombotic or hemorrhagic crises. Median survival times are 12 to 15 years. Thrombotic complications are most likely in patients with very high platelet counts and homozygous *JAK2* mutations. Therapy consists of "gentle" chemotherapeutic agents that suppress thrombopoiesis.

Primary Myelofibrosis

The hallmark of primary myelofibrosis is the development of obliterative marrow fibrosis. The replacement of the marrow by fibrous tissue reduces bone marrow hematopoiesis, leading to cytopenias and extensive extramedullary hematopoiesis. Histologically, the appearance is identical to the spent phase that occurs occasionally late in the course of other myeloproliferative neoplasms. The genetics of primary myelofibrosis is very similar to ET; approximately 90% cases have activating mutations of *JAK2, CALR,* or *MPL.*

Pathogenesis

The chief pathologic feature is the extensive deposition of collagen in the marrow by nonneoplastic fibroblasts. The fibrosis inexorably displaces hematopoietic elements, including stem cells, from the marrow and eventually leads to marrow failure. It is probably caused by the inappropriate release of fibrogenic factors from neoplastic megakaryocytes. Two factors synthesized by megakaryocytes have been implicated: *platelet-derived growth factor* and *TGF-β.* As you recall, platelet-derived growth factor and TGF-β are fibroblast mitogens. In addition, TGF-β promotes collagen deposition and causes angiogenesis, both of which are observed in myelofibrosis (Chapter 3).

As marrow fibrosis progresses, circulating HSCs take up residence in niches in secondary hematopoietic organs, such as the spleen, the liver, and the lymph nodes, leading to the appearance of extramedullary hematopoiesis. For incompletely understood reasons, red cell production at extramedullary sites is disordered. This factor and the concomitant suppression of marrow function result in moderate to severe anemia. It is not clear whether primary myelofibrosis is truly distinct from PCV and ET or merely reflects unusually rapid progression to the spent phase.

MORPHOLOGY

Early in the course, the marrow is often hypercellular due to increases in maturing cells of all lineages, a feature reminiscent of PCV. Morphologically, the erythroid and granulocytic precursors appear normal, but megakaryocytes are large, dysplastic, and abnormally clustered. At this stage fibrosis is minimal, and the blood may show leukocytosis and thrombocytosis. **With progression, the marrow becomes more hypocellular and diffusely fibrotic.** Clusters of atypical megakaryocytes with unusual nuclear shapes (described as "cloud-like") are seen, and hematopoietic elements are often found within dilated sinusoids, which is a manifestation of severe architectural distortion caused by the fibrosis. Very late in the course, the fibrotic marrow space may be converted into bone, a change called "osteosclerosis." These features are identical to those seen in the spent phase of other myeloproliferative neoplasms.

Fibrotic obliteration of the marrow space leads to extensive extramedullary hematopoiesis, principally in the spleen, which is usually markedly enlarged, sometimes up to 4000 g. Grossly, such spleens are firm and diffusely red to gray. As in CML, subcapsular infarcts are common (see Fig. 13.41). Initially, extramedullary hematopoiesis is confined to the sinusoids, but later it expands into the cords. The liver may be enlarged moderately by sinusoidal foci of extramedullary hematopoiesis. Hematopoiesis may also appear within lymph nodes, but significant lymphadenopathy is uncommon.

The marrow fibrosis is reflected in several characteristic blood findings (Fig. 13.38). Marrow distortion leads to the premature release of nucleated erythroid and early granulocyte progenitors **(leukoerythroblastosis)**, and immature cells also enter the circulation from sites of extramedullary hematopoiesis. **Teardrop-shaped red cells** (dacryocytes), cells that were probably damaged during the birthing process in the fibrotic marrow, are also often seen. Although characteristic of primary myelofibrosis, leukoerythroblastosis and teardrop red cells are seen in many infiltrative disorders of the marrow, including granulomatous diseases and metastatic tumors. Other common, albeit nonspecific, blood findings include abnormally large platelets and basophilia.

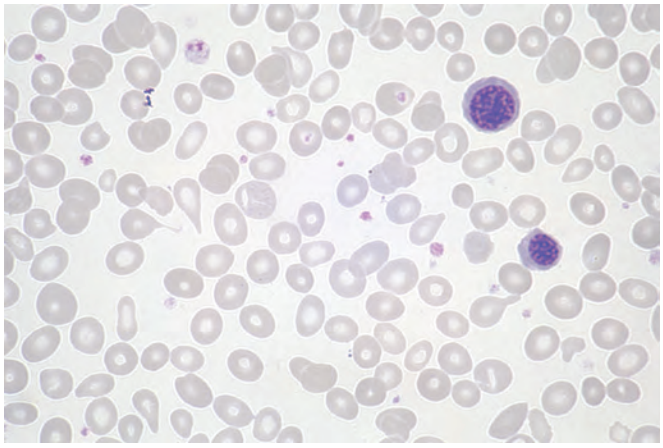


Figure 13.38 Primary myelofibrosis (peripheral blood smear). Two nucleated erythroid precursors and several teardrop-shaped red cells (dacryocytes) are evident. Immature myeloid cells were present in other fields. An identical picture can be seen in other diseases producing marrow distortion and fibrosis.

Clinical Features

Primary myelofibrosis is less common than PCV and ET and usually occurs in individuals older than 60 years of age. Except when preceded by another myeloproliferative neoplasm, it comes to attention because of progressive anemia and splenomegaly, which produces a sensation of fullness in the left upper quadrant. Nonspecific symptoms such as fatigue, weight loss, and night sweats result from an increase in metabolism associated with the expanding mass of hematopoietic cells. Hyperuricemia and secondary gout due to a high rate of cell turnover can complicate the picture.

Laboratory studies typically show a moderate to severe normochromic normocytic anemia accompanied by leukoerythroblastosis. The white cell count is usually normal or mildly reduced, but can be markedly elevated (80,000 to 100,000 cells/mm^3) early in the course. The platelet count is usually normal or elevated at the time of diagnosis, but thrombocytopenia may supervene as the disease progresses. These blood findings are not specific; bone marrow biopsy is essential for diagnosis.

Primary myelofibrosis is a much more difficult disease to treat than PCV or ET. The course is variable, but the median survival is in the range of 3 to 5 years. Threats to life include intercurrent infections, thrombotic episodes, bleeding related to platelet abnormalities, and transformation to AML, which occurs in 5% to 20% of cases. When myelofibrosis is extensive, AML sometimes arises at extramedullary sites, including lymph nodes and soft tissues. JAK2 inhibitors are approved to treat this disease and are effective at decreasing the splenomegaly and constitutional symptoms. HSC transplantation offers some hope for cure in patients young and fit enough to withstand the procedure.

KEY CONCEPTS

MYELOID NEOPLASMS

Myeloid tumors occur mainly in adults and fall into three major groups.

AML

- Aggressive tumor comprised of immature myeloid lineage blasts, which replace the marrow and suppress normal hematopoiesis.
- Associated with diverse acquired mutations that lead to expression of abnormal transcription factors, which interfere with myeloid differentiation.
- Often also associated with mutations in genes encoding growth factor receptor signaling pathway components or regulators of the epigenome.

Myeloproliferative Neoplasms

- Myeloid tumors in which production of formed myeloid elements is initially increased, leading to high blood counts and extramedullary hematopoiesis.
- Commonly associated with acquired mutations that lead to constitutive activation of tyrosine kinases, which mimic signals from normal growth factors. The most common pathogenic kinases are BCR-ABL (associated with CML) and mutated *JAK2* (associated with PCV and primary myelofibrosis).
- All can transform to acute leukemia and to a spent phase of marrow fibrosis associated with anemia, thrombocytopenia, and splenomegaly.

MDS

- Myeloid tumors characterized by disordered and ineffective hematopoiesis and dysmaturation.
- May occur de novo or following genotoxic exposures.
- Frequently harbor mutations in splicing factors, epigenetic regulators, and transcription factors.
- Manifest with one or more cytopenias and often progress to AML.

Langerhans Cell Histiocytosis

The term *histiocytosis* is an "umbrella" designation for a variety of proliferative disorders of dendritic cells or macrophages. Some, such as rare "histiocytic" sarcomas, are clearly malignant, whereas others, such as reactive proliferations of macrophages in lymph nodes, are clearly benign. Lying between these two extremes are the Langerhans cell histiocytoses, a spectrum of clonal proliferations of a special type of immature dendritic cell called the Langerhans cell. Because Langerhans cells are part of the innate immune system, these tumors (as well as other clonal histiocytoses) can be considered unusual myeloid neoplasms.

Pathogenesis

The origin and nature of the proliferating cells in Langerhans cell histiocytosis has been controversial, but it is now recognized that the majority of cases have driver mutations in cancer-associated genes and are probably best considered neoplasms, albeit with an unusual proclivity for spontaneous remission. The most common mutation is an activating valine-to-glutamate substitution at residue 600 in BRAF, already discussed for its role in hairy cell leukemia, which is present in 55% to 60% of cases. Less common mutations have also been detected in TP53, RAS, and the receptor tyrosine kinase MET.

One factor that contributes to the homing of neoplastic Langerhans cells is the aberrant expression of chemokine receptors. For example, while normal epidermal Langerhans cells express CCR6, their neoplastic counterparts express both CCR6 and CCR7. This allows the neoplastic cells to migrate into tissues that express the relevant chemokines—CCL20 (a ligand for CCR6) in skin and bone and CCL19 and CCL21 (ligands for CCR7) in lymphoid organs.

MORPHOLOGY

Regardless of the clinical picture, the proliferating Langerhans cells have abundant, often vacuolated cytoplasm and vesicular nuclei containing linear grooves or folds (Fig. 13.39A). The presence of *Birbeck granules* in the cytoplasm is characteristic. Birbeck granules are pentalaminar tubules, often with a dilated terminal end producing a tennis racket–like appearance (Fig. 13.39B), which contain the protein langerin. In addition, the tumor cells also typically express HLA-DR, S-100, and CD1a.

Clinical Features

Langerhans cell histiocytosis presents as several clinicopathologic entities:

- *Multifocal multisystem Langerhans cell histiocytosis (Letterer-Siwe disease)* occurs most frequently before 2 years of age

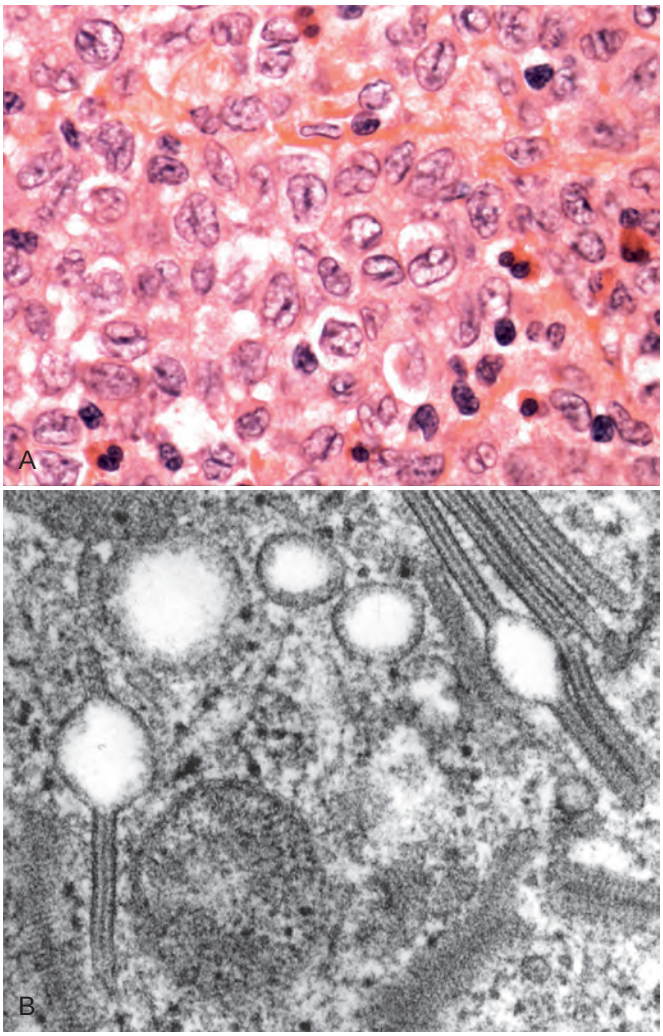


Figure 13.39 Langerhans cell histiocytosis. (A) Langerhans cells with folded or grooved nuclei and moderately abundant pale cytoplasm are mixed with a few eosinophils. (B) An electron micrograph shows rodlike Birbeck granules with characteristic periodicity and dilated terminal end. (B, Courtesy Dr. George Murphy, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

but occasionally affects adults. A dominant clinical feature is the development of cutaneous lesions resembling a seborrheic eruption, which is caused by infiltrates of Langerhans cells over the front and back of the trunk and on the scalp. Most of those affected have concurrent hepatosplenomegaly, lymphadenopathy, pulmonary lesions, and (eventually) destructive osteolytic bone lesions. Extensive infiltration of the marrow often leads to anemia, thrombocytopenia, and a predisposition to recurrent infections, such as otitis media and mastoiditis. In some instances the proliferating cells are quite anaplastic; such tumors are sometimes referred to as Langerhans cell sarcoma. The course of untreated disease is rapidly fatal. With intensive chemotherapy, 50% of patients survive 5 years.

- *Unifocal and multifocal unisystem Langerhans cell histiocytosis (eosinophilic granuloma)* is characterized by proliferations of Langerhans cells admixed with variable numbers of eosinophils, lymphocytes, plasma cells, and neutrophils. Eosinophils are usually, but not always, a prominent component of the infiltrate. It typically arises within the medullary cavities of bones, most commonly the calvaria, ribs, and femur. Less commonly, unisystem lesions of identical histology arise in the skin, lungs, or stomach. *Unifocal lesions* most commonly affect the skeletal system in older children or adults. Bone lesions can be asymptomatic or cause pain, tenderness, and, in some instances, pathologic fractures. Unifocal disease is indolent and may heal spontaneously or may be cured by local excision or irradiation. *Multifocal unisystem disease* usually affects young children, who present with multiple erosive bony masses that sometimes expand into adjacent soft tissue. Involvement of the posterior pituitary stalk of the hypothalamus leads to diabetes insipidus in about 50% of patients. The combination of calvarial bone defects, diabetes insipidus, and exophthalmos is referred to as the *Hand-Schüller-Christian* triad. Many patients experience spontaneous regression; others can be treated successfully with chemotherapy. BRAF inhibitors are active against *BRAF*-mutated disease but are not curative; combinations of BRAF inhibitors and chemotherapy are being tested in clinical trials.
- *Pulmonary Langerhans cell histiocytosis* represents a special category of disease, most often seen in adult smokers, which may regress spontaneously upon cessation of smoking. These lesions have been described as reactive proliferations of Langerhans cells, but fully 40% are associated with BRAF mutations, suggesting that in many instances they too are neoplastic in origin.

Spleen

The spleen is an ingeniously designed filter for the blood and a site of immune responses to blood-borne antigens. Normally in the adult it weighs about 150 g and is enclosed within a thin, glistening, slate-gray connective tissue capsule. Its cut surface reveals extensive red pulp dotted with gray specks, which are the white pulp follicles. These consist of an artery with an eccentric collar of T lymphocytes, the so-called periarteriolar lymphatic sheath. At intervals this sheath expands to form lymphoid follicles containing mainly B lymphocytes, which are capable of developing into germinal centers identical to those seen in lymph nodes in response to antigenic stimulation (Fig. 13.40).

The red pulp of the spleen is traversed by numerous thin-walled vascular sinusoids, separated by the splenic cords or "cords of Billroth." The endothelial lining of the sinusoid is discontinuous, providing a passage for blood cells between the sinusoids and cords. The cords contain a labyrinth of macrophages loosely connected through long dendritic processes to create both a physical and a functional filter. As it traverses the red pulp, the blood takes two routes

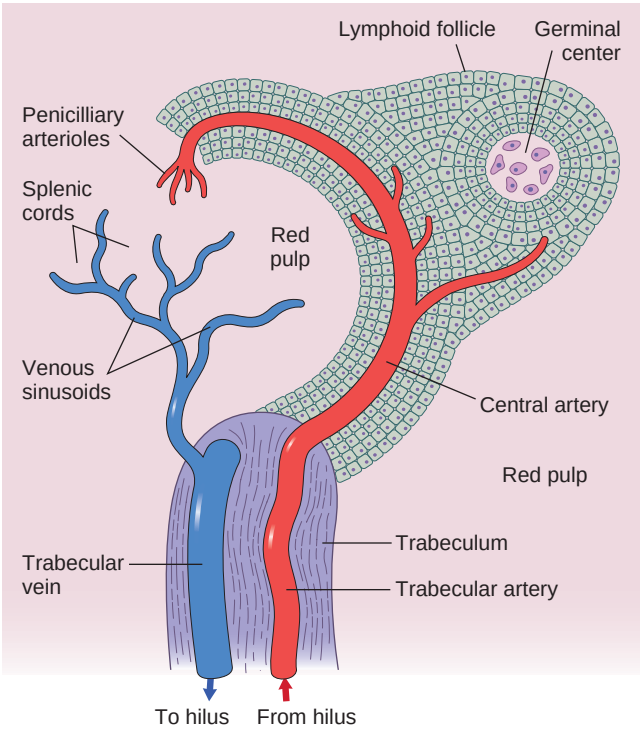


Figure 13.40 Normal splenic architecture. (Modified from Faller DV: Diseases of the spleen. In Wyngaarden JB, Smith LH, editors: *Cecil Textbook of Medicine*, ed 18, Philadelphia, 1988, WB Saunders, p 1036.)

to reach the splenic veins. Some flows through capillaries into the cords, from which blood cells squeeze through gaps in the discontinuous basement of the endothelial lining to reach the sinusoids; this is the so-called open circulation or slow compartment. In the other "closed circuit," blood passes rapidly and directly from the capillaries to the splenic veins. Although only a small fraction of the blood pursues the "open" route, during the course of a day the entire blood volume passes through the cords, where it is closely examined by macrophages.

The spleen has four functions that impact disease states:

- *Phagocytosis of blood cells and particulate matter.* As is discussed under the hemolytic anemias (Chapter 14), red cells undergo extreme deformation during passage from the cords into the sinusoids. In conditions in which red cell deformability is decreased, red cells become trapped in the cords and are more readily phagocytosed by macrophages. Splenic macrophages also are responsible for "pitting" of red cells, the process by which inclusions such as *Heinz bodies* and *Howell-Jolly bodies* are excised, and for the removal of particles, such as bacteria, from the blood.
- *Antibody production.* Dendritic cells in the periarterial lymphatic sheath trap antigens and present them to T lymphocytes. T- and B-cell interaction at the edges of white pulp follicles leads to the generation of antibody-secreting plasma cells, which are found mainly within the sinuses of the red pulp. The spleen seems to be an important site of production of antibodies against microbial polysaccharides as well as autoantibodies against a variety of self antigens.
- *Hematopoiesis.* During fetal development, the spleen may be a minor site of hematopoiesis, but this normally disappears by birth. However, the spleen can become a major site of compensatory extramedullary hematopoiesis in the setting of severe chronic anemia (e.g., in patients with thalassemia, described in Chapter 14) and in patients with myeloproliferative neoplasms, such as CML and primary myelofibrosis.
- *Sequestration of formed blood elements.* The normal spleen contains only about 30 to 40 mL of red cells, but this volume increases greatly with splenomegaly. The spleen also normally harbors approximately 30% to 40% of the total platelet mass in the body. With splenomegaly, up to 80% to 90% of the total platelet mass may be sequestered in the interstices of the red pulp, producing thrombocytopenia. Similarly, the enlarged spleen may trap white cells and thereby induce leukopenia.

As the largest unit of the mononuclear phagocyte system, the spleen is involved in all systemic inflammations, generalized hematopoietic disorders, and many metabolic disturbances. In each, the spleen undergoes enlargement *(splenomegaly)*, which is the major manifestation of disorders of this organ. It is rarely the primary site of disease. Splenic insufficiency due to splenectomy or autoinfarction (as in sickle-cell disease) has one major clinical manifestation, an increased susceptibility to sepsis caused by encapsulated bacteria such as pneumococci, meningococci, and *Haemophilus influenzae.* The decreases in phagocytic capacity and antibody production that result from asplenia both contribute to the increased risk of sepsis, which may be fatal. All asplenic individuals should be vaccinated against these agents to reduce the risk of this tragic complication.

SPLENOMEGALY

When sufficiently enlarged, the spleen causes a dragging sensation in the left upper quadrant and, through pressure on the stomach, discomfort after eating. In addition, enlargement can cause a syndrome known as *hypersplenism,* which is characterized by anemia, leukopenia, and/or thrombocytopenia. The probable cause of the cytopenias is increased sequestration of formed elements and the consequent enhanced phagocytosis by the splenic macrophages.

Table 13.12 lists major disorders associated with splenomegaly. Splenomegaly in virtually all the conditions mentioned has been discussed elsewhere. A few disorders are left to consider.

Nonspecific Acute Splenitis

Enlargement of the spleen occurs in any blood-borne infection. The nonspecific splenic reaction in these infections is caused both by the microbiologic agents themselves and by cytokines that are released as part of the immune response.

MORPHOLOGY

The spleen is enlarged (200 to 400 g) and soft. Microscopically, the major feature is acute congestion of the red pulp, which may

Table 13.12 Disorders Associated With Splenomegaly

I. Infections
Nonspecific splenitis of various blood-borne infections (particularly infectious endocarditis)
Infectious mononucleosis
Tuberculosis
Typhoid fever
Brucellosis
Cytomegalovirus
Syphilis
Malaria
Histoplasmosis
Toxoplasmosis
Kala-azar
Trypanosomiasis
Schistosomiasis
Leishmaniasis
Echinococcosis
II. Congestive States Related to Portal Hypertension
Cirrhosis of the liver
Portal or splenic vein thrombosis
Cardiac failure
III. Lymphohematogenous Disorders
Hodgkin lymphoma
Non-Hodgkin lymphomas and lymphocytic leukemias
Multiple myeloma
Myeloproliferative neoplasms
Hemolytic anemias
IV. Immunologic-Inflammatory Conditions
Rheumatoid arthritis
Systemic lupus erythematosus
V. Storage Diseases
Gaucher disease
Niemann-Pick disease
Mucopolysaccharidoses
VI. Miscellaneous Disorders
Amyloidosis
Primary neoplasms and cysts
Secondary neoplasms

encroach on and virtually efface the lymphoid follicles. Neutrophils, plasma cells, and occasionally eosinophils are usually present throughout the white and red pulp. At times the white pulp follicles may undergo necrosis, particularly when the causative agent is a hemolytic streptococcus. Rarely, abscess formation occurs.

Congestive Splenomegaly

Chronic venous outflow obstruction causes a form of splenic enlargement referred to as *congestive splenomegaly*. Venous obstruction may be caused by intrahepatic disorders that retard portal venous drainage or extrahepatic disorders that directly impinge upon the portal or splenic veins. All of these disorders ultimately lead to portal or splenic vein hypertension. Systemic, or central, venous congestion is encountered in cardiac decompensation involving the right side of the heart, as can occur in tricuspid or pulmonic valvular disease, chronic cor pulmonale, or following left-sided heart failure. Systemic congestion is associated with only moderately enlarged spleens that rarely exceed 500 g in weight.

Cirrhosis of the liver is the main cause of massive congestive splenomegaly. The "pipe-stem" hepatic fibrosis of schistosomiasis causes particularly severe congestive splenomegaly, while the diffuse fibrous scarring of alcoholic cirrhosis and pigment cirrhosis also evokes profound enlargements. Other forms of cirrhosis are less commonly implicated.

Congestive splenomegaly also may be caused by obstruction of the extrahepatic portal vein or splenic vein. This can stem from *portal vein thrombosis,* which is usually associated with some intrahepatic obstructive disease or inflammation of the portal vein *(pyelophlebitis)*, such as follows intraperitoneal infections. Thrombosis of the splenic vein can be caused by infiltrating tumors arising in neighboring organs, such as carcinomas of the stomach or pancreas.

MORPHOLOGY

Long-standing splenic congestion produces marked enlargement (1000 to 5000 g). The organ is firm, and the capsule is usually thickened and fibrous. Microscopically, the red pulp is congested early in the course but becomes increasingly fibrotic and cellular with time. The elevated portal venous pressure stimulates the deposition of collagen in the basement membrane of the sinusoids, which appear dilated because of the rigidity of their walls. The resultant slowing of blood flow from the cords to the sinusoids prolongs the exposure of the blood cells to macrophages, resulting in excessive destruction (hypersplenism).

Splenic Infarcts

Splenic infarcts are common lesions caused by the occlusion of the major splenic artery or any of its branches. The lack of extensive collateral blood supply predisposes to infarction following vascular occlusion. The spleen, along with kidneys and brain, ranks as one of the most frequent sites where emboli lodge. In normal-sized spleens, infarcts are most often caused by emboli that arise from the heart. The infarcts can be small or large, be single or multiple, or even involve the entire organ. They are usually bland except in individuals with infectious endocarditis of the mitral or aortic valves, in whom septic infarcts are common. Infarcts also are common in markedly enlarged spleens, regardless of cause, presumably because the blood supply is tenuous and easily compromised.

MORPHOLOGY

Bland infarcts are characteristically pale, wedge-shaped, and subcapsular in location. The overlying capsule is often covered with fibrin (Fig. 13.41). In septic infarcts this appearance is modified by the development of suppurative necrosis. In the course of healing, large depressed scars often develop.

NEOPLASMS

Neoplastic involvement of the spleen is rare except in myeloid and lymphoid tumors, which often cause splenomegaly

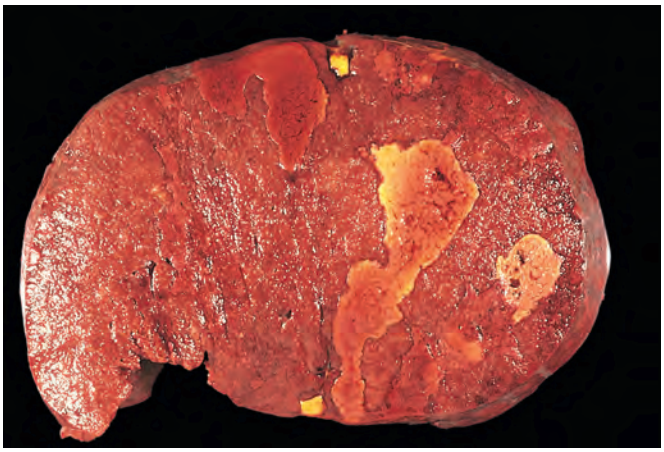


Figure 13.41 Splenic infarcts. Multiple well-circumscribed infarcts are present in this spleen, which is massively enlarged (2820 g; normal: 150 to 200 g) by extramedullary hematopoiesis secondary to a myeloproliferative neoplasm (myelofibrosis). Recent infarcts are hemorrhagic, whereas older, more fibrotic infarcts are a pale yellow-gray color.

(discussed earlier). Benign fibromas, osteomas, chondromas, lymphangiomas, and hemangiomas may arise in the spleen. Of these, lymphangiomas and hemangiomas are most common and often cavernous in type.

CONGENITAL ANOMALIES

Complete absence of the spleen is rare and is usually associated with other congenital abnormalities, such as situs inversus and cardiac malformations. Hypoplasia is a more common finding.

Accessory spleens (spleniculi) are common, being present singly or multiply in 20% to 35% of postmortem examinations. They are small, spherical structures that are histologically and functionally identical to the normal spleen. They can be found at any place within the abdominal cavity. Accessory spleens are of clinical importance in some hematologic disorders, such as hereditary spherocytosis (Chapter 14), where splenectomy may be used as a treatment. If an accessory spleen is overlooked, the therapeutic benefit of removal of the definitive spleen may be reduced or lost entirely.

RUPTURE

Splenic rupture is usually precipitated by blunt trauma. Much less often, it occurs in the apparent absence of a physical blow. Such "spontaneous ruptures" never involve truly normal spleens but rather stem from some minor physical insult to a spleen made fragile by an underlying condition. The most common predisposing conditions are infectious mononucleosis, malaria, typhoid fever, and lymphoid neoplasms, which may cause the spleen to enlarge rapidly, producing a thin, tense capsule that is susceptible to rupture. This dramatic event often precipitates intraperitoneal hemorrhage, which must be treated by prompt splenectomy to prevent death from blood loss. Chronically enlarged spleens are unlikely to rupture because of the toughening effect of extensive reactive fibrosis.

Thymus

Once an organ buried in obscurity, the thymus has a starring role in cell-mediated immunity (Chapter 6). Here, our interest centers on the disorders of the gland itself.

The thymus is embryologically derived from the third and, inconstantly, the fourth pair of pharyngeal pouches. At birth it weighs 10 to 35 g. It grows until puberty, achieving a maximum weight of 20 to 50 g, and thereafter undergoes progressive involution to little more than 5 to 15 g in older adults. The thymus also may involute in children and young adults in response to severe illness and HIV infection.

The fully developed thymus is composed of two fused, well-encapsulated lobes. Fibrous extensions of the capsule divide each lobe into numerous lobules, each with an outer cortical layer enclosing the central medulla. Diverse types of cells populate the thymus, but thymic epithelial cells and immature T lymphocytes, also called thymocytes, predominate. Epithelial cells in the cortex are polygonal in shape and have an abundant cytoplasm with dendritic extensions that contact adjacent cells. In contrast, the epithelial cells in the medulla are densely packed, often spindle-shaped, and have scant cytoplasm devoid of interconnecting processes. Whorls of medullary epithelial cells create *Hassall corpuscles,* with their characteristic keratinized cores.

As you know from consideration of the thymus in relation to immunity (Chapter 6), progenitor cells migrate from the marrow to the thymus and mature into T cells, which are exported to the periphery, but only after they have been educated in the "thymic university" to distinguish between self antigens and non–self antigens. During adulthood the thymic production of T cells slowly declines as the organ atrophies.

Macrophages, dendritic cells, a minor population of B lymphocytes, rare neutrophils and eosinophils, and scattered myoid (muscle-like) cells are also found within the thymus. The myoid cells are of particular interest because they likely play some role in the development of myasthenia gravis, a musculoskeletal disorder of immune origin.

Pathologic changes within the thymus are limited and will be described here. The changes associated with myasthenia gravis are considered in Chapter 27.

DEVELOPMENTAL DISORDERS

Thymic hypoplasia or *aplasia* is seen in DiGeorge syndrome, which is marked by severe defects in cell-mediated immunity and variable abnormalities of parathyroid development and function. As discussed in Chapter 5, DiGeorge syndrome is often associated with other developmental defects as part of the 22q11 deletion syndrome.

Isolated *thymic cysts* are uncommon lesions that are usually discovered incidentally postmortem or during surgery. They rarely exceed 4 cm in diameter, can be spherical or arborizing, and are lined by stratified to columnar epithelium. The fluid contents can be serous or mucinous and are often modified by hemorrhage.

While isolated cysts are not clinically significant, neoplastic thymic masses (whatever their origin) compress and distort adjacent normal thymus and sometimes cause cysts to form. Therefore the presence of a cystic thymic lesion in a symptomatic patient should provoke a thorough search for a neoplasm, particularly a lymphoma or a thymoma.

THYMIC HYPERPLASIA

The term thymic hyperplasia is misleading, since it usually applies to the appearance of B-cell germinal centers within the thymus, a finding that is referred to as *thymic follicular hyperplasia*. Such B-cell follicles are present in only small numbers in the normal thymus. Although follicular hyperplasia can occur in a number of chronic inflammatory and immunologic states, it is most frequently encountered in myasthenia gravis, in which it is found in 65% to 75% of cases (Chapter 27). Similar thymic changes are sometimes encountered in Graves disease, systemic lupus erythematosus, scleroderma, rheumatoid arthritis, and other autoimmune disorders.

In other instances, a morphologically normal thymus is simply large for the age of the patient. As mentioned, the size of the thymus varies widely, and whether this constitutes a true hyperplasia or is merely a variant of normal is unclear. The main significance of this form of thymic "hyperplasia" is that it may be mistaken radiologically for a thymoma, leading to unnecessary surgical procedures.

THYMOMA

A diversity of neoplasms may arise in the thymus—germ cell tumors, lymphomas, carcinoids, and others—but **the designation "thymoma" is restricted to tumors of thymic epithelial cells.** Such tumors typically also contain benign immature T cells (thymocytes).

The WHO has created a classification system based on histology for thymomas, but its clinical utility remains uncertain. We instead use a classification that relies on the most important prognostic features, the surgical stage and the presence or absence of overt cytologic features of malignancy. In this simple system there are only three histologic subtypes:

- Tumors that are cytologically benign and noninvasive
- Tumors that are cytologically benign but invasive or metastatic
- Tumors that are cytologically malignant (thymic carcinoma)

In all categories the tumors usually occur in adults older than 40 years of age; thymomas are rare in children. Males and females are affected equally. Most tumors arise in the anterior superior mediastinum, but sometimes they occur in the neck, thyroid, pulmonary hilus, or elsewhere. They are uncommon in the posterior mediastinum. Thymomas account for 20% to 30% of tumors in the anterosuperior mediastinum, which is also a common location for certain lymphomas.

MORPHOLOGY

Macroscopically, thymomas are lobulated, firm, gray-white masses of up to 15 to 20 cm in size. They sometimes have areas of cystic necrosis and calcification. Most are encapsulated, but 20% to 25% of tumors penetrate the capsule and infiltrate perithymic tissues and structures.

Noninvasive thymomas are most often composed of medullary-type epithelial cells or a mixture of medullary- and cortical-type epithelial cells. The medullary-type epithelial cells are elongated or spindle-shaped (Fig. 13.42A). There is usually a sparse infiltrate of thymocytes, which often recapitulate the phenotype of medullary thymocytes. In mixed thymomas there is an admixture of polygonal cortical-type epithelial cells and a denser infiltrate of thymocytes. The medullary and mixed patterns

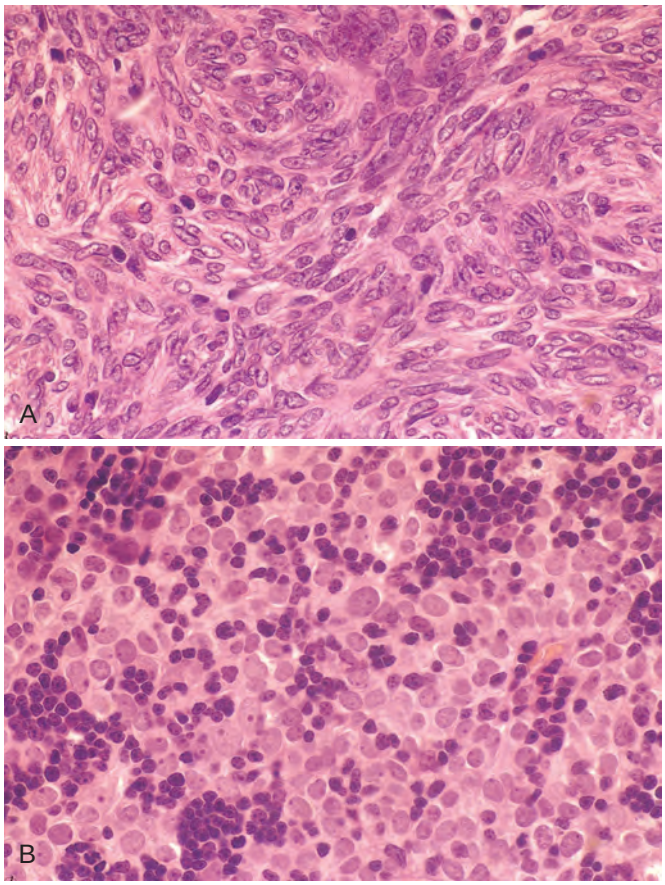


Figure 13.42 Thymoma. (A) Benign thymoma (medullary type). The neoplastic epithelial cells are arranged in a swirling pattern and have bland, oval to elongated nuclei with inconspicuous nucleoli. Only a few small, reactive lymphoid cells are interspersed. (B) Malignant thymoma, type I. The neoplastic epithelial cells are polygonal and have round to oval, bland nuclei with inconspicuous nucleoli. Numerous small, reactive lymphoid cells are interspersed. The morphologic appearance of this tumor is identical to that of benign thymomas of the cortical type. In this case, however, the tumor was locally aggressive, invading adjacent lung and pericardium.

together account for about 50% of all thymomas. Tumors that have a substantial proportion of medullary-type epithelial cells are usually noninvasive.

Invasive thymoma refers to a tumor that is cytologically benign but locally invasive. These tumors are much more likely to metastasize. The epithelial cells are most commonly of the cortical variety, with abundant cytoplasm and rounded vesicular nuclei (Fig. 13.42B), and are usually mixed with numerous thymocytes. In some cases the neoplastic cells show cytologic atypia, a feature that correlates with a propensity for more aggressive behavior. These tumors account for about 20% to 25% of all thymomas. **By definition, invasive thymomas penetrate through the capsule into surrounding structures.** The extent of invasion has been subdivided into various stages, which are beyond our scope. With minimal invasion, complete excision yields a 5-year survival rate of greater than 90%, whereas extensive invasion is associated with a 5-year survival rate of less than 50%.

Thymic carcinoma represents about 5% of thymomas. Macroscopically, they are usually fleshy, obviously invasive masses, sometimes accompanied by metastases to sites such as the lungs. Microscopically, most are **squamous cell carcinomas.** Another distinctive variant is **lymphoepithelioma-like carcinoma,** a tumor composed of sheets of cells that bears a close histologic resemblance to nasopharyngeal carcinoma. About 50% of lymphoepithelioma-like carcinomas contain monoclonal EBV genomes, consistent with a role for EBV in their pathogenesis. A variety of other, less common histologic patterns of thymic carcinoma have been described; all exhibit cytologic atypia seen in other carcinomas.

Clinical Features

About 40% of thymomas present with symptoms stemming from impingement on mediastinal structures. Another 30% to 45% are detected in the course of evaluating patients with myasthenia gravis. The rest are discovered incidentally during imaging studies or cardiothoracic surgery. In addition to myasthenia gravis, other associated autoimmune disorders include hypogammaglobulinemia, pure red cell aplasia, Graves disease, pernicious anemia, dermatomyositis-polymyositis, and Cushing syndrome. The basis for these associations is uncertain, but the thymocytes that arise within thymomas give rise to long-lived CD4+ and CD8+ T cells, and cortical thymomas rich in thymocytes are more likely to be associated with autoimmune disease. Hence it seems likely that abnormalities in the selection or "education" of T cells maturing within the environment of the neoplasm contribute to the development of diverse autoimmune disorders.

SUGGESTED READINGS

Hematopoietic Stem Cells

Gao X, Xu C, Asada N et al: The hematopoietic stem cell niche from embryo to adult, *Development* 145:dev139691, 2018. *[Discussion of origins of hematopoietic stem cells].*

Pinho S, Frenette PS: Haematopoietic stem cell activity and interactions with the niche, *Nat Rev Mol Cell Biol* 20:303, 2019. *[Discussion of the nature and biology of the marrow stem cell niche].*

White Cell Neoplasms

Allen CE, Merad M, McClain KL: Langerhans-cell histiocytosis, *N Engl J Med* 379:856, 2018. *[Up-to-date review of the pathogenesis and clinical features of Langerhans cell histiocytosis].*

Arber A, Orazi A, Hasserjian R et al: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia, *Blood* 127:2391, 2016. *[Overview of the most recent version of the classification system used for myeloid neoplasms].*

Hunger SP, Mullighan CG: Acute lymphoblastic leukemia in children, *N Engl J Med* 373:2015, 1541. *[Overview of the genetics, clinical features, and treatment of childhood acute lymphoblastic leukemia].*

Jain N, Keating M, Thompson P et al: Ibrutinib and venetoclax for first-line treatment of CLL, *N Engl J Med* 380:2095, 2019. *[Report of excellent response to therapy with agents that inhibit B-cell receptor signaling and BCL2 function, evidence for the importance of these factors in CLL cell growth and survival].*

Kumar SJ, Rajkuar V, Kyle RA et al: Multiple myeloma, *Nat Rev Dis Primers* 3:17046, 2017. *[Review of the pathogenesis, clinical features, and treatment of multiple myeloma].*

Küppers R, Stevenson FK: Critical influences on the pathogenesis of follicular lymphoma, *Blood* 131:2297, 2018. *[Discussion of new insights into the pathogenesis of this disease, including its genetics and its microenvironmental influences].*

Ogawa S: Genetics of MDS, *Blood* 133:1049, 2019. *[Comprehensive review of genetic drivers in myelodysplastic syndromes].*

Sallman DA, List A: The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes, *Blood* 133:1039, 2019. *[Recent review of emerging link between myelodysplastic syndromes and inflammation].*

Schmitz R, Ceribelli M, Pittaluga S et al: Oncogenic mechanisms in Burkitt lymphoma, *Cold Spring Harb Perspect Med* 4:a014282, 2014. *[Review of the molecular pathogenesis Burkitt lymphoma].*

Swerdlow SH, Campo E, Pileri SA et al: The 2016 revision of the World Health Organization classification of lymphoid neoplasms, *Blood* 127:2375, 2016. *[Overview of the most recent version of the classification system used for lymphoid neoplasms].*

Vainchenker W, Kralovics R: Genetic basis and molecular pathophysiology or classical myeloproliferative neoplasms, *Blood* 129:667, 2017. *[Update on the pathogenesis of the major subtypes of myeloproliferative neoplasms].*

Young RM, Phelan JD, Shaffer AL et al: Taming the heterogeneity of aggressive lymphomas for precision therapy, *Ann Rev Cancer Biol* 3:429, 2019. *[Excellent review of the molecular origins of diffuse large B-cell lymphoma and implications for therapy].*

CHAPTER 14

Red Blood Cell and Bleeding Disorders

CHAPTER CONTENTS

In this chapter, we will first consider diseases of red cells. By far, the most common and important are the anemias, red cell deficiency states that usually have a nonneoplastic basis. We will then complete our review of blood diseases by discussing the major bleeding disorders and complications of blood transfusion.

ANEMIAS

Anemia is defined as a reduction of the total circulating red cell mass below normal limits. Anemia reduces the oxygen-carrying capacity of the blood, leading to tissue hypoxia. In practice, the measurement of red cell mass is not easy, and anemia is usually diagnosed based on a reduction in the *hematocrit* (the ratio of packed red cells to total blood volume) and the *hemoglobin concentration* of the blood to levels that are below the normal range. These values correlate with the red cell mass except when there are changes in plasma volume caused by fluid retention or dehydration.

There are many classifications of anemia. We will follow one based on underlying mechanisms that is presented in Table 14.1. A second clinically useful approach classifies anemia according to alterations in red cell morphology, which often point to particular causes. Morphologic characteristics that provide etiologic clues include red cell size (normocytic, microcytic, or macrocytic); degree of hemoglobinization, reflected in the color of red cells (normochromic or hypochromic); and shape. Microcytic hypochromic anemias are caused by disorders of hemoglobin synthesis, and macrocytic anemias often stem from abnormalities that impair the maturation of erythroid precursors in the bone marrow. Normochromic, normocytic anemias have diverse etiologies; in some of these anemias, characteristic abnormalities of red cell shape provide an important clue as to the cause. Red cell shape is assessed through visual inspection of peripheral smears, whereas as other red cell indices are determined in clinical laboratories with special instrumentation. The most useful of these indices are as follows:

- *Mean cell volume:* the average volume of a red cell expressed in femtoliters (fL)
- *Mean cell hemoglobin:* the average content (mass) of hemoglobin per red cell, expressed in picograms (pg)
- *Mean cell hemoglobin concentration:* the average concentration of hemoglobin in a given volume of packed red cells, expressed in grams per deciliter (g/dL)
- *Red cell distribution width:* the coefficient of variation of red cell volume

Table 14.1 Classification of Anemia According to Underlying Mechanism

Mechanism	Specific Examples
Blood Loss	
Acute blood loss	Trauma
Chronic blood loss	Gastrointestinal tract lesions, gynecologic disturbances[a]
Increased Red Cell Destruction (Hemolysis)	
Inherited genetic defects	
Red cell membrane disorders	Hereditary spherocytosis, hereditary elliptocytosis
Enzyme deficiencies	
Hexose monophosphate shunt enzyme deficiencies	G6PD deficiency, glutathione synthetase deficiency
Glycolytic enzyme deficiencies	Pyruvate kinase deficiency, hexokinase deficiency
Hemoglobin abnormalities	
Deficient globin synthesis	Thalassemia syndromes
Structurally abnormal globins (hemoglobinopathies)	Sickle cell disease, unstable hemoglobins
Acquired genetic defects	
Deficiency of phosphatidylinositol-linked glycoproteins	Paroxysmal nocturnal hemoglobinuria
Antibody-mediated destruction	Hemolytic disease of the newborn (Rh disease), transfusion reactions, drug-induced, autoimmune disorders
Mechanical trauma	
Microangiopathic hemolytic anemias	Hemolytic uremic syndrome, disseminated intravascular coagulation, thrombotic thrombocytopenia purpura
Cardiac traumatic hemolysis	Defective cardiac valves
Repetitive physical trauma	Bongo drumming, marathon running, karate chopping
Infections of red cells	Malaria, babesiosis
Toxic or chemical injury	Clostridial sepsis, snake venom, lead poisoning
Membrane lipid abnormalities	Abetalipoproteinemia, severe hepatocellular liver disease
Sequestration	Hypersplenism
Decreased Red Cell Production	
Inherited genetic defects	
Defects leading to stem cell depletion	Fanconi anemia, telomerase defects
Defects affecting erythroblast maturation	Thalassemia syndromes
Nutritional deficiencies	
Deficiencies affecting DNA synthesis	B_{12} and folate deficiencies
Deficiencies affecting hemoglobin synthesis	Iron deficiency
Erythropoietin deficiency	Renal failure, anemia of chronic inflammation
Immune-mediated injury of progenitors	Aplastic anemia, pure red cell aplasia
Inflammation-mediated iron sequestration	Anemia of chronic inflammation
Primary hematopoietic neoplasms	Acute leukemia, myelodysplastic syndrome, myeloproliferative neoplasms (Chapter 13)
Space-occupying marrow lesions	Metastatic neoplasms, granulomatous disease
Infections of red cell progenitors	Parvovirus B19 infection
Unknown mechanisms	Endocrine disorders, hepatocellular liver disease

G6PD, Glucose-6-phosphate dehydrogenase.

[a]Most often anemia stems from iron deficiency, not bleeding per se.

Adult reference ranges for red cell indices are shown in Table 14.2.

Whatever its cause, when sufficiently severe anemia leads to manifestations related to the diminished hemoglobin and oxygen content of the blood. Patients appear pale and often report weakness, malaise, easy fatigability, and dyspnea on mild exertion. Hypoxia can cause fatty change in the liver, myocardium, and kidney. On occasion, myocardial hypoxia manifests as angina pectoris, particularly when complicated by pre-existing coronary artery disease. With acute blood loss and shock, oliguria and anuria can develop as a result of renal hypoperfusion. Central nervous system hypoxia can cause headache, dimness of vision, and faintness.

Anemias of Blood Loss

Acute Blood Loss

The effects of acute blood loss are mainly due to the loss of intravascular volume, which if massive can lead to cardiovascular collapse, shock, and death. The clinical features depend on the rate of hemorrhage and whether

Table 14.2 Adult Reference Ranges for Red Cells[a]

Measurement (Units)	Men	Women
Hemoglobin (g/dL)	13.6–17.2	12.0–15.0
Hematocrit (%)	39–49	33–43
Red cell count ($\times 10^6/\mu L$)	4.3–5.9	3.5–5.0
Reticulocyte count (%)	0.5–1.5	
Mean cell volume (fL)	82–96	
Mean cell hemoglobin (pg)	27–33	
Mean cell hemoglobin concentration (g/dL)	33–37	
Red cell distribution width	11.5–14.5	

[a]Reference ranges vary among laboratories. The reference ranges for the laboratory providing the result should always be used in interpreting test results.

the bleeding is external or internal. If the patient survives, the blood volume is rapidly restored by movement of water from the interstitial fluid compartment to the intravascular compartment. This fluid shift produces hemodilution and lowers the hematocrit. The resulting reduction in tissue oxygenation triggers increased secretion of erythropoietin from the kidney, which stimulates the proliferation of committed erythroid progenitors (colony-forming unit–erythroid [CFU–E]) in the marrow (see Fig. 13.1). It takes about 5 days for the progeny of these CFU–Es to mature and appear as newly released red cells (reticulocytes) in the peripheral blood. The iron in hemoglobin is recaptured if red cells extravasate into tissues, whereas bleeding into the gut or out of the body leads to iron loss and possible iron deficiency, which can hamper the restoration of normal red cell counts.

Significant bleeding results in predictable changes in the blood involving not only red cells, but also white cells and platelets. If the bleeding is sufficiently massive to cause a decrease in blood pressure, the compensatory release of adrenergic hormones mobilizes granulocytes from the intravascular marginal pool and results in leukocytosis. Initially, red cells appear normal in size and color (normocytic, normochromic). However, as marrow production increases, there is a striking increase in the reticulocyte count (reticulocytosis), which reaches 10% to 15% after 7 days. Reticulocytes are larger than normal red cells and have blue-red polychromatophilic cytoplasm due to the presence of RNA, a feature that allows them to be identified in the clinical laboratory. Early recovery from blood loss also is often accompanied by thrombocytosis, which results from an increase in platelet production.

Chronic Blood Loss

Chronic blood loss induces anemia only when the rate of loss exceeds the regenerative capacity of the marrow or when iron reserves are depleted and iron deficiency anemia appears (see later).

Hemolytic Anemias

Hemolytic anemias share the following features:

- *A shortened red cell life span* below the normal 120 days
- *Elevated erythropoietin levels and a compensatory increase in erythropoiesis*
- *Accumulation of hemoglobin degradation products* that are created as part of the process of red cell hemolysis

The physiologic destruction of senescent red cells takes place within macrophages, which are abundant in the spleen, liver, and bone marrow. This process appears to be triggered by **age-dependent changes** in red cell surface proteins, which lead to their recognition and removal by phagocytes. In the great majority of hemolytic anemias, the premature destruction of red cells also occurs within phagocytes, an event that is referred to as extravascular hemolysis. If persistent, extravascular hemolysis leads to a hyperplasia of phagocytes manifested by varying degrees of *splenomegaly*.

Extravascular hemolysis is most commonly caused by alterations that make red cells less deformable. Extreme changes in shape are required for red cells to navigate the splenic sinusoids successfully. Reduced deformability makes this passage difficult, leading to red cell sequestration and phagocytosis by macrophages located within the splenic cords. Regardless of the cause, the principal clinical features of extravascular hemolysis are anemia, splenomegaly, and jaundice. Some hemoglobin inevitably escapes from phagocytes, which leads to variable decreases in plasma haptoglobin, an α_2-globulin that binds free hemoglobin and prevents its excretion in the urine. Because much of the premature destruction of red cells occurs in the spleen, individuals with extravascular hemolysis often benefit from splenectomy.

Intravascular hemolysis of red cells may be caused by mechanical injury, complement fixation, intracellular parasites (e.g., falciparum malaria, Chapter 8), **or exogenous toxic factors.** Compared to extravascular hemolysis, it occurs less commonly; sources of mechanical injury include trauma caused by cardiac valves, narrowing of the microcirculation by thrombi, or repetitive physical trauma (e.g., marathon running and bongo drum beating). Complement fixation occurs in a variety of situations in which antibodies recognize and bind red cell antigens. Toxic injury is exemplified by clostridial sepsis, which results in the release of enzymes that digest the red cell membrane.

Whatever the mechanism, intravascular hemolysis is manifested by anemia, hemoglobinemia, hemoglobinuria, hemosiderinuria, and jaundice. Free hemoglobin released from lysed red cells is promptly bound by haptoglobin, producing a complex that is rapidly cleared by mononuclear phagocytes. As serum haptoglobin is depleted, free hemoglobin oxidizes to methemoglobin, which is brown in color. The renal proximal tubular cells reabsorb and break down much of the filtered hemoglobin and methemoglobin, but some passes out in the urine, imparting a red-brown color. Iron released from hemoglobin can accumulate within tubular cells, giving rise to *renal hemosiderosis*. Concomitantly, heme groups derived from hemoglobin-haptoglobin complexes are metabolized to bilirubin within mononuclear phagocytes, leading to jaundice. Unlike in extravascular hemolysis, splenomegaly is not seen.

In all types of uncomplicated hemolytic anemia, the excess serum bilirubin is unconjugated. The level of hyperbilirubinemia depends on the functional capacity of the liver and the rate of hemolysis. When the liver is normal, jaundice is rarely severe, but excessive bilirubin excreted by the liver into the biliary tract often leads to the formation of gallstones derived from heme pigments.

MORPHOLOGY

Certain changes are seen in hemolytic anemia regardless of cause or type. Anemia and lowered tissue oxygen tension trigger the production of erythropoietin, which stimulates erythroid differentiation and leads to the appearance of **increased numbers of erythroid precursors (normoblasts)** in the marrow (Fig. 14.1). Compensatory increases in erythropoiesis result in a **prominent reticulocytosis** in the peripheral blood. The phagocytosis of red cells leads to the accumulation of the iron-containing pigment **hemosiderin,** particularly in the spleen, liver, and bone marrow. Such iron accumulation is referred to as **hemosiderosis**. If the anemia is severe, **extramedullary hematopoiesis** can appear in the liver, spleen, and lymph nodes. With chronic hemolysis, elevated biliary excretion of bilirubin promotes the formation of **pigment gallstones** (cholelithiasis).

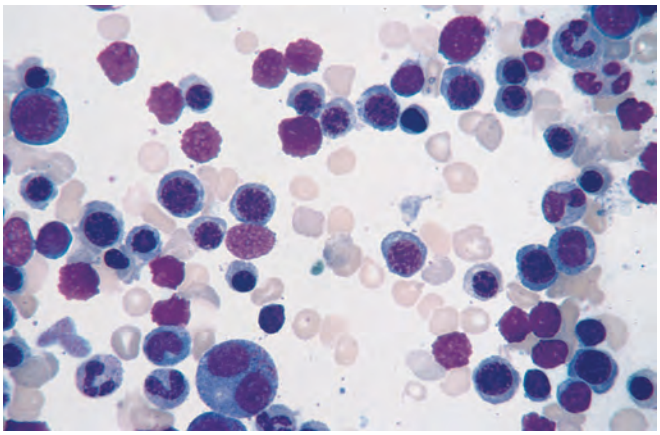


Figure 14.1 Marrow aspirate smear from a patient with hemolytic anemia. There is an increased number of maturing erythroid progenitors (normoblasts). (Courtesy Dr. Steven Kroft, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

The hemolytic anemias can be classified in a variety of ways; here, we rely on the underlying mechanisms (see Table 14.1). We begin by discussing the major inherited forms of hemolytic anemia, and then move on to the acquired forms that are most common or of particular pathophysiologic interest.

Hereditary Spherocytosis

Hereditary spherocytosis (HS) is an inherited disorder caused by intrinsic defects in the red cell membrane skeleton that render red cells spheroid, less deformable, and vulnerable to splenic sequestration and destruction. The prevalence of HS is highest in northern Europe, where rates of 1 in 5000 are reported. An autosomal dominant inheritance pattern is seen in about 75% of cases. The remaining patients have a more severe form of the disease that is usually caused by the inheritance of two different defects (a state known as compound heterozygosity).

Pathogenesis

The remarkable deformability and durability of the normal red cell are attributable to the physicochemical properties of its specialized membrane skeleton (Fig. 14.2), which lies closely apposed to the internal surface of the plasma membrane. Its chief protein component, spectrin, consists of two polypeptide chains, α and β, which form intertwined (helical) flexible heterodimers. The "head" regions of spectrin dimers self-associate to form tetramers, and the "tails" associate with actin oligomers. Each actin oligomer can bind multiple spectrin tetramers, thus creating a two-dimensional spectrin-actin skeleton that is connected to the cell membrane by two distinct interactions. The first, involving the proteins ankyrin and band 4.2, binds spectrin to the transmembrane ion transporter, band 3. The second, involving protein 4.1, binds the "tail" of spectrin to another transmembrane protein, glycophorin A.

HS is caused by diverse mutations that lead to an insufficiency of membrane skeletal components. As a result of these alterations, the life span of affected red cells is decreased on average to 10 to 20 days from the normal 120 days. The pathogenic mutations most commonly affect ankyrin, band 3, spectrin, or band 4.2, the proteins involved in one of the two tethering interactions. Most mutations cause frameshifts or introduce premature stop codons, such

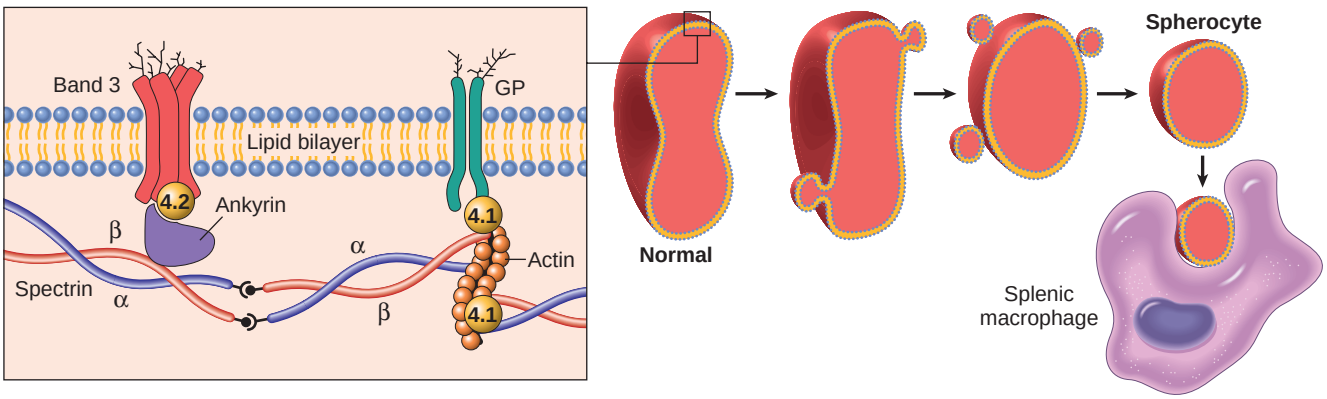


Figure 14.2 Role of the red cell membrane skeleton in hereditary spherocytosis. The *left panel* shows the normal organization of the major red cell membrane skeletal proteins. Various mutations involving α-spectrin, β-spectrin, ankyrin, band 4.2, or band 3 that weaken the interactions between these proteins cause red cells to lose membrane fragments as they age. To accommodate the resultant change in the ratio of surface area to volume, these cells adopt a spherical shape. Spherocytic cells are less deformable than normal ones and therefore become trapped in the splenic cords, where they are phagocytosed by macrophages. *GP,* Glycophorin.

that the mutated allele fails to produce any protein. The resulting deficiency of the affected protein reduces the assembly of the skeleton as a whole, destabilizing the overlying plasma membrane. Young HS red cells are normal in shape, but the destabilized lipid bilayer sheds membrane fragments as red cells age in the circulation. The loss of membrane relative to cytoplasm "forces" the cells to assume the smallest possible diameter for a given volume, namely, a sphere. Compound heterozygosity for two defective alleles understandably results in more profound membrane skeleton deficiency and more severe disease.

The invariably beneficial effects of splenectomy prove that the spleen has a cardinal role in the premature demise of spherocytes. The travails of spherocytic red cells are fairly well defined. In the life of the portly, inflexible spherocyte, the spleen is the villain. Normal red cells must undergo extreme deformation to leave the cords of Billroth and enter the sinusoids. Because of their spheroidal shape and reduced deformability, the hapless spherocytes are trapped in the splenic cords, where they are easy prey for macrophages. The splenic environment also exacerbates the tendency of HS red cells to lose membrane along with K^+ ions and H_2O; prolonged splenic exposure (erythrostasis), depletion of red cell glucose, and diminished red cell pH have all been suggested to contribute to these abnormalities (Fig. 14.3). After splenectomy the spherocytes persist, but the anemia is corrected.

MORPHOLOGY

The most specific morphologic finding is **spherocytosis,** apparent on smears as small, dark-staining (hyperchromic) red cells lacking the central zone of pallor (Fig. 14.4). Spherocytosis is distinctive but not pathognomonic, as spherocytes are also seen in other disorders associated with red cell membrane loss, such as in autoimmune hemolytic anemia. Other features are common to all hemolytic anemias. These include reticulocytosis, marrow erythroid hyperplasia, hemosiderosis, and mild jaundice. **Cholelithiasis** (pigment stones) occurs in 40% to 50% of affected adults. Moderate **splenomegaly** is characteristic (500 to 1000 g); in few other hemolytic anemias is the spleen enlarged as much or as consistently. Splenomegaly results from congestion of the cords of Billroth and increased numbers of phagocytes.

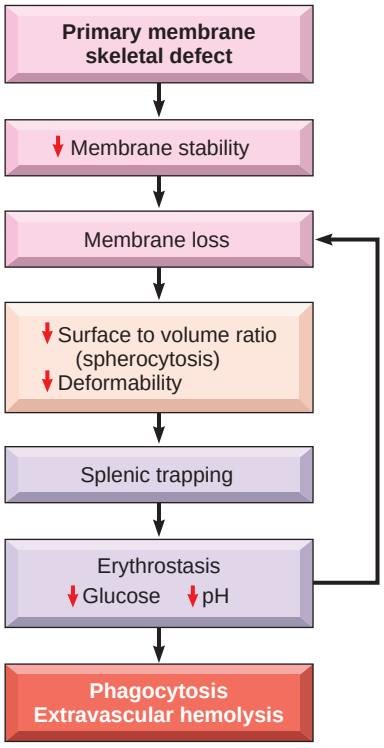


Figure 14.3 Pathophysiology of hereditary spherocytosis.

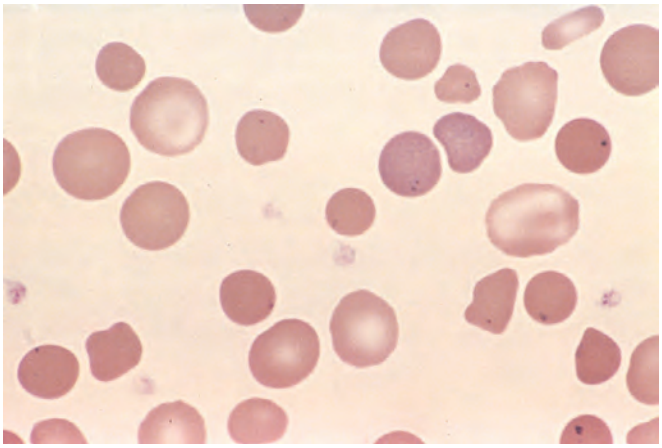


Figure 14.4 Hereditary spherocytosis (peripheral smear). Note the anisocytosis and several dark-appearing spherocytes with no central pallor. Howell-Jolly bodies (small, dark nuclear remnants) also are seen in some of the red cells of this asplenic patient. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Clinical Features

The diagnosis is based on family history, hematologic findings, and laboratory evidence. In two-thirds of cases, the red cells are abnormally sensitive to osmotic lysis when incubated in hypotonic salt solutions, which causes the influx of water into spherocytes with little margin for expansion. HS red cells also have an increased mean cell hemoglobin concentration, due to dehydration caused by the loss of K^+ and H_2O.

The characteristic clinical features are anemia, splenomegaly, and jaundice. The severity varies greatly. In a small minority (mainly compound heterozygotes), HS presents at birth with marked jaundice and requires exchange transfusions. In 20% to 30% of patients, the disease is so mild as to be virtually asymptomatic; here the decreased red cell survival is readily compensated for by increased erythropoiesis. In most, however, the compensatory changes are outpaced, producing a chronic hemolytic anemia of mild to moderate severity.

The generally stable clinical course is sometimes punctuated by *aplastic crises,* usually triggered by an acute parvovirus infection. Parvovirus infects and kills red cell progenitors, causing all red cell production to cease until an immune response clears the virus, generally in 1 to 2 weeks. Because of the reduced life span of HS red cells, cessation of erythropoiesis

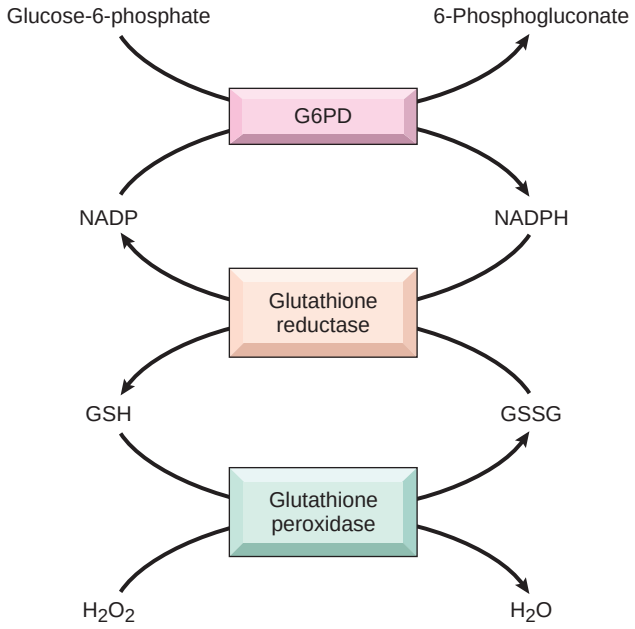


Figure 14.5 Role of glucose-6-phosphate dehydrogenase (G6PD) in defense against oxidant injury. Detoxification of H_2O_2, a potential oxidant, requires reduced glutathione (GSH), which is generated in a reaction that requires reduced nicotinamide adenine dinucleotide (NADPH). The synthesis of NADPH depends on the activity of G6PD. *GSSG*, Oxidized glutathione; *NADP*, nicotinamide adenine dinucleotide phosphate.

for even short periods leads to sudden worsening of the anemia. Transfusions may be necessary to support the patient during the acute phase of the infection. *Hemolytic crises* are produced by intercurrent events leading to increased splenic destruction of red cells (e.g., infectious mononucleosis and its attendant increase in spleen size); these are clinically less significant than aplastic crises. Gallstones, found in many patients, may also produce symptoms. Splenectomy treats the anemia and its complications, but brings with it an increased risk of sepsis because the spleen acts as an important filter for blood-borne bacteria.

Hemolytic Disease Due to Red Cell Enzyme Defects: Glucose-6-Phosphate Dehydrogenase Deficiency

Abnormalities in the hexose monophosphate shunt or glutathione metabolism resulting from deficient or impaired enzyme function reduce the ability of red cells to protect themselves against oxidative injuries and lead to hemolysis. The most important of these enzyme derangements is hereditary deficiency of glucose-6-phosphate dehydrogenase (G6PD) activity. G6PD reduces nicotinamide adenine dinucleotide phosphate (NADP) to NADPH while oxidizing glucose-6-phosphate (Fig. 14.5). NADPH then provides reducing equivalents needed for conversion of oxidized glutathione to reduced glutathione, which protects against oxidant injury by participating as a cofactor in reactions that neutralize compounds such as H_2O_2 (see Fig. 14.5).

G6PD deficiency is a recessive X-linked trait, placing males at much higher risk for symptomatic disease. Several hundred G6PD genetic variants exist, but most clinically significant hemolytic anemia is associated with only two variants, designated $G6PD^-$ and G6PD Mediterranean. $G6PD^-$ is present in about 10% of American blacks; G6PD Mediterranean, as the name implies, is prevalent in the Middle East. The high frequency of these variants in each population is believed to stem from a protective effect against *Plasmodium falciparum* malaria (discussed later). G6PD variants associated with hemolysis result in misfolding of the protein, making it more susceptible to proteolytic degradation. Compared with the most common normal variant, G6PD B, the half-life of $G6PD^-$ is moderately reduced, whereas that of G6PD Mediterranean is more markedly abnormal. Because mature red cells do not synthesize new proteins, as red cells age $G6PD^-$ and G6PD Mediterranean enzyme activities quickly fall to levels that are inadequate to protect against oxidant stress. Thus, older red cells are much more prone to hemolysis than younger ones.

The episodic hemolysis that is characteristic of G6PD deficiency is caused by exposures that generate oxidant stress. The most common triggers are infections, in which oxygen-derived free radicals are produced by activated leukocytes. Many infections can trigger hemolysis; viral hepatitis, pneumonia, and typhoid fever are among those most likely to do so. The other important initiators are drugs and certain foods. The drugs implicated are numerous, including antimalarials (e.g., primaquine and chloroquine), sulfonamides, nitrofurantoins, and others. Some drugs cause hemolysis only in individuals with the more severe Mediterranean variant. The most frequently cited food is the *fava bean,* which generates oxidants when metabolized. "Favism" is endemic in the Mediterranean, Middle East, and parts of Africa where consumption is prevalent. Uncommonly, G6PD deficiency presents as neonatal jaundice or a chronic low-grade hemolytic anemia in the absence of infection or known environmental triggers.

Oxidants cause both intravascular and extravascular hemolysis in G6PD-deficient individuals. Exposure of G6PD-deficient red cells to high levels of oxidants causes the cross-linking of reactive sulfhydryl groups on globin chains, which become denatured and form membrane-bound precipitates known as *Heinz bodies.* These are seen as dark inclusions within red cells stained with crystal violet (Fig. 14.6). Heinz bodies can damage the membrane sufficiently to cause intravascular hemolysis. Less severe membrane damage results in decreased red cell deformability. As inclusion-bearing red cells pass through the splenic cords, macrophages pluck out the Heinz bodies. As a result of membrane damage, some of these partially devoured cells retain an abnormal shape, appearing to have a bite taken out of them (see Fig. 14.6). Other less severely damaged cells become spherocytes due to loss of membrane surface area. Both bite cells and spherocytes are trapped in splenic cords and removed by phagocytes.

Acute intravascular hemolysis, marked by anemia, hemoglobinemia, and hemoglobinuria, usually begins 2 to 3 days following exposure of G6PD-deficient individuals to environmental triggers. Because only older red cells are at risk for lysis, the episode is self-limited, as hemolysis ceases when only younger G6PD-replete red cells remain (even if exposure to the trigger, e.g., an offending drug, continues). The recovery phase is heralded by reticulocytosis. Because hemolytic episodes related to G6PD deficiency occur intermittently, features related to chronic hemolysis (e.g., splenomegaly, cholelithiasis) are absent.

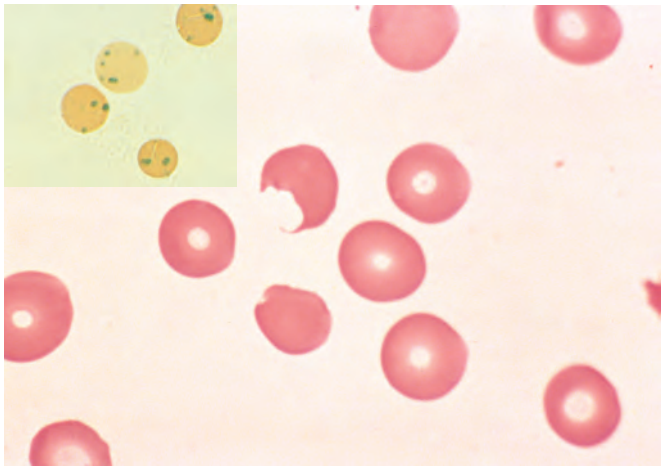


Figure 14.6 Glucose-6-phosphate dehydrogenase deficiency: effects of oxidant drug exposure (peripheral blood smear). *Inset,* Red cells with precipitates of denatured globin (Heinz bodies) revealed by supravital staining. As the splenic macrophages pluck out these inclusions, "bite cells" like the one in this smear are produced. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Sickle Cell Disease

Sickle cell disease is a common hereditary hemoglobinopathy caused by a point mutation in β-globin that promotes the polymerization of deoxygenated hemoglobin, leading to red cell distortion, hemolytic anemia, microvascular obstruction, and ischemic tissue damage. Several hundred hemoglobinopathies caused by various mutations in globin genes are known, but only those associated with sickle cell disease are prevalent enough in the United States to merit discussion. Hemoglobin (Hb) is a tetrameric protein composed of two pairs of globin chains, each with its own heme group. Normal adult red cells contain mainly HbA ($\alpha_2\beta_2$), along with small amounts of HbA_2 ($\alpha_2\delta_2$) and fetal hemoglobin (HbF; $\alpha_2\gamma_2$). Sickle cell disease is caused by a missense mutation in the β-globin gene that leads to the replacement of a charged glutamate residue with a hydrophobic valine residue. The abnormal physiochemical properties of the resulting sickle hemoglobin (HbS) are responsible for the disease.

About 8% to 10% of African Americans in the United States are heterozygous for HbS, a largely asymptomatic condition known as *sickle cell trait.* The offspring of two heterozygotes has a 1 in 4 chance of being homozygous for the sickle mutation, a state that produces symptomatic sickle cell disease, which afflicts 70,000 to 100,000 individuals in the United States. In affected individuals, almost all the hemoglobin in the red cell is HbS ($\alpha_2\beta^s_2$).

The high prevalence of sickle cell trait in certain African populations stems from its protective effects against falciparum malaria. Genetic studies have shown that the sickle hemoglobin mutation has arisen independently at least six times in areas of Africa in which falciparum malaria is endemic, providing clear evidence of strong Darwinian selection. Parasite densities are lower in infected, heterozygous HbAS children than in infected, normal HbAA children, and AS children are significantly less likely to have severe disease or to die from malaria. Although mechanistic details are lacking, two scenarios to explain these observations are favored:

- Metabolically active intracellular parasites consume oxygen and decrease intracellular pH, both of which promote sickling of AS red cells. These distorted, stiff cells may be cleared more rapidly by splenic and hepatic phagocytes, keeping parasite loads low.
- Sickling also impairs the formation of membrane knobs containing a protein made by the parasite called *PfEMP-1.* These membrane knobs are implicated in adhesion of infected red cells to endothelium, which is believed to have an important pathogenic role in the most severe form of the disease, cerebral malaria.

It has been suggested that G6PD deficiency and thalassemia also protect against malaria by increasing the clearance and decreasing the adherence of infected red cells, possibly by raising levels of oxidant stress and causing membrane damage in the parasite-bearing cells that leads to their rapid removal from the bloodstream.

Pathogenesis

The major pathologic manifestations—chronic hemolysis, microvascular occlusions, and tissue damage—all stem from the tendency of HbS molecules to stack into polymers when deoxygenated. Initially, this process converts the red cell cytosol from a freely flowing liquid into a viscous gel. With continued deoxygenation, HbS molecules assemble into long needlelike fibers within red cells, producing a distorted sickle or holly-leaf shape.

Several variables affect the rate and degree of sickling:

- *Interaction of HbS with the other types of hemoglobin.* In heterozygotes with sickle cell trait, about 40% of the hemoglobin is HbS and the rest is HbA, which interferes with HbS polymerization. As a result, red cells in heterozygous individuals only sickle if exposed to prolonged, relatively severe hypoxia. HbF inhibits the polymerization of HbS even more than HbA; hence, infants with sickle cell disease do not become symptomatic until they reach 5 or 6 months of age, when the level of HbF normally falls. However, in some individuals HbF expression remains relatively high, a condition known as *hereditary persistence of fetal hemoglobin;* in these individuals, sickle cell disease is much less severe. Another variant hemoglobin, HbC, also is common in regions where HbS is found; overall, about 2% to 3% of American blacks are HbC heterozygotes, and about 1 in 1250 are compound HbS/HbC heterozygotes. In HbSC red cells, the percentage of HbS is 50%, as compared with only 40% in HbAS cells. Moreover, with aging HbSC cells tend to lose salt and water and become dehydrated, an effect that increases the intracellular concentration of HbS. These factors increase the tendency for HbS to polymerize, and as a result compound HbSC heterozygotes have a symptomatic sickling disorder termed *HbSC disease* that is somewhat milder than sickle cell disease.
- *Mean cell hemoglobin concentration* (MCHC). Higher HbS concentrations increase the probability that aggregation and polymerization will occur during any given period of deoxygenation. Thus, intracellular dehydration, which increases the MCHC, facilitates sickling. Conversely, conditions that decrease the MCHC reduce disease

severity. This occurs when an individual who is homozygous for HbS also has coexistent α-thalassemia, which reduces Hb synthesis and leads to milder disease.

- *Intracellular pH.* A decrease in pH reduces the oxygen affinity of hemoglobin, thereby increasing the fraction of deoxygenated HbS at any given oxygen tension and augmenting the tendency for sickling.
- *Transit time of red cells through microvascular beds.* As will be discussed, much of the pathology of sickle cell disease is related to vascular occlusion caused by sickling within microvascular beds. Transit times in most normal microvascular beds are too short for significant aggregation of deoxygenated HbS to occur, and as a result sickling is confined to microvascular beds with slow transit times. Blood flow is sluggish in the normal spleen and bone marrow, which are prominently affected in sickle cell disease, and also in vascular beds that are inflamed. The movement of blood through inflamed tissues is slowed because of the adhesion of leukocytes to activated endothelial cells and the transudation of fluid through leaky vessels. As a result, inflamed vascular beds are prone to sickling and occlusion.

Sickling causes cumulative damage to red cells through several mechanisms. As HbS polymers grow, they herniate through the membrane skeleton and project from the cell ensheathed only by the lipid bilayer. This severe derangement in membrane structure causes an influx of Ca^{2+} ions, which induce the cross-linking of membrane proteins and activate an ion channel that leads to the efflux of K^+ and H_2O. As a result, with repeated sickling episodes, red cells become dehydrated, dense, and rigid (Fig. 14.7). Eventually, the most severely damaged cells are converted to nondeformable irreversibly sickled cells that retain a sickle shape, even when fully oxygenated. The severity of the hemolysis correlates with the percentage of irreversibly sickled cells, which are rapidly sequestered and removed by mononuclear phagocytes (extravascular hemolysis). Sickled red cells are also mechanically fragile, leading to some intravascular hemolysis as well.

The pathogenesis of the microvascular occlusions, which are responsible for the most serious clinical features, is far less certain. Microvascular occlusions are not related to the number of irreversibly sickled cells, but instead may be dependent on more subtle red cell membrane damage and local factors, such as inflammation or vasoconstriction, that tend to slow or arrest the movement of red cells through microvascular beds (see Fig. 14.7). As mentioned earlier, sickle red cells express higher than normal levels of adhesion molecules and are sticky. Mediators released from granulocytes during inflammatory reactions up-regulate the expression of adhesion molecules on endothelial cells (Chapter 3) and further enhance the tendency for sickle red cells to arrest during transit through the microvasculature. The stagnation of red cells within inflamed vascular beds results in extended exposure to low oxygen tension, sickling, and vascular obstruction. Once started, it is easy to envision how a vicious cycle of sickling, obstruction, hypoxia, and more sickling ensues. Depletion of nitric oxide (NO) also may play a part in the vascular occlusions. Free hemoglobin released from lysed sickle red cells can bind and inactivate NO, a potent vasodilator and inhibitor of platelet aggregation. This in turn may lead to increased vascular tone (narrowing of vessels) and enhanced platelet aggregation, both of which may contribute to red cell stasis, sickling, and (in some instances) thrombosis.

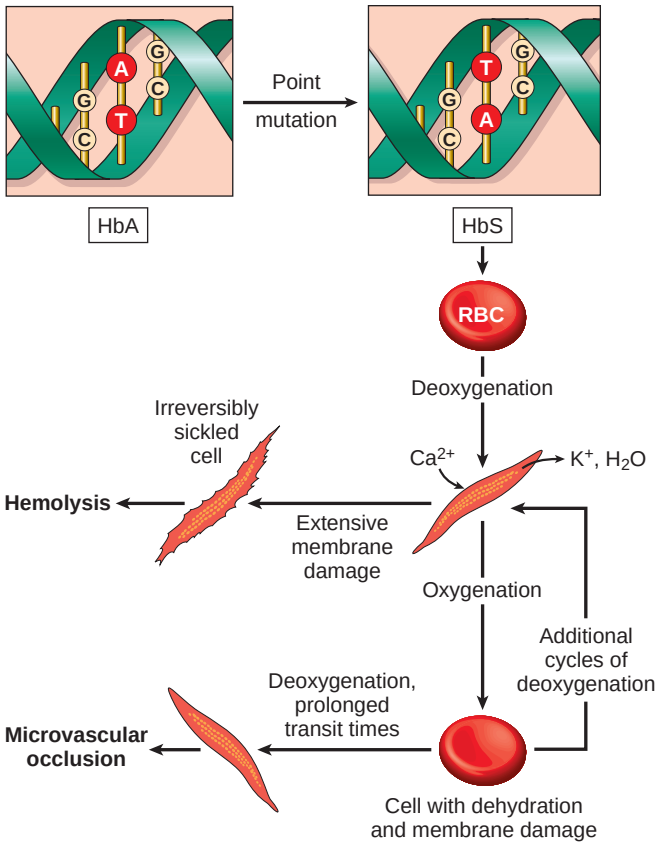


Figure 14.7 Pathophysiology of sickle cell disease. *HbA*, Hemoglobin A; *HbS*, hemoglobin S; *RBC*, red blood cell.

MORPHOLOGY

In sickle cell anemia, the peripheral blood demonstrates variable numbers of **irreversibly sickled cells,** reticulocytosis, and target cells, which result from red cell dehydration (Fig. 14.8). **Howell-Jolly bodies** (small nuclear remnants) also are present in red cells due to asplenia (see later). The bone marrow is hyperplastic as a result of a compensatory erythroid hyperplasia. Marked expansion of the marrow leads to bone resorption and secondary new bone formation, producing prominent cheekbones and changes in the skull that resemble a "crewcut" on radiographic studies. Extramedullary hematopoiesis may also appear. The increased breakdown of hemoglobin may cause hyperbilirubinemia and formation of pigment gallstones.

In early childhood, the spleen is enlarged (up to 500 g) by red pulp congestion caused by the trapping of sickled red cells in the cords and sinuses (Fig. 14.9). With time, however, chronic erythrostasis leads to splenic infarction, fibrosis, and progressive shrinkage, so that by adolescence or early adulthood only a small nubbin of fibrous splenic tissue is left, a process called **autosplenectomy** (Fig. 14.10). Infarctions caused by vascular occlusions may occur in many other tissues as well, including the bones, brain, kidney, liver, retina, and pulmonary vessels, the latter sometimes producing cor pulmonale. In adult patients, vascular stagnation in subcutaneous tissues often leads to leg ulcers; this complication is rare in children.

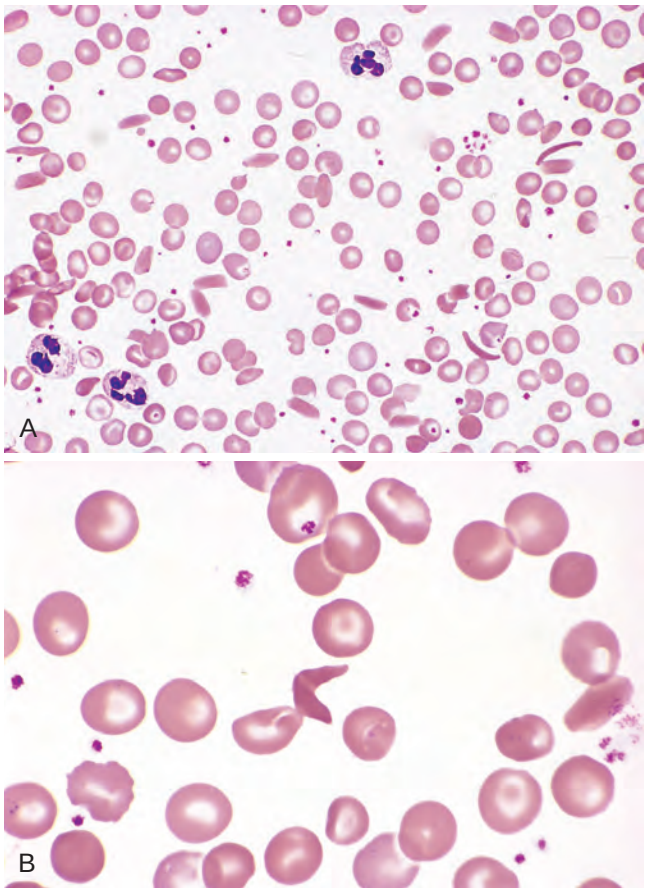


Figure 14.8 Sickle cell disease (peripheral blood smear). (A) Low magnification shows irreversibly sickled cells as well as target cells and red cell anisocytosis and poikilocytosis. (B) Higher magnification shows an irreversibly sickled cell in the center. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

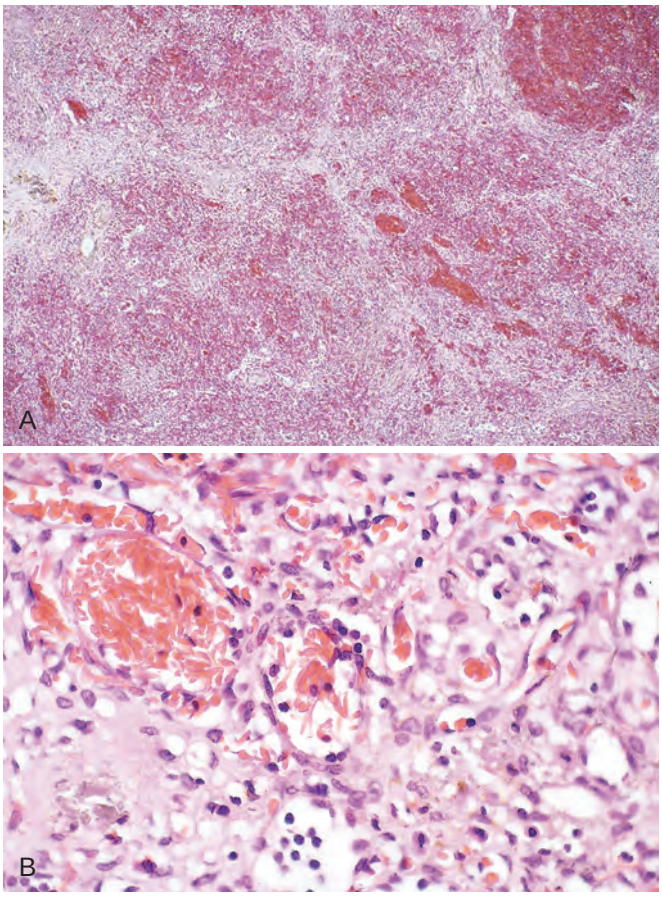


Figure 14.9 (A) Spleen in sickle cell disease (low power). Red pulp cords and sinusoids are markedly congested; between the congested areas, pale areas of fibrosis resulting from ischemic damage are evident. (B) Under high power, splenic sinusoids are dilated and filled with sickled red cells. (Courtesy Dr. Darren Wirthwein, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Clinical Features

Sickle cell disease causes a moderately severe hemolytic anemia (hematocrit 18% to 30%) associated with reticulocytosis, hyperbilirubinemia, and the presence of irreversibly sickled cells. Its course is punctuated by a variety of "crises." *Vaso-occlusive crises,* also called *pain crises,* are episodes of hypoxic injury and infarction that cause severe pain in the affected region. Although infection, dehydration, and acidosis (all of which favor sickling) may act as triggers, in most instances no predisposing cause is identified. The most commonly involved sites are the bones, lungs, liver, brain, spleen, and penis. In children, painful bone crises are extremely common and often difficult to distinguish from acute osteomyelitis. These frequently manifest as the *hand-foot syndrome* or dactylitis of the bones of the hands and feet. *Acute chest syndrome* is a particularly dangerous type of vaso-occlusive crisis involving the lungs that typically presents with fever, cough, chest pain, and pulmonary infiltrates. Pulmonary inflammation (such as may be induced by an infection) may cause blood flow to become sluggish and "spleenlike," leading to sickling and vaso-occlusion. This compromises pulmonary function, creating a potentially fatal cycle of worsening pulmonary and systemic hypoxemia, sickling, and vaso-occlusion. *Priapism* affects up to 45% of males after puberty and may lead to hypoxic damage and erectile dysfunction. Other disorders related to vascular obstruction, particularly stroke and retinopathy leading to loss of visual acuity and even blindness, can take a devastating toll. Factors proposed to contribute to stroke include the adhesion of sickle red cells

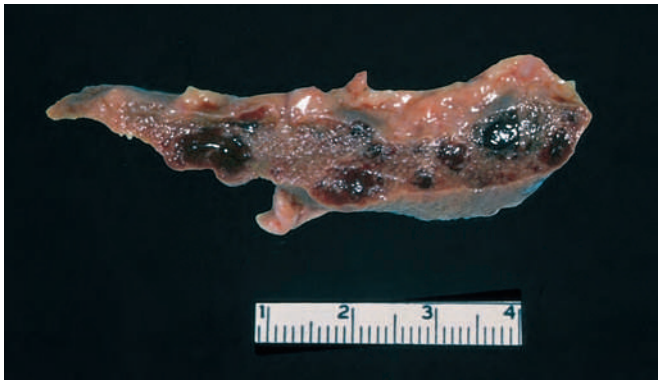


Figure 14.10 "Autoinfarcted" splenic remnant in sickle cell disease. (Courtesy Drs. Dennis Burns and Darren Wirthwein, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

to arterial vascular endothelium and vasoconstriction caused by the depletion of NO by free hemoglobin.

Although occlusive crises are the most common cause of patient morbidity and mortality, several other acute events complicate the course. *Sequestration crises* occur in children with intact spleens. Massive entrapment of sickled red cells leads to rapid splenic enlargement, hypovolemia, and sometimes shock. Both sequestration crises and the acute chest syndrome may be fatal and sometimes require prompt treatment with exchange transfusions. *Aplastic crises* stem from the infection of red cell progenitors by parvovirus B19, which causes a transient cessation of erythropoiesis and a sudden worsening of the anemia.

In addition to these dramatic crises, chronic tissue hypoxia takes a subtle but important toll. Chronic hypoxia is responsible for a generalized impairment of growth and development, as well as organ damage affecting the spleen, heart, kidneys, and lungs. Sickling provoked by hypertonicity in the renal medulla causes damage that eventually leads to *hyposthenuria* (the inability to concentrate urine), which increases the propensity for dehydration and its attendant risks.

Increased susceptibility to infection with encapsulated organisms is another threat. This is due in large part to altered splenic function, which is severely impaired in children by congestion and poor blood flow, and completely absent in adults because of splenic infarction. Defects of uncertain etiology in the alternative complement pathway also impair the opsonization of bacteria. *Pneumococcus pneumoniae* and *Haemophilus influenzae* septicemia and meningitis are common, particularly in children, but can be reduced by vaccination and prophylactic antibiotics.

It must be emphasized that there is great variation in the clinical manifestations of sickle cell disease. Some individuals suffer repeated vaso-occlusive crises, whereas others have only mild symptoms. The basis for this wide range in disease expression is not understood; both modifying genes and environmental factors are suspected.

The diagnosis is suggested by the clinical findings and the presence of irreversibly sickled red cells and is confirmed by various tests for sickle hemoglobin. Prenatal diagnosis is possible by analysis of fetal DNA obtained by amniocentesis or chorionic biopsy. Newborn screening for sickle hemoglobin is now routinely performed in all 50 states, typically using samples obtained by heel stick at birth.

The outlook for patients with sickle cell disease has improved considerably over the past 10 to 20 years. About 90% of patients survive to 20 years of age, and close to 50% survive beyond the fifth decade. The mainstay of treatment is an inhibitor of DNA synthesis, hydroxyurea, which has several beneficial effects. These include (1) an increase in red cell HbF levels, which occurs by unknown mechanisms; and (2) an anti-inflammatory effect, which stems from an inhibition of leukocyte production. These activities (and possibly others) are believed to act in concert to decrease crises related to vascular occlusions in both children and adults. When added to hydroxyurea, L-glutamine has been shown to decrease pain crises; the mechanism is uncertain, but it may involve changes in metabolism that decrease oxidant stress in red cells. Hematopoietic stem cell transplantation offers a chance at cure and is increasingly being explored as a therapeutic option. Another exciting new approach involves using gene editing (CRISPR technology) to reverse hemoglobin switching, so that hematopoietic stem cells produce red cells that express fetal hemoglobin instead of sickle hemoglobin. A clinical trial testing this approach is ongoing and has produced excellent responses.

Thalassemia

Thalassemia is a genetically heterogeneous disorder caused by germline mutations that decrease the synthesis of either α-globin or β-globin, leading to anemia, tissue hypoxia, and red cell hemolysis related to the imbalance in globin chain synthesis. The two α chains in HbA are encoded by an identical pair of α-globin genes on chromosome 16, and the two β chains are encoded by a single β-globin gene on chromosome 11. β-thalassemia is caused by deficient synthesis of β chains, whereas α-thalassemia is caused by deficient synthesis of α chains. The hematologic consequences of diminished synthesis of one globin chain stem not only from hemoglobin deficiency but also from a relative excess of the other globin chain, particularly in β-thalassemia (described later).

Thalassemia is endemic in the Mediterranean basin (indeed, *thalassa* means "sea" in Greek) as well as the Middle East, tropical Africa, the Indian subcontinent, and Asia, and in aggregate is among the most common inherited disorders of humans. As with sickle cell disease and other common inherited red cell disorders, its prevalence seems to be explained by the protection it affords heterozygous carriers against malaria. Although we discuss thalassemia with other inherited forms of anemia associated with hemolysis, it is important to recognize that the defects in globin synthesis that underlie these disorders cause anemia through two mechanisms: decreased red cell production, and decreased red cell lifespan.

β-Thalassemia

β-thalassemia is caused by mutations that diminish the synthesis of β-globin chains. Its clinical severity varies widely due to heterogeneity in the causative mutations. We will begin our discussion with the molecular lesions in β-thalassemia and then relate the clinical variants to specific underlying molecular defects.

Molecular Pathogenesis

The causative mutations fall into two categories: (1) β^0 mutations, associated with absent β-globin synthesis, and (2) β^+ mutations, characterized by reduced (but detectable) β-globin synthesis. Sequencing of β-thalassemia genes has revealed more than 100 different causative mutations, mostly consisting of point mutations, which fall into three major classes:

- *Splicing mutations.* These are the most common cause of β^+-thalassemia. Some of these mutations destroy normal RNA splice junctions and completely prevent the production of normal β-globin mRNA, resulting in β^0-thalassemia. Others create an "ectopic" splice site within an intron. Because the flanking normal splice site remains, both normal and abnormal splicing occurs and some normal β-globin mRNA is made, resulting in β^+-thalassemia.
- *Promoter region mutations.* These mutations reduce transcription by 75% to 80%. Some normal β-globin is synthesized; thus, these mutations are associated with β^+-thalassemia.

- *Chain terminator mutations.* These are the most common cause of β^0-thalassemia. They consist of either nonsense mutations that introduce a premature stop codon or small insertions or deletions that shift the mRNA reading frames (frameshift mutations; Chapter 5). Both block translation and prevent the synthesis of any functional β-globin.

Impaired β-globin synthesis results in anemia by two mechanisms (Fig. 14.11). The deficit in HbA synthesis produces "underhemoglobinized" hypochromic, microcytic red cells with subnormal oxygen transport capacity. Even more important is the diminished survival of red cells and their precursors, which results from the imbalance in α- and β-globin synthesis. Unpaired α chains precipitate within red cell precursors, forming insoluble inclusions. These inclusions cause a variety of untoward effects, but membrane damage is the proximal cause of most red cell pathology. Many red cell precursors succumb to membrane damage and undergo apoptosis. In severe β-thalassemia, it is estimated that 70% to 85% of red cell precursors suffer this fate, which leads to *ineffective erythropoiesis*. Those red cells that are released from the marrow also contain inclusions and have membrane damage, leaving theme prone to splenic sequestration and extravascular hemolysis.

In severe β-thalassemia, ineffective erythropoiesis creates several additional problems. Erythropoietic drive in the setting of severe uncompensated anemia leads to massive erythroid hyperplasia in the marrow and extensive extramedullary hematopoiesis. The expanding mass of red cell precursors erodes the bony cortex, impairs bone growth, and produces skeletal abnormalities (described later). Extramedullary hematopoiesis involves the liver, spleen, and lymph nodes, and in extreme cases produces extraosseous masses in the thorax, abdomen, and pelvis. The metabolically active erythroid progenitors steal nutrients from other tissues that are already oxygen-starved, causing severe cachexia in untreated patients.

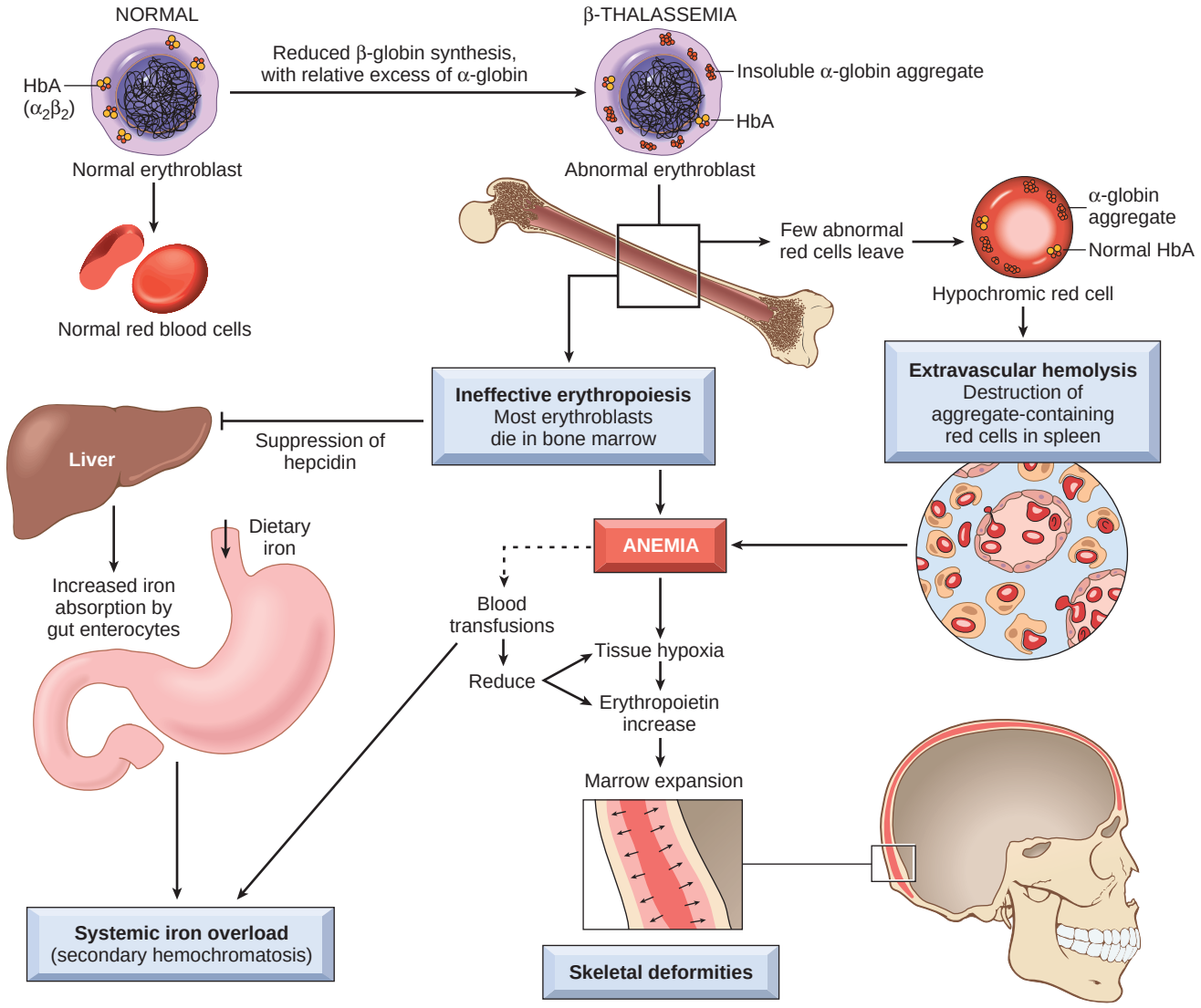


Figure 14.11 Pathogenesis of β-thalassemia major. Note that the aggregates of unpaired α-globin chains, a hallmark of the disease, are not visible in routinely stained blood smears. Blood transfusions are a double-edged sword, diminishing the anemia and its attendant complications, but also adding to the systemic iron overload. *HbA*, Hemoglobin A.

Another serious complication of ineffective erythropoiesis is excessive absorption of dietary iron. Erythroid precursors secrete a hormone called erythroferrone that inhibits production of hepcidin, a key negative regulator of iron uptake in the gut (described later in this chapter). In thalessemia, the marked expansion of erythroid precursors leads to increased absorption of iron from the gut (Fig. 14.12), and this together with repeated blood transfusions inevitably lead to severe iron accumulation *(secondary hemochromatosis)* unless preventive steps are taken. Injury to parenchymal organs, particularly the heart and liver, often follows (Chapter 18).

Clinical Syndromes

The relationships of clinical phenotypes to underlying genotypes are summarized in Table 14.3. Clinical classification of β-thalassemia is based on the severity of the anemia, which in turn depends on the genetic defect (β^+ or β^0) and the gene dosage (homozygous or heterozygous). In general, individuals with two β-thalassemia alleles (β^+/β^+, β^+/β^0, or β^0/β^0) have a severe, transfusion-dependent anemia called β-thalassemia major. Heterozygotes with one β-thalassemia gene and one normal gene (β^+/β or β^0/β) usually have a mild asymptomatic microcytic anemia. This condition is referred to as β-thalassemia minor or β-thalassemia trait. A third genetically heterogeneous variant of moderate severity is called β-thalassemia intermedia. This category includes milder variants of β^+/β^+ or β^+/β^0-thalassemia and unusual forms of heterozygous β-thalassemia. Some patients with β-thalassemia intermedia have two defective β-globin genes and an α-thalassemia gene defect, which improves the effectiveness of erythropoiesis and red cell survival by lessening the imbalance in α- and β-chain synthesis. In other rare but informative cases, affected individuals have a single β-globin defect and one or two extra copies of normal α-globin genes (stemming from a gene duplication event), which worsens the chain imbalance. These unusual forms of the disease emphasize the cardinal role of unpaired α-globin chains in the pathology. The clinical and morphologic features of β-thalassemia intermedia are not described separately but can be surmised from the following discussions of β-thalassemia major and β-thalassemia minor.

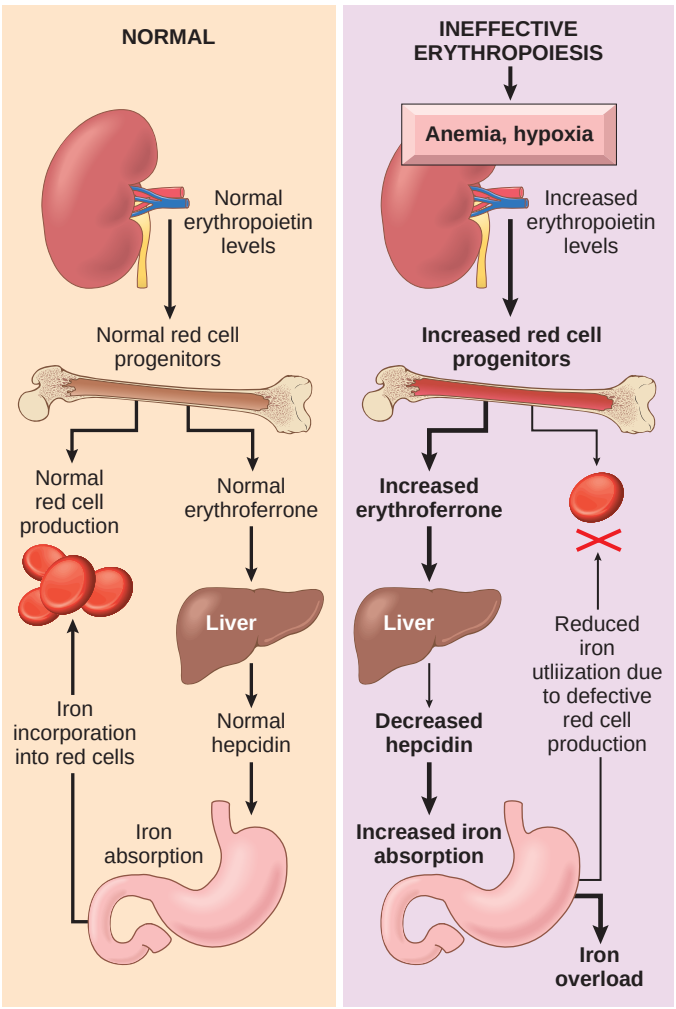


Figure 14.12 Mechanism of iron overload due to ineffective hematopoiesis. In the setting of ineffective erythropoiesis, such as in those with severed thalassemia, increased release of erythroferrone from the expanded mass of erythroid progenitors suppresses hepcidin production, leading to increased iron uptake from the gut.

Table 14.3 Clinical and Genetic Classification of Thalassemia

Clinical Syndrome	Genotype	Clinical Features	Molecular Genetics
β-Thalassemia			
β-Thalassemia major	Homozygous β-thalassemia (β^0/β^0, β^+/β^+, β^0/β^+)	Severe; requires blood transfusions	Mainly point mutations that lead to defects in the transcription, splicing, or translation of β-globin mRNA
β-Thalassemia intermedia	Variable (β^0/β^+, β^+/β^+, β^0/β, β^+/β)	Severe but does not require regular blood transfusions	
β-Thalassemia minor	Heterozygous β-thalassemia (β^0/β, β^+/β)	Asymptomatic with mild or absent anemia; red cell abnormalities seen	
α-Thalassemia			
Silent carrier	$-/\alpha\ \alpha/\alpha$	Asymptomatic; no red cell abnormality	Mainly gene deletions
α-Thalassemia trait	$-/-\ \alpha/\alpha$ (Asian) $-/\alpha\ -/\alpha$ (black African, Asian)	Asymptomatic, like β-thalassemia minor	
HbH disease	$-/-\ -/\alpha$	Severe; resembles β-thalassemia intermedia	
Hydrops fetalis	$-/-\ -/-$	Lethal in utero without transfusions	

β-Thalassemia Major. β-Thalassemia major is most common in Mediterranean countries, parts of Africa, and Southeast Asia. In the United States, the incidence is highest in immigrants from these areas. The anemia manifests 6 to 9 months after birth as hemoglobin synthesis switches from HbF to HbA. In untransfused patients, hemoglobin levels are 3 to 6 g/dL. The red cells may completely lack HbA (β^0/β^0 genotype) or contain small amounts (β^+/β^+ or β^0/β^+ genotypes). The major red cell hemoglobin is HbF, which is markedly elevated. HbA_2 levels are sometimes high but more often are normal or low.

MORPHOLOGY

Blood smears show severe red cell abnormalities, including marked variation in size **(anisocytosis)** and shape **(poikilocytosis), microcytosis,** and **hypochromia.** Target cells (so called because hemoglobin collects in the center of the cell), basophilic stippling, and fragmented red cells also are common. Inclusions of aggregated α chains are efficiently removed by the spleen and not easily seen. The reticulocyte count is elevated, but is lower than expected for the severity of anemia because of ineffective erythropoiesis. Variable numbers of poorly hemoglobinized nucleated red cell precursors (normoblasts) are seen in the peripheral blood as a result of "stress" erythropoiesis and abnormal release of red cell precursors from sites of extramedullary hematopoiesis.

Other major alterations involve the bone marrow and spleen. In untransfused patients, there is a striking expansion of hematopoietically active marrow. In the bones of the face and skull, the burgeoning marrow erodes existing cortical bone and induces new bone formation, giving rise to a "crewcut" appearance on radiographic studies (Fig. 14.13). Both phagocyte hyperplasia and extramedullary hematopoiesis contribute to enlargement of the spleen, which can weigh as much as 1500 g. The liver and the lymph nodes also may be enlarged by extramedullary hematopoiesis.

Hemosiderosis and secondary hemochromatosis, the two manifestations of iron overload (Chapter 18), inevitably occur unless chelation therapy is given. The deposited iron often damages organs, most notably the heart, liver, and pancreas.

The clinical course of β-thalassemia major is brief unless blood transfusions are given. Untreated children suffer from growth retardation and die at an early age from the effects of anemia. In those who survive long enough, the cheekbones and other bony prominences are enlarged and distorted. Hepatosplenomegaly due to extramedullary hematopoiesis is usually present. Although blood transfusions improve the anemia and suppress complications related to excessive erythropoiesis, they lead to complications of their own. Cardiac disease resulting from progressive iron overload and secondary hemochromatosis (Chapter 18) is an important cause of death, particularly in heavily transfused patients, who must be treated with iron chelators to prevent this complication. With transfusions and iron chelation, survival into the third decade is possible, but the overall outlook remains guarded. Hematopoietic stem cell transplantation is the only therapy offering a cure and is being used increasingly. Prenatal diagnosis is possible by molecular analysis of DNA.

β-Thalassemia Minor. β-Thalassemia minor is much more common than β-thalassemia major and understandably affects the same ethnic groups. Most patients are heterozygous carriers of a β^+ or β^0 allele. These patients are usually asymptomatic. Anemia, if present, is mild. The peripheral blood smear typically shows hypochromia, microcytosis, basophilic stippling, and target cells. Mild erythroid hyperplasia is seen in the bone marrow. Hemoglobin electrophoresis usually reveals an increase in HbA_2 ($\alpha_2\delta_2$) to 4% to 8% of the total hemoglobin (normal, 2.5% ± 0.3%), reflecting an elevated ratio of δ-chain to β-chain synthesis. HbF levels are generally normal or occasionally slightly increased.

Recognition of β-thalassemia trait is important for two reasons: (1) it may be mistaken for iron deficiency, and (2) it has implications for genetic counseling. Iron deficiency (the most common cause of microcytic anemia) can usually be excluded by measurement of serum iron, total iron-binding capacity, and serum ferritin (see the Iron Deficiency Anemia section later in this chapter). The increase in HbA_2 is diagnostically useful, particularly in individuals (such as women of childbearing age) who are at high risk of iron deficiency.

α-Thalassemia

α-Thalassemia is caused by inherited deletions that result in reduced or absent synthesis of α-globin chains. Normal individuals have four α-globin genes, and the severity of α-thalassemia depends on how many α-globin genes are affected. As in β-thalassemias, the anemia stems both from inadequate hemoglobin synthesis and the presence of excess, unpaired β, γ, and δ globin chains, which vary in type at different ages. In newborns with α-thalassemia, excess unpaired γ-globin chains form γ_4 tetramers known as *hemoglobin Barts,* whereas in older children and adults excess β-globin chains form β_4 tetramers known as *HbH.* Because free β and γ chains are more soluble than free α chains and form fairly stable homotetramers, hemolysis and ineffective erythropoiesis are less severe than in β-thalassemia. A variety of molecular lesions give rise to α-thalassemia, but gene deletion is the most common cause of reduced α-chain synthesis.

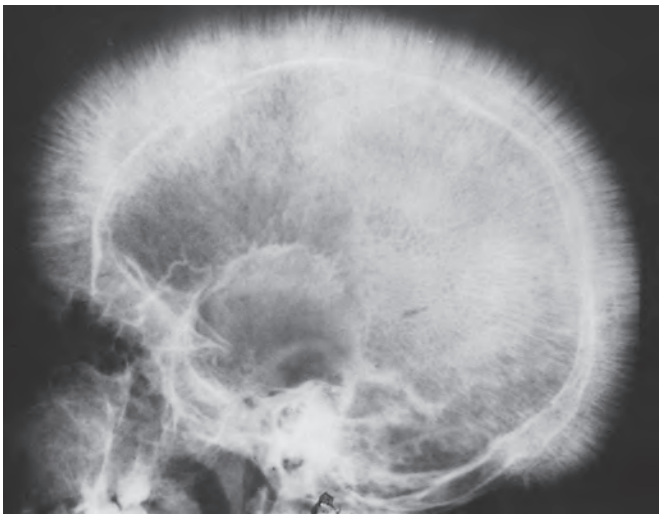


Figure 14.13 β-Thalassemia major. X-ray film of the skull showing new bone formation on the outer table, producing perpendicular radiations resembling a crewcut. (Courtesy Dr. Jack Reynolds, Department of Radiology, University of Texas Southwestern Medical School, Dallas, Tex.)

Clinical Syndromes

The clinical syndromes are determined and classified by the number of α-globin genes that are deleted. Each of the four α-globin genes normally contributes 25% of the total α-globin chains. α-Thalassemia syndromes stem from combinations of deletions that remove one to four α-globin genes. Not surprisingly, the severity of the clinical syndrome is proportional to the number of α-globin genes that are deleted. The different types of α-thalassemia and their salient clinical features are listed in Table 14.3.

Silent Carrier State. Silent carrier state is associated with the deletion of a single α-globin gene, which causes a barely detectable reduction in α-globin chain synthesis. These individuals are completely asymptomatic but have slight microcytosis.

α-Thalassemia Trait. α-Thalassemia trait is caused by the deletion of two α-globin genes from a single chromosome (α/α −/−) or the deletion of one α-globin gene from each of the two chromosomes (α/− α/−) (see Table 14.3). The former genotype is more common in Asian populations, the latter in regions of Africa. Both genotypes produce similar deficiencies of α-globin, but they have different implications for the children of affected individuals, who are at risk of clinically significant α-thalassemia (HbH disease or hydrops fetalis) only when at least one parent has the −/− haplotype. As a result, symptomatic α-thalassemia is relatively common in Asian populations and rare in African populations. The clinical picture in α-thalassemia trait is identical to that described for β-thalassemia minor, that is, small red cells (microcytosis), minimal or no anemia, and no abnormal physical signs. HbA_2 levels are normal or low.

Hemoglobin H (HbH) Disease. HbH disease is caused by deletion of three α-globin genes. It is most common in Asian populations. With only one normal α-globin gene, the synthesis of α chains is markedly reduced, and tetramers of β-globin, called HbH, form. HbH has an extremely high affinity for oxygen and therefore is not useful for oxygen delivery, leading to tissue hypoxia disproportionate to the level of hemoglobin. Additionally, HbH is prone to oxidation, which causes it to precipitate and form intracellular inclusions that promote red cell sequestration and phagocytosis in the spleen. The result is a moderately severe anemia resembling β-thalassemia intermedia.

Hydrops Fetalis. Hydrops fetalis, the most severe form of α-thalassemia, is caused by deletion of all four α-globin genes. In the fetus, excess γ-globin chains form tetramers (hemoglobin Barts) that have such a high affinity for oxygen that they deliver little to tissues. Survival in early development is due to the expression of ζ chains, an embryonic globin that pairs with γ chains to form a functional $\zeta_2\gamma_2$ Hb tetramer. Signs of fetal distress usually become evident by the third trimester of pregnancy. In the past, severe tissue anoxia led to death in utero or shortly after birth; with intrauterine transfusion many affected infants are now saved. The fetus shows severe pallor, generalized edema, and massive hepatosplenomegaly similar to that seen in hemolytic disease of the newborn (Chapter 10). There is a lifelong dependence on blood transfusions for survival, with the associated risk of iron overload. Hematopoietic stem cell transplantation can be curative.

Paroxysmal Nocturnal Hemoglobinuria

Paroxysmal nocturnal hemoglobinuria (PNH) is a disease that results from acquired mutations in the phosphatidylinositol glycan complementation group A gene *(PIGA)*, an enzyme that is essential for the synthesis of certain membrane-associated complement regulatory proteins. PNH has an incidence of 2 to 5 per million in the United States. Despite its rarity, it has fascinated hematologists because it is the only hemolytic anemia caused by an acquired genetic defect. Recall that proteins are anchored into the lipid bilayer in two ways. Most have a hydrophobic region that spans the cell membrane; these are called transmembrane proteins. The others are attached to the cell membrane through a covalent linkage to a specialized phospholipid called glycosylphosphatidylinositol (GPI). In PNH, these GPI-linked proteins are deficient because of somatic mutations that inactivate PIGA. *PIGA* is X-linked and subject to lyonization (random inactivation of one X chromosome in cells of females; Chapter 5). As a result, a single acquired mutation in the active *PIGA* gene of any given cell is sufficient to produce a deficiency state. Because the causative mutations occur in a hematopoietic stem cell, all of its clonal progeny (red cells, white cells, and platelets) are deficient in GPI-linked proteins. Typically, only a subset of stem cells acquires the mutation, and the mutant clone coexists with the progeny of normal stem cells that are not PIGA deficient.

Remarkably, most normal individuals harbor small numbers of bone marrow cells with *PIGA* mutations identical to those that cause PNH. It is hypothesized that these cells increase in numbers (thus producing clinically evident PNH) only in rare instances where they have a selective advantage, such as in the setting of autoimmune reactions against GPI-linked antigens. Such a scenario might explain the frequent association of PNH and aplastic anemia, a marrow failure syndrome (discussed later) that has an autoimmune basis in many individuals.

PNH blood cells are deficient in three GPI-linked proteins that regulate complement activity: (1) decay-accelerating factor, or CD55; (2) membrane inhibitor of reactive lysis, or CD59; and (3) C8-binding protein. Of these factors, the most important is CD59, a potent inhibitor of C3 convertase that prevents the spontaneous activation of the alternative complement pathway.

Red cells deficient in GPI-linked factors are abnormally susceptible to lysis or injury by complement. This manifests as intravascular hemolysis, which is caused by the C5b-C9 membrane attack complex. The hemolysis is paroxysmal and nocturnal in only 25% of cases; chronic hemolysis without dramatic hemoglobinuria is more typical. The tendency for red cells to lyse at night is explained by a slight decrease in blood pH during sleep, which increases the activity of complement. The anemia is variable but usually mild to moderate in severity. The loss of heme iron in the urine (hemosiderinuria) eventually leads to iron deficiency, which can exacerbate the anemia if untreated.

Thrombosis is the leading cause of disease-related death in individuals with PNH. About 40% of patients suffer from venous thrombosis, often involving the hepatic, portal, or

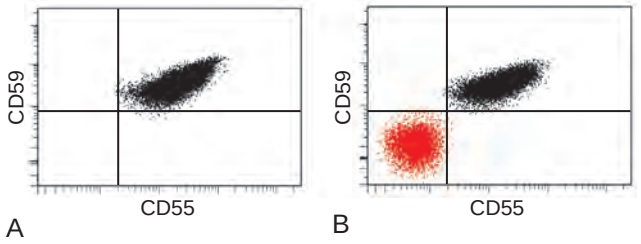


Figure 14.14 Paroxysmal nocturnal hemoglobinuria (PNH). (A) Flow cytogram of blood from a normal individual shows that the red cells express two phosphatidylinositol glycan (PIG)-linked membrane proteins, CD55 and CD59, on their surfaces. (B) Flow cytogram of blood from a patient with PNH shows a population of red cells that is deficient in both CD55 and CD59. As is typical of PNH, a second population of CD55+/CD59+ red cells that is derived from residual normal hematopoietic stem cells also is present. (Courtesy Dr. Scott Rodig, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

cerebral veins. How complement activation leads to thrombosis in patients with PNH is not clear; the absorption of NO by free hemoglobin (discussed in the Sickle Cell Disease section earlier in this chapter) may be one contributing factor, and a role for endothelial damage caused by the C5-9 membrane attack complex is also suspected.

About 5% to 10% of patients eventually develop acute myeloid leukemia or a myelodysplastic syndrome, indicating that PNH may arise in the context of genetic damage to hematopoietic stem cells.

PNH is diagnosed by flow cytometry, which provides a sensitive means for detecting red cells that are deficient in GPI-linked proteins such as CD59 (Fig. 14.14). The cardinal role of complement activation in PNH pathogenesis has been proven by therapeutic use of a monoclonal antibody called Eculizumab that prevents the conversion of C5 to C5a. This inhibitor not only reduces the hemolysis and attendant transfusion requirements, but also lowers the risk of thrombosis by up to 90%. The drawbacks to C5 inhibitor therapy are its high cost and an increased risk of serious or fatal meningococcal infection (as is true in individuals with inherited complement defects). Immunosuppressive drugs are sometimes beneficial for those with evidence of marrow aplasia. The only cure is hematopoietic stem cell transplantation.

Immunohemolytic Anemia

Immunohemolytic anemia is caused by antibodies that recognize red cells and lead to their premature destruction. Although these disorders are commonly referred to as autoimmune hemolytic anemias, the designation immunohemolytic anemia is preferred because the immune reaction is initiated in some instances by an ingested drug. Immunohemolytic anemia can be classified based on the characteristics of the responsible antibody (Table 14.4).

The diagnosis of immunohemolytic anemia requires the detection of antibodies and/or complement on red cells from the patient. This is done using the *direct Coombs antiglobulin test,* in which the patient's red cells are mixed with sera containing antibodies that are specific for human immunoglobulin or complement. If either immunoglobulin or complement is present on the surface of the red cells, the antibodies cause agglutination, which is appreciated visually as clumping. In the *indirect Coombs antiglobulin test,* the patient's serum is tested for its ability to agglutinate commercially available red cells bearing particular defined antigens. This test is used to characterize the antigen target and temperature dependence of the responsible antibody. Quantitative immunologic tests to measure such antibodies directly also are available.

Warm Antibody Type. This form constitutes approximately 80% of cases of immunohemolytic anemia. It is caused by antibodies that bind stably to red cells at 37°C. About 50% of cases are idiopathic (primary); others are related to a predisposing condition (see Table 14.4) or exposure to a drug. Most causative antibodies are of the IgG class; less commonly, IgA antibodies are the culprits. The red cell hemolysis is mostly extravascular. IgG-coated red cells bind to Fc receptors on phagocytes, which remove red cell membrane during "partial" phagocytosis. As in hereditary spherocytosis, the loss of membrane converts the red cells to spherocytes, which are sequestered and destroyed in the spleen. Moderate splenomegaly due to hyperplasia of splenic phagocytes is usually seen.

As with other autoimmune disorders, the cause of primary immunohemolytic anemia is unknown. In cases that are idiopathic, the antibodies are directed against red cell surface proteins, often components of the Rh blood group complex. In drug-induced cases, two mechanisms have been described.

- *Antigenic drugs.* In this setting hemolysis usually follows large, intravenous doses of the offending drug and occurs 1 to 2 weeks after therapy is initiated. These drugs, exemplified by penicillin and cephalosporins, bind to the red cell membrane and create a new antigenic determinant that is recognized by antibodies. The responsible antibodies sometimes fix complement and cause intravascular hemolysis, but more often they act as opsonins that promote extravascular hemolysis within phagocytes.
- *Tolerance-breaking drugs.* These drugs, of which the antihypertensive agent α-methyldopa is the prototype, break tolerance in some unknown manner that leads to the production of antibodies against red cell antigens, particularly the Rh blood group antigens. About 10% of patients taking α-methyldopa develop autoantibodies, as assessed by the direct Coombs test, and roughly 1% develop clinically significant hemolysis.

Table 14.4 Classification of Immunohemolytic Anemia

Warm Antibody Type (IgG Antibodies Active at 37°C)
Primary (idiopathic)
Secondary
Autoimmune disorders (particularly systemic lupus erythematosus)
Drugs
Lymphoid neoplasms
Cold Agglutinin Type (IgM Antibodies Active Below 37°C)
Acute (mycoplasmal infection, infectious mononucleosis)
Chronic
Idiopathic
Lymphoid neoplasms
Cold Hemolysin Type (IgG Antibodies Active Below 37°C)
Rare; occurs mainly in children following viral infections

Treatment of warm antibody immunohemolytic anemia centers on the removal of initiating factors (i.e., drugs); when this is not feasible, immunosuppressive drugs and splenectomy are the mainstays.

Cold Agglutinin Type. This type of immunohemolytic anemia is caused by IgM antibodies that bind to red cells avidly at low temperatures (0°C to 4°C) but not at 37°C. It accounts for 15% to 20% of cases. Cold agglutinin antibodies sometimes appear transiently following certain infections, such as with *Mycoplasma pneumoniae,* Epstein-Barr virus, cytomegalovirus, influenza virus, and human immunodeficiency virus (HIV). In these settings, the disorder is self-limited and the antibodies rarely induce clinically important hemolysis. Chronic cold agglutinin immunohemolytic anemia occurs in association with certain B-cell neoplasms or as an idiopathic condition.

Clinical symptoms result from binding of IgM to red cells in vascular beds where the temperature may fall below 30°C, such as in exposed fingers, toes, and ears. IgM binding agglutinates red cells and fixes complement rapidly. As the blood recirculates and warms, IgM is released, usually before complement-mediated hemolysis can occur; therefore, intravascular hemolysis is usually not seen. However, the transient interaction with IgM is sufficient to deposit sublytic quantities of C3b, an excellent opsonin, which leads to the removal of red cells by phagocytes in the spleen, liver, and bone marrow (extravascular hemolysis). The hemolysis is of variable severity. Vascular obstruction caused by agglutinated red cells may produce pallor, cyanosis, and Raynaud phenomenon (Chapter 11) in parts of the body that are exposed to cold temperatures. Chronic cold agglutinin immunohemolytic anemia caused by IgM antibodies may be difficult to treat. The best approach, when possible, is avoidance of cold temperatures.

Cold Hemolysin Type. Cold hemolysins are autoantibodies responsible for an unusual entity known as *paroxysmal cold hemoglobinuria.* This rare disorder may cause substantial, sometimes fatal, intravascular hemolysis and hemoglobinuria. The autoantibodies are IgGs that bind to the P blood group antigen on the red cell surface in cool, peripheral regions of the body. Complement-mediated lysis occurs when the cells recirculate to the body's warm core, where the complement cascade functions more efficiently. Most cases are seen in children following viral infections; in this setting the disorder is transient, and most of those affected recover within 1 month.

Hemolytic Anemia Resulting From Trauma to Red Cells

The most significant hemolysis caused by trauma to red cells is seen in individuals with cardiac valve prostheses and microangiopathic disorders. Artificial mechanical cardiac valves are more frequently implicated than are bioprosthetic porcine or bovine valves. The hemolysis stems from shear forces produced by turbulent blood flow and pressure gradients across damaged valves. *Microangiopathic hemolytic anemia* is most commonly seen with disseminated intravascular coagulation (DIC), but it also occurs in thrombotic thrombocytopenic purpura (TTP), hemolytic uremic syndrome (HUS), malignant hypertension, systemic lupus erythematosus, and disseminated cancer. The common pathogenic feature in these disorders is microvascular lesions that result in luminal narrowing, often due to the deposition of thrombi, producing shear stresses that mechanically injure passing red cells. Regardless of the cause, traumatic damage leads to intravascular hemolysis and the appearance of red cell fragments *(schistocytes),* "burr cells," "helmet cells," and "triangle cells" in blood smears (Fig. 14.15).

KEY CONCEPTS

Hereditary Spherocytosis

- Autosomal dominant disorder caused by mutations that affect the red cell membrane skeleton, leading to loss of membrane and eventual conversion of red cells to spherocytes, which are phagocytosed and removed in the spleen
- Manifested by anemia and splenomegaly

Thalassemias

- Autosomal codominant disorders caused by mutations in α- or β-globin that reduce hemoglobin synthesis, resulting in a microcytic, hypochromic anemia.
- In β-thalassemia, unpaired α-globin chains form aggregates that damage red cell precursors and further impair erythropoiesis.
- Ineffective erythropoesis increases iron absorption and can lead to systemic iron overload.

Sickle Cell Anemia

- Autosomal recessive disorder resulting from a mutation in β-globin that causes deoxygenated hemoglobin to self-associate into long polymers that distort (sickle) the red cell
- Episodic lockage of vessels by sickle red cells causes pain crises and tissue infarction, particularly of the marrow and spleen
- Red cell membrane damage caused by repeated bouts of sickling results in a moderate to severe hemolytic anemia

Glucose-6-Phosphate Dehydrogenase Deficiency

- X-linked disorder caused by mutations that destabilize G6PD, making red cells susceptible to oxidant damage

Immunohemolytic Anemias

- Caused by antibodies against either normal red cell constituents or antigens modified by haptens (e.g., drugs)
- Antibody binding results in either red cell opsonization and extravascular hemolysis or (uncommonly) complement fixation and intravascular hemolysis

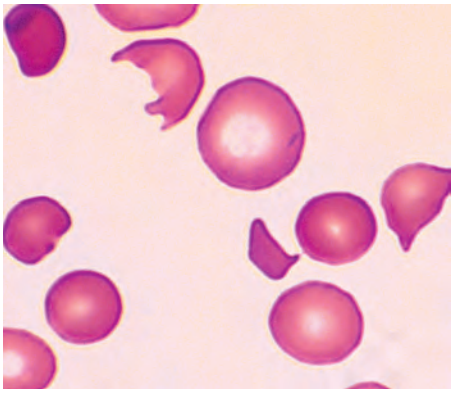


Figure 14.15 Microangiopathic hemolytic anemia. A peripheral blood smear from a patient with hemolytic uremic syndrome shows several fragmented red cells. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Anemias of Diminished Erythropoiesis

Although anemias caused by inadequate red cell production are heterogeneous, they can be classified into several major categories based on pathophysiology (see Table 14.1). The most common and important anemias associated with red cell underproduction are those caused by nutritional deficiencies, followed by those that arise secondary to chronic inflammation or renal failure. Also included are less common disorders that lead to generalized bone marrow failure, such as aplastic anemia, primary hematopoietic neoplasms (Chapter 13), and infiltrative disorders that replace the marrow (e.g., metastatic cancer and disseminated granulomatous disease). We first discuss the extrinsic causes of diminished erythropoiesis, which are more common and clinically important, and then non-neoplastic intrinsic causes.

Megaloblastic Anemia

The common theme among the various causes of megaloblastic anemia is an impairment of DNA synthesis that leads to ineffective hematopoiesis and distinctive morphologic changes, including abnormally large erythroid precursors and red cells. The causes of megaloblastic anemia are given in Table 14.5. The following discussion first describes the common features and then turns to the two principal subtypes: pernicious anemia (the major form of vitamin B_{12} deficiency anemia) and folate deficiency anemia.

Some of the metabolic roles of vitamin B_{12} and folate are considered later. For now it suffices that vitamin B_{12} and folic acid are coenzymes required for the synthesis of thymidine, one of the four bases found in DNA. A deficiency of these vitamins or impairment of their metabolism results in defective nuclear maturation due to deranged or inadequate DNA synthesis, with an attendant delay or block in cell division.

Table 14.5 Causes of Megaloblastic Anemia

Vitamin B_{12} Deficiency
Decreased Intake
Inadequate diet, vegetarianism
Impaired Absorption
Intrinsic factor deficiency Pernicious anemia Gastrectomy Malabsorption states Diffuse intestinal disease (e.g., lymphoma, systemic sclerosis) Ileal resection, ileitis Competitive parasitic uptake Fish tapeworm infestation Bacterial overgrowth in blind loops and diverticula of bowel
Folic Acid Deficiency
Decreased Intake
Inadequate diet, alcoholism, infancy
Impaired Absorption
Malabsorption states Intrinsic intestinal disease Anticonvulsants, oral contraceptives
Increased Loss
Hemodialysis
Increased Requirement
Pregnancy, infancy, disseminated cancer, markedly increased hematopoiesis
Impaired Utilization
Folic acid antagonists
Unresponsive to Vitamin B_{12} or Folic Acid Therapy
Metabolic inhibitors of DNA synthesis and/or folate metabolism (e.g., methotrexate)

Modified from Beck WS: Megaloblastic anemias. In Wyngaarden JB, Smith LH, editors: *Cecil Textbook of Medicine*, ed 18, Philadelphia, 1988, WB Saunders, p 900.

MORPHOLOGY

Certain peripheral blood findings are common to all forms of megaloblastic anemia. The presence of red cells that are macrocytic and oval **(macro-ovalocytes)** is highly characteristic. Because they are larger than normal and contain ample hemoglobin, most macrocytes lack the central pallor of normal red cells and even appear "hyperchromic," but the mean cell hemoglobin content is not elevated. There is marked variation in red cell size (anisocytosis) and shape (poikilocytosis). The reticulocyte count is low. Nucleated red cell progenitors occasionally appear in the circulating blood when anemia is severe. Neutrophils are also larger than normal and show **nuclear hypersegmentation,** having five or more nuclear lobules instead of the normal three to four (Fig. 14.16).

The marrow is usually markedly hypercellular as a result of increased numbers of hematopoietic precursors. **Megaloblastic changes** are detected at all stages of erythroid development**.** The most primitive cells (promegaloblasts) are large, with a deeply basophilic cytoplasm, prominent nucleoli, and a distinctive, fine nuclear chromatin pattern (Fig. 14.17). As these cells differentiate and begin to accumulate hemoglobin, the nucleus retains its finely distributed chromatin instead of developing the clumped pyknotic chromatin typical of normoblasts. Although nuclear maturation is delayed, cytoplasmic maturation and hemoglobin accumulation proceed at a normal pace, leading to nuclear-to-cytoplasmic asynchrony. Because DNA synthesis is impaired in all proliferating cells, granulocytic precursors also display dysmaturation in the form of **giant metamyelocytes** and **band forms.** Megakaryocytes also may be abnormally large and have bizarre, multilobate nuclei.

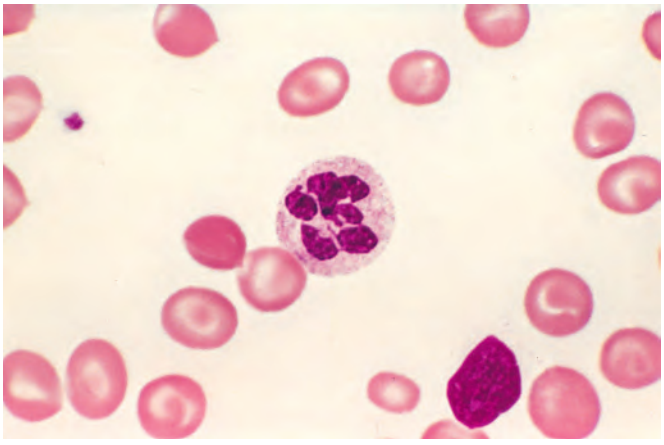


Figure 14.16 Megaloblastic anemia. A peripheral blood smear shows a hypersegmented neutrophil with a six-lobed nucleus. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

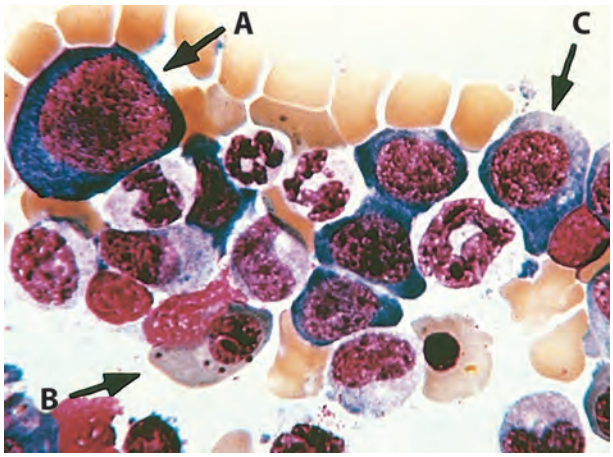


Figure 14.17 Megaloblastic anemia (bone marrow aspirate). *A to C,* Megaloblasts in various stages of differentiation. Note that the orthochromatic megaloblast *(B)* is hemoglobinized (as revealed by cytoplasmic color), but in contrast to normal orthochromatic normoblasts, the nucleus is not pyknotic. The early erythroid precursors *(A and C)* and the granulocytic precursors also are enlarged and have abnormally immature chromatin. (Courtesy Dr. Jose Hernandez, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

The marrow hyperplasia is a response to increased levels of growth factors, such as erythropoietin. However, the derangement in DNA synthesis causes most precursors to undergo apoptosis in the marrow (an example of ineffective hematopoiesis) and leads to pancytopenia. The anemia is further exacerbated by a mild degree of red cell hemolysis of uncertain etiology.

Vitamin B_{12} Deficiency Anemia: Pernicious Anemia

Pernicious anemia is a specific form of megaloblastic anemia caused by an autoimmune gastritis that impairs the production of intrinsic factor, which is required for vitamin B_{12} uptake from the gut.

Normal Vitamin B_{12} Metabolism. Vitamin B_{12} is a complex organometallic compound also known as cobalamin that is present in animal products such as meat, fish, milk, and eggs. The daily requirement is 2 to 3 μg. A diet that includes animal products contains significantly more than the minimal daily requirement and normally results in the accumulation of intrahepatic stores of vitamin B_{12} that are sufficient to last for several years. By contrast, plants and vegetables contain little cobalamin, and strictly vegetarian or macrobiotic diets do not provide adequate amounts of this essential nutrient.

Absorption of vitamin B_{12} requires intrinsic factor, which is secreted by the parietal cells of the fundic mucosa (Fig. 14.18). Vitamin B_{12} is freed from binding proteins in food through the action of pepsin in the stomach and binds to a salivary protein called *haptocorrin*. In the duodenum, bound vitamin B_{12} is released from haptocorrin by the action of pancreatic proteases and associates with intrinsic factor. This complex is transported to the ileum, where it is endocytosed by ileal enterocytes that express a receptor for intrinsic factor called *cubilin* on their surfaces. Within ileal cells, vitamin B_{12} associates with a major carrier protein, transcobalamin II, and is secreted into the plasma. Transcobalamin II delivers

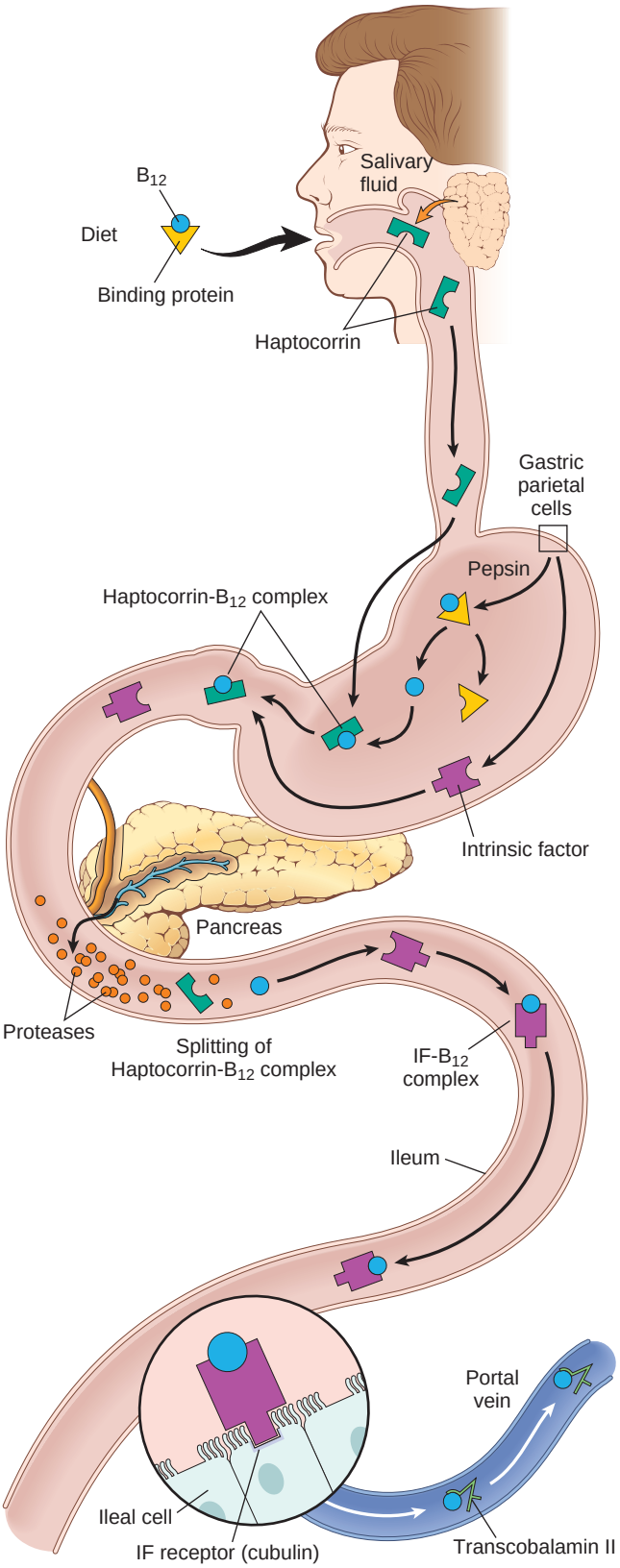


Figure 14.18 Schematic illustration of vitamin B_{12} absorption. *IF,* Intrinsic factor; *haptocorrin, cubilin,* see text.

vitamin B_{12} to the liver and other cells of the body, including rapidly proliferating cells in the bone marrow and the gastrointestinal tract. In addition to this major pathway, oral B_{12} can also be absorbed (albeit inefficiently) by passive diffusion, making it feasible to treat pernicious anemia with high doses of oral vitamin B_{12}.

Biochemical Functions of Vitamin B_{12}. Only two reactions in humans are known to require vitamin B_{12}. In one, methylcobalamin serves as an essential cofactor in the conversion of homocysteine to methionine by methionine synthase (Fig. 14.19). In the process, methylcobalamin yields a methyl group that is recovered from N^5-methyltetrahydrofolic acid (N^5-methyl FH_4), the principal form of folic acid in plasma. In the same reaction, N^5-methyl FH_4 is converted to tetrahydrofolic acid (FH_4). FH_4 is crucial because it is required (through its derivative $N^{5,10}$-methylene FH_4) for the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), a building block for DNA. It is postulated that impaired DNA synthesis in vitamin B_{12} deficiency stems from the reduced availability of FH_4, most of which is "trapped" as N^5-methyl FH_4. The FH_4 deficit may be exacerbated by an "internal" deficiency of metabolically active polyglutamylated forms of FH_4. This stems from a requirement for vitamin B_{12} in the synthesis of methionine, which contributes a carbon group that is needed in the metabolic reactions that create folate polyglutamates (Fig. 14.20). Whatever the mechanism, lack of folate is the proximate cause of anemia in vitamin B_{12} deficiency, as the anemia improves following the administration of folic acid.

The neurologic complications associated with vitamin B_{12} deficiency are more enigmatic, because they are not improved (and may actually be worsened) by folate administration. The other known reaction that depends on vitamin B_{12} is the isomerization of methylmalonyl coenzyme A to succinyl coenzyme A by the enzyme methylmalonyl-coenzyme A mutase, which requires adenosylcobalamin. A deficiency of vitamin B_{12} thus leads to increased plasma and urine levels of methylmalonic acid. Interruption of this reaction and the consequent buildup of methylmalonate and propionate (a precursor) could lead to the formation and incorporation of abnormal fatty acids into neuronal lipids. It has been suggested that this biochemical abnormality predisposes to myelin breakdown, producing subacute combined degeneration of the spinal cord (Chapter 28). However, rare individuals with hereditary deficiencies of methylmalonyl-coenzyme A mutase do not suffer from these abnormalities, casting doubt on this explanation.

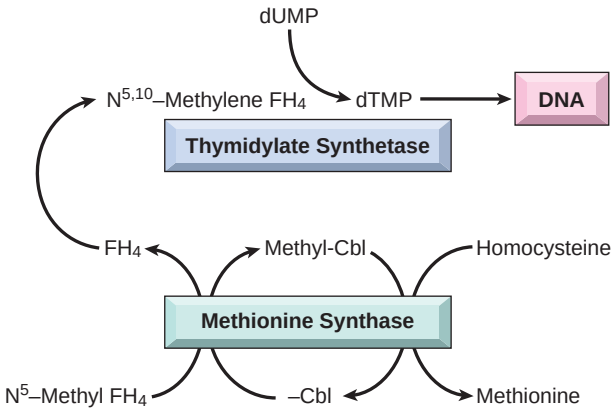


Figure 14.19 Relationship of N^5-methyl FH_4, methionine synthase, and thymidylate synthetase. In cobalamin (Cbl) deficiency, folate is sequestered as N^5-methyl FH_4. This ultimately deprives thymidylate synthetase of its folate coenzyme ($N^{5,10}$-methylene FH_4), thereby impairing DNA synthesis. *dTMP,* Deoxythymidine monophosphate; *dUMP,* deoxyuridine monophosphate; *FH_4,* tetrahydrofolic acid.

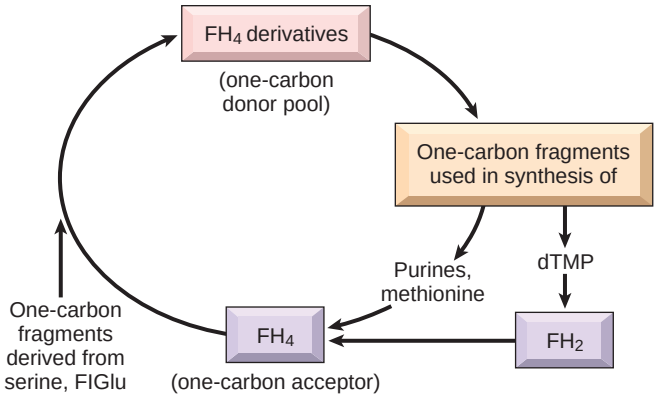


Figure 14.20 Role of folate derivatives in the transfer of one-carbon fragments for synthesis of biologic macromolecules. *dTMP,* Deoxythymidine monophosphate; *FH_4,* tetrahydrofolic acid; *FH_2,* dihydrofolic acid; *FIGlu,* formiminoglutamate.

Pernicious Anemia

Incidence. Although somewhat more prevalent in Scandinavian and other Caucasian populations, pernicious anemia occurs in all racial groups, including people of African descent and Hispanics. It is a disease of older adults; the median age at diagnosis is 60 years, and it is rare in people younger than 30 years of age. A genetic predisposition is strongly suspected, but no definable genetic pattern of transmission has been discerned. As described later, many affected individuals have a tendency to form antibodies against multiple self antigens.

Pathogenesis

Pernicious anemia is believed to result from an autoimmune attack on the gastric mucosa. Histologically, there is a chronic atrophic gastritis marked by a loss of parietal cells, a prominent infiltrate of lymphocytes and plasma cells, and megaloblastic changes in mucosal cells similar to those found in erythroid precursors. Three types of autoantibodies are present in many, but not all, patients. About 75% of patients have a type I antibody that blocks the binding of vitamin B_{12} to intrinsic factor. Type I antibodies are found in both plasma and gastric juice. Type II antibodies prevent binding of the intrinsic factor–vitamin B_{12} complex and also are present in a large proportion of patients with pernicious anemia. Both type I and type II antibodies are found in plasma and gastric juice. Type III antibodies are present in 85% to 90% of patients and recognize the α and β subunits of the gastric proton pump, a component of the microvilli of the canalicular system of the gastric parietal cell. Type III antibodies are not specific, as they are found in as many as 50% of older adults with idiopathic chronic gastritis.

Autoantibodies are of diagnostic utility, but are not thought to be the primary cause of the gastric pathology; rather, it seems that an autoreactive T-cell response initiates gastric mucosal injury and triggers the formation of autoantibodies. When the mass of intrinsic factor–secreting cells falls below a threshold (and reserves of stored vitamin B_{12} are depleted), anemia develops. Notably, pernicious anemia also is associated with other autoimmune disorders, particularly autoimmune thyroiditis and adrenalitis, suggesting that it arises in individuals with a predisposition to develop autoimmunity.

Vitamin B_{12} deficiency also may arise from causes other than pernicious anemia. Most of these impair absorption of the vitamin at one of the steps outlined earlier (see Table 14.5). With achlorhydria and loss of pepsin secretion (which occurs in some older adults), vitamin B_{12} is not readily released from proteins in food. With gastrectomy, intrinsic factor is lost. With insufficiency of the exocrine pancreas, vitamin B_{12} cannot be released from haptocorrin-vitamin B_{12} complexes. Ileal resection or diffuse ileal disease may prevent adequate absorption of intrinsic factor–vitamin B_{12} complex. Certain tapeworms (particularly those acquired by eating raw fish) compete with the host for B_{12} and can induce a deficiency state. In some settings, such as pregnancy, hyperthyroidism, disseminated cancer, and chronic infection, an increased demand for vitamin B_{12} may produce a relative deficiency, even with normal absorption.

MORPHOLOGY

The findings in the bone marrow and blood in pernicious anemia are similar to those described earlier for other forms of megaloblastic anemia. The stomach typically shows diffuse chronic gastritis (Chapter 17). The most characteristic alteration is **fundic gland atrophy,** affecting both chief cells and parietal cells, the latter being virtually absent. The glandular epithelium is replaced by mucus-secreting goblet cells that resemble those lining the large intestine, a form of metaplasia referred to as **intestinalization.** Some of the affected cells and their nuclei may be twice normal in size, a "megaloblastic" change analogous to that seen in the marrow. With time, the tongue may take on a shiny, glazed, "beefy" appearance **(atrophic glossitis).** Because the gastric atrophy and metaplastic changes are due to autoimmunity, they persist following parenteral administration of vitamin B_{12}, whereas the "megaloblastic" changes in the marrow and gut are readily reversible.

Central nervous system lesions are found in about three-fourths of all cases of florid pernicious anemia but can also be seen in the absence of overt hematologic findings. The principal alterations involve the cord, where there is **demyelination of the dorsal and lateral spinal tracts,** sometimes followed by loss of axons. These changes may give rise to spastic paraparesis, sensory ataxia, and severe paresthesias in the lower limbs. Less frequently, degenerative changes occur in the ganglia of the posterior roots and in peripheral nerves (Chapter 28).

Clinical Features

Pernicious anemia is insidious in onset, and the anemia is often quite severe by the time it comes to medical attention. The course is progressive unless halted by therapy.

The diagnosis is based on (1) a moderate to severe megaloblastic anemia, (2) leukopenia with hypersegmented granulocytes, (3) low serum vitamin B_{12}, and (4) elevated serum levels of homocysteine and methylmalonic acid. Serum antibodies to intrinsic factor are highly specific for pernicious anemia. The diagnosis is confirmed by an outpouring of reticulocytes and a rise in hematocrit levels beginning about 5 days after parenteral administration of vitamin B_{12}.

Persons with atrophy and metaplasia of the gastric mucosa due to pernicious anemia are at increased risk for gastric carcinoma (Chapter 17). With parenteral or high-dose oral vitamin B_{12}, the anemia is cured and the progression of the peripheral neurologic disease can be reversed or at least halted, but the changes in the gastric mucosa and the risk of carcinoma are unaffected.

Anemia of Folate Deficiency

A deficiency of folic acid (more properly, pteroylmonoglutamic acid) results in a megaloblastic anemia having the same pathologic features as that caused by vitamin B_{12} deficiency. FH_4 derivatives act as intermediates in the transfer of one-carbon units such as formyl and methyl groups to various compounds (see Fig. 14.20). FH_4 serves as an acceptor of one-carbon fragments from compounds such as serine and formiminoglutamic acid. The FH_4 derivatives so generated in turn donate the acquired one-carbon fragments in reactions synthesizing various metabolites. FH_4, then, can be viewed as the biologic "middleman" in a series of swaps involving one-carbon moieties. The most important metabolic processes depending on such transfers are (1) purine synthesis; (2) the conversion of homocysteine to methionine, a reaction also requiring vitamin B_{12}; and (3) dTMP synthesis. In the first two reactions, FH_4 is regenerated from its one-carbon carrier derivatives and is available to accept another one-carbon moiety and reenter the donor pool. In the synthesis of dTMP, dihydrofolic acid (FH_2) is produced that must be reduced by dihydrofolate reductase for reentry into the FH_4 pool. The reductase step is significant, because this enzyme is susceptible to inhibition by various drugs. Among the molecules whose synthesis is dependent on folates, dTMP is perhaps the most important biologically, because it is required for DNA synthesis. It should be apparent from this discussion that **suppressed synthesis of DNA, the common denominator of folic acid and vitamin B_{12} deficiency, is the immediate cause of megaloblastosis.**

Etiology

The three major causes of folic acid deficiency are (1) decreased intake, (2) increased requirements, and (3) impaired utilization (see Table 14.5). Humans depend on dietary sources for folic acid. Most normal diets contain ample amounts. The richest sources are green vegetables such as lettuce, spinach, asparagus, and broccoli. Certain fruits (e.g., lemons, bananas, melons) and animal sources (e.g., liver) contain lesser amounts. The folic acid in these foods is largely in the form of folylpolyglutamates. Although abundant in raw foods, polyglutamates are sensitive to heat; boiling, steaming, or frying food for 5 to 10 minutes destroys up to 95% of the folate content. Intestinal conjugases split the polyglutamates into monoglutamates that are absorbed in the proximal jejunum. During intestinal absorption they are modified to 5-methyltetrahydrofolate,

the transport form of folate. The body's reserves of folate are relatively modest, and a deficiency can arise within weeks to months if intake is inadequate.

Decreased intake can result from either a nutritionally inadequate diet or impairment of intestinal absorption. A normal diet contains folate in excess of the minimal daily adult requirement. Inadequate dietary intakes are almost invariably associated with grossly deficient diets, which are most frequently encountered in chronic alcoholics, the indigent, and the very old. In alcoholics with cirrhosis, other mechanisms of folate deficiency such as trapping of folate within the liver, excessive urinary loss, and disordered folate metabolism also have been implicated. Under these circumstances, the megaloblastic anemia is often accompanied by general malnutrition and manifestations of other avitaminoses, including cheilosis, glossitis, and dermatitis. Malabsorption syndromes, such as sprue, may lead to inadequate folate absorption, as may infiltrative diseases of the small intestine (e.g., lymphoma). In addition, certain drugs, particularly the anticonvulsant phenytoin and oral contraceptives, interfere with absorption.

Despite normal intake of folic acid, a relative deficiency can be encountered when requirements are increased. Conditions in which this is seen include pregnancy, infancy, derangements associated with hyperactive hematopoiesis (e.g., chronic hemolytic anemia), and disseminated cancer. In all of these circumstances, the demands of increased DNA synthesis render normal intake inadequate.

Folic acid antagonists, such as methotrexate, inhibit dihydrofolate reductase and lead to a deficiency of FH_4. Inhibition of folate metabolism affects all rapidly proliferating tissues, particularly the bone marrow and the gastrointestinal tract. Many chemotherapeutic drugs used in the treatment of cancer damage DNA or inhibit DNA synthesis through other mechanisms, and these also cause megaloblastic changes in rapidly dividing cells.

As mentioned earlier, the megaloblastic anemia that results from folic acid deficiency is identical to that encountered in vitamin B_{12} deficiency. Thus, the diagnosis of folate deficiency can be made only by demonstration of decreased serum or red cell folate levels. As in vitamin B_{12} deficiency, serum homocysteine levels are increased, but methylmalonate concentrations are normal. Importantly, neurologic changes do not occur.

Although prompt hematologic response heralded by reticulocytosis follows the administration of folic acid, it should be remembered that the hematologic symptoms of vitamin B_{12} deficiency anemia also respond to folate therapy. As mentioned earlier, folate does not prevent (and may even exacerbate) the neurologic deficits seen in vitamin B_{12} deficiency states. It is thus essential to exclude vitamin B_{12} deficiency as the cause of megaloblastic anemia before initiating therapy with folate.

Iron Deficiency Anemia

Deficiency of iron is the most common nutritional disorder in the world and results in clinical signs and symptoms that are mostly related to inadequate hemoglobin synthesis. Although the prevalence of iron deficiency anemia is higher in low income countries, this form of anemia is common in the United States, particularly in toddlers, adolescent girls, and women of childbearing age. The factors underlying iron deficiency differ somewhat in various population groups and can be best considered in the context of normal iron metabolism.

Table 14.6 Iron Distribution in Healthy Young Adults (mg)

Pool	Men	Women
Total	3450	2450
Functional		
Hemoglobin	2100	1750
Myoglobin	300	250
Enzymes	50	50
Storage		
Ferritin, hemosiderin	1000	400

Iron Metabolism. The normal daily Western diet contains about 10 to 20 mg of iron, mostly in the form of heme in animal products, with the remainder being inorganic iron in vegetables. About 20% of heme iron (in contrast with 1% to 2% of nonheme iron) is absorbable, so the average Western diet contains sufficient iron to balance fixed daily losses. The total body iron content is normally about 2.5 g in women and as high as 6 g in men, and it can be divided into functional and storage compartments (Table 14.6). About 80% of the functional iron is found in hemoglobin; the rest is in myoglobin and iron-containing enzymes such as catalase and cytochromes. The storage pool represented by hemosiderin and ferritin contains about 15% to 20% of total body iron. The major sites of storage iron are the liver and mononuclear phagocytes. Healthy young females have smaller iron stores than do males, primarily because of blood loss during menstruation, and often develop iron deficiency due to excessive loss or increased demand associated with menstruation and pregnancy, respectively.

Iron in the body is recycled between the functional and storage pools (Fig. 14.21). It is transported in plasma by an iron-binding protein called *transferrin,* which is synthesized in the liver. In normal individuals, transferrin is about one-third saturated with iron, yielding serum iron levels that average 120 μg/dL in men and 100 μg/dL in women. The major function of plasma transferrin is to deliver iron to cells, including erythroid precursors, which require iron to synthesize hemoglobin. Erythroid precursors possess high-affinity receptors for transferrin that mediate iron import through receptor-mediated endocytosis.

Free iron is highly toxic (Chapter 18), and storage iron must therefore be sequestered. This is achieved by the binding of storage iron to either ferritin or hemosiderin. Ferritin is a ubiquitous protein-iron complex that is found at highest levels in the liver, spleen, bone marrow, and skeletal muscles. In the liver, most ferritin is stored within the parenchymal cells; in other tissues, such as the spleen and the bone marrow, it is found mainly in macrophages. Hepatocyte iron is derived from plasma transferrin, whereas storage iron in macrophages is derived from the breakdown of red cells. Intracellular ferritin is located in the cytosol and in lysosomes, in which partially degraded protein shells of ferritin aggregate into hemosiderin granules. Iron in hemosiderin is chemically reactive and turns blue-black when exposed to potassium ferrocyanide, which is the basis for the *Prussian blue stain*. With normal iron stores, only

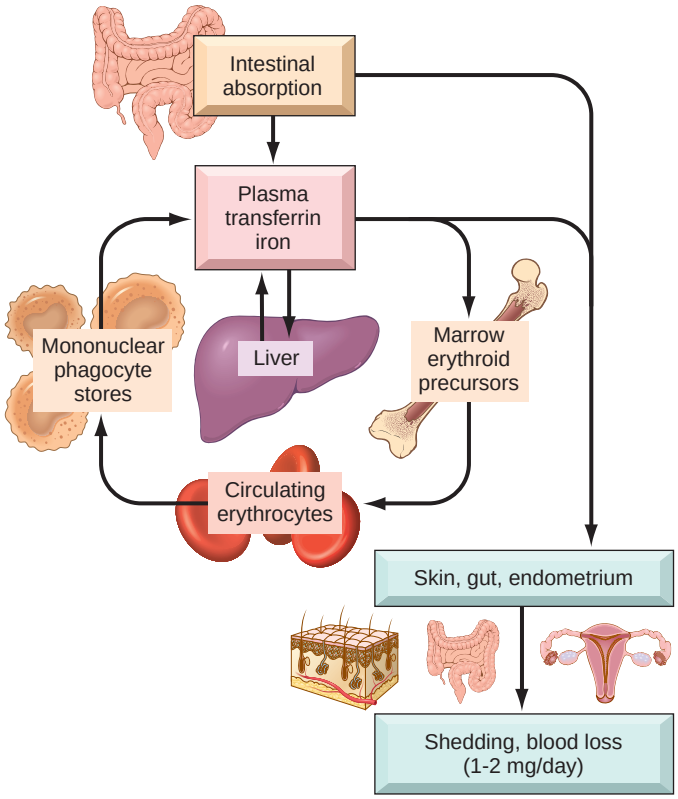


Figure 14.21 Iron metabolism. Iron absorbed from the gut is bound to plasma transferrin and transported to the bone marrow, where it is delivered to developing red cells and incorporated into hemoglobin. Mature red cells are released into the circulation and, after 120 days, are ingested by macrophages, primarily in the spleen, liver, and bone marrow. Here iron is extracted from hemoglobin and recycled to plasma transferrin. At equilibrium, iron absorbed from the gut is balanced by losses in shed keratinocytes, enterocytes, and (in women) endometrium.

trace amounts of hemosiderin are found in the body, principally in macrophages in the bone marrow, spleen, and liver, most being stored as ferritin. In iron-overloaded cells, most iron is stored in hemosiderin.

Because plasma ferritin is derived largely from the storage pool of body iron, its levels correlate with body iron stores. In iron deficiency, serum ferritin is below 12 µg/L, whereas in iron overload values approaching 5000 µg/L may be seen. Of physiologic importance, the storage iron pool can be readily mobilized if iron requirements increase, as may occur after loss of blood.

Because iron is essential for cellular metabolism and highly toxic in excess, total body iron stores must be regulated meticulously. Iron balance is maintained by regulating the absorption of dietary iron in the proximal duodenum. There is no regulated pathway for iron excretion, which is limited to the 1 to 2 mg lost each day through the shedding of mucosal and skin epithelial cells. By contrast, as body iron stores increase, absorption decreases, and vice versa.

The pathways responsible for the absorption of iron from the gut are now understood in reasonable detail (Fig. 14.22) and differ for nonheme and heme iron. Luminal nonheme iron is mostly in the Fe^{3+} (ferric) state and must first be reduced to Fe^{2+} (ferrous) iron by ferrireductases, such as b cytochromes and STEAP3. Fe^{2+} iron is then transported across the apical membrane by divalent metal transporter 1 (DMT1). Heme iron is moved across the apical membrane into the cytoplasm through transporters that are incompletely characterized. Here, it is metabolized to release Fe^{2+} iron, which enters a common pool with nonheme Fe^{2+} iron. The absorption of nonheme iron is variable and often inefficient, being inhibited by substances in the diet that bind and stabilize Fe^{3+} iron and enhanced by substances that stabilize Fe^{2+} iron (described later). Frequently, less than 5% of dietary nonheme iron is absorbed. In contrast, about 20% of the heme iron derived from hemoglobin, myoglobin, and other animal proteins is absorbed.

Once in duodenal cells, Fe^{2+} iron can follow one of two pathways: transport to the blood or storage as mucosal iron. Fe^{2+} iron destined for the circulation is transported from the cytoplasm across the basolateral enterocyte membrane by ferroportin. This process is coupled to the oxidation of Fe^{2+} iron to Fe^{3+} iron, which is carried out by the iron oxidases hephaestin and ceruloplasmin. Newly absorbed Fe^{3+} iron binds rapidly to transferrin, which delivers iron to red cell progenitors in the marrow (see Fig. 14.21). Both DMT1 and ferroportin are widely distributed in the body and are involved in iron transport in other tissues as well. For example, DMT1 mediates the uptake of "functional" iron (derived from endocytosed transferrin) across lysosomal membranes into the cytosol of red cell precursors in the bone marrow, and ferroportin plays an important role in the release of storage iron from macrophages.

Iron absorption in the duodenum is regulated by *hepcidin*, a small circulating peptide that is synthesized and released from the liver in response to increases in intrahepatic iron levels. Hepcidin inhibits iron transfer from the enterocyte to plasma by binding to ferroportin, causing it to be endocytosed and degraded. As a result, as hepcidin levels rise, iron becomes trapped within duodenal cells in the form of mucosal ferritin and is lost as these cells slough. Thus, when the body is replete with iron, high hepcidin levels inhibit its absorption into the blood. Conversely, with low body stores of iron, hepcidin levels fall, facilitating iron absorption. By inhibiting ferroportin, hepcidin not only reduces iron uptake from enterocytes but also suppresses iron release from macrophages, an important source of the iron that is used by erythroid precursors to make hemoglobin.

Alterations in hepcidin have a central role in diseases involving disturbances of iron metabolism. This is illustrated by the following examples.

- As described later in this chapter, the *anemia of chronic inflammation* is caused in part by inflammatory mediators that increase hepatic hepcidin production.
- A rare form of microcytic anemia is caused by mutations that disable TMPRSS6, a hepatic transmembrane serine protease that normally suppresses hepcidin production when iron stores are low. Affected patients have high hepcidin levels, resulting in reduced iron absorption and failure to respond to iron therapy.
- Conversely, hepcidin activity is inappropriately low in both primary and secondary *hemochromatosis*, a syndrome caused by systemic iron overload. As discussed in Chapter 18, the various inherited forms of primary hemochromatosis are associated with mutations in hepcidin or the genes that regulate hepcidin expression. Secondary hemochromatosis can occur in diseases associated

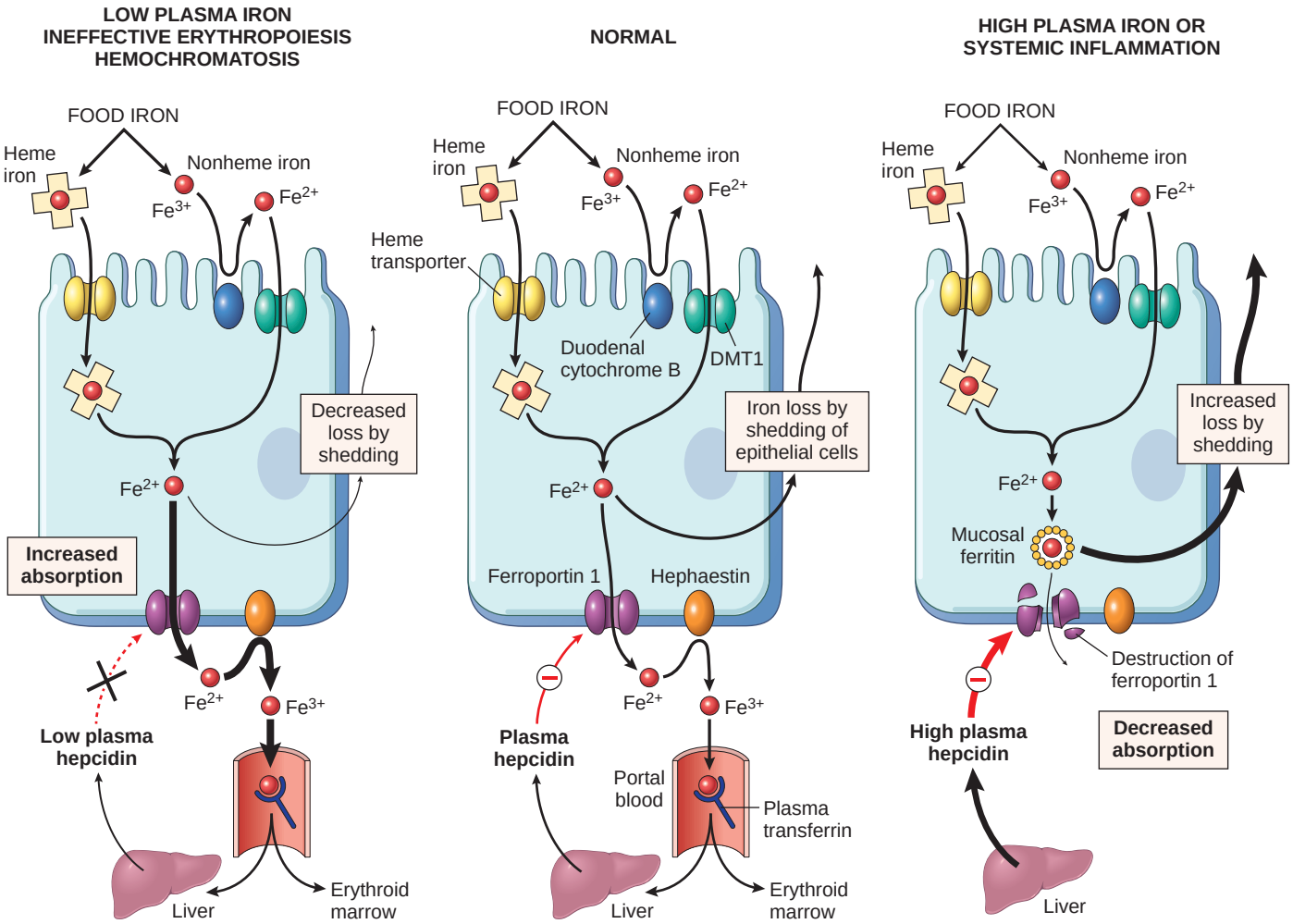


Figure 14.22 Regulation of iron absorption. Duodenal epithelial cell uptake of heme and nonheme iron is described in the text. When the storage sites of the body are replete with iron and erythropoietic activity is normal, plasma hepcidin balances iron uptake and loss to and maintains iron homeostasis by downregulating ferroportin and limiting iron uptake *(middle panel)*. Hepcidin rises in the setting of systemic inflammation or when liver iron levels are high, decreasing iron uptake and increasing iron loss during the shedding of duodenocytes *(right panel)*. Conversely, hepcidin levels fall in the setting of low plasma iron, primary hemochromatosis, or ineffective hematopoiesis *(left panel)*, leading to a rise in iron absorption. *DMT1*, Divalent metal transporter 1.

with ineffective erythropoiesis, such as β-thalassemia major (discussed earlier) and myelodysplastic syndrome (Chapter 13), due to expansion of erythroid progenitors and release of increased amounts of erythroferrone, which inhibits hepatic hepcidin production.

Etiology. Iron deficiency can result from (1) dietary lack, (2) impaired absorption, (3) increased requirement, or (4) chronic blood loss. To maintain a normal iron balance, about 1 mg of iron must be absorbed from the diet every day. Because only 10% to 15% of ingested iron is absorbed, the daily iron requirement is 7 to 10 mg for adult men and 7 to 20 mg for adult women. Because the average daily dietary intake of iron in the Western world is about 15 to 20 mg, most men ingest more than adequate iron, whereas many women consume marginal amounts of iron. The bioavailability of dietary iron is as important as the overall content. The absorption of inorganic iron is influenced by other dietary contents. It is enhanced by ascorbic acid, citric acid, amino acids, and sugars in the diet, and it is inhibited by tannates (found in tea), carbonates, oxalates, and phosphates.

Dietary lack is rare in high income countries, where on average about two-thirds of the dietary iron is in the form of heme, mainly in meat. The situation is different in low income countries, where food is less abundant and most dietary iron is found in plants in the poorly absorbable inorganic form. Dietary iron inadequacy occurs in even high income societies in the following groups:

- *Infants,* who are at high risk due to the very small amounts of iron in milk. Human breast milk provides only about 0.3 mg/L of iron. Cow's milk contains about twice as much iron, but its bioavailability is poor.
- *The impoverished*, who can have suboptimal diets for socioeconomic reasons at any age
- *Older adults*, who often have restricted diets with little meat because of limited income or poor dentition
- *Teenagers* who subsist on "junk" food

Impaired absorption is found in sprue, other causes of fat malabsorption (steatorrhea), and chronic diarrhea. Gastrectomy diminishes iron absorption by decreasing the acidity of the proximal duodenum (acidity enhances

uptake) and by increasing the speed with which gut contents pass through the duodenum. Specific items in the diet, as is evident from the preceding discussion, can also affect absorption.

Increased requirement is an important cause of iron deficiency in growing infants, children, and adolescents, as well as premenopausal women, particularly during pregnancy. Economically deprived women having multiple, closely spaced pregnancies are at exceptionally high risk.

Chronic blood loss is the most common cause of iron deficiency in high income societies. External hemorrhage or bleeding into the gastrointestinal, urinary, or genital tracts depletes iron reserves. Iron deficiency in adult men and postmenopausal women in high income countries must be attributed to gastrointestinal blood loss until proven otherwise. To prematurely ascribe iron deficiency in such individuals to any other cause is to run the risk of missing a gastrointestinal cancer or other bleeding lesion. An alert clinician investigating unexplained iron deficiency anemia occasionally discovers an occult bleeding source such as a cancer and thereby saves a life.

Pathogenesis

Whatever its basis, iron deficiency leads to inadequate hemoglobin production and hypochromic microcytic anemia. At the outset of chronic blood loss or other states of negative iron balance, reserves in the form of ferritin and hemosiderin may be adequate to maintain normal hemoglobin and hematocrit levels as well as normal serum iron and transferrin saturation. Progressive depletion of these reserves first lowers serum iron and transferrin saturation levels without producing anemia. In this early stage, there is increased erythroid activity in the bone marrow. Anemia appears only when iron stores are completely depleted and is accompanied by lower than normal serum iron, ferritin, and transferrin saturation levels.

MORPHOLOGY

The bone marrow reveals a mild to moderate increase in erythroid progenitors. A diagnostically significant finding is the **absence of stainable iron in macrophages,** which is best assessed by performing Prussian blue stains on smears of aspirated marrow. In peripheral blood smears, the red cells are small **(microcytic)** and pale **(hypochromic).** Normal red cells with sufficient hemoglobin have a zone of central pallor measuring about one third of the cell diameter. In established iron deficiency, the zone of pallor is enlarged; hemoglobin may be seen only in a narrow peripheral rim (Fig. 14.23). Poikilocytosis in the form of small, elongated red cells (pencil cells) also is characteristically seen.

Clinical Features

The clinical manifestations of the anemia are nonspecific and were detailed earlier. The dominating signs and symptoms frequently relate to the underlying cause, for example, gastrointestinal or gynecologic disease, malnutrition, pregnancy, or malabsorption. In severe long-standing iron deficiency, depletion of iron-containing enzymes in cells throughout the body also causes other changes, including koilonychia, alopecia, atrophic changes in the tongue and gastric mucosa, and intestinal malabsorption. Depletion of iron from the central nervous system may lead to the appearance of pica, in which affected individuals have a craving for non-foodstuffs such as clay or food ingredients such as flour, and periodically move their limbs during sleep. Pica is also seen in association with developmental disorders such as autism (in the absence of iron deficiency). Esophageal webs may appear together with microcytic hypochromic anemia and atrophic glossitis to complete the triad of findings in the rare *Plummer-Vinson syndrome* (Chapter 17).

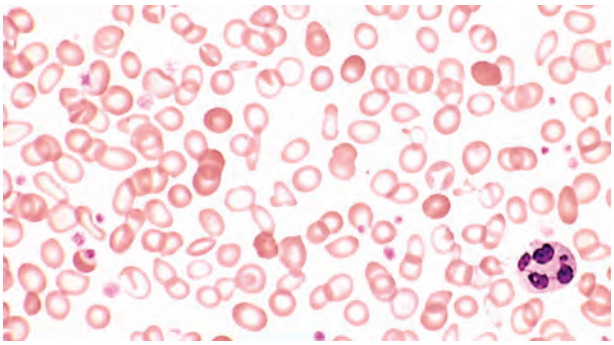


Figure 14.23 Iron deficiency (peripheral blood smear). Note the hypochromic microcytic red cells containing a narrow rim of peripheral hemoglobin. Scattered fully hemoglobinized cells, present due to recent blood transfusion, stand in contrast. (Courtesy Dr. Robert W. McKenna, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

The diagnosis of iron deficiency anemia ultimately rests on laboratory studies. Both the hemoglobin and hematocrit are depressed, usually to a moderate degree, in association with hypochromia, microcytosis, and modest poikilocytosis. The serum iron and ferritin are low, and the total plasma iron-binding capacity (reflecting elevated transferrin levels) is high. Low serum iron with increased iron-binding capacity results in a reduction of transferrin saturation to below 15%. Reduced iron stores inhibit hepcidin synthesis, and its serum levels fall. In uncomplicated cases, oral iron supplementation produces an increase in reticulocytes in about 5 to 7 days that is followed by a steady increase in blood counts and the normalization of red cell indices.

Anemia of Chronic Inflammation

Impaired red cell production associated with chronic diseases that produce systemic inflammation is a common cause of anemia in hospitalized patients. This form of anemia stems from a reduction in the proliferation of erythroid progenitors and impaired iron utilization. The chronic illnesses associated with this form of anemia can be grouped into three categories:

- *Chronic microbial infections,* such as osteomyelitis, bacterial endocarditis, and lung abscess
- *Chronic immune disorders,* such as rheumatoid arthritis and inflammatory bowel disease
- *Neoplasms,* such as carcinomas of the lung and breast, and Hodgkin lymphoma

The anemia of chronic inflammation is associated with low serum iron, reduced total iron-binding capacity, and abundant stored iron in tissue macrophages. Several effects

of inflammation contribute to the observed abnormalities. Most notably, certain inflammatory mediators, particularly interleukin-6 (IL-6), stimulate an increase in the hepatic production of hepcidin. As discussed earlier, hepcidin inhibits ferroportin function in macrophages and reduces the transfer of iron from the storage pool to developing erythroid precursors in the bone marrow. As a result, the erythroid precursors are starved for iron in the midst of plenty. In addition, these progenitors do not proliferate adequately because erythropoietin levels are inappropriately low for the degree of anemia. The precise mechanism underlying the reduction in erythropoietin is uncertain; direct suppression of renal erthropoietin production by inflammatory cytokines is suspected.

What might be the reason for iron sequestration in the setting of inflammation? The best guess is that it enhances the body's ability to fend off certain infections, particularly those caused by bacteria (e.g., *H. influenzae*) that require iron for pathogenicity. In this regard it is interesting to consider that hepcidin is structurally related to defensins, a family of peptides that have intrinsic antibacterial activity. This connection highlights the poorly understood but intriguing relationship between inflammation, innate immunity, and iron metabolism.

The anemia is usually mild, and the dominant symptoms are those of the underlying disease. The red cells can be normocytic and normochromic, or hypochromic and microcytic, as in anemia of iron deficiency. The presence of increased storage iron in marrow macrophages, a high serum ferritin level, and a reduced total iron-binding capacity readily rule out iron deficiency as the cause of anemia. Only successful treatment of the underlying condition reliably corrects the anemia, but some patients, particularly those with cancer, benefit from administration of erythropoietin.

Aplastic Anemia

Aplastic anemia refers to a syndrome of chronic primary hematopoietic failure and attendant pancytopenia (anemia, neutropenia, and thrombocytopenia). In the majority of patients, autoimmune mechanisms are suspected, but inherited or acquired abnormalities of hematopoietic stem cells also contribute in a subset of patients.

Etiology. The most common circumstances associated with aplastic anemia are listed in Table 14.7. Most cases of "known" etiology follow exposure to chemicals and drugs. Some of the associated agents (including many cancer chemotherapy drugs and the organic solvent benzene) cause marrow suppression that is dose related and reversible. In other instances, aplastic anemia arises in an unpredictable, idiosyncratic fashion following exposure to drugs that normally cause little or no marrow suppression. The implicated drugs in these idiosyncratic reactions include chloramphenicol and gold salts.

Persistent marrow aplasia can also appear after a variety of viral infections, most commonly viral hepatitis, which is associated with approximately 5% of cases. Why aplastic anemia develops in certain individuals is not understood.

Whole-body irradiation can destroy hematopoietic stem cells in a dose-dependent fashion. Persons who receive therapeutic irradiation or are exposed to radiation in nuclear accidents (e.g., Chernobyl) are at risk for marrow aplasia.

Table 14.7 Major Causes of Aplastic Anemia

Acquired
Idiopathic
Acquired stem cell defects
Immune mediated
Chemical Agents
Dose related
Alkylating agents
Antimetabolites
Benzene
Chloramphenicol
Inorganic arsenicals
Idiosyncratic
Chloramphenicol
Phenylbutazone
Organic arsenicals
Methylphenylethylhydantoin
Carbamazepine
Penicillamine
Gold salts
Physical Agents
Whole-body irradiation
Viral infections
Hepatitis (unknown virus)
Cytomegalovirus infections
Epstein-Barr virus infections
Herpes zoster (varicella zoster)
Inherited
Fanconi anemia
Telomerase defects

Specific abnormalities underlying some cases of aplastic aplasia are as follows.

- *Fanconi anemia* is a rare autosomal recessive disorder caused by defects in a multiprotein complex that is required for DNA repair (Chapter 7). Marrow hypofunction becomes evident early in life and is often accompanied by multiple congenital anomalies, such as hypoplasia of the kidney and spleen, and bone anomalies, commonly involving the thumbs or radii.
- Inherited defects in *telomerase* are found in 5% to 10% of adult-onset aplastic anemia. Telomerase is required for cellular immortality and limitless replication (Chapters 1 and 7). It might be anticipated, therefore, that deficits in telomerase activity could result in premature hematopoietic stem cell exhaustion and marrow aplasia.
- Even more common than telomerase mutations are abnormally short telomeres, which are found in the marrow cells of as many as one-half of those affected with aplastic anemia. It is unknown whether this shortening is due to other unappreciated telomerase defects or is a consequence of excessive stem cell replication.

In most instances, however, no initiating factor can be identified; about 65% of cases fall into this idiopathic category.

Pathogenesis

The pathogenesis of aplastic anemia is not fully understood. Indeed, it is unlikely that a single mechanism underlies all cases. However, two major etiologies have been invoked:

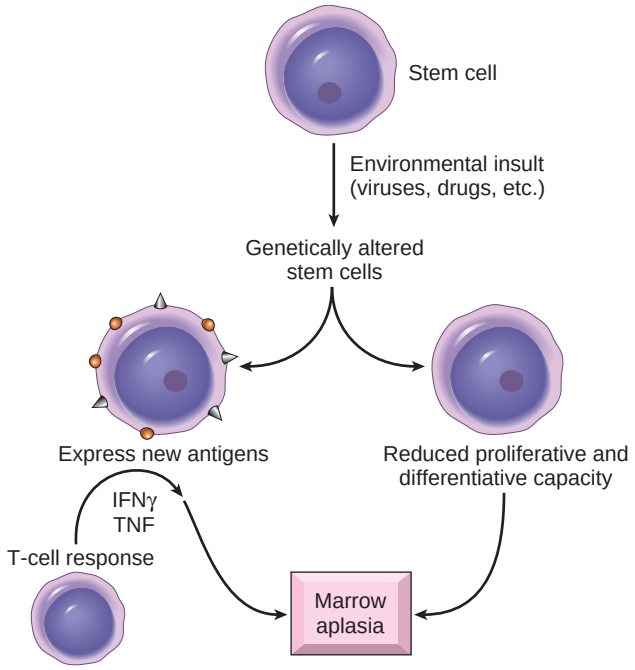


Figure 14.24 Pathophysiology of aplastic anemia. Damaged stem cells can produce progeny expressing neoantigens that evoke an autoimmune reaction, or give rise to a clonal population with reduced proliferative capacity. Either pathway could lead to marrow aplasia. See text for abbreviations.

an extrinsic, immune-mediated suppression of marrow progenitors, and an intrinsic abnormality of stem cells (Fig. 14.24).

Experimental studies have focused on a model in which activated T cells suppress hematopoietic stem cells. Stem cells may first be antigenically altered by exposure to drugs, infectious agents, or other unidentified environmental insults. This provokes a cellular immune response, during which activated Th1 cells produce cytokines such as interferon-γ (IFN-γ) and TNF that suppress and kill hematopoietic progenitors. This scenario is supported by several observations.

- Analysis of the few remaining marrow stem cells from aplastic anemia marrows has revealed that genes involved in apoptosis and death pathways are up-regulated; of note, the same genes are up-regulated in normal stem cells exposed to interferon-γ.
- Even more compelling (and clinically relevant) evidence comes from experience with immunosuppressive therapy. Antithymocyte globulin and other immunosuppressive drugs such as cyclosporine produce responses in 60% to 70% of patients. It is proposed that these therapies work by suppressing or killing autoreactive T-cell clones. The antigens recognized by the autoreactive T cells are not well defined. In some instances GPI-linked proteins may be the targets, possibly explaining the previously noted association of aplastic anemia and PNH.

Alternatively, the notion that aplastic anemia results from a fundamental stem cell abnormality is supported by the presence of karyotypic aberrations and acquired mutations involving cancer genes in many cases; the occasional transformation of aplasias into myeloid neoplasms, typically myelodysplastic syndrome or acute myeloid leukemia; and the association with abnormally short telomeres. Some marrow insult (or a predisposition to DNA damage) presumably results in sufficient injury to limit the proliferative and differentiation capacity of stem cells. If the damage is extensive enough, aplastic anemia results. These two mechanisms are not mutually exclusive, because genetically altered stem cells might also express "neoantigens" that could serve as targets for a T-cell attack.

MORPHOLOGY

The markedly hypocellular bone marrow is largely devoid of hematopoietic cells; often only fat cells, fibrous stroma, and scattered lymphocytes and plasma cells remain. Marrow aspirates often yield little material (a "dry tap"); hence, aplasia is best appreciated in marrow biopsies (Fig. 14.25). Other nonspecific pathologic changes are related to granulocytopenia and thrombocytopenia, such as mucocutaneous bacterial infections and abnormal bleeding, respectively. If the anemia necessitates multiple transfusions, systemic hemosiderosis can appear.

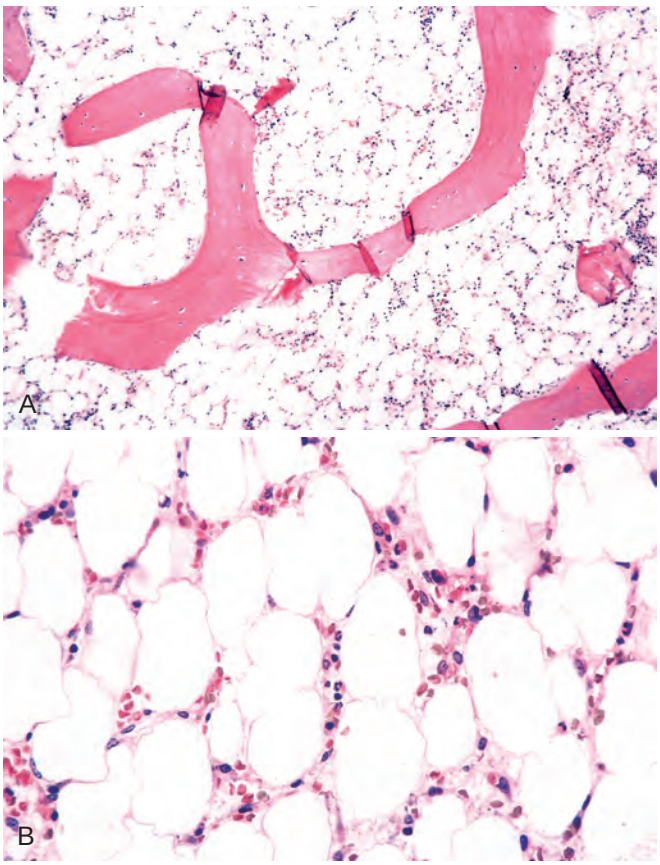


Figure 14.25 Aplastic anemia (bone marrow biopsy). Markedly hypocellular marrow contains mainly fat cells. (A) Low power. (B) High power. (Courtesy Dr. Steven Kroft, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

Clinical Features

Aplastic anemia can occur at any age and in either sex. The onset is usually insidious. Initial manifestations vary depending on which cell line is predominantly affected, but pancytopenia ultimately appears, with the expected consequences. Anemia leads to progressive weakness, pallor, and dyspnea; thrombocytopenia is heralded by petechiae and ecchymoses; and neutropenia manifests as frequent and persistent infections or the sudden onset of chills, fever, and prostration. Splenomegaly is characteristically absent; if present, the diagnosis of aplastic anemia should be seriously questioned. The red cells are usually slightly macrocytic and normochromic. Reticulocytopenia is the rule.

The diagnosis rests on examination of a bone marrow biopsy. It is important to distinguish aplastic anemia from other causes of pancytopenia, such as "aleukemic" leukemia and myelodysplastic syndrome (Chapter 13), which can have identical clinical manifestations. In aplastic anemia, the marrow is hypocellular (usually markedly so), whereas myeloid neoplasms are usually associated with hypercellular marrows filled with neoplastic progenitors.

The prognosis is variable. Stem cell transplantation is the treatment of choice in those with a suitable donor and provides a 5-year survival of more than 75%. Older patients or those without suitable donors often respond well to immunosuppressive therapy.

Pure Red Cell Aplasia

Pure red cell aplasia is a primary marrow disorder in which only erythroid progenitors are suppressed. In severe cases, red cell progenitors are completely absent from the marrow. It may occur in association with neoplasms, particularly thymoma and large granular lymphocytic leukemia (Chapter 13), drug exposures, autoimmune disorders, and parvovirus infection (see later). With the exception of those with parvovirus infection, it is likely that most cases have an autoimmune basis. When a thymoma is present, resection leads to hematologic improvement in about one-half of the patients. In patients without thymoma, immunosuppressive therapy is often beneficial. Plasmapheresis also may be helpful in unusual patients with neutralizing antibodies to erythropoietin, which may appear de novo or following the administration of recombinant erythropoietin.

A special form of red cell aplasia occurs in individuals infected with parvovirus B19, which preferentially infects and destroys red cell progenitors. Normal individuals clear parvovirus infections within 1 to 2 weeks; as a result, the aplasia is transient and clinically unimportant. However, as mentioned earlier, in persons with moderate to severe hemolytic anemias, even a brief cessation of erythropoiesis results in rapid worsening of the anemia, producing an aplastic crisis. In those who are severely immunosuppressed (e.g., persons with advanced HIV infection), an ineffective immune response sometimes permits the infection to persist, leading to chronic red cell aplasia and a moderate to severe anemia.

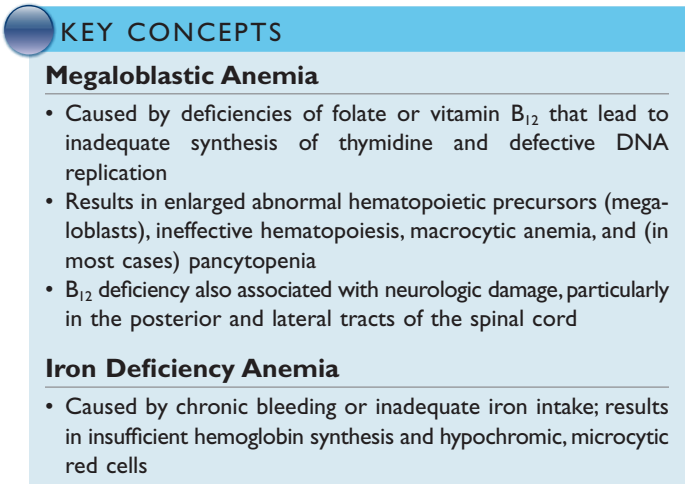
Other Forms of Marrow Failure

Myelophthisic anemia **describes a form of marrow failure in which space-occupying lesions replace normal marrow elements.** The most common cause is metastatic cancer, most often carcinoma arising in the breast, lung, and prostate. However, any infiltrative process (e.g., granulomatous disease) involving the marrow can produce identical findings. Myelophthisic anemia also is a feature of the spent phase of myeloproliferative neoplasms (Chapter 13). All of the responsible diseases cause marrow distortion and fibrosis, which displace normal marrow elements and disturb mechanisms that regulate the egress of red cells and granulocytes from the marrow. The latter effect causes the premature release of nucleated erythroid precursors and immature granulocytic forms *(leukoerythroblastosis)* into the circulation and the appearance of *teardrop-shaped red cells,* which are believed to be deformed during their tortuous escape from the fibrotic marrow.

Chronic renal failure, whatever its cause, is almost invariably associated with an anemia that tends to be roughly proportional to the severity of the uremia. The basis of anemia in renal failure is multifactorial, but the dominant cause is the diminished synthesis of erythropoietin by the damaged kidneys, leading to inadequate red cell production. Uremia also reduces red cell lifespan and impairs platelet function (both through uncertain mechanisms), and these effects may also contribute to anemia through extravascular hemolysis, abnormal bleeding, and eventually iron deficiency. Administration of recombinant erythropoietin and iron replacement therapy significantly improves the anemia.

Hepatocellular disease, whether toxic, infectious, or cirrhotic, is associated with anemia attributed to decreased marrow function. Folate and iron deficiencies caused by poor nutrition and excessive bleeding often exacerbate anemia in this setting. Erythroid progenitors are preferentially affected; depression of the white cell and platelet counts also occurs, but less frequently. The anemia is often slightly macrocytic due to lipid abnormalities associated with liver failure, which cause red cell membranes to acquire phospholipid and cholesterol as they circulate in the peripheral blood, thereby increasing cell size.

Endocrine disorders, particularly hypothyroidism, also may be associated with a mild normochromic, normocytic anemia.



Anemia of Chronic Inflammation

- Caused by inflammatory cytokines, which increase hepcidin levels, thereby reducing iron absorption and sequestering iron in macrophages, and also suppress erythropoietin production

Aplastic Anemia

- Caused by bone marrow failure (hypocellularity) due to diverse causes, including exposures to toxins and radiation, idiosyncratic reactions to drugs and viruses, and inherited defects in telomerase and DNA repair

Pure Red Cell Aplasia

- Acute: Parvovirus B19 infection
- Chronic: Associated with thymoma, large granular lymphocytic leukemia, presence of neutralizing antibodies against erythropoietin, and other autoimmune phenomenon

Other Causes of Underproduction Anemias

- Marrow replacement (tumors, granulomatous disease; so-called myelophthisic anemias), renal failure, endocrine disorders, liver failure

POLYCYTHEMIA

Polycythemia denotes an abnormally high number of circulating red cells, usually with a corresponding increase in the hemoglobin level. It may be relative (when there is hemoconcentration due to decreased plasma volume) or absolute (when there is an increase in the total red cell mass). Relative polycythemia results from dehydration, such as occurs with deprivation of water, prolonged vomiting or diarrhea, or excessive use of diuretics. It also is associated with a condition of unknown etiology called stress polycythemia, or *Gaisböck syndrome*. Affected individuals are usually males who are hypertensive, obese, and anxious ("stressed"). Absolute polycythemia is *primary* when it results from an intrinsic abnormality of hematopoietic precursors and *secondary* when it stems from the response of red cell progenitors to elevated levels of erythropoietin. A pathophysiologic classification of polycythemia divided along these lines is given in Table 14.8.

The most common cause of primary polycythemia is *polycythemia vera,* a myeloproliferative neoplasm associated with mutations that lead to erythropoietin-independent growth of red cell progenitors (Chapter 13). Much less commonly, primary polycythemia results from familial erythropoietin receptor mutations that induce erythropoietin-independent receptor activation. One such individual won Olympic gold medals in cross-country skiing, having benefited from this natural form of blood doping! Secondary polycythemia stems from compensatory or pathologic increases in erythropoietin secretion. Causes of the latter include erythropoietin-secreting tumors and rare (but illustrative) inherited defects in various components of the renal oxygen-sensing pathway. These defects stabilize HIF-1α, a transcription factor that stimulates the transcription of the erythropoietin gene.

Table 14.8 Pathophysiologic Classification of Polycythemia

Relative
Reduced plasma volume (hemoconcentration)
Absolute
Primary (Low Erythropoietin)
Polycythemia vera
Inherited erythropoietin receptor mutations (rare)
Secondary (High Erythropoietin)
Compensatory
Lung disease
High-altitude living
Cyanotic heart disease
Paraneoplastic
Erythropoietin-secreting tumors (e.g., renal cell carcinoma, hepatocellular carcinoma, cerebellar hemangioblastoma)
Hemoglobin mutants with high oxygen affinity
Inherited defects that stabilize HIF-1α
Chuvash polycythemia (homozygous *VHL* mutations)
Prolyl hydroxylase mutations

HIF-1α, Hypoxia-induced factor 1α.

BLEEDING DISORDERS: HEMORRHAGIC DIATHESES

Excessive bleeding can result from (1) increased fragility of vessels, (2) platelet deficiency or dysfunction, and (3) derangement of coagulation, alone or in combination. Before discussing specific bleeding disorders, it is helpful to review the common laboratory tests used in the evaluation of a bleeding diathesis. The normal hemostatic response involves the blood vessel wall, the platelets, and the clotting cascade (Chapter 4). The following tests are used to evaluate different aspects of hemostasis:

- *Prothrombin time (PT).* This test assesses the extrinsic and common coagulation pathways. The clotting of plasma after addition of an exogenous source of tissue thromboplastin (e.g., brain extract) and Ca^{2+} ions is measured in seconds. A prolonged PT can result from deficiency or dysfunction of factor V, factor VII, factor X, prothrombin, or fibrinogen.
- *Partial thromboplastin time (PTT).* This test assesses the intrinsic and common clotting pathways. The clotting of plasma after addition of kaolin, cephalin, and Ca^{2+} ions is measured in seconds. Kaolin activates the contact-dependent factor XII and cephalin substitutes for platelet phospholipids. Prolongation of the PTT can be due to deficiency or dysfunction of factors V, VIII, IX, X, XI, or XII, prothrombin, or fibrinogen, or to interfering antiphospholipid antibodies (Chapter 4).
- *Platelet counts.* These are obtained on anticoagulated blood using an electronic particle counter. The reference range is 150×10^3 to 350×10^3 platelets/μL. Abnormal platelet counts should be confirmed by inspection of a peripheral blood smear, as clumping of platelets during automated counting can cause spurious "thrombocytopenia." High counts may be indicative of a myeloproliferative neoplasm, such as essential thrombocythemia (Chapter 13), but are more likely to reflect reactive processes that increase platelet production (e.g., systemic inflammation).

- *Tests of platelet function.* At present, no single test provides an adequate assessment of the complex functions of platelets. Specialized tests that can be useful in particular clinical settings include tests of platelet aggregation, which measure the ability of platelets to adhere to one another in response to agonists like thrombin; and quantitative and qualitative tests of von Willebrand factor, which plays an important role in platelet adhesion to the extracellular matrix (Chapter 4). An older test, the bleeding time, is time-consuming and difficult to standardize and has been largely discarded. Instrument-based assays that provide quantitative measures of platelet function are used in some centers but remain imperfect at predicting bleeding risk, presumably because of difficulties in simulating in vivo clotting in the laboratory.

More specialized tests are available to measure the levels of specific clotting factors, fibrinogen, fibrin split products, and the presence of circulating anticoagulants.

Bleeding Disorders Caused by Vessel Wall Abnormalities

Disorders in this category are relatively common but do not usually cause serious bleeding problems. Most often, they present with small hemorrhages (petechiae and purpura) in the skin or mucous membranes, particularly the gingivae. On occasion, more significant hemorrhages occur into joints, muscles, and subperiosteal locations, or take the form of menorrhagia, nosebleeds, gastrointestinal bleeding, or hematuria. The platelet count and tests of coagulation (PT, PTT) are usually normal, pointing by exclusion to the underlying problem.

The clinical conditions in which vessel wall abnormalities cause bleeding include the following:

- *Infections* often induce petechial and purpuric hemorrhages, particularly meningococcemia, other forms of septicemia, infective endocarditis, and several of the rickettsioses. The involved mechanisms include microbial damage to the microvasculature (vasculitis) and disseminated intravscular coagulation.
- *Drug reactions* sometimes take the form of cutaneous petechiae and purpura without causing thrombocytopenia. In many instances the vascular injury is mediated by the deposition of drug-induced immune complexes in vessel walls, leading to hypersensitivity *(leukocytoclastic)* vasculitis (Chapter 11).
- *Scurvy* and the *Ehlers-Danlos syndrome* are associated with microvascular bleeding due to collagen defects that weaken vessel walls. Acquired vascular fragility accounts for the spontaneous purpura that are commonly seen in older adults and the skin hemorrhages that are seen with *Cushing syndrome,* in which the protein-wasting effects of excessive corticosteroid production cause loss of perivascular extracellular matrix.
- *Henoch-Schönlein purpura* is a systemic immune disorder characterized by purpura, colicky abdominal pain, polyarthralgia, and acute glomerulonephritis (Chapter 20). These changes result from the deposition of circulating immune complexes within vessels throughout the body and within the glomerular mesangial regions.
- *Hereditary hemorrhagic telangiectasia* (also known as *Weber-Osler-Rendu syndrome*) is an autosomal dominant disorder that can be caused by mutations in at least five different genes, most of which modulate TGF-β signaling. It is characterized by dilated, tortuous blood vessels with thin walls that bleed readily. Bleeding can occur anywhere, but it is most common under the mucous membranes of the nose (epistaxis), tongue, mouth, and eyes, and throughout the gastrointestinal tract.
- *Perivascular amyloidosis* can weaken blood vessel walls and cause bleeding. This complication is most common with amyloid light-chain (AL) amyloidosis (Chapter 6) and often manifests as mucocutaneous petechiae.

Among these conditions, serious bleeding is most often associated with hereditary hemorrhagic telangiectasia. The bleeding in each is nonspecific, and the diagnosis is based on the recognition of other more specific associated findings.

Bleeding Related to Reduced Platelet Number: Thrombocytopenia

Reduction in platelet number (thrombocytopenia) constitutes an important cause of generalized bleeding. A count less than 150,000 platelets/μL is generally considered to constitute thrombocytopenia. Platelet counts in the range of 20,000 to 50,000 platelets/μL can aggravate posttraumatic bleeding, and platelet counts less than 20,000 platelets/μL may be associated with spontaneous (nontraumatic) bleeding. When thrombocytopenia is isolated, the PT and PTT are normal.

You will recall that following a vascular injury, platelets adhere and aggregate to form the primary hemostatic plug and also promote key reactions in the coagulation cascade that lead to secondary hemostasis and formation of a fibrin clot (Chapter 4). Spontaneous bleeding associated with thrombocytopenia most often involves small vessels. Common sites for such hemorrhages are the skin and the mucous membranes of the gastrointestinal and genitourinary tracts. Most feared, however, is intracranial bleeding, which is a threat to any patient with a markedly depressed platelet count.

The causes of thrombocytopenia fall into four major categories (Table 14.9).

- *Decreased platelet production.* This can result from conditions that depress marrow output generally (such as aplastic anemia and leukemia) or affect megakaryocytes selectively. Examples of the latter include certain drugs and alcohol, which may suppress platelet production through uncertain mechanisms when taken in large amounts; HIV, which may infect megakaryocytes and inhibit platelet production; and myelodysplastic syndrome (Chapter 13), which occasionally presents with isolated thrombocytopenia.
- *Decreased platelet survival.* This important mechanism of thrombocytopenia may have an immunologic or nonimmunologic basis. In immune thrombocytopenia, destruction is caused by the deposition of antibodies or immune complexes on platelets. Autoimmune thrombocytopenia is discussed in the following section. Alloimmune thrombocytopenia can arise when platelets are transfused or when platelets cross the placenta from the fetus into the pregnant mother. In the latter case, IgG antibodies

Table 14.9 Causes of Thrombocytopenia

Decreased Production of Platelets
Selective impairment of platelet production
Drug-induced: alcohol, thiazides, cytotoxic drugs
Infections: measles, human immunodeficiency virus (HIV)
Nutritional deficiencies
B_{12}, folate deficiency (megaloblastic leukemia)
Aplastic anemia (see Table 14.7)
Bone marrow replacement
Leukemia, disseminated cancer, granulomatous disease
Ineffective hematopoiesis
Myelodysplastic syndromes (Chapter 13)
Decreased Platelet Survival
Immunologic destruction
Chromic immune thrombocytopenic purpura
Acute immune thrombocytopenic purpura
Systemic lupus erythematosus, B-cell lymphoid neoplasms
Alloimmune: posttransfusion and neonatal
Drug-associated: quinidine, heparin, sulfa compounds
Infections: HIV, infectious mononucleosis (transient, mild), dengue fever
Nonimmunologic destruction
Disseminated intravascular coagulation
Thrombotic microangiopathies
Giant hemangiomas
Sequestration
Hypersplenism
Dilution
Transfusions

made in the mother may cause clinically significant thrombocytopenia in the fetus. This is reminiscent of hemolytic disease of the newborn, in which red cells are the target (Chapter 10). The most important nonimmunologic causes are *disseminated intravascular coagulation (DIC)* and the *thrombotic microangiopathies,* in which unbridled, often systemic, platelet activation reduces platelet life span. Nonimmunologic destruction of platelets may also be caused by mechanical injury, such as in individuals with prosthetic heart valves.

- *Sequestration.* The spleen normally sequesters 30% to 35% of the body's platelets, but this can rise to 80% to 90% when the spleen is enlarged, producing moderate degrees of thrombocytopenia.
- *Dilution.* Massive transfusions can produce dilutional thrombocytopenia.

Chronic Immune Thrombocytopenic Purpura

Chronic immune thrombocytopenic purpura (ITP) is caused by autoantibody-mediated destruction of platelets. It can occur in the setting of a variety of predisposing conditions and exposures (secondary) or in the absence of any known risk factors (primary or idiopathic). The contexts in which chronic ITP occurs secondarily are numerous and include individuals with systemic lupus erythematosus (Chapter 6), HIV infection, and B-cell neoplasms such as chronic lymphocytic leukemia (Chapter 13). The diagnosis of primary chronic ITP is made only after secondary causes are excluded.

Pathogenesis

Autoantibodies, most often directed against platelet membrane glycoproteins IIb–IIIa or Ib–IX, can be demonstrated in the plasma and bound to the platelet surface in about 80% of patients. In the overwhelming majority of cases, the antiplatelet antibodies are of the IgG class. As in autoimmune hemolytic anemias, antiplatelet antibodies act as opsonins that are recognized by IgG Fc receptors expressed on phagocytes (Chapter 6), leading to increased platelet destruction. The thrombocytopenia is usually markedly improved by splenectomy, indicating that the spleen is the major site of removal of opsonized platelets. The splenic red pulp also is rich in plasma cells, and part of the benefit of splenectomy may stem from the removal of a source of autoantibodies. In some instances the autoantibodies may also bind to and damage megakaryocytes, leading to decreases in platelet production that further exacerbate the thrombocytopenia.

MORPHOLOGY

The principal changes of thrombocytopenic purpura are found in the spleen, bone marrow, and blood, but they are not specific. Secondary changes related to the bleeding diathesis may be found anywhere in the body. The spleen is of normal size. Typically, there is congestion of the sinusoids and enlargement of the splenic follicles, often associated with prominent reactive germinal centers. In many instances scattered megakaryocytes are found within the sinuses, possibly representing a mild form of extramedullary hematopoiesis driven by elevated levels of thrombopoietin. **The marrow reveals a modestly increased number of megakaryocytes.** Some are apparently immature, with large, nonlobulated, single nuclei. These findings are not specific but merely reflect accelerated thrombopoiesis, being found in most forms of thrombocytopenia resulting from increased platelet destruction. The importance of bone marrow examination is to rule out thrombocytopenias resulting from marrow failure or other primary marrow disorders. The secondary changes relate to hemorrhage, often in the form of **petechial bleeds into the skin and mucous membranes.** The **peripheral blood often reveals abnormally large platelets** (megathrombocytes), which are a sign of accelerated thrombopoiesis.

Clinical Features

Chronic ITP occurs most commonly in adult women younger than 40 years of age. The female-to-male ratio is 3:1. It is often insidious in onset and is characterized by bleeding into the skin and mucosal surfaces. Pinpoint hemorrhages *(petechiae)* are especially prominent in the dependent areas where the capillary pressure is higher. Petechiae can become confluent, giving rise to *ecchymoses.* Often there is a history of easy bruising, nosebleeds, gingival bleeding, and hemorrhages into soft tissues from relatively minor trauma. The disease may manifest first with melena, hematuria, or excessive menstrual flow. Subarachnoid hemorrhage and intracerebral hemorrhage are serious and sometimes fatal complications, but fortunately are rare in treated patients. Splenomegaly and lymphadenopathy are not seen in primary disease, and their presence should lead one to consider other diagnoses, such as ITP secondary to a B-cell neoplasm.

Typical laboratory findings are reflective of isolated thrombocytopenia. A low platelet count, normal or increased megakaryocytes in the bone marrow, and large platelets in the peripheral blood are taken as presumptive evidence of

accelerated platelet destruction. The PT and PTT are normal. Tests for platelet autoantibodies suffer from low sensitivity and specificity and are not clinically useful. Therefore, the diagnosis is one of exclusion and can be made only after other causes of thrombocytopenia (such as those listed in Table 14.9) have been ruled out.

Almost all patients respond to glucocorticoids (which inhibit phagocyte function), but many relapse following withdrawal of steroids. Those with moderately severe thrombocytopenia (platelet counts >30,000/mL) can be followed carefully, and in some of these individuals ITP may spontaneously remit. In individuals with severe thrombocytopenia, splenectomy normalizes the platelet count in about two-thirds of patients, but with the attendant increased risk of bacterial sepsis. Immunomodulatory agents such as intravenous immunoglobulin or anti-CD20 antibody (rituximab) are often effective in patients who relapse after splenectomy or for whom splenectomy is contraindicated. Peptides that mimic the effects of thrombopoietin (so-called *TPO-mimetics*) also may be effective in improving platelet counts in individuals with disease that is refractory to other treatments.

Acute Immune Thrombocytopenic Purpura

Like chronic ITP, this condition is caused by autoantibodies to platelets, but its clinical features and course are distinct. Acute ITP is mainly a disease of childhood occurring with equal frequency in both sexes. Symptoms appear abruptly, often 1 to 2 weeks after a self-limited viral illness, which appears to trigger the development of autoantibodies through uncertain mechanisms. Unlike chronic ITP, acute ITP is self-limited, usually resolving spontaneously within 6 months. Glucocorticoids are given only if the thrombocytopenia is severe. In about 20% of children, usually those without a viral prodrome, thrombocytopenia persists; these children have a childhood form of chronic ITP that follows a course similar to the adult disease.

Drug-Induced Thrombocytopenia

Drugs can induce thrombocytopenia through direct effects on platelets and secondary to immunologically mediated platelet destruction. The drugs most commonly implicated are quinine, quinidine, and vancomycin, all of which induce drug-dependent antibody binding to platelet glycoproteins. Much more rarely, drugs induce true autoantibodies through unknown mechanisms. Thrombocytopenia, which may be severe, can occur in those who are taking platelet inhibitory drugs that bind glycoprotein IIb/IIIa; it is hypothesized that these drugs induce conformational changes in glycoprotein IIb/IIIa and create an immunogenic epitope.

Heparin-induced thrombocytopenia (HIT) has a distinctive pathogenesis and is of particular importance because of its potential for severe clinical consequences (Chapter 4). Thrombocytopenia occurs in about 5% of persons receiving heparin and is of two types:

- *Type I HIT occurs rapidly after the onset of therapy and is of little clinical importance,* sometimes resolving despite the continuation of therapy. It most likely results from a direct platelet-aggregating effect of heparin.
- *Type II HIT is less common but is often life threatening.* It occurs 5 to 14 days after therapy begins (or sooner if the person has been sensitized to heparin) and, paradoxically, often leads to venous and arterial thrombosis. It is caused by antibodies that recognize complexes of heparin and platelet factor 4, a normal component of platelet granules. Binding of antibody to these complexes activates platelets and promotes thrombosis, even in the setting of thrombocytopenia. Failure to immediately stop heparin and institute an alternative nonheparin anticoagulant may lead to clot formation within large arteries, vascular insufficiency, and limb loss, as well as deep venous thrombosis with the attendant risk of pulmonary thromboembolism. The risk of type II HIT is lowered, but not completely eliminated, by the use of low-molecular-weight heparin preparations. Once type II HIT develops, even low-molecular-weight heparins exacerbate the thrombotic tendency and must be avoided.

HIV-Associated Thrombocytopenia

Thrombocytopenia is one of the most common hematologic manifestations of HIV infection. Both decreased platelet production and increased platelet destruction contribute. CD4 and CXCR4, the receptor and co-receptor, respectively, for HIV, are found on megakaryocytes, allowing these cells to be infected. HIV-infected megakaryocytes are prone to apoptosis, and their ability to produce platelets is impaired. HIV infection also causes B-cell hyperplasia and dysregulation (possibly due to its direct effects on CD4+ T cells), which predisposes to the development of autoantibodies. In some instances, the antibodies are directed against platelet membrane glycoprotein IIb–III complexes. As in other immune cytopenias, the autoantibodies opsonize platelets, promoting their destruction by mononuclear phagocytes in the spleen and elsewhere. The deposition of immune complexes on platelets also may contribute thrombocytopenia in some HIV-infected patients.

Thrombotic Microangiopathies: Thrombotic Thrombocytopenic Purpura (TTP) and Hemolytic Uremic Syndrome (HUS)

Thrombotic microangiopathies encompasses a spectrum of clinical syndromes that are caused by insults that lead to excessive activation of platelets, which deposit as thrombi in small blood vessels. This group of disorders includes TTP and HUS.

According to its original description, TTP was defined as the pentad of fever, thrombocytopenia, microangiopathic hemolytic anemia, transient neurologic deficits, and renal failure. HUS also is associated with microangiopathic hemolytic anemia and thrombocytopenia but is distinguished by the absence of neurologic symptoms, the prominence of acute renal failure, and its frequent occurrence in children. With time, experience, and increased mechanistic insight, however, these distinctions have blurred. Many adult patients with "TTP" lack one or more of the five criteria, and some patients with "HUS" have fever and neurologic dysfunction.

In both conditions, intravascular thrombi cause a microangiopathic hemolytic anemia and widespread organ dysfunction, and the attendant consumption of platelets leads to thrombocytopenia. It is believed that the varied clinical manifestations of TTP and HUS are related to differing proclivities for thrombus formation in tissues. Although DIC (discussed later) and thrombotic microangiopathies share features such as microvascular occlusion and

microangiopathic hemolytic anemia, they are pathogenically distinct. In TTP and HUS (unlike in DIC), activation of the coagulation cascade is not of primary importance, and hence laboratory tests of coagulation, such as the PT and PTT, are usually normal.

Although certain features of the various thrombotic microangiopathies overlap, the triggers for the pathogenic platelet activation are distinctive and provide a more satisfying and clinically relevant way of thinking about these disorders (summarized in Table 14.10). **TTP is caused by a deficiency in a plasma enzyme called ADAMTS13, also designated "vWF metalloprotease."** ADAMTS13 degrades very high-molecular-weight multimers of vWF. In its absence, large multimers accumulate in plasma and tend to promote spontaneous platelet activation and aggregation. Superimposition of endothelial cell injury (caused by some other condition) may further promote the formation of platelet aggregates, thus initiating or exacerbating clinically evident TTP.

ADAMTS13 deficiency may be inherited or acquired. In the acquired form, an autoantibody that inhibits the metalloprotease activity of ADAMTS13 is present. Less commonly, patients inherit an inactivating mutation in *ADAMTS13*. In those with hereditary ADAMTS13 deficiency, the onset is often delayed until adolescence, and the symptoms are episodic. Thus, factors other than ADAMTS13 deficiency (e.g., a superimposed vascular injury or prothrombotic state) must be involved in triggering full-blown TTP.

TTP is an important diagnosis to consider in any patient presenting with thrombocytopenia and microangiopathic hemolytic anemia, because delays in diagnosis can be fatal. With plasma exchange, which removes autoantibodies and provides functional ADAMTS13, TTP (which once was uniformly fatal) can be treated successfully in more than 80% of patients.

In contrast, HUS is associated with normal levels of ADAMTS13 and is initiated by several other distinct defects. **"Typical" HUS is strongly associated with infectious gastroenteritis caused by *Escherichia coli* strain O157:H7, which elaborates a Shiga-like toxin.** This toxin is absorbed from the inflamed gastrointestinal mucosa into the circulation, where it is believed to directly or indirectly alter endothelial cell function in some manner that provokes platelet activation and aggregation. Children and older adults are at highest risk. Those affected present with bloody diarrhea, and a few days later HUS makes its appearance.

"Atypical" HUS is often associated with defects in complement factor H, membrane cofactor protein (CD46), or factor I, proteins that act to prevent excessive activation of the alternative complement pathway. Deficiencies of these proteins may be caused by inherited defects or acquired inhibitory autoantibodies and are associated with a remitting, relapsing clinical course.

Unlike TTP, the basis for the platelet activation in typical and atypical HUS is unclear. As in paroxysmal nocturnal hemoglobinuria (discussed earlier), therapeutic antibodies that inhibit the activation of the complement factor C5 are effective in preventing thrombosis in patients with atypical HUS, proving that excessive complement activation has a central role in this form of the disease. Immunosuppression also may be beneficial to patients with inhibitory autoantibodies. Typical HUS is treated supportively. Patients who survive the acute insult usually recover, but some have permanent renal damage and require dialysis or renal transplantation. The impact of HUS and TTP on the kidneys is discussed further in Chapter 20.

Thrombotic microangiopathies resembling HUS also may be seen following exposures to other agents that damage endothelial cells (e.g., certain drugs and radiation therapy). The prognosis in these settings is guarded, because the HUS is often complicated by chronic, life-threatening conditions.

Table 14.10 Thrombotic Microangiopathies: Causes and Associations

Thrombotic Thrombocytopenic Purpura
Deficiency of ADAMTS13 Inherited Acquired (autoantibodies)
Hemolytic Uremic Syndrome
Typical: Escherichia coli *strain O157:H7 infection* Endothelial damage by Shiga-like toxin
Atypical: alternative complement pathway inhibitor deficiencies (complement factor H, membrane cofactor protein [CD46], or factor I) Inherited Acquired (autoantibodies)
Miscellaneous associations Drugs (cyclosporine, chemotherapeutic agents) Radiation, bone marrow transplantation Other infections (HIV, pneumococcal sepsis) Conditions associated with autoimmunity (systemic lupus erythematosus, HIV infection, lymphoid neoplasms)

HIV, Human immunodeficiency virus.

Bleeding Disorders Related to Defective Platelet Function

Qualitative defects of platelet function can be inherited or acquired. Several inherited disorders characterized by abnormal platelet function and normal platelet counts have been described. A brief discussion of these rare diseases is warranted because they provide insight into the molecular mechanisms of platelet function.

Inherited disorders of platelet function can be classified into three pathogenically distinct groups: (1) defects of adhesion, (2) defects of aggregation, and (3) disorders of platelet secretion (release reaction).

- *Bernard-Soulier syndrome* illustrates the consequences of defective adhesion of platelets to subendothelial matrix. It is an autosomal recessive disorder caused by an inherited deficiency of the platelet membrane glycoprotein complex Ib–IX. This glycoprotein is a receptor for vWF and is essential for normal platelet adhesion to the subendothelial extracellular matrix (Chapter 4). Affected patients have a variable, often severe, bleeding tendency.
- *Glanzmann thrombasthenia* exemplifies bleeding due to defective platelet aggregation. It is transmitted as an autosomal recessive trait. Thrombasthenic platelets fail to aggregate in response to adenosine diphosphate (ADP), collagen, epinephrine, or thrombin because of deficiency or dysfunction of glycoprotein IIb–IIIa, an integrin that participates in "bridge formation" between platelets by binding fibrinogen. The associated bleeding tendency is often severe.

- *Storage pool disorders* are characterized by the defective release of certain mediators of platelet activation, such as thromboxanes and granule-bound ADP. The biochemical defects underlying these disorders are varied, complex, and beyond the scope of our discussion.

Among the acquired defects of platelet function, two are clinically significant. The first is caused by ingestion of aspirin and other nonsteroidal anti-inflammatory drugs. Aspirin is a potent, irreversible inhibitor of cyclooxygenase, an enzyme that is required for the synthesis of thromboxane A_2 and prostaglandins (Chapter 3). These mediators play important roles in platelet aggregation and subsequent release reactions (Chapter 4). The antiplatelet effects of aspirin form the basis for its use in preventing coronary thrombosis (Chapter 12). Uremia (Chapter 20) is the second common condition associated with acquired defects in platelet function. The pathogenesis of platelet dysfunction in uremia is complex and involves defects in adhesion, granule secretion, and aggregation.

Hemorrhagic Diatheses Related to Abnormalities in Clotting Factors

Inherited or acquired deficiencies of virtually every coagulation factor have been reported as causes of bleeding diatheses. Bleeding due to coagulation factor deficiencies commonly manifest as large posttraumatic ecchymoses or hematomas, or prolonged bleeding from a laceration or after a surgical procedure. Unlike bleeding seen with thrombocytopenia, bleeding due to coagulation factor deficiencies often occurs into the gastrointestinal and urinary tracts and into weight-bearing joints (hemarthrosis). Typical stories include the patient who oozes blood for days after a tooth extraction or who develops a hemarthrosis after minor stress on a knee joint.

Hereditary deficiencies typically affect a single clotting factor. The most common and important inherited deficiencies of coagulation factors affect factor VIII (hemophilia A) and factor IX (hemophilia B). Deficiencies of vWF (von Willebrand disease) also are discussed here, as this factor influences both coagulation and platelet function.

Acquired deficiencies usually involve multiple coagulation factors and can be caused by decreased protein synthesis or a shortened protein half-life. Vitamin K deficiency (Chapter 9) impairs the synthesis of factors II, VII, IX, X and protein C. Many coagulatiion factors are made in the liver, and inadequate synthesis is often observed in severe parenchymal liver disease. By contrast, in DIC, multiple coagulation factors are consumed, leading to their deficiency. Acquired deficiencies of single factors occur, but are rare. These are usually caused by inhibitory autoantibodies.

Factor VIII–vWF Complex

The two most common inherited disorders of bleeding, hemophilia A and von Willebrand disease, are caused by qualitative or quantitative defects involving factor VIII and vWF, respectively. Before we discuss these disorders, it will be helpful to review the structure and function of these two proteins, which exist together in the plasma as part of a single large complex.

Factor VIII is an essential cofactor of factor IX, which converts factor X to factor Xa (Fig. 14.26; Chapter 4). It is made by endothelial cells, whereas vWF is made by both endothelial cells and megakaryocytes, which are the source of the vWF that is present in platelet α-granules. Once secreted into the blood, factor VIII binds to and is stabilized by vWF, an interaction that increases the half-life of factor VIII from about 2.4 hours to about 12 hours.

vWF secreted into the circulation by endothelial cells exists as multimers containing as many as 100 subunits that can exceed 20×10^6 daltons in molecular mass. Some of the

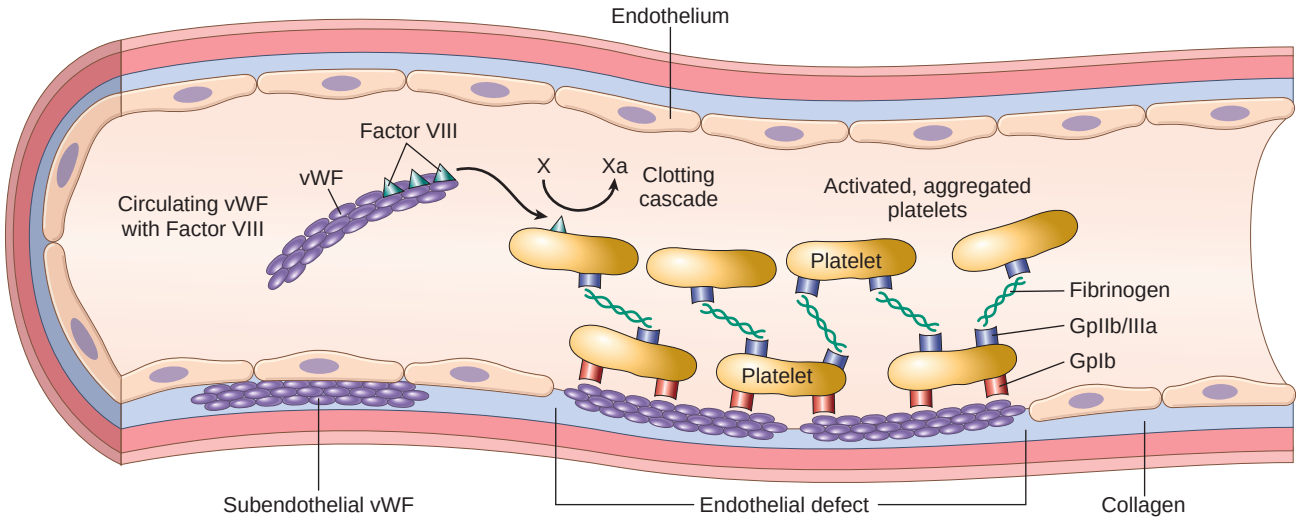


Figure 14.26 Structure and function of factor VIII–von Willebrand factor (vWF) complex. Factor VIII is synthesized in the liver and kidney, and vWF is made in endothelial cells and megakaryocytes. The two associate to form a complex in the circulation. vWF is also present in the subendothelial matrix of normal blood vessels and the α-granules of platelets. Following endothelial injury, exposure of subendothelial vWF causes adhesion of platelets, primarily via the glycoprotein Ib (GpIb) platelet receptor. Circulating vWF and vWF released from the α-granules of activated platelets can bind exposed subendothelial matrix, further contributing to platelet adhesion and activation. Activated platelets form hemostatic aggregates; fibrinogen participates in aggregation through bridging interactions with the glycoprotein IIb/IIIa (GpIIb/IIIa) platelet receptor. Factor VIII takes part in the coagulation cascade as a cofactor in the activation of factor X on the surface of activated platelets.

secreted vWF also is deposited in the subendothelial matrix, where it lies ready to promote platelet adhesion if the endothelial lining is disrupted (see Fig. 14.26). In addition to factor VIII, vWF interacts with several other proteins involved in hemostasis, including collagen, heparin, and platelet membrane glycoproteins. The most important hemostatic function of vWF is to promote the adhesion of platelets to the subendothelial matrix. This occurs through bridging interactions between platelet glycoprotein Ib–IX, vWF, and matrix components such as collagen. vWF also may promote platelet aggregation by binding to activated GpIIb/IIIa integrins; this activity may be of particular importance under conditions of high shear stress (such as occurs in small vessels).

Factor VIII and vWF protein levels are measured by immunologic techniques. Factor VIII function is measured by performing coagulation assays with mixtures of patient plasma and factor VIII–deficient plasma, and vWF function is assessed using the ristocetin agglutination test. The latter is performed by adding patient plasma to normal formalin-fixed platelets in the presence of ristocetin, a small molecule that binds and "activates" vWF. Ristocetin induces multivalent vWF multimers to bind platelet glycoprotein Ib–IX and form interplatelet "bridges." The resultant clumping (agglutination) of platelets reflects the vWF activity of the sample.

Von Willebrand Disease

Von Willebrand disease is the most common inherited bleeding disorder of humans, affecting about 1% of adults in the United States. The bleeding tendency is usually mild and often goes unnoticed until some hemostatic stress, such as surgery or a dental procedure, reveals its presence. The most common presenting symptoms are spontaneous bleeding from mucous membranes (e.g., epistaxis), excessive bleeding from wounds, or menorrhagia. It is usually transmitted as an autosomal dominant disorder, but rare autosomal recessive variants also exist.

Von Willebrand disease is clinically and molecularly heterogeneous; hundreds of vWF variants have been described, only a few of which have been formally proven to cause disease. Three types are recognized, each with a range of phenotypes:

- *Type 1 and type 3 von Willebrand disease are associated with quantitative defects in vWF.* Type 1, an autosomal dominant disorder characterized by a mild to moderate vWF deficiency, accounts for about 70% of all cases. Incomplete penetrance and variable expressivity are commonly observed, but it generally is associated with mild disease. Type 1 disease is associated with a spectrum of mutations, including point substitutions that interfere with maturation of the vWF protein or that result in rapid clearance from the plasma. Type 3 disease is an autosomal disorder usually caused by deletions or frameshift mutations involving both alleles, resulting in little to no vWF synthesis. Because vWF stabilizes factor VIII in the circulation, factor VIII levels are also reduced in type 3 disease and the associated bleeding disorder is often severe.
- *Type 2 von Willebrand disease is characterized by qualitative defects in vWF;* there are several subtypes, of which type 2A is the most common. It is inherited as an autosomal dominant disorder. vWF is expressed in normal amounts, but missense mutations are present that lead to defective multimer assembly. As a result, large- and intermediate-sized multimers, the most active forms of vWF, are missing from plasma. Type 2 von Willebrand disease accounts for 25% of all cases and is associated with mild to moderate bleeding.

Patients with von Willebrand disease have defects in platelet function despite having normal platelet counts. The plasma level of active vWF, measured as the ristocetin cofactor activity, is reduced. Because a deficiency of vWF decreases the stability of factor VIII, type 1 and type 3 von Willebrand disease are associated with a prolonged PTT.

Even within families in which a single defective vWF allele is segregating, wide variability in clinical expression is common. This is due in part to modifying genes that influence circulating levels of vWF, which have a broad range in normal populations. Persons with types 1 or 2 von Willebrand disease facing hemostatic challenges (dental work, surgery) can be treated with desmopressin (which stimulates vWF release), infusions of plasma concentrates containing factor VIII and vWF, or with recombinant vWF. By contrast, rare patients with type 3 disease must be treated prophylactically with plasma concentrates and factor VIII infusions to prevent severe "hemophilia-like" bleeding.

Hemophilia A (Factor VIII Deficiency)

Hemophilia A, the most common hereditary disease associated with life-threatening bleeding, is caused by mutations in factor VIII, an essential cofactor for factor IX in the coagulation cascade. Hemophilia A is inherited as an X-linked recessive trait and thus affects mainly males and homozygous females. Rarely, excessive bleeding occurs in heterozygous females, presumably as a result of inactivation of the X chromosome bearing the normal factor VIII allele by chance in most cells (unfavorable lyonization). About 30% of patients have no family history; their disease is caused by new mutations.

Hemophilia A exhibits a wide range of clinical severity that correlates well with the level of factor VIII activity. Those with less than 1% of normal levels have severe disease; those with 2% to 5% of normal levels have moderately severe disease; and those with 6% to 50% of normal levels have mild disease. The varying degrees of factor VIII deficiency are largely explained by heterogeneity in the causative mutations. The most severe deficiencies result from an inversion involving the X chromosome that completely abolishes the synthesis of factor VIII. Less commonly, severe hemophilia A is associated with point mutations in factor VIII that impair the function of the protein. In such cases, factor VIII protein levels may be normal by immunoassay. Mutations permitting some active factor VIII to be synthesized are associated with mild to moderate disease.

In all symptomatic cases, there is a tendency toward easy bruising and massive hemorrhage after trauma or operative procedures. In addition, "spontaneous" hemorrhages frequently occur in regions of the body that are susceptible to trauma, particularly the joints, where they are known as *hemarthroses.* Recurrent bleeding into the joints leads to progressive deformities that can be crippling. Petechiae are characteristically absent.

Patients with hemophilia A have a prolonged PTT and a normal PT, results that point to an abnormality of the

intrinsic coagulation pathway. Factor VIII–specific assays are required for diagnosis. As explained in Chapter 4, the bleeding diathesis reflects the pre-eminent role of the factor VIIIa/factor IXa complex in activation of factor X in vivo. The precise explanation for the tendency of hemophiliacs to bleed at particular sites (joints, muscles, and the central nervous system) remains uncertain.

Hemophilia A is treated with infusions of recombinant factor VIII. About 15% of patients with severe hemophilia A develop antibodies that bind and inhibit factor VIII, probably because the protein is perceived as foreign, having never been "seen" by the immune system. Recently, bispecific antibodies have been developed that bind factor IXa to factor X; these antibodies bypass the need for factor VIII and are particularly effective in patients with factor VIII antibody inhibitors. Before the development of recombinant factor VIII therapy, thousands of hemophiliacs received plasma-derived factor VIII concentrates containing HIV, and many developed AIDS (Chapter 6). The risk of HIV transmission has been eliminated, but tragically too late for an entire generation of hemophiliacs. Efforts to develop somatic gene therapy for hemophilia are ongoing.

Hemophilia B (Christmas Disease, Factor IX Deficiency)

Severe factor IX deficiency produces a disorder clinically indistinguishable from factor VIII deficiency (hemophilia A). This should not be surprising, given that factors VIII and IX function together to activate factor X. A wide spectrum of mutations involving the gene that encodes factor IX is found in hemophilia B. Like hemophilia A, it is inherited as an X-linked recessive trait and shows variable clinical severity. In about 15% of these patients, factor IX protein is present but is nonfunctional. As with hemophilia A, the PTT is prolonged and the PT is normal. Diagnosis of Christmas disease (named after the first patient identified with this condition, and not the holiday) is possible only by assay of the factor levels. The disease is treated with infusions of recombinant factor IX.

Disseminated Intravascular Coagulation (DIC)

DIC is an acute, subacute, or chronic thrombohemorrhagic disorder characterized by the excessive activation of coagulation and the formation of thrombi in the microvasculature. It occurs as a secondary complication of many disorders. Sometimes the coagulopathy is localized to a specific organ or tissue. As a consequence of the thrombotic diathesis, there is consumption of platelets, fibrin, and coagulation factors and, secondarily, activation of fibrinolysis. DIC can present with signs and symptoms relating to tissue hypoxia and infarction caused by microthrombi; hemorrhage, due to depletion of factors required for hemostasis and activation of fibrinolytic mechanisms; or both.

Etiology and Pathogenesis

At the outset, it must be emphasized that DIC is not a primary disease, but rather is an acquired coagulopathy that may occur in the course of a variety of clinical conditions. In discussing the general mechanisms underlying DIC, it is useful to briefly review the normal process of blood coagulation and clot removal (Chapter 4).

Clotting in vivo is initiated by exposure of tissue factor, which activates factor VII. The most important effect of tissue factor/factor VII complexes is activation of factor IX, which in turn activates factor X. Activation of factor X leads to the generation of thrombin, the central player in clotting. At sites where the endothelium is disrupted, thrombin converts fibrinogen to fibrin; feeds back to activate factors IX, VIII, and V; stimulates fibrin crosslinking; inhibits fibrinolysis; and activates platelets, all of which augment the formation of a stable clot. To prevent runaway clotting, this process must be sharply limited to the site of tissue injury. Remarkably, as thrombin is swept away in the bloodstream and encounters uninjured vessels, it is converted to an anticoagulant through binding to thrombomodulin, a protein found on the surface of endothelial cells. The thrombin-thrombomodulin complex activates protein C, an important inhibitor of factor V and factor VIII. Other activated coagulation factors are removed from the circulation by the liver, and, as you will recall, the blood also contains several potent fibrinolytic factors, such as plasmin. These and additional checks and balances normally ensure that just enough clotting occurs at the right place and time.

From this brief review, it should be clear that DIC could result from pathologic activation of coagulation or impairment of clot-inhibiting mechanisms. Because the latter rarely constitute primary mechanisms of DIC, we will focus on the abnormal initiation of clotting.

Two major mechanisms trigger DIC: (1) release of tissue factor or other procoagulants into the circulation, and (2) widespread injury of endothelial cells. Procoagulants such as tissue factor can be derived from a variety of sources, such as the placenta in obstetric complications or tissues injured by trauma or burns. Mucus released from certain adenocarcinomas may also act as a procoagulant by directly activating factor X.

Endothelial injury can initiate DIC in several ways. Injuries that cause endothelial cell necrosis expose the subendothelial matrix, leading to the activation of platelets and the coagulation pathway. However, even subtle endothelial injuries can unleash procoagulant activity. One mediator of endothelial injury is TNF, which is implicated in DIC occurring with sepsis. TNF induces endothelial cells to express tissue factor on their cell surfaces and to decrease the expression of thrombomodulin, tilting the checks and balances that govern hemostasis towards coagulation. In addition, TNF up-regulates the expression of adhesion molecules on endothelial cells, thereby promoting the adhesion of leukocytes, which can damage endothelial cells by releasing reactive oxygen species and preformed proteases. Widespread endothelial injury may also be produced by deposition of antigen-antibody complexes (e.g., systemic lupus erythematosus), temperature extremes (e.g., heat stroke, burns), or microorganisms (e.g., meningococci, rickettsiae).

DIC is most commonly associated with obstetric complications, malignant neoplasms, sepsis, and major trauma. The triggers in these conditions are often multiple and interrelated. For example, in bacterial infections, endotoxins can inhibit the endothelial expression of thrombomodulin directly or indirectly by stimulating immune cells to make TNF, and also can activate factor XII. Antigen-antibody complexes formed in response to infection can activate the classical complement pathway, giving rise to complement fragments

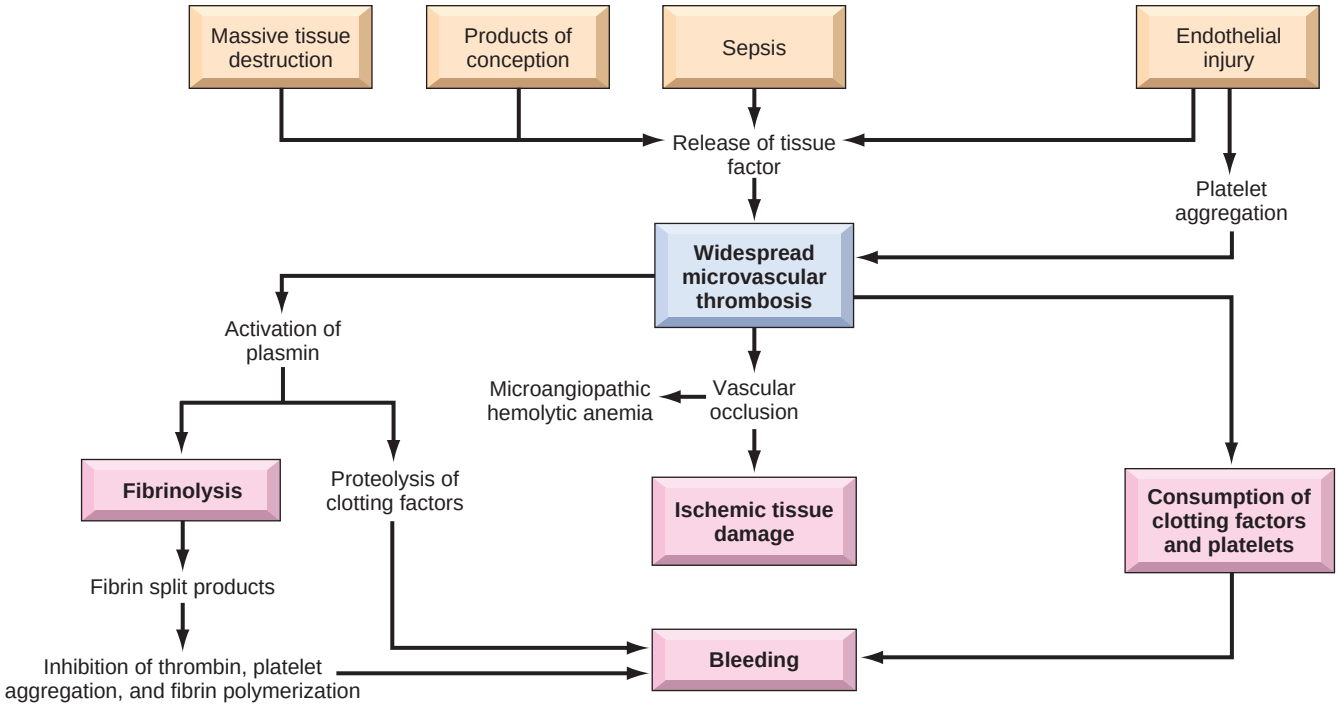


Figure 14.27 Pathophysiology of disseminated intravascular coagulation.

that secondarily activate platelets and granulocytes. In massive trauma, extensive surgery, and severe burns, the major trigger is the release of procoagulants such as tissue factor. In obstetric conditions, procoagulants derived from the placenta, dead retained fetus, or amniotic fluid may enter the circulation. Hypoxia, acidosis, and shock, which often coexist in very ill patients, also can cause widespread endothelial injury, and supervening infections can complicate problems further. Among cancers, acute promyelocytic leukemia and adenocarcinomas of the lung, pancreas, colon, and stomach are most frequently associated with DIC.

The possible consequences of DIC are twofold (Fig. 14.27).

- *Widespread deposition of fibrin* within the microcirculation. This leads to ischemia of the more severely affected or more vulnerable organs and a microangiopathic hemolytic anemia, which results from the fragmentation of red cells as they squeeze through the narrowed microvasculature.
- *Consumption of platelets and clotting factors* and the activation of plasminogen, leading to a hemorrhagic diathesis. Plasmin not only cleaves fibrin, but it also digests factors V and VIII, reducing their concentration further. In addition, fibrin degradation products resulting from fibrinolysis inhibit platelet aggregation, fibrin polymerization, and thrombin.

MORPHOLOGY

Thrombi are most often found in the brain, heart, lungs, kidneys, adrenals, spleen, and liver, in decreasing order of frequency, but any tissue can be affected. Affected kidneys may have small thrombi in the glomeruli that evoke only reactive swelling of endothelial cells or, in severe cases, microinfarcts or even **bilateral renal cortical necrosis.** Numerous fibrin thrombi may be found in alveolar capillaries, sometimes associated with pulmonary edema and fibrin exudation, creating "hyaline membranes" reminiscent of acute respiratory distress syndrome (Chapter 15). In the central nervous system, fibrin thrombi may cause microinfarcts, occasionally complicated by simultaneous hemorrhage, which can sometimes lead to variable neurologic signs and symptoms. In meningococcemia, fibrin thrombi within the microcirculation of the adrenal cortex are the probable basis for the massive adrenal hemorrhages seen in **Waterhouse-Friderichsen syndrome** (Chapter 24). An unusual form of DIC occurs in association with giant hemangiomas **(Kasabach-Merritt syndrome)** in which thrombi form within the neoplasm because of stasis and recurrent trauma to fragile blood vessels.

Clinical Features

The onset of DIC may be fulminant, as in sepsis or amniotic fluid embolism, or insidious and chronic, as in cases of carcinomatosis or retention of a dead fetus. It is almost impossible to detail all of the potential clinical presentations, but a few common patterns are worthy of description. These include microangiopathic hemolytic anemia; dyspnea, cyanosis, and respiratory failure; convulsions and coma; oliguria and acute renal failure; and sudden or progressive circulatory failure and shock. In general, acute DIC, associated (for example) with obstetric complications or major trauma, is dominated by a bleeding diathesis, whereas chronic DIC, such as occurs in cancer patients, tends to present with thrombotic complications. Diagnosis is based on clinical observation and laboratory studies, including measurement of fibrinogen levels, platelets, PT and PTT, and fibrin degradation products, particularly D-dimers.

The prognosis is highly variable and largely depends on the underlying disorder. The only definitive treatment is to remove or treat the inciting cause. Management requires meticulous maneuvering between the Scylla of thrombosis and the Charybdis of bleeding diathesis. Administration of anticoagulants or procoagulants has been advocated in specific settings, but not without controversy.

KEY CONCEPTS

Immune Thrombocytopenic Purpura

- Caused by autoantibodies against platelet antigens
- May be triggered by drugs, infections, or lymphomas, or may be idiopathic

Thrombotic Thrombocytopenic Purpura and Hemolytic Uremic Syndrome

- Both manifest with thrombocytopenia, microangiopathic hemolytic anemia, and renal failure; fever and CNS involvement are more typical of TTP.
- TTP: Caused by acquired or inherited deficiencies of ADAMTS 13, a plasma metalloprotease that cleaves very-high-molecular-weight multimers of vWF. Deficiency of ADAMTS 13 results in abnormally large vWF multimers that activate platelets.
- HUS: caused by deficiencies of complement regulatory proteins or agents that damage endothelial cells, such as a Shiga-like toxin elaborated by *E. coli* strain O157:H7. These abnormalities initiate platelet activation, platelet aggregation, and microvascular thrombosis.

Von Willebrand Disease

- Autosomal dominant disorder caused by mutations in vWF, a large protein that promotes the adhesion of platelets to subendothelial collagen and stabilizes factor VIII
- Typically causes a mild to moderate bleeding disorder resembling that associated with thrombocytopenia

Hemophilia

- Hemophilia A: X-linked disorder caused by mutations in factor VIII. Affected males typically present with severe bleeding into soft tissues and joints and have a prolonged PTT.
- Hemophilia B: X-linked disorder caused by mutations in coagulation factor IX. It is clinically identical to hemophilia A.

Disseminated Intravascular Coagulation

- Syndrome in which systemic activation of the coagulation leads to consumption of coagulation factors and platelets
- Can produce bleeding, vascular occlusion, and tissue hypoxemia in various combinations
- Common triggers: sepsis, major trauma, certain cancers, obstetric complications

Complications of Transfusion

Blood products are often rightly called the gift of life, permitting people to survive traumatic injuries and procedures such as hematopoietic stem cell transplantation and complex surgical procedures that would otherwise prove fatal. Over 5 million red cell transfusions are given in US hospitals each year. Thanks to improved screening of donors, blood products (packed red blood cells, platelets, and fresh-frozen plasma) are safer than ever before.

Nevertheless, complications still occur. Most are minor and transient. The most common is referred to as a *febrile nonhemolytic reaction,* which takes the form of fever and chills, sometimes with mild dyspnea, within 6 hours of a transfusion of red cells or platelets. These reactions are thought to be caused by inflammatory mediators derived from donor leukocytes. The frequency of these reactions increases with the storage age of the product and is decreased by measures that limit donor leukocyte contamination. Symptoms respond to antipyretics and are short-lived.

Other transfusion reactions are uncommon or rare, but can have severe and sometimes fatal consequences, and therefore merit discussion.

Allergic Reactions

Severe, potentially fatal allergic reactions may occur when blood products containing certain antigens are given to previously sensitized recipients. These are most likely to occur in patients with IgA deficiency, which has a frequency of 1:300 to 1:500 people. In this instance, the reaction is triggered by IgG antibodies that recognize IgA in the infused blood product. Fortunately, most patients with IgA deficiency do not develop such antibodies and these severe reactions are rare, occurring in 1 in 20,000 to 1 in 50,000 transfusions. Urticarial allergic reactions may be triggered by the presence an allergen in the donated blood product that is recognized by IgE antibodies in the recipient. These are considerably more common, occurring in 1% to 3% of transfusions, but are generally mild. In most instances, symptoms respond to antihistamines and do not require discontinuation of the transfusion.

Hemolytic Reactions

Acute hemolytic reactions are usually caused by preformed IgM antibodies against donor red cells that fix complement. They most commonly stem from an error in patient identification or tube labeling that allows a patient to receive an ABO-incompatible unit of blood. Preexisting high-affinity "natural" IgM antibodies, usually against polysaccharide blood group antigens A or B, bind to red cells and rapidly induce complement-mediated lysis, intravascular hemolysis, and hemoglobinuria. Fever, shaking chills, and flank pain appear rapidly. The direct Coombs test is typically positive, unless all of the donor red cells have lysed. The signs and symptoms are due to complement activation rather than intravascular hemolysis per se, as osmotic lysis of red cells (e.g., by mistakenly infusing red cells and 5% dextrose in water simultaneously) produces hemoglobinuria without any of the other symptoms of a hemolytic reaction. In severe cases, the process may rapidly progress to DIC, shock, acute renal failure, and occasionally death.

Delayed hemolytic reactions are caused by antibodies that recognize red cell antigens that the recipient was sensitized to previously, for example, through a prior blood transfusion. These are typically caused by IgG antibodies to foreign protein antigens and are associated with a positive direct Coombs test and laboratory features of hemolysis (e.g., low haptoglobin and elevated lactate dehydrogenase). Antibodies to antigens such as Rh, Kell, and Kidd often induce sufficient complement activation to cause severe and

potentially fatal reactions identical to those resulting from ABO mismatches. Other antibodies that do not fix complement typically result in red cell opsonization, extravascular hemolysis, and spherocytosis, and are associated with relatively minor signs and symptoms.

Transfusion-Related Acute Lung Injury

Transfusion-related acute lung injury (TRALI) is a severe, frequently fatal complication in which factors in a transfused blood product trigger the activation of neutrophils in the lung microvasculature. The incidence of TRALI is low, probably less than 1 in 10,000 transfusions, but it may occur more frequently in patients with preexisting lung disease. Though its pathogenesis is incompletely understood, current models favor a "two-hit" hypothesis. The first is neutrophil sequestration and priming in the microvasculature of the lung. It is postulated that endothelial cells are involved both in sequestration and priming; the former by up-regulation of adhesion molecules and the latter by release of cytokines. The second hit involves activation of primed neutrophils by a factor present in the transfused blood product.

A variety of factors have been implicated as "second hits," but the leading candidates are antibodies in the transfused blood product that recognize antigens expressed on neutrophils. By far the most common antibodies associated with TRALI are those that bind major histocompatibility complex (MHC) antigens. These antibodies are often found in multiparous women, who generate such antibodies in response to foreign MHC antigens expressed by the fetus. In other cases, donor antibodies to neutrophil-specific antigens have been implicated as triggers.

Although TRALI has been associated with virtually all plasma-containing blood products, it is more likely to occur following transfusion of products containing high levels of donor antibody, such as fresh-frozen plasma and platelets. The presentation is dramatic, consisting of sudden-onset respiratory failure, during or soon after a transfusion. Diffuse bilateral pulmonary infiltrates that do not respond to diuretics are seen on chest imaging. Other associated findings include fever, hypotension, and hypoxemia. The treatment is largely supportive, and the outcome is guarded; mortality is 5% in uncomplicated cases and up to 67% in the severely ill. TRALI is important to recognize, because donor products that induce the complication in one patient are much more likely to do so in a second. Indeed, measures taken to exclude multiparous women from plasma donation have sharply reduced the incidence of TRALI.

Infectious Complications

Virtually any infectious agent can be transmitted through blood products, but bacterial and viral infections are the dominant culprits. Most bacterial infections are caused by skin flora, indicating that the contamination occurred at the time that the product was taken from the donor. Significant bacterial contamination (sufficient to produce symptoms) is much more common in platelet preparations than in red cell preparations, due in large part to the fact that platelets (unlike red cells) must be stored at room temperature, a condition favorable for bacterial growth. Rates of bacterial infection following platelet transfusion are as high as 1 in 5000, with infections secondary to red cell transfusions being several orders of magnitude less frequent. Many of the symptoms (fever, chills, hypotension) resemble those of hemolytic and nonhemolytic transfusion reactions, and it may be necessary to start broad-spectrum antibiotics prospectively in symptomatic patients while awaiting laboratory results.

Advances in donor selection, donor screening, and infectious disease testing have dramatically decreased the incidence of viral transmission by blood products. However, on rare occasions when the donor is acutely infected but the virus is not yet detectable with current nucleic acid testing technology, there can be transfusion-related transmission of viruses such as HIV, hepatitis C, and hepatitis B. Rates of transmission of HIV, hepatitis C, and hepatitis B are estimated to be 1 in 1.5 million, 1 in 1.2 million, and 1 in 1 million, respectively. There also remains a low risk of "exotic" infectious agents such as West Nile virus, trypanosomiasis, and babesiosis.

SUGGESTED READINGS

Red Cell Disorders

Higgs DR, Engel JD, Stamatoyannopoulos G: Thalassemia, *Lancet* 379:373, 2012. *[A review of the molecular pathogenesis of thalassemia syndromes].*

Hill A, DeZern AE, Kinoshita T et al: Paroxysmal nocturnal hemoglobinuria, *Nat Rev Dis Primers* 3:17028, 2017. *[A discussion of pathophysiology of paroxysmal nocturnal hemoglobinuria and the therapeutic impact of antibodies that inhibit the C5b–C9 membrane attack complex].*

Liu N, Hargreaves VV, Zhu Q et al: Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch, *Cell* 173:430, 2018. *[A landmark paper defining the molecular basis of hemoglobin switching and how to reverse it, providing a therapeutic strategy in certain thalassemias and hemoglobinopathies].*

Muckenthaler MU, Rivella S, Hentze MW: A red carpet for iron metabolism, *Cell* 168:344, 2017. *[A review focused on normal iron metabolism and disorders that perturb it].*

Narla J, Mohandas N: Red cell membrane disorders, *Int J Lab Med* 39(47):2017. *[An excellent overview of common hemolytic anemias caused by inherited red cell membrane defects].*

Sundd P, Gladwin MT, Novelli EM: Pathophysiology of sickle cell disease, *Annu Rev Pathol* 14:263, 2019. *[An updated discussion of how sickle cell disease leads to tissue damage].*

Young NS: Aplastic anemia, *N Engl J Med* 379:2018, 1643. *[A discussion of the pathophysiology, diagnosis, and treatment of aplastic anemia].*

Bleeding Disorders

Gando S, Levi M, Toh CH: Disseminated intravascular coagulation, *Nat Rev Dis Primers* 2:16037, 2016. *[An overview of the pathophysiology of disseminated intravascular coagulation].*

Leebeek FWG, Eikenboom JCJ: von Willebrand's disease, *N Engl J Med* 375:2067, 2016. *[An update on the molecular pathogenesis, pathophysiology, clinical features, and treatment of von Willebrand disease].*

Noris M, Remuzzi G: Atypical hemolytic uremic syndrome, *N Engl J Med* 361:1676, 2009. *[An article focused on the role of excessive activation of the alternative complement pathway in some forms of hemolytic uremic syndrome].*

Saha M, McDaniel JK, Zheng XL: Thrombotic thrombocytopenia purpura: Pathogenesis, diagnosis, and potential novel therapeutics, *J Thromb Haemost* 15:1889, 2017. *[A current review of basic and clinical aspects of thrombotic thrombocytopenia purpura].*

Salter BS, Weiner MM, Trinh MA et al: Heparin-induced thrombocytopenia: A comprehensive clinical review, *J Am Coll Cardiol* 67:2519, 2016. *[A thorough discussion of pathogenesis, clinical features, diagnostic criteria, and therapeutic approaches in heparin-induced thrombocytopenia].*

Zuffery A, Kapur R, Semple JW: Pathogenesis and therapeutic mechanisms in immune thrombocytopenia, *J Clin Med* 6:16, 2017. *[A discussion of the pathophysiology and treatment of immune thrombocytopenia].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 15

The Lung

Aliya N. Husain

CHAPTER CONTENTS

The lungs are ingeniously constructed to carry out their cardinal function, the exchange of gases between inspired air and blood. Developmentally, the respiratory system is an outgrowth from the ventral wall of the foregut. The midline trachea develops two lateral outpocketings, the lung buds, which eventually divide into branches called lobar bronchi, three on the right and two on the left, thus giving rise to three lobes on the right and two on the left. The lobar bronchi allow air to pass in and out of the lung. They have firm cartilaginous walls that provide mechanical support and are lined by columnar ciliated epithelium with abundant subepithelial glands that produce mucus, which impedes the entry of microbes. The right mainstem bronchus is more vertical and directly in line with the trachea. Consequently, aspirated foreign materials such as vomitus, blood, and foreign bodies tend to enter the right lung more often than the left.

The right and left lobar bronchi branch to give rise to progressively smaller airways that are accompanied by a dual arterial supply from the pulmonary and bronchial arteries. Distally, the bronchi give way to *bronchioles*, which are distinguished by the lack of cartilage and the presence of submucosal glands within their walls. Further branching

of bronchioles leads to the *terminal bronchioles,* which are less than 2 mm in diameter. Beyond the terminal bronchiole is the *acinus,* a roughly spherical structure with a diameter of about 7 mm. An acinus is composed of a *respiratory bronchiole* (which gives off several alveoli from its sides), *alveolar ducts,* and *alveolar sacs,* the blind ends of the respiratory passages, whose walls are formed entirely of alveoli, the site of gas exchange. A cluster of three to five terminal bronchioles, each with its appended acinus, is referred to as the pulmonary *lobule*.

Except for the true vocal cords, which are covered by stratified squamous epithelium, the entire respiratory tree, including the larynx, trachea, and bronchioles, is normally lined mainly by tall, columnar, pseudostratified, ciliated epithelial cells and a smaller population of non-ciliated cells called club cells that secrete a number of protective substances into the airway. There also are scattered cells called ionocytes that express high levels of the cystic fibrosis transmembrane conductance regulator (CFTR) and appear to modulate the ion content and viscosity of bronchial secretions. In addition, the bronchial mucosa contains neuroendocrine cells with neurosecretory-type granules that can release a variety of factors including serotonin, calcitonin, and gastrin-releasing peptide (bombesin). Numerous mucus-secreting goblet cells and submucosal glands also are dispersed throughout the walls of the trachea and bronchi (but not the bronchioles).

The microscopic structure of the alveolar walls (or alveolar septa) consists of the following (Fig. 15.1):

- An intertwining network of *anastomosing capillaries* lined with endothelial cells.
- *Basement membrane and surrounding interstitium,* which separates the endothelial cells from the alveolar lining epithelial cells. In thin portions of the alveolar septum the basement membranes of epithelium and endothelium are fused, whereas in thicker portions they are separated by an interstitial space *(pulmonary interstitium)* containing fine elastic fibers, small bundles of collagen, a few fibroblast-like interstitial cells, smooth muscle cells, mast cells, and rare lymphocytes and monocytes.

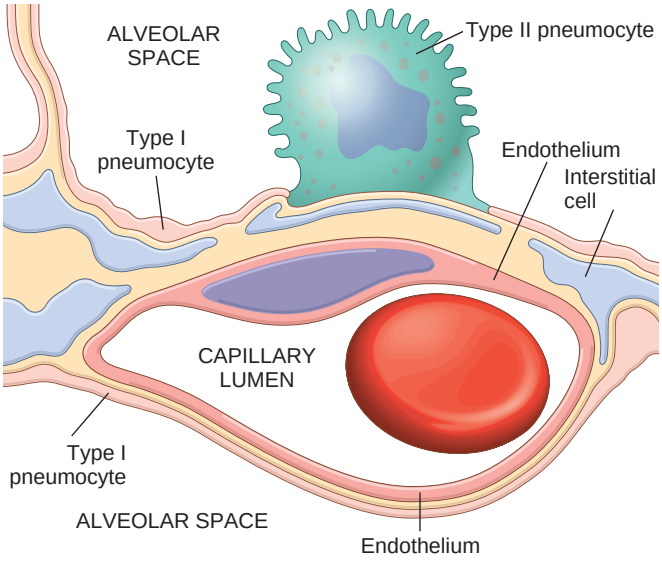


Figure 15.1 Microscopic structure of the alveolar wall. Note that the basement membrane *(yellow)* is thin on one side and widened where it is continuous with the interstitial space. Portions of interstitial cells are shown.

- *Alveolar epithelium,* a continuous layer of two cell types: flat, plate-like *type I pneumocytes,* covering 95% of the alveolar surface, and rounded *type II pneumocytes*. Type II cells synthesize *surfactant* (which forms a very thin layer over the alveolar cell membranes) and are involved in the repair of alveolar damage through their ability to proliferate and give rise to type I cells.

The alveolar walls are perforated by numerous *pores of Kohn,* which permit the passage of bacteria and exudate between adjacent alveoli (see Fig. 15.35B). Scattered resident *alveolar macrophages* also are present, either loosely attached to epithelial cells or lying free within the alveolar spaces.

CONGENITAL ANOMALIES

Developmental anomalies of the lung are rare. The most common of these include the following:

- *Pulmonary hypoplasia* is a defect in the development of both lungs (one may be more affected than the other) that results in decreased lung size. It is caused by abnormalities that compress the lung or impede lung expansion in utero, such as congenital diaphragmatic hernia and oligohydramnios. Severe hypoplasia is fatal in the early neonatal period.
- *Foregut cysts* arise from abnormal detachments of primitive foregut and are most often located in the hilum or middle mediastinum. Depending on the wall structure, these cysts are classified as bronchogenic (most common), esophageal, or enteric. A bronchogenic cyst is rarely connected to the tracheobronchial tree. It is lined by ciliated pseudostratified columnar epithelium and has a wall containing bronchial glands, cartilage, and smooth muscle. They may come to attention due to symptoms resulting from compression of nearby structures or superimposed infection or may be incidental findings.
- *Pulmonary sequestration* is a discrete area of lung tissue that (1) is not connected to the airways and (2) has an abnormal blood supply arising from the aorta or its branches. *Extralobar sequestration* is external to the lung and most commonly presents in infants as a mass lesion. It may be associated with other congenital anomalies. *Intralobar sequestration* occurs within the lung. It usually presents in older children, often due to recurrent localized infections or bronchiectasis.

Other, less common congenital abnormalities include tracheal and bronchial anomalies (atresia, stenosis, tracheoesophageal fistula), vascular anomalies, congenital pulmonary airway malformation, and congenital lobar overinflation (emphysema).

ATELECTASIS (COLLAPSE)

Atelectasis refers either to incomplete expansion of the lungs (neonatal atelectasis) or to the collapse of previously inflated lung, and results in areas of poorly aerated pulmonary parenchyma. The main types of acquired atelectasis, which is encountered principally in adults, are the following (Fig. 15.2).

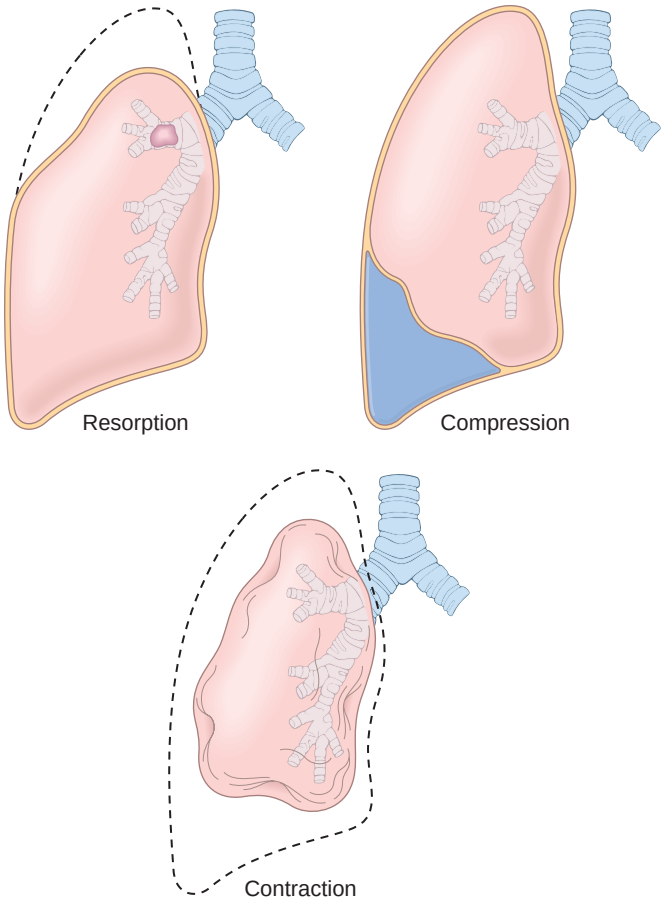


Figure 15.2 Various forms of acquired atelectasis. *Dashed lines* indicate normal lung volume.

- *Resorption atelectasis* stems from obstruction of an airway. Over time, air is resorbed from distal alveoli, which collapse. Since lung volume is diminished, the mediastinum shifts *toward* the atelectatic lung. Airway obstruction is most often caused by excessive secretions (e.g., mucus plugs) or exudates within smaller bronchi, as may occur in bronchial asthma, chronic bronchitis, bronchiectasis, and postoperative states. Aspiration of foreign bodies and intrabronchial tumors may also lead to airway obstruction and atelectasis.
- *Compression atelectasis* results whenever significant volumes of fluid (transudate, exudate, or blood), tumor, or air *(pneumothorax)* accumulate within the pleural cavity. With compression atelectasis, the mediastinum shifts *away* from the affected lung.
- *Contraction atelectasis* occurs when focal or generalized pulmonary or pleural fibrosis prevents full lung expansion.

Significant atelectasis reduces oxygenation and predisposes to infection. Except in cases caused by fibrosis, atelectasis is a reversible disorder.

PULMONARY EDEMA

Pulmonary edema (excessive interstitial fluid in the alveoli) can result from hemodynamic disturbances (cardiogenic pulmonary edema) or increased capillary permeability due to microvascular injury (non-cardiogenic pulmonary edema) (Table 15.1). A general consideration of edema is given in Chapter 4, and pulmonary congestion and edema are described briefly in the context of congestive heart failure (Chapter 12). Whatever the clinical setting, pulmonary congestion and edema produce heavy, wet lungs. Therapy and outcome depend on the underlying etiology.

Hemodynamic Pulmonary Edema

Hemodynamic pulmonary edema is caused by *increased hydrostatic pressure* and occurs most commonly in left-sided congestive heart failure. Edema accumulates initially in the basal regions of the lower lobes because hydrostatic pressure is greatest in these sites (dependent edema). Histologically, the alveolar capillaries are engorged, and there is an intra-alveolar transudate that appears as finely granular pale pink material. Alveolar microhemorrhages and *hemosiderin-laden macrophages ("heart failure" cells)* may be present. In long-standing pulmonary congestion (e.g., as seen in mitral stenosis), hemosiderin-laden macrophages are abundant, and fibrosis and thickening of the alveolar walls cause the soggy lungs to become firm and brown *(brown induration)*. These changes not only impair respiratory function but also predispose to infection.

Table 15.1 Classification and Causes of Pulmonary Edema

Hemodynamic Edema
Increased hydrostatic pressure (increased pulmonary venous pressure)
Left-sided heart failure (common)
Volume overload
Pulmonary vein obstruction
Decreased oncotic pressure (less common)
Hypoalbuminemia
Nephrotic syndrome
Liver disease
Protein-losing enteropathies
Lymphatic obstruction (rare)
Edema Due to Alveolar Wall Injury (Microvascular or Epithelial Injury)
Direct injury
Infections: bacterial pneumonia
Inhaled gases: high concentrations of oxygen, smoke
Liquid aspiration: gastric contents, near-drowning
Radiation
Lung trauma
Indirect injury
Systemic inflammatory response syndrome (e.g., associated with sepsis, burns, pancreatitis, extensive trauma)
Blood transfusion–related
Drugs and chemicals: chemotherapeutic agents (bleomycin), other medications (methadone, amphotericin B), heroin, cocaine, kerosene, paraquat
Edema of Undetermined Origin
High altitude
Neurogenic (central nervous system trauma)

Edema Caused by Microvascular (Alveolar) Injury

Non-cardiogenic pulmonary edema is caused by injury of the alveolar septa. Primary injury to the vascular endothelium or damage to alveolar epithelial cells (with secondary microvascular injury) produces an inflammatory exudate that leaks into the interstitial space and, in more severe cases, into the alveoli. Injury-related alveolar edema is an important feature of a serious and often fatal condition, *acute respiratory distress syndrome* (discussed below).

ACUTE LUNG INJURY AND ACUTE RESPIRATORY DISTRESS SYNDROME (DIFFUSE ALVEOLAR DAMAGE)

Acute lung injury (ALI) is characterized by the abrupt onset of hypoxemia and bilateral pulmonary edema in the absence of cardiac failure (non-cardiogenic pulmonary edema). Acute respiratory distress syndrome (ARDS) is a manifestation of severe ALI. Both ARDS and ALI are associated with inflammation-associated increases in pulmonary vascular permeability, edema, and epithelial cell death. The histologic manifestation of these diseases is *diffuse alveolar damage*.

ALI is a well-recognized complication of diverse conditions including both pulmonary and systemic disorders (Table 15.2). In many cases, several predisposing conditions are present (e.g., shock, oxygen therapy, and sepsis). In other uncommon instances, ALI appears acutely in the absence of known triggers and follows a rapidly progressive clinical course, a condition known as *acute interstitial pneumonia*.

Table 15.2 Conditions Associated With Development of Acute Respiratory Distress Syndrome

Infection
Sepsis[a]
Diffuse pulmonary infections[a]
Viral, *Mycoplasma*, and *Pneumocystis* pneumonia; miliary tuberculosis
Gastric aspiration[a]
Physical/Injury
Mechanical trauma including head injuries[a]
Pulmonary contusions
Near-drowning
Fractures with fat embolism
Burns
Ionizing radiation
Inhaled Irritants
Oxygen toxicity
Smoke
Irritant gases and chemicals
Chemical Injury
Heroin or methadone overdose
Acetylsalicylic acid
Barbiturate overdose
Paraquat
Hematologic Conditions
Transfusion-associated lung injury (TRALI)
Disseminated intravascular coagulation
Pancreatitis
Uremia
Cardiopulmonary Bypass
Hypersensitivity Reactions
Organic solvents
Drugs

[a]More than 50% of cases of acute respiratory distress syndrome are associated with these four conditions.

Pathogenesis

ALI/ARDS is initiated by injury of pneumocytes and pulmonary endothelium, setting in motion a vicious cycle of increasing inflammation and pulmonary damage (Fig. 15.3).

- *Endothelial activation* is an important early event. In some instances, endothelial activation is secondary to pneumocyte injury, which is sensed by resident alveolar macrophages. In response, these immune sentinels secrete mediators such as tumor necrosis factor (TNF) that act on the neighboring endothelium. Alternatively, circulating inflammatory mediators may activate pulmonary endothelium directly in the setting of severe tissue injury or sepsis. Some of these mediators injure endothelial cells, while others (notably cytokines) induce endothelial cells to express increased levels of adhesion molecules, procoagulant proteins, and chemokines.
- *Adhesion and extravasation of neutrophils.* Neutrophils adhere to the activated endothelium and migrate into the interstitium and the alveoli, where they degranulate and release inflammatory mediators including proteases, reactive oxygen species, and cytokines. Experimental evidence suggest that neutrophil extracellular traps (NETs) are released and also contribute directly to lung damage. These injuries and associated proinflammatory factors set in motion a vicious cycle of inflammation and endothelial damage that lies at the heart of ALI/ARDS.
- *Accumulation of intra-alveolar fluid and formation of hyaline membranes.* Endothelial activation and injury make pulmonary capillaries leaky, allowing interstitial and intra-alveolar edema fluid to form. Damage and necrosis of type II alveolar pneumocytes lead to surfactant abnormalities, further compromising alveolar gas exchange. Ultimately the inspissated protein-rich edema fluid and debris from dead alveolar epithelial cells organize into hyaline membranes, a characteristic feature of ALI/ARDS.
- *Resolution of injury* is impeded in ALI/ARDS due to epithelial necrosis and inflammatory damage that impairs the ability of remaining cells to assist with edema resorption. Eventually, however, if the inflammatory stimulus lessens, macrophages remove intra-alveolar debris and release fibrogenic cytokines such as transforming growth factor β (TGF-β) and platelet-derived growth factor. These factors stimulate fibroblast growth and collagen deposition, leading to fibrosis of alveolar walls. Residual type II pneumocytes proliferate to replace type I pneumocytes, reconstituting the alveolar lining. Endothelial restoration occurs through proliferation of uninjured capillary endothelium.

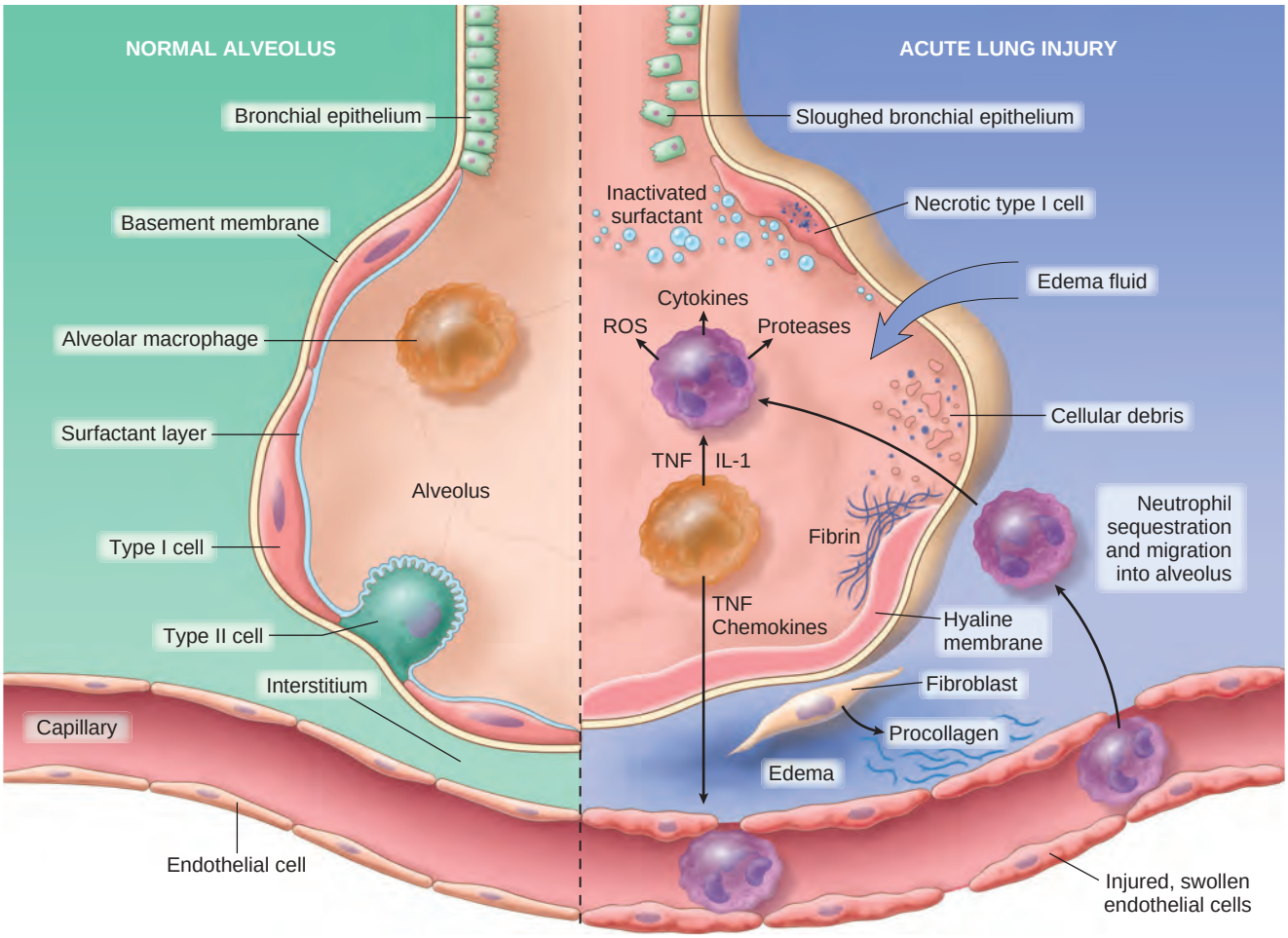


Figure 15.3 The normal alveolus *(left side)* compared with the injured alveolus in the early phase of acute lung injury and acute respiratory distress syndrome. *IL-1*, Interleukin-1; *ROS*, reactive oxygen species; *TNF*, tumor necrosis factor. (Modified with permission from Matthay MA, Ware LB, Zimmerman GA: The acute respiratory distress syndrome, *J Clin Invest* 122:2731, 2012.)

The lesions in ARDS are not evenly distributed, and as a result there are typically areas that are stiff and poorly aerated and regions that have nearly normal levels of compliance and ventilation. Because poorly aerated regions continue to be perfused, there is a mismatch of ventilation and perfusion, a phenomenon that exacerbates the hypoxemia and cyanosis.

Epidemiologic studies have shown that ALI/ARDS is more common and has a worse prognosis in chronic alcoholics and in smokers. Genome-wide association studies have identified a number of genetic variants that increase the risk of ARDS, some of which map to genes linked to inflammation and coagulation.

MORPHOLOGY

In the acute exudative stage, the lungs are heavy, firm, red, and boggy. They exhibit congestion, interstitial and intra-alveolar edema, inflammation, fibrin deposition, and **diffuse alveolar damage.** The alveolar walls become lined with waxy **hyaline membranes** (Fig. 15.4) that are morphologically similar to those seen in hyaline membrane disease of neonates (Chapter 10). Alveolar hyaline membranes consist of fibrin-rich edema fluid mixed with the remnants of necrotic epithelial cells. In the proliferative or organizing stage, type II pneumocytes proliferate, and granulation tissue forms in the alveolar walls and spaces. In most cases the granulation tissue resolves, leaving minimal functional impairment. Sometimes, however, fibrotic thickening (scarring) of the alveolar septa ensues (late fibrotic stage).

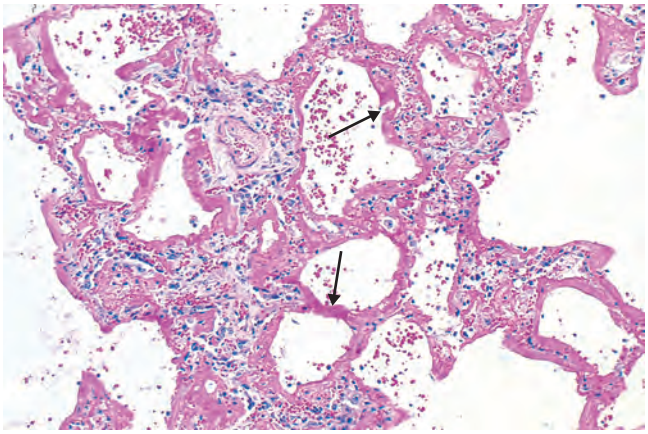


Figure 15.4 Diffuse alveolar damage (acute respiratory distress syndrome). Some of the alveoli are collapsed, while others are distended. Many are lined by hyaline membranes *(arrows)*.

Clinical Features

Profound dyspnea and tachypnea herald ALI/ARDS, followed by increasing respiratory failure, hypoxemia, cyanosis, and the appearance of diffuse bilateral infiltrates on radiographic examination. Hypoxemia may be refractory to oxygen therapy due to ventilation-perfusion mismatch, and respiratory acidosis can develop. Early in the course, the lungs become stiff due to loss of functional surfactant, leading to the need for intubation and high ventilatory pressures to maintain adequate gas exchange.

There are no proven specific treatments for ARDS, which is common in acutely ill patients and continues to take a high toll, even in patients receiving state-of-the-art supportive care. In a 2016 study of intensive care units in 50 countries, the incidence of ARDS was 10.4%, and mortality rates were 35% for mild, 40% for moderate, and 46% for severe ARDS. The majority of deaths are attributable to sepsis, multiorgan failure, or severe lung injury. Most survivors recover pulmonary function, but in a minority of patients, the lung damage results in interstitial fibrosis and chronic pulmonary disease.

KEY CONCEPTS

ACUTE RESPIRATORY DISTRESS SYNDROME

- ARDS is a clinical syndrome of progressive respiratory insufficiency caused by diffuse alveolar damage in the setting of sepsis, severe trauma, or diffuse pulmonary infection.
- Damage to endothelial and alveolar epithelial cells and secondary inflammation are the key initiating events and the basis of lung damage.
- The characteristic histologic finding is hyaline membranes lining alveolar walls, accompanied by edema, scattered neutrophils and macrophages, and epithelial necrosis.

OBSTRUCTIVE AND RESTRICTIVE LUNG DISEASES

Obstructive lung diseases are characterized by an increase in resistance to airflow due to diffuse airway disease, which may affect any level of the respiratory tract. These are contrasted with restrictive diseases, which are characterized by reduced expansion of lung parenchyma and decreased total lung capacity. The clinical distinction between these diseases is based primarily on pulmonary function tests. In individuals with diffuse obstructive disorders, pulmonary function tests show decreased maximal airflow rates during forced expiration, usually expressed as forced expiratory volume at 1 second (FEV_1) over forced ventilatory capacity (FVC). An FEV_1/FVC ratio of less than 0.7 generally indicates obstructive disease. Expiratory airflow obstruction may be caused by a variety of conditions (Table 15.3), each with characteristic pathologic changes and different mechanisms of airflow obstruction. As discussed later, however, the divisions between these entities are not "clean," and many patients have diseases with overlapping features. By contrast, restrictive diseases are associated with proportionate decreases in both total lung capacity and FEV_1, such that the FEV_1/FVC ratio remains normal. Restrictive defects occur in two broad kinds of conditions: (1) *chest wall disorders* (e.g., severe obesity, pleural diseases, kyphoscoliosis, and neuromuscular diseases such as poliomyelitis) and (2) *chronic interstitial and infiltrative diseases,* such as pneumoconioses and interstitial fibrosis.

OBSTRUCTIVE LUNG DISEASES

Common obstructive lung diseases include chronic obstructive pulmonary disease (COPD), asthma, and bronchiectasis (Table 15.3). COPD has two major clinicopathologic manifestations, emphysema and chronic bronchitis, which are often found together in the same patient, almost certainly because they share the same major etiologic factor—cigarette smoking. While asthma is distinguished from chronic bronchitis and emphysema by the presence of reversible bronchospasm, some patients with otherwise typical asthma also develop an irreversible component (Fig. 15.5). Conversely, some patients with otherwise typical COPD have a reversible component. Clinicians commonly label such patients as having COPD/asthma.

Chronic Obstructive Pulmonary Disease

COPD, a major public health problem, is defined by the World Health Organization (WHO) as "a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities caused by exposure to noxious particles or gases." It is currently the fourth

Table 15.3 Disorders Associated With Airflow Obstruction: The Spectrum of Chronic Obstructive Pulmonary Disease

Clinical Term	Anatomic Site	Major Pathologic Changes	Etiology	Signs/Symptoms
Chronic bronchitis	Bronchus	Mucous gland hyperplasia, hypersecretion	Tobacco smoke, air pollutants	Cough, sputum production
Bronchiectasis	Bronchus	Airway dilation and scarring	Persistent or severe infections	Cough, purulent sputum, fever
Asthma	Bronchus	Smooth muscle hyperplasia, excess mucus, inflammation	Immunologic or undefined causes	Episodic wheezing, cough, dyspnea
Emphysema	Acinus	Airspace enlargement; wall destruction	Tobacco smoke	Dyspnea
Small airways disease, bronchiolitis[a]	Bronchiole	Inflammatory scarring/obliteration	Tobacco smoke, air pollutants, miscellaneous	Cough, dyspnea

[a]Can be seen with any form of obstructive lung disease or as an isolated finding.

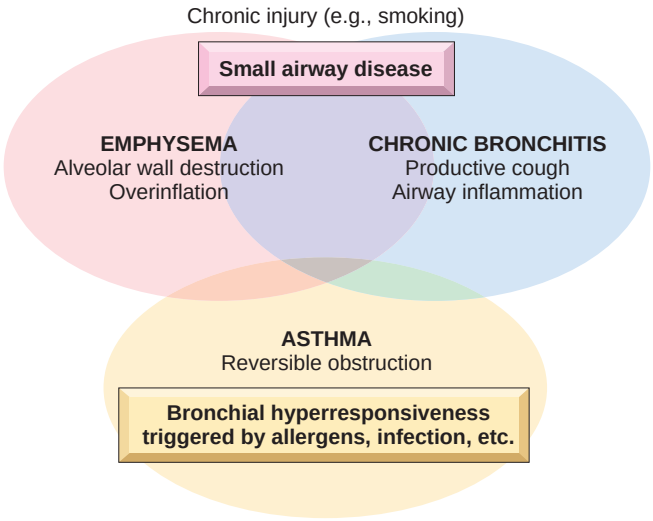


Figure 15.5 Schematic representation of overlap between chronic obstructive lung diseases.

leading cause of death in the world and is projected to rank third by 2020 due to increases in cigarette smoking in countries such as China. There is a strong association between heavy cigarette smoking and COPD. Overall, 35% to 50% of heavy smokers develop COPD; conversely about 80% of COPD is attributable to smoking. Women and African Americans who smoke heavily are more susceptible than other groups. Additional risk factors include poor lung development early in life, exposure to environmental and occupational pollutants, airway hyperresponsiveness, and certain genetic polymorphisms.

Recognizing that emphysema and chronic bronchitis often occur together in patients with COPD, it is still useful to discuss these patterns of lung injury and associated functional abnormalities individually to highlight the pathophysiologic basis of different causes of airflow obstruction. We will finish our discussion by returning to the clinical features of COPD.

Emphysema

Emphysema is defined by irreversible enlargement of the airspaces distal to the terminal bronchiole, accompanied by destruction of their walls. Subtle but functionally important small airway fibrosis (distinct from chronic bronchitis) is also present and is a significant contributor to airflow obstruction. Emphysema is classified according to its anatomic distribution within the lobule. Recall that the lobule is a cluster of acini, the terminal respiratory units. Based on the segments of the respiratory units that are involved, emphysema is subdivided into four major types: (1) *centriacinar,* (2) *panacinar,* (3) *paraseptal,* and (4) *irregular.* Of these, only the first two cause clinically significant airflow obstruction (Fig. 15.6).

- *Centriacinar (centrilobular) emphysema.* Centriacinar emphysema is the most common form, constituting more than 95% of clinically significant cases. It occurs predominantly in heavy smokers with COPD. In this type of emphysema the central or proximal parts of the acini, formed by respiratory bronchioles, are affected, whereas distal alveoli are spared (Figs. 15.6B and 15.7A). Thus, both emphysematous and normal airspaces exist within the same acinus and lobule. The lesions are more common and usually more pronounced in the upper lobes, particularly in the apical segments. In severe centriacinar emphysema, the distal acinus may also be involved, making differentiation from panacinar emphysema difficult.
- *Panacinar (panlobular) emphysema.* Panacinar emphysema is associated with *α1-antitrypsin deficiency* (Chapter 18) and is exacerbated by smoking. In this type the acini are uniformly enlarged from the level of the respiratory bronchiole to the terminal blind alveoli (Figs. 15.6C and 15.7B). In contrast to centriacinar emphysema, panacinar emphysema tends to occur more commonly in the lower zones and in the anterior margins of the lung, and it is usually most severe at the bases.
- *Distal acinar (paraseptal) emphysema.* Distal acinar emphysema probably underlies many cases of spontaneous pneumothorax in young adults. In this type the proximal portion of the acinus is normal, and the distal part is predominantly involved. The emphysema is more striking adjacent to the pleura, along the lobular connective tissue septa, and at the margins of the lobules. It occurs adjacent to areas of fibrosis, scarring, or atelectasis and is usually more severe in the upper half of the lungs. The characteristic finding is multiple enlarged airspaces, ranging from less than 0.5 cm to more than 2.0 cm in diameter, which sometimes form cyst-like structures.

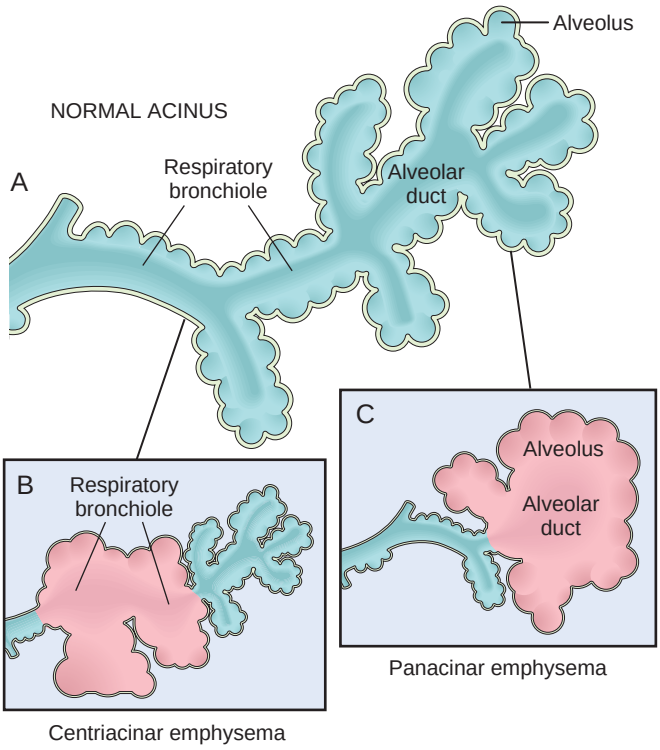


Figure 15.6 Clinically significant patterns of emphysema. (A) Structure of the normal acinus. (B) Centriacinar emphysema with dilation that initially affects the respiratory bronchioles. (C) Panacinar emphysema with initial distention of the alveolus and alveolar duct.

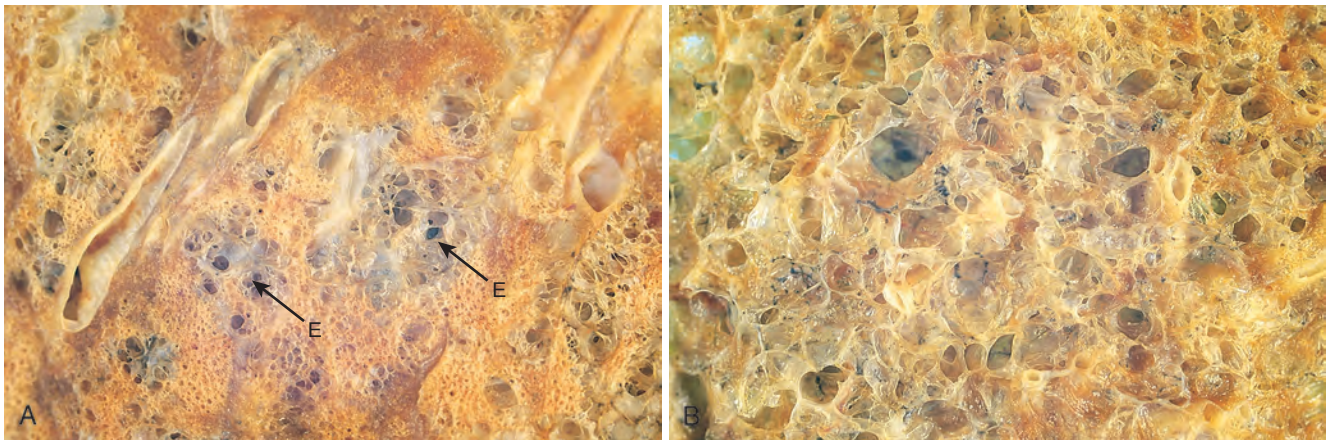


Figure 15.7 (A) Centriacinar emphysema. Central areas show marked emphysematous damage *(E)* surrounded by relatively spared alveolar spaces. (B) Panacinar emphysema involving the entire pulmonary lobule.

- *Airspace enlargement with fibrosis (irregular emphysema).* Irregular emphysema, so named because the acinus is irregularly involved, is almost invariably associated with scarring. In most instances it occurs in small foci and is clinically insignificant.

Pathogenesis

Clinically significant emphysema is largely confined to smokers and to patients with α_1-antitrypsin deficiency, highlighting the importance of these two etiologic factors. Mechanisms relating to these factors that contribute to the development of emphysema include the following (Fig. 15.8):

- *Toxic injury and inflammation.* Inhaled cigarette smoke and other noxious particles damage respiratory epithelium and cause inflammation, which results in variable degrees of parenchymal destruction. A wide variety of inflammatory mediators (including leukotriene B_4, interleukin [IL]-8, TNF, and others) are increased in the affected parts of the lung. These mediators are released by resident epithelial cells and macrophages and variously attract inflammatory cells from the circulation (chemotactic factors), amplify the inflammatory process (proinflammatory cytokines), and induce structural changes (growth factors). Chronic inflammation also leads to the accumulation of T and B cells in affected parts of the lung, but the role of adaptive immunity in emphysema is currently uncertain.
- *Protease-antiprotease imbalance.* Several proteases are released from the inflammatory cells and epithelial cells that break down connective tissue components. In patients who develop emphysema, there is a relative deficiency of protective antiproteases, which in some instances has a genetic basis (discussed later).
- *Oxidative stress.* Substances in tobacco smoke, alveolar damage, and inflammatory cells all produce oxidants, which may beget tissue damage, endothelial dysfunction, and inflammation. The role of oxidants is supported by studies of mice in which the *NRF2* gene is inactivated. NRF2 is a transcription factor that serves as a sensor for oxidants in many cell types including alveolar epithelial cells. Intracellular oxidants activate *NRF2,* which upregulates the expression of genes that protect cells from oxidant damage. Mice without *NRF2* are significantly more sensitive to tobacco smoke than normal mice, and genetic variants in *NRF2, NRF2* regulators, and *NRF2* target genes are all associated with smoking-related lung disease in humans.
- *Infection.* Although infection is not thought to play an initiating role in the tissue destruction, bacterial and/or viral infections may acutely exacerbate existing disease.

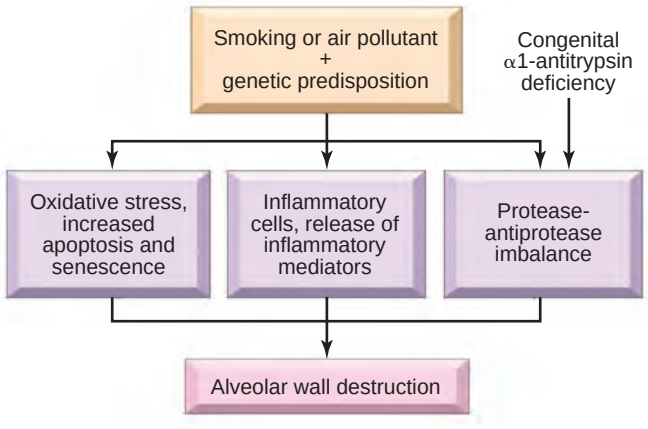


Figure 15.8 Pathogenesis of emphysema. See text for details.

The idea that proteases are important is based in part on the observation that patients with a genetic deficiency of the antiprotease α_1-antitrypsin have a markedly enhanced tendency to develop emphysema that is compounded by smoking. About 1% of all patients with emphysema have this defect. α_1-Antitrypsin, normally present in serum, tissue fluids, and macrophages, is a major inhibitor of proteases (particularly elastase) secreted by neutrophils during inflammation. It is encoded by the proteinase inhibitor *(Pi)* locus on chromosome 14. The *Pi* locus is polymorphic, and approximately 0.012% of the U.S. population is homozygous for the *Z* allele, a genotype associated with very low serum levels of α_1-antitrypsin. More than 80% of *ZZ* individuals develop symptomatic panacinar emphysema, which occurs at an earlier age and is of greater severity if the individual smokes. It is postulated that any injury (e.g., that induced by smoking) that increases the activation and influx of

neutrophils into the lung leads to local release of proteases, which in the absence of α_1-antitrypsin activity result in excessive digestion of elastic tissue and, with time, emphysema.

Several other genetic variants have also been linked to risk of emphysema. Among these are variants of the nicotinic acetylcholine receptor that are hypothesized to influence the addictiveness of tobacco smoke and thus the behavior of smokers. Not surprisingly, the same variants are also linked to lung cancer risk, emphasizing the importance of smoking in both of these diseases.

A number of factors contribute to airway obstruction in emphysema. Small airways are normally held open by the elastic recoil of the lung parenchyma. The loss of elastic tissue in the walls of alveoli that surround respiratory bronchioles reduces radial traction, leading to collapse of respiratory bronchioles during expiration and functional airflow obstruction in the absence of mechanical obstruction. In addition, even young smokers often have changes related to small airway inflammation that also contribute to airway narrowing and obstruction (described later).

MORPHOLOGY

Advanced emphysema produces voluminous lungs, often overlapping the heart anteriorly. Generally, in patients with smoking-related disease, the upper two-thirds of the lungs are more severely affected. Large alveoli can easily be seen on the cut surface of fixed lungs (see Fig. 15.7). Apical blebs or bullae characteristic of irregular emphysema may appear in patients with advanced disease.

Microscopically, abnormally large alveoli are separated by thin septa with focal centriacinar fibrosis. There is loss of attachments between alveoli and the outer wall of small airways. The pores of Kohn are so large that septa appear to be floating or protrude blindly into alveolar spaces with a club-shaped end. As alveolar walls are destroyed, there is a decrease in the capillary bed area. With advanced disease, there are even larger abnormal airspaces and possibly blebs or bullae, which often deform and compress the respiratory bronchioles and vasculature of the lung. Inflammatory changes in small airways are often superimposed (described next under chronic bronchitis), as are vascular changes related to pulmonary hypertension stemming from local hypoxemia and loss of capillary beds.

Chronic Bronchitis

Chronic bronchitis is defined clinically as persistent cough with sputum production for at least 3 months in at least 2 consecutive years in the absence of any other identifiable cause. Longstanding chronic bronchitis is associated with progressive lung dysfunction, which may be so severe as to lead to hypoxemia, pulmonary hypertension, and cor pulmonale.

Pathogenesis

The primary or initiating factor in the genesis of chronic bronchitis is exposure to noxious or irritating inhaled substances such as tobacco smoke (90% of those affected are smokers) and dust from grain, cotton, and silica. Several factors contribute to its pathogenesis.

- *Mucus hypersecretion.* The earliest feature of chronic bronchitis is hypersecretion of mucus in the large airways, associated with enlargement of the submucosal glands in the trachea and bronchi. The basis for mucus hypersecretion is incompletely understood, but it appears to involve inflammatory mediators such as histamine and IL-13. With time, there is also a marked increase in goblet cells in small airways—small bronchi and bronchioles—leading to excessive mucus production that contributes to airway obstruction. It is thought that both the enlargement of submucosal glands and the increase in numbers of goblet cells are protective reactions against tobacco smoke or other pollutants (e.g., sulfur dioxide and nitrogen dioxide).
- *Acquired cystic fibrosis transmembrane conductance regulator (CFTR) dysfunction.* There is substantial evidence that smoking leads to acquired CFTR dysfunction, which in turn causes the secretion of abnormal, dehydrated mucus that exacerbates the severity of chronic bronchitis.
- *Inflammation.* Inhalants that induce chronic bronchitis cause cellular damage, eliciting both acute and chronic inflammatory responses involving neutrophils, lymphocytes, and macrophages. Long-standing inflammation and accompanying fibrosis involving small airways (small bronchi and bronchioles, less than 2 to 3 mm in diameter) can also lead to chronic airway obstruction.
- *Infection.* Infection does not initiate chronic bronchitis, but is probably significant in maintaining it and may be critical in producing acute exacerbations.

Cigarette smoke predisposes to chronic bronchitis in several ways. Not only does it damage airway-lining cells, leading to chronic inflammation, but it also interferes with the ciliary action of the respiratory epithelium, preventing the clearance of mucus and increasing the risk of infection.

MORPHOLOGY

Grossly, there is hyperemia, swelling, and edema of the mucous membranes, frequently accompanied by excessive mucinous or mucopurulent secretions. Sometimes, heavy casts of secretions and pus fill the bronchi and bronchioles. The characteristic microscopic features are chronic inflammation of the airways (predominantly lymphocytes and macrophages); thickening of the bronchiolar wall due to smooth muscle hypertrophy, deposition of extracellular matrix in the muscle layer, and peribronchial fibrosis; goblet cell hyperplasia; and enlargement of the mucus-secreting glands of the trachea and bronchi. Of these, the most striking change is an increase in the size of the mucous glands. This increase can be assessed by the ratio of the thickness of the mucous gland layer to the thickness of the wall between the epithelium and the cartilage **(Reid index)**. The Reid index (normally 0.4) is increased in chronic bronchitis, usually in proportion to the severity and duration of the disease. The mucus plugging, inflammation, and fibrosis may lead to marked narrowing of bronchioles, and in the most severe cases, there may be obliteration of lumen due to fibrosis **(bronchiolitis obliterans)**. The bronchial epithelium may also exhibit squamous metaplasia and dysplasia due to the irritating and mutagenic effects of substances in tobacco smoke.

Clinical Features of COPD

Most affected patients have a smoking history of 40 pack-years or greater. COPD often presents insidiously with slowly

Table 15.4 Predominant Features of Emphysema and Chronic Bronchitis

	Bronchitis	Emphysema
Age, years	40–45	50–75
Dyspnea	Mild; late	Severe; early
Cough	Early; copious sputum	Late; scanty sputum
Infections	Common	Occasional
Respiratory insufficiency	Early, periodic	End-stage
Cor pulmonale	Common	Uncommon, end-stage
Airway resistance	Increased	Normal or slightly increased
Elastic recoil	Normal	Low
Chest radiograph	Prominent vessels; large heart size	Hyperinflation; normal heart size
Appearance	Blue bloater	Pink puffer

increasing dyspnea on exertion and chronic cough with sputum production, slight at first but increasing over time. Other patients present with exacerbations caused by superimposed infection that can lead to confusion with other disorders, such as asthma (due to wheezing). The most important diagnostic test is spirometry, which typically shows an FEV_1/FVC ratio of less than 0.7.

Once COPD appears, symptoms often wax and wane over time and are generally worse in the morning. The clinical picture varies according to the severity of the disease and the relative contributions of emphysematous and bronchitic changes (Table 15.4). At one extreme end of the spectrum lie "*pink puffers,*" patients in whom emphysema dominates. Classically, the patient is barrel-chested and dyspneic, with obviously prolonged expiration, sits forward in a hunched-over position, and breathes through pursed lips. Cough is often slight, overdistention of the lungs is severe, diffusion capacity is low, and blood gas values are relatively normal at rest. Weight loss is common and can be so severe as to suggest an occult cancer. At the other end lie patients with pure chronic bronchitis, ingloriously referred to as "*blue bloaters.*" Their cardinal symptom is a persistent cough productive of sputum, coupled with hypercapnia, hypoxemia, and mild cyanosis. Most patients are somewhere in the middle, with signs and symptoms stemming from both bronchitic and emphysematous changes.

Treatment options include smoking cessation, oxygen therapy, long-acting bronchodilators with inhaled corticosteroids, antibiotics, physical therapy, bullectomy, and, in selected patients, lung volume reduction surgery and lung transplantation. Even with intervention, however, COPD often progresses and frequently proves fatal. Long-standing COPD, particularly in patients with a bronchitic component, commonly leads to pulmonary hypertension, cor pulmonale, and death due to heart failure. Death may also result from acute respiratory failure due to acute infections superimposed on COPD. In patients with emphysematous changes, subpleural blebs may rupture, leading to fatal pneumothorax. The best hope for a major change in this dire picture is more effective programs aimed at prevention of smoking and other environmental exposures.

KEY CONCEPTS

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

- Most common in long-standing tobacco smokers (typically >40 pack-years); air pollutants also contribute.
- Underlying pulmonary pathology usually includes both chronic bronchitis and emphysema.
- Often fatal due to development of heart failure or of respiratory failure due to superimposed infection.

Emphysema

- In COPD, usually follows a centroacinar distribution characterized by permanent enlargement of airspaces distal to terminal bronchioles.
- Particularly severe in patients with α_1-antitrypsin deficiency, in which a panacinar pattern of emphysematous change may be seen.
- Tissue destruction is caused by elastases and oxidants released from inflammatory cells, particularly neutrophils, which are responding to cellular injury caused by tobacco smoke and pollutants.

Chronic Bronchitis

- Defined as persistent productive cough for at least 3 consecutive months in at least 2 consecutive years.
- Dominant pathologic features are mucus hypersecretion due to enlargement of mucus-secreting glands and chronic inflammation associated with bronchiolar wall fibrosis.

Other Forms of Emphysema

In addition to emphysema occurring in the setting of COPD, several other conditions may be associated with lung overinflation or focal emphysematous change and are mentioned here in brief.

- *Compensatory hyperinflation.* This term is used to designate dilation of alveoli in response to loss of lung substance elsewhere, for example, following surgical removal of a lung or lobe with cancer.
- *Obstructive overinflation.* In this condition the lung expands because air is trapped within it. A common cause is subtotal obstruction of an airway by a tumor or foreign object. In infants, it may be caused by *congenital lobar overinflation*, which most often results from hypoplasia of bronchial cartilage. Overinflation occurs either (1) because of an obstruction that acts as ball valve, allowing air to enter on inspiration while preventing its exodus on expiration, or (2) because collaterals bring in air behind the obstruction. These collaterals consist of the *pores of Kohn* and other direct accessory bronchioloalveolar connections (the *canals of Lambert*). Obstructive overinflation can be life-threatening if the affected portion distends sufficiently to compress the adjacent uninvolved lung.
- *Bullous emphysema.* This is a descriptive term for large subpleural blebs or bullae (spaces greater than 1 cm in diameter in the distended state) that can occur in any form of emphysema (Fig. 15.9), often near the apex. Rupture of the bullae may give rise to pneumothorax.
- *Interstitial emphysema.* Entrance of air into the connective tissue stroma of the lung, mediastinum, or subcutaneous

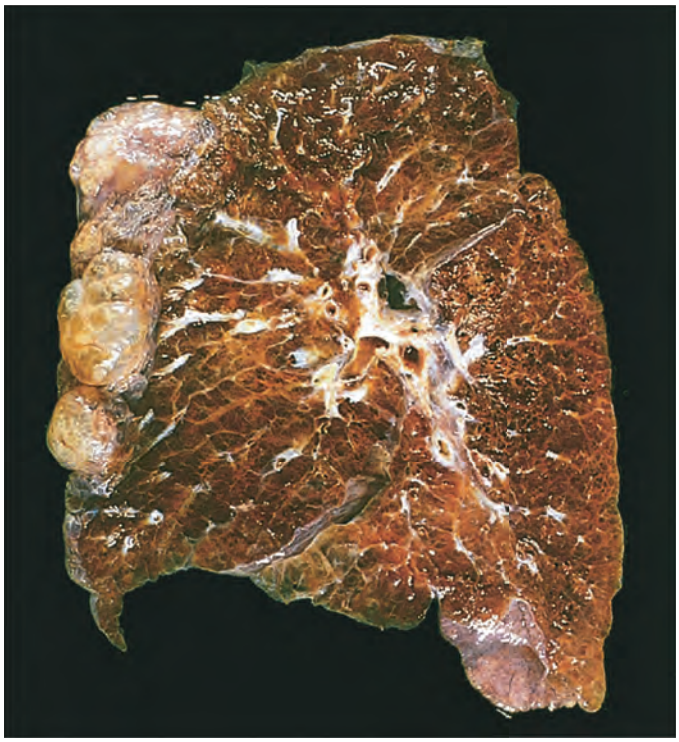


Figure 15.9 Bullous emphysema. Note the large subpleural bullae *(upper left)*.

tissue produces *interstitial emphysema.* In most instances, it is caused by alveolar tears that occur in patients with pulmonary emphysema due to transient increases in intra-alveolar pressure, for example, during coughing. Less commonly, chest wounds or fractured ribs that puncture the lung provide the portal for entrance of air into surrounding soft tissues.

Asthma

Asthma is a heterogeneous disease, usually characterized by chronic airway inflammation and variable expiratory airflow obstruction that produces symptoms such as wheezing, shortness of breath, chest tightness, and cough, which vary over time and in intensity. Symptomatic episodes are most likely to occur at night or in the early morning and are produced by bronchoconstriction that is at least partly reversible, either spontaneously or with treatment. Rarely, an unremitting attack called *acute severe asthma* (formerly known as *status asthmaticus*) may prove fatal; usually, such patients have a long history of asthma. Between the attacks, patients may be virtually asymptomatic. Of note, there has been a significant increase in the incidence of asthma in the Western world over the past 40 to 50 years, a trend that has now started to abate. However, the prevalence of asthma continues to increase in lower income countries and in some ethnic groups in which its prevalence was previously low.

Asthma has several distinct clinical phenotypes, each with different underlying pathogenic mechanisms. It may be categorized as *atopic* (evidence of allergen sensitization and immune activation, often in a patient with allergic rhinitis or eczema) or *nonatopic* (no evidence of allergen sensitization), of which several subtypes exist. In all types, episodes of bronchospasm may have diverse triggers, such as respiratory infections (especially viral infections), irritants (e.g., smoke, fumes), cold air, stress, and exercise. One biologically meaningful and clinically useful way to classify asthma is based on its triggers. We will first briefly describe the various major subtypes of asthma, and then will delve more deeply into its pathogenesis.

Atopic Asthma. This type of asthma is a classic example of an IgE-mediated (type I) hypersensitivity reaction (discussed in Chapter 6). The disease usually begins in childhood and is triggered by environmental allergens, such as dusts, pollens, cockroach or animal dander, and foods, which most frequently act in synergy with other proinflammatory environmental cofactors, most notably respiratory viral infections. A positive family history of asthma is common, and a skin test with the offending antigen in these patients results in an immediate wheal-and-flare reaction. Atopic asthma may also be diagnosed based on high total serum IgE levels or evidence of allergen sensitization by serum radioallergosorbent tests (RASTs), which can detect the presence of IgE antibodies that are specific for individual allergens.

Non-Atopic Asthma. Individuals with non-atopic asthma do not have evidence of allergen sensitization, and skin test results are usually negative. A positive family history of asthma is less common in these patients. Respiratory infections due to viruses (e.g., rhinovirus, parainfluenza virus, and respiratory syncytial virus) are common triggers in non-atopic asthma. Inhaled air pollutants such as tobacco smoke, sulfur dioxide, ozone, and nitrogen dioxide may also contribute to the chronic airway inflammation and hyperreactivity in some cases. As already mentioned, in some instances attacks may be triggered by seemingly innocuous events, such as exposure to cold and even exercise.

Drug-Induced Asthma. Several pharmacologic agents provoke asthma. Aspirin-sensitive asthma is an uncommon type, occurring in individuals with recurrent rhinitis and nasal polyps. These individuals are exquisitely sensitive to small doses of aspirin as well as other non-steroidal anti-inflammatory medications, and they experience not only asthmatic attacks but also urticaria. Aspirin and related drugs trigger asthma in these patients by inhibiting the cyclooxygenase pathway of arachidonic acid metabolism, leading to a rapid decrease in prostaglandin E_2. Normally prostaglandin E_2 inhibits enzymes that generate proinflammatory mediators such as leukotrienes B_4, C_4, D_4, and E_4, which are believed to have central roles in aspirin-induced asthma.

Occupational Asthma. This form of asthma may be triggered by fumes (epoxy resins, plastics), organic and chemical dusts (wood, cotton, platinum), gases (toluene), or other chemicals (formaldehyde, penicillin products). Only minute quantities of chemicals are required to induce the attack, which usually occurs after repeated exposure. The

underlying mechanisms vary according to stimulus and include type I reactions, direct liberation of bronchoconstrictor substances, and hypersensitivity responses of unknown origin.

Pathogenesis

Atopic asthma, the most common form of the disease, is caused by a Th2-mediated IgE response to environmental allergens in genetically predisposed individuals. Airway inflammation is central to the disease pathophysiology and causes airway dysfunction partly through the release of potent inflammatory mediators and partly through remodeling of the airway wall. As the disease becomes more severe, there is increased local secretion of growth factors, which induce mucous gland enlargement, smooth muscle proliferation, angiogenesis, and fibrosis. Varying combinations of these processes help explain the different asthma subtypes, their response to treatment, and their natural history over a person's lifetime.

The contributions of the immune response, genetics, and environment are discussed separately below, although they are closely intertwined.

Th2 Responses, IgE, and Inflammation. **A fundamental abnormality in asthma is an exaggerated Th2 response to normally harmless environmental antigens** (Fig. 15.10). Th2 cells secrete cytokines that promote inflammation and stimulate B cells to produce IgE and other antibodies. These cytokines include IL-4, which stimulates the production of IgE; IL-5, which activates locally recruited eosinophils; and IL-13, which stimulates mucus secretion from bronchial submucosal glands and also promotes IgE production by B cells. The T cells and epithelial cells secrete chemokines that recruit more T cells and eosinophils, thus exacerbating the reaction. As in other allergic reactions (Chapter 6), IgE binds to the Fc receptors on submucosal mast cells, and repeat exposure to the allergen triggers the mast cells to release granule contents and produce cytokines and other mediators, which collectively induce the early-phase (immediate hypersensitivity) reaction and the late-phase reaction.

The early-phase reaction is dominated by bronchoconstriction, increased mucus production, variable degrees of vasodilation, and increased vascular permeability. Bronchoconstriction is triggered by direct stimulation of subepithelial vagal (parasympathetic) receptors through both central and local reflexes triggered by mediators produced by mast cells and other cells in the reaction. The late-phase reaction is dominated by recruitment of leukocytes, notably eosinophils, neutrophils, and more T cells. Although Th2 cells are the dominant T-cell type involved in the disease, other T cells that contribute to the inflammation include Th17 (IL-17 producing) cells, which recruit neutrophils.

Many mediators produced by leukocytes and epithelial cells have been implicated in the asthmatic response. The long list of "suspects" in acute asthma can be ranked based on the clinical efficacy of pharmacologic intervention with antagonists of specific mediators.

- Mediators whose role in bronchospasm is clearly supported by efficacy of pharmacologic intervention are (1) *leukotrienes* C_4, D_4, and E_4, which cause prolonged bronchoconstriction as well as increased vascular permeability and increased mucus secretion; (2) *acetylcholine,* released from intrapulmonary parasympathetic nerves, which can cause airway smooth muscle constriction by directly stimulating muscarinic receptors; (3) *IL-5,* antagonists of which are effective in treating severe forms of asthma that are associated with peripheral blood eosinophilia; and (4) *galectin-10* (GAL10), which is released from eosinophils and forms crystals known as *Charcot-Leyden crystals*. Long recognized as a feature of asthma, recent studies have shown that these crystals are strong inducers of inflammation and mucus production.
- A second group of agents are present at the "scene of the crime" but seem to have relatively minor contributions on the basis of lack of efficacy of potent antagonists or synthesis inhibitors. These include (1) *histamine,* a potent bronchoconstrictor; (2) *prostaglandin* D_2, which elicits bronchoconstriction and vasodilation; and (3) *platelet-activating factor,* which causes aggregation of platelets and release of serotonin from their granules. These mediators might yet prove important in certain types of chronic or non-allergic asthma.
- Finally, a large third group comprises "suspects" for whom specific antagonists or inhibitors are not available or have been insufficiently studied as yet. These include IL-4, IL-13, TNF, chemokines (e.g., eotaxin, also known as CCL11), neuropeptides, nitric oxide, bradykinin, and endothelins.

It is thus clear that multiple mediators contribute to the acute asthmatic response. Moreover, the composition of this "mediator soup" likely varies among individuals or different types of asthma. The appreciation of the importance of inflammatory cells and mediators in asthma has led to greater emphasis on anti-inflammatory drugs, such as corticosteroids, in its treatment.

Genetic Susceptibility. Susceptibility to atopic asthma is multigenic and often associated with increased incidence of other allergic disorders, such as allergic rhinitis (hay fever) and eczema. Genetic polymorphisms linked to asthma and other allergic disorders were described in Chapter 6. Suffice it to say here that many of these are likely to influence immune responses and the subsequent inflammatory reaction. Some of the stronger or more interesting genetic variants associated with asthma include the following:

- A susceptibility locus for asthma located on chromosome 5q, near the gene cluster encoding the cytokines IL-3, IL-4, IL-5, IL-9, and IL-13 and the IL-4 receptor. Among the genes in this cluster, polymorphisms in the *IL13* gene have the strongest and most consistent associations with asthma or allergic disease, while IL-4 receptor gene variants are associated with atopy, elevated total serum IgE, and asthma.
- Particular class II HLA alleles linked to production of IgE antibodies against some antigens, such as ragweed pollen.
- Variants associated with the genes encoding IL-33, a member of the IL-1 family of cytokines, and its receptor, ST2, which induce the production of Th2 cytokines.
- Variants associated with the gene encoding thymic stromal lymphopoietin (TSLP), a cytokine produced by epithelium that may have a role in initiating allergic reactions.

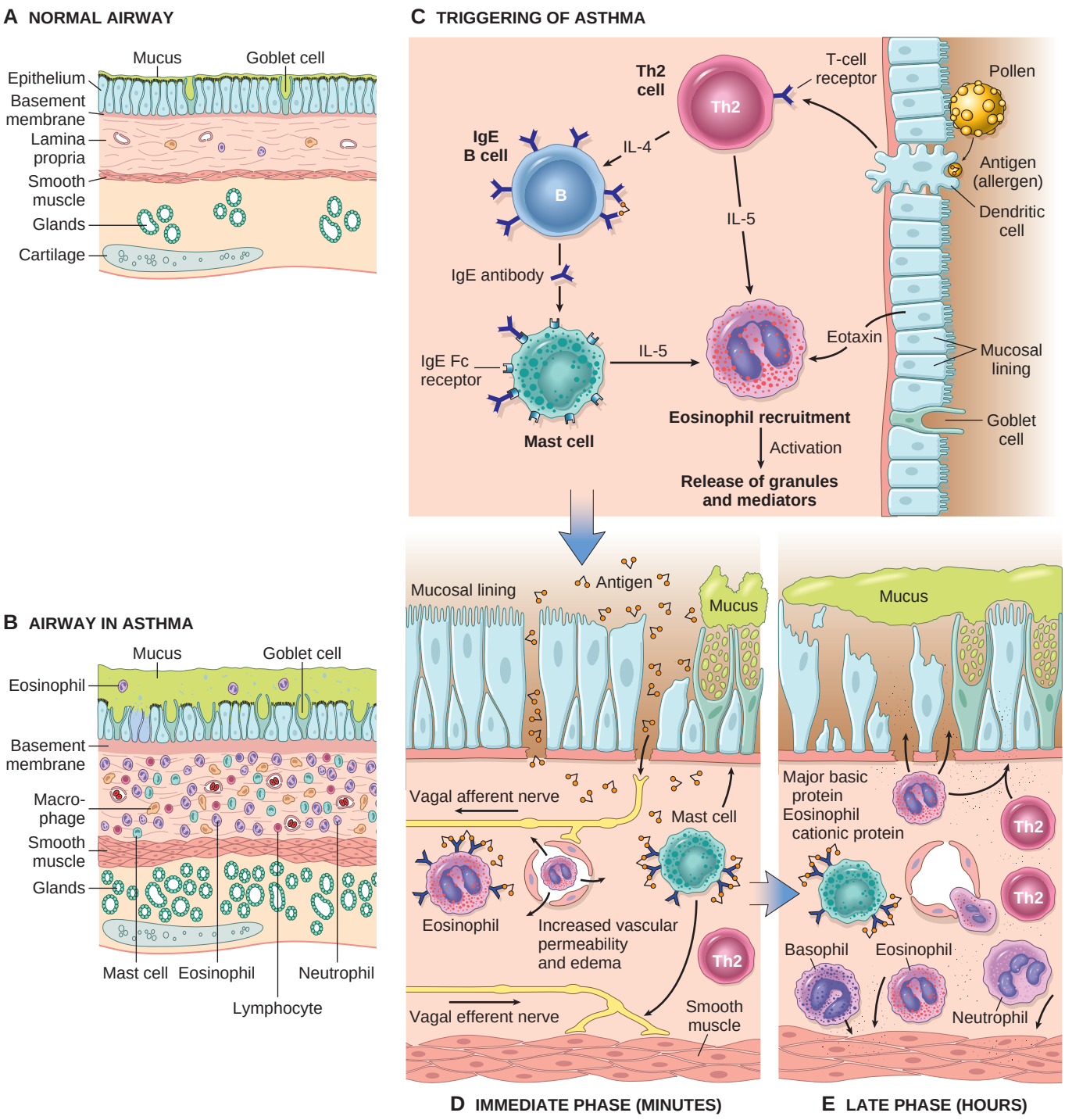


Figure 15.10 (A and B) Comparison of a normal airway and an airway involved by asthma. The asthmatic airway is marked by accumulation of mucus in the bronchial lumen secondary to an increase in the number of mucus-secreting goblet cells in the mucosa and hypertrophy of submucosal glands; intense chronic inflammation due to recruitment of eosinophils, macrophages, and other inflammatory cells; thickened basement membrane; and hypertrophy and hyperplasia of smooth muscle cells. (C) Inhaled allergens (antigen) elicit a Th2-dominated response favoring IgE production and eosinophil recruitment. (D) On re-exposure to antigen, the immediate reaction is triggered by antigen-induced cross-linking of IgE bound to Fc receptors on mast cells. These cells release preformed mediators that directly and via neuronal reflexes induce bronchospasm, increased vascular permeability, mucus production, and recruitment of leukocytes. (E) Leukocytes recruited to the site of reaction (neutrophils, eosinophils, and basophils; lymphocytes and monocytes) release additional mediators that initiate the late phase reaction. Several factors released from eosinophils (e.g., major basic protein, eosinophil cationic protein) also cause damage to the epithelium. *IL-5,* Interleukin-5.

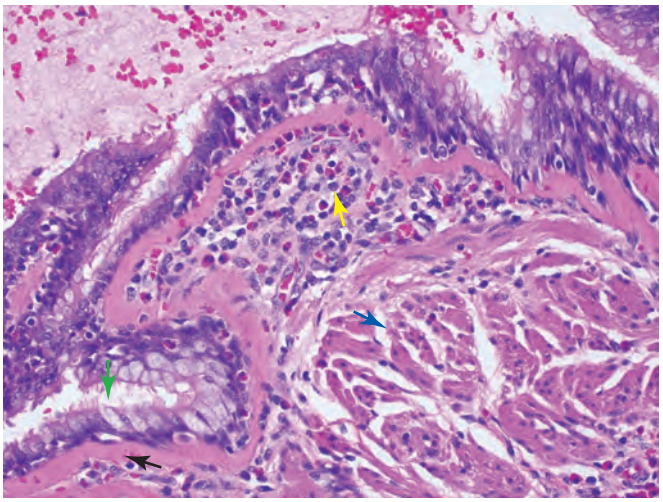


Figure 15.11 Bronchus from an asthmatic patient showing goblet cell hyperplasia *(green arrow)*, sub–basement membrane fibrosis *(black arrow)*, eosinophilic inflammation *(yellow arrow)*, and muscle hypertrophy *(blue arrow)*.

Environmental Factors. Asthma is a disease of industrialized societies where the majority of people live in cities. Two ideas, neither wholly satisfying, have been proposed to explain this association. First, industrialized environments contain many airborne pollutants that can serve as allergens to initiate the Th2 response. Second, city life tends to limit the exposure of very young children to certain antigens, particularly microbial antigens, and exposure to such antigens may protect children from asthma and atopy. The idea that microbial exposure during early life reduces the later incidence of allergic (and some autoimmune) diseases has been popularized as the hygiene hypothesis. Although the underlying mechanisms of this protective effect are unclear, it has spurred trials of probiotics and intentional early exposure of children to putative allergens to decrease their risk of later developing allergies.

Infections do not cause asthma by themselves, but may be important co-factors. Young children with aeroallergen sensitization who develop lower respiratory tract viral infections (rhinovirus type C, respiratory syncytial virus) have a 10- to 30-fold increased risk of developing persistent and/or severe asthma. Both viral and bacterial infections (identified by cultures and non-culture tools) are associated with acute exacerbations of the disease.

Over time, repeated bouts of allergen exposure and immune reactions result in structural changes in the bronchial wall, referred to as *airway remodeling*. These changes, described later in greater detail, include hypertrophy and hyperplasia of bronchial smooth muscle, epithelial injury, increased airway vascularity, subepithelial mucous gland enlargement, and subepithelial fibrosis.

A small subset of patients with asthma, many of whom have severe disease that is refractory to glucocorticoids, has inflammatory infiltrates that are enriched for neutrophils rather than eosinophils. This form of the disease may be driven by a Th17 T-cell response to chronic bacterial colonization of the lung.

MORPHOLOGY

In patients dying of acute severe asthma (status asthmaticus), the lungs are overinflated and contain small areas of atelectasis. The most striking gross finding is occlusion of bronchi and bronchioles by thick, tenacious mucus plugs, which often contain shed epithelium. A characteristic finding in sputum or bronchoalveolar lavage specimens of patients with atopic asthma is **Curschmann spirals,** which may result from extrusion of mucus plugs from subepithelial mucous gland ducts or bronchioles. Also present are numerous eosinophils and **Charcot-Leyden crystals** composed of the eosinophil-derived protein galectin-10. The other characteristic histologic findings of asthma, collectively called **airway remodeling** (Figs. 15.10B and 15.11), include:

- Thickening of airway wall
- Sub–basement membrane fibrosis (due to deposition of type I and III collagens)
- Increased vascularity
- Increase in the size of the submucosal glands and number of airway goblet cells
- Hypertrophy and/or hyperplasia of the bronchial wall muscle with increased extracellular matrix

While acute airflow obstruction is primarily attributed to muscular bronchoconstriction, acute edema, and mucus plugging, airway remodeling may contribute to chronic irreversible airway obstruction.

Clinical Features

A classic acute asthmatic attack lasts up to several hours. In some patients, however, the cardinal symptoms of chest tightness, dyspnea, wheezing, and coughing (with or without sputum production) are present at a low level constantly. In its most severe form, acute severe asthma, the paroxysm persists for days or even weeks, sometimes causing airflow obstruction that is so extreme that marked cyanosis or even death ensues.

The diagnosis is based on demonstration of an increase in airflow obstruction (from baseline levels); difficulty with exhalation (prolonged expiration, wheeze); and (in those with atopic asthma) identification of eosinophilia in the peripheral blood and eosinophils, Curschmann spirals, and Charcot-Leyden crystals in the sputum. In the usual case with intervals of freedom from respiratory difficulty, the disease is more discouraging and disabling than lethal, and most individuals are able to maintain a productive life.

Therapy is based on the severity of the disease. The centerpieces of standard therapy are bronchodilators, glucocorticoids, and leukotriene antagonists. For severe and difficult-to-treat asthma in adolescents and adults, novel biologic therapies targeting inflammatory mediators such as IL-5 blocking antibodies, which is effective in severe asthma associated with Th2 immune responses and peripheral blood eosinophilia, are now available. Up to 50% of childhood asthma remits in adolescence only to return in adulthood in a significant number of patients. In other cases there is a variable decline in baseline lung function over time.

KEY CONCEPTS

ASTHMA

- Asthma is characterized by reversible bronchoconstriction caused by airway hyperresponsiveness to a variety of stimuli.
- Atopic asthma is caused by a Th2 and IgE-mediated immunologic reaction to environmental allergens and is characterized by acute-phase (immediate) and late-phase reactions. The Th2 cytokines IL-4, IL-5, and IL-13 are important mediators.
- Triggers for non-atopic asthma are less clear but include viral infections and inhaled air pollutants, which can also trigger atopic asthma.
- Eosinophils are key inflammatory cells in atopic asthma; other inflammatory cells implicated in its pathogenesis include mast cells, neutrophils, and T lymphocytes.
- Airway remodeling (sub–basement membrane fibrosis, hypertrophy of bronchial glands, and smooth muscle hyperplasia) adds an irreversible component to the obstructive disease.

Bronchiectasis

Bronchiectasis is a disorder in which destruction of smooth muscle and elastic tissue by inflammation stemming from persistent or severe infections leads to permanent dilation of bronchi and bronchioles. Because of better control of lung infections, bronchiectasis is now uncommon, but may still develop in association with the following:

- *Congenital or hereditary conditions that predispose to chronic infections,* including cystic fibrosis, intralobar sequestration of the lung, immunodeficiency states, primary ciliary dyskinesia, and Kartagener syndrome.
- *Severe necrotizing pneumonia* caused by bacteria, viruses, or fungi; this may be a single severe episode or recurrent infections.
- *Bronchial obstruction,* due to tumor, foreign body, or mucus impaction; in each instance the bronchiectasis is localized to the obstructed lung segment.
- *Immune disorders,* including rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease, and the posttransplant setting (chronic rejection after lung transplant and chronic graft-versus-host disease after hematopoietic stem cell transplantation).
- Up to 50% of cases are *idiopathic,* lacking the aforementioned associations, in which there appears to be dysfunctional host immunity to infectious agents leading to chronic inflammation.

Pathogenesis

Obstruction and infection are the major conditions associated with bronchiectasis. **The infections that lead to bronchiectasis are usually the result of a defect in airway clearance.** Sometimes this defect stems from airway obstruction, leading to distal pooling of secretions.

Both mechanisms are readily apparent in a severe form of bronchiectasis that is associated with cystic fibrosis (Chapter 10). In cystic fibrosis the primary defect in ion transport results in thick viscous secretions that perturb mucociliary clearance and lead to airway obstruction. This sets the stage for chronic bacterial infections, which cause widespread damage to airway walls. With destruction of supporting smooth muscle and elastic tissue, the bronchi become markedly dilated, while smaller bronchioles are progressively obliterated as a result of fibrosis (bronchiolitis obliterans).

Primary ciliary dyskinesia is an autosomal recessive disease with a frequency of 1 in 10,000 to 20,000 births. The disease-causing mutations result in ciliary dysfunction due to defects in ciliary motor proteins (e.g., mutations involving dynein), again preventing mucociliary clearance, setting the stage for recurrent infections that lead to bronchiectasis. Ciliary function also is necessary during embryogenesis to ensure proper rotation of the developing organs in the chest and abdomen; in its absence, their location becomes a matter of chance. As a result, approximately half of patients with primary ciliary dyskinesia have *Kartagener syndrome,* marked by situs inversus or a partial lateralizing abnormality associated with bronchiectasis and sinusitis. Males with this condition also tend to be infertile as a result of sperm dysmotility.

Allergic bronchopulmonary aspergillosis occurs in patients with asthma or cystic fibrosis and frequently leads to the development of bronchiectasis. It is a hyperimmune response to the fungus *Aspergillus fumigatus.* Sensitization to *Aspergillus* leads to activation of Th2 helper T cells, which release cytokines that recruit eosinophils and other leukocytes. Characteristically, there are high serum IgE levels, serum antibodies to *Aspergillus,* intense airway inflammation with eosinophils, and formation of mucus plugs, which play a primary role in the development of bronchiectasis.

MORPHOLOGY

Bronchiectasis usually affects the lower lobes bilaterally, particularly air passages that are vertical, and is most severe in the more distal bronchi and bronchioles. When tumor or aspiration of foreign bodies leads to bronchiectasis, the involvement is localized. **The airways are dilated, sometimes up to four times normal size.** Characteristically, the bronchi and bronchioles are so dilated that they can be followed almost to the pleural surfaces. By contrast, in the normal lung, the bronchioles cannot be followed by eye beyond a point 2 to 3 cm from the pleural surfaces. On the cut surface of the lung, the dilated bronchi appear cystic and are filled with mucopurulent secretions (Fig. 15.12).

The histologic findings vary with the activity and chronicity of the disease. In the full-blown, active case there is an intense acute and chronic inflammatory exudation within the walls of the bronchi and bronchioles, associated with desquamation of the lining epithelium and extensive ulceration. There also may be squamous metaplasia of the remaining epithelium in response to chronic inflammation, further diminishing mucociliary clearance. In some instances, necrosis destroys the bronchial or bronchiolar walls and forms a lung abscess. Fibrosis of the bronchial and bronchiolar walls and peribronchiolar fibrosis develop in more chronic cases, leading to varying degrees of subtotal or total obliteration of bronchiolar lumens.

Haemophilus influenzae is found in approximately half and *Pseudomonas aeruginosa* in 12% to 30% of sputum cultures from patients with bronchiectasis, with four other types of bacteria

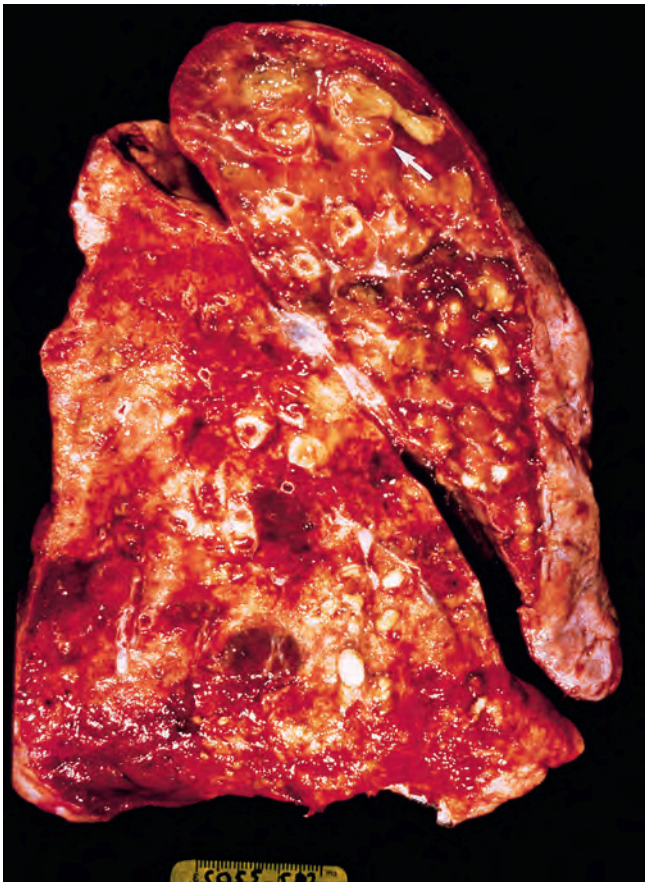


Figure 15.12 Bronchiectasis in a patient with cystic fibrosis, who underwent lung transplantation. Cut surface of lung shows markedly distended peripheral bronchi filled with mucopurulent secretions.

(including non-tuberculous mycobacteria) making up most of the remaining cases in patients from different geographic locations. This observation suggests that the inflamed, mucoid, sometimes anaerobic, bronchiectatic microenvironment favors colonization by a relatively few microbial species. In allergic bronchopulmonary aspergillosis, fungal hyphae can be seen on special stains within the mucoinflammatory contents of the dilated segmental bronchi. In late stages the fungus may infiltrate the bronchial wall.

Clinical Features

Bronchiectasis causes severe, persistent cough; expectoration of foul smelling, sometimes bloody sputum; dyspnea and orthopnea in severe cases; and, on occasion, hemoptysis, which may be massive. Symptoms are often episodic and are precipitated by upper respiratory tract infections or the introduction of new pathogenic agents. Paroxysms of cough are particularly frequent when the patient rises in the morning, as the change in position causes collections of pus and secretions to drain into the bronchi. Obstructive respiratory insufficiency can lead to marked dyspnea and cyanosis. However, current treatments with better antibiotics and physical therapy have improved outcomes considerably, and life expectancy has almost doubled. Hence, cor pulmonale, brain abscesses, and amyloidosis are less frequent complications of bronchiectasis currently than in the past.

CHRONIC DIFFUSE INTERSTITIAL (RESTRICTIVE) DISEASES

Restrictive lung disorders fall into two general categories: (1) *chronic interstitial and infiltrative diseases,* such as pneumoconioses and interstitial fibrosis of unknown etiology, and (2) *chest wall disorders* (e.g., neuromuscular diseases such as poliomyelitis, severe obesity, pleural diseases, and kyphoscoliosis), which are not discussed here.

Chronic interstitial pulmonary diseases are a heterogeneous group of disorders characterized predominantly by inflammation and fibrosis of the lung interstitium associated with pulmonary function studies indicative of restrictive lung disease. Diffuse restrictive diseases are categorized based on histology and clinical features (Table 15.5). Many of the entities are of unknown cause and pathogenesis, and some have an intra-alveolar as well as an interstitial component. Patients have dyspnea, tachypnea, end-inspiratory crackles, and eventual cyanosis, without wheezing or other evidence of airway obstruction. The classic functional abnormalities are reductions in diffusion capacity, lung volume, and lung compliance. Chest radiographs show bilateral lesions that take the form of small nodules, irregular lines, or *ground-glass shadows,* all corresponding to areas of interstitial fibrosis. Although the entities can often be distinguished in their early stages, advanced forms are hard to differentiate because all result in diffuse scarring of the lung, often referred to as *end-stage lung* or *honeycomb lung*. Eventually, secondary pulmonary hypertension and right-sided heart failure (cor pulmonale) may result.

Fibrosing Diseases

Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis (IPF) refers to a clinicopathologic syndrome marked by progressive interstitial pulmonary fibrosis and respiratory failure. In Europe the

Table 15.5 Major Categories of Chronic Interstitial Lung Disease

Fibrosing
Usual interstitial pneumonia (idiopathic pulmonary fibrosis)
Nonspecific interstitial pneumonia
Cryptogenic organizing pneumonia
Connective tissue disease–associated
Pneumoconiosis
Drug reactions
Radiation pneumonitis
Granulomatous
Sarcoidosis
Hypersensitivity pneumonitis
Eosinophilic
Smoking-Related
Desquamative interstitial pneumonia
Respiratory bronchiolitis–associated interstitial lung disease
Other
Langerhans cell histiocytosis
Pulmonary alveolar proteinosis
Lymphoid interstitial pneumonia

term *cryptogenic fibrosing alveolitis* is more popular. IPF has characteristic radiologic, pathologic, and clinical features. The histologic pattern of fibrosis is referred to as usual interstitial pneumonia (UIP), which can often be diagnosed based on its characteristic appearance on computed tomography scans. The UIP pattern can also be seen in other diseases, notably connective tissue diseases, chronic hypersensitivity pneumonia, and asbestosis; these must be distinguished from IPF based on other clinical, laboratory, and histologic features.

Pathogenesis

While the cause of IPF remains unknown, it appears that it arises in genetically predisposed individuals who are prone to aberrant repair of recurrent alveolar epithelial cell injuries caused by environmental exposures (Fig. 15.13). The implicated factors are as follows:

- *Environmental factors.* Most important among these is cigarette smoking, which increases the risk of IPF several-fold. IPF incidence is also increased in individuals who are exposed to air pollution, microaspiration, metal fumes, and wood dust, or who work in certain occupations, including farming, hairdressing, and stone polishing. It is hypothesized that exposure to environmental irritants or toxins in each of these contexts causes recurrent alveolar epithelial cell damage.
- *Genetic factors.* The vast majority of individuals who smoke or who have other environmental exposures linked to IPF do not develop the disorder, indicating that additional factors are required for its development. Mutations in the *TERT, TERC, PARN,* and *RTEL1* genes, all of which are involved with the maintenance of telomeres, are associated with increased risk of IPF. You will recall that maintenance of telomeres (the ends of chromosomes) is necessary to prevent cellular senescence. Up to 15% of familial IPF is associated with inherited defects in genes that maintain telomeres, while up to 25% of sporadic IPF cases are associated with abnormal telomere shortening in peripheral blood lymphocytes, a finding that also points to a problem with telomere maintenance. Other, rare familial forms of IPF are associated with mutations in genes encoding components of surfactant; these mutations create folding defects in the affected proteins, leading to activation of the unfolded protein response in type II pneumocytes. This in turn appears to make pneumocytes more sensitive to environmental insults, leading to cellular dysfunction and injury. Finally, roughly one-third of IPF cases are associated with a single-nucleotide polymorphism in the promoter of the *MUC5B* gene that greatly increases the secretion of MUC5B, a member of the mucin family. This may in turn alter mucociliary clearance, but precisely how this alteration relates to IPF risk is uncertain.
- *Age.* IPF is a disease of older individuals, rarely appearing before the age of 50 years. Whether this association stems from aging-related telomere shortening or from other acquired changes associated with aging is unknown.

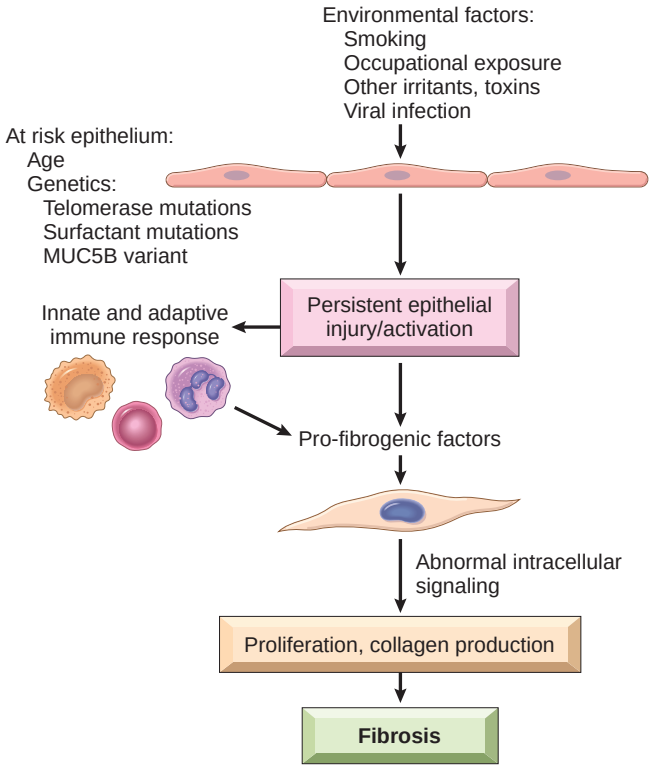


Figure 15.13 Proposed pathogenic mechanisms in idiopathic pulmonary fibrosis. Environmental factors that are potentially injurious to alveolar epithelium interact with genetic or aging-related factors that place epithelium at risk, creating a persistent epithelial injury. Factors secreted from injured/activated epithelium, possibly augmented by factors released from innate and adaptive immune cells responding to "danger" signals produced by damaged epithelium, activate interstitial fibroblasts. There is some evidence that these activated fibroblasts exhibit signaling abnormalities that lead to increased signaling through the PI3K/AKT pathway. The activated fibroblasts synthesize and deposit collagen, leading to interstitial fibrosis and eventual respiratory failure.

It is easy to imagine how some of the factors cited herein might combine to exacerbate alveolar epithelial cell damage and senescence, which seems to be the initiating event in IPF, but it must be admitted that the pathogenesis of IPF is complex and poorly understood. For example, it is unknown precisely how alveolar epithelial cell damage translates into interstitial fibrosis. One model holds that the injured epithelial cells are the source of profibrogenic factors such as TGF-β, whereas a second, non–mutually exclusive model proposes that innate and adaptive immune cells produce such factors as part of the host response to epithelial cell damage. Other work has described abnormalities in the fibroblasts themselves that involve changes in intracellular signaling and features reminiscent of epithelial mesenchymal transition (Chapter 7), but a causal link between these alterations and fibrosis has not been established.

MORPHOLOGY

Grossly, the pleural surfaces of the lung are cobblestoned as a result of the retraction of scars along the interlobular septa. The cut surface shows firm, rubbery white areas of fibrosis, which occurs preferentially in the lower lobes, the **subpleural regions,** and along the **interlobular septa.** Microscopically, the hallmark is **patchy interstitial fibrosis,** which varies in intensity (Fig. 15.14) and age. The earliest lesions contain an exuberant proliferation of fibroblasts **(fibroblastic foci).** With time these areas become more fibrotic and less cellular. Quite typical is the

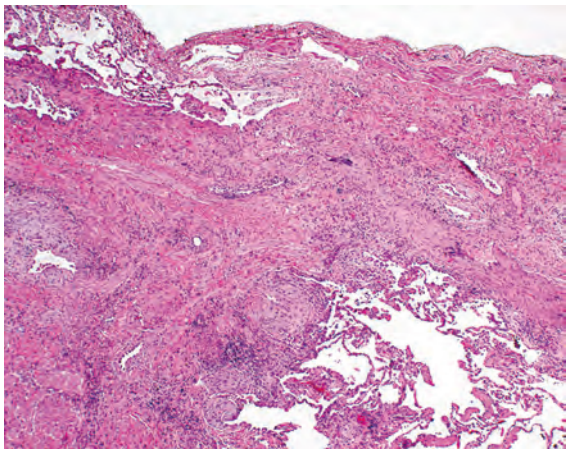


Figure 15.14 Usual interstitial pneumonia. The fibrosis is more pronounced in the subpleural region. (Courtesy Dr. Nicole Cipriani, Department of Pathology, University of Chicago, Chicago, Ill.)

coexistence of both early and late lesions (Fig. 15.15). The dense fibrosis causes the destruction of alveolar architecture and the formation of cystic spaces lined by hyperplastic type II pneumocytes or bronchiolar epithelium **(honeycomb fibrosis)**. With adequate sampling, these diagnostic histologic changes (i.e., areas of dense fibrosis and fibroblastic foci) can be identified even in advanced IPF. There is mild to moderate inflammation within the fibrotic areas, consisting of mostly lymphocytes admixed with a few plasma cells, neutrophils, eosinophils, and mast cells. Foci of squamous metaplasia and smooth muscle hyperplasia may be present, along with pulmonary arterial hypertensive changes (intimal fibrosis and medial thickening). In acute exacerbations, DAD may be superimposed on these chronic changes.

Clinical Features

IPF begins insidiously with gradually increasing dyspnea on exertion and dry cough. Most patients are 55 to 75 years old at presentation. Hypoxemia, cyanosis, and clubbing occur late in the course. The course in individual patients is unpredictable. Usually there is slowly progressive respiratory failure, but some patients have acute exacerbations and follow a rapid downhill clinical course. The median survival is about 3.8 years after diagnosis. Lung transplantation is the only definitive therapy; however, two drugs, a tyrosine kinase inhibitor and a TGF-β antagonist, have both been shown to slow disease progression and represent the first effective (albeit modestly so) targeted therapies for IPF.

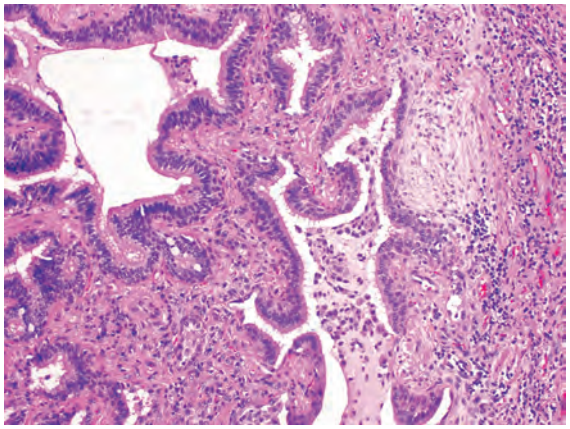


Figure 15.15 Usual interstitial pneumonia. Fibroblastic focus with fibers running parallel to surface and bluish myxoid extracellular matrix. Honeycombing is present on the left.

Nonspecific Interstitial Pneumonia

Despite its "nonspecific" name, this entity is important to recognize, since these patients have a much better prognosis than patients with UIP. Nonspecific interstitial pneumonia is most often associated with connective tissue disease but may also be idiopathic.

MORPHOLOGY

On the basis of its histology, nonspecific interstitial pneumonia is divided into cellular and fibrosing patterns. The cellular pattern consists primarily of mild to moderate chronic interstitial inflammation, containing lymphocytes and a few plasma cells, in a uniform or patchy distribution. The fibrosing pattern consists of diffuse or patchy interstitial fibrotic lesions of roughly the same stage of development, an important distinction from UIP. Fibroblastic foci, honeycombing, hyaline membranes, and granulomas are absent.

Clinical Features

Patients present with dyspnea and cough of several months' duration. They are more likely to be female nonsmokers in their sixth decade of life. On imaging, the lesions have the appearance of bilateral, symmetric, predominantly lower lobe reticular opacities. Patients having the cellular pattern are somewhat younger than those with the fibrosing pattern and have a better prognosis.

Cryptogenic Organizing Pneumonia

Cryptogenic organizing pneumonia is most often seen as a response to infection or inflammatory injury of the lungs. It has been associated with viral and bacterial pneumonias, inhaled toxins, drugs, connective tissue disease, and graft-versus-host disease in hematopoietic stem cell transplant recipients. Patients present with cough and dyspnea and have patchy subpleural or peribronchial areas of airspace consolidation radiographically. Histologically, it is characterized by the presence of polypoid plugs of loose organizing connective tissue (Masson bodies) within alveolar ducts, alveoli, and often bronchioles (Fig. 15.16). The connective tissue is all of the same age, and the underlying lung architecture is normal. There is no interstitial fibrosis or honeycomb lung. Some patients recover spontaneously, but most need treatment with oral steroids for 6 months or longer for complete recovery. The long-term prognosis is dependent on the underlying disorder.

Pulmonary Involvement in Autoimmune Diseases

Many autoimmune diseases (also referred to as connective tissue diseases because of their frequent association with arthritis) can involve the lung at some point in their course. Those that are well recognized for producing pulmonary disease include systemic lupus erythematosus, rheumatoid arthritis, progressive systemic sclerosis (scleroderma), and dermatomyositis-polymyositis. Pulmonary involvement can take different histologic patterns; nonspecific interstitial

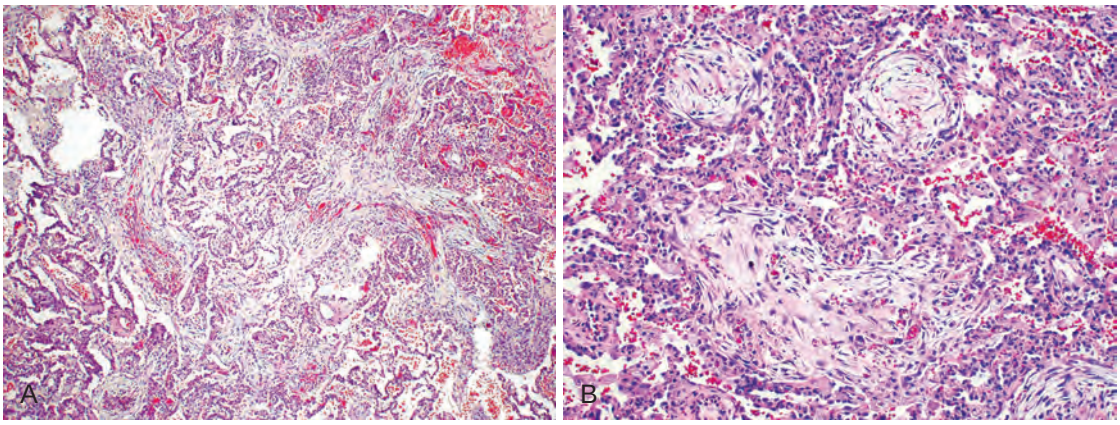


Figure 15.16 Cryptogenic organizing pneumonia. Some alveolar spaces are filled with balls of fibroblasts (Masson bodies), while the alveolar walls are relatively normal. (A) Low power. (B) High power.

pneumonia, usual interstitial pneumonia, organizing pneumonia, and bronchiolitis are the most common.

- *Rheumatoid arthritis* is associated with pulmonary involvement in 30% to 40% of patients as (1) chronic pleuritis, with or without effusion; (2) diffuse interstitial pneumonitis and fibrosis; (3) intrapulmonary rheumatoid nodules; (4) follicular bronchiolitis; or (5) pulmonary hypertension. When lung disease occurs in the setting of rheumatoid arthritis and a pneumoconiosis (described next), it is referred to as *Caplan syndrome*.
- *Systemic sclerosis* (scleroderma) is associated with diffuse interstitial fibrosis (nonspecific interstitial pattern more common than usual interstitial pattern) and pleural involvement.
- *Lupus erythematosus* may cause patchy, transient parenchymal infiltrates or occasionally severe lupus pneumonitis, as well as pleuritis and pleural effusions.

Pulmonary involvement in these diseases has a variable prognosis that is determined by the extent and histologic pattern of involvement.

KEY CONCEPTS

CHRONIC INTERSTITIAL LUNG DISEASES

- Diffuse interstitial fibrosis of the lung gives rise to restrictive lung diseases characterized by reduced lung compliance and reduced FVC. The ratio of FEV_1 to FVC is normal.
- Ididopathic pulmonary fibrosis is prototypic of restrictive lung diseases. It is characterized by patchy interstitial fibrosis, fibroblastic foci, and formation of cystic spaces (honeycomb lung). This histologic pattern is known as usual interstitial fibrosis.
- The cause of idiopathic pulmonary fibrosis is unknown, but genetic analyses point to roles for senescence of alveolar epithelium (due to telomere shortening), altered mucin production, and abnormal signaling in alveolar fibroblasts. Injury to alveolar epithelial cells sets in motion events that lead to increased local production of fibrogenic cytokines, such as TGF-β.

Pneumoconioses

The term *pneumoconiosis,* originally coined to describe the non-neoplastic lung reaction to inhalation of mineral dusts encountered in the workplace, now also includes disease induced by chemical fumes and vapors. A simplified classification is presented in Table 15.6. Where implemented, regulations limiting worker exposure have resulted in a marked decrease in dust-associated diseases.

Pathogenesis

Specific factors that influence the development of dust-borne pneumoconiosis include the following:

- *Dust retention,* which is determined by the dust concentration in ambient air, the duration of exposure, and the effectiveness of clearance mechanisms. Any influence, such as cigarette smoking, that impairs mucociliary clearance significantly increases the accumulation of dust in the lungs.
- *Particle size.* The most dangerous particles are from 1 to 5 μm in diameter because particles of this size can reach the terminal small airways and air sacs and deposit in their linings.
- *Particle solubility and cytotoxicity,* which are influenced by particle size. In general, small particles composed of injurious substances of high solubility are more likely to produce rapid-onset acute lung injury. By contrast, larger particles are more likely to resist dissolution and may persist within the lung parenchyma for years. These tend to evoke fibrosing collagenous pneumoconioses, such as is characteristic of *silicosis.*
- *Particle uptake by epithelial cells or egress across epithelial linings,* which allows direct interactions with fibroblasts and interstitial macrophages to occur. Some particles may reach the lymph nodes through lymphatic drainage directly or within migrating macrophages and thereby initiate an adaptive immune response to components of the particulates or to self-proteins modified by the particles or both.
- *Activation of the inflammasome* (Chapter 3), which occurs following the phagocytosis of certain particles by macrophages. This innate immune response amplifies the intensity and the duration of the local reaction.
- *Tobacco smoking,* which worsens the effects of all inhaled mineral dusts, but particularly those caused by asbestos.

Table 15.6 Lung Diseases Caused by Air Pollutants

Agent	Disease	Exposure
Mineral Dusts		
Coal dust	Anthracosis Macules Progressive massive fibrosis Caplan syndrome	Coal mining (particularly hard coal)
Silica	Silicosis Caplan syndrome	Metal casting work, sandblasting, hard rock mining, stone cutting, others
Asbestos	Asbestosis Pleural plaques Caplan syndrome Mesothelioma Carcinoma of the lung, larynx, stomach, colon	Mining, milling, manufacturing, and installation and removal of insulation
Beryllium	Acute berylliosis Beryllium granulomatosis Lung carcinoma (?)	Mining, manufacturing
Iron oxide	Siderosis	Welding
Barium sulfate	Baritosis	Mining
Tin oxide	Stannosis	Mining
Organic Dusts That Induce Hypersensitivity Pneumonitis		
Moldy hay	Farmer's lung	Farming
Bagasse	Bagassosis	Manufacturing wallboard, paper
Bird droppings	Bird breeder's lung	Bird handling
Organic Dusts That Induce Asthma		
Cotton, flax, hemp	Byssinosis	Textile manufacturing
Red cedar dust	Asthma	Lumbering, carpentry
Chemical Fumes and Vapors		
Nitrous oxide, sulfur dioxide, ammonia, benzene, insecticides	Bronchitis, asthma Pulmonary edema ARDS Mucosal injury Fulminant poisoning	Occupational and accidental exposure

ARDS, Acute respiratory distress syndrome.

In general, only a small percentage of exposed people develop occupational respiratory diseases, implying a genetic predisposition to their development. Many of the diseases listed in Table 15.6 are quite uncommon; hence only a select few that cause pulmonary fibrosis are presented next.

Coal Workers' Pneumoconiosis

Coal workers' pneumoconiosis is lung disease caused by inhalation of coal particles and other admixed forms of dust. Dust reduction measures in coal mines around the globe have drastically reduced its incidence. The spectrum of lung findings in coal workers is wide, varying from asymptomatic anthracosis, to simple coal workers' pneumoconiosis with little to no pulmonary dysfunction, to complicated coal workers' pneumoconiosis, or *progressive massive fibrosis,* in which lung function is compromised. Contaminating silica in the coal dust favors the development of progressive disease. In most cases, carbon dust itself is the major culprit, and studies have shown that complicated lesions contain much more dust than simple lesions. Coal workers may also develop emphysema and chronic bronchitis independent of smoking.

MORPHOLOGY

Carbon deposits are dark black in color and are readily visible grossly and microscopically. **Anthracosis** is the most innocuous coal-induced pulmonary lesion in coal miners and is also seen to some degree in urban dwellers and tobacco smokers. Inhaled carbon pigment is engulfed by alveolar or interstitial macrophages, which accumulate in the connective tissue adjacent to the lymphatics and in organized lymphoid tissue adjacent to the bronchi or in the lung hilus.

Simple coal workers' pneumoconiosis is characterized by **coal macules** (1 to 2 mm in diameter) and somewhat larger **coal nodules.** Coal macules consist of carbon-laden macrophages; nodules also contain a delicate network of collagen fibers. Although these lesions are scattered throughout the lung, the upper lobes and upper zones of the lower lobes are more heavily involved. They are located primarily adjacent to respiratory bronchioles, the site of initial dust accumulation. In due course dilation of adjacent alveoli occurs, sometimes giving rise to **centrilobular emphysema.**

Complicated coal workers' pneumoconiosis (progressive massive fibrosis) occurs on a background of simple disease and generally requires many years to develop. It is characterized by

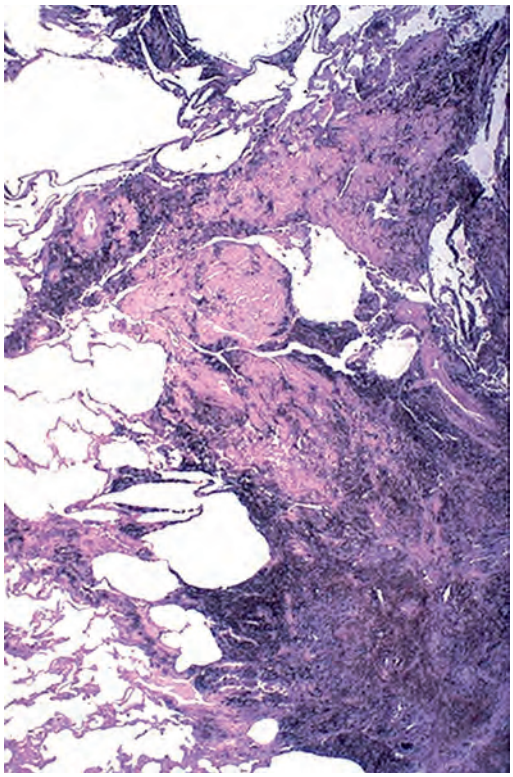


Figure 15.17 Progressive massive fibrosis in a coal worker. A large amount of black pigment is associated with dense interstitial fibrosis. (From Klatt EC: *Robbins and Cotran Atlas of Pathology*, ed 2, Philadelphia, 2010, Saunders, p 121.)

intensely blackened scars 1 cm or larger, sometimes up to 10 cm in greatest diameter. They are usually multiple. Microscopically, the lesions consist of dense collagen and pigment (Fig. 15.17). The center of the lesion is often necrotic, most likely due to local ischemia.

Clinical Features

Coal workers' pneumoconiosis is usually benign, causing little decrement in lung function. Even mild forms of complicated coal workers' pneumoconiosis do not affect lung function significantly. In a minority of cases (fewer than 10%), progressive massive fibrosis develops, leading to increasing pulmonary dysfunction, pulmonary hypertension, and cor pulmonale. Once progressive massive fibrosis develops, it may continue to worsen even if further exposure to dust is prevented. Unlike silicosis (discussed next), there is no convincing evidence that coal workers' pneumoconiosis increases susceptibility to tuberculosis, nor does it predispose to cancer in the absence of smoking. However, domestic indoor use of "smoky coal" (bituminous) for cooking and heating, a common practice in lower income parts of the world, is associated with an increased risk of lung cancer death, even in those who do not smoke.

Silicosis

Silicosis is a common lung disease caused by inhalation of proinflammatory crystalline silicon dioxide (silica). It usually presents after decades of exposure as slowly progressing, nodular, fibrosing pneumoconiosis. Currently, silicosis is the most prevalent chronic occupational disease in the world. Both dose and race are important in developing silicosis (African Americans are at higher risk than Caucasians). As shown in Table 15.6, workers in a large number of occupations are at risk, including individuals involved with the repair, rehabilitation, or demolition of concrete structures such as buildings and roads. The disease also occurs in workers producing stressed denim by sandblasting, stone carvers, and jewelers using chalk molds. Occasionally, heavy exposure over months to a few years can result in acute silicosis, a disorder characterized by the accumulation of abundant lipoproteinaceous material within alveoli (identical morphologically to alveolar proteinosis, discussed later).

Pathogenesis

Phagocytosis of inhaled silica crystals by macrophages activates the inflammasome and stimulates the release of inflammatory mediators, particularly IL-1 and IL-18. This in turn leads to the recruitment of additional inflammatory cells and activates interstitial fibroblasts, leading to collagen deposition. Silica occurs in both crystalline and amorphous forms, but crystalline forms (including quartz, cristobalite, and tridymite) are much more fibrogenic. Of these, quartz is most commonly implicated. The lack of severe responses to silica in coal and hematite miners is thought to be due to coating of silica with other minerals, especially clay components, which render the silica less toxic. Although amorphous silicates are biologically less active than crystalline silica, heavy lung burdens of these minerals may also produce lesions.

MORPHOLOGY

Silicosis is characterized grossly in its early stages by tiny, barely palpable, discrete pale to blackened (if coal dust is also present) nodules in the hilar lymph nodes and upper zones of the lungs. As the disease progresses, these nodules coalesce into **hard, collagenous scars** (Fig. 15.18). Some nodules may undergo central

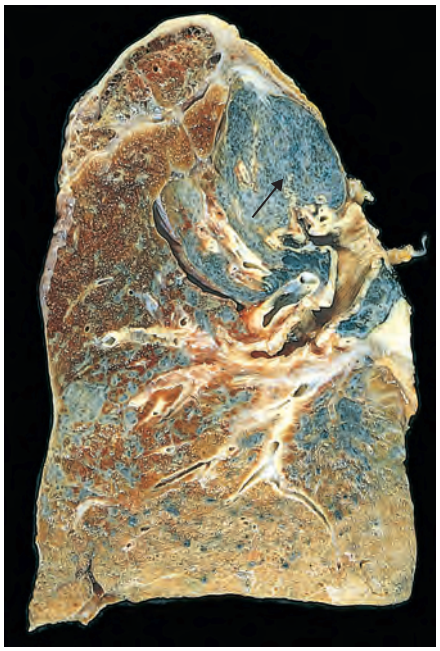


Figure 15.18 Advanced silicosis. Scarring has contracted the upper lobe into a small dark mass *(arrow)*. Note the dense pleural thickening. (Courtesy Dr. John Godleski, Brigham and Women's Hospital, Boston, Mass.)

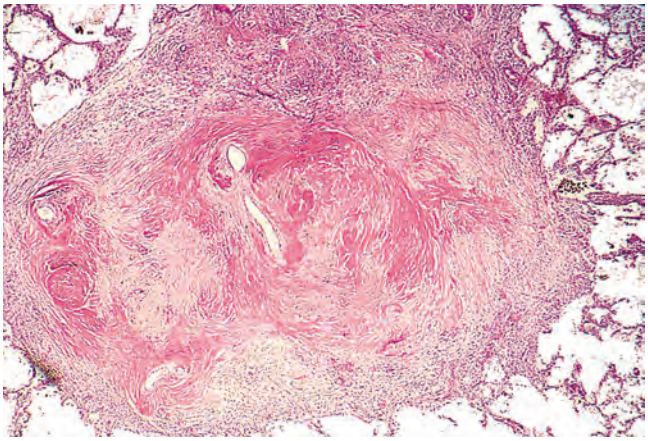


Figure 15.19 Several coalescent collagenous silicotic nodules. (Courtesy Dr. John Godleski, Brigham and Women's Hospital, Boston, Mass.)

softening and cavitation due to superimposed tuberculosis or to ischemia. Fibrotic lesions may also occur in the hilar lymph nodes and pleura. Sometimes, thin sheets of calcification occur in the lymph nodes and are seen radiographically as **eggshell calcification** (i.e., calcium surrounding a zone lacking calcification). If the disease continues to progress, expansion and coalescence of lesions may produce progressive massive fibrosis. Histologic examination reveals the hallmark lesion characterized by a central area of whorled collagen fibers with a more peripheral zone of dust-laden macrophages (Fig. 15.19). Examination of the nodules by polarized microscopy reveals weakly birefringent silicate particles.

Clinical Features

The onset of silicosis may be slow and insidious (10 to 30 years after exposure; this is most common), accelerated (within 10 years of exposure), or rapid (weeks or months after intense exposure to fine dust high in silica; this is rare). Chest radiographs typically show a fine nodularity in the upper zones of the lung. Pulmonary function is either normal or only moderately affected early in the course, and most patients do not develop shortness of breath until progressive massive fibrosis supervenes. The disease may continue to worsen even if the patient is no longer exposed. It is slow to kill, but impaired pulmonary function may severely limit activity.

Silicosis also is associated with an increased susceptibility to *tuberculosis* and a twofold increased risk of lung cancer. The former may be because crystalline silica inhibits the ability of pulmonary macrophages to kill phagocytosed mycobacteria. The link to cancer is not fully understood, but it is but one of many chronic inflammatory conditions that increase the risk of carcinoma in involved tissues (Chapter 7).

Asbestos-Related Diseases

Asbestos is a family of proinflammatory crystalline hydrated silicates that are associated with pulmonary fibrosis and various forms of cancer. Use of asbestos is tightly restricted in many higher income countries; however, there is little, if any, control in lower income parts of the world. Asbestos-related diseases include:

- Localized fibrous plaques or, rarely, diffuse pleural fibrosis
- Pleural effusions, recurrent
- Parenchymal interstitial fibrosis *(asbestosis)*
- Lung carcinoma
- Mesothelioma
- Laryngeal, ovarian, and perhaps other extrapulmonary neoplasms, including colon carcinoma
- Increased risks for systemic autoimmune diseases and cardiovascular disease also have been proposed

The increased incidence of asbestos-related cancers in family members of asbestos workers has alerted the general public to the potential hazards of even low-level exposure to asbestos. However, the necessity of expensive asbestos abatement programs for environments such as schools with low, but measurable, airborne asbestos fiber counts remains a matter of contention.

Pathogenesis

The disease-causing capabilities of the different forms of asbestos depend on concentration, size, shape, and solubility. Asbestos occurs in two distinct geometric forms, serpentine and amphibole. The serpentine chrysotile form accounts for 90% of the asbestos used in industry. Amphiboles, even though less prevalent, are more pathogenic than chrysotiles, particularly with respect to induction of mesothelioma, a malignant tumor derived from the lining cells of pleural surfaces (described later).

The greater pathogenicity of amphiboles is apparently related to their aerodynamic properties and solubility. Chrysotiles, with their more flexible, curled structure, are likely to become impacted in the upper respiratory passages and removed by the mucociliary elevator. Furthermore, once trapped in the lungs, chrysotiles are gradually leached from the tissues because they are more soluble than amphiboles. In contrast, the straight, stiff amphiboles may align themselves with the airstream and thus be delivered deeper into the lungs, where they can penetrate epithelial cells and reach the interstitium. Both amphiboles and serpentines are fibrogenic, and increasing doses are associated with a higher incidence of asbestos-related diseases.

In contrast to other inorganic dusts, asbestos acts as a tumor initiator and a tumor promoter (Chapter 7). Some of its oncogenic effects are mediated by reactive free radicals generated by asbestos fibers, which preferentially localize in the distal lung, close to the mesothelial cells of the pleura. Toxic chemicals adsorbed onto the asbestos fibers also likely contribute to the oncogenicity of the fibers. For example, the adsorption of carcinogens in tobacco smoke onto asbestos fibers may be the basis for the remarkable synergy between tobacco smoking and the development of lung carcinoma in asbestos workers. Smoking also enhances the effect of asbestos by interfering with the mucociliary clearance of fibers. One study of asbestos workers found a fivefold increase of lung carcinoma with asbestos exposure alone, while asbestos exposure and smoking together led to a 55-fold increase in the risk.

Once phagocytosed by macrophages, asbestos fibers activate the inflammasome and stimulate the release of proinflammatory factors and fibrogenic mediators. The initial injury occurs at bifurcations of small airways and ducts, where asbestos fibers land, penetrate, and are directly toxic to pulmonary parenchymal cells. Macrophages, both

alveolar and interstitial, attempt to ingest and clear the fibers. Long-term deposition of fibers and persistent release of mediators (e.g., reactive oxygen species, proteases, cytokines, and growth factors) eventually lead to generalized interstitial pulmonary inflammation and fibrosis.

MORPHOLOGY

Asbestosis is marked by **diffuse pulmonary interstitial fibrosis,** which is distinguished from diffuse interstitial fibrosis resulting from other causes only by the presence of **asbestos bodies.** Asbestos bodies are golden brown, fusiform or beaded rods with a translucent center that consist of asbestos fibers coated with an iron-containing proteinaceous material (Fig. 15.20). They arise when macrophages phagocytose asbestos fibers; the iron is presumably derived from phagocyte ferritin. Other inorganic particulates may become coated with similar iron-protein complexes and are called **ferruginous bodies.**

Asbestosis begins as fibrosis around respiratory bronchioles and alveolar ducts and extends to involve adjacent alveolar sacs and alveoli. The fibrosis distorts the architecture, creating enlarged airspaces enclosed by thick fibrous walls; eventually the affected regions become honeycombed. The pattern of fibrosis is histologically similar to that seen in usual interstitial fibrosis, with fibroblastic foci and varying degrees of fibrosis. In contrast to coal workers' pneumoconiosis and silicosis, asbestosis begins in the lower lobes and subpleurally, with the middle and upper lobes becoming affected as fibrosis progresses. The scarring may trap and narrow pulmonary arteries and arterioles, causing pulmonary hypertension and cor pulmonale.

Pleural plaques, the most common manifestation of asbestos exposure, are well-circumscribed plaques of dense collagen that are often calcified (Fig. 15.21). They develop most frequently on the anterior and posterolateral aspects of the parietal pleura and over the domes of the diaphragm. The size and number of pleural plaques do not correlate with the level of exposure to asbestos or the time since exposure. They also do not contain identifiable asbestos bodies; however, only rarely do they occur in individuals without a history or evidence of asbestos exposure. Uncommonly, asbestos exposure induces pleural effusions, which are usually serous but may be bloody. Rarely, diffuse visceral pleural fibrosis may occur and, in advanced cases, bind the lung to the thoracic wall.

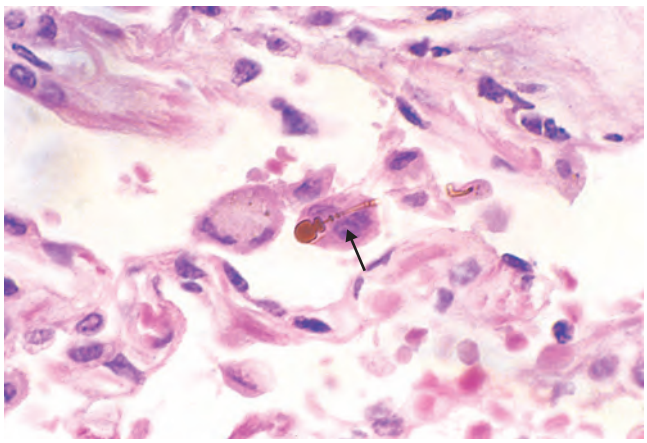


Figure 15.20 High-power detail of an asbestos body, revealing the typical beading and knobbed ends *(arrow)*.

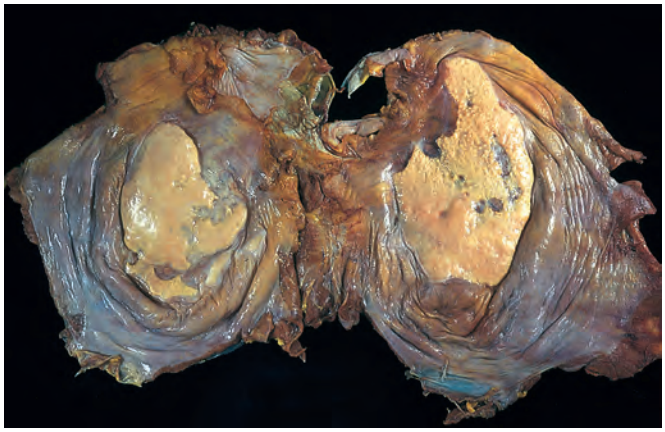


Figure 15.21 Asbestos-related pleural plaques. Large, discrete fibrocalcific plaques are seen on the pleural surface of the diaphragm. (Courtesy Dr. John Godleski, Brigham and Women's Hospital, Boston, Mass.)

Both lung carcinomas and mesotheliomas (pleural and peritoneal) develop in workers exposed to asbestos (see sections on Carcinomas and Pleural Tumors).

Clinical Features

The clinical findings in asbestosis are very similar to those caused by other diffuse interstitial lung diseases (discussed earlier). They rarely appear fewer than 10 years after first exposure and are more common after 20 to 30 years. Dyspnea is usually the first manifestation; at first, it is provoked by exertion, but later is present even at rest. Cough associated with production of sputum, when present, is likely to be due to smoking rather than asbestosis. Chest x-ray studies reveal irregular linear densities, particularly in both lower lobes. With advancement of the pneumoconiosis, a honeycomb pattern develops. The disease may remain static or progress to respiratory failure, cor pulmonale, and death. Pleural plaques are usually asymptomatic and are detected on radiographs as circumscribed densities. Asbestosis complicated by lung or pleural cancer is associated with a particularly grim prognosis.

KEY CONCEPTS

PNEUMOCONIOSES

- Pneumoconioses encompass a group of chronic fibrosing diseases of the lung resulting from exposure to organic and inorganic particulates, most commonly mineral dust.
- Pulmonary alveolar macrophages play a central role in the pathogenesis of lung injury by promoting inflammation and producing fibrogenic cytokines.
- Coal dust–induced disease varies from asymptomatic anthracosis to simple coal workers' pneumoconiosis (coal macules or nodules, and centrilobular emphysema), to progressive massive fibrosis, manifested by increasing pulmonary dysfunction, pulmonary hypertension, and cor pulmonale.
- Silicosis is the most common pneumoconiosis in the world, and crystalline silica (e.g., quartz) is the usual culprit. The lung disease is progressive even after exposure stops.

- The manifestations of silicosis range from asymptomatic silicotic nodules to large areas of dense fibrosis; persons with silicosis also have an increased susceptibility to tuberculosis. There is twofold increased risk of lung cancer.
- Asbestos fibers come in two forms; the stiff amphiboles have a greater fibrogenic and carcinogenic potential than the serpentine chrysotiles.
- Asbestos exposure is linked with six disease processes: (1) parenchymal interstitial fibrosis (asbestosis); (2) localized pleural plaques (asymptomatic) or rarely diffuse pleural fibrosis; (3) recurrent pleural effusions; (4) lung carcinoma; (5) malignant pleural and peritoneal mesotheliomas; and (6) laryngeal cancer.
- Cigarette smoking increases the risk of lung cancer in the setting of asbestos exposure; even family members of workers exposed to asbestos are at increased risk for lung carcinoma and mesothelioma.

Complications of Therapies

Drug-Induced Lung Diseases. An increasing number of prescription drugs have been found to cause a variety of acute and chronic alterations in lung structure and function, interstitial fibrosis, bronchiolitis obliterans, and eosinophilic pneumonia. For example, cytotoxic drugs used in cancer therapy (e.g., bleomycin) cause pulmonary damage and fibrosis as a result of direct toxicity and by stimulating the influx of inflammatory cells into the alveoli. Amiodarone, a drug used to treat cardiac arrhythmias, is preferentially concentrated in the lung and causes significant pneumonitis in 5% to 15% of patients receiving it. Cough induced by angiotensin-converting enzyme inhibitors is very common.

Illicit intravenous drug abuse most often causes lung infections. In addition, particulate matter used to cut drugs may lodge in the lung microvasculature, producing granulomatous inflammation and fibrosis.

Radiation-Induced Lung Diseases. Radiation pneumonitis is a well-known complication of radiotherapy for thoracic tumors (lung, esophageal, breast, mediastinal). It most often involves the lung within the radiation field and occurs in acute and chronic forms. Acute radiation pneumonitis (lymphocytic alveolitis or hypersensitivity pneumonitis) occurs 1 to 6 months after irradiation in 3% to 44% of patients, depending on dose and age. It manifests with fever, dyspnea out of proportion to the volume of lung irradiated, pleural effusion, and pulmonary infiltrates. Morphologic changes are those of diffuse alveolar damage associated with atypia of hyperplastic type II pneumocytes and fibroblasts. Epithelial cell atypia and foam cells within vessel walls are also characteristic of radiation damage. With steroid therapy, these symptoms may resolve completely, but other cases progress to chronic radiation pneumonitis (pulmonary fibrosis), which also may occur without antecedent, clinically apparent, acute radiation pneumonitis. In its most severe form, chronic radiation pneumonitis and associated progressive fibrosis can lead to cyanosis, pulmonary hypertension, and cor pulmonale.

Granulomatous Diseases

Sarcoidosis

Sarcoidosis is a systemic granulomatous disease of unknown cause that may involve many tissues and organs. Its various clinical presentations are protean, but the most common are bilateral hilar lymphadenopathy or parenchymal lung involvement, occurring in 90% of cases. Eye and skin lesions are next in frequency. Since other diseases, including mycobacterial and fungal infections and berylliosis, can also produce noncaseating granulomas, the diagnosis is one of exclusion.

Sarcoidosis usually occurs in adults younger than 40 years of age but can affect any age group. The prevalence is higher in women but varies widely in different countries and populations. In the United States the rates are highest in the Southeast and are 10 times higher in African-Americans than in Caucasians. In contrast, the disease is rare among the Chinese and Southeast Asians. Patterns of organ involvement also vary with race.

Pathogenesis

Although several lines of evidence suggest that sarcoidosis is a disease of disordered immune regulation in genetically predisposed individuals, its etiology is unknown. There are several immunologic abnormalities in the local milieu of sarcoid granulomas that suggest a cell-mediated immune response to an unidentified antigen. These abnormalities include:

- Intra-alveolar and interstitial accumulation of CD4+ T cells, resulting in CD4/CD8 T-cell ratios ranging from 5:1 to 15:1, suggesting pathogenic involvement of CD4+ helper T cells. There is oligoclonal expansion of T-cell subsets as determined by analysis of T-cell receptor rearrangement, consistent with an antigen-driven proliferation.
- Increased levels of T cell–derived Th1 cytokines such as IL-2 and interferon (IFN)-γ, which may be responsible for T-cell expansion and macrophage activation, respectively.
- Increased levels of several cytokines in the local environment (IL-8, TNF, macrophage inflammatory protein 1α) that favor recruitment of additional T cells and monocytes and contribute to the formation of granulomas. TNF in particular is released at high levels by activated alveolar macrophages, and the TNF concentration in the bronchoalveolar fluid is a marker of disease activity.
- Impaired dendritic cell function.

Additionally, there are systemic immunologic abnormalities in individuals with sarcoidosis. Both anergy to common skin test antigens, such as *Candida* or tuberculosis purified protein derivative (PPD), and polyclonal hypergammaglobulinemia, another manifestation of helper T-cell dysregulation, are frequently observed. Evidence of genetic influences includes familial and racial clustering of cases and the association with certain human leukocyte antigen (HLA) genotypes (e.g., HLA-A1 and HLA-B8).

MORPHOLOGY

Virtually every organ in the body has been described as being affected by sarcoidosis, at least on rare occasions. Involved tissues contain well-formed **non-necrotizing granulomas** (Fig. 15.22) composed of aggregates of tightly clustered epithelioid macrophages, often with giant cells. Central necrosis is unusual. With chronicity the granulomas may become enclosed within fibrous rims or may eventually be replaced by hyaline fibrous scars.

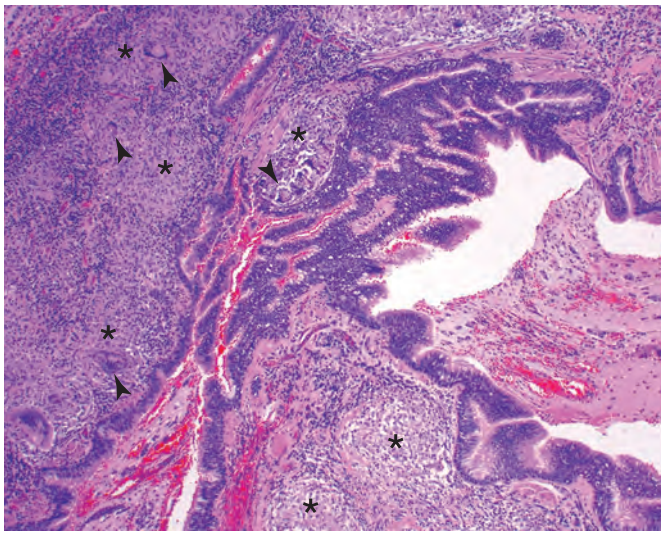


Figure 15.22 Bronchus with characteristic noncaseating sarcoidal granulomas *(asterisks)*, with many multinucleated giant cells *(arrowheads)*. Note subepithelial location of granulomas.

Laminated concretions composed of calcium and proteins known as **Schaumann bodies** and stellate inclusions known as **asteroid bodies** are found within giant cells in approximately 60% of the granulomas. Though characteristic, these microscopic features are not pathognomonic of sarcoidosis because asteroid and Schaumann bodies may be encountered in other granulomatous diseases (e.g., tuberculosis).

The **lung** is a common site of involvement. Macroscopically, there is usually no demonstrable alteration, although in advanced cases coalescence of granulomas produces small nodules that are palpable or visible as 1 to 2 cm, noncaseating, noncavitated consolidations. The lesions are distributed primarily along the lymphatics around bronchi and blood vessels, although alveolar lesions and pleural involvement are also seen. The relatively high frequency of granulomas in the bronchial submucosa accounts for the high diagnostic yield of bronchoscopic biopsies. There seems to be a strong tendency for lesions to heal in the lungs, so varying stages of fibrosis and hyalinization are often found.

Lymph nodes are involved in almost all cases, particularly the hilar and mediastinal nodes, but any node in the body may be affected. Nodes are characteristically enlarged, discrete, and sometimes calcified. Tonsillar granulomas are seen in about one-fourth to one-third of cases. The **spleen** is involved in about 75% of cases, but overt splenomegaly is seen in only 20% of cases. On occasion, granulomas may coalesce to form small nodules that are visible macroscopically. The **liver** is affected slightly less often than the spleen. It may be moderately enlarged and typically contains scattered granulomas, more in portal triads than in the lobular parenchyma.

The **bone marrow** is involved in about 20% of cases. Radiologically, visible bone lesions have a particular tendency to involve phalangeal bones of the hands and feet, creating small circumscribed areas of bone resorption within the marrow cavity and a diffuse reticulated pattern throughout the cavity, with widening of the bony shafts or new bone formation on the outer surfaces.

Skin lesions, encountered in 25% of cases, assume a variety of appearances, including discrete subcutaneous nodules; focal, slightly elevated, erythematous plaques; or flat lesions that are slightly reddened and scaling, resembling those of systemic lupus erythematosus. Lesions may also appear on the mucous membranes of the oral cavity, larynx, and upper respiratory tract. Other patients present with **erythema nodosum**, painful erythematous nodules on the shins that stem from septal panniculitis.

Ocular involvement, seen in 25% of cases, takes the form of iritis or iridocyclitis and may be bilateral or unilateral. Consequently, corneal opacities, glaucoma, and total loss of vision may occur. These ocular lesions are frequently accompanied by inflammation of the lacrimal glands and suppression of lacrimation (**sicca syndrome**). Bilateral sarcoidosis of the parotid, submaxillary, and sublingual glands constitutes the combined uveoparotid involvement designated as Mikulicz syndrome (Chapter 16).

Muscle involvement is underdiagnosed, since it may be asymptomatic. Muscle weakness, aches, tenderness, and fatigue should prompt consideration of occult sarcoid myositis, which can be diagnosed by muscle biopsy. Sarcoid granulomas occasionally occur in the heart, kidneys, central nervous system (neurosarcoidosis, seen in 5% to 15% of cases), and endocrine glands, particularly in the pituitary, as well as in other body tissues.

Clinical Features

Because of its varying severity and inconstant tissue distribution, sarcoidosis may present with diverse features. It may be discovered unexpectedly on routine chest films as bilateral hilar adenopathy or may present with peripheral lymphadenopathy, cutaneous lesions, eye involvement, splenomegaly, or hepatomegaly. In the great majority of cases, however, individuals seek medical attention because of the insidious onset of respiratory abnormalities (shortness of breath, cough, chest pain, hemoptysis) or of constitutional signs and symptoms (fever, fatigue, weight loss, anorexia, night sweats).

Sarcoidosis follows an unpredictable course. It may be inexorably progressive or marked by periods of activity interspersed with remissions, sometimes permanent, that may be spontaneous or induced by steroid therapy. Overall, 65% to 70% of affected patients recover with minimal or no residual manifestations. Twenty percent have permanent loss of some lung function or some permanent visual impairment. Of the remaining 10% to 15%, some die of cardiac or central nervous system damage, but most succumb to progressive pulmonary fibrosis and cor pulmonale.

KEY CONCEPTS

SARCOIDOSIS

- Sarcoidosis is a multisystem disease of unknown etiology; the diagnostic histopathologic feature is the presence of noncaseating granulomas in various tissues.
- Immunologic abnormalities include high levels of CD4+ T cells in the lung that secrete Th1-dependent cytokines such as IFN-γ and IL-2 locally.
- Clinical manifestations include lymph node enlargement, eye involvement (sicca syndrome [dry eyes], iritis, or iridocyclitis), skin lesions (e.g., erythema nodosum), and visceral (liver, skin, marrow) involvement. Lung involvement occurs in 90% of cases, with formation of granulomas and interstitial fibrosis.

Hypersensitivity Pneumonitis

The term *hypersensitivity pneumonitis* describes a spectrum of immunologically mediated, predominantly interstitial lung disorders caused by intense, often prolonged exposure to inhaled organic antigens. Affected individuals have an abnormal sensitivity or heightened reactivity to the causative antigen, which, in contrast to asthma, leads to pathologic changes that primarily involve the alveolar walls (thus the synonym *extrinsic allergic alveolitis*). It is important to recognize these diseases early in their course because progression to serious chronic fibrotic lung disease can be prevented by removal of the environmental agent.

Most commonly, hypersensitivity results from the inhalation of organic dust containing antigens made up of the spores of thermophilic bacteria, fungi, animal proteins, or bacterial products. Numerous syndromes are described, depending on the occupation or exposure of the individual. *Farmer's lung* results from exposure to dusts generated from humid, warm, newly harvested hay that permits the rapid proliferation of the spores of thermophilic actinomycetes. *Pigeon breeder's lung* (bird fancier's disease) is provoked by proteins from serum, excreta, or feathers of birds. *Humidifier* or *air-conditioner lung* is caused by thermophilic bacteria in heated water reservoirs. Pet birds and moldy basements are easily missed unless asked about specifically.

Several lines of evidence suggest that hypersensitivity pneumonitis is an immunologically mediated disease:

- Bronchoalveolar lavage specimens from the acute phase show increased levels of proinflammatory chemokines such as macrophage inflammatory protein 1α and IL-8.
- Bronchoalveolar lavage specimens also consistently demonstrate increased numbers of both CD4+ and CD8+ T lymphocytes.
- Most patients have specific antibodies against the causative antigen in their serum.
- Complement and immunoglobulins have been demonstrated within vessel walls by immunofluorescence.
- The presence of non-necrotizing granulomas in two-thirds of the patients suggests that T cell–mediated (type IV) hypersensitivity reactions against the implicated antigens have a pathogenic role.

MORPHOLOGY

Histologic changes depend on the phase of the disease; acute alveolar damage is seen in the first hours and few days after antigen exposure, while subacute changes are characteristically centered on bronchioles. They include interstitial pneumonitis, consisting primarily of lymphocytes, plasma cells, and macrophages (eosinophils are rare), as well as non-necrotizing granulomas (Fig. 15.23). Interstitial fibrosis with fibroblastic foci, honeycombing, and obliterative bronchiolitis together with granulomas is seen in the chronic phase.

Clinical Features

The clinical manifestations are varied. Acute attacks, which follow inhalation of antigenic dust in sensitized patients, consist of episodes of fever, dyspnea, cough, and leukocytosis. Micronodular interstitial infiltrates may appear in the chest radiograph, and pulmonary function tests show an acute restrictive disorder. Symptoms usually appear 4 to 6 hours after exposure and may last for 12 hours to several days. They recur with re-exposure. If exposure is continuous and protracted, a chronic form of the disease supervenes, leading to progressive fibrosis, dyspnea, and cyanosis—a picture similar to that seen in other forms of chronic interstitial disease.

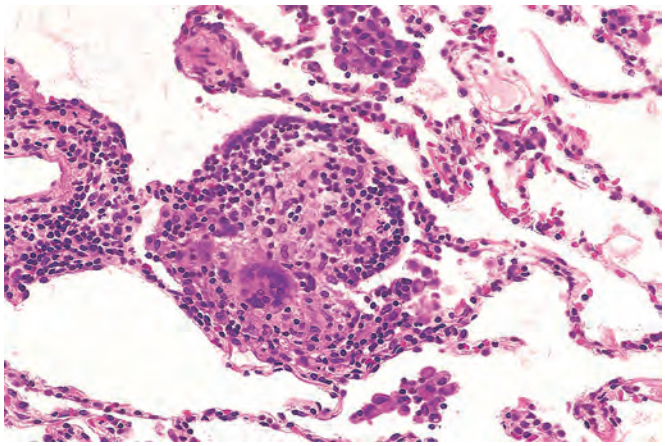


Figure 15.23 Hypersensitivity pneumonitis. Loosely formed interstitial granulomas and chronic inflammation are characteristic.

Pulmonary Eosinophilia

Although relatively rare, there are several clinical and pathologic pulmonary entities that are characterized by an infiltration of eosinophils, recruited in part by elevated alveolar levels of eosinophil attractants such as IL-5. Pulmonary eosinophilia is divided into the following categories:

- *Acute eosinophilic pneumonia with respiratory failure.* This is an acute illness of unknown cause. It has a rapid onset with fever, dyspnea, and hypoxemic respiratory failure. The chest radiograph shows diffuse infiltrates, and bronchoalveolar lavage fluid contains more than 25% eosinophils. Histology shows diffuse alveolar damage and many eosinophils. There is a prompt response to corticosteroids.
- *Secondary eosinophilia,* which occurs in a number of parasitic, fungal, and bacterial infections; in hypersensitivity pneumonitis; in drug allergies; and in association with asthma, allergic bronchopulmonary aspergillosis, or *Churg-Strauss syndrome,* a form of vasculitis.
- *Idiopathic chronic eosinophilic pneumonia,* characterized by focal areas of cellular consolidation of the lung substance distributed chiefly in the periphery of the lung fields. Prominent in these lesions are aggregates of lymphocytes and eosinophils within the septal walls and the alveolar spaces. Interstitial fibrosis and organizing pneumonia are often present. These patients have cough, fever, night sweats, dyspnea, and weight loss, all of which respond to corticosteroid therapy. Chronic eosinophilic pneumonia is diagnosed when other causes of pulmonary eosinophilia are excluded.

Smoking-Related Interstitial Diseases

Smoking-related diseases can be grouped into obstructive diseases (emphysema and chronic bronchitis, already

discussed) and restrictive or interstitial diseases. A majority of individuals with idiopathic pulmonary fibrosis are smokers; however, the role of cigarette smoking in its pathogenesis has not been clarified yet. Desquamative interstitial pneumonia and respiratory bronchiolitis–associated interstitial lung disease are two other smoking-associated interstitial lung diseases worthy of brief mention.

Desquamative Interstitial Pneumonia

Desquamative interstitial pneumonia is characterized by large collections of macrophages in the airspaces of a current or former smoker. The macrophages were originally thought to be desquamated pneumocytes—thus the misnomer "desquamative interstitial pneumonia."

MORPHOLOGY

The most striking finding is the accumulation of a large number of macrophages with abundant cytoplasm containing dusty brown pigment **(smokers' macrophages)** in the airspaces (Fig. 15.24). Some of the macrophages contain lamellar bodies (composed of surfactant) within phagocytic vacuoles, presumably derived from necrotic type II pneumocytes. The alveolar septa are thickened by a sparse inflammatory infiltrate of lymphocytes, plasma cells, and occasional eosinophils. The septa are lined by plump, cuboidal pneumocytes. Interstitial fibrosis, when present, is mild. Emphysema is often present.

Desquamative interstitial pneumonia usually presents in the fourth or fifth decade of life and is now equally common in men and women. Virtually all patients are cigarette smokers. Presenting symptoms include an insidious onset of dyspnea and dry cough over weeks or months, often associated with clubbing of digits. Pulmonary function tests usually show a mild restrictive abnormality and moderately decreased diffusing capacity. Patients with desquamative interstitial pneumonia typically have an excellent response to steroid therapy and cessation of smoking, but occasionally patients progress to interstitial fibrosis.

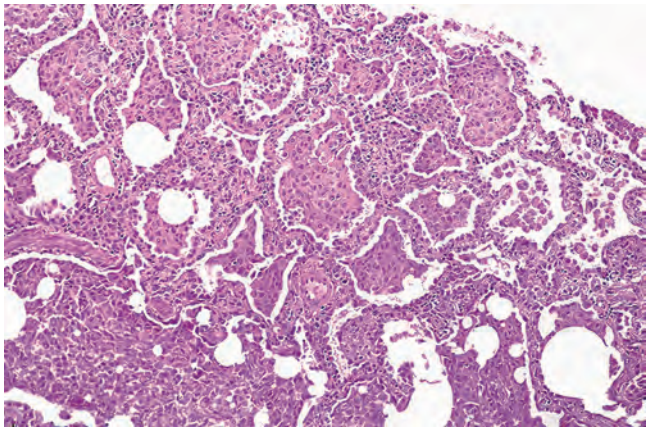


Figure 15.24 Desquamative interstitial pneumonia. Medium-power detail of lung demonstrates the accumulation of large numbers of macrophages within the alveolar spaces and only mild fibrous thickening of the alveolar walls.

Respiratory Bronchiolitis–Associated Interstitial Lung Disease

Respiratory bronchiolitis–associated interstitial lung disease is marked by chronic inflammation and peribronchiolar fibrosis. It is a common histologic lesion in cigarette smokers. It is characterized by the presence of pigmented intraluminal macrophages within first- and second-order respiratory bronchioles. In its mildest form, it is most often an incidental finding in the lungs of smokers or ex-smokers. The term *respiratory bronchiolitis–associated interstitial lung disease* is used for patients who develop significant pulmonary symptoms, abnormal pulmonary function, and imaging abnormalities.

MORPHOLOGY

The changes are patchy at low magnification and have a bronchiolocentric distribution. Respiratory bronchioles, alveolar ducts, and peribronchiolar spaces contain aggregates of dusty brown macrophages **(smokers' macrophages)** similar to those seen in desquamative interstitial pneumonia. There is a patchy submucosal and peribronchiolar infiltrate of lymphocytes and histiocytes. Mild peribronchiolar fibrosis is also seen, which expands contiguous alveolar septa. Centrilobular emphysema is common but not severe. Desquamative interstitial pneumonia is often found in different parts of the same lung.

Symptoms are usually mild, consisting of gradual onset of dyspnea and cough in patients who are typically current cigarette smokers with exposures of over 30 pack-years in the fourth or fifth decade of life. Cessation of smoking usually results in improvement.

Pulmonary Langerhans Cell Histiocytosis

Pulmonary Langerhans cell histiocytosis is a rare disease characterized by focal collections of Langerhans cells (often accompanied by eosinophils). As these lesions progress, scarring occurs, leading to airway destruction and alveolar damage that result in the appearance of irregular cystic spaces. Imaging of the chest shows characteristic cystic and nodular abnormalities. Langerhans cells are immature dendritic cells with grooved, indented nuclei and abundant cytoplasm. They are positive for S100, CD1a, and CD207 (langerin) and are negative for CD68.

More than 90% of affected patients are relatively young adult smokers or ex-smokers; among smokers, about half improve after smoking cessation, suggesting that in some cases the lesions are a reactive inflammatory process. However, in other cases the Langerhans cells have activating mutations in the serine/threonine kinase BRAF, a feature consistent with a neoplastic process that is also commonly seen in Langerhans cell histiocytosis involving other tissues (Chapter 13). A neoplastic basis may explain why the disease progresses in some patients, sometimes even necessitating lung transplantation.

Pulmonary Alveolar Proteinosis

Pulmonary alveolar proteinosis (PAP) is a rare disease caused by defects in pulmonary macrophage function due

to deficient granulocyte-macrophage colony-stimulating factor (GM-CSF) signaling, which results in the accumulation of surfactant in the intra-alveolar and bronchiolar spaces. PAP is characterized radiologically by bilateral patchy asymmetric pulmonary opacifications. There are three distinct classes of disease—autoimmune (formerly called acquired), secondary, and congenital—each with a similar spectrum of histologic changes.

- *Autoimmune PAP* is caused by autoantibodies that bind and neutralize the function of GM-CSF. It occurs primarily in adults, represents 90% of all cases of PAP, and lacks any familial predisposition. Knockout of the GM-CSF gene in mice induces PAP, and these mice are "cured" by treatment with GM-CSF. Loss of GM-CSF signaling blocks the terminal differentiation of alveolar macrophages, impairing their ability to catabolize surfactant.
- *Secondary PAP* is uncommon and is associated with diverse diseases, including hematopoietic disorders, malignancies, immunodeficiency disorders, lysinuric protein intolerance (an inborn error of amino acid metabolism), and acute silicosis and other inhalational syndromes. It is speculated that these diseases somehow impair GM-CSF–dependent signaling or downstream events involved in macrophage maturation or function, again leading to inadequate clearance of surfactant from alveolar spaces.
- *Hereditary PAP* is extremely rare, occurs in neonates, and is caused by loss-of-function mutations in the genes that encode GM-CSF or the GM-CSF receptor.

MORPHOLOGY

The disease is characterized by the accumulation of intra-alveolar precipitates containing surfactant proteins, causing focal-to-confluent consolidation of large areas of the lungs with minimal inflammatory reaction (Fig. 15.25). As a consequence there is a marked increase in the size and weight of the lung. The alveolar precipitate is pink, homogeneous, and periodic acid–Schiff–positive and contains cholesterol clefts and surfactant proteins (which can be demonstrated by immunohistochemical stains). Ultrastructurally, the surfactant lamellae in type II pneumocytes are normal, in contrast to surfactant dysfunction disorders (described next).

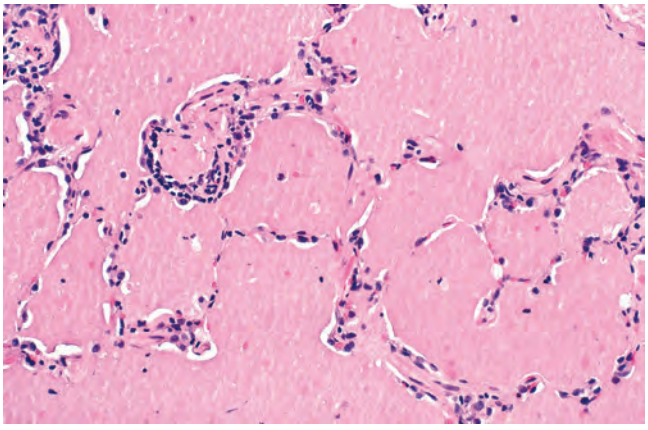


Figure 15.25 Pulmonary alveolar proteinosis. The alveoli are filled with a dense, amorphous, protein-lipid granular precipitate, while the alveolar walls are normal.

Clinical Features

Adult patients, for the most part, present with cough and abundant sputum that often contains chunks of gelatinous material. Some have symptoms lasting for years, often with intermittent febrile illnesses caused by secondary pulmonary infections with a variety of organisms. Progressive dyspnea, cyanosis, and respiratory insufficiency may occur, but other patients follow a benign course, with eventual resolution of the lesions. Whole-lung lavage is the standard of care and provides benefit regardless of the underlying defect. GM-CSF therapy is safe and effective in more than half of patients with autoimmune PAP, and therapy directed at the underlying disorder may also be helpful. Primary disease is treated with GM-CSF replacement therapy, sometimes followed by allogeneic hematopoietic stem cell transplantation, which can be curative.

Surfactant Dysfunction Disorders

Surfactant dysfunction disorders are diseases caused by diverse mutations in genes encoding proteins involved in surfactant trafficking or secretion. Clinical manifestations range from neonatal respiratory failure to adult-onset interstitial lung disease. The most commonly mutated genes are the following:

- *ATP-binding cassette protein member 3 (ABCA3)* is the most frequently mutated gene in surfactant dysfunction disorders. Mutations in *ABCA3* are associated with an autosomal recessive disorder that usually presents in the first few months of life with rapidly progressive respiratory failure followed by death. Less commonly, it comes to attention in older children and in adults with chronic interstitial lung disease.
- The second most commonly mutated gene in surfactant dysfunction disorders encodes *surfactant protein C*. This form has an autosomal dominant mode of inheritance and has a highly variable course.
- The third most commonly mutated gene in surfactant dysfunction disorders encodes *surfactant protein B*. This form has an autosomal recessive mode of inheritance. Typically, the infant is full term and develops progressive respiratory distress shortly after birth. Death ensues between 3 and 6 months of age unless lung transplantation is performed.

MORPHOLOGY

There is a variable amount of intra-alveolar pink granular material, type II pneumocyte hyperplasia, interstitial fibrosis, and alveolar simplification. Immunohistochemical stains show the lack of surfactant proteins C and B in their respective deficiencies. Ultrastructurally, abnormalities in lamellar bodies in type II pneumocytes can be seen in all three; small lamellar bodies with electron dense cores are diagnostic for *ABCA3* mutation (Fig. 15.26).

DISEASES OF VASCULAR ORIGIN

Pulmonary Embolism and Infarction

Pulmonary embolism is an important cause of morbidity and mortality, particularly in patients who are bedridden,

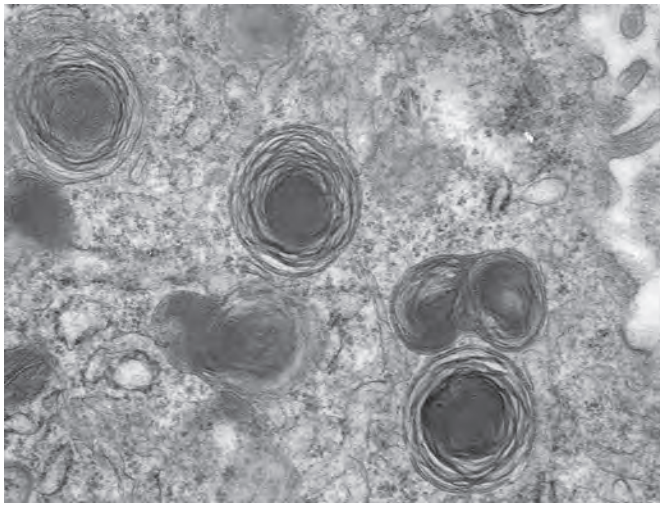


Figure 15.26 Pulmonary alveolar proteinosis associated with mutation of the *ABCA3* gene. An electron micrograph shows type 2 pneumocytes containing small surfactant lamellae with electron dense cores, an appearance that is characteristic of cases associated with *ABCA3* mutations.

but also in a wide range of conditions that are associated with hypercoagulability. Blood clots that occlude the large pulmonary arteries are almost always embolic in origin. The usual source—thrombi in the deep veins of the leg (>95% of cases)—and the magnitude of the clinical problem were discussed in Chapter 4. Pulmonary embolism causes more than 50,000 deaths in the United States each year. Its incidence at autopsy has varied from 1% in the general population of hospital patients to 30% in patients dying after severe burns, trauma, or fractures. It is the sole or major contributing cause of death in about 10% of adults who die acutely in hospitals. By contrast, large-vessel pulmonary thromboses are rare and develop only in the presence of pulmonary hypertension and heart failure.

Pathogenesis

Pulmonary embolism usually occurs in patients with a predisposing condition that produces an increased tendency to clot (thrombophilia). Patients often have cardiac disease or cancer or have been immobilized for several days or weeks prior to the appearance of a symptomatic embolism. Patients with hip fractures are at particularly high risk. Hypercoagulable states, either primary (e.g., factor V Leiden, prothrombin mutations, and antiphospholipid syndrome) or secondary (e.g., obesity, recent surgery, cancer, oral contraceptive use, pregnancy), are important risk factors. Indwelling central venous lines can be a nidus for formation of right atrial thrombi, which can embolize to the lungs. Rarely, pulmonary embolism may consist of fat, air, or tumor. Small bone marrow emboli are often seen in patients who die after chest compressions performed during resuscitative efforts.

The pathophysiologic response and clinical significance of pulmonary embolism depend on the extent to which pulmonary artery blood flow is obstructed, the size of the occluded vessels, the number of emboli, and the cardiovascular health of the patient. Emboli have two deleterious pathophysiologic consequences: *respiratory compromise* due to the nonperfused, although ventilated, segment; and *hemodynamic compromise* due to increased resistance to pulmonary blood flow caused by the embolic obstruction. Sudden death often ensues, largely as a result of the blockage of blood flow through the lungs. Death may also be caused by acute right-sided heart failure *(acute cor pulmonale)*.

MORPHOLOGY

Large emboli lodge in the main pulmonary artery or its major branches or at the bifurcation as a saddle embolus (Fig. 15.27). Smaller emboli travel out into the more peripheral vessels, where they may cause hemorrhage or infarction. In patients with adequate cardiovascular function, the bronchial arterial supply sustains the lung parenchyma; in this instance, hemorrhage may occur, but there is no infarction. In those in whom cardiovascular function is already compromised, such as patients with heart or lung disease, infarction is more likely. Overall, about 10% of emboli cause infarction. About 75% of infarcts affect the lower lobes, and in more than half, multiple lesions occur. They vary in size from barely visible to massive lesions involving large parts of a lobe. Typically, they extend to the periphery of the lung as a wedge with the apex pointing toward the hilus of the lung. In many cases, an occluded vessel is identified near the apex of the infarct. Pulmonary embolus can be distinguished from a postmortem clot by the presence of the lines of Zahn in the thrombus (Chapter 4).

The pulmonary infarct is classically hemorrhagic and appears as a raised, red-blue area in the early stages (Fig. 15.28). Often, the apposed pleural surface is covered by a fibrinous exudate. The red cells begin to lyse within 48 hours, and the infarct becomes paler and eventually red-brown as hemosiderin is produced. With the passage of time, fibrous replacement begins at the margins as a gray-white peripheral zone and eventually converts the infarct into a contracted scar. Histologically, the hemorrhagic area shows ischemic necrosis of the alveolar walls, bronchioles, and vessels. If the infarct is caused by an infected embolus, the neutrophilic inflammatory reaction can be intense. Such lesions are referred to as **septic infarcts,** some of which turn into abscesses.

Clinical Features

A large pulmonary embolus is one of the few causes of virtually instantaneous death. During cardiopulmonary resuscitation in such instances, the patient frequently is said

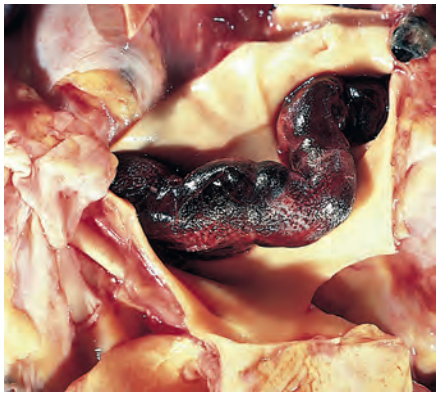


Figure 15.27 Large saddle embolus from the femoral vein lying astride the main left and right pulmonary arteries. (From the teaching collection of the Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

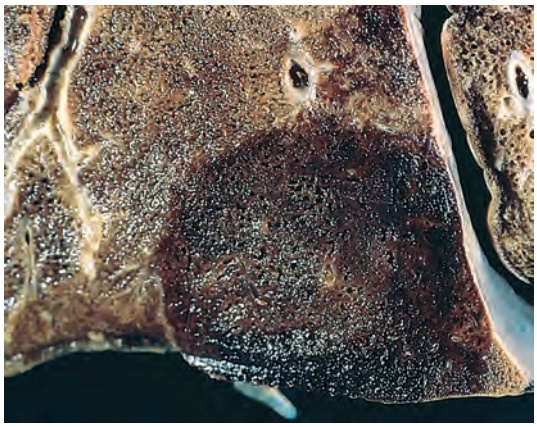


Figure 15.28 Acute hemorrhagic pulmonary infarct.

to have electromechanical dissociation, in which the electrocardiogram has a rhythm, but no pulses are palpated because no blood is entering the pulmonary arterial circulation. If the patient survives after a sizable pulmonary embolus, however, the clinical syndrome may mimic myocardial infarction, with severe chest pain, dyspnea, and shock. Small emboli are silent or induce only transient chest pain and cough. In the remaining group of patients with symptomatic pulmonary embolism, the most common presenting symptoms (in descending order) are dyspnea, pleuritic pain, and cough, accompanied in about half of cases by calf or thigh swelling or pain. Emboli that lead to pulmonary infarction may additionally produce fever and hemoptysis. An overlying fibrinous pleuritis may produce a pleural friction rub.

In hemodynamically stable patients with a low to moderate risk of pulmonary embolism, D-dimer measurement is a useful screening test, as a normal D-dimer level excludes pulmonary embolism. Definitive diagnosis is usually made by computed tomographic pulmonary angiogram, which identifies obstructed pulmonary arteries. Rarely, other diagnostic methods, such as ventilation-perfusion scanning, are required. Deep vein thrombosis can be diagnosed with duplex ultrasonography. Chest radiography may be normal or disclose a pulmonary infarct, usually 12 to 36 hours after it has occurred, as a wedge-shaped infiltrate.

After the initial acute insult, emboli often resolve via contraction and fibrinolysis, particularly in relatively young patients. If unresolved, with time multiple small emboli may lead to pulmonary hypertension and chronic cor pulmonale. Perhaps most important is that a small embolus may presage a larger one. In the presence of an underlying predisposing condition, patients with a pulmonary embolus have a 30% chance of suffering a second embolus.

Prevention of pulmonary embolism is a major clinical challenge for which there is no easy solution. Prophylactic therapy includes early ambulation in postoperative and postpartum patients, elastic stockings and graduated compression stockings for bedridden patients, and anticoagulation in high-risk individuals. Treatment of pulmonary embolism includes anticoagulation and supportive measures; thrombolysis may have some benefit in patients with severe complications (e.g., shock), but carries a high risk of bleeding. Those at risk of recurrent pulmonary embolism in whom anticoagulation is contraindicated may be fitted with an inferior vena cava filter (an "umbrella") that catches clots before they reach the lungs.

KEY CONCEPTS

PULMONARY EMBOLISM

- Almost all large pulmonary artery thrombi are embolic in origin, usually arising from the deep veins of the lower leg.
- Risk factors include prolonged bed rest, leg surgery, severe trauma, congestive heart failure, use of oral contraceptives (especially those with high estrogen content), disseminated cancer, and inherited forms of hypercoagulability.
- The vast majority (60% to 80%) of emboli are clinically silent, a minority (5%) cause acute cor pulmonale, shock, or death (typically from large "saddle emboli"), and the remainder cause symptoms related to ventilation-perfusion mismatch and/or pulmonary infarction, particularly dyspnea and pleuritic chest pain.
- Risk of recurrence is high, and recurrent embolism may eventually lead to pulmonary hypertension and cor pulmonale.

Pulmonary Hypertension

Pulmonary hypertension is defined as a mean pulmonary artery pressure greater than or equal to 25 mm Hg at rest. Based on underlying mechanisms, the WHO has classified pulmonary hypertension into five groups: (1) pulmonary arterial hypertension, a diverse collection of disorders that all primarily impact small pulmonary muscular arteries; (2) pulmonary hypertension due to left heart failure; (3) pulmonary hypertension due to lung diseases and/or hypoxia; (4) chronic thromboembolic pulmonary hypertension and other obstructions; and (5) pulmonary hypertension with unclear and/or multifactorial mechanisms.

Pathogenesis

As can be gathered from the above classification, pulmonary hypertension has diverse causes even within each group. It is most frequently associated with structural cardiopulmonary conditions that increase either pulmonary blood flow, pulmonary vascular resistance, or left heart resistance to blood flow. Some of the more common causes are the following:

- *Chronic obstructive or interstitial lung diseases* (group 3). These diseases obliterate alveolar capillaries, increasing pulmonary resistance to blood flow and, secondarily, pulmonary blood pressure.
- *Antecedent congenital or acquired heart disease* (group 2). Mitral stenosis, for example, causes an increase in left atrial pressure and pulmonary venous pressure that is eventually transmitted to the arterial side of the pulmonary vasculature, leading to hypertension.
- *Recurrent thromboemboli* (group 4). Recurrent pulmonary emboli may cause pulmonary hypertension by reducing the functional cross-sectional area of the pulmonary vascular bed, which in turn leads to an increase in pulmonary vascular resistance.
- *Autoimmune diseases* (group 1). Several of these diseases (most notably systemic sclerosis) involve the pulmonary vasculature and/or the interstitium, leading to increased vascular resistance and pulmonary hypertension.

- *Obstructive sleep apnea* (group 3) is a common disorder that is associated with obesity and hypoxemia. It is now recognized to be a significant contributor to the development of pulmonary hypertension and cor pulmonale.

Uncommonly, pulmonary hypertension is encountered in patients in whom all known causes are excluded; this is referred to as *idiopathic pulmonary arterial hypertension* and also falls into group 1 disease. However, "idiopathic" is a bit of a misnomer, as up to 80% of "idiopathic" cases (sometimes referred to as primary pulmonary hypertension) have a genetic basis, sometimes being inherited in families as an autosomal dominant trait. Within these families, there is incomplete penetrance, and only 10% to 20% of the family members actually develop overt disease.

As is often the case, much has been learned about the pathogenesis of pulmonary hypertension by investigating the molecular basis of the uncommon familial form of the disease. The first mutation to be discovered in familial pulmonary arterial hypertension was in the gene encoding bone morphogenetic protein receptor type 2 *(BMPR2)*. Inactivating germline mutations in the *BMPR2* gene are found in 75% of the familial cases of pulmonary hypertension and 25% of sporadic cases. Subsequently other mutations have been discovered that also converge on the BMPR2 pathway and affect intracellular signaling. It has also been demonstrated that *BMPR2* is downregulated in lungs from some patients with idiopathic pulmonary arterial hypertension without *BMPR2* mutations.

BMPR2 is a cell surface protein belonging to the TGF-β receptor superfamily, which binds a variety of cytokines, including TGF-β, bone morphogenetic protein (BMP), activin, and inhibin. Originally described as a pathway that regulates bone growth, BMP-BMPR2 signaling is now known to be important for embryogenesis, apoptosis, and cell proliferation and differentiation. Details remain to be worked out, but it appears that haploinsufficiency for BMPR2 leads to dysfunction and proliferation of endothelial cells and vascular smooth muscle cells. Because only 10% to 20% of individuals with *BMPR2* mutations develop disease, it is likely that modifier genes and/or environmental triggers also contribute to the pathogenesis of the disorder. A two-hit model has been proposed whereby a genetically susceptible individual with a *BMPR2* mutation requires additional genetic or environmental insults to develop the disease (Fig. 15.29).

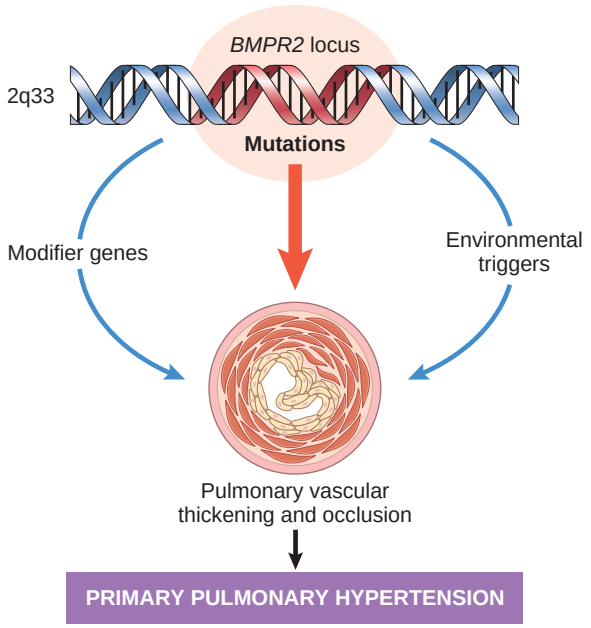


Figure 15.29 Pathogenesis of primary (idiopathic) pulmonary hypertension. See text for details.

MORPHOLOGY

All forms of pulmonary hypertension are associated with **medial hypertrophy of the pulmonary muscular and elastic arteries** and **right ventricular hypertrophy.** The presence of many organizing or recanalized thrombi favors recurrent pulmonary emboli as the cause, and the coexistence of diffuse pulmonary fibrosis, or severe emphysema and chronic bronchitis, points to chronic hypoxia and loss of capillary beds as initiating events. The vessel changes can involve the entire arterial tree, from the main pulmonary arteries down to the arterioles (Fig. 15.30). In the most severe cases, the thickening of the walls of the pulmonary artery and its major branches take on some features of systemic atherosclerosis, but classic atherosclerotic changes are not seen. Instead, the arterioles and small arteries (40 to 300 μm in diameter) are prominently affected by striking medial hypertrophy and intimal fibrosis, sometimes narrowing the lumens to pinpoint channels. One extreme in the spectrum of pathologic changes is the plexiform lesion, so called because a tuft of capillary formations is present, producing a network, or web, that spans the lumens of dilated thin-walled, small arteries and may extend outside the vessel. Plexiform lesions are most prominent in idiopathic and familial pulmonary hypertension (group 1), unrepaired congenital heart disease with left-to-right shunts (group 2), and pulmonary hypertension associated with human immunodeficiency virus (HIV) infection and drugs (also group 1).

Clinical Features

Idiopathic (often inherited) forms of pulmonary hypertension are most common in women who are 20 to 40 years of age, but may occasionally present in childhood. Clinical signs and symptoms in all types become evident only in advanced disease. In cases of idiopathic disease, the presenting features are usually dyspnea and fatigue, but some patients have chest pain of the anginal type. Over time, severe respiratory distress, cyanosis, and right ventricular hypertrophy occur, and death from decompensated cor pulmonale, often with superimposed thromboembolism and pneumonia, ensues within 2 to 5 years in 80% of patients.

Treatment choices depend on the underlying cause. For patients with secondary disease, therapy is directed at the trigger (e.g., thromboembolic disease or hypoxemia). A variety of vasodilators have been used with varying success in those with group 1 or refractory disease belonging to other groups. Lung transplantation provides definitive treatment for selected patients.

Diffuse Pulmonary Hemorrhage Syndromes

Pulmonary hemorrhage is a dramatic complication of some interstitial lung disorders. *Pulmonary hemorrhage syndromes*

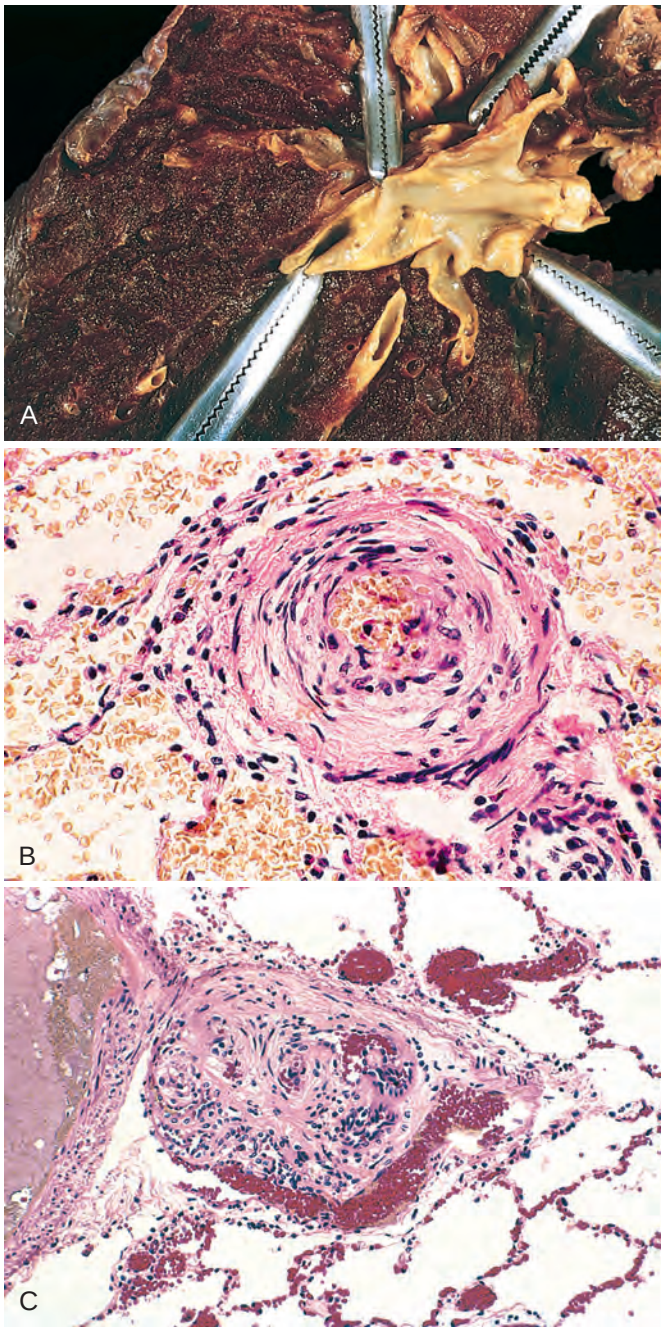


Figure 15.30 Vascular changes in pulmonary arterial hypertension. (A) Atheroma-like changes, a finding usually limited to large vessels. (B) Marked medial hypertrophy. (C) Plexiform lesion of small arteries that is characteristic of advanced pulmonary hypertension.

(Fig. 15.31) include (1) Goodpasture syndrome, (2) idiopathic pulmonary hemosiderosis, and (3) vasculitis-associated hemorrhage. The latter encompasses conditions such as hypersensitivity angiitis, polyangiitis with granulomatosis, and systemic lupus erythematosus (Chapter 11).

Idiopathic Pulmonary Hemosiderosis

Idiopathic pulmonary hemosiderosis is a rare disorder characterized by intermittent, diffuse alveolar hemorrhage. Most cases occur in young children, although the disease has been reported in adults as well. It usually presents with an insidious onset of productive cough, hemoptysis, and anemia associated with diffuse pulmonary infiltrates.

The cause and pathogenesis are unknown, and anti–basement membrane antibodies (the cause of Goodpasture syndrome) are undetectable. However, favorable response to long-term immunosuppression with prednisone and/or azathioprine indicates that an immunologic mechanism may be involved in the pulmonary capillary damage underlying alveolar bleeding. In addition, long-term follow-up shows that some affected patients develop other immune disorders.

Goodpasture Syndrome (Anti–Glomerular Basement Membrane Antibody Disease With Pulmonary Involvement)

Goodpasture syndrome is an uncommon autoimmune disease in which kidney and lung injury are caused by circulating autoantibodies against the noncollagenous domain of the α3 chain of collagen IV. When only renal disease is caused by this antibody, it is called anti–glomerular basement membrane disease. The term *Goodpasture syndrome* applies to the 40% to 60% of patients who develop pulmonary hemorrhage in addition to renal disease. Although any age can be affected, most cases occur in the teens or 20s, and in contrast to many other autoimmune diseases, there is a male preponderance. The majority of patients are active smokers.

Pathogenesis

The immunopathogenesis of the syndrome and the nature of the Goodpasture antigens are described in Chapter 20. The pathogenic antibodies initiate inflammatory destruction of the basement membrane in renal glomeruli and pulmonary alveoli, giving rise to rapidly progressive glomerulonephritis and a necrotizing hemorrhagic interstitial pneumonitis. The trigger that initiates the production of anti–basement

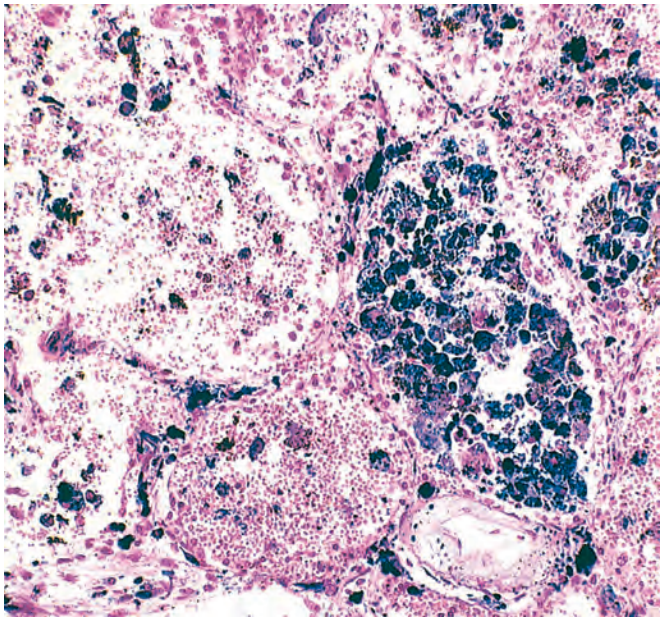


Figure 15.31 Diffuse pulmonary hemorrhage syndrome. There is acute intra-alveolar hemorrhage and hemosiderin-laden macrophages, reflecting previous hemorrhage (Prussian blue iron stain).

membrane antibodies is unknown. In addition to autoreactive B cells, some experimental evidence suggests that T cells also contribute, both by enhancing B-cell function and by participating directly in glomerular damage and crescent formation. As with other autoimmune disorders, there is an association with certain HLA subtypes (e.g., HLA-DRB1*1501 and HLA-DRB1*1502).

MORPHOLOGY

In the classic case, the lungs are heavy, with areas of red-brown consolidation. Histologically, there is focal necrosis of alveolar walls associated with intra-alveolar hemorrhages. Often the alveoli contain hemosiderin-laden macrophages (see Fig. 15.31). In later stages there may be fibrous thickening of the septa, hyperplasia of type II pneumocytes, and organization of blood in alveolar spaces. In many cases, immunofluorescence studies reveal linear deposits of immunoglobulins along the basement membranes of the septal walls. The kidneys have the characteristic findings of focal proliferative glomerulonephritis in early cases or crescentic glomerulonephritis in patients with rapidly progressive glomerulonephritis. Diagnostic linear deposits of immunoglobulins and complement are seen by immunofluorescence studies along the glomerular basement membranes even in the few patients without renal disease.

Clinical Features

Most cases begin with respiratory symptoms, principally hemoptysis, and radiographic evidence of focal pulmonary consolidations. Soon, manifestations of glomerulonephritis appear, leading to rapidly progressive renal failure. The most common cause of death is uremia. The once dismal prognosis for this disease has been markedly improved by intensive plasmapheresis. This procedure is thought to be beneficial by removing anti–basement membrane antibodies and possibly other mediators of immunologic injury. Simultaneous immunosuppressive therapy inhibits further antibody production, ameliorating both lung hemorrhage and glomerulonephritis.

Polyangiitis With Granulomatosis

Previously called Wegener granulomatosis, this autoimmune disease most often involves the upper respiratory tract and/or the lungs, with hemoptysis being the common presenting symptom. Its features are discussed in Chapter 11. Here, it suffices to emphasize that a transbronchial lung biopsy might provide the only tissue available for diagnosis. Since the amount of tissue is small, necrosis and granulomatous vasculitis might not be present. Rather, the diagnostically important histologic features are capillaritis and scattered, poorly formed granulomas (unlike those of sarcoidosis, which are rounded and well-defined).

PULMONARY INFECTIONS

Respiratory tract infections are more frequent than infections of any other organ and account for the largest number of workdays lost in the general population. The vast majority consist of upper respiratory tract infections caused by viruses (common cold, pharyngitis), but bacterial, viral, mycoplasmal, and fungal infections of the lung (pneumonia) account for an enormous amount of morbidity and are responsible for 2.3% of all deaths in the United States. Pneumonia can be very broadly defined as any infection of the lung parenchyma.

Pulmonary antimicrobial defense mechanisms are described in Chapter 8. Pneumonia can result whenever these local defense mechanisms are impaired or the systemic resistance of the host is lowered. Factors that impair resistance include chronic diseases, immunologic deficiencies, treatment with immunosuppressive agents, and leukopenia. Local pulmonary defense mechanisms may also be compromised by many factors, including:

- *Loss or suppression of the cough reflex,* as a result of altered sensorium (e.g., coma), anesthesia, neuromuscular disorders, drugs, or chest pain, any of which may lead to *aspiration* of gastric contents.
- *Dysfunction of the mucociliary apparatus,* which can be caused by cigarette smoke, inhalation of hot or corrosive gases, viral diseases, or genetic defects of ciliary function (e.g., immotile cilia syndrome).
- *Accumulation of secretions* in conditions such as cystic fibrosis and bronchial obstruction.
- *Interference with the phagocytic and bactericidal activities of alveolar macrophages* by alcohol, tobacco smoke, anoxia, or oxygen intoxication.
- *Pulmonary congestion and edema.*

Defects in innate immunity (including neutrophil and complement defects) and humoral immunodeficiency typically lead to an increased incidence of infections with pyogenic bacteria. Germline mutations in MyD88 (an adaptor for several Toll-like receptors [TLRs] that is important for activation of the transcription factor nuclear factor kappa B [NF-κB]) are also associated with destructive bacterial (pneumococcal) pneumonias. On the other hand, cell-mediated immune defects (congenital and acquired) lead to increased infections with intracellular microbes such as mycobacteria and herpesviruses as well as with microorganisms of very low virulence, such as the fungus *Pneumocystis jiroveci*.

Several other points should be emphasized. First, to paraphrase the French physician Louis Cruveilhier in 1919 (during the Spanish flu epidemic), "flu condemns, and additional infection executes." The most common cause of death in viral influenza epidemics is superimposed bacterial pneumonia. Second, although the portal of entry for most bacterial pneumonias is the respiratory tract, hematogenous seeding of the lungs from another organ may occur and may be difficult to distinguish from primary pneumonia. Finally, many patients with chronic diseases acquire terminal pneumonia while hospitalized *(nosocomial infection)* because of several factors: bacteria common to the hospital environment may have acquired resistance to antibiotics; opportunities for spread are increased; invasive procedures, such as intubations and injections, are common; and bacteria may contaminate equipment used in respiratory care units.

Pneumonia is classified based on the etiologic agent or, if no pathogen can be isolated (which occurs in about 50% of cases), by the clinical setting in which the infection occurs.

Table 15.7 Pneumonia Syndromes

Community-Acquired Acute Pneumonia
Streptococcus pneumoniae
Haemophilus influenzae
Moraxella catarrhalis
Staphylococcus aureus
Legionella pneumophila
Enterobacteriaceae *(Klebsiella pneumoniae)* and *Pseudomonas* spp.
Mycoplasma pneumoniae
Chlamydia spp. *(C. pneumoniae, C. psittaci, C. trachomatis)*
Coxiella burnetii (Q fever)
Viruses: respiratory syncytial virus, parainfluenza virus, and human metapneumovirus (children); influenza A and B (adults); adenovirus (military recruits)
Health Care–Associated Pneumonia
Staphylococcus aureus, methicillin-sensitive
Staphylococcus aureus, methicillin-resistant
Pseudomonas aeruginosa
Streptococcus pneumoniae
Hospital-Acquired Pneumonia
Gram-negative rods, Enterobacteriaceae (*Klebsiella* spp., *Serratia marcescens*, *Escherichia coli*) and *Pseudomonas* spp.
Staphylococcus aureus (usually methicillin-resistant)
Aspiration Pneumonia
Anaerobic oral flora *(Bacteroides, Prevotella, Fusobacterium, Peptostreptococcus)*, admixed with aerobic bacteria *(Streptococcus pneumoniae, Staphylococcus aureus, Haemophilus influenzae, Pseudomonas aeruginosa)*
Chronic Pneumonia
Nocardia
Actinomyces
Granulomatous: *Mycobacterium tuberculosis* and atypical mycobacteria, *Histoplasma capsulatum, Coccidioides immitis, Blastomyces dermatitidis*
Necrotizing Pneumonia and Lung Abscess
Anaerobic bacteria (extremely common), with or without mixed aerobic infection
Staphylococcus aureus, *Klebsiella pneumoniae*, *Streptococcus pyogenes*, and type 3 pneumococcus (uncommon)
Pneumonia in the Immunocompromised Host
Cytomegalovirus
Pneumocystis jiroveci
Mycobacterium avium-intracellulare complex
Invasive aspergillosis
Invasive candidiasis
"Usual" bacterial, viral, and fungal organisms (listed herein)

The latter considerably narrows the list of suspected pathogens, providing a guide for empirical antimicrobial therapy. As Table 15.7 indicates, pneumonia can arise in seven distinct clinical settings ("pneumonia syndromes"), and the implicated pathogens are fairly specific to each category.

Community-Acquired Bacterial Pneumonias

Community-acquired acute pneumonia refers to lung infection in otherwise healthy individuals that is acquired from the normal environment (in contrast to hospital-acquired pneumonia). It may be bacterial or viral. Clinical and radiologic features are usually insensitive in differentiating between viral and bacterial infections. One marker of inflammation, procalcitonin, an acute-phase reactant produced primarily in the liver, is more significantly elevated in bacterial than viral infections and has some predictive value, but is not specific, as it is also markedly elevated in other severe inflammatory disorders, such as systemic inflammatory response syndrome (SIRS) (Chapter 4).

Often, a bacterial infection follows an upper respiratory tract viral infection. Bacterial invasion of the lung parenchyma causes the alveoli to be filled with an inflammatory exudate, thus causing consolidation ("solidification") of the pulmonary tissue. Many variables, such as the specific etiologic agent, the host reaction, and the extent of involvement, determine the precise form of pneumonia. Predisposing conditions include extremes of age, chronic diseases (congestive heart failure, COPD, and diabetes), congenital or acquired immune deficiencies, and decreased or absent splenic function. The latter puts the patient at risk for infection with encapsulated bacteria such as pneumococcus.

Streptococcus pneumoniae

***Streptococcus pneumoniae*, or *pneumococcus*, is the most common cause of community-acquired acute pneumonia.** Examination of Gram-stained sputum is an important step in the diagnosis of acute pneumonia. The presence of numerous neutrophils containing the typical gram-positive, lancet-shaped diplococci supports the diagnosis of pneumococcal pneumonia, but it must be remembered that *S. pneumoniae* is a part of the endogenous flora in 20% of adults, and therefore false-positive results may be obtained. Isolation of pneumococci from blood cultures is more specific but less sensitive (in the early phase of illness, only 20% to 30% of patients have positive blood cultures). Pneumococcal vaccines containing capsular polysaccharides from the common serotypes are used in individuals at high risk for pneumococcal sepsis.

Haemophilus influenzae

Haemophilus influenzae is a pleomorphic, gram-negative organism that occurs in encapsulated and nonencapsulated forms. There are six serotypes of the encapsulated form (types a to f), of which type b is the most virulent. Antibodies against the capsule protect the host from *H. influenzae* infection; hence the capsular polysaccharide b is incorporated in the widely used vaccine against *H. influenzae*. With routine use of *H. influenzae* vaccines, the incidence of disease caused by the b serotype has declined significantly. By contrast, infections with nonencapsulated forms, also called *nontypeable forms*, are increasing. These are less virulent and tend to spread along the surface of the upper respiratory tract, producing otitis media (infection of the middle ear), sinusitis, and bronchopneumonia. Neonates and children with comorbidities such as prematurity, malignancy, and immunodeficiency are at high risk for development of invasive infection.

H. influenzae pneumonia, which may follow a viral respiratory infection, is a pediatric emergency and has a high mortality rate. Descending laryngotracheobronchitis results in airway obstruction as the smaller bronchi are plugged by dense, fibrin-rich exudates containing neutrophils, similar to that seen in pneumococcal pneumonias. Pulmonary consolidation is usually lobular and patchy but may be confluent and involve the entire lung lobe. Before a vaccine became widely available, *H. influenzae* was a common cause of suppurative meningitis in children up to

5 years of age. *H. influenzae* also causes an acute, purulent conjunctivitis (pink eye) in children and, in predisposed older patients, may cause septicemia, endocarditis, pyelonephritis, cholecystitis, and suppurative arthritis. Finally, *H. influenzae* is the most common bacterial cause of acute exacerbations of COPD.

Moraxella catarrhalis

Moraxella catarrhalis is recognized as a cause of bacterial pneumonia, especially in the elderly. It is the second most common bacterial cause of acute exacerbation of COPD. Along with *S. pneumoniae* and *H. influenzae, M. catarrhalis* is one of the three most common causes of otitis media in children.

Staphylococcus aureus

Staphylococcus aureus is an important cause of secondary bacterial pneumonia in children and healthy adults following viral respiratory illnesses (e.g., measles in children and influenza in both children and adults). Staphylococcal pneumonia is associated with a high incidence of complications, such as lung abscess and empyema. *Intravenous drug users* are at high risk for development of staphylococcal pneumonia in association with endocarditis. It is also an important cause of hospital-acquired pneumonia.

Klebsiella pneumoniae

Klebsiella pneumoniae is the most frequent cause of gram-negative bacterial pneumonia. It commonly afflicts debilitated and malnourished people, particularly *chronic alcoholics*. Thick, mucoid (often blood-tinged) sputum is characteristic because the organism produces an abundant viscid capsular polysaccharide, which the patient may have difficulty expectorating.

Pseudomonas aeruginosa

Although *Pseudomonas aeruginosa* most commonly causes hospital-acquired infections, it is mentioned here because of its occurrence in cystic fibrosis and immunocompromised patients. It is common in patients who are neutropenic, and it has a propensity to invade blood vessels with consequent extrapulmonary spread. *Pseudomonas* septicemia is a very fulminant disease.

Legionella pneumophila

Legionella pneumophila is the agent of legionnaires' disease, the form of pneumonia caused by this organism. It also causes Pontiac fever, a related self-limited upper respiratory tract infection. This organism flourishes in artificial aquatic environments, such as water-cooling towers and the tubing systems of domestic (potable) water supplies. It is transmitted by either inhalation of aerosolized organisms or aspiration of contaminated drinking water. *Legionella* pneumonia is common in individuals with predisposing conditions such as cardiac, renal, immunologic, or hematologic disease. Organ transplant recipients are particularly susceptible. It can be quite severe, frequently requiring hospitalization, and immunosuppressed patients have fatality rates of up to 50%. The diagnosis can be made rapidly by detecting *Legionella* DNA in sputum using a polymerase chain reaction (PCR)–based test or by identification of *Legionella* antigens in the urine; culture remains the diagnostic gold standard, but takes 3 to 5 days.

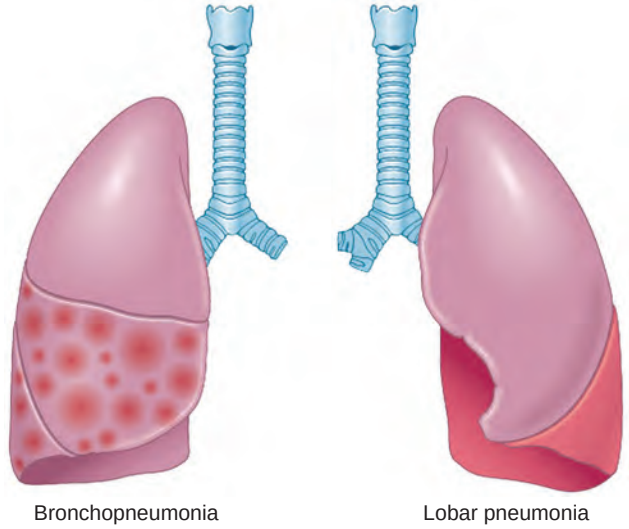


Figure 15.32 Comparison of bronchopneumonia and lobar pneumonia.

Mycoplasma pneumoniae

Mycoplasma infections are particularly common among children and young adults. They occur sporadically or as local epidemics in closed communities (schools, military camps, and prisons).

MORPHOLOGY

Bacterial pneumonia has two patterns of anatomic distribution: lobular bronchopneumonia and lobar pneumonia (Fig. 15.32). Patchy consolidation of the lung is the dominant characteristic of **bronchopneumonia** (Fig. 15.33), while consolidation of a large portion of a lobe or of an entire lobe defines **lobar**

Figure 15.33 Bronchopneumonia. Section of lung showing patches of consolidation *(arrows)*.

Figure 15.34 Lobar pneumonia—gray hepatization. The lower lobe is uniformly consolidated.

pneumonia (Fig. 15.34). These anatomic categorizations may be difficult to apply in individual cases because patterns overlap. The patchy involvement may become confluent, producing lobar consolidation. Moreover, the same organisms may produce either pattern depending on patient susceptibility. **Most important from the clinical standpoint are identification of the causative agent and determination of the extent of disease.**

In lobar pneumonia, four stages of the inflammatory response have classically been described: congestion, red hepatization, gray hepatization, and resolution. In the first stage of **congestion,** the lung is heavy, boggy, and red. It is characterized by vascular engorgement, intra-alveolar edema fluid containing a few neutrophils, and the presence of bacteria, which may be numerous. In the next stage of **red hepatization**, there is massive confluent exudation, as neutrophils, red cells, and fibrin fill the alveolar spaces (Fig. 15.35A). On gross examination, the lobe is red, firm, and airless, with a liver-like consistency, hence the name hepatization. The third stage of **gray hepatization** is marked by progressive disintegration of red cells and the persistence of a fibrinosuppurative exudate (Fig. 15.35B), resulting in a color change to grayish-brown. In the final stage of **resolution,** the exudate within the alveolar spaces is broken down by enzymatic digestion to produce granular, semifluid debris that is resorbed, ingested by macrophages, expectorated, or organized by fibroblasts growing into it (Fig. 15.35C). Pleural fibrinous reaction to the underlying inflammation, often present in the early stages if the consolidation extends to the lung surface **(pleuritis)**, may similarly resolve. More often it undergoes organization, leaving fibrous thickening or permanent adhesions.

Foci of **bronchopneumonia** are consolidated areas of acute suppurative inflammation. The consolidation may be confined to one lobe but is more often multilobar and frequently bilateral and basal because of the tendency of secretions to gravitate to the lower lobes. Well-developed lesions are slightly elevated, dry, granular, gray-red to yellow, and poorly delimited at their margins (see Fig. 15.33). Histologically, the reaction usually elicits a neutrophil-rich exudate that fills the bronchi, bronchioles, and adjacent alveolar spaces (see Fig. 15.35A).

Complications of pneumonia include (1) tissue destruction and necrosis, causing **abscess formation** (particularly common with pneumococcal or *Klebsiella* infections); (2) spread of infection

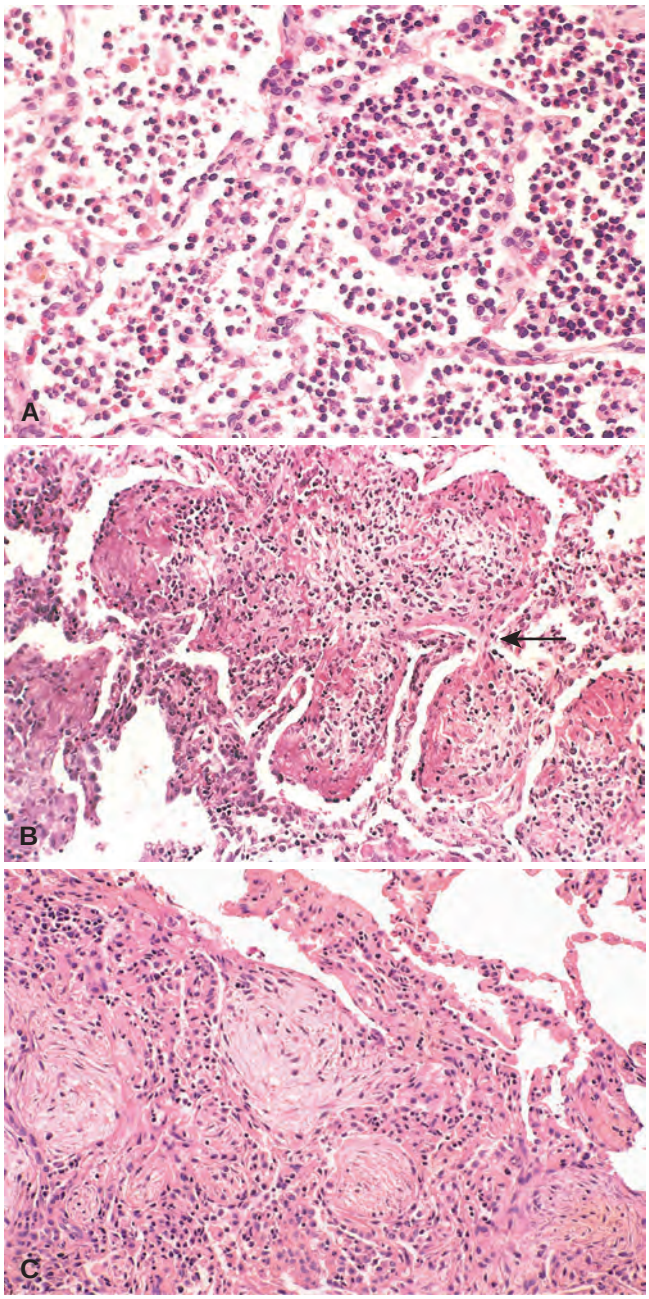


Figure 15.35 Stages of bacterial pneumonia. (A) Acute pneumonia. The congested septal capillaries and numerous intra-alveolar neutrophils are characteristic of early red hepatization. Fibrin nets have not yet formed. (B) Early organization of intra-alveolar exudate, seen focally to be streaming through the pores of Kohn *(arrow)*. (C) Advanced organizing pneumonia. The exudates have been converted to fibromyxoid masses rich in macrophages and fibroblasts.

to the pleural cavity, causing an intrapleural fibrinosuppurative reaction known as **empyema;** and (3) **bacteremic dissemination** to the heart valves, pericardium, brain, kidneys, spleen, or joints, causing abscesses, endocarditis, meningitis, or suppurative arthritis.

Clinical Features

The major symptoms of community-acquired acute bacterial pneumonia are abrupt onset of high fever, shaking chills, and cough producing mucopurulent sputum and occasionally hemoptysis. When pleuritis is present it is accompanied by pleuritic pain and pleural friction rub. The whole lobe is radiopaque in lobar pneumonia, whereas there are focal opacities in bronchopneumonia.

The clinical picture is markedly modified by the administration of effective antibiotics. Appropriately treated patients may become afebrile with few clinical signs 48 to 72 hours after the initiation of antibiotics. The identification of the organism and the determination of its antibiotic sensitivity are the keystones of therapy. Fewer than 10% of patients with pneumonia severe enough to merit hospitalization now succumb, and in most instances death results from a complication, such as empyema, meningitis, endocarditis, or pericarditis, or is attributable to some predisposing influence, such as debility or chronic alcoholism.

Community-Acquired Viral Pneumonia

Common viral infections include influenza virus types A and B, respiratory syncytial viruses, human metapneumovirus, adenovirus, rhinoviruses, rubeola, and varicella viruses. Any of these agents can cause a relatively mild upper respiratory tract infection, recognized as the common cold, or a more severe lower respiratory tract infection. Factors that favor extension of the infection to the lung include extremes of age, malnutrition, alcoholism, and underlying debilitating illnesses.

Although the molecular details vary, all of the viruses that cause pneumonia produce disease through similar general mechanisms. These viruses have tropisms that allow them to attach to and enter respiratory lining cells. Viral replication and gene expression leads to cytopathic changes, inducing cell death and secondary inflammation. The resulting damage and impairment of local pulmonary defenses, such as mucociliary clearance, may predispose to bacterial superinfections, which are often more serious than the viral infection itself.

Influenza

Influenza viruses of type A infect humans, pigs, horses, and birds and are the major cause of pandemic and epidemic influenza infections. The influenza genome encodes several proteins, but the most important from the vantage point of viral virulence are the hemagglutinin and neuraminidase proteins. Hemagglutinin has three major subtypes (H1, H2, H3), while neuraminidase has two (N1, N2). Both proteins are embedded in a lipid bilayer, which constitutes the influenza virus envelope. Hemagglutinin is particularly important, as it serves to attach the virus to its cellular target via sialic acid residues on surface polysaccharides. Following uptake of the virus into endosomal vesicles, acidification of the endosome triggers a conformation change in hemagglutinin that allows the viral envelope to fuse with the host cell membrane, releasing the viral genomic RNAs into the cytoplasm of the cell. Neuraminidase in turn facilitates the release of newly formed virions that are budding from infected cells by cleaving sialic acid residues. Neutralizing host antibodies against viral hemagglutinin and neuraminidase prevent and ameliorate, respectively, infection with influenza virus.

The viral genome is composed of eight single-stranded RNAs, each encoding one or more proteins. The RNAs are packaged into helices by nucleoproteins that determine the influenza virus type (A, B, or C). A single subtype of influenza virus A predominates throughout the world at a given time. Epidemics of influenza are caused by spontaneous mutations that alter antigenic epitopes on the viral hemagglutinin and neuraminidase proteins. These antigenic changes *(antigenic drift)* result in new viral strains that are sufficiently different to elude, at least in part, anti-influenza antibodies produced in members of the population in response to prior exposures to other flu strains. Usually, however, these new strains bear sufficient resemblance to prior strains that some members of the population are at least partially resistant to infection. By contrast, pandemics, which are longer and more widespread than epidemics, occur when both the hemagglutinin and the neuraminidase genes are replaced through recombination with animal influenza viruses *(antigenic shift)*. In this instance, essentially all individuals are susceptible to the new influenza virus.

If the host lacks protective antibodies, the virus infects pneumocytes and elicits several cytopathic changes. Shortly after entry into pneumocytes, the viral infection inhibits sodium channels, producing electrolyte and water shifts that lead to fluid accumulation in the alveolar lumen. This is followed by the death of the infected cells through several mechanisms, including inhibition of host cell messenger RNA translation and activation of caspases leading to apoptosis. The death of epithelial cells exacerbates the fluid accumulation and releases "danger signals" that activate resident macrophages. In addition, prior to their death, infected epithelial cells release a variety of inflammatory mediators, including several chemokines and cytokines, adding fuel to the inflammatory fire. In addition, mediators released from epithelial cells and macrophages activate the nearby pulmonary endothelium and serve as chemoattractants for neutrophils, which migrate into the interstitium within the first day or two of infection. In some cases viral infection may cause sufficient lung injury to produce ARDS, but more often severe pulmonary disease stems from a superimposed bacterial pneumonia. Of these, secondary pneumonias caused by *S. aureus* are particularly common and often life-threatening.

Control of the infection relies on several host mechanisms. The presence of viral products induces innate immune responses in infected cells, such as the production of α- and β-interferon. These mediators upregulate the expression of the *MX1* gene, which encodes a guanosine triphosphatase that interferes with viral gene transcription and viral replication. As with other viral infections, natural killer cells and cytotoxic T cells can recognize and kill infected host cells, limiting viral replication and viral spread to adjacent pneumocytes. The cellular immune response is eventually

augmented by development of antibody responses to the viral hemagglutinin and neuraminidase proteins.

Insight into future pandemics has come from studying past pandemics. DNA analysis of viral genomes retrieved from the lungs of a soldier who died in the great 1918 influenza pandemic that killed between 20 million and 40 million people worldwide identified swine influenza sequences, consistent with this virus having its origin in a "antigenic shift." The first flu pandemic of this century, in 2009, was also caused by an antigenic shift involving a virus of swine origin. It caused particularly severe infections in young adults, apparently because older adults had antibodies against past influenza strains that conveyed at least partial protection. Comorbidities such as diabetes, heart disease, lung disease, and immunosuppression were also associated with a higher risk of severe infection.

What then might be the source of the next great pandemic? There is no certainty, but one concern is centered on avian influenza, which normally infects birds. One such strain, type H5N1, has spread throughout the world in wild and domestic birds. Fortunately, the transmission of the current H5N1 avian virus is inefficient. However, if H5N1 influenza recombines with an influenza that is highly infectious for humans, a strain might result that is capable of sustained human-to-human transmission (and thus of causing the next great pandemic).

Human Metapneumovirus

Human metapneumovirus, a paramyxovirus discovered in 2001, is found worldwide and is associated with upper and lower respiratory tract infections. Infections can occur in any age group but are most common in young children, elderly adults, and immunocompromised patients. Some infections, such as bronchiolitis and pneumonia, are severe; overall, metapneumovirus is responsible for 5% to 10% of hospitalizations and 12% to 20% of outpatient visits of children suffering from acute respiratory tract infections. Such infections are clinically indistinguishable from those caused by human respiratory syncytial virus and are often mistaken for influenza. The first human metapneumovirus infection occurs during early childhood, but reinfections are common throughout life, especially in older subjects. Diagnostic methods include PCR tests for viral RNA. Treatment generally focuses on supportive measures. Although work is ongoing, a clinically effective and safe vaccine has yet to be developed.

Human Coronaviruses

Coronaviruses are enveloped, positive-sense RNA viruses that infect humans and several other vertebrate species. Weakly pathogenic coronaviruses cause mild cold-like upper respiratory tract infections, while highly pathogenic ones may cause severe, often fatal pneumonia. An example of a highly pathogenic type is SARS-CoV-2, a strain that emerged in late 2019 in China that is producing a still evolving pandemic as of early 2020 (discussed in Chapter 8). Highly pathogenic coronaviruses like SARS-CoV-2 bind the ACE2 protein on the surface of pulmonary alveolar epithelial cells, explaining the tropism of these viruses for the lung. With highly pathogenic forms in susceptible hosts, typically older individuals with comorbid conditions, the host immune response and locally released cytokines often produce acute lung injury and ARDS.

MORPHOLOGY

All viral infections produce similar morphologic changes. Upper respiratory infections are marked by mucosal hyperemia and swelling, infiltration of the submucosa by mononuclear cells (mainly lymphocytes and monocytes), and overproduction of mucus secretions. The swollen mucosa and viscous exudate may plug the nasal channels, sinuses, or the Eustachian tubes, leading to suppurative secondary bacterial infection. Virus-induced tonsillitis causing hyperplasia of the lymphoid tissue within the Waldeyer ring is frequent in children.

In viral **laryngotracheobronchitis** and **bronchiolitis** there is vocal cord swelling and abundant mucus production. Impairment of bronchociliary function invites bacterial superinfection with more marked suppuration. Plugging of small airways may give rise to focal lung atelectasis. With more severe bronchiolar involvement, widespread plugging of secondary and terminal airways by cell debris, fibrin, and inflammatory exudate may, if prolonged, lead to organization and fibrosis, resulting in obliterative bronchiolitis and permanent lung damage.

Lung involvement may be quite patchy or may involve whole lobes bilaterally or unilaterally. The affected areas are red-blue and congested. Pleuritis or pleural effusions are infrequent. The histologic pattern depends on the severity of the disease. **Predominant is an interstitial inflammatory reaction involving the walls of the alveoli.** The alveolar septa are widened and edematous and usually contain a mononuclear inflammatory infiltrate of lymphocytes, macrophages, and occasionally plasma cells. In severe cases, neutrophils may also be present. The alveoli may be free of exudate, but in many patients there is intra-alveolar proteinaceous material and a cellular exudate. When complicated by ARDS, pink hyaline membranes line the alveolar walls (see Fig. 15.4). Eradication of the infection is followed by reconstitution of the normal lung architecture.

Superimposed bacterial infection modifies this picture by causing ulcerative bronchitis, bronchiolitis, and bacterial pneumonia. Some viruses, such as herpes simplex, varicella, and adenovirus, may be associated with necrosis of bronchial and alveolar epithelium and acute inflammation. Characteristic viral cytopathic changes are described in Chapter 8.

Clinical Features

The clinical course of viral pneumonia is extremely varied. Many cases masquerade as severe upper respiratory tract infections or as chest colds. Even individuals with well-developed atypical pneumonia have few localizing symptoms. Cough may be absent, and the major manifestations may consist only of fever, headache, and myalgia. The edema and exudation often cause ventilation-perfusion mismatch leading to hypoxemia and thus evoke symptoms out of proportion to the scant physical findings.

Viral pneumonias are usually mild and resolve spontaneously without any lasting sequelae. However, interstitial viral pneumonias may assume epidemic proportions, and in such instances even a low rate of complications can lead to significant morbidity and mortality, as is typically true of influenza epidemics.

Health Care–Associated Pneumonia

Health care–associated pneumonia was recently described as a distinct clinical entity associated with several risk factors. These are hospitalization of at least 2 days within the recent

past; presentation from a nursing home or long-term care facility; attending a hospital or hemodialysis clinic; and recent intravenous antibiotic therapy, chemotherapy, or wound care. The most common organisms isolated are methicillin-resistant *S. aureus* and *P. aeruginosa.* These patients have a higher mortality than those with community-acquired pneumonia.

Hospital-Acquired Pneumonia

Hospital-acquired pneumonias are defined as pulmonary infections acquired in the course of a hospital stay. They are common in patients with severe underlying disease, immunosuppression, prolonged antibiotic therapy, or invasive access devices such as intravascular catheters. Patients on mechanical ventilation are at particularly high risk. Superimposed on an underlying disease (that caused hospitalization), hospital-acquired infections are serious and often life-threatening. Gram-positive cocci (mainly *S. aureus*) and gram-negative rods (Enterobacteriaceae and *Pseudomonas* species) are the most common isolates. The same organisms predominate in ventilator-associated pneumonia, with gram-negative bacilli being somewhat more common in this setting.

KEY CONCEPTS

ACUTE PNEUMONIA

- *S. pneumoniae* (the pneumococcus) is the most common cause of community-acquired acute pneumonia; the distribution of inflammation is usually lobar.
- Lobar pneumonias evolve through four stages: congestion, red hepatization, gray hepatization, and resolution.
- Other common causes of acute bacterial pneumonias in the community include *H. influenzae* and *M. catarrhalis* (both associated with acute exacerbations of COPD), *S. aureus* (usually secondary to viral respiratory infections), *K. pneumoniae* (observed in patients who are chronic alcoholics), *P. aeruginosa* (seen in persons with cystic fibrosis and in those with neutropenia), and *L. pneumophila,* seen particularly in individuals with co-morbid conditions (e.g., heart or lung disease) and in organ transplant recipients.
- Important causes of community-acquired viral pneumonia include influenza virus, metapneumonia virus, and coronavirus COVID-19, the latter a newly emergent pathogen.
- Bacterial pneumonias are characterized by predominantly intra-alveolar neutrophilic inflammation, while viral pneumonia shows interstitial lymphocytic inflammation.

Aspiration Pneumonia

Aspiration pneumonia occurs in markedly debilitated patients or those who aspirate gastric contents either while unconscious (e.g., after a stroke) or during repeated vomiting. These patients have abnormal gag and swallowing reflexes that predispose to aspiration. The resultant pneumonia is partly chemical due to the irritating effects of gastric acid and partly bacterial (from the oral flora). Typically, more than one organism is recovered on culture, aerobes being more common than anaerobes. This type of pneumonia is often necrotizing, pursues a fulminant clinical course, and is a frequent cause of death. In patients who survive, lung abscess is a common complication.

Microaspiration, in contrast, occurs frequently in almost all people, especially those with gastroesophageal reflux disease. It usually results in small, poorly formed non-necrotizing granulomas with multinucleated foreign body giant cell reaction. It is usually inconsequential, but may exacerbate other pre-existing lung diseases such as asthma, interstitial fibrosis, and lung rejection.

Lung Abscess

The term *pulmonary abscess* describes a local suppurative process that produces necrosis of lung tissue. Oropharyngeal surgical or dental procedures, sinobronchial infections, and bronchiectasis play important roles in their development.

Etiology and Pathogenesis

Under appropriate circumstances any bacterial pathogen can produce an abscess; those that do so most commonly include aerobic and anaerobic streptococci, *S. aureus,* and a host of gram-negative organisms. Mixed infections often occur because of the important causal role played by inhalation of foreign material. Anaerobic organisms normally found in the oral cavity, including members of the *Bacteroides, Fusobacterium,* and *Peptococcus* genera, are the exclusive isolates in about 60% of cases. The causative organisms are introduced by the following mechanisms:

- *Aspiration of infective material* (the most frequent cause). Risk factors include suppressed cough reflexes (e.g., acute alcohol intoxication, opioid abuse, coma, anesthesia, seizure disorders), severe dysphagia (e.g., neurologic deficits, esophageal disease), protracted vomiting, and poor dental hygiene. Aspiration first causes pneumonia, which progresses to tissue necrosis and formation of lung abscess.
- *Antecedent primary lung infection.* Postpneumonic abscess formations are usually associated with *S. aureus, K. pneumoniae,* and pneumococcus. Posttransplant or otherwise immunosuppressed individuals are at special risk.
- *Septic embolism.* Infected emboli may arise from thrombophlebitis in any portion of the systemic venous circulation or from the vegetations of infective bacterial endocarditis on the right side of the heart and lodge in the lung.
- *Neoplasia.* Secondary infection is particularly common in bronchopulmonary segments obstructed by a primary or secondary malignancy *(postobstructive pneumonia).*
- *Miscellaneous.* Traumatic penetrations of the lungs; direct extension of suppurative infections from the esophagus, spine, subphrenic space, or pleural cavity; and hematogenous seeding of the lung by pyogenic organisms all may lead to lung abscess formation.

When all these causes are excluded, there are still cases in which no discernible basis for the abscess formation can be identified. These are referred to as *primary cryptogenic lung abscesses.*

MORPHOLOGY

Abscesses vary in diameter from a few millimeters to large cavities of 5 to 6 cm (Fig. 15.36). They may affect any part of the lung and may be single or multiple. Pulmonary abscesses due to aspiration are more common on the right (because of the more vertical right main bronchus) and are most often single. Abscesses that

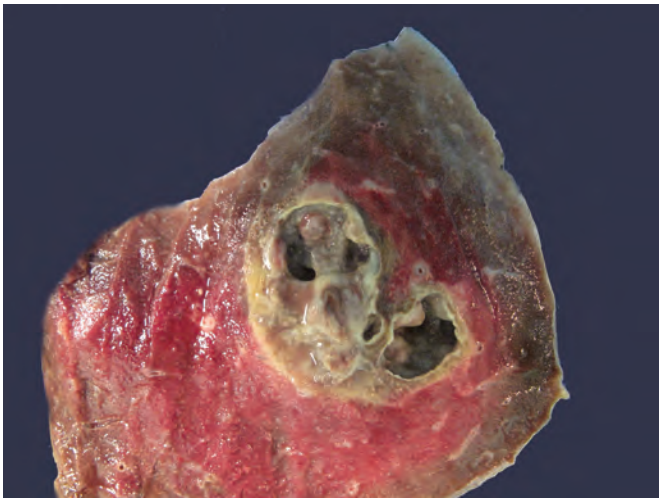


Figure 15.36 Cut surface of lung showing two abscesses. (Courtesy Dr. M. Kamran Mirza, University of Chicago, Chicago, Ill.)

develop in the course of pneumonia or bronchiectasis are usually multiple, basal, and diffusely scattered. Septic emboli and pyemic abscesses are multiple and may affect any region of the lungs.

The **cardinal histologic change in all abscesses is suppurative destruction of the lung parenchyma within the central area of cavitation.** The abscess cavity may be filled with suppurative debris or, if there is communication with an air passage, may be partially drained to create an air-containing cavity. Superimposed saprophytic infections are prone to develop within the necrotic debris. Continued infection leads to large, poorly demarcated, fetid, green-black, multilocular cavities designated gangrene of the lung. In chronic cases considerable fibroblastic proliferation produces a fibrous wall.

Clinical Features

The manifestations of pulmonary abscesses are much like those of bronchiectasis and characteristically include cough, fever, and copious amounts of foul-smelling purulent or sanguineous sputum. Fever, chest pain, and weight loss are common. Clubbing of the fingers and toes may appear. The diagnosis can be only suspected from the clinical findings and must be confirmed radiologically. Whenever an abscess is discovered in older individuals, it is important to rule out an underlying carcinoma, which is present in 10% to 15% of cases.

The course of abscesses is variable. With antimicrobial therapy, most resolve, leaving behind a scar. Complications include extension of the infection into the pleural cavity, hemorrhage, the development of *brain abscesses* or *meningitis* from septic emboli, and (rarely) secondary amyloidosis (type AA).

Chronic Pneumonia

Chronic pneumonia is most often a localized lesion in the immunocompetent patient, with or without regional lymph node involvement. Typically the inflammatory reaction is granulomatous and is caused by bacteria (e.g., *Mycobacterium tuberculosis*) or fungi (e.g., *Histoplasma capsulatum*). Tuberculosis of the lung and other organs is described in Chapter 8. Chronic pneumonias caused by fungi are discussed here.

Histoplasmosis

H. capsulatum infection is acquired by inhalation of dust particles from soil contaminated with bird or bat droppings that contain small spores (microconidia), the infectious form of the fungus. It is endemic along the Ohio and Mississippi rivers and in the Caribbean. It is also found in Mexico, Central and South America, parts of eastern and southern Europe, Africa, eastern Asia, and Australia. Like *M. tuberculosis, H. capsulatum* is an intracellular pathogen that is found mainly in phagocytes. The clinical presentations and morphologic lesions of histoplasmosis bear a striking resemblance to those of tuberculosis, including (1) a self-limited and often latent primary pulmonary involvement, which may result in coin lesions on chest radiography; (2) chronic, progressive, secondary lung disease, which is localized to the lung apices and causes cough, fever, and night sweats; (3) spread to extrapulmonary sites, including mediastinum, adrenal glands, liver, or meninges; and (4) widely disseminated disease in immunocompromised patients. Histoplasmosis can occur in immunocompetent individuals but as per usual is more severe in those with depressed cell mediated immunity.

The pathogenesis of histoplasmosis is incompletely understood. The portal of entry is virtually always the lung. Macrophages ingest but cannot kill the organism without T-cell help, and this allows the organism to multiply within phagolysosomes and disseminate prior to the development of T-cell immunity, which takes 1 to 2 weeks. In individuals with adequate cell-mediated immunity, the infection is controlled by Th1 helper T cells that recognize fungal antigens and subsequently secrete IFN-γ, which activates macrophages and enables them to kill intracellular yeasts. In addition, *Histoplasma* induces macrophages to secrete TNF, which recruits and stimulates other macrophages to kill *Histoplasma.*

MORPHOLOGY

In the lungs of otherwise healthy adults, *Histoplasma* infections produce **granulomas,** which usually become necrotic and may coalesce to produce areas of consolidation. With spontaneous resolution or effective treatment, these lesions undergo fibrosis and concentric calcification (tree-bark appearance) (Fig. 15.37A). Histologic differentiation from tuberculosis, sarcoidosis, and coccidioidomycosis requires identification of the 3- to 5-μm thin-walled yeast forms, which may persist in tissues for years. In **fulminant disseminated histoplasmosis,** which occurs in immunosuppressed individuals, granulomas do not form; instead, there are focal accumulations of mononuclear phagocytes filled with fungal yeasts throughout the body (Fig. 15.37B).

The diagnosis of histoplasmosis may be established by serologic tests for antibodies and fungal antigens, culture, or identification of the fungus in tissue biopsies. The majority of cases resolve spontaneously. Progressive disease or disease in immunocompromised patients is treated with antifungal agents.

Blastomycosis

Blastomyces dermatitidis is a soil-inhabiting dimorphic fungus. It causes disease in the central and southeastern United States; infection also occurs in Canada, Mexico, the Middle East, Africa, and India. There are three clinical forms: *pulmonary*

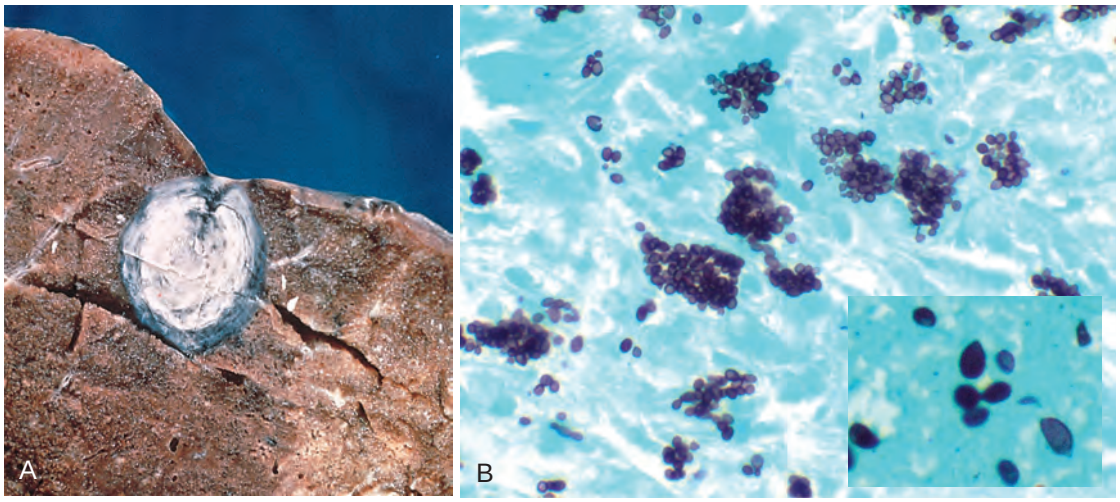


Figure 15.37 Histoplasmosis. (A) Laminated *Histoplasma* granuloma of the lung. (B) *Histoplasma capsulatum* yeast forms fill phagocytes in the lung of a patient with disseminated histoplasmosis; *inset* shows high-power detail of pear-shaped, thin-based budding yeasts (silver stain).

blastomycosis, disseminated blastomycosis, and a rare *primary cutaneous form* that results from direct inoculation of organisms into the skin. The pneumonia most often resolves spontaneously, but it may persist or progress to a chronic lesion.

MORPHOLOGY

In the normal host, the lung lesions of blastomycosis are suppurative granulomas. Macrophages have a limited ability to ingest and kill *B. dermatitidis*, and the persistence of the yeast cells leads to the recruitment of neutrophils. In tissue, *B. dermatitidis* is a round, 5- to 15-μm yeast cell that divides by broad-based budding. It has a thick, double-contoured cell wall, and visible nuclei (Fig. 15.38). Involvement of the skin and larynx is associated with marked epithelial hyperplasia, which may be mistaken for squamous cell carcinoma.

Coccidioidomycosis

Almost everyone who inhales the spores of *Coccidioides immitis* becomes infected and develops a delayed-type hypersensitivity reaction to the fungus, but most remain asymptomatic. Indeed, more than 80% of people in endemic areas of the southwestern and western United States and in Mexico have a positive skin test reaction. One reason for the infectivity of *C. immitis* is that infective arthroconidia, when ingested by alveolar macrophages, block fusion of the phagosome and lysosome and so resist intracellular killing. Approximately 10% of infected people develop lung lesions, and less than 1% of people develop disseminated *C. immitis* infection, which frequently involves the skin and meninges. Certain ethnic groups (e.g., Filipinos and African Americans) and the immunosuppressed are at high risk for disseminated disease.

MORPHOLOGY

Within macrophages or giant cells, *C. immitis* is present as thick-walled, nonbudding spherules 20 to 60 μm in diameter, often filled with small endospores. A pyogenic reaction is superimposed when the spherules rupture to release the endospores (Fig. 15.39). Rare progressive *C. immitis* disease involves the lungs, meninges, skin, bones, adrenals, lymph nodes, spleen, or liver. At all these sites, the inflammatory response may be purely granulomatous, pyogenic, or mixed.

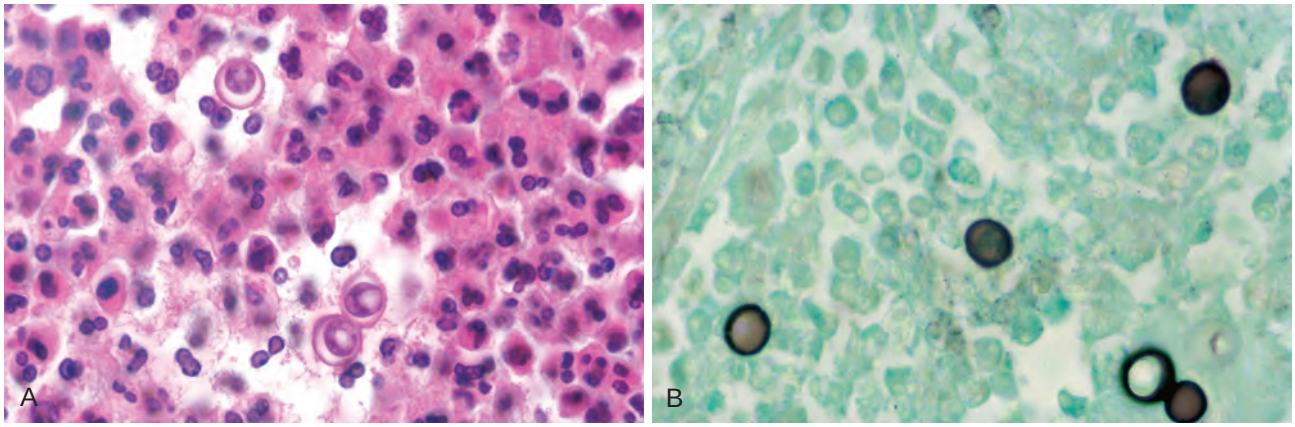


Figure 15.38 Blastomycosis. (A) Rounded budding yeasts, larger than neutrophils, are present. Note the characteristic thick wall and nuclei (not seen in other fungi). (B) Silver stain.

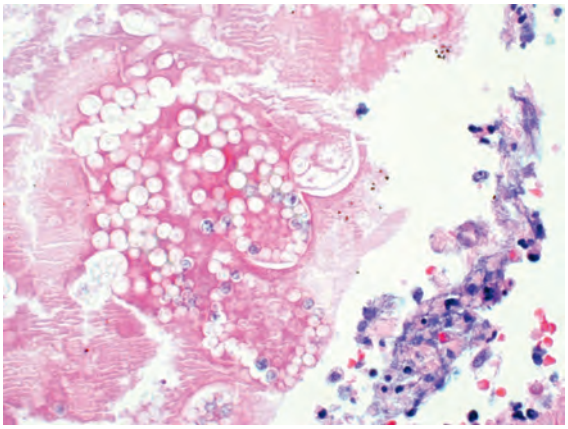


Figure 15.39 Coccidioidomycosis. Intact and ruptured spherules are seen.

Pneumonia in the Immunocompromised Host

The appearance of a pulmonary infiltrate, with or without signs of infection (e.g., fever), is one of the most common and serious complications in patients whose immune defenses are suppressed by disease, immunosuppressive therapy for organ or hematopoietic stem cell transplants, chemotherapy for tumors, or irradiation. In addition to the usual pathogens, a wide variety of so-called opportunistic infectious agents, many of which rarely cause infection in normal hosts, can cause pneumonia, and often more than one agent is involved. Mortality from these opportunistic infections is high. Table 15.8 lists some of the opportunistic agents according to their prevalence and whether they cause local or diffuse pulmonary infiltrates. The differential diagnosis of such infiltrates includes drug reactions and involvement of the lung by tumor. The specific infections are discussed in Chapter 8. Of these, the ones that commonly involve the lung can be classified according to the etiologic agent: (1) bacteria (*P. aeruginosa*, *Mycobacterium* species, *L. pneumophila*, and *Listeria monocytogenes*), (2) viruses (cytomegalovirus [CMV] and herpesvirus), and (3) fungi (*P. jiroveci*, *Candida* species, *Aspergillus* species, the Phycomycetes, and *Cryptococcus neoformans*).

Table 15.8 Causes of Pulmonary Infiltrates in Immunocompromised Hosts

Diffuse Infiltrates	Focal Infiltrates
Common	
Cytomegalovirus	Gram-negative bacterial infections
Pneumocystis jiroveci	*Staphylococcus aureus*
Drug reaction	*Aspergillus*
	Candida
	Malignancy
Uncommon	
Bacterial pneumonia	*Cryptococcus*
Aspergillus	*Mucor*
Cryptococcus	*Pneumocystis jiroveci*
Malignancy	*Legionella pneumophila*

Pulmonary Disease in Human Immunodeficiency Virus Infection

Pulmonary disease accounts for 30% to 40% of hospitalizations in HIV-infected individuals. Although the use of potent antiretroviral agents and effective chemoprophylaxis has markedly altered the incidence and outcome of pulmonary disease in HIV-infected persons, the plethora of infectious agents and other pulmonary lesions make diagnosis and treatment a distinct challenge. Some of the individual microbial agents afflicting HIV-infected individuals have already been discussed; this section focuses only on the general principles of HIV-associated pulmonary disease.

- *Despite the emphasis on opportunistic infections, it must be remembered that bacterial lower respiratory tract infections caused by the "usual" pathogens are among the most serious pulmonary disorders in HIV infection.* The implicated organisms include *S. pneumoniae, S. aureus, H. influenzae,* and gram-negative rods. Bacterial pneumonias in HIV-infected persons are more common, more severe, and more often associated with bacteremia than in those without HIV infection.
- *Not all pulmonary infiltrates in HIV-infected individuals are infectious in etiology.* A host of noninfectious diseases, including Kaposi sarcoma (Chapters 6 and 11), non-Hodgkin lymphoma (Chapter 13), and lung cancer, occur with increased frequency and must be excluded.
- *The CD4+ T-cell count determines the risk of infection with specific organisms.* As a rule of thumb, bacterial and tubercular infections are more likely at higher CD4+ counts (>200 cells/mm^3). *Pneumocystis* pneumonia usually strikes at CD4+ counts less than 200 cells/mm^3, while CMV, fungal, and *Mycobacterium avium-intracellulare* complex infections are uncommon until the disease is very advanced (CD4+ counts less than 50 cells/mm^3).

Finally, pulmonary disease in HIV-infected persons may result from more than one cause, and even common pathogens may present with atypical manifestations. Therefore the diagnostic work-up of these patients may be more extensive (and expensive) than would be necessary in an immunocompetent individual.

LUNG TRANSPLANTATION

Indications for transplantation may include almost all non-neoplastic terminal lung diseases, provided that the patient does not have any other serious disease that would preclude lifelong immunosuppressive therapy. The most common indications are end-stage COPD, idiopathic pulmonary fibrosis, cystic fibrosis, and idiopathic/familial pulmonary arterial hypertension. Although bilateral lung transplantation offers better survival as compared to single lung, the latter may be performed to benefit two recipients from a single (and all too scarce) donor. When bilateral chronic infection is present (e.g., cystic fibrosis, bronchiectasis), both lungs of the recipient must be replaced to remove the reservoir of infection.

With improving surgical and organ preservation techniques, postoperative complications (e.g., anastomotic

dehiscence, vascular thrombosis, primary graft dysfunction) are becoming rare. The transplanted lung is subject to two major complications: infection and rejection.

- *Pulmonary infections* in lung transplant patients are essentially those of any immunocompromised host, discussed earlier. In the early posttransplant period (the first few weeks), bacterial infections are most common. With ganciclovir prophylaxis and matching of donor-recipient CMV status, CMV pneumonia occurs less frequently and is less severe, although some resistant strains are emerging. Most infections occur in the third to twelfth month after transplantation. *P. jiroveci* pneumonia is rare, since almost all patients receive adequate prophylaxis, usually with trimethoprim-sulfamethoxazole (Bactrim). Fungal infections are mostly due to *Aspergillus* and *Candida* species, and they may involve the bronchial anastomotic site and/or the lung.
- *Acute lung allograft rejection* occurs to some degree in all patients despite routine immunosuppression. It most often appears several weeks to months after surgery but also may present years later or whenever immunosuppression is decreased. Patients present with fever, dyspnea, cough, and radiologic infiltrates. Since these are similar to the picture of infections, diagnosis often relies on transbronchial biopsy.
- *Chronic lung allograft rejection* is a significant problem in at least half of all patients by 3 to 5 years posttransplant. It is manifested by cough, dyspnea, and an irreversible decrease in lung function due to pulmonary fibrosis.

MORPHOLOGY

The morphologic features of acute rejection are primarily those of inflammatory infiltrates (lymphocytes, plasma cells, and few neutrophils and eosinophils) around small vessels, in the submucosa of airways, or both. The major morphologic correlate of chronic rejection is **bronchiolitis obliterans,** the partial or complete occlusion of small airways by fibrosis, with or without active inflammation (Fig. 15.40). Bronchiolitis obliterans is patchy and therefore difficult to diagnose via transbronchial biopsy. Bronchiectasis and pulmonary fibrosis may also develop with long-standing chronic rejection.

Acute cellular airway rejection (the presumed forerunner of later, fibrous obliteration of these airways) is generally responsive to therapy, but the treatment of established bronchiolitis obliterans has been disappointing. Its progress may be slowed or even halted for some time, but it cannot be reversed. Infrequent complications of lung transplantation include Epstein-Barr virus (EBV)–associated B-cell lymphoma, which most often arises within the lung allograft. With continuing improvement in surgical, immunosuppressive, and antimicrobial therapies, the outcome of lung transplantation has improved considerably. The overall median survival is 6 years, with younger patients and those undergoing bilateral lung transplantation having better outcomes.

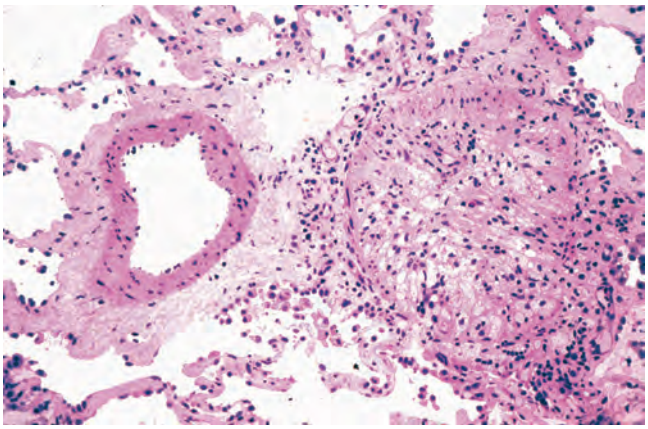


Figure 15.40 Chronic rejection of lung allograft associated with bronchiole (bronchiolitis obliterans). An adjacent pulmonary artery is normal. (Courtesy Dr. Thomas Krausz, Department of Pathology, The University of Chicago, Pritzker School of Medicine, Chicago, Ill.)

TUMORS

Of the wide variety of benign and malignant tumors that may arise in the lung, 90% to 95% are carcinomas, about 5% are carcinoid tumors, and 2% to 5% are mesenchymal and other miscellaneous neoplasms.

Carcinomas

Lung cancer is currently the most frequently diagnosed major cancer and the most common cause of cancer mortality worldwide. Globally, in 2018 there were an estimated 2.1 million new cases and 1.8 million lung cancer deaths. The number of new cases of lung cancer in 2018 in the United States is expected to number approximately 230,000 (note that in 1950 it was 18,000), accounting for about 14% of cancer diagnoses and taking more than 150,000 lives, which amounts to roughly 28% of all cancer-related deaths. Each year, lung cancer kills more people than colon, breast, and prostate cancer combined. It is generally a disease of older adults, occurring most often between ages 55 and 84 years, with a peak incidence between 65 and 74 years. Only 2% of all cases occur before the age of 40.

Because lung cancer is strongly linked to cigarette smoking, changes in smoking habits greatly influence lung cancer incidence and mortality as well as the prevalence of the various histologic types of lung cancer. Since the early 1990s, lung cancer incidence and mortality rates have been decreasing in men due to a decrease in male smoking over the past 35 years. However, the decrease in smoking among women has lagged behind that of men. Since 1987 more women have died each year of lung cancer than of breast cancer, which for more than 40 years had been the leading cause of cancer death in women.

Etiology and Pathogenesis

Most (but not all) lung cancers are associated with a well-known carcinogen—cigarette smoke. In addition, there are other genetic and environmental factors. Lung cancers are

broadly classified into small cell and non–small cell types, with the latter group including adenocarcinoma and squamous cell carcinoma. The driver mutations that cause lung cancer vary among these histologic subtypes, as will be described later.

Tobacco Smoking. **About 80% of lung cancers occur in active smokers or those who stopped recently, and there is a nearly linear correlation between the frequency of lung cancer and pack-years of cigarette smoking.** The increased risk is 60 times greater in habitual heavy smokers (two packs a day for 20 years) than in nonsmokers. However, since lung cancer develops in only 10% to 15% of smokers, there are likely to be other factors that interact with smoking to predispose individuals to this deadly disease. For unclear reasons, it appears that women are more susceptible to carcinogens in tobacco than men. Although cessation of smoking decreases the risk for lung cancer over time, it may never return to baseline levels. In fact, genetic changes that predate lung cancer can persist for many years in the bronchial epithelium of former smokers. Pipe and cigar smokers also incur an elevated risk, albeit only modestly. Chewing tobacco is not a safe substitute for smoking cigarettes or cigars, as these products spare the lung but cause oral cancers and can lead to nicotine addiction. The long-term effects of electronic cigarette aerosols are not known, as "vaping" is a relatively recent phenomenon (Chapter 9).

Unfortunately, the carcinogenic effects of tobacco smoke extend to those who live and work with smokers. *Secondhand smoke,* or environmental tobacco smoke, contains numerous human carcinogens for which there is no safe level of exposure. It is estimated that each year about 3000 nonsmoking adults die of lung cancer as a result of breathing secondhand smoke.

What of heavy smokers who never develop cancer? While some of this may be a matter of chance, the mutagenic effect of carcinogens in smoke is modified by genetic variants. Recall that many chemicals (procarcinogens) are converted into carcinogens via activation by the highly polymorphic P-450 monooxygenase enzyme system (Chapter 9). Specific P-450 variants have an increased capacity to activate procarcinogens in cigarette smoke, and smokers with these variants incur a greater risk of lung cancer. Similarly, individuals whose peripheral blood lymphocytes show more numerous chromosomal breakages after exposure to tobacco-related carcinogens (mutagen sensitivity genotype) have a greater than 10-fold higher risk of developing lung cancer as compared with controls, presumably because of genetic variation in genes involved in DNA repair.

The histologic changes that correlate with steps along the path to neoplastic transformation are best documented for squamous cell carcinoma and are described in more detail later. There is a linear correlation between the intensity of exposure to cigarette smoke and the appearance of ever more worrisome epithelial changes. These begin with rather innocuous-appearing basal cell hyperplasia and squamous metaplasia and progress to squamous dysplasia and carcinoma in situ, the last stage before progression to invasive cancer.

Industrial Hazards. Certain industrial exposures, such as asbestos, arsenic, chromium, uranium, nickel, vinyl chloride, and mustard gas, increase the risk of developing lung cancer. High-dose ionizing radiation is carcinogenic. There was an increased incidence of lung cancer among survivors of the Hiroshima and Nagasaki atomic bomb blasts, as well as in workers heavily involved in clean-up after the Chernobyl disaster. Uranium is weakly radioactive, but lung cancer rates among nonsmoking uranium miners are four times higher than those in the general population, and among smoking miners they are about 10 times higher. Asbestos exposure also increases the risk for lung cancer development. The latent period before the development of lung cancer is 10 to 30 years. Lung cancer is the most frequent malignancy in individuals exposed to asbestos, particularly when coupled with smoking. Asbestos workers who do not smoke have a five-fold greater risk of developing lung cancer than do nonsmoking control subjects, whereas those who smoke have a 55-fold greater risk.

Air Pollution. It is uncertain whether air pollution, by itself, increases the risk of lung cancer, but it likely adds to the risk in those who smoke or are exposed to secondhand smoke. It may do so through several different mechanisms. Chronic exposure to air particulates in smog may cause lung irritation, inflammation, and repair, and you will recall that chronic inflammation and repair increases the risk of a variety of cancers (Chapter 7). A specific form of air pollution that may contribute to an increased risk of lung cancer is radon gas. Radon is a ubiquitous radioactive gas that has been linked epidemiologically to increased lung cancer in uranium miners. Other underground miners and workers in locations below ground, such as subways, tunnels, and basements, are at increased risk for radon exposure. This has generated concern that low-level exposure (e.g., in well-insulated homes in areas with naturally high levels of radon in soil) may also increase the incidence of lung cancer.

Acquired Mutations. As with other cancers (Chapter 7), smoking-related carcinomas of the lung arise by a stepwise accumulation of oncogenic "driver" mutations that result in the neoplastic transformation of pulmonary epithelial cells. Some of the genetic changes associated with cancers can be found in the "benign" bronchial epithelium of smokers without lung cancers, suggesting that large areas of the respiratory mucosa are mutagenized by exposure to carcinogens in tobacco smoke ("field effect"). On this fertile soil, cells that accumulate just the "wrong" panoply of complementary driver mutations to acquire all of the hallmarks of cancer develop into carcinomas.

The major histologic subtypes of lung cancer each have distinctive molecular features, as follows:

- *Adenocarcinoma* is associated with tobacco smoking, but less so than other histologic subtypes; as a result, it is the most common subtype in never-smokers (described below). About one-third of adenocarcinomas have oncogenic gain-of-function mutations involving components of growth factor receptor signaling pathways; these are important to recognize because they often can be targeted with specific inhibitors (discussed later). These include gain-of-function mutations in genes encoding several different receptor tyrosine kinases, such as: *EGFR,* in 10% to 15% of tumors in Caucasians and a higher percentage of nonsmoking Asian women; *ALK,* in 3% to 5% of tumors;

ROS1, in 1% of tumors; *MET,* in 2% to 5% of tumors; and *RET,* in 1% to 2% of tumors. Other tumors have gain-of-function mutations in serine/threonine kinases (*BRAF,* 2% of tumors, and *PI3K,* 2% of tumors) or in the *KRAS* gene (roughly 30% of tumors), all of which encode signaling molecules that lie downstream of receptor tyrosine kinases in growth factor signaling pathways.

- *Squamous cell carcinoma* is highly associated with exposure to tobacco smoke and harbors diverse genetic aberrations, many of which are chromosome deletions involving tumor suppressor loci. These losses, especially those involving 3p, 9p (site of the *CDKN2A* gene), and 17p (site of the *TP53* gene), are early events in tumor evolution, being detected at an appreciable frequency in the histologically normal respiratory mucosal cells of smokers. Most tumors have mutations in *TP53*, and p53 protein overexpression (as seen by immunohistochemical staining), a marker of *TP53* mutations, is an early event, being reported in 10% to 50% of squamous dysplasias and 60% to 90% of squamous cell carcinoma in situ. The *CDKN2A* tumor suppressor gene, which encodes the cyclin-dependent kinase inhibitor p16, is mutated in 65% of tumors. Many squamous cell carcinomas also have amplification of *FGFR1,* a gene encoding the fibroblast growth factor receptor tyrosine kinase.
- *Small cell carcinoma* is virtually always smoking related and has the highest mutational burden among lung cancers. There is almost universal inactivation of both *TP53* and *RB,* and unusual transformations of non–small cell carcinoma to small cell carcinoma are often associated with acquisition of *RB* loss-of-function mutations, emphasizing the importance of *RB* inactivation in this lung cancer subtype. Loss of chromosome 3p also occurs in nearly all of these tumors and is seen even in histologically normal lung epithelium, suggesting that this also is a critical early event. This subtype is also commonly associated with amplification of genes of the *MYC* family.

Lung Cancer in Never-Smokers. The WHO estimates that 25% of lung cancer worldwide occurs in never-smokers. This percentage is probably closer to 10% to 15% in Western countries. These cancers occur more commonly in women, and most are adenocarcinomas, often with targetable mutations/co-mutations. Cancers in nonsmokers are more likely to have *EGFR* mutations and almost never have *KRAS* mutations; *TP53* mutations are not uncommon, but occur less frequently than in smoking-related cancers.

Precursor (Preinvasive) Lesions. Four types of morphologic precursor epithelial lesions are recognized: (1) atypical adenomatous hyperplasia, (2) adenocarcinoma in situ, (3) squamous dysplasia and carcinoma in situ, and (4) diffuse idiopathic pulmonary neuroendocrine cell hyperplasia. It should be remembered that the term *precursor* does not imply that progression to cancer is inevitable. Currently it is not possible to distinguish between precursor lesions that progress and those that remain localized or regress.

Classification

Tumor classification is important for consistency in patient treatment and provides a uniform basis for epidemiologic and biologic studies. The most recent classification of lung cancer is given in Table 15.9. Several histologic variants of each type of lung cancer are described; however, their clinical significance is still undetermined except as mentioned herein. The relative proportions of the major categories are:

- Adenocarcinoma (50%)
- Squamous cell carcinoma (20%)
- Small cell carcinoma (15%)
- Large cell carcinoma (2%)
- Other (13%)

There may be mixtures of histologic patterns, even in the same cancer. Thus combinations of squamous cell carcinoma and adenocarcinoma or small cell and squamous cell carcinoma occur in about 14% and 5% of patients, respectively.

The incidence of adenocarcinoma has increased significantly in the last 2 decades, and it is now the most common form of lung cancer in women and men. The basis for this change is unclear. One possible factor is the increase in women smokers, but this only highlights our ignorance about why women develop adenocarcinoma more frequently. Another possibility is that changes in cigarettes (altered filter tips and decreased tar and nicotine) may have caused smokers to inhale more deeply, increasing the exposure of peripheral airways and cells with a predilection to give rise to adenocarcinoma to carcinogens.

Table 15.9 Histologic Classification of Malignant Epithelial Lung Tumors

Tumor Classification
Adenocarcinoma
Lepidic, acinar, micropapillary, papillary, solid (according to predominant pattern)
Invasive mucinous adenocarcinoma
Minimally invasive adenocarcinoma (nonmucinous, mucinous)
Squamous cell carcinoma
Keratinizing, nonkeratinizing, basaloid
Neuroendocrine tumors
Small cell carcinoma
Combined small cell carcinoma
Large cell neuroendocrine carcinoma
Combined large-cell neuroendocrine carcinoma
Carcinoid tumor
Typical, atypical
Other uncommon types
Large cell carcinoma
Adenosquamous carcinoma
Sarcomatoid carcinoma
Pleomorphic, spindle cell, giant cell carcinoma, carcinosarcoma, pulmonary blastoma
Others such as lymphoepithelioma-like carcinoma and NUT carcinoma
Salivary gland–type tumors

MORPHOLOGY

Lung carcinomas may arise in the peripheral lung (more often adenocarcinomas) or in the central/hilar region (more often squamous cell carcinomas), sometimes in association with recognizable precursor lesions.

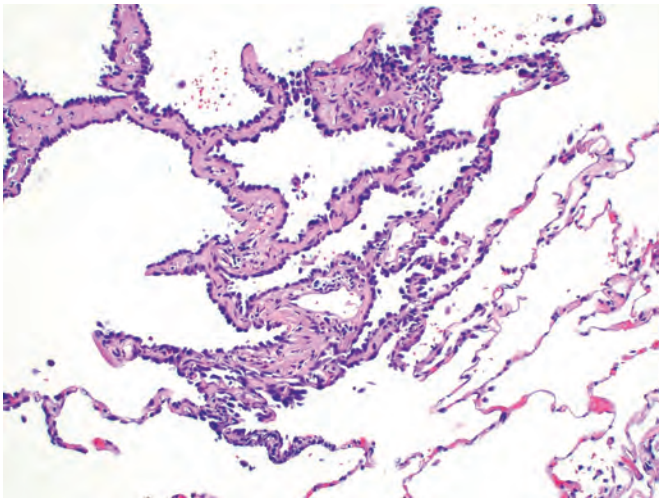


Figure 15.41 Atypical adenomatous hyperplasia. The epithelium is cuboidal, and there is mild interstitial fibrosis.

Atypical adenomatous hyperplasia is a small precursor lesion (≤5 mm) characterized by dysplastic pneumocytes lining alveolar walls that are mildly fibrotic (Fig. 15.41). It can be single or multiple and can be in the lung adjacent to invasive tumor or away from it.

Adenocarcinoma in situ (formerly called bronchioloalveolar carcinoma) is a lesion that is less than 3 cm in size and is composed entirely of dysplastic cells growing along pre-existing alveolar septa. The cells have more dysplasia than atypical adenomatous hyperplasia and may or may not have intracellular mucin (Fig. 15.42).

Adenocarcinoma is an invasive malignant epithelial tumor with glandular differentiation or mucin production by the tumor cells. Adenocarcinomas grow in various patterns, including acinar, lepidic, papillary, micropapillary, and solid. Compared with squamous cell cancers, these lesions are usually more peripherally located and tend to be smaller. They vary histologically from well-differentiated tumors with obvious glandular elements (Fig. 15.43A), to papillary lesions resembling other papillary carcinomas, to solid masses with only occasional mucin-producing glands and cells. The majority express thyroid transcription factor-1 (TTF-1) (Fig. 15.43A inset), a protein first identified in the thyroid that is required for normal lung development. At the periphery of the tumor there is often a lepidic pattern of spread, in which the tumor cells "crawl" along normal-appearing alveolar septa. Tumors (≤3 cm) with a small invasive component (≤5 mm) associated with scarring and a peripheral lepidic growth pattern are called **microinvasive adenocarcinoma.** These have a far better prognosis than invasive carcinomas of the same size. **Mucinous adenocarcinomas** tend to spread aerogenously, forming satellite tumors; thus, these are less likely to be cured by surgery. They may present as a solitary nodule or as multiple nodules, or an entire lobe may be consolidated by tumor, mimicking lobar pneumonia.

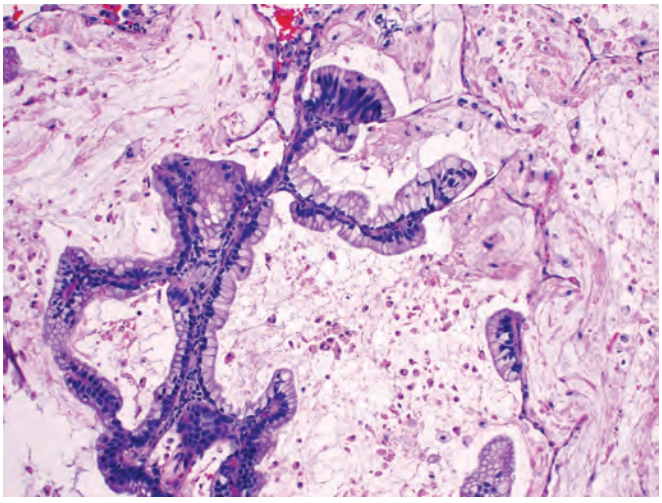


Figure 15.42 Adenocarcinoma in situ, mucinous subtype. Characteristic growth along pre-existing alveolar septa is evident, without invasion.

Squamous cell carcinoma is more common in men and is strongly associated with smoking. Precursor lesions that give rise to invasive squamous cell carcinoma are well characterized. Squamous cell carcinomas are often antedated by **squamous metaplasia** or **dysplasia** in the bronchial epithelium, which then transforms to **carcinoma in situ,** a phase that may last for years (Fig. 15.44). By this time, atypical cells may be identified in cytologic smears of sputum or in bronchial lavage fluids or brushings (Fig. 15.45), but the lesion is asymptomatic and undetectable on radiographs. Eventually, an invasive squamous cell carcinoma appears. The tumor may then follow a variety of paths. It may grow exophytically into the bronchial lumen, producing an intraluminal mass. With further enlargement the bronchus becomes obstructed, leading to distal atelectasis and infection. The tumor may also penetrate the wall of the bronchus and infiltrate along the peribronchial tissue (Fig. 15.46) into the adjacent carina or mediastinum. In other instances, the tumor grows along a broad front to produce a cauliflower-like intraparenchymal mass that compresses the surrounding lung. As in almost all types of lung cancer, the neoplastic tissue is gray-white and firm to hard. Especially when the tumors are bulky, focal areas of hemorrhage or necrosis may appear to produce red or yellow-white mottling and softening. Sometimes these necrotic foci cavitate.

Histologically, squamous cell carcinoma is characterized by the presence of keratinization and/or intercellular bridges. Keratinization may take the form of squamous pearls or individual cells with markedly eosinophilic cytoplasm (see Fig. 15.43B). These features are prominent in well-differentiated tumors, are easily seen but not extensive in moderately differentiated tumors, and are focally seen in poorly differentiated tumors. Mitotic activity is higher in poorly differentiated tumors. In the past, most squamous cell carcinomas arose centrally from the segmental or subsegmental bronchi, but the incidence of squamous cell carcinoma of the peripheral lung is increasing. Squamous metaplasia, epithelial dysplasia, and foci of frank carcinoma in situ may be seen in bronchial epithelium adjacent to the tumor mass (see Fig. 15.44).

Small cell carcinoma is a highly malignant tumor with a strong relationship to cigarette smoking; only about 1% occurs in nonsmokers. Tumors may arise in major bronchi or in the periphery of the lung. There is no known pre-invasive phase. They are the most aggressive of lung tumors, metastasizing widely and virtually always proving to be fatal.

Small cell carcinoma is comprised of relatively small cells with scant cytoplasm, ill-defined cell borders, finely granular nuclear

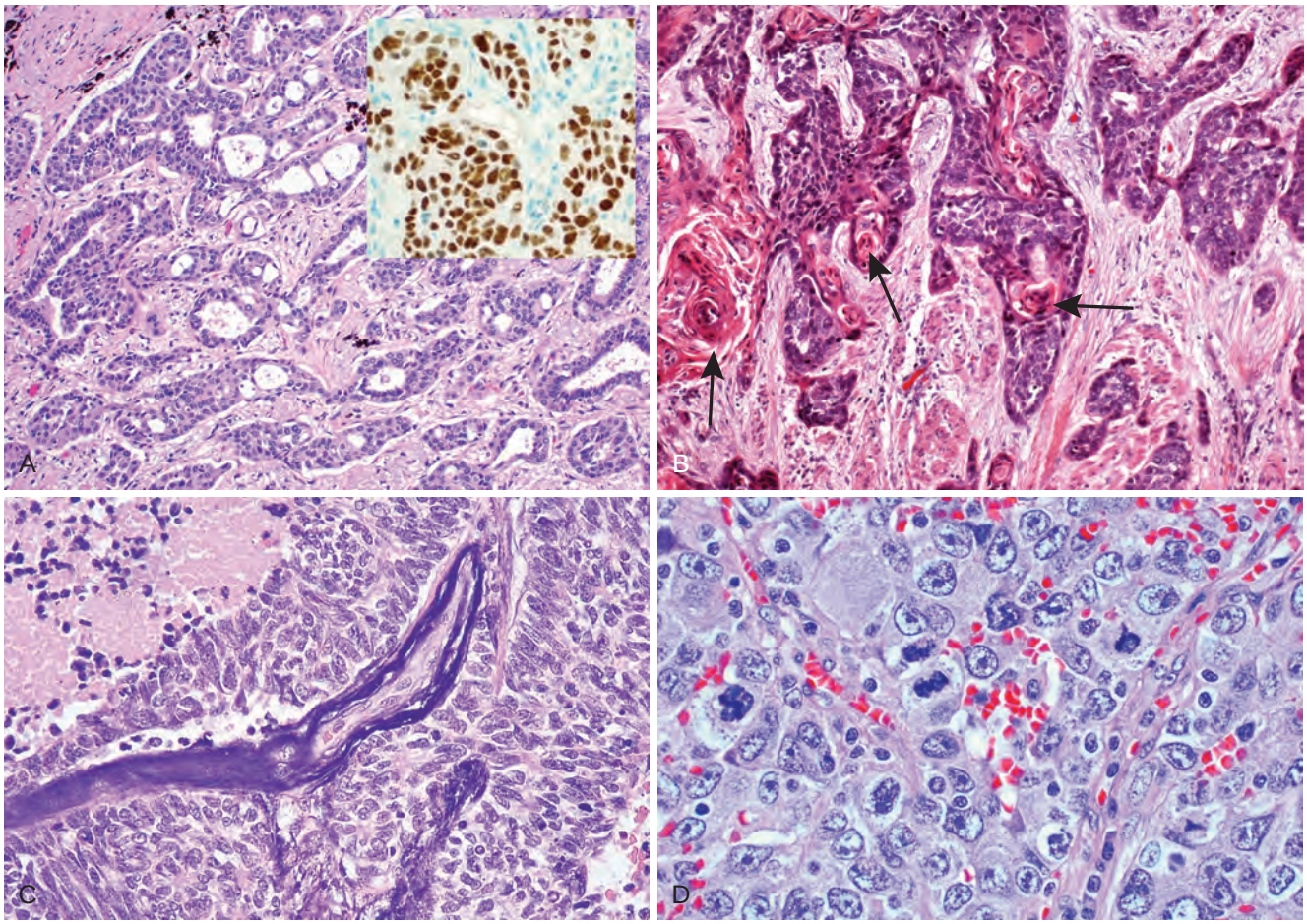


Figure 15.43 Histologic variants of lung carcinoma. (A) Gland-forming adenocarcinoma; *inset* shows thyroid transcription factor 1 (TTF-1) expression, as detected by immunohistochemistry. (B) Well-differentiated squamous cell carcinoma showing keratinization *(arrow)*. (C) Small cell carcinoma. There are islands of small, deeply basophilic cells and areas of necrosis. (D) Large cell carcinoma. The tumor cells are pleomorphic and show no evidence of squamous or glandular differentiation.

chromatin (salt and pepper pattern), and absent or inconspicuous nucleoli (see Fig. 15.43C). The cells are round, oval, or spindle-shaped, and nuclear molding is prominent. There is no absolute size for the tumor cells, but in general they are smaller than three times the diameter of a small resting lymphocyte (a size of about 25 μm). The mitotic count is high. The cells grow in clusters that exhibit neither glandular nor squamous organization. Necrosis is common and often extensive. Basophilic staining of vascular walls due to encrustation by DNA from necrotic tumor cells (Azzopardi effect) is frequently present. Combined small cell carcinoma is a variant in which typical small cell carcinoma is mixed with non–small cell histologies, such as large cell neuroendocrine carcinoma or even spindled cell morphologies resembling sarcoma.

Electron microscopy shows dense-core neurosecretory granules, about 100 nm in diameter, in two-thirds of cases of small cell carcinoma. The occurrence of neurosecretory granules; the expression of neuroendocrine markers such as chromogranin, synaptophysin, and CD56; and the ability of some of these tumors to secrete hormones (e.g., parathormone-related protein, a cause of paraneoplastic hypercalcemia) suggest that this tumor originates from neuroendocrine progenitor cells, which are present in the lining bronchial epithelium. This simplistic idea is challenged, however, by the existence of tumors comprised of a mixture of small cell carcinoma and other histologies and well-documented "transformations" of non–small cell carcinoma to small cell carcinoma. Among the various types of lung cancer, small cell carcinoma is the one that is most commonly associated with ectopic hormone production (discussed later).

Large cell carcinoma is an undifferentiated malignant epithelial tumor that lacks the cytologic features of other forms of lung cancer. The cells typically have large nuclei, prominent nucleoli, and a moderate amount of cytoplasm (see Fig. 15.43D). Large cell carcinoma is a diagnosis of exclusion since it expresses none of the markers associated with adenocarcinoma (TTF-1, napsin A) or squamous cell carcinoma (p40, p63). One histologic variant is large cell neuroendocrine carcinoma, which has molecular features similar to those of small cell carcinoma, but is comprised of tumor cells of larger size.

Combined Carcinoma. Approximately 4% to 5% of all lung carcinomas have a combined histology, including two or more of the aforementioned types.

Any type of lung carcinoma may extend to the pleural surface and then spread within the pleural cavity or into the pericardium. Metastases to the bronchial, tracheal, and mediastinal nodes can

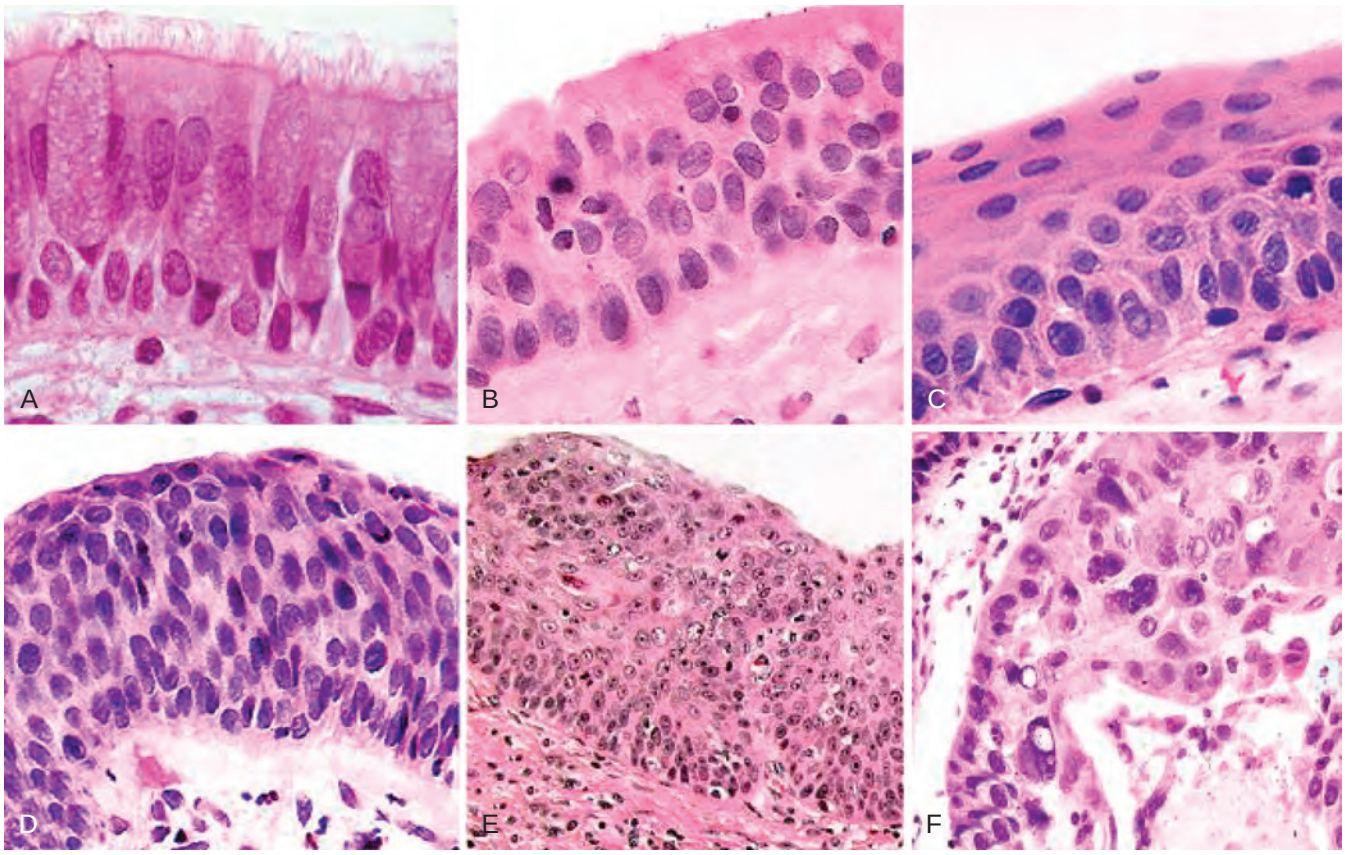


Figure 15.44 Precursor lesions of squamous cell carcinomas. Some of the earliest ("mild") changes in smoking-damaged respiratory epithelium include goblet cell hyperplasia (A), basal cell (or reserve cell) hyperplasia (B), and squamous metaplasia (C). More ominous changes include the appearance of squamous dysplasia (D), characterized by the presence of disordered squamous epithelium, with loss of nuclear polarity, nuclear hyperchromasia, pleomorphism, and mitotic figures. Squamous dysplasia may progress through the stages of mild, moderate, and severe dysplasia. Carcinoma in situ (E), the stage immediately preceding invasive squamous carcinoma (F), by definition has not penetrated the basement membrane and has cytologic features similar to those in frank carcinoma. (A–E, Courtesy Dr. Adi Gazdar, Department of Pathology, University of Texas, Southwestern Medical School, Dallas, Tex. F, Reproduced with permission from Travis WD, et al, editors: *World Health Organization Histological Typing of Lung and Pleural Tumors*, Heidelberg, 1999, Springer.)

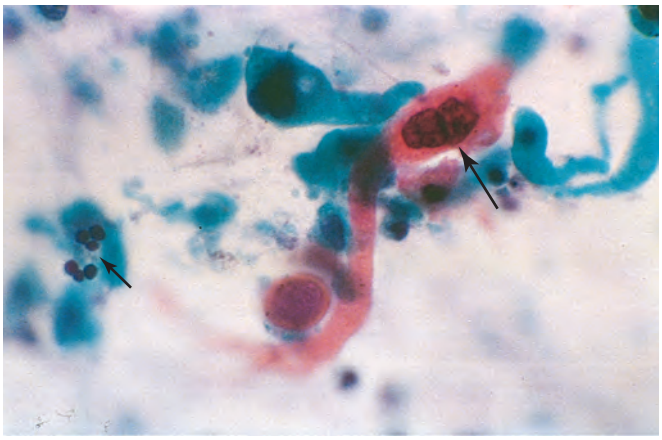


Figure 15.45 Cytologic diagnosis of lung cancer. A sputum specimen shows an orange-staining, keratinized squamous carcinoma cell with a prominent hyperchromatic nucleus *(large arrow)*. Note the size of the tumor cells compared with normal neutrophils *(small arrow)*.

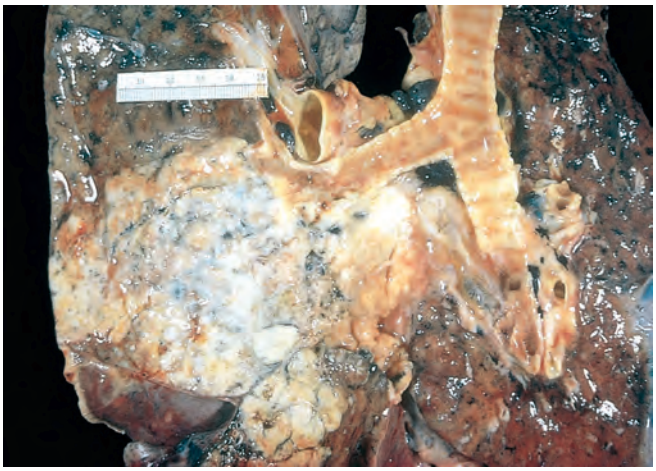


Figure 15.46 Lung carcinoma. The gray-white tumor infiltrates the lung parenchyma. Histologic sections identified this tumor as a squamous cell carcinoma.

be found in most cases. The frequency of nodal involvement varies slightly with the histologic pattern but averages greater than 50%.

Distant spread of lung carcinoma occurs through both lymphatic and hematogenous pathways. These tumors often spread early throughout the body except for squamous cell carcinoma, which metastasizes late outside the thorax. Metastasis may be the first manifestation of an underlying occult pulmonary lesion. No organ or tissue is spared, but the adrenal glands, for obscure reasons, are involved in more than half of the cases. The liver (30% to 50%), brain (20%), and bone (20%) are other favored sites of metastases.

Secondary Pathology. Lung carcinomas have local effects that may cause several pathologic changes in the lung distal to the point of bronchial involvement. Partial obstruction may cause marked **focal emphysema;** total obstruction may lead to **atelectasis.** The impaired drainage of the airways is a common cause for **severe suppurative or ulcerative bronchitis** or **bronchiectasis. Pulmonary abscesses** sometimes call attention to an otherwise silent carcinoma. Compression or invasion of the superior vena cava can cause venous congestion and edema of the head and arm and, ultimately, circulatory compromise—the **superior vena cava syndrome.** Extension to the pericardial or pleural sacs may cause **pericarditis** (Chapter 12) or **pleuritis** with significant effusions. Apical lung cancers in the superior pulmonary sulcus tend to invade the neural structures around the trachea, including the cervical sympathetic plexus, and produce a group of clinical findings that includes severe pain in the distribution of the ulnar nerve and *Horner syndrome* (enophthalmos, ptosis, miosis, and anhidrosis) on the same side as the lesion. Such tumors are also referred to as *Pancoast tumors*.

Staging. A uniform TNM system for staging cancer according to its anatomic extent at the time of diagnosis is useful, particularly for comparing treatment results from different centers (Table 15.10).

Table 15.10 International Staging System for Lung Cancer

TNM Staging	
Tis	Carcinoma in situ Adenocarcinoma in situ: adenocarcinoma with pure lepidic pattern, ≤3 cm Squamous cell carcinoma in situ
T1	Tumor ≤3 cm without pleural or mainstem bronchus involvement (T1mi, minimally invasive adenocarcinoma; T1a, <1 cm; T1b, 1–2 cm; T1c, 2–3 cm)
T2	Tumor 3–5 cm or involvement of mainstem bronchus but not of carina, visceral pleural involvement, or lobar atelectasis (T2a, 3–4 cm; T2b, 4–5 cm)
T3	Tumor >5–7 cm or one with involvement of parietal pleura, chest wall (including superior sulcus tumors), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium, or separate tumor nodules in the same lobe
T4	Tumor >7 cm or any tumor with invasion of mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, or carina, or separate tumor nodules in a different ipsilateral lobe
N0	No metastasis to regional lymph nodes
N1	Ipsilateral intraparenchymal or peribronchial or hilar nodal involvement
N2	Metastasis to ipsilateral mediastinal or subcarinal lymph nodes
N3	Metastasis to contralateral mediastinal or hilar lymph nodes, ipsilateral or contralateral scalene, or supraclavicular lymph nodes
M0	No distant metastasis
M1	Distant metastasis (M1a, separate tumor nodule in contralateral lobe or pleural nodules or malignant pleural or pericardial effusion; M1b, single extrathoracic metastasis in a single organ; M1c, multiple extrathoracic metastases)

Stage Grouping			
Stage 0	Tis	N0	M0
Stage IA	IA1, T1mi or T1a; IA2, T1b; IA3, T1c	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0 N1	M0 M0
Stage IIB	T2b T1a, T1b, T1c, T2a, T2b T3	N0 N1 N0	M0 M0 M0
Stage IIIA	T1a, T1b, T1c, T2a, T2b T3 T4	N2 N1 N0, N1	M0 M0 M0
Stage IIIB	T1a, T1b, T1c, T2a, T2b T3, T4	N3 N2	M0 M0
Stage IIIC	T3, T4	N3	M0
Stage IVA	T any	N any	M1a, M1b
Stage IVB	T any	N any	M1c

Clinical Features

Lung cancer is one of the most insidious and aggressive neoplasms in the realm of oncology. In the usual case it is discovered in patients in their 50s or older whose symptoms are of several months' duration. The major presenting complaints are cough (75%), weight loss (40%), chest pain (40%), and dyspnea (20%). Some of the more common local manifestations of lung cancer and their pathologic bases are listed in Table 15.11.

Not infrequently, lung cancer is recognized though biopsy of tissues involved by metastatic disease. Symptoms of metastases depend on the site, for example, back pain in bone metastases and headache, hemiparesis, cranial nerve damage, and seizures in brain metastases.

The best "treatment" for lung cancer is smoking prevention, which has lowered lung cancer incidence in the United States among men; however, 15% of adults still smoke, and even those who quit remain at elevated risk for an extended period of time. This reality has led to early detection trials in high-risk individuals using low-dose computed tomography, which is capable of detecting some early (resectable) non–small cell lung cancers, but at a cost of a high incidence of false-positive (noncancer) findings. Overall, the outlook is poor for most patients. Even with incremental improvements in thoracic surgery, radiation therapy, and chemotherapy, the overall 5-year survival rate is only 18.7%. The 5-year survival rate is 52% for cases detected when the disease is still localized, 22% when there is regional metastasis, and only 4% with distant metastases. In general, adenocarcinoma and squamous cell carcinoma tend to remain localized longer and have a slightly better prognosis than small cell carcinoma, which is usually advanced by the time it is discovered.

Table 15.11 Local Effects of Lung Tumor Spread

Clinical Feature	Pathologic Basis
Cough (50%–75%)	Involvement of central airways
Hemoptysis (25%–50%)	Hemorrhage from tumor in airway
Chest pain (20%)	Extension of tumor into mediastinum, pleura, or chest wall
Pneumonia, abscess, lobar collapse	Airway obstruction by tumor
Lipoid pneumonia	Tumor obstruction; accumulation of cellular lipid in foamy macrophages
Pleural effusion	Tumor spread into pleura
Hoarseness	Recurrent laryngeal nerve invasion
Dysphagia	Esophageal invasion
Diaphragm paralysis	Phrenic nerve invasion
Rib destruction	Chest wall invasion
SVC syndrome	SVC compression by tumor
Horner syndrome	Sympathetic ganglia invasion
Pericarditis, tamponade	Pericardial involvement

SVC, Superior vena cava.

Treatment of patients with adenocarcinoma and activating mutations in *EGFR* (present in about 15% of all cases) or in other tyrosine kinases with specific kinase inhibitors prolongs survival. Many tumors that recur carry new mutations that generate resistance to these inhibitors, proving that these drugs are "hitting" their target. In contrast, activating *KRAS* mutations (present in approximately 30% of cases of adenocarcinoma) appear to be associated with a worse prognosis, regardless of treatment, in an already grim disease. Because of the mutagenic effects of carcinogens in tobacco smoke, lung cancers have a high burden of potentially antigenic neoantigens. Accordingly, both adenocarcinoma and squamous cell carcinoma respond in subsets of cases to checkpoint inhibitor therapy, which has produced improvements in survival and is now approved for use.

Small cell carcinoma is quite sensitive to radiation therapy and chemotherapy, and approximately 10% of patients with limited disease survive for 5 years and may be cured. Unfortunately, however, most patients present with advanced stage disease; for these patients, despite excellent initial responses to chemotherapy, the median survival is approximately 10 months and the cure rate is close to zero. New approaches involving use of antibody-drug conjugates that deliver chemotherapy selectively to tumor cells and immune checkpoint inhibitors are being tested.

Paraneoplastic Syndromes. Lung carcinoma can be associated with several paraneoplastic syndromes (Chapter 7), some of which may antedate the development of a detectable pulmonary lesion. The hormones or hormone-like factors elaborated by lung cancer cells and associated syndromes include:

- *Antidiuretic hormone* (ADH), inducing hyponatremia due to inappropriate ADH secretion
- *Adrenocorticotropic hormone* (ACTH), producing Cushing syndrome
- *Parathormone, parathyroid hormone-related peptide, prostaglandin E, and some cytokines,* all implicated in the hypercalcemia often seen with lung cancer
- *Calcitonin,* causing hypocalcemia
- *Gonadotropins,* causing gynecomastia
- *Serotonin and bradykinin,* associated with the carcinoid syndrome

The incidence of clinically significant paraneoplastic syndromes related to these factors in lung cancer patients ranges from 1% to 10%, although a much higher proportion of patients show elevated serum levels of these (and other) peptide hormones. Any histologic type of tumor may occasionally produce any one of the hormones, but tumors that produce ACTH and ADH are predominantly small cell carcinomas, whereas those that produce hypercalcemia are mostly squamous cell carcinomas.

Other systemic manifestations of lung carcinoma include the *Lambert-Eaton myasthenic syndrome* (Chapter 27), in which muscle weakness is caused by autoantibodies (possibly elicited by tumor ionic channels) directed to the neuronal calcium channel; *peripheral neuropathy,* usually purely sensory; dermatologic abnormalities, including *acanthosis nigricans* (Chapter 25); hematologic abnormalities, such as *leukemoid reactions;* hypercoagulable states, such as *Trousseau syndrome* (deep vein thrombosis and thromboembolism); and finally, a peculiar abnormality of connective tissue called *hypertrophic pulmonary osteoarthropathy,* associated with clubbing of the fingers.

KEY CONCEPTS

CARCINOMAS OF THE LUNG

- The three major histologic subtypes are adenocarcinoma (most common), squamous cell carcinoma, and small cell carcinoma.
- Each of these is clinically and genetically distinct. Small cell lung carcinomas are best treated by chemotherapy because almost all are metastatic at presentation. The other carcinomas may be curable by surgery if limited to the lung. Combination chemotherapy also is available along with tyrosine kinase inhibitors for those with *EGFR*, *ALK*, *ROS*, and *MET* mutations.
- Smoking is the most important risk factor for lung cancer; the most common subtype related to smoking in men and women is adenocarcinoma. Adenocarcinoma also is the most common subtype in non-smokers.
- Precursor lesions include atypical adenomatous hyperplasia and adenocarinoma in situ (formerly bronchioloalveolar carcinoma) for adenocarcinomas and squamous dysplasia for squamous cell carcinoma.
- Tumors 3 cm or less in diameter characterized by pure growth along pre-existing structures (lepidic pattern) without stromal invasion are now called adenocarcinoma in situ.
- Lung cancers, particularly small cell lung carcinomas, often cause paraneoplastic syndromes.

Neuroendocrine Proliferations and Tumors

The normal lung contains neuroendocrine cells within the epithelium as single cells or as clusters, the neuroepithelial bodies. Virtually all pulmonary neuroendocrine cell hyperplasias are secondary to airway fibrosis and/or inflammation. The exception is a rare disorder called *diffuse idiopathic*

pulmonary neuroendocrine cell hyperplasia, in which hyperplasia occurs in the absence of an inflammatory stimulus.

Neoplasms of neuroendocrine cells in the lung include benign *tumorlets,* small, inconsequential, hyperplastic nests of neuroendocrine cells seen in areas of scarring or chronic inflammation; *carcinoids;* and the (already discussed) highly aggressive small cell carcinoma and large cell neuroendocrine carcinoma of the lung. Carcinoid tumors are classified separately, since they differ significantly from carcinomas with evidence of neuroendocrine differentiation in terms of incidence and clinical, epidemiologic, histologic, and molecular characteristics. For example, in contrast to small cell and large cell neuroendocrine carcinomas, carcinoids may occur in patients with multiple endocrine neoplasia type 1.

Carcinoid Tumors

Carcinoid tumors represent 1% to 5% of all lung tumors. Most patients with these tumors are younger than 60 years of age, and the incidence is equal for both sexes. Approximately 20% to 40% of patients are nonsmokers. Carcinoid tumors are low-grade malignant epithelial neoplasms that are subclassified into *typical* and *atypical carcinoids*.

MORPHOLOGY

Carcinoids may arise centrally or may be peripheral. On gross examination, the central tumors grow as finger-like or spherical polypoid masses that commonly project into the lumen of the bronchus and are usually covered by an intact mucosa (Fig. 15.47A). They rarely exceed 3 to 4 cm in diameter. Most are confined to the mainstem bronchi. Others, however, penetrate the bronchial wall to fan out in the peribronchial tissue, producing the so-called **collar-button lesion.** Peripheral tumors are solid and nodular.

Histologically, the tumor is composed of organoid, trabecular, palisading, ribbon, or rosette-like arrangements of cells separated by a delicate fibrovascular stroma. In common with the lesions of the gastrointestinal tract, the individual cells are quite regular and have uniform round nuclei and a moderate amount of eosinophilic cytoplasm (Fig. 15.47B). Typical carcinoids have fewer than two mitoses per 10 high-power fields and lack necrosis, while atypical carcinoids have between two and 10 mitoses per 10 high-power fields and/or foci of necrosis. Atypical carcinoids also show increased pleomorphism, have more prominent nucleoli, and are more likely to grow in a disorganized fashion and invade lymphatics. On electron microscopy the cells exhibit the dense-core granules characteristic of other neuroendocrine tumors and, by immunohistochemistry, are found to contain serotonin, neuron-specific enolase, bombesin, calcitonin, or other peptides.

Clinical Features

The clinical manifestations of bronchial carcinoids emanate from their intraluminal growth, their capacity to metastasize, and the ability of some of the lesions to elaborate vasoactive amines. Persistent cough, hemoptysis, impairment of drainage of respiratory passages with secondary infections, bronchiectasis, emphysema, and atelectasis are all by-products of the intraluminal growth of these lesions.

Most interesting are functioning lesions capable of producing the classic *carcinoid syndrome,* characterized by intermittent attacks of diarrhea, flushing, and cyanosis. Approximately, 10% of bronchial carcinoids give rise to this syndrome. Overall, most bronchial carcinoids do not have secretory activity and do not metastasize to distant sites but follow a relatively benign course for long periods and are therefore amenable to resection. The reported 5-year survival rates are 95% for typical carcinoids, 70% for atypical carcinoids, 30% for large cell neuroendocrine carcinoma, and 5% for small cell carcinoma, respectively.

Miscellaneous Tumors

Benign and malignant mesenchymal tumors, such as inflammatory myofibroblastic tumor, fibroma, fibrosarcoma, lymphangioleiomyomatosis, leiomyoma, leiomyosarcoma, lipoma, hemangioma, and chondroma, may occur in the lung but are rare. Hematolymphoid tumors similar to those described in other organs, may also affect the lung, either as isolated lesions or, more commonly, as part of a generalized disorder. These include Langerhans cell histiocytosis,

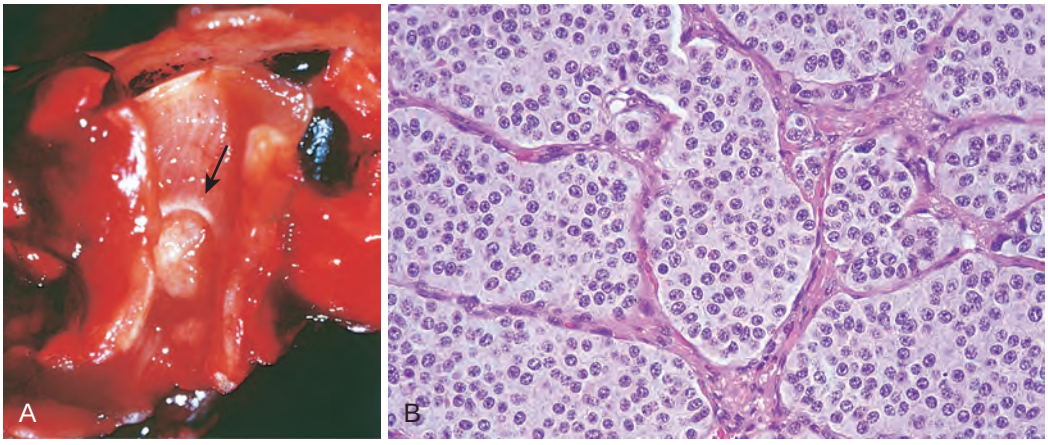


Figure 15.47 Bronchial carcinoid. (A) Carcinoid growing as a spherical mass *(arrow)* protruding into the lumen of the bronchus. (B) The tumor cells have small, rounded, uniform nuclei and moderate amounts of cytoplasm. (Courtesy Dr. Thomas Krausz, Department of Pathology, The University of Chicago, Pritzker School of Medicine, Chicago, Ill.)

Hodgkin lymphomas, lymphomatoid granulomatosis, an unusual EBV-positive B-cell lymphoma, and low-grade extranodal marginal zone B-cell lymphoma (Chapter 13).

Pulmonary *hamartoma* is a relatively common lesion that is usually discovered as an incidental, rounded radio-opacity *(coin lesion)* on a routine chest film. Most are solitary, less than 3 to 4 cm in diameter, and well circumscribed. Pulmonary hamartoma consists of nodules of connective tissue intersected by epithelial clefts. Cartilage is the most common connective tissue, but there may also be fibrous tissue and fat. The clefts are lined by ciliated or nonciliated epithelium (Fig. 15.48). The traditional term hamartoma is retained for this lesion, but it is in fact a clonal neoplasm associated with chromosomal aberrations involving either 6p21 or 12q14-q15. These aberrations are found in the mesenchymal component, while the epithelial component appears to represent entrapped respiratory epithelium.

Lymphangioleiomyomatosis is a pulmonary disorder that primarily affects young women of childbearing age. It is characterized by a proliferation of perivascular epithelioid cells that express markers of both melanocytes and smooth muscle cells. The proliferation distorts the involved lung, leading to cystic, emphysema-like dilation of terminal airspaces, thickening of the interstitium, and obstruction of lymphatic vessels. The lesional epithelioid cells frequently harbor loss-of-function mutations in the tumor suppressor *TSC2*, one of the loci linked to tuberous sclerosis (Chapter 28). The protein encoded by *TSC2,* tuberin, is a negative regulator of mammalian target of rapamycin (mTOR), a key regulator of cellular metabolism. While *TSC2* mutations point to increased mTOR activity as a pathogenic factor, the disorder remains poorly understood. The strong tendency to affect young women suggests that estrogen contributes to the proliferation of perivascular epithelioid cells, which often express estrogen receptors. Patients most commonly present with dyspnea or spontaneous pneumothorax, the latter related to emphysematous changes. The disease tends to be slowly progressive over a period of several decades. mTOR inhibitors slow or prevent the deterioration of lung function, but must be continued indefinitely. Only lung transplantation is curative.

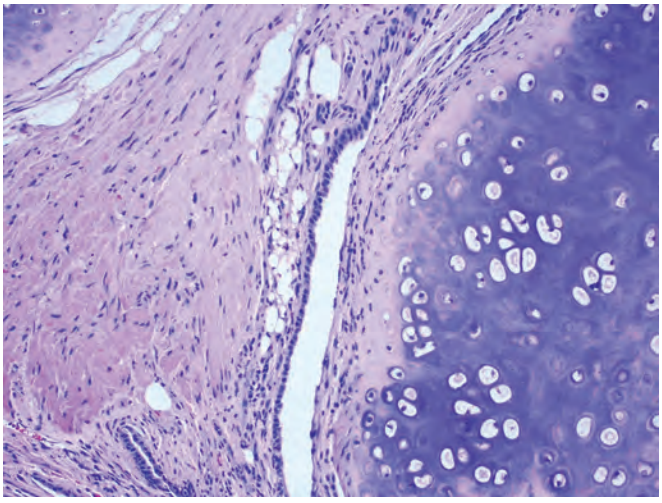


Figure 15.48 Pulmonary hamartoma. There are islands of cartilage, fat, smooth muscle, and entrapped respiratory epithelium.

Table 15.12 Mediastinal Neoplasms and Other Masses

Anterior Mediastinum
Thymoma
Teratoma
Lymphoma
Thyroid lesions
Parathyroid tumors
Metastatic carcinoma
Posterior Mediastinum
Neurogenic tumors (schwannoma, neurofibroma)
Lymphoma
Metastatic tumor (most are from the lung)
Bronchogenic cyst
Gastroenteric hernia
Middle Mediastinum
Bronchogenic cyst
Pericardial cyst
Lymphoma

Inflammatory myofibroblastic tumor is a rare tumor that is more common in children, with an equal male-to-female ratio. Presenting symptoms include fever, cough, chest pain, and hemoptysis. It may also be asymptomatic. Imaging studies usually show a round, well-defined, peripheral mass. Calcification is present in about a quarter of cases. Grossly, the lesion is firm, 3 to 10 cm in diameter, and grayish white. Microscopically, there is proliferation of spindle-shaped fibroblasts and myofibroblasts, lymphocytes, plasma cells, and peripheral fibrosis. Some of these tumors have activating rearrangements of the *ALK* receptor tyrosine kinase gene, located on 2p23, and treatment with ALK inhibitors have produced sustained responses in such cases.

Tumors in the mediastinum may arise in mediastinal structures or may be metastatic from the lung or other organs. They often invade or compress the lungs. Table 15.12 lists the most common tumors in the various compartments of the mediastinum. Specific tumor types are discussed in appropriate sections of this book.

Metastatic Tumors

The lung is the most common site of metastatic neoplasms. Both carcinomas and sarcomas arising anywhere in the body may spread to the lungs via the blood or lymphatics or by direct continuity. Esophageal carcinoma and mediastinal lymphoma may also invade the lung by direct extension.

MORPHOLOGY

The pattern of metastatic spread to the lungs is quite variable. In the usual case, multiple discrete nodules (cannonball lesions) are scattered throughout all lobes, particularly in the lung periphery (Fig. 15.49). Other times spread takes the form of a solitary nodule, endobronchial or pleural involvement, pneumonic consolidation, and/or some combination thereof. Foci of lepidic growth similar to adenocarcinoma in situ are seen occasionally with metastatic carcinomas and may be associated with any of the listed patterns.

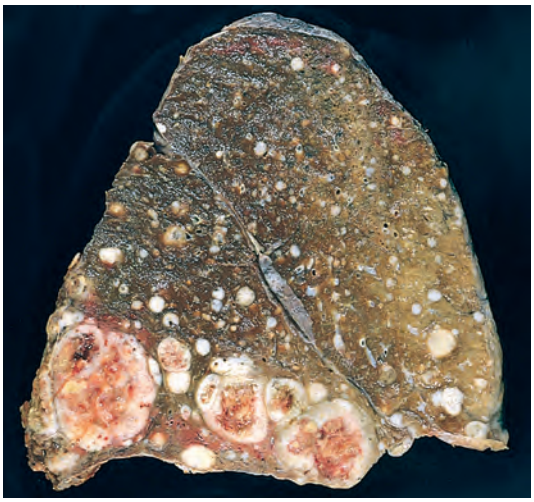


Figure 15.49 Numerous metastases to lung from a renal cell carcinoma. (Courtesy Dr. Michelle Mantel, Brigham and Women's Hospital, Boston, Mass.)

PLEURA

Most pleural disorders stem from complications of disease arising elsewhere in the body. Secondary infections and inflammations are particularly common findings at autopsy. Important primary disorders include (1) intrapleural bacterial infections, presumably the product of seeding from transient bacteremia; and (2) mesothelioma, a primary pleural neoplasm (discussed later).

Pleural Effusion

Pleural effusion is a common manifestation of both primary and secondary pleural diseases, which may be inflammatory or noninflammatory. Normally, no more than 15 mL of serous, relatively acellular, clear fluid lubricates the pleural surface. Accumulation of pleural fluid occurs in the following settings:

- Increased hydrostatic pressure, as in congestive heart failure
- Increased vascular permeability, as in pneumonia
- Decreased osmotic pressure, as in nephrotic syndrome
- Increased intrapleural negative pressure, as in atelectasis
- Decreased lymphatic drainage, as in mediastinal carcinomatosis

Inflammatory Pleural Effusions

Serous, serofibrinous, and fibrinous pleuritis all have an inflammatory basis, differing only in the intensity and duration of the process. The most common causes of pleuritis are disorders associated with inflammation of the underlying lung, such as tuberculosis, pneumonia, lung infarction, lung abscess, and bronchiectasis. Rheumatoid arthritis, systemic lupus erythematosus, uremia, diffuse systemic infections, and metastatic involvement of the pleura can also cause serous or serofibrinous pleuritis. Radiotherapy for tumors in the lung or mediastinum often causes a serofibrinous pleuritis. In most of these disorders, the pleural reaction is minimal, and the fluid exudate is resorbed with either resolution or organization of the fibrinous component. However, large amounts of fluid sometimes accumulate and compress the lung, causing respiratory distress.

A purulent pleural exudate *(empyema)* usually results from bacterial or mycotic seeding of the pleural space. Most commonly, this seeding occurs by contiguous spread of organisms from intrapulmonary infection, but occasionally it occurs through lymphatic or hematogenous dissemination from a more distant source. Rarely, infections below the diaphragm, such as the subdiaphragmatic or liver abscess, may extend by continuity through the diaphragm into the pleural spaces, more often on the right side.

Empyema is characterized by loculated, yellow-green, creamy pus composed of masses of neutrophils admixed with other leukocytes. Although empyema may accumulate in large volumes (up to 500 to 1000 mL), usually the volume is small, and the pus becomes walled off by fibrosis. Empyema may resolve, but more often the exudate organizes into dense, tough fibrous adhesions that frequently obliterate the pleural space or envelop the lungs; either can seriously restrict pulmonary expansion.

Hemorrhagic pleuritis manifested by sanguineous inflammatory exudates is infrequent and is most often associated with hemorrhagic diatheses, rickettsial infections, and neoplastic involvement of the pleural cavity. The sanguineous exudate must be differentiated from hemothorax (discussed later). When hemorrhagic pleuritis is encountered, a careful search should be made for the presence of tumor cells.

Noninflammatory Pleural Effusions

Noninflammatory collections of serous fluid within the pleural cavities are called *hydrothorax*. The fluid is clear and straw colored. Hydrothorax may be unilateral or bilateral, depending on the underlying cause. The most common cause of hydrothorax is heart failure, and for this reason it is usually accompanied by pulmonary congestion and edema. Hydrothorax may also be seen in any other systemic disease associated with generalized edema, such as patients with renal failure or cirrhosis of the liver.

The escape of blood into the pleural cavity is known as *hemothorax*. It is most commonly a complication of trauma or less commonly surgery, but can also accompany rupture of an aortic aneurysm, a setting in which it is almost invariably fatal.

Chylothorax is an accumulation of milky fluid, usually of lymphatic origin, in the pleural cavity. Chyle is milky white because it contains finely emulsified fats. Chylothorax is most often caused by thoracic duct trauma or by obstruction of a major lymphatic duct, usually by a malignancy. Such cancers most commonly arise within the thoracic cavity and invade the lymphatics locally, but occasionally more distant cancers metastasize via the lymphatics and grow within the right lymphatic or thoracic duct, producing obstruction.

Pneumothorax

Pneumothorax refers to air or gas in the pleural cavities and is most commonly associated with emphysema, asthma,

and tuberculosis. It may be spontaneous, traumatic, or therapeutic. Spontaneous pneumothorax may complicate any form of pulmonary disease that causes emphysematous changes. An abscess cavity that communicates either directly with the pleural space or with the lung interstitial tissue may also lead to the escape of air. In the latter circumstance the air may dissect through the lung substance or back through the mediastinum (interstitial emphysema), eventually entering the pleural cavity. Traumatic pneumothorax is usually caused by some perforating injury to the chest wall, but sometimes the trauma pierces the lung and thus provides two avenues for the accumulation of air within the pleural spaces. Resorption of the air in the pleural space occurs in spontaneous and traumatic pneumothorax, provided that the original communication seals itself.

Of the various forms of pneumothorax, the one that attracts the most clinical attention is *spontaneous idiopathic pneumothorax*. This entity is encountered in relatively young people; seems to be due to rupture of small, peripheral, usually apical subpleural blebs; and usually subsides spontaneously as the air is resorbed. Recurrent attacks are common and can be quite disabling.

Pneumothorax may cause marked respiratory distress due to collapse and atelectasis of the lung. In some instances the pleural defect acts as a flap valve and permits the entrance of air during inspiration but fails to permit its escape during expiration. The result is called a *tension pneumothorax,* in which progressively increasing intrapleural pressure may compress vital mediastinal structures and the contralateral lung.

Pleural Tumors

The pleura may be involved by primary or secondary tumors. Secondary metastatic involvement is far more common than are primary tumors. The most frequent metastatic malignancies arise from primary neoplasms of the lung and breast. In addition to these cancers, malignancy from any organ of the body may spread to the pleural spaces. Ovarian carcinomas, for example, tend to cause widespread implants in both the abdominal and thoracic cavities. Most metastatic implants produce a serous or serosanguineous effusion that often contains neoplastic cells. For this reason, careful cytologic examination of the sediment is of considerable diagnostic value.

Solitary Fibrous Tumor

Solitary fibrous tumor is a soft tissue tumor with a propensity to occur in the pleura and, less commonly, in the lung, as well as other sites. The tumor is often attached to the pleural surface by a pedicle. It may be small (1 to 2 cm in diameter) or may reach an enormous size, but it tends to remain confined to the surface of the lung (Fig. 15.50).

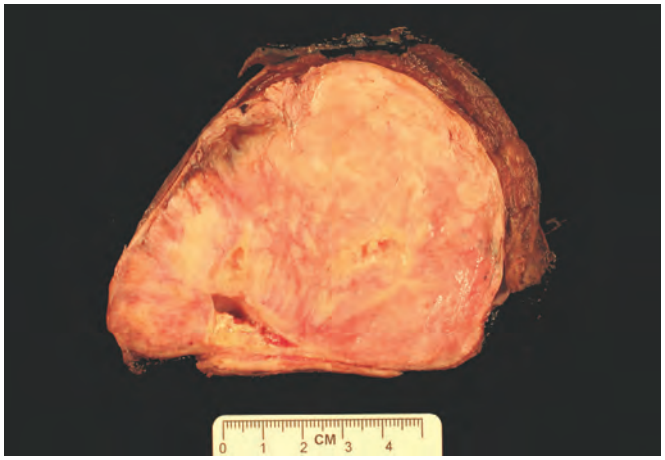


Figure 15.50 Solitary fibrous tumor. Cut surface is solid with a whorled appearance. (Courtesy Dr. Justine A. Barletta, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

MORPHOLOGY

Grossly, solitary fibrous tumor consists of dense fibrous tissue with occasional cysts filled with viscid fluid. Microscopically, the tumor shows whorls of reticulin and collagen fibers among which are interspersed spindle cells resembling fibroblasts. Rarely, this tumor may be malignant, marked by pleomorphism, mitotic activity, necrosis, and large size (>10 cm). The tumor cells are positive for CD34 and STAT6 and negative for keratin by immunostaining, features that help to distinguish this lesion from malignant mesothelioma (which has the opposite phenotype). The solitary fibrous tumor has no relationship to asbestos exposure.

Solitary fibrous tumor is highly associated with a cryptic inversion of chromosome 12 involving the genes *NAB2* and *STAT6*. This rearrangement creates a *NAB2-STAT6* fusion gene that appears to be virtually unique to solitary fibrous tumor. It encodes a chimeric transcription factor that is hypothesized to be a key driver of tumor development.

Malignant Mesothelioma

Malignant mesothelioma, although rare, has assumed great importance in the past few decades because of its increased incidence among people with heavy exposure to asbestos (see Pneumoconioses). Thoracic mesothelioma arises from either the visceral or the parietal pleura. In coastal areas with shipping industries in the United States and Great Britain, as well as in Canadian, Australian, and South African mining areas, as many as 90% of mesotheliomas are asbestos-related. The lifetime risk of developing mesothelioma in heavily exposed individuals is as high as 7% to 10%. There is a long latent period of 25 to 45 years for the development of asbestos-related mesothelioma, and there seems to be no increased risk of mesothelioma in asbestos workers who smoke. This is in contrast to the risk of asbestos-related lung carcinoma, which is markedly magnified by smoking. Thus, asbestos workers (particularly those who smoke) are at much higher risk of dying of lung carcinoma than mesothelioma.

Asbestos bodies (see Fig. 15.20) are found in increased numbers in the lungs of patients with mesothelioma. Another marker of asbestos exposure, the *asbestos plaque,* has been previously discussed (see Fig. 15.21).

Although several cytogenetic abnormalities have been detected, the most common is homozygous deletion of chromosome 9p leading to loss of the tumor suppressor gene *CDKN2A*, which occurs in about 80% of mesotheliomas. Sequencing of mesothelioma genomes has shown that driver mutations are also common in the *NF2* (neurofibromatosis-2) gene, which encodes a cell signaling regulator; and *BAP1*, which encodes a protein that interacts with the BRCA1 tumor suppressor and appears to function as a chromatin regulator. Of note, individuals with germline mutations in *BAP1* have a markedly elevated risk of mesothelioma, further implicating this tumor suppressor gene in the pathogenesis of the disease.

MORPHOLOGY

Malignant mesothelioma is a diffuse lesion arising from either the visceral or the parietal pleura that spreads widely in the pleural space and is usually associated with extensive pleural effusion and direct invasion of thoracic structures. The affected lung becomes ensheathed by a thick layer of soft, gelatinous, grayish pink tumor tissue (Fig. 15.51).

Microscopically, malignant mesothelioma may be epithelioid (60% to 80%), sarcomatoid (10% to 12%), or biphasic (10% to 15%). This is in keeping with the fact that mesothelial cells have the potential to develop as epithelium-like cells or mesenchymal stromal cells.

The **epithelioid type** of mesothelioma consists of cuboidal, columnar, or flattened cells forming tubular or papillary structures resembling adenocarcinoma (Fig. 15.52A). Immunohistochemical stains are very helpful in differentiating it from pulmonary adenocarcinoma. Most mesotheliomas show strong positivity for keratin proteins, calretinin (Fig. 15.52B), Wilms tumor 1 (WT-1), cytokeratin 5/6, and podoplanin, and unlike adenocarcinomas are negative for Claudin4. This panel of antibodies is diagnostic in a majority of cases when interpreted in the context of morphology and clinical presentation. The mesenchymal type of mesothelioma **(sarcomatoid type)** has an appearance resembling fibrosarcoma.

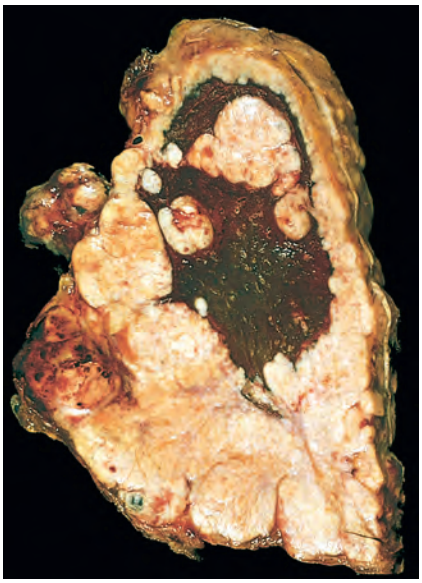


Figure 15.51 Malignant mesothelioma. Note the thick, firm, white pleural tumor tissue that ensheathes the lung.

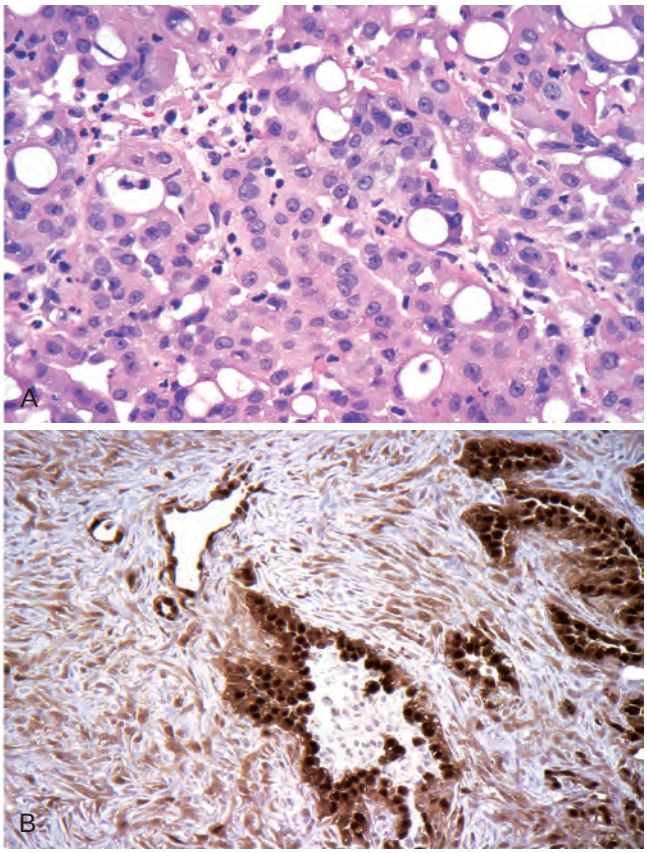


Figure 15.52 Histologic variants of malignant mesothelioma. (A) Epithelioid type. (B) Mixed type, stained for calretinin (immunoperoxidase method). The epithelial component is strongly positive (dark brown), while the sarcomatoid component is less so. (Courtesy Dr. Thomas Krausz, Department of Pathology, The University of Chicago, Pritzker School of Medicine, Chicago, Ill.)

Sarcomatoid mesotheliomas tend to have lower expression of many of the markers described previously, and some may be positive only for keratin. The **biphasic type** of mesothelioma contains both epithelioid and sarcomatoid patterns (see Fig. 15.52B).

Clinical Features

The presenting symptoms are chest pain, dyspnea, and, as noted, recurrent pleural effusions. Concurrent pulmonary asbestosis (fibrosis) is present in only 20% of individuals with pleural mesothelioma. The lung is invaded directly, and there is often metastatic spread to the hilar lymph nodes and, eventually, to the liver and other distant organs. Fifty percent of patients die within 12 months of diagnosis, and few survive longer than 2 years. Aggressive therapy (extrapleural pneumonectomy, chemotherapy, radiation therapy) seems to improve this poor prognosis in some patients.

Mesotheliomas also arise in the peritoneum, pericardium, tunica vaginalis, and genital tract (benign adenomatoid tumor) (see Chapter 21). *Peritoneal mesotheliomas* are related to heavy asbestos exposure in 60% of male patients (the number is much lower in females). Although in about half of cases the disease remains confined to the abdominal cavity, intestinal involvement frequently leads to death from intestinal obstruction or inanition.

SUGGESTED READINGS

Acute Lung Injury

Matthay MA et al: Acute respiratory distress syndrome, *Nat Rev Dis Primers* 5:18, 2019. *[This primer on adult and pediatric ARDS considers its epidemiology, mechanisms of lung injury, optimal approaches to diagnosis, management strategies, and quality of life after recovery from ARDS].*

Sharp C et al: Advances in understanding of the pathogenesis of acute respiratory distress syndrome, *Respiration* 89:420, 2015. *[Comprehensive review of pathogenic factors of acute respiratory distress syndrome and acute lung injury that underlie novel approaches being studied for treatment].*

Obstructive Pulmonary Diseases

Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management and Prevention of COPD*. 2019. Available from: http://goldcopd.org. *[Global strategy for the diagnosis, management and prevention of COPD which provides a non-biased review of the current evidence. A pocket guide is also available].*

Hogg JC et al: The contribution of small airway obstruction to the pathogenesis of chronic obstructive pulmonary disease, *Physiol Rev* 97:529, 2017. *[Comprehensive review of past and current understanding of small airways including anatomy, assessment methods, imaging, epithelial barrier and mucus secretion, and inflammation and remodeling].*

Rabe KF et al: Chronic obstructive pulmonary disease, *Lancet* 389:1931, 2017. *[An excellent review of epidemiology, causes, pathophysiology, and treatment of COPD].*

Emphysema

Barnes PJ: Cellular and molecular mechanisms of asthma and COPD, *Clin Science* 131:1541, 2017. *[Review of inflammatory cells, inflammatory mediators, and structural changes in asthma and chronic obstructive pulmonary disease and their implications for therapy].*

Boucherat O et al: Bridging lung development with chronic obstructive pulmonary disease. Relevance of developmental pathways in chronic obstructive pulmonary disease pathogenesis, *Am J Respir Crit Care Med* 193:362, 2016. *[Excellent outline of developmental insults and genes in the occurrence and pathogenesis of chronic obstructive pulmonary disease].*

Choudhury G et al: Role of inflammation and oxidative stress in the pathology of ageing in COPD: potential therapeutic interventions, *COPD* 14:122, 2017. *[Review discusses the evidence supporting a number of mechanisms including oxidative stress and aging in the pathogenesis of chronic obstructive pulmonary disease].*

Faner R et al: Multilevel, dynamic chronic obstructive pulmonary disease heterogeneity: a challenge for personalized medicine, *Ann Am Thorac Soc* 13(S5):S466, 2016. *[Succinct review of the various components of chronic obstructive pulmonary disease leading to the concept of treatable traits].*

Huang YJ et al: Understanding the role of the microbiome in chronic obstructive pulmonary disease: principles, challenges, and future directions, *Transl Res* 179:71, 2017. *[A comprehensive review of studies that examine the differences in the microbiome in health and COPD].*

Jones RL et al: Airway remodelling in COPD: it's not asthma!, *Respirology* 21:1347, 2016. *[A review summarizing studies of airway matrix deposition in COPD, which is different from that in asthma].*

Kligerman S et al: Clinical-radiologic-pathologic correlation of smoking-related diffuse parenchymal lung disease, *Radiol Clin North Am* 54:1047, 2016. *[Clinically oriented review of chronic obstructive pulmonary disease and other smoking-related lung injuries and how their pathologic manifestations lead to specific findings on CT].*

Chronic Bronchitis

Solomon GM et al: Therapeutic approaches to acquired cystic fibrosis transmembrane conductance regulator dysfunction in chronic bronchitis, *Ann Am Thorac Soc* 13(Suppl 2):S169, 2016. *[Review of the growing evidence for the role of* CFTR *gene dysfunction in patients with smoking-related chronic bronchitis, which has paved the way for* CFTR *modulating drugs].*

Asthma

Barnes PJ: Cellular and molecular mechanisms of asthma and COPD, *Clin Sci* 131:1541, 2017. *[Review comparing inflammatory and cellular mechanisms of asthma and chronic obstructive pulmonary disease and how the differences in inflammatory cells and mediators account for the differences in clinical manifestations and response to therapy].*

Global Initiative for Asthma. *Global strategy for asthma management and prevention*. 2019. Available from: www.ginasthma.org. *[An evidence-based educational resource for worldwide use for diagnosis, management and prevention of asthma. Pocket guide also available].*

Israel E, Reddel HK: Severe and difficult-to-treat asthma in adults, *N Engl J Med* 377:965, 2017. *[Review focuses on assessment, management, and pathobiologic features of the airway and new therapeutic agents that have been developed to target asthma].*

Leung JM et al: Asthma-COPD overlap syndrome: pathogenesis, clinical features, and therapeutic targets, *BMJ* 358:j3772, 2017. *[Review of asthma-COPD overlap syndrome as a clinically important phenotype with emphasis on factors that may lead to asthma in patients with chronic obstructive pulmonary disease and vice versa].*

Papi A et al: Asthma, *Lancet* 391:783, 2018. *[Clinically focused overview of asthma including epidemiology, pathophysiology, clinical diagnosis, asthma phenotypes and clinical management. Emerging therapies, controversies, and uncertainties in asthma management are also discussed].*

Bronchiectasis

Boyton RJ et al: Bronchiectasis: current concepts in pathogenesis, immunology, and microbiology, *Annu Rev Pathol* 11:523, 2016. *[Comprehensive review of diverse etiologies and how the interplay between immunogenetic susceptibility, immune dysregulation, and bacterial infection results in lung damage].*

Mirra V et al: Primary ciliary dyskinesia: an update on clinical aspects, genetics, diagnosis, and future treatment strategies, *Front Pediatr* 5:135, 2017. *[Comprehensive review of primary ciliary dyskinesia from its history to current understanding of the genetic causes and multidisciplinary approach to patient management].*

Muldoon EG et al: Allergic and noninvasive infectious pulmonary aspergillosis syndromes, *Clin Chest Med* 38:521, 2017. *[Article reviews noninvasive aspergillosis syndromes, focusing on clinical presentation, diagnosis, management, and potential complications].*

Idiopathic Pulmonary Fibrosis

Batra K et al: Pathology and radiology correlation of idiopathic interstitial pneumonias, *Hum Pathol* 72:1, 2018. *[Excellent review of the salient features of the idiopathic interstitial pneumonias including illustrative pathologic and radiologic images].*

Lederer DJ et al: Idiopathic pulmonary fibrosis, *N Engl J Med* 378:1811, 2018. *[Excellent review of the clinical features, pathobiologic features and risk factors, diagnosis, therapy, and complications of idiopathic pulmonary fibrosis].*

Nonspecific Interstitial Pneumonia

Wells AU et al: Nonspecific interstitial pneumonia: time to be more specific?, *Curr Opin Pulm Med* 22:450, 2016. *[Review of clinical, imaging, and serologic features supporting the concept that nonspecific interstitial pneumonia is a grouping of separate disorders with a common histologic pattern].*

Organizing Pneumonia

Torrealba JR et al: Pathology-radiology correlation of common and uncommon computed tomographic patterns of organizing pneumonia, *Hum Pathol* 71:30, 2018. *[Review of clinical, radiologic, and pathologic features of organizing pneumonia including differential diagnoses].*

Pneumoconiosis

Barmania S: Deadly denim: sandblasting-induced silicosis in the jeans industry, *Lancet Respir Med* 4:543, 2016. *[Spotlight on denim sandblasting, a relatively new workplace exposure, resulting in silicosis first recognized in Turkey and now in other countries].*

Cullinan P et al: Occupational lung diseases: from old and novel exposures to effective preventive strategies, *Lancet Respir Med* 5:445,

2017. *[Review summarizing the predominant lung diseases caused by a wide range of hazards in the workplace and possible avenues for their recognition and prevention]*.

Perret JL et al: Coal mine dust lung disease in the modern era, *Respirology* 22:662, 2017. *[Excellent review of epidemiology; prevention; molecular biomarkers; and clinical, pathologic, and radiologic features of coal workers' pneumoconiosis]*.

Radiation Pneumonitis

Bledsoe TJ et al: Radiation pneumonitis, *Clin Chest Med* 38:201, 2017. *[Review of pathophysiology, incidence, diagnosis, treatment, and prognosis of radiation pneumonitis]*.

Sarcoidosis

Esteves T et al: Is there any association between sarcoidosis and infectious agents? A systematic review and meta-analysis, *BMC Pulm Med* 16:165, 2016. *[Review of 58 case-control studies suggesting that some infectious agents can be associated with sarcoidosis and that more than one agent might be implicated in its pathogenesis]*.

O'Regan A, Berman JS: Sarcoidosis, *Ann Intern Med* 156:ITC5-1, 2012. *[Comprehensive review of pathogenesis, clinical manifestations, diagnosis, and prognosis of sarcoidosis]*.

Salamo O et al: Noncoding RNAs: new players in pulmonary medicine and sarcoidosis, *Am J Respir Cell Mol Biol* 58:147, 2018. *[Review provides insight into the role of noncoding RNAs in pathogenesis of pulmonary diseases in general and sarcoidosis in particular as well as their possible use as a diagnostic and prognostic tool]*.

Hypersensitivity Pneumonia

Miller R et al: Hypersensitivity pneumonitis: a perspective from members of the Pulmonary Pathology Society, *Arch Pathol Lab Med* 142:120, 2018. *[Review of clinical forms, pathogenesis, diagnostic criteria, radiology, pathology, and treatment of hypersensitivity pneumonitis]*.

Eosinophilic Lung Diseases

Cottin V: Eosinophilic lung diseases, *Clin Chest Med* 37:535, 2016. *[Review of the classification, clinical features, general diagnostic work-up, and management of eosinophilic lung diseases]*.

Smoking-Related Interstitial Diseases

Franks T, Galvin JR: Smoking-related "interstitial" lung disease, *Arch Pathol Lab Med* 139:974, 2015. *[Brief review of smoking-related diseases that emphasizes the fact that histologic findings in cigarette smokers are diverse]*.

Pulmonary Langerhans Cell Histiocytosis

Vassallo R et al: Current understanding and management of pulmonary Langerhans cell histiocytosis, *Thorax* 72:937, 2017. *[Review of epidemiology, pathogenesis, clinical features, pathology, and management of pulmonary Langerhans cell histiocytosis]*.

Pulmonary Alveolar Proteinosis

Kumar A et al: Pulmonary alveolar proteinosis in adults: pathophysiology and clinical approach, *Lancer Respir Med* 2018. [*ePub ahead of print*]. *[Review of the pathogenesis, classification, and treatment of pulmonary alveolar proteinosis]*.

Surfactant Dysfunction Disorders

Gupta A, Zheng SL: Genetic disorders of surfactant protein dysfunction: when to consider and how to investigate, *Arch Dis Child* 102:84, 2017. *[Review of the clinical features, pathogenesis, molecular genetics, pathology, and natural history of these rare disorders]*.

Pulmonary Hypertension

Thenappan T et al: Pulmonary arterial hypertension: pathogenesis and clinical management, *BMJ* 360:j5492, 2018. *[State-of-the-art review of epidemiology, pathogenesis, diagnosis, and treatment of pulmonary hypertension]*.

Pulmonary Hemorrhage Syndromes

Lally L, Spiera RF: Pulmonary vasculitis, *Rheum Dis Clin North Am* 41:315, 2015. *[Review of the most common inflammatory conditions involving pulmonary vessels]*.

Pneumonia

Kradin RL, Digumarthi S: The pathology of pulmonary bacterial infection, *Semin Diagn Pathol* 34:498, 2017. *[Excellent review of salient features of various bacterial pneumonias]*.

Pritt BS, Aubry MC: Histopathology of viral infections of the lung, *Semin Diagn Pathol* 34:510, 2017. *[Review of etiology, epidemiology, clinical presentation, and pathologic features of viral pneumonias]*.

Wunderink RG, Waterer G: Advances in the causes and management of community acquired pneumonia in adults, *BMJ* 358:j2471, 2017. *[Review of incidence, clinical diagnosis, causes, and optimal management of community-acquired pneumonia]*.

Lung Transplantation

Yusen RD et al: The registry of the International Society for Heart and Lung Transplantation: 32nd Adult Lung and Heart-Lung Transplantation Report 2015, *J Heart Lung Transplant* 34:1264, 2015. *[Annual report of registry data including numbers of transplants, indications, and outcomes]*.

Lung Cancer

Calvayrac O et al: Molecular biomarkers for lung adenocarcinoma, *Eur Respir J* 49:1601734, 2017. *[Review discusses the oncogenic drivers of lung adenocarcinoma, the associated epidemiologic features, and ways in which their identification can inform therapeutic strategies]*.

Dautzenberg B et al: Patients with lung cancer: are electronic cigarettes harmful or useful?, *Lung Cancer* 105:42, 2017. *[Review focuses on useful factual information about electronic cigarettes for professionals who deal with lung cancer patients]*.

Gazdar A et al: Small-cell lung cancer: what we know, what we need to know and the path forward, *Nat Rev Cancer* 17:725, 2017. *[Excellent review of small cell lung cancer including history of clinical and laboratory research, genetics, molecular pathogenesis, targeted therapy and immunotherapy, and mechanisms of resistance]*.

Hirsh V: New developments in the treatment of advanced squamous cell lung cancer: focus on afatinib, *Onco Targets Ther* 10:2513, 2017. *[Review summarizing the biologic basis for treating squamous cell carcinoma with checkpoint inhibitors, anti-VEGFR2 antibody, and ErbB-family blocker]*.

McIntyre A et al: Lung cancer—a global perspective, *J Surg Oncol* 115:550, 2017. *[Review summarizing global variations in incidence, etiology, genetics, screening, and patterns of treatment of lung cancer]*.

Melosky B et al: Breaking the biomarker code: PD-L1 expression and checkpoint inhibition in advanced NSCLC, *Cancer Treat Rev* 65:65, 2018. *[Review of six phase III clinical trials showing that checkpoint inhibitors represent a major advance in the treatment of advanced non–small cell lung cancer and that PD-L1 status can inform treatment decisions]*.

Sabari JK et al: Unravelling the biology of SCLC: implications for therapy, *Nat Rev Clin Oncol* 14:549, 2017. *[Review of advances in the understanding of the biology of small cell lung carcinoma that have led to the development of novel experimental therapies]*.

Thakrar R et al: Preinvasive disease of the airway, *Cancer Treat Rev* 58:77, 2017. *[Review describes the natural history of preinvasive disease of the airway, the current understanding of molecular changes, and treatment approaches]*.

Travis WD et al: The 2015 World Health Organization Classification of Lung Tumors: impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification, *J Thorac Oncol* 10:1243, 2015. *[Concise review highlighting updates in the most recent World Health Organization classification of lung tumors incorporating the most recent molecular and genetic knowledge]*.

Pleura

Galateau-Salle F et al: The 2015 World Health Organization Classification of Tumors of the Pleura: advances since the 2004 Classification, *J Thorac Oncol* 11:142, 2016. *[Concise review highlighting updates in the most recent World Health Organization classification of malignant mesothelioma]*.

Husain AN et al: Guidelines for pathologic diagnosis of malignant mesothelioma: 2017 update of the consensus statement from the International Mesothelioma Interest Group, *Arch Pathol Lab Med* 142:89, 2018. *[Practical guide for pathologic diagnosis of mesothelioma]*.

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 16

Head and Neck

Mark W. Lingen • Nicole A. Cipriani

CHAPTER CONTENTS

Diseases of the head and neck range from the common cold to uncommon neoplasms of the ear and nose. The examples discussed in the following sections are organized based on the primary anatomic site at which they occur, that is, oral cavity, upper airways (including the nose, pharynx, larynx, and nasal sinuses), ears, neck, and salivary glands.

Oral Cavity

Diseases of Teeth and Supporting Structures

Caries (Tooth Decay)

Dental caries is caused by focal demineralization of tooth structure (enamel and dentin) by acidic products of bacterial sugar fermentation. Caries is the principal cause of tooth loss before age 35. Caries was previously most common in industrialized countries with abundant, carbohydrate-rich processed and refined foods. Demographics have, however, shifted due to both reduced incidence in industrialized countries and increased incidence in developing nations. The former reflects improved oral hygiene and, in the United States, drinking water fluoridation. Within the crystalline tooth enamel, fluoroapatite (calcium fluorophosphate) is more resistant to acid degradation than hydroxylapatite (calcium apatite). Increasing processed food consumption is associated with greater caries rates in low income countries, where incidence is expected to continue to grow.

Gingivitis

Gingivitis is inflammation of the oral mucosa surrounding the teeth, caused by accumulation of dental plaque and calculus. It may occur at any age but is most prevalent and severe in adolescence. Dental plaque is a sticky, colorless biofilm that collects between and on the surface of the teeth because of poor oral hygiene. It contains a mixture of bacteria (which produce the acid that contributes to caries development), salivary proteins, and desquamated epithelial cells. If not removed, plaque can become mineralized to form calculus (tartar). Gingivitis is characterized by erythema, edema, bleeding, changes in contour, and loss of soft tissue

adaptation to the teeth. Fortunately, it is reversible; therapy is primarily aimed at reducing the accumulation of plaque and calculus via regular oral hygiene.

Periodontitis

Periodontitis is an inflammatory process that affects the supporting structures of the teeth (periodontal ligaments), alveolar bone, and cementum. Sequelae of periodontitis include destruction of the periodontal ligament, which attaches teeth to the alveolar bone. This leads to loosening and eventual tooth loss. Like caries, periodontitis is associated with poor oral hygiene but is also characterized by altered composition of the oral microbiome, which is believed to be important in pathogenesis. For the most part, facultative gram-positive organisms colonize healthy gingival sites; plaque within areas of active periodontitis contains anaerobic and microaerophilic gram-negative flora. Of the 300 bacterial species within the oral cavity, adult periodontitis is associated primarily with *Aggregatibacter (Actinobacillus) actinomycetemcomitans, Porphyromonas gingivalis,* and *Prevotella intermedia*.

Periodontal disease is most frequently localized, but can also be a component of systemic diseases, including acquired immunodeficiency syndrome (AIDS), leukemia, Crohn disease, diabetes, Down syndrome, sarcoidosis, and syndromes associated with neutrophil defects (Chédiak-Higashi syndrome, agranulocytosis, and cyclic neutropenia). Periodontal infections can serve as a site of origin for infective endocarditis and abscesses within the lungs and brain.

INFLAMMATORY/REACTIVE LESIONS

Aphthous Ulcers (Canker Sores)

Aphthous ulcers are common, often recurrent, and painful. The causes are not known, but oral aphthous ulcers affect up to 40% of the population, and are most frequent in the first two decades of life. Aphthous ulcers tend to be clustered within some families and may be associated with immunologic disorders including celiac disease, inflammatory bowel disease, and Behçet disease. The lesions appear as single or multiple, shallow, hyperemic, mucosal ulcerations covered by a thin exudate and rimmed by a narrow zone of erythema (Fig. 16.1). The underlying inflammatory infiltrate is initially mononuclear but becomes neutrophil-rich upon secondary bacterial infection. Lesions typically resolve spontaneously in 7 to 10 days but sometimes persist for weeks, particularly in immunocompromised patients.

Fibrous Proliferative Lesions

The ***irritation fibroma,*** also called *traumatic fibroma* or *focal fibrous hyperplasia,* is a submucosal nodular mass of fibrous connective tissue stroma that occurs primarily on the buccal mucosa along the bite line or the gingiva (Fig. 16.2). It is thought to be a reactive process induced by repetitive trauma. Treatment is complete surgical excision.

Pyogenic granuloma **is typically found on the gingiva of children, young adults, and pregnant women** *(pregnancy tumor).* This exophytic inflammatory lesion is red to purple in color and frequently ulcerated (Fig. 16.3). In some cases, rapid growth can be alarming and elicit concerns of malignancy. Histologically, pyogenic granulomas are a highly vascularized proliferation of organizing granulation tissue. Pyogenic granulomas can regress, mature into dense fibrous masses, or develop into a peripheral ossifying fibroma. Complete surgical excision is the definitive treatment for these lesions.

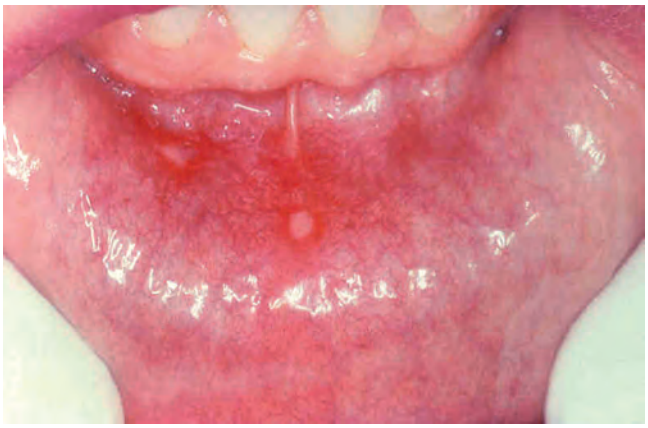


Figure 16.1 Aphthous ulcer. Single ulceration with an erythematous halo surrounding a yellowish fibrinopurulent membrane.

Peripheral ossifying fibroma is a common gingival growth that is most likely reactive in nature rather than neoplastic. As mentioned earlier, some arise from a long-standing pyogenic granuloma, and others develop *de novo* from cells of the periodontal ligament. Peripheral ossifying fibromas appear as red, ulcerated, nodular lesions of the gingiva. The peak incidence is in young females. Since the lesions have a recurrence rate of 15% to 20%, complete surgical excision down to the periosteum is required.

Peripheral giant cell granuloma is an uncommon oral cavity lesion that likely represents a reactive inflammatory process. It is generally covered by intact gingival mucosa, but may be ulcerated. Histologically, peripheral giant cell granulomas contain aggregates of multinucleate, foreign body–like giant cells separated by a fibroangiomatous stroma. Although not encapsulated, the lesions are usually well delimited and easily excised. They should be differentiated from central giant-cell tumors found within the jaws and from

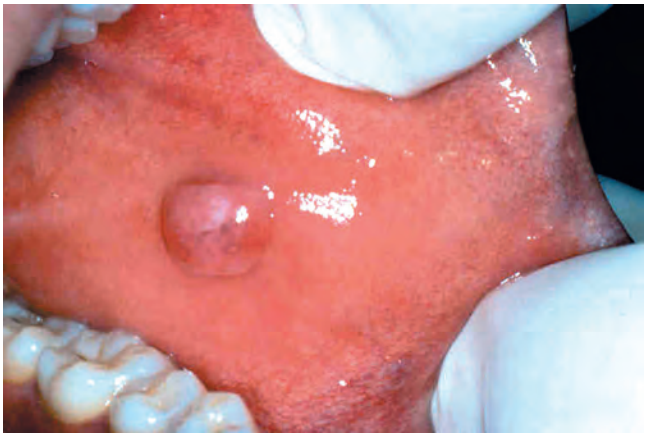


Figure 16.2 Irritation fibroma. Smooth pink exophytic nodule on the buccal mucosa.

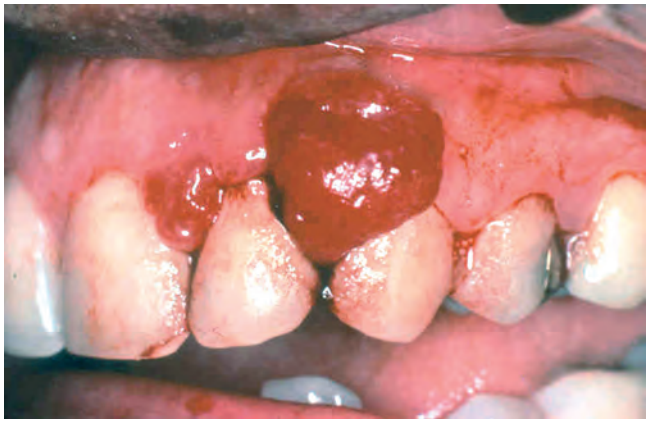


Figure 16.3 Pyogenic granuloma. Erythematous and hemorrhagic exophytic mass arising from the gingival mucosa.

the histologically similar but frequently multiple "brown tumors" seen in hyperparathyroidism (Chapter 24).

INFECTIONS

Herpes Simplex Virus Infections

Oral herpes usually presents as gingivostomatitis in children, pharyngitis in adults, and chronic mucocutaneous infection in immunocompromised individuals. Most orofacial herpetic infections are caused by herpes simplex virus type 1 (HSV-1), but oral HSV-2 (genital herpes) infections do occur. In children, primary infections are most common between 2 and 4 years of age. These are often asymptomatic but may present as acute herpetic gingivostomatitis, with abrupt onset of vesicles and ulcerations of the oral mucosa, in up to 20% of cases. These lesions can be accompanied by lymphadenopathy, fever, and anorexia. In adults, acute herpes pharyngitis is common and may recur.

MORPHOLOGY

Herpes vesicles range in size from a few millimeters to large bullae filled with clear, serous fluid. They rapidly rupture to become painful, red-rimmed, shallow ulcerations. Intracellular and intercellular edema and acantholysis create clefts that may become macroscopic vesicles. Individual epidermal cells at vesicle margins or free within the fluid may contain eosinophilic **intranuclear viral inclusions,** or several cells may fuse to produce giant cells **(multinucleate polykaryons).** These can be demonstrated by the diagnostic **Tzanck test** based on microscopic examination of the vesicle fluid. Vesicles and shallow ulcers usually clear spontaneously within 3 to 4 weeks, but the virus treks along regional nerves and eventually becomes dormant (i.e., latent) within local ganglia (e.g., the trigeminal ganglion).

Most individuals who have been infected harbor latent virus in epithelial cells or ganglia. Reactivation of latent HSV causes *recurrent herpetic stomatitis* and is associated with trauma, allergies, exposure to ultraviolet light, upper respiratory tract infection, pregnancy, menstruation, immunosuppression, and exposure to temperature extremes. In contrast to acute herpetic gingivostomatitis, recurrent herpetic stomatitis occurs at the site of primary inoculation or in mucosa associated with the same ganglion. The lesions appear as groups of small (1 to 3 mm) vesicles on the lips *(herpes labialis),* nasal orifices, buccal mucosa, gingiva, and hard palate. Although they typically resolve in 7 to 10 days, lesions can persist in immunocompromised patients and may require systemic antiviral therapy.

Other viral infections that can involve the oral cavity as well as the head and neck region include herpes zoster, Epstein-Barr virus (EBV; mononucleosis, nasopharyngeal carcinoma, lymphoma), cytomegalovirus, enterovirus (herpangina, hand-foot-and-mouth disease, acute lymphonodular pharyngitis), and rubeola (measles).

Oral Candidiasis (Thrush)

***Candida albicans* is a normal component of the oral flora in approximately 50% of the population and is the most common fungal infection of the oral cavity.** Factors that influence the likelihood of infection include the strain of *C. albicans*, oral microbiome composition, and the individual's immune status. Oral candidiasis can be *pseudomembranous, erythematous,* or *hyperplastic.* The pseudomembranous form, *thrush,* is characterized by a superficial, gray to white inflammatory membrane composed of matted organisms enmeshed in a fibrinosuppurative exudate that can be readily scraped off to reveal an underlying erythematous inflammatory base. The infection typically remains superficial except in the setting of immunosuppression, such as in individuals with organ or bone marrow transplants, neutropenia, chemotherapy-induced immunosuppression, AIDS, or diabetes. Broad-spectrum antibiotics that eliminate or alter the normal bacterial flora of the mouth can also promote thrush.

Deep Fungal Infections

In addition to their usual sites of infection, some deep fungal infections, including histoplasmosis, blastomycosis, coccidioidomycosis, cryptococcosis, zygomycosis, and aspergillosis, have a predilection for the oral cavity, head, and neck. The incidence of oral fungal infections has grown along with increasing numbers immunocompromised patients as a result of diseases such as AIDS, therapies for cancer, and organ transplantation.

ORAL MANIFESTATIONS OF SYSTEMIC DISEASE

Oral lesions are often the first sign of underlying systemic conditions. Some of the more common disease associations and their oral manifestations are cited in Table 16.1. Only hairy leukoplakia is considered in more detail here.

Hairy Leukoplakia

Hairy leukoplakia is a distinctive oral lesion on the lateral border of the tongue caused by EBV that usually occurs

Table 16.1 Oral Manifestations of Some Systemic Diseases

Systemic Disease	Associated Oral Changes
Infectious Diseases	
Scarlet fever	Fiery red tongue with prominent papillae (raspberry tongue); white-coated tongue through which hyperemic papillae project (strawberry tongue)
Measles	Spotty enanthema in the oral cavity often precedes the skin rash; ulcerations on the buccal mucosa about Stensen duct produce Koplik spots
Infectious mononucleosis	Acute pharyngitis and tonsillitis that may cause coating with a gray-white exudative membrane; enlargement of lymph nodes in the neck, palatal petechiae
Diphtheria	Characteristic dirty white, fibrinosuppurative, tough, inflammatory membrane over the tonsils and retropharynx
Human immunodeficiency virus	Predisposition to opportunistic oral infections, particularly herpes virus, Candida, and other fungi; oral lesions of Kaposi sarcoma and hairy leukoplakia (see text)
Dermatologic Conditions[a]	
Lichen planus	Reticulate, lacelike, white keratotic lesions that sometimes ulcerate and rarely form bullae; seen in more than 50% of patients with cutaneous lichen planus; rarely is the sole manifestation
Pemphigus	Vesicles and bullae prone to rupture, leaving hyperemic erosions covered with exudates
Bullous pemphigoid	Oral lesions (mucus membrane pemphigoid) resemble those of pemphigus but can be differentiated histologically
Erythema multiforme	Maculopapular, vesiculobullous eruption that sometimes follows an infection elsewhere, ingestion of drugs, development of cancer, or a collagen vascular disease; when there is widespread mucosal and skin involvement it is referred to as Stevens-Johnson syndrome
Hematologic Disorders	
Pancytopenia (agranulocytosis, aplastic anemia)	Severe oral infections in the form of gingivitis, pharyngitis, tonsillitis; may extend to produce cellulitis of the neck (Ludwig angina)
Leukemia	With depletion of functioning neutrophils, oral lesions similar to those in pancytopenia may develop
Monocytic leukemia	Leukemic infiltration and enlargement of the gingivae, often with accompanying periodontitis
Miscellaneous	
Melanotic pigmentation	May appear in Addison disease, hemochromatosis, fibrous dysplasia of bone (Albright syndrome), and Peutz-Jeghers syndrome (gastrointestinal polyposis)
Phenytoin (Dilantin) ingestion	Striking fibrous enlargement of the gingivae
Pregnancy	A friable, red, pyogenic granuloma protruding from the gingiva (pregnancy tumor)
Rendu-Osler-Weber syndrome	Autosomal dominant disorder with multiple congenital aneurysmal telangiectasias beneath mucosal surfaces of the oral cavity and lips

[a]See Chapter 25.

in immunocompromised patients. In patients infected with the human immunodeficiency virus (HIV), hairy leukoplakia may portend development of AIDS. However, the lesions are increasingly observed in patients who are immunocompromised for other reasons including cancer therapy, transplant-associated immunosuppression, and advanced age. Hairy leukoplakia takes the form of white, confluent patches of fluffy ("hairy"), hyperkeratotic thickenings, almost always situated on the lateral border of the tongue. Unlike thrush, the lesion cannot be scraped off. The distinctive microscopic appearance consists of hyperparakeratosis and acanthosis with "balloon cells" in the upper spinous layer. EBV RNA transcripts and proteins can be detected within the lesional cells. Superimposed candidal infection can add to the "hairiness."

PRECANCEROUS AND CANCEROUS LESIONS

Many epithelial and connective tissue tumors of the head and neck region (e.g., papillomas, hemangiomas, lymphomas) occur elsewhere in the body and are described in other chapters. This discussion considers only the most common oral cancer, squamous cell carcinoma (SCC), and its associated precancerous lesions.

Leukoplakia and Erythroplakia

Leukoplakia is defined by the World Health Organization as "a white patch or plaque that cannot be scraped off and cannot be characterized clinically or pathologically as any other disease." This clinical term is reserved for lesions that are present in the oral cavity for no apparent reason. As such, white patches caused by obvious irritation or entities such as lichen planus and candidiasis are not considered to be leukoplakia. Approximately 3% of the world's population have leukoplakia; 5% to 25% of these lesions are premalignant. Thus, until proven otherwise by histologic evaluation, all leukoplakias must be considered precancerous.

Related to leukoplakia, but much less common and much more ominous, is *erythroplakia,* which is a red, velvety, possibly eroded area within the oral cavity that usually remains level with or may be slightly depressed relative to the surrounding mucosa (Fig. 16.4). The epithelium in

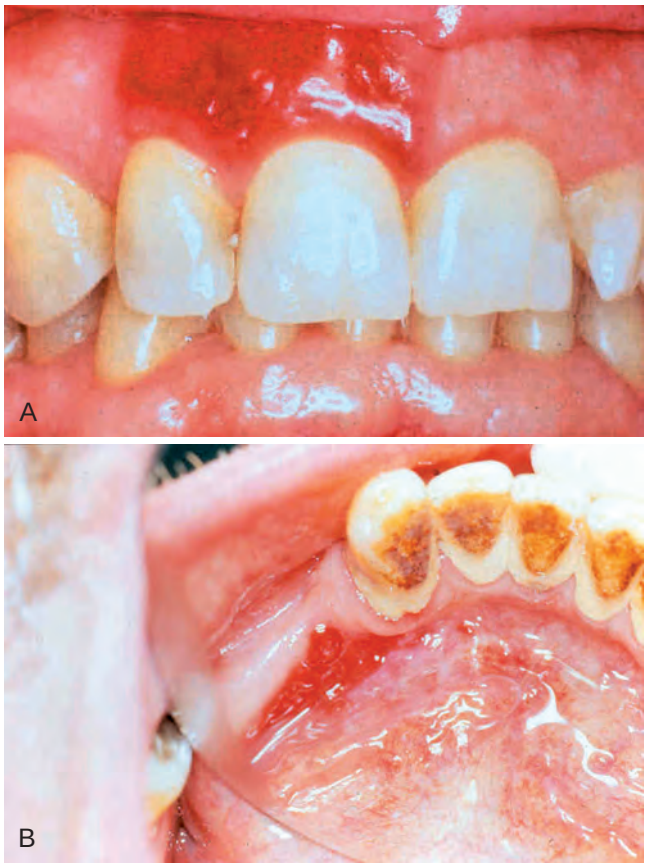


Figure 16.4 Erythroplakia. (A) Red discoloration of the maxillary gingiva. (B) Red lesion of the mandibular alveolar ridge. Biopsy of both lesions revealed carcinoma in situ.

such lesions tends to be markedly atypical, and the risk of malignant transformation is much higher than is leukoplakia. Intermediate forms that have the characteristics of both leukoplakia and erythroplakia are termed *speckled leukoerythroplakia.*

Both leukoplakia and erythroplakia may be seen in adults at any age, but they are usually found in persons 40 to 70 years of age, with a 2:1 male preponderance. Although these lesions have multifactorial origins, the use of tobacco (cigarettes, pipes, cigars, and certain forms of smokeless tobacco) is a common antecedent.

MORPHOLOGY

Leukoplakia may occur anywhere in the oral cavity (favored locations are buccal mucosa, floor of the mouth, ventral surface of the tongue, palate, and gingiva). They appear as solitary or multiple white patches or plaques, often with sharply demarcated borders. They may be slightly thickened and smooth, wrinkled and fissured, or may appear as raised, sometimes corrugated, verrucous plaques (Fig. 16.5A). On histologic examination, they present a spectrum of epithelial changes ranging from hyperkeratosis overlying a thickened, acanthotic but orderly mucosal epithelium to lesions with markedly dysplastic changes, sometimes merging into carcinoma in situ (Fig. 16.5B).

The histologic changes in **erythroplakia** only rarely demonstrate orderly epidermal maturation; virtually all (approximately 90%) display severe dysplasia, carcinoma in situ, or minimally invasive carcinoma. An intense subepithelial inflammatory reaction with vascular dilation that contributes to the reddish appearance of these lesions is often present.

Squamous Cell Carcinoma

Approximately 95% of cancers of the head and neck are SCCs, with the remainder largely consisting of adenocarcinomas of salivary gland origin. Head and neck squamous cell carcinoma (HNSCC) is the sixth most common neoplasm in the world. At current rates, approximately 50,000 cases in the United States and more than 650,000 cases worldwide will be diagnosed each year. The term *head and neck cancer* includes tumors of the oral cavity, pharynx, larynx, and nasal cavities, which are discussed here, and also tumors of the thyroid (Chapter 24) and salivary glands (discussed later in this chapter).

The pathogenesis of SCC is multifactorial.

- *Infection with high-risk human papillomavirus (HPV)* is now the primary cause of SCC of the oropharynx.
- Within North America and Europe, oral cavity SCC has classically been a disease of middle-aged individuals who have been chronic users of *smoked tobacco* and *alcohol*.
- In India and Asia, the chewing of betel quid and paan is a major regional predisposing influence. This concoction, considered a delicacy by some, contains ingredients such as areca nut, slaked lime, and tobacco, wrapped in a betel leaf; many of the ingredients of paan can give rise to carcinogens.
- *Actinic radiation* (sunlight) and pipe smoking are known predisposing influences for cancer of the lower lip.
- The incidence of oral cavity SCC (particularly involving the tongue) in individuals younger than 40 years of age, who have no known risk factors, has been on the rise. The pathogenesis in this group of patients, who are nonsmokers and are not infected with HPV, is unknown.

In the oropharynx, as many as 80% of SCCs, particularly those involving the tonsils, the base of the tongue, and the pharynx, harbor oncogenic variants of HPV, particularly HPV-16. HPV-associated SCC of the oropharynx has increased more than twofold over the past two decades. It is hoped that this trend will be reversed by the HPV vaccine, which is protective against cervical cancer. Early detection of HPV-associated head and neck SCC can be particularly challenging because the anatomic sites of origin (tonsillar crypts, base of tongue, and oropharynx) are not readily accessible or amenable to cytologic screening (unlike the cervix) for premalignant lesions. Unlike the oropharynx, HPV-associated SCC of the oral cavity is relatively uncommon.

Prognosis is dependent on a number of factors including the specific etiology of SCC. The 5-year survival rate of "classic" smoking and alcohol related early-stage SCC is approximately 80%, but survival drops to 20% for late-stage disease. Long-term survival is better in patients with HPV-positive SCC. The dismal outlook in classic SCC reflects diagnosis at advanced stage as well as the frequent presence of multiple primary tumors. Notably, second primary

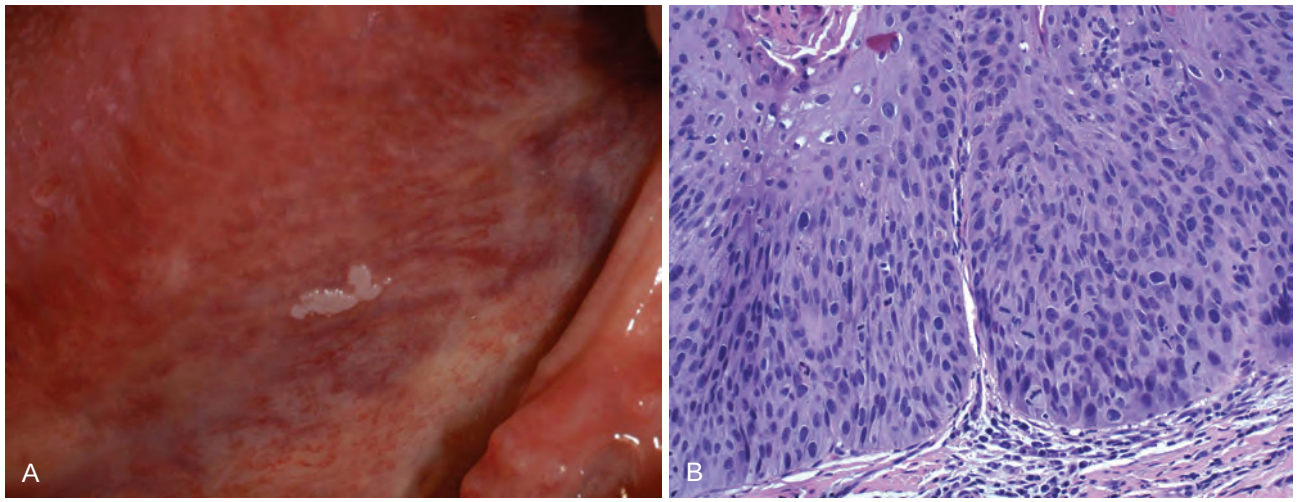


Figure 16.5 Leukoplakia. (A) Clinical appearance of leukoplakia is highly variable. In this example, the lesion is relatively smooth and thin with well-demarcated borders. (B) Histologic appearance of leukoplakia showing severe dysplasia that is characterized by nuclear and cellular pleomorphism, numerous mitotic figures, and a loss of normal maturation.

tumors occur at 3% to 7% per year, the highest rate of among all malignancies. This observation has led to the concept of *field cancerization*, which postulates that multiple individual primary tumors develop independently in the upper aerodigestive tract as a result of years of mucosal exposure to carcinogens. This may explain why there is an almost 35% chance of developing at least one new primary tumor in individuals who survive 5 years after detection of the initial primary tumor. Development of these tumors can be particularly devastating for individuals whose initial lesions were small. The second primary tumors are commonly fatal. Early detection of premalignant lesions is therefore critical for long-term survival of these patients.

Molecular Biology of Squamous Cell Carcinoma. As with other cancers, SCC development is driven by the accumulation of mutations and epigenetic changes that alter the expression and function of oncogenes and tumor suppressor genes. A number of reproducible molecular alterations, some definitively characterized and some inferred, have been identified. In general, these can be categorized as loss of heterozygosity (LOH), copy number alterations, hypermethylation, changes in RNA expression, and somatic DNA mutations.

Several large-scale exome-sequencing studies, including those performed as part of the Cancer Genome Atlas (TCGA), have defined the mutational landscape of SCC. A number of important findings have come from these studies. First, several genes previously proposed as playing critical roles in SCC *(TP53, CDKN2A, PIK3CA)* were shown to be mutated with frequencies sufficient to suggest that they are drivers of cancer development. Second, a number of novel and potentially targetable genetic alterations were identified, particularly *NOTCH1* and its associated pathways and the tumor suppressor *FAT1*, a member of the cadherin family. Third, although *TP53* was the most frequently mutated gene, the mutational frequency for the next most common genes ranged from 1% to 23%, suggesting that there is considerable intertumor variability with respect to the specific mutations harbored in a given tumor. Finally, the mutations identified in HPV-positive and HPV-negative tumors were different. For example, HPV-negative tumors harbored more somatic mutations than HPV-positive tumors. In addition, owing to the expression of the HPV oncoproteins E6 and E7, p53 and RB pathways are typically inactivated in HPV-positive tumors, similar to what has been observed in cervical cancer (Chapter 22). Despite this advanced understanding, targeted therapies for reducing SCC morbidity and mortality remain limited.

Lesions are considered premalignant and at risk for progressing to HNSCC when histologic dysplasia is present. However, the criteria for diagnosing dysplasia are subjective and open to a wide range of interpretation, even among highly qualified pathologists. This, in part, reflects the absence of validated histologic findings that predict malignant transformation of dysplastic lesions. In general, ~15% of premalignant lesions will progress to HNSCC with a mean time interval of 5 years. Our inability to accurately prognosticate on the basis of histologic alterations emphasizes the need to develop molecular criteria for predicting outcomes.

MORPHOLOGY

SCC may arise anywhere in the head and neck region that is lined by stratified squamous epithelium. For the "classic" HPV-negative keratinizing SCC, the favored locations are the ventral surface of the tongue, floor of the mouth, lower lip, soft palate, and gingiva (Table 16.2). Conversely, HPV-associated SCCs are most often nonkeratinizing neoplasms arising in the reticulated epithelium of the tonsillar crypts within the lingual tonsils, base of tongue, soft palate, and pharynx. Although keratinizing SCCs are typically preceded by premalignant lesions, such as leukoplakia and erythroplakia, HPV-associated SCCs tend to develop without a readily identified premalignant (i.e., dysplastic) component.

Early-stage keratinizing SCCs appear as raised, firm, pearly plaques or irregular, roughened, or verrucous areas of mucosal thickening. Either pattern may be superimposed on a background of apparent leukoplakia or erythroplakia. As these lesions enlarge, they typically create ulcerated and protruding masses with irregular and indurated, or rolled, borders (Fig. 16.6A). Conversely, HPV-associated SCCs typically present as small primary tumors that

Table 16.2 Differences Between HPV-Associated and Non-Associated SCC

	Associated With HPV	Not Associated With HPV
Patient Age	Younger	Older
Risk Factors	Number of oral sex partners	Tobacco, alcohol
Location	Oropharynx	Oral cavity
Clinical Presentation	Small primary lesion with bulky nodal disease	Large primary lesion with variable nodal disease
Histology	Nonkeratinizing SCC	Keratinizing SCC
Distant Metastasis	Rare	Common
Clinical Outcomes	Good	Poor
Risk of Second Primary	Low	High

HPV, Human papillomavirus; *SCC,* squamous cell carcinoma.

lack obvious surface mucosal lesions but are accompanied by significant cervical lymphadenopathy.

Keratinizing SCCs begin as dysplastic lesions, which may or may not progress to full-thickness dysplasia (carcinoma in situ) before invading the underlying connective tissue stroma (Fig. 16.6B). This difference in progression should be contrasted with cervical cancer (Chapter 22), in which full-thickness dysplasia, representing carcinoma in situ, typically develops before invasion. Keratinizing SCCs range from well-differentiated to anaplastic, and sometimes sarcomatoid, tumors. However, the degree of histologic differentiation, as determined by the relative degree of keratinization, is not correlated with behavior, including growth rate. SCC tends to infiltrate locally before metastasizing. The routes of extension depend on the primary site. The cervical lymph nodes are favored sites of local metastasis, and the most common sites of distant metastasis are mediastinal lymph nodes, lungs, liver, and bones. Unfortunately, distant metastases are often already present at the time of discovery of the primary lesion.

The histology of **HPV-associated SCC is characterized by the proliferation of nests and lobules of nonkeratinizing and basaloid cells** growing within sheets of lymphocytes (Fig. 16.7A). Immunohistochemical detection of strong p16 protein expression can serve as a marker for HPV-associated SCC (Fig. 16.7B). However, further testing, using PCR or in situ hybridization (ISH) (Fig. 16.7C), may be required in certain cases.

ODONTOGENIC CYSTS AND TUMORS

The overwhelming majority of odontogenic cysts are derived from remnants of odontogenic epithelium present within the jaws. In contrast to the rest of the skeleton, epithelial-lined cysts are common in the jaws. These cysts are subclassified as either inflammatory or developmental (Table 16.3), with only the most common described here.

The *dentigerous cyst* originates around the crown of an unerupted tooth as a result of fluid accumulation between the developing tooth and the dental follicle. Radiographically, these are unilocular lesions most often associated with impacted third molar (wisdom) teeth. Histologically, they are lined by a thin layer of stratified squamous epithelium, and there is frequently a dense chronic inflammatory cell infiltrate in surrounding connective tissue. Complete surgical excision is curative.

The *keratocystic odontogenic tumor,* which was previously referred to as *odontogenic keratocyst*, must be differentiated from other odontogenic cysts because of its aggressive behavior. Keratocystic odontogenic tumors can be seen at any age but are most common between 10 and 40 years of age, in males, and within the posterior mandible. The lesions are well-defined unilocular or multilocular radiolucencies with a lining consisting of a thin layer of keratinized stratified squamous epithelium with a prominent basal cell layer and a corrugated epithelial surface. Treatment requires complete excision because of locally aggressive behavior. Recurrence rates for inadequately removed lesions can be as high as 60%. About 80% of the lesions are solitary, but patients with multiple cysts should be evaluated for nevoid basal cell carcinoma syndrome (Gorlin syndrome), which is associated with mutations in the tumor suppressor gene *PTCH* (Patched) on chromosome 9q22 (Chapter 25).

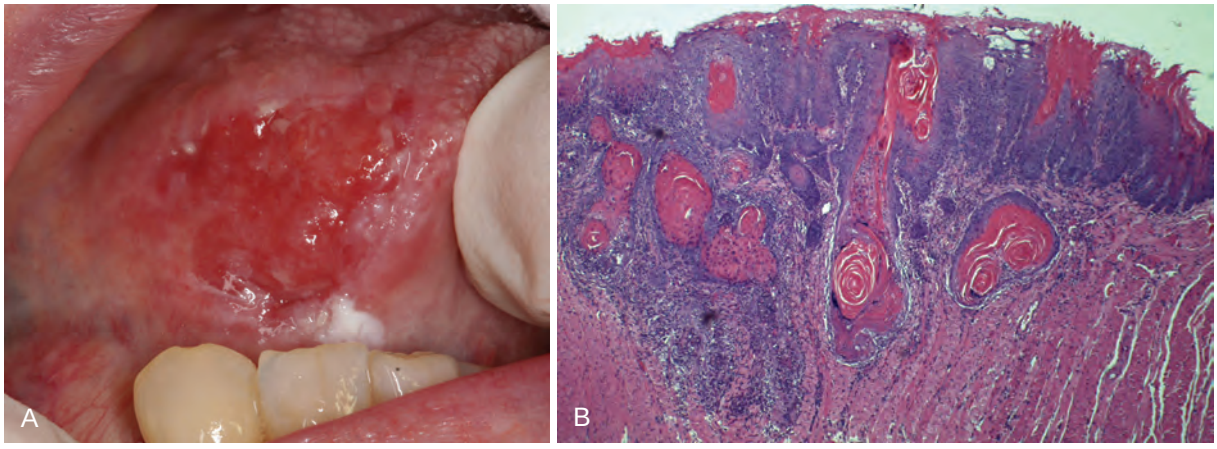


Figure 16.6 Keratinizing squamous cell carcinoma. (A) Clinical appearance demonstrating ulceration and induration of the oral mucosa. (B) Histologic appearance demonstrating numerous nests and islands of malignant keratinocytes with keratin whorls that invade the underlying connective tissue stroma and skeletal muscle.

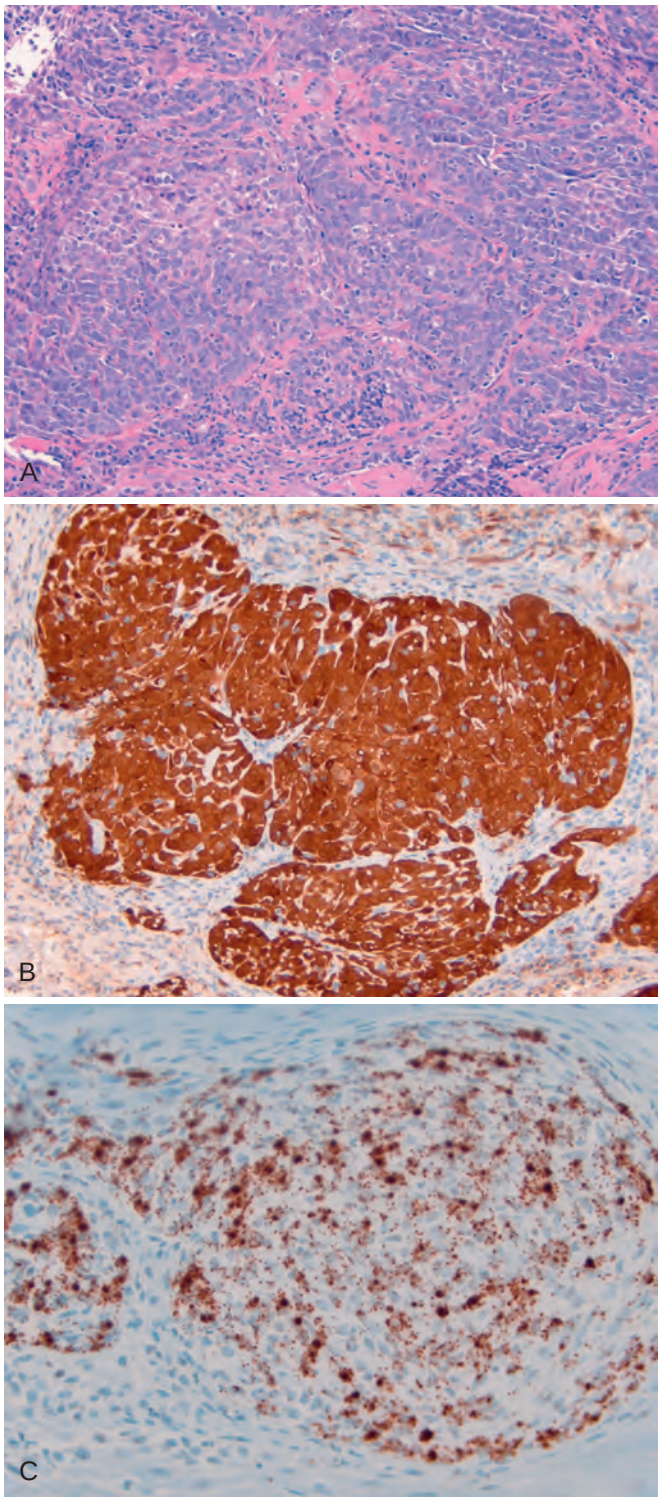


Figure 16.7 Nonkeratinizing squamous cell carcinoma. (A) Histologic appearance demonstrating nests and cords of basaloid appearing cells. (B) Immunohistochemistry demonstrating strong nuclear and cytoplasmic tumor cell staining p16. (C) In situ hybridization for high-risk human papillomavirus E6 and E7 mRNA expression.

The *radicular cyst* is a common inflammatory lesion found at the tooth apex. The cysts develop as a result of long-standing inflammation of the tooth *(pulpitis)*, often secondary to advanced carious lesions or local trauma. Necrosis of the pulp tissue may occur and can traverse the length of the root to exit at the apex into the surrounding alveolar bone. Over time, granulation tissue may develop, with subsequent epithelialization that results in a radicular cyst. The older term *periapical granuloma* is still used, even though the lesion does not show true granulomatous inflammation. Periapical inflammatory lesions persist due to continued presence of bacteria or other irritating agents. Successful treatment, therefore, necessitates the complete removal of offending material and restoration or extraction of the tooth.

Odontogenic tumors are a group of lesions with diverse histologic appearances and clinical behavior. Some are true neoplasms (both benign and malignant), and others are more likely hamartomas. Odontogenic tumors are derived from odontogenic epithelium, odontogenic mesenchyme, or both (Table 16.4). The two most common and clinically significant tumors are odontoma and ameloblastoma.

Table 16.3 Histologic Classification of Odontogenic Cysts

Inflammatory
Radicular cyst
Residual cyst
Paradental (inflammatory collateral) cyst
Developmental
Dentigerous cyst
Odontogenic keratocyst
Gingival cyst
Eruption cyst
Lateral periodontal cyst
Glandular odontogenic cyst
Calcifying epithelial odontogenic cyst
Nasopalatine duct cyst

Table 16.4 Histologic Classification of Odontogenic Tumors

Tumors of Odontogenic Epithelium
Benign
Ameloblastoma
Calcifying epithelial odontogenic tumor
Squamous odontogenic tumor
Adenomatoid odontogenic tumor
Malignant
Ameloblastic carcinoma
Sclerosing odontogenic carcinoma
Clear-cell odontogenic carcinoma
Ghost cell odontogenic carcinoma
Primary intraosseous squamous cell carcinoma
Tumors of Odontogenic Mesenchyme
Odontogenic fibroma
Odontogenic myxoma
Cementoblastoma
Tumors of Odontogenic Epithelium and Mesenchyme
Benign
Odontoma
Odontoma, compound type
Odontoma, complex type
Ameloblastic fibroma
Ameloblastic fibro-odontoma
Dentinogenic ghost cell tumor
Malignant
Ameloblastic fibrosarcoma

- *Odontoma,* the most common odontogenic tumor, arises from epithelium and is associated with extensive enamel and dentin deposition. These lesions probably represent hamartomas rather than true neoplasms and are cured by local excision.
- *Ameloblastoma* arises from odontogenic epithelium and does not display ectomesenchymal differentiation. It is cystic, slow growing, and locally invasive with a typically indolent course. Treatment requires wide surgical resection to prevent recurrence.

KEY CONCEPTS

ORAL CAVITY

- Caries is the most common cause of tooth loss in persons younger than 35 years of age. The primary cause is tooth destruction by acid end-products of bacterial sugar fermentation.
- Gingivitis is a common, reversible inflammation of the mucosa surrounding the teeth.
- Periodontitis is a chronic inflammatory condition that results in destruction of the supporting structures and eventual tooth loss. It is associated with poor oral hygiene and altered oral microbiota.
- Aphthous ulcers are painful superficial ulcers of unknown etiology.
- Fibromas and pyogenic granulomas are common reactive lesions of the oral mucosa.
- Leukoplakias and erythroplakias are oral mucosal lesions that may undergo malignant transformation.
- The majority of oral cavity and oropharyngeal cancers are SCCs. Oral cavity SCCs are linked to tobacco and alcohol use, but the incidence of HPV-associated lesions has risen dramatically in oropharyngeal SCCs.

Upper Airways

The term *upper airway* is used here to include the nose, pharynx, and larynx. Disorders of these structures are among the most common afflictions of humans, but fortunately the overwhelming majority are more nuisance than threat.

NOSE

Inflammatory diseases, such as the common cold, are the most frequent disorders of the nose and accessory air sinuses. These are most often viral in origin, but can be complicated by superimposed bacterial infections. Destructive inflammatory nasal diseases and primary tumors of the nasal cavity or maxillary sinus occur less frequently.

Inflammatory Lesions

Infectious Rhinitis. Infectious rhinitis, also known as the *common cold,* is caused by one or more viruses. Major offenders are adenoviruses, echoviruses, and rhinoviruses, all of which evoke a profuse catarrhal discharge. During the initial acute stages, the nasal mucosa is thickened, edematous, and red; the nasal cavities are narrowed; and the turbinates are enlarged. These changes may extend, to produce pharyngotonsillitis. Secondary bacterial infection enhances the inflammatory reaction and produces a mucopurulent or sometimes frankly suppurative exudate. Fortunately, these infections clear rapidly. As the saying goes, "in a week if treated, but in 7 days if ignored."

Allergic Rhinitis. Allergic rhinitis (hay fever) is initiated by hypersensitivity to one of a large group of allergens, most commonly plant pollens, fungi, animal allergens, and dust mites. It affects 20% of the US population. As with asthma, allergic rhinitis is an IgE-mediated immune reaction with an early- and late-phase response (see "Immediate [Type I] Hypersensitivity" in Chapter 6). The allergic reaction is characterized by mucosal edema, erythema, and mucus secretion accompanied by eosinophil-rich leukocytic infiltrates.

Nasal Polyps. Recurrent attacks of rhinitis may eventually lead to focal protrusions of the mucosa, producing *nasal polyps,* which may reach 3 to 4 cm in length. Histologically, these polyps consist of edematous mucosa with loose stroma, hyperplastic or cystic mucus glands, and infiltrates of neutrophils, eosinophils, and plasma cells (Fig. 16.8). In the absence of bacterial infection, the mucosal surface is intact, but it may become ulcerated. When multiple or large, nasal polyps may encroach on the airway and impair sinus drainage. Despite features indicating an allergic etiology, most people with nasal polyps are not atopic, and only 0.5% of atopic patients develop nasal polyps.

Chronic Rhinitis. Chronic rhinitis follows repeated episodes of microbial or allergic rhinitis with the eventual development of superimposed bacterial infection. A deviated nasal septum or nasal polyps that impair drainage of secretions contribute to the likelihood of microbial invasion. Superficial desquamation or ulceration of the mucosal epithelium is common. Inflammatory infiltrates include variable numbers of neutrophils, lymphocytes, and plasma cells. Infections may extend into the sinuses.

Sinusitis. Most cases of acute sinusitis are preceded by rhinitis, but maxillary sinusitis occasionally arises by extension of a periapical infection through the bony sinus floor. Responsible microorganisms are usually components of the oral microbiome that elicit an inflammatory reaction. The resulting mucosal edema impairs sinus drainage and may result in *empyema* of the sinus. Outflow obstruction occurs most often in the frontal and, less commonly, anterior ethmoid sinuses. This may occasionally lead to mucus accumulation, producing a so-called *mucocele.*

Acute sinusitis may progress to *chronic sinusitis,* particularly when drainage is disrupted. A polymicrobial infection,

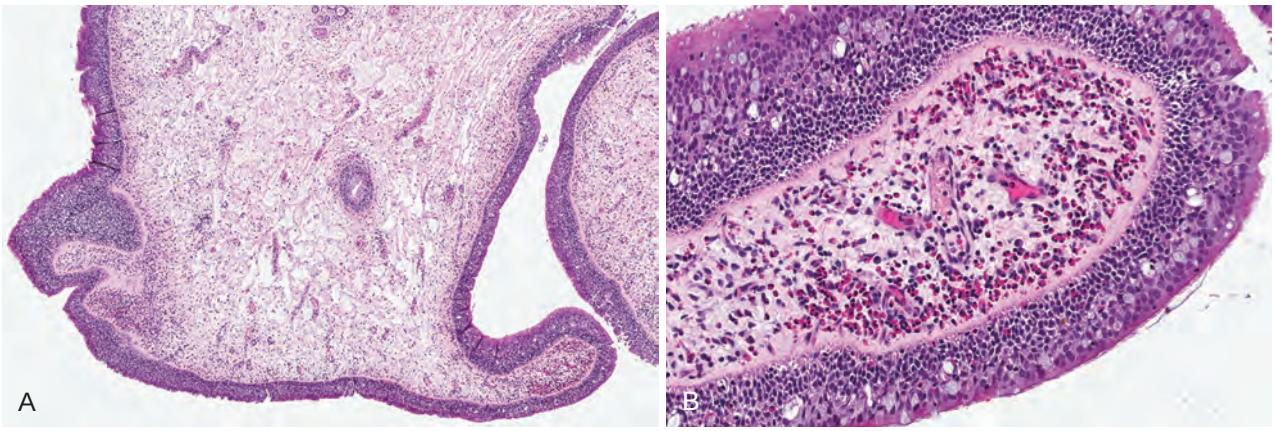


Figure 16.8 (A) Nasal polyps. Low-power magnification showing edematous stroma lined by epithelium. (B) High-power view showing stromal edema, eosinophil-rich inflammatory infiltrate, and respiratory epithelial lining.

largely composed of normal oral microbiome components, is present in most cases. Fungi may cause severe chronic sinusitis (e.g., in mucormycosis), especially in diabetic patients. Sinusitis can, uncommonly, present as a component of *Kartagener syndrome,* which also includes bronchiectasis and situs inversus (Chapter 15), as a result of defective ciliary action. Although most cases of chronic sinusitis are more uncomfortable than disabling or serious, the infections can occasionally spread into the orbit or penetrate into surrounding bone leading to osteomyelitis. Spread into the cranial vault can cause septic thrombophlebitis within a dural venous sinus.

Necrotizing Lesions of the Nose and Upper Airways

Necrotizing ulcerating lesions of the nose and upper respiratory tract may be produced by the following:

- *Acute fungal infections* (including mucormycosis; Chapter 8), particularly in diabetic and immunosuppressed patients
- *Granulomatosis with polyangiitis,* previously called Wegener granulomatosis (Chapter 11)
- *Extranodal NK/T-cell lymphoma, nasal type,* in which the tumor cells harbor EBV (Chapter 13) typically occurs in males of Asian or Latin American descent in the fifth or sixth decade of life; ulceration and bacterial infection frequently complicate the process. Untreated, these neoplasms are almost always rapidly fatal as a result of uncontrolled lymphoma spread and cranial vault penetration or secondary bacterial infection with septic dissemination. Localized cases can often be controlled with radiotherapy, but relapse and recurrence are common and associated with poor outcomes.

NASOPHARYNX

The nasopharyngeal mucosa, related lymphoid structures, and glands may be involved in a wide variety of infections (e.g., diphtheria, infectious mononucleosis) as well as neoplasms.

Inflammatory Lesions

Pharyngitis and tonsillitis frequently accompany viral upper respiratory infections. Rhinoviruses, echoviruses, and adenoviruses are the most common causes; remaining cases are primarily due to various strains of influenza or respiratory syncytial virus. Erythema and edema of the nasopharyngeal mucosa with reactive enlargement of nearby tonsils and lymph nodes are characteristic. Bacterial infection may cause pharyngitis and tonsillitis or, alternatively, occur as a secondary process superimposed on viral infection. β-hemolytic streptococci are the most frequent pathogens, but *Staphylococcus aureus* or other bacteria may also be present. When bacteria are present, exudate and exudative membranes (pseudomembranes) may cover the nasopharyngeal mucosa as well as enlarged nasopalatine and palatine tonsils. *Follicular tonsillitis,* in which enlarged tonsils (due to reactive lymphoid hyperplasia) are dotted by exudate emanating from the tonsillar crypts, is common.

The major importance of streptococcal "sore throat" lies in the possible development of late sequelae, such as rheumatic fever (Chapter 12) and glomerulonephritis (Chapter 20). Whether recurrent episodes of acute tonsillitis favor the development of chronic tonsillar enlargement is open to debate, but regardless of the cause of the enlargement, patients may benefit from surgical excision.

TUMORS OF THE NOSE, SINUSES, AND NASOPHARYNX

Tumors in the nose, sinuses, and nasopharynx are infrequent but include the wide spectrum of mesenchymal and epithelial neoplasms.

Nasopharyngeal Angiofibroma. **Nasopharyngeal angiofibroma is a benign, highly vascular tumor that occurs almost exclusively in adolescent males who are often fair-skinned and red headed.** It is believed to arise within the fibrovascular stroma of the posterolateral wall of the roof of the nasal cavity. Surgical removal, often with preoperative embolization to decrease bleeding, is the treatment

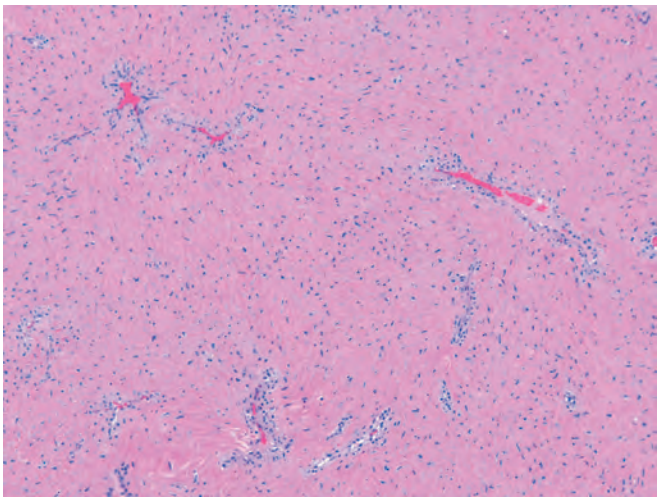


Figure 16.9 Nasopharyngeal angiofibroma showing thin-walled muscular vessels separated by dense collagenous stroma containing small, bland fibroblasts.

of choice. Because nasopharyngeal angiofibroma is often locally aggressive, with intracranial extension, recurrence rates can approach 20%, and 9% of cases can be fatal (Fig. 16.9). Mutations in *CTNNB1*, which encodes *beta-catenin*, are present in the majority of nasopharyngeal angiofibromas. This tumor can also develop sporadically or syndromically in association with familial adenomatous polyposis.

Sinonasal (Schneiderian) Papilloma. **Sinonasal papilloma is a benign neoplasm arising from the respiratory or Schneiderian mucosa lining the nasal cavity and paranasal sinuses.** These lesions occur in three forms: *exophytic* (most common), *endophytic* (or *inverted*), and *oncocytic* (formerly known as *cylindrical cell)*. Sinonasal papillomas are most common in males between 30 and 60 years of age. The endophytic form may invaginate into the underlying stroma (Fig. 16.10). Although this is a benign neoplasm, it can display locally aggressive behavior within both the nose and the paranasal sinuses, including invasion into the orbit or cranial vault, and has a high rate of recurrence if not adequately excised. Malignant transformation occurs in approximately 10% of cases. Most endophytic sinonasal papillomas have *EGFR* gene mutations. The remaining cases generally harbor HPV DNA, often low-risk types 6 and 11.

Olfactory Neuroblastoma (Esthesioneuroblastoma). **Olfactory neuroblastomas arise from neuroectodermal olfactory cells within the mucosa, particularly in the superior aspect of the nasal cavity.** The age distribution is bimodal, with peaks at 15 and 50 years. Patients typically present with nasal obstruction and/or epistaxis. Histologically, olfactory neuroblastomas are *small, blue, round cell neoplasms*, a category that includes lymphoma, small cell carcinoma, Ewing sarcoma/peripheral neuroectodermal tumor, rhabdomyosarcoma, melanoma, and sinonasal undifferentiated carcinoma. Olfactory neuroblastomas are typically composed of well-circumscribed nests and lobules of cells separated by fibrovascular stroma (Fig. 16.11). A fibrillary matrix, representing tangled neuronal cell processes, is often present. Consistent with their neuroendocrine origin, electron microscopy demonstrates membrane-bound secretory granules, and immunohistochemistry is positive for neuron-specific enolase, synaptophysin, CD56, and chromogranin. Depending on tumor stage and grade, combinations of surgery, radiation therapy, and chemotherapy yield 5-year survival rates of 40% to 90%. Genetic aberrations in olfactory neuroblastoma are heterogenous, with cases demonstrating varying chromosomal gains and/or losses.

NUT Midline Carcinoma. **This uncommon tumor occurs in the nasopharynx, salivary gland, or other midline structures of the thorax or abdomen.** It can occur at any age, from infancy to late adulthood, but the true incidence of NUT midline carcinoma is not known, as it is easily mistaken for SCC. Despite this, NUT midline carcinoma is clinically distinctive due to its extremely aggressive behavior and resistance to conventional therapy; most patients survive for less than 1 year following diagnosis. NUT midline carcinomas are uniformly associated with translocations that fuse genes encoding chimeric proteins composed of most of NUT, a chromatin regulator, and a portion of a "chromatin reader" protein, usually BRD4. Drugs that displace BRD4-NUT from chromatin induce NUT midline carcinoma cells to terminally differentiate, and targeted therapy with BRD4-NUT inhibitors is now being tested clinically.

Nasopharyngeal Carcinoma. **Nasopharyngeal carcinoma is characterized by a distinctive geographic distribution, a close anatomic relationship to lymphoid tissue, and an association with EBV infection.** The nomenclature for nasopharyngeal carcinomas continues to evolve, but, at

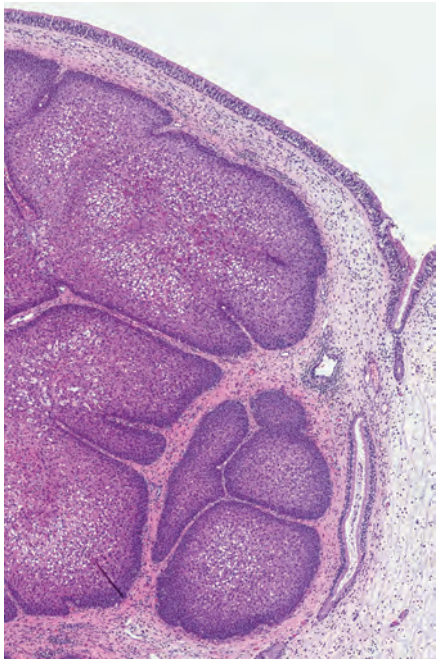


Figure 16.10 Inverted papilloma. The overlying surface is lined by thin respiratory epithelium; however, the underlying stroma contains multiple nodules of thick neoplastic epithelium growing inward; hence, the term *inverted*.

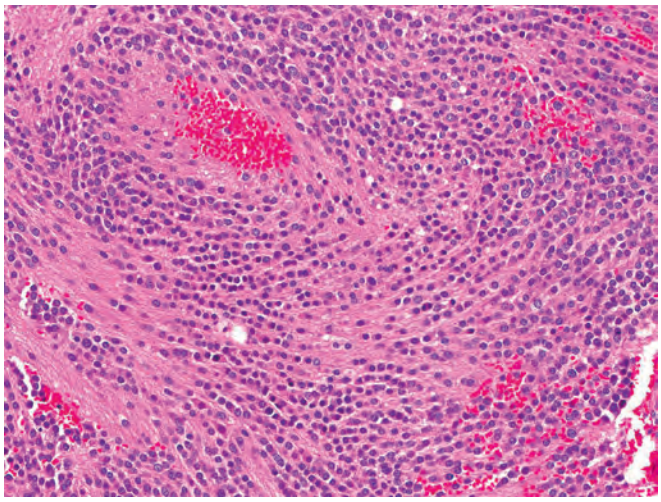


Figure 16.11 Olfactory neuroblastoma is composed of small monotonous cells with round nuclei in an eosinophilic fibrillary stroma. The cells appear to swirl around a blood-filled vascular channel.

present, is described as one of three patterns: (1) keratinizing SCCs, (2) nonkeratinizing SCCs, and (3) basaloid SCCs. *Lymphoepithelioma* or *lymphoepithelial carcinoma* are historic names for this tumor that are no longer used.

The three principal factors that influence development of nasopharyngeal carcinomas are heredity, age, and EBV infection. For example, these neoplasms are particularly common in some parts of Africa, where they are the most frequent childhood cancer. In contrast, nasopharyngeal carcinomas are common in adults in southern China but rarely occur in children. In the United States, nasopharyngeal carcinomas are rare in all age groups. Beyond EBV infection, diets high in nitrosamines, which are found in fermented foods and salted fish, and environmental insults, including smoking and chemical fumes, have been linked to these tumors. In the nonkeratinizing form, most patients have antibodies against EBV early antigens or viral capsid antigens, and PCR can be used to detect EBV DNA in the serum.

MORPHOLOGY

Histologically, keratinizing and nonkeratinizing nasopharyngeal carcinomas resemble well-differentiated and poorly differentiated SCCs arising in other locations. The undifferentiated variant is composed of large epithelial cells with oval or round vesicular nuclei, prominent nucleoli, and indistinct cell borders; the latter impart the appearance of a syncytium (Fig. 16.12B). Abundant, normal-appearing lymphocytes, predominantly T cells, infiltrate the tumors. EBV-encoded RNAs such as EBER-1 or proteins such as LMP-1 can be identified in the malignant epithelial cells by in situ hybridization (Fig. 16.12C) or immunohistochemistry, respectively.

Primary nasopharyngeal carcinomas are often clinically occult until they present at advanced stages with nasal obstruction, epistaxis, and metastases to the cervical lymph nodes in up to 70% of patients. Radiotherapy is the

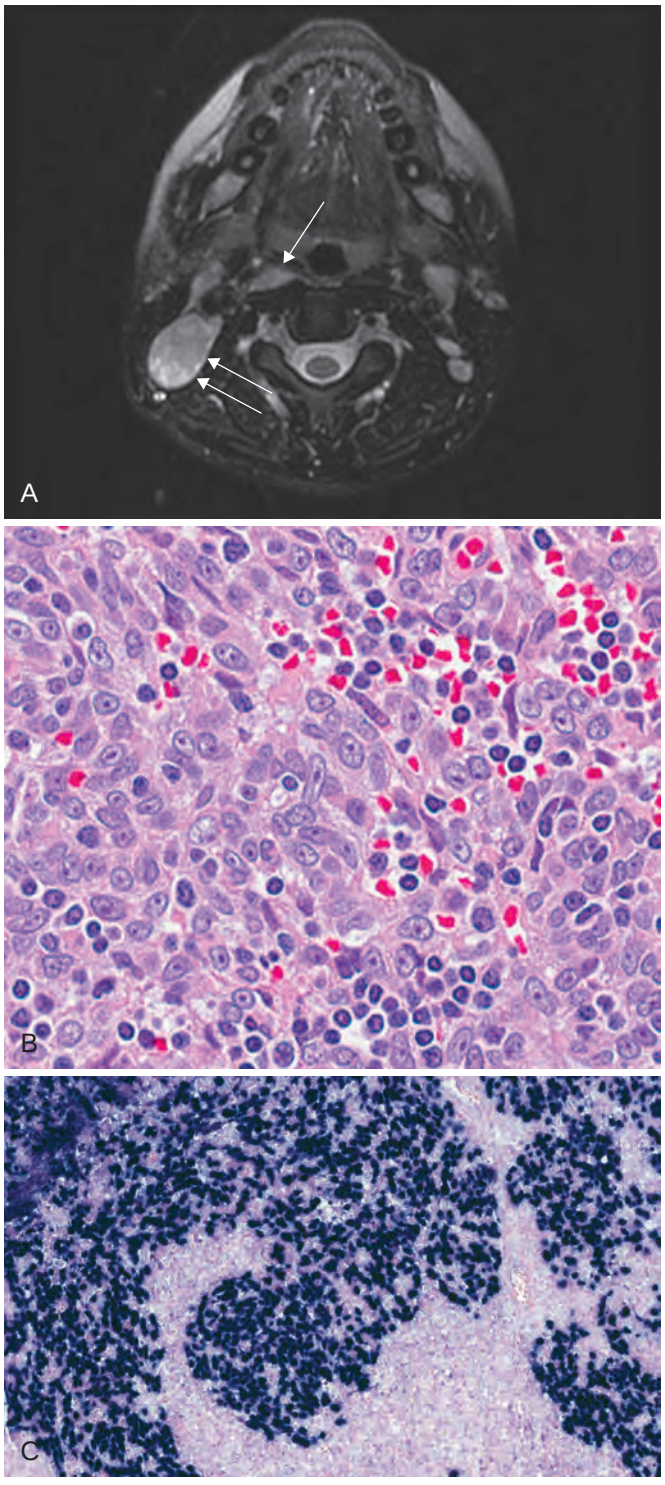


Figure 16.12 Nasopharyngeal carcinoma, nonkeratinizing undifferentiated type. (A) Computed tomography study demonstrating thickening of the nasopharyngeal region *(arrow)* and an enlarged cervical lymph node *(double arrow)*. (B) The syncytial clusters of malignant epithelium are infiltrated by benign lymphocytes. (C) In situ hybridization for EBER-1, a small nuclear RNA encoded by Epstein-Barr virus.

standard treatment and results in an overall 5-year survival of approximately 60%. This, however, varies with histology; the stage-dependent 5-year survival for nonkeratinizing nasopharyngeal carcinoma ranges from 70% to 98%, but that for the keratinizing form is only 20%. These stark differences

have been attributed to divergent therapeutic responses, as undifferentiated carcinomas are the most radiosensitive, and the keratinizing carcinomas are least radiosensitive.

LARYNX

The most common disorders of the larynx are inflammatory. Tumors are uncommon but are amenable to resection, though often at the price of loss of natural voice.

Inflammatory Lesions

Laryngitis may occur as the sole manifestation of allergic, viral, bacterial, or chemical insult, but it is more commonly associated with generalized upper respiratory tract infection, heavy environmental toxin exposure (e.g., tobacco smoke), or gastroesophageal reflux due to irritating effects of gastric contents. The larynx may also be affected in systemic infections, such as tuberculosis and diphtheria. Although most infections are self-limited, the sequelae can be serious, especially in infancy or childhood, when mucosal congestion, exudation, or edema may cause laryngeal obstruction. In particular, laryngoepiglottitis caused by respiratory syncytial virus, *Haemophilus influenzae,* or β-hemolytic streptococci may induce such sudden swelling of the epiglottis and vocal cords in the small airways of infants and young children as to constitute a medical emergency. This is uncommon in adults because of the larger size of the larynx as well as stronger accessory respiratory muscles. *Croup* refers to laryngotracheobronchitis in children that produces a characteristic inspiratory stridor due to airway narrowing. The laryngitis that is common in heavy smokers predisposes to squamous epithelial metaplasia and sometimes overt carcinoma.

Reactive Nodules (Vocal Cord Nodules and Polyps)

Reactive nodules, also called polyps, develop on the vocal cords, most often in heavy smokers or those who impose great strain on their vocal cords (so-called *singer's nodules*) (Fig. 16.13). Because of their strategic location and accompanying inflammation, they characteristically change the character of the voice and often cause progressive hoarseness. Despite this, the risk of developing cancer in these lesions is almost nonexistent. By convention, vocal cord nodules are bilateral and polyps are unilateral. Adults are most often affected. Both nodules and polyps are smooth, rounded, sessile or pedunculated excrescences, generally only a few millimeters in greatest dimension, located usually on the true vocal cords. The histologic features of nodules and polyps are similar; they are covered by squamous epithelium that may become keratotic, hyperplastic, or even slightly dysplastic overlying a loose myxoid connective tissue core. The latter may be variably fibrotic, fibrinous, or highly vascularized. Nodules on opposing vocal cords may impinge on each other and cause ulceration.

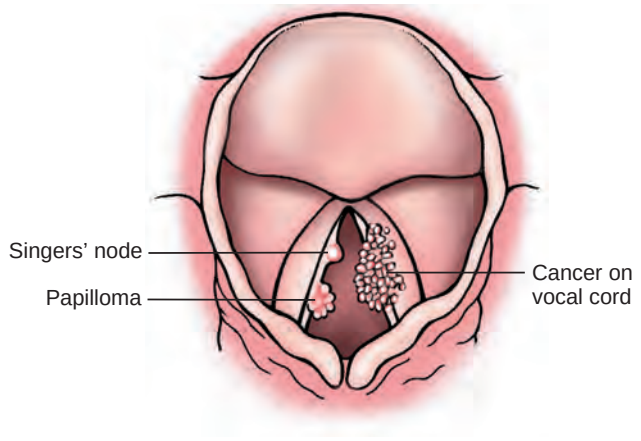


Figure 16.13 Diagrammatic comparison of a singer's nodule, a benign papilloma, and an exophytic carcinoma of the larynx to highlight differences in their clinical appearance.

Squamous Papilloma and Papillomatosis

Laryngeal squamous papillomas are benign neoplasms, usually located on the true vocal cords, that form soft, exophytic proliferations that rarely exceed 1 cm in diameter (see Fig. 16.13). The growths are composed of multiple slender, fingerlike projections supported by central fibrovascular cores and covered by an orderly stratified squamous epithelium. When papillomas develop on the free edge of a vocal cord, trauma may lead to ulceration that can be accompanied by hemoptysis.

Papillomas are usually solitary in adults but are often multiple in children, a condition referred to as *juvenile laryngeal papillomatosis*. The lesions are caused by HPV types 6 and 11 acquired via the maternal birth canal. They frequently recur but only rarely transform to SCC. Although the polyps often regress spontaneously at puberty, the regularity of recurrence requires some children to undergo numerous surgeries.

Carcinoma of the Larynx

Carcinoma of the larynx is typically a squamous cell carcinoma seen in male chronic smokers.

Hyperplasia-Dysplasia-Carcinoma Sequence. **Epithelial alterations in the larynx range from hyperplasia to dysplasia, carcinoma in situ, and invasive carcinoma.** Grossly, epithelial surfaces may be smooth, white or reddened with focal thickenings, roughened by keratosis, or punctuated by irregular verrucous or ulcerated white-pink lesions (see Figs. 16.13 and 16.14).

Nondysplastic hyperplasias have almost no potential for malignant transformation. For lesions with dysplasia, the risk of malignant transformation is directly proportional to the initial grade of dysplasia. For example, the risk of transformation is only 1% to 2% over 5 to 10 years for lesions with mild dysplasia, but increases to 5% to 10% for those with severe dysplasia. Histologic evaluation is the only way to grade dysplasia.

Laryngeal carcinoma is most often related to tobacco smoke; risk is proportional to exposure. Prior to malignant transformation, epithelial changes frequently regress after smoking cessation. Alcohol synergizes with tobacco to increase risk substantially. Nutritional factors, asbestos

exposure, irradiation, and HPV infection can also increase cancer risk.

MORPHOLOGY

About 95% of laryngeal carcinomas are typical SCCs. Most arise on the vocal cords, but they may develop above or below the cords and involve the epiglottis, aryepiglottic folds, or pyriform sinuses. SCCs of the larynx develop similarly to those at other sites, beginning as in situ lesions that become pearly gray, wrinkled mucosal plaques and, ultimately, ulcerated fungating masses (Fig. 16.14). Histologically, the degree of anaplasia is highly variable and can include tumor giant cells and bizarre mitotic figures. As expected of sites in which neoplasia arises following recurrent exposure to environmental carcinogens, squamous hyperplasia with foci of dysplasia or carcinoma in situ may be present within adjacent mucosae.

Laryngeal carcinoma presents most commonly in men within the sixth decade of life. The initial manifestation is often persistent hoarseness, dysphagia, and dysphonia. Prognosis is directly related to clinical stage. In early disease, organ-preserving techniques (e.g., laser surgery, microsurgery, and radiation therapy) are being used with greater frequency. Combined chemotherapy and radiation therapy, with or without laryngectomy, is typically required for more advanced or recurrent disease.

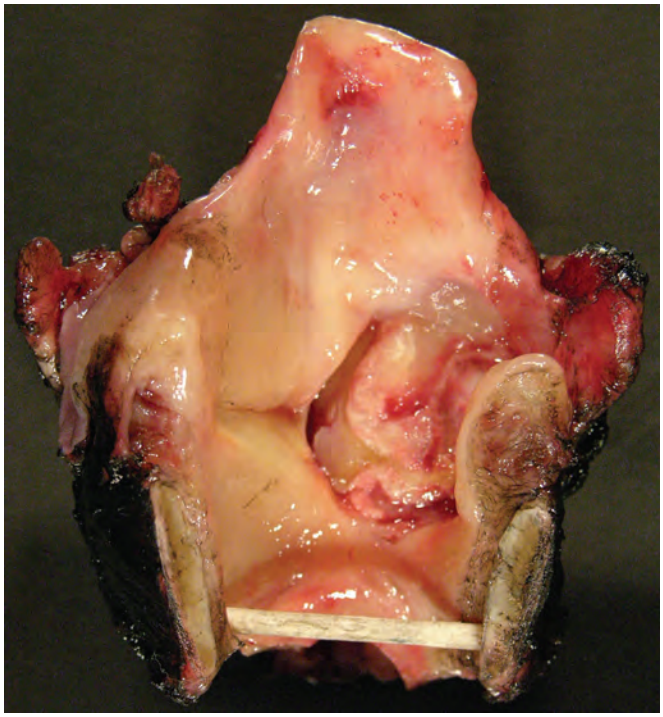


Figure 16.14 Laryngeal carcinoma. Note the large, ulcerated, fungating lesion involving the right vocal cord and pyriform sinus.

KEY CONCEPTS

UPPER AIRWAYS

- Rhinitis can be infectious or allergic in nature and, over time, can lead to chronic rhinitis, sinusitis, and development of nasal polyps.
- Pharyngitis and tonsillitis are typically caused by common upper respiratory tract infection by rhinoviruses, echoviruses, and adenoviruses.
- Nasopharyngeal carcinoma is most often caused by EBV; tumors are most common among African children and Asian adults.
- Laryngitis can be caused by a host of etiologies, including allergic, viral, bacterial, and chemical insults.
- Vocal cord nodules and polyps are reactive lesions seen in smokers or individuals who strain their vocal cords.
- Laryngeal SCC is linked to smoking and alcohol use and is most prevalent in older males.

Ears

Although they rarely shorten life, disorders of the ear often impact its quality. The most common aural disorders, in descending order of frequency, are (1) acute and chronic otitis (most often involving the middle ear and mastoid), sometimes leading to a cholesteatoma; (2) symptomatic otosclerosis; (3) aural polyps; (4) labyrinthitis; (5) carcinomas, largely of the external ear; and (6) paragangliomas, mostly in the middle ear. Only those conditions that have distinctive morphologic features (save for labyrinthitis) are described. Paragangliomas are discussed later.

INFLAMMATORY LESIONS

Acute and chronic otitis media occur most often in infants and children. The origin is typically viral infection that induces a serous exudate. Superimposed bacterial infection, most frequently by *Streptococcus pneumoniae,* nontypeable *H. influenzae,* or *Moraxella catarrhalis,* can lead to suppurative inflammation.

Repeated bouts of acute otitis media with failure of resolution lead to chronic disease. The causative agents are usually *Pseudomonas aeruginosa, S. aureus,* fungus, or, in some cases, a polymicrobial infection. Chronic infection has the potential to perforate the eardrum, encroach on the ossicles or labyrinth, spread into the mastoid spaces, and even penetrate the cranial vault to produce temporal cerebritis or abscess. In individuals with diabetes, otitis media caused by *P. aeruginosa* is especially aggressive and can spread widely, resulting in destructive necrotizing otitis media.

Cholesteatomas are non-neoplastic, cystic lesions associated with chronic otitis media. The cysts are typically 1 to 4 cm in diameter, lined by keratinizing squamous or metaplastic mucus-secreting epithelium, and filled with

amorphous, keratinous debris. Cholesterol spicules may be present. Although the pathogenesis is not clear, it is thought that chronic inflammation and perforation of the eardrum, with ingrowth of the squamous epithelium or metaplasia of the secretory epithelial lining of the middle ear, promote formation of a squamous cell nest that becomes cystic. Cyst rupture induces a brisk inflammatory reaction that includes giant cells engulfing partially necrotic squamous cells and other particulate debris. These lesions, by progressive enlargement, can erode into the ossicles, labyrinth, adjacent bone, or surrounding soft tissue and sometimes produce visible neck masses.

OTOSCLEROSIS

Otosclerosis refers to abnormal bone deposition in the middle ear about the rim of the oval window into which the footplate of the stapes fits. Both ears are usually affected. At first, there is fibrous ankylosis of the footplate. This often occurs in the early decades of life. Over time, bony overgrowth anchors the footplate into the oval window. The severity of the hearing loss reflects the degree of immobilization. Minimal degrees of otosclerosis are common in young to middle-aged adults, but fortunately more severe symptomatic disease is relatively rare. Otosclerosis is familial in most instances, with an autosomal dominant transmission and variable penetrance. The basis for the osseous overgrowth is obscure, but it appears to represent uncoupling of normal bone resorption and bone formation. Thus, it begins with bone resorption, followed by fibrosis and vascularization of the temporal bone in the immediate vicinity of the oval window. Over time the fibrosis is replaced by dense new bone anchoring the footplate of the stapes. In most instances, the process is slowly progressive over the span of decades and can eventually lead to marked hearing loss.

TUMORS

Epithelial and mesenchymal tumors of the external, middle, or internal ear are rare, with the exception of basal cell carcinomas or SCCs arising on the pinna (external ear). These tend to occur in elderly men and are associated with sun exposure. In contrast, SCCs of the ear canal occur most often in middle-aged to elderly women and are not associated with sun exposure. SCCs involving the ear resemble their counterparts at other skin locations, beginning as papules that extend and eventually erode and invade. Despite local invasion, basal cell carcinomas and SCCs involving the pinna rarely spread. The outlook is more bleak when SCCs arise in the external ear canal, as these may invade the cranial cavity or metastasize to regional lymph nodes, accounting for a 5-year survival of only about 50%.

KEY CONCEPTS

EARS

- Infections of the ear are common in children and typically viral in etiology. Chronic infections can be complicated by bacterial infections, which, in turn, can lead to secondary complications including perforated eardrum as well as spreading to the ossicles or mastoid spaces.
- Otosclerosis, with its associated hearing loss, is caused by the abnormal deposition of bone in the middle ear.

Neck

Most of the conditions that involve the neck are described elsewhere (e.g., primary squamous cell and basal cell carcinomas of the skin, melanomas, lymphomas) or are components of systemic disorders (e.g., generalized rashes, lymphadenopathy of infectious mononucleosis, tonsillitis). What remains to be considered here are a few uncommon lesions unique to the neck.

BRANCHIAL CYST (CERVICAL LYMPHOEPITHELIAL CYST)

The vast majority of branchial cysts are thought to arise from remnants of the second branchial arch and are most commonly observed in young adults between 20 and 40 years of age. These benign lesions usually appear on the upper lateral aspect of the neck along the sternocleidomastoid muscle. The slowly-enlarging cysts are well circumscribed, 2 to 5 cm in diameter, and usually lined by stratified squamous or pseudostratified columnar epithelium. The fibrous cyst walls typically contain lymphoid tissue with prominent germinal centers. Cyst contents may be clear and watery or mucinous and may also contain desquamated cells and granular cellular debris. The cysts are readily excised. Similar lesions may appear in the parotid gland or oral cavity beneath the tongue. Branchial cysts do not undergo malignant transformation. Most cystic SCCs in the neck are metastases from cancers of the upper airways or digestive tract.

THYROGLOSSAL DUCT CYST

The thyroid anlage begins in the region of the foramen cecum at the base of the tongue; as the gland develops, it descends to its definitive midline location in the anterior neck. Remnants of this developmental process may persist, resulting in 1- to 4-cm cysts that are lined by stratified squamous epithelium when located near the base of the tongue or by pseudostratified columnar epithelium in lower locations. Transitional epithelial differentiation patterns also occur. The fibrous cyst wall often includes lymphoid aggregates or thyroid remnants. Malignant transformation of the lining epithelium is exceedingly rare. The definitive treatment is surgical excision.

PARAGANGLIOMA (CAROTID BODY TUMOR)

Paragangliomas arise from neuroendocrine cells associated with the sympathetic and parasympathetic nervous systems and occur at many sites. Adrenal medullary pheochromocytomas are the most common paraganglioma (Chapter 24). Approximately 70% of extra-adrenal paragangliomas occur in the head and neck. Although the pathogenesis is not fully understood, loss-of-function mutations in genes encoding succinate dehydrogenase subunits (which are also common in gastrointestinal stromal tumors) or cofactors in mitochondrial oxidative phosphorylation occur frequently in both hereditary and spontaneous paragangliomas. How these mutations contribute to tumor development is not yet clear, but it is suspected that they do so by altering cellular metabolism, which you will recall is one of the hallmarks of neoplasia (Chapter 7). Interestingly, the incidence of these tumors is greater in people living at high altitudes.

Paragangliomas typically develop in two locations:

- *Paravertebral paraganglia* (e.g., organ of Zuckerkandl and, rarely, bladder). Such tumors have sympathetic connections and stain positively for chromaffin, which detects cells producing catecholamines.
- Paraganglia related to the great vessels of the head and neck, the so-called *aorticopulmonary chain,* including the *carotid bodies* (i.e., carotid body tumor); aortic bodies; jugulotympanic ganglia (sometimes referred to as *glomus tympanicum* or *jugulare*); ganglion nodosum of the vagus nerve; and clusters located about the oral cavity, nose, nasopharynx, larynx, and orbit. These are innervated by the parasympathetic nervous system and only rarely produce catecholamines.

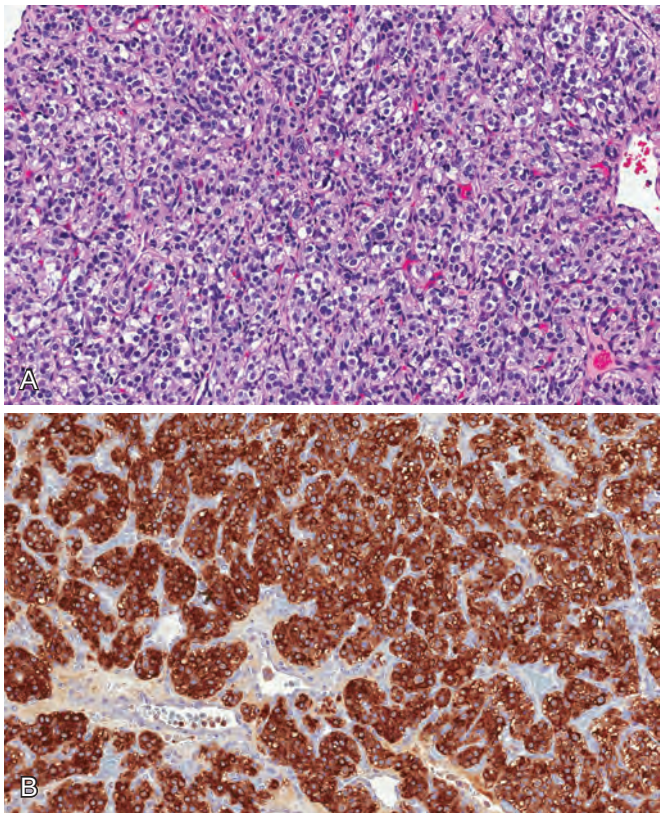


Figure 16.15 Carotid body tumor. (A) Low-power view showing tumor clusters separated by septa (zellballen). The septae are marked by capillaries containing bright red blood cells. (B) Immunohistochemistry demonstrating positivity for chromogranin in the tumor cells.

MORPHOLOGY

The **carotid body tumor** is a typical parasympathetic paraganglioma. It rarely exceeds 6 cm in diameter and arises close to or envelops the bifurcation of the common carotid artery. The tumor tissue is red-pink to brown, and the microscopic features are characteristic of paragangliomas at all sites. Nests (zellballen) of round to oval **chief cells** with abundant, clear or granular, eosinophilic cytoplasm and uniform, round to ovoid, sometimes vesicular, nuclei are surrounded by delicate vascular septae (Fig. 16.15). There is little cellular pleomorphism, and mitoses are scant. The chief cells, which are neuroectodermal in origin, stain strongly for neuroendocrine markers including chromogranin, synaptophysin, neuron-specific enolase, CD56, and CD57. The supporting network of spindle-shaped stromal cells, collectively called *sustentacular cells,* are positive for S-100 protein. Electron microscopy often detects well-demarcated neuroendocrine granules in paravertebral tumors, but the number of these granules is variable, and they tend to be scant in nonfunctioning tumors.

Carotid body tumors, and paragangliomas in general, are rare, slow-growing, painless masses that usually arise in the fifth and sixth decades of life. They commonly occur singly and sporadically but also may be familial. Tumors associated with multiple endocrine neoplasia type 2 most often occur within the adrenal (pheochromocytomas) (Chapter 24), whereas tumors associated with hereditary paraganglioma syndromes (PGL types 1 to 4, involving the succinate dehydrogenase genes) typically involve the head and neck. Approximately one-third of head and neck paragangliomas are thought to be associated with germline mutations. Carotid body tumors frequently recur after incomplete resection and, despite their benign appearance, may metastasize to regional lymph nodes and distant sites. About 50% ultimately prove fatal, largely because of infiltrative growth. Unfortunately, histologic features do not predict the clinical course of a carotid body tumor—mitoses, pleomorphism, and even vascular invasion are not reliable indicators. However, succinate dehydrogenase B *(SDHB)* mutations are associated with the highest rates of metastasis (30% to 50%).

KEY CONCEPTS

NECK

- Branchial cysts arise from the region of the second branchial pouch and typically present in young adults.
- Thyroglossal duct cysts arise due to incomplete descent of the thyroid analog from the foramen cecum in the base of the tongue.
- Seventy percent of extra-adrenal paragangliomas occur in the head and neck. These tumors are most common in the fifth and sixth decades of life and can be hereditary or sporadic. Hereditary head and neck paragangliomas are associated with succinate dehydrogenase gene mutations, occur at younger ages, are often multiple, and can be malignant.

Salivary Glands

There are three major salivary glands, parotid, submandibular, and sublingual, as well as innumerable minor salivary glands distributed throughout the mucosa of the oral cavity; all may be involved by inflammatory or neoplastic disease.

XEROSTOMIA

Xerostomia is defined as a dry mouth resulting from a decrease in the production of saliva. Its incidence has been reported to be as high as 20% in individuals older than 70 years of age in some populations. Xerostomia is characteristic of the autoimmune disorder *Sjögren syndrome* (Chapter 6), in which it is usually accompanied by dry eyes (keratoconjunctivitis sicca). Lack of salivary secretions is also a major complication of radiation therapy for head and neck cancers. Xerostomia is, however, most often a side-effect of commonly-prescribed medications, including anticholinergic, antidepressant/antipsychotic, diuretic, antihypertensive, sedative, muscle relaxant, analgesic, and antihistamine drugs. Xerostomia may present as dry mucosa and/or atrophy of the papillae of the tongue, with fissuring and ulcerations. In Sjögren syndrome, inflammatory enlargement of the salivary glands may also occur. Complications of xerostomia include increased rates of dental caries, candidiasis, and difficulty in swallowing and speaking.

INFLAMMATION (SIALADENITIS)

Sialadenitis may be induced by trauma, viral or bacterial infection, or autoimmune disease. The most prevalent form of sialadenitis is the **mucocele.** The most common viral cause of sialadenitis is **mumps,** which affects the major salivary glands, particularly the parotids, and frequently involves other glandular organs such as the pancreas and testes (Chapter 8).

Mucocele. **Mucoceles are most often found on the lower lip as the result of trauma** (Fig. 16.16A), **occur at all ages, and present as fluctuant lower lip swellings with a blue translucent hue.** This common salivary gland lesion results from either blockage or rupture of a salivary gland duct, with consequent leakage of saliva into the surrounding connective tissue stroma. Patients may report a history of changes in the size of the lesion, particularly in association with meals. Histologically, mucoceles are pseudocysts lined by inflammatory granulation tissue or fibrous connective tissue and filled with mucin and inflammatory cells, particularly macrophages (Fig. 16.16B). Complete excision of the cyst and accompanying minor salivary gland lobule is required, as incomplete excision may lead to recurrence.

Ranula is a term reserved for epithelial-lined cysts that arise when the duct of the sublingual gland has been damaged. A ranula may become so large that it develops into a *plunging ranula*, a colorful description of a cyst that has dissected through the connective tissue stroma connecting the two bellies of the mylohyoid muscle.

Sialolithiasis and Nonspecific Sialadenitis. **Nonspecific bacterial sialadenitis, involving the submandibular and other major salivary glands, is common and most often caused by infection with *S. aureus* and *Streptococcus viridans* following ductal obstruction by stones *(sialolithiasis).*** Stone formation is sometimes related to obstruction of salivary gland orifices by impacted food debris or local edema after injury, but no underlying cause is detected in many cases. Reduced secretion of saliva, as sometimes occurs in patients receiving long-term phenothiazines, may predispose to secondary bacterial invasion. Similarly, decreased salivary secretions due to dehydration may facilitate bacterial

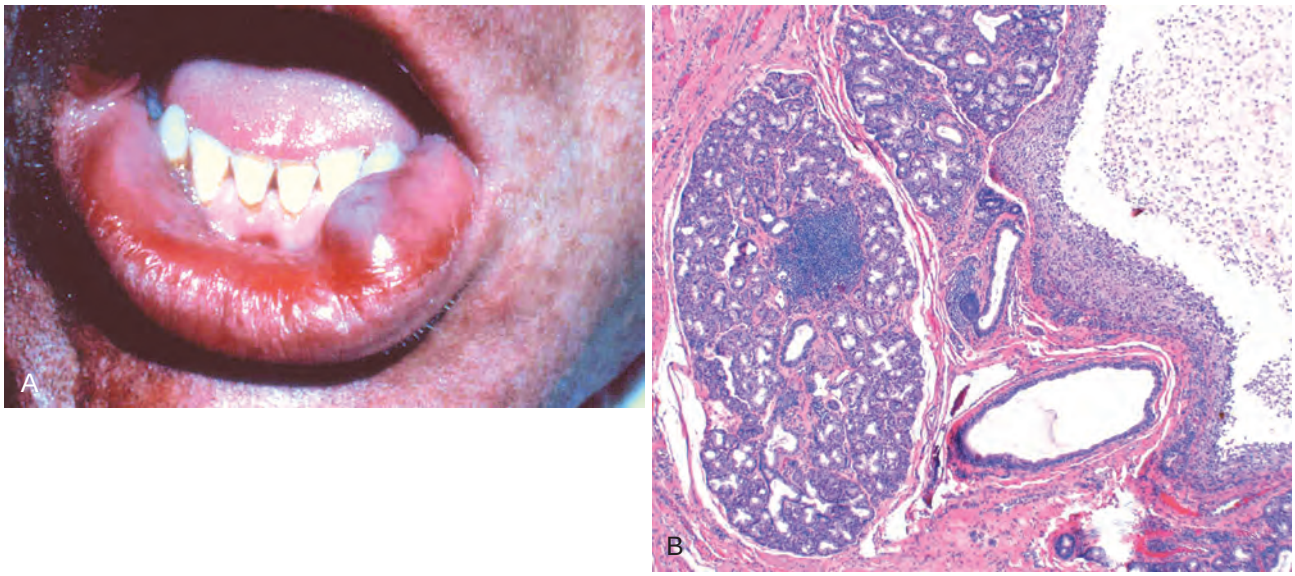


Figure 16.16 Mucocele. (A) Fluctuant fluid-filled lesion on the lower lip subsequent to trauma. (B) Cystlike cavity filled with mucinous material and lined by organizing granulation tissue.

suppurative parotitis, particularly in elderly patients with a recent history of major thoracic or abdominal surgery.

Obstructive processes and bacterial invasion lead to inflammation of the affected glands that is often interstitial or, when induced by staphylococcal or other pyogens, associated with overt suppurative necrosis and abscess formation. Inflammation causes painful enlargement and, sometimes, a purulent ductal discharge. Disease is almost always unilateral and involves a single gland.

NEOPLASMS

Despite their relatively simple morphology, the salivary glands give rise to more than 30 histologically distinct tumors. A classification and the relative incidence of benign and malignant tumors is listed in Table 16.5; not included are the rare benign and malignant mesenchymal neoplasms. Despite this diversity, a small number of neoplasms makes up more than 90% of salivary gland tumors. These will be the focus of the discussion that follows.

Salivary gland neoplasms represent less than 2% of all tumors in humans. About 65% to 80% arise within the parotid, 10% in the submandibular gland, and the remainder in the minor salivary glands, including the sublingual glands. The majority of parotid tumors are benign, but 40% of submandibular, 50% of minor salivary gland, and 70% to 90% of sublingual tumors are malignant. Thus, in general terms, the malignant potential of salivary gland tumors is inversely proportional to gland size, with smaller glands harboring greater numbers of cancers.

Salivary gland tumors occur in adults, with only approximately 5% occurring before 16 years of age. There is a slight female predominance overall, but this can vary with histology. For example, Warthin tumors are much more common in males, perhaps reflecting the historically higher prevalence of smoking, a predisposing factor, among men. Whatever the histologic pattern, parotid gland neoplasms produce distinctive swellings in front of and below the ear. When first diagnosed, both benign and malignant lesions range from 4 to 6 cm in diameter and, except in the case of highly invasive malignant tumors, are mobile on palpation. Benign tumors are often present for many months to several years before coming to clinical attention, but cancers are detected more quickly due to their rapid growth. Despite these generalities, there are no reliable clinical criteria to differentiate benign from malignant lesions.

Table 16.5 Histologic Classification and Incidence of the Most Common Benign and Malignant Tumors of the Salivary Glands

Benign	Malignant
Pleomorphic adenoma (mixed tumor)	Mucoepidermoid carcinoma
Warthin tumor	Acinic cell carcinoma
Oncocytoma	Adenoid cystic carcinoma
Canalicular adenoma	Adenocarcinoma, not otherwise specified
Basal cell adenoma	Carcinoma ex pleomorphic adenoma
Other adenomas	Squamous cell carcinoma
Ductal papillomas	Other carcinomas

Pleomorphic Adenoma

Pleomorphic adenomas are the most common salivary gland neoplasms. They represent about 60% of tumors in the parotid, are less common in the submandibular glands, and are relatively rare in the minor salivary glands. Pleomorphic adenomas are benign tumors that consist of a mixture of ductal (epithelial), myoepithelial, and mesenchymal cells, which explains why they are also termed *mixed tumors.*

Little is known about the origin of pleomorphic adenomas, but radiation exposure increases risk. Equally uncertain is the histogenesis of the various components. One view is that all tumor elements, including those that appear mesenchymal, are of myoepithelial or ductal reserve cell (stem cell) origin. Many cases are associated with chromosomal rearrangements that induce overexpression of PLAG1, a transcription factor that promotes expression of genes that increase cell growth, including components of growth factor receptor signaling pathways. Mutations of the *HMGA2* gene, which encodes a DNA-binding protein, are associated with many of the cases that lack PLAG overexpression.

MORPHOLOGY

Most pleomorphic adenomas present as rounded, well-demarcated masses that rarely exceed 6 cm in greatest dimension (Fig. 16.17). Although encapsulated, in some locations (particularly the palate) the capsule of pleomorphic adenomas is not fully developed, and expansile growth produces protrusions into the surrounding gland. These may lead to recurrences if the tumor is merely enucleated. The cut surface is gray-white with myxoid and blue translucent areas of chondroid (cartilage-like) stroma.

The dominant histologic feature is morphologic heterogeneity. The epithelial elements resembling ductal cells or myoepithelial cells are arranged as ducts, acini, irregular tubules, strands, or sheets of cells. These elements are typically dispersed within a background of loose myxoid and hyaline tissue containing islands of cartilage and, rarely, foci of bone (Fig. 16.18). Sometimes the epithelial cells form well-developed ducts lined by cuboidal to columnar cells with an underlying layer of deeply chromatic, small myoepithelial cells. In other instances, there may be strands or sheets of myoepithelial cells. Islands of well-differentiated squamous epithelium may also be present. In most cases, there is no epithelial dysplasia or mitotic activity. Tumors behave similarly whether they are primarily composed of epithelial or mesenchymal elements.

Pleomorphic adenomas present as painless, slow-growing, mobile, discrete masses within the parotid or submandibular areas or in the buccal cavity. The rate of recurrence following parotidectomy is about 4%. Recurrences occur months to years after surgery. In contrast, recurrence after simple enucleation approaches 25%, likely due to tumor extension beyond the capsule.

Rates of malignant transformation in pleomorphic adenoma correlate with age of the lesion; 2% of tumors that have been present for less than 5 years harbor cancers, but this increases to 10% for those present for more than 15 years. Cancers are usually adenocarcinomas or undifferentiated carcinomas. They are highly infiltrative and may completely

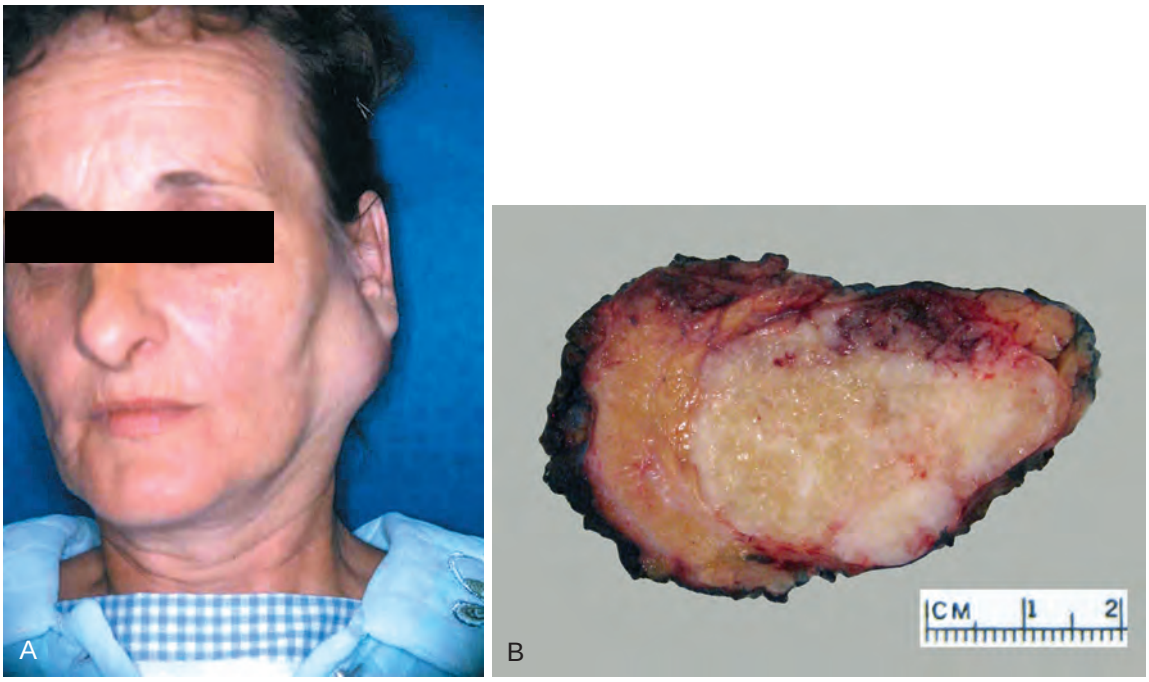


Figure 16.17 Pleomorphic adenoma. (A) Slowly enlarging neoplasm in the parotid gland of many years' duration. (B) Grossly, this representative cross-section of a parotidectomy specimen shows a circumscribed, yellow-white tumor and adjacent normal salivary gland tissue on the left.

replace the precursor lesion, making it difficult to substantiate the diagnosis of carcinoma arising in pleomorphic adenoma, which requires the presence of residual benign pleomorphic adenoma. Regrettably, when they appear, these cancers are among the most aggressive of all salivary gland neoplasms, with mortality rates of 30% to 50% at 5 years.

Warthin Tumor (Papillary Cystadenoma Lymphomatosum)

The *Warthin tumor* is the second most common salivary gland neoplasm. It is unique in that it arises almost exclusively in the parotid gland. Warthin tumors are benign and occur more commonly in males than in females, usually in the fifth to seventh decades of life. The risk is increased eightfold in smokers. Most Warthin tumors are unifocal, but about 10% are multifocal and 10% bilateral.

MORPHOLOGY

Warthin tumors are round to oval encapsulated masses, 2 to 5 cm in diameter, and readily palpable within the superficial parotid gland. Transection reveals a pale gray surface punctuated by narrow cystic or cleftlike spaces filled with mucinous or serous secretions and frequently narrowed by polypoid projections of lymphoepithelial elements. The lining is composed of a double layer of oncocytic cells; the innermost layer is columnar, while cuboidal cells occupy

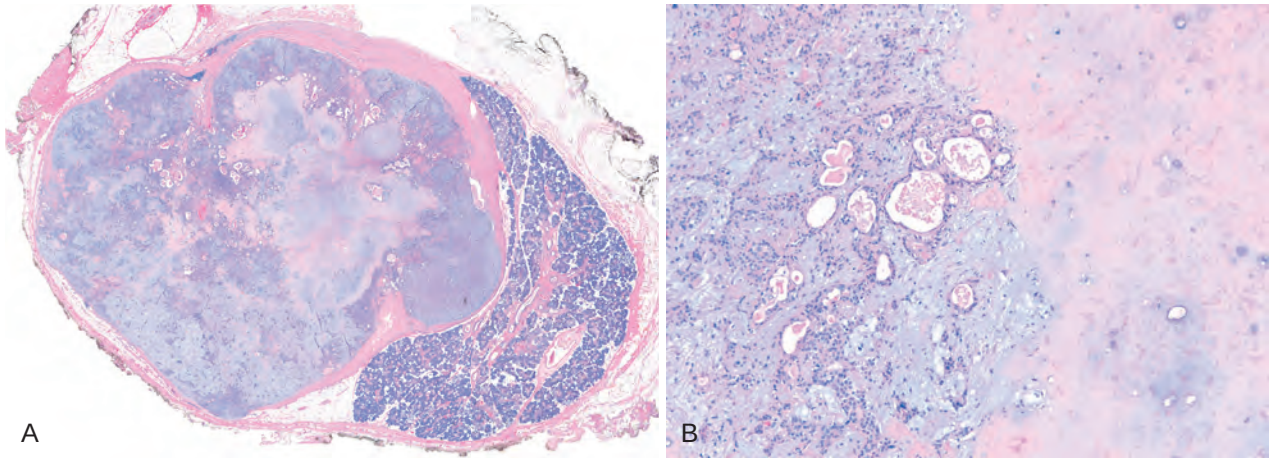


Figure 16.18 Pleomorphic adenoma. (A) Low-power histologic view showing a well-demarcated tumor with adjacent normal salivary gland parenchyma on the right. (B) High-power view showing epithelial and myoepithelial cells forming ducts *(left)* and a chondromyxoid matrix *(right)*.

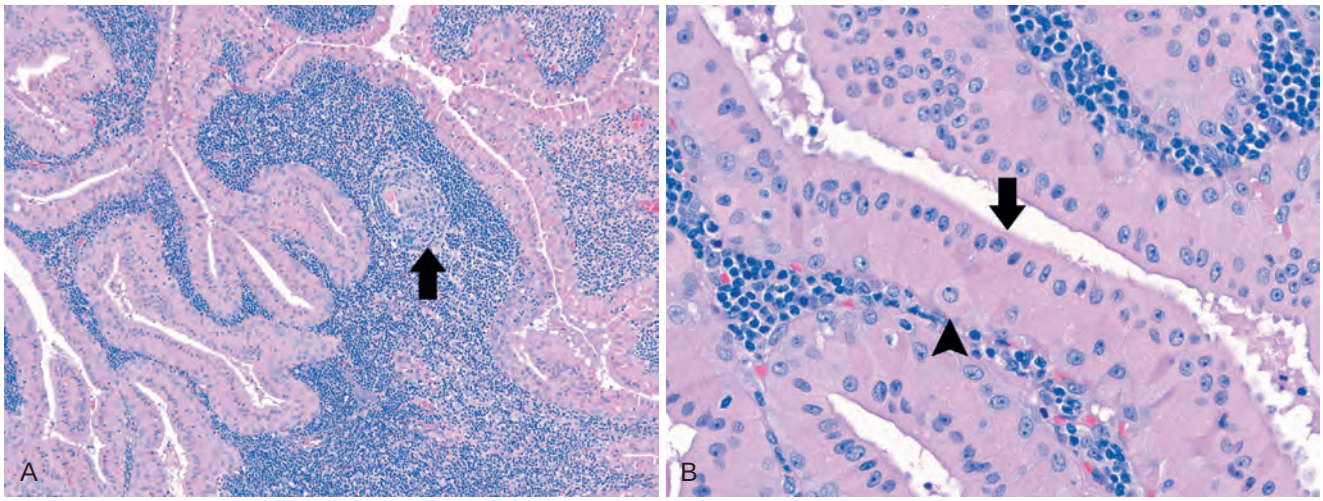


Figure 16.19 Warthin tumor. (A) Low-power view showing epithelial and lymphoid elements. Note the lymphoid follicular germinal center beneath the epithelium *(arrow)*. (B) The epithelium consists of a double layer of oncocytic epithelial cells on a lymphoid stroma, with a columnar luminal layer *(arrow)* associated with a cuboidal layer *(arrowhead)*.

the outer layer (Fig. 16.19). The term **oncocyte** (or swollen cell in Greek) refers to large cells containing numerous mitochondria that impart a granular, eosinophilic appearance to the cytoplasm and have large nuclei with prominent nucleoli. Secretory cells dispersed in the inner layer account for secretions within the dilated lumen. Foci of squamous metaplasia may be present.

The histogenesis of Warthin tumors is debated; it is not clear whether the epithelium is truly neoplastic (monoclonal) or is a reactive proliferation (polyclonal). Regardless, the epithelial cells are thought to secrete chemoattractants for the lymphoid cells, which are reactive. Although Warthin tumors have occasionally arisen within cervical lymph nodes, these neoplasms are benign and have low recurrence rates of only 2% after resection.

Mucoepidermoid Carcinoma

Mucoepidermoid carcinoma is the most common primary malignant tumor of salivary glands, representing about 15% of all salivary gland tumors. Most (60% to 70%) occur in the parotids, but mucoepidermoid carcinomas account for a large fraction of salivary gland neoplasms in the other glands, particularly the minor salivary glands. More than one-half of cases are associated with a balanced (11;19)(q21;p13) chromosomal translocation that creates a fusion gene composed of portions of the *CRTC1* (also known as *MECT1*) and *MAML2* genes. The CRTC1-MAML2 fusion protein is believed to play a key role in the genesis of this tumor, possibly by perturbing Notch and cAMP-dependent signaling pathways.

MORPHOLOGY

Mucoepidermoid carcinomas can grow as large as 8 cm in diameter. Although they appear circumscribed, well-defined capsules are not present, and the tumors are often infiltrative at the margins. On transection, mucoepidermoid carcinomas are pale, gray-white and frequently contain small, mucin-containing cysts. Histology demonstrates cords, sheets, or cystic configurations of squamous, mucous, or intermediate cells that have squamous features and small to large mucus-filled vacuoles (Fig. 16.20A–B). The cytology may be monotonous and bland or, alternatively, highly anaplastic and unmistakably malignant. Accordingly, mucoepidermoid carcinomas are subclassified as low, intermediate, or high grade.

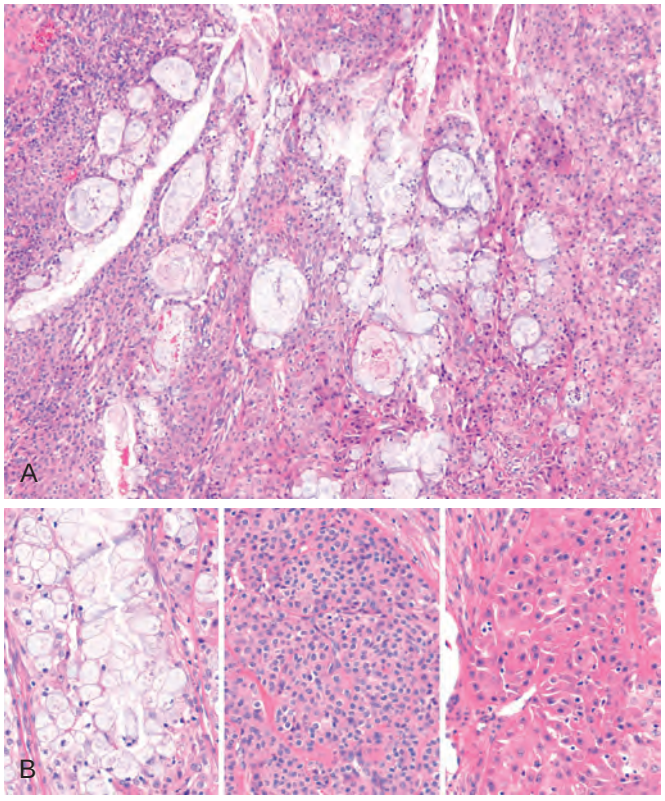


Figure 16.20 Mucoepidermoid carcinoma. (A) Low-power view showing sheets and microcysts containing cells of varying morphologies. (B) These cells vary from mucinous *(left)*, to intermediate *(center)*, to squamous *(right)*.

The clinical course and prognosis of mucoepidermoid carcinomas reflects the histologic grade. Low-grade tumors may invade locally and recur in about 15% of cases, but they rarely metastasize. The 5-year survival rate is more than 90%. In contrast, high-grade neoplasms and, to a lesser extent, intermediate-grade tumors are invasive and difficult to excise. Up to 30% recur and 30% metastasize to distant sites, yielding a 5-year survival of only 50%.

OTHER SALIVARY GLAND TUMORS

Two less common neoplasms merit brief description: adenoid cystic carcinoma and acinic cell carcinoma.

Adenoid cystic carcinomas. 50% of these relatively uncommon tumors occur in the minor salivary glands (particularly the palatine glands). Among major salivary glands, parotid and submandibular glands are most commonly affected. Adenoid cystic carcinomas also occur in the nose, sinuses, upper airways, lung, breast, and other sites. Although the pathogenesis is not defined, *MYB-NFIB* gene rearrangements are present in a subset of adenoid cystic carcinomas.

MORPHOLOGY

Grossly, adenoid cystic carcinomas are small, poorly encapsulated, infiltrative, gray-pink lesions composed of small cells with dark, compact nuclei and scant cytoplasm. The tumor cells are organized in a cribriform growth pattern that resembles swiss cheese. The spaces between the tumor cells are often filled with hyaline material thought to represent excess basement membrane (Fig. 16.21A). Less common histologic patterns are tubular and solid.

Adenoid cystic carcinomas are unpredictable, slow-growing tumors that tend to invade perineural spaces (see Fig. 16.21B). They regularly recur, and 50% or more eventually disseminate to distant sites including bone, liver, and brain, sometimes decades after attempted removal. Thus, although the 5-year survival rate is about 60% to 70%, it drops to about 30% at 10 years and 15% at 15 years. Adenoid cystic carcinomas arising in the minor salivary glands tend to have a poorer prognosis than those arising in the parotid glands.

Acinic cell carcinomas. These represent only 2% to 3% of salivary gland tumors and are generally small and well-circumscribed. Most develop in the parotid glands, with the remainder arising in the submandibular glands. Minor salivary glands, which normally have only a scant number of serous acinic cells, are rarely involved. Like Warthin tumors, acinic cell carcinomas can be bilateral or multicentric.

MORPHOLOGY

Histologically, acinic cell carcinomas are composed of cells that resemble normal serous acinar cells of the salivary glands, with small and round nuclei and variable morphology (Fig. 16.22). The cytoplasm typically contains purple granules (**zymogen granules**

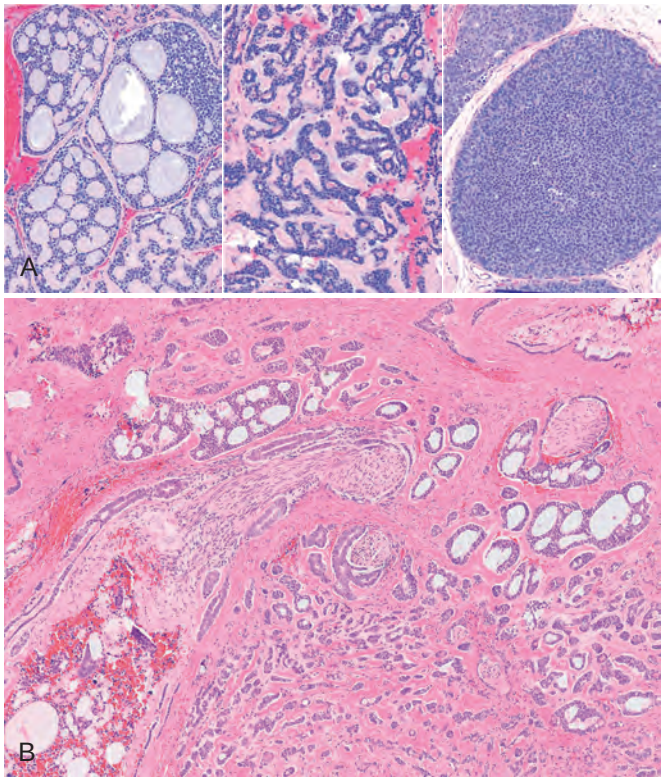


Figure 16.21 Adenoid cystic carcinoma. (A) Three typical growth patterns including cribriform *(left)*, tubular *(center)*, and solid *(right)*. (B) Perineural invasion by adenoid cystic carcinoma.

recapitulating the normal serous acinar digestive enzymes), but cells can also be clear or vacuolated. The neoplastic cells are organized into sheets, microcysts, glands, follicles, or papillae. Tumors with a sheetlike growth pattern, nuclear pleomorphism, mitoses, and necrosis are referred to as dedifferentiated or having undergone high-grade transformation.

The clinical course of acinic cell carcinomas reflects the degree of cytologic pleomorphism. Although recurrence is uncommon after resection, 10% to 15% of tumors metastasize to lymph nodes. The survival rate is approximately 90% at 5 years and 60% at 20 years.

KEY CONCEPTS

DISEASES OF SALIVARY GLANDS

- Sialadenitis is inflammation of the salivary glands and can be caused by trauma, infection, or autoimmune disease.
- Mucoceles are the result of trauma to or blockage of a salivary gland duct that allows saliva leakage into surrounding connective tissue.
- Pleomorphic adenoma is a benign, slow-growing neoplasm composed of a heterogenous mixture of epithelial and mesenchymal cells.
- Mucoepidermoid carcinoma is a malignant neoplasm of variable biological aggressiveness that is composed of a mixture of squamous and mucous cells.

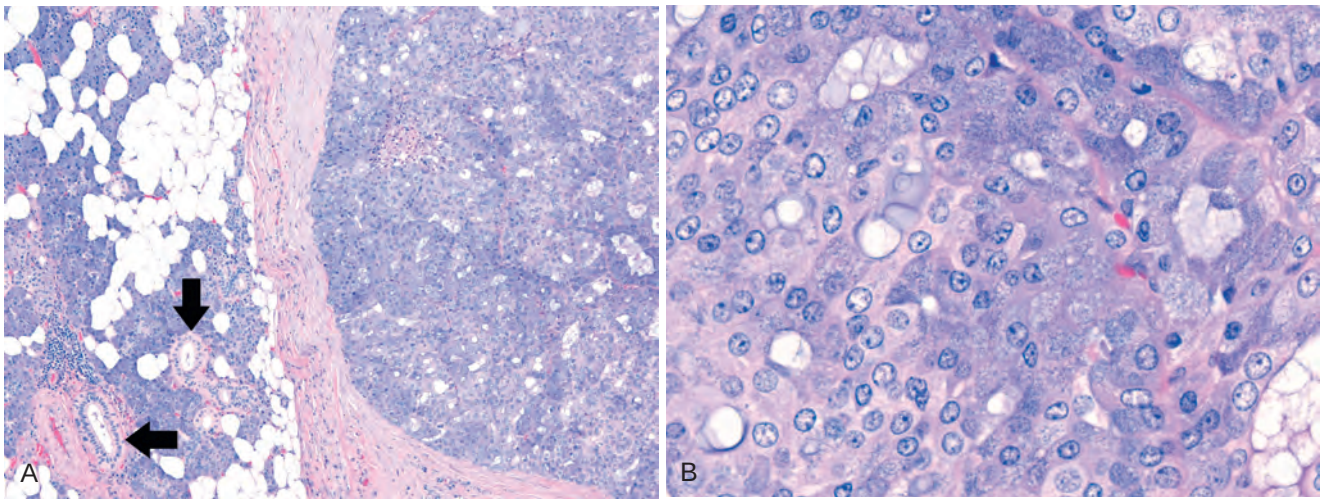


Figure 16.22 Acinic cell carcinoma. (A) Low-power view showing normal serous parotid tissue with admixed fat and ducts *(arrows)* on the left and sheets of serous cells in acinic cell carcinoma on the right. (B) High-power image of tumor cells with purple cytoplasmic granules and monotonous round nuclei.

SUGGESTED READINGS

Comprehensive genomic characterization of head and neck squamous cell carcinomas, *Nature* 517(7536):576–582, 2015. *[A comprehensive TCGA manuscript that defines the mutational landscape of head and neck cancer].*

El-Naggar AK, Chan JKC, Grandis JR et al: Tumors of salivary glands, Chapter 7. In *WHO classification of head and neck tumors*, ed 4, 2017, WHO, IARC, pp 159–201. *[This chapter provides a comprehensive up-to-date summary of the current classification of benign and malignant salivary gland tumors].*

Gillison ML, Broutian T, Pickard RK et al: Human papilloma virus and rising oropharyngeal cancer incidence in the United States, *J Clin Oncol* 29:4294, 2011. *[A study using the Surveillance, Epidemiology, and End Results (SEER) Residual Tissue Repositories Program, which demonstrated that increases in the population-level incidence and survival of oropharyngeal cancers in the United States since 1984 are caused by HPV infection].*

Lingen MW, Xiao W, Schmitt A et al: Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas, *Oral Oncol* 49:1, 2013. *[This manuscript estimated the etiologic fraction for HPV among consecutive, incident oral cavity squamous cell carcinomas at four North American hospitals and determined that the etiologic fraction of HPV-positive cases was significantly lower than the oropharynx].*

Neville BW, Damm DD, Allen CM et al: *Oral and maxillofacial pathology*, ed 4, 2016, Elsevier. *[A combined clinical and pathologic approach in an atlas-style text describing the major diseases of the oral cavity].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 17

The Gastrointestinal Tract

CHAPTER CONTENTS

The gastrointestinal (GI) tract is a hollow tube extending from the oral cavity to the anus that consists of anatomically distinct segments including the esophagus, stomach, small intestine, colon, rectum, and anus. Each of these segments has unique, complementary, and highly integrated functions, which together serve to regulate the intake, processing, and absorption of ingested nutrients and the disposal of waste products. The regional variations in structure and function are reflected in diseases of the GI tract, which often affect one or another segment preferentially. Accordingly, following consideration of several important congenital abnormalities, the discussion is organized anatomically. Disorders affecting more than one segment of the GI tract, such as Crohn disease, are discussed with the region that is involved most frequently.

Congenital Abnormalities

Depending on both the nature and the timing of the insult during gestation, a variety of developmental anomalies can affect the GI tract. Importantly, because many organs develop simultaneously during embryogenesis and are susceptible to the same insults, the presence of congenital GI anomalies should prompt evaluation of other organs.

ATRESIA, FISTULAE, AND DUPLICATIONS

Atresia, fistulae, and duplications may occur in any part of the GI tract. When present within the esophagus they are discovered shortly after birth, usually due to regurgitation during feeding. Most of the lesions are incompatible with survival without prompt surgical repair. Absence, or agenesis, of the esophagus is extremely rare, but atresia, in which development is incomplete, is present in 1 to 5 infants per 10,000 live births. In atresia, a segment of the esophagus does not develop, leaving only a thin, noncanalized cord and causing a mechanical obstruction (Fig. 17.1A). Atresia occurs most commonly at or near the tracheal bifurcation and is usually associated with a fistula connecting the upper or lower esophageal segments to a bronchus or the trachea (Fig. 17.1B). In other cases a fistula can be present without atresia (Fig. 17.1C). Any of the three primary forms of tracheoesophageal fistula can lead to aspiration, suffocation, pneumonia, and severe fluid and electrolyte imbalances. Developmental abnormalities of the esophagus are associated with a range of congenital abnormalities, including VACTERL (malformations including vertebral, anal, cardiac, tracheoesophageal, renal, and limb defects); TACRD (tracheal agenesis/atresia, complex congenital cardiac abnormalities, radial ray defects, and duodenal atresia); and other abnormalities involving the heart, lung, great vessels, genitourinary tract, and anus. The last-mentioned includes imperforate anus, which is due to a failure of the cloacal diaphragm to involute.

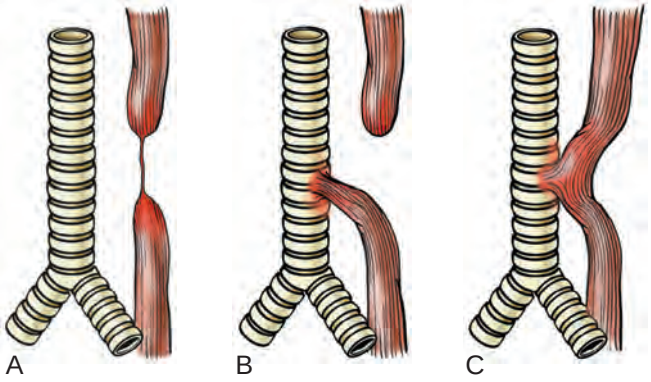


Figure 17.1 Esophageal atresia and tracheoesophageal fistula. (A) Blind upper and lower esophagus with thin cord of connective tissue linking the two segments. (B) Blind upper segment with fistula between lower segment and trachea. (C) Fistula (without atresia) between patent esophagus and trachea. The developmental anomaly shown in (B) is the most common. (Modified from Morson BC, Dawson IMP, editors: *Gastrointestinal Pathology*, Oxford, 1972, Blackwell Scientific Publications, p 8.)

Stenosis is an incomplete form of atresia in which the lumen is markedly reduced in caliber as a result of fibrous thickening of the wall. This results in either partial or complete obstruction. In addition to congenital forms, stenosis can be acquired as a consequence of inflammatory scarring, such as that caused by chronic gastroesophageal reflux, irradiation, systemic sclerosis, or caustic injury. It can involve any part of the GI tract, but because of their smaller caliber, clinically significant stenosis most often affects the esophagus and small intestine.

DIAPHRAGMATIC HERNIA, OMPHALOCELE, AND GASTROSCHISIS

Diaphragmatic hernia occurs when incomplete formation of the diaphragm allows the abdominal viscera to herniate into the thoracic cavity, most commonly on the left side. When severe, the space-filling effect of displaced viscera can lead to potentially fatal pulmonary hypoplasia.

Omphalocele occurs when the extraembryonic gut fails to return to the abdominal cavity and closure of the abdominal musculature is incomplete. As a result, abdominal viscera (including the liver) herniate ventrally into a membrane covered by amnion and peritoneum separated by Wharton jelly. Omphalocele is often associated with other birth defects as well as chromosomal abnormalities.

Gastroschisis is similar to omphalocele except that it involves all the layers of the abdominal wall, from the peritoneum to the skin, is usually limited to the intestine, and occurs as an isolated defect without other abnormalities. The incidence of gastroschisis is rising, possibly due to environmental factors including smoking and exposure to agricultural chemicals. Surgical repair of omphalocele and gastroschisis is generally successful.

ECTOPIA

Ectopic tissues (developmental rests) are common in the GI tract. The most frequent site of ectopic gastric mucosa is the upper third of the esophagus, where it is referred to as an inlet patch. While generally asymptomatic, acid released by gastric mucosa within the esophagus can result in dysphagia, esophagitis, Barrett esophagus, or, rarely, adenocarcinoma. Ectopic pancreatic tissue occurs less frequently and is found in the esophagus or stomach. These nodules are most often asymptomatic but may produce local damage and inflammation. When ectopic pancreatic tissue is present in the pylorus, inflammation and scarring may lead to obstruction. Because rests may be present within any layer of the gastric wall, they can mimic invasive cancer. Gastric heterotopia, small patches of ectopic gastric mucosa in the small bowel or colon, may present with occult blood loss due to peptic ulceration of adjacent mucosa.

MECKEL DIVERTICULUM

A true diverticulum is a blind outpouching of the alimentary tract that communicates with the lumen and includes all three layers of the bowel wall. The most common true diverticulum, and the most common congenital anomaly of the GI tract, is Meckel diverticulum, which occurs in the ileum. Meckel diverticulum occurs due to failed involution of the vitelline duct, which connects the lumen of the developing gut to the yolk sac. This solitary diverticulum extends from the antimesenteric side of the bowel (Fig. 17.2). The "rule of 2's" is often used to help remember characteristics of Meckel diverticula, which:

- Occur in approximately 2% of the population
- Are generally present within 2 feet (60 cm) of the ileocecal valve
- Are approximately 2 inches (5 cm) long
- Are twice as common in males
- Are most often symptomatic by age 2 (only approximately 4% are ever symptomatic).

The mucosal lining of Meckel diverticula may resemble that of normal small intestine, but ectopic pancreatic or gastric tissue may also be present. The latter may secrete acid, cause peptic ulceration of adjacent small intestinal mucosa, and present with occult bleeding or abdominal pain resembling acute appendicitis or obstruction.

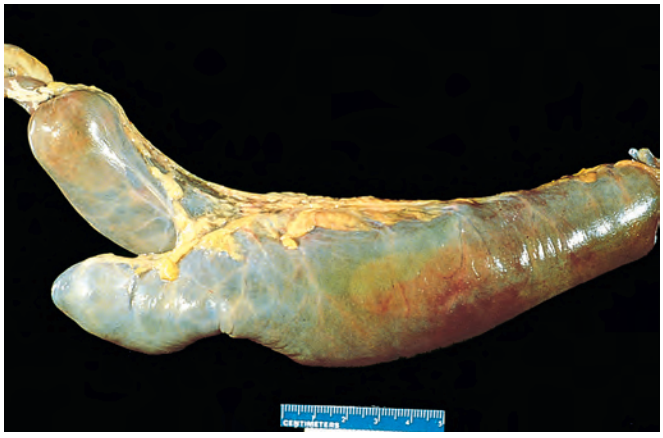


Figure 17.2 Meckel diverticulum. The blind pouch is located on the antimesenteric side of the small bowel.

Virtually all other diverticula are acquired and either lack muscularis entirely or have an attenuated muscularis propria. These are, technically, pseudodiverticula, but are generally referred to simply as divericula. They occur most often in the sigmoid colon (discussed later).

PYLORIC STENOSIS

Pyloric stenosis may be either congenital or acquired. The most common congenital form is congenital hypertrophic pyloric stenosis. It is three to five times more common in males and has an overall incidence of 1 per 300 to 900 live births. Monozygotic twins have a high rate of concordance; risk is increased to a lesser degree in dizygotic twins and siblings of affected individuals. Turner syndrome and trisomy 18 also confer an increased risk. While the underlying mechanisms are not understood, erythromycin or azithromycin exposure, either orally or via mother's milk, in the first 2 weeks of life is associated with this disorder.

Congenital hypertrophic pyloric stenosis generally presents between the third and sixth weeks of life as new-onset regurgitation, projectile nonbilious vomiting after feeding, and frequent demands for refeeding. Physical examination reveals a firm, ovoid, 1- to 2-cm abdominal mass in up to 90% of cases; abnormal left-to-right hyperperistalsis just before vomiting is also common. Diagnosis is primarily by ultrasonography. Anatomically, the gastric outflow tract is obstructed by hyperplasia of the pyloric muscularis propria. Edema and inflammatory changes in the mucosa and submucosa may aggravate the narrowing. Surgical splitting of the muscularis (myotomy) is curative.

Acquired pyloric stenosis occurs in adults as a consequence of antral gastritis or peptic ulcers close to the pylorus. Carcinomas of the distal stomach and pancreas may also narrow the pyloric channel due to fibrosis or malignant infiltration.

HIRSCHSPRUNG DISEASE

Hirschsprung disease causes functional obstruction of the colon due to failure of ganglion cells to migrate to the wall of the colon resulting from a mutation in the receptor tyrosine kinase. Hirschsprung disease occurs in approximately 1 in 5000 live births. It may be isolated or occur in combination with other developmental abnormalities; 10% of all cases occur in children with Down syndrome, and serious neurologic abnormalities are present in 5% of infants with Hirschsprung disease.

Pathogenesis

The enteric neuronal plexus develops from neural crest cells that migrate into the bowel wall during embryogenesis. Hirschsprung disease, also known as congenital aganglionic megacolon, results when the migration of neural crest cells from cecum to rectum is arrested prematurely or when the ganglion cells undergo premature death. This produces a

distal intestinal segment that lacks both the Meissner submucosal plexus and the Auerbach myenteric plexus, termed aganglionosis. Coordinated peristaltic contractions are absent, and functional obstruction occurs, resulting in dilation proximal to the affected segment.

Heterozygous loss-of-function mutations in the receptor tyrosine kinase *RET* account for the majority of familial and approximately 15% of sporadic Hirschsprung disease cases. The association between Down syndrome and Hirschsprung disease may be linked to the Down syndrome cell adhesion molecule gene on chromosome 21, overexpression of which leads to neural defects in experimental models. In patients without trisomy 21, germline mutations in several genes that are required for normal gut innervation have been implicated, the most common of which involve *RET* and *EDNRB* or *EDN3* (which encode a receptor-ligand pair). These proteins appear to participate in signaling pathways that regulate development of the enteric nervous system. Defects in known genes cannot explain all cases, however, and penetrance in those carrying pathogenic mutations is incomplete, suggesting that modifying genes or environmental factors are also important. Some of these modifying genes may be sex-linked, since the disease is more common in males, but, when present in females, the affected aganglionic segment tends to be longer.

MORPHOLOGY

Diagnosis of Hirschsprung disease requires histologic confirmation that ganglion cells are absent within the affected segment. In addition to their characteristic morphology in hematoxylin and eosin–stained sections, ganglion cells can be identified using immunohistochemical stains for acetylcholinesterase.

The rectum is always affected, but the length of the additional involved segments varies widely. In most cases the abnormality extends to the sigmoid colon, but in severe cases, the entire colon may be involved. Disease limited to the rectosigmoid is termed *short-segment Hirschsprung disease*, while cases with more proximal extension are termed *long-segment Hirschsprung disease*.

The aganglionic region may have a grossly normal or contracted appearance. In contrast, the normally innervated proximal colon undergoes progressive dilation (Fig. 17.3) and may become massively distended **(megacolon)** as a result of the distal obstruction. The colonic wall may be stretched to the point of rupture, which occurs most frequently near the cecum. Mucosal inflammation or shallow ulcers may also be present in normally innervated segments, making gross identification of the extent of aganglionosis difficult. Hence, intraoperative histopathologic analysis (frozen section) is commonly used to confirm the presence of ganglion cells at the anastomotic site.

Clinical Features

Hirschsprung disease typically presents with a failure to pass meconium in the immediate postnatal period. Obstruction or constipation follows, often with visible but ineffective peristalsis. When only a few centimeters of rectum are involved, some stool may be passed, obscuring the diagnosis. More often, the obstruction is obvious and is marked by abdominal distention and bilious vomiting as luminal material backs up into the proximal colon. The major threats to life are enterocolitis, fluid and electrolyte disturbances, perforation, and peritonitis. Treatment consists of surgical resection of the aganglionic segment and anastomosis of the normal proximal colon to the rectum. Even after successful surgery, it may take years to attain normal bowel function and continence.

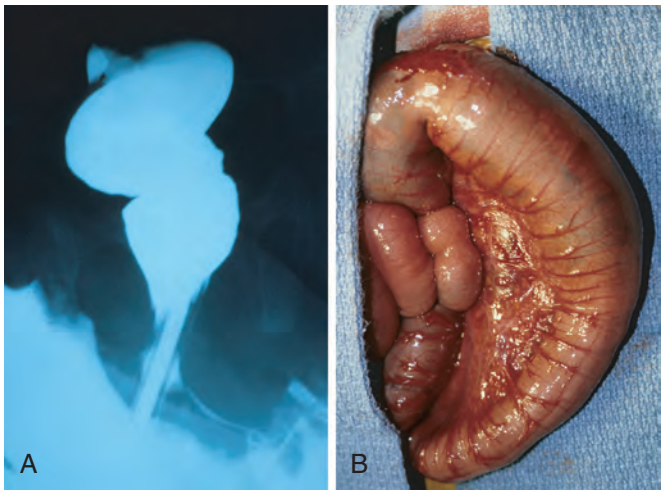


Figure 17.3 Hirschsprung disease. (A) Preoperative barium enema study showing constricted rectum (bottom of the image) and dilated sigmoid colon. (B) Corresponding intraoperative photograph showing constricted rectum and dilation of the sigmoid colon. (Courtesy Dr. Aliya Husain, The University of Chicago, Chicago, Ill.)

In contrast to the congenital megacolon of Hirschsprung disease, acquired megacolon may occur at any age as a result of Chagas disease, obstruction by a neoplasm or inflammatory stricture, a complication of ulcerative colitis, visceral myopathy, or in association with psychosomatic disorders. Of these, only Chagas disease (discussed later) shares the same pathophysiology as Hirschsprung disease; the loss of ganglion cells.

KEY CONCEPTS

CONGENITAL MALFORMATIONS OF THE GASTROINTESTINAL TRACT

- The GI tract is a common site of developmental abnormalities, often in association with congenital defects of other organ systems.
- Atresia and fistulae are developmental anomalies that typically present at birth. Imperforate anus is the most common form of congenital intestinal atresia, while the esophagus is the most common site of fistulization.
- Stenosis may be developmental or acquired. Both forms are characterized by a thickened wall and partial or complete luminal obstruction. Acquired forms are often due to inflammatory scarring.
- Diaphragmatic hernia is characterized by incomplete diaphragm development and herniation of abdominal organs into the thorax, often producing pulmonary hypoplasia. Omphalocele and gastroschisis refer to ventral herniation of abdominal organs.
- Ectopia refers to normally formed tissues at an abnormal site. This is common in the GI tract, with ectopic gastric mucosa in the upper third of the esophagus being the most common form.
- Meckel diverticulum is a true diverticulum, defined by the presence of all three layers of the bowel wall, that stems from failed

involution of the vitelline duct. It is a frequent site of gastric ectopia, which may result in peptic injury and occult bleeding.
- Congenital hypertrophic pyloric stenosis presents between the third and sixth weeks of life and is more common in males.
- Hirschsprung disease is caused by the failure of neural crest–derived ganglion cells to migrate into the distal colon. The defect, which always involves the rectum, extends proximally for variable distances.

Esophagus

The esophagus develops from the cranial portion of the foregut and is recognizable by the third week of gestation. It is a hollow, highly distensible muscular tube that extends from the epiglottis in the pharynx to the gastroesophageal junction. Acquired diseases of the esophagus run the gamut from highly lethal cancers to the heartburn of gastroesophageal reflux, which may be chronic and incapacitating or merely an occasional annoyance.

ESOPHAGEAL OBSTRUCTION

The primary function of the esophagus is to deliver ingested solid food and fluids to the stomach. This can be impeded by physical or functional obstruction. The latter results from disruption of coordinated peristalsis after swallowing. Esophageal dysmotility falls into three principal patterns.

- *Nutcracker esophagus* describes functional obstruction of the esophagus by intense, high amplitude, uncoordinated contractions of inner circular and outer longitudinal smooth muscle. Barium swallow can be normal or have a nutcracker-like appearance, with narrow and distended regions. Manometry is, however, required for diagnosis.
- *Diffuse esophageal spasm*, also known as corkscrew esophagus due to the appearance on barium swallow, is characterized by repetitive, simultaneous contractions of the distal esophageal smooth muscle. Unlike nutcracker esophagus, these contractions are of normal amplitude.
- *Lower esophageal sphincter dysfunction,* including high resting pressure or incomplete relaxation, may be present as an isolated anomaly or accompany nutcracker esophagus or diffuse esophageal spasm.

Because wall stress is increased, esophageal dysmotility may result in development of small diverticula, particularly in the epiphrenic region immediately above the lower esophageal sphincter. Similarly, impaired relaxation and spasm of the cricopharyngeus muscle after swallowing can result in increased pressure within the distal pharynx and development of a Zenker (pharyngoesophageal) diverticulum, located immediately above the upper esophageal sphincter. Zenker diverticula are uncommon, typically develop after age 50, and may reach several centimeters in size. When small they may be asymptomatic, but larger Zenker diverticula may accumulate significant amounts of food, producing a mass and symptoms that include regurgitation and halitosis. Mechanical obstruction caused by strictures or cancer typically presents as progressive dysphagia that begins with inability to swallow solids.

Benign esophageal stenosis is generally caused by fibrous thickening of the submucosa and is associated with atrophy of the muscularis propria and secondary epithelial damage. Although occasionally congenital, stenosis is most often due to inflammation and scarring that may be caused by chronic gastroesophageal reflux, irradiation, or caustic injury. In general, patients with functional obstruction or benign strictures maintain their appetite and weight. In contrast, malignant strictures are often associated with weight loss.

Esophageal mucosal webs are uncommon ledge-like mucosal protrusions. These usually occur in women older than age 40 and may be associated with gastroesophageal reflux, chronic graft-versus-host disease, or blistering skin diseases. In the upper esophagus, webs may be accompanied by iron deficiency anemia, glossitis, and cheilosis as part of *Plummer-Vinson syndrome.* Morphologically, esophageal webs are recognized as semi-circumferential lesions composed of a fibrovascular connective tissue and overlying epithelium. The main symptom of webs is partial obstruction, primarily of incompletely chewed food, and nonprogressive dysphagia.

Esophageal rings, or *Schatzki rings,* are similar to webs but are circumferential, are thicker, and include mucosa, submucosa, and occasionally hypertrophic muscularis propria. When present in the distal esophagus, above the gastroesophageal junction, they are termed *A rings* and are covered by squamous mucosa; in contrast, those located at the squamocolumnar junction of the lower esophagus are designated *B rings* and may have gastric cardia–type mucosa on their undersurface.

ACHALASIA

Achalasia is characterized by the triad of incomplete lower esophageal sphincter relaxation, increased lower esophageal sphincter tone, and aperistalsis of the esophagus. Symptoms include dysphagia, difficulty in belching, and chest pain. Although there is some increased risk for esophageal cancer, it is not considered great enough to warrant surveillance endoscopy.

Primary achalasia is rare and results from degeneration of nitric oxide–producing neurons that normally induce lower esophageal sphincter relaxation. Degenerative changes in the extraesophageal vagus nerve or the dorsal motor nucleus of the vagus may also occur. The cause is unknown; rare familial cases have been described.

Secondary achalasia may arise in Chagas disease, in which *Trypanosoma cruzi* infection causes destruction of the myenteric plexus, failure of peristalsis, and esophageal dilation. Duodenal, colonic, and ureteric myenteric plexuses can also be affected in Chagas disease. Other causes of achalasia-like disease include diabetic autonomic neuropathy; infiltrative

disorders such as malignancy, amyloidosis, and sarcoidosis; systemic sclerosis; or lesions of dorsal motor nuclei (e.g., following polio). Lower esophageal sphincter dysfunction also occurs in association with Down syndrome or as part of Allgrove (triple-A) syndrome, an autosomal recessive disorder characterized by achalasia, alacrima, and adrenocorticotropic hormone–resistant adrenal insufficiency. The association of some achalasia cases with remote herpes simplex virus 1 (HSV1) infection, linkage of immunoregulatory gene polymorphisms to achalasia, and occasional coexistence of Sjögren syndrome or autoimmune thyroid disease suggest that achalasia may also be driven by immune-mediated destruction of inhibitory esophageal neurons. Treatments for both primary and secondary achalasia aim to overcome the obstruction and include laparoscopic myotomy, pneumatic balloon dilation, and botulinum neurotoxin (Botox) injection to inhibit contraction-promoting cholinergic neurons.

ESOPHAGITIS AND RELATED DISORDERS

Lacerations

Longitudinal mucosal tears near the gastroesophageal junction, called Mallory-Weiss tears, are most often associated with severe retching or vomiting secondary to acute alcohol intoxication. Normally, a reflex relaxation of the gastroesophageal musculature precedes the antiperistaltic contractile wave associated with vomiting. It is speculated that this relaxation fails during prolonged vomiting, with the result that refluxing gastric contents overwhelm the gastric inlet and cause the esophageal wall to stretch and tear. The roughly linear lacerations of Mallory-Weiss syndrome are longitudinally oriented and range in length from millimeters to several centimeters. These tears usually cross the gastroesophageal junction and may also be located in the proximal gastric mucosa. Up to 10% of upper GI bleeding, which often presents as hematemesis (Table 17.1), is due to superficial esophageal lacerations such as those associated with Mallory-Weiss syndrome. These do not generally require surgical intervention, and healing tends to be rapid and complete. In contrast, Boerhaave syndrome is much less common but far more serious and is characterized by transmural tearing and rupture of the distal esophagus. This catastrophic event produces severe mediastinitis and generally requires surgical intervention. Because patients can present with severe chest pain, tachypnea, and shock, the initial differential diagnosis can include myocardial infarction.

Table 17.1 Esophageal Causes of Hematemesis

Lacerations (Mallory-Weiss syndrome)
Esophageal perforation (cancer or Boerhaave syndrome)
Varices (cirrhosis)
Esophageal-aortic fistula (usually with cancer)
Chemical and pill esophagitis
Infectious esophagitis (*Candida*, herpes)
Benign strictures
Vasculitis (autoimmune, cytomegalovirus)
Reflux esophagitis (erosive)
Eosinophilic esophagitis
Esophageal ulcers (many etiologies)
Barrett esophagus
Adenocarcinoma
Squamous cell carcinoma
Hiatal hernia

Chemical and Infectious Esophagitis

The stratified squamous mucosa of the esophagus may be damaged by a variety of irritants including alcohol, corrosive acids or alkalis, excessively hot fluids, and heavy smoking. Symptoms range from self-limited pain, particularly on swallowing, called *odynophagia,* to hemorrhage, stricture, or perforation in severe cases.

In children, esophageal chemical injury is often secondary to accidental ingestion of household cleaning products; severe damage may follow attempted suicide in adults. Less severe chemical injury to the esophageal mucosa can occur when medicinal pills lodge and dissolve in the esophagus, a condition termed *pill-induced esophagitis.* Iatrogenic esophageal injury may be caused by cytotoxic chemotherapy, radiation therapy, or graft-versus-host disease. The esophagus may also be involved by the desquamative skin diseases bullous pemphigoid, epidermolysis bullosa, and, rarely, Crohn disease.

Esophageal infections in otherwise healthy individuals are most often due to herpes simplex virus. Infections are more common in patients who are debilitated or immunosuppressed and may be caused by herpes simplex virus, cytomegalovirus (CMV), or fungal organisms. Among fungi, candidiasis is most common, although mucormycosis and aspergillosis also occur.

MORPHOLOGY

The morphology of chemical and infectious esophagitis varies with etiology. Dense infiltrates of neutrophils are usually present but may be absent following injury induced by chemicals (lye, acids, or detergent) that cause outright necrosis of the esophageal wall. Pill-induced esophagitis frequently occurs at the site of strictures that impede passage of luminal contents.

Esophageal irradiation causes damage similar to that seen in other tissues that is marked by intimal proliferation and luminal narrowing of submucosal and mural blood vessels. Concomitant mucosal damage is, in part, secondary to this radiation-induced vascular injury, as discussed in Chapter 9.

Infection by fungi or bacteria can either cause injury or complicate a preexisting ulcer. Nonpathogenic oral bacteria are frequently found in ulcer beds, while pathogenic bacteria and fungi, which account for about 10% of infectious esophagitis, may invade and cause mucosal necrosis. Candidiasis is characterized by adherent, gray-white **pseudomembranes** composed of densely matted fungal hyphae and inflammatory cells covering the esophageal mucosa.

The endoscopic appearance frequently indicates the cause of viral esophagitis. Herpesviruses typically cause punched-out ulcers (Fig. 17.4A). Biopsy specimens demonstrate viral nuclear inclusions in degenerating, multinucleate, epithelial cells at the margin of the ulcer (Fig. 17.4B). In contrast, CMV infection causes shallower ulcerations and is marked by characteristic nuclear and cytoplasmic

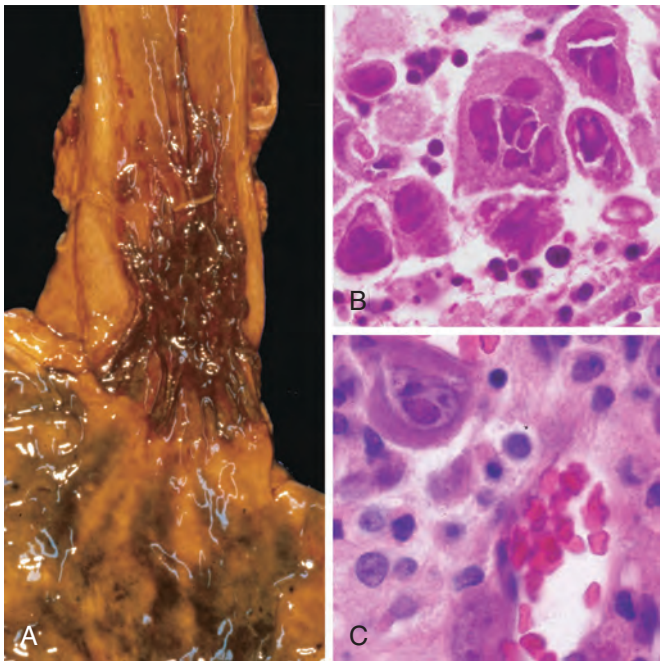


Figure 17.4 Viral esophagitis. (A) Postmortem specimen with multiple, overlapping herpetic ulcers in the distal esophagus. (B) Multinucleate squamous cells containing herpesvirus nuclear inclusions. (C) Cytomegalovirus-infected endothelial cells with nuclear and cytoplasmic inclusions.

inclusions within capillary endothelium and stromal cells (Fig. 17.4C). Although the histologic appearance is characteristic, immunohistochemical stains for virus-specific antigens are sensitive and specific ancillary diagnostic tools.

Histologic features of esophageal **graft-versus-host disease** are similar to those in the skin and include basal epithelial cell apoptosis, mucosal atrophy, and submucosal fibrosis without significant acute inflammatory infiltrates. The microscopic appearances of esophageal involvement in bullous pemphigoid, epidermolysis bullosa, and Crohn disease are also similar to those in the skin (Chapter 25).

Reflux Esophagitis

Reflux of gastric contents into the lower esophagus is the most frequent cause of esophagitis and the most common outpatient GI diagnosis in the United States. The associated clinical condition, termed gastroesophageal reflux disease (GERD), occurs because the esophageal epithelium is sensitive to acid despite being resistant to abrasive injury.

Pathogenesis

Transient lower esophageal sphincter relaxation is thought to be a major cause of GERD. This relaxation is mediated via vagal pathways and can be triggered by gastric distention. Gastroesophageal reflux can also occur following abrupt increases in intra-abdominal pressure, e.g., after coughing, straining, or bending. Other conditions that are associated with GERD include alcohol and tobacco use, obesity, central nervous system depressants, pregnancy, hiatal hernia (discussed later), delayed gastric emptying, and increased gastric volume.

MORPHOLOGY

Simple hyperemia, evident to the endoscopist as redness, may be the only alteration in mild GERD; mucosal histology is often unremarkable. More significant gastric reflux is associated with erosions (Fig. 17.5A) and influx of eosinophils into the squamous mucosa (Fig. 17.6A). Basal zone hyperplasia and elongation of lamina propria papillae are also common. Neutrophil infiltration is less frequent and is generally associated with bacterial or fungal infection or chemical damage.

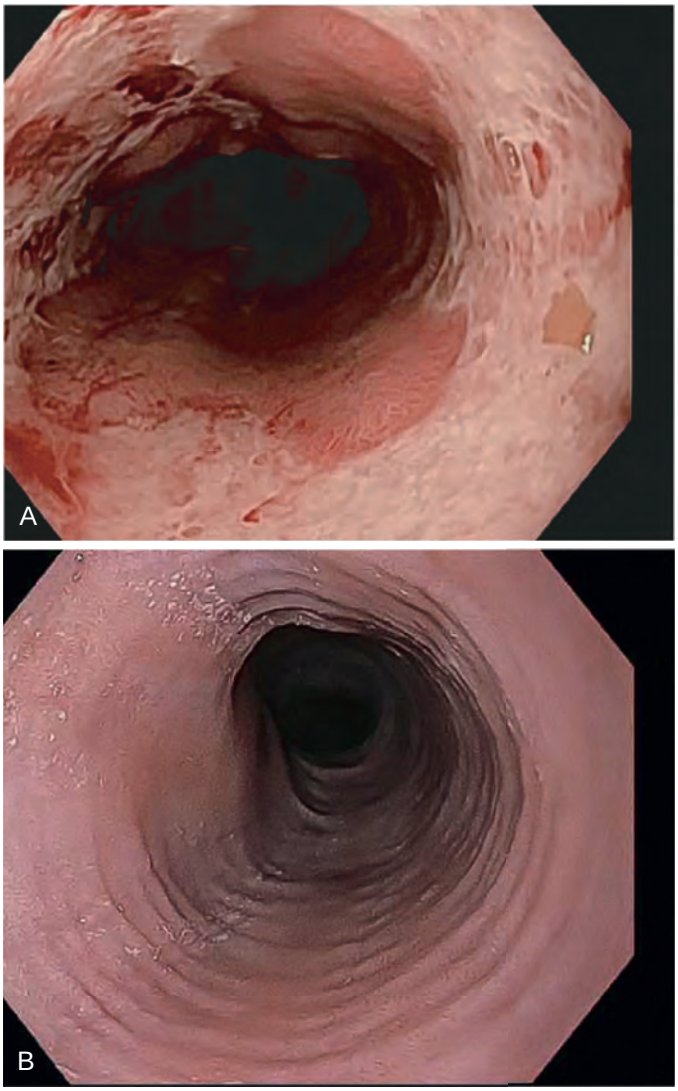


Figure 17.5 Esophagitis. (A) Endoscopic view of reflux esophagitis with multiple erosions within the squamous-lined esophagus, a metaplastic zone (Barrett esophagus, discussed later), and distal gastric mucosa. Note the tan islands of metaplastic epithelium within the white squamous mucosa. (B) The "feline" or "ringed" endoscopic appearance of the esophagus is typical of eosinophilic esophagitis. (Courtesy Dr. Linda S. Lee, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass.)

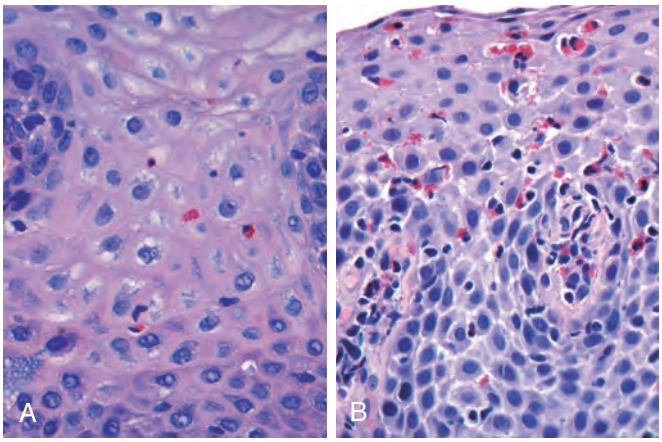


Figure 17.6 Esophagitis. (A) Reflux esophagitis with scattered intraepithelial eosinophils and mild basal zone expansion. (B) Eosinophilic esophagitis is characterized by numerous intraepithelial eosinophils.

Clinical Features

GERD has an estimated prevalence of 10% to 20% in the West, but less than 5% in Asia, and is most common in individuals older than age 40. Typical symptoms include heartburn, dysphagia, and regurgitation of sour-tasting gastric contents, most often postprandially. Rarely, GERD-associated chest pain may be mistaken for ischemic heart disease. Treatment with proton pump inhibitors to reduce gastric acidity typically provides symptomatic relief. Endoscopy is not required for the diagnosis and is usually reserved for those who are refractory to proton pump inhibitors. Endoscopic biopsy can also detect eosinophilic esophagitis as well as Barrett esophagus, strictures, ulcerations, and other GERD complications. Notably, the severity of symptoms is not closely related to the degree of histologic damage, and substantial histologic abnormalities may be found in those without typical GERD symptoms.

Hiatal hernia may cause lower esophageal sphincter incompetence and symptoms that are similar to GERD but it is symptomatic in fewer than 10% of adults. The majority of hiatal hernias are of the sliding type, in which the diaphragmatic crura are spread and the gastroesophageal junction herniates into the thorax. Less common paraesophageal hiatal hernias penetrate through a defect in the phrenoesophageal membrane, can involve the stomach and other organs, and generally require surgical repair.

Eosinophilic Esophagitis

Eosinophilic esophagitis is a form of acute eosinophil-dominated esophageal inflammation associated with atopic disease. Many patients have atopic dermatitis, allergic rhinitis, asthma, or modest peripheral eosinophilia. Consistent with this, recent data suggest that mast cells may also be important in pathogenesis. The incidence of eosinophilic esophagitis has increased dramatically, particularly in urban areas and among white males. In addition to GERD-like symptoms, patients experience food impaction, dysphagia, and vomiting. Infants may also suffer from feeding intolerance. On endoscopic examination the esophagus can take the appearance of stacked circular rings (referred to as feline esophagus because of supposed endoscopic resemblance to a striped cat's tail), strictures, and linear furrows (Fig. 17.5B). The cardinal histologic feature is large numbers of intraepithelial eosinophils, particularly superficially, that can form clusters and sheets (Fig. 17.6B). Their abundance can help to distinguish eosinophilic esophagitis from GERD, Crohn disease, and other causes of esophagitis. However, these histologic findings are not entirely specific, and some patients with extensive intraepithelial eosinophil infiltration may respond to proton pump inhibitors, particularly when given at high doses. Dietary restrictions to prevent exposure to food allergens are often helpful and can be targeted on the basis of skin patch or immunoglobulin E (IgE) testing. Alternatively, empirical elimination of cow's milk, eggs, soy or legumes, and wheat from the diet can be beneficial. When dietary treatment fails, topical or, less frequently, systemic corticosteroids can be used and are sometimes combined with proton pump inhibitors. In contrast to GERD, eosinophilic esophagitis is not associated with increased risk of Barrett esophagus.

Esophageal Varices

Esophageal varices are dilated veins within the lower esophagus. Although most small varices never bleed, rupture of large varices can result in exsanguination.

Pathogenesis

Esophageal varices are caused by portal hypertension, which is due to impaired blood flow through the portal venous system and liver. Increased portal venous pressure results in the development of collateral channels at sites where the portal and caval systems communicate. These collateral veins allow some drainage to occur, but also result in congestion and dilation of subepithelial and submucosal venous plexuses within the distal esophagus and proximal stomach. These dilated vessels, termed *varices,* are common in patients with cirrhosis, most frequently due to alcoholic liver disease. Worldwide, hepatic schistosomiasis is the second most common cause of varices. A more detailed consideration of portal hypertension is given in Chapter 18.

MORPHOLOGY

Esophageal varices are tortuous dilated veins within the mucosa and submucosa of the distal esophagus and proximal stomach (Fig. 17.7). Variceal rupture can result in hemorrhage into the lumen or the esophageal wall and may be associated with mucosal ulceration and necrosis. If rupture has occurred in the past, venous thrombosis, inflammation, and evidence of prior sclerotherapy may also be present.

Clinical Features

Gastroesophageal varices are present in 30% of those with compensated cirrhosis and in 60% of those with decompensated cirrhosis. Variceal hemorrhage is an emergency that can be treated medically by inducing splanchnic vasoconstriction or endoscopically by sclerotherapy (injection of thrombotic agents), balloon tamponade, or variceal ligation. Despite

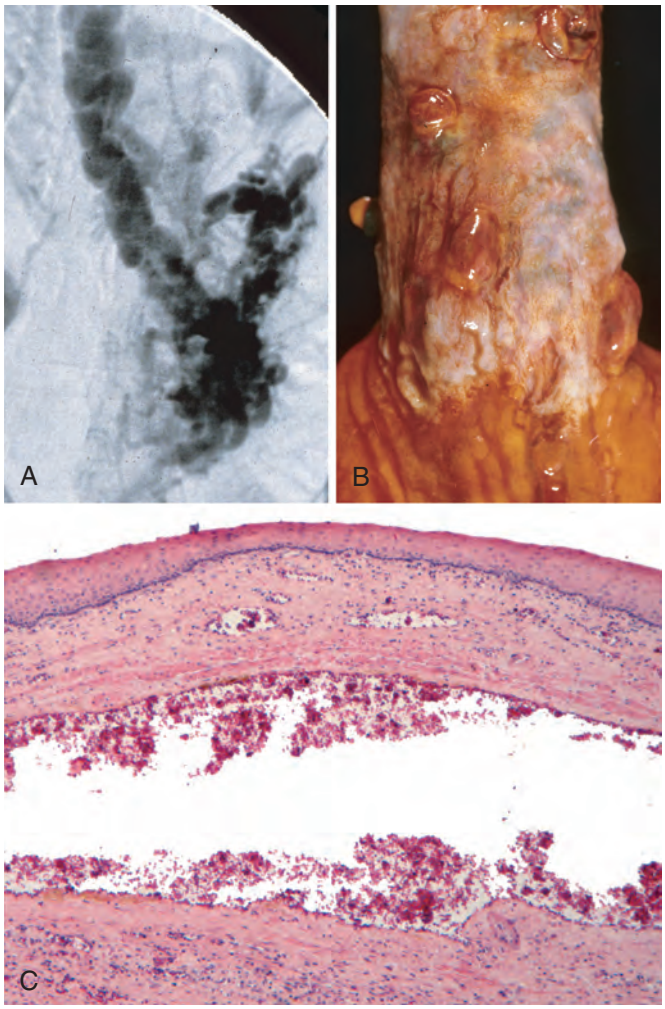


Figure 17.7 Esophageal varices. (A) Although no longer used as a diagnostic approach, this angiogram demonstrates several tortuous esophageal varices. (B) Collapsed varices are present in this postmortem specimen corresponding to the angiogram in (A). The polypoid areas represent previous sites of variceal hemorrhage that have been ligated with bands. (C) Dilated varices beneath intact squamous mucosa.

these interventions, each episode of variceal hemorrhage confers a 15% to 20% risk of mortality, and over half of patients who survive a first variceal bleed have recurrent hemorrhage within 1 year, with mortality risk similar to the first episode. Risk factors for hemorrhage include large or tortuous varices, elevated hepatic venous pressure gradient, previous bleeding, and advanced liver disease. Treatments include beta-blockers to reduce portal blood flow and endoscopic variceal ligation.

Barrett Esophagus

Barrett esophagus is a complication of chronic GERD that is characterized by intestinal metaplasia within the esophageal squamous mucosa and is associated with an increased risk of cancer. The incidence of Barrett esophagus is rising; it is estimated to occur in as many as 10% of individuals with symptomatic GERD and up to 2% of the general population. Barrett esophagus is most common in white males and typically presents between 40 and 60 years of age. The greatest clinical concern is that it confers an increased risk of esophageal adenocarcinoma. Genomic sequencing of biopsies involved by Barrett esophagus has revealed the presence of pathogenic driver mutations in cancer genes that are also found in esophageal adenocarcinoma. Potentially oncogenic mutations are more numerous when biopsies demonstrate dysplasia, a preinvasive neoplastic change that is associated with prolonged symptoms, longer segment length, increased patient age, and Caucasian race. The vast majority of esophageal adenocarcinomas occur in association with Barrett esophagus. Nevertheless, most individuals with Barrett esophagus do not develop esophageal tumors.

MORPHOLOGY

Barrett esophagus can be recognized as tongues of red, velvety mucosa extending upward from the gastroesophageal junction. This metaplastic mucosa alternates with residual smooth, pale squamous (esophageal) mucosa (see Fig. 17.5A) and interfaces with light-brown columnar (gastric) mucosa distally (Fig. 17.8A, B). Barrett esophagus can be subclassified as long segment (≥3 cm), or short segment (<3 cm). Patients with short-segment disease may not experience GERD symptoms and are at lower risk of developing dysplasia or carcinoma relative to those with long-segment disease.

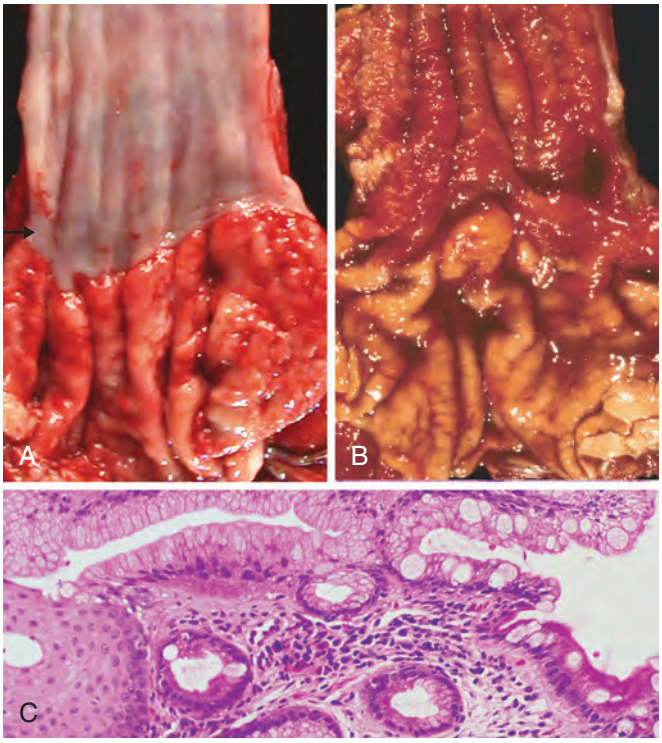


Figure 17.8 Barrett esophagus. (A) Normal gastroesophageal junction. (B) Barrett esophagus. Note the small islands of residual pale squamous mucosa within the Barrett mucosa. (C) Histologic appearance of the gastroesophageal junction in Barrett esophagus. Note the transition between esophageal squamous mucosa (left) and Barrett metaplasia, with abundant metaplastic goblet cells (right).

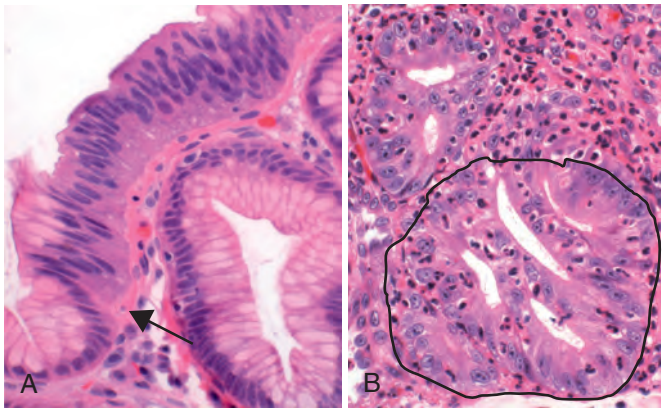


Figure 17.9 Dysplasia in Barrett esophagus. (A) Abrupt transition from metaplasia to low-grade dysplasia *(arrow)*. Note the nuclear stratification and hyperchromasia. (B) Architectural irregularities, including gland-within-gland, or cribriform, profiles in high-grade dysplasia *(encircled)*.

Diagnosis of Barrett esophagus requires endoscopic evidence of metaplastic columnar mucosa above the gastroesophageal junction. Microscopically, intestinal-type metaplasia is seen as replacement of the squamous esophageal epithelium with goblet cells. Goblet cells are diagnostic of Barrett esophagus and have distinct mucous vacuoles that stain pale blue and impart the shape of a wine goblet to the remaining cytoplasm (Fig. 17.8C). Non-goblet columnar cells, such as gastric-type foveolar cells, may also be present, but there is not universal agreement as to whether they are sufficient for the diagnosis.

Dysplasia is classified as low grade or high grade. Atypical mitoses, nuclear hyperchromasia, irregularly clumped chromatin, increased nuclear-to-cytoplasmic ratio, and failure of epithelial cells to mature as they migrate to the esophageal surface are present in both grades (Fig. 17.9A). Dysplastic glands display budding, irregular shapes, and cellular crowding. High-grade dysplasia (Fig. 17.9B) exhibits more severe cytologic and architectural changes than low-grade dysplasia. With progression, epithelial cells may invade the lamina propria, a feature that defines progression to intramucosal carcinoma.

Clinical Features

Barrett esophagus can be identified only thorough endoscopy and biopsy, which are usually prompted by GERD symptoms. Once diagnosed, the best course of management is a matter of debate. Many support surveillance by periodic endoscopy with biopsy, but randomized trials have failed to demonstrate that this improves overall survival. Furthermore, uncertainties regarding the potential of dysplasia, particularly low-grade dysplasia, to regress spontaneously and limited information on the risk of progression complicate clinical decisions.

Intramucosal or invasive carcinoma requires therapeutic intervention. Treatment options include surgical resection, or esophagectomy, as well as newer modalities such as photodynamic therapy, laser ablation, and endoscopic mucosectomy. Multifocal high-grade dysplasia, which carries a significant risk of progression to intramucosal or invasive carcinoma, is treated as intramucosal carcinoma. At present, many physicians follow low-grade dysplasia or a single focus of high-grade dysplasia with endoscopy and biopsy at regular intervals.

ESOPHAGEAL TUMORS

Nearly all esophageal cancers are either adenocarcinoma or squamous cell carcinoma. Squamous cell carcinoma is more common worldwide, but adenocarcinoma is on the rise in the United States and other Western countries. Other malignancies of the esophagus are far less common and include unusual forms of adenocarcinoma, undifferentiated carcinoma, neuroendocrine carcinoma, melanoma, lymphoma, and sarcoma; these are not discussed here. Benign tumors of the esophagus are generally mesenchymal, and arise within the esophageal wall, with leiomyomas being most common.

Adenocarcinoma

Esophageal adenocarcinoma typically arises in the background of Barrett esophagus and long-standing GERD. Thus, increased rates of esophageal adenocarcinoma may be partly due to the increased incidence of obesity-related gastroesophageal reflux and associated Barrett esophagus. Other major risk factors include tobacco use and exposure to radiation. Conversely, risk is reduced by diets rich in fresh fruits and vegetables. Some serotypes of *Helicobacter pylori* are associated with decreased risk of esophageal adenocarcinoma, perhaps because they cause gastric atrophy, which decreases acid secretion and reduces reflux-induced esophageal damage.

Esophageal adenocarcinoma occurs most frequently in Caucasians and shows a strong gender bias, being sevenfold more common in men. The incidence varies widely worldwide, with rates being highest in the United States, the United Kingdom, Canada, Australia, the Netherlands, and Brazil and lowest in Korea, Thailand, Japan, and Ecuador. In countries where esophageal adenocarcinoma is more common, the incidence has increased markedly since 1970, more rapidly than almost any other cancer. For unknown reasons, these increases have been restricted to white and Hispanic men and white women in the United States. As a result, esophageal adenocarcinoma, which represented less than 5% of esophageal cancers before 1970, now accounts for more than half of all esophageal cancers in the United States.

Pathogenesis

Molecular studies indicate that progression of Barrett esophagus to adenocarcinoma occurs over an extended period through the stepwise acquisition of genetic and epigenetic changes. Chromosomal abnormalities and mutations of the tumor suppressor genes *TP53* and *CDKN2A* are detected at early stages. In the case of *CDKN2A*, which you will recall encodes two tumor suppressor proteins, p16 and p19-ARF, both allelic loss and hypermethylation-induced epigenetic silencing have been described. With progression, there may be amplification of several oncogenes, including the *EGFR, ERBB2, MET, cyclin D1,* and *cyclin E* genes.

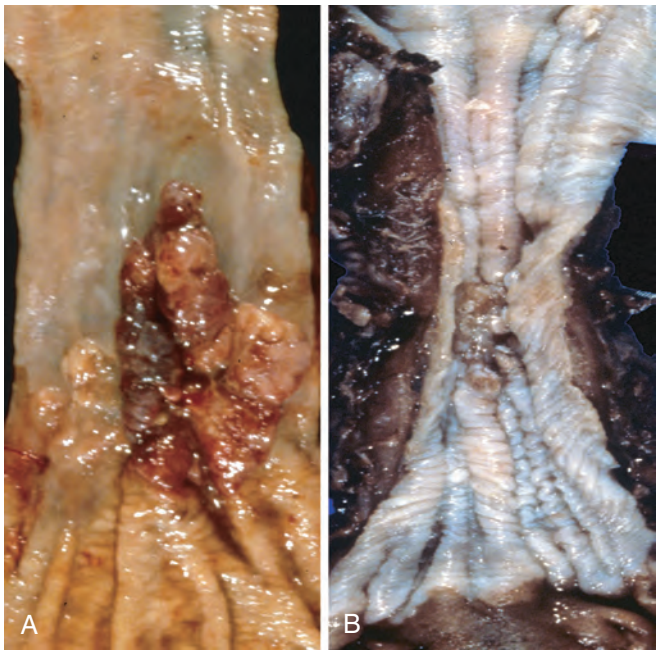


Figure 17.10 Esophageal cancer. (A) Adenocarcinoma usually occurs distally and, as in this case, often involves the gastric cardia. (B) Squamous cell carcinoma is most frequently found in the mid-esophagus, where it commonly causes strictures.

MORPHOLOGY

Esophageal adenocarcinoma usually occurs in the distal third of the esophagus and may invade the adjacent gastric cardia (Fig. 17.10A). It initially appears as flat or raised patches in otherwise intact mucosa, and may grow into large masses of 5 cm or more in diameter. Alternatively, tumors may infiltrate diffusely or ulcerate and invade deeply.

Microscopically, tumors typically produce mucin and form glands (Fig. 17.11A), often with intestinal-type morphology; less frequently, tumors are composed of diffusely infiltrative signet-ring cells (similar to those seen in diffuse gastric cancers) or, in rare cases, small poorly differentiated cells (similar to small cell carcinoma of the lung). Barrett esophagus is frequently present adjacent to the tumor.

Clinical Features

Although esophageal adenocarcinomas are occasionally discovered incidentally during the evaluation of GERD or surveillance of Barrett esophagus, they more commonly present with pain or difficulty in swallowing, progressive weight loss, hematemesis, chest pain, or vomiting. By the time symptoms appear, the tumor has often spread to submucosal lymphatic vessels. As a result, overall 5-year survival is less than 25%. In contrast, 5-year survival is approximately 80% in cases in which adenocarcinoma is limited to the mucosa or submucosa.

Squamous Cell Carcinoma

In the United States, esophageal squamous cell carcinoma occurs in adults older than age 45 and affects males four times more frequently than females. Risk factors include alcohol and tobacco use, poverty, caustic esophageal injury, achalasia, Plummer-Vinson syndrome, diets that are deficient in fruits or vegetables, and frequent consumption of very hot beverages. Previous radiation to the mediastinum also predisposes individuals to esophageal carcinoma, with most cases occurring 5 to 10 or more years after exposure.

Esophageal squamous cell carcinoma incidence varies up to 180-fold between and within countries, being more common in rural and low income populations. The regions with highest incidence are Iran, central China, Hong Kong, Brazil, and South Africa. A pocket of extremely high incidence in western Kenya has been linked to consumption of a traditional fermented milk, termed *mursik,* which contains the carcinogen acetaldehyde (Chapter 9). In other geographic areas with high esophageal squamous cell carcinoma rates, regular ingestion of very hot tea is common. In the United States, it is nearly eightfold more frequent in African Americans than in Caucasians, a striking risk disparity that partially reflects differences in rates of alcohol and tobacco use but likely also reflects contributions of other factors that remain poorly understood.

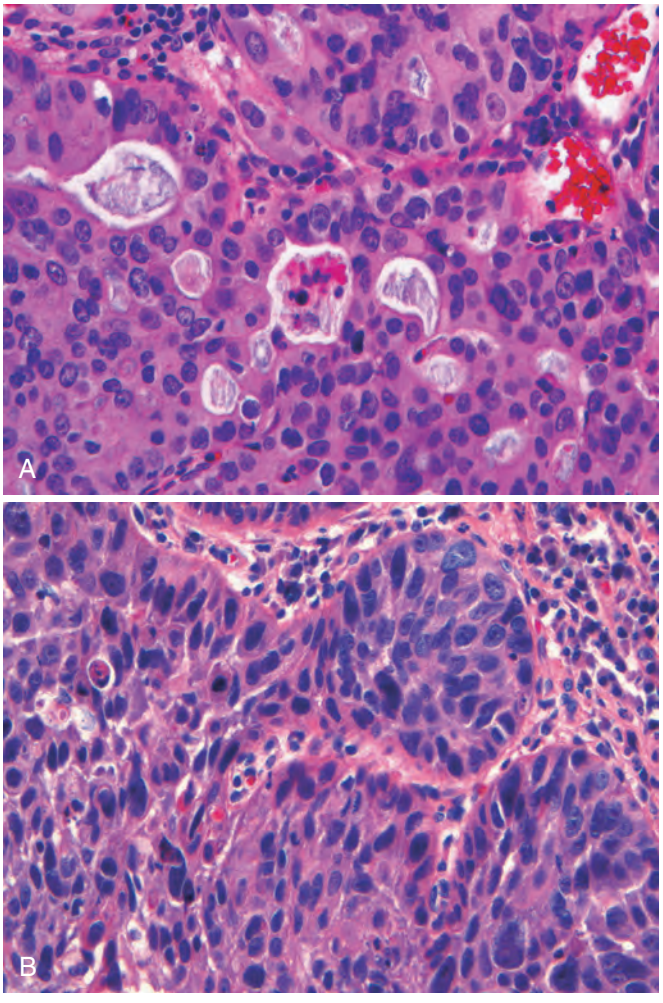


Figure 17.11 Esophageal cancer. (A) Esophageal adenocarcinoma organized into back-to-back glands. (B) Squamous cell carcinoma composed of nests of malignant cells that partially recapitulate the organization of squamous epithelium.

Pathogenesis

The majority of esophageal squamous cell carcinomas in Europe and the United States are linked to alcohol and tobacco use, which synergize to increase the risk. However, esophageal squamous cell carcinoma is also common in some regions where alcohol and tobacco use is uncommon. Thus, nutritional deficiencies, as well as polycyclic hydrocarbons, nitrosamines, and other mutagenic compounds, such as those found in fungus-contaminated foods must be considered. Some data suggest that human papillomavirus (HPV) infection may contribute to esophageal squamous cell carcinoma in high-risk, but not in low-risk, regions. The molecular pathogenesis of esophageal squamous cell carcinoma remains incompletely defined, but recurrent abnormalities include amplification of the transcription factor gene *SOX2,* which is involved in stem cell self-renewal; overexpression of the cell cycle regulator cyclin D1; and loss-of-function mutations in the tumor suppressor genes *TP53, CDH1* (which encodes E-cadherin), and *NOTCH1*.

Clinical Features

The onset of esophageal squamous cell carcinoma is insidious. Patients typically have dysphagia, odynophagia (pain on swallowing), or obstruction at presentation and may have unknowingly adjusted to progressive esophageal obstruction by altering their diet from solid to liquid foods. Weight loss and debilitation result from both impaired nutrition and tumor cachexia. Hemorrhage and sepsis may accompany tumor ulceration, and symptoms of iron deficiency are often present. Occasionally, the first symptoms are caused by aspiration of food via a tracheoesophageal fistula (caused by extension of the tumor into the trachea).

Increased prevalence of endoscopic screening has led to earlier detection of esophageal squamous cell carcinoma. This is significant because 5-year survival rates are 75% in individuals with superficial lesions but much lower in patients with more advanced tumors. Lymph node metastases are associated with poor prognosis. The overall 5-year survival rate in the United States remains less than 20% and varies by tumor stage and patient age, race, and sex.

MORPHOLOGY

In contrast to the distal location of adenocarcinoma, half of squamous cell carcinomas occur in the middle third of the esophagus (Fig. 17.10B). Squamous cell carcinoma begins as an in situ lesion termed **squamous dysplasia** (this histopathology is referred to as intraepithelial neoplasia or carcinoma in situ at sites outside the GI tract, e.g., uterine cervix). Early lesions appear as small, gray-white, plaque-like thickenings. Over months to years they grow into tumor masses that may be polypoid, or exophytic, and protrude into and obstruct the lumen. Other tumors are ulcerated or diffusely infiltrative lesions that spread within the esophageal wall and cause thickening, rigidity, and luminal narrowing. Tumors may invade surrounding structures including the respiratory tree (causing pneumonia), the aorta (causing catastrophic exsanguination), or the mediastinum and pericardium.

Most squamous cell carcinomas are moderately to well differentiated (Fig. 17.11B). Less common histologic variants include verrucous squamous cell carcinoma, spindle cell carcinoma, and basaloid squamous cell carcinoma. Regardless of histology, symptomatic tumors are generally large and invasive at diagnosis. The rich lymphatic network of the esophagus promotes metastases as well as circumferential and longitudinal spread; intramural satellite tumor nodules may be present several centimeters away from the principal mass. The sites of lymph node metastases vary with tumor location: cancers in the upper third of the esophagus favor cervical lymph nodes; those in the middle third favor mediastinal, paratracheal, and tracheobronchial nodes; and those in the lower third spread to gastric and celiac nodes.

KEY CONCEPTS

ESOPHAGEAL DISEASES

- Abnormalities of esophageal motility include nutcracker esophagus and diffuse esophageal spasm.
- Achalasia, characterized by incomplete lower esophageal sphincter relaxation, increased lower esophageal sphincter tone, and esophageal aperistalsis, is a common form of functional esophageal obstruction. It can be primary or secondary, with the latter form most commonly due to *Trypanosoma cruzi* infection.
- Mallory-Weiss tears of mucosa at the gastroesophageal junction develop as a result of severe retching or vomiting.
- Esophagitis can result from chemical or infectious mucosal injury. Infection is most common in immunocompromised individuals.
- The most prevalent cause of esophagitis is reflux of gastric acid into the esophagus (GERD).
- Eosinophilic esophagitis is strongly associated with food allergy, allergic rhinitis, or asthma. It is a common cause of GERD-like symptoms in children living in high income countries.
- Gastroesophageal varices are a consequence of portal hypertension and are present in half of cirrhosis patients.
- Barrett esophagus develops in patients with chronic GERD and represents columnar metaplasia of the esophageal squamous mucosa.
- Barrett esophagus is a risk factor for development of esophageal adenocarcinoma.
- Esophageal squamous cell carcinoma is associated with alcohol and tobacco use, poverty, caustic esophageal injury, achalasia, tylosis, and Plummer-Vinson syndrome.

Stomach

Disorders of the stomach are a frequent cause of clinical disease, with inflammatory and neoplastic lesions being most common. In the United States, diseases related to the stomach account for nearly one-third of all health care spending on GI disease. In addition, despite decreasing incidence in certain locales such as the United States, gastric cancer remains a leading cause of death worldwide.

The stomach is divided into four major anatomic regions: cardia, fundus, body, and antrum. The cardia and antrum are lined mainly with mucin-secreting foveolar cells that form small glands. The antral glands are similar but also contain endocrine cells such as G cells that release gastrin to stimulate luminal acid secretion by parietal cells within the gastric fundus and body. The well-developed glands

of the body and fundus contain parietal cells as well as chief cells that produce and secrete digestive enzymes.

GASTROPATHY AND ACUTE GASTRITIS

Inflammation of the gastric mucosa occurs in many conditions, and is called acute gastritis when neutrophils are present and gastropathy when inflammatory cells are rare or absent. Irritants including non-steroidal anti-inflammatory drugs (NSAIDs), alcohol, and bile are the most common causes of gastropathy. Both gastropathy and gastritis may be asymptomatic or associated with epigastric pain, nausea, and vomiting. In more severe cases there may be mucosal erosion, ulceration, hemorrhage, hematemesis, melena, or, rarely, massive blood loss.

Pathogenesis

The gastric lumen has a pH of close to 1, more than a million times more acidic than the blood. This harsh environment contributes to digestion but also has the potential to cause damage. Multiple mechanisms have evolved to protect the gastric mucosa (Fig. 17.12). Foveolar cell secretions form a thin layer of mucus and phospholipids that prevents large food particles from directly touching the epithelium. Mucus also promotes formation of an "unstirred" layer of fluid with a neutral pH as a result of bicarbonate ion secretion by surface epithelial cells. Beneath the mucus, gastric epithelial cells form a physical barrier that limits back diffusion of acid and leakage of other luminal materials, including digestive enzymes, into the lamina propria. Surface foveolar cells are replaced every 3 to 7 days, while parietal and chief cells are more long-lived. The rich mucosal vasculature delivers oxygen and nutrients and removes whatever gastric acid may have diffused into the lamina propria.

Gastropathy and gastritis occur when the damaging forces overwhelm the protective factors (Fig. 17.12). Disruption of protective mechanisms is illustrated by the following examples:

- *Inhibition of cyclooxygense (COX) by NSAIDs.* NSAIDs inhibit COX-dependent synthesis of prostaglandins E_2 and I_2, which contribute to nearly all of the above defense mechanisms, including: mucus, bicarbonate, and phospholipid secretion; mucosal blood flow; and epithelial restitution. They also reduce acid secretion. Although COX-1 plays a larger role than COX-2, both isoenzymes contribute to mucosal protection. Thus, while the risk of NSAID-induced gastric injury is greatest with nonselective inhibitors such as aspirin, ibuprofen, and naproxen, selective COX-2 inhibitors such as celecoxib can also cause gastropathy or gastritis.
- *Inhibition of gastric bicarbonate transporters* by ammonium ions can result in gastric injury in uremic patients and those infected with urease-secreting *H. pylori*.
- *Reduced mucin and bicarbonate secretion* have been suggested as factors that explain the increased susceptibility of older adults to gastritis.
- *Decreased oxygen delivery* may account for an increased incidence of acute gastritis at high altitudes.

Ingestion of harsh chemicals, particularly acids or bases, either accidentally or as a suicide attempt, causes severe gastric injury. Direct cellular damage also contributes to

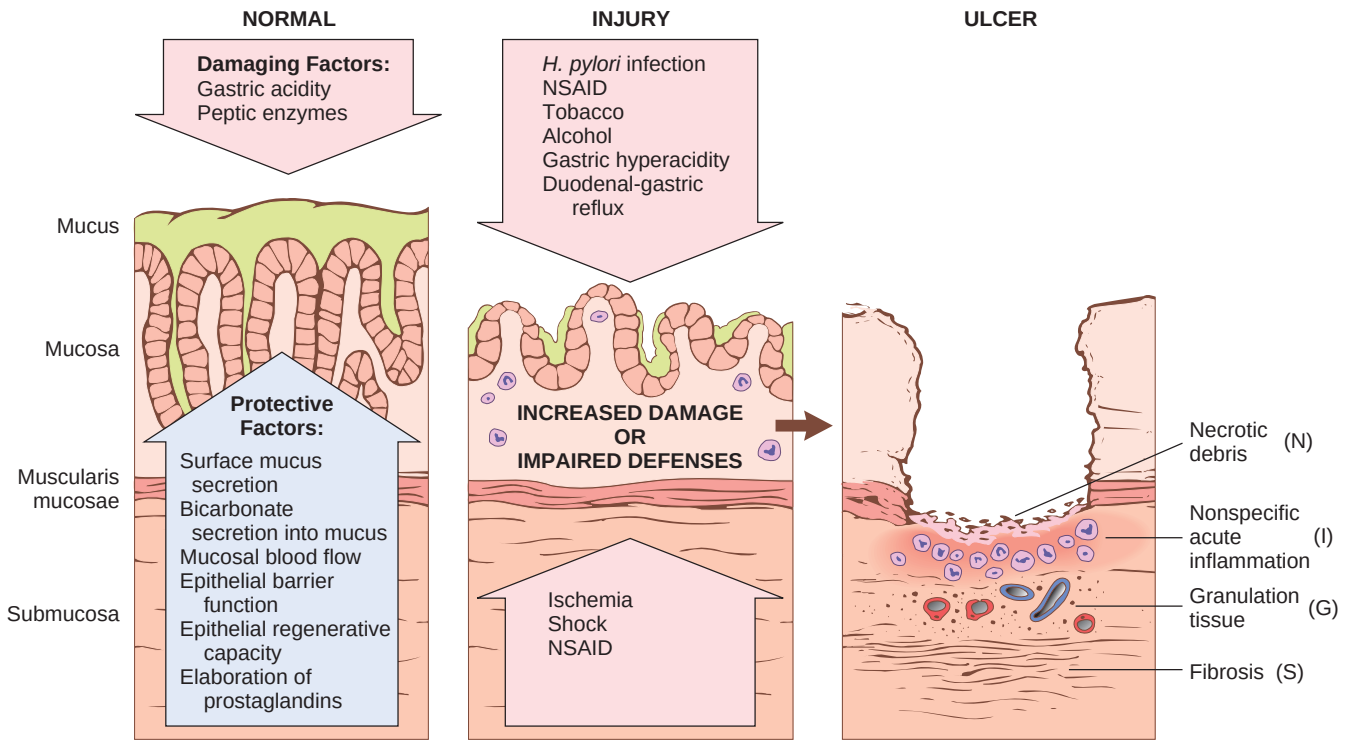


Figure 17.12 Mechanisms of gastric injury and protection. This diagram illustrates the progression from more mild forms of injury to ulceration that may occur with acute or chronic gastritis. Ulcers include layers of necrosis *(N)*, inflammation *(I)*, and granulation tissue *(G)*, but a fibrotic scar *(S)*, which takes time to develop, is present only in chronic lesions. *NSAID*, Non-steroidal anti-inflammatory drug.

gastritis induced by alcohol consumption, NSAIDs, radiation therapy, and chemotherapy. Agents that inhibit DNA synthesis or the mitotic apparatus, including those used in cancer chemotherapy, may cause generalized mucosal damage by inhibiting epithelial renewal.

MORPHOLOGY

Histologically, gastropathy and mild acute gastritis may be difficult to recognize, since the lamina propria shows only moderate edema and slight vascular congestion. The surface epithelium is intact, but foveolar cell hyperplasia, with characteristic corkscrew profiles and epithelial proliferation, is often present. Neutrophils may be found among the epithelial cells or within mucosal gland lumina in gastritis. The lamina propria contains only a few lymphocytes and plasma cells.

The presence of neutrophils above the basement membrane in direct contact with epithelial cells is abnormal in all parts of the GI tract and signifies active inflammation, or, in this case, gastritis (rather than gastropathy). The term *active inflammation* is preferred over acute inflammation, since neutrophils may be present in both acute and chronic disease states. Thus, within the GI tract, the term acute refers to disease duration rather than the inflammatory pattern.

As injury progresses, erosions, i.e., superficial mucosal defects, are accompanied by neutrophilic infiltrates and fibrin-containing exudate within the lumen. Hemorrhage may cause dark puncta within hyperemic mucosa. Concurrent erosion and hemorrhage is termed **acute erosive hemorrhagic gastritis.** Although large areas of the gastric surface may be denuded, the involvement is typically superficial.

STRESS-RELATED MUCOSAL DISEASE

Stress-related mucosal disease occurs in patients with severe trauma, extensive burns, intracranial disease, major surgery, serious medical disease, and other forms of severe physiologic stress. In some cases, stress-related ulcers are given specific names based on location and clinical associations. For example:

- *Stress ulcers* are most common in the setting of shock, sepsis, or severe trauma.
- *Curling ulcers* refer to ulcers occurring in the proximal duodenum in the context of severe burns or trauma.
- *Cushing ulcers* refer to gastric, duodenal, and esophageal ulcers arising in those with intracranial disease. They have an elevated risk of perforation.

Pathogenesis

The pathogenesis of stress-related gastric mucosal injury is most often related to local ischemia. This may be due to systemic hypotension or reduced blood flow caused by stress-induced splanchnic vasoconstriction. Upregulation and increased release of the vasoconstrictor endothelin-1 also contribute to ischemic gastric mucosal injury.

Lesions associated with intracranial injury are thought to be caused by direct stimulation of vagal nuclei, which causes hypersecretion of gastric acid. Systemic acidosis, a frequent finding in these settings, may also contribute to mucosal injury by lowering the intracellular pH of mucosal cells and reducing the pH gradient that promotes washout of acid that has back-diffused into the lamina propria.

MORPHOLOGY

Stress-related gastric mucosal injury ranges from shallow erosions to lesions that penetrate the mucosa. Ulcers are rounded, less than 1 cm in diameter, and have a base that is frequently stained brown to black by acid digestion of extravasated blood. Unlike peptic ulcers, which are usually solitary and located in the antrum, acute stress ulcers are found anywhere in the stomach and are often multiple. Microscopically, acute stress ulcers are sharply demarcated, with relatively normal adjacent mucosa. Serositis may be present, but the scarring and blood vessel thickening that characterize chronic peptic ulcers are conspicuously absent. Healing with complete re-epithelialization occurs within days to weeks after successful treatment of the underlying condition.

Clinical Features

Most critically ill patients admitted to hospital intensive care units have histologic evidence of gastric mucosal damage. Bleeding from these lesions is sufficiently severe to require red cell transfusions in 1% to 4% of such patients. Other complications, including perforation, can also occur. Prophylactic proton pump inhibitors may blunt the impact of stress ulceration, but the most important determinant of clinical outcome is whether the underlying condition can be corrected.

Non–stress-related causes of gastric bleeding include the following two conditions:

- *Dieulafoy lesion* consists of an abnormal submucosal arteriole that is usually found within the lesser curvature of the stomach near the gastroesophageal junction. Erosion of the overlying epithelium can cause recurrent gastric bleeding that, while usually self-limited, can be copious. NSAID use may enhance bleeding risk.
- *Gastric antral vascular ectasia (GAVE)* is responsible for 4% of non-variceal upper GI bleeding. It can be recognized endoscopically as longitudinal stripes of edematous, erythematous mucosa that alternate with less severely injured, paler mucosa and is sometimes referred to as watermelon stomach. The erythematous stripes are created by ectatic mucosal vessels. Histologically, the antral mucosa shows reactive gastropathy with dilated capillaries containing fibrin thrombi. While most often idiopathic, GAVE can be associated with cirrhosis and systemic sclerosis. Recurrent bleeding may produce a positive test for fecal blood and lead to iron deficiency anemia.

CHRONIC GASTRITIS AND ITS COMPLICATIONS

Chronic gastritis is most often caused by *H. pylori* infection. Other causes include autoimmune gastritis; radiation injury;

chronic bile reflux; mechanical injury (e.g., an indwelling nasogastric tube); and involvement by systemic diseases such as Crohn disease, amyloidosis, or graft-versus-host disease.

Relative to acute gastritis, the symptoms associated with chronic gastritis are typically less severe but more persistent. Nausea and upper abdominal pain are typical, sometimes with vomiting, but hematemesis is uncommon.

Helicobacter pylori Gastritis

H. pylori are spiral-shaped or curved bacilli present in gastric biopsy specimens of almost all patients with duodenal ulcers as well as most individuals with gastric ulcers or chronic gastritis. Acute *H. pylori* infection does not produce sufficient symptoms to come to medical attention in most cases; it is the chronic gastritis that ultimately causes the individual to seek treatment. *H. pylori* organisms are present in the majority of individuals with chronic antral gastritis.

Epidemiology. In the United States, *H. pylori* infection is associated with poverty, household crowding, limited education, residence in rural areas, birth outside of the United States, and age over 60 years. The prevalence of infection is also greater in those of African American or Mexican American ethnicity as well as some immigrant populations. Humans are the primary carriers, suggesting that transmission is primarily by the fecal-oral route. Infection is typically acquired in childhood and persists for life without treatment. Improved sanitation explains the marked reduction in *H. pylori* infection rates among younger people. In well-resourced countries, *H. pylori* infection is currently uncommon before 10 years of age, is found in only 10% of those between the ages of 18 and 30 years, but is present in up to 50% of those older than 65. It follows that environment during childhood is a critical risk factor for *H. pylori* colonization.

Pathogenesis

***H. pylori* infection most often presents as a predominantly antral gastritis with normal or increased acid production.** When inflammation remains limited to the antrum, modestly increased local gastrin production can augment parietal cell mass within the gastric body and increase acid secretion that leads to greater risk of gastric or duodenal peptic ulcer disease (see later). Alternatively, long-standing *H. pylori* gastritis may progress to involve the gastric body and fundus. This may result in atrophic gastritis with reduced parietal cell mass and intestinal metaplasia. In contrast to autoimmune gastritis (see later), atrophy induced by *H. pylori* is not associated with autoantibodies and is typically patchy. The loss of parietal cells leads to reduced acid secretion that, in turn, stimulates gastrin production. However, because some parietal cells survive, limited acid secretion continues and gastrin increases are not as great as those in autoimmune gastritis. Nevertheless, decreased acid secretion in *H. pylori* gastritis with atrophy reduces the risk of gastric and duodenal ulcers. This results in an inverse relationship between gastric adenocarcinoma, which is associated with atrophy and intestinal metaplasia, and duodenal ulcers, which are associated with increased acid secretion.

H. pylori organisms have adapted to the ecologic niche provided by gastric mucus. Its virulence is linked to the following factors:

- *Flagella,* which allow the bacteria to be motile in viscous mucus
- *Urease,* which generates ammonia from endogenous urea and thereby elevates local gastric pH and enhances bacterial survival
- *Adhesins* that enhance bacterial adherence to surface foveolar cells
- *Toxins,* such as cytotoxin-associated gene A *(CagA)*

Variation in these and other bacterial factors are strongly linked to outcome. For example, the *CagA* gene is present in 50% of *H. pylori* isolates overall but in 90% of *H. pylori* isolates in populations with an increased prevalence of gastric cancer. This may, in part, be because *CagA*-expressing strains colonize the gastric body and induce proinflammatory cytokine secretion, atrophy, and intestinal metaplasia more effectively than *CagA*-negative *H. pylori*.

Host factors also play an important role in the outcome of *H. pylori* infection. Genetic polymorphisms that lead to increased expression of the proinflammatory cytokines tumor necrosis factor (TNF) and interleukin (IL)-1β or decreased expression of the anti-inflammatory cytokine IL-10 seem to be associated with development of pangastritis, atrophy, intestinal metaplasia, and gastric cancer. The course of *H. pylori* gastritis is, therefore, the result of interplay between gastroduodenal mucosal defenses, host immune responses, and bacterial virulence factors.

MORPHOLOGY

Gastric biopsy specimens generally demonstrate *H. pylori* in infected individuals. The organism is concentrated within the superficial mucus overlying epithelial cells in the surface and neck regions. The distribution can be irregular, with areas of heavy colonization adjacent to those with few organisms. Organisms are most easily demonstrated with immunostains or histochemical stains (Fig. 17.13A).

The antrum is the preferred biopsy site for evaluation of *H. pylori* gastritis because it is most commonly infected. In dense colonization, organisms may also be found in oxyntic (acid-producing) mucosa of the fundus and body. *H. pylori*–infected antral mucosa is usually erythematous and has a coarse or even nodular appearance. The inflammatory infiltrate includes large numbers of plasma cells, often in clusters or sheets, within the superficial lamina propria. These are accompanied by increased numbers of lymphocytes, macrophages, and neutrophils within the lamina propria. Neutrophils infiltrate across the basement membrane (Fig. 17.13B) and accumulate in the lumens of gastric glands, or pits, to create pit abscesses. When intense, inflammatory infiltrates may create thickened rugal folds, mimicking the endoscopic appearance of early cancers. Lymphoid aggregates, some with germinal centers, are frequently present (Fig. 17.13C) and represent induced **mucosa-associated lymphoid tissue** (MALT) that has the potential to transform into lymphoma. Thus, chronic *H. pylori* gastritis is associated with increased risk of both gastric adenocarcinoma and lymphoma.

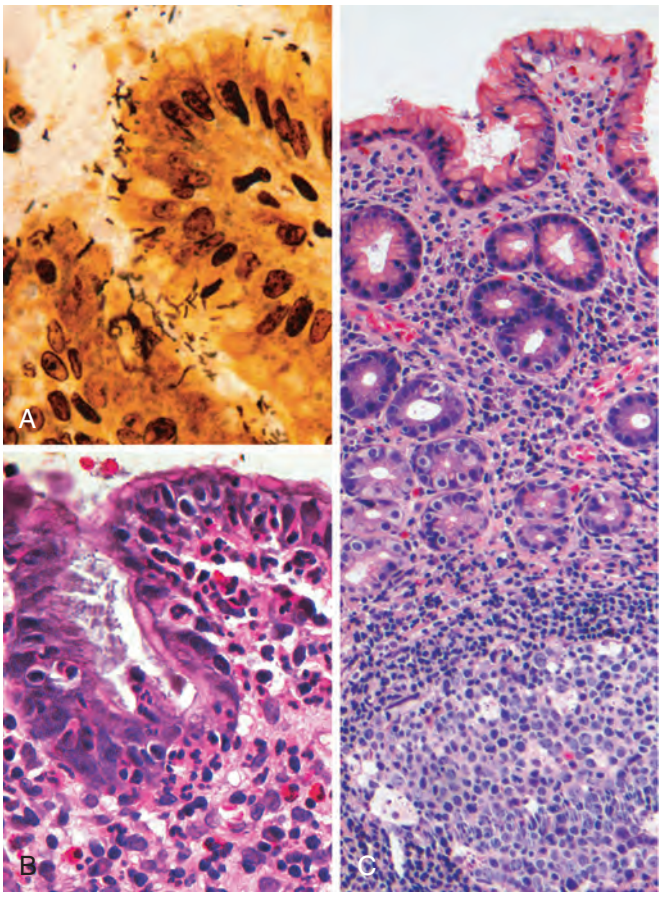


Figure 17.13 *Helicobacter pylori* gastritis. (A) Spiral-shaped *H. pylori* are highlighted in this Warthin-Starry silver stain. Organisms are abundant within surface mucus. (B) Intraepithelial and lamina propria neutrophils are prominent. (C) Lymphoid aggregates with germinal centers and abundant subepithelial plasma cells within the superficial lamina propria are characteristic of *H. pylori* gastritis.

Table 17.2 Characteristics of *Helicobacter pylori*–Associated and Autoimmune Gastritis

	H. pylori–Associated	Autoimmune
Location	Antrum	Body
Inflammatory infiltrate	Neutrophils, subepithelial plasma cells	Lymphocytes, macrophages
Acid production	Increased to slightly decreased	Decreased
Gastrin secretion	Normal to increased	Increased to markedly increased
Other lesions	Hyperplastic/ inflammatory polyps	Neuroendocrine hyperplasia
Serology	Antibodies to *H. pylori*	Antibodies to parietal cells (H^+,K^+-ATPase, intrinsic factor)
Sequelae	Peptic ulcer, adenocarcinoma, MALToma	Atrophy, pernicious anemia, adenocarcinoma, carcinoid tumor
Associations	Low socioeconomic status, poverty, residence in rural areas	Autoimmune disease: thyroiditis, diabetes mellitus, Graves disease

H^+,K^+ ATPase, Hydrogen potassium adenosine triphosphatase; *MALToma*, mucosa-associated lymphoid tissue lymphoma.

Clinical Features

The changes seen in atrophic *H. pylori* gastritis are sometimes referred to as environmental metaplastic atrophic gastritis, to distinguish it from autoimmune metaplastic atrophic gastritis (described next). In addition to histologic identification of the organism, several noninvasive diagnostic tests have been developed, including serology for antibodies to *H. pylori,* fecal bacterial detection, and the urea breath test, which is positive due to ammonia produced by the bacterial urease. Other diagnostic tests performed on biopsies include the rapid urease test, bacterial culture, or polymerase chain reaction (PCR)-based detection of *H. pylori* DNA.

Effective treatments for *H. pylori* infection include combinations of antibiotics and proton pump inhibitors. Individuals with *H. pylori* gastritis usually improve after treatment, although relapses can occur after incomplete eradication or re-infection, particularly in regions with high endemic colonization rates.

Autoimmune Atrophic Gastritis

Autoimmune atrophic gastritis, in contrast to *H. pylori*–associated gastritis, typically spares the antrum and is often associated with marked hypergastrinemia (Table 17.2). It accounts for less than 10% of cases of chronic gastritis and has an estimated prevalence of 2% in those over 60 years of age. Autoimmune atrophic gastritis is characterized by:

- Antibodies to parietal cells and intrinsic factor that can be detected in serum and gastric secretions
- Reduced serum pepsinogen I concentration
- Endocrine cell hyperplasia
- Vitamin B_{12} deficiency
- Defective gastric acid secretion (achlorhydria)

Pathogenesis

Autoimmune atrophic gastritis is associated with loss of parietal cells, which are responsible for secretion of gastric acid and intrinsic factor. The absence of acid production stimulates gastrin release, resulting in hypergastrinemia and hyperplasia of antral gastrin-producing G cells. Intrinsic factor loss results in defective ileal vitamin B_{12} absorption, which ultimately leads to vitamin B_{12} deficiency and *pernicious anemia* (a form of megaloblastic anemia described in Chapter 14). The high frequency of associated intestinal metaplasia has led to the term *autoimmune metaplastic atrophic gastritis*.

CD4+ T cells directed against parietal cell components, including hydrogen potassium adenosine triphosphatase (H^+,K^+-ATPase), are considered to be the principal agents of injury in autoimmune atrophic gastritis. This is supported by the observation that transfer of H^+,K^+-ATPase–reactive CD4+ T cells into naïve mice results in gastritis and production of H^+,K^+-ATPase autoantibodies. There is no evidence of an autoimmune reaction to chief cells, suggesting that these may be lost through gastric gland destruction

during autoimmune attack on parietal cells. If autoimmune destruction is controlled by immunosuppression, the glands can repopulate, demonstrating that gastric stem cells survive and are able to differentiate into parietal and chief cells.

Autoantibodies to parietal cell components, most prominently H^+,K^+-ATPase, or proton pump, and intrinsic factor are present in up to 80% of patients with autoimmune atrophic gastritis. However, these antibodies are not thought to be pathogenic because neither secreted intrinsic factor nor the luminally oriented proton pump is accessible to circulating antibodies, and passive transfer of these antibodies does not produce gastritis in experimental animals.

MORPHOLOGY

Autoimmune atrophic gastritis is characterized by diffuse damage to the oxyntic (acid-producing) mucosa within the body and fundus. The antrum and cardia are typically spared. With diffuse atrophy, oxyntic mucosa of the body and fundus appears markedly thinned, and rugal folds are lost. If vitamin B_{12} deficiency is severe, nuclear enlargement (megaloblastic change) occurs within epithelial cells. Neutrophils may be present, but the inflammatory infiltrate is typically composed of lymphocytes, macrophages, and plasma cells, often in association with lymphoid aggregates and follicles. In contrast to the superficial lamina propria inflammation that is typical of *H. pylori* gastritis, inflammation in autoimmune atrophic gastritis is deeper and centered on the gastric glands (Fig. 17.14A). Loss of parietal and chief cells can be extensive. When atrophy is incomplete, residual islands of oxyntic mucosa may give the appearance of multiple small polyps or nodules. In other areas, small surface elevations may represent sites of intestinal metaplasia, characterized by the presence of goblet cells and columnar absorptive cells (Fig. 17.14B). Endocrine and enterochromaffin-like cell hyperplasia is commonly present. Rarely, endocrine cell hyperplasia may progress to form small, multicentric, low-grade neuroendocrine (carcinoid) tumors.

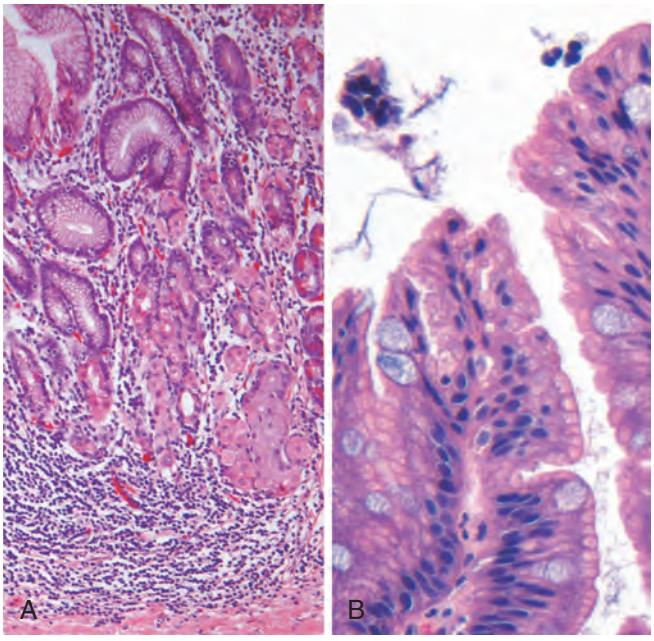


Figure 17.14 Autoimmune gastritis. (A) Low-magnification image of gastric body demonstrating deep inflammatory infiltrates, primarily composed of lymphocytes, and glandular atrophy. (B) Intestinal metaplasia, recognizable as the presence of goblet cells within gastric foveolar epithelium.

Clinical Features

Antibodies to parietal cells and intrinsic factor are present 2 to 3 decades before the appearance of gastric atrophy and other complications such as pernicious anemia; hence, progression to clinically evident disease is slow. The median age at diagnosis is 60. Slightly more women than men are affected, and there does not appear to be an ethnic or racial bias. Pernicious anemia and autoimmune atrophic gastritis are often associated with other autoimmune diseases including Hashimoto thyroiditis, type 1 diabetes mellitus, Addison disease, and others. These associations, along with concordance in some monozygotic twins and clustering of disease in families, support a genetic predisposition. Nevertheless, there is little evidence of linkage between autoimmune gastritis and specific human leukocyte antigen (HLA) alleles.

The clinical features are mainly related to vitamin B_{12} deficiency, including megaloblastic anemia (Chapter 14); atrophic glossitis, in which the tongue becomes smooth and beefy red; epithelial megaloblastosis; malabsorptive diarrhea; and many others (see Chapter 28). Clinical features and pathologic changes related to the effects of vitamin B_{12} deficiency in the marrow and the CNS (e.g., subacute combined degeneration of the spinal cord) are described in detail in Chapters 14 and 28, respectively.

Uncommon Forms of Gastritis

Eosinophilic Gastritis. Eosinophilic gastritis is characterized by tissue damage associated with dense infiltrates of eosinophils in the mucosa and muscularis, usually in the antral or pyloric region. The lesion may also be present at other sites within the GI tract and is associated with peripheral eosinophilia and increased serum IgE levels. Allergic reactions are one cause of eosinophilic gastritis, with cow's milk and soy protein being the most common allergens in children. Eosinophilic gastritis can also occur in association with immune disorders such as systemic sclerosis and inflammatory myopathy, parasitic infections, and even *H. pylori* infection.

Lymphocytic Gastritis. This disease preferentially affects women and produces nonspecific abdominal symptoms. It is idiopathic, but approximately 40% of cases are associated with celiac disease, suggesting an immune-mediated pathogenesis. Lymphocytic gastritis typically affects the entire stomach and is often referred to as *varioliform gastritis* based on the distinctive endoscopic appearance (characterized by thickened folds covered by small nodules with central aphthous ulceration). Histologically there is a marked increase in the number of intraepithelial T lymphocytes.

Granulomatous Gastritis. The descriptive term granulomatous gastritis is applied to any gastritis that contains

granulomas. It encompasses a diverse group of diseases with widely varying clinical and pathologic features. Many cases are idiopathic. In Western populations, gastric involvement by Crohn disease is the most common specific cause of granulomatous gastritis followed by sarcoidosis and infections (including mycobacteria, fungi, CMV, and *H. pylori*). In addition to the presence of histologically evident granulomas, narrowing and rigidity of the gastric antrum may occur secondary to transmural granulomatous inflammation.

Peptic Ulcer Disease

Peptic ulcer disease (PUD) refers to chronic mucosal ulceration affecting the duodenum or stomach and is almost always associated with *H. pylori* infection, NSAIDs, or cigarette smoking. The most common form of PUD occurs within the gastric antrum or duodenum as a result of chronic *H. pylori*–induced antral gastritis, which is associated with increased gastric acid secretion and decreased duodenal bicarbonate secretion. As mentioned earlier, *H. pylori* infection involving the gastric fundus or body is usually accompanied by more modestly increased acid secretion due to associated gastric atrophy. Because of reduced acid secretion, individuals with gastric mucosal atrophy are generally protected from antral and duodenal ulcers. Acid secreted by ectopic gastric mucosa within the duodenum, an ileal Meckel diverticulum, or esophageal ectopic gastric mucosa (an inlet patch) may also cause PUD.

Epidemiology. The incidence of PUD is falling along with reduced prevalence of *H. pylori* infection. However, PUD in patients older than 60 years has increased due to growing NSAID use. This can be amplified by *H. pylori* infection, which synergizes with low-dose aspirin (for cardiovascular benefits) to induce gastric injury. PUD is also associated with cigarette use and cardiovascular disease, likely due to reduced mucosal blood flow, oxygenation, and healing. Other risk factors for PUD are listed in Table 17.3.

Pathogenesis

PUD results from imbalances between defense mechanisms and damaging factors that cause chronic gastritis (discussed earlier). Thus, PUD generally develops on a background of chronic gastritis. The reasons why some people develop only chronic gastritis while others develop PUD are poorly understood. However, as with *H. pylori* gastritis, it is likely that host factors as well as variation in pathogenicity of bacterial strains are involved.

Table 17.3 Risk Factors for Peptic Ulcer Disease

- *Helicobacter pylori* infection
- Cigarette use (synergizes with *H. pylori* for gastric PUD)
- Chronic obstructive pulmonary disease
- Illicit drugs, e.g., cocaine, that reduce mucosal blood flow
- NSAIDs (potentiated by corticosteroids)
- Alcoholic cirrhosis (primarily duodenal PUD)
- Psychological stress (can increase gastric acid secretion)
- Endocrine cell hyperplasia (can stimulate parietal cell growth and gastric acid secretion)
- Zollinger-Ellison syndrome (PUD of stomach, duodenum, and jejunum)
- Viral infection (CMV, herpes simplex virus)

CMV, Cytomegalovirus; *NSAIDs,* non-steroidal anti-inflammatory drugs; *PUD,* peptic ulcer disease.

MORPHOLOGY

Peptic ulcers occur in the context of chronic gastritis but are most frequently found in the proximal duodenum, within a few centimeters of the pyloric valve. Foveolar metaplasia, in which gastric-type mucus cells are present, is common in chronic duodenal peptic disease and may be a protective response, as gastric epithelia are less sensitive to acid than intestinal epithelia. Gastric peptic ulcers are predominantly located along the lesser curvature near the interface of the body and antrum.

Peptic ulcers are solitary in more than 80% of patients and form a round to oval, **sharply punched-out defect** (Fig. 17.15A). The mucosal margin is usually level with the surrounding mucosa but may overhang the base, particularly on the proximal side. In contrast, **heaped-up margins are more characteristic of cancers.** The depth of ulcers correlates with diameter, and deep extension may be limited by the thick gastric muscularis propria, adherent pancreas, omental fat, or the liver. **Perforation** into the

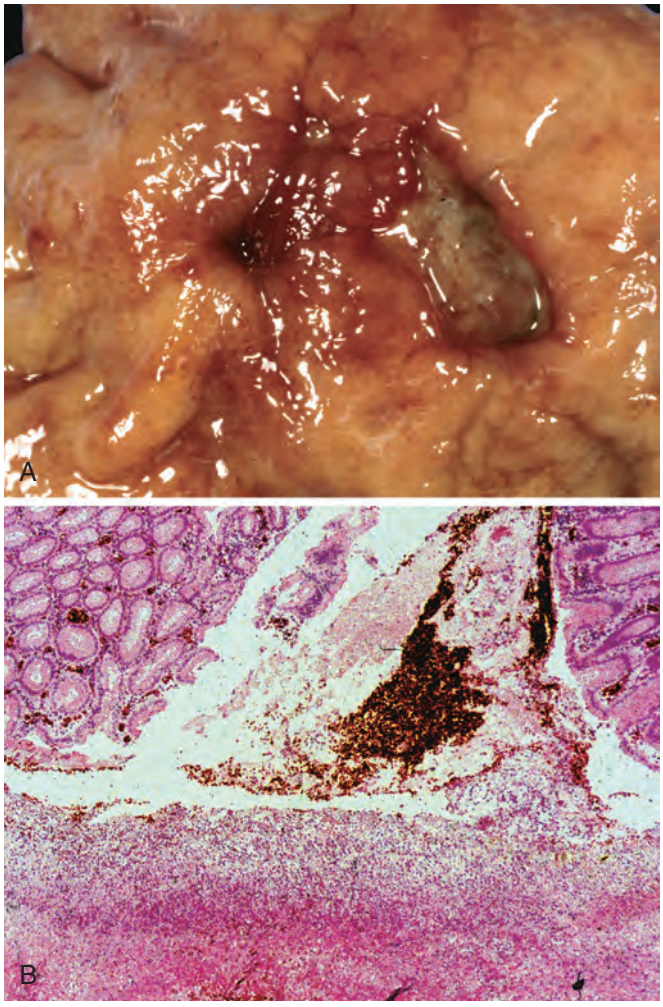


Figure 17.15 Acute gastric perforation in a patient presenting with free air under the diaphragm. (A) Mucosal defect with clean edges. (B) The necrotic ulcer base is composed of granulation tissue.

peritoneal cavity is a surgical emergency that may be identified as free air under the diaphragm on upright abdominal radiographs.

The base of peptic ulcers is smooth and clean as a result of peptic digestion of exudate. Active ulcers may be lined by a thin layer of fibrinoid debris with a predominantly neutrophilic inflammatory infiltrate. Beneath this, **granulation tissue** with immature vessels, mononuclear leukocytes, and a fibrous or collagenous scar forms the ulcer base (Fig. 17.15B). Larger vessels within the scarred area are typically thickened and are occasionally thrombosed. Bleeding from these vessels may cause life-threatening **hemorrhage.** Scarring may involve the entire thickness of the wall and draw the surrounding mucosa into folds that radiate outward.

Clinical Features

Peptic ulcers typically come to clinical attention because of epigastric burning or aching pain. Others present with iron deficiency anemia, hemorrhage, or perforation (Table 17.4). The pain tends to occur 1 to 3 hours after meals during the day, is worse at night (usually between 11 p.m. and 2 a.m.), and is relieved by alkali or food. Nausea, vomiting, bloating, belching, and significant weight loss are additional manifestations. With penetrating ulcers the pain is occasionally referred to the back, left upper quadrant, or chest, where it may be misinterpreted as cardiac in origin.

Current therapies for PUD are aimed at *H. pylori* eradication and neutralization of gastric acid, primarily with proton pump inhibitors. It is also important to withdraw other offending agents such as NSAIDs, including selective COX-2 inhibitors, that may interfere with mucosal healing. While peptic ulcers were previously notoriously difficult to eradicate, the recurrence rate is now less than 20% following successful clearance of *H. pylori*.

Other Complications of Chronic Gastritis

Mucosal Atrophy and Intestinal Metaplasia

As noted above, chronic gastritis may lead to significant loss of parietal cell mass. This oxyntic atrophy is frequently associated with intestinal metaplasia, recognized by the presence of goblet cells, and is strongly associated with increased risk of gastric adenocarcinoma. The risk of adenocarcinoma is greatest in autoimmune metaplastic atrophic gastritis. This may be because achlorhydria of gastric mucosal atrophy permits overgrowth of bacteria that produce carcinogenic nitrosamines. Intestinal metaplasia also occurs in chronic *H. pylori* gastritis but may regress after clearance of the organism.

Table 17.4 Complications of Peptic Ulcer Disease

Bleeding
Occurs in 15%–20% of patients
Most frequent complication
May be life threatening
Accounts for 25% of ulcer-related deaths
May be the first indication of an ulcer
Perforation
Occurs in up to 5% of patients
Accounts for two-thirds of ulcer-related deaths
Is rarely first indication of an ulcer
Obstruction
Mostly in chronic ulcers
Secondary to edema or scarring
Occurs in about 2% of patients
Most often associated with pyloric channel ulcers
May occur with duodenal ulcers
Causes incapacitating, crampy abdominal pain
Can rarely cause total obstruction and intractable vomiting

Dysplasia

Chronic gastritis exposes the epithelium to inflammation-related free radical damage and proliferative stimuli that include regenerative responses to injury and increased production of gastrin, a gastric epithelial mitogen. Over time this combination of stressors can lead to the accumulation and amplification of genetic alterations that result in carcinoma. Preinvasive in situ lesions can be recognized histologically as dysplasia. The morphologic hallmarks of dysplasia are variations in epithelial size, shape, and orientation along with coarse chromatin texture, hyperchromasia, and nuclear enlargement. The distinction between dysplasia and regenerative epithelial changes induced by active inflammation can be a challenge for the pathologist, since increased epithelial proliferation and mitotic figures may be prominent in both. One clue is that reactive epithelial cells mature as they reach the mucosal surface, while dysplastic lesions remain cytologically immature.

Gastritis Cystica

Gastritis cystica is an exuberant reactive epithelial proliferation associated with entrapment of epithelial-lined cysts. These may be found within the submucosa (gastritis cystica polyposa) or deeper layers of the gastric wall (gastritis cystica profunda). Because of the association with chronic gastritis and partial gastrectomy, it is presumed that gastritis cystica is trauma-induced, but the reasons for the development of epithelial cysts within deeper portions of the gastric wall are not clear. Regenerative epithelial changes can be prominent in the entrapped epithelium, and gastritis cystica can therefore be mistaken for invasive adenocarcinoma.

HYPERTROPHIC GASTROPATHIES

Hypertrophic gastropathies are uncommon diseases characterized by giant "cerebriform" enlargement of the rugal folds due to epithelial hyperplasia without inflammation. As might be expected, the hypertrophic gastropathies are linked to excessive growth factor release. Two well-defined examples are Ménétrier disease and Zollinger-Ellison syndrome, the morphologic features of which are compared with other gastric proliferations in Table 17.5.

Ménétrier Disease

Ménétrier disease is a rare disorder associated with excessive secretion of transforming growth factor (TGF)-α, which is an EGF receptor (EGFR) ligand, and hyperactivation of the epidermal growth factor receptor on gastric epithelial cells. Some cases occur in association with viral or *H. pylori* infection. Diffuse hyperplasia of foveolar epithelium within the body and fundus as well as hypoproteinemia due to

Table 17.5 Hypertrophic Gastropathies and Gastric Polyps

Parameter	Ménétrier Disease (Adult)	Zollinger-Ellison Syndrome	Inflammatory and Hyperplastic Polyps	Gastritis Cystica	Fundic Gland Polyps	Gastric Adenomas
Mean patient age, years	30–60	50	50–60	Variable	50	50–60
Location	Body and fundus	Fundus	Antrum > body	Body	Body and fundus	Antrum > body
Predominant cell type	Mucous	Parietal > mucous, endocrine	Mucous	Mucous, cyst-lining	Parietal and chief	Dysplastic, intestinal
Inflammatory infiltrate	Limited, lymphocytes	Neutrophils	Neutrophils and lymphocytes	Neutrophils and lymphocytes	None	Variable
Symptoms	Hypoproteinemia, weight loss, diarrhea	Peptic ulcers	Similar to chronic gastritis	Similar to chronic gastritis	None, nausea	Similar to chronic gastritis
Risk factors	None	Multiple endocrine neoplasia	Chronic gastritis, *H. pylori*	Trauma, prior surgery	PPIs, FAP	Chronic gastritis, atrophy, intestinal metaplasia
Association with adenocarcinoma	Yes	No	Occasional	No	Syndromic (FAP) only	Frequent

FAP, Familial adenomatous polyposis; *PPIs*, proton pump inhibitors.

albumin loss, which can approach 10 g/day, across the gastric mucosa characterize Ménétrier disease. Secondary symptoms such as weight loss, diarrhea, and peripheral edema are commonly present. Systemic effects of TGF-α, including hyperplasia of the liver, pancreas, and GI tract (beyond the stomach), muscle and adipose tissue loss, and psoriasis, may also be present. Symptoms and pathologic features of Ménétrier disease in children are similar to those in adults, but pediatric disease is usually self-limited and often follows CMV or another respiratory infection. Risk of gastric adenocarcinoma is increased in adults with Ménétrier disease.

MORPHOLOGY

Ménétrier disease is characterized by irregular enlargement of the gastric rugae. Some areas may appear polypoid. Enlarged rugae are present in the body and fundus (Fig. 17.16A), but the antrum is generally spared. Histologically, the most characteristic feature is hyperplasia of foveolar mucous cells. The glands are elongated with a corkscrew-like appearance, and cystic dilation is common (Fig. 17.16B). Inflammation is usually modest, although some cases show marked intraepithelial lymphocytosis. Diffuse or patchy glandular atrophy, evident as hypoplasia of parietal and chief cells, is typically intermixed with epithelial hyperplasia.

Treatment of Ménétrier disease is supportive, with intravenous albumin and parenteral nutritional supplementation. In cases associated with herpesvirus, CMV, or *H. pylori,* treatment of the infection may be helpful. Antibodies that block epidermal growth factor receptor activation are effective in many cases. In severe cases gastrectomy remains a therapeutic option.

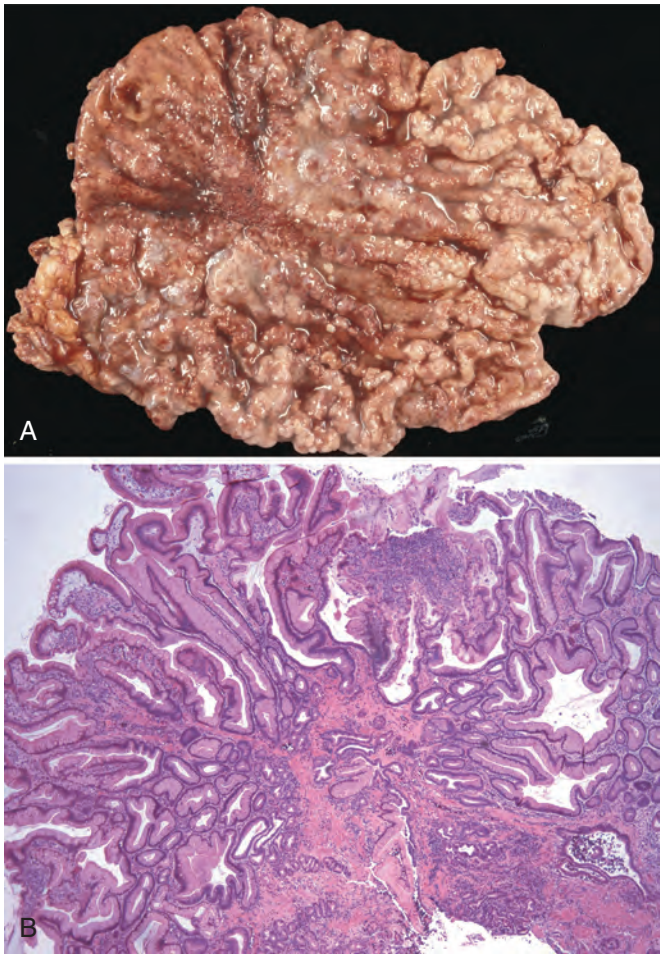


Figure 17.16 Ménétrier disease. (A) Marked hypertrophy of rugal folds. (B) Foveolar hyperplasia with elongated and focally dilated glands. (Courtesy Dr. M. Kay Washington, Vanderbilt University, Nashville, Tenn.)

Zollinger-Ellison Syndrome

Zollinger-Ellison syndrome is caused by gastrin-secreting neuroendocrine tumors. These gastrinomas are almost always found in the small intestine or pancreas. Patients often present with duodenal ulcers or chronic diarrhea. Within the stomach, the most remarkable feature is a marked increase in oxyntic mucosal thickness due to massive parietal cell hyperplasia. Gastrin also induces hyperplasia of mucous neck and endocrine cells within oxyntic mucosa.

Treatment of individuals with Zollinger-Ellison syndrome includes blockade of acid hypersecretion, typically with proton pump inhibitors. Acid suppression allows peptic ulcers to heal and prevents gastric perforation, allowing treatment to focus on the gastrinoma, which then becomes the main determinant of long-term survival.

Gastrinomas are sporadic, solitary lesions in 75% of patients. The remaining 25% of patients with gastrinomas have multiple endocrine neoplasia type 1 (MEN1). These patients are younger and primarily have duodenal tumors that commonly metastasize. *MEN1* gene mutations in sporadic gastrinomas are also associated with aggressive tumor behavior. Somatostatin receptor scintigraphy or endoscopic ultrasonography can help with identification. Patients with metastatic disease may benefit from treatment with somatostatin analogues.

KEY CONCEPTS

GASTRITIS

- Gastritis is a mucosal inflammatory process. When inflammatory cells are absent or rare, the term gastropathy can be applied.
- The clinical spectrum of acute gastritis ranges from asymptomatic disease to mild epigastric pain, nausea, and vomiting. Causative factors include any agent or disease that interferes with gastric mucosal protective mechanisms.
- The most common cause of chronic gastritis is *H. pylori* infection. Other injurious agents include NSAIDs and alcohol.
- *H. pylori* gastritis typically affects the antrum and is associated with increased gastric acid production. Later in the course of disease, the body can be involved, and resulting glandular atrophy can lead to mildly reduced acid production. Host immune responses and bacterial characteristics determine whether the infection remains antral or progresses to pangastritis and atrophy.
- *H. pylori* gastritis induces MALT that can give rise to B-cell lymphomas (MALTomas).
- After *H. pylori* and NSAIDs, autoimmune atrophic gastritis is the most frequent cause of chronic gastritis. It results in atrophy of the gastric body oxyntic glands, which leads to decreased gastric acid production, antral G-cell hyperplasia, and vitamin B_{12} deficiency. Anti–parietal cell and anti–intrinsic factor antibodies are typically present.
- Intestinal metaplasia develops in all forms of chronic gastritis and is a risk factor for gastric adenocarcinoma.
- Peptic ulcer disease is usually secondary to chronic *H. pylori*–induced gastritis and the resulting hyperchlorhydria. Ulcers can develop in the stomach or duodenum and usually heal after suppression of gastric acid production and *H. pylori* eradication.
- Ménétrier disease is a rare disorder caused by excessive secretion of TGF-α and characterized by diffuse foveolar hyperplasia and protein-losing enteropathy.
- Zollinger-Ellison syndrome is caused by gastrin-secreting tumors that cause parietal cell hyperplasia and acid hypersecretion; 60% to 90% of these gastrinomas are malignant.

GASTRIC POLYPS AND TUMORS

Polyps, nodules or masses that project above the level of the surrounding mucosa, are identified in up to 5% of upper GI endoscopies. Polyps may develop as a result of epithelial or stromal cell hyperplasia, inflammation, ectopia, or neoplasia. Only the most common types of gastric polyps will be discussed here (Peutz-Jeghers and juvenile polyps are discussed with intestinal polyps later).

Inflammatory and Hyperplastic Polyps

Up to 75% of all gastric polyps are inflammatory or hyperplastic polyps. Their incidence is correlated with the regional prevalence of *H. pylori* infection. These polyps are most common in individuals between 50 and 60 years of age and usually develop in association with chronic gastritis, which initiates the injury that leads to reactive hyperplasia and polyp growth. The risk of dysplasia in inflammatory polyps ranges from 1% to 20% and increases sharply in pedunculated polyps greater than 1 cm in diameter.

MORPHOLOGY

The majority of inflammatory or hyperplastic polyps are smaller than 1 cm in diameter. Polyps are frequently multiple, particularly in individuals with atrophic gastritis. Hyperplastic polyps are ovoid in shape and have a smooth surface; superficial erosions are common in inflammatory polyps. Microscopically, polyps have irregular, cystically dilated, and elongated foveolar glands (Fig. 17.17A). The lamina propria is typically edematous with variable degrees of acute and chronic inflammation (Fig. 17.17B).

Fundic Gland Polyps

Fundic gland polyps occur sporadically and in individuals with germline mutations in the *APC* gene (the cause of familial adenomatous polyposis) or the DNA repair gene *MUTYH* (both discussed later). The prevalence of sporadic fundic gland polyps has increased markedly in recent years as a result of increased use of proton pump inhibitors. These drugs inhibit acid production, which leads to increased gastrin secretion and, in turn, oxyntic gland growth. Fundic gland polyps may be asymptomatic or associated with nausea, vomiting, or epigastric pain.

MORPHOLOGY

Fundic gland polyps occur in the gastric body and fundus and are well-circumscribed lesions with a smooth surface. They may be

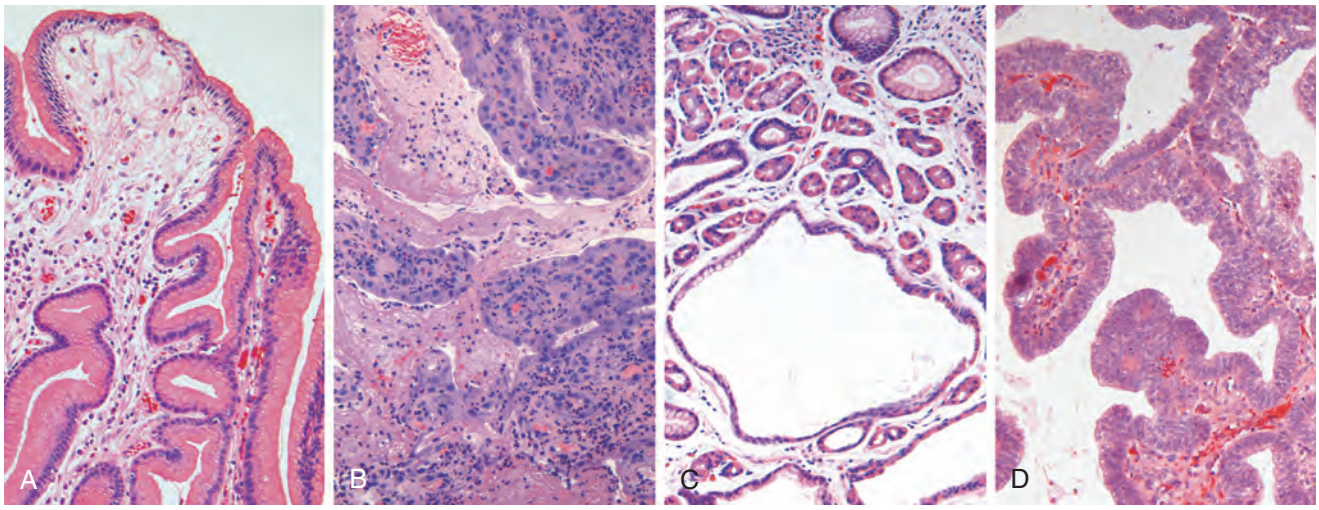


Figure 17.17 Gastric polyps. (A) Hyperplastic polyp containing corkscrew-shaped foveolar glands. (B) Hyperplastic polyp with extensive ulceration. (C) Fundic gland polyp composed of cystically dilated glands lined by parietal, chief, and foveolar cells. (D) Gastric adenoma with epithelial nuclear hyperchromasia and pseudostratification.

single or multiple and are composed of cystically dilated, irregular glands lined by parietal, chief, and foveolar mucus cells. Inflammation is typically absent or minimal (Fig. 17.17C). Dysplasia may occur in FAP-associated fundic gland polyps, and rare cancers have been reported. In contrast, sporadic fundic gland polyps, including those arising in association with proton pump inhibitor therapy, do not carry a significant risk of neoplasia.

Gastric Adenoma

Most gastric adenomas develop in a background of chronic gastritis with atrophy and intestinal metaplasia, with the exception of those associated with germline mutations in *APC* or *MUTYH*. They represent up to 10% of all gastric polyps (see Table 17.5). Their frequency increases progressively with age, and there is a marked variation in prevalence among different populations that parallels the incidence of gastric adenocarcinoma. Patients are usually between 50 and 60 years of age, and males are affected three times more often than females. The risk of adenocarcinoma within gastric adenomas correlates with size and is particularly elevated in lesions greater than 2 cm in diameter. Gastric adenomas have a greater risk of cancer (up to 30% when large) and must be managed more aggressively than colonic adenomas.

MORPHOLOGY

Gastric adenomas are usually solitary antral lesions. All GI adenomas exhibit dysplasia, which is classified as low- or high-grade (Fig. 17.17D). Both grades of dysplasia are characterized by nuclear enlargement, elongation, pseudostratification, and hyperchromasia. High-grade dysplasia is marked by more severe cytologic atypia and architectural abnormalities, including glandular budding and gland-within-gland, or cribriform, structures.

Gastric Adenocarcinoma

Adenocarcinoma is the most common malignancy of the stomach, comprising more than 90% of all gastric cancers. As discussed in more detail later, gastric adenocarcinoma is separated morphologically into intestinal type, which tends to form bulky masses, and diffuse type, which infiltrates and thickens the gastric wall. Early symptoms of both types of gastric adenocarcinoma resemble those of chronic gastritis and peptic ulcer disease, including dyspepsia, dysphagia, and nausea. As a result, these tumors are often discovered at advanced stages, when symptoms such as weight loss, anorexia, early satiety (primarily in diffuse cancers), anemia, and hemorrhage appear.

Epidemiology. Gastric cancer incidence varies markedly with geography. In Japan, Chile, Costa Rica, and Eastern Europe, the incidence is up to 20-fold higher than in North America, northern Europe, Africa, and Southeast Asia. Mass endoscopic screening programs have been successful in regions where the incidence is high, such as Japan, where 35% of newly detected cases are early gastric cancers, limited to the mucosa and submucosa. Unfortunately, mass screening programs are not cost-effective in low-incidence regions, and fewer than 20% of cases are detected at an early stage in North America and northern Europe. Metastases are often present at diagnosis. Sites most commonly involved include supraclavicular sentinel lymph nodes (Virchow node), periumbilical lymph nodes (Sister Mary Joseph nodule), ovaries (Krukenberg tumor), the left axillary lymph node, and the pouch of Douglas.

Gastric cancer is more common in lower socioeconomic groups and in individuals with multifocal mucosal atrophy and intestinal metaplasia. As noted earlier, peptic ulcer disease is associated with reduced risk of gastric cancer, but patients who have had partial gastrectomies for peptic ulcer disease have a slightly higher risk of developing cancer in the residual gastric stump, possibly due to hypochlorhydria and bile reflux.

Adenocarcinoma of the stomach was the most common cause of cancer death in the United States in 1930. However, since that time the incidence has fallen by 85%, and, gastric adenocarcinoma now accounts for only 2.5% of cancer deaths in the United States. Decreased gastric cancer incidence is largely attributed to reduced rates of *H. pylori* infection and primarily relates to intestinal-type cancers. Other environmental and dietary factors, including decreased consumption of dietary carcinogens such as N-nitroso compounds and benzo[*a*]pyrene (because of the reduced use of salt and smoking for food preservation) and the widespread availability of refrigeration may have also contributed to the decreased incidence. Consistent with an environmental, rather than genetic, cause, migrants from high- to low-risk regions maintain the risk of their original country, but their children have gastric cancer rates similar to those in the new country of residence.

Although the overall incidence of gastric adenocarcinoma is falling, cancer of the gastric cardia is on the rise. This is probably related to Barrett esophagus and may reflect the increasing incidence of chronic GERD and obesity. Consistent with this presumed shared pathogenesis, gastric cardia adenocarcinomas and distal esophageal adenocarcinomas are similar in morphology, clinical behavior, and therapeutic responses.

Pathogenesis

While the majority of gastric cancers are not hereditary, mutations identified in familial gastric cancer have provided important insights into mechanisms of carcinogenesis in sporadic cases. Familial gastric cancer is strongly associated with germline loss-of-function mutations in the tumor suppressor gene *CDH1,* which encodes the cell adhesion protein E-cadherin (discussed in Chapter 7). Loss-of-function mutations in *CDH1* are also present in about 50% of sporadic diffuse gastric tumors, while E-cadherin expression is drastically decreased in the remainder of diffuse tumors, often by hypermethylation and silencing of the *CDH1* promoter. **Thus, E-cadherin loss is a key step in the development of diffuse gastric cancer.** *CDH1* mutations are also common in sporadic and familial lobular carcinoma of the breast, which, like diffuse gastric cancer (see later), tends to infiltrate as single cells.

In contrast to diffuse gastric cancers, **intestinal-type gastric cancers are strongly associated with mutations that result in increased signaling via the Wnt pathway.** These include loss-of-function mutations in the adenomatous polyposis coli *(APC)* tumor suppressor gene and gain-of-function mutations in the gene encoding β-catenin. Other genes commonly affected by loss-of-function mutations or silencing include those involved in TGF-β signaling *(TGFβRII),* regulation of apoptosis *(BAX),* and cell cycle control *(CDKN2A),* all of which are discussed in more detail in Chapter 7. FAP patients with germline *APC* mutations have an increased risk of intestinal-type gastric cancer, particularly in those who reside in high-risk areas like Japan. Variants of pro-inflammatory genes such as IL-1β and IL-1 receptor, are associated with elevated risk of gastric cancer in those who have *H. pylori* gastritis. Thus, both host genetic background and environmental factors affect risk. Other associations between chronic inflammation and cancer are discussed later in the context of inflammatory bowel disease (IBD) and in Chapter 7.

MORPHOLOGY

Gastric adenocarcinomas are classified according to their location and gross and histologic morphology. Most distal gastric adenocarcinomas occur in the gastric antrum; the lesser curvature is involved more often than the greater curvature. Gastric tumors with an **intestinal morphology form bulky tumors** (Fig. 17.18A) and are composed of glandular structures (Fig. 17.19A), while cancers with a **diffuse infiltrative growth pattern** (Fig. 17.18B) are typically composed of **signet-ring cells** (Fig. 17.19B). Although they may penetrate the gastric wall, intestinal-type adenocarcinomas frequently grow along broad cohesive fronts to form either an exophytic mass or an ulcerated infiltrative tumor. The neoplastic cells often contain apical mucin vacuoles, and abundant mucin may be present in gland lumina. In contrast, diffuse cancers permeate the gastric wall as small clusters and individual discohesive cells due to the absence of E-cadherin. These cells do not form glands but instead have large mucin vacuoles that expand the cytoplasm and push the nucleus to the periphery, creating a signet-ring cell morphology. They be mistaken for inflammatory cells at low magnification. Release of extracellular

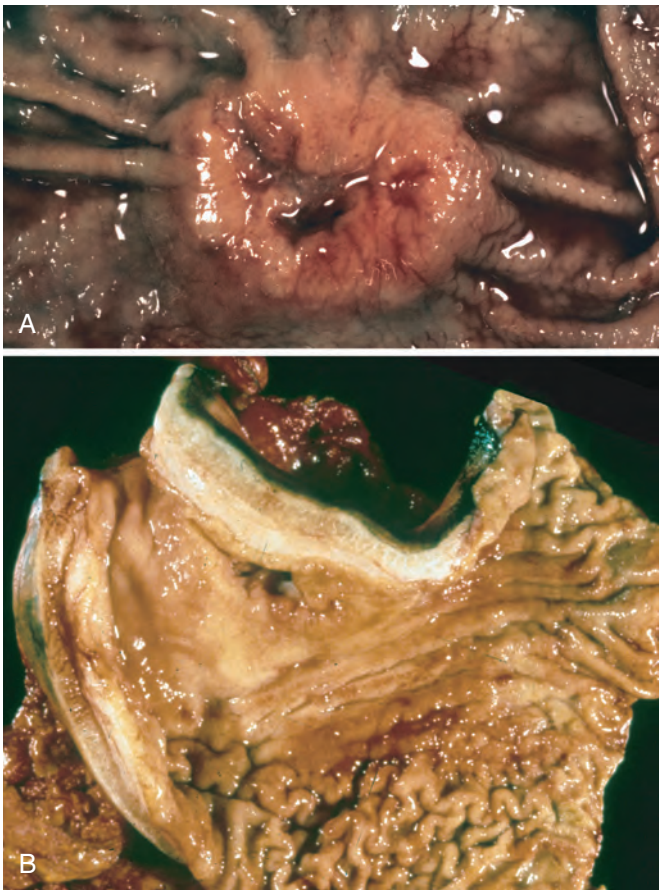


Figure 17.18 Gastric adenocarcinoma. (A) Intestinal-type adenocarcinoma consisting of an elevated mass with heaped-up borders and central ulceration. Compare to the peptic ulcer in Fig. 17.15A. (B) Infiltrative type (linitis plastica) gastric cancer. The gastric wall is markedly thickened, and rugal folds are partially lost, but there is no dominant mass.

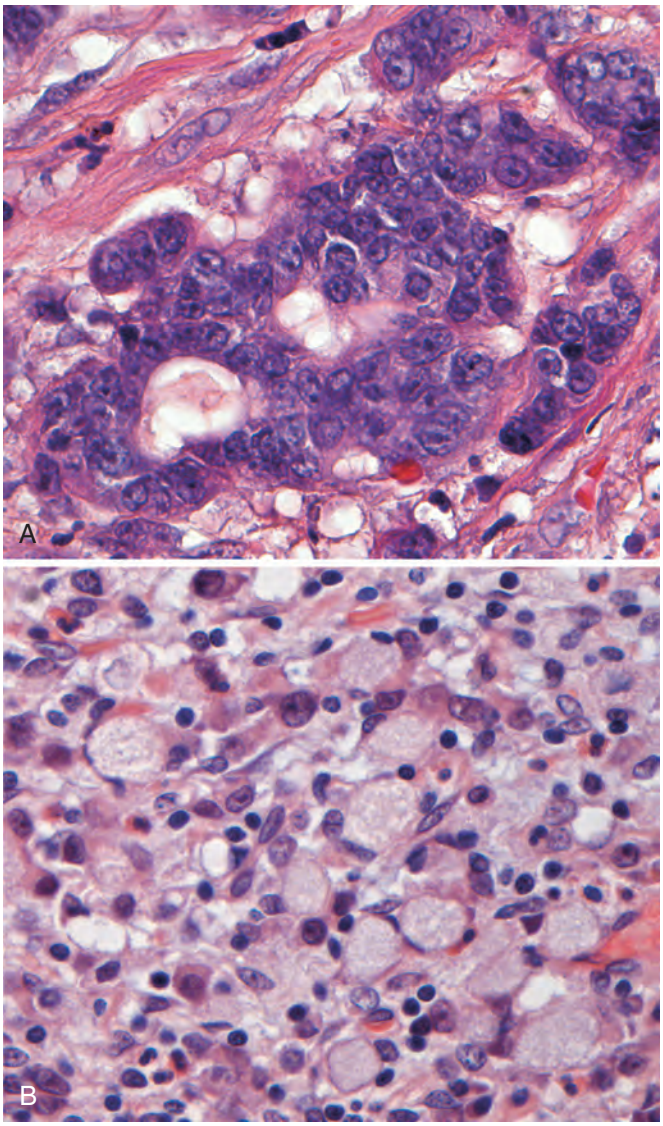


Figure 17.19 Gastric adenocarcinoma. (A) Intestinal-type adenocarcinoma composed of columnar, gland-forming cells infiltrating through desmoplastic stroma. (B) In infiltrative tumors, signet-ring cells can be recognized by their large cytoplasmic mucin vacuoles and peripherally displaced, crescent-shaped, thin nuclei. Note the absence of gland formation.

mucin in either type of gastric cancer can result in formation of large mucin lakes that dissect tissue planes.

A mass may be difficult to appreciate in diffuse gastric cancer, but these infiltrative tumors often evoke a **desmoplastic** reaction that stiffens the gastric wall. When there are large areas of infiltration, diffuse rugal flattening and a rigid, thickened wall may impart a leather bottle appearance termed **linitis plastica** (Fig. 17.18B).

Clinical Features

Intestinal-type gastric cancer predominates in high-risk areas and develops from precursor lesions including flat dysplasia and adenomas. The mean age of presentation is 55 years, and the male-to-female ratio is 2:1. In contrast, the incidence of diffuse gastric cancer is relatively uniform across countries, and there are no identified precursor lesions; prevalence is similar in males and females. Notably, the remarkable decrease in gastric cancer incidence applies only to the intestinal type. As a result, the incidences of intestinal- and diffuse-type gastric cancers are now similar in many regions.

The depth of invasion and the extent of nodal and distant metastases remain the most powerful prognostic indicators in gastric cancer. Local invasion into the duodenum, pancreas, and retroperitoneum is common. In such cases efforts are usually focused on chemotherapy or radiation therapy and palliative care. When possible, surgery remains the preferred treatment approach. With surgical resection, the 5-year survival rate of early gastric cancer can exceed 90%, even if lymph node metastases are present. In contrast, the 5-year survival rate for advanced gastric cancer remains less than 20%. Because of the advanced stage at which most gastric cancers are discovered in the United States, the overall 5-year survival is less than 30%.

Lymphoma

Although extranodal lymphomas can arise in virtually any tissue, they do so most commonly in the GI tract, particularly the stomach. In allogeneic hematopoietic stem cell and organ transplant recipients, the bowel is also the most frequent site for Epstein-Barr virus–positive B-cell lymphoproliferations. This preferential location is most likely because the deficits in T-cell function caused by oral immunosuppressive agents (e.g., cyclosporine) are greatest at intestinal sites of drug absorption. Nearly 5% of all gastric malignancies are primary lymphomas, the most common of which are indolent extranodal marginal zone B-cell lymphomas. In the gut these tumors are often referred to as lymphomas of MALT, or MALTomas. These and other lymphomas of the gut are discussed in Chapter 13.

Pathogenesis

Extranodal marginal zone B-cell lymphomas usually arise at sites of chronic inflammation. They can originate in the GI tract at sites of preexisting MALT such as the Peyer patches of the small intestine, but more commonly arise within tissues that are normally devoid of organized lymphoid tissue. MALT is not present in the normal stomach but can be induced, typically as a result of chronic gastritis. *H. pylori* infection is the most common inducer of gastric MALT and therefore is found in association with most gastric MALTomas.

Three translocations are associated with gastric MALToma: t(11;18)(q21;q21), the less common t(1;14)(p22;q32), and t(14;18)(q32;q21). The t(11;18)(q21;q21) translocation brings together the apoptosis inhibitor 2 *(API2)* gene on chromosome 11 with the "mutated in MALT lymphoma" *(MLT1)* gene on chromosome 18. This creates a chimeric *API2-MLT1* fusion gene that encodes an API2-MALT1 fusion protein. The t(14;18)(q32;q21) and t(1;14)(p22;q32) translocations cause increased expression of intact MALT1 and BCL-10 proteins, respectively.

Each of the three translocations leads to constitutive activation of nuclear factor kappa B (NF-κB), a transcription factor that promotes B-cell growth and survival. Antigen-dependent activation of NF-κB in normal B and T cells requires both BCL-10 and MALT1, which work together in a signal transduction pathway downstream of lymphocyte

antigen receptors. Thus, *H. pylori*–induced inflammation may trigger NF-κB activation through the MALT1/BCL-10 pathway in MALTomas that lack these translocations. Removal of this stimulus may explain why these tumors tend to respond to *H. pylori* eradication. In contrast, tumors bearing translocations involving *MALT1* or *BCL10* are generally insensitive to *H. pylori* clearance. Other tumor characteristics such as invasion into or through the muscularis propria and lymph node involvement also correlate with resistance to *H. pylori* eradication.

As with other low-grade lymphomas, MALTomas can transform into more aggressive tumors that are histologically identical to diffuse large B-cell lymphomas. This is often associated with additional genetic changes such as inactivation of the tumor suppressor genes that encode p53 and p16.

MORPHOLOGY

Histologically, gastric MALToma takes the form of a dense lymphocytic infiltrate in the lamina propria (Fig. 17.20A). The neoplastic lymphocytes infiltrate gastric glands to create **diagnostic lymphoepithelial lesions** (Fig. 17.20A, inset). Reactive-appearing B-cell follicles may be present, and in about 40% of tumors plasmacytic differentiation is observed. At other sites GI lymphomas may disseminate as discrete small nodules (Fig. 17.20B) or infiltrate the wall diffusely (Fig. 17.20C).

Like other tumors of mature B cells, MALTomas express the B-cell markers CD19 and CD20. They do not express CD5 or CD10, but are positive for CD43 in about 25% of cases, an unusual feature that can be diagnostically helpful. In cases lacking lymphoepithelial lesions, monoclonality may be demonstrated by restricted expression of either κ or λ immunoglobulin light chains or by molecular detection of clonal IgH rearrangements.

Clinical Features

The most common presenting symptoms are dyspepsia and epigastric pain. Hematemesis, melena, and constitutional symptoms such as weight loss can also be present. Because gastric MALTomas and *H. pylori* gastritis often coexist and have overlapping clinical symptoms and endoscopic appearances, diagnostic difficulties may arise, particularly in small biopsy specimens.

Neuroendocrine Neoplasms

These tumors arise from the diffuse components of the endocrine system. Those previously termed carcinoid tumors are now properly referred to as well-differentiated neuroendocrine tumors, although the term carcinoid continues to be used informally in the GI tract and formally at other sites. Most are found in the GI tract, and more than 40% occur in the small intestine. The tracheobronchial tree and lungs are the next most commonly involved sites. Gastric neuroendocrine neoplasms may be associated with endocrine cell hyperplasia, autoimmune chronic atrophic gastritis, MEN1, and Zollinger-Ellison syndrome. Gastric endocrine and enterochromaffin-like cell hyperplasia has been linked to proton pump inhibitor therapy, but the risk of progression to a neuroendocrine neoplasm in this circumstance is negligible.

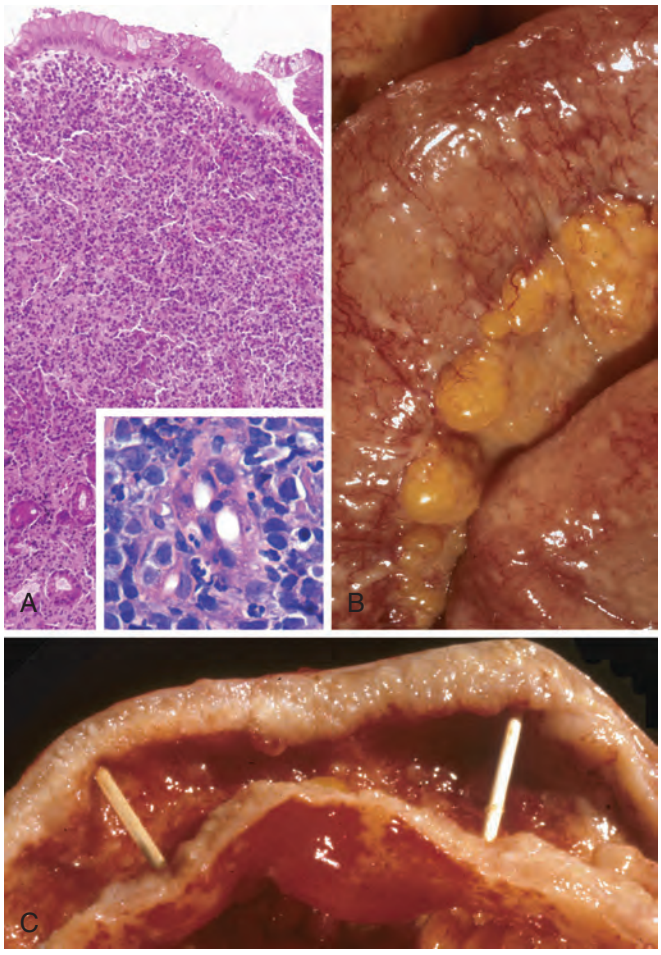


Figure 17.20 Lymphoma. (A) Gastric mucosa-associated lymphoid tissue lymphoma replacing much of the gastric epithelium. *Inset* shows lymphoepithelial lesions with neoplastic lymphocytes surrounding and infiltrating gastric glands. (B) Disseminated lymphoma within the small intestine with numerous small serosal nodules. (C) Large B-cell lymphoma infiltrating the small intestinal wall and producing diffuse thickening.

MORPHOLOGY

Grossly, neuroendocrine neoplasms are intramural or submucosal masses that create polypoid lesions (Fig. 17.21A). In the stomach they typically arise within oxyntic mucosa. The overlying mucosa may be intact or ulcerated, and in the intestines the tumors may invade deeply to involve the mesentery. Neuroendocrine neoplasms tend to be yellow or tan in color and are very firm as a consequence of intense desmoplasia, which may cause kinking and obstruction of the small bowel. Histologically, well-differentiated neuoroendocrine tumors are composed of islands, trabeculae, strands, glands, and sheets of uniform cells with scant, pink granular cytoplasm and a round to oval nucleus with a "salt and pepper" chromatin pattern (Fig. 17.21). Immunohistochemical stains are positive for endocrine granule markers such as synaptophysin and chromogranin A. Poorly differentiated neoplasms with high mitotic rates and Ki-67 proliferative indices are termed neuroendocrine carcinomas and subclassified as small cell or large cell types. Similar to neuroendocrine carcinomas at extraintesintal sites, mutation of *TP53* and *RB* is common.

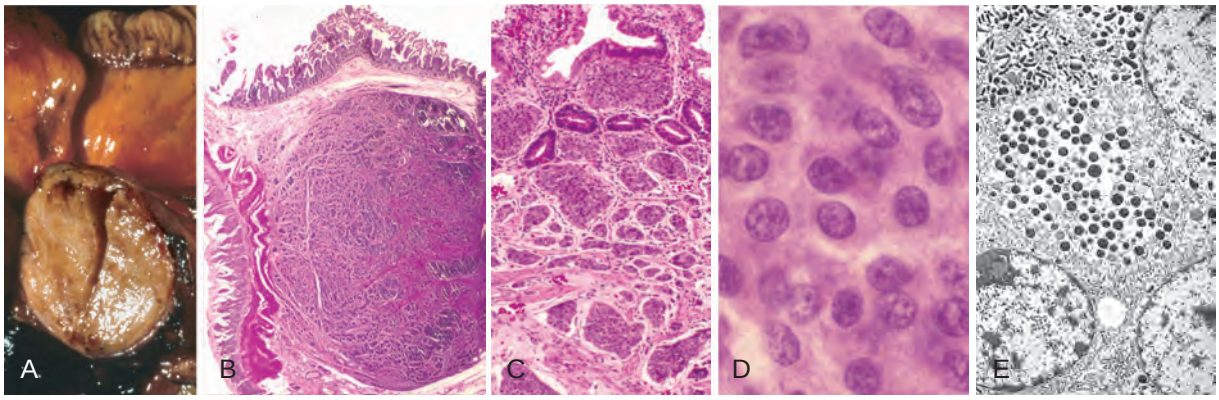


Figure 17.21 Neuroendocrine tumor (carcinoid tumor). (A) Gross cross section of a submucosal tumor nodule. (B) Microscopically the nodule is composed of tumor cells embedded in dense fibrous tissue. (C) In other areas the tumor has spread extensively within mucosal lymphatic channels. (D) High magnification shows the characteristically bland cytology. The chromatin texture, with fine and coarse clumps, is frequently described as a "salt and pepper" pattern. Despite their innocuous appearance, these tumors can be clinically aggressive. (E) Electron microscopy reveals cytoplasmic dense core neurosecretory granules.

Clinical Features

The peak incidence of GI neuroendocrine neoplasms is in the sixth decade, but they may appear at any age. Symptoms reflect hormones released by the tumor cells. For example, those that produce gastrin cause Zollinger-Ellison syndrome. Carcinoid syndrome, which develops in fewer than 10% of patients, is caused by vasoactive substances secreted by the tumor into the systemic circulation. This can lead to cutaneous flushing, sweating, bronchospasm, colicky abdominal pain, diarrhea, and right-sided cardiac valvular fibrosis. When neuroendocrine neoplasms are confined to the intestine, vasoactive substances released are metabolized to inactive forms by the liver, a "first-pass" effect similar to that exerted on oral drugs. Although this can be overcome by a large tumor burden or when tumors secrete hormones into a nonportal venous circulation, the carcinoid syndrome remains strongly associated with metastatic disease to sizes that allow tumor products to bypass the portal circulation and liver. Notably, although a marker of advanced disease, metastatic behavior alone is insufficient to classify a neoplasm as a neuorendocrine carcinoma.

The most important prognostic factors for GI neuroendocrine neoplasms are degree of histological differentiation, mitotic rate, and Ki-67 proliferative index. Behavior is also affected by size and location, and staging is region-specific.

- *Foregut neuroendocrine tumors* within the stomach and duodenum rarely metastasize and are generally cured by resection. This is particularly true for tumors that arise in association with atrophic gastritis. Gastric neuroendocrine tumors without predisposing factors are often more aggressive.
- *Midgut neuroendocrine tumors* arise in the jejunum and ileum, are often multiple, and tend to be aggressive.
- *Hindgut neuroendocrine tumors* arising in the appendix and colorectum are typically discovered incidentally. Those in the appendix occur at any age and are generally located at the tip. These tumors are rarely more than 2 cm in diameter and are almost always benign. Rectal neuroendocrine tumors tend to produce polypeptide hormones and, when symptomatic, present with abdominal pain and weight loss. Because they are usually discovered when small, metastasis of rectal neuroendocrine tumors is uncommon.

Gastrointestinal Stromal Tumor

Gastrointestinal stromal tumor (GIST) is the most common mesenchymal tumor of the abdomen, with an annual incidence between 11 and 20 per 1 million people. More than half of these tumors occur in the stomach. The term *stromal* reflects historical confusion about the origin of this tumor, which is now recognized to arise from the interstitial cells of Cajal, or pacemaker cells. A wide variety of other mesenchymal neoplasms may arise in the stomach. Many are named according to the cell type they most resemble; for example, smooth muscle tumors are called leiomyomas or leiomyosarcomas, nerve sheath tumors are termed schwannomas, and tumors resembling glomus bodies (vascular structures found in the nail beds and other sites) are referred to as glomus tumors. These are all uncommon in the gut and are discussed in Chapter 26.

Epidemiology. Clinically silent, microscopic proliferations that may represent precursors to GIST are common. These foci have low mitotic rates and extremely low risk of neoplastic transformation. Changes associated with progression to overt GIST are not well defined, but loss or partial deletion of chromosomes 9, 14, or 22 is common.

GISTs are diagnosed at a peak age of 60 years, with fewer than 10% occurring before 40 years of age. Of the uncommon GISTs in children, some are related to the Carney triad, a nonhereditary syndrome of GIST, paraganglioma, and pulmonary chondroma that occurs primarily in females. There is also an increased incidence of GIST in individuals with neurofibromatosis type 1.

Pathogenesis

Approximately 75% of all GISTs have oncogenic, gain-of-function mutations in the receptor tyrosine kinase KIT. Another 8% of GISTs have mutations that activate the related receptor tyrosine kinase, platelet-derived growth factor receptor α (PDGFRA). Constitutively active KIT and PDGFRA trigger the same downstream signaling pathways. Mutations of the correspondence genes are therefore mutually exclusive. Rare individuals with germline mutations of these genes have diffuse hyperplasia of Cajal cells and develop multiple GISTs. GISTs without mutated *KIT* or *PDGFRA* often have mutations in genes encoding components of the mitochondrial succinate dehydrogenase complex *(SDHA, SDHB, SDHC, SDHD),* an interesting (and still poorly understood) example of oncogenic mutations affecting components of a metabolic pathway, specifically the Krebs cycle. These mutations, which cause loss of succinate dehydrogenase function, are often inherited in the germline and confer an increased risk for GIST and paraganglioma (Carney-Stratakis syndrome), with the second copy of the affected gene being either mutated or lost in the tumor.

MORPHOLOGY

Primary gastric GISTs can reach 30 cm in diameter. They usually form a solitary, well-circumscribed, fleshy mass. The overlying mucosa is frequently ulcerated, and the cut tumor surface has a whorled appearance. Metastases can form multiple serosal nodules throughout the peritoneal cavity or within the liver, but spread outside of the abdomen is uncommon. GISTs composed of thin elongated cells are classified as **spindle cell type** (Fig. 17.22A), while GISTs dominated by plumper, epithelial-appearing cells are termed **epithelioid type** (Fig. 17.22B); mixtures of the two patterns are common. The most useful diagnostic marker is KIT (CD117), which is immunohistochemically detectable in Cajal cells and 95% of gastric GISTs. Nuclear pleomorphism is uncommon.

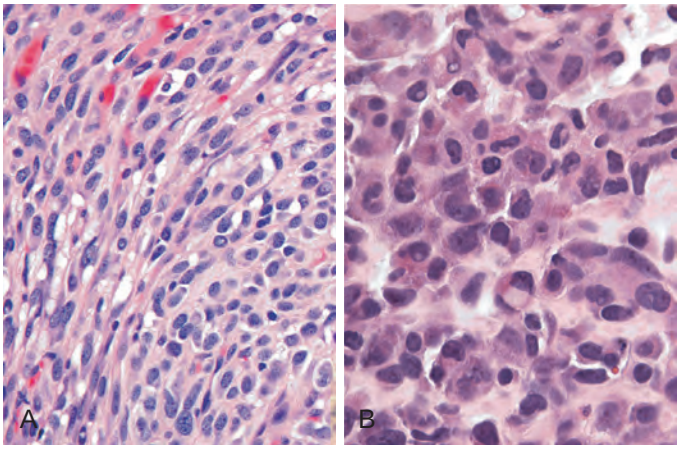


Figure 17.22 Gastrointestinal stromal tumor. (A) Histologically, typical spindle cell gastrointestinal stromal tumors are composed of bundles, or fascicles, of spindle-shaped tumor cells. (B) Epithelioid cell gastrointestinal stromal tumors are composed of plump, epithelial-appearing cells without well-defined fascicles. These tumors can mimic metastatic adenocarcinoma.

Clinical Features

Symptoms of GISTs at presentation are typically related to mass effects. Mucosal ulceration can cause blood loss, and approximately half of individuals with GISTs present with anemia or related symptoms. GISTs may also be discovered incidentally during radiologic imaging, endoscopy, or abdominal surgery performed for other reasons. Complete surgical resection is the primary treatment for localized gastric GIST. The prognosis correlates with tumor size, mitotic index, and location, with gastric GISTs being less aggressive than those arising in the small intestine. Recurrence or metastasis is rare for gastric GISTs smaller than 5 cm but common for mitotically active tumors larger than 10 cm. Many tumors fall into an intermediate category where the malignant potential of the lesion cannot be predicted.

The molecular phenotype is an important consideration in the treatment of patients with unresectable, recurrent, or metastatic GISTs. Tumors with mutations in *KIT* or *PDGFRA* often respond to the tyrosine kinase inhibitor imatinib, whereas tumors without these mutations are generally resistant. In treated patients, imatinib resistance often develops as a result of secondary *KIT* or *PDGFRA* mutations; such tumors may respond to other tyrosine kinase inhibitors that bypass the effects of mutations that convey resistance to imatinib.

KEY CONCEPTS

NEOPLASTIC AND NON-NEOPLASTIC PROLIFERATIONS OF THE STOMACH

- The majority of gastric polyps are inflammatory or hyperplastic polyps. These reactive lesions are associated with chronic gastritis.
- Fundic gland polyps occur sporadically, most often as a consequence of proton pump inhibitor therapy and in patients with familial adenomatous polyposis and *MUTYH*-associated polyposis.
- Gastric adenomas develop in a background of chronic gastritis and are strongly linked to intestinal metaplasia and mucosal (glandular) atrophy. Adenocarcinoma is frequent in gastric adenomas, which therefore require more aggressive therapy than adenomas of the colon.
- Gastric adenocarcinoma incidence varies markedly with geography. Individual tumors are classified according to location and gross and histologic morphology. Gastric tumors with an intestinal histology tend to form bulky tumors and may be ulcerated, while those composed of signet-ring cells typically display a diffuse infiltrative growth pattern that may thicken the gastric wall without forming a discrete mass. Gastric adenocarcinomas are linked to chronic gastritis.
- Primary gastric lymphomas are most often indolent B-cell tumors of mucosa-associated lymphoid tissue (MALTomas) that arise in the setting of chronic *H. pylori* gastritis.
- Neuroendocrine neoplasms arise from diffuse components of the endocrine system and are most common in the GI tract, particularly the small intestine. They can be subdivided into neuroendocrine tumors (previously termed carcinoid tumors)

and neuroendocrine carcinomas on the basis of histology and immunohistochemistry for the proliferative marker Ki-67. Prognosis is also affected by location; tumors of the small intestine tend to be most aggressive, while those of the appendix are typically benign.

- GIST is the most common mesenchymal tumor of the abdomen, occurs most often in the stomach, and develops from Cajal (pacemaker) cells. Most tumors have activating mutations in either *KIT* or *PDGFRA* tyrosine kinases and respond to targeted tyrosine kinase inhibitors.

Small Intestine and Colon

The small intestine and colon make up the majority of the GI tract and are the sites of a broad array of diseases. Some of these relate to nutrient and water transport, perturbation of which results in malabsorption and diarrhea. The intestines are also the principal site where the immune system interfaces with a diverse array of antigens present in food and gut microbes. Thus, it is not surprising that the small intestine and colon are frequently affected by infectious and inflammatory disorders. Finally, the colon is the most common site of GI neoplasia in Western populations.

INTESTINAL OBSTRUCTION

Obstruction of the GI tract may occur at any level, but the small intestine is most often involved because of its relatively narrow lumen. Collectively, hernias, intestinal adhesions, intussusception, and volvulus account for 80% of mechanical obstructions (Fig. 17.23). Tumors, infarction, and other causes of strictures, for example, Crohn disease, make up the remainder. The clinical manifestations of intestinal obstruction include abdominal pain and distention, vomiting, and constipation. Surgical intervention is usually required in cases where the obstruction has a mechanical basis or is associated with bowel infarction.

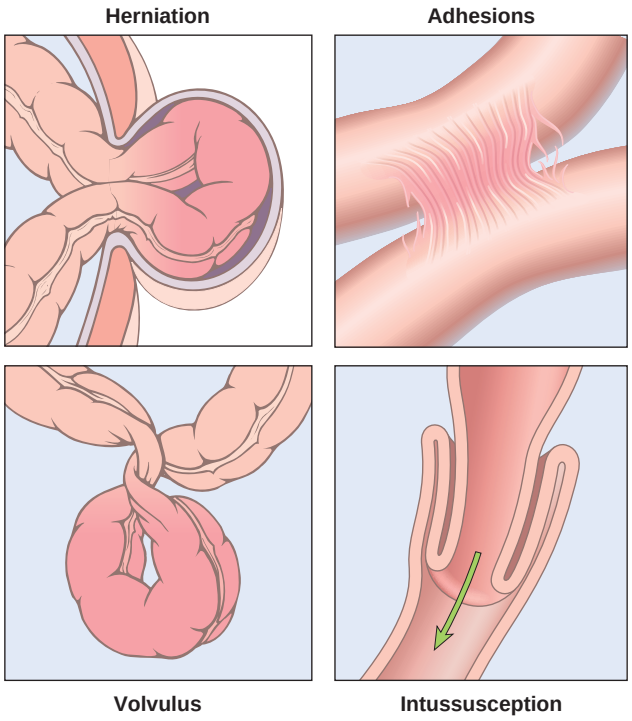


Figure 17.23 Intestinal obstruction. The four major causes of intestinal obstruction are (1) herniation of a segment in the umbilical or inguinal regions, (2) adhesion between loops of intestine, (3) volvulus, and (4) intussusception.

Hernias

Any weakness or defect in the abdominal wall may permit protrusion of a serosa-lined pouch of peritoneum called a hernia sac. Acquired hernias typically occur anteriorly, via the inguinal and femoral canals, via the umbilicus, or at sites of surgical defects and occur in up to 5% of the population. **Hernias are the most frequent cause of intestinal obstruction worldwide** and the third most common cause of obstruction in the United States. Obstruction usually occurs because of visceral protrusion (external herniation) and is most frequently associated with inguinal hernias, which tend to have narrow orifices and large sacs. Small bowel loops are typically involved, but omentum or large bowel may also protrude, and any of these may become entrapped. Pressure at the neck of the pouch may impair venous drainage of the entrapped viscus. The resultant stasis and edema increase the bulk of the herniated loop, leading to permanent entrapment (incarceration) and, over time, arterial and venous compromise (strangulation) and infarction (Fig. 17.24).

Adhesions

Surgical procedures, infection, or other causes of peritoneal inflammation, such as endometriosis, may lead to

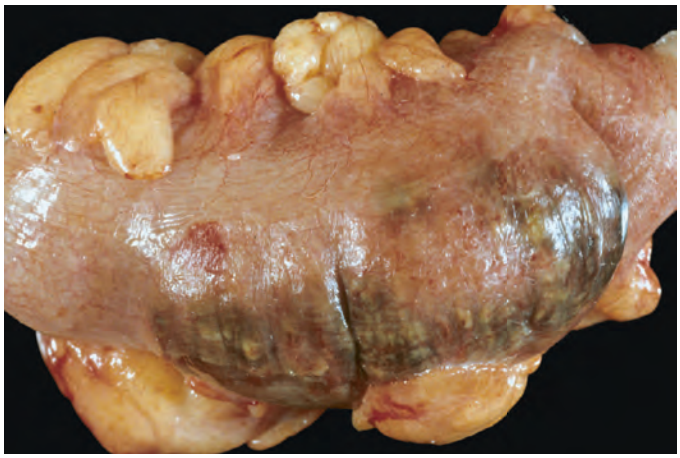


Figure 17.24 Intestinal obstruction. Portion of bowel incarcerated within an inguinal hernia. Note dusky areas of serosa and associated hemorrhage that indicate ischemic damage.

development of adhesions between bowel segments, the abdominal wall, or operative sites. These fibrous bridges can create closed loops through which other viscera may slide and become entrapped, resulting in internal herniation. Sequelae, including obstruction and strangulation, are similar to external hernias.

Volvulus

Volvulus occurs when a loop of bowel twists about its mesenteric point of attachment and results in both luminal and vascular compromise. Thus, volvulus presents with features of both obstruction and infarction. It occurs most often in large redundant loops of sigmoid colon, followed in frequency by the cecum, small bowel, stomach, or, rarely, transverse colon. Because it is rare, volvulus can be overlooked clinically, often with catastrophic results.

Intussusception

Intussusception occurs when a segment of the intestine, constricted by a wave of peristalsis, telescopes into the immediately distal segment. Once trapped, the invaginated segment is propelled by peristalsis and pulls the mesentery along. Untreated, intussusception may progress to intestinal obstruction, compression of mesenteric vessels, and infarction. **Intussusception is the most common cause of intestinal obstruction in children younger than 2 years of age.** In idiopathic cases there is no underlying anatomic defect, and the patient is otherwise healthy. Other cases have been associated with viral infection and rotavirus vaccines, perhaps due to reactive hyperplasia of Peyer patches and other mucosa-associated lymphoid tissues that can act as the leading edge. Intussusception is uncommon in older children and adults and is generally caused by an intraluminal mass or tumor that serves as the initiating point of traction. Approximately 1% of patients with cystic fibrosis develop intussusception; here it is believed that inspissated stool serves as a point of traction. Contrast enemas can be used both diagnostically and therapeutically for idiopathic intussusception in infants and young children, in whom air enemas may also be used to reduce the intussusception. Surgical intervention is necessary when a mass is present.

ISCHEMIC BOWEL DISEASE

The majority of the GI tract is supplied by the celiac, superior mesenteric, and inferior mesenteric arteries. As they approach the intestinal wall the superior and inferior mesenteric arteries ramify into the mesenteric arcades. Interconnections between arcades, as well as collateral vessels from the proximal celiac and distal pudendal and iliac circulations, make it possible for the small intestine and colon to tolerate slowly progressive loss of blood supply from one artery. In contrast to chronic, progressive hypoperfusion, acute compromise of any major vessel can lead to infarction of several meters of intestine. Damage can range from mucosal infarction extending no deeper than the muscularis mucosae to mural infarction of mucosa and submucosa to transmural infarction involving all three wall layers. While mucosal or mural infarctions can follow acute or chronic hypoperfusion, transmural infarction is typically due to acute vascular obstruction. In a large majority of cases, acute obstruction is caused by thrombosis or embolism. The most important risk factor for thrombosis is severe atherosclerosis (which is often prominent at the origin of mesenteric vessels). Less common causes of thrombosis include systemic vasculitides (Chapter 11). Obstructive emboli most commonly originate from aortic atheromas or cardiac mural thrombi. Mesenteric venous thrombosis, which can also lead to ischemic disease, is uncommon but can result from inherited or acquired hypercoagulable states, invasive neoplasms, cirrhosis, trauma, or abdominal masses that compress the portal drainage. Intestinal hypoperfusion may also occur in the absence of vascular obstruction in the setting of cardiac failure, shock, dehydration, or use of vasoconstrictive drugs.

Pathogenesis

Intestinal responses to ischemia occur in two phases. The initial hypoxic injury occurs at the onset of vascular compromise, but the greatest damage occurs during the second phase, reperfusion injury, which is initiated by restoration of the blood supply. While the underlying mechanisms of reperfusion injury are incompletely understood, they include leakage of gut lumen bacterial products (e.g., lipopolysaccharide) into the systemic circulation, free radical production, neutrophil infiltration, and release of additional inflammatory mediators (Chapter 2).

The severity of vascular compromise, the time frame during which it develops, and the vessels affected are the major variables that determine the severity of ischemic bowel disease. Intestinal vascular anatomy also contributes to the distribution of ischemic damage in two ways:

- *Intestinal segments at the end of their respective arterial supplies are particularly susceptible to ischemia.* These *watershed zones* include the splenic flexure where the superior and inferior mesenteric arterial circulations terminate and, to a lesser extent, the sigmoid colon and rectum, where inferior mesenteric, pudendal, and iliac arterial circulations end. Generalized hypotension or hypoxemia can cause localized injury at watershed zones, and ischemic disease should be considered in the differential diagnosis of focal colitis of the splenic flexure or rectosigmoid colon.
- *Intestinal capillaries run alongside the glands, from crypt to surface (villus), before making a hairpin turn and descending to the post-capillary venules. This arrangement makes the surface epithelium particularly vulnerable to ischemic injury,* relative to the crypts. This pattern of circulation protects the epithelial stem cells, which are located within the crypts and are necessary for recovery from epithelial injury. Surface epithelial atrophy with normal or hyperproliferative crypts is therefore a morphologic signature of ischemic intestinal disease.

MORPHOLOGY

Although the colon is the most common site of GI ischemia, mucosal and mural infarction may involve any level of the gut from stomach to anus. The lesions can be continuous but are most often segmental and patchy (Fig. 17.25A). The mucosa is hemorrhagic and often ulcerated (Fig. 17.25B). Edema may thicken

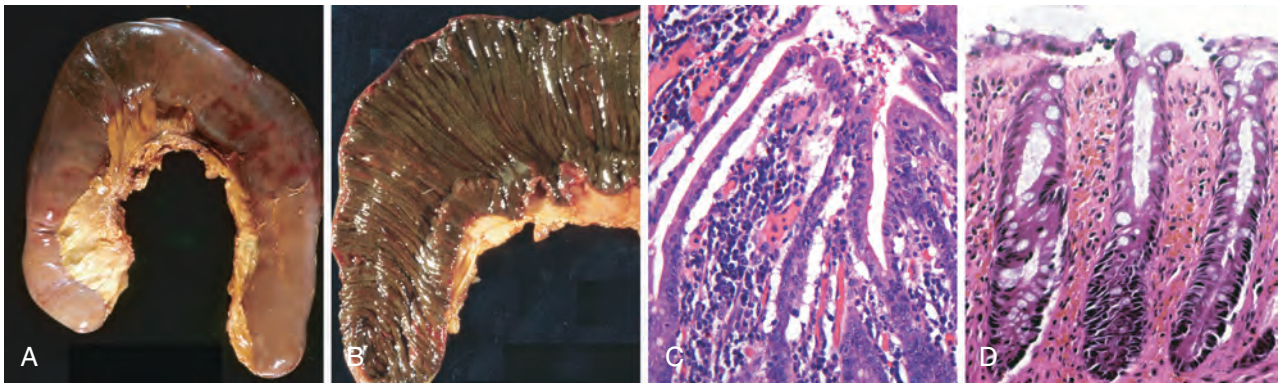


Figure 17.25 Ischemic bowel disease. (A) Jejunal resection with dusky serosa of acute ischemia (mesenteric thrombosis). (B) Mucosa is dark colored because of hemorrhage. (C) Characteristic attenuated villous epithelium in this case of acute mesenteric thrombosis. (D) Chronic colonic ischemia with atrophic surface epithelium and fibrotic lamina propria.

the bowel wall, particularly when damage extends into the submucosa and muscularis propria.

Substantial portions of the bowel are generally involved in **transmural infarction caused by acute arterial obstruction.** The demarcation between normal and ischemic bowel is sharply defined, and initially the infarcted bowel is intensely congested and dusky to purple-red. Later, blood-tinged mucus or frank blood accumulates in the lumen, and the wall becomes edematous, thickened, and rubbery. There is coagulative necrosis of the muscularis propria within 1 to 4 days, and perforation may occur. Serositis, with purulent exudates and fibrin deposition, may be prominent.

In **mesenteric venous thrombosis,** arterial blood continues to flow for a time, resulting in a less abrupt transition from affected to normal bowel. Propagation of the thrombus may lead to secondary involvement of the splanchnic bed. The ultimate result is similar to acute arterial obstruction because impaired venous drainage eventually prevents oxygenated arterial blood from entering the capillaries.

Microscopic examination of ischemic intestine demonstrates the characteristic atrophy or sloughing of surface epithelium (Fig. 17.25C), often with hyperproliferative crypts. Inflammatory infiltrates are initially absent, but neutrophils are recruited within hours of reperfusion. Chronic ischemia is accompanied by fibrous scarring of the lamina propria (Fig. 17.25D) and, uncommonly, stricture formation. In both acute and chronic ischemia, bacterial superinfection and enterotoxin release may induce **pseudomembrane formation** that resembles *Clostridioides* (formerly *Clostridium*) *difficile*–associated pseudomembranous colitis (discussed later).

Clinical Features

Ischemic disease of the colon is most common in patients older than 70 years of age and occurs slightly more often in women. **Acute colonic ischemia typically presents with sudden onset of cramping, left lower abdominal pain, a desire to defecate, and passage of blood or bloody diarrhea.** Patients may progress to shock and vascular collapse within hours in severe cases. Surgical intervention is necessary in approximately 10% of cases and should be considered if evidence of infarction (e.g., diminished peristaltic sounds, guarding, or rebound tenderness) develops. Because these physical signs overlap with those of other abdominal emergencies (e.g., acute appendicitis, perforated ulcer, and acute cholecystitis), the diagnosis of intestinal infarction may be delayed or missed, with disastrous consequences.

With appropriate management, mortality in the first 30 days is approximately 10%. Mortality is doubled in patients with right-sided colonic disease, who have a more severe course in general. This may be because the right side of the colon is supplied by the superior mesenteric artery, which also supplies much of the small intestine. Thus, right-sided colonic ischemia may be the initial presentation of more widespread compromise of intestinal perfusion. Other poor prognostic indicators include coexisting chronic obstructive pulmonary disease and persistence of symptoms for more than 2 weeks. Ischemic bowel injury can take a variety of subacute or chronic forms.

- *Limited zones of mucosal and mural infarctions* may progress to more extensive and transmural infarction if the vascular supply is not restored by correction of the insult or, in chronic disease, by development of adequate collateral supplies. The diagnosis of nonocclusive ischemic enteritis and colitis can be particularly difficult, as symptoms including intermittent bloody diarrhea and intestinal obstruction are often nonspecific.
- *Chronic ischemia* may masquerade as inflammatory bowel disease, with episodes of bloody diarrhea interspersed with periods of healing.
- *CMV infection* causes ischemic GI disease due to viral tropism for endothelial cells and resulting localized vascular obstruction. CMV infection, which can be a complication of immunosuppressive therapy, is discussed further in Chapter 8.
- *Radiation enterocolitis* occurs when the GI tract is irradiated and is due to a combination of epithelial and endothelial injury. Beyond clinical history, the presence of highly atypical "radiation fibroblasts" within the stroma may provide an important clue to the etiology. Acute radiation enteritis manifests as anorexia, abdominal cramps, and malabsorptive diarrhea, while chronic radiation enteritis or colitis is often more indolent and may present as an inflammatory enterocolitis.
- *Necrotizing enterocolitis* is an acute disorder of the small and large intestines that can result in transmural necrosis. It is the most common acquired GI emergency of neonates, particularly those who are premature or of low birth

weight, and frequently presents when oral feeding is initiated. It is discussed in more detail in Chapter 10 but is noted here because ischemic injury is thought to contribute to the pathogenesis.

ANGIODYSPLASIA

Angiodysplasia is characterized by malformed submucosal and mucosal blood vessels that are dilated and thin walled. They are most common in the cecum or right colon and, usually, come to clinical attention after the sixth decade of life. Although the prevalence is less than 1% in adults, angiodysplasia accounts for 20% of major episodes of lower intestinal bleeding in older populations. Presentation can range from chronic, intermittent to acute, massive hemorrhage.

The pathogenesis of angiodysplasia remains undefined but has been attributed to mechanical and congenital factors. Normal distention and contraction of the gut may intermittently occlude the submucosal veins that penetrate through the muscularis propria, leading to focal dilation and tortuosity of overlying submucosal and mucosal vessels. Because the cecum has the largest diameter of any colonic segment, it develops the greatest wall tension. This may explain the preferential distribution of angiodysplastic lesions in the cecum and right colon.

Morphologically, angiodysplastic lesions are characterized by ectatic nests of tortuous veins, venules, and capillaries. The vascular channels are often separated from the intestinal lumen by only the thin vessel wall and a layer of attenuated epithelial cells, making the vessels vulnerable to disruption by even minor injuries.

MALABSORPTION AND DIARRHEA

Malabsorption is characterized by defective absorption of fats, fat- and water-soluble vitamins, proteins, carbohydrates, electrolytes and minerals, and water. A hallmark of malabsorption is steatorrhea, characterized by excessive fecal fat and bulky, frothy, greasy, yellow or clay-colored stools. The chronic malabsorptive disorders most commonly encountered in the United States are pancreatic insufficiency, celiac disease, and Crohn disease (Table 17.6). Intestinal graft-versus-host disease is an important cause of malabsorption and diarrhea after allogeneic hematopoietic stem cell transplantation.

Malabsorption results from disturbance in at least one of the four phases of nutrient absorption:

- *Intraluminal digestion,* in which proteins, carbohydrates, and fats are broken down into forms suitable for absorption
- *Terminal digestion,* involving hydrolysis of carbohydrates and peptides by disaccharidases and peptidases in the brush border of the small intestinal mucosa
- *Transepithelial transport,* in which nutrients, fluid, and electrolytes are transported across and processed within the small intestinal epithelium
- *Lymphatic transport* of absorbed lipids

In many malabsorptive disorders a defect in one of these processes predominates, but several usually contribute. As a result, the symptoms and conseqences of malabsorption syndromes resemble each other more than they differ. Chronic malabsorption can lead to diarrhea, flatus, abdominal pain, weight loss, and muscle wasting, and is often associated with anorexia, abdominal distention, and borborygmi. Inadequate absorption of vitamins and minerals can result in anemia and mucositis due to pyridoxine, folate, or vitamin B_{12} deficiency; bleeding due to vitamin K deficiency; osteopenia and tetany due to calcium, magnesium, or vitamin D deficiencies; or peripheral neuropathy due to vitamin A or B_{12} deficiencies. A variety of endocrine and skin disturbances (e.g., goiter and vitiligo) may also develop as a consequence of nutrient, micronutrient, and vitamin deficiencies.

Table 17.6 Defects in Malabsorptive and Diarrheal Disease

Disease	Intraluminal Digestion	Terminal Digestion	Transepithelial Transport	Lymphatic Transport
Celiac disease		+	+	
Environmental enteropathy		+	+	
Chronic pancreatitis	+			
Cystic fibrosis	+			
Primary bile acid malabsorption	+		+	
Carcinoid syndrome			+	
Autoimmune enteropathy		+	+	
Disaccharidase deficiency		+		
Whipple disease				+
Abetalipoproteinemia			+	
Viral gastroenteritis		+	+	
Bacterial gastroenteritis		+	+	
Parasitic gastroenteritis		+	+	
Inflammatory bowel disease	+	+	+	

"+" indicates that the process is abnormal in the disease indicated.

Diarrhea is defined as an increase in stool mass, frequency, or fluidity, typically greater than 200 g per day. Both nutrient malabsorption and increased secretion of fluid by the intestine may contribute to diarrhea. In severe cases stool volume can exceed 14 L per day and, without fluid resuscitation, result in death. Painful, bloody, small-volume diarrhea is known as dysentery. Diarrhea can be classified into four major categories:

- *Secretory diarrhea* is characterized by isotonic stool and persists during fasting.
- *Osmotic diarrhea,* such as occurs with lactase deficiency, is due to the excessive osmotic force exerted by unabsorbed luminal solutes. The diarrhea fluid is more than 50 mOsm more concentrated than plasma, and diarrhea abates with fasting.
- *Malabsorptive diarrhea* follows generalized failure of nutrient absorption, is associated with steatorrhea, and is relieved by fasting.
- *Exudative diarrhea* due to inflammatory disease is characterized by purulent, often bloody stools that continue during fasting.

Cystic Fibrosis

Cystic fibrosis affects many organ systems, primarily the lungs, and is discussed in greater detail elsewhere (Chapter 10). Only the malabsorption associated with cystic fibrosis is considered here. Due to the absence of the epithelial cystic fibrosis transmembrane conductance regulator (CFTR), individuals with cystic fibrosis have defects in chloride and, in certain tissues, bicarbonate ion transport across epithelia. This interferes with bicarbonate, sodium, and water secretion, ultimately resulting in decreased hydration of luminal contents. The thick viscous stool may occasionally lead to intestinal obstruction, but more commonly results in formation of pancreatic intraductal concretions. The latter can begin in utero and result in duct obstruction, low-grade chronic pancreatitis with autodigestion of the pancreas, and eventually exocrine pancreatic insufficiency in more than 80% of patients. This in turn impairs the intraluminal phase of nutrient absorption, a defect that can be effectively treated by oral enzyme supplementation.

Celiac Disease

Celiac disease, also known as celiac sprue or gluten-sensitive enteropathy, is an immune-mediated disorder triggered by the ingestion of gluten-containing foods such as wheat, rye, or barley in genetically predisposed individuals. Celiac disease has an overall worldwide incidence of 0.6% to 1%, but its prevalence varies widely among countries and regions. Some of these differences correlate with variation in wheat consumption. While previously uncommon in many countries, the incidence of celiac disease is growing, possibly as a result of adoption of Western diets.

Pathogenesis

The alcohol-soluble fraction of gluten, gliadin, contains most of the disease-producing components. Gluten is digested by luminal and brush-border enzymes into amino acids and peptides, including a 33-amino acid α-gliadin peptide that is resistant to degradation by gastric, pancreatic, and small intestinal proteases (Fig. 17.26). Some gliadin peptides may induce epithelial cells to express IL-15, which in turn triggers activation and proliferation of CD8+ intraepithelial lymphocytes. These lymphocytes express NKG2D, a natural killer cell marker and receptor for MIC-A. Enterocytes that have been induced to express surface MIC-A in response

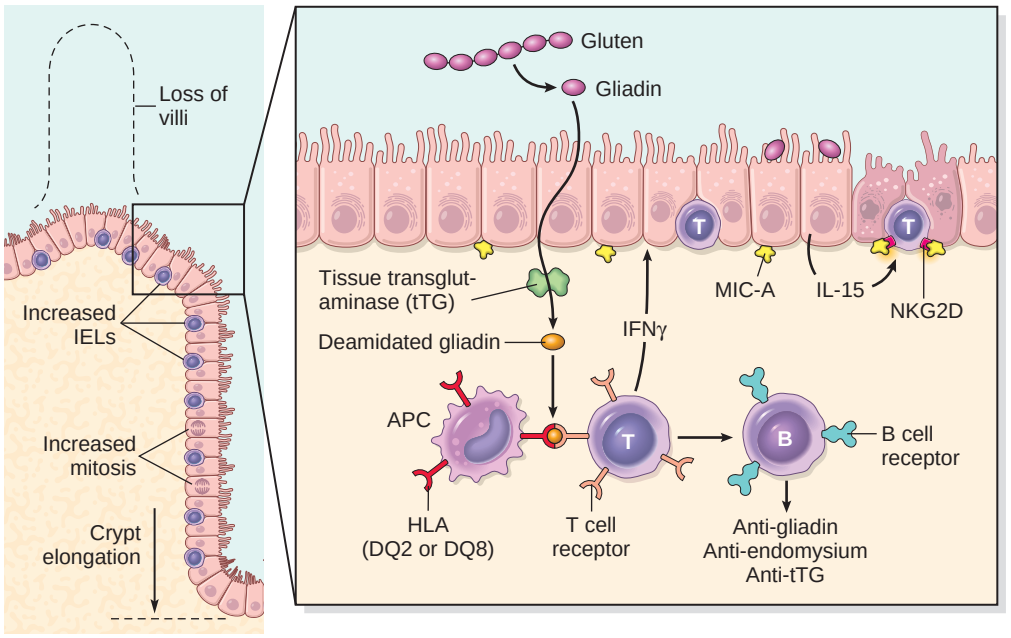


Figure 17.26 The *left panel* illustrates the morphologic alterations that may be present in celiac disease including villous atrophy, increased numbers of intraepithelial lymphocytes *(IELs)*, and epithelial proliferation with crypt elongation. The *right panel* depicts a model for the pathogenesis of celiac disease. Note that both innate (CD8+ intraepithelial T cells, activated by IL-15) and adaptive (CD4+ T cells and B cells sensitization to gliadin) immune mechanisms are involved in the tissue responses to gliadin.

to stress are then attacked by NKG2D-expressing intraepithelial lymphocytes. The resulting epithelial damage may enhance passage of other gliadin peptides into the lamina propria where they are deamidated by tissue transglutaminase. These gliadin peptides interact with HLA-DQ2 or HLA-DQ8 on antigen-presenting cells and in turn stimulate CD4+ T cells to produce cytokines that exacerbate tissue damage.

While most people eat grain and are exposed to gluten and gliadin, very few develop celiac disease. Thus, host factors determine whether disease develops. Among these, HLA proteins seem to be critical, since almost all people with celiac disease carry the class II HLA-DQ2 or HLA-DQ8 allele. However, the HLA locus accounts for less than half of the genetic component of celiac disease. Remaining genetic factors may include polymorphisms of genes involved in immune regulation and epithelial function. These genetic variables may also contribute to associations between celiac disease and other immune diseases including type 1 diabetes, thyroiditis, Sjögren syndrome, and IgA nephropathy.

MORPHOLOGY

Biopsy specimens from the second portion of the duodenum or proximal jejunum, which are exposed to the highest concentrations of dietary gluten, are generally diagnostic in celiac disease. The histopathology is characterized by increased numbers of intraepithelial CD8+ T lymphocytes (intraepithelial lymphocytosis), crypt hyperplasia, and villous atrophy (Fig. 17.27). This loss of mucosal and brush-border surface area contributes to malabsorption. Increased rates of epithelial turnover, reflected in increased crypt mitotic activity, may also limit the ability of absorptive enterocytes to fully differentiate and express proteins necessary for terminal digestion and transepithelial transport. Other features of fully developed celiac disease include increased numbers of plasma cells, mast cells, and eosinophils, especially within the upper part of the lamina propria. With increased serologic screening and early detection of disease-associated antibodies, it is now appreciated that increased numbers of intraepithelial lymphocytes, particularly within the villus, is a sensitive marker of celiac disease, even in the absence of epithelial damage and villous atrophy. However, intraepithelial lymphocytosis and villous atrophy are not specific for celiac disease and can be present in other diseases, including viral enteritis. The combination of histology and serology therefore is most specific for diagnosis of celiac disease.

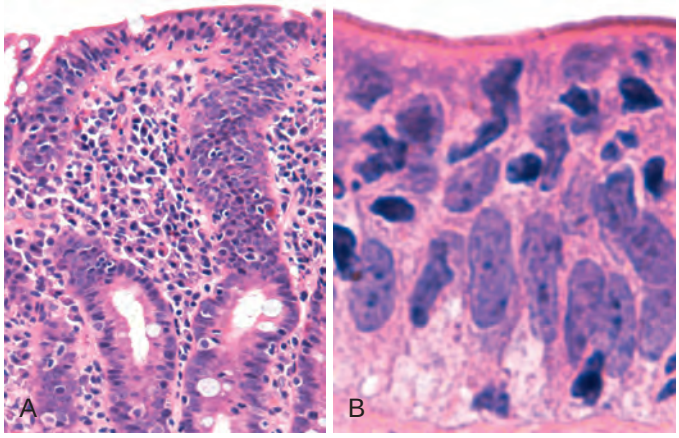


Figure 17.27 Celiac disease. (A) Advanced cases of celiac disease show complete loss of villi, or total villous atrophy. Note the dense plasma cell infiltrates in the lamina propria. (B) Infiltration of the surface epithelium by T lymphocytes, which can be recognized by their small, densely stained nuclei relative to larger, pale-staining epithelial nuclei.

Clinical Features

In adults, celiac disease presents most commonly between the ages of 30 and 60. Many cases of celiac disease escape clinical attention for extended periods because of atypical presentations. Other patients may have silent celiac disease, defined as positive serology and villous atrophy without symptoms, or latent celiac disease, in which positive serology is not accompanied by villous atrophy. Celiac disease may be associated with chronic diarrhea, bloating, or chronic fatigue, but can also be asymptomatic. These cases may present with anemia due to chronic iron and vitamin malabsorption. In adults, celiac disease is detected twice as frequently in women, perhaps because menstrual bleeding accentuates the effects of impaired iron absorption. A characteristic itchy, blistering skin lesion, dermatitis herpetiformis (Chapter 25), can be present in as many as 10% of patients.

Pediatric celiac disease, which affects males and females equally, may present with malabsorption or atypical symptoms affecting almost any organ. In those with classic symptoms, disease typically begins after introduction of gluten to the diet, between 6 and 24 months of age, and manifests as irritability, abdominal distention, anorexia, chronic diarrhea, failure to thrive, weight loss, or muscle wasting. Children with nonclassic symptoms tend to present at older ages with complaints of abdominal pain, nausea, vomiting, bloating, or constipation. Common extraintestinal complaints include arthritis or joint pain, aphthous stomatitis, iron deficiency anemia, delayed puberty, and short stature.

Unfortunately, the only treatment currently available for celiac disease is a gluten-free diet. Adherence to a gluten-free diet typically results in resolution of symptoms, reduced titers of anti–tissue transglutaminase or other celiac disease–associated antibodies, and restoration of normal or near-normal mucosal histology within 6 to 24 months. Therapeutic efforts are targeted at development of agents that minimize the impact of small quantities of ingested gluten. A gluten-free diet may also reduce the risk of long-term complications including anemia, female infertility, osteoporosis, and cancer (discussed below).

Noninvasive serologic tests are generally performed prior to biopsy. The most sensitive test is the measurement of IgA antibodies against tissue transglutaminase. IgA anti-endomysial antibodies can also be present. IgG anti–tissue transglutaminase antibodies may be detected in patients with IgA deficiency. The absence of HLA-DQ2 and HLA-DQ8 is useful for its negative predictive value, but the presence of these alleles is not helpful in confirming the diagnosis.

Individuals with celiac disease have an increased risk of malignancy. The most common cancer is enteropathy-associated T-cell lymphoma, an aggressive lymphoma of intraepithelial T lymphocytes. Small intestinal adenocarcinoma is also more frequent in individuals with celiac disease. Thus, when symptoms such as abdominal pain, diarrhea, and

weight loss develop despite a strict gluten-free diet, cancer or refractory sprue, in which the response to a gluten-free diet is lost, must be considered.

Environmental Enteric Dysfunction

Environmental enteric dysfunction, which has also been referred to as *environmental enteropathy, tropical enteropathy, or tropical sprue,* is a disorder prevalent in areas and populations with poor sanitation and hygiene, including severely under-resourced regions in parts of sub-Saharan Africa, such as Zambia; aboriginal populations in northern Australia; and some impoverished groups in South America and Asia. Affected individuals often suffer from malabsorption, malnutrition, and stunted growth. The relatively high oral vaccine failure rates in regions where environmental enteric dysfunction is endemic have been proposed to be due to defective mucosal immune function in environmental enteric dysfunction.

There are presently no accepted clinical, laboratory, or histopathologic criteria that allow diagnosis of environmental enteric dysfunction. Intestinal biopsy specimens have been examined in a small number of cases, and reported histologic features include more pronounced immune cell infiltration but less advanced villous atrophy relative to advanced celiac disease.

The underlying causes of environmental enteric dysfunction are unknown. Defective intestinal barrier and transport functions secondary to chronic exposure to fecal pathogens and other microbial contaminants, mucosal immune abnormalities, and repeated bouts of diarrhea within the first 2 or 3 years of life are likely involved. Although many pathogens are endemic in affected communities, no single infectious agent has been linked to environmental enteric dysfunction. Moreover, oral antibiotic treatment and nutritional supplementation do not reverse or prevent environmental enteric dysfunction. Irreversible failures in physical development, e.g., height and weight, may be accompanied by uncorrectable cognitive deficits. Thus, the global impact of environmental enteric dysfunction, which is estimated to affect more than 150 million children worldwide and may contribute to a large number of childhood deaths, is difficult to overstate.

Autoimmune Enteropathy

Autoimmune enteropathy is an X-linked disorder characterized by severe persistent diarrhea and autoimmune disease that occurs most often in young children. A particularly severe familial form, termed IPEX, an acronym denoting immune dysregulation, polyendocrinopathy, enteropathy, and X-linkage, is caused by germline loss of function mutations in the *FOXP3* gene, which is located on the X chromosome. FOXP3 is a transcription factor expressed in CD4+ regulatory T cells, and individuals with *FOXP3* mutations have defective development and function of these cells. Other defects in regulatory T-cell function have been linked to less severe forms of autoimmune enteropathy. Autoantibodies to enterocytes and goblet cells are common, and some patients have antibodies to parietal or islet cells. Within the small intestine, intraepithelial lymphocytes may be increased but not to the extent seen in celiac disease. In contrast to celiac disease, neutrophils are often seen infiltrating the intestinal mucosa. Therapy includes immunosuppressive drugs such as cyclosporine and hematopoietic stem cell transplantation.

Lactase (Disaccharidase) Deficiency

The disaccharidases, including lactase, are located in the apical brush-border membrane of the villus absorptive epithelial cells. In the absence of lactase, dietary lactose cannot be broken into glucose and galactose. Lactose cannot be absorbed and remains in the lumen where it exerts an osmotic force that attracts fluid and causes diarrhea. Because the defect is biochemical, histology is generally unremarkable. Lactase deficiency is of two types:

- *Congenital lactase deficiency,* caused by a mutation in the gene encoding lactase, is an autosomal recessive disorder. The disease is rare and presents as explosive diarrhea with watery, frothy stools and abdominal distention upon milk ingestion. Symptoms abate when exposure to milk and milk products is terminated. Congenital lactase deficiency was often fatal prior to the availability of soy-based infant formula.
- *Acquired lactase deficiency* is caused by downregulation of lactase gene expression and is particularly common among Native American, African American, and Chinese populations. Acquired lactase deficiency can develop with increased age, potentially because, until relatively recently (in evolutionary terms), people did not ingest milk after weaning. Lactase deficiency can also present following enteric viral or bacterial infections, where it may resolve over time. Symptoms, including abdominal fullness and diarrhea, follow ingestion of lactose-containing dairy products. Flatulence also occurs due to fermentation of the unabsorbed sugars by colonic bacteria.

Microvillus Inclusion Disease

Microvillus inclusion disease, sometimes referred to as Davidson disease, is an autosomal recessive disorder of vesicular transport that leads to deficient brush-border assembly. It is caused by mutations in the *MYO5B* gene, which encodes a motor protein that is required for delivery and retrieval of plasma membrane components and for normal nutrient, ion, and water transport. The disease occurs most commonly in European, Middle Eastern, and Navajo Native American populations. Severe, intractable diarrhea develops before 3 months of age (often in the first few days of life) as a result of defective terminal digestion and transepithelial transport functions. The factors responsible for early or later presentation have not been defined. The defect in plasma membrane trafficking leads to the accumulation of abnormal apical vesicles containing microvilli and various membrane components. The abnormal vesicles can be identified by electron microscopy or by immunostaining for the brush border protein villin. CD10 immunohistochemistry, which labels the microvillus inclusions as well as other cytoplasmic structures, is also used for diagnosis. Total parenteral nutrition and small bowel transplantation are the only treatments available for microvillus inclusion disease.

Abetalipoproteinemia

Abetalipoproteinemia is a rare autosomal recessive disease characterized by an inability to assemble triglyceride-rich lipoproteins. It is caused by a mutation in microsomal triglyceride transfer protein (MTP), which is required for transfer of lipids to nascent apolipoprotein B polypeptide in the endoplasmic reticulum. Without MTP, lipids accumulate intracellularly. Triglyceride accumulation is evident as epithelial vacuolization and can be highlighted by special stains, such as oil red-O, particularly after a fatty meal.

Abetalipoproteinemia presents in infancy with failure to thrive, diarrhea, and steatorrhea. Because of the defect in lipoprotein assembly, the plasma is completely devoid of lipoproteins containing apolipoprotein B. Failure to absorb essential fatty acids leads to deficiencies of fat-soluble vitamins as well as lipid membrane defects that can be recognized by the presence of acanthocytes, red cells with spiky membrane protrusions, in peripheral blood smears.

KEY CONCEPTS

CONGENITAL AND ACQUIRED (NONINFECTIOUS) DISORDERS OF THE INTESTINES

- Abdominal hernias may occur through any weakness or defect in the wall of the peritoneal cavity including inguinal and femoral canals, the umbilicus, and sites of surgical scars.
- Intussusception occurs when a segment of intestine telescopes into the immediately distal segment. It is the most common idiopathic cause of intestinal obstruction in children younger than 2 years of age, but is driven by leading edge mass lesions in older individuals.
- Ischemic bowel disease of the colon is most common at the splenic flexure, sigmoid colon, and rectum; these are watershed zones where two arterial circulations terminate.
- Angiodysplasia is a malformation of submucosal and mucosal blood vessels and a common cause of lower intestinal bleeding in those older than 60 years of age.
- Diarrhea can be characterized as secretory, osmotic, malabsorptive, or exudative.
- The malabsorption associated with cystic fibrosis is the result of pancreatic insufficiency, leading to inadequate pancreatic digestive enzymes, and deficient luminal breakdown of nutrients.
- Celiac disease is an immune-mediated enteropathy triggered by the ingestion of gluten-containing grains. The malabsorptive diarrhea in celiac disease is due to loss of brush-border surface area including villous atrophy and, possibly, deficient enterocyte maturation as a result of immune-mediated epithelial damage.
- Environmental enteric dysfunction is prevalent in areas with poor sanitation. It is estimated to affect more than 150 million children worldwide and may contribute to a very large number of childhood deaths.
- Autoimmune enteropathy is characterized by persistent diarrhea associated with other forms of autoimmune disease. A particularly severe form is IPEX, caused by mutation in the *FOXP3* gene, which is necessary for regulatory T-cell development and function.
- Lactase deficiency causes an osmotic diarrhea due to the inability to break down lactose. The autosomal recessive form is rare and severe; the acquired form is common and usually presents in adulthood.
- Microvillus inclusion disease is a rare autosomal recessive disorder due to mutations in *MYO5B* that impair vesicular transport and brush-border assembly.

INFECTIOUS ENTEROCOLITIS

Enterocolitis can present with a broad range of symptoms including diarrhea, abdominal pain, urgency, perianal discomfort, incontinence, and hemorrhage (Table 17.7). This global problem is responsible for over 1 million deaths each year. Half of the mortality occurs in children under the age of 5, in whom diarrheal disease is the fourth most common cause of death worldwide. The specific pathogens vary widely as a function of age, nutrition, and host immune status as well as geography and other environmental influences such as the availability of clean drinking water. Bacterial infections, such as enterotoxigenic *E. coli* and *Salmonella* spp., are frequently responsible for acute diarrheal illnesses. Rotavirus and adenovirus are common causes of death in children under 5 years of age (see Table 17.7). Mycobacterial infections of the GI tract are considered in detail in Chapter 8.

Cholera

***Vibrio cholerae* are comma-shaped, gram-negative bacteria that cause cholera, a disease that has been endemic in the Ganges Valley of India and Bangladesh throughout history.** Since 1817 seven great pandemics have spread along trade routes to Europe, Australia, and the Americas, after each of which cholera retreated back to the Ganges Valley. Cholera also persists within the Gulf of Mexico but causes only rare cases of seafood-associated disease. This occurs because shell fish and plankton can be reservoirs of *Vibrio cholerae*. The incidence of cholera increases during summer, reflecting more rapid growth of *Vibrio* bacteria at warm temperatures. It is primarily transmitted by drinking contaminated water and may cause epidemics in areas where disasters such as earthquakes or war have destroyed sewage systems with resultant fecal contamination of drinking water supplies. For example, the 2010 Haitian earthquake led to a cholera epidemic that affected more than 5% of the population. Over half of cases required hospitalization, and 1% were fatal.

Pathogenesis

Severe diarrhea is caused by the toxin released by the bacteria. *Vibrio* organisms are noninvasive and remain within the intestinal lumen. Proteins involved in motility and attachment are necessary for efficient colonization. Hemagglutinin, a metalloproteinase, is important for bacterial detachment and shedding into the stool, which is key to dissemination.

Cholera toxin is encoded by a virulence phage. It is composed of five B subunits and a single A subunit. The B subunits bind GM1 ganglioside on the surface of intestinal epithelial cells, allowing the toxin to be carried by endocytosis to the endoplasmic reticulum (Fig. 17.28). Here, the A subunit

Table 17.7 Features of Bacterial Enterocolitides

Infection Type	Geography	Reservoir	Transmission	Epidemiology	Affected GI Sites	Symptoms	Complications
Cholera	India, Africa	Shellfish	Fecal-oral, water	Sporadic, endemic, epidemic	Small intestine	Severe watery diarrhea	Dehydration, electrolyte imbalances
Campylobacter spp.	High income countries	Chickens, sheep, pigs, cattle	Poultry, milk, other foods	Sporadic; children, travelers	Colon	Watery or bloody diarrhea	Reactive arthritis, Guillain-Barré syndrome
Shigellosis	Worldwide, endemic in under-resourced countries	Humans	Fecal-oral, food, water	Children, migrant workers, travelers, nursing home residents	Left colon, ileum	Bloody diarrhea	Reactive arthritis, urethritis, conjunctivitis, hemolytic-uremic syndrome
Salmonellosis	Worldwide	Poultry, farm animals, reptiles	Meat, poultry, eggs, milk	Children, older adults	Colon and small intestine	Watery or bloody diarrhea	Sepsis, abscess
Enteric (typhoid) fever	India, Mexico, Philippines	Humans	Fecal-oral, water	Children, adolescents, travelers	Small intestine	Bloody diarrhea, fever	Chronic infection, carrier state, encephalopathy, myocarditis, intestinal perforation
Yersinia spp.	Northern and central Europe	Pigs, cows, puppies, cats	Pork, milk, water	Clustered cases	Ileum, appendix, right colon	Abdominal pain, fever, diarrhea	Reactive arthritis, erythema nodosum
Escherichia coli							
ETEC	Under-resourced countries	Unknown	Food or fecal-oral	Infants, adolescents, travelers	Small intestine	Severe watery diarrhea	Dehydration, electrolyte imbalances
EPEC	Worldwide	Humans	Fecal-oral	Infants	Small intestine	Watery diarrhea	Dehydration, electrolyte imbalances
EHEC	Worldwide	Widespread, includes cattle	Beef, milk, produce	Sporadic and epidemic	Colon	Bloody diarrhea	Hemolytic-uremic syndrome
EIEC	Under-resourced countries	Unknown	Cheese, other foods, water	Young children	Colon	Bloody diarrhea	Unknown
EAEC	Worldwide	Unknown	Unknown	Children, adults, travelers	Colon	Nonbloody diarrhea, afebrile	Poorly defined
Pseudomembranous colitis (*C. difficile*)	Worldwide	Humans, hospitals	Antibiotics allow emergence	Immunosuppressed, antibiotic-treated	Colon	Watery diarrhea, fever	Relapse, toxic megacolon
Whipple disease	Rural > urban	Unknown	Unknown	Rare	Small intestine	Malabsorption	Arthritis, CNS disease
Mycobacterial infection	Worldwide	Unknown	Unknown	Immunosuppressed, endemic	Small intestine	Malabsorption	Pneumonia, infection at other sites

CNS, Central nervous system; *EAEC,* enteroaggregative *E. coli; EIEC,* enteroinvasive *E. coli; EHEC,* enterohemorrhagic *E. coli; EPEC,* enteropathogenic *E. coli; ETEC,* enterotoxigenic *E. coli; GI,* gastrointestinal.

of the toxin is reduced by protein disulfide isomerase, and a fragment of the A subunit is unfolded. This peptide fragment is then transported into the cytosol using host cell machinery that normally moves misfolded proteins from the endoplasmic reticulum to the cytosol. Once in the cytosol, the A subunit fragment refolds and then interacts with cytosolic ADP ribosylation factors (ARFs) to activate the stimulatory G protein $G_s\alpha$. This stimulates adenylate cyclase and the resulting increase in intracellular cyclic adenosine monophosphate (cAMP) opens the cystic fibrosis conductance regulator (CFTR), which releases chloride ions into the lumen. Chloride and sodium absorption are also inhibited by cAMP. The resulting accumulation of chloride, bicarbonate, and sodium within the intestinal lumen creates an osmotic force

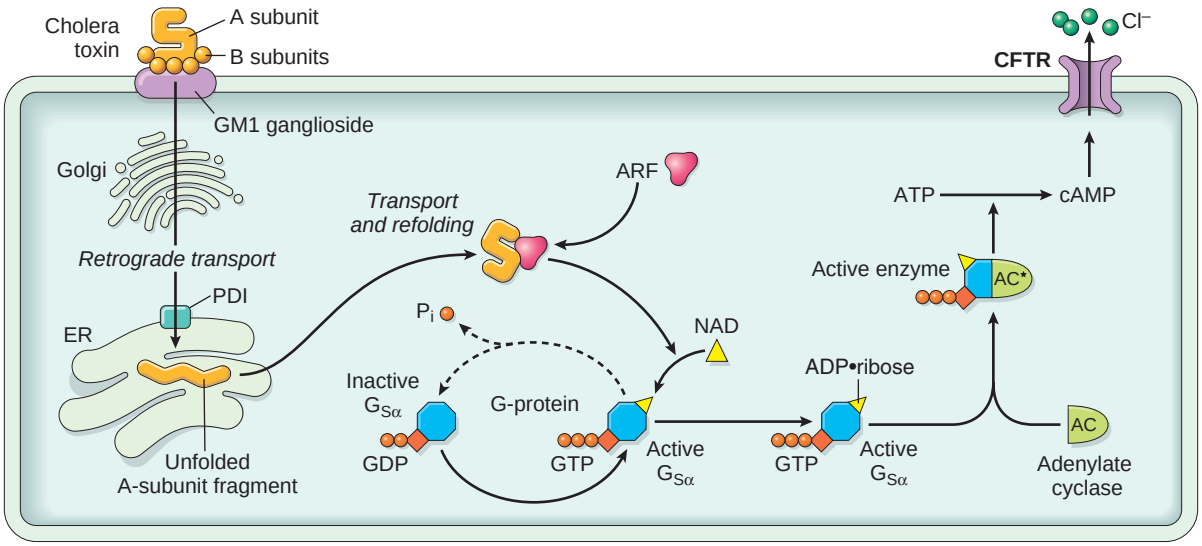


Figure 17.28 Cholera toxin transport and signaling. After retrograde toxin transport to the endoplasmic reticulum *(ER)*, the A subunit is released by the action of protein disulfide isomerase *(PDI)* and is then able to access the epithelial cell cytoplasm. In concert with an ADP-ribosylation factor *(ARF)*, the A subunit then ADP-ribosylates $G_s\alpha$, which locks it in the active, GTP-bound state. This leads to adenylate cyclase *(AC)* activation, and the cyclic adenosine monophosphate *(cAMP)* produced opens cystic fibrosis transmembrane conductance regulator *(CFTR)* to drive chloride secretion and diarrhea.

that draws water into the lumen and causes massive diarrhea. Remarkably, mucosal biopsies show only minimal histologic alterations.

Clinical Features

Most individuals exposed to *V. cholerae* are asymptomatic or develop only mild diarrhea. In those with severe disease there is an abrupt onset of watery diarrhea and vomiting following an incubation period of 1 to 5 days. The voluminous stools resemble rice water and are sometimes described as having a fishy odor. The rate of diarrhea may reach 1 L per hour, leading to dehydration, hypotension, muscular cramping, anuria, shock, loss of consciousness, and death. Mortality for severe cholera is about 50% without treatment, but timely fluid replacement can save more than 99% of patients. Oral rehydration is often sufficient but is unfortunately underutilized. Promising investigational therapies for acute therapy include CFTR inhibitors, which prevent diarrhea by blocking chloride secretion. Vaccination of at-risk individuals during cholera outbreaks can limit disease spread and is recommended, along with other prevention control strategies, by the World Health Organization (WHO). Prophylactic vaccination is a long-term goal, and some data indicate that this approach will convey sufficient herd immunity to reduce the incidence of cholera.

Campylobacter Enterocolitis

***Campylobacter jejuni* is the most common bacterial enteric pathogen in high income countries and is an important cause of traveler's diarrhea and food poisoning.** Infections are most often associated with ingestion of improperly cooked chicken, but outbreaks can also be caused by unpasteurized milk or contaminated water.

Pathogenesis

The pathogenesis of *Campylobacter* infection remains poorly defined, but the four major properties that contribute to virulence are motility, adherence, toxin production, and invasion. Flagella allow *Campylobacter* to be motile, which facilitates adherence and colonization. Some *C. jejuni* isolates also release cytotoxins that cause direct epithelial damage or a cholera toxin–like enterotoxin. Dysentery, i.e., bloody diarrhea, is generally associated with bacterial invasion and is caused by only a minority of *Campylobacter* strains. Enteric fever occurs when bacteria proliferate within the lamina propria and mesenteric lymph nodes.

***Campylobacter* infection can result in reactive arthritis, primarily in patients with HLA-B27 genotype.** Other extraintestinal complications, including erythema nodosum and Guillain-Barré syndrome, a flaccid paralysis caused by immunologically mediated inflammation of peripheral nerves (Chapter 27), are not HLA-linked. Molecular mimicry has been implicated in the pathogenesis of Guillain-Barré syndrome, as serum antibodies to *C. jejuni* lipopolysaccharide cross-react with peripheral and central nervous system gangliosides. The risk of developing Guillain-Barré syndrome following *Campylobacter* infection is estimated to be less than 0.1%, but up to 40% of Guillain-Barré cases have documented *Campylobacter* infection in the 2 weeks prior to presentation, and up to 50% have positive stool cultures or circulating antibodies to *Campylobacter*.

MORPHOLOGY

Campylobacter are comma-shaped, flagellated, gram-negative organisms. Diagnosis is primarily by stool culture, since biopsy findings are nonspecific. Mucosal and intraepithelial neutrophil infiltrates are prominent, particularly within the superficial mucosa (Fig. 17.29A); cryptitis (neutrophil infiltration of the crypt epithelium) and crypt abscesses (accumulations of neutrophils within crypt lumens) may also be present. Importantly, crypt architecture is preserved (Fig. 17.29D), although this can be difficult to assess in cases with severe mucosal damage.

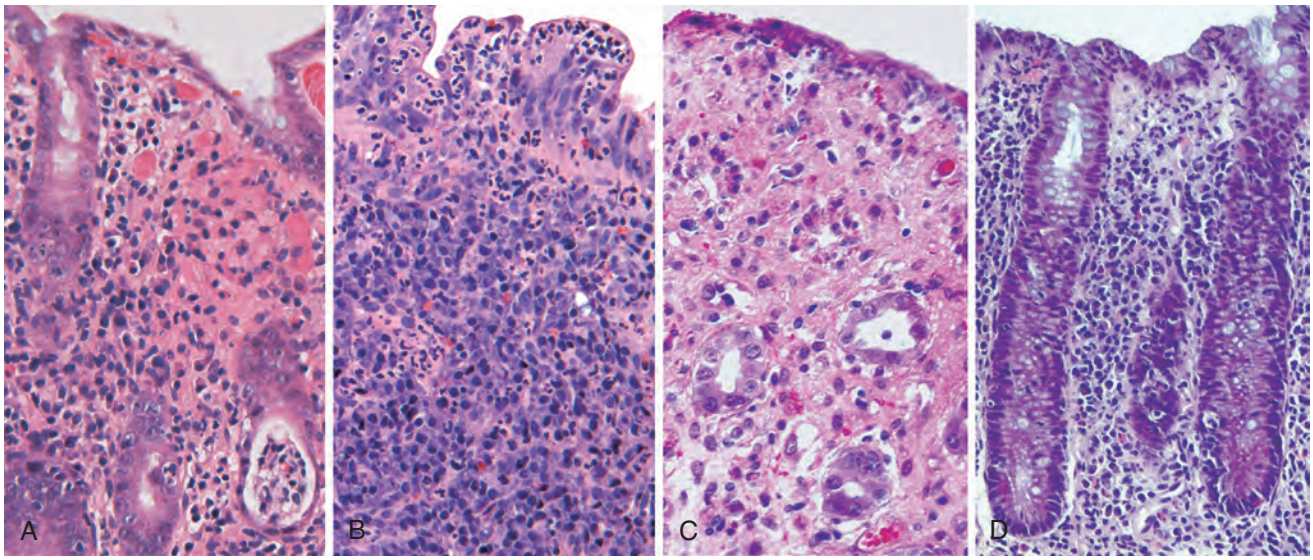


Figure 17.29 Bacterial enterocolitis. (A) *Campylobacter jejuni* infection produces acute, self-limited colitis. Neutrophils can be seen within the surface and crypt epithelia, and a crypt abscess is present at the lower right. (B) In *Yersinia* infection the surface epithelium can be eroded by neutrophils, and the lamina propria is densely infiltrated by sheets of plasma cells admixed with lymphocytes and neutrophils. (C) Enterohemorrhagic *Escherichia coli* O157:H7 results in an ischemia-like morphology with surface atrophy and erosion. (D) Enteroinvasive *E. coli* infection is similar to other acute, self-limited colitides such as those caused by *C. jejuni*. Note the maintenance of normal crypt architecture and spacing despite abundant intraepithelial neutrophils.

Clinical Features

Ingestion of as few as 500 *C. jejuni* organisms can cause disease after an incubation period of up to 8 days. Watery diarrhea, either acute or following an influenza-like prodrome, is the primary symptom, but dysentery develops in 15% of adults and more than 50% of children. Patients may shed bacteria for 1 month or more after clinical resolution. Antibiotic therapy is generally not required.

Shigellosis

***Shigella* are gram-negative unencapsulated, nonmotile, facultative anaerobes that belong to the *Enterobacteriaceae* family and are closely related to enteroinvasive *E. coli*.** Although humans are the only known reservoir, *Shigella* spp. remain one of the most common causes of bloody diarrhea, with a worldwide incidence of up to 165 million cases each year. Given the extremely low infective dose of several hundred organisms and the presence of as many as 10^9 organisms in each gram of stool during acute disease, *Shigella* are highly transmissible by the fecal-oral route or via contaminated water and food.

In the United States and Europe, children in daycare centers, migrant workers, travelers to low resource countries, and individuals in nursing homes are most commonly affected. Deaths are generally limited to children younger than 5 years of age. *Shigella* is endemic in countries with poor sanitation, where it is responsible for 10% of pediatric diarrheal disease and up to 75% of diarrheal deaths.

Pathogenesis

Shigella are resistant to gastric acid, thereby explaining the low infective dose. Once in the intestine, organisms are taken up by M, or microfold, cells, which are specialized for luminal antigen sampling and presentation. The bacteria proliferate within the M cells and then escape into the lamina propria, where they are phagocytosed by and induce apoptosis of macrophages. The inflammatory response damages surface epithelium and allows *Shigella* within the intestinal lumen as well as those within the lamina propria to gain access to the basolateral membranes of colonic epithelial cells, across which bacteria invade the cytoplasm more efficiently than from the apical surface. All *Shigella* spp. carry virulence plasmids, some of which encode a type III secretion system capable of directly injecting bacterial proteins into the host cytoplasm. *Shigella dysenteriae* serotype 1 also release the Shiga toxin Stx, which inhibits eukaryotic protein synthesis, resulting in host cell damage and death.

MORPHOLOGY

Shigella infections are most prominent in the left colon, but the ileum may also be involved, perhaps reflecting the abundance of M cells in the **dome epithelium** overlying Peyer patches. The mucosa is hemorrhagic and ulcerated, and pseudomembranes may be present. The histology of early cases is similar to other acute self-limited colitides such as *Campylobacter* colitis, but **because of the tropism for M cells, aphthous ulcers similar to those seen in Crohn disease may occur.** This and other factors lead to significant potential for confusion with inflammatory bowel disease.

Clinical Features

After an incubation period of up to 1 week, *Shigella* causes self-limited disease characterized by 7 to 10 days of diarrhea, fever, and abdominal pain. The initially watery diarrhea progresses to a dysenteric phase in approximately 50%

of patients, and constitutional symptoms can persist for as long as 1 month. Duration of symptoms is typically shorter in children, but severity is often much greater. In adults, an uncommon subacute presentation includes weeks of waxing and waning diarrhea that can mimic new-onset ulcerative colitis. Confirmation of *Shigella* infection requires stool culture. Antibiotic treatment shortens both the clinical course and the duration of bacterial shedding in stools. Conversely, antidiarrheal medications can prolong symptoms and delay *Shigella* clearance and are therefore contraindicated.

Long-term complications of *Shigella* infection are uncommon but include a triad of sterile reactive arthritis, urethritis, and conjunctivitis that preferentially affects HLA-B27–positive men between 20 and 40 years of age. *S. dysenteriae* serotype 1 that secrete Shiga toxin may also trigger hemolytic-uremic syndrome, although enterohemorrhagic *E. coli* (EHEC) are more commonly responsible (Chapter 20). Toxic megacolon and intestinal obstruction are uncommon complications.

Salmonella

***Salmonella*, which are classified within the Enterobacteriaceae family of gram-negative bacilli, are divided into *S. typhi*, the causative agent of typhoid fever (discussed in the next section), and nontyphoid *Salmonella* such as *S. enteritidis*.** The latter are the most common cause of salmonellosis, of which more than 1 million cases occur each year in the United States; the prevalence is even greater in under-resourced countries. Infection is most common in young children and older adults, with peak incidence in the summer and fall. Salmonella is transmitted via contaminated food, particularly raw or undercooked meat, poultry, eggs, and milk. Centralized food processing can lead to large outbreaks. Vaccines are available for both humans and farm animals, e.g., egg-laying hens.

Pathogenesis

Very few viable *Salmonella* are necessary to cause infection; the absence of gastric acid, in individuals with atrophic gastritis or those on acid-suppressive therapy, further reduces the required inoculum. *Salmonella* possess virulence genes that encode a type III secretion system capable of transferring bacterial proteins into M cells and enterocytes. The transferred proteins activate host Rho guanosine triphosphatases (GTPases), thereby triggering actin rearrangement and bacterial endocytosis which, in turn, facilitates bacterial growth within endosomes. *Salmonella* also secrete a molecule that induces epithelial cell release of the eicosanoid hepoxilin A3, a potent chemoattractant that stimulates neutrophil chemotaxis into the intestinal lumen and potentiates mucosal damage. Flagellin, the core protein of bacterial flagella, and bacterial lipopolysaccharide may activate TLR5 and TLR4, respectively, further augmenting inflammation and tissue damage. Both Th1 and Th17 T cells contribute to pathogen clearance, explaining why those with genetic defects in Th17 immunity are at risk for disseminated salmonellosis.

The gross and microscopic features of *Salmonella* enteritis are nonspecific and are similar to the acute self-limited colitis of *Campylobacter* and *Shigella*. Stool cultures are essential for diagnosis.

Clinical Features

Salmonella infections are clinically indistinguishable from those caused by other enteric pathogens. Symptoms range from loose stools to cholera-like profuse diarrhea to dysentery. Fever often resolves within 2 days, but diarrhea can persist for a week, and organisms can be shed in the stool for several weeks after resolution. Antibiotic therapy is not generally recommended because it can prolong the carrier state or even cause relapse and does not typically shorten the duration of diarrhea. Although *Salmonella* infections are typically self-limited, some complications, including reactive arthritis and meningitis, and even death may occur, particularly in patients with malignancies, immunosuppression, alcoholism, cardiovascular dysfunction, sickle cell disease, or hemolytic anemia.

Typhoid Fever

Typhoid fever, also referred to as enteric fever, has an annual worldwide incidence of 30 million individuals. The disease is caused by *Salmonella enterica* and its two subtypes, *typhi* and *paratyphi*. The majority of cases in endemic countries are due to *S. typhi,* while infection by *S. paratyphi* is more common among travelers, perhaps because they are often vaccinated against *S. typhi.* In endemic areas, children and adolescents are affected most often, but there is no age preference in non-endemic countries. Infection is strongly associated with travel to India, Mexico, the Philippines, Pakistan, El Salvador, and Haiti. Humans are the sole reservoir for *S. typhi* and *S. paratyphi;* transmission occurs from person to person or via food or contaminated water. Gallbladder colonization with *S. typhi* or *S. paratyphi* may be associated with gallstones and a chronic carrier state.

Pathogenesis

In contrast to *S. enteritidis*, *S. typhi* disseminate via lymphatic and blood vessels, which causes systemic reactive hyperplasia of phagocytes and lymphoid tissues. Similar to *Shigella, S. typhi* are resistant to gastric acid and, initially, invade via small intestinal M cells.

MORPHOLOGY

Salmonella infection causes **Peyer patches in the terminal ileum** to enlarge into sharply delineated, plateau-like elevations up to 8 cm in diameter. Draining mesenteric lymph nodes are also enlarged. Neutrophils accumulate within the superficial lamina propria together with macrophages containing bacteria, red cells, nuclear debris, lymphocytes, and plasma cells. Mucosal damage creates oval ulcers, oriented along the axis of the ileum, that may perforate.

In disseminated *S. typhi* infection, the spleen is enlarged and soft, with uniformly pale red pulp, obliterated follicular markings, and prominent phagocyte hyperplasia. The liver is punctuated by small, randomly scattered foci of parenchymal necrosis in which hepatocytes are replaced by macrophage aggregates, called **typhoid nodules,** which may also be found in the bone marrow and lymph nodes.

Clinical Features

Acute infection results in anorexia, abdominal pain, bloating, nausea, vomiting, and bloody diarrhea followed by a short asymptomatic phase that gives way to bacteremia and fever with flu-like symptoms. Abdominal tenderness may mimic appendicitis. Rose spots, small erythematous maculopapular lesions, develop on the chest and abdomen. Symptoms abate after several weeks but relapse can occur. **Systemic dissemination may cause extraintestinal complications including encephalopathy, meningitis, seizures, endocarditis, myocarditis, pneumonia, and cholecystitis.** Patients with sickle cell disease are particularly susceptible to *Salmonella* osteomyelitis. Blood cultures are positive in more than 90% of affected individuals during the febrile phase. Antibiotic treatment can prevent disease progression.

Yersinia

Three *Yersinia* species are human pathogens. *Y. enterocolitica* and *Y. pseudotuberculosis* cause GI disease and are discussed here; *Y. pestis,* the agent of pulmonic and bubonic plague, is discussed in Chapter 8. The incidence of GI *Yersinia* infections in Europe is greater than in North America, perhaps as a result of differing dietary practices. Infections are most frequently linked to ingestion of pork, raw milk, and contaminated water and tend to cluster in the winter.

Pathogenesis

Yersinia invade M cells and use specialized bacterial proteins, called adhesins, to bind to host cell β_1 integrins. A pathogenicity island encodes proteins that mediate iron capture and transport; similar iron transport systems are also present in *E. coli, Klebsiella, Salmonella,* and enterobacteria. In *Yersinia,* **iron enhances virulence and stimulates systemic dissemination, explaining why individuals with increased non-heme iron, such as those with certain chronic forms of anemia or hemochromatosis, are at greater risk for sepsis and death.**

MORPHOLOGY

Yersinia infections preferentially involve the ileum, appendix, and right colon (see Fig. 17.29B). The organisms proliferate extracellularly in lymphoid tissue, resulting in regional lymph node and Peyer patch hyperplasia as well as bowel wall thickening. The mucosa overlying lymphoid tissue may become hemorrhagic, and aphthous erosions and ulcers may appear along with neutrophil infiltrates (Fig. 17.29B) and granulomas. As with *Shigella,* these factors can cause diagnostic confusion with Crohn disease.

Clinical Features

Yersinia infection generally presents as abdominal pain, nausea, vomiting, and abdominal tenderness. **Extraintestinal symptoms of pharyngitis, arthralgia, and erythema nodosum occur frequently.** Fever and diarrhea are less common. Peyer patch invasion with subsequent involvement of regional lymphatics can mimic acute appendicitis in teenagers and young adults, but enteritis and colitis predominate in younger children. *Yersinia* can be detected by stool culture on *Yersinia*-selective agar. In cases with extraintestinal disease, lymph nodes or blood cultures may also be positive. Postinfectious complications include reactive arthritis with urethritis and conjunctivitis, myocarditis, erythema nodosum, and kidney disease.

Escherichia coli

E. coli are gram-negative bacilli that colonize the healthy GI tract; most are nonpathogenic, but a subset cause human disease. The latter are classified according to morphology, pathogenesis, and in vitro behavior. Subgroups with major clinical relevance include enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), and enteroaggregative *E. coli* (EAEC).

***Enterotoxigenic* Escherichia coli. ETEC are the principal cause of traveler's diarrhea and spread via contaminated food or water.** In under-resourced countries, children younger than 2 years of age are particularly susceptible. ETEC are noninvasive but produce heat-labile (LT) and heat-stable (ST) toxins. LT is similar to cholera toxin and activates adenylate cyclase, resulting in increased intracellular cAMP and chloride secretion. ST binds to guanylate cyclase, and increases intracellular cyclic guanosine monophosphate (cGMP); the effects on epithelial ion transport are similar to those induced by LT. Like cholera, the secretory, non-inflammatory diarrhea induced by ETEC results in only mild histologic changes, despite having the potential to cause dehydration and, in severe cases, shock.

***Enteropathogenic* Escherichia coli. EPEC are prevalent worldwide and are an important cause of endemic diarrhea** as well as diarrheal outbreaks, particularly in children less than 2 years of age. EPEC are characterized by their ability to produce attaching and effacing (A/E) lesions in which bacteria attach tightly to small intestinal enterocyte apical membranes and cause local loss, i.e., effacement, of the microvilli. The proteins necessary for creating A/E lesions are encoded by a large genomic pathogenicity island, the locus of enterocyte effacement (LEE), which is also present in many EHEC strains. These proteins include Tir, which is inserted into the intestinal epithelial cell plasma membrane. Tir acts as a receptor for the bacterial outer membrane protein intimin, which is encoded by the *espE* gene and can be used for molecular detection and diagnosis of EPEC infection. The locus of enterocyte effacement also encodes a type III secretion system, similar to that in *Shigella,* that injects bacterial effector proteins into the epithelial cell cytoplasm. EPEC strains do not produce Shiga toxins.

***Enterohemorrhagic* Escherichia coli. EHEC are categorized as *E. coli* O157:H7 and non-O157:H7 serotypes.** *E. coli* O157:H7 outbreaks are often caused by consumption of inadequately cooked ground beef, reflecting the natural reservoir in cows. Contaminated milk and vegetables are also vehicles for infection. Both O157:H7 and non-O157:H7 serotypes produce Shiga-like toxins, and therefore lesions (see Fig. 17.29C) and clinical symptoms are similar to those resulting from *S. dysenteriae* infection. O157:H7 serotypes are more likely than non-O157:H7 serotypes to cause outbreaks of bloody diarrhea, hemolytic-uremic syndrome, and ischemic colitis. Importantly, antibiotics are not recommended

because killing the bacteria can lead to release of Shiga-like toxins and increased risk of hemolytic uremic syndrome, especially in children.

Enteroinvasive* Escherichia coli. EIEC organisms are bacteriologically similar to *Shigella and are transmitted via food or water or by person-to-person contact. While EIEC do not produce toxins, they invade epithelial cells and cause nonspecific features of acute self-limited colitis (see Fig. 17.29D). EIEC infections are most common among young children in under-resourced countries and are occasionally associated with outbreaks in more affluent regions.

***Enteroaggregative* Escherichia coli. EAEC has a unique "stacked brick" morphology when bound to epithelial cells.** These organisms cause diarrhea in children and adults worldwide, including traveler's diarrhea. The organisms attach to enterocytes via adherence fimbriae and are aided by dispersin, a bacterial surface protein that neutralizes the negative surface charge of lipopolysaccharide. EAEC organisms produce minimal histologic changes and cause nonbloody diarrhea that may be prolonged in individuals with immunodeficiency.

Pseudomembranous Colitis

Pseudomembranous colitis is generally caused by *Clostridioides* (formerly *Clostridium*) *difficile* and is also referred to as antibiotic-associated colitis or antibiotic-associated diarrhea. While other organisms such as *Salmonella, Clostridium perfringens* type A, or *Staphylococcus aureus* may also produce diarrhea in the context of antibiotic therapy, only *C. difficile* causes pseudomembranous colitis.

Pathogenesis

Disruption of the normal colonic microbiota by antibiotics allows *C. difficile* overgrowth. Almost any antibiotic may be responsible; the most important determinants of the disease are frequency of use and the effect on colonic microbiota. Toxins released by *C. difficile* cause the ribosylation of small GTPases such as Rho, leading to disruption of the epithelial cytoskeleton, tight junction barrier loss, cytokine release, and apoptosis. Toxin-mediated cytoskeletal disruption in other cell types, particularly neutrophils, may also contribute to development of purulent pseudomembranes, but the processes that produce pseudomembranous colitis remain incompletely understood.

MORPHOLOGY

The surface epithelium is denuded, and the superficial lamina propria contains a dense infiltrate of neutrophils and occasional fibrin thrombi within capillaries. **Pseudomembranes** (Fig. 17.30A, B), made up of an adherent layer of inflammatory cells and debris are not specific and may also be present in ischemia or necrotizing infections. However, **histopathology of *C. difficile*–associated colitis is pathognomonic.** Specific features include the mucopurulent exudate that characteristically erupts from damaged crypts to form "volcano" lesions (Fig. 17.30C) and coalesces to form pseudomembranes.

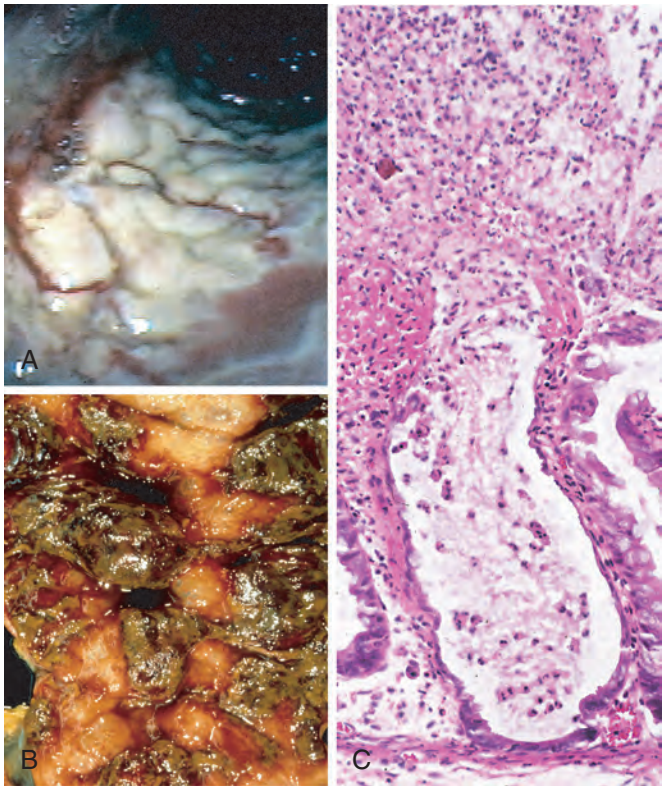


Figure 17.30 *Clostridioides* (formerly *Clostridium*) *difficile* colitis. (A) The colon is coated by tan pseudomembranes composed of neutrophils, dead epithelial cells, and inflammatory debris (endoscopic view). (B) Pseudomembranes are easily appreciated on gross examination. (C) Typical pattern of neutrophils emanating from a damaged crypt is reminiscent of a volcanic eruption.

Clinical Features

Risk factors for *C. difficile*–associated colitis include antibiotic treatment, advanced age, hospitalization, and immunosuppression. The organism is particularly prevalent in hospitals; as many as 30% of hospitalized adults are colonized with *C. difficile* (a rate 10-fold greater than the general population), but most colonized patients are free of disease. Individuals with *C. difficile*–associated colitis present with fever, leukocytosis, abdominal pain, cramps, watery diarrhea, and dehydration. Protein loss can give rise to hypoalbuminemia. Fecal leukocytes and occult blood may be present, but grossly bloody diarrhea is uncommon. Toxic magacolon, characterized by marked dilatation of the colon, is a potentially fatal complication resulting from marked injury to the colonic wall. The diagnosis is usually made by detection of *C. difficile* toxin, rather than by culture, and is supported by the characteristic histopathology. Metronidazole and vancomycin are generally effective therapies, but the prevalence of antibiotic-resistant and hypervirulent *C. difficile* strains is increasing. Recurrent *C. difficile*–associated colitis occurs in up to 40% of patients. Fecal microbial transplantation successfully prevents recurrent infection in some patients and is becoming an accepted therapy. Many investigators are working to develop defined microbial populations that can be administered in place of fecal preparations. New antibiotics as well as monoclonal antibodies against toxins A and B can also be effective, but are not yet in widespread use.

Whipple Disease

Whipple disease is a rare, multivisceral chronic disease first described as intestinal lipodystrophy in 1907 by George Hoyt Whipple, a pathologist who won the Nobel Prize for his work on pernicious anemia.

Pathogenesis

Whipple's original case report described an individual with malabsorption, lymphadenopathy, and arthritis of undefined origin. Postmortem examination demonstrated the presence of foamy macrophages filled with gram-positive microbes. The organism was identified as an actinomycete by PCR in 1992 and named *Tropheryma whipplei*. Clinical symptoms occur because bacteria-laden macrophages accumulate within the small intestinal lamina propria and mesenteric lymph nodes, causing lymphatic obstruction. Thus, **the malabsorptive diarrhea of Whipple disease is due to impaired lymph drainage.**

MORPHOLOGY

The morphologic hallmark of Whipple disease is a **dense accumulation of distended, foamy macrophages in the small intestinal lamina propria** (Fig. 17.31A). The macrophages contain periodic acid–Schiff (PAS)–positive, diastase-resistant granules that represent partially digested bacteria within lysosomes (Fig. 17.31B). Intact rod-shaped bacilli can also be identified by electron microscopy (Fig. 17.31C). A similar infiltrate of foamy macrophages is present in intestinal mycobacterial infections (Fig. 17.31D), and the organisms are PAS-positive in both diseases. The acid-fast stain can be helpful, since mycobacteria stain positively (Fig. 17.31E), while *T. whipplei* does not.

Villous expansion caused by dense macrophage infiltrates imparts a shaggy gross appearance to the mucosal surface. Lymphatic dilation and mucosal lipid deposition account for the endoscopically visible white to yellow mucosal plaques. In Whipple disease, bacteria-laden macrophages can accumulate within **mesenteric lymph nodes, synovial membranes, cardiac valves, brain,** and other sites.

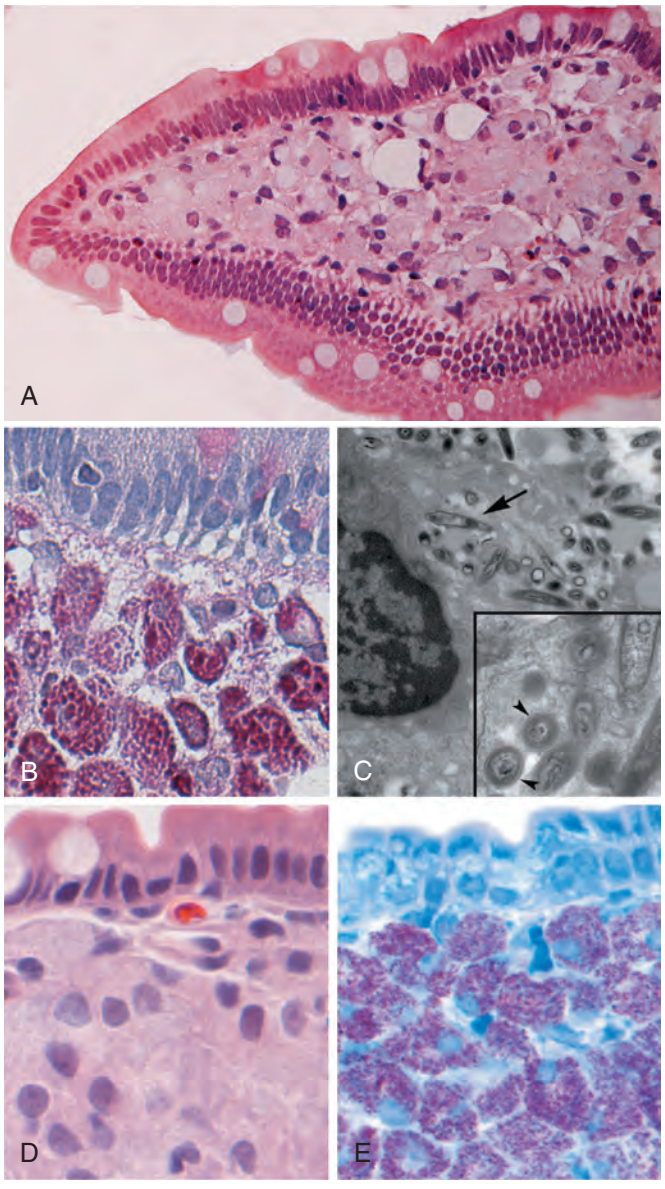


Figure 17.31 Whipple disease and mycobacterial infection. (A) Hematoxylin and eosin staining shows effacement of normal lamina propria by a sheet of distended macrophages. (B) Periodic acid–Schiff stain highlights macrophage lysosomes full of bacilli. (C) Electron micrograph of part of a macrophage shows bacilli within the cell *(top arrow);* also seen at higher magnification *(arrowheads in inset)*. (D) The morphology of mycobacterial infection can be similar to Whipple disease, particularly in the immunocompromised host. Compare with (A). (E) Unlike *T. whippelii,* mycobacteria are positive with stains for acid-fast bacteria. (C, Courtesy George Kasnic and Dr. William Clapp, University of Florida, Gainesville, Fla.)

Whipple disease is most common in Caucasian men, particularly farmers and others with occupational exposure to soil or animals. While there is no consistent familial clustering, the rarity of infection despite a large number of healthy carriers suggests that genetic risk factors exist.

The clinical presentation of Whipple disease is usually a triad of diarrhea, weight loss, and arthralgia. Extraintestinal symptoms may precede malabsorption by months or years and include arthritis; arthralgia; fever; lymphadenopathy; and neurologic, cardiac, or pulmonary disease.

Viral Gastroenteritis

Symptomatic human infection is caused by many viruses. The most common are discussed here.

Norovirus. This was previously known as Norwalk-like virus and is a common cause of nonbacterial gastroenteritis. Humans are the only known reservoir for these are small icosahedral viruses with a single-stranded RNA genome. **Approximately half of all gastroenteritis outbreaks worldwide are thought to be due to norovirus.** In the United States, noroviruses are the most common cause of acute gastroenteritis requiring medical attention. They cause more than 200,000 childhood deaths annually in under-resourced countries. With increasing use of the rotavirus vaccine, norovirus has become the most common cause of diarrhea worldwide.

Local norovirus outbreaks are usually initiated via contaminated food or water, but dissemination throughout a population occurs via person-to-person, fecal-oral transmission. Fecal-oral transmission is responsible for most sporadic cases. Infections spread easily within schools, hospitals, nursing homes, and other large groups in close quarters, such as on cruise ships. In these environments, vehicles include airborne droplets, environmental surfaces, and fomites. Some data suggest that strict enforcement of personal hygiene, enteric precautions, and regular environmental decontamination with alcohol-based hand sanitizers and quaternary ammonium wipes may reduce transmission. However, prevention is difficult to achieve due to the resistance of the virus to a wide range of temperatures, from freezing to 60°C, and the extraordinarily low infective dose of less than 20 particles.

Following an incubation period of up to 2 days, affected individuals develop vomiting; cramping abdominal pain; watery diarrhea; and nonspecific symptoms including headache, chills, and myalgias. The symptoms resolve within 2 to 3 days in most cases. Infectious viral particles can be shed prior to symptoms and for up to 8 weeks after infection in healthy, immunocompetent individuals.

Norovirus infection in immunocompromised patients is a significant problem. Some data suggest that nearly 20% of patients on immunosuppression after renal transplantation or as treatment for graft-versus-host disease after hematopoietic stem cell transplantation are infected with norovirus and have intermittent diarrhea. Many of these patients fail to clear the infection, suffer from chronic diarrhea, and shed infectious virus for up to 1 year. The malnutrition and dehydration that accompany infection can also increase morbidity of the underlying disease.

Some individuals are naturally resistant to norovirus infection due to mutations that inactivate galactoside 2-alpha-L-fucosyltransferase, *FUT2,* which contributes to the glycosylation of proteins destined for the cell surface in the Golgi complex. Approximately 20% of individuals of European and African descent are homozygous for these mutations. These individuals, termed non-secretors, neither shed blood group antigens in saliva nor express them on epithelial cells at many mucosal sites including the GI tract. Non-secretors are resistant to infection because they lack the carbohydrate groups on these antigens that serve as ligands for viral infection. In others, acquired immunity is difficult to achieve because there is a great diversity of norovirus strains, which limits cross protection, and because the viral genome regularly undergoes recombination that results in antigenic shift. As a result, a new pandemic strain emerges every 2 to 4 years.

MORPHOLOGY

Morphology is nonspecific but can include mild villous shortening, epithelial vacuolization, crypt hypertrophy, and lamina propria infiltration by neutrophils, lymphocytes, and monocytes (Fig. 17.32A). The microvillus brush border is also compromised, and reduced expression of digestive enzymes can lead to mild steatorrhea and carbohydrate malabsorption.

Rotavirus. **This encapsulated virus with a segmented, double-stranded RNA genome is highly prevalent and a significant cause of diarrheal deaths worldwide.** Children between 6 and 24 months of age are most vulnerable, probably because antibodies in breast milk confer protection during the first 6 months of life. Rotavirus vaccines are becoming widely used but are incompletely effective in under-resourced regions, possibly as a consequence of malnutrition. The WHO recommends that the first dose be given at 6 weeks of age.

Rotavirus outbreaks in hospitals and daycare centers are common, and infection spreads easily; the estimated minimal infective inoculum is only 10 viral particles. Rotavirus selectively infects and destroys mature enterocytes in the small intestine, and the villus surface is repopulated by immature epithelial cells. Enterocyte damage may be mediated by a viral factor called nonstructural protein 4 (NSP4), which can induce apoptosis. The loss of absorptive function and net secretion of water and electrolytes is compounded by an osmotic diarrhea caused by the incomplete absorption of nutrients. Like norovirus, rotavirus has a short incubation period followed by several days of vomiting and watery diarrhea.

Adenovirus. A common cause of pediatric diarrhea, adenovirus also affects immunocompromised patients. Small intestinal biopsy specimens can show epithelial degeneration but more often exhibit nonspecific villous atrophy and compensatory crypt hyperplasia. Viral nuclear inclusions are uncommon. Disease typically presents after an incubation period of 1 week with symptoms that include diarrhea, vomiting, and abdominal pain. Fever and weight loss may also be present. Symptoms generally resolve within 10 days.

Parasitic Enterocolitis

Although viruses and bacteria are the predominant enteric pathogens in the United States, parasitic disease and protozoal infections affect more than one-half of the world's population on a chronic or recurrent basis. The small intestine can provide a home for many as 20 species of parasites including nematodes, such as the roundworms *Ascaris* and *Strongyloides;* hookworms and pinworms; cestodes, including flatworms and tapeworms; trematodes, or flukes; and protozoa. Parasitic infections are covered in Chapter 8; those that are common in the intestinal tract are discussed briefly here.

Ascaris lumbricoides. This nematode infects 1 billion individuals worldwide. *Ascaris* eggs are shed in the feces of infected individuals, allowing spread via the direct fecal-oral route. Eggs may also be present in improperly cleaned and peeled fruits and vegetables that have grown in contaminated soil, explaining why *Ascaris*, *Strongyloides,* hookworm (*Necator duodenale* and *Ancylostoma duodenale*), and whipworm *(Trichuris trichiura)* are known as soil-transmitted helminths. Ingested eggs hatch in the intestine, and larvae penetrate the intestinal mucosa. Larvae then migrate from splanchnic to systemic circulations and finally enter the lungs to grow within the alveoli. Approximately 3 weeks later, the larvae are coughed up and swallowed.

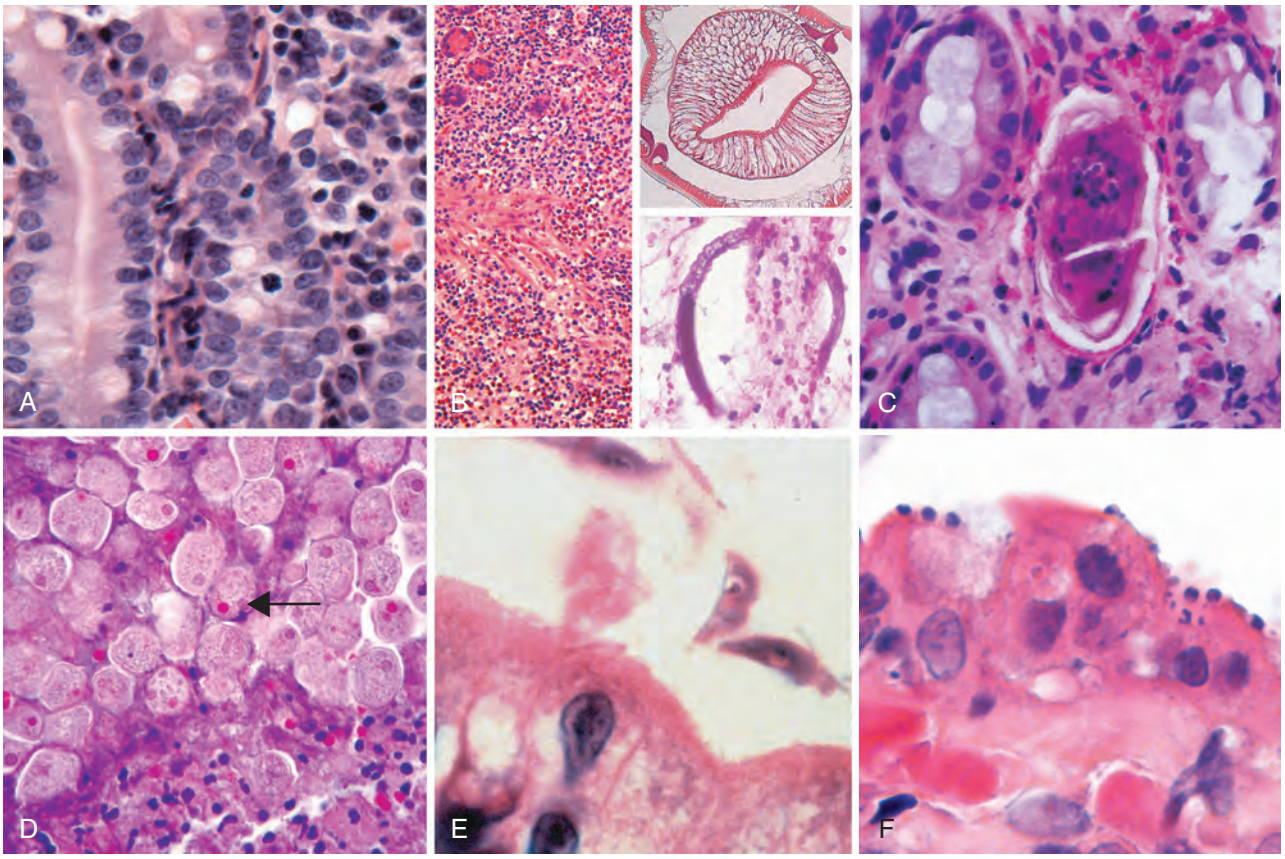


Figure 17.32 Infectious enteritis. (A) Histologic features of viral enteritis include increased numbers of intraepithelial and lamina propria lymphocytes and crypt hypertrophy. (B) Diffuse eosinophilic infiltrates in parasitic infection. This case was caused by *Ascaris (upper inset)*, but a similar tissue reaction could be caused by *Strongyloides (lower inset)*. (C) Schistosomiasis can induce an inflammatory reaction to eggs trapped within the lamina propria. (D) *Entamoeba histolytica* in a colon biopsy specimen. Note some organisms ingesting red blood cells *(arrow)*. (E) *Giardia lamblia*, which are present in the luminal space over nearly normal-appearing villi, are easily overlooked. (F) *Cryptosporidium* organisms are seen as small blue spheres that appear to lie on top of the brush border but are actually enveloped by a thin layer of host cell cytoplasm.

Upon return to the small intestine, the larvae mature into adult worms, which induce an eosinophil-rich inflammatory reaction (Fig. 17.32B) that can cause physical obstruction of the intestine or biliary tree. Larvae can also form hepatic abscesses and cause pneumonitis. Diagnosis is usually made by detection of eggs in stool samples.

Strongyloides. The larvae of *Strongyloides* live in fecally contaminated ground soil and can penetrate unbroken skin. They migrate through the lungs, where they induce inflammatory infiltrates, and then reside in the intestine while maturing into adult worms. Unlike other intestinal worms, which require an ova or larval stage outside the human, the eggs of *Strongyloides* can hatch within the intestine and release larvae that penetrate the mucosa, causing autoinfection (Fig. 17.32B). Hence *Strongyloides* infection can persist for life, and immunosuppressed individuals can develop overwhelming autoinfection. *Strongyloides* incite a strong tissue reaction and peripheral eosinophilia.

Necator duodenale *and* Ancylostoma duodenale. These hookworms infect 1 billion people worldwide and cause significant morbidity. Infection is initiated by larval penetration through the skin. After further development in the lungs, they migrate up the trachea and are swallowed. Worms attach to the duodenal mucosa, suck blood, and reproduce. This causes multiple superficial erosions, focal hemorrhage, inflammatory infiltrates, and, in chronic infection, iron deficiency anemia. Diagnosis can be made by detection of the eggs in fecal smears.

Trichuris trichiura. Whipworms primarily infect young children. Similar to *Enterobius vermicularis, T. trichiura* does not penetrate the intestinal mucosa and rarely causes serious disease. Heavy infections may cause bloody diarrhea and rectal prolapse.

Enterobius vermicularis *and* Enterobius gregorii. These organisms, also known as pinworms, infect people in both high income and low income countries; over 1 billion are infected worldwide. *E. vermicularis* is the more common species, but *Enterobius gregorii* have also been reported in Europe, Africa, and Asia. Because they do not invade host tissue and live their entire life within the intestinal lumen, pinworms rarely cause serious illness. Infection is primarily by the fecal-oral route. Adult worms living in the intestine migrate to the anal orifice at night, where the female deposits eggs on the perirectal mucosa. The eggs cause irritation and rectal and perineal pruritus. Scratching leads to contamination of the fingers, which promotes human-to-human transmission.

Both eggs and adult pinworms remain viable outside the body, and repeat infection is common. Diagnosis can be made by applying cellophane tape to the perianal skin and examining the tape under a microscope to identify eggs.

Schistosomiasis. This disease involving the intestines most commonly takes the form of adult worms residing within the mesenteric veins. Eggs are trapped within the mucosa and submucosa (Fig. 17.32C). The resulting immune reaction is often granulomatous and can cause bleeding and even obstruction. More details are presented in Chapter 8.

Intestinal Cestodes. The three primary species of cestodes that affect humans are *Diphyllobothrium latum,* fish tapeworms; *Taenia solium,* pork tapeworms; and *Hymenolepis nana,* dwarf tapeworms. They reside exclusively within the intestinal lumen and are transmitted by ingestion of raw or undercooked fish, meat, or pork that contains encysted larvae. Release of the larva allows attachment to the intestinal mucosa through its head, or scolex. The worm derives its nutrients from the food stream and enlarges by formation of egg-filled segments termed proglottids. Humans are usually infected by a single worm, and since the worm does not penetrate the intestinal mucosa, peripheral eosinophilia does not generally occur. Nevertheless, the parasite burden can be staggering. Adult worms can grow to many meters in length and shed large numbers of proglottids and eggs into the feces. Clinical symptoms include abdominal pain, diarrhea, and nausea, but most cases are asymptomatic. Occasionally, *D. latum* causes B_{12} deficiency and megaloblastic anemia because it competes with the host for dietary B_{12}. Identification of proglottids and eggs in stools is the most efficient method of diagnosis.

Entamoeba histolytica. This protozoan causes amebiasis and is spread by fecal-oral transmission. *E. histolytica* infects approximately 500 million people in countries such as India, Mexico, and Colombia and causes 40 million cases of dysentery and liver abscess annually. The cysts have four nuclei and a chitin wall that is resistant to gastric acid, a characteristic that allows them to pass through the stomach without harm. *E. histolytica* then colonize the epithelial surface of the colon and release trophozoites, ameboid forms that reproduce under anaerobic conditions.

Amebiasis affects the cecum and ascending colon most often, but sigmoid colon, rectum, and appendix can also be involved. Dysentery develops when the amebae attach to the colonic epithelium, induce apoptosis, invade crypts, and burrow laterally into the lamina propria. This recruits neutrophils, causes tissue damage, and creates a flask-shaped ulcer with a narrow neck and broad base. Histologic diagnosis can be difficult, since amebae can be similar to macrophages in size and appearance (Fig. 17.32D). Parasites penetrate splanchnic vessels and embolize to the liver to produce abscesses in about 40% of patients with amebic dysentery. Amebic liver abscesses can exceed 10 cm in diameter and have a scant inflammatory reaction at their margins with a shaggy fibrin lining. The abscesses persist after the acute intestinal illness has passed and may, rarely, reach the lung and the heart by direct extension. Amebae may also spread to the kidneys and brain via the bloodstream.

Individuals with amebiasis may present with abdominal pain, bloody diarrhea, or weight loss. Occasionally, acute necrotizing colitis and megacolon occur; both are associated with significant mortality. The parasites lack mitochondria or Krebs cycle enzymes and are thus obligate fermenters of glucose. Metronidazole, which inhibits pyruvate oxidoreductase, an enzyme required for fermentation, is the most effective treatment for systemic disease.

Giardia lamblia. These organisms, also referred to as *G. duodenalis* or *G. intestinalis,* were initially described by van Leeuwenhoek, the inventor of the microscope, who discovered the pathogen in his own stool. *G. lamblia* is the most common parasitic pathogen in humans and is spread by fecally contaminated water or food. Infection may occur after ingestion of as few as 10 cysts. Because cysts are resistant to chlorine, *Giardia* are endemic in unfiltered public water supplies. They are also common in rural streams, explaining infection in campers who use these as a water source. Infection may occur by the fecal-oral route, and because the cysts are stable, they may be accidentally swallowed while swimming in contaminated water.

Giardia are flagellated protozoans that cause decreased expression of brush-border enzymes, microvillous damage, and apoptosis of small intestinal epithelial cells. Secretory IgA and mucosal IL-6 responses are important for clearance of *Giardia* infections. Immunosuppressed, agammaglobulinemic, or malnourished individuals are often severely affected. *Giardia* can evade immune clearance through continuous modification of the major surface antigen, variant surface protein, and can persist for months or years, causing intermittent symptoms.

Giardia trophozoites can be identified in duodenal biopsies based on their characteristic pear shape and the presence of two equally sized nuclei. Despite large numbers of trophozoites, some of which are tightly bound to the brush border of villous enterocytes, there is no invasion, and small intestinal morphology may be normal (Fig. 17.32E); villous blunting with increased numbers of intraepithelial lymphocytes and mixed lamina propria inflammatory infiltrates can accompany heavy infections.

Giardiasis may be subclinical or accompanied by acute or chronic diarrhea, malabsorption, and weight loss. Infection is usually documented by immunofluorescent detection of cysts in stool samples. Oral antimicrobial therapy is effective, but recurrence is common.

Cryptosporidium. Like *Giardia,* cryptosporidia are an important cause of diarrhea worldwide. Cryptosporidiosis was first discovered in the 1980s as an agent of chronic diarrhea in acquired immunodeficiency syndrome (AIDS) patients and is now recognized as a cause of acute, self-limited disease in immunologically normal hosts. Cryptosporidiosis also causes persistent diarrhea in residents of under-resourced countries. The organisms are present worldwide, with the exception of Antarctica, perhaps, because the oocysts are killed by freezing. Oocysts are resistant to chlorine and may therefore persist in treated, but unfiltered, water. Contaminated drinking water continues to be the most common means of transmission. The largest documented outbreak, a result of inadequate water purification, occurred in 1993 in Milwaukee, Wisconsin, and affected

more than 400,000 people. Like giardiasis, cryptosporidiosis is readily spread to water sport participants. Food-borne infection occurs less frequently.

Humans are infected by several different *Cryptosporidium* species, including *C. hominis* and *C. parvum*. All are able to go through an entire life cycle, with asexual and sexual reproductive phases, in a single host. Once ingested, as few as 10 encysted oocytes are sufficient to cause symptomatic infection. Sporozoites are released following protease activation by gastric acid. They are motile and have a specialized organelle that attaches to the brush border and causes changes in the enterocyte cytoskeleton. These changes induce the enterocyte to engulf the parasite, which takes up residence in an endocytic vacuole. Sodium malabsorption, chloride secretion, and increased tight junction permeability cause the nonbloody, watery diarrhea that ensues.

Mucosal histology is often only minimally altered, but persistent cryptosporidiosis in children and heavy infection in immunosuppressed patients can result in villous atrophy, crypt hyperplasia, and inflammatory infiltrates. Although the sporozoite is intracellular, it appears, by light microscopy, to sit on top of the epithelial apical membrane (Fig. 17.32F). Organisms are typically most concentrated in the terminal ileum and proximal colon, but can be present throughout the gut, biliary tract, and even the respiratory tract of immunodeficient hosts. Diagnosis is based on finding oocysts in the stool.

KEY CONCEPTS

INFECTIOUS ENTEROCOLITIS

- *V. cholerae* secrete a preformed toxin that causes massive chloride secretion. Water follows the resulting osmotic gradient, leading to secretory diarrhea.
- *C. jejuni* is the most common bacterial enteric pathogen in high income countries and also causes traveler's diarrhea. Most isolates are noninvasive.
- *Salmonella* and *Shigella* spp. are invasive and associated with exudative bloody diarrhea (dysentery).
- *Salmonella enteritidis* infection is a common cause of food poisoning.
- *S. typhi* cause systemic disease (typhoid fever).
- Pseudomembranous colitis is often triggered by antibiotic therapy. The responsible organism, *C. difficile*, releases toxins that disrupt epithelial function and induce an inflammatory response that includes characteristic volcano-like eruptions of neutrophils from colonic crypts. These spread to form mucopurulent pseudomembranes.
- Norovirus is a common cause of self-limited diarrhea in adults and children. It spreads from person to person in sporadic cases and via contaminated water in epidemics.
- Rotavirus causes severe childhood diarrhea and significant mortality worldwide. Widespread vaccination has reduced the incidence of rotaviral infection.
- Parasitic and protozoal infections affect more than one-half of the world's population on a chronic or recurrent basis. Each parasite has a distinctive life cycle and tissue reaction. Those that invade are typically associated with tissue and systemic eosinophilia.

IRRITABLE BOWEL SYNDROME

Irritable bowel syndrome (IBS) is characterized by chronic, relapsing abdominal pain, bloating, and changes in bowel habits. Despite very real symptoms, the endoscopic and microscopic evaluations are normal in IBS patients. Thus, the diagnosis depends on clinical symptoms and functional testing. It should be recognized that IBS is a syndrome and that multiple illnesses may be represented under this global descriptor. IBS is currently divided into diarrhea-predominant, constipation-predominant, and mixed subtypes, as defined by successive revisions of the Rome criteria.

Pathogenesis

The pathogenesis of IBS remains poorly defined, although there is clearly interplay between psychologic stressors, diet, perturbation of the gut microbiome, increased enteric sensory responses to GI stimuli, and abnormal GI motility. For example, patients with constipation-predominant or diarrhea-predominant IBS tend to have decreased or increased colonic contractions and transit rates, respectively. Excess bile acid synthesis or bile acid malabsorption has been identified as one cause of diarrhea-predominant IBS, likely due to the effects of bile acids on intestinal motility and epithelial ion transport.

Other data link disturbances in enteric nervous system function to IBS, suggesting a role for defective brain-gut axis signaling. Consistent with this, deep sequencing and genome-wide association studies have linked several candidate genes to IBS including serotonin reuptake transporters, cannabinoid receptors, and TNF-related inflammatory mediators. Further, 5-HT3 receptor antagonists are effective in many cases of diarrhea-predominant IBS. Opioids and psychoactive drugs with anti-cholinergic effects are also used to treat diarrhea-predominant IBS.

A separate group of IBS patients relate onset to a bout of infectious gastroenteritis, suggesting that immune activation or, alternatively, a shift in the gut microbiome may trigger disease.

Clinical Features

The peak prevalence of IBS is between 20 and 40 years of age, and there is a significant female predominance. Variability in diagnostic criteria makes it difficult to establish the incidence, but most authors report a prevalence in high income countries of between 5% and 10%. IBS is presently diagnosed using clinical criteria that require the occurrence of abdominal pain or discomfort at least 3 days per month over 3 months with improvement following defecation and a change in stool frequency or form. Other causes, such as enteric infection or inflammatory bowel disease, must be excluded.

INFLAMMATORY BOWEL DISEASE (IBD)

Inflammatory bowel disease (IBD) is a chronic condition resulting from complex interactions between intestinal microbiota and host immunity in genetically predisposed individuals that leads to inappropriate mucosal immune

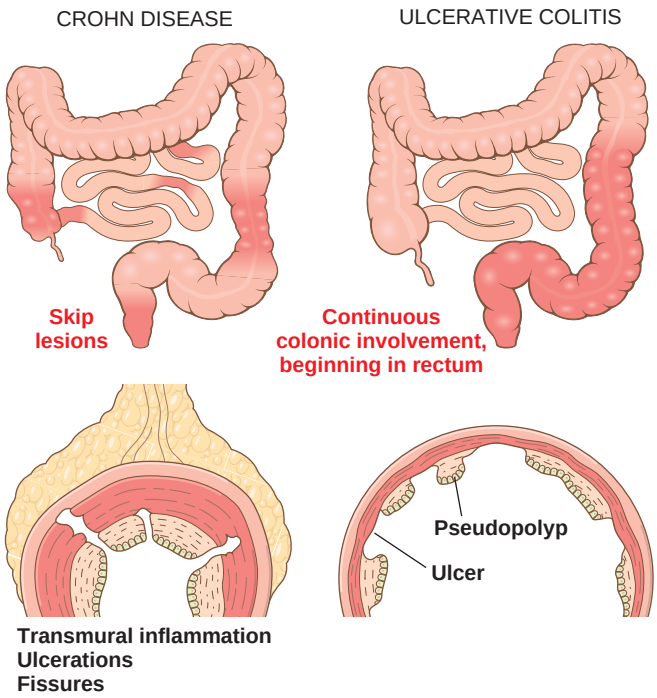


Figure 17.33 Distribution of lesions in inflammatory bowel disease. The distinction between Crohn disease and ulcerative colitis is primarily based on morphology.

Table 17.8 Features That Differ Between Crohn Disease and Ulcerative Colitis

Feature	Crohn Disease	Ulcerative Colitis
Macroscopic		
Bowel region	Ileum ± colon	Colon only
Distribution	Skip lesions	Diffuse
Stricture	Yes	Rare
Wall appearance	Thick	Normal
Microscopic		
Inflammation	Transmural	Limited to mucosa
Pseudopolyps	Moderate	Marked
Ulcers	Deep, knife-like	Superficial, broad-based
Lymphoid reaction	Marked	Moderate
Fibrosis	Marked	Mild to none
Serositis	Marked	Mild to none
Granulomas	Yes (~35%)	No
Fistulae/sinuses	Yes	No
Clinical		
Perianal fistula	Yes (in colonic disease)	No
Fat/vitamin malabsorption	Yes	No
Malignant potential	With colonic involvement	Yes
Recurrence after surgery	Common	No
Toxic megacolon	No	Yes

All features may not be present in a single case.

activation. The two disorders that comprise IBD are ulcerative colitis and Crohn disease. Descriptions of ulcerative colitis and Crohn disease date back to antiquity and at least the 16th century, respectively, but it took modern microbiologic techniques to exclude conventional infectious etiologies for these diseases. As will be discussed later, however, the luminal microbiota likely play an important role in IBD pathogenesis.

The distinction between ulcerative colitis and Crohn disease is primarily based on the distribution of affected sites (Fig. 17.33) and the morphologic expression of disease (Table 17.8) at those sites. **Ulcerative colitis involves only the colon and rectum and is generally limited to the mucosa and submucosa. In contrast, Crohn disease, which has also been referred to as regional enteritis (because of frequent ileal involvement) may involve any area of the GI tract and is typically transmural.**

Epidemiology. Ulcerative colitis and Crohn disease most frequently present in the teens and early 20s but can develop at any age. IBD is most common among Caucasians and in the United States occurs three to five times more often among eastern European (Ashkenazi) Jews than the general population. This is at least partly due to genetic factors, as discussed later. IBD is most common in North America, northern Europe, and Australia, but incidence worldwide is on the rise and is becoming significant in Africa, South America, and Asia, where prevalence was historically low. The hygiene hypothesis suggests that this increasing incidence is related to improved food storage conditions, decreased food contamination, and changes in gut microbiome composition that result in inadequate development of regulatory processes that limit mucosal immune responses. This in turn allows some mucosa-associated microbes to trigger persistent and chronic inflammation in susceptible hosts. Other potential explanations for increased IBD prevalence include the idea that preservatives and other materials added to processed foods induce low-grade mucosal damage that predisposes to IBD.

Pathogenesis

IBD results from a combination of abnormalities in immune regulation, host-microbe interactions, and epithelial barrier functions in genetically susceptible individuals. Over 200 IBD-associated genetic polymorphisms have been identified, but these account for less than 50% of disease risk in Crohn disease and make even smaller contributions to ulcerative colitis. For example, polymorphisms of *NOD2,* the strongest risk gene for Crohn disease, are associated with only a 10-fold increased risk of disease. Further, risk-associated *NOD2* alleles are found in about 35% of Caucasians with Crohn disease, or twice as often as in healthy Caucasians. Thus, while genetic predisposition is important, environmental factors are also critical to pathogenesis. The genetic and environmental elements that contribute to disease can be thought of in terms of immunity, autophagy and cellular stress responses, and host-microbial interactions.

- *Mucosal immunity.* A plethora of immune signaling and regulatory genes including those encoding HLA molecules and cytokines have been associated with IBD. In the case of the latter, polymorphisms in genetic loci that include genes in both proinflammatory, e.g., interferon-γ, and

anti-inflammatory (immunoregulatory), e.g., IL-10 and the IL-10 receptor, signaling are involved. Th1 polarization is present in both diseases, although there is also evidence of Th2 activation in ulcerative colitis, perhaps reflecting the underlying differences in genes associated with disease. Th17 signaling is also important to IBD pathogenesis, as both Crohn disease and ulcerative colitis are linked to polymorphisms in the IL-23 receptor and other molecules involved in Th17 signaling. These genetic data are consistent with the observation that Th17 T-cell populations are expanded within diseased intestine.

Many of the genes linked to IBD are also associated with other autoimmune diseases including diabetes mellitus, rheumatoid arthritis, and psoriasis. In most cases the risk alleles are consistent across diseases, but in a few cases alleles linked to increased risk in other diseases are protective in IBD. This may explain why some immunomodulatory therapies that are effective in IBD enhance progression of other diseases, while, conversely, some therapies that are effective in other diseases may exacerbate IBD.

Overall, multiple genetic polymorphisms contribute to IBD pathogenesis, and the specific genes involved vary between Crohn disease and ulcerative colitis, among individuals, and, in a few genes, across ethnic groups. For example, *NOD2* polymorphisms have not been associated with Crohn disease in Asian populations. Within ethnic groups, network analyses have demonstrated that the constellation of risk alleles present correlates with disease phenotype including the presence of colonic involvement, strictures, or fistulae in Crohn disease.

- *Autophagy and cellular stress responses.* Genetic associations as well as molecular analyses indicate that defects in autophagy and cellular stress responses contribute to IBD pathogenesis. The most studied of these genes include *ATG16L1* and *IRGM,* both of which are involved in autophagosome formation. Autophagy is a normal homeostatic mechanism that clears damaged organelles and is upregulated in response to cellular stress including nutrient deprivation and endoplasmic reticulum stress. Autophagy is also a means of clearing sources of reactive oxygen species and intracellular pathogens. The *ATG16L1* mutation that is linked to Crohn disease promotes ATG16L1 degradation, thereby limiting autophagy. The precise pathways by which loss of ATG16L1 function leads to disease are still being defined, but it is noteworthy that Paneth cell granules, which contain antibacterial peptides that are released into the crypt lumen, are structurally and functionally abnormal in patients and mice with *ATG16L1* defects.
- *Host-microbial interactions.* The gut microbiome, composed of bacteria, fungi, and viruses, has been the subject of intense investigation over the last decade. Because many of these organisms are not culturable by standard techniques, the studies have largely relied on high-throughput sequencing technologies.

  In general, the microbiome is populated by a small number of species at birth. This evolves to become a far more complex ecosystem with many more species after weaning and through childhood. Microbial complexity then declines in old age. Data from human subjects as well as experimental models have shown that microbes express proteins and other molecules and generate metabolites that can protect against or promote disease. Specific examples include *Clostridia* species, which stimulate development of regulatory T cells, and *Bifidobacterium,* which promote Th17 differentiation. In addition, multiple species produce butyrate, whose activities include enhancement of mucosal barrier function, increased epithelial proliferation, and immune regulation.

  Soluble microbial products activate host sensing proteins, including surface TLRs and dectin receptors and intracellular microbial-associated pattern receptors, e.g., *NOD2*. This is one means by which the microbiome can modulate mucosal immunity. Conversely, the immune system can markedly alter the microbiome by mechanisms that include luminal secretion of antimicrobial peptides and antibacterial IgA. This interplay between microbes and the immune system may contribute to the evolution of dysbiosis, characterized by shifts in microbial populations and reduced species diversity, in disease. Thus, although dysbiosis is often established at disease presentation, it is difficult to differentiate between microbial changes that trigger disease and those that are caused by disease.

  A notable exception to the polygenic nature of IBD is very early onset IBD, which can be driven by mutations in single genes required for epithelial ion transport, immune signaling, or host defense. Studies of these monogenic forms of IBD have shed light on mechanisms involved in more common forms of IBD. For example, it was learned that macrophages produce excessive IL-1β in infants with IL-10 receptor mutations (the most common mutation in very early onset IBD). These patients have benefited from IL-1 receptor antagonist treatment, and the efficacy of similar treatment is now being evaluated in polygenic IBD.

Crohn Disease

The eponym *Crohn disease* is based on a 1932 publication, but the entity was described centuries earlier. For example, Louis XIII of France (1601–1643) suffered relapsing bloody diarrhea, fever, rectal abscess, small intestinal and colonic ulcers, and fistulae beginning at age 20 years, all of which are classic features of Crohn disease.

MORPHOLOGY

Crohn disease may occur in any area of the GI tract but most commonly involves the terminal ileum, ileocecal valve, and cecum. Disease is limited to the small intestine in about 40% of cases; the small intestine and colon are both involved in 30% of patients; the remainder have only colonic involvement. The presence of multiple, separate, sharply delineated areas of disease, resulting in **skip lesions,** is characteristic, and when present, differentiates Crohn disease from ulcerative colitis. The presence of strictures, which occur commonly in Crohn disease but only rarely in long-standing ulcerative colitis, may also be helpful (Fig. 17.34A).

The earliest lesion of Crohn disease, the **aphthous ulcer,** may progress, and multiple lesions often coalesce into elongated, serpentine ulcers oriented along the axis of the bowel (Fig. 17.34B). Edema and loss of the normal mucosal folds are common. Ulceration with sparing of interspersed mucosa, a result of the patchy distribution of Crohn disease, results in an irregular, **cobblestone** appearance of the mucosa (Fig. 17.34B). **Fissures**

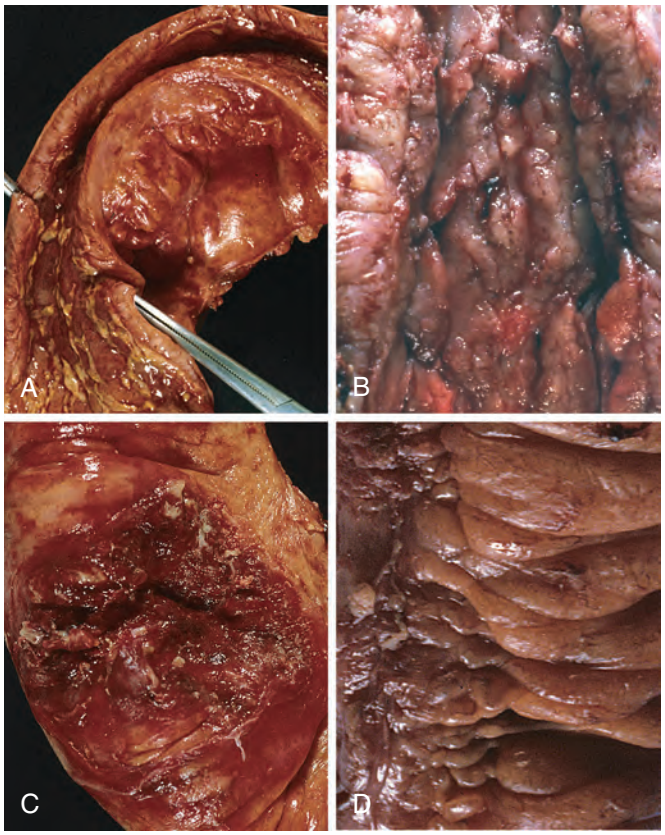


Figure 17.34 Gross pathology of Crohn disease. (A) Small intestinal stricture. (B) Linear mucosal ulcers, which impart a cobblestone appearance to the mucosa, and thickened intestinal wall. (C) Perforation and associated serositis. (D) Creeping fat.

frequently develop and may extend deeply to become fistula tracts or sites of perforation (Fig. 17.34C). The intestinal wall is thickened and rubbery as a consequence of transmural edema, inflammation, submucosal fibrosis, and hypertrophy of the muscularis propria, all of which contribute to stricture formation (see Fig. 17.34A). In cases with extensive transmural disease, mesenteric adipose tissue frequently extends over the serosal surface **(creeping fat)** (Fig. 17.34D).

The microscopic features of active Crohn disease include abundant neutrophils that infiltrate and damage crypt epithelium. Clusters of neutrophils within a crypt are referred to as **crypt abscesses** and are often associated with crypt destruction. Ulceration is common in Crohn disease, and there may be an abrupt transition between ulcerated and adjacent normal mucosa. Even in areas where gross examination suggests diffuse disease, microscopic pathology can appear patchy. Repeated cycles of crypt destruction and regeneration lead to **distortion and disorganzation of mucosal glands;** the normally straight and parallel crypts take on bizarre branching shapes and unusual orientations to one another (Fig. 17.35A). Epithelial metaplasia is another consequence of chronic relapsing injury. One form, **pseudopyloric metaplasia,** refers to the presence of gastric antral-appearing glands. **Paneth cell metaplasia** may also occur in the left colon, where Paneth cells are normally absent. These architectural and metaplastic changes may persist even when active inflammation has resolved. Mucosal atrophy, with loss of crypts, may also occur after years of disease. **Noncaseating granulomas** (Fig. 17.35C), a hallmark of Crohn disease, are found in approximately 35% of cases and may occur in areas of active disease (Fig. 17.35C) or uninvolved regions within any layer of the intestinal wall (Fig. 17.35D). Granulomas may also be present in draining mesenteric lymph nodes.

Clinical Features

The clinical manifestations of Crohn disease are extremely variable. In most patients, disease begins with intermittent attacks of relatively mild diarrhea, fever, and abdominal pain. Approximately 20% of patients present acutely with right lower quadrant pain, fever, and bloody diarrhea that may mimic acute appendicitis or bowel perforation. Periods of active disease are typically interrupted by asymptomatic periods that last for weeks to many months. Disease reactivation can be associated with a variety of external triggers, including physical or emotional stress, specific dietary items, and cigarette smoking. The latter is a strong risk factor for development of Crohn disease, and in some cases, disease onset is associated with initiation of smoking. Unfortunately, smoking cessation does not result in disease remission.

Iron deficiency anemia due to blood loss may develop in individuals with colonic disease, while extensive small bowel disease may result in protein loss sufficient to cause hypoalbuminemia and malabsorption of nutrients, vitamin B_{12}, and bile salts. Fibrosing strictures, particularly of the terminal ileum, are common and require surgical resection. Disease often recurs at the site of anastomosis, and as many as 40% of patients require additional resections within 10 years. Fistulae develop between loops of bowel and may also involve the urinary bladder, vagina, and abdominal or perianal skin. Perforation and peritoneal abscesses are common. Therapies include anti-inflammatory agents, e.g., salicylates; immunosuppressive drugs, e.g., corticosteroids; and biologic therapies, e.g., anti-TNF antibodies. Over the last two decades, anti-TNF antibodies have revolutionized treatment of Crohn disease. More recently, other biologic therapies including antibodies against other cytokines and cell adhesion proteins that are necessary for inflammatory cell migration as well as targeted kinase inhibitors are in various stages of testing and approved clinical use. It is not difficult to envision treatment algorithms that use the genetics, immune function, and microbial composition of individual patients to guide selection of targeted therapies.

Extraintestinal manifestations of Crohn disease include cutaneous nodules formed by granulomas, uveitis, migratory polyarthritis, sacroileitis, ankylosing spondylitis, erythema nodosum, cutaneous granulomas, and clubbing of the fingertips, any of which may develop before intestinal disease is recognized. Pericholangitis and primary sclerosing cholangitis occur in individuals with Crohn disease with a higher frequency than in those without Crohn disease, but are even more common in individuals with ulcerative colitis (see below and Chapter 18). As discussed later, risk of colonic adenocarcinoma is increased in patients with long-standing colonic disease.

Ulcerative Colitis

Ulcerative colitis is closely related to Crohn disease, but its intestinal involvement is limited to the colon and rectum.

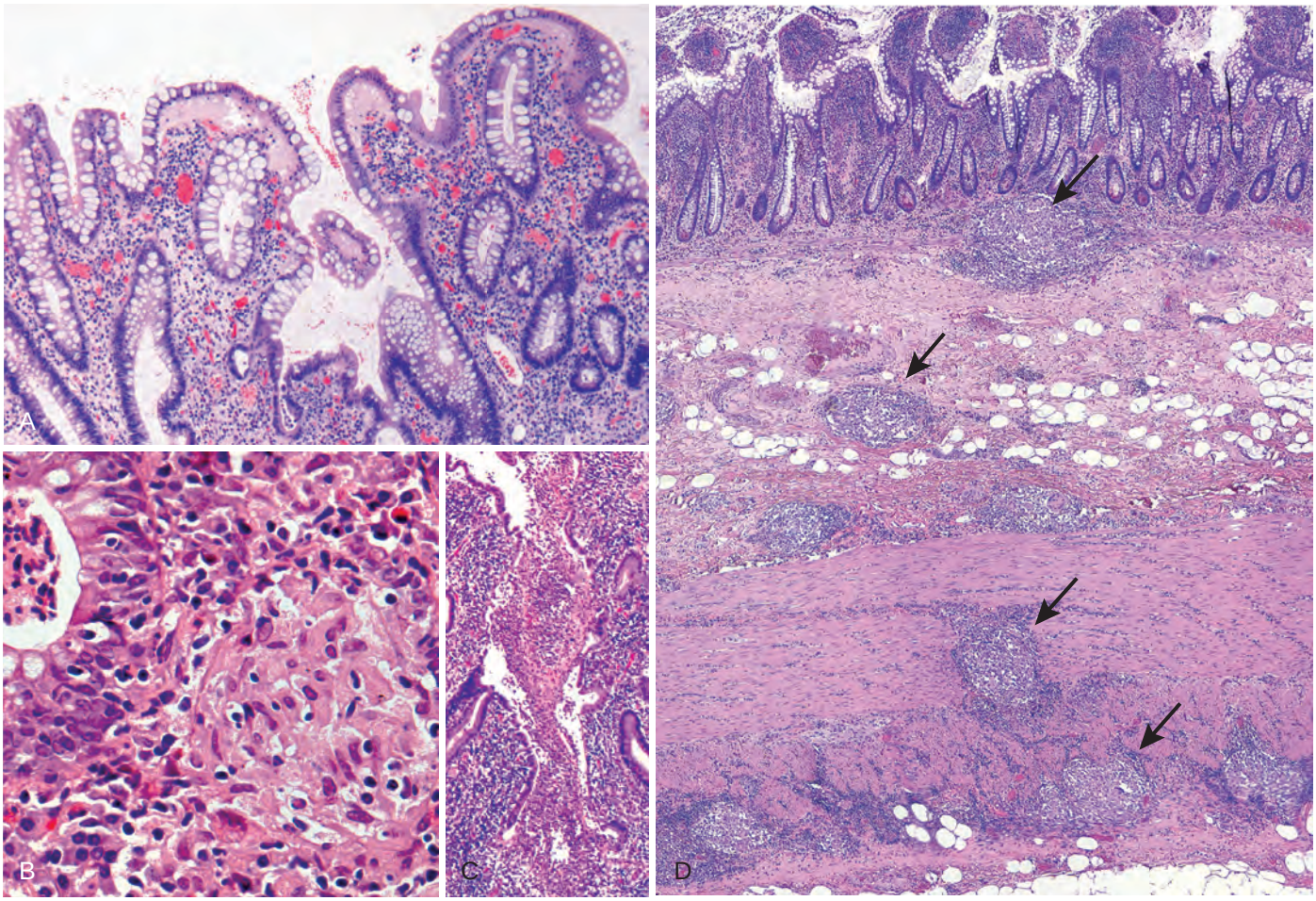


Figure 17.35 Microscopic pathology of Crohn disease. (A) Haphazard crypt organization results from repeated injury and regeneration. (B) Noncaseating granuloma. (C) Active Crohn disease, with ulceration and purulent exudate. (D) Transmural Crohn disease with submucosal and serosal granulomas *(arrows)*.

Extraintestinal manifestations of ulcerative colitis overlap with those of Crohn disease and include migratory polyarthritis, sacroiliitis, ankylosing spondylitis, uveitis, and skin lesions. Approximately 2.5% to 7.5% of individuals with ulcerative colitis also have primary sclerosis cholangitis (Chapter 18). The long-term outlook for ulcerative colitis patients depends on the severity of active disease and disease duration. Many of the therapies, including anti-TNF antibodies, that are effective in Crohn disease are also effective in ulcerative colitis. Some Crohn disease treatments are, however, ineffective in ulcerative colitis suggesting that the molecular mechanisms underlying these two forms of IBD are not entirely overlapping.

MORPHOLOGY

Grossly, ulcerative colitis always involves the rectum and extends proximally in a continuous fashion to involve part or all of the colon. Disease of the entire colon is termed **pancolitis** (Fig. 17.36A). Limited distal disease may be referred to descriptively as **ulcerative proctitis** or **ulcerative proctosigmoiditis.** The small intestine is normal, although mild mucosal inflammation of the distal ileum, termed *backwash ileitis,* may be present in severe cases of pancolitis. Skip lesions are not seen (although focal appendiceal or cecal inflammation may occasionally be present in ulcerative colitis that is otherwise limited to the distal colon).

Grossly, involved colonic mucosa may be slightly red and granular or have extensive, broad-based ulcers. There can be an abrupt transition between diseased and uninvolved colon (Fig. 17.36B). Ulcers are aligned along the long axis of the colon but do not typically replicate the serpentine ulcers of Crohn disease. Isolated islands of regenerating mucosa often bulge into the lumen to create **pseudopolyps** (Fig. 17.36C), and the tips of these polyps may fuse to create **mucosal bridges** (Fig. 17.36D). Chronic disease may lead to **mucosal atrophy** with a smooth mucosal surface that lacks normal folds. Unlike Crohn disease, ulcerative colitis is not transmural. As a result **the colon wall is not thickened, the serosal surface is normal, and strictures do not occur.** Uncommonly, severe cases are associated with inflammation of the muscularis propria and neuromuscular dysfunction leading to colonic dilation and **toxic megacolon,** which carries a significant risk of perforation.

Histologic features of mucosal disease in ulcerative colitis are similar to colonic Crohn disease and include inflammatory infiltrates, crypt abscesses (Fig. 17.37A), crypt distortion, and pseudopyloric epithelial metaplasia (Fig. 17.37B). In contrast to

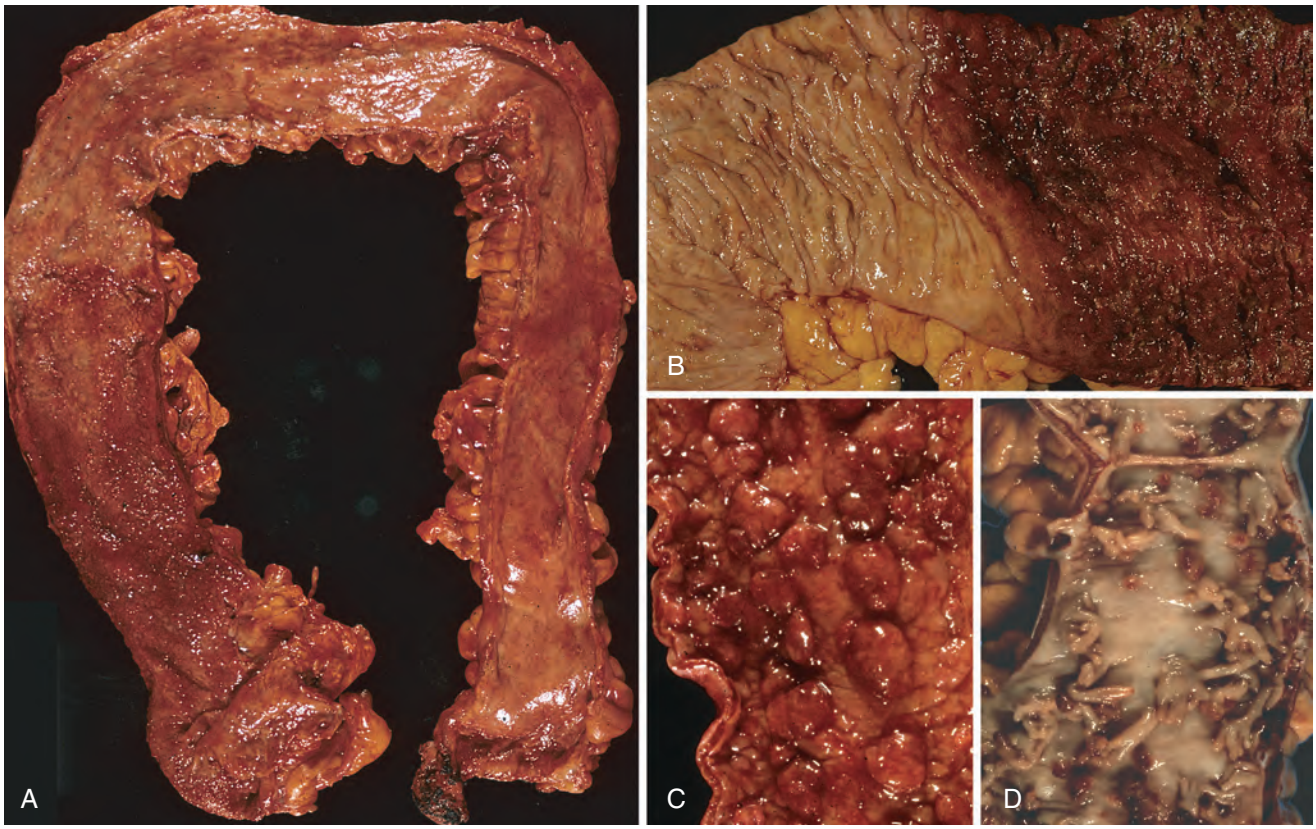


Figure 17.36 Gross pathology of ulcerative colitis. (A) Total colectomy with pancolitis showing active disease, with red, granular mucosa in the cecum (left) and smooth, atrophic mucosa distally (right). (B) Sharp demarcation between active ulcerative colitis (right) and normal mucosa (left). (C) Inflammatory polyps. (D) Mucosal bridges can join inflammatory polyps.

Crohn disease, granulomas are not present, and the diffuse inflammation is generally limited to the mucosa and superficial submucosa (Fig. 17.37C). In severe cases, extensive mucosal destruction may be accompanied by ulcers that extend into the submucosa, but the muscularis propria is rarely involved. Submucosal fibrosis, mucosal atrophy, and distorted mucosal architecture remain as residua of healed disease, but histology may also revert to near normal after prolonged remission.

Clinical Features

Ulcerative colitis is a relapsing disorder characterized by attacks of bloody diarrhea with stringy mucoid material, lower abdominal pain, and cramps that are temporarily relieved by defecation. These symptoms may persist for days, weeks, or months before they subside. The initial attack may, in some cases, be severe enough to constitute a medical or surgical emergency. More than half of patients have clinically mild disease, although almost all experience at least one relapse during a 10-year period. Historically, up to 30% of those affected required colectomy within the first 3 years after presentation because of uncontrollable symptoms, but the incidence of colectomy has fallen sharply with improvements in medical management. Colectomy effectively cures intestinal disease in ulcerative colitis, but extraintestinal manifestations may persist.

The factors that trigger development of ulcerative colitis in previously healthy individuals are not known. However, infectious enteritis precedes disease onset in some cases. It has been hypothesized that enteritis triggers mucosal immune activation and microbial changes that lead to disease in susceptible individuals. In other patients the first episode of disease is preceded by psychologic stress, which may also be linked to relapse during remission. The initial onset of symptoms has also been reported to occur shortly after smoking cessation in some patients; in these smoking may partially relieve symptoms.

Indeterminate Colitis

Because of the extensive genetic, pathologic, and clinical overlap between ulcerative colitis and Crohn disease (see Table 17.8), definitive diagnosis is not possible in up to 10% of IBD patients. These cases, termed indeterminate colitis, do not involve the small bowel and have colonic disease in a continuous pattern typical of ulcerative colitis but also have features suggestive of Crohn disease. These include patchy histologic disease, a family history of Crohn disease, and perianal lesions. Serologic studies can be useful in cases with overlapping features, as perinuclear anti-neutrophil cytoplasmic antibodies are found in 75% of individuals with ulcerative colitis and only 11% of individuals with Crohn disease. In contrast, antibodies to *Saccharomyces cerevisiae* are

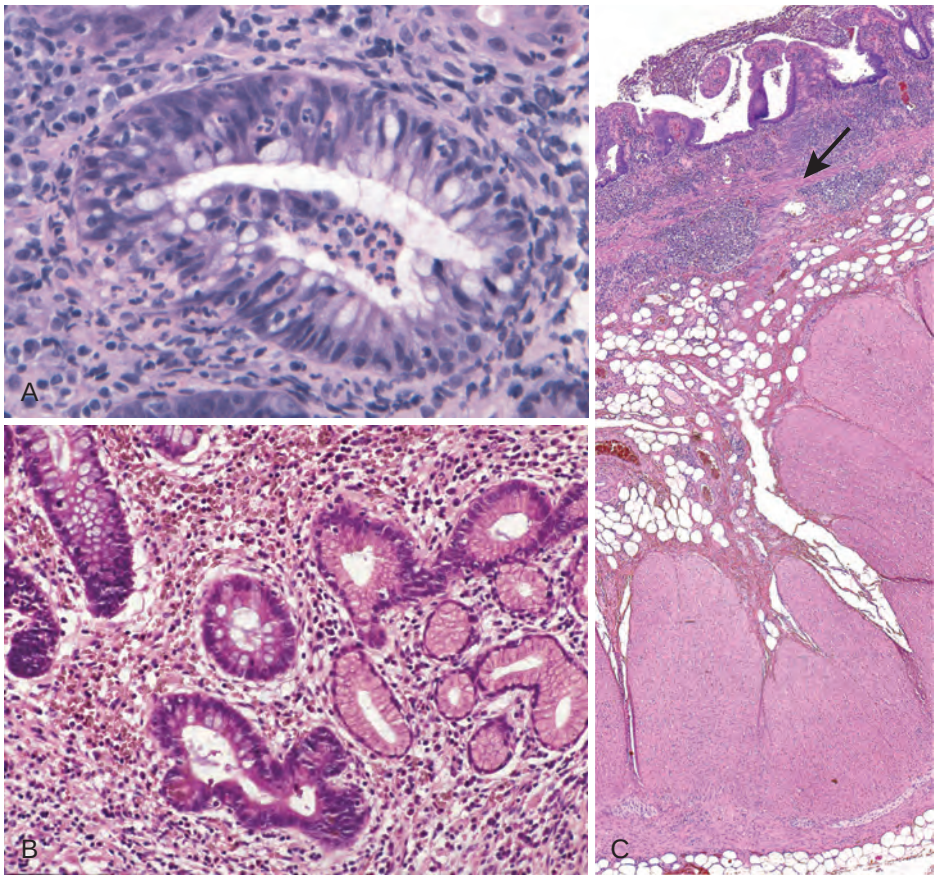


Figure 17.37 Histopathology of ulcerative colitis. (A) Crypt abscess. (B) Pseudopyloric metaplasia *(right)*. (C) Disease is limited to the mucosa (above the *arrow*). Compare to Fig. 17.35D.

often lacking in patients with ulcerative colitis and present in those with Crohn disease. However, the serologic results are frequently ambiguous in cases that are indeterminate on clinical grounds. Fortunately, overlap in medical management allows indeterminate colitis to be treated effectively.

Colitis-Associated Neoplasia

One of the most feared long-term complications of ulcerative colitis and colonic Crohn disease is the development of neoplasia. The risk of dysplasia and subsequent cancer is related to several factors:

- *Duration of the disease.* Risk increases sharply 8 to 10 years after disease onset.
- *Extent of the disease.* Patients with pancolitis are at greater risk than those with only left-sided disease. Crohn disease patients without colonic involvement are not at increased risk.
- *Nature of the inflammatory response.* Greater frequency and severity of active inflammation (characterized by the presence of neutrophils) confers increased risk.

To facilitate early detection of neoplasia, patients are typically enrolled in surveillance programs approximately 8 years after diagnosis of IBD. The major exception to this is patients with IBD and primary sclerosing cholangitis, who have a greater risk of developing cancer and are generally enrolled for surveillance at the time of diagnosis.

The goal of surveillance biopsies is to identify dysplastic epithelium, which is a precursor of colitis-associated carcinoma. Dysplasia can develop in flat areas of mucosa that are not grossly recognized as abnormal. High-resolution endoscopes and specialized endoscopic techniques, including chromoendoscopy, have helped target biopsies toward areas suspicious for dysplasia, but multiple surveillance biopsies and histological examination are still required. IBD-associated dysplasia is classified histologically as low grade or high grade (Fig. 17.38A, B) and may be multifocal. High-grade dysplasia may be associated with invasive carcinoma at the same site (Fig. 17.38C) or elsewhere in the colon and therefore often prompts colectomy. Low-grade dysplasia may be followed closely with frequent endoscopic surveillance, but colectomy is also common, particularly when multiple foci of flat (non-polypoid) dysplasia are present, in those with extensive or longstanding disease, and in older patients. Colonic adenomas (discussed later) also occur in IBD patients, and in some cases these may be difficult to differentiate from a polypoid focus of IBD-associated dysplasia. Fortunately, solitary polypoid lesions in ulcerative colitis without foci of flat dysplasia can often be managed by endoscopic resection and careful follow-up.

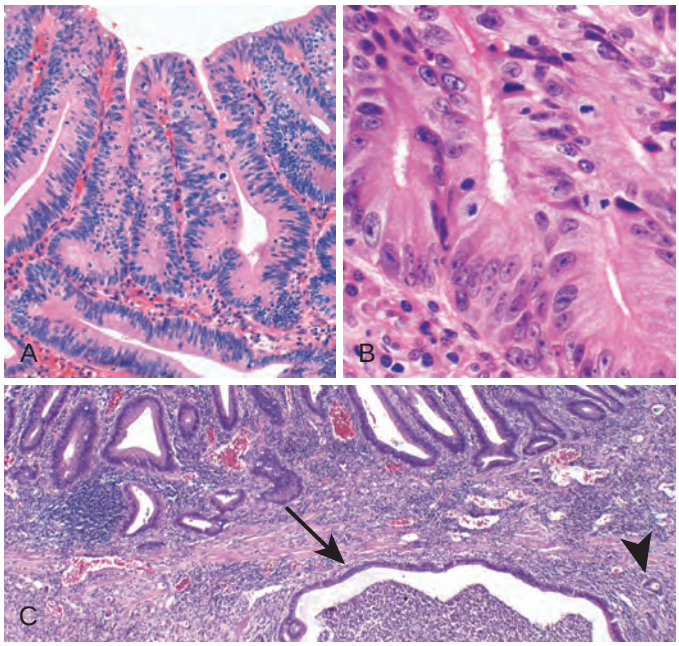


Figure 17.38 Colitis-associated dysplasia. (A) Dysplasia with extensive nuclear stratification and marked nuclear hyperchromasia. (B) Cribriform glandular arrangement in high-grade dysplasia. (C) Colectomy specimen with high-grade dysplasia on the surface and underlying invasive adenocarcinoma. A large cystic, neutrophil-filled space lined by invasive adenocarcinoma is apparent *(arrow)* beneath the muscularis mucosae. Also seen are small invasive glands *(arrowhead)*.

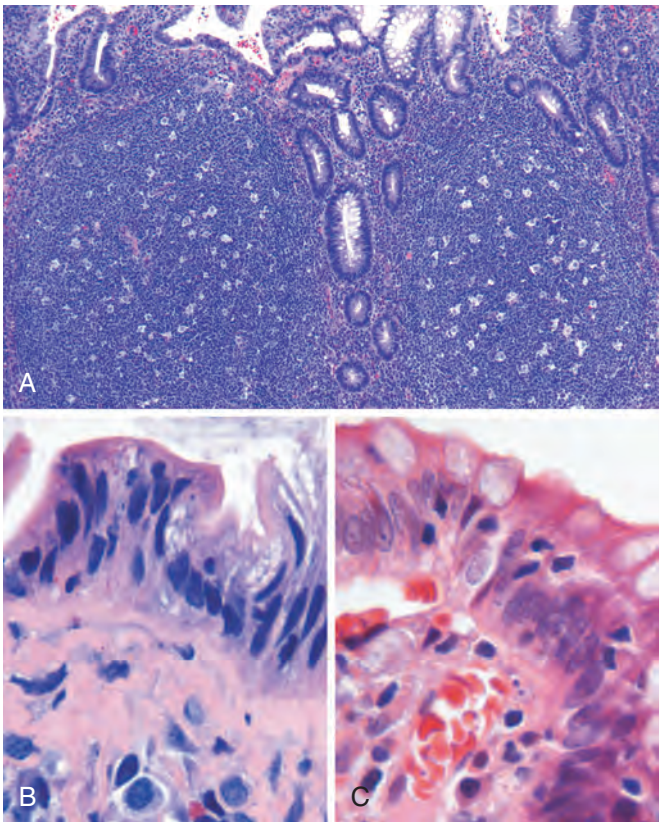


Figure 17.39 Uncommon causes of colitis. (A) Diversion colitis. Note the large lymphoid follicles with germinal centers. (B) Collagenous colitis with a dense subepithelial collagen band. (C) Lymphocytic colitis. Intraepithelial lymphocytes can be recognized by their densely stained, small nuclei.

OTHER CAUSES OF CHRONIC COLITIS

Diversion Colitis

Surgical treatment of ulcerative colitis, Hirschsprung disease, colon cancer, and other intestinal disorders sometimes requires creation of a temporary or permanent ostomy, resulting in a blind distal segment of colon from which the normal fecal flow is diverted. Colitis can develop within the diverted segment. Besides mucosal erythema and friability, the most striking feature of diversion colitis is the development of numerous mucosal lymphoid follicles (Fig. 17.39A). Increased numbers of lamina propria lymphocytes, monocytes, macrophages, and plasma cells may also be present. In severe cases the histopathology may resemble IBD and include crypt abscesses, mucosal architectural distortion, or, rarely, granulomas. The mechanisms responsible for diversion colitis are not well understood, but changes in the luminal microbiota and diversion of the fecal stream that provides nutrients to colonic epithelial cells have been proposed. Consistent with this, enemas containing short-chain fatty acids, a product of colonic bacterial metabolism that signals to and serves as an energy source for colonic epithelial cells, can promote mucosal recovery. Such enemas are, however, used infrequently because of butyrate's foul odor. Diversion colitis resolves after anastomosis and restoration of fecal flow.

Microscopic Colitis

Microscopic colitis encompasses two entities, collagenous colitis and lymphocytic colitis. These idiopathic diseases both present with chronic, nonbloody, watery diarrhea without weight loss. Radiologic and endoscopic studies are typically normal. Collagenous colitis, which occurs primarily in middle-aged and older women, is characterized by the presence of a dense subepithelial collagen layer, increased numbers of intraepithelial T lymphocytes, and a mixed inflammatory infiltrate within the lamina propria (Fig. 17.39B). Lymphocytic colitis is histologically similar, but the subepithelial collagen layer is of normal thickness and the increase in intraepithelial T lymphocytes is greater, frequently exceeding one per five colonocytes (Fig. 17.39C). Lymphocytic colitis shows a strong association with celiac disease and autoimmune diseases including Graves disease, rheumatoid arthritis, and autoimmune or lymphocytic gastritis. Similar histologic changes can be induced by NSAIDs and, possibly, proton pump inhibitors.

GRAFT-VERSUS-HOST DISEASE

Graft-versus-host disease occurs following allogeneic hematopoietic stem cell transplantation. The small bowel and colon are involved in most cases. Although graft-versus-host disease is secondary to donor T cells targeting antigens

on the recipient's GI epithelial cells, the lamina propria lymphocytic infiltrate is typically sparse. Epithelial apoptosis, particularly of crypt cells, is the most common histologic finding. The crypts may be completely destroyed in severe cases. Intestinal graft-versus-host disease often presents with watery diarrhea that may become bloody in severe cases. Intriguingly, the morphologic changes in graft-versus-host disease overlap with the enterocolitis and diarrhea that frequently develop in oncology patients treated with immune checkpoint inhibitors, complications that are becoming more frequent with increased use of these therapies. While the mechanisms of checkpoint inhibitor–induced enterocolitis are not well defined, they are likely to overlap, at least partially, with those of graft-versus-host disease.

SIGMOID DIVERTICULAR DISEASE

Diverticular disease is due to acquired pseudodiverticular outpouchings of the colonic mucosa and submucosa. Unlike true diverticula, such as Meckel diverticulum, these do not include all three layers of the colonic wall. Colonic diverticula are rare in persons younger than age 30, but the prevalence approaches 50% in Western adult populations older than age 60. Diverticulosis is less common in Japan as well as under-resourced countries, probably because of dietary differences. Moreover, most diverticula in Asia and Africa occur in the right colon, while right-sided diverticula are uncommon in Western countries. The reasons for this difference in distribution are not well defined.

Pathogenesis

Colonic diverticula are the product of the unique structure of the muscularis propria and elevated intraluminal pressure. Focal discontinuities are created in the inner circular muscle layer at sites where nerves and arterial vasa recta penetrate it. In the small intestine these gaps are reinforced by the external longitudinal layer of the muscularis propria, but in the colon this outer muscle layer is gathered into three bands, the taeniae coli. Intraluminal pressure, particularly in the sigmoid colon, is elevated by exaggerated peristaltic contraction, an effect that is enhanced by low-fiber diets with reduced stool bulk. This increased pressure leads to mucosal herniation through the anatomic weak points in the muscularis propria.

MORPHOLOGY

Colonic diverticula are small, flask-like outpouchings, usually 0.5 to 1 cm in diameter, that occur in a regular distribution (reflecting the anatomic contribution to pathogenesis) alongside the taeniae coli (Fig. 17.40A). These are most common in the sigmoid colon, but other parts of the colon may be affected. Colonic diverticula have a thin wall composed of a flattened or atrophic mucosa that can be surrounded by compressed submucosa and attenuated or absent muscularis (Fig. 17.40B, C). Hypertrophy of the circular layer of the muscularis propria in the affected bowel segment is common. Diverticular obstruction induces inflammation, termed diverticulitis. Because the wall of the diverticulum is supported only by the muscularis mucosae and a thin layer of subserosal adipose tissue, inflammation and increased pressure within an obstructed diverticulum can lead to perforation. With or without perforation, diverticulitis may cause segmental diverticular disease–associated colitis, fibrotic thickening in and around the colonic wall, or stricture formation. Perforation can result in pericolonic abscesses, sinus tracts, and, occasionally, peritonitis.

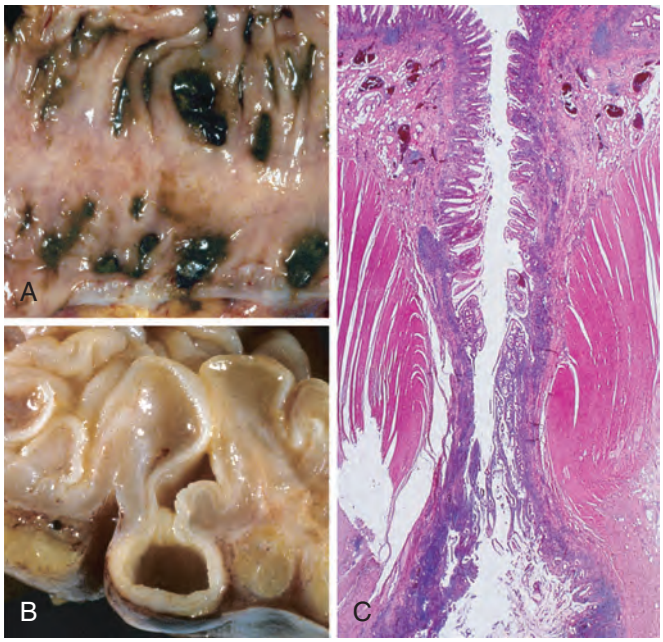


Figure 17.40 Sigmoid diverticular disease. (A) Stool-filled diverticula are regularly arranged. (B) Cross section showing the outpouching of mucosa through the muscularis propria. (C) Low-power photomicrograph of a sigmoid diverticulum showing protrusion of the mucosa through the muscularis propria.

Clinical Features

Most individuals with diverticular disease remain asymptomatic throughout their lives. However, about 20% develop manifestations such as intermittent cramping, continuous lower abdominal discomfort, constipation, distention, or a sensation of never being able to completely empty the rectum. Patients sometimes experience alternating constipation and diarrhea that mimics IBS. Whether a high-fiber diet prevents the development of diverticulosis or protects against diverticulitis is unclear, but diets supplemented with fiber may provide symptomatic improvement. Even when diverticulitis occurs, it most often resolves spontaneously. Surgical intervention is generally reserved for those with severe or recurrent diverticulitis.

KEY CONCEPTS

IRRITABLE BOWEL SYNDROME AND COLITIS

- Irritable bowel syndrome is characterized by chronic, relapsing abdominal pain, bloating, and changes in bowel habits without gross or histologic pathology. The pathogenesis is not defined, but likely includes contributions by psychologic stressors, diet,

the gut microbiome, abnormal GI motility, and increased enteric sensory responses to GI stimuli.
- IBD is an umbrella term for ulcerative colitis and Crohn disease. Indeterminate colitis is used for cases of IBD without definitive features of either ulcerative colitis or Crohn disease.
- Ulcerative colitis is limited to the colonic mucosa and submucosa, is continuous from the rectum, and ranges from only rectal disease to pancolitis; neither skip lesions nor granulomas are present.
- Crohn disease most commonly affects the terminal ileum and cecum, but can affect any site within the GI tract. Involvement is discontinuous, resulting in skip lesions. In affected areas disease can be transmural, affecting the full thickness of the entire bowel wall. Noncaseating granulomas are common.
- Both forms of IBD present most often in the teens and early 20s and are associated with extraintestinal manifestations.
- IBD is thought to arise from a synergy between genetic risk, mucosal immune dysfunction, and dysbiosis, i.e., disease-associated changes in microbial composition.
- The risk of colonic epithelial dysplasia and adenocarcinoma is increased in IBD patients who have had colonic disease for more than 8 to 10 years.
- The two forms of microscopic colitis, collagenous colitis and lymphocytic colitis, both cause chronic watery diarrhea. The intestines are endoscopically normal, and the diseases are identified by their characteristic histologic features.
- Diverticular disease of the sigmoid colon is common in Western populations older than age 60. The causes include low-fiber diets, colonic spasm, and the unique anatomy of the colon. Inflammation of diverticula, diverticulitis, affects a minority of those with diverticulosis, but can cause perforation in its most severe form.

POLYPS

Polyps are most common in the colon and rectum but, as discussed earlier, may occur in the esophagus, stomach, or small intestine. Most, if not all, polyps begin as small elevations of the mucosa. These are referred to as *sessile,* a term borrowed from botanists who use it to describe flowers and leaves that grow directly from the stem without a stalk. As sessile polyps enlarge, proliferation of cells adjacent to the mass and the effects of traction on the luminal protrusion may combine to create a stalk. Polyps with stalks are termed *pedunculated.* In general, intestinal polyps can be classified as nonneoplastic or neoplastic. The most common neoplastic polyp is the adenoma, which has the potential to progress to cancer. Nonneoplastic polyps can be classified as inflammatory, hamartomatous, or hyperplastic.

Hyperplastic Polyps

Colonic hyperplastic polyps are benign epithelial proliferations that are typically discovered in the sixth and seventh decades of life. The pathogenesis of hyperplastic polyps is incompletely understood, but they are thought to result from decreased epithelial cell turnover and delayed shedding of surface epithelial cells, leading to a "piling up" of goblet cells and absorptive cells. These lesions are without malignant potential. Their chief significance is that they must be distinguished from sessile serrated adenomas, which are histologically similar but have malignant potential, as described later. It is also important to remember that epithelial hyperplasia can occur as a nonspecific reaction adjacent to or overlying any mass or inflammatory lesion and therefore can be a clue to the presence of an adjacent, clinically important lesion.

MORPHOLOGY

Hyperplastic polyps are most commonly found in the left colon and are typically less than 5 mm in diameter. They are smooth, nodular protrusions of the mucosa, often on the crests of mucosal folds. They may occur singly but are more frequently multiple, particularly in the sigmoid colon and rectum. Histologically, hyperplastic polyps are composed of mature goblet and absorptive cells with a serrated surface architecture that is the morphologic hallmark of these lesions (Fig. 17.41). Serration is typically restricted to the upper third of the crypt.

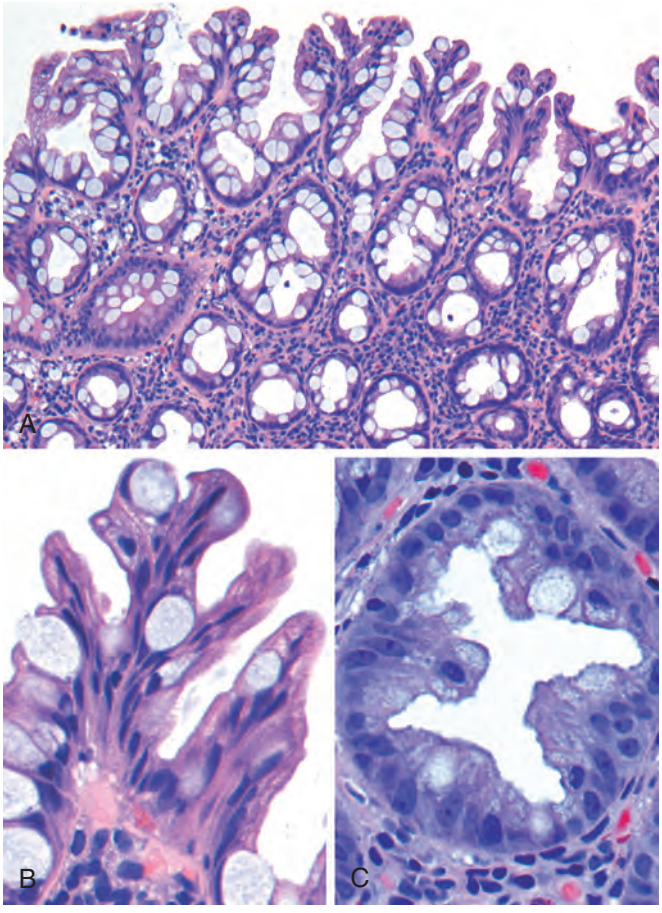


Figure 17.41 Hyperplastic polyp. (A) Polyp surface with irregular tufting of epithelial cells. (B) Tufting results from epithelial overcrowding. (C) Epithelial crowding produces a serrated architecture when crypts are cut in cross section.

Inflammatory Polyps

Polyps that form as part of the solitary rectal ulcer syndrome are an example of a purely inflammatory lesion. Patients present with a clinical triad of rectal bleeding, mucus discharge, and an inflammatory lesion of the anterior rectal wall. The underlying cause is impaired relaxation of the anorectal sphincter that creates a sharp angle at the anterior rectal shelf and leads to recurrent abrasion and ulceration of the overlying rectal mucosa. An inflammatory polyp may ultimately form as a result of chronic cycles of injury and healing. Entrapment of this polyp in the fecal stream leads to mucosal prolapse. The distinctive histologic features include mixed inflammatory infiltrates, erosion, and epithelial hyperplasia together with prolapse-induced lamina propria fibromuscular hyperplasia (Fig. 17.42).

Hamartomatous Polyps

Hamartomatous polyps occur sporadically or as components of genetically determined or acquired syndromes (Table 17.9).

Although they were originally thought to reflect developmental abnormalities, many hamartomatous polyp syndromes are caused by germline mutations in tumor suppressor genes or proto-oncogenes. Some of these syndromes are associated with increased cancer risk, either within the polyps or at other intestinal or extraintestinal sites. Thus, a subset of hamartomatous polyps can be considered to be premalignant, neoplastic lesions, much like adenomas. It is also important to recognize these polyps because of associated extraintestinal manifestations and the possibility that other family members are affected. Several of these syndromes are discussed below and others are summarized in Table 17.9.

Juvenile Polyps

Juvenile polyps are focal malformations of the epithelium and lamina propria that may be sporadic or syndromic. The majority occur in children younger than 5 years of age, but both sporadic and syndromic polyps can present at older ages. Many juvenile polyps are located in the rectum and cause rectal bleeding. In some cases intussusception, intestinal obstruction, or polyp prolapse (through the anal sphincter) may occur.

Sporadic juvenile polyps, which are also referred to as retention polyps, are usually solitary. In contrast, the autosomal dominant syndrome of juvenile polyposis is characterized by from three to hundreds of hamartomatous polyps. In these cases, colectomy may be necessary to limit the chronic and sometimes severe hemorrhage associated with polyp ulceration. A minority of patients also have polyps in the stomach and small bowel that can undergo malignant transformation. Pulmonary arteriovenous malformations and other congenital malformations are recognized extraintestinal manifestations of juvenile polyposis.

MORPHOLOGY

Most juvenile polyps are less than 3 cm in diameter. They are typically pedunculated, smooth-surfaced, reddish lesions with characteristic cystic spaces. Microscopic examination shows these cysts to be dilated glands filled with mucin and inflammatory debris (Fig. 17.43). The remainder of the polyp is composed of lamina propria expanded by mixed inflammatory infiltrates. The muscularis mucosae may be normal or attenuated.

Although the morphogenesis of juvenile polyps is incompletely understood, it has been proposed that mucosal hyperplasia is the initiating event. This hypothesis is consistent with the discovery that mutations in pathways that regulate cellular growth cause autosomal dominant juvenile polyposis. The most common mutation identified is of *SMAD4,* which encodes a signaling intermediate in the TGF-β pathway. These patients often have both juvenile polyposis and hereditary hemorrhagic telangiectasia. In other cases, mutations in *BMPR1A,* a kinase that is a member of the TGF-β superfamily, cause disease (see Table 17.9). Together these mutations account for fewer than half of patients. Thus, other genes responsible for autosomal dominant juvenile polyposis remain to be discovered.

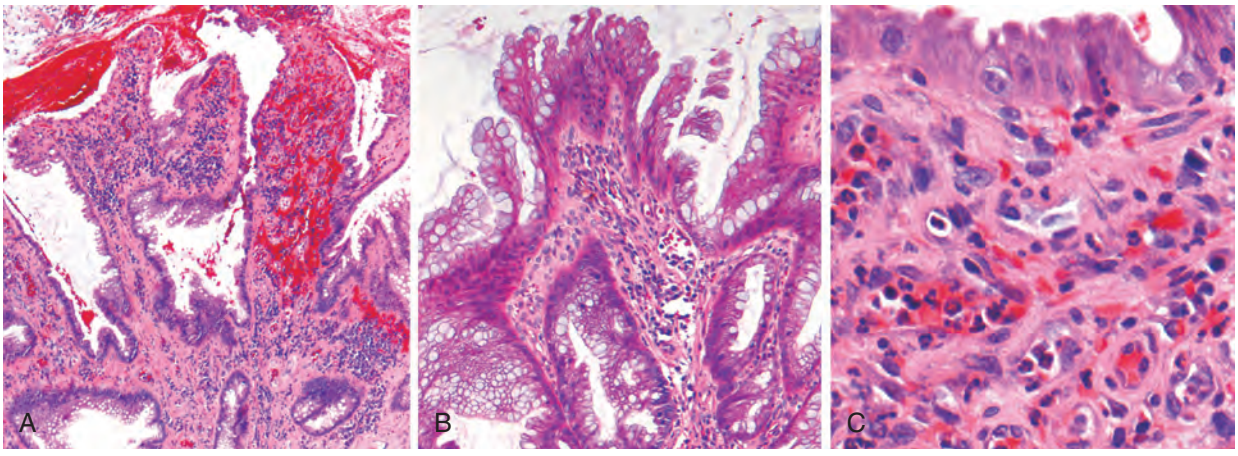


Figure 17.42 Solitary rectal ulcer syndrome. (A) The dilated glands, proliferative epithelium, superficial erosions, and inflammatory infiltrate are typical of an inflammatory polyp. However, the smooth muscle hyperplasia within the lamina propria indicates that mucosal prolapse has also occurred. (B) Epithelial hyperplasia. (C) Granulation tissue–like capillary proliferation within the lamina propria caused by repeated erosion.

Table 17.9 Gastrointestinal Polyposis Syndromes

Syndrome	Mean Age at Presentation (years)	Mutated Genes; Pathway	GI Lesions	Selected Extra-GI Manifestations
Juvenile polyposis	<5	*SMAD4,BMPR1A*; TGF-β signaling pathway	Juvenile polyps; risk of gastric, small intestinal, colonic, and pancreatic adenocarcinoma	Congenital malformations, digital clubbing
Peutz-Jeghers syndrome	10–15	*STK11*; AMP kinase-related pathways	Arborizing polyps; small intestine > colon > stomach; colonic adenocarcinoma	Pigmented macules; risk of colon, breast, lung, pancreatic, and thyroid cancer
Cowden syndrome, Bannayan-Ruvalcaba-Riley syndrome[a]	<15	*PTEN*: PI3K/AKT pathway	Hamartomatous/ inflammatory intestinal polyps, lipomas, ganglioneuromas	Benign skin tumors, benign and malignant thyroid and breast lesions; no increase in GI cancers
Cronkhite-Canada syndrome	>50	Nonhereditary, unknown cause	Hamartomatous polyps of stomach, small intestine colon; abnormalities in nonpolypoid mucosa	Nail atrophy, hair loss, abnormal skin pigmentation, cachexia, and anemia. Fatal in up to 50%
Tuberous sclerosis		*TSC1* (hamartin), *TSC2* (tuberin); mTOR pathway	Hamartomatous polyps	Intellectual disability, epilepsy, facial angiofibroma, cortical (CNS) tubers, renal angiomyolipoma
FAP				
Classic FAP	10–15	*APC*	Multiple adenomas	Congenital RPE hypertrophy
Attenuated FAP	40–50	*APC*	Multiple adenomas	
Gardner syndrome	10–15	*APC*	Multiple adenomas	Osteomas, thyroid and desmoid tumors, skin cysts
Turcot syndrome	10–15	*APC*	Multiple adenomas	Medulloblastoma, glioblastoma
MUTYH-associated polyposis	30–50	*MUTYH*	Multiple adenomas	Gastric and duodenal polyps

[a]Also called PTEN hamartoma-tumor syndrome.

CNS, Central nervous system; *FAP*, familial adenomatous polyposis; *GI*, gastrointestinal; *mTOR*, mammalian target of rapamycin; *RPE*, retinal pigmented epithelium; *TGF-β*, transforming growth factor beta.

Dysplasia is rare in sporadic juvenile polyps. In contrast, juvenile polyposis syndrome is associated with dysplasia, both within the juvenile polyps and in separate adenomas. As a result, 30% to 50% of patients with juvenile polyposis develop colonic adenocarcinoma by age 45.

Peutz-Jeghers Syndrome

This rare autosomal dominant syndrome presents at a median age of 11 years with multiple GI hamartomatous polyps and mucocutaneous hyperpigmentation. The latter takes the form of dark blue to brown macules on the lips, nostrils, buccal mucosa, palmar surfaces of the hands, genitalia, and perianal region. These lesions are similar to freckles but are distinguished by their presence in the buccal mucosa. Peutz-Jeghers polyps can initiate intussusception, which is occasionally fatal. Of greater importance, **Peutz-Jeghers syndrome is associated with a markedly increased risk of several malignancies.** Lifetime risk is approximately 40%, and regular surveillance is recommended beginning at birth for sex cord tumors of the testes; in late childhood for gastric and small intestinal cancers; and in the second and third decades of life for colon, pancreatic, breast, lung, ovarian, and uterine cancers.

Pathogenesis

Germline heterozygous loss-of-function mutations in the gene *STK11* are present in approximately half of individuals with familial Peutz-Jeghers syndrome as well as a subset of patients with sporadic Peutz-Jeghers syndrome. *STK11* is a tumor suppressor gene that regulates AMP-activated protein kinases (AMPK), which control cell polarization and act as a brake on growth and anabolic metabolism (Chapter 7). As is common with other tumor suppressor genes, the function of the second "normal" copy of *STK11* is often lost through somatic mutation in cancers occurring in Peutz-Jeghers syndrome, providing an explanation for the high risk of neoplasia in affected patients.

MORPHOLOGY

The polyps of Peutz-Jeghers syndrome are most common in the small intestine, although they may occur in the stomach and colon and, with much lower frequency, in the bladder and lungs. Grossly, the polyps are large and pedunculated with a lobulated contour. Histologic examination demonstrates a characteristic arborizing network of connective tissue, smooth muscle, lamina propria, and glands lined by normal-appearing intestinal epithelium (Fig. 17.44).

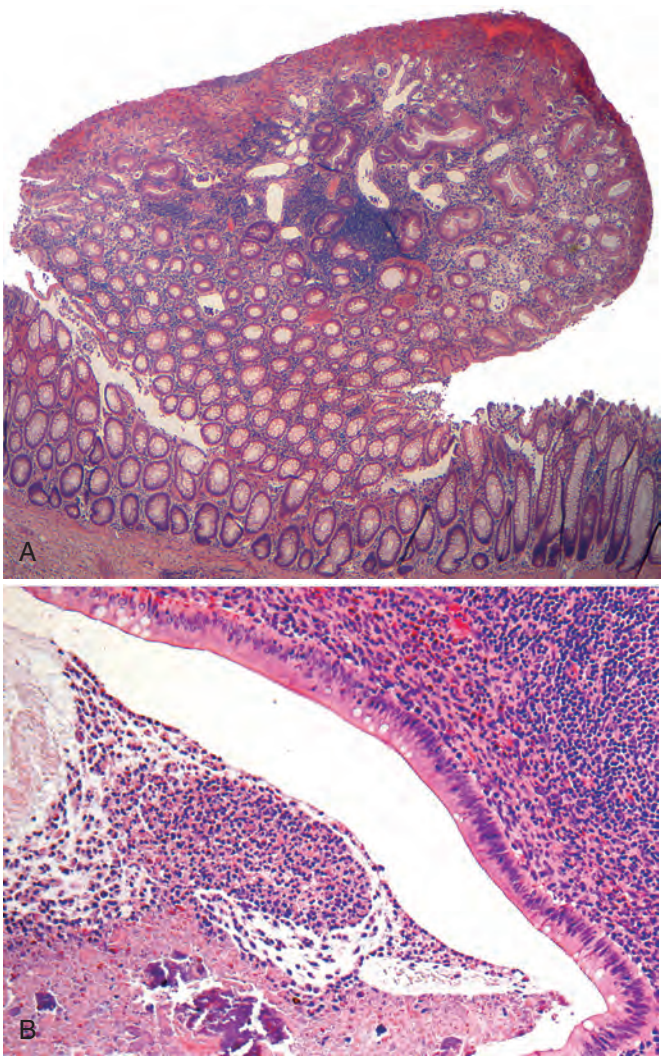


Figure 17.43 Juvenile polyposis. (A) Juvenile polyp. Note the surface erosion and cystically dilated crypts. (B) Inspissated mucus, neutrophils, and inflammatory debris can accumulate within dilated crypts.

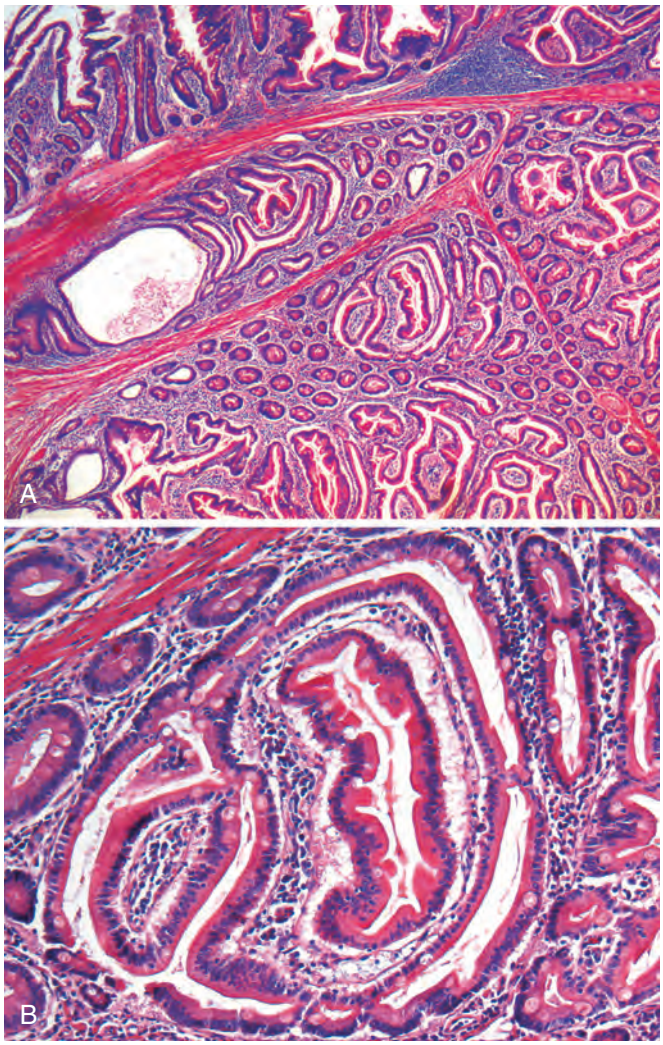


Figure 17.44 Peutz-Jeghers polyp. (A) Polyp surface (top) overlies stroma composed of smooth muscle bundles cutting through the lamina propria. (B) Complex glandular architecture and the presence of smooth muscle are features that distinguish Peutz-Jeghers polyps from juvenile polyps.

The arborization and presence of smooth muscle intermixed with lamina propria are helpful in distinguishing polyps of Peutz-Jeghers syndrome from juvenile polyps.

Clinical Features

Because the morphology of Peutz-Jeghers polyps can overlap with that of sporadic hamartomatous polyps, the presence of multiple polyps in the small intestine, mucocutaneous hyperpigmentation, and a positive family history are critical to the diagnosis. Detection of *STK11* mutations can be helpful diagnostically in patients with polyps who lack mucocutaneous hyperpigmentation. However, the absence of *STK11* mutations does not exclude the diagnosis, as these mutations are not present in all patients.

Neoplastic Polyps

The most common neoplastic polyps are colonic adenomas, which are precursors to the majority of colorectal adenocarcinomas. It is important, however, to realize that any neoplastic mass lesion in the GI tract may produce a polyp. This includes adenocarcinomas, neuroendocrine tumors, stromal tumors, lymphomas, and even metastatic cancers from distant sites.

Adenomas are epithelial neoplasms that range from small, often pedunculated, polyps to large sessile lesions. Adenomas develop in approximately 30% of adults living in the Western world by age 60. Because these polyps are precursors to colorectal adenocarcinoma, surveillance is recommended beginning at age 45 to 50. Individuals at increased risk, including those with a family history of colorectal adenocarcinoma, are typically screened colonoscopically at least 10 years before the youngest age at which a relative was diagnosed. The preferred approach to surveillance varies, but colonoscopy is the most frequent mode. Adenomas are less common in Asia, but their incidence has risen (in parallel with increasing incidence of colorectal adenocarcinoma) as Western diets and lifestyles become more prevalent.

Consistent with their being precursor lesions, the prevalence of colorectal adenomas within populations correlates with that of colorectal adenocarcinoma, and the distributions of adenomas and adenocarcinoma within the colon are similar. Regular surveillance colonoscopy and polyp removal reduce the incidence of colorectal adenocarcinoma. Despite this strong relationship, it must be emphasized that majority of adenomas do not progress to become adenocarcinomas. There are no tools presently available to identify adenomas that are destined to undergo malignant transformation. Most adenomas are clinically silent, with the exception of large polyps that can produce occult bleeding and anemia.

MORPHOLOGY

Colorectal adenomas are characterized by the presence of epithelial dysplasia. Typical adenomas range from 0.3 to 10 cm in diameter and can be pedunculated (Fig. 17.45A) or sessile, with the surface of both types having a texture resembling velvet or a raspberry (Fig. 17.45B). Histologically, **the hallmarks of epithelial dysplasia are nuclear hyperchromasia, elongation, and stratification** (Fig. 17.46C). These changes are most easily appreciated at the surface of the adenoma and are often accompanied by prominent nucleoli, eosinophilic cytoplasm, and a reduction in the number of goblet cells. Pedunculated adenomas have slender fibromuscular stalks (Fig. 17.45C) containing prominent blood vessels derived from the submucosa. The stalk is usually covered by nonneoplastic epithelium, but dysplasia can be present.

Adenomas can be classified as **tubular, tubulovillous,** or **villous** based on their architecture. These categories, however, have little clinical significance in isolation. Tubular adenomas tend to be small, pedunculated polyps composed of rounded, or tubular, glands (Fig. 17.46A). In contrast, villous adenomas, which are often larger and sessile, are covered by slender villi (Fig. 17.46B). Villous adenomas harbor cancers more frequently than tubular adenomas, but this also correlates with the larger size of villous adenomas. As the name implies, tubulovillous adenomas have a mixed architecture.

Sessile serrated lesions, also referred to as sessile serrated adenomas, overlap histologically with hyperplastic polyps but are more commonly found in the right colon. Despite their malignant potential, these polyps **lack typical cytologic features of dysplasia** that are present in other adenomas. Histologically, these lesions can be differentiated from hyperplastic polyps by the presence of serrated architecture throughout the full length of the glands including the crypt base, crypt dilation, and lateral crypt growth (Fig. 17.46D).

Intramucosal carcinoma occurs when dysplastic epithelial cells breach the basement membrane to invade the lamina propria. These may extend into, but not through, the muscularis mucosae. Because functional lymphatic channels are absent in the colonic mucosa, intramucosal carcinomas have little or no metastatic potential and complete polypectomy is generally curative (Fig. 17.47A). Invasion beyond the muscularis mucosae, including into the submucosal stalk of a pedunculated polyp (Fig. 17.47B), constitutes invasive adenocarcinoma and carries a risk of spread to other sites. In such cases, several factors, such as the histologic grade of the invasive component, the presence of vascular or lymphatic invasion, and the distance from the invasive component to the margin of resection are considered in planning further therapy.

Although most colorectal adenomas are benign lesions, a small proportion harbor invasive cancer at the time of detection. **Size is the most important characteristic that correlates with risk of malignancy.** For example, while cancer is extremely rare in adenomas less than 1 cm in diameter, nearly 40% of lesions larger than 4 cm in diameter contain foci of invasive cancer. High-grade dysplasia is also a risk factor for cancer in an individual polyp, but does not confer an increased risk of cancer in other polyps within the same patient.

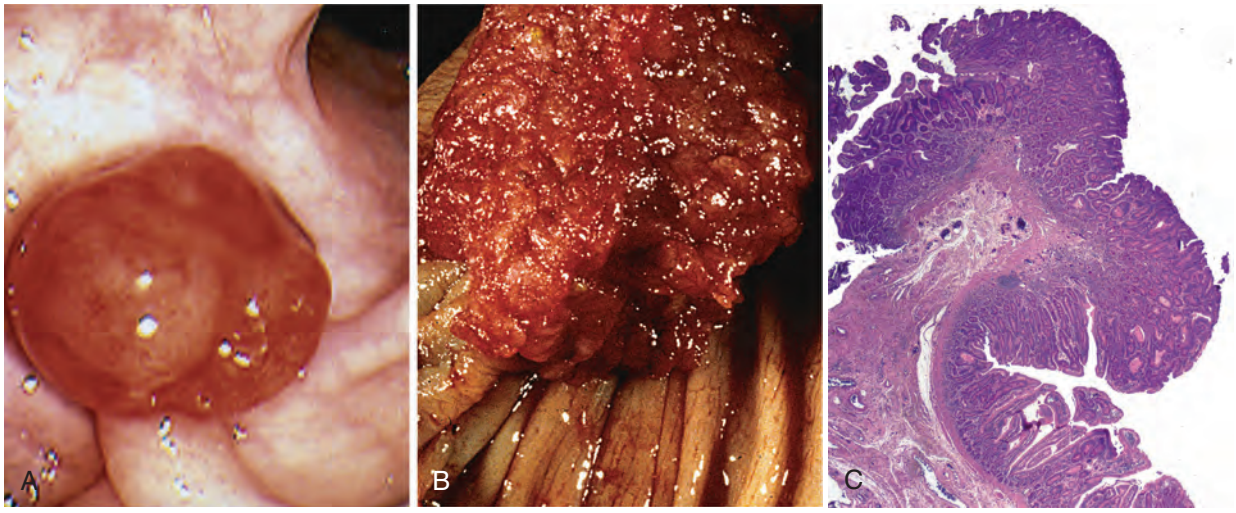


Figure 17.45 Colonic adenomas. (A) Pedunculated adenoma (endoscopic view). (B) Adenoma with a velvety surface. (C) Low-magnification photomicrograph of a pedunculated tubular adenoma.

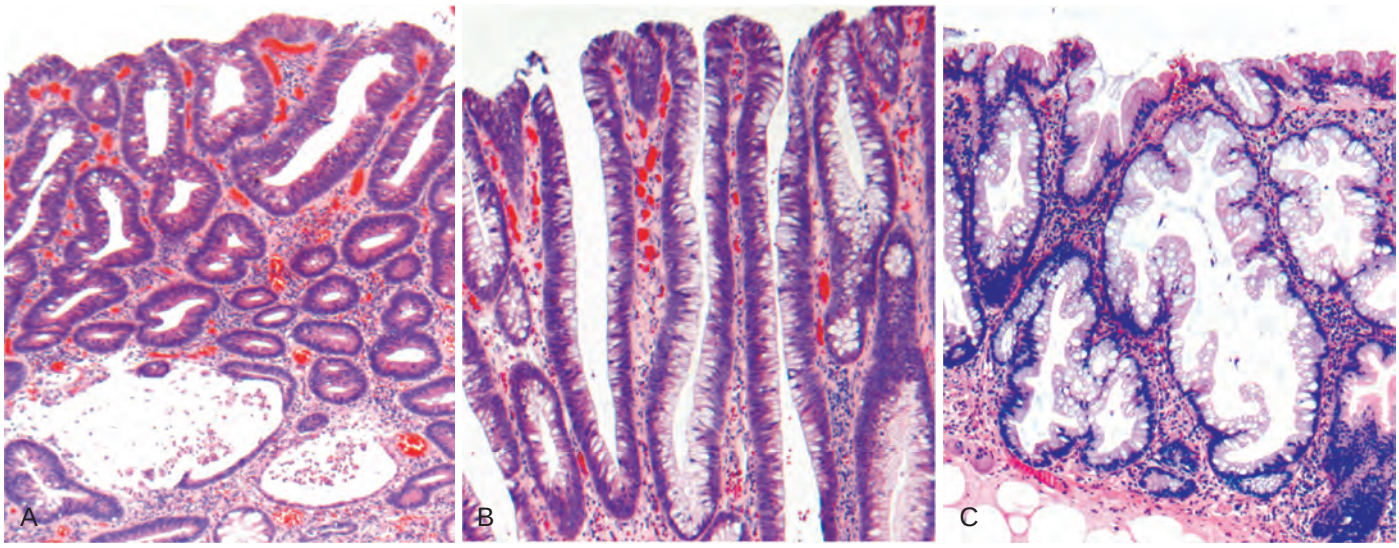


Figure 17.46 Histologic appearance of colonic adenomas. (A) Tubular adenoma with a smooth surface and rounded glands. Active inflammation is occasionally present in adenomas; in this case, crypt dilation and rupture can be seen at the bottom of the field. (B) Villous adenoma with long, slender projections that are reminiscent of small intestinal villi. (C) Sessile serrated adenoma lined by goblet cells without cytologic features of dysplasia. This lesion is distinguished from a hyperplastic polyp by extension of the neoplastic process to the crypts, resulting in lateral growth. Compare to the hyperplastic polyp in Fig. 17.41.

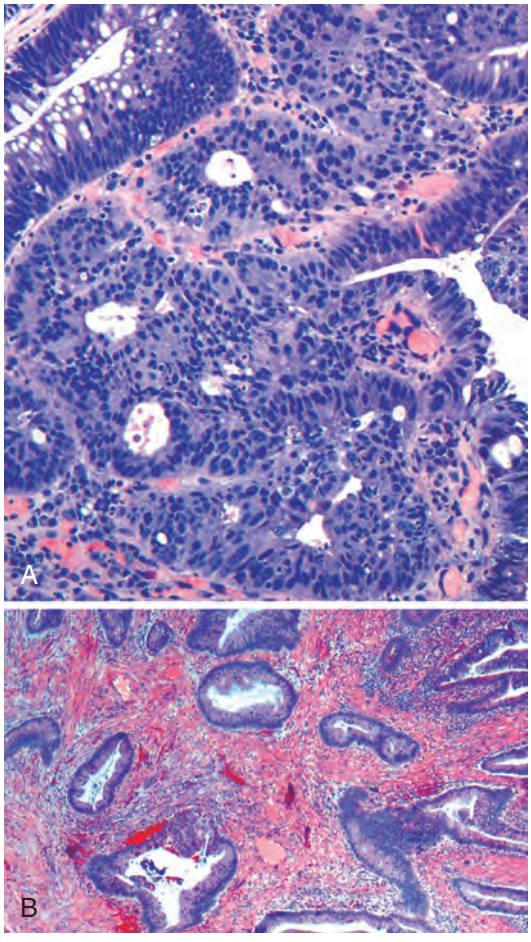


Figure 17.47 Adenoma with carcinoma. (A) Cribriform glands interface directly with the lamina propria without an intervening basement membrane in this intramucosal carcinoma. (B) Invasive adenocarcinoma (left) beneath a villous adenoma (right). Note the desmoplastic response to the invasive components.

FAMILIAL ADENOMATOUS POLYPOSIS

Familial adenomatous polyposis (FAP) is an autosomal dominant disorder in which patients develop numerous colorectal adenomas as teenagers. It is caused by somatic mutations of the adenomatous polyposis coli, or *APC*, gene, which is a key negative regulator of the Wnt signaling pathway (Chapter 7). Approximately 75% of cases are inherited, while the remainder appear to be caused by de novo mutations.

At least 100 polyps are necessary for a diagnosis of classic FAP, but thousands may be present (Fig. 17.48). Except for their numbers, FAP-associated polyps are morphologically indistinguishable from sporadic adenomas. However, flat or depressed adenomas are also prevalent in FAP, and microscopic adenomas, consisting of only one or two dysplastic crypts, are common in otherwise normal-appearing mucosa.

Colorectal adenocarcinoma develops in 100% of untreated FAP patients, often before age 30 and nearly always by age 50. As a result, prophylactic colectomy is the standard of care. Colectomy prevents colorectal cancer, but patients remain at risk for neoplasia at other sites. The ampulla of Vater and the stomach are common extracolonic sites of adenomas in FAP patients.

FAP is associated with a variety of extraintestinal manifestations including congenital hypertrophy of the retinal pigment epithelium, which can be detected at birth and may be an adjunct to early screening. Specific *APC* mutations have been associated with other manifestations of FAP and partly explain variants such as Gardner syndrome and Turcot syndrome (Table 17.10).

Table 17.10 Common Patterns of Sporadic and Familial Colorectal Neoplasia

Etiology	Molecular Defect	Target Genes	Transmission	Predominant Site	Histology
Familial adenomatous polyposis	APC/WNT pathway	*APC*	Autosomal dominant	None	Tubular, villous; typical adenocarcinoma
MYH-associated polyposis	DNA mismatch repair	*MYH*	Autosomal recessive	None	Sessile serrated adenoma; mucinous adenocarcinoma
Hereditary non-polyposis colorectal cancer	DNA mismatch repair	*MSH2, MLH1*	Autosomal dominant	Right side	Sessile serrated adenoma; mucinous adenocarcinoma
Sporadic colon cancer (70%–80%)	APC/WNT pathway	*APC*	None	Left side	Tubular, villous; typical adenocarcinoma
Sporadic colon cancer (10%–15%)	DNA mismatch repair	*MSH2, MLH1*	None	Right side	Sessile serrated adenoma; mucinous adenocarcinoma
Sporadic colon cancer (5%–10%)	Hypermethylation	*MLH1, BRAF*	None	Right side	Sessile serrated adenoma; mucinous adenocarcinoma

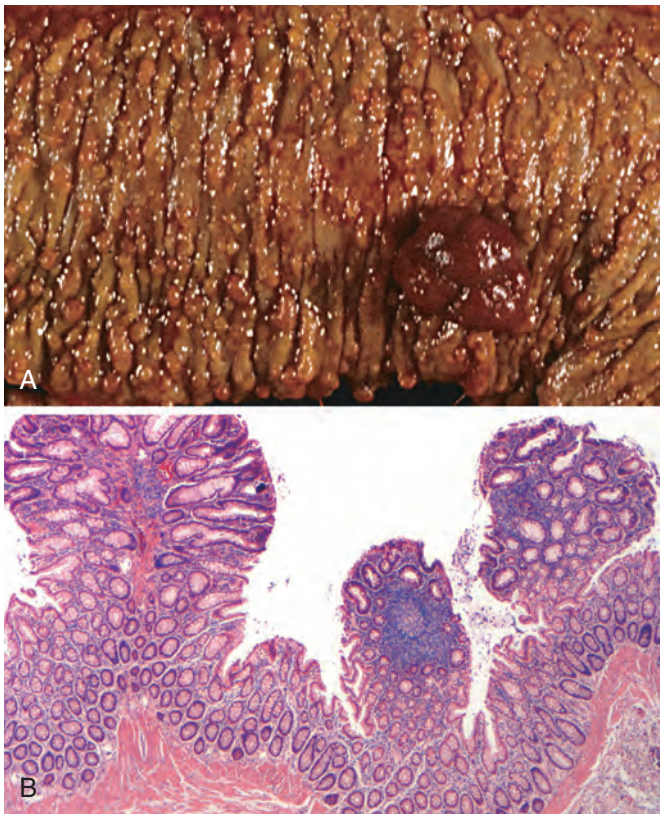


Figure 17.48 Familial adenomatous polyposis. (A) Hundreds of small polyps are present throughout this colon with a dominant polyp (right). (B) Three tubular adenomas are present in this single microscopic field.

Some polyposis patients without *APC* loss have bi-allelic mutations of the base-excision repair gene *MUTYH* (also referred to as *MYH*). This autosomal recessive disorder is termed *MUTYH-associated polyposis, or MAP.* In contrast to FAP, MAP is characterized by fewer than 100 polyps, which appear at later ages. Colon cancer development is also delayed. In addition, sessile serrated adenomas and hyperplastic polyps, often with *KRAS* mutations, are frequently present in those with MAP.

HEREDITARY NON-POLYPOSIS COLORECTAL CANCER (HNPCC)

HNPCC is caused by inherited mutations in mismatch repair genes that encode proteins responsible for the detection, excision, and repair of errors that occur during DNA replication (Chapter 7). HNPCC, also known as Lynch syndrome, was originally described based on familial clustering of cancers at several sites, including the colorectum, endometrium, stomach, ovary, ureters, brain, small bowel, hepatobiliary tract, pancreas, and skin. HNPCC is thought to account for 2% to 4% of all colorectal cancers, making it the most common syndromic form of colon cancer. Colon cancers in HNPCC patients tend to occur at younger ages than sporadic colon cancers and are often located in the right colon (see Table 17.10).

Just as identification of *APC* mutations in FAP provided molecular insight into the pathogenesis of the majority of sporadic colon cancers, unraveling the defects in HNPCC has shed light on the mechanisms responsible for most of the remaining sporadic cases. There are at least five mismatch repair genes, but the majority of patients with HNPCC have mutations in *MSH2* or *MLH1*. Patients with HNPCC inherit one mutant gene and one normal allele. When the second copy is lost through mutation or epigenetic silencing, defects in mismatch repair lead to the accumulation of mutations at rates up to 1000 times higher than normal, mostly in regions containing short repeating sequences referred to as microsatellites. The human genome contains approximately 50,000 to 100,000 microsatellites, which are prone to undergo expansion during DNA replication and represent the most frequent sites of mutations in HNPCC. The consequences of mismatch repair deficiency and the resulting microsatellite instability (MSI) are discussed next in the context of colonic adenocarcinoma.

ADENOCARCINOMA

Adenocarcinoma of the colon is the most common malignancy of the GI tract and is a major cause of morbidity and mortality worldwide. In contrast, the small intestine,

which accounts for 75% of the overall length of the GI tract, is an uncommon site for benign and malignant tumors. Among malignant small intestinal tumors, adenocarcinomas and well-differentiated neuroendocrine tumors have roughly equal incidence, followed by lymphomas and sarcomas.

Epidemiology. **Colorectal adenocarcinoma is responsible for nearly 10% of all cancer deaths worldwide.** Approximately 1.2 million new cases of colorectal adenocarcinoma and 600,000 associated deaths occur each year worldwide. The incidence of these tumors is highest in North America, with the United States accounting for approximately 10% of worldwide cases and cancer deaths. This represents nearly 15% of all cancer-related deaths in the United States, second only to lung cancer. Australia, New Zealand, Europe, and (with changes in lifestyle and diet) Japan also have high incidences of colorectal adenocarcinoma. In contrast, rates are lower in South America, India, Africa, and South Central Asia. Colorectal cancer incidence peaks at 60 to 70 years of age. Fewer than 20% of cases occur before age 50, but recent data suggest that incidence of disease before age 40 is increasing.

The dietary factors most closely associated with increased risk of colorectal cancer are low intake of unabsorbable vegetable fiber and high intake of refined carbohydrates and fat. Although these associations are clear, the mechanistic relationship between diet and risk is poorly understood. It is theorized that reduced fiber content leads to decreased stool bulk and altered composition of the intestinal microbiota. This change may increase synthesis of potentially toxic oxidative by-products of bacterial metabolism, which would be expected to remain in contact with the colonic mucosa for longer periods of time as a result of reduced stool bulk. High fat intake also enhances hepatic synthesis of cholesterol and bile acids, which can be converted into carcinogens by intestinal bacteria.

In addition to dietary modification, pharmacologic chemoprevention is possible. Aspirin and other NSAIDs have a protective effect. This is consistent with studies showing that some NSAIDs cause polyp regression in FAP patients in whom the rectum was left in place after colectomy. It is suspected that this effect is mediated by inhibition of the enzyme COX-2, which is highly expressed in 90% of colorectal carcinomas and 40% to 90% of adenomas. COX-2 is necessary for production of prostaglandin E_2, which promotes epithelial proliferation, particularly after injury. TLR4, which recognizes lipopolysaccharide and is also overexpressed in adenomas and carcinomas, upregulates COX-2 expression.

Pathogenesis

Studies of colorectal carcinogenesis have provided fundamental insight into the general mechanisms of cancer evolution. These were discussed in Chapter 7; concepts that pertain specifically to colorectal carcinogenesis will be reviewed here.

The combination of molecular events that lead to colonic adenocarcinoma is heterogeneous and includes genetic and epigenetic abnormalities. **At least two genetic pathways have been described: the APC/β-catenin pathway, which is activated in the classic adenoma-carcinoma sequence, and the MSI pathway, which is associated with defects in DNA mismatch repair and accumulation of mutations in microsatellite repeat regions of the genome** (Table 17.10). Both pathways involve the stepwise accumulation of multiple mutations, but differ in the genes involved and the mechanisms by which mutations occur. Epigenetic events, the most common of which is methylation-induced gene silencing, may enhance progression along either pathway.

- *The classic adenoma-carcinoma sequence accounts for up to 80% of sporadic cancers and typically includes mutation of APC early in the neoplastic process* (Fig. 17.49). Both copies of the *APC* gene must be functionally inactivated, either by mutation or by epigenetic events, for adenomas to develop. The APC protein normally binds to and promotes degradation of β-catenin, a component of the Wnt signaling pathway (Chapter 7). With loss of APC function, β-catenin accumulates and translocates to the nucleus, where it forms a complex with the DNA-binding factor TCF and activates the transcription of genes, including *MYC* and cyclin D1, that promote proliferation. The critical role of β-catenin in this pathway is demonstrated by the fact that many colon cancers without *APC* mutations harbor β-catenin mutations that prevent APC-dependent degradation, thereby allowing β-catenin to accumulate and activate Wnt signaling. Additional mutations accumulate including activating mutations in *KRAS* that promote growth and prevent apoptosis. *KRAS* mutations are present in fewer than 10% of adenomas less than 1 cm in diameter but are found in 50% of adenomas greater than 1 cm in diameter and in 50% of invasive adenocarcinomas, indicating that this is a late event in neoplastic progression. Carcinogenesis is also associated with mutations in tumor suppressor genes such as those encoding SMAD2 and SMAD4, which are effectors of TGF-β signaling. Because TGF-β signaling normally inhibits the cell cycle, loss of these genes may allow unrestrained cell growth. The tumor suppressor gene *TP53* is mutated in 70% to 80% of colon cancers but is uncommonly affected in adenomas, suggesting that *TP53* mutations also occur at later stages of tumor progression. Loss of function of *TP53* and other tumor suppressor genes is often caused by chromosomal deletions, demonstrating the chromosomal instability that is a hallmark of the APC/β-catenin pathway. Alternatively, tumor suppressor genes may be silenced by methylation of CpG-rich zones, or CpG islands, within the 5′ region that frequently includes the promoter and transcriptional start site of some genes. Telomerase expression also increases as lesions become more advanced.
- *In patients with DNA mismatch repair deficiency, mutations accumulate in microsatellite repeats, a condition referred to as microsatellite instability (MSI).* These are referred to as MSI high, or MSI-H, tumors. Some microsatellite sequences are located in the coding or promoter regions of genes involved in regulation of cell growth, such as the type II TGF-β receptor and the pro-apoptotic protein BAX (Fig. 17.50). Because TGF-β inhibits colonic epithelial cell proliferation, mutation of *TGFBR2,* which encodes the type II TGF-β receptor, can contribute to uncontrolled cell growth, while loss of BAX may enhance the survival of genetically abnormal clones.

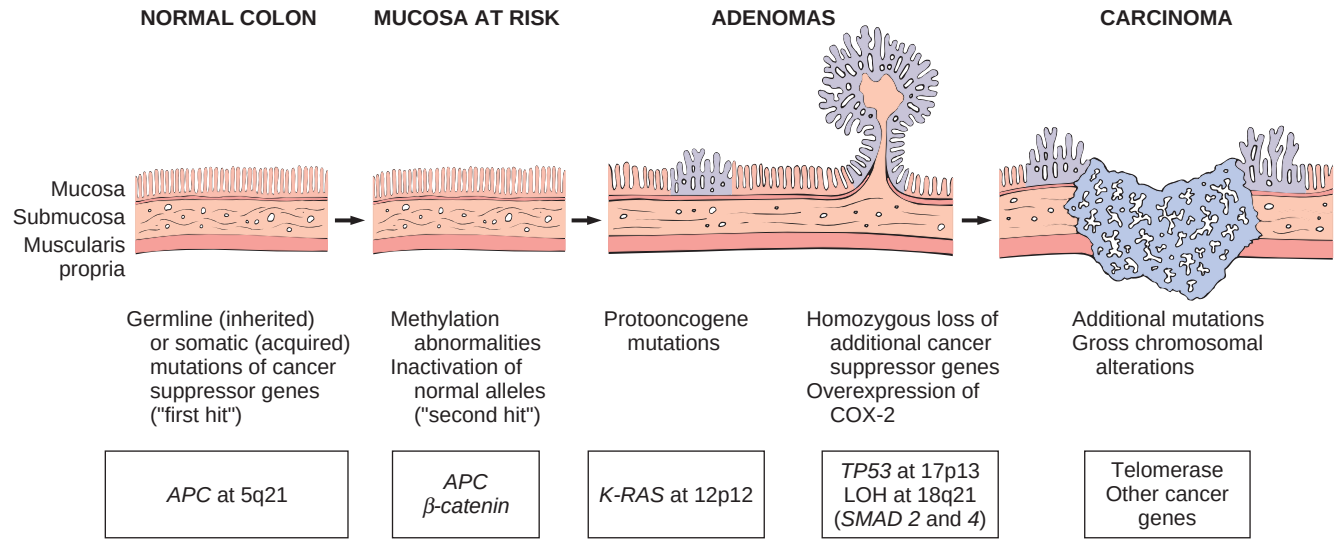


Figure 17.49 Morphologic and molecular changes in the adenoma-carcinoma sequence. Loss of one normal copy of the tumor suppressor gene *APC* occurs early. Individuals born with one mutant allele are therefore at increased risk of developing colon cancer. Alternatively, inactivation of *APC* in colonic epithelium may occur later in life. This is the "first hit" according to the Knudson hypothesis (Chapter 7). The loss of the intact second copy of *APC* follows ("second hit"). Other changes, including mutation of *KRAS*, losses at 18q21 involving *SMAD2* and *SMAD4*, and inactivation of the tumor suppressor gene *TP53*, lead to the emergence of carcinoma. Although there seems to be a temporal sequence of changes, the accumulation of mutations, rather than their occurrence in a specific order, is most critical.

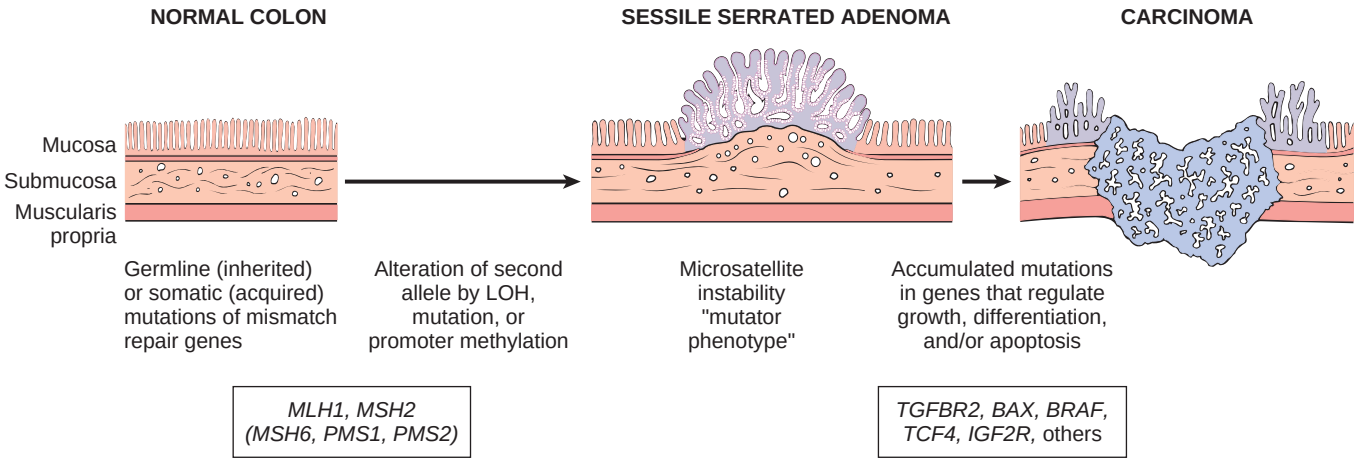


Figure 17.50 Morphologic and molecular changes in the mismatch repair pathway of colon carcinogenesis. Defects in mismatch repair genes result in microsatellite instability and permit accumulation of mutations in numerous genes. If these mutations affect genes involved in cell survival and proliferation, cancer may develop.

- *A subset of microsatellite unstable colon cancers without mutations in DNA mismatch repair enzymes demonstrate the CpG island hypermethylation phenotype (CIMP).* In these tumors, the *MLH1* promoter region is typically hypermethylated, thereby reducing MLH1 expression and repair function. Activating mutations in the oncogene *BRAF* are common in these cancers. In contrast, *KRAS* and *TP53* are not typically mutated. The combination of MSI, *BRAF* mutation, and methylation of specific targets, such as *MLH1,* is the signature of this pathway of carcinogenesis.
- *A small group of colon cancers display increased CpG island methylation in the absence of MSI.* Many of these tumors harbor *KRAS* mutations, but *TP53* and *BRAF* mutations are uncommon. In contrast, *TP53* mutations are common in colon cancers that do not display a CpG island methylator phenotype.

While morphology cannot reliably define the underlying molecular events that lead to carcinogenesis, certain correlations have been associated with mismatch repair deficiency and MSI. These molecular alterations are common in sessile serrated lesions and cancers that arise from them. In addition, invasive carcinomas with MSI often have prominent mucinous differentiation and peritumoral lymphocytic infiltrates. These tumors as well as those with a CpG island hypermethylation phenotype are frequently located in the right colon. MSI can be identified by the absence of immunohistochemical staining for mismatch repair proteins or by molecular genetic analysis of microsatellite sequences. It is important to identify patients

with HNPCC because of the implications for genetic counseling, the elevated risk of a second malignancy of the colon or other organs, and, in some settings, differences in prognosis and therapy.

MORPHOLOGY

Overall, adenocarcinomas are distributed approximately equally over the entire length of the colon. Tumors in the **proximal colon often grow as polypoid, exophytic masses** that extend along one wall of the large-caliber cecum and ascending colon; these tumors rarely cause obstruction. In contrast, **carcinomas in the distal colon tend to be annular lesions that produce "napkin-ring" constrictions and luminal narrowing** (Fig. 17.51), sometimes to the point of obstruction. Both forms grow into the bowel wall over time. The general microscopic characteristics of right- and left-sided colonic adenocarcinomas are similar. Most tumors are composed of tall columnar cells that resemble dysplastic epithelium found in adenomas (Fig. 17.52A). The invasive component of these tumors elicits a strong stromal desmoplastic response, which is responsible for their characteristic firm consistency. Some poorly differentiated tumors form few glands (Fig. 17.52B). Others may produce abundant mucin that accumulates within the intestinal wall; these are associated with poor prognosis. Uncommonly, tumors may be composed of signet-ring cells similar to those in gastric cancer (Fig. 17.52C).

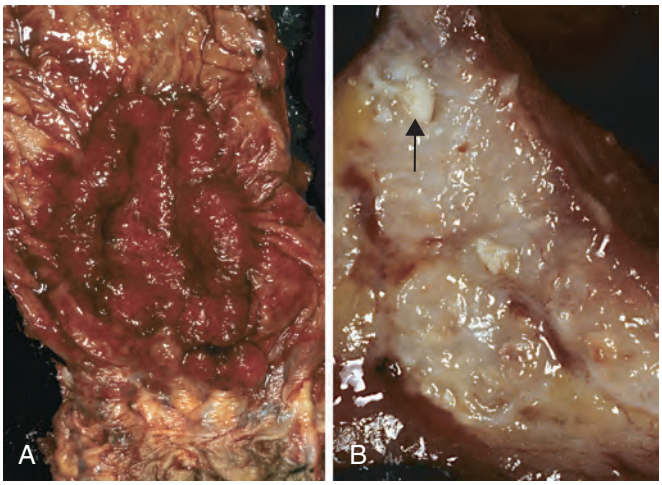


Figure 17.51 Colorectal carcinoma. (A) Circumferential, ulcerated rectal cancer. Note the anal mucosa at the bottom of the image. (B) Cancer of the sigmoid colon that has invaded through the muscularis propria and is present within subserosal adipose tissue (left). Areas of chalky necrosis are present within the colon wall *(arrow)*.

Clinical Features

The availability of endoscopic screening combined with the knowledge that most carcinomas arise within adenomas presents a unique opportunity for cancer prevention. Unfortunately, colorectal cancers develop insidiously and may go undetected for long periods. Cecal and other right-sided colon cancers are most often called to clinical attention by the appearance of fatigue and weakness due to iron deficiency anemia. Thus, it is a clinical maxim that the underlying cause of iron deficiency anemia in an older man or postmenopausal woman is GI cancer until proven otherwise. Left-sided colorectal adenocarcinomas may produce occult bleeding, changes in bowel habits, or cramping and left lower quadrant discomfort.

Although poorly differentiated and mucinous histologies are associated with worse prognosis, **the two most important prognostic factors are depth of invasion and the presence of lymph node metastases.** Invasion into the muscularis propria significantly reduces the probability of survival, which is decreased further by the presence of lymph node metastases (Fig. 17.53A). Metastases may involve regional lymph nodes, lungs (Fig. 17.53B), and bones, but as a result of portal drainage of the colon, the liver is the most common site of distant metastatic lesions (Fig. 17.53C). The rectum does not drain via the portal circulation; hence anal carcinomas that metastasize often circumvent the liver.

These prognostic factors were originally recognized by Dukes and Kirklin and form the core of the TNM (tumor-nodes-metastasis) system (Table 17.11). The TNM classification is used to define tumor stage (Table 17.12). Regardless of stage, it must be remembered that some patients with small numbers of metastases do well for years following resection of distant tumor nodules.

Five-year survival rates vary widely worldwide. The overall 5-year survival rate in the United States is 65% and

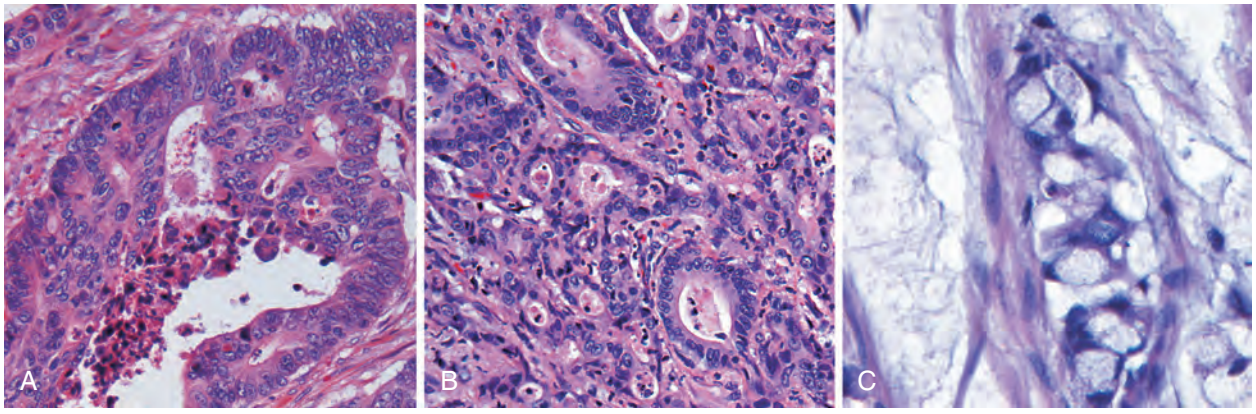


Figure 17.52 Histologic appearance of colorectal carcinoma. (A) Well-differentiated adenocarcinoma. Note the elongated, hyperchromatic nuclei. Necrotic debris, present in the gland lumen, is typical. (B) Poorly differentiated adenocarcinoma forms a few glands but is largely composed of infiltrating nests of tumor cells. (C) Mucinous adenocarcinoma with signet-ring cells and extracellular mucin pools.

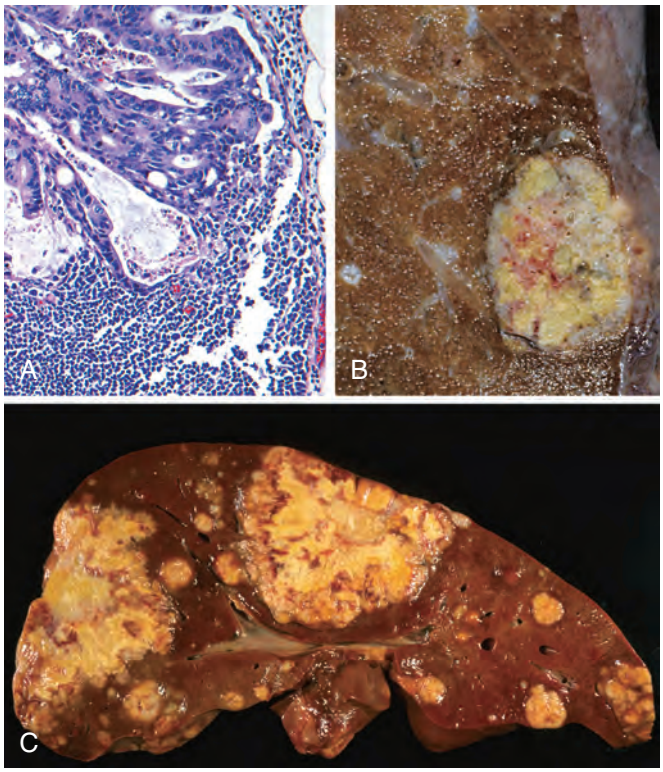


Figure 17.53 Metastatic colorectal carcinoma. (A) Lymph node metastasis. Note the glandular structures within the subcapsular sinus. (B) Solitary subpleural nodule of colorectal carcinoma metastatic to the lung. (C) Liver containing two large and many smaller metastases. Note the central necrosis within metastases.

ranges from 90% to 40% depending on stage. Survival rates in Europe, Japan, and Australia are similar, ranging from 60% (Switzerland, Japan) to 40% (Poland). Overall survival rates are somewhat lower in other countries, such as China, India, the Philippines, and Thailand (30% to 42%). In contrast, the 5-year survival rate in Gambia is only 4%, a measure of the challenges of delivering healthcare to under-resourced parts of the world.

Table 17.11 American Joint Committee on Cancer (AJCC) TNM Classification of Colorectal Carcinoma

TNM	
Tumor	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ, intramucosal carcinoma
T1	Tumor invades the submucosa
T2	Tumor invades the muscularis propria
T3	Tumor invades through the muscularis propria into pericolorectal tissues
T4	Tumor invades visceral peritoneum or invades or adheres to adjacent organ or structure
T4a	Tumor invades visceral peritoneum
T4b	Tumor directly invades or adheres to adjacent organs or structures
Regional Lymph Nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in one to three regional lymph nodes
N1a	Metastasis in one regional lymph node
N1b	Metastasis in two to three regional lymph nodes
N1c	Tumor deposit(s) in subserosa or in non-peritonealized pericolic or perirectal soft tissue without regional nodal metastasis
N2	Metastasis in four or more regional lymph nodes
N2a	Metastasis in four to six regional lymph nodes
N2b	Metastasis in seven or more regional lymph nodes
Distant Metastasis	
M0	No distant metastasis
M1	Distant metastasis to one or more distant sites or organs or peritoneal metastasis
M1a	Metastasis to one site or organ without peritoneal metastasis
M1b	Metastases to two or more sites or organs without peritoneal metastasis
M1c	Metastasis to the peritoneal surface, alone or with other site or organ metastases

Used with the permission of the American College of Surgeons. Amin MB, Edge SB, Greene FL, et al, editors: *AJCC Cancer Staging Manual,* ed 8, New York, Springer, 2017.

KEY CONCEPTS

BENIGN AND MALIGNANT PROLIFERATIVE LESIONS OF THE COLON

- Intestinal polyps can be classified as nonneoplastic or neoplastic. The nonneoplastic polyps can be further subclassified as hyperplastic, inflammatory, or hamartomatous.
- Hyperplastic polyps are benign epithelial proliferations most commonly found in the left colon and rectum. They have no malignant potential and must be distinguished from sessile serrated polyps.
- Inflammatory polyps form as a result of chronic cycles of injury and healing.
- Hamartomatous polyps occur sporadically or as a part of genetic diseases. The latter include juvenile polyposis and Peutz-Jeghers syndrome, which are associated with increased risk of malignancy.
- Benign epithelial neoplastic polyps of the intestines are termed adenomas. The hallmark of these lesions, which are the precursors of colonic adenocarcinomas, is cytologic dysplasia.
- In contrast to traditional adenomas, sessile serrated polyps lack cytologic dysplasia and share morphologic features with hyperplastic polyps.
- Familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC) are the most common forms of familial colon cancer.
- FAP is caused by *APC* mutations. Patients typically have more than 100 adenomas, and all patients develop colon cancer before 30 years of age.
- HNPCC is caused by mutations in DNA mismatch repair enzymes. HNPCC patients have far fewer polyps and develop cancer at older ages than FAP patients but younger ages than those with sporadic colon cancer.
- FAP and HNPCC typify distinct pathways of neoplastic transformation and progression that are also involved in the majority of sporadic colon cancers. A third pathway characterized by

CpG island hypermethylation underlies most remaining colon cancers.

- Nearly all colonic cancers are adenocarcinomas. The two most important prognostic factors are depth of invasion and the presence or absence of lymph node metastases.

TUMORS OF THE ANAL CANAL

The anal canal can be divided into thirds. The upper zone is lined by columnar rectal epithelium; the middle third by transitional epithelium; and the lower third by stratified squamous epithelium. Carcinomas of the anal canal may have typical glandular or squamous patterns of differentiation, recapitulating the normal epithelium of the upper and lower thirds, respectively (Fig. 17.54A). An additional differentiation pattern of squamous cell carcinomas, termed basaloid, is present in tumors populated by immature cells derived from the basal layer of transitional epithelium (Fig. 17.54B). These tumors are classified as squamous, but the archaic term cloacogenic carcinoma is often applied when the entire tumor displays a basaloid pattern. Alternatively, basaloid differentiation may be mixed with mucinous differentiation. Squamous cell carcinoma of the anal canal is frequently associated with infection by high-risk strains of human papilloma virus (HPV), whereas infections with low-risk HPV strains are associated with condyloma acuminatum (Fig. 17.54C). The major risk factor for these HPV associated lesions is anal sex.

Table 17.12 Colorectal Cancer Staging System

Stage	T	N	M
0	Tis	N0	M0
I	T1–T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1–T2 T1	N1/N1c N2a	M0 M0
IIIB	T3–T4a T2–T3 T1–T2	N1/N1c N2a N2b	M0 M0 M0
IIIC	T4a T3–T4a T4b	N2a N2b N1/N2	M0 M0 M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b
IVC	Any T	Any N	M1c

Used with the permission of the American College of Surgeons. Amin MB, Edge SB, Greene FL, et al, editors: *AJCC Cancer Staging Manual*, ed 8, New York, Springer, 2017.

HEMORRHOIDS

Hemorrhoids affect about 5% of the general population and develop secondary to persistently elevated venous pressure within the hemorrhoidal plexus. The most frequent predisposing influences are straining at defecation, e.g., in constipation, and venous stasis of pregnancy. Hemorrhoids may also develop secondary to portal hypertension. The pathogenesis of hemorrhoids (anal varices) in portal hypertension is similar to that of esophageal varices; anal varices are both more common and much less serious. Anal and perianal varices connect portal and caval venous systems, thereby relieving the venous hypertension.

Hemorrhoids often present with pain and rectal bleeding, particularly bright red blood seen on toilet tissue. Except for pregnant women, hemorrhoids are rarely encountered in persons younger than age 30. Hemorrhoidal bleeding is not generally a medical emergency and can be treated by sclerotherapy, rubber band ligation, or infrared coagulation. Extensive or severe internal or external hemorrhoids may be removed surgically by hemorrhoidectomy.

ACUTE APPENDICITIS

The appendix is a normal true diverticulum of the cecum that is prone to acute and chronic inflammation. Acute

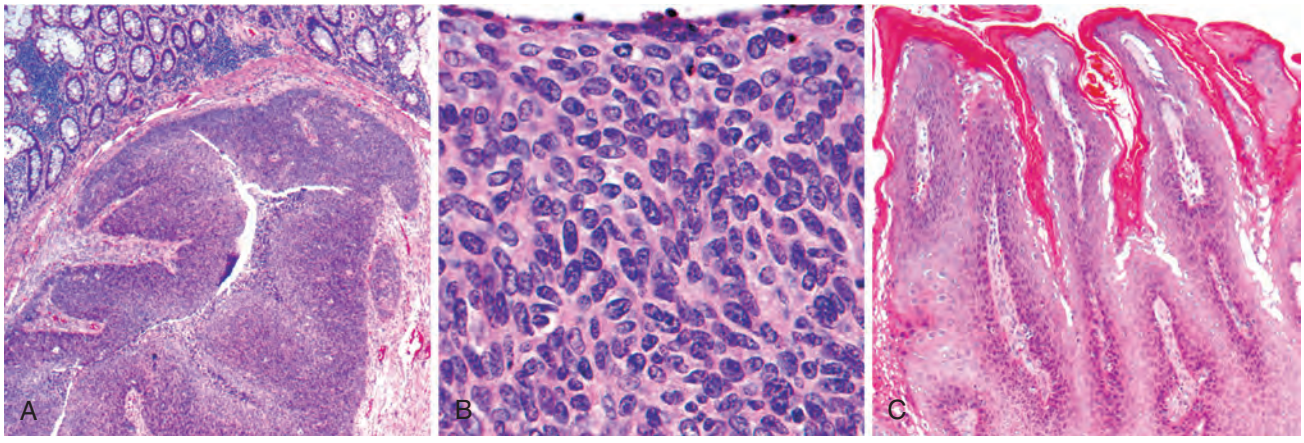


Figure 17.54 Anal tumors. (A) This anal squamous cell carcinoma demonstrates a multilayered organization reminiscent of squamous mucosa. The adjacent rectal mucosa is normal. (B) This anal squamous cell carcinoma with basaloid features is composed of hyperchromatic cells that resemble the basal layer of normal squamous mucosa. (C) Condyloma acuminatum with verrucous architecture.

appendicitis is most common in adolescents and young adults, with a lifetime risk of 7%; males are affected slightly more often than females. Despite the prevalence of acute appendicitis, the diagnosis can be difficult to confirm preoperatively and may be confused with mesenteric lymphadenitis (often secondary to unrecognized *Yersinia* infection or viral enterocolitis), acute salpingitis, ectopic pregnancy, mittelschmerz pain (caused by minor pelvic bleeding at the time of ovulation), and Meckel diverticulitis.

Pathogenesis

Acute appendicitis is thought to be initiated by progressive increases in intraluminal pressure that compromise venous outflow. In 50% to 80% of cases, acute appendicitis is associated with overt luminal obstruction, usually caused by a small stone-like mass of stool, or fecalith, or, less commonly, a gallstone, tumor, or mass of worms (oxyuriasis vermicularis). Peri-appendiceal lymphoid hyperplasia, e.g., after viral infection, and the resulting appendiceal compression has also been proposed as a mechanism of venous and luminal obstruction. Stasis of luminal contents allows bacterial proliferation that leads to ischemia and inflammation.

MORPHOLOGY

In early acute appendicitis, subserosal vessels are congested, and there is a modest perivascular neutrophilic infiltrate within all layers of the wall. The inflammatory reaction transforms the normal glistening serosa into a dull, granular, erythematous surface. Although mucosal neutrophils and focal superficial ulceration are often present, these are not specific markers of acute appendicitis. Diagnosis of acute appendicitis requires **neutrophilic infiltration of the muscularis propria.** In more severe cases a prominent neutrophilic exudate generates a serosal fibrinopurulent reaction. As the process continues, focal abscesses may form within the wall (acute suppurative appendicitis). Further compromise of appendiceal vessels can result in ischemia, hemorrhagic ulceration, and gangrenous necrosis that extends to the serosa, creating acute gangrenous appendicitis, which can be followed by rupture and suppurative peritonitis.

Clinical Features

Typically, early acute appendicitis produces periumbilical pain that ultimately localizes to the right lower quadrant, followed by nausea, vomiting, low-grade fever, and a mildly elevated peripheral white cell count. A classic physical finding is the *McBurney sign,* deep tenderness located two-thirds of the distance from the umbilicus to the right anterior superior iliac spine (McBurney point).

Regrettably, classic signs and symptoms of acute appendicitis are often absent. In some cases, a retrocecal appendix may generate right flank or pelvic pain, while a malrotated colon may give rise to appendiceal pain in the left upper quadrant. As with other causes of acute inflammation, there is neutrophilic leukocytosis. In some cases the peripheral leukocytosis may be minimal or, alternatively, so great that other causes are considered. The diagnosis of acute appendicitis in young children and the very old is particularly problematic, since other causes of abdominal emergencies are prevalent in these populations, and the very young and old are also more likely to have atypical clinical presentations.

Given these diagnostic challenges, it should be no surprise that even highly skilled surgeons remove normal appendices. This is preferred to delayed resection of a diseased appendix, given the significant morbidity and mortality associated with appendiceal perforation. Other complications of appendicitis include pyelophlebitis, portal venous thrombosis, liver abscess, and bacteremia.

TUMORS OF THE APPENDIX

Several tumors occur in the appendix. The most common is the well-differentiated neuroendocrine (carcinoid) tumor. It is usually discovered incidentally at the time of surgery or pathologic examination of a resected appendix. This neoplasm, which is almost always benign, most frequently forms a solid bulbous swelling at the distal tip of the appendix, where it can reach 2 to 3 cm in diameter. Although intramural and transmural extension may be evident, nodal metastases are very infrequent, and distant spread is exceptionally rare.

Conventional adenomas, mucin-producing and non–mucin-producing adenocarcinomas, and the uncommon adenocarcinoid tumor also occur in the appendix and may cause obstruction and enlargement that mimics acute appendicitis.

Mucinous neoplasms of the appendix occur in adults, most often in the sixth decade of life. They are characterized by mucinous epithelial proliferation, extracellular mucin, and pushing tumor margins; an infiltrating invasive pattern classifies the tumor as a mucin-producing adenocarcinoma.

Mucin can dissect through the wall to the peritoneal surface, causing appendiceal rupture. Tumors are classified as low-grade or high-grade appendiceal mucinous neoplasms (LAMN, HAMN) on the basis of histological features. *GNAS,* which encodes Gαs, is frequently mutated and thought to contribute to excessive mucin production. *KRAS* mutations are also common. Genes involved in colon cancer, such as *APC, TP53,* and *SMAD4,* are typically intact in LAMN but may be mutated in HAMN. When confined to the appendix, prognosis of LAMN is excellent. Outcomes are more variable in HAMN. Peritoneal spread of LAMN, HAMN, or mucin-producing adenocarcinomas (of the appendix or colon) can lead to pseudomyxoma peritonei, in which tenacious, semisolid mucin fills the abdomen. This disseminated intraperitoneal disease may be held in check for years by repeated surgical debulking but, in most instances, follows an inexorably fatal course.

KEY CONCEPTS

HEMORRHOIDS AND APPENDICEAL DISEASE

- Hemorrhoids are collateral vessels that develop secondary to persistently elevated venous pressure within the hemorrhoidal plexus. They also occur in portal hypertension.
- Acute appendicitis is most common in children and adolescents. It is thought to be initiated by increased intraluminal pressure and compromised venous outflow.
- The most common tumor of the appendix is the well-differentiated neuroendocrine (carcinoid) tumor, which is almost always benign.
- Peritoneal dissemination of mucinous neoplasms can cause pseudomyxoma peritonei.

Peritoneal Cavity

The peritoneal cavity houses the abdominal viscera and is lined by a single layer of mesothelial cells; these cover the visceral and parietal surfaces and are supported by a thin layer of connective tissue to form the peritoneum. Here we discuss inflammatory, infectious, and neoplastic disorders of the peritoneal cavity and retroperitoneal space.

INFLAMMATORY DISEASE

Peritonitis may result from bacterial invasion or chemical irritation and is most often due to:

- Leakage of bile or pancreatic enzymes, which produces *sterile peritonitis.*
- *Perforation or rupture of the biliary system* that evokes a highly irritating peritonitis, usually complicated by bacterial superinfection
- *Acute hemorrhagic pancreatitis* (Chapter 19), which is associated with leakage of pancreatic enzymes and fat necrosis. Damage to the bowel wall may allow bacteria to spread to the peritoneal cavity.
- *Foreign material,* including that introduced surgically (e.g., talc and sutures), may induce foreign body–type granulomas and fibrous scarring.
- *Endometriosis,* which causes hemorrhage into the peritoneal cavity, that acts as an irritant.
- *Ruptured dermoid cysts,* which release keratins and induce an intense granulomatous reaction.
- *Perforation* of abdominal viscera, which results in leakage of luminal materials that can lead to infection and immune activation.

Peritoneal Infection

Bacterial peritonitis occurs when bacteria from the GI lumen are released into the abdominal cavity, most commonly following intestinal perforation. *E. coli*, streptococci, *S. aureus,* enterococci, and *C. perfringens* are implicated most often.

Spontaneous bacterial peritonitis develops in the absence of an obvious source of contamination. It is seen most often in patients with cirrhosis and ascites and less frequently in children with nephrotic syndrome. Diagnosis is based on the presence of neutrophils in ascitic fluid and positive bacterial cultures; *E. coli,* streptococci, and *Klebsiella* species are the most frequently involved organisms.

MORPHOLOGY

The cellular inflammatory response is composed primarily of dense collections of neutrophils and fibrinopurulent debris that coat the viscera and abdominal wall. Serous or slightly turbid fluid begins to accumulate and becomes suppurative as infection progresses. Subhepatic and subdiaphragmatic abscesses may be formed. With the exception of tuberculous peritonitis, the reaction usually remains superficial.

Sclerosing Retroperitonitis

Sclerosing retroperitonitis, also known as idiopathic retroperitoneal fibrosis or Ormond disease, is characterized by dense fibrosis that may extend to involve the mesentery. Although the cause of sclerosing retroperitonitis is unknown, many cases are now thought to fall within the spectrum of IgG4-related sclerosing disease, an immunoinflammatory disorder that can lead to fibrosis in a wide variety of tissues. Because the process frequently compresses the ureters, this entity is described in more detail in Chapters 6 and 21.

TUMORS

Primary malignant tumors arising from the peritoneal lining, mesotheliomas, are similar to tumors of the pleura and pericardium. Peritoneal mesotheliomas are almost always associated with high levels of asbestos exposure. Rarely, primary benign and malignant soft tissue tumors may also develop within the peritoneum and retroperitoneum. The most common of these is desmoplastic small round cell tumor. This is an aggressive tumor that occurs in children and young adults. It bears a strong morphologic resemblance to Ewing sarcoma (Chapter 26) and like Ewing sarcoma is also characterized by a reciprocal translocation, t(11;22) (p13;q12), that produces a *EWS* and *WT1* fusion gene.

Secondary tumors may involve the peritoneum by direct spread or metastatic seeding, resulting in peritoneal carcinomatosis. As discussed earlier, mucinous carcinomas, particularly those of the appendix, may cause pseudomyxoma peritonei.

SUGGESTED READINGS

Peery AF, Crockett SD, Murphy CC et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018, *Gastroenterology* 156:254–272, e11, 2019. *[Current analysis of trends in gastrointestinal disease epidemiology and costs].*

Congenital Abnormalities

Duggan CP, Jaksic T: Pediatric intestinal failure, *N Engl J Med* 377:666–675, 2017. *[Review of short bowel syndrome and intestinal failure].*

Heuckeroth RO: Hirschsprung disease—integrating basic science and clinical medicine to improve outcomes, *Nat Rev Gastroenterol Hepatol* 15:152–167, 2018. *[Review of mechanisms and emerging therapies in Hirschsprung disease].*

Jobson M, Hall NJ: Contemporary management of pyloric stenosis, *Semin Pediatr Surg* 25:219–224, 2016. *[Review focusing on mechanisms and treatment of pyloric stenosis].*

Esophagus

Esophageal Obstruction

Pandolfino JE, Gawron AJ: Achalasia: a systematic review, *JAMA* 313:1841–1852, 2015. *[Comprehensive review of achalasia].*

Esophagitis

Dellon ES, Hirano I: Epidemiology and natural history of eosinophilic esophagitis, *Gastroenterology* 154:319–332, e3, 2018. *[Comprehensive review of eosinophilic esophagitis].*

Gyawali CP, Fass R: Management of gastroesophageal reflux disease, *Gastroenterology* 154:302–318, 2018. *[Detailed approach to patients with esophageal reflux].*

Barrett Esophagus

McDonald SA, Graham TA, Lavery DL et al: The Barrett's gland in phenotype space, *Cell Mol Gastroenterol Hepatol* 1:41–54, 2015. *[Detailed review of mechanisms underlying Barrett metaplasia].*

Moyes LH, Oien KA, Foulis AK et al: Prevalent low-grade dysplasia: the strongest predictor of malignant progression in Barrett's columnar-lined oesophagus, *Gut* 65:360–361, 2016. *[Large clinical analysis of the risk of dysplasia and cancer in Barrett esophagus]*.

Esophageal Tumors

Coleman HG, Xie SH, Lagergren J: The epidemiology of esophageal adenocarcinoma, *Gastroenterology* 154:390–405, 2018. *[Trends in esophageal adenocarcinoma epidemiology]*.

Reid BJ: Genomics, endoscopy, and control of gastroesophageal cancers: a perspective, *Cell Mol Gastroenterol Hepatol* 3:359–366, 2017. *[Analysis of current understanding and approaches in esophageal adenocarcinoma]*.

Stomach

Gastropathy and Acute Gastritis

Barletta JF, Bruno JJ, Buckley MS et al: Stress ulcer prophylaxis, *Crit Care Med* 44:1395–1405, 2016. *[Detailed literature review with recommendations regarding stress ulcer prophylaxis]*.

Goldstein JL, Scheiman JM, Fort JG et al: Aspirin use in secondary cardiovascular protection and the development of aspirin-associated erosions and ulcers, *J Cardiovasc Pharmacol* 68:121–126, 2016. *[Analysis of non-steroidal anti-inflammatory drugs and proton pump inhibitors in gastric ulcers]*.

Lanas A, Chan FKL: Peptic ulcer disease, *Lancet* 390:613–624, 2017. *[Analysis of peptic ulcers not related to* Helicobacter pylori *or non-steroidal anti-inflammatory drug use]*.

Chronic Gastritis

Anderson WF, Rabkin CS, Turner N et al: The changing face of noncardia gastric cancer incidence among US Non-Hispanic whites, *J Natl Cancer Inst* 2018. doi:10.1093/jnci/djx262. *[Epidemiologic analysis of non-cardia gastric cancer incidence and race]*.

Bockerstett KA, DiPaolo RJ: Regulation of gastric carcinogenesis by inflammatory cytokines, *Cell Mol Gastroenterol Hepatol* 4:47–53, 2017. *[Review of relationship between inflammatory cytokine responses and gastric carcinogenesis]*.

Crowe SE: *Helicobacter pylori* infection, *N Engl J Med* 380:1158–1165, 2019. *[Overarching review of* Helicobacter pylori *infection and gastritis pathogenesis]*.

Neumann WL, Coss E, Rugge M et al: Autoimmune atrophic gastritis-pathogenesis, pathology and management, *Nat Rev Gastroenterol Hepatol* 10:529, 2013. *[All-encompassing review of atrophic gastritis]*.

Hypertrophic Gastropathies

Fiske WH, Tanksley J, Nam KT et al: Efficacy of cetuximab in the treatment of Menetrier's disease, *Sci Transl Med* 1:8ra18, 2009. *[Analysis of series of patients treated for Ménétrier disease]*.

Simmons LH, Guimaraes AR, Zukerberg LR: Case records of the Massachusetts General Hospital. Case 6-2013. A 54-year-old man with recurrent diarrhea, *N Engl J Med* 368:757, 2013. *[Case report and discussion of Zollinger-Ellison syndrome]*.

Gastric Polyps and Tumors

Bannon AE, Klug LR, Corless CL et al: Using molecular diagnostic testing to personalize the treatment of patients with gastrointestinal stromal tumors, *Expert Rev Mol Diagn* 17:445–457, 2017. *[Updated discussion of molecular diagnostics and treatment of gastrointestinal stromal tumors]*.

Boland CR, Yurgelun MB: Historical perspective on familial gastric cancer, *Cell Mol Gastroenterol Hepatol* 3:192–200, 2017. *[Perspective on familial gastric cancer syndromes]*.

Polk DB, Peek RM Jr: *Helicobacter pylori*: gastric cancer and beyond, *Nat Rev Cancer* 10:403, 2010. *[Good review of* Helicobacter pylori *and mechanisms by which it is linked to gastric cancer]*.

Sagaert X, Van Cutsem E, De Hertogh G et al: Gastric MALT lymphoma: a model of chronic inflammation-induced tumor development, *Nat Rev Gastroenterol Hepatol* 7:336, 2010. *[Discussion of gastric mucosa-associated lymphoid tissue lymphoma pathogenesis]*.

Small Intestine and Colon

Intestinal Obstruction

Langer JC: Intestinal rotation abnormalities and midgut volvulus, *Surg Clin North Am* 97:147–159, 2017. *[Review of volvulus and its complications]*.

Ischemic Bowel Disease and Angiodysplasia

Barnert J, Messmann H: Diagnosis and management of lower gastrointestinal bleeding, *Nat Rev Gastroenterol Hepatol* 6:637, 2009. *[Discussion of clinical approaches to lower gastrointestinal bleeding]*.

Beg S, Ragunath K: Review on gastrointestinal angiodysplasia throughout the gastrointestinal tract, *Best Pract Res Clin Gastroenterol* 31:119–125, 2017. *[Review of angiodysplasia]*.

Malabsorption and Diarrhea

Das S, Jayaratne R, Barrett KE: The role of ion transporters in the pathophysiology of infectious diarrhea, *Cell Mol Gastroenterol Hepatol* 6:33–45, 2018. *[Physiology-oriented review of ion transport in diarrhea]*.

Guerrant RL, DeBoer MD, Moore SR et al: The impoverished gut—a triple burden of diarrhoea, stunting and chronic disease, *Nat Rev Gastroenterol Hepatol* 10:220, 2013. *[Review of mechanisms underlying growth failure (stunting)]*.

Kelly T, Buxbaum J: Gastrointestinal manifestations of cystic fibrosis, *Dig Dis Sci* 60:1903–1913, 2015. *[Review of obstruction and other gastrointestinal complications of cystic fibrosis]*.

Kelly CP, Bai JC, Liu E et al: Advances in diagnosis and management of celiac disease, *Gastroenterology* 148:1175–1186, 2015. *[Expert review of current and potential future means of managing celiac disease]*.

Thiagarajah JR, Donowitz M, Verkman AS: Secretory diarrhoea: mechanisms and emerging therapies, *Nat Rev Gastroenterol Hepatol* 12:446–457, 2015. *[Review of secretory diarrhea with a view toward future, targeted therapies]*.

Infectious Enterocolitis

Barton Behravesh C, Mody RK, Jungk J et al: 2008 outbreak of *Salmonella* Saintpaul infections associated with raw produce, *N Engl J Med* 364:918, 2011. *[Analysis of a* Salmonella *epidemic]*.

Barzilay EJ, Schaad N, Magloire R et al: Cholera surveillance during the Haiti epidemic—the first 2 years, *N Engl J Med* 368:599, 2013. *[Study of cholera outbreak following the 2010 Haitian earthquake]*.

DuPont HL: Acute infectious diarrhea in immunocompetent adults, *N Engl J Med* 370:1532–1540, 2014. *[Expert discussion of acute infectious diarrhea]*.

Ooijevaar RE, Terveer EM, Verspaget HW et al: Clinical application and potential of fecal microbiota transplantation, *Annu Rev Med* 70:335–351, 2019. *[Review of current status of and future prospects for therapeutic fecal transplantation]*.

Irritable Bowel Syndrome

Ford AC, Lacy BE, Talley NJ: Irritable bowel syndrome, *N Engl J Med* 376:2566–2578, 2017. *[Comprehensive review of irritable bowel syndrome]*.

Ghoshal UC, Gwee KA: Post-infectious IBS, tropical sprue and small intestinal bacterial overgrowth: the missing link, *Nat Rev Gastroenterol Hepatol* 14:435–441, 2017. *[Discussion of infection and microbial dysbiosis as contributors to irritable bowel syndrome pathogenesis]*.

Inflammatory Bowel Disease

Honda K, Littman DR: The microbiota in adaptive immune homeostasis and disease, *Nature* 535:75–84, 2016. *[Overarching review of host–microbial interactions focused on immune regulation]*.

Jostins L, Ripke S, Weersma RK et al: Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease, *Nature* 491:119, 2012. *[Comprehensive analysis of 163 IBD-associated genes with new insight into the relationship to pathogen responses]*.

Kanneganti TD: Inflammatory bowel disease and the NLRP3 inflammasome, *N Engl J Med* 377:694–696, 2017. *[Review of advances in understanding inflammasome signaling and relevance to inflammatory bowel disease]*.

Kelsen JR, Dawany N, Moran CJ et al: Exome sequencing analysis reveals variants in primary immunodeficiency genes in patients with very early onset inflammatory bowel disease, *Gastroenterology* 149:1415–1424, 2015. *[Identification of monogenic causes of inflammatory bowel disease]*.

Lavoie S, Conway KL, Lassen KG et al: The Crohn's disease polymorphism, ATG16L1 T300A, alters the gut microbiota and enhances the local TH1/TH17 response, *Elife* 8:2019. *[Analysis of relationship between*

risk alleles, gut microbiome, and mucosal immunity in inflammatory bowel disease].

Waldner MJ, Neurath MF: Mechanisms of immune signaling in colitis-associated cancer, *Cell Mol Gastroenterol Hepatol* 1:6–16, 2015. *[Review of immune contributions to colitis-associated cancers in the context of inflammatory bowel disease].*

Other Causes of Colitis

Collins M, Michot JM, Danlos FX et al: Inflammatory gastrointestinal diseases associated with PD-1 blockade antibodies, *Ann Oncol* 28:2860–2865, 2017. *[Discussion of inflammatory and diarrheal diseases associated with immunotherapy for cancers].*

Pardi DS: Diagnosis and management of microscopic colitis, *Am J Gastroenterol* 112:78–85, 2017. *[Review of collagenous and lymphocytic colitis pathobiology and diagnosis].*

Diverticulitis

Crowe FL, Balkwill A, Cairns BJ et al: Source of dietary fibre and diverticular disease incidence: a prospective study of UK women, *Gut* 63:1450–1456, 2014. *[Prospective study of diet and diverticular disease].*

Young-Fadok TM: Diverticulitis, *N Engl J Med* 380:500–501, 2019. *[Overarching review of diverticular disease].*

Polyps

Anderson JC, Baron JA, Ahnen DJ et al: Factors associated with shorter colonoscopy surveillance intervals for patients with low-risk colorectal adenomas and effects on outcome, *Gastroenterology* 152:1933–1943, 2017. *[Analysis of practice patterns and colonoscopy frequency].*

Click B, Pinsky PF, Hickey T et al: Association of colonoscopy adenoma findings with long-term colorectal cancer incidence, *JAMA* 319:2021–2031, 2018. *[Study of colorectal cancer risk as a function of colonoscopy findings in nearly 16,000 patients].*

East JE, Atkin WS, Bateman AC et al: British Society of Gastroenterology position statement on serrated polyps in the colon and rectum, *Gut* 66:1181–1196, 2017. *[Position statement on management of serrated polyps].*

Sengupta S, Bose S: Peutz-Jeghers syndrome, *N Engl J Med* 380:472, 2019. *[Review of Peutz-Jeghers disease etiology and management].*

Adenocarcinoma

Corley DA, Jensen CD, Marks AR et al: Adenoma detection rate and risk of colorectal cancer and death, *N Engl J Med* 370:1298, 2014. *[Large study of relationship between adenoma detection rates and colon cancer incidence].*

Dalerba P, Sahoo D, Paik S et al: CDX2 as a prognostic biomarker in stage II and stage III colon cancer, *N Engl J Med* 374:211–222, 2016. *[Identification of transcription factor CDX2 as a prognostic factor in colorectal cancer that could guide use of adjuvant therapy].*

Medema JP: Targeting the colorectal cancer stem cell, *N Engl J Med* 377:888–890, 2017. *[Short review of potential new therapies in treating colorectal cancer].*

Hemorrhoids

Jacobs D: Clinical practice. Hemorrhoids, *N Engl J Med* 371:944–951, 2014. *[Clinically oriented review of hemorrhoid management].*

Acute Appendicitis and Tumors of the Appendix

Flum DR: Clinical practice. Acute appendicitis—appendectomy or the "antibiotics first" strategy, *N Engl J Med* 372:1937–1943, 2015. *[Discussion of alternative approaches to management of acute appendicitis].*

Lee LH, McConnell YJ, Tsang E et al: Simplified 2-tier histologic grading system accurately predicts outcomes in goblet cell carcinoid of the appendix, *Hum Pathol* 46:1881–1889, 2015. *[Clinicopathologic analysis of histologic features associated with outcome in appendiceal goblet cell carcinoid tumors].*

Peritoneal Cavity

Kim J, Bhagwandin S, Labow DM: Malignant peritoneal mesothelioma: a review, *Ann Transl Med* 5:236, 2017. *[Overarching review of peritoneal mesothelioma and evolving therapies].*

Misdraji J: Mucinous epithelial neoplasms of the appendix and pseudomyxoma peritonei, *Mod Pathol* 28(Suppl 1):S67–S79, 2015. *[Systematic review of pseudomyxoma peritonei].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 18

Liver and Gallbladder

Ryan M. Gill • Sanjay Kakar[a]

CHAPTER CONTENTS

The Liver and Bile Ducts

The normal adult liver weighs 1400 to 1600 g. It has a dual blood supply, with the portal vein providing 60% to 70% of hepatic blood flow and the hepatic artery supplying the remaining 30% to 40%. The portal vein and the hepatic artery enter the inferior aspect of the liver through the hilum, or *porta hepatis*. Within the liver, the branches of the portal veins, hepatic arteries, and bile ducts travel in parallel within portal tracts.

The lobule model provides one useful way to consider the anatomic organization of the liver (Fig. 18.1). In this model, the liver is divided into 1- to 2-mm hexagonal lobules

[a]The contributions of Drs. James Crawford and Neil Theise to this chapter in the last several editions of this book are gratefully acknowledged.

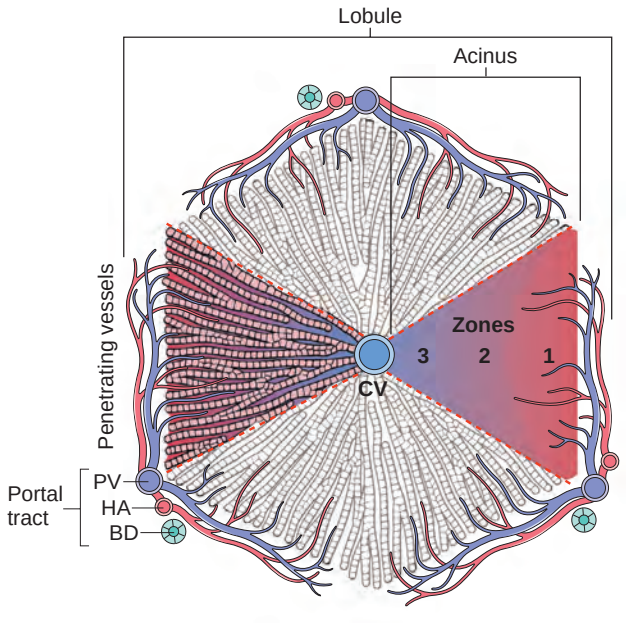


Figure 18.1 Model of liver anatomy. In the hexagonal lobular model, the terminal hepatic vein *(CV)* is at the center of a "lobule," while the portal tracts are at the periphery; these landmarks serve to identify "periportal" and "pericentral" parenchyma. In the acinar model, on the basis of blood flow, three zones can be defined, zone 1 being the closest to the blood supply and zone 3 being the farthest. *BD,* Bile duct; *HA,* hepatic artery; *PV,* portal vein.

that are oriented around the terminal tributaries of the hepatic vein. The veins lie at the center of the lobules and hence are termed central veins, and portal tracts are located at the periphery. Each lobule can be further subdivided into six triangular acini. The hepatocytes in the vicinity of the central vein are called "pericentral" and are said to be in "zone 3," while those near the portal tract are periportal and belong to "zone 1." Notably, certain types of hepatic injury preferentially affect hepatocytes in particular zones (see Fig. 18.1). These differing sensitivities result from variation in hepatocyte oxygenation (highest in zone 1, lowest in zone 3) and metabolic activities from the periphery to the center of the lobule.

Within the lobule, hepatocytes are organized into anastomosing sheets or "plates" extending from portal tracts to the terminal hepatic veins. Between the trabecular plates of hepatocytes are vascular sinusoids. Blood traverses the sinusoids and exits into the terminal hepatic veins through numerous orifices in the vein wall. Hepatocytes are thus bathed on two sides by a mixture of portal venous and hepatic arterial blood. The sinusoids are lined by fenestrated endothelium. Beneath the endothelial cells lies the *space of Disse,* into which protrude abundant hepatocyte microvilli. Scattered *Kupffer cells* of the mononuclear phagocyte system are attached to the luminal face of endothelial cells, and fat-containing *hepatic stellate cells* are found in the space of Disse. Between abutting hepatocytes are bile canaliculi, which are channels 1 to 2 μm in diameter, formed by grooves in the plasma membranes of facing hepatocytes and separated from the vascular space by tight junctions. These channels drain into the *canals of Hering* that, in turn, connect to bile ductules. The ductules empty into the interlobular bile ducts within portal tracts. Large numbers of lymphocytes (mostly gamma delta T cells, but also natural killer [NK] cells) are also present in normal liver, comprising as many as 22% of cells other than hepatocytes.

GENERAL FEATURES OF LIVER DISEASE

The liver is vulnerable to a wide variety of metabolic, toxic, microbial, circulatory, and neoplastic insults. Hepatic damage also occurs secondary to other diseases, such as heart failure, disseminated cancer, and extrahepatic infections. The enormous functional reserve of the liver masks the clinical impact of mild liver damage, but with progression of diffuse disease or disruption of bile flow, the consequences of deranged liver function may become life threatening.

With the exception of acute liver failure, liver disease is an insidious process in which clinical detection and symptoms of hepatic decompensation may occur weeks, months, or many years after the onset of injury. The ebb and flow of hepatic injury may be imperceptible to the patient and detectable only by abnormal laboratory tests (Table 18.1).

Mechanisms of Injury and Repair

Hepatocyte and Parenchymal Responses

Injured or dysfunctional hepatocytes in a variety of disorders may demonstrate several potentially reversible morphologic changes. These include accumulation of fat (steatosis) and bilirubin (cholestasis), as well as ballooning, a change marked by cell swelling, cytoplasmic clearing, and clumping of intermediate filaments, which when prominent, may form *Mallory hyaline.* Ballooned hepatocytes are a hallmark of

Table 18.1 Laboratory Evaluation of Liver Disease

Test Category	Serum Measurement
Hepatocyte integrity	Cytosolic hepatocellular enzymes[a] Serum aspartate aminotransferase (AST) Serum alanine aminotransferase (ALT) Serum lactate dehydrogenase (LDH)
Biliary excretory function	Substances normally secreted in bile[a] Serum bilirubin Total: unconjugated plus conjugated Direct: conjugated only Urine bilirubin Serum bile acids Plasma membrane enzymes (from damage to bile canaliculus)[a] Serum alkaline phosphatase Serum γ-glutamyl transpeptidase (GGT)
Hepatocyte synthetic function	Proteins secreted into the blood Serum albumin[b] Coagulation factors[b] Prothrombin time (PT) and partial thromboplastin time (PTT): fibrinogen, prothrombin, factors V, VII, IX, and X Hepatocyte metabolism Serum ammonia[a] Aminopyrine breath test (hepatic demethylation)[b]

[a]Increased in liver disease.

[b]Decreased in liver disease.

alcohol-induced or nonalcoholic steatohepatitis (see later discussion), but may also occur with ischemic or toxic injury, or with cholestasis.

When injury is irreversible, hepatocytes may die by necrosis or apoptosis. In hepatocyte necrosis, ion imbalances due to defective plasma membrane transporter function cause water to flow into the cell, which swells and ruptures. Membrane abnormalities also lead to accumulation of intracellular calcium and several events that trigger mitochondrial dysfunction. Even before membrane ruptures, loss of membrane integrity leads to the release of cytoplasmic constituents into the extracellular compartment, including substances that alert innate immune cells to an ongoing threat (danger signals). The remnants of the necrotic cells are rapidly phagocytosed by macrophages, which tend to cluster and mark sites of hepatocyte necrosis (Fig. 18.2). This form of injury is the predominant mode of death in ischemic/hypoxic injury and a significant part of hepatic injury in the setting of oxidative stress.

Hepatocyte apoptosis is an active form of "programmed" cell death resulting in hepatocyte shrinkage, nuclear chromatin condensation *(pyknosis),* fragmentation *(karyorrhexis),* and cellular fragmentation into acidophilic *apoptotic bodies*. These changes are a result of caspase cascades described in detail in Chapter 2. Apoptotic hepatocytes were first clearly described in yellow fever by William Thomas Councilman and therefore have often been referred to as *Councilman bodies;* while apoptosis occurs in many forms of liver disease, by convention this eponym is restricted to that disease. In the more frequent settings in which apoptotic hepatocytes are seen (e.g., acute and chronic hepatitis), the term *acidophil bodies* is used, due to their deeply eosinophilic staining characteristics (Fig. 18.3).

If parenchymal injury is widespread, it may lead to confluent necrosis, a zonal loss of contiguous hepatocytes. This may be seen in acute toxic or ischemic injuries or in severe viral or autoimmune hepatitis. Confluent necrosis may begin with hepatocyte dropout in zone 3 near the central vein. The resulting space is filled by cellular debris, macrophages, and remnants of the reticulin meshwork. In *bridging necrosis,* the area of necrosis may extend from the central veins to the portal tracts, or across adjacent portal tracts (often with an inapparent central vein within the area of injury). In pan-acinar necrosis, the entire lobule is obliterated. Even in diseases such as viral hepatitis, in which hepatocytes are the principal targets of attack, secondary vascular insults—due to inflammation or thrombosis—may produce large areas of confluent hepatocyte necrosis. This process occurs in many types of liver diseases in which there is extensive hepatocyte loss and collapse of the supporting reticulin framework. The resultant scarring (known in the liver as *cirrhosis*) in which hepatocytes are surrounded by fibrosis (with or without a regenerative response) represents a common end point for chronic liver disease. In some cases, there is scar regression (described in the next section).

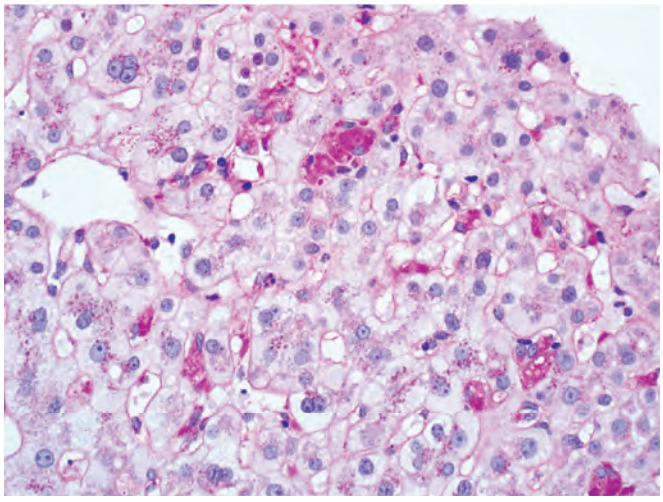


Figure 18.2 Acute hepatitis B. Clusters of macrophages with PAS-positive intracellular material derived from necrotic hepatocytes.

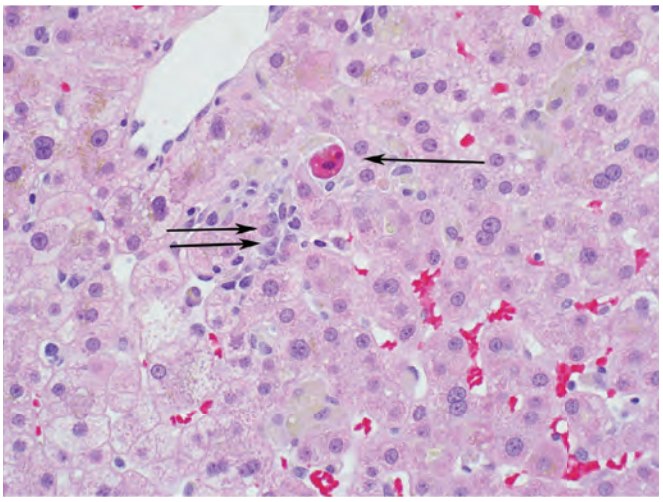


Figure 18.3 Foci of lobular hepatitis in chronic hepatitis C show an apoptotic hepatocyte ("acidophil body"; *arrow*) and a focus of mononuclear infiltration surrounding a more darkly stained, injured hepatocyte *(double arrows).*

Replacement of lost hepatocytes occurs primarily by replication of mature hepatocytes adjacent to those that have died, even when there is significant confluent necrosis. Hepatocytes are almost stem cell–like in their ability to continually replicate, even after years of chronic injury, and thus replenishment of parenchyma from tissue stem cells is usually not a significant part of the repair process. Eventually, however, in individuals with chronic disease, the hepatocytes enter into replicative senescence; once this occurs, stem cell populations may begin to expand and differentiate, an event marked by *ductular reactions*. These ductlike structures, sometimes without lumens, contain multipotent cells that may contribute significantly to parenchymal restoration.

Scar Formation and Regression

The principal cell type involved in scar deposition in the liver is the hepatic stellate cell. When quiescent, the main function of stellate cells is lipid storage, including vitamin A. However, in several forms of acute and chronic injury, stellate cells become activated and differentiate into highly fibrogenic myofibroblasts. The stimuli for stellate cell activation are varied and include: (1) inflammatory cytokines, such as tumor necrosis factor alpha (TNF-α), produced by Kupffer cells, macrophages, and other cell types; (2) altered

interactions with extracellular matrix (ECM); and (3) toxins and reactive oxygen species (ROS). Following activation, the conversion of stellate cells into myofibroblasts is stimulated by signals transmitted by platelet-derived growth factor receptor β and by cytokines, such as transforming growth factor β (TGF-β) and interleukin-17 (IL-17), as well as chemokines, which may be released from resident Kupffer cells or recruited macrophages and lymphocytes. The activated stellate cells in turn also release cytokines, growth factors or and chemotactic and vasoactive factors. If the injury and inflammatory stimuli persist, extracellular matrix deposition and scarring commence, often in the space of Disse; there is concurrent loss of sinusoidal endothelial cell fenestration, a change termed *sinusoidal capillarization* that is a particularly prominent feature in nonalcoholic steatohepatitis and also is a feature of the abnormal sinusoids seen in hepatocellular carcinoma (both described later).

Areas of hepatocyte loss in chronic liver disease, perhaps related to vascular compromise, are transformed into dense fibrous septa through collapse of the underlying reticulin framework and deposition of collagen by myofibroblasts. Portal fibroblasts may also play a role in the scarring that accompanies chronic liver injury in some disorders. Eventually, the fibrous septa encircle surviving hepatocytes and give rise to diffuse scarring (cirrhosis). In chronic liver disease, surviving hepatocytes replicate in an effort to restore the parenchyma, forming regenerative nodules that are a predominant feature in most cirrhotic livers.

If the chronic injury leading to scar formation is interrupted (e.g., clearance of hepatitis virus infection or cessation of alcohol use), then stellate cell activation and scarring ceases and fibrous septae may begin to be broken down by metalloproteinases produced by hepatocytes, leading to partial resolution and an appearance termed *incomplete septal cirrhosis*. Unfortunately, vascular remodeling and other architectural changes that occur in cirrhosis may not revert to normal, even with extensive scar resorption, which may explain why vascular abnormalities such as portal hypertension fail to improve in some patients.

Liver Failure

The most severe form of liver disease is liver failure, which may be acute (due to sudden and massive hepatic destruction) or (more commonly) chronic, following years of insidious, progressive liver injury. In some cases, individuals with chronic liver disease develop *acute-on-chronic liver failure,* in which an unrelated acute injury is superimposed on late-stage chronic disease or there is a "flare" of the chronic disease that leads to liver decompensation. Whatever the sequence, 80% to 90% of functional capacity must be lost before hepatic failure appears. When the liver can no longer maintain homeostasis, liver transplantation offers the best hope for survival; without transplantation, the mortality rate in persons with hepatic failure is about 80%.

Acute Liver Failure

Acute liver failure is defined as an acute liver illness associated with encephalopathy and coagulopathy that occurs within 26 weeks of the initial liver injury in the absence of preexisting liver disease. Usually the syndrome manifests within 8 weeks of injury, and many patients progress to coma within only a week. Within this 26-week window, it is useful to know the interval between the onset of symptoms and liver failure, since this may provide helpful clues to the etiology. Very rapid onset acute liver failure is most often induced by drugs or toxins and is typically the result of *massive hepatic* necrosis, This form of liver failure has been referred to as "fulminant liver failure" in the past, a term that remains entrenched in the literature and is still used interchangeably. Accidental or deliberate ingestion of acetaminophen (Chapter 9) accounts for almost 50% of adult cases in the United States, while autoimmune hepatitis, other drugs/toxins, and acute hepatitis A and B infections account for most remaining cases. In Asia, acute hepatitis B and E are the predominant causes. With acetaminophen toxicity, the liver failure occurs within 1 week of the onset of symptoms, whereas failure due to hepatitis viruses takes longer to develop. The mechanism of hepatocellular necrosis may be direct toxic damage (as with acetaminophen), but more often is a variable combination of toxicity and immune-mediated hepatocyte destruction (e.g., hepatitis virus infection). Rare causes of acute liver failure include abnormalities of blood flow, metabolic disorders, and malignancies, most commonly leukemia or lymphoma (33%), followed by breast cancer (30%) and colon cancer (7%). An etiology cannot be established in about 15% of adult and 50% of pediatric cases.

MORPHOLOGY

Acute liver failure usually is associated with **massive hepatic necrosis,** with broad regions of parenchymal loss surrounding islands of preserved or regenerating hepatocytes (Fig. 18.4). Affected livers are small and shrunken. The prominence of hepatocellular dropout and of ductular reactions in these livers depends on the nature and duration of the insult. Toxic injuries, such as acetaminophen overdoses, usually take place within hours to days, too brief a period to allow for scar formation or regeneration. Acute viral infections may cause failure over weeks to a few months, so that although hepatocyte injury continues to outpace repair, regeneration is often evident along with scarring.

Rarely, acute liver failure is associated with widespread dysfunction of liver cells without obvious cell death, such as in **diffuse microvesicular steatosis** related to fatty liver of pregnancy or in idiosyncratic reactions to toxins (e.g., valproate, tetracycline). In these settings, hepatocyte metabolism is severely affected, usually due to mitochondrial dysfunction, preventing the liver from carrying out its normal functions. In states of immunodeficiency, such as untreated infection with human immunodeficiency virus (HIV), posttransplant immunosuppression, and certain lymphoid malignancies, nonhepatotropic viruses, particularly cytomegalovirus, herpes simplex viruses, and adenovirus, can cause acute liver failure. With better treatments for HIV infection, these forms of viral hepatitis are declining in incidence.

Clinical Features

Acute liver failure manifests first with nausea, vomiting, and jaundice, followed by life-threatening encephalopathy, and coagulation defects. Typically, serum liver transaminases are markedly elevated. The liver is initially enlarged due to hepatocyte swelling, inflammatory infiltrates, and edema;

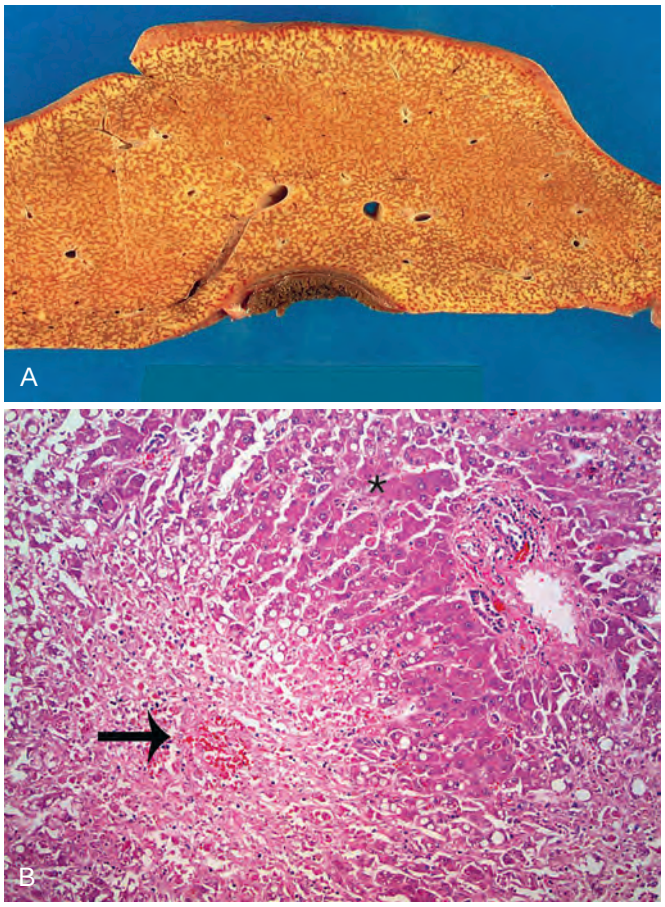


Figure 18.4 (A) Massive necrosis, cut section of liver. The liver is small (700 g), bile-stained, soft, and congested. (B) Hepatocellular necrosis caused by acetaminophen overdose. Confluent necrosis is seen in the perivenular region (zone 3) *(large arrow)*. Residual normal tissue is indicated by an *asterisk*. (Courtesy Dr. Matthew Yeh, University of Washington, Seattle, Wash.)

as parenchyma is destroyed, however, the liver shrinks dramatically. Decline of serum transaminases (as hepatocytes die) often is not, therefore, a sign of improvement, but is rather an indication that there are few viable hepatocytes left; this suspicion is confirmed if there is worsening jaundice, coagulopathy, and encephalopathy. With unabated progression, multiorgan system failure occurs and, if transplantation is not possible, death ensues. Other manifestations of acute liver failure are as follows:

- *Cholestasis* occurs due to alterations of bile formation and flow leading to retention of bilirubin and other solutes normally eliminated in bile. Yellow discoloration of the skin and sclera (*jaundice* and *icterus,* respectively) occur with bilirubin retention, and there is increased risk of potentially life-threatening bacterial infection.
- *Hepatic encephalopathy* is a spectrum of disturbances in consciousness, ranging from subtle behavioral abnormalities, to marked confusion and stupor, to deep coma and death. Encephalopathy may progress over days, weeks, or months following acute liver injury. Associated fluctuating neurologic signs include rigidity and hyperreflexia. *Asterixis,* a particularly characteristic sign, is manifested as nonrhythmic, rapid extension-flexion movements of the head and extremities, best seen when the arms are held in extension with dorsiflexed wrists. Hepatic encephalopathy is believed to be caused by elevated ammonia levels, which correlate with impaired neuronal function and cerebral edema. The principal source of the ammonia is the gastrointestinal tract, where it is produced by microorganisms and also by enterocytes during glutamine metabolism. Normally, ammonia is transported in the portal vein to the liver, where it is metabolized in the urea cycle; in severe liver disease, this detoxification mechanism fails. Several nonexclusive mechanisms of CNS toxicity for ammonia have been proposed, including direct effects on neurons and indirect effects mediated through its metabolism in astocytes to glutamine, which appears to contribute to edema by serving as an osmolyte.
- *Coagulopathy*. Hepatocytes are responsible for synthesis of clotting factors II (prothrombin), V, VII, IX, X, XI, and XII, as well as fibrinogen (Chapter 4). Thus, with liver failure, factor deficiencies and coagulopathy develops. Easy bruisability is an early sign, which can progress to life-threatening or fatal intracranial bleeding. The liver is also responsible for helping to remove activated coagulation factors from the circulation, and loss of this function in some instances leads to disseminated intravascular coagulation (Chapter 14), further exacerbating the bleeding tendency.
- *Portal hypertension* arises when there is diminished flow through the portal venous system, which may occur because of obstruction at the prehepatic, intrahepatic, or post-hepatic level. Although it can occur in acute liver failure, portal hypertension is more commonly seen with chronic liver failure and is described later. In acute liver failure, if portal hypertension develops within days to weeks, obstruction is predominantly intrahepatic and the major clinical consequence is *ascites*. In chronic liver disease, portal hypertension develops over months to years, and its effects are more complex and widespread.
- *Hepatorenal syndrome* is a form of renal failure occurring in individuals with liver failure in whom there is no intrinsic morphologic or functional cause for kidney dysfunction. Its onset is marked by a drop in urine output and increasing levels of urea and creatinine in the blood. Its pathophysiology is incompletely understood, but the triggering event is hypothesized to be portal hypertension and secondary increased production of vasodilators such as nitric oxide by endothelial cells in the splanchnic vasculature. This in turn leads to systemic vasodilation and diminished renal perfusion, which is sensed by the kidney, provoking activation of the renin/angiotensin axis. In the setting of portal hypertension and persistent vasodilator production, the principal effect of renin/angiotensin activation is to further diminish renal perfusion and glomerular filtration rate, leading to renal failure. The decline in renal function is reversible if liver function is restored, for example, by liver transplantation. Patients who develop the hepatorenal syndrome usually have portal hypertension due to cirrhosis, severe alcoholic hepatitis, or (less often) metastatic tumors. However, patients with fulminant hepatic failure from any cause may develop hepatorenal syndrome.

Chronic Liver Failure and Cirrhosis

Liver failure in chronic liver disease is most often associated with advanced fibrosis/cirrhosis, a condition marked by diffuse remodeling of the liver into parenchymal nodules (often regenerative) surrounded by fibrous bands and a variable degree of vascular (often portosystemic) shunting. The leading causes of chronic liver failure worldwide are chronic hepatitis B, chronic hepatitis C, nonalcoholic fatty liver disease, and alcoholic liver disease. In the United States, chronic liver disease is the twelfth most common cause of mortality and accounts for most liver-related deaths.

Although cirrhosis and chronic liver failure are often associated, they are not synonymous; not all cirrhosis leads inexorably to chronic liver failure, and not all end-stage chronic liver disease is cirrhotic. For example, chronic diseases such as primary biliary cholangitis, primary sclerosing cholangitis, nodular regenerative hyperplasia, chronic schistosomiasis, and fibropolycystic liver disease are often not accompanied by fully established cirrhosis, even at end stage. On the other hand, patients with well-treated autoimmune hepatitis or those with suppressed hepatitis B or cured hepatitis C often do not progress to end-stage liver disease, even if cirrhosis is present.

In some diseases that give rise to cirrhosis, such as untreated viral hepatitis, alcoholic liver disease, nonalcoholic fatty liver disease, and metabolic diseases, the morphology and pathophysiology of cirrhosis show some distinctive features (described in subsequent sections of this chapter). Thus, while the term cirrhosis implies the presence of severe chronic disease, it is not a specific diagnosis and has variable prognostic implications. There are also some instances in which cirrhosis arises without any clear cause; the term *cryptogenic cirrhosis* is sometimes applied to such cases.

MORPHOLOGY

Cirrhosis is marked by the presence of parenchymal nodules surrounded by dense bands of fibrosis throughout the liver, converting the normally smooth liver capsule into a bumpy surface with depressed areas of scarring and bulging regenerative nodules (Fig. 18.5). Microscopically the extent of the fibrosis is highlighted by the use of special stains for collagen (Fig. 18.6A). The size of the nodules, the pattern of scarring (linking portal tracts to each other versus linking portal tracts to central veins), the degree of parenchymal collapse, and the extent of vascular thrombosis (particularly of the portal vein) all vary between diseases and, to some degree, between individuals with the same disease. Morphologic features of regression include thin incomplete scars (see Fig. 18.6B) associated with variable ductular reaction and architectural changes. As mentioned earlier, not all end-stage livers are cirrhotic, but all reveal varying degrees of injury, repair, regeneration, and fibrosis. Thus, end-stage liver disease is best viewed as a spectrum of ineffective attempts to repair ongoing, intermittent, or past injury.

Clinical Features

About 40% of individuals with cirrhosis are asymptomatic until the most advanced stages of the disease. When symptomatic, they present with nonspecific manifestations:

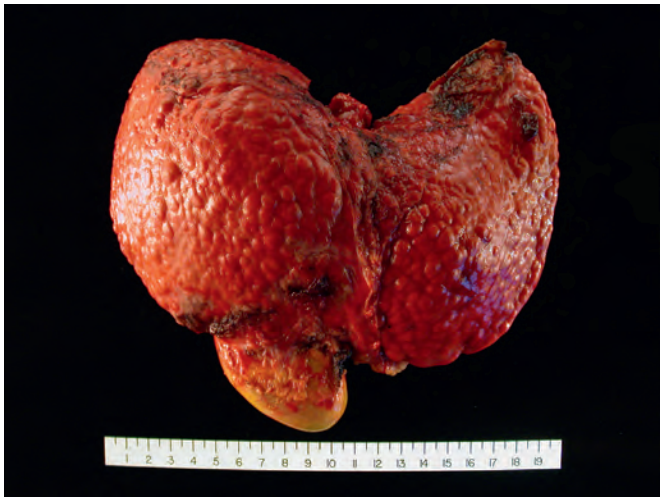


Figure 18.5 Cirrhosis resulting from chronic viral hepatitis. Note the depressed areas of dense scar separating bulging regenerative nodules over the liver surface.

anorexia, weight loss, weakness, and, in advanced disease, symptoms and signs of liver failure discussed earlier. The common causes of death are hepatic encephalopathy, bleeding from esophageal varices, bacterial infections (resulting from damage to the gut mucosal barrier and Kupffer cell dysfunction), and hepatocellular carcinoma.

In addition to jaundice, encephalopathy, and coagulopathy (all also seen with acute liver failure), chronic liver failure is associated with several other significant features.

- *Through unclear mechanisms, persistent cholestasis can lead to pruritus* (itching), the intensity of which can be profound. Some patients may scratch their skin raw, risking repeated bouts of potentially life-threatening infection. Relief may only be found in liver transplantation.
- *Impaired estrogen metabolism leads to hyperestrogenemia, which has several effects.* It produces vascular changes that may lead to palmar erythema (a reflection of local vasodilation) and spider angiomas of the skin. Each angioma is a central, pulsating, dilated arteriole from which small vessels radiate. In males, hyperestrogenemia may also produce *hypogonadism* and *gynecomastia*.
- *Hypogonadism can also occur in women due to disruption of hypothalamic-pituitary axis function,* either through nutritional deficiencies associated with the chronic liver disease or primary hormonal alterations.

Portal Hypertension

Increased resistance to portal blood flow may develop in a variety of circumstances, which can be divided into *prehepatic, intrahepatic, and posthepatic* (Table 18.2). The major prehepatic conditions are obstructive thrombosis, narrowing of the portal vein before it ramifies within the liver, or massive splenomegaly with increased splenic vein blood flow. The main posthepatic causes are severe right-sided heart failure, constrictive pericarditis, and hepatic vein outflow obstruction. The dominant intrahepatic cause is cirrhosis, accounting for most cases of portal hypertension. Far less frequent intrahepatic causes are schistosomiasis, massive fatty change, diffuse fibrosing granulomatous disease such as sarcoidosis, and diseases affecting the portal

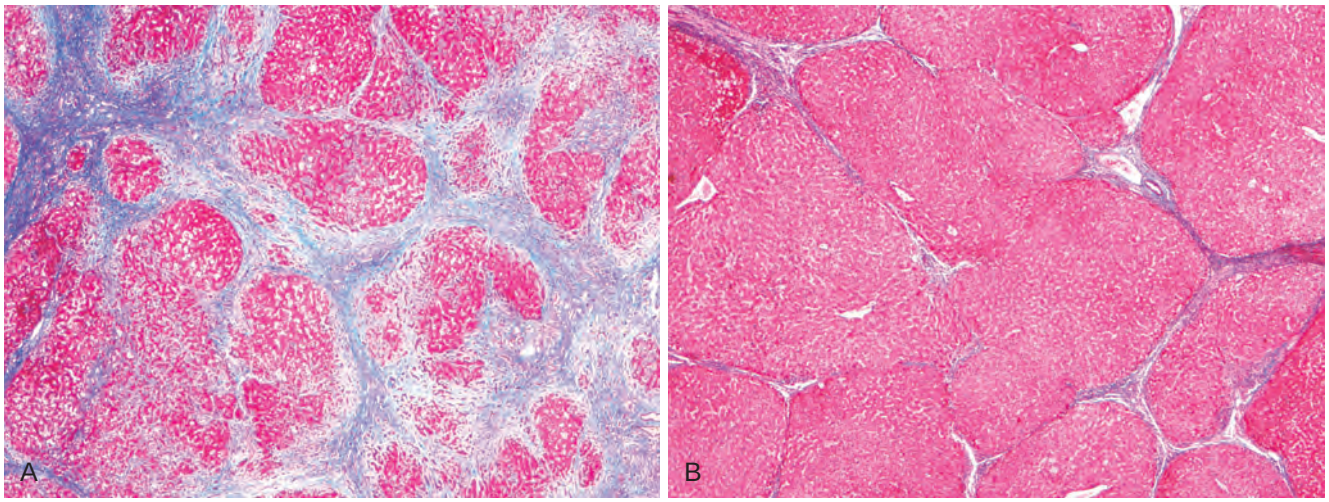


Figure 18.6 Alcoholic cirrhosis in an active drinker (A) and after long-term abstinence (B). (A) Thick bands of collagen separate rounded cirrhotic nodules. (B) After 1 year of abstinence, most scars are gone. (Masson trichrome stain) (Courtesy Drs. Hongfa Zhu and Isabel Fiel, Mount Sinai School of Medicine, NY.)

Table 18.2 Location and Causes of Portal Hypertension

Prehepatic Causes
Obstructive thrombosis of portal vein
Structural abnormalities such as narrowing of the portal vein before it ramifies in the liver
Intrahepatic Causes
Cirrhosis from any cause
Nodular regenerative hyperplasia
Primary biliary cholangitis (even in the absence of cirrhosis)
Schistosomiasis
Massive fatty change
Diffuse, fibrosing granulomatous disease (e.g., sarcoid)
Infiltrative malignancy, primary or metastatic
Focal malignancy with invasion into portal vein (particularly hepatocellular carcinoma)
Amyloidosis
Post-Hepatic Causes
Severe right-sided heart failure
Constrictive pericarditis
Hepatic vein outflow obstruction

microcirculation such as nodular regenerative hyperplasia (discussed later).

The pathophysiology of portal hypertension is complex and involves resistance to portal flow at the level of sinusoids and an increase in portal flow caused by hyperdynamic circulation. The increased resistance to portal flow at the level of the sinusoids is caused by contraction of vascular smooth muscle cells and myofibroblasts, and disruption of blood flow by scarring and the formation of parenchymal nodules. Alterations in sinusoidal endothelial cells that contribute to the intrahepatic vasoconstriction associated with portal hypertension include a decrease in nitric oxide (NO) production and increased release of endothelin-1, angiotensinogen, and eicosanoids. Sinusoidal remodeling and anastomoses between the arterial and portal system in the fibrous septa contribute to portal hypertension by imposing arterial pressures on the low-pressure portal venous system. Sinusoidal remodeling and intrahepatic shunts also interfere with the metabolic exchange between sinusoidal blood and hepatocytes.

Another major factor in the development of portal hypertension is an increase in portal venous blood flow resulting from a hyperdynamic circulation. This is caused by arterial vasodilation, primarily in the splanchnic circulation. The increased splanchnic arterial blood flow in turn leads to increased venous efflux into the portal venous system. While various mediators such as prostacyclin and TNF have been implicated in the causation of the splanchnic arterial vasodilation, NO has emerged as the most significant one.

The four major consequences of portal hypertension are (1) hepatic encephalopathy (described under liver failure), (2) ascites, (3) the formation of portosystemic venous shunts, and (4) congestive splenomegaly. These are illustrated in Fig. 18.7 and are described next.

Ascites

The accumulation of fluid in the peritoneal cavity is called ascites, which in 85% of cases is caused by cirrhosis. Ascites usually becomes clinically detectable when at least 500 mL have accumulated. The fluid is generally serous, having less than 3 g/dL of protein (largely albumin), and a serum-to-ascites albumin gradient of ≥1.1 g/dL. The fluid may contain a scant number of mesothelial cells and mononuclear leukocytes. The presence of neutrophils suggests infection, whereas the presence of red cells points to possible disseminated intra-abdominal cancer. With long-standing ascites, seepage of peritoneal fluid through trans-diaphragmatic lymphatics may produce hydrothorax, usually on the right side. The pathogenesis of ascites is complex, involving sinusoidal hypertension, hypoalbuminemia, increased hepatic lymph flow, and splanchnic vasodilation and hyperdynamic circulation.

Portosystemic Shunts

In the setting of chronic portal hypertension, vascular dilation and remodeling often leads to the development of thin-walled venous shunts between the portal and systemic circulations that bypass the liver. These shunts

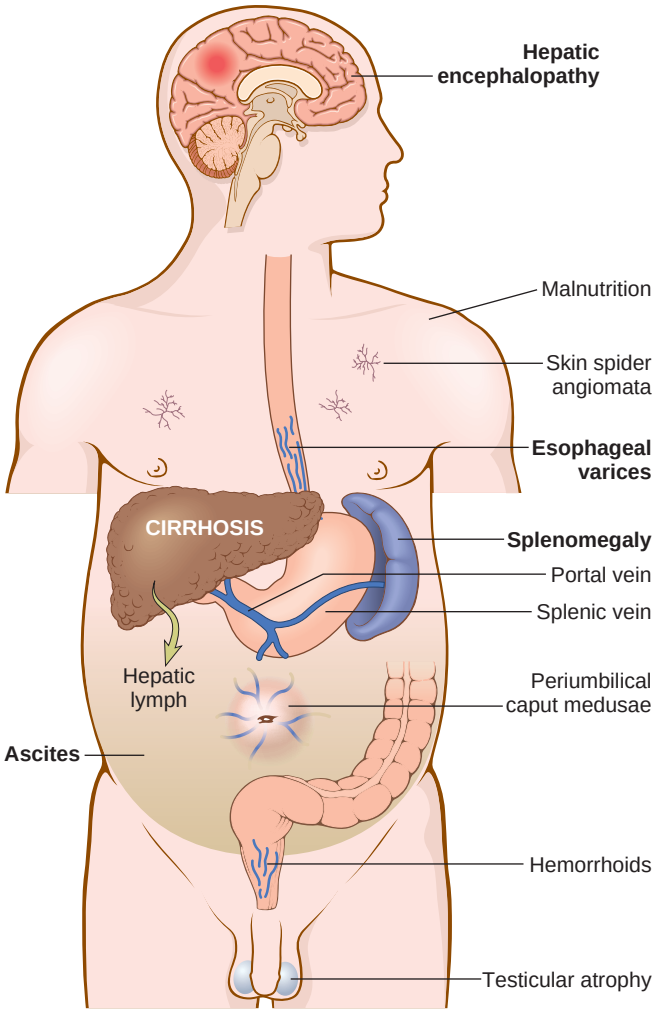


Figure 18.7 Major clinical consequences of portal hypertension in the setting of cirrhosis, shown for the male. In women, oligomenorrhea, amenorrhea, and sterility as a result of hypogonadism are frequent. Clinically significant findings are in boldface.

may appear wherever the systemic and portal circulations share capillary beds (see Fig. 18.7). The principal sites are veins around and within the rectum (manifest as hemorrhoids), the esophagogastric junction (producing varices), the retroperitoneum, and the falciform ligament of the liver (involving periumbilical and abdominal wall collaterals). Abdominal wall collaterals appear as dilated subcutaneous veins extending from the umbilicus toward the rib margins (caput medusae) and constitute a clinical hallmark of portal hypertension. Although hemorrhoidal bleeding may occur, it is rarely massive or life threatening. Much more important are esophagogastric varices, which appear in about 40% of individuals with advanced cirrhosis of the liver and cause massive hematemesis and death in about half of those affected. Each episode of bleeding is associated with ~30% mortality.

Splenomegaly

Long-standing portal hypertension may cause congestive splenomegaly. The degree of splenic enlargement varies widely, and the splenic weight may reach as much as 1000 g (five to six times normal), but it is not necessarily correlated with other features of portal hypertension. Splenomegaly may secondarily induce hematologic abnormalities attributable to "hypersplenism," particularly thrombocytopenia or even pancytopenia, in large part because of sequestration of blood elements in the expanded splenic red pulp.

Pulmonary Complications of Liver Failure and Portal Hypertension. We conclude this discussion with two pulmonary syndromes that occur in the setting of chronic liver failure and portal hypertension:

- *Hepatopulmonary syndrome* is seen in up to ~30% patients with cirrhosis of the liver and portal hypertension. The syndrome is caused by the dilatation of intrapulmonary capillary and precapillary vessels up to 500 μM in size. Right-to-left shunting of blood through the dilated vessels produces ventilation-perfusion mismatch and impairs oxygenation of blood, which manifests as hypoxemia. The resultant dyspnea is worse in the upright position rather than in the recumbent position, as gravity exacerbates the ventilation-perfusion mismatch. Patients with this syndrome have a poorer prognosis than patients without hepatopulmonary syndrome. The pathogenesis is unclear, although it has been postulated that the diseased liver may not clear factors such as endothelin-1 that stimulate endothelial cells to produce vasodilators such as nitric oxide.
- *Portopulmonary hypertension* refers to pulmonary arterial hypertension arising in liver disease. It also is not well understood, but it seems to depend on concomitant portal hypertension and excessive pulmonary vasoconstriction and vascular remodeling. The most common clinical manifestations are dyspnea on exertion and clubbing of the fingers.

Acute-on-Chronic Liver Failure

Some individuals with stable but well-compensated, advanced chronic liver disease suddenly develop signs of acute liver failure. In such patients, there is often established cirrhosis with extensive vascular shunting. Thus, large volumes of functioning liver parenchyma have a borderline vascular supply, leaving them highly vulnerable to superimposed insults. The short-term mortality of patients with this form of liver failure is around 50%.

Patients with chronic hepatitis B infection who become superinfected with hepatitis D may suddenly decompensate, as may patients with medically suppressed hepatitis B infection in whom viral mutants arise that are resistant to therapy, both due to acute flares of disease. Ascending cholangitis in a patient with primary sclerosing cholangitis or fibropolycystic liver disease (described later) may also cause rapid decompensation of liver function. Rare patients with nonalcoholic steatohepatitis may develop severe hepatic dysfunction following rapid weight loss or malnutrition; the mechanism of injury in such cases is unknown.

Other causes of decompensation are systemic rather than intrahepatic insults. For example, sepsis and its attendant hypotension may undermine hepatic parenchyma with borderline vascular supply. Likewise, acute cardiac failure or a superimposed drug or toxic injury might tip a well-compensated cirrhotic patient into failure. Finally, the possibility of malignancy, either a primary liver tumor,

particularly hepatocellular carcinoma or cholangiocarcinoma, or liver metastases from some other site (e.g., colon), should always be kept in mind as a possible cause of acute liver failure in previously well-compensated chronic liver disease.

KEY CONCEPTS

LIVER FAILURE

- Liver failure may follow acute injury or chronic injury, but may also occur as an acute insult superimposed on an otherwise well-compensated chronic liver disease.
- The mnemonic for causes of acute liver failure is as follows:
  - A: Acetaminophen, hepatitis A, autoimmune hepatitis
  - B: Hepatitis B
  - C: Hepatitis C, cryptogenic
  - D: Drugs/toxins, hepatitis D
  - E: Hepatitis E, esoteric causes (e.g., Wilson disease, Budd-Chiari syndrome, lymphoma, carcinoma)
  - F: Fatty change of the microvesicular type (e.g., fatty liver of pregnancy, valproate, tetracycline, Reye syndrome)
- Serious and sometimes fatal sequelae of liver failure include coagulopathy, encephalopathy, portal hypertension, bleeding esophageal varices, hepatorenal syndrome, and portopulmonary hypertension.

INFECTIOUS DISORDERS

Viral Hepatitis

The term "viral hepatitis" is most commonly used in the context of liver involvement by hepatotropic viruses, which include hepatitis A, B, C, D, and E. Nonhepatotropic viruses such as Epstein-Barr virus, cytomegalovirus, herpes simplex virus, adenovirus, and yellow fever virus can also cause hepatitis, often in association with systemic infection. Liver involvement can occur in a wide variety of other systemic viral infections, but it is generally mild and often subclinical.

In the sections that follow, we will first discuss individual hepatotropic viruses, and then review certain clinicopathologic syndromes that are common to all.

Hepatitis A Virus

Hepatitis A virus (HAV) infection is a self-limited disease and does not lead to chronic hepatitis or a carrier state. It accounts for about 25% of clinically evident acute hepatitis worldwide and is responsible for an estimated 2000 new cases of viral hepatitis per year in the United States.

Epidemiology

HAV is spread by ingestion of contaminated water and foods, and is endemic in countries with poor hygiene and sanitation. In high income countries, the prevalence of seropositivity (indicative of previous exposure) increases gradually with age, reaching 50% by 50 years of age in the United States. HAV is shed in the stool for 2 to 3 weeks before and 1 week after the onset of jaundice. Close personal contact with an infected individual or fecal-oral contamination during this period accounts for most cases and can result in outbreaks in institutional settings such as schools and nurseries. Water-borne epidemics can occur in places with overcrowded, unsanitary conditions. In high income countries, sporadic infections may be contracted by the consumption of raw or steamed shellfish (oysters, mussels, clams), which concentrate the virus from seawater contaminated with human sewage. Sexual transmission may occur, but maternal-fetal transmission does not. Since HAV viremia is transient, blood-borne transmission of HAV is rare, and donated blood is not screened for this virus.

Pathogenesis

Discovered in 1973, HAV is a small, nonenveloped, positive-strand RNA picornavirus that occupies its own genus, *Hepatovirus*. Ultrastructurally, HAV is an icosahedral capside that is 27 nm in diameter. The receptor for HAV on hepatocytes is HAVcr-1 (also known as TIM-1), a glycoprotein that is a member of a family of proteins that serve as receptors for several other viruses as well. HAV is not cytopathic, and the hepatocellular injury is inflicted by cytotoxic T-lymphocytes and NK cells that recognize and kill hepatocytes infected by the virus.

Clinical Features

The incubation period for HAV is 2 to 6 weeks. IgM antibody against HAV appears with the onset of symptoms and persists for 3 to 6 months (Fig. 18.8). IgG anti-HAV appear during recovery from the acute infection and persists for years, conferring lifelong immunity against reinfection. Affected individuals have nonspecific symptoms such as fatigue and loss of appetite, and they often develop jaundice. Most patients recover within 3 months, and disease resolution occurs in nearly all patients by 6 months. Acute liver failure occurs in 0.1% to 0.3% of patients, especially those with chronic liver disease due to another etiology. Other

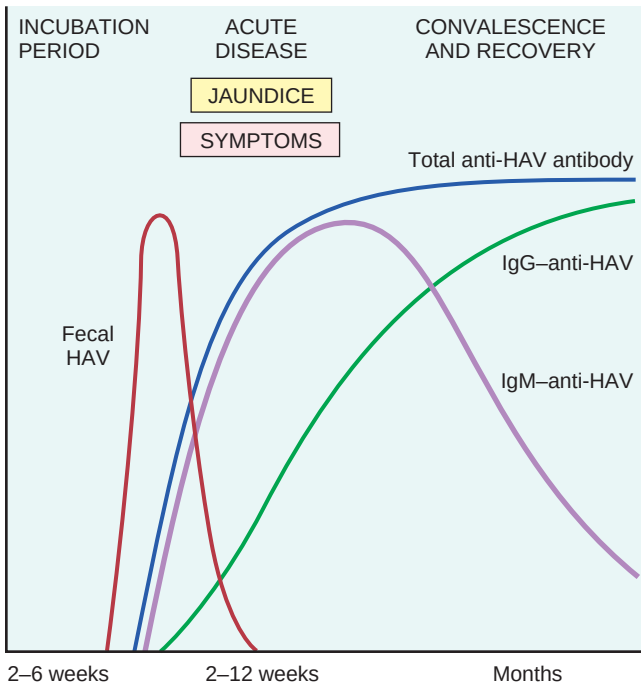


Figure 18.8 Temporal changes in serologic markers in acute hepatitis A infection *(HAV)*. *IgM*, immunoglobulin M.

uncommon complications include prolonged cholestasis and relapse of disease within 6 months of the original onset. Extrahepatic manifestations include rash, arthralgia and immune complex mediated complications like leukocytoclastic vasculitis, glomerulonephritis and cryoglobulinemia. HAV vaccine is effective in preventing infection.

Hepatitis B Virus

Hepatitis B virus (HBV) infection has varied clinical outcomes, which depend on the age of exposure, comorbid conditions (including exposure to other infectious agents), and host immunity. **The major clinical presentations include: (1) acute hepatitis followed by recovery and clearance of the virus, (2) acute hepatic failure with massive liver necrosis, (3) chronic hepatitis with or without progression to cirrhosis, and (4) an asymptomatic, "healthy" carrier state** (Fig. 18.9).

Epidemiology

Chronic HBV infection affects 400 million people worldwide, with the highest prevalence (>8%) in Africa, Asia, and the Western Pacific rim. The prevalence is also relatively high (2% to 7%) in Southern and Eastern Europe, and lower (<2%) in Western Europe, North America, and Australia. It is estimated that there are 60,000 new cases of HBV infection in the United States every year, and that nearly 2 million people have chronic HBV infection. The virus has a parenteral mode of transmission through unprotected sex, blood transfusion, and sharing of needles and syringes for intravenous drug use. Transfusion-related transmission has been sharply curtailed by screening of donated blood for HBsAg and exclusion of paid blood donors. In high-prevalence regions, transmission during childbirth accounts for 90% of cases. Horizontal transmission can also occur in children through minor breaks in the skin or mucous membranes or with close bodily contact.

Pathogenesis

HBV was first linked to hepatitis in the 1960s when Australia antigen (now known as hepatitis B surface antigen) was identified. The virus is a member of the *Hepadnaviridae,* a family of DNA viruses that cause hepatitis in multiple animal species. The mature virion has an outer surface envelope composed of viral proteins and host-derived lipids that surrounds a core consisting of nucleocapsid proteins, viral polymerase, and viral DNA. The HBV genome is a partially double-stranded circular DNA molecule with several open reading frames encoding the following proteins:

- *Hepatitis B surface antigen (HBsAg)*, which refers to three related viral envelope glycoproteins, large, middle, and small HBsAg. Large HBsAg is usually associated with complete virions, while noninfective viral envelope proteins (mainly small HBsAg) are released in large quantities by infected hepatocytes free of viral core elements.
- *Hepatitis B core antigen (HBcAg),* the nucleocapsid protein, which plays a role in assembly of complete virions; and a longer polypeptide transcript with a precore and core region, designated *hepatitis Be antigen (HBeAg)*.
- *HBV polymerase (Pol),* which exhibits both DNA polymerase and reverse transcriptase activities.
- *Hepatitis B X protein (HBx),* which is not necessary for virus replication, but may act as a transcriptional transactivator of both viral genes and a subset of host genes. It has been implicated in the pathogenesis of HBV-related hepatocellular carcinoma.

Entry of HBV into hepatocytes is enabled by binding of large HbsAg to the bile salt transporter known as sodium taurocholate cotransporting polypeptide (NTCP). The viral genome enters the nucleus, where the plus strand is synthesized to form covalently closed circular DNA (ccc DNA). HBV replication occurs through reverse transcription via an RNA intermediate.

The host immune response to the virus is the main determinant of the outcome of the infection. High-level viral replication and high production of viral proteins can produce cytopathic changes in infected cells; however, most hepatocyte injury is caused by an attack of CD8+ cytotoxic T cells on infected cells. A strong response by virus-specific

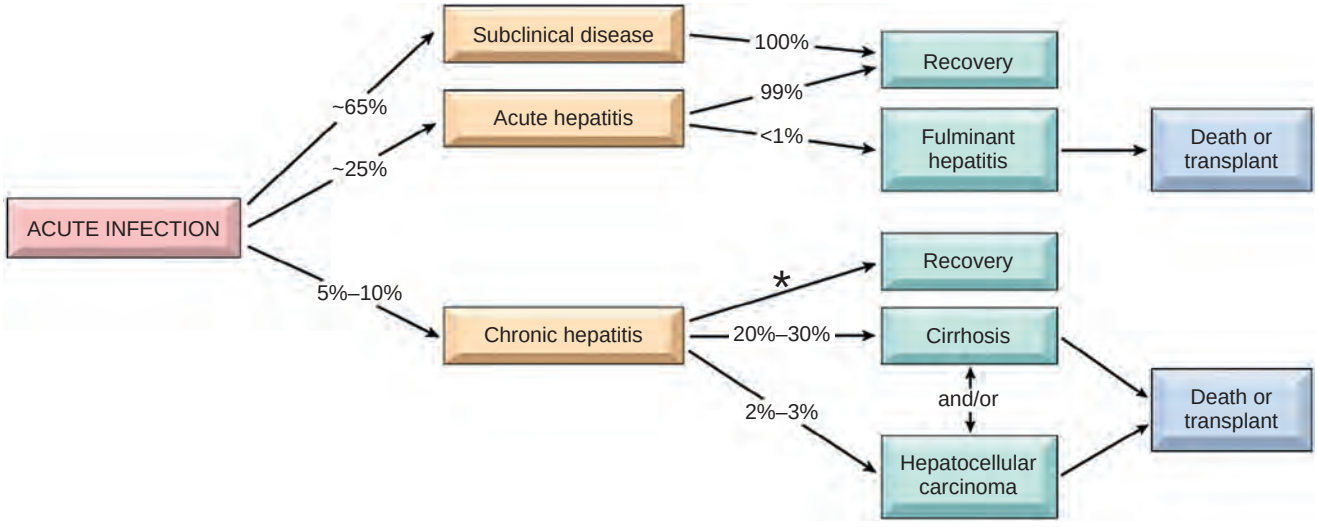


Figure 18.9 Potential outcomes of hepatitis B infection in adults and their approximate frequencies in the United States. *Spontaneous hepatitis B surface antigen clearance occurs during chronic hepatitis B viral infection at an estimated annual incidence of 1% to 2% in Western countries.

CD4+ and CD8+ interferon (IFN)-γ–producing cells is associated with resolution of acute infection.

Clinical Features

HBV has a prolonged incubation period (4 to 26 weeks). The course of the infection is followed by measuring viral antigens and serologic responses in the blood, as shown in Fig. 18.10. A few salient points about these markers is merited:

- *HBsAg* appears before the onset of symptoms and peaks in acute, symptomatic disease. In those who clear the infection, it often declines to undetectable levels in 12 weeks, but can be present for as long as 24 weeks. By contrast, HBsAg persists in cases that progress to chronicity.
- *Anti-HBs antibody* begins to rise following the resolution of acute disease, generally after the disappearance of HBsAg (see Fig. 18.10A). In some cases, the appearance of anti-HBs antibody is delayed until weeks to months after disappearance of HBsAg; in such instances, the serologic diagnosis can be made by detection of IgM anti-HBc antibody. Anti-HBs antibodies tend to persist for life, conferring protection; this is the basis for current vaccination strategies using noninfectious HBsAg. By contrast, anti-HBs antibodies are not produced in cases that progress to chronic liver disease, which is also associated with variable persistent elevations of serum transaminases in most (but not all) cases (see Fig. 18.10B).
- *HBeAg, HBV DNA, and HBV DNA polymerase* are detectable in the serum soon after HBsAg and signify active viral replication. Persistence of HBeAg is an important indicator of continued virus replication, infectivity, and probable progression to chronic hepatitis. One caveat is that mutated strains of HBV can emerge that do not produce HBeAg, but are replication competent and express HBcAg. In such patients, the HBeAg may be low or undetectable despite the presence of serum HBV DNA.
- *Anti-HBe antibody* signifies that acute infection has peaked and is on the wane, whereas in cases that progress to chronic infection, anti-HBe antibody is not produced or appears only late in the disease course.

Acute HBV infection is mild or subclinical in nearly two-thirds of adults, while others have nonspecific constitutional symptoms such as anorexia, fever, jaundice, and upper right quadrant pain. Immune complex–mediated phenomena such as glomerulonephritis may occur, as well as other immunologic disorders, such as *polyarteritis nodosa* (Chapter 11). Acute liver failure is rare, occurring in approximately 0.1% to 0.5% of acutely infected individuals. In most cases, the infection is self-limited and resolves without treatment. The infection persists and becomes chronic in 5% to 10% of infected individuals. The risk of chronic infection is inversely related to age and is greatest (approximately 90%) in infants who are exposed to the virus through transmission from their mothers at the time of birth. Individuals who acquire HBV at birth and who subsequently develop chronic hepatitis are also at greatest risk of developing hepatocellular carcinoma (Fig. 18.9), which is one of the most common and deadliest cancers in parts of the world such as China where perinatal transmission occurs frequently.

Vaccination is highly effective, as it induces a protective anti-HBs antibody response in 95% of infants, children, and adolescents. Although vaccine-induced escape mutants that replicate in the presence of vaccine-induced immunity have been described, these do not appear to be increasing in prevalence. In those with chronic HBV, treatment with interferon and antiviral agents such as reverse transcriptase inhibitors (entecavir, tenofovir) can slow disease progression, reduce liver damage, and lower the risk of cirrhosis and hepatocellular carcinoma, but complete cure is difficult.

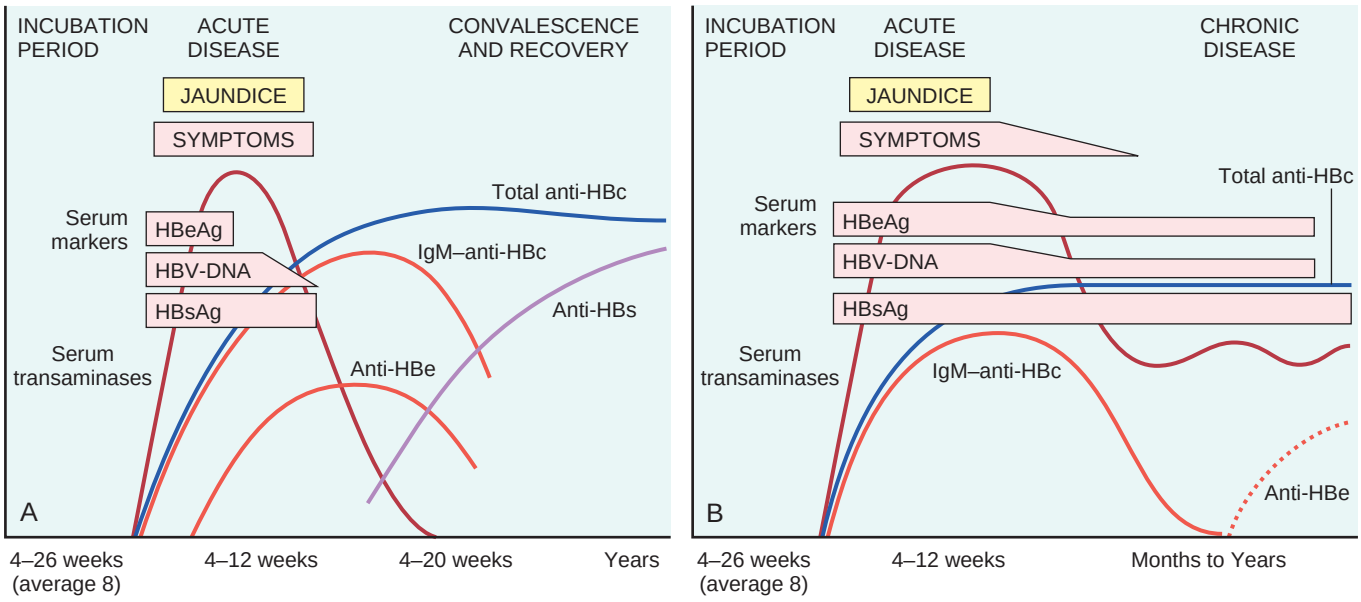


Figure 18.10 Temporal changes in serologic markers in hepatitis B viral infection. (A) Acute infection with resolution. (B) Progression to chronic infection. In some cases of chronic hepatitis B virus, serum transaminases may become normal.

Hepatitis C Virus

Hepatitis C Virus (HCV) rarely causes symptomatic acute hepatitis, but it is the most common cause of chronic viral hepatitis.

Epidemiology

HCV affects approximately 170 million people worldwide, and there are 2.7 million chronic HCV cases in the United States. Screening of donor blood has decreased the annual incidence of infection from 230,000 new infections per year in the mid-1980s to 17,000 new infections per year currently. The mode of transmission is parenteral; intravenous drug abuse, multiple sex partners, needle-stick injury, and multiple contacts with an HCV-infected person are the most common risk factors. Transfusion-related transmission has largely disappeared in the United States. Perinatal transmission can occur in children. One-third of infected individuals have no identifiable risk factors, an enduring hepatologic mystery.

Pathogenesis

HCV, discovered in 1989, is a member of the *Flaviviridae* family. It is a small, enveloped, single-stranded RNA virus with one open reading frame that codes for a single polyprotein, which is subsequently processed into several functional proteins. These include two envelope proteins, E1 and E2, and five core proteins: p7, NS2, NS3/4a (protease), NS5A (replication complex), and NS5B (RNA polymerase). The role of these proteins in the life cycle of HCV is summarized in Fig. 18.11.

The inability of the host immune response to eliminate HCV is related to rapid emergence of genetic variants, both within the population and within infected individuals. HCV RNA polymerase has low fidelity, giving rise to variants that fall into seven genotypes and many subtypes. As the disease progresses, multiple "personal" genetic variants, known as *quasispecies,* emerge in infected individuals. Anti-HCV antibodies directed against E2 envelope protein are common and have neutralizing activity, but emergent variants with altered E2 epitopes escape from this host defense. NS3/NS4A protease also impairs the interferon-mediated cellular antiviral response, defeating another potential means of host defense.

Clinical Features

The incubation period for HCV hepatitis ranges from 4 to 26 weeks, with a mean of 9 weeks. HCV RNA is detectable in blood for 1 to 3 weeks in acute infection and coincides with increase in transaminases. Anti-HCV antibodies emerge 3 to 6 weeks after infection. Spontaneous clearance of the virus after 4 to 6 months is rare, hence persistent detection of HCV RNA after this period is an indicator of chronic HCV infection (Fig. 18.12). Assay for anti-HCV antibodies is used as a screening test for HCV, but it can be negative in early infection and in immunosuppressed individuals. After successful treatment, or in cases with spontaneous viral clearance, anti-HCV antibodies may be detected in the absence of HCV RNA, but can disappear with time.

Acute HCV infection is asymptomatic in about 85% of individuals, and severe acute hepatitis is rare. Although a few fortunate people clear the acute infection (for unclear reason), persistent infection and chronic hepatitis are the usual outcome. In contrast with HBV, chronic disease occurs in 80% to 90% of HCV-infected individuals, and approximately 20% of these individuals progress to cirrhosis over 20 to 30 years. Fluctuating elevations of serum aminotransferases are typically present during the chronic phase of the disease. Even those with normal transaminases are at a risk for

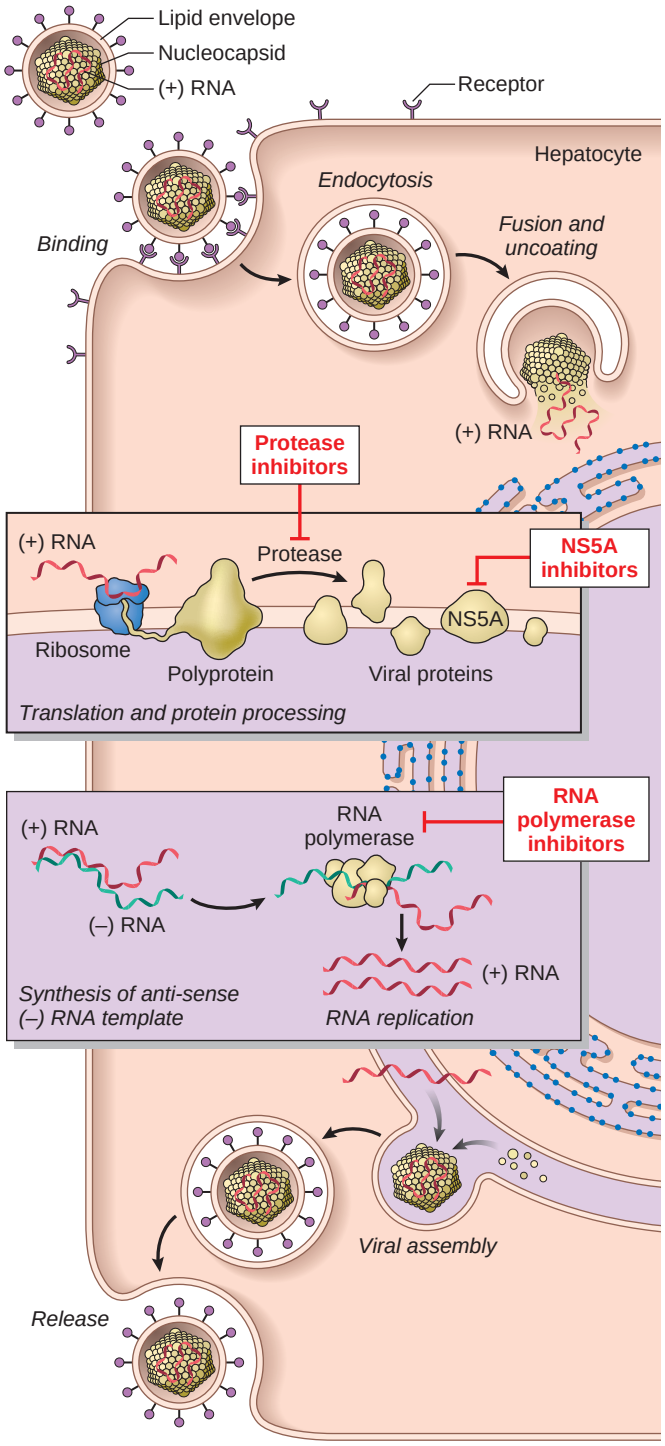


Figure 18.11 Life cycle of hepatitis C. Viral entry, replication, assembly, and budding are shown, emphasizing steps that can be effectively targeted with antiviral drugs.

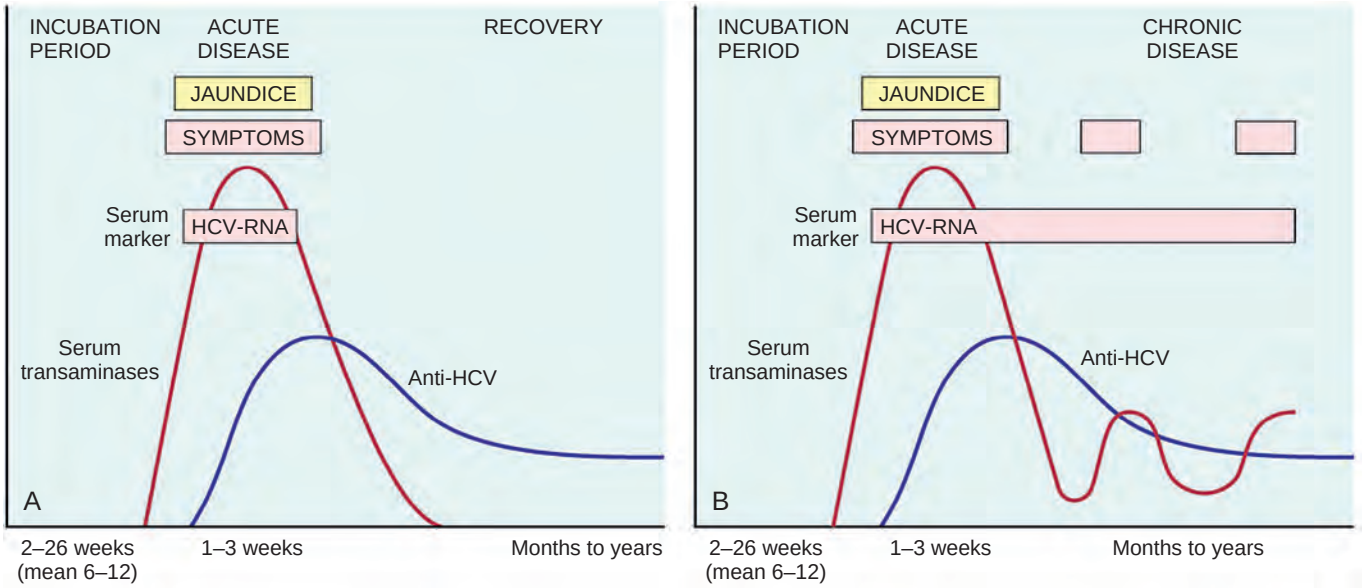


Figure 18.12 Temporal changes in serologic markers in hepatitis C viral infection. (A) Acute infection with resolution. (B) Progression to chronic infection.

developing permanent liver damage, and hence all those with detectable serum HCV RNA levels need treatment and long-term follow-up. Older age, male gender, alcohol use, immunosuppressive drugs, hepatitis B/HIV co-infection, and diseases associated with insulin resistance, including obesity, type 2 diabetes, and metabolic syndrome have been associated with progression. Those who develop cirrhosis are at risk for development of hepatocellular carcinoma. Although the overall risk is small, in the United States, HCV is responsible for about one-third of the cases of liver cancer.

Fortunately, the treatment of hepatitis C has been revolutionized in the past few years by the availability of highly effective antiviral agents. This remarkable advance has been made possible by understanding the role of viral proteins in the virus' life cycle, which in turn has enabled the development of specific drug antagonists (see Fig. 18.11). As in HIV infection, these drugs are used in combination to overcome resistance, and include agents that target the NS3/4a protease, the NS5A replication complex, and the NS5B polymerase. Unlike HIV, HCV does not integrate into the genome where it can lie latent, and as a result true cures are often achieved. Sustained therapeutic response is defined as undetectable HCV RNA after 12 to 24 weeks of drug therapy; overall, such sustained responses are obtained in 80% to 90% of cases, with lower response rates in patients infected with genotype 3 HCV. HCV appears to be cured in >99% of patients who achieve a sustained viral response. The major downside of these advances is their very high cost; a curative course of drug therapy costs over $100,000, and it is estimated that treatment of HCV infections in the United States alone may generate expenses of greater than $50 billion over the next 5 years.

Hepatitis D Virus

Hepatitis D virus (HDV) or "the delta agent," is a unique RNA virus that is dependent for its life cycle on HBV.

Epidemiology

It is estimated that 5% of HBV-infected individuals are co-infected with HDV, which amounts to approximately 15 million people worldwide. The highest prevalence is in the Amazon basin, central Africa, the Middle East, and the Mediterranean basin. HDV is uncommon in Southeast Asia and China. Disease transmission occurs through a parenteral route, often in association with intravenous drug use and multiple blood transfusions.

Pathogenesis

HDV, discovered in 1977, is a 35-nm, double-shelled particle. The external coat antigen of HDV surrounds an internal polypeptide assembly, designated delta antigen (HDAg), the only protein produced by the virus. Associated with the virus particle is a small circular molecule of single-stranded RNA, whose length is the shortest of any known animal virus. Replication of the virus is through RNA-directed RNA synthesis by host RNA polymerase. The basis for disease flares associated with HDV superinfection is not well understood; both cytopathic effects and enhanced cytotoxic host responses have been proposed.

Clinical Features

Infection with HDV arises in two settings, each with somewhat different clinical courses.

- *Co-infection* occurs following exposure to serum containing both HDV and HBV. Co-infection can result in acute hepatitis that is indistinguishable from acute hepatitis B. It is self -limited and is usually followed by clearance of both viruses. However, there is a higher rate of acute hepatic failure in intravenous drug users.
- *Superinfection* occurs when a chronic carrier of HBV is exposed to a new inoculum of HDV. This results in disease 30 to 50 days later, presenting either as severe acute hepatitis in an asymptomatic HBV carrier, or as

an exacerbation of preexisting chronic hepatitis B infection. Chronic HDV infection occurs in more than 80% of superinfections and may have two phases, an acute phase with active HDV replication and suppression of HBV associated with high transaminase levels, and a chronic phase in which HDV replication decreases, HBV replication increases, and transaminase levels fluctuate.

HDV RNA is detectable in the blood and liver just before and in the early days of acute symptomatic disease. IgM anti-HDV antibody is the most reliable indicator of recent HDV exposure, although its appearance is late and frequently short-lived. With chronic delta hepatitis arising from HDV superinfection, HBsAg is present in serum, and anti-HDV antibodies persist for months or longer.

Co-infection with HDV and HBV increases the risk of progression to cirrhosis and HCC. Chronic HDV infection is treated with interferon-γ, but clearance of the virus occurs in only a minority of cases. Newer agents targeting viral entry and replication are in clinical trials. Vaccination for HBV also prevents HDV infection.

Hepatitis E Virus

Hepatitis E virus (HEV) is an enterically transmitted, waterborne infection that occurs primarily in young to middle-aged adults.

Epidemiology

HEV is a zoonotic disease with animal reservoirs, such as monkeys, cats, pigs, and dogs. Epidemics have been reported in Asia and the Indian subcontinent, sub-Saharan Africa, the Middle East, China, and Mexico. HEV infection accounts for more than 30% of cases of sporadic acute hepatitis in India, exceeding the frequency of HAV. Sporadic cases can occur in travelers to these regions, and are also seen in high income nations in association with pig farming or consumption of organ meat.

Pathogenesis

Discovered in 1983, HEV is an unenveloped, positive-stranded RNA virus in the *Hepevirus* genus. The RNA genome is 7.3 kb in length and contains four open reading frames that encode multiple proteins, including a viral protease and viral RNA polymerase. HEV is not cytopathic, and hepatic damage is thought to stem from the host response to cells infected with virus.

Clinical Features

Virions are shed in stool during the acute illness, and infection is typically by the fecal-oral route. The average incubation period following exposure is 4 to 5 weeks, followed by a self-limited acute hepatitis in most cases, with resolution in 2 to 4 weeks. Before the onset of clinical illness, HEV RNA and HEV virions can be detected by PCR in stool and serum. The onset of clinical symptoms, rise of serum aminotransferases, and appearance of IgM anti-HEV are virtually simultaneous. IgM is replaced with persistent IgG anti-HEV antibodies during recovery.

Chronic liver disease or persistent viremia in immunocompetent patients is not observed. However, acute HEV infection is associated with a mortality rate approaching 20% among pregnant women. Chronic HEV infection can occur in the setting of immunosuppression, such as in those with AIDS and in transplant recipients.

Clinicopathologic Syndromes of Viral Hepatitis

Viral hepatitis may pursue several clinical courses: (1) acute asymptomatic infection with recovery; (2) acute symptomatic hepatitis, anicteric or icteric, with recovery; (3) acute liver failure with massive to submassive hepatic necrosis; and (4) chronic hepatitis, with or without progression to cirrhosis. Table 18.3 provides a summary of the salient features of infection by various hepatitis viruses. Acute infections by all of the hepatotropic viruses can either be asymptomatic or symptomatic. HAV and HEV (in immunocompetent

Table 18.3 Hepatitis Viruses

Virus	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis D	Hepatitis E
Type of virus	ssRNA	Partially dsDNA	ssRNA	Circular defective ssRNA	ssRNA
Route of transmission	Fecal-oral (contaminated food or water)	Parenteral, sexual contact, perinatal	Parenteral; intranasal cocaine use	Parenteral	Fecal-oral
Mean incubation period	2–6 weeks	2–26 weeks	4–26 weeks	Same as HBV	4–5 weeks
Frequency of chronic liver disease	Never	5%–10%	>80%	10% (co-infection); 90%-100% for superinfection	In immunocompromised hosts only
Diagnosis	Serum IgM antibodies	HBsAg or HBcAg antibodies; PCR for HBV DNA	ELISA for HCV antibodies; PCR for HCV RNA	Serum IgM and IgG antibodies; PCR for HDV RNA	Serum IgM and IgG antibodies; PCR for HEV RNA

dsDNA, Double-stranded DNA; *ELISA,* enzyme-linked immunosorbent assay; *HBcAg,* hepatitis B core antigen; *HBsAg,* hepatitis B surface antigen; *HBV,* hepatitis B virus; *HCV,* hepatitis C virus; *HDAg,* hepatitis D antigen; *HDV,* hepatitis D virus; *HEV,* hepatitis E virus; *IgG,* immunoglobulin G; *IgM,* immunoglobulin M; *IV,* intravenous; *PCR,* polymerase chain reaction; *ssRNA,* single-stranded RNA.

From Washington K: Inflammatory and infectious diseases of the liver. In Iacobuzio-Donahue CA, Montgomery EA, editors: *Gastrointestinal and Liver Pathology,* Philadelphia, 2005, Churchill Livingstone.

hosts) do not cause chronic hepatitis, and only a small number of HBV-infected adult patients develop chronic hepatitis. In contrast, HCV is notorious for progressing to chronic infection. Acute liver failure is unusual and is seen primarily with HAV, HBV, or HDV infection, depending on region. HEV can cause acute liver failure in pregnant women. Although HBV and HCV are responsible for most cases of chronic hepatitis, many other disorders have similar clinicopathologic features, especially autoimmune and drug/toxin-induced hepatitis, described later. Therefore, serologic and molecular studies are essential for the diagnosis of viral hepatitis and for distinguishing among the various types.

Major features of the main clinicopathologic syndromes associated with hepatropic virus infections are as follows.

- *Acute asymptomatic infection with recovery*. Here, infection is identified due to minimally elevated serum transaminases or, after recovery, by the presence of antiviral antibodies. HAV and HBV infection can be subclinical events, verified only by the presence of anti-HAV or anti-HBV antibodies.
- *Acute symptomatic infection with recovery*. Symptomatic disease can be divided into four phases: (1) incubation period, (2) symptomatic preicteric phase, (3) symptomatic icteric phase, and (4) convalescence. The incubation period for the different viruses is given in Table 18.3. Peak infectivity occurs during the last asymptomatic days of the incubation period and the early days of acute symptoms.
- *Acute liver failure*. Viral hepatitis accounts for approximately 10% of cases of acute hepatic failure. Hepatitis A and E are the most common causes worldwide, while HBV is more common in Asia and the Mediterranean. The treatment is to provide supportive care and allow for replication of residual hepatocytes to lead to restitution of the liver. Liver transplantation is the only option if the disease does not resolve before secondary infection and failure of other organs develops.
- *Chronic hepatitis*. **This is defined as symptomatic, biochemical, or serologic evidence of continuing or relapsing hepatic disease for more than 6 months.** In some patients, persistent elevation of serum transaminases may be the only clinical evidence of chronicity. Prolongation of the prothrombin time, hyperglobulinemia, hyperbilirubinemia, and mild elevations in alkaline phosphatase level can occur. In symptomatic individuals, the most common finding is fatigue; less common symptoms are malaise, loss of appetite, and occasional bouts of mild jaundice. Immune complex disease can develop due to circulating antibody-antigen complexes in chronic HBV and HCV infection, and can manifest as vasculitis, glomerulonephritis, and cryoglobulinemia.
- *Carrier state*. A "carrier" is an individual who harbors and can transmit an organism, but has no symptoms. For hepatotropic viruses, carrier state has been used to described two separate scenarios: (1) individuals who harbor the virus but have no liver disease; and (2) individuals who harbor the virus and have asymptomatic nonprogressive liver damage. In both cases, particularly the latter, affected individuals constitute reservoirs of infection. For HBV infection, the term "healthy carrier" has been used for an individual with HBsAg and anti-HBe, but without HBeAg. These patients have normal aminotransferases, low or undetectable serum HBV DNA, and lack of significant inflammation or liver injury on biopsy (see Fig. 18.11). For HCV, a state equivalent to the HBV "healthy carrier" is not recognized.

HIV is an important comorbid factor in hepatotropic viral infections. Co-infection of HIV and hepatitis viruses is common in the United States, as between 10% and 25% of HIV-infected individuals are also infected with HBV and HCV, respectively. Chronic HBV and HCV infection are leading causes of morbidity and mortality in HIV-infected individuals, although the severity and progression in immunocompetent HIV patients is similar to those who are HIV negative. Untreated HIV infection significantly exacerbates the severity of liver disease caused by HBV or HCV.

MORPHOLOGY

The general morphologic features of viral hepatitis are depicted schematically in Fig. 18.13. The morphologic changes in acute and chronic viral hepatitis due to hepatotropic viruses are largely similar and overlap with those associated with autoimmune hepatitis, adverse drug reactions, and Wilson disease.

In acute viral hepatitis, the liver may be normal in size, enlarged (due to inflammation), or shrunken in those cases that are associated with acute liver failure and massive liver necrosis (see Fig. 18.4). Microscopically, there is a portal and lobular inflammatory infiltrate comprised predominantly of lymphocytes and variably admixed with plasma cells and eosinophils. The hepatocyte injury may result in necrosis or apoptosis (see Figs. 18.2 and 18.3), often with pigmented macrophages scavenging the dead cell debris. Necrosis of groups of hepatocytes (i.e., confluent necrosis) may be seen in severe cases and can progress to necrosis of the entire lobule (i.e., panlobular or panacinar necrosis) or connect vascular structures (i.e., bridging necrosis). Liver failure can develop with massive hepatic necrosis.

The defining histologic feature of chronic viral hepatitis is portal lymphocytic, or lymphoplasmacytic, inflammation with fibrosis. The inflammatory cells often cross the limiting plate and lead to injury of periportal hepatocytes **(interface activity).** This may be accompanied by a variable degree of lobular inflammation. Fibrosis develops with increasing liver damage manifesting initially as portal and periportal fibrosis. Fibrous septa develop and lead to portoportal bridging fibrosis, and eventually cirrhosis.

There are some morphologic features that are distinctive for particular subtypes of chronic viral hepatitis. In chronic hepatitis B, the swollen endoplasmic reticulum of hepatocytes are swollen and filled with HBsAg, leading to a **"ground-glass" appearance**. Immunostaining for hepatitis B surface and core antigens can confirm the HBV infection (Fig. 18.14). Chronic hepatitis C typically shows prominent lymphoid aggregates or fully formed lymphoid follicles in the portal tracts (Fig. 18.15). Steatosis is common in chronic hepatitis C, and can be marked with genotype 3 infection. Bile duct injury can be seen in hepatitis C, mimicking biliary disease, but it is typically a focal finding.

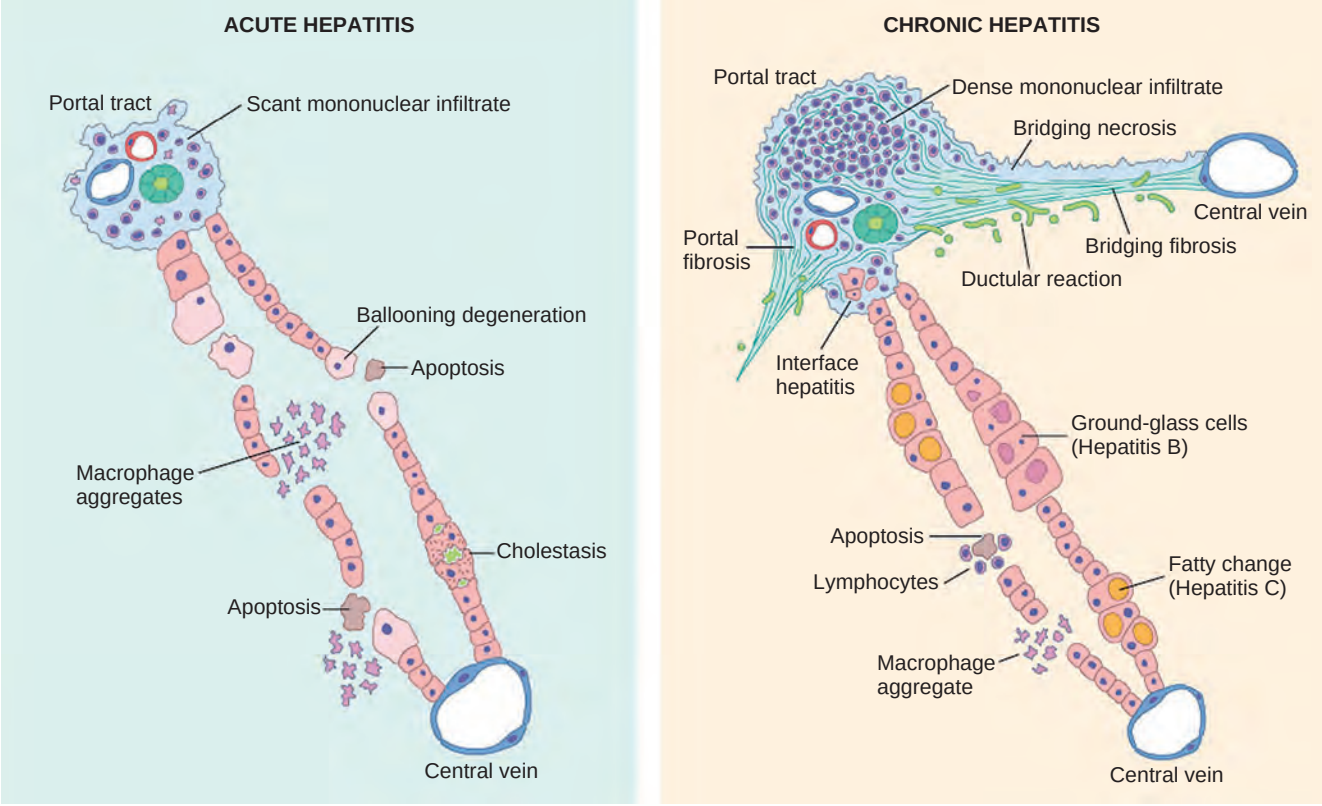


Figure 18.13 Diagrammatic representation of the morphologic features of acute and chronic hepatitis. Acute hepatitis is characterized by lobular inflammation and hepatocellular injury, and chronic hepatitis shows dense portal inflammation. Bridging necrosis can occur in severe acute hepatitis, and fibrosis is seen in chronic hepatitis. Ductular reaction is often present in areas of fibrosis in chronic hepatitis.

Liver biopsy is often performed to confirm the diagnosis of chronic hepatitis and to assess inflammatory activity (grade) and fibrosis (stage). Stage of the disease is used along with other clinical parameters to determine the therapeutic approach.

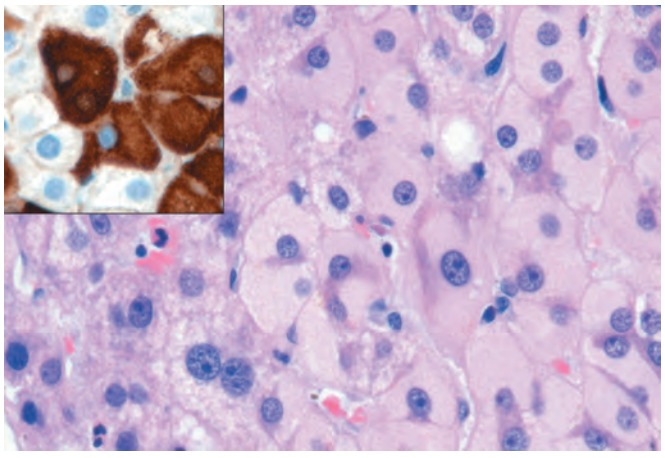


Figure 18.14 Ground-glass hepatocytes in chronic hepatitis B caused by accumulation of hepatitis B surface antigen in the endoplasmic reticulum. The cytoplasmic inclusions are light pink and finely granular on hematoxylin and eosin staining; immunostaining *(inset)* confirms that they contain hepatitis B surface antigen *(brown)*.

KEY CONCEPTS

VIRAL HEPATITIS

- Hepatitis A is transmitted through the fecal-oral route, causes acute hepatitis, and does not lead to chronic liver disease.
- Hepatitis B is transmitted parenterally with most infections being subclinical, but can cause acute hepatitis, chronic hepatitis, and cirrhosis.
- Hepatitis C is most often associated with progression to chronic liver disease (80% or more), while acute infections are almost always subclinical.
- Hepatitis D is a defective virus, requiring hepatitis B co-infection for replication and infection.
- Hepatitis E is endemic in equatorial regions and frequently epidemic; it causes acute disease, which can be severe in pregnancy and can lead to chronic hepatitis in immunocompromised individuals.
- The inflammatory cells in both acute and chronic viral hepatitis are mainly T cells; the morphologic findings overlap with other hepatitic disorders such as autoimmune hepatitis, drug-induced liver injury, and Wilson disease.
- Biopsy assessment in chronic viral hepatitis provides the extent of fibrosis (stage), which can determine the therapeutic course of action.
- Patients with HBV or HCV are at increased risk for development of HCC, especially in the setting of cirrhosis.

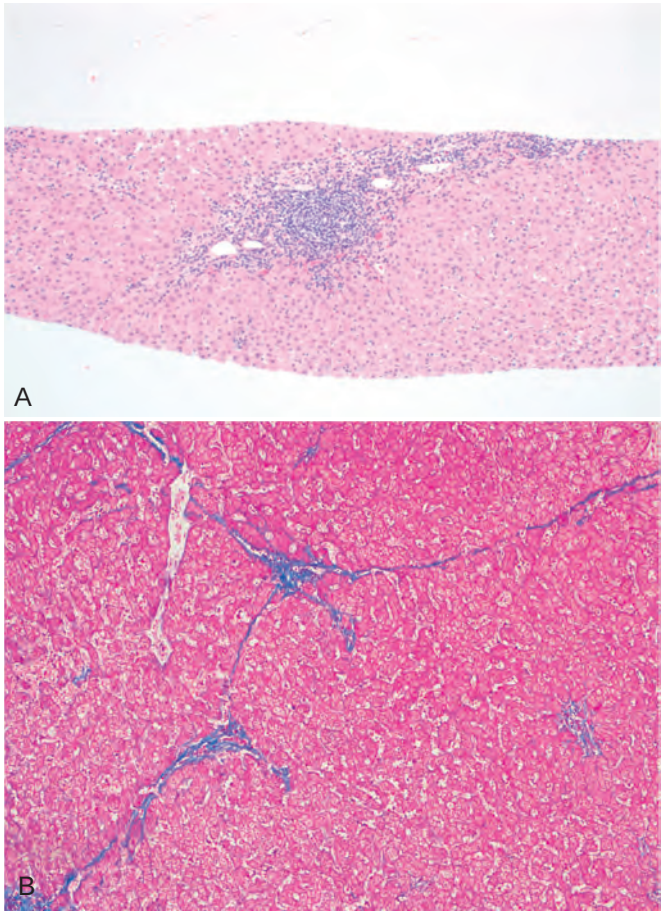


Figure 18.15 Chronic viral hepatitis due to hepatitis C virus. (A) Characteristic portal tract expansion by a lymphoid aggregate. (B) Delicate bridging fibrosis, seen later in the disease course.

Bacterial, Fungal, and Parasitic Infections

Bacteria, fungi, helminths, and other parasites/protozoa can involve the liver and biliary tree as localized infections or as part of a systemic disease. Examples of bacterial infections include *Staphylococcus aureus* in toxic shock syndrome, *Salmonella typhi* in typhoid fever, *Treponema pallidum* in secondary or tertiary syphilis, and *Bartonella henselae* in cat scratch disease. Biliary obstruction creates a milieu for bacterial proliferation leading to infection of the biliary tree, often called ascending cholangitis. When severe, this may extend to the liver and produce intrahepatic abscesses. Spread of bacteria through the hematogenous route or direct spread from adjacent infected tissues can also lead to abscess formation. Liver abscesses are associated with fever, right upper quadrant pain, and tender hepatomegaly. Jaundice may result from extrahepatic biliary obstruction. Antibiotic therapy is generally effective, but surgical drainage may be necessary for large lesions. Extrahepatic bacterial infections, particularly intraabdominal infections, may lead to nonspecific inflammatory changes in the liver. Sepsis can be associated with ductular reaction and cholestasis presenting as bile plugs in ductules (ductular or *cholangiolar cholestasis*), a typical feature of sepsis.

The liver can be involved in disseminated fungal (e.g., histoplasmosis) and mycobacterial infections. The liver histology shows epithelioid granulomas, with or without necrosis, at sites of involvement. The diagnosis rests on serologic findings, blood cultures, or demonstrating the organisms in liver biopsies.

Parasitic and helminthic infections that can involve the liver include malaria, schistosomiasis, strongyloidiasis, cryptosporidiosis, leishmaniasis, echinococcosis, amebiasis, and infections by the liver flukes *Fasciola hepatica, Opisthorchis* species, and *Clonorchis sinensis* (Chapter 8). Schistosomiasis is most commonly found in Asia, Africa, and South America in areas where the water contains numerous freshwater snails as a vector, and it is particularly likely to cause chronic liver disease. Liver fluke infections, most common in Southeast Asia, are notorious for increasing the risk of cholangiocarcinoma (discussed later). Hydatid cysts are usually caused by echinococcal infections (Chapter 8). They often have characteristic calcifications in the cyst walls, which may enable radiologic diagnosis. Hydatid cysts are uncommon in high income countries. Cystic liver degeneration or abscesses can be caused by amebic and other protozoal and helminthic infections. The incidence of amebic infections is low in high income countries and is usually found in immigrants from endemic regions.

AUTOIMMUNE HEPATITIS

Autoimmune hepatitis is a chronic, progressive hepatitis associated with genetic predisposition, autoantibodies, and therapeutic response to immunosuppression. There is a female predominance (78%).

Pathogenesis

Autoimmune hepatitis has a strong association with specific HLA alleles in Caucasians (DR3), Japanese (DR4), and in South Americans (DRB1). Proposed triggers for the immune reaction include viral infections, drugs/toxins, and vaccination, but the antigens that are the target of autoimmunity are unknown, and as with other autoimmune disorders, that basis for the immune attack on hepatocytes remains obscure. The lymphocytic infiltrate in the liver is composed predominantly of CD4+ helper T cells with CD8+ cytotoxic T cells at the interface. CD4+ cells play an important role in activating B lymphocytes and their differentiation into plasma cells, which are responsible for production of autoantibodies. The mechanism by which the interplay of lymphocytes, autoantibodies, and HLA types lead to liver injury is unclear. A small number of drugs such as minocycline, nitrofurantoin, and α-methyl dopa can trigger the formation of autoantibodies and hepatocellular injury that mimics autoimmune hepatitis.

MORPHOLOGY

The features of autoimmune hepatitis overlap with acute and chronic hepatitis of other etiologies. Extensive inflammation and hepatocellular injury at the interface, as well as in the hepatic parenchyma, is characteristic of autoimmune hepatitis. Numerous plasma cells in clusters are typical (Fig. 18.16). Lymphocytes and plasma cells may be seen within the cytoplasm of hepatocytes,

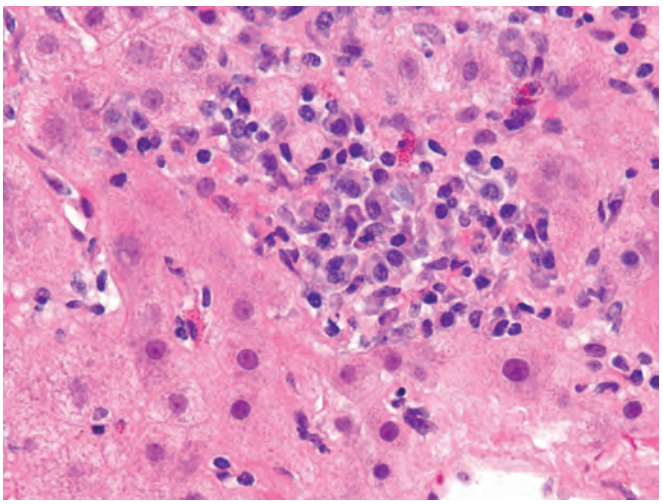


Figure 18.16 Autoimmune hepatitis. A focus of lobular hepatitis with prominent plasma cells typical for this disease is shown.

often at the interface, a curious phenomenon called emperipolesis. Necrosis of groups of hepatocytes (confluent necrosis) can involve the perivenular areas, the entire acinus (panacinar necrosis), or connect vascular structures (bridging necrosis). The resultant regenerative features can manifest as "rosettes," a circular arrangement of hepatocytes around a dilated canaliculus. Most cases have some degree of fibrosis at initial presentation, which increases with disease progression. Cirrhosis with limited inflammatory activity (i.e., "burnt out cirrhosis") is seen at initial presentation in some cases, the end result of subclinical disease.

Clinical Features

Untreated autoimmune hepatitis leads to death or progresses to cirrhosis in a majority of cases. It has a wide range of presentations ranging from asymptomatic disease detected by elevated transaminases during screening, to acute and chronic presentations. Acute presentations may be indistinguishable from acute viral or drug-induced hepatitis and may lead to acute liver failure. More commonly, the presentation is insidious with nonspecific symptoms such as fatigue, anorexia, nausea, and abdominal pain. Patients without symptoms or apparently acute presentations often prove to have fibrosis and even cirrhosis on biopsy, indicating the existence of subclinical disease that went unrecognized.

In a small subset of patients, there may be overlapping features with primary biliary cholangitis or primary sclerosing cholangitis, the latter occurring most often in the pediatric setting. The diagnosis of overlap syndromes requires the unequivocal presence of serologic, biochemical, and/or histologic evidence of both diseases. In some instances, the second component of the overlap syndrome becomes evident months or years after treatment of the first component. Autoimmune hepatitis may also be associated with other autoimmune diseases like type 1 diabetes mellitus, thyroiditis, and celiac sprue. Autoimmune hepatits-like injury induced by drugs is important to recognize because discontinuation of the drug typically leads to clinical recovery.

The diagnosis of autoimmune hepatitis is based on the combination of four features: autoantibodies, elevation of serum IgG, exclusion of other etiologies (e.g., viral hepatitis, drugs), and supportive histologic findings on liver biopsy. These features have been combined to create a scoring scheme, which enables categorization of cases into definite and probable autoimmune hepatitis (Table 18.4). Based on the types of autoantibodies present, autoimmune hepatitis is subclassified as types 1 and 2. *Type 1* is more common and is typically characterized by the presence of antinuclear antibodies (ANAs) and anti–smooth muscle actin (SMA) antibodies. Antibodies against soluble liver antigen/liver-pancreas antigen (SLA/LPA) and, less commonly, anti-mitochondrial antibodies (AMA, more typical of primary biliary cholangitis) also may be seen. *Type 2* disease is more common in children and is characterized by anti–liver kidney microsome-1 (LKM-1) antibodies, which are directed against CYP2D6, and anti–liver cytosol-1 (ACL-1) antibodies. The other clinical features, natural history, and pathologic findings are similar in the two types. Serum IgG is often elevated in autoimmune hepatitis and may provide another clue to the diagnosis.

Autoantibody titers do not correlate well with disease severity or outcome, and may not be present in a small subset of cases (i.e., seronegative autoimmune hepatitis). Autoantibodies also may be present in other disease processes such as steatohepatitis and chronic viral hepatitis (e.g., with HBV and HCV); hence, it is important to make the diagnosis based on overall clinical, serologic, and histologic features.

The treatment of choice is immunosuppression with prednisone, with or without azathioprine, leading to remission in 80% to 90% of patients, generally within 12 months of treatment. Other immunosuppressants are used if the side effects of steroids cannot be tolerated. Patients with incomplete responses or with multiple relapses are at higher risk of progression to cirrhosis and also at risk for developing

Table 18.4 Simplified Diagnostic Criteria (2008) of the International Autoimmune Hepatitis Group

		Points[a]
Autoantibodies	ANA or ASMA or LKM > 1:80	2
	ANA or ASMA or LKM > 1:40	1
	SLA/LP Positive (>20 units)	0
IgG (or gamma-globulins)	>1.10 times normal limit	2
	Upper normal limit	1
Liver histology[b]	Typical for autoimmune hepatitis	2
	Compatible with autoimmune hepatitis	1
	Atypical for autoimmune hepatitis	0
Absence of viral hepatitis	Yes	2
	No	0

[a]Definite autoimmune hepatitis (AIH): N = 7; probable AIH: N = 6.

[b]Typical: (1) Interface hepatitis, lymphocytic/lymphoplasmacytic infiltrates in portal tracts and extending in the lobule; (2) emperipolesis (active penetration by one cell into and through larger cell); (3) hepatic rosette formation.

Compatible: Chronic hepatitis with lymphocytic infiltration without features considered typical.

Atypical: Showing signs of another diagnosis like nonalcoholic fatty liver disease.

AIH, Autoimmune hepatitis; *ANA*, antinuclear antibody; *ASMA*, anti–smooth muscle actin; *IgG*, immunoglobulin G; *LKM*, anti–liver kidney microsomal antibodies; *LP*, liver pancreas; *SLA*, soluble liver antigen.

Modified from Hennes EM, Zeniya M, Czaja AJ, et al: Simplified criteria for the diagnosis of autoimmune hepatitis, *Hepatology* 48(1):169–176, 2008.

hepatocellular carcinoma. Liver transplantation may be necessary for cirrhotic patients. The 10-year survival rate after liver transplant is 75%, with disease recurrence in 20% of patients.

KEY CONCEPTS

AUTOIMMUNE HEPATITIS

- The diagnosis of autoimmune hepatitis is based on a combination of four features: autoantibodies, elevated serum IgG, pathologic findings, and exclusion of viral/drug etiologies.
- The most common autoantibodies in type 1 autoimmune hepatitis are ANAs and anti–smooth muscle antibodies (ASMAs), while type 2 autoimmune hepatitis is characterized by anti-LKM1 autoantibodies.
- Autoimmune hepatitis can have varied presentations: asymptomatic liver enzyme elevation, acute liver failure, chronic hepatitis, and cirrhosis.
- Typical histologic features of autoimmune hepatitis include a high degree of necroinflammatory activity and numerous plasma cells.

DRUG- AND TOXIN-INDUCED LIVER INJURY

Drug or toxin-induced liver injury is a major cause of acute liver failure in the United States. Due to its central role in metabolism, the liver is susceptible to injury by a wide array of compounds found in medicines (Table 18.5), herbal products, dietary supplements, poisonous plants and fungi (e.g., *Amanita phylloides* mushrooms), and household and industrial products (e.g., ointments, perfumes, shampoo, cleaning solvents, pesticides). Based on U.S. Drug Induced Liver Injury Network data, 10% of the patients with drug-induced liver injury die or require liver transplantation, and 17% go on to develop chronic liver disease.

Pathogenesis

Principles of drug and toxic injury are discussed in Chapter 9. Liver injury may develop immediately after exposure to the responsible agent or manifest after weeks or even months. In most instances, the injury is mediated by reactive metabolites generated in the liver. The cytochrome family of enzymes, particularly cytochrome P-450, is involved in most of these metabolic reactions. Since these enzymes are more active in the central zone of the lobule, necrosis of perivenular hepatocytes is a typical feature of drug-induced liver injury. Agents that induce the cytochrome system, such as rifampicin, phenytoin, isoniazid, tobacco smoke, and ethanol, may exacerbate the toxicity of other drugs.

Drug-induced liver injury can be idiosyncratic (unpredictable) or dose-dependent (predictable). Idiosyncratic reactions are the most common form of drug-induced liver injury and usually occur after 1 to 3 months of exposure (see Table 18.5). In some instances, they appear to stem from a hypersensitivity response to the drug or its metabolite(s). It is thought that there is genetic susceptibility in those who have idiosyncratic reactions. For example, the metabolism of isoniazid (an antituberculous agent) is slow in individuals who have a variant of N-acetyltransferase (NAT2) with decreased enzymatic activity, raising drug levels in the liver and leading to susceptibility to isoniazid hepatotoxicity. Antimicrobial drugs are the most common culprits in idiosyncratic reactions, accounting for nearly one-half of cases. Other commonly implicated agents include cardiovascular drugs, central nervous system agents, antineoplastic drugs, and analgesics like nonsteroidal anti-inflammatory

Table 18.5 Patterns of Drug- and Toxin-Induced Hepatic Injury

Pattern of Injury	Morphologic Findings	Examples of Associated Agents
Cholestatic	Bland hepatocellular cholestasis, without inflammation	Contraceptive and anabolic steroids, antibiotics, ART
Cholestatic hepatitis	Cholestasis with lobular necroinflammatory activity; may show bile duct destruction	Antibiotics, phenothiazines, statins
Hepatocellular necrosis	Spotty hepatocyte necrosis	Methyldopa, phenytoin
	Massive necrosis	Acetaminophen, halothane
	Chronic hepatitis	Isoniazid
Fatty liver disease	Large and small droplet fat	Ethanol, corticosteroids, methotrexate, total parenteral nutrition
	"Microvesicular steatosis" (diffuse small droplet fat)	Valproate, tetracycline, aspirin (Reye syndrome), ART
	Steatohepatitis with Mallory hyaline	Ethanol, amiodarone, irinotecan
Fibrosis and cirrhosis	Periportal and pericellular fibrosis	Alcohol, methotrexate, enalapril, vitamin A and other retinoids
Granulomas	Noncaseating epithelioid granulomas	Sulfonamides, amiodarone, isoniazid
	Fibrin ring granulomas	Allopurinol
Vascular lesions	Sinusoidal obstruction syndrome (veno-occlusive disease): obliteration of central veins	High-dose chemotherapy, bush teas
	Budd-Chiari syndrome	Oral contraceptives
	Peliosis hepatis: blood-filled cavities, not lined by endothelial cells	Anabolic steroids, tamoxifen
Neoplasms	Hepatocellular adenoma	Oral contraceptives, anabolic steroids
	Hepatocellular carcinoma	Alcohol, thorotrast
	Cholangiocarcinoma	Thorotrast
	Angiosarcoma	Thorotrast, vinyl chloride

ART, Antiretroviral therapy.

Modified from Washington K: Metabolic and toxic conditions of the liver. In Iacobuzio-Donahue CA, Montgomery EA, editors: *Gastrointestinal and Liver Pathology*, Philadelphia, 2005, Churchill Livingstone.

drugs. A wide variety of herbal and nutritional agents also have been implicated in idiosyncratic reactions.

A classic, dose-dependent, predictable hepatotoxin is acetaminophen, now the most common cause of acute liver failure necessitating transplantation in the United States. The toxic agent is not acetaminophen itself but rather a metabolite, N-acetyl_P-benzquinone immine (NAPQI), produced by the cytochrome P-450 system. Pericentral zone 3 hepatocytes are most sensitive to NAPQI, but in severe overdoses the injury affects all parts of the lobules, resulting in acute hepatic failure. While suicide attempts with acetaminophen are common, so are accidental overdoses. This is because the activity of the cytochrome P-450 system may be up-regulated by other agents taken in combination with acetaminophen, such as alcohol (beware acetaminophen as a hangover prophylactic) or codeine in acetaminophen compound tablets. Other examples of direct-acting hepatotoxins that may produce severe liver injury include organic solvents and toxins in mushrooms.

MORPHOLOGY

Drugs can lead to one or more patterns of injury. Hepatocellular injury accounts for almost one-half of cases, and the remaining are divided approximately equally between cholestatic and mixed hepatocellular/cholestatic patterns. Idiosyncratic reactions with a hepatocellular pattern of injury show typical features of acute hepatitis dominated by inflammation and varying degrees of necrosis. The picture is similar to viral and autoimmune hepatitis, and centrizonal necrosis is a common feature. Progression to chronic hepatitis, and even cirrhosis, can be seen in a small minority of cases. Injury caused by intrinsic hepatotoxins is dominated by necrosis with minimal inflammation. Injury centered on bile ducts is characterized by varying combinations of cholestasis and ductular reactions, which can progress to chronic cholestasis and duct "dropout" in a minority of cases.

Some agents produce other patterns of injury. Drugs like amiodarone (antiarrhythmic), tamoxifen (anti-estrogen), irinotecan (antineoplastic), and methotrexate (immunosuppressive) can cause a **steatohepatitis-like** pattern of injury. Mitochondrial dysfunction caused by drugs such as tetracycline (antibiotic), valproic acid (anticonvulsant), and zidovudine (antiretroviral) can result in **microvesicular steatosis.** Endothelial injury to sinusoids and central veins can be caused by cytotoxic agents (azathioprine, oxaliplatin), leading to **sinusoidal obstruction syndrome** (formerly called veno-occlusive disease).

Clinical Features

Drug-induced liver injury can have a wide variety of presentations and should always be included in the differential diagnosis of liver disease. Since there are no specific clinical or pathologic features, the diagnosis is established based on temporal association of drug or toxin exposure with onset of liver injury. Liver enzyme tests can be used to gauge whether the injury is primarily (1) hepatocellular (alanine aminotransferase [ALT] ≥5 times the upper limit of normal, or ALT/alkaline phosphatase [ALP] ratio >5), (2) cholestatic (ALP ≥2 times the upper limit of normal, or ALT/ALP ratio <2), or (3) mixed (increased ALT and ALP with ALT/ALP ratio between 2 and 5). Exclusion of other etiologies and disease recovery (in most cases) on withdrawal of the offending agent support the presumptive diagnosis. Recurrence with rechallenge with the drug can be confirmatory but is rarely done in practice for obvious reasons.

KEY CONCEPTS

DRUG-INDUCED LIVER INJURY

- There are two mechanisms of liver injury:
  - Direct hepatotoxicity, a dose-dependent phenomenon of the drug, or its metabolite, which predictably affects exposed individuals, a typical example being acetaminophen, which is the most common cause of acute liver failure in the United States.
  - Idiosyncratic (hypersensitivity) response, which is not dose-dependent, accounts for the majority of drug-induced liver injury, but typically occurs in a minority of individuals.
- Drugs can mimic any clinical or histologic pattern of injury and should be included in the differential diagnosis of liver diseases in diverse clinical settings.
- Correlation of temporal profile of drug intake and onset of disease is necessary for diagnosis.

FATTY LIVER DISEASE

Alcoholic liver disease and nonalcoholic fatty liver disease share many similarities and are considered together in this section.

Alcoholic Liver Disease

Excessive alcohol (ethanol) consumption is a major cause of liver disease in most Western countries; it accounts for 5.9% of deaths globally and more often leads to death and disability earlier in life than other forms of chronic liver injury. There are three distinctive, albeit overlapping, forms of alcohol-induced liver injury: (1) steatosis, or fatty change, (2) alchoholic steato-hepatitis, and (3) fibrosis, which leads to cirrhosis.

Pathogenesis

The pharmacokinetics and metabolism of alcohol are described in Chapter 9. Pertinent to this discussion are the detrimental effects of alcohol and its byproducts on hepatocellular function. Short-term ingestion of as little as 80 g of alcohol (six beers or 8 ounces of 80-proof liquor) over one to several days generally produces mild, reversible, hepatic steatosis. The risk of severe hepatic injury becomes significant with intake of 80 g or more of ethanol per day, and daily ingestion of 160 g or more for 10 to 20 years is frequently associated with severe liver injury. However, only 10% to 15% of alcoholics develop cirrhosis. Thus, other factors also influence the development and severity of alcoholic liver disease. These include:

- *Gender.* Although the majority of patients with alcoholic liver disease are men, on a dose-for-dose basis, women are more susceptible to alcohol-induced hepatic injury.

Gender differences in alcohol pharmacokinetics and metabolism may contribute, as well as estrogen-dependent responses of the liver to gut-derived endotoxin (lipopolysaccharide [LPS]). Although exact mechanisms are not known, it appears that estrogen increases gut permeability to endotoxins. These are transported in the portal system to the liver, where they bind the LPS receptor CD14 expressed on Kupffer cells, an event that stimulates Toll-like receptor 4. Toll-like receptor signaling triggers the release of cytokines and chemokines, which may contribute to the inflammation that accompanies alcoholic liver disease.

- *Ethnic and genetic differences.* In the United States, cirrhosis rates are higher for African Americans than for Caucasian Americans despite similar levels of alcohol consumption. Studies with twins suggest that there is a genetic component in alcohol-induced liver disease, although it remains difficult to separate genetic from environmental influences. Genetic variation in alcohol detoxifying enzymes and cytokine promoters may play significant roles and contribute to differences across populations. *ALDH**2, a variant of aldehyde dehydrogenase *(ALDH)* found in 50% of Asians, has a very low enzyme activity. Individuals homozygous for *ALDH**2 are unable to oxidize acetaldehyde and are intolerant of alcohol, which produces upper body flushing and variable levels of nausea and lethargy.
- *Comorbid conditions.* Iron overload, nonalcoholic steatohepatitis, and infection with HCV and HBV synergize with alcohol to increase the severity of liver disease.

Excessive alcohol intake causes steatosis, dysfunction of mitochondria, microtubules and cellular membranes, and oxidative stress, and the resulting injury leads to varying degrees of inflammation and hepatocyte death. Several factors appear to contribute to steatosis. Alcohol metabolism by alcohol dehydrogenase and acetaldehyde dehydrogenase generate large amounts of reduced nicotinamide adenine dinucleotide (NADH). This alters the redox balance in hepatocytes and has a plethora of effects that favor lipogenesis, including suppression of fatty acid oxidation and increased expression of enzymes that carry out fatty acid synthesis. The accumulation of intrahepatic lipids may be further exacerbated by impaired lipoprotein assembly and secretion. Alcohol also enhances peripheral catabolism of fat, increasing the circulating pool of lipids that are available for uptake by hepatocytes.

The exact mechanisms underlying hepatocyte injury and alcoholic hepatitis are uncertain, but several likely contributory factors have been identified.

- *Acetaldehyde* (the product of alcohol dehydrogenase) induces lipid peroxidation and acetaldehyde-protein adduct formation, disrupting cytoskeleton and membrane function and possibly producing neoantigens.
- *CYP2E1 induction.* High levels of alcohol consumption induce liver microsomes that contain CYP2E1, a component of the cytochrome P-450 system. CYP2E1 metabolism of alcohol produces ROS that damage cellular protein, membranes, and mitochondria, effects that may promote apoptosis.
- *Methoinine metabolism.* Alcohol impairs hepatic metabolism of methionine, which decreases glutathione levels, thereby sensitizing the liver to oxidative injury, and contributes to homocysteine production, which may induce the endoplasmic reticulum stress response.

The induction of cytochrome P-450 enzymes in the liver by alcohol also enhances the conversion of other drugs (e.g., acetaminophen) to toxic metabolites. Furthermore, as already mentioned, alcohol has been linked to increased uptake of bacterial endotoxin from the gut, which induces inflammatory responses in the liver.

Damage caused by these and other factors lead to inflammation and, with chronicity, liver fibrosis and deranged vascular perfusion (summarized in Fig. 18.17). In essence, alcoholic liver disease can be a regarded as a maladaptive state in which cells in the liver respond in an increasingly pathologic manner to a stimulus (alcohol) that originally was only marginally harmful.

MORPHOLOGY

Characteristic changes in alcoholic liver disease begin in centrilobular zone 3 and extend outward toward the portal tracts as the injury increases in severity. **Hepatic steatosis (fatty liver)** is an early, predictable effect of alcohol consumption. After even moderate intake of alcohol, lipid droplets accumulate in hepatocytes. Lipid accumulation begins as small droplets that coalesce into large droplets, which distend the hepatocyte and push the nucleus aside (Fig. 18.18). Macroscopically, the fatty liver of chronic alcoholism is enlarged (as heavy as 4 to 6 kg), soft, yellow, and greasy. Steatosis may be separated into microvesicular and macrovesicular forms. **Macrovesicular steatosis** is the predominant form in alcoholic liver disease. An unusual exception is **alcoholic foamy degeneration,** a form of microvesicular steatosis sometimes seen with chronic heavy alcohol use that is associated with endoplasmic reticulum and mitochondrial damage. In general, fatty change is completely reversible if there is abstention from further intake of alcohol.

In a subset of patients with alcoholic liver disease, liver inflammation **(alcoholic hepatitis)** and fibrosis (see Fig. 18.18) are prominent features. The morphologic findings in alcoholic hepatitis include the following:

- *Ballooned hepatocytes* (Fig. 18.19). These are injured, swollen hepatocytes with cleared-out cytoplasm and cytoskeletal damage, which when extensive, results in formation of **Mallory hyaline,** tangled skeins of intermediate filaments such as keratin 8 and keratin 18 that are partially degraded and ubiquitinated (see Fig. 18.19B). Mallory hyaline may reflect a failed attempt to sequester and degrade damaged cytoplasmic proteins. Other ballooned hepatocytes lack Mallory hyaline and contain fat droplets instead. While ballooned hepatocytes are not entirely specific for alcoholic steatohepatitis, they are essential for this diagnosis. Other conditions with ballooned hepatocytes include nonalcoholic steatohepatitis (NASH), Wilson disease, and chronic biliary tract diseases.
- *Inflammation and necrosis.* Neutrophils are generally more prominent in alcoholic hepatitis than in NASH and can satellite around ballooned hepatocytes, particularly when Mallory hyaline is present. Lobular lymphocytic infiltrates are common, and portal lymphocytic infiltrates also may occur, along with increases in macrophages in both portal tracts and lobular parenchyma.

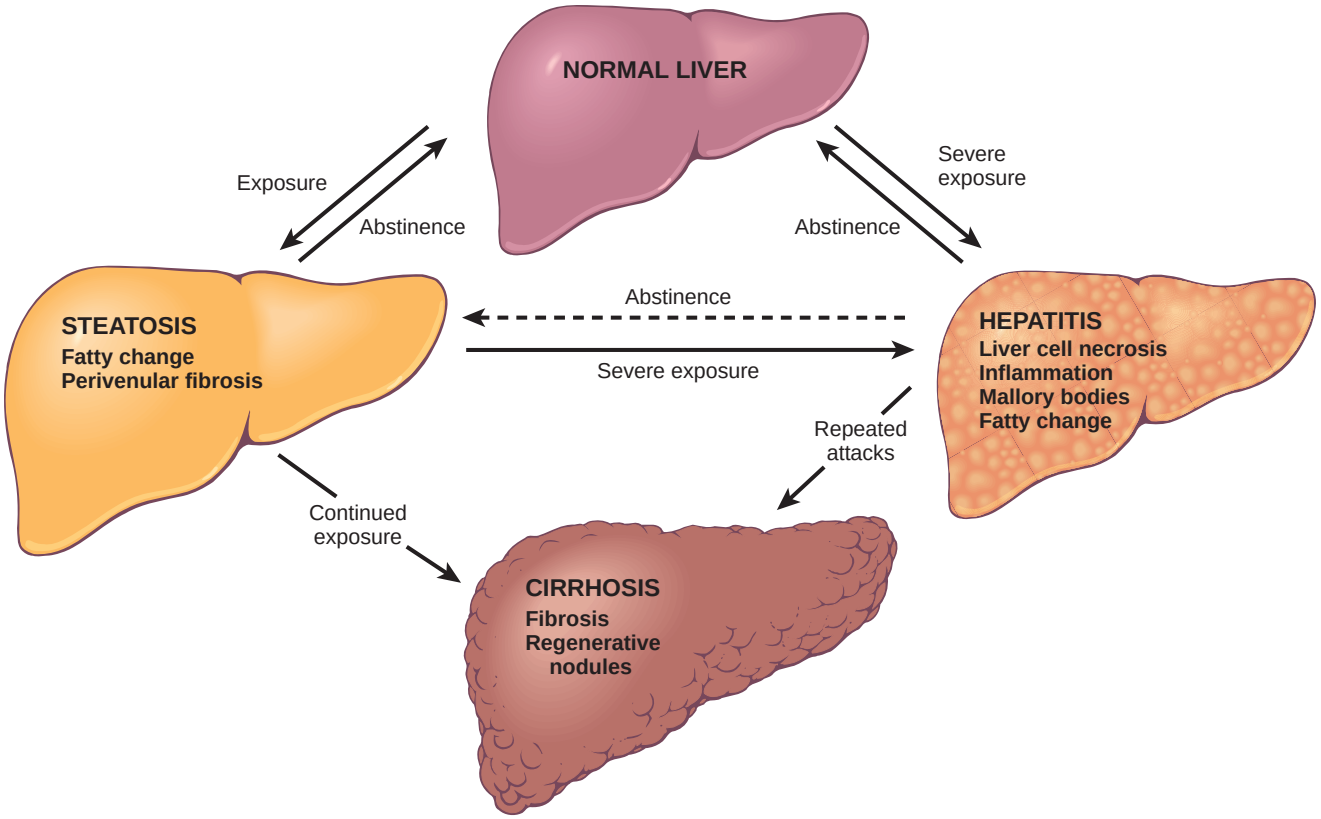


Figure 18.17 Alcoholic liver disease. The interrelationships among hepatic steatosis, alcoholic hepatitis, and alcoholic cirrhosis are shown, along with depictions of key morphologic features. It should be noted that some patients present initially with cirrhosis without any of the other forms of alcoholic liver disease.

Necrosis/apoptosis is usually spotty, but more prominent hepatocellular injury and confluent necrosis can occur in some cases.

- *Perivenular/pericellular fibrosis.* Steatohepatitis is often accompanied by fibrosis. This usually starts in acinar Zone 3 as pericellular, or perisinusoidal, fibrosis with a "chicken wire" appearance (see Fig. 18.18). With continued damage, this progresses to portal/periportal fibrosis, and then to bridging fibrosis and cirrhosis, which in many cases is micronodular ("Laennec cirrhosis") (Fig. 18.20). Early stages of scarring can regress with cessation of alcohol use, but with the development of cirrhosis and its vascular derangements (related to perivenular fibrosis and fibrous obliteration, termed phlebosclerosis, and veno-occlusive lesions), the chances of full restoration of normal function diminish. Complete regression of alcoholic cirrhosis, although reported, is rare.

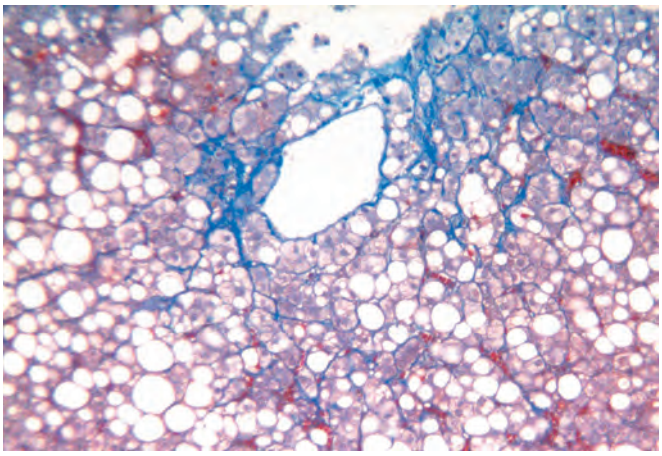


Figure 18.18 Alcoholic steatosis and fibrosis. A mix of small and large fat droplets (seen as clear vacuoles) is most prominent around the central vein and extends outward to the portal tracts. Some fibrosis *(stained blue)* is present in a characteristic perisinusoidal chicken wire pattern. (Masson trichrome stain). (Courtesy Dr. Elizabeth Brunt, Washington University, St. Louis, Mo.)

Clinical Features

Hepatic steatosis may cause hepatomegaly, with mild elevation of serum bilirubin and alkaline phosphatase levels. Severe hepatic dysfunction is unusual. Alcohol withdrawal and consumption of an adequate diet are sufficient treatment. In contrast, alcoholic hepatitis tends to appear acutely, usually following a bout of heavy drinking. Typically, there is malaise, anorexia, weight loss, upper abdominal discomfort, and tender hepatomegaly, accompanied by the laboratory findings of hyperbilirubinemia, elevated serum aminotransferases and alkaline phosphatase, and often a neutrophilic leukocytosis. In contrast to other chronic liver diseases, in which serum ALT tends to be higher than serum AST, in alcoholic liver disease serum AST tend to be higher

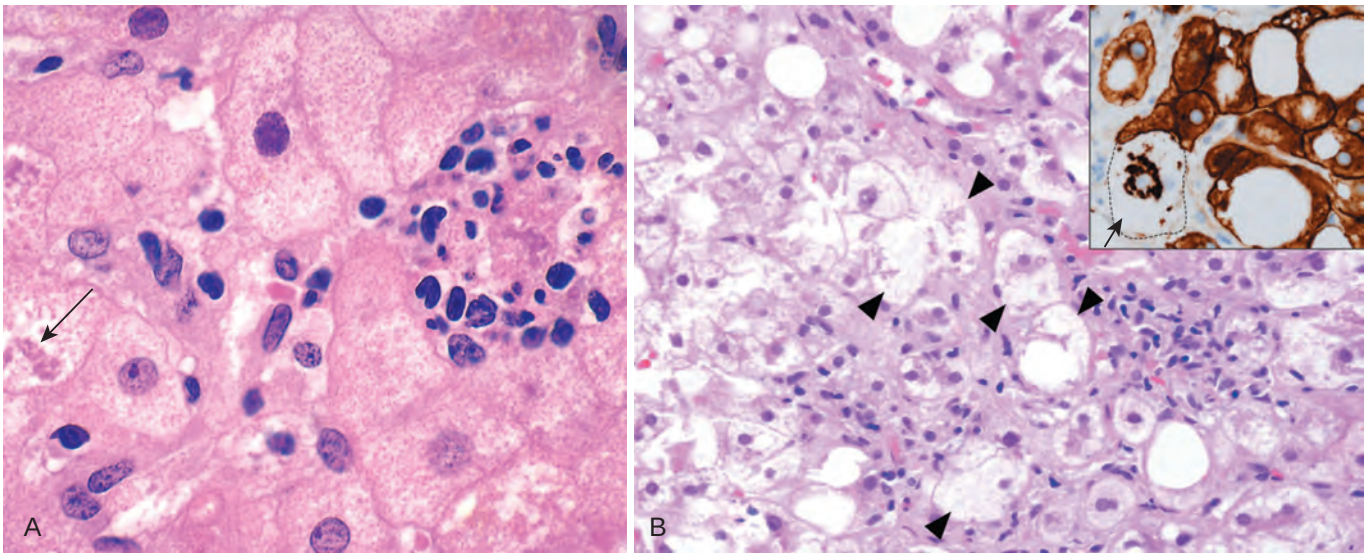


Figure 18.19 (A) Alcoholic hepatitis with clustered inflammatory cells marking the site of a necrotic hepatocyte. Mallory hyaline is present in another hepatocyte *(arrow).* (B) Alcoholic hepatitis with many ballooned hepatocytes *(arrowheads).* Clusters of inflammatory cells are also present; *inset* shows immunostaining for keratins 8 and 18 *(brown),* with most hepatocytes, including those with fat vacuoles, showing normal cytoplasmic staining, but in the ballooned cell *(arrow)* the ubiquinated keratins are collapsed into Mallory hyaline, leaving the cytoplasm "empty." (Courtesy Dr. Elizabeth Brunt, Washington University, St. Louis, Mo.)

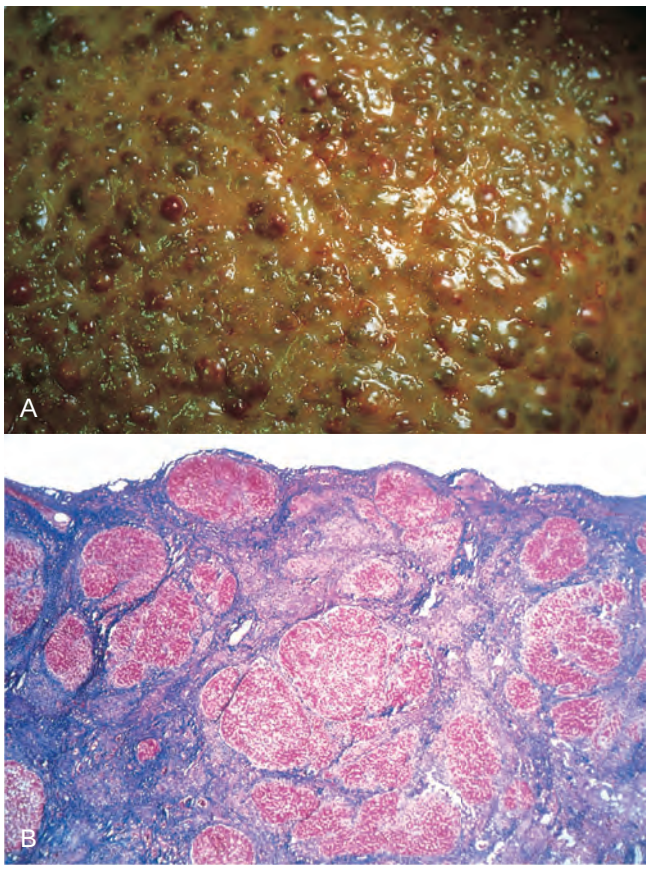


Figure 18.20 Alcoholic cirrhosis. (A) The characteristic diffuse nodularity of the surface is induced by the underlying fibrous scarring. The average nodule size is 3 mm in this close-up view, typical of the "micronodular" cirrhosis of alcoholic liver disease. The greenish tint is caused by cholestasis. (B) Microscopically, this cirrhosis is marked by small nodules entrapped in blue-staining fibrous tissue; fatty accumulation is no longer seen in this "burned-out" stage (Masson trichrome stain).

than serum ALT levels by a ratio of 2:1 or greater. This finding can be particularly helpful in the setting of occult alcoholism. On the other end of the spectrum, in severe cases symptoms and laboratory findings may mimic those of acute liver failure. In other instances, an acute cholestatic syndrome picture is seen that resembles large bile duct obstruction.

The outlook is unpredictable; each episode of hepatitis incurs about a 10% to 20% risk of death. With repeated bouts, cirrhosis develops in about one-third of patients within a few years. Alcoholic hepatitis also may be superimposed on established cirrhosis. With proper nutrition and total cessation of alcohol consumption, alcoholic hepatitis may clear. However, in some patients, the hepatitis persists despite abstinence, and progression to cirrhosis is seen.

The manifestations of alcoholic cirrhosis are similar to those of other forms of cirrhosis. Laboratory findings reflect hepatic dysfunction and include elevated serum aminotransferases, hyperbilirubinemia, variable elevation of serum alkaline phosphatase, hypoproteinemia (globulins, albumin, and clotting factors), and anemia. In some instances, liver biopsy may be indicated, since in about 10% to 20% of cases of presumed alcoholic cirrhosis, another disease process is found. Finally, cirrhosis may be clinically silent, discovered only at autopsy or when a stress such as infection or trauma tips the balance toward hepatic insufficiency.

The long-term outlook for alcoholics with liver disease is variable. The 5-year survival rate approaches 90% in abstainers who are free of jaundice, ascites, or hematemesis; it drops to 50% to 60% in those who continue to imbibe. In advanced disease, common proximate causes of death include (1) hepatic coma, (2) massive gastrointestinal hemorrhage, (3) intercurrent infection (to which these patients are predisposed), (4) hepatorenal syndrome (often following a bout of alcoholic hepatitis), and (5) hepatocellular carcinoma (the risk of developing this tumor in alcoholic cirrhosis is 1% to 6% annually).

KEY CONCEPTS

ALCOHOLIC LIVER DISEASE

- Alcoholic liver disease is a chronic disorder that can give rise to steatosis, alcoholic hepatitis, progressive fibrosis, and marked derangement of vascular perfusion leading eventually to cirrhosis.
- Consumption of 80 g/day of alcohol is considered to be the threshold for development of alcoholic liver disease, but may be lower in women.
- It may take 10 to 15 years of drinking for development of cirrhosis, which occurs only in a small proportion of chronic alcoholics.
- The pathologic effects of alcohol on hepatocytes include changes in lipid metabolism related to altered redox potential, injury caused by ROS generated by metabolism of alcohol by the P450 system, and protein adducts formed by acetaldehyde, a major metabolite of alcohol.

Table 18.6 World Health Organization Criteria for the Metabolic Syndrome

One of:	Diabetes mellitus or impaired glucose tolerance or impaired fasting glucose or insulin resistance
and two of:	Blood pressure: ≥140/90 mm Hg Dyslipidemia: triglycerides (TG): ≥1.695 mmol/L and high-density lipoprotein cholesterol (HDL-C) ≤0.9 mmol/L (male), ≤1 mmol/L (female) Central obesity: waist-hip ratio >0.90 (male); >0.85 (female), or body mass index >30 kg/m^2 Microalbuminuria: urinary albumin excretion rate of ≥20 μg/min or albumin-to-creatinine ratio ≥30 mg/g

Nonalcoholic Fatty Liver Disease (NAFLD)

NAFLD is defined as the presence of hepatic steatosis (fatty liver) in individuals who do not consume alcohol or do so in small quantities and who do not have another cause of secondary hepatic fat accumulation (e.g., HCV, Wilson disease, medications). NAFLD is associated with obesity, diabetes mellitus type 2, and hyperlipidemia, all components of the metabolic syndrome (Table 18.6). It has become the most common cause of chronic liver disease in the United States and is projected to exceed 30% prevalence in the adult population by 2030. The term *nonalcoholic steatohepatitis* (or its common acronym *NASH*) is reserved for NAFLD patients who demonstrate steatohepatitic injury with histologic features similar to those seen with alcoholic hepatitis. The diagnosis of NASH confers an increased risk of developing advanced fibrotic liver disease, and the rise in NAFLD prevalence is expected to produce a concomitant increase in the incidence of NASH and its serious complications (e.g., decompensated cirrhosis, hepatocellular carcinoma).

Pathogenesis

The precise mechanisms underlying NAFLD are unknown. The strong association with insulin resistance suggests that this factor is particularly important in disease development. However, even among those with insulin resistance, there is significant variability in the severity of NAFLD, and complex associations with genetic variants, diet, and the intestinal microbiome have also been suggested. As in alcoholic liver disease, there is an established association between increased gut-derived endotoxin production and liver inflammation and injury. High-fructose diets have also been associated with increased risk of NAFLD-related fibrosis, and dietary fat, particularly trans-fat, may have a role in producing liver injury. Obstructive sleep apnea, usually occurring in the setting of obesity, has been associated with disease progression, possibly related to intermittent hypoxia. Fibrosis may be accelerated when injury from another liver disease (e.g., hemochromatosis) is superimposed on NAFLD.

In individuals with established insulin resistance and metabolic syndrome, visceral adipose tissue not only increases in mass but also becomes dysfunctional. Resistance to insulin leads to increased release of free fatty acids from adipocytes due to overactivity of lipoprotein lipase (Chapter 24). This is associated with reduced production of the hormone adiponectin from adipocytes, which decreases oxidation of free fatty acids by skeletal muscle and increases free fatty acid uptake into hepatocytes (Chapter 9), where the fatty acids are stored as triglycerides. The dysfunctional adipocytes also synthesize pro-inflammatory cytokines such as TNF-α. At the same time, there is evidence that hepatocytes in NAFLD down-regulate lipolysis, an alteration that may further contribute to the accumulation of lipids. Lipolysis occurs through a mechanism called lipophagy that closely resembles macroautophagy, which you will recall is the process by which cells remove excessive or dysfunctional cellular components and (in times of starvation) generate metabolites for energy production.

The net effect of these changes is to cause lipid to accumulate in hepatocytes. Fat-laden cells are highly sensitive to lipid peroxidation products generated by oxidative stress, which can damage mitochondrial and plasma membranes, potentially leading to apoptosis or necrosis. These tendencies may be exacerbated by the pro-inflammatory state that accompanies insulin resistance. Once cell injury becomes established, release of cytokines such as TNF-α and TGF-β locally from Kupffer cells leads to the activation of stellate cells, collagen deposition, and scarring.

MORPHOLOGY

NASH shares many morphologic features with alcoholic hepatitis; steatosis (≥5% of hepatocytes), lobular inflammation, and ballooned hepatocytes are required for its diagnosis. It is not possible to reliably distinguish alcoholic hepatitis from NASH based on histologic findings, though alcoholic hepatitis has on average less steatosis and more ballooned hepatocytes, lobular inflammation, Mallory hyaline, neutrophilic infiltrates, cholestasis, and obliterated central veins (Fig. 18.21).

Determination of the extent of fibrosis is important for clinical management. Fibrosis typically develops around the central vein as a fine "spider web" of pericellular collagen deposition (also

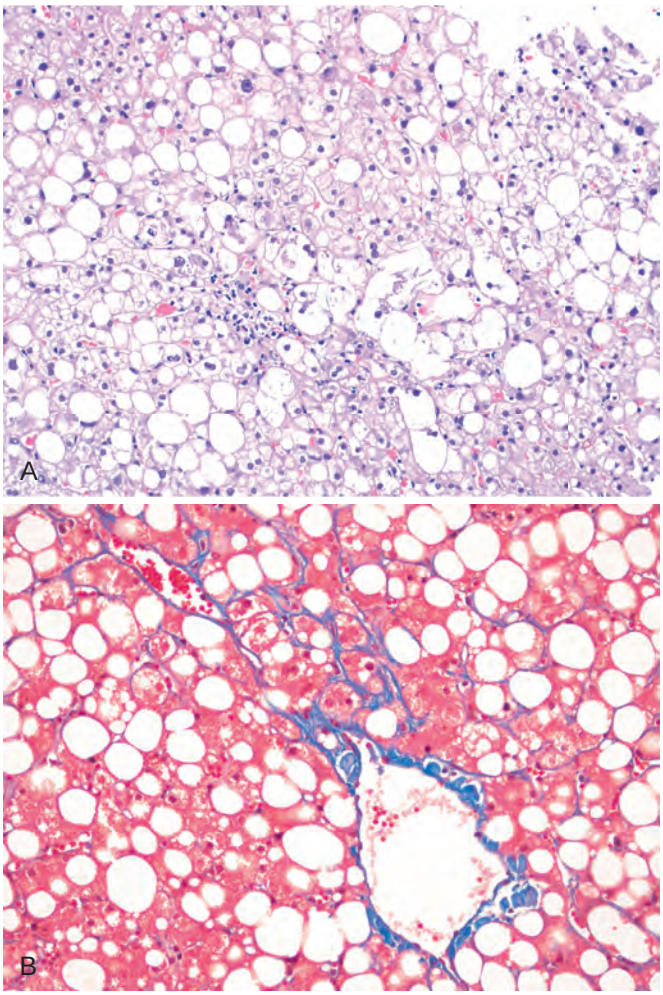


Figure 18.21 Nonalcoholic fatty liver disease. (A) Liver with mixed small and large fat droplets as well as ballooned hepatocytes. (B) Steatosis and fibrosis extending along sinusoids in a chicken wire pattern in which individual and clustered hepatocytes are surrounded by thin scars *(blue fibers)*. Note the resemblance to alcoholic hepatitis depicted in Fig. 18.18. (Masson trichrome stain.)

referred to as a chicken wire pattern), which may only be appreciated with a trichrome stain. Progression of fibrosis usually manifests as periportal fibrosis, followed by bridging fibrosis and cirrhosis. Cirrhosis is often subclinical for years, and, when established, the steatosis or ballooned hepatocytes may be reduced or absent. More than 90% of cases labeled as "cryptogenic cirrhosis" (i.e., cirrhosis of unknown cause) are thought to be due to NASH (i.e., "burned out NASH").

Pediatric NAFLD differs significantly from adult NAFLD. Typically, children show more diffuse steatosis and portal (rather than central) fibrosis, and ballooned hepatocytes may not be present.

Clinical Features

The varied clinical course of individuals with NAFLD is summarized in Fig. 18.22. Individuals with only steatosis are generally asymptomatic. Clinical presentation is often related to other signs and symptoms of the metabolic syndrome, in particular insulin resistance or diabetes mellitus. Imaging studies may reveal fat accumulation in the liver. Liver biopsy is required for diagnosis of NASH and aids in assessment of fibrosis. Viral, autoimmune, and other metabolic diseases of the liver must be excluded, and biopsy may be helpful when there is more than one potential etiology for liver injury. Serum AST and ALT are elevated in most patients with NASH. Despite the enzyme elevations, patients may be asymptomatic. Others have nonspecific symptoms such as fatigue, or complain of right-sided abdominal discomfort caused by hepatomegaly. Because of the association with metabolic syndrome, cardiovascular disease is a frequent cause of death in patients with NASH. NASH also increases the risk of hepatocellular carcinoma, as do other metabolic diseases (discussed later).

The goal of treating individuals with NASH is to reverse the histologic features of disease and prevent or reverse fibrosis by correcting underlying risk factors, such as obesity and hyperlipidemia, and to treat insulin resistance. Weight loss, diet, and exercise can potentially reverse the histologic

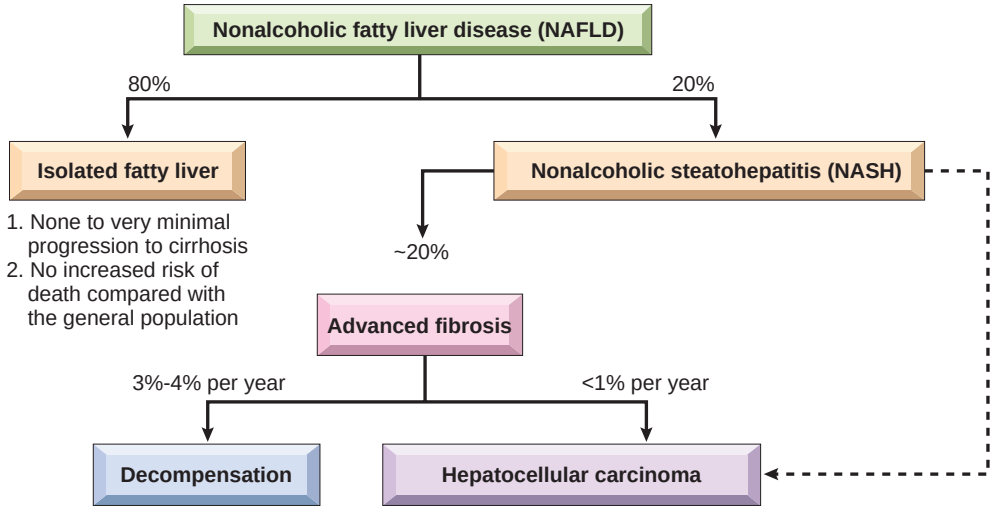


Figure 18.22 Natural history of nonalcoholic fatty liver disease phenotypes. Isolated fatty liver, shows minimal risk for progression to cirrhosis or increased mortality, and nonalcoholic steatohepatitis shows increased overall mortality as well as increased risk for cirrhosis and hepatocellular carcinoma.

abnormalities in NASH. There are numerous ongoing clinical trials aimed at developing pharmacologic approaches for treating NASH and its complications.

KEY CONCEPTS

NONALCOHOLIC FATTY LIVER DISEASE

- The most common metabolic liver disorder is NAFLD, which is associated with metabolic syndrome, obesity, type 2 diabetes mellitus, and hyperlipidemia.
- NAFLD may show all the histologic changes associated with alcoholic liver disease (e.g., steatosis, steatohepatitis, and steatofibrosis). The diagnosis of NASH requires biopsy, and NASH cannot be reliably distinguished from alcoholic hepatitis without clinical history.
- Pediatric NAFLD is being increasingly recognized as the obesity epidemic involves pediatric age groups. Its histologic features differ somewhat from those seen in adults.

INHERITED LIVER DISEASE

Among inherited metabolic diseases, hemochromatosis, Wilson disease, and α_1-antitrypsin deficiency are most prominent and are discussed in the following sections.

Hemochromatosis

Hemochromatosis is caused by excessive iron absorption, most of which is deposited in the liver and pancreas, followed by the heart, joints, and endocrine organs. When hemochromatosis results from an inherited disorder, it is referred to as *hereditary hemochromatosis*. When accumulation occurs as a consequence of parenteral administration of iron, usually in the form of transfusions, or other causes, it is called *secondary hemochromatosis*. The classification of the various causes of iron overload is shown in Table 18.7.

Table 18.7 Classification of Iron Overload

I. Hereditary Hemochromatosis
Mutations of genes encoding HFE, transferrin receptor 2 (TfR2), or hepcidin
Mutations of genes encoding HJV (hemojuvelin: juvenile hemochromatosis)
II. Hemosiderosis (Secondary Hemochromatosis)
Parenteral iron overload due to red cell transfusions
Severe chronic hemolytic anemias (e.g., sickle cell disease)
Severe forms of thalassemia
Bone marrow failure (e.g., aplastic anemia)
Conditions associated with increased iron uptake
β-Thalassemia
Myelodysplastic syndrome
Increased oral intake of iron
African iron overload (Bantu siderosis)
Congenital atransferrinemia
Chronic liver disease
Alcoholic liver disease
Porphyria cutanea tarda
Neonatal hemochromatosis[a]

[a]Neonatal hemochromatosis develops in utero but does not appear to be a hereditary condition.

As discussed in Chapter 14, the total body iron pool ranges from 2 to 6 g in normal adults; about 0.5 g is stored in the liver, 98% of which is in hepatocytes. In severe hemochromatosis, the total body iron pool may exceed 50 g, more than one-third of which accumulates in the liver. The following features characterize severe iron overload:

- Fully developed cases exhibit (1) micronodular cirrhosis (all patients); (2) diabetes mellitus (75% to 80% of patients); and (3) abnormal skin pigmentation (75% to 80% of patients).
- Iron accumulation in hereditary forms is lifelong, but the injury caused by excessive iron is slow and progressive; hence symptoms usually first appear in the fourth to fifth decades of life in men and later in women since menstrual bleeding counterbalances the accumulation until menopause.
- Because many women do not accumulate clinically relevant amounts of iron within their lifetime, hereditary hemochromatosis affects more males than females (ratio of 5 to 7:1).

Pathogenesis

Because there is no regulated iron excretion from the body, the total body content of iron is tightly regulated by intestinal absorption. In hereditary hemochromatosis, regulation of intestinal absorption of dietary iron is abnormal, leading to net iron accumulation of 0.5 to 1 g/year. The disease typically manifests after 20 g of stored iron has accumulated. Mechanisms of liver injury include (1) lipid peroxidation via iron-catalyzed free radical reactions, (2) stimulation of collagen formation by activation of hepatic stellate cells, and (3) interaction of ROS and iron itself with DNA, leading to lethal cell injury and predisposition to hepatocellular carcinoma.

The main regulator of iron absorption is the protein hepcidin, encoded by the *HAMP* gene and produced and secreted by the liver (Fig. 18.23). Transcription of *HAMP* is increased by inflammatory cytokines and iron, and decreased by iron deficiency, hypoxia, and erythroferrone, a hormone produced by marrow erythroblasts. These inputs are integrated to regulate the synthesis and plasma levels of hepcidin (see Chapter 14). Hepcidin binds to the cellular iron efflux channel ferroportin, causing its internalization and proteolysis, thereby inhibiting the release of iron from intestinal cells and macrophages. Due to these activities, an increase in hepcidin lowers plasma iron levels. Conversely, an abnormal deficiency of hepcidin causes iron overload. Other proteins involved in iron metabolism do so by regulating hepcidin levels. Decreased hepcidin synthesis or activity may be caused by loss-of-function mutations in the genes *HAMP*, *HJV*, *TFR2*, and *HFE*, all of which are associated with hereditary forms of hemochromatosis. Deficiencies of hepcidin also occur when erythroferrone levels are chronically elevated, as occurs in disorders marked by ineffective hematopoiesis, such as β-thalassemia and myelodysplastic syndrome.

The adult form of hereditary hemochromatosis is usually caused by mutations of *HFE;* mutation of *TFR2* is a rare cause. Mutations in *HAMP* and *HJV* genes are much less common and give rise to juvenile forms of hereditary hemochromatosis. *HFE* encodes an HLA class I–like molecule that governs intestinal absorption of dietary iron by regulating

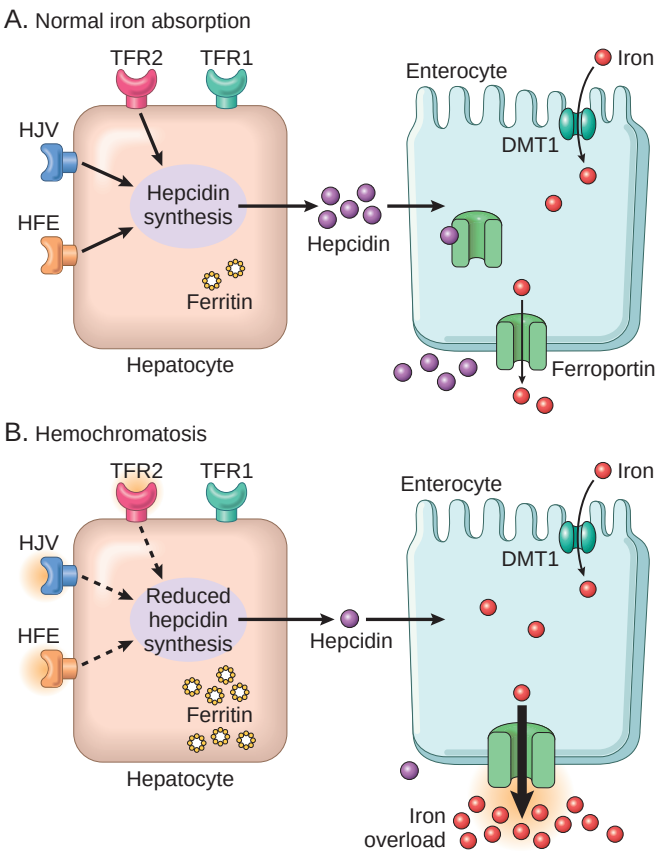


Figure 18.23 Normal iron metabolism and altered iron metabolism in hereditary hemochromatosis. (A) In normal state, HFE, HJV, and TFR2 regulate hepcidin synthesis by hepatocytes maintaining normal circulating hepcidin levels. Hepcidin binds to ferroportin on enterocytes leading to internalization of the complex and ferroportin degradation. This in turn reduces efflux of iron from enterocytes. Through these regulatory interactions, normal iron absorption is maintained. (B) In hereditary hemochromatosis, *HFE, HJV,* or *TFR2* gene mutations reduce hepcidin synthesis, diminishing circulating hepcidin levels. Due to decreased hepcidin-ferroportin interaction, ferroportin activity and iron efflux from enterocytes increases, giving rise to systemic iron overload.

hepcidin synthesis. The most common *HFE* mutation produces an inactivating cysteine-to-tyrosine substitution at amino acid 282 (C282Y) and is present in 70% or more of patients diagnosed with hereditary hemochromatosis. The other common disease-associated *HFE* mutation results in a H63D (histidine at position 63 to aspartate) substitution.

The C282Y mutation of *HFE* is largely confined to Caucasian populations of European origin, while the H63D mutation has a worldwide distribution. The frequency of C282Y homozygosity is 0.45% (1 in 220 persons), and the heterozygous frequency is 11%, making hereditary hemochromatosis one of the most common genetic disorders in humans. The penetrance of the disorder is low in patients with the homozygous C282Y mutation, and is even lower in homozygous H63D and C282Y/H63D compound heterozygous individuals.

MORPHOLOGY

Severe hemochromatosis (hereditary or secondary) is characterized by (1) **deposition of hemosiderin** in the following organs (in decreasing order of severity): liver, pancreas, myocardium, pituitary gland, adrenal gland, thyroid and parathyroid glands, joints, and skin; (2) cirrhosis; and (3) pancreatic fibrosis. In the **liver,** iron becomes evident first as golden-yellow hemosiderin granules in the cytoplasm of periportal hepatocytes that stain with Prussian blue (Fig. 18.24). With increasing iron load, there is progressive deposition in the rest of the lobule, along with bile duct epithelium and Kupffer cells. Iron is a direct hepatotoxin, and inflammation is characteristically absent. In early stages of disease, the liver is typically slightly larger than normal, dense, and chocolate brown. Fibrous septa develop slowly, leading ultimately to a small, shrunken liver with a micronodular pattern of cirrhosis. The liver parenchyma in later stages is often dark brown to nearly black due to extensive iron accumulation.

Other organs that are particularly susceptible to the toxic effects of iron also show morphologic changes. The **pancreas** becomes intensely pigmented and may undergo parenchymal atrophy associated with interstitial fibrosis. The **heart** is often enlarged and exhibits hemosiderosis, producing a striking brown coloration. Both organs may develop fibrosis. Skin pigmentation predominantly results from increased epidermal melanin production, the mechanism of which is unknown. The combination of these pigments imparts a characteristic slate-gray color to the skin. With hemosiderin deposition in joint linings, an **acute synovitis** may develop. Excessive deposition of calcium pyrophosphate damages the articular cartilage, producing a disabling polyarthritis referred to as **pseudogout.** The **testes may be small and atrophic,** secondary to pituitary involvement and reduced gonadotropin and testosterone levels.

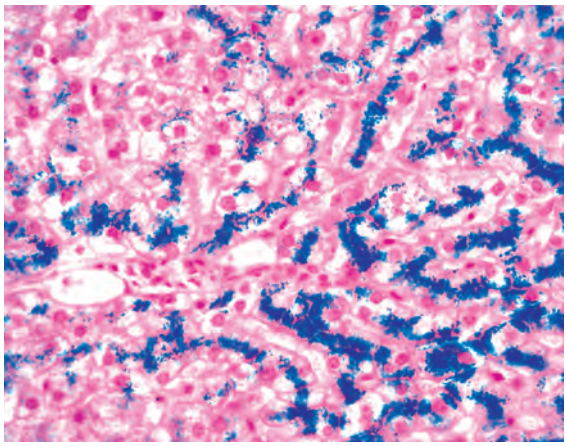


Figure 18.24 Hereditary hemochromatosis. In this Prussian blue–stained section, hepatocellular iron appears blue. The parenchymal architecture is normal.

Clinical Features

The principal manifestations of hemochromatosis include hepatomegaly, abdominal pain, abnormal skin pigmentation (particularly in sun-exposed areas), deranged glucose homeostasis or diabetes mellitus (due to destruction of pancreatic islets), cardiac dysfunction (arrhythmias, cardiomyopathy), and atypical arthritis. In some patients, the presenting complaint is hypogonadism (e.g., amenorrhea in the female, impotence and loss of libido in the male). It is more often a disease of males, for reasons described earlier,

and rarely becomes evident before 40 years of age. The classic tetrad of cirrhosis with hepatomegaly, abnormal skin pigmentation, diabetes mellitus, and cardiac dysfunction may not develop until late in the course. Death may result from cirrhosis or cardiac disease.

Another significant cause of death is hepatocellular carcinoma; the risk is 200-fold greater than in the general population. Treatment that reduces iron overload does not fully remove the cancer risk, presumably because DNA damage occurs prior to diagnosis and treatment initiation.

Fortunately, hemochromatosis is often diagnosed before irreversible tissue damage has occurred. Currently most patients with hemochromatosis are identified in the subclinical, precirrhotic stage due to routine serum iron measurements (as part of other diagnostic workups), and the diagnosis is confirmed by DNA sequencing and detection of causative mutations, usually in *HFE*. Further evaluation includes exclusion of secondary causes of iron overload. A liver biopsy may also provide useful information (e.g., about the presence of fibrosis), but is not done routinely, as genetic testing and imaging have negated the need for quantitative assessment of tissue iron content. Screening of the family members of probands is important, as identification of asymptomatic carriers can prevent disease development.

Prevention of the development or progression of disease in patients with adult onset hereditary is remarkably simple, as regular phlebotomy steadily depletes tissue iron stores. With treatment, life expectancy is normal.

The most common causes of secondary (or acquired) hemochromatosis are disorders associated with ineffective erythropoiesis, such as thalassemia (Chapter 14) and myelodysplastic syndrome (Chapter 13). In these disorders, the excess iron results not only from transfusions, but also from increased absorption. Erythroferrone released from an expanded population of erythroid progenitors in the bone marrow suppress hepcidin production by the liver, leading to increased iron absorption. Transfusions alone, when given repeatedly over a period of years (e.g., in patients with chronic hemolytic anemias), can also lead to systemic hemosiderosis and parenchymal injury, independent of ineffective hematopoiesis. Finally, other forms of cirrhosis may also diminish hepcidin production due to loss of hepatocyte mass, leading again to increased iron uptake from the gut and increased tissue levels of iron.

Wilson Disease

Wilson disease is an autosomal recessive disorder caused by mutation of the *ATP7B* gene, resulting in impaired copper excretion into bile and a failure to incorporate copper into ceruloplasmin. This disorder is marked by the accumulation of toxic levels of copper in many tissues and organs, principally the liver, brain, and eye. Free copper is taken up by hepatocytes and incorporated into apoceruloplasmin to form *ceruloplasmin,* which is secreted into the blood. Ceruloplasmin accounts for 90% to 95% of plasma copper. Excess copper within hepatocytes that is not incorporated into ceruloplasmin is sequestered in lysosomes and transported into the bile, from which it is eventually excreted in the feces.

Pathogenesis

Wilson disease results from loss-of-function mutations in the *ATP7B* gene, which encodes a transmembrane copper-transporting ATPase, expressed on the hepatocyte canalicular membrane. More than 300 sequence variants in the *ATP7B* gene have been identified, but not all cause the disease. The overwhelming majority of patients are compound heterozygotes containing different mutations on each *ATP7B* allele. The overall frequency of mutated alleles is 1:100, and the prevalence of the disease is approximately 1:30,000 to 1:50,000 (approximately 9000 patients in the United States). Deficiency of ATP7B decreases copper transport into bile, impairs its incorporation into ceruloplasmin, and inhibits ceruloplasmin secretion into the blood. These defects lead to a decrease in circulating ceruloplasmin accompanied by the accumulation of copper in hepatocytes. Total serum copper may be lower than normal due to the deficiency in ceruloplasmin, particularly early in the disease course.

The excess hepatic copper causes toxic injury by three mechanisms: (1) promoting the formation of free radicals by the Fenton reaction (Chapter 2); (2) binding to sulfhydryl groups of cellular proteins; and (3) displacing other metals from hepatic metalloenzymes. Hepatocyte injury causes nonceruloplasmin-bound copper to be spilled into the blood and accumulate in certain tissues, particularly the basal ganglia of the brain and the cornea. Concomitantly, urinary excretion of copper markedly increases, but not to levels sufficient to prevent deposition of copper in tissues.

MORPHOLOGY

The liver often bears the brunt of injury, but the disease may also present as a neurologic disorder. The hepatic changes are variable, ranging from relatively minor to massive damage, and mimic many other disease processes. **Fatty change (steatosis)** may be present with focal hepatocyte necrosis. Acute liver failure can mimic acute viral hepatitis. Chronic hepatitis in Wilson disease typically exhibits moderate to severe portal inflammation and hepatocyte necrosis, admixed with fatty change and features of steatohepatitis (i.e., ballooned hepatocytes, prominent Mallory hyaline, and perisinusoidal fibrosis). Eventually cirrhosis supervenes. Histochemical copper staining is not sensitive or specific for diagnosis of Wilson disease. Toxic brain injury primarily involves the basal ganglia, and nearly all patients with neurologic involvement develop eye lesions called **Kayser-Fleischer rings,** green to brown deposits of copper in Descemet membrane in the limbus of the cornea.

Clinical Features

The age at onset of Wilson disease ranges from 6 to 40 years of age (average age is 11.4 years). Its clinical presentation is extremely variable. Some patients present with acute or chronic liver disease. Neurologic involvement may lead to movement disorders (tremor, poor coordination, chorea or choreoathetosis) or rigid dystonia (spastic dystonia, masklike facies, rigidity, and gait disturbances). Patients may also have psychiatric symptoms. Hemolytic anemia may occur due to red cell membrane damage caused by oxidants

produced by free copper. The biochemical diagnosis of Wilson disease is usually based on the presence of decreased serum ceruloplasmin, increased hepatic copper content (the most sensitive and accurate test), and increased urinary excretion of copper (the most specific test). Sequencing of the *ATP7B* gene is now available for cases in which results of biochemical tests are indeterminate. Such problems may arise in the setting of liver injury, which may cause serum ceruloplasmin levels to rise into the normal range. Demonstration of Kayser-Fleischer rings by slit examination also is diagnostically helpful. Early recognition and long-term copper chelation therapy (with D-penicillamine or trientine) or zinc-based therapy (which blocks uptake of dietary copper in the gut) is effective. Individuals with hepatitis or unmanageable cirrhosis require liver transplantation, which can be curative.

α_1-Antitrypsin Deficiency

α_1-Antitrypsin deficiency is an autosomal recessive disorder of protein folding marked by very low levels of circulating α_1-antitrypsin (α_1AT). The major function of this protein is the inhibition of proteases, particularly neutrophil elastase, cathepsin G, and proteinase 3, which are normally released from neutrophils at sites of inflammation. α_1AT deficiency leads to the development of pulmonary emphysema, because the activity of neutrophil elastases is not inhibited (Chapter 15). It also causes liver disease as a consequence of hepatocellular accumulation of the misfolded protein, an example of a "toxic gain-of-function" mutation.

α_1AT is a small 394–amino acid plasma glycoprotein synthesized predominantly by hepatocytes. It is a member of the serine protease inhibitor (serpin) family. The gene is very polymorphic, and at least 75 forms of α_1AT have been identified, denoted alphabetically by their relative migration in an isoelectric gel. The general notation is "Pi" for "protease inhibitor" and an alphabetic letter for the position in the gel; two letters denote the genotype of an individual's two alleles. The most common genotype is PiMM, occurring in 90% of individuals (the "wild type").

The most common clinically significant mutation is PiZ; homozygotes for the PiZZ protein have circulating α_1AT levels that are only 10% of normal. These individuals are at high risk for developing clinical disease. Expression of alleles is autosomal codominant, and, consequently, PiMZ heterozygotes have intermediate plasma levels of α_1AT. Among people of northern European descent, the PiZZ state affects 1 in 1800 live births. Because of its early presentation with liver disease, α_1AT deficiency is the most commonly diagnosed inherited hepatic disorder in infants and children.

Pathogenesis

Disease-associated variants show a selective defect in transport of the protein from the endoplasmic reticulum to the Golgi apparatus; this is particularly characteristic of the PiZ polypeptide, resulting from the substitution of lysine for glutamine at amino acid position 342. The mutant polypeptide (α_1AT-Z) misfolds and aggregates, creating endoplasmic reticulum stress and triggering the unfolded protein response, a signaling cascade that may lead to apoptosis (Chapter 2). All individuals with the PiZZ genotype accumulate α_1AT-Z in the endoplasmic reticulum of hepatocytes, but only a small subset develops overt liver disease. Other genetic factors or environmental factors are thus posited to play a role in the development of liver disease.

MORPHOLOGY

α_1AT deficiency is characterized by the presence of round-to-oval **cytoplasmic globular inclusions in hepatocytes,** which are strongly periodic acid–Schiff (PAS)-positive and diastase-resistant (Fig. 18.25). The inclusions appear first in periportal hepatocytes in early and mild forms of the disease and with progression appear in central hepatocytes in more severe disease, such as that associated with the PiZZ variant. However, the number of globule-containing hepatocytes is not tightly correlated with disease severity, and patients may present with neonatal hepatitis prior to the appearance of the globules (generally 12 weeks of age or older for PiZZ-associated disease).

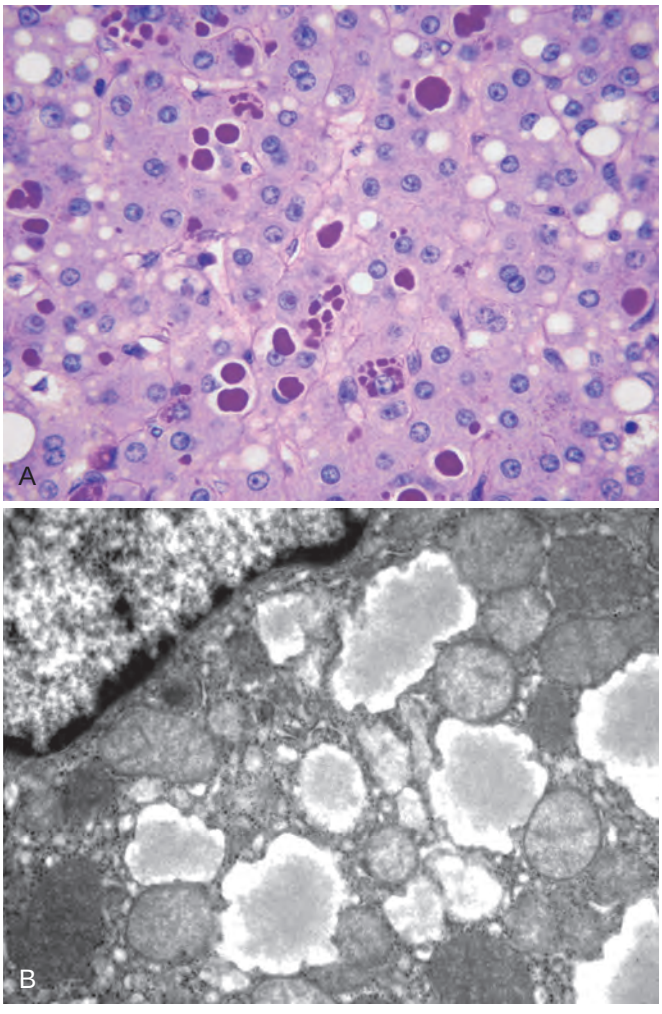


Figure 18.25 α_1-Antitrypsin deficiency. (A) Periodic acid–Schiff stain after diastase digestion of the liver, highlighting the characteristic magenta cytoplasmic granules. (B) Electron micrograph showing endoplasmic reticulum dilated by aggregates of misfolded protein.

Clinical Features

The clinical findings and course are quite variable. Unusual neonatal presentations tend to be associated with severe disease and rapid progression to cirrhosis. Other patients present in adulthood with chronic hepatitis, cirrhosis, or hepatocelllar carcinoma, which develops in 2% to 3% of PiZZ adults, usually (but not always) in the setting of cirrhosis. Liver transplantation is curative of hepatic disease, but it has no effect on the development and course of pulmonary disease (Chapter 15).

KEY CONCEPTS

INHERITED LIVER DISEASE

- Hereditary hemochromatosis is most commonly caused by loss-of-function mutations in the *HFE* gene, whose product regulates intestinal iron uptake by increasing hepcidin synthesis by the liver. It is characterized by increased absorption of dietary iron, and accumulation of iron in the liver and pancreas; injury to these organs results in cirrhosis and diabetes.
- Wilson disease is caused by mutations that abolish the function of the metal ion transporter ATP7B, which results in accumulation of copper in the liver, brain (particularly basal ganglia), and eyes ("Kayser-Fleisher rings").
- α_1-Antitrypsin deficiency is a disease of protein misfolding that results in impaired secretion of α_1-antitrypsin into the serum and creates endoplasmic reticulum stress, leading to hepatocyte injury through the unfolded protein response pathway. The main consequence of α_1-antitrypsin deficiency is pulmonary emphysema because of unchecked elastase activity.

CHOLESTATIC DISEASE

Bile Formation and Secretion

Bile plays a critical role in elimination of bilirubin, excess cholesterol, xenobiotics, and trace metals such as copper, arsenic, selenium, and zinc, and its detergent action emulsifies dietary fat in the intestinal lumen, enabling its absorption by the gut. The major components of bile are bilirubin, bile salts, cholesterol, and phospholipids (mainly phosphatidylcholine).

Bilirubin is a toxic end product of heme degradation that is processed by the liver and excreted in the bile (Fig. 18.26). Hepatic handling of bilirubin involves uptake from the circulation, intracellular storage, conjugation with glucuronic acid, and excretion into bile. The majority of the bilirubin produced daily (0.2 to 0.3 g, 85%) is derived from breakdown of senescent red cells by macrophages in the spleen, liver, and bone marrow. The remaining bilirubin is produced by the turnover of hepatic proteins containing heme groups (e.g., P-450 cytochromes). Heme is converted to bilirubin by the action of several phagocyte enzymes and released into the blood, where it binds albumin, a necessary step for transport, as bilirubin is insoluble at physiologic pH. Uptake by hepatocytes at the sinusoidal membrane is followed by conjugation of bilirubin with one or two molecules of glucuronic acid in the endoplasmic reticulum, and excretion of the water-soluble, nontoxic bilirubin glucuronides into bile. Most bilirubin glucuronides are deconjugated in the gut lumen by bacterial β-glucuronidases and degraded to colorless urobilinogens. The urobilinogens and the residue of intact pigment are largely excreted in feces. Approximately 20% of the urobilinogens formed are reabsorbed in the ileum and colon, returned to the liver, and re-excreted into bile. A small amount of reabsorbed urobilinogen is excreted in the urine.

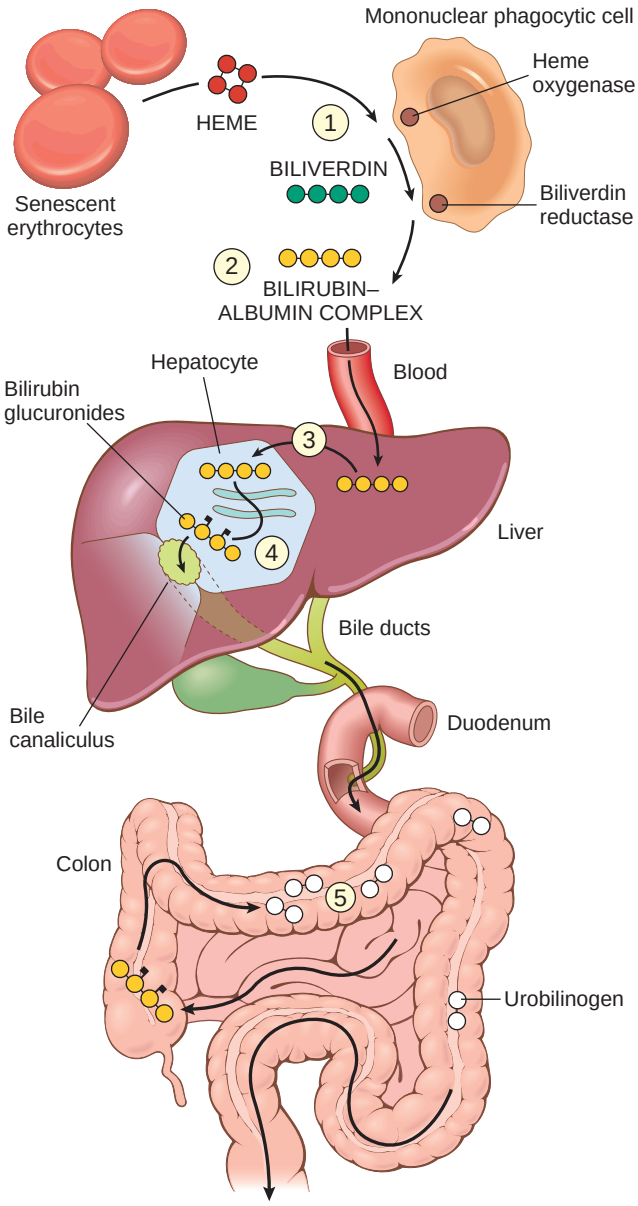


Figure 18.26 Bilirubin metabolism and elimination. (1) Normal bilirubin production from heme (0.2 to 0.3 g/day) is derived primarily from the breakdown of senescent circulating erythrocytes. (2) Extrahepatic bilirubin is bound to serum albumin and delivered to the liver. (3) Hepatocellular uptake and (4) glucuronidation in the endoplasmic reticulum generates bilirubin monoglucuronides and diglucuronides, which are water soluble and readily excreted into bile. (5) Gut bacteria deconjugate the bilirubin and degrade it to colorless urobilinogens. The urobilinogens and the residue of intact pigments are excreted in the feces, with some reabsorption and excretion into urine.

Bile salts are formed by the conjugation of bile acids with taurine or glycine. The predominant bile acids in humans are cholic acid and chenodeoxycholic acid, both of which are highly effective detergents. Bile acids combine with cholesterol and phospholipids to form micelles that solubilize cholesterol and reduce the toxic effect of bile acids on biliary epithelium. Most of the secreted bile acids, conjugated or unconjugated, are reabsorbed from the intestine and recirculate to the liver (enterohepatic circulation), where they are taken up by hepatocytes, thus helping to maintain the endogenous pool of bile acids.

The constituents of bile are ferried across the canalicular membrane of hepatocytes by a variety of transporter proteins. Some of the important ones are MRP2 (multidrug-resistant protein-2) for conjugated bilirubin, bile salt export pump (BSEP) for bile salts, MDR3 (multi-drug resistance-3) for phosphatidylcholine, and sterolins 1 and 2 for cholesterol.

Pathophysiology of Hyperbilirubinemia

Serum bilirubin levels in the normal adult vary between 0.3 and 1.2 mg/dL, and the rate of bilirubin production is equal to the rate of hepatic uptake, conjugation, and biliary excretion. Jaundice becomes evident when the serum bilirubin levels rise above 2 to 2.5 mg/dL. Depending on the underlying etiology (summarized in Table 18.8), the elevation can predominantly involve unconjugated (indirect) or conjugated (direct) bilirubin. Testing for conjugated and unconjugated plasma bilirubin helps determine the cause of hyperbilirubinemia. Excess bilirubin production (e.g., due to hemolytic anemia or ineffective erythropoiesis) or defective conjugation (due to immaturity or hereditary causes) leads to the accumulation of unconjugated bilirubin. This form is largely insoluble and cannot be excreted in the urine. Although most unconjugated bilirubin is tightly bound to albumin in the blood, at excessive levels the unbound fraction rises and may diffuse into tissues, particularly the brain in infants, and produce neurologic damage (kernicterus). Conjugated hyperbilirubinemia most often results from hepatocellular disease, bile duct injury, and biliary obstruction. Since this form is water-soluble and loosely bound to serum albumin, it can be excreted in the urine.

With this as an overview, we next discuss hepatobiliary disorders that lead to hyperbilirubinemia. Many of these are characterized by cholestasis, which refers to the retention of bilirubin and other solutes eliminated in bile due to impaired bile formation or obstruction of flow. Before delving into specific entities, we will briefly review the morphologic and clinical features of cholestasis that are common to all.

MORPHOLOGY

The hallmark of cholestasis is accumulation of green-brown plugs of bile pigment in hepatocytes and dilated canaliculi (Fig. 18.27). Rupture of canaliculi can lead to extravasation of bile, which is phagocytosed by Kupffer cells. Accumulation of bile salts in hepatocytes results in a swollen, foamy appearance of the cytoplasm ("feathery degeneration").

Table 18.8 Causes of Jaundice

Predominantly Unconjugated Hyperbilirubinemia
Excess production of bilirubin
Hemolytic anemias
Resorption of blood from internal hemorrhage (e.g., alimentary tract bleeding, hematoma)
Ineffective erythropoiesis (e.g., pernicious anemia, thalassemia)
Reduced hepatic uptake
Drug interference with membrane carrier systems
Some cases of Gilbert syndrome
Impaired bilirubin conjugation
Physiologic jaundice of the newborn (decreased UGT1A1 activity, decreased excretion)
Breast milk jaundice (β-glucuronidases in milk)
Genetic deficiency of UGT1A1 activity (Crigler-Najjar syndrome types I and II, some cases of Gilbert syndrome)
Predominantly Conjugated Hyperbilirubinemia
Deficiency of canalicular membrane transporters (Dubin-Johnson syndrome, Rotor syndrome)
Hepatocellular disease (e.g., viral or drug-induced hepatitis, cirrhosis)
Impaired bile flow from duct obstruction or autoimmune cholangiopathies

UGT1A1, Uridine diphosphate-glucuronyltransferase family, peptide A1.

Clinical Features

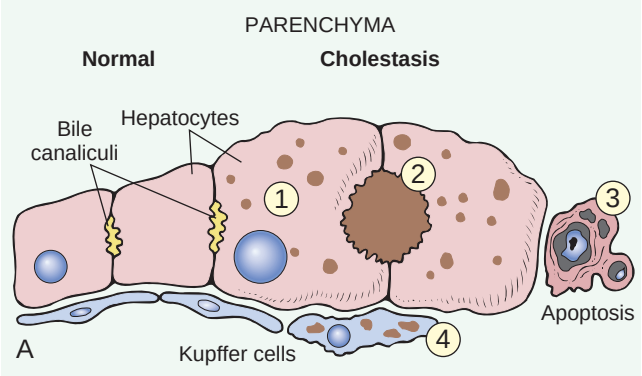
Elevated bilirubin becomes clinically evident as yellow discoloration of the skin (jaundice) and sclera (icterus). Other manifestations include pruritus, skin xanthomas (focal accumulation of cholesterol), or symptoms related to intestinal malabsorption, including deficiencies of fat-soluble vitamins (vitamins A, D, and K). Elevated serum alkaline phosphatase and γ-glutamyl transpeptidase (GGT), enzymes present on the apical (canalicular) membranes of hepatocytes and bile duct epithelial cells, are characteristic laboratory findings of cholestatic disease.

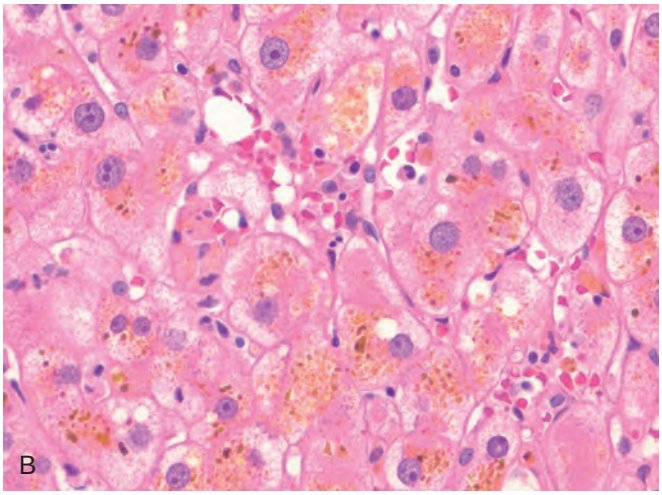
Physiologic Jaundice of the Newborn

In physiologic jaundice of the newborn, levels of UGT1A1 (uridine diphosphate-glucuronyltransferase family, peptide A1), the enzyme that is responsible for bilirubin glucuronidation, are low at birth and do not reach adult levels until 3 to 4 months of age. Hence, transient and mild unconjugated hyperbilirubinemia is nearly universal in the first week. Breastfeeding may exacerbate unconjugated hyperbilirubinemia, possibly because of the presence of bilirubin-deconjugating enzymes in breast milk. In most infants, phototherapy with blue light (which converts bilirubin to a soluble isomer that is readily excreted in the urine) is sufficient to keep the levels of unconjugated bilirubin within a safe range until the hepatic machinery for conjugation matures sufficiently.

Hereditary Hyperbilirubinemia

Genetic mutations can result in impaired uptake, conjugation, or secretion of bilirubin. *Crigler-Najjar syndrome type 1* is caused by severe UGT1A1 deficiency and is fatal around birth; in *Crigler-Najjar type 2* and Gilbert syndrome, there is some UGT1A1 activity, leading to a milder phenotype. In contrast, *Dubin-Johnson syndrome* and *Rotor syndrome* lead





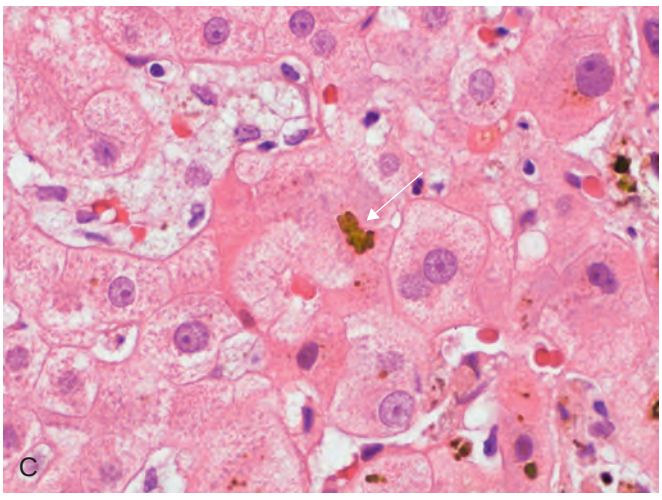


Figure 18.27 Cholestasis. (A) Morphologic features of cholestasis (right) and comparison with normal liver (left). Cholestatic hepatocytes (1) are enlarged with dilated canalicular spaces (2). Apoptotic cells (3) may be seen, and Kupffer cells (4) frequently contain regurgitated bile pigments. (B) Intracellular cholestasis showing the bile pigments in the cytoplasm. (C) Bile plug *(arrow)* showing the expansion of bile canaliculus by bile.

to conjugated hyperbilirubinemia. Both are autosomal recessive disorders and clinically innocuous. Dubin-Johnson syndrome is caused by mutation in the *MRP2* gene (multidrug resistant protein 2), which is required for transport of non–bile salt organic anions at the canalicular membranes. Deposition of brown-black melanin-like pigment in the hepatocytes is a striking feature in this disease and can lead to blackening of the liver.

Large Bile Duct Obstruction

Obstruction of the large bile ducts has diverse causes. In adults, the common causes are stones (choledocholithiasis), malignant neoplasms of the biliary tree or head of pancreas (usually adenocarcinoma), and strictures resulting from previous surgical procedures or ischemic injury. Primary sclerosing cholangitis (described later) may lead to an obstructive picture due to inflammatory injury to the large intrahepatic or extrahepatic bile ducts. In the pediatric setting, the common culprits are biliary atresia, cystic fibrosis, and choledochal cysts. The resulting cholestatic changes are reversible if the obstruction is corrected early in the disease course, but persistent obstruction can lead to fibrosis and so-called biliary cirrhosis. Biliary obstruction also predisposes to ascending cholangitis, a bacterial infection of the biliary tree most commonly caused by enteric organisms such as coliforms and enterococci. Cholangitis usually presents with fever, chills, abdominal pain, and jaundice. Severe cases can result in abscess formation, sepsis, and death.

MORPHOLOGY

Obstruction of extrahepatic or large intrahepatic bile ducts leads to proximal duct dilation. The hallmark on liver biopsies is portal expansion due to edema, prominent ductular reaction at the portal-parenchymal interface, and infiltrating neutrophils associated with ductules **("pericholangitis")** (Fig. 18.28). In ascending cholangitis, the neutrophils also involve the bile duct epithelium and lumens (Fig. 18.29). Persistent obstruction leads to fibrosis, which can eventually proceed to biliary cirrhosis (Fig. 18.30). Swelling of periportal hepatocytes **("feathery degeneration"),** bile pigment, and Mallory hyaline is seen in periportal hepatocytes in advanced disease. Superimposed ascending cholangitis in advanced disease can precipitate acute-on-chronic liver failure.

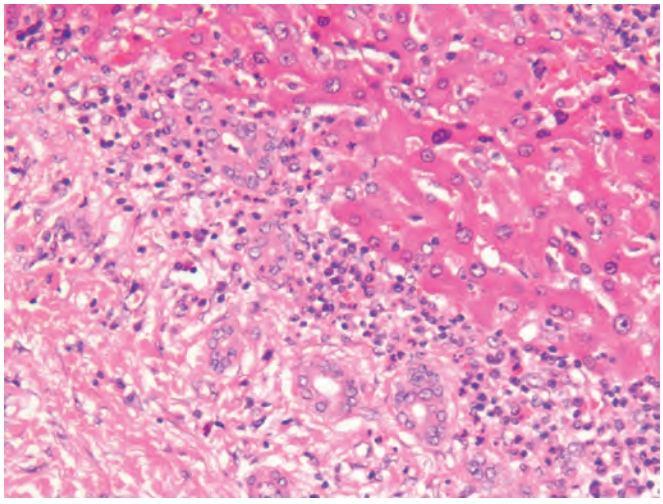


Figure 18.28 Acute large duct obstruction. There is marked edema of the portal tract stroma *(white spaces)* and a ductular reaction with admixed neutrophils at the interface between portal tract and hepatocellular parenchyma.

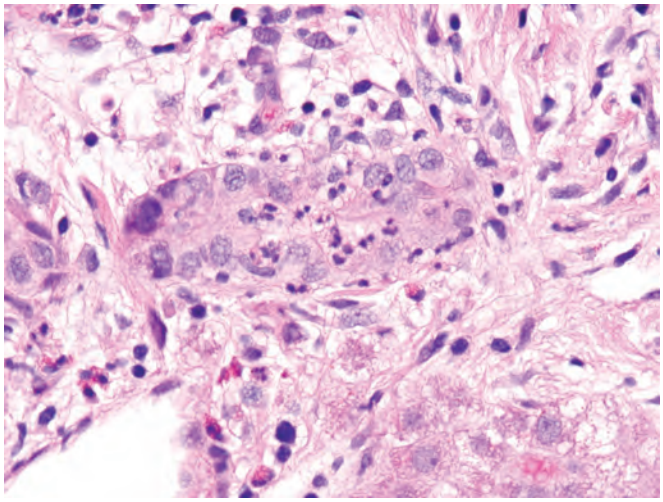


Figure 18.29 Ascending cholangitis. Individuals with large bile duct obstruction are at risk for bacterial infections within the biliary tree. Neutrophils are seen within the bile duct epithelial lining and within the lumen.

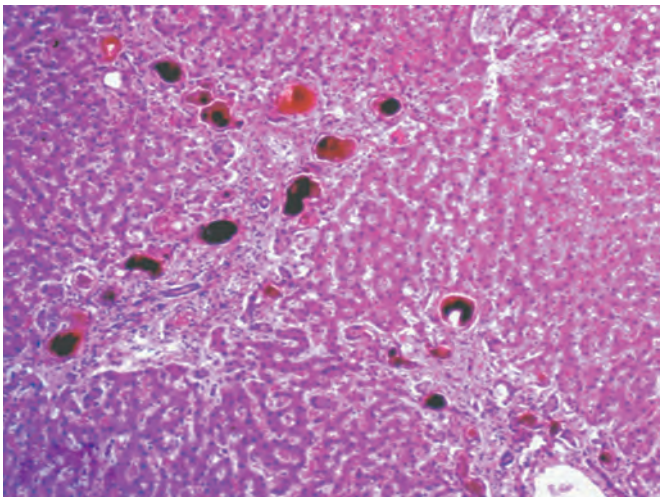


Figure 18.31 Ductular cholestasis of sepsis. Large, dark bile concretions within markedly dilated canals of Hering and ductules at the portal-parenchymal interface. (Courtesy Dr. Jay Lefkowitch, Columbia University College of Physicians and Surgeons, NY.)

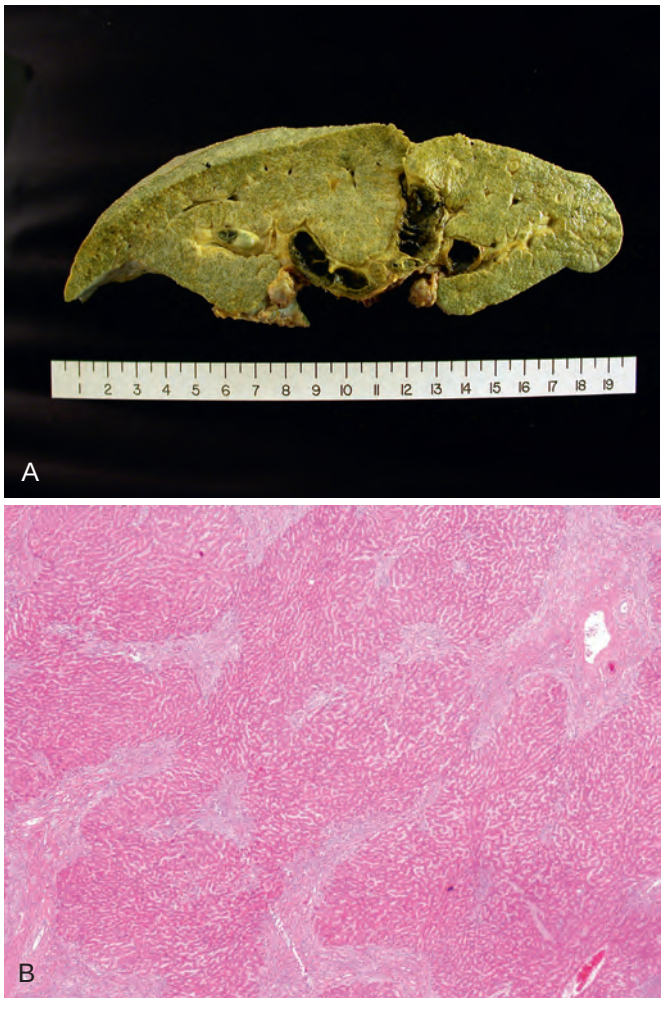


Figure 18.30 Biliary cirrhosis. (A) Sagittal section through the liver demonstrates the nodularity (most prominent at the right) and bile staining of end-stage biliary cirrhosis. (B) Unlike other forms of cirrhosis, nodules of liver cells in biliary cirrhosis are often not round but irregular, like jigsaw puzzle shapes.

Cholestasis of Sepsis

Sepsis may affect the liver by several mechanisms: (1) through direct effects of intrahepatic bacterial infection (e.g., abscess formation or bacterial cholangitis); (2) ischemia relating to hypotension caused by sepsis (particularly when the liver is cirrhotic); or (3) in response to circulating microbial products. The latter effect is most likely to lead to the cholestasis of sepsis, particularly when the systemic infection is caused by gram-negative organisms. Characteristic morphologic findings in the setting of severe sepsis include canalicular cholestasis and bile plugs within dilated canals of Hering and bile ductules at the portal-parenchymal interface ("ductular" or "cholangiolar cholestasis") (Fig. 18.31). Inflammation and hepatocellular injury is typically mild.

Primary Hepatolithiasis

At one time called "recurrent pyogenic cholangitis," hepatolithiasis refers to stones in the intrahepatic bile ducts that can lead to repeated bouts of ascending cholangitis and progressive inflammatory destruction of hepatic parenchyma. The disease is highly prevalent in East Asia, but is rare in other parts of the world. Its cause is uncertain; congenital duct abnormalities, diet, and chronic infection with bacteria or parasites have all been suggested as possible etiologies. Chronic inflammatory injury associated with hepatolithiasis is a risk factor for cholangiocarcinoma, particularly in Taiwan and to a lesser extent in Japan.

MORPHOLOGY

Pigmented calcium bilirubinate stones are present in distended intrahepatic bile ducts (Fig. 18.32). The ducts show chronic inflammation, mural fibrosis, and peribiliary gland hyperplasia. Obstruction of extrahepatic ducts is not present. Repeated bouts of inflammation, parenchymal collapse, and fibrosis can lead to a mass-like lesion, which can mimic a tumor on imaging.

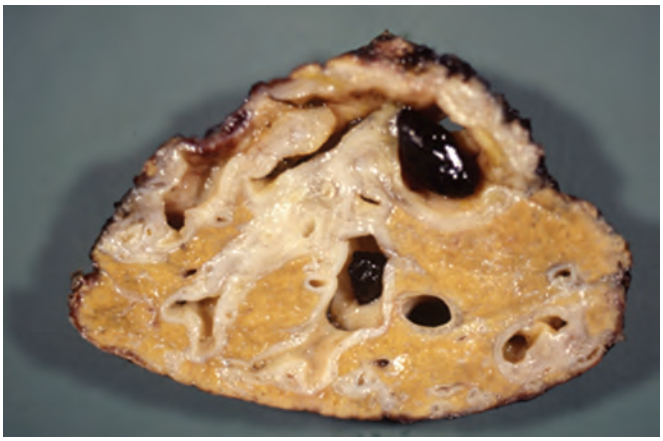


Figure 18.32 Hepatolithiasis. A resected, atrophic right hepatic lobe with characteristic findings including markedly dilated and distorted bile ducts containing large pigment stones and broad areas of collapsed liver parenchyma. (Courtesy Dr. Wilson M.S. Tsui, Caritas Medical Centre, Hong Kong.)

Neonatal Cholestasis

Physiologic jaundice of the newborn (discussed earlier) resolves by 2 weeks, and infants who have jaundice beyond 14 to 21 days after birth must be evaluated for neonatal cholestasis. The major causes can be grouped under the two broad categories: (1) obstructive biliary disease such as biliary atresia; and (2) nonobstructive etiologies that include paucity of bile ducts, infectious/metabolic diseases, bile transporter defects, and idiopathic neonatal hepatitis.

Extrahepatic Biliary Atresia

Extrahepatic biliary atresia is characterized by complete or partial obstruction of the extrahepatic biliary tree within the first 3 months of life. It accounts for one-third of neonatal cholestasis cases and for 50% to 60% of children referred for liver transplantation. In its most common perinatal form (80% of cases), the biliary tree is normally formed and the disease onset is after birth. Infections, toxic agents, and autoimmune injury have been invoked, but the cause is unknown. The less common fetal form likely results from aberrant development of the extrahepatic biliary tree. Infants with extrahepatic biliary atresia present with jaundice, dark urine, light or acholic stools, and hepatomegaly.

MORPHOLOGY

Inflammation and fibrosis of the hepatic or common bile ducts is the hallmark of the disease and can extend to involve the intrahepatic ducts. Typical features of biliary obstruction are seen on liver biopsy, including portal edema, ductular reaction, and neutrophil infiltrates. If left uncorrected, cirrhosis can develop by 3 to 6 months of age.

Clinical Features

Since extrahepatic biliary atresia requires surgical intervention (portoenterostomy or Kasai procedure), its differentiation from nonobstructive neonatal cholestasis is critical. Other etiologies of obstructive biliary diseases like cystic fibrosis also must be excluded. The combination of clinical presentation, imaging (ultrasound and hepatospecific iminodiacetic-acid [HIDA] scans), and biopsy can confirm the diagnosis in a vast majority of cases. Ultrasound findings of a small or absent gallbladder and fibrosis at the porta hepatis support the diagnosis of extrahepatic biliary atresia, as do HIDA scans with technetium-99m (^{99m}Tc), which is excreted in bile. In extrahepatic biliary atresia, there is total lack of secretion of ^{99m}Tc into the bile, and the biliary tree is not visualized. Involvement of ducts proximal to the porta hepatis, intrahepatic progression of disease, and ascending cholangitis are impediments to successful surgical therapy. Liver transplantation is the only option when surgical intervention is not feasible.

Nonobstructive Neonatal Cholestasis

A diverse group of disorders are associated with nonobstructive neonatal cholestasis. These can be divided into those in which cholestasis appears as part of a syndrome or as an isolated liver abnormality.

- *Alagille syndrome* is an autosomal dominant disorder that is associated with cholestasis and paucity of bile ducts as well as other abnormalities such as dysmorphic facies, butterfly-shaped vertebra, eye defects, and cardiac defects. It is caused by loss-of-function mutations in the Notch pathway involving either the genes encoding the ligand JAG1 or the receptor NOTCH2, both of which are required for normal development of the biliary tree.
- *Certain inborn errors of metabolism,* particularly galactosemia (Chapter 10) and Niemann-Pick disease (Chapter 5), can present with nonobstructive cholestasis.
- *Nonsyndromic causes* of nonobstructive cholestasis include α_1-antitrypsin deficiency (discussed earlier) and diverse disorders of bile acid synthesis and bile transport.

With recent advances, the underlying etiology can be determined in 85% to 90% of cases (summarized in Table 18.9).

Table 18.9 Major Causes of Neonatal Cholestasis

Obstructive Biliary Disease
Extrahepatic biliary atresia
Neonatal Infection
Cytomegalovirus Bacterial sepsis Urinary tract infection Syphilis
Genetic Disorders
Metabolic diseases: Tyrosinemia, galactosemia Lipid storage diseases: Niemann-Pick disease Bile synthesis defects: Bile transport abnormalities (progressive intrahepatic cholestasis), bile acid synthesis defects Cystic fibrosis α_1-Antitrypsin deficiency Alagille syndrome (syndromic paucity of bile ducts)
Miscellaneous
Shock/hypoperfusion, drugs, total parenteral nutrition, hypopituitarism Idiopathic neonatal hepatitis

MORPHOLOGY

Decreased numbers of bile ducts within portal regions profiles is the most characteristic and constant feature. The biliary insufficiency often is associated with hepatitis, characterized by inflammation and hepatocellular apoptosis/necrosis. The most striking features are hepatocanalicular cholestasis and giant-cell or syncytial cell change with multinucleated hepatocytes (Fig. 18.33). Reactive Kupffer cells and extramedullary hematopoiesis are usually present.

Autoimmune Cholangiopathies

This section discusses the two main autoimmune disorders of bile ducts: primary biliary cholangitis and primary sclerosing cholangitis. The features of these two conditions are contrasted in Table 18.10.

Primary Biliary Cholangitis (PBC)

PBC is an autoimmune disease characterized by inflammatory destruction of small- and medium-sized intrahepatic bile ducts. Large intrahepatic ducts and the extrahepatic biliary tree are not involved. Most patients are diagnosed in the early stages of disease, and hence the former name "primary biliary cirrhosis" is no longer used. PBC has a striking female predilection of 9:1, with a peak incidence between 40 and 50 years of age. The disease is most common in the United States and Northern Europe, and the incidence is low in Africa and the Indian subcontinent. Family members of PBC patients have an increased risk for development of the disease.

Pathogenesis

PBC is thought to be an autoimmune disorder resulting from a T lymphocyte–mediated attack on small interlobular bile ducts. The trigger for the attack is not known, but may involve exposure to environmental factors like infections and toxic chemicals in genetically susceptible individuals. This may lead to expression of "auto-antigens" on bile duct epithelial cells and the resultant destruction by T lymphocytes. The retention of bile salts due to bile duct injury leads to secondary hepatocellular injury in PBC, which can eventually produce a cirrhotic picture.

Antimitochondrial antibodies directed against the E2 component of the pyruvate dehydrogenase complex (PDC-E2) are the most characteristic finding in PBC. PDC-E2–specific T cells are present in PBC, further supporting the notion of an immune-mediated process. The role of antimitochondrial antibodies in the pathogenesis of PBC is unclear, as 5% of patients with otherwise typical PBC are antimitochondrial antibody (AMA)-negative. Moreover, antibody titers do not correlate with disease severity or disease progression, and they are not predictive of response to therapy (discussed later). Other autoantibodies against nuclear pore proteins and centromeric proteins may also be present.

Table 18.10 Main Features of Primary Biliary Cholangitis and Primary Sclerosing Cholangitis

Parameter	Primary Biliary Cholangitis	Primary Sclerosing Cholangitis
Age	Median age 50 years	Median age 30 years
Gender	90% female	70% male
Associated conditions	Sjögren syndrome (70%), thyroid disease, scleroderma	Inflammatory bowel disease (70%)
Serology	95% AMA-positive, 40%–50% ANA positive	65% ANCA-positive; ANA variable, AMA typically negative 6% ANA-positive
Radiology	Normal	Strictures and beading of large bile ducts; pruning of smaller ducts
Duct lesion	Florid duct lesions; loss of small ducts	Inflammatory destruction of extrahepatic and large intrahepatic ducts; fibrotic obliteration of medium and small intrahepatic ducts; ductular reaction in smaller portal tracts

AMA, Antimitochondrial antibody; *ANA*, antinuclear antibody; *ANCA*, antineutrophil cytoplasmic antibody.

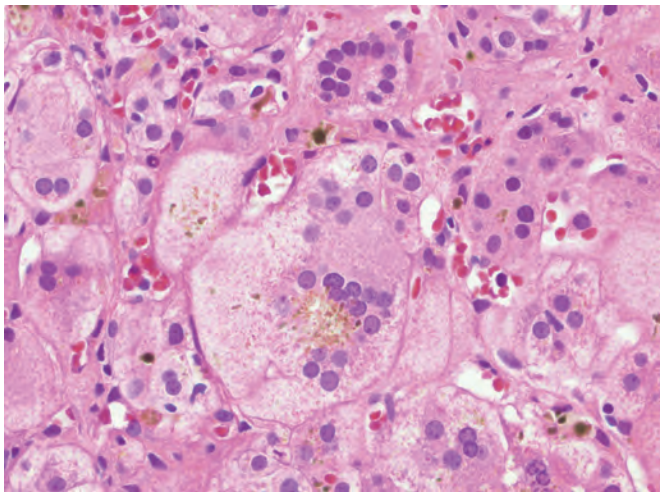


Figure 18.33 Neonatal hepatitis. Note the multinucleated giant hepatocytes.

MORPHOLOGY

The hallmark of PBC is lymphocytic infiltration and epithelial injury involving the small interlobular bile ducts. Poorly formed epithelioid granulomas are often present in the portal tracts and can be centered on bile ducts. **The histologic picture of lymphocytic and/or granulomatous bile duct destruction (florid duct lesion) is highly characteristic of PBC** (Fig. 18.34). Portal lymphoplasmacytic inflammation and ductular reaction are often present. The bile duct involvement has a patchy distribution, with involvement of a minority of portal tracts being a common feature in early disease. Disease progression leads to loss of small intrahepatic bile ducts ("ductopenia"). Unlike obstructive, drug-induced, or sepsis-associated cholestasis, the bile

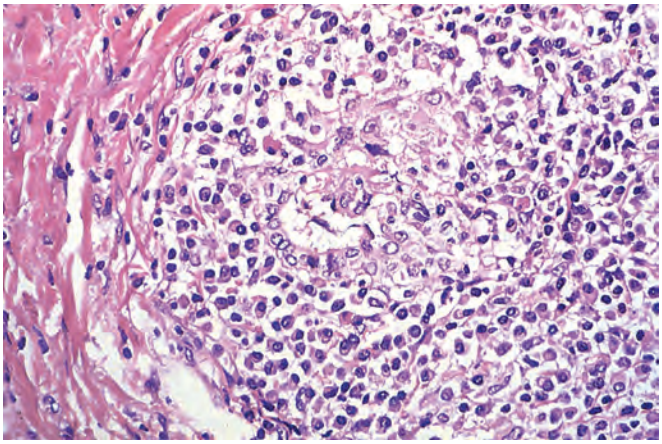


Figure 18.34 Primary biliary cholangitis. A portal tract is markedly expanded by an infiltrate of lymphocytes and plasma cells surrounding a destructive granulomatous reaction centered on a bile duct ("florid duct lesion").

accumulation in PBC is not centrizonal, but is seen in periportal/periseptal regions. Stasis of bile salts leads to swelling of periportal hepatocytes, which have clear cytoplasm with granular strands **(feathery degeneration)** and may develop Mallory hyaline. These periportal changes are referred to as cholate stasis and can be seen in any chronic biliary disease. End-stage disease culminates in cirrhosis. Like other biliary diseases, cirrhotic nodules in PBC tend to be elongated ("garland-shaped"), in contrast to the round nodules in hepatitic diseases. Some patients develop portal hypertension due to regenerative nodules without fibrosis, a feature called nodular regenerative hyperplasia. The basis of this phenomenon in PBC is not known.

Clinical Features

Symptomatic PBC usually presents with fatigue and pruritus, which increase slowly over time. Hypercholesterolemia is common. Splenomegaly and jaundice are seen in advanced disease. Other features that may be seen include skin hyperpigmentation, xanthelasmas, steatorrhea, and vitamin D malabsorption–related osteomalacia and/or osteoporosis. Other autoimmune diseases, such as Sjögren syndrome, systemic sclerosis, thyroiditis, rheumatoid arthritis, Raynaud phenomenon, and celiac disease, may also be present in affected patients.

An increasing number of cases are being detected in asymptomatic patients with elevated serum alkaline phosphatase levels. Antimitochondrial antibodies are present in 90% to 95% of patients at all stages of the disease process and are highly characteristic. For uncertain reasons, many patients also have elevated serum levels of IgM antibody. Antimitochondrial antibodies can be detected by indirect immunofluorescence or enzyme-linked immunosorbent assay (ELISA); the latter has higher sensitivity and specificity as it selectively detects antibodies directed against the E2 subunit of the pyruvate dehydrogenase complex. The diagnosis can be established if two of the following are present: elevated alkaline phosphatase for more than 6 months; positive test for antimitochondrial antibodies; and characteristic histologic findings. The AMA-negative form of the disease has only two of these findings and is sometimes termed "autoimmune cholangiopathy" or "autoimmune cholangitis." Its clinical presentation, pathology, and natural history are otherwise similar.

PBC is a slowly progressive disease, and progression to end-stage liver disease occurs in 20% to 25% of patients over 15 to 20 years. Treatment with oral ursodeoxycholic acid, a naturally occurring bile acid, slows disease progression in most patients, but the response is inadequate in up to 40% of those affected; other medical therapies, such as with obeticholic acid (a synthetic bile acid) are being evaluated in such patients. In those in whom medical treatment fails, liver transplantation yields excellent results, with survival of more than 70% at 7 years. Like many chronic forms of hepatic injury, PBC is associated with an increased risk of hepatocellular carcinoma.

Primary Sclerosing Cholangitis (PSC)

PSC is characterized by inflammation and obliterative fibrosis of extrahepatic and large intrahepatic ducts and dilation of preserved segments. PSC tends to occur in the third through fifth decades and has a 2:1 male predominance.

Pathogenesis

The presence of circulating autoantibodies and T lymphocytes in the periductal stroma, and associations with HLA-B8 and other MHC antigens and ulcerative colitis, all support the idea that PSC is an immunologically mediated process. A combination of environmental and genetic factors is thought to trigger the inflammatory injury of bile ducts. First-degree relatives of patients with PSC have an increased risk of developing the disease, suggesting a genetic component. It has been proposed that T cells, activated in the damaged mucosa of patients with ulcerative colitis, migrate to the liver, where they recognize a cross-reacting bile duct antigen. Infections or changes in the intestinal microbiome have also been postulated to lead to alterations in cholangiocytes that incite inflammatory injury. Unlike PBC, the autoantibody profile in PSC is not characteristic, but atypical perinuclear antineutrophil cytoplasmic antibodies (pANCA) targeting a nuclear envelope protein are found in approximately 65% of patients.

MORPHOLOGY

The large intrahepatic bile ducts and extrahepatic ducts show epithelial injury and neutrophilic infiltrates superimposed on chronic inflammation. The edema and inflammation cause narrowing of the lumen, leading to fibrosis and strictures. Lithiasis can develop in the dilated ducts. The inflammatory injury can result in circumferential "onion skin" fibrosis (periductal fibrosis/sclerosis) around an increasingly atrophic duct lumen (Fig. 18.35), eventually leading to obliteration by a "tombstone" scar. The smaller intrahepatic bile ducts, which are more commonly sampled on biopsy, are not directly involved by inflammation, but may exhibit mild injury and prominent ductular reaction due to cholestasis. Progression of cholestasis and fibrosis culminates in biliary cirrhosis. Biliary intraepithelial neoplasia can develop and eventually lead to cholangiocarcinoma, a dreaded complication of PSC.

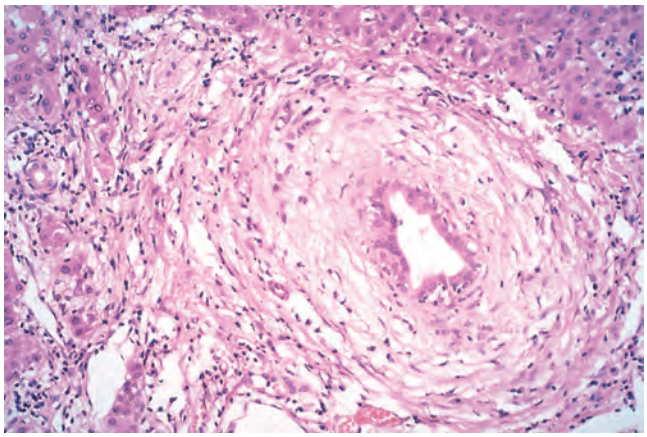


Figure 18.35 Primary sclerosing cholangitis. A degenerating bile duct is entrapped in a dense, "onion-skin" concentric scar.

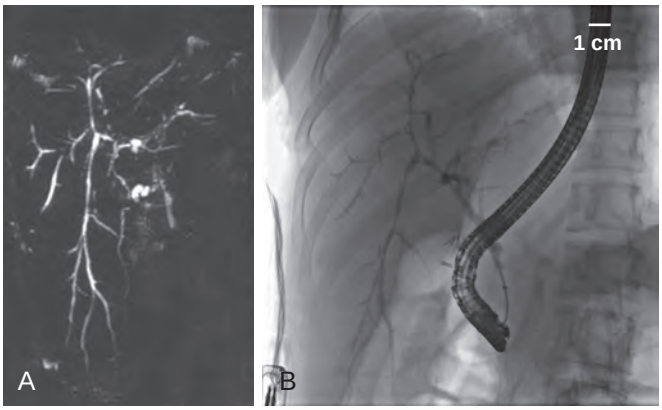


Figure 18.36 Imaging studies of a patient with primary sclerosing cholangitis. (A) Magnetic resonance cholangiography shows focal dilation in some bile ducts *(bright, broad areas)* and strictures in others *(thinning or absence)*. (B) Endoscopic retrograde cholangiography of the same patient shows nearly identical features as in A. The endoscope is visible, giving a sense of scale. (Courtesy Dr. M. Edwyn Harrison, MD, Mayo Clinic, Scottsdale, Ariz.)

Clinical Features

The most common presenting symptoms are fatigue, pruritus, and jaundice. Nearly one-half of the patients are asymptomatic at presentation and come to attention because of persistent elevation of serum alkaline phosphatase, particularly in ulcerative colitis patients who are routinely screened. Ascending cholangitis can also be the initial presentation. Chronic pancreatitis and chronic cholecystitis, from involvement of pancreatic ducts and gallbladder, are also seen. Fibrosis and eventually cirrhosis can occur.

The gold standard for diagnosis of PSC is the characteristic "beading" seen in the large intrahepatic and extrahepatic biliary tree by endoscopic/magnetic retrograde cholangiopancreaticography (ERCP/MRCP), involving attributable to irregular biliary strictures and dilations (Fig. 18.36). Inflammatory bowel disease, particularly ulcerative colitis, affects approximately 70% of individuals with PSC, whereas 8% of inflammatory bowel disease patients develop PSC. In a small subset of cases, mostly in association with inflammatory bowel disease, there is involvement of smaller bile ducts only and the ERCP/MRCP examination is normal. This is referred to as small-duct PSC and can progress to typical large-duct PSC. Classic PSC needs to be distinguished from cholangitis that occurs in IgG_4 sclerosing disease (Chapter 6), which is steroid-responsive and often associated with pancreatitis. In 5% to 10% of cases, autoimmune hepatitis is present along with PSC.

The disease typically follows a protracted course leading to cirrhosis over 10 to 15 years. The lifetime risk of developing cholangiocarcinoma is 20%. There is no established medical therapy for PSC; clinical trials of ursodeoxycholic acid are ongoing. Cholestyramine, a bile acid–binding resin, is used to alleviate pruritus, and endoscopic dilation or stenting is used to relieve biliary obstruction. Liver transplantation is the treatment of choice for end-stage liver disease.

KEY CONCEPTS

CHOLESTATIC DISEASES

- Cholestasis occurs with impaired bile flow, leading to accumulation of bile pigment in the hepatic parenchyma. Hepatic etiologies include metabolic defects in bile formation or secretion and inflammatory injury to bile ducts, while posthepatic etiologies include mechanical obstruction or inflammatory destruction of the extrahepatic bile ducts.
- Large bile duct obstruction is most commonly associated with gallstones and carcinomas involving the head of the pancreas. Ascending cholangitis may develop. Chronic obstruction can lead to cirrhosis.
- Cholestasis in sepsis may result from direct effects of intrahepatic bacterial infection, circulating microbial products, or ischemia relating to hypotension.
- Primary hepatolithiasis is a disorder of intrahepatic gallstone formation that is most common in East Asia and leads to repeated bouts of ascending cholangitis and inflammatory parenchymal destruction. It predisposes to cholangiocarcinoma.
- PBC is an autoimmune disease with inflammatory and often granulomatous destruction of small- to medium-sized intrahepatic bile ducts. It occurs most often in middle-aged women and is typically associated with antimitochondrial antibodies and with other autoimmune diseases (e.g., Sjögren syndrome and Hashimoto thyroiditis).
- PSC is an autoimmune disease that is more common in males and is strongly associated with inflammatory bowel disease, particularly ulcerative colitis. The diagnosis is established by visualization of the biliary tree. Histologically, the typical features are inflammation, fibrosis, and strictures involving large intrahepatic and extrahepatic bile ducts. Patients are at risk for developing cholangiocarcinoma.

STRUCTURAL ANOMALIES OF THE BILIARY TREE

Choledochal Cyst

Choledochal cysts are congenital dilations of the common bile duct. They present most often in children before age 10 as jaundice and/or recurrent abdominal pain, symptoms

that are typical of biliary colic. Approximately 20% of cases become symptomatic only in adulthood. In some cases, choledochal cysts occur in conjunction with cystic dilation of the intrahepatic biliary tree (*Caroli disease*, described later). The female-to-male ratio is 3:1 to 4:1. These uncommon cysts may take the form of segmental or cylindric dilation of the common bile duct, diverticula of the extrahepatic ducts, or choledochoceles, which are cystic lesions that protrude into the duodenal lumen. Choledochal cysts predispose to stone formation, stenosis and stricture, pancreatitis, and obstructive biliary complications within the liver. There also is an increased risk of developing bile duct carcinoma.

Fibropolycystic Disease

Fibropolycystic disease of the liver is a heterogeneous group of lesions in which the primary abnormality is congenital malformation of the biliary tree. Lesions may be found incidentally during radiographic studies, surgery, or autopsy. The most severe forms may manifest as hepatosplenomegaly or portal hypertension in the absence of hepatic dysfunction, starting in late childhood or adolescence. These lesions are part of the spectrum of developmental ductal plate malformations, which result from persistence of the fetal periportal ductal plates. The caliber of involved portal tracts determines the size, morphology, and distributions of the lesions. Fibropolycystic liver disease often occurs in association with autosomal recessive polycystic renal disease (Chapter 20). Persons with fibropolycystic liver disease are at increased risk for cholangiocarcinoma.

Three sets of pathologic findings may be seen with polyfibrocystic disease, sometimes overlapping with each other:

- *Von Meyenburg complexes* are small bile duct hamartomas (Fig. 18.37). Occasional von Meyenburg complexes are common in otherwise normal individuals. When they are diffuse, they signal the presence of underlying, clinically important fibropolycystic disease.
- *Single or multiple intrahepatic or extrahepatic biliary cysts*. When present in isolation, these may be symptomatic due to ascending cholangitis. Multifocal cystic dilation of the large intrahepatic bile ducts is referred to as *Caroli disease*. When the cystic dilation of the biliary tree occurs along with congenital hepatic fibrosis, the term *Caroli syndrome* is applied (Fig. 18.38). Ducts may be cystically dilated, but true cysts are also present. These may be intrahepatic cysts or choledochal cysts, as described earlier.
- *Congenital hepatic fibrosis* is characterized by portal tracts enlarged by broad bands of collagenous tissue forming septa that divide the liver into irregular islands (Fig. 18.39). Variable numbers of abnormally shaped bile ducts are embedded in the fibrous tissue, although they remain in continuity with the biliary tree. As with cirrhosis, individuals with congenital hepatic fibrosis are at risk for developing portal hypertension and its complications, particularly bleeding varices.

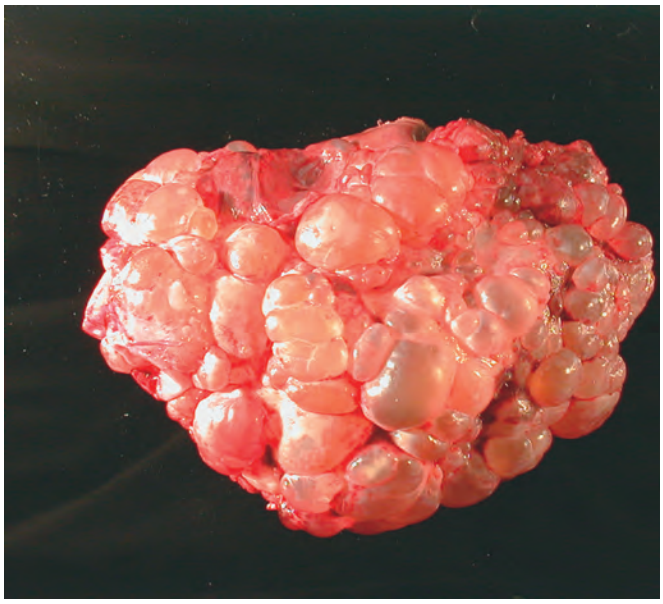


Figure 18.38 Congenital hepatic fibrosis with multiple biliary cysts.

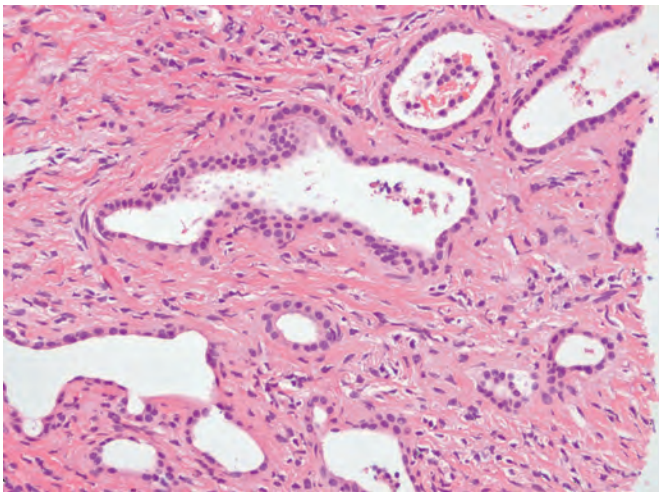


Figure 18.37 Von Meyenburg complex (bile duct hamartoma). Dilated irregular bile ducts with curvilinear outlines thought to represent ductal plate formation.

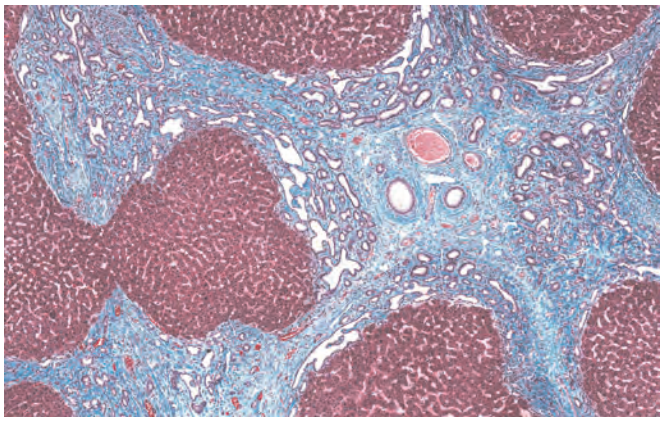


Figure 18.39 Congenital hepatic fibrosis. Broad bands of fibrosis with dilated remnants of ductal plates. The intervening parenchyma is relatively normal (Masson Trichrome stain).

CIRCULATORY DISORDERS

Given the large volume of blood that flows through the liver, it is not surprising that circulatory disturbances have considerable impact on the organ. In most instances, however, clinically significant liver function abnormalities do not develop, but hepatic morphology may be strikingly affected. These disorders can be grouped according to whether blood flow into, through, or from the liver is impaired (Fig. 18.40).

Impaired Blood Flow Into the Liver

Hepatic Artery Compromise

The liver's dual blood supply mitigates against infarction; however, obstruction of an intrahepatic branch of the hepatic artery by emboli, thrombi, or compression may result in a localized infarct that is either pale and anemic, or hemorrhagic if there is suffusion with portal blood (Fig. 18.41). Underlying causes may include neoplasia, polyarteritis nodosa (Chapter 11), or sepsis. Interruption of blood flow through the main hepatic artery does not always produce ischemic necrosis of the organ, particularly if the liver is otherwise normal, as retrograde arterial flow through accessory vessels, when coupled with the portal venous supply, is usually sufficient to sustain the liver parenchyma.

Several conditions increase the vulnerability of the liver to infarction in the setting of hepatic artery obstruction. Hepatic artery thrombosis in a transplanted liver generally leads to infarction of the major ducts of the biliary tree, since their blood supply is entirely arterial, and may also lead to hepatic injury and regenerative changes. Large infarcts may occur with combined portal vein and hepatic artery thrombosis, which eventually result in subcapsular scarring.

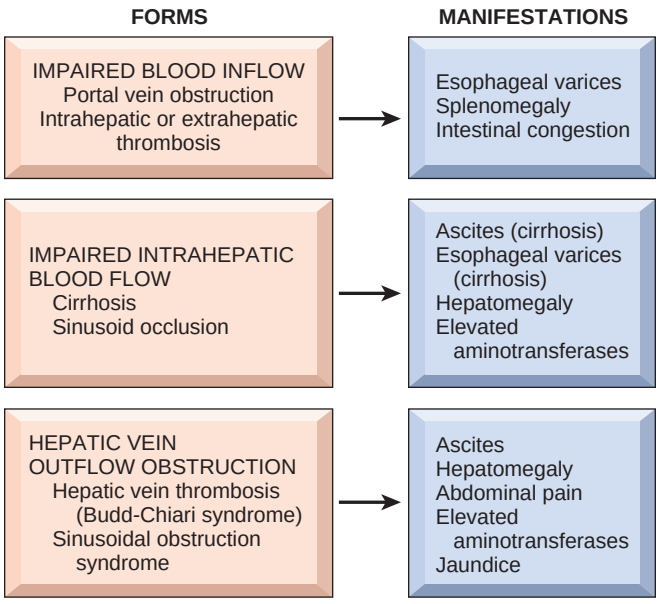


Figure 18.40 Forms and clinical manifestations of hepatic circulatory disorders.

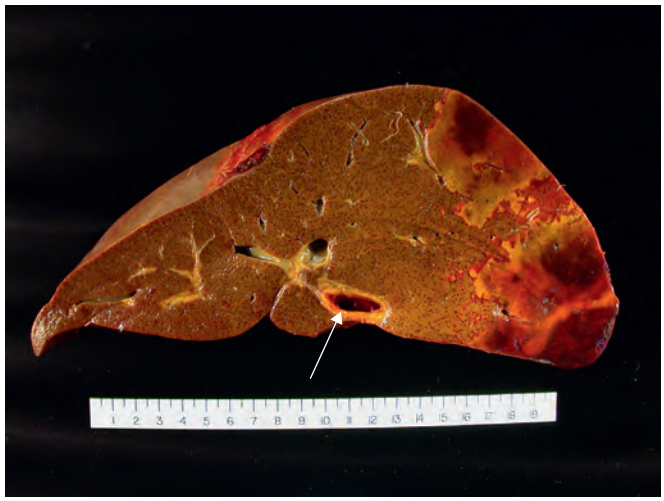


Figure 18.41 Liver infarct. A thrombus is lodged in a peripheral branch of the hepatic artery *(arrow)* and compresses the adjacent portal vein; the distal necrotic infarcted tissue has pale margins and multifocal areas of hemorrhage.

Portal Vein Obstruction and Thrombosis

Blockage of the portal vein and its tributaries may be insidious and well tolerated, or may be a catastrophic, potentially lethal event; most cases fall somewhere in between. Occlusive disease of the portal vein or its major branches typically produces abdominal pain and, in most instances, manifestations of portal hypertension, principally esophageal varices that are prone to rupture.

The obstruction may occur at the level of the extrahepatic portal vein, intrahepatic portal vein radicles, or small portal vein branches. Extrahepatic portal vein obstruction may be idiopathic (approximately one-third of cases) or may arise from the following conditions:

- *Neonatal umbilical sepsis or umbilical vein catheterization,* often producing subclinical occlusion of the portal vein that presents as variceal bleeding and ascites years later
- *Intraabdominal infection* caused by acute diverticulitis or appendicitis, leading to pyelophlebitis in the splanchnic circulation
- *Inherited or acquired hypercoagulable states,* including myeloproliferative neoplasms such as polycythemia vera (Chapter 13), in which platelet abnormalities predispose to portal vein thrombosis
- *Trauma,* surgical or otherwise
- *Pancreatitis- and pancreatic cancer–associated splenic vein thrombosis,* which propagates into the portal vein
- *Invasion of the portal vein by hepatocellular carcinoma*
- *Cirrhosis,* which is associated with portal vein thrombosis in about 25% of patients. Many such patients have an underlying thrombophilic genotype (e.g., Factor V Leiden, Chapter 4).

Intrahepatic portal vein radicles are most likely to be obstructed by acute thrombosis, often in the setting of a malignancy or some other hypercoagulable state. By contrast, small portal vein branch obstruction may be seen in a variety of pathogenically distinct conditions characterized by noncirrhotic portal hypertension, as follows:

- *The most common cause of small portal vein branch obstruction is schistosomiasis;* the eggs of the parasites and the associated granulomatous inflammatory response obstruct the smallest portal vein branches.
- *Other diseases associated with obstruction of small portal vein branches are collectively referred to as idiopathic noncirrhotic portal hypertension.* The pathogenesis is unknown; associations with prothrombotic states, infections (e.g., HIV), drugs, toxins, immune deficiencies, chronic biliary obstruction, and autoimmune diseases have been noted. Geographic variation in the incidence of noncirrhotic portal fibrosis and hypertension has been reported. It is particularly common in India, although the incidence appears to be declining. Patients often present with upper gastrointestinal bleeding. In East Asia, particularly Japan, there is a female predominance and patients present with splenomegaly, often in association with rheumatologic diseases. The disease is seen in untreated HIV disease and in those being treated with antiretroviral therapy, in whom it may represent a complication of treatment. Liver transplantation may be necessary to avoid fatal sequelae of portal hypertension in all these settings.

Impaired Blood Flow Through the Liver

The most common intrahepatic cause of blood flow impairment is cirrhosis, as described earlier. In addition, physical occlusion of sinusoidal blood flow occurs in a small group of diseases, including the following:

- *Sickle cell disease* (Fig. 18.42), due to obstruction by sickled red cells
- *Disseminated intravascular coagulation*, due to myriad small thrombi
- *Eclampsia* (discussed later)
- *Diffuse intrasinusoidal metastatic tumor,* due to physical obstruction by plugs of tumor, sometimes with superimposed thrombosis

In all of these scenarios, obstruction of blood flow may lead to massive necrosis of hepatocytes and acute liver failure.

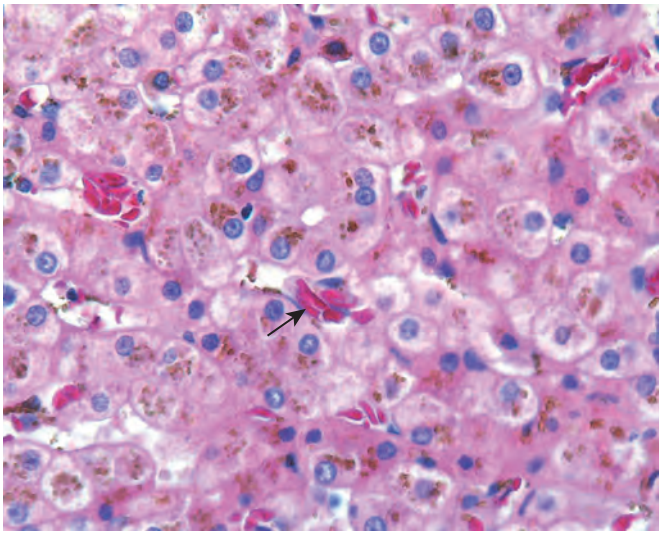


Figure 18.42 Sickle cell crisis in liver. The photomicrograph shows several sinusoids containing "sickled" red cells *(arrow).*

Peliosis hepatis is a peculiar form of sinusoidal dilation that occurs in any condition in which efflux of hepatic blood is impeded. The liver contains blood-filled cystic spaces, either unlined or lined with sinusoidal endothelial cells. The pathogenesis is unknown. *Bartonella* species have been seen in the sinusoidal endothelial cells in AIDS-associated peliosis, but peliosis is also seen in cancer and with other infections, such as tuberculosis. Sex hormone administration (e.g., anabolic steroids, oral contraceptives, danazol) sometimes causes peliosis as well. While clinical signs are generally absent, potentially fatal intraabdominal hemorrhage or hepatic failure may occur. Lesions usually disappear after correction of the underlying cause.

Hepatic Venous Outflow Obstruction

Hepatic Vein Thrombosis

The obstruction of major hepatic veins produces liver enlargement, pain, and ascites, a condition known as Budd-Chiari syndrome. Obstruction of a single main hepatic vein by thrombosis is clinically silent. Hepatic damage is the consequence of increased intrahepatic blood pressure. Hepatic vein thrombosis is associated with myeloproliferative neoplasms such as polycythemia vera (Chapter 13), inherited disorders of coagulation (Chapter 4), antiphospholipid antibody syndrome, paroxysmal nocturnal hemoglobinuria (Chapter 14), and intraabdominal cancers, particularly hepatocellular carcinoma. In pregnancy or with oral contraceptive use, it occurs through interaction with an underlying thrombophilic disorder.

MORPHOLOGY

In Budd-Chiari syndrome, the liver is swollen and red-purple and has a tense capsule (Fig. 18.43). There may be areas of hemorrhagic collapse alternating with preserved or regenerating parenchyma, with the patterns being dependent on which small and large hepatic veins are obstructed. Microscopically, the affected hepatic parenchyma reveals severe centrilobular congestion and necrosis. Pericentral/sinusoidal fibrosis develops in instances in which thrombosis develops more slowly. The major veins contain thrombi showing various degrees of organization.

Clinical Features

The mortality of untreated acute hepatic vein thrombosis is high. Prompt surgery to create a portosystemic venous

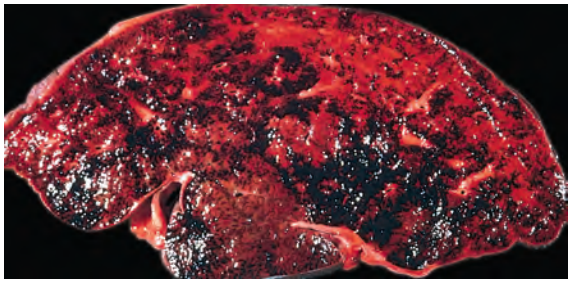


Figure 18.43 Budd-Chiari syndrome. Thrombosis of the major hepatic veins has caused hemorrhagic liver necrosis.

shunt permits reverse flow through the portal vein and improves the prognosis. The chronic form is far less lethal, and more than two-thirds of patients are alive after 5 years.

Sinusoidal Obstruction Syndrome

Originally described in Jamaican drinkers of pyrrolizidine alkaloid–containing bush tea and named *veno-occlusive disease,* sinusoidal obstruction syndrome now occurs primarily in two settings: (1) following allogeneic (or less commonly, autologous) hematopoietic stem cell transplantation, usually within the first 3 weeks (though changes in conditioning regimens have lowered the incidence); and (2) in cancer patients receiving certain forms of chemotherapy. The mortality rate is up to 80% in severe disease.

Pathogenesis

Sinusoidal obstruction syndrome arises from toxic injury to the sinusoidal endothelium. Injured, sloughed endothelium obstructs sinusoidal blood flow, and associated debris accumulates in the terminal hepatic vein. Red cells enter the space of Disse in the disrupted sinusoids, and the cessation of blood flow often leads to necrosis of perivenular hepatocytes.

MORPHOLOGY

Sinusoidal obstruction syndrome is characterized by obliteration of the terminal hepatic venules by swollen or necrotic endothelium, edema, and eventual collagen deposition. In acute disease, there is centrilobular congestion, hepatocellular necrosis, and accumulation of hemosiderin-laden macrophages. In advanced lesions, the obliterated lumens of venules can be highlighted by special stains for connective tissue (Fig. 18.44). In chronic or healed sinusoidal obstruction syndrome, complete fibrous obliteration of the venule may follow.

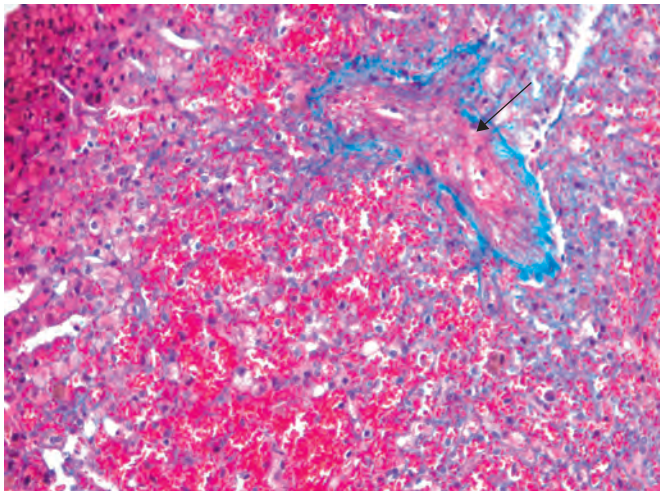


Figure 18.44 Sinusoidal obstruction syndrome. Collagen stain reveals marked sinusoidal congestion, hepatocyte atrophy and loss, and organizing thrombus within the vein lumen *(arrow).* (Masson trichrome stain.)

Clinical Features

Although histology is the gold standard for diagnosis, liver biopsy is often risky in these patients. As a result, the diagnosis is typically made on clinical grounds based on the findings of tender hepatomegaly, ascites, weight gain, and jaundice, as well as reversed or attenuated hepatic venous flow on Doppler ultrasonography. Early results suggest that treatment with anticoagulants and ursodeoxycholate may lower the incidence and severity of sinusoidal obstruction syndrome in patients undergoing hematopoietic stem cell transplantation.

Passive Congestion and Centrilobular Necrosis

These hepatic manifestations of systemic circulatory compromise—passive congestion and centrilobular necrosis—are considered together because they represent a morphologic continuum. Both changes are commonly seen at autopsy because there is an element of preterminal circulatory failure in virtually every nontraumatic death.

MORPHOLOGY

Right-sided cardiac decompensation leads to passive congestion of the liver. The liver is slightly enlarged, tense, and cyanotic, with rounded edges. Microscopically, there is **congestion of dilated centrilobular sinusoids.** With time, centrilobular hepatocytes become atrophic, resulting in markedly attenuated liver cell plates.

Left-sided cardiac failure or shock may lead to hepatic hypoperfusion and hypoxia of hepatocytes around central veins. The combination of hypoperfusion and retrograde congestion acts synergistically to cause **centrilobular hemorrhagic necrosis.** The liver takes on a variegated mottled appearance, reflecting hemorrhage and necrosis in the centrilobular regions (Fig. 18.45*A*). This finding is known as **nutmeg liver** due to its resemblance to the cut surface of a nutmeg. In most instances, the only clinical evidence of centrilobular necrosis or its variants is transient elevation of serum aminotransferases, but the parenchymal damage may be sufficient to induce mild to moderate jaundice.

By microscopy, there is a sharp demarcation of viable periportal and necrotic or atrophic pericentral hepatocytes, with suffusion of blood through the centrilobular region (see Fig. 18.45B). With sustained chronic severe congestive heart failure, **cardiac sclerosis** develops with centrilobular pericellular fibrosis, sometimes with bridging fibrous septa.

KEY CONCEPTS

CIRCULATORY DISORDERS

- Circulatory disorders of the liver can be caused by impaired blood inflow, defects in intrahepatic blood flow, and obstruction of blood outflow.
- Portal vein obstruction by intrahepatic or extrahepatic thrombosis may cause portal hypertension, esophageal varices, and ascites.
- The most common cause of impaired intrahepatic blood flow is cirrhosis.

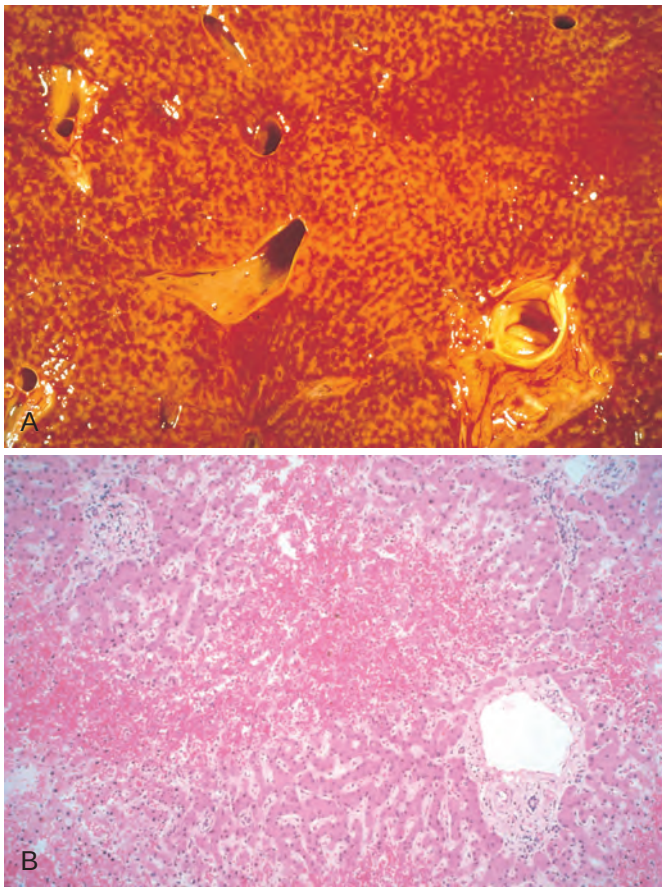


Figure 18.45 Acute passive congestion ("nutmeg liver"). (A) The cut surface of the liver has a variegated mottled red appearance, representing congestion and hemorrhage in the centrilobular regions of the parenchyma. (B) On microscopic examination, the centrilobular region is suffused with red blood cells, and atrophied hepatocytes are not easily seen. Portal tracts and the periportal parenchyma are intact.

- Obstructions of blood outflow include hepatic vein thrombosis (Budd-Chiari syndrome) and sinusoidal obstruction syndrome, previously known as veno-occlusive disease.

HEPATIC DISEASE ASSOCIATED WITH PREGNANCY

Some forms of hepatic disease may be exacerbated by pregnancy. Viral hepatitis (HAV, HBV, HCV, or HBV + HDV) is the most common cause of jaundice in pregnancy. While these women require careful clinical management, pregnancy does not specifically alter the course of viral hepatitis, with exception of HEV infection, which, for unknown reasons, runs a more severe course and has a fatality rate approaching 20% in pregnant patients. The liver may also be secondarily involved by other infections during pregnancy. These include hepatitis caused by herpes simplex virus, a rare cause of acute liver failure in pregnancy, and liver abscess caused by *Listeria monocytogenes,* an organism that thrives in placental tissue, from where it may seed the liver.

Abnormal liver tests occur in 3% to 5% of pregnancies, but in most cases these are of no clinical importance. In a very small subgroup of pregnant women (0.1%), more serious hepatic complications develop. These disorders include preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy. In extreme cases, eclampsia and acute fatty liver of pregnancy may be fatal.

Preeclampsia and Eclampsia

Preeclampsia affects up to 10% of pregnancies and is characterized by maternal hypertension, proteinuria, peripheral edema, and coagulation abnormalities (see Chapter 22 for detailed discussion). When hyperreflexia and convulsions occur, the condition is called *eclampsia* and may be life-threatening. Alternatively, subclinical hepatic disease may be the primary manifestation of preeclampsia, as part of a syndrome of *h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelets, dubbed the *HELLP syndrome.* Here we focus on the hepatic pathology of these entities.

MORPHOLOGY

In preeclampsia, **periportal sinusoids contain fibrin deposits associated with hemorrhage into the space of Disse,** leading to periportal hepatocellular coagulative necrosis. Blood under pressure may coalesce and expand to form a hepatic hematoma; dissection of blood under Glisson capsule may lead to catastrophic hepatic rupture in eclampsia (Fig. 18.46).

Clinical Features

Patients with hepatic involvement in preeclampsia may show modest to severe elevation of serum aminotransferases and mild elevation of serum bilirubin. Hepatic dysfunction sufficient to cause a coagulopathy signifies advanced and potentially lethal disease. Mild cases may be managed conservatively. Termination of pregnancy is required in severe cases. Women who survive mild or severe preeclampsia recover without sequelae.

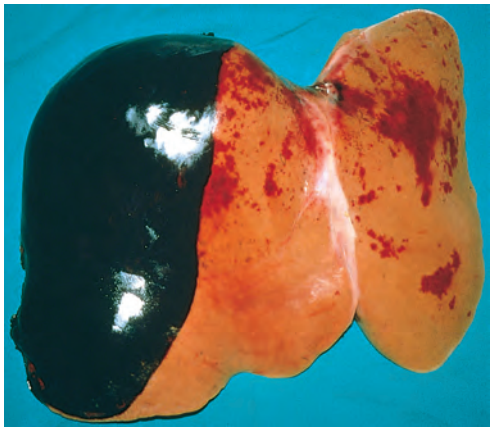


Figure 18.46 Eclampsia. Subcapsular hematoma dissecting under Glisson capsule in a fatal case. (Courtesy Dr. Brian Blackbourne, Office of the Medical Examiner, San Diego, Calif.)

Acute Fatty Liver of Pregnancy

Acute fatty liver of pregnancy presents with a spectrum of disease ranging from subclinical or modest hepatic dysfunction (evidenced by elevated serum aminotransferase levels) to hepatic failure, coma, and death. It is a rare disease affecting approximately 1 in 16,000 pregnant women.

Pathogenesis

The pathogenesis of this disease in some cases may involve an inherited disorder of metabolism. In a subset of patients, both mother and father carry a heterozygous deficiency in mitochondrial long-chain 3-hydroxyacyl coenzyme A (CoA) dehydrogenase. The homozygous-deficient fetuses and placenta fare well during pregnancy but cause hepatic dysfunction in the mother, because long-chain 3-hydroxylacyl metabolites produced by the fetus, or placenta, enter the maternal circulation and cause hepatic toxicity. Thus, this may be a rare instance of the fetus causing metabolic disease in the mother.

MORPHOLOGY

The diagnosis of acute fatty liver of pregnancy rests on biopsy identification of characteristic **diffuse microvesicular steatosis** in hepatocytes. In severe cases, there may be lobular disarray with hepatocyte dropout, reticulin collapse, and portal tract inflammation, making distinction from viral hepatitis difficult. Cholestasis may also be present.

Clinical Features

While this condition most commonly runs a mild course, women with acute fatty liver of pregnancy can progress within days to hepatic failure and death. The primary treatment is termination of pregnancy. Affected women present in the latter half of pregnancy, usually in the third trimester. Symptoms are directly attributable to incipient hepatic failure, including bleeding, nausea and vomiting, jaundice, and coma. In 20% to 40% of cases, the presenting symptoms may be those of coexistent preeclampsia.

Intrahepatic Cholestasis of Pregnancy

The onset of pruritus in the second or third trimester, followed in some cases (10% to 25%) by darkening of the urine, and occasionally light stools and jaundice, heralds the development of this enigmatic syndrome, which resolves within 2 to 3 weeks of delivery. Serum bilirubin (mostly conjugated) rarely exceeds 5 mg/dL; alkaline phosphatase may be slightly elevated. The level of bile salts is increased greatly. The altered hormonal state of pregnancy likely combines with environmental factors and transporter protein defects to engender canalicular cholestasis. There is a modest risk of fetal loss, and the condition may recur in subsequent pregnancies. The pruritus resulting from retention of bile salts can be extremely distressing for the pregnant mother.

NODULES AND TUMORS

Hepatic mass lesions include non-neoplastic processes like focal nodular hyperplasia, regenerative nodules, and abscess or another infectious process. Neoplastic mass lesions include benign hepatocellular nodules such as hepatocellular adenoma; malignant primary liver epithelial neoplasms, such as hepatocellular carcinoma or intrahepatic cholangiocarcinoma; and nonepithelial neoplasms, such as angiosarcoma and metastatic tumors.

Non-Neoplastic Mass Lesions

Focal Nodular Hyperplasia

Focal nodular hyperplasia (FNH) is a benign, non-neoplastic lesion that is most common in adult women.

Pathogenesis

FNH is thought to result from altered blood flow that leads to hyperplastic changes in hepatocytes. FNH-like nodules can occur adjacent to other tumors and in vascular conditions such as hemangiomas and Budd-Chiari syndrome. Multifocal FNH can occur in association with hepatic or extrahepatic hemangioma, vascular malformations, or (curiously) brain tumors, such as meningioma and astrocytoma.

MORPHOLOGY

FNH is well circumscribed, typically lacks a capsule, and has a **central stellate scar** in up to 80% of cases (Fig. 18.47A). Most lesions are 5 cm or less in diameter. Microscopically, fibrous septa radiate from the central scar dividing the hepatic parenchyma into nodules. The fibrous septa and central scar contain thick-walled arteries with intimal and fibromuscular hyperplasia and poorly formed elastic lamina (Fig. 18.47B). A prominent ductular reaction is often present, but interlobular bile ducts are absent. The hepatocytes in the nodules are arranged in one- to two-cell-thick plates; cytologic and architectural atypia is not present. Immunohistochemistry for glutamine synthetase (GS), an enzyme normally present in centrizonal hepatocytes, is very useful for diagnosis. In FNH, a highly characteristic "maplike pattern" of strong cytoplasmic GS staining is seen within anastomosing groups of hepatocytes.

Clinical Features

Most cases are asymptomatic and come to clinical attention during imaging or surgery. The presence of a central scar is a highly characteristic finding and helps in establishing the diagnosis on imaging. FNH is a benign lesion and does not need any treatment. Resection is performed for large symptomatic lesions.

Other Non-Neoplastic Mass Lesions

Non-neoplastic mass forming lesions like abscesses, granulomas, and inflammatory pseudotumors can mimic true neoplasms. Large regenerative nodules can occur in non-cirrhotic liver in vascular conditions such as Budd-Chiari

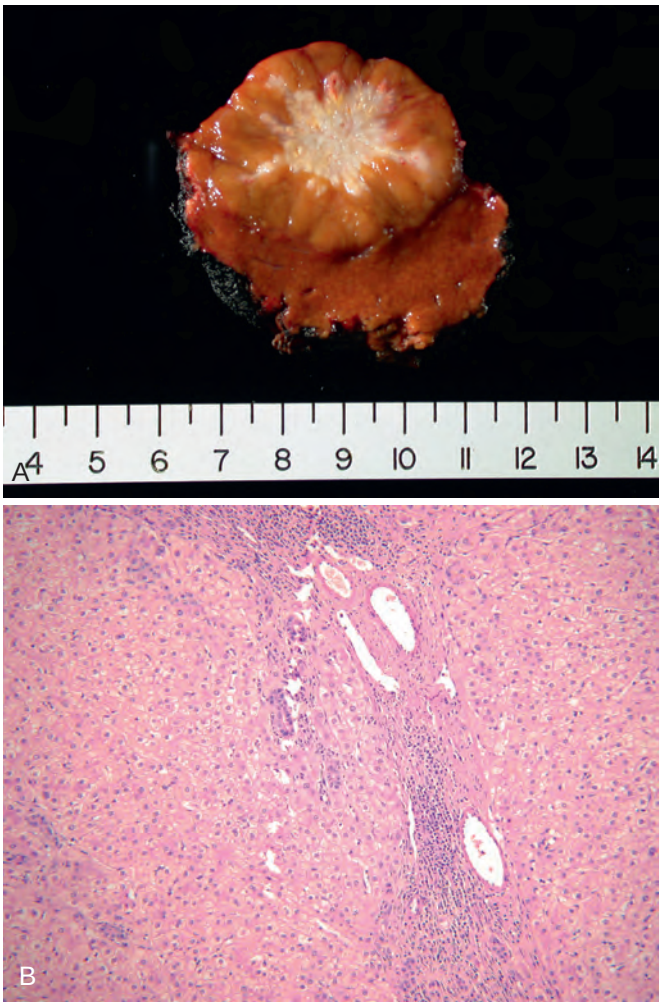


Figure 18.47 Focal nodular hyperplasia. (A) Resected specimen showing lobulated contours and a central stellate scar. (B) Low-power micrograph showing a broad fibrous scar with thick-walled arteries and ductular reaction, but no interlobular bile ducts.

syndrome, hemangioma, and vascular malformations. In some conditions, the entire liver becomes nodular without significant fibrosis, a condition referred to as nodular regenerative hyperplasia (NRH). NRH can lead to portal hypertension and mimic cirrhosis on imaging.

Benign biliary proliferations include bile duct adenoma and biliary hamartoma (von Meyenburg complex, described earlier). Both are asymptomatic and generally less than 2 cm. These lesions come to clinical attention as tiny nodules or cysts on imaging, or as small subcapsular lesions observed during abdominal surgery when they can be mistaken for metastatic disease. Both lesions show haphazard bile ducts in a fibrous stroma, and lack atypical cytologic or architectural features. The ductular profiles in bile duct adenoma tend to be round and nondilated, while those in biliary hamartoma are dilated, curvilinear and often show inspissated bile or secretions. The latter are thought to be a remnant of ductal plate malformation. The recent identification of pathogenic acquired *BRAF* mutations in the majority of bile duct adenomas indicates that this is a benign neoplasm, settling a longstanding debate.

Benign Neoplasms

Cavernous Hemangioma

Cavernous hemangioma is the most common benign liver tumor. It appears as a discrete red-blue, soft nodule, usually less than 2 cm in diameter, generally in a subcapsular location. Histologically, the tumor consists of dilated thin-walled vascular channels (Fig. 18.48). Most are asymptomatic and are detected incidentally by imaging.

Hepatocellular Adenoma

Hepatocellular adenoma is a benign neoplasm that typically occurs in young women and is strongly associated with use of oral contraceptives and anabolic steroids. The incidence has increased in the past decade, presumably related to obesity and metabolic syndrome. Multiple hepatocellular adenomas (termed *hepatic adenomatosis* when 10 or more tumors are present) can occur both in familial and acquired settings.

Pathogenesis

Three molecular subtypes have been defined, each associated with distinct clinicopathologic features and varying risk of transformation into hepatocellular carcinoma.

- *Hepatocyte nuclear factor 1-alpha (HNF1α)-inactivated hepatocellular adenoma*. All of these tumors by definition have loss-of-function mutations in the *HNF1-α* gene, which encodes a transcription factor that regulates many genes in hepatocytes and in pancreatic islets. Heterozygous germline mutations are responsible for autosomal dominant maturity onset diabetes of the young, type 3, and are also associated with 10% of HNF1α-inactivated hepatocellular adenomas. This subtype accounts for 40% to 50% of cases, has a strong female predilection, and is associated with minimal risk of transformation to hepatocellular carcinoma.
- *Inflammatory hepatocellular adenoma* is a subtype that results from acquired activating mutations in gp130, a co-receptor for IL-6, that lead to constitutive activation of JAK-STAT signaling. This subtype accounts for 40% to 50% of cases,

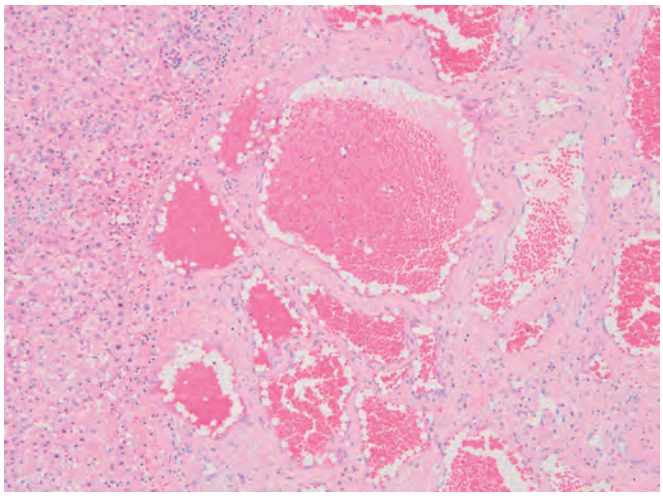


Figure 18.48 Cavernous hemangioma. Blood-filled vascular channels separated by a dense fibrous stroma.

is more common in women, and is associated with obesity and metabolic syndrome. β-catenin–activating mutations also are present in 10% of inflammatory hepatocellular adenomas, and these tumors have a higher risk of malignant transformation.

- *β-Catenin–activated hepatocellular adenoma* is defined by activating mutations in the β-catenin gene *(CTNNB1)* or in other components of the Wnt pathway (such as *APC*). These tumors are at high risk for malignant transformation to hepatocellular carcinoma and are associated with oral contraceptive and anabolic steroid use. Nearly 40% occur in men.

MORPHOLOGY

Hepatocellular adenomas appear as discrete masses, sometimes with associated hemorrhage, composed of hepatocytes arranged in one- to two-cell-thick plates (Fig. 18.49). Large arteries are present within a small amount of fibrous stroma, while interlobular bile ducts and normal portal tracts are absent. *HNF1α*-mutated tumors typically show prominent fat in the lesional hepatocytes and by immunohistochemistry, liver fatty acid binding protein (LFABP), a protein regulated by *HNF1α*, is absent (Fig. 18.50A).

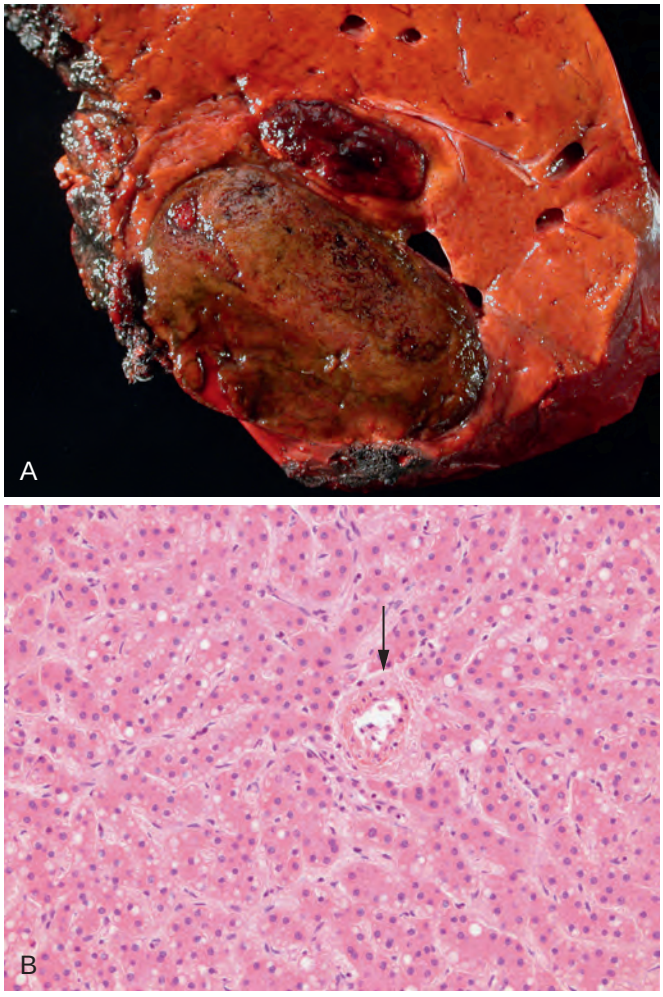


Figure 18.49 Hepatocellular adenoma. (A) Resected specimen showing a well-define tan mass in the liver. (B) Microscopic view showing thin cords of hepatocytes, with an arterial vascular supply *(arrow)* and no portal tracts.

β-catenin–mutated tumors often show atypical cytologic or architectural features, and areas showing overt transformation to hepatocellular carcinoma may be present. Due to β-catenin mutations or other mutations that activate Wnt signaling, strong nuclear localization of β-catenin is seen within at least a subset of the tumor cells (see Fig. 18.50B). Inflammatory tumors typically show sinusoidal dilation and may have features mimicking focal nodular hyperplasia, such as fibrous septa and ductular reaction. Because of hyperactive JAK-STAT signaling, these tumors overexpress acute-phase reactants, such as C-reactive protein and serum amyloid A (see Fig. 18.50C).

Clinical Features

Hepatocellular adenoma may be detected incidentally by imaging or come to attention due to symptoms related abdominal pain or hemorrhagic necrosis. Rupture of hepatocellular adenomas can cause massive abdominal bleeding and represents a surgical emergency. Resection is recommended for β-catenin–activated tumors as well as tumors 5 cm or larger due to the risk of hemorrhage and malignant transformation. For smaller tumors without β-catenin activation, close follow-up and cessation of exposure to oral contraceptive or anabolic steroids may suffice.

Primary Malignant Neoplasms

Malignant tumors in the liver can be primary or metastatic. Among primary epithelial tumors, the most common are hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Hepatoblastoma is a rare hepatocellular tumor that occurs in the pediatric setting. Nonepithelial tumors, such as angiosarcoma, are exceedingly rare.

Hepatoblastoma

Hepatoblastoma is the most common liver tumor of early childhood. It rarely occurs over the age of 3 years, and its incidence is increasing. Activation of the Wnt signaling pathway is characteristic of hepatoblastoma, leading to nuclear β-catenin and increased expression of Wnt target genes such as glutamine synthetase in nearly all cases. These tumors are associated with several syndromes, including *familial adenomatous polyposis* (caused by germline loss-of-function mutations in *APC;* Chapter 17) and *Beckwith-Wiedemann syndrome*. The latter is a disorder associated with congenital growth abnormalities caused by abnormal epigenetic imprinting of a region of chromosome 11 that leads to overexpression of insulin-like growth factor-2 and loss of expression of *CDKN1C,* a tumor suppressor gene that encodes the cyclin-dependent kinase inhibitor p57.

MORPHOLOGY

There are a variety of histologic subtypes of hepatoblastoma, which can be broadly grouped into two categories: (1) **epithelial type,** composed of small polygonal fetal cells or smaller embryonal cells forming trabeculae, acini, tubules, or papillary structures vaguely recapitulating the developing liver (Fig. 18.51); and (2) **mixed epithelial and mesenchymal type,** which contains additional foci of mesenchymal differentiation that may consist of primitive mesenchyme, osteoid, cartilage, or striated muscle.

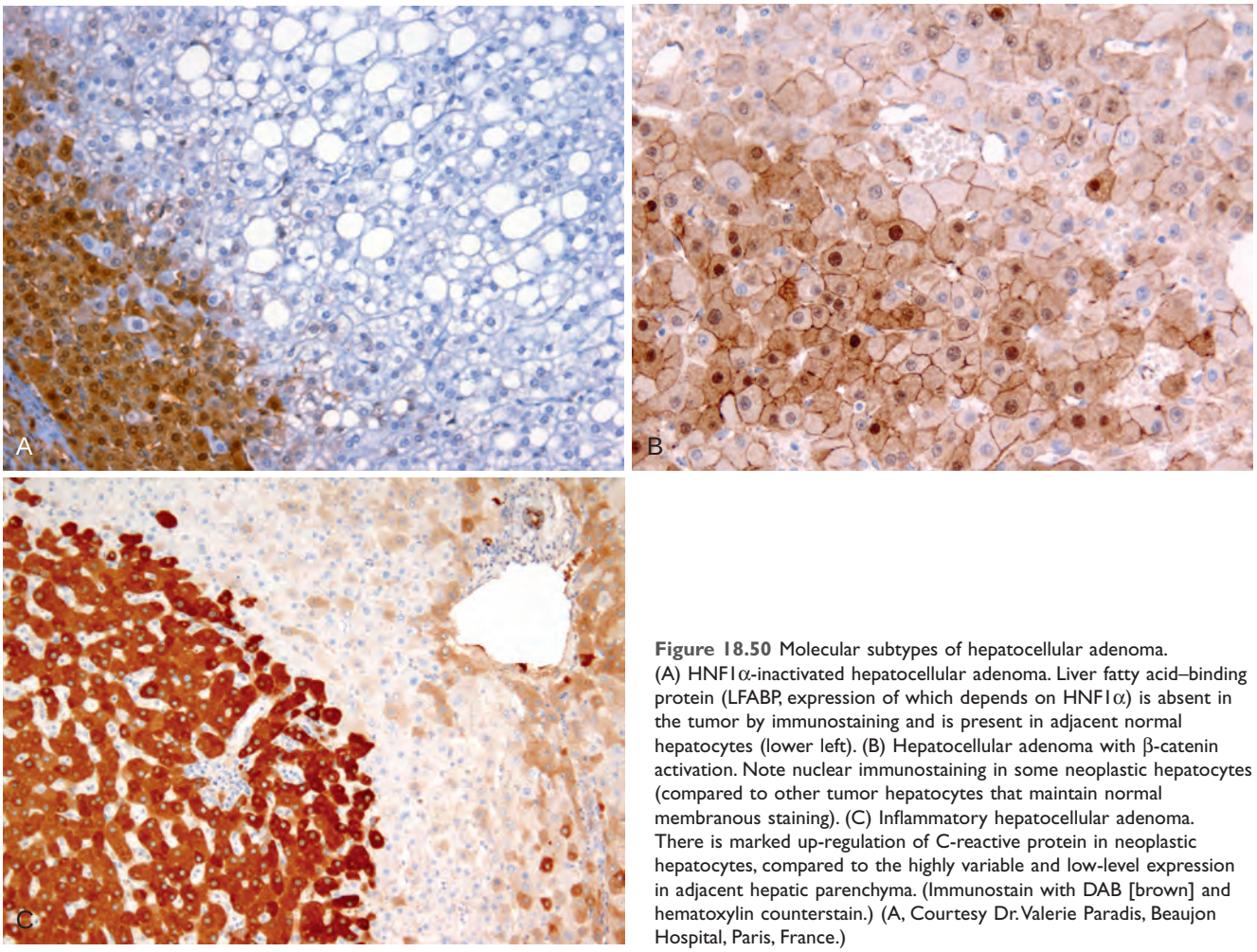


Figure 18.50 Molecular subtypes of hepatocellular adenoma. (A) HNF1α-inactivated hepatocellular adenoma. Liver fatty acid–binding protein (LFABP, expression of which depends on HNF1α) is absent in the tumor by immunostaining and is present in adjacent normal hepatocytes (lower left). (B) Hepatocellular adenoma with β-catenin activation. Note nuclear immunostaining in some neoplastic hepatocytes (compared to other tumor hepatocytes that maintain normal membranous staining). (C) Inflammatory hepatocellular adenoma. There is marked up-regulation of C-reactive protein in neoplastic hepatocytes, compared to the highly variable and low-level expression in adjacent hepatic parenchyma. (Immunostain with DAB [brown] and hematoxylin counterstain.) (A, Courtesy Dr. Valerie Paradis, Beaujon Hospital, Paris, France.)

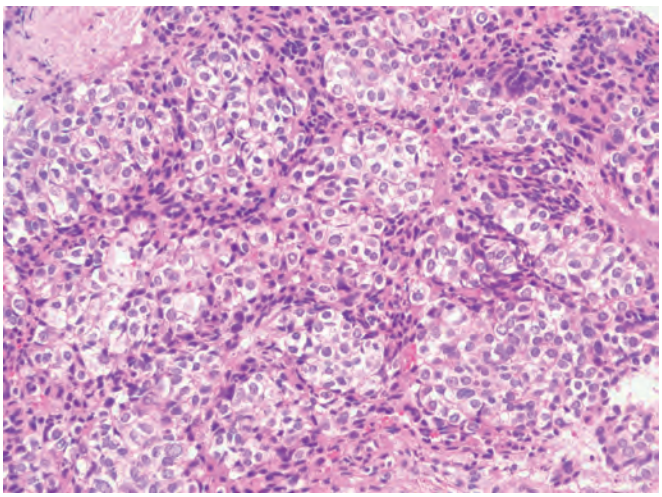


Figure 18.51 Hepatoblastoma. The tumor cells are arranged in sheets and resemble embryonal and fetal hepatocytes.

Clinical Features

Hepatoblastoma usually comes to clinical attention due to abdominal swelling in an asymptomatic infant or child. Symptoms related to liver dysfunction (jaundice, pruritis) are seen in a subset of patients, and about 20% of tumors will have metastasized to the lungs by the time of diagnosis. The tumor is treated with surgical resection and chemotherapy, which has improved outcomes. The overall 5-year survival rate is approximately 80%.

Hepatocellular Carcinoma (HCC)

HCC accounts for approximately 5.4% of all cancers worldwide and is one of the most common cancers in geographic regions with high rates of hepatitis B infection. More than 85% of cases occur in countries in Asia (southeast China, Korea, Taiwan) and sub-Saharan Africa, where chronic HBV infection is common. The peak incidence of HCC in these areas is in young adults between 20 and 40 years of age who acquired hepatitis B virus by maternal-fetal transmission. Encouragingly, the incidence of HCC is decreasing in Asia due to hepatitis B vaccination, but at the same time the incidence is increasing in the Western countries owing to rising rates of hepatitis C infection and metabolic syndrome. For unclear reasons, there is a pronounced male predominance, as high as 8:1, in high-incidence areas.

Pathogenesis

Most HCCs occur in the setting of chronic liver disease with cirrhosis, while 15% to 20% arise in noncirrhotic livers (Fig. 18.52). The most common underlying diseases are chronic viral hepatitis (B and C), metabolic diseases

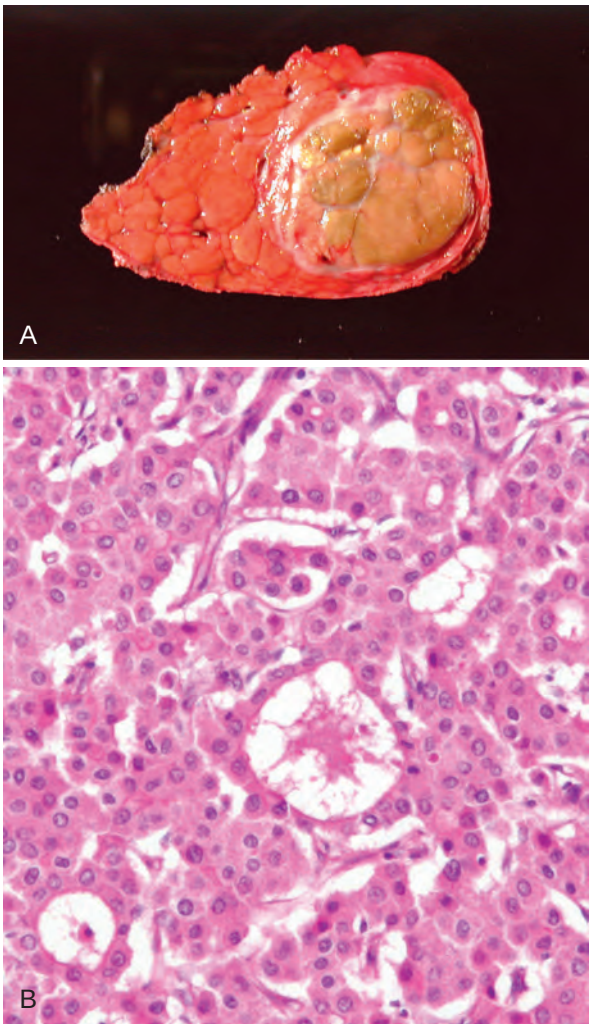


Figure 18.52 Hepatocellular carcinoma. (A) Liver removed at autopsy showing a unifocal neoplasm replacing most of the right hepatic lobe. (B) Malignant hepatocytes growing in distorted versions of normal architecture, including large pseudoacinar spaces (malformed, dilated bile canaliculi) and thickened hepatocyte trabeculae.

such as hereditary hemochromatosis and α_1-antitrypsin deficiency, and alcoholic liver disease. Nonalcoholic fatty liver disease also increases the risk of HCC, even in the absence of cirrhosis. Although details are not clearly worked out, it is believed that the chronic injury, inflammation, and hepatocyte regeneration that are seen in these disorders contribute to the acquisition of driver mutations that lead to HCC development (described later). Part of the risk in Africa and Asia appears to be related to contamination of crops by aflatoxin, a mycotoxin produced by *Aspergillus* species that acts synergistically with alcohol and hepatitis B. The risk for HCC in cirrhosis related to other etiologies, like Wilson disease and chronic biliary diseases, is somewhat lower but still elevated above the population average.

As with other cancers, HCC is associated with complementary sets of driver mutations that lead to the acquisition of cancer hallmarks (Chapter 7). Among the most common are activating mutations in the β-catenin gene (40% of tumors), mutations in the *TERT* (telomerase transcriptase) gene promoter that up-regulate telomerase activity (50% to 60% of tumors), and inactivating mutations in *TP53* (up to 60% of tumors). One unusual histologic subtype that often occurs in adolescents and young adults in the absence of preexisting liver disease, *fibrolamellar HCC,* is strongly associated with a fusion gene that leads to aberrant activity of protein kinase A, an enzyme that participates in a signaling pathway regulated by cAMP.

MORPHOLOGY

Several precursor lesions for HCC have been described. As discussed earlier, in noncirrhotic liver, HCC can arise in hepatocellular adenoma, especially those with β-catenin–activating mutations. In chronic liver disease, the earliest morphologic alterations that appear to correlate with the presence of "at-risk" hepatocytes are called "large cell change" and "small cell change" (Fig. 18.53). **Large cell change** refers to hepatocytes that are larger than normal and often have enlarged, multiple, pleomorphic nuclei, without an increase in nuclear-to-cytoplasmic ratio (see Fig. 18.53A). In **small cell change,** the hepatocytes have a high nuclear-to-cytoplasmic ratio and mild nuclear hyperchromasia and/or pleomorphism (see Fig. 18.53B). More ominous are nodular lesions

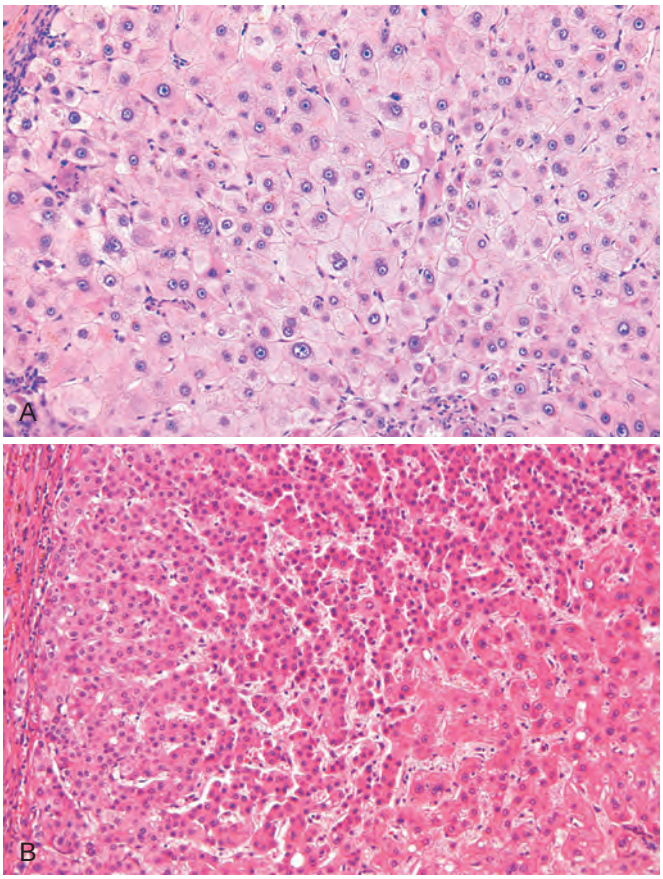
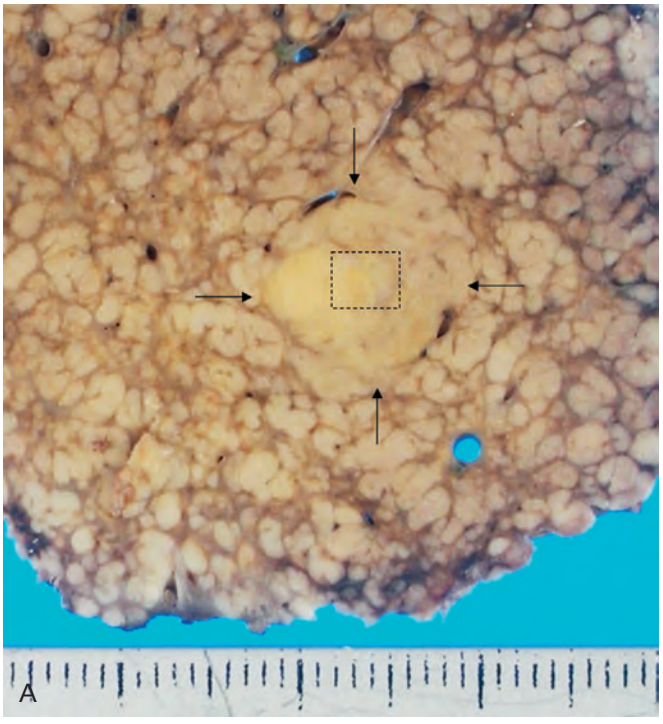


Figure 18.53 Premalignant changes in hepatocytes. (A) Large cell change. Large hepatocytes with large, often atypical nuclei are scattered among normal-size hepatocytes with round, typical nuclei. (B) Small cell change. The abnormal cells have a high nuclear-to-cytoplasmic ratio and are separated by thickened plates. Normal-appearing hepatocytes are in the lower-right corner. (Courtesy Dr. Young Nyun Park, Yonsei Medical College, Seoul, South Korea.)

in cirrhotic liver referred to as **dysplastic nodules** (Fig. 18.54), which are often associated with small cell change. Dysplastic nodules differ from adjacent cirrhotic nodules in size, color, and vascularization, show varying degrees of dysplasia, and have clonal aberrations associated with full-blown HCC. Small areas of HCC may sometimes be seen in high-grade dysplastic nodules ("nodule in nodule appearance") (see Fig. 18.54B).

HCC may form a single mass or multiple discrete masses, or it may diffusely infiltrate the liver. They may be pale and yellow due to fatty change or green due to cholestasis. Tumors larger than 2 cm are more likely to be associated with vascular invasion and intrahepatic metastases. Invasion of veins with extension into the portal vein, inferior vena cava, and even the right side of the heart may occur.



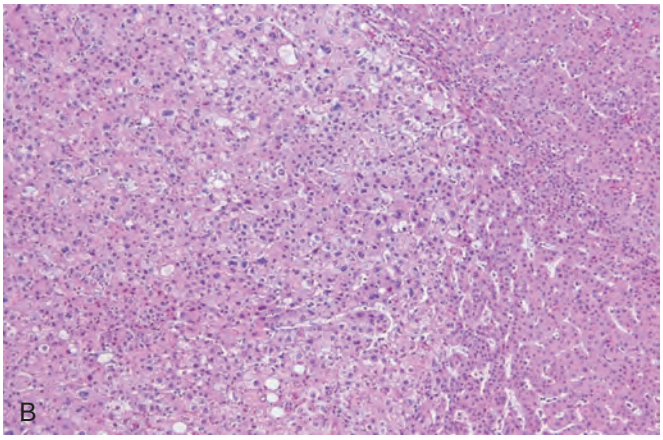
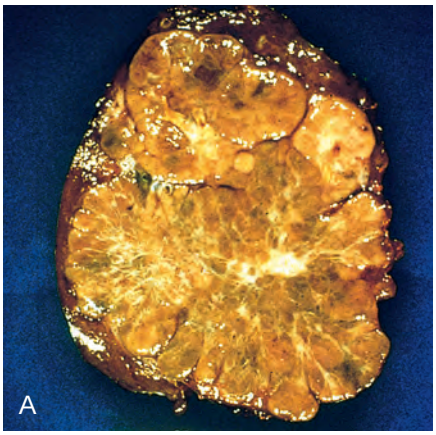


Figure 18.54 (A) Hepatitis C–related cirrhosis with a distinctively large nodule *(arrows)*. Nodule-in-nodule growth suggests an evolving cancer. (B) Histologically, the region with in the box in A shows a well-differentiated hepatocellular carcinoma (HCC) (right side) and a subnodule of moderately differentiated HCC within it (center, left). (Courtesy Dr. Masamichi Kojiro, Kurume University, Kurume, Japan.)



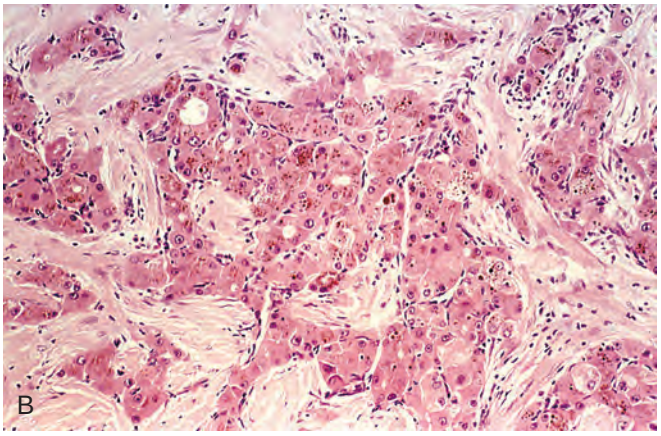


Figure 18.55 Fibrolamellar carcinoma. (A) Resected specimen showing a well-demarcated nodule. (B) Microscopic view showing nests and cords of malignant-appearing, oncocytic hepatocytes separated by dense bundles of collagen.

Microscopically, well and moderately differentiated HCC are composed of cells that resemble normal hepatocytes, while poorly differentiated tumors show marked cytologic atypia. The tumor cells grow in thick plates or trabeculae, pseudoglandular structures with bile plugs, or sheets (see Fig. 18.52B). The distinctive fibrolamellar variant shows a characteristic triad of features: large polygonal cells with granular (oncocytic) cytoplasm due to abundant mitochondria; vesicular nuclei with a prominent nucleolus; and parallel lamellae of dense collagen bundles (Fig. 18.55).

Clinical Features

The clinical manifestations of HCC are nonspecific and include abdominal pain, malaise, fatigue, weight loss, and hepatomegaly. Features of underlying chronic liver disease can be present. Elevated levels of serum α-fetoprotein is a frequent finding in advanced disease, but it is not sensitive as a screening test for early tumors and is not associated with the fibrolamellar variant. Ultrasonography is used for screening high-risk patients such as those with cirrhosis. Computed tomography and magnetic resonance imaging with contrast studies yield highly characteristic findings. Early enhancement of the tumor due to contrast uptake in the arterial phase, followed by rapid venous washout, is considered diagnostic of HCC.

Surgical resection when possible is the treatment of choice for tumors in noncirrhotic livers and in cirrhotic livers with

adequate function. Liver transplantation is considered for HCC in the setting of advanced cirrhosis. Image-guided tumor ablation with alcohol or radiofrequency waves can be done for unresectable tumors or those that do not meet criteria for transplantation. Hematogenous metastases, especially to the lung, tend to occur late in the disease. Lymph node metastases occur in <5% of cases.

Overall outcomes in HCC are poor due to underlying liver disease and the intrinsic resistance of HCC to conventional chemotherapy. The overall 5-year survival rate is 30% for tumor confined to the liver and only 5% to 10% for cases with extrahepatic spread. Outcomes are better for the unusual fibrolamellar variant, with up to 40% of patients surviving 10 years or longer. This is in large part because, in the absence of underlying liver disease, extensive surgical resection is possible thanks to the regenerative capacity of the remaining liver.

Malignant Biliary Tumors

Adenocarcinomas arising from the intrahepatic biliary tree are referred to as *intrahepatic cholangiocarcinoma,* while similar tumor arising from the extrahepatic bile ducts are referred to as *biliary adenocarcinoma*. Intrahepatic cholangiocarcinoma is the most common primary malignant tumor of the liver after HCC. Its incidence is rising in the United States. It accounts for 7.6% of cancer deaths worldwide and 3% of cancer deaths in the United States. It is very common in Southeast Asian countries such as Thailand, Laos, and Cambodia, where liver fluke infestation is endemic.

Pathogenesis

Developmental disorders, like fibropolycystic liver disease, and chronic inflammatory conditions involving the bile ducts, including primary sclerosing cholangitis, infestation by liver flukes (particularly *Opisthorchis* and *Clonorchis* species), and hepatolithiasis are risk factors for biliary tract neoplasms. Chronic liver diseases that predispose to HCC, such as hepatitis B, hepatitis C, and nonalcoholic fatty liver disease, also increase the risk for intrahepatic cholangiocarcinoma. As with HCC, chronic injury, inflammation, and regeneration of biliary epithelium in these conditions may set the stage for acquisition of driver mutations leading to cancer.

Driver mutations in intrahepatic cholangiocarcinoma and extrahepatic biliary adenocarcinoma only partially overlap. *KRAS* mutations are common in both tumors. In addition, subsets of intrahepatic cholangiocarcinoma have distinctive driver mutations in *IDH1* and *IDH2* (isocitrate dehydrogenase) that generate "oncometabolites" (Chapter 7), and in chromatin modifier genes like *BAP1* and *PBRM1*. Fusion genes involving *FGFR2* (fibroblast growth factor receptor 2) are also common. By contrast, in addition to *KRAS* mutations, extrahepatic biliary adenocarcinomas are most likely to have mutations in *TP53* and *SMAD4,* genetic features that are similar to those of pancreatic adenocarcinoma.

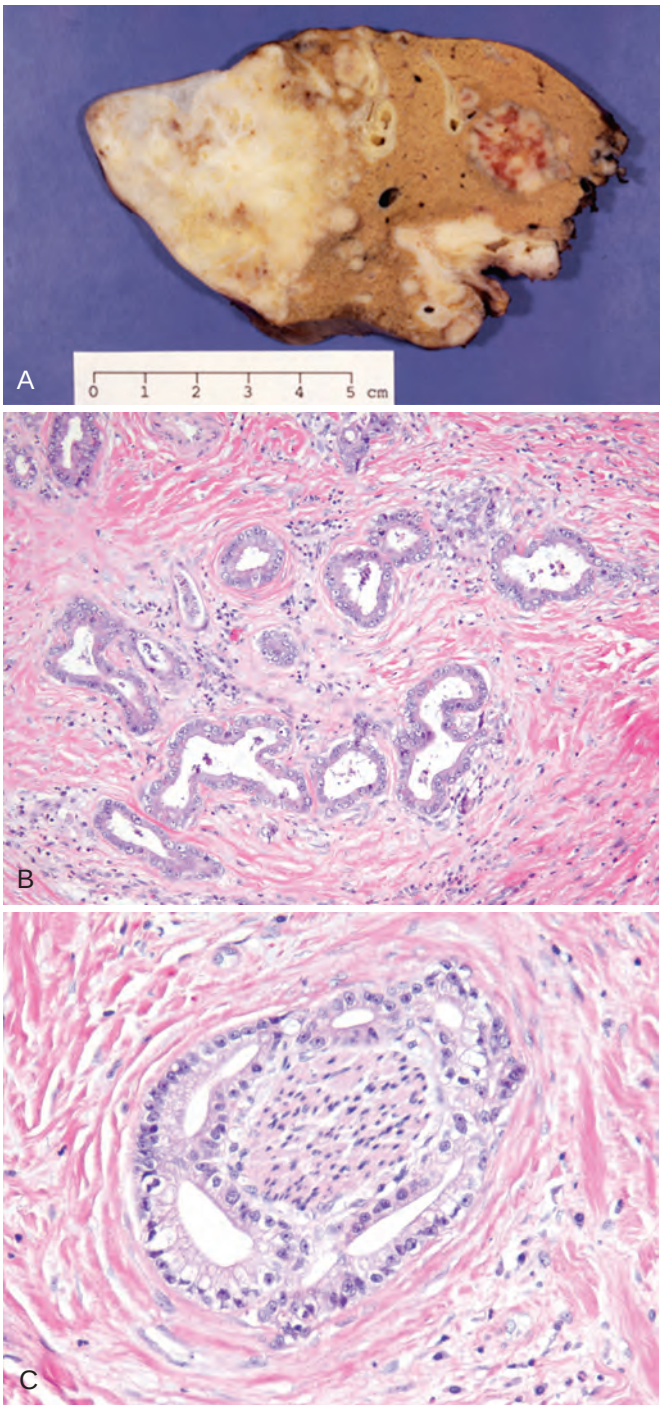


Figure 18.56 Cholangiocarcinoma. (A) Multifocal cholangiocarcinoma in a liver from a patient with infestation by the liver fluke *Clonorchis sinensis*. (B) Invasive malignant glands in a reactive, sclerotic stroma. (C) Perineural invasion by malignant glands, forming a wreathlike pattern around the central, trapped nerve. (A, Courtesy Dr. Wilson M.S. Tsui, Caritas Medical Centre, Hong Kong.)

MORPHOLOGY

Intrahepatic cholangiocarcinoma usually occurs in the noncirrhotic liver (Fig. 18.56) and typically forms a firm gray-white mass. Extrahepatic biliary adenocarcinomas include perihilar tumors known as **Klatskin tumors,** which are located at the junction of the right and left hepatic ducts. These tumors comprise 60% to 70% of extrahepatic biliary adenocarcinomas, while the remaining 30% to 40% involve the common bile duct. Extrahepatic biliary adenocarcinomas are generally small at diagnosis, as they are detected early due to obstructive features. These tumors may

form a firm nodule in the bile duct wall, may be diffusely infiltrative, or may form papillary lesions.

Microscopically, both intrahepatic and extrahepatic tumors show features of adenocarcinomas. Most are well to moderately differentiated and are arranged in clearly defined glandular/tubular structures lined by malignant epithelial cells embedded in an abundant fibrous stroma (see Fig. 18.56B). Lymphovascular and perineural invasion (see Fig. 18.56C) are common. These tumors are thought to usually arise from premalignant lesions termed biliary intraepithelial neoplasia. Other less common precursor lesions are mucinous cystic neoplasms and intraductal papillary neoplasms.

Clinical Features

Intrahepatic cholangiocarcinoma may be detected incidentally on imaging or may present with a cholestatic picture or symptomatic liver mass, whereas extrahepatic adenocarcinoma typically presents with symptoms related to biliary obstruction. Tumor resection with negative margins and absence of lymph node involvement are favorable factors. Lymph node metastasis is present in 50% to 60% of patients at presentation. Adjuvant chemotherapy is commonly given after resection, but tumor responses are usually transient. The overall prognosis is poor, as recurrences are common, and the 5-year survival rate is 20% to 40% after surgical resection.

Other Primary Hepatic Malignant Tumors

A number of other cancers may arise within the liver. Among the most common or pathogenically interesting are the following:

- *Combined hepatocellular-cholangiocarcinoma* is composed of cells resembling both HCC and intrahepatic cholangiocarcinoma. These tumors have the same risk factors as HCC.
- *Angiosarcoma* is a malignant vascular tumor that is historically associated with exposure to vinyl chloride, arsenic, or the imaging agent Thorotrast (Chapters 9 and 11). This tumor has become rare with reduced exposures to these agents.
- *Epithelioid hemangioendothelioma* is another vascular tumor with intermediate malignant potential and a somewhat better prognosis than angiosarcoma, which is almost uniformly fatal.
- *Primary hepatic lymphoma* is a rare type of liver malignancy. The most common subtype is *diffuse large B-cell lymphoma,* a form of non-Hodgkin lymphoma that frequently occurs at extranodal sites (Chapter 13). Another rare subtype is *hepatosplenic T-cell lymphoma,* most common in young adult males, which has a predilection for growth within the sinusoids of the liver, spleen, and bone marrow.

Metastasis

Involvement of the liver by metastatic malignancy is far more common than primary hepatic neoplasia. Although the most common primary sources are the colon, breast, lung, and pancreas, virtually any may spread to the liver. Typically, multiple nodular metastases are found that often cause striking hepatomegaly and replace much of the normal liver parenchyma. The liver weight can exceed several kilograms. Metastasis may also appear as a single nodule, in which case it may be resected surgically. Always surprising is the extent of metastatic involvement that may be present in the absence of clinical or laboratory evidence of hepatic insufficiency; often the only telltale sign is hepatomegaly. However, with massive destruction of hepatocytes or direct obstruction of major bile ducts, jaundice and other evidence of liver dysfunction inevitably appears.

KEY CONCEPTS

LIVER TUMORS

- The liver is the most common site of metastatic cancers from primary tumors of the colon, lung, and breast.
- Hepatocellular adenomas are benign tumors of neoplastic hepatocytes, which can be subclassified on the basis of different sets of driver mutations.
- HCC is the most common primary hepatic malignant neoplasm, often occurring in the setting of cirrhosis related to chronic liver disease such as chronic hepatitis B and C, autoimmune hepatitis, alcoholic/nonalcoholic fatty liver disease, and hemochromatosis.
- Intrahepatic cholangiocarcinoma and extrahepatic biliary adenocarcinoma show similar histologic features and share risk factors like primary sclerosing cholangitis and infestation by liver flukes such as *Opisthorchis* and *Clonorchis*. These tumors have a poor prognosis.

Gallbladder

As much as 1 L of bile is secreted by the liver per day. Between meals, bile is stored in the gallbladder, where it is concentrated. The adult gallbladder has a capacity of about 50 mL. The organ is not essential for biliary function, since humans do not suffer from indigestion or malabsorption of fat after cholecystectomy.

CONGENITAL ANOMALIES

The gallbladder may be congenitally absent, or there may be gallbladder duplication with conjoined or independent cystic ducts. A longitudinal or transverse septum may create a bilobed gallbladder. Aberrant locations of the gallbladder occur in 5% to 10% of the population, most commonly with partial or complete embedding in the liver substance. A folded fundus is the most common anomaly, creating a *phrygian cap* (Fig. 18.57). *Agenesis* of all or any portion of the hepatic or common bile ducts and hypoplastic narrowing of biliary channels (true "biliary atresia") may also occur. *Choledochal cysts,* described earlier, may be isolated findings in the gallbladder or associated with other cysts in the extrahepatic biliary tree or with fibropolycystic disease.

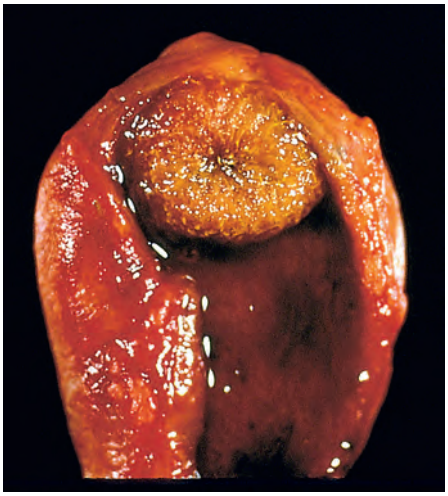


Figure 18.57 Phrygian cap of the gallbladder; the fundus is folded inward.

CHOLELITHIASIS (GALLSTONES)

More than 95% of biliary tract disease is attributable to gallstones. Gallstones afflict 10% to 20% of adult populations in high income countries. It is estimated that more than 20 million persons in the United States have gallstones, totaling some 25 to 50 tons in weight, leading to more than 700,000 cholecystectomies performed annually at a cost of approximately $6 billion.

Epidemiology

There are two general classes of gallstones: cholesterol stones, containing more than 50% of crystalline cholesterol monohydrate, and pigment stones composed predominantly of bilirubin calcium salts, each with different risk factors. Cholesterol gallstones are more prevalent in the United States and Western Europe (90%) and uncommon in low income countries. The prevalence rates of cholesterol gallstones approach 75% in Native Americans of the Pima, Hopi, and Navajo groups, while pigment stones are rare in these populations. Pigment gallstones, the predominant type of gallstone in non-Western populations, arise primarily in the setting of bacterial infections or parasitic infestations of the biliary tree, and as well as in individuals with diseases that lead to chronic red cell hemolysis.

The major risk factors associated with the development of gallstones are listed in Table 18.11 and are briefly described here:

- *Age and sex.* The prevalence of cholesterol gallstones increases throughout life, but they predominantly affect individuals of middle to older age. Prevalence is higher in females in any region or ethnicity; in Caucasian women, it is about twice as high as in men. Hypersecretion of biliary cholesterol seems to play the major role in both age and gender differences. Significant associations are also seen with metabolic syndrome and obesity.
- *Environmental factors.* Estrogen exposure, including through oral contraceptive use and during pregnancy, increases expression of hepatic lipoprotein receptors and stimulates hepatic HMG-CoA reductase activity, enhancing both cholesterol uptake and biosynthesis, respectively. The net result is excess biliary secretion of cholesterol. Obesity and rapid weight loss are also strongly associated with increased biliary cholesterol secretion.
- *Acquired disorders.* Gallbladder stasis, either neurogenic or hormonal, fosters a local environment that is favorable for both cholesterol and pigment gallstone formation.
- *Hereditary factors.* Genes encoding hepatocyte proteins that transport biliary lipids, known as ATP-binding cassette (ABC) transporters, have associations with gallstone formation. In particular, a common variant of the sterol transporter encoded by the *ABCG8* gene is associated with an increased risk of cholesterol gallstones.

Pathogenesis of Cholesterol Stones

Cholesterol is rendered soluble in bile by forming micelles with bile salts and lecithins, both of which act as detergents. When cholesterol concentrations exceed the solubilizing capacity of bile (supersaturation), cholesterol can no longer remain dispersed and nucleates into solid cholesterol monohydrate crystals. Four conditions appear to contribute to formation of cholesterol gallstones: (1) supersaturation of bile with cholesterol; (2) hypomotility of the gallbladder; (3) accelerated cholesterol crystal nucleation; and (4) hypersecretion of mucus in the gallbladder, which traps the nucleated crystals, leading to accretion of more cholesterol and the appearance of macroscopic stones.

Pathogenesis of Pigment Stones

Pigment gallstones are complex mixtures of insoluble calcium salts of unconjugated bilirubin and inorganic calcium salts. Disorders that are associated with elevated levels of unconjugated bilirubin in bile increase the risk of developing pigment stones. These include chronic hemolytic anemia, severe ileal dysfunction or bypass, and bacterial contamination of the biliary tree, Unconjugated bilirubin is normally a minor component of bile, but it increases when infection of the biliary tract leads to release of microbial β-glucuronidases, which hydrolyze bilirubin glucuronides. Thus, infection of the biliary tract with *Escherichia coli, Ascaris lumbricoides,* or the liver fluke *C. sinensis* increases the

Table 18.11 Risk Factors for Gallstones

Cholesterol Stones
Demography: northern Europeans, North and South Americans, Native Americans, Mexican Americans
Advancing age
Female sex hormones
Female gender
Oral contraceptives
Pregnancy
Obesity and metabolic syndrome
Rapid weight reduction
Gallbladder stasis
Inborn disorders of bile acid metabolism
Hyperlipidemia syndromes
Pigment Stones
Demography: Asians more than Westerners, rural more than urban
Chronic hemolytic anemias
Biliary infection
Gastrointestinal disorders: ileal disease (e.g., Crohn disease), ileal resection or bypass, cystic fibrosis with pancreatic insufficiency

Figure 18.58 Cholesterol gallstones. The wall of the gallbladder is thickened and fibrotic due to chronic cholecystitis.

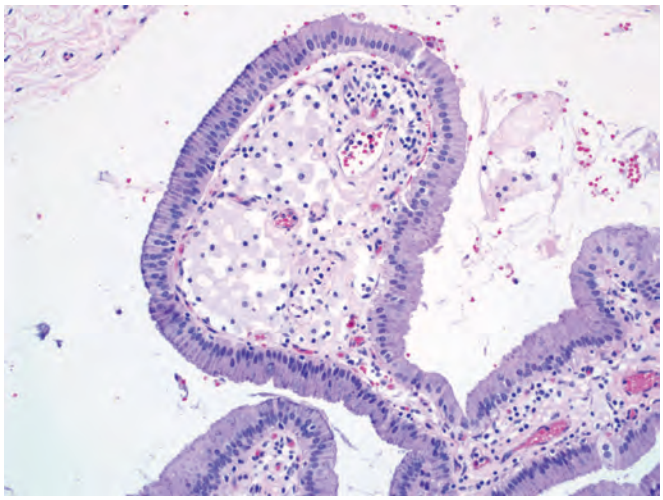


Figure 18.59 Cholesterolosis. Gallbladder mucosa demonstrates lamina propria distended by foamy macrophages.

likelihood of pigment stone formation. In hemolytic anemias, the secretion of conjugated bilirubin into bile increases. About 1% of bilirubin glucuronides are deconjugated in the biliary tree, and in the setting of chronically increased secretion of conjugated bilirubin, a sufficiently large amount of deconjugated bilirubin is generated to allow pigment stones to form.

MORPHOLOGY

Cholesterol stones arise exclusively in the gallbladder and range from 100% pure (which is rare) down to around 50% cholesterol. Pure cholesterol stones are pale yellow, round to ovoid, and have a finely granular, hard external surface (Fig. 18.58), which on transection reveals a glistening radiating crystalline palisade. With increasing proportions of calcium carbonate, phosphates, and bilirubin, the stones take on a gray-white to black color and may be lamellated. Multiple stones are usually present that range up to several centimeters in diameter. Rarely, a very large stone may virtually fill the fundus. Surfaces of stones may be rounded or faceted because of tight apposition to adjacent stones. Stones composed largely of cholesterol are radiolucent; sufficient calcium carbonate is found in 10% to 20% of cholesterol stones to render them radiopaque. The supersaturated cholesterol in bile that forms cholesterol stones may also diffuse into the mucosa and manifest as *cholesterolosis* (Fig. 18.59).

Pigment gallstones are brown to black. In general, black pigment stones are found in sterile gallbladder bile, and brown stones are found in infected large bile ducts. Black stones contain oxidized polymers of calcium salts of unconjugated bilirubin, small amounts of calcium carbonate, calcium phosphate, mucin glycoprotein, and some cholesterol monohydrate crystals. Brown stones contain similar compounds along with some cholesterol and calcium salts of palmitate and stearate. Black stones are rarely greater than 1.5 cm in diameter, are almost invariably present in great number (with an inverse relationship between size and number; Fig. 18.60), and are quite friable. Their contours are usually spiculated and molded. Brown stones tend to be laminated and soft and may have a soaplike or greasy consistency. Approximately 50% to 75% of black stones are radiopaque due to calcium salts, while brown stones, which contain calcium soaps, are radiolucent. Mucin glycoproteins constitute the scaffolding and interparticle cement of all types of stones.

Clinical Features

Gallstones may be present for decades before symptoms develop, and 70% to 80% of patients remain asymptomatic throughout their lives. Asymptomatic individuals convert to being symptomatic at an average rate of up to 4% per year, although the risk diminishes with time. Symptomatic patients with biliary "colic" experience excruciating pain, although this name is a misnomer, as the pain is usually constant, not colicky. Pain often follows a fatty meal that induces gallbladder contraction, which presses a stone against the gall bladder outlet, leading to increased pressure and eventually pain. The pain is localized to right upper quadrant

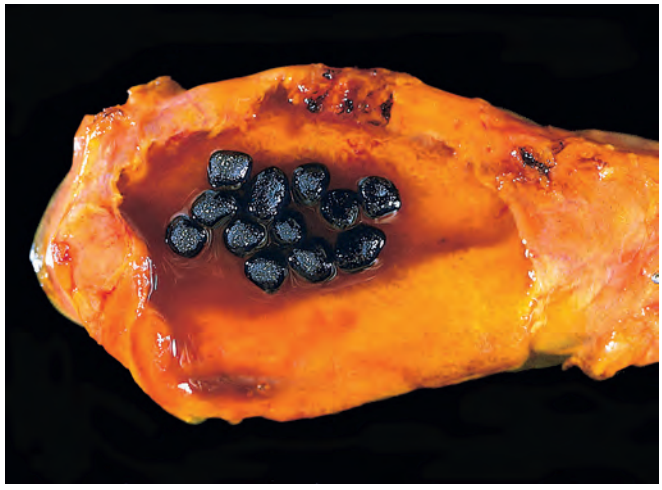


Figure 18.60 Pigment gallstones. Several faceted black gallstones are present in this otherwise unremarkable gallbladder from a patient with a mechanical mitral valve prosthesis, leading to chronic intravascular hemolysis.

or epigastrium and may radiate to the right shoulder or the back. Inflammation of the gallbladder (cholecystitis, discussed later) in association with stones also generates pain. More severe complications include empyema, perforation, fistula, inflammation of the biliary tree (cholangitis), obstructive cholestasis, and pancreatitis. The larger the calculi, the less likely they are to enter the cystic or common ducts to produce obstruction; it is very small stones, or "gravel," that are most dangerous. Occasionally a large stone may erode directly into an adjacent loop of small bowel, generating intestinal obstruction ("gallstone ileus" or Bouveret syndrome). Gallstones also are associated with an increased risk of gallbladder carcinoma (discussed later).

CHOLECYSTITIS

Inflammation of the gallbladder may be acute, chronic, or acute superimposed on chronic. It almost always occurs in association with gallstones. In the United States, cholecystitis is one of the most common indications for abdominal surgery. Its epidemiology closely parallels that of gallstones.

Acute Cholecystitis

Acute cholecystitis is precipitated in 90% of cases by obstruction of the neck of the gallbladder or the cystic duct by a stone. Acute calculous cholecystitis is the primary complication of gallstones and the most common reason for emergency cholecystectomy. Cholecystitis without gallstones *(acalculous cholecystitis)* may also occur in severely ill patients and accounts for the remaining 10% of cases.

Pathogenesis

Acute calculous cholecystitis results from chemical irritation and inflammation of a gallbladder obstructed by stones. The action of mucosal phospholipases hydrolyzes luminal lecithins to toxic lysolecithins. The normally protective glycoprotein mucus layer is disrupted, exposing the mucosal epithelium to the direct detergent action of bile salts. Prostaglandins released within the wall of the distended gallbladder contribute to mucosal and mural inflammation; distention and increased intraluminal pressure compromise blood flow to the mucosa. These events initially occur in the absence of bacterial infection, but later in the course bacterial infection may be superimposed and exacerbate the inflammatory process. Acute calculous cholecystitis is particularly common in diabetic patients who have symptomatic gallstones.

Acute acalculous cholecystitis is thought to result from ischemia. The cystic artery is an end artery without a collateral circulation. Contributing factors may include inflammation and edema of the wall (compromising blood flow) and gallbladder stasis due to accumulation of microcrystals of cholesterol (biliary sludge), viscous bile, and mucus, causing cystic duct obstruction in the absence of stones. It usually occurs in acutely ill patients who are hospitalized for unrelated conditions. Risk factors for acute acalculous cholecystitis include: (1) sepsis with hypotension and multisystem organ failure; (2) immunosuppression; (3) major trauma and burns; (4) diabetes mellitus; and (5) infections.

MORPHOLOGY

In **acute cholecystitis** the gallbladder is usually enlarged and tense, and it may assume a bright red or blotchy, violaceous to green-black discoloration, imparted by subserosal hemorrhages. The serosa is frequently covered by a fibrinous exudate that may be fibrinopurulent in severe cases. There are no specific morphologic differences between acute acalculous and calculous cholecystitis, save the absence of stones in the acalculous form. In **calculous cholecystitis,** an obstructing stone is usually present in the neck of the gallbladder or the cystic duct. The gallbladder lumen contains one or more stones and is filled with cloudy or turbid bile mixed with fibrin, pus, and hemorrhage. When the exudate is virtually pure pus, the condition is referred to as **gallbladder empyema.** In mild cases, the gallbladder wall is thickened, edematous, and hyperemic. In more severe cases, it is transformed into a green-black necrotic organ, termed **gangrenous cholecystitis,** with small-to-large perforations. The invasion of gas-forming organisms, notably clostridia and coliforms, may cause an **acute "emphysematous" cholecystitis.** Histologically, early changes of acute cholecystitis include edema, congestion, and mucosal erosion. Neutrophils are typically sparse, unless there is superimposed infection.

Clinical Features

Individuals with acute calculous cholecystitis usually (but not always) have experienced previous episodes of pain related to the gallbladder. An attack begins with progressive right upper quadrant or epigastric pain that lasts for more than 6 hours. It is frequently associated with mild fever, anorexia, tachycardia, sweating, nausea, and vomiting. Most patients are free of jaundice; the presence of hyperbilirubinemia suggests obstruction of the common bile duct. Mild to moderate leukocytosis may be accompanied by modestly elevated serum alkaline phosphatase. Acute calculous cholecystitis may appear with remarkable suddenness and constitute an acute surgical emergency, or it may present with mild symptoms that resolve without medical intervention. In the absence of medical attention, the attack usually subsides in 7 to 10 days and frequently within 24 hours. However, as many as 25% of patients develop progressively more severe symptoms and require immediate surgical intervention. Recurrence is common in patients who recover without surgery.

Clinical symptoms of acute acalculous cholecystitis tend to be more insidious, since they are obscured by the underlying conditions precipitating the attack. A higher proportion of patients have no symptoms referable to the gallbladder; diagnosis therefore rests on a high index of suspicion. In the severely ill patient, early recognition of the condition is crucial, since failure to do so almost ensures a fatal outcome. As a result of either delay in diagnosis or the disease itself, the incidence of gangrene and perforation is much higher in acalculous cholecystitis than in calculous cholecystitis. In rare instances, primary bacterial infection by agents such as *S. typhi* and staphylococci can give rise to acute acalculous cholecystitis. A more indolent form of acute acalculous cholecystitis may occur in the setting of systemic vasculitis, severe atherosclerotic ischemic disease in the elderly, AIDS (usually related to *Cryptosporidium* infection), or ascending biliary tract infection.

Chronic Cholecystitis

Chronic cholecystitis may be a sequel to repeated bouts of mild to severe acute cholecystitis, but in many instances it develops in the apparent absence of antecedent attacks. Since it is associated with cholelithiasis in more than 90% of cases, the at-risk patient population is the same as that for gallstones. The evolution of chronic cholecystitis is obscure; it is not clear that gallstones play a direct role in the initiation of inflammation or the development of pain, particularly since chronic acalculous cholecystitis shows symptoms and histology similar to those of the calculous form. Rather, supersaturation of bile predisposes to both chronic inflammation and, in most instances, stone formation. Microorganisms, usually *E. coli* and enterococci, are cultured from the bile in about one-third of cases. Unlike acute calculous cholecystitis, obstruction of gallbladder outflow is not a prerequisite.

MORPHOLOGY

The morphologic changes in chronic cholecystitis are extremely variable and sometimes minimal. The serosa is usually smooth and glistening but may be dulled by **subserosal fibrosis.** Dense fibrous adhesions may be present that represent the sequelae of prior acute inflammation. On sectioning, the wall is variably thickened and has an opaque gray-white appearance. In uncomplicated cases, the lumen contains green-yellow, mucoid bile and usually stones. The mucosa itself is generally preserved.

On histologic examination, the degree of inflammation is variable. In the mildest cases, only scattered lymphocytes, plasma cells, and macrophages are found in the mucosa and in the subserosal fibrous tissue (Fig. 18.61A). In more advanced cases, there is marked subepithelial and subserosal fibrosis, accompanied by mononuclear cell infiltration. Reactive proliferation of the mucosa and fusion of the mucosal folds may give rise to buried crypts of epithelium within the gallbladder wall. Outpouchings of the mucosal epithelium through the wall **(Rokitansky-Aschoff sinuses)** may be quite prominent (see Fig. 18.61B).

In rare instances, there is extensive dystrophic calcification of the gallbladder wall **(porcelain gallbladder).** More often, there is complete replacement of the gallbladder wall and mucosa by dense fibrosis **(hyalinizing cholecystitis),** with or without calcification. This change is notable for being associated with an increased incidence of gallbladder carcinoma (discussed later). In **xanthogranulomatous cholecystitis,** the gallbladder has a massively thickened wall and is shrunken, nodular, and chronically inflamed with foci of necrosis and hemorrhage. It is triggered by rupture of Rokitansky-Aschoff sinuses into the wall of the gallbladder followed by an accumulation of macrophages that have ingested biliary phospholipids. Such lipid-containing cells with foamy cytoplasm are called xanthoma cells, hence the name of this condition. Finally, an atrophic, chronically obstructed, often dilated gallbladder may contain only clear secretions, a condition known as **hydrops of the gallbladder.** Other rare forms of chronic cholecystitis include IgG_4-related sclerosing cholecystitis, another manifestation of IgG_4-related fibrosing disease.

Clinical Features

Chronic cholecystitis does not have the striking manifestations of the acute forms and is usually characterized by recurrent attacks of either steady epigastric or right upper quadrant pain. Nausea, vomiting, and intolerance for fatty foods are frequent accompaniments.

Diagnosis of both acute and chronic cholecystitis is important because of the following complications:

- *Bacterial superinfection* with cholangitis or sepsis
- *Gallbladder perforation* and local abscess formation
- *Gallbladder rupture* with diffuse peritonitis
- *Biliary enteric (cholecystenteric) fistula,* with drainage of bile into adjacent organs, entry of air and bacteria into the biliary tree, and potentially, gallstone-induced intestinal obstruction (ileus)
- *Aggravation of preexisting medical illness,* with cardiac, pulmonary, renal, or liver decompensation

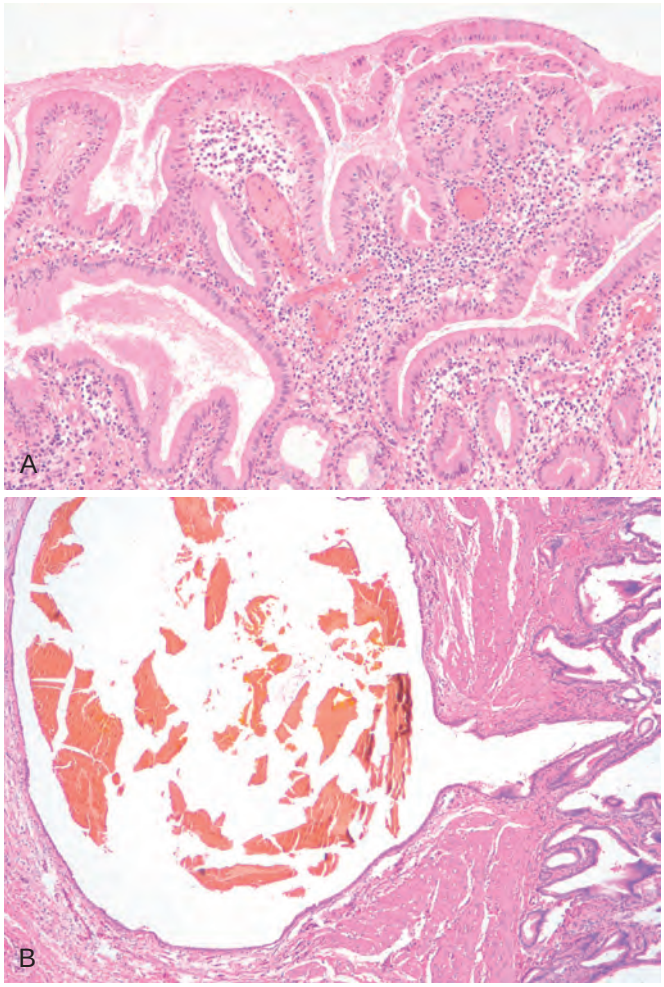


Figure 18.61 Chronic cholecystitis. (A) The gallbladder mucosa is infiltrated by inflammatory cells. (B) Outpouching of the mucosa through the wall forms Rokitansky-Aschoff sinus (contains bile).

GALLBLADDER CARCINOMA

Carcinoma of the gallbladder is the most common malignancy of the extrahepatic biliary tract. Approximately 6000 new cases of gallbladder cancer are diagnosed each year in the United States. There are wide variations in the incidence of gallbladder cancer worldwide, with regions

such as Chile, Bolivia, and Northern India harboring the highest numbers of cases. Even within the United States, areas with large Native American or Hispanic populations, such as the southwest, have a higher incidence of gallbladder cancer than the rest of the country. Gallbladder cancer is at least twice as common in women as in men; this gender disparity can be several-fold greater in regions of highest incidence. The overwhelming majority of patients are diagnosed at an advanced, surgically unresectable, stage, and the mean 5-year survival rate for affected patients is less than 10%.

Pathogenesis

The most important risk factor for gallbladder cancer (besides gender and ethnicity) is gallstones, which are present in 95% of cases. However, it should be noted that only 1% to 2% of patients with gallstones develop gallbladder cancer. In Asia, chronic bacterial or parasitic infections have been implicated as risk factors, and the coexistence of gallstones with gallbladder cancer is much lower. Nonetheless, the common thread that ties gallstones or chronic infections together with gallbladder cancer is chronic inflammation. Relatively common driver mutations include gain-of-function aberrations affecting members of the EGF receptor gene family (including *HER2*) and genes enocding downstream signaling components such as RAS, and loss-of-function mutations in the *TP53* tumor suppressor gene, which are present in up to one-half of tumors. As is typical of *TP53*-mutated cancers, most gallbladder carcinomas exhibit aneuploidy.

MORPHOLOGY

Several precursor lesions of gallbladder carcinoma have been described, indicating that (as is typical of carcinoma generally) full-blown cancer likely arises from a lengthy stepwise process. These precursors include flat in situ lesions with varying degrees of dysplasia, mass-forming adenoma-like lesions termed intracholecystic papillary tubular neoplasm, and intestinal metaplasia. Gallbladder cancers are mainly adenocarcinomas and are most often detected in the fundus (Fig. 18.62). They may produce a firm, poorly circumscribed mass or may diffusely infiltrate the wall of gallbladder, simulating the gross appearance of chronic cholecystitis. Microscopically, they are usually characterized by the presence of glands embedded in desmoplastic stroma, but in some cases cytologic atypia and stromal response are minimal; in these instances, identification of perineural and vascular invasion help establish the diagnosis. With progression, direct extension into other organs, fistula formation, peritoneal and biliary spread, and metastasis to the liver and portahepatic lymph nodes occur. Rarely, other types of malignant tumors arise primarily within the gallbladder, including neuroendocrine tumors, squamous cell carcinoma, and sarcomas.

Clinical Features

Preoperative diagnosis of carcinoma of the gallbladder is the exception rather than the rule, occurring in fewer than 20% of patients. Presenting symptoms are insidious and typically indistinguishable from those associated with cholelithiasis, including abdominal pain, jaundice, anorexia, and nausea and vomiting. If the tumor is detected prior to invasion and spread, such as when carcinoma is an incidental finding during surgery for symptomatic gallstones or acute cholecystitis, cure by surgical resection is possible. However, most patients have advanced disease at diagnosis, and the overall prognosis for such patients is poor, even with apparently complete resection of evident tumor. Adjuvant chemotherapy is often offered to such patients but is not curative, and most patients eventually succumb to their disease.

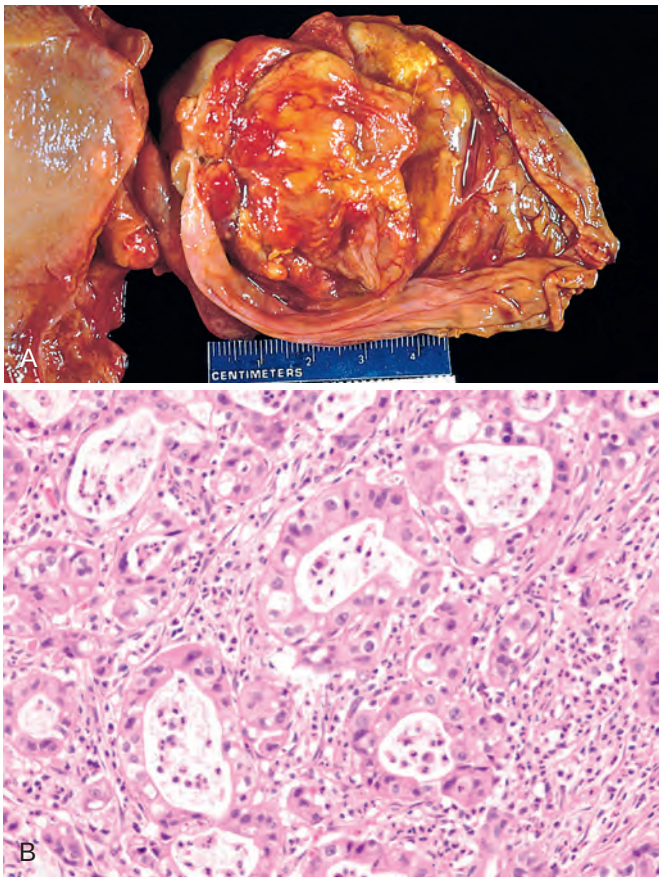


Figure 18.62 Gallbladder adenocarcinoma. (A) The opened gallbladder contains a large, exophytic tumor that virtually fills the lumen. (B) Malignant glands are seen infiltrating a densely fibrotic gallbladder wall.

KEY CONCEPTS

DISEASES OF THE GALLBLADDER

- Gallbladder diseases include cholelithiasis, acute and chronic cholecystitis, and gallbladder cancer.
- Gallstones are common in Western countries. The great majority are cholesterol stones. Pigmented stones containing bilirubin and calcium are most common in Asian countries.
- Risk factors for the development of cholesterol stones are advancing age, female gender, estrogen use, obesity, and heredity.
- Cholecystitis almost always occurs in association with cholelithiasis, although in about 10% of cases it occurs in the absence of gallstones. Gallstones are also a risk factor for gallbladder cancer.

- Acute calculous cholecystitis is the most common reason for emergency cholecystectomy.
- Gallbladder cancers are associated with gallstones in the vast majority of cases. Typically, they are detected late because of nonspecific symptoms and hence carry a poor prognosis.

SUGGESTED READINGS

Mechanisms of Liver Injury and Repair

Gouw ASW, Clouston AD, Theise ND: Ductular reactions in human livers: diversity at the interface, *Hepatology* 54:1853, 2011. *[A review of ductular reactions, the stem cell response of human livers in all liver diseases, that are related to mechanisms of regeneration, fibrogenesis, and neoplasia].*

Kocabayoglu P, Friedman SL: Cellular basis of hepatic fibrosis and its role in inflammation and cancer, *Front Biosci (Schol Ed)* 5:217, 2013. *[Interweaving what is known about hepatic stellate cells and other myofibroblastic cells of the liver with inflammatory, fibrosing, and neoplastic disease processes].*

Koyama Y, Brenner DA: Liver inflammation and fibrosis, *J Clin Invest* 127(1):55–64, 2017. doi:10.1172/JCI88881. [*Epub 2017 Jan 3*]. PubMed PMID: 28045404. PubMed Central PMCID: PMC5199698, Review, *[A recent review of the complex interplay between the immune system and hepatic fibrosis].*

Quaglia A, Alves VA, Balabaud C et al: International Liver Pathology Study Group. Role of aetiology in the progression, regression, and parenchymal remodelling of liver disease: implications for liver biopsy interpretation, *Histopathology* 68(7):953–967, 2016. doi:10.1111/his.12957. [*Epub 2016 Apr 28*]. PubMed PMID: 26918878. Review, *[This paper challenges pathologists to evolve past use of the term* cirrhosis *to communicate more useful information about fibrosis in relation to disease etiology].*

Theise ND, Saxena R, Portmann BC et al: The canals of Hering and hepatic stem cells in humans, *Hepatology* 30(6):1425–1433, 1999. PubMed PMID: 10573521. *[The meticulous histopathologic identification of a stem cell niche in liver].*

Wanless IR, Nakashima E, Sherman M: Regression of human cirrhosis. Morphologic features and the genesis of incomplete septal cirrhosis, *Arch Pathol Lab Med* 124:1599, 2000. *[A revolutionary paper that began the reassessment of whether fibrosis/cirrhosis could regress; take a look also, in the same volume of the journal, at the critiques by the original reviewers of the paper (published despite their objections) and the authors' response to these. Pathology history happening before your eyes].*

Acute, Chronic, Acute-On-Chronic Liver Failure

Bernal W, Wendon J: Acute liver failure, *N Engl J Med* 369:2525, 2013. *[An excellent clinical review].*

Berzigotti A, Seijo S, Reverter E et al: Assessing portal hypertension in liver diseases, *Expert Rev Gastroenterol Hepatol* 7:141, 2013. *[Evolving definitions and methodologies for evaluating portal hypertension are at the leading edge of advances in hepatology].*

Chun LJ, Tong MJ, Busuttil RW et al: Acetaminophen hepatotoxicity and acute liver failure, *J Clin Gastroenterol* 43:342, 2009. *[About the most common cause of acute liver failure leading to transplantation].*

Hytiroglou P, Snover DC, Alves V: Beyond "cirrhosis": a proposal from the International Liver Pathology Study Group, *Am J Clin Pathol* 137:5, 2012. *[New understandings of diverse cirrhoses and a new look at the old concept of cirrhosis and advanced-stage liver disease].*

Khungar V, Poordad F: Hepatic encephalopathy, *Clin Liver Dis* 16:301, 2012. *[A good overview of the one of the most ominous sequelae of all forms of liver failure].*

Laleman W, Verbeke L, Meersseman P: Acute-on-chronic liver failure: current concepts on definition, pathogenesis, clinical manifestations and potential therapeutic interventions, *Expert Rev Gastroenterol Hepatol* 5:523, 2011.

Lee WM, Squires RH Jr, Nyberg SL et al: Acute liver failure: summary of a workshop, *Hepatology* 47(4):1401–1415, 2008. doi:10.1002/hep.22177. PubMed PMID: 18318440. PubMed Central PMCID: PMC3381946, *[A discussion of findings from the Acute Liver Failure Study Group].*

Lefkowitch JH: The pathology of acute liver failure, *Adv Anat Pathol* 23(3):144–158, 2016. doi:10.1097/PAP.0000000000000112. PubMed PMID: 27058243. Review, *[An excellent review of the pathology of acute liver failure].*

van Leeuwen DJ, Alves V, Balabaud C et al: Acute-on-chronic liver failure 2018: a need for (urgent) liver biopsy? *Expert Rev Gastroenterol Hepatol* 12(6):565–573, 2018. *[Recent review of terminology and clinicopathologic considerations around acute on chronic liver failure].*

Viral Hepatitis

Chung RT, Baumert TF: Curing chronic hepatitis C—The arc of medical triumph, *N Engl J Med* 370:1576, 2014. *[An excellent perspective on why and how hepatitis C can be cured].*

European Association for the Study of the Liver: EASL recommendations on treatment of hepatitis C 2018, *J Hepatol* 69(2):461–511, 2018. *[Review of diagnosis, screening and treatment guidelines for hepatitis C].*

Joyce MA, Tyrrell DL: The cell biology of hepatitis C virus, *Microbes Infect* 12:263, 2010. *[What we understand and do not understand about hepatitis C virus and the cells it infects].*

Lamontagne RJ, Bagga S, Bouchard MJ: Hepatitis B virus molecular biology and pathogenesis, *Hepatoma Res* 2:163–186, 2016. *[Review of hepatitis B virus infection].*

Lenggenhager D, Weber A: Clinicopathologic features and pathologic diagnosis of hepatitis E, *Hum Pathol* 2019. doi:10.1016/j.humpath.2019.10.003. pii: S0046-8177(19)30179-0. [Epub ahead of print], *[Review of hepatitis E virus infection].*

Autoimmune Liver Diseases

Czaja AJ: Examining pathogenic concepts of autoimmune hepatitis for cues to future investigations and interventions, *World J Gastroenterol* 25(45):6579–6606, 2019. *[Review of pathogenesis of AIH].*

Eisenmann de Torres B, and members of the International Autoimmune Hepatitis Group: Simplified criteria for the diagnosis of autoimmune hepatitis, *Hepatology* 48:169, 2008. *[A still relevant set of criteria for diagnosing autoimmune hepatitis].*

European Association for the Study of the Liver: EASL Clinical Practice Guidelines: the diagnosis and management of patients with primary biliary cholangitis, *J Hepatol* 67(1):145–172, 2017. *[Review of diagnosis and treatment guidelines for PBC].*

Karlsen TH, Folseraas T, Thorburn D, et al: Primary sclerosing cholangitis – a comprehensive review, *J Hepatol* 67(6):1298–1323, 2017. *[Review of all aspects of PSC].*

Drug- and Toxin-Induced Liver Injury

Crawford JM: Histologic findings in alcoholic liver disease, *Clin Liver Dis* 16:699, 2012. *[A thorough look at the morphologic changes in alcoholic liver disease and the underlying mechanisms that produce them].*

Gunawan B, Kaplowitz N: Clinical perspectives on xenobiotic-induced hepatotoxicity, *Drug Metab Rev* 36:301, 2004. *[A comprehensive and current review].*

Kleiner DE: Drug-induced liver injury: the hepatic pathologist's approach, *Gastroenterol Clin North Am* 46(2):273–296, 2017. *[Review of the manner in which morphologic changes are evaluated in the setting of suspected drug-induced liver injury].*

Fatty Liver Disease

Alpert L, Hart J: The pathology of alcoholic liver disease, *Clin Liver Dis* 20(3):473–489, 2016. *[Recent review of the pathology of alcoholic liver disease].*

Kleiner DE, Brunt EM: Nonalcoholic fatty liver disease: pathologic patterns and biopsy evaluation in clinical research, *Semin Liver Dis* 32:3, 2012. *[As authoritative as one can be on the topic].*

Inherited Liver Diseases

Perlmutter DH, Silverman GA: Hepatic fibrosis and carcinogenesis in α1-antitrypsin deficiency: a prototype for chronic tissue damage in gain-of-function disorders, *Cold Spring Harb Perspect Biol* 1:3, 2011.

[A broader perspective on some chronic hereditary diseases using α_1-antitrypsin as the paradigmatic example].

Pietrangelo A: Genetics, genetic testing, and management of hemochromatosis: 15 years since hepcidin, *Gastroenterology* 149(5):1240–1251, 2015. *[Comprehensive review of genetics of hemochromatosis with emphasis on hepcidin in mechanism of disease].*

Rosencrantz R, Schilsky M: Wilson disease: pathogenesis and clinical considerations in diagnosis and treatment, *Semin Liver Dis* 31:245, 2011. *[Review of mechanisms of copper-induced liver injury in Wilson disease with a useful summary of diagnosis and treatment considerations].*

Cholestatic Syndromes

Desmet VJ: Congenital diseases of intrahepatic bile ducts: variations on the theme "ductal plate malformation", *Hepatology* 16:1069, 1992. *[A classic paper in liver pathology].*

Hirschfield GM, Heathcote EJ, Gershwin ME: Pathogenesis of cholestatic liver disease and therapeutic approaches, *Gastroenterology* 139:1481, 2010. *[A thorough overview of the most common features of liver disease].*

Paumgartner G: Biliary physiology and disease: reflections of a physician-scientist, *Hepatology* 51:1095, 2010. *[How bench-top work comes to exert an impact on clinical medicine, sometimes slowly over decades].*

Tsui WM, Lam PW, Lee WK: Primary hepatolithiasis, recurrent pyogenic cholangitis, and oriental cholangiohepatitis: a tale of 3 countries, *Adv Anat Pathol* 18:318, 2011. *[A review that reveals much about the diseases discussed as well as the variable histories of medical understanding in different regions of the globe].*

Circulatory Disorders

Guido M, Alves VAF, Balabaud C et al: Histology of portal vascular changes associated with idiopathic non-cirrhotic portal hypertension: nomenclature and definition, *Histopathology* 74(2):219–226, 2019. *[Provides a review of the spectrum of histologic changes and a unifying nomenclature].*

Benign and Malignant Liver Tumors

Choi WT, Kakar S: Atypical hepatocellular neoplasms: review of clinical, morphologic, immunohistochemical, molecular, and cytogenetic features. *Adv Anat Pathol* 25(4):254–262, 2018. *[Review of pathologic and molecular aspects of adenomas and atypical tumors with borderline HCC features].*

International Consensus Group for Hepatocellular Neoplasia: Pathologic diagnosis of early hepatocellular carcinoma, *Hepatology* 49:658, 2009. *[A good example of how change comes to medicine, individual efforts combining, over years, to achieve a new consensus].*

Massarweh NN, El-Serag HB: Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma, *Cancer Control* 24(3): 1073274817729245, 2017. *[Review of epidemiology ad trends in hepatocellular carcinoma and cholangiocarcinoma].*

Rizvi S, Khan SA, Hallemeier CL et al: Cholangiocarcinoma—evolving concepts and therapeutic strategies, *Nat Rev Clin Oncol* 15(2):95–111, 2018. *[Review of epidemiology, pathogenesis, and management of cholangiocarcinoma].*

Sempoux C, Chang C, Gouw A et al: Benign hepatocellular nodules: what have we learned using the patho-molecular classification, *Clin Res Hepatol Gastroenterol* 37(4):322–327, 2013. *[From the team largely responsible for the new molecular subclassifications of hepatocellular adenomas].*

Vogel A, Cervantes A, Chau I et al: ESMO Guidelines Committee: hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, *Ann Oncol* 29(Suppl 4):iv238–iv255, 2018. *[Review of guidelines for diagnosis and management of HCC].*

WHO Classification of Tumours Editorial Board: *Digestive System Tumors*, Lyon, France, 2019, International Agency for Research on Cancer. *[WHO classification of tumours series, ed 5, vol 1].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 19

The Pancreas

Anirban Maitra[a]

CHAPTER CONTENTS

The adult pancreas is a transversely oriented retroperitoneal organ extending from the C-loop of the duodenum to the hilum of the spleen (Fig. 19.1A). Although the organ gets its name from the Greek *pankreas* ("all flesh"), it is in fact a complex lobulated organ with distinct exocrine and endocrine components.

The *exocrine pancreas* constitutes 80% to 85% of the organ and is composed of acinar cells. These pyramidally shaped epithelial cells contain membrane-bound granules rich in proenzymes (zymogens), including trypsinogen, chymotrypsinogen, procarboxypeptidase, proelastase, kallikreinogen, and prophospholipase A and B, all of which contribute to digestion. Upon secretion, these proenzymes and enzymes are carried by a series of ductules and ducts to the duodenum, where they are activated by proteolytic cleavage (described later).

The *endocrine pancreas* is composed of about 1 million endocrine cell clusters, the islets of Langerhans, that are scattered throughout the gland. Islet cells secrete insulin, glucagon, somatostatin, and pancreatic polypeptide. Although they constitute only 1% to 2% of the organ mass, the hormones released by the islet cells are essential regulators of systemic metabolism. Diseases of the endocrine pancreas are described in detail in Chapter 24.

CONGENITAL ANOMALIES

The complex process by which the dorsal and ventral pancreatic primordia fuse during pancreatic development is imperfect, giving rise to variation in pancreatic anatomy. The pancreas normally arises from the fusion of dorsal and ventral outpouchings of the foregut. The body, the tail, and the superior/anterior aspect of the head of the pancreas, as well as the accessory duct of Santorini, are derived from the dorsal primordium. Typically, the ventral primordium gives rise to the posterior/inferior part of the head of the pancreas and drains through the main pancreatic duct into the papilla of Vater.

Pancreas Divisum. **Pancreas divisum is the most common congenital anomaly of the pancreas** with an incidence of 3% to 10%. In most individuals, the main pancreatic duct (the duct of Wirsung) joins the common bile duct just proximal to the papilla of Vater, and the accessory pancreatic duct (the duct of Santorini) drains into the duodenum through a separate minor papilla (see Fig. 19.1A). Pancreas divisum is caused by a failure of fusion of the fetal duct systems of the dorsal and ventral pancreatic primordia. As a result, in this anatomic variant the bulk of the pancreas (formed by the dorsal pancreatic primordium) drains into the duodenum through the small-caliber minor papilla (Fig. 19.1B), while the duct of Wirsung drains only a small portion of the head of the gland through the papilla of Vater. Although controversial, it has been suggested that inadequate drainage of the pancreatic secretions through the minor papilla, especially when combined with genetic defects that also increase susceptibility to pancreatitis (described later), predisposes individuals with pancreatic divisum to chronic pancreatitis.

Annular Pancreas. Annular pancreas is a band-like ring of normal pancreatic tissue that completely encircles the second portion of the duodenum. Annular pancreas can produce duodenal obstruction.

Ectopic Pancreas. Pancreatic tissue that is aberrantly situated, or ectopic, is found in about 2% of careful routine postmortem examinations. The favored sites for ectopia are

[a]I am grateful for the immense contributions of Ralph H. Hruban and Christine A. Iacobuzio-Donahue, who authored this chapter in the prior edition of this book. Many of the photomicrographs used in this chapter are contributed by Dr. Hruban.

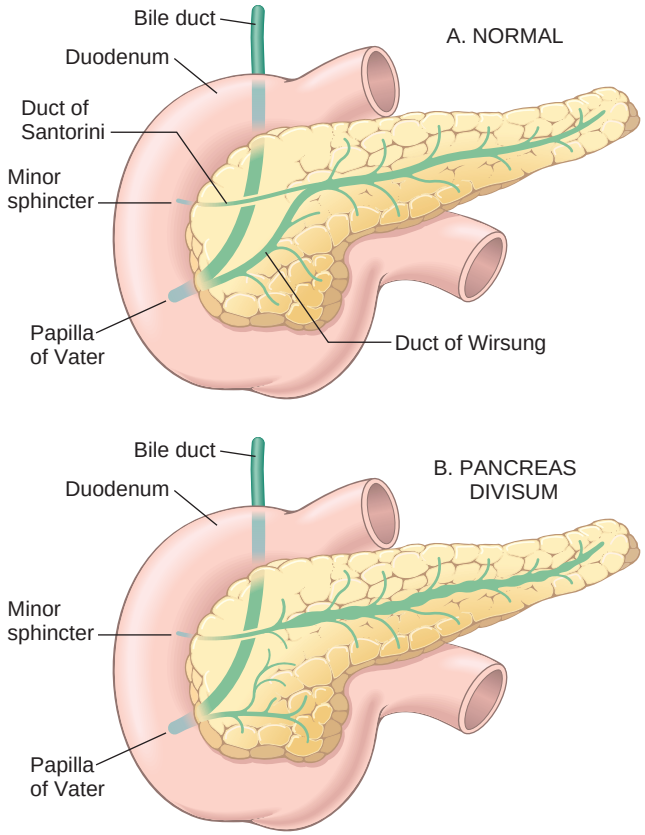


Figure 19.1 Pancreatic ductal anatomy. (A) Normal ductal anatomy. (B) Ductal anatomy in pancreatic divisum. (Modified from Gregg JA, et al: Pancreas divisum: results of surgical intervention, *Am J Surg* 145:488–492, 1983.)

the stomach and duodenum, followed by the jejunum, Meckel diverticula, and ileum. Although usually incidental findings, these embryologic rests, composed of normal-appearing pancreatic acini, glands, and sometimes islets of Langerhans, may cause pain from localized inflammation or may incite mucosal bleeding.

Agenesis. Very rarely the pancreas fails to develop (agenesis). Some cases of agenesis are caused by homozygous germline mutations involving *PDX1,* a gene encoding a homeobox transcription factor that is critical for pancreatic development.

PANCREATITIS

Pancreatitis is divided into two forms, acute and chronic, each with its own characteristic pathologic and clinical features; both are initiated by injuries that lead to autodigestion of the pancreas by its own enzymes. Under normal circumstances, several factors protect the pancreas from autodigestion.

- *Most digestive enzymes are synthesized as inactive proenzymes* (zymogens) and packaged within secretory granules.
- *Proenzymes are typically activated by trypsin,* which itself is activated by duodenal enteropeptidase (enterokinase) in the small bowel; as a result, intrapancreatic activation of proenzymes is normally minimal.
- *Acinar and ductal cells secrete trypsin inhibitors,* including serine protease inhibitor Kazal type 1 (SPINK1), which further limit intrapancreatic trypsin activity.

Pancreatitis occurs when these protective mechanisms are disrupted or overwhelmed. Acute attacks may range from mild and self-limited to life-threatening events. Recurrent or persistent pancreatitis may lead to permanent loss of pancreatic function.

Acute Pancreatitis

Acute pancreatitis is characterized by reversible pancreatic parenchymal injury and inflammation and has many causes, including toxic exposures (e.g., alcohol), pancreatic duct obstruction (e.g., biliary calculi), inherited genetic defects, vascular injury, and infections. Acute pancreatitis is relatively common with an annual incidence in Western countries of 10 to 20 cases per 100,000 people. Biliary tract disease and alcoholism account for approximately 80% of these cases (Table 19.1). The proportion of cases caused by excessive alcohol intake varies from 65% in the United States to 20% in Sweden to 5% or less in southern France and the United Kingdom. The male-to-female ratio is 6:1 in those with alcoholism and 1:3 in patients with biliary tract disease. Gallstones are present in 35% to 60% of acute pancreatitis cases and cause "gallstone pancreatitis" in about 5% of patients with gallstones.

Pathogenesis

Acute pancreatitis results from inappropriate release and activation of pancreatic enzymes that, in turn, destroy

Table 19.1 Etiologic Factors in Acute Pancreatitis

Metabolic
Alcoholism[a]
Hyperlipoproteinemia
Hypercalcemia
Drugs (e.g., azathioprine, statins, GLP-1 agonists, DPP-4 inhibitors)
Genetic
Mutations in genes encoding trypsin *(PRSS1)*, trypsin regulators *(SPINK1)*, or proteins that regulate calcium metabolism *(CASR)*
Cystic fibrosis *(CFTR)*
Mechanical
Gallstones[a]
Trauma
Iatrogenic injury
Operative injury
Endoscopic procedures with dye injection (e.g., ERCP)
Vascular
Shock
Atheroembolism
Vasculitis (e.g., polyarteritis nodosa)
Infectious
Mumps
Coxsackievirus

[a]Most common etiologies in the United States.

DPP-4, Dipeptidyl peptidase-4; *ERCP,* endoscopic retrograde cholangiopancreatography; *GLP-1,* glucagon-like peptide-1.

pancreatic tissue and elicit an acute inflammatory reaction. Inappropriate trypsin activation within the pancreas can activate other proenzymes such as prophospholipase and proelastase, which degrade fat cells and damage the elastic fibers of blood vessels, respectively. Once tissue damage commences, trypsin can also directly or indirectly activate factors found in the blood, including components of the coagulation, complement, kallikrein, and fibrinolytic pathways (Chapters 3 and 4). The resulting inflammation and small-vessel thrombosis causes further damage to acinar cells, amplifying intrapancreatic enzyme activation.

The triggers that cause intrapancreatic enzyme activation in sporadic acute pancreatitis remain a topic of investigation, but there is evidence for at least three major initiating events (Fig. 19.2):

- *Pancreatic duct obstruction* is most commonly caused by gallstones and biliary sludge but can also stem from periampullary neoplasms (e.g., pancreatic cancer), choledochoceles (congenital cystic dilation of the common bile duct), parasites (particularly *Ascaris lumbricoides* and *Clonorchis sinensis*), and possibly pancreas divisum. Obstruction raises intrapancreatic ductal pressure and leads to accumulation of enzyme-rich fluid within the interstitium. Unlike other pancreatic enzymes, lipase is secreted in an active form and has the potential to cause local fat necrosis. The death of adipocytes is hypothesized to produce "danger" signals locally that are sensed by pancreatic stellate cells and leukocytes, which release proinflammatory mediators that promote microvascular leak and development of interstitial edema. Edema may further compromise local blood flow, causing vascular insufficiency and ischemic injury to acinar cells.
- *Primary acinar cell injury* leads to release of digestive enzymes, inflammation, and tissue autodigestion. Acinar cells can be damaged by a variety of endogenous, exogenous, and iatrogenic insults. Many of these result in oxidative stress and generation of intracellular free radicals that lead to membrane lipid oxidation and transcription factor activation. The latter include AP1 and NF-κB, which induce expression of chemokines that attract mononuclear cells. Increased calcium flux is another important trigger for inappropriate enzyme activation. Calcium has a key role in trypsin regulation. When calcium levels are low, trypsin tends to cleave and inactivate itself, but this autoinhibition is abrogated and trypsin autoactivation is favored when calcium levels are increased. Any factor causing increased calcium within acinar cells may therefore trigger excessive trypsin activation. Examples include inherited abnormalities that affect calcium levels (Table 19.2).

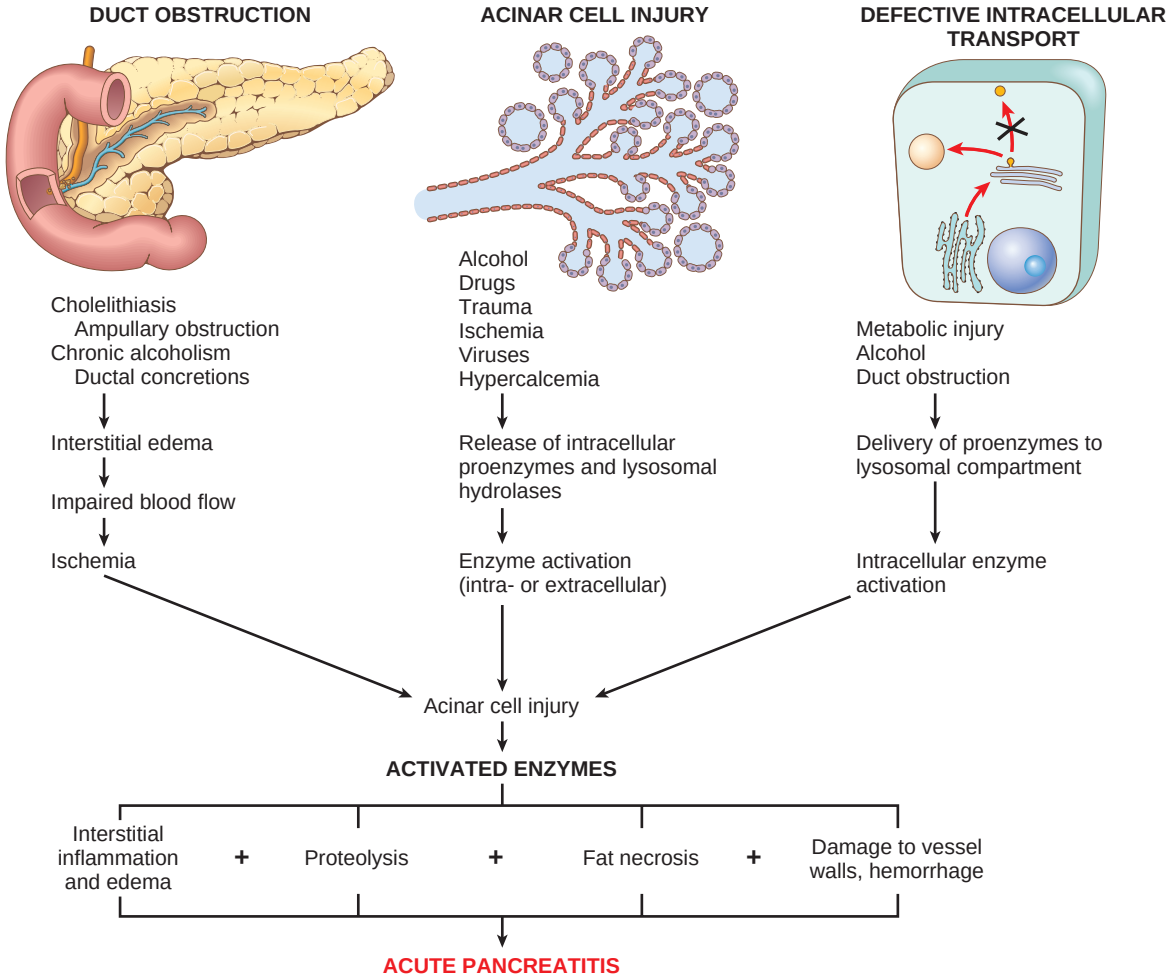


Figure 19.2 Three proposed pathways in the pathogenesis of acute pancreatitis.

Table 19.2 Inherited Predisposition to Pancreatitis

Gene (Chromosome Location)	Protein Product	Function
CFTR (7q31)	Cystic fibrosis transmembrane conductance regulator	Epithelial anion channel. Loss-of-function mutations alter fluid pressure and limit bicarbonate secretion, leading to inspissation of secreted fluids and duct obstruction
PRSS1 (7q34)	Serine protease 1 (trypsinogen 1)	Cationic trypsin. Gain-of-function mutations prevent self-inactivation of trypsin
SPINK1 (5q32)	Serine peptidase inhibitor, Kazal type 1	Inhibitor of trypsin. Mutations cause loss-of-function, increasing trypsin activity
CASR (3q13)	Calcium-sensing receptor	Membrane-bound receptor that senses extracellular calcium levels and controls luminal calcium levels. Mutations may alter calcium concentrations and activate trypsin
CTRC (1p36)	Chymotrypsin C (caldecrin)	Degrades trypsin, protects the pancreas from trypsin-related injury
CPA1 (7q32)	Carboxypeptidase A1	Exopeptidase involved in regulating zymogen activation

- *Defective intracellular transport of proenzymes within acinar cells.* In normal acinar cells, digestive enzymes and lysosomal hydrolases are transported via separate pathways. In animal models of acinar injury, pancreatic proenzymes, of which the *cathepsin* family of hydrolases is most important, are aberrantly delivered to lysosomes. The hydrolases then activate pancreatic proenzymes that disrupt lysosomal membranes, ultimately leading to release of activated enzymes. It is not clear if this mechanism is relevant to human acute pancreatitis.

Two factors have been proposed to contribute to alcohol-indcuced pancreatitis: duct obstruction and acinar cell damage. Alcohol consumption transiently increases contraction of the sphincter of Oddi (the muscle at the papilla of Vater), and chronic alcohol ingestion results in the secretion of protein-rich pancreatic fluid that is prone to form inspissated protein plugs that obstruct small pancreatic ducts. Within acinar cells, alcohol-induced oxidative stress may generate free radicals, leading to membrane lipid oxidation and activation of the proinflammatory transcription factors AP1 and NF-κB. Oxidative stress also may promote fusion of lysosomes and zymogen granules and increase intracellular calcium levels, possibly through mitochondrial damage, thereby promoting the intracellular activation of trypsin and other digestive enzymes. Nevertheless, it should be noted that most alcoholics never develop pancreatitis and those who do usually do so after many years of alcohol abuse. Thus, key aspects of the pathophysiology of alcohol-induced pancreatitis remain obscure.

Other proven or suspected triggers of acute pancreatitis in sporadic cases include (see Table 19.1):

- *Metabolic disorders* such as hypertriglyceridemia and hypercalcemic states, as in hyperparathyroidism.
- *Genetic lesions,* as described below.
- *Medications.* Hundreds of drugs have been anecdotally linked to acute pancreatitis, but only about 50 have a definite association. The latter include azathioprine, angiotensin-converting enzyme (ACE) inhibitors, statins, and antidiabetic drugs such as glucagon-like peptide-1 (GLP-1) agonists and dipeptidyl peptidase-4 (DPP-4) inhibitors.
- *Traumatic acinar cell injury,* by either blunt abdominal trauma or iatrogenic injury during surgery or endoscopic retrograde cholangiopancreatography.
- *Ischemic acinar cell injury* due to shock, vascular thrombosis, embolism, or vasculitis.
- *Infections,* including mumps, can lead to acute pancreatitis through direct acinar cell injury,

Hereditary factors are increasingly recognized as causes of pancreatitis. Affected individuals have recurrent attacks of severe acute pancreatitis, which often begin in childhood, and can ultimately lead to chronic pancreatitis. **The feature shared by most forms of hereditary pancreatitis is a defect that increases or sustains the activity of trypsin** (see Table 19.2). Three genes implicated in hereditary pancreatitis deserve special note: *PRSS1, SPINK1,* and *CFTR*. Most hereditary cases are due to gain-of-function mutations in the trypsinogen gene (known as *PRSS1*). Some of these *PRSS1* gene mutations make trypsin resistant to self-inactivation, abrogating an important negative feedback mechanism; other mutations make trypsinogen more prone to proteolytic activation. Hereditary pancreatitis associated with trypsinogen mutation has an autosomal dominant mode of inheritance, as is typically true of disorders associated with gain-of-function mutations.

Hereditary pancreatitis can also be caused by loss-of-function mutations in *SPINK1,* a gene encoding a trypsin inhibitor. Because one functional copy of *SPINK1* produces sufficient inhibitor to maintain adequate trypsin control, this form of hereditary pancreatitis has an autosomal recessive mode of inheritance.

As discussed in detail in Chapter 10, cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane conductance regulator *(CFTR)* gene that encodes an apical chloride channel. Loss of *CFTR* leads to abnormal secretions that promote protein plugging, duct obstruction, and the development of pancreatitis. Disease can occur in those with homozygous and even heterozygous *CFTR* gene mutations, particularly in patients who also have *SPINK1* mutations.

Of note, patients with *PRSS1*-associated hereditary pancreatitis have a 40% lifetime risk of developing pancreatic cancer, yet another example of the nefarious association between neoplasia and chronic tissue injury and inflammation.

MORPHOLOGY

The morphology of acute pancreatitis ranges from limited inflammation and edema to extensive necrosis and hemorrhage. The basic alterations are **(1) microvascular leak and edema, (2) fat necrosis, (3) acute inflammation, (4) damage, including autodigestion, of pancreatic parenchyma, and (5) blood vessel destruction with interstitial hemorrhage.** The extent of each of these depends on the duration and severity of disease.

In more limited forms of **acute interstitial pancreatitis,** histologic alterations are limited to mild inflammation, interstitial edema, and focal fat necrosis within the pancreas and the peripancreatic fat (Fig. 19.3). Fat necrosis, triggered by lipase activity, leads to **saponification,** a process in which fatty acids combine with calcium to form insoluble calcium soaps that impart a granular blue microscopic appearance to surviving fat cells (Chapter 2).

In **acute necrotizing pancreatitis,** acini, ducts, and even islets undergo necrosis. In the most severe form, **hemorrhagic pancreatitis,** extensive parenchymal necrosis is accompanied by intraparenchymal hemorrhage due to vascular injury. This imparts a red-black color with interspersed foci of yellow-white, chalky fat necrosis (Fig. 19.4). Fat necrosis may also occur in the omentum and the mesentery of the bowel adjacent to the pancreas and beyond the abdominal cavity, such as in the subcutaneous fat, as a result of systemic lipase release. The peritoneal cavity typically contains a serous, slightly turbid, brown-tinged fluid with fat globules that reflect digestion of adipose tissue.

Clinical Features

Abdominal pain is the cardinal manifestation of acute pancreatitis. Characteristically the pain is constant, intense, and referred to the upper or mid back and, occasionally, the left shoulder. The severity ranges from mild discomfort to incapacitating pain. Anorexia, nausea, and vomiting are common. Elevated plasma levels of amylase and lipase support the diagnosis of acute pancreatitis, as does the exclusion of other causes of abdominal pain.

Severe acute pancreatitis is a medical emergency. Patients usually present with an acute abdomen and systemic findings caused by the release of toxic enzymes, cytokines, and other mediators into the circulation. These activate a systemic inflammatory response, resulting in leukocytosis, disseminated intravascular coagulation, edema, and acute respiratory distress syndrome. Shock, due to the systemic inflammatory response syndrome (Chapter 4), and acute renal tubular necrosis may occur.

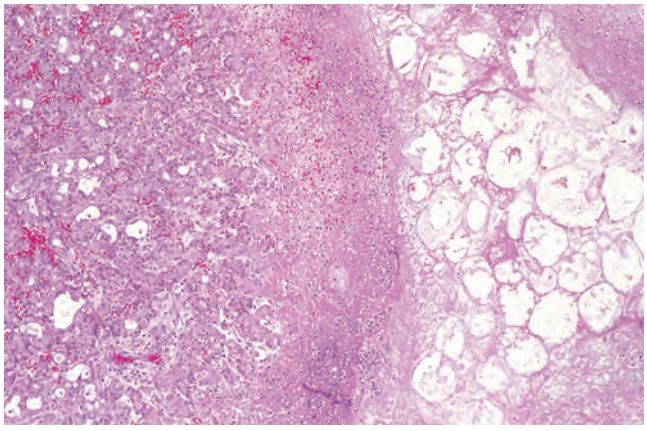


Figure 19.3 The microscopic field shows a region of fat necrosis on the right and focal pancreatic parenchymal necrosis (center).

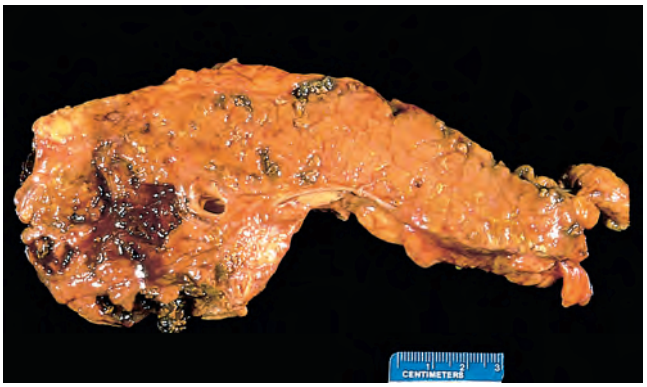


Figure 19.4 The pancreas has been sectioned longitudinally to reveal dark areas of hemorrhage in the head of the pancreas and a focal area of pale fat necrosis in the peripancreatic fat (upper left).

Laboratory findings include elevation of serum amylase and lipase levels during the first 4 to 12 hours following the onset of pain. Serum lipase is the most specific and sensitive marker of acute pancreatitis, as serum amylase has a short half-life and may return to normal in 3 to 5 days, whereas lipase levels remain elevated for 8 to 14 days. Glycosuria occurs in 10% of cases, and hypocalcemia may result from saponification of necrotic fat. Direct visualization of the enlarged inflamed pancreas by computed tomography (CT) scanning is useful in the diagnosis of pancreatitis.

Treatment centers on "resting" the pancreas by total elimination of oral intake and supportive therapy with intravenous fluids and analgesia. Most individuals with acute pancreatitis recover fully but about 5% with severe acute pancreatitis die in the first week of illness. Acute respiratory distress syndrome and acute renal failure are ominous complications. Possible sequelae include sterile pancreatic abscesses and pancreatic pseudocysts. In 40% to 60% of patients with acute necrotizing pancreatitis the acellular debris becomes infected, usually by gram-negative organisms from the gut, further complicating the clinical course. Systemic organ failure and pancreatic necrosis are both adverse prognostic indicators.

KEY CONCEPTS

ACUTE PANCREATITIS

- Acute pancreatitis is a form of reversible pancreatic parenchymal injury associated with inflammation.
- Acute pancreatitis may be caused by:
  - Excessive alcohol intake
  - Pancreatic duct obstruction (e.g., gallstones)
  - Genetic factors (e.g., *PRSS1*, *SPINK1*, *CFTR*)
  - Traumatic injuries
  - Medications
  - Infections (e.g., mumps or coxsackievirus)
  - Metabolic disorders leading to hypercalcemia
  - Ischemia

- The key feature common to all these causes is that they promote inappropriate activation of digestive enzymes within the pancreas.
- Clinical features include acute abdominal pain, systemic inflammatory response syndrome, and elevated serum lipase and amylase levels.

Chronic Pancreatitis

Chronic pancreatitis is defined as prolonged inflammation of the pancreas associated with irreversible destruction of exocrine parenchyma, fibrosis, and, in the late stages, loss of endocrine parenchyma. The prevalence of chronic pancreatitis is between 0.04% and 5%; most affected patients are middle-aged males. **The most common cause of chronic pancreatitis is long-term alcohol use.** In addition to alcohol, chronic pancreatitis has been associated with the following conditions:

- Long-standing *obstruction* of the pancreatic duct by calculi or neoplasms
- *Autoimmune injury*
- *Hereditary factors,* as discussed above; up to 25% of chronic pancreatitis has a genetic basis

Pathogenesis

Chronic pancreatitis often follows repeated episodes of acute pancreatitis. It has been proposed that acute pancreatitis initiates a sequence of perilobular fibrosis, duct distortion, and altered secretions that, as a result of recurrent injury, leads to loss of exocrine parenchyma and fibrosis.

Chronic pancreatic injury of any cause leads to local production of inflammatory mediators that promote fibrosis and acinar cell loss. While there is overlap between the cytokines released during chronic and acute pancreatitis, fibrogenic factors tend to predominate in chronic pancreatitis. These fibrogenic cytokines, including transforming growth factor β (TGFβ) and platelet-derived growth factor (PDGF), induce activation and proliferation of pancreatic stellate cells (periacinar myofibroblasts), collagen deposition, and fibrosis (Fig. 19.5).

Autoimmune pancreatitis is a pathogenically distinct form of chronic pancreatitis that comes in two distinct forms, each with its own characteristic histopathology. Autoimmune pancreatitis type 1 is associated with the presence of immunoglobulin G4 (IgG4)–secreting plasma cells in the pancreas and is one manifestation of a systemic IgG-related disease (Chapter 6). In contrast, autoimmune pancreatitis type 2 is restricted to the pancreas with the exception of a subset of patients with ulcerative colitis. Both variants of autoimmune pancreatitis may mimic pancreatic carcinoma, including presentation as a "mass lesion" in the pancreatic head on imaging. Autoimmune pancreatitis, which responds to steroid therapy, is therefore important to distinguish from neoplasia.

MORPHOLOGY

Chronic pancreatitis is characterized by parenchymal fibrosis, acinar atrophy and dropout, and variable ductal dilation (Fig. 19.6A). Grossly, the gland is hard, sometimes with visibly dilated ducts containing calcified concretions. These changes are typically accompanied by a histologically evident chronic inflammatory infiltrate that surrounds lobules and ducts. Acinar loss is a constant feature, but there is usually relative sparing of the islets of Langerhans, which become embedded in the sclerotic tissue and may fuse and appear enlarged. Ductal epithelium may

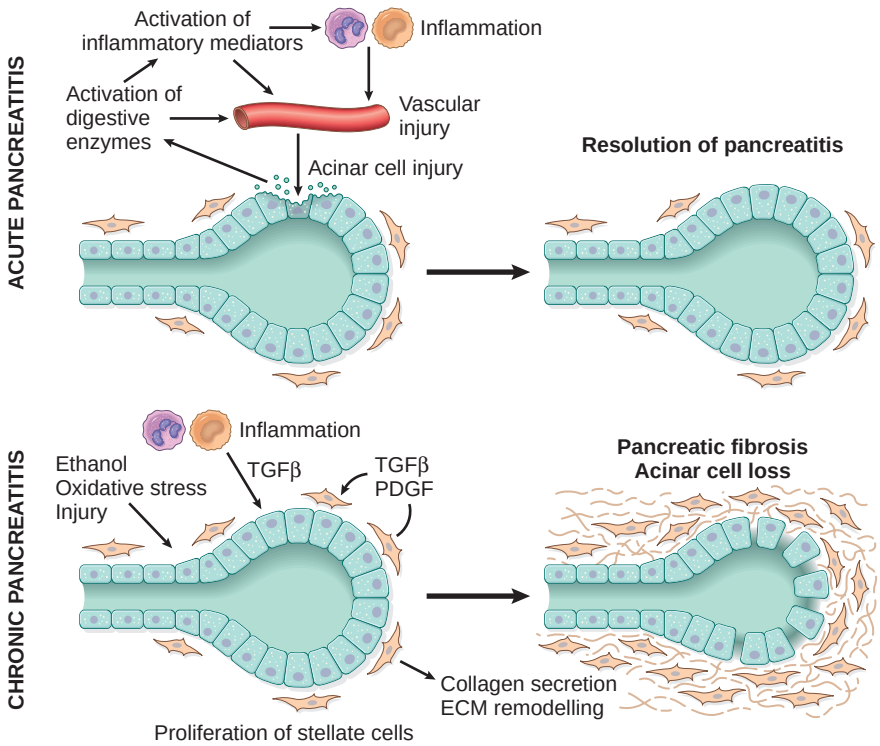


Figure 19.5 Comparison of the mediators in acute and chronic pancreatitis. In acute pancreatitis, acinar injury results in release of digestive enzymes, leading to a cascade of events including activation of the clotting cascade, acute and chronic inflammation, vascular injury, and edema. In most patients, complete resolution of the acute injury occurs with restoration of acinar cell mass. In chronic pancreatitis, repeated episodes of acinar cell injury lead to the production of profibrogenic cytokines such as transforming growth factor β (TGFβ) and platelet-derived growth factor (PDGF), resulting in myofibroblast proliferation, collagen synthesis, and extracellular matrix (ECM) remodeling. Repeated injury leads to irreversible loss of acinar cell mass, fibrosis, and (exocrine) pancreatic insufficiency.

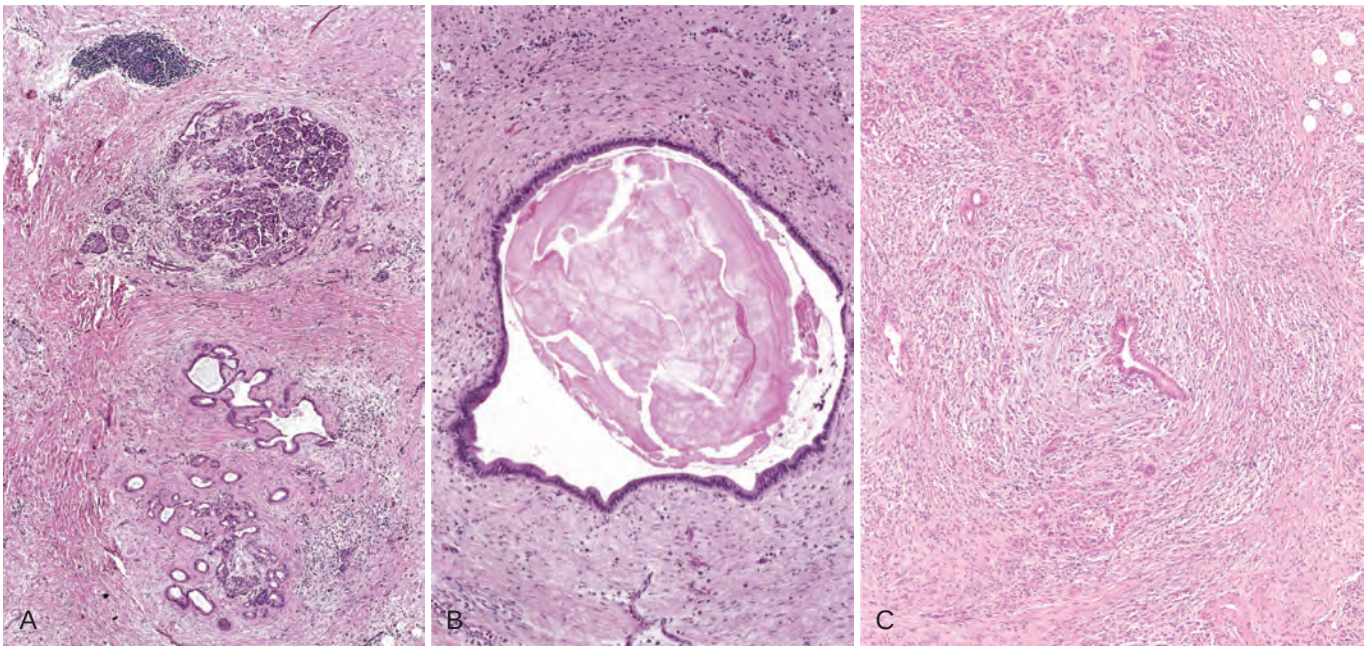


Figure 19.6 Chronic pancreatitis. (A) Extensive fibrosis and atrophy has left only residual islets (top) and ducts (bottom), with a sprinkling of chronic inflammatory cells and a few islands of acinar tissue. (B) Higher power view demonstrating dilated ducts with inspissated eosinophilic ductal concretions in a person with alcoholic chronic pancreatitis. (C) Example of autoimmune pancreatitis with extensive lymphoplasmacytic infiltrates and storiform fibrotic stroma. In this case, an IgG4 stain confirmed abundant IgG4-expressing plasma cells. (C, Photomicrograph courtesy Aatur D. Singhi, University of Pittsburgh Medical Center, Pittsburgh, Pa.)

be atrophic, hyperplastic, or metaplastic (squamous). Chronic pancreatitis caused by alcohol abuse is characterized by ductal dilation and intraluminal protein plugs and calcifications (Fig. 19.6B). **Autoimmune pancreatitis** type 1 typically displays swirling or storiform fibrosis, obliterative inflammation of the veins (phlebitis), and a dense lymphoplasmacytic inflammation that is enriched in IgG4-secreting plasma cells (Fig. 19.6C). The last-mentioned may also be seen in other organs. In contrast, autoimmune pancreatitis type 2 is characterized by neutrophilic infiltrates within the epithelium and lumen of medium-sized pancreatic ducts (granulocytic epithelial lesions). While lymphoplasmacytic infiltrates are seen in both variants, type 2 autoimmune pancreatitis lacks the abundant IgG4-secreting plasma cells that characterize type 1.

Clinical Features

Chronic pancreatitis may follow multiple bouts of acute pancreatitis with repeated attacks of mild to moderately severe abdominal pain or persistent abdominal and back pain. Attacks may be precipitated by excessive alcohol consumption, overeating (which increases demand on the pancreas), or use of opiates and other drugs that increase sphincter of Oddi tone. In other patients the disease may be entirely silent until pancreatic insufficiency and diabetes mellitus develop due to destruction of the exocrine and endocrine pancreas.

Diagnosis of chronic pancreatitis requires a high degree of suspicion. During an attack there may be mild fever and mild-to-moderate elevations of serum amylase. In chronic disease, however, acinar cell dropout may be so great that this diagnostic clue is absent. Gallstone-induced obstruction may present with jaundice or elevated serum alkaline phosphatase. Visualization of calcifications within the pancreas by CT and ultrasonography can be very helpful. Weight loss (due to pancreatic exocrine insufficiency and malabsorption) and edema (secondary to hypoalbuminemia) may also support the diagnosis.

Chronic pancreatitis is usually not an immediately life-threatening condition, but the long-term outlook is poor, with a 20- to 25-year mortality rate of 50%. Pancreatic exocrine insufficiency, chronic malabsorption, and diabetes mellitus all can lead to significant morbidity and contribute to mortality. In other patients, severe, chronic pain is a dominant problem. Pancreatic pseudocysts (described later) develop in about 10% of patients. Patients with hereditary pancreatitis associated with *PRSS1*-mutations have a 40% lifetime risk of developing pancreatic cancer; the risk of pancreatic cancer is only modestly elevated with other forms of chronic pancreatitis.

KEY CONCEPTS

CHRONIC PANCREATITIS

- Chronic pancreatitis is characterized by irreversible injury of the pancreas leading to fibrosis, loss of pancreatic parenchyma, loss of exocrine and endocrine function, and pseudocyst development.
- Chronic pancreatitis is most often caused by:
  - Repeated bouts of acute pancreatitis
  - Chronic alcohol use
  - Germline mutations in genes such as *CFTR* (the gene encoding the transporter that is defective in cystic fibrosis), particularly when combined with environmental stressors

- Chronic pancreatitis may be the manifestation of a systemic or localized autoimmune etiology.
- Clinical features include intermittent or persistent abdominal pain, intestinal malabsorption, and diabetes.

NONNEOPLASTIC CYSTS

A variety of cysts can arise in the pancreas. Most are nonneoplastic pseudocysts, but congenital cysts and neoplastic cysts also occur.

Congenital Cysts

Congenital cysts are unilocular, thin-walled cysts that likely result from anomalous pancreatic duct development. They range in size from microscopic to 5 cm in diameter and are lined by a uniform cuboidal epithelium or, if intracystic pressure is high, flattened and attenuated epithelial cells. A thin, fibrous capsule filled by clear serous fluid typifies congenital cysts that may be sporadic or part of inherited conditions such as *autosomal-dominant polycystic kidney disease* (Chapter 20) and *von Hippel–Lindau disease* (Chapter 28). Cysts in the kidney, liver, and pancreas frequently coexist in polycystic kidney disease. In von Hippel–Lindau disease vascular neoplasms are found in the retina and cerebellum or brain stem in association with congenital cysts (and also neoplasms) in the pancreas, liver, and kidney.

Pseudocysts

Pseudocysts are formed when areas of intrapancreatic or peripancreatic hemorrhagic fat necrosis are walled off by fibrosis and granulation tissue. These lesions, which account for 75% of all pancreatic cysts, are referred to as pseudocysts because they lack an epithelial lining. Pseudocysts typically arise following a bout of acute pancreatitis, particularly one superimposed on chronic alcoholic pancreatitis. Traumatic injury to the pancreas can also give rise to pseudocysts. While many pseudocysts resolve spontaneously, they may also become secondarily infected, and larger pseudocysts may compress or even perforate into adjacent structures.

MORPHOLOGY

Pseudocysts are usually solitary and may be situated within the pancreas or, more commonly, in the lesser omental sac or in the retroperitoneum between the stomach and transverse colon or between the stomach and liver. They can even be subdiaphragmatic (Fig. 19.7A). Pseudocysts are lined by fibrous tissue and granulation tissue (Fig. 19.7B) and range in size from 2 to 30 cm in diameter.

NEOPLASMS

A broad spectrum of exocrine neoplasms arises in the pancreas. These neoplasms may be cystic or solid; some are benign, while others are among the most lethal of all malignancies. Endocrine tumors also occur in the pancreas and are discussed in Chapter 24.

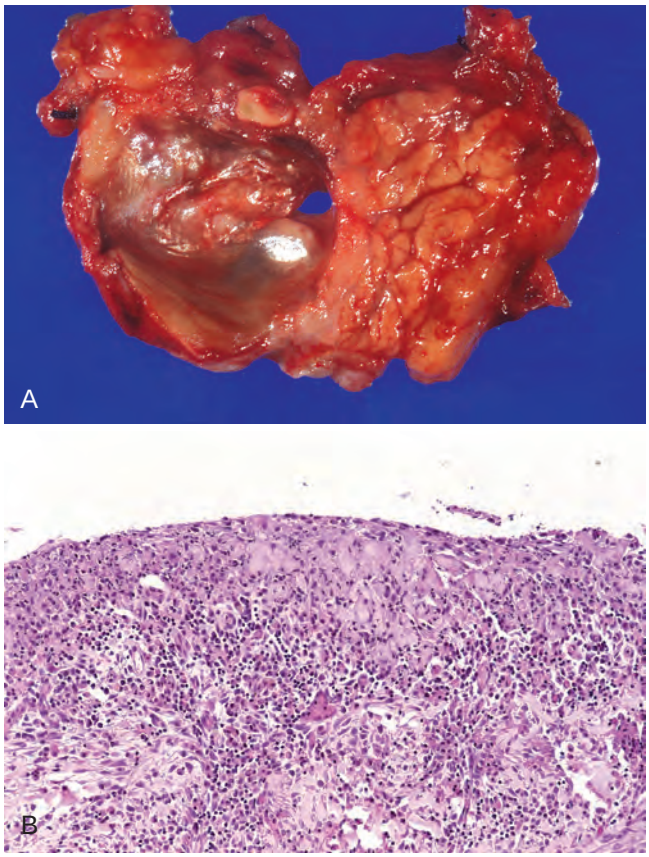


Figure 19.7 Pancreatic pseudocyst. (A) Cross-section revealing a poorly defined cyst with a necrotic brown-black wall. (B) The cyst lacks a true epithelial lining and instead is lined by fibrin and granulation tissue.

Cystic Neoplasms

Cystic neoplasms are diverse tumors that range from harmless benign cysts to lesions that are precursors to invasive, potentially lethal, cancers. Only 5% to 15% of all pancreatic cysts are neoplastic; conversely, fewer than 5% of all pancreatic neoplasms are cystic. Serous cystic neoplasms are almost always benign, whereas others, such as intraductal papillary mucinous neoplasms and mucinous cystic neoplasms, are precancerous. The genomic landscape of each type of cystic neoplasm is distinct, underscoring differences in pathogenesis and natural history.

Serous cystic neoplasms usually occur in the tail of the pancreas. The cysts are small (1 to 3 mm) and can be solitary, multiple, or present as a honeycomb of microcystic lesions. Serous cysts are lined by glycogen-rich cuboidal cells, and contain clear, thin, straw-colored fluid (Fig. 19.8). They account for 15% to 25% of all cystic neoplasms of the pancreas and are twice as common in women. Serous cystic neoplasms typically present in the sixth to seventh decade of life with nonspecific symptoms such as abdominal pain, but many are detected incidentally during imaging for another indication. Surgical resection is curative in the vast majority of patients. Inactivation of the *VHL* tumor suppressor gene

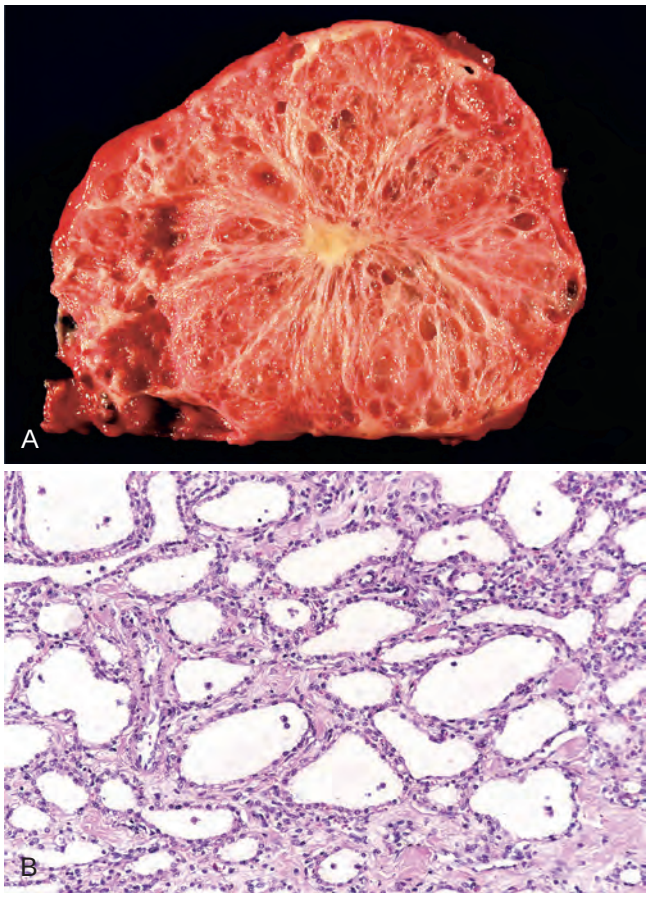


Figure 19.8 Serous cystic neoplasm (serous cystadenoma). (A) Cross-section through a microcystic serous cystic neoplasm. Only a thin rim of normal pancreatic parenchyma remains. The cysts are relatively small and contain clear, straw-colored fluid. (B) The cysts are lined by cuboidal epithelium without atypia.

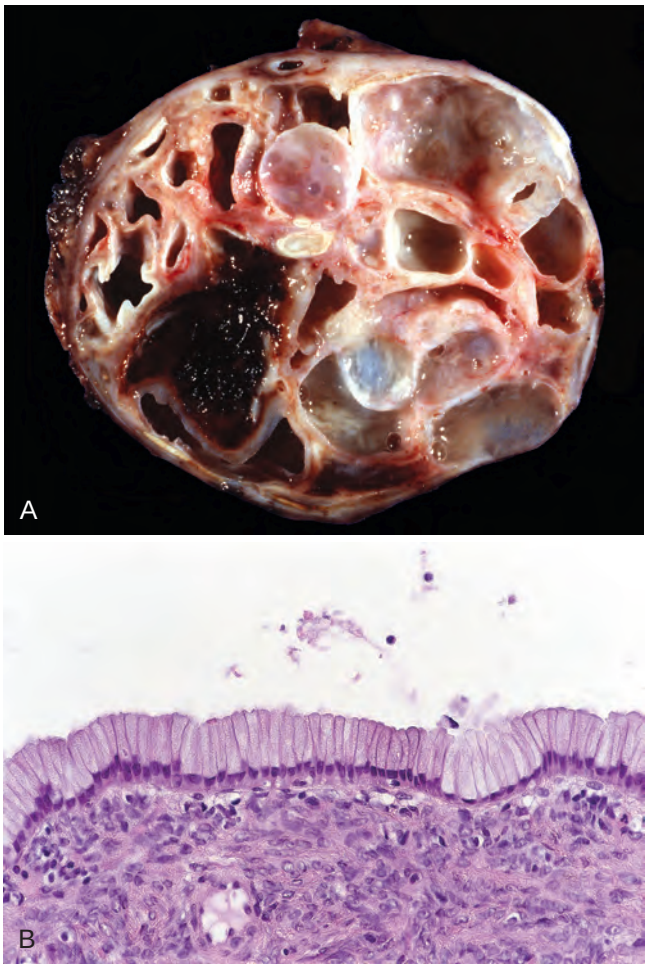


Figure 19.9 Pancreatic mucinous cystic neoplasm with low-grade dysplasia. (A) Cross-section through a mucinous multiloculated cyst in the tail of the pancreas. The cysts are large and filled with tenacious mucin. (B) The cysts are lined by columnar mucinous epithelium with a dense "ovarian" stroma.

on chromosome 3p is the most common genetic abnormality in serous cystic neoplasms.

Close to 95% of *mucinous cystic neoplasms* arise in women and, in contrast to serous cystic neoplasms, they are precursors to invasive carcinomas. These neoplasms usually arise in the tail of the pancreas and present as painless, slow-growing masses. The cystic cavities are larger than those in serous cystic neoplasms and are filled with thick, tenacious mucin. The columnar mucin-producing epithelium that lines the cysts is associated with a dense stroma similar to ovarian stroma (Fig. 19.9). The latter frequently expresses estrogen and progesterone receptors as well as other markers of the ovarian stroma, such as inhibin. Surgical resection is curative for noninvasive mucinous cystic neoplasms, but up to one-third harbor an invasive adenocarcinoma. Up to 50% of patients with an invasive adenocarcinoma arising in a mucinous cystic neoplasm will succumb to their disease within 5 years; therefore, early detection and treatment, before invasive cancer develops, is critical. Mucinous cystic neoplasms harbor oncogenic *KRAS* mutations in approximately half of cases, while mutations of *TP53* and *SMAD4* are typically observed in invasive neoplasms arising from these cysts. Recently, loss-of-function mutations of *RNF43,* which encodes an E3 ubiquitin ligase that normally downregulates Wnt signaling, have been described in up to one-third of mucinous cystic neoplasms. Similar *RNF43* mutations also occur in colorectal cancers.

Intraductal papillary mucinous neoplasms (IPMNs) are mucin-producing neoplasms that involve the larger ducts of the pancreas. In contrast to mucinous cystic neoplasms, IPMNs arise more frequently in men and tend to involve the head of the pancreas. Up to 20% are multifocal. Two features are useful in distinguishing IPMNs from mucinous cystic neoplasms: (1) absence of the dense "ovarian" stroma seen in mucinous cystic neoplasms and (2) involvement of a pancreatic duct (Fig. 19.10). Just as with mucinous cystic neoplasms, IPMNs can progress to an invasive cancer; early detection and treatment before progression to invasive cancer is thus critical. Similar to mucinous cystic neoplasms, intraductal mucinous papillary neoplasms harbor oncogenic *KRAS* mutations in approximately 80% of tumors and loss-of-function *RNF43* mutations in up to 50%. *TP53* and *SMAD4* mutations typically occur only with transition to invasive cancer. Notably, oncogenic mutations of *GNAS,* which encodes the alpha subunit of the stimulatory G protein,

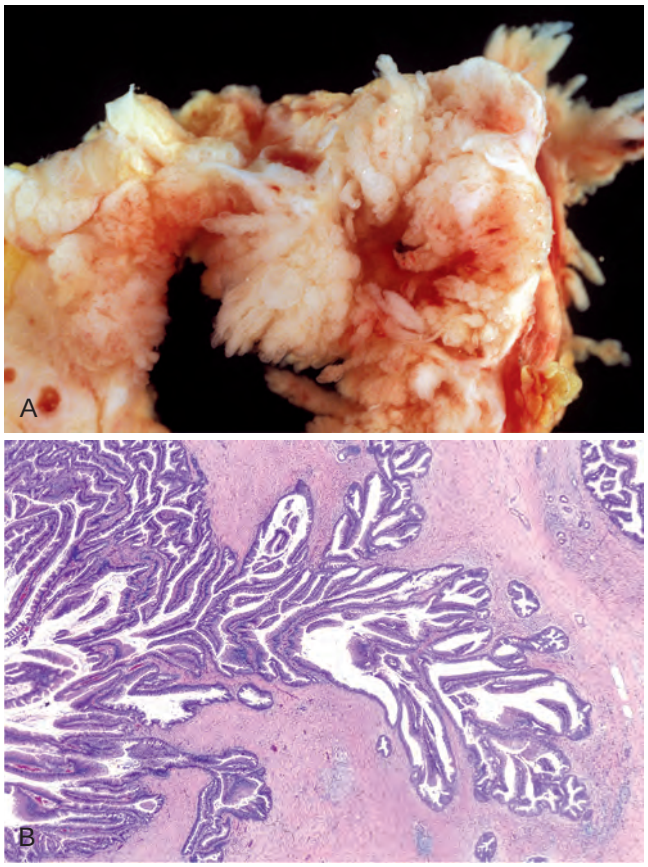


Figure 19.10 Intraductal papillary mucinous neoplasm. (A) Cross-section through the head of the pancreas showing a prominent papillary neoplasm distending the main pancreatic duct. (B) The neoplasm involves the main pancreatic duct (left) and extends down into the smaller ducts and ductules (right).

Gsα, are present in approximately two-thirds of IPMNs but are not found in other pancreatic cysts.

The uncommon *solid-pseudopapillary neoplasm* occurs mainly in young women and due to its large size often presents with abdominal discomfort. As implied by the name, these well-circumscribed neoplasms have solid and cystic components, with the latter filled with hemorrhagic debris. The neoplastic cells grow in solid sheets or, as the name suggests, as pseudopapillary projections and often appear to be poorly cohesive. Activating mutations of *CTNNB1* (β-catenin) are present in nearly all of these tumors, while mutations of *KRAS, RNF43, GNAS,* or *VHL* (characteristic of other pancreatic cystic neoplasms) are not observed. Surgical resection is the treatment of choice and proves curative in most patients.

KEY CONCEPTS

CYSTIC NEOPLASMS

- Virtually all serous cystic neoplasms are benign.
- Noninvasive intraductal papillary mucinous neoplasms and mucinous cystic neoplasms are almost always curable, but both may progress to invasive, potentially lethal carcinomas.
- Each of the major cystic neoplasms has a relatively specific mutational profile.

Pancreatic Carcinoma

Infiltrating ductal adenocarcinoma of the pancreas, more commonly known as pancreatic cancer, is the third leading cause of cancer deaths in the United States, trailing only lung and colon cancers, and has one of the highest mortality rates of any cancer. It was estimated that in 2020 pancreatic cancer would strike approximately 57,600 Americans, virtually all of whom would die of their disease. The 5-year survival rate is a dismal 10%.

Precursors to Pancreatic Cancer

Invasive pancreatic cancers arise from noninvasive precursor lesions referred to as pancreatic intraepithelial neoplasia (PanIN) (Fig. 19.11). These lesions develop in small ducts and are often microscopic, although some, such as the two variants of cystic mucinous neoplasms described above, can be detected macroscopically. Overall, it is thought that greater than 90% of pancreatic cancers arise from PanIN (with the remainder arising from cystic lesions). This conclusion is supported by the following observations:

- The genetic and epigenetic alterations identified in PanIN are similar to those found in invasive cancers (described later).
- PanIN is often found in pancreatic parenchyma adjacent to infiltrating carcinoma. A corollary of this observation is that while isolated low-grade PanIN lesions can be observed in the pancreata of elderly individuals, high-grade PanIN (also known as carcinoma in situ) is almost never observed in the absence of an invasive cancer.
- PanIN precedes the development of invasive cancer in genetically engineered mouse models of pancreatic cancer.
- Isolated case reports have documented individuals with PanIN who later developed an invasive pancreatic cancer.

The epithelial cells in PanIN show dramatic telomere shortening, which may predispose these lesions to accumulate chromosomal abnormalities and progress to invasive carcinoma (Chapter 7). As previously mentioned, PanIN is separated into low-grade and high-grade based on

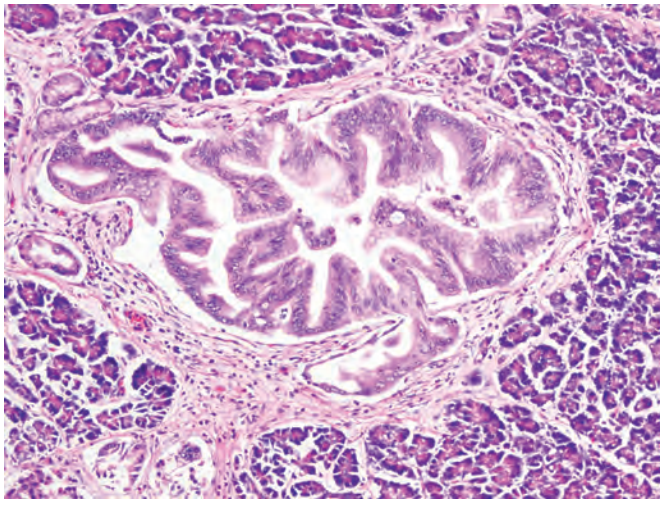


Figure 19.11 High-grade pancreatic intraepithelial neoplasia involving a small pancreatic duct.

morphologic features, which also tend to correlate with the stepwise accumulation of genetic abnormalities (Fig. 19.12).

Pathogenesis

As with other carcinomas, pancreatic cancer is the product of complementary mutations and epigenetic alterations that alter the expression of oncogenes and tumor suppressor genes. The most common molecular alterations in pancreatic carcinogenesis are summarized in Table 19.3 and include the following.

- ***KRAS* (chromosome 12p) is the most frequently altered oncogene in pancreatic cancer, with activating point mutations present in greater than 90% of cases.** These result in constitutive activation of Ras, a small, guanosine triphosphate (GTP)–binding protein enzyme that normally participates in signaling events downstream of growth factor receptors with intrinsic tyrosine kinase activity (Chapters 1 and 7). Ras signaling activates several downstream pathways that augment cell growth and survival, most notably the mitogen-activated protein kinase (MAPK) and PI3K/AKT pathways (Chapter 7).
- *CDKN2A* (chromosome 9p) is inactivated in 30% of pancreatic cancers through point mutations or homozygous deletions. This complex locus encodes two tumor suppressor proteins (Chapter 7): p16/INK4a, a cyclin-dependent kinase inhibitor that antagonizes cell cycle progression, and ARF, a protein that augments the function of the p53 tumor suppressor protein.
- *SMAD4* (chromosome 18q) is inactivated in 55% of pancreatic cancers. *SMAD4* encodes a tumor suppressor that plays an important role in signal transduction from the TGFβ family of cell surface receptors. *SMAD4* is only rarely inactivated in other cancer types.
- *TP53* (chromosome 17p) is inactivated in 70% to 75% of pancreatic cancers. This gene encodes the tumor suppressor p53, a nuclear DNA-binding protein that can respond to DNA damage by arresting cell growth, inducing cell death (apoptosis), or causing cellular senescence (Chapter 7).

Also seen are less common mutations that affect genes involved in DNA repair or regulation of chromatin structure (Table 19.3). In addition, other work has identified pathogenic epigenetic changes, which together with the aforementioned genetic changes produce several distinct molecular subtypes of pancreatic cancer.

- *DNA methylation (epigenetic) abnormalities.* DNA methylation abnormalities are widespread in pancreatic cancer. For example, promoter hypermethylation causes transcriptional silencing of tumor suppressor genes including *CDKN2A*. Conversely, promoter hypomethylation leads to overexpression of oncogenes such as *GATA6* and *BRD4*.
- *Transcriptomic profiles and pancreatic cancer subtypes.* Global analyses of gene expression have identified two distinct pancreatic cancer subtypes, termed basal-like and classical. Basal-like pancreatic cancers are highly aggressive, whereas the classical subtype has a somewhat more favorable prognosis. How these subtypes respond differentially to specific therapies is an area of active investigation.

Epidemiology and Inheritance. Pancreatic cancer is primarily a disease of older adults, with 80% of cases occurring after age 60. The incidence is increased in African Americans, Japanese Americans, Native Hawaiian Islanders, and Ashkenazi Jews.

The strongest environmental influence is cigarette smoking, which doubles the risk of pancreatic cancer. Chronic pancreatitis is an additional risk factor that may reflect the association between chronic inflammation, tissue repair, and neoplasia. Pancreatic cancer risk is also greater in those with visceral obesity and high body mass index. Diabetes mellitus is also a modest risk factor. Conversely, new-onset diabetes mellitus may be the first indication of an occult pancreatic cancer in older patients; abnormal glucose tolerance or frank diabetes is present in as many as half of pancreatic cancer patients up to 3 years before clinical signs of cancer appear. Thus, the appearance of diabetes may present an opportunity for early diagnosis of pancreatic cancer.

Approximately 10% of patients with pancreatic cancer have a deleterious germline mutation in a cancer predisposition gene (Table 19.4) or report one or more first-degree relatives with pancreatic cancer. *BRCA2* mutations are the

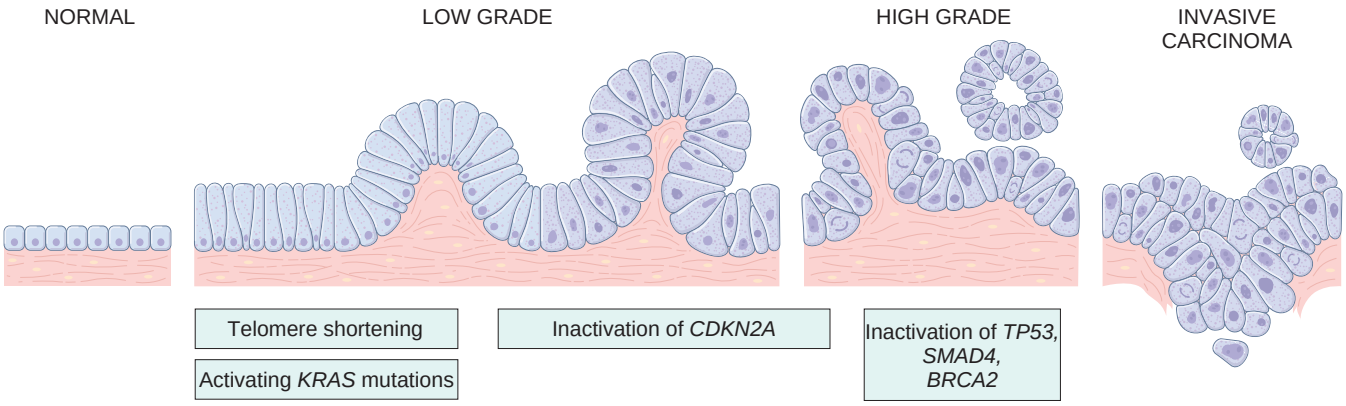


Figure 19.12 Model for the progression from normal ducts (far left) through low- and high-grade pancreatic intraepithelial neoplasia (center) to invasive carcinoma (far right). Telomere shortening and mutations of the oncogene *KRAS* are the earliest discernible alterations in multistep progression, followed by inactivation of the *CDKN2A* tumor suppressor gene that encodes the cell cycle regulator p16. Inactivation of the *TP53*, *SMAD4*, and *BRCA2* tumor suppressor genes occurs in high-grade pancreatic intraepithelial neoplasia that precedes invasive cancer. It is important to note that while there is a general temporal sequence of changes, the accumulation of multiple mutations is more important than their occurrence in a specific order.

Table 19.3 Somatic Molecular Alterations in Pancreatic Ductal Adenocarcinoma

Gene	Chromosomal Region	Percentage of Cases With Genetic Alteration[a]	Gene Function
Oncogenes (Frequency Includes Point Mutations and Amplifications)			
KRAS	12p	>90%	GTP binding enzyme and growth factor signal transducer
AKT2	19q	6%	Growth factor signal transducer
MYC	8q	5%	Transcription factor
GATA6	18q	9%	Transcription factor
FGFR1	8p	5%	Growth factor receptor
BRAF	7q	3%	Serine threonine kinase, regulates MAPK signaling
Tumor Suppressor Genes (Frequency Includes Point Mutations and Homozygous Deletions)			
p16/CDKN2A	9p	30%	Negative cell-cycle regulator
TP53	17p	75%	Response to DNA damage
SMAD4	18q	55%	TGFβ pathway
BRCA2	13q	4%	DNA damage response
ATM	11q	5%	DNA damage response
ARID1A	1p	6%	Chromatin regulator
MLL3/KMT2C	7q	4%	Chromatin regulator
KDM6A	Xp	3%	Chromatin regulator

[a]Frequency of alterations abstracted from The Cancer Genome Atlas Research Network: Integrated Molecular Characterization of Pancreatic Ductal Adenocarcinoma, *Cancer Cell* 32(2):185–203.e13, 2017.

GTP, Guanosine triphosphate; *MAPK*, mitogen-activated protein kinase; *TGFβ*, transforming growth factor β.

most common known cause of familial pancreatic cancer; germline mutations in other genes that are linked to hereditary breast and ovarian cancer (*BRCA1, PALB2, ATM*) are also associated with a 4- to 10-fold elevated risk. With the availability of therapeutic agents that specifically target cancers with DNA repair defects, most commonly *BRCA1* or *BRCA2* mutations, germline testing is now recommended for all patients with pancreatic cancer. Patients with defects in DNA mismatch repair genes who have hereditary nonpolyposis colorectal cancer are also at increased risk for pancreatic cancer. Identification of microsatellite instability, the molecular signature of DNA mismatch repair in pancreatic cancer, is also therapeutically important, as such cancers are more likely to respond to immune checkpoint therapy.

Table 19.4 Inherited Predisposition to Pancreatic Cancer

Disorder	Gene	Increased Risk of Pancreatic Cancer (Fold)	Risk of Pancreatic Cancer by Age 70 (%)
Peutz-Jeghers syndrome	*STK11*	130	30–60
Hereditary pancreatitis	*PRSS1*, *SPINK1*	50–80	25–40
Familial atypical multiple-mole melanoma syndrome	*CDKN2A*	20–35	10–17
Strong family history (3 or more relatives with pancreatic cancer)	Unknown	14–32	8–16
Hereditary breast and ovarian cancer	Multiple, including *BRCA1*, *BRCA2*, *PALB2*, *ATM*	4–10	5
Hereditary non-polyposis colorectal cancer	Multiple, including *MLH1*, *MSH2*, *PMS2*	8–10	4

MORPHOLOGY

Approximately 60% of pancreatic cancers arise in the head, 15% in the body, and 5% in the tail; in 20% of cases the entire pancreas is involved.

The vast majority of pancreatic neoplasms are adenocarcinomas that recapitulate normal ductal epithelium by forming glands and secreting mucin. Two features are characteristic: (1) these cancers are highly invasive, often extending into peripancreatic tissues; and (2) they elicit an intense desmoplastic response that results in the deposition of dense collagen. As a result of these characteristics, most pancreatic cancers are hard, stellate, gray-white, poorly defined masses (Fig. 19.13A).

Most carcinomas of the head of the pancreas obstruct the distal common bile duct resulting in distention of the biliary tree in about 50% of patients. In contrast, carcinomas of the pancreatic body and tail do not impinge on the biliary tract and may be large and widely disseminated at diagnosis.

Pancreatic cancers tend to grow along nerves and invade into blood vessels and the retroperitoneum. The spleen, adrenals, transverse colon, and stomach are often involved by direct invasion, and peripancreatic, gastric, mesenteric, omental, and portohepatic lymph nodes are frequently involved. Perineural, lymphatic, and large vessel invasion are common. Distant metastases are principally to the liver and lungs.

Microscopically, carcinomas of the head, body, and tail of the pancreas are indistinguishable. They form abortive tubular structures or cell clusters with an aggressive, deeply infiltrative growth pattern

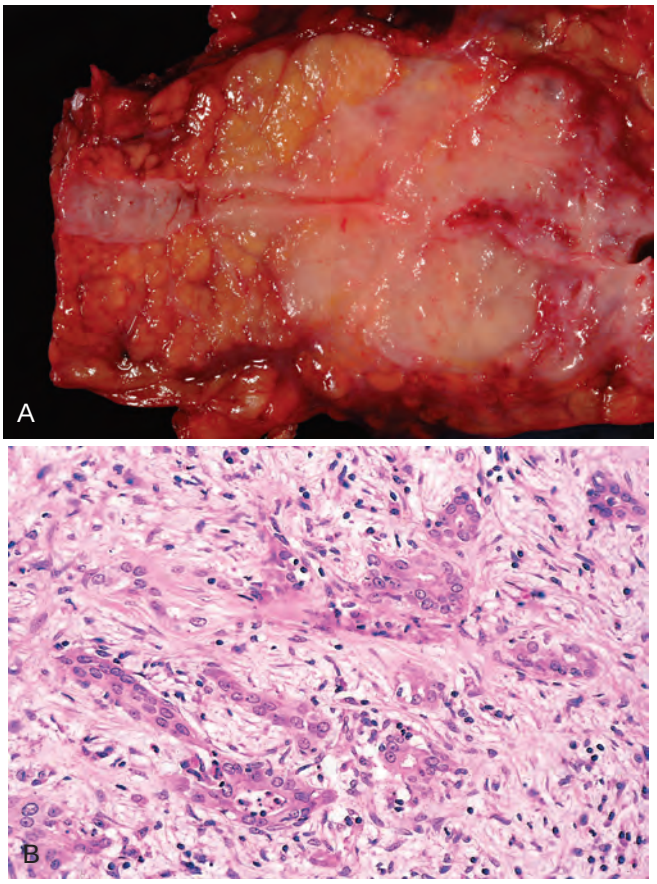


Figure 19.13 Carcinoma of the pancreas. (A) Cross-section through the tail of the pancreas showing normal pancreatic parenchyma and a normal pancreatic duct (*left*) and a pale mass centered on the duct (*right*). (B) Poorly formed glands composed of malignant epithelial cells are dispersed within densely fibrotic stroma; some inflammatory cells are also present.

(Fig. 19.13B). The poorly formed malignant glands are lined by pleomorphic cuboidal-to-columnar epithelial cells. It can, however, occasionally be difficult to discriminate between well-differentiated carcinomas and benign glands with epithelial atypia and stromal fibrosis that are common in chronic pancreatitis. The intense desmoplasia elicited by pancreatic cancers can also interfere with diagnostic interpretation of needle biopsies, regardless of tumor differentiation, as much of the tumor mass is nonneoplastic.

Less common morphologic variants of pancreatic cancer include adenosquamous carcinomas, colloid carcinoma, hepatoid carcinoma, medullary carcinoma, signet-ring cell carcinoma, undifferentiated carcinoma, and undifferentiated carcinoma with osteoclast-like giant cells.

Clinical Features

Carcinomas of the pancreas typically remain silent until they cause obstruction or invade into adjacent structures. Pain is usually the first symptom, but by the time pain appears these cancers are usually beyond cure. Obstructive jaundice is associated with most cases involving the pancreatic head, as these tend to block the common bile duct; this is exemplified by the *Courvoisier sign,* a palpably enlarged, nontender gallbladder with mild painless jaundice. Weight loss, anorexia, and generalized malaise and weakness are often signs of advanced disease. Migratory thrombophlebitis, known as the *Trousseau sign,* occurs in about 10% of patients and is attributable to the elaboration of platelet-activating factors and procoagulants from the carcinoma or its necrotic products (Chapter 4). Sadly, Armand Trousseau (1801–1867, physician at Hotel Dieu, Paris), for whom this sign is named, correctly suspected that he had carcinoma when he developed spontaneously appearing and disappearing (migratory) thromboses.

Survival after diagnosis of advanced pancreatic carcinoma is typically short. Greater than 80% of pancreatic cancers are unresectable at the time of diagnosis due to invasion of vessels and other structures or distant metastases. In contrast, patients who undergo successful resection are increasingly surviving longer (some greater than 5 years), underscoring the critical importance of early detection. To facilitate diagnosis when pancreatic cancers are resectable, many have sought tests that could be used for early detection. Although serum levels of several antigens (e.g., carcinoembryonic antigen and CA19-9 antigen) are often elevated in individuals with pancreatic cancer and can be useful in following individual patient's response to treatment, these markers lack the specificity and sensitivity needed for tests applied to larger populations. Although imaging techniques such as endoscopic ultrasonography and CT can be used to establish the diagnosis, they are not useful as screening tests. Thus, the US Screening and Prevention Task Force (USPSTF) does not currently recommend screening the general population for pancreatic cancer due to the overarching concern for false positives and potential harm. Screening is however recommended for individuals who harbor deleterious germline mutations (Table 19.4) that place them at increased risk for pancreatic cancer.

KEY CONCEPTS

PANCREATIC CARCINOMA

- Pancreatic cancer is one of the most aggressive of the solid cancers.
- Cigarette smoking is a significant cause of pancreatic cancer.
- Cancer-causing germline mutations are present in 10% of patients.
- Invasive pancreatic cancer arises from histologically well-defined precursor lesions, the most common of which is pancreatic intraepithelial neoplasia (PanIN).
- Ductal adenocarcinomas are highly invasive and elicit an intense desmoplastic response.
- The genes most frequently mutated or otherwise altered in pancreatic cancer include *KRAS, p16/CDKN2A, TP53,* and *SMAD4;* transcriptional profiles can be used to define the highly aggressive basal-like and somewhat less aggressive classical subtypes.
- Patients often present with abdominal pain and weight loss, sometimes accompanied by jaundice and deep vein thrombosis. New-onset diabetes is detected in up to half of cases.

Acinar Cell Carcinoma

Like normal acinar cells, acinar cell carcinomas form zymogen granules and produce exocrine enzymes such as trypsin and lipase. Up to 15% of those with acinar cell carcinoma develop metastatic fat necrosis syndrome due to lipase release

into the circulation. Acinar cell carcinomas demonstrate aberrant Wnt pathway activation, due to either loss-of-function mutations of the *APC* tumor suppressor gene or activating point mutations of *CTNNB1*, which encodes beta-catenin.

Pancreatoblastoma

Pancreatoblastomas are rare neoplasms that occur primarily in children 1 to 15 years of age. They have a distinct microscopic appearance consisting of squamous islands admixed with acinar cells. They are malignant, but survival is better than with pancreatic ductal adenocarcinomas. Like acinar cell carcinoma, pancreatoblastomas frequently harbor mutations that activate the Wnt signaling pathway.

SUGGESTED READINGS

Congenital Anomalies

Stanger BZ, Hebrok M: Control of cell identity in pancreas development and regeneration, *Gastroenterology* 144:1170–1179, 2013. *[Review of the normal development of the pancreas and how development explains anomalies].*

Acute Pancreatitis

Forsmark CE, Vege SS, Wilcox CM: Acute pancreatitis, *N Engl J Med* 375:1972–1981, 2017. *[Clinically oriented authoritative overview of acute pancreatitis, especially guidelines for classification and management].*

Kang R, Lotze MT, Zeh HJ et al: Cell death and DAMPs in acute pancreatitis, *Mol Med* 20:466–477, 2014. *[Mechanistic review of acute pancreatitis highlighting how local tissue damage translates into the systemic inflammatory response that is characteristic of severe disease].*

Lankisch PG, Apte M, Banks PA: Acute pancreatitis, *Lancet* 386:85–96, 2015. *[Another clinically oriented review of acute pancreatitis, this time with more of an international focus].*

Mounzer R, Whitcomb DC: Genetics of acute and chronic pancreatitis, *Curr Opin Gastroenterol* 29:544–551, 2013. *[Review of genetic factors that predispose to acute pancreatitis and recurrent acute pancreatitis leading to chronic pancreatitis from one of the leading experts in this area].*

Chronic Pancreatitis

Hart PA, Zen Y, Chari ST: Recent advances in autoimmune pancreatitis, *Gastroenterology* 149:39–51, 2015. *[Review of this uncommon entity including the two subtypes of autoimmune pancreatitis, clinical and imaging features, and management].*

Hegyi E, Sahin-Toth M: Genetic risk in chronic pancreatitis—the trypsin dependent pathway, *Dig Dis Sci* 62:1692–1701, 2017. *[Mechanistic review that ties together how mutations in genes responsible for trypsin homeostasis lead to chronic pancreatitis].*

Kleef J, Whitcomb DC, Shimosegawa T et al: Chronic pancreatitis, *Nat Rev Dis Primers* 3:17060, 2017. *[The primer series is wonderful for a quick "one-stop shop" overview of any entity, for example, chronic pancreatitis. All the authors are world-renowned experts in this area].*

Majumder S, Chari ST: Chronic pancreatitis, *Lancet* 387:1957–1966, 2016. *[Excellent treatise on pathophysiology, mechanisms, and management of chronic pancreatitis].*

Pseudocysts

Klöppel G: Pseudocysts and other non-neoplastic cysts of the pancreas, *Semin Diagn Pathol* 17:7–15, 2000. *[Discussion of pseudocysts by a leading expert in pancreatic pathology].*

Cystic Neoplasms

Basturk O, Hong SM, Wood LD et al: A revised classification system and recommendations from the Baltimore consensus meeting for neoplastic precursor lesions in the pancreas, *Am J Surg Pathol* 39:1730–1741, 2015. *[Guidelines defining the diagnostic criteria and dysplasia grading in the common precursor lesions of pancreatic cancer].*

Singhi AD, McGrath K, Brand RE et al: Preoperative next generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia, *Gut* 2017. *[Outstanding example of how molecular pathology is integrated into the diagnostic algorithm of the management of pancreatic cysts to differentiate conservative management versus surgery].*

Springer S, Masica DL, Dal Molin M et al: A multimodality test to guide the management of patients with a pancreatic cyst, *Sci Trans Med* 11:eaav4772, 2019. *[Real-world application of molecular assays in pancreatic cyst fluid to classify the underlying cyst type in the pancreas].*

Tanaka M, Fernandez-del Castillo C, Kamisawa T et al: Revisions of International Consensus Fukuoka Guidelines for the Management of IPMN of the Pancreas, *Pancreatology* 17:738–753, 2017. *[Guidelines on the clinical management of cysts].*

Wu J, Jiao Y, Dal Molin M et al: Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways, *Proc Natl Acad Sci USA* 108:21188–21193, 2011. *[Sequencing the exomes of the most common cystic neoplasms of the pancreas reveals a mutation profile specific for each type of cyst].*

Pancreatic Carcinoma

Cancer Genome Atlas Research Network: Integrated genomic characterization of pancreatic ductal adenocarcinoma, *Cancer Cell* 32:185–203, 2017. *[Definitive article on molecular characterization of pancreatic cancer, using multiple platforms and integrative data analysis].*

Chari ST, Kelly K, Hollingsworth MA et al: Early detection of pancreatic cancer: a summative review, *Pancreas* 44:693–712, 2015. *[Roadmap for early detection approaches in pancreatic cancer that also highlights the importance of new-onset diabetes as an avenue for enriching for patients at risk for this cancer].*

Hu C, Hart SN, Polley EC et al: Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer, *JAMA* 19:2401–2409, 2018. *[Largest case control series on identifying deleterious germline mutations that correlate with risk of pancreatic cancer].*

Kleeff J, Korc M, Apte M et al: Pancreatic cancer, *Nat Rev Dis Prim* 2:16022, 2016. *[A relatively comprehensive review on pathology, genetics, clinical features, and treatment of pancreatic cancer].*

Neoptolemos JP, Kleeff J, Michl P et al: Therapeutic developments in pancreatic cancer: current and future perspectives, *Nat Rev Gastroenterol Hepatol* 15:333–348, 2018. *[Clinically oriented review on recent advances in treatment options for localized and metastatic pancreatic cancer and emerging therapies like immunotherapies].*

Ying H, Dey P, Yao W et al: Genetics and biology of pancreatic ductal adenocarcinoma, *Genes Dev* 30:355–385, 2016. *[Exhaustive review of molecular aberrations in pancreatic cancer, both genomic and nongenomic, and potential translational implications].*

CHAPTER 20

The Kidney

Anthony Chang • Zoltan G. Laszik

CHAPTER CONTENTS

What is a human but an ingenious machine designed to turn, with "infinite artfulness, the red wine of Shiraz into urine?" So said the storyteller in Isak Dinesen's *Seven Gothic Tales*. More accurately but less poetically, human kidneys serve to convert more than 1700 L of blood per day into about 1 L of a highly concentrated fluid called *urine*. In so doing, the kidney excretes the waste products of metabolism; precisely regulates the body's concentration of water, salt, calcium, phosphorus, and other anions and cations; and maintains the appropriate acid-base balance of plasma. The kidney also serves as an endocrine organ, secreting such hormones as erythropoietin, renin, and prostaglandins, and regulating vitamin D metabolism. The physiologic mechanisms that the kidney has developed to carry out these functions require a high degree of structural complexity.

Renal diseases are responsible for a great deal of morbidity and mortality. According to the 2015 U.S. Renal Data System Annual Data Report, more than 660,000 Americans had end-stage renal disease (ESRD), of whom two-thirds are maintained on dialysis, at a cost of approximately $72,000 per person annually. The 1-year mortality rate of ESRD, when the enhanced risk for cardiovascular disease conferred by ESRD is considered, exceeds that of most newly diagnosed cancers. Acute kidney injury occurs in more than 2 million people worldwide, and it is a major risk factor for the development of chronic kidney disease and ESRD. In addition, millions of people are affected annually by nonfatal kidney diseases, most notably infections of the kidney or lower urinary tract, kidney stones, and urinary obstruction. The availability of dialysis and the success of renal transplantation have improved the outlook for patients.

The study of kidney diseases is facilitated by dividing them into those that affect the four basic morphologic components: glomeruli, tubules, interstitium, and blood vessels. This approach is useful because the early manifestations of disease affecting each of these components tend to be distinct. Further, some components seem to be more vulnerable to specific forms of renal injury; for example, **most glomerular diseases are immunologically mediated, whereas tubular and interstitial disorders are frequently caused by toxic or infectious agents.** However, some disorders affect more than one structure, and the anatomic and functional interdependence of the components of the kidney means that damage to one usually secondarily affects the others. Primary disorders of the blood vessels, for example, inevitably affect all structures supplied by these vessels. Severe glomerular damage impairs the flow through the peritubular vascular system; conversely, tubular destruction, by increasing intraglomerular pressure, induces glomerular injury. Thus, whatever the origin, all forms of chronic kidney disease ultimately damage all four components of the kidney, culminating in what are called *end-stage kidneys*. The functional reserve of the kidney is large, and much damage may occur before there is evident functional impairment. For these reasons, recognizing the early signs and symptoms is of particular clinical importance.

CLINICAL MANIFESTATIONS OF RENAL DISEASES

The clinical manifestations of renal disease can be grouped into reasonably well-defined syndromes. Some are unique to glomerular diseases, and others are present in diseases that affect any one of the components.

- *Azotemia* is a biochemical abnormality that refers to an elevation of blood urea nitrogen (BUN) and creatinine levels, and is related largely to a decreased glomerular filtration rate (GFR). Azotemia is a consequence of many renal disorders, but it also arises from extrarenal disorders. It is a typical feature of both acute and chronic kidney injury. *Prerenal azotemia* is encountered when there is hypoperfusion of the kidneys that impairs renal function in the absence of parenchymal damage. It may be caused by hypotension, excessive fluid losses from any cause, or if the effective intravascular volume is decreased due to shock, volume depletion, congestive heart failure, or cirrhosis of the liver. *Postrenal azotemia* is seen whenever urine flow is obstructed distal to the kidney. Relief of the obstruction is followed by correction of the azotemia.
- When azotemia leads to clinical signs and symptoms associated with biochemical abnormalities, it is termed *uremia*. Uremia is characterized not only by failure of renal excretory function but also by a host of metabolic and endocrine alterations resulting from renal damage. Uremic patients frequently manifest secondary involvement of the gastrointestinal system (e.g., uremic gastroenteritis), peripheral nerves (e.g., peripheral neuropathy), and heart (e.g., uremic fibrinous pericarditis).
- *Nephritic syndrome* is a clinical entity caused by inflammatory glomerular disease and is dominated by the acute onset of either grossly visible hematuria (red blood cells in urine) or microscopic hematuria with dysmorphic red cells and red cell casts on urinalysis, diminished GFR, mild to moderate proteinuria, and hypertension. It is the classic presentation of acute poststreptococcal glomerulonephritis. *Rapidly progressive glomerulonephritis* (RPGN) is a form of nephritic syndrome in which there is rapid decline in GFR (within hours to days).
- *Nephrotic syndrome,* also due to glomerular disease, is characterized by heavy proteinuria (more than 3.5 g/day), hypoalbuminemia, severe edema, hyperlipidemia, and lipiduria (lipid in the urine). The clinical features of nephritis and the nephrotic syndrome are discussed in more detail later.
- *Asymptomatic hematuria or proteinuria,* or a combination of these two, is usually a manifestation of subtle or mild glomerular abnormalities.
- *Acute kidney injury* (previously called acute renal failure) is characterized by rapid decline in GFR (within hours to days) with concurrent dysregulation of fluid and electrolyte balance, and retention of metabolic waste products normally excreted by the kidney including urea and creatinine. In its most severe forms, it is manifested by *oliguria* or *anuria* (reduced or no urine flow). It can result from glomerular, interstitial, vascular, or acute tubular injury (ATI).
- *Chronic kidney disease* (previously called chronic renal failure) is defined as the presence of a diminished GFR that is persistently less than 60 mL/min/1.73 m^2 for at least 3 months, from any cause, and/or persistent albuminuria. It may present with a clinically silent decline in renal excretory function in milder forms or with prolonged symptoms and signs of uremia in more severe

cases. It is the end result of all chronic renal parenchymal diseases.

- *In end-stage renal disease,* the GFR is less than 5% of normal; this is the terminal stage of uremia.
- *Renal tubular defects* are dominated by polyuria (excessive urine formation), nocturia, and electrolyte disorders (e.g., metabolic acidosis). They are the result of diseases that either directly affect tubular structures (e.g., the nephronophthisis) or cause defects in specific tubular functions. The latter can be inherited (e.g., familial nephrogenic diabetes, cystinuria, renal tubular acidosis) or acquired (e.g., lead nephropathy).
- *Urinary tract obstruction* and *renal tumors* have varied clinical manifestations based on the specific anatomic location and nature of the lesion. *Urinary tract infection* is characterized by bacteriuria and pyuria (bacteria and leukocytes in the urine). The infection may be symptomatic or asymptomatic, and it may affect the kidney *(pyelonephritis)* or the bladder *(cystitis)*.
- *Nephrolithiasis (renal stones)* is manifested by spasms of severe pain (renal colic) and hematuria, often with recurrent stone formation.

Chronic kidney disease is estimated to affect 11% of all adults in the United States, with a particular predominance among older adults. It is the end result of a variety of renal diseases, but most commonly diabetes and hypertension, and the major cause of death from renal disease. The evolution from normal renal function to symptomatic chronic kidney injury progresses through a series of stages that are defined by measures of serum creatinine from which estimates of reduction in GFR are derived. Chronic kidney disease causes significant systemic abnormalities, which are listed in Table 20.1.

Table 20.1 Principal Systemic Manifestations of Chronic Kidney Disease and Uremia

Fluid and Electrolytes
Dehydration
Edema
Hyperkalemia
Metabolic acidosis
Calcium Phosphate and Bone
Hyperphosphatemia
Hypocalcemia
Secondary hyperparathyroidism
Renal osteodystrophy
Hematologic
Anemia
Bleeding diathesis
Cardiopulmonary
Hypertension
Congestive heart failure
Cardiomyopathy
Pulmonary edema
Uremic pericarditis
Gastrointestinal
Nausea and vomiting
Bleeding
Esophagitis, gastritis, colitis
Neuromuscular
Myopathy
Peripheral neuropathy
Encephalopathy
Dermatologic
Sallow color
Pruritus
Dermatitis

KEY CONCEPTS

CLINICAL MANIFESTATIONS OF RENAL DISEASES

- Azotemia is the biochemical manifestation of acute or chronic kidney injury and is characterized by elevated BUN or alternately by an elevated serum creatinine. It reflects a reduction in GFR.
- Kidney injury that results in azotemia can be either acute or chronic. Acute kidney injury can be reversible or progress to chronic kidney disease, which is generally irreversible.
- One major manifestation of kidney injury is the nephrotic syndrome, in which injury to the glomerulus results in abnormal filtration, leading to heavy proteinuria, edema, and metabolic disturbances.
- Nephritic syndromes are those in which hematuria, azotemia, hypertension, and sub-nephrotic proteinuria are the major manifestations.
- Diseases involving the tubules and interstitium may have clinical manifestations of the nephritic syndrome, or of specific defects in tubular function, or of acute or chronic kidney disease without more specific defining features.

GLOMERULAR DISEASES

Glomerular diseases constitute some of the major problems in nephrology. Glomeruli may be injured by a variety of factors and in the course of several systemic diseases. Systemic immunologic diseases such as systemic lupus erythematosus (SLE), vascular disorders such as hypertension, metabolic diseases such as diabetes mellitus, and some hereditary conditions such as Fabry disease often affect the glomerulus. These are termed *secondary glomerular diseases*. Disorders in which the kidney is the only or predominant organ involved constitute the various types of *primary glomerulonephritis* or, because some do not have a cellular inflammatory component, *primary glomerulopathy*. However, both the clinical manifestations and glomerular histologic changes in primary and secondary forms can be similar.

In the following sections, we discuss the normal glomerulus and mechanisms of glomerular injury, and then the various types of primary glomerulopathies. We briefly review the secondary forms covered in other parts of this book. Table 20.2 lists the most common forms of glomerulonephritis that have reasonably well-defined morphologic and clinical characteristics. The clinical manifestations of glomerular disease are clustered into the five major glomerular syndromes summarized in Table 20.3. Both the primary glomerulopathies and the systemic diseases affecting the glomerulus can result in these syndromes. Because glomerular diseases are often associated with systemic disorders, mainly diabetes mellitus, SLE, vasculitis, and amyloidosis, in any

Table 20.2 Glomerular Diseases

Primary Glomerulopathies
Acute proliferative glomerulonephritis
Postinfectious
Other
Rapidly progressive (crescentic) glomerulonephritis
Membranous nephropathy
Minimal change disease
Focal segmental glomerulosclerosis
Membranoproliferative glomerulonephritis
Dense deposit disease
IgA nephropathy
Systemic Diseases With Glomerular Involvement
Systemic lupus erythematosus
Diabetes mellitus
Amyloidosis
Goodpasture syndrome
Microscopic polyarteritis/polyangiitis
Granulomatosis with polyangiitis
Henoch-Schönlein purpura
Hereditary Disorders
Alport syndrome
Thin basement membrane nephropathy
Fabry disease

patient with manifestations of glomerular disease it is essential to consider these systemic conditions.

Structure of the Glomerulus

Many clinical manifestations of glomerular diseases result from perturbations of specific components of the glomerular tuft, so before discussing these diseases we describe the key anatomic structures of glomeruli. The glomerulus consists of an anastomosing network of capillaries lined by fenestrated endothelium invested by two layers of epithelial cells (Fig. 20.1). The visceral epithelial cells (commonly referred to as *podocytes*) are incorporated into and become an intrinsic part of the capillary wall, separated from endothelial cells by a basement membrane. The parietal epithelium, situated on the Bowman capsule, lines the urinary space, the cavity in which plasma filtrate first collects.

The glomerular capillary wall is the filtering membrane and consists of the following structures (Fig. 20.2):

- There is a thin layer of fenestrated *endothelial cells,* with each fenestra being about 70 to 100 nm in diameter.

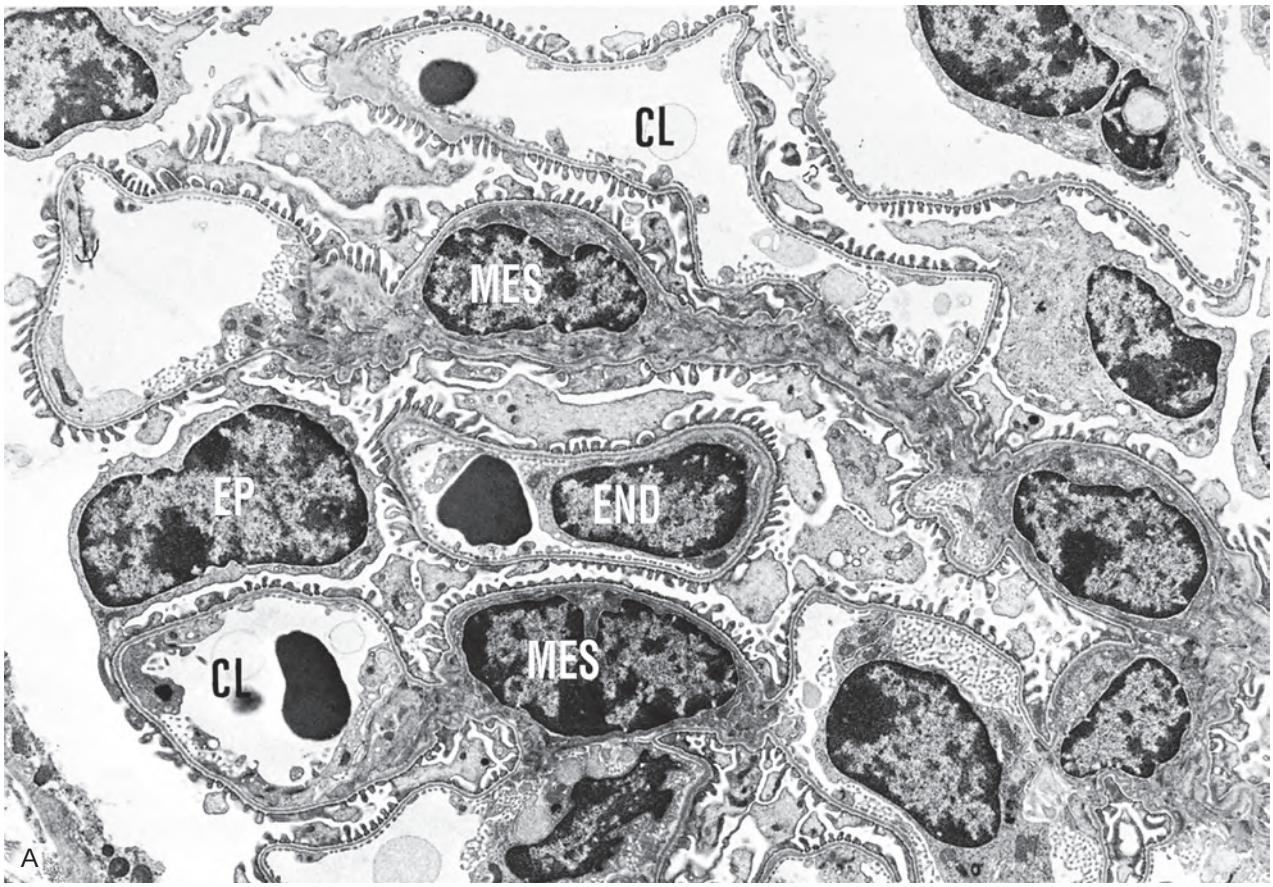
Table 20.3 Clinical Manifestations of Glomerular Diseases

Syndrome	Manifestations
Nephritic syndrome	Hematuria, azotemia, variable proteinuria, oliguria, edema, and hypertension
Rapidly progressive glomerulonephritis	Acute nephritis, proteinuria, and acute renal failure
Nephrotic syndrome	>3.5 g/day proteinuria, hypoalbuminemia, hyperlipidemia, lipiduria
Chronic kidney disease	Azotemia → uremia progressing for months to years
Isolated urinary abnormalities	Glomerular hematuria and/or subnephrotic proteinuria

- A *glomerular basement membrane* (GBM) with a thick electron-dense central layer, the *lamina densa,* and thinner electron-lucent peripheral layers, the *lamina rara interna* and *lamina rara externa*. The GBM consists of collagen (mostly type IV), laminin, polyanionic proteoglycans (mostly heparan sulfate), fibronectin, entactin, and several other glycoproteins. Type IV collagen forms a network suprastructure to which other glycoproteins attach. The building block (monomer) of this network is a triple-helical molecule composed of one or more of six types of α chains (α_1 to α_6 or COL4A1 to COL4A6). Each molecule consists of a 7S domain at the N terminus, a triple-helical domain in the middle, and a globular noncollagenous (NC1) domain at the C terminus. The NC1 domain is important for helix formation and for assembly of collagen monomers into the basement membrane suprastructure. Glycoproteins (laminin, entactin) and proteoglycans (heparan sulfate, perlecan) attach to the collagenous suprastructure. The biochemical properties of these structural components are critical to understanding glomerular diseases. For example, antigens in the NC1 domain are the targets of antibodies in anti-GBM nephritis; genetic defects in the α-chains underlie some forms of hereditary nephritis; and the proteoglycan content of the GBM may contribute to its permeability characteristics.
- The *visceral epithelial cells* (podocytes) possess interdigitating processes embedded in and adherent to the lamina rara externa of the basement membrane (see Fig. 20.1). Adjacent *foot processes* are separated by 20- to 30-nm–wide *filtration slits,* which are bridged by a thin diaphragm (see Fig. 20.2).
- The entire glomerular tuft is supported by the mesangium, the tissue that lies between the capillaries and is composed of basement membrane–like *mesangial matrix* that forms a meshwork in which the *mesangial cells* are embedded (see Fig. 20.1). These cells, of mesenchymal origin, are contractile, phagocytic, and capable of proliferation, of laying down both matrix and collagen, and of secreting several biologically active mediators. Biologically, they are most akin to vascular smooth muscle cells and pericytes. They are important in many forms of glomerulonephritis.

The normal glomerulus is highly permeable to water and small solutes because of the fenestrated nature of the endothelium, and it is impermeable to proteins of the size of albumin (~3.6-nm radius; 70 kilodaltons [kD] molecular weight) or larger. The permeability characteristics of the *glomerular filtration barrier* allow discrimination among various protein molecules, depending on their size (the larger, the less permeable) and charge (the more cationic, the more permeable). This size- and charge-dependent barrier function is accounted for by the structure of the capillary wall. The charge-dependent restriction is important in the virtually complete exclusion of albumin from the filtrate, because albumin is an anionic molecule.

The visceral epithelial cell is important for the maintenance of glomerular barrier function; its slit diaphragm presents a size-selective distal diffusion barrier to the filtration of proteins, and this cell type is largely responsible, under normal circumstances, for synthesis of GBM components. Proteins located in the slit diaphragm or present in assemblies of molecules within visceral epithelial cells



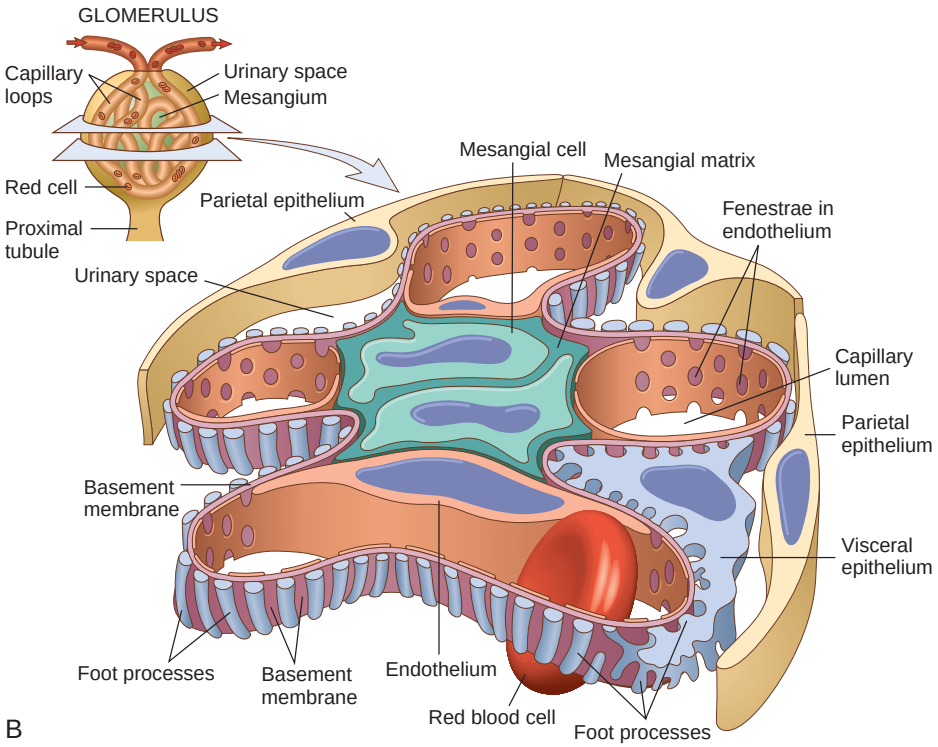


Figure 20.1 (A) Low-power electron micrograph of renal glomerulus. *CL,* Capillary lumen; *EP,* visceral epithelial cells with foot processes; *END,* endothelium; *MES,* mesangium. (B) Schematic representation of a glomerular lobe. (A, Courtesy Dr. Vicki Kelley, Brigham and Women's Hospital, Boston, Mass.)

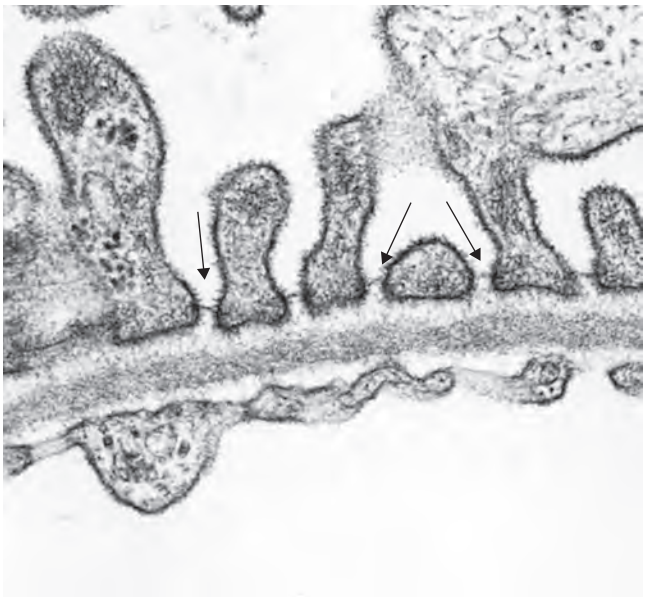


Figure 20.2 Glomerular filter consisting *(from bottom to top)* of fenestrated endothelium, basement membrane, and foot processes of epithelial cells. Note the filtration slits *(arrows)* and diaphragm between the foot processes. Note also that the basement membrane consists of a central lamina densa, sandwiched between two looser layers, the lamina rara interna and lamina rara externa. (Courtesy Dr. Helmut Rennke, Brigham and Women's Hospital, Boston, Mass.)

that are attached to the slit diaphragm are illustrated in Fig. 20.3. Nephrin is a transmembrane protein with a large extracellular portion made up of immunoglobulin (Ig)-like domains. Nephrin molecules extend toward each other from neighboring foot processes and dimerize across the slit diaphragm. Within the cytoplasm of the foot processes, nephrin forms molecular connections with podocin, CD2-associated protein, and ultimately the actin cytoskeleton of the visceral epithelial cells. More slit diaphragm proteins continue to be identified, and comprehensive descriptions of their structure and interactions have been published. The importance of the slit diaphragm proteins in maintaining glomerular permeability is demonstrated by the observation that mutations in the genes encoding them give rise to defects in permeability and the nephrotic syndrome (discussed later).

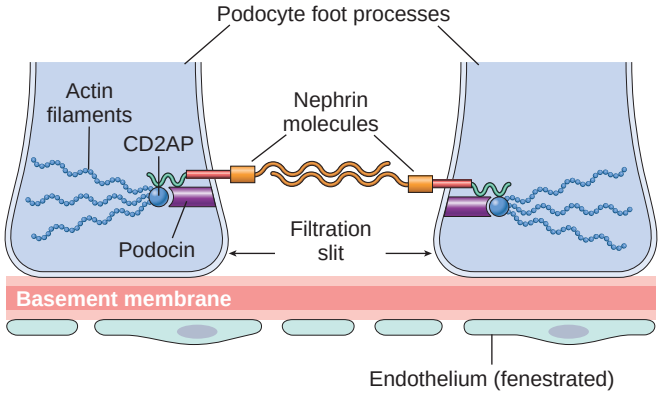


Figure 20.3 A simplified schematic diagram of some of the best-studied proteins of the glomerular slit diaphragm. *CD2AP,* CD2-associated protein.

Pathologic Responses of the Glomerulus to Injury

Various types of glomerulopathies are characterized by one or more of four basic tissue reactions.

Hypercellularity

Some *inflammatory diseases* of the glomerulus are characterized by an increase in the number of cells in the glomerular tufts. This hypercellularity results from one or more of the following:

- *Proliferation* of mesangial or endothelial cells.
- *Infiltration of leukocytes,* including neutrophils, monocytes, and, in some diseases, lymphocytes. The combination of infiltration of leukocytes and swelling and proliferation of mesangial and/or endothelial cells is often referred to as *endocapillary proliferation*.
- *Formation of crescents.* These are accumulations of cells composed of proliferating glomerular epithelial cells and infiltrating leukocytes. The epithelial cell proliferation that characterizes crescent formation occurs following an immune/inflammatory injury involving the capillary walls. Plasma proteins leak into the urinary space, where it is believed that exposure to procoagulants such as tissue factor leads to fibrin deposition. Activation of coagulation factors such as thrombin is suspected of being a trigger for crescent formation, but the actual mechanisms are still poorly understood. Molecules that have been implicated in recruitment of leukocytes into crescents include multiple proinflammatory cytokines.

Basement Membrane Thickening

By light microscopy, this change appears as thickening of the capillary walls, best seen in sections stained with periodic acid–Schiff (PAS). By electron microscopy, such thickening takes one of three forms:

- Deposition of amorphous electron-dense material, most often immune complexes, on the endothelial or epithelial side of the basement membrane or within the GBM itself. Fibrin, amyloid, cryoglobulins, and abnormal fibrillary proteins may also deposit in the GBM.
- Increased synthesis of the protein components of the basement membrane, as occurs in diabetic glomerulosclerosis.
- Formation of additional layers of basement membrane matrices, which most often occupy subendothelial locations and may range from poorly organized matrix to fully duplicated lamina densa, as occurs in membranoproliferative glomerulonephritis (MPGN).

Hyalinosis and Sclerosis

Hyalinosis, as applied to the glomerulus, denotes the accumulation of material that is homogeneous and eosinophilic by light microscopy. *Hyalin* is an extracellular, amorphous material composed of plasma proteins that have insudated from the circulation into glomerular structures. When extensive, these deposits may obliterate the capillary lumens of the glomerular tuft. Hyalinosis is usually a consequence of endothelial or capillary wall injury and typically the end result of various forms of glomerular damage.

Sclerosis is characterized by deposition of extracellular collagenous matrix. It may be confined to mesangial areas, as is often the case in diabetic glomerulosclerosis, involve

the capillary loops, or both. The sclerosing process may also result in obliteration of some or all of the capillary lumens in affected glomeruli.

Many primary glomerulopathies are classified by their histology, as seen in Table 20.2. The histologic changes can be further subdivided by their distribution into the following categories: *diffuse,* involving all of the glomeruli in the kidney; *global,* involving the entirety of individual glomeruli; *focal,* involving only a fraction of the glomeruli in the kidney; *segmental,* affecting a part of each glomerulus; and *capillary loop* or *mesangial,* affecting predominantly capillary or mesangial regions.

KEY CONCEPTS

INJURY OF GLOMERULAR STRUCTURES

- The GBM is composed of type IV collagen molecules and other matrix proteins. These proteins can be the target of antibodies in some types of glomerulonephritis; genetic abnormalities in their composition are the basis for some forms of hereditary nephritis.
- Visceral epithelial cells (podocytes) are a critical component of the glomerular filtration barrier, and their injury leads to protein leakage into the urinary space (proteinuria).
- The acute glomerular response to injury includes hypercellularity with proliferation of mesangial and/or endothelial cells, influx of leukocytes, and, when severe, formation of crescents.
- Chronic glomerular responses to injury include basement membrane thickening, hyalinosis, and sclerosis.

Pathogenesis of Glomerular Injury

Although much remains unknown about etiologic agents and triggering events, it is clear that **immune mechanisms underlie most forms of primary glomerulopathy and many secondary glomerular disorders** (Table 20.4). Glomerulonephritis can be induced experimentally by antigen-antibody reactions. Furthermore, glomerular deposits of immunoglobulins, often with components of complement, are found in the majority of individuals with glomerulonephritis.

Table 20.4 Immune Mechanisms of Glomerular Injury

Antibody-Mediated Injury
In Situ Immune Complex Deposition
Fixed intrinsic tissue antigens
NC1 domain of type IV collagen antigen (anti-GBM nephritis)
PLA_2R antigen (membranous glomerulopathy)
Mesangial antigens
Others
Planted antigens
Exogenous (infectious agents, drugs)
Endogenous (DNA, nuclear proteins, immunoglobulins, immune complexes, IgA)
Circulating Immune Complex Deposition
Endogenous antigens (e.g., DNA, tumor antigens)
Exogenous antigens (e.g., infectious products)
Cell-Mediated Immune Injury
Activation of Alternative Complement Pathway

GBM, Glomerular basement membrane.

Cell-mediated immune reactions also may play a role, usually in concert with antibody-mediated events. We begin this discussion with a review of antibody-instigated injury.

Two forms of antibody-associated injury have been established: (1) injury by *antibodies reacting in situ within the glomerulus,* either binding to insoluble fixed (intrinsic) glomerular antigens or extrinsic molecules planted within the glomerulus, and (2) injury resulting from *deposition of circulating antigen-antibody complexes in the glomerulus.* It is clear that the major cause of glomerulonephritis resulting from formation of antigen-antibody complexes is the consequence of in situ immune complex formation, and not deposition of circulating complexes as was once thought.

Diseases Caused by In Situ Formation of Immune Complexes

In this form of injury, immune complexes are formed locally by antibodies that react with intrinsic tissue antigens or with extrinsic antigens "planted" in the glomerulus from the circulation. Membranous nephropathy is the classic example of glomerular injury resulting from local formation of immune complexes by antibodies reactive with endogenous antigens.

The pattern of immune deposition by immunofluorescence microscopy is granular, reflecting the very localized antigen-antibody interaction. On electron microscopy, the glomerulopathy is characterized by the presence of numerous discrete subepithelial electron-dense deposits (made up of immune reactants). These subepithelial complexes, with resultant host responses, can result in a thickened basement membrane appearance by light microscopy; hence the term *membranous nephropathy*.

Antibodies can react in situ with antigens that are not normally present in the glomerulus but are "planted" there. Such antigens may localize in the kidney by interacting with various intrinsic components of the glomerulus. Planted antigens include cationic molecules that bind to anionic components of the glomerulus; DNA, nucleosomes, and other nuclear proteins, which have an affinity for GBM components; bacterial products; large aggregated proteins (e.g., aggregated immunoglobulins), which deposit in the mesangium because of their size; and immune complexes themselves, because they continue to have reactive sites for further interactions with free antibody, free antigen, or complement. There is no dearth of other possible planted antigens, including viral, bacterial, and parasitic products and drugs.

Disease Caused by Antibodies Directed Against Normal Components of the Glomerular Basement Membrane

In anti-GBM antibody–induced glomerulonephritis, antibodies bind to intrinsic antigens homogeneously distributed along the entire length of the GBM, resulting in a diffuse linear pattern of staining for the antibodies by immunofluorescence techniques (Fig. 20.4B and E). This contrasts with the granular pattern of immunofluorescence staining corresponding to the discrete immune complexes seen in membranous nephropathy, or other glomerular diseases in which large complexes of antigens and antibodies form in situ. Although anti-GBM antibody–induced glomerulonephritis accounts for less than 5% of cases of human glomerulonephritis, it causes severe necrotizing and crescentic glomerular damage and the clinical syndrome of RPGN.

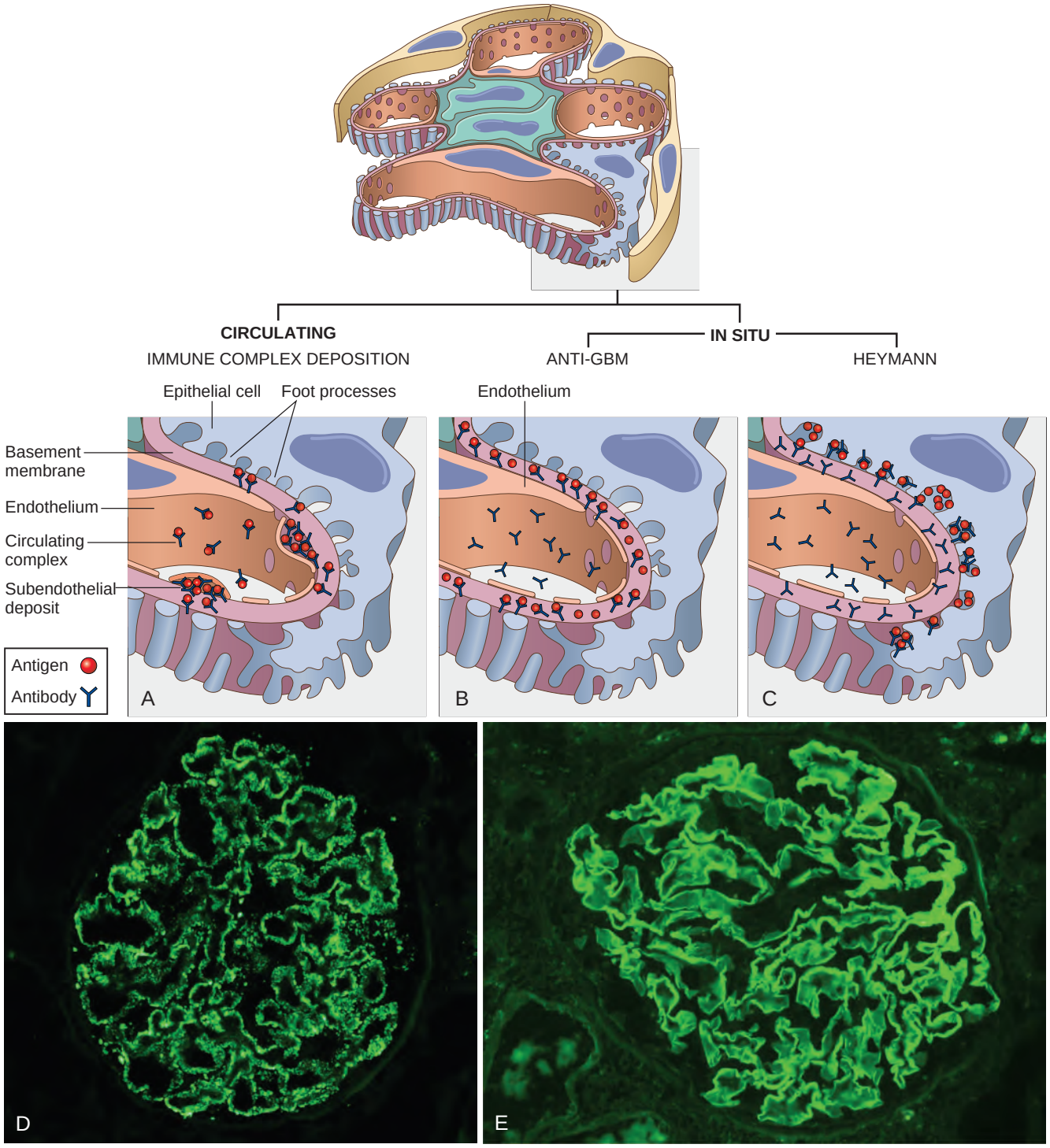


Figure 20.4 Antibody-mediated glomerular injury can result either from the deposition of circulating immune complexes (A) or, more commonly, from in situ formation of complexes exemplified by anti–glomerular basement membrane (anti-GBM) disease (B) or Heymann nephritis (C). (D and E) Two patterns of deposition of immune complexes as seen by immunofluorescence microscopy: granular, characteristic of circulating and in situ immune complex nephritis (D), and linear, characteristic of classic anti-GBM disease (E). (D, Courtesy Dr. J. Kowalewska, Department of Pathology, University of Washington, Seattle, Washington. E, From Kumar V, Abbas AK, Aster JC: *Robbins Basic Pathology,* ed 10, Philadelphia, 2018, Elsevier.)

Glomerulonephritis Resulting From Deposition of Circulating Immune Complexes

In this type of nephritis, glomerular injury is caused by the trapping of circulating antigen-antibody complexes within glomeruli. The antibodies have no immunologic specificity for glomerular constituents, and the complexes localize within the glomeruli because of their physicochemical properties and the hemodynamic factors peculiar to the glomerulus (see Fig. 20.4A).

The antigens that trigger the formation of circulating immune complexes may be of endogenous origin, as in the glomerulonephritis associated with SLE or in IgA nephropathy, **or they may be exogenous,** as may occur in the glomerulonephritis that follows certain infections. Microbial antigens that are implicated include bacterial products (streptococcal proteins), the surface antigen of hepatitis B virus, hepatitis C virus antigens, and antigens of *Treponema pallidum, Plasmodium falciparum,* and several viruses. Some tumor antigens are also thought to cause immune complex–mediated nephritis. In many cases, the inciting antigen is unknown.

Mechanisms of Glomerular Injury Following Immune Complex Formation

The pathogenesis of immune complex diseases is discussed in Chapter 6. Here we briefly review the salient features that relate to glomerular injury. **Whatever the antigen may be, antigen-antibody complexes formed or deposited in the glomeruli may elicit a local inflammatory reaction that produces injury.** The antibodies may activate complement and engage Fc receptors on leukocytes and perhaps glomerular mesangial or other cells, leading to inflammation. The glomerular lesions may exhibit leukocytic infiltration and proliferation of mesangial and endothelial cells.

Electron microscopy reveals electron-dense deposits, presumably containing immune complexes, that may lie in the mesangium, between the endothelial cells and the GBM (subendothelial deposits), or between the outer surface of the GBM and the podocytes (subepithelial deposits). Deposits may be located at more than one site in a given case. By immunofluorescence microscopy, the immune complexes are seen as granular deposits along the basement membrane (see Fig. 20.4D), in the mesangium, or in both locations. Once deposited in the kidney, immune complexes may eventually be degraded, mostly by infiltrating neutrophils and monocytes/macrophages, mesangial cells, and endogenous proteases, and the inflammatory reaction may then subside. Such a course occurs when the exposure to the inciting antigen is short-lived and limited, as in most cases of poststreptococcal glomerulonephritis. However, if immune complexes are deposited repeatedly for prolonged periods, as may be seen in SLE or viral hepatitis, many cycles of injury may occur, leading to a more chronic membranous or membranoproliferative type of glomerulonephritis.

Several factors affect glomerular localization of antigen, antibody, or immune complexes. The molecular charge and size of these reactants are clearly important. Highly cationic antigens tend to cross the GBM, and the resultant complexes eventually reside in a subepithelial location. Highly anionic macromolecules are excluded from the GBM and are trapped subendothelially or are not nephritogenic at all. Molecules of neutral charge and immune complexes containing these molecules tend to accumulate in the mesangium. Large circulating complexes are not usually nephritogenic, because they are cleared by the mononuclear phagocyte system and do not enter the GBM in significant quantities. The pattern of localization is also affected by changes in glomerular hemodynamics, mesangial function, and integrity of the charge-selective barrier in the glomerulus. These influences may underlie the variable pattern of immune-reactant deposition in various forms of glomerulonephritis (Fig. 20.5). In turn, the distinct patterns of localization of immune complexes is a key determinant of the injury response and the histologic features that subsequently develop. Immune complexes located in subendothelial portions of capillaries and in mesangial regions are accessible to the circulation and more likely to be involved in inflammatory processes that require interaction and activation of circulating leukocytes. Diseases in which immune complexes are confined to the subepithelial locations and for which the capillary basement membranes may be a barrier to interaction with circulating leukocytes, as in the case of membranous nephropathy, typically have a noninflammatory pathology.

In summary, **most cases of immune complex–mediated glomerulonephritis are a consequence of deposition of discrete immune complexes, which give rise to granular immunofluorescence staining along the basement membranes or in the mesangium.** However, it may be difficult to determine whether the deposition has occurred in situ, by circulating complexes, or by both mechanisms

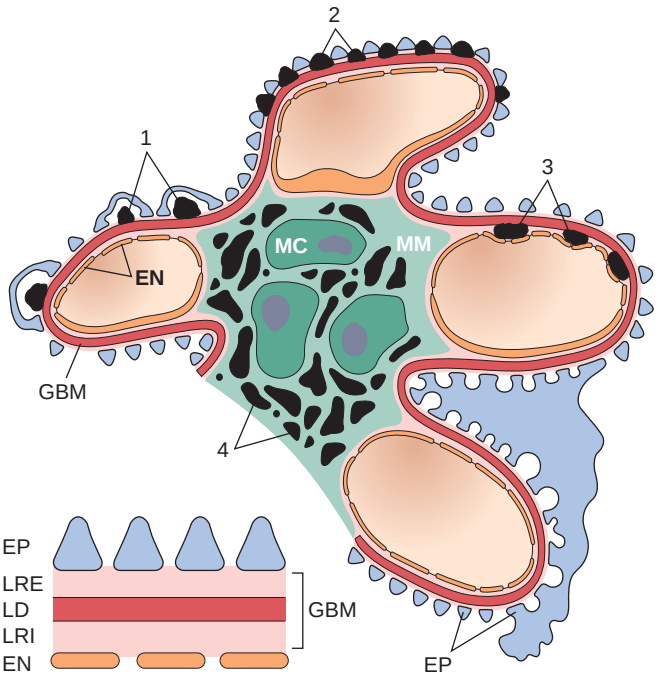


Figure 20.5 Localization of immune complexes in the glomerulus: (1) subepithelial humps, as in acute glomerulonephritis; (2) membranous deposits, as in membranous nephropathy and Heymann nephritis; (3) subendothelial deposits, as in lupus nephritis and membranoproliferative glomerulonephritis; (4) mesangial deposits, as in IgA nephropathy. *EN,* Endothelium; *EP,* epithelium; *GBM,* glomerular basement membrane; *LD,* lamina densa; *LRE,* lamina rara externa; *LRI,* lamina rara interna; *MC,* mesangial cell; *MM,* mesangial matrix. (Modified from Couser WG: Mediation of immune glomerular injury, *J Am Soc Nephrol* 1[1]: 13–29, 1990.)

because, as discussed earlier, trapping of circulating immune complexes can initiate further in situ complex formation. Single etiologic agents, such as hepatitis B and C viruses, can cause either a membranous pattern of glomerulonephritis, with subepithelial in situ deposition of antigens, or a membranoproliferative pattern, more indicative of subendothelial deposition of antigens or deposition of circulating complexes. It is best then to consider that **antigen-antibody deposition in the glomerulus is a major pathway of glomerular injury and that in situ immune reactions, trapping of circulating complexes, interactions between these two events, and local hemodynamic and structural determinants in the glomerulus all contribute to the diverse morphologic and functional alterations in glomerulonephritis.**

Cell-Mediated Immunity in Glomerulonephritis

Although antibody-mediated mechanisms may initiate most forms of glomerulonephritis, there is evidence that sensitized T cells cause glomerular injury and are involved in the progression of some glomerulonephritides. Clues to the role of cellular immunity include the presence of activated macrophages and T cells and their products in the glomerulus in some forms of human and experimental glomerulonephritis; in vitro and in vivo evidence of lymphocyte activation on exposure to antigen in human and experimental glomerulonephritis; abrogation of glomerular injury by lymphocyte depletion; and experiments in which glomerular injury may be induced by transfer of T cells from nephritic animals to normal recipients. The evidence is most compelling for certain types of experimental crescentic glomerulonephritis in which antibodies to GBM initiate glomerular injury and activated T lymphocytes propagate the inflammation. Despite this body of suggestive evidence, proof of glomerulonephritis in humans resulting primarily from T-cell activation remains lacking.

Activation of Alternative Complement Pathway

Alternative complement pathway activation occurs in the clinicopathologic entity called *dense-deposit disease,* until recently referred to as *membranoproliferative glomerulonephritis (MPGN type II),* and in *C3 glomerulopathies.* This is discussed later in the sections describing these diseases.

Mediators of Glomerular Injury

Once immune reactants or sensitized T cells have localized in the glomerulus, how does the glomerular damage ensue? The mediators—both cells and molecules—are the usual suspects involved in acute and chronic inflammation, described in Chapter 3, and only a few are highlighted here (Fig. 20.6).

Cells

- *Neutrophils* and *monocytes* infiltrate glomeruli in certain types of glomerulonephritis, largely as a result of activation of complement, resulting in generation of chemotactic agents (mainly C5a), but also by Fc receptor–mediated activation. Neutrophils release proteases, which cause GBM degradation; oxygen-derived free radicals, which cause cell damage; and arachidonic acid metabolites, which contribute to the reductions in GFR.
- *Macrophages and T lymphocytes,* which infiltrate glomeruli in antibody- and cell-mediated reactions, when activated, release a vast number of biologically active molecules.
- *Platelets* may aggregate in glomeruli during immune-mediated injury. Their release of eicosanoids, growth factors, and other mediators may contribute to vascular injury and proliferation of glomerular cells. Antiplatelet agents have beneficial effects in both human and experimental glomerulonephritis.
- *Resident glomerular cells,* particularly mesangial cells, can be stimulated to produce inflammatory mediators, including reactive oxygen species (ROS), cytokines, chemokines,

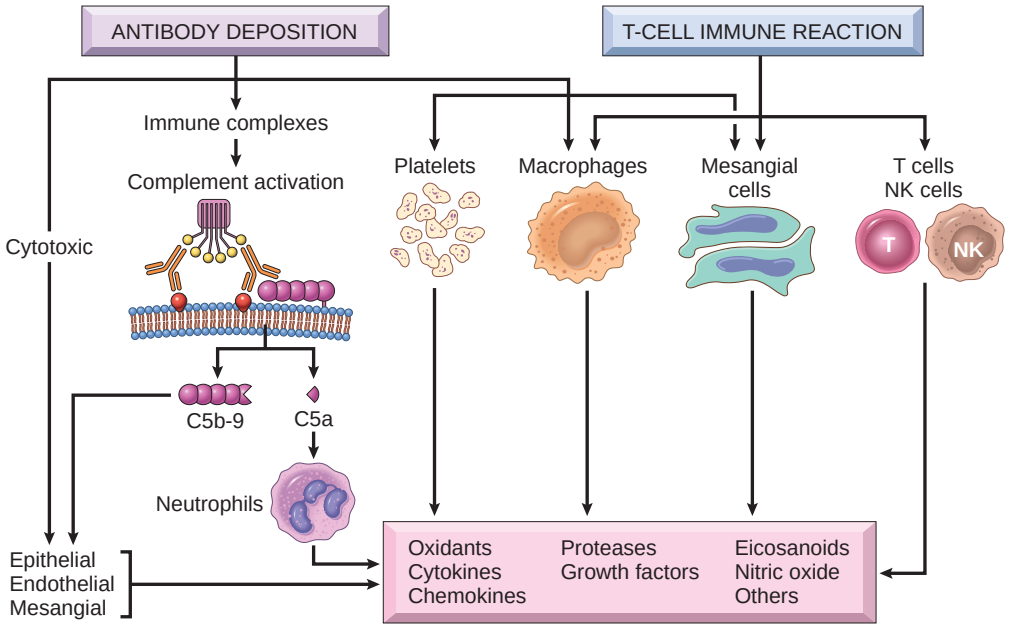


Figure 20.6 Mediators of immune glomerular injury.

growth factors, eicosanoids, nitric oxide, and endothelin. They may initiate inflammatory responses in glomeruli, even in the absence of leukocytic infiltration.

Soluble Mediators

Virtually all the known inflammatory chemical mediators (Chapter 3) have been implicated in glomerular injury.

- *Complement activation* leads to the generation of chemotactic products that induce leukocyte influx (complement-neutrophil–dependent injury) and the formation of C5b–C9, the membrane attack complex. C5b–C9 causes cell lysis but, in addition, it may stimulate mesangial cells to produce oxidants, proteases, and other mediators. Thus, even in the absence of neutrophils, C5b–C9 can cause proteinuria, as has been demonstrated in experimental membranous glomerulopathy. In some diseases, collectively called *C3 glomerulopathies,* there is evidence of complement activation resulting not from antibody or immune complex deposition but from defective regulation of the complement system. As expected, because antibodies are not involved, in these disorders complement activation is by the alternative pathway. The end result is the same as in classical pathway activation.
- *Eicosanoids, nitric oxide, angiotensin,* and *endothelin* are involved in the hemodynamic changes.
- *Cytokines,* particularly IL-1 and TNF, which may be produced by infiltrating leukocytes and resident glomerular cells, induce leukocyte adhesion and a variety of other effects.
- *Chemokines* such as monocyte chemoattractant protein 1 promote monocyte and lymphocyte influx. *Growth factors* such as platelet-derived growth factor (PDGF) are involved in mesangial cell proliferation. TGF-β, connective tissue growth factor, and fibroblast growth factor seem to be critical in the ECM deposition and hyalinization leading to glomerulosclerosis in chronic injury. Vascular endothelial growth factor (VEGF) seems to maintain endothelial integrity and may help regulate capillary permeability.
- The *coagulation system* is also a mediator of glomerular damage. Fibrin is frequently present in the glomeruli and Bowman space in glomerulonephritis, indicative of coagulation cascade activation, and activated coagulation factors, particularly thrombin, may be a stimulus for crescent formation.

Epithelial Cell Injury

Podocyte injury is common to many forms of both primary and secondary glomerular diseases, of both immune and nonimmune etiologies. The term *podocytopathy* has been applied to diseases with disparate etiologies whose principal manifestation is injury to podocytes. This can be induced by antibodies to podocyte antigens; by toxins, as in an experimental model of proteinuria induced by puromycin aminonucleoside; conceivably by certain cytokines; by certain viral infections such as human immunodeficiency virus (HIV); or by incompletely characterized circulating factors, postulated in minimal change disease and focal segmental glomerulosclerosis (FSGS). Such injury is reflected by morphologic changes in podocytes, which include effacement of foot processes, vacuolization, and retraction and detachment of cells from the GBM, and functionally by proteinuria (Fig. 20.7).

Loss of podocytes, which have only a very limited capacity for replication and repair, may be a feature of multiple types of glomerular injury, including FSGS and diabetic nephropathy. Such loss typically cannot be recognized in pathologic specimens unless morphometric techniques are applied. In most forms of glomerular injury, loss of normal slit diaphragms is a key event in the development of proteinuria (see Fig. 20.7). Functional abnormalities of the slit diaphragm may also result from mutations in its components, such as nephrin and podocin, without actual inflammatory damage to the glomerulus. Such mutations are the cause of rare hereditary forms of the nephrotic syndrome.

Mechanisms of Progression in Glomerular Diseases

Thus far, the immunologic mechanisms and mediators that *initiate* glomerular injury have been discussed. The outcome of such injury depends on several factors, including the severity of renal damage, the nature and persistence of the antigens, and the immune status, age, and genetic predisposition of the host.

It has long been known that **once any renal disease, glomerular or otherwise, destroys functioning nephrons**

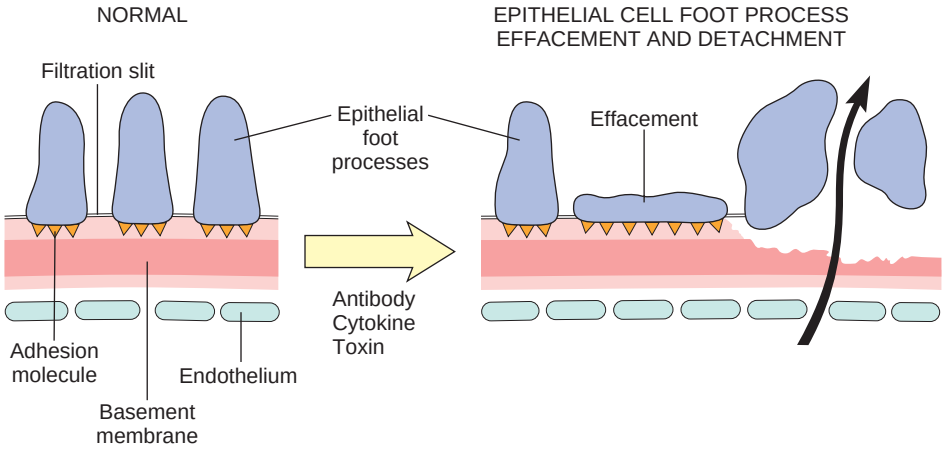


Figure 20.7 Epithelial cell injury. The postulated sequence is a consequence of antibodies specific to epithelial cell antigens, toxins, cytokines, or other factors causing injury; this results in foot process effacement and sometimes detachment of epithelial cells and protein leakage through defective glomerular basement membrane and filtration slits.

and reduces the GFR to about 30% to 50% of normal, progression to end-stage renal failure proceeds at a steady rate, independent of the original insult or activity of the underlying disease. The secondary factors that lead to progression are of great clinical interest because they can be targets of therapy that delays or even prevents the inexorable journey to dialysis or transplantation.

The two major histologic characteristics of such progressive renal damage are *glomerulosclerosis* and *tubulointerstitial fibrosis*.

Glomerulosclerosis

Sclerosis involving portions of some glomeruli (also referred to as secondary FSGS develops after many types of renal injury and can lead to proteinuria and increasing functional impairment. Glomerular sclerosis may be seen even in cases in which the primary disease was nonglomerular. The mechanisms of glomerulosclerosis in this setting are described in the discussion of FSGS later in this chapter.

Tubular Injury and Interstitial Fibrosis

Tubulointerstitial injury, manifested by tubular damage and interstitial inflammation, is a component of many acute and chronic glomerulonephritides. Tubulointerstitial fibrosis contributes to progression in both immune and nonimmune glomerular diseases, for example, diabetic nephropathy. Indeed, **there is often a much better correlation of decline in renal function with the extent of tubulointerstitial damage than with the severity of glomerular injury.** Many factors may lead to such tubulointerstitial injury, including ischemia of tubular segments downstream from sclerotic glomeruli, acute and chronic inflammation in the adjacent interstitium, and damage or loss of the peritubular capillary blood supply. It appears that proteinuria also can cause *direct injury to and activation of tubular cells*. Activated tubular cells in turn express adhesion molecules and elaborate proinflammatory cytokines, chemokines, and growth factors that contribute to interstitial fibrosis. Filtered proteins that may produce these tubular effects include cytokines, complement products, the iron in hemoglobin, immunoglobulins, lipid moieties, and oxidatively modified plasma proteins.

KEY CONCEPTS

PATHOGENESIS OF GLOMERULAR INJURY AND PROGRESSION OF GLOMERULAR DISEASE

- Antibody-mediated injury is an important mechanism of glomerular damage, mainly via complement- and leukocyte-mediated pathways. Antibodies may also be directly cytotoxic to glomerular cells.
- The most common forms of antibody-mediated glomerulonephritis are caused by the formation of immune complexes, which may involve either endogenous antigens (e.g., PLA_2R in membranous nephropathy) or exogenous antigens (e.g., microbial). Immune complexes show a granular pattern of deposition by immunofluorescence.
- Autoantibodies against components of the GBM are the cause of anti–GBM antibody–mediated disease, often associated with severe injury. The immunofluorescence pattern of antibody deposition is linear.
- Activation of the alternative complement pathway is an important mechanism of injury in C3 glomerulopathies that include dense deposit disease and C3 glomerulonephritis.
- Soluble inflammatory mediators, such as cytokines, chemokines, growth factors, eicosanoids, nitric oxide, and activated coagulation factors, also contribute to the glomerular injury.
- Epithelial cell (podocyte) injury induced by antibodies, toxins, cytokines, infections, and poorly characterized circulating factors is a common manifestation of various forms of glomerular diseases.
- Progressive glomerular injury can be the result of either primary or secondary glomerular injuries, of diseases that are either renal limited or systemic, and of diseases that initially involve renal structures other than glomeruli.
- Progressive injury ensues from a cycle of glomerular and nephron loss, compensatory changes that lead to further glomerular injury and glomerulosclerosis, and eventually ESRD.
- Progressive glomerular injury is accompanied by chronic injuries to other renal structures, typically manifest as tubulointerstitial fibrosis.

Having discussed the factors involved in the initiation and progression of glomerular injury, we now turn to a discussion of individual glomerular diseases. Table 20.5 summarizes the main clinical and pathologic features of the major forms of primary glomerulopathies.

Nephritic Syndrome

Glomerular diseases presenting with a nephritic syndrome are characterized by inflammation in the glomeruli. The main clinical features of nephritic syndrome include the following:

- *Hematuria* (red blood cells and red cell casts in urine)
- *Proteinuria* (usually subnephrotic range) with or without edema
- *Azotemia*
- *Hypertension*

Nephritic syndrome is the typical clinical presentation of most proliferative types of GN such as postinfectious GN, crescentic GN, and proliferative lupus GN. The lesions that cause the nephritic syndrome have in common proliferation of the cells within the glomeruli, often accompanied by an inflammatory leukocytic infiltrate. This inflammatory reaction severely injures the capillary walls, permitting blood to pass into the urine and inducing hemodynamic changes that lead to a reduction in GFR. The reduced GFR is manifested clinically by oliguria, fluid retention, and azotemia. Hypertension probably is a result of both the fluid retention and renin release from the ischemic kidneys.

Acute Proliferative (Postinfectious and Infection-Associated) Glomerulonephritis

As the name implies, **this cluster of diseases is characterized histologically by diffuse proliferation of glomerular cells associated with influx (exudation) of leukocytes, typically caused by immune complexes.** The inciting antigen may be exogenous or endogenous. The prototypic exogenous antigen-induced disease pattern is postinfectious

Table 20.5 Summary of Major Primary Glomerulonephritides

Disease	Most Frequent Clinical Presentation	Pathogenesis	Glomerular Pathology		
			Light Microscopy	Fluorescence Microscopy	Electron Microscopy
Postinfectious glomerulonephritis	Nephritic syndrome	Immune complex mediated; circulating or planted antigen	Diffuse endocapillary proliferation; leukocytic infiltration	Granular IgG and C3 in GBM and mesangium; Granular IgA in some cases	Primarily subepithelial humps; subendothelial deposits in early disease stages
Crescentic (rapidly progressive) glomerulonephritis	Nephritic syndrome; rapid progression	Anti-GBM antibody mediated; immune complex mediated; ANCA mediated and unknown	Extracapillary proliferation with crescents; necrosis	Linear IgG and C3 in anti-GBM antibody mediated GN; granular IgG, other Igs, and/or complement in immune complex mediated GN; or no deposits in ANCA mediated GN	No deposits in anti-GBM and ANCA mediated GN; immune complexes at various locations in immune complex mediated GN
Membranous nephropathy	Nephrotic syndrome	In situ immune complex PLA_2R antigen in most cases of primary disease	Diffuse capillary wall thickening	Granular IgG and C3; diffuse	Subepithelial deposits
Minimal change disease	Nephrotic syndrome	Unknown; loss of glomerular polyanion; podocyte injury	Normal; lipid in tubules	Negative	Effacement of foot processes; no deposits
Focal segmental glomerulosclerosis	Nephrotic syndrome; non-nephrotic proteinuria	Unknown Ablation nephropathy Plasma factor (?); podocyte injury	Focal and segmental sclerosis and hyalinosis	Focal; IgM + C3 in many cases	Effacement of foot processes; epithelial denudation
Membranoproliferative glomerulonephritis (MPGN) type I	Nephritic/ nephrotic syndrome	Immune complex	Mesangial proliferative or membranoproliferative patterns of proliferation; GBM thickening; splitting	IgG ++ C3; C1q ++ C4	Subendothelial deposits
Dense-deposit disease (MPGN type II)	Hematuria Chronic renal failure	Acquired or genetic dysregulation of the alternative complement pathway	Mesangial proliferative or membranoproliferative patterns of proliferation; GBM thickening; splitting	C3; no C1q or C4	Dense deposits
IgA nephropathy	Recurrent hematuria or proteinuria	Unknown	Focal mesangial proliferative glomerulonephritis; mesangial widening	IgA ± IgG, IgM, and C3 in mesangium	Mesangial and paramesangial dense deposits

ANCA, Antineutrophil cytoplasmic antibodies; *GBM*, glomerular basement membrane.

glomerulonephritis, whereas an example of an endogenous antigen-induced disease is the nephritis of SLE, described in Chapter 6. The most common underlying infections are streptococcal, but the disorder may also be associated with other infections. Poststreptococcal GN is decreasing in frequency in the United States but continues to be a fairly common disorder worldwide. It usually appears 1 to 4 weeks after a streptococcal infection of the pharynx or skin (impetigo). Skin infections are commonly associated with overcrowding and poor hygiene. Poststreptococcal glomerulonephritis occurs most frequently in children 6 to 10 years of age, but children and adults of any age can also be affected.

Pathogenesis

Poststreptococcal GN is caused by immune complexes containing streptococcal antigens and specific antibodies. Only certain strains of group A β-hemolytic streptococci are nephritogenic, more than 90% of cases being traced to types 12, 4, and 1, which can be identified by typing of the M protein of the bacterial cell walls.

Many lines of evidence support an immunologic basis for poststreptococcal glomerulonephritis. The latent period between infection and onset of nephritis is compatible with the time required for the production of antibodies and the formation of immune complexes. Elevated titers of antibodies

against one or more streptococcal antigens are present in a great majority of patients. Serum complement levels are low, compatible with activation of the complement system and consumption of complement components. There are granular immune deposits in the glomeruli, indicative of an immune complex–mediated mechanism.

The streptococcal antigenic component responsible for the immune reaction had long eluded identification, but the preponderance of evidence identifies streptococcal pyogenic exotoxin B (SpeB) as the principal antigenic determinant in most but not all cases of poststreptococcal glomerulonephritis. This protein can directly activate complement, is commonly secreted by nephritogenic strains of streptococci, and has been localized to the "humplike" deposits characteristic of this disease (described later). At the outset, the inciting antigens are exogenously planted from the circulation in subendothelial locations in glomerular capillary walls, leading to in situ formation of immune complexes, where they elicit an inflammatory response. Subsequently, through mechanisms that are not well understood, the antigen-antibody complexes dissociate, migrate across the GBM, and reform on the subepithelial side of the GBM. Deposition of circulating immune complexes also may contribute to the lesions. A similar form of glomerulonephritis occurs sporadically in association with other infections, including those of bacterial (e.g., staphylococcal endocarditis, pneumococcal pneumonia, and meningococcemia), viral (e.g., hepatitis B, hepatitis C, mumps, HIV infection, varicella, and infectious mononucleosis), and parasitic (malaria, toxoplasmosis) origin. In these settings, granular immunofluorescent deposits and subepithelial humps characteristic of immune complex nephritis are also present.

MORPHOLOGY

The classic histologic picture is one of **enlarged, hypercellular glomeruli** (Fig. 20.8B). The hypercellularity is caused by (1) infiltration by leukocytes, both neutrophils and monocytes; (2) proliferation of endothelial and mesangial cells; and (3) in severe cases by crescent formation. The proliferation and leukocyte infiltration are typically global and diffuse, that is, involving all lobules of all glomeruli. There is also swelling of endothelial cells, and the combination of proliferation, swelling, and leukocyte

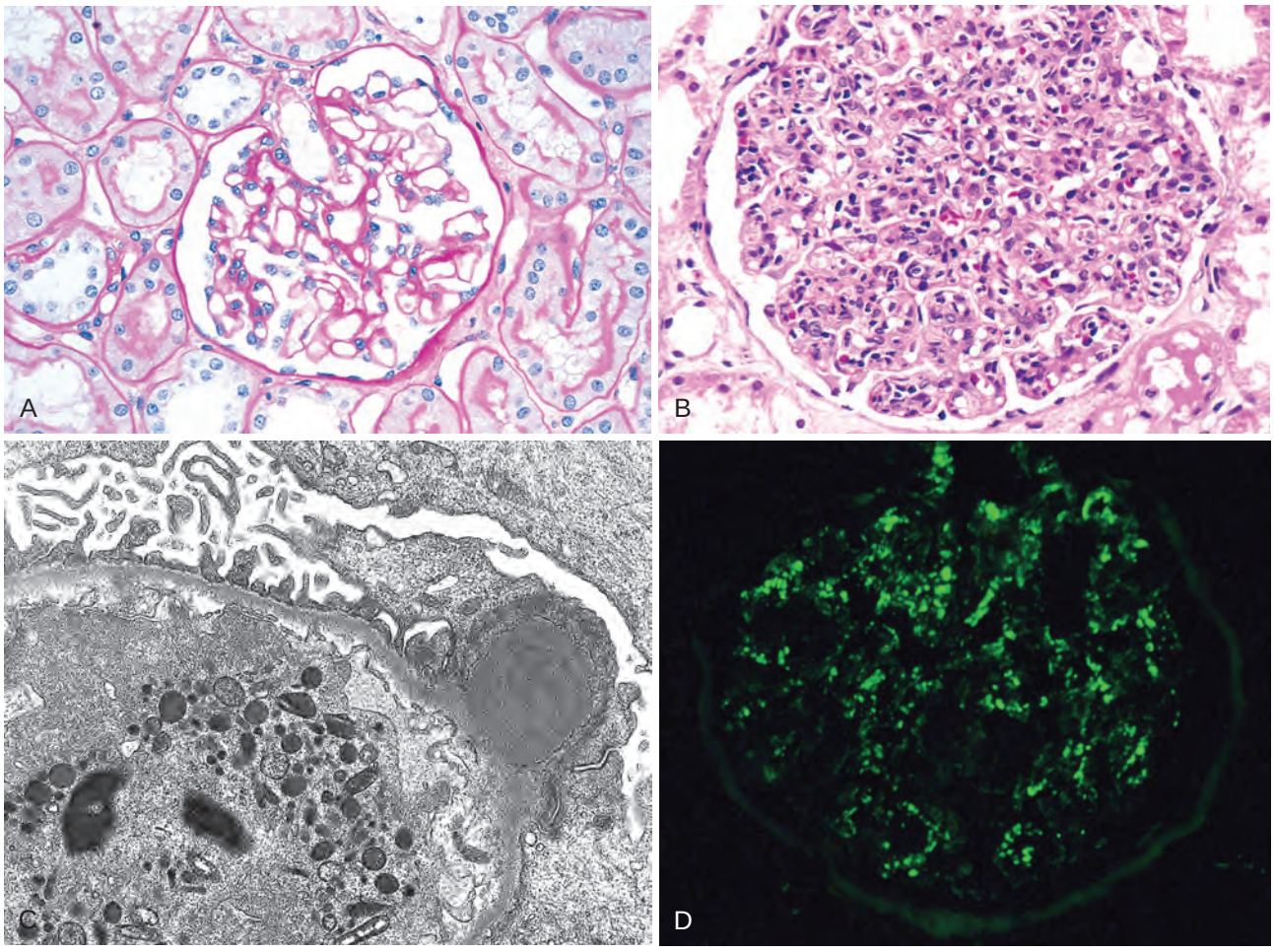


Figure 20.8 Acute proliferative glomerulonephritis. (A) Normal glomerulus. (B) Glomerular hypercellularity is due to intracapillary leukocytes and proliferation of intrinsic glomerular cells. (C) Typical electron-dense subepithelial "hump" and a neutrophil in the lumen. (D) Immunofluorescent stain demonstrates discrete, coarsely granular deposits of complement protein C3 (stain for IgG was similar), corresponding to "humps" illustrated in part C. (*A–C,* Courtesy Dr. H. Rennke, Brigham and Women's Hospital, Boston, Mass. *D,* Courtesy D. J. Kowalewska, Cedars-Sinai Medical Center, Los Angeles, Calif.)

infiltration obliterates the capillary lumens. There may be interstitial edema and inflammation, and the tubules often contain red cell casts.

By **immunofluorescence microscopy,** there are granular deposits of IgG and C3, and sometimes IgM in the mesangium and along the GBM (Fig. 20.8D). Although immune complex deposits are almost universally present, they are often focal and sparse. The characteristic **electron microscopic findings** are discrete, amorphous, electron-dense deposits on the epithelial side of the membrane, often having the appearance of "humps" (Fig. 20.8C), presumably representing the antigen-antibody complexes at the subepithelial cell surface. Subendothelial deposits are also commonly seen, typically early in the disease course, and mesangial and intramembranous deposits may be present.

Clinical Features

In the typical case, a young child abruptly develops malaise, fever, nausea, oliguria, and hematuria (smoky or cola-colored urine) 1 to 2 weeks after recovery from a sore throat. The patients have dysmorphic red cells or red cell casts in the urine, mild proteinuria (usually less than 1 g/day), periorbital edema, and mild to moderate hypertension. In adults the onset is more likely to be atypical, such as the sudden appearance of hypertension or edema, frequently with elevation of BUN. The glomerulonephritis is subclinical in some infected individual, and is discovered only on screening for microscopic hematuria carried out during epidemic outbreaks. Important laboratory findings include elevations of antistreptococcal antibody titers and a decline in the serum concentration of C3 and other components of the complement cascade.

More than 95% of affected children eventually recover renal function with conservative therapy aimed at maintaining sodium and water balance. A small minority of children (perhaps less than 1%) do not improve, become severely oliguric, and develop a rapidly progressive form of glomerulonephritis (described later). Some of the remaining patients may undergo slow progression to chronic glomerulonephritis with or without recurrence of an active nephritic picture. Prolonged and persistent heavy proteinuria and abnormal GFR mark patients with an unfavorable prognosis.

In adults the disease is less benign. Although the overall prognosis in epidemics is good, in only about 60% of sporadic cases do the patients recover promptly. In the remainder, the glomerular lesions fail to resolve quickly, as manifested by persistent proteinuria, hematuria, and hypertension. In some of these patients, the lesions eventually clear, but others develop chronic glomerulonephritis or even RPGN.

Crescentic (Rapidly Progressive) Glomerulonephritis

RPGN is a clinical syndrome associated with severe glomerular injury, but it does not denote a specific etiology. It is characterized by relatively rapid and progressive loss of renal function associated with severe oliguria and signs of nephritic syndrome; if untreated, death from renal failure may occur within weeks to months. *The most common histologic picture is the presence of crescents in most of the glomeruli, hence the name crescentic glomerulonephritis.* As discussed earlier, these are produced predominantly by the proliferation of the epithelial cells lining the Bowman capsule and by the infiltration of monocytes and macrophages.

Pathogenesis

Crescentic GN may be a manifestation of a number of different diseases, some restricted to the kidney and others systemic. Although no single mechanism can explain all cases, there is little doubt that **in most cases the glomerular injury is immunologically mediated.** A practical classification divides crescentic GN into three groups on the basis of immunologic findings (Table 20.6), which have been described as follows:

- *Anti-GBM antibody–mediated disease, characterized by linear deposits of IgG and, in many cases, C3 in the GBM.* In some of these patients, the anti-GBM antibodies cross-react with pulmonary alveolar basement membranes to produce the clinical picture of pulmonary hemorrhage associated with renal failure *(Goodpasture syndrome)*. Plasmapheresis to remove the pathogenic circulating antibodies is usually part of the treatment, which also includes therapy to suppress the underlying immune response.

  The antigen common to the alveoli and GBM is a peptide within the noncollagenous portion of the α_3 chain of collagen type IV. What triggers the formation of these autoantibodies is unclear in most patients. Exposure to viruses or hydrocarbon solvents (found in paints and dyes) has been implicated in some patients, as have various drugs and cancers. There is a high prevalence of certain HLA alleles (e.g., HLA-DRB1) in affected patients, a finding consistent with the genetic predisposition to autoimmunity.
- *Diseases caused by immune complex deposition, with granular deposits of antibodies and complement by immunofluorescence.* Crescentic GN can be a complication of any of the immune complex nephritides, including postinfectious glomerulonephritis, lupus nephritis, IgA nephropathy, and Henoch-Schönlein purpura. These patients usually cannot be helped by plasmapheresis, and they require treatment for the underlying disease.
- *Pauci-immune crescentic GN, defined by the lack of detectable anti-GBM antibodies or immune complexes by immunofluorescence and electron microscopy.* Most patients with this type of RPGN have circulating antineutrophil cytoplasmic

Table 20.6 Crescentic (Rapidly Progressive) Glomerulonephritides

Type I (Anti-GBM Antibody)
Renal limited
Goodpasture syndrome
Type II (Immune Complex)
Idiopathic
Postinfectious glomerulonephritis
Lupus nephritis
Henoch-Schönlein purpura
IgA nephropathy
Others
Type III (Pauci-Immune)
ANCA-associated
Idiopathic
Granulomatosis with polyangiitis (formerly Wegener granulomatosis)
Microscopic polyangiitis

ANCA, Antineutrophil cytoplasmic antibodies; *GBM*, glomerular basement membrane.

antibodies (ANCAs) that produce a cytoplasmic (c-ANCA) or perinuclear (p-ANCA) staining pattern and are known to play a role in some vasculitides (Chapter 11). This type of RPGN may be a component of a systemic vasculitis such as granulomatosis with polyangiitis (formerly called Wegener granulomatosis) or microscopic polyangiitis. In many cases, however, pauci-immune crescentic glomerulonephritis is limited to the kidneys and hence idiopathic. More than 90% of such idiopathic cases have c-ANCAs (now called PR3-ANCA because the antibodies are specific for the neutrophil granule protein proteinase-3) or p-ANCAs (now called MPO-ANCA indicating reactivity with neutrophil myeloperoxidase) in the serum. The presence of circulating ANCAs in both idiopathic crescentic glomerulonephritis and cases of crescentic glomerulonephritis that occur as a component of systemic vasculitis, and the similar pathologic features in both settings, have led to the idea that these disorders are pathogenetically related. According to this concept, all cases of crescentic glomerulonephritis of the pauci-immune type are manifestations of small-vessel vasculitis or polyangiitis, which is limited to glomerular and perhaps peritubular capillaries in cases of idiopathic crescentic glomerulonephritis. ANCAs have proved to be invaluable as a highly sensitive diagnostic marker for pauci-immune crescentic glomerulonephritis, but proof of their role as a direct cause of this glomerulonephritis has been elusive. Recent evidence of their pathogenic potential has been obtained by studies in mice showing that transferring antibodies against myeloperoxidase (the target antigen of most p-ANCAs) induces a form of RPGN.

To summarize, about one-fifth of patients with RPGN have anti-GBM antibody–mediated glomerulonephritis without lung involvement; another one-fourth have immune complex–mediated crescentic glomerulonephritis; and the remainder are of the pauci-immune type.

MORPHOLOGY

The kidneys are enlarged and pale, often with petechial hemorrhages on the cortical surfaces. Depending on the underlying cause, the glomeruli often show focal and segmental necrosis, and variably show diffuse or focal endothelial proliferation, and mesangial proliferation. Segmental glomerular necrosis and distinctive **crescents** (Fig. 20.9) adjacent to glomerular segments uninvolved by inflammatory or proliferative changes is the feature most typical of pauci-immune RPGN. Crescents are formed by proliferation of glomerular epithelial cells and by migration of monocytes and macrophages into the urinary space. Neutrophils and lymphocytes may be present. The crescents may obliterate the urinary space and compress the glomerular tuft. **Fibrin strands are frequently prominent between the cellular layers in the crescents;** indeed, as discussed earlier, the escape of procoagulant factors, fibrin, and cytokines into Bowman space may contribute to crescent formation. By immunofluorescence microscopy, immune complex–mediated cases show granular immune deposits; Goodpasture syndrome cases show linear GBM fluorescence for Ig and complement, and pauci-immune cases have little or no deposition of immune reactants. Electron microscopy discloses deposits in those cases due to immune complex deposition. Regardless of type, electron microscopy may show **ruptures in the GBM,** a severe injury that allows leukocytes, plasma proteins such as coagulation factors and complement, and inflammatory mediators to reach the urinary space, where they trigger crescent formation (Fig. 20.10). In time, most crescents undergo organization and foci of segmental necrosis resolve as segmental scars (a type of segmental sclerosis), but restoration of normal glomerular architecture may be achieved with early aggressive therapy.

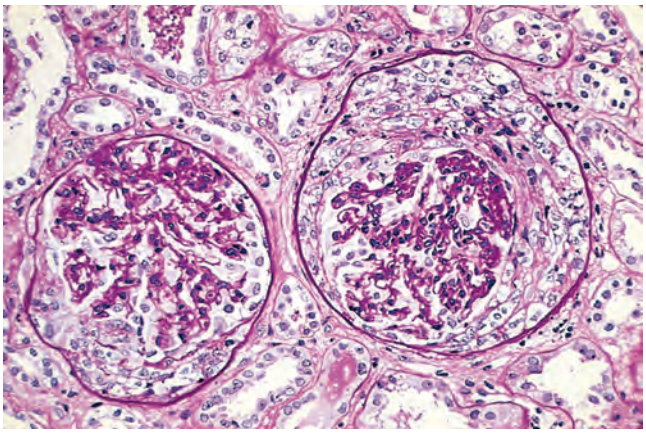


Figure 20.9 Crescentic glomerulonephritis (periodic acid–Schiff stain). Note the compressed glomerular tufts and the crescent-shaped mass of proliferating epithelial cells and leukocytes within the Bowman capsule. (Courtesy Dr. M. A. Venkatachalam, University of Texas Health Sciences Center, San Antonio, Tex.)

Clinical Features

The renal manifestations of all forms of crescentic glomerulonephritis include hematuria with red blood cell casts in the urine, moderate proteinuria occasionally reaching the

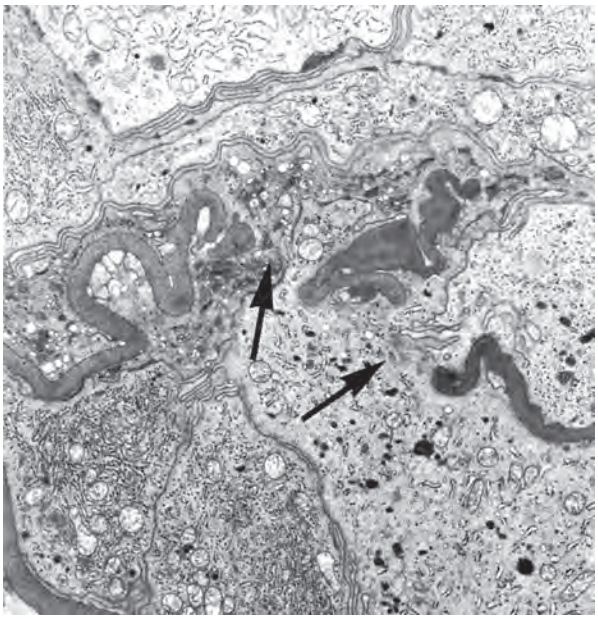


Figure 20.10 Crescentic glomerulonephritis. Electron micrograph showing characteristic wrinkling of glomerular basement membrane with focal disruptions *(arrows)*.

nephrotic range, and variable hypertension and edema. In Goodpasture syndrome, the course may be dominated by recurrent hemoptysis or even life-threatening pulmonary hemorrhage. Serum analyses for anti-GBM antibodies, antinuclear antibodies, and ANCAs are helpful in the diagnosis of specific subtypes. Although milder forms of glomerular injury may subside, the renal involvement is usually progressive over a matter of weeks and culminates in severe oliguria. Recovery of renal function may follow early intensive plasmapheresis (plasma exchange) combined with steroids and cytotoxic agents in Goodpasture syndrome. This therapy can reverse both pulmonary hemorrhage and renal failure. Other forms of RPGN also respond well to steroids and cytotoxic agents. However, despite therapy, many patients eventually require chronic dialysis or transplantation, particularly if the disease is discovered at a late stage.

KEY CONCEPTS

THE NEPHRITIC SYNDROME

- The nephritic syndrome is characterized by hematuria, oliguria with azotemia, proteinuria, and hypertension.
- The most common causes are immunologically mediated glomerular injury; lesions are characterized by proliferative changes and leukocyte infiltration.
- Acute postinfectious glomerulonephritis typically occurs after streptococcal infection in children and young adults but may occur following infection with many other organisms; it is caused by deposition of immune complexes, mainly in the subepithelial spaces, with abundant neutrophils and proliferation of glomerular cells. Most affected children recover; the prognosis is worse in adults.
- RPGN is a clinical entity with features of the nephritic syndrome and rapid loss of renal function.
- RPGN is commonly associated with severe glomerular injury with necrosis and GBM breaks and subsequent proliferation of parietal epithelial cells forming crescents, also called crescentic GN.
- Crescentic GN may be antibody mediated, caused by autoantibodies to the GBM, or as a result of immune complex deposition. It can also occur in the absence of significant antibody deposition; most patients with this type of RPGN have circulating ANCAs.

Nephrotic Syndrome

Certain glomerular diseases virtually always produce the nephrotic syndrome. In addition, many other forms of primary and secondary glomerulopathies discussed in this chapter may underlie the syndrome. Before the major diseases associated with nephrotic syndrome are presented, the causes and pathophysiology of this clinical complex are briefly discussed.

Pathophysiology

Nephrotic syndrome is caused by a derangement in glomerular capillary walls resulting in increased permeability to plasma proteins. The manifestations of the syndrome include the following:

- *Massive proteinuria,* with the daily loss of 3.5 g or more of protein (less in children)
- *Hypoalbuminemia,* with plasma albumin levels less than 3 g/dL
- *Generalized edema*
- *Hyperlipidemia and lipiduria*

The various components of nephrotic syndrome bear a logical relationship to one another. The glomerular capillary wall, with its endothelium, GBM, and visceral epithelial cells, acts as a size and charge barrier through which the plasma filtrate passes. Increased permeability resulting from either structural or physicochemical alterations in this barrier allows proteins to escape from the plasma into the urinary space, resulting in proteinuria.

Heavy proteinuria depletes serum albumin levels at a rate beyond the compensatory synthetic capacity of the liver, resulting in hypoalbuminemia. Increased renal catabolism of filtered albumin also contributes to the hypoalbuminemia. The generalized edema is a direct consequence of *decreased intravascular colloid osmotic pressure*. There is also *sodium and water retention,* which aggravates the edema (Chapter 4). This seems to be due to several factors, including compensatory secretion of aldosterone, mediated by the hypovolemia-enhanced renin secretion; stimulation of the sympathetic system; and a reduction in the secretion of natriuretic factors such as atrial peptides. Edema is characteristically soft and pitting, and it is most marked in the periorbital regions and dependent portions of the body. If severe, it may also lead to pleural effusions and ascites.

The largest proportion of protein lost in the urine is albumin, but globulins are also excreted in some diseases. The ratio of low- to high-molecular-weight proteins in the urine in various cases of nephrotic syndrome is a manifestation of the selectivity of proteinuria. A highly selective proteinuria consists mostly of low-molecular-weight proteins (albumin, 70 kD; transferrin, 76 kD molecular weight), whereas a poorly selective proteinuria consists of higher-molecular-weight globulins in addition to albumin.

The genesis of the hyperlipidemia is complex. Most patients with nephrotic syndrome have increased blood levels of cholesterol, triglyceride, very-low-density lipoprotein, low-density lipoprotein, lipoprotein(a), and apoprotein, and there is a decrease in high-density lipoprotein concentration in some patients. These defects seem to be due to a combination of increased synthesis of lipoproteins in the liver, abnormal transport of circulating lipid particles, and decreased lipid catabolism. Lipiduria follows the hyperlipidemia, because lipoproteins also leak across the glomerular capillary wall. The lipid appears in the urine either as free fat or as *oval fat bodies,* representing lipoprotein resorbed by tubular epithelial cells and then shed along with injured tubular cells that have detached from the basement membrane.

Nephrotic patients are particularly vulnerable to *infection,* especially staphylococcal and pneumococcal infections, probably due to loss of immunoglobulins in the urine. *Thrombotic and thromboembolic complications* are also common in nephrotic syndrome, due in part to loss of endogenous anticoagulants (e.g., antithrombin III) in the urine. *Renal vein thrombosis,* once thought to be a cause of nephrotic syndrome, is most often a consequence of this hypercoagulable state, particularly in patients with membranous nephropathy (see later).

Pathogenesis

The incidences of the several causes of the nephrotic syndrome vary according to age and geography. In children younger than 17 years of age in North America, for example, nephrotic syndrome is almost always caused by a lesion primary to the kidney; among adults, in contrast, it is often associated with a systemic disease. Table 20.7 represents a composite derived from several studies of the causes of the nephrotic syndrome and is therefore only approximate. The most frequent *systemic causes* of the nephrotic syndrome are diabetes, amyloidosis, and SLE. The most important of the *primary glomerular lesions* are minimal change disease, membranous nephropathy, and FSGS. The first is most common in children in North America, the second is most common in older adults, and FSGS occurs at all ages. These three lesions are discussed individually in the following sections. Other less common causes of nephrotic syndrome include the various proliferative glomerulonephritides such as MPGN and IgA nephropathy.

Membranous Nephropathy

Membranous nephropathy is characterized by diffuse thickening of the glomerular capillary wall due to the accumulation of deposits containing Ig along the subepithelial side of the basement membrane. Approximately 75% of cases of membranous nephropathy are primary. The remaining cases occur in association with other systemic diseases and have identifiable etiologic agents, and hence are referred to as secondary membranous nephropathy. The most notable of these associations are as ollows:

- *Drugs* (penicillamine, captopril, gold, nonsteroidal antiinflammatory drugs [NSAIDs]). From 1% to 7% of patients with rheumatoid arthritis treated with penicillamine or gold (drugs now used infrequently for this disease) develop membranous nephropathy.
- *Underlying malignant tumors,* particularly carcinomas of the lung and colon, and melanoma. According to some studies, these are present in as many as 5% to 10% of adults with membranous nephropathy.
- *SLE.* About 10% to 15% of glomerulonephritis in SLE is of the membranous type.
- *Infections* (chronic hepatitis B, hepatitis C, syphilis, schistosomiasis, malaria)
- *Other autoimmune disorders* such as thyroiditis can be associated with secondary membranous nephropathy.

Table 20.7 Cause of Nephrotic Syndrome

Causes	Approximate Prevalence (%)[a]	
	Children	Adults
Primary Glomerular Disease		
Membranous nephropathy	3	30
Minimal change disease	75	8
Focal segmental glomerulosclerosis	10	35
Membranoproliferative glomerulonephritis and dense deposit disease[b]	10	10
Other proliferative glomerulonephritides (focal, "pure mesangial," IgA nephropathy)[b]	2	17
Systemic Diseases		
Diabetes mellitus		
Amyloidosis		
Systemic lupus erythematosus		
Drugs (nonsteroidal antiinflammatory, penicillamine, heroin)		
Infections (malaria, syphilis, hepatitis B and C, HIV)		
Malignant disease (carcinoma, lymphoma)		
Miscellaneous (bee-sting allergy, hereditary nephritis)		

[a]Approximate prevalence of primary disease = 95% of nephrotic syndrome in children, 60% in adults. Approximate prevalence of systemic disease = 5% in children, 40% in adults.

[b]Membranoproliferative and other proliferative glomerulonephritides may result in mixed nephrotic/nephritic syndromes.

Pathogenesis

Membranous nephropathy is a form of chronic immune complex–mediated disease. In secondary membranous nephropathy, the inciting antigens can sometimes be identified in the immune complexes. The antigens may be endogenous or exogenous. The endogenous antigens may be renal or nonrenal (described later). Membranous nephropathy in SLE is associated with deposition of complexes of self nuclear proteins and autoantibodies. Exogenous antigens include those derived from hepatitis B virus and *Treponema pallidum* in patients infected with these microbes.

Primary (formerly idiopathic) membranous nephropathy is an *autoimmune disease caused in most cases by antibodies to a renal autoantigen.* A major recent advance came from the identification of the M-type phospholipase A_2 receptor (PLA_2R) as the antigen that underlies 60% to 70% of human membranous nephropathy. Autoantibody binding to PLA_2R, a membrane protein at the basal surface of the glomerular epithelial cell, is followed by complement activation and then shedding of the immune aggregates from the cell surface to form characteristic deposits of immune complexes along the *subepithelial aspect* of the basement membrane (see Fig. 20.4C). Less often, the target antigen is thrombospondin type-1 domain–containing 7A (THSD7A) or neutral endopeptidase (CD10), a membrane protein recognized by placentally transferred maternal antibodies in cases of neonatal membranous nephropathy. The lesions bear a striking resemblance to those of experimental Heymann nephritis, which is induced by antibodies to the megalin antigenic complex present in the rat podocyte, the antigenic counterpart of the human PLA_2R.

How does the glomerular capillary wall become leaky in membranous nephropathy? There is a paucity of neutrophils, monocytes, or platelets in glomeruli. The virtually uniform presence of complement and corroborating experimental work suggest that the complement C5b–C9 membrane attack complex has an important role. It is postulated that C5b–C9 activates glomerular epithelial and mesangial cells, inducing them to liberate proteases and oxidants, which cause capillary wall injury and increased protein leakage. A subclass of IgG, IgG4, which differs from other IgG subclasses in being a poor activator of the classical complement pathway, is the principal immunoglobulin deposited in most cases of membranous nephropathy. How IgG4 may activate the complement system is not clear.

MORPHOLOGY

By light microscopy, the glomeruli either appear normal in the early stages of the disease or exhibit **uniform, diffuse thickening of the glomerular capillary wall** (Fig. 20.11A). By electron microscopy, the thickening is seen to be caused by irregular electron-dense deposits containing immune complexes between the basement membrane and the overlying epithelial cells, with effacement of podocyte foot processes (Fig. 20.11B and D). Basement membrane material is laid down between these deposits, appearing as irregular spikes protruding from the GBM. These spikes are best seen by silver stains, which color the basement membrane, but not the deposits, black. In time, these spikes thicken to produce domelike protrusions and eventually close over the immune deposits, burying them within a markedly thickened, irregular membrane. Immunofluorescence microscopy demonstrates that the granular deposits contain both immunoglobulins and complement (Fig. 20.11C). Immunostains also reveal PLA_2R or THSD7A glomerular positivity in the majority of patients; however, a significant proportion of patients with secondary membranous nephropathy can also be positive. As the disease advances, segmental sclerosis may occur; in the course of time, glomeruli may become totally sclerosed. The epithelial cells of the proximal tubules contain protein reabsorption droplets, and there may be considerable interstitial mononuclear cell inflammation.

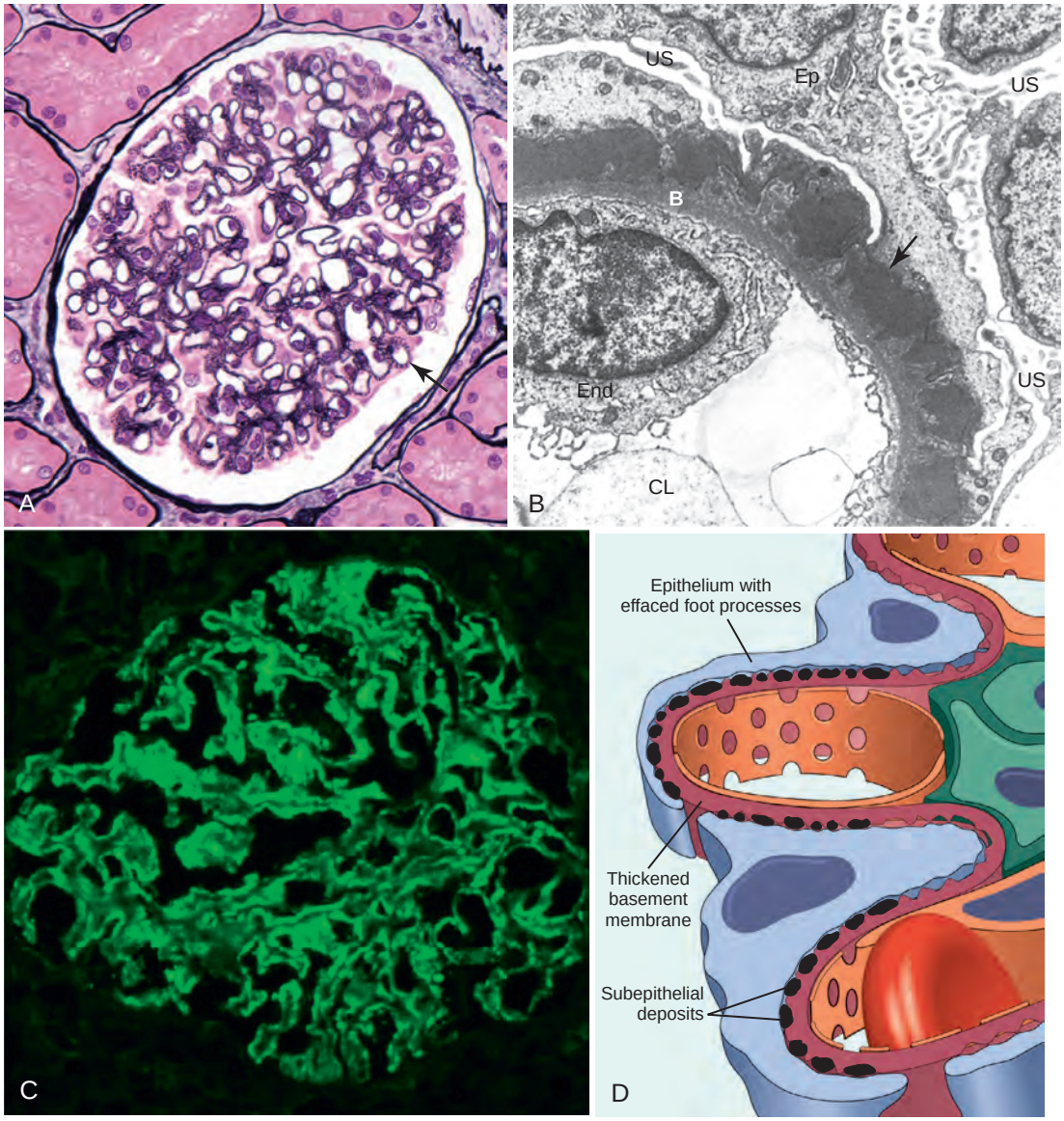


Figure 20.11 Membranous nephropathy. (A) Silver methenamine stain. Note the marked diffuse thickening of the capillary walls without an increase in the number of cells. There are prominent "spikes" of silver-staining matrix *(arrow)* projecting from the basement membrane lamina densa toward the urinary space, which separate and surround deposited immune complexes that lack affinity for the silver stain. (B) Electron micrograph showing electron-dense deposits *(arrow)* along the epithelial side of the basement membrane (B). Note the effacement of foot processes overlying deposits. (C) Characteristic granular immunofluorescent deposits of IgG along the glomerular basement membrane. (D) Diagrammatic representation of membranous nephropathy. *CL,* Capillary lumen; *End,* endothelium; *Ep,* epithelium; *US,* urinary space. (A, Courtesy Dr. Charles Lassman, UCLA School of Medicine, Los Angeles, Calif.)

Clinical Features

This disorder usually presents with the insidious onset of the nephrotic syndrome or, in 15% of patients, with non-nephrotic proteinuria. Hematuria and mild hypertension are present in 15% to 35% of cases. It is necessary in any patient to first rule out the secondary causes described earlier, because treatment of the underlying condition (malignant neoplasm, infection, or SLE) or discontinuance of the offending drug can reverse or ameliorate the injury.

The course of the disease is variable but generally indolent. In contrast to minimal change disease, described later, the proteinuria is nonselective and usually does not respond well to corticosteroid therapy. Complete or partial remissions may occur in up to 40% of patients, even in some patients without therapy. Progression is associated with increasing sclerosis of glomeruli, rising serum creatinine reflecting renal insufficiency, and development of hypertension. Although proteinuria persists in more than 60% of patients, only about 10% die or progress to renal failure within 10 years, and no more than 40% eventually develop severe chronic kidney disease or ESRD. The disease recurs in up to 40% of patients who undergo transplantation for ESRD. Spontaneous remissions and a relatively benign outcome occur more commonly in women and in those with proteinuria in the non-nephrotic range.

Because of the variable course of the disease, it has been difficult to evaluate the overall effectiveness of corticosteroids or other immunosuppressive therapy in controlling the proteinuria or progression. It is thought that circulating antibodies to PLA_2R and THSD7A may be useful biomarkers of disease activity and thereby aid in the diagnosis and management of membranous nephropathy.

Minimal Change Disease

This relatively benign disorder is characterized by diffuse effacement of foot processes of visceral epithelial cells (podocytes), detectable only by electron microscopy, in glomeruli that appear normal by light microscopy. It is the most frequent cause of nephrotic syndrome in children, but it is less common in adults (see Table 20.7). The peak incidence is between 2 and 6 years of age. The disease sometimes follows a respiratory infection or routine prophylactic immunization.

Pathogenesis

Although the absence of immune deposits in the glomerulus excludes immune complex–mediated injury, several features of the disease point to an immunologic basis, including (1) the clinical association with respiratory infections and prophylactic immunization; (2) the response to corticosteroids and/or other immunosuppressive therapy; (3) the association with other atopic disorders (e.g., eczema, rhinitis); (4) the increased prevalence of certain HLA haplotypes in cases associated with atopy (suggesting a genetic predisposition); and (5) the increased incidence of minimal change disease in patients with Hodgkin lymphoma, in whom defects in T cell–mediated immunity are well recognized.

The current leading hypothesis is that minimal change disease involves some immune dysfunction that results in the elaboration of factors that damage visceral epithelial cells and cause proteinuria. Candidate pathogenic factors such as angiopoietin-like-4 have been identified in animal models, but none of these factors have been proven to cause the human disease. The ultrastructural changes point to a primary *visceral epithelial cell injury* (podocytopathy), and studies in animal models suggest the loss of glomerular polyanions. Thus, defects in the charge barrier may contribute to the proteinuria. The actual route by which protein traverses the epithelial cell portion of the capillary wall remains an enigma. Possibilities include transcellular passage through epithelial cells, passage through residual spaces between remaining but damaged foot processes or through abnormal spaces developing underneath the portion of the foot process that directly abuts the basement membrane, or leakage through foci in which the epithelial cells have become detached from the basement.

MORPHOLOGY

The glomeruli are normal by light microscopy (Fig. 20.12A). By electron microscopy the GBM appears normal, and no electron-dense material is deposited. **The principal lesion is in the visceral epithelial cells, which show a uniform and diffuse effacement of foot processes,** these being reduced to a rim of cytoplasm with loss of recognizable intervening slit diaphragms (Fig. 20.12B). This change, often incorrectly termed "fusion" of foot processes, actually represents simplification of the epithelial cell architecture with flattening, retraction, and swelling of foot processes. Foot process effacement is also present in other proteinuric states (e.g., membranous glomerulopathy, diabetic nephropathy); it is only when effacement is associated with normal glomeruli by light microscopy that the diagnosis of minimal change disease can be made. The visceral epithelial changes are completely reversible after corticosteroid therapy, concomitant with remission of the proteinuria. The cells of the proximal tubules are often laden with lipid and protein, reflecting tubular reabsorption of lipoproteins passing through diseased glomeruli (thus, the historical name **lipoid nephrosis** for this disease). Immunofluorescence studies show no Ig or complement deposits.

Clinical Features

Despite massive proteinuria, renal function remains good, and there is commonly no hypertension or hematuria. The proteinuria usually is highly selective, most of the protein being albumin. **A characteristic feature is its usually dramatic response to corticosteroid therapy.** Most children (>90%) with minimal change disease respond rapidly to this treatment. However, proteinuria may recur in a significant proportion of patients, and some patients may become steroid-dependent or resistant. Nevertheless, the long-term prognosis for patients is excellent, and even steroid-dependent disease usually resolves when children reach puberty. Although adults are slower to respond, their long-term prognosis is also excellent.

As has been noted, minimal change disease in adults can be associated with Hodgkin lymphoma and, less frequently, other lymphomas and leukemias. In addition, secondary minimal change disease may follow NSAID therapy, usually in association with acute interstitial nephritis, to be described later in this chapter.

Focal Segmental Glomerulosclerosis

Primary FSGS is the most common cause of nephrotic syndrome in adults in the United States. It is sometimes

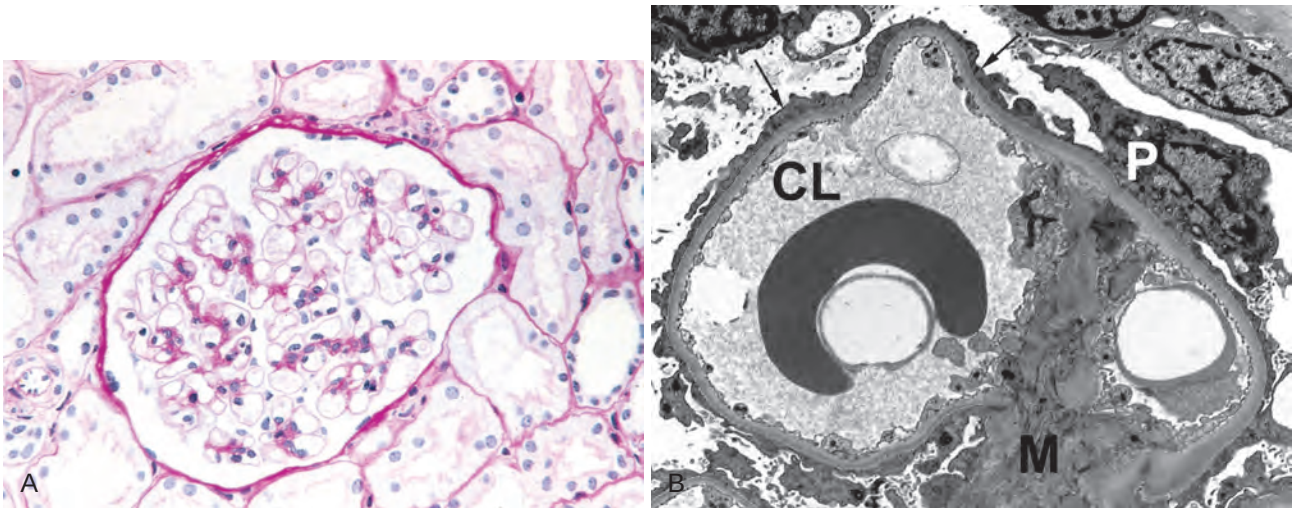


Figure 20.12 Minimal change disease. (A) Glomerulus stained with periodic acid–Schiff. Note normal basement membranes and absence of proliferation. (B) Ultrastructural characteristics of minimal change disease include effacement of foot processes *(arrows)* and absence of deposits. *CL,* Capillary lumen; *M,* mesangium; *P,* podocyte cell body.

considered to be a primary disorder of podocytes, like minimal change disease. As the name implies, this lesion is characterized by sclerosis of some, but not all, glomeruli (thus, it is focal); and in the affected glomeruli, only a portion of the capillary tuft is involved (thus, it is segmental). FSGS is frequently manifest clinically by the acute or subacute onset of nephrotic syndrome or non-nephrotic proteinuria. Hypertension, microscopic hematuria, and some degree of azotemia are commonly present when the disease is first clinically recognized.

Classification and Types

FSGS occurs in the following settings:

- As a primary disease (idiopathic FSGS)
- In association with other known conditions, such as HIV infection (HIV-associated nephropathy), heroin addiction (heroin nephropathy), sickle cell disease, and morbid obesity
- As a secondary event, reflecting scarring of previously active necrotizing lesions, in cases of focal glomerulonephritis (e.g., IgA nephropathy)
- As a component of the adaptive response to loss of renal tissue (renal ablation, described earlier), whether from congenital anomalies (e.g., unilateral renal agenesis or renal dysplasia) or acquired causes (e.g., reflux nephropathy), or in advanced stages of other renal disorders, such as hypertensive nephropathy.
- In uncommon inherited forms of nephrotic syndrome where the disease may be caused by mutations in genes that encode proteins localized to the slit diaphragm, e.g., podocin, α-actinin 4, and TRPC6 (transient receptor potential calcium channel-6)

Idiopathic FSGS accounts for 10% and 35% of cases of nephrotic syndrome in children and adults, respectively. FSGS (both primary and secondary forms) has increased in incidence and is now the most common cause of nephrotic syndrome in adults in the United States, particularly in Hispanic and African-American patients. The clinical signs differ from those of minimal change disease in the following respects: (1) there is a higher incidence of hematuria, reduced GFR, and hypertension; (2) proteinuria is more often nonselective; (3) there is poor response to corticosteroid therapy; and (4) there is progression to chronic kidney disease, with at least 50% developing ESRD within 10 years.

Pathogenesis

The characteristic degeneration and focal disruption of visceral epithelial cells with effacement of foot processes resemble the diffuse epithelial cell change typical of minimal change disease and other podocytopathies. *This epithelial damage is the ultrastructural hallmark of FSGS.* Multiple different mechanisms can cause such epithelial damage, including circulating factors and genetically determined defects affecting components of the slit diaphragm complex. The hyalinosis and sclerosis stem from entrapment of plasma proteins in extremely hyperpermeable foci and increased ECM deposition. The recurrence of proteinuria after transplantation, sometimes within 24 hours, with subsequent progression to overt lesions of FSGS, suggests that an unknown circulating factor is the cause of the epithelial damage in some patients.

The discovery of a genetic basis for some cases of FSGS and other causes of the nephrotic syndrome has improved the understanding of the pathogenesis of proteinuria in the nephrotic syndrome and has provided new methods for diagnosis and prognosis. Inherited forms of FSGS may be associated with the following:

- The first relevant gene to be identified, *NPHS1,* maps to chromosome 19q13 and encodes the protein *nephrin*. Nephrin is a key component of the slit diaphragm (see Fig. 20.3), the structure that controls glomerular permeability. Several mutations of the *NPHS* gene have been identified that give rise to *congenital nephrotic syndrome* of the Finnish type, producing a minimal change disease–like glomerulopathy with extensive foot process effacement.
- A distinctive pattern of autosomal recessive FSGS results from mutations in the *NPHS2* gene, which maps to chromosome 1q25–q31 and encodes the protein product

podocin. Podocin has also been localized to the slit diaphragm. Mutations in *NPHS2* result in a syndrome of steroid-resistant nephrotic syndrome of childhood onset.

- Mutations in the gene encoding the podocyte actin-binding protein α-actinin 4 underlie some cases of autosomal dominant FSGS, which can be insidious in onset but has a high rate of progression to renal insufficiency.
- Mutations in the gene encoding TRPC6 have been found in some kindreds with adult-onset FSGS. This protein is widely expressed, including in podocytes, and the pathogenic mutations may perturb podocyte function by increasing calcium flux in these cells.

What these proteins have in common is their localization to the slit diaphragm and to adjacent podocyte cytoskeletal structures. Their specific functions and interactions are incompletely understood, but it is clear that the integrity of each is necessary to maintain the normal glomerular filtration barrier. Recently, two sequence variants (G1/G2 risk alleles) in the apolipoprotein L1 gene *(APOL1)* on chromosome 22 have been strongly associated with an increased risk of FSGS and renal failure in individuals of African descent, although the mechanisms underlying this association are not yet known. These sequence variants are particularly remarkable because the selective pressures for their conservation in people of African descent is a result of resistance to trypanosome infection conferred by these polymorphisms.

Renal ablation FSGS, a secondary form of FSGS, occurs as a complication of glomerular and nonglomerular diseases causing reduction in functioning renal tissue. Particularly striking examples in which this occurs are reflux nephropathy and unilateral agenesis. These may lead to progressive glomerulosclerosis and renal failure. The glomerulosclerosis seems to be initiated by the adaptive change that occurs in the relatively unaffected glomeruli of diseased kidneys. Such a mechanism is suggested by experiments in rats subjected to subtotal nephrectomy. *Compensatory hypertrophy* of the remaining glomeruli initially maintains renal function in these animals, but proteinuria and segmental glomerulosclerosis soon develop, leading eventually to total glomerular sclerosis and uremia. The glomerular hypertrophy is associated with *hemodynamic changes,* including increases in glomerular blood flow, filtration, and transcapillary pressure (glomerular hypertension), and often with systemic hypertension.

The sequence of events (Fig. 20.13) that is thought to lead to sclerosis in this setting entails endothelial and visceral epithelial cell injury; visceral epithelial cell loss leading to segments of GBM denuded of overlying foot processes and consequent increased glomerular permeability to proteins; and accumulation of proteins in the mesangial matrix. This is followed by proliferation of mesangial cells, infiltration by macrophages, increased accumulation of extracellular matrix (ECM), and segmental and eventually global sclerosis of glomeruli. With increasing reductions in nephron mass and ongoing compensatory changes, a vicious cycle of continuing glomerulosclerosis sets in. Most of the mediators of chronic inflammation and fibrosis, particularly TGF-β, play a role in the development of sclerosis. Currently, the most successful interventions to interrupt these mechanisms of progressive glomerulosclerosis involve treatment with inhibitors of the renin-angiotensin system, which not only reduce intraglomerular hypertension, but also have direct effects on each of the mechanisms identified earlier. Importantly, these agents have been shown to ameliorate progression of sclerosis in both animal and human studies.

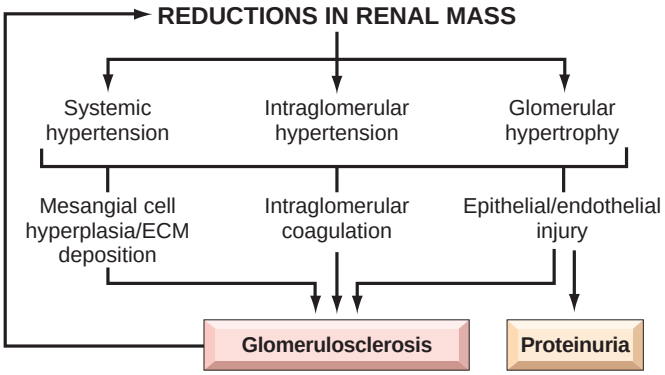


Figure 20.13 Focal segmental glomerulosclerosis associated with loss of renal mass. The adaptive changes in glomeruli (hypertrophy and glomerular capillary hypertension), as well as systemic hypertension, cause epithelial and endothelial injury and resultant proteinuria. The mesangial response, involving mesangial cell proliferation and extracellular matrix production, together with intraglomerular coagulation, causes glomerulosclerosis. This results in further loss of functioning nephrons and a vicious circle of progressive glomerulosclerosis.

MORPHOLOGY

By light microscopy, the **focal and segmental lesions may involve only a minority of the glomeruli** and may be missed if the biopsy specimen contains an insufficient number of glomeruli (Fig. 20.14A). In the sclerotic segments, there is collapse of capillary loops, increase in matrix, and segmental deposition of plasma proteins along the capillary wall (hyalinosis), which may become so pronounced as to occlude capillary lumens. Lipid droplets and foam cells are often present (Fig. 20.14B). Glomeruli that do not show segmental lesions usually appear normal on light microscopy but may show increased mesangial matrix. On electron microscopy, both sclerotic and nonsclerotic areas show **diffuse effacement of foot processes,** and there may also be focal detachment of the epithelial cells and denudation of the underlying GBM. By immunofluorescence microscopy, IgM and C3 may be present in the sclerotic areas and/or in the mesangium. In addition to the focal sclerosis, there may be pronounced hyalinosis and thickening of afferent arterioles. With progression of the disease, increased numbers of glomeruli become involved, and sclerosis spreads within each glomerulus. In time, this leads to total (i.e., global) sclerosis of glomeruli, with pronounced tubular atrophy and interstitial fibrosis.

A morphologic variant of FSGS, called **collapsing glomerulopathy,** is characterized by retraction and/or collapse of the entire glomerular tuft, with or without additional FSGS lesions of the type described earlier (Fig. 20.15). A characteristic feature is proliferation and hypertrophy of glomerular visceral epithelial cells. Collapsing glomerulopathy may be idiopathic, but it also has been associated with some drug toxicities (e.g., pamidronate), and it is the most characteristic lesion of HIV-associated nephropathy. Collapsing glomerulopathy is typically associated with prominent tubular injury with formation of microcysts and prominent interstitial inflammation. It has a particularly poor prognosis.

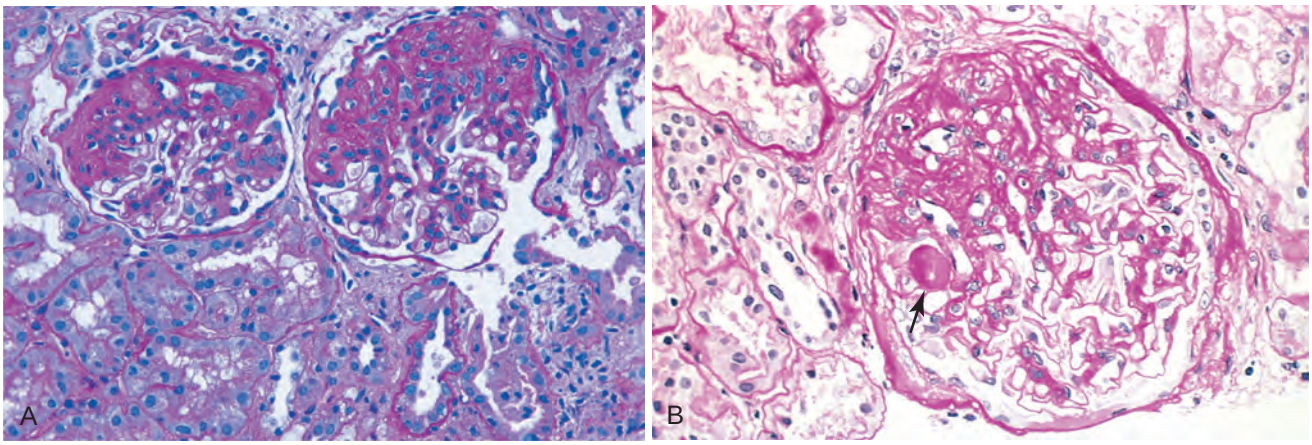


Figure 20.14 Focal segmental glomerulosclerosis, periodic acid–Schiff stain. (A) Segmental sclerosis involves the upper half of both glomeruli. (B) High-power view showing hyaline insudation *(arrow)* and lipid (small vacuoles) in sclerotic area.

Clinical Features

There is little tendency for spontaneous remission in idiopathic FSGS, and responses to corticosteroid therapy are variable. In general, children have a better prognosis than adults do. Progression to renal failure occurs at variable rates. About 20% of patients follow an unusually rapid course, with intractable massive proteinuria ending in renal failure within 2 years. Factors associated with rapid progression include the degree of proteinuria, the degree of renal insufficiency at diagnosis, and histologic subtype (the collapsing variant has an unfavorable course; the tip variant has a relatively good prognosis). Recurrences are seen in 25% to 50% of patients receiving allografts.

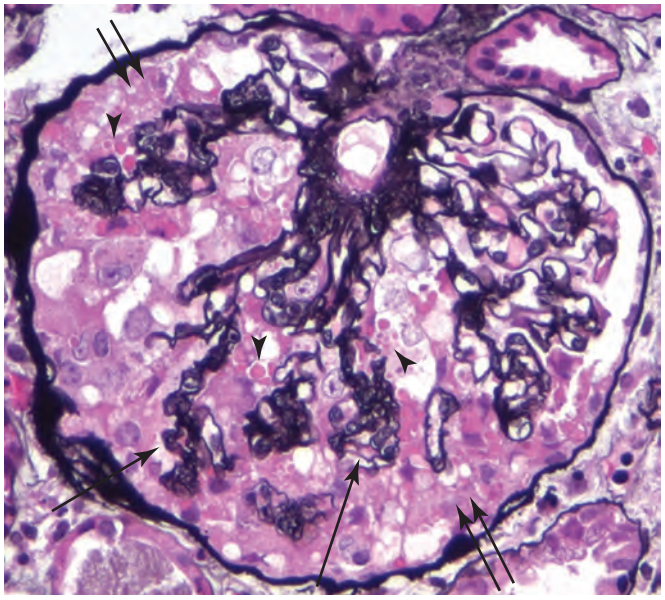


Figure 20.15 Collapsing glomerulopathy. Visible are retraction of the glomerular tuft *(arrows)*, narrowing of capillary lumens, proliferation and swelling of visceral epithelial cells *(double arrows)*, and prominent accumulation of intracellular protein absorption droplets in the visceral epithelial cells *(arrowheads)*. Silver methenamine stain. (Courtesy Dr. Jolanta Kowalewska, Cedars-Sinai Medical Center, Los Angeles, Calif.)

HIV-Associated Nephropathy

HIV infection can directly or indirectly cause several renal complications, including acute renal failure or acute interstitial nephritis induced by drugs or infection, thrombotic microangiopathies, postinfectious glomerulonephritis, and, most commonly, a severe form of the collapsing variant of FSGS, termed *HIV-associated nephropathy*. The latter has been reported in 5% to 10% of HIV-infected individuals in some older series, more frequently in blacks than in whites. With the advent of antiretroviral therapy for HIV infection, the incidence of this lesion has been much reduced. The morphologic features of HIV-associated nephropathy are:

- A high frequency of the collapsing variant of FSGS (see Fig. 20.15)
- A striking focal cystic dilation of tubule segments, which are filled with proteinaceous material, and inflammation and fibrosis
- The presence of large numbers of tubuloreticular inclusions within endothelial cells, detected by electron microscopy. Such inclusions, also present in SLE, have been shown to be modifications of endoplasmic reticulum induced by circulating interferon-α. They are not usually present in idiopathic FSGS and therefore may have diagnostic value in a biopsy specimen.

The pathogenesis of HIV-associated FSGS is unclear, but may be primarily due to the presence of the G1/G2 risk alleles for *APOL1*. There is some data to suggest that HIV can infect tubular epithelial cells and podocytes, but much remains to be known.

Membranoproliferative Glomerulonephritis

MPGN is best considered a pattern of immune-mediated injury rather than a specific disease. An emerging consensus on classification separates MPGN into two groups, one (type I) characterized by deposition of immune complexes containing IgG and complement, and a second (type II, now called *dense deposit disease*) in which activation of complement appears to be the most important factor. The latter belong to a group of disorders called *C3 glomerulopathies*. The criteria that define this group are still evolving.

MPGN is characterized histologically by alterations in the GBM, accumulation of mesangial matrix, proliferation of glomerular cells, leukocyte infiltration, and the presence of deposits in mesangial regions and glomerular capillary walls. As we will describe later, these deposits are made up of immune complexes in type I MPGN and some unknown material in type II MPGN. In type II, C3 is present on the GBM but not in the dense deposits. This difference is important and suggests that although morphologically similar, types I and II MPGN are pathogenically distinct. In both types, because the proliferation is predominantly in the mesangium but also may involve the capillary loops, an older synonym is *mesangiocapillary glomerulonephritis*.

MPGN accounts for up to 10% of cases of nephrotic syndrome in children and young adults. Some patients present only with hematuria or proteinuria in the non-nephrotic range, but many others have a combined nephrotic-nephritic picture. MPGN is increasingly recognized to be associated with other systemic disorders and known etiologic agents (secondary MPGN), but there is still a residue of cases of unknown etiology (primary MPGN).

Pathogenesis

In most cases of type I MPGN, there is evidence of immune complexes in the glomerulus and activation of both classical and alternative complement pathways. The antigens involved in idiopathic MPGN are unknown. In many cases, they are believed to be proteins derived from infectious agents such as hepatitis C and B viruses, which presumably behave either as "planted" antigens after first binding to or becoming trapped within glomerular structures or are contained in preformed immune complexes deposited from the circulation.

MORPHOLOGY

The **glomeruli are large and hypercellular.** The hypercellularity is produced both by proliferation of cells in the mesangium and so-called endocapillary proliferation involving capillary endothelium and infiltrating leukocytes. The glomeruli have an accentuated "lobular" appearance due to the **proliferating mesangial cells and increased mesangial matrix** (Fig. 20.16). The GBM is thickened, and often shows a **"double-contour" or "tram-track" appearance**, especially evident in silver or PAS stains. This is caused by **"duplication" of the basement membrane** (also commonly referred to as splitting), usually as the result of new basement membrane synthesis in response to subendothelial deposits of immune complexes. Between the duplicated basement membranes, there is inclusion or interposition of cellular elements, which can be of mesangial, endothelial, or leukocytic origin. Such interposition also gives rise to the appearance of "split" basement membranes (Fig. 20.17A). Crescents are present in many cases.

Type I MPGN is characterized by the presence of discrete **subendothelial electron-dense deposits.** Mesangial and occasional subepithelial deposits may also be present (see Fig. 20.17A). By immunofluorescence, IgG and C3 are deposited in a granular pattern, and early complement components (C1q and C4) are often also present, indicative of an immune complex pathogenesis.

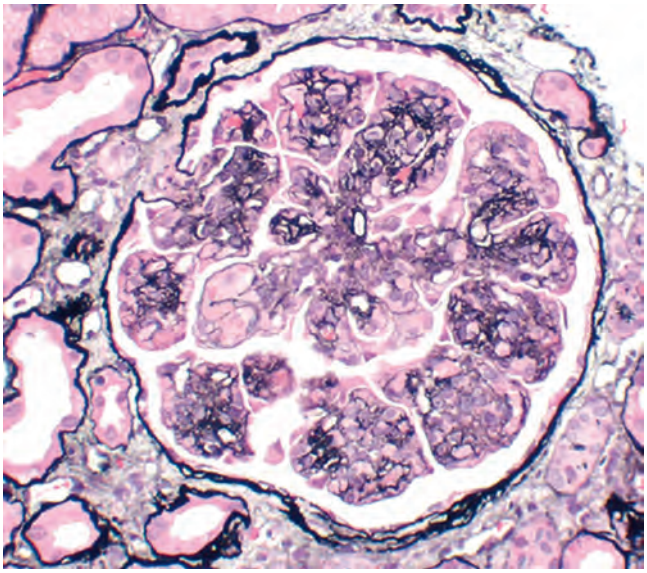


Figure 20.16 Membranoproliferative glomerulonephritis, showing mesangial cell proliferation, increased mesangial matrix (staining black with silver stain), basement membrane thickening with segmental splitting, accentuation of lobular architecture, swelling of cells lining peripheral capillaries, and influx of leukocytes (endocapillary proliferation).

Clinical Features

Most patients with primary MPGN present in adolescence or as young adults with nephrotic syndrome and a nephritic component manifested by hematuria or, more insidiously, as mild proteinuria. Few remissions occur spontaneously in either type, and the disease follows a slowly progressive but unremitting course. Some patients develop numerous crescents and a clinical picture of RPGN. About 50% develop chronic renal failure within 10 years. Treatments with steroids, immunosuppressive agents, and antiplatelet drugs have not proven to be of any benefit.

Secondary Membranoproliferative Glomerulonephritis

Secondary MPGN (invariably type I) is more common in adults and arises in the following settings:

- Chronic immune complex disorders, such as SLE; hepatitis B infection; hepatitis C infection, usually with cryoglobulinemia; endocarditis; infected ventriculoatrial shunts; chronic visceral abscesses; HIV infection; and schistosomiasis
- α_1-Antitrypsin deficiency
- Malignant diseases, particularly lymphoid tumors such as chronic lymphocytic leukemia, which are commonly complicated by development of autoantibodies

Dense Deposit Disease

Most patients with dense-deposit disease (formerly called type II MPGN) have abnormalities resulting in excessive activation of the alternative complement pathway. These patients have a consistently decreased serum C3 but normal C1 and C4, the early components of complement. They also have diminished serum levels of Factor B and properdin, components of the alternative complement pathway. The glomeruli contain deposits of C3 and properdin but not IgG. Recall that in the alternative complement pathway, C3 is directly cleaved to C3b (Fig. 20.18; see also Chapter 3,

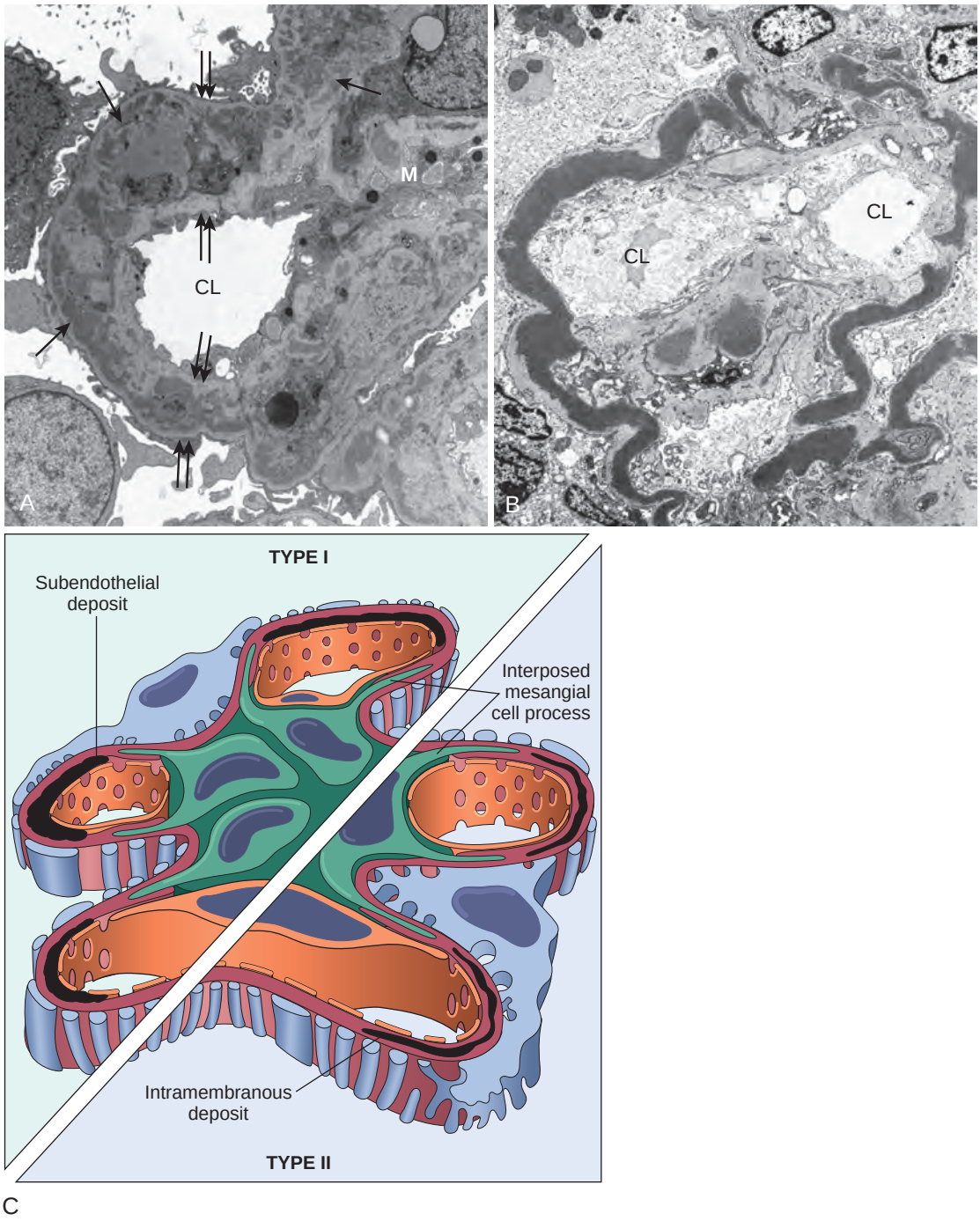


Figure 20.17 (A) Membranoproliferative glomerulonephritis, type I. Note discrete electron-dense deposits *(arrows)* incorporated into the glomerular capillary wall between duplicated (split) basement membranes *(double arrows)*, and in mesangial regions *(M);* (B) Dense-deposit disease (type II membranoproliferative glomerulonephritis). There are dense homogeneous deposits within the basement membrane. In both, mesangial interposition gives the appearance of split basement membranes when viewed in the light microscope. (C) Schematic representation of patterns in the two types of membranoproliferative glomerulonephritis. In type I, there are subendothelial deposits; type II is characterized by intramembranous dense deposits (dense-deposit disease). In both, the basement membranes appear split when viewed in the light microscope. *CL,* Capillary lumen. (A, Courtesy Dr. Jolanta Kowalewska, Cedars-Sinai Medical Center, Los Angeles, Calif.)

Fig. 3.12). The reaction depends on the initial activation of C3 by such substances as bacterial polysaccharides, endotoxin, and aggregates of IgA via a pathway involving Factors B and D. This leads to the generation of C3bBb, the alternative pathway C3 convertase. Normally, this C3 convertase is labile, but more than 70% of patients with dense-deposit disease have a circulating autoantibody termed *C3 nephritic factor (C3NeF)* that binds the alternative pathway C3 convertase and protects it from inactivation (see Fig. 20.18). This favors persistent C3 activation and hypocomplementemia. There is also decreased C3 synthesis by the liver, further contributing to the profound hypocomplementemia.

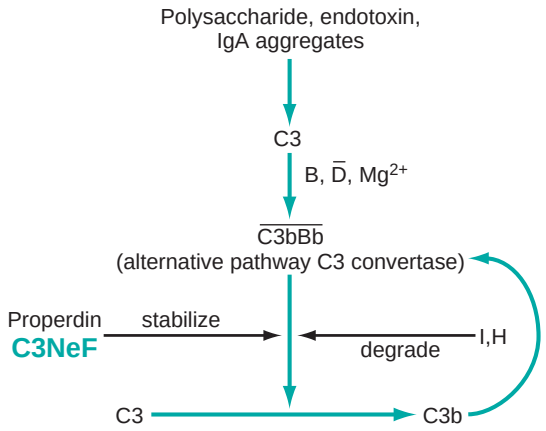


Figure 20.18 The alternative complement pathway in C3 glomerulopathy and atypical hemolytic uremic syndrome. Note that C3NeF, an antibody present in the serum of individuals with membranoproliferative glomerulonephritis, acts at the same step as properdin, serving to stabilize the alternative pathway C3 convertase, thus enhancing C3 activation and consumption, causing hypocomplementemia.

The precise nature of the dense deposits is unknown. Mutations in components of the alternate pathway such as Factor H have also been associated with dense deposit disease.

MORPHOLOGY

While some cases of dense deposit disease share histologic features with MPGN, there is a wider spectrum of histologic alterations in dense deposit disease. Many cases have a predominately mesangial proliferative pattern of injury, while others have an inflammatory and focally crescentic appearance. In some cases, dense deposits of a cellular material can be seen permeating the GBMs in histologic sections. The defining feature is revealed by election microscopy, which demonstrates permeation of the lamina densa of the GBM by a ribbonlike, homogeneous, extremely electron-dense material of unknown composition (see Fig. 20.17B). By immunofluorescence, C3 is present in irregular granular or linear foci in the basement membranes on either side but not within the dense deposits. C3 is also present in the mesangium in characteristic circular aggregates (mesangial rings). IgG is usually absent, as are components of the classical pathway of complement activation (such as C1q and C4). C3 glomerulopathies other than dense deposit disease can have a similar distribution, with mesangial and capillary wall involvement, but lack the extremely electron dense deposits that define dense deposit disease.

Clinical Features

Dense deposit disease affects primarily children and young adults. The clinical presentation of nephritic syndrome with hematuria and/or nephrotic syndrome with proteinuria overlaps with that of MPGN. The prognosis is poor, with over one-half of patients progressing to ESRD. There is a high incidence of recurrence in transplant recipients; dense deposits recur in 90% of such patients.

Fibrillary Glomerulonephritis

Fibrillary glomerulonephritis is a rare disease characterized by fibrillary deposits in the mesangium and glomerular capillary walls that resemble amyloid fibrils superficially but differ ultrastructurally and do not stain with Congo red. The glomerular lesions usually show mesangioproliferative or membranoproliferative patterns by light microscopy. By immunofluorescence microscopy, there is selective deposition of polyclonal IgG, often of the IgG4 subclass, complement C3, and Igκ and Igλ light chains. Clinically, patients develop nephrotic syndrome, hematuria, and progressive renal insufficiency. The disease recurs in kidney transplants. The pathogenesis is unknown, but DNAJB9, a co-chaperone for heat shock protein 70s, recently has been identified as a highly sensitive and specific marker for fibrillary GN.

KEY CONCEPTS

THE NEPHROTIC SYNDROME

- Membranous nephropathy is caused by an autoimmune response, most often directed against the PLA_2R on podocytes; it is characterized by granular subepithelial deposits of antibodies with GBM thickening and loss of foot processes but little or no inflammation; the disease is often resistant to steroid therapy.
- The nephrotic syndrome is characterized by proteinuria, which results in hypoalbuminemia and edema.
- Podocyte injury is an underlying mechanism of proteinuria, and may be the result of nonimmune causes (as in minimal change disease and FSGS) or immune mechanisms (as in membranous nephropathy).
- Minimal change disease is the most frequent cause of nephrotic syndrome in children; it is manifested by proteinuria and effacement of glomerular foot processes without antibody deposits; the pathogenesis is unknown; secondary forms can be triggered by infections, immunization, drugs, and certain neoplastic lesions; the disease responds well to steroid therapy.
- FSGS may be primary (podocyte injury by unknown mechanisms) or secondary (e.g., as a consequence of prior glomerulonephritis, hypertension, or infection such as HIV); glomeruli show focal and segmental obliteration of capillary lumens, and loss of foot processes; the disease is often resistant to therapy and may progress to ESRD.
- MPGN in most cases is the result of immune complex deposition in both mesangial regions and capillary walls. It may be associated with systemic infections.
- Dense deposit disease (formerly type II MPGN), defined by a unique permeation of GBMs by electron dense material, primarily affects children and young adults. It is associated with acquired or genetic dysregulation of the alternate pathway of complement.
- Fibrillary glomerulonephritis is a rare disease characterized by Congo-red negative fibrillary glomerular deposits of unknown etiology.

Other Glomerular Diseases

IgA Nephropathy (Berger Disease)

IgA nephropathy, characterized by the presence of prominent IgA deposits in the mesangial regions and recurrent hematuria, is the most common type of glomerulonephritis worldwide. The disease can be suspected by light microscopic examination, but the diagnosis is made only by the detection of glomerular IgA deposition (Fig. 20.19). Mild proteinuria is usually present, and the nephrotic syndrome

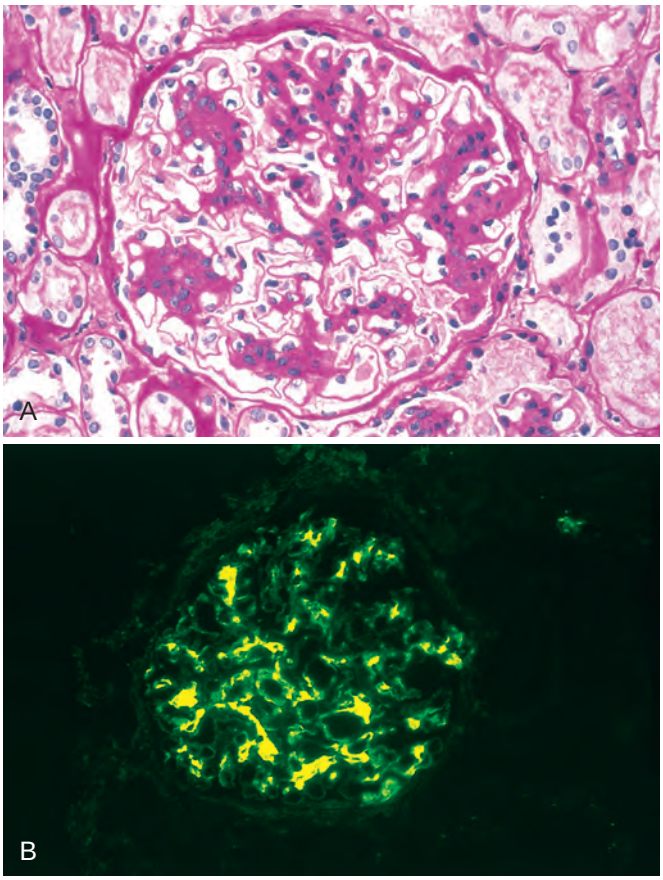


Figure 20.19 IgA nephropathy. (A) Light microscopy showing mesangial proliferation and matrix increase. (B) Characteristic deposition of IgA, principally in mesangial regions, detected by immunofluorescence.

may occasionally develop. Rarely, patients may present with crescentic GN.

Whereas IgA nephropathy is typically an isolated renal disease, similar IgA deposits are present in a systemic disorder of children, *Henoch-Schönlein purpura,* to be discussed later, which has many overlapping features with IgA nephropathy. In addition, *secondary IgA nephropathy* occurs in patients with liver and intestinal diseases, as discussed later.

Pathogenesis

Current evidence favors a "multi-hit" etiology for this disorder involving several steps. IgA, the main Ig in mucosal secretions, is present in plasma at low concentrations, mostly in monomeric form, the polymeric forms being catabolized in the liver. In patients with IgA nephropathy, levels of plasma polymeric IgA are increased, but increased production is not sufficient to cause this disease. A clue comes from the observation that in IgA nephropathy, the glomerular deposits consist predominantly of polymeric IgA molecules with aberrant glycosylation. It is believed that a key facet of IgA nephropathy is a hereditary or acquired defect in the normal formation or attachment of galactose-containing sugar chains called O-linked glycans to the hinge region of the IgA molecule (particularly to those of the IgA1 subclass) prior to their secretion by B cells. This aberrantly glycosylated IgA is either deposited by itself in glomeruli, or it elicits an autoimmune response and forms immune complexes in the circulation with IgG autoantibodies directed against the abnormal IgA molecules. The immune complexes are deposited in the mesangium; alternatively, the abnormal IgA1 is deposited in the mesangium with subsequent formation of immune complexes in situ. The mesangial immune deposits then activate mesangial cells to proliferate, produce increased amounts of ECM, and secrete numerous cytokines and growth factors. These secreted mediators may not only participate in further mesangial cell activation but may also recruit inflammatory cells into the glomeruli. The recruited leukocytes contribute to glomerular injury and also to a reparative response, which can include opsonization and removal of the immune complexes. The deposited IgA and IgA-containing immune complexes activate the complement system via the alternate pathway, and hence the presence of C3 and the absence of C1q and C4 in glomeruli are typical of this disorder. A genetic influence is suggested by the occurrence of this condition in families and in HLA-identical siblings, the increased frequency of certain HLA and complement genotypes in some populations, and the findings of genomewide association studies linking specific MHC Class II loci to disease susceptibility.

Epidemiologic features of this disorder indicate that the increased synthesis of abnormal IgA may occur in response to respiratory or gastrointestinal exposure to environmental agents (e.g., viruses, bacteria, food proteins). The specific initiating antigens are unknown, and several infectious agents and food products have been implicated. IgA nephropathy occurs with increased frequency in individuals with *gluten enteropathy* (celiac disease), in whom intestinal mucosal defects are well defined, and in *liver disease,* in which there is defective hepatobiliary clearance of IgA complexes *(secondary IgA nephropathy)*.

MORPHOLOGY

On histologic examination, the lesions vary considerably. The glomeruli may be normal or may show mesangial widening and endocapillary proliferation (mesangioproliferative glomerulonephritis) (Fig. 20.19A), segmental proliferation confined to some glomeruli (focal proliferative glomerulonephritis), or rarely, overt crescentic glomerulonephritis. The presence of leukocytes within glomerular capillaries is a variable feature. The mesangial widening may be the result of cell proliferation, accumulation of matrix, immune deposits, or some combination of these abnormalities. Healing of the focal proliferative lesion may lead to secondary focal segmental sclerosis. The characteristic immunofluorescent picture is of **mesangial deposition of IgA** (Fig. 20.19B), often with C3 and properdin and lesser amounts of IgG or IgM. Early complement components are usually absent. Electron microscopy confirms the presence of electron-dense deposits predominantly in the mesangium; capillary wall deposits, if present, are usually sparse.

Clinical Features

The disease affects people of any age, most commonly older children and young adults. Many patients present with gross hematuria after an infection of the respiratory or, less commonly, gastrointestinal or urinary tract; 30% to 40% have only microscopic hematuria, with or without proteinuria;

and 5% to 10% develop acute nephritic syndrome, including some with RPGN. The hematuria typically lasts for several days and then subsides, only to return every few months. The subsequent course is highly variable. Many patients maintain normal renal function for decades. Slow progression to chronic renal failure occurs in 15% to 40% of cases over a period of 20 years. Onset in old age, heavy proteinuria, hypertension, and the extent of glomerulosclerosis on biopsy are clues to an increased risk of progression. Recurrence of IgA deposits in transplanted kidneys is frequent, and in approximately 15% of those with recurrent IgA deposits, the disease runs the same slowly progressive course as that of primary IgA nephropathy.

Hereditary Nephritis

Hereditary nephritis refers to a group of heterogeneous familial renal diseases associated with mutations in collagen genes that manifest primarily with glomerular injury. Two deserve discussion: *Alport syndrome,* because the lesions and genetic defects have been well studied, and *thin basement membrane nephropathy,* the most common cause of *benign familial hematuria*.

Alport Syndrome

Alport syndrome is caused by mutations affecting type IV collagen that result in hematuria with progression to chronic renal failure, accompanied by nerve deafness and various eye disorders, including lens dislocation, posterior cataracts, and corneal dystrophy. The disease is inherited as an X-linked trait in approximately 85% of cases. In the X-linked form, males express the full syndrome, while female heterozygotes typically present with hematuria. Approximately 90% of affected males progress to ESRD before 40 years of age. Autosomal recessive and autosomal dominant pedigrees also exist, in which males and females are equally susceptible to the full syndrome.

Pathogenesis

The disease manifestations are due to mutations in one of several genes coding for subunits of the collagen IV molecule. More than 500 mutations resulting in disease have been identified, all of which interfere with assembly of type IV collagen, which is crucial for function of the GBM, the lens of the eye, and the cochlea. Because the GBM consists of networks of trimeric collagen IV molecules composed of α_3, α_4, and α_5 chains, mutations affecting any one chain result in defective assembly of the collagen network. Because type IV collagen chains are encoded on autosomes (chromosomes 2 and 13) and the X-chromosome, the inheritance pattern can be autosomal or X-linked. Missense and splice site mutations, insertions, and deletions have all been identified. Genetic analysis has shown that in patients with X-linked disease, large deletions in the collagen IV α_5 chain (COL4A5) are associated with ESRD at an earlier age.

MORPHOLOGY

Fully developed Alport syndrome has characteristic electron microscopic findings. The GBM shows irregular foci of thickening alternating with attenuation (thinning), and pronounced splitting and lamination of the lamina densa, often producing a distinctive basket-weave appearance (Fig. 20.20). Similar alterations can be found in the tubular basement membranes.

Immunohistochemistry can be helpful in cases with absent or borderline basement membrane lesions, because antibodies to α_3, α_4, and α_5 collagen fail to stain both glomerular and tubular basement membranes in the classic X-linked form. There is also absence of α_5 staining in skin biopsy specimens from these patients. As the disease progresses, there is development of focal segmental and global glomerulosclerosis and other changes of progressive renal injury, including vascular sclerosis, tubular atrophy, and interstitial fibrosis.

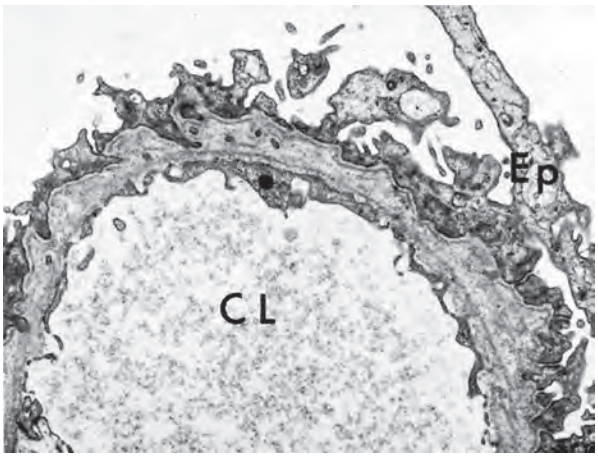


Figure 20.20 Hereditary nephritis (Alport syndrome). Electron micrograph of a glomerular capillary with irregular thickening of the basement membrane, lamellation of the lamina densa, and foci of rarefaction. Such changes may be present in other diseases but are most pronounced and widespread in hereditary nephritis. *CL,* Capillary lumen; *Ep,* epithelium.

Clinical Features

The most common presenting sign is gross or microscopic hematuria, frequently accompanied by red cell casts. Proteinuria may develop later, and rarely, the nephrotic syndrome develops. Symptoms appear at 5 to 20 years of age, and the onset of overt renal failure is between 20 and 50 years of age in men. The auditory defects may be subtle, requiring sensitive testing.

Thin Basement Membrane Nephropathy (Benign Familial Hematuria)

This is a fairly common hereditary entity manifested *clinically by familial asymptomatic hematuria* – usually uncovered on routine urinalysis – *and morphologically by diffuse thinning of the GBM* to widths between 150 and 225 nm (compared with 300 to 400 nm in healthy adults). Although mild or moderate proteinuria may also be present, renal function is normal and prognosis is excellent. The abnormality is estimated to affect 1% of the general population.

The disorder should be distinguished from IgA nephropathy, another common cause of hematuria, and X-linked Alport syndrome. In contrast to Alport syndrome, hearing loss, ocular abnormalities, and a family history of renal failure are absent.

The anomaly in thin basement membrane nephropathy has also been traced to mutations in genes encoding α_3 or

α_4 chains of type IV collagen. The disease most often has an autosomal inheritance, and most patients are heterozygous for the defective gene. The disorder in homozygotes resembles autosomal recessive Alport syndrome. Homozygotes or compound heterozygotes may progress to renal failure. Thus, these diseases illustrate a continuum of changes resulting from mutations in collagen type IV genes.

KEY CONCEPTS

GLOMERULAR DISEASES WITH ISOLATED HEMATURIA

- IgA nephropathy, characterized by mesangial deposits of IgA-containing immune complexes, is the most common cause of glomerulonephritis worldwide. It is a common cause of both isolated and frequently recurrent hematuria and of various degrees of proteinuria and sometimes of nephritic syndrome; it commonly affects children and young adults and has a variable course.
- Alport syndrome, a form of hereditary nephritis, is caused by mutations in genes encoding GBM type IV collagen. It manifests as hematuria and slowly progressing proteinuria and declining renal function; glomeruli appear normal by light microscopy until late in the disease course.
- Thin basement membrane nephropathy has a benign clinical course and is also the result of mutations in genes coding for GBM type IV collagen and hence may be considered as part of a spectrum of diseases that includes hereditary nephritis.

Glomerular Lesions Associated With Systemic Diseases

Many immunologically mediated, metabolic, or hereditary systemic disorders are associated with glomerular injury; in some (e.g., SLE and diabetes), the glomerular involvement is a major clinical manifestation. Most of these diseases are discussed elsewhere in this book. Here we briefly recall some of the lesions and discuss only those not considered in other sections.

Lupus Nephritis

The various types of lupus nephritis are described in Chapter 6. As discussed, SLE gives rise to a wide variety of renal lesions and clinical presentations. Clinical manifestations can include recurrent microscopic or gross hematuria, nephritic syndrome, RPGN, nephrotic syndrome, acute and chronic renal failure, and hypertension.

Henoch-Schönlein Purpura

This childhood syndrome consists of purpuric skin lesions, abdominal pain and intestinal bleeding, and arthralgias along with renal abnormalities. Skin lesions characteristically involve the extensor surfaces of arms and legs as well as buttocks; abdominal manifestations include pain, vomiting, and intestinal bleeding. Renal manifestations occur in one-third of patients and include gross or microscopic hematuria, nephritic syndrome, nephrotic syndrome, or some combination of these. A small number of patients, mostly adults, develop a rapidly progressive form of glomerulonephritis with many crescents. Not all components of the syndrome need to be present for the diagnosis, and individual patients may have purpura, abdominal pain, or urinary abnormalities as the dominant feature. The disease is most common in children 3 to 8 years of age, but it also occurs in adults, in whom the renal manifestations are usually more severe. There is a strong background of atopy in about one-third of patients, and onset often follows an upper respiratory infection. IgA is deposited in the glomerular mesangium in a distribution similar to that of IgA nephropathy. This has led to the concept that *IgA nephropathy and Henoch-Schönlein purpura are manifestations of the same disease.* The finding of Ig and C3 deposits in glomeruli suggests that immune complexes are involved in the disease.

MORPHOLOGY

On histologic examination, the renal lesions vary from mild focal mesangial proliferation to diffuse mesangial proliferation and/or endocapillary proliferation to crescentic glomerulonephritis. Whatever the histologic lesions, the pathognomonic feature by fluorescence microscopy is the **deposition of IgA, sometimes with IgG and C3, in the mesangial region**, sometimes with deposits extending to the capillary loops. The skin lesions consist of subepidermal hemorrhages and a necrotizing vasculitis involving the small vessels of the dermis. Deposits of IgA, along with IgG and C3, are also present in such vessels. Vasculitis also occurs in other organs, such as the gastrointestinal tract, but is rare in the kidney.

The course of the disease is variable, but recurrences of hematuria may persist for many years after onset. Most children have an excellent prognosis. Patients with the more diffuse lesions, crescents, or the nephrotic syndrome have a somewhat poorer prognosis.

Diabetic Nephropathy

Diabetes is a major cause of renal morbidity and mortality, and diabetic nephropathy is the leading cause of chronic kidney failure in the United States. Advanced or end-stage kidney disease occurs in as many as 40% of both type 1 diabetics and type 2 diabetics. The pathology and pathogenesis of this disorder are discussed in Chapter 24.

Other Systemic Disorders

Goodpasture syndrome (Chapter 15), *microscopic polyangiitis,* and *granulomatosis with polyangiitis* (formerly called *Wegener granulomatosis)* (Chapter 11) are commonly associated with glomerular lesions, as described in the discussion of these diseases. Suffice it to say here that the glomerular lesions in these three conditions can be histologically similar and are principally characterized by foci of glomerular necrosis and crescent formation. In the early or mild forms of renal involvement, there is focal and segmental, sometimes necrotizing, glomerulonephritis, and most of these patients will have hematuria with a mild decline in GFR. In the more severe cases, which may be associated with RPGN, there is more extensive necrosis, fibrin deposition, and extensive formation of epithelial (cellular) crescents, which can become organized to form fibrocellular and fibrous crescents if the glomerular injury evolves into segmental or global scarring (sclerosis).

Essential mixed cryoglobulinemia is another systemic condition in which deposits of cryoglobulins composed principally of IgG-IgM complexes induce cutaneous vasculitis, synovitis, and a proliferative glomerulonephritis, typically MPGN. Most cases of essential mixed cryoglobulinemia have been associated with infection with hepatitis C virus, and this condition in particular is associated with glomerulonephritis, usually MPGN type I.

Immunoglobulins secreted by plasma cell neoplasms may also induce glomerular lesions, including amyloidosis.

TUBULAR AND INTERSTITIAL DISEASES

Most forms of tubular injury involve the interstitium as well; therefore, diseases affecting these two components are discussed together. Under this heading, we consider two major processes: (1) ischemic or toxic tubular injury, and (2) inflammation of the tubules and interstitium *(tubulointerstitial nephritis)*.

Acute Tubular Injury/Necrosis

ATI is characterized by acute renal failure and often, but not invariably, morphologic evidence of tubular injury, in the form of necrosis of tubular epithelial cells. Because necrosis is often not present, the term ATI is preferred by pathologists over the older name *acute tubular necrosis* (ATN). It is the most common cause of acute kidney injury. ATI can be caused by a variety of conditions, including the following:

- *Ischemia, due to decreased or interrupted blood flow*. Examples include diffuse involvement of the intrarenal blood vessels such as in microscopic polyangiitis, microangiopathies (e.g., hemolytic uremic syndrome [HUS] or thrombotic thrombocytopenic purpura [TTP]), or decreased effective circulating blood volume, as occurs in hypovolemic shock (Chapter 4).
- *Direct toxic injury to the tubules*. This may be caused by *endogenous* agents, e.g., myoglobin, hemoglobin, monoclonal light chains, bile/bilirubin, *or exogenous agents,* e.g., drugs, radiocontrast dyes, heavy metals, organic solvents.

ATI accounts for some 50% of cases of acute kidney injury in hospitalized patients. Other causes of acute renal failure are discussed elsewhere in this chapter.

ATI is a reversible process that arises in a variety of clinical settings. Most of these, ranging from severe trauma to acute pancreatitis, have in common a period of inadequate blood flow to the peripheral organs, usually accompanied by marked hypotension and shock. This pattern is called *ischemic ATI*. The second pattern, called *nephrotoxic ATI,* is caused by a multitude of drugs, such as gentamicin, radiographic contrast agents, poisons, including heavy metals (e.g., mercury), and organic solvents (e.g., carbon tetrachloride). Combinations of ischemic and nephrotoxic ATI also can occur, exemplified by mismatched blood transfusions and other hemolytic crises causing *hemoglobinuria* and skeletal muscle injuries causing *myoglobinuria*. Such injuries result in characteristic intratubular hemoglobin or myoglobin casts, respectively; the toxic iron content of these globin molecules contributes to the ATI. In addition to its frequency, the potential reversibility of ATI adds to its clinical importance. Proper management can make the difference between recovery and death.

Pathogenesis

The critical events in both ischemic and nephrotoxic ATI are (1) tubular injury and (2) persistent and severe disturbances in blood flow (Fig. 20.21).

- *Tubular cell injury:* Tubular epithelial cells, especially those of proximal tubules, are particularly sensitive to ischemia and are also vulnerable to toxins. Several factors predispose the tubules to toxic injury, including an increased surface area for tubular reabsorption, active transport systems for ions and organic acids, a high rate of metabolism and oxygen consumption that is required to perform these transport and reabsorption functions, and the capability for resorption and concentration of toxins. Ischemia causes numerous structural and functional

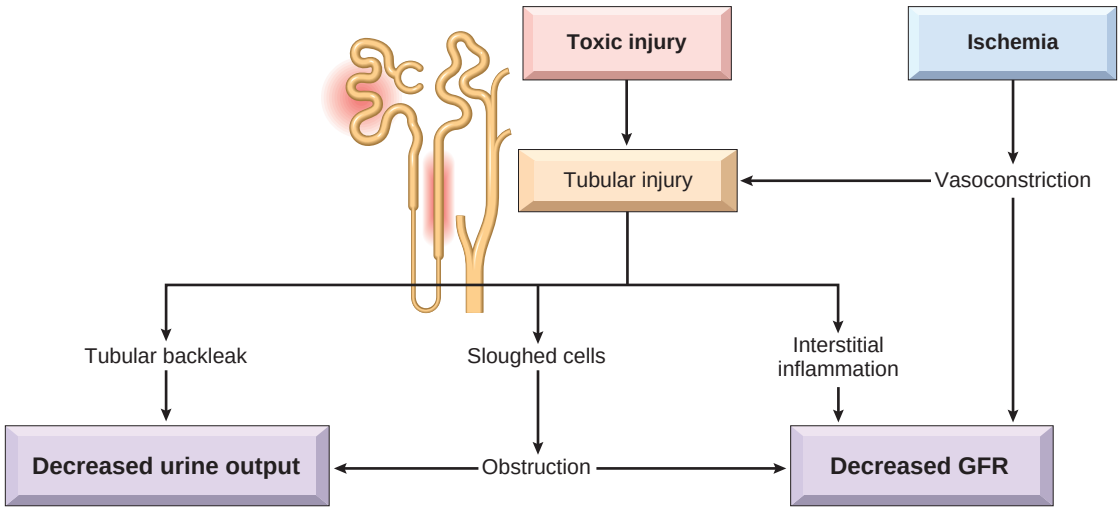


Figure 20.21 Postulated sequence in ischemic or toxic acute tubular injury. *GFR,* Glomerular filtration rate.

alterations in epithelial cells, as discussed in Chapter 2. One early reversible result of ischemia is loss of cell polarity due to redistribution of membrane proteins (e.g., the enzyme Na,K^+-ATPase) from the basolateral to the luminal surface of the tubular cells, resulting in abnormal ion transport across the cells and increased sodium delivery to distal tubules, which incites vasoconstriction via *tubuloglomerular feedback*. Initially, this response to the increased sodium lowers the GFR to maintain distal blood flow. In addition, ischemic tubular cells express cytokines and adhesion molecules, thus recruiting leukocytes that appear to participate in the subsequent injury. In time, injured cells detach from the basement membranes and cause *luminal obstruction,* increased intratubular pressure, and further decrease in GFR. In addition, glomerular filtrate in the lumen of the damaged tubules can leak back into the interstitium, resulting in interstitial edema, increased interstitial pressure, and further damage to the tubule. All of these effects, as shown in Fig. 20.21, contribute to the decreased GFR.

- *Disturbances in blood flow:* Ischemic renal injury is also characterized by *hemodynamic alterations* that cause reduced GFR. The major one is *intrarenal vasoconstriction,* which results in both reduced glomerular blood flow and reduced oxygen delivery to the functionally important tubules in the outer medulla (thick ascending limb and straight segment of the proximal tubule). Several vasoconstrictor pathways have been implicated, including the renin-angiotensin system, stimulated by decreased sodium in the tubules as a result of decreased blood pressure, and *sublethal endothelial injury,* leading to increased release of the vasoconstrictor *endothelin* and decreased production of the vasodilators *nitric oxide* and *prostacyclin (prostaglandin I_2)*. There is also some evidence of a direct effect of ischemia or toxins on the glomerulus, causing a reduced glomerular ultrafiltration coefficient.

The patchiness of tubular necrosis and maintenance of the integrity of the basement membrane along many segments allow repair of the injured foci and recovery of function if the precipitating cause is removed. This repair is dependent on the capacity of reversibly injured epithelial cells to proliferate and differentiate. Re-epithelialization is mediated by a variety of growth factors and cytokines produced locally by the tubular cells themselves or by inflammatory cells in the vicinity of necrotic foci.

MORPHOLOGY

ATI is characterized by **tubular epithelial injury** at multiple points along the nephron, with large skip areas in between, often accompanied by rupture of basement membranes (tubulorrhexis) and **occlusion of tubular lumens by casts** (Figs. 20.22 and 20.23). The distinct patterns of tubular injury in ischemic and toxic ATI are shown in Fig. 20.22. The straight portion of the proximal tubule and the ascending thick limb in the renal medulla are especially vulnerable, but focal lesions may also occur in the distal tubule, often in conjunction with casts. It should be noted that the severity of the morphologic findings often does not correlate with the severity of clinical manifestations.

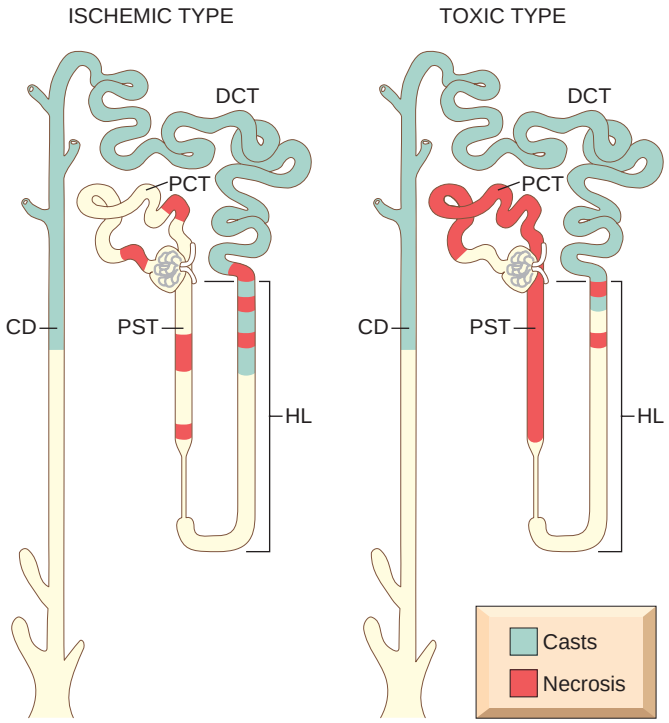


Figure 20.22 Patterns of tubular damage in ischemic and toxic acute tubular injury. In the ischemic type, tubular necrosis is patchy, relatively short lengths of tubules are affected, and proximal straight tubule *(PST)* segments and ascending limbs of Henle's loop *(HL)* are most vulnerable. In toxic acute tubular injury, extensive necrosis is present along the proximal convoluted tubule *(PCT)* segments with many toxins (e.g., mercury), but necrosis of the distal tubule, particularly ascending HL, also occurs. In both types, lumens of the distal convoluted tubules *(DCT)* and collecting ducts *(CD)* contain casts.

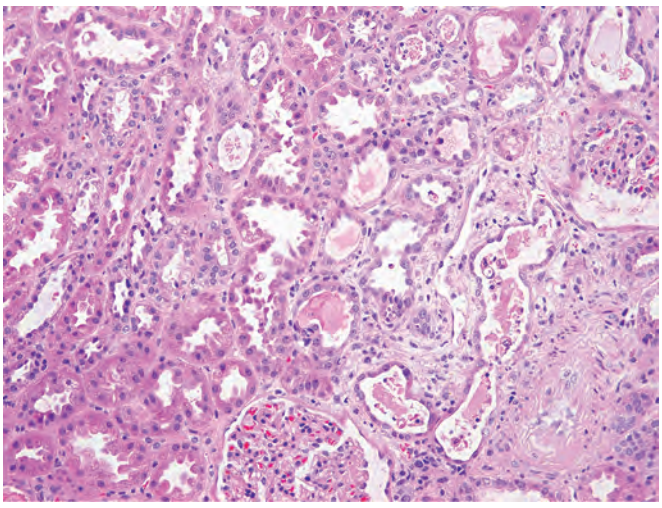


Figure 20.23 Acute tubular injury. Some tubular epithelial cells in the tubules are necrotic, and many have detached (from their basement membranes) and been sloughed into the tubular lumens.

Other findings in ischemic ATI are interstitial edema and accumulations of leukocytes within dilated vasa recta. There is also evidence of epithelial regeneration in the form of flattened epithelial cells with hyperchromatic nuclei and mitotic figures. In time, this regeneration repopulates the tubules so that no residual evidence of damage is seen.

Toxic ATI is manifested by ATI, most obvious in the proximal convoluted tubules. On histologic examination, the tubular injury can be distinctive in poisoning with certain agents. With mercuric chloride, for example, severely injured cells may contain large acidophilic inclusions. Later, these cells become necrotic, are desquamated into the lumen, and may undergo calcification. Ethylene glycol poisoning produces marked ballooning and hydropic or vacuolar degeneration of proximal convoluted tubules, and calcium oxalate crystals are often also found in the tubular lumens.

Clinical Features

The clinical course of ATI is highly variable, but the classic case may be divided into *three* stages.

- *Initiation phase,* lasting about 36 hours, is dominated by the inciting medical, surgical, or obstetric event. The only indication of renal involvement is a slight decline in urine output with a rise in BUN. At this point, oliguria could be explained by a transient decrease in blood flow and declining GFR.
- *Maintenance phase* is characterized by sustained decreases in urine output to between 40 and 400 mL/day (oliguria), salt and water overload, rising BUN concentrations, hyperkalemia, metabolic acidosis, and other manifestations of uremia. With appropriate management, the patient can overcome this oliguric crisis.
- *Recovery phase* is ushered in by a steady increase in urine volume that may reach up to 3 L/day. The tubules are still damaged, so large amounts of water, sodium, and potassium are lost in the flood of urine. *Hypokalemia, rather than hyperkalemia, becomes a clinical problem.* There is a peculiar increased vulnerability to infection at this stage. Eventually, renal tubular function is restored and concentrating ability improves. At the same time, BUN and creatinine levels begin to return to normal. Subtle tubular functional impairment may persist for months, but most patients who reach this phase eventually recover completely.

The prognosis of ATI depends on the magnitude and duration of injury. Recovery is expected with nephrotoxic ATI when the toxin has not caused serious damage to other organs, such as the liver or heart. With supportive care, 95% of those who do not succumb to the precipitating cause recover. Conversely, in shock related to sepsis, extensive burns, or other causes of multiorgan failure, the mortality rate can exceed 50%.

KEY CONCEPTS

ACUTE TUBULAR INJURY

- ATI is the most common cause of acute kidney injury and attributed to ischemia and/or toxicity from an endogenous or exogenous substance.
- Tubular epithelial cell injury and altered intrarenal hemodynamics are the primary contributors to ATI.
- The clinical outcome is determined by the magnitude and duration of ATI.

Tubulointerstitial Nephritis

This group of renal diseases involves inflammatory injuries of the tubules and interstitium that are often insidious in onset and are principally manifest by azotemia. Tubulointerstitial nephritis can be acute or chronic. *Acute tubulointerstitial nephritis* has a rapid clinical onset and is characterized histologically by interstitial edema, often accompanied by leukocytic infiltration of the interstitium and tubules, and tubular injury. In *chronic interstitial nephritis,* there is infiltration with predominantly mononuclear leukocytes, prominent interstitial fibrosis, and widespread tubular atrophy. *Secondary tubulointerstitial nephritis* is also present in a variety of vascular, cystic (polycystic kidney disease), and metabolic (diabetes) renal disorders, in which it may also contribute to progressive damage. Here we discuss primary causes of tubulointerstitial injury (Table 20.8). Glomerular and vascular abnormalities may also be present in advanced stages of these diseases. Conversely, chronic tubulointerstitial damage is an important consequence of progression in diseases that primarily affect the glomerulus.

Tubulointerstitial disorders are distinguished clinically from the glomerular diseases by the following hallmarks:

- Absence of nephritic or nephrotic syndrome
- Presence of defects in tubular function. The latter may be subtle and include impaired ability to concentrate urine, evidenced clinically by polyuria or nocturia; salt wasting; diminished ability to excrete acids (metabolic acidosis); and isolated defects in tubular reabsorption or secretion. Advanced forms, however, may be difficult to distinguish clinically from other causes of renal insufficiency.

Table 20.8 Causes of Tubulointerstitial Nephritis

Infections
Acute bacterial pyelonephritis
Chronic pyelonephritis (including reflux nephropathy)
Other infections (e.g., viruses, parasites)
Toxins
Drugs
Acute-hypersensitivity interstitial nephritis
Analgesics
Heavy metals
Lead, cadmium
Metabolic Diseases
Urate nephropathy
Nephrocalcinosis (hypercalcemic nephropathy)
Oxalate nephropathy
Physical Factors
Chronic urinary tract obstruction
Neoplasms
Multiple myeloma (light-chain cast nephropathy)
Immunologic Reactions
Transplant rejection
Sjögren syndrome
Sarcoidosis
Vascular Diseases
Miscellaneous
Nephronophthisis
"Idiopathic" interstitial nephritis

Specific conditions are listed in Table 20.8 and are discussed elsewhere in this book. This section deals principally with pyelonephritis and drug-induced tubulointerstitial nephritis.

Pyelonephritis and Urinary Tract Infection

Pyelonephritis is one of the most common diseases of the kidney and is defined as inflammation affecting the tubules, interstitium, and renal pelvis. It occurs in two forms. *Acute pyelonephritis* is generally caused by bacterial infection and is associated with urinary tract infection. *Chronic pyelonephritis* is a more complex disorder; bacterial infection plays a dominant role, but other factors (vesicoureteral reflux, obstruction) predispose to repeat episodes of acute pyelonephritis.

Pyelonephritis is a serious complication of *urinary tract infections* that affect the bladder (cystitis), the kidneys and their collecting systems (pyelonephritis), or both. Bacterial infections of the lower urinary tract may be asymptomatic (asymptomatic bacteriuria) and remain localized to the bladder, but lower urinary tract infection can potentially spread to the kidney.

Pathogenesis

More than 85% of urinary tract infections are caused by the gram-negative bacilli that are normal inhabitants of the intestinal tract. By far the most common is *Escherichia coli,* followed by *Proteus, Klebsiella,* and *Enterobacter. Streptococcus faecalis,* also of enteric origin, staphylococci, and virtually every other bacterial and fungal agent can also cause lower urinary tract and renal infection. Mycobacterial and fungal organisms induce caseating and non-caseating granulomatous inflammation, respectively. In immunocompromised persons, particularly those with transplanted organs, viruses such as polyomavirus, cytomegalovirus, and adenovirus can also cause renal infection.

There are two routes by which bacteria can reach the kidneys: (1) through the bloodstream (hematogenous infection) and (2) from the lower urinary tract (ascending infection) (Fig. 20.24). The hematogenous route is less common and results from seeding of the kidneys by bacteria from distant foci in the course of septicemia or localized infections such as infective endocarditis. Hematogenous infection is more likely to occur in the presence of ureteral obstruction, and in debilitated patients. Typically, in patients receiving immunosuppressive therapy, nonenteric organisms, such as staphylococci and certain fungi and viruses, are involved.

Ascending infection is the most common cause of clinical pyelonephritis. Normal human bladder and bladder urine are sterile; therefore, a number of steps must occur for renal infection to occur:

- The first step is *colonization of the distal urethra and introitus* (in the female) by coliform bacteria. This colonization is influenced by the degree of bacterial adherence to urethral mucosal epithelia, as discussed in Chapter 8, which involves adhesive molecules (adhesins) on the P-fimbriae (pili) of bacteria that interact with receptors on the surface of urothelial cells.
- *From the urethra to the bladder*, organisms gain entrance during urethral catheterization or other instrumentation. Long-term catheterization, in particular, carries a risk of infection. In the absence of instrumentation, *urinary infections are much more common in females,* and this has been ascribed to the shorter urethra in females, the absence of antibacterial properties in prostatic fluid, hormonal changes affecting bacterial adherence to the mucosa, and urethral trauma during sexual intercourse, or any combination of these factors.

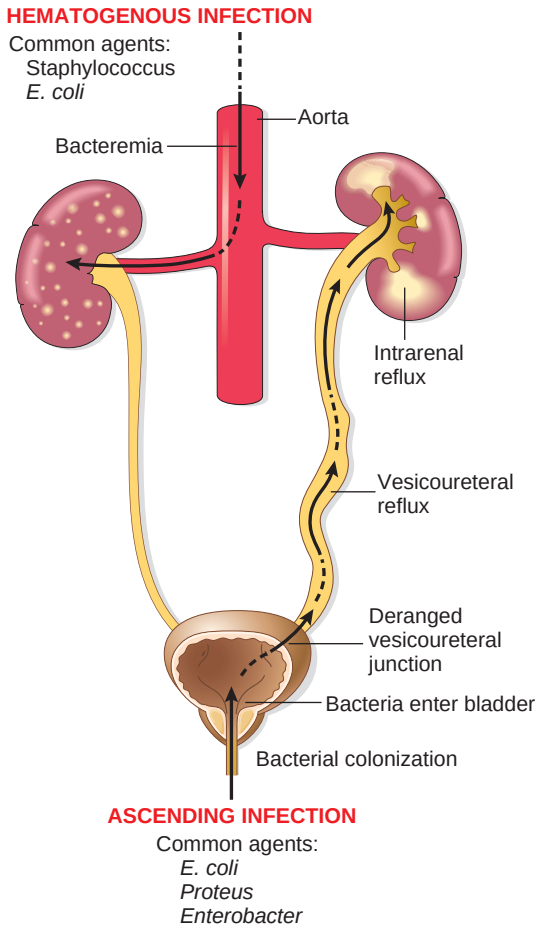


Figure 20.24 Schematic representation of pathways of renal infection. Hematogenous infection results from bacteremic spread. More common is ascending infection, which results from a combination of urinary bladder infection, vesicoureteral reflux, and intrarenal reflux.

Several conditions predispose to the movement of microbes from the bladder to the kidneys:

- *Urinary tract obstruction and stasis of urine.* Ordinarily, organisms introduced into the bladder are cleared by continual voiding and by antibacterial mechanisms. However, outflow obstruction or bladder dysfunction results in incomplete emptying and residual urine. In the presence of stasis, bacteria introduced into the bladder can multiply unhindered. Accordingly, urinary tract infection is frequent among patients with lower urinary tract obstruction, such as may occur with benign prostatic hypertrophy, tumors, calculi, or neurogenic bladder dysfunction caused by diabetes or spinal cord injury.
- *Vesicoureteral reflux.* Although obstruction is an important predisposing factor in ascending infection, *incompetence of the vesicoureteral valve* allows bacteria to ascend the ureter into the renal pelvis. The normal ureteral insertion

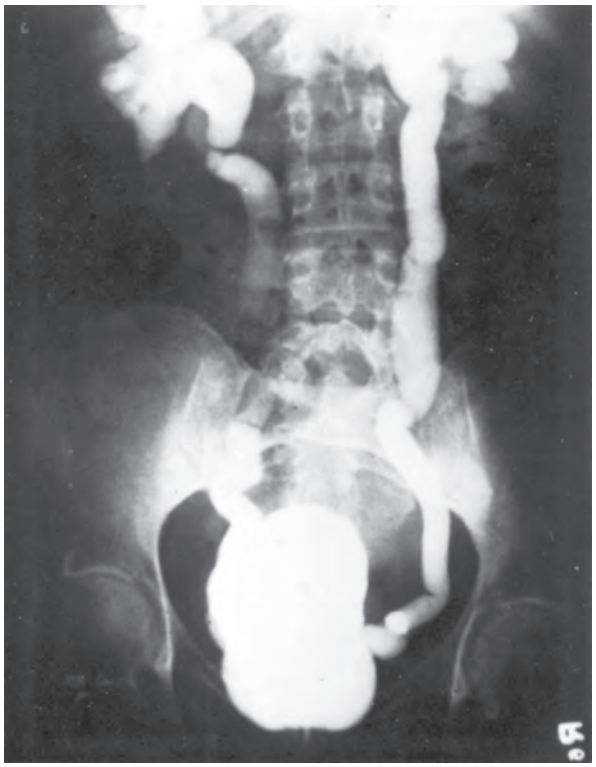


Figure 20.25 Vesicoureteral reflux demonstrated by a voiding cystourethrogram. Dye injected into the bladder refluxes into both dilated ureters, filling the pelvis and calyces.

into the bladder is a one-way valve that prevents retrograde flow of urine when the intravesical pressure rises, as in micturition. An incompetent vesicoureteral orifice allows the reflux of bladder urine into the ureters *(vesicoureteral reflux)* (Fig. 20.25). Reflux is most often due to a congenital absence or shortening of the intravesical portion of the ureter, such that the ureter is not compressed during micturition. In addition, it may be acquired by bladder infection itself. It is postulated that bacteria themselves or the associated inflammation can promote reflux by affecting ureteral contractility, particularly in children. Vesicoureteral reflux is estimated to affect 1% to 2% of otherwise normal children. *Acquired vesicoureteral reflux* in adults can result from persistent bladder atony caused by spinal cord injury. Similar to that of an obstruction, the residual urine in the urinary tract after voiding favors bacterial growth.

- *Intrarenal reflux.* Vesicoureteral reflux also affords a mechanism to propel infected bladder urine up to the renal pelvis and deep into the renal parenchyma through open ducts at the tips of the papillae (intrarenal reflux). Intrarenal reflux is most common in the upper and lower poles of the kidney, where papillae tend to have flattened or concave tips rather than the convex pointed type present in the midzones of the kidney (and depicted in most textbooks). Reflux can be demonstrated by voiding cystourethrography, in which the bladder is filled with a radiopaque dye and images are taken during micturition. Vesicoureteral reflux can be demonstrated in about 30% of infants and children with urinary tract infection (see Fig. 20.25).

In the absence of vesicoureteral reflux, infection usually remains localized in the bladder. Thus, the majority of individuals with repeated or persistent bacterial colonization of the urinary tract suffer from cystitis and urethritis *(lower urinary tract infection)* rather than pyelonephritis.

Acute Pyelonephritis

Acute pyelonephritis is a suppurative inflammation of the kidney caused by bacterial and sometimes viral (e.g., polyomavirus) infection, which can reach the kidney by hematogenous spread or, more commonly, through the ureters in association with vesicoureteral reflux.

MORPHOLOGY

The hallmarks of acute pyelonephritis are **patchy interstitial suppurative inflammation, intratubular aggregates of neutrophils, neutrophilic tubulitis, and tubular injury.** The suppuration may occur as discrete focal abscesses or large wedgelike areas and can involve one or both kidneys (Fig. 20.26).

In the early stages, the neutrophilic infiltration is limited to the tubules. The tubular lumens are a conduit for the extension of the infection, and the infection can extend to the interstitium and produces abscesses that destroy the involved tubules (Fig. 20.27). Characteristically, glomeruli are relatively resistant to the infection. Extensive disease, however, eventually also destroys the glomeruli, and fungal pyelonephritis (e.g., *Candida*) often results in granulomatous interstitial inflammation.

Three complications may be superimposed on acute pyelonephritis.

- *Papillary necrosis* is seen mainly in diabetics, sickle cell disease, and in those with urinary tract obstruction. Papillary necrosis is usually bilateral but may be unilateral. One or all of the pyramids of the affected kidney may be involved. On cut section, the tips or distal two thirds of the pyramids have areas of gray-white to yellow necrosis (Fig. 20.28). On microscopic examination, the necrotic tissue shows characteristic ischemic

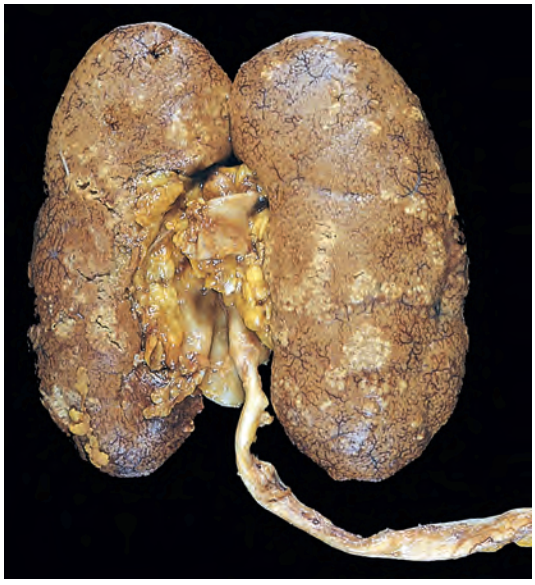


Figure 20.26 Acute pyelonephritis. Cortical surface shows grayish white areas of inflammation and abscess formation.

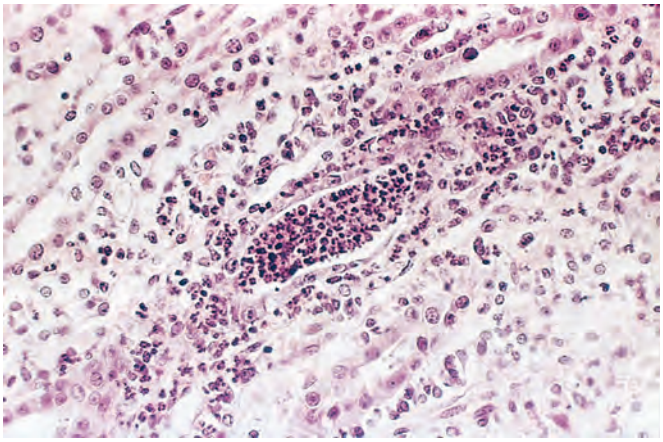


Figure 20.27 Acute pyelonephritis marked by a neutrophilic exudate within tubules and interstitial inflammation.

coagulative necrosis, with preservation of outlines of tubules. The leukocytic response is limited to the junctions between preserved and destroyed tissue.

- *Pyonephrosis* is seen when there is total or almost complete obstruction, particularly when it is high in the urinary tract. The suppurative exudate is unable to drain and fills the renal pelvis, calyces, and ureter with pus.
- *Perinephric abscess* is an extension of suppurative inflammation through the renal capsule into the perinephric tissue.

After the acute phase of pyelonephritis, healing occurs. The neutrophilic infiltrate is replaced by one that is predominantly composed of macrophages, plasma cells, and lymphocytes. The inflammatory foci are eventually replaced by irregular scars that can be seen on the cortical surface as fibrous depressions. Such scars are characterized microscopically by tubular atrophy, interstitial fibrosis, and a lymphocytic infiltrate in a characteristic patchy, jigsaw pattern with intervening preserved parenchyma. **The pyelonephritic scar is almost always associated with inflammation, fibrosis, and deformation of the underlying calyx and pelvis,** reflecting the role of ascending infection and vesicoureteral reflux in the pathogenesis of the disease.

Clinical Features

Acute pyelonephritis is often associated with the following:

- *Urinary tract obstruction,* either congenital or acquired
- *Instrumentation* of the urinary tract, most commonly catheterization
- *Vesicoureteral reflux*
- *Pregnancy.* Between 4% and 6% of pregnant women develop bacteriuria sometime during pregnancy, and 20% to 40% of these eventually develop symptomatic urinary infection if not treated.
- *Gender and age.* After the first year of life (when congenital anomalies in males commonly become evident) and up to around 40 years of age, infections are much more frequent in females. With increasing age, the incidence in males rises as a result of prostatic hypertrophy and instrumentation.
- *Preexisting renal lesions,* causing intrarenal scarring and obstruction
- *Diabetes,* in which increased susceptibility to infection, neurogenic bladder dysfunction, and more frequent instrumentation are predisposing factors
- *Immunosuppression and immunodeficiency*

Acute pyelonephritis usually presents with a sudden onset of pain at the costovertebral angle and systemic evidence of infection, such as fever and malaise. There are often indications of bladder and urethral irritation, such as dysuria, frequency, and urgency. The urine contains many leukocytes (pyuria) derived from the inflammatory infiltrate, but pyuria does not differentiate upper from lower urinary tract infection. The finding of leukocyte *casts,* typically rich in neutrophils (pus casts), indicates renal involvement, because casts are formed only in tubules. The diagnosis of infection is established by quantitative urine culture.

Uncomplicated acute pyelonephritis follows a benign course, and symptoms disappear within a few days after the institution of appropriate antibiotic therapy. Bacteria, however, may persist in the urine, or there may be recurrence of infection with new serologic types of *E. coli* or other

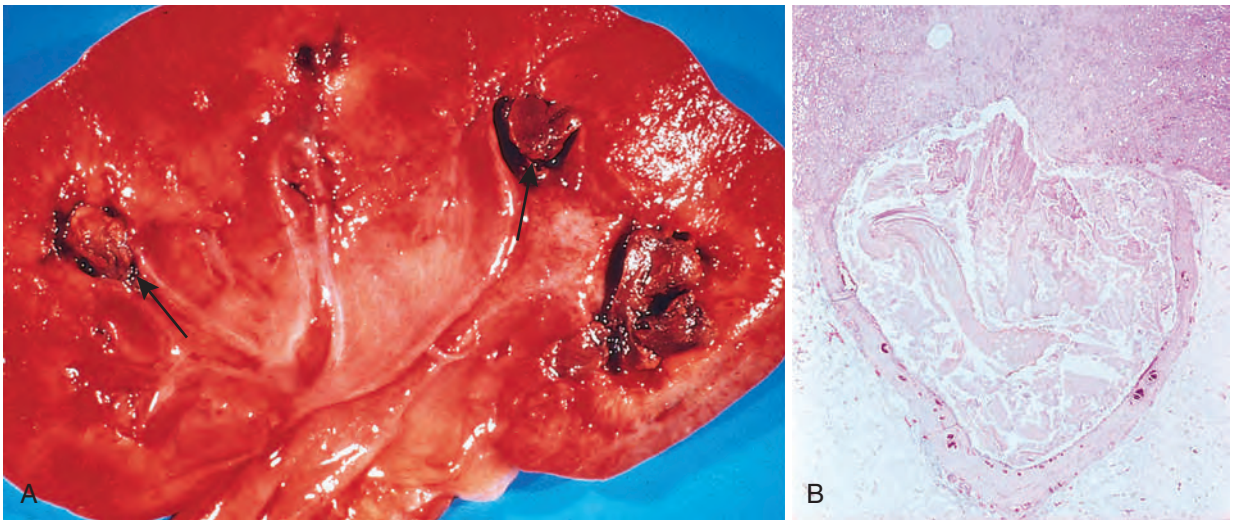


Figure 20.28 Papillary necrosis. Areas of pale-gray hemorrhage and necrosis involve the papillae *(arrows)*.

organisms. Such bacteriuria then either disappears or may persist, sometimes for years. In the presence of unrelieved urinary obstruction, diabetes mellitus, or immunodeficiency, acute pyelonephritis may be more serious, leading to repeated septicemic episodes. The superimposition of *papillary necrosis* may lead to acute renal failure.

A viral pathogen causing pyelonephritis in kidney allografts is *polyomavirus*. Latent infection with polyomavirus is widespread in the general population, and immunosuppression of the allograft recipient can lead to reactivation of latent infection and the development of nephropathy resulting in allograft failure in up to 5% of kidney transplant recipients. This form of pyelonephritis, now referred to as polyomavirus nephropathy, is characterized by infection of tubular epithelial cell nuclei, leading to nuclear enlargement and intranuclear inclusions visible by light microscopy (viral cytopathic effect). The inclusions are composed of virions arrayed in distinctive crystalline-like lattices when visualized by electron microscopy (Fig. 20.29). An interstitial inflammatory response is invariably present. Treatment consists of a reduction in immunosuppression.

Chronic Pyelonephritis and Reflux Nephropathy

Chronic pyelonephritis is a disorder in which chronic tubulointerstitial inflammation and scarring involve the calyces and pelvis (Fig. 20.30). Although several diseases produce chronic tubulointerstitial alterations (see Table 20.8), only chronic pyelonephritis and analgesic nephropathy affect the calyces, making pelvocalyceal damage an important diagnostic clue. Chronic pyelonephritis at one time accounted for 10% to 20% of patients in renal transplant or dialysis units, until predisposing conditions such as reflux became better recognized. This condition remains an important cause of kidney destruction in children with severe lower urinary tract abnormalities.

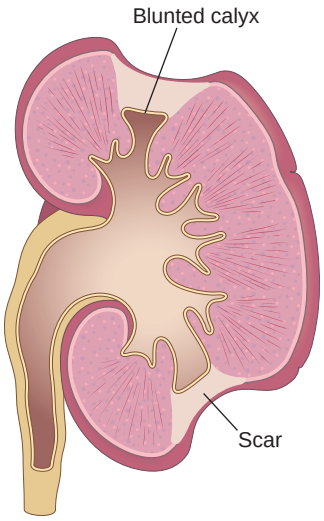


Figure 20.30 Typical coarse scars of chronic pyelonephritis associated with vesicoureteral reflux. The scars are usually polar and are associated with underlying blunted calyces.

MORPHOLOGY

The characteristic changes of chronic pyelonephritis are seen on gross examination (Figs. 20.30 and 20.31A). The kidneys usually are irregularly scarred; if bilateral, the involvement is asymmetric. In contrast, both kidneys in chronic glomerulonephritis are diffusely and symmetrically scarred. The hallmarks of chronic pyelonephritis are **coarse, discrete, corticomedullary scars overlying dilated, blunted, or deformed calyces, and flattening of the papillae.** The scars vary from one to several, and most are in the upper and lower poles, consistent with the frequency of reflux in these sites.

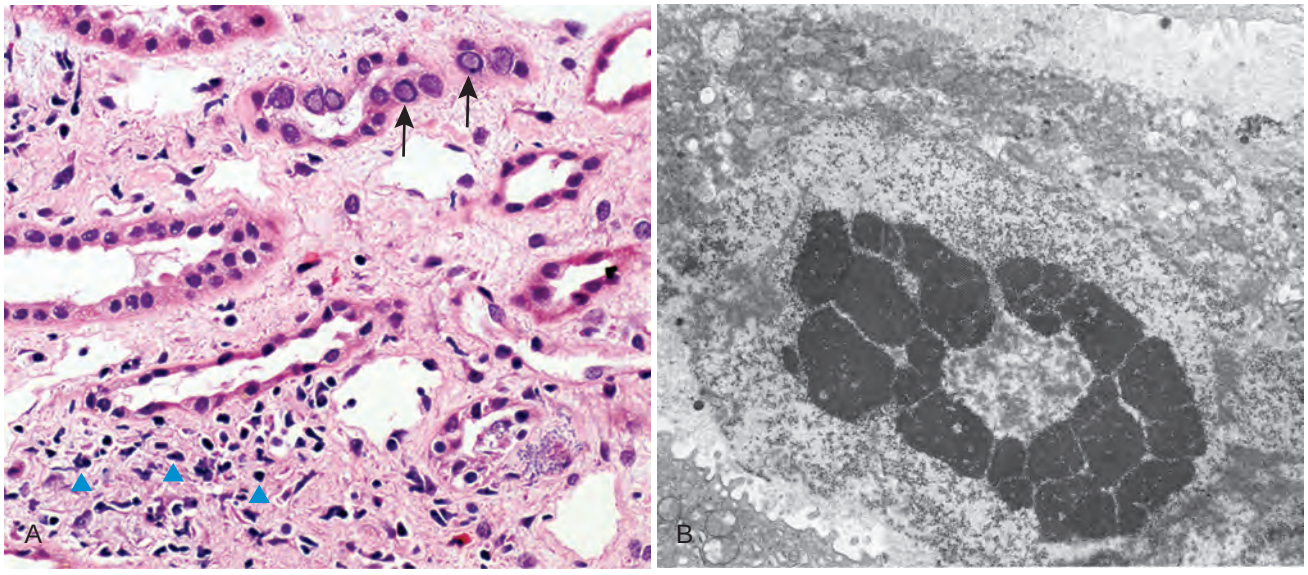


Figure 20.29 Polyomavirus nephropathy. (A) The kidney shows enlarged tubular epithelial cells with nuclear inclusions *(arrows)* and interstitial inflammation *(arrowheads)*. (B) Intranuclear viral inclusions visualized by electron microscopy. (Courtesy Dr. Jean Olson, Department of Pathology, University of California San Francisco, San Francisco, Calif.)

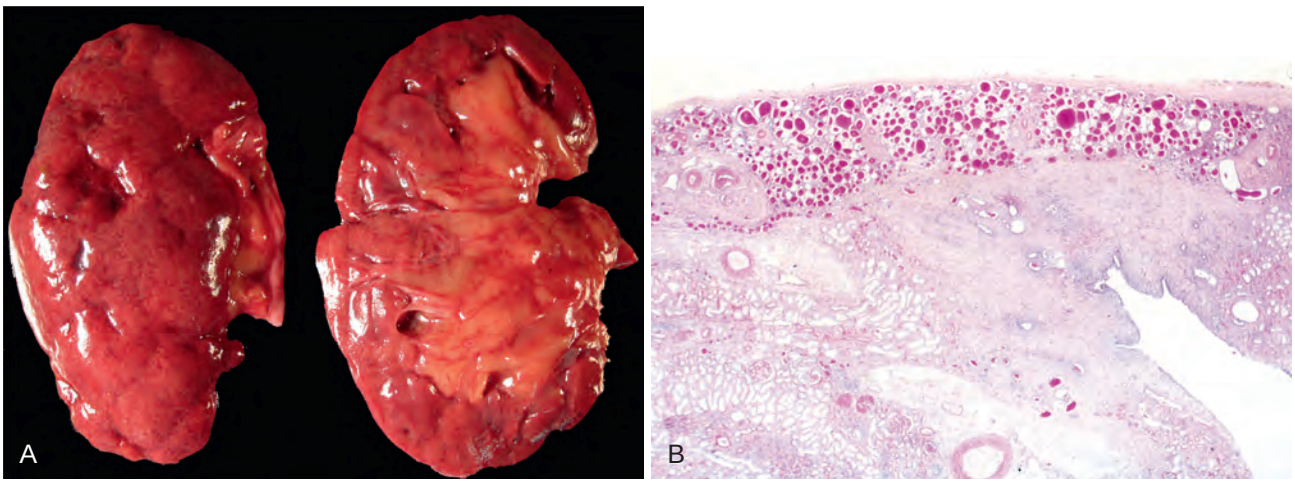


Figure 20.31 (A) Chronic pyelonephritis. The surface *(left)* is irregularly scarred. The cut section *(right)* reveals blunting and loss of several papillae. (B) Low-power view shows a corticomedullary renal scar with an underlying dilated deformed calyx. Note the thyroidization of tubules in the cortex.

The microscopic changes involve predominantly tubules and interstitium. The tubules show atrophy in some areas and hypertrophy or dilation in others. Dilated tubules with flattened epithelium may be filled with casts resembling thyroid colloid (thyroidization). There are varying degrees of chronic interstitial inflammation and fibrosis in the cortex and medulla. There is often fibrosis around the calyceal epithelium as well as a marked chronic inflammatory infiltrate. Glomeruli may appear normal except for a variety of ischemic changes, including periglomerular fibrosis, fibrous obliteration, and secondary changes related to hypertension. Individuals with chronic pyelonephritis and reflux nephropathy who develop proteinuria in advanced stages show secondary FSGS, as described later.

Xanthogranulomatous pyelonephritis is a rare form of chronic pyelonephritis characterized by accumulation of foamy macrophages intermingled with plasma cells, lymphocytes, polymorphonuclear leukocytes, and occasional giant cells. Often associated with *Proteus* infections and obstruction, the lesions sometimes produce large, yellowish orange nodules that may be grossly confused with renal cell carcinoma.

Clinical Features

Chronic obstructive pyelonephritis may have a silent onset or present with manifestations of acute recurrent pyelonephritis, such as back pain, fever, pyuria, and bacteriuria. These patients receive medical attention relatively late in their disease course because of the gradual onset of renal insufficiency and hypertension. Reflux nephropathy is often discovered in children when the cause of hypertension is investigated. Loss of tubular function—in particular of concentrating ability—gives rise to polyuria and nocturia. Radiographic studies show asymmetrically contracted kidneys with characteristic coarse scars and blunting and deformity of the calyceal system. Significant bacteriuria may be present, but it is often absent in the late stages.

Although proteinuria is usually mild, some individuals with pyelonephritic scars develop secondary *FSGS* with significant proteinuria, even in the nephrotic range, usually several years after the scarring has occurred and often in the absence of continued infection or persistent vesicoureteral reflux. The onset of proteinuria is a poor prognostic sign since it increases the likelihood of progression to ESRD. The glomerulosclerosis, as discussed, is attributable to the adaptive glomerular alterations secondary to loss of renal mass caused by pyelonephritic scarring.

KEY CONCEPTS

PYELONEPHRITIS

- Both acute and chronic pyelonephritis may be caused by infection via the ascending (more common) or hematogenous route. Obstructive lesions of the urinary tract are important predisposing factors.
- Bacteria are the most common infectious agent in acute pyelonephritis and induce a prominent neutrophilic inflammatory response; granulomatous interstitial inflammation is characteristic of fungal or mycobacterial infections.
- Chronic pyelonephritis ensues when anatomic anomalies result in urine reflux or urine outflow obstruction; multiple episodes of this injury leads to irregular scarring of the kidney that is typically more prominent at the upper or lower poles where reflux is more common.

Tubulointerstitial Nephritis Induced by Drugs and Toxins

Drug- and toxin-induced tubulointerstitial nephritis is the second most common cause of acute kidney injury (after pyelonephritis). Toxins and drugs can injure kidneys in at least three ways: (1) trigger an interstitial immunologic reaction, exemplified by the acute hypersensitivity nephritis induced by drugs such as methicillin; (2) cause ATI, as described earlier; and (3) cause subclinical but cumulative injury to tubules that takes years to result in chronic kidney disease. The last type of damage is especially worrisome, because it may be unrecognized until the renal damage is irreversible.

Acute Drug-Induced Interstitial Nephritis

First reported after the use of sulfonamides, acute tubulointerstitial nephritis most frequently occurs with synthetic penicillins (methicillin, ampicillin), other synthetic antibiotics (rifampin), diuretics (thiazides), nonsteroidal antiinflammatory drugs (NSAIDs), and miscellaneous other drugs (allopurinol, cimetidine, checkpoint inhibitors). The chronic tubulointerstitial nephritis caused by phenacetin-containing analgesics, termed *analgesic nephropathy,* is of historical importance as its incidence has substantially diminished due to the withdrawal or restriction of phenacetin in most countries.

Drug-induced acute interstitial nephritis begins 2 to 40 days after drug exposure and is characterized by *fever, eosinophilia* (which may be transient), *rash* in about 25% of patients, and *renal abnormalities*. The latter takes the form of hematuria, mild proteinuria, and leukocyturia (often including eosinophils). A rising serum creatinine or acute kidney injury with oliguria develops in about 50% of cases.

Pathogenesis

Many features of the disease suggest an idiosyncratic immune mechanism. Clinical evidence of hypersensitivity includes the latent period, eosinophilia and rash, the fact that the onset of nephropathy is not dose-related, and recurrence of clinical and pathologic manifestations after re-exposure to the same or a chemically related drug. In some patients, serum IgE levels are increased, and IgE-containing plasma cells and basophils are present in the lesions, suggesting that the *late-phase reaction of an IgE-mediated (type I) hypersensitivity* may be involved in the pathogenesis (Chapter 6). In other cases, a mononuclear or granulomatous reaction, together with positive results of skin tests to drugs, suggest a T cell–mediated (type IV) delayed-hypersensitivity reaction.

The most likely sequence of events is that the drugs function as haptens and covalently bind to some plasma membrane or extracellular component of tubular cells. These modified self antigens then become immunogenic. The resultant injury is due to IgE or cell-mediated immune reactions directed against the tubular cells or their basement membranes.

MORPHOLOGY

The interstitium shows variable but frequently pronounced **edema and infiltration by mononuclear cells,** principally lymphocytes and macrophages. Eosinophils and neutrophils may be present (Fig. 20.32), often in clusters and large numbers, and smaller numbers of plasma cells and mast cells are sometimes also present. Inflammation may be more prominent in the medulla where the inciting agent is concentrated. With some drugs (e.g., methicillin, thiazides), interstitial non-necrotizing granulomas may be seen when tubules rupture. Tubulitis, the infiltration of tubules by lymphocytes, is common. Variable degrees of tubular injury and regeneration are present. The glomeruli are normal except in some cases caused by NSAIDs, when minimal change disease and the nephrotic syndrome develop concurrently (see below).

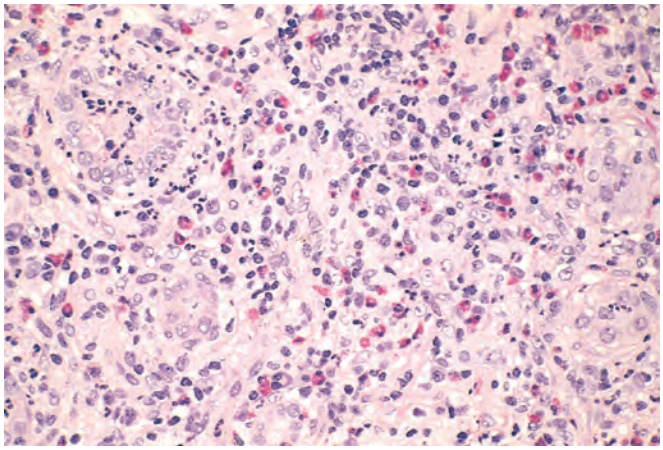


Figure 20.32 Drug-induced interstitial nephritis, with prominent eosinophilic and mononuclear cell infiltrate. (Courtesy Dr. H. Rennke, Brigham and Women's Hospital, Boston, Mass.)

Clinical Features

It is important to recognize drug-induced acute interstitial nephritis because withdrawal of the offending drug is followed by recovery, although it may take several months, and irreversible damage can occur. While drugs are the leading cause of acute interstitial nephritis, in many affected patients (approximately 30% to 40%) an offending drug or mechanism cannot be identified.

On occasion, necrotic papillae are excreted and may cause gross hematuria or renal colic due to ureteric obstruction. Papillary necrosis is not specific for analgesic nephropathy and is also seen in diabetes, as well as in urinary tract obstruction, sickle cell disease or trait (described later), and focally in renal tuberculosis. In all cases, it is caused by ischemia resulting from compression or obstruction of small blood vessels in the medulla. Such compression may be caused by interstitial edema (as in inflammatory reactions and urinary tract obstruction) or microvascular disease (as in diabetes). Table 20.9 lists the main features of papillary necrosis in these conditions.

Nephropathy Associated With NSAIDs

NSAIDs, one of the most commonly used classes of drugs, produce several forms of renal injury. Although these complications are uncommon, they should be kept in mind because NSAIDs are frequently administered to patients with other potential causes of renal disease. Many NSAIDs are nonselective cyclo-oxygenase inhibitors, and their adverse renal effects are related to their ability to inhibit cyclo-oxygenase–dependent prostaglandin synthesis. The selective COX-2 inhibitors, while sparing the gastrointestinal tract, do affect the kidneys because COX-2 is expressed in human kidneys. NSAID-associated renal syndromes include the following:

- *Acute kidney injury,* due to the decreased synthesis of vasodilatory prostaglandins and resultant ischemia. This is particularly likely to occur in the setting of other renal diseases or conditions causing volume depletion.
- *Acute hypersensitivity interstitial nephritis,* resulting in renal failure, as described earlier.
- *Acute interstitial nephritis and minimal change disease.* This curious association of two diverse renal conditions, one

Table 20.9 Causes of Papillary Necrosis

	Diabetes	Analgesic Nephropathy[a]	Sickle Cell Disease	Obstruction
Male-to-female ratio	1:3	1:5	1:1	9:1
Time course	10 years	>5 years of abuse	Variable	Variable
Infection	80%	25%	±	90%
Calcification	Rare	Frequent	Rare	Frequent
Number of papillae affected	Several; all of same stage	Almost all; different stages of necrosis	Few	Variable

[a]The incidence of this disease has greatly decreased since the 1990s because its main cause, phenacetin, has been withdrawn from the market in most countries, and combination analgesics are no longer available without prescription.

Data from Seshan S, et al, editors: *Classification and Atlas of Tubulointerstitial and Vascular Diseases,* Baltimore, 1999, Williams & Wilkins.

leading to renal failure and the other to nephrotic syndrome, suggests a hypersensitivity reaction affecting the interstitium and possibly the glomeruli, but also is consistent with injury to podocytes mediated by cytokines released as part of the inflammatory process.

- *Membranous nephropathy,* with the nephrotic syndrome, is a recently appreciated association, also of unclear pathogenesis.

KEY CONCEPTS

TUBULOINTERSTITIAL NEPHRITIS INDUCED BY DRUGS AND TOXINS

- Drug-induced tubulointerstitial nephritis is the second most common cause of acute kidney injury.
- Prominent interstitial inflammation with associated tubular injury, which may or may not be accompanied by eosinophils or granulomatous inflammation, can be induced by almost any pharmacologic agent.
- NSAIDs can cause tubulointerstitial nephritis and/or glomerular injury, such as minimal change disease or membranous nephropathy.

Other Tubulointerstitial Diseases

Urate Nephropathy

Three types of nephropathy can occur in persons with hyperuricemic disorders:

- *Acute uric acid nephropathy* is caused by the precipitation of uric acid crystals in the renal tubules, principally in collecting ducts, leading to obstruction of nephrons and the development of acute renal failure. This is particularly likely to occur in individuals with leukemias or lymphomas who are undergoing chemotherapy (tumor lysis syndrome); the drugs kill tumor cells, and uric acid is produced as released nucleic acids are broken down. Precipitation of uric acid is favored by the acidic pH in collecting tubules.
- *Chronic urate nephropathy,* or gouty nephropathy, occurs rarely in protracted forms of hyperuricemia. The monosodium urate crystals deposit in the acidic milieu of the distal tubules and collecting ducts, and form distinct birefringent needlelike crystals either in the tubular lumens or in the interstitium (Fig. 20.33). The urate deposits evoke a mononuclear response that contains foreign-body giant cells. This lesion is called a *tophus* (Chapter 26). Tubular obstruction by the urates causes cortical atrophy and scarring. Clinically, urate nephropathy is a subtle disease associated with tubular defects that may progress slowly. Some individuals with gout who develop a chronic nephropathy have evidence of increased exposure to lead.
- *Nephrolithiasis:* uric acid stones are present in 22% of individuals with gout and 42% of those with secondary hyperuricemia (see later discussion of renal stones).

Hypercalcemia and Nephrocalcinosis

Disorders associated with hypercalcemia, such as hyperparathyroidism, multiple myeloma, vitamin D intoxication, metastatic cancer, or excess calcium intake (milk-alkali syndrome), may induce the formation of calcium stones and deposition of calcium in the kidney (nephrocalcinosis). Extensive degrees of calcinosis, under certain conditions, may lead to chronic tubulointerstitial disease and renal insufficiency.

The earliest functional defect is an inability to concentrate the urine. Other tubular defects, such as tubular acidosis and salt-losing nephritis, may also occur. With further damage, a slowly progressive chronic kidney disease develops. This is usually due to nephrocalcinosis, but many of these patients also have calcium stones and secondary pyelonephritis.

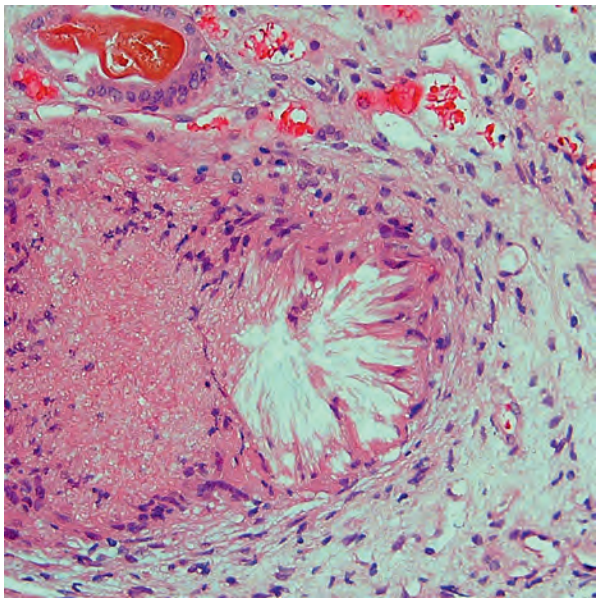


Figure 20.33 Granulomatous inflammation outlines the slender urate crystals in the renal medulla.

Autosomal Dominant Tubulointerstitial Kidney Disease

Autosomal dominant tubulointerstitial kidney disease (ADTKD) was previously known as medullary cystic kidney disease. Given the variable presence of cysts, which often are not located in the medulla, and the distinctive genetics, it is now recognized as ADTKD due to following genetic mutations:

- *MUC1* encodes mucin-1 (expressed in distal nephrons)
- *UMOD* encodes uromodulin (expressed in thick ascending limb of loop of Henle)
- *REN* encodes preprorenin (expressed by juxtaglomerular apparatus)
- *HNF1β* encodes hepatocyte nuclear factor 1β, a transcription factor that regulates multiple genes, including *UMOD*.

ADTKD is associated with nonspecific clinical and pathologic findings and causes progressive renal failure in adult life with an autosomal dominant pattern of transmission. The pathogenic mechanisms are unknown.

Light-Chain Cast Nephropathy ("Myeloma Kidney")

Nonrenal malignant tumors, particularly those of hematopoietic origin, affect the kidneys in several ways (Table 20.10). The most common involvements are tubulointerstitial, caused by complications of the tumor (hypercalcemia, ureteral obstruction) or therapy (irradiation, hyperuricemia, chemotherapy, hematopoietic cell transplantation, infections in immunosuppressed patients). We limit the discussion here to the tubulointerstitial lesions in *multiple myeloma* patients.

Overt renal insufficiency occurs in half of those with multiple myeloma and related lymphoplasmacytic disorders. Several factors contribute to renal damage:

- *Bence-Jones proteinuria and cast nephropathy*. The main cause of renal dysfunction is related to Bence-Jones (light-chain) proteinuria and correlates with the degree of proteinuria. Two mechanisms seem to account for the renal toxicity of Bence-Jones proteins. First, some Ig light chains are directly toxic to epithelial cells, because of their intrinsic physicochemical properties. Second, Bence-Jones proteins combine with urinary glycoprotein (Tamm-Horsfall protein) under acidic conditions to form large, histologically distinct tubular casts that obstruct the tubular lumens and induce a characteristic inflammatory reaction (light-chain cast nephropathy).
- *Amyloidosis* of AL type, formed from free light chains (usually of λ type), occurs in 6% to 24% of individuals with myeloma.
- *Light-chain deposition disease*. In some patients, light chains (usually of κ type) deposit in GBMs and mesangium in nonfibrillar forms, causing a glomerulopathy (described earlier), and in tubular basement membranes, which may cause tubulointerstitial nephritis.
- *Hypercalcemia* and *hyperuricemia* are often present in these patients.

Table 20.10 Renal Disease Related to Nonrenal Neoplasms

Direct or Metastatic Tumor Invasion of Renal Parenchyma
Ureters (obstruction) Artery (renovascular hypertension)
Hypercalcemia
Hyperuricemia
Amyloidosis (AL, Light-Chain Type)
Excretion of Abnormal Proteins (Multiple Myeloma)
Glomerulopathies
Membranous nephropathy, secondary (carcinomas) Minimal change disease (Hodgkin disease) Membranoproliferative glomerulonephritis (leukemias and lymphomas) Monoclonal immunoglobin/light-chain deposition disease (multiple myeloma)
Effects of Radiation Therapy, Chemotherapy, Hematopoietic Cell Transplantation, Secondary Infection

MORPHOLOGY

The tubulointerstitial changes in light-chain cast nephropathy are characteristic. The Bence-Jones tubular casts appear as pink to blue amorphous masses, sometimes concentrically laminated and often fractured, which fill and distend the tubular lumens. Some casts are surrounded by multinucleated giant cells that are derived from activated macrophages (Fig. 20.34). The adjacent interstitial tissue usually shows an inflammatory response and fibrosis. On occasion, the casts rupture the tubules, evoking a granulomatous inflammatory reaction. Amyloidosis, light-chain deposition disease, nephrocalcinosis, and infection may also be present.

Clinical Features

Clinically, the renal manifestations are of several types. In the most common form, *chronic kidney disease* develops insidiously and progresses slowly during a period of several months to years. Another form occurs suddenly and is manifested by *acute kidney injury* with oliguria. Precipitating factors include dehydration, hypercalcemia, acute infection, and treatment with nephrotoxic antibiotics. *Bence-Jones proteinuria* occurs in 70% of individuals with multiple myeloma; the presence of significant non–light-chain proteinuria (e.g., albuminuria) suggests AL amyloidosis or light-chain deposition disease.

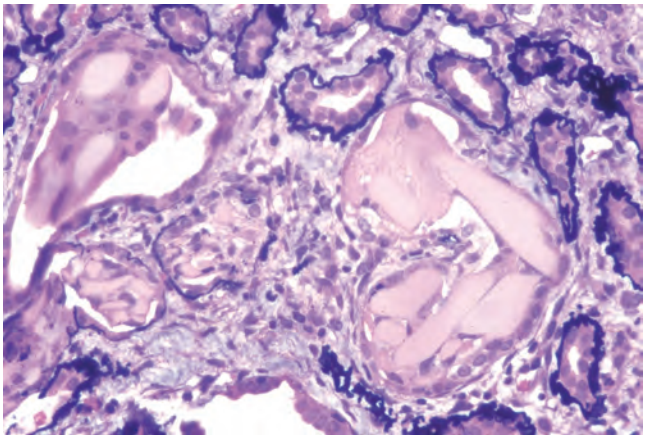


Figure 20.34 Light-chain cast nephropathy. Note the angulated and tubular casts, surrounded by macrophages, including multinucleated cells.

Bile Cast Nephropathy

Impairment of renal function often occurs in patients with severe acute or advanced chronic liver disease. In this setting, serum bilirubin levels can be markedly elevated, particularly in jaundiced patients, with bile cast formation (also known as cholemic nephrosis) in distal nephron segments. The casts can extend to proximal tubules, resulting in both direct toxic effects on tubular epithelial cells and obstruction of the involved nephron. This mechanism of injury is analogous to that with monoclonal immunoglobulin and myoglobin casts. The tubular bile casts can range from yellowish-green to red-pink and contain variable degrees of sloughed cells or cellular debris. The reversibility of the renal injury depends on the severity and duration of the liver dysfunction.

VASCULAR DISEASES

Nearly all diseases of the kidney involve the renal blood vessels secondarily. Systemic vascular diseases, such as various forms of vasculitis, also affect renal vessels, and their effects on the kidney are clinically important. Hypertension, as discussed in Chapter 11, is intimately linked with the kidney, because kidney disease can be both a cause and consequence of increased blood pressure. In this chapter, we discuss nephrosclerosis and renal artery stenosis, lesions associated with hypertension, and sundry lesions involving mostly smaller vessels of the kidney.

Nephrosclerosis

***Nephrosclerosis* is the term used for the renal pathology associated with sclerosis of renal arterioles and small arteries; it is strongly associated with hypertension, which can be both a cause and a consequence of nephrosclerosis.** Nephrosclerosis at autopsy is associated with advanced age, is more frequent in blacks than in whites, and may be seen in the absence of hypertension. Hypertension and diabetes, however, increase the incidence and severity of the lesions.

Pathogenesis

Two processes participate in the arterial lesions:

- Medial and intimal thickening, a response to hemodynamic changes, aging, genetic defects, or some combination of these
- Hyalinization of arteriolar walls, caused by extravasation of plasma proteins through injured endothelium and by increased deposition of basement membrane matrix

Because of thickened walls, the affected vessels have narrowed lumens, which results in focal parenchymal ischemia. Ischemia leads to glomerulosclerosis and chronic tubulointerstitial injury, and it produces a reduction in functional renal mass.

MORPHOLOGY

The kidneys are either normal or moderately reduced in size, with average weights between 110 and 130 g. The cortical surfaces have a fine, even granularity that resembles grain leather (Fig. 20.35). The loss of mass is due mainly to **cortical scarring and shrinking.**

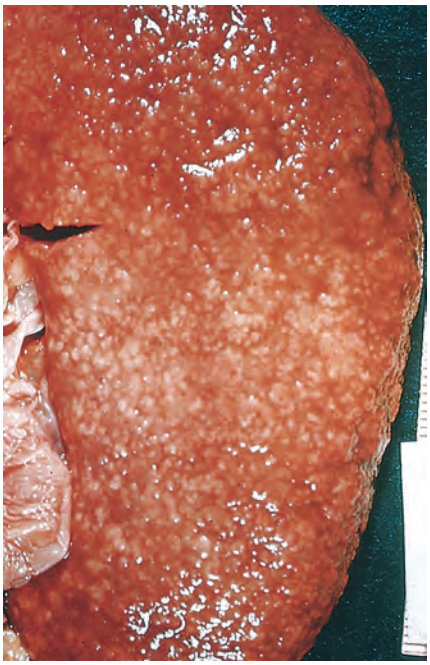


Figure 20.35 Close-up of the gross appearance of the cortical surface in benign nephrosclerosis illustrating the fine, leathery granularity of the surface.

On histologic examination, there is narrowing of the lumens of arterioles and small arteries, caused by thickening and hyalinization of the walls **(hyaline arteriolosclerosis)** (Fig. 20.36). Corresponding to the finely granular surface are microscopic subcapsular scars with sclerotic glomeruli and tubular dropout, alternating with better preserved parenchyma. In addition, the interlobular and arcuate arteries show medial hypertrophy, replication of the internal elastic lamina, and increased myofibroblastic tissue in the intima, all of which narrow the lumen. This change, called fibroelastic hyperplasia, often accompanies hyaline arteriolosclerosis and increases in severity with age and in the presence of hypertension.

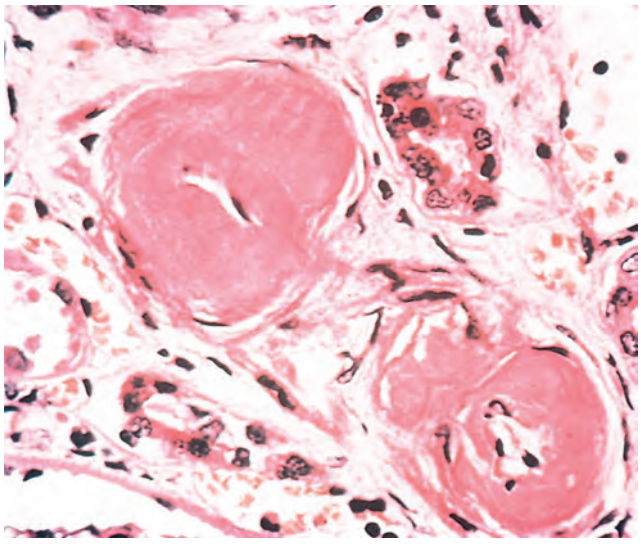


Figure 20.36 Hyaline arteriolosclerosis. High-power view of two arterioles with hyaline deposition, marked thickening of the walls, and a narrowed lumen. (Courtesy Dr. M. A. Venkatachalam, Department of Pathology, University of Texas Health Sciences Center, San Antonio, Tex.)

Consequent to the vascular narrowing, there is patchy ischemic atrophy, which consists of (1) foci of **tubular atrophy and interstitial fibrosis** and (2) a variety of **glomerular alterations.** The latter include collapse of the GBM, deposition of collagen within the Bowman space, periglomerular fibrosis, and total sclerosis of glomeruli. When the ischemic changes are pronounced and affect large areas of parenchyma, they can produce wedge-shaped infarcts or regional scars with histologic alterations that may resemble those in renal ablation injury, mentioned earlier.

Clinical Features

It is unusual for uncomplicated nephrosclerosis to cause renal insufficiency or uremia. However, three groups of hypertensive patients with nephrosclerosis are at increased risk of developing renal failure: people of African descent, people with severe blood pressure elevations, and persons with a second underlying disease, especially diabetes. In these groups, renal insufficiency may supervene after prolonged hypertension, but rapid renal failure may result from the development of the malignant or accelerated phase of hypertension.

A small percentage of hypertensive persons (as many as 5%) show a rapidly rising blood pressure that, if untreated, leads to death within 1 to 2 years. This form of hypertension, called *malignant hypertension,* is characterized by severe pressure elevations (i.e., systolic pressure more than 200 mm Hg, diastolic pressure more than 120 mm Hg), renal failure, and retinal hemorrhages and exudates, with or without papilledema (swelling of the optic nerve that reflects increased intracranial pressures). Renal lesions associated with malignant hypertension have been referred to as malignant nephrosclerosis. It should be noted, however, that there is considerable clinical and morphologic overlap between the renal pathology of malignant hypertension and thrombotic microangiopathies. About 30% of cases of malignant hypertension have microangiopathic hemolytic anemia (discussed below) and conversely, severe hypertension can occur in primary forms of hemolytic uremic syndrome. Endothelial injury is a common pathogenic factor in these disorders.

KEY CONCEPTS

NEPHROSCLEROSIS

- Nephrosclerosis, which is commonly associated with hypertension, is defined by the presence of varying degrees of glomerulosclerosis, interstitial fibrosis and tubular atrophy, arteriosclerosis, and arteriolosclerosis.
- Luminal reduction of the renal vasculature (arteries and arterioles) contributes to glomerulosclerosis (both global and segmental), which can subsequently cause interstitial fibrosis and tubular atrophy.

Renal Artery Stenosis

Unilateral renal artery stenosis is responsible for 2% to 5% of hypertension cases, and it is important to recognize because it is potentially curable by surgery. Furthermore, important insights into renal mechanisms of hypertension came from studies of experimental and human renal artery stenosis.

Pathogenesis

Hypertension secondary to renal artery stenosis is caused by increased production of renin from the ischemic kidney. The classic experiments of Goldblatt and colleagues showed that constriction of one renal artery in dogs results in hypertension and that the magnitude of the effect is proportional to the amount of narrowing. Elevation in blood pressure, at least initially, is due to stimulation of renin secretion by the juxtaglomerular apparatus and the subsequent production of the vasoconstrictor angiotensin II. A large proportion of individuals with renovascular hypertension have elevated renin levels, and almost all show a reduction of blood pressure when given drugs that block angiotensin II activity. Other factors, however, may contribute to the maintenance of renovascular hypertension after the renin-angiotensin system has initiated it, including *sodium retention*.

MORPHOLOGY

The most common cause of renal artery stenosis (70% of cases) is narrowing at the origin of the renal artery by an **atheromatous plaque.** This occurs more frequently in men, and the incidence increases with advancing age and diabetes. The plaque is usually concentrically placed, and superimposed thrombosis often occurs.

The second most frequent cause of stenosis is **fibromuscular dysplasia** of the renal artery. This heterogeneous entity is characterized by fibrous or fibromuscular thickening that may involve the intima, the media, or the adventitia of the artery (Fig. 20.37). The stenoses, as a whole, are more common in women and tend to occur in younger age groups (i.e., in the third and fourth decades).

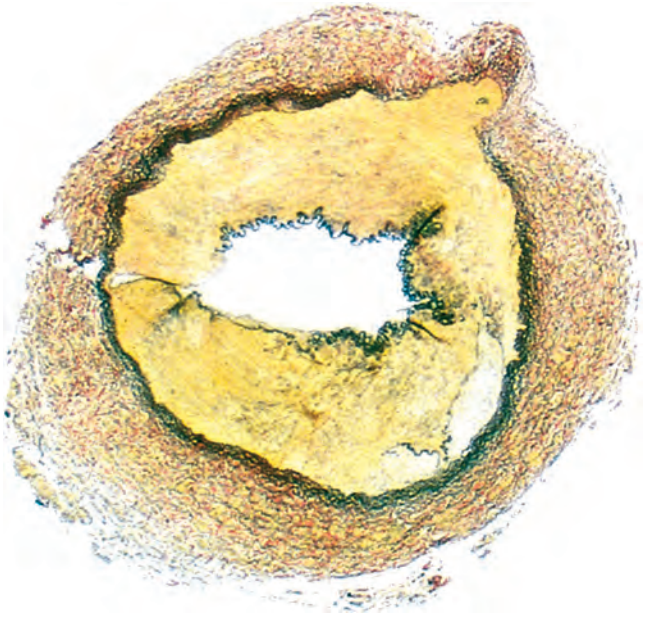


Figure 20.37 Fibromuscular dysplasia of the renal artery, medial type (elastic tissue stain). The media shows marked fibrous thickening, and the lumen is stenotic. (Courtesy Dr. Seymour Rosen, Beth Israel Hospital, Boston, Mass.)

The ischemic kidney is reduced in size and shows signs of **diffuse ischemic atrophy,** with crowded glomeruli, atrophic tubules, interstitial fibrosis, and focal inflammatory infiltrates. The arterioles in the ischemic kidney are usually protected from the effects of high pressure, thus showing only mild arteriolosclerosis. In contrast, the contralateral nonischemic kidney may show more severe arteriolosclerosis, depending on the severity of the hypertension.

Clinical Features

Few distinctive features suggest the presence of renal artery stenosis, and in general, these patients resemble those with essential hypertension. On occasion, a bruit can be heard on auscultation of the affected kidneys. Elevated plasma or renal vein renin, response to angiotensin-converting enzyme inhibitor, renal scans, and intravenous pyelography may aid with diagnosis, but arteriography is required to localize the stenotic lesion. The cure rate after surgery is 70% to 80% in well-selected cases.

Thrombotic Microangiopathies

The term *thrombotic microangiopathy* encompasses a spectrum of clinical syndromes that includes thrombotic thrombocytopenic purpura (TTP) and hemolytic-uremic syndrome (HUS). As discussed in Chapter 14, **HUS and TTP are caused by diverse insults that lead to thrombi in capillaries and/or arterioles in various tissue beds, including those of the kidney** (Fig. 20.38). The thrombi create flow abnormalities that shear red cells, producing a microangiopathic hemolytic anemia. Of greater importance, the thrombi produce microvascular occlusions that cause tissue ischemia and organ dysfunction. Widespread "consumption" of platelets leads to thrombocytopenia.

This group of disorders is classified according to the current understanding of their causes or associations (Table 20.11):

- *Typical HUS (synonyms: epidemic, classic, diarrhea-positive)* is most frequently associated with consumption of food contaminated by bacteria producing Shiga-like toxins.

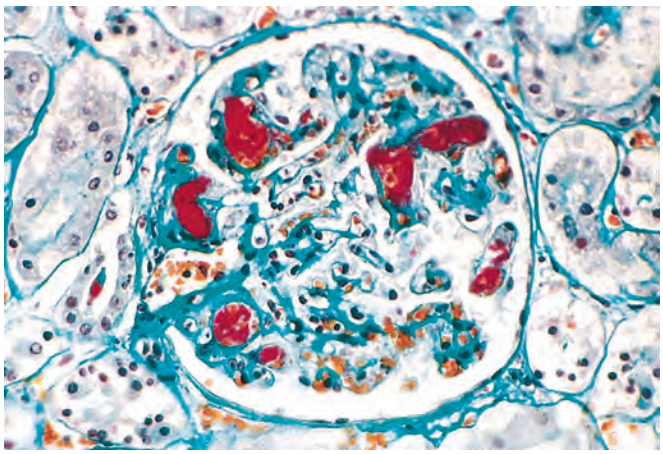


Figure 20.38 Fibrin stain showing thrombi *(red)* in the glomerular capillaries, characteristic of thrombotic microangiopathic disorders.

Table 20.11 Etiologic Classification of the Major Forms of Primary Thrombotic Microangiopathy

	Forms	Etiology
Shiga toxin–mediated HUS	Acquired	Shiga toxin–producing *E. coli*, *Shigella dysenteriae* serotype 1
Atypical HUS (complement-mediated TMA)	Inherited	Complement dysregulation due to genetic abnormalities (relatively common)
	Acquired	Acquired complement dysregulation due to autoantibodies
TTP	Inherited	Genetic ADAMTS13 deficiency (rare)
	Acquired	ADAMTS13 deficiency due to autoantibodies (relatively common)

ADAMTS13, von Willebrand factor cleaving protease; *HUS*, hemolytic uremic syndrome; *TMA*, thrombotic microangiopathy; *TTP*, thrombotic thrombocytopenic purpura.

- *Atypical HUS (synonyms: non-epidemic, diarrhea-negative)* is associated with the following:
  - Inherited mutations of or autoantibodies targeting complement-regulatory proteins
  - Diverse acquired causes of endothelial injury, including: antiphospholipid antibodies; complications of pregnancy and oral contraceptives; vascular renal diseases such as scleroderma and malignant hypertension; chemotherapeutic and immunosuppressive drugs; and radiation
- *TTP* is characterized by inherited or acquired deficiencies of ADAMTS13, a plasma metalloprotease that regulates the function of von Willebrand factor (vWF).

Pathogenesis

Within the thrombotic microangiopathies, two pathogenetic triggers dominate: (1) *endothelial injury* and (2) *excessive platelet activation and aggregation*. As we will discuss, endothelial injury appears to be the primary cause of HUS, whereas platelet activation may be the inciting event in TTP.

Endothelial Injury. In typical (epidemic, classic, diarrhea-positive) HUS, the trigger for endothelial injury and activation is usually a Shiga-like toxin, while for inherited forms of atypical HUS the cause of the endothelial injury appears to be excessive, inappropriate activation of complement. Many other exposures and conditions can occasionally precipitate a HUS-like picture, presumably also by injuring the endothelium. The endothelial injury in HUS appears to cause platelet activation and thrombosis within microvascular beds. There is evidence that reduced endothelial production of prostaglandin I_2 and nitric oxide (both inhibitors of platelet aggregation) contributes to thrombosis. The reduction in these two factors and increased production of endothelium-derived endothelin may also promote vasoconstriction, exacerbating the hypoperfusion of tissues.

Platelet Aggregation. In contrast to HUS, in TTP the initiating event appears to be platelet aggregation induced by very large multimers of vWF, which accumulate due to a deficiency of ADAMTS13, a plasma protease that cleaves vWF multimers into smaller sizes. The deficiency of ADAMTS13 is

most often caused by autoantibodies that inhibit ADAMTS13 function. Less commonly, a chronic relapsing and remitting form of TTP is associated with inherited deficiencies of ADAMTS13. Very large vWF multimers can bind platelet surface glycoproteins and activate platelets spontaneously, providing a pathophysiologic explanation for the microthrombi that are observed in vascular beds.

With this as an introduction, we will now briefly delve into the various subtypes of HUS/TTP and then return to the morphologic features that are common to all.

Typical (Epidemic, Classic, Diarrhea-Positive) Hemolytic Uremic Syndrome

This is the best-characterized form of HUS. Most cases occur after intestinal infection with strains of *E. coli* (the most common being O157:H7) that produce Shiga-like toxins, so-called because they resemble those made by *Shigella dysenteriae* (Chapter 17). Epidemics have been traced to various sources, most commonly the ingestion of contaminated ground meat (as in hamburgers), but also drinking water, raw milk, and person-to-person transmission. However, most cases of typical HUS caused by *E. coli* are sporadic. Less commonly, infections by other agents, including *S. dysenteriae*, can give rise to a similar clinical picture.

Typical HUS can occur at any age, but children and older adults are at highest risk. Following a prodrome of influenza-like or diarrheal symptoms, there is a sudden onset of bleeding manifestations (especially hematemesis and melena), severe oliguria, and hematuria, associated with microangiopathic hemolytic anemia, thrombocytopenia, and (in some patients) prominent neurologic changes. Hypertension is present in about half of patients.

Precisely how Shiga-like toxin exposure causes HUS is not well understood. According to one model, the toxin "activates" endothelial cells, which respond by increasing their expression of leukocyte adhesion molecules and endothelin and decreasing nitric oxide production. In the presence of cytokines such as TNF, Shiga-like toxin may cause endothelial apoptosis. These alterations lead to platelet activation and induce vasoconstriction, resulting in the characteristic microangiopathy. But other possibilities remain. For example, there is some evidence that Shiga-like toxins may bind and activate platelets directly; or alternatively, may bind the regulatory complement protein Factor H and inhibit its activity, causing hyperactivation of complement, an intriguing idea given the clear role of complement activation in some forms of atypical HUS (described in the next section).

In typical HUS, if the renal failure is managed properly with dialysis, most patients recover normal renal function in a matter of weeks. However, due to underlying renal damage, the long-term (15- to 25-year) outlook is more guarded. In one study, only 10 of 25 patients with prior typical HUS had normal renal function, and 7 had chronic kidney disease.

Atypical (Non-Epidemic, Diarrhea-Negative) Hemolytic Uremic Syndrome

Atypical HUS occurs in a number of different settings. More than one-half of those affected have an inherited deficiency of complement-regulatory proteins, most commonly Factor H, which breaks down the alternative pathway C3 convertase and protects cells from damage by uncontrolled complement activation (Chapter 3). A number of patients have mutations in other proteins that regulate complement, complement Factor I and CD46 (membrane cofactor protein). Having multiple genetic mutations in complement-regulatory proteins may cause atypical HUS to manifest at a younger age. Autoantibodies against complement regulatory proteins also result in atypical HUS.

The remaining cases of atypical HUS arise in association with a variety of miscellaneous conditions or exposures. These include the following:

- The *antiphospholipid syndrome,* either primary or secondary to SLE (lupus anticoagulant). The syndrome is described in detail in Chapter 4. In this setting, the microangiopathy tends to follow a chronic course.
- Complications of pregnancy or the postpartum period. So-called *postpartum renal failure* is a form of HUS that usually occurs after an uneventful pregnancy, 1 day to several months after delivery. The condition has a grave prognosis, although recovery can occur in milder cases.
- *Vascular diseases affecting the kidney,* such as systemic sclerosis and malignant hypertension (Fig. 20.39).

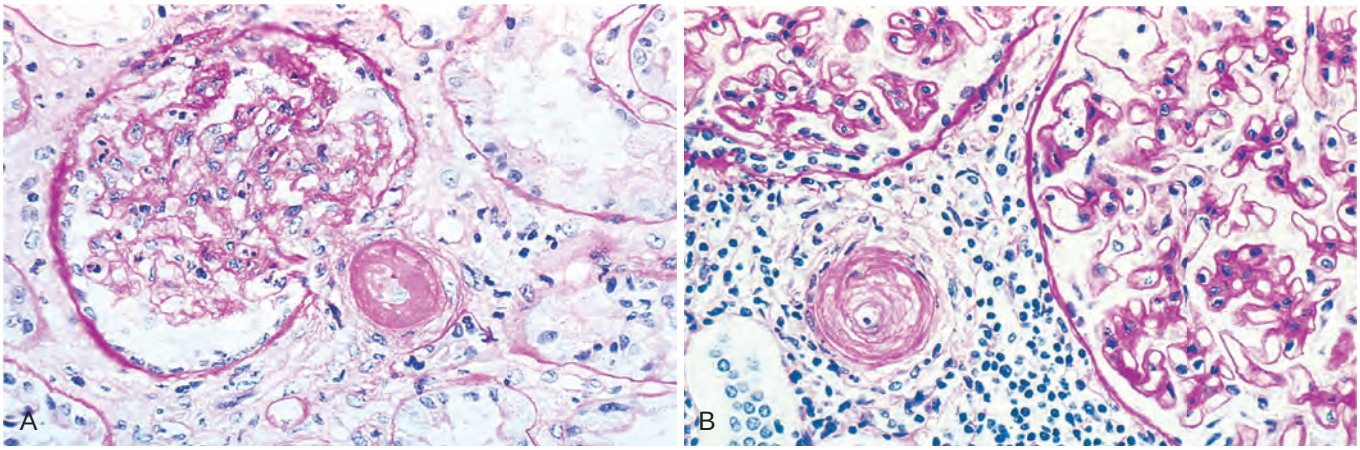


Figure 20.39 Thrombotic microangiopathy associated with malignant hypertension. (A) Fibrinoid necrosis of afferent arteriole (periodic acid–Schiff stain). (B) Hyperplastic arteriolitis (onion-skin lesion). (Courtesy Dr. H. Rennke, Brigham and Women's Hospital, Boston, Mass.)

- Chemotherapeutic and immunosuppressive drugs, such as mitomycin, cyclosporine, cisplatin, gemcitabine, and antagonists of VEGF.
- Irradiation of the kidney.

Patients with atypical HUS do not fare as well as those with typical HUS, in large part because the underlying conditions may be chronic and difficult to treat. As in typical HUS, some patients have neurologic symptoms; the disease in these patients can be distinguished from TTP by the presence of ADAMTS13 levels in the plasma of greater than 10% (see later).

Thrombotic Thrombocytopenic Purpura

TTP is classically manifested by the pentad of fever, neurologic symptoms, microangiopathic hemolytic anemia, thrombocytopenia, and renal failure. The most common cause of deficient ADAMTS13 activity is inhibitory autoantibodies, and the majority of individuals with such antibodies are women. Regardless of the cause, most patients present as adults younger than 40 years of age. Less commonly, patients inherit an inactivating mutation in *ADAMTS13*. In those with hereditary ADAMTS13 deficiency, the onset is often delayed until adolescence, and the symptoms are episodic. Thus, factors other than ADAMTS13 (e.g., some superimposed vascular injury or prothrombotic state) must be involved in triggering full-blown TTP.

For unknown reasons, in TTP, central nervous system involvement is the dominant feature, whereas renal involvement is seen in about 50% of patients. The clinical findings are dictated by the distribution of the microthrombi, which are found in arterioles throughout the body. With plasma exchange, which removes autoantibodies and provides functional ADAMTS13, TTP (which once was uniformly fatal) can be treated successfully in more than 80% of patients.

MORPHOLOGY

The morphologic findings in the various forms of HUS/TTP are indistinguishable and vary mainly in chronicity rather than cause. In acute, active disease, the kidney may show patchy or diffuse cortical necrosis and subcapsular petechiae. The glomerular capillaries are distended and occluded by thrombi. Disruption of the mesangial matrix and damage to the mesangial cells often results in mesangiolysis. Interlobular arteries and arterioles often show occlusive thrombi. Chronic disease is common in patients with atypical HUS, and has features that stem from continued injury and attempts at healing. The renal cortex reveals various degrees of scarring. By light microscopy, the glomeruli are mildly hypercellular and have marked thickening of the capillary walls associated with splitting or reduplication of the basement membrane (so called double contours or tram tracks). The walls of arteries and arterioles often exhibit increased layers of cells and connective tissue ("onion-skinning") that narrow the vessel lumens. The morphologic lesions are similar to those seen in malignant hypertension (Fig. 20.39). These changes lead to persistent hypoperfusion and potentially diffuse cortical necrosis, which manifests clinically as renal failure and hypertension. The gross alterations of massive ischemic necrosis are sharply limited to the cortex (Fig. 20.40). The histologic appearance is that of acute ischemic infarction. The lesions may be patchy, with areas of coagulative necrosis and apparently better preserved cortex.

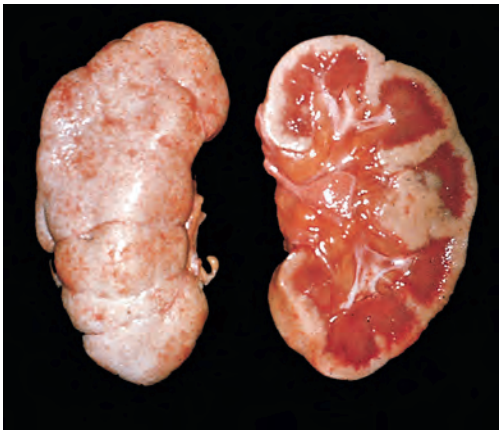


Figure 20.40 Diffuse cortical necrosis. The pale ischemic necrotic areas are confined to the cortex and columns of Bertin.

KEY CONCEPTS

THROMBOTIC MICROANGIOPATHY

- Thrombotic microangiopathy encompasses a diverse set of conditions that all lead to deposition of thrombi in the microvasculature, accompanied by red cell hemolysis, tissue ischemia and organ dysfunction, and a consumptive thrombocytopenia.
- In typical HUS, Shiga-like toxin produced by bacteria, most commonly *E. coli* strain O157:H7, is responsible for producing platelet activation and thrombosis.
- In most cases of atypical HUS, aberration activation of complement due to inherited mutations or acquired autoantibodies is the key pathogenic abnormality.
- In TTP, deficiencies of ADAMTS13, a negative regulator of vWF, permits the formation of abnormally large multimers of vWF that are capable of activating platelets.

Other Vascular Disorders

Atherosclerotic Ischemic Renal Disease

We have seen that atherosclerotic unilateral renal artery stenosis can lead to hypertension. *Bilateral renal artery disease,* usually diagnosed definitively by arteriography, is a fairly common cause of chronic ischemia with renal insufficiency in older individuals, sometimes in the absence of hypertension. The importance of recognizing this condition is that surgical revascularization can prevent further decline in renal function.

Atheroembolic Renal Disease

Embolization of fragments of atheromatous plaques from the aorta or renal artery into intrarenal vessels occurs in older adults with severe atherosclerosis, especially after surgery on the abdominal aorta, aortography, or intra-aortic cannulization. These emboli can be recognized in the lumens of arcuate and interlobular arteries by their content of cholesterol crystals, which appear as rhomboid clefts (Fig. 20.41). The clinical consequences of atheroemboli vary according to the number of emboli and the preexisting state of renal function. Frequently they are of no significance. However, acute renal injury or failure may develop in older adults in whom renal function is already compromised.

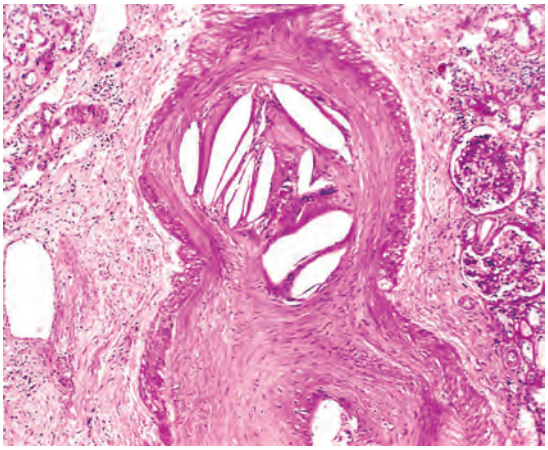


Figure 20.41 Atheroembolus with typical cholesterol clefts in an interlobar artery.

Sickle Cell Nephropathy

Sickle cell disease (homozygous) or trait (heterozygous) may lead to a variety of alterations in renal morphology and function, some of which produce clinically significant abnormalities. The various manifestations are grouped under *sickle cell nephropathy*.

The most common abnormalities are *hematuria* and a *diminished concentrating ability* (hyposthenuria). These are thought to be due to accelerated sickling in the hypertonic hypoxic milieu of the renal medulla; the hyperosmolarity dehydrates red cells and increases intracellular sickle hemoglobin (HbS) concentrations, which likely explains why even those with sickle cell trait are affected. Patchy *papillary necrosis* may occur in both homozygotes and heterozygotes; this is sometimes associated with cortical scarring. *Proteinuria* is also common in sickle cell disease, occurring in about 30% of patients. It is usually mild to moderate, but on occasion the overt nephrotic syndrome arises, associated with sclerosing glomerular lesions.

Renal Infarcts

The kidneys are common sites for the development of infarcts. Contributing to this predisposition is the extensive blood flow to the kidneys (one-fourth of the cardiac output), but probably more important is the limited collateral circulation from extrarenal sites (small blood vessels penetrating from the renal capsule supply only the very outer rim of the cortex). Although thrombosis in advanced atherosclerosis and the acute vasculitis of polyarteritis nodosa may occlude arteries, most infarcts are due to embolism. A major source of such emboli is mural thrombosis in the left atrium and ventricle as a result of myocardial infarction. Vegetative endocarditis, aortic aneurysms, and aortic atherosclerosis are less frequent sources of emboli.

MORPHOLOGY

Because of the lack of a collateral blood supply, most renal infarcts are of the "white" anemic variety. Within 24 hours, infarcts become sharply demarcated, pale, yellow-white areas that may contain small irregular foci of hemorrhagic discoloration. They are usually ringed by a zone of intense hyperemia.

The infarcts are wedge-shaped, with the base against the cortical surface and the apex pointing toward the medulla. In time, these acute areas of ischemic necrosis undergo progressive fibrous scarring, giving rise to depressed, pale, gray-white scars that assume a *V*-shape on section. The histologic changes in renal infarcts are those of ischemic coagulative necrosis, described in Chapter 2.

Many renal infarcts are clinically silent. Sometimes, pain with tenderness localized to the costovertebral angle occurs, associated with showers of red cells in the urine. Large infarcts of one kidney are probably associated with narrowing of the renal artery or one of its major branches, which, in turn, may cause hypertension.

CONGENITAL AND DEVELOPMENTAL ANOMALIES

About 10% of people are born with significant malformations of the urinary system. Renal dysplasias and hypoplasias account for 20% of chronic kidney disease in children.

Congenital renal disease can be hereditary but most often results from an acquired developmental defect during gestation. As discussed in Chapter 10, defects in genes involved in normal renal development, including the Wilms tumor–associated genes, understandably cause urogenital anomalies. As a rule, the resulting developmental abnormalities involve structural components of both the kidney and urinary tract. Other genetic defects primarily produce functional abnormalities in tubular transport, such as cystinuria and renal tubular acidosis. Here, we restrict the discussion to structural anomalies involving primarily the kidney. All except horseshoe kidney are uncommon. Anomalies of the lower urinary tract are discussed in Chapter 21.

Agenesis of the Kidney. Bilateral agenesis is incompatible with life and usually encountered in stillborn infants. It is often associated with other congenital disorders (e.g., limb defects, hypoplastic lungs). Unilateral agenesis is uncommon and compatible with normal life if no other abnormalities exist. The solitary kidney enlarges as a result of compensatory hypertrophy. Some patients eventually develop progressive glomerular sclerosis in the remaining kidney as a result of the adaptive changes in hypertrophied nephrons, discussed earlier in the chapter, and, in time, chronic kidney disease ensues.

Hypoplasia. *Hypoplasia* refers to failure of the kidneys to develop to a normal size. This anomaly may occur bilaterally, resulting in renal failure in early childhood, but it is more commonly encountered as a unilateral defect. True renal hypoplasia is observed in low-birth-weight infants and may contribute to their increased lifetime risk for chronic kidney disease. Differentiation between congenital and acquired atrophic kidneys may be impossible, but a truly hypoplastic kidney shows no scars and has a reduced number of renal lobes and pyramids, usually six or fewer.

Ectopic Kidneys. The development of the metanephros into the kidneys may occur in ectopic foci. These kidneys lie either just above the pelvic brim or sometimes within the pelvis.

They are usually normal or slightly small in size but otherwise are not remarkable. Because of their abnormal position, kinking or tortuosity of the ureters may cause obstruction to urinary flow, which predisposes to bacterial infections.

Horseshoe Kidneys. Fusion of the upper (10%) or lower poles (90%) of the kidneys produces a horseshoe-shaped structure that is continuous across the midline anterior to the great vessels. This anomaly is found in 1 in 500 to 1000 autopsies.

CYSTIC DISEASES OF THE KIDNEY

Cystic diseases of the kidney are heterogeneous, comprising hereditary, developmental, and acquired disorders. They are important for several reasons: (1) They are reasonably common and often represent diagnostic problems for clinicians, radiologists, and pathologists; (2) some forms, such as adult polycystic kidney disease, are major causes of chronic kidney disease; and (3) they can occasionally be confused with malignant tumors. A useful classification of renal cysts is summarized in Table 20.12.

Autosomal Dominant (Adult) Polycystic Kidney Disease

Autosomal dominant (adult) polycystic kidney disease is a hereditary disorder characterized by multiple expanding cysts of both kidneys that ultimately destroy the renal parenchyma and cause renal failure. It is a common condition affecting roughly 1 in 400 to 1000 live births and accounting for about 5% to 10% of cases of ESRD requiring transplantation or dialysis. The inheritance pattern is autosomal dominant with high penetrance. Although the susceptibility to develop this disease is inherited as an autosomal dominant trait, as with tumor suppressor genes, both alleles of the involved genes have to be nonfunctional for development of the disease. Thus, individuals prone to autosomal dominant polycystic kidney disease inherit one copy of a mutated *APKD* gene, and mutation of the other allele is acquired in the somatic cells of the kidney. The disease is bilateral; reported unilateral cases probably represent multicystic dysplasia. The cysts initially involve a minority of the nephrons, so renal function is retained until about the fourth or fifth decade of life.

Genetics and Pathogenesis

A wide range of different mutations in *PKD1* and *PKD2* has been described, and this allelic heterogeneity has complicated genetic diagnosis of this disorder.

- The *PKD1* gene is located on chromosome 16p13.3. It encodes a large (460-kD) integral membrane protein named *polycystin-1,* which has a large extracellular region, multiple transmembrane domains, and a short cytoplasmic tail. Polycystin-1 is expressed in tubular epithelial cells, particularly those of the distal nephron. At present, its precise function is not known, but it contains domains that are usually involved in cell-cell and cell-matrix interactions. Mutations in *PKD1* account for about 85% of cases. In individuals with these mutations, the likelihood of developing renal failure is less than 5% by 40 years of age, rising to more than 35% by 50 years of age, more than 70% at 60 years of age, and more than 95% by 70 years of age.

Table 20.12 Summary of Renal Cystic Diseases

Disease	Inheritance	Pathologic Features	Clinical Features or Complications	Typical Outcome	Diagrammatic Representation
Adult polycystic kidney disease	Autosomal dominant	Large multicystic kidneys, liver cysts, berry aneurysms	Hematuria, flank pain, urinary tract infection, renal stones, hypertension	Chronic renal failure beginning at 40–60 years of age	
Childhood polycystic kidney disease	Autosomal recessive	Enlarged, cystic kidneys at birth	Hepatic fibrosis	Variable, death in infancy or childhood	
Medullary sponge kidney	None	Medullary cysts on excretory urography	Hematuria, urinary tract infection, recurrent renal stones	Benign	
Familial juvenile nephronophthisis	Autosomal recessive	Corticomedullary cysts, shrunken kidneys	Salt wasting, polyuria, growth retardation, anemia	Progressive renal failure beginning in childhood	
Multicystic renal dysplasia	None	Irregular kidneys with cysts of variable size	Association with other renal anomalies	Renal failure if bilateral, surgically curable if unilateral	
Acquired renal cystic disease	None	Cystic degeneration in end-stage kidney disease	Hemorrhage, erythrocytosis, neoplasia	Dependence on dialysis	
Simple cysts	None	Single or multiple cysts in normal-sized kidneys	Microscopic hematuria	Benign	

- The *PKD2* gene, located on chromosome 4q21, accounts for most of the remaining cases of polycystic disease. Its product, *polycystin-2,* is an integral membrane protein that is expressed in all segments of the renal tubules and in many extrarenal tissues. Polycystin-2 functions as a Ca^{2+}-permeable cation channel. Overall, the disease is less severe than that associated with *PKD1* mutations. Renal failure occurs in less than 5% of patients with *PKD2* mutations at 50 years of age, but this rises to 15% at 60 years of age, and 45% at 70 years of age.

The pathogenesis of polycystic disease is not established, but the currently favored hypothesis places the cilia-centrosome complex of tubular epithelial cells at the center of the disorder (Fig. 20.42). The tubular epithelial cells of the kidney contain a single nonmotile primary cilium, a 2- to 3-μm-long hairlike organelle that projects into the tubular lumen from the apical surface of the cells. The cilium is made up of microtubules, and it arises from and is attached to a basal body derived from the centriole. The cilia are part of a system of organelles and cellular structures that sense mechanical signals. The apical cilia function in the kidney tubule as a mechanosensor to monitor changes in fluid flow and shear stress, while intercellular junctional complexes monitor forces between cells, and focal adhesions sense attachment to extracellular matrices. In response to external signals, these sensors regulate ion flux (cilia can induce Ca^{2+} flux in cultured kidney epithelial cells) and cellular behavior, including cell polarity and proliferation. The idea that defects in mechanosensing, Ca^{2+} flux, and signal transduction underlie cyst formation is supported by several observations.

- Both polycystin-1 and polycystin-2 are localized to the primary cilium.
- Other genes that are mutated in cystic diseases (e.g., the nephrocystin genes, described later) encode proteins that are also localized to cilia and/or basal bodies.

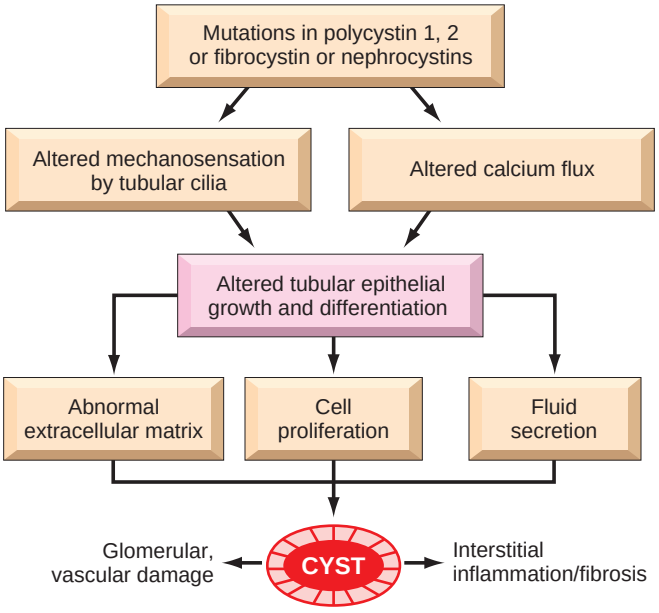


Figure 20.42 Possible mechanisms of cyst formation in cystic kidney diseases *(see text).*

- Knockout of the *PKD1* gene in one model organism (the worm *Caenorhabditis elegans*) results in ciliary abnormalities and cyst formation.
- Tubular cells obtained from mice with a deletion of the *PKD1* gene (which causes embryonic lethality in this species) retain normal architecture of cilia but lack the flow-induced Ca^{2+} flux that occurs in normal tubular cells.

Polycystin-1 and polycystin-2 appear to form a protein complex that regulates intracellular Ca^{2+} in response to fluid flow, perhaps because fluid moving through the kidney tubules causes ciliary bending that opens Ca^{2+} channels. Mutation of either of the *PKD* genes leads to loss of the polycystin complex or formation of an aberrant complex. The consequent disruption of normal polycystin activity results in alterations of intracellular Ca^{2+}, which (you will recall) regulates many downstream signaling events, including pathways that directly or indirectly impact cellular proliferation, apoptosis, and secretory functions. The increase in calcium is thought to stimulate proliferation and secretion from epithelial cells lining the cysts, which together result in progressive cyst formation and enlargement. In addition, cyst fluids have been shown to harbor mediators derived from epithelial cells that enhance fluid secretion and induce inflammation. Finally, the calcium-induced signals also alter the interaction of epithelial cells with ECM, and this too is thought to contribute to the cyst formation and interstitial fibrosis that are characteristic of progressive polycystic kidney disease.

MORPHOLOGY

In gross appearance, the kidneys are bilaterally enlarged and may achieve enormous sizes; weights as much as 4 kg for each kidney have been reported. The external surface appears to be composed solely of a mass of cysts, up to 3 to 4 cm in diameter, with no intervening parenchyma (Fig. 20.43A and B). However, microscopic examination reveals functioning nephrons dispersed between the cysts. The cysts may be filled with a clear, serous fluid or with turbid, red to brown, sometimes hemorrhagic fluid. As these cysts enlarge, they may encroach on the calyces and pelvis to produce pressure defects. The cysts arise from the tubules throughout the nephron and therefore have variable lining epithelia. On occasion, papillary epithelial formations and polyps project into the lumen.

Clinical Features

Many patients remain asymptomatic until renal insufficiency announces the presence of the disease. In others, hemorrhage or progressive dilation of cysts may produce pain. Excretion of blood clots causes renal colic. The enlarged kidneys, usually apparent on abdominal palpation, may induce a dragging sensation. The disease occasionally begins with the insidious onset of hematuria, followed by other features of progressive chronic kidney disease, such as proteinuria (rarely more than 2 g/day), polyuria, and hypertension. Patients with *PKD2* mutations tend to have an older age at onset and later development of renal failure. Both genetic and environmental factors influence disease severity. Progression is accelerated in blacks (particularly in those with sickle cell trait), in males, and in the presence of hypertension.

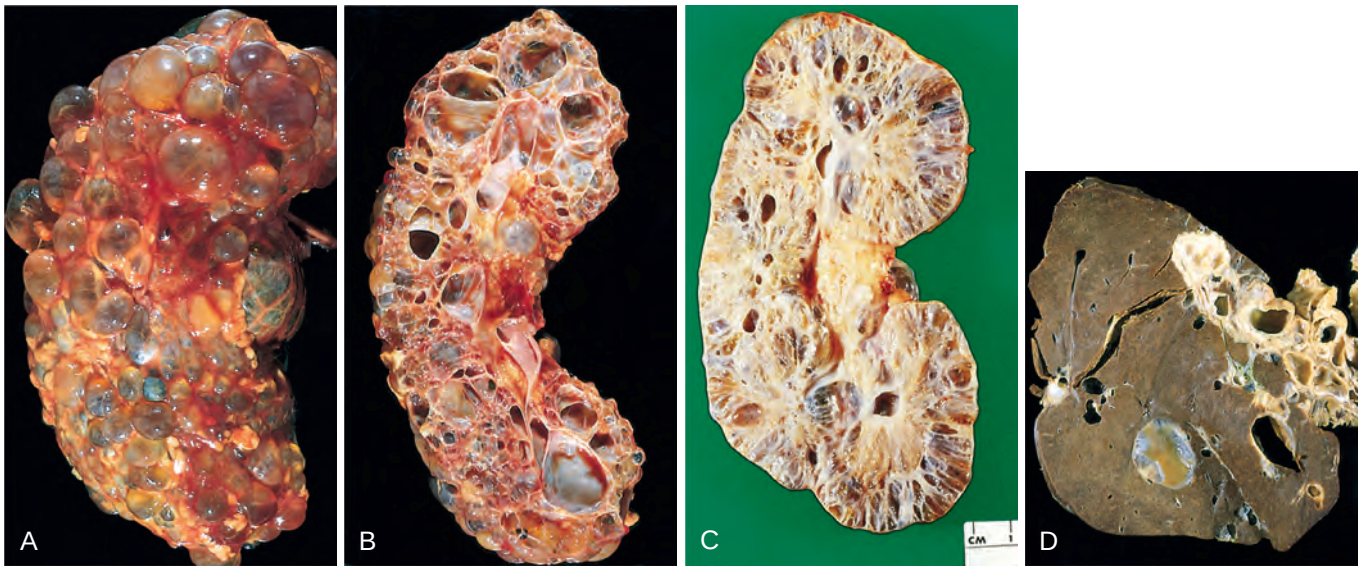


Figure 20.43 (A and B) Autosomal dominant adult polycystic kidney disease (ADPKD) viewed from the external surface and bisected. The kidney is markedly enlarged and contains numerous dilated cysts. (C) Autosomal recessive childhood PKD, showing smaller cysts and dilated channels at right angles to the cortical surface. (D) Liver cysts in adult PKD.

Individuals with polycystic kidney disease also tend to have extrarenal congenital anomalies. About 40% have one to several cysts in the liver (polycystic liver disease) that are usually asymptomatic. The cysts are derived from biliary epithelium. Cysts occur much less frequently in the spleen, pancreas, and lungs. Intracranial berry aneurysms, presumably from altered expression of polycystin in vascular smooth muscle, arise in the circle of Willis, and subarachnoid hemorrhages from these account for death in 4% to 10% of individuals. *Mitral valve prolapse* and other cardiac valvular anomalies occur in 20% to 25% of patients, but most are asymptomatic. The diagnosis is made by radiologic imaging techniques.

This form of chronic kidney disease is remarkable in that patients may survive for many years with azotemia slowly progressing to uremia. Ultimately, about 40% of adult patients die of coronary or hypertensive heart disease, 25% of infection, 15% of a ruptured berry aneurysm or hypertensive intracerebral hemorrhage, and the rest of other causes.

Autosomal Recessive (Childhood) Polycystic Kidney Disease

Autosomal recessive (childhood) polycystic kidney disease is genetically distinct from adult polycystic kidney disease. *Perinatal, neonatal, infantile,* and *juvenile* subcategories have been defined, depending on the time of presentation and presence of associated hepatic lesions. The first two are the most common; serious manifestations are usually present at birth, and the young infant might succumb rapidly to renal failure.

Genetics and Pathogenesis

In most cases, the disease is caused by mutations in the *PKHD1* gene, which maps to chromosome region 6p21–p23. The gene is highly expressed in adult and fetal kidney and also in liver and pancreas. The *PKHD1* gene encodes *fibrocystin*, a 447-kD integral membrane protein with a large extracellular region, a single transmembrane component, and a short cytoplasmic tail. The extracellular region contains multiple copies of a domain forming an Ig-like fold. Like polycystins 1 and 2, fibrocystin also has been localized to the primary cilium of tubular cells. The function of fibrocystin is unknown, but its putative conformational structure indicates it may be a cell surface receptor with a role in collecting duct and biliary differentiation.

Analysis of autosomal recessive polycystic disease patients has revealed a wide range of different mutations. The vast majority of cases are compound heterozygotes (i.e., inherit a different mutant allele from each of the two parents). This complicates molecular diagnosis of the disorder.

MORPHOLOGY

The kidneys are enlarged and have a smooth external appearance. On cut section, numerous small cysts in the cortex and medulla give the kidney a spongelike appearance. Dilated elongated channels are present at right angles to the cortical surface, completely replacing the medulla and cortex (Fig. 20.43C). On microscopic examination, there is cylindrical or, less commonly, saccular dilation of all collecting tubules. The cysts have a uniform lining of cuboidal cells, reflecting their origin from the collecting ducts. In almost all cases, the liver has cysts (Fig. 20.43D) associated with portal fibrosis and proliferation of portal bile ducts.

Patients who survive infancy (infantile and juvenile forms) may develop a peculiar hepatic injury characterized by bland periportal fibrosis and proliferation of well-differentiated biliary ductules, now termed *congenital hepatic fibrosis*. In older children, hepatic disease is the predominant clinical concern. Such patients may develop portal hypertension with splenomegaly. Curiously, congenital hepatic fibrosis

sometimes occurs in the absence of polycystic kidneys or has been reported in the presence of adult polycystic kidney disease.

Cystic Diseases of the Renal Medulla

The two major types of medullary cystic disease are *medullary sponge kidney,* a relatively common and usually innocuous structural change, and *nephronophthisis,* which is almost always associated with renal dysfunction.

Medullary Sponge Kidney

The term *medullary sponge kidney* is restricted to multiple cystic dilations of the collecting ducts in the medulla. The condition occurs in adults and is usually discovered radiographically. Renal function is usually normal. On gross inspection, the papillary ducts in the medulla are dilated, and small cysts may be present. The cysts are lined by cuboidal epithelium or occasionally by transitional epithelium. Unless there is superimposed pyelonephritis, cortical scarring is absent. The pathogenesis is unknown.

Nephronophthisis

This group of progressive renal disorders is characterized by variable number of cysts in the medulla, usually concentrated at the corticomedullary junction. Initial injury probably involves the distal tubules with tubular basement membrane disruption, followed by chronic and progressive tubular atrophy involving both medulla and cortex and interstitial fibrosis. Although the medullary cysts are important, the **cortical tubulointerstitial damage is the cause of the eventual renal insufficiency.**

Three variants of the nephronophthisis disease complex are recognized: (1) sporadic, nonfamilial; (2) familial juvenile nephronophthisis (most common); and (3) renal-retinal dysplasia (15%) in which the kidney disease is accompanied by ocular lesions. The familial forms are inherited as autosomal recessive traits and usually become manifest in childhood or adolescence. As a group, the nephronophthisis complex is now the most common genetic cause of ESRD in children and young adults.

Children affected with nephronophthisis present first with polyuria and polydipsia, which reflect a marked defect in the concentrating ability of renal tubules. Sodium wasting and tubular acidosis are also prominent. Some syndromic variants of nephronophthisis (e.g., Senior-Loken syndrome, Joubert syndrome, Bardet Biedl syndrome, Jeune syndrome, Meckel Gruber syndrome, Mainzer-Saldino syndrome, Sensenbrenner syndrome) can have extrarenal associations, including ocular motor abnormalities, retinal dystrophy, liver fibrosis, and cerebellar abnormalities. The expected course is progression to ESRD in 5 to 10 years.

Genetics and Pathogenesis

Sixteen responsible gene loci, *NPHP1* to *NPHP11* (that encode proteins called nephrocystins, *JBTS2, JBTS3, JBTS9,* and *JBTS11,* are mutated in the juvenile forms of nephronophthisis, and the list continually expands as additional loci that contribute to this ciliopathy are identified. These proteins are present in the primary cilia, basal bodies attached to these cilia, or the centrosome organelle from which the basal bodies originate. The *NPHP2* gene product has been identified as *inversin,* which mediates left-right patterning during embryogenesis.

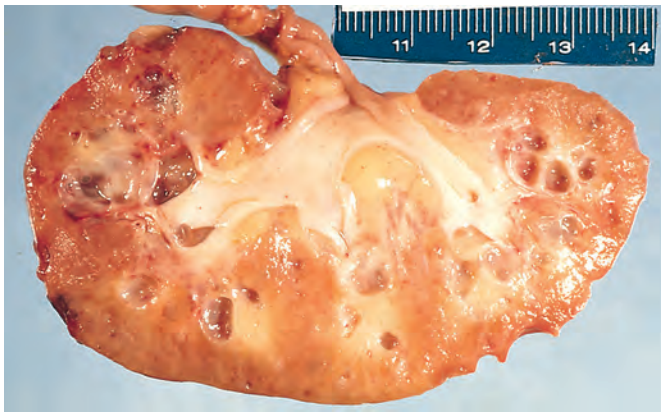


Figure 20.44 Medullary cystic disease. Cut section of kidney showing cysts at the corticomedullary junction and in the medulla.

MORPHOLOGY

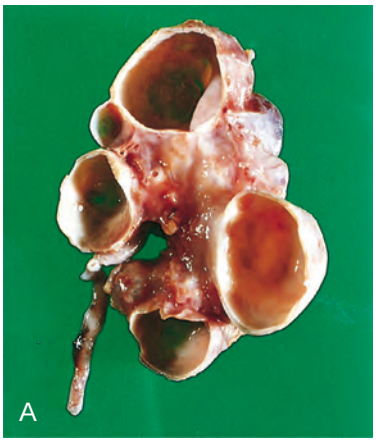
In nephronophthisis, the kidneys are small, have contracted granular surfaces, and show cysts in the medulla, most prominently at the corticomedullary junction. The cysts are similar to those in medullary cystic disease, a rare, usually autosomal dominant, disease (Fig. 20.44). Small cysts are also seen in the cortex. The cysts are lined by flattened or cuboidal epithelium and are usually surrounded by either inflammatory cells or fibrous tissue. In the cortex, there is widespread atrophy and thickening of the tubular basement membranes, together with interstitial fibrosis. In general, glomerular structure is preserved.

There are few specific clues to diagnosis, because the medullary cysts might be too small to be visualized radiographically. The disease should be strongly considered in children or adolescents with otherwise unexplained chronic renal failure, a positive family history, and chronic tubulointerstitial nephritis on biopsy.

Multicystic Renal Dysplasia

Dysplasia is a sporadic disorder that can be unilateral or bilateral and is often cystic. The kidney is usually enlarged, extremely irregular, and multicystic (Fig. 20.45A). The cysts vary in size from several millimeters to centimeters in diameter. On histologic examination, they are lined by flattened epithelium. Although normal nephrons are present, the characteristic histologic feature is the presence of islands of undifferentiated mesenchyme, often with cartilage, and immature collecting ducts (Fig. 20.45B). Most cases are associated with ureteropelvic obstruction, ureteral agenesis or atresia, and other anomalies of the lower urinary tract.

When unilateral, the dysplasia may mimic a neoplasm and lead to surgical exploration and nephrectomy. The opposite kidney functions normally, and such patients have an excellent prognosis after surgical removal of the affected kidney. In bilateral multicystic renal dysplasia, renal failure may ultimately result.



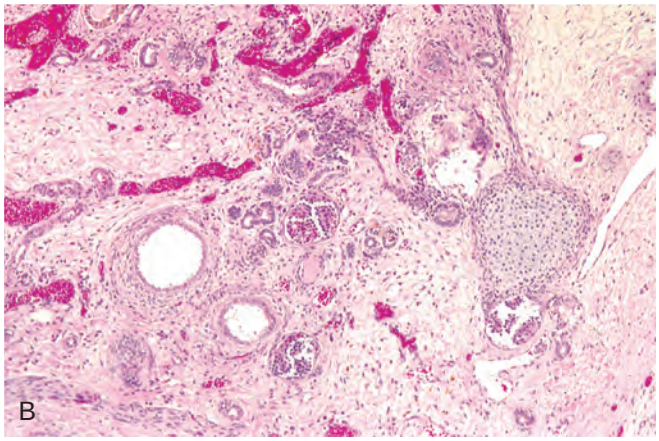


Figure 20.45 Multicystic renal dysplasia. (A) Gross appearance. (B) Histologic section showing disorganized architecture, dilated tubules with cuffs of primitive stroma, and an island of cartilage (hematoxylin and eosin stain). (A, Courtesy Dr. D. Schofield, Children's Hospital, Los Angeles, Calif.; B, courtesy Dr. Laura Finn, Children's Hospital, Seattle, Wash.)

Acquired Cystic Disease

Patients with ESRD who have undergone prolonged dialysis sometimes show numerous cortical and medullary renal cysts. The cysts measure 0.1 to 4 cm in diameter, contain clear fluid, are lined by either hyperplastic or flattened tubular epithelium, and often contain calcium oxalate crystals. They probably form as a result of obstruction of tubules by interstitial fibrosis or by oxalate crystals. Most are asymptomatic, but sometimes the cysts bleed, causing hematuria. There is up to 100-fold increased risk of renal cell carcinoma, which develops in 7% of patients observed for 10 years.

Simple Cysts

Simple cysts may be single or multiple and usually involve the cortex. They are commonly 1 to 5 cm but may reach 10 cm or more in size. They are translucent, lined by a gray, glistening, smooth membrane, and filled with clear fluid. On microscopic examination, these membranes are composed of a single layer of cuboidal or flattened cuboidal epithelium, which in many instances may be completely atrophic.

Simple cysts are common postmortem findings without clinical significance. On occasion, hemorrhage into them may cause sudden distention and pain, and calcification of the hemorrhage may give rise to bizarre radiographic shadows. The main importance of cysts lies in their differentiation from kidney tumors. Radiologic studies show that in contrast to renal tumors, renal cysts have smooth contours, are almost always avascular, and give fluid rather than solid signals on ultrasonography.

KEY CONCEPTS

CYSTIC DISEASES

- Autosomal dominant polycystic kidney disease accounts for a small yet significant subset of ESRD.
- Ciliopathies or abnormalities of the cilium-centrosome complex underlie the major cystic kidney diseases, including polycystic kidney disease (both autosomal dominant and autosomal recessive forms) and nephronophthisis.
- Dysfunction of the primary cilium of tubular epithelial cells results in alterations in ion flux and changes in cell proliferation and function, culminating in renal cyst formation.

URINARY TRACT OBSTRUCTION (OBSTRUCTIVE UROPATHY)

Obstructive lesions of the urinary tract increase susceptibility to infection and to stone formation, and unrelieved obstruction almost always leads to permanent renal atrophy, termed *hydronephrosis* or *obstructive uropathy*. Fortunately, many causes of obstruction are surgically correctable or medically treatable.

Obstruction may be sudden or insidious, partial or complete, unilateral or bilateral; it may occur at any level of the urinary tract from the urethra to the renal pelvis. It can be caused by *intrinsic* lesions of the urinary tract or *extrinsic* lesions that compress the ureter. The common causes are as follows (Fig. 20.46):

- *Congenital anomalies:* posterior urethral valves and urethral strictures, meatal stenosis, bladder neck obstruction; ureteropelvic junction narrowing or obstruction; severe vesicoureteral reflux
- *Urinary calculi*
- *Benign prostatic hypertrophy*
- *Tumors:* carcinoma of the prostate, bladder tumors, contiguous malignant disease (retroperitoneal lymphoma), carcinoma of the cervix or uterus
- *Inflammation:* prostatitis, ureteritis, urethritis, retroperitoneal fibrosis
- *Sloughed papillae or blood clots*
- *Pregnancy*
- *Uterine prolapse and cystocele*
- *Functional disorders:* neurogenic (spinal cord damage or diabetic nephropathy) and other functional abnormalities of the ureter or bladder (often termed *dysfunctional obstruction*)

***Hydronephrosis* is the term used to describe dilation of the renal pelvis and calyces associated with progressive atrophy of the kidney due to obstruction to the outflow of urine.** Even with complete obstruction, glomerular filtration

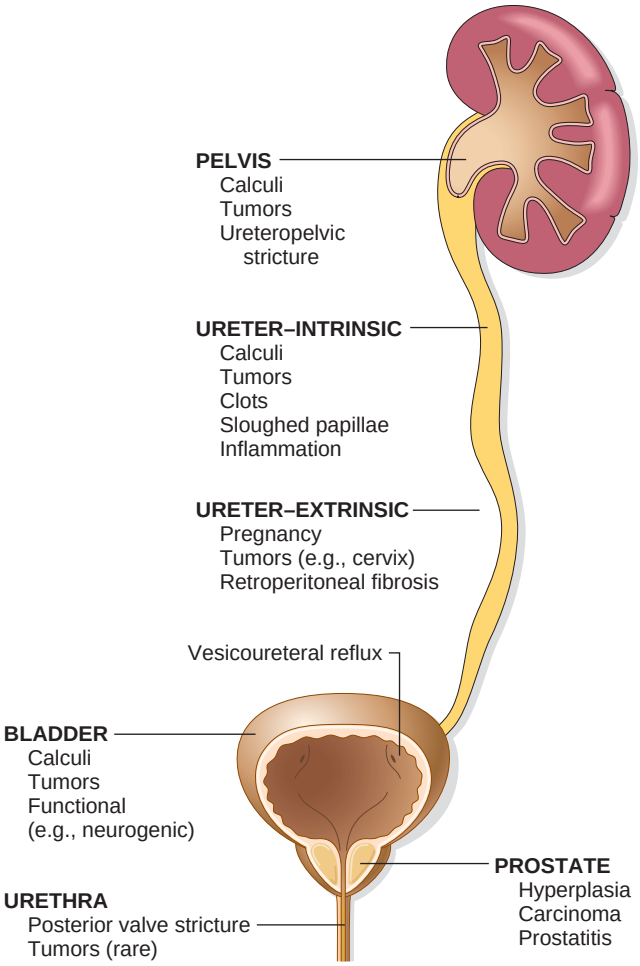


Figure 20.46 Obstructive lesions of the urinary tract.

persists for some time because the filtrate subsequently diffuses back into the renal interstitium and perirenal spaces, from where it ultimately returns to the lymphatic and venous systems. Because of this continued filtration, the affected calyces and pelvis become dilated, often markedly so. The high pressure in the pelvis is transmitted back through the collecting ducts into the cortex, causing renal atrophy, but it also compresses the renal vasculature of the medulla, causing a diminution in inner medullary blood flow. The medullary vascular defects are initially reversible, but lead to medullary functional disturbances. Accordingly, the initial functional alterations caused by obstruction are largely tubular, manifested primarily by impaired concentrating ability. Only later does the GFR begin to fall. Obstruction also triggers an interstitial inflammatory reaction, leading eventually to interstitial fibrosis, by mechanisms similar to those discussed earlier (see Fig. 20.8).

MORPHOLOGY

When the obstruction is sudden and complete, it leads to mild dilation of the pelvis and calyces and sometimes to atrophy of the renal parenchyma. When the obstruction is subtotal or intermittent, progressive dilation ensues, giving rise to hydronephrosis (Fig. 20.47). Depending on the level of urinary block, the dilation may affect the bladder first, or the ureter and then the kidney.

The kidney may be slightly to massively enlarged, depending on the degree and the duration of the obstruction. The earlier features are those of simple dilation of the pelvis and calyces, but in addition there is often significant interstitial inflammation, even in the absence of infection. In chronic cases, the picture is one of cortical tubular atrophy with marked diffuse interstitial fibrosis. Progressive blunting of the apices of the pyramids occurs, and these eventually become cupped. In far-advanced cases, the kidney may become transformed into a thin-walled cystic structure having a diameter of up to 15 to 20 cm (see Fig. 20.47) with striking parenchymal atrophy, total obliteration of the pyramids, and thinning of the cortex.

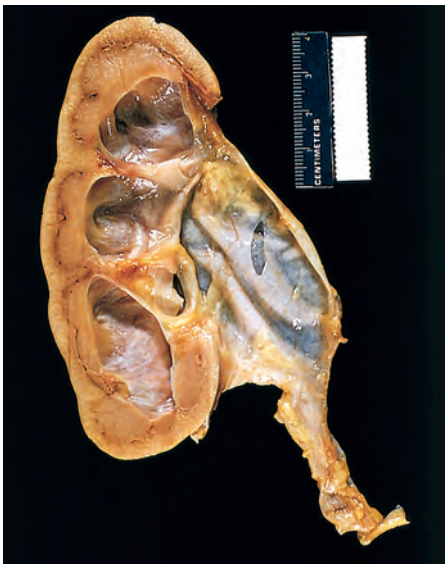


Figure 20.47 Hydronephrosis of the kidney, with marked dilation of the pelvis and calyces and thinning of the renal parenchyma.

Clinical Features

Acute obstruction may provoke pain attributed to distention of the collecting system or renal capsule. Most of the early symptoms are produced by the underlying cause of the hydronephrosis. Thus, calculi lodged in the ureters may give rise to renal colic, and prostatic enlargements may give rise to bladder symptoms.

Unilateral complete or partial hydronephrosis may remain silent for long periods because the unaffected kidney can maintain adequate renal function. Sometimes its existence first becomes apparent in the course of imaging studies. In its early stages, perhaps the first few weeks, relief of obstruction leads to reversion to normal function. *Ultrasonography* is a useful noninvasive technique in the diagnosis of obstructive uropathy.

In *bilateral partial obstruction,* the earliest manifestation is inability to concentrate urine, reflected by polyuria and nocturia. Some patients develop distal tubular acidosis, renal salt wasting, secondary renal calculi, and chronic tubulointerstitial nephritis with scarring and atrophy of the papilla and medulla. Hypertension is common.

Complete bilateral obstruction of rapid onset results in oliguria or anuria and is incompatible with survival unless the obstruction is relieved. Curiously, after relief of complete urinary tract obstruction, postobstructive *diuresis* occurs. This can often be massive, with the kidney excreting large amounts of urine that is rich in sodium chloride.

UROLITHIASIS (RENAL CALCULI, STONES)

Urolithiasis affects 5% to 10% of people in the United States in their lifetime, and the stones may form anywhere in the urinary tract, but most arise in the kidney. Men are affected more often than women, and the peak age at onset is between 20 and 30 years. Familial and hereditary predisposition to stone formation has long been known. Many inborn errors of metabolism, such as cystinuria and primary hyperoxaluria, provide examples of hereditary disease characterized by excessive production and excretion of stone-forming substances.

Etiology and Pathogenesis

There are four main types of calculi (Table 20.13): (1) *calcium stones* (about 70%), composed largely of calcium oxalate or calcium oxalate mixed with calcium phosphate; (2) another 15% are so-called *triple stones* or *struvite stones,* composed of magnesium ammonium phosphate; (3) 5% to 10% are *uric acid stones;* and (4) 1% to 2% are made up of *cystine*. An organic mucoprotein matrix, making up 1% to 5% of the stone by weight, is present in all calculi. **Although there are many causes for the initiation and propagation of stones, the most important determinant is an increased urinary concentration of the stones' constituents, such that it exceeds their solubility (supersaturation).** A low urine volume in some metabolically normal patients may also favor supersaturation.

Table 20.13 Prevalence of Various Types of Renal Stones

Stone Type	Percentage of All Stones
Calcium Oxalate and Phosphate	**70**
Idiopathic hypercalciuria (50%)	
Hypercalciuria and hypercalcemia (10%)	
Hyperoxaluria (5%)	
Enteric (4.5%)	
Primary (0.5%)	
Hyperuricosuria (20%)	
Hypocitraturia	
No known metabolic abnormality (15% to 20%)	
Magnesium Ammonium Phosphate (Struvite)	**5–10**
Uric Acid	**5–10**
Associated with hyperuricemia	
Associated with hyperuricosuria	
Idiopathic (50% of uric stones)	
Cystine	**1–2**
Others or Unknown	**±5**

- *Calcium oxalate stones* are associated in about 5% of patients with hypercalcemia and hypercalciuria, such as occurs with hyperparathyroidism, diffuse bone disease, sarcoidosis, and other hypercalcemic states. About 55% have hypercalciuria without hypercalcemia, which is caused by several factors, including hyperabsorption of calcium from the intestine (absorptive hypercalciuria), an intrinsic impairment in renal tubular reabsorption of calcium (renal hypercalciuria), or idiopathic fasting hypercalciuria with normal parathyroid function. As many as 20% of calcium oxalate stones are associated with increased uric acid secretion *(hyperuricosuric calcium nephrolithiasis),* with or without hypercalciuria. The mechanism of stone formation in this setting involves "nucleation" of calcium oxalate by uric acid crystals in the collecting ducts. Five percent are associated with *hyperoxaluria,* either hereditary (primary oxaluria) or, more commonly, acquired by intestinal overabsorption in patients with enteric diseases. *Hypocitraturia,* which can be idiopathic or associated with acidosis and chronic diarrhea of unknown cause, may produce calcium stones. In a variable proportion of individuals with calcium stones, no cause can be found (idiopathic calcium stone disease).
- *Magnesium ammonium phosphate stones* are formed largely after infections by urea-splitting bacteria (e.g., *Proteus* and some staphylococci) that convert urea to ammonia. The resultant alkaline urine causes the precipitation of magnesium ammonium phosphate salts. These form some of the largest stones, as the amount of urea excreted normally is very large. Indeed, so-called *staghorn calculi* occupying large portions of the renal pelvis are frequently a consequence of infection.
- *Uric acid stones* are common in individuals with hyperuricemia, such as patients with gout, and diseases involving rapid cell turnover, such as the leukemias. However, **more than half of all patients with uric acid calculi have neither hyperuricemia nor increased urinary excretion of uric acid.** In this group, it is thought that a tendency to excrete urine of pH below 5.5 may predispose to uric acid stones, because uric acid is insoluble in acidic urine. In contrast to the radiopaque calcium stones, uric acid stones are radiolucent.
- *Cystine stones* are caused by genetic defects in the renal reabsorption of amino acids, including cystine, leading to cystinuria. These stones also form at low urinary pH.

It can therefore be appreciated that **increased concentration of stone constituents, changes in urinary pH, decreased urine volume, and the presence of bacteria influence the formation of calculi.** However, many calculi occur in the absence of these factors; conversely, individuals with hypercalciuria, hyperoxaluria, and hyperuricosuria often do not form stones. It has been postulated that stone formation is enhanced by a deficiency in inhibitors of crystal formation in urine. The list of such inhibitors is long, including pyrophosphate, diphosphonate, citrate, glycosaminoglycans, osteopontin, and a glycoprotein called nephrocalcin.

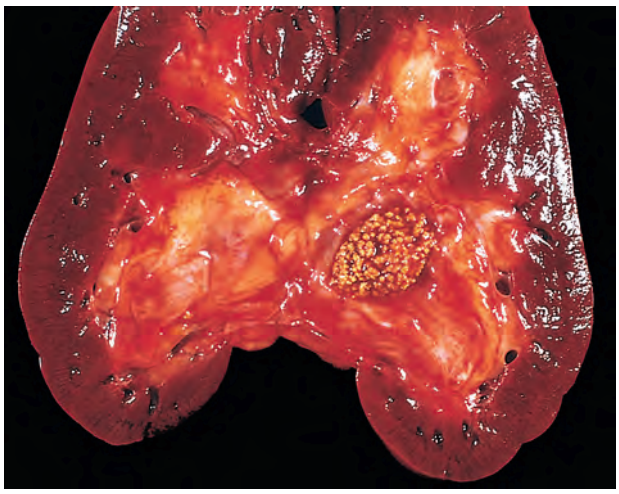


Figure 20.48 Nephrolithiasis. A large stone impacted in the renal pelvis. (Courtesy Dr. E. Mosher, Brigham and Women's Hospital, Boston, Mass.)

MORPHOLOGY

Stones are unilateral in about 80% of patients. The favored sites for their formation are within the renal calyces and pelves (Fig. 20.48) and in the bladder. If formed in the renal pelvis, they tend to remain small, having an average diameter of 2 to 3 mm. These may have smooth contours or may take the form of an irregular, jagged mass of spicules. Often many stones are found within one kidney. On occasion, progressive accretion of salts leads to the development of branching structures known as staghorn calculi, which create a cast of the pelvic and calyceal system.

Clinical Features

Urolithiasis may be asymptomatic, produce severe renal colic and abdominal pain, or may cause significant renal damage. Larger stones often manifest themselves by hematuria. Stones also predispose to superimposed infection, both by their obstructive nature and by the trauma they produce.

NEOPLASMS OF THE KIDNEY

Both benign and malignant neoplasms occur in the kidney. Malignant neoplasms are of great importance clinically. By far the most common malignant tumor is renal cell carcinoma, followed by Wilms tumor, which is found in children and described in Chapter 10, and finally urothelial carcinoma of the calyces and pelves.

Benign Neoplasms

Renal Papillary Adenoma

Small, discrete adenomas arising from the renal tubular epithelium are found commonly (7% to 22%) at autopsy. They are most frequently papillary and therefore called *papillary adenomas*.

MORPHOLOGY

These small tumors are less than 1.5 cm in diameter. They are present invariably within the cortex and appear grossly as pale yellow-gray, discrete, well-circumscribed nodules. On microscopic examination, they are composed of complex, branching, papillomatous structures with numerous complex fronds. Cells may also grow as tubules, glands, cords, and sheets of cells. The cells are cuboidal to polygonal in shape and have regular, small central nuclei, scanty cytoplasm, and no atypia.

By histologic criteria, these tumors do not differ from low-grade papillary renal cell carcinoma and share some immunohistochemical and cytogenetic features (trisomies 7 and 17) with papillary cancers, to be discussed later. The tumor size is used as a prognostic feature, with a cutoff of 3 cm separating those that metastasize from those that rarely do. However, because of occasional reports of small tumors that have metastasized, some regard all adenomas, regardless of size, as potentially malignant.

Angiomyolipoma

This benign neoplasm consists of vessels, smooth muscle, and fat originating from perivascular epithelioid cells. **Angiomyolipomas are present in 25% to 50% of patients with tuberous sclerosis,** a disease caused by loss-of-function mutations in the *TSC1* or *TSC2* tumor suppressor genes. Tuberous sclerosis is characterized by lesions of the cerebral cortex that produce epilepsy and intellectual disability, a variety of skin abnormalities, and unusual benign tumors at other sites, such as the heart (Chapters 12 and 28). The clinical importance of angiomyolipoma is due largely to their susceptibility to spontaneous hemorrhage.

Oncocytoma

This is an epithelial neoplasm composed of large eosinophilic cells having small, round, benign-appearing nuclei that have large nucleoli. It is thought to arise from the intercalated cells of collecting ducts, and accounts for approximately 5% to 15% of renal neoplasms. Ultrastructurally the eosinophilic cells have numerous mitochondria. In gross appearance, the tumors are tan or mahogany brown, relatively homogeneous, and usually well encapsulated with a central scar in one-third of cases. However, they may achieve a large size (up to 12 cm in diameter). There are some familial cases in which these tumors are multicentric rather than solitary.

Malignant Neoplasms

Renal Cell Carcinoma

Renal cell carcinomas represent about 3% of all newly diagnosed cancers in the United States and account for 85% of renal cancers in adults. There are approximately 65,000 new cases per year and 13,000 deaths from the disease. The tumors occur most often in older individuals, usually in the sixth and seventh decades of life, and show a 2:1 male preponderance.

Epidemiology

Tobacco is the most significant risk factor. Cigarette smokers have double the incidence of renal cell carcinoma, and pipe and cigar smokers are also more susceptible. An international

study has identified additional risk factors, including obesity (particularly in women); hypertension; unopposed estrogen therapy; and exposure to asbestos, petroleum products, and heavy metals. There is also an increased risk in patients with ESRD, chronic kidney disease, acquired cystic disease (see earlier), and tuberous sclerosis.

Most renal cancer is sporadic, but unusual forms of autosomal dominant familial cancers occur, usually in younger individuals. Although they account for only 4% of renal cancers, familial variants have been instructive in understanding renal carcinogenesis.

- *Von Hippel-Lindau (VHL) syndrome.* One-half to two-thirds of individuals with VHL (nearly all, if they live long enough) (Chapter 28) develop renal cysts and bilateral, often multiple, renal cell carcinomas. *Current studies implicate the* VHL *gene in the development of both familial and sporadic clear cell carcinomas.*
- *Hereditary leiomyomatosis and renal cell cancer syndrome.* This autosomal dominant disease is caused by mutations of the *FH* gene, which expresses fumarate hydratase, and is characterized by cutaneous and uterine leiomyomata and an aggressive type of papillary carcinoma with increased propensity for metastatic spread.
- *Hereditary papillary carcinoma.* This autosomal dominant form is manifested by multiple bilateral tumors with papillary histology. These tumors show a series of cytogenetic abnormalities and, as we will describe, mutations in the *MET* proto-oncogene.
- *Birt-Hogg-Dubé syndrome.* The autosomal dominant inheritance pattern of this disease is due to mutations involving the *BHD* gene, which expresses folliculin. The syndrome features a constellation of skin (fibrofolliculomas, trichodiscomas, and acrochordons), pulmonary (cysts or blebs), and renal tumors with a wide range of histologic subtypes.

Classification of Renal Cell Carcinoma: Histology, Cytogenetics, and Genetics

The classification of renal cell carcinoma is based on correlative cytogenetic, genetic, and histologic studies of both familial and sporadic tumors. The major types of tumor are as follows (Fig. 20.49):

- *Clear cell carcinoma.* This is the most common type, accounting for 70% to 80% of renal cell cancers. The tumors are made up of cells with clear or granular cytoplasm and are *nonpapillary*. They can be familial, but in most cases (95%) are sporadic. In 98% of these tumors, whether familial, sporadic, or associated with VHL syndrome, there is loss of sequences on the short arm of chromosome 3. The deleted region harbors the *VHL* gene (3p25.3). A second nondeleted allele of the *VHL* gene shows somatic mutations or hypermethylation-induced inactivation in up to 80% of clear cell cancers, indicating that the *VHL* gene acts as a tumor suppressor gene in both sporadic and familial cancers (Chapter 7). The *VHL* gene encodes a protein that is part of a ubiquitin ligase complex involved in targeting other proteins for degradation. Important among the targets of the VHL protein is the transcription factor hypoxia-inducible factor-1 (HIF-1). When *VHL* is inactive, HIF-1 levels remain high, even under normoxic conditions, causing inappropriate expression of a number of genes that are turned on by HIF. These include genes that promote angiogenesis, such as VEGF, and genes that stimulate cell growth, such as insulin-like growth factor-1 (IGF-1). In addition, HIF collaborates in complex ways with the oncogenic factor MYC to "reprogram" cellular metabolism in a way that favors growth. Deep sequencing of renal carcinoma genomes has revealed frequent mutations in a number of genes that regulate histone modifications, indicating that dysregulation of

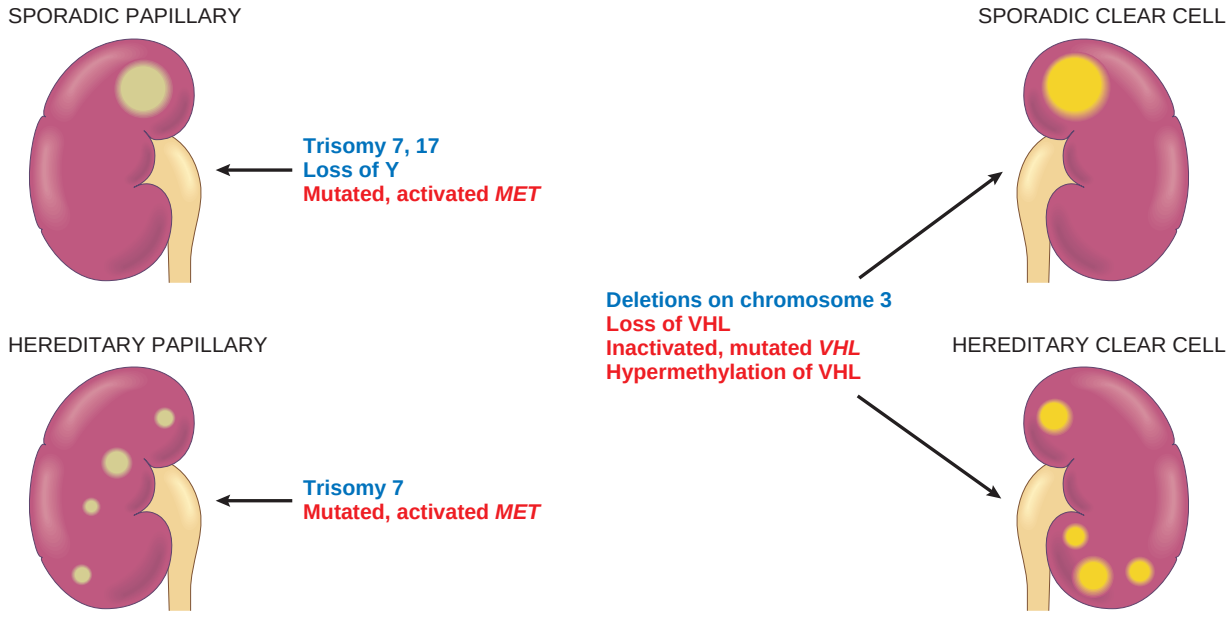


Figure 20.49 Cytogenetics *(blue)* and genetics *(red)* of clear cell versus papillary renal cell carcinoma. (Courtesy Dr. Keith Ligon, Brigham and Women's Hospital, Boston, Mass.)

the epigenome also has an important role in clear cell carcinoma.

- *Papillary carcinoma* accounts for 10% to 15% of renal cancers. It is characterized by a papillary growth pattern and also occurs in both familial and sporadic forms. These tumors are not associated with 3p deletions. The most common cytogenetic abnormalities are trisomies 7 and 17 and loss of Y in male patients in the sporadic form, and trisomy 7 in the familial form. The gene on chromosome 7 for the familial form has been mapped to *MET,* a proto-oncogene that encodes the tyrosine kinase receptor for *hepatocyte growth factor*. *MET* is also mutated in a small proportion of sporadic papillary carcinomas. Hepatocyte growth factor (also called *scatter factor*) mediates growth, cell mobility, invasion, and morphogenetic differentiation. Unlike clear cell carcinomas, papillary carcinomas are frequently multifocal in origin.
- *Chromophobe carcinoma* represents 5% of renal cell cancers and is composed of cells with prominent cell membranes and pale eosinophilic cytoplasm, usually with a halo around the nucleus. On cytogenetic examination, these tumors show multiple chromosome losses and extreme hypodiploidy. Like the benign oncocytoma, they are thought to grow from intercalated cells of collecting ducts and have an excellent prognosis compared with that of the clear cell and papillary cancers. Histologic distinction from oncocytoma can be difficult.
- *Xp11 translocation carcinoma* is a genetically distinct subtype of renal cell carcinoma. It often occurs in young patients and is defined by translocations of the *TFE3* gene located at Xp11.2 with a number of partner genes, all of which result in overexpression of the TFE3 transcription factor. The neoplastic cells consist of clear cytoplasm with a papillary architecture.
- *Collecting duct (Bellini duct) carcinoma* represents approximately 1% or less of renal epithelial neoplasms. They arise from collecting duct cells in the medulla. Several chromosomal losses and deletions have been described, but a distinct pattern has not been identified. Histologically these tumors are characterized by malignant cells forming glands enmeshed within a prominent fibrotic stroma, typically in a medullary location. *Medullary carcinoma* is a morphologically similar neoplasm that is seen in patients with sickle cell trait.

MORPHOLOGY

Renal cell carcinomas may arise in any portion of the kidney, but more commonly affects the poles. **Clear cell carcinomas** most likely arise from proximal tubular epithelium, and usually occur as solitary unilateral lesions. They are bright yellow-gray-white spherical masses of variable size that distort the renal outline. The yellow color is a consequence of the prominent lipid accumulations in tumor cells. There are commonly large areas of gray-white necrosis and foci of hemorrhagic discoloration. The margins are usually sharply defined and confined within the renal capsule (Fig. 20.50). In clear cell carcinoma, the growth pattern varies from solid to trabecular (cordlike) or tubular (resembling tubules). The tumor cells have a rounded or polygonal shape and abundant clear or granular cytoplasm, which contains glycogen and lipids (Fig. 20.51A). The tumors have delicate branching vasculature and may show cystic as well as solid areas. Most tumors are well differentiated, but some show nuclear atypia with bizarre nuclei and giant cells.

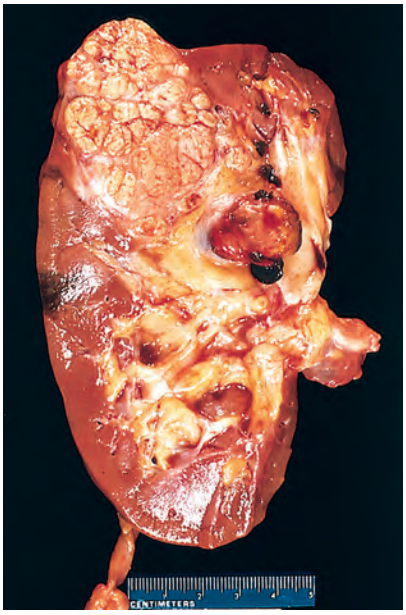


Figure 20.50 Renal cell carcinoma. Typical cross-section of yellowish, spherical neoplasm in one pole of the kidney. Note the tumor in the dilated thrombosed renal vein.

As tumors enlarge, they may bulge into the calyces and pelvis and eventually fungate through the walls of the collecting system to extend into the ureter. **One of the striking characteristics of renal cell carcinoma is its tendency to invade the renal vein** (see Fig. 20.50), in which it may grow as a solid column of cells that extends up the inferior vena cava, sometimes as far as the right side of the heart.

Papillary carcinomas, thought to arise from distal convoluted tubules, can be multifocal and bilateral. They are typically hemorrhagic and cystic, especially when large. The tumor is composed of cuboidal or low columnar cells arranged in papillary formations. Interstitial foam cells are common in the papillary cores (see Fig. 20.51B). Psammoma bodies may be present. The stroma is usually scanty but highly vascularized. **Chromophobe renal carcinoma** is made up of pale eosinophilic cells, often with a perinuclear halo, arranged in solid sheets with a concentration of the largest cells around blood vessels (see Fig. 20.51C). **Collecting duct carcinoma** is a rare variant showing irregular channels lined by highly atypical epithelium with a hobnail pattern. Sarcomatoid changes arise infrequently in all types of renal cell carcinoma and are a decidedly ominous feature.

Clinical Features

The classic clinical features of renal cell carcinoma are *costovertebral pain, palpable mass,* and *hematuria,* but all three are seen in only 10% of cases. The most reliable clue is hematuria, but it is usually intermittent and may be microscopic; thus, the tumor may remain silent until it attains a large size, often greater than 10 cm. At this time, it is often associated with generalized constitutional symptoms, such as fever, malaise, weakness, and weight loss. This pattern of asymptomatic growth occurs in many patients, so the

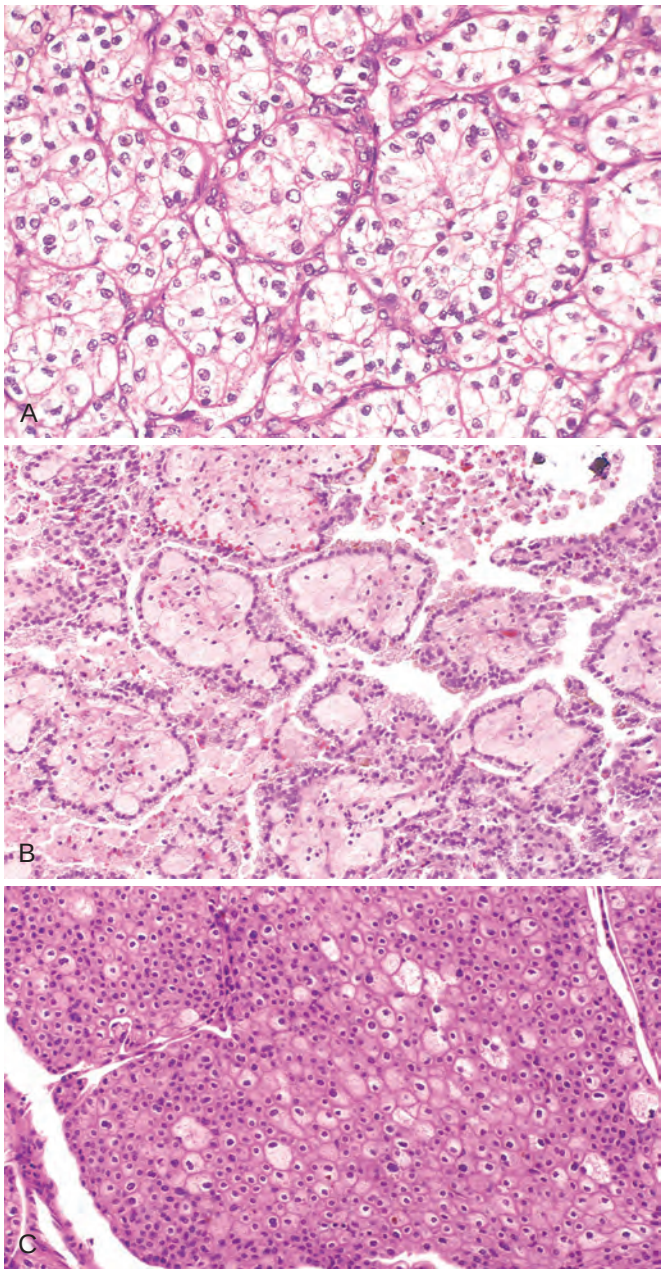


Figure 20.51 Renal cell carcinoma. (A) Clear cell type. (B) Papillary type. Note the papillae and foamy macrophages in the stalk. (C) Chromophobe type. (Courtesy Dr. A. Renshaw, Baptist Hospital, Miami, Fla.)

tumor may have reached a diameter of more than 10 cm when it is discovered. Currently, an increasing number of tumors are being discovered in the asymptomatic state by incidental radiologic studies (e.g., computed tomography or magnetic resonance imaging) performed for other indications.

Renal cell carcinoma is considered one of the great mimics in medicine, because it tends to produce a diversity of systemic symptoms not related to the kidney. In addition to fever and constitutional symptoms mentioned earlier, renal cell carcinomas produce a number of syndromes ascribed to abnormal hormone production, including polycythemia, hypercalcemia, hypertension, hepatic dysfunction, feminization or masculinization, Cushing syndrome, eosinophilia, leukemoid reactions, and amyloidosis.

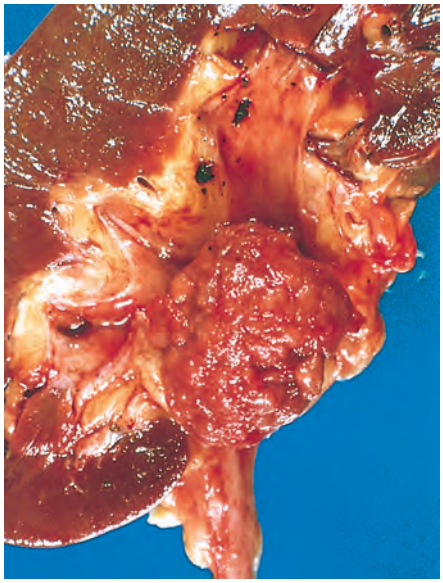


Figure 20.52 Urothelial carcinoma of the renal pelvis. The pelvis has been opened to expose the nodular irregular neoplasm, just proximal to the ureter.

A particularly troublesome feature of renal cell carcinoma is its tendency to metastasize widely before giving rise to any local symptoms or signs. In 15% of new patients with renal cell carcinoma, there is radiologic evidence of metastases at the time of presentation. The most common locations of metastasis are the lungs (more than 50%) and bones (33%), followed in frequency by the regional lymph nodes, liver, adrenal glands, and brain.

The average 5-year survival rate of persons with renal cell carcinoma is about 70% and as high as 100% in the absence of distant metastases. With renal vein invasion or extension into the perinephric fat, the figure is reduced to approximately 60%. Radical nephrectomy has been the treatment of choice, but nephron-sparing surgery to preserve renal function is recommended for T1a tumors (<4 cm) as well as larger tumors when technically feasible. Drugs that inhibit VEGF and various tyrosine kinases are used as an adjunct to therapy for metastatic disease.

Urothelial Carcinoma of the Renal Pelvis

Approximately 5% to 10% of primary renal tumors originate from the urothelium of the renal pelvis (Fig. 20.52). These tumors range from apparently benign papillomas to invasive urothelial (transitional cell) carcinomas.

Renal pelvic tumors usually become clinically apparent within a relatively short time, because they lie within the pelvis and, by fragmentation, produce noticeable hematuria. They are almost invariably small when discovered. These tumors may block urinary outflow and lead to palpable hydronephrosis and flank pain. On histologic examination, pelvic tumors are the exact counterpart of those found in the urinary bladder; further details are in Chapter 21.

Urothelial tumors may occasionally be multiple, involving the pelvis, ureters, and bladder. In 50% of renal pelvic tumors, there is a preexisting or concomitant bladder urothelial

tumor. On histologic examination, there are also foci of atypia or carcinoma in situ in grossly normal urothelium remote from the pelvic tumor. There is an increased incidence of urothelial carcinomas of the renal pelvis in individuals with Lynch syndrome.

Infiltration of the wall of the pelvis and calyces is common. For this reason, despite their apparently small, deceptively benign appearance, the prognosis for these tumors is not good. Reported 5-year survival rates vary from 50% to 100% for low-grade noninvasive lesions to 10% with high-grade infiltrating tumors.

KEY CONCEPTS

KIDNEY NEOPLASMS

- Clear cell renal cell carcinoma is the most common subtype of malignant renal neoplasms, which often involves *VHL*, a tumor suppressor gene.
- Papillary renal cell carcinoma is the second most common subtype of malignant renal neoplasms, which may involve the *MET* proto-oncogene.
- Hereditary forms of renal cell carcinoma have led to the discovery of important genes (e.g., *VHL*, *BHD*) in renal carcinogenesis.
- Urothelial tumors resembling similar tumors in the urinary bladder can also originate in the renal pelvis. These tumors have a poor prognosis.

SUGGESTED READINGS

Pathogenesis of Immune-Mediated Glomerular Injury

Couser WG: Basic and translational concepts of immune-mediated glomerular diseases, *J Am Soc Nephrol* 23:381–399, 2012. *[An outstanding and comprehensive review of the immunopathogenesis of glomerular diseases].*

Pickering M, Cook HT: Complement and glomerular disease: new insights, *Curr Opin Nephrol Hypertens* 20:271–277, 2011. *[An insightful review into newly emerging concepts of the contribution of complement to renal disease, with an emphasis on dysregulation of the alternate pathway of complement activation].*

Mechanisms of Progression in Glomerular Diseases

Schlondorff DO: Overview of factors contributing to the pathophysiology of progressive renal disease, *Kidney Int* 74:860–866, 2008. *[An excellent summary of complex and interacting pathways that lead to progressive chronic kidney injury].*

Nephrotic Syndrome

Rosenberg AZ, Kopp JB: Focal segmental glomerulosclerosis, *Clin J Am Soc Nephrol* 12:502–517, 2017. *[An excellent and well-illustrated review of pathologic and clinical characteristics of this group of diseases].*

Assady S, Wanner N, Skorecki KL et al: New insights into podocyte biology in glomerular health and disease, *J Am Soc Nephrol* 28:1707–1715, 2017. *[An excellent synopsis of alternations in podocytes that underlie glomerular diseases].*

Ronco P, Debiec H: Pathophysiological advances in membranous nephropathy, *Lancet* 385:1983–1992, 2015. *[A comprehensive review of the immunopathology of membranous nephropathy by two of the foremost investigators in the identification of the target antigens in this disease].*

Sethi S, Fervenza FC: Membranoproliferative glomerulonephritis–a new look at an old entity, *N Engl J Med* 366:1119–1131, 2012. *[An excellent reassessment of this entity with a review of new classification schema that are based on underlying pathogenesis].*

Nephritic Syndrome

Nast CC: Infection-related glomerulonephritis: changing demographics and outcomes, *Adv Chronic Kidney Dis* 19:68–75, 2012. *[A thoughtful review that integrates long-established concepts in infection-related glomerulonephritis with changing patterns of disease expression in both First and Third World societies].*

Rapidly Progressive (Crescentic) Glomerulonephritis

Jennette JC, Nachman PH: ANCA glomerulonephritis and vasculitis, *Clin J Am Soc Nephrol* 12:1680–1691, 2017. *[A thorough review of this major form of vasculitis with an emphasis on underlying pathophysiology].*

Kambham N: Crescentic Glomerulonephritis: an update on Pauci-immune and Anti-GBM diseases, *Adv Anat Pathol* 19(2):111–124, 2012. *[A current review of the clinicopathologic manifestations of these diseases].*

Isolated Urinary Abnormalities

Gubler MC: Inherited diseases of the glomerular basement membrane, *Nat Clin Pract Nephrol* 4:24–37, 2008. *[A detailed review of the pathophysiology of hereditary nephritides and a clear guide to their identification and differential diagnosis].*

Roberts IS: Pathology of IgA nephropathy, *Nat Rev Nephrol* 10:445–454, 2014. *[An excellent review that details emerging concepts of the pathophysiology of this disorder].*

Acute Tubular Injury/Necrosis

Bellomo R, Kellum JA, Ronco C: Acute kidney injury, *Lancet* 380:756–766, 2012. *[An excellent review of acute kidney injury with an emphasis on clinical aspects and management issues].*

Zuk A, Bonventre JV: Acute kidney injury, *Annu Rev Med* 67:293–307, 2016. *[A comprehensive review of acute kidney injury].*

Tubulointerstitial Nephritis

Johnson JR, Russo TA: Acute pyelonephritis in adults, *N Engl J Med* 378:48–59, 2018. *[A review of the clinical and pathologic aspects of pyelonephritis].*

Perazella MA, Markowitz GS: Drug-induced acute interstitial nephritis, *Nat Rev Nephrol* 6:461–470, 2010. *[A review of the various drugs that can induce acute interstitial nephritis along with relevant clinical and pathologic issues].*

Vascular Diseases

Brocklebank V, Wood KA, Kavanagh D: Thrombotic microangiopathy and the kidney, *Clin J Am Soc Nephrol* 13:300–317, 2018. *[An excellent review of the pathogenesis and clinical features of various forms of TMA].*

Chang A: Thrombotic microangiopathy and the kidney, *Diagn Histopathol* 23:101–108, 2017. *[An overview of the dysregulation of the alternative pathway of complement cascade and other mechanisms that contribute to thrombotic microangiopathy].*

Hill GS: Hypertensive nephrosclerosis, *Curr Opin Nephrol Hypertens* 17:266–270, 2008. *[A review of the pathophysiologic underpinning of hypertensive nephrosclerosis].*

Cystic Diseases of the Kidney

Benoit G, Machuca E, Heidet L et al: Hereditary kidney diseases: highlighting the importance of classical Mendelian phenotypes, *Ann N Y Acad Sci* 1214:83–98, 2010. *[A review of the genetic mutations that affect a variety of kidney structures ranging from the podocyte to glomerular basement membranes to mitochondria].*

Hildebrandt F, Benzing T, Katsanis N: Ciliopathies, *N Engl J Med* 364:1533–1543, 2011. *[An excellent review of the spectrum of single-gene mutations involving cilia and their clinical manifestations and underlying disease mechanisms].*

Neoplasms of the Kidney

Jonasch E, Futreal A, Davis I et al: State of the science: an update on renal cell carcinoma, *Mol Cancer Res* 10:859–880, 2012. *[An excellent review of the molecular basis of renal cell carcinoma and their subtypes with an emphasis on potential therapeutic implications].*

Srigley JR, Delahunt B, Eble JN et al: The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia, *Am J Surg Pathol* 37(10):1469–1489, 2013. *[An updated pathologic classification of renal cell carcinoma].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 21

The Lower Urinary Tract and Male Genital System

George Jabboure Netto • Mahul B. Amin

CHAPTER CONTENTS

The Lower Urinary Tract

The renal pelves, ureters, urinary bladder, and urethra (except the terminal portion) are lined by a special form of transitional epithelium called *urothelium*. Urothelium is composed of five to six layers of cells with oval nuclei, often with linear nuclear grooves, and a surface layer consisting of large, flattened "umbrella cells" with abundant cytoplasm. This epithelium rests on a well-developed basement membrane, beneath which is a lamina propria. The lamina propria in the bladder contains wisps of smooth muscle that form a discontinuous muscularis mucosa. It is important to differentiate the muscularis mucosa from the deeper well-defined larger muscle bundles of the detrusor muscle (muscularis propria), since bladder cancers are staged on the basis of invasion of the latter. If urine flow is obstructed and intravesical pressures rise, the bladder musculature undergoes hypertrophy.

The ureters lie throughout their course in a retroperitoneal position. Retroperitoneal tumors or fibrosis may entrap and obstruct the ureters. In middle-aged and older women, relaxation of pelvic support leads to prolapse (descent) of the uterus, pulling with it the floor of the bladder. In this fashion the bladder is protruded into the vagina, creating a pouch *(cystocele)* that fails to empty readily with micturition. In males, the seminal vesicles and prostate have similar close relationships, being situated just posterior and inferior to the neck of the bladder. Thus, enlargement of the prostate, so common in middle to later life, constitutes an important cause of urinary tract obstruction. Subsequent sections discuss the major pathologic lesions of the ureters, urinary bladder, and urethra.

URETERS

Congenital Anomalies

Congenital anomalies of the ureters are found in about 2% to 3% of all autopsies (Table 21.1). Although most have little clinical significance, certain anomalies may lead to obstruction of the flow of urine and thus cause clinical disease.

Ureteropelvic junction (UPJ) obstruction is the most common cause of hydronephrosis in infants and children. Cases that present early in life preferentially affect males, are bilateral in 20% of cases, and are often associated with other anomalies. There is agenesis of the contralateral kidney in a minority of cases. In adults, UPJ obstruction is more common in women and is most often unilateral. The condition is most commonly ascribed to abnormal organization of smooth muscle bundles or excess stromal deposition of collagen between smooth muscle bundles at the UPJ, or in rare cases to extrinsic compression of the UPJ by abnormal renal vessels.

Tumors and Tumor-Like Lesions

Primary tumors of the ureter are rare. Benign tumors are generally of mesenchymal origin. *Fibroepithelial polyp,* a tumor-like lesion often occurring in children, is composed of loose, vascularized connective tissue overlaid by urothelium. Primary malignant tumors of the ureter resemble those arising in the renal pelvis, calyces, and bladder. The majority are *urothelial carcinomas* (Fig. 21.1). They occur most frequently during the sixth and seventh decades of life and cause obstruction of the ureteral lumen. They commonly occur concurrently with urothelial carcinomas of bladder or renal pelvis.

Obstructive Lesions

Intrinsic and extrinsic lesions may obstruct the ureters and may give rise to *hydroureter, hydronephrosis,* and *pyelonephritis* (Chapter 20). Unilateral obstruction typically results from proximal intrinsic or extrinsic causes (e.g., stones, neoplasms etc.), whereas bilateral obstruction arises from distal causes, such as nodular hyperplasia of the prostate (Table 21.2).

Sclerosing Retroperitoneal Fibrosis

This rare lesion is characterized by a fibrotic proliferative inflammatory process that encases retroperitoneal structures and causes hydronephrosis. The disorder occurs in middle to late age and is more common in males than females. At least a subset of these cases is related to *IgG4-related disease,* a recently described entity associated with elevated levels of serum immunoglobulin G4 (IgG4) and fibroinflammatory lesions rich in IgG4-secreting plasma cells, that also affects other organs (Chapter 6). Other etiologies for retroperitoneal fibrosis include drug exposures (ergot derivatives, β-adrenergic blockers), inflammatory conditions (vasculitis, diverticulitis, Crohn disease), or malignancies (lymphomas, urinary tract carcinomas). Most, however, have no obvious cause and are considered primary or idiopathic *(Ormond disease).*

Microscopic examination typically reveals fibrous tissue containing a prominent infiltrate of lymphocytes, often with germinal centers, plasma cells (frequently IgG4-positive), and eosinophils. Treatment initially involves corticosteroids, although many patients eventually become resistant and require ureteral stents or surgical extrication of the ureters from the surrounding fibrous tissue (ureterolysis).

Figure 21.1 Papillary transitional cell carcinoma extensively involving the ureter. (Courtesy Dr. Cristina Magi-Galluzzi, The Johns Hopkins Hospital, Baltimore, Md.)

Table 21.1 Congenital Anomalies of Ureter

- Vesicoureteral reflux
- Double and bifid ureters
- Ureteropelvic junction obstruction
- Diverticula

Table 21.2 Major Causes of Ureteral Obstruction

Cause of Obstruction	Characteristics/Mechanisms
Intrinsic	
Calculi	Typically of renal origin Usually small (5 mm in diameter or less) Impact at loci of ureteral narrowing—ureteropelvic junction, where ureters cross iliac vessels and where they enter bladder—and cause excruciating "renal colic"
Strictures	May be congenital or acquired
Tumors	Urothelial carcinomas arising in ureters Rarely, benign tumors or fibroepithelial polyps
Blood clots	Massive hematuria from renal calculi, tumors, or papillary necrosis
Neurogenic	Interruption of the neural pathways to the bladder
Extrinsic	
Pregnancy	Physiologic relaxation of smooth muscle or pressure on ureters at pelvic brim from enlarging fundus
Periureteral inflammation	Salpingitis, diverticulitis, peritonitis, sclerosing retroperitoneal fibrosis
Endometriosis	With pelvic lesions associated with scarring
Tumors	Cancers of the rectum, bladder, prostate, ovaries, uterus, cervix; lymphomas, sarcomas

KEY CONCEPTS

DISORDERS OF THE URETERS

- Ureteral obstruction is clinically significant because it can lead to hydroureter, hydronephrosis, or pyelonephritis, thus compromising renal function.
- In children, congenital UPJ obstruction is the most common obstructive lesion.
- In adults, ureteral obstruction may be acute (e.g., due to obstructing calculi) or chronic (e.g., due to intrinsic or extrinsic tumors or rarely idiopathic conditions such as sclerosing retroperitoneal fibrosis).

URINARY BLADDER

Congenital Anomalies

Several different types of congenital abnormalities of the bladder are recognized; these are variously associated with increased risk of infection or neoplasia.

- *Vesicoureteral reflux* is the most common and serious congenital anomaly. As discussed in Chapter 20, vesicoureteral reflux predisposes to ascending pyelonephritis and loss of renal function. Abnormal connections between the bladder and the vagina, rectum, or uterus may create *congenital vesicouterine fistulae.*
- *Diverticula* are pouch-like invaginations of the bladder wall that vary from less than 1 cm to 10 cm in diameter and may be congenital or acquired. *Congenital diverticula* may be due to a focal failure of development of the normal musculature or to a urinary tract obstruction during fetal development. *Acquired diverticula* are most often associated with prostatic hyperplasia, producing urinary outflow obstruction. The resulting increase in intravesical pressure causes outpouching of the bladder wall and the formation of diverticula. Although most diverticula are small and asymptomatic, they may come to clinical attention since they constitute sites of urinary stasis and predispose to infection and the formation of bladder calculi. Rarely, carcinoma may arise in bladder diverticula; such tumors are on average more advanced in stage due to the thin or absent muscularis propria layer of diverticula.
- *Exstrophy of the bladder* is a developmental failure in the anterior wall of the abdomen and the bladder. As a result, the bladder communicates directly with the abdominal surface (Fig. 21.2). The exposed bladder mucosa may undergo colonic glandular metaplasia and is subject to chronic infection that often spreads to the upper urinary tract. Exstrophy is associated with an increased risk of adenocarcinoma in the bladder remnant.
- *Urachal anomalies.* The urachal canal connects the fetal bladder with the allantois and normally is obliterated at birth. Rarely, it remains fully or partially patent. The former creates a fistulous urinary tract connection between the bladder and umbilicus. When only a central region of patent urachus persists, a *urachal cyst* lined by urothelial or metaplastic glandular epithelium is formed. Urachal cysts are at increased risk for neoplastic transformation, mostly presenting as adenocarcinomas (see Neoplasms). These account for only a small minority of bladder cancers overall (0.1% to 0.3%), but constitute 20% to 40% of bladder adenocarcinomas.

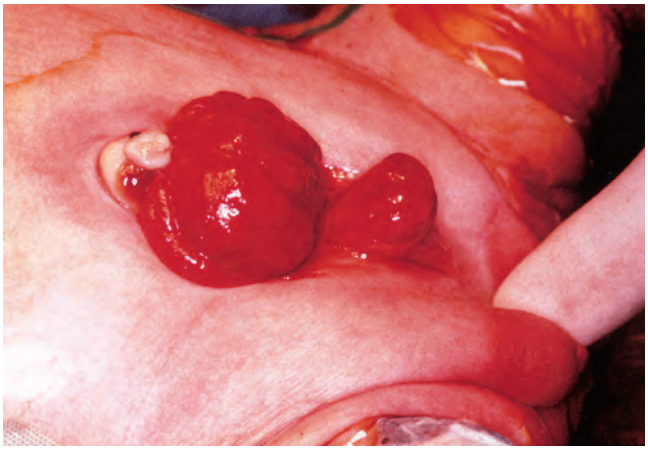


Figure 21.2 Exstrophy of the bladder in a newborn boy. The tied umbilical cord is seen above the hyperemic mucosa of the everted bladder. Below is an incompletely formed penis with marked epispadias. (Courtesy Dr. John Gearhart, The Johns Hopkins Hospital, Baltimore, Md.)

Inflammation

Acute and Chronic Cystitis

Bacterial pyelonephritis is frequently preceded by infection of the urinary bladder, with retrograde spread of microorganisms into the kidneys and their collecting systems (discussed in Chapter 20). Predisposing factors for cystitis include bladder calculi, urinary obstruction, diabetes mellitus, instrumentation, and immune deficiency. The most common etiologic agents of cystitis are the coliforms: *Escherichia coli,* followed by *Proteus, Klebsiella,* and *Enterobacter.* Women are more likely to develop cystitis as a result of their shorter urethras. *Tuberculous cystitis* is almost always a sequel to renal tuberculosis. *Candida albicans* and, much less often, cryptococcal agents may cause cystitis, particularly in immunosuppressed patients or those receiving long-term antibiotics. Schistosomiasis *(Schistosoma haematobium)* remains an important cause of cystitis in certain African and Middle Eastern countries. Viruses (e.g., adenovirus), *Chlamydia,* and *Mycoplasma* may also cause cystitis. Adenovirus and BK virus infections may result in hemorrhagic cystitis. Gas-forming bacteria (such as *Clostridium perfringens*) result in *emphysematous cystitis* (gas-filled vesicles in the bladder wall).

MORPHOLOGY

Most cases of cystitis produce nonspecific acute or chronic inflammation of the bladder. In **acute cystitis,** there is hyperemia of the mucosa and neutrophilic infiltrate, sometimes associated with exudate. Persistence of the bacterial infection leads to **chronic cystitis** associated with mononuclear inflammatory infiltrates.

Other types of non-infectious cystitis include iatrogenic, follicular, and eosinophilic cystitis. Patients receiving systemic chemotherapy or pelvic irradiation may develop **iatrogenic** cystitis.

Cytotoxic agents, such as cyclophosphamide, may cause **hemorrhagic cystitis.** Acute and chronic **radiation cystitis** may occur following the irradiation of the bladder region. **Follicular cystitis** is characterized by the presence of lymphoid follicles within the bladder mucosa and underlying wall. **Eosinophilic cystitis,** manifested by infiltration of the submucosa by eosinophils, typically is a nonspecific subacute inflammation but may also be a manifestation of a systemic allergic disorder.

Clinical Features

All forms of cystitis are characterized by a triad of symptoms: (1) frequency, which in acute cases may necessitate urination every 15 to 20 minutes; (2) lower abdominal pain localized over the bladder region or in the suprapubic region; and (3) dysuria (pain or burning on urination). The local symptoms of cystitis may be merely disturbing, but these infections may be antecedents to pyelonephritis, a more serious disorder (Chapter 20). Cystitis is sometimes a secondary complication of an underlying disorder associated with urinary stasis, such as prostatic hypertrophy, cystocele, calculi, or bladder neoplasms. These primary lesions must be corrected before the cystitis can be relieved.

Special Forms of Cystitis

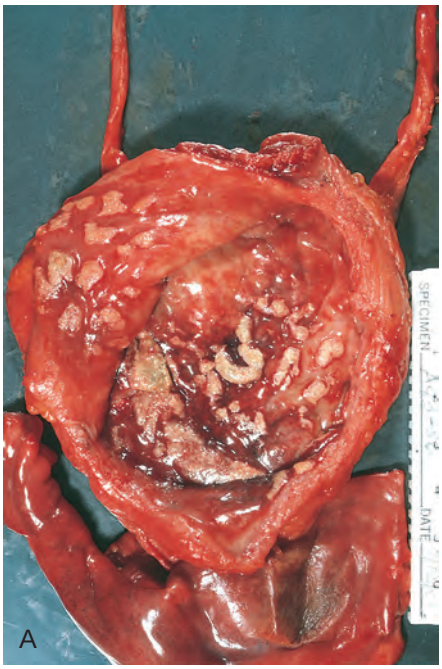
Several variants of cystitis have distinctive causes or morphologic appearances.

Interstitial Cystitis (Chronic Pelvic Pain Syndrome). Interstitial cystitis is a disorder of unknown etiology that occurs most frequently in women. The associated pain syndrome is defined by the American Urological Association as "an unpleasant sensation (pain, pressure, discomfort) perceived to be related to the urinary bladder, associated with urinary tract symptoms of more than six weeks duration, in the absence of infection or other identifiable causes." It is characterized by intermittent, often severe, suprapubic pain; urinary frequency; urgency; hematuria; and dysuria. Typical cystoscopic findings include mucosal fissures and punctate hemorrhages (glomerulations). Microscopically, the pathologic findings are nonspecific; mast cells are often increased in the submucosa, but the pathogenic significance of this finding is uncertain. The main role of biopsy is to rule out carcinoma in situ (CIS), which clinically mimics interstitial cystitis. Treatment is largely empiric. Some cases are associated with chronic mucosal ulcers (Hunner ulcers); this is termed the late (classic, ulcerative) phase. Late in the course, transmural fibrosis may lead to a contracted bladder.

Malakoplakia. Malakoplakia is a distinctive chronic inflammatory reaction that appears to stem from acquired defects in phagocyte function. It arises in the setting of chronic bacterial infection, mostly by *E. coli* or occasionally *Proteus* species, and occurs with increased frequency in the setting of immunosuppression, such as in renal transplant recipients.

MORPHOLOGY

Cystoscopically, malakoplakia takes the form of soft yellow, slightly raised mucosal plaques, 3 to 4 cm in diameter (Fig. 21.3A). Histologically, the lesion is composed of aggregates of large foamy macrophages with occasional multinucleate giant cells and lymphocytes. The **macrophages have abundant granular cytoplasm** due to phagosomes stuffed with particulate and membranous debris of bacterial origin, which are believed to accumulate due to defective phagolysosome function. In addition, laminated mineralized concretions resulting from deposition of calcium in enlarged lysosomes, known as **Michaelis-Gutmann bodies,** are typically present within the macrophages (Fig. 21.3B). Similar lesions have been described in other organs including the colon and other parts of the gastrointestinal tract, brain, lungs, bones, endometrium, kidneys, prostate, and epididymis.



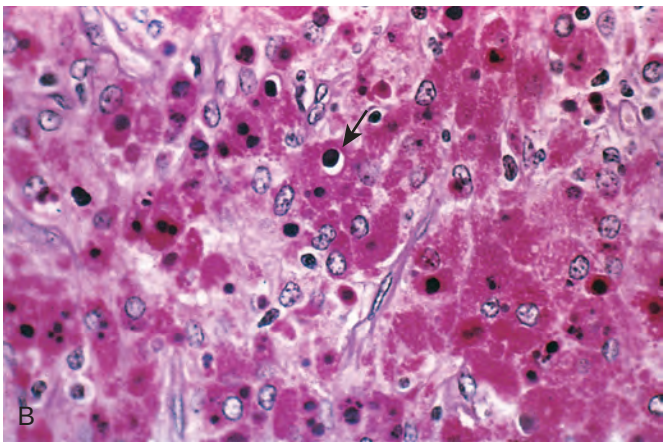


Figure 21.3 Malakoplakia. (A) Bladder involved by malakoplakia showing the characteristic yellow-orange mucosal lesions. (B) Periodic acid–Schiff (PAS) stain. Note the large macrophages with granular PAS-positive cytoplasm and several dense, round Michaelis-Gutmann bodies surrounded by artifactual cleared holes in the upper middle field *(arrow)*.

Polypoid Cystitis. Polypoid cystitis is an inflammatory lesion resulting from irritation of the bladder mucosa, most commonly as a result of instrumentation, including

indwelling catheters. The urothelium is thrown into broad bulbous polypoid projections as a result of marked submucosal edema. Polypoid cystitis may be mistaken for papillary urothelial carcinoma, both clinically and histologically.

Metaplastic Lesions

Several lesions that stem from true metaplasia or conditions that mimic metaplasia affect the bladder.

- *Cystitis glandularis and cystitis cystica.* These are common lesions of the urinary bladder in which nests of urothelium (von Brunn nests) grow downward into the lamina propria. Here, epithelial cells in the center of the nest undergo metaplasia and take on a cuboidal or columnar appearance *(cystitis glandularis)* or retract to produce cystic spaces lined by flattened urothelium *(cystitis cystica)*. Because the two processes often coexist, the condition is typically referred to as *cystitis cystica et glandularis*. In a variant of cystitis, glandularis goblet cells are present, and the epithelium resembles intestinal mucosa *(intestinal or colonic metaplasia)*. Both variants can arise in the setting of inflammation and metaplasia. Extensive and multifocal intestinal metaplasia is a precursor to adenocarcinoma.
- *Squamous metaplasia.* As a response to chronic injury, the urothelium is often replaced by nonkeratinizing or keratinizing squamous epithelium, which is a more durable lining. This should be distinguished from glycogenated squamous epithelium that is normally found in women at the trigone region of bladder. Extensive multifocal keratinizing squamous metaplasia is a precursor to dysplastic lesions and in situ and invasive squamous cell carcinoma. A classic example of this sequence is seen with bladder schistosomiasis, which commonly produces squamous metaplasia and is associated with squamous cell carcinoma in areas in which it is endemic.
- *Nephrogenic adenoma.* Nephrogenic adenoma is an unusual lesion that may not be a form of true metaplasia. There is convincing evidence from renal transplant recipients showing that at least some of these lesions are caused by implantation and growth of renal tubular cells at sites of bladder mucosa erosion. The overlying urothelium is focally replaced by cuboidal epithelium, which can assume a papillary growth pattern. Although the lesions are typically less than 1 cm in size, larger lesions have been reported that can produce signs and symptoms that raise a suspicion of cancer. In addition, the tubular proliferation can infiltrate the underlying lamina propria and superficial detrusor muscle, microscopically mimicking a malignant process.

KEY CONCEPTS

INFLAMMATORY DISORDERS AND METAPLASIAS OF THE BLADDER

- The bladder can be involved by a number of inflammatory lesions, many of which manifest with frequency and dysuria.
- Acute or chronic bacterial cystitis is extremely common, particularly in women, and results from retrograde spread of colonic bacteria in most cases.
- Other forms of cystitis have iatrogenic causes, such as radiation cystitis and hemorrhagic cystitis due to chemotherapy.
- Some inflammatory or metaplastic bladder lesions are significant in that they may clinically and/or histologically mimic bladder cancer, including malakoplakia, polypoid cystitis, cystitis cystica et glandularis, and nephrogenic adenoma.

Neoplasms

Bladder cancer is the ninth most common cancer type worldwide and is responsible for significant morbidity and mortality. It is the fourth most common cancer in American men (7% of all new cases). The overwhelming majority (>95%) of bladder tumors are of epithelial origin (Table 21.3), with urothelial neoplasms being by far the most common type followed by squamous and glandular neoplasms.

Urothelial Neoplasms

Urothelial neoplasms represent about 90% of all bladder tumors and run the gamut from small benign lesions that do not recur to aggressive cancers that are often fatal. Many of these tumors are multifocal at presentation. Though most common in the bladder, all of the urothelial lesions described here may be seen at any site where there is urothelium, from the renal pelvis to the distal urethra.

There are two distinct precursor lesions to invasive urothelial carcinoma: noninvasive papillary tumors and flat noninvasive urothelial carcinoma in situ (CIS) (Fig. 21.4). The most common precursor lesions are the noninvasive

Table 21.3 Tumors of the Urinary Bladder

Urothelial (transitional) tumors
• Noninvasive urothelial (transitional cell) tumors
• Infiltrating urothelial carcinoma
• Variants: nested, microcystic, micropapillary, plasmacytoid, sarcomatoid, giant cell, poorly differentiated, lipid-rich, and clear cell
Adenocarcinoma
Squamous cell carcinoma
Mixed carcinoma
Small-cell carcinoma
Sarcomas

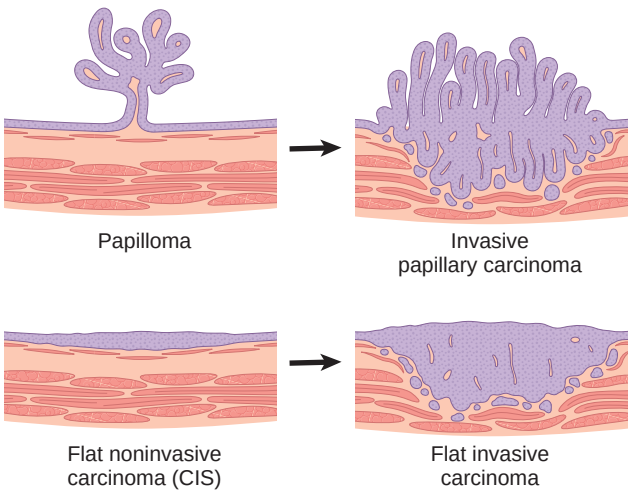


Figure 21.4 Four morphologic patterns of bladder tumors. *CIS*, Carcinoma in situ.

papillary tumors, which originate from papillary urothelial hyperplasia. These tumors have a range of atypical changes and are graded according to their biologic behavior. The other precursor lesion to invasive carcinoma, flat noninvasive urothelial carcinoma, is referred to as CIS. As discussed in Chapter 7, CIS is a term used to describe epithelial lesions that have the cytologic features of malignancy but are confined to the epithelium, showing no evidence of basement membrane invasion. Such lesions are considered to be high grade. In about one-half of individuals with invasive bladder cancer, the tumor has already invaded the bladder wall at the time of presentation, and precursor lesions are not detected. It is presumed in such cases that the precursor lesion was destroyed by the high-grade invasive component, which typically appears as a large, ulcerated mass. Although invasion into the lamina propria worsens the prognosis, the major decrease in survival is associated with invasion of the muscularis propria (detrusor muscle). Once muscularis propria invasion occurs, there is a 30% 5-year mortality rate.

Epidemiology. The incidence of bladder cancer is higher in men (male-to-female ratio of 3:1), in higher income nations, and in urban dwellers. About 80% of patients are between 50 and 80 years of age. Bladder cancer, with rare exceptions, is not familial.

Several factors have been implicated in the causation of urothelial carcinoma, including the following:

- *Cigarette smoking* is clearly the most important influence, increasing the risk threefold to sevenfold, depending on the duration and type of tobacco use. Between 50% and 80% of all bladder cancers among men are associated with the use of cigarettes. Cigars, pipes, and smokeless tobacco are associated with a smaller risk.
- *Industrial exposure to aryl amines,* particularly 2-naphthylamine and related compounds, as pointed out in our earlier discussion of chemical carcinogenesis (Chapter 7). Cancer appears 15 to 40 years after the first exposure.
- *Schistosoma haematobium* infections in endemic areas (Egypt, Sudan) are an established risk. The ova are deposited in the bladder wall and incite a brisk chronic inflammatory response that induces progressive mucosal squamous metaplasia, dysplasia, and, in some instances, neoplasia. Seventy percent of the cancers are squamous, the remainder being urothelial or glandular (the least common type).
- *Long-term use of analgesics* is implicated, as it is in analgesic nephropathy (Chapter 20).
- *Heavy long-term exposure to cyclophosphamide,* an immunosuppressive agent, induces, as noted, hemorrhagic cystitis and increases the risk of bladder cancer.
- *Irradiation,* often administered for other pelvic malignancies, increases the risk of urothelial carcinoma occurring several years following exposure.

Pathogenesis

Analysis of the genomes of precursor lesions and invasive bladder carcinomas has identified two relatively distinct molecular pathways of tumor progression (Fig. 21.5). Non–muscle-invasive papillary cancers often have gain-of-function alterations that increase signaling through growth factor receptor pathways, such as amplifications of the *FGFR3* tyrosine kinase receptor gene and activating mutations in the genes encoding RAS and PI 3-kinase. These tumors frequently recur but progress to muscle-invasive bladder cancer in only about 20% of cases. The majority of muscle-invasive bladder cancers develop by progression from "flat" CIS. Mutations that disrupt the function of p53 and RB are prevalent in all muscle-invasive cancers, but occur early in the development of CIS and later in the progression of papillary cancers.

Additional work is ongoing to further characterize the genetic and molecular landscape of bladder cancer. Consistent with epidemiologic data suggesting that environmental carcinogens have an important role, bladder cancers have a high burden of somatic mutations, comparable to other carcinogen-induced cancers, such as lung cancer and melanoma. As in other cancers, recurrently mutated

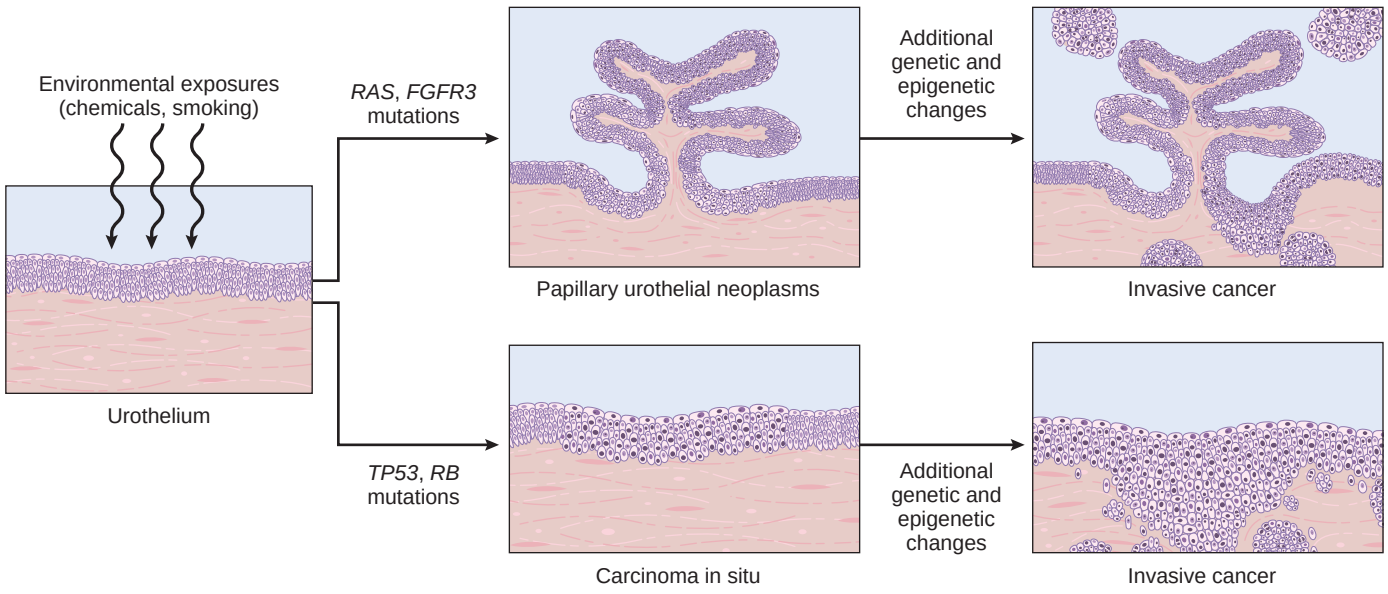


Figure 21.5 Salient environmental risk factors and major molecular pathways of bladder cancer development from carcinoma in situ and papillary bladder neoplasms. See text for details.

oncogenes and tumor suppressor genes include genes involved in cell cycle regulation, chromatin regulation, DNA repair, and growth factor signaling pathways. Analyses of RNA expression suggest the existence of several relatively distinct molecular subtypes; further work is needed to fully understand their biologic and clinical significance.

MORPHOLOGY

The appearance of urothelial tumors varies from purely papillary to nodular or flat. Papillary lesions are red, elevated excrescences ranging in size from less than 1 cm in diameter to large masses up to 5 cm in diameter (Fig. 21.6). Multiple discrete tumors are often present.

Table 21.3 lists the grading system of urothelial tumors. **Papillomas** represent 1% or less of bladder tumors and are often seen in younger patients. These tumors typically arise singly as small (0.5 to 2 cm), delicate structures superficially attached to the mucosa by a stalk and are referred to as **exophytic papillomas.** The individual finger-like papillae have a central core of loose fibrovascular tissue covered by epithelium that is histologically identical to normal urothelium (Fig. 21.7A). Recurrences and progression are rare but

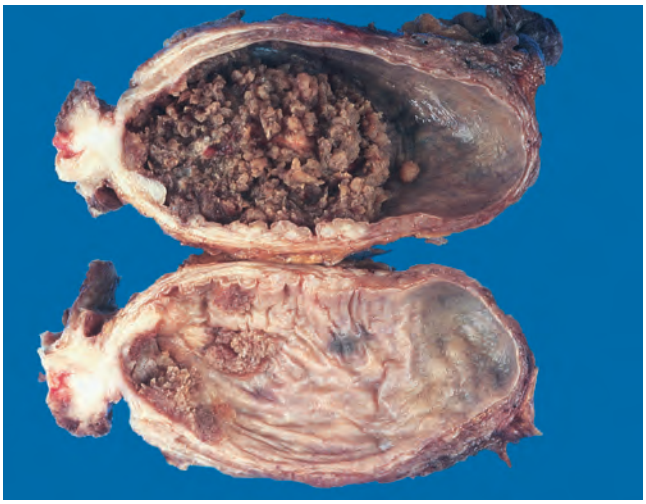


Figure 21.6 Cross-section of bladder with the upper section showing a large papillary tumor. The lower section demonstrates multifocal smaller papillary neoplasms. (Courtesy Dr. Fred Gilkey, Sinai Hospital, Baltimore, Md.)

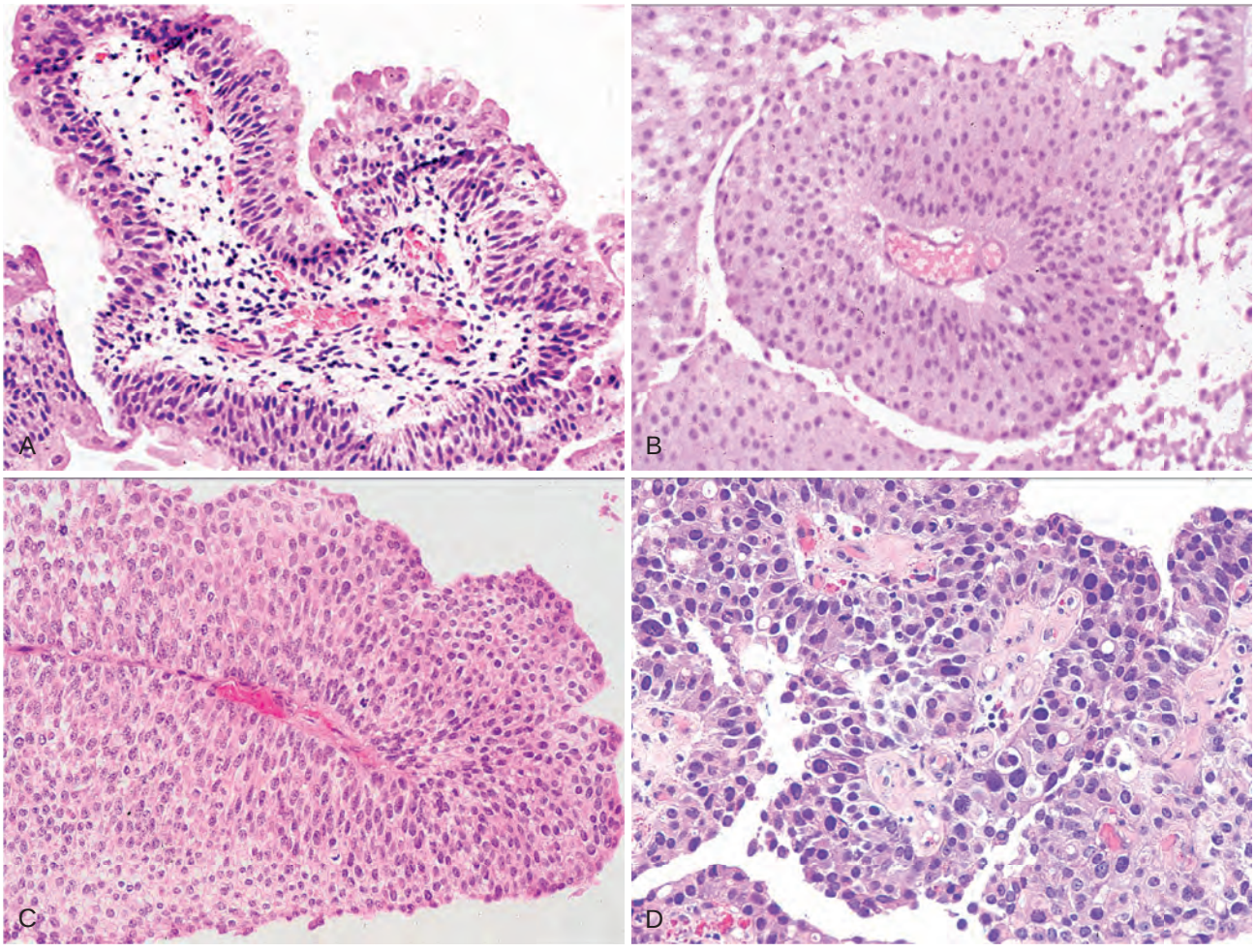


Figure 21.7 Papillary urothelial neoplasms. (A) Papilloma consisting of small papillary fronds lined by normal-appearing urothelium. (B) Papillary urothelial neoplasms of low malignant potential (PUNLMP) showing thicker urothelium with greater density of cells. (C) Low-grade papillary urothelial carcinoma with an overall orderly appearance, with scattered hyperchromatic nuclei and mitotic figures *(upper left)*. (D) High-grade papillary urothelial carcinoma with marked cytologic atypia.

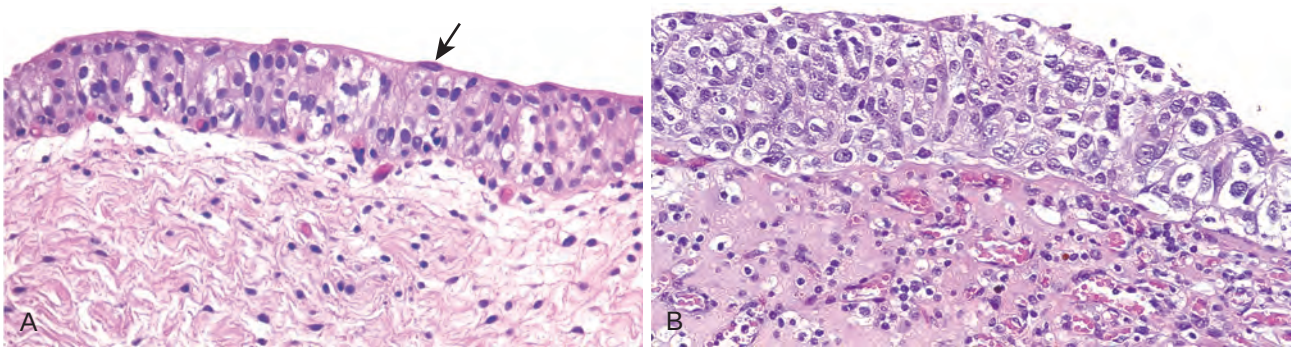


Figure 21.8 Bladder carcinoma in situ. (A) Normal urothelium with uniform nuclei and well-developed umbrella cell layer. (B) Carcinoma in situ showing a disorganized urothelium containing numerous cells having markedly enlarged and pleomorphic nuclei.

may occur. In contrast to exophytic papillomas, **inverted papillomas** are completely benign lesions consisting of inter-anastomosing cords of cytologically bland urothelium that extend down into the lamina propria; they simulate an invasive process.

Papillary urothelial neoplasms of low malignant potential (PUNLMP) share many histologic features with papillomas, differing only in having thicker urothelium with greater density of cells (Fig. 21.7B) At cystoscopy, these tumors tend to be larger than papillomas and may be indistinguishable from papillary cancers. Recurrent tumors usually show the same morphology; progression to tumors of higher grade may occur but is rare.

Low-grade papillary urothelial carcinomas have an orderly architectural appearance and low-grade cytologic atypia. The cells are evenly spaced (i.e., maintain polarity) and cohesive. There are scattered hyperchromatic nuclei, infrequent mitotic figures predominantly toward the base, and slight variation in nuclear size and shape (Fig. 21.7C). These low-grade cancers may recur and, infrequently, may also invade. Only rarely do these tumors pose a threat to life.

High-grade papillary urothelial carcinomas contain dyscohesive cells with large hyperchromatic nuclei, irregular nuclear chromatin, and prominent nucleoli. Some of the tumor cells are highly anaplastic (Fig. 21.7D). Mitotic figures, including atypical ones, are frequent. Architecturally, there is disarray and loss of polarity. As compared to low-grade lesions, these tumors have a much higher incidence of progression to muscle-invasive bladder cancer and have a significant potential for metastasis to regional lymph nodes and systemic spread (e.g., to liver and lung).

CIS (or flat urothelial carcinoma) is defined by the presence of cytologically malignant cells within a flat urothelium (Fig. 21.8). CIS may range from full-thickness cytologic atypia to scattered malignant cells in an otherwise normal urothelium, the latter termed **pagetoid spread.** A common feature shared with high-grade papillary urothelial carcinoma is a lack of cohesiveness, which leads to shedding of malignant cells into the urine. When shedding is extensive, only a few CIS cells may be left clinging to a largely denuded basement membrane. On cystoscopy CIS usually appears as an area of mucosal reddening, granularity, or thickening without an evident intraluminal mass. It is commonly multifocal and may involve most of the bladder surface and extend into the ureters and urethra. **If untreated, 50% to 75% of CIS progresses to invasive cancer.**

Invasive urothelial carcinoma (Fig. 21.9) may be associated with papillary urothelial cancer, usually high grade, or adjacent CIS. The extent of spread (stage), based primarily on depth of invasion in the bladder wall, at the time of initial diagnosis is the most important prognostic factor. Stage also determines treatment modality, with invasion of the muscularis propria layer being an indication for radical cystectomy or radiation therapy with neoadjuvant or adjuvant chemotherapy (Table 21.4).

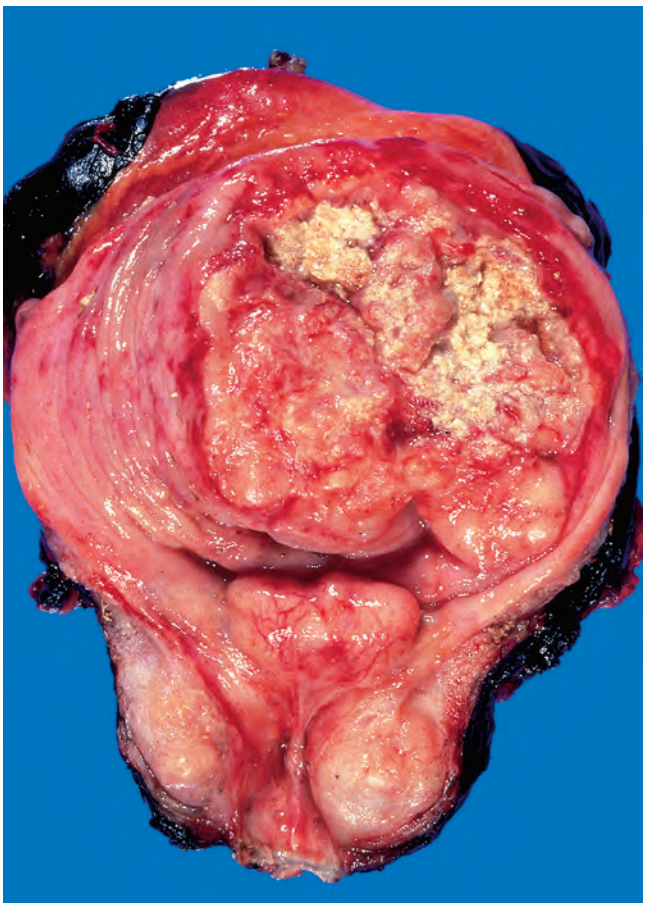


Figure 21.9 Opened bladder showing a high-grade invasive transitional cell carcinoma at an advanced stage. The aggressive multinodular neoplasm has fungated into the bladder lumen and spread over a wide area. The yellow regions represent areas of ulceration and necrosis.

Table 21.4 Grading of Noninvasive Urothelial (Transitional Cell) Tumors

WHO/ISUP Grades (2016)
Flat Lesions
• Urothelial proliferation of uncertain malignant potential (flat hyperplasia) • Urothelial dysplasia • Urothelial carcinoma in situ
Exophytic Papillary Lesions
• Papilloma • Urothelial proliferation of uncertain malignant potential (papillary hyperplasia) • Papillary urothelial neoplasms of low malignant potential • Papillary urothelial carcinoma, low grade • Papillary urothelial carcinoma, high grade

ISUP, International Society of Urological Pathology; *WHO*, World Health Organization.

Unusual variants of urothelial cancer include a nested variant with deceptively bland cytology, lymphoepithelioma-like carcinoma, micropapillary carcinoma, and sarcomatoid carcinoma (see Table 21.3).

Clinical Features

Painless hematuria is the most common symptom of bladder cancer. Frequency, urgency, and dysuria may accompany hematuria. Occasionally, obstruction of the ureteral orifice may lead to pyelonephritis or hydronephrosis.

The subsequent treatment and course of bladder carcinoma depend on tumor grade and stage (Table 21.5), with the key variable being whether the tumor is muscle-invasive or not. Noninvasive papillary urothelial tumors and those that solely invade the lamina propria (together referred to as non–muscle-invasive) constitute the majority (70% to 80%) of urothelial neoplasms. Twenty percent to 30% of bladder cancers are muscle-invasive with only a minority of these originating in patients with a prior history of non–muscle-invasive disease.

Initial treatment of non–muscle-invasive tumors is guided by the pathologic findings. For small, localized low-grade papillary tumors, diagnostic transurethral resection is the only procedure needed. CIS and papillary tumors that are large, high grade, multifocal, have a history of recurrence, or are associated with lamina propria invasion are treated with intravesical instillation of an attenuated strain of *Mycobacterium bovis* called bacillus Calmette-Guérin (BCG), which elicits a local inflammatory reaction that destroys the tumor.

Unfortunately, non–muscle-invasive bladder cancers have a high tendency to recur (up to 70%) and are at risk for progression to a higher grade or stage. The risk of recurrence and progression is related to several variables, including tumor size, stage, grade, multifocality, prior recurrence, and the presence of CIS in the surrounding mucosa. Most recurrences appear at sites other than that of the original lesion, yet are clonally related, apparently arising from shedding and implantation of cells from the original tumor at a distant site. Other recurrences are clonally distinct and represent a second primary tumor. Papillomas, PUNLMP, and low-grade papillary urothelial cancer have a 98% 10-year survival rate regardless of the number of recurrences, and only a few tumors (<10%) progress to higher-grade lesions. In contrast, high-grade papillary urothelial carcinomas invade and lead to death in about 25% of cases. Patients with CIS only, as opposed to CIS associated with infiltrating urothelial carcinoma, are less likely to progress to muscle-invasive cancer (28% versus 59%) or die of disease (7% versus 45%).

Because of the risk of recurrence and progression, patients with non–muscle-invasive urothelial neoplasms need lifelong surveillance and follow-up. Surveillance cystoscopy, biopsies, and urine cytologic examinations contribute mightily to the huge burden on health care resources (over 3 billion dollars annually in the United States alone) imposed by bladder cancer. Several noninvasive assays to screen for recurrence have been developed, but none have been proven to have sufficient sensitivity and specificity to replace cystoscopy.

Muscle-invasive bladder carcinoma is treated with radical cystectomy or cystoprostatectomy, or radiation therapy with neoadjuvant and/or adjuvant chemotherapy. In addition

Table 21.5 Pathologic Stage Classification of Urinary Bladder Carcinoma (Primary Tumor)

pTNM, AJCC 8th Edition	
Tumor	
pTX	Primary tumor cannot be assessed
pT0	No evidence of primary tumor
Ta	Noninvasive papillary carcinoma
Tis	Urothelial carcinoma in situ: "flat tumor"
T1	Tumor invades lamina propria (subepithelial connective tissue)
T2	Tumor invades muscularis propria
T2a	Tumor invades superficial muscularis propria (inner half)
T2b	Tumor invades deep muscularis propria (outer half)
T3	Tumor invades perivesical soft tissue
T3a	Tumor invades perivesical soft tissue microscopically
T3b	Tumor invades perivesical soft tissue macroscopically (extravesical mass)
T4	Extravesical tumor directly invades any of the following: prostatic stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall
T4a	Extravesical tumor invades directly into prostatic stroma, seminal vesicles, uterus, or vagina
T4b	Extravesical tumor invades pelvic wall or abdominal wall
Regional Lymph Nodes (pN)	
NX	Lymph nodes cannot be assessed
N0	No lymph node metastasis
N1	Single regional lymph node metastasis in the true pelvis (perivesical, obturator, internal and external iliac, or sacral lymph node)
N2	Multiple regional lymph node metastasis in the true pelvis (perivesical, obturator, internal and external iliac, or sacral lymph node metastasis)
N3	Lymph node metastasis to the common iliac lymph nodes
Metastasis	
M1	Distant metastasis
M1a	Distant metastasis limited to lymph nodes beyond the common iliac arteries
M1b	Non–lymph node distant metastases

AJCC, American Joint Commission on Cancer.

to muscle-invasive cancers, radical cystectomy is also indicated in cases of (1) CIS or high-grade papillary cancer refractory to BCG and other intravesical therapies; (2) CIS extending into the prostatic urethra and ducts, sites where instilled BCG does not come into contact with the neoplastic cells; and (3) occasional cases of non–muscle-invasive papillary urothelial high-grade carcinoma, such as multifocal lesions that are too large and extenisive to completely eradicate by transurethral resection.

Once bladder carcinoma metastasizes, treatment options are limited. Most metastatic tumors respond poorly to chemotherapy, which produces 5-year survival rates of only 15%. As with other cancers linked to environmental carcinogens with high mutational loads, a subset (roughly 30%) of metastatic bladder carcinomas respond to immune checkpoint inhibitors, sometimes dramatically, providing some hope for this group of patients.

Other Epithelial Bladder Tumors

Squamous cell carcinomas resembling squamous cell cancers occurring at other sites make up 3% to 7% of bladder cancers in the United States but are much more frequent in countries where urinary schistosomiasis is endemic. Pure squamous cell carcinomas arise from atypical keratinizing mucosa (squamous dysplasia and CIS) and are nearly always associated with chronic bladder irritation and infection. *Mixed urothelial carcinoma with areas of squamous carcinoma* is more frequent than pure squamous cell carcinoma. Most are invasive, fungating tumors or are infiltrative and ulcerative. The level of cellular differentiation varies widely, from well-differentiated lesions producing abundant keratin to anaplastic tumors with only focal evidence of squamous differentiation. *Adenocarcinoma* of the bladder is rare and histologically identical to adenocarcinomas seen in the gastrointestinal tract. Some arise from urachal remnants or in association with extensive intestinal metaplasia (discussed earlier).

Small-cell carcinoma, indistinguishable from small-cell carcinomas of the lung, occasionally arise in the bladder, often in association with urothelial, squamous, or adenocarcinoma (see Table 21.3). As with small-cell carcinomas in other anatomic sites, these very aggressive tumors are strongly associated with loss-of-function mutations in the *TP53* and *RB* tumor suppressor genes.

Mesenchymal Tumors

Benign Tumors. A great variety of benign mesenchymal tumors may arise in the bladder, having the histologic features of their counterparts elsewhere. Even collectively, they are rare. The most common is *leiomyoma.* They tend to grow as isolated, intramural (submucosal), encapsulated, oval-to-spherical masses up to several centimeters in diameter.

Sarcomas. True sarcomas are distinctly uncommon in the bladder. Inflammatory myofibroblastic tumors and various carcinomas that assume sarcomatoid growth patterns are more common than true sarcomas. As a group, sarcomas tend to produce large masses (up to 15 cm in diameter) that protrude into the vesicle lumen. Their soft, fleshy, gray-white gross appearance suggests their mesenchymal nature. The most common bladder sarcoma in infancy or childhood is *embryonal rhabdomyosarcoma.* In some of these cases they manifest as a polypoid grape-like mass *(sarcoma botryoides).* The most common bladder sarcoma in adults is *leiomyosarcoma* (Chapter 26).

Secondary Tumors

Secondary malignant involvement of the bladder is most often due to direct extension of cancers arising in adjacent organs, mainly the cervix, uterus, prostate, and rectum. Lymphoma may involve the bladder as a component of systemic disease. Metastatic spread of solid tumors to the bladder may occur but is very rare.

KEY CONCEPTS

BLADDER NEOPLASMS

- Bladder cancer is most common in older males, and cigarette smoking constitutes one of the most important risk factors.
- Painless hematuria is a common presenting symptom of bladder cancer and requires clinical investigation by cystoscopy and/or urine cytology analysis to rule out urothelial neoplasia.
- Non–muscle-invasive bladder cancer is associated with gain-of-function mutations in growth factor signaling pathways and frequently recur but rarely progress in stage or grade.
- Muscle-invasive bladder cancers are associated with inactivation of *TP53* and *RB* tumor suppressor genes and often develop from "flat" carcinoma in situ, with or without a high-grade papillary component.
- Molecular subtyping of bladder cancer has identified several molecular subtypes that are being evaluated for prognostic and potential therapeutic significance.
- Other epithelial bladder tumor variants may occur, either alone or mixed with urothelial carcinoma, including squamous cell carcinoma, adenocarcinoma, and small cell carcinoma.

Obstruction

Obstruction of the bladder outlet is of major clinical importance because of its eventual effect on the kidney. In males, the most common cause is enlargement of the prostate gland due to benign prostatic hyperplasia (BPH). Bladder obstruction is less common in females and is most often caused by cystocele of the bladder. Infrequent causes are (1) congenital urethral strictures; (2) inflammatory urethral strictures; (3) inflammatory fibrosis and contraction of the bladder; (4) bladder tumors, either benign or malignant; (5) invasion of the bladder neck by tumors arising in contiguous organs; (6) mechanical obstructions caused by foreign bodies and calculi; and (7) injury of nerves controlling bladder contraction (neurogenic bladder).

MORPHOLOGY

In the early stages there is only thickening of the bladder wall due to smooth muscle hypertrophy. With progressive hypertrophy the individual muscle bundles greatly enlarge and produce trabeculation of the bladder wall (see Fig. 21.9). In the course of time, crypts form and may be converted into diverticula.

In some cases of acute obstruction or in terminal disease when the patient's normal reflex mechanisms are depressed, the bladder may become extremely dilated. The enlarged bladder may reach the brim of the pelvis or even the level of the umbilicus. In such cases, the bladder wall is markedly thinned and lacks trabeculations.

URETHRA

Inflammation

Urethritis is classically divided into gonococcal and nongonococcal causes. *Gonococcal urethritis* is one of the earliest manifestations of this venereal infection. *Nongonococcal urethritis* is common and can be caused by several different organisms. Various strains of *Chlamydia* (e.g., *Chlamydia trachomatis*) are the cause of 25% to 60% of nongonococcal urethritis in men and about 20% in women. *Mycoplasma (Ureaplasma urealyticum)* also accounts for symptoms of urethritis in many cases. Urethritis is often accompanied by cystitis in women and by prostatitis in men. In many instances of suspected bacterial urethritis, no organism can be isolated. Some urethritis is truly noninfectious in origin. An example of such an inflammatory urethritis is a disorder known as *reactive arthritis* (formerly Reiter syndrome), which is associated with the clinical triad of arthritis, conjunctivitis, and urethritis (Chapter 26).

The morphologic changes are entirely typical of inflammation in other sites within the urinary tract. The urethral involvement is not itself a serious clinical problem but may cause considerable local pain, itching, and frequency and may warn of more serious disease at higher levels of the urogenital tract.

Tumors and Tumor-Like Lesions

Urethral caruncle is an inflammatory lesion that presents as a small, red, painful mass about the external urethral meatus, typically in older females. It consists of inflamed granulation tissue covered by an intact but friable mucosa, which may ulcerate and bleed with the slightest trauma. Surgical excision affords prompt relief and cure.

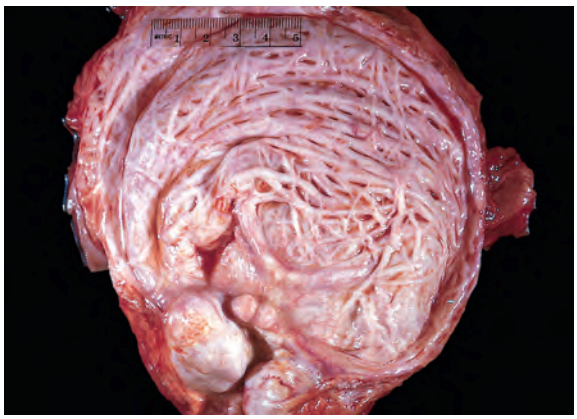


Figure 21.10 Hypertrophy and trabeculation of bladder wall secondary to benign prostatic hyperplasia.

Benign epithelial tumors of the urethra include squamous and urothelial papillomas, inverted urothelial papillomas, and condylomas. *Primary carcinoma of the urethra* is an uncommon lesion (Fig. 21.10). Cancers arising within the proximal urethra tend to show urothelial differentiation and are analogous to those occurring within the bladder, whereas lesions originating within the distal urethra are more often squamous cell carcinomas and HPV-related. Adenocarcinomas are infrequent in the urethra and generally occur in women. Neoplasms of the prostatic urethra are addressed later in the section on the prostate.

The Male Genital Tract

PENIS

Congenital Anomalies

The penis is involved by many congenital anomalies; only those that are clinically significant are discussed here.

Hypospadias and Epispadias

Malformation of the urethral groove and canal may create an abnormal opening either on the ventral surface of the penis *(hypospadias)* or on the dorsal surface *(epispadias)*. Either of these anomalies can be associated with failure of normal descent of the testes and with malformations of the urinary tract. Hypospadias, the more common of the two, occurs in approximately 1 in 300 live male births. Even when isolated, these urethral defects may have clinical significance because the abnormal opening is often constricted, resulting in urinary tract obstruction and an increased risk of ascending infections. When the orifices are situated near the base of the penis, normal ejaculation and insemination are hampered and may be a cause of sterility.

Phimosis

When the orifice of the prepuce is too small to permit its normal retraction, the condition is designated *phimosis*. An abnormally small orifice may result from anomalous development but is more frequently the result of repeated bouts of infection that cause scarring of the preputial ring. Phimosis is important because it interferes with cleanliness and permits the accumulation of secretions and detritus under the prepuce, favoring the development of secondary infections and increasing the risk for penile carcinoma.

Inflammation

Inflammations of the penis almost invariably involve the glans and prepuce and include a wide variety of specific and nonspecific infections. The specific infections—syphilis, gonorrhea, chancroid, granuloma inguinale, lymphopathia venerea, genital herpes—are sexually transmitted and are discussed in Chapter 8. Only the nonspecific infections causing so-called balanoposthitis are described here.

Balanoposthitis refers to infection of the glans and prepuce caused by a wide variety of organisms. Among the more

common agents are *C. albicans,* anaerobic bacteria, *Gardnerella,* and pyogenic bacteria. Most cases occur as a consequence of poor local hygiene in uncircumcised males, in whom the accumulation of desquamated epithelial cells, sweat, and debris, termed *smegma,* acts as a local irritant. Persistence of such infections leads to inflammatory scarring and, as mentioned earlier, is a common cause of phimosis.

Tumors

Tumors of the penis are uncommon. The most frequent neoplasms are squamous cell carcinoma and benign genital warts (condyloma acuminatum).

Benign Tumors and Tumor-Like Conditions

Condyloma Acuminatum

Condyloma acuminatum is a benign sexually transmitted wart caused by human papillomavirus (HPV). It is related to the common wart and may occur on any moist mucocutaneous surface of the external genitals in either sex. "Low-risk" HPV serotypes (HPV 6 and, less frequently, HPV 11) are the most frequent cause of condylomata acuminata.

MORPHOLOGY

Condylomata acuminata may arise in the external genitalia or perineal areas. Penile lesions usually occur in the coronal sulcus and inner surface of the prepuce. They consist of **single or multiple sessile or pedunculated, red papillary excrescences** that are several millimeters in diameter (Fig. 21.11). Histologically, a branching, villous, papillary connective tissue stroma is covered by epithelium that may have considerable superficial hyperkeratosis and thickening of the underlying epidermis **(acanthosis)** (Fig. 21.12A). The normal orderly maturation of the epithelial cells is preserved. The lining cells frequently display perinuclear cytoplasmic vacuolization (koilocytosis), characteristic of HPV infection (Fig. 21.12B).

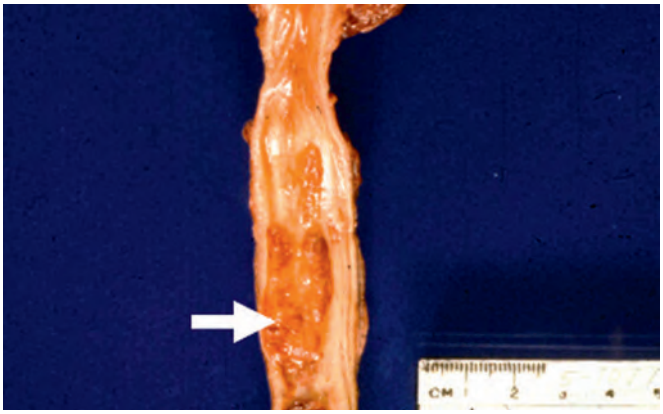


Figure 21.11 Carcinoma of urethra with typical fungating growth pattern *(arrow).*

Peyronie Disease

This disorder, which appears to be reactive rather than neoplastic, is characterized by hard penile plaques that result from the deposition of collagen in the connective tissue between the corpora cavernosa and the tunica albuginea. The fibrosis is believed to be the product of microvascular trauma and subsequent organizing sclerosing chronic inflammation. Clinically, the lesion results in penile curvature toward the side of the lesion and pain during intercourse. Treatments include surgery and injection of collagenase to lyse the fibrous plaques.

Malignant Tumors

Squamous Carcinoma in situ/Penile Intraepithelial Neoplasia

As at other sites, such as uterine cervix, in situ squamous lesions span a range of morphologies from CIS to less severe derangements. These lesions are encompassed by the umbrella term *penile intraepithelial neoplasia (PeIN).* All are squamous lesions confined to the epidermis by an intact basement membrane. PeIN may be HPV-related (undifferentiated PeIN) or non–HPV-related (differentiated). The latter is associated with balanitis xerotica obliterans, occurs on

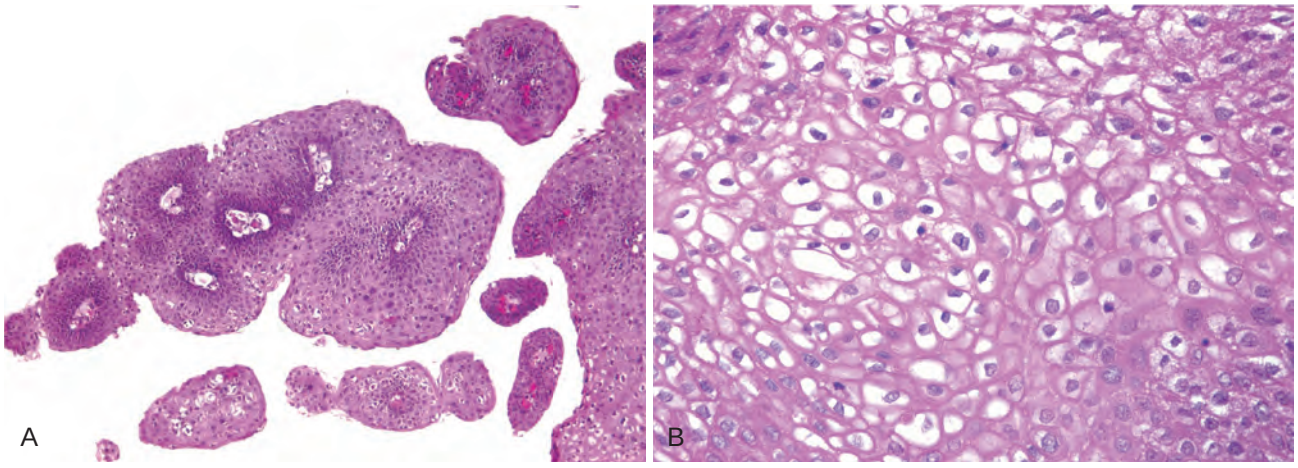


Figure 21.12 Condyloma acuminatum of the penis. (A) Low magnification reveals the papillary architecture and thickening of the epidermis. (B) On high magnification the epithelium shows perinuclear vacuolization (koilocytosis) characteristic of human papillomavirus infection.

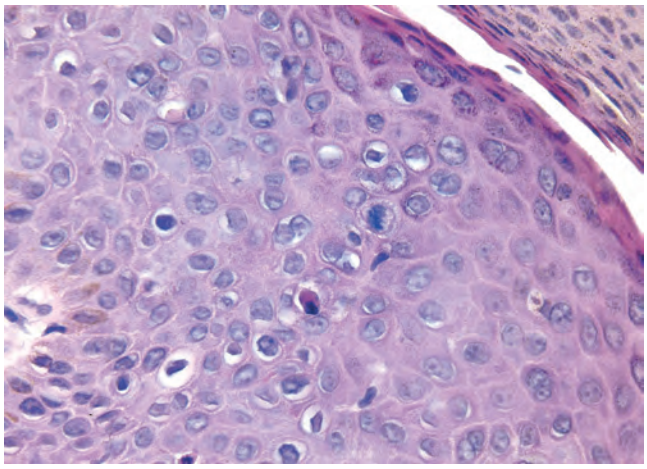


Figure 21.13 Bowen disease (carcinoma in situ) of the penis. Note the hyperchromatic, dysplastic dyskeratotic epithelial cells and scattered mitoses *(mid-right)* well above the basal layer.

the foreskin of older patients, and as the name implies retains a degree of squamous maturation. Undifferentiated PeIN is composed of more overtly malignant cells and may manifest clinically as two distinct lesions: Bowen disease and bowenoid papulosis. Both are associated with high-risk HPV, most commonly HPV 16.

- *Bowen disease* most commonly affects the penile shaft and scrotum of older men. At these sites, it appears as a solitary, thickened, gray-white, opaque plaque. In less common cases when it affects the glans, the lesion acquires a velvety red appearance. Histologically, the lesion consists of dysplastic squamous cells containing large hyperchromatic irregular nuclei and lacking orderly maturation (Fig. 21.13). Mitoses, some atypical, are numerous. In 10% of patients, Bowen disease gives rise to infiltrating squamous cell carcinoma.
- *Bowenoid papulosis* occurs in sexually active adults. It is distinguished from Bowen disease by the younger age of affected patients and its presentation as multiple (rather than solitary) reddish brown papular lesions. Although etiologically related to HPV 16 and histologically indistinguishable from Bowen disease, Bowenoid papulosis virtually never develops into invasive carcinoma and usually regresses spontaneously.

Invasive Squamous Cell Carcinoma

Squamous cell carcinoma of the penis is associated with poor genital hygiene and high-risk HPV infection. Penile carcinoma affects middle-aged and older patients (40 to 70 years of age). In the United States, it accounts for less than 1% of cancers in males. In contrast, in some parts of Asia, Africa, and South America, penile cancer accounts for 10% to 20% of male malignancies. Low income status and poor hygiene habits are salient risk factors. Circumcision confers protection, and hence penile cancer is extremely rare among Jews and Muslims and is correspondingly more common in populations in which circumcision is not practiced routinely. It is postulated that circumcision reduces exposure to carcinogens that may be concentrated in smegma and decreases the likelihood of infections with potentially oncogenic types of HPV. Only a minority of penile cancer is HPV-related; other factors such as poor hygiene and chronic inflammation presumably contribute to those that are not HPV-related. The availability of vaccines to both low-risk and high-risk subtypes of HPV may help reduce the incidence of penile cancer and condyloma acuminatum. Other risk factors include cigarette smoking and chronic inflammatory conditions such as *lichen sclerosis et atrophicus (balanitis xerotica obliterans).*

Pathogenesis

The contribution of high-risk HPV to penile carcinoma is identical to its role in cervical cancer (Chapters 7 and 22). These forms of HPV encode E6 and E7 proteins that inactivate the p53 and RB tumor suppressor proteins, leading to genomic instability and increased proliferation, respectively. E6 protein also stimulates telomerase expression, leading to cellular immortalization. E7 protein induces feedback loops that increase levels of the cyclin-dependent kinase inhibitor p16, a feature that can be used as a surrogate for the presence of high-risk HPV in tumor cells.

MORPHOLOGY

Squamous cell carcinoma of the penis usually originates in glans or inner surface of the prepuce near the coronal sulcus. Macroscopically the tumors may be irregular, fungating cauliflower-like masses; flat, indurated lesions; or large verruciform/papillary tumors (Fig. 21.14). The World Health Organization (WHO) 2016 classification of penile squamous cell carcinoma recognizes HPV-related and non–HPV-related categories, each with a number of different histologies (Table 21.6). Conventional (usual) squamous cell carcinoma is the most common HPV-negative type, encompassing almost half of all penile cancers. Pathologic prognostic factors in penile carcinoma include stage, grade, histologic subtype, vascular invasion, and perineural invasion. Many of the histologic subtypes are associated with distinct grades (e.g., verrucous and papillary carcinomas are well differentiated/grade 1 tumors, while sarcomatoid and basaloid carcinomas are poorly differentiated/grade 3 tumors. Among the special subtypes of penile carcinoma, several merit brief description. **Verrucous carcinoma** is an exophytic, warty well-differentiated, non–HPV-related variant that invades

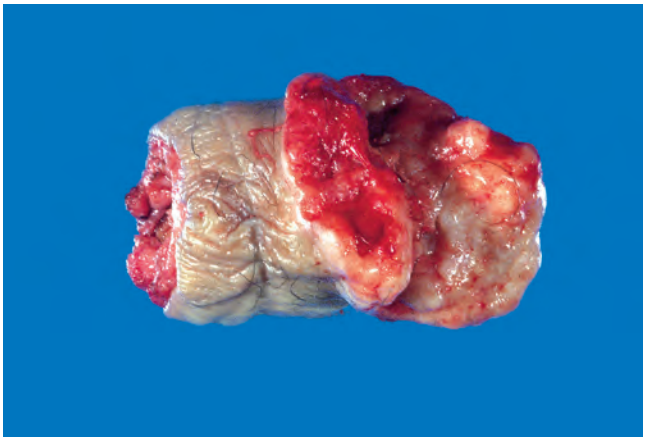


Figure 21.14 Carcinoma of the penis. The glans penis is deformed by a firm, ulcerated, infiltrative mass.

Table 21.6 World Health Organization 2016 Pathologic Classification of Squamous Cell Carcinoma of the Penis

Non–HPV-Related Squamous Cell Carcinoma
Squamous cell carcinoma, usual type
Pseudohyperplastic carcinoma
Pseudoglandular carcinoma
Verrucous carcinoma
Carcinoma cuniculatum
Papillary squamous cell carcinoma
Adenosquamous carcinoma
Sarcomatoid (spindle cell) carcinoma
Mixed squamous cell carcinoma
HPV-Related Squamous Cell Carcinoma
Basaloid squamous cell carcinoma
Papillary-basaloid carcinoma
Warty carcinoma
Warty-basaloid carcinoma
Clear cell squamous carcinoma
Lymphoepithelioma-like carcinoma
Others

HPV, Human papillomavirus.

From Cubilla A, Velazquez EF, Amin MB, et al: The World Health Organisation 2016 classification of penile carcinomas: a review and update from the International Society of Urological Pathology expert-driven recommendations, *Histopathology* 72(6):893–904, 2018.

locally along a broad pushing border, but rarely metastasizes. In contrast, **basaloid carcinoma** is an HPV-related tumor comprised of relatively small hyperchromatic cells that has a destructive pattern of invasion and usually pursues an aggressive course.

Clinical Features

Invasive squamous cell carcinoma of the penis is a slowly growing, locally invasive lesion that often has been present for a year or more before it is brought to medical attention. The lesions are nonpainful until they undergo secondary ulceration and infection. Metastases to inguinal lymph nodes may occur early in its course, but widespread dissemination is rare until the lesion is far advanced. Clinical assessment of regional lymph node involvement is notoriously inaccurate; only 50% of enlarged inguinal nodes detected in men with penile squamous cell carcinoma contain cancer, with the remainder showing only reactive lymphoid hyperplasia when examined histologically. As mentioned above, prognosis is strongly correlated with tumor stage at diagnosis. Patients with penile cancer metastasis to multiple, bilateral lymph or pelvic lymph nodes have a guarded prognosis with reported 5-year disease-specific survival rates ranging from 7% to 60%.

KEY CONCEPTS

CARCINOMA OF THE PENIS

- Most common in lower income parts of the world, such as Africa, Asia, and South America, due in part to a lower incidence of circumcision.
- Two pathogenic pathways, one related to HPV and the other unrelated to HPV.
- Squamous cell carcinoma occurs on the glans or shaft of the penis as an ulcerated infiltrative lesion that may spread to inguinal nodes and infrequently to distant sites.
- Other important lesions of the penis include congenital abnormalities involving the position of the urethra (epispadias, hypospadias) and inflammatory disorders (balanitis, phimosis).

TESTIS AND EPIDIDYMIS

Distinct pathologic conditions affect the testis and epididymis. In the epididymis, the most frequent conditions are inflammatory diseases, whereas tumors dominate in the testis.

Congenital Anomalies

With the exception of undescended testes (cryptorchidism), congenital anomalies are extremely rare and include absence of one or both testes and fusion of the testes (so-called *synorchism*).

Cryptorchidism

Cryptorchidism is a complete or partial failure of the intra-abdominal testes to descend into the scrotal sac and is associated with testicular dysfunction and an increased risk of testicular cancer. It is found in approximately 1% of 1-year-old boys. It usually occurs as an isolated anomaly but may be accompanied by other malformations of the genitourinary tract, such as hypospadias.

Testicular descent occurs in two phases. During the first transabdominal phase, the testis comes to lie within the lower abdomen or brim of the pelvis. This phase is controlled by the hormone müllerian-inhibiting substance. In the second inguinoscrotal phase, the testes descend through the inguinal canal into the scrotal sac. This phase is androgen-dependent and is thought to be mediated by androgen-induced release of calcitonin gene–related peptide from the genitofemoral nerve. The testes may arrest anywhere along their pathway of descent; the most common site is in the inguinal canal, while arrest within the abdomen is uncommon, accounting for approximately 5% to 10% of cases. Even though testicular descent is controlled by hormonal factors, cryptorchidism is only rarely associated with a well-defined hormonal disorder.

MORPHOLOGY

Cryptorchidism is usually unilateral, being bilateral in 25% of patients. Cryptorchid testes are small and firm. The histologic changes in the malpositioned testis begin as early as 2 years of age. Early on, **thickening of the basement membrane** of the spermatic tubules is seen (Fig. 21.15). Subsequent loss of spermatogonia leaves the tubules with only Sertoli cells. The scarred tubules may appear as dense cords of hyaline connective tissue associated with a concomitant increase in interstitial stroma. Leydig cells are spared and therefore appear relatively prominent. Similar histologic changes may also be seen in the contralateral (descended) testis in males with unilateral cryptorchidism, suggesting that cryptorchidism is a marker of an intrinsic defect in gonadal development.

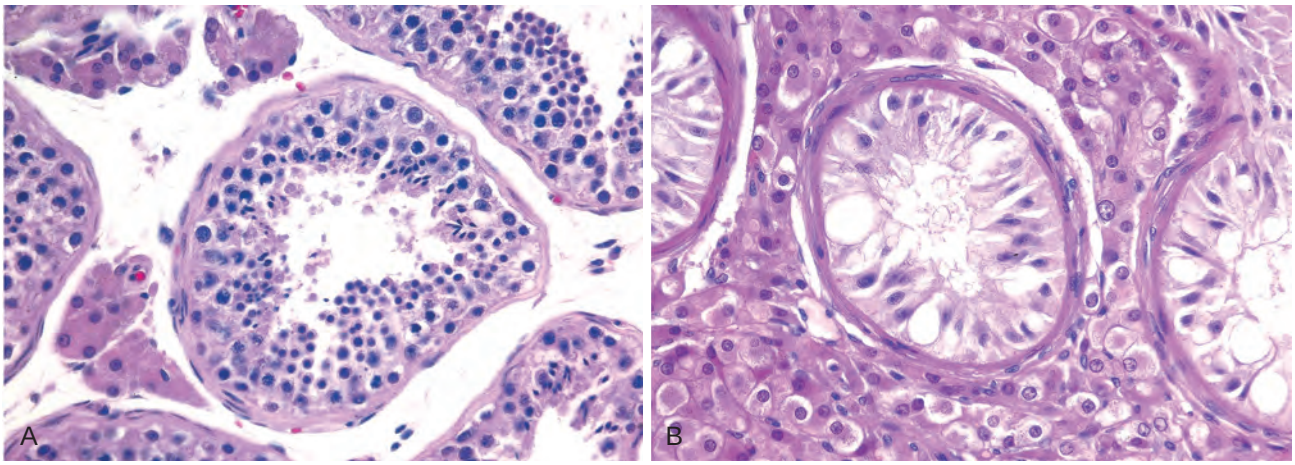


Figure 21.15 Cryptorchidism. (A) Normal testis shows tubules with active spermatogenesis. (B) Testicular atrophy in cryptorchidism. The tubules show Sertoli cells but no spermatogenesis. There is thickening of basement membranes and an apparent increase in interstitial Leydig cells.

Clinical Features

Cryptorchidism is asymptomatic and comes to attention when the scrotal sac is discovered to be empty by the patient, a parent, or a physician. In addition to sterility, cryptorchidism may be associated with other morbidity such as crushing injuries due to trauma to the inguinal region and inguinal hernia (10% to 20% of cases).

During the first year of life the majority of inguinal cryptorchid testes descend spontaneously into the scrotum. If it remains undescended, surgical correction by orchiopexy (placement in the scrotal sac) is required, preferably prior to development of histologic changes. Current recommendations are for orchiopexy to be performed at 6 to 12 months of age. Even with this corrective procedure, deficient spermatogenesis has been reported in 10% to 60% of patients, and repositioning has not been proven to completely eliminate the risk of cancer. Since the contralateral normally descended testis is also at higher risk for malignancy, it is believed that cryptorchidism and the associated risk for germ cell neoplasm is linked to an in utero defect in gonadal cell development (see discussion of testicular dysgenesis syndrome later), rather than the abnormal anatomic position.

KEY CONCEPTS

CRYPTORCHIDISM

- Cryptorchidism refers to incomplete descent of the testis from the abdomen to the scrotum and is present in about 1% of 1-year-old male infants.
- Bilateral or, in some cases, even unilateral cryptorchidism is associated with tubular atrophy and sterility.
- The cryptorchid testis carries a three- to fivefold higher risk for testicular cancer.
- Orchiopexy reduces but does not completely eliminate the risk of sterility and cancer.

Regressive Changes

Atrophy and Decreased Fertility

Testicular atrophy may be caused by one of several conditions, including (1) progressive atherosclerotic narrowing of the blood supply in old age; (2) end-stage inflammatory orchitis; (3) cryptorchidism; (4) hypopituitarism; (5) generalized malnutrition or cachexia; (6) irradiation; (7) prolonged administration of antiandrogens (e.g., for treatment of carcinoma of the prostate); (8) cirrhosis; (9) primary failure of genetic origin, such as in Klinefelter syndrome; and (10) exhaustion atrophy, which may follow persistent stimulation by high levels of follicle-stimulating pituitary hormone. The gross and microscopic alterations follow the pattern already described for cryptorchidism. In extreme cases there is complete regression ("vanishing testis").

Inflammation and Infections

Inflammatory disorders are distinctly more common in the epididymis. Several, including tuberculosis and gonorrhea, arise first in the epididymis and only involve the testis secondarily, while others, such as syphilis, involve the testis first.

Nonspecific Epididymitis and Orchitis

Epididymitis and possible subsequent orchitis are commonly related to infections in the urinary tract (cystitis, urethritis, prostatitis), which reach the epididymis and the testis through the vas deferens or the lymphatics of the spermatic cord. The cause of epididymitis varies with the age of the patient. Though uncommon in children, epididymitis in childhood is usually associated with a congenital genitourinary abnormality and infection with gram-negative rods. In sexually active men younger than age 35 years, the sexually transmitted pathogens *C. trachomatis* and *Neisseria gonorrhoeae* are most frequent. In men older than age 35, common urinary tract pathogens, such as *E. coli* and *Pseudomonas,* are responsible for most infections.

MORPHOLOGY

As elsewhere in the body, bacterial infection induces acute inflammation characterized by congestion, edema, and infiltration by neutrophils. The infection usually starts in the interstitial connective tissue, but then rapidly extends to involve the tubules, sometimes culminating in abscess formation or suppurative necrosis of the entire epididymis (Fig. 21.16). From here infection often extends

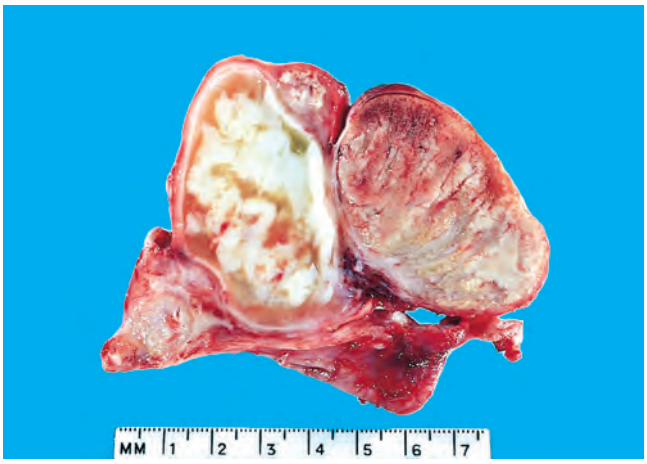


Figure 21.16 Acute epididymitis caused by gonococcal infection. The epididymis is replaced by an abscess. Normal testis is seen on the right.

to the testis, where it evokes a similar inflammatory reaction. Inflammation of the epididymis and testis is often followed by fibrous scarring, which may produce sterility. Usually the Leydig cells are not destroyed, and as a result androgen production by the testis is relatively unaffected.

Granulomatous (Autoimmune) Orchitis

Idiopathic granulomatous orchitis usually presents in middle age as a moderately tender testicular mass of sudden onset, sometimes associated with fever. It may sometimes appear insidiously, however, as a painless testicular mass mimicking a testicular tumor—hence its importance. Histologically, the orchitis is distinguished by granulomas restricted to spermatic tubules. The process must be distinguished from mycobacterial infection (see below), which often involves the epididymus and produces caseous necrosis. Although an autoimmune basis is suspected, the cause of these lesions remains unknown.

Infections

Gonorrhea

Extension of infection from the posterior urethra to the prostate, seminal vesicles, and then to the epididymis is a frequent complication of neglected gonococcal infection. In severe cases epididymal abscesses may develop, leading to extensive destruction and scarring. The infection may also spread to the testis and produce suppurative orchitis.

Mumps

Mumps is a systemic viral disease that most commonly affects school-aged children. Testicular involvement is extremely uncommon in this age group. In postpubertal males, however, orchitis occurs in 20% to 30% of cases. Most often, acute interstitial orchitis develops 1 week after the onset of swelling of the parotid glands.

Tuberculosis

When it involves the male genital tract, tuberculosis almost invariably begins in the epididymis, from where it may spread to the testis. The infection invokes the classic morphologic reactions of caseating granulomatous inflammation characteristic of tuberculosis elsewhere.

Syphilis

The testis and epididymis may be affected in both acquired and congenital syphilis, but almost invariably the testis is involved first. The morphologic pattern of the reaction takes two forms: (1) *obliterative endarteritis* associated with perivascular cuffs of lymphocytes and plasma cells, the histologic hallmark of syphilis, or (2) granulomatous inflammation, a lesion known as a *gumma*.

Other Infections

Leprosy, sarcoidosis, Crohn disease, malakoplakia, toxoplasmosis, fungi, parasites and brucellosis are other rare causes of orchitis.

Vascular Disorders

Torsion

Twisting of the spermatic cord typically cuts off the venous drainage of the testis. If untreated, it frequently leads to testicular infarction and thus represents one of the few true urologic emergencies. The thick-walled arteries remain patent, producing intense vascular engorgement followed by hemorrhagic infarction.

There are two settings in which testicular torsion occurs. Neonatal torsion occurs either in utero or shortly after birth. It lacks any associated anatomic defect to account for its occurrence. "Adult" torsion is typically seen in adolescence and presents with the sudden onset of testicular pain. It often occurs without any inciting injury. If the testis is manually untwisted within approximately 6 hours of the onset of torsion, the affected testis may be spared an orchiectomy. In adults, torsion results from a bilateral anatomic defect that leads to increased mobility of the testes *(bell-clapper abnormality)*. Contralateral *orchiopexy* is performed to prevent recurrence in the unaffected testis.

MORPHOLOGY

Depending on the duration of torsion, the morphologic changes range from intense congestion to widespread hemorrhage and testicular infarction (Fig. 21.17). In advanced stages, the swollen testis consists entirely of soft, necrotic, hemorrhagic tissue.

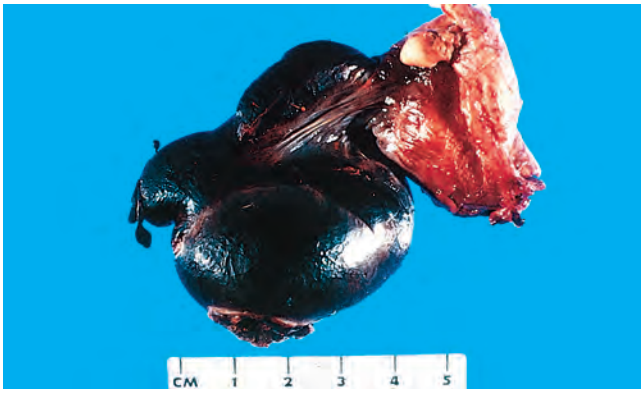


Figure 21.17 Torsion of testis. The dark discoloration is the result of hemorrhage and infarction.

Spermatic Cord and Paratesticular Tumors

Lipomas are common lesions involving the proximal spermatic cord, identified at the time of inguinal hernia repair. Although diagnosed as "lipomas," many of these lesions probably represent retroperitoneal adipose tissue that has been pulled into the inguinal canal along with the hernia sac, rather than a true neoplasm.

The most common benign paratesticular tumor is *adenomatoid tumor*. Although these lesions are mesothelial in nature, they are not referred to as mesotheliomas to distinguish them from other mesothelial lesions that may occur at this site. Adenomatoid tumors are usually small nodules, typically occurring near the upper pole of the epididymis. The importance of this lesion is that it is one of the few benign tumors that occur near the testis. If the pathologist can identify the nature of this lesion in intraoperative frozen sections, local excision of the adenomatoid tumor can spare the patient from an orchiectomy.

The most common malignant paratesticular tumors are rhabdomyosarcomas in children and liposarcomas in adults.

Testicular Tumors

Testicular neoplasms span an amazing gamut of histologic types. The 2016 WHO classification (Table 21.7) lists numerous categories, two of which are responsible for the overwhelming majority of lesions: germ cell tumors (GCTs) (95%) and sex cord–stromal tumors, both of which are discussed here.

Table 21.7 Pathologic Classification of Common Testicular Tumors

Germ Cell Tumors Derived From Germ Cell Neoplasia in Situ
Noninvasive germ cell neoplasia Germ cell neoplasia in situ
Tumors of a single histologic type (pure forms) Seminoma
Nonseminomatous germ cell tumors Embryonal carcinoma Yolk sac tumor, postpubertal type Choriocarcinoma Teratoma, postpubertal type Teratoma with somatic-type malignancy
Nonseminomatous germ cell tumors of more than one histologic type Mixed germ cell tumor
Germ Cell Tumors Unrelated to Germ Cell Neoplasia in Situ
Spermatocytic tumor
Teratoma, prepubertal type
Yolk sac tumor, prepubertal type
Mixed teratoma and yolk sac tumor, prepubertal type
Sex Cord–Stromal Tumors
Pure tumors
Leydig cell tumor
Sertoli cell tumor
Other
Tumor Containing Both Germ Cell and Sex Cord–Stromal Elements
Gonadoblastoma

From World Health Organization (WHO): *Histologic Classification of Testicular Tumors*, Geneva, 2016, WHO.

Germ Cell Tumors

GCTs predominantly affect Caucasian males between 15 and 45 years of age and are the most common cancer in this age group. GCTs are increasing in incidence and occur at variable rates across populations, both for unclear reasons. The lifetime risk is highest in Northern Europe and New Zealand and is lowest in Africa and Asia. In the United States in 2019, it was predicted that there would be 9560 new cases diagnosed and 410 deaths from GCTs.

Pathogenesis

Both environmental exposures and inherited and acquired genetic abnormalities contribute to the development of germ cell neoplasia. Below we touch on important environmental and genetic factors and how their interplay appears to lead to the stepwise development of GCTs (Fig. 21.18).

Environmental Factors. The role of environmental factors is inferred in part from population migration studies. For example, the incidence of testicular GCTs in Finland is about two times lower than in Sweden, but second-generation Finnish immigrants to Sweden have a tumor incidence that approaches that of the Swedish population. Testicular GCTs

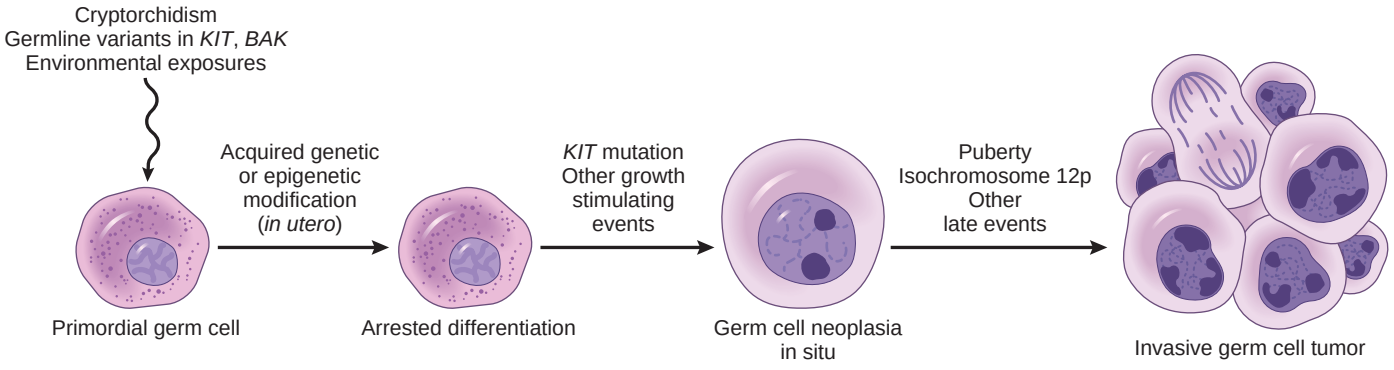


Figure 21.18 Salient environmental and genetic risk factors and acquired genetic and epigenetic alterations leading to germ cell tumor development and progression. See text for details.

also are associated with a spectrum of disorders collectively known as *testicular dysgenesis syndrome*. Components of this syndrome include cryptorchidism, hypospadias, and poor sperm quality. It has been proposed that these conditions are increased by in utero exposures to pesticides and nonsteroidal estrogens. The most important association is with cryptorchidism, which is seen with approximately 10% of testicular GCTs. Curiously, Klinefelter syndrome is associated with a greatly increased risk (50 times normal) for development of mediastinal GCTs, but these patients do not develop testicular tumors.

Genetic Factors. The relative risk of GCTs is four times higher than normal in fathers and sons of affected patients and is 8 to 10 times higher in brothers. Several genetic loci have been linked to familial GCT risk, including variants in genes encoding the ligand for the receptor tyrosine kinase KIT and BAK, which you will recall is an important inducer of apoptotic cell death (Chapter 1). Interestingly, these genes are also thought to play a role in gonadal development. Associations with variants in genes involved in sex hormone metabolism have also been found.

Steps in the Development of Germ Cell Tumors. A preponderance of evidence points to a primordial germ cell with an acquired defect in differentiation as the cell of origin of GCTs. The next step down the road to transformation consists of activation of growth factor receptor signaling, often by activating mutations in the KIT receptor tyrosine kinase, which stimulate proliferation and appear to be involved in the genesis of a precursor lesion termed *germ cell neoplasia in situ*. This precursor lesion is found in about 90% of testes involved by germ cell neoplasms and is associated with all types of GCTs except spermatocytic tumor and unusual types that arise in infancy. Germ cell neoplasia in situ also is frequently found in testes at high risk for developing GCTs, such as cryptorchid testes. Germ cell neoplasia in situ is believed to arise in utero and stay dormant until puberty, when hormonal influences may stimulate germ cell growth. The lesional cells retain the expression of the transcription factors OCT3/4 and NANOG, which are important in maintenance of pluripotent stem cells. Progression to full-blown GCTs is strongly associated with reduplication of the short arm of chromosome 12 (isochromosome 12p), a cytogenetic alteration that is invariably found in invasive GCTs regardless of histologic type. Progression is very likely, if not inevitable, as about 70% of individuals with germ cell neoplasia in situ (e.g., in cryptorchid testes) develop invasive GCTs within 7 years of diagnosis.

Classification. A simple classification of the most common types of testicular tumors is presented in Table 21.7. Two broad groups are recognized. *Seminomatous tumors* are composed of cells that resemble primordial germ cells or early gonocytes. *Nonseminomatous tumors* may be composed of undifferentiated cells that resemble embryonic stem cells, as in the case of embryonal carcinoma, but the malignant cells may also differentiate along other lineages, generating yolk sac tumors, choriocarcinomas, and teratomas. GCTs may be composed of a single cell type or mixtures of seminomatous and/or nonseminomatous components.

Seminoma

Seminoma is the most common type of GCT, making up about 50% of these tumors overall. The peak incidence is in the fourth decade. An identical tumor arises in the ovary, where it is called *dysgerminoma* (Chapter 22), and in the central nervous system, usually in midline structures such as the pineal gland, where it is referred to as *germinoma*.

MORPHOLOGY

Seminomas produce bulky masses, sometimes ten times the size of the normal testis. The typical seminoma has a homogeneous, gray-white, lobulated cut surface, usually devoid of hemorrhage or necrosis (Fig. 21.19). Generally the tunica albuginea is not penetrated, but occasionally extension to the epididymis, spermatic cord, or scrotal sac occurs.

The lesion is composed of sheets of uniform cells divided into poorly demarcated lobules by delicate fibrous septa containing a lymphocytic infiltrate (Fig. 21.20A); in some tumors, ill-defined granulomas also are present, presumably as part of a host response to the neoplasm. **The classic seminoma cell is round to polyhedral and has a distinct cell membrane; clear or watery-appearing cytoplasm; and a large, central nucleus with one or two prominent nucleoli** (Fig. 21.20B). The cytoplasm contains varying amounts of glycogen. By immunohistochemistry, the tumor cells are typically positive for KIT, OCT3/4, and podoplanin and negative for cytokeratin. Approximately 15% of seminomas contain syncytiotrophoblasts. In this subset of patients, serum human chorionic gonadotropin (hCG) levels are elevated, though not to the extent seen in patients with choriocarcinoma.

Spermatocytic Tumor

Spermatocytic tumors are uncommon, representing 1% to 2% of all testicular germ cell neoplasms. In contrast to other GCTs, affected individuals are generally older (usual more than 65 years old). This slow-growing tumor does not metastasize, and when treated by surgical resection it has an excellent prognosis. Although it bears some morphologic resemblance to seminoma, the origin and pathogenesis of spermatocytic tumor are quite distinct, as the tumor is not

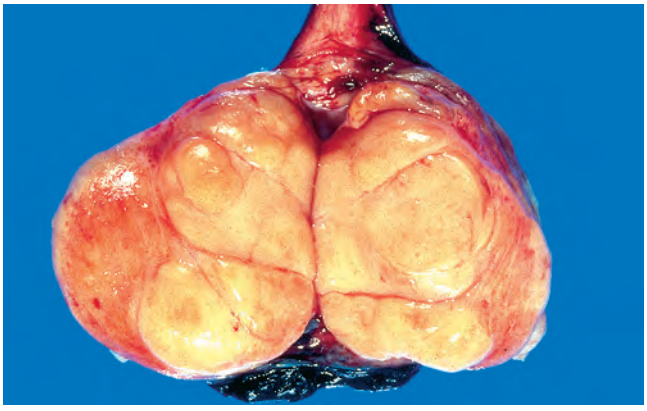


Figure 21.19 Seminoma of the testis, appearing as a well-circumscribed, pale, fleshy, homogeneous mass on cut surface.

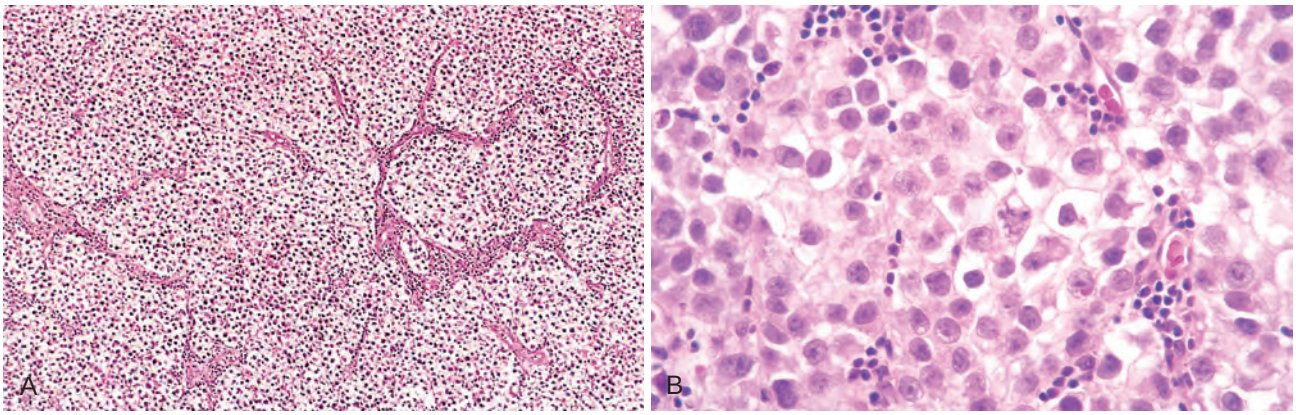


Figure 21.20 Seminoma. (A) Low magnification shows sheets of pale tumor cells divided into poorly demarcated lobules by delicate septa containing reactive lymphocytes. (B) High magnification reveals large cells with distinct cell borders, pale nuclei, prominent nucleoli, and abundant cytoplasm.

associated with germ cell neoplasia in situ, lacks isochromosome 12p, and instead is characteristically associated with gain of chromosome 9q.

MORPHOLOGY

Grossly, spermatocytic tumor tends to be circumscribed and fleshy and is occasionally myxoid, sometimes with cystic regions. The tumor characteristically contains three intermixed cell populations: (1) medium-sized cells containing a round nucleus, spireme-type chromatin (similar to that seen in the meiotic phase of normal spermatocytes) with occasional nucleoli, and eosinophilic cytoplasm; (2) smaller cells with dense chromatin and with a narrow rim of eosinophilic cytoplasm resembling secondary spermatocytes; and (3) scattered giant cells, either uninucleate or multinucleate. In contrast to seminoma, spermatocytic tumor uniformly lacks inflammatory infiltrates and syncytiotrophoblasts, does not occur at extratesticular sites, and is never admixed with other GCTs.

Embryonal Carcinoma

The peak incidence for embryonal carcinoma is in the 20- to 30-year-old age group, about a decade earlier than that of seminoma. Embryonal carcinoma is more aggressive than seminoma and may occur as a pure tumor or mixed with other germ cell components.

MORPHOLOGY

Embryonal carcinomas tend to be locally aggressive, frequently extending through the tunica albuginea into the epididymis or spermatic cord. On cut surface, the tumor often has a variegated appearance due to the presence of foci of hemorrhage or necrosis (Fig. 21.21). Histologically, the cells grow in alveolar or tubular patterns, sometimes with papillary folds. More undifferentiated lesions may display sheets of cells with cleft-like spaces (Fig. 21.22). The neoplastic cells have an epithelial appearance, are large and anaplastic, and have hyperchromatic nuclei with prominent nucleoli. The cell borders are usually indistinct, and there is considerable variation in cell and nuclear size and shape (pleomorphism). Mitotic figures and tumor giant cells are frequently seen. Vascular-lymphatic invasion is common. Like seminoma, embryonal carcinoma stains positively for OCT3/4, but differs in that it is typically also positive for cytokeratin and negative for KIT and podoplanin.

Yolk Sac Tumor

Also known as *endodermal sinus tumor,* prepubertal yolk sac tumors are the most common testicular tumor in infants and children up to 3 years of age. In this age group it has a very good prognosis. By contrast, postpubertal yolk sac tumor is rarely "pure" and more frequently occurs in combination with embryonal carcinoma or other germ cell components.

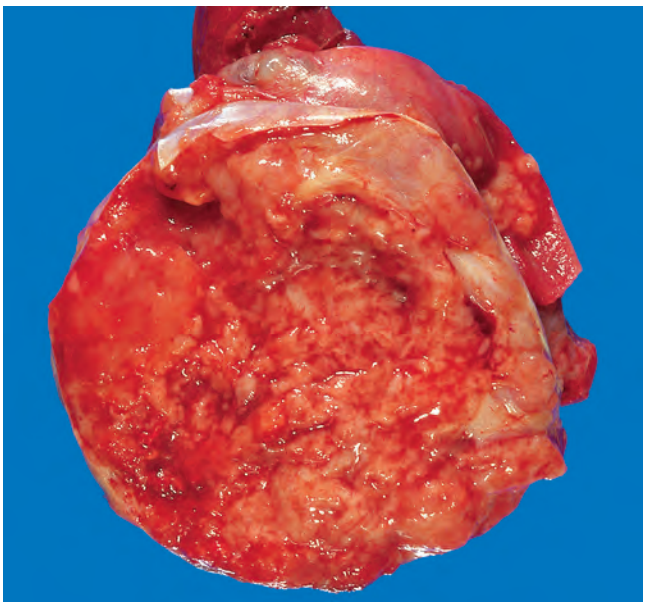


Figure 21.21 Embryonal carcinoma. In contrast to the seminoma illustrated in Fig. 21.19, as shown here, embryonal carcinoma often produces a hemorrhagic mass.

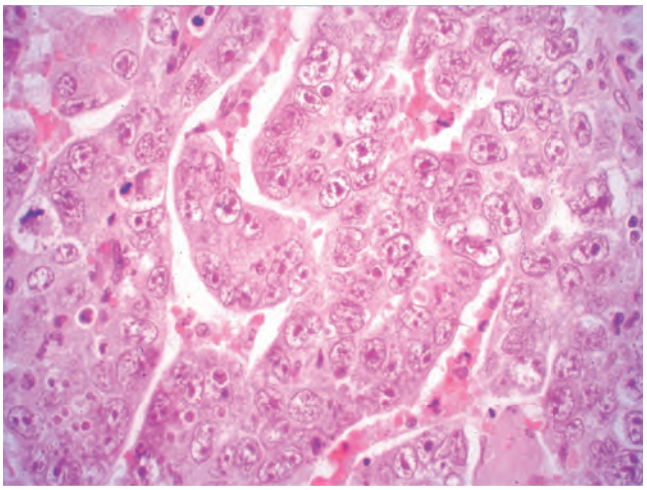


Figure 21.22 Embryonal carcinoma composed of cells with large, hyperchromatic nuclei arranged in sheets and poorly formed glands.

MORPHOLOGY

Pre- and postpubertal yolk sac tumors have similar histology; the chief difference is that prepubertal tumors are not known to be associated with germ cell neoplasia in situ and lack other germ cell elements. Yolk sac tumors are typically composed of a lace-like (reticular) network of medium-sized cuboidal, flattened, or spindled cells, which generally show less cytologic atypia than embryonal carcinoma. In addition, papillary structures, solid cords of cells, and a multitude of other, less common growth patterns may be found. In approximately 50% of tumors, structures resembling endodermal sinuses **(Schiller-Duval bodies)** may be seen; these consist of a mesodermal core with a central capillary and a visceral and parietal layer of cells resembling primitive glomeruli. Eosinophilic, hyaline-like globules containing α-fetoprotein (AFP) and α_1-antitrypsin are typically seen in yolk sac tumor cells. The expression of AFP is characteristic of yolk sac tumor cells and underscores their resemblance to yolk sac; stains for cytokeratin are also positive.

Choriocarcinoma

Choriocarcinoma is a highly malignant type of GCT. In its pure form, choriocarcinoma is rare, constituting less than 1% of all GCTs. Serum hCG is invariably elevated, often markedly so. Widespread metastasis, often associated with hemorrhage at sites of involvement, may be seen.

MORPHOLOGY

Typically, choriocarcinoma primary tumors are small, rarely larger than 5 cm in diameter. Hemorrhage and necrosis are extremely common. The tumor is composed of two intimately juxtaposed cell types, **syncytiotrophoblasts** and **cytotrophoblasts** (Fig. 21.23). Syncytiotrophoblasts are large multinucleated cells with abundant eosinophilic vacuolated cytoplasm containing hCG, which is readily detected by immunohistochemistry. Cytotrophoblasts are more regular and tend to be polygonal, with distinct borders and clear cytoplasm; they grow in cords or sheets and have a single, fairly uniform nucleus. This neoplasm can also arise in the female genital tract (Chapter 22).

Teratoma

The designation *teratoma* refers to GCTs having various cellular or organoid components reminiscent of the normal derivatives of more than one germ layer. They may occur at any age from infancy to adulthood. Pure teratoma is fairly common in infants and children and is referred to as the "prepubertal type." Among GCTs, teratomas are second in frequency in infants and children only to yolk sac tumors. In adults, pure teratomas are rare, constituting 2% to 3% of GCTs, and most often teratomas are found mixed with other histologies (a phenomenon that is discussed later). Teratomas that occur early in life also differ biologically and clinically from those that occur in adults. Prepubertal teratomas are not associated with germ cell neoplasia in situ or isochromosome 12p and pursue a benign course. Only a minor fraction of teratomas occurring in adults share these features, and postpubertal adult teratomas are generally taken to be malignant.

MORPHOLOGY

Teratomas presenting in adult males are usually large, ranging from 5 to 10 cm in diameter. The presence of a variety of tissues imparts a heterogeneous appearance, with solid, sometimes cartilaginous, and cystic areas (Fig. 21.24). Microscopically, collections of differentiated cells or organoid structures, such as neural tissue, muscle bundles, islands of cartilage, squamous epithelium lining epidermal-like surfaces with or without skin adnexal structures, structures reminiscent of thyroid gland, bronchial epithelium, and bits of intestinal wall or brain substance, are seen embedded in a fibrous or myxoid stroma (Fig. 21.25). Elements may be mature (resembling various adult tissues) or immature (sharing histologic features with fetal or embryonic tissue).

Rarely, malignant non–germ cell (somatic) tumors arise in postpubertal teratomas, a phenomenon referred to as *teratoma with somatic-type malignant transformation*. Transformation may take the form of a squamous cell carcinoma, mucin-secreting

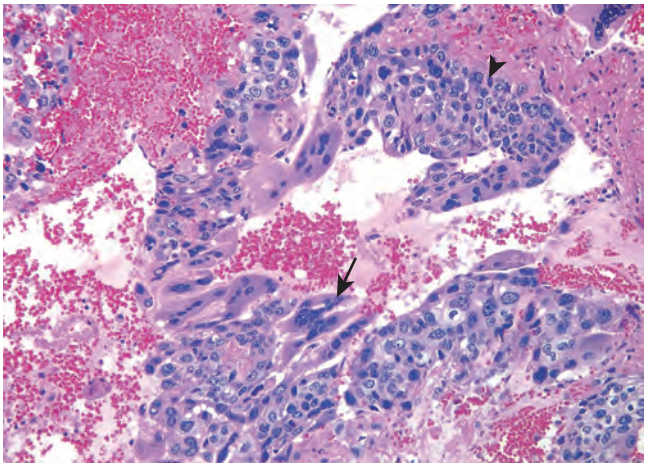


Figure 21.23 Choriocarcinoma. Both cytotrophoblastic cells *(arrowhead)* with central nuclei and syncytiotrophoblastic cells *(arrow)* with multiple dark nuclei embedded in eosinophilic cytoplasm are present. Hemorrhage and necrosis are seen in the upper right field.

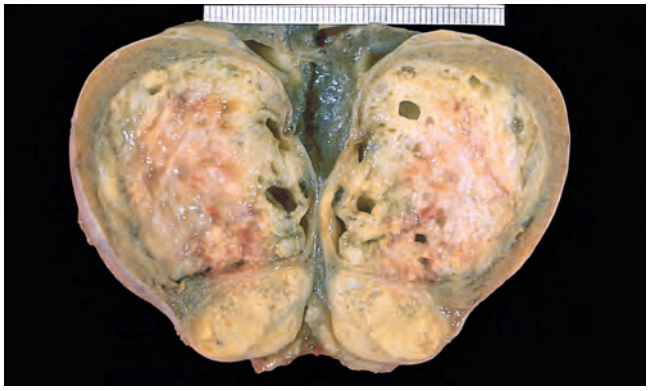


Figure 21.24 Teratoma of testis. The variegated cut surface with cysts reflects the presence of multiple tissue types.

adenocarcinoma, sarcoma, or other cancers. The importance of recognizing a somatic-type malignancy arising in a teratoma is that these secondary tumors are chemoresistant; thus, the only hope for cure resides in resection of the tumor. These non–germ cell malignancies retain isochromosome 12p, proving a clonal relationship to the preceding teratoma.

In contrast to their postpubertal counterparts, prepubertal teratomas produce distinct elements that are arranged into structures that more closely resemble normal tissue. Special types include dermoid cysts, which typically contain hair, teeth, and skin. On rare occasions, prepubertal teratoma may be admixed with yolk sac tumor.

Mixed Tumors

About 60% of GCTs are composed of more than one of the above types. Common mixtures include teratoma, embryonal carcinoma, and yolk sac tumor; seminoma with embryonal carcinoma; and embryonal carcinoma with teratoma.

Clinical Features of Testicular Germ Cell Tumors. Although painless enlargement of the testis is a characteristic feature of GCTs, any solid testicular mass should be considered neoplastic until proven otherwise. Biopsy of a testicular neoplasm is associated with a risk of tumor spillage, which would necessitate excision of the scrotal skin in addition to orchiectomy. Consequently, the standard management of a solid testicular mass is radical orchiectomy based on the presumption of malignancy. Occasionally, when a benign lesion is suspected and the lesion can be excised by partial orchiectomy, an intraoperative frozen section examination may be performed in an effort to avoid an unnecessary radical orchiectomy.

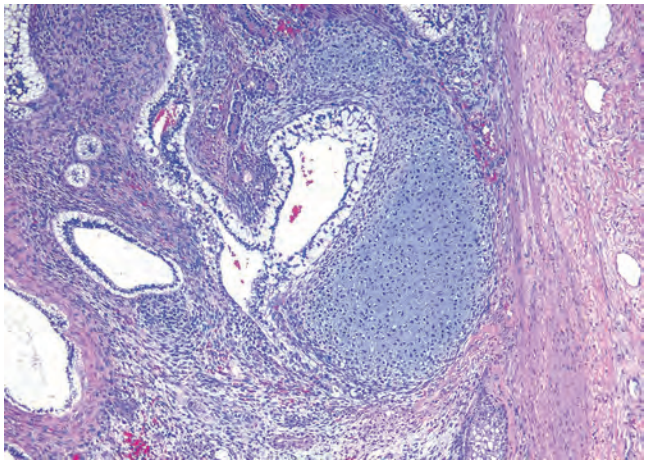


Figure 21.25 Teratoma of the testis, consisting of a disorganized collection of glands, cartilage, smooth muscle, and immature stroma.

Testicular tumors have a characteristic mode of spread. *Lymphatic spread* is common to all forms. In general, retroperitoneal para-aortic nodes are the first to be involved. Subsequent spread may occur to mediastinal and supraclavicular nodes. *Hematogenous spread* is primarily to the lungs, but liver, brain, and bones may also be involved. Metastases from seminomas typically involve lymph nodes. Hematogenous spread occurs later in the course of dissemination. Nonseminomatous GCTs tend to metastasize earlier and more frequently via a hematogenous route. The histology of metastases and distant recurrences may differ from that of the testicular lesion. This is a reflection of the fact that GCTs are derived from pluripotent germ cells. Furthermore, a minor component in the primary tumor that is unresponsive to chemotherapy may survive and subsequently become the dominant histologic type in the metastatic site.

Serum Biomarkers in Germ Cell Tumors. GCTs secrete polypeptide hormones and enzymes that can be detected in the patient's peripheral blood. These include hCG, AFP, and lactate dehydrogenase, all of which can serve as valuable biomarkers in the diagnosis and management of testicular cancer. The elevation of lactate dehydrogenase correlates with the mass of tumor cells and provides a tool to assess tumor burden. Marked elevation of serum AFP or hCG levels is produced by yolk sac tumor and syncytiotrophoblastic elements, respectively. Both of these markers are elevated in more than 80% of individuals with nonseminomatous GCTs at the time of diagnosis. As stated earlier, approximately 15% of seminomas have syncytiotrophoblastic giant cells and minimal elevation of hCG levels. In the context of testicular tumors, serum markers are valuable in several different ways:

- In the initial evaluation of testicular masses
- In the staging of testicular GCTs; for example, after orchiectomy, persistent elevation of hCG or AFP concentrations indicates metastatic spread
- In assessing tumor burden
- In monitoring the response to therapy

The therapy and prognosis of testicular tumors depend largely on the clinical and pathologic stage and the histologic type. Seminoma, which is radiosensitive and chemosensitive, has the best prognosis. More than 95% of patients with stage I and II seminoma are cured by orchiectomy with or without chemotherapy or radiotherapy. In nonseminomatous GCT patients, approximately 90% achieve complete remission with aggressive chemotherapy, and most can be cured. Pure embryonal carcinoma behaves more aggressively than mixed GCTs. Pure choriocarcinoma and mixed GCT with predominantly choriocarcinoma have a poor prognosis.

Tumors of Sex Cord–Gonadal Stroma

As indicated in Table 21.7, sex cord–gonadal stromal tumors are subclassified based on their presumed histogenesis and differentiation. The two most important members of this group—Leydig cell tumors and Sertoli cell tumors—are described here. Pathologic features correlated with malignancy in sex cord–stromal tumors of the testis include large size (greater than 5 cm), necrosis, infiltrative borders, anaplasia, mitotic activity, vascular-lymphatic invasion, and extratesticular extension.

Leydig Cell Tumors

Tumors of Leydig cells often elaborate androgens and in some cases estrogens and corticosteroids as well. They may arise in children or adults. As with other testicular tumors, the most common presenting feature is testicular swelling, but in some patients gynecomastia brings them to clinical attention. In children, hormonal effects, manifested primarily as sexual precocity, may be the dominant feature. Leydig cell tumor is associated with Klinefelter syndrome, cryptorchidism, and hereditary leiomyomatosis and renal cell carcinoma syndrome (caused by germline mutations in the metabolic enzyme fumarate hydratase). Approximately 10% of Leydig cell tumors in adults are malignant and can produce metastases.

MORPHOLOGY

Leydig cell tumors form circumscribed nodules, usually less than 5 cm in diameter, that have a distinctive golden brown, homogeneous appearance on cut surfaces. Histologically, the tumor cells resemble their normal counterparts. They are large in size and have round or polygonal cell outlines, abundant granular eosinophilic cytoplasm, and a round central nucleus. The cytoplasm frequently contains lipid droplets, vacuoles, or lipofuscin pigment and, most characteristically, rod-shaped **crystalloids of Reinke,** which are seen in about 25% of the tumors.

Sertoli Cell Tumors

Most Sertoli cell tumors are hormonally silent and present as a testicular mass. They are associated with Carney complex (caused by germline mutations in the gene *PRKAR1A,* which encodes a cyclic adenosine monophosphate–dependent protein kinase), Peutz-Jeghers syndrome (Chapter 17), and familial adenomatosis polyposis syndrome (Chapter 17). These neoplasms appear as firm, small nodules with a homogeneous gray-white to yellow cut surface. Histologically, the tumor cells are arranged in distinctive trabeculae that tend to form cord-like structures and tubules. Most Sertoli cell tumors are benign, but approximately 10% pursue a malignant course.

Gonadoblastoma

Gonadoblastomas are rare neoplasms comprised of a mixture of germ cells and gonadal stromal elements that almost always arise in gonads with some form of testicular dysgenesis (discussed earlier). In some cases, the germ cell component becomes malignant, giving rise to seminoma.

Testicular Lymphoma

Primary testicular lymphomas are rare, accounting for 5% of testicular neoplasms, but are the most common form of testicular neoplasm in men older than 60 years of age. In most cases, these are aggressive tumors that are disseminated at the time of detection. In contrast to GCTs, tumors are frequently bilateral and involve the spermatic cord. The most common testicular lymphomas, in decreasing order of frequency, are diffuse large B-cell lymphoma, Burkitt lymphoma, and Epstein-Barr virus–positive extranodal NK/T-cell lymphoma (Chapter 13). Testicular lymphomas have a high propensity for central nervous system involvement, which is a frequent site of recurrence.

KEY CONCEPTS

TESTICULAR TUMORS

- GCTs are by far the most common tumor types and account for 95% of testicular neoplasms.
- Cryptorchidism, infertility, and prior history of GCT in contralateral testis are risk factors.
- Germ cell neoplasia in situ is a precursor lesion associated with most GCTs. Seventy percent of patients with documented germ cell neoplasia in situ will develop invasive GCTs.
- Seminoma, embryonal carcinoma, yolk sac tumors, choriocarcinoma, and teratoma are the most common types of GCTs. Mixed tumors that contain more than one histologic type account for 40% of GCTs.
- Clinically, testicular GCTs are divided into two groups: seminomas and nonseminomatous tumors. Seminomas spread mainly to para-aortic nodes and are radiosensitive. Nonseminomatous tumors tend to spread earlier via lymphatic and hematogenous routes.
- hCG, AFP, and lactate dehydrogenase can serve as valuable blood biomarkers in the diagnosis and management of GCTs.
- Non-GCTs include sex cord–gonadal stromal tumors and non-Hodgkin lymphoma, which is the most common testicular tumor in men older than 60 years.

Lesions of Tunica Vaginalis

The tunica vaginalis is a mesothelial-lined surface exterior to the testis that may accumulate serous fluid *(hydrocele)* causing considerable enlargement of the scrotal sac. By transillumination it is usually possible to define the clear, translucent character of the contained fluid. Rarely, malignant mesotheliomas and ovarian epithelial-type tumors can arise from the tunica vaginalis.

Hematocele is a collection of blood in the tunica vaginalis. It is an uncommon condition usually encountered following testicular trauma or torsion, or in individuals with systemic bleeding disorders. *Chylocele* refers to the accumulation of lymph in the tunica and is almost always found in patients with elephantiasis who have widespread, severe lymphatic obstruction caused, for example, by filariasis (Chapter 8). *Spermatocele* refers to a small cystic accumulation of semen in dilated efferent ducts or ducts of the rete testis. *Varicocele* is a dilated vein in the spermatic cord. Varicoceles may be asymptomatic but have also been implicated in some men as a contributing factor to infertility. They can be corrected by surgical repair.

PROSTATE

The prostate is a retroperitoneal organ encircling the neck of the bladder and urethra that lacks a distinct capsule. In a normal adult, the prostate weighs approximately 20 g and can be divided into four biologically and anatomically distinct regions, the peripheral, central, transition, and periurethral zones (Fig. 21.26). These zones are at risk for different types of proliferative lesions. For example, most hyperplasias arise in the transition zone, whereas most carcinomas originate in the peripheral zone.

Histologically, the prostate consists of glands separated by abundant fibromuscular stroma. The glands are lined by two layers of cells: a basal layer of low cuboidal basal epithelium covered by a layer of columnar secretory cells (Fig. 21.27), which often contain small papillary infoldings. Testicular androgens control the growth and survival of prostatic cells, and castration leads to widespread apoptosis of prostatic epithelium and atrophy of the prostate.

Only three pathologic processes affect the prostate gland with sufficient frequency to merit discussion: inflammation, benign prostatic hypertrophy (BPH), and tumors. Of these, BPH is the most common and occurs so often in older males that it can almost be viewed as a "normal" part of aging. Prostatic carcinoma is also extremely common in older men and is an important cause of morbidity and mortality. We begin our discussion with consideration of inflammatory processes.

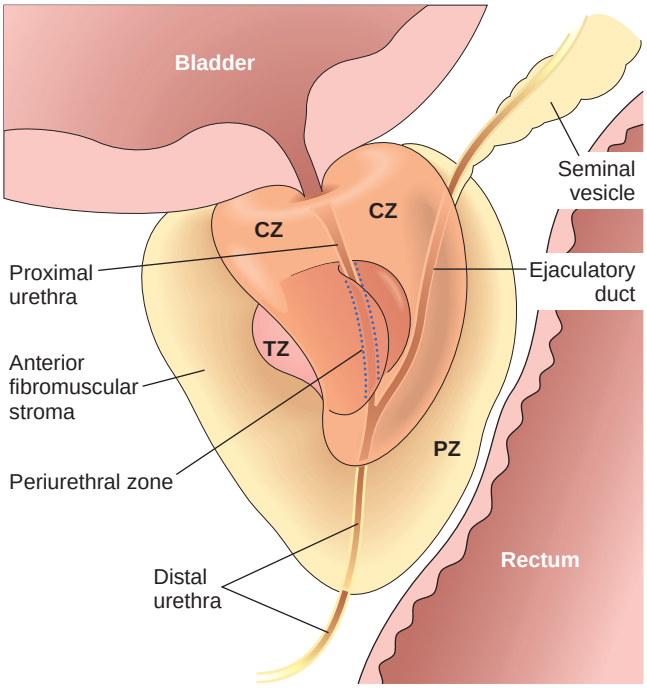


Figure 21.26 Adult prostate. The normal prostate contains several distinct regions, including a central zone *(CZ)*, a peripheral zone *(PZ)*, a transitional zone *(TZ)*, and a periurethral zone. Most carcinomas arise from the peripheral zone and may be palpable during digital examination of the rectum. Benign prostatic hyperplasia, in contrast, arises from the more centrally situated transitional zone and often produces urinary obstruction.

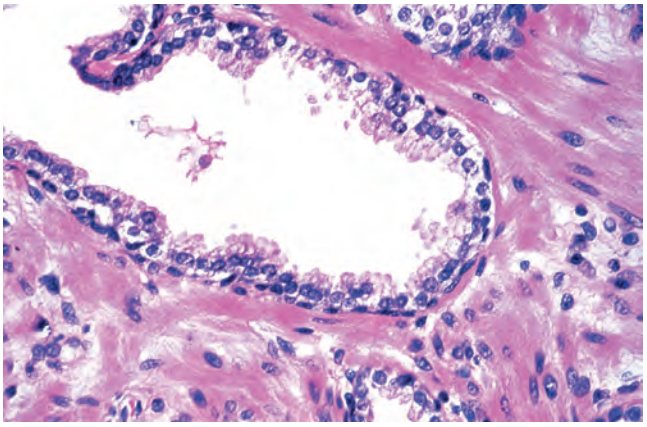


Figure 21.27 Normal prostate gland, with basal cell and secretory cell layers.

Inflammation

Prostatitis may be divided into several categories, depending on cause, patterns of tissue reaction, and clinical course.

- *Acute bacterial prostatitis* typically results from bacteria similar to those that cause urinary tract infections. Thus, most cases are caused by various strains of *E. coli,* other gram-negative rods, enterococci, and staphylococci. The organisms usually reach the prostate through the reflux of contaminated urine from the posterior urethra or the urinary bladder, but occasionally distant foci of infection seed the prostate through lymphohematogenous routes. Prostatitis sometimes follows surgical manipulation of the urethra or prostate gland itself (e.g., catheterization, cystoscopy, or urethral dilation). Clinically, acute bacterial prostatitis is associated with fever, chills, and dysuria. On rectal examination the prostate is exquisitely tender and boggy. The diagnosis can be established by urine culture and clinical features. Biopsy of a gland in which acute prostatitis is suspected is contraindicated, as this may lead to sepsis.
- *Chronic bacterial prostatitis* usually presents with low back pain, dysuria, and perineal and suprapubic discomfort, though it also may be asymptomatic. Patients often have a history of recurrent urinary tract infections (cystitis, urethritis) caused by the same organism. Because most antibiotics penetrate the prostate poorly, bacteria find haven in the parenchyma and constantly seed the urinary tract. Diagnosis of chronic bacterial prostatitis depends on the demonstration of leukocytosis in expressed prostatic secretions and positive bacterial cultures. The implicated organisms are the same as those that cause acute prostatitis.
- *Chronic abacterial prostatitis,* the most common form of prostatitis, is often referred to clinically as *chronic pelvic pain syndrome.* It is indistinguishable from chronic bacterial prostatitis in terms of signs and symptoms, but there is no history of recurrent urinary tract infection. Expressed prostatic secretions contain more than 10 leukocytes per high-power field, but bacterial cultures are uniformly negative.
- *Granulomatous prostatitis* may be caused by a specific infectious agent or may refer to a pattern of tissue reaction to noninfectious stimuli. In the United States, the most common cause is instillation of BCG for treatment of

bladder cancer. In this setting, the finding of granulomas in the prostate is of no clinical significance and requires no treatment. Fungal granulomatous prostatitis is typically seen in immunocompromised hosts. Nonspecific granulomatous prostatitis is relatively common and represents a reaction to secretions from ruptured prostatic ducts and acini.

- *Other forms of prostatitis.* Adenoviral and IGg4-associated (Chapter 6) autoimmune prostatitis have also been described.

Benign Enlargement

Benign Prostatic Hyperplasia

BPH (also referred to as nodular hyperplasia) is the most common benign prostatic disease in men older than age 50 years. Approximately 30% of white American men in that age group have moderate to severe symptoms of BPH, and histologic evidence of BPH is found in up to 90% of men by age 80. It is not a premalignant lesion.

Etiology and Pathogenesis

Dihydrotestosterone (DHT) is the main androgen in the prostate, where it is formed from testosterone through the action of type 2 5α-reductase (Fig. 21.28A). This enzyme is expressed primarily in stromal cells and is not expressed in prostatic epithelial cells. Type 1 5α-reductase is another enzyme that mediates DHT production from testosterone in extraprostatic locations (e.g., liver and skin) and provides an additional source of DHT that reaches the prostate through the blood.

DHT binds to and activates androgen receptors (ARs) found in both stromal and epithelial prostate cells. It is more potent than testosterone because it has a higher affinity for ARs and forms a more stable complex with them. Binding of DHT stimulates ARs to translocate from the cytoplasm to the nucleus and activate the transcription of androgen-dependent genes, which encode several growth factors and their receptors. Most important among the upregulated factors are members of the fibroblast growth factor (FGF) family and transforming growth factor (TGF) β (Chapter 3). FGFs, produced by stromal cells, are paracrine regulators of androgen-stimulated epithelial growth during embryonic prostatic development, and some of these pathways may be "reawakened" in adulthood to produce prostatic growth in BPH. TGFβ serves as a mitogen for fibroblasts and other mesenchymal cells but inhibits epithelial proliferation. Although the ultimate cause of BPH is unknown, it is believed that DHT-induced growth factors act by increasing the proliferation of stromal cells and decreasing the death of epithelial cells.

While it is recognized that androgens play a permissive role in BPH pathogenesis, multiple lines of evidence support a role for estrogens as well. Two different forms of estrogen receptor (ER), ERα and ERβ, have opposing proliferative and antiproliferative effects on prostate cells, respectively. Effects of estrogens on the prostate are associated with multiple mechanisms including apoptosis, aromatase expression, and paracrine regulation via prostaglandin E_2. Estrogens thus contribute to BPH pathogenesis by tipping the balance toward proliferation (Fig. 21.28B).

MORPHOLOGY

In BPH, the weight of the enlarged prostate often increases three- to fivefold (60 to 100 g), and even greater enlargement

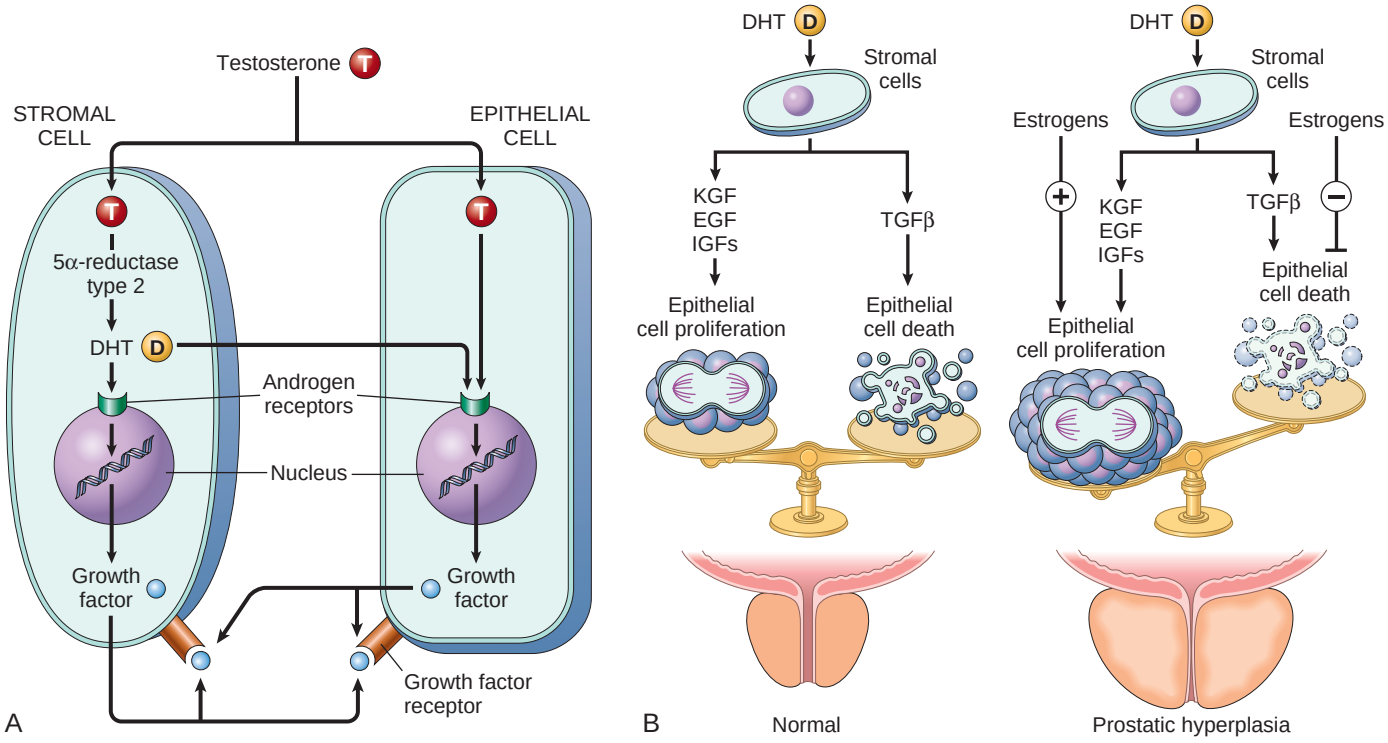


Figure 21.28 Simplified scheme of the pathogenesis of prostatic hyperplasia. (A) The central role of the stromal cells in generating dihydrotestosterone *(DHT)* is depicted. DHT may also be produced in skin and liver by both type 1 and type 2 5α-reductase. (B) The contribution of estrogen in tipping the balance of cell proliferation and cell death toward the former is illustrated. *EGF,* Epidermal growth factor; *IGFs,* insulin-like growth factors; *KGF,* keratinocyte growth factor; *TGFβ,* transforming growth factor β.

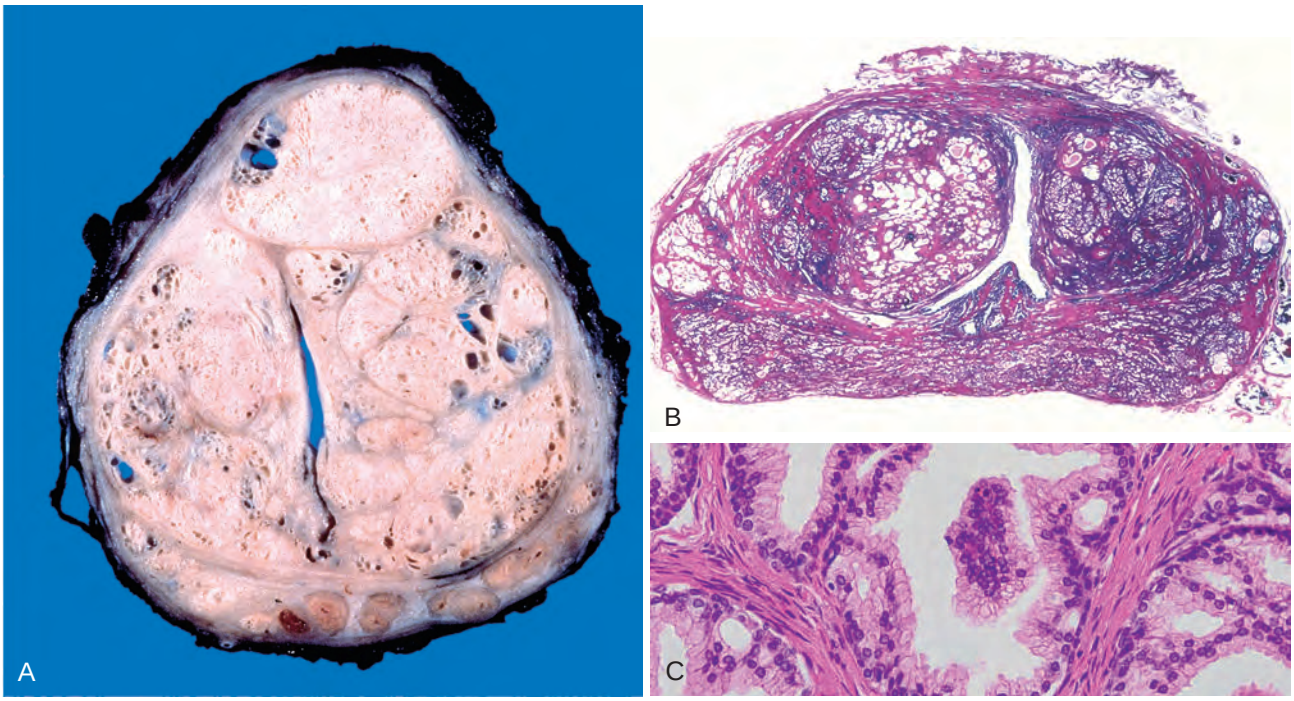


Figure 21.29 Benign prostatic hyperplasia (BPH). (A) Well-defined nodules of BPH compress the urethra into a slit-like lumen. (B) Microscopic view of a whole mount of the prostate shows nodules of hyperplastic glands on both sides of the urethra. (C) Under high power a nodule composed of hyperplastic glands with papillary epithelial infoldings is seen.

may be seen. BPH affects the transition zone and thus may encroach on the urethra, compressing it to a slit-like orifice (Fig. 21.29A). On cross-section, hyperplastic nodules are seen that vary in color and consistency depending on their cellular content (Fig. 21.29B). Nodules that contain mostly glands are yellow-pink and soft and exude a milky white prostatic fluid. Nodules composed primarily of fibromuscular stroma are pale gray and firm.

Microscopically, individual nodules contain small to large to cystically dilated glands that are separated by bland spindle-shaped stromal cells. The glands are lined by two layers of cells, an inner columnar secretory cell layer and an outer layer of cuboidal or flattened basal epithelium (Fig. 21.29C), and infolding of the glands may produce a papillary architecture. In markedly enlarged glands, compromise of the vascular supply may produce prostatic infarcts, which may have adjacent areas of squamous metaplasia.

Clinical Features

The main symptoms of BPH are due to urinary obstruction caused by prostatic enlargement and stromal smooth muscle–mediated contraction. The increased resistance to urinary outflow leads to bladder hypertrophy and distention, accompanied by urine retention. The inability to empty the bladder completely creates a reservoir of residual urine that is a common source of infection. Patients experience increased urinary frequency, nocturia, difficulty in starting and stopping the stream of urine, overflow dribbling, and dysuria (painful micturition) and have an increased risk of developing bacterial infections of the bladder and kidney. In many cases, sudden, acute urinary retention occurs that requires emergency catheterization for relief.

Symptomatic BPH is usually managed medically with α-adrenergic blockers and 5α-reductase inhibitors. The former decrease prostate smooth muscle tone via inhibition of α_1-adrenergic receptors, while the latter physically shrink the prostate by decreasing DHT synthesis. For moderate to severe cases recalcitrant to medical therapy, a wide range of invasive procedures exist. Transurethral resection of the prostate (TURP) was for long the gold standard, but alternative procedures to destroy excessive prostatic tissue have been developed with lower morbidity and lower costs. These procedures include high-intensity focused ultrasound (HIFU), laser therapy, hyperthermia, transurethral electrovaporization, and radiofrequency ablation.

KEY CONCEPTS

BENIGN PROSTATIC HYPERPLASIA

- BPH is characterized by proliferation of benign stromal and glandular elements. DHT, an androgen derived from testosterone, is the major hormonal stimulus for proliferation.
- BPH most commonly affects the inner periurethral zone and transition zone of the prostate, producing nodules that compress the prostatic urethra.
- Clinical symptoms and signs are related to urinary obstruction that also predisposes to recurrent urinary tract infections.
- Medical management is based on α-adrenergic blockers and 5α-reductase inhibitors, which decrease prostatic smooth muscle tone and inhibit DHT production, respectively.

Neoplasms

Adenocarcinoma

In the United States, adenocarcinoma of the prostate is the most common form of cancer in men, with an expected 174,650 new cases in 2019, accounting for 20% of all male cancers. Prostate cancer is the second cause of cancer-related

death in men, surpassed only by lung cancer. The biologic behavior of prostate cancer spans the gamut from clinically insignificant "histologic" cancers to aggressive tumors that are rapidly fatal. It is largely a disease of aging. Based on autopsy studies, the incidence of prostate cancer increases from 20% in men in their 50s to approximately 70% in men between the ages of 70 and 80 years.

Epidemiology and Pathogenesis. An interplay of both environmental exposures and inherited genetic factors contributes to striking differences in the incidence of prostate cancer across geographic locales (Chapter 7). For example, the incidence of prostate cancer in individuals of Japanese descent living in the United States is substantially higher than in Japanese living in Japan, but also is only about 50% of the reported incidence in African Americans. As explained below, some genetic factors appear to increase risk due to intrinsic effects on prostatic epithelium, while others may act by modifying the risk associated with environmental exposures. Furthermore, as might be expected of an androgen-sensitive organ, androgens and AR function have central roles in the development, progression, and treatment of prostate cancer.

Environmental Factors. It is hypothesized that exposure to carcinogens, estrogens, and oxidants damage prostatic epithelium, setting the stage for acquisition of genetic and epigenetic changes that lead to cancer development (Fig. 21.30). Broadly speaking, the "Western diet" is a suspect, given the relatively high incidence of prostate cancer in the United States, South America, Western Europe, and Australia. One dietary component that epidemiologic and animal model studies suggest contributes to risk is consumption of charred red meats and animal fats, which leads to formation of carcinogenic heterocyclic aromatic amines and polycyclic aromatic hydrocarbons. These and other dietary exposures may create oxidant stress in prostatic epithelium, leading to cell injury and inflammation. One piece of evidence linking these exposures to prostate cancer development is the observation that polymorphisms in glutathione-*S*-transferase *(GSTP1),* an enzyme involved in detoxification of polycyclic aromatic hydrocarbons, are linked to prostate cancer risk.

Inherited Genetic Factors. Studies of twins and families support the existence of important genetic predisposing factors. Men with first-degree relatives affected by the disease have a twofold increased risk, and numerous germline variants have been identified that are associated with risk. Among the relatively common variants conveying a modest increase in risk are variants in regulatory regions that influence the expression of *MYC,* an important oncogene in prostate cancer. Other rare variants have been identified that are linked to a high risk of early-onset, aggressive disease. Among these are mutations that disrupt the function of several DNA repair genes, including loss-of-function mutations in *BRCA2* (required for repair by homologous recombination) and in DNA mismatch repair genes (as part of Lynch syndrome; see Chapter 17). A variant in the gene encoding the transcription factor HOXB13, which plays a role in prostatic development during embryogenesis, also has been described that is associated with a severalfold increase in risk.

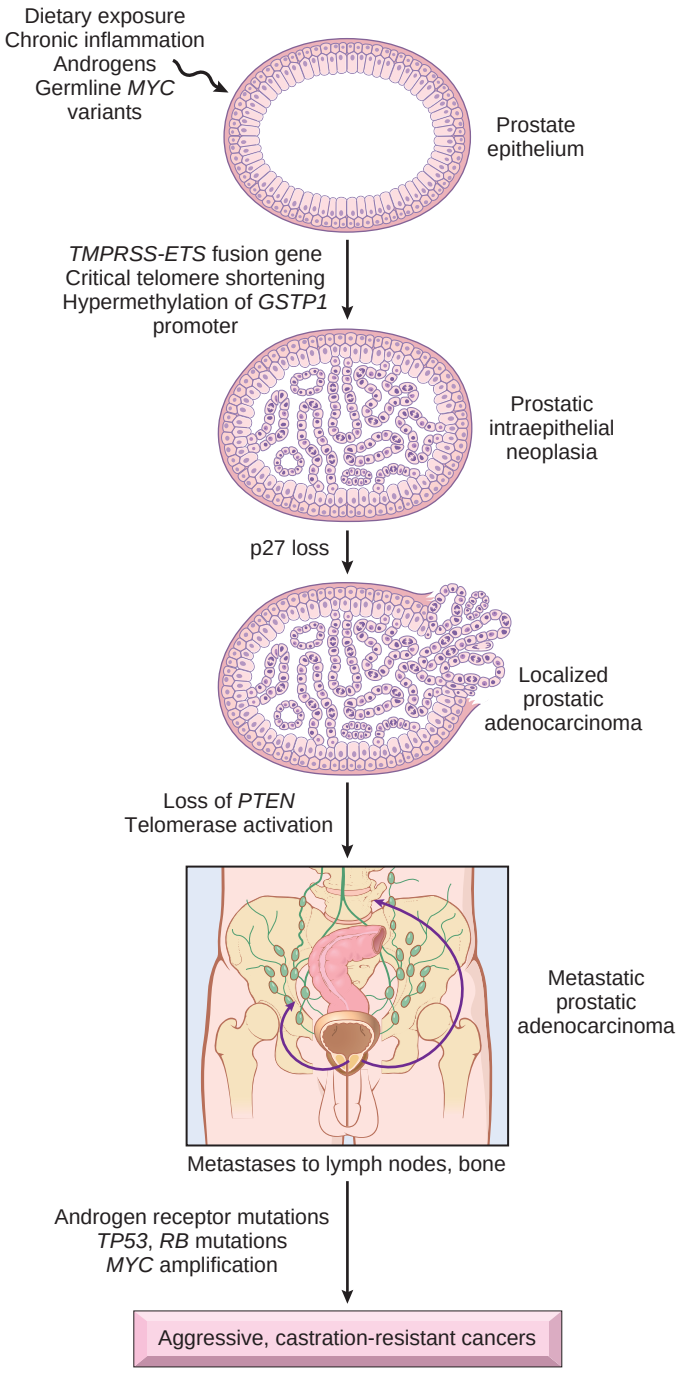


Figure 21.30 Salient environmental and genetic risk factors and acquired genetic and epigenetic alterations leading to prostate cancer development and progression. See text for details.

Androgens. **Like their normal counterparts, the growth and survival of prostate cancer cells depend on androgens, which bind to the AR and induce the expression of pro-growth and pro-survival genes.** The importance of androgens in maintaining the growth and survival of prostate cancer cells can be seen in the therapeutic effect of castration or treatment with antiandrogens, which usually induce disease regression. Unfortunately, most tumors eventually become resistant to androgen blockade. Tumors escape through a variety of mechanisms, including acquisition of hypersensitivity to low levels of androgen (e.g., through AR gene amplification);

ligand-independent AR activation (e.g., via splice variants that lack the ligand binding domain); mutations in AR that allow it to be activated by non-androgen ligands; and other mutations or epigenetic changes that activate alternative signaling pathways and bypass the need for AR altogether.

Acquired Genetic and Epigenetic Alterations. As with other cancers, prostate cancer is caused by genetic and epigenetic alterations that altered the expression of tumor suppressor genes and oncogenes, ultimately leading to acquisition of the hallmarks of cancer (Chapter 7). **The most common genetic alteration in the prostate is a chromosomal rearrangement that juxtaposes the coding sequence of an ETS family transcription factor gene (most commonly *ERG* or *ETV1*) next to the androgen-regulated *TMPRSS2* promoter.** The rearrangements, which occur in approximately half of prostate cancers, place the involved ETS oncogene under the control of the androgen-regulated *TMPRSS2* promoter and lead to its overexpression in an androgen-dependent fashion. Silencing of the gene encoding p27, an inhibitor of cyclin-dependent kinases, is another common event. DNA sequencing has demonstrated many other recurrent genomic rearrangements, primarily deletions and amplifications, that involve other cancer genes in prostate carcinoma. Among these are amplification of *MYC* and deletion of *PTEN,* both of which accelerate cell growth and may contribute to resistance to antiandrogen therapy. In late-stage disease, loss of *TP53* (by deletion or mutation), and deletions of *RB* are common, as are amplifications of the AR gene.

Epigenetic events that modify gene expression are also common in prostate cancer. Work carried out as part of The Cancer Genome Atlas project has identified several distinct subsets of prostate cancers that are defined by differences in DNA methylation. One particularly frequent early event is epigenetic silencing by DNA methylation of the *GSTP1* gene, which downregulates *GSTP1* expression, possibly enhancing the genotoxic effects of environmental carcinogens. Other genes silenced by epigenetic modifications in a subset of prostate cancers include genes involved in cell cycle regulation *(RB, CDKN2A),* maintenance of genomic stability *(MLH1, MSH2),* and Wnt signaling *(APC).* The cumulative salient genetic and epigenetic alterations during prostate cancer development and progression are illustrated in Fig. 21.30.

Precursor Lesions. Evidence in favor of stepwise development of prostate cancer (as outlined in Fig. 21.30) includes the existence of a putative precursor lesion, *prostatic intraepithelial neoplasia (PIN)*. A precursor role for PIN in at least some cases of prostate cancer is supported by several observations. Both PIN and cancer typically predominate in the peripheral zone and are relatively uncommon in other zones, and prostates containing cancer have a higher frequency and a greater extent of PIN, which is often seen in proximity to cancer. More directly, many of the molecular changes seen in invasive cancers are also present in PIN, strongly supporting the argument that PIN is a precursor.

MORPHOLOGY

When the term *prostate cancer* is used without qualification, it refers to the common or acinar variant of prostatic adenocarcinoma. In approximately 70% of cases, carcinoma of the prostate arises in the peripheral zone of the gland, classically in a posterior location, where it may be palpable on rectal examination. Characteristically, on cross-section the neoplastic tissue is gritty and firm to palpation, but it is sometimes extremely difficult to visualize by eye (Fig. 21.31). Histologically, most adenocarcinomas consist of glands arranged in well-defined, easily recognized patterns, which are used to grade these tumors (discussed later). The glands are typically smaller than benign glands and are lined by a single uniform layer of cuboidal or low columnar epithelium. In contrast to benign glands, malignant glands have tightly packed cells and characteristically lack branching and papillary infoldings. **The outer basal cell layer typical of benign glands is absent.** The cytoplasm of the tumor cells ranges from pale-clear to a distinctive amphophilic appearance. Nuclei are enlarged and often contain one or more large nucleoli. There is some variation in nuclear size and shape, but in general pleomorphism is not marked. Mitotic figures are uncommon.

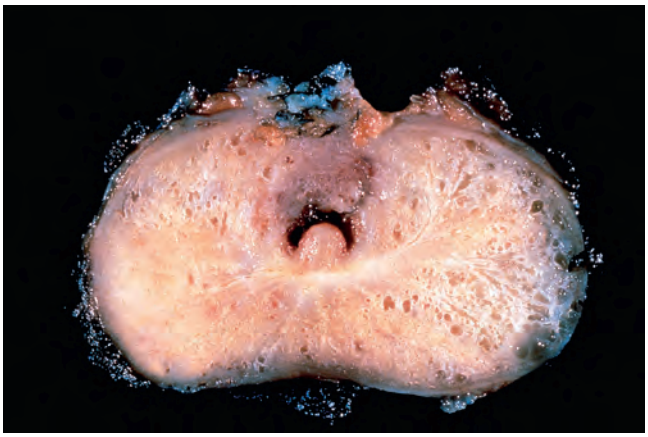


Figure 21.31 Adenocarcinoma of the prostate. An area of cancer is present in the posterior aspect *(lower left)* that has solid gray-white appearance, in contrast to the spongy appearance of benign peripheral zone on the contralateral side.

When prostate cancer invades locally, it most commonly involves periprostatic tissue, seminal vesicles, and the base of the urinary bladder, which in advanced disease may produce ureteral obstruction. Metastases spread via lymphatics to the obturator nodes and eventually to the para-aortic nodes. **Hematogenous spread occurs chiefly to the bones,** particularly the axial skeleton. Bony metastases are typically osteoblastic, a feature that in men points strongly to a prostatic origin (Fig. 21.32). The bones that are most commonly involved, in descending order of frequency, are lumbar spine, proximal femur, pelvis, thoracic spine, and ribs. Tumors may also spread to viscera, but extensive visceral dissemination is the exception rather than the rule.

The diagnosis of prostate cancer on biopsy specimens can be challenging due to several factors. There is often only a scant amount of tissue available for histologic examination in needle biopsies, and malignant glands may be admixed with numerous benign glands (Fig. 21.33). Moreover, the histologic findings may be subtle (leading to underdiagnosis), and there are benign mimickers of cancer that can lead to a misdiagnosis. A few findings are specific, such as perineural invasion (Fig. 21.34), but in general the diagnosis is made based on a constellation of architectural, cytologic, and ancillary findings. As discussed earlier, one distinguishing feature

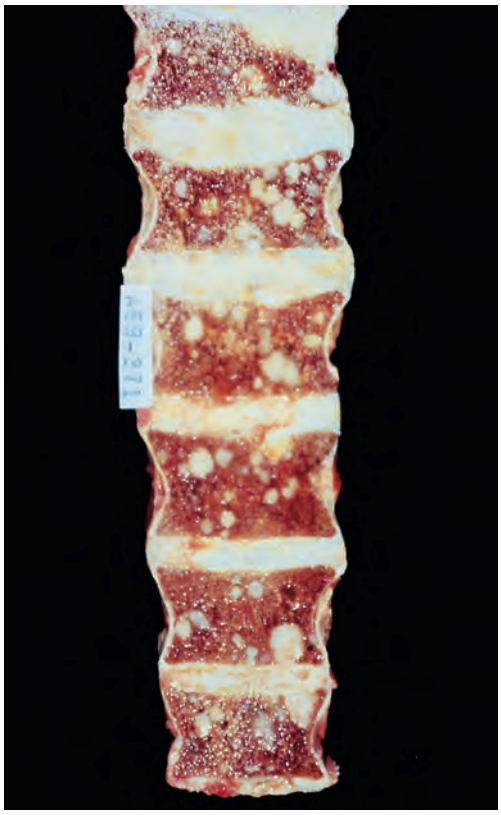


Figure 21.32 Metastatic osteoblastic prostatic carcinoma within vertebral bodies.

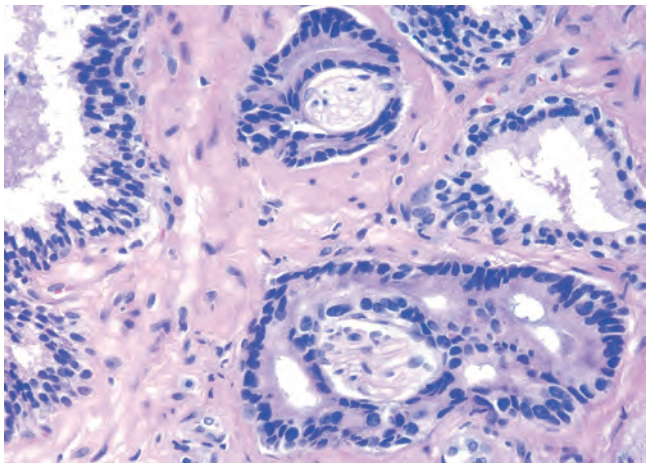


Figure 21.34 Carcinoma of prostate, showing perineural invasion by malignant glands. Compare to benign gland *(left)*.

is that benign glands contain basal cells, which are absent in cancer (compare benign normal glands in Fig. 21.33A and benign hyperplastic glands in Fig. 21.29C with cancerous glands in Fig. 21.33B). This distinction can be brought out by using various immunohistologic markers that stain basal cells. Another useful marker is α-methylacyl coenzyme A racemase (AMACR), which is upregulated in prostate cancer. Most prostate cancers are positive for AMACR, the sensitivity varying among studies from 82% to 100%. Such markers, while improving diagnostic accuracy, are still prone to false-positive and false-negative results and must be used in conjunction with the routine hematoxylin and eosin–stained sections.

As already discussed, in approximately 80% of cases prostatic tissue removed for carcinoma also harbors **prostatic intraepithelial neoplasia (PIN)**. PIN consists of architecturally benign large, branching prostatic acini lined by atypical cells with prominent nucleoli that may be cytologically identical to carcinoma. Unlike malignant glands, glands involved by PIN retain, at least partially, a layer of basal cells and have an intact basement membrane.

Grading and Staging. **Grade and stage are the most important prognostic factors in prostate cancer.** Grading is performed using the Gleason system, which stratifies prostate cancer into five grades on the basis of glandular patterns of growth. Grade 1 corresponds to well-differentiated tumors in which the neoplastic glands are uniform and round in appearance and are packed into well-circumscribed nodules (Fig. 21.35A). In contrast, grade 5 tumors do not form glands, with tumor cells infiltrating the stroma in cords, sheets, and solid nests (Fig. 21.35C). Other grades fall between

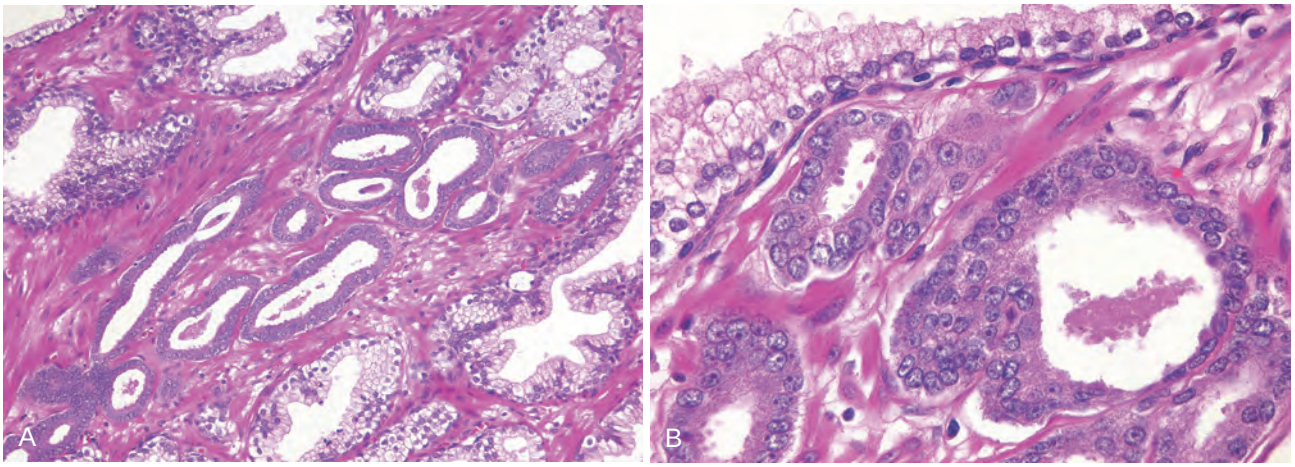


Figure 21.33 Prostatic adenocarcinoma. (A) Photomicrograph of small focus of adenocarcinoma of the prostate demonstrating small glands crowded in between larger benign glands. (B) Higher magnification shows several small malignant glands with enlarged nuclei, prominent nucleoli, and dark cytoplasm compared with larger benign gland *(top)*.

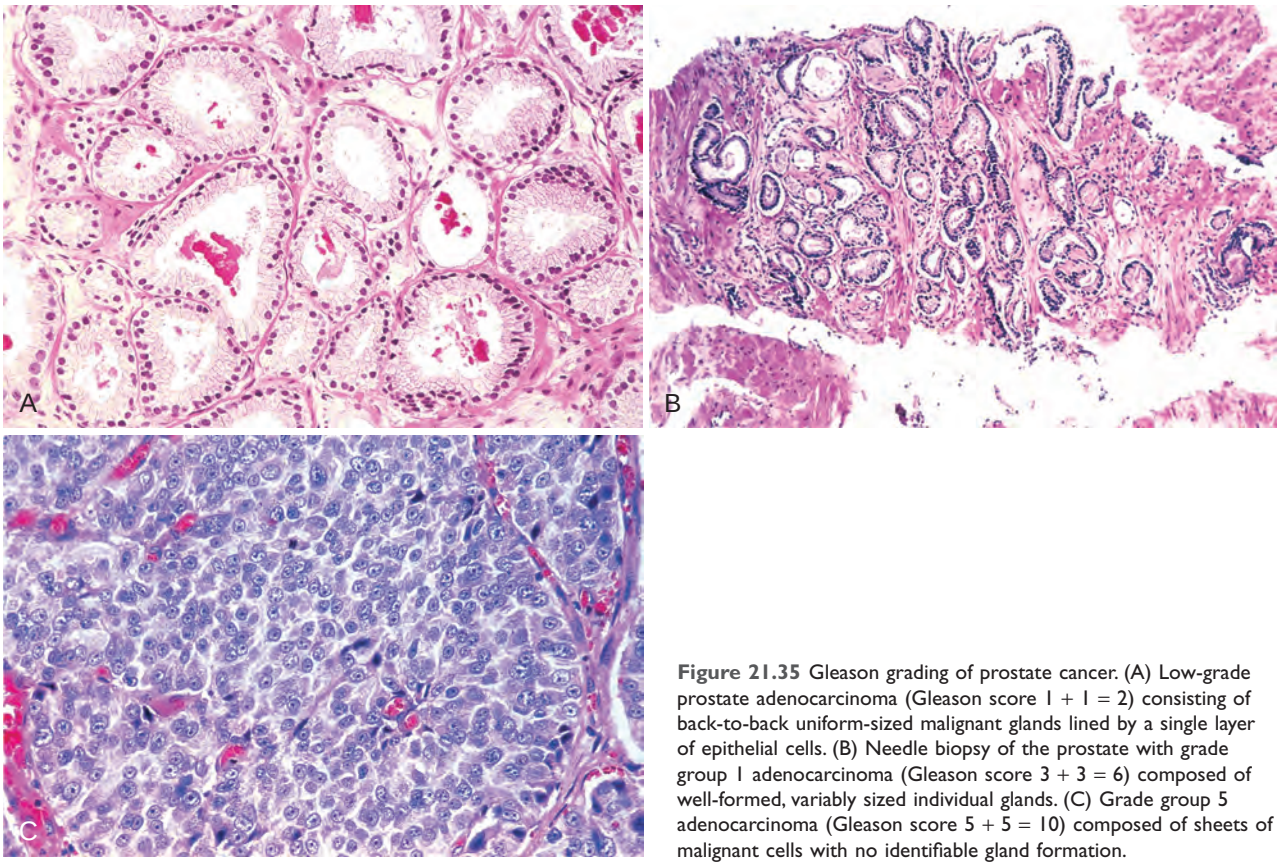


Figure 21.35 Gleason grading of prostate cancer. (A) Low-grade prostate adenocarcinoma (Gleason score 1 + 1 = 2) consisting of back-to-back uniform-sized malignant glands lined by a single layer of epithelial cells. (B) Needle biopsy of the prostate with grade group 1 adenocarcinoma (Gleason score 3 + 3 = 6) composed of well-formed, variably sized individual glands. (C) Grade group 5 adenocarcinoma (Gleason score 5 + 5 = 10) composed of sheets of malignant cells with no identifiable gland formation.

these extremes (Fig. 21.35B). Most tumors contain more than one pattern; in such instances, a primary grade is assigned to the dominant pattern, and a secondary grade is assigned to the second most frequent pattern. The two numeric grades are then added to obtain a combined Gleason score. Tumors with only one pattern are treated as if their primary and secondary grades are the same, and hence the number is doubled for the score. The majority of potentially treatable cancers detected on needle biopsy as a result of screening have Gleason scores of 6 or 7. Tumors with Gleason scores of 8 through 10 tend to be advanced cancers that are less likely to be cured. Although there is some evidence that prostate cancers can become more aggressive with time, most commonly, the Gleason score remains stable over a period of several years. Presently, Gleason scores are combined into five Grade Groups (Table 21.8), each with a different prognosis.

Pathologic staging of prostatic cancer is used in combination with the grade to stratify management of prostate cancer. As shown in Table 21.9, pTNM staging is based on tumor extent (T) and presence of nodal or distant metastasis (N and M). Grade is now integrated with TNM pathologic stage data to derive combined prognostic groups that result in "upstaging" of some tumors.

Clinical Features

Localized prostate cancer is asymptomatic and is usually discovered by the detection of a suspicious nodule on rectal examination or elevated serum prostate-specific antigen (PSA) level (discussed later). Digital rectal examination may detect some early prostatic carcinomas because of their posterior location, but suffers from both low sensitivity and specificity. Patients with clinically advanced prostatic cancer may present with symptoms of urinary obstruction. Typically, a transrectal needle biopsy is required to confirm the diagnosis.

Measurement of serum PSA levels is widely used to assist with the diagnosis and management of prostate cancer but is controversial. PSA is a product of prostatic epithelium and is normally secreted in the semen. It is an androgen-regulated serine protease whose function is to cleave and liquefy the seminal coagulum formed after ejaculation. In normal men, only minute amounts of PSA circulate in the serum. Elevated blood levels of PSA occur in association with localized as well as advanced cancer. However, as a screening test for prostate cancer, PSA measurement has suboptimal sensitivity and specificity.

Table 21.8 Prostate Cancer Gleason Grade Groups

- Grade Group 1 (≤6)
  Only individual discrete well-formed glands
- Grade Group 2 (3 + 4)
  Predominantly well-formed glands with a lesser component of poorly formed, fused or cribriform glands
- Grade Group 3 (4 + 3)
  Predominantly poorly formed/fused/cribriform glands with a lesser component of well-formed glands
- Grade Group 4 (4 + 4/3 + 5/5 + 3)
  Only poorly formed/fused/cribriform glands or predominantly mix of well-formed glands and lack of glands
- Grade Group 5 (4 + 5/5 + 4/5 + 5)
  Lack gland formation (or with necrosis) with or without poorly formed/fused/cribriform glands

Table 21.9 Pathologic Staging of Prostatic Adenocarcinoma Using the pTNM System

pTNM, AJCC 8th Edition	
pTNM Designation	**Anatomic Findings**
Extent of Primary Tumor (T)	
pT2	Organ confined
pT3	Extraprostatic extension
pT3a	Extraprostatic extension (unilateral or bilateral) or microscopic invasion of bladder neck
pT3b	Tumor invades seminal vesicles
pT4	Tumor is fixed or invades adjacent structures other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, or pelvic wall
Definition of Regional Lymph Nodes (N)	
Nx	Regional nodes not accessed
N0	No regional nodal metastases
N1	Metastasis in regional lymph nodes
Definition of Distant Metastases (M)	
M0	No distant metastases
M1	Distant metastases present
M1a	Metastases to distant lymph nodes
M1b	Bone metastases
M1c	Other distant sites

AJCC, American Joint Commission on Cancer.

PSA is organ specific, but not cancer specific. Although serum levels of PSA are elevated on average to a lesser extent in patients with BPH than in those with prostate cancer, there is considerable overlap in these two groups of patients. Other factors such as prostatitis, prostate gland infarction (e.g., in the setting of BPH), instrumentation of the prostate, and ejaculation also increase serum PSA levels. Of even greater concern, many prostate cancers are so indolent that they are clinically insignificant, and detection of such cancers may lead to overtreatment, with its associated morbidity and economic costs. As a result of these concerns, some countries with national health care systems (e.g., the United Kingdom) have recommended against using PSA as a screening test. By contrast, once the diagnosis of prostate cancer has been established, serial measurements of PSA after treatment has clear-cut value in monitoring recurrence and disease progression.

Following the diagnosis of prostate cancer, the next challenge is to determine whether the cancer is biologically significant, that is, one that is likely to progress locally or metastasize. In addition to tumor grade, a number of genomic tests using gene panels have recently been developed that may enhance the ability to predict which cancers (e.g., low-volume Grade Group 1) can be followed by active surveillance alone. Patients with localized prostate cancer that is considered to be at high risk for local and systemic spread are treated with surgery or radiation, with or without hormonal manipulation. More than 90% of patients who receive such therapy can expect to live for 15 years. Currently, the most common treatment is radical prostatectomy. The prognosis following radical prostatectomy is based on the pathologic stage, margin status, and Gleason grade. For patients who are not candidates for surgery because of age or other medical conditions, radiotherapy with either external-beam radiation or placement of radioactive seeds in the prostate (brachytherapy) often provides excellent local control of disease.

Metastatic carcinoma is treated with androgen deprivation therapy. Androgen deprivation is usually achieved by orchiectomy or by administration of synthetic analogs of luteinizing hormone–releasing hormone (LHRH); chronic administration of LHRH agonists desensitizes pituitary cells expressing LHRH receptors, suppressing the release of luteinizing hormone (LH), which is required for testosterone production by Leydig cells. Although androgen deprivation therapy induces remissions, most tumors eventually become resistant to testosterone withdrawal, an event that is a harbinger of disease progression and death.

KEY CONCEPTS

CARCINOMA OF THE PROSTATE

- Carcinoma of the prostate is very common in older men. In the United States, it is the most common malignancy in men.
- Prostate carcinomas range from very indolent lesions, which are being increasingly managed by active surveillance, to lethal disease that requires definitive therapy.
- Carcinomas of the prostate arise most commonly in the outer, peripheral gland and may be palpable by rectal examination.
- The most common driver mutations in prostate cancer are gene rearrangements that result in androgen dependent overexpression of Ets family transcription factors, most commonly ERG or ETV1.
- Grading of prostate cancer by the Gleason system strongly correlates with pathologic stage and is a strong prognosticator. The system is compressed into Grade Groups that also influence the combined prognostic stage of the tumor.
- Most localized cancers are clinically silent and are detected by routine monitoring of PSA concentrations in older men. Bone metastases, often osteoblastic, typify advanced prostate cancer.
- Serum PSA measurement is a useful but imperfect cancer-screening test, with significant rates of false-negative and false-positive results. Evaluation of PSA concentrations after treatment has great value in monitoring progressive or recurrent disease.

Miscellaneous Tumors

Prostate adenocarcinomas may also arise from prostatic ducts. *Ductal adenocarcinoma* arising in peripheral ducts may present in a fashion similar to ordinary prostate cancer, whereas those arising in the larger periurethral ducts may show signs and symptoms similar to urothelial cancer, causing hematuria and urinary obstructive symptoms. Ductal adenocarcinomas are associated with a relatively poor prognosis. Prostate cancers may show squamous differentiation, either following hormone therapy or de novo, resulting in either adenosquamous or pure squamous cancer. Prostate cancers that reveal abundant mucinous secretions in greater than 25% of the tumor are termed *colloid carcinoma of the prostate.*

The most aggressive variant of prostate cancer is *small-cell carcinoma* (also known as neuroendocrine carcinoma). Almost all cases of small-cell carcinoma are rapidly fatal. This cancer is increasing in frequency and is most often seen as a form of recurrent disease in patients with typical prostate cancer undergoing treatment with antiandrogen therapies. While these therapies control conventional prostate cancers, in some cases they provoke the emergence of an androgen-independent sub-clone with a neuroendocrine phenotype.

The most common tumor to secondarily involve the prostate is urothelial cancer. Two distinct patterns of involvement exist. Large invasive urothelial cancers can directly invade from the bladder into the prostate. Alternatively, CIS of the bladder can extend into the prostatic urethra and down into the prostatic ducts and acini.

The same mesenchymal tumors described earlier that involve the bladder may also manifest in the prostate. In addition, there exist unique mesenchymal tumors of the prostate derived from the prostatic stroma.

SUGGESTED READINGS

Penile Disorders

Alemany L, Cubilla A, Halec G et al.; HPV VVAP Study Group: Role of human papillomavirus in penile carcinomas worldwide, *Eur Urol* 69(5):953–961, 2016. *[Review of the role of HPV in penile carcinoma].*

Cubilla AL, Velazquez EF, Amin MB et al.; Members of the ISUP Penile Tumor Panel: the World Health Organisation 2016 classification of penile carcinomas: a review and update from the International Society of Urological Pathology expert-driven recommendations, *Histopathology* 72(6):893–904, 2018. *[Update of the classification of penile carcinoma].*

Moch H, Cubilla AL, Humphrey PA et al: The 2016 WHO classification of tumours of the Urinary System and Male Genital Organs-Part A: renal, penile, and testicular tumours, *Eur Urol* 70(1):93–105, 2016. *[Latest classification system for male urogenital neoplasms].*

Testicular Disorders

Idrees MT, Ulbright TM, Oliva E et al; Members of the International Society of Urological Pathology Testicular Tumour Panel: the world health organization 2016 classification of testicular non-germ cell tumours: a review and update from the international society of urological pathology testis consultation panel, *Histopathology* 70(4):513–521, 2017. *[Latest classification system for non–germ cell testicular neoplasms].*

Litchfield K, Levy M, Huddart RA et al: The genomic landscape of testicular germ cell tumours: from susceptibility to treatment, *Nat Rev Urol* 13:409, 2016. *[Discussion of the pathophysiology of testicular germ cell tumors and implications for therapy].*

Looijenga LH, Stoop H, Biermann K: Testicular cancer: biology and biomarkers, *Virchows Arch* 464(3):301–313, 2014. *[Review of salient features of germ cell tumor biology and associated biomarkers].*

Taylor-Weiner A, Zack T, O'Donnell E et al: Genomic evolution and chemoresistance in germ-cell tumours, *Nature* 540:114, 2016. *[A landmark paper describing genetic drivers in testicular germ cell tumors and additional genomic changes that emerge with resistance to chemotherapy].*

Bladder Disorders

Cancer Genome Atlas Research Network: Comprehensive molecular characterization of urothelial bladder carcinoma, *Nature* 507(7492): 315–322, 2014. *[Landmark paper providing the first genome-wide view of driver genetic events in bladder carcinoma].*

Eich ML, Dyrskjøt L, Netto GJ: Toward personalized management in bladder cancer: the promise of novel molecular taxonomy, *Virchows Arch* 471(2):271–280, 2017. *[Discussion of the implications of the genetics and transcriptomics of bladder carcinoma for classification and treatment].*

Humphrey PA, Moch H, Cubilla AL et al: The 2016 WHO classification of tumours of the Urinary System and Male Genital Organs-Part B: prostate and bladder tumours, *Eur Urol* 70(1):106–119, 2016. *[Update on the classification of bladder carcinoma].*

Robertson AG, Kim J, Al-Ahmadie H et al: Comprehensive molecular characterization of muscle-invasive bladder cancer, *Cell* 171(3):540–556. e25, 2017. *[A first look at the molecular characterization of invasive bladder carcinoma].*

Prostatic Disorders

Cancer Genome Atlas Research Network: The molecular taxonomy of primary prostate cancer, *Cell* 163(4):1011–1025, 2015. *[Landmark paper providing the first genome-wide view of driver genetic events in prostate carcinoma].*

Epstein JI, Egevad L, Amin MB et al.; Grading Committee: The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system, *Am J Surg Pathol* 40(2):244–252, 2016. *[Current recommendations for prostate cancer grading].*

Humphrey PA, Moch H, Cubilla AL et al: The 2016 WHO classification of tumours of the Urinary System and Male Genital Organs-Part B: prostate and bladder tumours, *Eur Urol* 70(1):106–119, 2016. *[Update on classification of prostate and bladder carcinoma].*

Netto GJ, Eich ML, Varambally S: Prostate cancer: an update on molecular pathology with clinical implications, *Eur Urol Suppl* 16:253–271, 2017. *[Discussion of the implications of the genetics and transcriptomics of prostate cancer for classification and treatment].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 22

The Female Genital Tract

Lora Hedrick Ellenson • Edyta C. Pirog

CHAPTER CONTENTS

A brief review of the development and anatomy of the female genital tract is fundamental to understanding the diseases that affect this complex organ system. Normal development of the female genital tract proceeds through a series of tightly choreographed events involving the primordial germ cells, the müllerian (paramesonephric) ducts, the wolffian (mesonephric) ducts, and the urogenital sinus (Fig. 22.1).

- *Germ cells* arise in the wall of the yolk sac by the fourth week of gestation. By the fifth or sixth week, they migrate into the urogenital ridge and induce proliferation of the mesodermal epithelium, which gives rise to the epithelium and stroma of the ovary.
- The lateral *müllerian ducts* form at about the sixth week of development through invagination and fusion of the coelomic lining epithelium. The ducts progressively grow caudally into the pelvis, where they swing medially to fuse with the urogenital sinus at the müllerian tubercle (see Fig. 22.1A). Further caudal growth brings these fused ducts into contact with the urogenital sinus. The unfused upper portions of the müllerian ducts mature into the fallopian tubes, while the fused lower portion develops into the uterus, cervix, and upper vagina.
- The *urogenital sinus* develops when the cloaca is subdivided by the urorectal septum; it eventually forms the lower part of the vagina and the vestibule of the external genitalia (see Fig. 22.1B).
- The *mesonephric ducts* normally regress in the female, but remnants may persist into adult life as epithelial inclusions adjacent to the ovaries, tubes, and uterus. In the cervix and vagina, these rests may be cystic and are termed *Gartner duct cysts*.

The epithelial lining of the female genital tract as well as the ovarian surface share a common origin from coelomic epithelium (mesothelium), which may explain why morphologically similar benign and malignant lesions arise in various sites within the female genital tract and the adjacent peritoneal surfaces.

Diseases of the female genital tract are extremely common and include complications of pregnancy, infections, tumors, and hormonally induced abnormalities. The following discussion presents the pathology of the major diseases that result in clinical problems. Additional details can be found in current textbooks of gynecologic pathology and clinical obstetrics and gynecology. We will discuss the pathologic conditions peculiar to each segment of the female genital tract separately, but before doing so we will briefly review infections and pelvic inflammatory disease because they can affect many of the various anatomic structures concomitantly.

INFECTIONS

A large variety of organisms can infect the female genital tract. Some infections with microorganisms such as *Candida, Trichomonas,* and *Gardnerella* are very common and may cause significant discomfort, but are without serious sequelae. Others, such as *Neisseria gonorrhoeae* and *Chlamydia* infections, are major causes of infertility, while organisms such as *Group B Streptococcus* infections are implicated in preterm deliveries, stillbirths, and neonatal infections. Viruses, especially herpes simplex viruses (HSVs) and human papillomaviruses (HPVs), also account for considerable morbidity; HSVs cause painful genital ulcerations, whereas HPVs are involved in the pathogenesis of cervical, vaginal, and vulvar cancers (described later).

Many of these infections are sexually transmitted, including trichomoniasis, gonorrhea, *Mycoplasma, Chlamydia,* HSV, and HPV as well as less common infections such as syphilis, chancroid, granuloma inguinale, and lymphogranuloma venereum. Most of these conditions are considered in Chapter 8; HPV is also discussed in Chapter 7 due to its important

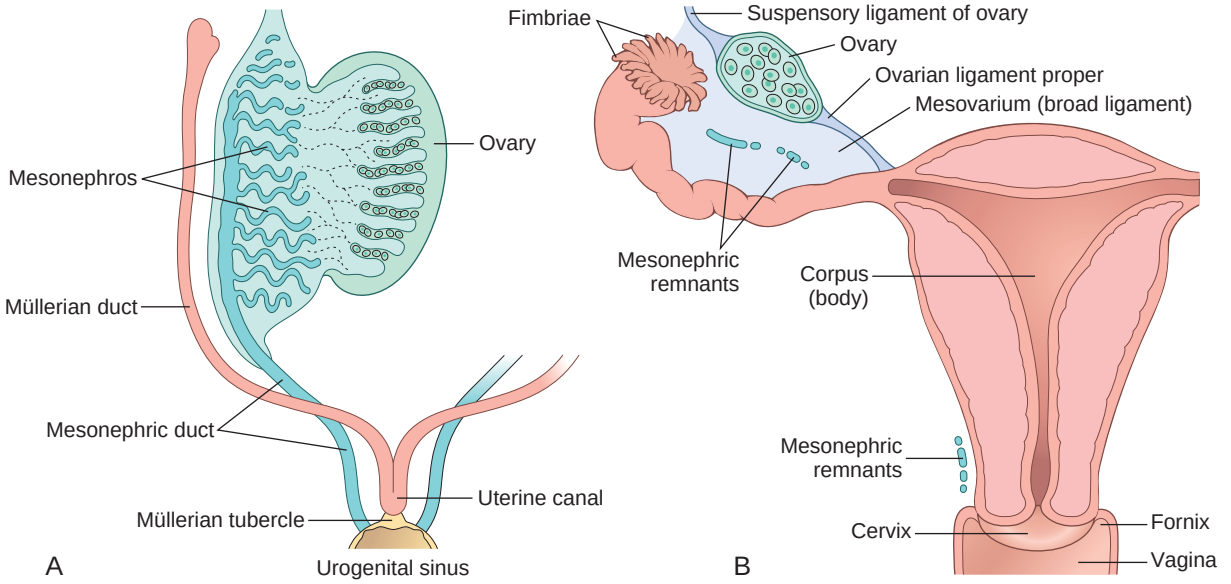


Figure 22.1 Embryology and anatomy of the female genital tract. (A) Early in development, the mesonephric *(blue)* and müllerian *(red)* ducts merge at the urogenital sinus to form the müllerian tubercle. (B) By birth, the müllerian ducts have fused to form the fallopian tubes, uterus, and endocervix *(red)*, merging with the vaginal squamous mucosa. The mesonephric ducts regress, but may be found as a remnant in the ovary, adnexa, and cervix (Gartner duct cysts). (Modified from Langman J: *Medical Embryology*, Baltimore, 1981, Williams and Wilkins.)

role as a transforming virus. Here we touch only on aspects relevant to the female genital tract, including pathogens confined to the lower genital tract (vulva, vagina, and cervix) and those that involve the entire genital tract and are implicated in pelvic inflammatory disease.

Infections of the Lower Genital Tract

Herpes Simplex Virus

Genital herpes simplex virus (HSV) infection is common and may involve the vulva, vagina, or cervix. HSVs are DNA viruses that include two serotypes, HSV-1 and HSV-2. HSV-1 typically results in perioral infection, whereas HSV-2 usually involves genital mucosa and skin; however, depending on sexual practices, HSV-1 may be detected in the genital region and HSV-2 may cause oral infections as well (Chapter 8). By 40 years of age, approximately 30% of women are seropositive for antibodies against HSV-2.

About one-third of newly infected individuals are symptomatic. Lesions typically develop 3 to 7 days after transmission and are often associated with systemic symptoms such as fever, malaise, and tender inguinal lymph nodes. The earliest lesions usually consist of red papules that progress to vesicles and then to painful coalescent ulcers. The lesions are easily visible on vulvar skin and mucosa, while cervical and vaginal lesions present with purulent discharge and pelvic pain. Lesions around the urethra may cause painful urination and urine retention. The vesicles and ulcers contain numerous viral particles, accounting for the high transmission rate during active infection. The mucosal and cutaneous lesions heal spontaneously in 1 to 3 weeks, but during the acute infection the virus migrates to the regional lumbosacral nerve ganglia and establishes a latent infection. Because of viral latency, HSV infections persist indefinitely, and any decrease in immune function due to stress, trauma, concurrent viral infection, or hormonal changes can trigger reactivation of the virus and recurrence of the skin and mucosal lesions. As expected, recurrences are much more common in immunosuppressed individuals.

MORPHOLOGY

By the time an HSV lesion is biopsied, it typically has ulcerated. The epithelium is desquamated, and marked acute inflammation is present in the ulcer bed. Smears of the inflammatory exudate from active lesions show characteristic cytopathic changes consisting of multinucleated squamous cells containing eosinophilic to basophilic viral inclusions with a "ground-glass" appearance (Fig. 22.2).

Transmission of HSV takes place mainly during the active phase but occasionally may occur during the latent phase due to subclinical virus shedding. Condoms and antiviral therapies reduce the risk of transmission, but do not prevent it completely. As with other sexually transmitted diseases, the virus is more readily transmitted to women than to men. Previous infection with HSV-1 seems to reduce susceptibility to HSV-2 infection. The gravest consequence of HSV infection is transmission to the neonate during birth. This risk is highest if the infection is active during delivery and particularly if it is a primary (initial) infection in the mother. Cesarean delivery is warranted in such cases to prevent transmission to the neonate. Another important consequence of active genital HSV infection is that it increases the risk of HIV-1 acquisition and transmission.

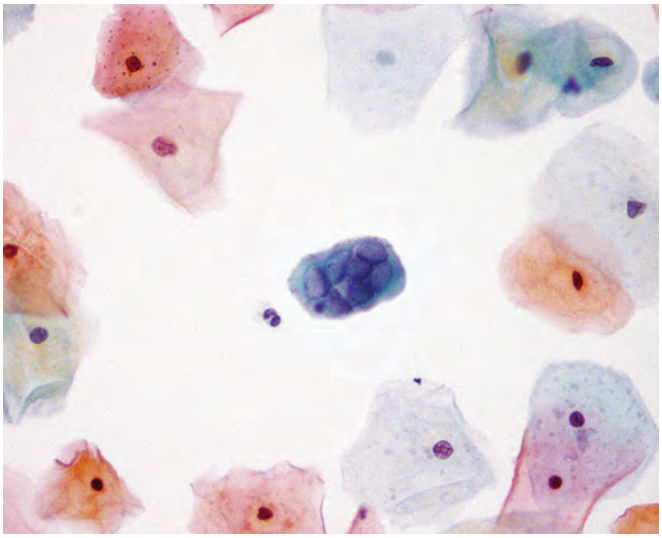


Figure 22.2 Herpes simplex virus (HSV) infection (cervical smear). The cell in the center shows HSV cytopathic effect. Infected cells become multinucleated and contain intranuclear viral inclusions with a characteristic "ground-glass" appearance.

The diagnosis is based on typical clinical findings and HSV detection. The purulent exudate is aspirated from the lesions and inoculated into a tissue culture. The viral cytopathic effect can be seen after 48 to 72 hours, and the virus can then be serotyped. In addition, some laboratories offer more sensitive polymerase chain reaction tests, enzyme-linked immunosorbent assays, and direct immunofluorescent antibody tests for detection of HSV in the lesional secretions. Individuals with primary, acute HSV infection do not have serum anti-HSV antibodies. Detection of anti-HSV antibodies in the serum is indicative of recurrent/latent infection.

There is no effective treatment for latent HSV; however, antiviral agents like acyclovir or famciclovir shorten the length of the initial and recurrent symptomatic phases. The ultimate solution is an effective vaccine, a tantalizing goal yet to be realized.

Other Lower Female Genital Tract Infections

As mentioned earlier, a variety of other viruses, fungi, and bacteria can also cause symptomatic infections of the lower genital tract. Those that are most common include the following:

- *Molluscum contagiosum* is a cutaneous or mucosal lesion caused by poxvirus (Fig. 22.3A). There are four types of molluscum contagiosum viruses (MCVs), MCV-1 to MCV-4, with MCV-1 being the most prevalent and MCV-2 being most often sexually transmitted. The infections are common in young children between 2 and 12 years of age, in whom they are transmitted through direct contact or shared articles (e.g., towels). Molluscum may affect any area of the skin but is most common on the trunk, arms, and legs. In adults, molluscum infections are typically sexually transmitted and affect the genitals, lower abdomen, buttocks, and inner thighs. The average

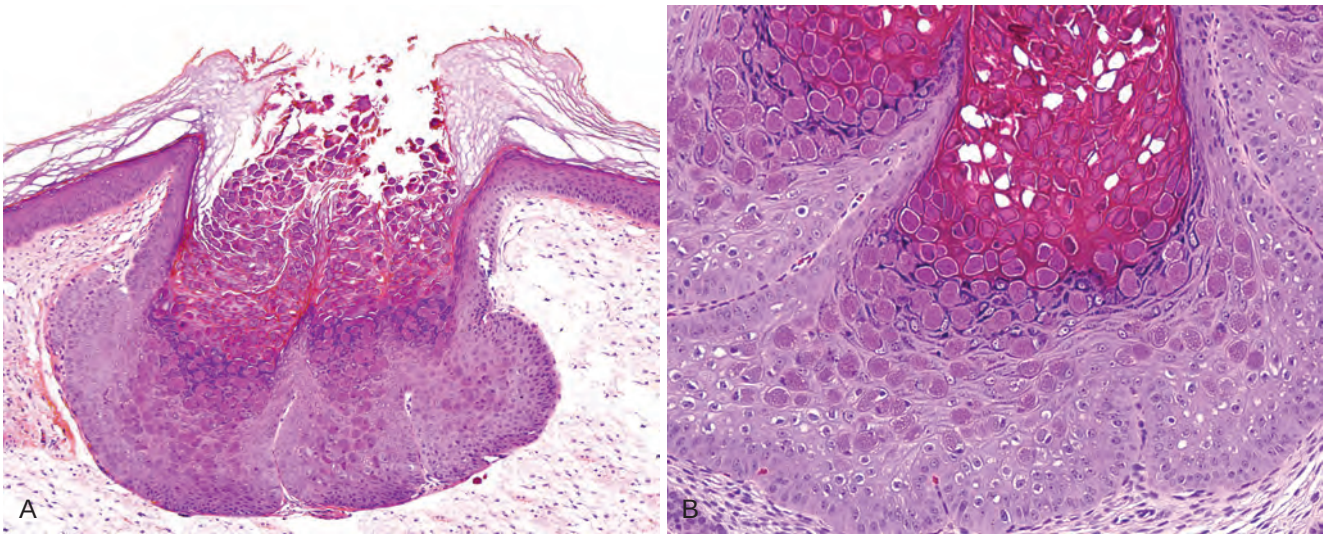


Figure 22.3 Molluscum contagiosum infection. (A) Low-power appearance of a dome-shaped papule with dimpled center. (B) High-power magnification reveals intracytoplasmic viral inclusions.

incubation period is 6 weeks. Diagnosis is based on the characteristic clinical appearance of pearly, dome-shaped papules with a dimpled center. The papules measure 1 to 5 mm in diameter, and their central waxy core contains cells with *cytoplasmic viral inclusions* (see Fig. 22.3B).

- *Fungal infections,* especially those caused by yeast *(Candida),* are extremely common; in fact, yeast are part of many women's normal vaginal microflora, and the development of symptomatic candidiasis is typically a result of a disturbance in the patient's vaginal microbial ecosystem. Diabetes mellitus, antibiotics, pregnancy, and conditions that compromise neutrophil or Th17 T-cell function increase the risk of symptomatic candidal infection, which manifests itself by marked vulvovaginal pruritus, erythema, swelling, and curdlike vaginal discharge. Severe infection may result in mucosal ulcerations. The diagnosis is made by finding the pseudospores or filamentous fungal hyphae in wet KOH mounts of the discharge or on Papanicolaou (Pap) smear. Even though sexual transmission of yeast infection has been documented, candidiasis is not considered a sexually transmitted disease.
- *Trichomonas vaginalis* is a large, flagellated ovoid protozoan that is usually transmitted by sexual contact. Infected patients may be asymptomatic or may complain of yellow, frothy vaginal discharge, vulvovaginal discomfort, dysuria (painful urination), and dyspareunia (painful intercourse). The vaginal and cervical mucosa typically has a fiery-red appearance, with marked dilation of cervical mucosal vessels resulting in characteristic colposcopic appearance of "strawberry cervix."
- *Gardnerella vaginalis* is a gram-negative coccobacillus that is implicated as the main cause of *bacterial vaginosis* (vaginitis). Patients typically present with thin, green-gray, malodorous (fishy) vaginal discharge. Pap smears reveal superficial and intermediate squamous cells covered with a shaggy coat of coccobacilli. Bacterial cultures in such cases reveal *G. vaginalis* and other bacteria, including anaerobic peptostreptococci and aerobic α-hemolytic streptococci. In pregnant patients, bacterial vaginosis has been implicated in premature labor.
- *Ureaplasma urealyticum* and *Mycoplasma hominis* species account for some cases of vaginitis and cervicitis and have been implicated in chorioamnionitis and premature delivery in pregnant patients.
- *Chlamydia trachomatis* infections mainly take the form of cervicitis. However, in some patients the infection may ascend to the uterus and fallopian tubes, resulting in endometritis and salpingitis; thus Chlamydia is one of the causes of pelvic inflammatory disease, as discussed next.

Infections Involving the Lower and Upper Genital Tract

Pelvic Inflammatory Disease (PID)

PID is an infection that begins in the vulva or vagina and spreads upward to involve most of the structures in the female genital system, resulting in pelvic pain, adnexal tenderness, fever, and vaginal discharge. *Neisseria gonorrhoeae* continues to be a common cause of PID, the most serious complication of gonorrhea in women. *Chlamydia* infection is another well-recognized cause of PID. Infections after spontaneous or induced abortions and normal or abnormal deliveries (called *puerperal infections*) are also important causes of PID. In these situations, the infections are typically polymicrobial and may be caused by staphylococci, streptococci, coliforms, and *Clostridium perfringens*.

With gonococcus, inflammatory changes start to appear approximately 2 to 7 days after inoculation. The initial infection most commonly involves the endocervical mucosa, but it may also begin in the Bartholin gland and other vestibular, or periurethral, glands. From these sites, the organisms may spread upward to involve the fallopian tubes and tuboovarian region. The non-gonococcal bacterial infections that follow induced abortion, dilation and curettage of the uterus, and other surgical procedures are thought to spread upward from the uterus through the lymphatics or venous

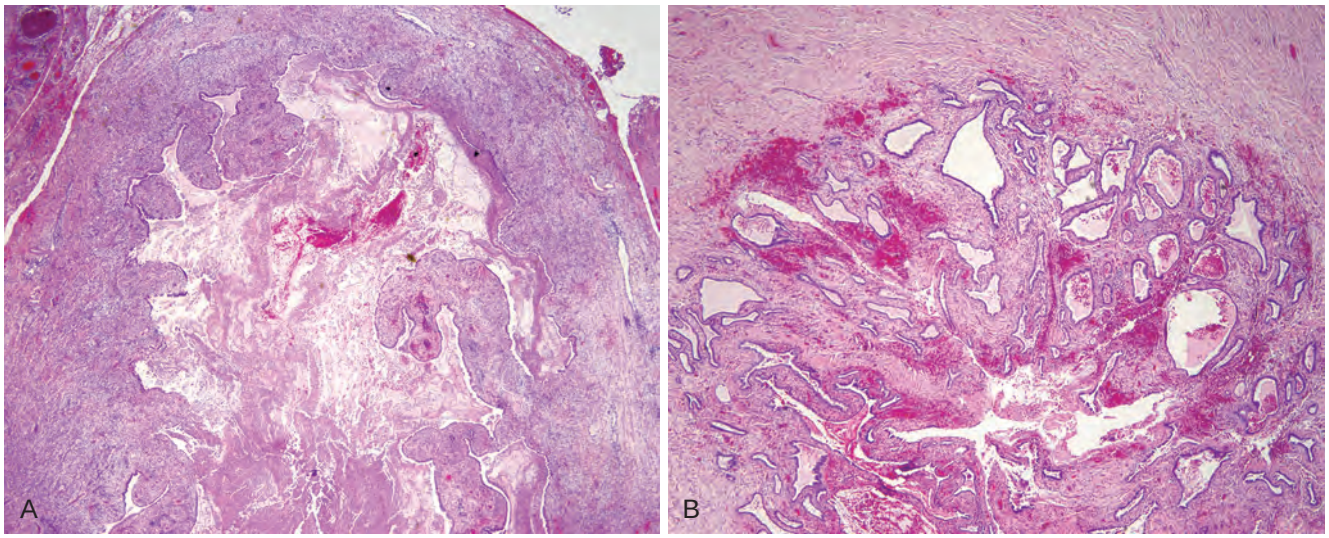


Figure 22.4 Salpingitis. (A) Acute salpingitis; note the dilated tube lumen and edematous tubal plicae expanded by inflammatory cell infiltrates. Pus fills the center of the fallopian tube. (B) Chronic salpingitis showing scarring and fusion of the plicae with formation of glandlike spaces. Such scarring may cause infertility or ectopic tubal pregnancy.

channels rather than on the mucosal surfaces. Therefore, these infections tend to produce more inflammation within the deeper layers of the organs than gonococcal infections.

MORPHOLOGY

Gonococcal infection is characterized by marked acute inflammation of involved mucosal surfaces. Smears of the inflammatory exudate disclose phagocytosed gram-negative diplococci within neutrophils; however, definitive diagnosis requires culture or detection of gonococcal RNA or DNA. If infection spreads, the endometrium is usually spared (for unclear reasons), but within the fallopian tubes, an **acute suppurative salpingitis** ensues (Fig. 22.4A). The tubal mucosa becomes congested and diffusely infiltrated by neutrophils, plasma cells, and lymphocytes, resulting in epithelial injury and sloughing of the plicae. The tubal lumen fills with purulent exudate that may leak out of the fimbriated end. The infection may then spread to the ovary to create a **salpingo-oophoritis.** Collections of pus may accumulate within the ovary and tube **(tubo-ovarian abscess)** or tubal lumen **(pyosalpinx)** (see Fig. 22.4A). With time, the infecting organisms may disappear, but the tubal plicae, denuded of epithelium, adhere to one another and slowly fuse in a reparative, scarring process that forms glandlike spaces and blind pouches, referred to as **chronic salpingitis** (see Fig. 22.4B). The scarring of the tubal lumen and fimbriae may prevent the uptake and passage of oocytes, leading to infertility or ectopic pregnancy. **Hydrosalpinx** may also develop as a consequence of the fusion of the fimbriae and the subsequent accumulation of the tubal secretions and tubal distention.

As compared to gonococcal infections, PID caused by staphylococci, streptococci, and the other puerperal invaders tends to show less involvement of the mucosa and the tube lumen, and more inflammation within the deeper tissue layers. These infections often spread throughout the wall to involve the serosa and the broad ligaments, pelvic structures, and peritoneum. Bacteremia is a more frequent complication of streptococcal or staphylococcal PID than of gonococcal infections.

The acute complications of PID include peritonitis and bacteremia, which in turn may result in endocarditis, meningitis, and suppurative arthritis. The chronic sequelae of PID include infertility and tubal obstruction, ectopic pregnancy, pelvic pain, and intestinal obstruction due to adhesions between the bowel and pelvic organs. In the early stages, gonococcal infections are usually readily controlled with antibiotics, although penicillin-resistant strains have emerged. Infections that become walled off in tubo-ovarian abscesses are difficult to eradicate with antibiotics, and it sometimes becomes necessary to remove the organs surgically. Postabortion and postpartum PIDs may also be amenable to treatment with antibiotics, but are far more difficult to control because of the broad spectrum of pathogens that may be involved.

Vulva

Diseases of the vulva in the aggregate constitute only a small fraction of gynecologic practice. Many inflammatory diseases that affect skin elsewhere on the body also occur on the vulva, such as psoriasis, eczema, and allergic dermatitis. Because it is constantly exposed to secretions and moisture, the vulva is more prone to superficial infections than skin elsewhere on the body. Nonspecific vulvitis is particularly likely to occur in the setting of immunosuppression. Most skin cysts (epidermal inclusion cysts) and skin tumors such as squamous cell carcinoma, basal cell carcinoma, and melanoma can also occur in the vulva. Here we discuss vulvar disorders that are relatively specific and

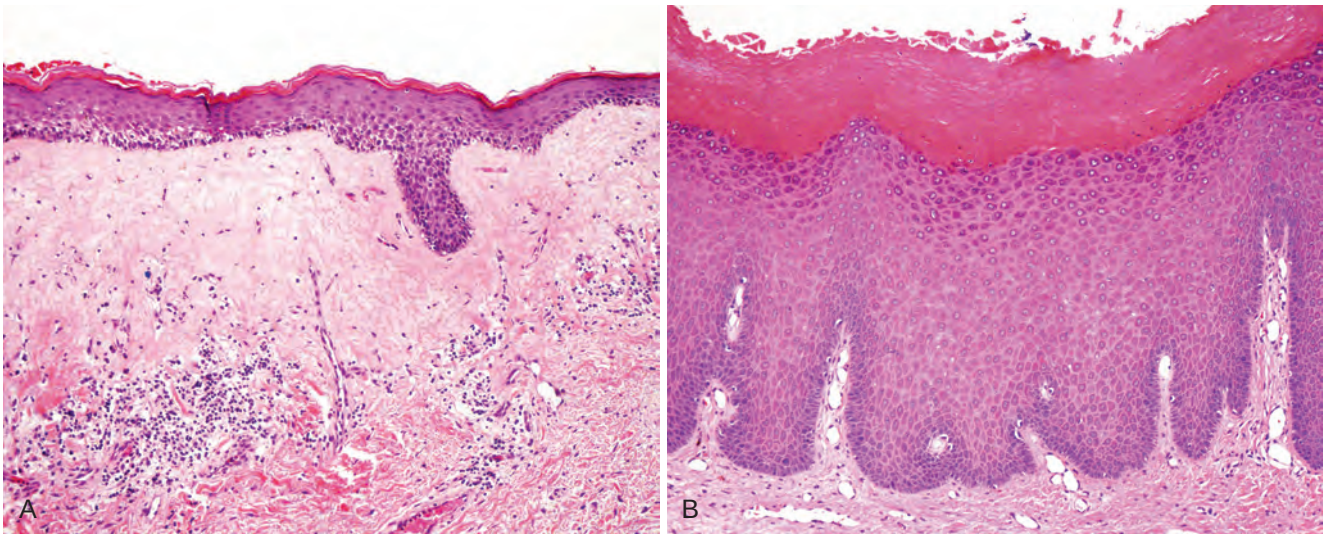


Figure 22.5 Non-neoplastic epithelial vulvar disorders. (A) Lichen sclerosus. There is marked thinning of the epidermis, sclerosis of the superficial dermis, and chronic inflammatory cells in deeper dermis. (B) Squamous cell hyperplasia, displaying thickened epidermis and hyperkeratosis.

common, including Bartholin cyst, non-neoplastic epithelial disorders, benign exophytic lesions, and tumors of the vulva.

BARTHOLIN CYST

Infection of the Bartholin gland produces an acute inflammation (adenitis) and may result in an abscess. Bartholin duct cysts are relatively common, occur at all ages, and result from obstruction of the duct by an inflammatory process. These cysts are usually lined by transitional or squamous epithelium. They may become large, up to 3 to 5 cm in diameter, and produce pain and local discomfort. Bartholin duct cysts are either excised or opened permanently (marsupialization).

NON-NEOPLASTIC EPITHELIAL DISORDERS

Leukoplakia is a descriptive clinical term for opaque, white, plaquelike epithelial thickening that may produce pruritus and scaling. Leukoplakia (literally, *white plaques*) may be caused by a variety of benign, premalignant, or malignant disorders, including the following:

- Inflammatory dermatoses (e.g., psoriasis, chronic dermatitis)
- Lichen sclerosus and squamous cell hyperplasia
- Neoplasias, such as vulvar intraepithelial neoplasia (VIN), Paget disease, and invasive carcinoma

Inflammatory dermatoses associated with leukoplakia are described in Chapter 25, while neoplastic disorders are discussed later in this chapter. Here the major non-neoplastic causes of leukoplakia—lichen sclerosis and squamous cell hyperplasia—are briefly discussed.

Lichen Sclerosus

Lichen sclerosis presents as smooth, white plaques or macules that in time may enlarge and coalesce, producing a surface that resembles porcelain or parchment. When the entire vulva is affected, the labia become atrophic and agglutinated, and the vaginal orifice constricts. Histologically, the lesion is characterized by marked thinning of the epidermis (Fig. 22.5A), degeneration of the basal epithelial cells, excessive keratinization (hyperkeratosis), sclerotic changes of the superficial dermis, and a bandlike lymphocytic infiltrate in the underlying dermis. The disease occurs in all age groups but is most common in postmenopausal women. It may also be encountered elsewhere on the skin. Its pathogenesis is uncertain, but the presence of activated T cells in the subepithelial inflammatory infiltrate and the increased frequency of autoimmune disorders in affected women suggest that an autoimmune reaction is involved. Although lichen sclerosus is not itself a premalignant lesion, women with symptomatic lichen sclerosus have a slightly increased chance of developing squamous cell carcinoma of the vulva.

Squamous Cell Hyperplasia

Previously called hyperplastic dystrophy or *lichen simplex chronicus,* squamous cell hyperplasia is a nonspecific condition resulting from rubbing or scratching of the skin to relieve pruritus. Clinically it presents as leukoplakia, and histologic examination reveals thickening of the epidermis (acanthosis) and hyperkeratosis (see Fig. 22.5B). Lymphocytic infiltration of the dermis is sometimes present. The hyperplastic epithelium may show mitotic activity but lacks cellular atypia. While squamous cell hyperplasia is not considered premalignant, it is sometimes present at the margins of vulvar cancers.

BENIGN EXOPHYTIC LESIONS

Benign raised (exophytic) or wartlike lesions of the vulva may be caused by infection or may be reactive conditions of unknown etiology. *Condyloma acuminatum,* a papillomavirus-induced lesion, also called a *genital wart,* and syphilitic *condyloma latum* (described in Chapter 21) are consequences

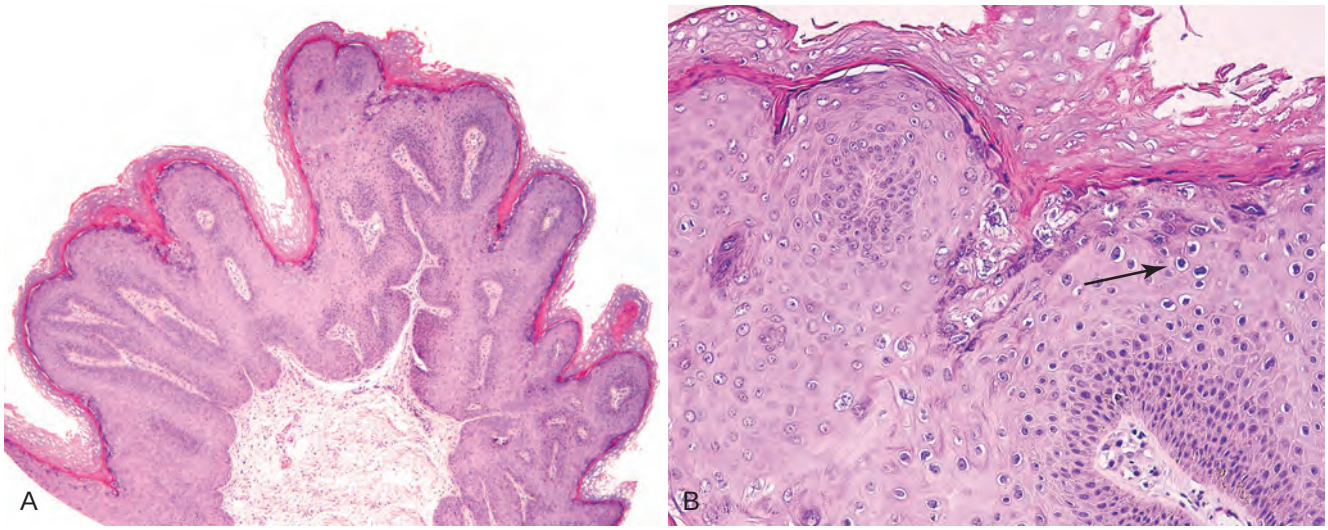


Figure 22.6 Condyloma acuminatum. (A) Low-power view showing exophytic, papillary architecture. (B) High-power view reveals human papillomavirus cytopathic effect (koilocytic atypia) characterized by atypical, enlarged, hyperchromatic nuclei with perinuclear halos *(arrow)*.

of sexually transmitted infections. Vulvar *fibroepithelial polyps,* or skin tags, are similar to skin tags occurring elsewhere on the skin. Vulvar *squamous papillomas* are benign exophytic proliferations covered by nonkeratinized squamous epithelium that develop on vulvar surfaces; they may be single or numerous (vulvar papillomatosis). The etiology of fibroepithelial polyps and squamous papillomas is unknown.

Condyloma Acuminatum

Condylomata acuminata are benign genital warts caused by low-risk HPV, mainly types 6 and 11. They may be solitary, but are more frequently multifocal, and they may involve vulvar, perineal, and perianal regions as well as the vagina and, less commonly, the cervix. The lesions are identical to those found on the penis and around the anus in males (Chapter 21). On histologic examination, they consist of papillary, exophytic, treelike cores of stroma covered by thickened squamous epithelium (Fig. 22.6A). The surface epithelium shows characteristic viral cytopathic changes referred to as *koilocytic atypia* (see Fig. 22.6B), which manifest as nuclear enlargement, hyperchromasia, and a cytoplasmic perinuclear halo (see the Cervix section later in this chapter). Condylomata acuminata are not precancerous lesions. HPV vaccines (described later) provide excellent protection against infection by low-risk HPV and genital warts.

SQUAMOUS NEOPLASTIC LESIONS

Vulvar Intraepithelial Neoplasia and Vulvar Carcinoma

Carcinoma of the vulva is an uncommon malignant neoplasm (approximately one-eighth as frequent as cervical cancer) representing about 3% of all genital cancers in the female; approximately two-thirds occur in women older than 60 years of age. Squamous cell carcinoma is the most common histologic type of vulvar cancer. In terms of etiology, pathogenesis, and histologic features, vulvar squamous cell carcinomas are divided into two groups:

- *Basaloid and warty carcinomas* are related to infection with high-risk HPVs, most commonly HPV-16. These are less common (30% of cases) and occur in younger women (average 60 years of age).
- *Keratinizing squamous cell carcinomas* are unrelated to HPV infection. These are more common (70% of cases) and occur in older women (average 75 years of age).

Basaloid and *warty carcinomas* develop from an in situ precursor lesion called *classic vulvar intraepithelial neoplasia* (VIN). This form of VIN occurs mainly in reproductive age women and includes lesions formerly designated as *carcinoma in situ* or *Bowen disease.* The risk factors for VIN are the same as those associated with cervical squamous intraepithelial lesions (e.g., young age at first intercourse, multiple sexual partners, male partner with multiple sexual partners), as both are related to HPV infection. VIN is frequently multicentric, and 10% to 30% of patients with VIN also have vaginal or cervical HPV-related lesions. Spontaneous regression of classic VIN has been reported, usually in younger women. The risk of progression to invasive carcinoma is higher in women who are older than 45 years of age or who are immunosuppressed.

Keratinizing squamous cell carcinoma occurs most often in individuals with long-standing lichen sclerosus or squamous cell hyperplasia and is not related to HPV. It arises from a precursor lesion referred to as *differentiated vulvar intraepithelial neoplasia* (differentiated VIN). It is postulated that chronic epithelial irritation in lichen sclerosus or squamous cell hyperplasia may contribute to a gradual evolution to the malignant phenotype, presumably through acquisition of driver mutations in oncogenes and tumor suppressor genes. In line with this idea, some investigators have reported a high frequency of *TP53* mutations in differentiated VIN.

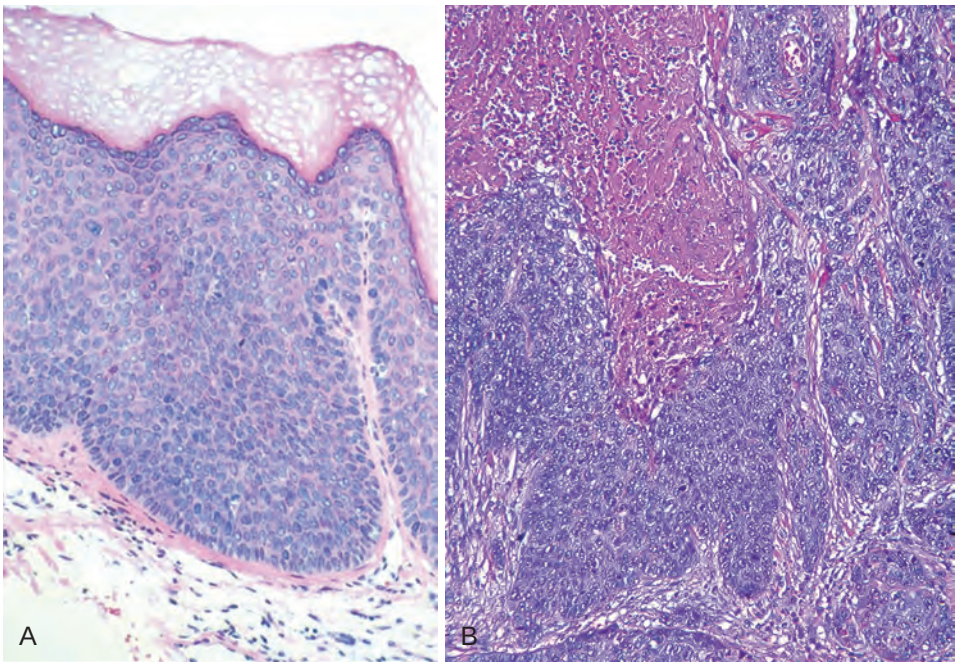


Figure 22.7 Human papillomavirus (HPV)-related vulvar pre-neoplastic and malignant lesions. (A) Classic vulvar intraepithelial neoplasia (HPV positive), showing small, immature basaloid cells encompassing full thickness of the epithelium. No invasion is present. (B) Basaloid vulvar carcinoma (HPV positive), composed of invasive small, immature (basaloid) cells. There is a focus of necrosis *(red area)*.

MORPHOLOGY

The lesions of **classic VIN** may be discrete and white (hyperkeratotic) or slightly raised and pigmented. Microscopically, it is characterized by epidermal thickening, nuclear atypia, increased mitoses, and lack of cellular maturation (Fig. 22.7A), features analogous to those seen in cervical squamous intraepithelial lesions (SIL, see the Cervix section later in this chapter). Invasive carcinomas that arise from classic VIN may be exophytic or indurated with central ulceration. On histologic examination, basaloid carcinoma (see Fig. 22.7B) consists of nests and cords of small, tightly packed cells that lack maturation and resemble the basal layer of the normal epithelium. The tumor may have foci of central necrosis. By contrast, warty carcinoma is characterized by exophytic, papillary architecture and prominent koilocytic atypia.

Differentiated VIN is characterized by marked atypia of the basal layer of the squamous epithelium and normal-appearing differentiation of the more superficial layers (Fig. 22.8A). Invasive keratinizing squamous cell carcinomas that arise in differentiated VIN contain nests and tongues of malignant squamous epithelium with prominent central keratin pearls (see Fig. 22.8B).

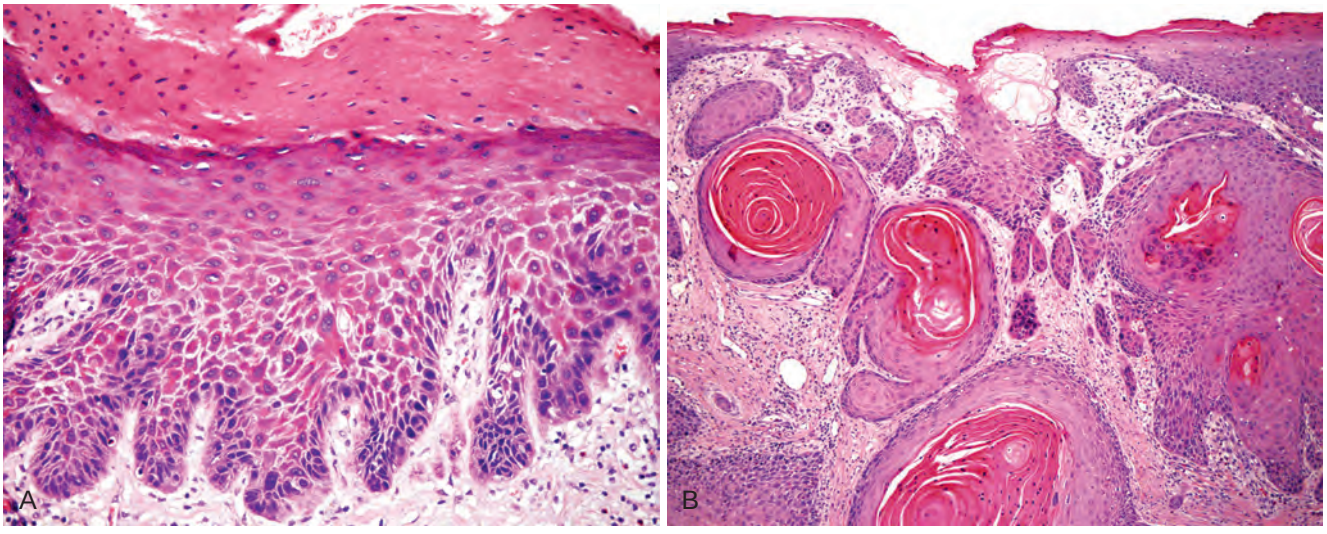


Figure 22.8 Non–human papillomavirus (HPV) vulvar pre-neoplastic and malignant lesions. (A) Differentiated vulvar intraepithelial neoplasia (HPV negative), showing maturation of the superficial layers, hyperkeratosis, and atypia of basal epithelial cells. No invasion is present. (B) Well-differentiated, keratinizing squamous cell carcinoma of the vulva (HPV negative) composed of invasive tumor nests with central keratin pearls.

The risk of cancer development in VIN depends on duration and extent of disease, and the immune status of the patient. Invasive carcinomas associated with lichen sclerosus, squamous cell hyperplasia, and differentiated VIN may develop in an insidious fashion and may be misinterpreted as dermatitis or leukoplakia for long periods. Once invasive cancer develops, the risk of metastatic spread is determined by the size of tumor, the depth of invasion, and whether there is lymphatic invasion. The initial spread is to inguinal, pelvic, iliac, and periaortic lymph nodes. Ultimately, hematogenous dissemination to the lungs, liver, and other internal organs may occur. Patients with lesions less than 2 cm in diameter have a 90% 5-year survival rate after treatment with vulvectomy and lymphadenectomy; however, larger lesions with lymph node involvement have a poor prognosis.

GLANDULAR NEOPLASTIC LESIONS

Like the breast, the vulva contains modified apocrine sweat glands. Presumably because of these "breastlike" features, the vulva may be involved by two tumors with counterparts in the breast, papillary hidradenoma and extramammary Paget disease.

Papillary Hidradenoma

Papillary hidradenoma presents as a sharply circumscribed nodule, most commonly on the labia majora or interlabial folds, and may be confused clinically with carcinoma because of its tendency to ulcerate. Its histologic appearance is identical to that of intraductal papilloma of the breast and consists of papillary projections covered by two cell layers, an upper layer of columnar secretory cells and a deeper layer of flattened myoepithelial cells. These myoepithelial elements are characteristic of sweat glands and sweat gland tumors (Fig. 22.9).

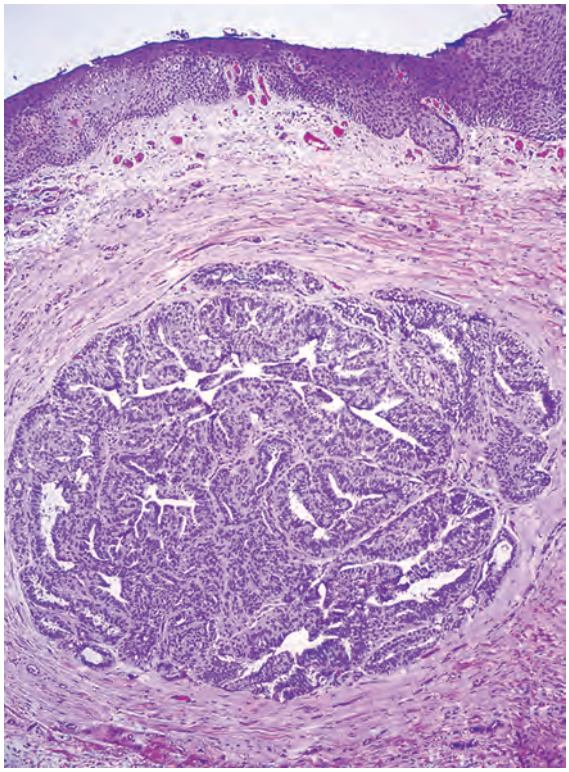


Figure 22.9 Papillary hidradenoma of the vulva, a well-circumscribed tumor composed of benign papillary projections covered with columnar secretory epithelium and underlying myoepithelial cells.

Extramammary Paget Disease

This curious and rare lesion of the vulva is similar in its manifestations to Paget disease of the breast (Chapter 23). In the vulva, it presents as a pruritic, red, crusted, maplike area, usually on the labia majora.

In contrast to Paget disease of the nipple, in which 100% of patients have an underlying ductal breast carcinoma, vulvar Paget is typically not associated with underlying cancer and is confined to the vulvar epidermis. Treatment consists of wide local excision, which is necessary because Paget cells spread laterally within the epidermis and may be present beyond the confines of the grossly visible lesion. Intra-epidermal Paget disease may persist for many years or even decades without invasion or metastases. In the rare instances when invasion develops, the prognosis is poor.

MORPHOLOGY

Paget disease is a distinctive intraepithelial proliferation of malignant cells. Paget cells are larger than surrounding keratinocytes and are seen singly or in small clusters within the epidermis (Fig. 22.10A). The cells have pale cytoplasm containing mucopolysaccharide that stains with periodic acid–Schiff (PAS), Alcian blue, or mucicarmine stains. In addition, unlike squamous epithelium, the cells express cytokeratin 7 (see Fig. 22.10B). Ultrastructurally, Paget cells display apocrine, eccrine, and keratinocyte differentiation and presumably arise from multipotent cells found within the mammary gland–like ducts of the vulvar skin.

KEY CONCEPTS

- Approximately 30% of vulvar cancers are caused by infection with high-risk HPVs, principally HPV-16. These cancers develop from an in situ lesion termed classic vulvar intraepithelial neoplasia (classic VIN).
- Most vulvar cancers (70%) are not related to HPV and develop in a background of lichen sclerosus or squamous cell hyperplasia from the premalignant lesion called differentiated vulvar intraepithelial neoplasia (differentiated VIN).

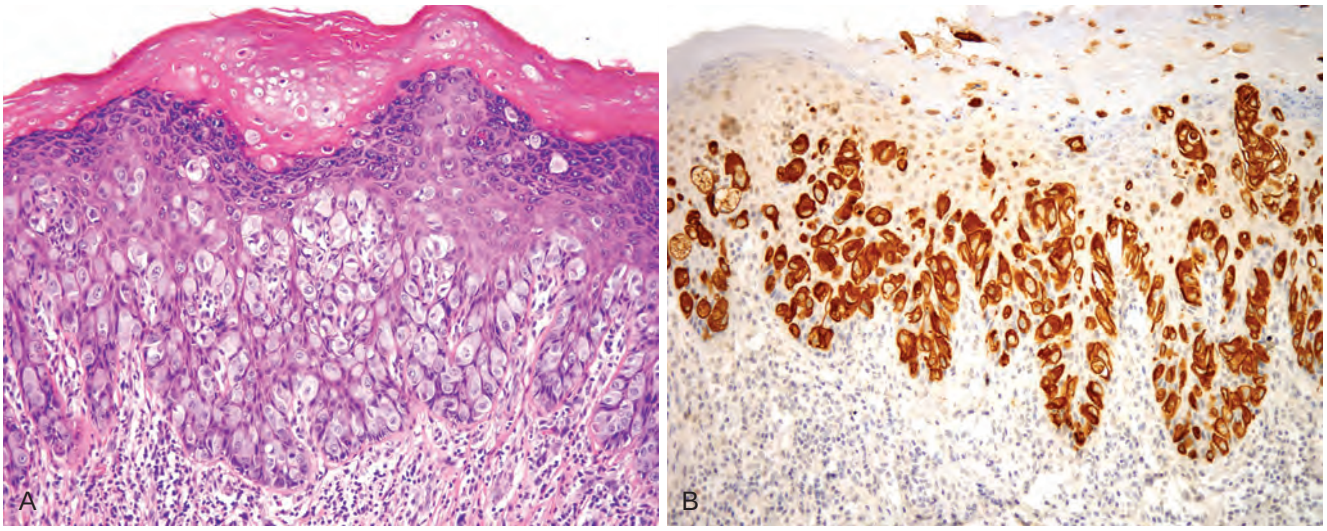


Figure 22.10 Paget disease of the vulva. (A) The epidermis is infiltrated by large cells with pale-pink cytoplasm that are spreading along the basal portion of the squamous epithelium. There is inflammation in the underlying dermis. (B) Immunostaining for cytokeratin 7 highlights the intraepidermal Paget cells.

Vagina

The vagina is a portion of the female genital tract that is remarkably resistant to diseases that affect nearby structures. For example, in the adult, inflammation that begins in the vulva and perivulvar structures often spreads to the cervix without significant involvement of the vagina. Primary lesions of the vagina are rare, the most serious being vaginal squamous cell carcinoma; they are discussed only briefly here.

DEVELOPMENTAL ANOMALIES

Septate, or double, vagina is an uncommon anomaly that arises from a failure of müllerian duct fusion and is accompanied by a double uterus *(uterus didelphys)*. These and other anomalies of the external genitalia may be the manifestations of genetic syndromes, in utero exposure to diethylstilbestrol (DES, used to prevent threatened abortions in the 1940s to 1960s), or other unknown factors that perturb reciprocal epithelial-stromal signaling during fetal development.

During embryonal development, the vagina is initially covered by columnar, endocervical-type epithelium. This is normally replaced by squamous epithelium advancing upward from the urogenital sinus. Small patches of residual glandular epithelium may persist into adult life; such areas are recognized as *vaginal adenosis,* which on examination presents as red, granular areas that stand out from the surrounding normal pale-pink vaginal mucosa. Microscopically, adenosis consists of columnar mucinous epithelium indistinguishable from endocervix. Adenosis is found in only a small percentage of adult women, but has been reported in 35% to 90% of women exposed to DES in utero. Rare cases of clear cell carcinoma arising in DES-related adenosis were recorded in teenagers and young adult women in the 1970s and 1980s, resulting in discontinuation of DES treatment.

Gartner duct cysts are relatively common lesions found along the lateral walls of the vagina that are derived from wolffian (mesonephric) duct rests. They consist of 1- to 2-cm fluid-filled submucosal cysts. Other cysts, including mucus cysts, which occur in the proximal vagina, are derived from müllerian epithelium. Another müllerian-derived lesion, endometriosis (described later), may occur in the vagina and clinically simulate a neoplasm.

PREMALIGNANT AND MALIGNANT NEOPLASMS OF THE VAGINA

Most benign tumors of the vagina occur in reproductive-age women and include stromal tumors (stromal polyps), leiomyomas, and hemangiomas. The most common malignant tumor to involve the vagina is carcinoma spreading from the cervix, followed by primary squamous cell carcinoma of the vagina. Infants may develop a unique, rare malignancy—embryonal rhabdomyosarcoma (sarcoma botryoides).

Vaginal Intraepithelial Neoplasia and Squamous Cell Carcinoma

Virtually all primary carcinomas of the vagina are squamous cell carcinomas associated with high-risk HPV infection. Vaginal carcinoma is extremely uncommon (about 0.6 per 100,000 women yearly) and accounts for about 1% of malignant neoplasms in the female genital tract. The greatest risk factor is a previous carcinoma of the cervix or vulva; 1% to 2% of women with an invasive cervical carcinoma eventually develop a vaginal squamous cell carcinoma. Squamous cell carcinoma of the vagina arises from a premalignant lesion, *vaginal intraepithelial neoplasia,* analogous to cervical squamous intraepithelial lesion (SIL; see the Cervix section later in this chapter). Most often the invasive tumor affects the upper vagina, particularly the posterior wall at the junction with the ectocervix. The lesions in the lower two-thirds of the vagina metastasize to the inguinal nodes, whereas lesions in the upper vagina tend to spread to regional iliac nodes.

Embryonal Rhabdomyosarcoma

Also called *sarcoma botryoides,* this uncommon vaginal tumor composed of malignant embryonal rhabdomyoblasts is most frequently found in infants and children younger than 5 years of age. These tumors tend to grow as polypoid, rounded, bulky masses that have the appearance and consistency of grapelike clusters (hence the designation *botryoides,* or grapelike) (Fig. 22.11). The tumor cells are small and have oval nuclei, with small protrusions of cytoplasm from one end, resembling a tennis racket. Rarely, striations (indicative of muscle differentiation) can be seen within the cytoplasm. Beneath the vaginal epithelium, the tumor cells are crowded in a so-called cambium layer, but in the deep regions they lie within a loose edematous fibromyxomatous stroma that may contain many inflammatory cells. Such lesions can be mistaken for benign inflammatory polyps. The tumors tend to invade locally and cause death by penetration into the peritoneal cavity or by obstruction of the urinary tract. Conservative surgery coupled with chemotherapy offer the best hope, particularly in cases diagnosed sufficiently early.

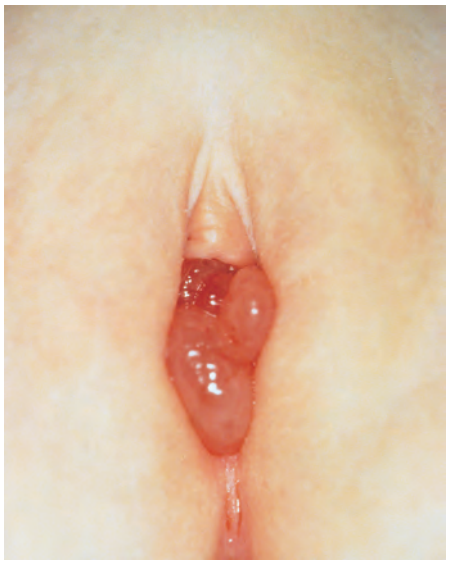


Figure 22.11 Sarcoma botryoides (embryonal rhabdomyosarcoma) of the vagina appearing as a polypoid mass protruding from the vagina. (Courtesy Dr. Michael Donovan, Children's Hospital, Boston, Mass.)

Cervix

Anatomically the cervix consists of the external vaginal portio (ectocervix) and the endocervical canal. The ectocervix is visible on vaginal examination and is covered by a mature squamous epithelium that is continuous with the vaginal wall. The squamous epithelium converges centrally at a small opening termed the *external os* that leads to the endocervical canal. The endocervix is lined by columnar, mucus-secreting epithelium. The point where the squamous and the columnar epithelium meet is referred to as the *squamocolumnar junction* (Fig. 22.12). The position of the junction is variable and changes with age and hormonal influence, but in general the junction moves upward into the endocervical canal with time. The replacement of the glandular epithelium by advancing squamous epithelium is a process called squamous metaplasia. The area of the cervix where the columnar epithelium coexists with the squamous epithelium is termed the "*transformation zone.*" The unique epithelial environment of the cervix renders it highly susceptible to infection with HPV, the main cause of cervical cancer. Immature squamous cells in the transformation zone are most susceptible to HPV infection, and as a result this is where the majority of cervical precursor lesions and cervical cancers develop.

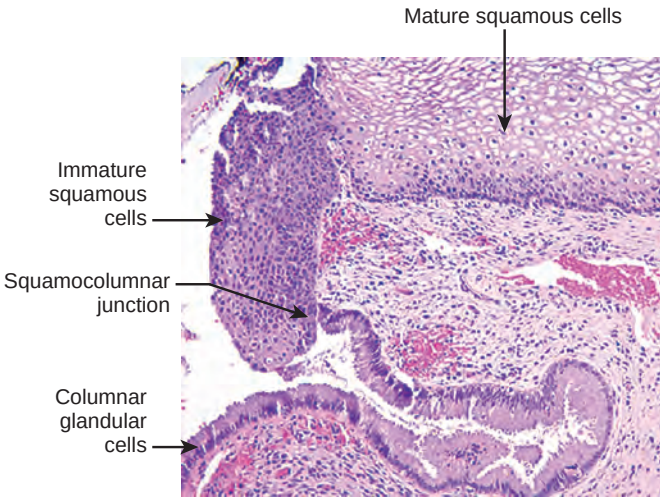


Figure 22.12 Cervical squamocolumnar junction showing a transition from mature, glycogenized squamous epithelium to columnar endocervical glandular epithelium. The superficial, mature squamous epithelial cells are not susceptible to human papillomavirus (HPV) infection. The HPV-susceptible cells include immature basal squamous cells and endocervical glandular cells.

INFLAMMATIONS

Acute and Chronic Cervicitis

At the onset of menarche, the production of estrogens by the ovary stimulates maturation of the cervical and vaginal squamous mucosa and formation of intracellular glycogen vacuoles in the squamous cells. As these cells are shed, the glycogen provides a substrate for various endogenous vaginal aerobes and anaerobes, but particularly lactobacilli, which are the dominant microbial species in the normal vagina. Lactobacilli produce lactic acid, which maintains the vaginal pH below 4.5, suppressing the growth of other saprophytic and pathogenic organisms. In addition, at low pH, lactobacilli produce bacteriotoxic hydrogen peroxide (H_2O_2). If the pH becomes alkaline due to bleeding, sexual intercourse, or vaginal douching, H_2O_2 production by lactobacilli decreases. Antibiotic therapy that suppresses lactobacilli can also cause the pH to rise. In each of these settings, the altered vaginal environment promotes the overgrowth

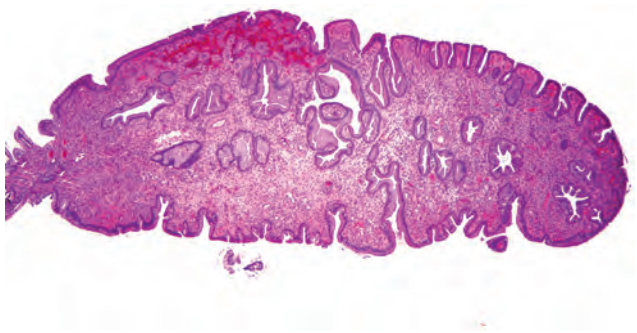


Figure 22.13 Endocervical polyp composed of a dense fibrous stroma covered with endocervical columnar epithelium.

of other microorganisms, which may result in cervicitis or vaginitis. Some degree of cervical inflammation may be found in virtually all women, and it is usually of little clinical consequence. However, infections by gonococci, chlamydiae, mycoplasmas, and HSV may produce significant acute or chronic cervicitis and are important to identify due to their association with upper genital tract disease, complications during pregnancy, and sexual transmission. Marked cervical inflammation produces reparative and reactive changes of the epithelium and shedding of atypical-appearing squamous cells, and therefore may cause an abnormal Pap test result.

ENDOCERVICAL POLYPS

Endocervical polyps are common benign exophytic growths that arise within the endocervical canal. They vary from small, sessile "bumps" to large polypoid masses that may protrude through the cervical os. Histologically, they are composed of a fibrous stroma covered by mucus-secreting endocervical glands, often accompanied by inflammation (Fig. 22.13). Their main significance is that they may be the source of irregular vaginal "spotting" or bleeding that arouses suspicion of some more ominous lesion. Simple curettage or surgical excision is curative.

PREMALIGNANT AND MALIGNANT NEOPLASMS OF THE CERVIX

Worldwide, cervical carcinoma is the fourth most common cancer in women, with an estimated 570,000 new cases in 2018, of which more than one-half will prove fatal. In the United States, it was estimated in 2019 that 13,179 women would be diagnosed with cervical cancer and that 4250 women would die of the disease. Fifty years ago, carcinoma of the cervix was the leading cause of cancer death in women in the United States, but the death rate has declined by 75% to its present rank as the thirteenth cause of cancer mortality. No form of cancer better documents the remarkable benefits of effective screening, early diagnosis, and curative therapy than does cancer of the cervix. Most credit for these dramatic gains belongs to the effectiveness of the Pap test in detecting cervical precursor lesions, some of which would have progressed to cancer if not treated; in addition, the Pap test can also detect low-stage, highly curable cancers. The accessibility of the cervix to Pap testing and visual examination (colposcopy) as well as the slow progression from precursor lesions to invasive carcinoma (typically over the course of decades) provides ample time for screening, detection, and preventive treatment.

Pathogenesis

High-risk HPVs are by far the most important factor in the development of cervical cancer. HPVs are DNA viruses that are grouped into those of high and low oncogenic risk based on their genotypes. There are 15 high-risk HPVs that are currently identified, but HPV-16 alone accounts for almost 60% of cervical cancer cases, and HPV-18 accounts for another 10% of cases; other HPV types contribute to less than 5% of cases, individually. High-risk HPVs are also implicated in squamous cell carcinomas arising at many other sites, including the vagina, vulva, penis, anus, tonsil, and other oropharyngeal locations. As noted earlier, low oncogenic risk HPVs are the cause of sexually transmitted vulvar, perineal, and perianal warts (condyloma acuminatum).

Genital HPV infections are extremely common; most of them are asymptomatic, do not cause any tissue changes, and therefore are not detected on Pap test. The prevalence of HPV in cervical smears in women with normal Pap test results peaks between 20 and 24 years of age, around the onset of sexual activity, and subsequently declines as protective immunity is established and women enter into monogamous relationships. Most HPV infections are transient; overall, 50% of HPV infections are cleared within 8 months, and 90% of infections are cleared within 2 years. The duration of the infection is related to HPV type, as infections with high-risk HPVs take longer to clear on average than infections with low-risk HPVs (13 months vs. 8 months, respectively). Persistent infection increases the risk of the development of cervical precursor lesions and subsequent carcinoma.

Productive, persistent HPV infection requires viral entry into immature basal epithelial cells. As a result, surfaces covered with mature, intact squamous epithelium, such as the ectocervix, vagina, vulva, penis, and oropharynx, are normally resistant to HPV infection. Sites in the female genital tract that are susceptible to infection include areas of squamous epithelial trauma and repair, where the virus may access basal cells, and the immature metaplastic squamous cells that are present at the squamocolumnar junction of the cervix (see Fig. 22.12). The cervix, with its relatively large areas of immature squamous metaplastic epithelium, is particularly vulnerable to HPV infection. Other sites in the body that are vulnerable to HPV infection include the squamocolumnar junction of the anus and the squamous cells of oropharyngeal tonsillar crypts, both relatively common sites of HPV-associated cancers in individuals who practice anal or oral sex, respectively.

The ability of HPV to act as a carcinogen depends on the viral E6 and E7 proteins, which interfere with the activity of the key tumor suppressor proteins, p53 and RB, respectively. Although HPV infects immature squamous cells, viral replication occurs in maturing squamous cells. Normally, these more mature cells are arrested in the G_1 phase of the cell cycle, but they continue to actively progress through the cell cycle when infected with HPV, which uses

the host cell DNA synthesis machinery to replicate its own genome. As you will recall from Chapter 7, viral E7 protein binds the hypophosphorylated (active) form of RB and promotes its degradation via the proteasome pathway, and also binds and inhibits p21 and p27, two important cyclin-dependent kinase inhibitors. Removal of these controls not only enhances cell cycle progression, but also impairs the ability of cells to repair DNA damage. The DNA repair defect in infected cells is exacerbated by E6 proteins encoded by high-risk HPV subtypes, which bind p53 and promote its degradation by the proteasome. In addition, E6 up-regulates the expression of telomerase, which leads to cellular immortalization. The net effect is increased proliferation of cells that are prone to acquire additional mutations that may lead to cancer development. By contrast to high-risk HPVs, the E7 proteins of low-risk HPVs bind RB with lower affinity, while the E6 proteins of low-risk HPVs fail to bind p53 altogether and instead appear to dysregulate growth and survival by interfering with the Notch signaling pathway.

Another factor that contributes to malignant transformation by HPV is the physical state of the virus. The viral DNA is integrated into the host cell genome in most cancers. This configuration increases the expression of *E6* and *E7* genes, and may also dysregulate oncogenes near the sites of viral insertion, such as *MYC*. By contrast, viral DNA is extrachromosomal (episomal) in precursor lesions associated with high-risk HPVs and in condylomata associated with low-risk HPVs.

Even though HPV is firmly established as the major cause of cervical cancer, it is worth remembering that although a high percentage of young women are infected with one or more HPV types during their reproductive years, only a few develop cancer. Thus, other factors, such as exposure to co-carcinogens and host immune status, influence whether an HPV infection regresses or persists and eventually leads to cancer.

Cervical Intraepithelial Neoplasia (Squamous Intraepithelial Lesions)

The classification of cervical precursor lesions has evolved over time, and the terms from the various classification systems are used interchangeably. Hence a brief review of the terminology is warranted. The oldest classification system grouped lesions as having mild dysplasia on one end and severe dysplasia/carcinoma in situ on the other. This was followed by the *cervical intraepithelial neoplasia* (CIN) classification, with mild dysplasia termed *CIN I,* moderate dysplasia termed *CIN II,* and severe dysplasia termed *CIN III*. Because the decision with regard to patient management is two-tiered (observation vs. surgical treatment), the three-tier classification system has been recently simplified to a two-tiered system, with CIN I renamed low-grade squamous intraepithelial lesion (LSIL) and CIN II and CIN III combined into one category referred to as high-grade squamous intraepithelial lesion (HSIL) (Table 22.1).

Table 22.1 Classification Systems for Squamous Cervical Precursor Lesions

Dysplasia/Carcinoma in situ	CIN	SIL, Current Classification
Mild dysplasia	CIN I	Low-grade SIL (LSIL)
Moderate dysplasia	CIN II	High-grade SIL (HSIL)
Severe dysplasia	CIN III	High-grade SIL (HSIL)
Carcinoma in situ	CIN III	High-grade SIL (HSIL)

CIN, Cervical intraepithelial neoplasia; *SIL,* squamous intraepithelial lesion.

LSIL does not progress directly to invasive carcinoma, and, in fact, most cases regress spontaneously; only a small percentage progress to HSIL. LSIL represents a productive HPV infection in which there is a high level of viral replication, but only mild alterations in the growth of host cells. For these reasons, LSIL is not treated like a premalignant lesion. LSIL is approximately ten times more common than HSIL.

By contrast to LSIL, HSIL is considered to be at high risk for progression to carcinoma. In HSIL, there is a progressive deregulation of the cell cycle by HPV, which results in increased cellular proliferation, decreased or arrested epithelial maturation, and a lower rate of viral replication as compared with LSIL. Derangement of the cell cycle in HSIL may become irreversible and lead to a fully transformed malignant phenotype.

MORPHOLOGY

The diagnosis of SIL is based on identification of nuclear atypia characterized by nuclear enlargement, hyperchromasia (dark staining), coarse chromatin granules, and variation in nuclear size and shape (Fig. 22.14). The nuclear changes are often accompanied by cytoplasmic "halos." At an ultrastructural level, these "halos" consist of perinuclear vacuoles, a cytopathic change created in part by an HPV-encoded protein E5 that localizes to the membranes of the endoplasmic reticulum. Nuclear alterations associated with perinuclear halos are termed **koilocytic atypia.** The grading of SIL into low or high grade is based on expansion of the immature cell layer from its normal, basal location. If the immature squamous cells are confined to the lower one-third of the epithelium, the lesion is called LSIL; if they expand to the upper two-thirds of the epithelial thickness, it is called HSIL.

The histologic features of LSIL correlate with HPV replication and changes in host cell growth and gene expression (Fig. 22.15). The highest viral loads (assessed by in situ hybridization for HPV DNA; see Fig. 22.15B) are found in maturing keratinocytes in the upper half of the epithelium. HPV E6 and E7 proteins prevent cell cycle arrest. As a result, cells in the upper portion of the epithelium express markers of actively dividing cells, such as Ki-67 (see Fig. 22.15C), that are normally confined to the basal layer of the epithelium. Disturbed growth regulation also leads to overexpression of p16, a cyclin-dependent kinase inhibitor (see Fig. 22.15D). Both Ki-67 and p16 staining are highly correlated with HPV infection and are useful for confirmation of the diagnosis in equivocal cases of SIL.

More than 80% of LSILs and 100% of HSILs are associated with high-risk HPVs, with HPV-16 being the most

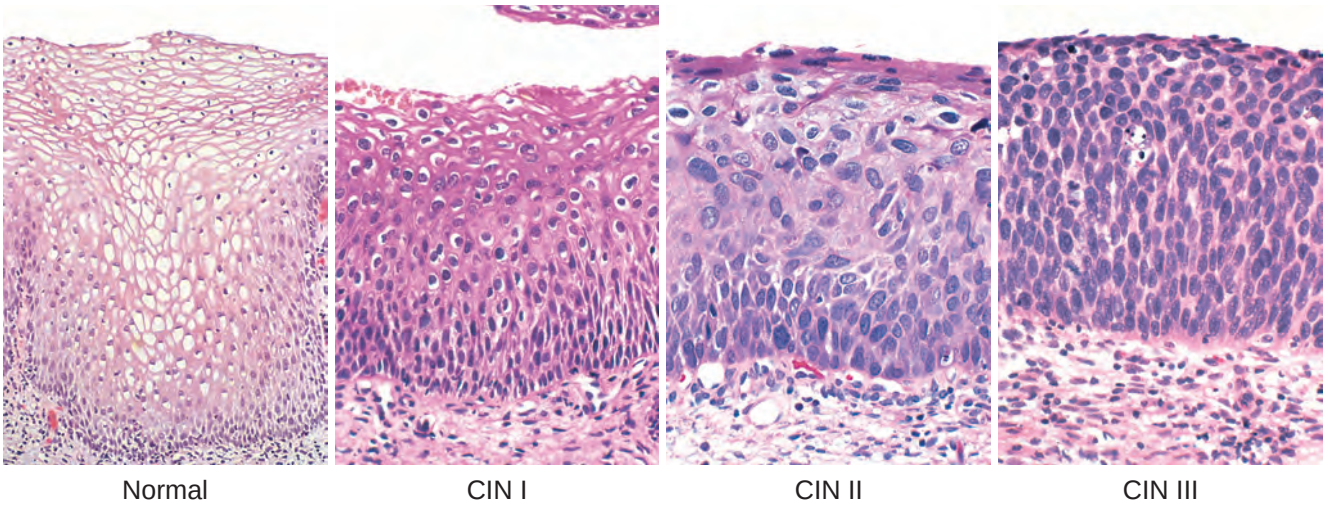


Figure 22.14 Spectrum of cervical intraepithelial neoplasia: normal squamous epithelium for comparison; low-grade squamous intraepithelial lesion (cervical intraepithelial neoplasia *[CIN]* I) with koilocytic atypia; high-grade squamous intraepithelial lesion (HSIL) (CIN II) with progressive atypia and expansion of the immature basal cells above the lower third of the epithelial thickness; HSIL (CIN III) with diffuse atypia, loss of maturation, and expansion of the immature basal cells to the epithelial surface.

common HPV type in both lesions. Table 22.2 shows rates of regression and progression of SILs within a 2-year follow-up. Although the majority of HSILs develop from LSILs, approximately 20% of cases of HSIL develop de novo, independent of any preexisting LSIL. The rates of progression are by no means uniform, and although HPV type—especially HPV 16—is associated with increased risk, it is difficult to predict the outcome in an individual patient. These findings underscore that the risk of precursor lesions and cancer is conferred only in part by HPV type. Progression to invasive carcinoma, when it occurs, on average takes place over several decades.

Cervical Carcinoma

The average age of patients with invasive cervical carcinoma is between 45 and 50 years. Squamous cell carcinoma is the most common histologic subtype, accounting for approximately 80% of cases. The second most common tumor type is adenocarcinoma, which constitutes about 15% of cervical cancer cases and develops from a precursor lesion called *adenocarcinoma in situ*. Adenosquamous and neuroendocrine carcinomas are rare cervical tumors that account for the remaining 5% of cases. All of the aforementioned tumor types are caused by high-risk HPVs. The progression time

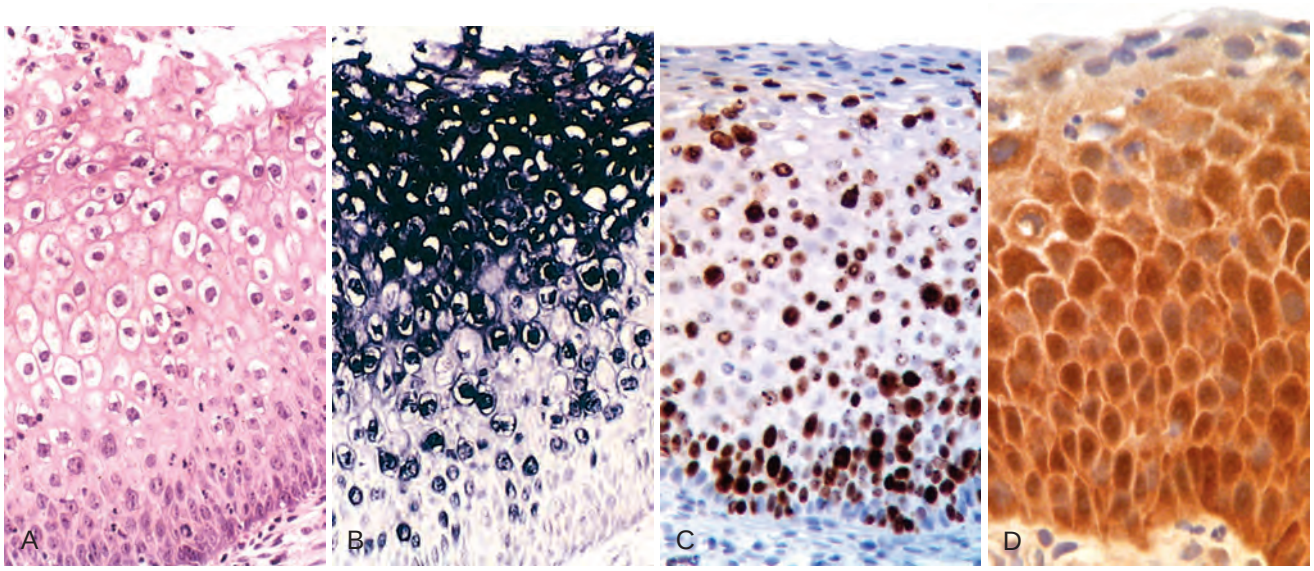


Figure 22.15 (A) Low-grade squamous intraepithelial lesion: routine hematoxylin and eosin staining shows koilocytic change, seen as nuclear enlargement and perinuclear "halos." (B) In situ hybridization test for human papillomavirus deoxyribonucleic acid (HPV DNA). The dark granular staining denotes HPV DNA, which is most abundant in the superficial koilocytes. (C) Diffuse positivity for the proliferation marker Ki-67 (seen as brown nuclear staining), illustrates expansion of the proliferating cells from the normal basal location to the superficial layers of the epithelium. (D) Up-regulation of the cyclin-dependent kinase inhibitor p16 (seen here as brown staining) characterizes high-risk HPV infections.

Table 22.2 Natural History of Squamous Intraepithelial Lesions With Approximate 2-Year Follow-Up

Lesion	Regress	Persist	Progress
LSIL	60%	30%	10% to HSIL
HSIL	30%	60%	10% to carcinoma[a]

HSIL, High-grade squamous intraepithelial lesion; *LSIL*, low-grade squamous intraepithelial lesion.

[a]Progression within 2 to 10 years.

from in situ to invasive adenosquamous and neuroendocrine carcinomas is shorter than in squamous cell carcinoma, and patients with these tumors often present with advanced disease and have a less favorable prognosis.

MORPHOLOGY

Invasive cervical carcinoma may manifest as fungating (exophytic) or infiltrative masses. **Squamous cell carcinoma** is composed of nests and tongues of malignant squamous epithelium, either keratinizing or nonkeratinizing, which invade the underlying cervical stroma (Fig. 22.16A–B). **Adenocarcinoma** is characterized by proliferation of glandular epithelium composed of malignant endocervical cells with large, hyperchromatic nuclei and relatively mucin-depleted cytoplasm, resulting in a dark appearance of the glands, as compared with the normal endocervical epithelium

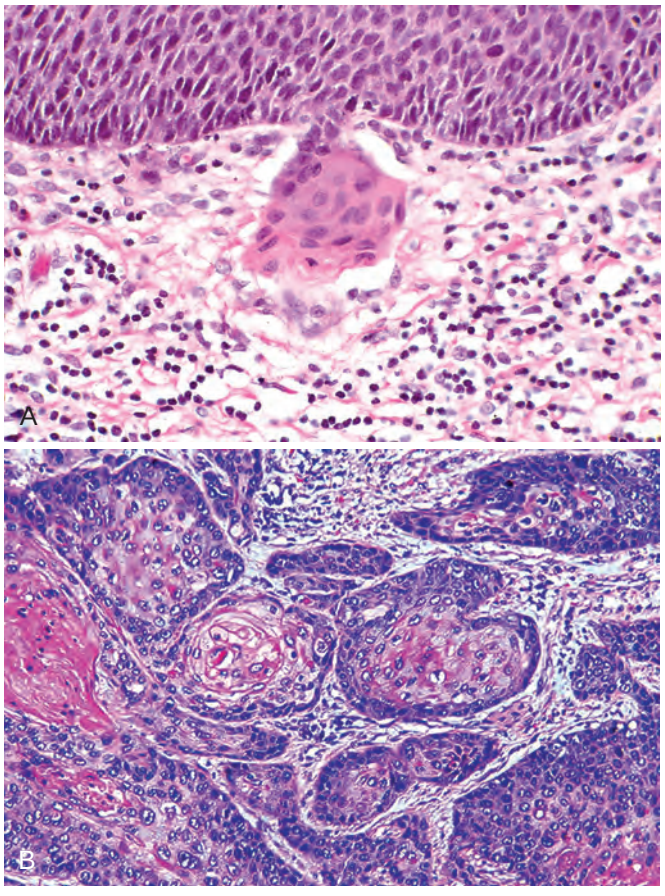


Figure 22.16 Squamous cell carcinoma of the cervix. (A) Early invasion in squamous cell carcinoma showing an invasive nest breaking through the basement membrane of a high-grade squamous intraepithelial lesion. (B) Invasive squamous cell carcinoma.

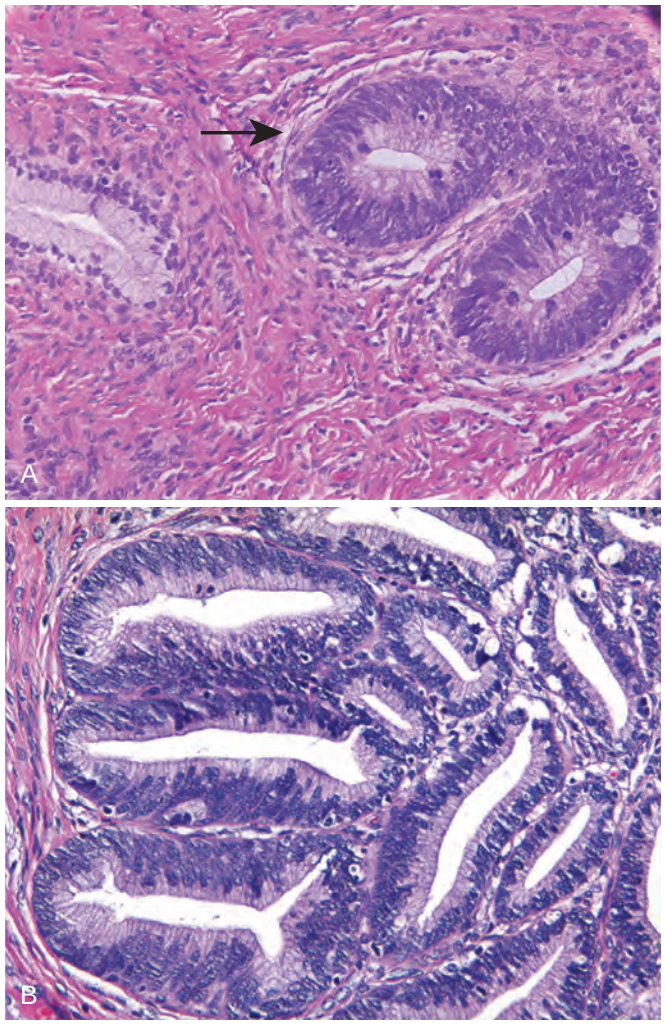


Figure 22.17 Adenocarcinoma of the cervix. (A) Adenocarcinoma in situ *(arrow)* showing dark glands with atypical, enlarged nuclei adjacent to a normal, pale endocervical gland. (B) Invasive adenocarcinoma.

(Fig. 22.17A–B). Adenosquamous carcinoma is composed of intermixed malignant glandular and squamous epithelium. Neuroendocrine cervical carcinoma has an appearance similar to small cell carcinoma of the lung (Chapter 15) but differs in being positive for high-risk HPVs.

Advanced cervical carcinoma spreads by direct extension to contiguous tissues, including paracervical soft tissue, urinary bladder, ureters (resulting in hydronephrosis), rectum, and vagina. Lymphovascular invasion results in local and distant lymph nodes metastases. Distant metastases may also be found in the liver, lungs, bone marrow, and other organs.

Cervical cancer is staged as follows:

Stage 0—Carcinoma in situ (CIN III, HSIL)

Stage I—Carcinoma confined to the cervix

- Ia—Preclinical carcinoma, that is, diagnosed only by microscopy
- Ia1—Stromal invasion no deeper than 3 mm and no wider than 7 mm (so-called superficially invasive squamous cell carcinoma)
- Ia2—Maximum depth of invasion of stroma deeper than 3 mm and no deeper than 5 mm taken from base of epithelium; horizontal invasion no more than 7 mm

Ib—Histologically invasive carcinoma confined to the cervix and greater than stage Ia2

Stage II—Carcinoma extends beyond the cervix but not to the pelvic wall. Carcinoma involves the vagina but not the lower third.

Stage III—Carcinoma has extended to the pelvic wall. On rectal examination, there is no cancer-free space between the tumor and the pelvic wall. The tumor involves the lower third of the vagina.

Stage IV—Carcinoma has extended beyond the true pelvis or has involved the mucosa of the bladder or rectum. This stage also includes cancers with metastatic dissemination.

Clinical Features

More than one-half of invasive cervical cancers are detected in women who did not participate in regular screening. While superficially invasive squamous cell carcinomas may be treated by cervical cone excision alone, most invasive cancers are managed by hysterectomy with lymph node dissection and, for advanced lesions, radiation and chemotherapy. The prognosis for invasive carcinomas depends on the stage of the cancer at diagnosis and to some degree on histologic subtype, with small-cell neuroendocrine tumors having a very poor prognosis. With current treatments, the 5-year survival rate is 100% for superficially invasive squamous cell carcinomas and less than 20% for tumors extending beyond the pelvis. Most patients with advanced cervical cancer die of the consequences of local tumor invasion (e.g., ureteral obstruction, pyelonephritis, and uremia) rather than distant metastases.

CERVICAL CANCER SCREENING AND PREVENTION

As is well known, cytologic cancer screening has significantly reduced mortality from cervical cancer. In countries where such screening is not widely practiced, cervical cancer continues to exact a high toll. The reason that cytologic screening is so effective in preventing cervical cancer is that most cancers arise from precursor lesions over the course of years. These lesions shed abnormal cells that can be detected on cytologic examination. Using a spatula or brush, the transformation zone of the cervix is circumferentially scraped and the cells are smeared or spun down onto a slide. Following fixation and staining with the Papanicolaou method, the smears are screened microscopically by eye or (increasingly) with automated image analysis systems. The cellular changes seen on the Pap test, illustrating the spectrum from LSIL to HSIL, are shown in Fig. 22.18.

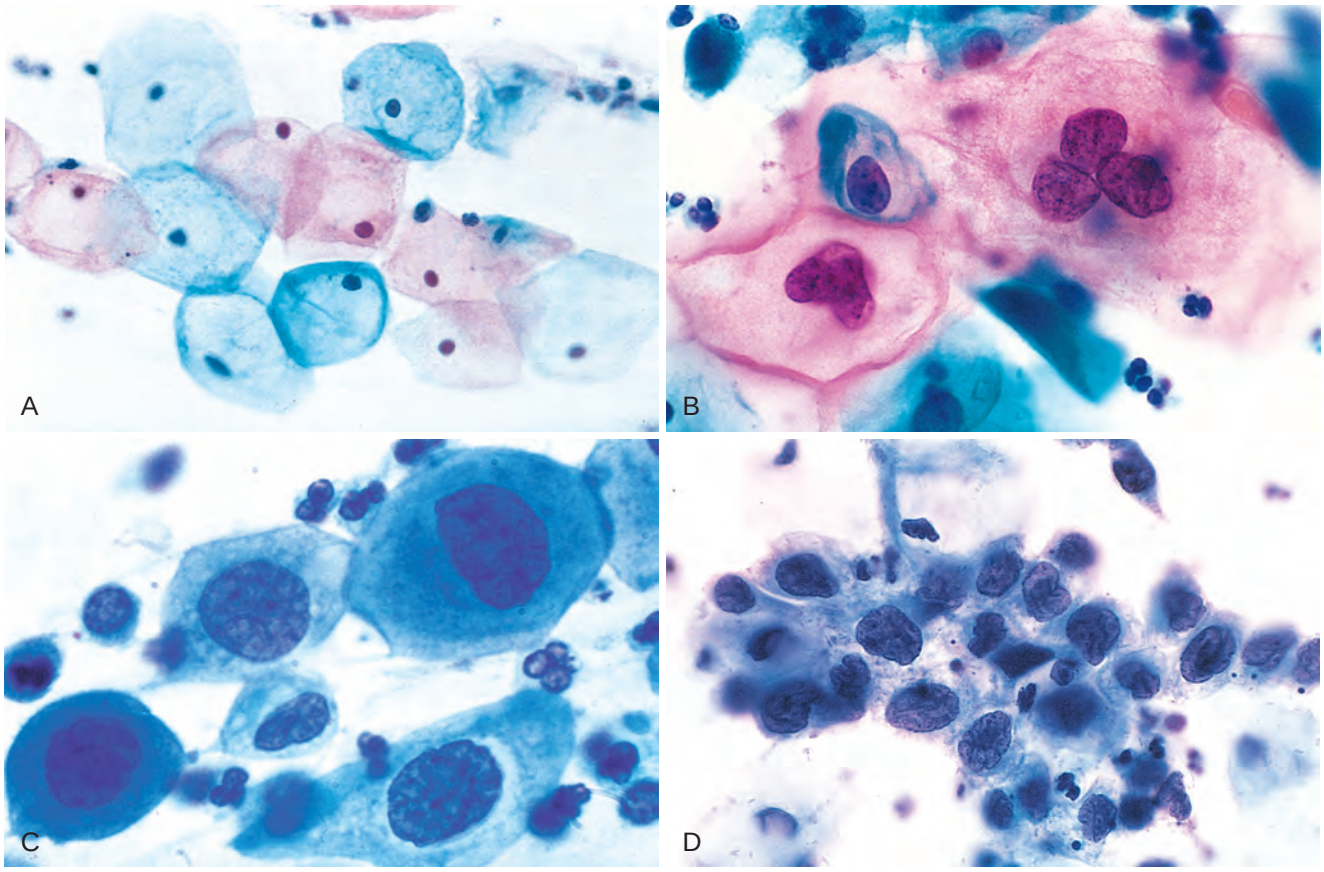


Figure 22.18 The cytology of cervical intraepithelial neoplasia as seen on the Papanicolaou smear. The cytoplasmic staining of desquamated cells may be either red or blue. (A) Normal exfoliated superficial squamous cells. (B) Low-grade squamous intraepithelial lesion—koilocytes. (C) High-grade squamous intraepithelial lesion (HSIL; cervical intraepithelial neoplasia [CIN] II). (D) HSIL (CIN III). Note the reduction in cytoplasm and the increase in the nucleus-to-cytoplasm ratio, which occurs as the grade of the lesion increases. This reflects the progressive loss of cellular differentiation on the surface of the lesions from which these cells are exfoliated. (Courtesy Dr. Edmund S. Cibas, Brigham and Women's Hospital, Boston, Mass.)

Testing for the presence of HPV DNA in the cervical scrape is a molecular method of cervical cancer screening. HPV testing has a higher sensitivity but lower specificity, as compared to the Pap test. HPV DNA testing may be added to cervical cytology for screening in women 30 years of age or older. HPV testing of women younger than 30 years of age is not recommended because of the high incidence of infection, and thus the particularly low specificity of HPV test results in this age group.

Cervical cancer screening and preventive measures are carried out in a stepwise fashion. Recommendations for the frequency of Pap screening vary, but in general the first smear should be at 21 years of age or within 3 years of onset of sexual activity, and thereafter every 3 years. After 30 years of age, women who have had normal cytology results and are negative for HPV may be screened every 5 years. Women who have a normal cytology result but test positive for high-risk HPV DNA should have cervical cytology repeated every 6 to 12 months.

When the result of a Pap test is abnormal, a colposcopic examination of the cervix and vagina is performed to identify the lesion. The mucosa is examined with a magnifying glass following application of acetic acid, which highlights abnormal epithelium as white spots *(aceto-white areas)*. Abnormal appearing areas are biopsied. Women with biopsy-confirmed LSIL can be followed in a conservative fashion. Some gynecologists will perform local ablation (e.g., cryotherapy) of LSIL, particularly if there is concern about the reliability of patient follow-up. HSIL is treated with cervical conization (superficial excision).

An additional important aspect of cervical cancer prevention is vaccination against high-risk oncogenic HPV, which is now recommended for all girls and boys by 11 to 12 years of age, as well as young men and women up to 26 years of age. Two HPV vaccines are now FDA-licensed. Both provide nearly complete protection against high-risk oncogenic HPV types 16 and 18 (together accounting for approximately 70% of cervical cancers), and one also provides protection against additional five high oncogenic risk HPVs as well as two low oncogenic risk HPV types 6 and 11, which are responsible for genital warts. Vaccination is now recommended for boys as well as girls due to the role that males play in the spread of HPV to women and the toll that HPV-related anal and oropharyngeal cancers take in men. The vaccines offer protection for up to 10 years; longer follow-up studies are still pending. Because the HPV vaccines do not protect against all high-risk HPV types, current guidelines recommend that cervical cancer screening be continued as in the past.

KEY CONCEPTS

- Cervical LSIL is a productive HPV infection that usually regresses spontaneously but occasionally progresses to HSIL.
- HSIL is characterized by progressive deregulation of the cell cycle and increasing cellular atypia. HSIL may progress to invasive carcinoma.
- Almost all cervical precursor lesions and cervical carcinomas are caused by high-risk HPV types, most commonly HPV-16.

Body of Uterus and Endometrium

The uterus has two major components: the myometrium and the endometrium. The myometrium is composed of tightly interwoven bundles of smooth muscle that form the wall of the uterus. The internal cavity of the uterus is lined by the endometrium, which is composed of glands embedded in a cellular stroma. The uterus is affected by a variety of disorders, the most common of which results from endocrine imbalances, complications of pregnancy, and neoplastic proliferation.

ENDOMETRIAL HISTOLOGY IN THE MENSTRUAL CYCLE

The endometrium undergoes dynamic physiologic and morphologic changes during the menstrual cycle in response to sex steroid hormones coordinately produced in the ovary. The ovary is influenced by hormones produced by the pituitary gland due to signals from the hypothalamus. Together, hypothalamic, pituitary, and ovarian factors and their interactions regulate maturation of ovarian follicles, ovulation, and menstruation.

The histologic appearance of the endometrium may be used to assess hormonal status, document ovulation, and determine causes of endometrial bleeding and infertility (Fig. 22.19). Progression through a normal menstrual cycle is correlated with the following histologic features:

- The cycle commences with *menses,* during which the superficial portion of the endometrium, referred to as the functionalis, is shed.
- The *proliferative phase* is marked by rapid growth of glands and stroma arising from the deeper portion of the endometrium (basalis). During the proliferative phase, the glands are straight, tubular structures lined by regular, tall, pseudostratified columnar cells. Mitotic figures are numerous, and there is no evidence of mucus secretion or vacuolation. The endometrial stroma is composed of spindle cells with scant cytoplasm that are also actively proliferating (see Fig. 22.19A).
- At *ovulation,* endometrial proliferation ceases and differentiation commences in response to the effects of progesterone made by the corpus luteum in the ovary.
- *Postovulation* is initially marked by the appearance of *secretory vacuoles* beneath the nuclei in the glandular epithelium (see Fig. 22.19B). Secretory activity is most prominent during the third week of the menstrual cycle, when the basal vacuoles progressively move to the apical surface. By the fourth week, the glands are tortuous, producing a serrated appearance. This serrated or "sawtooth" appearance is accentuated by secretory exhaustion and shrinkage of the glands.
- *Stromal changes in the late secretory phase,* due predominantly to progesterone, are the most significant features. Prominent spiral arterioles appear accompanied by an

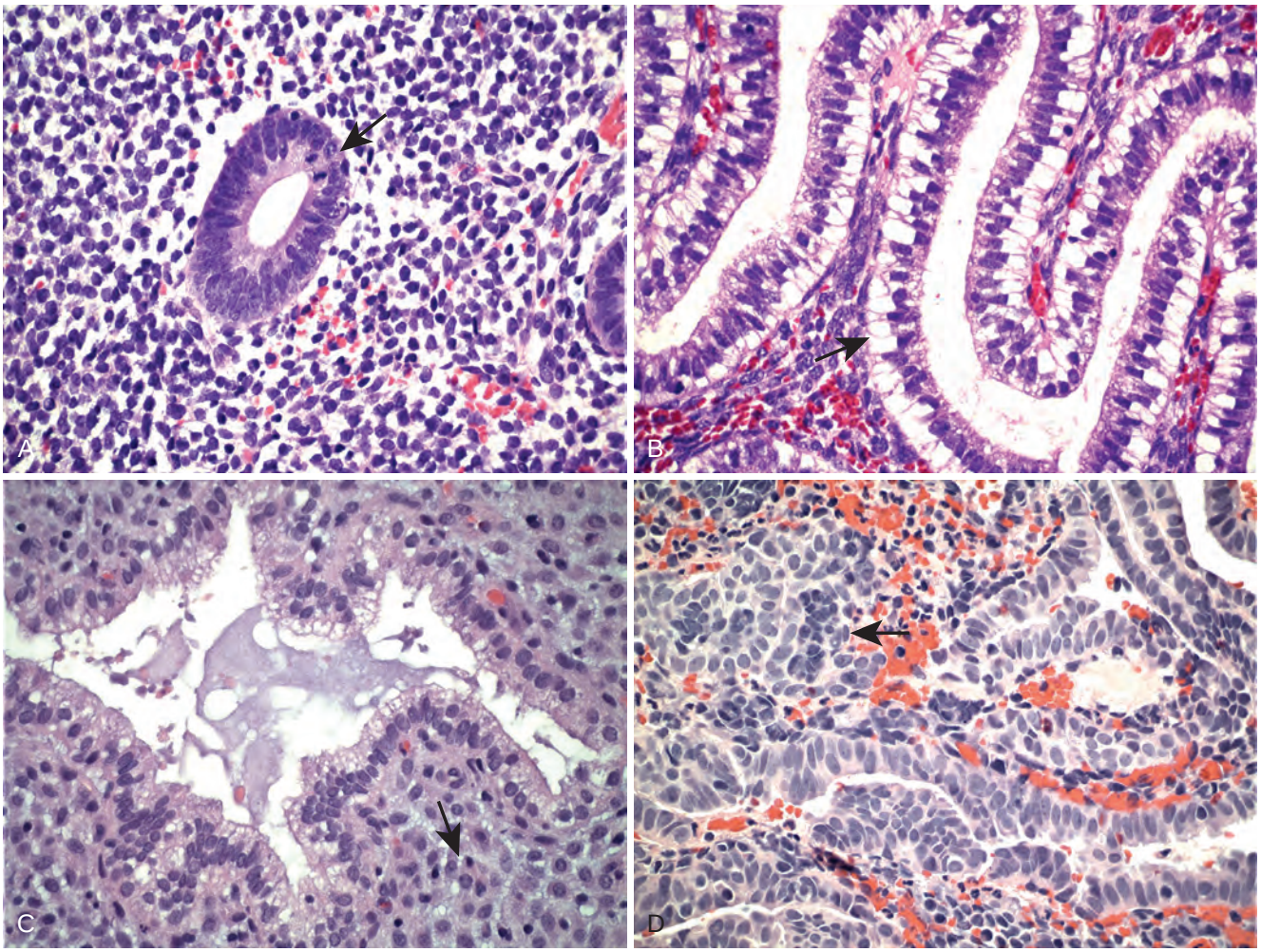


Figure 22.19 Histology of the menstrual cycle. (A) Proliferative phase with mitoses *(arrow)*. (B) Early secretory phase with subnuclear vacuoles *(arrow)*. (C) Late secretory exhaustion and predecidual changes *(arrow)*. (D) Menstrual endometrium with stromal breakdown *(arrow) (see text)*.

increase in ground substance and edema between the stromal cells followed by stromal cell hypertrophy, increased cytoplasmic eosinophilia *(predecidual change)*, and a resurgence of stromal mitoses (see Fig. 22.19C). Predecidual changes spread throughout the functionalis and are accompanied by a sparse infiltrate of neutrophils and lymphocytes, which in this context are considered normal.

- With the dissolution of the corpus luteum and the subsequent drop in progesterone levels, the functionalis degenerates and bleeding into the stroma occurs, followed by stromal breakdown and onset of the next menstrual cycle (see Fig. 22.19D).

The action of the ovarian hormones on the endometrium primarily occurs through their cognate nuclear receptors. During the proliferative phase, estrogen drives the proliferation of both glands and stroma, sometimes by promoting "cross-talk" between these two cell types. For example, much of the effect of estrogen on glandular proliferation occurs via stromal cells, which in response to estrogen produce growth factors (e.g., insulin-like growth factor-1 and epidermal growth factor) that bind receptors expressed on the epithelial cells. During the secretory phase, progesterone down-regulates the expression of estrogen receptor in both the glands and the stroma, and as a result endometrial proliferation is suppressed. Progesterone also promotes the differentiation of the glands and causes functional changes in the stromal cells. Endometrial stem cells have been identified that likely have a central role in the regeneration of the endometrium after menses. They may also contribute to the development of ectopic endometrial tissue and endometrial cancer.

FUNCTIONAL ENDOMETRIAL DISORDERS (DYSFUNCTIONAL UTERINE BLEEDING)

Although abnormal uterine bleeding can be caused by well-defined pathologic conditions, such as chronic endometritis, endometrial polyps (Fig. 22.20C), submucosal leiomyomas (see Fig. 22.20D), or endometrial neoplasms, it most commonly stems from hormonal disturbances that produce *dysfunctional uterine bleeding* (Table 22.3). This is a clinical term for uterine bleeding that lacks an underlying structural abnormality. As discussed earlier, the normal

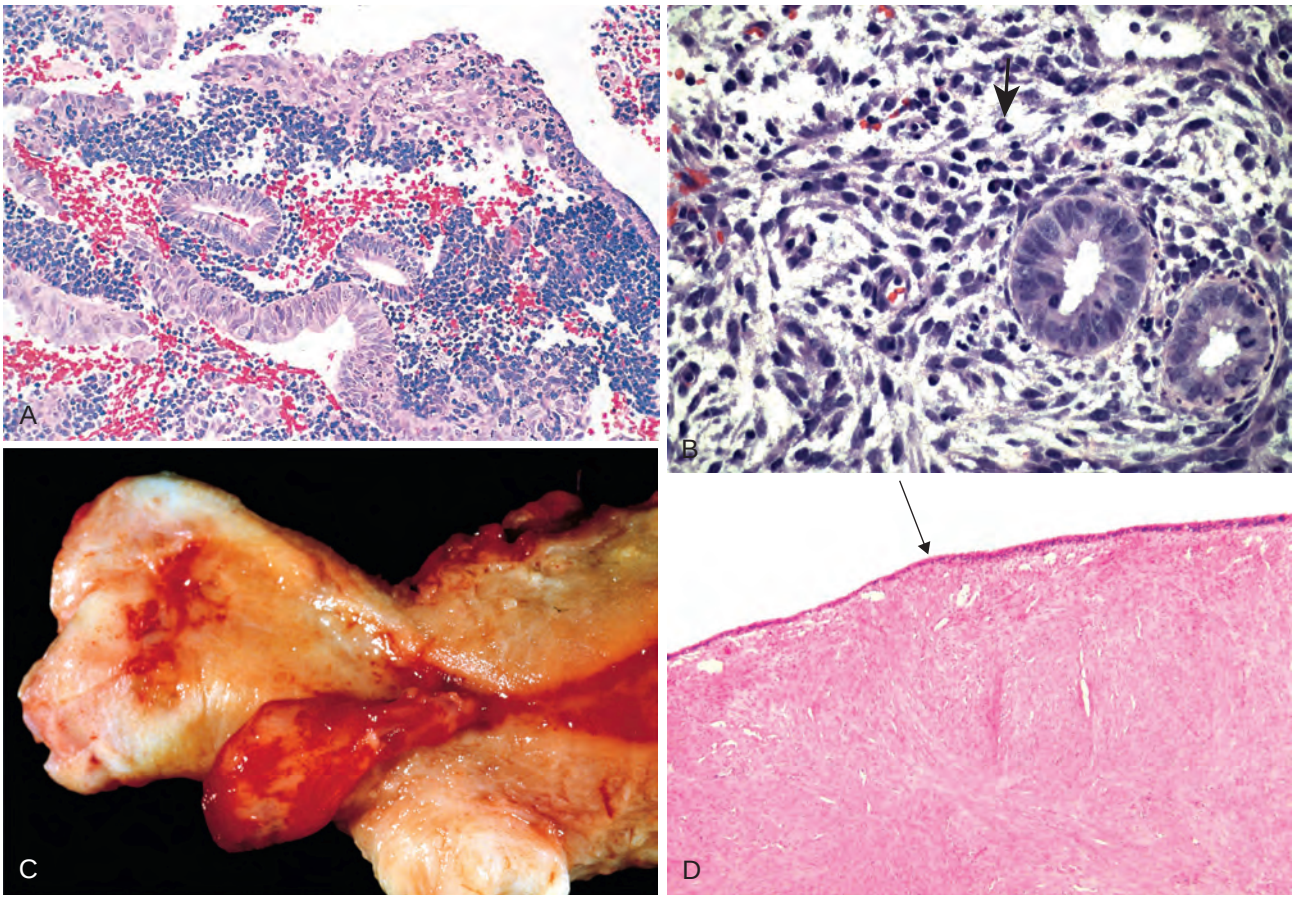


Figure 22.20 Common causes of abnormal uterine bleeding. (A) The most common is dysfunctional uterine bleeding, seen here as anovulatory endometrium with stromal breakdown. Note breakdown associated with proliferative glands. (B) Chronic endometritis with plasma cells *(arrow)*. (C) Endometrial polyp. (D) Submucosal leiomyoma with attenuation of the endometrial lining *(arrow)*.

cyclical proliferation, differentiation, and shedding of the endometrium requires that all the involved pituitary and ovarian hormones be released at the proper time in the right amounts. Any disturbance of this finely tuned system may result in dysfunctional uterine bleeding, the most common causes of which are discussed in the following sections.

Table 22.3 Causes of Abnormal Uterine Bleeding by Age Group

Age Group	Causes
Prepuberty	Precocious puberty (hypothalamic, pituitary, or ovarian origin)
Adolescence	Anovulatory cycle, coagulation disorders
Reproductive age	Complications of pregnancy (abortion, trophoblastic disease, ectopic pregnancy) Anatomic lesions (leiomyoma, adenomyosis, polyps, endometrial hyperplasia, carcinoma) Dysfunctional uterine bleeding Anovulatory cycle Ovulatory dysfunctional bleeding (e.g., inadequate luteal phase)
Perimenopausal	Dysfunctional uterine bleeding Anovulatory cycle Anatomic lesions (carcinoma, hyperplasia, polyps)
Postmenopausal	Endometrial atrophy Anatomic lesions (carcinoma, hyperplasia, polyps)

Anovulatory Cycle

The most frequent cause of dysfunctional bleeding is anovulation (failure to ovulate). Anovulatory cycles result from hormonal imbalances and are most common at menarche and in the perimenopausal period. Less commonly, anovulation is the result of the following:

- *Endocrine disorders,* such as thyroid disease, adrenal disease, or pituitary tumors
- *Ovarian lesions,* such as a functioning ovarian tumor (granulosa cell tumors) or polycystic ovaries (see the Ovaries section later in this chapter)
- *Generalized metabolic disturbances*, such as obesity, malnutrition, or other chronic systemic diseases

Failure of ovulation results in excessive endometrial stimulation by estrogens that is unopposed by progesterone. Under these circumstances, the endometrial glands undergo mild architectural changes, including cystic dilation, that usually resolve due to a subsequent ovulatory cycle. However, repeated anovulation may result in bleeding that, in certain clinical situations, may prompt an endometrial biopsy. In this setting, biopsies reveal stromal condensation and eosinophilic

epithelial metaplasia, features similar to those seen in menstrual endometrium. However, unlike menstrual endometrium, progesterone-dependent morphologic features (e.g., glandular secretory changes and stromal pre-decidualization) are absent because the source of progesterone, the corpus luteum, does not develop without ovulation. Most commonly, the endometrium is composed of pseudostratified glands and contains scattered mitotic figures (see Fig. 22.20A). More severe consequences of repeated anovulation are discussed in the Endometrial Hyperplasia section later in this chapter.

INFLAMMATORY DISORDERS

The endometrium and myometrium are relatively resistant to infections, primarily because the endocervix forms a barrier to ascending infection. Thus, although chronic inflammation in the cervix is common and usually insignificant, it is of concern in the endometrium.

Acute Endometritis

Acute endometritis is uncommon and limited to bacterial infections that arise after delivery or miscarriage. Retained products of conception are the usual predisposing factors; the causative agents include group A hemolytic streptococci, staphylococci, and other bacteria. The inflammatory response is chiefly limited to the stroma and is entirely nonspecific. Removal of the retained gestational products by curettage and antibiotic therapy promptly clears these infections.

Chronic Endometritis

Chronic endometritis occurs in association with the following disorders:

- Chronic pelvic inflammatory disease
- Retained gestational tissue, postpartum or postabortion
- Intrauterine contraceptive devices
- Tuberculosis, either from miliary spread or, more often, from drainage of tuberculous salpingitis. Endometrial tuberculosis is rare in high income countries.

The diagnosis of chronic endometritis rests on the identification of plasma cells in the stroma (see Fig. 22.20B), which are not seen in normal endometrium. In about 15% of cases, no cause is apparent. Some women with this so-called "nonspecific" chronic endometritis have gynecologic complaints such as abnormal bleeding, pain, discharge, and infertility. *Chlamydia* may be involved and is commonly associated with both acute (e.g., neutrophils) and chronic (e.g., lymphocytes, plasma cells) inflammatory infiltrates. The responsible organisms may not be detected by culture. If infection is suspected on clinical grounds, antibiotic therapy is indicated, even in the face of negative cultures, as it may prevent other sequelae (e.g., salpingitis).

ENDOMETRIOSIS AND ADENOMYOSIS

Endometriosis is defined by the presence of "ectopic" endometrial tissue at a site outside of the uterus. The abnormal tissue most commonly includes both endometrial glands and stroma, but may consist only of stroma in some cases. It occurs in the following sites, in descending order of frequency: (1) ovaries, (2) uterine ligaments, (3) rectovaginal septum, (4) cul de sac, (5) pelvic peritoneum, (6) serosa of the large and small bowel and appendix, (7) mucosa of the cervix, vagina, and fallopian tubes, and (8) laparotomy scars.

Endometriosis can have significant clinical consequences; it often causes infertility, dysmenorrhea (painful menstruation), pelvic pain, and other problems. The disorder is principally a disease of women in active reproductive life, most often in the third and fourth decades, and affects approximately 10% of women. There are three types of endometriosis: superficial peritoneal endometriosis, ovarian endometriosis, and deep infiltrating endometriosis (Fig. 22.21). The superficial and ovarian forms of endometriosis are uncommonly associated with the development of malignancy, whereas it is extremely rare for the deep infiltrating form to undergo malignant transformation.

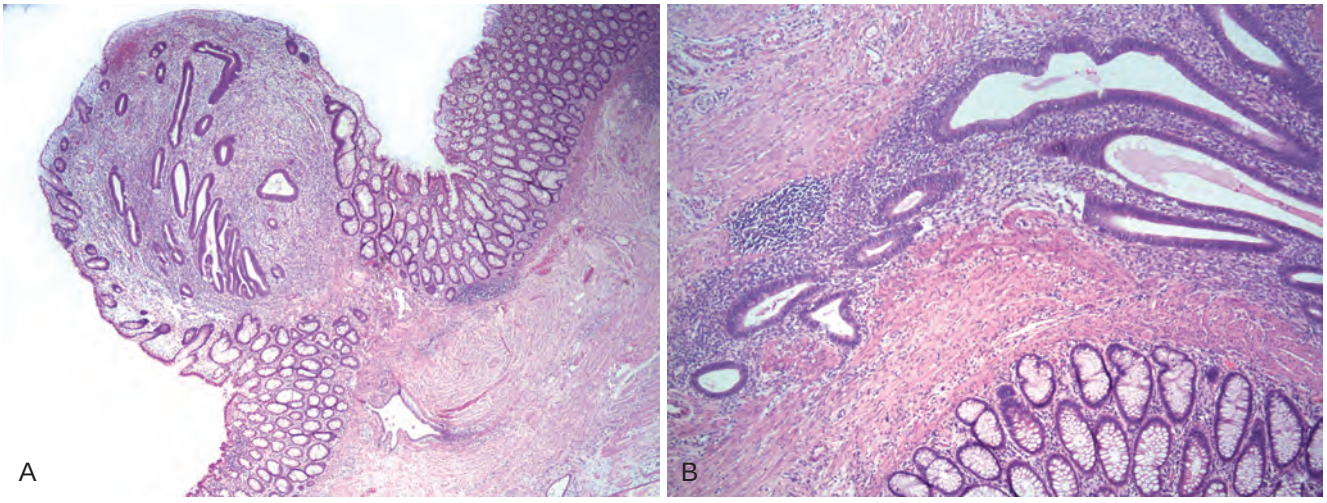


Figure 22.21 Endometriosis. (A) Endometriosis involving the mucosa of the colon. (B) Higher magnification reveals endometrial glands and stroma adjacent to normal colonic mucosa.

Pathogenesis

The pathogenesis of endometriosis remains elusive. Proposed origins of endometriotic lesions fall into two main categories: (1) those that propose an origin from the uterine endometrium and (2) those that propose an origin from cells outside the uterus that have the capacity to give rise to endometrial tissue. The leading theories are as follows:

- *The regurgitation theory* proposes that endometrial tissue implants at ectopic sites via retrograde flow of menstrual endometrium.
- *The benign metastasis theory* holds that endometrial tissue from the uterus can "spread" to distant sites (e.g., bone, lung, and brain) via blood vessels and lymphatic channels.
- *The metaplastic theory* suggests that endometrium arises directly from coelomic epithelium (mesothelium of pelvis or abdomen), from which the müllerian ducts and ultimately the endometrium itself originate during embryonic development. In addition, mesonephric remnants may undergo endometrial differentiation and give rise to ectopic endometrial tissue.
- *The extrauterine stem/progenitor cell theory* proposes that stem/progenitor cells from the bone marrow differentiate into endometrial tissue.

Retrograde menstruation through the fallopian tubes is common, and the regurgitation theory provides a plausible explanation for the origin of ectopic endometrial tissue in the peritoneal cavity, which constitutes the vast majority of cases. However, it cannot explain endometriosis in women who are amenorrheic because of a variety of underlying etiologies (e.g., gonadal dysgenesis); endometriosis in the urogenital tract of men treated with high-dose estrogens for prostate cancer; and endometriosis in distant sites like the brain, lung, and bone. In addition, the relatively low incidence of endometriosis, despite the common occurrence of retrograde menstruation (up to 90% of women), suggests that additional factors are involved in the pathogenesis of the disorder.

Molecular analyses have provided additional insights. The endometriotic implants show certain differences when compared to the endometria of women without endometriosis (Fig. 22.22). These include the following:

- *Release of proinflammatory and angiogenic factors,* including PGE_2, IL-1β, TNFα, IL-6 and IL-8, NGF, VEGF, MCP-1, MMPs, and TIMPs. Some of these factors are released by associated macrophages, which are recruited to endometriotic implants by proinflammatory factors. Like metastatic tumors, the ability of endometriotic implants to survive and grow is dependent on angiogenesis, which is mediated by typical proangiogenic factors such as VEGF. Similarly, the ability to implant requires remodeling of the extracellular matrix, which is carried out by factors such as matrix metalloproteases. These factors may also contribute to avoidance of immune clearance.
- *Increased estrogen and retinoic acid production by endometriotic stromal cells,* due in large part to high levels of the key steroidogenic enzyme aromatase, which is absent in normal endometrial stroma. Estrogen enhances the survival and persistence of endometriotic tissue, and inhibitors of aromatase are beneficial in the treatment of endometriosis. A link between inflammation and estrogen production is made plausible by the ability of prostaglandin E_2 to stimulate local synthesis of estrogen. It also has been shown that retinoic acid produced in the stromal cells promotes epithelial cell survival via paracrine signaling. In addition, epigenetic alterations (DNA methylation) have been described that lead to increased responsiveness to estrogen and decreased responsiveness to progesterone, alterations that promote endometrial proliferation and survival. These abnormalities are present not only in ectopic endometriotic tissue, but also, albeit to a lesser degree, in the uterine endometrium of patients with endometriosis, suggesting that there is a fundamental defect in the endometrium.
- *Mutations in tumor suppressor genes and oncogenes,* such as *KRAS, PIK3CA, PPP2R1A,* and *ARID1A,* have been identified in the epithelial cells of deeply infiltrating endometriosis. The significance of these mutations in cancer driver genes is unknown, but it is possible that they contribute to the locally aggressive behavior of deeply infiltrating endometriosis. An association between endometriosis and ovarian cancer of the endometrioid and clear cell types (discussed later) has been noted in a number of epidemiologic studies with an approximate threefold increase in women with endometriosis. These studies suggest a common origin of abnormal endometriotic tissue and ovarian cancers in some cases. Consistent with this idea, in some instances it has been shown that carcinomas harbor the same mutations as associated endometriosis in the same patient, strongly supporting a pathogenic role for these driver mutations in the development of both the premalignant and malignant lesions.

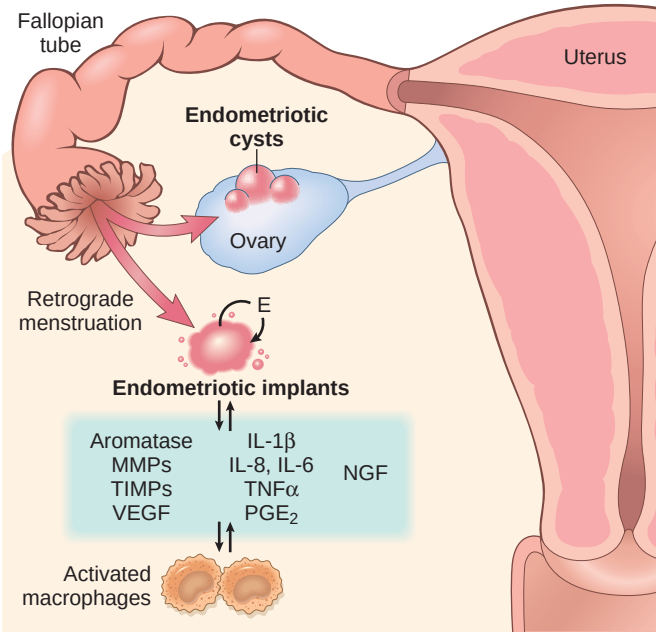


Figure 22.22 Pathogenesis of endometriosis. Depicted is the interplay between factors expressed in endometriotic implants and activated macrophages that are hypothesized to play a role in the establishment and maintenance of endometriotic implants. *E,* Estrogen.

MORPHOLOGY

Endometriotic lesions bleed periodically in response to both extrinsic cyclic (ovarian) and intrinsic hormonal stimulation. This bleeding produces nodules with a red-blue to yellow-brown appearance on or just beneath the mucosal and/or serosal surfaces at sites of involvement. When lesions are widespread, organizing hemorrhage causes extensive fibrous adhesions between tubes, ovaries, and other structures and obliterates the pouch of Douglas. The ovaries may become markedly distorted by large cystic masses (3 to 5 cm in diameter) filled with brown fluid resulting from previous hemorrhage; these are referred to clinically as **chocolate cysts** or **endometriomas.** Deeply infiltrating endometriosis can invade tissues and also cause fibrosis and adhesions.

The histologic diagnosis of endometriosis is usually straightforward but may be difficult in long-standing cases in which the endometrial tissue is obscured by secondary fibrosis. The diagnosis is readily made when both endometrial glands and stroma are present (see Fig. 22.21B), with or without the presence of hemosiderin. In rare cases only stroma is identified. If only glands are present, other diagnoses with different clinical ramifications, such as endosalpingiosis, must be considered.

Atypical endometriosis, the likely precursor to endometriosis-related ovarian carcinoma, has two morphologic appearances. One shows cytologic atypia of the epithelium lining the endometriotic cyst, without major architectural changes. The second is marked by glandular crowding due to excessive epithelial proliferation, often associated with cytologic atypia, producing an appearance that resembles complex atypical endometrial hyperplasia (discussed later).

Clinical Features

Clinical signs and symptoms usually include severe dysmenorrhea, dyspareunia (pain with intercourse), and pelvic pain due to the intrapelvic bleeding and periuterine adhesions. Menstrual irregularities are common, and infertility is the presenting complaint in 30% to 40% of women. In addition, although uncommon, malignancies can develop within endometriomas, suggesting that these lesions contain "at-risk" epithelium.

A related disorder, *adenomyosis,* is defined as the presence of endometrial tissue within the uterine wall (myometrium). Adenomyosis remains in continuity with the endometrium, presumably signifying down-growth of endometrial tissue into and between the smooth muscle fascicles of the myometrium. Adenomyosis occurs in up to 20% of uteri. On microscopic examination, irregular nests of endometrial stroma, with or without glands, are arranged within the myometrium. Like endometriosis, the clinical symptoms of adenomyosis include menometrorrhagia (irregular and heavy menses), colicky dysmenorrhea, dyspareunia, and pelvic pain, particularly during the premenstrual period. It can coexist with endometriosis.

KEY CONCEPTS

ENDOMETRIOSIS

- Endometriosis is defined as endometrial glands and stroma outside of the uterus. The "ectopic" endometrial tissue may bleed cyclically.
- Most common sites of endometriosis are within the abdominal cavity, but occasionally it is found at distant sites.
- Several theories (regurgitation, metaplasia, metastasis, and stem cell origin) are proposed to explain the distribution of endometriosis.
- It commonly results in dysmenorrhea, pelvic pain, and infertility.
- Peritoneal and ovarian endometriosis may be a precursor to carcinoma (endometrioid and clear cell carcinoma).

Endometrial Polyps

Endometrial polyps are exophytic masses of variable size that project into the endometrial cavity. They may be single or multiple and are usually sessile and relatively small, measuring 0.5 to 3 cm in diameter, but are occasionally large and pedunculated. Polyps may be asymptomatic or may cause abnormal bleeding and infertility.

Cytogenetic studies indicate that the stromal cells in endometrial polyps contain acquired chromosomal rearrangements that are similar to those found in other benign mesenchymal tumors, suggesting that these lesions are best considered epithelial neoplasms. Polyps are responsive to estrogen but show little or no response to progesterone (see Fig. 22.20C). Endometrial polyps may occur in association with the administration of tamoxifen, which is often used in the therapy of breast cancer due to its anti-estrogenic activity on the breast. However, tamoxifen has weak pro-estrogenic effects in the endometrium. Atrophic polyps, which mainly occur in postmenopausal women, likely represent the atrophic remnants of polyps that developed prior to menopause. Rarely, adenocarcinoma arises within endometrial polyps.

ENDOMETRIAL HYPERPLASIA

Endometrial hyperplasia is an important cause of abnormal bleeding and a frequent precursor to the most common type of endometrial carcinoma. It is defined as an abnormal proliferation of the endometrial glands relative to the stroma, resulting in an increased gland-to-stroma ratio when compared with normal proliferative endometrium. Clinicopathologic and epidemiologic studies have pointed to the malignant potential of endometrial hyperplasia, and molecular studies have confirmed this relationship, as endometrial hyperplasia and carcinoma share specific acquired driver mutations in cancer genes (described later).

Endometrial hyperplasia is associated with *prolonged estrogenic stimulation of the endometrium,* which can be due to anovulation, increased estrogen production from endogenous sources, or exogenous estrogen. Associated conditions include the following:

- Obesity (peripheral conversion of androgens to estrogens)
- Menopause
- Polycystic ovarian syndrome
- Functioning granulosa cell tumors of the ovary
- Excessive ovarian cortical function (cortical stromal hyperplasia)
- Prolonged administration of estrogenic substances (estrogen replacement therapy)

These are the same influences postulated to be of pathogenic significance in some endometrial carcinomas, discussed later.

Inactivation of the *PTEN* tumor suppressor gene is a common genetic alteration in both endometrial hyperplasias and endometrioid endometrial carcinoma. As discussed in Chapter 7, *PTEN* encodes a lipid phosphatase that is an important negative regulator of phosphatidylinositol 3-kinase (PI3K)/AKT growth-regulatory pathway. When PTEN function is lost, the PI3K/AKT pathway becomes overactive. Mutations in *PTEN* are found in more than 20% of hyperplasias, both with and without atypia, and in 30% to 80% of endometrial carcinomas, suggesting that alterations in *PTEN* occur at an early stage of endometrial tumorigenesis (although they are not predictive of progression of hyperplasia to carcinoma). Of note, patients with *Cowden syndrome,* which is caused by germline mutations in *PTEN,* have a high incidence of endometrial carcinoma and certain other tumors, particularly breast cancer. As with many other tumor suppressors, it is not entirely clear why the loss of *PTEN* (which is expressed in many tissues) is so highly associated with particular tumors. It is interesting to note, however, that PI3K/AKT signaling enhances the ability of the estrogen receptor to turn on the expression of its target genes. Thus, loss of PTEN function may stimulate estrogen-dependent gene expression, leading to overgrowth of cell types that depend on estrogen for trophic signals, such as endometrial and mammary epithelial cells.

MORPHOLOGY

The classification of endometrial hyperplasia according to the World Health Organization (WHO) includes two major categories, hyperplasia and atypical hyperplasia (also referred to as endometrial intraepithelial neoplasia), which differ in appearance and propensity to progress to carcinoma. **Typical hyperplasia** has a wide-range of appearances, but the cardinal feature is an increased gland-to-stroma ratio. The glands show variation in size and shape and may be dilated (Fig. 22.23A). Although there may be back-to-back glands focally, some intervening stroma is usually retained (see Fig. 22.23B). These lesions are caused by persistent estrogen stimulation and rarely progress to adenocarcinoma (approximately 1% to 3%). Hyperplasia may evolve into cystic atrophy when estrogen is withdrawn.

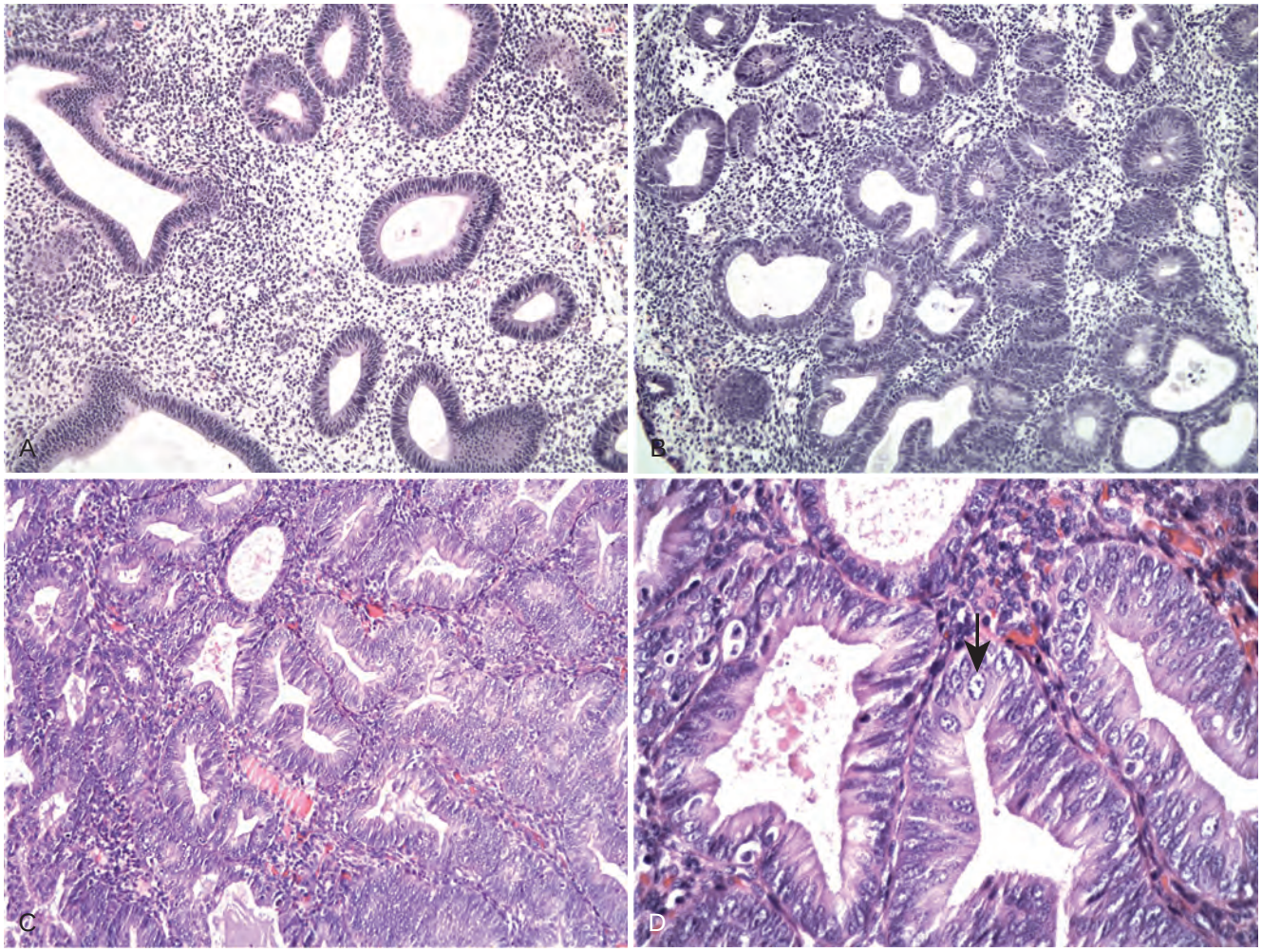


Figure 22.23 Endometrial hyperplasia. (A) Typical hyperplasia. Note the architectural abnormalities including mild glandular crowding and cystic glandular dilation. (B) Typical hyperplasia demonstrating increased glandular crowding with areas of back-to-back glands and cytologic features similar to proliferative endometrium. (C) Atypical hyperplasia with further increase in glandular crowding and abnormal cytologic features. (D) High magnification of atypical hyperplasia showing rounded, vesicular nuclei with prominent nucleoli *(arrow)*.

Atypical hyperplasia (endometrial intraepithelial neoplasia) is composed of complex patterns of proliferating glands displaying nuclear atypia. The glands are commonly back-to-back and often have complex outlines due to branching structures. Individual cells are rounded and lose the normal perpendicular orientation to the basement membrane. In addition, the nuclei have open (vesicular) chromatin and conspicuous nucleoli. The features of atypical hyperplasia have considerable overlap with those of well-differentiated endometrioid adenocarcinoma (discussed later), and accurate distinction from cancer may not be possible without hysterectomy (Fig. 22.23C–D). Indeed, up to 50% of women with a diagnosis of atypical hyperplasia are found to have carcinoma when a hysterectomy is performed.

Currently, atypical hyperplasia is managed by hysterectomy or, in young women who desire fertility, a trial of progestin therapy and close follow-up. Most often, hopefully after a successful pregnancy, lack of regression prompts removal of the uterus.

KEY CONCEPTS

ENDOMETRIAL HYPERPLASIA

- Endometrial hyperplasia is defined as an increase in the number of glands relative to the stroma, appreciated as crowded glands, often with abnormal shapes.
- It is most commonly caused by unopposed estrogen stimulation and is an important cause of abnormal vaginal bleeding.
- It is divided into typical and atypical hyperplasia based on nuclear morphology. Atypical hyperplasia is associated with an increased risk of endometrial carcinoma.
- The *PTEN* tumor suppressor gene is mutated in approximately 20% of endometrial hyperplasias.

MALIGNANT TUMORS OF THE ENDOMETRIUM

Carcinoma of the Endometrium

Endometrial carcinoma is the most common invasive cancer of the female genital tract. It accounts for 7% of all invasive cancer in women, excluding skin cancer. At one time endometrial carcinoma was far less common than cancer of the cervix, but earlier detection and eradication of the precursor lesions of cervical carcinoma, coupled with an increase in endometrial carcinomas in younger women, have reversed this ratio. In 2019 in the United States, 61,880 new endometrial cancers and 12,160 deaths were predicted. Worldwide there were over 380,000 new cases diagnosed in 2018.

Pathogenesis

Clinicopathologic studies and molecular analyses support the existence of two broad categories of endometrial carcinoma referred to as type I and type II, each with distinct risk factors, precursor lesions, molecular genetics, and clinical behaviors (summarized in Table 22.4). More recently, genomic sequencing of endometrial carcinoma has revealed four major molecular subtypes among endometrioid and serous carcinomas, the most common morphologies of type 1 and type 2 tumors, respectively. The salient molecular features of these four subtypes are as follows:

- *Ultramutated/POLE tumors,* which are defined by the presence of mutations in DNA polymerase ε (POLE) that produce an exceptionally high burden of somatic mutations, the large majority of which are passenger mutations.
- *Hypermutated/MSI (microsatellite instability) tumors,* defined by mutations in or epigenetic silencing of mismatch repair genes, also leading to genomic instability and a high burden of somatic mutations.
- *Copy number low/MSS (microsatellite stable) tumors,* a common subtype also associated with endometrioid morphology that is frequently associated with mutations that upregulate signaling through the PI3K/AKT pathway.
- *Copy number high/serous-like tumors,* aggressive tumors with serous or high-grade endometrioid morphology that are often associated with *TP53* mutations and numerous genomic copy number variants.

Table 22.4 Characteristics of Type I and Type II Endometrial Carcinoma

Characteristics	Type I	Type II
Age	55–65 years	65–75 years
Clinical setting	Unopposed estrogen Obesity Hypertension Diabetes	Atrophy Thin physique
Morphology	Endometrioid	Serous Clear cell Mixed müllerian tumor
Precursor	Hyperplasia	Serous endometrial intraepithelial carcinoma
Mutated genes/ genetic abnormalities	*PTEN* *ARID1A* (regulator of chromatin) *PIK3CA* (PI3K) *KRAS* *FGF2* (growth factor) MSI[a] *CTNNB1* (Wnt signaling) *POLE* *TP53* (progressed tumors)	*TP53* Aneuploidy *PIK3CA* (PI3K) *FBXW7* (regulator of MYC, cyclin E) *CCNE1* *PPP2R1A* (PP2A)
Behavior	Indolent Spreads via lymphatics	Aggressive Intraperitoneal and lymphatic spread

[a]Microsatellite instability.

Endometrioid Endometrial Carcinoma

This is the most common type of endometrial carcinoma, accounting for approximately 80% to 85% of cases. The majority of these tumors fall in to the type I category. Most are well differentiated and mimic proliferative endometrial glands, features that are the basis for their name. As discussed earlier, it typically arises in the setting of endometrial hyperplasia, and, like endometrial hyperplasia, is associated with conditions in which unopposed estrogenic stimulation

of the endometrium occurs. Most notable of these is obesity, a rapidly increasing problem in high income countries. The association with obesity underlies other conditions that commonly co-occur with endometrial carcinoma, particularly hypertension and type 2 diabetes.

As with other cancers, development of endometrioid carcinoma stems from the stepwise acquisition of genetic alterations in tumor suppressor genes and oncogenes. In hysterectomy specimens containing both atypical hyperplasia and carcinoma, identical *PTEN* mutations have been identified in each component, supporting the view that atypical hyperplasia is a precursor to carcinoma and that *PTEN* mutations occur before the development of overt carcinoma (Fig. 22.24A).

Sequencing of the genomes of endometrioid carcinomas has shown that the most common mutations act to increase signaling through the PI3K/AKT pathway, which is a hallmark of this particular tumor type. As mentioned earlier, PI3K/AKT signaling augments expression of estrogen receptor–dependent target genes in endometrial cells. Type I endometrial carcinomas are somewhat unique in that individual tumors may harbor multiple mutations that increase PI3K/AKT signaling, suggesting that tumor development and progression is fostered by successive increases in signal strength. Among the mutations that impact the PI3K/AKT pathway in endometrial carcinomas are the following:

- Mutations in the *PTEN* tumor suppressor gene are found in 30% to 80% of endometrioid carcinomas.
- *PIK3CA*, an oncogene that encodes the catalytic subunit of PI3K, harbors activating mutations in approximately 40% of endometrioid carcinomas. *PIK3CA* mutations rarely occur in atypical hyperplasias, suggesting that mutations in *PIK3CA* play a role in malignant transformation.
- Mutations that activate *KRAS*, which also stimulates PI3K/AKT signaling, are found in approximately 25% of cases.
- Loss-of-function mutations in *ARID1A*, a regulator of chromatin structure, occur in approximately one-third of tumors. Of interest, *ARID1A* is also frequently mutated in ovarian endometrioid and clear cell carcinomas, tumors that arise within endometriosis. Through unclear mechanisms, it appears that loss of ARID1A function also enhances pro-oncogenic changes in gene expression that are mediated by PI3K/AKT signaling.

Other commonly observed mutations in endometrioid carcinoma disrupt genes that are required for the maintenance of genomic stability, presumably acting in part to create a mutator phenotype that increases the rate of acquisition of mutations in oncogenes and tumor suppressor genes.

- *Defects involving DNA mismatch repair genes* are found in about 20% of sporadic tumors and are particularly prevalent in endometrial carcinomas arising in women with hereditary nonpolyposis colorectal carcinoma syndrome (HNPCC, also referred to as Lynch syndrome, discussed in Chapter 17). In sporadic endometrioid

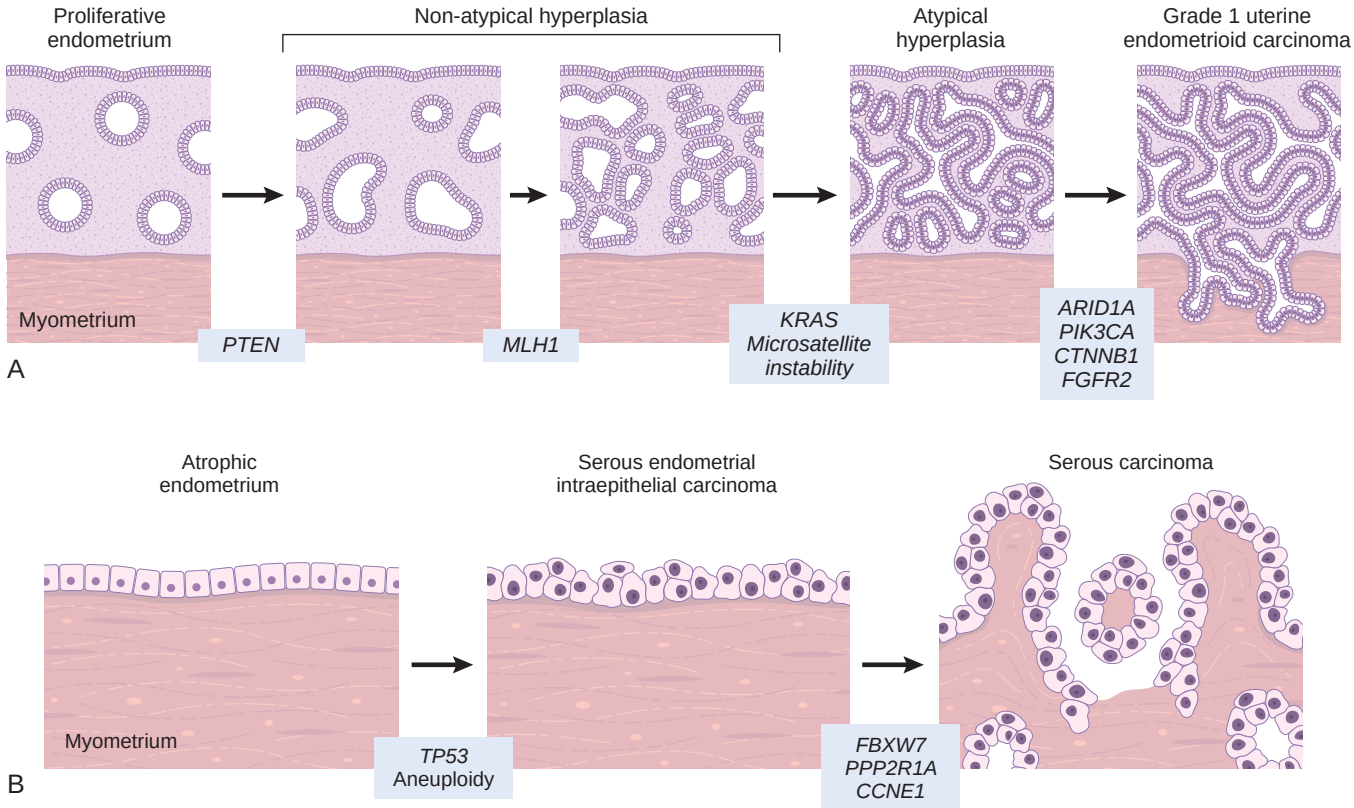


Figure 22.24 Schematic diagram depicting the development of type I endometrial carcinoma (A) and type II endometrial carcinoma (B). The most common molecular genetic alterations are shown at the time they are most likely to occur during the progression of the disease. *MLH1* encodes a component of the mismatch repair complex of proteins, and its loss of function results in microsatellite instability, found in about 20% of endometrioid carcinomas.

carcinoma, loss of expression of DNA mismatch repair genes is commonly caused by epigenetic silencing (via promoter hypermethylation).

- *A small subset (less than 10%) of endometrial carcinomas have mutations that disrupt the proofreading function of DNA polymerase ε, encoded by the POLE gene*. These tumors have a remarkably high burden of somatic point mutations, possibly the highest of any human cancer.
- *In approximately 50% of poorly differentiated carcinomas, loss-of-function mutations in TP53 are present*. Because *TP53* mutations are lacking in the vast majority of well-differentiated endometrioid carcinomas, these mutations are thought to be late events involved in tumor progression.

MORPHOLOGY

Endometrioid carcinoma can take the form of a localized polypoid mass or diffusely involve the endometrial lining (Fig. 22.25A). Spread generally occurs by myometrial invasion followed by direct extension to adjacent structures/organs. Invasion of the broad ligaments may create a palpable mass. Eventually, dissemination to regional lymph nodes occurs, and in the late stages, distant metastatic spread to the lungs, liver, bones, and other organs may be seen.

Endometrioid adenocarcinomas demonstrate glandular growth patterns that fall into three histologic grades: **well differentiated** (grade 1) (Fig. 22.25B), composed almost entirely of well-formed glands; **moderately differentiated** (grade 2) (Fig. 22.25C), showing well-formed glands mixed with areas composed of solid sheets of cells, which by definition make up 50% or less of the tumor; and **poorly differentiated** (grade 3) (Fig. 22.25D), characterized by greater than 50% solid growth pattern. Well-differentiated tumors may be distinguished from hyperplasias by the presence of desmoplastic stroma or complex epithelial growth patterns (e.g., confluent glandular or papillary growth patterns).

Tumors with polymerase ε mutations or defects in DNA mismatch repair are frequently associated with large numbers of infiltrating T cells. It is believed that these cells represent an ineffective host response to neoantigens created by mutations that lead to tumor-specific amino acid substitutions in proteins.

Up to 20% of endometrioid carcinomas contain foci of squamous differentiation. Squamous elements may be histologically benign-appearing when they are associated with well-differentiated

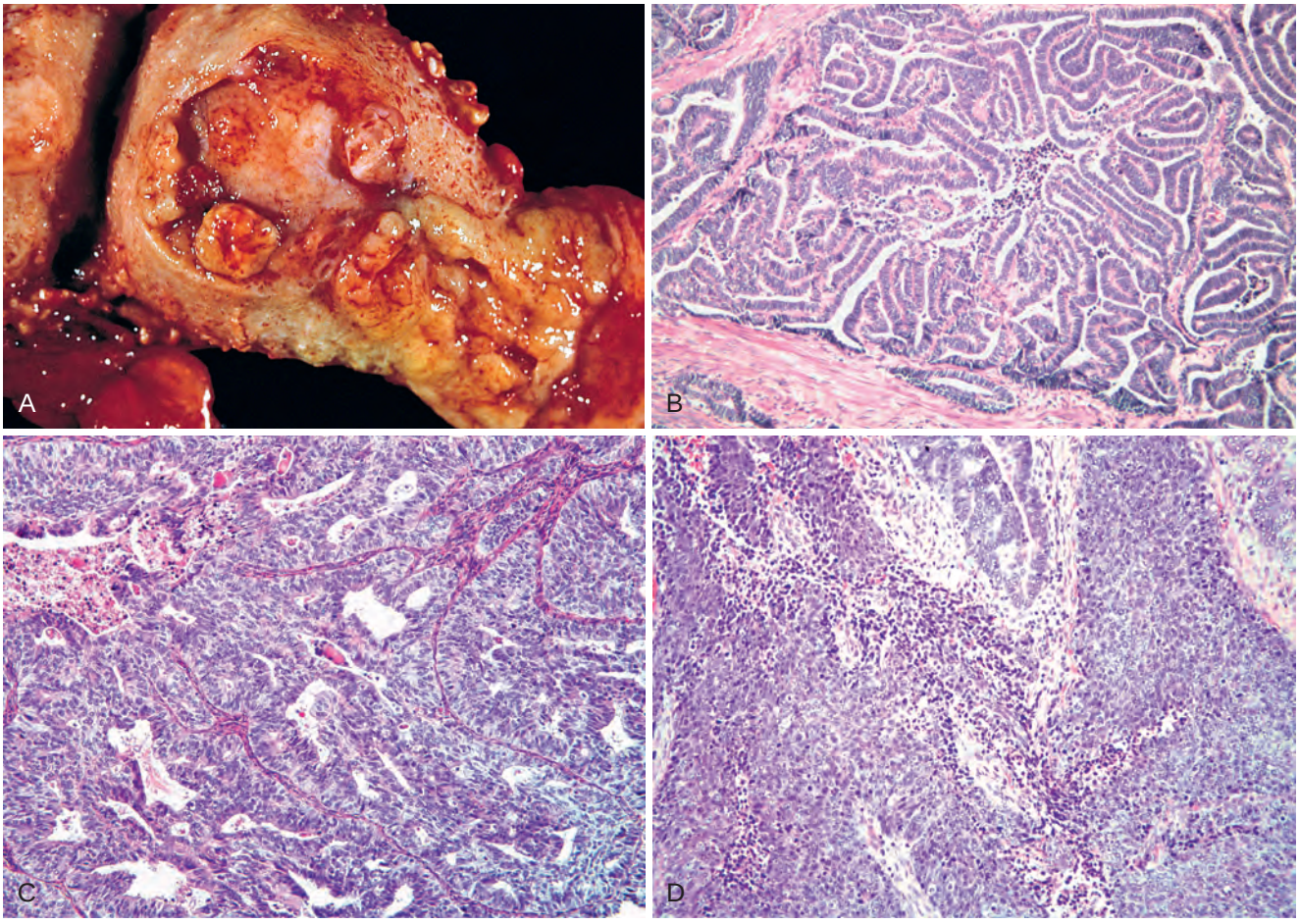


Figure 22.25 Type I endometrial carcinoma. (A) Endometrial adenocarcinoma presenting as a fungating mass in the fundus of the uterus. (B) Well-differentiated (grade 1) endometrioid adenocarcinoma with preserved glandular architecture but lack of intervening stroma. (C) Moderately differentiated (grade 2) endometrioid adenocarcinoma with glandular architecture admixed with solid areas. (D) Poorly differentiated (grade 3) endometrioid adenocarcinoma with a predominantly solid growth pattern.

adenocarcinomas. Less commonly, moderately or poorly differentiated endometrioid carcinomas contain squamous elements that appear frankly malignant. Current classification systems grade the carcinomas based on glandular differentiation alone and ignore areas of squamous differentiation.

Serous Endometrial Carcinoma

Serous endometrial carcinoma generally occurs in women who are about 10 years older than those with endometrioid carcinomas, and in contrast with endometrioid carcinoma, they usually arise in the setting of endometrial atrophy (see Fig. 22.24B). Serous tumors are considered type II tumors and are by definition poorly differentiated (grade 3). They account for approximately 15% of cases of endometrial carcinoma. This subtype shows significant morphologic and biologic overlap with ovarian serous carcinoma.

Pathogenesis

Serous endometrial carcinoma is highly associated with disruptive mutations in the *TP53* tumor suppressor gene.

TP53 mutations are present in greater than 90% of tumors. Most consist of missense mutations that result in the accumulation of the altered p53 protein (Fig. 22.26B and D). Genomic sequencing studies have detected driver mutations in a number of other cancer genes (see Table 22.4). In addition to point mutations, serous carcinomas typically are associated with significant chromosomal instability and numerous copy number alterations, a general feature of *TP53*-mutated cancers. All serous carcinomas belong to the copy number high/serous-like molecular category.

Mutations in *TP53* are also found in approximately 75% of endometrial intraepithelial carcinomas, an in situ precursor lesion, suggesting that mutation of *TP53* is an early event in the evolution of serous endometrial carcinoma. The tumor presumably arises as a surface epithelial neoplasm that then extends into adjacent gland structures and later invades endometrial stroma. Their generally poorer prognosis is thought to be a consequence of a propensity to exfoliate, travel through the fallopian tubes, and implant on peritoneal surfaces like their ovarian counterparts. As a result, they have often spread outside of the uterus at the time of diagnosis.

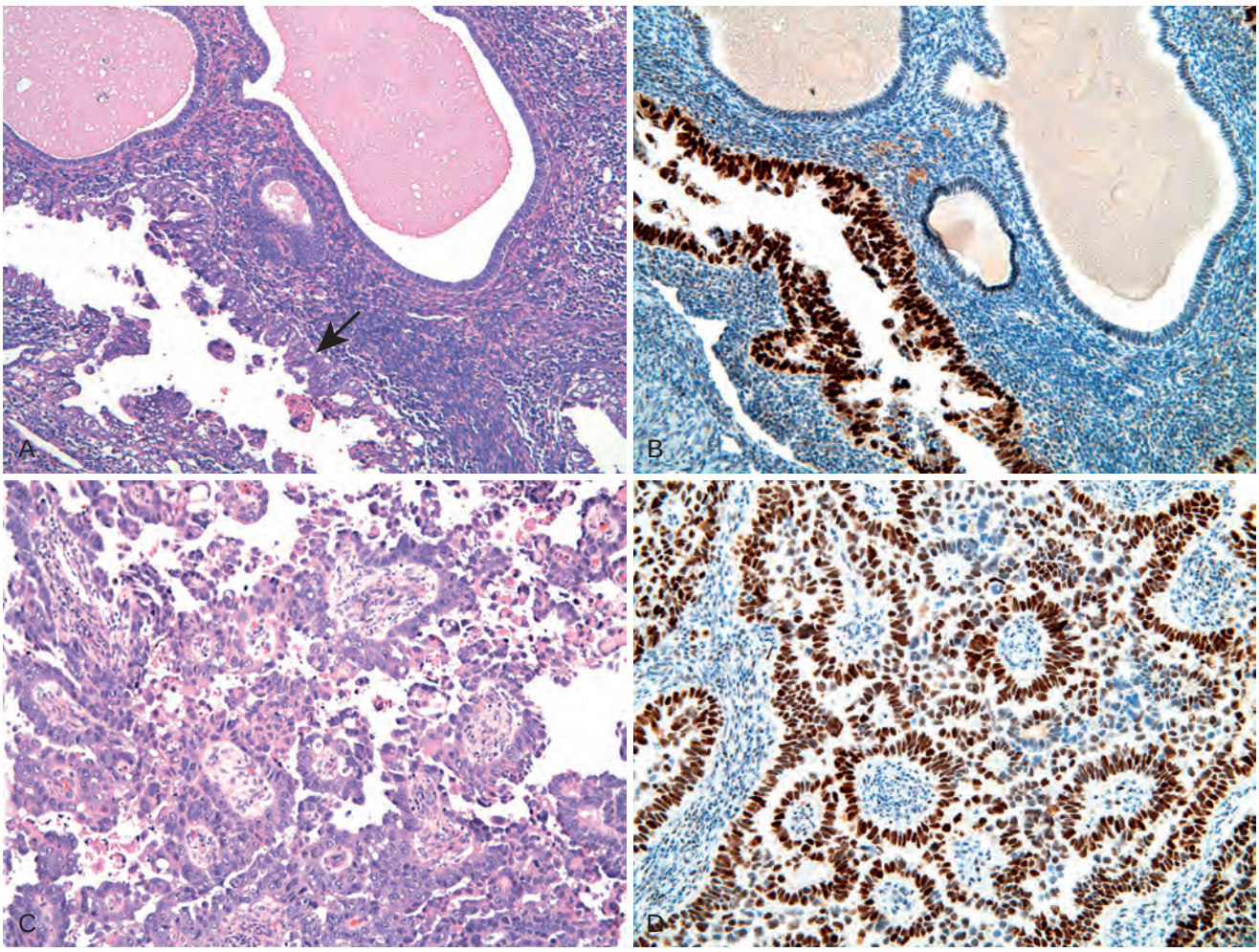


Figure 22.26 Type II endometrial carcinoma. (A) Endometrial intraepithelial carcinoma, the precursor to serous carcinoma showing malignant cells *(arrow)* with morphologic features identical to serous carcinoma lining the surfaces of the endometrial glands without obvious stromal invasion. (B) Strong, diffuse expression of p53 as detected by immunohistochemistry in endometrial intraepithelial carcinoma. (C) Serous carcinoma of the endometrium with papillary growth pattern consisting of malignant cells with marked cytologic atypia including high nuclear-to-cytoplasmic ratio, atypical mitotic figures, and hyperchromasia. (D) As with the previous lesion, there is an accumulation of p53 protein in the nucleus.

MORPHOLOGY

Generally, serous carcinomas arise in small atrophic uteri and are often large bulky tumors or deeply invasive into the myometrium. The precursor lesion, **serous endometrial intraepithelial carcinoma,** consists of malignant cells identical to those of serous carcinoma that are confined to the epithelial surface (see Fig. 22.26A–B).The invasive lesions may have a papillary growth pattern composed of cells with marked cytologic atypia including high nuclear-to-cytoplasmic ratio, atypical mitotic figures, hyperchromasia, and prominent nucleoli (see Fig. 22.26C–D). However, they can also have a predominantly glandular growth pattern; in such cases, they are distinguished from endometrioid carcinoma by the marked cytologic atypia. All of the tumors in this category are classified as grade 3 irrespective of architectural pattern. Serous carcinoma, despite relatively superficial endometrial involvement, may be associated with extensive peritoneal disease, suggesting spread by routes (i.e., tubal regurgitation) other than direct invasion.

Clinical Features

Carcinoma of the endometrium is uncommon in women younger than 40 years of age; the peak incidence is in postmenopausal women 55 to 65 years of age. There is no currently available screening test. Although it may be asymptomatic for a period, it usually produces irregular or postmenopausal vaginal bleeding. Fortunately, postmenopausal bleeding often leads to early detection, and cures are possible in most patients. The diagnosis must be established by histologic examination of tissue obtained by biopsy or curettage. At the time of diagnosis, endometrioid carcinomas are analyzed for evidence of DNA mismatch repair defects because approximately 3% to 5% of women with endometrial cancer have Lynch syndrome and are at high risk for colon carcinoma.

As would be anticipated, the prognosis depends heavily on the stage at diagnosis, as well as histologic grade and subtype. The staging system for endometrial adenocarcinoma is as follows:

Stage I—Carcinoma is confined to the corpus uteri itself.
Stage II—Carcinoma involves the corpus and the cervix.
Stage III—Carcinoma extends outside the uterus but not outside the true pelvis.
Stage IV—Carcinoma extends outside the true pelvis or involves the mucosa of the bladder or the rectum.

In the United States, most tumors (about 80%) are stage I well-differentiated or moderately differentiated endometrioid carcinomas. Surgery, alone or in combination with irradiation, gives about 90% 5-year survival in stage I (grade 1 or 2) disease. This rate drops to approximately 75% for stage I/grade 3 tumors and to 50% or less for stage II and III endometrial carcinomas.

As mentioned, serous carcinoma has a propensity for extrauterine (lymphatic or transtubal) spread. For unknown reasons, it occurs more frequently in women of African-American descent, a difference that accounts for a twofold higher mortality rate in African-American women with endometrial carcinoma compared with Caucasian women. Overall, the 5-year survival for women with serous carcinoma is 18% to 27%, and even when it is confined to the uterus the recurrence rate is as high as 80%.

Treatment varies according to tumor type. Women with endometrioid carcinoma often receive adjuvant radiation to reduce local recurrence and are treated with chemotherapy when the tumor has spread beyond the uterus. By contrast, because of the aggressive nature of serous carcinoma, women may be treated with chemotherapy, even in the absence of detectable extrauterine spread. Inhibitors of the PI3K/AKT pathway are being tested in clinical trials, and the continued identification of biologic targets is likely to expand the roster of rational therapies in the future. In addition, immune checkpoint inhibitors are under investigation for the treatment of women with tumors in the hypermutated MSI and "ultramutated" polymerase ε molecular categories.

Carcinosarcoma (Malignant Mixed Müllerian Tumors)

Carcinosarcomas (also referred to as malignant mixed müllerian tumors) are mixed epithelial and mesenchymal tumors. The epithelial component most often resembles poorly differentiated endometrioid or serous carcinoma, while the mesenchymal component can take a number of forms. Some contain uterine mesenchymal elements (stromal sarcoma, leiomyosarcoma), while others contain heterologous malignant cell types (rhabdomyosarcoma, chondrosarcoma). Based on molecular studies showing the presence of shared driver mutations, the epithelial and mesenchymal elements appear to be derived from a single founding cancer cell. The mutations found in carcinosarcomas involve the same genes that are mutated in endometrial carcinoma, such as *PTEN, TP53,* and *PIK3CA,* whereas alterations typical of sarcomas are absent, suggesting that these tumors are carcinomas that have acquired the capacity for mesenchymal differentiation. At present, the mechanisms underlying the sarcomatous transformation are unknown, but some abnormality of epigenetic regulation seems likely.

MORPHOLOGY

Carcinosarcomas are often bulky and polypoid, and they may protrude through the cervical os. These tumors usually contain areas of adenocarcinoma (endometrioid, serous, or clear cell) mixed with the malignant mesenchymal (sarcomatous) elements (Fig. 22.27A); alternatively, the tumor may contain two distinct and separate epithelial and mesenchymal components. Sarcomatous components may also mimic extrauterine tissues (e.g., striated muscle, cartilage, adipose tissue, and bone). Metastases usually contain only epithelial components (see Fig. 22.27B).

Carcinosarcomas occur in postmenopausal women and present with bleeding. Outcome is determined primarily by depth of invasion and stage. The only other known prognostic factor is the differentiation of the mesenchymal component; patients with tumors that have heterologous mesenchymal components have a poorer prognosis than those whose tumors do not. Overall 5-year survival rates are 25% to 30% for patients with high-stage disease.

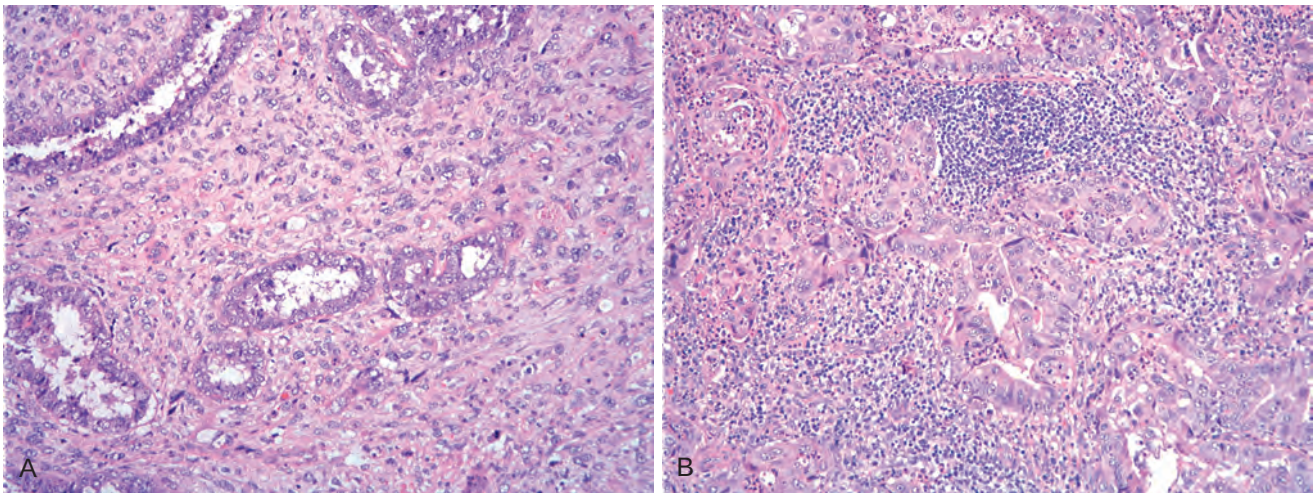


Figure 22.27 Carcinosarcoma. (A) Micrograph showing both malignant epithelial and stromal components. (B) Lymph node metastasis from a carcinosarcoma showing only the epithelial component, as is typically the case.

KEY CONCEPTS

ENDOMETRIAL CARCINOMA

- Endometrial carcinoma is the most common malignancy of the female genital tract.
- There are two major types of endometrial carcinoma: type I and type II. Type I tumors are low-grade and usually indolent; type II tumors are high-grade aggressive tumors and have a poor prognosis.
- Four molecular subtypes of endometrioid and serous carcinoma are currently recognized.
- Endometrioid (type I) carcinoma is often preceded by atypical hyperplasia and commonly has mutations that upregulate PI3K/AKT signaling.
- Serous (type II) carcinoma is associated with serous endometrial intraepithelial carcinoma, and the most common mutations are in *TP53*. *TP53* mutations are also found in precursor lesions.
- Stage remains the most important factor in outcome; serous tumors are much more likely to present at advanced stage and have a decidedly worse prognosis.
- Carcinosarcomas are aggressive tumors that resemble endometrial carcinoma genetically and have poor outcomes with current therapies.

TUMORS OF ENDOMETRIAL STROMA

These relatively uncommon tumors comprise less than 5% of endometrial cancers and include stromal neoplasms admixed with benign glands (adenosarcomas) and pure stromal neoplasms.

Adenosarcoma

Adenosarcoma presents most commonly as large broad-based endometrial polypoid growths that may prolapse through the cervical os. The diagnosis is based on the presence of malignant-appearing stroma, which coexists with benign but abnormally shaped endometrial glands. These tumors predominate in women between the fourth and fifth decades and are generally considered to be a low-grade malignancy; recurrences develop in one-fourth of cases and are nearly always confined to the pelvis. The principal diagnostic dilemma is distinguishing these tumors from large benign polyps.

Stromal Tumors

The endometrium occasionally gives rise to neoplasms that resemble normal stromal cells. Endometrial stromal neoplasms are divided into two categories: (1) benign stromal nodules and (2) endometrial stromal sarcoma. Stromal sarcoma may be further divided into low-grade and high-grade types depending on their differentiation.

Clues to the pathogenesis of stromal sarcoma have come from the identification of several recurrent chromosomal aberrations that are quite specific for these malignancies. As with many sarcomas, stromal sarcoma is associated with chromosomal translocations that create fusion genes. Low-grade endometrial stromal sarcoma usually has a translocation in which a portion of the *JAZF1* gene, which encodes a transcriptional repressor, is fused to a second gene belonging to the polycomb gene family, such as *SUZ12*. Polycomb proteins participate in complexes that introduce repressive histone marks into chromatin, thereby silencing genes, and it is hypothesized that the JAZF1 fusion proteins act by disrupting the function of the polycomb complex, leading to misexpression of oncogenic genes. Recently, high-grade endometrial stromal sarcomas have been observed to contain different chromosomal translocations that also result in the formation of fusion genes, which are presumed to be pathogenically significant but are currently of unknown function.

About one-half of stromal sarcomas recur; relapse rates range from 36% to more than 80% for stage I and stage III/IV tumors, respectively. Unfortunately, relapse is not reliably predicted by either mitotic index or the degree of cytologic atypia. Distant metastases may announce their

presence decades after the initial diagnosis, and death from metastatic tumor occurs in about 15% of cases. The 5-year survival rates average 50% for low-grade tumors and are lower for high-grade tumors.

TUMORS OF THE MYOMETRIUM

Leiomyoma

Uterine leiomyoma (commonly called *fibroid*) is perhaps the most common tumor in women. They are benign smooth muscle neoplasms that may occur singly, but more often are multiple. Most leiomyomas have normal karyotypes, but approximately 40% have a simple chromosomal abnormality. Several cytogenetic subgroups are recognized, including tumors with rearrangements of chromosomes 12q14 and 6p involving the *HMGIC* and *HMGIY* genes, respectively, which are also implicated in a variety of other benign neoplasms. Both genes encode closely related DNA-binding factors that regulate chromatin structure. Additionally, mutations in the gene *MED12* occur in roughly 70% of uterine leiomyomas. *MED12* encodes a component of Mediator, a multiprotein complex that stimulates gene expression by serving as a bridge between transcription factors and RNA polymerase. Mice expressing mutated forms of *MED12* develop uterine leiomyoma, proving that such mutations can be causative, but precisely how they contribute to tumor development remains to be defined.

MORPHOLOGY

Leiomyomas are sharply circumscribed, discrete, round, firm, gray-white tumors varying in size from small, barely visible nodules to massive tumors that fill the pelvis. Except in rare instances, they are found within the myometrium of the corpus. Only infrequently do they involve the uterine ligaments, lower uterine segment, or cervix. They can occur within the myometrium (intramural), just beneath the endometrium (submucosal) or beneath the serosa (subserosal) (Fig. 22.28A). The characteristic whorled pattern of smooth muscle bundles on cut section usually makes these lesions readily identifiable. Large tumors may develop areas of yellow-brown to red softening.

Leiomyomas are composed of bundles of smooth muscle cells that resemble the uninvolved myometrium (Fig. 22.28B). Usually, the individual muscle cells are uniform in size and shape and have a characteristic oval nucleus and long, slender bipolar cytoplasmic processes. Mitotic figures are scarce. Morphologic variants include leiomyoma with bizarre nuclei, which has nuclear atypia and giant cells, and cellular leiomyomas. Both have a low mitotic index, helping to distinguish these benign tumors from leiomyosarcoma. An extremely rare variant, **intravenous leiomyomatosis,** is a uterine leiomyoma that extends into vessels and spreads hematogenously to other sites, most commonly the vena cava and the right atrium. Another variant, **disseminated peritoneal leiomyomatosis,** presents as multiple small peritoneal nodules. Both are considered benign despite their unusual behavior.

Uterine leiomyomas, even when large or numerous, may be asymptomatic. Common signs and symptoms include abnormal bleeding, urinary frequency due to compression of the bladder, sudden pain from infarction of a large or pedunculated tumor, and impaired fertility. In pregnant women, leiomyomas may increase the frequency of spontaneous abortion, fetal malpresentation, uterine inertia (failure to contract with sufficient force), and postpartum hemorrhage. Malignant transformation to leiomyosarcoma is extremely rare.

Leiomyosarcoma

These uncommon malignant neoplasms are thought to arise from the myometrium or endometrial stromal precursor cells, rather than leiomyomas. In contrast to leiomyomas, leiomyosarcomas have complex, highly variable karyotypes that frequently include deletions. Like leiomyomas, a subset contains *MED12* mutations, a genetic aberration that appears to be virtually unique to uterine smooth muscle tumors.

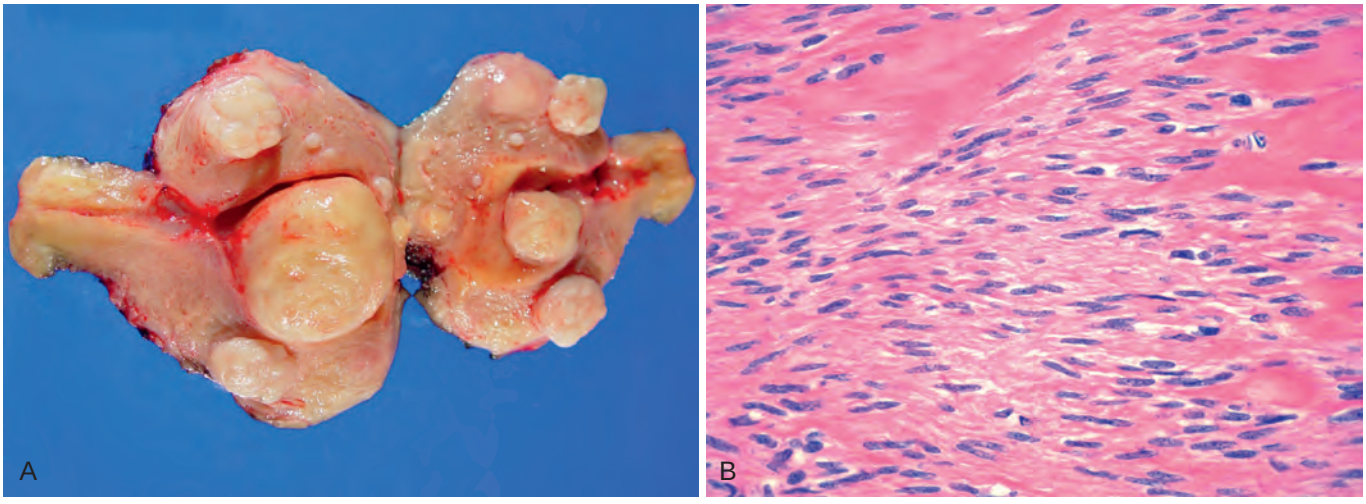


Figure 22.28 Leiomyomas of the uterine myometrium (A) The uterus is opened to reveal multiple tumors in submucosal (bulging into the endometrial cavity), intramural, and subserosal locations that display a firm white appearance on sectioning. (B) Leiomyoma showing well-differentiated, regular, spindle-shaped smooth muscle cells associated with hyalinization.

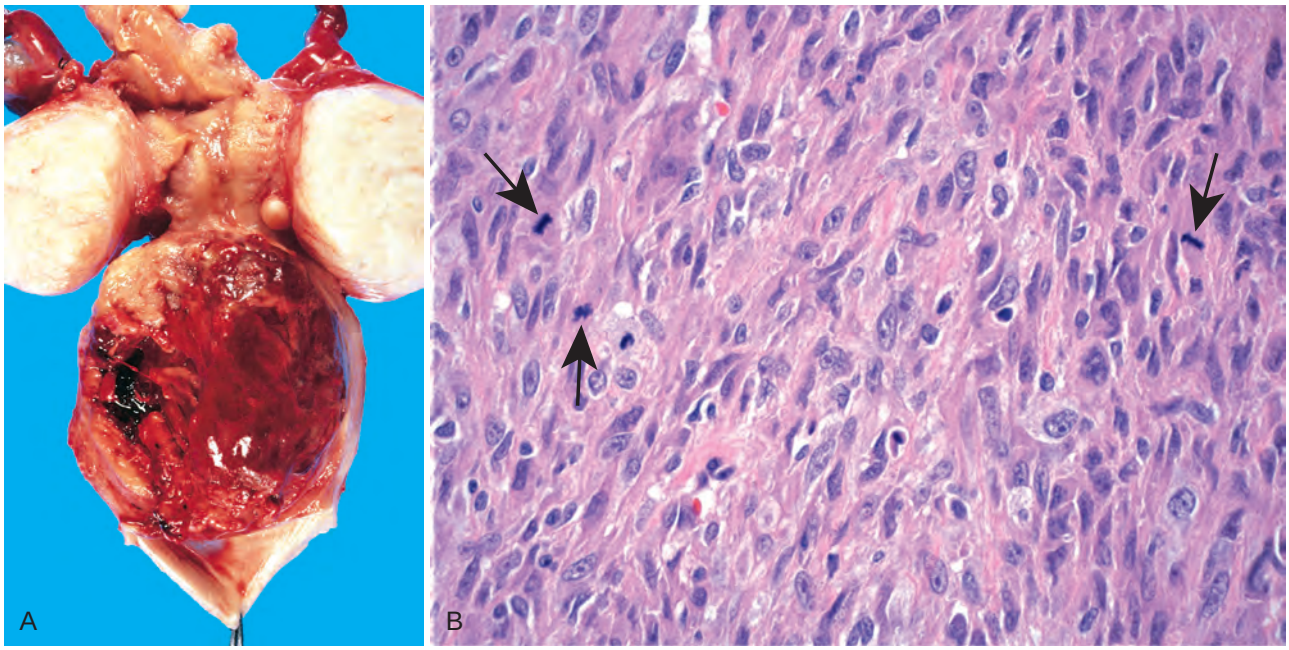


Figure 22.29 Leiomyosarcoma. (A) A large hemorrhagic tumor mass distends the lower corpus and is flanked by two leiomyomas. (B) The tumor cells are irregular in size and have hyperchromatic nuclei. Numerous mitotic figures are present *(arrows)*.

MORPHOLOGY

Leiomyosarcomas grow within the uterus in two somewhat distinctive patterns: (1) bulky, fleshy masses that invade the uterine wall or (2) polypoid masses that project into the uterine lumen (Fig. 22.29A). They exhibit a wide range of cytologic atypia, from extremely well differentiated to highly anaplastic (Fig. 22.29B). The distinction from leiomyoma is based on nuclear atypia, mitotic index, and tumor necrosis. With few exceptions, the presence of 10 or more mitoses per 10 high-power (400×) fields indicates malignancy, particularly if accompanied by cytologic atypia and/or necrosis. If the tumor contains nuclear atypia or large (epithelioid) cells, five mitoses per 10 high-power (400×) fields are sufficient to justify a diagnosis of malignancy. Rare exceptions include mitotically active leiomyomas in young or pregnant women, and caution should be exercised in interpreting such neoplasms as malignant. A proportion of smooth muscle neoplasms may be impossible to classify and are called smooth muscle tumors of "uncertain malignant potential."

Leiomyosarcoma occurs both before and after menopause, with a peak incidence at 40 to 60 years of age. These tumors often recur following surgery, and more than one-half eventually metastasize hematogenously to distant organs, such as lungs, bone, and brain. Dissemination throughout the abdominal cavity is also encountered. The overall 5-year survival rate is about 40%, but the anaplastic lesions have a 5-year survival rate of only 10% to 15%.

KEY CONCEPTS

- Endometrial stromal tumors include stromal nodules, low-grade stromal sarcoma, and high-grade stromal sarcoma.
  - Stromal nodules are benign, well-circumscribed tumors.
  - Low-grade stromal sarcoma resembles stromal nodules but infiltrates into the surrounding myometrium. It is associated with fusion of the *JAZF1* gene and various polycomb factor genes, usually *SUZ12*.
  - High-grade stromal sarcoma shows marked atypia and is associated with other gene fusions.
  - Both low- and high-grade stromal sarcoma are prone to late recurrences.
- Leiomyomas are very common benign smooth muscle tumors that cause significant morbidity, but are not prone to malignant transformation.
- Leiomyosarcoma (a malignant smooth muscle tumor) is an uncommon, highly malignant myometrial tumor that usually arises de novo.

FALLOPIAN TUBES

The most common disorders affecting the fallopian tube are infections and associated inflammatory conditions, followed in frequency by ectopic (tubal) pregnancy and endometriosis.

Inflammations

Suppurative salpingitis may be caused by any pyogenic organism; in some cases more than one organism is involved. *Gonococcus* is the causative organism in more than 60% of cases, with *Chlamydiae* being responsible for many of the remaining cases. These tubal infections are a part of pelvic inflammatory disease, described earlier in this chapter.

Tuberculous salpingitis is rare in the United States, accounting for not more than 1% to 2% of all forms of salpingitis. It is more common, however, in parts of the world where tuberculosis is prevalent and is an important cause of infertility in these areas.

TUMORS AND CYSTS

The most common primary lesions of the fallopian tube (excluding endometriosis) are minute, 0.1- to 2-cm translucent cysts filled with clear serous fluid, called *paratubal cysts*. Larger varieties are found near the fimbriated end of the tube or in the broad ligaments and are referred to as *hydatids of Morgagni*. These cysts, lined by benign serous (tubal type) epithelium, are presumed to arise in remnants of the müllerian duct and are of little significance.

Benign tumors of the fallopian tube are rare and include *adenomatoid tumor* (mesothelioma), which occurs subserosally on the tube or sometimes in the mesosalpinx. These small nodules are counterparts of the adenomatoid tumors that occur in the testes or epididymis (Chapter 21).

Historically, fallopian tube carcinoma was considered to be a rare entity. However, substantial data have accumulated indicating that at least a subset of "serous ovarian cancers" actually arise from the epithelium of the fallopian tube (discussed later). This idea is supported by the frequent identification of serous tubal in situ carcinoma in women at risk for serous carcinoma (e.g., women with germline *BRCA1* mutations) and observations showing that failure to remove the fallopian tubes at the time of oophorectomy is associated with a significant residual risk of "ovarian" cancer. We will return to this issue when discussing ovarian cancer in the next section.

Ovaries

The most common lesions encountered in the ovary are functional or benign cysts and tumors. Neoplastic disorders can be grouped according to their origin from each of the three main ovarian cell types: (1) müllerian epithelium, (2) germ cells, and (3) sex cord–stromal cells. Primary inflammations of the ovary (oophoritis) are uncommon, and on rare occasions may have an autoimmune basis (autoimmune oophoritis); the autoimmune reactions affect the ovarian follicles and may lead to infertility.

NON-NEOPLASTIC AND FUNCTIONAL CYSTS

Follicle and Luteal Cysts

Cystic follicles are very common in the ovary. They originate from unruptured graafian follicles or in follicles that have ruptured and immediately sealed.

MORPHOLOGY

These cysts are usually multiple. They range in size up to 2 cm in diameter, are filled with a clear serous fluid, and are lined by a gray, glistening membrane. On occasion, larger cysts exceeding 2 cm (follicle cysts) may be diagnosed by palpation or ultrasonography; these may cause pelvic pain. Granulosa lining cells are present if the intraluminal pressure has not been so great as to cause their atrophy. The outer theca cells may be conspicuous due to increased amounts of pale cytoplasm (a change referred to as luteinization). As discussed later, when luteinization is pronounced (hyperthecosis), it may be associated with increased estrogen production and endometrial abnormalities.

Luteal cysts (corpora lutea) are present in the normal ovaries of women of reproductive age. They are lined by a rim of bright yellow tissue containing luteinized granulosa cells and are prone to rupture, which may produce a peritoneal reaction. Sometimes the combination of old hemorrhage and fibrosis may make their distinction from endometriotic cysts difficult.

Polycystic Ovaries and Stromal Hyperthecosis

***Polycystic ovarian syndrome* (PCOS) is a complex endocrine disorder characterized by hyperandrogenism, menstrual abnormalities, polycystic ovaries, chronic anovulation, and decreased fertility.** Formerly called Stein Leventhal syndrome, it affects 6% to 10% of reproductive age women worldwide. It is also associated with obesity, type 2 diabetes, and premature atherosclerosis, all of which may be indicative of an underlying metabolic disorder. The etiology of PCOS remains incompletely understood. It is marked by a dysregulation of enzymes involved in androgen biosynthesis and excessive androgen production, which is considered to be a central feature of this disorder. In addition, women with PCOS show insulin resistance and altered adipose tissue metabolism, which contribute to the development of both diabetes and obesity.

The central morphologic abnormality of PCOS is numerous cystic follicles or follicle cysts that enlarge the ovaries. However, polycystic ovaries are detected in 20% to 30% of all women, so this finding is not specific. In addition, due to an increase in free serum estrone levels, women with PCOS are at risk for endometrial hyperplasia and carcinoma.

Stromal hyperthecosis, also called cortical stromal hyperplasia, is a disorder of ovarian stroma most often seen in postmenopausal women, but it may overlap with PCOS in younger women. The disorder is characterized by uniform, usually bilateral, enlargement of the ovary (up to 7 cm), which has a white to tan appearance on sectioning. Microscopic examination shows hypercellular stroma and luteinization of the stromal cells, which are visible as discrete nests of cells with vacuolated cytoplasm. The clinical presentation and effects on the endometrium are similar to those of PCOS, although virilization may be even more striking.

A physiologic condition mimicking the aforementioned syndromes is *theca lutein hyperplasia of pregnancy*. In response to pregnancy hormones (gonadotropins), theca cells proliferate and the perifollicular zone expands. As the follicles regress, the concentric theca-lutein hyperplasia may appear nodular. This change is not to be confused with true luteomas of pregnancy (see later).

OVARIAN TUMORS

There are numerous types of ovarian tumors. About 80% are benign, and these occur mostly in young women between 20 and 45 years of age. So-called borderline tumors (tumors

of indeterminate malignancy) occur at slightly older ages. Malignant tumors are more common in women between 45 and 65 years of age. Ovarian cancer accounts for 3% of all cancers in females and is the fifth most common cause of death due to cancer in women in the United States. **Because most ovarian cancers have spread beyond the ovary and fallopian tube by the time of diagnosis, they account for a disproportionate number of deaths from cancer of the female genital tract.**

Classification

The classification of ovarian tumors given in Table 22.5 is a simplified version of the World Health Organization Histological Classification, which separates ovarian neoplasms according to the most probable tissue of origin. It is now believed that most tumors of the ovary arise ultimately from one of three ovarian components:

Table 22.5 WHO Classification of Ovarian Neoplasms

Surface Epithelial-Stromal Tumors

- Serous tumors
  - Benign (cystadenoma, cystadenofibroma)
  - Borderline (serous borderline tumor)
  - Malignant (low- and high-grade serous adenocarcinoma)
- Mucinous tumors, endocervical-like and intestinal type
  - Benign (cystadenoma, cystadenofibroma)
  - Borderline (mucinous borderline tumor)
  - Malignant (mucinous adenocarcinoma)
- Endometrioid tumors
  - Benign (cystadenoma, cystadenofibroma)
  - Borderline (endometrioid borderline tumor)
  - Malignant (endometrioid adenocarcinoma)
- Clear cell tumors
  - Benign
  - Borderline
  - Malignant (clear cell adenocarcinoma)
- Transitional cell tumors
  - Benign Brenner tumor
  - Brenner tumor of borderline malignancy
  - Malignant Brenner tumor
- Epithelial-stromal
  - Adenosarcoma
  - Malignant mixed müllerian tumor

Sex Cord–Stromal Tumors

- Granulosa tumors
- Fibromas
- Fibrothecomas
- Thecomas
- Sertoli-Leydig cell tumors
- Steroid (lipid) cell tumors

Germ Cell Tumors

- Teratoma
  - Immature
  - Mature
  - Solid
  - Cystic (dermoid cyst)
- Monodermal (e.g., struma ovarii, carcinoid)
- Dysgerminoma
- Yolk sac tumor
- Mixed germ cell tumors

Metastatic Cancer From Non-Ovarian Primary

- Colonic, appendiceal
- Gastric
- Pancreaticobiliary
- Breast

Table 22.6 Frequency of Major Ovarian Tumors

Type	Percentage of Malignant Ovarian Tumors	Percentage That Are Bilateral
Serous		
Benign (60%)	47	25
Borderline (15%)		30
Malignant (25%)		65
Mucinous		
Benign (80%)	3	5
Borderline (10%)		10
Malignant (10%)		<5
Endometrioid carcinoma	20	40
Undifferentiated carcinoma	10	—
Clear cell carcinoma	6	40
Granulosa cell tumor	5	5
Teratoma	1	15
Benign (96%)		Rare
Malignant (4%)		
Metastatic	5	>50
Others	3	—

- Surface/fallopian tube epithelium and endometriosis
- Germ cells, which migrate to the ovary from the yolk sac and are pluripotent
- Stromal cells, including the sex cords, which are forerunners of the endocrine apparatus of the postnatal ovary

There is also a group of miscellaneous tumors, and finally there are secondary or metastatic tumors to the ovary.

Although some of the specific tumors have distinctive features and are hormonally active, most are nonfunctional and produce few symptoms until they reach a large size. Some, principally epithelial tumors, are often bilateral. Table 22.6 lists the tumors and their subtypes. Abdominal pain and distention, urinary and gastrointestinal tract symptoms due to compression by the tumor or cancer invasion, and vaginal bleeding are the most common symptoms. The benign forms may be entirely asymptomatic and occasionally are found unexpectedly on abdominal or pelvic examination or during surgery.

Epithelial Tumors

Most primary ovarian neoplasms arise from müllerian epithelium. The classification of these tumors is based on both differentiation and extent of proliferation of the epithelium. There are three major histologic types based on the differentiation of the neoplastic epithelium: serous, mucinous, and endometrioid tumors. These epithelial proliferations are classified as benign, borderline, and malignant. The benign tumors are often further subclassified based on the components of the tumors, which may include cystic areas (cystadenoma), cystic and fibrous areas (cystadenofibroma), and predominantly fibrous areas (adenofibroma). The borderline tumors and the malignant tumors can also have a cystic component, and when malignant they are sometimes referred to as *cystadenocarcinoma*. The tumors can be relatively small, or they can grow to fill the entire pelvis before they are detected.

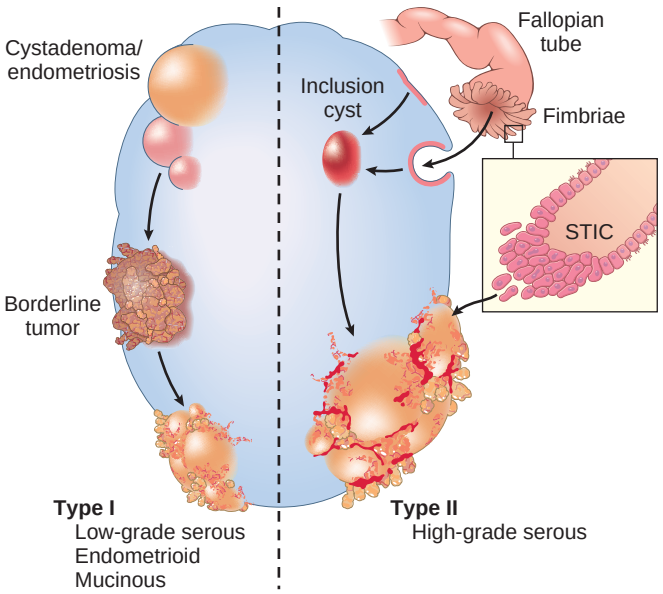


Figure 22.30 Schematic diagram of the pathogenesis of ovarian epithelial tumors. Type I tumors progress from benign tumors through borderline tumors that may give rise to a low-grade carcinoma. These include low-grade serous, endometrioid, and mucinous carcinomas. Type II tumors arise from inclusions cysts/fallopian tube epithelium via intraepithelial precursors that are often not identified. They demonstrate high-grade features and are most commonly of serous histology. *STIC*, Serous tubal intraepithelial carcinoma.

Clinicopathologic and molecular studies have suggested that ovarian carcinoma may be broadly categorized into two different types, referred to as type I and type II (Fig. 22.30). Type I carcinomas are low-grade tumors that often arise in association with borderline tumors or endometriosis. These tumors encompass several histologic subtypes, including low-grade serous, endometrioid, and mucinous tumors, as discussed later. Type II tumors are most often high-grade serous carcinomas that arise from serous intraepithelial carcinoma (see later).

Serous Tumors

These cystic neoplasms include the most common malignant ovarian tumors and account for approximately 40% of all cancers of the ovary. Although the term *serous* appropriately describes the cyst fluid, it has become synonymous with the tubal-like epithelium in these tumors. Together, benign, borderline, and malignant serous tumors account for about 30% of all ovarian tumors and just over 50% of ovarian epithelial tumors. About 70% are benign or borderline, and 30% are malignant. Benign and borderline tumors are most common between 20 and 45 years of age. Serous carcinomas occur later in life on average, but often occur at earlier ages in familial cases.

Pathogenesis

Little is known about the risk factors for benign and borderline tumors. Risk factors for malignant serous tumors (serous carcinomas) are also not completely understood, but nulliparity, family history, and heritable mutations play a role in tumor development. There is a higher frequency of carcinoma in women with low parity. Women 40 to 59 years of age who have taken oral contraceptives or undergone tubal ligation have a reduced risk of developing ovarian cancer. The most intriguing risk factors are genetic. As discussed in Chapters 7 and 23, inherited germline mutations in both *BRCA1* and *BRCA2* increase susceptibility to both ovarian cancer and breast cancer. *BRCA1* mutations are present in about 5% of patients younger than 70 years of age with ovarian cancer. The estimated risk of ovarian cancer in women bearing *BRCA1* or *BRCA2* mutations is 20% to 60% by 70 years of age.

Serous ovarian carcinoma is divided into two major groups: (1) low-grade carcinoma and (2) high-grade carcinoma. This distinction is made based on the degree of nuclear atypia and correlates with patient survival. Low-grade carcinomas may arise in association with serous borderline tumors, while high-grade carcinomas arise from in situ lesions in the fallopian tube fimbriae or from serous inclusion cysts within the ovary.

The concept of a fallopian tube origin for high-grade serous carcinomas was initially sparked by the observation that women with *BRCA1/2* germline mutations were often discovered at the time of prophylactic salpingo-oophorectomy to have areas of marked epithelial atypia in their fallopian tubes. The lesions, called serous tubal intraepithelial carcinoma (STIC), have since been described in association with sporadic high-grade serous ovarian cancers, suggesting that at least some high-grade serous carcinomas arise from the fallopian tube. What then is the origin of high-grade serous carcinomas that involve the ovary, without concomitant involvement of the fallopian tube? Historically, the source of these tumors was hypothesized to be cortical inclusion cysts (Fig. 22.31), which were thought to arise through invagination of the surface epithelium, followed by serous metaplasia. A recent alternative idea is that the cysts arise from implantation of detached fallopian tube epithelium at sites where ovulation has disrupted the surface of the ovary (see Fig. 22.30).

The percentage of sporadic high-grade serous carcinomas that arise in the fallopian tube or from ovarian inclusion cysts is currently uncertain, as is the origin of cortical inclusion cysts. However, this paradigm shift has already altered the management of women at high-risk for ovarian

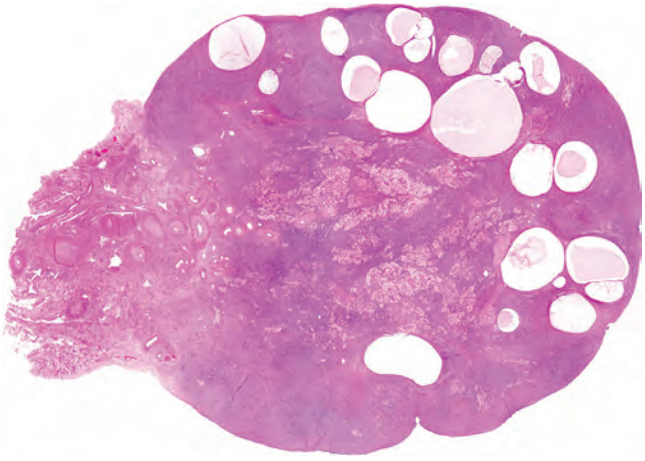


Figure 22.31 Cortical inclusion cysts of the ovary.

carcinoma (*BRCA* mutation carriers and women with a strong family history of breast/ovarian cancer), as these women now undergo salpingo-oophorectomy, instead of simple oophorectomy.

Regardless of their origin, studies have shown that low- and high-grade serous carcinoma have distinct mutational profiles, as follows:

- Low-grade tumors arising in serous borderline tumors have mutations in the *KRAS, BRAF,* or *ERBB2* oncogenes, and usually have wild-type *TP53* genes.
- High-grade tumors have a high frequency of *TP53* mutations and lack mutations in either *KRAS* or *BRAF*. Genomic imbalances are very common and include amplifications of a number of oncogenes (e.g., *PIK3CA,* the gene encoding the catalytic subunit of PI3K) and deletions of tumor suppressor genes (e.g., *RB*). Almost all ovarian carcinomas arising in women with *BRCA1* or *BRCA2* mutations are high-grade serous carcinomas with *TP53* mutations. Curiously, *BRCA1* and *BRCA2* mutations are rare in sporadic high-grade serous carcinoma.

MORPHOLOGY

Serous tumors may present as a multicystic lesion in which papillary epithelium is contained within fibrous walled cysts (intracystic) (Fig. 22.32A) or as a mass projecting from the ovarian surface. Benign tumors typically have a smooth glistening cyst wall with no epithelial thickening or with small papillary projections. Borderline tumors contain an increased number of papillary projections (Fig. 22.32A and C). Larger areas of solid or papillary tumor growth, tumor irregularity, and fixation or nodularity of the capsule are features associated with malignancy (Fig. 22.32B). Bilaterality is common, occurring in 20% of benign serous cystadenomas, 30% of serous borderline tumors, and approximately 66% of serous carcinomas. A significant proportion of borderline and malignant serous tumors involve the surface of the ovary (Fig. 22.32C).

Microscopically, the cysts are lined by columnar epithelium. In benign tumors (Fig. 22.33A), the epithelial cells retain abundant cilia, and microscopic papillae may be found. **Serous borderline tumors** exhibit increased complexity of the stromal papillae, stratification of the epithelium, and mild nuclear atypia, but invasion of the stroma is not seen (Fig. 22.33B). The epithelial cells often grow in a delicate, papillary pattern referred to as "micropapillary carcinoma," which is thought to be the precursor to **low-grade serous carcinoma** (Fig. 22.33C). **High-grade serous carcinoma** is distinguished by having more complex growth patterns and widespread infiltration or frank effacement of the underlying stroma (Fig. 22.33D). The individual tumor cells display marked nuclear atypia, including pleomorphism, atypical mitotic figures, and multinucleation; occasionally, tumors may be so undifferentiated that serous features are no longer recognizable. Serous tubal intraepithelial carcinoma consists of cells morphologically identical to high-grade serous carcinoma that do not invade the underlying stroma. Concentric calcifications (psammoma bodies) are common in all types of serous tumors, but are not specific for neoplasia.

Ovarian serous tumors, both low- and high-grade, have a propensity to spread to the peritoneal surfaces and omentum and are commonly associated with the presence of ascites. As with other tumors, the extent of the spread outside the ovary determines the stage of the disease.

The biologic behavior of serous tumors depends on the degree of differentiation and the extent and distribution of peritoneal disease. Serous tumors may occur on the surface of the ovaries; if unencapsulated, such tumors are likely to spread to the peritoneum. As discussed, at least some of these carcinomas likely originate from the fallopian tube, from which they may exfoliate and spread to the peritoneum. In rare instances, they appear to originate on the peritoneal surface (referred to as primary peritoneal serous carcinoma).

Prognosis is closely related to the histologic appearance of the tumor and the presence and extent of peritoneal disease. Borderline serous tumors may arise from or secondarily involve the peritoneal surfaces as noninvasive implants, remaining localized and causing no symptoms, or slowly spread, producing intestinal obstruction or other complications after many years. Similarly, low-grade carcinomas, even after spread outside the ovary, often progress slowly, and patients may survive for relatively long periods before dying of disease. In contrast, high-grade tumors are often widely metastatic throughout the abdomen at the time

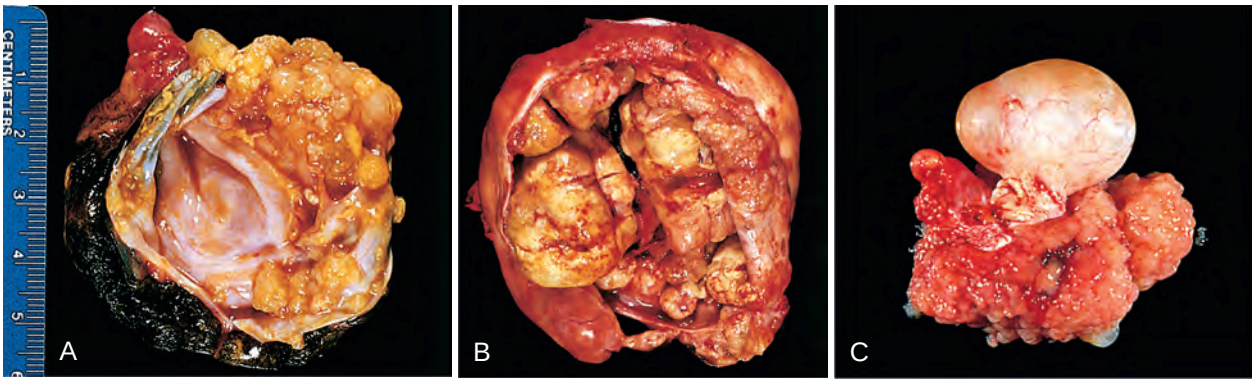


Figure 22.32 Gross appearances of serous tumors of the ovary. (A) Serous borderline tumor opened to display a cyst cavity lined by delicate papillary tumor growths. (B) Carcinoma. The cyst is opened to reveal a large, bulky tumor mass. (C) Another borderline tumor growing on the ovarian surface *(lower).*

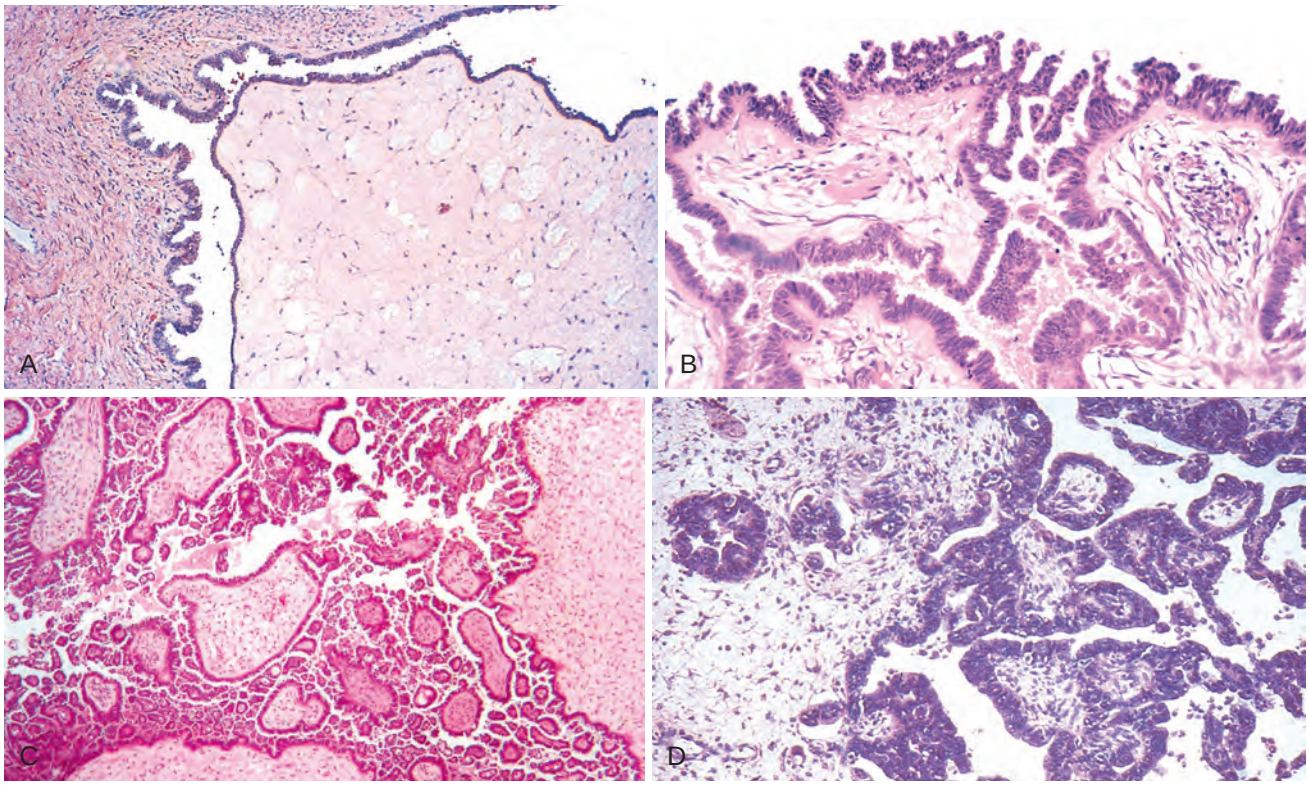


Figure 22.33 Microscopic appearances of serous tumors of the ovary. (A) Serous cystadenoma revealing stromal papillae with a columnar epithelium. (B) Borderline serous tumor showing increased architectural complexity and epithelial cell stratification. (C) Complex micropapillary growth defines a low-grade "micropapillary" serous carcinoma. (D) High-grade serous carcinoma of the ovary with invasion of underlying stroma.

of presentation, a picture associated with rapid clinical deterioration. Consequently, grading of the tumor, even after extension to the peritoneum, impacts both prognosis and selection of therapy. The 5-year survival rate for borderline and malignant tumors confined to the ovary is 100% and 70%, respectively, whereas the 5-year survival rate for the same tumors involving the peritoneum is about 90% and 25%, respectively. Because of their protracted course, borderline tumors may recur after many years, and 5-year survival is not synonymous with cure.

Mucinous Tumors

Mucinous tumors account for about 20% to 25% of all ovarian neoplasms. They occur principally in middle adult life and are rare before puberty and after menopause. The vast majority are benign or borderline tumors. Primary ovarian mucinous carcinomas are uncommon and account for approximately 3% of all ovarian cancers.

Pathogenesis

Mutation of the *KRAS* proto-oncogene is a consistent genetic alteration in mucinous tumors of the ovary, including the majority of benign mucinous cystadenomas (58%), mucinous borderline tumors (75% to 86%), and ovarian mucinous carcinomas (85%). Interestingly, one study showed that several tumors with distinct areas of epithelium showing benign, borderline, and carcinoma had identical *KRAS* mutations in each area. Thus, *KRAS* mutations may initiate the development of these neoplasms.

MORPHOLOGY

Mucinous tumors differ from the serous variety in several ways. The surface of the ovary is rarely involved, and only 5% of primary mucinous cystadenomas and mucinous carcinomas are bilateral. Mucinous tumors also tend to produce larger cystic masses; some have been recorded with weights of more than 25 kg. They are multiloculated tumors filled with sticky, gelatinous fluid rich in glycoproteins (Fig. 22.34A).

Microscopically, benign mucinous tumors are characterized by a lining of tall, columnar epithelial cells with apical mucin that lack cilia. The vast majority demonstrate gastric or intestinal type differentiation; uncommonly, tumors may show endocervical type mucinous differentiation instead (Fig. 22.34B). Mucinous borderline tumors are distinguished from cystadenomas by epithelial stratification, tufting, and/or papillary intraglandular growth, often producing an appearance strikingly similar to tubular adenomas or villous adenomas of the intestine. **Mucinous carcinoma** characteristically demonstrates confluent glandular growth that is now recognized as a form of "expansile" invasion. Some authors use the term intraepithelial carcinomas for tumors with marked epithelial atypia that lack invasive features. Approximate 10-year survival rates for stage I noninvasive "intraepithelial carcinoma" and for frankly invasive malignant tumors are greater than 95% and 90%, respectively. Mucinous carcinomas that have spread beyond the ovary are usually fatal, but as stated earlier, these tumors are uncommon and must be distinguished from metastatic mucinous adenocarcinomas.

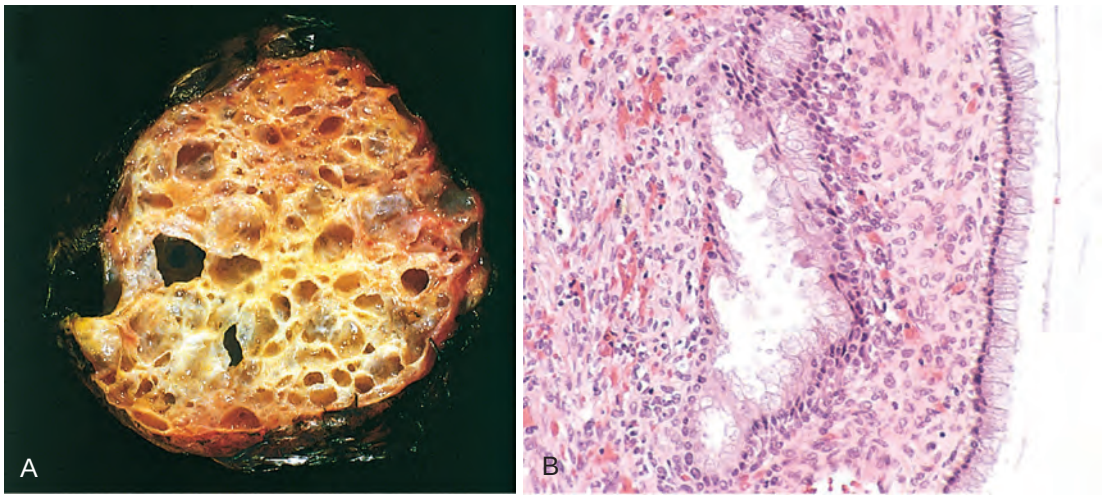


Figure 22.34 Mucinous cystadenoma (A) Note the multicystic appearance, delicate septa, and the presence of glistening mucin within the cysts. (B) Columnar cells lining the cysts.

A clinical condition referred to as *pseudomyxoma peritonei* is marked by mucinous ascites, cystic epithelial implants on the peritoneal surfaces, adhesions, and frequent involvement of the ovaries (Fig. 22.35). Pseudomyxoma peritonei, if extensive, may result in intestinal obstruction and death. Historically, it was thought that many cases in women were due to the spread of primary ovarian mucinous neoplasms. However, recent evidence points in almost all cases to an extra-ovarian source, usually the appendix (Chapter 17). Because the majority of primary mucinous ovarian tumors are unilateral, bilateral presentation of mucinous tumors always requires exclusion of metastasis from a non-ovarian primary tumor.

Endometrioid Ovarian Tumors

Endometrioid carcinoma accounts for approximately 10% to 15% of all ovarian cancers. Benign endometrioid tumors, called *endometrioid adenofibroma,* and borderline endometrioid tumors also occur, but are uncommon. Approximately 15% to 30% of ovarian endometrioid carcinomas are accompanied by carcinoma of the endometrium; although such cases have a relatively good prognosis, molecular data suggest that in most instances of concomitant endometrial and ovarian disease, the ovarian disease arises from metastatic spread of a primary endometrial tumor.

Pathogenesis

In about 15% to 20% of cases, endometrioid carcinoma coexists with endometriosis. The peak incidence of tumors associated with endometriosis is a decade earlier than that of endometrioid carcinomas that are not associated, suggesting that ovarian endometriosis serves as a precursor to ovarian endometrioid carcinoma in some instances. Molecular studies have found striking similarities to endometrial endometrioid carcinoma; shared features include relatively frequent alterations that increase PI3K/AKT pathway signaling (mutations in *PTEN, PIK3CA, ARID1A,* and *KRAS*) and mutations in mismatch DNA repair genes. Also similar to endometrioid carcinomas of the endometrium, *TP53* mutations are common in poorly differentiated tumors.

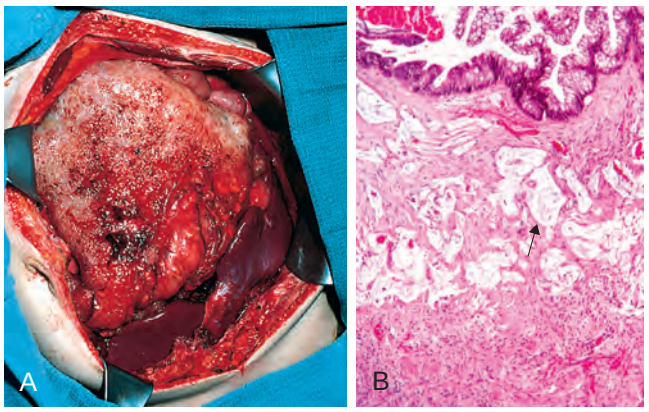


Figure 22.35 Pseudomyxoma peritonei. (A) View at laparotomy revealing massive overgrowth of a gelatinous metastatic tumor. (B) Histology of peritoneal implants from an appendiceal tumor, showing mucin-producing epithelium and free mucin *(arrow)*. (A, Courtesy Dr. Paul H. Sugarbaker, Washington Hospital Cancer Center, Washington, DC.)

MORPHOLOGY

Endometrioid tumors are distinguished from serous and mucinous tumors by the presence of tubular glands resembling benign or malignant endometrium. Endometrioid carcinomas typically present with solid and cystic areas of growth. Forty percent involve both ovaries, and such bilaterality usually implies extension of the neoplasm beyond the genital tract. These are low-grade tumors that reveal glandular patterns bearing a strong resemblance to those of endometrial origin. The 5-year survival rate for patients with stage I tumors is approximately 75%.

Clear Cell Carcinoma

Benign and borderline clear cell tumors are exceedingly rare, and clear cell carcinomas are uncommon. They are composed of large epithelial cells with abundant clear cytoplasm, an appearance that resembles hypersecretory

gestational endometrium. Because these tumors sometimes occur in association with endometriosis or endometrioid carcinoma of the ovary and resemble clear cell carcinoma of the endometrium, they are now thought to be variants of endometrioid adenocarcinoma. In line with this idea, the most common genetic aberrations (*PIK3CA, ARID1A, KRAS, PTEN,* and *TP53*) are shared with endometrioid carcinoma, albeit at somewhat different frequencies. Clear cell tumors of the ovary can be predominantly solid or cystic. In the solid neoplasms, the clear cells are arranged in sheets or tubules, while in the cystic variety, the neoplastic cells line the spaces. Clear cell carcinoma confined to the ovaries has a 90% 5-year survival, but in advanced stage disease it appears that clear cell morphology portends a poor outcome. Clear cell carcinoma is treated like other types of ovarian carcinoma.

Cystadenofibroma

Cystadenofibroma is an uncommon variant in which there is pronounced proliferation of the fibrous stroma that underlies the columnar lining epithelium. These benign tumors are usually small and multilocular and have simple papillary processes that are not as complicated and branching as those found in the ordinary cystadenoma. They may contain mucinous, serous, endometrioid, or transitional (Brenner tumor) epithelium. Borderline lesions with cellular atypia and, rarely, tumors with focal areas of carcinoma occur, but metastatic spread of either is extremely uncommon.

Transitional Cell Tumors

Transitional cell tumors contain neoplastic epithelial cells resembling urothelium and are usually benign. They comprise roughly 10% of ovarian epithelial tumors and are also referred to as *Brenner tumors*. Uncommon transitional cell carcinomas also occur in the ovary.

MORPHOLOGY

Brenner tumors may be solid or cystic, are usually unilateral (approximately 90%), and vary in size from small lesions less than 1 cm in diameter to massive tumors up to 20 to 30 cm in diameter (Fig. 22.36A). The fibrous stroma, resembling that of the normal ovary, is marked by sharply demarcated nests of epithelial cells resembling the epithelium of the urinary tract, often with mucinous glands in their center (Fig. 22.36B). Infrequently, the stroma is composed of plump fibroblasts resembling theca cells; such neoplasms may have hormonal activity. Most Brenner tumors are benign, but borderline and malignant counterparts have been reported. Tumors with benign Brenner nests admixed with malignant tumor cells are referred to as malignant Brenner tumors, while tumors with greater than 50% malignant transitional type epithelium are considered transitional cell carcinomas of the ovary.

Brenner tumors are often detected incidentally and even when large behave in a benign fashion. Malignant Brenner tumors generally present in stage 1 and for prognostic purposes are considered to be equivalent to low-grade (type 1) carcinomas. The uncommon transitional cell carcinomas are considered to be equivalent to high-grade (type II) ovarian carcinomas; these often present at advanced stage and are treated like high-grade serous carcinomas.

Clinical Course, Detection, and Prevention of Ovarian Epithelial Tumors

All ovarian carcinomas produce similar clinical manifestations, most commonly lower abdominal pain and abdominal enlargement. Gastrointestinal complaints, urinary frequency, dysuria, pelvic pressure, and many other symptoms may appear. Benign lesions are easily resected and cured. The malignant tumors tend to cause progressive weakness, weight loss, and cachexia. Once carcinomas extend through the capsule of the tumor and seed the peritoneal cavity, they often cause ascites, which may be massive. Characteristically, the ascitic fluid is filled with exfoliated tumor cells. The peritoneal pattern of spread is distinctive: serosal surfaces are diffusely seeded with 0.1- to 0.5-cm nodules of tumor that only rarely invade deeply into the underlying parenchyma. The regional nodes are often involved, and metastases may be found in the liver, lungs, gastrointestinal tract, and elsewhere. Metastasis across the midline to the opposite ovary is discovered in about one-half the cases at surgery and heralds a progressive downhill course and death within a few months or years.

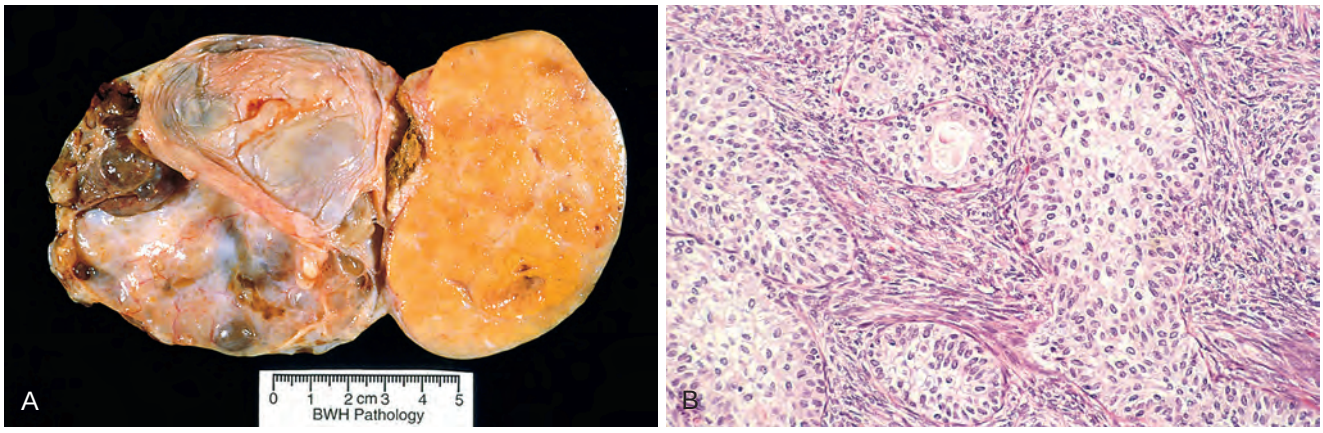


Figure 22.36 Brenner tumor (A) Brenner tumor *(right)* associated with a benign cystic teratoma *(left)*. (B) Histologic detail of characteristic epithelial nests within the ovarian stroma. (Courtesy Dr. M. Nucci, Brigham and Women's Hospital, Boston, Mass.)

Most women with ovarian carcinoma present with high-stage disease. This is the primary reason for the relatively poor 5- and 10-year survival rates of patients with these tumors, as compared with rates for patients with cervical or endometrial carcinoma. For these reasons, development of new assays that permit early diagnosis is a top priority. Biochemical tests for tumor antigens or tumor products in the plasma of these patients are being sought vigorously, but none proposed to date has sufficient sensitivity and specificity to be useful. The serum marker CA-125 is used in patients with known disease to monitor disease recurrence/progression.

Prevention of ovarian cancer also remains an elusive goal. Screening to identify women at risk (positive for *BRCA* mutations or with strong family histories) and treatment with risk-reducing salpingo-oophorectomy are currently standard. The long-term impact of these approaches in women at high risk remains to be determined; recent data suggest that there remains a 3% to 4% risk of developing ovarian cancer after salpingo-oophorectomy.

KEY CONCEPTS

- Epithelial ovarian tumors are classified into benign, borderline, or malignant.
- About 80% of all ovarian epithelial tumors are benign and occur in young women. The malignant tumors occur most commonly in older women and account for approximately 3% of all cancers in women in the United States.
- The majority of the malignant epithelial tumors are high-grade serous carcinomas, which have a poor prognosis in large part because they are usually detected after they have spread beyond the ovary and/or the fallopian tube.
- There are three major histologic types of epithelial ovarian tumors: serous, mucinous, and endometrioid, all of which have a benign, borderline, and malignant category.
- Benign tumors are composed of well-differentiated epithelial cells with minimal proliferation. Borderline tumors show increased cell proliferation, but lack stromal invasion. Malignant tumors show increased epithelial atypia and are defined by the presence of stromal invasion.
- Ovarian carcinomas are currently divided into type I (low-grade) and type II (high-grade) tumors.
- The origin of ovarian tumors is still under investigation, but it appears that *BRCA1*- and *BRCA2*-related tumors as well as the majority of sporadic ovarian serous tumors likely arise from fallopian tube epithelium rather than ovarian epithelium.

Germ Cell Tumors

Germ cell tumors constitute 15% to 20% of all ovarian tumors and include multiple subtypes (Fig. 22.37). Most are *benign cystic teratomas,* but others, found principally in children and young adults, may show malignant behavior and pose problems in histologic diagnosis and in therapy. They bear a high degree of morphologic and molecular similarity to germ cell tumors in the male testis (Chapter 21).

Teratoma

Teratomas are divided into three categories: (1) mature (benign), (2) immature (malignant), and (3) monodermal or highly specialized.

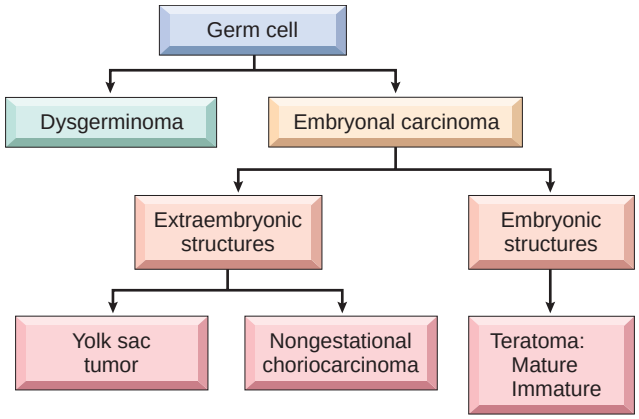


Figure 22.37 Histogenesis and interrelationships of ovarian tumors of germ cell origin.

Mature (Benign) Teratomas

Most benign teratomas are cystic and are often referred to as *dermoid cysts* because they are almost always lined by skinlike structures. Cystic teratomas are usually found in young women. They may be discovered incidentally, but are occasionally associated with clinically important paraneoplastic syndromes, such as *inflammatory limbic encephalitis,* which may remit following removal of the tumor.

MORPHOLOGY

Benign teratoma is bilateral in 10% to 15% of cases. Characteristically it consists of a unilocular cyst containing hair and sebaceous material (Fig. 22.38). Sectioning reveals a thin wall lined by opaque, gray-white, wrinkled epidermis, frequently with protruding hair shafts. Within the wall, it is common to find grossly evident tooth structures and areas of calcification.

Microscopically, the cyst wall is composed of stratified squamous epithelium with underlying sebaceous glands, hair shafts, and other skin adnexal structures (Fig. 22.39). In most cases, tissues from other germ layers can be identified, such as cartilage, bone, thyroid, and neural tissue. Dermoid cysts are sometimes incorporated within the wall of a mucinous cystadenoma. **About 1% of the**

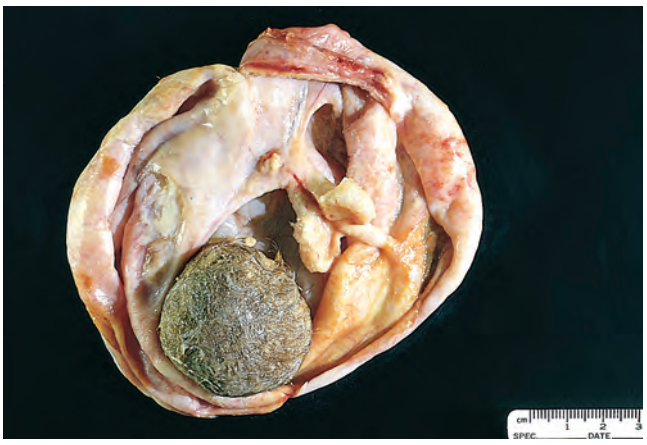


Figure 22.38 Opened mature cystic teratoma (dermoid cyst) of the ovary. Hair *(bottom)* and a mixture of tissues are evident.

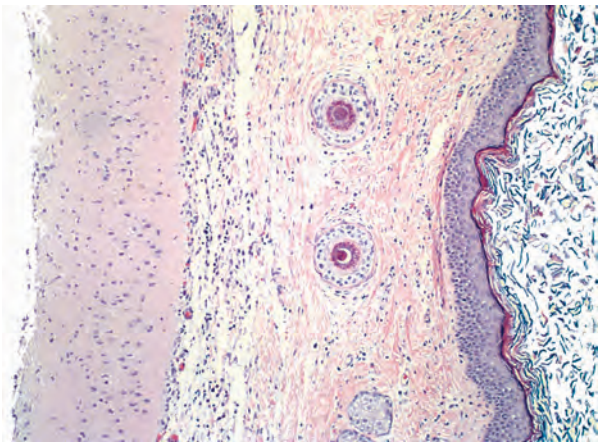


Figure 22.39 Benign cystic teratoma. Low-power view of skin *(right edge)*, beneath which there is brain tissue *(left edge)*.

dermoid cysts undergo malignant transformation, most commonly to squamous cell carcinoma, but also to other cancers (e.g., thyroid carcinoma, melanoma).

In rare instances, a benign teratoma is solid and composed entirely of benign-looking heterogeneous collections of tissues and organized structures derived from all three germ layers. These tumors presumably have the same histogenetic origin as dermoid cysts but lack preponderant differentiation into ectodermal derivatives. Such neoplasms may be difficult to distinguish from malignant immature teratomas on gross inspection.

The origin of teratomas has been a matter of fascination for centuries. Some common beliefs blamed witches, nightmares, or adultery with the devil. The karyotype of almost all benign ovarian teratomas is 46,XX. Genetic analyses indicate that the majority of teratomas arise from an ovum after the first meiotic division, while the remainder arise before the first division.

Monodermal or Specialized Teratomas

Specialized teratomas are a rare but remarkable group of tumors, the most common of which are struma ovarii and carcinoid. They are always unilateral, although a contralateral teratoma may be present. Struma ovarii is composed entirely of mature thyroid tissue, which may be functional and cause hyperthyroidism. Carcinoid tumors, which presumably arise from intestinal tissue found in teratomas, may also be functional; particularly if large (>7 cm), they can produce sufficient 5-hydroxytryptamine to cause the carcinoid syndrome even in the absence of hepatic metastases because ovarian veins connect directly to the systemic circulation. Primary ovarian carcinoid must be distinguished from metastatic intestinal carcinoid, which virtually always involves the ovaries bilaterally. Even rarer is strumal carcinoid, a combination of struma ovarii and carcinoid in the same ovary. Only about 2% of carcinoids in teratomas metastasize.

Immature Malignant Teratomas

These are rare tumors that differ from benign teratomas in that the component tissues resemble embryonal and immature fetal tissue. The tumor is found chiefly in prepubertal adolescents and young women, the mean age being 18 years.

MORPHOLOGY

Immature malignant teratomas are bulky, have a smooth external surface, and tend to be solid on sectioning. Hair, sebaceous material, cartilage, bone, and calcification may be present, along with areas of necrosis and hemorrhage. On microscopic examination, there are varying amounts of immature neuroepithelium, cartilage, bone, muscle, and other elements. An important risk for subsequent extraovarian spread is the histologic grade (I to III) of the tumor. Grading is based on the proportion of the tumor that is comprised of immature neuroepithelium (Fig. 22.40).

Immature teratomas grow rapidly, frequently penetrate the capsule, and spread either locally or distantly. Stage I tumors, particularly those with low-grade (grade 1) histology, have an excellent prognosis. Higher-grade tumors confined to the ovary are generally treated with adjuvant chemotherapy. Most recurrences develop in the first 2 years, and absence of disease beyond this period carries an excellent chance of cure.

Dysgerminoma

Dysgerminoma is the ovarian counterpart of testicular seminoma. Dysgerminoma accounts for about 2% of ovarian cancers and roughly 50% of malignant ovarian germ cell tumors. They may occur in childhood, but 75% occur in the second and third decades of life. Some occur in patients with gonadal dysgenesis, including pseudohermaphroditism. Most of these tumors have no endocrine function. A few produce elevated levels of chorionic gonadotropin, a finding that correlates with the presence of syncytiotrophoblastic giant cells. Like seminoma, dysgerminoma expresses stem cell markers such as OCT3, OCT4, and NANOG, transcription factors implicated in maintenance of pluripotency. They also express the receptor tyrosine kinase KIT, and approximately one-third have activating mutations in the *KIT* gene. These proteins are useful diagnostic markers, and mutated,

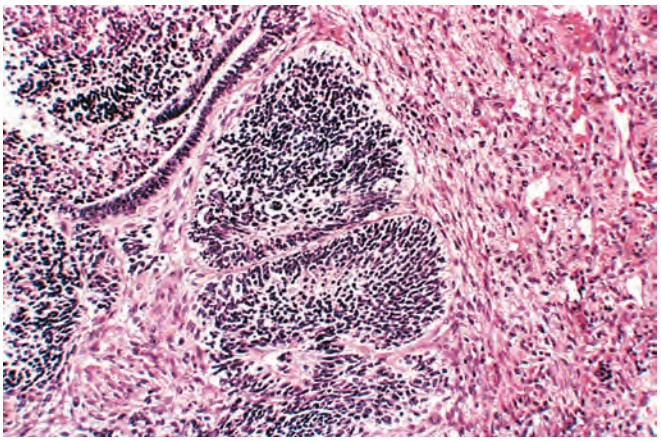


Figure 22.40 Immature teratoma of the ovary illustrating primitive neuroepithelium.

constitutively active KIT also represents a potential therapeutic target.

MORPHOLOGY

Most dysgerminomas (80% to 90%) are unilateral tumors ranging in size from barely visible nodules to masses that virtually fill the abdomen. On cut surface they have a solid yellow-white to gray-pink appearance and are often soft and fleshy. Like seminoma, it is composed of large vesicular cells with clear cytoplasm, well-defined cell boundaries, and centrally placed regular nuclei. The tumor cells grow in sheets or cords separated by scant fibrous stroma (Fig. 22.41), which is infiltrated by lymphocytes and may contain noncaseating granulomas. On occasion, small nodules of dysgerminoma are encountered in the wall of an otherwise benign cystic teratoma; conversely, a predominantly dysgerminomatous tumor may contain a small cystic teratoma.

All dysgerminomas are malignant, but the degree of histologic atypia is variable, and only about one-third are aggressive. A unilateral tumor that has not broken through the capsule or spread outside the ovary has an excellent prognosis (up to 96% cure rate) after simple salpingo-oophorectomy. These neoplasms are responsive to chemotherapy, and even those that have extended beyond the ovary can often be cured. Overall survival exceeds 80%.

Yolk Sac Tumors

Though rare, yolk sac tumor (also known as *endodermal sinus tumor*) still ranks as the second most common malignant ovarian tumor of germ cell origin. It is thought to be derived from malignant germ cells that are differentiating along the extraembryonic yolk sac lineage (see Fig. 22.37). Similar to the normal yolk sac, the tumor cells elaborate *α-fetoprotein*. Its characteristic histologic feature is a glomerulus-like structure composed of a central blood vessel enveloped by tumor cells within a space that is also lined by tumor cells *(Schiller-Duval body)* (Fig. 22.42). Conspicuous intracellular and extracellular hyaline droplets are typically present, some of which stain for α-fetoprotein by immunoperoxidase techniques.

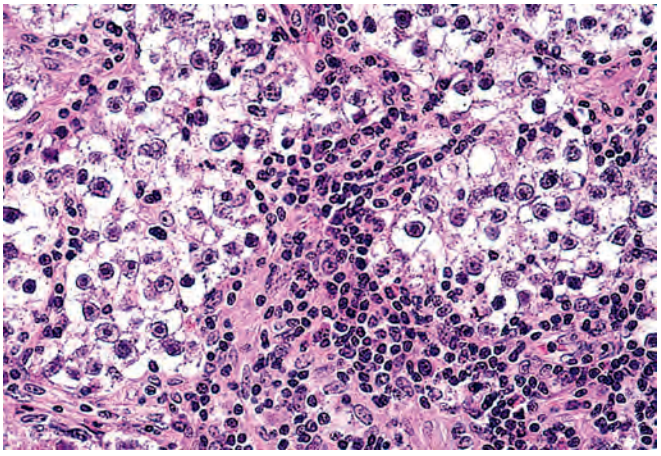


Figure 22.41 Dysgerminoma showing polyhedral tumor cells with round nuclei and adjacent inflammation.

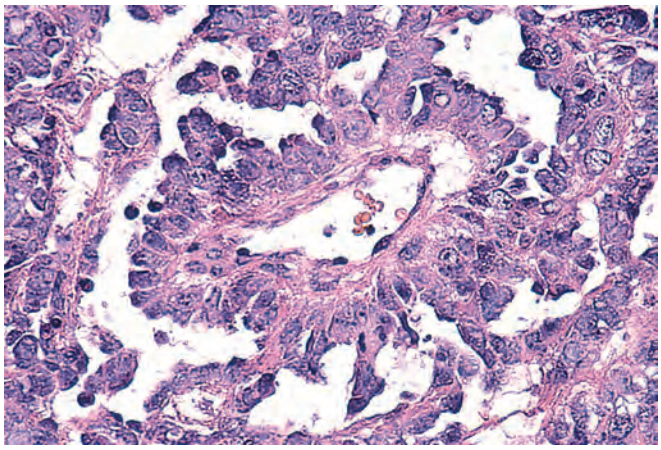


Figure 22.42 A Schiller-Duval body in a yolk sac carcinoma.

Most patients are children or young women presenting with abdominal pain and a rapidly growing pelvic mass that involves a single ovary. With combination chemotherapy, there is greater than 80% survival, independent of disease stage.

Choriocarcinoma

More commonly of placental origin, ovarian choriocarcinoma, like yolk sac tumor, is an example of a malignant germ cell that exhibits extraembryonic differentiation. Most ovarian choriocarcinomas exist in combination with other germ cell tumors, and pure choriocarcinoma is extremely rare. They are histologically identical to the more common placental lesions (described later). The ovarian tumors are aggressive and have usually metastasized hematogenously to the lungs, liver, bone, and other sites by the time of diagnosis. Like all choriocarcinomas, they elaborate high levels of *chorionic gonadotropins,* which may be helpful in establishing the diagnosis or detecting recurrences. In contrast to choriocarcinoma arising in placental tissue, those arising in the ovary are generally unresponsive to chemotherapy and are often fatal.

Other Germ Cell Tumors

These include (1) *embryonal carcinoma,* a highly malignant tumor of primitive embryonal elements that is histologically similar to embryonal carcinoma arising in the testes (Chapter 21); (2) *polyembryoma,* a malignant tumor containing so-called embryoid bodies; and (3) *mixed germ cell tumors* containing various combinations of dysgerminoma, teratoma, yolk sac tumor, and choriocarcinoma.

KEY CONCEPTS

GERM CELL TUMORS

- Germ cell tumors constitute 15% to 20% of ovarian tumors.
- The majority are mature cystic teratomas (dermoid cysts) in women of reproductive age.
- The remainder occur in young women and children; in these age groups, malignant tumors dominate.

- Immature teratomas are distinguished from mature teratomas by the presence of embryonal or fetal elements, most often consisting of primitive neuroepithelium.
- Germ cell tumors show various lines of differentiation toward oogonia (dysgerminoma), extra-embryonic yolk sac (yolk sac tumors), placenta (choriocarcinoma), or multiple germ layers (teratoma).

Sex Cord–Stromal Tumors

These ovarian neoplasms arise from the ovarian stroma, which is derived from the sex cords of the embryonic gonad. The undifferentiated gonadal mesenchyme eventually produces specific types of cells in male (Sertoli and Leydig) and female (granulosa and theca) gonads, and tumors resembling all of these cell types can be identified in the ovary. Moreover, because some of these cells normally secrete estrogens (granulosa and theca cells) or androgens (Leydig cells), their corresponding tumors may be either feminizing (granulosa/theca cell tumors) or masculinizing (Leydig cell tumors).

Granulosa Cell Tumors

Granulosa cell tumors are composed of cells that resemble granulosa cells of a developing ovarian follicle. They are broadly divided into adult and juvenile granulosa cell tumors, based largely on the age of the patient. Collectively, these neoplasms account for about 5% of all ovarian tumors, and 95% of granulosa cell tumors are of the adult type. Although they may be discovered at any age, approximately two-thirds occur in postmenopausal women.

MORPHOLOGY

Granulosa cell tumors are usually unilateral and vary from microscopic foci to large, solid, and cystic encapsulated masses. Tumors that are hormonally active have a yellow coloration to their cut surfaces, due to intracellular lipids.

The granulosa cell component of these tumors has many histologic patterns. The small, cuboidal to polygonal cells may grow in anastomosing cords, sheets, or strands (Fig. 22.43A). In occasional cases, small, distinctive, glandlike structures filled with an acidophilic material recall immature follicles **(Call-Exner bodies)**. When these structures are evident, the diagnosis is straightforward. Occasionally, there is a predominant thecoma component that consists of clusters or sheets of cuboidal to polygonal cells (see later). In some tumors, the granulosa or theca cells may appear plumper and have ample cytoplasm characteristic of luteinization (i.e., luteinized granulosa–theca cell tumors).

Granulosa cell tumors are of clinical importance for two reasons: (1) they sometimes elaborate large amounts of estrogen, and (2) they may behave like low-grade malignancies. Functionally active tumors in prepubertal girls (juvenile granulosa cell tumors) may produce precocious sexual development. In adult women, they may be associated with proliferative breast disease, endometrial hyperplasia, and endometrial carcinoma, which eventually develops in about 10% to 15% of women with steroid-producing tumors. Occasionally, granulosa cell tumors produce androgens, masculinizing the patient.

All granulosa cell tumors are potentially malignant. It is difficult to predict their biologic behavior from histology. The likelihood of malignant behavior (recurrence, extension) ranges from 5% to 25%. In general, malignant tumors pursue an indolent course in which local recurrences may be amenable to surgical therapy. Recurrences within the pelvis and abdomen may appear 10 to 20 years after removal of the original tumor. The 10-year survival rate is approximately 85%. Tumors composed predominantly of theca cells are almost never malignant.

Elevated tissue and serum levels of *inhibin,* a product of granulosa cells, are associated with granulosa cell tumors. Detection and measurement of inhibin is useful in diagnosing granulosa and other sex cord–stromal tumors (see Fig. 22.43B), and for monitoring patients being treated for these neoplasms. The most common driver mutations are found in the *FOXL2* gene, which is mutated in 97% of adult granulosa cell tumors. *FOXL2* encodes a transcription factor that is important in granulosa cell development, which presumably explains its strong association with this tumor type. Interestingly, mutations in *FOXL2* appear to be less

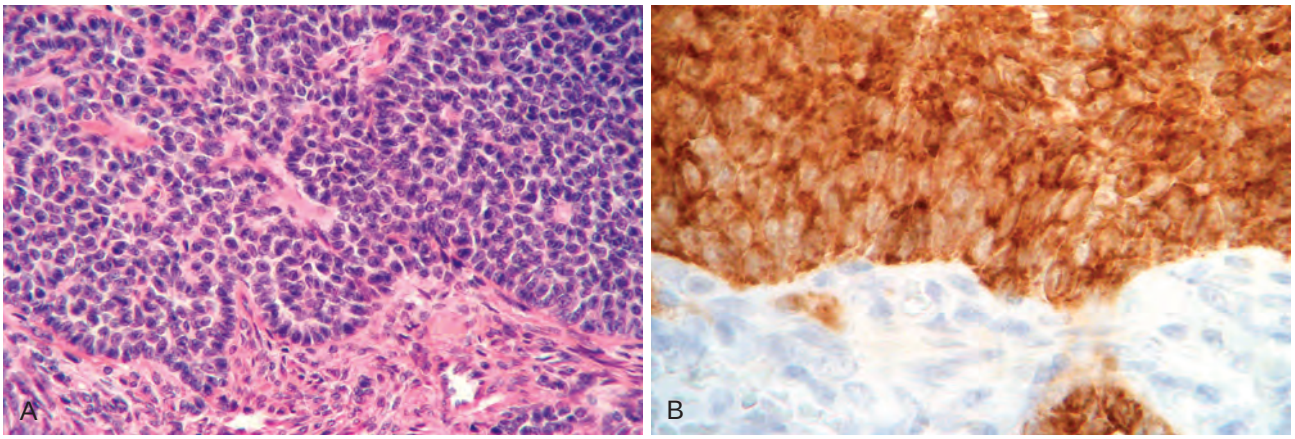


Figure 22.43 Granulosa cell tumor. (A) Tumor cells are arranged in sheets punctuated by small follicle-like structures (Call-Exner bodies). (B) Strong immunohistochemical positivity with an antibody to inhibin characterizes these tumors.

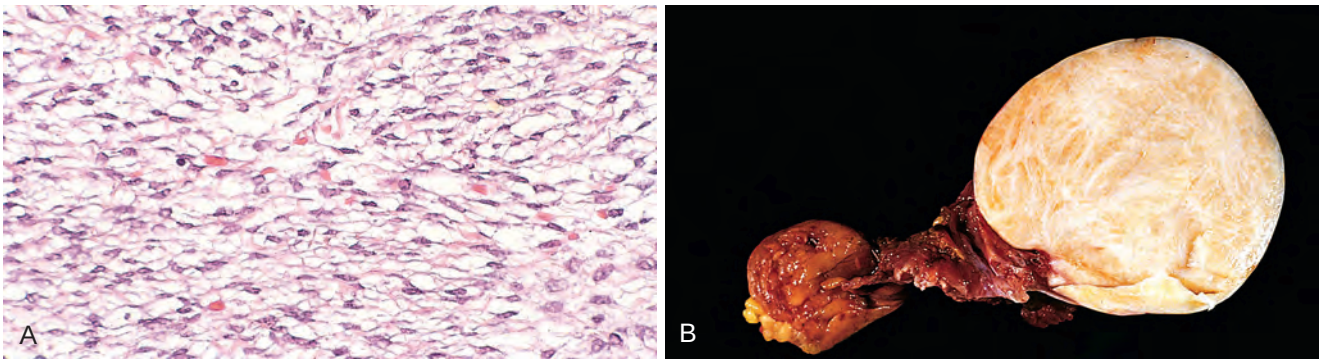


Figure 22.44 Ovarian fibroma. (A) Thecoma-fibroma composed of plump, differentiated stromal cells with thecal appearance. (B) Large bisected fibroma of the ovary apparent as a white, firm mass *(right)*. The fallopian tube is attached.

common in juvenile granulosa tumor, suggesting that it is genetically distinct from the adult type.

Fibromas, Thecomas, and Fibrothecomas

Tumors arising in the ovarian stroma that are composed of either fibroblasts (fibromas) or plump spindle cells with lipid droplets (thecomas) are relatively common, accounting for about 4% of all ovarian tumors (Fig. 22.44A). Many tumors contain a mixture of these cells and are termed *fibrothecomas*. Fibrothecomas and pure thecomas (a rare subtype) may be hormonally active. By contrast, pure fibromas as a rule are hormonally inactive.

Fibromas of the ovary are unilateral in about 90% of cases and are usually solid, spherical or slightly lobulated, encapsulated, hard, gray-white masses covered by glistening, intact ovarian serosa (see Fig. 22.44B). On histologic examination, they are composed of well-differentiated fibroblasts and scant interspersed collagenous stroma. Focal areas of thecal differentiation may be identified.

Most of these tumors come to attention as a pelvic mass, sometimes accompanied by pain and two decidedly curious associations. The first is ascites, found in about 40% of cases in which the tumors measure more than 6 cm in diameter. Uncommonly there is also a hydrothorax, usually only on the right side. This combination of findings (ovarian tumor, hydrothorax, and ascites) is designated *Meigs syndrome*. Its genesis is unknown. The second association is with the basal cell nevus syndrome, described in Chapter 25. The vast majority of fibromas, fibrothecomas, and thecomas are benign. Rarely, cellular fibromas with mitotic activity and increased nuclear-to-cytoplasmic ratio are identified; because they may pursue a malignant course, they are termed *fibrosarcomas*.

Sertoli-Leydig Cell Tumors

These tumors are often functional; they most commonly produce masculinization or defeminization, but a few have estrogenic effects. The tumor cells recapitulate, to a certain extent, testicular Sertoli or Leydig cells at various stages of development. They occur in women of all ages, although the peak incidence is in the second and third decades. In over one-half of cases, the tumor cells have mutations in *DICER1,* a gene that you will recall encodes an endonuclease that is essential for proper processing of micro-RNAs (Chapter 1). The presence of *DICER1* mutations suggests that genesis of these tumors involves abnormalities of gene expression related to dysregulation of micro-RNAs.

MORPHOLOGY

These tumors are unilateral and may resemble granulosa cell tumors grossly. The cut surface is usually solid and varies from gray to golden brown in appearance (Fig. 22.45A). Microscopically,

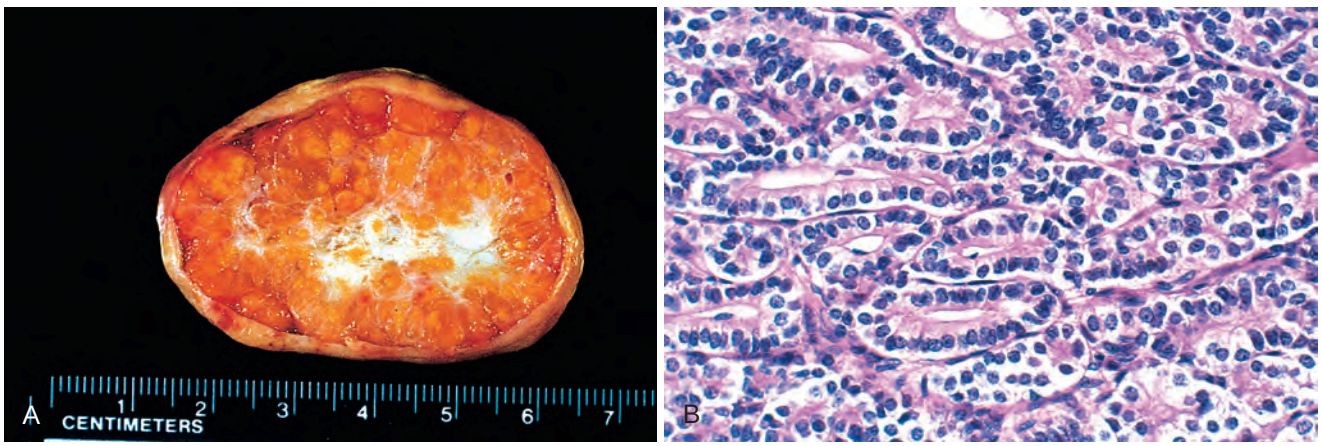


Figure 22.45 Sertoli cell tumor. (A) Gross photograph illustrating characteristic golden-yellow appearance of the tumor. (B) Photomicrograph showing well-differentiated Sertoli cell tubules. (Courtesy Dr. William Welch, Brigham and Women's Hospital, Boston, Mass.)

a range of differentiation is seen. Well-differentiated tumors show tubules composed of Sertoli cells or Leydig cells interspersed with stroma (Fig. 22.45B). The intermediate forms show only immature tubules and large eosinophilic Leydig cells. The poorly differentiated tumors have a sarcomatous pattern with a disorderly disposition of epithelial cell cords. Leydig cells may be absent. Heterologous elements, such as mucinous glands, bone, and cartilage, may be present in some tumors.

The incidence of recurrence or metastasis by Sertoli-Leydig cell tumors is less than 5%. These neoplasms may block normal female sexual development in children and may cause defeminization of women, manifest by atrophy of the breasts, amenorrhea, sterility, and loss of hair. The syndrome may progress to striking virilization (hirsutism) associated with male hair distribution, hypertrophy of the clitoris, and voice changes.

Other Sex Cord–Stromal Tumors

There are several other uncommon but distinctive ovarian tumors of sex cord or stromal origin that often produce steroid hormones.

- *Hilus cell tumors (pure Leydig cell tumors)* are rare, unilateral tumors comprised of large lipid-laden Leydig cells with distinct borders and characteristic cytoplasmic structures called *Reinke crystalloids*. Women with hilus cell tumors usually present with evidence of masculinization (hirsutism, voice changes, and clitoral enlargement), but these changes are milder than those seen in association with Sertoli–Leydig cell tumors. The tumors produce predominantly testosterone. Treatment consists of surgical excision. True hilus cell tumors are almost always benign.
- *Pregnancy luteoma* refers to a rare tumor that closely resembles the corpus luteum of pregnancy. These tumors may produce virilization in pregnant patients and their female infants.
- *Gonadoblastoma* is an uncommon tumor composed of germ cells and sex cord–stroma derivatives resembling immature Sertoli and granulosa cells. It occurs in individuals with abnormal sexual development and in gonads of indeterminate nature. Eighty percent of patients are phenotypic females, and 20% are phenotypic males with undescended testicles and female internal secondary organs. A coexistent dysgerminoma occurs in 50% of the cases. The prognosis is excellent if the tumor is completely excised.

KEY CONCEPTS

SEX CORD–STROMAL TUMORS

- Granulosa cells tumors are the most common malignant tumor in this category. They are indolent tumors, but may recur 10 to 20 years after resection of the primary tumor. They are often hormonally active; such tumors are associated with hyperestrinism, leading to an elevated risk of endometrial carcinoma.
- Fibromas are relatively common benign tumors composed of fibroblasts. They are predominantly unilateral and are generally hormonally inactive.
- Pure thecomas are rare but may be hormonally active.
- Sertoli–Leydig cell tumors commonly present with masculinization; less than 5% recur or metastasize.

Metastatic Tumors

The most common metastatic tumors of the ovary are derived from tumors of müllerian origin: the uterus, fallopian tube, contralateral ovary, or pelvic peritoneum. The most common extra-müllerian tumors metastatic to the ovary are carcinomas of the breast and gastrointestinal tract, including colon, stomach, biliary tract, and pancreas. Also included in this group are rare cases of pseudomyxoma peritonei, derived from appendiceal tumors. A classic metastatic gastrointestinal carcinoma involving the ovaries is termed *Krukenberg tumor,* characterized by bilateral metastases composed of mucin-producing cancer cells with a "signet-ring" appearance, most often of gastric origin.

Gestational and Placental Disorders

Diseases of pregnancy and pathologic conditions of the placenta are important causes of fetal intrauterine or perinatal death, congenital malformations, intrauterine growth restriction, maternal death, and morbidity for both the mother and child. Only those disorders for which recognition of morphologic features contribute to an understanding of the clinical problem are discussed here. These include selected disorders of early pregnancy, late pregnancy, and trophoblastic neoplasia.

Understanding placental disorders requires a working knowledge of normal placental anatomy. The placenta is composed of chorionic villi (Fig. 22.46A and B) that sprout from the chorion to provide a large contact area between the fetal and maternal circulations. In the mature placenta, the maternal blood enters the intervillous space through endometrial arteries (spiral arteries) and circulates around the villi to allow gas and nutrient exchange (Fig. 22.47). The deoxygenated blood flows back from the intervillous space to the decidua and enters the endometrial veins. Deoxygenated fetal blood enters the placenta through two umbilical arteries that branch radially to form chorionic arteries. Chorionic arteries branch further as they enter the villi. In the chorionic villi, they form an extensive capillary system, bringing fetal blood in close proximity to maternal blood. The gas and nutrient diffusion occurs through the villous capillary endothelial cells and thinned-out syncytiotrophoblasts and cytotrophoblasts. Under normal circumstances, there is little or no mixing of fetal and maternal blood, though sufficient free fetal DNA reaches the maternal circulation to permit prenatal genetic testing (Chapter 5). Blood oxygenated in the placenta returns to the fetus through the single umbilical vein.

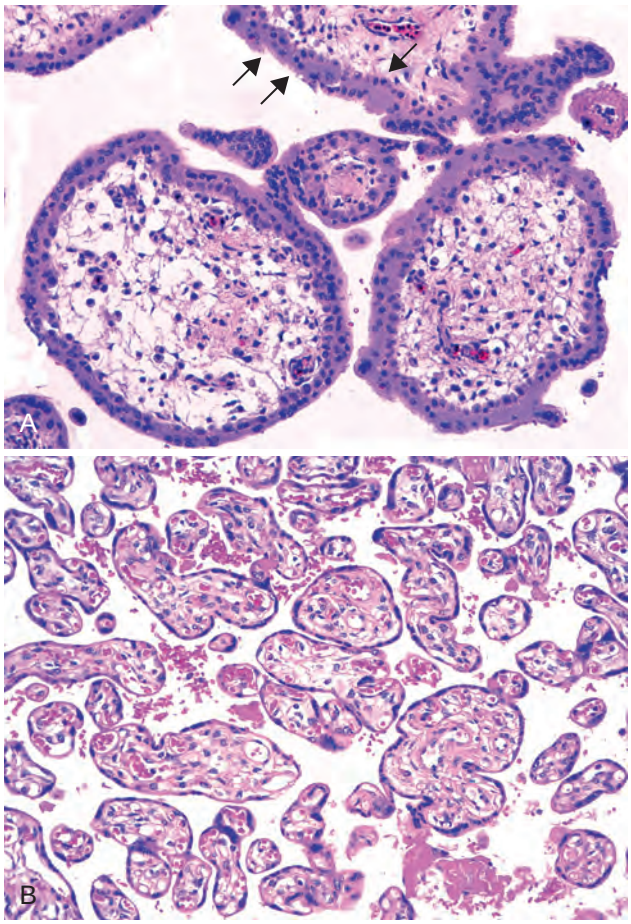


Figure 22.46 Normal placenta. (A) First-trimester chorionic villi composed of delicate mesh of central stroma surrounded by two discrete layers of epithelium—the outer layer consisting of syncytiotrophoblast *(double arrows)* and the inner layer consisting of cytotrophoblast *(arrow)*. (B) Third-trimester chorionic villi composed of stroma with a dense network of dilated capillaries surrounded by markedly thinned-out syncytiotrophoblast and cytotrophoblast (same magnification as A).

DISORDERS OF EARLY PREGNANCY

Spontaneous Abortion

Spontaneous abortion, or "miscarriage," is defined as pregnancy loss before 20 weeks of gestation. Most of these occur before 12 weeks. Ten to fifteen percent of clinically recognized pregnancies terminate in spontaneous abortion. However, using sensitive chorionic gonadotropin assays, it has been determined that an additional 20% of early pregnancies in otherwise healthy women terminate spontaneously, many without notice. In most individual instances, the mechanisms leading to early loss of pregnancy are unknown. However, multiple fetal and maternal causes of spontaneous abortion have been identified. Among the most important are the following:

- *Fetal chromosomal anomalies,* such as aneuploidy, polyploidy, and translocations, are present in approximately 50% of early abortuses. More subtle genetic defects, for which routine genetic testing is not available, account for an additional fraction of abortions.
- *Maternal endocrine factors,* including luteal-phase defect, poorly controlled diabetes, and other endocrine disorders
- *Physical defects of the uterus,* such as submucosal leiomyomas, uterine polyps, or uterine malformations, may prevent or disrupt implantation.
- *Systemic disorders affecting the maternal vasculature,* such as antiphospholipid antibody syndrome, coagulopathies, and hypertension
- *Infections* with protozoa *(Toxoplasma),* bacteria *(Mycoplasma, Listeria),* or a number of viruses. Ascending infection is particularly common in second-trimester losses.

Ectopic Pregnancy

Ectopic pregnancy refers to implantation of the fetus in a site other than the normal intrauterine location; the most common site is the extrauterine fallopian tube (approximately 90% of cases). Other sites include the ovary, the abdominal cavity, and the intrauterine portion of the fallopian tube (cornual pregnancy). Ectopic pregnancies account for 2% of confirmed pregnancies. The most important predisposing condition, present in 35% to 50% of patients, is prior pelvic inflammatory disease resulting in intraluminal fallopian tube scarring (chronic salpingitis). The risk of ectopic pregnancy is also increased with peritubal scarring and adhesions, which may be caused by appendicitis, endometriosis, and previous surgery. In some cases, however, the fallopian tubes are apparently normal. Another risk factor is use of an intrauterine contraceptive device, which is associated with a twofold increase in ectopic pregnancy.

Ovarian pregnancy results from the fertilization and trapping of the ovum within the follicle just at the time of its rupture. Abdominal pregnancies occur when the fertilized ovum fails to enter or drops out of the fimbriated end of the tube. In each abnormal location, the fertilized ovum develops as usual, forming placental tissue, amniotic sac, and fetus. The host implantation site may also develop decidual changes.

MORPHOLOGY

Tubal pregnancy is the most common cause of **hematosalpinx (blood-filled fallopian tube)** and should always be suspected when a tubal hematoma is present. Initially the embryonal sac, surrounded by immature chorionic villi, implants within the lumen of the fallopian tube. Trophoblastic cells and chorionic villi then invade the wall of the fallopian tube, as in the uterus during normal pregnancy. With time, the growth of the gestational sac distends the fallopian tube, causing thinning of the wall and rupture. **Rupture frequently results in massive intraperitoneal hemorrhage,** which sometimes is fatal. Less commonly, the tubal pregnancy may undergo spontaneous regression and resorption, or may be extruded through the fimbriated end of the tube into the abdominal cavity (tubal abortion).

Clinical Features

Rupture of a tubal pregnancy is a medical emergency. The clinical course of ectopic tubal pregnancy is characterized by the onset of moderate to severe abdominal pain and vaginal bleeding 6 to 8 weeks after the last menstrual period,

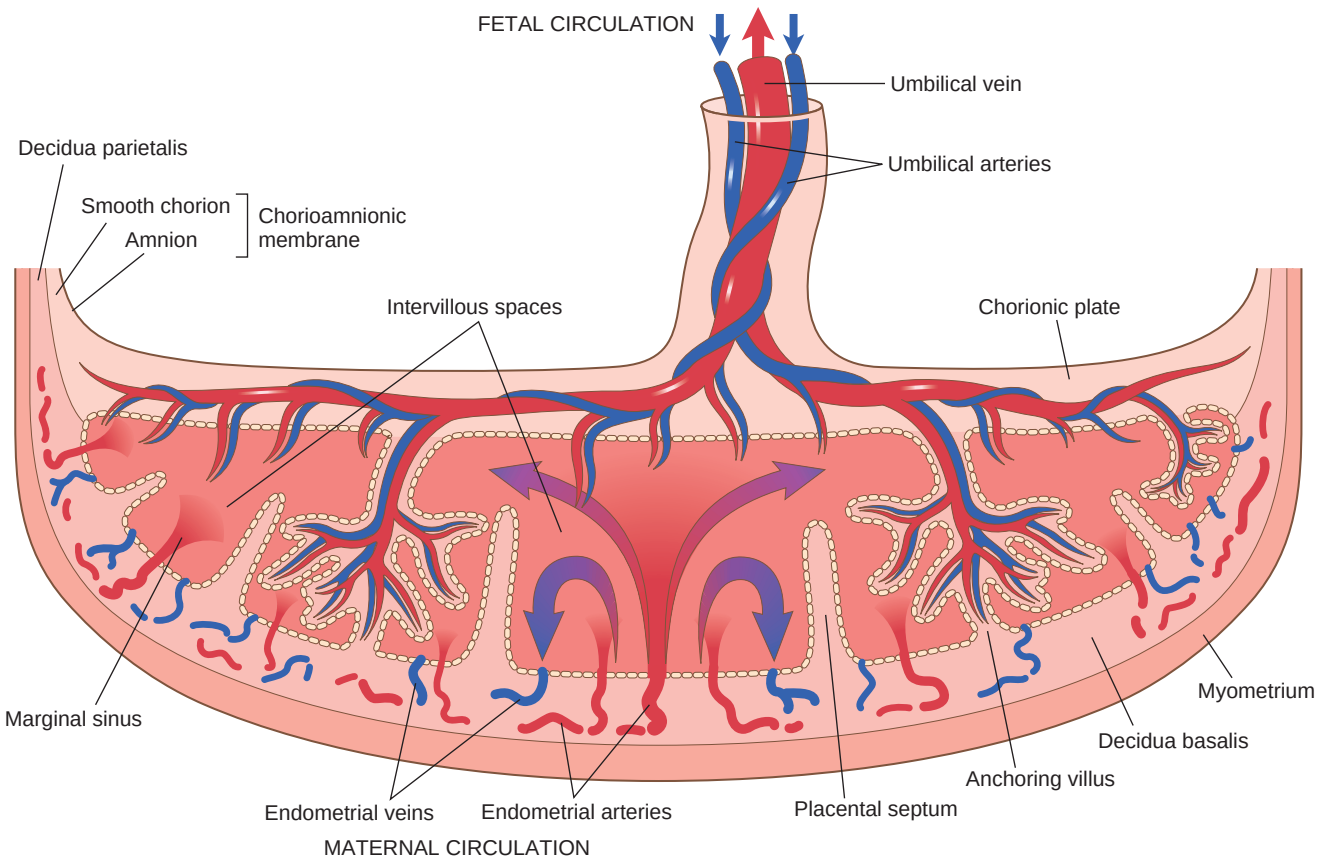


Figure 22.47 Diagram of placental anatomy. Within the outer boundary of myometrium is a layer of decidua, from which the maternal vessels originate and deliver blood to and from the intervillous spaces. Umbilical vessels branch and terminate in placental villi, where nutrient exchange takes place.

correlating with distention and then rupture of the fallopian tube. In such cases, the patient may rapidly develop hemorrhagic shock and signs of an acute abdomen. Rapid diagnosis is critical and is based on determination of chorionic gonadotropin titers, pelvic sonography, endometrial biopsy (which shows decidua without chorionic villi or implantation site), and/or laparoscopy. Despite advances in early diagnosis, ectopic pregnancy still accounts for 4% to 10% of pregnancy-related deaths.

DISORDERS OF LATE PREGNANCY

Disorders that occur in the third trimester of pregnancy are related to the complex anatomy of the maturing placenta. Complete interruption of blood flow through the umbilical cord from any cause (e.g., constricting knots or compression) can be lethal to the fetus. Ascending infections involving the chorioamnionic membranes may lead to premature rupture of amniotic membranes and delivery. Retroplacental hemorrhage at the interface of the placenta and myometrium *(abruptio placentae)* threatens both the mother and fetus. Disruption of the fetal vessels in terminal villi may produce a significant loss of fetal blood, resulting in fetal injury or death. Uteroplacental malperfusion can be precipitated by abnormal placental implantation or development, or maternal vascular disease; the effects may range from mild intrauterine growth restriction to severe uteroplacental ischemia and maternal preeclampsia (described later).

Twin Placentas

Twin pregnancies arise from fertilization of two ova (dizygotic) or from division of one fertilized ovum (monozygotic). There are three types of twin placentas (Fig. 22.48): diamnionic dichorionic (which may be fused); diamnionic monochorionic; and monoamnionic monochorionic.

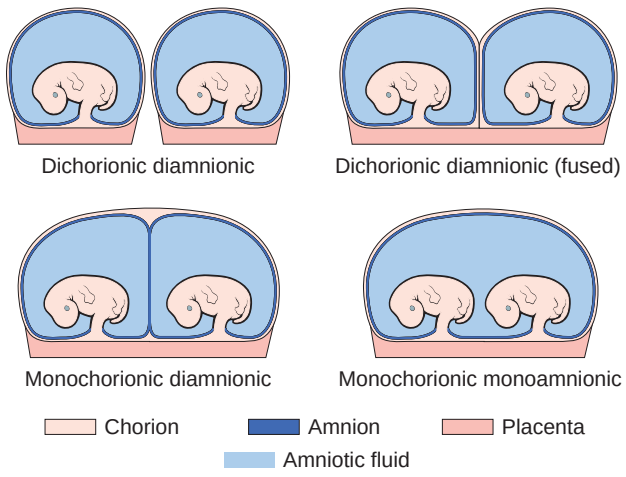


Figure 22.48 Diagrammatic representation of the various types of twin placentation and their membrane relationships. (Modified from Gersell D, et al: Diseases of the placenta. In Kurman R, editor: *Blaustein's Pathology of the Female Genital Tract,* New York, 1994, Springer-Verlag.)

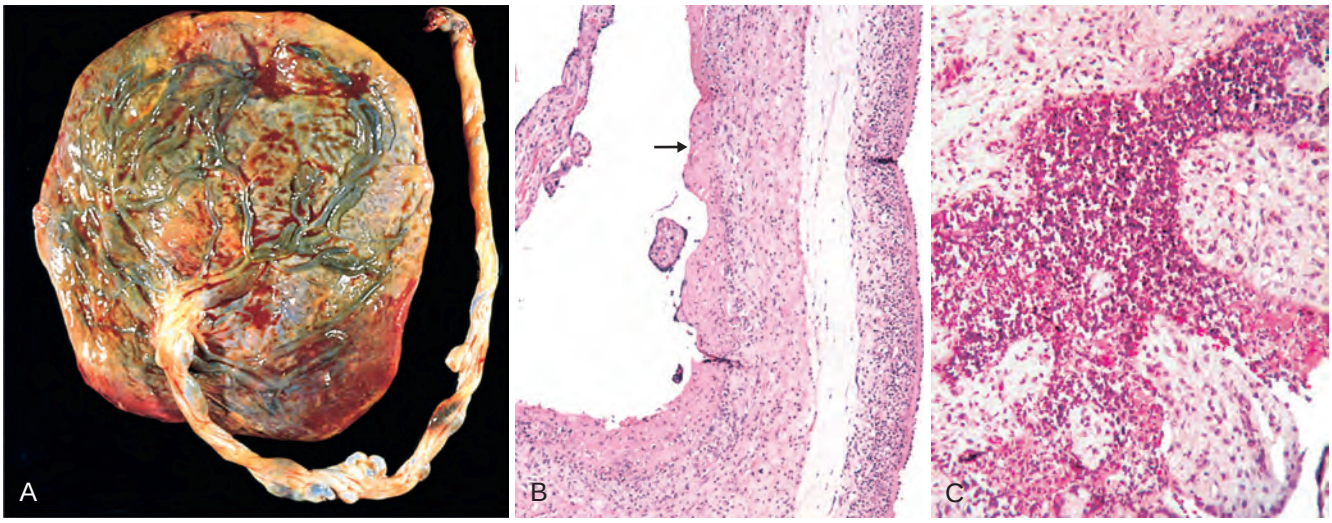


Figure 22.49 Placental infections derived from ascending and blood-borne routes. Acute chorioamnionitis. (A) On gross examination, the placenta contains greenish opaque membranes. (B) A photomicrograph illustrates a dense bandlike inflammatory exudate on the amniotic surface *(arrow)*. (C) Acute necrotizing intervillositis from *Listeria* infection. The space between the villi is filled with blood and acute inflammatory cells (neutrophils). The villi appear necrotic.

Monochorionic placentas imply monozygotic (identical) twins, and the time at which splitting of the developing embryo occurs determines whether one or two amnions are present. Dichorionic placentation may occur with either monozygotic or dizygotic twins and is not specific.

One complication of monochorionic twin pregnancy is *twin-to-twin transfusion syndrome*. Monochorionic twin placentas have vascular anastomoses that connect the circulations of the twins, and in some cases these connections include one or more arteriovenous shunts. Twin-to-twin perfusion syndrome results when these shunts preferentially increase blood flow to one twin at the expense of the second, causing one twin to be underperfused and the second to be fluid overloaded. If severe, it may result in the death of one or both fetuses.

Abnormalities of Placental Implantation

Several types of abnormal placental implantations are associated with significant complications. *Placenta previa* is a condition in which the placenta implants in the lower uterine segment or cervix, often leading to serious third-trimester bleeding. A complete placenta previa covers the internal cervical os and thus requires cesarean delivery to avert placental rupture and fatal maternal hemorrhage during vaginal delivery. *Placenta accreta* is caused by partial or complete absence of the decidua, such that the placental villous tissue adheres directly to the myometrium, which leads to a failure of placental separation at birth. It is an important cause of severe, potentially life-threatening postpartum bleeding. Common predisposing factors are placenta previa (in up to 60% of cases) and history of previous cesarean delivery.

Placental Infections

Infections in the placenta develop by two pathways: (1) ascending infection through the birth canal and (2) hematogenous (transplacental) infection. Ascending infections are by far the most common and are virtually always bacterial; in many such instances, localized infection of the membranes produces premature rupture of membranes and preterm delivery. The amniotic fluid may be cloudy with purulent exudate, and histologically the chorion-amnion contains an infiltrate of neutrophils accompanied by edema and congestion of the vessels (Fig. 22.49A–B). The infection frequently elicits a fetal response consisting of a "vasculitis" of the umbilical and fetal chorionic plate vessels. Uncommonly, bacterial infections may result from hematogenous spread to the placenta, leading to acute villitis (Fig. 22.49C).

Several hematogenous infections, classically components of the TORCH group (*t*oxoplasmosis and *o*thers [syphilis, tuberculosis, listeriosis], *r*ubella, *c*ytomegalovirus, *h*erpes simplex), can affect the placenta. They give rise to chronic inflammatory cell infiltrates in the chorionic villi (chronic villitis) and are described in Chapter 10.

Preeclampsia and Eclampsia

Preeclampsia is a systemic syndrome characterized by widespread maternal endothelial dysfunction that presents during pregnancy with hypertension, edema, and proteinuria. It occurs in about 3% to 5% of pregnant women, usually in the last trimester and more commonly in women pregnant for the first time (primiparas). Some of these women become seriously ill, developing convulsions; this particularly severe form of the disorder is termed *eclampsia*. Other complications stemming from systemic endothelial dysfunction include hypercoagulability, acute renal failure, and pulmonary edema. Approximately 10% of women with preeclampsia develop microangiopathic hemolytic anemia, elevated liver enzymes, and low platelets, referred to as the HELLP syndrome (Chapter 18). Preeclampsia should be distinguished from gestational hypertension that can develop in pregnancy without proteinuria.

Pathogenesis

The placenta plays a central role in the pathogenesis of preeclampsia, as symptoms disappear rapidly after delivery of the placenta. The critical abnormalities in preeclampsia are diffuse endothelial dysfunction, vasoconstriction (leading to hypertension), and increased vascular permeability (resulting in proteinuria and edema). Recent work has demonstrated that these effects are most likely mediated by placenta-derived factor(s) released into the maternal circulation. Although the release of these factors and the clinical syndrome develop late in gestation, the pathogenesis of the disease appears to be closely tied to the earliest events of pregnancy and placentation. The principal pathophysiologic aberrations are proposed to be the following:

- *Abnormal placental vasculature*. The precipitating events in preeclampsia are abnormal trophoblastic implantation and a failure of physiologic remodeling of the maternal vessels, which is required for adequate perfusion of the placenta. At the implantation site of a normal pregnancy, fetal extravillous trophoblasts (cells not associated with chorionic villi) invade the maternal decidua and decidual vessels, destroy the vascular smooth muscle, and replace the maternal endothelial cells with fetal trophoblastic cells, forming hybrid fetomaternal blood vessels. This process converts the decidual spiral arteries from small-caliber resistance vessels to large-capacity uteroplacental vessels lacking a smooth muscle coat (Fig. 22.50). In preeclampsia, this vascular remodeling fails to occur, leaving the placenta ill equipped to meet the increased circulatory demands of late gestation and setting the stage for the development of placental ischemia.
- *Endothelial dysfunction and imbalance of angiogenic and antiangiogenic factors*. It is postulated that in response to hypoxia, the ischemic placenta releases factors into the maternal circulation that cause an imbalance in circulating angiogenic and anti-angiogenic factors; this, in turn, leads to systemic maternal endothelial dysfunction and the clinical symptoms of the disease. In support of this, the blood levels of two placenta-derived antiangiogenic factors, soluble FMS-like tyrosine kinase (sFltl) and endoglin, which antagonize the effects of VEGF and TGFβ, respectively, are several orders of magnitude higher in women with preeclampsia than in healthy controls. In preeclampsia, high levels of sFlt1 and soluble endoglin bring about a decrease in angiogenesis much earlier than in normal pregnancy. The result is defective vascular development in the placenta. Furthermore, TGFβ induces endothelial production of nitric oxide, a potent vasodilator; thus, inhibition of TGFβ by endoglin may directly contribute to systemic vasoconstriction, hypertension, and tissue hypoperfusion.

  Studies in animal models also implicate sFltl and soluble endoglin in the pathogenesis of endothelial dysfunction. When sFlt and endoglin are overexpressed together, rats develop nephrotic-range proteinuria, severe hypertension, and fetal growth restriction, the hallmarks of severe preeclampsia, as well as features of the HELLP syndrome, including elevated liver enzymes, decreased platelet count, and hemolysis. Thus, it seems that sFlt1 and soluble endoglin are key mediators that link the placenta to the characteristic maternal endothelial dysfunction of preeclampsia.

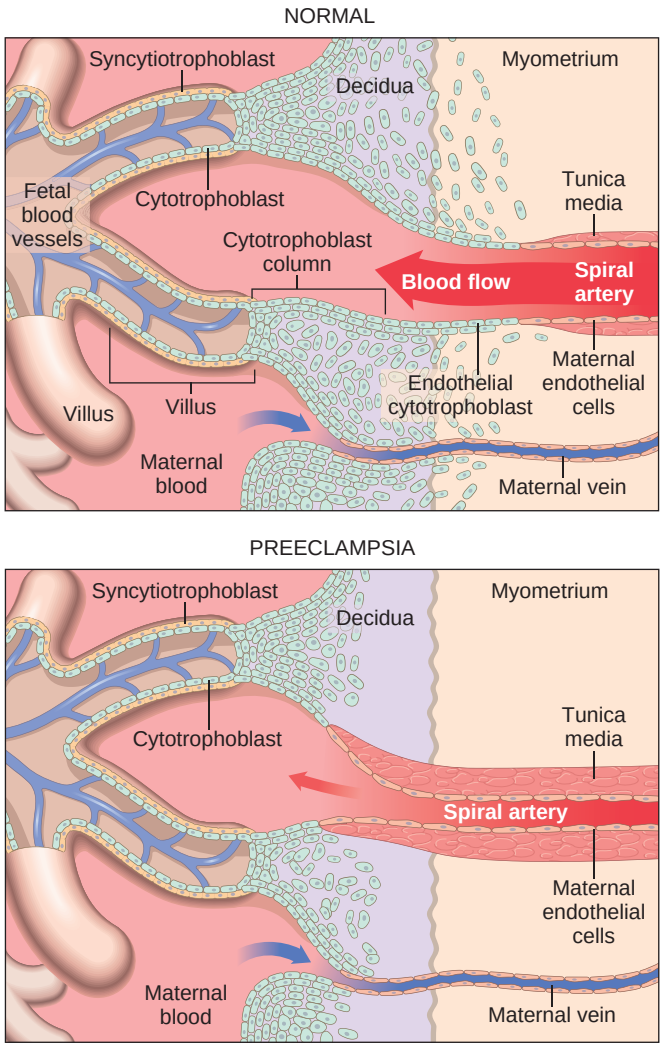


Figure 22.50 The physiologic alterations in the uterine spiral arteries and the failure of their remodeling in preeclampsia. (Modified from Maynard S, Epstein FH, Karumanchi SA: Preeclampsia and angiogenic imbalance, *Annu Rev Med* 59:61–78, 2008.)

- *Coagulation abnormalities*. Preeclampsia is associated with a hypercoagulable state that may lead to the formation of thrombi in arterioles and capillaries throughout the body, but particularly in the liver, kidneys, brain, and pituitary gland. Hypercoagulability is likely related to the reduced endothelial production of PGI_2, a potent antithrombotic factor, and increased release of procoagulant factors. Production of PGI_2 is stimulated by VEGF, and women with preeclampsia have decreased endothelial production of PGI_2, presumably due to antagonism of VEGF by sFlt1.

MORPHOLOGY

The placenta reveals several microscopic changes, most of which reflect malperfusion, ischemia, and vascular injury. These include (1) **infarcts,** which are larger and more numerous than those that may be seen in normal full-term placentas, (2) **exaggerated ischemic changes** in the chorionic villi and trophoblast, consisting of increased syncytial knots, (3) frequent **retroplacental**

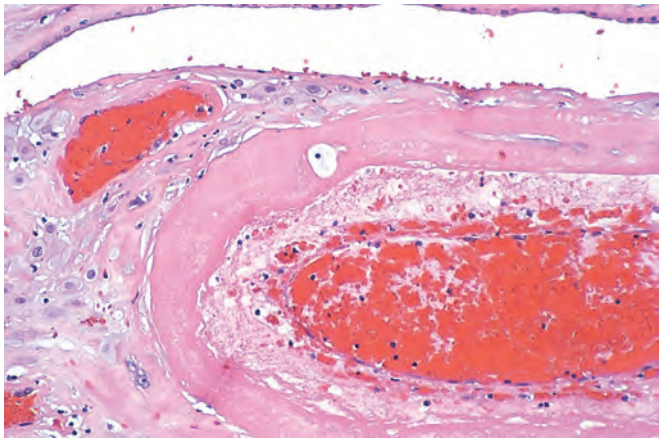


Figure 22.51 Acute atherosis of uterine vessels in eclampsia. Note fibrinoid necrosis of the vessel wall and subendothelial macrophages. (Courtesy Dr. Drucilla J. Roberts, Massachusetts General Hospital, Boston, Mass.)

hematomas due to bleeding and instability of uteroplacental vessels, and (4) **abnormal decidual vessels,** which may show thrombi, lack of normal physiologic conversion (described earlier), fibrinoid necrosis, or intimal lipid deposition (acute atherosis) (Fig. 22.51).

The liver lesions, when present, take the form of irregular, focal, subcapsular, and intraparenchymal hemorrhages. On histologic examination, there are fibrin thrombi in the portal capillaries and foci of hemorrhagic necrosis.

The kidney lesions are variable. The glomeruli show marked swelling of endothelial cells, amorphous dense deposits on the endothelial side of the basement membrane, and mesangial cell hyperplasia. Immunofluorescent studies show abundant fibrin deposition in glomeruli. In advanced cases, fibrin thrombi are present in the glomeruli and capillaries of the cortex. If widespread and severe, these thrombi may produce **bilateral renal cortical necrosis** (Chapter 20), leading to complete destruction of the cortices. The brain may have gross or microscopic foci of hemorrhage along with small-vessel thromboses. Similar changes are often found in the heart and the anterior pituitary gland.

Clinical Features

Preeclampsia most commonly starts after 34 weeks of gestation but begins earlier in women with hydatidiform mole (discussed later) or preexisting kidney disease, hypertension, or coagulopathies. The onset is typically marked by hypertension and edema, with proteinuria appearing several days later. Headaches and visual disturbances may follow and are indicative of severe preeclampsia, often requiring delivery. Eclampsia is heralded by more severe central nervous system involvement, including convulsions and eventual coma.

Management of preeclampsia differs depending on the gestational age and severity of disease. For term pregnancies, delivery is the treatment of choice regardless of disease severity. For preterm pregnancies, in which delivery is not in the best interest of the fetus, patients with mild disease can be managed expectantly by closely monitoring the mother and fetus. However, eclampsia, severe preeclampsia with maternal end-organ dysfunction, fetal compromise, or the HELLP syndrome are indications for delivery regardless of gestational age. Antihypertensive therapy does not affect the disease course or improve outcome. Proteinuria and hypertension usually disappear within 1 to 2 weeks after delivery except when they predate the pregnancy. Although in most instances preeclampsia has no lasting sequelae, about 20% of affected women develop hypertension and microalbuminuria within 7 years of a pregnancy complicated by preeclampsia. There is also a twofold increase in the long-term risk of vascular disease of the heart and the brain.

GESTATIONAL TROPHOBLASTIC DISEASE

Gestational trophoblastic disease encompasses a spectrum of tumors and tumorlike conditions characterized by proliferation of placental tissue, either villous or trophoblastic. The major disorders of this type are hydatidiform mole (complete and partial), invasive mole, choriocarcinoma, and placental site trophoblastic tumor (PSTT).

Hydatidiform Mole

Hydatidiform moles are important to recognize because they are associated with an increased risk of persistent trophoblastic disease (invasive mole) or choriocarcinoma. Moles are characterized histologically by cystic swelling of the chorionic villi and variable trophoblastic proliferation. They are usually diagnosed during early pregnancy (average 9 weeks) by pelvic sonography. Molar pregnancy can develop at any age, but the risk is higher at the two ends of reproductive life, in teenagers and between 40 and 50 years of age. For unknown reasons, the incidence varies considerably in different parts of the world. Hydatidiform mole occurs in about 1 in 1000 to 2000 pregnancies in the United States, but it is twice as common in Southeast Asia. Two types of benign, noninvasive moles—complete and partial—can be identified by cytogenetic and histologic studies.

Complete Mole

Complete mole results from fertilization of an egg that has lost its female chromosomes, and as a result the genetic material is completely paternally derived (Fig. 22.52A–B). Ninety percent have a 46,XX karyotype stemming from the duplication of the genetic material of one sperm (a phenomenon called androgenesis). The remaining 10% result from the fertilization of an empty egg by two sperm; these may have a 46,XX or 46,XY karyotype. In complete moles, the embryo dies very early in development and therefore is usually not identified. Patients have 2.5% risk of subsequent choriocarcinoma and a 15% risk of persistent or invasive mole.

Partial Mole

Partial moles result from fertilization of an egg with two sperm (see Fig. 22.52C). In these moles, the karyotype is triploid (e.g., 69,XXY) or occasionally tetraploid (92,XXXY). Fetal tissues are typically present. Partial moles have an increased risk of persistent molar disease but are not associated with choriocarcinoma.

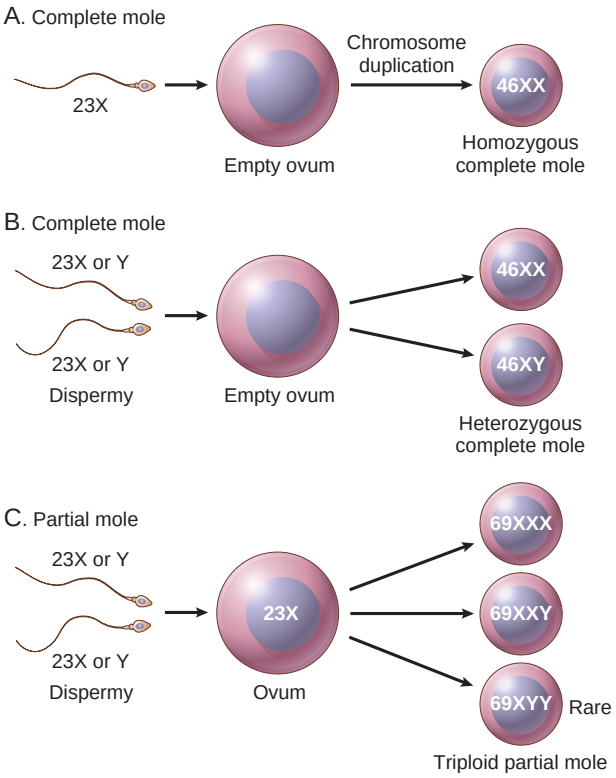


Figure 22.52 Origin of complete and partial hydatidiform moles. (A) Complete moles most commonly arise from fertilization of an empty ovum by a single sperm that undergoes duplication of its chromosomes. (B) Less commonly, complete moles arise from dispermy in which two sperm fertilize an empty ovum. (C) Partial moles arise from two sperm fertilizing a single ovum.

MORPHOLOGY

Grossly, hydatidiform mole appears as a delicate, friable mass of thin-walled, translucent, cystic, grapelike structures consisting of swollen edematous (hydropic) villi (Figs. 22.53 and 22.54). In complete mole, the microscopic abnormalities involve all or most of the villous tissue. The chorionic villi are enlarged, scalloped in shape with central cavitation (cisterns), and covered by extensive trophoblast proliferation that involves the entire circumference of the villi. In contrast, in partial moles, only a fraction of the villi are enlarged and edematous, and the trophoblastic hyperplasia is focal and less marked than in complete moles.

Clinical Features

Most women with partial or early complete mole present with spontaneous miscarriage or undergo curettage because of ultrasound findings of abnormal villous enlargement. In cases of complete mole, human chorionic gonadotropin (hCG) levels greatly exceed those of a normal pregnancy of similar gestational age. In addition, the rate at which hCG levels rise over time in molar pregnancy exceeds what is seen with normal single or even multiple pregnancies. Most moles are successfully removed by curettage. Patients are subsequently monitored for 6 months to 1 year to ensure that hCG levels decrease to nonpregnant levels. Continuous elevation of hCG may be indicative of persistent or invasive mole (described next).

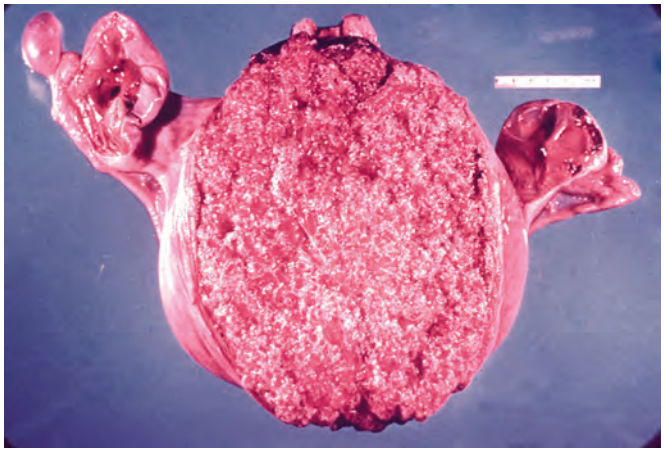


Figure 22.53 Complete hydatidiform mole. Note marked distention of the uterus by enlarged, vesicular chorionic villi. Adnexa (ovaries and fallopian tubes) are visible on the left and right side of the uterus.

Invasive Mole

Invasive mole is an infiltrative lesion that penetrates or even perforates the uterine wall. There is invasion of the myometrium by hydropic chorionic villi, accompanied by proliferation of both cytotrophoblasts and syncytiotrophoblasts. In some instances, the villi invade parametrial tissue and blood vessels and may even embolize to distant sites, such as the lungs and brain. However, these emboli do not grow, as do true metastases, and even without chemotherapy they eventually regress. The tumor manifests clinically with vaginal bleeding and irregular uterine enlargement. It is always associated with a persistently elevated serum hCG. Invasive mole responds well to chemotherapy but may result in uterine rupture, necessitating hysterectomy.

Choriocarcinoma

Gestational choriocarcinoma is a malignant neoplasm of trophoblastic cells derived from a previously normal or abnormal pregnancy, such as an extrauterine ectopic pregnancy. Gestational choriocarcinoma is an uncommon condition that arises in 1 in 20,000 to 30,000 pregnancies in the

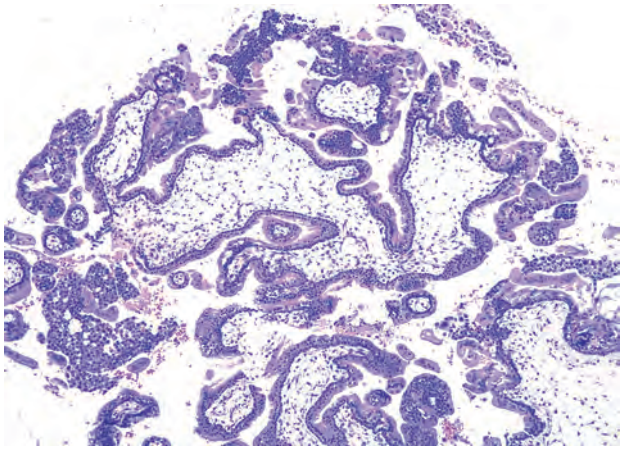


Figure 22.54 Complete hydatidiform mole demonstrating marked villous enlargement, edema, and circumferential trophoblast proliferation.

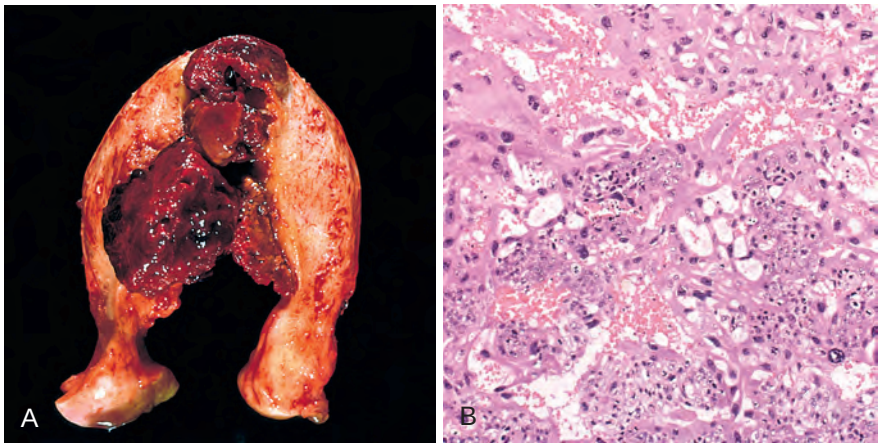


Figure 22.55 Choriocarcinoma. (A) Choriocarcinoma presenting as a bulky hemorrhagic mass invading the uterine wall. (B) Photomicrograph illustrating a tumor composed of neoplastic cytotrophoblasts and syncytiotrophoblasts. (Courtesy Dr. David R. Genest, Brigham and Women's Hospital, Boston, Mass.)

United States. It may be preceded by several conditions; 50% arise in complete hydatidiform moles, 25% in previous abortions, approximately 22% follow normal pregnancies, with the remainder occurring in ectopic pregnancies. Gestational choriocarcinoma is rapidly invasive and metastasizes widely, but once identified responds well to chemotherapy. Rarely, nongestational choriocarcinoma develops from germ cells in the ovaries or the mediastinum. These tumors are morphologically identical to gestational choriocarcinoma but can be distinguished by the absence of paternally derived DNA, which is present in all gestational choriocarcinomas.

MORPHOLOGY

Choriocarcinoma is a soft, fleshy, yellow-white tumor that usually has large pale areas of necrosis and extensive hemorrhage (Fig. 22.55A). Histologically, it consists entirely of proliferating syncytiotrophoblasts and cytotrophoblasts (Fig. 22.55B); chorionic villi are absent. Mitoses are abundant and sometimes abnormal. The tumor invades the underlying myometrium, frequently penetrates blood vessels, and in some cases extends out onto the uterine serosa and into adjacent structures.

Clinical Features

Gestational choriocarcinoma usually manifests as irregular vaginal spotting of a bloody, brown fluid. This discharge may appear in the course of an apparently normal pregnancy, after a miscarriage, or after curettage. Sometimes the tumor does not appear until months after these events. This tumor has high propensity for hematogenous spread, and by the time it is discovered, radiographs of the chest and bones may disclose the presence of metastatic lesions. The hCG levels are typically elevated to levels above those encountered in hydatidiform moles, but occasional tumors produce little hormone, and some tumors are so necrotic that hCG levels are low. Widespread metastases are characteristic; the most common sites are the lungs (50%) and vagina (30% to 40%), followed by, in descending order of frequency, the brain, liver, bone, and kidney.

Treatment of gestational choriocarcinoma depends on the stage of the tumor and usually consists of evacuation of the contents of the uterus and chemotherapy. The results of chemotherapy are spectacular and result in nearly 100% remission and a high rate of cures. Many of the cured patients have had normal subsequent pregnancies and deliveries. By contrast, rare nongestational choriocarcinomas that arise outside of the uterus are more resistant to therapy, although cures are possible with regimens used to treat other forms of germ cell neoplasms.

Placental Site Trophoblastic Tumors

PSTTs comprise less than 2% of gestational trophoblastic neoplasms. They are neoplastic proliferations of extravillous trophoblasts, also called *intermediate trophoblasts*. Normal extravillous trophoblasts are polygonal mononuclear cells that have abundant cytoplasm and produce human placental lactogen. In normal pregnancy, extravillous trophoblasts are found in the implantation site, as clusters of cells within the placental parenchyma, and in the placental membranes. PSTT presents as a uterine mass, accompanied by either abnormal uterine bleeding or amenorrhea and moderately elevated hCG. The malignant trophoblastic cells typically diffusely infiltrate the endomyometrium. PSTT may follow a normal pregnancy (one-half of the cases), spontaneous abortion, or hydatidiform mole. Patients with localized disease have an excellent prognosis, but 10% to 15% of women die of disseminated disease.

SUGGESTED READINGS

Infectious Diseases of Lower Genital Tract

Johnston C, Gottlieb SL, Wald A: Status of vaccine research and development of vaccines for herpes simplex virus, *Vaccine* 34:2948, 2016. *[An overview of HSV vaccine development].*

Vulva

Nooij LS, Ter Haar NT, Ruano D et al: Genomic characterization of vulvar (pre)cancers identifies distinct molecular subtypes with prognostic significance, *Clin Cancer Res* 23:6781, 2017. *[A comprehensive molecular analysis of vulvar cancer].*

Vagina

Schrager S, Potter BE: Diethylstilbestrol exposure, *Am Fam Physician* 69:2395, 2004. *[A review of clinical outcomes in patients with intrauterine DES exposure].*

Cervix

Brotherton JML, Bloem PN: Population-based HPV vaccination programmes are safe and effective: 2017 update and the impetus for achieving better global coverage, *Best Pract Res Clin Obstet Gynaecol* S1521-6934:30128, 2017. *[A review of current HPV vaccines].*

Munoz N, Bosch FX, de Sanjose S et al: Epidemiologic classification of human papillomavirus types associated with cervical cancer, *N Engl J Med* 348:518, 2003. *[A seminal study of HPV detection in cervical carcinoma].*

Ostor AG: Natural history of cervical intraepithelial neoplasia: a critical review, *Int J Gynecol Pathol* 12:186, 1993. *[A comprehensive literature review of the natural history of cervical intra-epithelial neoplasia].*

Schiffman M, Castle PE, Jeronimo J et al: Human papillomavirus and cervical cancer, *Lancet* 370:890, 2007. *[A review of HPV-related cervical carcinogenesis].*

Serrano B, Brotons M, Bosch FX et al: Epidemiology and burden of HPV-related disease, *Best Pract Res Clin Obstet Gynaecol* S1521-6934:30124, 2017. *[The epidemiology of HPV-related disease].*

Wright TC Jr, Schiffman M, Solomon D et al: Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening, *Obstet Gynecol* 103:304, 2004. *[Current recommendations for HPV testing].*

Uterus and Endometrium

Bulun SE: Endometriosis, *N Engl J Med* 360:268, 2009. *[A seminal review of endometriosis pathogenesis].*

Croce S, Chibon F: *MED12* and uterine smooth muscle oncogenesis: State of the art and perspectives, *Eur J Cancer* 51:1603, 2015. *[A review of the pathogenesis of leiomyomas].*

Giudice LC: Endometriosis, *N Engl J Med* 362:2389, 2010. *[A review of the pathophysiology of endometriosis].*

Hufnagel D, Li F, Cosar E et al: The role of stem cells in the etiology and pathophysiology of endometriosis, *Semin Reprod Med* 33:333, 2015. *[A review of stem cells in endometriosis].*

The Cancer Genome Atlas Research Network: Integrated genomic characterization of endometrial carcinoma, *Nature* 500:242, 2013. *[A comprehensive genomic study of endometrioid and serous endometrial carcinoma].*

Ovary

Azzia R et al: Polycystic ovary syndrome, *Nat Rev Dis Primers* 11:16057, 2016. *[A comprehensive review of PCOS].*

Heravi-Moussavi A et al: Recurrent somatic dicer mutations in non-epithelial ovarian cancers, *N Engl J Med* 366:234, 2012. *[A landmark study of mutations in non-epithelial ovarian tumors].*

Kraggerud SM et al: Molecular characteristics of malignant ovarian germ cell tumors and comparison with testicular counterparts: implications for pathogenesis, *Endocr Rev* 34:339, 2013. *[A detailed review of the molecular genetics of ovarian germ cell tumors].*

Kurman RJ, Shih IM: The dualistic model of ovarian carcinogenesis: revisted, revised and expanded, *Am J Pathol* 186:733, 2016. *[A recent comprehensive review of the pathogenesis of ovarian cancer].*

Mathias-Guiu X, Stewart CJR: Endometriosis-associated ovarian neoplasia, *Pathology* 17:30384, 2017. *[An overview of ovarian tumors associated with endometriosis].*

Placenta

Hod T, Cerdeira AS, Karumanchi SA: Molecular mechanisms of preeclampsia, *Cold Spring Harb Perspect Med* 5:a023473, 2015. *[A review of angiogenic alterations in preeclampsia].*

Lurain JR: Gestational trophoblastic disease I: epidemiology, pathology, clinical presentation and diagnosis of gestational trophoblastic disease, and management of hydatidiform mole, *Am J Obstet Gynecol* 203:531, 2010. *[A review of the clinicopathologic characteristics of molar pregnancies].*

Wilcox AJ, Weinberg CR, O'Connor JF et al: Incidence of early loss of pregnancy, *N Engl J Med* 319:189, 1988. *[An older but still useful paper on early pregnancy loss].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 23

The Breast

Susan C. Lester

CHAPTER CONTENTS

The breast is distinguished from other organs by three important characteristics. First, its major function is to provide for the nutritional support and survival of another individual, the infant. Second, it undergoes dynamic structural changes throughout life: expansion of the lobular system after menarche; periodic remodeling during adulthood, especially during and after pregnancy; and ultimately involution and regression. Finally, as symbols of femininity, breasts have social, cultural, and personal importance that is unique from other organs. All of these features impact the origin, presentation, and treatment of breast disease.

Understanding diseases of the breast requires knowledge of its normal anatomy and cellular constituents, which include two major structures (ducts and lobules), two types of epithelial cells (luminal and myoepithelial), and two types of stroma (interlobular and intralobular). Each of these elements is the source of both benign and malignant lesions (Fig. 23.1). Six to 10 major ducts extend down from the nipple and branch into multiple smaller ducts and finally terminate as lobules. Superficial portions of the major ducts are lined by keratinizing squamous epithelium that abruptly transitions to a double-layered epithelium (luminal and myoepithelial cells), which lines the remainder of the duct/lobular system. Following puberty, the duct system expands and proliferates, giving rise to numerous terminal duct lobular units. These units consist of grape-like clusters of small acini surrounded by a specialized intralobular stroma (see Fig. 23.1). In some women, ducts and lobules extend into the subcutaneous tissue of the chest wall and into the axilla. As a result, diseases such as breast cancer sometimes arise outside of the confines of the grossly evident breast tissue.

Changes in the female breast are most dynamic and profound during the reproductive years. Further branching of ducts and lobule formation occur under the influence of hormones such as estrogen, progesterone, and prolactin. Just as the endometrium grows and ebbs with each menstrual cycle, so does the breast. In the first half of the menstrual cycle the lobules are relatively quiescent. After ovulation, under the influence of estrogen and rising progesterone levels, cell proliferation increases, as does the number of acini per lobule. Upon menstruation, the fall in hormone levels induces regression of the lobules.

Only with pregnancy does the breast completely mature and become fully functional. Lobules increase progressively in number and size. By the end of a full-term pregnancy the breast is composed almost entirely of lobules separated by scant stroma. After parturition, the lobules initially produce colostrum (high in protein), changing to milk (higher in fat

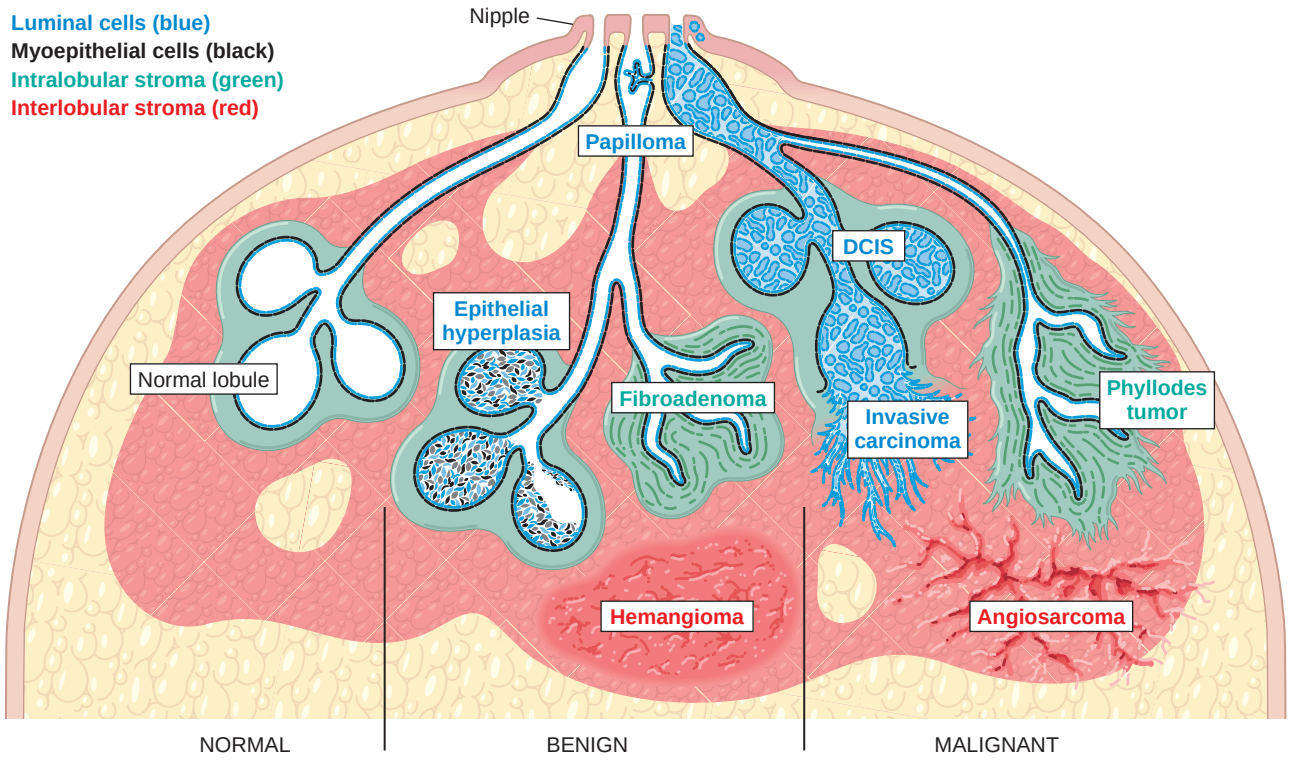


Figure 23.1 The normal cells and structures of the breast, including epithelial cells and myoepithelial cells, intralobular stromal cells and interlobular stromal cells, and large ducts and terminal duct lobular units, can give rise to both benign and malignant tumors. *DCIS*, Ductal carcinoma in situ.

and calories) over the next 10 days as progesterone levels drop. Upon cessation of lactation, epithelial cells undergo apoptosis and lobules regress, but only partially. The permanent changes that are produced by pregnancy may explain the reduction in breast cancer risk that is observed in women who give birth to children at young ages.

After the third decade, long before menopause, lobules and their specialized stroma start to involute, and the interlobular stroma is converted from radiodense fibrous stroma to radiolucent adipose tissue. These changes may be attenuated or delayed by hormones from endogenous (e.g., estrogen from fat stores in obese women) or exogenous (e.g., postmenopausal hormone replacement therapy) sources.

DISORDERS OF DEVELOPMENT

Milk Line Remnants

Supernumerary nipples or breasts result from the persistence of epidermal thickenings along the milk line, which extends from the axilla to the perineum. Disorders of normally situated breasts only very rarely arise in these heterotopic foci, which most commonly come to attention due to painful swelling prior to menstruation.

Accessory Axillary Breast Tissue

In some women the normal ductal system extends into the subcutaneous tissue of the chest wall or the axillary fossa (the axillary tail of Spence). Because this tissue may not be removed by prophylactic mastectomies, such procudures markedly reduce, but do not completely eliminate, the risk of breast cancer.

Congenital Nipple Inversion

The failure of the nipple to evert during development is common and may be unilateral. Congenitally inverted nipples are usually of little significance, since they correct spontaneously during pregnancy or can sometimes be everted by simple traction. Acquired nipple retraction is of more concern, since it may indicate the presence of an invasive cancer or an inflammatory nipple disease.

CLINICAL PRESENTATIONS OF BREAST DISEASE

The most common breast signs and symptoms reported by women are pain, inflammatory changes, nipple discharge, "lumpiness," or a palpable mass (Fig. 23.2A). Few symptoms are so severe as to require treatment, and the primary reason for investigating their cause is to evaluate the possibility of malignancy. Greater than 90% of symptomatic breast lesions are benign. Of women with cancer, about 45% have symptomatic disease, while the remainder come to attention through mammographic screening (Fig. 23.2B). Among the most common or significant signs and symptoms that bring breast lesions to clinical attention are the following:

- *Pain* (mastalgia or mastodynia) is a common symptom that may be cyclic with menses or noncyclic. Diffuse

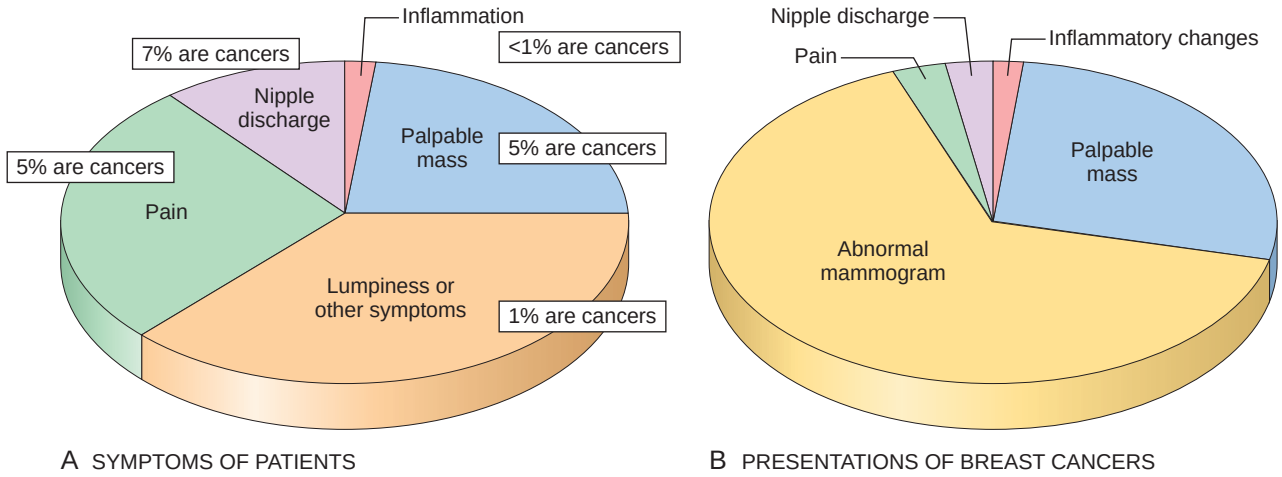


Figure 23.2 Symptoms of breast disease and presentations of breast cancer. (A) Common breast-related symptoms that bring patients to clinical attention include pain, inflammation, nipple discharge, "lumpiness," and palpable masses. Although these symptoms often cause concern, they are associated with cancer in less than 10% of affected women. (B) Presentations of breast cancer.

cyclic pain may be due to premenstrual edema. Noncyclic pain is usually localized to one area of the breast and may be caused by ruptured cysts, physical injury, and infections, but often no specific lesion is identified. Although most painful masses are benign, in about 5% of cases the underlying cause is breast cancer.

- *Inflammation* causes erythema and edema involving all or part of a breast. This is a rare symptom and is most often caused by infections, which only occur with any frequency during lactation and breastfeeding. An important mimic of reactive inflammation is inflammatory breast carcinoma (discussed later).
- *Nipple discharge* may be normal when small in quantity and bilateral. Milky discharge (galactorrhea) is associated with elevated prolactin levels (e.g., by a pituitary adenoma), hypothyroidism, or endocrine anovulatory syndromes and also may occur in patients taking oral contraceptives, tricyclic antidepressants, methyldopa, or phenothiazines. Repeated nipple stimulation also may induce lactation. Galactorrhea is not a feature of malignancy. Bloody or serous discharges are most commonly due to large duct papillomas and cysts. During pregnancy, rapid growth and remodeling of the breast may produce a bloody discharge. Discharge associated with malignancy is most commonly due to ductal carcinoma in situ (DCIS) and increases in prevalence with age; it is caused by an underlying carcinoma in 7% of women under age 60 and in 30% of women 60 years of age or older. Discharge in older women that is spontaneous, unilateral, and bloody is likely to have a malignant origin.
- *Lumpiness,* or diffuse nodularity of the breast, usually is a manifestation of normal glandular tissue. When pronounced, imaging studies may be needed to exclude the presence of a discrete mass.
- *Palpable masses* can arise from proliferations of stromal cells or epithelial cells and are generally detected when 2 to 3 cm in size. Most (~95%) are benign: these tend to be round or oval in shape, rubbery, and mobile and to have circumscribed borders. The most common benign masses are fibroadenomas and cysts. In contrast, malignant tumors invade across tissue planes, are often hard (scirrhous) in consistency, and usually have irregular borders. The likelihood that a mass is malignant increases with age, rising from 10% for women younger than age 40 to 60% for women older than age 50. Approximately 50% of carcinomas are located in the upper outer quadrant, 10% in each of the remaining quadrants, and about 20% in the central or subareolar region. Although about one-third of cancers are first detected as a palpable mass, screening by palpation has little effect on reducing breast cancer mortality. Unfortunately, the majority of cancers that have the capacity to metastasize do so before they reach a palpable size.

Mammographic screening was introduced in the 1980s in an effort to detect nonpalpable asymptomatic breast carcinomas before metastatic spread and is currently the most commonly used screening test for breast cancer (Fig. 23.2B). The sensitivity and specificity of mammography increase with age. At age 40, the probability that a mammographic lesion is cancer is only 10%, but this rises to greater than 25% in women older than 50. The principal mammographic signs of breast carcinoma are densities and calcifications:

- *Densities*. Breast lesions that replace adipose tissue with radiodense tissue form mammographic densities. Rounded densities are most commonly benign lesions such as fibroadenomas or cysts, whereas invasive carcinomas generally form irregular masses. The average size of invasive carcinomas detected by mammography is about 1 cm (significantly smaller than carcinomas detected by palpation), and only 15% will have metastasized to regional lymph nodes at the time of detection.
- *Calcifications*. Calcifications form on secretions, necrotic debris, or hyalinized stroma and are often associated with benign lesions such as apocrine cysts, hyalinized fibroadenomas, and sclerosing adenosis. Calcifications associated with malignancy are usually small, irregular, numerous, and clustered. A marked increase in the

diagnosis of DCIS, a form of breast cancer that is often associated with calcifications, was observed after the introduction of mammographic screening.

Approximately 10% of invasive carcinomas are not visible by standard two-dimensional mammographic techniques. Causes include the presence of surrounding radiodense tissue that obscures the tumor (especially in younger women); small tumor size; a diffuse infiltrative growth pattern with little or no tissue response; or a location close to the chest wall or in the periphery of the breast. Other imaging modalities can be useful adjuncts in such circumstances, particularly in evaluating palpable masses that are not seen well with standard mammography. Digital breast tomosynthesis (three-dimensional mammography) integrates additional views of the breast and can detect subtle changes in breast parenchymal texture. Ultrasonography distinguishes between solid and cystic lesions and more precisely defines the borders of solid lesions. Magnetic resonance imaging (MRI) detects cancers by the rapid uptake of contrast agents due to increased tumor vascularity and blood flow and can be particularly helpful in the evaluation of breasts of high density.

Although the recent downward trend in deaths from breast cancer is partially attributed to earlier diagnosis due to mammography, the beneficial effect of screening is smaller than originally anticipated for several reasons. By the time of detection, most (70% to 80%) cancers are invasive and some have metastasized. Even more troubling, the cancers that are most likely to cause death are least likely to be detected by mammography. These lethal cancers arise in young women of prescreening age or grow so rapidly that they present during the interval between mammograms. On the other side of the ledger, many cancers detected by mammography are clinically unimportant, being so indolent that they would never have caused harm had they gone undetected (a situation reminiscent of many prostate cancers in men; see Chapter 21). It is estimated that between 10% and 30% of invasive cancers detected by mammography fall into this category. Although this issue is sometimes referred to as a problem of overdiagnosis, it is more accurate to say that it reflects a need for reliable predictors of the behavior of invasive carcinomas; some predictors of outcome in current use are discussed later.

INFLAMMATORY DISORDERS

Inflammatory diseases of the breast are rare (accounting for less than 1% of breast symptoms) and are caused by infections, autoimmune disease, or foreign body–type reactions to extravasated keratin or secretions. Inflammatory breast cancer mimics inflammation by obstructing dermal vasculature with tumor emboli and should always be considered in a woman with an erythematous swollen breast (discussed later).

Acute Mastitis

Acute bacterial mastitis typically occurs during the first month of breastfeeding and is caused by a local infection when the breast is most vulnerable due to cracks and fissures in the nipples. From this portal of entry, *Staphylococcus aureus* or, less commonly, streptococci invade the breast tissue. The breast is erythematous and painful, and fever is often present. At the outset only one duct system or sector of the breast is involved. If not treated, the infection may spread to the entire breast. Staphylococci infection often leads to single or multiple abscesses, whereas streptococci cause spreading infection in the form of cellulitis.

Most cases of lactational mastitis are easily treated with appropriate antibiotics and continued expression of milk from the breast. Only rarely is surgical drainage required.

Squamous Metaplasia of Lactiferous Ducts

Squamous metaplasia of lactiferous ducts is known by a variety of names, including recurrent subareolar abscess, periductal mastitis, and Zuska disease. Women, and sometimes men, present with a painful erythematous subareolar mass that mimics a bacterial abscess. In recurrent cases, a fistula tract often develops under the smooth muscle of the nipple and opens onto the skin at the edge of the areola. In many affected women the nipple inverts due to traction produced by inflammation and scarring. More than 90% of afflicted individuals are smokers. It has been suggested that a relative deficiency of vitamin A associated with smoking or toxic substances in tobacco smoke alters the differentiation of the ductal epithelium.

MORPHOLOGY

The key feature is **keratinizing squamous metaplasia,** which extends into the nipple duct well past the usual point of transition from squamous to glandular epithelium (Fig. 23.3). Keratin shed from these cells is trapped and plugs the ductal system, causing

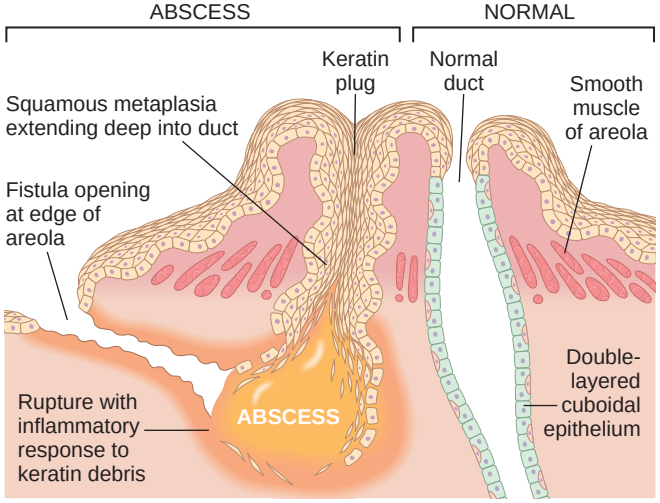


Figure 23.3 Squamous metaplasia of lactiferous ducts. When squamous metaplasia extends deep into a nipple duct, keratin becomes trapped and accumulates. If the duct ruptures, the ensuing intense inflammatory response to keratin results in an erythematous painful mass. A fistula tract may burrow beneath the smooth muscle of the nipple to open at the edge of the areola.

dilation and eventually rupture of the duct. An intense chronic granulomatous inflammatory response develops once keratin spills into the surrounding periductal tissue. With recurrences, a secondary anaerobic bacterial infection may supervene and cause acute inflammation.

Simple incision drains the abscess cavity, but the offending keratinizing epithelium remains, and recurrences are common. In most cases, en bloc surgical removal of the involved duct and contiguous fistula tract is curative. If secondary bacterial infection is present, antibiotics also have a therapeutic role.

Duct Ectasia

Duct ectasia presents as a palpable periareolar mass that is often associated with thick, white nipple secretions and occasionally with skin retraction. Pain and erythema are uncommon. This disorder tends to occur in the fifth or sixth decade of life, usually in multiparous women. Unlike squamous metaplasia of lactiferous ducts, it is not associated with cigarette smoking.

MORPHOLOGY

Ectatic dilated ducts are filled with inspissated secretions and numerous lipid-laden macrophages. With duct rupture, a marked periductal and interstitial chronic inflammatory reaction ensues, consisting of lymphocytes, macrophages, and variable numbers of plasma cells (Fig. 23.4). Granulomas may form around cholesterol deposits and secretions. Subsequent fibrosis produces an irregular mass with skin and nipple retraction.

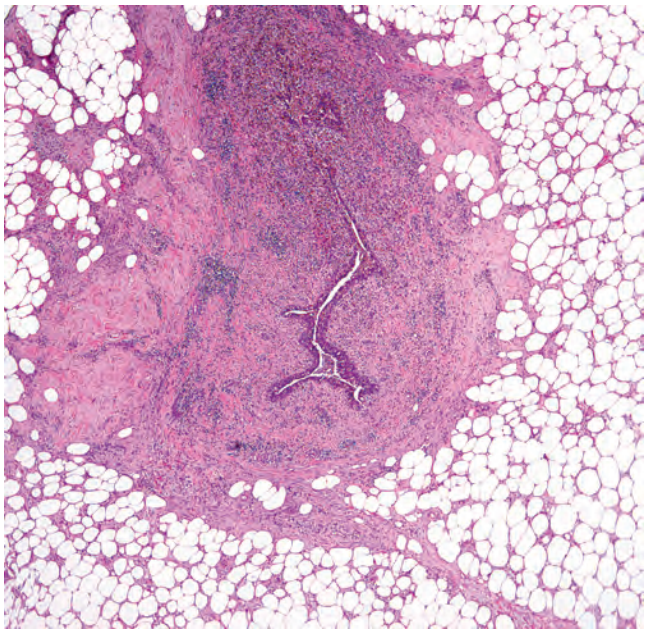


Figure 23.4 Duct ectasia. Chronic inflammation and fibrosis surround an ectatic duct filled with inspissated debris. The fibrotic response can produce a firm irregular mass that mimics invasive carcinoma on palpation or mammogram.

The principal significance of duct ectasia is that it may be difficult to distinguish on clinical and radiologic grounds from invasive carcinoma.

Fat Necrosis

The presentations of fat necrosis are protean and may closely mimic cancer. It may present as a painless palpable mass, skin thickening or retraction, or mammographic densities or calcifications. About half of affected women have a history of breast trauma or surgery.

MORPHOLOGY

Acute lesions may be hemorrhagic and contain central areas of liquefactive fat necrosis with neutrophils and macrophages. Over the next few days proliferating fibroblasts and chronic inflammatory cells surround the injured area. Subsequently, giant cells, calcifications, and hemosiderin appear, and eventually the lesion is replaced by scar tissue or is encircled and walled off by fibrous tissue. Ill-defined, firm, gray-white nodules containing small chalky-white foci are seen grossly.

Lymphocytic Mastopathy (Sclerosing Lymphocytic Lobulitis)

This condition presents with single or multiple hard palpable masses or mammographic densities. The masses are associated with areas of densely collagenized stroma, a feature that may make it difficult to obtain lesional tissue by needle biopsy. Within the stroma are atrophic ducts and lobules with thickened basement membranes that are surrounded by a prominent lymphocytic infiltrate. This condition is most common in women with type 1 (insulin-dependent) diabetes or autoimmune thyroid disease and is hypothesized to have an autoimmune basis. Its only clinical significance is that it must be distinguished from breast cancer.

Granulomatous Mastitis

Granulomatous inflammation of the breast may be a manifestation of systemic granulomatous diseases (e.g., granulomatosis with polyangiitis, sarcoidosis, tuberculosis) or of inflammatory or infectious disorders that are localized to the breast. *Granulomatous lobular mastitis* is an uncommon disease that occurs only in parous women. The granulomas are closely associated with lobules and may contain lipid vacuoles surrounded by neutrophils. A similar histologic pattern is seen in *cystic neutrophilic granulomatous mastitis,* which is often caused by lipophilic *Corynebacteria*. These histologic patterns may be manifestations of the same disease. Treatment includes antibiotics and sometimes steroids. Localized infections by mycobacteria or fungi are very rare and are most common in immunocompromised patients or adjacent to foreign objects such as breast prostheses or nipple piercings.

KEY CONCEPTS

INFLAMMATORY DISORDERS

- Inflammatory diseases of the breast are rare outside of the lactational period.
- The specific cause must be determined, as appropriate treatment may be antibiotics, steroids, or surgery.
- The possibility of inflammatory carcinoma mimicking a non-neoplastic inflammatory disorder should always be considered.

BENIGN EPITHELIAL LESIONS

Benign epithelial lesions are classified into three groups, each with a different risk for subsequent development of breast cancer: (1) *nonproliferative breast changes,* (2) *proliferative breast disease,* and (3) *atypical hyperplasia.* Most come to clinical attention when detected by mammography or as incidental findings in surgical specimens.

Nonproliferative Breast Changes (Fibrocystic Changes)

This group includes common morphologic alterations that are often grouped under the term *fibrocystic changes*. To the clinician, the term might mean "lumpy bumpy" breasts on palpation; to the radiologist, a dense breast with cysts; and to the pathologist, benign histologic findings. These lesions are termed nonproliferative to indicate that they are not associated with an increased risk of breast cancer; this name is somewhat unfortunate because some of these changes do involve increased proliferation and may even be associated with clonal genetic aberrations.

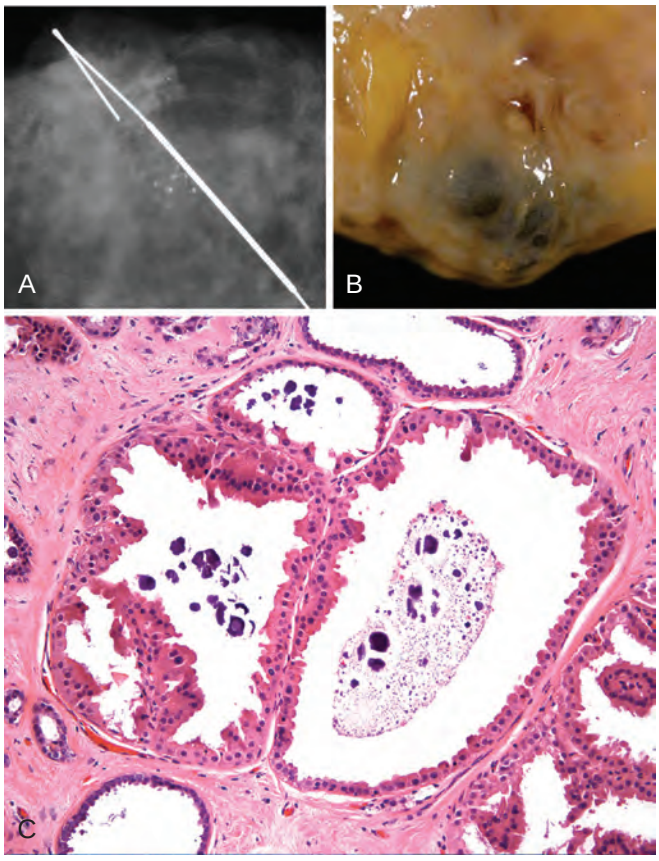


Figure 23.5 Apocrine cysts. (A) Clustered, rounded calcifications are seen in a specimen radiograph. (B) Gross appearance of typical cysts filled with dark, turbid fluid contents. (C) Cysts are lined by apocrine cells with round nuclei and abundant granular cytoplasm. Note the luminal calcifications, which form on secretory debris.

MORPHOLOGY

There are three principal nonproliferative morphologic changes: (1) cystic change, often with apocrine metaplasia; (2) fibrosis; and (3) adenosis.

- **Cysts.** Small cysts form by the dilation of lobules and in turn may coalesce to form larger cysts. Unopened cysts contain turbid, semitranslucent brown- or blue-colored fluid (blue-dome cysts) (Fig. 23.5B). Cysts are lined either by a flattened atrophic epithelium or by metaplastic apocrine cells. The latter cells have abundant granular, eosinophilic cytoplasm and closely resemble the normal apocrine epithelium of sweat glands (Fig. 23.5C). Calcifications are common (Fig. 23.5A). Cysts may cause concern when they are solitary and firm. The diagnosis is confirmed by the disappearance of the mass after fine-needle aspiration of its contents.
- **Fibrosis.** Cysts frequently rupture, releasing secretory material into the adjacent stroma. The resulting chronic inflammation and fibrosis contribute to palpable nodularity of the breast.
- **Adenosis.** Adenosis is defined as an increase in the number of acini per lobule. It is a normal feature of pregnancy. In nonpregnant women, adenosis can occur as a focal change. The acini are lined by columnar epithelial cells, and calcifications are occasionally present within the lumens.

Several other morphologic alterations fall into the category of "nonproliferative" changes. **Lactational adenomas** present as palpable masses in pregnant or lactating women and regress after cessation of breastfeeding. They consist of normal-appearing breast tissue with lactational changes. Their name may be a misnomer, as these lesions may be an exaggerated local response to gestational hormones rather than true neoplasms. **Flat epithelial atypia** is a clonal process characterized by the presence of dilated acini and cysts lined by epithelial cells that display mild cytologic atypia. It is associated with deletions of chromosome 16q and is the earliest morphologically recognizable clonal lesion of the breast. Flat epithelial atypia is often associated with lesions that increase the risk of cancer (e.g., atypical hyperplasia, described later) but has not been shown to increase risk in isolation.

Proliferative Breast Disease Without Atypia

Lesions characterized by proliferation of epithelial cells, without atypia, are associated with a small increase in the risk of subsequent carcinoma in either breast. They are commonly detected as mammographic densities, calcifications, or

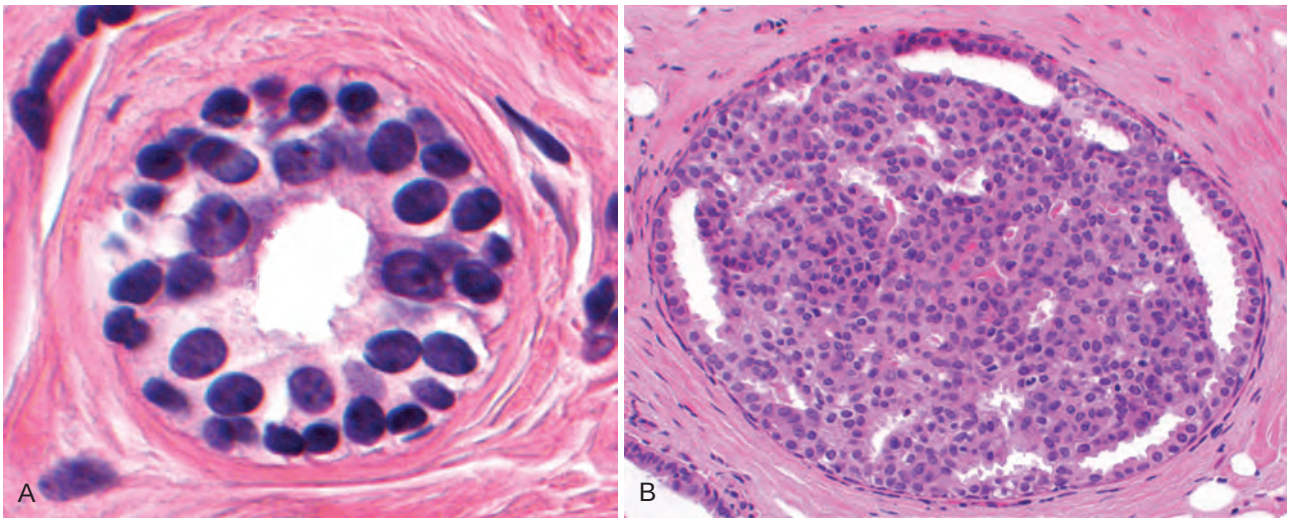


Figure 23.6 (A) Normal duct or acinus with a single basally located myoepithelial cell layer (cells with dark, compact nuclei and scant cytoplasm) and single luminal cell layer (cells with larger open nuclei, small nucleoli, and more abundant cytoplasm). (B) Epithelial hyperplasia. The lumen is filled with a heterogeneous, mixed population of luminal and myoepithelial cell types. Irregular slit-like fenestrations are prominent at the periphery.

incidental findings in biopsies performed for other reasons. These lesions are considered predictors of risk, rather than direct precursors, of carcinoma.

MORPHOLOGY

Several morphologically distinct patterns of proliferative breast disease without atypia are recognized:

- **Epithelial hyperplasia.** Normal breast ducts and lobules are lined by a double layer of myoepithelial cells and luminal cells (Fig. 23.6A). In epithelial hyperplasia, increased numbers of both luminal and myoepithelial cell types fill and distend ducts and lobules. Irregular lumens can often be discerned at the periphery of the cellular masses (Fig. 23.6B). Epithelial hyperplasia is usually an incidental finding.
- **Sclerosing adenosis.** There are an increased number of acini that are compressed and distorted in the central portion of the lesion. On occasion, stromal fibrosis completely compresses the lumens to create the appearance of solid cords or double strands of cells lying within dense stroma, a pattern that superficially resembles invasive carcinoma (Fig. 23.7). Sclerosing adenosis may come to attention as a palpable mass, a radiologic density, or calcifications.
- **Complex sclerosing lesion.** These lesions have components of sclerosing adenosis, papilloma, and epithelial hyperplasia. One member of this group, the radial sclerosing lesion ("radial scar"), has an irregular shape and closely mimics invasive carcinoma mammographically, grossly, and histologically (Fig. 23.8). A central nidus of entrapped glands in a hyalinized stroma is surrounded by long radiating projections into stroma. The term radial scar is a misnomer, as it is not associated with prior trauma or surgery.
- **Papilloma.** Papillomas grow within a dilated duct and are composed of multiple branching fibrovascular cores (Fig. 23.9). Epithelial hyperplasia and apocrine metaplasia are frequently present. Large duct papillomas are situated in the lactiferous sinuses of the nipple and are usually solitary. Small duct papillomas are commonly multiple and located deeper within the ductal system. More than 80% of large duct papillomas produce a nipple discharge, which may be bloody due to torsion of the stalk, leading to infarction. Serous discharge results from intermittent blockage and release of normal breast secretions or irritation of the duct by the papilloma. Most small duct papillomas come to clinical attention as small palpable masses, or as densities or calcifications seen on mammograms.

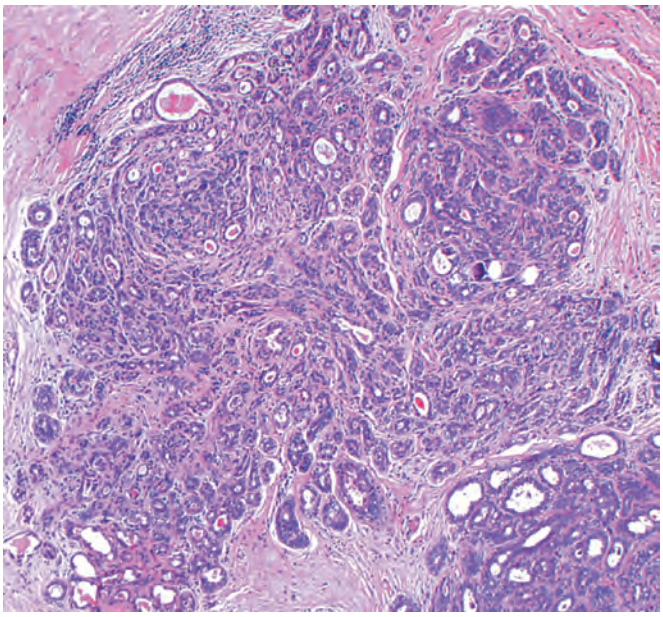


Figure 23.7 Sclerosing adenosis. The involved terminal duct lobular unit is enlarged, and the acini are compressed and distorted by dense stroma. Calcifications are present within some of the lumens. Unlike carcinomas, the acini are arranged in a swirling pattern, and the outer border is well circumscribed.

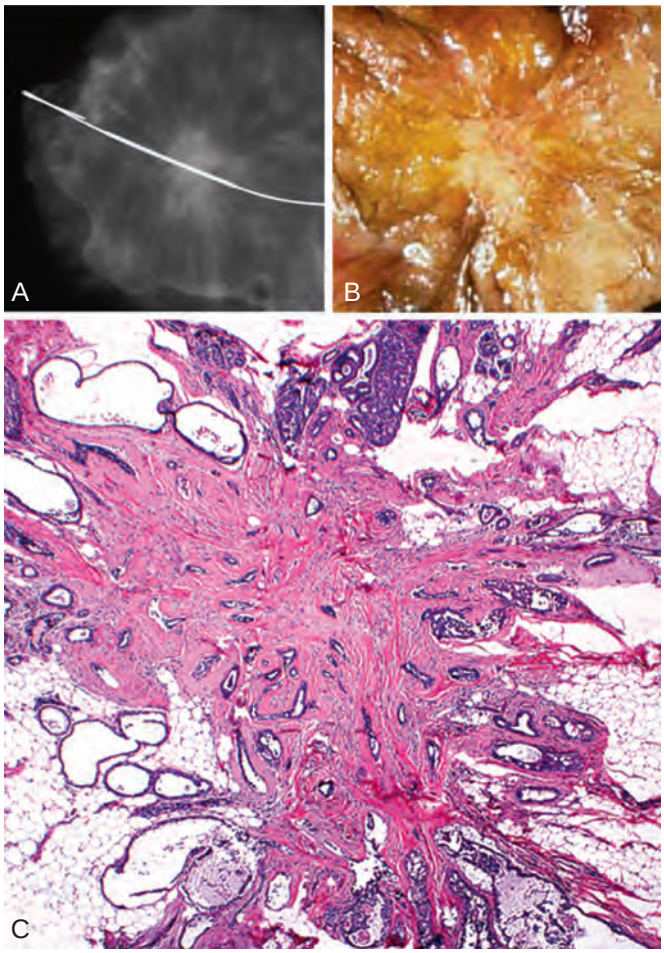


Figure 23.8 Radial sclerosing lesion. (A) Radiograph shows an irregular central mass with long radiodense projections. (B) Grossly the mass appears fibrotic and has irregular borders, but it is not as firm as an invasive carcinoma. (C) The mass consists of a central nidus of small tubules entrapped in a densely fibrotic stroma and numerous projections containing epithelium with varying degrees of cyst formation and hyperplasia.

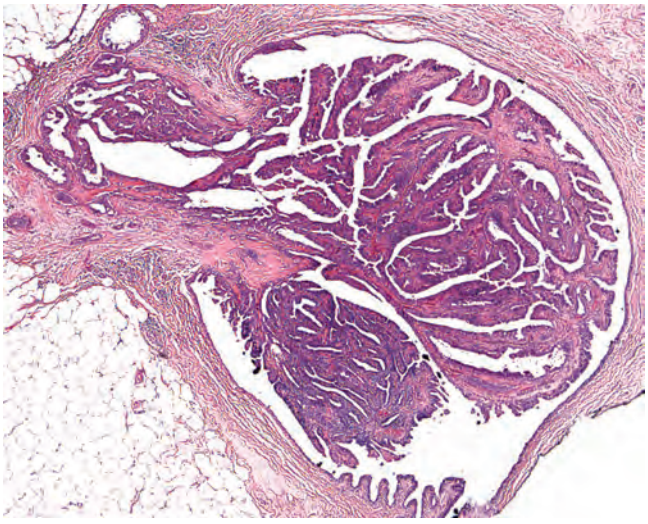


Figure 23.9 Intraductal papilloma. A central fibrovascular core extends from the wall of a duct. The papillae arborize within the lumen and are lined by myoepithelial and luminal cells.

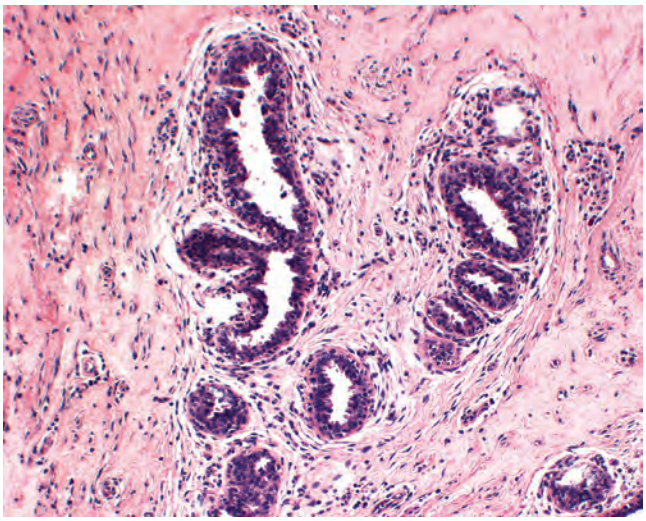


Figure 23.10 Gynecomastia. Breast enlargement in males is due to an increase in the number of ducts accompanied by dense cellular stroma. Lobule formation is absent.

Gynecomastia

Gynecomastia (enlargement of the male breast) is the only benign lesion seen with any frequency in the male breast. It presents as a button-like subareolar enlargement and may be unilateral or bilateral. Microscopically, there is an increase in dense collagenous connective tissue associated with epithelial hyperplasia of the duct lining (Fig. 23.10). Lobule formation is almost never observed.

Gynecomastia occurs as a result of an imbalance between estrogens, which stimulate breast tissue, and androgens, which counteract these effects. It may appear during puberty, in the very aged, or at any time during adult life when there is cause for hyperestrinism. The most important of these is cirrhosis of the liver, since this organ is responsible for metabolizing estrogen. In older males, gynecomastia may stem from a relative increase in estrogens as testicular androgen production falls. Drugs such as alcohol, marijuana, heroin, antiretroviral therapy, and anabolic steroids have been associated with gynecomastia. Rarely, gynecomastia occurs as part of Klinefelter syndrome (XXY karyotype) or in association with functioning testicular neoplasms, such as Leydig cell or Sertoli cell tumors.

Proliferative Breast Disease With Atypia

Atypical hyperplasia is a clonal proliferation having some, but not all, of the histologic features of carcinoma in situ. It is associated with a moderately increased risk of carcinoma and is divided into two forms, atypical ductal hyperplasia and atypical lobular hyperplasia. Atypical ductal hyperplasia is present in 5% to 17% of specimens from biopsies performed for calcifications. Atypical lobular hyperplasia is an incidental finding and is found in fewer than 5% of biopsies.

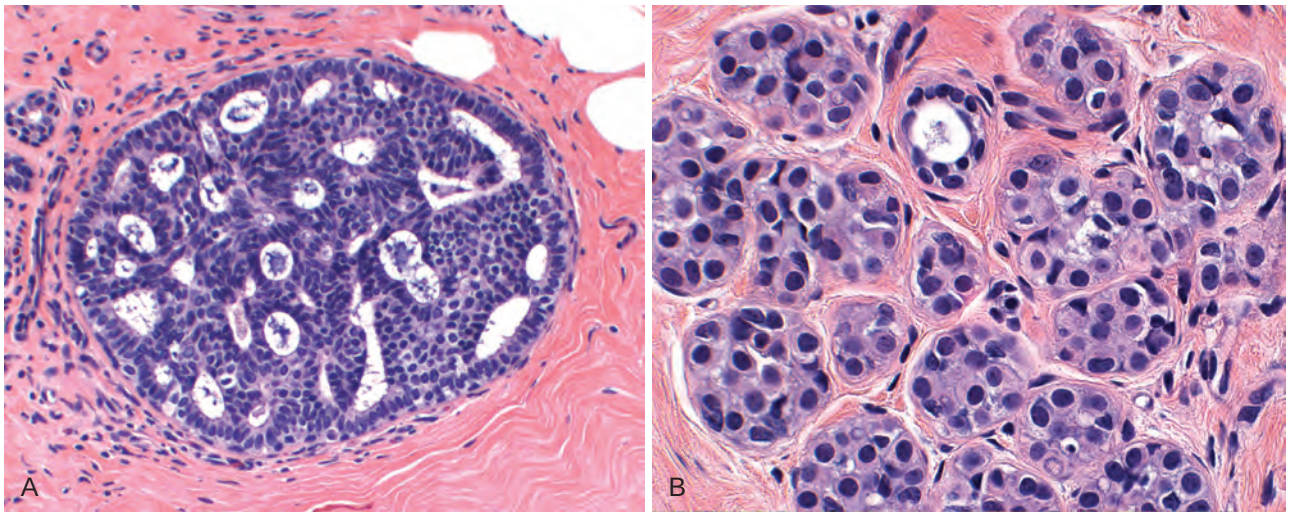


Figure 23.11 (A) Atypical ductal hyperplasia. A duct is filled with a mixed population of cells consisting of oriented columnar cells at the periphery and more rounded cells within the central portion. Although some of the spaces are round and regular, the peripheral spaces are irregular and slit-like. These features are highly atypical, but fall short of a diagnosis of ductal carcinoma in situ. (B) Atypical lobular hyperplasia. A population of monomorphic small, round, loosely cohesive cells partially fills a lobule. Although the cells are morphologically identical to the cells of lobular carcinoma in situ, the extent of involvement is not sufficient for this diagnosis.

MORPHOLOGY

Atypical ductal hyperplasia is recognized by its histologic resemblance to DCIS. It consists of a relatively monomorphic proliferation of regularly spaced cells, sometimes forming cribriform spaces. It is distinguished from DCIS in that it only partially fills involved ducts (Fig. 23.11A).

Atypical lobular hyperplasia consists of cells identical to those of lobular carcinoma in situ (LCIS) (described later), but the cells do not fill or distend more than 50% of the acini within a lobule (Fig. 23.11B).

Atypical ductal hyperplasia and atypical lobular hyperplasia express high levels of estrogen receptor (ER), have a low rate of proliferation, and may have acquired chromosomal aberrations such as losses of 16q and 17p or gains of 1q, features also found in low-grade carcinoma in situ and ER-positive invasive breast cancer. Atypical lobular hyperplasia also shows loss of E-cadherin expression, a feature it shares with LCIS (discussed later).

Clinical Significance of Benign Epithelial Changes

Epidemiologic studies have established the association of benign histologic changes with the later development of invasive cancer (Table 23.1). Proliferative changes are associated with a 1.5- to 2-fold increased risk compared to nonproliferative changes, while proliferative disease with atypia confers a 4- to 5-fold increased risk. Both breasts are at increased risk. Risk reduction can be achieved by bilateral prophylactic mastectomy or treatment with estrogen antagonists, such as tamoxifen. However, fewer than 20% of women with atypical hyperplasia develop breast cancer, and therefore many choose close clinical and radiologic surveillance over intervention.

Table 23.1 Epithelial Breast Lesions and Risk of Developing Invasive Carcinoma

Pathologic Lesions	Relative Risk (Absolute Lifetime Risk)[a]
Nonproliferative breast changes (mild hyperplasia, duct ectasia, cysts, apocrine metaplasia, adenosis, fibroadenoma without complex features)	1.0 (~3%)
Proliferative disease without atypia (moderate or florid hyperplasia, sclerosing adenosis, complex sclerosing lesion, fibroadenoma with complex features)	1.5–2 (~5%–7%)
Proliferative disease with atypia (atypical ductal hyperplasia, atypical lobular hyperplasia)	4–5 (~13%–17%)
Carcinoma in situ (lobular carcinoma in situ, ductal carcinoma in situ)	8–10 (~25%–30%)

[a]Relative risk is the likelihood of developing invasive carcinoma compared to women without any risk factors. Absolute lifetime risk is the percentage of women expected to develop invasive carcinoma in the absence of an intervention.

KEY CONCEPTS

BENIGN EPITHELIAL LESIONS

- Benign epithelial lesions usually do not cause symptoms but are frequently detected as mammographic calcifications or densities.
- The majority are not precursors of cancer.
- These lesions are classified according to the subsequent risk of cancer in either breast.
- Although risk reduction can be achieved by surgery or chemoprevention, the majority of women will not develop cancer, and many women choose surveillance instead of intervention.

CARCINOMA OF THE BREAST

Breast carcinoma is the most common and deadly malignancy of women globally; each year, 1.7 million women are diagnosed, and one in three of those afflicted die of disease. Although the incidence of breast cancer is four to seven times higher in the United States and Europe than elsewhere, the worldwide incidence and mortality is increasing at an alarming rate, and by 2020 it is estimated that 70% of cases will be in lower income countries. The factors underlying this trend are thought to be social changes that increase breast cancer risk—specifically delayed childbearing, fewer pregnancies, and reduced breastfeeding—combined with a lack of access to optimal health care.

The lifetime risk of breast cancer is 1 in 8 for women living to age 90 in the United States. In 2019, over 260,000 women in the United States were diagnosed with invasive breast cancer, and more than 40,000 women died of the disease—a toll among cancers second only to lung cancer. It is both ironic and tragic that a neoplasm arising in an exposed organ, readily accessible to self-examination and clinical surveillance, continues to exact such a heavy toll.

All breast cancers can be separated into three major groups defined by the expression of two proteins, ER and HER2 (also known as ERBB2). In this chapter, "luminal" cancers are defined as being positive for ER and negative for HER2. "HER2" cancers are defined as cancers overexpressing HER2 and can be either ER-positive or ER-negative. "Triple negative breast cancers" (TNBCs) are cancers that are negative for ER and HER2. These cancers are termed "triple negative" because they also fail to express progesterone receptor (PR), which is under the control of ER. These three groups of cancers differ with regard to patient characteristics, pathologic features, treatment response, metastatic patterns, time to relapse, and outcome and will be discussed in more detail later. However, it is important to note that breast cancer is biologically heterogeneous and that each of these three groups is comprised of numerous clinically important subtypes.

Incidence and Epidemiology

Breast cancer is rare in women younger than age 25 and increases in incidence rapidly after age 30 (Fig. 23.12). The incidence of TNBC and HER2 cancer plateaus in middle age, whereas the incidence of luminal cancer peaks later in life. As a result, TNBCs and HER2 cancers comprise almost half of cancers in young women and fewer than 20% of cancers in older women.

Starting in the 1980s, a marked increase in the apparent incidence of breast cancer was noted in older women. This change appears to be attributable to two factors: the introduction of mammographic screening and the increased use of postmenopausal hormonal therapy in older women. After 2000, the number of newly diagnosed breast cancers fell in older women, probably due to a leveling off of the rate of mammographic screening (approximately 65% to 75% of eligible women) and decreased use of postmenopausal hormone therapy. In contrast, over the same time the incidence of breast cancer remained relatively constant in younger women.

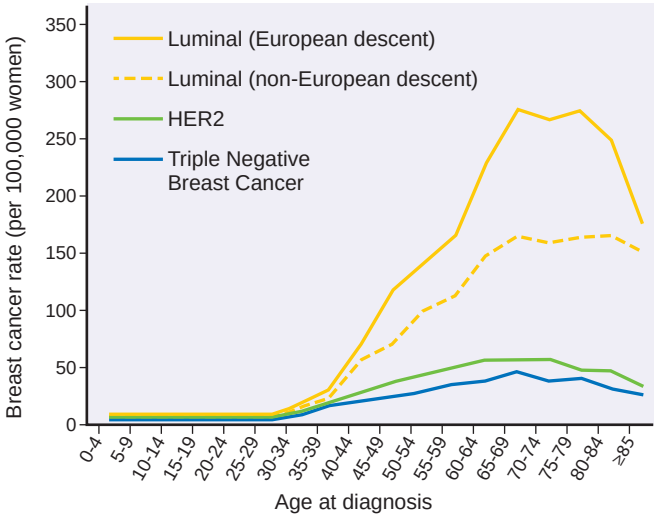


Figure 23.12 Incidence of luminal (ER-positive/HER2-negative), HER2 (HER2-positive), and triple negative (ER-negative/HER2-negative) breast cancers according to age. Rates are per 100,000 women. Triple negative *(solid blue line)* and HER2 *(solid green line)* cancers have a relatively constant incidence after age 40 years. In contrast, luminal cancers show a marked increase in incidence with age. This increase is greatest for women of European descent *(solid yellow line)* and less pronounced for women other ethnic backgrounds (African, Hispanic, and Asian) *(broken yellow line)*. *TNBC*, Triple negative breast cancer.

Breast cancer incidence and biology vary with ethnicity. The incidence of breast cancer is highest in women of European descent; in this group, the average age at diagnosis is 63 years, and only 20% of cases are diagnosed at an age younger than 50 years. In contrast, the average age at diagnosis for women of African descent is 59 years, and 35% of cancers are diagnosed at ages below 50. For Hispanic women, the average age at diagnosis is 56, and 20% are diagnosed at ages below 50. The "excess" cancers in women of European descent are mostly of luminal type (Fig. 23.12). Thus, TNBCs and HER2 cancers make up a greater proportion of cancers occurring in other ethnic groups.

The risk of death in those who develop invasive breast cancer has gradually declined in both younger and older women, most recently by 1% to 2% per year; the current overall risk of death is about 20%. This decrease is attributed to mammographic screening as well as to more effective treatment modalities. However, the decline in the death rate has lagged behind in women of African descent, who have the highest mortality rate. Although this difference is explained partly by unequal access to care, breast cancers in women of African descent are also more likely to be biologically aggressive (e.g., TNBC) and to fall into molecular subtypes that are difficult to treat.

Risk Factors

The most important risk factors are gender (99% of those affected are female), age, lifetime exposure to estrogen, genetic inheritance, and, to a lesser extent, environmental and lifestyle factors (Table 23.2). There is concern (but no proof) that environmental contaminants, such as organochlorine pesticides and certain plastics, have estrogenic effects on humans that may increase the risk of breast cancer. The

Table 23.2 Risk Factors for Developing Breast Cancer

Risk Factors	Relative Risk[a]
Female gender Increasing age Germline mutations of high penetrance Strong family history (>1 first-degree relative, young age, multiple cancers) Personal history of breast cancer High breast density	>4.0
Germline mutations of moderate penetrance High-dose radiation to chest at young age Family history (1 first-degree relative)	2.1–4.0
Early menarche (age <12 years) Late menopause (age >55 years) Late first pregnancy (age >35 years) Nulliparity Absence of breastfeeding Exogenous hormone therapy Postmenopausal obesity Physical inactivity High alcohol consumption	1.1–2.0

[a]Relative risk is the likelihood of developing invasive carcinoma compared to women without any risk factors.

major factors that decrease risk are early pregnancy (prior to 20 years of age) and prolonged breastfeeding. Societal changes in these factors in low income countries are thought to be contributing to the rise in breast cancer incidence.

Surgical and medical interventions also can decrease risk. Bilateral prophylactic mastectomy decreases risk by about 90%. Chemoprevention using ER antagonists decreases the incidence of ER-positive cancers. These interventions are mainly offered to women at very high risk for breast cancer.

Molecular Classification and Pathogenesis

Several different approaches have been used to subclassify breast cancer into clinically meaningful subtypes. Based on gene expression profiling, breast cancers cluster into three main groups: "luminal" (predominantly ER-positive/HER2-negative), "HER2-enriched" (predominantly HER2-positive), and "basal-like" (predominantly ER-negative/HER2-negative). Because these molecular subtypes correlate reasonably well with ER and HER2 protein expression (Fig. 23.13), which is easily assessed by standard clinical assays, molecular subtyping by expression profiling is not a routine part of breast cancer classification.

Like other cancers, breast cancers arise through several pathways that involve the stepwise acquisition of driver mutations in epithelial cells. Breast cancer develops in a hormonal milieu that facilitates mutagenesis and the outgrowth of abnormal clones. Carcinomas associated with germline mutations in cancer genes make up the minority of carcinomas, but provide generally important insights into the molecular pathogenesis of breast cancer; thus, we begin by discussing tumors with a strong familial basis and then turn to sporadic breast cancers.

Pathogenesis of Familial Breast Cancer

It is believed that one-quarter to one-third of breast cancers occur due to inheritance of a susceptibility gene or genes. Single gene mutations with moderate to high penetrance ("penetrance" refers to the risk of an individual with the mutated gene developing cancer) accounts for 8% to 17% of breast carcinomas (Table 23.3). Inheritance also is thought to play a role in an additional 15% to 20% of women based on a positive family history, defined as an affected first-degree relative (mother, sister, or daughter), cancer in multiple relatives, and early-onset cancers. In these women, inheritance of a single susceptibility gene with low penetrance or combinations of genes that interact to increase risk is believed to be the culprit.

The most important high penetrance susceptibility genes for familial breast cancer are tumor suppressor genes that regulate genomic stability or are involved in pro-growth signaling pathways (see Table 23.3). Risk is inherited in an autosomal dominant fashion and stems from alleles with loss-of-function mutations. As with other tumor suppressor genes, cancer development is associated with sporadic loss-of-function mutations in the single normal copy of the gene.

Mutations in *BRCA1* and *BRCA2* are responsible for 80% to 90% of single gene familial breast cancers and about 3% to 6% of all breast cancers. Penetrance, age of onset, and susceptibility to other types of cancers vary according to the specific *BRCA1* and *BRCA2* mutation that is inherited, but most carriers develop breast cancer by the age of 70 years. Mutations in *BRCA1* also markedly increase the risk of ovarian carcinoma, which occurs in 20% to 40% of carriers. *BRCA2* confers a smaller risk for ovarian carcinoma (10% to 20%) but is associated more frequently with male breast cancer. *BRCA1* and *BRCA2* carriers also are at higher risk for other epithelial cancers, such as prostatic and pancreatic carcinoma.

BRCA1 (on chromosome 17q21) and *BRCA2* (on chromosome 13q12.3) are both large genes, and hundreds of different mutations distributed throughout their coding regions have been associated with familial breast cancer. The frequency of mutations that increase breast cancer risk is about 1 in 400 persons in the general population, and inconsequential polymorphisms are common. As a result, genetic testing is complex and is generally restricted to individuals with a strong family history or those belonging to certain ethnic groups. For example, in Ashkenazi Jewish populations, about 1 in 40 individuals carry one of three specific mutations, two in *BRCA1* and one in *BRCA2*. Identification of carriers is important, since increased surveillance and prophylactic mastectomy and salpingo-oophorectomy can reduce cancer-related morbidity and mortality.

BRCA1-associated breast cancers are commonly poorly differentiated, often have characteristic morphologic features (described later), and usually fall in the TNBC subgroup. *BRCA2*-associated breast carcinomas also tend to be poorly differentiated, but are more often ER-positive than *BRCA1* cancers.

Other tumor suppressor genes associated with germline mutations that convey a high risk for breast cancer, often as part of well-described syndromes, include *TP53* (Li-Fraumeni syndrome), *PTEN* (Cowden syndrome), *STK11* (Peutz-Jeghers syndrome), *CDH1* (hereditary diffuse gastric cancer syndrome), and *PALPB2* (see Table 23.3). Moderately penetrant (10% to 30% risk) mutations occur in the tumor suppressor genes *ATM* (a cause of ataxia telangiectasia when homozygous mutations are present) and *CHEK2*.

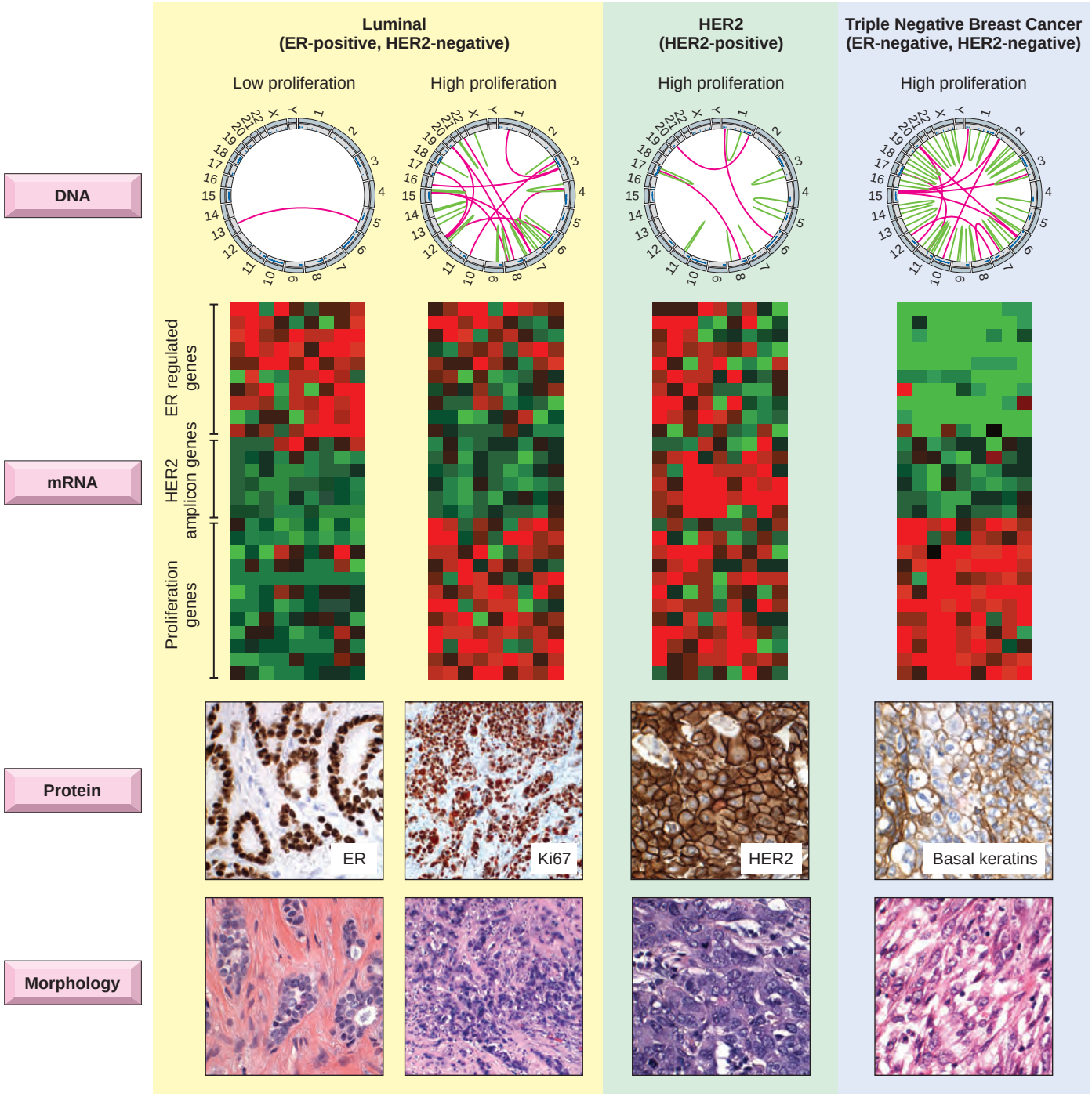


Figure 23.13 Molecular classification of invasive breast cancer. Three major groups of breast cancer are distinguished by characteristic changes in DNA, mRNA, protein, and morphology. Genomic abnormalities are shown in circos plots (Chapter 7), which present a snapshot of all of the genomic abnormalities within a particular tumor; these abnormalities are mapped onto the chromosomes, which are displayed at the periphery of a circle. *Green loops* show intrachromosomal rearrangements, while *red loops* show interchromosomal rearrangements. Gene expression (mRNA) profiling measures relative levels of mRNA expression. *Red* indicates a relative increase; *green,* a relative decrease; and *black,* no change in levels. Genes are arrayed from top to bottom and tumors from left to right. Immunohistochemical studies detect proteins using specific antibodies visualized with a brown chromogen. *Luminal (ER-positive/HER2-negative) cancers* are diverse, ranging from well-differentiated cancers with low proliferative rates and few chromosomal changes to poorly differentiated cancers with high proliferative rates and large numbers of chromosomal rearrangements. All of these cancers express ER (an estrogen-dependent transcription factor). Proliferation is estimated by counting mitoses or by staining for cell cycle–specific proteins such as Ki-67. *HER2-positive cancers* may be ER-positive or ER-negative, but when ER is present, levels are typically low. HER2 positivity can be detected as an increase in *HER2* gene copy number, an increase in HER2 mRNA, or an increase in HER2 protein, as shown here. *Triple negative breast cancers (ER-negative/HER2-negative; this group largely overlaps with "basal-like" carcinomas defined by mRNA expression)* are characterized by genomic instability (denoted by numerous chromosomal changes), a high proliferative rate, and expression of many proteins typical of myoepithelial cells (e.g., basal keratins).

Table 23.3 Most Common Single Gene Mutations Associated With Hereditary Susceptibility to Breast Cancer

Gene (Syndrome)	% of Single Gene Cancers[a]	Risk of Breast Cancer to Age 70[b]	Other Cancers	Comments
High Penetrance Germline Mutations (>4-fold increased risk; 3%–7% of breast cancers)				
BRCA1 (familial breast and ovarian cancer)	~55%	~40%–90%, females; 1%, males	Ovarian (~20%–40%), fallopian tube, pancreas, prostate, others	Majority of cancers are TNBC
BRCA2 (familial breast and ovarian cancer)	~35%	~30%–60%, females; 6%, males	Ovarian (~10%–20%), pancreas, prostate, others	Majority of cancers are ER positive. Biallelic mutations cause a form of Fanconi anemia.
TP53 (Li-Fraumeni)	<1%	~50%–60%, females; <1%, males	Sarcoma, leukemia, brain tumors, others	Majority of cancers are ER and HER2 positive
PTEN (Cowden)	<1%	~20%–80%, females; <1%, males	Thyroid, endometrium, others	Also associated with benign tumors
STK11 (Peutz-Jeghers)	<1%	~40%–60%, females	Ovarian, colon, pancreas, others	Also associated with benign colon polyps
CDH1 (hereditary diffuse gastric cancer)	<1%	~50%, females	Gastric signet ring cell carcinoma, colon	Majority of cancers are lobular in type
PALPB2 (hereditary breast cancer)	<1%	~30%–60%, females; <1%, males	Pancreas, prostate	Biallelic mutations cause a form of Fanconi anemia
Moderate Penetrance Germline Mutations (2- to 4-fold increased risk; 5% to 10% of breast cancers)				
ATM (ataxia-telangiectasia)	~5%	~15%–30%, females		Biallelic mutations cause ataxia-telangiectasia
CHEK2 (hereditary breast cancer)	~5%	~10%–30%, females	Prostate, thyroid, colon, kidney	Majority of cancers are ER positive

[a]The percentage of all breast cancers that are associated with a germline mutation conferring an increased risk of breast cancer.
[b]Risk for specific patients can vary with the specific mutation and the presence of other gene mutations.
ER, Estrogen receptor; *TNBC*, triple negative breast cancer.

Most of these genes play complex and interrelated roles in maintaining genomic integrity. Under normal circumstances, cells that sustain DNA damage undergo cell cycle arrest and either repair their DNA or die by apoptosis. *ATM* senses DNA damage and "activates" p53, the so-called guardian of the genome, which has a direct role in inducing cell cycle arrest and, if DNA repair is unsuccessful, apoptosis. *BRCA1, BRCA2,* and *CHEK2* all have important functions in repair of double-stranded DNA breaks. If any of these functions are impaired, the likelihood of permanent DNA damage increases, leading to potentially oncogenic mutations that will be passed to daughter cells.

What is mysterious is why malfunction of these genes, particularly *BRCA1* and *BRCA2*, is more highly associated with breast cancer than other cancers. BRCA1 and BRCA2 are components of a large complex of proteins that are required to repair double-stranded DNA breaks through a process called *homologous recombination,* in which a normal sister chromatid is used as a template for repairing the DNA break. BRCA1 and BRCA2 are expressed ubiquitously, so the link to breast cancer cannot be explained by tissue-specific patterns of gene expression. An alternative possibility is that breast (and ovarian) epithelial cells may be particularly prone to suffer the type of DNA damage that BRCA1 and BRCA2 are required to repair. BRCA1 also interacts with protein complexes that regulate chromatin structure and has a role in transcription, and it also remains possible that its tumor suppressive role, at least in part, involves functions that are independent of DNA repair.

Pathogenesis of Sporadic Breast Cancer

As with other forms of carcinoma, multiple pathways resulting in different types of breast carcinoma have been identified. The mutation that initiates the process, as demonstrated by inherited cancers, appears to strongly influence the phenotype of the cancer that ultimately develops, as may the specific cell type in which the initiating event occurs. Pathways leading to the three major molecular types of breast carcinoma, as well as how the biologic features of these types relate to response to treatment, recurrence, metastatic pattern, and survival, are summarized in Fig. 23.14 and Table 23.4 and are described next.

Luminal (ER-positive/HER2-negative) cancers arise via the dominant pathway of breast cancer development, constituting 50% to 65% of cases. By gene expression profiling, ER-positive cancers fall into the "luminal" subgroup, a pattern of gene expression that is dominated by a large number of genes that are regulated by estrogen. Luminal breast cancer has the widest spectrum of histologic grades and proliferation rates of the three molecular types. Cancers with high expression of ER usually also express high levels of PR, which is itself upregulated by estrogen and ER; such ER-positive/PR-positive tumors are usually well differentiated and slow-growing. In contrast, carcinomas with low ER and absent PR tend to lie at the other end of the spectrum, as they are typically poorly differentiated and have a high proliferative rate.

The major risk factor for luminal breast cancer is estrogen exposure (see Table 23.2), which may promote the

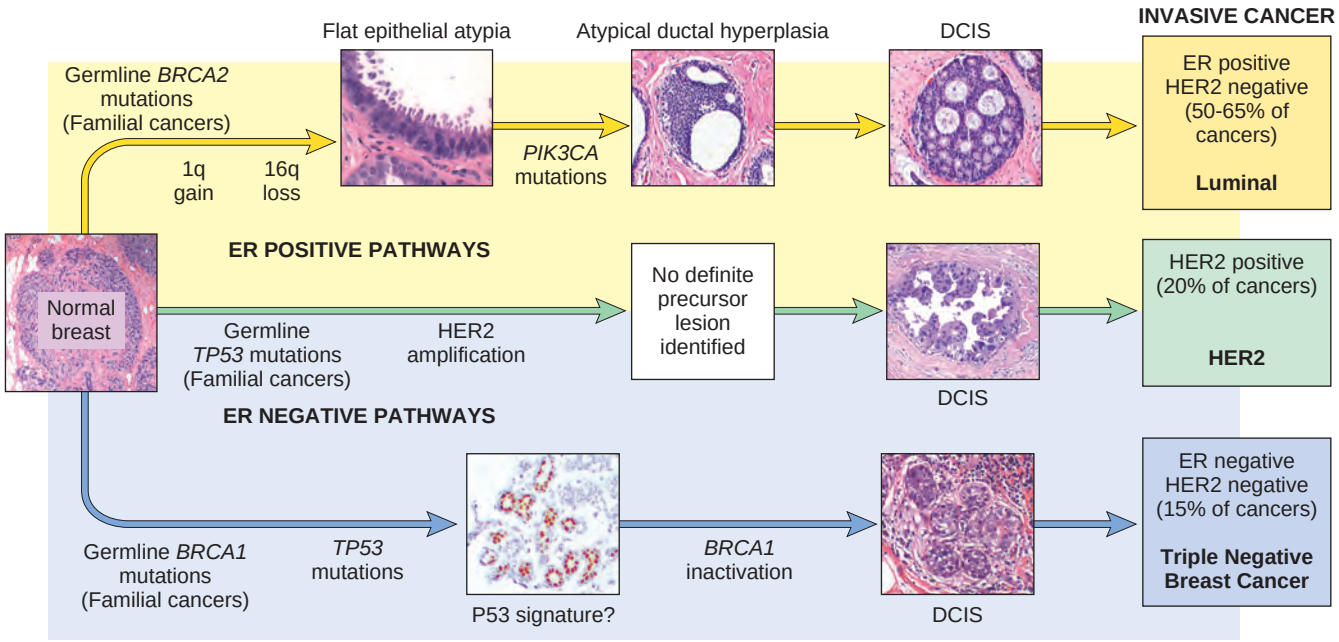


Figure 23.14 Major pathways of breast cancer development. Three main pathways have been identified. The most common pathway *(yellow arrow)* leads to luminal (ER-positive) carcinomas. Recognizable non-obligate precursor lesions include flat epithelial atypia and atypical hyperplasia. A less common pathway *(blue arrow)* leads to triple negative breast cancer (ER-negative/HER2-negative). A possible precursor lesion consisting of morphologically normal cells that overexpress p53 has been identified (analogous to the "p53 signature lesions" for ovarian carcinoma). The third pathway *(green arrow)* consists of HER2-positive cancers. Amplification of HER2 can occur in either ER-positive or ER-negative lesions. A definite HER2-positive precursor lesion has not been identified. See text for other details. *DCIS,* Ductal carcinoma in situ.

Table 23.4 Molecular Subtypes of Invasive Breast Cancer

Defining Features	Luminal (ER-Positive/HER2-Negative)		HER2 (HER2 Positive)	TNBC (ER-Negative/ HER2-Negative)[a]
Percent of breast cancers	~40%–55% (low to moderate proliferation)	~10% (high proliferation)	~20%	~15%
Most similar group defined by mRNA profiling[b]	Luminal A	Luminal B	HER2-enriched (ER-negative), luminal B (ER-positive)	Basal-like
Most common gene mutations	*PIK3CA* (45%), *TP53* (12%)	*PIK3CA* (29%), *TP53* (29%)	*PIK3CA* (39%), *TP53* (70%–80%)	*PIK3CA* (9%), *TP53* (70%–80%)
Typical special histologic types	Tubular, grade 1 or 2 lobular, mucinous, papillary	Grade 3 lobular	Some apocrine, some micropapillary	Medullary features, metaplastic[c]
Typical patient groups	Older women, men, cancers detected by mammographic screening	*BRCA2* mutation carriers	Young women, *TP53* mutation carriers (ER positive)	Young women, women of African heritage, *BRCA1* mutation carriers
Complete response to chemotherapy	<10%	~10%	ER positive ~15%; ER negative ~30%–60%	~30%
Metastatic pattern	Bone (70%), more common than viscera (25%) or brain (<10%)	Bone (80%) more common than viscera (30%) or brain (10%)	Bone (70%), viscera (45%), and brain (30%) all are common	Bone (40%), viscera (35%), and brain (25%) all are common
Relapse pattern	Low rate over many years, long survival possible with bone metastases	Early peak at <10 years, late recurrence possible	Bimodal with early and late (10 years) peaks	Early peak at <8 years, late recurrence rare, survival with metastases rare

[a]TNBC lacks expression of ER, progesterone receptor, and HER2.
[b]The three major groups of cancer identified by protein expression or mRNA profiling largely overlap but are not identical. "Luminal B" can refer to ER-positive cancers with high proliferation with or without HER2 expression.
[c]Some rare special histologic types have a more favorable prognosis than this group as a whole (e.g., adenoid cystic carcinoma, secretory carcinoma, low-grade adenosquamous carcinoma).

ER, Estrogen receptor; *mRNA,* messenger RNA; *TNBC,* triple negative breast cancer.

development of breast cancer through several mechanisms. Estrogen increases the local production of growth factors, such as transforming growth factor α, platelet-derived growth factor, and fibroblast growth factor, and regulates the expression of dozens of genes in breast epithelial cells that may directly contribute to tumor growth and development. Estrogen exposure also stimulates the proliferation of breast epithelial cells during puberty, menstrual cycles, and pregnancy, thereby increasing the number of cells that are "at risk" for transformation. The DNA replication that attends cellular proliferation is conducive to the accumulation of mutations, and the lull in cell division that occurs during the latter part of the menstrual cycle may allow time for defective DNA repair to occur and for mutations to become "fixed" in the genome. Repetition of this process during each cycle may underlie the association between the cumulative number of menstrual cycles a woman experiences and her risk of developing breast cancer, as well as the strong association between luminal cancers and age (see Fig. 23.12). A clear measure of the importance of estrogen is found in the therapeutic benefits of estrogen antagonists, which reduce the development of luminal cancers in women at high risk, and the increased incidence of luminal cancers in women treated with postmenopausal hormone therapy.

Once a clonal population of mutated ER-positive cells emerges, estrogen may also enhance the transformation of such precursor lesions to fully malignant cancers. ER-positive precursor lesions include flat epithelial atypia and atypical ductal and lobular hyperplasia. All of these may be associated with genomic changes found in invasive carcinomas, including gains of chromosome 1q, losses of chromosome 16q, and activating mutations in *PIK3CA,* the gene that encodes phosphoinositide-3 kinase (PI3K), which you will recall is an important component of signaling pathways downstream of growth factor receptors (Chapter 7). These are considered the earliest recognizable precursors of luminal breast cancers. It must be emphasized, however, that few of these lesions progress to malignancy, presumably because multiple additional events are needed before the cells acquire all of the hallmarks of cancer.

The precursor lesions mentioned above are mainly associated with the development of luminal cancers with low growth fractions and indolent clinical behavior; such tumors by gene expression profiling belong to the "luminal A" subgroup (see Fig. 23.13). More aggressive, high-grade luminal cancers also occur and cluster within the "luminal B" subgroup, defined by high levels of expression of genes associated with proliferation. These aggressive forms of luminal cancer share some genomic changes with low-grade carcinomas but tend to have a greater burden of chromosomal aberrations. They sometimes arise from low-grade luminal cancers following the acquisition of mutations in genes that regulate genomic stability such as *TP53* and also frequently occur in patients with germline *BRCA2* mutations.

Cancers detected by mammographic screening are usually small luminal cancers limited to the breast (see Table 23.4). These cancers typically respond well for many years to antiestrogen therapy and have a more favorable outcome than the other types of breast cancer. The rate of recurrence is low, and even when these carcinomas metastasize (most often to bone), they often can be held in check for a decade or longer by treatment with antiestrogens (Fig. 23.15). By contrast, cytotoxic chemotherapy provides little benefit, particularly in treating "luminal A" cancer with low growth fractions. Some luminal cancers eventually escape from estrogen dependence through several mechanisms, including outgrowth of clones that lack ER expression, compensatory alterations in related growth factor signaling pathways, or acquisition of mutations in the ER gene *(ESR1)* that lead to estrogen-independent ER function.

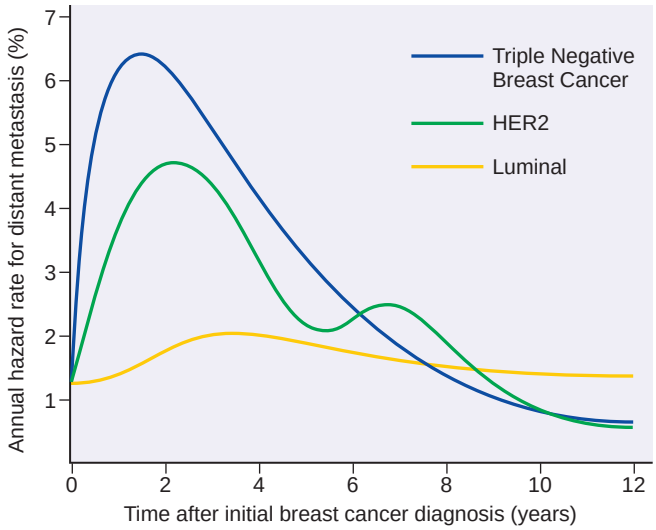


Figure 23.15 Each molecular group of breast cancer has characteristic patterns of recurrence over time. Luminal (ER-positive) cancers have the lowest rate of recurrence in the first 10 years, but recurrences continue with a steady rate over a long period of time. In contrast, almost all recurrences of triple negative breast cancer (ER-negative/HER2-negative) occur within the first 8 years. Recurrences after this time are rare. HER2-positive cancers show a mixed pattern with both early and late peaks. Although not yet fully explained, a late peak may be due to acquired resistance to targeted therapy or to selection of HER2-negative tumor cell populations.

HER2-positive cancers arise through a pathway that is strongly associated with amplification of the *HER2* gene on chromosome 17q. HER2 (also known as ERBB2) is a receptor tyrosine kinase that promotes cell proliferation and opposes apoptosis by stimulating the RAS- and PI3K-AKT signaling pathways. Cancers with *HER2* amplification constitute approximately 20% of all breast cancers and may be either ER-positive or ER-negative. This is the most common subtype of breast cancer in patients with germline *TP53* mutations (Li-Fraumeni syndrome).

The only common molecular mechanism for *HER2* overexpression is gene amplification, found in >95% of HER2-positive carcinomas. The size of the amplicon varies and typically includes at least 10 adjacent genes, some of which may cooperate with HER2 to promote tumor growth. In addition, these cancers characteristically have complex interchromosomal rearrangements and a high mutational load. The gene expression profile of HER2 cancers is dominated by transcripts encoded by *HER2,* co-amplified genes, and genes involved in pro-growth signaling pathways and proliferation. Variation in gene expression among HER2 cancers is largely based on their ER status and differing levels of expression of ER-regulated genes.

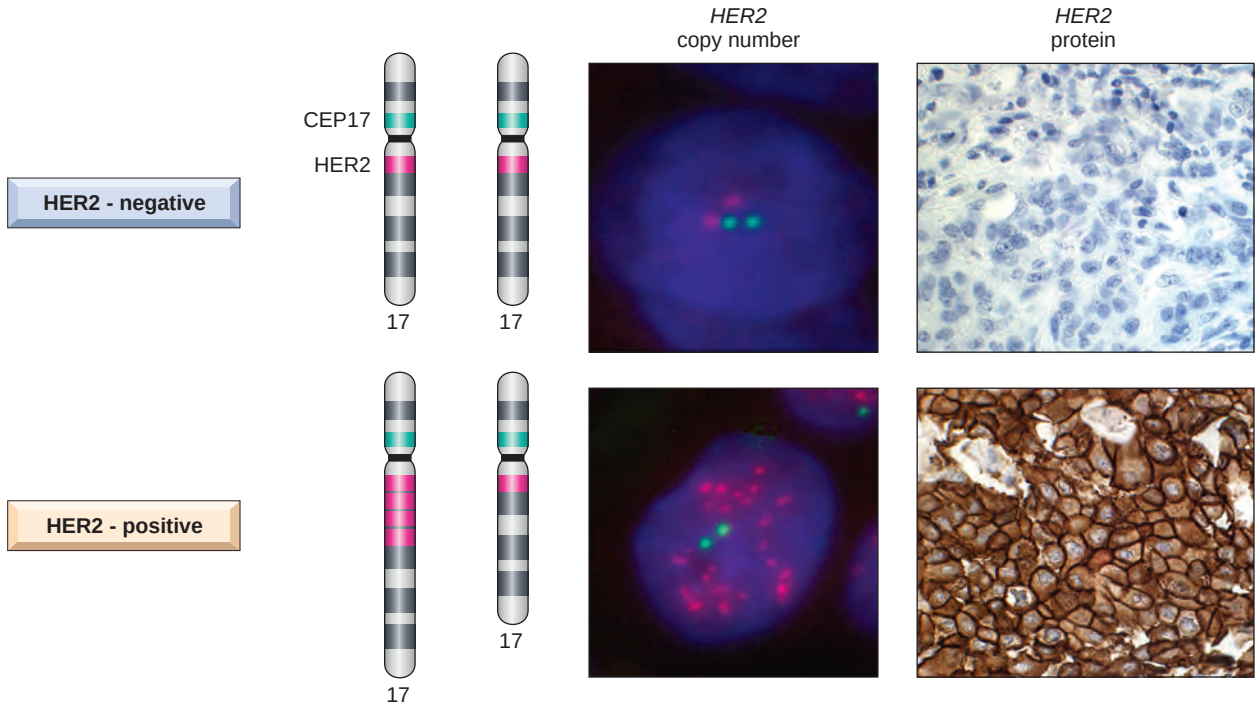


Figure 23.16 Identification of HER2-positive breast cancer. HER2 protein overexpression is virtually always caused by amplification of the region of chromosome 17q that contains the *HER2* gene. The increase in *HER2* gene copy number is detected by fluorescence in situ hybridization using a *HER2*-specific probe (red signal), which is typically co-hybridized to tumor cell nuclei with a second probe specific for the centromeric region of chromosome 17 (green signal), allowing the chromosome 17 copy number to be determined. Alternatively, HER2 protein overexpression in tumor cells can be detected by immunohistochemical staining with antibodies specific for HER2.

Clinically, HER2 carcinomas are diagnosed by detecting HER2 overexpression by immunohistochemistry or *HER2* gene amplification by in situ hybridization (Fig. 23.16). Before the implementation of HER2-targeted therapy, HER2-positive cancers had a poor clinical outcome. However, greater than half of patients with HER2 carcinomas have complete remissions when treated with antibodies that bind and block HER2 activity; such patients have an excellent prognosis (see Table 23.4). The remarkable efficacy of this form of therapy proves the importance of HER2 as an oncogenic "driver." However, not all HER2-positive carcinomas respond to targeted therapy, and some that initially respond become resistant (see Fig. 23.15). Multiple mechanisms of resistance to HER2 antagonists have been defined in cancers, and numerous therapeutic agents are under investigation in an effort to both improve response and overcome resistance.

TNBCs arise through an estrogen-independent pathway that is not associated with *HER2* gene amplification. They comprise about 15% of breast cancers and have a "basal-like" gene expression profile, so-called because many of the genes that comprise this signature are normally expressed in basally located myoepithelial cells. A possible precursor lesion has been identified that consists of lobular epithelial cells that are ER-negative and p53-positive, the latter being a feature that correlates with the presence of *TP53* mutations. These lesions resemble serous tubal intraepithelial carcinoma, the proposed precursor to serous carcinoma that is seen in the fallopian tubes of women with germline *BRCA1* mutations (see Chapter 22). As might be expected based on this similarity, TNBC also is associated with germline *BRCA1* mutations.

TNBC also shares other genetic features with serous ovarian carcinoma. Both are characterized by marked genomic instability due to defects in DNA repair by homologous recombination that manifests as complex copy number alterations and chromosomal rearrangements. In familial breast cancer, this defect is often related to germline *BRCA1* and *BRCA2* mutations. *BRCA1* mutations are rare in sporadic TNBC, but *BRCA1* is instead often silenced through epigenetic mechanisms and allelic loss. The only gene that is mutated in the majority of these cancers is *TP53*; presumably, other genes that contribute to the malignant phenotype are dysregulated by copy number variation or gene rearrangements.

Compared with luminal cancers, TNBC is more likely to present as a palpable mass and is less likely to be detected by mammographic screening (see Table 23.4). Cytotoxic therapy combined with agents that are selectively active against cancers with defective homologous recombination results in complete or almost complete responses in about a third of cases. The cancers that recur usually do so in the first 8 years after diagnosis (see Fig. 23.15). Metastases are often to visceral sites and brain and frequently result in death. Patients who survive 10 years are likely cured, as late recurrences are unusual. Development of more effective treatments for this breast cancer subtype has been challenging, as obvious drug targets (such as ER or HER2) are lacking. In addition, the genomic instability of these tumors leads to profound genetic heterogeneity, increasing the likelihood of emergence of more aggressive, therapy-resistant subclones. However, genomic instability also may lead to expression of tumor neoantigens, and one approach that shows some

promise in TNBC is use of immune checkpoint inhibitors (see Chapter 7), which are under evaluation in numerous clinical trials.

Finally, it must be said that as with other carcinomas that begin as in situ lesions, it is not clear what drives transition of any of the molecular subtypes of breast cancer from innocuous in situ lesions to potentially lethal invasive cancers. Maintenance of the basement membrane depends on the interaction of luminal cells, myoepithelial cells, and stromal cells in the local microenvironment. The majority of genomic changes observed in invasive carcinoma also are already present in carcinoma in situ, including major driver mutations, and genomic changes specific to invasive carcinoma have not been identified, suggesting that invasion may depend on changes in stromal cells rather than tumor cells. For example, in some in situ cancers, myoepithelial cells are mislocalized and reduced in number, and the associated basement membrane is thin and has gaps. Such changes may set the stage for stromal invasion by breast cancer cells. Stromal invasion also may expose cancer cells to increased host immune surveillance, and it may also be that alterations that enable cancer cells to evade the host immune response are necessary before effective invasion can occur.

With this as background, we next turn to discussion of the pathology of breast cancer.

KEY CONCEPTS

CARCINOMA OF THE BREAST

- Breast cancer is the most common non-skin malignancy in women and the second most common cause of cancer deaths in the United States.
- The most important risk factors for sporadic cancers in women are estrogenic stimulation and age.
- Approximately a quarter to a third of breast cancers are familial, being related to inheritance of genetic variants that increase breast cancer risk.
- High-risk genes associated with familial breast cancer include several involved with DNA repair and genomic stability, most notably *BRCA1*, *BRCA2*, and *TP53*.
- Breast cancers cluster into three major molecular groups, luminal (ER-positive), HER2, and triple negative, each with distinctive biologic and clinical features.
- Luminal cancers are further divided into two groups, A and B, that differ mainly in terms of proliferation, which is low in group A and high in group B.
- HER2 cancers are defined by overexpression of the HER2 receptor, usually due to *HER2* gene amplification, and respond well to HER2 inhibitors.
- TNBCs lack ER and HER2 expression, are often associated with defects in DNA repair or genomic stability (e.g., due to silencing of *BRCA1* or *TP53* mutation), and carry a relatively poor prognosis.

Types of Breast Carcinoma

Almost all breast malignancies are adenocarcinomas. The terms *ductal* and *lobular* are still used to describe subsets of both in situ and invasive carcinomas, but most evidence suggests all breast carcinomas arise from cells in the terminal duct lobular unit. Carcinoma in situ was originally classified as ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS) based on the resemblance of the involved spaces to normal ducts or lobules, but it is now recognized that these growth patterns are not related to the cell of origin, but rather reflect differences in tumor cell genetics and biology. By current convention, "lobular" refers to invasive carcinomas that are biologically related to LCIS, and "ductal" is used more generally for adenocarcinomas that cannot be classified as a special histologic type.

Carcinoma in Situ

Carcinoma in situ (literally, "carcinoma in its original place") refers to cancer cells confined within ducts and lobules by a basement membrane. This type of cancer has no capacity to metastasize, as the location precludes access to blood vessels and lymphatics. Because many, if not all, of the molecular changes found in invasive carcinomas also are found in these lesions, carcinoma in situ is an apt description.

Ductal Carcinoma in Situ. DCIS is a clonal proliferation of epithelial cells limited to ducts and lobules by the basement membrane. Myoepithelial cells are preserved in involved ducts/lobules, although they may be diminished in number. DCIS can spread throughout the ductal system and produce extensive lesions involving an entire sector of a breast.

DCIS is almost always detected by mammography. Without mammography, fewer than 5% of carcinomas detected are in situ lesions, but this rises to 15% to 30% in screened populations. Most are identified as a result of calcifications associated with secretory material or necrosis; less commonly, periductal fibrosis surrounding DCIS results in a mammographic density or creates a vaguely palpable mass. Rarely, DCIS (often of micropapillary or papillary types) produces a nipple discharge or is detected as an incidental finding upon biopsy for another lesion.

MORPHOLOGY

DCIS grows in several architectural patterns, which vary in nuclear grade and the presence and extent of necrosis (Fig. 23.17). Some cases of DCIS have a single growth pattern, but most are comprised of a mixture of patterns. **Comedo DCIS** may occasionally produce a vague nodularity, but more often is detected as clustered or linear and branching areas of calcification (Fig. 23.17A). It is defined by two features: (1) tumor cells with pleomorphic, high-grade nuclei and (2) areas of central necrosis (Fig. 23.17B). **Cribriform DCIS** has rounded (cookie cutter–like) spaces, often filled with calcified secretory material (Fig. 23.17C). **Micropapillary DCIS** produces complex bulbous protrusions without fibrovascular cores (Fig. 23.17D). **Papillary DCIS** produces true papillae with fibrovascular cores that lack a myoepithelial cell layer. Varying degrees of necrosis can be associated with each architectural pattern as well as calcifications, which develop in association with intraluminal secretions or necrosis.

Paget disease of the nipple is a rare manifestation of breast cancer (1% to 4% of cases) that presents as a unilateral erythematous eruption with a scale crust. Pruritus is common, and the lesion may be mistaken for eczema. Malignant cells (Paget cells) extend from DCIS within the ductal system via the lactiferous

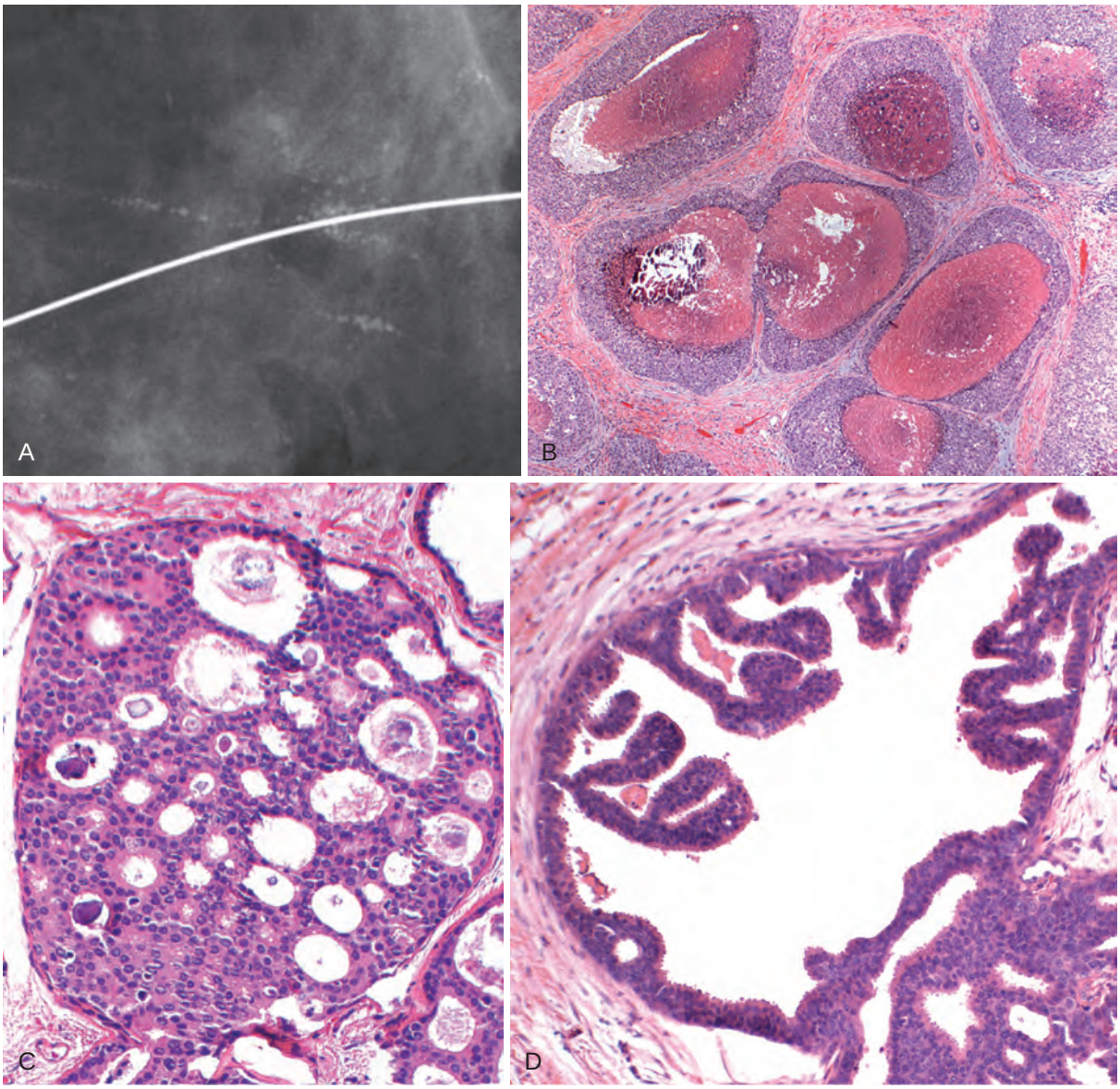


Figure 23.17 Ductal carcinoma in situ (DCIS). (A and B) Comedo type. (A) Specimen radiograph reveals linear and branching calcifications within the ductal system. (B) High-grade proliferation associated with large central zones of necrosis and calcifications fills several ducts. (C) Cribriform DCIS. Note the round, regular ("cookie cutter") spaces containing calcifying secretory material. (D) Micropapillary DCIS. The papillary projections lack fibrovascular cores.

sinuses into nipple skin without crossing the basement membrane (Fig. 23.18). The tumor cells disrupt the normal epithelial barrier, allowing extracellular fluid to seep out onto the nipple surface. Paget cells are readily detected by nipple biopsy or cytologic preparations of the exudate.

A palpable mass is present in 50% to 60% of women with Paget disease, and almost all of these masses prove to be invasive carcinoma. The carcinomas are usually ER-negative and overexpress HER2. In contrast, the majority of women without a palpable mass have only DCIS. Prognosis depends on the features of the underlying carcinoma and is not affected by the presence or absence of Paget disease.

Clinical Features

The natural history of DCIS has been difficult to determine because in the past all women were treated with mastectomy, and the current practice of surgical excision, usually followed by radiation, is largely curative. If untreated, women with small, low-grade DCIS develop invasive cancer at a rate of about 1% per year. When invasive cancer develops in the same breast quadrant, it tends to have a similar grade and expression pattern of ER and HER2 as the associated DCIS. Patients with high-grade or extensive DCIS are believed to have a higher risk for progression to invasive carcinoma.

Remarkably, the overall death rate for women with DCIS is lower than that for women in the population as a whole,

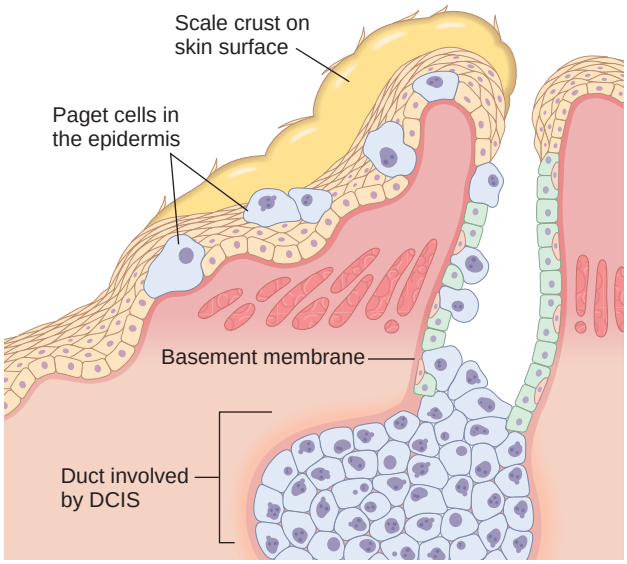


Figure 23.18 Paget disease of the nipple. Ductal carcinoma in situ *(DCIS)* arising within the ductal system of the breast can extend up the lactiferous ducts and into the skin of the nipple without crossing the basement membrane. The malignant cells disrupt the normally tight squamous epithelial cell barrier, allowing extracellular fluid to seep out and form an oozing scaly crust.

possibly because mammographic screening is a "marker" for better access to medical care or other socioeconomic factors that are associated with longevity. Death from metastatic breast cancer after a diagnosis of DCIS occurs in 1% to 3% of women. The origin of metastatic disease may be from a subsequent invasive carcinoma in the ipsilateral or contralateral breast or occult foci of invasion that were not detected at the time of DCIS diagnosis.

Mastectomy is curative in greater than 95% of women. Breast conservation is appropriate for most women but has a slightly higher risk of recurrence—about half of which are DCIS and half invasive carcinoma. The major risk factors for recurrence are (1) high nuclear grade and necrosis, (2) extent of disease, and (3) positive surgical margins. Ensuring complete excision of DCIS is not straightforward, since its distribution in the breast is not reliably predicted by imaging, and it is usually not grossly evident at surgery. Postoperative radiation therapy and tamoxifen also reduce the risk of recurrence; such treatments are reserved for cases that are deemed to be at particularly high risk of recurrence. Trials to identify patients who can safely undergo observation rather than treatment are ongoing.

Lobular Carcinoma in Situ. LCIS is a clonal proliferation of cells within ducts and lobules that grow in a dyscohesive fashion. It is almost always an incidental biopsy finding, since it is rarely associated with calcifications or stromal reactions that produce mammographic densities. As a result, the incidence of LCIS (1% to 6% of all carcinomas) has been unchanged by the introduction of mammographic screening. When both breasts are biopsied, LCIS is bilateral in 20% to 40% of cases compared with 10% to 20% of cases of DCIS.

The cells of atypical lobular hyperplasia, LCIS, and invasive lobular carcinoma are morphologically identical. The observed loss of cellular adhesion is usually due to dysfunction of E-cadherin, a transmembrane protein that contributes to the cohesion of normal epithelial cells in the breast and other glandular tissues. E-cadherin functions as a tumor suppressor protein in such tissues and may be lost in neoplastic proliferations through a variety of mechanisms, including mutation of the E-cadherin gene *(CDH1)*. In rare cases, there is dysregulation of other proteins, such as catenins, that are also needed for E-cadherin–mediated cellular cohesion. LCIS associated with invasive carcinoma shares the same mutations and, thus, in some cases is a true precursor lesion.

MORPHOLOGY

Classically, LCIS consists of a uniform population of cells with oval or round nuclei and small nucleoli involving ducts and lobules (Fig. 23.19A). Mucin-positive signet ring cells are commonly present. The lack of E-cadherin results in a rounded shape without attachment to adjacent cells (Fig. 23.19B). Pagetoid spread, defined by the presence of neoplastic cells between the basement membrane and overlying luminal cells, is commonly seen in breast ducts, but Paget disease (involvement of nipple skin) does not occur. Necrosis and secretory activity are not seen, and thus calcifications are absent. LCIS almost always expresses ER and PR and is HER2-negative.

LCIS is a risk factor for developing invasive carcinoma in either breast, with a slightly higher risk to the ipsilateral breast. Invasive carcinoma develops at a rate of about 1% per year, similar to that observed for untreated DCIS. However, unlike DCIS, it is unclear if surgical removal of the identified lesion lowers risk. Invasive carcinomas developing in women after LCIS are three-fold more likely to be lobular carcinoma; however, most are of other morphologies. Treatment choices include bilateral prophylactic mastectomy, tamoxifen, or, more typically, close clinical follow-up and mammographic screening.

Invasive (Infiltrating) Carcinoma

Breast carcinoma has a wide variety of morphologic appearances. About one-third can be classified into special histologic types that merit discussion because they have important biologic and clinical associations. We will first cover infiltrating carcinomas of "no special type" (typical ductal carcinomas) and will then discuss those that fall into special categories.

MORPHOLOGY

The majority of invasive breast cancers are ductal adenocarcinomas that are not classified further into a special type. In the absence of mammographic screening, these carcinomas usually present as a mass of at least 2 to 3 cm in size. The mammographic and gross appearance varies widely depending on the stromal reaction to the tumor (Fig. 23.20). They most commonly present as a hard, irregular radiodense mass (Fig. 23.20A, B) associated with a desmoplastic stromal reaction (Fig. 23.20C). When cut or scraped, such tumors typically produce a characteristic grating sound (similar to cutting a water chestnut) due to small, central pinpoint foci or streaks of chalky-white desmoplastic stroma and

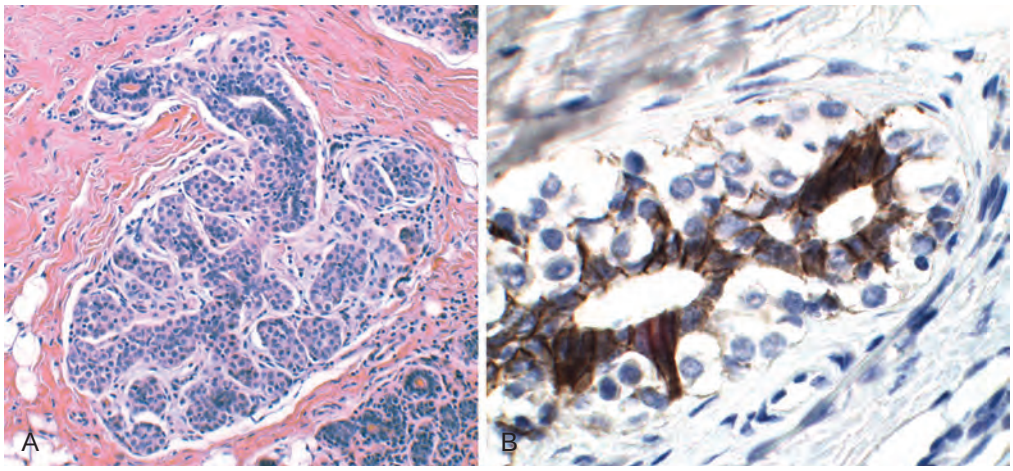


Figure 23.19 Lobular carcinoma in situ (LCIS). (A) Monomorphic population of small, rounded, loosely cohesive cells fills and expands the acini of a lobule. The underlying lobular architecture can still be recognized. The cells extend into the adjacent duct by pagetoid spread. (B) Immunoperoxidase study shows E-cadherin–positive normal luminal cells that have been undermined by E-cadherin–negative LCIS cells spreading along the basement membrane.

occasional foci of calcification. Less commonly, tumors present as deceptively well-circumscribed (Fig. 23.20D, E) masses composed of sheets of tumor cells with scant stromal reaction (Fig. 23.20F) or may be almost imperceptible (Fig. 23.20G, H), being comprised of scattered neoplastic glands or single tumor cells infiltrating otherwise unremarkable fibrofatty tissue (Fig. 23.20I).

Larger carcinomas may invade the pectoralis muscle and become fixed to the chest wall or invade the dermis and cause retraction (dimpling) of the skin. When the tumor involves the central portion of the breast, retraction of the nipple may develop. Rarely, breast cancer presents as metastasis to an axillary lymph node or a distant site before cancer is detected in the breast. In such cases, the primary carcinoma may be small, may be obscured by dense breast tissue, or may fail to produce a desmoplastic response. In most cases, these "occult" primary tumors (which are easily missed by palpation or mammography) can be detected by imaging studies using ultrasound or MRI.

Invasive carcinoma is graded using the Nottingham Histologic Score. Carcinomas are scored for tubule formation, nuclear pleomorphism, and mitotic rate. Grade 1 (well differentiated) carcinomas grow in a tubular or cribriform pattern, have small uniform nuclei, and have a low proliferative rate (Fig. 23.21A). Grade 2 (moderately differentiated) carcinomas have areas where cells grow as solid clusters or single infiltrating cells and show greater nuclear pleomorphism and high numbers of mitotic figures (Fig. 23.21B). Grade 3 (poorly differentiated) carcinomas invade as ragged nests or solid sheets of cells and have enlarged irregular nuclei. A high proliferative rate and areas of tumor necrosis are common in high-grade tumors (Fig. 23.21C).

Special Histologic Types of Invasive Carcinoma

As with all breast cancers, these special tumors can be organized into molecular groups based on expression of ER and HER2, which carry their usual therapeutic implications. However, special histologic types of breast cancer often harbor unique genetic aberrations, sometimes have distinct gene signatures, and frequently show associations with clinical behavior and prognosis that break the established "rules" for ductal carcinomas of no special type. Although relatively uncommon, study of these tumors has also provided important insights into breast cancer pathogenesis.

Lobular carcinoma is the subtype with the clearest association of phenotype and genotype. Like LCIS, most cases show biallelic loss of expression of *CDH1,* the gene that encodes E-cadherin. Lobular carcinomas are dyscohesive, typically infiltrate as single cells, and sometimes fail to produce a desmoplastic response, making it difficult to detect these cancers by palpation and imaging. They also have distinctive patterns of metastatic spread, often involving the peritoneum and retroperitoneum, the leptomeninges (carcinomatous meningitis), the gastrointestinal tract, and the ovaries and uterus. Males and females with heterozygous germline mutations in *CDH1* are at increased risk for developing lobular carcinoma and have a greatly increased risk for signet ring carcinoma of the stomach (Chapter 17).

Carcinomas with medullary pattern are of interest due to the finding that over half of *BRCA1*-associated carcinomas have this appearance (see Table 23.3). Although the majority of carcinomas with medullary pattern are not associated with germline *BRCA1* mutations, hypermethylation of the *BRCA1* promoter leading to downregulation of *BRCA1* expression is observed in 67% of these tumors. Of interest, this subtype has a better prognosis than other poorly differentiated carcinomas. Notably, these tumors also have unusually large number of infiltrating T lymphocytes, suggesting that improved outcomes may be related to a host immune response to tumor antigens.

Many other special histologic types of breast cancer (too numerous to list) have been described. There is much that remains to be learned about the biology and pathogenesis of these tumors, some of which are described below.

MORPHOLOGY

Some special histologic types of cancer almost always fall within the luminal (ER-positive/HER2-negative) group. These include lobular carcinoma, mucinous carcinoma, tubular carcinoma, and papillary carcinoma. **Lobular carcinoma** often (as already mentioned) insidiously infiltrates the breast while producing minimal

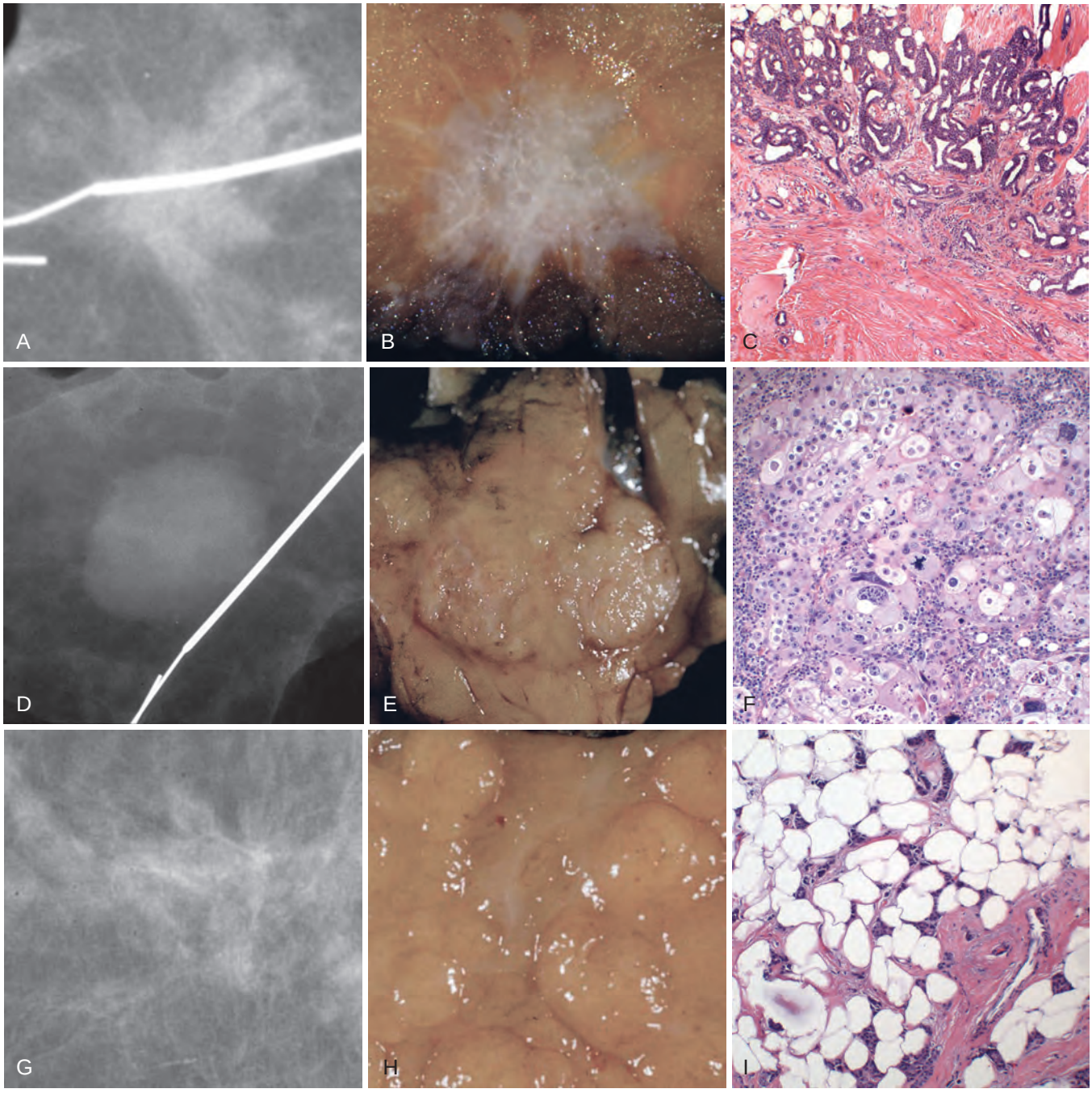


Figure 23.20 Invasive carcinoma of no special type. The majority of invasive carcinomas have a haphazard pattern of stromal invasion that produces masses with irregular margins on imaging (A) and gross examination (B). Microscopically, such tumors are marked by an exuberant desmoplastic stromal response (C). A subset of carcinomas grow as masses that appear to be well circumscribed or lobulated on imaging (D) and gross inspection (E). Microscopically, such cancers typically take on the appearance of expansile masses of cells with pushing borders; stromal response is often limited to a narrow zone of fibrosis at the tumor margin (F). Rarely, invasive cancers produce little or no stromal response. Such cancers may show only subtle architectural distortion on mammography (G) and may not produce palpable masses or be identifiable grossly (H). Microscopically, tumor cells are found scattered within normal-appearing fibroadipose tissue (I). (B, Courtesy Dr. David Hicks, University of Rochester Medical Center, Rochester, NY.)

desmoplasia. The histologic hallmark is the presence of dyscohesive infiltrating tumor cells, often including signet ring cells containing intracytoplasmic mucin droplets (Fig. 23.22A). Tubule formation is absent. **Mucinous (colloid) carcinoma** is soft or rubbery and has the appearance and consistency of pale gray-blue gelatin. The borders are pushing or circumscribed. The tumor cells are arranged in clusters and small islands of cells within large lakes of mucin (Fig. 23.22B). **Tubular carcinoma** consists exclusively of well-formed tubules and is sometimes mistaken for a benign sclerosing lesion (Fig. 23.22C). A cribriform pattern may also be present. Apocrine snouts are typical, and calcifications may be present within the lumens. **Papillary carcinoma,** as the name

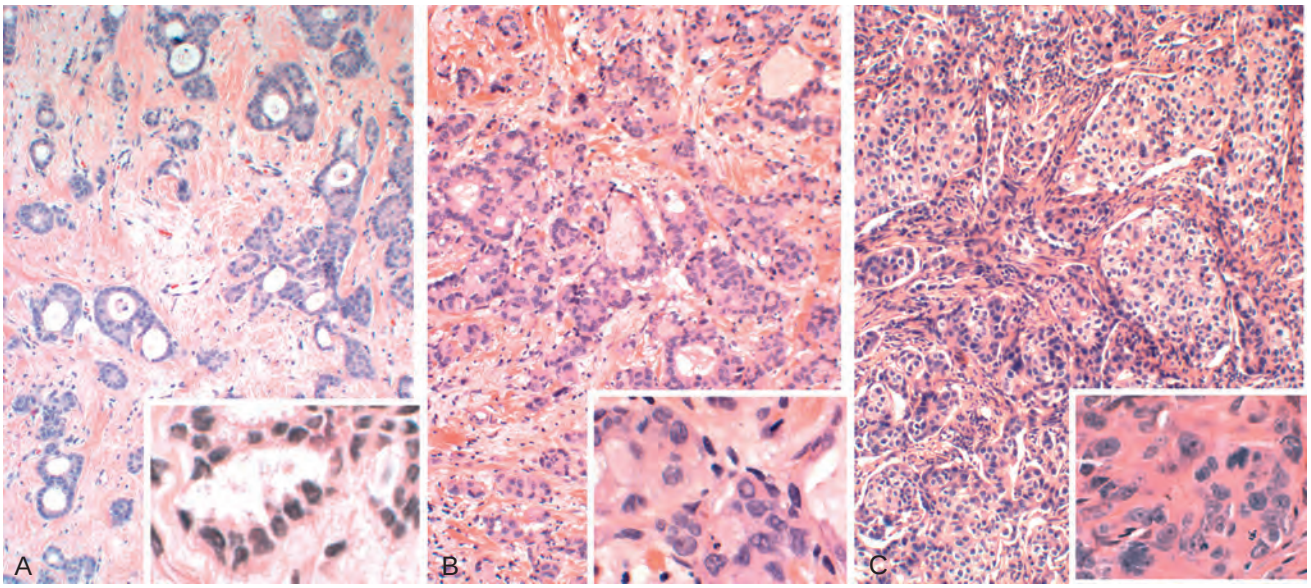


Figure 23.21 Invasive carcinomas are separated into three grades based on tubule formation, nuclear pleomorphism, and number of mitoses. (A) This example of a grade 1 well-differentiated carcinoma shows frequent tubules formed of cells with small monomorphic nuclei. Only rare mitoses are present. (B) In contrast, this grade 2 moderately differentiated carcinoma shows less tubule formation and some solid nests of cells with pleomorphic nuclei. Occasional mitotic figures are seen. (C) Grade 3 poorly differentiated carcinomas infiltrate as ragged sheets of cells with enlarged pleomorphic nuclei.

implies, produces true papillae, fronds of fibrovascular tissue lined by tumor cells (Fig. 23.22D).

Two special histologic types frequently overexpress HER2. The tumor cells of **apocrine carcinoma** resemble the cells that line sweat glands. These cells have enlarged round nuclei with prominent nucleoli and abundant eosinophilic, occasionally granular, cytoplasm (Fig. 23.22E). **Micropapillary carcinoma** (a misnomer) forms hollow balls of cells that float within intercellular fluid, creating structures that mimic the appearance of true papillae (Fig. 23.22F).

TNBC (ER-negative, HER2-negative) often corresponds to one of several special histologic types. Chief among these is **carcinoma with medullary pattern.** These carcinomas are softer than other carcinomas (*medulla* is Latin for "marrow") due to minimal desmoplasia and often form well-circumscribed masses. Histologic features include (1) solid sheets of large cells with pleomorphic nuclei and prominent nucleoli, (2) frequent mitotic figures, (3) a moderate to marked lymphoplasmacytic infiltrate surrounding and within the tumor, and (4) a pushing (noninfiltrative) border (Fig. 23.22G). DCIS is minimal or not seen. **Metaplastic carcinoma** includes spindle cell carcinomas and matrix-producing carcinomas. These carcinomas often have gene expression profiles resembling those of myoepithelial cells.

Rare special histologic types of TNBC have a favorable prognosis compared to other carcinomas in this molecular group and include secretory carcinoma, low-grade adenosquamous carcinoma, and adenoid cystic carcinoma. **Secretory carcinoma** mimics lactating breast by forming dilated spaces filled with eosinophilic material (Fig. 23.22H). These carcinomas rarely metastasize.

Another special subtype that merits mention is **inflammatory carcinoma.** This form of carcinoma has a characteristic gross appearance caused by extensive plugging of the lymphovascular spaces of the dermis with carcinoma cells and carries a very poor prognosis, as most patients prove to have distant metastases. It presents as breast erythema, swelling, and skin thickening. The edematous skin is tethered to the breast by Cooper ligaments and mimics the surface of an orange peel, an appearance referred to as *peau d'orange*. The name "inflammatory" is a misnomer, as typically no inflammation is present. The underlying carcinoma is usually diffusely infiltrative and typically does not form a discrete palpable mass. The presentation can be confused with a breast infection, leading to delayed diagnosis. These tumors are usually of high grade but do not belong to any particular molecular subtype.

Prognostic and Predictive Factors for Invasive Carcinoma

The outcome for women with breast cancer depends on the biologic features of the carcinoma (the molecular or histologic type) and the stage of disease at the time of diagnosis. Based on these factors, some women with breast cancer have a normal life expectancy, whereas others have only a 10% chance of being alive in 5 years. Patients who present with distant metastasis (5% of patients) or with inflammatory carcinoma (1% to 5% of patients) have a particularly poor prognosis. For other cancers, prognosis is determined by pathologic evaluation of the primary tumor and the axillary lymph nodes (Table 23.5).

Prognostic factors are important in counseling patients about the likely outcome of their disease, choosing the most appropriate treatment, and designing clinical trials. Predictive factors help to determine the likely response of a cancer to a specific type of treatment. Factors related to tumor biology are usually both prognostic and predictive (e.g., ER and HER2 expression, proliferation), whereas factors related to tumor extent (e.g., tumor size, lymph node metastases, distant metastases) are primarily prognostic.

The major prognostic factors are as follows (Table 23.5):

- *Lymph node metastases.* Axillary lymph node status is the most important prognostic factor for invasive carcinoma

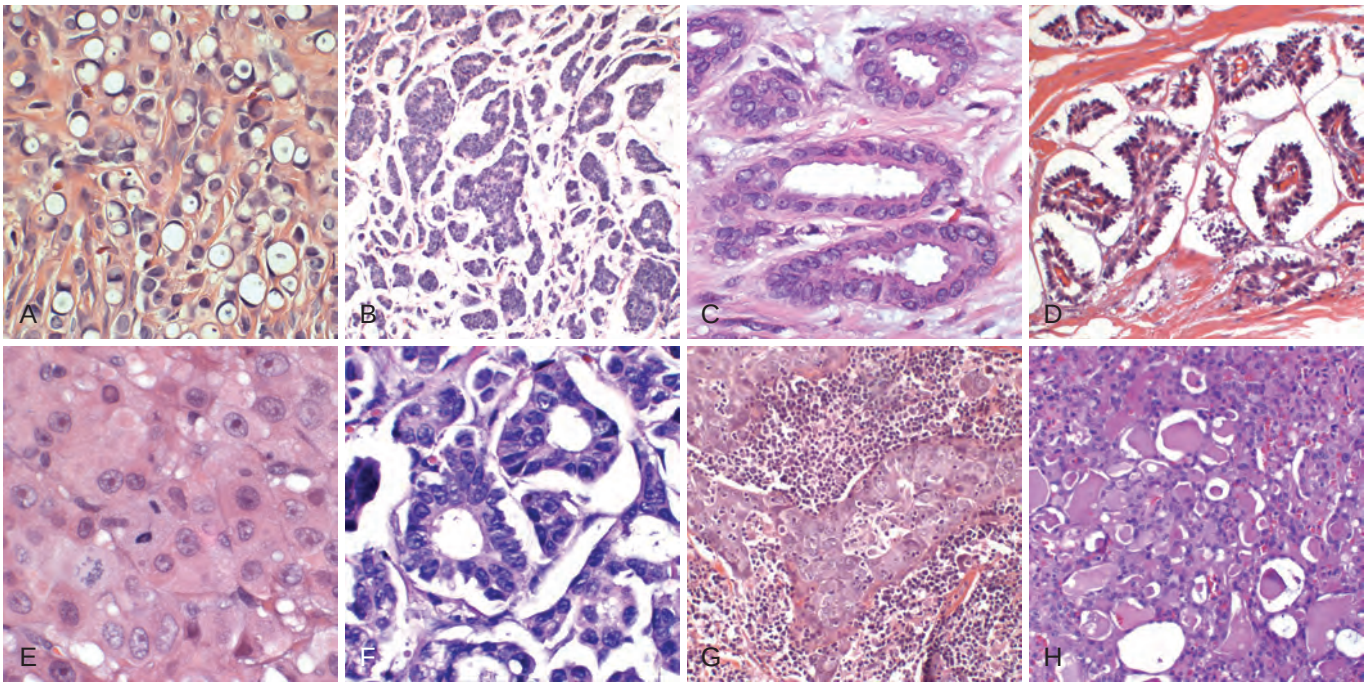


Figure 23.22 Special histologic types of invasive carcinoma. (A) Lobular carcinoma. (B) Mucinous carcinoma. (C) Tubular carcinoma. (D) Papillary carcinoma. (E) Apocrine carcinoma. (F) Micropapillary carcinoma. (G) Carcinoma with medullary pattern. (H) Secretory carcinoma. See text for morphologic descriptions.

Table 23.5 Prognostic Factors for Invasive Breast Carcinoma

Prognostic Factors	Comments
Elements of AJCC 8th Edition Staging	
Distant metastasis (M)	Metastasis beyond regional lymph nodes is the most important prognostic factor.
Regional lymph nodes (N)	Nodal metastasis (including the number of involved nodes) is the second most important prognostic factor.
Tumor (T)	Size, involvement of skin (e.g., ulceration or dermal metastases), invasion into chest wall, and presentation as inflammatory carcinoma are important features.
Histologic grade	Survival diminishes with higher histologic grade.
Expression of ER, PR, and HER2	Survival is highest for the most favorable combination (high ER and PR and absent HER2) and is lowest for the least favorable combination (absent ER, PR, and HER2).
Other Prognostic Factors	
Lymphovascular invasion	Tumor cells seen in vascular spaces at the periphery of carcinomas are a poor prognostic factor.
Special histologic types	Some histologic types of cancer are strongly correlated with very favorable survival (e.g., tubular, adenoid cystic).
Response to chemotherapy	The degree of response is a strong prognostic factor for TNBC and HER2 cancers, but not the majority of luminal cancers.
Gene expression profiling	The most important clinical value of these assays is to identify patients with antiestrogen-responsive cancers who do not need chemotherapy.

AJCC, American Joint Committee on Cancer; *ER*, estrogen receptor; *PR*, progesterone receptor; *TNBC*, triple negative breast cancer.

in the absence of distant metastases. The clinical assessment of lymph node status is unreliable due to both false-positives (e.g., palpable reactive nodes) and false-negatives (e.g., lymph nodes with small metastatic deposits). Therefore, biopsy is necessary for accurate assessment. With no nodal involvement, the 10-year disease-free survival rate is 70% to 80%; the rate falls to 35% to 40% with one to three positive nodes and to 10% to 15% when more than 10 nodes are positive. It is important to note that the presence of nodal metastases is correlated with the probability of distant metastasis, and that removal of involved lymph nodes does not lower the risk of future metastatic disease.

Lymphatic vessels in most breast carcinomas drain first to one or two *sentinel nodes*, which can be identified with radiotracer or colored dyes. If a biopsy of the sentinel nodes is negative for metastasis, it is unlikely that other, more distant nodes will be involved; this approach is used to spare patients the morbidity of a complete axillary dissection. Approximately 10% to 20% of women without axillary lymph node metastases experience recurrence with distant metastasis. In these patients, metastasis may occur via the internal mammary lymph nodes or hematogenously.

- *Distant metastases.* Once distant metastases are present, cure is unlikely, although long-term remissions and palliation can be achieved, especially in women with ER-positive tumors. As discussed earlier, the molecular group influences the timing and location of metastases (see Table 23.4).
- *Tumor size.* The risk of axillary lymph node metastasis increases with the size of the primary tumor, but size also is an independent prognostic factor. Women with node-negative carcinomas less than 1 cm in size have a

10-year survival rate of more than 90%, whereas survival drops to 77% for cancers greater than 2 cm.

- *Locally advanced disease.* Carcinomas invading the skin or skeletal muscle are usually large, making it difficult to achieve local control. Such cases are now rare in the United States but continue to be common in countries with limited access to care.
- *Lymphovascular invasion.* Tumor cells are present within vascular spaces (either lymphatics or small capillaries) in about half of all invasive carcinomas. This finding is strongly associated with the presence of lymph node metastases. It is a poor prognostic factor for overall survival in women without lymph node metastases and a risk factor for local recurrence.
- *Inflammatory carcinoma.* As already discussed, this subtype is characterized by extensive invasion of dermal lymphovascular channels and portends a poor outcome. The 3-year survival rate is only 3% to 10%. Only 1% to 5% of cancers are in this group, but the incidence is higher in women of African descent and younger women.

Additional prognostic factors are related to tumor biology (see Table 23.5). Molecular subtype, histologic grade (including proliferation), and special histologic types have already been discussed. A few other prognostic factors are clinically useful and merit brief mention.

- *Gene expression profiling.* A number of proprietary assays that quantify mRNA levels in breast cancer cells have been developed. Most are heavily weighted toward inclusion of genes that are involved in proliferation. The greatest clinical value of these assays is to identify patients with slowly growing, antiestrogen-responsive cancers who can be spared the toxicity of chemotherapy.
- *Response to neoadjuvant chemotherapy.* Treating patients prior to surgery provides the opportunity to observe the tumor response to chemotherapy. A third or more of TNBCs and HER2 cancers regress completely (termed a pathologic complete response). Patients with responsive cancers belonging to these subtypes have a better prognosis than patients with nonresponsive cancers. In contrast, very few luminal cancers respond completely to chemotherapy. However, these tumors are typically slow growing and can often be controlled with endocrine-based treatment for many years.

The most important prognostic factors are included in the latest American Joint Committee on Cancer (AJCC) staging system, which for the first time includes both an anatomic staging system (Table 23.6) and a prognostic staging system that integrates the anatomic stage and the molecular characteristics of individual breast cancers. The anatomic stage (0 to IV) considers the nature of the disease in the breast (T, for tumor), the involvement of regional lymph nodes (N, for nodes), and the presence of distant metastases (M, for metastases). The prognostic stage then takes into consideration biologic features that portend a more or less favorable outcome. For example, in some cases TNBC is "up-staged" in the new prognostic staging system to reflect its more aggressive behavior.

The main goals of breast cancer therapy are to control local disease and to prolong survival by treating known or potential distant metastases. Local control is achieved in the majority of patients with breast-conserving surgery and radiation therapy. Mastectomy is generally only necessary for locally advanced disease or for women at high risk of a second primary cancer who wish to reduce the risk of recurrence. In the absence of adequate control, some breast cancers can advance to *carcinoma en cuirasse* (literally "carcinoma of the breastplate"), a dreaded complication of infiltration of the skin and ulceration. Fortunately, this is now rarely seen in the United States, but it remains a common presentation for women living in areas with limited resources.

Systemic therapy is used to treat known or likely distant disease and also reduces the likelihood of local recurrence. The first effective systemic treatment for any cancer was the discovery that oophorectomy caused regression of breast cancer in the late 1800s. This remains a treatment modality, but there are now many other options to inhibit the growth of hormonally responsive cancers (Table 23.7). For many luminal cancers, endocrine therapy is the best and most effective therapeutic option. Chemotherapy is used for highly proliferative carcinomas, regardless of molecular subtype. For HER2 cancers, targeted therapy with HER2 antagonists has markedly improved prognosis. TNBC remains a therapeutic

Table 23.6 American Joint Committee on Cancer 8th Edition: Anatomic Stage[a]

Stage[b]	T: Primary Cancer (Tumor)	N: Lymph Nodes	M: Distant Metastasis	10-Year Survival (%)
0	Ductal carcinoma in situ	No metastases	Absent	97
I	Invasive carcinoma ≤2 cm	No metastases or only micrometastases	Absent	87
II	Invasive carcinoma >2 cm Invasive carcinoma >5 cm but ≤5 cm	1–3 positive LNs 0–3 positive LNs	Absent Absent	65
III	Invasive carcinoma >5 cm Any size invasive carcinoma Invasive carcinoma with skin or chest wall involvement or inflammatory carcinoma	Negative or positive LNs ≥4 positive LNs Negative or positive LNs	Absent Absent Absent	40
IV	Any size invasive carcinoma	Negative or positive LNs	Present	5

[a]In the 8th edition, prognostic stages are assigned using T, N, M, grade, ER, PR, and HER2. Pathologic prognostic stage is assigned for patients who undergo surgical excision prior to other treatment. A multigene assay, when available, can be used to assign stage in this setting. Clinical prognostic stage is assigned for all other patients including patients prior to surgery, patients not eligible for surgery, and patients undergoing systemic therapy prior to surgery.
[b]The anatomic stages listed are used only when information on grade, ER, PR, and HER2 are not available. The survival estimates include the average survival for patients with all biologic types of cancer.

ER, Estrogen receptor; *LNs,* lymph nodes; *PR,* progesterone receptor.

Table 23.7 Targeted Treatment of Breast Cancer

Target	Treatment	Companion Assay	Comments
ER	Estrogen deprivation (oophorectomy, aromatase inhibitors) Blockage of ER (tamoxifen) Degradation of ER (fulvestrant)	IHC for nuclear ER	Effective cytostatic (but not cytotoxic) therapy for ER-positive cancer
Cyclin-dependent kinases 4 and 6 (CDK4/6)	Kinase inhibitors (palbociclib, abemaciclib, ribociclib)		Used for ER-positive cancers, usually in conjunction with an aromatase inhibitor
HER2	Antibodies to HER2 Cytotoxic therapy linked to HER2 antibody Tyrosine kinase inhibitors Vaccines	IHC for membrane HER2 ISH for HER2 amplification DNA sequencing for *HER2* mutations	Effective for HER2-positive cancer
Defects in HRR[a]	Chemotherapy with agents causing DNA damage requiring HRR (e.g., platinum agents) Inhibition of alternative DNA repair pathway PARP inhibitors)	DNA sequencing to identify *BRCA1* and *BRCA2* mutations	May be effective in carcinomas with germline *BRCA1* or *BRCA2* mutations or carcinomas with somatic loss of BRCA function
P13K/AKT/mTOR pathway	Inhibition of proteins in the pathway	Activating mutations or pathway activation—ability to predict response under investigation	>80% of breast cancers have alterations in this pathway
Immune checkpoint proteins	Blocking antibodies to PD-L1, PD-I, and other immune checkpoint proteins such as TIM-1 and LAG-3	IHC for immune checkpoint proteins—ability to predict response under investigation	Under investigation for high-grade ER-negative carcinomas

[a]Mutations in *BRCA1* and *BRCA2* cause defects in HRR.

ER, Estrogen receptor; *HRR*, homologous recombination repair; *IHC*, immunohistochemistry; *ISH*, in situ hybridization; *PARP*, poly-ADP ribose polymerase

challenge. There is hope that the genetic instability of these cancers will make them susceptible to agents that inhibit DNA repair and to immune system–based therapies (see Table 23.7).

KEY CONCEPTS

TYPES OF CARCINOMA AND PROGNOSTIC FACTORS

- DCIS is treated locally, as subsequent invasive carcinomas usually occur at the same site, whereas LCIS confers bilateral risk.
- Special histologic types of carcinomas have prognostic importance and provide additional clues linking biologic changes to clinical behavior.
- Prognosis is dependent on both biologic features and the extent of cancer at the time of diagnosis (anatomic stage).
- In the AJCC 8th edition staging system, molecular group and anatomic stage are combined to create prognostic stage groups that provide better estimates of likely survival.
- Effective treatment requires both local and systemic control of disease.
- Improvements in treatment are being made as new targeted therapies are being developed and response to treatment is better understood.

Male Breast Cancer

The incidence of breast cancer in men is only 1% of that in women, which translates to a lifetime risk of 0.11%. There are about 2670 cases and 500 deaths in the United States each year. Risk factors are similar to those in women and include increasing age, first-degree relatives with breast cancer, exposure to exogenous estrogens or ionizing radiation, alcohol consumption, infertility, obesity, prior benign breast disease, Klinefelter syndrome, and residency in Western countries.

The most important familial factor conferring an increased risk for male breast cancer is germline mutation of the *BRCA2* tumor suppressor gene. Approximately 6% of male carriers develop breast cancer. Of men with breast cancer, 4% to 40% have germline *BRCA2* mutations depending on the population tested. A lower risk of male breast cancer is conferred by germline mutations in *BRCA1, PTEN, TP53,* and *PALB2* (see Table 23.3).

More than 90% of breast cancers in males are of luminal type, while TNBCs and HER2 cancers are very rare (<5%). Because breast epithelium in men is limited to large ducts near the nipple, carcinomas usually present as a palpable subareolar mass 2 to 3 cm in size and/or as nipple discharge. Male breast carcinomas are situated close to the skin and thoracic wall, and even small tumors can invade these structures and cause skin ulceration. Dissemination follows the same pattern as in women. Axillary lymph node involvement is present in about half of cases at the time of diagnosis, and distant metastases to the lungs, brain, bone, and liver are common. Although men present with higher stage disease, the prognosis is similar to that of women when matched for stage. Most cancers are treated locally with mastectomy and axillary node dissection. The same systemic treatment guidelines are used for men as for women, and response rates are similar.

STROMAL TUMORS

The two types of stroma in the breast, intralobular and interlobular, give rise to distinct types of neoplasms. Two closely related, breast-specific tumors, fibroadenoma and phyllodes tumor, arise from cells of the intralobular stroma.

These tumors are termed "biphasic" because they also include a non-neoplastic epithelial component, the proliferation of which may be stimulated by growth factors elaborated by the stromal cells. Both fibroadenoma and phyllodes tumor are driven by somatic mutations in *MED12*, a component of a multiple protein complex called mediator that links RNA polymerase II to specific DNA-binding transcription factors. It is no doubt not coincidental that the other tumor that is strongly associated with *MED12* mutations, uterine leiomyoma, also arises from stromal cells within an organ that is responsive to female sex hormones. Perhaps, by deranging mediator function, *MED12* mutations alter the expression of sex hormone–regulated genes that control the proliferation and survival of certain types of stromal cells. In contrast, interlobular stroma is the source of the same types of tumors found in connective tissue in other sites of the body (e.g., lipomas and angiosarcomas), as well as tumors arising more commonly in the breast (e.g., myofibroblastoma and fibrous tumors), and consist only of stromal cells.

Fibroadenoma

Fibroadenoma is the most common benign tumor of the female breast. Two-thirds of fibroadenomas harbor driver mutations in *MED12*. The pathogenesis of the remainder is uncertain.

MORPHOLOGY

Fibroadenomas vary in size from less than 1 cm to large tumors that replace most of the breast. They usually present as a palpable mass in young women and as a mammographic density (Fig. 23.23A) or clustered calcifications in older women. The tumors are well-circumscribed, rubbery, grayish white nodules that bulge above the surrounding tissue and often contain slit-like spaces lined by epithelium (Fig. 23.23B). The delicate and often myxoid stroma resembles normal intralobular stroma. The epithelium may be surrounded by stroma (pericanalicular pattern) or compressed and distorted by it (intracanalicular pattern) (Fig. 23.23C). In older women, the stroma typically becomes densely hyalinized and the epithelium atrophic.

Clinical Features

Most fibroadenomas occur in women in their 20s and 30s, and they are frequently multiple and bilateral. These tumors are hormonally responsive and may grow in size during pregnancy and regress after menopause. Rapid growth and infarction during pregnancy may raise a false suspicion of carcinoma. Curiously, almost half of women receiving cyclosporin A after renal transplantation develop multiple, bilateral fibroadenomas that regress after cessation of treatment. Whether these lesions are true neoplasms or reactive hyperplasias is unclear.

Fibroadenomas are associated with a slightly increased risk of carcinoma, a risk that may be higher if "complex" features are present (cysts larger than 0.3 cm, sclerosing adenosis, epithelial calcifications, or papillary apocrine change) (see Table 23.1). However, these changes are also associated with a higher likelihood of finding other lesions in the surrounding breast tissue (e.g., atypical hyperplasia), and these may be the true drivers of the increased risk.

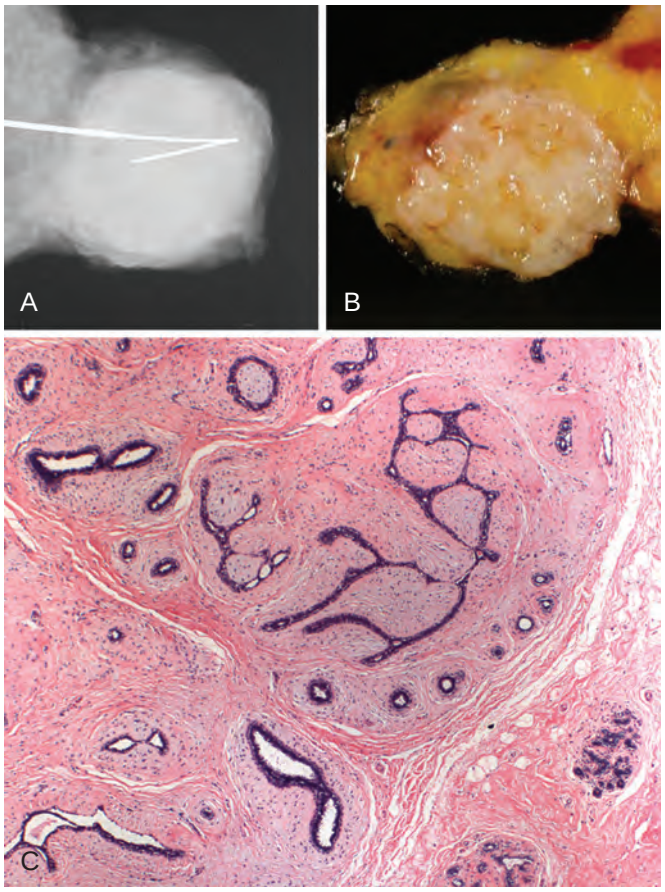


Figure 23.23 Fibroadenoma. (A) Radiograph shows a characteristically well-circumscribed mass. (B) Grossly, a rubbery, white, well-circumscribed mass is clearly demarcated from the surrounding yellow adipose tissue. The absence of adipose tissue accounts for the radiodensity of the lesion. (C) The proliferation of intralobular stroma surrounds, pushes, and distorts the associated epithelium. The border is sharply delimited from the surrounding tissue.

Phyllodes Tumor

Phyllodes tumor, like fibroadenoma, arises from intralobular stroma but is much less common. *Cystosarcoma phyllodes* is a term sometimes used for these lesions, but *phyllodes tumor* is preferred, since most behave in a benign fashion and are not cystic. Like fibroadenomas, the majority of phyllodes tumors have *MED12* mutations. Benign-appearing phyllodes tumors that have only a slight propensity to recur often have *MED12* mutations and few other genetic changes. In contrast, tumors that display malignant behavior are more likely to have mutations in additional genes, such as *TERT*, the gene that encodes telomerase.

MORPHOLOGY

Most phyllodes tumors are detected as palpable masses, while a few are found by mammography. The tumors vary in size from a few centimeters to massive lesions involving the entire breast. The larger lesions often have bulbous protrusions (*phyllodes* is Greek for "leaf-like") due to the presence of nodules of proliferating stroma covered by epithelium (Fig. 23.24). In some tumors these

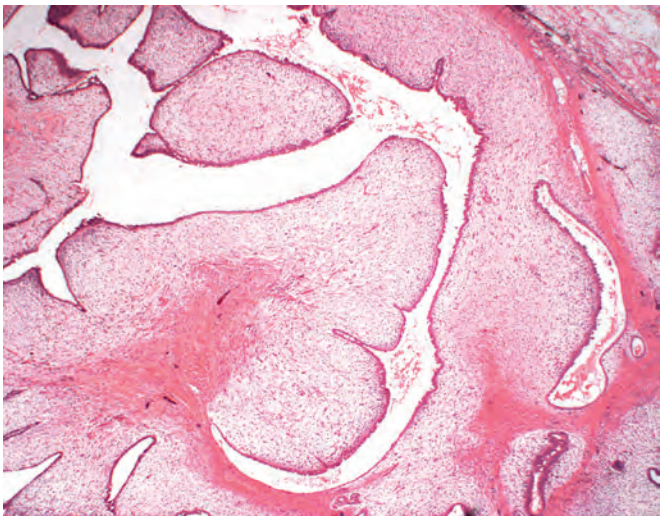


Figure 23.24 Phyllodes tumor. Compared to a fibroadenoma, there is increased stromal cellularity and overgrowth, giving rise to the typical leaf-like architecture.

protrusions extend into a cystic space. This growth pattern also may occasionally be seen in large fibroadenomas and is not an indication of malignancy. Phyllodes tumor is distinguished from fibroadenoma on the basis of higher cellularity, higher mitotic rate, nuclear pleomorphism, stromal overgrowth, and infiltrative borders. Low-grade (benign) lesions resemble fibroadenomas but are more cellular and are mitotically active. High-grade (malignant) lesions may be difficult to distinguish from sarcomas.

Clinical Features

Most phyllodes tumors present in the sixth decade, 10 to 20 years later than the peak age for fibroadenoma. Most have low-grade (benign) cytologic features; these occasionally recur locally but do not metastasize. In contrast, borderline and high-grade (malignant) phyllodes tumors often recur locally unless they are treated with wide excision or mastectomy. Regardless of grade, lymphatic spread is rare, and axillary lymph node dissection is contraindicated. The uncommon high-grade lesions give rise to distant hematogenous metastases in about one-third of cases. Only the stromal component metastasizes.

Lesions of Interlobular Stroma

Tumors of the interlobular stroma of the breast are composed of stromal cells without an accompanying epithelial component. These include benign as well as malignant tumors, all uncommon and hence considered briefly. *Myofibroblastoma* consists of myofibroblasts and is unusual in that it is the only breast tumor that is equally common in males. *Lipomas* are often palpable and can also be detected mammographically as fat-containing lesions. The only importance of these lesions is to distinguish them from malignancies.

Fibromatosis is a clonal proliferation of fibroblasts and myofibroblasts. It presents as an irregular, infiltrating mass that can involve muscle. Though locally aggressive, this lesion does not metastasize. Some cases are associated with prior trauma or surgery. Other cases occur as part of familial adenomatous polyposis, hereditary desmoid syndrome, and Gardner syndrome.

Malignant Tumors of Interlobular Stroma

Malignant stromal tumors of the breast are rare. The only sarcoma that occurs with any frequency in the breast is angiosarcoma—however, it accounts for less than 0.05% of breast malignancies. Angiosarcoma of the breast may be sporadic or arise as a complication of therapy. Most sporadic angiosarcomas occur in the breast parenchyma of young women (mean age 35) and have a poor prognosis. Tumors occurring after treatment for breast cancer typically arise in older women and are associated with radiation therapy or chronic edema. After radiation therapy, approximately 0.3% of women develop angiosarcomas in breast skin, with most cases being diagnosed 5 to 10 years after treatment.

OTHER MALIGNANT TUMORS OF THE BREAST

Malignancies of the breast arising from lymphocytes or skin, or metastatic from another site, comprise less than 5% of breast cancers. *Non-Hodgkin lymphoma* may arise primarily in the breast, or the breasts may be secondarily involved by systemic disease. Most primary breast lymphomas are of B-cell type, while rare T-cell lymphomas may arise in association with breast implants, possibly due to chronic inflammation, which is known to stimulate lymphoma development in other contexts. Young women with Burkitt lymphoma may present with massive bilateral breast involvement, often while pregnant or lactating. Malignant tumors may arise from the skin and dermis of the breast; these tumors are identical to their counterparts found in skin elsewhere (Chapter 25). Metastases to the breast are rare and most commonly arise from melanomas and ovarian cancers.

KEY CONCEPTS

STROMAL TUMORS

- Intralobular stroma is the origin of two biphasic tumors, fibroadenoma and phyllodes tumor.
- Fibroadenomas are the most common benign tumor of the breast.
- Tumors of interlobular stroma consist only of stromal cells and include both benign and malignant lesions.
- Angiosarcoma is the most common stromal malignancy and can either be sporadic or associated with radiation exposure or lymphedema.

SUGGESTED READINGS

The Social, Historical, and Political Significance of Breast Cancer

Mukherjee S: *The Emperor of all maladies: a biography of cancer*, 2011, Scribner. *[This look back at the evolution of our understanding of cancer includes the history of the discovery of HER2 and the remarkable impact of HER2 targeted therapy. It also includes a cautionary look at how clinical trials can go awry, as in the solid tumor autonomous marrow transplant program for breast cancer in the 1990s].*

Olson JS: *Bathsheba's breast: women, cancer, and history*, 2002, The Johns Hopkins University Press. *[From Queen Atossa in Babylonia in 490 BCE to Dr. Jeri Nielsen in Antarctica in 1999, women have diagnosed their own breast cancers for thousands of years. This is a chronicle of breast cancer as told by these women as well as an informative history of the important milestones in cancer treatment].*

Yalom M: *History of the breast*, 1998, Ballantine Books. *[Scholarly work on the cultural, political, psychologic, artistic, physical, and religious significance of the breast].*

General Sources of Information About Breast Cancer

Breast Cancer Facts & Figures: *American Cancer Society*, 2018. Available at https://www.cancer.org. *[Excellent resource for up-to-date information about breast cancer.].*

International Association of Cancer Registries (IACR): *GLOBOCAN*, 2012. Available at http://globocan.iarc.fr. *[The International Association of Cancer Registries provides information about the incidence of cancer and mortality worldwide].*

Large Internet Breast Cancer Datasets

The Cancer Genome Atlas project. https://cancergenome.nih.gov. (Accessed 1 January 2018). *[This is a National Institutes of Health–sponsored initiative that includes information on over 800 breast cancers].*[a]

The International Cancer Genome Consortium. https://icgc.org. (Accessed 1 January 2018). *[This is a voluntary international group of scientific researchers and organizations].*[a]

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC). www.cbioportal.org. (Accessed 1 January 2018). *[The original Molecular Taxonomy of Breast Cancer International Consortium included >2000 breast cancer samples from tumor repositories in the United Kingdom and Canada. This site also provides links to other databases].*[a]

National Cancer Database (NCDB). https://www.facs.org. (Accessed 1 January 2018). *[The National Cancer Database includes information on approximately 70% of newly diagnosed breast cancers in the United States from American College of Surgeons Commission on Cancer–accredited hospitals].*[b]

Surveillance, Epidemiology, and End Results (SEER) program. http://seer.cancer.gov. (Accessed 1 January 2018). *[This is a program of the National Cancer Institute that collects population-based information from cancer registries covering approximately 28% of the U.S. population].*[b]

Breast Cancer Biology

Casasent AK, Schalck A, Gao R et al: Multiclonal invasion in breast tumors identified by topographic single cell sequencing, *Cell* 172:205, 2018. *[The powerful combination of morphology and single-cell sequencing allows the separate evaluation of ductal carcinoma in situ (DCIS) and invasive carcinoma and reveals that multiple different subclones arise in DCIS, invade the stroma, and contribute to the invasive tumor].*

Davalos V, Martinez-Cardus A, Esteller M: The epigenetic revolution in breast cancer: from single-gene to genome-wide next-generation approaches, *Am J Pathol* 187:2163, 2017. *[Epigenetics plays an important role in every aspect of breast cancer – from the determination of molecular type (expression of estrogen receptor), to the activation/inactivation of growth signaling pathways, to the development of resistance to treatment].*

Del Alcazar CRG, Huh SJ, Ekram MB et al: Immune escape in breast cancer during in situ to invasive carcinoma transition, *Cancer Discov* 7:1098, 2017. *[Some breast cancers, especially ER-negative and HER2-positive types, are associated with prominent T-cell infiltrates. How these cancers escape from immune surveillance and how to modulate this response to enhance response to therapy are promising areas of investigation].*

Ellsworth RE, Blackburn HL, Shriver CD et al: Molecular heterogeneity in breast cancer: state of the science and implications for patient care, *Semin Cell Dev Biol* 64:65, 2017. *[Single-cell sequencing has shown that every cell in a breast cancer may be genetically unique. Heterogeneity is one of the major challenges in developing effective treatment strategies for cancer].*

Fleischer T, Tekpli X, Mathelier A et al: DNA methylation at enhancers identifies distinct breast cancer lineages, *Nat Commun* 8:1379, 2018. *[Methylation is an important mechanism of genomic regulation that underlies major patterns of gene expression such as the expression of estrogen receptor].*

Geyer FC, Pareja F, Weigelt B et al: The spectrum of triple-negative breast disease: high- and low-grade lesions, *Am J Pathol* 187:2139, 2017. *[Although carcinomas that lack expression of hormone receptors and HER2 (triple negative carcinoma) generally confer a poor prognosis, there are many subtypes within this group, including some with a very good prognosis].*

Heng YJ, Lester SC, Tse GM et al: The molecular basis of breast cancer pathological phenotypes, *J Pathol* 241:375, 2017. *[In this study using data from The Cancer Genome Atlas, the gene expression patterns underlying pathologic features are explored. For example, tubule formation correlates strongly with genes expressed by luminal cancers].*

Mardamshina M, Geiger T: Next-generation proteomics and its application to clinical breast cancer research, *Am J Pathol* 187:2175, 2017. *[DNA sequencing and messenger RNA profiling provide only part of the profile of a cancer cell. Evaluation of the functional molecules (proteins) has been more challenging, but new techniques should allow large-scale analysis].*

Russnes HG, Lingjaerde OC, Borresen-Dale A-L et al: Breast cancer molecular stratification: from intrinsic subtypes to integrative clusters, *Am J Surg Pathol* 187:2152, 2017. *[Modern classification of breast cancers includes alterations in DNA and messenger RNA patterns, protein expression, and the resulting phenotype as recognized as histologic type].*

Resources for Classification, Staging, Reporting, and Treating Breast Cancer

Bossuyt V, Provenzano E, Symmans WF et al: Recommendations for standardized pathological characterization of residual disease for neoadjuvant clinical trials of breast cancer by the BIG-NABCG collaboration, *Ann Oncol* 26:1280–1291, 2015. *[Neoadjuvant therapy is a powerful tool for both patient care and research, as it is currently the only method to directly measure the degree to which carcinomas respond to different types of therapy].*

Hortobagyi GN, Connolly JL, D'Orsi CJ et al: Breast. In Amin MB, Edge SB, et al; for the American Joint Commission on Cancer, editors: *AJCC Cancer Staging Manual*, New York, NY, 2017, Springer, pp 589–628. Revised Breast Chapter available at: https://cancerstaging.org/About/news/Pages/Updated-Breast-Chapter-for-8th-Edition.aspx. (Accessed 23 December 2017). *[The American Joint Commission on Cancer issues guidelines on cancer staging. After publication of the book, a significantly revised Breast Chapter was provided online. This edition introduces two new staging systems that for the first time combine biologic features of breast cancers with anatomic stage].*

International Agency for Research on Cancer (IARC): *World Health Organization Classification of Tumours: Breast Tumours*, ed 5, Lyon, France, December 2019, IARC. *[Over 100 experts from around the world collaborate with the World Health Organization to issue guidelines on the classification of breast cancers].*

Lester SC, Bose S, Chen YY et al: Protocol for the examination of specimens from patients with invasive carcinoma of the breast, *Arch Pathol Lab Med* 133:1515, 2009. *[The College of American Pathologists (CAP) develops national standards for the reporting of breast carcinoma and is currently working on developing a worldwide consensus. Current CAP protocols are available at www.cap.org].*

National Comprehensive Cancer Network (NCCN). https://www.nccn.org. *[The National Comprehensive Cancer Network is an alliance of cancer centers that issues guidelines for the treatment of breast cancer as well as other cancers].*

[a]The largest datasets with molecular information on breast cancer include these.

[b]These are large data repositories for clinical information about breast cancer patients including types of cancer, stage, and survival.

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 24

The Endocrine System

Anirban Maitra

CHAPTER CONTENTS

The endocrine system consists of a highly integrated and widely distributed group of organs, called glands, that orchestrate a state of metabolic equilibrium among the various organs of the body. Signaling by secreted molecules can be classified into three types—autocrine, paracrine, or endocrine—on the basis of the distance over which the signal acts. **In endocrine signaling, the secreted molecules, also known as hormones, act on target cells that are distant from their sites of synthesis.** An endocrine hormone is frequently carried by the blood from its site of release to its target. The production of several hormones from endocrine glands is stimulated by trophic factors released from the

pituitary. The endocrine hormone inhibits production of the trophic factors, a process known as *feedback inhibition*, thus maintaining physiologic levels of the hormone.

Several processes can disturb the normal activity of the endocrine system, including impaired synthesis or release of hormones, abnormal interactions between hormones and their target tissues, and abnormal responses of target organs. Endocrine diseases can be generally classified as (1) diseases of underproduction or overproduction of hormones and their resulting biochemical and clinical consequences, and (2) diseases associated with the development of mass lesions. Such lesions might be nonfunctional, or they might be associated with overproduction or underproduction of hormones. The study of endocrine diseases requires integration of morphologic findings with biochemical measurements of the levels of hormones, their regulators, and other metabolites.

Pituitary Gland

The pituitary gland is a small, bean-shaped structure that lies at the base of the brain within the sella turcica. Its function is controlled by the hypothalamus, to which it is connected by a stalk containing axons extending from the hypothalamus and a rich venous plexus. Along with the hypothalamus, the pituitary has a central role in regulating the function of most of the other endocrine glands.

The pituitary gland is composed of two morphologically and functionally distinct components: the anterior lobe (adenohypophysis) and the posterior lobe (neurohypophysis). **The anterior pituitary, or adenohypophysis, which constitutes about 80% of the gland, produces trophic hormones that stimulate the production of hormones from the thyroid, adrenal, and other glands.** The anterior pituitary is composed of epithelial cells derived embryologically from the developing oral cavity. In routine histologic sections, it contains a colorful array of cells that variously have eosinophilic cytoplasm *(acidophils)*, basophilic cytoplasm *(basophils)*, or poorly staining cytoplasm *(chromophobe cells)* (Fig. 24.1). Detailed studies have demonstrated that the distinct staining properties of these cells are related to the presence of different polypeptide hormones within their cytoplasm that control the activity of other endocrine glands. There are six terminally differentiated cell types in the anterior pituitary, each of which is defined by the hormones that it synthesizes:

- *Somatotrophs* produce growth hormone (GH).
- *Mammosomatotrophs* produce GH and prolactin (PRL).
- *Lactotrophs* produce PRL.
- *Corticotrophs* produce adrenocorticotropic hormone (ACTH), pro-opiomelanocortin (POMC), and melanocyte-stimulating hormone (MSH).
- *Thyrotrophs* produce thyroid-stimulating hormone (TSH).
- *Gonadotrophs* produce follicle-stimulating hormone (FSH) and luteinizing hormone (LH). In women, FSH stimulates the formation of graafian follicles in the ovary, and LH induces ovulation and the formation of corpora lutea in the ovary. The same two hormones also regulate spermatogenesis and testosterone production in males.

The production of most pituitary hormones is controlled by positively and negatively acting factors from the hypothalamus (Fig. 24.2), which are carried to the anterior pituitary by the portal venous plexus. While most hypothalamic factors promote pituitary hormone release, others (e.g., somatostatin and dopamine) are inhibitory. Rarely, signs and symptoms of pituitary disease may be caused by overproduction or underproduction of hypothalamic factors, rather than a primary pituitary abnormality. During embryogenesis, specific transcription factors regulate the differentiation of multipotent stem cells within the Rathke pouch into the various cell types of the anterior pituitary. For example, somatotrophs, mammosomatotrophs, lactotrophs, and thyrotrophs are all derived from a common precursor that expresses the transcription factor PIT-1 (lactrotrophs also express the alpha subunit of the estrogen receptor, ERα). By contrast, corticotrophs are derived from progenitor cells expressing the transcription factor TPIT (also known as T-box protein 19 or Tbx19), and gonadotrophs are derived from precursor cells expressing steroidogenic factor-1 (SF-1) and GATA-2. The expression of these lineage-specific transcription factors is retained in pituitary adenomas and is used to classify these tumors (see later).

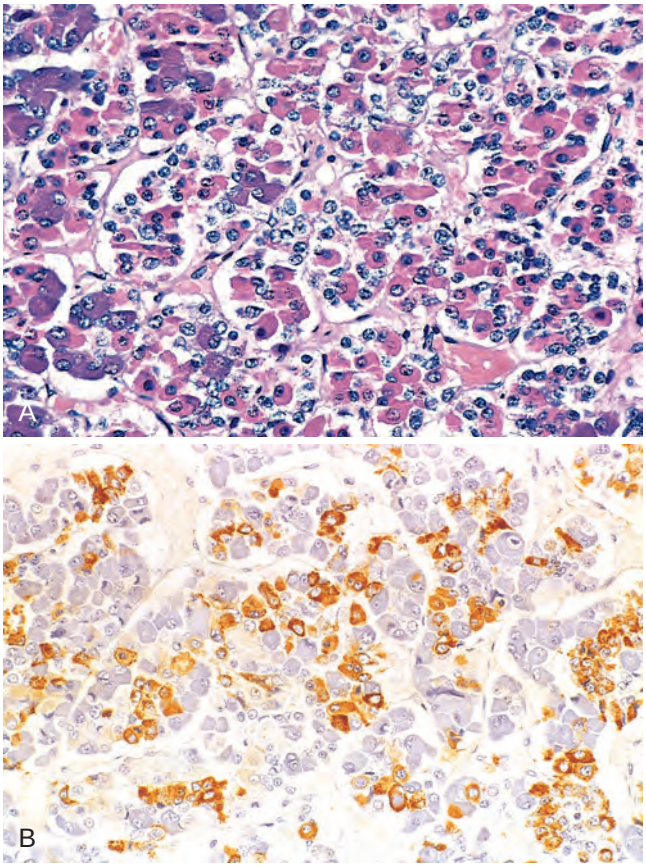


Figure 24.1 (A) Photomicrograph of a normal pituitary. The gland is populated by several distinct cell populations containing a variety of stimulating (tropic) hormones. Each hormone has different staining characteristics, resulting in a mixture of cell types in routine histologic preparations. (B) Immunostain for human growth hormone.

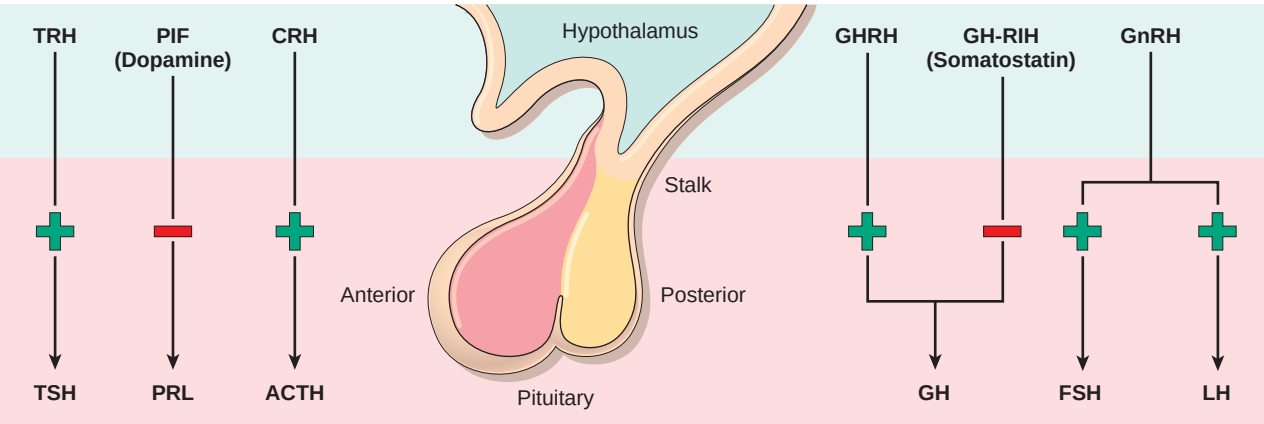


Figure 24.2 Hormones released by the anterior pituitary. The adenohypophysis (anterior pituitary) releases five hormones that are, in turn, under the control of various stimulatory and inhibitory hypothalamic releasing factors. The stimulatory releasing factors are TRH (thyrotropin-releasing hormone), CRH (corticotropin-releasing hormone), GHRH (growth hormone–releasing hormone), and GnRH (gonadotropin-releasing hormone). The inhibitory hypothalamic influences comprise PIF (prolactin inhibitory factor or dopamine) and growth hormone–release inhibiting hormone (GH-RIH or somatostatin). *TSH*, Thyroid-stimulating hormone (thyrotropin); *PRL*, prolactin; *ACTH*, adrenocorticotropic hormone (corticotropin); *GH*, growth hormone (somatotropin); *FSH*, follicle-stimulating hormone; *LH*, luteinizing hormone.

The posterior pituitary consists of modified glial cells (termed pituicytes) and axonal processes extending from the hypothalamus through the pituitary stalk to the posterior lobe (axon terminals). Two peptide hormones are secreted from the posterior pituitary, *oxytocin* and *antidiuretic hormone* (ADH, also called *arginine vasopressin* or AVP). These hormones are actually synthesized in the hypothalamus and are transported through axons to the posterior pituitary. In response to appropriate stimuli, the preformed hormones are then released directly into the systemic circulation. For example, dilation of the cervix in pregnancy results in oxytocin release, leading to contraction of the uterine smooth muscle, facilitating parturition (uterine labor). Similarly, oxytocin released on nipple stimulation in the postnatal period acts on the smooth muscles surrounding the lactiferous ducts of the mammary glands and facilitates lactation. Synthetic oxytocin can be given to pregnant women to induce labor. The most important function of ADH is to conserve water by restricting diuresis during periods of dehydration and hypovolemia. Decreased blood pressure, sensed by baroreceptors (pressure-sensing receptors) in the cardiac atria and carotids, stimulates ADH release. An increase in plasma osmotic pressure detected by osmoreceptors also triggers ADH secretion. In contrast, states of hypervolemia and increased atrial distention result in inhibition of ADH secretion.

CLINICAL MANIFESTATIONS OF PITUITARY DISEASE

The manifestations of pituitary disorders are related to either excess or deficiency of pituitary hormones, or to mass effects.

- *Hyperpituitarism* arises from excess secretion of trophic hormones. The causes of hyperpituitarism include hyperplasias, adenomas, and carcinomas of the anterior pituitary, secretion of hormones by nonpituitary tumors, and certain hypothalamic disorders. The symptoms of hyperpituitarism are discussed later in the context of individual tumors.
- *Hypopituitarism* results from deficiency of trophic hormones. It may be caused by ischemic injury, surgery or radiation, inflammatory disorders, and the mass effects of nonfunctional pituitary adenomas.
- *Symptoms related to local mass effects.* Because of the close proximity of the optic nerves and chiasm to the sella, expanding pituitary lesions often compress decussating fibers in the optic chiasm. This gives rise to visual field abnormalities, classically in the form of defects in both lateral (temporal) visual fields, so-called bitemporal hemianopsia. Like any expanding intracranial mass, pituitary adenomas can produce signs and symptoms of elevated intracranial pressure, including headache, nausea, and vomiting. On occasion, acute hemorrhage into an adenoma is associated with clinical evidence of rapid enlargement of the lesion, a situation appropriately termed *pituitary apoplexy*. Acute pituitary apoplexy is a neurosurgical emergency because it can cause sudden death (see later).

Diseases of the posterior pituitary often come to clinical attention because of increased or decreased secretion of ADH and associated changes in fluid and electrolyte balances.

PITUITARY ADENOMAS AND HYPERPITUITARISM

The most common cause of hyperpituitarism is an adenoma arising in the anterior lobe. Pituitary adenomas are classified on the basis of the hormones and cell type–specific transcription factors that are expressed by the tumor cells (Table 24.1). Some pituitary adenomas secrete two hormones (GH and prolactin being the most common combination), and rarely, pituitary adenomas are plurihormonal. Less common causes of hyperpituitarism include pituitary carcinomas and

Table 24.1 Classification of Pituitary Adenomas

Adenoma Type	Hormone	Transcription Factor	Morphologic Variant	Associated Syndrome[a]
Somatotroph adenoma	GH GH GH, PRL (in same cells) GH, PRL (in different cells)	PIT-1 PIT-1 PIT-1, ERα PIT-1, ERα	Densely granulated adenoma Sparsely granulated adenoma Mammosomatotroph adenoma Mixed somatotroph-lactotroph adenoma	Gigantism (children) Acromegaly (adults)
Lactotroph adenoma	PRL PRL PRL, GH (focal and variable)	PIT-1, ERα PIT-1, ERα PIT-1, ERα	Sparsely granulated adenoma Densely granulated adenoma Acidophilic stem cell adenoma	Galactorrhea and amenorrhea (in females) Sexual dysfunction, infertility
Thyrotroph adenoma	TSH	PIT-1	Thyrotroph adenoma	Hyperthyroidism
Corticotroph adenoma	ACTH ACTH ACTH	TPIT TPIT TPIT	Densely granulated adenoma Sparsely granulated adenoma Crooke cell adenoma (prominent intracytoplasmic cytokeratin filaments)	Cushing syndrome Nelson syndrome Mass effects (20% of corticotroph adenomas are hormonally silent)
Gonadotroph adenoma	FSH, LH	SF-1, GATA-2, ERα	Gonadotroph adenoma	Mass effects and hypopituitarism (most gonadotroph adenomas are hormonally silent)
Null cell adenoma	None	None		Mass effects
Plurihormonal adenoma	GH, PRL, TSH	PIT-1		

[a]Note that nonfunctional ("silent") adenomas in each category express the corresponding hormone(s) and transcription factor within the neoplastic cells, as determined by special immunohistochemical staining on tissues. However, these adenomas do not produce the associated clinical syndrome and typically present with *mass effects* accompanied by *hypopituitarism* due to destruction of normal pituitary parenchyma. These features are particularly common with gonadotroph adenomas and up to one-fifth of corticotroph adenomas, while rarely observed with somatotroph or lactotroph adenomas. Null cell adenomas, by definition, only present with mass effects.

ACTH, Adrenocorticotropic hormone; *FSH*, follicle-stimulating hormone; *GH*, growth hormone; *LH*, luteinizing hormone; *PRL*, prolactin; *TSH*, thyroid-stimulating hormone.

Partially modified from Lopes MBS: The 2017 World Health Organization classification of tumors of the pituitary gland: a summary, *Acta Neuropathol* 134:521–535, 2017.

some hypothalamic disorders. By contrast, large pituitary adenomas, and particularly nonfunctioning ones, may cause hypopituitarism by destroying the adjacent normal anterior pituitary parenchyma.

Pituitary adenomas are usually found in adults; the peak incidence is from 35 to 60 years of age. They are designated, somewhat arbitrarily, *microadenomas* if they are less than 1 cm in diameter and *macroadenomas* if they exceed 1 cm in diameter. Nonfunctional adenomas are likely to come to clinical attention at a later stage than those associated with endocrine abnormalities and are therefore more likely to be macroadenomas. Based on autopsy studies, the prevalence of pituitary adenomas in the population is estimated to be about 14%, but the vast majority of these lesions are clinically silent microadenomas ("pituitary incidentaloma").

Pathogenesis

As with other neoplasms, pituitary adenomas are caused by mutations in cancer genes, which are most commonly acquired somatic mutations but which may also be germline mutations associated with an inherited predisposition to pituitary neoplasms (Table 24.2):

- *Activating G-protein mutations* are one of the most common alterations in pituitary adenomas. As described in Chapter 1, G-proteins normally play a critical role in signal transduction, transmitting signals from particular cell surface receptors (e.g., GHRH receptor) to intracellular effectors (e.g., adenyl cyclase), which then generate second messengers (e.g., cyclic adenosine monophosphate, cAMP). They are heterotrimeric proteins, composed of α-subunits that bind guanine nucleotide and interact with cell surface receptors and intracellular effectors (Fig. 24.3), and β- and γ-subunits that noncovalently bind α-subunits. G_s is a stimulatory G-protein with a pivotal role in signal transduction in several endocrine organs, including the pituitary. The α-subunit of G_s ($G_s\alpha$) is encoded by the *GNAS* gene, located on chromosome 20q13. In the basal state, G_s exists in an inactive state, with guanosine diphosphate (GDP) bound to the guanine nucleotide-binding site of $G_s\alpha$. On interaction with the ligand-bound cell surface receptor, GDP dissociates, and guanosine triphosphate (GTP) binds to $G_s\alpha$, activating the G-protein. The activation of $G_s\alpha$ generates cAMP, a potent mitogen for several types of endocrine cells (e.g., pituitary somatotrophs and corticotrophs, thyroid follicular cells, parathyroid cells). Normally, $G_s\alpha$ activation is transient because of an intrinsic GTPase activity in the α-subunit, which hydrolyzes GTP into GDP. Approximately 40% of somatotroph cell adenomas bear somatic *GNAS* mutations that abrogate the GTPase activity of $G_s\alpha$, leading to constitutive activation of $G_s\alpha$, persistent generation of cAMP, and unchecked cellular proliferation (see Table 24.2). *GNAS* mutations have also been described in a minority of corticotroph adenomas; in contrast, *GNAS* mutations are absent in thyrotroph, lactotroph, and gonadotroph adenomas, because their trophic hypothalamic release hormones act via other signaling pathways.
- Activating mutations of *ubiquitin-specific protease 8 (USP8)* also occur in 30% to 60% of corticotroph adenomas. The encoded protein is an enzyme that removes ubiquitin residues from proteins like epidermal growth factor receptor (EGFR), protecting them from proteasome-dependent

Table 24.2 Genetic Alterations in Pituitary Tumors

Gene	Protein Function	Oncogenic Mutations	Most Commonly Associated Pituitary Tumor
GNAS	α subunit of stimulatory G-protein, Gsα	*Somatic* activating mutation	Somatotroph adenoma
USP8	Deubiquitinase	*Somatic* activating mutation	Corticotroph adenoma
Protein kinase A, regulatory subunit 1α *(PRKAR1A)*[a]	Negative regulator of protein kinase A (PKA), leading to increased cAMP production	*Germline* inactivating mutations (Carney complex)	Somatotroph or lactotroph adenoma
MEN1[a]	Transcription regulator	*Germline* inactivating mutations (multiple endocrine neoplasia, type 1)	Somatotroph, lactotroph, or corticotroph adenoma
CDKN1B (p27/KIP1)[a]	Negative cell cycle regulator	*Germline* inactivating mutations ("MEN-1-like" syndrome)	Corticotroph adenoma
Aryl hydrocarbon receptor interacting protein *(AIP)*[a]	Receptor for aryl hydrocarbons and a ligand-activated transcription factor	*Germline* inactivating mutations (familial isolated pituitary adenoma syndrome)	Somatotroph or lactotroph adenoma (especially in patients younger than 35 years of age)
HRAS	Mitogenic signaling, cell growth and survival	*Somatic* activating mutation	Pituitary carcinoma
DICER1[a]	MicroRNA processing	*Germline* inactivating mutation	Pituitary blastoma

[a]Genetic alterations associated with *familial* or *germline* predisposition to pituitary adenomas.

degradation. Aberrant activation of USP8 thus enhances the activity of EGFR and other pro-growth signaling pathways in pituitary adenomas.

- Approximately 5% of pituitary adenomas are caused by germline loss-of-function mutations in genes such as *MEN1, CDKN1B, PRKAR1A,* or *AIP* (see Table 24.2). Of note, somatic mutations of these four genes are rarely encountered in sporadic pituitary adenomas.

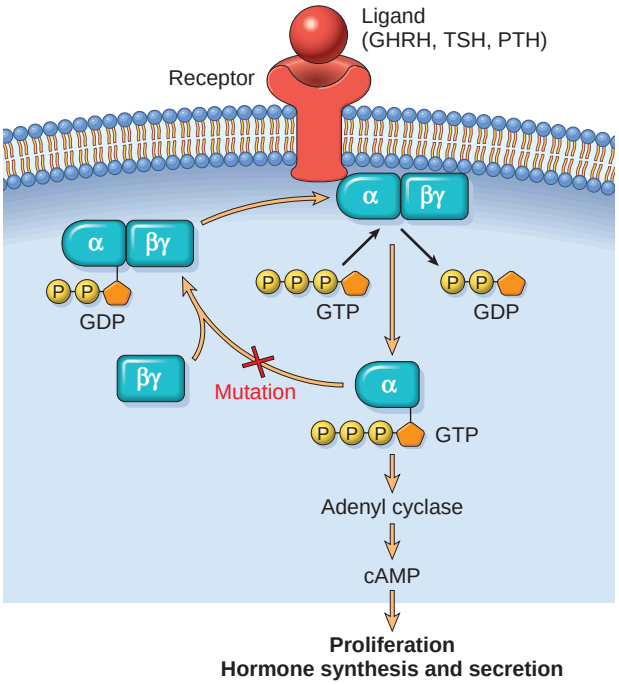


Figure 24.3 G-protein signaling in endocrine neoplasia. Mutations that lead to G-protein hyperactivity are seen in a variety of endocrine neoplasms, including pituitary, thyroid, and parathyroid adenomas. G-proteins (composed of α and βγ subunits) play a critical role in signal transduction, transmitting signals from cell surface receptors (GHRH, TSH, or PTH receptor) to intracellular effectors (e.g., adenyl cyclase), which then generate second messengers (*cAMP,* cyclic adenosine monophosphate) that stimulate cellular responses. *GDP,* Guanosine diphosphate; *GTP,* guanosine triphosphate; *Pi,* inorganic phosphate. See Fig. 24.2 for other abbreviations.

MORPHOLOGY

The **typical pituitary adenoma** is soft and well-circumscribed. Small adenomas may be confined to the sella turcica, but with expansion they frequently erode the sella turcica and anterior clinoid processes. Larger lesions may extend superiorly through the diaphragm sella into the suprasellar region, compressing the optic chiasm and adjacent structures, such as cranial nerves (Fig. 24.4). In as many as 30% of cases, the adenomas are not encapsulated and infiltrate neighboring tissues such as the cavernous and sphenoid sinuses, dura, and on occasion, the brain itself. Such lesions are termed **aggressive adenomas.** Not unexpectedly, macroadenomas are more likely to be invasive and to have foci of hemorrhage and necrosis.

Histologically, typical pituitary adenomas are composed of uniform (monomorphic), polygonal cells arrayed in sheets or cords.

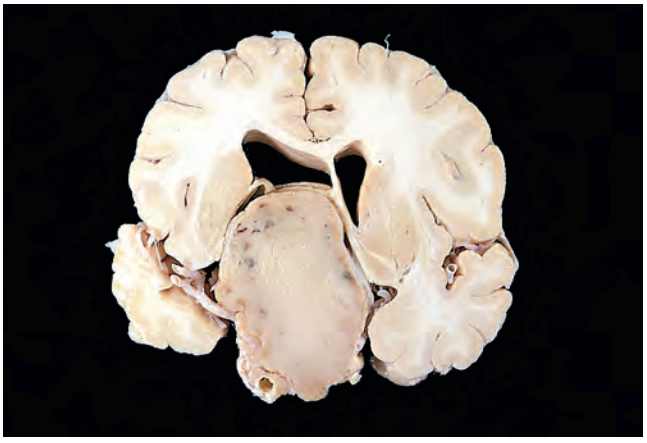


Figure 24.4 Pituitary adenoma. This massive, nonfunctional adenoma has grown beyond the confines of the sella turcica, distorting the overlying brain. On average, nonfunctional adenomas are larger at time of diagnosis.

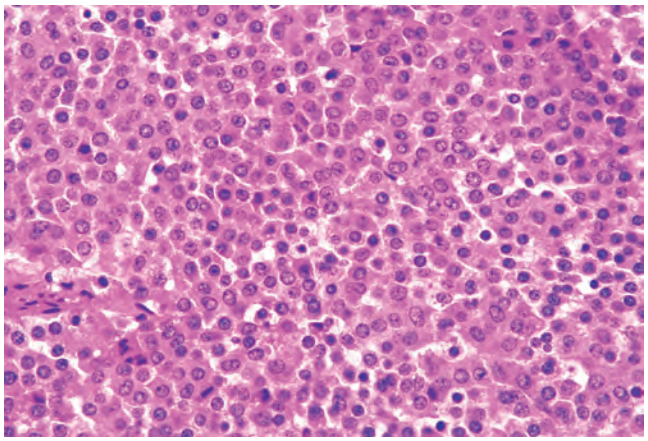


Figure 24.5 Pituitary adenoma. The monomorphism of these cells contrasts markedly with the mixture of cells seen in the normal anterior pituitary. Note also the absence of reticulin network.

The supporting connective tissue, or reticulin, is sparse, accounting for the soft, gelatinous consistency of many of these tumors. This cellular monomorphism and the absence of a significant reticulin network distinguish pituitary adenomas from normal anterior pituitary parenchyma (Fig. 24.5). Immunohistochemical stains for hormones and lineage-specific transcription factors are used to identify specific pituitary adenoma subtypes (see Table 24.1). Mitotic activity and expression of Ki-67 (MIB-1, a marker of cycling cells) are typically low in pituitary adenomas; higher-than-normal rates of cell division are associated with more aggressive tumors.

Clinical Features

The signs and symptoms of pituitary adenomas are related to endocrine abnormalities and mass effects. The effects of excessive secretion of anterior pituitary hormones are mentioned later, when the specific types of pituitary adenoma are described. Local mass effects may be produced by any type of pituitary tumor. As mentioned earlier, these effects include radiographic abnormalities of the sella turcica, visual field abnormalities, signs and symptoms of elevated intracranial pressure, and occasionally hypopituitarism. Acute hemorrhage into an adenoma is sometimes associated with *pituitary apoplexy,* as noted earlier. The biologic behavior of a pituitary adenoma cannot always be reliably predicted from its histologic appearance. This has led some to suggest that the term "adenoma" (which implies a benign course) should be replaced by *pituitary neuroendocrine tumor*, a name that does not reflect any specific expectation about a tumor's behavior.

The following is a description of the individual types of tumors.

Lactotroph Adenoma

Prolactin-secreting lactotroph adenomas are the most common type of hyperfunctioning pituitary adenoma, accounting for about 30% of clinically recognized cases. These lesions range from small microadenomas to large, expansile tumors associated with symptomatic mass effects.

MORPHOLOGY

The large majority of lactotroph adenomas are composed of chromophobe cells with juxtanuclear localization of the transcription factor PIT-1; these are known as **sparsely granulated lactotroph adenomas** (Fig. 24.6A). Much rarer are the eosinophilic **densely granulated lactotroph adenomas,** characterized by diffuse cytoplasmic PIT-1 localization (Fig. 24.6B). Prolactin can be demonstrated within cytoplasmic secretory granules with immunohistochemical stains, and estrogen receptor alpha (ERα) is co-expressed along with PIT-1, consistent with lactotroph differentiation. Lactotroph adenomas often undergo dystrophic calcification, ranging from isolated psammoma bodies to extensive calcification ("pituitary stone"). Prolactin secretion by functioning adenomas is usually efficient and proportional, in that serum prolactin concentrations tend to correlate with the size of the adenoma.

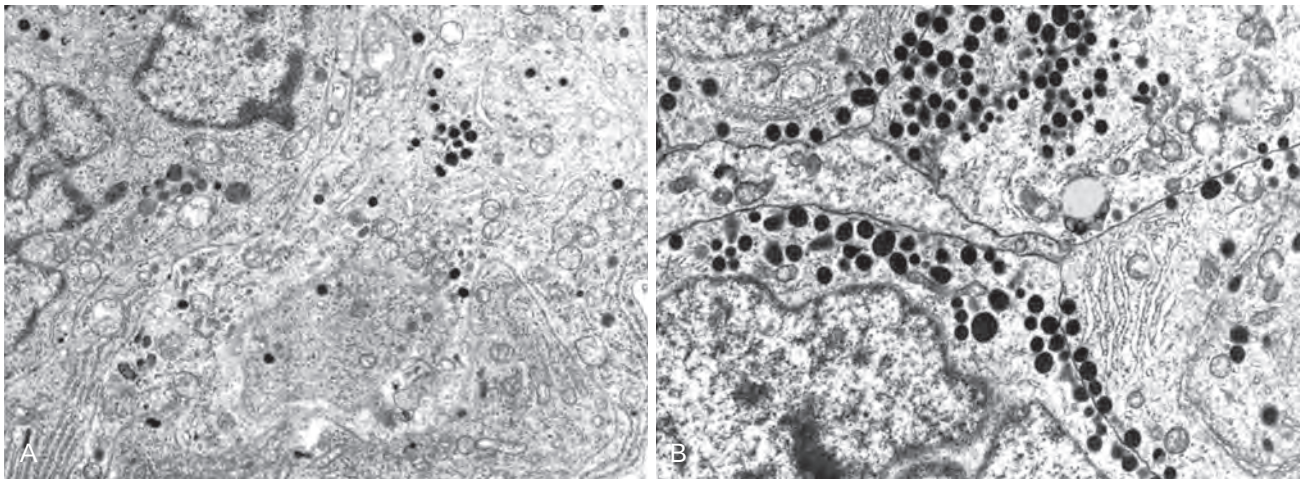


Figure 24.6 Ultrastructural features of prolactinoma. (A) Electron micrograph of a sparsely granulated prolactinoma. The tumor cells contain abundant granular endoplasmic reticulum (indicative of active protein synthesis) and small numbers of electron-dense secretory granules. (B) Electron micrograph of a densely granulated growth hormone–secreting adenoma. The tumor cells are filled with numerous large, electron-dense secretory granules. (Courtesy Dr. Eva Horvath, St. Michael's Hospital, Toronto, Ontario, Canada.)

Clinical Features

Increased serum levels of prolactin, or *prolactinemia,* cause amenorrhea, galactorrhea, loss of libido, and infertility. The diagnosis of an adenoma is made more readily in women than in men, especially between 20 and 40 years of age, because hyperprolactinemia disrupts the menstrual cycle. Lactotroph adenoma is the cause of almost one-fourth of cases of amenorrhea. In contrast, in men and older women, the hormonal manifestations may be subtle, allowing the tumors to reach considerable size (macroadenomas) before being detected clinically.

Hyperprolactinemia may result from causes other than prolactin-secreting pituitary adenomas. Physiologic hyperprolactinemia occurs in pregnancy. Prolactin levels are also elevated by nipple stimulation, as occurs during suckling in lactating women, and as a response to many types of stress. Pathologic hyperprolactinemia can also result from lactotroph hyperplasia caused by loss of dopamine-mediated inhibition of prolactin secretion. This may occur due to damage of the dopaminergic neurons of the hypothalamus or the pituitary stalk (e.g., due to head trauma), or exposure to drugs that block dopamine receptors on lactotroph cells. Any suprasellar mass (e.g., a pituitary adenoma) may disturb the inhibitory influence of the hypothalamus on prolactin secretion, resulting in hyperprolactinemia. Therefore, mild hyperprolactenemia in a person with a pituitary adenoma does not necessarily indicate a prolactin-secreting tumor. Other causes of hyperprolactinemia include renal failure and hypothyroidism. Lactotroph adenomas are treated by surgery or, more commonly, with bromocriptine, a dopamine receptor agonist that causes the lesions to diminish in size.

Somatotroph Adenoma

Growth hormone (GH)-secreting somatotroph adenomas are the second most common type of functioning pituitary adenoma, and cause gigantism in children and acromegaly in adults. Somatotroph adenomas may be quite large by the time they come to clinical attention because the manifestations of excessive GH may be subtle.

MORPHOLOGY

Histologically, pure somatotroph adenomas are also classified into densely granulated and sparsely granulated subtypes. **Densely granulated somatotroph adenomas** are composed of monomorphic, eosinophilic cells with granular cytoplasm and a large central nucleus with prominent nucleoli. The cells have diffuse, strong immunoreactivity for GH. In contrast, the **sparsely granulated variants** are composed of chromophobe cells with patchy, weak staining for GH and a characteristic paranuclear glossy inclusion known as a fibrous body, composed of intermediate filaments that stain for cytokeratin. Bihormonal **mammosomatotroph** adenomas that synthesize GH and prolactin in the same cell are being increasingly recognized; morphologically, most resemble densely granulated somatotroph adenomas, but have immunohistochemical reactivity for prolactin and GH. These are to be contrasted with mixed **somatotroph-lactotroph adenomas,** which have GH and prolactin expression in different cells.

Clinical Features

In contrast to corticotroph or gonadotroph adenomas, silent somatotroph adenomas are rare. Persistently elevated levels of GH stimulate the hepatic secretion of insulin-like growth factor-1 (IGF-1), which causes many of the clinical manifestations.

- If a somatotroph adenoma appears in children before the epiphyses have closed, the elevated levels of GH (and IGF-1) result in *gigantism*. This is characterized by a generalized increase in body size with disproportionately long arms and legs.
- If the levels of GH are increased after closure of the epiphyses, *acromegaly* develops. In this condition, growth is most conspicuous in skin and soft tissues, viscera (thyroid, heart, liver, and adrenals), and the bones of the face, hands, and feet. Bone density may increase (*hyperostosis*) in both the spine and the hips. Enlargement of the jaw results in its protrusion (*prognathism*) and broadening of the lower face. The feet and hands are enlarged, and the fingers become thickened and sausage-like. In most instances, gigantism is also accompanied by evidence of acromegaly. These changes may develop slowly over decades before being recognized, and hence the underlying adenoma may reach substantial size.
- GH excess can also be associated with a variety of other disturbances, including gonadal dysfunction, diabetes mellitus, generalized muscle weakness, hypertension, arthritis, congestive heart failure, and an increased risk of gastrointestinal cancers.

The diagnosis relies on documenting elevated serum GH and IGF-1 levels. In addition, failure to suppress GH production in response to an oral load of glucose is one of the most sensitive tests for acromegaly. The causative pituitary adenoma can be removed surgically or treated by pharmacologic means. The latter includes somatostatin analogues (recall that somatostatin inhibits pituitary GH secretion) and GH receptor antagonists, which prevent hormone binding to target organs such as the liver. When effective control of high GH levels is achieved, the characteristic tissue overgrowth and related symptoms gradually recede, and the metabolic abnormalities improve. Overall, sparsely granulated somatotroph adenomas tend to have a more aggressive course than densely granulated adenomas and may be less responsive to somatostatin analogues. Thus, accurate subtyping of somatotroph adenomas is of prognostic importance.

Corticotroph Adenoma

Excess production of ACTH by functioning corticotroph adenomas leads to adrenal hypersecretion of cortisol and the development of hypercortisolism (also known as *Cushing syndrome*).

MORPHOLOGY

Corticotroph adenomas are usually microadenomas at the time of diagnosis. These tumors are most often basophilic **(densely granulated)** and occasionally chromophobic **(sparsely**

granulated). Both variants stain positively with periodic acid-Schiff (PAS) because of the presence of carbohydrate in proopiomelanocortin (POMC), the precursor of ACTH. Nuclear TPIT is also positive in the neoplastic cells, consistent with corticotroph lineage. A third uncommon variant, called **Crooke cell adenoma,** is characterized by ringlike deposition of cytokeratin called Crooke change. This variant has an aggressive natural history compared with other subtypes of corticotroph adenomas.

Clinical Features

The manifestations of Cushing syndrome are discussed in more detail later with the diseases of the adrenal gland. The syndrome can be caused by a wide variety of conditions in addition to ACTH-producing pituitary tumors. When the hypercortisolism is due to excessive production of ACTH by the pituitary, it is designated *Cushing disease*. Large destructive pituitary adenomas can develop in patients after surgical removal of the adrenal glands for treatment of Cushing syndrome. This condition, known as *Nelson syndrome,* occurs because of loss of the inhibitory effect of adrenal corticosteroids on a corticotroph microadenoma. Because the adrenals are absent in persons with Nelson syndrome, hypercortisolism does not develop. Patients present with mass effects due to the pituitary tumor, and there can be hyperpigmentation because melanotropin, which has trophic effects on melanocytes, is also derived from POMC.

Other Anterior Pituitary Tumors

Several less frequent types of pituitary adenoma also merit brief comment (see Table 24.1).

- *Gonadotroph (LH-producing and FSH-producing) adenomas* can be difficult to recognize because they secrete small amounts of hormones that usually do not cause a recognizable clinical syndrome (i.e., most are nonfunctioning). Gonadotroph adenomas most frequently present in middle-aged men and women with neurologic symptoms, such as impaired vision, headaches, diplopia, or pituitary apoplexy. Pituitary hormone deficiencies can also be found, most commonly impaired secretion of LH, which causes decreased energy and libido in men (due to reduced testosterone) and amenorrhea in premenopausal women. The neoplastic cells express the common gonadotropin α-subunit and the specific β-FSH and β-LH subunits; FSH is usually the predominant secreted hormone. Gonadotroph adenomas also usually express steroidogenic factor-1 (SF-1), GATA-2, and ERα, transcription factors associated with normal gonadotroph differentiation.
- *Thyrotroph (TSH-producing)* adenomas are uncommon, accounting for approximately 1% of pituitary adenomas. They are a rare cause of hyperthyroidism. Due to their shared lineage with lactotroph and somatotroph adenomas, these tumors also express PIT-1.
- Pituitary adenomas may also secrete multiple hormones; such "plurihormonal" adenomas are usually aggressive. Most are derived from cells of PIT-1–expressing lineage.
- *Null cell adenomas* do not express any markers of hormonal or lineage differentiation. Not surprisingly, null cell adenomas typically present with symptoms stemming from mass effects. These lesions may also compromise the residual anterior pituitary and cause hypopituitarism, which may appear slowly due to gradual enlargement of the adenoma or abruptly because of acute intratumor hemorrhage (pituitary apoplexy).

Pituitary carcinoma is rare, accounting for less than 1% of pituitary tumors. The presence of craniospinal or systemic metastases is a sine qua non of a pituitary carcinoma. Most pituitary carcinomas are functional, with prolactin and ACTH being the most commonly secreted products. Metastases usually appear late in the course, following multiple local recurrences.

Pituitary blastoma is an entity that occurs in children (typically younger than 2 years of age) who carry germline mutations of *DICER1,* the gene encoding a microRNA-processing protein. Morphologically, these tumors are composed of immature "blastema-like" cells (so-called "small round blue cells") and rosette-like formations resembling the primitive Rathke epithelium from which the pituitary develops. Pituitary blastoma presents with signs and symptoms of Cushing disease. These children also develop primitive "blastema-like" neoplasms in other organs, most commonly pleuropulmonary blastoma.

KEY CONCEPTS

HYPERPITUITARISM

- The most common cause of hyperpituitarism is an anterior lobe pituitary adenoma.
- Pituitary adenomas can be macroadenomas (greater than 1 cm in diameter) or microadenomas.
- Functioning adenomas are associated with distinct endocrine signs and symptoms, while nonfunctioning (silent) adenomas typically present with mass effects, including visual disturbances.
- Lactotroph adenomas secrete prolactin and can present with amenorrhea, galactorrhea, loss of libido, and infertility.
- Somatotroph adenomas secrete GH and present with gigantism in children and acromegaly in adults, impaired glucose tolerance, and diabetes mellitus.
- Corticotroph adenomas secrete ACTH and present with Cushing syndrome and hyperpigmentation.
- The two distinctive morphologic features of most adenomas are their cellular monomorphism and absence of a reticulin network.

HYPOPITUITARISM

Hypopituitarism refers to decreased secretion of pituitary hormones, which can result from diseases of the hypothalamus or of the pituitary. Hypofunction of the anterior pituitary occurs when approximately 75% of the parenchyma is lost or absent. When accompanied by evidence of posterior pituitary dysfunction in the form of diabetes insipidus (see later), hypopituitarism is almost always of hypothalamic origin.

Most cases of hypopituitarism arise from destructive processes involving the anterior pituitary, as follows:

- *Tumors and other mass lesions:* Pituitary adenomas, other benign tumors arising within the sella, primary and metastatic malignancies, and cysts can cause hypopituitarism. Any mass lesion in the sella can cause damage by exerting pressure on adjacent normal pituitary cells.
- *Traumatic brain injury and subarachnoid hemorrhage* are among the most common causes of pituitary hypofunction.
- *Pituitary surgery or radiation:* Surgical excision of a pituitary adenoma may inadvertently include the nonadenomatous pituitary. Radiation of the pituitary, used to prevent regrowth of residual tumor after surgery, can damage the nonadenomatous pituitary.
- *Pituitary apoplexy:* As mentioned earlier, this is caused by a sudden hemorrhage into the pituitary gland, often occurring into a pituitary adenoma. In its most dramatic presentation, apoplexy causes the abrupt onset of excruciating headache, diplopia due to pressure on the oculomotor nerves, and acute hypopituitarism. In severe cases, it can cause cardiovascular collapse, loss of consciousness, and even sudden death. The combination of mass effect from the hemorrhage and acute hypopituitarism makes pituitary apoplexy a neurosurgical emergency.
- *Ischemic necrosis of the pituitary (Sheehan syndrome)*, also known as postpartum necrosis, is the most common form of ischemic necrosis of the anterior pituitary. During pregnancy, the anterior pituitary enlarges to almost twice its normal size. This physiologic expansion of the gland is not accompanied by an increase in blood supply from the low-pressure venous system; hence, there is relative hypoxia. Any further reduction in blood supply caused by obstetric hemorrhage or shock may precipitate infarction of the anterior lobe. Because the posterior pituitary is supplied directly from arterial branches, it is much less susceptible to ischemic injury and is therefore usually unaffected. Pituitary necrosis may also be encountered in disseminated intravascular coagulation and, less commonly, in sickle cell anemia, elevated intracranial pressure, traumatic injury, and shock of any origin. Whatever the pathogenesis, the ischemic area is resorbed and replaced by a nubbin of fibrous tissue attached to the wall of an empty sella.
- *Rathke cleft cyst:* These cysts, lined by ciliated cuboidal epithelium with occasional goblet cells and anterior pituitary cells, can accumulate proteinaceous fluid and expand, compromising the normal gland.
- *Empty sella syndrome:* Any condition or treatment that destroys part or all of the pituitary gland, such as ablation of the pituitary by surgery or radiation, can result in an empty sella and the empty sella syndrome. There are two types: (1) In a primary empty sella, a defect in the diaphragma sella allows the arachnoid mater and cerebrospinal fluid to herniate into the sella, expanding the sella and compressing the pituitary. Classically, this occurs in obese women with a history of multiple pregnancies. Affected individuals often present with visual field defects and occasionally with endocrine anomalies, such as hyperprolactinemia, due to interruption of inhibitory hypothalamic inputs. Sometimes the loss of functioning parenchyma is sufficient to produce hypopituitarism. (2) In secondary empty sella, a mass, such as a pituitary adenoma, enlarges the sella and is then either surgically removed or undergoes infarction, leading to loss of pituitary function.
- *Hypothalamic lesions:* As mentioned earlier, hypothalamic lesions can also affect the pituitary by causing a deficiency of pituitary hormone–releasing factors. In contrast to diseases that involve the pituitary directly, hypothalamic abnormalities can also diminish the secretion of ADH, resulting in diabetes insipidus (discussed later). Hypothalamic lesions that cause hypopituitarism include tumors, which may be benign (e.g., craniopharyngioma) or malignant; most of the latter are metastases from tumors such as breast and lung carcinoma. Hypothalamic insufficiency may also appear following irradiation of the brain.
- *Inflammatory disorders and infections,* such as sarcoidosis or tuberculous meningitis, can involve the hypothalamus and cause deficiencies of anterior pituitary hormones and diabetes insipidus.
- *Genetic defects:* Congenital deficiency of transcription factors required for normal pituitary function is a rare cause of hypopituitarism. For example, mutation of the pituitary-specific gene *PIT1* results in combined pituitary hormone deficiency, characterized by deficiencies of GH, prolactin, and TSH.

The clinical manifestations of anterior pituitary hypofunction vary depending on the specific hormones that are lacking.

- Children can develop growth failure *(pituitary dwarfism)* due to growth hormone deficiency.
- Gonadotropin (LH and FSH) deficiency leads to amenorrhea and infertility in women and decreased libido, impotence, and loss of pubic and axillary hair in men.
- TSH and ACTH deficiencies result in symptoms of hypothyroidism and hypoadrenalism, respectively, and are discussed later in this chapter.
- Prolactin deficiency results in failure of postpartum lactation.
- The anterior pituitary is also a rich source of melanotropins (also known as melanocyte-stimulating hormone), synthesized from the same precursor molecule that produces ACTH; therefore, one of the manifestations of hypopituitarism includes pallor due to a loss of stimulatory effects on melanocytes.

POSTERIOR PITUITARY SYNDROMES

The clinically relevant posterior pituitary syndromes involve ADH and include diabetes insipidus and syndrome of inappropriate secretion of ADH.

- *Diabetes insipidus.* ADH deficiency causes diabetes insipidus, a condition characterized by excessive urination (polyuria) due to an inability of the kidney to resorb water properly from the urine. Diabetes insipidus can occur in a variety of conditions, including head trauma, tumors, inflammatory disorders of the hypothalamus and pituitary, and surgical complications. Rarely, diabetes insipidus has a genetic basis, either because of autosomal dominant mutations of the arginine vasopressin (AVP) gene, or mutations of arginine vasopressin receptor type

2 (AVPR2), an X-linked condition that usually presents in young boys. Diabetes insipidus from ADH deficiency is designated as central to differentiate it from nephrogenic diabetes insipidus, which is a result of renal tubular unresponsiveness to circulating ADH. The clinical manifestations of these two disorders are similar and include the excretion of large volumes of dilute urine with a lower than normal specific gravity. Serum sodium and osmolality are increased by the excessive renal loss of free water, resulting in thirst and polydipsia. Patients who can drink water generally compensate for the urinary losses, but patients who are obtunded, bedridden, or otherwise limited in their ability to obtain water may develop life-threatening dehydration.

- *Syndrome of inappropriate ADH (SIADH) secretion.* ADH excess causes over-resorption of free water, resulting in hyponatremia. The most frequent causes of SIADH are the secretion of ectopic ADH by malignant neoplasms (particularly small-cell carcinoma of the lung), drugs that increase ADH secretion, and a variety of central nervous system disorders, including infections and trauma. The clinical manifestations of SIADH are dominated by hyponatremia, cerebral edema, and resultant neurologic dysfunction. Although total body water is increased, blood volume remains normal, and peripheral edema does not develop.

HYPOTHALAMIC SUPRASELLAR TUMORS

Neoplasms in this location may induce hypofunction or hyperfunction of the anterior pituitary, diabetes insipidus, or combinations of these manifestations. The most commonly implicated tumors are glioma (sometimes arising in the chiasm; Chapter 28) and craniopharyngioma. Craniopharyngioma is thought to arise from vestigial remnants of Rathke pouch. These slow-growing tumors account for 1% to 5% of intracranial tumors. A small minority of these lesions occurs within the sella, but most are suprasellar, with or without intrasellar extension. A bimodal age distribution is observed, with one peak in childhood (5 to 15 years) and a second peak in adults 65 years of age or older. Patients usually come to attention because of headaches and visual disturbances, while children sometimes present with growth retardation due to pituitary hypofunction and GH deficiency.

MORPHOLOGY

Craniopharyngiomas average 3 to 4 cm in diameter; they may be encapsulated and solid, but more commonly they are cystic and sometimes multiloculated. They often encroach on the optic chiasm or cranial nerves, and not infrequently they bulge into the floor of the third ventricle and base of the brain. Two distinct histologic variants are recognized: adamantinomatous craniopharyngioma (most often observed in children) and papillary craniopharyngioma (most often observed in adults). The adamantinomatous type frequently contains radiologically demonstrable calcifications; the papillary variant calcifies only rarely.

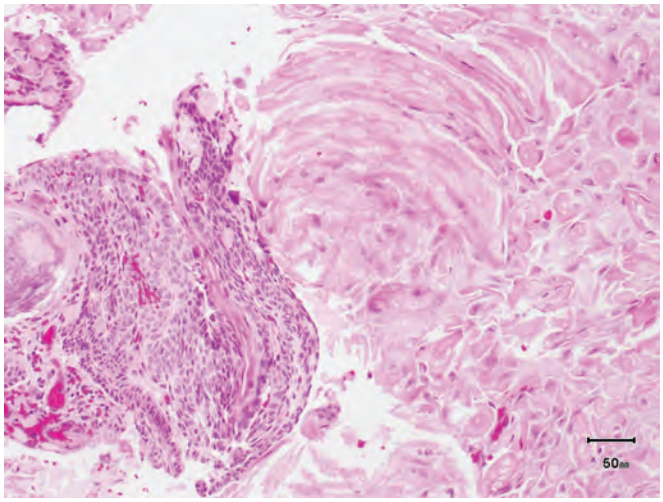


Figure 24.7 Adamantinomatous craniopharyngioma, demonstrating characteristic compact, lamellar "wet" keratin *(right half of photomicrograph)* and cords of squamous epithelium with peripheral palisading on the left. (Courtesy Dr. Charles Eberhart, Department of Pathology, Johns Hopkins University, Baltimore, Md.)

Adamantinomatous craniopharyngioma consists of nests or cords of stratified squamous epithelium embedded in a spongy "reticulum" that becomes more prominent in the internal layers. "Palisading" of the squamous epithelium is frequently observed at the periphery. Compact, lamellar keratin formation ("wet keratin") is a diagnostic feature of this tumor (Fig. 24.7). As mentioned earlier, **dystrophic calcification** is a frequent finding. Additional features include cyst formation, fibrosis, and chronic inflammation. The cysts of adamantinomatous craniopharyngiomas often contain a cholesterol-rich, thick brownish-yellow fluid that has been compared to "machine oil." These tumors extend fingerlets of epithelium into adjacent brain, where they elicit a brisk glial reaction. This subtype of craniopharyngioma is characterized by recurrent mutations of the *CTNNB1 (β-catenin)* gene, which leads to aberrant activation of the Wnt signaling pathway.

Papillary craniopharyngiomas contain both solid sheets of cells and papillae lined by well-differentiated squamous epithelium. These tumors usually are distinguished from the adamantinous type by the lack of lamellar keratin, calcification, cysts, peripheral palisading of squamous cells, and a spongy reticulum. Also unlike adamantinomatous craniopharyngioma, this subtype is characterized by activating mutations of the *BRAF* oncogene at codon 600. The identification of $BRAF^{V600E}$ mutations has therapeutic implications due to availability of small-molecule BRAF inhibitor drugs that inhibit the BRAF serine-threonine kinase (Chapter 7).

Patients with craniopharyngiomas, especially those less than 5 cm in diameter, have an excellent recurrence-free and overall survival. Larger lesions are more invasive, but this does not impact on the prognosis. Malignant transformation of craniopharyngiomas into squamous carcinomas is rare and usually occurs only after irradiation.

Thyroid Gland

The thyroid gland consists of two lateral lobes connected by a thin isthmus, usually located below and anterior to the larynx. It develops embryologically from an evagination of the pharyngeal epithelium that descends from the foramen cecum at the base of the tongue to its normal position in the anterior neck. This pattern of descent explains the occasional presence of ectopic thyroid tissue at the base of the tongue (lingual thyroid) or at other sites high in the neck. The thyroid is divided by thin fibrous septae into lobules composed of 20 to 40 follicles, lined by a cuboidal to low columnar epithelium and filled with PAS-positive thyroglobulin.

In response to hypothalamic factors, TSH *(thyrotropin)* is released by thyrotrophs in the anterior pituitary into the circulation. The binding of TSH to its receptor on thyroid follicular epithelial cells activates the receptor, which then associates with a G_s protein (Fig. 24.8). Activation of the G protein triggers downstream events that increase intracellular cAMP levels, which, in turn, stimulates thyroid growth and thyroid hormone synthesis and release via cAMP-dependent protein kinases.

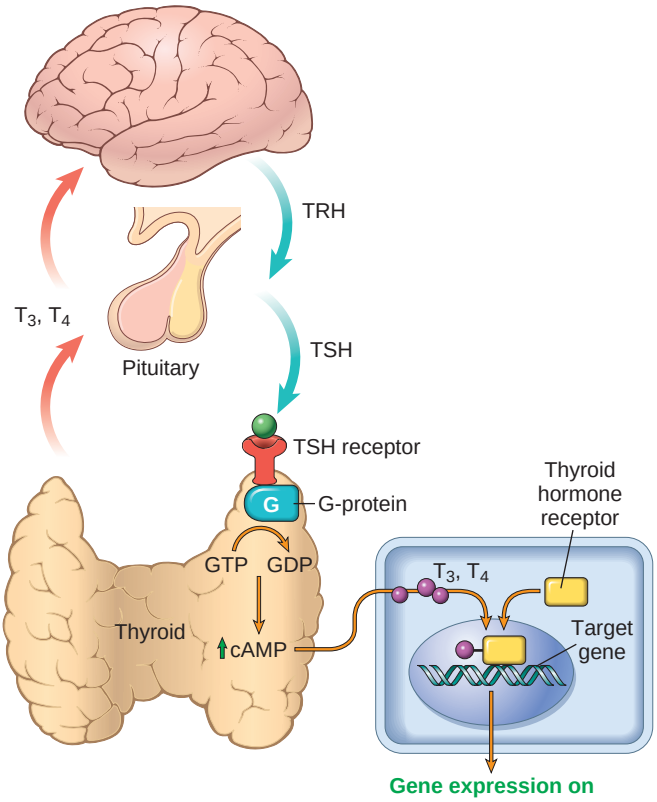


Figure 24.8 Homeostasis in the hypothalamus-pituitary-thyroid axis and mechanism of action of thyroid hormones. Secretion of thyroid hormones (T_3 and T_4) is controlled by trophic factors secreted by both the hypothalamus and the anterior pituitary. Decreased levels of T_3 and T_4 stimulate the release of thyrotropin-releasing hormone *(TRH)* from the hypothalamus and thyroid-stimulating hormone *(TSH)* from the anterior pituitary, causing T_3 and T_4 levels to rise. Elevated T_3 and T_4 levels, in turn, feed back to suppress the secretion of both TRH and TSH. TSH binds to the TSH receptor on the thyroid follicular epithelium, which causes activation of G-proteins, and cAMP-mediated synthesis and release of thyroid hormones (T_3 and T_4). In the periphery, T_3 and T_4 interact with the thyroid hormone receptor to form a hormone-receptor complex that translocates to the nucleus and binds to so-called thyroid response elements on target genes to initiate transcription.

Thyroid follicular epithelial cells convert thyroglobulin into *thyroxine (T_4)* and lesser amounts of *triiodothyronine (T_3)*. T_4 and T_3 are released into the systemic circulation, where most of these peptide hormones are reversibly bound to circulating plasma proteins, such as thyroxine-binding globulin and transthyretin. The binding proteins act as a buffer that maintains the serum unbound ("free") T_3 and T_4 concentrations within narrow limits, while ensuring that adequate amounts of the hormones are readily available to the tissues. In the periphery, the majority of free T_4 is deiodinated to T_3; the latter binds to thyroid hormone nuclear receptors in target cells with tenfold greater affinity than does T_4 and has proportionately greater activity. Binding of thyroid hormone to its nuclear thyroid hormone receptor (TR) results in the assembly of a multiprotein hormone-receptor complex on thyroid hormone response elements (TREs) near target genes, increasing their transcription (see Fig. 24.8). Thyroid hormone has diverse cellular effects, including the stimulation of carbohydrate and lipid catabolism and protein synthesis in a wide range of cells. The net result is an increase in the basal metabolic rate. In addition, thyroid hormone has a critical role in brain development in the fetus and neonate (see later).

The function of the thyroid gland can be inhibited by a variety of chemical agents, collectively referred to as *goitrogens*. Because they suppress T_3 and T_4 synthesis, the level of TSH increases, causing subsequent hyperplastic enlargement of the gland *(goiter)*. The antithyroid agent propylthiouracil inhibits the oxidation of iodide and thus blocks the production of thyroid hormones; propylthiouracil also inhibits the peripheral deiodination of circulating T_4 into T_3, thus ameliorating symptoms of thyroid hormone excess (see later). Iodide, when given in large doses to individuals with thyroid hyperfunction, blocks the release of thyroid hormones by inhibiting the proteolysis of thyroglobulin. Thus, thyroid hormone is synthesized and incorporated into colloid, but it is not released into the blood.

The thyroid gland follicles also contain a population of parafollicular cells, or C cells, which synthesize and secrete the hormone calcitonin. This hormone promotes the absorption of calcium by the skeletal system and inhibits the resorption of bone by osteoclasts.

Diseases of the thyroid include conditions associated with excessive release of thyroid hormones (hyperthyroidism), thyroid hormone deficiency (hypothyroidism), and mass lesions of the thyroid. We will first consider the clinical consequences of disturbed thyroid function, and then turn to the disorders that generate these problems.

HYPERTHYROIDISM

Thyrotoxicosis is a hypermetabolic state caused by elevated circulating levels of free T_3 and T_4. Because it is caused

Table 24.3 Disorders Associated With Thyrotoxicosis

Associated With Hyperthyroidism
Primary
Diffuse hyperplasia (Graves disease)
Hyperfunctioning ("toxic") multinodular goiter
Hyperfunctioning ("toxic") adenoma
Iodine-induced hyperthyroidism
Neonatal thyrotoxicosis associated with maternal Graves disease
Secondary
TSH-secreting pituitary adenoma (rare)[a]
Not Associated With Hyperthyroidism
Granulomatous (de Quervain) thyroiditis (painful)
Subacute lymphocytic thyroiditis (painless)
Struma ovarii (ovarian teratoma with ectopic thyroid)
Factitious thyrotoxicosis (exogenous thyroxine intake)

[a]Associated with increased thyroid-stimulating hormone (TSH); all other causes of thyrotoxicosis associated with decreased TSH.

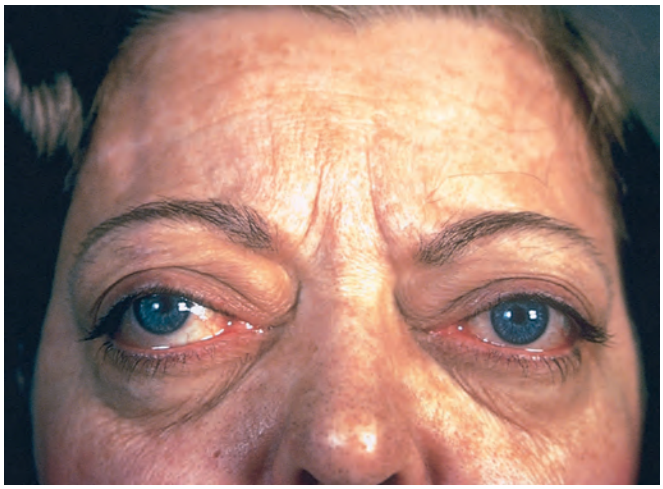


Figure 24.9 A person with hyperthyroidism. A wide-eyed, staring gaze, caused by overactivity of the sympathetic nervous system, is one feature of this disorder. In Graves disease, an important cause of hyperthyroidism, accumulation of loose connective tissue behind the orbit also adds to the protuberant appearance of the eyes.

most commonly by hyperfunction of the thyroid gland, it is often referred to as *hyperthyroidism*. However, in certain conditions the oversupply is related to either excessive release of preformed thyroid hormone (e.g., in thyroiditis) or to an extrathyroidal source, rather than hyperfunction of the gland (Table 24.3). Thus, strictly speaking, hyperthyroidism is only one (albeit the most common) cause of thyrotoxicosis. The terms *primary* and *secondary hyperthyroidism* are sometimes used to designate hyperthyroidism arising from an intrinsic thyroid abnormality and that arising from processes outside of the thyroid, such as a TSH-secreting pituitary tumor. With this caveat, we follow the common practice of using the terms thyrotoxicosis and hyperthyroidism interchangeably.

The most common causes of thyrotoxicosis are associated with hyperfunction of the gland and include the following:

- *Diffuse hyperplasia* of the thyroid associated with Graves disease (approximately 85% of cases)
- Hyperfunctional *multinodular goiter*
- Hyperfunctional thyroid *adenoma*

Clinical Features

The manifestations of hyperthyroidism are protean and include changes referable to the hypermetabolic state induced by excess thyroid hormone and to overactivity of the sympathetic nervous system (i.e., an increase in the β-adrenergic "tone").

- Excessive levels of thyroid hormone result in *an increase in the basal metabolic rate*. The skin of thyrotoxic patients tends to be soft, warm, and flushed because of increased blood flow and peripheral vasodilation, adaptations that serve to increase heat loss. Heat intolerance is common. Sweating is increased because of higher levels of calorigenesis. Heightened catabolic metabolism results in weight loss despite increased appetite.
- *Cardiac manifestations* are among the earliest and most consistent features. Individuals with hyperthyroidism can have elevated cardiac contractility and output, a response to increased peripheral oxygen requirements. Tachycardia, palpitations, and cardiomegaly are common. Arrhythmias such as atrial fibrillation occur frequently, particularly in older patients. Congestive heart failure may develop, especially in older adults with preexisting cardiac disease. Myocardial changes, including focal lymphocytic and eosinophilic infiltrates, mild fibrosis, myofibril fatty change, and an increase in size and number of mitochondria, have been described. Some individuals with thyrotoxicosis develop reversible left ventricular dysfunction and "low-output" heart failure, so-called *thyrotoxic* or *hyperthyroid cardiomyopathy*.
- *Overactivity of the sympathetic nervous system* produces tremor, hyperactivity, emotional lability, anxiety, inability to concentrate, and insomnia. Proximal muscle weakness and decreased muscle mass are common *(thyroid myopathy)*. In the gastrointestinal system, hypermotility and reduced transit times lead to hyperdefecation (not to be confused with diarrhea), which typically responds to beta-blockers, suggesting sympathetic hyperstimulation is the likely underlying pathophysiology. Most patients also develop some degree of fat malabsorption.
- *Ocular changes* often call attention to hyperthyroidism. A wide, staring gaze and lid lag are present because of sympathetic overstimulation of the superior tarsal muscle (also known as Müller's muscle), which functions alongside the levator palpebrae superioris muscle to raise the upper eyelid (Fig. 24.9). However, true *thyroid ophthalmopathy* associated with proptosis occurs only in Graves disease (see later).
- The *skeletal system* is also affected. Thyroid hormone stimulates bone resorption, increasing cortical bone porosity and reducing trabecular bone volume. The net effect is osteoporosis and an increased risk of fractures. Other findings include atrophy of skeletal muscle, with fatty infiltration and focal interstitial lymphocytic infiltrates; hepatic fatty change (steatosis) leading to mild hepatomegaly; and generalized lymphoid hyperplasia and lymphadenopathy in patients with Graves disease.
- *Thyroid storm* refers to the abrupt onset of severe hyperthyroidism. This condition occurs most commonly in patients with Graves disease and probably results from an acute elevation in catecholamine levels, as might be

encountered during infection, surgery, cessation of antithyroid medication, or any form of stress. Patients are often febrile and present with tachycardia out of proportion to the fever. Thyroid storm is a medical emergency. A significant number of untreated patients die of cardiac arrhythmias.

- *Apathetic hyperthyroidism* refers to thyrotoxicosis occurring in older adults, in whom advanced age and various comorbidities blunt the typical features of thyroid hormone excess. The diagnosis in these individuals is often made during laboratory workup for unexplained weight loss or worsening cardiovascular disease.

Laboratory findings that support the diagnosis of hyperthyroidism include a low TSH value accompanied by an increase in free T_4. The measurement of serum TSH concentration is the most useful single screening test for hyperthyroidism, because its levels are decreased even at the earliest stages, when the disease may still be subclinical. In occasional patients, hyperthyroidism results predominantly from increased circulating levels of T_3 ("T_3 toxicosis"). In these cases, free T_4 levels may be decreased, and direct measurement of serum T_3 is useful. In rare cases of pituitary-associated (secondary) hyperthyroidism, TSH levels are either normal or raised. Determining TSH levels after the injection of thyrotropin-releasing hormone (TRH stimulation test) is used in the evaluation of cases of suspected hyperthyroidism with equivocal changes in the baseline serum TSH level. A normal rise in TSH after administration of TRH excludes secondary hyperthyroidism. Once the diagnosis of thyrotoxicosis has been confirmed by a combination of TSH assays and free thyroid hormone levels, measurement of radioactive iodine uptake by the thyroid gland can help determine the etiology. For example, there may be diffusely increased uptake in the gland (Graves disease), increased uptake in a solitary nodule (toxic adenoma), or decreased uptake (thyroiditis).

The therapeutic options for hyperthyroidism include several medications, each with a different mechanism of action. Typically, these include a β-blocker to control symptoms induced by increased adrenergic tone, a thionamide (e.g., methimazole or propylthiouracil) to block new hormone synthesis, an iodine solution to block the release of thyroid hormone, and agents that inhibit peripheral conversion of T_4 to T_3. Radioactive iodine, which is incorporated into thyroid tissues, resulting in ablation of thyroid function over a period of 6 to 18 weeks, may also be used.

HYPOTHYROIDISM

Hypothyroidism is a condition caused by a structural or functional derangement that interferes with the production of thyroid hormone. Hypothyroidism is a common disorder. By some estimates, the prevalence of overt hypothyroidism is 0.3%, while that of subclinical hypothyroidism is greater than 4%. The prevalence increases with age, and it is nearly 10-fold more common in women than in men. Hypothyroidism can result from a defect anywhere in the hypothalamic-pituitary-thyroid axis. Like hyperthyroidism, it can be divided into *primary* and *secondary* forms, depending on whether it arises from an intrinsic thyroid abnormality or as a result of pituitary and hypothalamic disease (Table 24.4). Primary hypothyroidism accounts for the vast majority of cases and may be accompanied by enlargement of the thyroid gland (goiter).

Table 24.4 Causes of Hypothyroidism

Primary
Genetic defects in thyroid development (*PAX8*, *FOXE1*, TSH receptor mutations; rare)
Thyroid hormone resistance syndrome (*THRB* mutations; rare)
Postablative
Surgery, radioiodine therapy, or external irradiation
Autoimmune hypothyroidism
Hashimoto thyroiditis[a]
Iodine deficiency[a]
Drugs (lithium, iodides, p-aminosalicylic acid)[a]
Congenital biosynthetic defect (dyshormonogenetic goiter; rare)[a]
Secondary ("Central Hypothyroidism")
Pituitary failure (rare)
Hypothalamic failure (rare)

[a]Associated with enlargement of thyroid ("goitrous hypothyroidism"). Hashimoto thyroiditis and postablative hypothyroidism account for the majority of cases of hypothyroidism in high-income countries. *FOXE1*, Forkhead box E1; *PAX8*, paired box 8; *THRB*, thyroid hormone receptor β.

Primary hypothyroidism can be congenital, autoimmune, or iatrogenic.

- *Congenital hypothyroidism.* **Worldwide, congenital hypothyroidism is most often the result of endemic iodine deficiency in the diet** (see later). Other rare forms of congenital hypothyroidism include inborn errors of thyroid metabolism *(dyshormonogenetic goiter),* wherein one of the several steps leading to thyroid hormone synthesis is defective, such as (1) iodide transport into thyrocytes, (2) "organification" of iodine (covalent binding of iodine to tyrosine residues in thyroglobulin), and (3) further processing to form hormonally active T_3 and T_4. In rare instances, there may be complete absence of thyroid parenchyma *(thyroid agenesis),* or the gland may be greatly reduced in size *(thyroid hypoplasia)* due to germline mutations in genes responsible for thyroid development (see Table 24.4).
- *Autoimmune hypothyroidism.* **Autoimmune hypothyroidism is the most common cause of hypothyroidism in iodine-sufficient areas of the world.** The vast majority of cases of autoimmune hypothyroidism are due to Hashimoto thyroiditis. Circulating autoantibodies, including antimicrosomal, antithyroid peroxidase, and antithyroglobulin antibodies, are found in this disorder, and the thyroid is typically enlarged (goitrous). Autoimmune hypothyroidism can occur in isolation or in conjunction with autoimmune polyendocrine syndrome types 1 and 2 (APS-1 and APS-2; see discussion in the Adrenal Glands section later in this chapter).
- *Iatrogenic hypothyroidism.* This can be caused by either surgical or radiation-induced ablation. Resection of a large portion of the gland for the treatment of hyperthyroidism or a primary neoplasm can lead to hypothyroidism. The gland may also be ablated by radiation, whether in the form of radioiodine administered for the treatment of hyperthyroidism, or exogenous irradiation, such as external radiation therapy to the neck. Drugs given to decrease thyroid secretion (e.g., thionamides like methimazole and

propylthiouracil) can also cause acquired hypothyroidism, as can agents used to treat nonthyroid conditions (e.g., lithium, *p*-aminosalicylic acid).

Secondary (or central) hypothyroidism is caused by deficiencies of TSH or, far more uncommonly, TRH. Any of the causes of hypopituitarism (e.g., pituitary tumor, postpartum pituitary necrosis, trauma, and nonpituitary tumors) or hypothalamic damage (e.g., tumors, trauma, radiation therapy, or infiltrative diseases) can cause central hypothyroidism.

Cretinism

Cretinism refers to hypothyroidism that develops in infancy or early childhood. The term *cretin* was derived from the French *chrétien,* meaning "Christian" or "Christlike," and was applied to those affected because they were felt to be too mentally challenged to be capable of sinning. In the past, this disorder occurred fairly commonly in regions of the world where dietary iodine deficiency is endemic, such as the Himalayas, inland China, Africa, and other mountainous areas. It is now much less prevalent as a result of the widespread supplementation of foods with iodine. On rare occasions, cretinism results from genetic defects that interfere with the biosynthesis of thyroid hormone (dyshormonogenetic goiter).

Clinical features of cretinism include severe intellectual disability, short stature, coarse facial features, a protruding tongue, and umbilical hernia. The severity of the mental impairment seems to be related to the time at which thyroid deficiency occurs in utero. Normally, maternal T_3 and T_4 cross the placenta and are critical for fetal brain development. If there is maternal thyroid deficiency before the development of the fetal thyroid gland, intellectual disability is severe. In contrast, maternal thyroid hormone deficiency later in pregnancy, after the fetal thyroid has become functional, does not affect normal brain development.

Myxedema

The term myxedema is applied to hypothyroidism developing in the older child or adult. Myxedema was first linked with thyroid dysfunction in 1873 by Sir William Gull in an article addressing the development of a "cretinoid state" in adults. The clinical findings vary with age of onset. Older children show signs and symptoms intermediate between those of the cretin and those of the adult with hypothyroidism. In the adult, the condition appears insidiously and may take years before arousing clinical suspicion.

Myxedema is marked by a slowing of physical and mental activity. Early symptoms include generalized fatigue, apathy, and mental sluggishness, which may mimic depression. Speech and intellectual functions are slowed, and patients are listless, cold intolerant, and frequently overweight. Decreased sympathetic activity results in constipation and decreased sweating. The skin is cool and pale because of decreased blood flow. Reduced cardiac output probably contributes to shortness of breath and decreased exercise capacity, two frequent complaints. Thyroid hormones regulate the transcription of several sarcolemmal genes, such as calcium ATPases and the β-adrenergic receptor, and diminished expression of these genes results in a decrease in cardiac output. In addition, changes in lipid metabolism produce an increase in total cholesterol and low-density lipoprotein (LDL) levels that is pro-atherogenic, contributing to increased cardiovascular mortality. Histologically, there is an accumulation of matrix substances, such as glycosaminoglycans and hyaluronic acid, in skin, subcutaneous tissue, and visceral sites. This results in nonpitting edema, a broadening and coarsening of facial features, enlargement of the tongue, and deepening of the voice.

Laboratory evaluation plays a vital role in the diagnosis of suspected hypothyroidism because of the nonspecific nature of symptoms. Patients with unexplained increases in body weight or hypercholesterolemia should be assessed for potential hypothyroidism. **Measurement of the serum TSH level is the most sensitive screening test for this disorder.** The TSH level is increased in primary hypothyroidism as a result of a loss of feedback inhibition of TRH and TSH production by the hypothalamus and pituitary, respectively. The TSH level is not increased in persons with hypothyroidism due to primary hypothalamic or pituitary disease. T_4 levels are decreased in individuals with hypothyroidism of any origin.

THYROIDITIS

Thyroiditis encompasses a diverse group of disorders characterized by some form of thyroid inflammation. This discussion focuses on the three most common and clinically significant subtypes: (1) Hashimoto thyroiditis, (2) granulomatous (de Quervain) thyroiditis, and (3) subacute lymphocytic thyroiditis.

Hashimoto Thyroiditis

Hashimoto thyroiditis is an autoimmune disease that results in destruction of the thyroid gland and gradual and progressive thyroid failure. It is the most common cause of hypothyroidism in areas of the world where iodine levels are sufficient. The name is derived from the 1912 report by Hashimoto describing patients with goiter and lymphocytic infiltration of the thyroid *(struma lymphomatosa)*. It is most prevalent between 45 and 65 years of age and is more common in women than in men, with a female predominance of 10:1 to 20:1. It can also occur in children and is a major cause of nonendemic goiter in the pediatric population.

Pathogenesis

Hashimoto thyroiditis is caused by a breakdown in self-tolerance to thyroid autoantigens. This is exemplified by the presence of circulating autoantibodies against thyroglobulin and thyroid peroxidase in the vast majority of patients. The inciting events have not been elucidated, but possibilities include abnormalities of regulatory T cells (Tregs), or exposure of normally sequestered thyroid antigens (Chapter 6). Similar to other autoimmune diseases, Hashimoto thyroiditis predisposition has a strong genetic component. Increased susceptibility is associated with polymorphisms in several immune regulation–associated genes, such as cytotoxic T lymphocyte–associated antigen-4

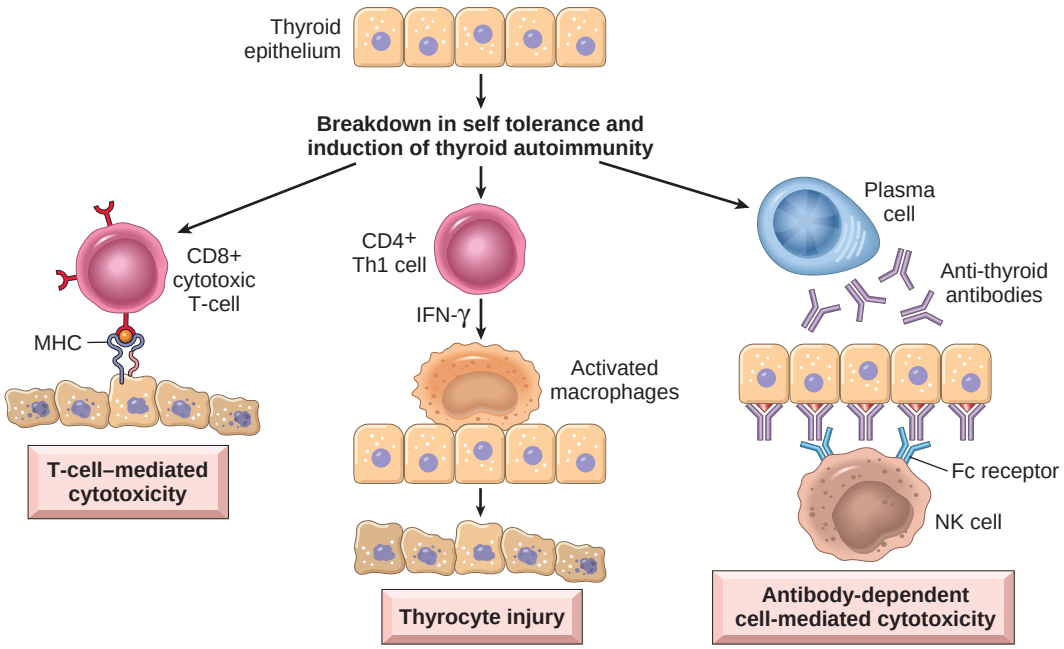


Figure 24.10 Pathogenesis of Hashimoto thyroiditis. Breakdown of peripheral tolerance to thyroid autoantigens results in progressive autoimmune destruction of thyroid cells by infiltrating cytotoxic T cells, locally released cytokines, or antibody-dependent cytotoxicity.

(CTLA4), protein tyrosine phosphatase-22 *(PTPN22),* and interleukin-2 receptor α chain *(IL2RA),* all of which encode regulators of T-cell responses. Not surprisingly, genome-wide association studies (GWAS) have found similar associations between *CTLA4, PTPN22,* and *IL2RA* polymorphisms and Graves disease, another type of autoimmune thyroid disease (discussed later), as well as with other systemic autoimmune diseases such as type 1 diabetes (T1D). The genetic associations support the idea that breakdown in immune tolerance is a common pathophysiologic theme for many autoimmune diseases, more than one of which can coexist in the same individual.

Induction of thyroid autoimmunity is accompanied by a progressive depletion of thyroid epithelial cells and replacement of the thyroid parenchyma by lymphocytic infiltrates and fibrosis. Multiple immunologic mechanisms may contribute to thyroid cell death, including (Fig. 24.10):

- *CD8+ cytotoxic T cell–mediated cell death:* CD8+ cytotoxic T cells may destroy thyroid follicular cells.
- *Cytokine-mediated cell death:* Activation of CD4+ Th1 cells leads to the production of inflammatory cytokines such as interferon-γ in the thyroid gland, with resultant recruitment and activation of macrophages and damage to follicles.
- A less likely mechanism involves binding of antithyroid antibodies (antithyroglobulin, and antithyroid peroxidase antibodies) followed by antibody-dependent cell-mediated cytotoxicity (Chapter 6).

MORPHOLOGY

The thyroid is usually diffusely enlarged. The capsule is intact, and the gland is well demarcated from adjacent structures. The cut surface is pale, yellow-tan, firm, and somewhat nodular. There is extensive infiltration of the parenchyma by a **mononuclear inflammatory infiltrate** containing small lymphocytes, plasma cells, and well-developed **germinal centers** (Fig. 24.11). The thyroid follicles are atrophic and are lined in many areas by epithelial cells with abundant eosinophilic, granular cytoplasm, termed **Hürthle cells**, which represent a metaplastic response of the normally low cuboidal follicular epithelium to chronic injury. In fine-needle aspiration biopsy samples, the presence of Hürthle cells in conjunction with a heterogeneous population of lymphocytes is characteristic of Hashimoto thyroiditis. In "classic" Hashimoto thyroiditis, interstitial connective tissue is increased and may be abundant. Unlike Reidel thyroiditis (see later), the fibrosis does not extend beyond the capsule of the gland.

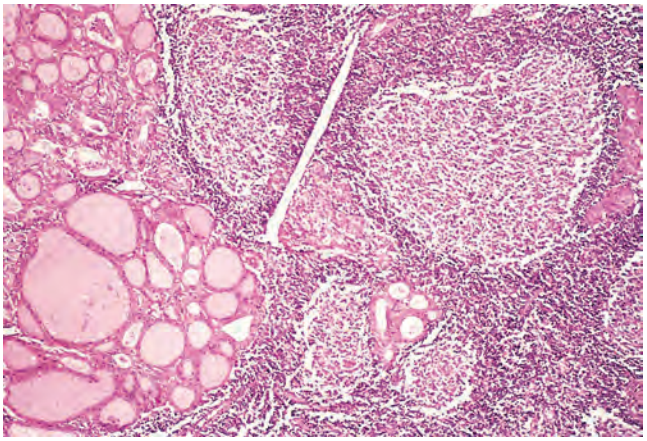


Figure 24.11 Hashimoto thyroiditis. The thyroid parenchyma contains a dense lymphocytic infiltrate with germinal centers. Residual thyroid follicles lined by deeply eosinophilic Hürthle cells are also seen.

Clinical Features

Hashimoto thyroiditis most often comes to attention as painless enlargement of the thyroid, usually associated with some degree of hypothyroidism, in a middle-aged woman. The enlargement of the gland is generally symmetric and diffuse, but in some cases it may be sufficiently localized to raise the suspicion of a neoplasm. In the typical case, hypothyroidism develops gradually. In some patients, however, it may be preceded by transient thyrotoxicosis caused by disruption of thyroid follicles and release of thyroid hormones *(hashitoxicosis)*. During this phase, free T_4 and T_3 levels are elevated, TSH is diminished, and radioactive iodine uptake is decreased. As hypothyroidism supervenes, T_4 and T_3 levels fall, accompanied by a compensatory increase in TSH.

Individuals with Hashimoto thyroiditis are at increased risk for developing other autoimmune diseases, both endocrine (type 1 diabetes, autoimmune adrenalitis) and non-endocrine (systemic lupus erythematosus, myasthenia gravis, and Sjögren syndrome; Chapter 6). They are also at increased risk for development of extranodal marginal zone B-cell lymphoma within the thyroid gland (Chapter 13). The relationship between Hashimoto disease and thyroid epithelial cancers remains controversial, with some morphologic and molecular studies suggesting a predisposition to papillary carcinoma.

Subacute Lymphocytic (Painless) Thyroiditis

Subacute lymphocytic thyroiditis, which is also referred to as *painless thyroiditis,* is a presumed autoimmune disease. Most patients have circulating antithyroid peroxidase antibodies or a family history of other autoimmune disorders. A similar disease process can occur during the postpartum period in up to 5% of women *(postpartum thyroiditis)*.

MORPHOLOGY

Except for possible mild symmetric enlargement, the thyroid appears grossly normal. Microscopic examination reveals lymphocytic infiltration with large germinal centers within the thyroid parenchyma and patchy disruption and collapse of thyroid follicles. Unlike Hashimoto thyroiditis, however, fibrosis and Hürthle cell metaplasia are not prominent.

Clinical Features

Painless thyroiditis usually comes to attention because of mild, transient hyperthyroidism, painless goiter, or both. Although it can occur at any age, it is most often seen in middle-aged adults and is more common in women. Some patients transition from hyperthyroidism to hypothyroidism before recovery. As many as one-third of cases evolve into overt hypothyroidism over a 10-year period, with pathogic features that resemble Hashimoto thyroiditis.

Granulomatous Thyroiditis

Granulomatous thyroiditis (also called *De Quervain thyroiditis*) occurs much less frequently than does Hashimoto thyroiditis. The disorder is most common between 40 and 50 years of age and, like other forms of thyroiditis, affects women more often than men (4:1).

Pathogenesis

Granulomatous thyroiditis is believed to be triggered by a viral infection. The majority of patients have a history of an upper respiratory infection just before the onset of thyroiditis. The disease is seasonal, with occurrences peaking in the summer, and clusters of cases have been reported in association with coxsackievirus, mumps, measles, adenovirus, and other viral infections. Although the pathogenesis of the disease is unclear, one model suggests that it results from virus-induced host tissue damage that stimulates a cytotoxic T-lymphocyte response to one or more thyroid antigens that damages thyroid follicle cells. In contrast to autoimmune thyroid disease, the immune response is virus-initiated and not self-perpetuating, so the process is limited.

MORPHOLOGY

The gland may be unilaterally or bilaterally enlarged and firm, with an intact capsule that may adhere to surrounding structures. On cut section, the involved areas are firm and yellow-white and stand out from adjacent normal rubbery, brown thyroid gland. Histologic changes are patchy and depend on the stage of the disease. Early in the active inflammatory phase, scattered follicles may be disrupted and replaced by neutrophils forming microabscesses. Later, more characteristic features appear in the form of aggregates of lymphocytes, activated macrophages, and plasma cells associated with collapsed and damaged thyroid follicles. **Multinucleate giant cells** enclose pools of colloid (Fig. 24.12), hence the designation **granulomatous thyroiditis.** In later stages of the disease, a chronic inflammatory infiltrate and fibrosis may replace the foci of injury. Different histologic stages are sometimes found in the same gland, suggesting waves of destruction over a period of time.

Clinical Features

Granulomatous thyroiditis is the most common cause of thyroid pain. There is variable enlargement of the gland. Inflammation of the thyroid and hyperthyroidism are

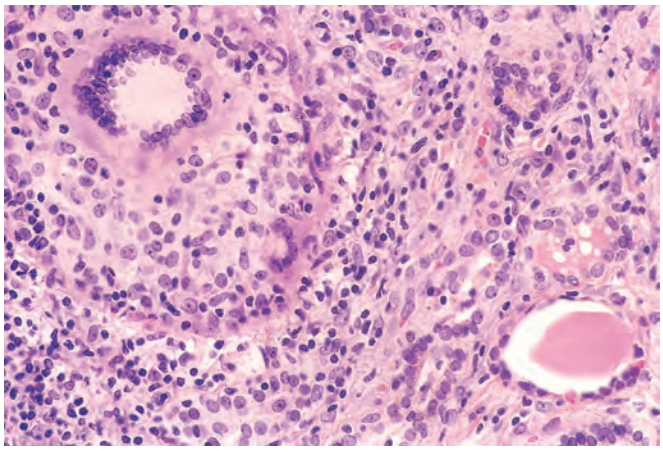


Figure 24.12 Granulomatous thyroiditis. The thyroid parenchyma contains a chronic inflammatory infiltrate with a multinucleate giant cell *(top left)* and a colloid follicle *(bottom right)*.

transient, usually diminishing in 2 to 6 weeks, even if the patient is not treated. Nearly all patients have high serum T_4 and T_3 levels and low serum TSH levels during this phase. However, unlike in hyperthyroid states such as Graves disease, radioactive iodine uptake is diminished. After recovery, generally in 6 to 8 weeks, normal thyroid function returns.

Other, less common forms of thyroiditis include *Riedel thyroiditis,* a rare disorder characterized by extensive fibrosis involving the thyroid and contiguous neck structures. The presence of a hard and fixed thyroid mass clinically simulates a thyroid carcinoma. It may be associated with fibrosis in other sites in the body, such as the retroperitoneum, and appears to be another manifestation of IgG4-related disease, which is associated with fibrosis and tissue infiltration by plasma cells producing IgG4 (Chapter 6).

KEY CONCEPTS

THYROIDITIS

- Hashimoto thyroiditis is the most common cause of hypothyroidism in regions where dietary iodine levels are sufficient.
- Hashimoto thyroiditis is an autoimmune thyroiditis characterized by progressive destruction of thyroid parenchyma, Hürthle cell change, and mononuclear (lymphoplasmacytic) infiltrates, with germinal centers and variable degrees of fibrosis.
- Subacute lymphocytic thyroiditis often occurs after pregnancy (postpartum thyroiditis), typically is painless, and is characterized by lymphocytic inflammation in the thyroid. It is another type of autoimmune thyroiditis.
- Granulomatous (de Quervain) thyroiditis is a self-limited disease, probably secondary to a viral infection, and is characterized by pain and granulomatous inflammation of the thyroid.

GRAVES DISEASE

Graves disease is the most common cause of endogenous hyperthyroidism. Graves reported in 1835 his observations of a disease characterized by "violent and long continued palpitations in females" associated with enlargement of the thyroid gland. The disease is characterized by a triad of clinical findings:

- *Hyperthyroidism* associated with diffuse enlargement of the gland
- Infiltrative *ophthalmopathy* and resultant exophthalmos
- Localized, infiltrative *dermopathy,* sometimes called *pretibial myxedema,* which is present in a minority of patients

Graves disease has a peak incidence between 20 and 40 years of age. Women are affected as much as 10 times more frequently than men. This disorder is said to affect 1.5% to 2% of women in the United States.

Pathogenesis

Graves disease is an autoimmune disorder characterized by the production of autoantibodies against multiple thyroid proteins, most importantly the TSH receptor. A variety of antibodies that either stimulate or block the TSH receptor are detected in the circulation. The most common antibody subtype, known as *thyroid-stimulating immunoglobulin* (TSI), is observed in approximately 90% of patients. In contrast to antibodies against thyroglobulin and thyroid peroxidase, TSI is almost never observed in other autoimmune diseases of the thyroid. TSI binds to the TSH receptor and mimics its actions, stimulating adenyl cyclase and increasing the release of thyroid hormones. As stated earlier, TSH receptor blocking antibodies may also be present, and in a minority of patients these lead to hypothyroidism.

Perhaps unsurprisingly, factors that predispose to Graves disease overlap with risk factors for Hashimoto thyroiditis, the other major form of autoimmune thyroid disease. The concordance rate in monozygotic twins is 30% to 40%, compared with less than 5% among dizygotic twins, and similar to Hashimoto thyroiditis, genetic susceptibility is associated with polymorphisms in immune-function genes like *CTLA4, PTPN22,* and *IL2RA*. In addition, GWAS data has also implicated variants in the TSH receptor *(TSHR)* gene locus in susceptibility to Graves disease.

Autoimmunity also plays a role in the development of the *infiltrative ophthalmopathy* that is characteristic of Graves disease. Activated CD4+ helper T cells secrete cytokines that stimulate fibroblast proliferation and synthesis of extracellular matrix proteins (glycosaminoglycans), leading to progressive infiltration of the retro-orbital space and ophthalmopathy. The protrusion of the eyeball (exopthalmos) is caused by an increase in the volume of the retro-orbital connective tissues and extraocular muscles, which occurs for several reasons: (1) marked infiltration of connective tissue by mononuclear cells, predominantly T cells; (2) inflammation, edema, and swelling of extraocular muscles; (3) accumulation of extracellular matrix components, specifically hydrophilic glycosaminoglycans such as hyaluronic acid and chondroitin sulfate; and (4) increased numbers of adipocytes (fatty infiltration). These changes not only displace the eyeball but also may interfere with the function of the extraocular muscles.

MORPHOLOGY

The thyroid gland is usually symmetrically enlarged due to **diffuse hypertrophy and hyperplasia** of thyroid follicular epithelial cells (Fig. 24.13A). Increases in weight to over 80 g are not uncommon. On cut section, the parenchyma has a soft, meaty appearance resembling muscle. Histologically, the follicle epithelial cells in untreated cases are tall and more crowded than usual. This crowding often results in the formation of small papillae that project into the follicle lumen and encroach on the colloid, sometimes filling the follicles (Fig. 24.13B). Such papillae lack fibrovascular cores, in contrast to those of papillary carcinoma (see later). The colloid within the follicle lumen is pale, with scalloped margins. Lymphoid infiltrates, consisting predominantly of T cells, along with scattered B cells and mature plasma cells, are present throughout the interstitium. Germinal centers are common.

Preoperative therapy alters the morphology of the thyroid in Graves disease. Administration of iodine causes involution of the epithelium and the accumulation of colloid by blocking thyroglobulin secretion. Treatment with the antithyroid drug propylthiouracil exaggerates the epithelial hypertrophy and hyperplasia by stimulating TSH secretion.

Changes in extrathyroidal tissue include lymphoid hyperplasia, especially enlargement of the thymus in younger patients. The

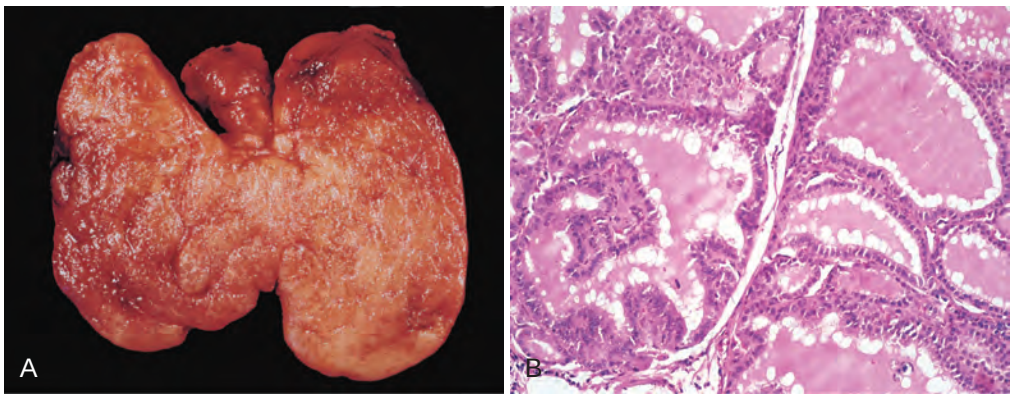


Figure 24.13 Graves disease. (A) There is diffuse symmetric enlargement of the gland and a beefy deep red parenchyma. Compare with gross photograph of multinodular goiter in Fig. 24.15. (B) Diffusely hyperplastic thyroid in a case of Graves disease. The follicles are lined by tall, columnar epithelium. The crowded, enlarged epithelial cells project into the lumens of the follicles. These cells actively resorb the colloid in the centers of the follicles, resulting in the scalloped appearance of the edges of the colloid. (A, Reproduced with permission from Lloyd RV, Douglas BR, Young WF Jr, editors: *Atlas of Nontumor Pathology: Endocrine Diseases*, Washington, DC, 2002, American Registry of Pathology.)

heart may be hypertrophied, and ischemic changes may be present, particularly in patients with preexisting coronary artery disease. In patients with ophthalmopathy, the tissues of the orbit are edematous because of the presence of hydrophilic mucopolysaccharides. In addition, there is infiltration by lymphocytes and fibrosis. Orbital muscles are edematous initially but may undergo fibrosis late in the course of the disease. The dermopathy, if present, is characterized by thickening of the dermis due to deposition of glycosaminoglycans and lymphocyte infiltration.

Clinical Features

The findings in Graves disease include changes associated with thyrotoxicosis and others associated uniquely with Graves disease, such as diffuse hyperplasia of the thyroid, ophthalmopathy, and dermopathy. The degree of thyrotoxicosis varies from case to case and is sometimes less conspicuous than other manifestations of the disease. Diffuse enlargement of the thyroid is present in all cases. The thyroid enlargement may be accompanied by increased flow of blood through the hyperactive gland, often producing an audible "bruit." Sympathetic overactivity produces a characteristic wide, staring gaze and lid lag, while the ophthalmopathy causes exophthalmos. The exophthalmos may persist or progress despite successful treatment of the thyrotoxicosis, sometimes resulting in corneal injury. The extraocular muscles are often weak. The infiltrative dermopathy *(pretibial myxedema)* is most common in the skin overlying the shins, where it presents as scaly thickening and induration. The basis of such localization is not clear, and it is present only in a minority of patients. Sometimes individuals spontaneously develop thyroid hypofunction. Patients are at increased risk for other autoimmune diseases, such as systemic lupus erythematosus, pernicious anemia, type 1 diabetes, and Addison disease.

Laboratory findings in Graves disease include elevated free T_4 and T_3 levels and depressed TSH levels. Because of ongoing stimulation by thyroid-stimulating immunoglobulins, radioiodine scans show a diffusely increased uptake of iodine in the thyroid gland.

Graves disease is treated with β-blockers, which dampen symptoms related to increased sympathetic nervous system activity (e.g., tachycardia, palpitations, tremulousness, and anxiety), and by measures that decrease thyroid hormone synthesis, such as the administration of thionamides (e.g., propylthiouracil), radioiodine ablation, and thyroidectomy. Surgery is used mostly in patients who have large goiters that are compressing surrounding structures.

KEY CONCEPTS

GRAVES DISEASE

- Graves disease, the most common cause of endogenous hyperthyroidism, is characterized by the triad of thyrotoxicosis, ophthalmopathy, and dermopathy.
- Graves disease is an autoimmune disorder caused by activation of thyroid epithelial cells by autoantibodies to the TSH receptor that mimic TSH action (thyroid-stimulating immunoglobulins).
- The thyroid in Graves disease shows diffuse hypertrophy and hyperplasia of follicles and lymphoid infiltrates; glycosaminoglycan deposition and lymphoid infiltrates are responsible for the ophthalmopathy and dermopathy.
- Laboratory features include elevations in serum free T_3 and T_4 and decreased serum TSH.

DIFFUSE AND MULTINODULAR GOITER

Enlargement of the thyroid, or goiter, is caused by impaired synthesis of thyroid hormone, which is most often the result of dietary iodine deficiency. Reduced thyroid hormone production leads to a compensatory rise in the serum TSH level, which, in turn, causes hypertrophy and hyperplasia of thyroid follicular cells and, ultimately, enlargement of the thyroid gland. This increase in functional mass of the gland usually overcomes the hormone deficiency, ensuring a euthyroid metabolic state in most individuals. If the underlying disorder is sufficiently severe

(e.g., a congenital biosynthetic defect or endemic iodine deficiency, discussed later), the compensatory responses may be inadequate, resulting in *goitrous hypothyroidism*. The degree of thyroid enlargement is proportional to the level and duration of thyroid hormone deficiency. Goiters can broadly be divided into two types: *diffuse nontoxic* and *multinodular*.

Diffuse Nontoxic (Simple) Goiter

Diffuse nontoxic (simple) goiter causes enlargement of the entire gland without producing nodularity. Because the enlarged follicles are filled with colloid, the term *colloid goiter* has been applied to this condition. This disorder occurs in both an endemic and a sporadic distribution.

- *Endemic goiter* occurs in geographic areas where the soil, water, and food supply contain low levels of iodine. The term endemic is used when goiters are present in more than 10% of the population in a given region. Such conditions are particularly common in mountainous areas of the world, including the Andes and Himalayas, where iodine deficiency is widespread. The lack of iodine leads to decreased synthesis of thyroid hormone and a compensatory increase in TSH, leading to follicular cell hypertrophy and hyperplasia and goitrous enlargement. With increasing dietary iodine supplementation, the frequency and severity of endemic goiter have declined significantly, although as many as 200 million people worldwide continue to be at risk for severe iodine deficiency.

  Variations in the prevalence of endemic goiter in regions with similar levels of iodine deficiency point to the existence of other causative influences, particularly dietary substances, referred to as *goitrogens*. The ingestion of substances that interfere with thyroid hormone synthesis at some level, such as vegetables belonging to the *Brassicaceae (Cruciferae)* family (e.g., cabbage, cauliflower, Brussels sprouts, turnips, and cassava), has been documented to be goitrogenic. Native populations subsisting on cassava root are particularly at risk. Cassava contains a thiocyanate that inhibits iodide transport within the thyroid, worsening any concurrent iodine deficiency.
- *Sporadic goiter* occurs less frequently than does endemic goiter. There is a striking female preponderance and a peak incidence at puberty or in young adult life. Sporadic goiter can be caused by several conditions, including the ingestion of substances that interfere with thyroid hormone synthesis. In other instances, goiter may result from hereditary enzymatic defects that interfere with thyroid hormone synthesis, all transmitted as autosomal recessive conditions (dyshormonogenetic goiter; see earlier). In most cases, however, the cause of sporadic goiter is not apparent.

MORPHOLOGY

Two phases can be identified in the evolution of diffuse nontoxic goiter: the **hyperplastic phase** and the phase of **colloid involution.** In the hyperplastic phase, the thyroid gland is diffusely and symmetrically enlarged, although the increase is usually modest, and the gland rarely exceeds 100 to 150 g. The follicles are lined by crowded columnar cells, which may pile up and form projections similar to those seen in Graves disease. The accumulation is not uniform throughout the gland, and some follicles are hugely distended, whereas others remain small. If dietary iodine subsequently increases or if the demand for thyroid hormone decreases, the stimulated follicle epithelium involutes to form an enlarged, colloid-rich gland **(colloid goiter).** In these cases, the cut surface of the thyroid is usually brown, somewhat glassy, and translucent. Histologically the follicle epithelium is flattened and cuboidal, and colloid is abundant during periods of involution.

Clinical Features

As stated earlier, most persons with simple goiters are clinically euthyroid. Therefore, the clinical manifestations are primarily related to *mass effects* from the enlarged thyroid gland (Fig. 24.14). Although serum T_3 and T_4 levels are normal, the serum TSH is usually elevated or at the upper range of normal, as is expected in marginally euthyroid individuals. In children, dyshormonogenetic goiter, caused by a congenital biosynthetic defect, may induce cretinism.

Multinodular Goiter

With time, recurrent episodes of hyperplasia and involution combine to produce a more irregular enlargement of the thyroid, termed *multinodular goiter*. Virtually all long-standing simple goiters convert into multinodular goiters. **Multinodular goiters produce the most extreme thyroid enlargements and are more frequently mistaken for neoplasms than any other form of thyroid disease.** Because they derive from simple goiter, they occur in both sporadic and endemic

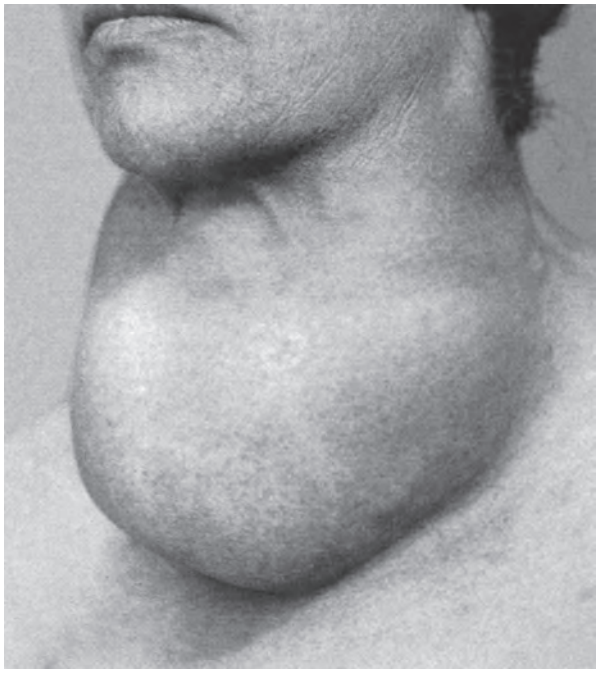


Figure 24.14 A 52-year-old woman with a huge colloid goiter who developed compressive symptoms. (Reproduced with permission from Lloyd RV, Douglas BR, Young WF Jr, editors: *Atlas of Nontumor Pathology: Endocrine Diseases*, Washington, DC, 2002, American Registry of Pathology.)

forms, having the same female-to-male distribution and presumably the same origins but affecting older individuals because they are late complications.

Pathogenesis

It is believed that multinodular goiters arise because of variations among follicular cells in their response to external stimuli, such as trophic hormones. If some cells in a follicle have a growth advantage, perhaps because of acquired genetic abnormalities similar to those that give rise to adenomas, they may begin to proliferate autonomously, producing a nodule. Consistent with this model, both polyclonal and monoclonal nodules coexist within the same multinodular goiter, the latter presumably having arisen because of the acquisition of a genetic abnormality favoring growth. Not surprisingly, activating mutations affecting proteins of the TSH signaling pathway have been identified in a subset of autonomous thyroid nodules (TSH signaling pathway mutations and their implications in adenomas are discussed later). The follicle hyperplasia and accumulation of colloid produce physical stress that may lead to rupture of follicles and vessels followed by hemorrhages, scarring, and calcifications. With scarring, nodularity appears, which may be accentuated by the preexisting stromal framework of the gland.

MORPHOLOGY

Multinodular goiters are multilobulated, asymmetrically enlarged glands that can reach weights of more than 2000 g. The pattern of enlargement is unpredictable and may involve one lobe far more than the other, producing pressure on midline structures, such as the trachea and esophagus. In other instances, the goiter grows behind the sternum and clavicles to produce the **intrathoracic** or **plunging goiter.** Occasionally, most of it is hidden behind the trachea and esophagus; in other instances, one nodule may stand out, mimicking a solitary tumor. On cut section, irregular nodules containing variable amounts of brown, gelatinous colloid are present (Fig. 24.15A). Older lesions have areas of hemorrhage, fibrosis, calcification, and cystic change. The microscopic appearance includes colloid-rich follicles lined by flattened, inactive epithelium and areas of **follicle hyperplasia,** accompanied by degenerative changes related to physical stress. In contrast to follicular neoplasms, a prominent capsule between the hyperplastic nodules and residual compressed thyroid parenchyma is not present (Fig. 24.15B).

Clinical Features

The dominant clinical features of multinodular goiter are those caused by mass effects. In addition to their cosmetic effects, goiters may cause airway obstruction, dysphagia, and compression of large vessels in the neck and upper thorax *(superior vena cava syndrome)*. Most patients are euthyroid or have subclinical hyperthyroidism (identified by reduced TSH levels), but about 10% of patients over a 10-year period develop an autonomous nodule within a long-standing goiter that produces hyperthyroidism *(toxic multinodular goiter)*, also known as *Plummer syndrome*. The incidence of malignancy in long-standing multinodular goiters is low (<5%) but not zero, and concern for malignancy arises in goiters that demonstrate sudden changes in size or symptoms (e.g., hoarseness). Dominant nodules in a multinodular goiter can present as a "solitary thyroid nodule," mimicking a thyroid neoplasm. A radioiodine scan demonstrates uneven iodine uptake (including the occasional "hot" autonomous nodule), consistent with an admixture of hyperplastic and involuting nodules. A fine-needle aspiration biopsy is helpful and can often, albeit not always, distinguish between follicular hyperplasia and thyroid neoplasm (see later).

NEOPLASMS OF THE THYROID

The solitary thyroid nodule is a palpably discrete swelling within an otherwise apparently normal thyroid gland. The estimated incidence of solitary palpable nodules in the adult population of the United States varies between 1% and 10%, but it is significantly higher in endemic goitrous regions. Single nodules are about four times more common in women than in men. The incidence of thyroid nodules increases throughout life.

From a clinical standpoint, the major concern in persons who present with thyroid nodules is the possibility of a malignant neoplasm. Fortunately, the overwhelming majority of solitary nodules of the thyroid prove to be localized, nonneoplastic lesions (e.g., a dominant nodule in multinodular goiter, simple cysts, or foci of thyroiditis) or benign neoplasms such as follicular adenoma. In fact, benign neoplasms outnumber thyroid carcinomas by a ratio of nearly 10:1. While only 1% of solitary thyroid nodules are

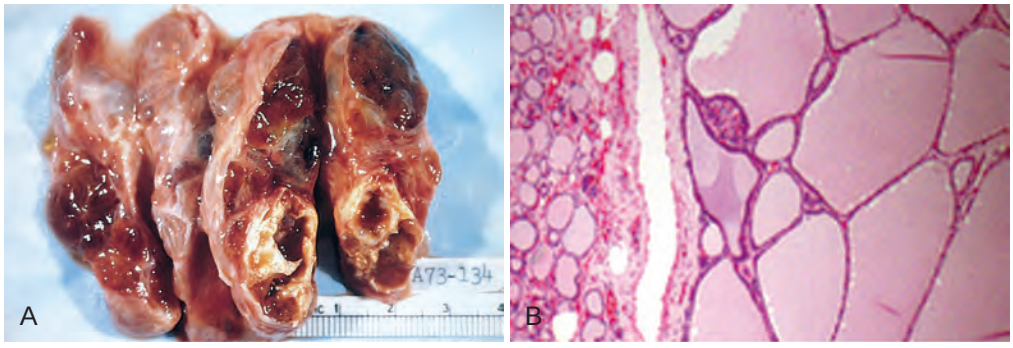


Figure 24.15 Multinodular goiter. (A) Gross morphology demonstrating a coarsely nodular gland containing areas of fibrosis and cystic change. (B) Photomicrograph of a hyperplastic nodule with compressed residual thyroid parenchyma on the periphery. Note the absence of a prominent capsule, a distinguishing feature from follicular neoplasms. (B, Courtesy Dr. William Westra, Department of Pathology, Johns Hopkins University, Baltimore, Md.)

malignant, this still represents about 15,000 new cases of thyroid carcinoma per year in the United States. Fortunately, most of these cancers are indolent; more than 90% of affected patients are alive 20 years after being diagnosed.

Several criteria provide clues to the nature of a thyroid nodule:

- *Solitary nodules, nodules in younger patients, and nodules in males* are more likely to be neoplastic.
- A history of *radiation treatment* to the head and neck region is associated with an increased incidence of thyroid malignancy.
- Functional nodules that take up radioactive iodine in imaging studies *(hot nodules)* are much more likely to be benign.

These associations and statistics, however, are of little comfort to a patient, in whom the timely recognition of a malignancy can be lifesaving. Ultimately, morphologic evaluation of a given thyroid nodule by fine-needle aspiration or surgical resection provides the most definitive information about its nature. The following sections consider the major thyroid neoplasms.

Thyroid Adenomas

Adenomas of the thyroid are typically discrete, solitary masses, derived from follicle epithelium, and hence they are also known as follicular adenomas. Clinically, follicular adenomas can be difficult to distinguish from dominant nodules of follicular hyperplasia or from the less common follicular carcinomas. In general, follicular adenomas are *not* forerunners to carcinomas; nevertheless, shared genetic alterations raise the possibility that at least a subset of follicular carcinomas arises in preexisting adenomas (see later). Although the vast majority of adenomas are nonfunctional, a small minority produces thyroid hormones and causes clinically apparent thyrotoxicosis. Hormone production in functional adenomas (called *toxic adenomas*) is independent of TSH stimulation.

Pathogenesis

Somatic mutations that lead to constitutive activation of the TSH receptor signaling pathway are found in toxic adenomas and toxic multinodular goiter. Gain-of-function mutations—most often in the gene encoding the TSH receptor *(TSHR),* and less commonly, the α-subunit of G_s *(GNAS)*—cause follicular cells to secrete thyroid hormone independent of TSH stimulation ("thyroid autonomy"). This leads to hyperthyroidism and produces a functional "hot" nodule on imaging. Overall, such mutations are present in 50% to 80% of toxic thyroid adenomas. Approximately one-third of toxic adenomas also harbor activating mutations of enhancer of zeste, homolog 1 (*EZH1*), which encodes a histone methyltransferase that functions as an epigenetic regulator of gene expression. Notably, *TSHR, GNAS,* and *EZH1* mutations are rarely seen in follicular carcinomas; thus, toxic adenomas (and toxic multinodular goiters) do not seem to be precancerous lesions.

In contrast, a minority (20% to 40%) of nonfunctioning follicular adenomas harbor oncogenic mutations of *RAS,* and a smaller percentage (5% to 10%) express fusion proteins containing the nuclear receptor PPARγ. Notably, both genetic alterations are seen at greater frequencies in follicular carcinomas, suggesting that nonfunctioning adenomas carrying these alterations are precursors to carcinomas. These are discussed in further detail later in this chapter.

MORPHOLOGY

The typical thyroid adenoma is a solitary, spherical, encapsulated lesion that is demarcated by a well-defined, intact capsule (Fig. 24.16A). By contrast, as the name implies, multinodular goiter contains numerous nodules, even in patients presenting with a solitary dominant nodule. Follicular adenomas average about 3 cm in diameter, but some are much larger (≥10 cm in diameter). In freshly resected specimens, the adenoma bulges from the cut surface and compresses the adjacent thyroid. The color ranges from gray-white to red-brown, depending on the cellularity of the adenoma and its colloid content. Areas of hemorrhage, fibrosis, calcification, and cystic change, similar to those encountered in multinodular goiters, are common in follicular adenomas, particularly within larger lesions.

Microscopically, the constituent cells often form uniform-appearing follicles that contain colloid (Fig. 24.16B). The follicular growth pattern is usually quite distinct from the adjacent nonneoplastic thyroid. The neoplastic cells show little variation in cell size, shape, or nuclear morphology, and mitotic figures are rare. Occasionally the neoplastic cells acquire brightly eosinophilic granular cytoplasm (**oxyphil** or **Hürthle cell change**) (Fig. 24.17). The hallmark of all follicular adenomas is the presence of an

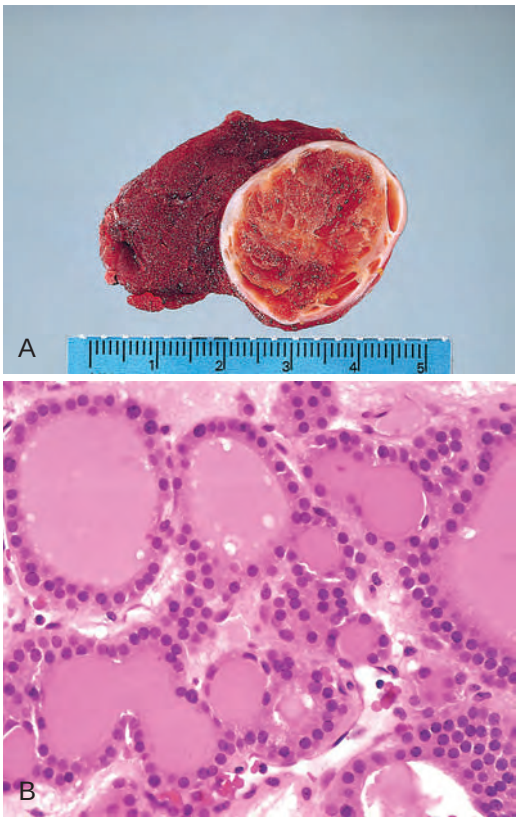


Figure 24.16 Follicular adenoma of the thyroid. (A) A solitary, well-circumscribed nodule is seen. (B) The photomicrograph shows well-differentiated follicles resembling normal thyroid parenchyma.

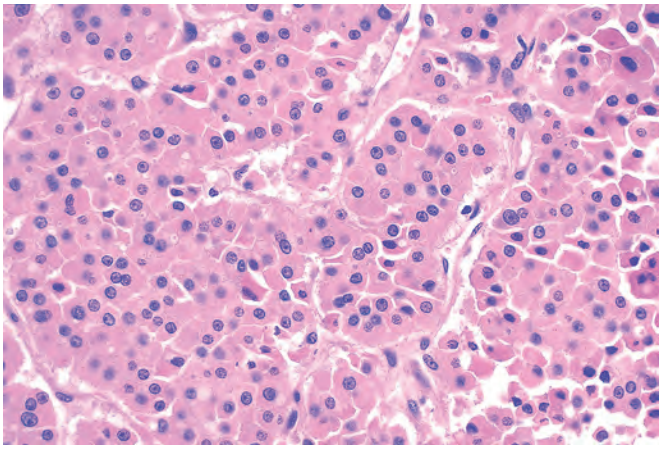


Figure 24.17 Hürthle cell (oxyphil) adenoma. A high-power view showing that the tumor is composed of cells with abundant eosinophilic cytoplasm and small regular nuclei. (Courtesy Dr. Mary Sunday, Duke University, Durham, NC.)

intact, well-formed capsule encircling the tumor. **Careful evaluation of the integrity of the capsule is therefore critical in distinguishing follicular adenoma from follicular carcinoma,** which demonstrates capsular and/or vascular invasion (see later). Mitotic activity, necrosis, or high cellularity also warrants close inspection to exclude follicular carcinoma and the encapsulated follicular variant of papillary carcinoma (see later).

Clinical Features

Many follicular adenomas present as solitary painless masses that are discovered on routine physical examination. Larger masses may produce local symptoms, such as difficulty in swallowing. Nonfunctioning adenomas take up less radioactive iodine than does normal thyroid parenchyma and on radionuclide scanning appear as cold nodules. Other techniques used to evaluate suspected adenomas are ultrasonography and fine-needle aspiration biopsy. **Because of the need for evaluating capsular integrity, the definitive diagnosis of adenoma can be made only after careful histologic examination of the resected specimen.** Suspected adenomas of the thyroid are therefore removed surgically to exclude malignancy. Follicular adenomas do not recur or metastasize and have an excellent prognosis.

Thyroid Carcinomas

It is estimated that there were than 52,000 new cases of thyroid cancer in the United States in 2019, and about 2000 individuals succumbed to the disease. The rate of newly diagnosed thyroid cancer has been rising dramatically (currently, at the rate of 3% increase in incidence each year in the United States), and in countries like South Korea, which has seen a veritable "epidemic" of thyroid cancer, it is the most common cause of newly diagnosed cancer in women. In contrast to the rising rates of newly diagnosed cancer, mortality from this disease is one of the lowest among solid tumors, with a 5-year survival greater than 98%. The major reason for this good outcome is that many of these cancers are found incidentally during neck palpation or by ultrasonography, and consist of small, localized lesions (typically papillary cancers) that have an excellent prognosis. To address this so-called "overdiagnosis" of thyroid cancer, the nomenclature of a subset of papillary cancers has undergone an important change in the past 2 years (see later).

The major subtypes of thyroid carcinoma and their relative frequencies are as follows:

- *Papillary thyroid carcinoma* (PTC), including "conventional" PTC and the "encapsulated follicular variant" (80% to 85%% of cases). This category includes the newly described lesion known as *"noninvasive follicular thyroid neoplasm with papillary-like nuclear features,"* a low-grade neoplasm with such a minimal risk of recurrence that "carcinoma" has been removed from its name.
- *Follicular carcinoma* (10% to 15% of cases)
- *Poorly differentiated* and *anaplastic (undifferentiated) carcinoma* (<5% of cases)
- *Medullary carcinoma* (5% of cases)

Most thyroid carcinomas are derived from thyroid follicle epithelium (except medullary carcinoma, which is derived from parafollicular C cells), and, of these, the vast majority are well-differentiated lesions. In the past decade, with increasing knowledge of the molecular pathology of follicular neoplasms, the concept of "multistep pathogenesis" from precursor lesions, which is well established in other epithelial cancers, has gained greater acceptance. In particular, the following three precursor lesion relationships have been recognized:

- Papillary microcarcinoma as a precursor to conventional PTC
- Noninvasive thryoid neoplasma with papillary-like nuclear features as a precursor to invasive encapsulated follicular variant PTC
- Nonfunctioning follicular adenoma as a precursor to follicular carcinoma

Further, it is now accepted that most poorly differentiated and anaplastic (undifferentiated) carcinomas arise from well-differentiated PTC or follicular carcinomas, through acquisition of additional mutations. These aggressive neoplasms and medullary thyroid carcinomas are the major causes of mortality from thyroid cancer.

We begin with a discussion of the molecular pathogenesis of thyroid neoplasms.

Pathogenesis

Driver Mutations. Distinct sets of driver mutations in cancer genes are found in the four major histologic variants of thyroid cancer. Among follicular neoplasms, there are both unique and shared genetic alterations between the histologic subtypes (Fig. 24.18), most of which involve components of the receptor tyrosine kinase (RTK) pathway, which you will recall is frequently mutated in many different cancers (Chapter 7). As is also true with other epithelial neoplasms, precursor lesions (listed earlier) share genetic alterations with the corresponding variant of carcinoma, albeit at a lower frequency.

- *Conventional papillary thyroid carcinomas.* Conventional PTCs have two defining genetic abnormalities: translocations that result in gene fusions of *RET* or *NTRK,* and point mutations in *BRAF* (see Fig. 24.18).

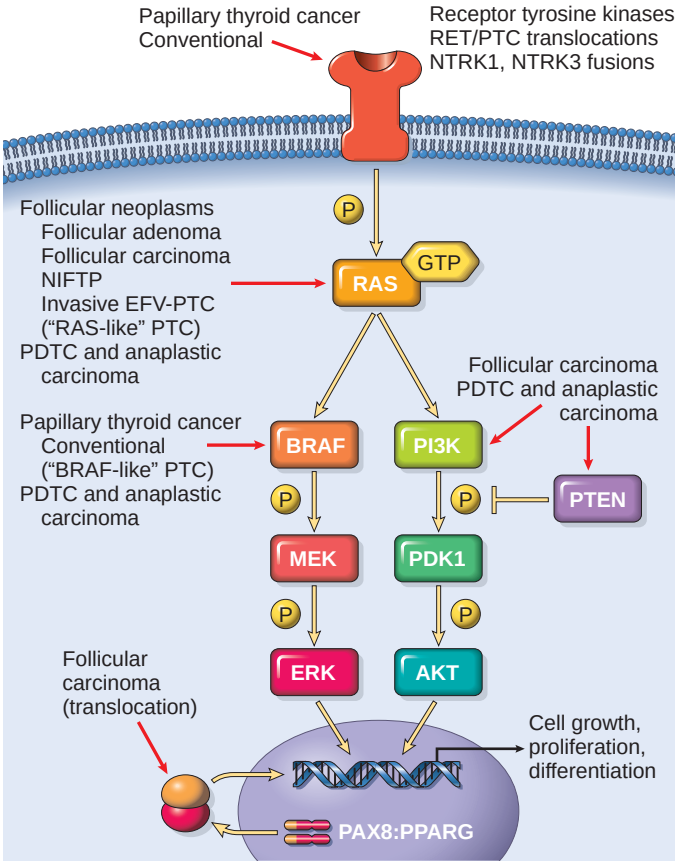


Figure 24.18 Genetic alterations in follicle cell–derived neoplasms of the thyroid gland. *EFV-PTC*, Encapsulated follicular variant of papillary thyroid carcinoma; *NIFTP*, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; *PDTC*, poorly differentiated thyroid carcinoma.

- The *RET* gene is located on chromosome 10q11, and encodes an RTK that is normally not expressed in thyroid follicular cells. As with many other RTKs implicated in cancer, PTC rearrangements involving *RET* create fusion genes that encode constitutively active forms of RET tyrosine kinase. There are more than 15 fusion partners of *RET*, and two—designated *PTC1* and *PTC2*—are most commonly observed in sporadic papillary cancers. Similarly, paracentric inversions or translocations of *NTRK1*, another gene encoding a receptor tyrosine kinase on chromosome 1q21, are present in ~5% of PTCs. These rearrangements also produce constitutively active NTRK1 fusion proteins.
- Between 50% and 80% of conventional PTCs harbor gain-of-function mutations in the *BRAF* gene, most commonly a valine-to-glutamate change in codon 600 ($BRAF^{V600E}$). BRAF encodes a serine/threonine kinase that lies downstream of receptor tyrosine kinases in growth factor signaling pathways. In a subtype of conventional PTCs, known as "tall cell variant" (see later), $BRAF^{V600E}$ mutations are virtually a diagnostic *sine qua non*. The presence of *BRAF* mutations in conventional PTCs is associated with reduced expression of thyroid differentiation markers (such as thyroglobulin and thyroid peroxidase) and may be associated with a higher risk of extrathyroidal extension and recurrence. As discussed elsewhere, the $BRAF^{V600E}$ mutation is found in many other cancers including melanoma and hairy cell leukemia (Chapter 7).
- *Follicular neoplasms.* In contrast to papillary carcinomas, follicular neoplasms are often associated with gain-of-function mutations in *RAS*. Neoplasms with *RAS* mutations retain expression of thyroid differentiation factors (e.g., thyroglobulin, thyroid peroxidase), which may contribute to their follicular growth pattern. Overall, approximately 20% to 40% of follicular adenomas and 30% to 50% of follicular carcinomas harbor *RAS* mutations (see Fig. 24.18). Similarly, approximately one-half of invasive encapsulated follicular variants of PTC and their putative precursor, noninvasive follicular thyroid neoplasm with papillary-like nuclear features, demonstrate *RAS* point mutations.

  A unique (2;3)(q13;p25) translocation has been described in 20% to 50% of follicular carcinomas and ~10% of follicular adenomas. This translocation creates a fusion gene composed of portions of *PAX8*, a paired homeobox gene that is important in thyroid development, and the peroxisome proliferator–activated receptor gene *(PPARG)*, which encodes a nuclear hormone receptor implicated in terminal differentiation of thyrocytes. Some noninvasive follicular thyroid neoplasms with papillary-like nuclear features and invasive encapsulated follicular variant of PTC (up to one-third in some series) also harbor *PAX8-PPARG* fusion genes, which are almost never seen in poorly differentiated or anaplastic carcinomas. Finally, up to 10% of follicular carcinomas exhibit gain-of-function mutations of *PIK3CA* (the gene that encodes phosphoinositide 3-kinase [PI3K]) or loss-of-function mutations of *PTEN*, a tumor suppressor gene and negative regulator of this pathway (see Fig. 24.18).
- *Poorly differentiated and anaplastic (undifferentiated) carcinomas.* These highly aggressive and lethal tumors can arise *de novo*, or, much more commonly, by "de-differentiation" of a papillary or follicular thyroid carcinoma. In addition to driver mutations that are also seen in well-differentiated thyoid cancers, three recurrent genetic "hits"—point mutations of *TP53*, beta-catenin *(CTNNB1)*, and *TERT* (which encodes the catalytic component of the enzyme telomerase)—are essentially restricted to poorly differentiated and anaplastic carcinomas and likely have central roles in their genesis and aggressive behavior.
- *Medullary thyroid carcinomas.* Familial medullary thyroid carcinomas occur in multiple endocrine neoplasia, type 2 (MEN-2, see later) and are associated with germline *RET* mutations that lead to constitutive activation of the receptor. *RET* mutations are also seen in approximately one-half of nonfamilial (sporadic) medullary thyroid cancers. Chromosomal rearrangements involving *RET*, such as the *RET/PTC* translocations reported in papillary cancers, are not seen in medullary carcinomas.

Environmental Factors. The major risk factor predisposing to thyroid cancer is exposure to *ionizing radiation*, particularly during the first two decades of life. In keeping with this, there was a marked increase in the incidence of conventional PTCs among children exposed to ionizing radiation after the Chernobyl nuclear disaster in 1986. Of

interest, radiation-induced PTCs have a higher frequency of chromosome rearrangements that produce gene fusions, possibly because of the ability of ionizing radiation to induce double-stranded DNA breaks. Deficiency of dietary iodine (and by extension, goiter) is associated with a higher frequency of follicular lesions.

Papillary Carcinoma and Follicular Variants, Including Invasive Encapsulated Follicular Variant of PTC and Noninvasive Follicular Thyroid Neoplasm With Papillary-like Nuclear Features

Papillary carcinomas are the most common form of thyroid cancer, accounting for nearly 85% of cases in the United States. They occur throughout life but most often between 25 and 50 years of age, and account for the majority of thyroid carcinomas associated with previous exposure to ionizing radiation. The incidence of papillary carcinoma has increased markedly in the last 30 years, partly because of the rise in detection of papillary microcarcinomas and noninvasive encapsulated variants found incidentally during neck examination.

MORPHOLOGY

Conventional papillary carcinomas may be solitary or multifocal. Some tumors are well circumscribed, while others infiltrate the adjacent parenchyma and have ill-defined margins. The tumors may contain areas of fibrosis and calcification and are often cystic. The cut surface sometimes reveals papillary foci that point to the diagnosis. **Papillary microcarcinoma** is defined as an otherwise conventional papillary carcinoma, but less than 1 cm in size. As stated earlier, they are considered putative precursor lesions of more typical papillary carcinomas.

The microscopic hallmarks of papillary neoplasms include the following (Fig. 24.19):

- Branching **papillae** having a fibrovascular stalk covered by a single to multiple layers of cuboidal epithelial cells. In most neoplasms, the epithelium covering the papillae consists of well-differentiated, uniform, orderly cuboidal cells, but at the other extreme are those with fairly anaplastic epithelium showing considerable variation in cell and nuclear morphology. When present, the papillae differ from those seen in areas of hyperplasia in being more complex and having dense fibrovascular cores.
- Nuclei with finely dispersed chromatin and an optically clear or empty appearance, giving rise to the designation **ground-glass** or **Orphan Annie eye nuclei.** In addition, invaginations of the nuclear membrane may give the appearance of nuclear inclusions ("pseudo-inclusions") or grooves. **The diagnosis of papillary carcinoma can be made based on these nuclear features,** even in the absence of papillary architecture.
- Concentrically calcified structures termed **psammoma bodies** are often present, usually within the cores of papillae. These

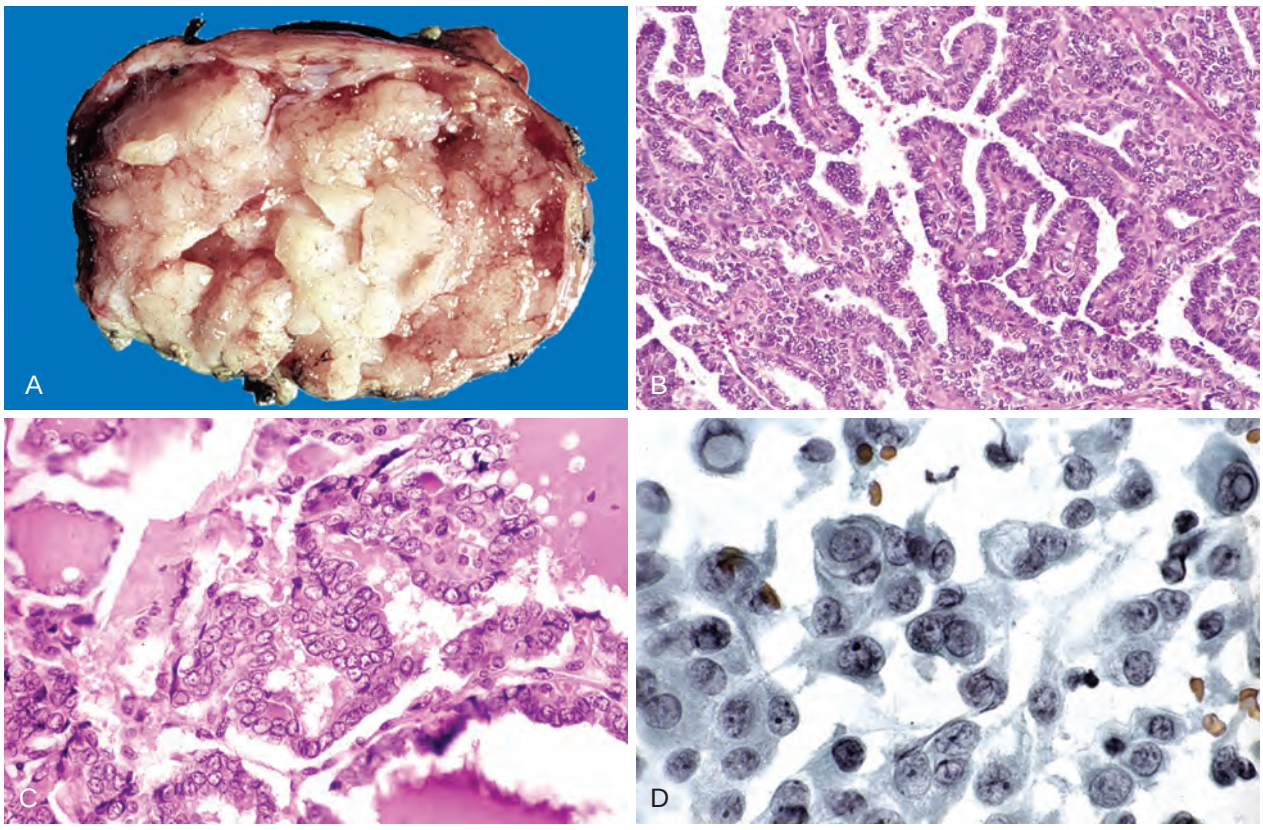


Figure 24.19 Papillary carcinoma of the thyroid. (A) The macroscopic appearance of a papillary carcinoma with grossly discernible papillary structures. (B) This particular example contains well-formed papillae. (C) High power shows cells with characteristic empty-appearing nuclei, sometimes called "Orphan Annie eye" nuclei. (D) Cells obtained by fine-needle aspiration of a papillary carcinoma. Characteristic intranuclear inclusions are visible in some of the aspirated cells.

structures are almost never found in follicular and medullary carcinomas, and they are a strong indication that the lesion is a papillary carcinoma when present in fine-needle aspiration material.

- Foci of lymphatic invasion by tumor are often present, but involvement of blood vessels is uncommon, particularly in smaller lesions. Metastases to adjacent cervical lymph nodes occur in up to one-half of cases.

There are over a dozen histologic variants of papillary carcinoma that can mimic other thyroid lesions or harbor distinct prognostic implications; most are beyond the scope of this book. The **tall cell variant** has tall columnar cells with intensely eosinophilic cytoplasm. These tumors tend to occur in older individuals and have higher frequencies of vascular invasion, extrathyroidal extension, and cervical and distant metastases than conventional PTC. Tall cell variant papillary carcinomas almost always harbor *BRAF* mutations, and often have *RET/PTC* translocations as well. The co-occurrence of these two aberrations may contribute to the aggressive behavior of this variant.

An unusual **diffuse sclerosing variant** of papillary carcinoma occurs in younger individuals, including children. The tumor has a prominent papillary growth pattern intermixed with solid areas containing nests of squamous metaplasia. As the name suggests, there is extensive, diffuse fibrosis throughout the thyroid gland, often associated with a prominent lymphocytic infiltrate, simulating Hashimoto thyroiditis. Lymph node metastases are present in almost all cases. The diffuse sclerosing variant carcinomas lack *BRAF* mutations, but *RET/PTC* translocations are found in approximately one-half of cases.

The **follicular variant of PTC** has the characteristic nuclear features of papillary carcinoma and an almost totally follicular architecture. As mentioned earlier, genetic analyses have shown that encapsulated follicular variants of PTC harbor distinct molecular abnormalities from conventional PTCs. As already discussed, encapsulated follicular variants of papillary thyroid cancer without capsular invasion are designated noninvasive follicular thyroid carcinoma with papillary-like nuclear features and have a very low risk of recurrence or metastasis, whereas invasive tumors are referred to as invasive encapsulated follicular variant of papillary thyroid carcinoma.

Clinical Features

Most conventional papillary carcinomas present as asymptomatic thyroid nodules, but the first manifestation may be a mass in a cervical lymph node. Interestingly, the presence of isolated cervical nodal metastases does not have a significant impact on prognosis, which is generally good. Most carcinomas are single nodules that move freely with the thyroid gland during swallowing and are not distinguishable on examination from benign nodules. Hoarseness, dysphagia, cough, or dyspnea suggests advanced disease. In a minority of patients, hematogenous metastases are present at the time of diagnosis, most commonly in the lung.

A variety of diagnostic tests have been used to help separate benign from malignant thyroid nodules, including radionuclide scanning and fine-needle aspiration. Papillary carcinomas are cold masses on scintigraphy. Improvements in cytologic analysis have made fine-needle aspiration cytology a reliable test for distinguishing between benign and malignant nodules.

Papillary thyroid cancers have an excellent prognosis, with a 10-year survival rate in excess of 95%. Patients with papillary microcarcinomas and noninvasive follicular thyroid neoplasms with papillary-like nuclear features have an outstanding prognosis on lobectomy alone, and they typically do not require total thyroidectomy. This has become especially important since thyroidectomy may lead to vocal cord palsy (due to injury to the laryngeal nerve) and iatrogenic hypoparathyroidism.

Between 5% and 20% of patients with more typical PTCs have local or regional recurrences, and 10% to 15% have distant metastases. The prognosis is dependent on several factors including age (in general, being less favorable among patients older than 40 years), presence of extrathyroidal extension, and presence of distant metastases (stage).

Follicular Carcinoma

Follicular carcinoma accounts for 5% to 15% of primary thyroid cancers; it is more frequent in areas with dietary iodine deficiency, where it constitutes 25% to 40% of thyroid cancers. It is more common in women (3:1) and presents more often in older patients than does papillary carcinoma; the peak incidence is between 40 and 60 years of age.

MORPHOLOGY

Follicular carcinomas are single nodules that may be well circumscribed or widely infiltrative (Fig. 24.20A). Sharply demarcated lesions may be exceedingly difficult to distinguish from follicular adenomas by gross examination. Larger lesions may penetrate the capsule and infiltrate into the adjacent neck. They are gray to tan to pink on cut section and may be translucent due to the presence of large, colloid-filled follicles. Degenerative changes, such as central fibrosis and foci of calcification, may be present.

Microscopically, most follicular carcinomas are composed of fairly uniform cells forming small follicles containing colloid, quite reminiscent of normal thyroid (Fig. 24.20B). In other cases, follicular differentiation may be less apparent, and there may be nests or sheets of cells without colloid. Occasional tumors are dominated by cells with abundant granular, eosinophilic cytoplasm **(Hürthle cell or oncocytic variant of follicular carcinoma).** Whatever the pattern, the nuclei lack the features typical of papillary carcinoma, and psammoma bodies are not present. **There is no reliable cytologic difference between follicular adenomas and minimally invasive follicular carcinomas**. Making this distinction requires extensive histologic sampling of the tumor capsule to exclude capsular and/or vascular invasion (Fig. 24.21). The criterion for vascular invasion is applicable only to capsular vessels and vascular spaces beyond the capsule; the presence of tumor plugs within intra-tumoral blood vessels has little prognostic significance. Unlike in papillary cancers, lymphatic spread is uncommon in follicular cancers. By contrast, the diagnosis of carcinoma is obvious in widely invasive follicular carcinomas, which infiltrate the thyroid parenchyma and extrathyroidal soft tissues. Histologically, these cancers tend to have a greater proportion of solid or trabecular growth pattern, less evidence of follicular differentiation, and increased mitotic activity.

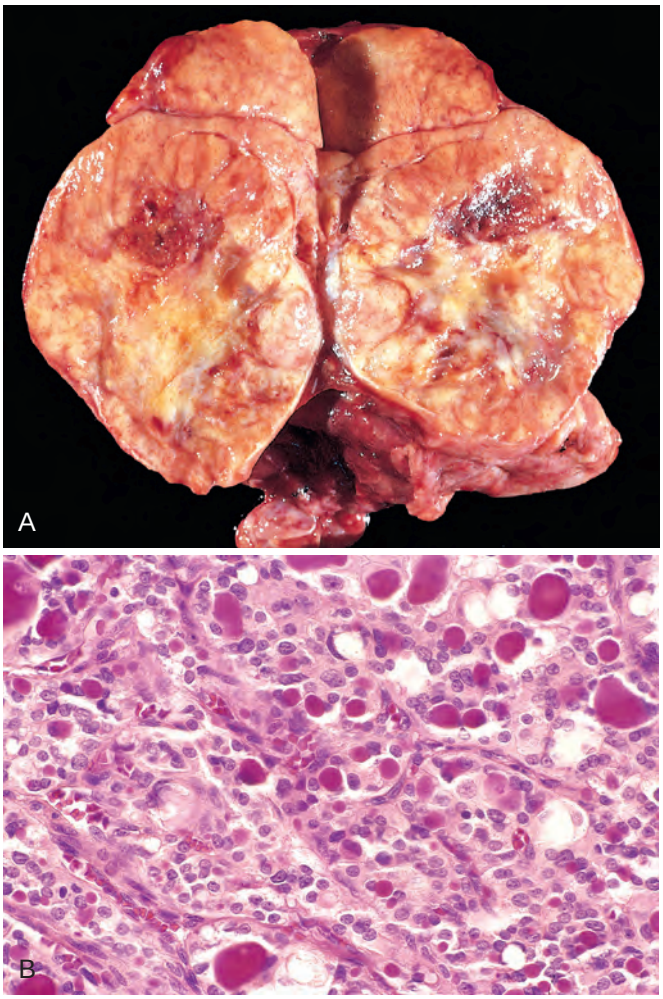


Figure 24.20 Follicular carcinoma. (A) Cut surface of a follicular carcinoma with substantial replacement of the lobe of the thyroid. The tumor has a light-tan appearance and contains small foci of hemorrhage. (B) A few of the glandular lumens contain recognizable colloid.

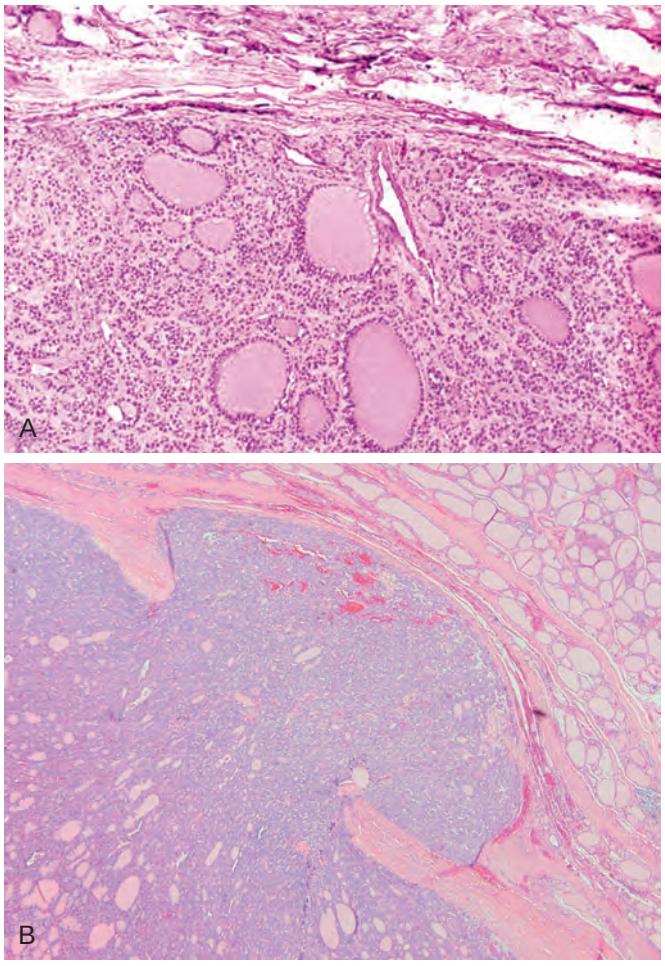


Figure 24.21 Capsular integrity in follicular neoplasms. In adenomas (A), a fibrous capsule, usually thin but occasionally more prominent, circumferentially surrounds the neoplastic follicles, and no capsular invasion is seen; compressed normal thyroid parenchyma is usually present external to the capsule *(top of panel)*. In contrast, follicular carcinomas demonstrate capsular invasion (B) that may be minimal, as in this case, or widespread. The presence of vascular invasion is another feature of follicular carcinomas.

Clinical Features

Follicular carcinomas present as slowly enlarging painless nodules. Most frequently they are cold nodules on scintigrams, although rare, better-differentiated lesions may be hyperfunctional, take up radioactive iodine, and appear warm on scintigraphy. Because follicular carcinomas have little propensity for invading lymphatics, regional lymph nodes are typically not involved, while vascular (hematogenous) dissemination is common, with metastases to bone, lungs, liver, and elsewhere.

The prognosis depends largely on the extent of invasion and stage at presentation. Widely invasive follicular carcinoma often presents with systemic metastases, and as many as one-half of affected patients succumb to their disease within 10 years. This is in sharp contrast to minimally invasive follicular carcinomas, which have a 10-year survival rate of greater than 90%. Most follicular carcinomas are treated with total thyroidectomy followed by the administration of radioactive iodine, which can be used to identify metastases and to ablate such lesions. In addition, because any residual follicular carcinoma may respond to TSH stimulation, patients are usually treated with thyroid hormone after surgery to suppress endogenous TSH levels. Serum thyroglobulin levels are used for monitoring tumor recurrence, because this thyroid protein should be barely detectable in a patient who is free of disease.

Poorly Differentiated and Anaplastic (Undifferentiated) Carcinoma

Poorly differentiated and anaplastic (undifferentiated) carcinomas represent aggressive neoplasms of the thyroid follicular epithelium, cumulatively accounting for less than 5% of thyroid tumors, but with a high mortality rate (approaching 100% for anaplastic lesions). Patients with either carcinoma are older than those with other types of thyroid cancer, with a mean age of 65 years. Approximately

one-fourth of patients with poorly differentiated or anaplastic thyroid carcinomas have a past history of a well-differentiated follicular neoplasm, and another one-fourth harbor a concurrent well-differentiated neoplasm in the resected specimen.

MORPHOLOGY

Microscopically, poorly differentiated carcinomas are composed of cells with minimal follicular differentiation, arranged in insular or trabecular growth patterns; intratumoral necrosis and frequent mitoses are often seen. Anaplastic carcinomas demonstrate variable morphology, including (1) large, pleomorphic **giant cells,** including occasional osteoclast-like multinucleate giant **cells;** (2) **spindle cells** with a sarcomatous appearance; and (3) **mixed spindle and giant cells.** Foci of papillary or follicular differentiation may be present in some tumors, suggesting an origin from a well-differentiated carcinoma. The neoplastic cells express epithelial markers like cytokeratin, but are usually negative for markers of thyroid differentiation, like thyroglobulin.

Clinical Features

Poorly differentiated and anaplastic carcinomas usually present as rapidly enlarging bulky neck masses. In most cases, the disease has spread beyond the thyroid into adjacent neck structures or has metastasized to the lungs at the time of presentation. Symptoms related to compression and invasion of adjacent structures, such as dyspnea, dysphagia, hoarseness, and cough, are common. Anaplastic carcinomas of the thyroid are one of the most aggressive cancers known; in most cases death occurs in less than 1 year. Poorly differentiated carcinomas fare somewhat better with radical surgery, external beam radiotherapy, and radioactive iodine. More recently, immunotherapy using immune checkpoint inhibitors has been used in these neoplasms in an attempt to improve their dire prognosis.

Medullary Carcinoma

Medullary carcinomas of the thyroid are neuroendocrine neoplasms derived from the parafollicular cells (C cells). They account for approximately 5% of thyroid neoplasms. Medullary carcinomas, like normal C cells, secrete *calcitonin,* the measurement of which plays an important role in the diagnosis and postoperative follow-up of patients. Calcitonin is a regulator of calcium metabolism. It is normally produced in response to hypercalcemia, and it reduces serum calcium by inhibiting osteoclast activity and renal tubular reabsorption of calcium. In some instances, the tumor cells elaborate other polypeptide hormones, such as serotonin, ACTH, and vasoactive intestinal peptide (VIP). About 70% of tumors arise sporadically. The remainder occur in the setting of MEN-2A or MEN-2B syndromes. Recall that activating point mutations in the *RET* proto-oncogene play an important role in the development of both familial and sporadic medullary carcinomas. Cases associated with MEN-2A or MEN-2B occur in younger patients, and they may even arise during the first decade of life. In contrast, sporadic and familial medullary carcinomas are neoplasms of adulthood, with a peak incidence in the 40s and 50s.

MORPHOLOGY

Sporadic medullary thyroid carcinomas present as a solitary nodule (Fig. 24.22A). In contrast, **bilaterality and multicentricity are common in familial cases.** Larger lesions often contain areas of necrosis and hemorrhage and may extend through the capsule of the thyroid. The tumors are firm, pale gray to tan, and infiltrative. There may be foci of hemorrhage and necrosis in larger lesions.

Microscopically, medullary carcinomas are composed of polygonal to spindle-shaped cells, which may form nests, trabeculae, and even follicles. Small, more anaplastic cells are present in some tumors and may be the predominant cell type. **Amyloid deposits** derived from calcitonin polypeptides are present in the stroma in many cases (Fig. 24.22B). Calcitonin is readily demonstrable within the cytoplasm of the tumor cells as well as in the stromal amyloid by immunohistochemical methods. As with all neuroendocrine tumors, electron microscopy reveals variable numbers of membrane-bound electron-dense granules within the cytoplasm of the neoplastic cells (Fig. 24.23). One of the features of familial

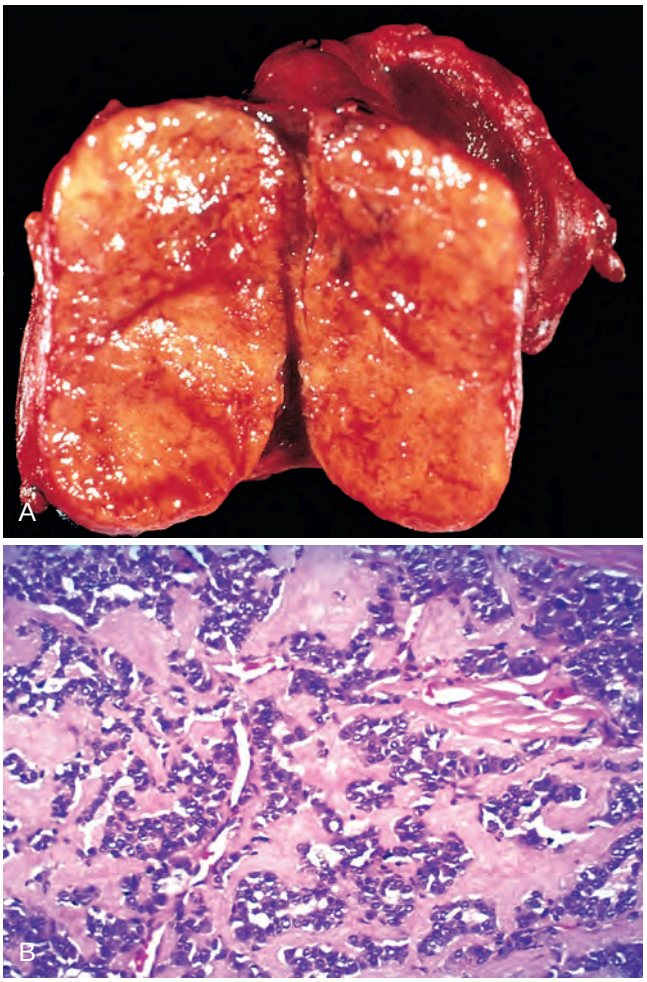


Figure 24.22 Medullary carcinoma of the thyroid. (A) These tumors typically show a solid pattern of growth and do not have connective tissue capsules. (B) Histology demonstrates abundant deposition of amyloid, visible here as homogeneous extracellular material, derived from calcitonin secreted by the neoplastic cells. (A, Courtesy Dr. Joseph Corson, Brigham and Women's Hospital, Boston, Mass.)

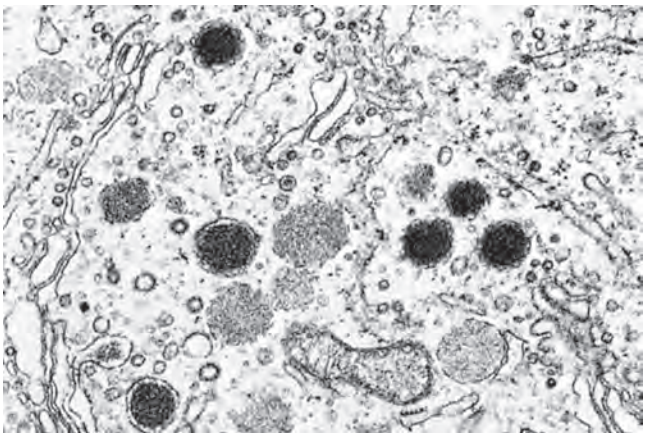


Figure 24.23 Electron micrograph of medullary thyroid carcinoma. These cells contain membrane-bound secretory granules that are the sites of storage of calcitonin and other peptides.

medullary cancers is the presence of multicentric **C-cell hyperplasia** in the surrounding thyroid parenchyma, a feature that is usually absent in sporadic lesions, and that is believed to be a precursor lesion in familial cases. Thus, the presence of multiple prominent clusters of C cells scattered throughout the parenchyma should raise the specter of an inherited predisposition, even if a family history is not present.

Clinical Features

Sporadic cases of medullary carcinoma come to medical attention most often as a mass in the neck, sometimes associated with dysphagia or hoarseness. In some instances, the initial manifestations are those of a paraneoplastic syndrome caused by the secretion of a peptide hormone (e.g., diarrhea due to the secretion of VIP, or Cushing syndrome due to ACTH). Notably, hypocalcemia is not a prominent feature, despite the presence of raised calcitonin levels. In addition to circulating calcitonin, secretion of carcinoembryonic antigen by the neoplastic cells is a useful biomarker, especially for presurgical assessment of tumor load and in calcitonin-negative tumors.

Patients with familial syndromes may come to attention because of symptoms localized to the thyroid or as a result of endocrine neoplasms in other organs (e.g., adrenal or parathyroid glands). Medullary carcinomas arising in the context of MEN-2B are generally more aggressive and metastasize more frequently than those occurring in patients with sporadic tumors or MEN-2A. As discussed later, asymptomatic MEN-2 individuals carrying germline *RET* mutations are offered prophylactic thyroidectomy as early as possible to prevent the otherwise inevitable development of medullary carcinoma, the major risk factor for poor outcome in these individuals. Sometimes the only histologic finding in the resected thyroid of asymptomatic carriers is the presence of C-cell hyperplasia or small (<1 cm) "micromedullary" carcinomas. Several small-molecule inhibitors of RET tyrosine kinase have recently been developed and have shown considerable promise in clinical trials of patients with medullary carcinoma.

KEY CONCEPTS

THYROID NEOPLASMS

- Most thyroid neoplasms manifest as solitary thyroid nodules; only 1% of all thyroid nodules are neoplastic.
- Multiple genetic pathways are involved in thyroid carcinogenesis. Conventional papillary carcinomas harbor either fusions *(RET/PTC, NTRK)* or *BRAF* point mutations. Follicular neoplasms of all subtypes are characterized by oncogenic *RAS* mutations and *PAX8/PPARG* fusions. Progression to aggressive neoplasms is associated with *TP53*, *CTNNB1*, and *TERT* mutations.
- Follicular adenomas and carcinomas both are composed of well-differentiated follicular epithelial cells; the latter are distinguished by evidence of capsular and/or vascular invasion.
- Conventional papillary carcinomas are recognized based on nuclear features (ground-glass nuclei, pseudoinclusions), even in the absence of papillae. Psammoma bodies are a characteristic feature of papillary cancers; these neoplasms often metastasize by way of lymphatics, but the prognosis is excellent.
- Encapsulated variants of papillary carcinoma without capsular or lymphatic invasion have recently been named noninvasive follicular thyroid neoplasm with papillary-like nuclear features to emphasize their ultra–low-risk potential for metastasis or recurrence.
- Poorly differentiated and anaplastic carcinomas are thought to arise from more differentiated thyroid carcinomas. They are highly aggressive, often lethal cancers.
- Medullary cancers are neoplasms arising from parafollicular C cells and can occur in either sporadic (70%) or familial (30%) settings. Multicentricity and C cell hyperplasia are features of familial cases. Amyloid deposits are a characteristic histologic finding.

CONGENITAL ANOMALIES

Thyroglossal duct cyst is the most common clinically significant congenital anomaly of the thyroid. A sinus tract may persist as a vestige of the tubular development of the thyroid gland. Parts of this tube may be obliterated, leaving small segments that form cysts. These occur at any age and might not become evident until adult life. Mucinous, clear secretions may collect within the cysts to form either spherical masses or fusiform swellings, rarely over 2 to 3 cm in diameter, that present in the midline of the neck anterior to the trachea. Segments of the duct and cysts that occur high in the neck are lined by stratified squamous epithelium resembling the covering of the posterior portion of the tongue in the region of the foramen cecum. Anomalies that occur in the lower neck more proximal to the thyroid gland are lined by epithelium resembling the thyroidal acinar epithelium. Characteristically, subjacent to the lining epithelium, there is an intense lymphocytic infiltrate. Superimposed infection may convert these lesions into abscess cavities, and rarely, they give rise to cancers.

Parathyroid Glands

The parathyroid glands are derived developmentally from pharyngeal pouches that also give rise to the thymus. They are most commonly located in close proximity to the upper and lower poles of each thyroid lobe but may be found anywhere along the pathway of descent of the pharyngeal pouches, including the carotid sheath, the thymus, and elsewhere in the anterior mediastinum. The four parathyroid glands are composed of two cell types: chief cells and oxyphil cells. *Chief cells* predominate; they are polygonal, 12 to 20 μm in diameter, and have central, round, uniform nuclei and light to dark pink cytoplasm. Chief cells have secretory granules containing *parathyroid hormone (PTH)*. *Oxyphil cells* and transitional oxyphils are found throughout the normal parathyroid, either singly or in small clusters. They are slightly larger than the chief cells, have acidophilic cytoplasm, and are tightly packed with mitochondria. Glycogen granules are also present in these cells, but secretory granules are sparse or absent. In early infancy and childhood, the parathyroid glands are composed almost entirely of solid sheets of chief cells. The amount of stromal fat increases up to 25 years of age, reaching a maximum of approximately 30% of the gland, and then plateaus.

The function of the parathyroid glands is to regulate calcium homeostasis. The activity of the parathyroid glands is controlled by the level of free (ionized) calcium in the blood. Normally, decreased levels of free calcium stimulate the synthesis and secretion of PTH. Several metabolic functions of PTH regulate serum calcium levels:

- Increased renal tubular reabsorption of calcium, thereby conserving free calcium
- Increased conversion of vitamin D to its active dihydroxy form in the kidneys, which, in turn, augments gastrointestinal calcium absorption
- Increased urinary phosphate excretion, thereby lowering serum phosphate levels and further increasing calcium (since phosphate binds to ionized calcium)
- Enhanced osteoclastic activity (i.e., bone resorption, thus releasing ionized calcium), by promoting the differentiation of osteoclast progenitor cells into mature osteoclasts

The net result of these activities is to elevate the level of free calcium, which, in turn, inhibits further PTH secretion in a classic feedback loop.

Similar to the other endocrine organs, abnormalities of the parathyroid glands include both hyperfunction and hypofunction. Tumors of the parathyroid glands, in contrast to thyroid tumors, usually come to attention because of excessive secretion of PTH rather than mass effects.

HYPERPARATHYROIDISM

Hyperparathyroidism is caused by elevated PTH and is classified into primary, secondary, and tertiary types.

- *Primary hyperparathyroidism:* an autonomous overproduction of PTH, usually resulting from an adenoma or hyperplasia of parathyroid tissue
- *Secondary hyperparathyroidism:* compensatory hypersecretion of PTH in response to prolonged hypocalcemia, most commonly from chronic renal failure
- *Tertiary hyperparathyroidism:* persistent hypersecretion of PTH, even after the cause of prolonged hypocalcemia is corrected (e.g., after renal transplant)

Primary Hyperparathyroidism

Primary hyperparathyroidism is one of the most common endocrine disorders, and it is an important cause of hypercalcemia. The frequency of the various parathyroid lesions underlying hyperfunction is as follows:

- Adenoma: 85% to 95%
- Primary hyperplasia (diffuse or nodular): 5% to 10%
- Parathyroid carcinoma: ~1%

Primary hyperparathyroidism is usually a disease of adults and is more common in women than in men by a ratio of nearly 4:1. The annual incidence is now estimated to be about 25 cases per 100,000 in the United States and Europe; as many as 80% of patients with this condition are identified in the outpatient setting, when hypercalcemia is discovered incidentally on a serum electrolyte panel. Most cases occur in the 50s or later in life.

The most common cause of primary hyperparathyroidism is a solitary sporadic parathyroid adenoma (Fig. 24.24). Most, if not all, sporadic parathyroid adenomas are monoclonal, consistent with a neoplastic origin. As with nodules in goitrous thyroids, sporadic parathyroid "hyperplasia" is also monoclonal in many instances, particularly when associated with a persistent stimulus for parathyroid growth (refractory secondary or tertiary parathyroidism; see later), suggesting that these lesions lie in the gray zone between reactive hyperplasia and neoplasia. There are two molecular

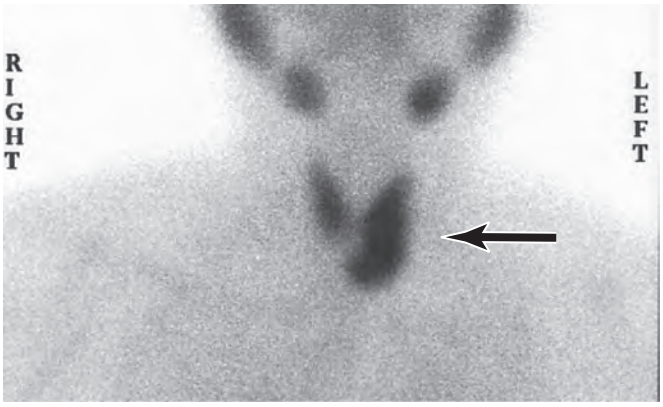


Figure 24.24 Parathyroid adenoma imaging. Technetium-99m-sestamibi radionuclide scan demonstrates an area of increased uptake corresponding to the left inferior parathyroid gland *(arrow)*, which contained a parathyroid adenoma. Preoperative scintigraphy is useful in localizing and distinguishing adenomas from parathyroid hyperplasia, where more than one gland demonstrates increased uptake.

defects that have an established role in the development of sporadic adenomas:

- *Cyclin D1 (CCDN1)* gene inversions leading to overexpression of cyclin D1 protein, a major regulator of the cell cycle. A pericentromeric inversion on chromosome 11 results in relocation of *CCDN1* (normally on 11q), so that it is positioned adjacent to the 5′-flanking region of the *PTH* gene (on 11p). As a consequence of these changes, a 5′-regulatory element flanking the *PTH* gene directs overexpression of cyclin D1, causing the cells to proliferate. Between 10% and 20% of sporadic parathyroid adenomas have this clonal rearrangement. In addition, cyclin D1 is overexpressed in approximately 40% of parathyroid adenomas, suggesting that mechanisms other than *CCDN1* gene inversion can lead to its overexpression.
- *MEN1 mutations:* Approximately 30% to 35% of sporadic parathyroid tumors have somatic mutations of the *MEN1* gene, a tumor suppressor gene on chromosome 11q13, with loss of heterozygosity (LOH) of the second allele. Germline mutations of *MEN1* are also found in patients with familial parathyroid adenomas (see later). The spectrum of *MEN1* mutations in sporadic tumors is virtually identical to that in familial parathyroid adenomas.
- In addition, *CDC73,* which encodes a protein known as parafibromin, is mutated in ~70% of sporadic parathyroid carcinomas, but rarely in adenomas. Germline mutations of *CDC73* lead to a rare syndrome known as hyperparathyroidism–jaw tumor syndrome, which includes parathyroid carcinomas and ossifying jaw tumors as part of the disease spectrum.

Familial syndromes are a distant second to sporadic parathyroid adenomas as causes of primary hyperparathyroidism. The genetic syndromes associated with *familial parathyroid adenomas* include MEN-1, MEN-2, and MEN-4 (discussed in further detail later), as well as familial hypocalciuric hypercalcemia, a rare autosomal dominant disorder caused by loss-of-function mutations in the parathyroid calcium-sensing receptor (*CASR*) gene, which results in decreased sensitivity to extracellular calcium.

MORPHOLOGY

The morphologic changes seen in primary hyperparathyroidism involve the parathyroid glands and organs that are affected by elevated levels of PTH and calcium. Parathyroid **adenomas** are almost always solitary and, like normal parathyroid glands, may be in close proximity to the thyroid or in an ectopic site (e.g., the mediastinum). The typical adenoma averages 0.5 to 5 g and consists of a well-circumscribed, soft, tan to reddish-brown nodule invested by a delicate capsule. In contrast to primary hyperplasia, the glands outside the adenoma are usually normal in size or shrunken because of feedback inhibition by elevated levels of serum calcium. Microscopically, parathyroid adenomas are mostly composed of uniform, polygonal chief cells with small, centrally placed nuclei (Fig. 24.25). At least a few nests of larger oxyphil cells are present as well; uncommonly, adenomas are composed entirely of this cell type **(oxyphil adenomas).** These may resemble Hürthle cell tumors in the thyroid. A rim of compressed, nonneoplastic parathyroid tissue, generally separated by a fibrous capsule, is often visible at the edge of the adenoma. Mitotic figures are rare, but it is not uncommon to find bizarre and pleomorphic nuclei even within adenomas (so-called **endocrine atypia**); this is not a criterion for malignancy. In contrast to the normal parathyroid parenchyma, adipose tissue is inconspicuous.

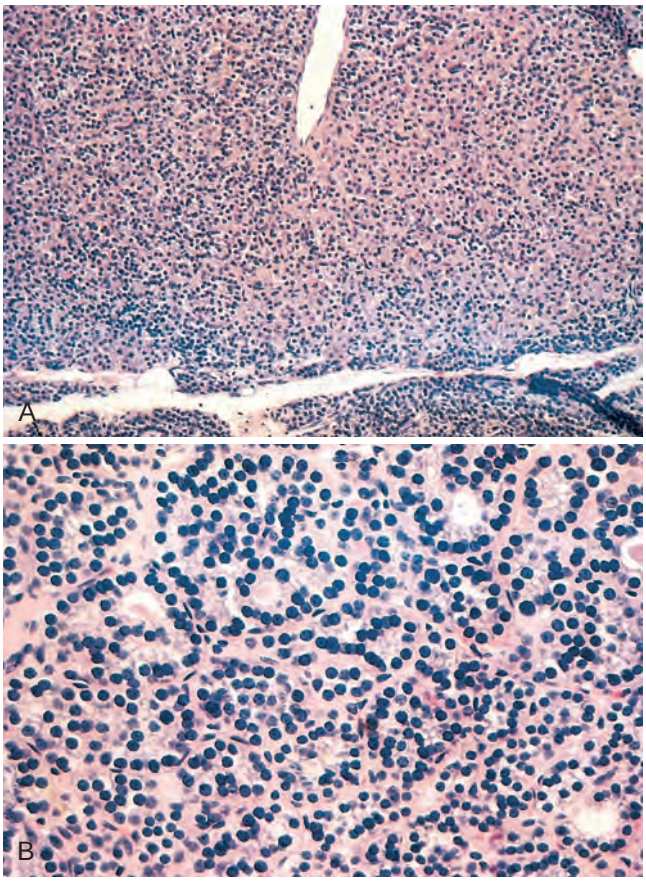


Figure 24.25 Parathyroid adenoma. (A) Solitary chief cell parathyroid adenoma (low-power photomicrograph) revealing delineation from the residual gland below. (B) High-power detail of a chief cell parathyroid adenoma. There is slight variation in nuclear size but no anaplasia and a tendency for follicle formation.

Primary hyperplasia may occur sporadically or as a component of MEN syndrome. Although classically all four glands are involved, there is frequently asymmetry with apparent sparing of one or two glands, making the distinction between hyperplasia and adenoma difficult. The combined weight of all glands rarely exceeds 1 g and is often less. Microscopically, the most common pattern seen is that of chief cell hyperplasia, which may involve the glands in a diffuse or multinodular pattern. Less commonly, the constituent cells contain abundant glycogen, giving them a clear appearance in histologic sections ("water-clear cell hyperplasia"). In many instances, there are islands of oxyphils, and poorly developed, delicate fibrous strands may envelop the nodules. As in the case of adenomas, stromal fat is inconspicuous within hyperplastic glands.

Parathyroid carcinomas may be circumscribed lesions that are difficult to distinguish from adenomas, or they may be clearly invasive neoplasms. These tumors enlarge one parathyroid gland and consist of gray-white, irregular masses that sometimes exceed 10 g in weight. The cells are usually uniform and resemble normal parathyroid cells, and are arrayed in nodular or trabecular patterns.

The mass is usually enclosed by a dense, fibrous capsule. **Diagnosis of carcinoma based on cytologic detail is unreliable, and invasion of surrounding tissues and metastasis are the only reliable criteria.** Local recurrence occurs in one-third of cases, and more distant dissemination occurs in another third.

Morphologic changes of hyperparathyroidism in the skeletal system (Chapter 26) and the urinary tract deserve special mention. Symptomatic, untreated primary hyperparathyroidism manifests with three interrelated skeletal abnormalities: osteoporosis, brown tumors, and osteitis fibrosa cystica. The osteoporosis results in decreased bone mass, with preferential involvement of the phalanges, vertebrae, and proximal femur. For unknown reasons, the increased osteoclast activity in hyperparathyroidism affects cortical bone (subperiosteal and endosteal surfaces) more severely than medullary bone. In medullary bone, osteoclasts tunnel centrally along the length of the trabeculae, creating the appearance of railroad tracks and producing what is known as dissecting osteitis (Fig. 24.26). The marrow spaces around the affected surfaces are replaced by fibrovascular tissue.

The bone loss predisposes to microfractures and secondary hemorrhages that elicit an influx of macrophages and an ingrowth of reparative fibrous tissue, creating a mass of reactive tissue, known as a **brown tumor** (see Fig. 26.13, Chapter 26). The brown color is the result of the vascularity, hemorrhage, and hemosiderin deposition, and it is not uncommon for the lesions to undergo cystic degeneration. The combination of increased osteoclast activity, peritrabecular fibrosis, and cystic brown tumors is the hallmark of severe hyperparathyroidism and is known as **generalized osteitis fibrosa cystica (von Recklinghausen disease of bone).** Osteitis fibrosa cystica is now rarely encountered because hyperparathyroidism is usually diagnosed on routine blood tests and treated at an early, asymptomatic stage (see later).

PTH-induced hypercalcemia favors formation of **urinary tract stones** (nephrolithiasis) as well as calcification of the renal interstitium and tubules (nephrocalcinosis). Metastatic calcification secondary to hypercalcemia may also be seen in other sites, including the stomach, lungs, myocardium, and blood vessels.

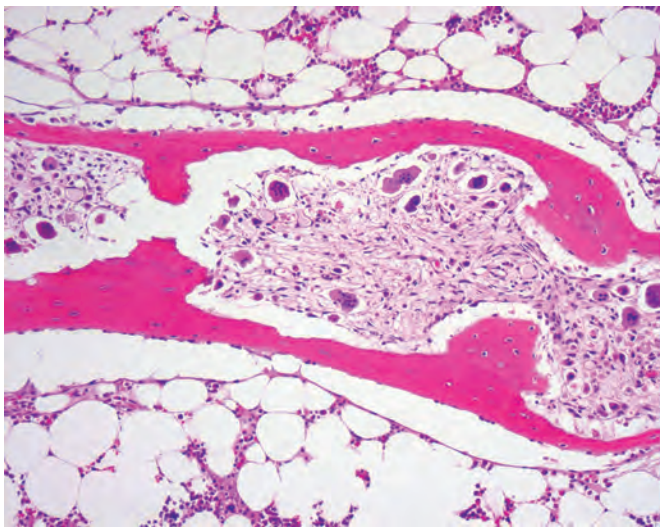


Figure 24.26 Hyperparathyroidism with osteoclasts boring into the center of the trabeculum (dissecting osteitis). (Photomicrograph reproduced from Horvai A: *Bone and Soft Tissue Pathology: A Volume in the High Yield Pathology Series,* Philadelphia, 2012, Elsevier.)

Clinical Features

Primary hyperparathyroidism may be (1) asymptomatic and identified on routine blood chemistry profile, or (2) associated with the classic clinical manifestations of primary hyperparathyroidism.

- *Asymptomatic hyperparathyroidism.* Because serum calcium levels are routinely assessed, most patients with primary hyperparathyroidism are diagnosed incidentally. In fact, primary hyperparathyroidism is the most common cause of asymptomatic hypercalcemia. Hence, many of the classic manifestations, particularly those referable to bone and renal disease, are now seen infrequently in clinical practice. Among other causes of hypercalcemia (Table 24.5), malignancy stands out as the most frequent cause of symptomatic hypercalcemia in adults, and must be excluded by appropriate clinical and laboratory investigations. As discussed in Chapter 7, hypercalcemia can occur with solid tumors, such as lung, breast, head and neck, and renal cancers, and with hematologic malignancies, notably multiple myeloma. The most common mechanism (in ~80% of cases) through which osteolytic tumors induce hypercalcemia is by secretion of PTH-related peptide (PTHrP), which (like PTH) induces osteoclastic bone resorption and hypercalcemia; the remaining 20% induce hypercalcemia through metastases to the bone and subsequent cytokine-induced bone resorption. In individuals with primary hyperparathyroidism, serum PTH levels are inappropriately elevated for the level of serum calcium, whereas PTH levels are low to undetectable in hypercalcemia caused by nonparathyroid diseases (see Table 24.5). Radioimmunoassays specific for PTH and PTHrP are available and can be useful in distinguishing primary hyperparathyroidism and malignancy-associated hypercalcemia. Other laboratory alterations referable to PTH excess include hypophosphatemia and increased urinary excretion of both calcium and phosphate. Secondary renal disease may in turn lead to phosphate retention, "normalizing" the serum phosphate levels.
- *Symptomatic primary hyperparathyroidism.* The signs and symptoms of hyperparathyroidism reflect the combined effects of increased PTH secretion and hypercalcemia. Primary hyperparathyroidism is associated with "painful bones, renal stones, abdominal groans, and psychic moans." The constellation of symptoms includes the following:

Table 24.5 Causes of Hypercalcemia

Increased PTH	Decreased PTH
Hyperparathyroidism	Hypercalcemia of malignancy[a]
Primary (adenoma > hyperplasia)[a]	Vitamin D toxicity
Secondary[†]	Immobilization
Tertiary[†]	Thiazide diuretics
Familial hypocalciuric hypercalcemia	Granulomatous disease (sarcoidosis)

PTH, Parathyroid hormone.

[a]Primary hyperparathyroidism is the most common cause of hypercalcemia overall. Malignancy is the most common cause of symptomatic hypercalcemia. Primary hyperparathyroidism and malignancy account for nearly 90% of cases of hypercalcemia.

[†]Secondary and tertiary hyperparathyroidism are most commonly associated with progressive renal failure.

- *Bone disease* and bone pain secondary to fractures of bones weakened by osteoporosis or osteitis fibrosa cystica
- *Nephrolithiasis* (renal stones) in 20% of newly diagnosed patients, with attendant pain and obstructive uropathy. Chronic renal insufficiency and abnormalities in renal function lead to polyuria and secondary polydipsia.
- Gastrointestinal disturbances, including constipation, nausea, peptic ulcers, pancreatitis, and gallstones
- Central nervous system alterations, including depression, lethargy, and eventually seizures
- Neuromuscular abnormalities, including weakness and fatigue
- Cardiac manifestations, including aortic or mitral valve calcifications (or both)

The abnormalities most directly related to hyperparathyroidism are nephrolithiasis and bone disease, whereas those attributable to hypercalcemia include fatigue, weakness, pancreatitis, metastatic calcifications, and constipation.

Secondary Hyperparathyroidism

Secondary hyperparathyroidism is caused by any condition that gives rise to chronic hypocalcemia, which, in turn, leads to compensatory overactivity of the parathyroid glands. Renal failure is by far the most common cause of secondary hyperparathyroidism, although several other diseases, including inadequate dietary intake of calcium, steatorrhea, and vitamin D deficiency, may also cause this disorder. The mechanisms by which chronic renal failure induces secondary hyperparathyroidism are complex and not fully understood. Chronic renal insufficiency is associated with decreased phosphate excretion, which, in turn, results in hyperphosphatemia. The elevated serum phosphate levels directly depress serum calcium levels and thereby stimulate parathyroid gland activity. In addition, loss of renal substance reduces the availability of α_1-hydroxylase, which is necessary for the synthesis of the active form of vitamin D, leading in turn to reduced intestinal absorption of calcium (Chapter 9). Because vitamin D has suppressive effects on parathyroid growth and PTH secretion, its deficiency compounds the hyperparathyroidism in renal failure.

MORPHOLOGY

The parathyroid glands in secondary hyperparathyroidism are hyperplastic. As in primary hyperparathyroidism, the degree of glandular enlargement may be asymmetric. Microscopically, the hyperplastic glands contain an increased number of chief cells, or water-clear cells in a diffuse or multinodular distribution. Fat cells are decreased in number. **Metastatic calcification** may be seen in many tissues, including the lungs, heart, stomach, and blood vessels.

Clinical Features

The clinical features of secondary hyperparathyroidism are usually dominated by the inciting chronic renal failure. Secondary hyperparathyroidism is usually not as severe or as prolonged as primary hyperparathyroidism, hence the skeletal abnormalities (referred to as *renal osteodystrophy)* tend to be milder. Control of the hyperparathyroidism allows the bony changes to regress significantly or disappear completely. The vascular calcification associated with secondary hyperparathyroidism may occasionally result in significant ischemic damage to skin and other organs, a process referred to as *calciphylaxis*. Patients with secondary hyperparathyroidism often respond to dietary vitamin D supplementation as well as phosphate binders, which decrease the prevailing hyperphosphatemia.

In a minority of patients, parathyroid activity may become autonomous and excessive, with resultant hypercalcemia, a process termed *tertiary hyperparathyroidism*. Parathyroidectomy may be necessary to control the hyperparathyroidism in such patients.

KEY CONCEPTS

HYPERPARATHYROIDISM

- Primary hyperparathyroidism is the most common cause of asymptomatic hypercalcemia.
- In a majority of cases, primary hyperparathyroidism is caused by a sporadic parathyroid adenoma and, less commonly, by parathyroid hyperplasia.
- Parathyroid adenomas are solitary, while hyperplasia typically is a multiglandular process.
- Skeletal manifestations of hyperparathyroidism include bone resorption, osteitis fibrosa cystica, and brown tumors. Renal changes include nephrolithiasis (stones) and nephrocalcinosis.
- The clinical manifestations of hyperparathyroidism have been classically summarized as "painful bones, renal stones, abdominal groans, and psychic moans."
- Secondary hyperparathyroidism most often is caused by renal failure, which lowers serum calcium levels, resulting in reactive hyperplasia of parathyroid glands.
- Malignancies are the most important cause of symptomatic hypercalcemia, which results from osteolytic metastases or release of PTH-related protein from nonparathyroid tumors.

HYPOPARATHYROIDISM

Hypoparathyroidism is far less common than is hyperparathyroidism. Acquired hypoparathyroidism is almost always an inadvertent consequence of surgery; in addition, there are several genetic causes of hypoparathyroidism.

- *Surgically induced hypoparathyroidism* occurs with inadvertent removal of all parathyroid glands during thyroidectomy, excision of parathyroid glands in the mistaken belief that they are lymph nodes during radical neck dissection for some form of malignant disease, or removal of too large a proportion of parathyroid tissue in the treatment of primary hyperparathyroidism.
- *Autoimmune hypoparathyroidism* is often associated with chronic mucocutaneous candidiasis and primary adrenal insufficiency; this syndrome is known as APS-1 and is caused by mutations in the *autoimmune regulator (AIRE)* gene. The syndrome typically presents in childhood with the onset of candidiasis, followed several years later by hypoparathyroidism and then adrenal insufficiency during adolescence. APS-1 is discussed further later in this chapter.

- *Autosomal-dominant hypoparathyroidism* is caused by gain-of-function mutations in the *calcium-sensing receptor (CASR)* gene. Inappropriate CASR activity due to heightened calcium sensing suppresses PTH, resulting in hypocalcemia and hypercalciuria. Recall that loss-of-function *CASR* mutations are a rare cause of familial parathyroid adenomas.
- *Familial isolated hypoparathyroidism* (FIH) is a rare condition with either autosomal dominant or autosomal recessive patterns of inheritance. Autosomal dominant FIH is caused by a mutation in the gene encoding PTH that impairs PTH processing to the mature active hormone. Autosomal recessive FIH is caused by loss-of-function mutations in the transcription factor gene *GCM2,* which is essential for development of the parathyroid.
- *Congenital absence* of parathyroid glands can occur in conjunction with other malformations, such as thymic aplasia and cardiovascular defects, or as a component of the 22q11 deletion syndrome. As discussed in Chapter 6, when thymic defects are present, the condition is called *DiGeorge syndrome.*

Clinical Features

The major manifestations of hypoparathyroidism are related to the severity and chronicity of the hypocalcemia.

- The hallmark of hypocalcemia is *tetany,* which is characterized by neuromuscular irritability, resulting from decreased serum calcium levels. The symptoms range from circumoral numbness or paresthesias (tingling) of the distal extremities and carpopedal spasm, to life-threatening laryngospasm and generalized seizures. The classic findings on physical examination are *Chvostek sign* and *Trousseau sign*. Chvostek sign is elicited in subclinical disease by tapping along the course of the facial nerve, which induces contractions of the muscles of the eye, mouth, or nose. Trousseau sign refers to carpal spasms produced by occlusion of the circulation to the distal arm with a blood pressure cuff for several minutes.
- *Mental status changes* include emotional instability, anxiety and depression, confusional states, hallucinations, and frank psychosis.
- *Intracranial manifestations* include calcifications of the basal ganglia, parkinsonian-like movement disorders, and increased intracranial pressure with resultant papilledema. The paradoxical association of hypocalcemia with calcifications may be because of an increase in phosphate levels, leading to deposition of calcium phosphate in vulnerable tissues.
- *Ocular disease* takes the form of calcification of the lens and cataract formation.
- *Cardiovascular manifestations* include a conduction defect that produces a characteristic prolongation of the QT interval in the electrocardiogram.
- *Dental abnormalities* occur when hypocalcemia is present during early development. These findings are highly characteristic of hypoparathyroidism and include dental hypoplasia, failure of eruption, defective enamel and root formation, and abraded carious teeth.

Pseudohypoparathyroidism

In this condition, hypoparathyroidism occurs because of end-organ resistance to the actions of PTH. Indeed, serum PTH levels are normal or elevated. In one form of pseudohypoparathyroidism, there is end-organ resistance to TSH and FSH/LH as well as PTH. All of these hormones signal via G-protein–coupled receptors, and the disorder results from genetic defects in components of this pathway that are shared across endocrine tissues. PTH resistance is the most obvious clinical manifestation. It presents as hypocalcemia, hyperphosphatemia, and elevated circulating PTH. TSH resistance is generally mild, while LH/FSH resistance manifests as hypergonadotropic hypogonadism in females.

The Endocrine Pancreas

The endocrine pancreas consists of about 1 million clusters of cells, the *islets of Langerhans,* which contain four major and two minor cell types. The four main types are β, α, δ, and PP (pancreatic polypeptide) cells. They can be differentiated by the ultrastructural characteristics of their granules, and by their hormone content (Fig. 24.27).

- The *β cells* produce insulin, which regulates glucose utilization in tissues and reduces blood glucose levels, as will be detailed in the discussion of diabetes.
- The *α cells* secrete glucagon, which stimulates glycogenolysis in the liver and thus increases blood sugar.
- The *δ cells* secrete somatostatin, which suppresses both insulin and glucagon release.
- The *PP cells* secrete pancreatic polypeptide, which exerts several gastrointestinal effects, such as stimulation of secretion of gastric and intestinal enzymes and inhibition of intestinal motility. These cells not only are present in islets but also are scattered throughout the exocrine pancreas.
- There are also two rare cell types, *D1 cells* and *enterochromaffin cells*. D1 cells elaborate vasoactive intestinal polypeptide (VIP), a hormone that induces glycogenolysis and hyperglycemia; it also stimulates gastrointestinal fluid secretion and causes secretory diarrhea.
- *Enterochromaffin cells* synthesize serotonin and are the source of pancreatic tumors that cause the carcinoid syndrome (Chapter 19).

The following discussion focuses on the two main disorders of islet cells: diabetes mellitus and pancreatic endocrine tumors.

DIABETES MELLITUS

Diabetes mellitus is a group of metabolic disorders sharing the common feature of hyperglycemia caused by defects in insulin secretion, insulin action, or, most

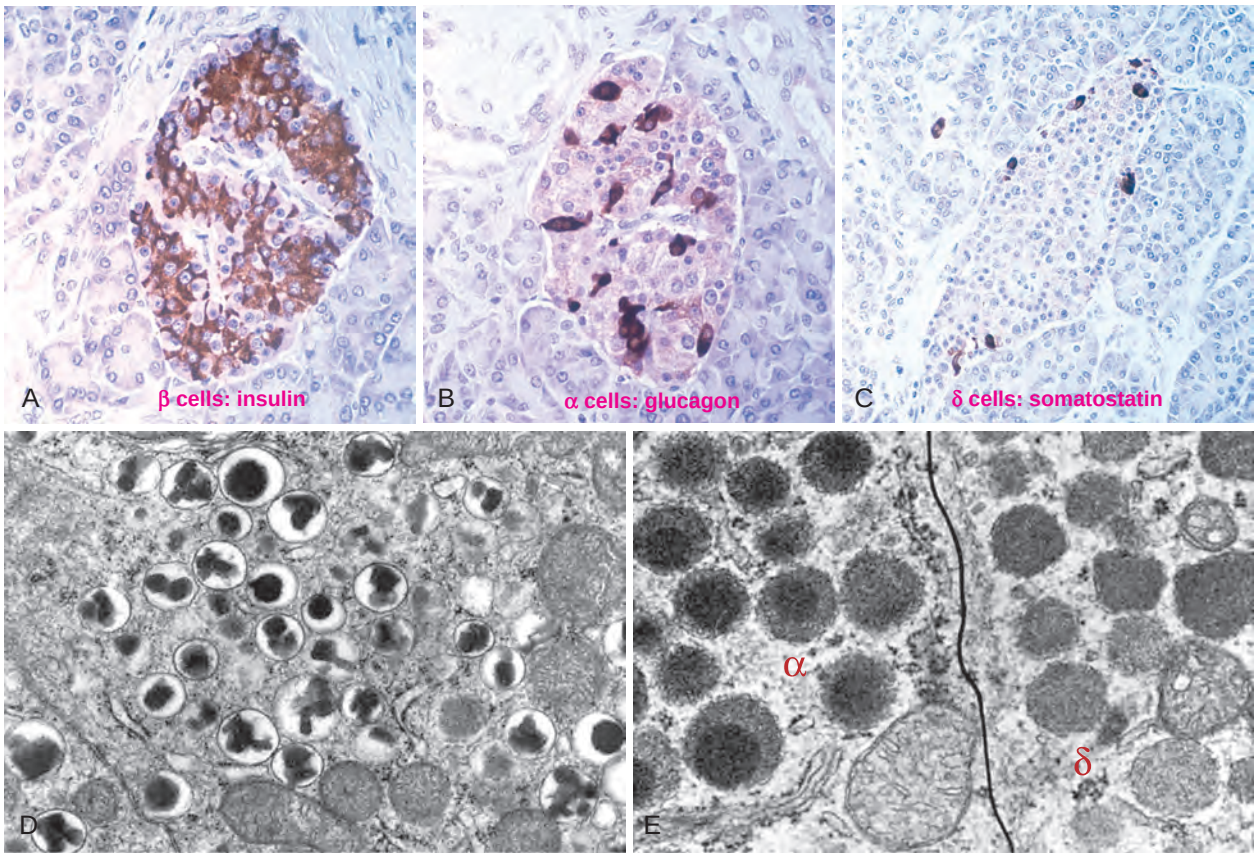


Figure 24.27 Hormone production in pancreatic islet cells. Immunoperoxidase staining shows insulin in β cells (A), glucagon in α cells (B), and somatostatin in δ cells (C). (D) Electron micrograph of a β cell shows the characteristic membrane-bound granules, each containing a dense, often rectangular core and distinct halo. (E) Portions of an α cell *(left)* and a δ cell *(right)* also show granules, but with closely apportioned membranes. The α-cell granule shows a dense, round center. (Electron micrographs courtesy Dr. Arthur Like, University of Massachusetts Medical School, Worcester, Mass.)

commonly, both. The chronic hyperglycemia and attendant metabolic deregulation may be associated with secondary damage in multiple organ systems, especially the kidneys, eyes, nerves, and blood vessels. **In the United States, diabetes is the leading cause of end-stage renal disease, adult-onset blindness, and nontraumatic lower extremity amputations**.

Diabetes and related disorders of glucose metabolism are common. According to the American Diabetes Association, diabetes affects more than 30 million children and adults, or more than 9% of the population, in the United States, of which about 1.2 million have the form of diabetes called type 1 and the remainder have type 2. Astonishingly, nearly one-fourth of these individuals are currently unaware that they have hyperglycemia. Approximately 1.5 million new cases of adult diabetes are diagnosed each year in the United States. Furthermore, a staggering 84 million adults in this country have impaired glucose tolerance or "prediabetes," which is defined as elevated blood sugar that does not reach the criterion used for an outright diagnosis of type 2 diabetes (T2D; see later), and individuals with prediabetes are at risk for developing frank T2D. Compared to non-Hispanic Caucasians, Native Americans, African Americans, and Hispanics are 1.5 to 2 times more likely to develop diabetes in their lifetimes. The World Health Organization estimates that as many as 422 million people suffer from diabetes worldwide, with India and China being the largest contributors to the world's diabetic burden. Increasingly sedentary lifestyles and poor eating habits have contributed to the simultaneous escalation of T2D and obesity, which some have called the *diabesity epidemic*. Sadly, this epidemic has now spread to children living in "food deserts" who subsist on highly processed foods rich in carbohydrates and sugar and do not exercise adequately.

The mortality rate from diabetes varies across countries, with middle- and low-income nations accounting for almost 80% of diabetes-related deaths and nearly double the mortality rates observed in high-income nations. Nonetheless, diabetes remains in the top 10 "killers" in the United States. The total yearly cost related to diabetes in the United States is estimated to be an astounding $327 billion, including $237 billion in direct medical costs and $90 billion in indirect costs stemming from the reduced productivity of individuals with diabetes.

Diagnosis

Blood glucose is normally maintained in a very narrow range of 70 to 120 mg/dL. According to the ADA and WHO, diagnostic criteria for diabetes include the following:

1. A fasting plasma glucose ≥126 mg/dL
2. A random plasma glucose ≥200 mg/dL (in a patient with classic hyperglycemic signs, as discussed later)

3. A 2-hour plasma glucose ≥200 mg/dL during an oral glucose tolerance test (OGTT) with a loading dose of 75 g
4. A glycated hemoglobin (HbA1c) level ≥6.5% (glycated hemoglobin is further discussed later in the chapter)

All tests, except the random blood glucose test in a patient with classic hyperglycemic signs, need to be repeated and confirmed on a separate day. If there is discordance between two assays (e.g., fasting glucose and HbA1c level), the result with the greatest degree of abnormality is considered the "readout." Of note, many acute stresses, such as severe infections, burns, or trauma, can lead to transient hyperglycemia due to secretion of hormones such as catecholamines and cortisol that oppose the action of insulin. The diagnosis of diabetes requires persistence of hyperglycemia following resolution of the acute illness.

Prediabetes, a state of dysglycemia that often precedes development of frank T2D, is defined by one or more of the following:

1. A fasting plasma glucose between 100 and 125 mg/dL ("impaired fasting glucose"),
2. A 2-hour plasma glucose between 140 and 199 mg/dL following a 75-g oral glucose tolerance test (OGTT) ("impaired glucose tolerance"), and/or
3. A glycated hemoglobin (HbA1c) level between 5.7% and 6.4%

As many as one-fourth of individuals with impaired glucose tolerance will develop overt diabetes over 5 years, with additional factors such as obesity and family history compounding the risk. In addition, individuals with prediabetes also are at significant risk for cardiovascular complications.

Classification

Although all forms of diabetes have hyperglycemia as a common feature, the underlying abnormalities involved in its development vary widely. The previous classification schemes of diabetes were based on clinical features, such as the age of onset of disease and the mode of therapy; in contrast, the current classification reflects our greater understanding of the pathogenesis of each variant (Table 24.6). The vast majority of cases of diabetes fall into one of two broad classes:

- *Type 1 diabetes (T1D)* is an autoimmune disease characterized by pancreatic β-cell destruction and an absolute deficiency of insulin. It accounts for approximately 5% to 10% of diabetes and is the most common subtype diagnosed in patients younger than 20 years of age.
- *Type 2 diabetes (T2D)* is caused by a combination of peripheral resistance to insulin action and a secretory response by pancreatic β cells that is inadequate to overcome insulin resistance ("relative insulin deficiency"). Approximately 90% to 95% of diabetes patients have T2D, and the vast majority of such individuals are overweight. Although classically considered "adult-onset," the prevalence of T2D in children and adolescents has been increasing at an alarming pace due to the increasing rates of obesity in children and young adults, particularly in Hispanic, Native American, and Asian ethnic groups.

Table 24.6 Classification of Diabetes

Type 1 Diabetes (β-Cell Destruction, Usually Leading to Absolute Insulin Deficiency)
Immune-mediated
Idiopathic (autoantibody-negative)
Type 2 Diabetes (Combination of Insulin Resistance and β-Cell Dysfunction)
Other Types
Genetic Defects of β-Cell Function
Maturity-onset diabetes of the young (MODY) caused by mutations in:
Hepatocyte nuclear factor 4α *(HNF4A)* (MODY1)
Glucokinase *(GCK)* (MODY2)
Hepatocyte nuclear factor 1α *(HNF1A)*, (MODY3)
Pancreatic and duodenal homeobox 1 *(PDX1)* (MODY4)
Hepatocyte nuclear factor 1β *(HNF1B)* (MODY5)
Neurogenic differentiation factor 1 *(NEUROD1)* (MODY6)
Neonatal diabetes (activating mutations in *KCNJ11* and *ABCC8*, encoding Kir6.2 and SUR1, respectively)
Maternally inherited diabetes and deafness (MIDD) due to mitochondrial DNA mutations (m.3243A→G)
Defects in proinsulin conversion
Insulin gene mutations
Genetic Defects in Insulin Action
Type A insulin resistance
Lipoatrophic diabetes
Exocrine Pancreatic Defects ("Pancreatogenic" or Type 3C Diabetes)
Chronic pancreatitis
Pancreatectomy/trauma
Pancreatic cancer
Cystic fibrosis
Hemochromatosis
Fibrocalculous pancreatopathy
Endocrinopathies
Acromegaly
Cushing syndrome
Hyperthyroidism
Pheochromocytoma
Glucagonoma
Infections
Cytomegalovirus
Coxsackie B virus
Congenital rubella
Drugs
Glucocorticoids
Thyroid hormone
Interferon-α
Protease inhibitors
β-adrenergic agonists
Thiazides
Nicotinic acid
Phenytoin (Dilantin)
Vacor
Genetic Syndromes Associated With Diabetes
Down syndrome
Klinefelter syndrome
Turner syndrome
Prader-Willi syndrome
Gestational Diabetes Mellitus

Modified from American Diabetes Association: Diagnosis and classification of diabetes mellitus, *Diabetes Care* 37(Suppl 1):S81–S90, 2014.

Table 24.7 Comparative Features of Type 1 and Type 2 Diabetes

Type 1 Diabetes	Type 2 Diabetes
Clinical	
Onset: usually childhood and adolescence	Onset: usually adult; increasing incidence in childhood and adolescence
Normal weight or weight loss preceding diagnosis	Vast majority are obese (80%)
Progressive decrease in insulin levels	Increased blood insulin (early); normal or moderate decrease in insulin (late)
Circulating islet autoantibodies (anti-insulin, anti-GAD, anti-ICA512)	No islet autoantibodies
Diabetic ketoacidosis in absence of insulin therapy	Nonketotic hyperosmolar coma more common
Genetics	
Major linkage to MHC class II genes; also linked to polymorphisms in *CTLA4* and *PTPN22*, and insulin gene VNTRs	No HLA linkage; linkage to candidate diabetogenic and obesity-related genes (e.g., *TCF7L2*, *PPARG*, *FTO*)
Pathogenesis	
Dysfunction in T-cell selection and regulation leading to breakdown in self-tolerance to islet autoantigens	Insulin resistance in peripheral tissues, failure of compensation by β cells
Pathology	
Insulitis (inflammatory infiltrate of T cells and macrophages) β-cell depletion, islet atrophy	No insulitis; amyloid deposition in islets Mild β-cell depletion

HLA, Human leukocyte antigen; *MHC*, major histocompatibility complex; *VNTRs*, variable number of tandem repeats.

The important similarities and differences between T1D and T2D are summarized in Table 24.7. A variety of monogenic and secondary causes are responsible for the remaining cases (discussed later). Before discussing the pathogenesis of the two major types, we will first briefly review normal insulin secretion and the mechanism of insulin action, since these are critical to understanding the pathogenesis of diabetes.

Glucose Homeostasis

Glucose homeostasis is tightly regulated by three interrelated processes: glucose production in the liver; glucose uptake and utilization by peripheral tissues, chiefly skeletal muscle; and actions of insulin and counter-regulatory hormones, including glucagon, on glucose uptake and metabolism. Insulin and glucagon have opposing effects on glucose homeostasis. During fasting states, low insulin and high glucagon levels facilitate hepatic gluconeogenesis and glycogenolysis (glycogen breakdown) while decreasing glycogen synthesis, thereby preventing hypoglycemia. Thus, fasting plasma glucose levels are determined primarily by hepatic glucose output. Following a meal, insulin levels rise and glucagon levels fall in response to the large glucose load. Insulin promotes glucose uptake and utilization in tissues (discussed later). The skeletal muscle is the major

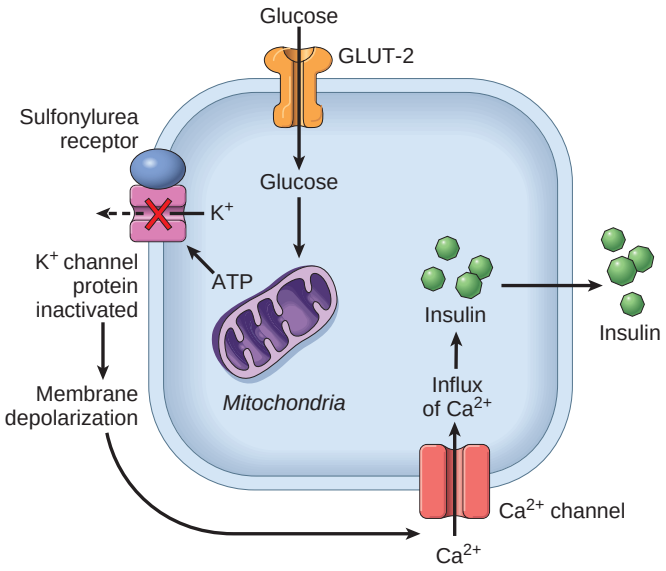


Figure 24.28 Insulin synthesis and secretion. The influx of glucose into β cells through the GLUT-2 receptors initiates a cascade of signaling events that culminates in Ca^{2+}-induced release of stored insulin (see text for details).

insulin-responsive site for postprandial glucose utilization, and it is critical for preventing hyperglycemia and maintaining glucose homeostasis. Although less dependent on insulin, brain and adipose tissues also extract a significant amount of glucose from the circulation.

Regulation of Insulin Release

Insulin is produced in the β cells of the pancreatic islets (see Fig. 24.27) as a precursor protein and is proteolytically cleaved in the Golgi complex to generate the mature hormone and a peptide byproduct, *C-peptide*. Both insulin and C-peptide are then stored in secretory granules and secreted in equimolar quantities after physiologic stimulation; thus, C-peptide levels serve as a surrogate for β-cell function, decreasing with loss of β-cell mass in T1D and increasing with insulin resistance–associated hyperinsulinemia.

The most important stimulus for insulin synthesis and release is glucose. An increase in blood glucose levels results in glucose uptake into pancreatic β cells, facilitated by an insulin-independent glucose transporter, GLUT2 (Fig. 24.28). Metabolism of glucose generates ATP, which leads to the influx of Ca^{2+} through plasma membrane calcium channels. The resultant increase in intracellular Ca^{2+} stimulates secretion of insulin, presumably from hormone stored in β-cell granules. This is the phase of immediate insulin release, sometimes called the first phase of β-cell insulin secretion. If the secretory stimulus persists, a delayed and protracted response follows that involves active synthesis of insulin, the second phase.

Oral intake of food leads to secretion of multiple hormones that play a role in glucose homeostasis and satiety. Of these, the most important class of hormones responsible for promoting insulin secretion from pancreatic β cells following feeding is the *incretins*, which act by binding G-protein–coupled receptors that are expressed on pancreatic β cells. The two most important incretins are glucose-dependent insulinotropic polypeptide (GIP) and

glucagon-like peptide-1 (GLP-1), both secreted by cells in the intestines following oral food intake. The elevation in GIP and GLP-1 levels is known as the "incretin effect." In addition to increasing insulin secretion from β cells, these hormones reduce glucagon secretion from pancreatic α cells and delay gastric emptying, which promotes satiety. Once released, circulating GIP and GLP-1 are degraded in the circulation by a class of enzymes known as dipeptidyl peptidases (DPPs), especially DPP-4. The "incretin effect" is significantly blunted in patients with T2D, and efforts to restore incretin function can improve glycemic control and promote weight loss (through restoration of satiety). These insights have led to the recent development of two classes of drugs for treating T2D: GLP-1 receptor agonists, which are synthetic GLP-1 mimetics that bind to and activate the GLP-1 receptor on islet and extrapancreatic cells; and DPP-4 inhibitors, which enhance levels of endogenous incretins by delaying their degradation. GLP-1 also increases energy expenditure, so the weight-loss–inducing effects are likely multifactorial. Indeed, GLP-1 receptor agonists are also now approved for treatment of obesity.

Insulin Action and Insulin-Signaling Pathways

Insulin is the most potent anabolic hormone known, with multiple synthetic and growth-promoting effects (Fig. 24.29). The principal metabolic function of insulin is to increase the rate of glucose transport into certain cells in the body, thus providing a major source of energy and metabolic intermediates derived from glucose that are used in the biosynthesis of cellular building blocks such as lipids, nucleotides, and amino acids. The most important targets of insulin action are striated muscle cells (including cardiomyocytes) and, to a lesser extent, adipocytes, which together normally represent about two-thirds of the body's weight. The type of adipose tissue that utilizes the most glucose is "beige" adipose tissue, which develops with exercise, and not the "white" adipose tissue that accumulates in obese individuals. This is one reason why exercise is beneficial and obesity detrimental to glucose control. In muscle cells, glucose is either stored as glycogen or oxidized to generate ATP. In adipose tissue, glucose is primarily used as a substrate for synthesis of lipids, which are stored as triglycerides. Besides promoting lipid synthesis, insulin also inhibits triglyceride hydrolysis and lipid release by adipocytes. Similarly, insulin promotes amino acid uptake and protein synthesis, while inhibiting protein degradation. Thus, the anabolic effects of insulin are attributable to increased synthesis and reduced degradation of glycogen, lipids, and proteins. In addition, insulin has several mitogenic activities, including initiation of DNA synthesis in certain cells and stimulation of their growth and differentiation.

The metabolic effects of insulin are exerted through its binding to the insulin receptor, which, in turn, sets into motion a series of signaling events through an array of mediators, the more pertinent of which are summarized in Fig. 24.30. The insulin receptor is a tetrameric protein composed of two α-subunits and two β-subunits. The β-subunit cytosolic domain possesses tyrosine kinase activity. Insulin binding to the α-subunit extracellular domain activates the β-subunit tyrosine kinase, which autophosphorylates itself and also phosphorylates several intracellular docking or bridging proteins, including so-called insulin receptor substrate (IRS) proteins. These molecules in turn activate downstream factors such as PI-3-kinase and Akt, a serine/threonine kinase that serves as a central signaling hub that mediates many insulin-dependent activities, including increased glucose uptake, reduced glucose synthesis, and increased glycogen and protein synthesis.

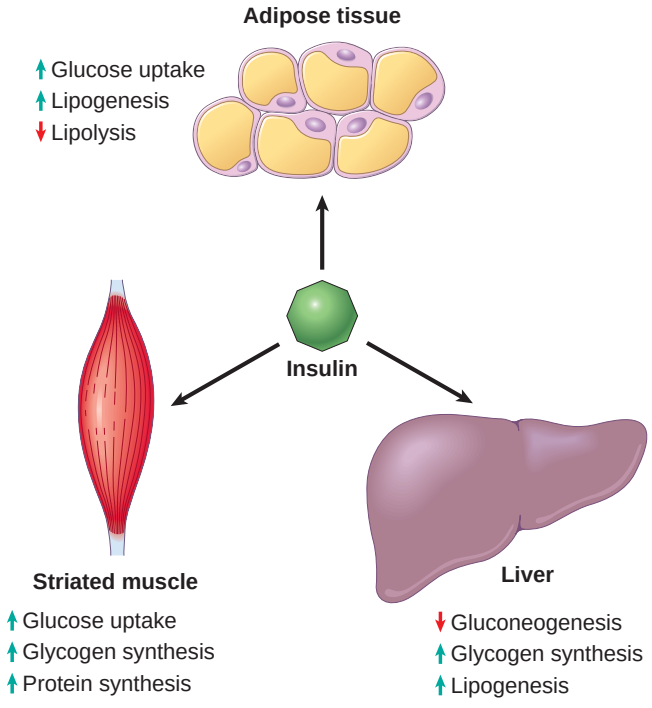


Figure 24.29 Metabolic actions of insulin in striated muscle, adipose tissue, and liver.

Pathogenesis of Type 1 Diabetes

T1D is an autoimmune disease in which islet destruction is caused primarily by immune effector cells reacting against endogenous β-cell antigens. T1D most commonly develops in childhood, becomes manifest at puberty, and progresses with age. Because the disease can develop at any age, including late adulthood, the old moniker "juvenile-onset diabetes" is no longer used. Similarly, "insulin-dependent diabetes mellitus" has been excluded from the current classification of diabetes because many forms of diabetes eventually require treatment with insulin. Nevertheless, most patients with T1D require insulin for survival; without insulin, they may develop serious metabolic complications such as ketoacidosis and coma.

As with most autoimmune diseases, the pathogenesis of T1D involves an interplay of genetic and environmental factors.

Genetic Susceptibility

Epidemiologic studies, such as those demonstrating higher concordance rates in monozygotic versus dizygotic twins, have convincingly established a genetic basis for T1D. More recently, GWAS have identified multiple genetic susceptibility loci for T1D, as well as for T2D (see later). **Of these, the most important locus is the *HLA* gene cluster, which according to some estimates contributes as much as 50% of the genetic susceptibility for T1D.** Ninety percent to 95%

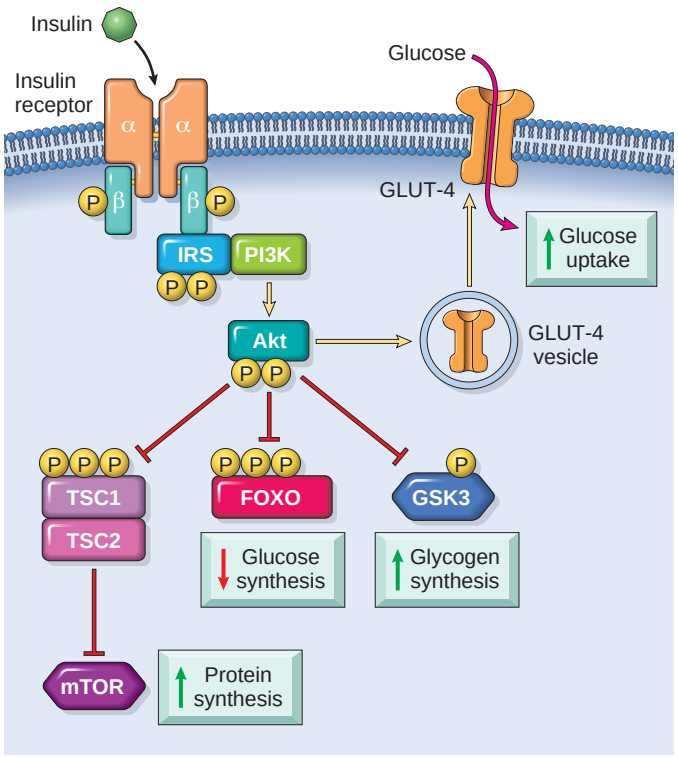


Figure 24.30 Insulin action on a target cell. Insulin binding to the tetrameric receptor initiates a cascade of phosphorylation events that result in activation of PI-3-kinase/Akt signaling. Akt is a serine threonine kinase that mediates its effector functions via phosphorylation-dependent events. For example, Akt phosphorylates and inhibits the function of the tuberous sclerosis complex *(TSC)* proteins, leading to activation of the downstream mammalian TOR *(mTOR)* complex, which enhances protein synthesis. Akt also inhibits the function of Forkhead box O *(FOXO)* protein, which, in turn, reduces glucose synthesis, while inhibition of glycogen synthase kinase 3 *(GSK3)* enhances glycogen production. Finally, Akt enhances intracellular glucose uptake by translocation of GLUT-4 vesicles to the cell membrane. *IRS*, Insulin receptor substrate; *PI3K*, phosphoinositide 3-kinase. (Modified from Brendan Manning, Harvard T.H. Chan School of Public Health.)

of Caucasians with this disease have either an *HLA-DR3* or *HLA-DR4* haplotype, in contrast to about 40% of normal subjects; moreover, 40% to 50% of patients with T1D are *DR3/DR4* compound heterozygotes, in contrast to 5% of normal subjects. Individuals who have either *DR3* or *DR4* concurrently with a *DQ8* haplotype demonstrate one of the highest inherited risks for T1D in sibling studies. Predictably, the polymorphisms in the HLA molecules that are associated with risk are located in or adjacent to the peptide-binding pockets, consistent with the notion that disease-associated alleles code for HLA molecules that have the capacity to display particular antigens. However, as discussed in Chapter 6, it is still not known how particular HLA alleles contribute to the pathogenesis of T1D (and other autoimmune diseases).

Several *non-HLA genes* also confer susceptibility to T1D. The first disease-associated non-*MHC* gene variant to be identified consisted of variable number of tandem repeats in the promoter region of the *insulin* gene. The mechanism underlying this association is unknown. It is possible that these polymorphisms influence insulin expression by thymic antigen-presenting cells, thus affecting the negative selection of insulin-reactive T cells (Chapter 6). The association between polymorphisms in *CTLA4* and *PTPN22* and autoimmune thyroiditis was mentioned earlier; not surprisingly, these genes have also been linked with susceptibility to T1D. The relationship of T1D to altered T-cell selection and regulation is also underscored by the striking prevalence of this disease in individuals with rare germline defects in genes that code for immune regulators, such as *AIRE*, mutations of which cause APS-1 (discussed later).

Environmental Factors

As in other autoimmune diseases, genetic susceptibility contributes to only a part of diabetes risk, and the concordance rate in monozygotic twins is only about 50%, so environmental factors must play a role. The nature of these environmental influences remains an enigma. Although antecedent viral infections have been suggested as triggers, neither the type of virus nor how it promotes islet-specific autoimmunity is established. Some studies suggest that viruses might share epitopes with islet antigens, and the immune response to the virus results in cross-reactivity and destruction of islet tissues, a phenomenon known as *molecular mimicry*. On the other hand, certain infections are also thought to be protective against T1D.

Mechanisms of β-Cell Destruction

While the clinical onset of T1D is often abrupt, there is a lengthy lag period between initiation of the autoimmune process and the appearance of symptomatic disease, during which there is progressive loss of insulin reserves. Three distinct stages of T1D are now recognized (Fig. 24.31). In stage 1 *(autoimmunity positive, normoglycemia, presymptomatic T1D)*, individuals have developed two or more islet autoantibodies but are still normoglycemic. In stage 2 *(autoimmunity positive, dysglycemia, presymptomatic T1D)*, there is increasingly severe loss of glucose tolerance due to progressive loss of β-cell mass, but frank symptoms are absent. Nonetheless, the 5-year risk of developing symptomatic T1D increases from less than 50% in stage 1 to 75% in stage 2. Finally, in stage 3 *(autoimmunity positive, dysglycemia, symptomatic T1D)*, classic manifestations of the disease (polyuria, polydipsia, polyphagia, ketoacidosis; see later) appear, typically after more than 90% of the β cells have been destroyed.

The fundamental immune abnormality in T1D is a failure of self-tolerance in T cells specific for islet antigens. This failure of tolerance may be a result of some combination of defective clonal deletion of self-reactive T cells in the thymus, as well as defects in the functions of regulatory T cells or abnormal resistance of effector T cells to suppression by regulatory cells. Thus, autoreactive T cells not only survive, but are poised to respond to self antigens. The initial activation of these cells is thought to occur in the peripancreatic lymph nodes, perhaps in response to antigens that are released from damaged islets. The activated T cells then traffic to the pancreas, where they cause β-cell injury. Multiple T-cell populations have been implicated in this damage, including Th1 cells (which may secrete cytokines, including IFN-γ and TNF, that injure β cells), and CD8+ CTLs (which kill β cells directly). The islet autoantigens that are the targets of immune attack may include insulin, the β-cell enzyme glutamic acid decarboxylase (GAD), and others. Consistent with the idea that failure of self-tolerance is fundamental to the pathogenesis of T1D, cancer patients

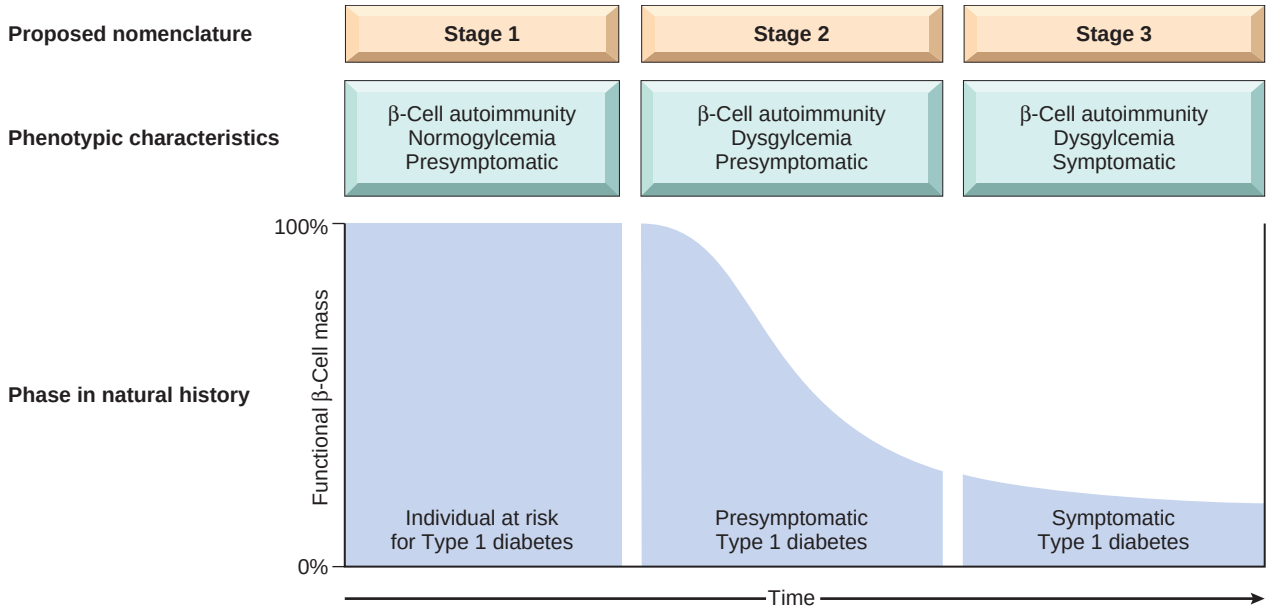


Figure 24.31 The three stages of type 1 diabetes. (Modified with permission from Insel RA, Dunne JL, Atkinson MA, et al: Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association, *Diabetes Care* 38(10):1964–1974, 2015.)

treated with immune checkpoint blockade therapy, which disrupts tolerance mechanisms, sometimes develop the disease.

A role for antibodies in T1D is suspected because autoantibodies against islet antigens are found in the vast majority of patients with T1D, including at the presymptomatic stages of disease, as described earlier. However, it is not clear if the autoantibodies cause injury or are merely a consequence of islet injury.

Pathogenesis of Type 2 Diabetes

T2D is a complex disease that involves the interplay of genetic and environmental factors and a pro-inflammatory state. Unlike T1D, there is no evidence of an autoimmune basis.

Genetic Factors

Genetic susceptibility contributes to the pathogenesis, as evidenced by the disease concordance rate of greater than 90% in monozygotic twins, a rate higher than in T1D. Furthermore, first-degree relatives have 5- to 10-fold higher risk of developing T2D than those without a family history, when matched for age and weight. GWAS performed over the past decade have identified at least 30 loci that individually confer a minimal to modest increase in the lifetime risk for T2D. Many of these genes are involved in adipose tissue function (through effects on bodily fat distribution [visceral vs. subcutaneous]), islet β-cell function, and obesity. It is believed that together, these genetic polymorphisms conspire to provide the genetic basis for T2D risk. However, heritable risk remains a minor player impacting disease susceptibility, and environmental factors are the major contributors.

Environmental Factors

The most important environmental risk factor for T2D is obesity, particularly central or visceral obesity. Greater than 80% of individuals with T2D are obese, and the incidence of diabetes worldwide has risen in proportion to obesity. Obesity contributes to the cardinal metabolic abnormalities of diabetes (see later) and to insulin resistance early in disease. In fact, even modest weight loss through dietary modifications can reduce insulin resistance and improve glucose tolerance. A sedentary lifestyle (typified by lack of exercise) is another risk factor for diabetes, independent of obesity. Weight loss and exercise usually increase insulin sensitivity additively and are often first-line interventions in patients with milder T2D. The combination of obesity, hyperglycemia, increased serum cholesterol and triglycerides, and hypertension is called the *metabolic syndrome*.

Despite this general risk, several populations worldwide in which T2D rates are increasing most rapidly (e.g., East Asian, South Asian, and Middle Eastern) do not show comparable increases in obesity (increased body mass index [BMI], a measure of total body fat). This has suggested that risk is related not only to the amount of body fat but also to its anatomic distribution, as discussed later.

Sleep disorders (such as obstructive sleep apnea) and circadian disruption are additional environmental risk factors for T2D. Circadian disruption is defined as misalignment between the endogenous circadian rhythm and the cycle or rhythm created by individual behaviors. Those at risk for circadian disruption include shift workers and those with sleep disorders or other conditions that restrict nighttime sleep and daytime wakefulness. Studies have shown that circadian disruption impairs glucose homeostasis by affecting both insulin secretion and insulin action. In addition, GWAS have shown an association between circadian-controlled genes and T2D. Disruption of "clock" genes not only affects insulin secretion and action but also activity level and feeding behaviors, resulting in increased risk for hyperglycemia and diabetes.

Metabolic Defects in Type 2 Diabetes

The development of T2D involves two key abnormalities:

- *Insulin resistance:* Decreased response of peripheral tissues, especially skeletal muscle, adipose tissue, and liver, to insulin
- *β-cell dysfunction:* Inadequate insulin secretion in the face of insulin resistance and hyperglycemia

Insulin resistance predates the development of hyperglycemia and is usually accompanied by compensatory β-cell hyperfunction and hyperinsulinemia in the early stages of the evolution of T2D (Fig. 24.32). Over time, the inability of β cells to adapt to increasing secretory needs for maintaining a euglycemic state results in chronic hyperglycemia and the resulting long-standing complications of diabetes.

Insulin Resistance

Insulin resistance is the failure of target tissues to respond normally to insulin. The liver, skeletal muscle, and adipose tissue are the major tissues where insulin resistance gives rise to abnormal glucose tolerance. Insulin resistance results in the following:

- Failure to inhibit endogenous glucose production (gluconeogenesis) in the liver, which contributes to high fasting blood glucose levels
- Failure of glucose uptake and glycogen synthesis to occur in skeletal muscle following a meal, which contributes to high postprandial blood glucose level
- Failure to inhibit activation of "hormone-sensitive" lipases in adipose tissue, leading to excess triglyceride breakdown in adipocytes and high levels of circulating free fatty acids (FFAs).

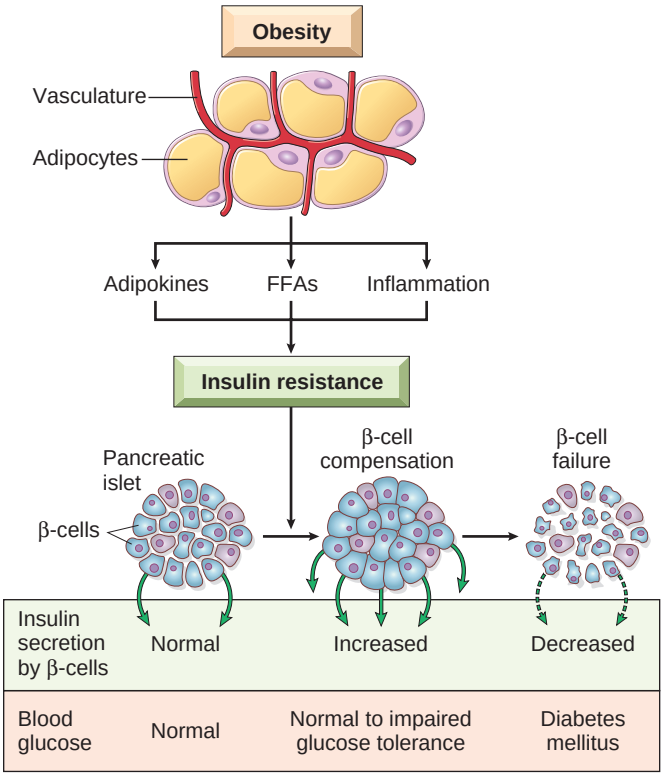


Figure 24.32 Development of type 2 diabetes. Insulin resistance associated with obesity is induced by adipokines, free fatty acids *(FFAs)*, and chronic inflammation in adipose tissue. Pancreatic β cells compensate for insulin resistance by hypersecretion of insulin. However, at some point, β-cell compensation is followed by β-cell failure, and diabetes ensues. (Reproduced with permission from Kasuga M: Insulin resistance and pancreatic β-cell failure, *J Clin Invest* 116(7):1756–1760, 2006.)

A variety of functional defects in the insulin-signaling pathway underlie insulin resistance. For example, reduced tyrosine phosphorylation of the insulin receptor and IRS proteins is observed in peripheral tissues, which compromises insulin signaling and reduces the level of the glucose transporter GLUT-4 on the cell surface (see Fig. 24.30). In fact, one of the mechanisms by which exercise improves insulin sensitivity is by increasing the translocation of GLUT-4 to the plasma membrane of skeletal muscle cells.

Obesity

Multiple factors contribute to insulin resistance, of which obesity is probably the most important. The risk for diabetes rises as the BMI increases. Not only the absolute amount of fat but also its distribution determines insulin sensitivity: central obesity (abdominal fat) is more likely to be linked with insulin resistance than is peripheral (gluteal/subcutaneous) obesity. In fact, individuals from Asia and the Middle East who develop diabetes without overt obesity have primarily visceral adiposity, and it is this increase in visceral fat that appears to engender T2D risk for them. By contrast, individuals who develop primarily subcutaneous adiposity may be relatively protected from T2D. Studies of these so-called "metabolically healthy obese" individuals is an emerging field.

Obesity can adversely impact insulin sensitivity in numerous ways (see Fig. 24.32):

- *Free fatty acids (FFAs).* Cross-sectional studies have demonstrated an inverse correlation between fasting plasma FFAs and insulin sensitivity. Central adipose tissue is more "lipolytic" than peripheral sites, which might explain the particularly deleterious consequences of this pattern of fat distribution. Excess FFAs overwhelm the intracellular fatty acid oxidation pathways, leading to the accumulation of cytoplasmic intermediates like diacylglycerol (DAG), phospholipids, and sphingolipids, including ceramides. These "toxic" lipid metabolites can attenuate signaling through the insulin receptor and activate inflammatory pathways in the islets, which further promote β-cell abnormalities. In liver cells, insulin normally inhibits gluconeogenesis by blocking the activity of phosphoenolpyruvate carboxykinase, the first enzymatic step in this process. Attenuated insulin signaling allows phosphoenolpyruvate carboxykinase to "ramp up" gluconeogenesis. Excess FFAs also compete with glucose as substrates for oxidation, further exacerbating the reduced glucose utilization.
- *Adipokines.* You will recall that adipose tissue is not merely a storage depot for fat but is also an endocrine organ that releases hormones in response to changes in metabolism (Chapter 9). A variety of proteins secreted into the circulation by adipose tissue have been identified that are collectively termed *adipokines* (or adipose cytokines). Some of these promote hyperglycemia, while others (such

as leptin and adiponectin) decrease blood glucose, in part by increasing insulin sensitivity in peripheral tissues. Adiponectin levels are reduced in obesity, thus contributing to insulin resistance.

- *Inflammation:* Over the past several years, inflammation has emerged as an important factor in the pathogenesis of T2D. It is now known that an inflammatory milieu—mediated not by an autoimmune process, as in T1D, but rather by pro-inflammatory cytokines that are secreted in response to excess nutrients such as FFAs and glucose—results in both insulin resistance and β-cell dysfunction. Excess FFAs within macrophages and β cells can activate the inflammasome, a multiprotein cytoplasmic complex that leads to secretion of the cytokine interleukin IL-1β (Chapter 3). IL-1β, in turn, mediates the secretion of additional pro-inflammatory cytokines from macrophages, islet cells, and other cells. IL-1 and other cytokines act on the major sites of insulin action to promote insulin resistance. Thus, excess FFAs can impede insulin signaling directly within peripheral tissues, as well as indirectly through the release of pro-inflammatory cytokines.
- *Liver steatosis:* High circulating levels of FFAs may result in the accumulation of excess fat (steatosis) in hepatocytes. This form of nonalcoholic fatty liver disease (NAFLD) ranges in severity from hepatic steatosis without evidence of liver injury to nonalcoholic steatohepatitis (NASH) with evidence of inflammation and hepatocyte injury with or without fibrosis (Chapter 18). NAFLD is common in those with metabolic syndrome and T2D, an association that cuts both ways: NAFLD promotes the development of T2D, which in turn increases the risk of developing the more severe forms of NAFLD.

β-Cell Dysfunction

While insulin resistance by itself can lead to impaired glucose tolerance, **β-cell dysfunction is a requirement for the development of overt diabetes.** In contrast to the severe genetic defects in β-cell function that occur in monogenic forms of diabetes (see later), β-cell function actually increases early in the disease process in most patients with "sporadic" T2D as a compensatory measure to counter insulin resistance and maintain euglycemia. Eventually, however, β cells seemingly exhaust their capacity to adapt to the long-term demands posed by insulin resistance, and the hyperinsulinemic state gives way to a state of relative insulin deficiency, that is, insulin levels are deficient for the level of blood glucose.

Several mechanisms have been implicated in promoting β-cell dysfunction in T2D, including the following:

- Excess FFAs that compromise β-cell function and attenuate insulin release *(lipotoxicity)*
- The impact of chronic hyperglycemia *(glucotoxicity)*
- An abnormal *incretin effect,* leading to reduced secretion of GIP and GLP-1, hormones that promote insulin release (see earlier)
- Amyloid deposition within islets. This is a characteristic finding in individuals with long-standing T2D, being present in more than 90% of diabetic islets examined, but it is unclear whether it is a cause or an effect of β-cell "burnout."
- Finally, the impact of genetics cannot be discounted, as many of the polymorphisms associated with an increased lifetime risk for T2D occur in genes that control insulin secretion (see earlier).

Monogenic Forms of Diabetes

Although genetically defined causes of diabetes are uncommon, they have been intensively studied in the hope of gaining insights into the disease. As Table 24.6 illustrates, monogenic forms of diabetes are classified separately from types 1 and 2 diabetes. Monogenic forms of diabetes result from either a primary defect in β-cell function or a defect in insulin receptor signaling (described later).

Genetic Defects in β-Cell Function

Approximately 1% to 2% of patients with diabetes harbor a primary defect in β-cell function that affects either β-cell mass and/or insulin production. This form of monogenic diabetes is caused by a heterogeneous group of genetic defects. The largest subgroup of patients in this category was designated as having "maturity-onset diabetes of the young" (MODY) because of its superficial resemblance to T2D and its occurrence in younger patients. MODY can result from germline loss-of-function mutations in one of six genes (see Table 24.6), of which mutations of *glucokinase (GCK)* are the most common. Glucokinase is a rate-limiting step in oxidative glucose metabolism, which, in turn, is coupled to insulin secretion within islet β cells (see Fig. 24.28).

Genetic Defects That Impair Tissue Response to Insulin

Rare insulin receptor mutations that affect receptor synthesis, insulin binding, or RTK activity can cause severe insulin resistance, accompanied by hyperinsulinemia and diabetes (*type A insulin resistance*). Such patients often show a velvety hyperpigmentation of the skin known as *acanthosis nigricans*. Females with type A insulin resistance also frequently have polycystic ovaries and elevated androgen levels.

Diabetes and Pregnancy

Pregnancy can be complicated by diabetes in one of two settings: when women with preexisting diabetes become pregnant ("pregestational" or overt diabetes) or women who were previously euglycemic develop impaired glucose tolerance and diabetes for the first time during pregnancy ("gestational" diabetes). Approximately 5% to 9% of pregnancies occurring in the United States are complicated by hyperglycemia, and the incidence of both pregestational and gestational diabetes is increasing in the general population. Pregnancy is a "diabetogenic" state in which the prevailing hormonal milieu favors insulin resistance. In a previously euglycemic woman who is otherwise susceptible due to concurrent genetic and environmental factors, the consequence may be gestational diabetes. Of even greater concern, women with pregestational diabetes have an increased risk of *stillbirth* and *congenital malformations* in the fetus. Poorly controlled diabetes that arises later in pregnancy, regardless of prior history, can lead to excessive birth weight in the newborn *(macrosomia)* and may have long-term sequelae for the child later in life, including an increased risk of obesity and diabetes. Gestational diabetes typically resolves following delivery; however, the majority of affected women develop overt diabetes over the next 10 to 20 years.

Clinical Features of Diabetes

It is difficult to sketch with brevity the diverse clinical presentations of diabetes. We will discuss the most common initial presentation or mode of diagnosis for each of the two major subtypes, followed by a discussion of acute and then chronic (long-term) complications of the disease.

T1D may arise at any age. In the initial 1 or 2 years following the onset of overt T1D, exogenous insulin requirements may be minimal because of residual endogenous insulin secretion (referred to as the *honeymoon period*). Eventually, however, β-cell function declines to a tipping point, and insulin requirements increase dramatically. Although β-cell destruction is a prolonged process, the transition from impaired glucose tolerance (stage 2, see earlier) to overt diabetes (stage 3) may be abrupt and is often brought on by a superimposed stress, such as infection, because of associated increase in insulin requirements.

In contrast to T1D, T2D is typically seen in obese patients older than 40 years of age; however, it is now being diagnosed in children and adolescents with increasing frequency due to increases in obesity and sedentary lifestyle. In some cases, medical attention is sought because of unexplained fatigue, dizziness, or blurred vision. Most frequently, however, the diagnosis of T2D is made after routine blood testing in asymptomatic persons. In fact, in light of the large number of asymptomatic individuals with undiagnosed hyperglycemia in the United States, routine blood glucose testing is recommended for everyone older than 45 years of age, and in younger individuals with obesity, family history, or the presence of the metabolic syndrome.

The Classic Triad of Diabetes

The onset of T1D is usually marked by the triad of polyuria, polydipsia, polyphagia, and, when severe, diabetic ketoacidosis, all resulting from metabolic derangements. Because insulin is a major anabolic hormone, its deficiency results in a catabolic state that affects glucose, fat, and protein metabolism. Unopposed secretion of counterregulatory hormones (such as glucagon) also plays a role in these metabolic derangements. The assimilation of glucose into muscle and adipose tissue is sharply diminished or abolished. Not only does storage of glycogen in liver and muscle cease, but reserves are depleted by glycogenolysis. The resultant *hyperglycemia* leads to filtration of so much glucose in the kidney that the renal tubular threshold for reabsorption is exceeded. This leads to glycosuria, which induces an osmotic diuresis and thus *polyuria,* causing a profound loss of water and electrolytes (Fig. 24.33). The renal water loss combined with the hyperosmolarity owing to increased levels of glucose in the blood depletes intracellular water, triggering the osmoreceptors of the thirst centers of the brain. Thus, intense thirst *(polydipsia)* appears. With a deficiency of insulin, the scales swing from insulin-promoted anabolism to catabolism of proteins and fats. Proteolysis follows, releasing gluconeogenic amino acids that are removed by the liver and used as building blocks for glucose. The catabolism of proteins and fats tends to induce a negative energy balance, which in turn leads to increasing appetite *(polyphagia),* thus completing the classic triad of polyuria, polydipsia, and polyphagia. Despite the increased appetite, catabolic effects prevail, resulting in weight loss and muscle weakness. The combination of polyphagia and weight loss is paradoxical and should always raise the suspicion of diabetes.

Acute Metabolic Complications of Diabetes

Diabetic ketoacidosis is a severe acute metabolic complication of T1D; it is not as common or as severe in T2D. The most frequent precipitating factor is a failure to take insulin, although other stressors such as infections, other illnesses, trauma, and certain drugs may also serve as triggers. It may also occur less commonly in T2D but only under condition of very severe stress such as caused by serious infections and trauma. Many of these factors are associated with the release of the catecholamine epinephrine, which blocks residual insulin action and stimulates the secretion of glucagon. The insulin deficiency coupled with glucagon excess decreases peripheral utilization of glucose while increasing gluconeogenesis, severely exacerbating hyperglycemia (the plasma glucose levels are usually in the range of 250 to 600 mg/dL). The hyperglycemia causes an osmotic diuresis and dehydration characteristic of the ketoacidotic state.

A second major effect of insulin deficiency is increased synthesis of ketone bodies. Insulin deficiency stimulates hormone-sensitive lipase, with a resultant breakdown of adipose stores and an increase in levels of FFAs. When these FFAs reach the liver, they are esterified to fatty acyl coenzyme A. Oxidation of fatty acyl coenzyme A molecules within the hepatic mitochondria produces *ketone bodies* (acetoacetic acid and β-hydroxybutyric acid). The rate at which ketone bodies are formed may exceed the rate at which they can be utilized by peripheral tissues, leading to *ketonemia* and *ketonuria*. If the urinary excretion of ketones is compromised by dehydration, the result is a systemic *metabolic ketoacidosis*. Release of ketogenic amino acids by protein catabolism aggravates the ketotic state.

The clinical manifestations of diabetic ketoacidosis include fatigue, nausea and vomiting, severe abdominal pain, a characteristic fruity odor, and deep, labored breathing (also known as *Kussmaul breathing*). Persistence of the ketotic state eventually leads to depressed consciousness and coma. Reversal of ketoacidosis requires administration of insulin, correction of metabolic acidosis, and treatment of any underlying precipitating factors, such as infection.

The lower frequency of ketoacidosis in T2D is believed to be due to higher portal vein insulin levels in these patients, which prevents unrestricted hepatic fatty acid oxidation and keeps the formation of ketone bodies in check. Instead, patients with T2D may develop a condition known as *hyperosmolar hyperglycemic state* due to severe dehydration resulting from sustained osmotic diuresis (particularly in patients who do not drink enough water to compensate for urinary losses from chronic hyperglycemia). Typically, this occurs in an older patient who has diabetes and is disabled by a stroke or an infection and thus unable to maintain adequate water intake. Furthermore, the absence of ketoacidosis and its symptoms (nausea, vomiting, Kussmaul breathing) delays the seeking of medical attention until severe dehydration and impairment of mental status occur. The hyperglycemia is usually more severe than in diabetic ketoacidosis, in the range of 600 to 1200 mg/dL.

Once treatment commences, ironically, the most common acute metabolic complication in either type of diabetes is

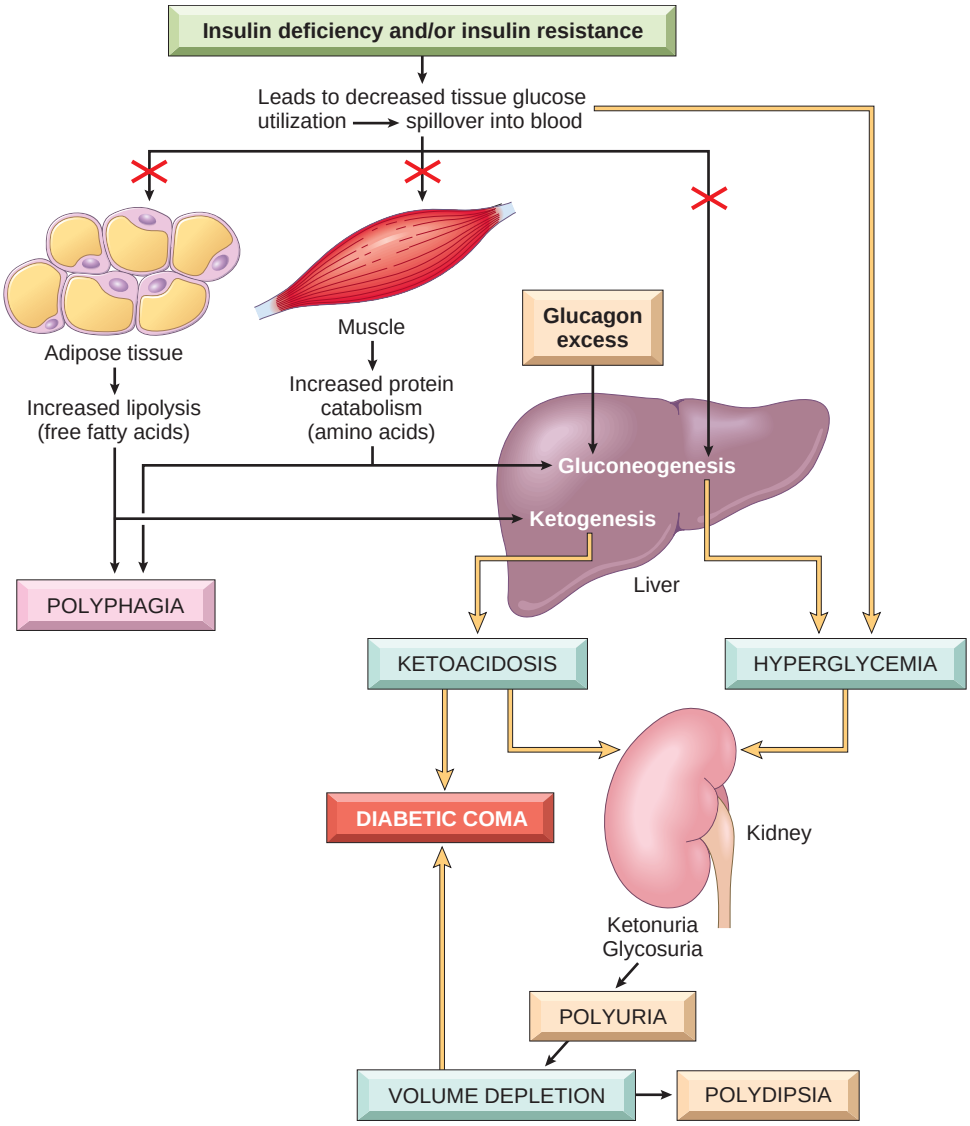


Figure 24.33 Sequence of metabolic derangements underlying the clinical manifestations of diabetes. An absolute insulin deficiency leads to a catabolic state, culminating in ketoacidosis and severe volume depletion. These cause sufficient central nervous system compromise to lead to coma and eventual death if left untreated.

hypoglycemia. Causes include missing a meal, excessive physical exertion, excessive insulin administration, or "misdosing" during the phase of dose finding for antidiabetic agents such as sulfonylureas. The signs and symptoms of hypoglycemia include dizziness, confusion, sweating, palpitations, and tachycardia; if hypoglycemia persists, loss of consciousness may occur. Rapid reversal of hypoglycemia through oral or intravenous glucose intake is critical for preventing permanent neurologic damage.

Chronic Complications of Diabetes

The morbidity associated with long-standing diabetes of either type is due to damage induced in large- and medium-sized muscular arteries (diabetic macrovascular disease) and in small vessels (diabetic microvascular disease) by chronic hyperglycemia. Macrovascular disease causes accelerated atherosclerosis among patients with diabetes, resulting in increased risk of myocardial infarction, stroke, and lower extremity ischemia. The effects of microvascular disease are most profound in the retina, kidneys, and peripheral nerves, resulting in *diabetic retinopathy, nephropathy, and neuropathy,* respectively (see later).

Pathogenesis

Persistent hyperglycemia (glucotoxicity) seems to be responsible for the long-term complications of diabetes. Much of the evidence supporting a role for glycemic control in ameliorating the long-term complications of diabetes has come from large randomized trials. The assessment of glycemic control in these trials has been based on the percentage of *glycated hemoglobin,* also known as HbA1c, which is formed by nonenzymatic covalent addition of glucose moieties to hemoglobin in red cells. Unlike blood glucose levels, HbA1c provides a measure of glycemic control over the lifespan of a red cell (120 days) and is affected little by day-to-day variation in glucose levels. It is recommended that HbA1c be maintained below 7% in patients with diabetes. The emergence of new technology, including

continuous glucose-monitoring systems, has introduced a new goal, increasing "time-in-range" (set at 70 to 180 mg/dL), which may be a better predictor of the risk of chronic complications than HbA1c level. It is important to stress, however, that hyperglycemia is not the only factor responsible for the long-term complications of diabetes, and that other underlying abnormalities, such as insulin resistance, and comorbidities like obesity, also play an important role.

At least four distinct mechanisms have been implicated in the deleterious effects of persistent hyperglycemia on peripheral tissues. **In each proposed mechanism, increased flux through metabolic pathways due to hyperglycemia is thought to generate harmful precursors that contribute to end-organ damage.**

- *Formation of advanced glycation end products.* Advanced glycation end products (AGEs) are formed as a result of nonenzymatic reactions between glucose-derived metabolites (glyoxal, methylglyoxal, and 3-deoxyglucosone) and the amino groups of intracellular and extracellular proteins. The rate of AGE formation is accelerated by hyperglycemia. AGEs bind to a specific receptor (RAGE) that is expressed on inflammatory cells (macrophages and T cells), endothelium, and vascular smooth muscle. The detrimental effects of the AGE-RAGE signaling axis within the vascular compartment include the following:
  - Release of *cytokines and growth factors,* including transforming growth factor-β (TGF-β), which leads to deposition of excess basement membrane material, and vascular endothelial growth factor (VEGF), implicated in diabetic retinopathy (see later)
  - Generation of *reactive oxygen species* (ROS) in endothelial cells
  - Increased *procoagulant activity* on endothelial cells and macrophages
  - Enhanced *proliferation of vascular smooth muscle cells* and *synthesis of extracellular matrix*

  Not surprisingly, endothelium-specific overexpression of RAGE in diabetic mice accelerates large vessel injury and microangiopathy, while RAGE-null mice show attenuation of these features. Antagonists of RAGE have emerged as therapeutic agents in diabetes and are being tested in clinical trials.

  In addition to receptor-mediated effects, AGEs can directly cross-link extracellular matrix proteins. Cross-linking of collagen type I molecules in large vessels decreases their elasticity, which may predispose these vessels to shear stress and endothelial injury (Chapter 11). Similarly, AGE-induced cross-linking of type IV collagen in basement membrane decreases endothelial cell adhesion and increases extravasation of fluid. Proteins cross-linked by AGEs are resistant to proteolytic digestion. Thus, cross-linking decreases protein removal, enhancing protein accumulation. AGE-modified matrix components also trap nonglycated plasma or interstitial proteins. In large vessels, trapping of LDL, for example, retards its efflux from the vessel wall and contributes to the deposition of cholesterol in the intima, thus accelerating atherogenesis (Chapter 11). In capillaries, including those of renal glomeruli, plasma proteins such as albumin bind to the glycated basement membrane, accounting in part for the basement membrane thickening that is characteristic of diabetic microangiopathy.
- *Activation of protein kinase C.* Calcium-dependent activation of intracellular protein kinase C (PKC) and the second messenger diacyl glycerol (DAG) is an important signal transduction pathway. Intracellular hyperglycemia stimulates the *de novo* synthesis of DAG from glycolytic intermediates, and hence causes excessive PKC activation. The downstream effects of PKC activation are numerous, including production of VEGF, TGF-β, and the procoagulant protein plasminogen activator inhibitor-1 (PAI-1) (Chapter 4) by the vascular endothelium. It should be evident that some effects of AGEs and activated PKC are overlapping, and both likely contribute to diabetic microangiopathy.
- *Oxidative stress and disturbances in polyol pathways.* Even in tissues that do not require insulin for glucose transport (e.g., nerves, lenses, kidneys, blood vessels), persistent hyperglycemia leads to an increase in intracellular glucose. This excess glucose is metabolized by the enzyme aldose reductase to sorbitol, a polyol, and eventually to fructose, in a reaction that uses NADPH (the reduced form of nicotinamide dinucleotide phosphate) as a cofactor. NADPH is also required by the enzyme glutathione reductase in a reaction that regenerates reduced glutathione (GSH). GSH is one of the important antioxidant mechanisms in the cell (Chapter 2), and any reduction in GSH increases cellular susceptibility to reactive oxygen species ("oxidative stress"). In the face of sustained hyperglycemia, progressive depletion of intracellular NADPH by aldose reductase compromises GSH regeneration, increasing cellular susceptibility to oxidative stress. Sorbitol accumulation in the lens contributes to cataract formation.
- *Hexosamine pathways and generation of fructose-6-phosphate.* Finally, it is postulated that hyperglycemia induces flux of glycolytic intermediates through the hexosamine pathway, which results in cell damage and enhanced oxidative stress.

Morphology of Chronic Complications of Diabetes

The important morphologic changes are related to the many late systemic complications of diabetes. As discussed earlier, these changes are seen in both T1D and T2D (Fig. 24.34).

MORPHOLOGY

PANCREAS

Alterations in the pancreas are inconstant and often subtle. Distinctive changes are more commonly associated with T1D than with T2D. One or more of the following alterations may be present:

- **Reduction in the number and size of islets.** This is most often seen in T1D, particularly with rapidly advancing disease. Most of the islets are small and inconspicuous.
- **Leukocytic infiltrates in the islets** (insulitis) are principally composed of T lymphocytes and are also seen in animal models of autoimmune diabetes (Fig. 24.35A). Lymphocytic infiltrates may be present in T1D at the time of clinical presentation.
- **In T2D there may be a subtle reduction in islet cell mass,** demonstrated only by special morphometric studies.
- **Amyloid deposition within islets in T2D** begins in and around capillaries and between cells. At advanced stages, the

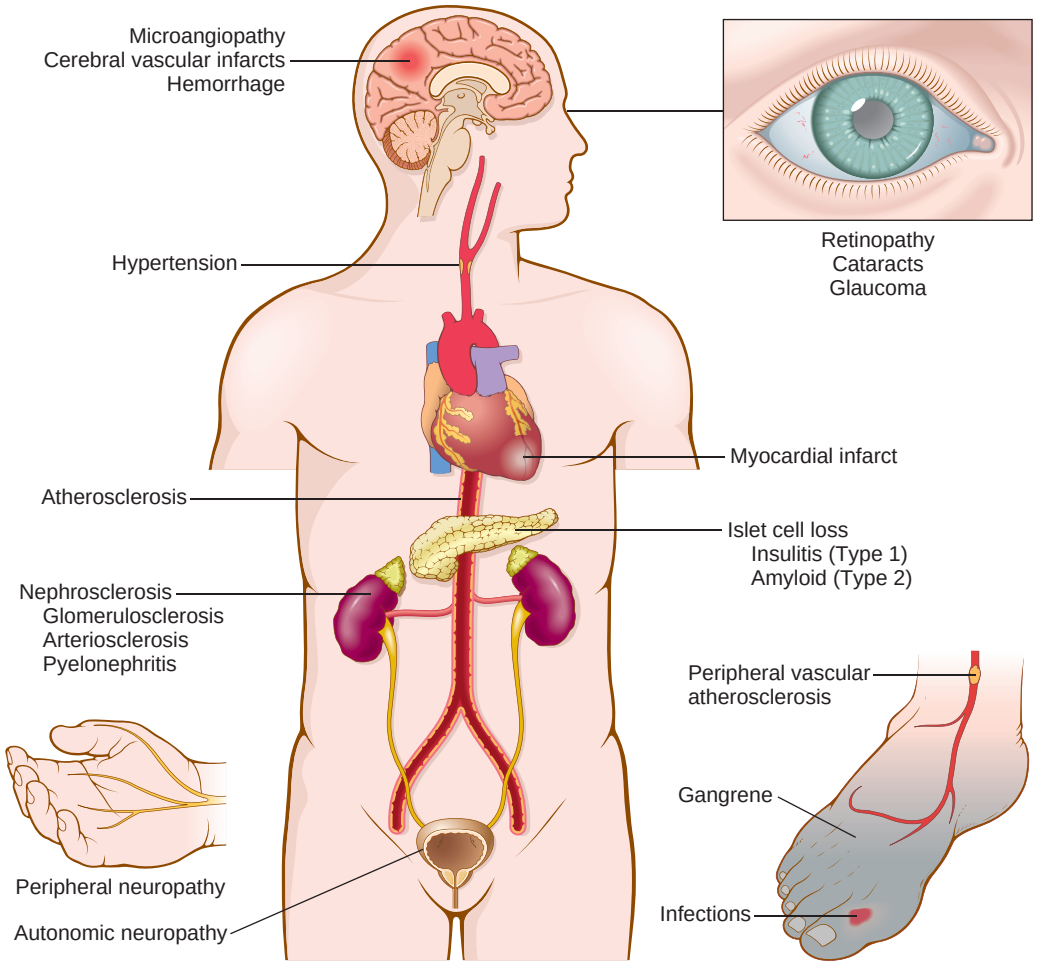


Figure 24.34 Long-term complications of diabetes.

islets may be virtually obliterated (see Fig. 24.35B); fibrosis may also be observed. Similar lesions may be found in older individuals without diabetes, apparently as part of normal aging.

- **An increase in the number and size of islets** is especially characteristic of nondiabetic newborns of mothers with diabetes. Presumably, fetal islets undergo hyperplasia in response to the maternal hyperglycemia.

Diabetic Macrovascular Disease

Diabetes exacts a heavy toll on the vascular system. *Endothelial dysfunction* (Chapter 11), which predisposes to atherosclerosis and other cardiovascular morbidities, is widespread in diabetes, as a consequence of the deleterious effects of persistent hyperglycemia and insulin resistance on the vascular compartment. The hallmark of diabetic macrovascular disease is *accelerated atherosclerosis* involving the aorta and large- and medium-sized arteries. The morphology of atherosclerosis in patients with diabetes is indistinguishable from that in individuals without diabetes (Chapter 11). **Myocardial infarction, caused by atherosclerosis of the coronary arteries, is the most common cause of death in diabetes.** *Gangrene of the lower extremities,* as a result of advanced vascular disease, is about 100 times more common in diabetes patients than in the general population. The larger renal arteries are also subject to severe atherosclerosis, but the most damaging effect of diabetes on the kidneys is exerted at the level of the glomeruli and the microcirculation (discussed later).

Hyaline arteriolosclerosis, the vascular lesion associated with essential hypertension (Chapters 11 and 20), is both more prevalent and more severe in patients with diabetes than those without, but it is not specific for diabetes and may be seen in older patients without hypertension. It takes the form of an amorphous, hyaline thickening of the wall of the arterioles, which causes narrowing of the lumen (Fig. 24.36). Not surprisingly, in diabetic patients, it is related not only to the duration of the disease but also to the level of blood pressure.

Diabetic Microangiopathy

One of the most consistent morphologic features of diabetes is *diffuse thickening of basement membranes.* The thickening is most evident in the capillaries of the skin, skeletal muscle, retina, renal glomeruli, and renal medulla. However, it may also be seen in such nonvascular structures as renal tubules, the Bowman capsule, peripheral nerves, and placenta. It should be noted that despite the increase in the thickness of basement membranes, capillaries in patients with diabetes are leakier than normal to plasma proteins. **The microangiopathy underlies the development of**

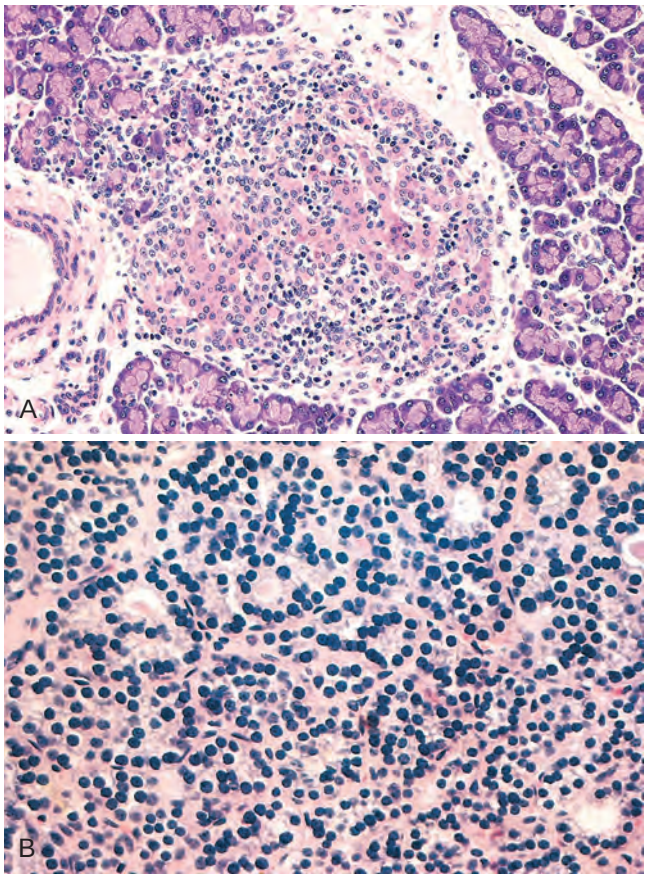


Figure 24.35 (A) Insulitis, shown here from a rat (BB) model of autoimmune diabetes, also seen in type 1 human diabetes. (B) Amyloidosis of a pancreatic islet in type 2 diabetes. (A, Courtesy Dr. Arthur Like, University of Massachusetts, Worchester, Mass.)

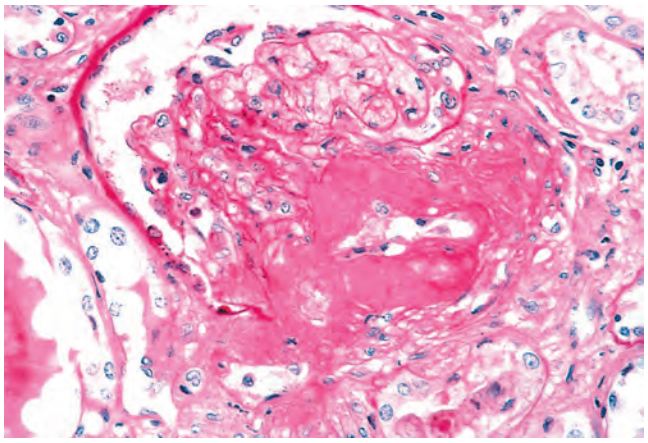


Figure 24.36 Severe renal hyaline arteriolosclerosis. Note a markedly thickened, tortuous afferent arteriole. The amorphous nature of the thickened vascular wall is evident (periodic acid–Schiff stain). (Courtesy M.A. Venkatachalam, MD, Department of Pathology, University of Texas Health Science Center, San Antonio, Texas.)

diabetic nephropathy, retinopathy, and some forms of neuropathy.

Diabetic Nephropathy

The kidneys are prime targets of diabetes. Renal failure is second only to myocardial infarction as a cause of death from this disease. **Three lesions are encountered: (1) glomerular lesions; (2) renal vascular lesions, principally arteriolosclerosis; and (3) pyelonephritis, including necrotizing papillitis.**

The most important glomerular lesions are capillary basement membrane thickening, diffuse mesangial sclerosis, and nodular glomerulosclerosis.

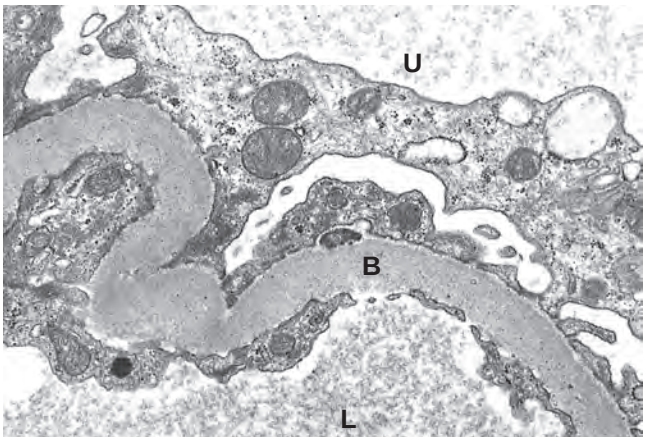


Figure 24.37 Electron micrograph of a renal glomerulus showing markedly thickened glomerular basement membrane *(B)* in a diabetic. *L,* Glomerular capillary lumen; *U,* urinary space. (Courtesy Dr. Michael Kashgarian, Department of Pathology, Yale University School of Medicine, New Haven, Conn.)

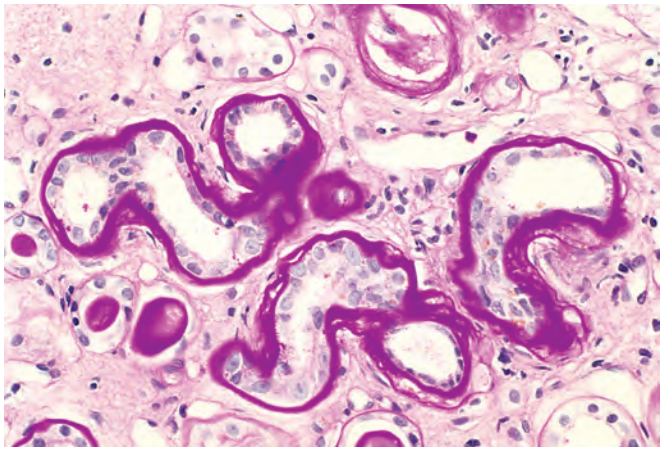


Figure 24.38 Renal cortex showing thickening of tubular basement membranes in a diabetic patient (periodic acid–Schiff stain).

Capillary Basement Membrane Thickening. Widespread thickening of the glomerular capillary basement membrane (GBM) occurs in virtually all cases of diabetic nephropathy and is part and parcel of diabetic microangiopathy. Capillary basement membrane thickening is best appreciated by electron microscopy (Fig. 24.37). Morphometric studies demonstrate that thickening begins as early as 2 years after the onset of T1D and by 5 years amounts to about a 30% increase. These progressive changes in the GBM are usually accompanied by mesangial widening and thickening of the tubular basement membranes (Fig. 24.38).

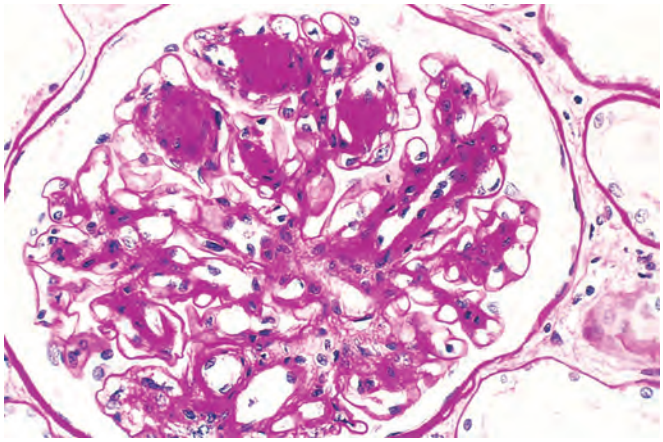


Figure 24.39 Diffuse and nodular diabetic glomerulosclerosis (periodic acid–Schiff [PAS] stain). Note the diffuse increase in mesangial matrix and characteristic acellular PAS-positive nodules.

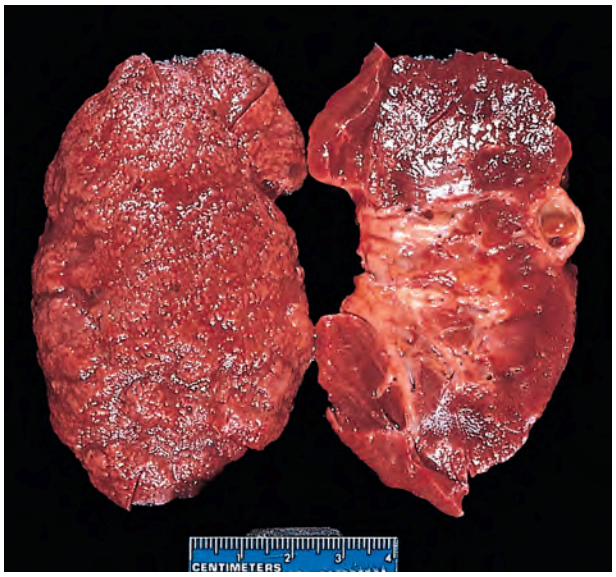


Figure 24.40 Nephrosclerosis in a patient with long-standing diabetes. The kidney has been bisected to demonstrate both diffuse granular transformation of the surface *(left)* and marked thinning of the cortical tissue *(right)*. Additional features include some irregular depressions, the result of pyelonephritis, and an incidental cortical cyst *(far right)*.

Diffuse Mesangial Sclerosis. This lesion consists of diffuse increase in mesangial matrix. The matrix depositions are PAS-positive (Fig. 24.39). As the disease progresses, the mesangial matrix deposits may take on a nodular appearance. The progressive expansion of the mesangium has been shown to correlate well with measures of deteriorating renal function such as increasing proteinuria.

Nodular Glomerulosclerosis. This is also known as intercapillary glomerulosclerosis or *Kimmelstiel-Wilson disease.* The glomerular lesions take the form of ovoid or spherical, often laminated, PAS-positive nodules of matrix situated in the periphery of the glomerulus. They lie within the mesangial core of the glomerular lobules and may be surrounded by patent peripheral capillary loops (see Fig. 24.39) or loops that are markedly dilated. The nodules often show features of mesangiolysis, defined by disruption of anchoring interactions between the capillaries and the mesangial stalks. The loss of these support structures may lead to the formation of capillary microaneurysms. While not all the lobules in individual glomeruli are involved by nodular lesions, even uninvolved lobules and glomeruli show striking diffuse mesangial sclerosis. The nodular lesions are frequently accompanied by prominent accumulations of hyaline material in capillary loops ("fibrin caps") or adherent to Bowman capsules ("capsular drops"). Both afferent and efferent glomerular hilar arterioles show hyalinosis. As a consequence of the glomerular and arteriolar lesions, the kidney becomes ischemic, develops tubular atrophy and interstitial fibrosis, and usually contracts in size (Fig. 24.40). Approximately 15% to 30% of individuals with long-term diabetes develop nodular glomerulosclerosis, and in most instances it is associated with renal failure.

Renal atherosclerosis and arteriolosclerosis also contribute to renal dysfunction in patients with diabetes. Hyaline arteriolosclerosis affects not only the afferent but also the efferent arteriole. Such efferent arteriolosclerosis is rarely, if ever, encountered in individuals who do not have diabetes.

Pyelonephritis is an acute or chronic inflammation of the kidneys that usually begins in the interstitial tissue and then spreads to affect the tubules. Both the acute and chronic forms of this disease are more common and more severe in patients with diabetes than in the general population. One special pattern of acute pyelonephritis, papillary necrosis, is much more prevalent in diabetes patients than in those without diabetes.

Diabetic Ocular Complications

The eye is profoundly affected by diabetes mellitus. The architecture and microanatomy of the eye are discussed in Chapter 29.

Diabetes-induced hyperglycemia leads to acquired opacification of the lens, a condition known as *cataract.* Long-standing diabetes is also associated with increased intraocular pressure (**glaucoma;** see later), and resulting damage to the optic nerve.

The most profound ocular changes of diabetes are seen in the retina. The retinal vasculopathy of diabetes mellitus can be classified into background (preproliferative) diabetic retinopathy and proliferative diabetic retinopathy (Chapter 29).

Diabetic Neuropathy

The prevalence of peripheral neuropathy in individuals with diabetes depends on the duration of the disease; up to 50% of diabetes patients overall have peripheral neuropathy clinically, and up to 80% of those who have had the disease for more than 15 years. This is discussed further in Chapter 27.

Clinical Manifestations of Chronic Diabetes

Table 24.7 summarizes some of the pertinent clinical, genetic, and histopathologic features that distinguish T1D and T2D. **In both types, it is the long-term effects of diabetes, rather than acute metabolic complications, that are responsible for much of the morbidity and mortality.** In most instances, these complications appear approximately 15 to 20 years after the onset of hyperglycemia. The severity of chronic

complications is related to both the degree and the duration of hyperglycemia, as evidenced by the attenuation of end-organ damage by effective glycemic control in prospective studies.

- **Macrovascular complications such as myocardial infarction, renal vascular insufficiency, and cerebrovascular accidents are the most common causes of mortality in long-standing diabetes.** Patients with diabetes have a two to four times greater incidence of coronary artery disease and a fourfold higher risk of dying from cardiovascular complications than age-matched individuals without diabetes. An elevated risk for cardiovascular disease is even observed in patients with prediabetes. Significantly, myocardial infarction is almost as common in women with diabetes as in men. In contrast, myocardial infarction is uncommon in women of reproductive age without diabetes. Diabetes is often accompanied by underlying conditions that favor the development of adverse cardiovascular events. For example, *hypertension* is found in approximately 75% of individuals with T2D and potentiates the effects of hyperglycemia and insulin resistance on endothelial dysfunction and atherosclerosis. Another cardiovascular risk frequently seen in diabetes patients is *dyslipidemia,* which includes both increased triglycerides and LDL levels and decreased levels of the "protective" lipoprotein, high-density lipoprotein (Chapter 11). Insulin resistance is believed to contribute to "diabetic dyslipidemia" by favoring the hepatic production of atherogenic lipoproteins and by suppressing the uptake of circulating lipids in peripheral tissues.
- **Diabetic nephropathy is a leading cause of end-stage renal disease in the United States.** Approximately 30% to 40% of all patients with diabetes develop clinical evidence of nephropathy. Development of end-stage renal disease is more likely with T1D than T2D, but because of the greater prevalence of T2D, these patients constitute slightly over one-half of the patients with diabetes starting dialysis each year. The progression from overt nephropathy to end-stage renal disease is highly variable, but by 20 years, more than 75% of patients with T1D and approximately 20% of those with T2D with overt nephropathy will develop end-stage renal disease, requiring dialysis or renal transplantation.

  The likelihood of diabetic nephropathy is greatly influenced by ethnicity; for example, Native Americans, Hispanics, and African Americans have a greater risk of developing end-stage renal disease than do non-Hispanic Caucasians with T2D. These differences are suspected to be genetic in origin, but responsible genes have yet to be identified. The earliest manifestation of diabetic nephropathy is the appearance of low amounts of albumin (*microalbuminuria*) in the urine (>30 mg and <300 mg/day). Without specific interventions, approximately 80% of patients with T1D and 20% to 40% of patients with T2D will develop *overt nephropathy with macroalbuminuria* (>300 mg/day of urinary albumin) over 10 to 15 years, usually accompanied by the appearance of hypertension.
- **Visual impairment, including total blindness, is one of the more feared consequences of long-standing diabetes.** Approximately 60% to 80% of patients develop some form of *diabetic retinopathy* approximately 15 to 20 years after diagnosis, and diabetic retinopathy is the leading cause of adult blindness in the United States. The fundamental lesion of retinopathy—neovascularization—is attributable to hypoxia-induced expression of VEGF in the retina. Current treatment for this condition includes administration of antiangiogenic agents that block the action of VEGF. As stated earlier, patients with diabetes also have an increased propensity for *glaucoma* and *cataract formation,* which also contribute to visual impairment.
- **Diabetic neuropathy can result in damage to the central nervous system, peripheral sensorimotor nerves, and the autonomic nervous system.** It most frequently takes the form of a *distal symmetric polyneuropathy* of the lower extremities that affects both motor and sensory function. Over time, the upper extremities may be involved as well, thus approximating a "glove-and-stocking" pattern of polyneuropathy. Other forms include *autonomic neuropathy,* which produces disturbances in bowel and bladder function and sometimes erectile dysfunction, and *diabetic mononeuropathy,* which may manifest as sudden footdrop, wristdrop, or isolated cranial nerve palsies.
- **Patients with diabetes are plagued by enhanced susceptibility to infections of the skin and to tuberculosis, pneumonia, and pyelonephritis.** Such infections cause the deaths of about 5% of these patients. In an individual with diabetic neuropathy, a trivial infection in a toe may be the first event in a long succession of complications (gangrene, bacteremia, pneumonia) that ultimately leads to death. The basis of enhanced susceptibility is multifactorial and includes decreased neutrophil function (chemotaxis, adherence to the endothelium, phagocytosis, and microbicidal activity) and impaired cytokine production by macrophages. The vascular compromise also impairs the delivery of immune cells and molecules to sites of infection.

The staggering societal and economic impact of diabetes has already been discussed. For the most part, diabetes remains a lifelong disease, although pancreatic islet cell transplantation has the potential to ameliorate T1D for many patients. For some individuals with T2D, dietary modifications, exercise, and weight-loss regimens can reduce insulin resistance and hyperglycemia at least early in the disease. However, all patients will ultimately require some form of therapeutic intervention to maintain glycemic control.

KEY CONCEPTS

DIABETES MELLITUS: PATHOGENESIS AND LONG-TERM COMPLICATIONS

- T1D is an autoimmune disease characterized by progressive destruction of islet β cells, leading to absolute insulin deficiency. T1D stems from a failure of self-tolerance in T cells, and circulating autoantibodies to islet cell antigens (including insulin) often are detected in affected patients.
- T2D has no autoimmune basis; instead, features central to its pathogenesis are insulin resistance and β-cell dysfunction, resulting in relative insulin deficiency.
- Obesity has an important relationship with insulin resistance (and hence T2D), mediated through multiple factors including excess FFAs, cytokines released from adipose tissues (adipocytokines), and inflammation.

- Monogenic forms of diabetes are uncommon and are caused by single-gene defects that result in primary β-cell dysfunction (e.g., glucokinase mutation) or lead to abnormalities of insulin-insulin receptor signaling (e.g., insulin receptor gene mutations).
- The long-term complications of diabetes are similar in both types and involve four potential mechanisms resulting from sustained hyperglycemia: formation of AGEs, activation of PKC, disturbances in the polyol pathways leading to oxidative stress, and overload of the hexosamine pathway.
- Long-term complications of diabetes include both large-vessel disease (macroangiopathy), such as atherosclerosis, ischemic heart disease, and lower extremity ischemia, as well as small vessel disease (microangiopathy), the latter manifesting mainly as retinopathy, nephropathy, and neuropathy.

PANCREATIC NEUROENDOCRINE TUMORS

The preferred term for tumors of the pancreatic islet cells ("islet cell tumors") is *pancreatic neuroendocrine tumors* or *PanNETs*. They are rare in comparison with tumors of the exocrine pancreas, accounting for 2% of all pancreatic neoplasms. PanNETs can occur anywhere within the pancreas or in the immediate peripancreatic tissues. They resemble their counterparts, carcinoid tumors, found elsewhere in the alimentary tract (Chapter 17). These tumors may be single or multiple and benign or malignant. Pancreatic endocrine neoplasms often elaborate pancreatic hormones, but are sometimes nonfunctional.

Like other endocrine neoplasms, it is difficult to predict the behavior of a PanNET based on its light microscopic appearance alone, although tumors with a higher proliferation index (measured as 3% or more neoplastic nuclei expressing Ki-67) can have an aggressive biological potential. Unequivocal criteria for malignancy include metastases, vascular invasion, and local infiltration. The functional status of the tumor has some impact on prognosis, in that approximately 90% of insulin-producing tumors are benign, and 60% to 90% of other functioning and nonfunctioning pancreatic endocrine neoplasms are malignant. Fortunately, insulinomas are the most common subtype of pancreatic endocrine neoplasms.

Pathogenesis

The genome of sporadic PanNETs recently has been sequenced, with identification of recurrent somatic alterations in three major genes or pathways:

- *MEN1,* which causes familial MEN syndrome, type 1, also is mutated in some sporadic neuroendocrine tumors
- Loss-of-function mutations in tumor suppressor genes such *PTEN* and *TSC2* (Chapter 7), which result in activation of the oncogenic mammalian TOR (mTOR) signaling pathway.
- Inactivating mutations in two genes, alpha-thalassemia/mental retardation syndrome, X-linked *(ATRX)* and death-domain–associated protein *(DAXX),* which have multiple cellular functions, including telomere maintenance. PanNETs with *DAXX* or *ATRX* mutations demonstrate a phenomenon known as "alternative lengthening of telomeres" (ALT), which allows telomeres to be maintained in noeplastic cells that do not express telomerase (Chapter 7). Of note, nearly one-half of PanNETs have a somatic mutation in either *ATRX* or *DAXX,* but not both, consistent with their function in the same oncogenic pathway.

The three most common and distinctive clinical syndromes associated with functional pancreatic endocrine neoplasms are (1) *hyperinsulinism,* (2) *hypergastrinemia and the Zollinger-Ellison syndrome,* and (3) *MEN* (described later).

Hyperinsulinism (Insulinoma)

β-cell tumors (insulinomas) are the most common pancreatic endocrine neoplasms and often produce sufficient insulin to induce clinically significant hypoglycemia. The characteristic clinical picture is dominated by hypoglycemic episodes, which occur when the blood glucose level falls below 50 mg/dL. Clinical manifestations include confusion, stupor, and loss of consciousness. These episodes are precipitated by fasting or exercise and are promptly relieved by feeding or parenteral administration of glucose.

MORPHOLOGY

Insulinomas are most often found within the pancreas and are usually benign. Most are solitary, although multiple tumors may be encountered. *Bona fide* carcinomas, making up only about 10% of cases, are diagnosed on the basis of local invasion and distant metastases. On rare occasions an insulinoma may arise in ectopic pancreatic tissue. In such cases, electron microscopy reveals the distinctive granules of β cells (see Fig. 24.27).

Solitary tumors are usually small (often <2 cm in diameter), encapsulated, pale to red-brown nodules located anywhere in the pancreas. Histologically, these benign tumors look remarkably like giant islets, with preservation of the regular cords of monotonous cells and their orientation to the vasculature. Malignant tumors are also well-differentiated, and they may be deceptively encapsulated. **Deposition of amyloid** is a characteristic feature of many insulinomas (Fig. 24.41).

Hyperinsulinism may also be caused by **focal or diffuse hyperplasia of the islets.** This change is found occasionally in adults but is far more commonly encountered as congenital hyperinsulinism with hypoglycemia in neonates and infants. Several clinical scenarios may result in islet hyperplasia (previously known as nesidioblastosis), including maternal diabetes, Beckwith-Wiedemann syndrome (Chapter 10), and rare mutations in the β-cell K^+-channel protein or sulfonylurea receptor. In maternal diabetes, the fetal islets respond to hyperglycemia by increasing their size and number. In the postnatal period, these hyperactive islets may be responsible for serious episodes of hypoglycemia. This phenomenon is usually transient.

Clinical Features

While up to 80% of islet cell tumors demonstrate excessive insulin secretion, the hypoglycemia is mild in all but about 20%, and many cases never become clinically symptomatic. The critical laboratory findings in insulinomas are high circulating levels of insulin and a high insulin-to-glucose ratio. Surgical removal of the tumor is usually followed by prompt reversal of the hypoglycemia.

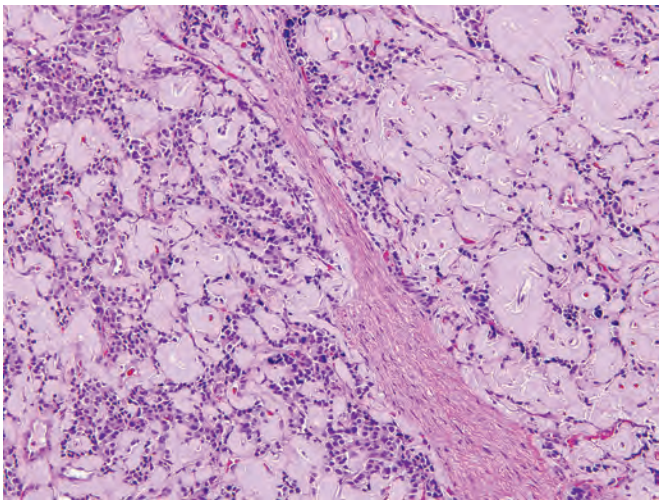


Figure 24.41 Pancreatic endocrine neoplasm ("islet cell tumor"). The neoplastic cells are monotonous and demonstrate minimal pleomorphism or mitotic activity. There is abundant amyloid deposition, characteristic of an insulinoma. Clinically, the patient had episodic hypoglycemia.

It is important to note that there are many other causes of hypoglycemia besides insulinoma. The differential diagnosis includes such conditions as abnormal insulin sensitivity, diffuse liver disease, inherited glycogenoses, and ectopic production of insulin by certain fibromas and fibrosarcomas. Depending on the clinical circumstances, hypoglycemia induced by self-injection of insulin should also be considered.

Zollinger-Ellison Syndrome (Gastrinoma)

Marked hypersecretion of gastrin usually has its origin in gastrin-producing tumors *(gastrinomas)*, which are just as likely to arise in the duodenum and peripancreatic soft tissues as in the pancreas (so-called gastrinoma triangle). There has been lack of agreement regarding the cell of origin of these tumors, although it seems likely that endocrine cells of either the gut or the pancreas could be the source. Zollinger and Ellison first called attention to the **association of pancreatic islet cell lesions, hypersecretion of gastric acid and severe peptic ulceration**, which are present in 90% to 95% of patients.

MORPHOLOGY

More than half of gastrin-producing tumors are locally invasive or have already metastasized at the time of diagnosis. In approximately 25% of patients, gastrinomas arise in conjunction with other endocrine tumors, as part of the MEN-1 syndrome (see later). MEN-1-associated gastrinomas are frequently multifocal, while sporadic gastrinomas are usually single. As with insulin-secreting tumors of the pancreas, gastrin-producing tumors are histologically bland and rarely show marked anaplasia.

In the Zollinger-Ellison syndrome, hypergastrinemia gives rise to excessive gastric acid secretion, which in turn causes **peptic ulceration** (Chapter 17). The duodenal and gastric ulcers are often multiple; although they are identical to those found in the general population, they are often unresponsive to therapy. In addition, ulcers may occur in unusual locations such as the jejunum; **when intractable jejunal ulcers are found, Zollinger-Ellison syndrome should be considered**.

Clinical Features

Treatment of Zollinger-Ellison syndrome involves control of gastric acid secretion by use of H^+K^+-ATPase inhibitors and excision of the neoplasm. Total resection of the neoplasm, when possible, eliminates the syndrome. Patients with hepatic metastases have a shortened life expectancy, with progressive tumor growth leading to liver failure, usually within 10 years.

Other Rare Pancreatic Endocrine Neoplasms

- α*-cell tumors (glucagonomas)* are associated with increased serum levels of glucagon and a syndrome consisting of mild diabetes mellitus, a characteristic skin rash *(necrolytic migratory erythema)*, and anemia. They occur most frequently in perimenopausal and postmenopausal women and are characterized by extremely high plasma glucagon levels.
- δ*-cell tumors (somatostatinomas)* are associated with diabetes mellitus, cholelithiasis, steatorrhea, and hypochlorhydria. They may be very difficult to localize preoperatively. High plasma somatostatin levels are required for diagnosis.
- *VIPoma* induces a characteristic syndrome *(watery diarrhea, hypokalemia, achlorhydria, or WDHA syndrome)* that is caused by release of vasoactive intestinal peptide (VIP) from the tumor. Some of these tumors are locally invasive and metastatic. A VIP assay should be performed on all patients with severe secretory diarrhea. Neural crest tumors, such as neuroblastomas, ganglioneuroblastomas, and ganglioneuromas (Chapter 10), and pheochromocytomas (see later) can also be associated with the VIPoma syndrome.
- *Pancreatic carcinoid tumors* producing serotonin and an atypical carcinoid syndrome are exceedingly rare.
- *Pancreatic polypeptide-secreting endocrine tumors* present as mass lesions as even high plasma levels of this hormone fail to cause symptoms.
- Some pancreatic and extrapancreatic endocrine tumors produce two or more hormones. In addition to insulin, glucagon, and gastrin, pancreatic endocrine tumors may produce ACTH, MSH, ADH, serotonin, and norepinephrine. These *multihormonal tumors* are to be distinguished from the MEN syndromes (discussed later), in which a multiplicity of hormones is produced by tumors in several different glands.

Adrenal Glands

The adrenal glands are paired endocrine organs consisting of a cortex and a medulla, which differ in their development, structure, and function. In essence, the cortex and medulla are two glands packaged into one structure. The adrenal cortex

has three zones. Beneath the capsule is the narrow layer of zona glomerulosa. An equally narrow zona reticularis abuts the medulla. Intervening is the broad zona fasciculata, which makes up about 75% of the total cortex. The *adrenal cortex* synthesizes three different types of steroids: (1) *glucocorticoids* (principally cortisol), which are synthesized primarily in the zona fasciculata and to a lesser degree in the zona reticularis; (2) *mineralocorticoids,* the most important being aldosterone, which is generated in the zona glomerulosa; and (3) *sex steroids* (estrogens and androgens), which are produced largely in the zona reticularis. The *adrenal medulla* is composed of chromaffin cells, which synthesize and secrete *catecholamines,* mainly epinephrine. Catecholamines have many effects that allow rapid adaptations to changes in the environment.

ADRENAL CORTEX

Diseases of the adrenal cortex can be conveniently divided into those associated with hyperfunction and those associated with hypofunction.

Adrenocortical Hyperfunction (Hyperadrenalism)

The syndromes of adrenal hyperfunction are caused by overproduction of the three major hormones of the adrenal cortex (1) *Cushing syndrome,* characterized by an excess of cortisol; (2) *hyperaldosteronism* as a result of excessive aldosterone; and (3) *adrenogenital or virilizing syndromes* caused by an excess of androgens. The clinical features of these syndromes overlap somewhat because of the overlapping functions of some of the adrenal steroids.

Hypercortisolism (Cushing Syndrome)

Pathogenesis

This disorder is caused by conditions that produce elevated glucocorticoid levels. Cushing syndrome can be broadly divided into *exogenous* and *endogenous* causes. The vast majority of cases of Cushing syndrome are the result of the administration of exogenous glucocorticoids (*iatrogenic* Cushing syndrome). The endogenous causes can, in turn, be divided into those that are *ACTH dependent* and those that are *ACTH independent* (Table 24.8).

ACTH-secreting pituitary adenomas account for 60% to 70% of cases of endogenous hypercortisolism. The pituitary form of this syndrome is referred to as *Cushing disease*. The disorder affects women about four times more frequently than men and occurs most frequently in young adults. In the vast majority of cases, it is caused by an *ACTH-producing pituitary microadenoma*. In some cases there is an underlying macroadenoma, and rarely there is *corticotroph cell hyperplasia* without a discrete adenoma. Corticotroph cell hyperplasia may be primary or arise secondarily from excessive stimulation of ACTH release by a hypothalamic corticotrophin-releasing hormone (CRH)-producing tumor. The adrenal glands in individuals with Cushing disease are characterized by variable degrees of *nodular cortical hyperplasia* (discussed later) caused by the elevated levels of ACTH. The cortical hyperplasia, in turn, is responsible for hypercortisolism.

Secretion of ectopic ACTH by nonpituitary tumors accounts for about 5% to 10% of ACTH-dependent Cushing syndrome cases. In many instances, the responsible tumor is a *small-cell carcinoma of the lung,* although other neoplasms, including carcinoids, medullary carcinomas of the thyroid, and PanNETs, have been associated with the syndrome. In addition to tumors that elaborate ectopic ACTH, occasionally a neuroendocrine neoplasm may produce ectopic CRH, which, in turn, causes ACTH secretion and hypercortisolism. As in the pituitary variant, the adrenal glands undergo bilateral cortical hyperplasia, but the rapid downhill course of patients with these cancers often limits the extent of the adrenal enlargement. This variant of Cushing syndrome is more common in men and usually occurs in the 40s and 50s.

Primary adrenal neoplasms, such as adrenal adenoma (~10% to 20%) and carcinoma (~5% to 7%) are the most common underlying causes of ACTH-independent Cushing syndrome. The biochemical *sine qua non* of ACTH-independent Cushing syndrome is elevated serum levels of cortisol with low levels of ACTH. Cortical carcinomas tend to produce more marked hypercortisolism than adenomas or hyperplasias. Recent large-scale gene-sequencing studies have helped define the molecular features of adrenocortical carcinomas, including recurrent activating mutations of *beta-catenin* (*CTNNB1*) and inactivating mutations of *TP53, MEN1,* and *PRKAR1A*. Interestingly, *PRKAR1A* mutations are observed in both adrenocortical adenomas and carcinomas, as well as in micronodular hyperplasia (see earlier), suggesting shared molecular mechanisms and ontogeny are at play in these cortisol-secreting lesions.

Primary cortical hyperplasia (i.e., ACTH-independent hyperplasia) is much less common than ACTH-dependent adrenal cortical hyperplasia. In *bilateral macronodular adrenal hyperplasia* (BMAH), the nodules are usually greater than 10 mm in diameter. Familial forms associated with germline mutations in the armadillo repeat containing 5 *(ARMC5)* gene, a putative tumor suppressor gene, have been reported; of note, up to 50% of apparently sporadic BMAH cases harbor germline *ARMC5* mutations with somatic loss of the second allele in the hyperplastic nodules. In true sporadic cases of BMAH, nodular hyperplasia is not entirely "autonomous." Specifically, cortisol production is regulated by hormones other than ACTH because of ectopic overexpression of the corresponding G-protein–coupled hormone receptors in the hyperplastic adrenocortical cells. Implicated hormones other than ACTH include gastric inhibitory peptide (GIP), LH, and ADH (vasopressin). The mechanism by which these receptors are overexpressed is not known, but epigenetic changes may be involved. A subset of BMAH arises in the setting of McCune-Albright syndrome (Chapter 26), a multisystemic disease characterized by germline mutations that activate *GNAS,* which encodes a stimulatory $G_s\alpha$. This $G_s\alpha$ mutation causes hyperplasia by increasing intracellular levels of cAMP, which you will recall is an important second messenger in many endocrine cell types.

ACTH-independent bilateral hyperplasia can also be *micronodular* (<10 mm in size), and these arise mainly in two settings: primary pigmented nodular adrenocortical disease or as part of the so-called Carney complex, a multisystem syndrome resulting in both endocrine and non-endocrine neoplasms. Both primary pigmented nodular adrenocortical disease and Carney complex are most

Table 24.8 Causes of Endogenous Cushing Syndrome

	Proportion (%)	Age (Peak)	Female:Male	Features
ACTH-Dependent	**70–80**	..	..	..
Cushing disease	60–70	..	..	..
Corticotroph adenoma	60–70	3rd–4th decades	3–5:1	Roughly 50% nonvisible on MRI
Corticotroph hyperplasia	Very rare	..	..	..
Ectopic ACTH[a]	5–10	..	..	..
Malignant neuroendocrine tumours	~4	5th–6th decades	0.6–1:1	May have very high ACTH
Benign neuroendocrine tumours	~6	3rd–4th decades	..	May respond to dexamethasone, CRH, desmopressin
Occult neuroendocrine tumours	~2	..	..	..
Ectopic CRH	Very rare	..	..	Causes pituitary corticotroph hyperplasia
ACTH-Independent	**20–30**	..	..	..
Unilateral adrenal	..	..	..	..
Adenoma	10–22	4th–5th decades	4–8:1	Most secrete only cortisol
Carcinoma	5–7	1st, 5th–6th decades	1.5–3:1	Mixed cortisol and androgen secretion frequent
Bilateral macronodular adrenal hyperplasia[b]	<2	5th–6th decades	2–3:1	Modest cortisol secretion compared with size; may also secrete androgen and mineralocorticoid
Aberrant G-protein-coupled receptors	..	..	..	..
Autocrine ACTH production	..	..	..	..
Sporadic or familial (*ARMC*$_5$)	..	..	..	..
Bilateral micronodular adrenal hyperplasias	<2	..	..	Adrenal size often normal
Primary pigmented nodular adrenocortical disease	Rare	1st–3rd decades	0.5:1<12 years 2:1>12 years	Frequent paradoxical increase of urine free cortisol with oral dexamethasone suppression test
Isolated or familial with Carney complex	Rare	1st–3rd decades	..	..
Isolated micronodular adrenocortical disease	Very rare	Infants	..	Nonpigmented adrenal micronodules
Primary bimorphic adrenocortical disease	Very rare	Infants	..	..
McCune-Albright syndrome	Rare	Infants (< 6 months)	1:1	Internodular adrenal atrophy
Bilateral adenomas or carcinomas	Rare	4th–5th decades	2–4:1	..

ACTH, Adrenocorticotropic hormone; *ARMC*$_5$, armadillo repeat containing 5; *CRH,* corticotrophin-releasing hormone.

[a]Most frequent sources of ectopic ACTH syndromes are small cell lung carcinoma and neuroendocrine tumors of lung, thymus, and pancreas. Less frequent causes include medullary thyroid carcinoma, gastrinoma, phaeochromocytoma, prostate carcinoma, and several others.
[b]In bilateral macronodular adrenal hyperplasia tissues, autocrine and paracrine ACTH might be produced and contribute to cortisol secretion. If confirmed by in vivo studies, the ACTH-independent classification will need to be modified in the future.

From Lacroix A, Feelders RA, Stratakis CA, Nieman LK: Cushing syndrome, *Lancet* 386(9996):913–927, 2015.

commonly associated with mutations of the regulatory subunit of cAMP-dependent protein kinase (encoded by the *PRKAR1A* gene), which, like activating *GNAS* mutations, act by increasing intracellular cAMP levels.

MORPHOLOGY

The main lesions of Cushing syndrome are found in the pituitary and adrenal glands. The **pituitary** shows changes regardless of the cause. The most common alteration, resulting from high levels of endogenous or exogenous glucocorticoids, is termed **Crooke hyaline change.** In this condition, the normal granular, basophilic cytoplasm of the ACTH-producing cells in the anterior pituitary becomes homogeneous and paler. This alteration is the result of the accumulation of intermediate keratin filaments in the cytoplasm, a finding that is also seen in "Crooke cell" corticotroph adenoma (discussed earlier).

Depending on the cause of the hypercortisolism, the **adrenals** show one of the following abnormalities: (1) cortical atrophy, (2) diffuse hyperplasia, (3) macronodular or micronodular hyperplasia, and (4) an adenoma or carcinoma. In patients in whom the syndrome results from exogenous glucocorticoids, suppression of endogenous ACTH results in bilateral **cortical atrophy,** due to a lack of stimulation of the zonae fasciculata and reticularis by ACTH. The zona glomerulosa is of normal thickness in such cases, because this portion of the cortex functions independently of ACTH. In contrast, in cases of endogenous hypercortisolism, the adrenals either are hyperplastic or contain a cortical neoplasm.

Diffuse hyperplasia is found in individuals with ACTH-dependent Cushing syndrome (Fig. 24.42). Both glands are enlarged, either subtly or markedly, weighing up to 30 g. The adrenal cortex is diffusely thickened and variably nodular, although the latter is not as pronounced as seen in cases of ACTH-independent nodular hyperplasia. Microscopically, the hyperplastic cortex demonstrates an expanded "lipid-poor" zona reticularis, comprising compact, eosinophilic cells, surrounded by an outer zone of vacuolated "lipid-rich" cells, resembling those seen in the zona fasciculata. Any nodules present are usually composed of vacuolated "lipid-rich" cells, which account for the yellow color of diffusely hyperplastic glands. In contrast, in **macronodular hyperplasia** the adrenals are almost entirely replaced by prominent nodules of varying sizes (10 to 30 mm), which contain an admixture of lipid-poor and lipid-rich cells. Unlike diffuse hyperplasia, the areas between the macroscopic nodules also demonstrate evidence of microscopic nodularity. **Micronodular hyperplasia** is composed of 1- to 3-mm (typically <10 mm) darkly pigmented (brown to black) micronodules with atrophic intervening areas (Fig. 24.43). The pigment is believed to be lipofuscin, a wear-and-tear pigment (Chapter 2).

Primary adrenocortical neoplasms causing Cushing syndrome may be malignant or benign. Functional adenomas or carcinomas are not morphologically distinct from nonfunctioning adrenal neoplasms (described later). Both the benign and malignant lesions are more common in women in their 30s to 50s. Adrenocortical **adenomas** are yellow tumors surrounded by thin or well-developed capsules, and most weigh less than 30 g. Microscopically, they are composed of cells that are similar to those encountered in the normal zona fasciculata. The **carcinomas** associated with Cushing syndrome, by contrast, tend to be larger than the adenomas. These tumors (detailed later) are unencapsulated masses frequently exceeding 200 to 300 g in weight that have all of the anaplastic characteristics of cancer. With functioning tumors, both benign and malignant, the adjacent adrenal cortex and that of the contralateral adrenal gland are atrophic, as a result of suppression of endogenous ACTH by high cortisol levels.

Figure 24.42 Diffuse hyperplasia of the adrenal gland contrasted with a normal adrenal gland (top). In cross-section, the hyperplastic adrenal cortex is yellow and thickened, and a subtle nodularity is seen (contrast with Fig. 24.43) in this gland from a patient with ACTH dependent Cushing syndrome.

Clinical Features

Cushing syndrome develops slowly and its onset may be subtle. Early stages may present with hypertension and weight gain (Table 24.9). With time, the more characteristic central pattern of adipose tissue deposition becomes apparent in the form of truncal obesity, moon facies, and accumulation of fat in the posterior neck and back *(buffalo hump)*. Hypercortisolism causes selective atrophy of fast-twitch (type 2) myofibers, resulting in decreased muscle mass and proximal limb weakness. Glucocorticoids induce gluconeogenesis and inhibit the uptake of glucose by cells, with resultant hyperglycemia, glucosuria, and polydipsia *(secondary diabetes)*. The catabolic effects cause loss of collagen and resorption of bones. Consequently, the skin is thin, fragile, and easily bruised; wound healing is poor; and cutaneous striae are particularly common in the abdominal area (Fig. 24.44). Bone resorption results in the development of *osteoporosis,* with consequent backache and increased susceptibility to fractures. Persons with Cushing syndrome are at increased risk for a variety of infections, because glucocorticoids suppress the immune response. Additional manifestations

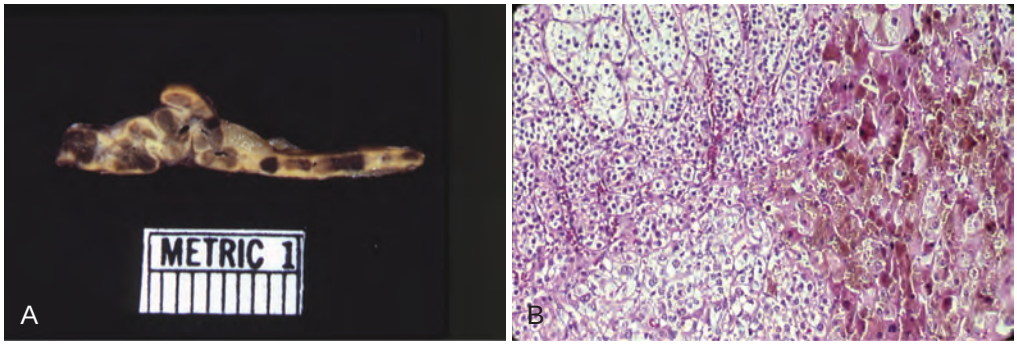


Figure 24.43 (A) Micronodular adrenocortical hyperplasia with prominent pigmented nodules in the adrenal gland. (B) On histologic examination, the nodules are composed of cells containing lipofuscin pigment, seen in the right part of the field. (Photographs courtesy Dr. Aidan Carney, Department of Medicine, Mayo Clinic, Rochester, Minn.)

Table 24.9 Clinical Features of Cushing Syndrome

Feature	Percent
Obesity or weight gain	95%[a]
Facial plethora	90%
Rounded face	90%
Decreased libido	90%
Thin skin	85%
Decrease in linear growth in children	70%–80%
Menstrual irregularity	80%
Hypertension	75%
Hirsutism	75%
Depression/emotional liability	70%
Easy bruising	65%
Glucose intolerance	60%
Weakness	60%
Osteopenia or fracture	50%
Nephrolithiasis	50%

[a]100% in children.

Modified from Newell-Price J, Bertagna X, Grossman AB, Nieman LK: Cushing syndrome, *Lancet* 367(9522):1605–1617, 2006.

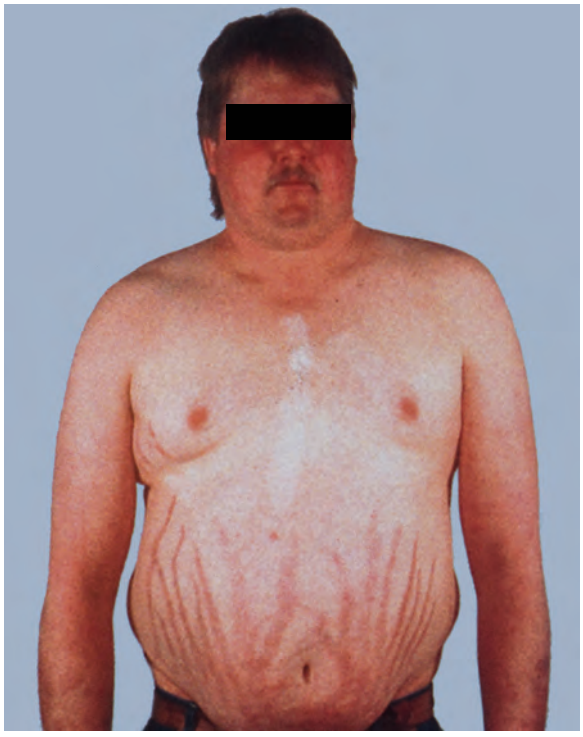


Figure 24.44 A patient with Cushing syndrome demonstrating central obesity, "moon facies," and abdominal striae. (Reproduced with permission from Lloyd RV, Douglas BR, Young WF Jr, editors: *Atlas of Nontumor Pathology: Endocrine Diseases*, Washington, DC, 2002, American Registry of Pathology.)

include several *mental disturbances,* including mood swings, depression, and frank psychosis, as well as *hirsutism* and *menstrual abnormalities*.

The laboratory diagnosis of Cushing syndrome is based on the following: (1) the 24-hour urine free-cortisol concentration, which is increased; and (2) loss of normal diurnal pattern of cortisol secretion. Determining the cause of Cushing syndrome depends on measurement of the serum ACTH and the dexamethasone suppression test, in which urinary steroid excretion is measured after administration of the glucocorticoid dexamethasone. The results of these tests fall into three patterns:

- In pituitary Cushing syndrome, the most common form, ACTH levels are elevated and are not suppressed by the administration of a low dose of dexamethasone. Hence, there is no reduction in urinary excretion of 17-hydroxycorticosteroids. After higher doses of injected dexamethasone, however, the pituitary responds by reducing ACTH secretion, which is reflected by suppression of urinary steroid secretion.
- Ectopic ACTH secretion results in an elevated level of ACTH, but its secretion is completely insensitive to low or high doses of exogenous dexamethasone.
- When Cushing syndrome is caused by an adrenal tumor, the ACTH level is quite low because of feedback inhibition of the pituitary. As with ectopic ACTH secretion, both low-dose and high-dose dexamethasone fail to suppress cortisol excretion.

KEY CONCEPTS

HYPERCORTISOLISM (CUSHING SYNDROME)

- The most common cause of hypercortisolism is exogenous administration of steroids.
- Endogenous hypercortisolism most often is secondary to an ACTH-producing pituitary microadenoma (Cushing disease), followed by primary adrenal neoplasms (ACTH-independent hypercortisolism) and paraneoplastic ACTH production by tumors (e.g., small cell lung cancer).
- The morphologic features in the adrenal vary from bilateral cortical atrophy (in exogenous steroid-induced disease), to bilateral diffuse or nodular hyperplasia (most common finding in endogenous Cushing syndrome), to an adrenocortical neoplasm.

Primary Hyperaldosteronism

Hyperaldosteronism is the generic term for a group of closely related conditions characterized by chronic excess aldosterone secretion. Hyperaldosteronism may be primary, or it may be secondary to an extra-adrenal cause. *Primary hyperaldosteronism* stems from an autonomous overproduction of aldosterone, with resultant hypertension, suppression of the renin-angiotensin system, and decreased plasma renin activity. Primary hyperaldosteronism is caused by one of three conditions (Fig. 24.45):

- *Bilateral idiopathic hyperaldosteronism,* characterized by bilateral nodular hyperplasia of aldosterone-secreting zona glomerulosa cells of the adrenal glands, is the most common underlying cause, accounting for about 60% of cases, most of which are sporadic. Individuals with bilateral idiopathic hyperaldosteronism tend to be older and to have less severe hypertension than those presenting with adrenal neoplasms. The molecular basis of this entity remains unclear, although recent studies suggest that a minority of patients harbor germline mutations of *KCNJ5,* consistent with their being part of a subtype that is now classified as *familial hyperaldosteronism, type III* (see later).

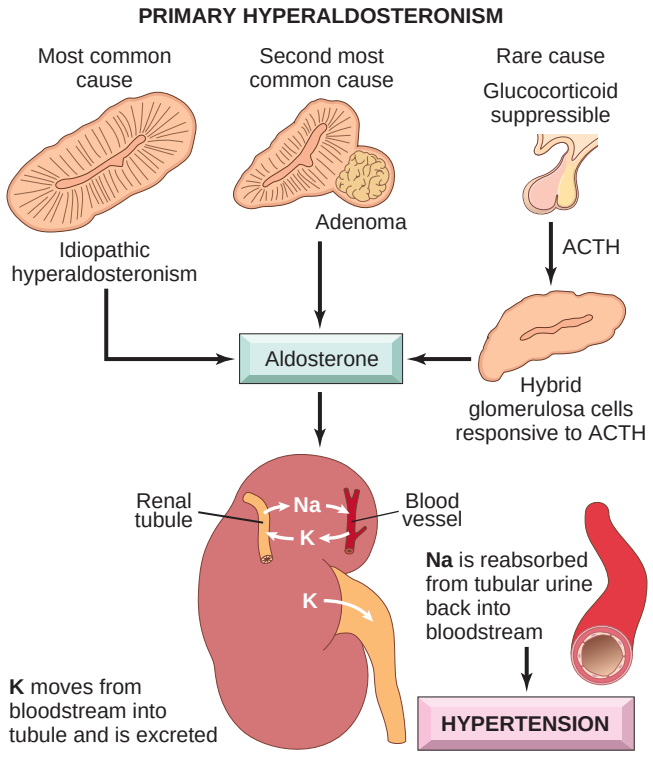


Figure 24.45 The major causes of primary hyperaldosteronism and its principal effects on the kidney.

- *Adrenocortical neoplasm,* either an aldosterone-producing adenoma (the most common cause) or, rarely, an adrenocortical carcinoma. In approximately 35% of cases, primary hyperaldosteronism is caused by a solitary aldosterone-secreting adenoma, a condition referred to as *Conn syndrome.* This condition occurs most frequently in mid-adult life and is more common in women than in men (2:1). Recent gene-sequencing studies have helped elucidate the genomic landscape of aldosterone-producing adenomas. Up to 50% harbor somatic mutations of *KCNJ5,* which encodes a potassium ion channel protein known as GIRK4 expressed in zona glomerulosa cells. GIRK4 mutations result in loss of selectivity of the channel to potassium ions, leading to nonspecific influx of sodium, and subsequent calcium-dependent activation of the aldosterone synthase enzyme. Disruption of intracellular ion homeostasis in zona glomerulosa cells is a common theme with other somatic mutations observed in aldosterone-producing adenomas as well. These include mutations in *CACNA1H,* which encodes a calcium channel, and *ATP1A1,* which encodes a sodium/potassium exchanging ATPase. The functional consequence of these mutations is to maintain cells in a chronic state of depolarization, resulting in autonomous aldosterone synthesis.
- *Familial hyperaldosteronism* accounts for ~5% of cases. Four distinct subtypes (FH-I to FH-IV) have been described, of which FH-I, or *glucocorticoid-remediable aldosteronism,* is the most common. FH-I stems from a rearrangement in chromosome 8 that places *CYP11B2* (the gene encoding aldosterone synthase, the enzyme that carries out the last step in aldosterone synthesis) under the control of the ACTH-responsive *CYP11B1* gene promoter. ACTH thus stimulates the synthesis of aldosterone synthase from the chimeric gene. Because in this unusual circumstance aldosterone production is under the control of ACTH, it is suppressible by dexamethasone. The remaining three subtypes of FH are rare, with FH-III being ascribed to germline *KCNJ5* mutations and FH-IV to germline *CACNA1H* mutations, two genes that are also somatically mutated in subsets of aldosterone-secreting adenomas.

In *secondary hyperaldosteronism,* in contrast, aldosterone release occurs in response to activation of the renin-angiotensin system (Chapter 11). It is characterized by increased levels of plasma renin and is encountered in conditions such as the following:

- Decreased renal perfusion (arteriolar nephrosclerosis, renal artery stenosis)
- Arterial hypovolemia and edema (congestive heart failure, cirrhosis, nephrotic syndrome)
- Pregnancy (due to estrogen-induced increases in plasma renin substrate)

MORPHOLOGY

Aldosterone-producing adenomas are almost always solitary, small (<2 cm in diameter), well-circumscribed lesions, more often found on the left than on the right. They tend to occur in the 30s and 40s, and in women more often than in men. They are often buried within the gland and do not produce visible enlargement, making them difficult to locate by imaging. They are bright yellow on cut section and, surprisingly, are composed of lipid-laden cortical cells that more closely resemble fasciculata cells than glomerulosa cells (the normal source of aldosterone). In general, the cells tend to be uniform in size and shape; occasionally, there is modest nuclear and cellular pleomorphism (see Fig. 24.51). A characteristic feature of aldosterone-producing adenomas is the presence of eosinophilic, laminated cytoplasmic inclusions, known as **spironolactone bodies,** found after treatment with the antihypertensive drug spironolactone. In contrast to cortical adenomas associated with Cushing syndrome, those associated with hyperaldosteronism do not usually suppress ACTH secretion. Therefore, the adjacent adrenal cortex and that of the contralateral gland are not atrophic.

Bilateral idiopathic hyperaldosteronism is marked by diffuse and focal hyperplasia of cells resembling those of the normal zona glomerulosa. The hyperplasia is often wedge-shaped, extending from the periphery toward the center of the gland. The enlargement may be subtle, and as a rule an adrenocortical adenoma must be carefully excluded as the cause for hyperaldosteronism.

Clinical Features

With an estimated prevalence rate of 5% to 10% among nonselected hypertensive patients, primary hyperaldosteronism may be the most common cause of secondary hypertension (i.e., hypertension secondary to an identifiable cause). The prevalence of hyperaldosteronism increases with the severity of hypertension, reaching nearly 20% in patients who are classified as having treatment-resistant hypertension. Through its effects on the renal mineralocorticoid receptor, aldosterone

promotes sodium reabsorption, which secondarily increases the reabsorption of water, expanding the extracellular fluid volume and elevating cardiac output.

The long-term effects of hyperaldosteronism-induced hypertension are cardiovascular compromise (e.g., left ventricular hypertrophy and reduced diastolic volumes) and an increase in the prevalence of adverse events such as stroke and myocardial infarction. *Hypokalemia* was considered a mandatory feature of primary hyperaldosteronism, but increasing numbers of normokalemic patients are now diagnosed. Hypokalemia results from renal potassium wasting and, when present, can cause a variety of neuromuscular manifestations, including weakness, paresthesias, visual disturbances, and occasionally frank tetany.

The diagnosis of primary hyperaldosteronism is confirmed by elevated ratios of plasma aldosterone concentration to plasma renin activity; if this screening test is positive, a confirmatory aldosterone suppression test is performed, as many other disorders can alter the plasma aldosterone and renin ratios.

In primary hyperaldosteronism, the therapy varies according to cause. Adenomas are amenable to surgical excision. In contrast, surgical intervention is not very beneficial in patients with bilateral hyperplasia, which often occurs in children and young adults. These patients are best managed medically with an aldosterone antagonist such as spironolactone. The treatment of secondary hyperaldosteronism rests on correcting the underlying cause of the hyperactivity of the renin-angiotensin system.

Adrenogenital Syndromes

Disorders of sexual differentiation, such as *virilization* or *feminization,* can be caused by primary gonadal disorders (Chapter 22) and several primary adrenal disorders. The adrenal cortex secretes dehydroepiandrosterone and androstenedione, two compounds that can be converted to testosterone in peripheral tissues. Unlike gonadal androgens, ACTH regulates adrenal androgen formation (Fig. 24.46); thus, excess secretion can occur either as a "pure" syndrome or as a component of Cushing disease. The adrenal causes of androgen excess include *adrenocortical neoplasms* and a group of disorders that have been designated *congenital adrenal hyperplasia* (CAH).

Adrenocortical neoplasms associated with virilization are more likely to be androgen-secreting adrenal carcinomas than adenomas. Such tumors are often also associated with hypercortisolism ("mixed syndrome"). They are morphologically identical to other cortical neoplasms and are discussed later.

Congenital adrenal hyperplasia stems from several autosomal recessive, inherited metabolic errors, each characterized by a deficiency of a particular enzyme

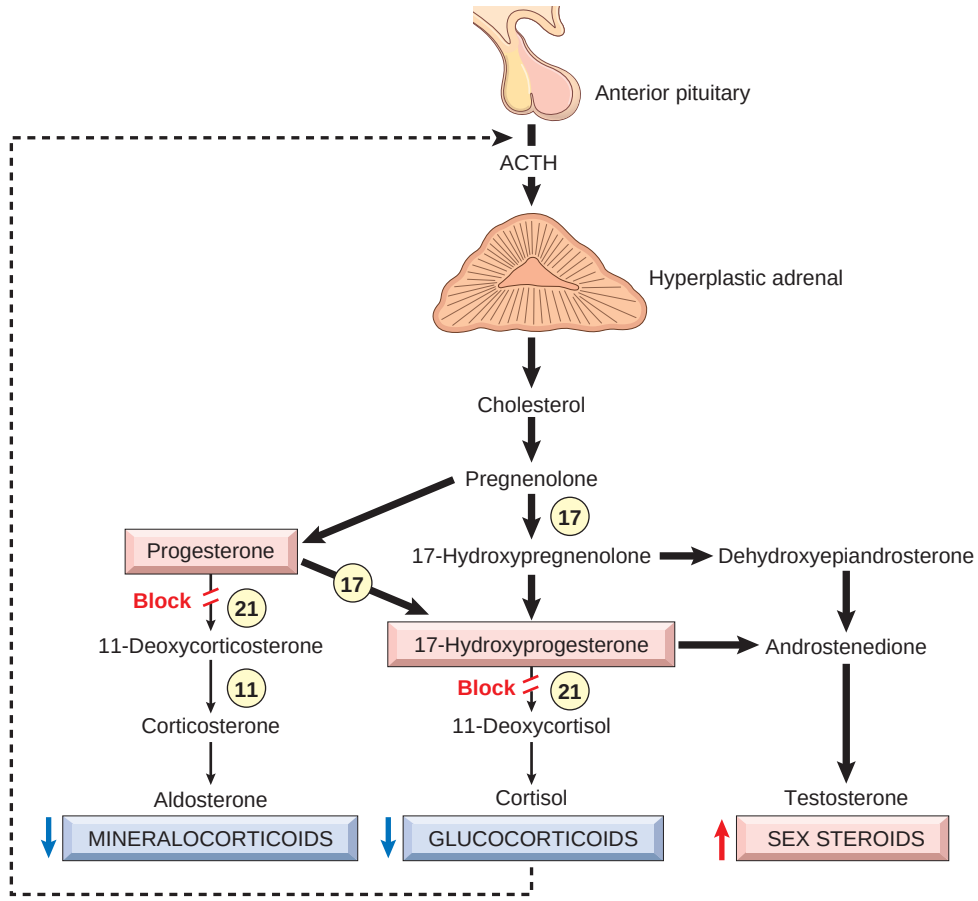


Figure 24.46 Consequences of C-21 hydroxylase deficiency. 21-Hydroxylase deficiency impairs the synthesis of both cortisol and aldosterone at different steps (shown as "Block" in the biosynthesis pathway). The resultant decrease in feedback inhibition *(dashed line)* causes increased secretion of adrenocorticotropic hormone, resulting ultimately in adrenal hyperplasia and increased synthesis of testosterone. The sites of action of 11-, 17-, and 21-hydroxylase are shown as numbers in circles.

involved in the biosynthesis of cortisol (see Fig. 24.46). Steroid precursors that build behind the defective step in the pathway are channeled into other pathways, resulting in increased production of androgens, which accounts for *virilization*. Simultaneously, the deficiency of cortisol leads to increased secretion of ACTH, culminating in adrenal hyperplasia. Certain enzyme defects may also impair aldosterone secretion, adding *salt wasting* to the virilizing syndrome. Other enzyme deficiencies may be incompatible with life or, in rare instances, may involve only the aldosterone pathway.

21-hydroxylase deficiency (caused by mutations of *CYP21A2*) is by far the most common genetic cause of CAH, accounting for over 90% of cases. Fig. 24.46 illustrates normal adrenal steroidogenesis and the consequences of 21-hydroxylase deficiency, which varies in severity depending on the nature of the *CYP21A2* mutation. Three distinctive syndromes have been described: (1) salt-wasting ("classic") adrenogenitalism, (2) simple virilizing adrenogenitalism, and (3) "nonclassic" adrenogenitalism.

- The *salt-wasting syndrome* results from an inability to convert progesterone into deoxycorticosterone because of a total lack of 21-hydroxylase. Thus, there is virtually no synthesis of mineralocorticoids, and, concomitantly, there is a block in the conversion of hydroxyprogesterone into deoxycortisol resulting in deficient cortisol synthesis. This pattern usually comes to light soon after birth, as in utero, electrolyte and fluid balance can be maintained by the maternal kidneys. There is salt wasting, hyponatremia, and hyperkalemia, which induce acidosis, hypotension, cardiovascular collapse, and possibly death. The concomitant block in cortisol synthesis and excess production of androgens, however, lead to virilization, which is easily recognized in the female at birth or in utero. Males with this disorder are generally unrecognized at birth but come to clinical attention 5 to 15 days later because of salt loss, hypotension, and other abnormalities.
- *Simple virilizing adrenogenital syndrome without salt wasting* (presenting as genital ambiguity) occurs in approximately one-third of patients with 21-hydroxylase deficiency. These patients generate sufficient mineralocorticoid to prevent a salt-wasting "crisis." However, the lowered glucocorticoid level fails to cause feedback inhibition of ACTH secretion. Thus, the level of testosterone is increased, with resultant progressive virilization.
- *Nonclassic or late-onset adrenal virilism* is significantly more common than the classic patterns already described. There is only a partial deficiency of 21-hydroxylase, which accounts for the later onset. Individuals with this syndrome may present with mild manifestations, such as hirsutism, acne, and menstrual irregularities. Nonclassical CAH cannot be diagnosed on routine newborn screening, and the diagnosis is usually made by demonstration of biosynthetic defects in steroidogenesis.

MORPHOLOGY

In all cases of CAH, the adrenals are bilaterally hyperplastic, sometimes increasing to 10 to 15 times their normal weight because of the sustained elevation in ACTH. The adrenal cortex is thickened and nodular, and on cut section the widened cortex appears brown because of lipid depletion. The proliferating cells are mostly compact, eosinophilic, lipid-depleted cells intermixed with lipid-laden clear cells. Hyperplasia of corticotroph (ACTH-producing) cells is present in the anterior pituitary in most persons with CAH.

Clinical Features

The clinical features of these disorders are determined by the specific enzyme deficiency and include abnormalities related to androgen excess, with or without aldosterone and glucocorticoid deficiency. CAH also affects the function of the adrenal medulla. Specifically, high levels of intra-adrenal glucocorticoids are required to facilitate medullary catecholamine (epinephrine and norepinephrine) synthesis. In patients with severe salt-wasting 21-hydroxylase deficiency, a combination of low cortisol levels and developmental defects of the medulla *(adrenomedullary dysplasia)* profoundly affects catecholamine secretion, further predisposing these individuals to hypotension and circulatory collapse.

Depending on the nature and severity of the enzymatic defect, the onset of clinical symptoms may occur in the perinatal period, later childhood, or, less commonly, adulthood. For example, in 21-hydroxylase deficiency excessive androgenic activity causes signs of masculinization in females, ranging from clitoral hypertrophy and pseudohermaphroditism in infants, to oligomenorrhea, hirsutism, and acne in postpubertal females. In males, androgen excess is associated with enlargement of the external genitalia and other evidence of precocious puberty in prepubertal patients and oligospermia in older males.

CAH should be suspected in any neonate with ambiguous genitalia. Severe enzyme deficiency in infancy can be a life-threatening condition with vomiting, dehydration, and salt wasting. Individuals with CAH are treated with exogenous glucocorticoids, which, in addition to providing glucocorticoids, also suppress ACTH levels and thus decrease the excessive synthesis of the steroid hormones responsible for many of the clinical abnormalities. Mineralocorticoid supplementation is required in the salt-wasting variants of CAH. With the availability of routine neonatal metabolic screens for CAH and the feasibility of molecular testing for antenatal detection of 21-hydroxylase mutations, the outcome for even the most severe variants has improved significantly.

KEY CONCEPTS

ADRENOGENITAL SYNDROMES

- The adrenal cortex can secrete excess androgens in either of two settings: adrenocortical neoplasms (usually virilizing carcinomas) or congenital adrenal hyperplasia (CAH).
- CAH consists of a group of autosomal recessive disorders characterized by defects in steroid biosynthesis, usually cortisol; the most common subtype is caused by deficiency of the enzyme 21-hydroxylase.
- Reduction in cortisol production causes a compensatory increase in ACTH secretion, which, in turn, stimulates androgen production. Androgens have virilizing effects, including masculinization

in females (ambiguous genitalia, oligomenorrhea, hirsutism), precocious puberty in males, and, in some instances, salt (sodium) wasting and hypotension.
- Bilateral hyperplasia of the adrenal cortex is characteristic, and a subset of 21-hydroxylase–deficient patients also demonstrates adrenomedullary dysplasia.

Adrenocortical Insufficiency

Adrenocortical insufficiency, or hypofunction, may be caused by primary adrenal disease (primary hypoadrenalism) or decreased stimulation of the adrenals due to a deficiency of ACTH (secondary hypoadrenalism) (Table 24.10). Three major patterns of adrenocortical insufficiency are recognized: (1) primary acute adrenocortical insufficiency (adrenal crisis), (2) primary chronic adrenocortical insufficiency *(Addison disease),* and (3) secondary adrenocortical insufficiency.

Primary Acute Adrenocortical Insufficiency

Acute adrenal cortical insufficiency occurs in a variety of clinical settings.
- In individuals with chronic adrenocortical insufficiency, a *crisis* may be precipitated by any form of stress that requires an immediate increase in steroid output to maintain homeostasis
- In patients maintained on exogenous corticosteroids, in whom *rapid withdrawal of steroids* or failure to increase steroid doses in response to an acute stress may precipitate an adrenal crisis because of the inability of the atrophic adrenals to produce glucocorticoid hormones
- As a result of *massive adrenal hemorrhage,* which damages the adrenal cortex sufficiently to cause acute adrenocortical insufficiency—as occurs in newborns following prolonged and difficult delivery with considerable trauma and hypoxia. It also occurs in some patients maintained on anticoagulant therapy, in postsurgical patients who develop disseminated intravascular coagulation and consequent hemorrhagic infarction of the adrenals, and as a complication of disseminated bacterial infection; in this last setting, it is called *Waterhouse-Friderichsen syndrome*.

Waterhouse-Friderichsen Syndrome

This uncommon but catastrophic syndrome is characterized by the following:
- Overwhelming bacterial infection, classically *Neisseria meningitidis* septicemia but occasionally caused by other highly virulent organisms, such as *Pseudomonas* species, pneumococci, *Haemophilus influenzae,* or even staphylococci
- Rapidly progressive hypotension leading to shock
- Disseminated intravascular coagulation associated with widespread purpura, particularly of the skin (Fig. 24.47)
- Rapidly developing adrenocortical insufficiency associated with massive bilateral adrenal hemorrhage

Waterhouse-Friderichsen syndrome can occur at any age but is more common in children. The basis for the adrenal hemorrhage is uncertain but may be due to direct bacterial seeding of small vessels in the adrenal, the development of disseminated intravascular coagulation, or endothelial dysfunction caused by microbial products and inflammatory mediators. Whatever the basis, the adrenals are converted to sacs of clotted blood, which obscures virtually all of the

Table 24.10 Adrenocortical Insufficiency

Primary Insufficiency
Loss of Cortical Cells
Congenital adrenal hypoplasia
X-linked adrenal hypoplasia (*NR0B1* mutation)
Adrenoleukodystrophy (*ABCD1* mutation)
Autoimmune adrenal insufficiency
Autoimmune polyendocrinopathy syndrome type 1 (*AIRE1* gene)
Autoimmune polyendocrinopathy syndrome types 2 and 4 (polygenic)
Infection
Acquired immune deficiency syndrome
Tuberculosis
Fungi
Acute hemorrhagic necrosis (Waterhouse-Friderichsen syndrome)
Amyloidosis, sarcoidosis, hemochromatosis
Metastatic carcinoma
Metabolic Failure in Hormone Production
Congenital adrenal hyperplasia (cortisol and aldosterone deficiency with virilization)
Drug- and steroid-induced inhibition of ACTH or cortical cell function
Secondary Insufficiency
Hypothalamic Pituitary Disease
Neoplasm, inflammation (sarcoidosis, tuberculosis, pyogens, fungi)
Hypothalamic Pituitary Suppression
Long-term steroid administration
Steroid-producing neoplasms

ACTH, Adrenocorticotropic hormone.

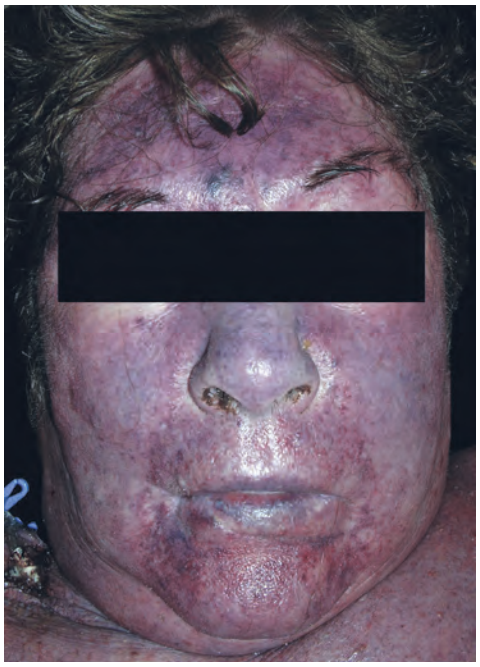


Figure 24.47 Diffuse purpuric rash in a patient with Waterhouse-Friderichsen syndrome. (Reproduced with permission from Vincentelli C, Molina EG, Robinson MJ: Fatal pneumococcal Waterhouse-Friderichsen syndrome in a vaccinated adult with congenital asplenia, *Am J Emerg Med* 27(6):751.e3–751.e5, 2009.)

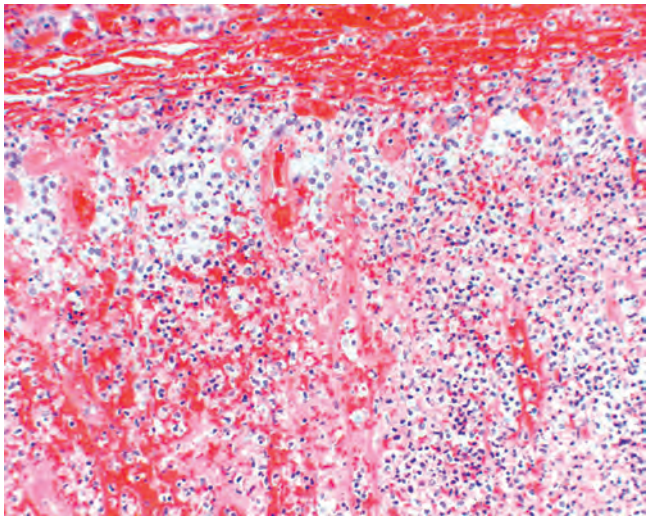


Figure 24.48 Waterhouse-Friderichsen syndrome. At autopsy, the adrenals were grossly hemorrhagic and shrunken; microscopically, little residual cortical architecture is discernible.

underlying detail. Histologic examination reveals that the hemorrhage starts within the medulla near thin-walled venous sinusoids, then suffuses peripherally into the cortex, often leaving islands of recognizable cortical cells (Fig. 24.48). Prompt recognition and appropriate therapy must be instituted immediately, or death follows within hours to a few days.

Primary Chronic Adrenocortical Insufficiency (Addison Disease)

In an article published in 1855, Thomas Addison described a group of patients suffering from a constellation of symptoms, including "general languor and debility, remarkable feebleness of the heart's action, and a peculiar change in the color of the skin" associated with disease of the "suprarenal capsules" or, in more modern parlance, the adrenal glands. Addison disease, or chronic adrenocortical insufficiency, is an uncommon disorder resulting from progressive destruction of the adrenal cortex. Typically, clinical manifestations of adrenocortical insufficiency do not appear until at least 90% of the adrenal cortex has been compromised. The causes of chronic adrenocortical insufficiency are listed in Table 24.10. Although all races and both sexes may be affected, certain causes of Addison disease (e.g., autoimmune adrenalitis) are much more common in Caucasians and in women.

Pathogenesis

A large number of diseases may affect the adrenal cortex, including lymphoma, amyloidosis, sarcoidosis, hemochromatosis, fungal infections, and adrenal hemorrhage, but more than 90% of all cases are attributable to one of four disorders: autoimmune adrenalitis, tuberculosis, AIDS, or metastatic cancers.

- *Autoimmune adrenalitis* accounts for 80% to 90% of cases of primary adrenal insufficiency in high-income countries. As the name implies, there is autoimmune destruction of steroidogenic cells, and autoantibodies to several key steroidogenic enzymes (21-hydroxylase, 17-hydroxylase) are detected in the majority of patients. Autoimmune adrenalitis occurs as an isolated condition in 40% of cases, or as part of an autoimmune polyendocrinopathy syndrome in the remaining 60%. The latter encompasses several distinct entities:
  - *APS-1,* also known as autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy, is characterized by chronic mucocutaneous candidiasis and abnormalities of skin, dental enamel, and nails (ectodermal dystrophy) in association with a combination of organ-specific autoimmune disorders (autoimmune adrenalitis, autoimmune hypoparathyroidism, idiopathic hypogonadism, pernicious anemia) that result in immune destruction of target organs. APS-1 is caused by mutations in the autoimmune regulator *(AIRE)* gene on chromosome 21q22. AIRE is expressed primarily in the thymus, where it functions as a transcription factor that promotes the expression of many peripheral tissue antigens. Self-reactive T cells that recognize these antigens are eliminated (Chapter 6). In the absence of AIRE function, central T-cell tolerance to peripheral tissue antigens is compromised, promoting autoimmunity. Individuals with APS-1 develop autoantibodies against IL-17 and IL-22, which are the principal effector cytokines secreted by Th17 T cells (Chapter 6). Because these two Th17-derived cytokines are crucial for defense against fungal infections, it is not surprising that patients develop chronic mucocutaneous candidiasis.
  - *APS-2* usually starts in the fourth decade of life and presents as a combination of adrenal insufficiency and autoimmune thyroiditis, with or without T1D. Unlike in APS-1, mucocutaneous candidiasis, ectodermal dysplasia, and autoimmune hypoparathyroidism do not develop. Overall, APS-2 is more prevalent than APS-1. In contrast, APS-4 is a rare condition characterized by adrenalitis plus other autoimmune phenomena (e.g., gastritis, vitiligo, alopecia, pernicious anemia) but neither thyroiditis nor T1D. The pathogenesis of these other autoimmune polyendocrinopathies is not understood.
- *Infections,* particularly tuberculosis and those produced by fungi, may also cause primary chronic adrenocortical insufficiency. *Tuberculous adrenalitis,* which once accounted for as much as 90% of cases of Addison disease, has become less common with the development of antimycobacterial drugs. With the resurgence of tuberculosis in most urban centers and the persistence of the disease in lower-income countries, however, this cause of adrenal insufficiency must be kept in mind. When present, tuberculous adrenalitis is usually associated with active infection in other sites, particularly in the lungs and genitourinary tract. Among fungi, disseminated infections caused by *Histoplasma capsulatum* and *Coccidioides immitis* may result in chronic adrenocortical insufficiency. AIDS patients are at risk for developing adrenal insufficiency from several infectious (cytomegalovirus, *Mycobacterium avium-intracellulare*) and noninfectious (Kaposi sarcoma) complications.
- *Metastatic neoplasms* involving the adrenals are another cause of adrenal insufficiency. The adrenals are a fairly common site for metastases in patients with disseminated carcinomas. Although adrenal function is preserved in

most such patients, the metastatic tumors occasionally destroy enough adrenal cortex to produce a degree of adrenal insufficiency. Carcinomas of the lung and breast are the source of a majority of metastases, although many other neoplasms, including gastrointestinal carcinomas, malignant melanoma, and hematopoietic neoplasms, may also metastasize to the adrenals.

- *Genetic causes of adrenal insufficiency* include congenital adrenal hypoplasia *(adrenal hypoplasia congenita)* and *adrenoleukodystrophy*. Adrenoleukodystrophy is caused by mutations of the ATP-binding cassette, subfamily D, member 1 *(ABCD1)* gene and is further described in Chapter 28. Congenital adrenal hypoplasia is a rare X-linked disease caused by mutations in the *NR0B1* gene that encodes DAX1, a transcription factor implicated in adrenal development.

MORPHOLOGY

The anatomic changes in the adrenal glands depend on the underlying disease. **Primary autoimmune adrenalitis** is characterized by irregularly shrunken glands, which may be difficult to identify within the suprarenal adipose tissue. Histologically, the cortex contains only scattered residual cortical cells in a collapsed network of connective tissue. A variable lymphoid infiltrate is present in the cortex and may extend into the adjacent medulla, although the medulla is otherwise preserved (Fig. 24.49). In cases of **tuberculous and fungal disease,** the adrenal architecture is effaced by a granulomatous inflammatory reaction identical to that encountered in other sites of infection. When hypoadrenalism is caused by **metastatic carcinoma,** the adrenals are enlarged and the normal architecture is obscured by the infiltrating neoplasm.

Clinical Features

Addison disease begins insidiously and does not come to attention until the levels of circulating glucocorticoids and mineralocorticoids are significantly decreased. The initial manifestations include progressive weakness and easy fatigability, which may be dismissed as nonspecific complaints. Gastrointestinal disturbances are common and include anorexia, nausea, vomiting, weight loss, and diarrhea. In individuals with primary adrenal disease, hyperpigmentation of the skin, particularly of sun-exposed areas and at pressure points, such as the neck, elbows, knees, and knuckles, is quite characteristic. This is caused by elevated levels of pro-opiomelanocortin (POMC), which is derived from the anterior pituitary and is a precursor of both ACTH and melanocyte-stimulating hormone (MSH). By contrast, hyperpigmentation is not seen in persons with adrenocortical insufficiency caused by primary pituitary or hypothalamic disease. Decreased mineralocorticoid activity in persons with primary adrenal insufficiency results in potassium retention and sodium loss, with consequent hyperkalemia, hyponatremia, volume depletion, and hypotension. Hypoglycemia may occasionally occur as a result of glucocorticoid deficiency and impaired gluconeogenesis. Stresses such as infections, trauma, or surgical procedures in such patients can precipitate an acute adrenal crisis, manifested by intractable vomiting, abdominal pain, hypotension, coma, and vascular collapse. Death occurs rapidly unless corticosteroid therapy begins immediately.

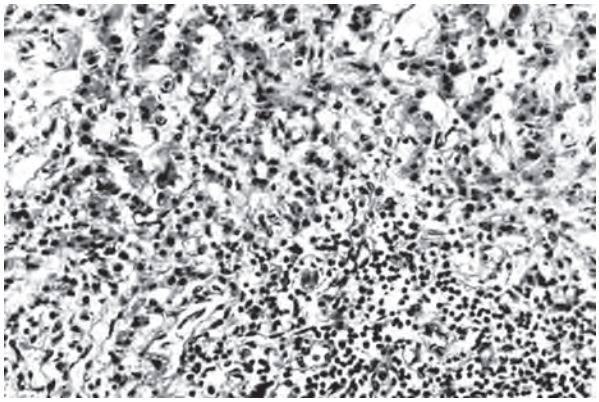


Figure 24.49 Autoimmune adrenalitis. In addition to loss of all but a subcapsular rim of cortical cells, there is an extensive mononuclear cell infiltrate.

KEY CONCEPTS

ADRENOCORTICAL INSUFFICIENCY (HYPOADRENALISM)

- Primary adrenocortical insufficiency can be acute (Waterhouse-Friderichsen syndrome) or chronic (Addison disease)
- Chronic adrenal insufficiency in the developed world most often is secondary to autoimmune adrenalitis, which occurs most commonly in the context of one of two autoimmune polyendocrine syndromes: APS-1 (caused by mutations in the *AIRE* gene) or APS-2. APS-1 is characterized by an autoimmune attack against multiple endocrine organs and autoantibodies against IL-17 and IL-22.
- Tuberculosis and infections due to opportunistic pathogens associated with the human immunodeficiency virus and tumors metastatic to the adrenals are the other important causes of chronic hypoadrenalism.
- Patients typically present with fatigue, weakness, and gastrointestinal disturbances. Primary adrenocortical insufficiency also is characterized by high ACTH levels with associated skin pigmentation.

Secondary Adrenocortical Insufficiency

Any disorder of the hypothalamus and pituitary, such as metastatic cancer, infection, infarction, or irradiation, that reduces the output of ACTH leads to a syndrome of hypoadrenalism that has many similarities to Addison disease. Analogously, prolonged administration of exogenous glucocorticoids suppresses the output of ACTH and adrenal function. With secondary disease, the hyperpigmentation of primary Addison disease is lacking, because levels of melanocyte-stimulating hormone are not elevated. The manifestations also differ in that secondary hypoadrenalism is characterized by deficient cortisol and androgen output but normal or near-normal aldosterone synthesis. Thus, in adrenal insufficiency secondary to pituitary malfunction, marked hyponatremia and hyperkalemia are not seen.

ACTH deficiency can occur alone, but in some instances, it is only one component of panhypopituitarism, associated with multiple trophic hormone deficiencies. Secondary

disease can be differentiated from Addison disease by demonstration of low levels of plasma ACTH in the former. In patients with primary disease, the destruction of the adrenal cortex reduces the response to exogenously administered ACTH, whereas in those with secondary hypofunction, there is a prompt rise in plasma cortisol levels.

MORPHOLOGY

In cases of hypoadrenalism secondary to hypothalamic or pituitary disease **(secondary hypoadrenalism),** depending on the severity of ACTH deficiency, the adrenals may be moderately to markedly decreased in size. The small, flattened glands usually retain their yellow color as a result of a small amount of residual lipid. The cortex may be reduced to a thin ribbon composed largely of zona glomerulosa. The medulla is not affected.

Adrenocortical Neoplasms

It should be evident from the preceding sections that functional adrenal neoplasms may be responsible for any of the various forms of hyperadrenalism. Adenomas and carcinomas are about equally common in adults; in children, carcinomas predominate. While most cortical neoplasms are sporadic, two familial cancer syndromes are associated with a predisposition for developing adrenocortical carcinomas: Li-Fraumeni syndrome, caused by germline *TP53* mutations (Chapter 7), and Beckwith-Wiedemann syndrome, a disorder of epigenetic imprinting involving the gene for insulin-like growth factor 2 (IGF-2) (Chapter 10). Interestingly, somatic *TP53* mutations and overexpression of IGF-2 also occur in sporadic adrenocortical carcinomas, reiterating an oft-observed phenomenon of shared molecular mechanisms in familial and sporadic neoplasms.

Functional adenomas are most commonly associated with hyperaldosteronism and Cushing syndrome, whereas a virilizing neoplasm is more likely to be a carcinoma. However, not all adrenocortical neoplasms elaborate steroid hormones. Determination of functionality is based on clinical evaluation, and measurement of hormones or hormone metabolites in the blood.

MORPHOLOGY

Most **adrenocortical adenomas** are clinically silent and are usually incidental findings at autopsy or during abdominal imaging for an unrelated cause (see the discussion of adrenal "incidentalomas" later). The typical cortical adenoma is a well-circumscribed, nodular lesion up to 2.5 cm in diameter that expands the adrenal (Fig. 24.50). In contrast to a functional adenoma, which is associated with atrophy of the adjacent cortex, the cortex adjacent to a nonfunctional adenoma is normal. On cut surface, adenomas are usually yellow to yellow-brown because of the presence of lipid.

Microscopically, adenomas are composed of cells similar to those populating the normal adrenal cortex. The nuclei tend to be small, although some degree of pleomorphism may be encountered, even in benign lesions ("endocrine atypia"). The cytoplasm of the neoplastic cells ranges from eosinophilic to vacuolated, depending on their lipid content (Fig. 24.51). Mitotic activity is generally inconspicuous, with a Ki-67 labeling index less than 5% (in contrast to adrenocortical carcinomas). Adrenocortical lesions—both adenomas and carcinomas—express steroidogenic factor-1 (SF-1) and inhibin-alpha.

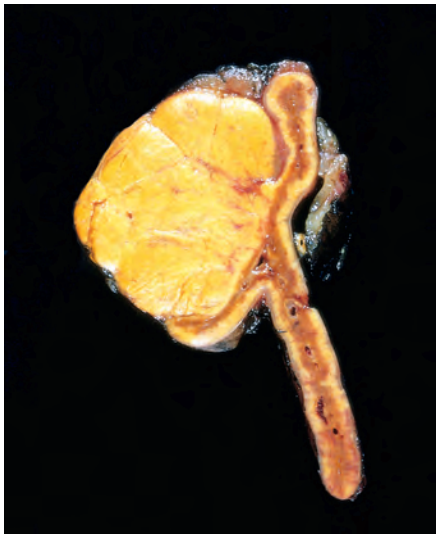


Figure 24.50 Adrenal cortical adenoma. The adenoma is distinguished from nodular hyperplasia by its solitary, circumscribed nature. The functional status of an adrenal cortical adenoma cannot be predicted from its gross or microscopic appearance.

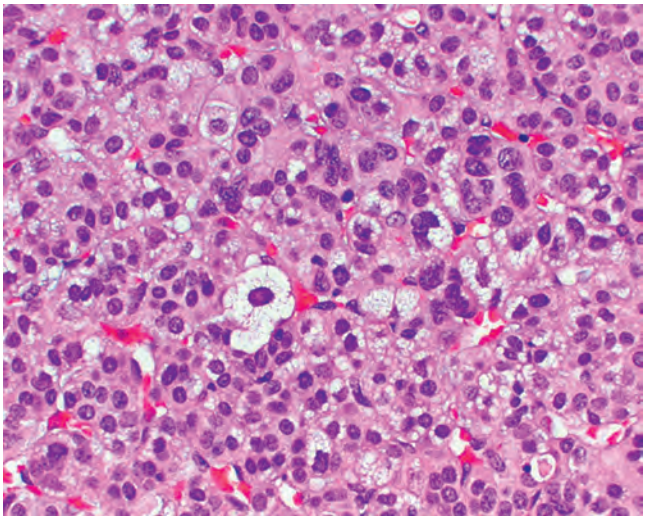


Figure 24.51 Histologic features of an adrenal cortical adenoma. The neoplastic cells are vacuolated because of the presence of intracytoplasmic lipid. There is mild nuclear pleomorphism. Mitotic activity and necrosis are not seen.

Adrenocortical carcinomas are rare neoplasms that have a bimodal distribution in the first and fifth decades of life. They are more likely to be functional than adenomas and are often associated with virilism or other clinical manifestations of hyperadrenalism. In most cases, adrenocortical carcinomas are large, invasive lesions, many exceeding 20 cm in diameter, which efface the native adrenal gland (Fig. 24.52). The less common, smaller, and better-circumscribed lesions may be difficult to distinguish from an adenoma. On cut surface, adrenocortical carcinomas are typically variegated, poorly demarcated lesions containing areas of necrosis, hemorrhage, and cystic change. Adrenal cancers have

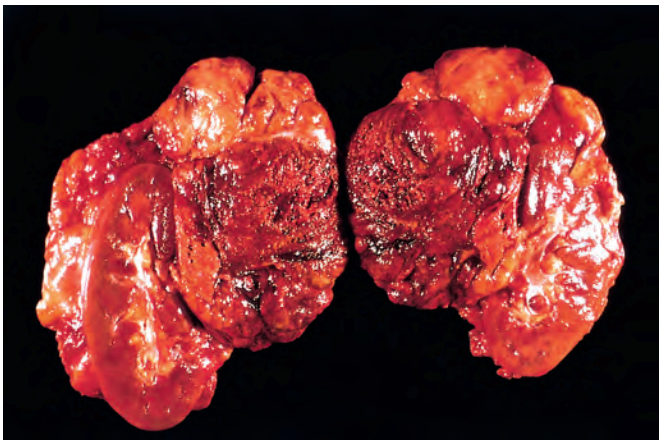


Figure 24.52 Adrenal carcinoma. The hemorrhagic and necrotic tumor dwarfs the kidney and compresses the upper pole.

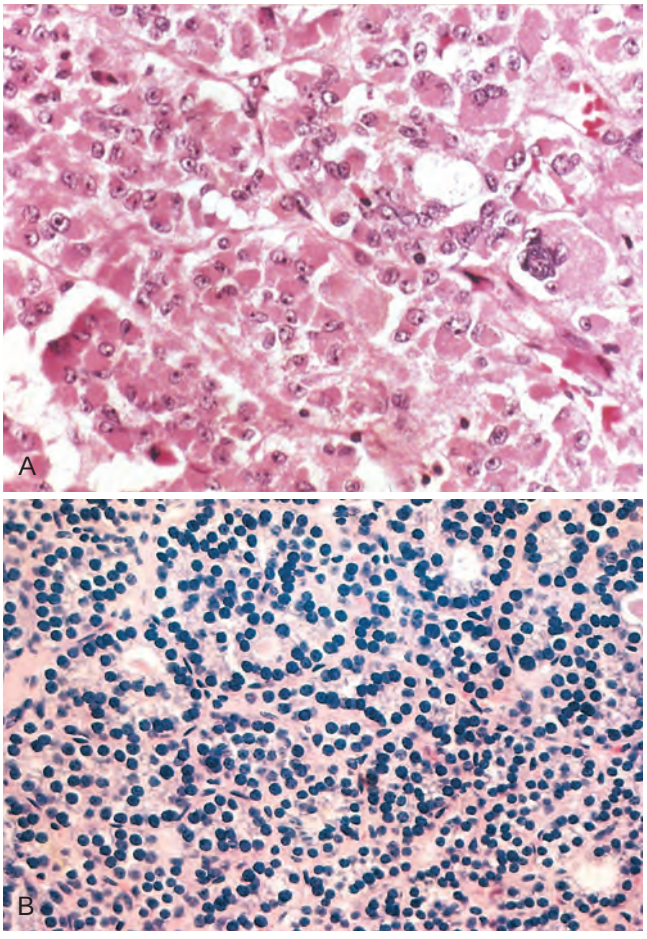


Figure 24.53 Adrenal carcinoma (A) revealing marked anaplasia, contrasted with normal adrenal cortical cells (B).

a strong tendency to invade the adrenal vein, vena cava, and lymphatics. Metastases to regional and periaortic nodes are common, as is distant hematogenous spread to the lungs and other viscera. The median patient survival is about 2 years.

Microscopically, adrenocortical carcinomas may be composed of well-differentiated cells, resembling those seen in cortical adenomas, or bizarre, monstrous giant cells (Fig. 24.53), which may be difficult to distinguish from those of an undifferentiated carcinoma metastatic to the adrenal. Between these extremes are cancers with moderate degrees of anaplasia, some composed predominantly of spindle cells. In addition to higher Ki-67 index, carcinomas also have significantly higher IGF-2 levels compared to adrenocortical adenomas.

Carcinomas, particularly those of bronchogenic origin, may metastasize to the adrenals and may be difficult to differentiate from primary cortical carcinomas. Of note, metastases to the adrenal cortex are significantly more common than primary adrenocortical carcinomas.

Other Adrenal Lesions

Adrenal cysts are relatively uncommon; however, with the use of sophisticated abdominal imaging techniques, the frequency of detection of these lesions is increasing. Larger cysts may produce an abdominal mass and flank pain. Both cortical and medullary neoplasms may undergo necrosis and cystic degeneration and may present as "nonfunctional cysts."

Adrenal myelolipomas are unusual benign lesions composed of mature fat and hematopoietic cells. Although most of these lesions represent incidental findings, occasional myelolipomas may reach massive proportions. Histologically, mature adipocytes are admixed with aggregates of differentiating hematopoietic cells belonging to all three lineages. Foci of myelolipomatous change may be seen in cortical tumors and in adrenals with cortical hyperplasia.

The term *adrenal incidentaloma* is a half-facetious moniker that has crept into the medical lexicon as advancements in medical imaging have led to the incidental discovery of adrenal masses in asymptomatic individuals or in individuals in whom the presenting complaint is not directly related to the adrenal gland. The estimated population prevalence of "incidentalomas" discovered by imaging is approximately 4%, with an age-dependent increase in prevalence. Fortunately, the vast majority of these lesions are small nonsecreting cortical adenomas of no clinical importance.

ADRENAL MEDULLA

The adrenal medulla is developmentally, functionally, and structurally distinct from the adrenal cortex. It is composed of specialized neural crest (neuroendocrine) cells, termed *chromaffin* cells, and their supporting (sustentacular) cells. The adrenal medulla is the major source of catecholamines (epinephrine, norepinephrine) in the body. Neuroendocrine cells similar to chromaffin cells are widely dispersed in an extra-adrenal system of clusters and nodules that, together with the adrenal medulla, make up the *paraganglion system*. These extra-adrenal paraganglia are closely associated with the autonomic nervous system and can be divided into three groups based on their anatomic distribution: (1) branchiomeric, (2) intravagal, and (3) aorticosympathetic. The branchiomeric and intravagal paraganglia associated with the parasympathetic system are located close to the major arteries and cranial nerves of the head and neck and

include the carotid bodies (Chapter 16). The intravagal paraganglia, as the term implies, are distributed along the vagus nerve. The aorticosympathetic chain is found in association with segmental ganglia of the sympathetic system and therefore is distributed mainly alongside of the abdominal aorta. The organs of Zuckerkandl, close to the aortic bifurcation, belong to this group.

The most important diseases of the adrenal medulla are neoplasms, which include neoplasms of chromaffin cells *(pheochromocytomas)* and neuronal neoplasms *(neuroblastic tumors)*. Neuroblastomas and other neuroblastic tumors are discussed in Chapter 10.

Pheochromocytoma

Pheochromocytomas are neoplasms composed of chromaffin cells, which synthesize and release catecholamines and, in some instances, peptide hormones. It is important to recognize these tumors because they are a rare cause of surgically correctable hypertension. Traditionally, the features of pheochromocytomas have been summarized by the "rule of 10s".

- *Ten percent of pheochromocytomas are extra-adrenal*, occurring in sites such as the organs of Zuckerkandl and the carotid body. Pheochromocytomas that develop in extra-adrenal paraganglia are designated *paragangliomas* and are discussed in Chapter 16.
- *Ten percent of sporadic adrenal pheochromocytomas are bilateral*; this figure may rise to as high as 50% in cases that are associated with familial tumor syndromes (see later).
- *Ten percent of adrenal pheochromocytomas are biologically malignant,* defined by the presence of metastatic disease. Malignancy is more common (20% to 40%) in extra-adrenal paragangliomas, and in tumors arising in the setting of certain germline mutations (see later).
- *Ten percent of adrenal pheochromocytomas are not associated with hypertension*. Of the 90% that present with hypertension, approximately two-thirds have "paroxysmal" episodes associated with a sudden rise in blood pressure and palpitations, which can, on occasion, be fatal.

One "traditional" 10% rule that has now been modified pertains to familial cases. It is now recognized that as many as 25% of individuals with pheochromocytomas and paragangliomas harbor an oncogenic germline mutation. These mutations can involve at least a dozen genes, the most common of which are listed in Table 24.11. Patients with germline mutations are typically younger at presentation than those with sporadic tumors and more often have bilateral tumors. The affected genes fall into two broad classes, those that enhance growth factor receptor pathway signaling (e.g., *RET, NF1*) and those that increase the activity and stability of two hypoxia-induced transcription factors, HIF-1α and HIF-2α (thus, creating a so-called "pseudohypoxia" phenotype). You will recall that the *VHL* gene encodes a tumor-suppressor protein that is needed for the oxygen-dependent degradation of HIF-1α and that is disrupted by loss-of-function mutations in patients with von Hippel-Lindau (VHL) syndrome. VHL syndrome is associated with a number of tumors, including pheochromocytomas and paragangliomas. Similarly, *EPAS1* encodes HIF-2α, and gain-of-function mutations are found in individuals with the "polycythemia paraganglioma syndrome" that stabilize HIF-2α by mutating amino acid residues that are required for its degradation. Other familial cases of pheochromocytoma are associated with germline mutations in genes encoding several components of the succinate dehydrogenase complex (most commonly, *SDHB, SDHC,* and *SDHD*). This complex is involved in mitochondrial electron transport and oxygen sensing, and it is believed that these mutations also lead to up-regulation of HIF protein, recapitulating "pseudohypoxia." Recent integrated molecular analyses of pheochromocytomas and paragangliomas have shown that many of the genes implicated in a germline predisposition to these neoplasms (e.g., *NF1, VHL, EPAS1,* and *RET*) also are involved by somatic mutations in sporadic tumors.

Table 24.11 Familial Syndromes Associated With Pheochromocytoma and Extra-Adrenal Paragangliomas

Syndrome	Gene	Associated Lesion	Other Features
Multiple endocrine neoplasia, type 2A (MEN-2A)	*RET*	Pheochromocytoma/paraganglioma	Medullary thyroid carcinoma Parathyroid hyperplasia
Multiple endocrine neoplasia, type 2B (MEN-2B)	*RET*	Pheochromocytoma/paraganglioma	Medullary thyroid carcinoma Marfanoid habitus Mucocutaneous GNs
Neurofibromatosis, type 1 (NF1)	*NF1*	Pheochromocytoma	Neurofibromatosis Café-au-lait spots Optic nerve glioma
Von Hippel-Lindau (VHL)	*VHL*	Pheochromocytoma/paraganglioma	Renal cell carcinoma Hemangioblastoma Pancreatic endocrine neoplasm
Hereditary paraganglioma 1	*SDHD*	Pheochromocytoma, paraganglioma	GIST
Hereditary paraganglioma 3	*SDHC*	Paraganglioma	GIST
Hereditary paraganglioma 4	*SDHB*	Pheochromocytoma, paraganglioma	GIST
Polycythemia paraganglioma syndrome	*EPAS1*	Pheochromocytoma, paraganglioma	Polycythemia

GIST, Gastrointestinal stromal tumor; *GN,* ganglioneuroma; *NF1*, neurofibromin; *SDHB*, succinate dehydrogenase complex, subunit B; *SDHC*, succinate dehydrogenase complex, subunit C; *SDHD*, succinate dehydrogenase complex, subunit D; *EPAS1*, endothelial PAS-domain containing 1 (encodes for HIF-2alpha).
Note: Additional rare syndromes associated with pheochromocytoma or paragangliomas are not included in this table.

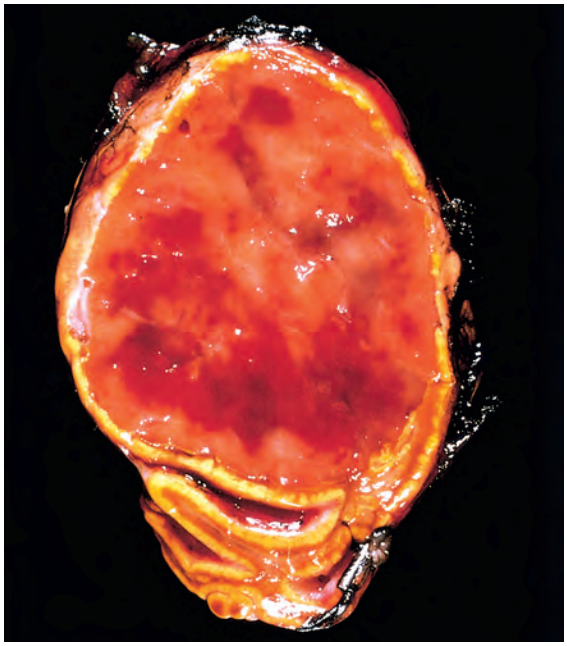


Figure 24.54 Pheochromocytoma. The tumor is enclosed within an attenuated cortex and demonstrates areas of hemorrhage. The comma-shaped residual adrenal is seen below. (Courtesy Dr. Jerrold R. Turner, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

MORPHOLOGY

Pheochromocytomas range from small, circumscribed lesions confined to the adrenal (Fig. 24.54) to large hemorrhagic masses weighing kilograms. The average weight of a pheochromocytoma is 100 g, but weights from just over 1 g to almost 4000 g have been reported. The larger tumors are well demarcated by either connective tissue or compressed cortical or medullary tissue. Richly vascularized fibrous trabeculae within the tumor produce a lobular pattern. In many tumors, remnants of the adrenal gland can be seen, stretched over the surface or attached at one pole. On section, the cut surfaces of smaller pheochromocytomas are yellow-tan. Larger lesions tend to be hemorrhagic, necrotic, and cystic and typically efface the adrenal gland. Incubation of fresh tissue with a potassium dichromate solution turns the tumor a dark brown color due to oxidation of stored catecholamines, thus the term **chromaffin.**

The histologic pattern in pheochromocytoma is quite variable. The tumors are composed of clusters of polygonal to spindle-shaped chromaffin cells or chief cells that are surrounded by supporting sustentacular cells, creating small nests or alveoli **(zellballen)** that are supplied by a rich vascular network (Fig. 24.55). Uncommonly, the dominant cell type is a spindle or small cell; various patterns can be found in any one tumor. The cytoplasm has a finely granular appearance, best demonstrated with silver stains, due to the presence of granules containing catecholamines. The nuclei are usually round to ovoid, with a stippled "salt-and-pepper" chromatin that is characteristic of neuroendocrine tumors. Electron microscopy reveals variable numbers of membrane-bound, electron-dense secretory granules (Fig. 24.56). Immunoreactivity for neuroendocrine markers (chromogranin and synaptophysin) is seen in the chief cells, while the peripheral sustentacular cells stain with antibodies against S-100, a calcium-binding protein expressed by a variety of mesenchymal cell types.

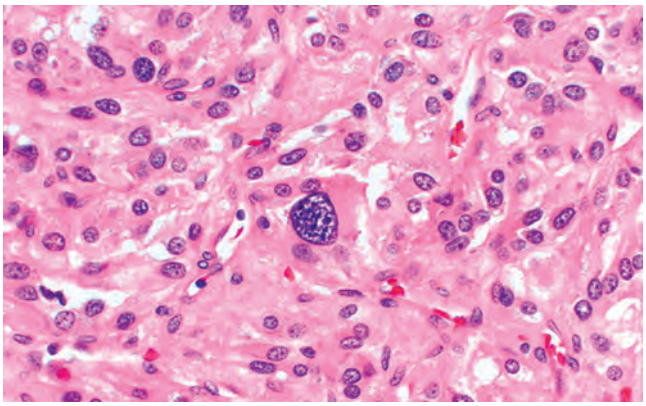


Figure 24.55 Pheochromocytoma demonstrating characteristic nests of cells ("zellballen") with abundant cytoplasm. Granules containing catecholamine are not visible in this preparation. It is not uncommon to find bizarre cells, even in pheochromocytomas that are biologically benign. (Courtesy Dr. Jerrold R. Turner, Department of Pathology, Brigham and Women's Hospital, Boston, Mass.)

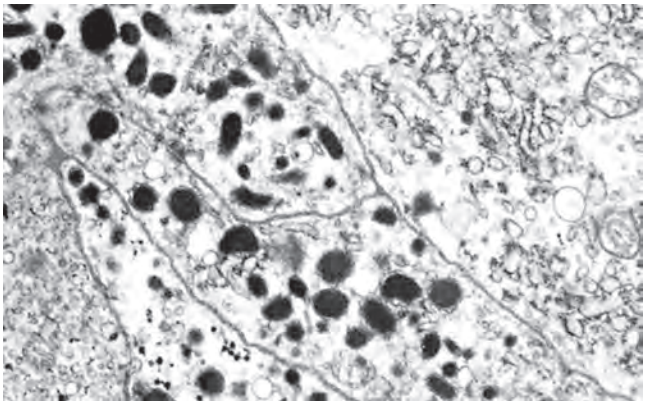


Figure 24.56 Electron micrograph of pheochromocytoma. This tumor contains membrane-bound secretory granules in which catecholamines are stored (30,000×).

Determining malignancy in pheochromocytomas can be vexing. Several histologic features, such as numbers of mitoses, confluent tumor necrosis, and spindle cell morphology, have been associated with an aggressive behavior and increased risk of metastasis, but are not entirely reliable. Tumors with "benign" histologic features may metastasize, while bizarrely pleomorphic tumors may remain confined to the adrenal gland. In fact, cellular and nuclear pleomorphism, including the presence of giant cells, and mitotic figures are often seen in benign pheochromocytomas, while cellular monotony is paradoxically associated with aggressive behavior. Even capsular and vascular invasion may be encountered in benign lesions. Therefore, the definitive diagnosis of malignancy in pheochromocytomas is based exclusively on the presence of metastases. These may involve regional lymph nodes as well as more distant sites, including liver, lung, and bone.

Clinical Features

The dominant clinical manifestation of pheochromocytoma is *hypertension,* observed in 90% of patients. Approximately two-thirds of patients with hypertension demonstrate *paroxysmal episodes,* which are described as an abrupt, precipitous

elevation in blood pressure associated with tachycardia, palpitations, headache, sweating, tremor, and a sense of apprehension. These episodes may also be associated with pain in the abdomen or chest, nausea, and vomiting. Isolated paroxysmal episodes of hypertension occur in less than one-half of patients; more commonly, patients demonstrate chronic, sustained elevation in blood pressure punctuated by the aforementioned paroxysms. The paroxysms may be precipitated by emotional stress, exercise, changes in posture, and palpation in the region of the tumor; urinary bladder paragangliomas occasionally precipitate a paroxysm during micturition. The elevations of blood pressure are induced by the sudden release of catecholamines that may acutely precipitate congestive heart failure, pulmonary edema, myocardial infarction, ventricular fibrillation, and cerebrovascular accident.

The cardiac complications have been attributed to what has been called *catecholamine cardiomyopathy,* or catecholamine-induced myocardial instability and ventricular arrhythmias. Nonspecific myocardial changes, such as focal necrosis, mononuclear infiltrates, and interstitial fibrosis, have been attributed either to ischemic damage secondary to catecholamine-induced constriction of myocardial blood vessels or to direct catecholamine toxicity. In some cases, pheochromocytomas secrete other hormones, such as ACTH and somatostatin, and may therefore be associated with clinical features related to the secretion of these or other peptide hormones. Laboratory diagnosis of pheochromocytoma is based on the demonstration of increased urinary excretion of free catecholamines and their metabolites, such as vanillylmandelic acid and metanephrines.

Isolated benign tumors are treated with surgical excision after preoperative and intraoperative medication with adrenergic-blocking agents to prevent a hypertensive crisis. Multifocal lesions require long-term medical treatment for hypertension.

Multiple Endocrine Neoplasia Syndromes

The MEN syndromes are a group of inherited diseases resulting in proliferative lesions (hyperplasia, adenomas, and carcinomas) of multiple endocrine organs. Like other inherited cancer disorders (Chapter 7), endocrine tumors arising in the context of MEN syndromes have certain distinct features that contrast with their sporadic counterparts.

- Tumors occur at a *younger age* than sporadic tumors.
- They arise in *multiple endocrine organs,* either *synchronously* (at the same time) or *metachronously* (at different times).
- Even in one organ, the tumors are often *multifocal*.
- The tumors are usually preceded by an *asymptomatic stage of hyperplasia* involving the cell of origin. For example, individuals with MEN-2 almost universally demonstrate C-cell hyperplasia in the thyroid parenchyma adjacent to medullary thyroid carcinomas.
- These tumors are usually *more aggressive* and *recur* in a higher proportion of cases than do similar sporadic endocrine tumors.

MULTIPLE ENDOCRINE NEOPLASIA, TYPE 1

MEN-1, or *Wermer syndrome,* is a rare heritable disorder with a prevalence of about 2 per 100,000. MEN-1 is characterized by abnormalities involving the parathyroid, pancreas, and pituitary gland; thus the mnemonic device, the three P's:

- *Parathyroid:* Primary hyperparathyroidism is the most common manifestation of MEN-1 (80% to 95% of patients) and is the initial manifestation of the disorder in most patients, appearing in almost all patients by 40 to 50 years of age. Parathyroid abnormalities include both hyperplasia and adenomas.
- *Pancreas:* Endocrine tumors of the pancreas are a leading cause of morbidity and mortality in persons with MEN-1. These tumors are usually aggressive and often present with metastatic disease. It is not uncommon to find multiple "microadenomas" scattered throughout the pancreas in conjunction with one or two dominant lesions. MEN-1–associated pancreatic endocrine tumors are often functional; however, because pancreatic polypeptide is the most commonly secreted product, many tumors fail to produce an endocrine hypersecretion syndrome. Among those that do, Zollinger-Ellison syndrome (associated with gastrinomas) and hypoglycemia and neurologic manifestations (associated with insulinomas) are most common.
- *Pituitary:* The most frequent anterior pituitary tumor encountered in MEN-1 is a prolactinoma; some patients develop acromegaly from somatotropin-secreting tumors.
- It is now recognized that the spectrum of this disease extends beyond the three *P's*. The duodenum is the most common site of gastrinomas in individuals with MEN-1 (far in excess of the frequency of pancreatic gastrinomas), and synchronous duodenal and pancreatic tumors may be present in the same individual. In addition, carcinoid tumors, thyroid and adrenocortical adenomas, and lipomas are more frequent than in the general population.

MEN-1 syndrome is caused by germline mutations in the *MEN1* tumor suppressor gene, which encodes a protein called menin. Menin is a component of several different transcription factor complexes, which (depending on the specific binding partner) may either promote or inhibit tumorigenesis. This dichotomy in menin function is best exemplified in the interactions of menin with two oncogenic transcription factors—JunD and KMT2A (previously known as MLL), a chromatin-modifying histone methyltransferase. When menin partners with JunD, it blocks transcriptional activation and cellular proliferation induced by JunD; in fact, loss of this tumor-suppressor interaction is believed to contribute to the MEN observed in the setting of *MEN-1* inactivating mutations. On the contrary, the association of wild-type menin with KMT2A leads to the formation of a tumor-promoting transcriptional complex in a subset of acute leukemias through upregulated expression of *HOX* genes (Chapter 13).

The dominant manifestations of MEN-1 usually result from the peptide hormones that are overproduced and include such abnormalities as recurrent hypoglycemia due to insulinomas, intractable peptic ulcers in persons with Zollinger-Ellison syndrome, nephrolithiasis caused by PTH-induced hypercalcemia, or symptoms of prolactin excess from a pituitary tumor. As expected, malignant behavior by one or more of the endocrine tumors arising in these patients is often the proximate cause of death.

MULTIPLE ENDOCRINE NEOPLASIA, TYPE 2

MEN-2 is subclassified into three distinct syndromes: MEN-2A, MEN-2B, and MEN-4.

- *MEN-2A*, or *Sipple syndrome,* is characterized by *pheochromocytoma, medullary carcinoma of the thyroid,* and *parathyroid hyperplasia* (see Table 24.11). Medullary carcinomas of the thyroid occur in almost 100% of patients. They are usually multifocal and are virtually always associated with foci of C-cell hyperplasia in the adjacent thyroid. The medullary carcinomas may elaborate calcitonin and other active products and are usually clinically aggressive. Among individuals with MEN-2A, 40% to 50% have pheochromocytomas, which are often bilateral and may arise in extra-adrenal sites. Parathyroid hyperplasia and evidence of hypercalcemia or renal stones occur in 10% to 20% of patients. **MEN-2A is clinically and genetically distinct from MEN-1 and is caused by germline gain-of-function mutations in the *RET* proto-oncogene**. As noted earlier, the *RET* proto-oncogene encodes a receptor tyrosine kinase that binds glial-derived neurotrophic factor and other related ligands and transmits growth and differentiation signals (Chapter 7). Loss-of-function mutations in *RET* result in intestinal aganglionosis and Hirschsprung disease (Chapter 17). In contrast, in MEN-2A (as well as in MEN-2B), germline mutations constitutively activate the RET receptor.
- *MEN-2B* has significant clinical overlap with MEN-2A. Patients develop medullary thyroid carcinomas, which are usually multifocal and more aggressive than in MEN-2A, and pheochromocytomas. However, unlike in MEN-2A, primary hyperparathyroidism is not present. In addition, MEN-2B is accompanied by neuromas or ganglioneuromas involving the skin, oral mucosa, eyes, respiratory tract, and gastrointestinal tract, and a marfanoid habitus, with long axial skeletal features and hyperextensible joints (see Table 24.11). Previously, families with isolated medullary thyroid carcinomas and germline *RET* mutations without additional features of MEN-2B were classified as a separate entity ("familial medullary thyroid carcinoma"), but have recently been grouped under the umbrella of MEN-2B. A germline missense mutation leading to a single amino acid change in RET, distinct from the mutations that are seen in MEN-2A, seems to be responsible for virtually all cases of MEN-2B. This point substitution affects a critical region of the tyrosine kinase domain of the protein and leads to constitutive activation of RET in the absence of ligand.

  In contrast to MEN-1, in which the long-term benefit of early diagnosis by genetic screening is not well established, diagnosis via screening of at-risk family members in MEN-2A kindred is important because medullary thyroid carcinoma is a life-threatening disease that can be prevented by early thyroidectomy. According to updated recommendations issued in 2015, all patients with medullary thyroid carcinoma are advised to undergo germline *RET* mutation testing, since even "apparently sporadic" cases might harbor a germline mutation, which has implications for genetic counseling and testing for unaffected family members. Routine genetic testing identifies *RET* mutation carriers earlier and more reliably in MEN-2 kindreds; all individuals carrying germline *RET* mutations are advised to undergo prophylactic thyroidectomy to prevent the inevitable development of medullary carcinomas.
- *MEN-4* has emerged as a new entity in the past decade. These patients have clinical features that phenocopy MEN-1 patients, but in contrast to that syndrome, harbor germline *CDKN1B* mutations, leading to reduced levels of the cell-cycle checkpoint protein, p27.

Pineal Gland

The rarity of clinically significant lesions (virtually only tumors) justifies brevity in the consideration of the pineal gland. It is a minute, pinecone-shaped organ (hence its name), weighing 100 to 180 mg and lying between the superior colliculi at the base of the brain. It is composed of a loose, neuroglial stroma enclosing nests of epithelial-appearing *pineocytes,* cells with photosensory and neuroendocrine functions (hence the designation of the pineal gland as the "third eye"). Silver impregnation stains reveal that these cells have long, slender processes reminiscent of primitive neuronal precursors intermixed with the processes of astrocytic cells. The principal secretory product of the pineal gland is melatonin, which is involved in the control of circadian rhythms, including the sleep-wake cycle; hence the popular use of melatonin for the treatment of jet lag.

All tumors involving the pineal gland are rare; most (50% to 70%) arise from sequestered embryonic germ cells (Chapter 28). They most commonly take the form of so-called *germinomas,* resembling testicular seminoma (Chapter 21) or ovarian dysgerminoma (Chapter 22). Other lines of germ cell differentiation include embryonal carcinomas; choriocarcinomas; mixtures of germinoma, embryonal carcinoma, and choriocarcinoma; and, uncommonly, typical teratomas (usually benign). Whether to characterize these germ cell neoplasms as pinealomas is debated, but most "pinealophiles" favor restricting the term *pinealoma* to neoplasms arising from the pineocytes.

PINEALOMA

These neoplasms are divided into two categories, *pineoblastomas* and *pineocytomas*, based on their level of differentiation, which, in turn, correlates with their aggressiveness. These tumors are rare and are described in specialized texts.

SUGGESTED READINGS

Pituitary

Asa SL, Mete O: What's new in pituitary pathology? *Histopathology* 72:133–141, 2018. *[A review on the updated WHO classification of pituitary tumors and other non-neoplastic entities].*

deKock L, Sabbaghian N, Plourde F et al: Pituitary blastoma: a pathognomonic feature of germline DICER1 mutations, *Acta Neuropathol* 128:111–122, 2014. *[DICER1 is a critical node in microRNA processing, and this study identifies one of the neoplastic manifestations of germline* DICER1 *mutations that arise in the pituitary].*

Lopes MBS: The 2017 World Health Organization classification of tumors of the pituitary gland: a summary, *Acta Neuropathol* 134:521–535, 2017. *[The WHO classification of pituitary tumors was updated in 2017, and while a full description of the changes is beyond the scope of this book, this review provides the reader with a quick overview].*

Schertnhaner-Reiter MH, Stratakis CA, Luger A: Genetics of diabetes insipidus, *Endocrinol Metab Clin North Am* 46:305–334, 2017. *[An updated review on the molecular mechanisms underlying diabetes insipidus].*

Thyroid

Acquaviva G, Visani M, Repaci A et al: Molecular pathology of thyroid tumors of follicular cells: a review of genetic alterations and their clinic-pathologic relevance, *Histopathology* 72:6–31, 2018. *[A review on genetics of thyroid neoplasms on how the landscape of thyroid neoplasms has changed over time, with excellent summary figures on the multistep pathogenesis of invasive thyroid cancers from precursor lesions].*

Cabanillas ME, McFadden DG, Durante C: Thyroid cancer, *Lancet* 388:2783–2795, 2016. *[A clinically oriented review on diagnosis and management of thyroid neoplasms].*

DeLeo S, Lee SY, Braverman LE: Hyperthyroidism, *Lancet* 388:906–918, 2016. *[A clinically oriented review on the causes and main therapeutic avenues in hyperthyroidism].*

Ganly I, Makarov V, Deraje S et al: Integrated genomic analysis of Hurthle cell cancer reveals oncogenic drivers, recurrent mitochondrial mutations, and unique chromosomal landscape, *Cancer Cell* 34:256–270, 2018. *[The molecular profiles of Hurthle cell carcinomas, including unique genetic features that underlie their distinctive morphology].*

Graves RJ: A newly observed affection of the thyroid gland in females, *London Med Surg J* 7:516–517, 1835. *[A classic publication describing the index case of the disease named after the author].*

Gull WW: On a cretinoid state supervening in adult life in women, *Trans Clin Soc Lond* 7:180–185, 1873. *[The original description of myxedema, hence also known as "Gull disease"].*

Hashimoto H: Zur Kenntnis der lymphomatösen Veränderung der Schilddrüse (Struma lymphomatosa), *Arch Klin* 97:1912. *[Dr. Hashimoto, a Japanese physician, also trained in pathology in Germany, which might explain why his paper describing the eponymously named condition in four patients was originally published in German].*

Landa I, Ibrahimpasic T, Boucai L et al: Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers, *J Clin Invest* 126:1052–1066, 2016. *[The molecular characterization of these highly aggressive thyroid cancers confirms that most arise via de-differentiation from well-differentiated, more indolent neoplasms].*

Nikiforov YE, Seethala RR, Tallini G et al: Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors, *JAMA Oncol* 2:1023–1029, 2016. *[A seminal paper that introduced the term "noninvasive follicular thyroid neoplasm with papillary-like nuclear features" in light of the overdiagnosis and overtreatment of this subtype of essentially benign papillary "cancers"].*

Smith TJ, Hegedus L: Graves' disease, *N Engl J Med* 375:1552–1565, 2016. *[A comprehensive review on the pathophysiology, clinical manifestations, and treatment strategies in Graves disease. Your one-stop shop for this topic].*

Parathyroid

Machado NN, Wilhelm SM: Diagnosis and evaluation of primary hyperparathyroidism, *Surg Clin North Am* 99:649–666, 2019. *[A clinically oriented and up-to-date review on the causes, symptoms, laboratory findings, and treatment of hyperthyroidism].*

Mannstadt M, Bilezikian JP, Thakker RV et al: Hypoparathyroidism, *Nat Rev Dis Primers* 3:17055, 2017. *[Similar to other reviews in this "Primer" series, this is a well-illustrated overview that helps supplement the text we have covered in this chapter with additional details].*

Simonds WF: Genetics of hyperparathyroidism, including parathyroid cancer, *Endocrinol Metab Clin North Am* 46:405–418, 2017. *[A comprehensive review on the underlying genetic abnormalities contributing to hyperthyroidism, including neoplasms].*

Pancreas – Diabetes

Czech MP: Insulin action and resistance in obesity and type 2 diabetes, *Nat Med* 23:804–814, 2017. *[A comprehensive review on the pathophysiology of obesity and insulin resistance in type 2 diabetes with some wonderful summary figures on the interplay between the pancreas, liver, adipose tissue, and muscle].*

Defronzo RA, Ferrannini E, Groop L et al: Type 2 diabetes mellitus, *Nat Rev Dis Primers* 1:15019, 2015. *[Although mildly dated, this illustrated review is a great overview of type 2 diabetes that supplements many of the aspects we have covered in this chapter].*

Drucker DJ, Habener JF, Holst JJ: Discovery, characterization, and clinical development of glucagon-like peptides, *J Clin Invest* 127:4217–4227, 2017. *[GLP-1 receptor agonists have emerged as one of the most important therapeutic options in diabetes, and this review provides a historical timeline and mechanisms of action of GLP agonists in human diseases. This is worth reading to understand how complex and drawn-out the process of successful drug development can be!].*

Gloyn AL, Drucker DJ: Precision medicine in the management of type 2 diabetes, *Lancet Diabetes Endocrinol* 2018. S2213-8587(18)30052-4, *[A review on how genetic underpinnings of type 2 diabetes can be used to tailor its management in patients].*

Goodarzi MO: Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its associations, *Lancet Diabetes Endocrinol* 6:223–236, 2018. *[No other parameter is as paramount to the pathogenesis of type 2 diabetes as obesity and associated insulin resistance. This review discusses the genetic underpinnings of obesity and how this reflects on the obesity-diabetes axis].*

Katsarou A, Gudbjornsdottir S, Rawshani A et al: Type 1 diabetes mellitus, *Nat Rev Dis Primers* 3:17016, 2017. *[A wonderful overview of type 1 diabetes, including epidemiology, genetics, clinical symptoms, and treatment, complete with easy-to-absorb graphs and tables].*

Langenberg C, Lotta LA: Genomic insights into the causes of type 2 diabetes, *Lancet* 391:2463–2474, 2018. *[A recent review on the polygenic architecture of type 2 diabetes, including genetic features implicated in each of the underlying pathophysiologic mechanisms, such as beta-cell dysfunction, insulin resistance, and adiposity].*

Lee YS, Wollam J, Olefsky JM: An integrated view of immunometabolism, *Cell* 172:22–40, 2018. *[Many of the consequences of obesity, including insulin resistance and type 2 diabetes, have their bases in a state of chronic inflammatory dysfunction. This review ties together how inflammation eventually leads to metabolic deregulation].*

Pugliese A: Autoreactive T cells in type 1 diabetes, *J Clin Invest* 127:2881–2891, 2017. *[This review summarizes the evidence on autoreactive T cells in type 1 diabetes and the putative cellular targets of this autoimmune response in patients].*

Samuel VT, Shulman GI: The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux, *J Clin Invest* 126:12–22, 2016. *[There are many comprehensive reviews on insulin resistance in type 2 diabetes, given its paramount importance to the pathogenesis of hyperglycemia. A couple are included in the bibliography. Take your pick!]*.

Pancreas—Neuroendocrine Tumors

Inzani F, Petrone G, Rindi G: The new World Health Organization classification for pancreatic neuroendocrine neoplasia, *Endocrinol Metab Clin North Am* 47:463–470, 2018. *[The WHO has recently updated the classification of neuroendocrine neoplasms of the GI tract, including PanNETs, and this review summarizes the key distinctions from previous schema]*.

Pea A, Hruban RH, Wood LD: Genetics of pancreatic neuroendocrine tumors: implications for the clinic, *Expert Rev Gastroenterol Hepatol* 9:1407–1419, 2015. *[A relatively recent review on newly identified genomic aberrations in pancreatic neuroendocrine tumors and how these can form the basis for personalized therapy]*.

Adrenal Cortex

Addison T: *On the constitutional and local effects of disease of the supra-renal capsules*, London, 1855, Samuel Highley. *[The original description of Addison disease]*.

Charmandari E, Nicolaides NC, Chrousos GP: Adrenal insufficiency, *Lancet* 383:2152–2167, 2014. *[A comprehensive review on the epidemiology, pathogenesis, clinical features, diagnosis, and management of adrenal insufficiency]*.

Hernandez-Ramirez LC, Stratakis CA: Genetics of Cushing syndrome, *Endocrinol Metab Clin North Am* 47:275–297, 2018. *[An exhaustive review on the genetics of Cushing syndrome including somatic mutations underlying pituitary and adrenal neoplasms]*.

Husebye ES, Anderson MS, Kampe O: Autoimmune polyendocrine syndromes, *N Engl J Med* 378:1132–1141, 2018. *[A review on the genetics, pathophysiology, and clinical manifestations of autoimmune polyendocrine syndromes]*.

Loriaux DL: Diagnosis and differential diagnosis of Cushing syndrome, *N Engl J Med* 376:1451–1459, 2017. *[A fairly recent clinically oriented review on the etiopathogenesis, laboratory diagnosis, and treatment of Cushing syndrome. This review is a great place to get started]*.

Mackay I, Aspinall S: Advances in our understanding of the prognosis of adrenal incidentaloma, *Expert Rev Endocrinol Metab* 11:529–541, 2016. *[A review that addresses the prevalence, etiology, diagnostic workup, and treatment of incidentally discovered adrenal nodules]*.

Monticone S, Buffolo F, Tetti M et al: The expanding genetic horizon of primary hyperaldosteronism, *Eur J Endocrinol* 178:101–111, 2018. *[A review on the genetic factors underlying primary hyperaldosteronism, including sporadic and familial disease]*.

Rushworth RL, Torpy DJ, Falhammar H: Adrenal crisis, *N Engl J Med* 381:852–861, 2019. *[A review on the causes, pathophysiology, and treatment of acute adrenal crises]*.

Adrenal Medulla

Friderichsen C: Nebennierenapoplexie bei kleinen Kindern, *Jahrbuch für Kinderheilkunde* 87:109–125, 1918. *[One of the two original descriptions of the eponymously named Waterhouse-Friderichsen syndrome; Carl Friderichsen was a Danish pediatrician]*.

Lam AK: Update on adrenal tumors in 2017 World Health Organization (WHO) of endocrine tumors, *Endocr Pathol* 28:213–227, 2018. *[This commentary summarizes the major changes to the classification and genetics of adrenal neoplasms, including data from the recently completed integrated molecular analysis of pheochromocytoma and paraganglioma]*.

Waterhouse R: A case of suprarenal apoplexy, *Lancet* 1:577–578, 1911. *[The second of two papers, this one authored by Rupert Waterhouse, an English physician, that led to recognition of the eponymous syndrome]*.

Multiple Endocrine Neoplasia

Li JWY, Hua X, Reidy-Lagunes D et al: MENIN loss as a tissue specific driver of tumorigenesis, *Mol Cell Endocrinol* 469:98–106, 2018. *[A recent review on the cellular functions of menin and the consequences of its abrogation leading to neuroendocrine tumorigenesis. Potential therapeutic approaches are also discussed]*.

Pacheco MC: Multiple endocrine neoplasia: a genetically diverse group of familial tumor syndromes, *J Pediatr Genet* 5:89–97, 2016. *[A review on the subtypes, genetics, and pathology of MEN syndromes]*.

CHAPTER 25

The Skin

Alexander J. Lazar

CHAPTER CONTENTS

THE SKIN: MORE THAN A MECHANICAL BARRIER

More than a century and a half ago, the noted pathologist Rudolph Virchow described the skin as a mere protective covering for more delicate and functionally sophisticated internal viscera. Subsequently, the skin was viewed as a necessary but rather uninteresting barrier to fluid loss and mechanical injury. Over the last several decades, however, the skin has come to be appreciated as a surprisingly complicated organ—the largest in the body—in which precisely regulated cellular and molecular interactions govern many essential processes.

Although the human integument may appear drab compared with the skin and pelage of other members of the animal kingdom, it is extraordinarily vibrant with regard to the diversity of functions that it carries out. Chief among these is its role as one of the first lines of defense against potentially harmful infectious and physical agents. However, the skin is also a highly sophisticated sensory organ and even has important endocrine roles, particularly the synthesis of vitamin D (Chapter 9), which is "powered" by sun exposure. It is composed of several cell types and structures that function interdependently and cooperatively (Fig. 25.1).

- *Squamous epithelial cells (keratinocytes)* are normally "glued" tightly together by cell junctions known as desmosomes and produce abundant amounts of keratin protein, both of which serve to create a tough, durable physical barrier. In addition, keratinocytes secrete soluble molecules such as cytokines and defensins that augment and regulate cutaneous immune responses (described later).
- *Melanocytes* within the epidermis are responsible for the production of melanin, a brown pigment that absorbs and protects against potentially injurious ultraviolet (UV) radiation in sunlight.
- *Dendritic cells.* Skin serves as one of the first lines of defense against microorganisms and is constantly exposed to microbial and nonmicrobial antigens, which are processed by intraepidermal dendritic cells known as *Langerhans*

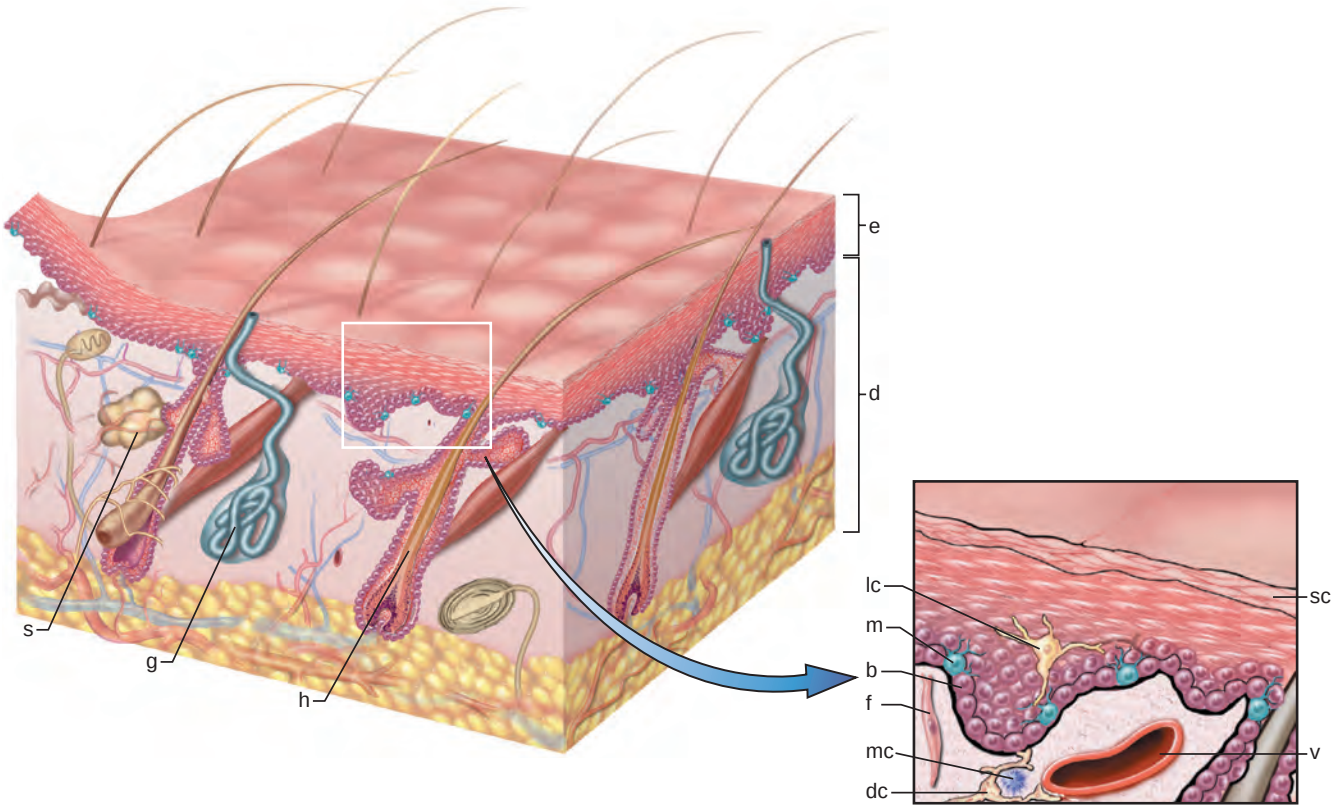


Figure 25.1 *Left*, The skin is composed of an epidermal layer *(e)*, from which specialized adnexa—hair follicles *(h)*, sweat glands *(g)*, and sebaceous glands *(s)*—descend into the underlying dermis *(d)*. *Inset*, This projection of the epidermal layer and underlying superficial dermis demonstrates the progressive upward maturation of basal cells *(b)* into cornified squamous epithelial cells of the stratum corneum *(sc)*. Melanin-containing dendritic melanocytes *(m)* (and rare Merkel cells containing neurosecretory granules) and mid-epidermal dendritic Langerhans cells *(lc)* are also present. The underlying dermis contains small vessels *(v)*, fibroblasts *(f)*, perivascular mast cells *(mc)*, and dendrocytes *(dc)*, which participate in dermal immune responses and repair.

cells. Langerhans cells secrete factors that augment innate immune responses and migrate from the skin to regional lymph nodes where they present their antigenic cargoes to T lymphocytes, thereby stimulating the adaptive immune system. Specialized *dendrocytes,* another type of dendritic cell found within the dermis, perform similar functions.

- *Lymphocytes.* Following their stimulation by dendritic cells in regional lymph nodes, T cells expressing an adhesion molecule called cutaneous lymphocyte-associated antigen (CLA) and chemokine receptors such as CCR4 and CCR10 leave the lymph node and home back to the dermis, a process that is directed in part by chemokines secreted by activated keratinocytes. Cytokines produced by these T cells mediate the microscopic patterns and clinical expressions of cutaneous inflammatory and infectious diseases. In addition, small numbers of B cells are found in the dermis that can participate in humoral responses to antigens encountered in the skin.
- *The skin is a large and complex ecosystem that provides niches for a broad spectrum of organisms, including bacteria, fungi, viruses, and mites.* These organisms have evolved symbiotic relationships with their human hosts and appear to contribute to health in a number of ways. By occupying skin niches, the normal "zoo" of organisms (the skin microbiome) prevents colonization of the skin by other potentially harmful organisms, just as the gut microbiome may protect against gastrointestinal pathogens (Chapter 17). In addition, the skin fauna primes and "educates" the cutaneous immune system in a manner that is believed to enhance immune responses to potential pathogens.
- *Afferent nerve fibers and a diverse set of associated specialized structures referred to as neural end organs* are responsible for physical sensations that run that gamut from pleasurable to painful, including touch, vibration, itchiness, cold, and heat. In addition, autonomic efferent nerve fibers regulate adnexal components such as sweat glands and effector pili muscles (see later) and can also influence the function of innate and adaptive immune cells in the dermis. Another cell type found in skin that remains cloaked in mystery is the *Merkel cell;* these cells are located in the epithelial basal cell layer and may have neuroendocrine or mechanoreceptor functions.
- *Adnexal components.* Sweat glands guard against deleterious variations in body temperature, and hair follicles, in addition to manufacturing hair shafts, have protected niches harboring epithelial stem cells capable of regenerating superficial epithelial skin structures following their disrupted by trauma, burns, and other types of injuries.

Perturbations that disrupt the delicate homeostasis that exists among skin cells may produce conditions as varied as wrinkles and hair loss, blisters and rashes, life-threatening cancers, and disorders of immune regulation. For example,

long-term exposure to sunlight fosters premature cutaneous aging, blunts immunologic responses to environmental antigens, and favors the development of a variety of premalignant and malignant cutaneous neoplasms. Ingested agents, such as therapeutic drugs, can cause an enormous number of rashes or exanthems. Systemic disorders, such as diabetes mellitus, amyloidosis, and lupus erythematosus, can also have important manifestations in the skin.

Skin conditions are very common, affecting about one-third of the US population each year. Since skin is uniquely accessible to visual examination, for the experienced observer it can yield numerous insights into the functional state of the body (if not the very soul of a patient). Close attention to the appearance and distribution of skin lesions is critical, as these characteristics are essential in formulating diagnoses and in understanding pathogenesis. To underscore this point, special emphasis is placed on the gross appearance of skin lesions under each specific entity.

Thousands of diseases affect the skin. Only those that are common or that illustrate important pathologic mechanisms are described here. Dermatologists and dermatopathologists have developed a set of terms to describe the gross and microscopic appearance of skin lesions that every student must be familiar with in order to be fluent in dermatopathology; the most important of these terms and their respective definitions are given in Table 25.1.

Table 25.1 Nomenclature of Skin Lesions

	Definition
Macroscopic Lesions	
Excoriation	Traumatic lesion breaking the epidermis and causing a raw linear defect (i.e., deep scratch); often self-induced.
Lichenification	Thickened, rough skin (similar to lichen on a rock); usually the result of repeated rubbing.
Macule, Patch	Circumscribed, flat lesion distinguished from surrounding skin by color. Macules are 5 mm in diameter or less; patches are greater than 5 mm.
Onycholysis	Separation of nail plate from nail bed.
Papule, Nodule	Elevated dome-shaped or flat-topped lesion. Papules are 5 mm or less across; nodules are greater than 5 mm in size.
Plaque	Elevated flat-topped lesion, usually greater than 5 mm across (may be caused by coalescent papules).
Pustule	Discrete, pus-filled, raised lesion.
Scale	Dry, horny, plate-like excrescence; usually the result of imperfect cornification.
Vesicle, Bulla, Blister	Fluid-filled raised lesion 5 mm or less across (vesicle) or greater than 5 mm across (bulla). Blister is the common term for either lesion.
Wheal	Itchy, transient, elevated lesion with variable blanching and erythema formed as the result of dermal edema.
Microscopic Lesions	
Acanthosis	Diffuse epidermal hyperplasia.
Dyskeratosis	Abnormal, premature keratinization within cells below the stratum granulosum.
Erosion	Discontinuity of the skin showing incomplete loss of the epidermis.
Exocytosis	Infiltration of the epidermis by inflammatory cells.
Hydropic swelling (ballooning)	Intracellular edema of keratinocytes, often seen in viral infections.
Hypergranulosis	Hyperplasia of the stratum granulosum, often due to intense rubbing.
Hyperkeratosis	Thickening of the stratum corneum, often associated with a qualitative abnormality of the keratin.
Lentiginous	Linear pattern of melanocyte proliferation within the epidermal basal cell layer.
Papillomatosis	Surface elevation caused by hyperplasia and enlargement of contiguous dermal papillae.
Parakeratosis	Keratinization with retained nuclei in the stratum corneum. On mucous membranes, parakeratosis is normal.
Spongiosis	Intercellular edema of the epidermis.
Ulceration	Discontinuity of the skin marked by complete loss of the epidermis revealing dermis or subcutis.
Vacuolization	Formation of vacuoles within or adjacent to cells; often refers to basal cell–basement membrane zone area.

DISORDERS OF PIGMENTATION AND MELANOCYTES

Focal or widespread loss of normal protective pigmentation can make individuals extraordinarily vulnerable to the harmful effects of sunlight (as in albinism). Changes in preexisting skin pigmentation may signify important primary skin disorders (e.g., malignant transformation of a mole) or point to the existence of an underlying systemic disorder (e.g., Addison disease; see Chapter 24).

Freckle (Ephelis)

Freckles are the most common pigmented lesions of childhood in lightly pigmented individuals. It is unclear whether freckles result from a focal abnormality in pigment production in a discrete field of melanocytes, enhanced melanin transfer to adjacent basal keratinocytes, or some combination thereof. The *café au lait* spots seen in neurofibromatosis type 1 (Chapter 27) are similar to freckles histologically, but differ in that they are larger, arise independently of sun exposure, and contain aggregated melanosomes (macromelanosomes), which can be seen within the cytoplasm of melanocytes in electron micrographs.

MORPHOLOGY

Freckles are generally small (1 mm to several millimeters in diameter), tan-red or light brown macules that appear after sun exposure. Hyperpigmentation of freckles results from increased amounts of melanin pigment within basal keratinocytes. Associated melanocytes may be slightly enlarged but are normal in density.

Once present, freckles fade and darken in a cyclic fashion during winter and summer, respectively. This is not because of changes in the number of melanocytes, but in the degree of pigmentation.

Lentigo

The term *lentigo* (plural, lentigines) refers to a common benign localized hyperplasia of melanocytes. It occurs at all ages, but most commonly appears in infants and children. There is no sex or racial predilection, and the cause and pathogenesis are unknown.

MORPHOLOGY

Lentigines may involve mucous membranes as well as the skin and consist of small (5 to 10 mm across), oval, tan-brown macules or patches. The essential histologic feature is **linear (nonnested) melanocytic hyperplasia** restricted to the cell layer immediately above the basement membrane that produces a hyperpigmented basal cell layer. Unlike freckles, lentigines do not darken when exposed to sunlight.

Melanocytic Nevus (Pigmented Nevus, Mole)

Melanocytic nevi (known colloquially as moles) are common benign neoplasms caused in most cases by acquired activating mutations in components of the RAS signaling pathway. Most of us have at least a few "moles" and probably regard them as mundane and uninteresting. However, in truth, moles (or melanocytic nevi) are diverse, dynamic, and biologically fascinating neoplasms. There are numerous subtypes of melanocytic nevi that are distinguished based on their clinical and histologic features; Table 25.2 provides a summary of salient features of some commonly encountered forms. Acquired melanocytic nevi are the most common type and are found in virtually all individuals.

Pathogenesis

Proof that nevi are neoplasms comes from studies showing that many have acquired mutations that lead to constitutive activation of RAS or the serine/threonine kinase BRAF, which lies immediately downstream of RAS (described in Chapter 7 and later under Melanoma). Given that RAS signals have potent transforming activity and have key roles in many full-blown cancers, it is reasonable to ask why nevi only rarely give rise to melanomas. One answer appears to lie in the phenomenon referred to as oncogene-induced senescence. Expression of either activated RAS or BRAF in normal human melanocytes causes only a limited period of proliferation that is followed by a permanent growth arrest mediated by the accumulation of p16/INK4a, a potent inhibitor of several cyclin-dependent kinases, including CDK4 and CDK6 (Chapter 7). This protective response is disrupted in melanoma and some precursor lesions that give rise to melanoma.

MORPHOLOGY

Common acquired melanocytic nevi are tan to brown, uniformly pigmented, small (usually less than 6 mm across), relatively flat macules or elevated papules with well-defined, rounded borders (Figs. 25.2A and 25.3A). They may become more prominent during pregnancy, indicating a degree of hormone sensitivity. Melanocytic nevi are thought to progress through a series of morphologic changes over time. The earliest lesions are believed to be **junctional nevi,** which consist of aggregates or nests of round cells that grow along the dermoepidermal junction (Fig. 25.2B). Nuclei of nevus cells are uniform and rounded in contour, contain inconspicuous nucleoli, and show little or no mitotic activity. Eventually, most junctional nevi grow into the underlying dermis as nests or cords of cells to form **compound nevi** (Fig. 25.3B). In older lesions the epidermal nests may be lost entirely to form pure **intradermal nevi.** Compound and dermal nevi are often more elevated than junctional nevi.

Growth of nevus cells from the dermoepidermal junction into the underlying dermis is accompanied by morphologic changes that are taken to reflect oncogene-induced senescence (Fig. 25.4). Whereas superficial nevus cells are larger, tend to produce melanin, and grow in nests, deeper nevus cells are smaller, produce little or no pigment, and appear as cords and single cells. At the deepest extent of the lesions, these cells often acquire fusiform contours and grow in fascicles resembling neural tissue (**neurotization;** see Fig. 25.4E). This striking metamorphosis correlates with enzymatic changes (progressive loss of tyrosinase activity and acquisition of cholinesterase activity) in deeper, nonpigmented, "nerve-like" nevus cells. These changes are helpful in distinguishing benign nevi from melanomas, which lack such features.

Table 25.2 Representative Variant Forms of Melanocytic Nevi

Nevus Variant	Diagnostic Architectural Features	Cytologic Features	Clinical Significance
Congenital nevus	Deep dermal and sometimes subcutaneous growth around adnexa, neurovascular bundles, and blood vessel walls	Identical to ordinary acquired nevi	Present at birth; large variants have increased melanoma risk
Blue nevus	Non-nested dermal infiltration, often with associated fibrosis	Highly dendritic, heavily pigmented nevus cells	Black-blue nodule; often confused with melanoma clinically
Spindle and epithelioid cell nevus (Spitz nevus)	Fascicular growth	Large, plump cells with pink-blue cytoplasm; fusiform cells	Common in children; red-pink nodule; often confused with hemangioma clinically
Halo nevus	Lymphocytic infiltration surrounding nevus cells	Identical to ordinary acquired nevi	Host immune response against nevus cells and surrounding normal melanocytes
Dysplastic nevus	Coalescent intraepidermal nests	Cytologic atypia	Potential marker or precursor of melanoma

Although melanocytic nevi are common, their clinical and histologic diversity necessitates thorough knowledge of their appearance and natural history, lest they be confused with other skin conditions, most notably melanoma. The biologic importance of some nevi, however, resides in their possible transformation to melanoma or as markers of increased risk for melanoma (described next).

Dysplastic Nevi

Dysplastic nevi are important because they can be direct precursors of melanoma and when multiple in number are a marker of an increased risk for melanoma. The association of melanocytic nevi with melanoma was made almost 200 years ago, but a potential precursor of melanoma was not identified until 1978 when Clark and colleagues described lesions now referred to as dysplastic nevi. Several lines of evidence support the concept that some dysplastic nevi are precursors of melanoma. One of the most compelling involves studies of families affected by *dysplastic nevus syndrome* (also known as familial atypical mole and melanoma syndrome), an autosomal dominant disorder in which a tendency to develop multiple dysplastic nevi and melanoma are co-inherited. The probability that a person with dysplastic nevus syndrome will develop melanoma is over 50% by

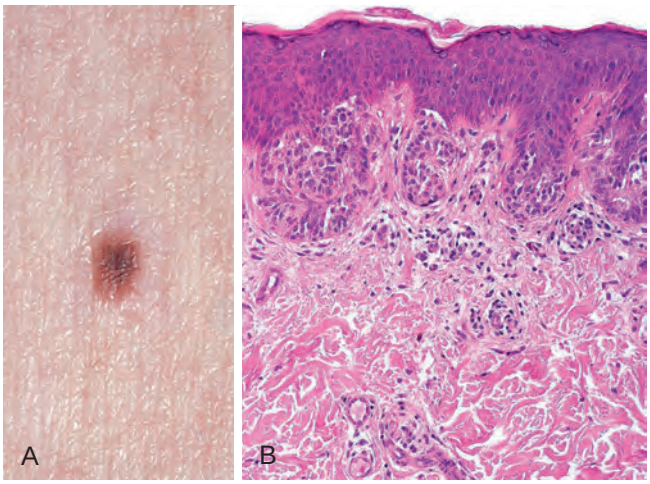


Figure 25.2 Melanocytic nevus, junctional type. (A) Grossly, lesions are small, relatively flat, symmetric, and uniform. (B) On histologic examination, junctional nevi are characterized by rounded nests of nevus cells originating at the tips of rete ridges along the dermoepidermal junction.

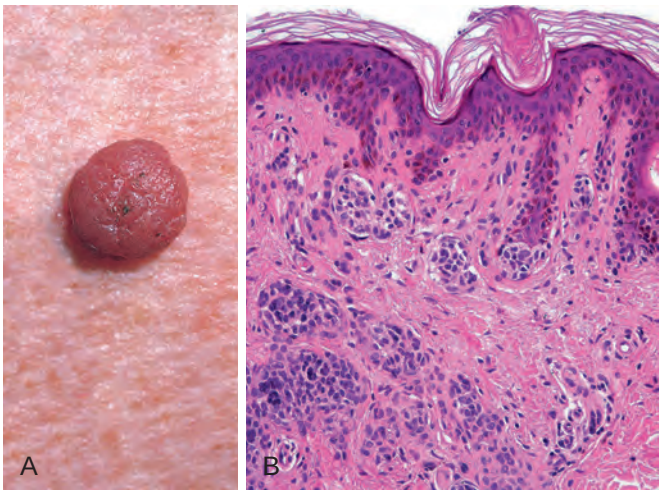


Figure 25.3 Melanocytic nevus, compound type. (A) In contrast to the junctional nevus, the compound nevus is raised and dome-shaped. The symmetry and uniform pigment distribution suggest a benign process. (B) Histologically, compound nevi combine the features of junctional nevi (intraepidermal nevus cell nests) with nests and cords of dermal nevus cells.

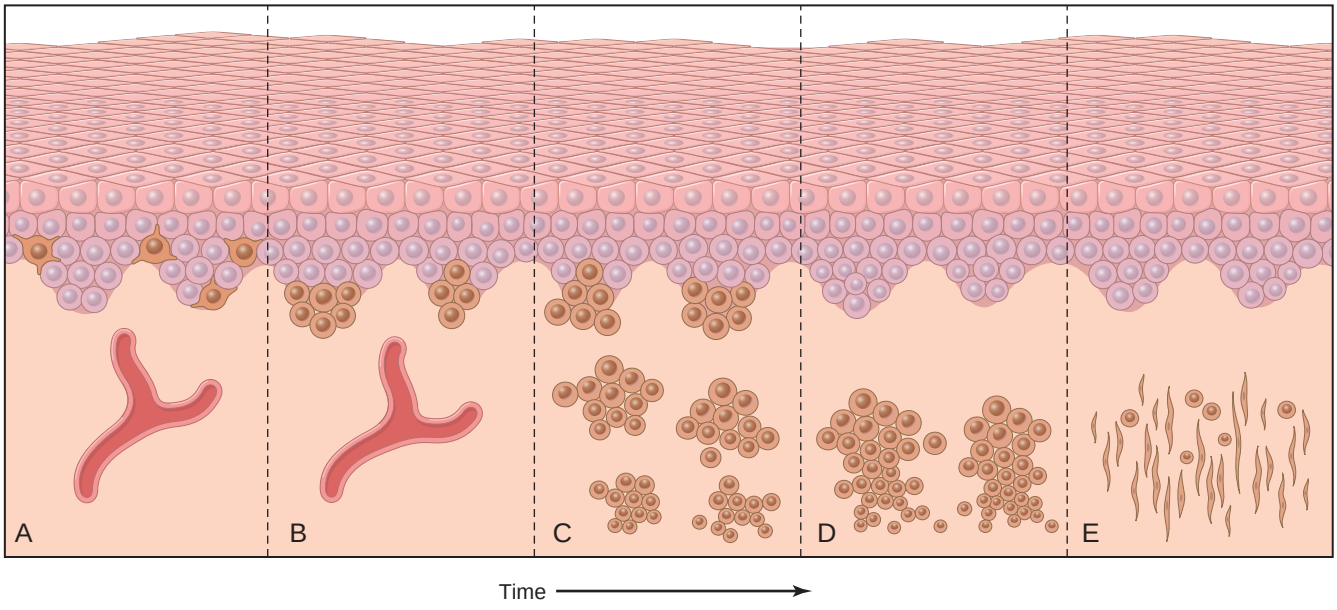


Figure 25.4 Maturation sequence of nondysplastic melanocytic nevi. (A) Normal skin shows only scattered dendritic melanocytes within the epidermal basal cell layer. (B) Junctional nevus. (C) Compound nevus. (D) Dermal nevus. (E) Dermal nevus with neurotization, a change that is also referred to as maturation. Nevi may exist at any stage in this sequence for variable periods of time, although many are believed to progress through this sequence.

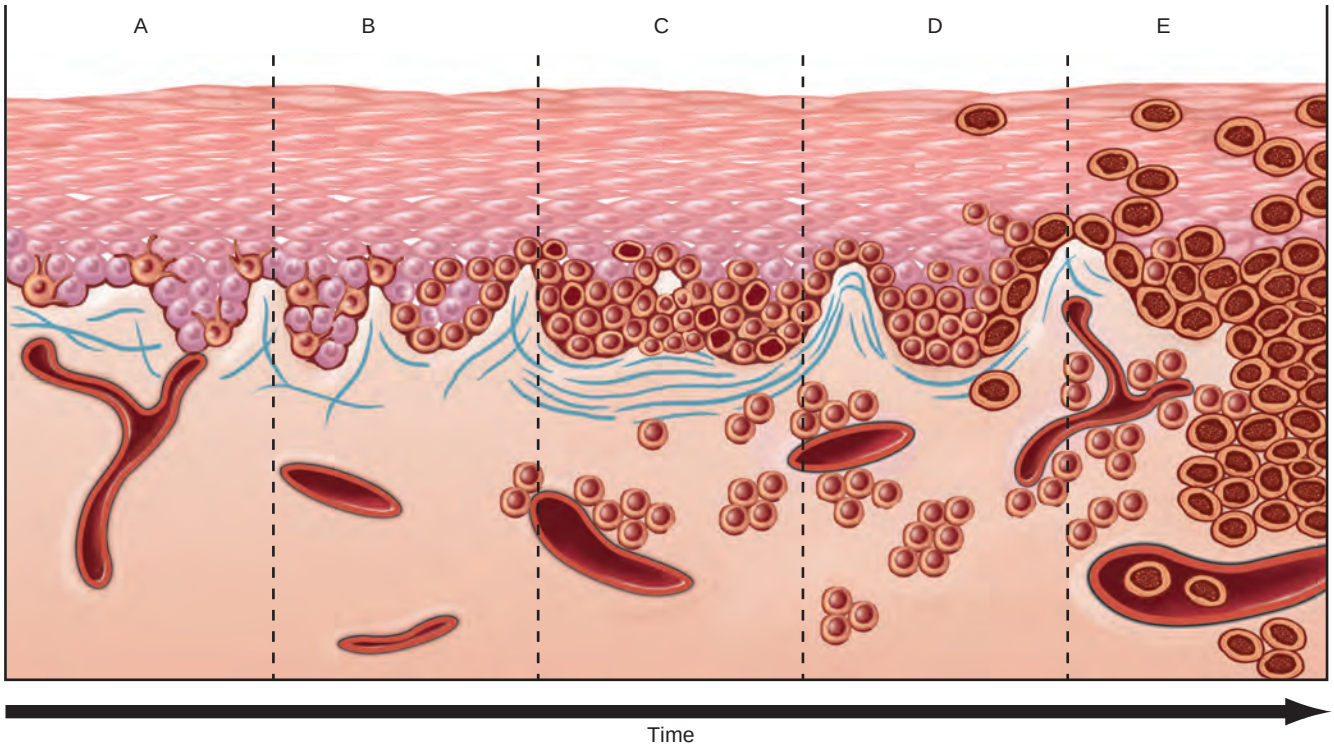


Figure 25.5 Potential steps of tumor progression in dysplastic nevi. (A) Lentiginous melanocytic hyperplasia. (B) Lentiginous junctional nevus. (C) Lentiginous compound nevus with abnormal architectural and cytologic features (dysplastic nevus). (D) Early melanoma, or melanoma in radial growth phase (large dark cells in epidermis). (E) Advanced melanoma (vertical growth phase) with malignant spread into the dermis and vessels. The risk of malignant transformation of any single dysplastic nevus is small, but appears to be higher than that of typical nevi.

age 60, and at-risk individuals sometimes develop melanomas at multiple sites. Even more directly, apparent transformation of dysplastic nevi to melanoma has been documented histologically.

Although dysplastic nevi can give rise to melanoma, the vast majority of such lesions are clinically stable and never progress. Conversely, not all melanomas in individuals with dysplastic nevus syndrome arise from dysplastic nevi, suggesting that these lesions are best viewed as indicators of increased melanoma risk. Dysplastic nevi may also occur as isolated lesions in otherwise normal individuals, in which case the risk of malignant transformation is very low.

Pathogenesis

Clark and associates first proposed morphologic stages in the development of dysplastic nevi and their eventual progression to melanoma (Fig. 25.5), and other workers have subsequently correlated these morphologic changes with the stepwise acquisition of oncogenic mutations and epigenetic changes. Like conventional nevi, dysplastic nevi also frequently have acquired activating mutations in the *NRAS* and *BRAF* genes. What then distinguishes dysplastic nevi from typical melanocytic nevi? An important clue comes from individuals with dysplastic nevus syndrome. Such individuals often have inherited loss-of-function mutations in *CDKN2A*. *CDKN2A* encodes several proteins including p16 (described in more detail under Melanoma), which you will recall is a negative regulator of cyclin-dependent kinase 4 (CDK4) and CDK6. Other affected families have mutations in the *CDK4* gene that make the CDK4 protein resistant to inhibition by p16. Thus, it appears that RAS or BRAF activation and increased CDK4 activity contribute to the development of dysplastic nevi. However, not all patients with germline mutations in *CDKN2A* or *CDK4* have dysplastic nevi, and not all familial dysplastic nevi are associated with mutations in these genes. As a result it is suspected that additional genes influence whether dysplastic nevi occur in a particular individual; the identities of these modifier genes as well as the other genes that are responsible for the syndrome are being sought.

MORPHOLOGY

Dysplastic nevi are larger than most acquired nevi (often greater than 5 mm across) and in individuals with the dysplastic nevus syndrome may number in the hundreds (Fig. 25.6A). They may appear as flat macules, slightly raised plaques with a "pebbly" surface, or target-like lesions with a darker raised center and irregular flat periphery. They can be recognized based on their large size, variability in pigmentation (variegation), and irregular borders. Most seem to be acquired rather than congenital. Unlike ordinary moles, dysplastic nevi occur on both sun-exposed and protected body surfaces.

Microscopically, dysplastic nevi usually involve both the epidermis and the dermis and exhibit architectural and cytologic atypia (Fig. 25.6A, B). **Nevus cell nests within the epidermis may be enlarged and often fuse or coalesce with adjacent nests.** As part of this process, single nevus cells begin to replace the normal basal cell layer along the dermoepidermal junction, producing **lentiginous hyperplasia.** Cytologic atypia takes the form of nuclear enlargement; irregular, often angulated, nuclear

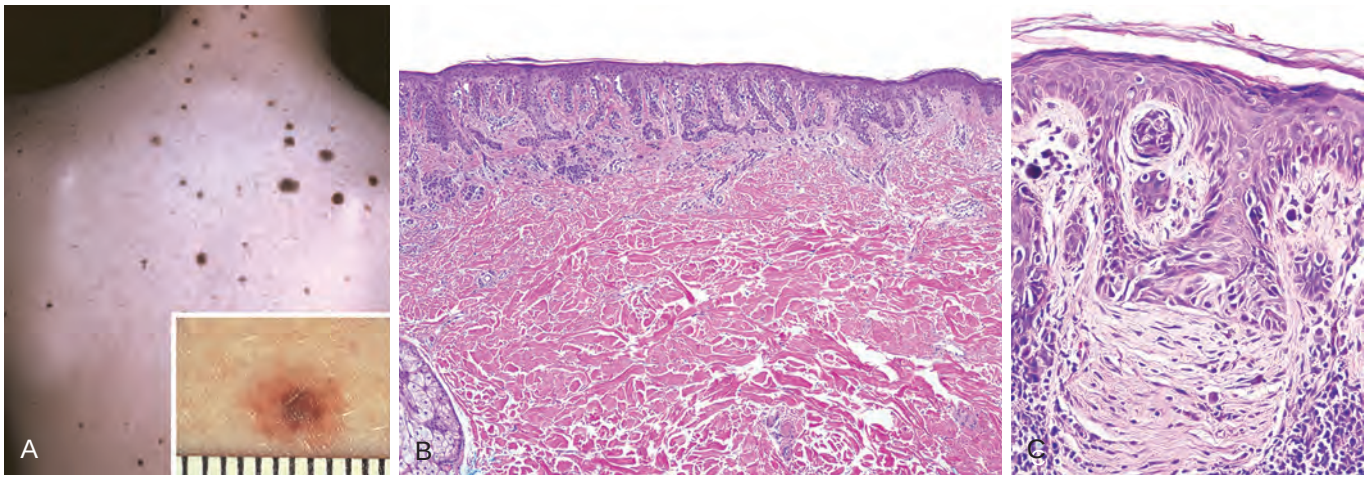


Figure 25.6 Dysplastic nevus. (A) Numerous atypical nevi on the back. (B) One such lesion (A, *inset*) has a compound nevus component *(left)* and an asymmetric junctional nevus component *(right)*. The former corresponds to the more pigmented and raised central zone, and the latter corresponds to the less pigmented, flat peripheral rim of the lesion shown in A. (C) An important feature is the presence of cytologic atypia (irregularly shaped, dark-staining nuclei). The dermis underlying the atypical cells characteristically shows linear, or lamellar, fibrosis.

contours; and hyperchromasia (Fig. 25.6C). Associated alterations in the superficial dermis include lymphocytic infiltrates (usually sparse); release of melanin from dead nevus cells into the dermis (melanin incontinence), where it is phagocytosed by dermal macrophages; and a peculiar **linear fibrosis** surrounding the epidermal rete ridges that are involved by the nevus. The diagnosis is based on this constellation of features, rather than any single finding.

Melanoma

Melanoma is the most deadly of all skin cancers and is strongly linked to acquired mutations caused by exposure to UV radiation in sunlight. Melanoma is a relatively common neoplasm that can be cured if it is detected and treated when it is in its earliest stages. The great preponderance of melanoma arises in the skin; other sites of origin include the oral and anogenital mucosal surfaces (i.e., oropharynx, gastrointestinal and genitourinary tracts), the esophagus, the meninges, and the uvea of the eye (Chapter 29). The following comments apply to cutaneous melanomas.

Today, as a result of increased public awareness of the signs of cutaneous melanoma, most are cured surgically. Nevertheless, the reported incidence of melanoma is increasing; more than 100,000 cases and more than 6800 deaths are expected in the United States in 2020. Notably, there has been a decrease in death rates over the past several years, which may reflect the effectiveness of immune checkpoint inhibitor therapy in this disease (Chapter 7).

Pathogenesis

In about 10% to 15% of affected patients, the risk of melanoma is inherited as an autosomal dominant trait with variable penetrance. As mentioned when discussing dysplastic nevi, some of these familial cases are associated with germline mutations affecting genes that regulate cell cycle progression, whereas others are associated with germline mutations affecting telomerase expression (described later). In the remaining patients, melanoma is sporadic and is strongly related to a single predisposing environmental factor: UV radiation exposure from sunlight. Because repair of UV radiation–induced DNA damage is imperfect, UV radiation leads to the accumulation of mutations in melanocytes over time. Sequencing of melanoma genomes has demonstrated a very high rate of point mutations that bear the signature of the damaging effects of UV radiation on DNA. In line with this molecular evidence, melanomas most commonly arise on sun-exposed surfaces, particularly the upper back in men and the back and legs in women, and lightly pigmented individuals are at higher risk than darkly pigmented individuals. Other inherited genetic variants linked to an increased risk of melanoma in fair-skinned populations act by diminishing melanin production in skin, thus increasing the amount of damage that sun exposure wreaks on melanocytes.

Nevertheless, the relationship between sun exposure and melanoma is not as straightforward as with other skin cancers, such as squamous cell carcinoma (discussed later). Some studies suggest that periodic severe sunburns early in life are the most important risk factor. Furthermore, since melanomas sometimes occur in dark-skinned individuals and at body sites that are not sun-exposed, sunlight is not always an essential predisposing factor, and other environmental factors may also contribute to risk.

The most frequent "driver" mutations in melanoma affect cell cycle control, pro-growth pathways, and telomerase. Some of the more common mutations are as follows:

- *Mutations that disrupt cell cycle control genes.* The *CDKN2A* gene is mutated in approximately 40% of pedigrees with autosomal dominant familial melanoma. *CDKN2A* is a complex locus that encodes three different tumor suppressors, p15, p16, and ARF. Of these, loss of p16 is clearly implicated in human melanoma, and experimental evidence also supports a role for loss of ARF. As already mentioned, p16 inhibits CDK4 and CDK6, thus reinforcing the ability of the RB tumor suppressor to block cells in the G_1 phase of the cell cycle. By contrast, ARF enhances the activity of the p53 tumor suppressor by inhibiting MDM2, an oncoprotein that stimulates p53 degradation.

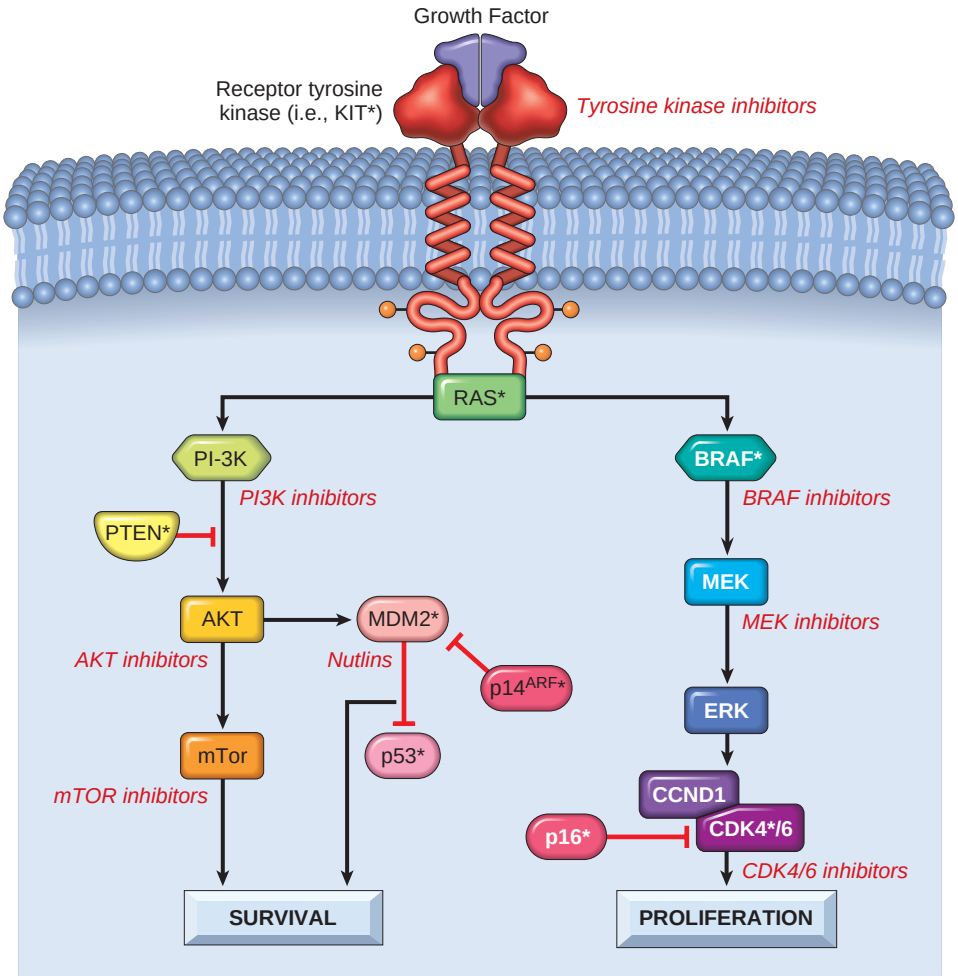


Figure 25.7 Pathways important in melanoma. Growth factors activate signaling circuits involving receptor tyrosine kinases (e.g., KIT), RAS, and two key downstream pathways that include the serine/threonine kinase BRAF and the phospholipid kinase PI3K. Proteins indicated by *asterisks* are mutated in melanoma. Components of these pathways that are being targeted by drugs are indicated.

CDKN2A is commonly mutated in sporadic melanomas, though other mechanisms controlling the G_1 checkpoint may also be involved. The net effect of all of these alterations is the same: increased melanocytic proliferation due to loss of cell cycle control and escape from oncogene-induced cellular senescence.

- *Mutations that activate pro-growth signaling pathways.* A second common group of molecular lesions in sporadic melanoma leads to aberrant increases in RAS and PI3K/AKT signaling (Fig. 25.7), which promote cell growth and survival (Chapter 7). Activating mutations in BRAF, a serine/threonine kinase that is downstream of RAS, are seen in 40% to 50% of melanomas, while activating mutations in RAS occur in an additional 15% to 20% of tumors. Melanomas with BRAF mutations also often show loss of the *PTEN* tumor suppressor gene, leading to heightened activation of the PI3K/AKT pathway. For reasons that are unclear, melanomas arising in non–sun-exposed cutaneous sites rarely have mutations in BRAF or RAS and are more likely to have activating mutations in the receptor tyrosine kinase KIT, which sits upstream of both RAS and PI3K/AKT. *PTEN* is also silenced in 20% of melanomas arising in non–sun-exposed sites. Other melanomas have loss-of-function mutations in the tumor-suppressor gene encoding neurofibromin 1 (*NF1*), a negative regulator of RAS, yet another mechanism of unleashing RAS signaling.
- *Mutations that activate telomerase.* Reactivation of telomerase, the enzyme activity that preserves telomeres and protects cells from senescence, has long been known to be important in cancer (Chapter 7). Sequencing of sporadic melanomas has identified mutations in the promoter of *TERT*, the gene that encodes the catalytic subunit of telomerase, in roughly 70% of tumors, making *TERT* the most commonly mutated gene yet identified in this cancer. Rare melanoma-prone families have also been identified that have germline *TERT* promoter mutations. As might be anticipated, the mutations increase *TERT* expression, suggesting that they act as an antidote to senescence.

MORPHOLOGY

Unlike benign nevi, melanomas show striking **variations in color,** appearing in shades of black, brown, red, dark blue, and gray (Fig. 25.8A). On occasion, zones of white or flesh-colored hypopigmentation also appear, sometimes due to focal regression of the tumor. **The borders of melanomas are irregular and**

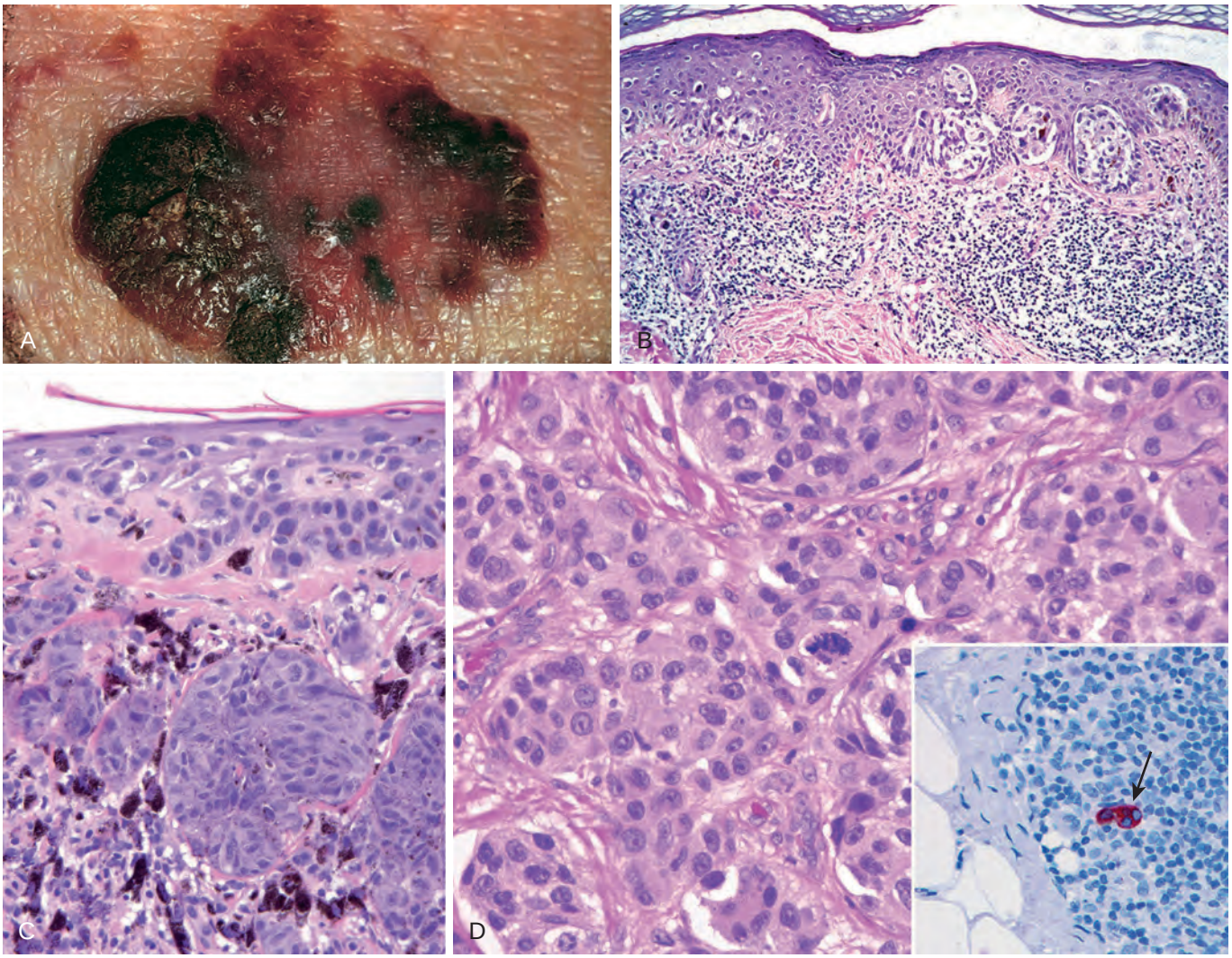


Figure 25.8 Melanoma. (A) Typical lesions are irregular in contour and pigmentation. Macular areas correlate with the radial growth phase, while raised areas correspond to nodular aggregates of malignant cells in vertical growth phase. (B) Radial growth phase showing irregular nested and single-cell growth of melanoma cells within the epidermis and an underlying inflammatory response within the dermis. (C) Vertical growth phase demonstrating nodular aggregates of infiltrating cells. (D) High-power view of melanoma cells. The *inset* shows a sentinel lymph node with a tiny cluster of melanoma cells *(arrow)* staining for the melanocytic marker HMB-45. Even small numbers of malignant cells in a draining lymph node may confer a worse prognosis.

often notched, unlike the smooth, round, and uniform borders of melanocytic nevi.

Central to understanding the progression of melanoma is the concept of radial and vertical growth phases. **Radial growth** describes the horizontal spread of melanoma within the epidermis and superficial dermis (Fig. 25.8B). During this initial stage the tumor cells seem to lack the capacity to metastasize. Tumors in radial growth phase fall into several clinicopathologic classes, including **lentigo maligna,** usually presenting as an indolent lesion on the face of older men that may remain in the radial growth phase for several decades; **superficial spreading,** the most common type of melanoma, usually involving sun-exposed skin; and **acral/mucosal lentiginous** melanoma, which is unrelated to sun exposure.

After a variable (and unpredictable) period of time, melanoma shifts from the radial phase to a **vertical growth phase,** during which the tumor cells invade downward into the deep dermis as an expansile mass (Fig. 25.8C). **The vertical growth phase is often heralded by the appearance of a nodule and correlates with the emergence of a tumor subclone with metastatic potential.** Unlike melanocytic nevi, maturation or "neurotization" is absent from the deep invasive portion of melanoma.

Individual melanoma cells are usually larger than normal melanocytes or cells found in melanocytic nevi. They have enlarged nuclei with irregular contours, chromatin that is characteristically clumped at the periphery of the nuclear membrane, and prominent red (eosinophilic) nucleoli (Fig. 25.8D). The appearance of the tumor cells is similar in the radial and vertical phases of growth. While most nevi and melanomas are easily distinguished based on their appearance, a small fraction of "atypical" lesions fall in a histologic gray zone and have been termed melanocytic tumors of uncertain malignant potential; such lesions require complete excision and close clinical follow-up.

Once a melanoma is excised, a number of pathologic features are used to gauge the probability of metastatic spread and prognosis. One model used to predict outcome is based in part on the following pathologic variables: (1) tumor depth of invasion

(Breslow thickness), (2) number of mitoses, (3) evidence of tumor regression (presumably due to the host immune response), (4) ulceration of overlying skin, (5) presence and number of tumor-infiltrating lymphocytes, and (6) location (central body or extremity). Determinants of a more favorable prognosis in this model include thinner tumor depth, no or few mitoses (<1/mm^2), a brisk tumor-infiltrating lymphocyte response, absence of regression, and lack of ulceration. Since most melanomas initially metastasize to regional lymph nodes, additional prognostic information may be obtained by performing a sentinel lymph node biopsy, as in breast cancer (Chapter 23). Microscopic involvement of a sentinel node by even a small number of melanoma cells (micrometastases, Fig. 25.8D, *inset*) confers a worse prognosis. The degree of involvement and the total number of lymph nodes involved correlate well with overall survival.

Clinical Features

The most important warning signs, sometimes called the ABCDEs of melanoma, are (1) asymmetry; (2) irregular borders; and (3) variegated color, (4) increasing diameter, and (5) evolution or change over time, especially if rapid. Because locally advanced melanomas often metastasize, early recognition and complete excision are critical. Melanoma of the skin is usually asymptomatic, although itching or pain may be early manifestations. The majority of lesions are greater than 10 mm in diameter at diagnosis. The most consistent clinical signs are changes in the color, size, or shape of a pigmented lesion.

Molecular insights into the pathogenesis of melanoma have led to success in treating this cancer with drugs that target the RAS and PI3K/AKT pathways (see Fig. 25.7). Such approaches are urgently needed, as metastatic melanoma is resistant to both conventional chemotherapy and radiation treatment. Ultimately, it is likely that these types of targeted therapies will be used in combinations tailored to fit the oncogenic molecular lesions found in individual tumors. This idea is based on the observation that a high fraction of tumors with BRAF mutations respond to BRAF inhibitors, whereas tumors belonging to other molecular subtypes do not.

Recognition that melanoma is inherently immunogenic has also resulted in successful development of therapeutic regimens using antibodies that inhibit immune checkpoint proteins, such as CTLA4 and PD1. Melanoma is one of the most responsive cancers to these checkpoint inhibitors, which take the brakes off an immune system that is poised to respond to tumor neoantigens (Chapter 7). Current efforts are focused at maximizing tumor response to checkpoint inhibitors without undue toxicity, which stems from autoimmune attack on normal host tissues.

KEY CONCEPTS

MELANOCYTIC LESIONS, BENIGN AND MALIGNANT

- Most melanocytic nevi have activating mutations in *BRAF*, or less often *RAS*, but the vast majority never undergo malignant transformation.
- Most sporadic dysplastic nevi are best regarded as markers of melanoma risk rather than premalignant lesions.
- Melanoma is a highly aggressive malignancy linked to sun exposure; risk of spread is predicted by several tumor characteristics, particularly the vertical thickness of excised tumors.
- Melanoma often incites a host immune response and often shows dramatic responses to immune checkpoint therapies that enhance T-cell immunity.

BENIGN EPITHELIAL TUMORS

Benign cutaneous epithelial neoplasms are common tumors that are derived from the keratinizing stratified squamous epithelium of the epidermis and hair follicles and the ductular epithelium of cutaneous glands. They often recapitulate the structures from which they arise. Their appearance sometimes raises a concern of malignancy, particularly when they are pigmented or inflamed, and biopsy is frequently required to establish a definitive diagnosis. In very rare instances they are a telltale sign of syndromes associated with potentially life-threatening visceral malignancies, such as multiple trichilemmomas in *Cowden syndrome* or multiple sebaceous neoplasms in *Muir-Torre syndrome.* Diagnosis of epithelial tumors in these instances can facilitate recognition of the underlying syndrome and implementation of appropriate clinical interventions.

Seborrheic Keratosis

These common epidermal tumors occur most frequently in middle-aged or older individuals. They arise spontaneously and are particularly numerous on the trunk, although the extremities, head, and neck may also be involved. In people of color, multiple small seborrheic keratoses on the face are termed *dermatosis papulosa nigra,* a condition that is present in up to 35% of African-American adults.

Pathogenesis

Activating mutations in fibroblast growth factor receptor-3 (FGFR3), a receptor tyrosine kinase, are found in many sporadic seborrheic keratoses and are thought to drive the growth of the tumor. Seborrheic keratoses may suddenly appear in large numbers as part of a paraneoplastic syndrome *(Leser-Trélat sign),* possibly due to stimulation of keratinocytes by transforming growth factor-α (TGF-α) produced by tumor cells, most commonly carcinomas of the gastrointestinal tract.

MORPHOLOGY

Seborrheic keratosis characteristically appears as round, flat, coin-like, waxy plaque that varies in diameter from millimeters to several centimeters (Fig. 25.9, *inset*). It is uniformly tan to dark brown and usually has a velvety to granular surface. Inspection with a hand lens typically reveals small, round, pore-like ostia impacted with keratin, a feature helpful in differentiating these pigmented lesions from melanoma.

On histologic examination, these neoplasms are exophytic and sharply demarcated from the adjacent epidermis. They are composed of sheets of small cells that closely resemble basal

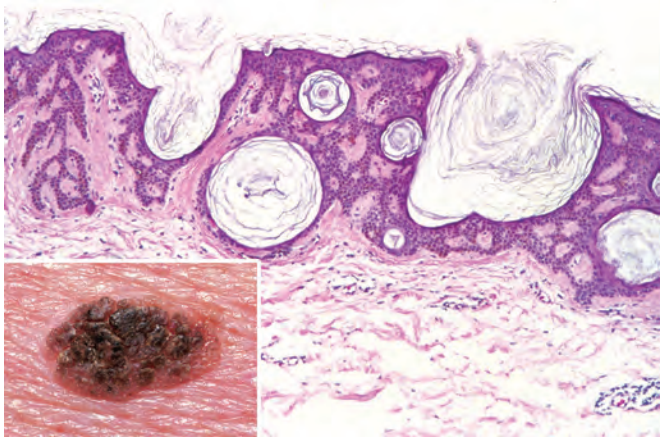


Figure 25.9 Seborrheic keratosis. A well-demarcated coin-like pigmented lesion containing dark keratin-filled surface plugs *(inset)* is composed of benign basaloid cells associated with prominent keratin-filled "horn" cysts, some of which communicate with the surface (pseudohorn cysts).

cells of the normal epidermis (Fig. 25.9). Variable melanin pigmentation is present within these basaloid cells, accounting for the brown coloration. Exuberant keratin production (hyperkeratosis) occurs at the surface, and small keratin-filled cysts (horn cysts) and invaginations of keratin into the main mass (invagination cysts) are characteristic features. If irritated and inflamed, seborrheic keratoses develop whirling foci of squamous differentiation resembling eddy currents in a stream.

Acanthosis Nigricans

Acanthosis nigricans can be an important cutaneous sign of several underlying benign and malignant conditions. It is a condition marked by thickened, hyperpigmented skin with a velvet-like texture that most commonly appears in the flexural areas (axillae, skin folds of the neck, groin, and anogenital regions). It is divided into two types based on the underlying condition.

- *In at least 80% of cases, acanthosis nigricans is associated with benign conditions and develops gradually,* usually during childhood or puberty. It may occur (1) as an autosomal dominant trait with variable penetrance, (2) in association with obesity or endocrine abnormalities (particularly with pituitary or pineal tumors, or with diabetes), and (3) as part of several rare congenital syndromes. The most common associations are with obesity and diabetes.
- *In the remaining cases, acanthosis nigricans arises in association with cancers,* most commonly gastrointestinal adenocarcinoma, usually in middle-aged and older individuals. In this setting, acanthosis nigricans is best viewed as a paraneoplastic phenomenon that is likely caused by growth factors released from tumors.

Pathogenesis

The unifying feature in all types of acanthosis nigricans is a disturbance that leads to increased growth factor receptor signaling in the skin. The familial form is associated with germline activating mutations in the receptor tyrosine kinase FGFR3, the same receptor that is frequently mutated in seborrheic keratosis. Depending on the mutation, acanthosis may be an isolated finding or be seen together with skeletal deformities, including achondroplasia and thanatophoric dysplasia. Why in some cases FGFR3 mutation gives rise to seborrheic keratosis and in others acanthosis nigricans is not clear. In those with type 2 diabetes, hyperinsulinemia is believed to provoke increased stimulation of insulin-like growth factor receptor-1 (IGFR1), another receptor tyrosine kinase that activates the same signaling pathways as FGFR3. Factors responsible for paraneoplastic acanthosis nigricans are uncertain; some cases have been linked to high levels of TGF-α, which may result in excessive activation of epidermal growth factor receptor (EGFR), yet another receptor tyrosine kinase, in the skin.

MORPHOLOGY

All forms of acanthosis nigricans have similar histologic features. The epidermis and underlying enlarged dermal papillae undulate sharply to form numerous repeating peaks and valleys. Variable hyperplasia may be seen, along with hyperkeratosis and slight basal cell layer hyperpigmentation (but no melanocytic hyperplasia).

Fibroepithelial Polyp

The fibroepithelial polyp has many names (acrochordon, squamous papilloma, skin tag) and is one of the most common cutaneous lesions. It usually comes to attention in middle-aged and older individuals on the neck, trunk, face, and intertriginous areas. Rarely, fibroepithelial polyps and tumors of perifollicular mesenchyme (specialized fibroblasts associated with the hair bulb) are seen together in *Birt-Hogg-Dubé syndrome,* but the vast majority of polyps are sporadic.

MORPHOLOGY

Fibroepithelial polyps are soft, flesh-colored, bag-like tumors that are often attached to the surrounding skin by a slender stalk. On histologic examination, they consist of fibrovascular cores covered by benign squamous epithelium. It is not uncommon for the polyps to undergo ischemic necrosis due to torsion, which may cause pain and precipitate their removal.

Fibroepithelial polyps are usually inconsequential, but can occasionally be associated with diabetes, obesity, and intestinal polyposis. Of interest, like melanocytic nevi and hemangiomas, they often become more numerous or prominent during pregnancy, presumably related to hormonal stimulation.

Epithelial or Follicular Inclusion Cyst (Wen)

Epithelial cysts are common lesions formed by the invagination and cystic expansion of the epidermis or, perhaps more commonly, a hair follicle. The lay term, *wen,* derives from the Anglo-Saxon *wenn,* meaning a lump or tumor. When large, they may be subject to traumatic rupture, which can spill keratin into the dermis and lead to an extensive and often painful granulomatous inflammatory response.

ADNEXAL (APPENDAGE) TUMORS

There are literally hundreds of neoplasms arising from or showing differentiation toward cutaneous appendages. Their significance varies according to type and clinical context, as follows:

- *Some are benign* but may be confused with cutaneous cancers such as basal cell carcinoma.
- *Other appendage tumors are associated with germline mutations in tumor suppressor genes.* In some such instances, the disorder presents with multiple adnexal tumors that may be disfiguring. In other cases, the appendage tumors are relatively trivial but warn of a predisposition for internal malignancy; such is the relationship between multiple *trichilemmomas* and *Cowden syndrome,* a disorder caused by germline mutations in the tumor suppressor gene *PTEN* that is associated with an increased risk of endometrial cancer, breast cancer, and other malignancies.

Appendage tumors often appear as nondescript, flesh-colored solitary or multiple papules and nodules. Some have a predisposition to occur on specific body surfaces. Selected examples are provided here to illustrate neoplasms of hair follicles and sebaceous, eccrine, and apocrine glands.

- *Eccrine poroma* occurs predominantly on the palms and soles where sweat glands are numerous.
- *Cylindroma,* an appendage tumor with ductal (apocrine or eccrine) differentiation, usually occurs on the forehead and scalp (Fig. 25.10A), where coalescence of nodules with time may produce a hat-like growth—hence the name *turban tumor.* These lesions can be dominantly inherited; in such cases they appear early in life and are associated with inactivating mutations in the tumor suppressor gene *CYLD,* which encodes a deubiquitinating enzyme that negatively regulates the oncogenic transcription factor NF-κB and other factors that contribute to cell cycle progression. In addition to familial cylindromatosis, germline mutations in *CYLD* are associated with two other genetic syndromes marked by the occurrence of multiple adnexal tumors, multiple *familial trichoepithelioma* (a follicular tumor) and *Brooke-Spiegler syndrome* (associated with both trichoepithelioma and cylindroma).
- *Syringoma,* a lesion with eccrine differentiation, usually occurs as multiple, small, tan papules in the vicinity of the lower eyelids.
- *Sebaceous adenoma* can be associated with internal malignancy in *Muir-Torre syndrome* and also is associated in a subset of cases with hereditary nonpolyposis colorectal carcinoma syndrome (also known as *Lynch syndrome;* see Chapter 17), which is caused by germline defects in DNA mismatch repair genes.
- *Pilomatricoma,* showing hair follicle differentiation, is associated with activating mutations in *CTNNB1,* the gene encoding β-catenin. Mutations in this gene are seen in numerous neoplasms but are of interest here since Wnt signaling through β-catenin is critical for early hair development and regulates hair growth and maintenance.
- *Adnexal tumors can also show apocrine differentiation;* these usually arise in body areas where apocrine glands are most prevalent such as the axilla and scalp.

MORPHOLOGY

The **cylindroma** is composed of islands of cells resembling the normal epidermal or adnexal basal cell layer (basaloid cells). These islands fit together like pieces of a jigsaw puzzle within a fibrous

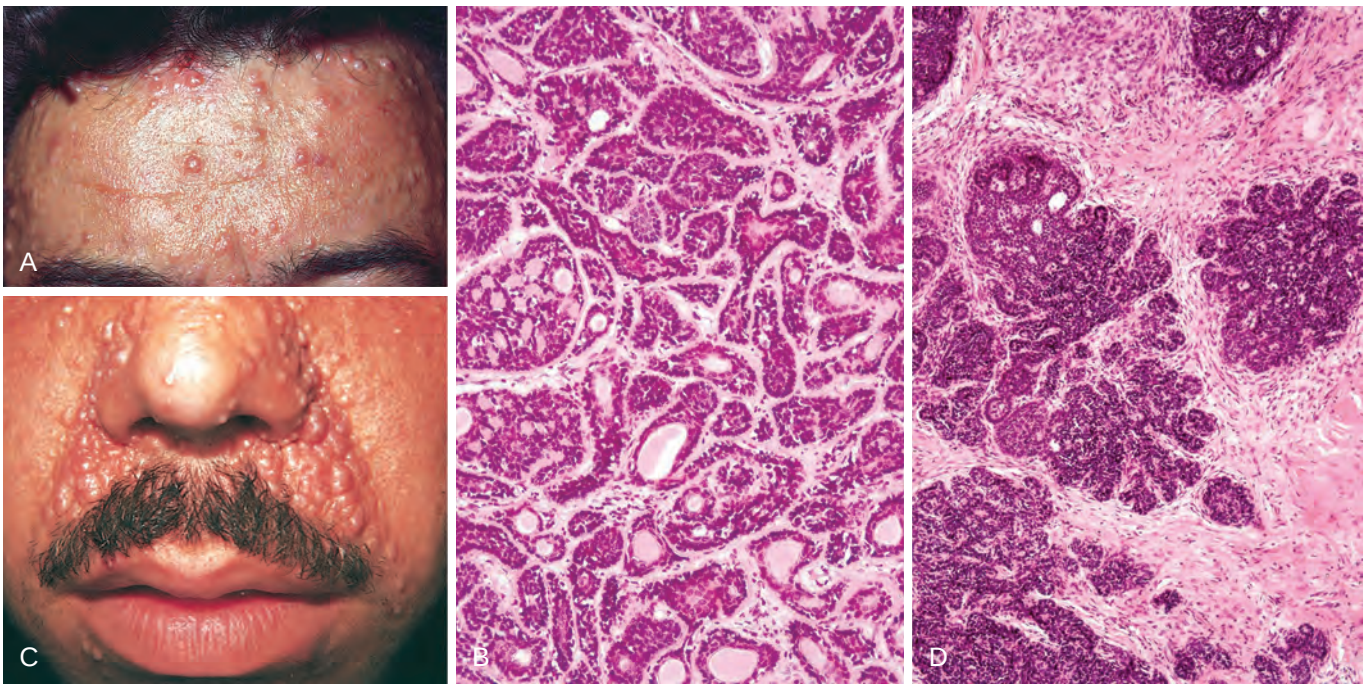


Figure 25.10 Cylindroma and trichoepithelioma. (A, B) Multiple cylindromas (papules) on the forehead (A) are composed of islands of basaloid cells (B) containing occasional ducts that fit together like pieces of a jigsaw puzzle. (C, D) Perinasal papules and small nodules of trichoepithelioma (C) are composed of buds of basaloid cells (D) that resemble primitive hair follicles.

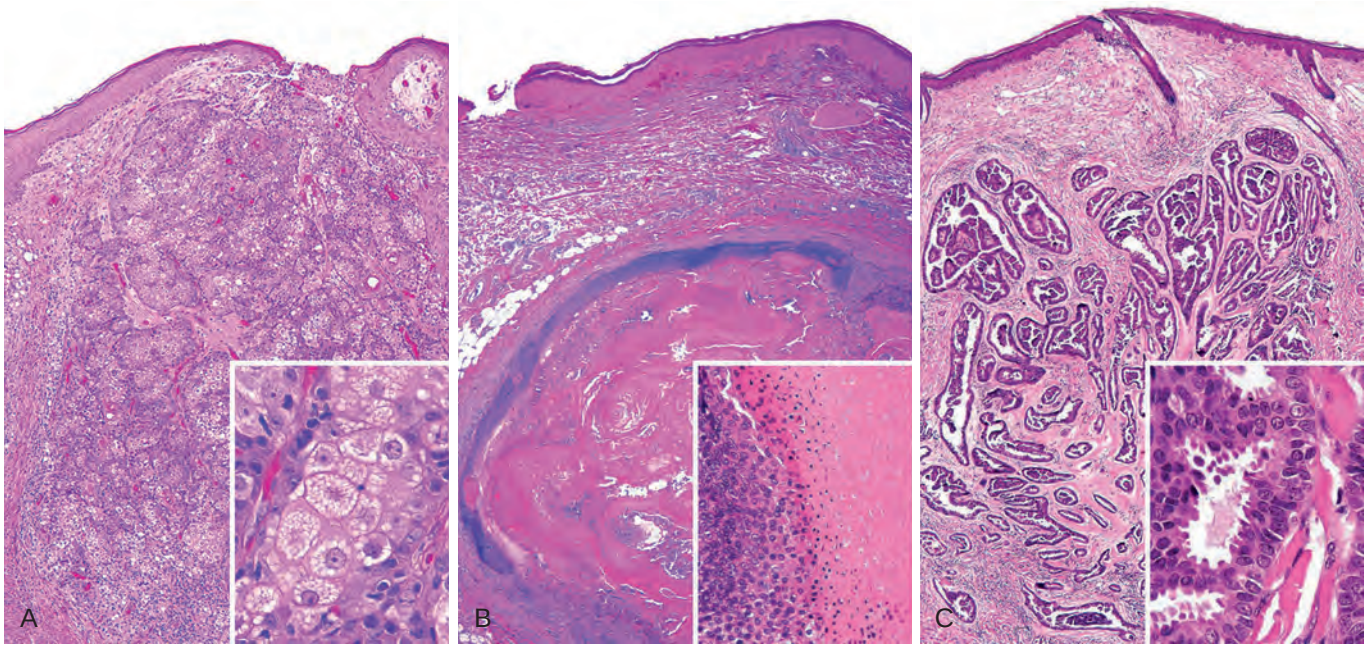


Figure 25.11 Diverse adnexal tumors. (A) Sebaceous adenoma; *inset* demonstrates sebaceous differentiation. (B) Pilomatrixoma; *inset* shows hair matrix differentiation to anucleate "ghost cells." (C) Apocrine carcinoma (well differentiated); *inset* shows apocrine differentiation and luminal secretions produced by decapitation of the lining cells.

dermal matrix (Fig. 25.10B). **Trichoepithelioma** is a proliferation of basaloid cells that forms primitive structures resembling hair follicles (Fig. 25.10C, D). **Sebaceous adenoma** shows a lobular proliferation of sebocytes with increased peripheral basaloid cells and more mature sebocytes in the central portion that have frothy or bubbly cytoplasm due to the presence of lipid vesicles (Fig. 25.11A). **Pilomatrixomas** are composed of basaloid cells that show trichilemmal or hair-like differentiation similar to that seen in the germinal portion of the normal hair bulb in the anagen growth phase (Fig. 25.11B). **Apocrine carcinoma** shows ductal differentiation with prominent decapitation secretion similar to that seen in the normal apocrine gland (Fig. 25.11C). The infiltrative growth pattern is a hint of malignancy in this otherwise well-differentiated tumor.

Although most appendage tumors are benign, malignant variants exist. Apocrine tumors are unusual in that malignant forms appear to be more common than benign forms. *Sebaceous carcinoma* arises from the meibomian glands of the eyelid and may follow an aggressive course replete with systemic metastases. *Eccrine* and *apocrine carcinomas* can be confused with metastatic adenocarcinoma because of their tendency to form gland-like structures.

PREMALIGNANT AND MALIGNANT EPIDERMAL TUMORS

Actinic Keratosis

Actinic keratosis (as the name implies) usually occurs in sun-damaged skin and exhibits hyperkeratosis. As expected, these tumors have a particularly high incidence in lightly pigmented individuals. Exposure to ionizing radiation, industrial hydrocarbons, and arsenicals may induce similar lesions. These lesions may show progressively worsening dysplastic changes that culminate in cutaneous squamous cell carcinoma and are analogous in this regard to the precursor lesions that give rise to squamous carcinomas of the uterine cervix (Chapter 22).

MORPHOLOGY

Actinic keratoses are usually less than 1 cm in diameter. They are typically tan-brown, red, or skin-colored and have a rough, sandpaper-like consistency. Some lesions produce so much keratin that a "cutaneous horn" develops (Fig. 25.12A), which in extreme cases may become so prominent that they resemble the actual horns of animals! Sun-exposed sites (face, arms, dorsum of hands) are most frequently affected. The lips may also develop similar lesions (termed **actinic cheilitis**).

Cytologic atypia is seen in the lowermost layers of the epidermis and may be associated with hyperplasia of basal cells (Fig. 25.12B) or, alternatively, with atrophy that results in thinning of the epidermis. The atypical basal cells usually have pink or reddish cytoplasm due to dyskeratosis. Intercellular bridges are present, in contrast to basal cell carcinoma, in which they are not visible. The superficial dermis contains thickened, blue-gray elastic fibers **(elastosis),** a probable result of abnormal elastic fiber synthesis by sun-damaged fibroblasts. The stratum corneum is thickened, and, unlike normal skin, the cells in this layer often retain their nuclei **(parakeratosis).**

Actinic keratoses may regress or remain stable throughout life, but enough transform to malignancy that local eradication is warranted. This can be accomplished by gentle curettage,

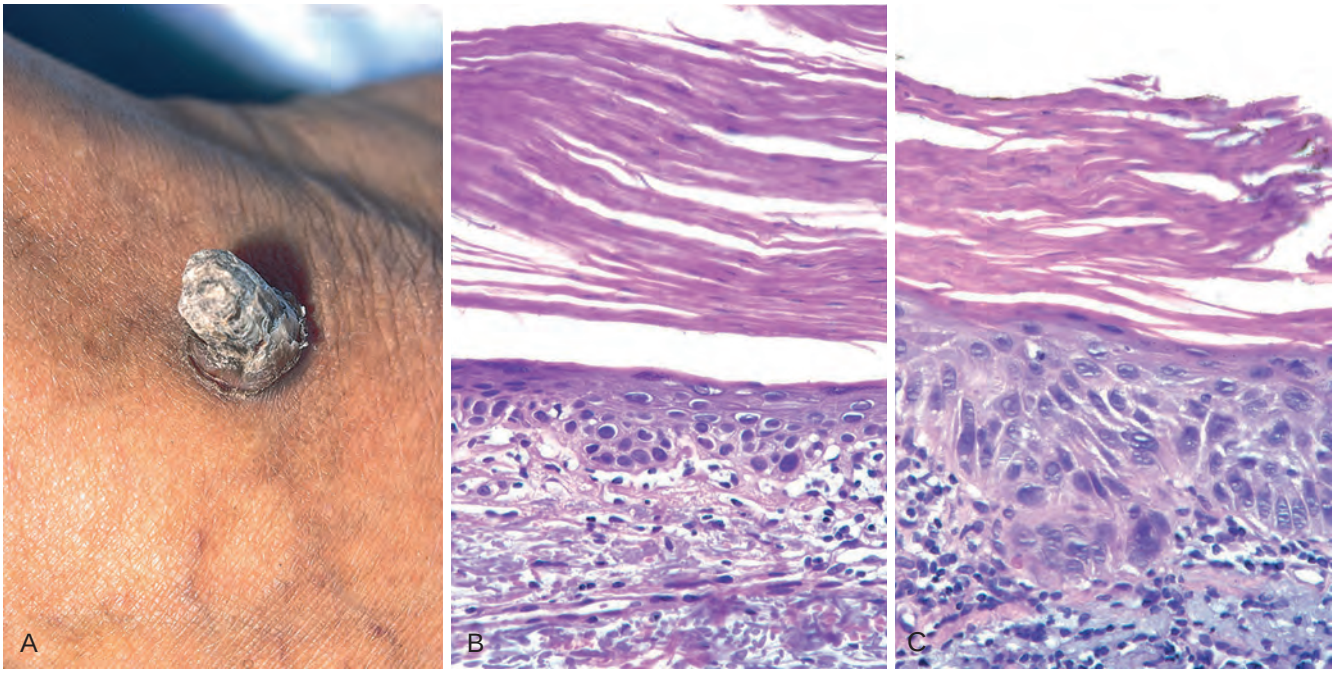


Figure 25.12 Actinic keratosis. (A) Excessive keratotic scale in this lesion has produced a "cutaneous horn." (B) Basal cell layer atypia (dysplasia) is associated with marked hyperkeratosis and parakeratosis. (C) Progression to full-thickness nuclear atypia, with or without the presence of superficial epidermal maturation, heralds the development of squamous cell carcinoma in situ.

freezing, or topical application of chemotherapeutic agents. Of interest, topical administration of imiquimod, a drug that activates Toll-like receptors (TLRs), eradicates up to 50% of lesions, a rate considerably higher than the spontaneous regression rate of approximately 5%. By stimulating TLR signaling, imiquimod activates cutaneous innate immune cells, which may recognize and eradicate precancerous lesions.

Squamous Cell Carcinoma

Squamous cell carcinoma is the second most common tumor arising on sun-exposed sites in older people, exceeded only by basal cell carcinoma. Except for lesions on the lower legs, these tumors have a higher incidence in men than in women. Invasive squamous cell carcinomas are usually discovered while they are small and resectable. Less than 5% of these tumors metastasize to regional nodes; these lesions are generally deeply invasive and involve the subcutis.

Pathogenesis

The most important cause of cutaneous squamous cell carcinoma is DNA damage induced by exposure to UV light. Tumor incidence is proportional to the degree of lifetime sun exposure. A second common association is with immunosuppression, most notably chronic immunosuppression as a result of chemotherapy or organ transplantation. Immunosuppression may contribute to carcinogenesis by reducing host surveillance and increasing the susceptibility of keratinocytes to infection and transformation by viruses, particularly human papillomavirus (HPV) subtypes 5 and 8. These same HPVs have been implicated in tumors arising in patients with the rare autosomal recessive condition *epidermodysplasia verruciformis,* which is marked by a high susceptibility to cutaneous squamous cell carcinomas. In addition to its damaging effect on DNA, sunlight, through uncertain mechanisms, seems to cause a transient defect in cutaneous innate immunity that may diminish immune-mediated elimination of sun-damaged cells. Other risk factors for squamous cell carcinoma include industrial carcinogens (tars and oils), chronic ulcers and draining osteomyelitis, old burn scars, ingestion of arsenicals, ionizing radiation, and (in the oral cavity) tobacco and betel nut chewing.

Most studies on the genetics of squamous cell carcinoma have focused on acquired defects in sporadic tumors and their precursors (actinic keratoses) and the relationships between these defects and sun exposure. The incidence of *TP53* mutations in actinic keratoses found in Caucasians is high, suggesting that p53 dysfunction is an early event in the development of tumors induced by sunlight. Normally, DNA damaged by UV light is sensed by checkpoint kinases such as ATM and ATR, which send out signals that upregulate the expression and stability of p53. p53 in turn arrests cells in the G_1 phase of the cell cycle and promotes either "high-fidelity" DNA repair or the elimination of cells that are damaged beyond repair (Chapter 7). When these protective functions of p53 are lost, DNA damage induced by UV light is more likely to be "repaired" by error-prone mechanisms, creating mutations that are passed down to daughter cells. Of note, the mutations that are seen in *TP53* often occur at pyrimidine dimers, indicating that they, too, stem from damage caused by UV light. A similar story underlies the remarkable susceptibility of patients with *xeroderma pigmentosum* to squamous cell carcinoma. This disorder is caused by inherited mutations in genes in the nucleotide excision repair pathway, which is required for accurate repair of pyrimidine dimers; when this pathway is defective,

error-prone repair pathways take over, leading to the rapid accumulation of mutations and eventual carcinogenesis.

As with all other forms of cancer, cutaneous squamous cell carcinoma stems from multiple driver mutations. In addition to defects in p53, mutations that increase RAS signaling and decrease Notch signaling are common and also contribute to the transformation process.

MORPHOLOGY

Squamous cell carcinomas that have not invaded through the basement membrane of the dermoepidermal junction (termed in situ carcinoma) appear as sharply defined, red, scaling plaques. More advanced, invasive lesions are nodular, show variable keratin production (appreciated grossly as hyperkeratotic scale), and may ulcerate (Fig. 25.13A).

Unlike actinic keratoses, in squamous cell carcinoma in situ, cells with atypical (enlarged and hyperchromatic) nuclei involve all levels of the epidermis (Fig. 25.12C). Invasive squamous cell carcinoma (Fig. 25.13B, C) shows variable degrees of differentiation, ranging from tumors composed of polygonal cells arranged in orderly lobules and having numerous large areas of keratinization to neoplasms consisting of highly anaplastic cells that exhibit only abortive, single-cell keratinization (dyskeratosis). The latter tumors may be so poorly differentiated that immunohistochemical stains for keratins are needed to confirm the diagnosis.

Basal Cell Carcinoma

Basal cell carcinoma is a distinctive locally aggressive cutaneous tumor that is associated with mutations that activate the Hedgehog signaling pathway. Basal cell carcinoma is the most common invasive cancer in humans, numbering nearly 1 million cases per year in the United States. These are slow-growing tumors that rarely metastasize. The vast majority are recognized at an early stage and cured by local excision. However, a small number of tumors (<0.5%) are locally aggressive and potentially disfiguring, and on rare occasions metastasis to distant sites is seen. Basal cell carcinoma occurs at sun-exposed sites in lightly pigmented elderly adults. As with squamous cell carcinoma, the incidence of basal cell carcinoma is increased in the setting of immunosuppression and in disorders of DNA repair, such as *xeroderma pigmentosum* (Chapter 7).

Pathogenesis

Most basal cell carcinomas have mutations that lead to unbridled Hedgehog signaling. As is often the case in biology and medicine, study of a rare genetic syndrome associated with a high risk of a common disease (basal cell carcinoma) has led to the elucidation of pathogenic mechanisms of general importance. The syndrome in question, *nevoid basal cell carcinoma syndrome* (NBCCS) (also known as basal cell nevus or *Gorlin syndrome*), is an autosomal dominant disorder characterized by the development of multiple basal cell carcinomas, often before age 20, accompanied by various other tumors (especially medulloblastomas and ovarian fibromas), odontogenic keratocysts, pits of the palms and soles, and certain developmental abnormalities. NBCCS is one of a number of cancer syndromes associated with skin manifestations (Table 25.3). The gene associated with NBCCS is *PTCH,* a tumor suppressor gene that is the human homologue of the *Drosophila* developmental gene *patched*. Individuals with NBCCS are born with a germline loss-of-function mutation in one *PTCH* allele; the second normal allele is inactivated in tumors by an acquired

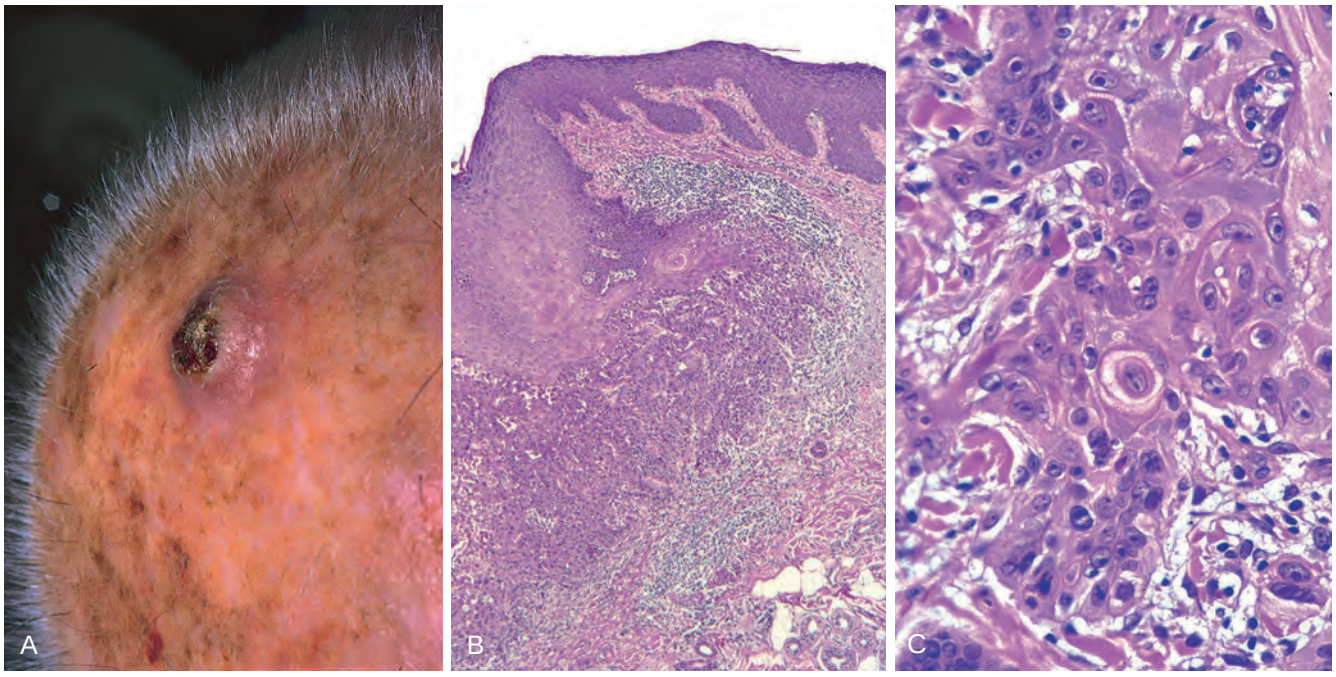


Figure 25.13 Invasive squamous cell carcinoma. (A) Lesions are often nodular and ulcerated, as seen in this scalp tumor. (B) Tongues of atypical squamous epithelium have transgressed the basement membrane and invaded deeply into the dermis. (C) Invasive tumor cells show enlarged nuclei with angulated contours and prominent nucleoli.

Table 25.3 Survey of Familial Cancer Syndromes With Cutaneous Manifestations

Disease	Inheritance	Chromosomal Location	Gene/Protein	Normal Function/Manifestation of Loss
Ataxia-telangiectasia	AR	11q22.3	*ATM*/ATM	DNA repair after radiation injury/neurologic and vascular lesions
Nevoid basal cell carcinoma syndrome	AD	9q22	*PTCH*/PTCH	Developmental patterning gene/multiple basal cell carcinomas; medulloblastoma; jaw cysts
Cowden syndrome	AD	10q23	*PTEN*/PTEN	Lipid phosphatase/benign follicular appendage tumors (trichilemmomas); internal adenocarcinoma (often breast or endometrial)
Familial atypical mole and melanoma syndrome	AD	9p21	*CDKN2A*/p16, INK4 *CDKN2A*/p14, ARF	Inhibits CDK4/6 phosphorylation of RB, promoting cell cycle arrest/melanoma; pancreatic carcinoma Binds MDM2, promoting p53 function/melanoma; pancreatic carcinoma
Muir-Torre syndrome	AD	2p22 3p21	*MSH2*/MSH2 *MLH1*/MLH1	Involved in DNA mismatch repair/sebaceous neoplasia; internal malignancy (colon and others)
Neurofibromatosis 1	AD	17q11	*NF1*/neurofibromin	Negatively regulates RAS signaling/neurofibromas
Neurofibromatosis 2	AD	22q12	*NF2*/merlin	Integrates cytoskeletal signaling/neurofibromas; acoustic neuromas
Tuberous sclerosis	AD	9q34 16p13	*TSC1*/hamartin *TSC2*/tuberin	Work together in a complex that negatively regulates mTOR/angiofibromas; intellectual disability
Xeroderma pigmentosum	AR	9q22 and others	*XPA*/XPA and others	Nucleotide excision repair/melanoma; nonmelanoma skin cancers

AD, Autosomal dominant; *AR*, autosomal recessive.

From Tsai KY, Tsao H: The genetics of skin cancer, *Am J Med Genet C Semin Med Genet* 131C:82, 2004.

mutation, usually caused by exposure to mutagens (particularly UV light).

PTCH protein is a receptor for *sonic hedgehog* (SHH), a component of the Hedgehog signaling pathway, which controls polarity and CNS development during embryogenesis and also regulates hair follicle formation and hair growth. In the "off" state, PTCH exists in a complex with another transmembrane protein called SMO (for "smoothened"). Binding of SHH to PTCH releases SMO, which in turn activates the transcription factor GLI1 (Fig. 25.14), thus turning on the expression of genes that support tumor cell growth and survival. Mice engineered to have excessive GLI1 activation are prone to develop skin tumors resembling basal cell carcinomas. Similarly, in NBCCS the loss of PTCH function causes constitutive activation of SMO and GLI1, leading to the development of basal cell carcinoma.

Mutations that activate Hedgehog signaling are also prevalent in sporadic basal cell carcinomas. Loss-of-function *PTCH* mutations are common, and about one-third of these mutations consist of C→T transitions that are hallmarks of UV damage. This insight has paved the way for the development and clinical implementation of small molecule inhibitors of the Hedgehog pathway.

MORPHOLOGY

Basal cell carcinomas usually present as **pearly papules** containing prominent dilated subepidermal blood vessels **(telangiectasias)** (Fig. 25.15A). Some tumors contain melanin and superficially resemble melanocytic nevi or melanomas. Advanced lesions may ulcerate, and extensive local invasion of bone or facial sinuses may occur after many years of neglect or in unusually aggressive tumors, explaining the archaic designation **rodent ulcers.** One common and important variant, the superficial basal cell carcinoma, presents as an erythematous, occasionally pigmented plaque that may resemble early forms of melanoma.

Histologically, the tumor cells resemble those in the normal basal cell layer of the epidermis. They arise from the epidermis or follicular epithelium and do not occur on mucosal surfaces. Two patterns are seen: **multifocal growths** originating from the epidermis and sometimes extending over several square centimeters or more of skin surface (multifocal superficial type) and **nodular lesions** growing downward deeply into the dermis as cords and islands of variably basophilic cells with hyperchromatic nuclei, embedded in a mucinous matrix and often surrounded by many fibroblasts and lymphocytes (Fig. 25.15B). The cells at the periphery of the tumor cell islands tend to be arranged radially with their long axes in parallel alignment **(palisading).** In sections, the stroma retracts away from the carcinoma (Fig. 25.15C), creating clefts or separation artifacts that assist in differentiating basal cell carcinomas from certain appendage tumors that are also characterized by proliferation of basaloid cells, such as trichoepithelioma.

KEY CONCEPTS

MALIGNANT EPIDERMAL TUMORS

- The incidence of both basal cell carcinoma and squamous cell carcinoma is strongly correlated with increasing lifetime sun exposure.
- Cutaneous squamous cell carcinoma can progress from actinic keratoses but also arises from chemical exposure, at thermal

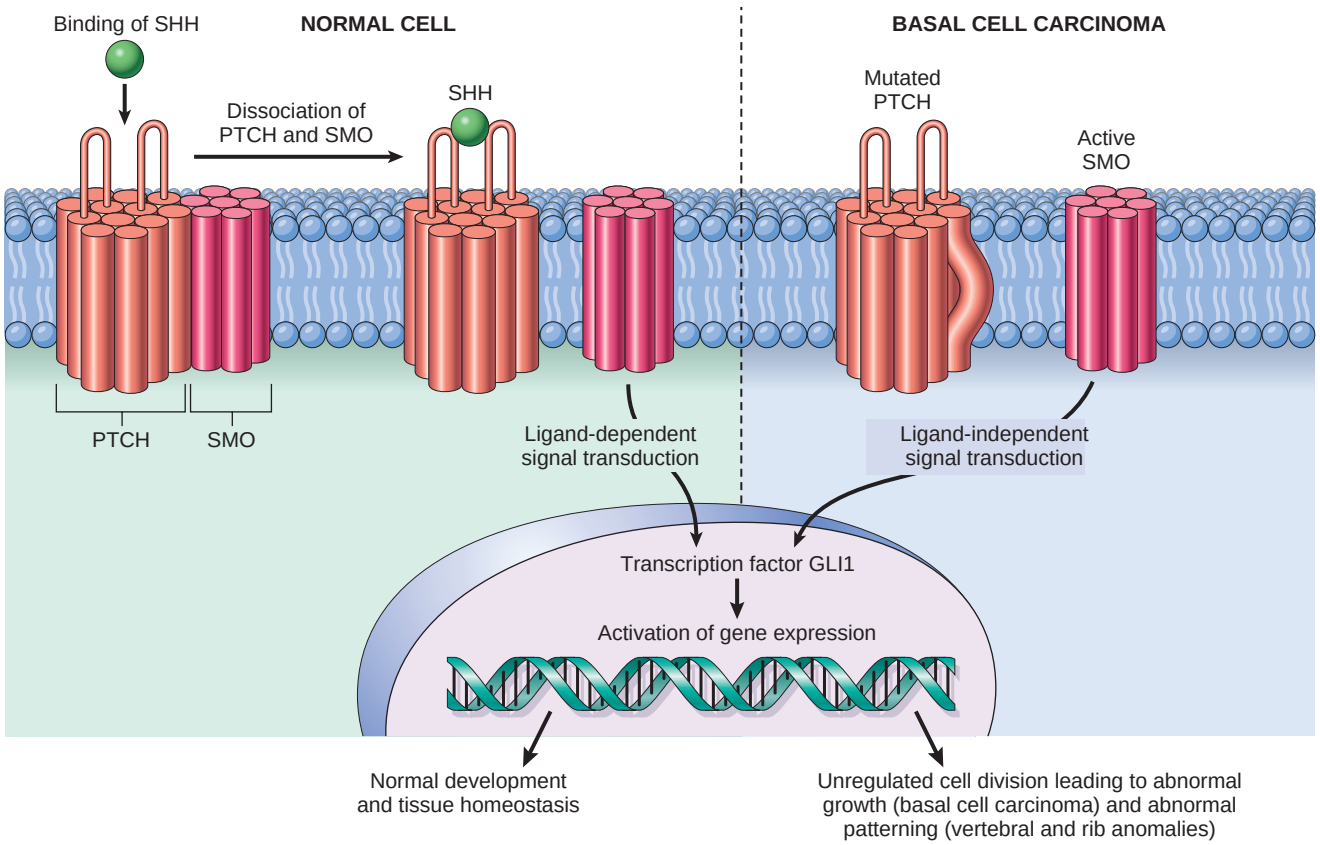


Figure 25.14 Normal and oncogenic hedgehog signaling. *Left,* Normally, PTCH and SMO form a receptor complex that can bind sonic hedgehog *(SHH).* In the absence of SHH, PTCH blocks SMO activity. When SHH binds PTCH, SMO is released to trigger a signal transduction cascade that leads to activation of GLI1 and other transcription factors. *Right,* Mutations in *PTCH,* and less often in *SMO,* allow SMO to signal without SHH binding and produce constitutive activation of GLI1. GLI signaling is a characteristic feature of sporadic basal cell carcinomas and tumors associated with nevoid basal cell carcinoma (Gorlin) syndrome.

burn sites, or in association with HPV infection in the setting of immunosuppression.
- Cutaneous squamous cell carcinoma has potential for metastasis but is much less aggressive than squamous cell carcinoma at mucosal sites.
- Basal cell carcinoma, the most common malignancy worldwide, is a locally aggressive tumor associated with mutations that activate Hedgehog signaling. Metastasis is very rare.

TUMORS OF THE DERMIS

The dermis contains a variety of elements such as smooth muscle, pericytes, fibroblasts, neural tissue, and endothelium. Neoplasms comprised of cells resembling all of these elements occur in the skin, but most also involve other soft tissues and viscera and are discussed elsewhere or are too rare to merit mention. This section discusses two dermal neoplasms—one benign, one malignant—that arise primarily in the skin.

Benign Fibrous Histiocytoma (Dermatofibroma)

Benign fibrous histiocytoma refers to a heterogeneous family of morphologically and histogenetically related benign dermal neoplasms of uncertain lineage. These tumors are usually seen in adults and often occur on the legs of young and middle-aged women. Lesions are asymptomatic or tender and may increase and decrease slightly in size over time. Their biologic behavior is indolent.

The cause of fibrous histiocytomas remains a mystery. Some cases have a history of antecedent trauma, suggesting an abnormal response to injury and inflammation, perhaps analogous to the deposition of increased amounts of altered collagen in a hypertrophic scar or keloid. However, several fusion genes, including one in which one partner is the gene for the receptor tyrosine kinase ALK, have been identified in a subset of cases, suggesting that these proliferations are best considered true neoplasms. These tumors appear to be composed at least partially of dermal dendritic cells.

MORPHOLOGY

These neoplasms appear as firm, tan to brown papules (Fig. 25.16A). Most are less than 1 cm in diameter, but actively growing lesions may reach several centimeters in diameter; with time they often become flattened.

The most common form of fibrous histiocytoma is referred to as a **dermatofibroma.** These tumors consist of benign, spindle-shaped cells that are usually arranged in a well-defined, nonencapsulated mass within the mid-dermis (Fig. 25.16B, C). Extension of these cells into the subcutaneous fat is sometimes

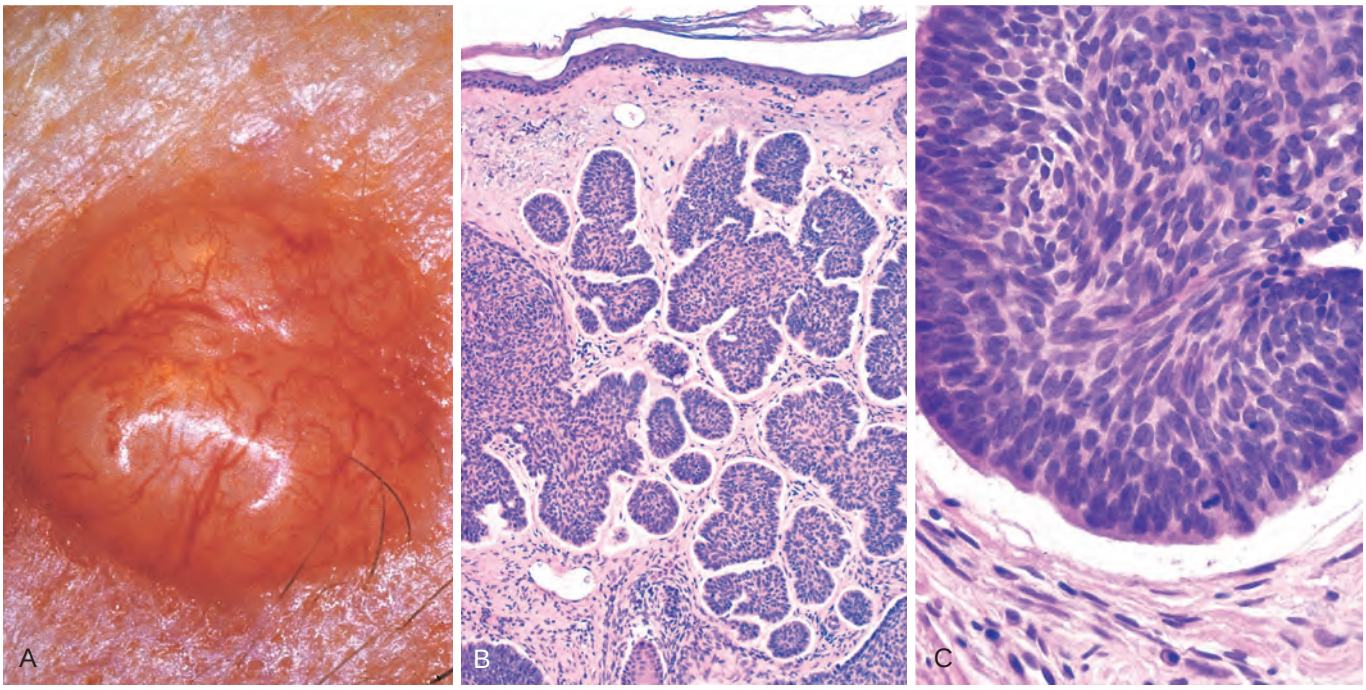


Figure 25.15 Basal cell carcinoma. Pearly, telangiectatic nodules (A) are composed of nests of uniform basaloid cells within the dermis (B) that are often separated from the adjacent stroma by thin clefts (C), an artifact of sectioning.

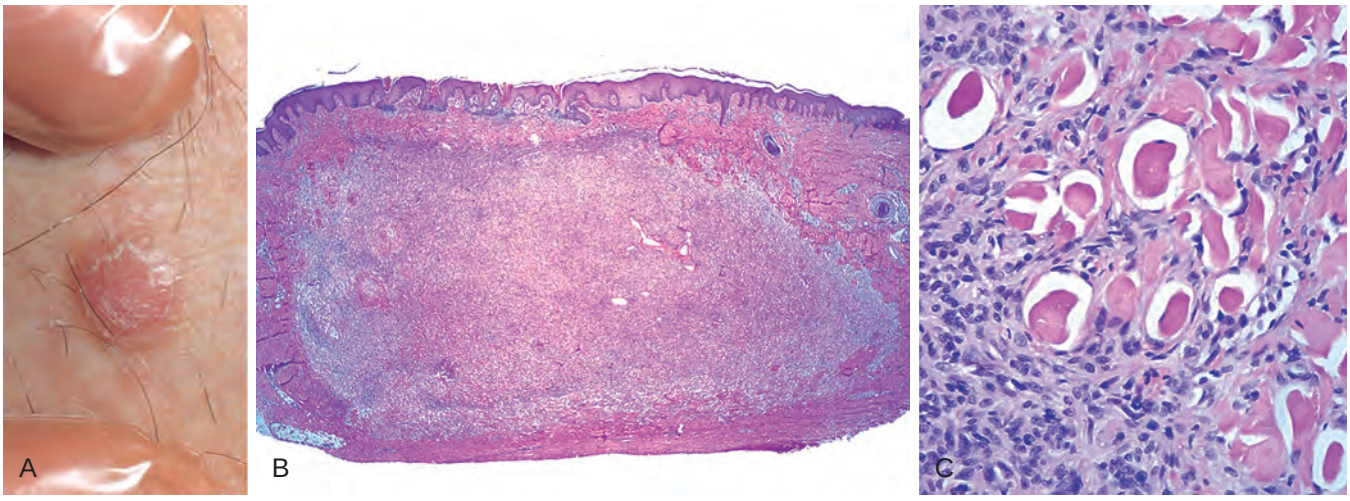


Figure 25.16 Benign fibrous histiocytoma (dermatofibroma). This firm, tan papule on the leg (A) contains a circumscribed dermal proliferation of benign-appearing spindle cells (B). Note the characteristic overlying epidermal hyperplasia (B) and the tendency of fibroblasts to surround individual collagen bundles (C).

observed. Many cases demonstrate a peculiar form of overlying epidermal hyperplasia, characterized by downward elongation of hyperpigmented rete ridges **(pseudoepitheliomatous hyperplasia).** Numerous histologic variants are noted, such as more cellular forms or tumors with pools of extravascular blood and hemosiderin (aneurysmal variants).

Dermatofibrosarcoma Protuberans

Dermatofibrosarcoma protuberans is best regarded as a well-differentiated, primary fibrosarcoma of the skin. These tumors are slow growing, and although they are locally aggressive and can recur, they rarely metastasize. Metastasis most often is seen with tumors that exhibit greater cytologic atypia.

Pathogenesis

The molecular hallmark of dermatofibrosarcoma protuberans is a translocation involving the genes encoding collagen 1A1 *(COL1A1)* and platelet-derived growth factor-β *(PDGFB).* The resulting rearrangement juxtaposes the *COL1A1* promoter sequences and the coding region of *PDGFB* and leads to overexpression and increased secretion of PDGFβ, which drives tumor cell growth through an autocrine loop. While the primary mode of treatment is wide local

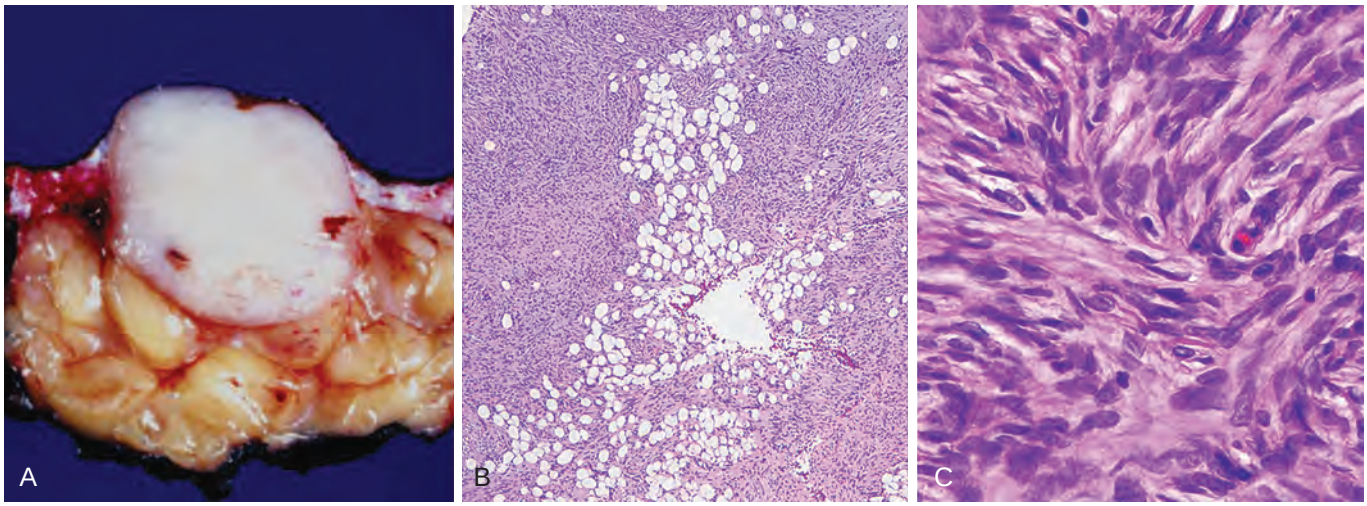


Figure 25.17 Dermatofibrosarcoma protuberans. (A) The tumor consists of a flesh-colored fibrotic nodule on sectioning. (B) The lesion often infiltrates the subcutis in a manner reminiscent of "Swiss cheese" to aficionados. (C) Characteristic storiform (swirling) alignment of the spindled cells is apparent.

excision, rare cases that are unresectable due to their location or because of metastatic spread can be treated with inhibitors of the PDGFβ receptor tyrosine kinase. Withdrawal of tyrosine kinase inhibitor therapy leads to regrowth of the tumor, so treatment must be lifelong.

MORPHOLOGY

Dermatofibrosarcoma protuberans usually appears as a "protuberant" nodule, most often on the trunk, within a firm (indurated) plaque that may sometimes ulcerate (Fig. 25.17A). These neoplasms are composed of closely packed fibroblasts arranged radially, reminiscent of blades of a pinwheel, a pattern referred to as **storiform.** Mitoses are rare. In contrast to dermatofibroma, the overlying epidermis is generally thinned. Deep extension from the dermis into subcutaneous fat, producing a characteristic "honeycomb" pattern, is frequently seen (Fig. 25.17B, C). These tumors may extend down into the subcutis and thus require wide excision to prevent local recurrence.

TUMORS OF CELLULAR MIGRANTS TO THE SKIN

Aside from tumors that arise directly from epidermal and dermal cells, several proliferative disorders of the skin involve cells whose progenitors arise elsewhere and then specifically home to the cutaneous microenvironment.

Mycosis Fungoides (Cutaneous T-Cell Lymphoma)

Cutaneous T-cell lymphoma (CTCL) spans a spectrum of lymphoproliferative disorders affecting the skin (Chapter 13), many with distinctive presentations. This section focuses on *mycosis fungoides,* a lymphoma of skin-homing CD4+ T-helper cells that presents in the skin. In most affected individuals, the disease remains localized to the skin for many years, but it may eventually evolve into a systemic lymphoma. This tumor may occur at any age, but most commonly afflicts persons older than age 40.

Lesions of mycosis fungoides usually involve truncal areas and include scaly, red-brown patches; raised, scaling plaques that may be confused with psoriasis; and fungating nodules. Prognosis is related to the percentage of body surface involved and progression from patch to plaque to nodular forms. Eczema-like lesions typify early stages of disease when obvious visceral or nodal spread has not occurred. Raised, indurated, irregularly outlined erythematous plaques may then supervene. Development of multiple tumor nodules correlates with systemic spread. Sometimes the plaques and nodules ulcerate (Fig. 25.18A). Ultimately, lesions may affect numerous body surfaces including the trunk, extremities, face, and scalp. In some individuals, seeding of the blood by malignant T cells is accompanied by diffuse erythema and scaling of the entire body surface (erythroderma), a condition known as *Sézary syndrome* (Chapter 13).

The proliferating cells in CTCL are usually clonal populations of CD4+ T-helper cells that home to the skin due to expression of cutaneous lymphocyte antigen. A wide variety of driver gene mutations in oncogenes and tumor suppressor genes have been described, none of which appear to be specific for CTCL. The neoplastic cells have clonal T-cell receptor gene rearrangements and sometimes express aberrant combinations of T-cell surface antigens. Topical therapy with steroids or UV light is often used for early skin lesions, whereas more aggressive systemic chemotherapy is indicated for advanced disease.

MORPHOLOGY

The histologic hallmark of CTCL of the mycosis fungoides type is the presence of atypical cells that characteristically form band-like aggregates within the superficial dermis (Fig. 25.18B) and invade the epidermis as single cells and small clusters **(Pautrier microabscesses).** These cells have markedly infolded nuclear membranes, imparting a convoluted or cerebriform contour. Although patches and plaques show pronounced epidermal infiltration by neoplastic cells (epidermotropism), in more advanced nodular lesions the malignant T cells often lose their epidermotropism, grow deeply into the dermis to produce nodules, and eventually spread systemically.

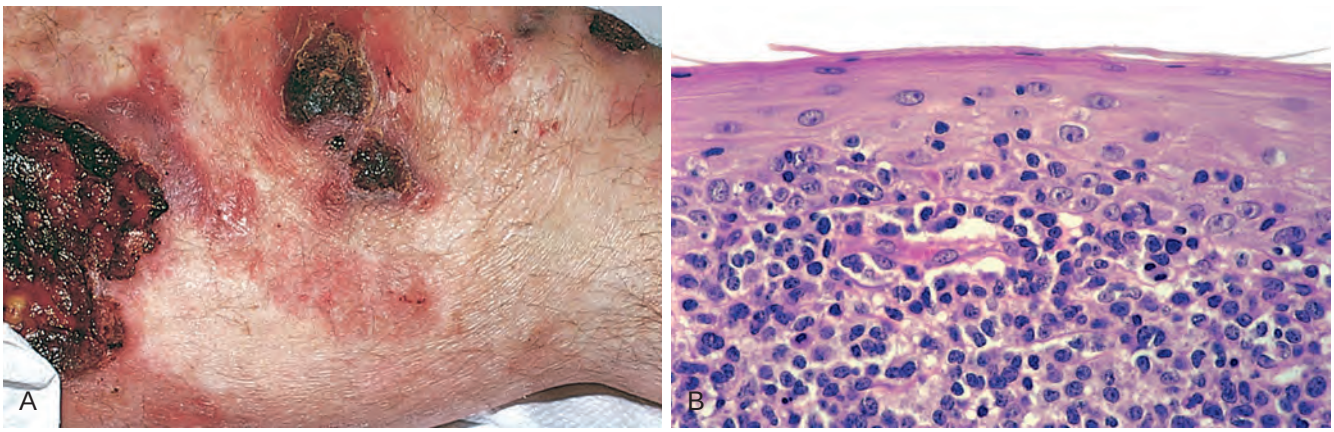


Figure 25.18 Cutaneous T-cell lymphoma. (A) Several erythematous plaques with scaling and ulceration are evident. (B) Microscopically, there is an infiltrate of atypical lymphocytes that accumulates beneath and invades the epidermis.

Mastocytosis

The term *mastocytosis* encompasses a spectrum of rare disorders characterized by increased numbers of mast cells in the skin and, in some instances, in other organs as well. A cutaneous form of the disease that affects predominantly children and accounts for more than 50% of all cases is termed *urticaria pigmentosa.* The cutaneous lesions are usually multiple, although solitary mastocytomas may also occur in very young children. About 10% of individuals with mast cell disease have systemic disease, with mast cell infiltration of many organs. These individuals are often adults, and unlike those with localized cutaneous disease, the prognosis in individuals with systemic disease is more guarded.

Many of the signs and symptoms of mastocytosis are due to the release of histamine, heparin, and other substances when mast cells degranulate. *Darier sign* refers to a localized area of dermal edema and erythema (wheal) that occurs when lesional skin is rubbed. *Dermatographism* refers to an area of dermal edema resembling a hive that occurs as a result of localized stroking of apparently normal skin with a pointed instrument. In systemic disease, all of the following may be seen: pruritus and flushing, triggered by certain foods, temperature changes, alcohol, and certain drugs (morphine, codeine, aspirin); watery nasal discharge (rhinorrhea); rarely, gastrointestinal or nasal bleeding, possibly due to the anticoagulant effects of heparin; and bone pain, which may be caused by mast cell infiltration or by pathologic fractures stemming from osteoporosis. Osteoporosis is caused by excessive histamine release in the marrow microenvironment and can be a clue to the diagnosis, particularly in premenopausal women and in men.

Pathogenesis

Many cases of mastocytosis have acquired activating point mutations in the KIT, or less frequently, the PDGFR-α receptor tyrosine kinases. The resulting increase in KIT signaling drives mast cell growth and survival. This insight has led to the clinical development of KIT kinase inhibitors, which often produce dramatic tumor regression, even in patients with advanced aggressive systemic disease.

MORPHOLOGY

The pathologic findings are highly variable. In **urticaria pigmentosa,** lesions are multiple and widely distributed, consisting of round to oval, red-brown, nonscaling papules and small plaques. Solitary **mastocytoma** presents as a pink to tan-brown nodule that may be pruritic or show blister formation (Fig. 25.19A). The histologic picture in urticaria pigmentosa or solitary mastocytoma varies from a subtle increase in mast cells around superficial dermal blood vessels to large numbers of tightly packed mast cells in the upper to mid-dermis (Fig. 25.19B). Fibrosis, edema, and eosinophils may also be present. Mast cells may be difficult to differentiate from lymphocytes in routine, hematoxylin and eosin–stained sections, and special metachromatic stains (toluidine blue or Giemsa) are used to visualize their granules (Fig. 25.19C). Even with these stains, extensive degranulation may result in failure to recognize these cells by light microscopy, but their identity can be readily confirmed with immunohistochemical stains for mast cell markers, such as mast cell tryptase and KIT.

DISORDERS OF EPIDERMAL MATURATION

Ichthyosis

Of the numerous disorders that impair epidermal maturation, ichthyosis is one of the most striking. The term is derived from the Greek root *ichthy,* meaning "fishy," and accordingly, this group of inherited disorders is associated with chronic, excessive keratin buildup (hyperkeratosis) that results in fish-like scaliness (Fig. 25.20A). Ichthyosis is subtyped according to the mode of inheritance, histology, and clinical features; the primary categories include *ichthyosis vulgaris* (autosomal dominant or acquired), *congenital ichthyosiform erythroderma* (autosomal recessive), *lamellar ichthyosis* (autosomal recessive), and *X-linked ichthyosis*. Most ichthyoses become apparent either at or around the time of birth. Acquired (noninherited) variants also exist; one such variant,

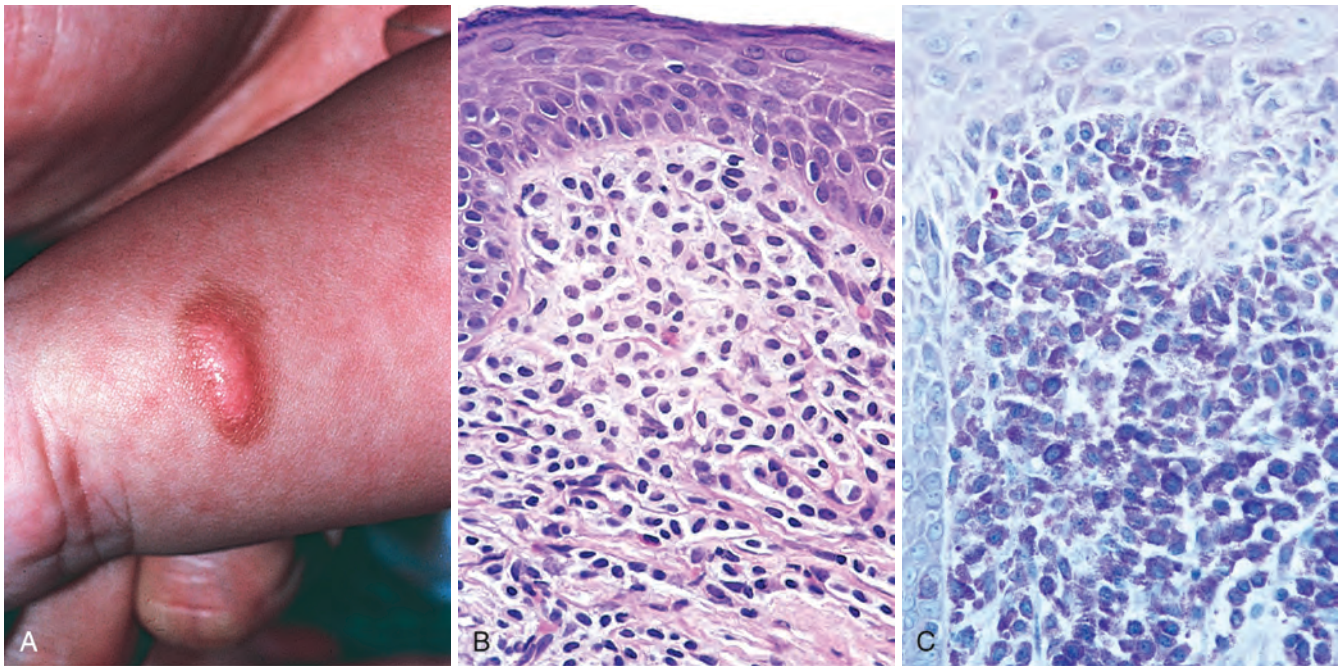


Figure 25.19 Mastocytosis. (A) Solitary mastocytoma in a 1-year-old child. (B) By histology, numerous ovoid cells with uniform, centrally located nuclei are observed in the dermis. (C) Giemsa staining reveals purple "metachromatic" granules within the cytoplasm of the mast cells.

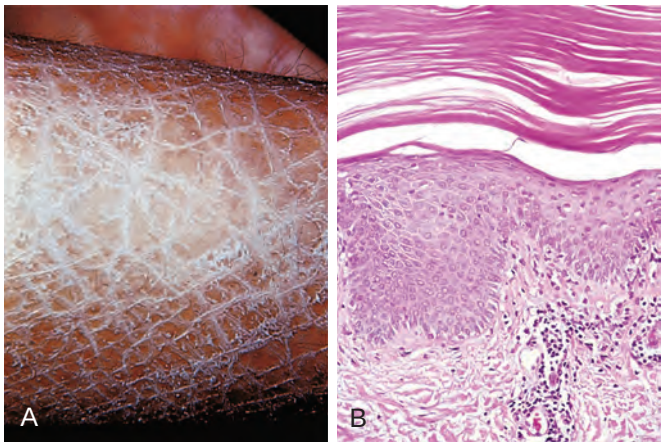


Figure 25.20 Ichthyosis. Note prominent fish-like scales (A) and compacted, thickened stratum corneum (B).

ichthyosis vulgaris, may be associated with lymphoid and visceral malignancies.

Pathogenesis

The primary abnormality in ichthyosis is defective desquamation, leading to retention of abnormally formed scales. For example, X-linked ichthyosis is caused by a deficiency of steroid sulfatase, an enzyme that helps to remove proadhesive cholesterol sulfate from intercellular spaces. In its absence, cholesterol sulfate accumulates, resulting in persistent cell-to-cell adhesion within the stratum corneum and a failure of desquamation.

MORPHOLOGY

All forms of ichthyosis exhibit a buildup of compacted stratum corneum that is associated with loss of the normal basket-weave pattern (Fig. 25.20B). There is generally little or no inflammation. Variations in the thickness of the epidermis and the stratum granulosum and the gross appearance and distribution of lesions are used to subclassify these disorders.

ACUTE INFLAMMATORY DERMATOSES

Innumerable inflammatory dermatoses have been described and can be broadly classified as acute and chronic. Acute lesions last from days to weeks and are characterized by inflammatory infiltrates (usually composed of lymphocytes and macrophages rather than neutrophils), edema, and variable degrees of epidermal, vascular, or subcutaneous injury. Chronic lesions, on the other hand, persist for months to years and are often associated with changes in epidermal growth (atrophy or hyperplasia) or dermal fibrosis. Discussed here are examples of the more commonly encountered acute dermatoses.

Urticaria

Urticaria (hives) is a common disorder of the skin that is usually caused by localized mast cell degranulation and is uniformly associated with dermal microvascular hyperpermeability. This combination of effects produces pruritic edematous plaques called *wheals.* Angioedema is closely

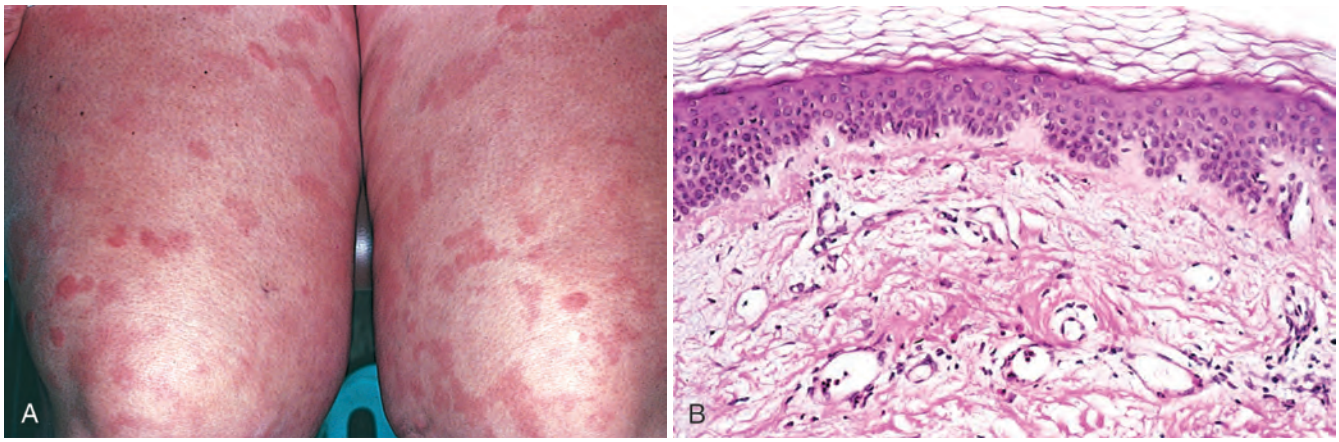


Figure 25.21 Urticaria. (A) Erythematous, edematous, often circular plaques are characteristic. (B) Histologically, there is superficial dermal edema, manifested by spaces between collagen bundles, and dilated lymphatic and blood-filled vascular spaces; the epithelium is normal.

related to urticaria and is characterized by edema of the deeper dermis and the subcutaneous fat.

Urticaria most often occurs between ages 20 and 40, but all age groups are susceptible. Individual lesions develop and fade within hours (usually less than 24 hours), and episodes may last for days or persist for months. Sites of predilection for urticarial eruptions include any area exposed to pressure, such as the trunk, distal extremities, and ears. Persistent episodes of urticaria may herald an underlying disease (e.g., collagen vascular disorders, Hodgkin lymphoma), but in the majority of cases no underlying cause is identified.

Pathogenesis

Urticaria is most commonly the result of antigen-induced release of vasoactive mediators from mast cells. The various forms can be classified based on their dependence on IgE antibody and mast cells, as follows:

- *Mast cell–dependent, immunoglobulin E (IgE)–dependent*. Urticaria of this type follows exposure to many different antigens (pollens, foods, drugs, insect venom) and is an example of a localized immediate hypersensitivity (type I) reaction triggered by the binding of antigen to IgE antibodies that are attached to mast cells through Fc receptors (Chapter 6).
- *Mast cell–dependent, IgE-independent*. This subset results from substances that directly incite the degranulation of mast cells, such as opiates, certain antibiotics, and radiographic contrast media.
- *Mast cell–independent, IgE-independent*. These forms of urticaria are triggered by local factors that increase vascular permeability. One form is initiated by exposure to chemicals or drugs, such as aspirin, that inhibit cyclooxygenase and arachidonic acid production. The precise mechanism of aspirin-induced urticaria is unknown. A second form is *hereditary angioneurotic edema* (Chapter 6), caused by an inherited deficiency of C1 inhibitor that results in excessive activation of the early components of the complement system and production of vasoactive mediators.

MORPHOLOGY

Lesions vary from small, pruritic papules to large edematous plaques (Fig. 25.21A). Individual lesions may coalesce to form annular, linear, or arciform configurations. The histologic features of urticaria may be very subtle. There is usually a sparse superficial perivenular infiltrate consisting of mononuclear cells and rare neutrophils. Eosinophils may also be present. Collagen bundles are more widely spaced than in normal skin, a result of dermal edema (Fig. 25.21B). Dermal lymphatic channels may also be dilated due to increased absorption of edema fluid. There are no changes in the epidermis.

Acute Eczematous Dermatitis

The Greek word *eczema,* meaning "to boil over," vividly describes the appearance of acute eczematous dermatitis—one of the most common skin disorders. Based on initiating factors, eczematous dermatitis can be subdivided into the following categories: (1) allergic contact dermatitis, (2) atopic dermatitis, (3) drug-related eczematous dermatitis, (4) photoeczematous dermatitis, and (5) primary irritant dermatitis.

The causes of eczema are sometimes broadly separated into "inside and outside jobs": disease resulting from external application of an antigen (e.g., poison ivy) or a reaction to an internal circulating antigen (which may be derived from ingested food or a drug). Treatment involves a search for offending substances that can be removed from the environment. Topical steroids can be used to block the inflammatory response. While such treatments are only palliative and do not cure, they are nevertheless helpful in interrupting acute exacerbations of eczema, which can become self-perpetuating if unchecked.

Pathogenesis

Eczematous dermatitis typically results from T cell–mediated inflammatory reactions (type IV hypersensitivity). This has been well studied in dermatitis triggered by contact

antigens (e.g., urushiol from poison ivy). It is believed that chemicals rubbed onto the skin react with self proteins, thereby acting as "haptens" that create neoantigens. These antigens are taken up by Langerhans cells, which then migrate through dermal lymphatics to draining lymph nodes. Here they present the neoantigens to naive CD4+ T cells, which are activated and develop into effector and memory cells (Chapter 6). On antigen reexposure, memory T cells expressing homing molecules, such as common lymphocyte antigen and particular chemokine receptors, migrate to cutaneous sites where antigen is localized. Here they release cytokines and chemokines, which serve to recruit numerous additional inflammatory cells. This process occurs within 24 hours and accounts for the initial erythema and pruritus that characterize the acute, spongiotic phase of eczema.

Langerhans cells within the epidermis play a central role in contact dermatitis, and understandably factors that affect Langerhans cell function impact the inflammatory reaction. For example, chronic exposure to UV light is injurious to epidermal Langerhans cells and can prevent sensitization to contact antigens, but UV light can also alter antigens and generate forms that are more likely to induce sensitivity reactions.

MORPHOLOGY

All types of eczematous dermatitis are characterized by red, papulovesicular, crusted lesions that, if persistent, develop reactive **acanthosis** and **hyperkeratosis** and take on the appearance of raised scaling plaques (Fig. 25.22). A striking example of eczema is an acute contact reaction to topical antigens such as urushiol in poison ivy/oak *(Rhus toxicodendron)*, which is characterized by pruritic, edematous, oozing plaques that often contain small and large blisters (vesicles and bullae) (Fig. 25.23A). Such lesions are prone to bacterial superinfection, which produces a yellow crust (impetiginization). With time, persistent lesions become less "wet" (fail to ooze or form vesicles) and become progressively more hyperkeratotic and acanthotic. **Spongiosis** characterizes acute eczematous dermatitis—hence the histologic synonym spongiotic dermatitis. Unlike urticaria, in which edema is restricted to the superficial dermis, edema seeps into the intercellular spaces of the epidermis, splaying apart keratinocytes, particularly in the stratum spinosum. Mechanical shearing of intercellular attachment sites (desmosomes) and cell membranes by progressive accumulation of intercellular fluid may result in the formation of intraepidermal vesicles (Fig. 25.23B).

During the earliest stages of eczematous dermatitis, there is a superficial, perivascular, lymphocytic infiltrate associated with papillary dermal edema and mast cell degranulation. The pattern and composition of this infiltrate may provide clues to the underlying cause. For example, eczema resulting from certain ingested drugs is marked by perivascular infiltrates that often contain eosinophils in the superficial and deep dermis. By contrast, eczematous dermatitis resulting from contact antigens tends to produce a mononuclear inflammatory reaction without eosinophils that preferentially affects the superficial dermis.

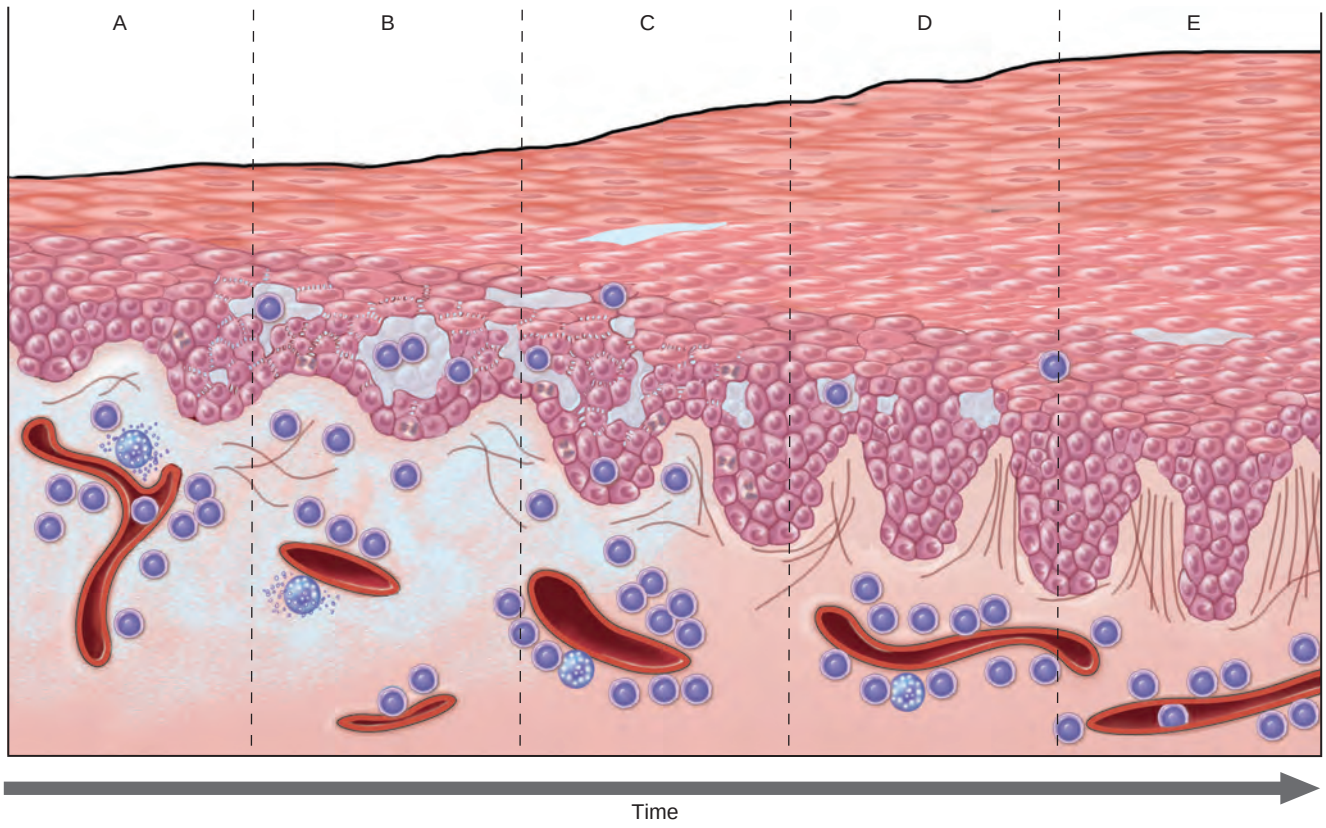


Figure 25.22 Stages of eczema development. (A) Initial dermal edema and perivascular infiltration by inflammatory cells is followed within 24 to 48 hours by (B) epidermal spongiosis and microvesicle formation. (C) Abnormal scale, including parakeratosis, as well as progressive acanthosis (D) and hyperkeratosis (E) appear as the lesion becomes chronic.

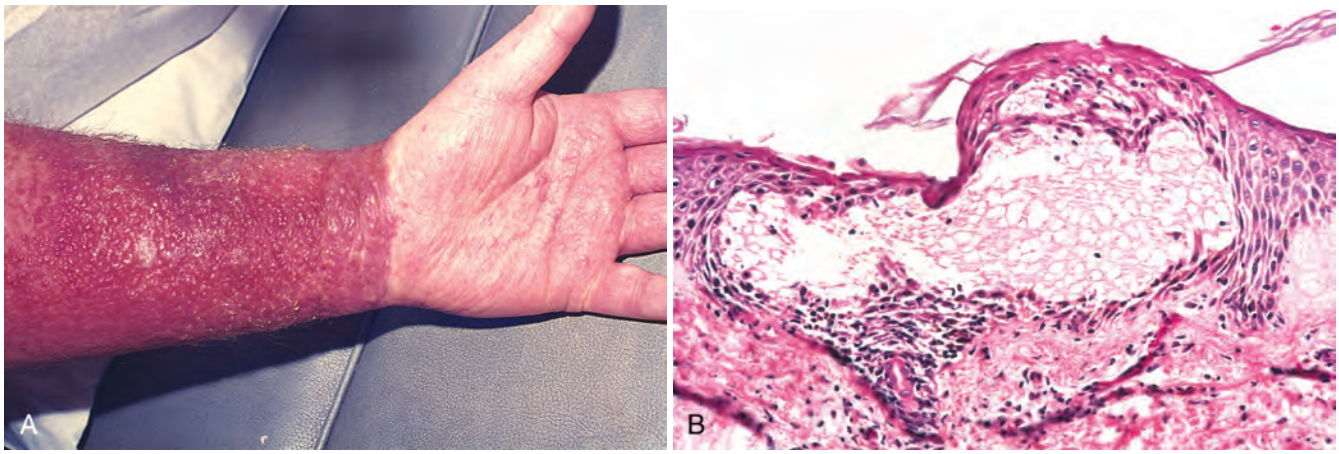


Figure 25.23 Eczematous dermatitis. (A) Acute allergic contact dermatitis due to antigen exposure (in this case, laundry detergent in clothing) marked by numerous vesicular lesions on erythematous skin. (B) Edema within the epidermis creates fluid-filled intraepidermal vesicles.

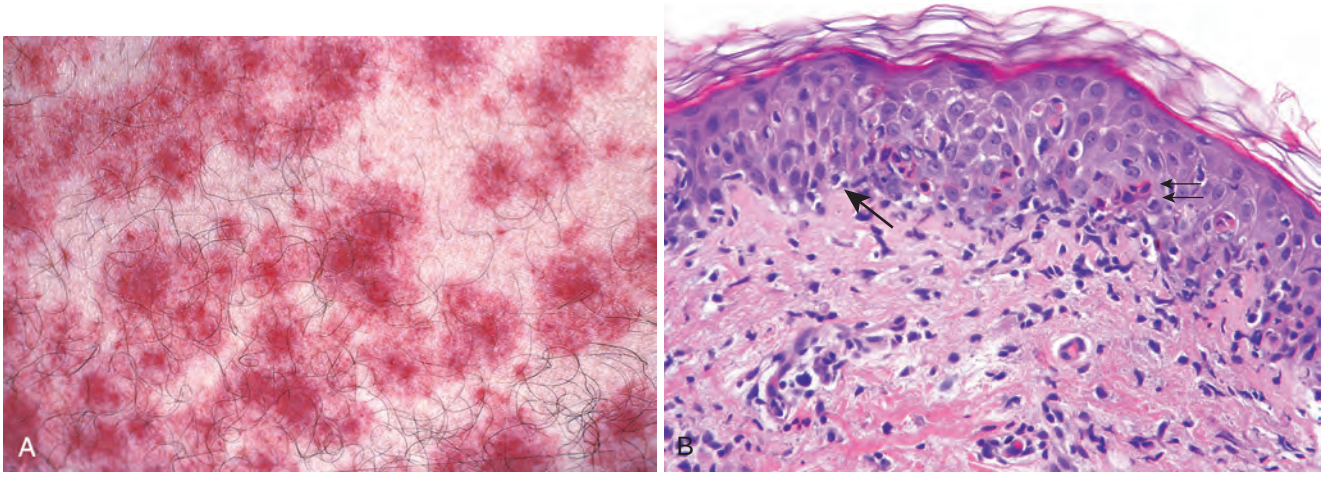


Figure 25.24 Erythema multiforme. (A) Target-like lesions consist of a central blister or zone of epidermal necrosis surrounded by macular erythema. (B) An early lesion shows lymphocytes accumulating along the dermoepidermal junction where basal keratinocytes are becoming vacuolated *(arrow)*. With time, necrotic/apoptotic keratinocytes appear in the overlying epithelium *(double arrow)*.

Erythema Multiforme

Erythema multiforme is an uncommon self-limited hypersensitivity reaction to certain infections and drugs. It affects individuals of any age and is associated with the following conditions: (1) infections such as herpes simplex, mycoplasmal infections, histoplasmosis, coccidioidomycosis, typhoid, and leprosy, among others; (2) exposure to certain drugs (sulfonamides, penicillin, barbiturates, salicylates, hydantoins, and antimalarials); (3) cancer (carcinomas and lymphomas); and (4) collagen vascular diseases (lupus erythematosus, dermatomyositis, and polyarteritis nodosa).

Pathogenesis

Erythema multiforme is characterized by keratinocyte injury mediated by skin-homing CD8+ cytotoxic T lymphocytes. This mechanism of injury is shared with a number of other conditions, including acute graft-versus-host disease, skin allograft rejection, and fixed drug eruptions. In erythema multiforme, CD8+ cytotoxic T cells are more prominent in the central portion of the lesions, while CD4+ helper T cells and Langerhans cells are more prevalent in the peripheral portions. The epidermal antigens that are recognized by the infiltrating T cells in erythema multiforme remain unknown.

MORPHOLOGY

Affected individuals present with a diverse array of lesions (hence the term multiforme), including macules, papules, vesicles, bullae, and characteristic targetoid (target-like) lesions (Fig. 25.24A). The lesions may occur in a variety of distributions. Cases that are limited in extent often show symmetric involvement of the extremities. A febrile form associated with extensive involvement of the skin is called **Stevens-Johnson syndrome,** which is often (but not exclusively) seen in children. In Stevens-Johnson syndrome, lesions involve not only the skin but also the lips and oral mucosa, conjunctiva, urethra, and genital and perianal areas. Secondary infection of involved areas due to loss of skin integrity may result in life-threatening sepsis. Another variant termed **toxic epidermal necrolysis** is characterized by diffuse necrosis and sloughing of

cutaneous and mucosal epithelial surfaces. The widespread epidermal damage produces a clinical picture similar to that seen in patients with extensive burns.

On histologic examination, the "targetoid" lesions show a superficial perivascular, lymphocytic infiltrate associated with dermal edema and accumulation of lymphocytes along the dermoepidermal junction, where they are intimately associated with degenerating and necrotic keratinocytes, a pattern termed *interface dermatitis* (Fig. 25.24B). With time there is upward migration of lymphocytes into the epidermis. Discrete and confluent zones of epidermal necrosis occur with concomitant blister formation. Epidermal sloughing leads to shallow erosions.

CHRONIC INFLAMMATORY DERMATOSES

This category includes inflammatory skin disorders that persist for months to years. The skin surface in chronic inflammatory dermatoses is often roughened as a result of excessive or abnormal scale formation and shedding.

Psoriasis

Psoriasis is a chronic inflammatory dermatosis that appears to have an autoimmune basis. It is a common disorder, affecting as many as 1% to 2% of people in the United States. Persons of all ages develop the disease. Approximately 15% of patients with psoriasis have associated arthritis, which may be mild or may produce severe deformities resembling the joint changes seen in rheumatoid arthritis. It can affect any joint in the body and may be symmetric or asymmetric. In addition, psoriasis may be associated with myopathy, enteropathy, and acquired immunodeficiency syndrome (AIDS).

Pathogenesis

As with other suspected disorders of autoimmunity, psoriasis is believed to be the product of environmental and genetic factors, including particular HLA gene variants. The culprit antigens remain elusive, but it appears that sensitized populations of CD4+ Th1 and Th17 cells and activated CD8+ cytotoxic effector T cells enter the skin and accumulate in the epidermis. These T cells may create an abnormal microenvironment by stimulating the secretion of cytokines and growth factors that induce keratinocyte proliferation, resulting in the characteristic lesions. The interactions between CD4+ T cells, CD8+ T cells, dendritic cells, and keratinocytes give rise to a cytokine "soup" dominated by Th1-type and Th17-type cytokines such as IL-12, interferon-γ, tumor necrosis factor (TNF), and IL-17. The importance of these factors is highlighted by the generally excellent clinical responses that are observed in patients treated with TNF inhibitors. Lymphocytes also produce growth factors for keratinocytes that may contribute to epidermal thickening. Psoriatic lesions can be induced in susceptible individuals by local trauma, a process known as the Koebner phenomenon, presumably because trauma sets in motion a local inflammatory response that becomes self-perpetuating.

MORPHOLOGY

Psoriasis most frequently affects the skin of the elbows, knees, scalp, lumbosacral areas, intergluteal cleft, and glans penis. The typical lesion is a well-demarcated, **pink to salmon-colored plaque covered by loosely adherent silver-white scale** (Fig. 25.25A). Variations exist, with some lesions occurring in annular, linear, gyrate, or serpiginous configurations. Psoriasis is one cause of total body erythema and scaling known as erythroderma. **Nail changes** occur in 30% of cases of psoriasis and consist of yellow-brown discoloration (often likened to an oil slick), with pitting, dimpling, separation of the nail plate from the underlying bed (onycholysis), thickening, and crumbling.

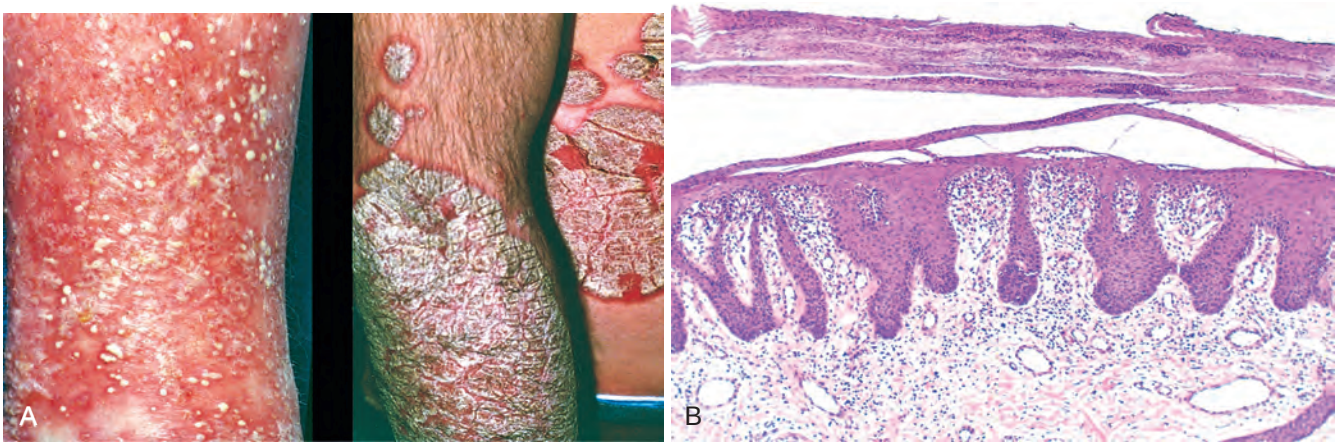


Figure 25.25 Psoriasis. (A) Early lesions may be dominated by inflammation, marked by the presence of small pustules and erythema *(left)*. Established chronic lesions are erythematous and covered by a characteristic silver-white scale *(right)*. (B) Microscopically, there is epidermal hyperplasia, parakeratotic scale, and accumulation of neutrophils within the superficial epidermis.

Established lesions of psoriasis have a characteristic histologic picture. Increased epidermal cell proliferation results in marked epidermal thickening (acanthosis), with regular downward elongation of the rete ridges sometimes described as appearing like test tubes in a rack (Fig. 25.25B). Mitotic figures are easily identified well above the basal cell layer, where mitotic activity is confined in normal skin. **The stratum granulosum is thinned or absent, and extensive overlying parakeratotic scale is seen.** Typical of psoriatic plaques is thinning of the portion of the epidermal cell layer that overlies the tips of the dermal papillae (suprapapillary plates), which contain dilated, tortuous blood vessels. This abnormal proximity of vessels within the dermal papillae to the overlying parakeratotic scale accounts for the characteristic clinical phenomenon of multiple, minute, bleeding points when the scale is lifted from the plaque **(Auspitz sign).** Neutrophils form small aggregates within the superficial epidermis **(spongiform pustules)** and the parakeratotic stratum corneum **(Munro microabscesses).** In pustular psoriasis, larger abscess-like accumulations of neutrophils are present directly beneath the stratum corneum.

Seborrheic Dermatitis

Seborrheic dermatitis is a chronic inflammatory dermatosis that is even more common than psoriasis, affecting up to 5% of the general population. It classically involves regions with a high density of sebaceous glands, such as the scalp, forehead (especially the glabella), external auditory canal, retroauricular area, nasolabial folds, and presternal area. Despite this association and its name, seborrheic dermatitis is associated with inflammation of the epidermis and is not a disease of the sebaceous glands per se.

Pathogenesis

The precise etiology of seborrheic dermatitis is unknown. Increased sebum production, often in response to androgens, is one possible contributory factor. Involvement of sebum is supported by clinical observations of patients with Parkinson disease, who typically show increased sebum production secondary to dopamine deficiency and have a markedly increased incidence of seborrheic dermatitis. Once treated with levodopa, the oiliness of the skin decreases and the seborrheic dermatitis improves. However, other conditions associated with increased sebum production such as acne (discussed later) are not associated with seborrheic dermatitis, and sebum production is probably best viewed as being necessary but not sufficient to cause the disorder. Other work has suggested a relationship with colonization of the skin by certain fungal species of the genus *Malassezia,* but there is no definitive evidence of a cause-and-effect relationship. A severe form of seborrheic dermatitis that is difficult to treat was once common in many human immunodeficiency virus (HIV)–infected individuals with low CD4 counts, but its incidence has fallen with the advent of effective antiviral therapy.

MORPHOLOGY

The individual lesions are macules and papules on an erythematous-yellow, often greasy base, typically in association with extensive scaling and crusting. Fissures may also be present, particularly behind the ears. Dandruff is the common clinical expression of seborrheic dermatitis of the scalp. Microscopically, seborrheic dermatitis shares features with spongiotic dermatitis and psoriasis, with earlier lesions being more spongiotic and later ones more acanthotic. Typically, mounds of parakeratosis containing neutrophils and serum are present at the ostia of hair follicles (so-called **follicular lipping**). A superficial perivascular inflammatory infiltrate generally consists of a mixture of lymphocytes and neutrophils.

Lichen Planus

"Pruritic, purple, polygonal, planar, papules, and plaques" are the tongue-twisting "six Ps" of lichen planus, a disorder of skin and mucosa. Lichen planus is usually self-limited, most commonly resolving spontaneously 1 to 2 years after onset. Resolution often leaves a residuum of postinflammatory hyperpigmentation. Oral lesions, however, may persist for years. Squamous cell carcinoma sometime arises from chronic mucosal and paramucosal lichen planus and may be an example of carcinogenesis fostered by chronic inflammation. As in psoriasis, the Koebner phenomenon may be seen in lichen planus.

Pathogenesis

The pathogenesis of lichen planus is not known. It is possible that expression of altered antigens in basal epidermal cells or the dermoepidermal junction elicits a cell-mediated cytotoxic (CD8+) T-cell response.

MORPHOLOGY

Cutaneous lesions consist of itchy, violaceous, flat-topped papules that may coalesce to form plaques (Fig. 25.26A). The papules are often highlighted by white dots or lines called **Wickham striae,** which are created by areas of hypergranulosis. In darkly pigmented individuals, lesions may acquire a dark brown color due to release of melanin into the dermis as the basal cell layer is destroyed. Lesions are usually multiple and symmetrically distributed, particularly on the extremities and often about the wrists and elbows. The glans penis is another common site of involvement. In 70% of cases, oral mucosal lesions are present that have a white, reticulated, or net-like appearance.

Lichen planus is characterized histologically by a dense, continuous infiltrate of lymphocytes along the dermoepidermal junction, a prototypic example of **interface dermatitis** (Fig. 25.26B). The lymphocytes are intimately associated with basal keratinocytes, which show degeneration, necrosis, and a resemblance in size and contour to more mature cells of the stratum spinosum (squamatization). As a consequence of this destructive lymphocytic infiltrate, the dermoepidermal interface takes on an angulated zigzag contour (sawtoothing). Anucleate, necrotic basal cells may become incorporated into the inflamed papillary dermis, where they are referred to as **colloid** or **Civatte bodies.** Although the lesions bear some similarities to those in erythema multiforme, lichen planus shows changes of chronicity, namely, epidermal hyperplasia (or rarely atrophy) and thickening of the granular cell layer and stratum corneum (hypergranulosis and hyperkeratosis, respectively).

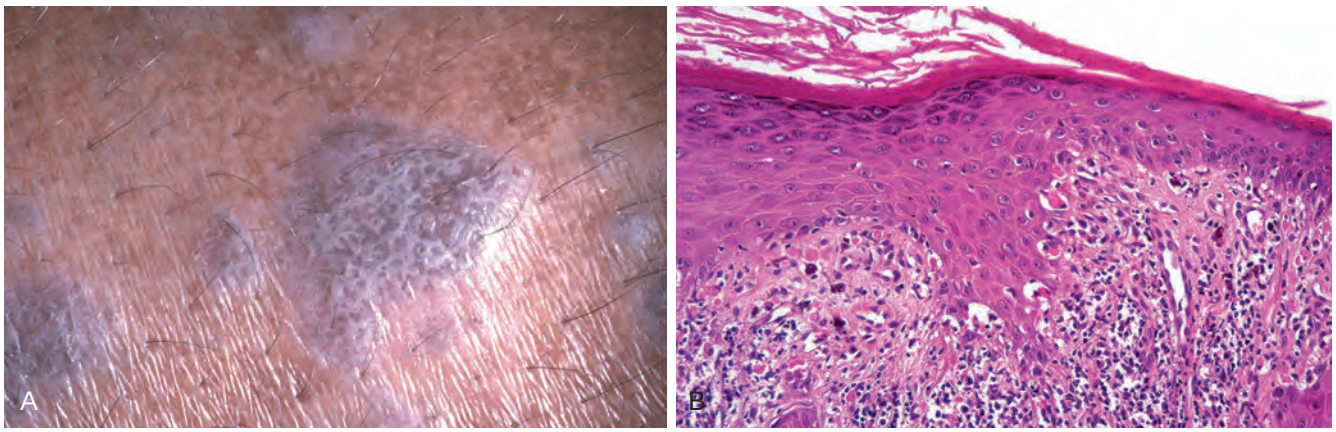


Figure 25.26 Lichen planus. (A) This flat-topped pink-purple, polygonal papule has a white lace-like pattern of lines that are referred to as Wickham striae. (B) There is a band-like infiltrate of lymphocytes at the dermoepidermal junction, hyperkeratosis, and pointed rete ridges (sawtoothing), the last as a result of chronic basal cell layer injury.

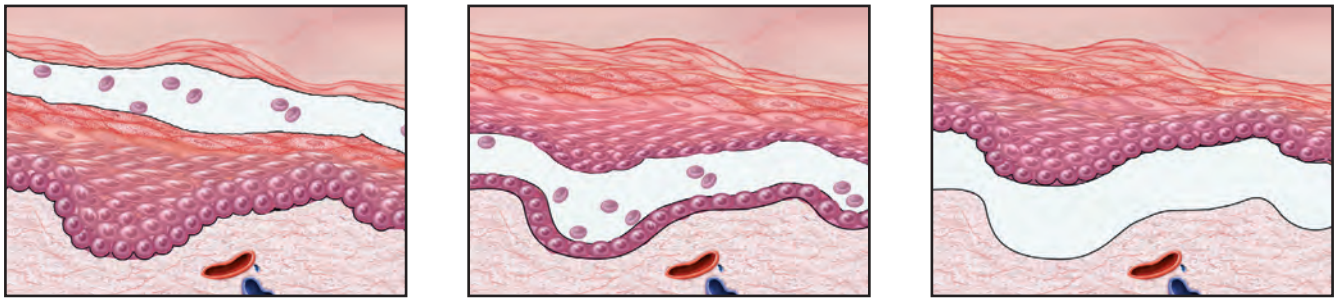


Figure 25.27 Schematic representation of different types of blisters. (A) In subcorneal blisters, the stratum corneum forms the roof of the bulla (as in pemphigus foliaceus). (B) In a suprabasilar blister, a portion of the epidermis, including the stratum corneum, forms the roof (as in pemphigus vulgaris). (C) In a subepidermal blister, the entire epidermis separates from the dermis (as in bullous pemphigoid).

KEY CONCEPTS

INFLAMMATORY DERMATOSES

- Many specific inflammatory dermatoses exist, which can be mediated by IgE antibodies (urticaria) or antigen-specific T cells (eczema, erythema multiforme, and psoriasis).
- These disorders are diagnosed based on the distribution and gross appearance of skin lesions and the microscopic patterns of inflammation (e.g., interface dermatitis in lichen planus and erythema multiforme).

BLISTERING (BULLOUS) DISEASES

Although vesicles and bullae (blisters) occur in several unrelated conditions such as herpesvirus infection, spongiotic dermatitis, erythema multiforme, and thermal burns, there exists a group of disorders in which blisters are the primary and most distinctive features. These *bullous diseases,* as they are called, produce dramatic lesions and in some instances are fatal if untreated. Blisters in the various disorders occur at different levels within the skin (Fig. 25.27); histologic assessment is essential for accurate diagnosis and provides insight into the pathogenic mechanisms. Knowledge of the structure of desmosomes and hemidesmosomes (described in Chapter 1), which, you will recall, provide the skin with mechanical stability, is helpful in understanding these diseases, as they are often caused by acquired or inherited defects in proteins that make up or bind to these structures (Fig. 25.28).

Inflammatory Blistering Disorders

Pemphigus

Pemphigus is a blistering disorder caused by autoantibodies that result in the dissolution of intercellular attachments within the epidermis and mucosal epithelium. The pathobiology of blistering disorders provides important insights into the molecular underpinnings of keratinocyte adhesion. The majority of individuals who develop pemphigus are in the fourth to sixth decades of life, and men and women are affected equally. There are multiple variants: (1) pemphigus vulgaris, (2) pemphigus vegetans, (3) pemphigus foliaceus, (4) pemphigus erythematosus, and (5) paraneoplastic pemphigus. These disorders are usually benign, but in extreme cases can be fatal without treatment.

- *Pemphigus vulgaris,* by far the most common type (accounting for more than 80% of cases worldwide), involves the

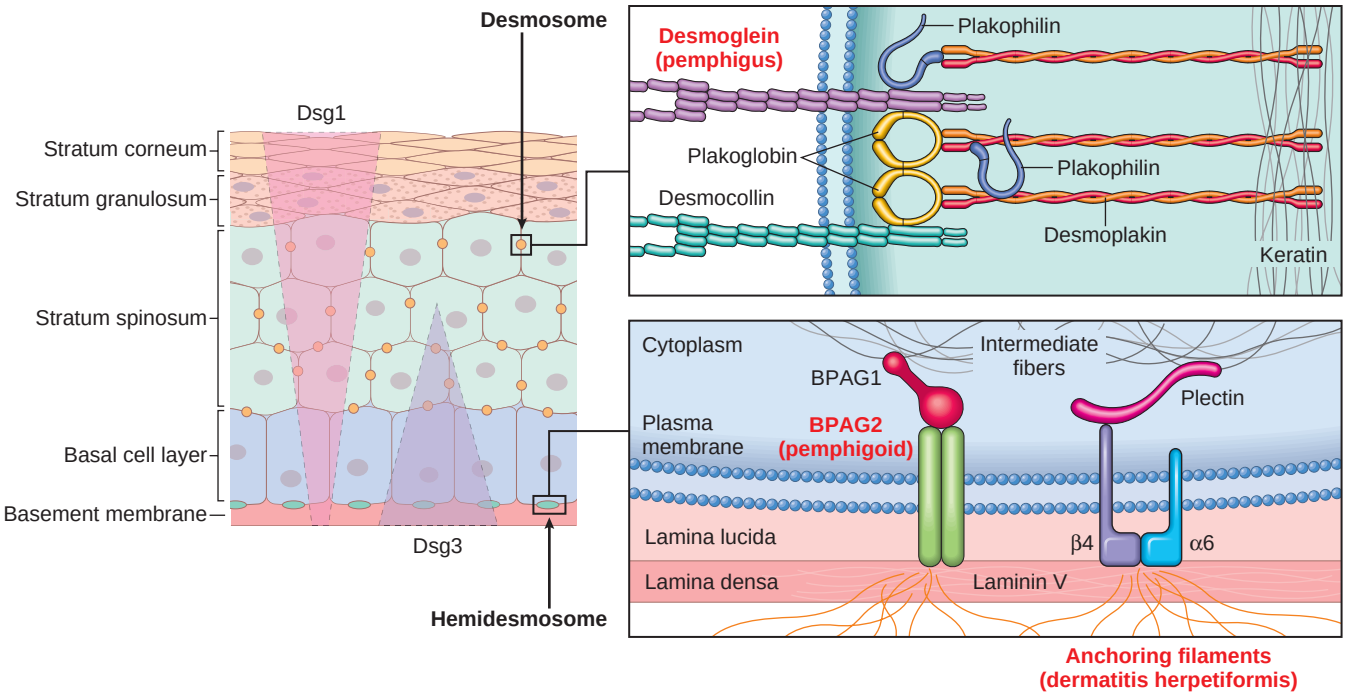


Figure 25.28 Keratinocyte adhesion molecules and blistering inflammatory disorders. Knowledge of the proteins composing desmosomes and hemidesmosomes is key to understanding blistering disorders. Desmogleins 1 and 3 *(Dsg1, Dsg3)* are functionally interchangeable components of desmosomes but have different distributions within the epidermis *(left panel)*. The major structural proteins of desmosomes and hemidesmosomes are shown in the *right panel*. In pemphigus vulgaris, autoantibodies against Dsg1 and Dsg3 cause blisters in the deep suprabasal epidermis, whereas in pemphigus foliaceus, autoantibodies are against Dsg1 alone, leading to superficial, subcorneal blisters. In bullous pemphigoid, autoantibodies bind BPAG2, a component of the hemidesmosomes, leading to blister formation at the level of the lamina lucida of the basement membrane. Dermatitis herpetiformis is caused by IgA autoantibodies to the fibrils that anchor hemidesmosomes to the dermis.

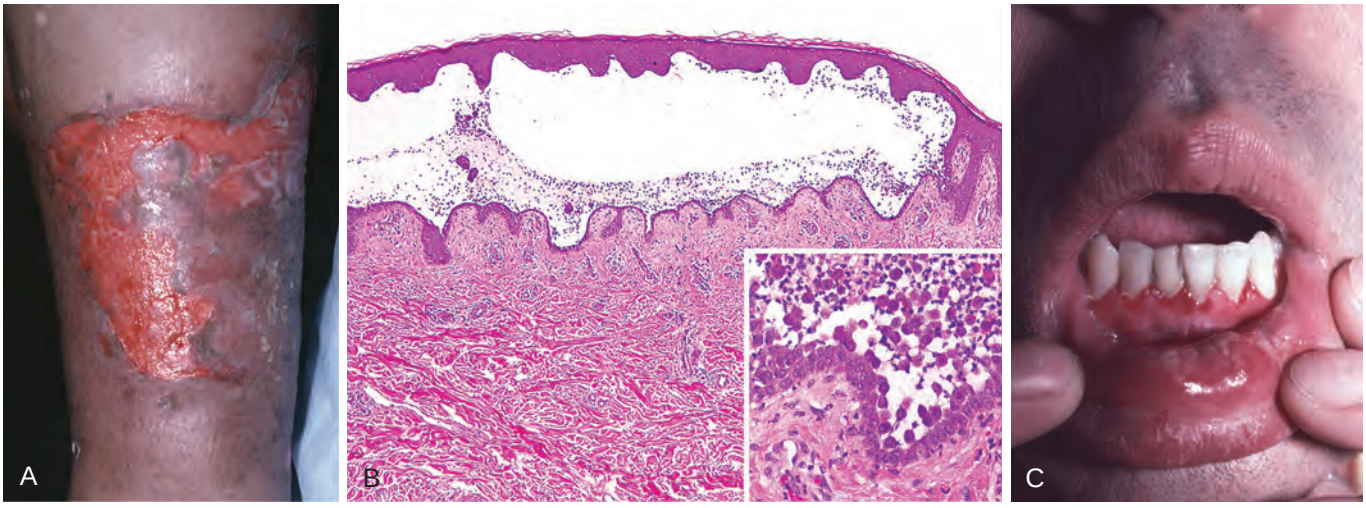


Figure 25.29 Pemphigus vulgaris. (A) Eroded plaques are formed following the rupture of confluent, thin-roofed bullae. (B) Suprabasal acantholysis results in an intraepidermal blister in which dyscohesive (acantholytic) epidermal cells are present *(inset)*. (C) Ulcerated blisters in the oral mucosa are also common, as seen here on the lip.

mucosa and skin, especially on the scalp, face, axilla, groin, trunk, and points of pressure. It may present as oral ulcers that may persist for months before skin involvement appears. Primary lesions are superficial vesicles and bullae that rupture easily, leaving shallow erosions covered with dried serum and crust (Fig. 25.29A).

- *Pemphigus vegetans* is a rare form that usually presents not with blisters but with large, moist, verrucous (wart-like), vegetating plaques studded with pustules on the groin, axillae, and flexural surfaces.
- *Pemphigus foliaceus* is a more benign form that is endemic in Brazil (where it is called *fogo selvagem*) and occurs sporadically in other geographic regions. Lesions are most common on the scalp, face, chest, and back; the mucous membranes are only rarely affected. Bullae are so superficial that they mainly present as areas of

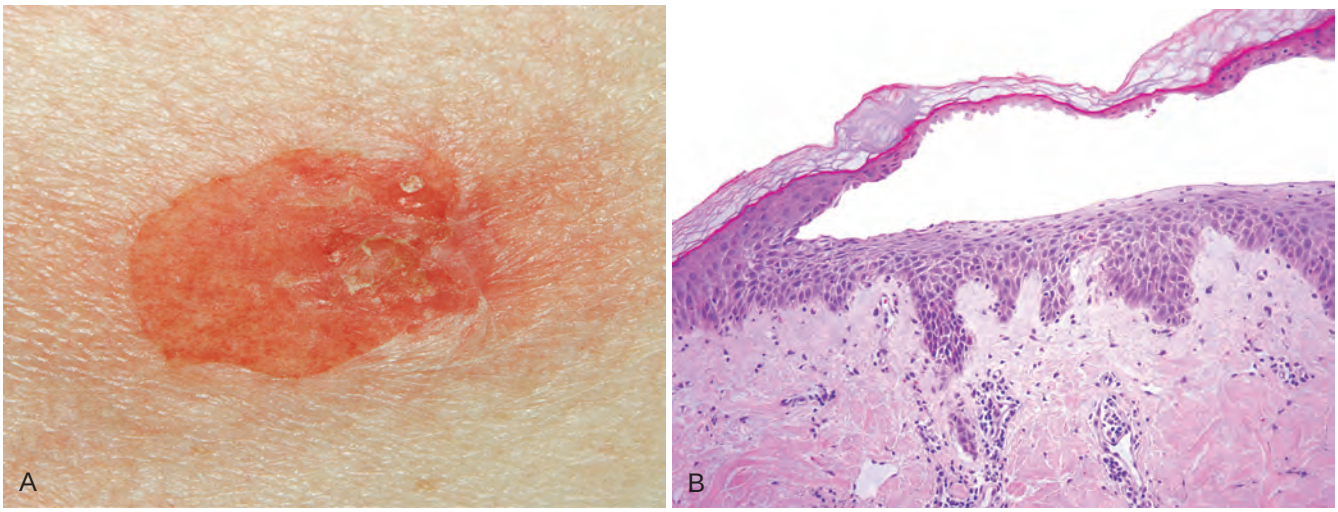


Figure 25.30 Pemphigus foliaceus. (A) The delicate superficial (subcorneal) blisters are much less erosive than those seen in pemphigus vulgaris. (B) Subcorneal separation of the epithelium is seen.

erythema and crusting; these represent superficial erosions at sites of blister rupture (Fig. 25.30A).

- *Pemphigus erythematosus* is considered to be a localized, less severe form of pemphigus foliaceus that may selectively involve the malar area of the face in a lupus erythematosus–like fashion.
- *Paraneoplastic pemphigus* occurs in association with various malignancies, most commonly non-Hodgkin lymphoma.

Pathogenesis

All forms of pemphigus are autoimmune diseases caused by IgG autoantibodies against desmogleins that disrupt intercellular adhesions and result in the formation of blisters. By direct immunofluorescence, lesions show a characteristic net-like pattern of intercellular IgG deposits. IgG is usually seen at all levels of the epithelium in pemphigus vulgaris, but tends to be more superficial in pemphigus foliaceus (Fig. 25.31). The distribution of desmoglein 1 and 3 in the epidermis and whether there are autoantibodies to one or both proteins appear to explain the position and severity of the blisters (see Fig. 25.28). The antibodies cause these lesions primarily by disrupting the intercellular adhesive function of the desmosomes; they may also act indirectly by activating intercellular proteases. Paraneoplastic pemphigus arises most often in the setting of lymphoid neoplasms and is also caused by autoantibodies that recognize desmogleins or other proteins involved in intercellular adhesion.

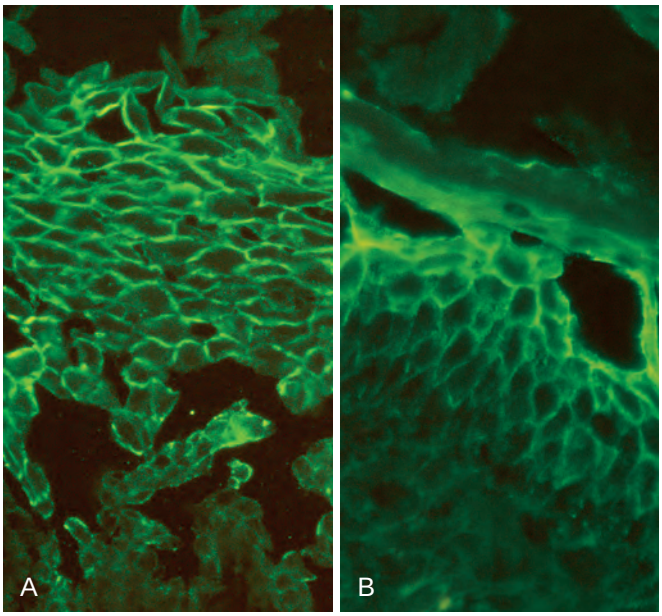


Figure 25.31 Direct immunofluorescence staining for immunoglobulin of epidermis involved by pemphigus. (A) In pemphigus vulgaris there is deposition of immunoglobulin along the plasma membranes of keratinocytes in a reticular or fishnet-like pattern accompanied by suprabasalar loss of cell-to-cell adhesion (acantholysis). (B) In pemphigus foliaceus the immunoglobulin deposits and acantholysis are more superficial.

MORPHOLOGY

The common histologic feature in all forms of pemphigus is **acantholysis,** the dissolution of the intercellular bridges that connect squamous epithelial cells. Acantholytic cells dissociate from one another, lose their polyhedral shape, and become rounded. In pemphigus vulgaris (see Fig. 25.29B) and pemphigus vegetans, acantholysis selectively involves the cells immediately above the basal cell layer, producing a suprabasal blister. In the vegetans variant, there is also overlying epidermal hyperplasia. The single layer of intact basal cells that forms the blister base in pemphigus vulgarus has been likened to a row of tombstones. In pemphigus foliaceus, blisters form by similar mechanisms but are found in the superficial epidermis at the level of the stratum granulosum (see Fig. 25.30B). Variable superficial dermal infiltration by lymphocytes, macrophages, and eosinophils accompanies each type of pemphigus.

The mainstay of treatment in all forms of pemphigus is immunosuppressive agents, which decrease the titers of the pathogenic antibodies.

Bullous Pemphigoid

Generally affecting elderly individuals, bullous pemphigoid shows a wide range of clinical presentations. Sites of involvement include the inner aspects of the thighs, flexor surfaces of the forearms, axillae, groin, and lower abdomen. Oral lesions are present in 10% to 15% of affected individuals, usually appearing after the cutaneous lesions. Some patients present with urticarial plaques and severe pruritus.

MORPHOLOGY

The lesions are tense bullae filled with clear fluid involving erythematous or normal-appearing skin (Fig. 25.32A). The bullae are usually less than 2 cm in diameter but occasionally may reach 4 to 8 cm in diameter. They do not rupture easily, unlike the blisters seen in pemphigus, and heal without scarring unless they become infected secondarily. The separation of bullous pemphigoid from pemphigus is based on the identification of **subepidermal, nonacantholytic** blisters. Early lesions show a superficial and sometimes deep perivascular infiltrate of lymphocytes and variable numbers of eosinophils, occasional neutrophils, superficial dermal edema, and basal cell layer vacuolization (Fig. 25.32B). Eosinophils are typically present directly beneath the epidermal basal cell layer. The vacuolated basal cell layer eventually lifts away, allowing space for a fluid-filled blister to form.

Pathogenesis

Bullous pemphigoid is caused by autoantibodies that bind to proteins that are required for adherence of basal keratinocytes to the basement membrane. Most antibody deposition occurs in a continuous linear pattern at the dermoepidermal junction (Fig. 25.33A), where specialized structures called hemidesmosomes link basal keratinocytes to the underlying basement membrane (Fig. 25.33B). The so-called bullous pemphigoid antigens (BPAGs) are components of hemidesmosomes

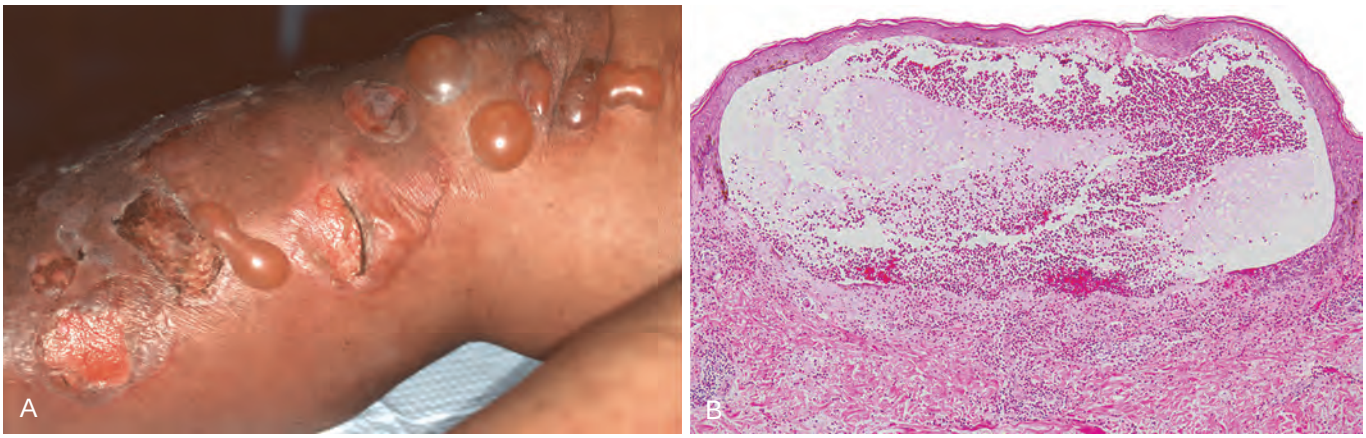


Figure 25.32 Bullous pemphigoid. (A) Bullae consist of tense blisters that usually fail to rupture, as their roof consists of intact epidermis. Ulcers form if the blisters rupture. (B) Intact subepidermal blister associated with eosinophils, lymphocytes, and occasional neutrophils.

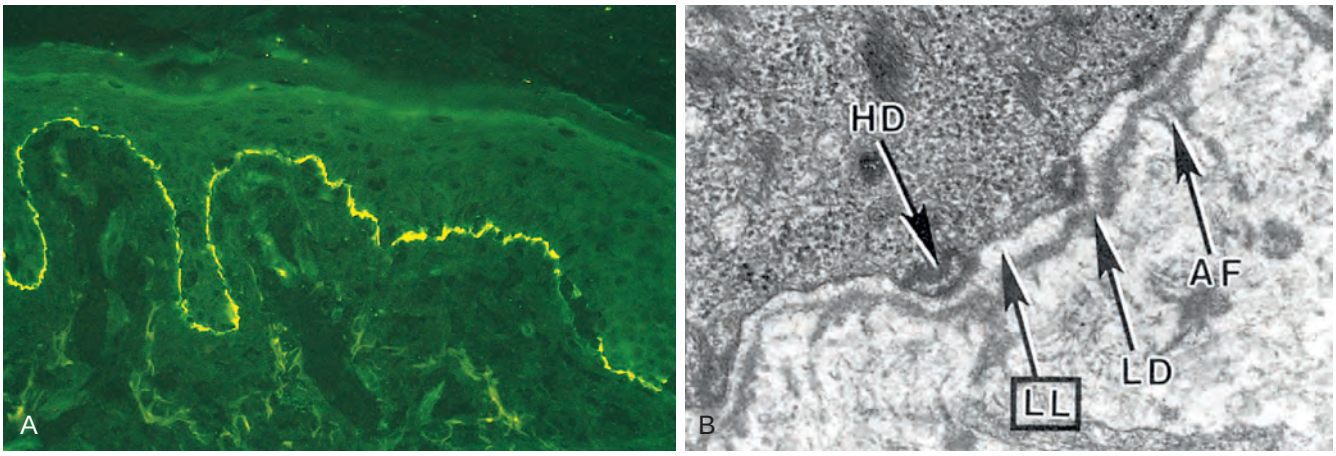


Figure 25.33 (A) Linear deposition of complement along the dermoepidermal junction in bullous pemphigoid. (B) Electron micrograph showing the ultrastructural features of the dermoepidermal junction. The bullous pemphigoid antigen (BPAG) is located in the basal portion of basal keratinocytes in association with hemidesmosomes *(HD)*, which attach the epidermis to the lamina lucida *(LL)* of the basement membrane. *AF,* Anchoring fibrils; *LD,* lamina densa. (See also Fig. 25.31.)

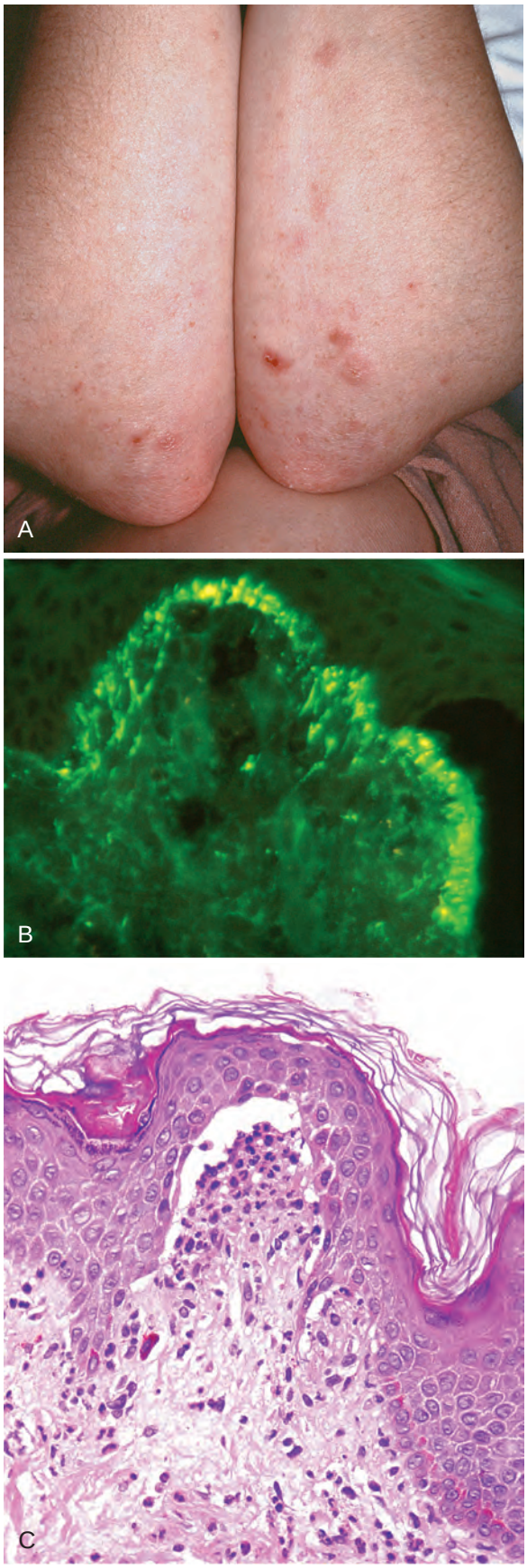


Figure 25.34 Dermatitis herpetiformis. (A) Lesions consist of groups of intact and eroded erythematous blisters (seen here on elbows and arms). (B) Selective deposition of IgA autoantibody at the tips of dermal papillae is characteristic. (C) The blisters are associated with the accumulation of neutrophils (microabscesses) at the tips of dermal papillae. (B, Courtesy Dr. Victor G. Prieto, Houston, Tex.)

(see Fig. 25.29). Antibodies against one such component called BPAG2 are proven to cause blistering. Pathogenic autoantibodies also activate complement, leading to the recruitment of neutrophils and eosinophils, inflammation, and disruption of epidermal attachments.

Dermatitis Herpetiformis

Dermatitis herpetiformis is a rare disorder characterized by urticaria and grouped vesicles. The disease affects predominantly males, most often in the third and fourth decades of life. In some cases it occurs in association with intestinal celiac disease (Chapter 17). The plaques and vesicles are extremely pruritic.

Pathogenesis

The association of dermatitis herpetiformis with celiac disease provides a clue to its pathogenesis. Genetically predisposed individuals develop IgA antibodies to dietary gluten (derived from the wheat protein gliadin). The antibodies cross-react with reticulin, a component of the anchoring fibrils that tether the epidermal basement membrane to the superficial dermis. The resultant injury and inflammation produce a subepidermal blister. In some people with dermatitis herpetiformis and gluten-sensitive enteropathy, both disorders respond to a gluten-free diet.

MORPHOLOGY

The lesions are bilateral, symmetric, and grouped, involving preferentially the extensor surfaces, elbows, knees, upper back, and buttocks (Fig. 25.34C). Fibrin and neutrophils accumulate selectively at the **tips of dermal papillae,** forming small microabscesses (Fig. 25.34A). The basal cells overlying these microabscesses show vacuolization and focal dermoepidermal separation that ultimately coalesce to form a true **subepidermal blister.** By direct immunofluorescence, dermatitis herpetiformis shows discontinuous, **granular deposits of IgA** that selectively localize in the tips of dermal papillae (Fig. 25.34B).

Noninflammatory Blistering Disorders

Epidermolysis Bullosa and Porphyria

Some disorders characterized by vesicles and bullae are mediated by inherited or, in some cases, acquired defects involving structural proteins that maintain the integrity of the skin. Two such disorders are epidermolysis bullosa and porphyria.

Epidermolysis Bullosa. Epidermolysis bullosa is a blanket term for a group of disorders caused by inherited defects

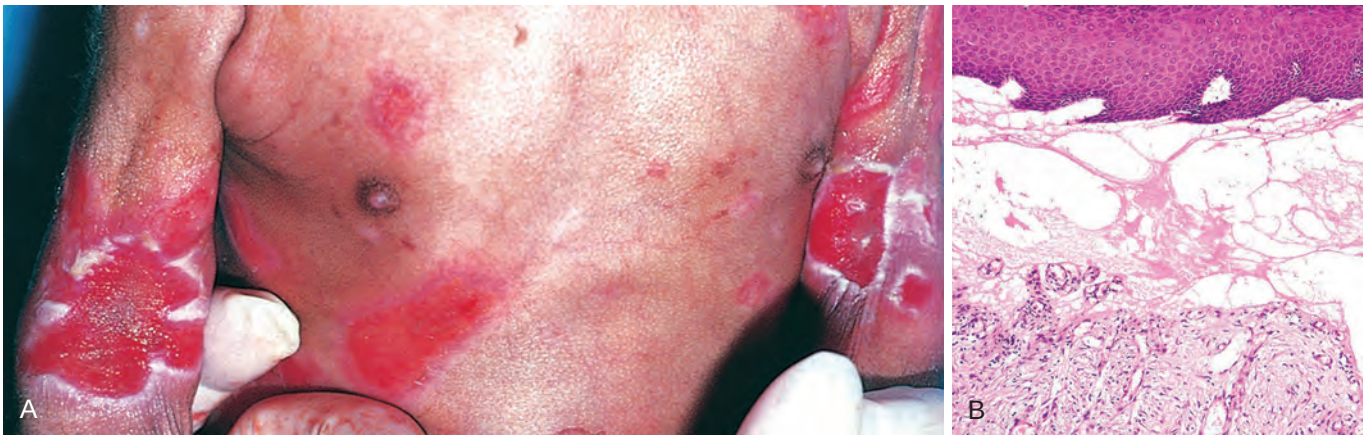


Figure 25.35 Epidermolysis bullosa. (A) Junctional epidermolysis bullosa showing typical erosions in flexural creases. (B) Subepidermal blister at the level of the lamina lucida. There is no associated inflammation.

in structural proteins that lend mechanical stability to the skin. The common feature is a proclivity to form blisters at sites of pressure, rubbing, or trauma, at or soon after birth. The histologic changes in all forms are so subtle that electron microscopy may be required to differentiate among the various types.

- In the *simplex type,* defects of the basal cell layer of the epidermis almost always result from mutations in the genes encoding keratin 14 or keratin 5. These two proteins normally pair with one another to make a functional keratin fiber, thus explaining the similar phenotype resulting from mutations in either gene. The mutated proteins have a dominant negative activity, and as a result the disorder shows an autosomal dominant mode of inheritance. This is the most common type of epidermolysis bullosa, encompassing 75% to 85% of cases.
- In the *junctional type,* blisters occur in otherwise histologically normal skin at precisely the level of the lamina lucida (Fig. 25.35; see Fig. 25.28). Most cases are caused by autosomal recessive defects in one of the subunits of laminin, a multicomponent protein located in the lamina lucida that binds to both hemidesmosomes and anchoring filaments. Some of the remaining cases are caused by mutations in BPAG2, the same protein that is targeted by autoantibodies in bullous pemphigoid.
- In the scarring *dystrophic types,* blisters develop beneath the lamina densa in association with rudimentary or defective anchoring fibrils. Dystrophic epidermolysis bullosa usually results from mutations in the *COL7A1* gene, which encodes type VII collagen (Chapter 3), a major component of the basement membrane anchoring fibrils. Depending on the mutation, the disorder may follow an autosomal dominant or autosomal recessive mode of inheritance.
- *Mixed types,* marked by defects at several levels, are also recognized.

Porphyria. Porphyria refers to a group of uncommon inborn or acquired disturbances of porphyrin metabolism. Porphyrins are pigments that are normally present in hemoglobin, myoglobin, and cytochromes. The classification of porphyria is based on both clinical and biochemical features. The five major types are (1) congenital erythropoietic porphyria, (2) erythrohepatic protoporphyria, (3) acute intermittent porphyria, (4) porphyria cutanea tarda, and (5) mixed porphyria. Cutaneous manifestations consist of urticaria and vesicles associated with scarring that are exacerbated by exposure to sunlight. The vesicles are subepidermal in location, and the adjacent dermis contains vessels with walls that are thickened by glassy deposits of serum proteins, including immunoglobulins (Fig. 25.36). The pathogenesis of these alterations is not understood.

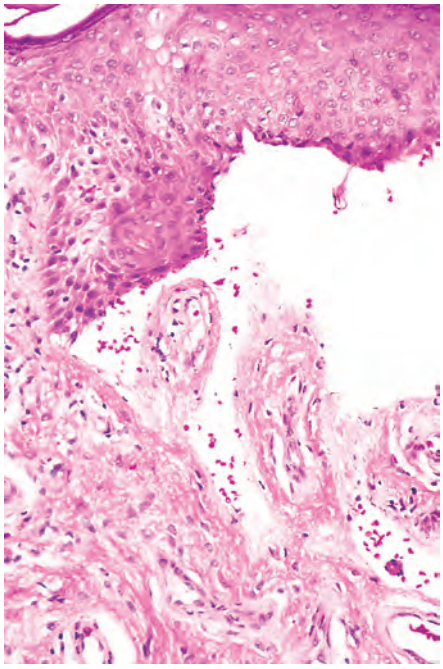


Figure 25.36 Porphyria. A noninflammatory blister at the dermoepidermal junction; note the seemingly rigid dermal papillae at the base that contain abnormal superficial vessels.

KEY CONCEPTS

BLISTERING DISORDERS

- Blistering disorders are classified based on the level of epidermal separation.

- These disorders are often caused by autoantibodies specific for epithelial or basement membrane proteins that lead to unmooring of keratinocytes (acantholysis).
- Pemphigus is associated with IgG autoantibodies to various intercellular desmogleins, resulting in bullae that are either subcorneal (pemphigus foliaceus) or suprabasilar (pemphigus vulgaris).
- Bullous pemphigoid is associated with IgG autoantibodies to basement membrane proteins and produces a subepidermal blister.
- Dermatitis herpetiformis is associated with IgA autoantibodies to fibrils that bind the epidermal basement membrane to the dermis and also produces subepidermal blisters.
- Noninflammatory blistering disorders include inherited defects in proteins that stabilize the epidermis (e.g., epidermolysis bullosa), as well as defects in porphyrin synthesis (the porphyrias) that lead to sun-induced skin damage through uncertain mechanisms.

DISORDERS OF EPIDERMAL APPENDAGES

Acne Vulgaris

Virtually universal in the middle to late teenage years, acne vulgaris affects males and females, with males tending to have more severe disease. Acne is seen in all races but is usually milder in people of Asian descent. It may be induced or exacerbated by drugs (corticosteroids, adrenocorticotropic hormone, testosterone, gonadotropins, contraceptives, trimethadione, iodides, and bromides), occupational exposures (cutting oils, chlorinated hydrocarbons, and coal tars), and conditions that favor occlusion of sebaceous glands, such as heavy clothing, cosmetics, and tropical climates. Some families seem to be particularly prone to acne, suggesting a hereditary component.

Acne is divided into noninflammatory and inflammatory types, although both types may coexist. Noninflammatory acne may take the form of open and closed comedones.

- *Open comedones* are small follicular papules containing a central black keratin plug. This color is the result of oxidation of melanin pigment (not dirt).
- *Closed comedones* are follicular papules without a visible central plug. Because the keratin plug is trapped beneath the epidermal surface, these lesions are potential sources of follicle rupture and inflammation.

Pathogenesis

The pathogenesis of acne is incompletely understood and is likely multifactorial. At least four factors contribute to its development: (1) keratinization of the lower portion of the follicular infundibulum and development of a keratin plug that blocks outflow of sebum to the skin surface; (2) hypertrophy of sebaceous glands during puberty under the influence of androgens; (3) lipase-synthesizing bacteria *(Propionibacterium acnes)* colonizing the upper and mid-portion of the hair follicle, converting lipids within sebum to proinflammatory fatty acids; and (4) secondary inflammation of the involved follicle. Androgens were first implicated in times past when it was noted that young castrated males generally did not develop the condition (a questionable tradeoff). Elimination of *P. acnes* is the rationale for administration of antibiotics to individuals with inflammatory acne. The synthetic vitamin A derivative 13-*cis*-retinoic acid (isotretinoin) brings about remarkable improvement in some cases of severe acne through its strong antisebaceous action.

MORPHOLOGY

Inflammatory acne is marked by erythematous papules, nodules, and pustules (Fig. 25.37A). Severe variants (e.g., **acne conglobata**) result in sinus tract formation and dermal scarring. Depending on the stage of the disease, open or closed comedones, papules, pustules, or deep inflammatory nodules may develop. **Open comedones** have large, patulous orifices, whereas the orifices of **closed comedones** are identifiable only microscopically (Fig. 25.37B, C). Variable infiltrates of lymphocytes and macrophages are present in and around affected follicles, and extensive acute inflammation accompanies follicle rupture. Dermal abscesses may form in association with rupture (Fig. 25.37B) and lead to scarring.

Rosacea

Rosacea is a common disease of middle age and beyond, affecting up to 3% of the U.S. population, with a predilection for females. Four stages are recognized: (1) flushing episodes (pre-rosacea), (2) persistent erythema and telangiectasia, (3) pustules and papules, and (4) rhinophyma (permanent thickening of the nasal skin by confluent erythematous papules and prominent follicles).

Pathogenesis

A clue to the cause comes from observations showing that individuals with rosacea have high cutaneous levels of the antimicrobial peptide cathelicidin, an important mediator of the cutaneous innate immune response. The cathelicidin peptides present in affected individuals are qualitatively distinct from those seen in individuals without rosacea as a result of alternative processing by proteases such as kallikrein 5 (also known as stratum corneum tryptic enzyme). Injection of cathelicidin peptides from patients into mice induces some of the cutaneous changes seen in rosacea, including inflammation and vascular dilation. In addition, it has been noted that activation of TLR2 upregulates kallikrein 5 expression in keratinocytes, suggesting that factors that stimulate TLR2 are involved. Several microbial triggers have been proposed, but none are proven.

MORPHOLOGY

Rosacea is characterized by a nonspecific perifollicular infiltrate composed of lymphocytes surrounded by dermal edema and telangiectasia. In the pustular phase, neutrophils may colonize the follicles, and folicle rupture may occur and cause a granulomatous dermal response. The development of rhinophyma is associated with hypertrophy of sebaceous glands and follicular plugging by keratotic debris.

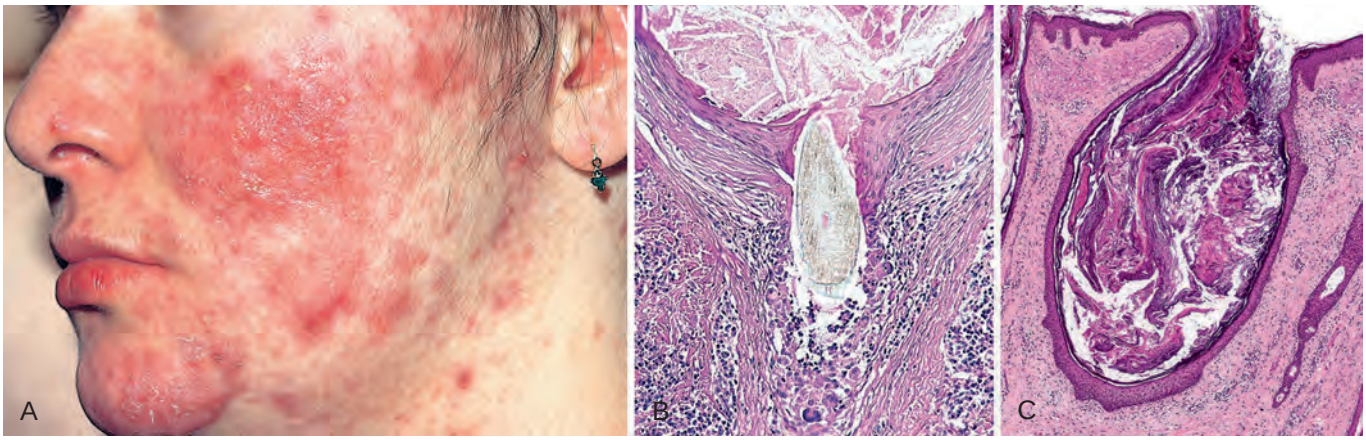


Figure 25.37 Acne. (A) Inflammatory acne associated with erythematous papules and pustules. (B) A hair shaft pierces the follicular epithelium, eliciting inflammation and fibrosis. (C) An open comedone.

PANNICULITIS

Erythema Nodosum and Erythema Induratum

Panniculitis is an inflammatory reaction in the subcutaneous adipose tissue that may preferentially affect (1) the lobules of fat or (2) the connective tissue that separates fat into lobules. Panniculitis often involves the lower legs. *Erythema nodosum* is the most common form and usually has a subacute presentation. A second somewhat distinctive form, erythema induratum, also merits brief discussion.

- *Erythema nodosum* presents as poorly defined, exquisitely tender, erythematous plaques and nodules that may be more readily palpated than seen. Its occurrence is often associated with infections (β-hemolytic streptococcal infection and tuberculosis and, less commonly, coccidioidomycosis, histoplasmosis, and leprosy), drug administration (sulfonamides and oral contraceptives), sarcoidosis, inflammatory bowel disease, and certain malignant neoplasms, but many times a cause cannot be identified. Fever and malaise may accompany the cutaneous signs. It is thought to be caused by a delayed hypersensitivity reaction to microbial or drug-related antigens. In some cases, immune complexes have been implicated, but in many cases the pathogenesis is mysterious. Over the course of weeks, lesions usually flatten and become bruise-like, leaving no residual clinical scars, while new lesions develop. Biopsy of a deep wedge of tissue to generously sample the subcutis is usually required for histologic diagnosis.
- *Erythema induratum* is an uncommon type of panniculitis that affects primarily adolescents and menopausal women. Although the cause is not known, most observers regard this as a primary vasculitis of deep vessels supplying the fat lobules of the subcutis; the associated vascular compromise leads to fat necrosis and inflammation. Erythema induratum presents as an erythematous, slightly tender nodule that usually goes on to ulcerate. Originally considered a hypersensitivity response to tuberculosis, erythema induratum today most commonly occurs without an associated underlying disease.

MORPHOLOGY

The histopathology of **erythema nodosum** is distinctive. In early lesions the connective tissue septa are widened by edema, fibrin exudation, and neutrophilic infiltration. Later, infiltration by lymphocytes, histiocytes, multinucleated giant cells, and occasional eosinophils is associated with septal fibrosis. Vasculitis is not present. In **erythema induratum,** on the other hand, granulomatous inflammation and zones of caseous necrosis involve the fat lobule. Early lesions show necrotizing vasculitis affecting small- to medium-sized arteries and veins in the deep dermis and subcutis.

Other Types of Panniculitis

Many other types of panniculitis have also been described, a few of which merit brief mention.

- *Weber-Christian disease (relapsing febrile nodular panniculitis)* is a rare form of lobular, nonvasculitic panniculitis seen in children and adults. It is marked by crops of erythematous plaques or nodules, predominantly on the lower extremities, created by deep-seated foci of inflammation containing aggregates of foamy macrophages admixed with lymphocytes, neutrophils, and giant cells.
- *Factitial panniculitis* is a form of secondary panniculitis caused by self-inflicted trauma or injection of foreign or toxic substances.
- Rare types of *T-cell lymphoma* home to fat lobules, producing fat necrosis and superimposed inflammation that mimics panniculitis.
- *Systemic lupus erythematosus* may occasionally cause inflammation of the subcutis and an associated panniculitis.

INFECTION

Despite its ample protective mechanisms, the exposure of the skin to frequent trauma and the external environment leaves it vulnerable to infection by microorganisms, parasites, and insects. We have already discussed the possible role of bacteria in the pathogenesis of common acne, and the

dermatoses resulting from viruses are too numerous to list. In the setting of the immunocompromised individual, ordinarily trivial cutaneous infections may become life threatening. Many disorders, such as herpes simplex and herpes zoster, viral exanthems, deep fungal infections, and immune reactions in skin provoked by infectious agents, are discussed in Chapter 8. Here we cover a representative sampling of common infections whose primary clinical manifestations are in the skin.

Verrucae (Warts)

Verrucae are squamoproliferative lesions caused by human papillomaviruses (HPVs). They are common lesions of children and adolescents, although they may be encountered at any age. Transmission of disease usually involves direct contact between individuals or autoinoculation. Verrucae are generally self-limited, regressing spontaneously within 6 months to 2 years.

Pathogenesis

More than 150 types of papillomavirus have been identified, many of them capable of producing warts in humans. The clinical variants of warts are often associated with distinct HPV subtypes. For example, anogenital warts are caused predominantly by HPV types 6 and 11. HPV type 16 has been associated with in situ squamous cell carcinoma of the genitalia and with bowenoid papulosis; the latter is a genital lesion of young adults with the histologic appearance of carcinoma in situ, but which usually regresses spontaneously (see also Chapter 21). The relationship of HPV subtypes 5 and 8 to squamous cell carcinomas, particularly in individuals affected by the rare condition epidermodysplasia verruciformis, was mentioned earlier. These patients develop multiple flat warts that contain HPV genomes, some of which progress to carcinoma. Viral typing can be accomplished by either in situ hybridization (Fig. 25.38D) or polymerase chain reaction.

HPVs that are associated with a high risk of cancer produce E6 proteins that abolish p53 function (Chapter 7). By contrast, HPV subtypes 5 and 8 produce variant E6 proteins that do not affect p53, explaining why these forms of HPV have low oncogenic potential. By contrast, the E6 proteins of low-risk HPVs interfere with Notch signaling, which is required for the normal maturation of keratinocytes; this effect likely contributes to the epidermal hyperplasia that characterizes warts.

MORPHOLOGY

The classification of verrucae is based largely on appearance and location. **Verruca vulgaris** is the most common type of wart. The lesions of verruca vulgaris may occur anywhere but are most often found on the hands, particularly on the dorsal surfaces and periungual areas, where they appear as gray-white to tan, flat to

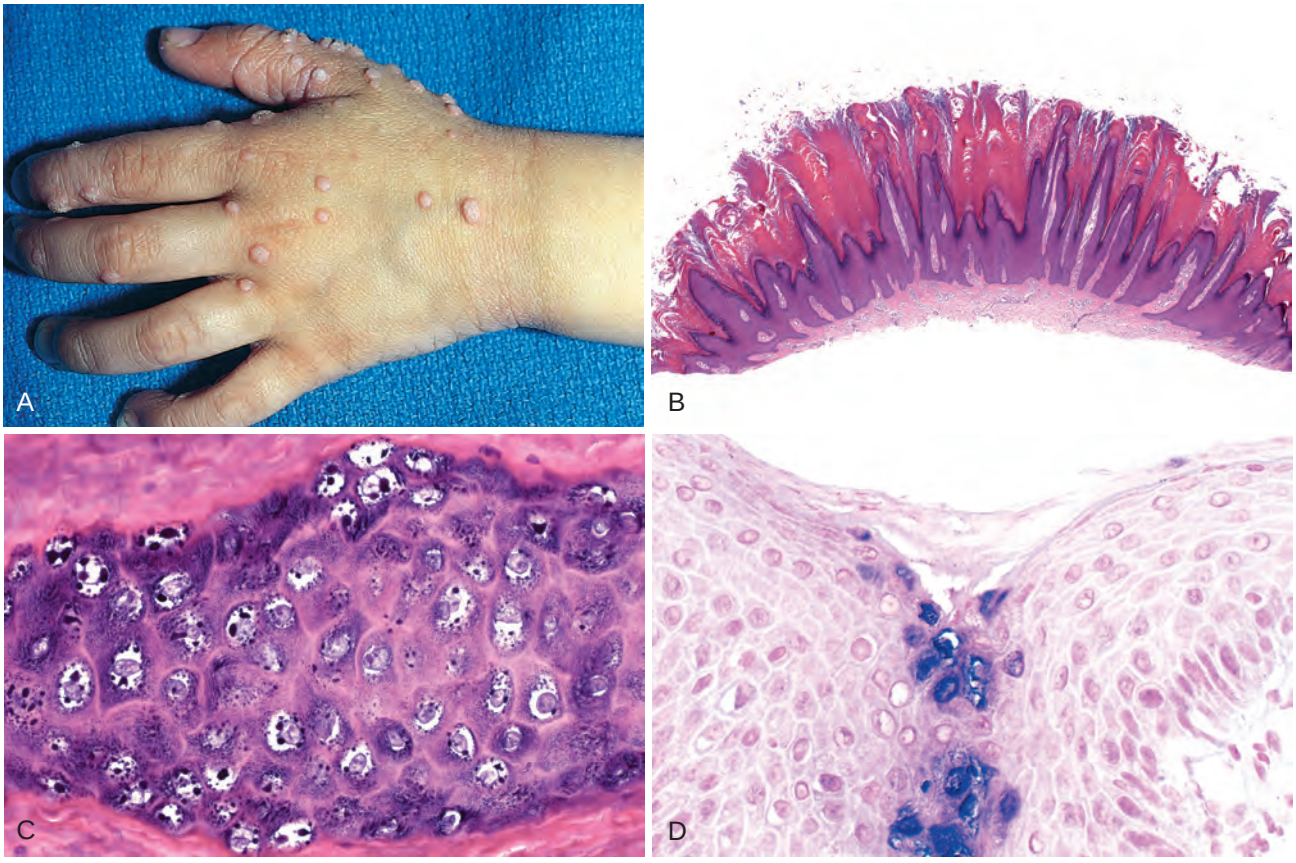


Figure 25.38 Verruca vulgaris. (A) Multiple papules with rough pebble-like surfaces. Low-power (B) and high-power (C) lesions showing papillomatous epidermal hyperplasia and cytopathic alterations including nuclear pallor and prominent keratohyaline granules. (D) In situ hybridization demonstrating human papillomavirus DNA within epidermal cells.

convex, 0.1- to 1-cm papules with a rough, pebble-like surface (Fig. 25.38A). **Verruca plana,** or **flat wart,** is common on the face or the dorsal surfaces of the hands. The warts are slightly elevated, flat, smooth, tan papules that are generally smaller than verruca vulgaris. **Verruca plantaris** and **verruca palmaris** occur on the soles and palms, respectively. These rough, scaly lesions may reach 1 to 2 cm in diameter, may sometimes coalesce, and may be confused with ordinary calluses. **Condyloma acuminatum (venereal wart)** occurs on the penis, female genitalia, urethra, perianal areas, and rectum. Venereal warts appear as soft, tan, cauliflower-like masses that occasionally reach many centimeters in diameter.

Histologic features common to verrucae include epidermal hyperplasia that is often undulant or spire-like in character, termed **verrucous** or **papillomatous epidermal hyperplasia** (Fig. 25.38B), and **cytoplasmic vacuolization (koilocytosis)** involving the more superficial epidermal layers, producing haloes of pallor surrounding infected nuclei. Electron microscopy of these zones reveals numerous HPV virions within nuclei. Infected cells may also demonstrate prominent, condensed keratohyaline granules and jagged eosinophilic intracytoplasmic keratin aggregates as a result of viral cytopathic effects (Fig. 25.38C). These cellular alterations are not as prominent in condylomas; hence, their diagnosis is based primarily on hyperplastic papillary architecture containing wedge-shaped zones of koilocytosis.

Molluscum Contagiosum

Molluscum contagiosum is a common, self-limited viral disease of the skin caused by a poxvirus. The virus is characteristically brick shaped, has a dumbbell-shaped DNA core, and measures 300 nm in maximal dimension, making it one of the largest pathogenic poxviruses in humans and one of the largest viruses in nature. Infection is usually spread by direct contact, particularly among children and young adults.

MORPHOLOGY

Multiple lesions may occur on the skin and mucous membranes, with a predilection for the trunk and anogenital areas. Individual lesions are firm, often pruritic, pink to skin-colored umbilicated papules, generally ranging in diameter from 0.2 to 0.4 cm. Rarely, "giant" forms occur measuring up to 2 cm in diameter. A curd-like material can be expressed from the central umbilication. Smearing this material onto a glass slide and staining with Giemsa often shows diagnostic molluscum bodies.

On microscopic examination, lesions show cup-like verrucous epidermal hyperplasia. The diagnostically specific structure is the **molluscum body,** which is a large (up to 35 μm), ellipsoid, homogeneous, cytoplasmic inclusion that is found in the cells of the stratum granulosum and the stratum corneum (Fig. 25.39). In hematoxylin and eosin stains, the inclusions are eosinophilic in the stratum granulosum and acquire a pale blue hue in the stratum corneum. Numerous virions are present within molluscum bodies.

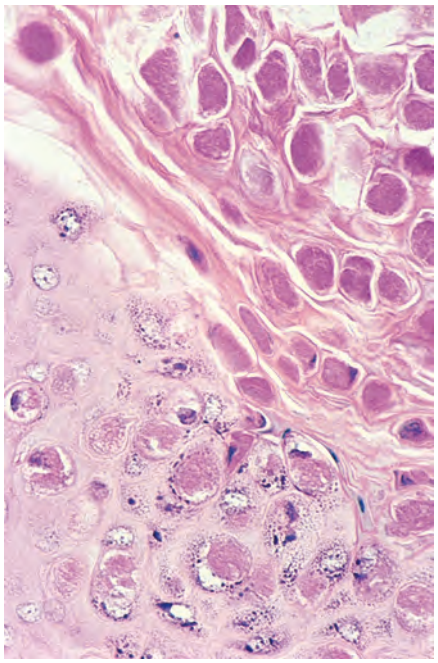


Figure 25.39 Molluscum contagiosum. A focus of verrucous epidermal hyperplasia contains numerous cells with ellipsoid cytoplasmic inclusions (molluscum bodies) within the stratum granulosum and stratum corneum.

Impetigo

Impetigo is a common superficial bacterial infection of skin. It is highly contagious and is frequently seen in otherwise healthy children as well as occasionally in adults in poor health. The infection usually involves exposed skin, particularly that of the face and hands. Two forms exist, classically referred to as *impetigo contagiosa* and *impetigo bullosa;* they differ from each other simply by the size of the pustules. Over the past several decades a remarkable shift in etiology has been observed. Whereas in the past impetigo contagiosa was almost exclusively caused by group A β-hemolytic streptococci and impetigo bullosa by *Staphylococcus aureus,* both are now usually caused by *S. aureus.*

Pathogenesis

Bacteria in the epidermis evoke an innate immune response that causes epidermal injury, leading to local serous exudate and formation of a scale crust (scab). The pathogenesis of blister formation in impetigo is related to bacterial production of a toxin that specifically cleaves desmoglein 1, the protein responsible for cell-to-cell adhesion within the uppermost epidermal layers. Recall that in pemphigus foliaceus, which has a similar plane of blister formation, desmoglein 1 is compromised not by a toxin but by an autoantibody (see Fig. 25.28). Because there is virtually no involvement of the dermis, once the bacteria are eliminated the lesions heal without scarring.

MORPHOLOGY

Impetigo presents as an erythematous macule, but multiple small pustules rapidly supervene. As pustules break, shallow erosions

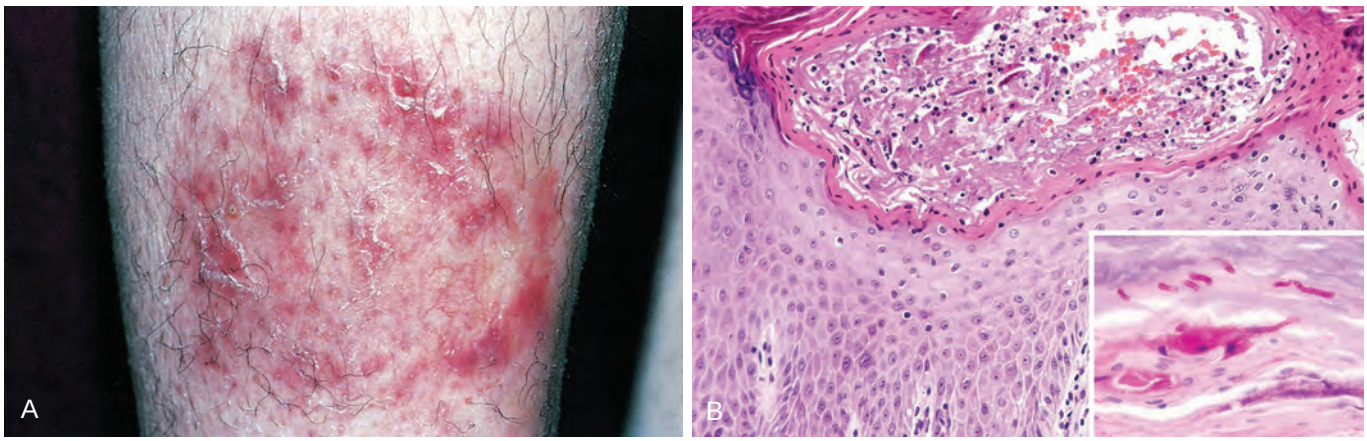


Figure 25.40 Tinea. (A) Characteristic plaque of tinea corporis. (B) Routine histology shows a mild eczematous (spongiotic) dermatitis and focal neutrophilic abscesses. Periodic acid–Schiff stain *(inset)* reveals deep red hyphae within the stratum corneum.

form, covered with drying serum, giving the characteristic appearance of **honey-colored crust.** If the crust is not removed, new lesions form about the periphery, and extensive epidermal damage may ensue. A bullous form of impetigo mainly occurs in children.

The characteristic microscopic feature of impetigo is **accumulation of neutrophils beneath the stratum corneum,** often producing a subcorneal pustule containing serum proteins and inflammatory cells. Special stains reveal the presence of bacteria in these foci. Nonspecific, reactive epidermal alterations and superficial dermal inflammation accompany these findings. Rupture of pustules releases serum, neutrophils, and cellular debris, which layer out and dry to form the characteristic crust.

Superficial Fungal Infections

As opposed to deep fungal infections of the skin, where the dermis or subcutis is primarily involved, superficial fungal infections of the skin are confined to the stratum corneum and are caused primarily by dermatophytes. These organisms grow in the soil and on animals and produce a number of diverse lesions with characteristic distributions, as follows:

- *Tinea capitis* usually occurs in children and is only rarely seen in infants and adults. It is a dermatophytosis of the scalp characterized by asymptomatic, patchy skin lesions associated with mild erythema, crust formation, scaling, and frequent hair loss.
- *Tinea barbae* is a dermatophyte infection of the beard area that affects adult men; it is relatively uncommon.
- *Tinea corporis,* on the other hand, is a common superficial fungal infection of skin that affects persons of all ages, but particularly children. Predisposing factors include excessive heat and humidity, exposure to infected animals, and chronic dermatophytosis of the feet or nails. The most common type of tinea corporis is an expanding, round, slightly erythematous plaque with an elevated scaling border (Fig. 25.40A).
- *Tinea cruris* occurs most frequently in the inguinal areas of obese men during warm weather. Heat, friction, and maceration all predispose to its development. The infection usually first appears on the upper inner thighs as moist, red patches with raised scaly borders.
- *Tinea pedis (athlete's foot)* affects 30% to 40% of the population at some time in their lives. There is diffuse erythema and scaling, often initially localized to the web spaces. Most of the inflammatory reaction, however, appears to be the result of bacterial superinfection and is not directly related to the primary dermatophytosis. Spread to (or primary infection of the nails) is referred to as *onychomycosis*. This produces discoloration, thickening, and deformity of the nail plate.
- *Tinea versicolor* usually occurs on the upper trunk and is highly distinctive in appearance. Caused by *Malassezia furfur* (a yeast, not a dermatophyte), the lesions consist of groups of macules of varied size and color with a fine peripheral scale.

MORPHOLOGY

The histologic features of dermatophytoses are variable, depending on the properties of the organism, the host response, and the degree of bacterial superinfection. There may be mild eczematous dermatitis associated with intraepidermal neutrophils (Fig. 25.40B). Due to cell walls rich in mucopolysaccharides, fungi stain bright pink to red with periodic acid–Schiff stain. They are found in the anucleate cornified layer of lesional skin, hair, or nails (Fig. 25.40B, *inset*). Culture of material scraped from these areas usually permits identification of the offending species.

SUGGESTED READINGS

Melanoma and Dysplastic Nevi

Breslow A: Prognosis in cutaneous melanoma: tumor thickness as a guide to treatment, *Pathol Annu* 15:1, 1980. *[Classic article describing the importance of vertical growth of melanoma, which remains the best predictor of outcome in this disease].*

Cancer Genome Atlas Network: Genomic classification of melanoma, *Cell* 16:1681, 2015. *[Comprehensive genomic characterization of melanoma demonstrating categories relevant to targeted treatment].*

Clark WH Jr, Remmer RR, Greene M et al: Origin of familial malignant melanomas from heritable melanocytic lesions. "The B-K mole syndrome.", *Arch Dermatol* 114:732, 1978. *[Classic article describing familial dysplastic nevi and their possible relationship to melanoma].*

Elder DE: Dysplastic nevi: an update, *Histopathology* 56:112, 2010. *[Updated discussion of histology and pathogenesis of dysplastic nevi and their relationship to melanoma]*.

Keung EZ, Wargo JA: The current landscape of immune checkpoint inhibition for solid malignancies, *Surg Oncol Clin N Am* 28:369, 2019. *[Excellent review on immunotherapy using melanoma as a paradigm of a responsive cancer]*.

Luke JJ, Flaherty KT, Ribas A et al: Targeted agents and immunotherapies: optimizing outcomes in melanoma, *Nat Rev Clin Oncol* 14: 463, 2017. *[Update on latest treatment agents and paradigms in melanoma]*.

Shain AH, Joseph NM, Yu R et al: Genomic and transcriptomic analysis reveals incremental disruption of key signaling pathways during melanoma evolution, *Cancer Cell* 34:45, 2018. *[Documents the key molecular features of primary cutaneous melanoma progression]*.

Epidermal Skin Tumors

Epstein EH: Basal cell carcinomas: attack of the hedgehog, *Nat Rev Cancer* 8:743, 2008. *[Epidemiology, clinical presentation, molecular pathogenesis, and novel treatment options are succinctly reviewed]*.

Hafner C, Hartmann H, Vogt T et al: High frequency of FGFR3 mutations in adenoid seborrheic keratoses, *J Invest Dermatol* 126:2404, 2006. *[Description of activating mutations in FGFR3 in seborrheic keratoses]*.

Pickering CR, Zhou JH, Lee JJ et al: Mutational landscape of aggressive cutaneous squamous cell carcinoma, *Clin Cancer Res* 20:2682, 2014. *[Documents key driver events in advanced cutaneous squamous cell carcinoma]*.

Mastocytosis and Cutaneous T-Cell Lymphoma

Stadler R, Stranzenbach R: Molecular pathogenesis of cutaneous lymphomas, *Exp Dermatol* 2018. [epub ahead of print]. *[Summary of epidemiology, staging, natural history, and immunopathogenesis of cutaneous T-cell lymphoma]*.

Volertas S, Schuler CF 4th, Akin C: New insights into clonal mast cell disorders including mastocytosis, *Immunol Allergy Clin North Am* 38:341, 2018. *[Update on pathogenesis and treatment of mast cell diseases]*.

Autoimmune and Inflammatory Skin Disorders

Coenraads PJ: Hand eczema, *N Engl J Med* 367:1829, 2012. *[Thorough discussion of the causes and management of a common form of eczema]*.

Nestle FO, Kaplan DH, Barker J: Psoriasis, *N Engl J Med* 361:496, 2009. *[Pathogenesis, clinical features, and targeted treatment options are discussed]*.

Schaefer P: Urticaria: evaluation and treatment, *Am Fam Physician* 83:1078, 2011. *[Practical discussion of the clinical evaluation and treatment of urticaria]*.

Sharma A, Bialynicki-Birula R, Schwatz RA et al: Lichen planus: an update and review, *Cutis* 90:17, 2012. *[Review of pathology and clinical features, natural history, and treatment]*.

Blistering Disorders

Bonciani D, Verdelli A, Bonciolini V et al: Dermatitis herpetiformis: from genetics to the development of skin lesions, *Clin Dev Immunol* 2012. doi:10.1155/2012/239691. *[Description of factors and disease mechanisms underlying this rare disorder]*.

Pollmann R, Schmidt T, Eming R et al: Pemphigus: a comprehensive review on pathogenesis, clinical presentation and novel therapeutic approaches, *Clin Rev Allergy Immunol* 54:1, 2018. *[An update on basic and clinical aspects of pemphigus]*.

Ujiie H, Shibaki A, Nishie W et al: What's new in bullous pemphigoid, *J Dermatol* 37:194, 2010. *[Review of bullous pemphigoid pathogenesis]*.

Disorders of Skin Appendages and Subcutis

Knutsen-Larson S, Dawson AL, Dunnick CA et al: Acne vulgaris: pathogenesis, treatment, and needs assessment, *Dermatol Clin* 30:99, 2012. *[Review of epidemiology, pathogenesis, and treatment of acne in the United States]*.

Nakatsuji T, Gallo RL: Antimicrobial peptides: old molecules with new ideas, *J Invest Dermatol* 132:887, 2012. *[Discussion of the possible role of antimicrobial peptides such as cathelicidin in rosacea, psoriasis, and atopic dermatitis]*.

CHAPTER 26

Bones, Joints, and Soft Tissue Tumors

Andrew Horvai

CHAPTER CONTENTS

Bone

BASIC STRUCTURE AND FUNCTION OF BONE

The adult human skeleton is composed of 206 bones that account for ~12% of body weight. Bone functions include mechanical support, force transmission, internal organ protection, and mineral homeostasis, and it also serves as the major site of hematopoiesis during postnatal life. Bone consists of extracellular matrix and several cell types.

Matrix

The extracellular component of bone, the matrix, is composed of osteoid (35%) and minerals (65%). The latter,

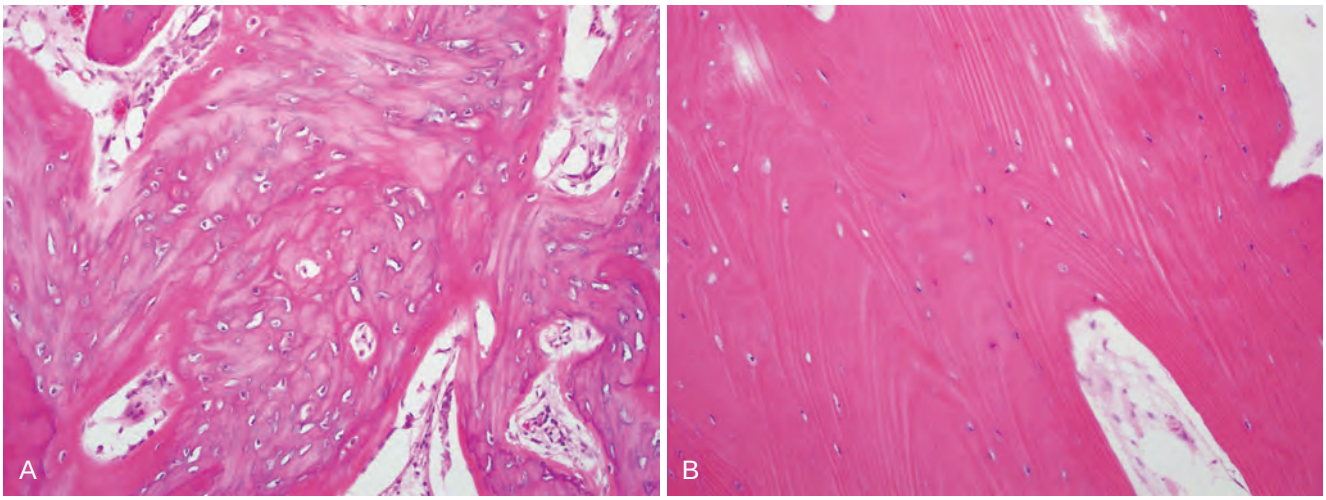


Figure 26.1 (A) Woven bone. (B) Lamellar bone contrasts with the more cellular, disorganized appearance of woven bone.

primarily hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$], give bone its hardness and serves as a repository for 99% of the calcium and 85% of the phosphorus in the body.

Osteoid consists of type I collagen and smaller amounts of glycosaminoglycans and other proteins. One of these proteins, osteopontin (also called osteocalcin), is produced by osteoblasts and contributes to the regulation of bone formation, mineralization, and calcium homeostasis. Serum osteopontin levels are used as a specific marker of osteoblast activity. Bone maturation and metabolism are also sensitive to cytokines and growth factors and are thus regulated by diverse inputs, which include locally and systemically produced factors, as well as mechanical force.

Bone matrix can be woven or lamellar (Fig. 26.1). Woven bone is produced rapidly (e.g., during fetal development or fracture repair), but the haphazard arrangement of collagen fibers imparts less structural integrity than the parallel collagen fibers of lamellar bone. Woven bone is always abnormal in adults, but its presence is not specific for any particular disease. Long bones are composed of a dense outer cortex and a central medulla. The latter is supported by bony trabeculae interspersed with marrow, which may be fatty (white) or hematopoietic (red).

Cells

The cellular components of bone include osteoblasts, osteocytes, and osteoclasts.

- *Osteoblasts* on the surface of the osteoid matrix synthesize, transport, and assemble matrix and regulate mineralization (Fig. 26.2A). Osteoblast activity is tightly regulated by hormonal and local mediators. Quiescent osteoblasts, which can be recognized by a decrease in cytoplasmic volume, may remain on the trabecular surface or become embedded within the matrix as osteocytes.
- *Osteocytes* are interconnected by an intricate network of dendritic cytoplasmic processes through tunnels (canaliculi) within the matrix. Osteocytes help control calcium and phosphate levels in the microenvironment, detect mechanical forces, and translate those forces into biologic activity—a process called mechanotransduction.

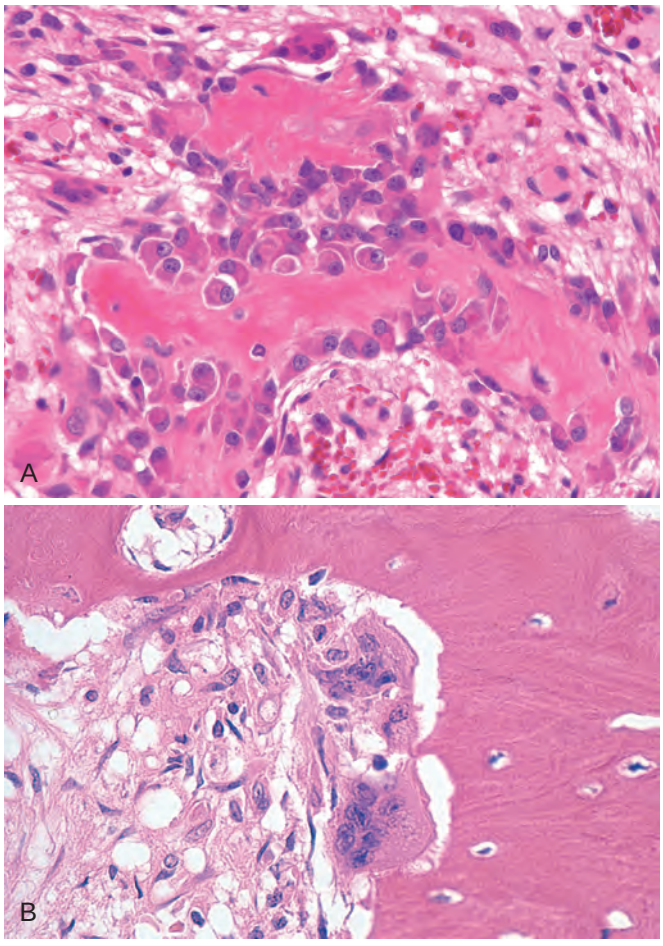


Figure 26.2 (A) Active osteoblasts synthesizing bone matrix. The surrounding spindle cells represent osteoprogenitor cells. (B) Two osteoclasts resorbing bone.

- *Osteoclasts* are specialized multinucleated macrophages that are derived from circulating monocytes and resorb bone (Fig. 26.2B). Surface integrin proteins allow osteoclasts to attach to the matrix and create a sealed extracellular trench (resorption pit). Secretion of acid and neutral

proteases, predominantly matrix metalloproteases (MMPs), into the pit results in dissolution of inorganic and organic bone components.

Development

Most bones that form during embryogenesis develop from a cartilage mold via endochondral ossification. The cartilage mold (anlagen) is synthesized by mesenchymal precursor cells. A central medullary canal within the anlagen is created by chondroblasts at approximately 8 weeks of gestation. Simultaneously, osteoblasts begin to deposit the cortex beneath the nascent periosteum of the midshaft (diaphysis). This forms a primary center of ossification resulting in radial bone growth. At the longitudinal ends (epiphysis), endochondral ossification forms secondary centers of ossification. Eventually, plates of cartilage anlage become entrapped between the expanding centers of ossification forming physes or growth plates (Fig. 26.3). Chondrocytes within the growth plates undergo sequential proliferation, hypertrophy, and apoptosis. Matrix mineralizes during apoptosis and is invaded by capillaries, providing the nutrients for activation of osteoblasts and osteoid synthesis. Most calcified cartilage matrix is ultimately resorbed leaving only strut-shaped remnants that serve as scaffolding for bone deposits known as primary spongiosa, the earliest bone trabeculae (see Fig. 26.3). Over time, this process produces longitudinal bone growth.

Flat bones, for example the cranium, are formed by intramembranous ossification, in which a dense layer of mesenchyme is directly ossified by osteoblasts without a cartilage anlagen. Bones enlarge by deposition of new bone on a preexisting surface, a process called appositional growth.

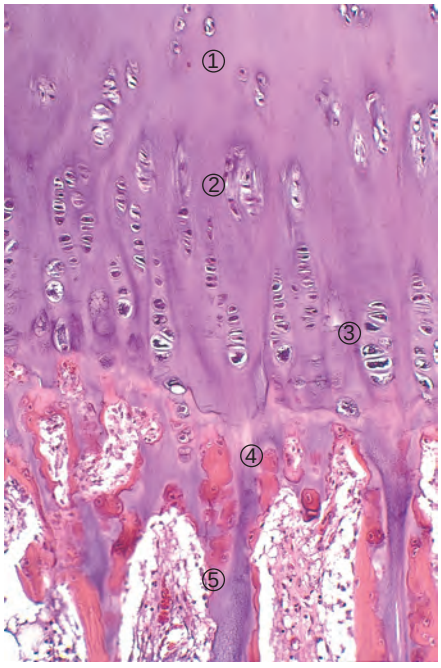


Figure 26.3 Active growth plate with ongoing endochondral ossification. 1, Reserve zone. 2, Zone of proliferation. 3, Zone of hypertrophy. 4, Zone of mineralization. 5, Primary spongiosa.

Local and systemic factors that regulate bone development include the following:

- Growth hormone (GH) secreted by the anterior pituitary gland induces and maintains chondrocyte proliferation.
- Thyroid hormone secreted by the thyroid gland acts on proliferating chondrocytes to induce hypertrophy.
- Indian hedgehog (Ihh) is secreted locally by prehypertrophic chondrocytes and coordinates chondrocyte proliferation and differentiation with osteoblast proliferation.
- Parathyroid hormone–related protein (PTHrP) produced by perichondrial stromal cells and early proliferating chondrocytes activates the PTH receptor to maintain chondrocyte proliferation.
- Wnt growth factors are expressed in the growth plate–proliferating zone and, via Frizzled and LRP5/6 receptors, activate β-catenin to promote chondrocyte proliferation and maturation.
- SOX9 is a transcription factor expressed by proliferating, but not hypertrophic, chondrocytes that is essential for differentiation of chondrocyte precursors.
- RUNX2 is a transcription factor expressed in early hypertrophic chondrocytes and immature mesenchymal cells that controls terminal chondrocyte and osteoblast differentiation.
- Fibroblast growth factors (FGFs) are secreted by a variety of mesenchymal cells. FGFs (most notably FGF-3) act on hypertrophic chondrocytes to inhibit proliferation and promote differentiation.
- Bone morphogenic proteins (BMPs), members of the TGF-β family, are expressed at various stages of chondrocyte development and have diverse effects on chondrocyte proliferation and hypertrophy at the growth plate.

Homeostasis and Remodeling

The adult skeleton appears static but actually undergoes continuous change via a tightly regulated process known as remodeling. This process, which turns over approximately 10% of the skeleton each year, repairs damage, and may change the shape of bones in response to mechanical forces. Remodeling occurs within the *bone (*or *basic) multicellular unit (BMU),* which consists of a unit of coupled osteoblast and osteoclast activity on the bone surface. Osteoclast attachment, bone resorption, osteoblast attachment and proliferation, and, finally, matrix synthesis occur sequentially at the BMU.

Events at the BMU are regulated by cell-to-cell interactions and cytokines, and several signaling pathways (Fig. 26.4). One pathway involves three factors: (1) the transmembrane receptor activator for NF-κB (RANK), which is expressed on osteoclast precursors; (2) RANK ligand (RANKL), which is expressed on osteoblasts and marrow stromal cells; and (3) osteoprotegerin (OPG), a secreted decoy receptor made by osteoblasts and several other types of cells that binds RANKL and thus prevents its interaction with RANK. When stimulated by RANKL, RANK signaling activates NF-κB, which is essential for the generation and survival of osteoclasts. A second important pathway involves macrophage colony-stimulating factor (M-CSF), a factor produced by osteoblasts that is also crucial for the generation of osteoclasts. Finally, WNT proteins produced by osteoprogenitor cells bind to LRP5 and LRP6 on osteoblasts to activate β-catenin signaling and osteoprotogerin (OPG) synthesis (Fig. 26.5).

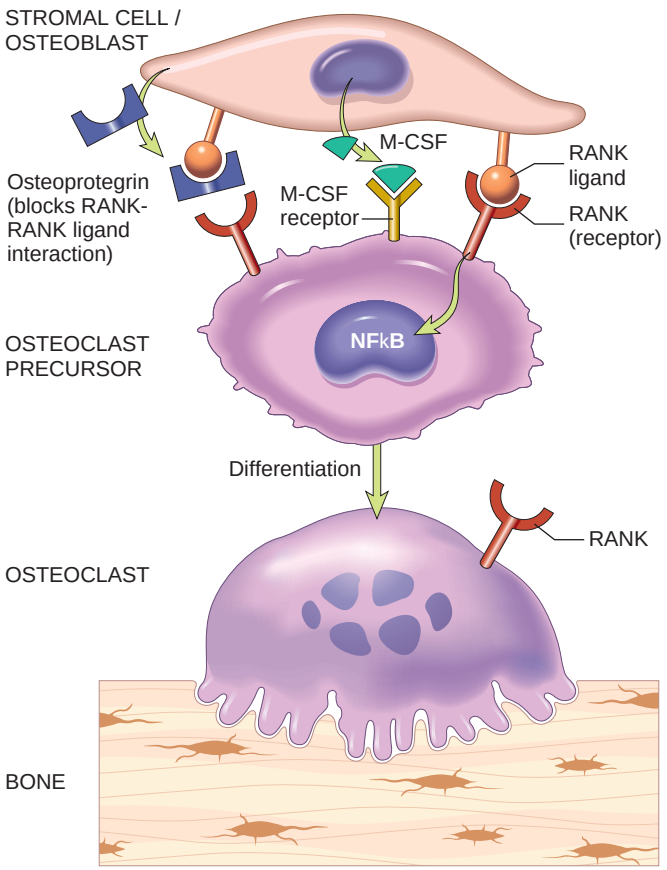


Figure 26.4 Paracrine molecular mechanisms that regulate osteoclast formation and function. Osteoclasts are derived from the same mononuclear cells that differentiate into macrophages. Osteoblast/stromal cell membrane-associated RANKL binds to its receptor RANK located on the cell surface of osteoclast precursors functions in concert with macrophage colony-stimulating factor (M-CSF). RANK activation induces precursor cells to become functional osteoclasts. Osteoblasts and stromal cells also secrete osteoprotegerin (OPG), which acts as a decoy receptor for RANKL, preventing it from binding the RANK on osteoclast precursors. Consequently, OPG prevents bone resorption by inhibiting osteoclast differentiation.

Conversely, osteocytes produce sclerostin, which inhibits WNT/β-catenin signaling and promotes bone formation. The importance of these pathways is emphasized by rare germline mutations in *OPG, RANK, RANKL, LRP5,* and *sclerostin* genes, which severely disrupt bone metabolism and produce congenital bone disorders (described later).

The balance between bone formation and resorption is modulated by RANK and WNT signaling. For example, because OPG and RANKL oppose one another, bone resorption or formation can be favored by increasing or decreasing the RANK-to-OPG ratio, respectively. Systemic factors affecting this balance include hormones (parathyroid hormone, estrogen, testosterone, and glucocorticoids), vitamin D, inflammatory cytokines (e.g., IL-1), and growth factors (e.g., bone morphogenetic factors). Each of these acts by altering RANK/NF-κB and WNT/β-catenin signaling. Parathyroid hormone, IL-1, and glucocorticoids promote osteoclast differentiation and bone turnover, while bone morphogenic proteins and sex hormones generally block osteoclast differentiation or activity by promoting OPG expression, alterations that favor bone deposition.

Another level of control involves paracrine signaling between osteoblasts and osteoclasts. Matrix breakdown by osteoclasts liberates and activates growth factors, cytokines, and enzymes (e.g., collagenase), some of which stimulate osteoblasts. Thus, substances that initiate bone deposition are released into the microenvironment during bone resorption (see Fig. 26.5).

Peak bone mass is achieved in early adulthood after the cessation of skeletal growth and is determined by factors that include vitamin D receptor and *LRP5/6* polymorphisms, nutrition, physical activity, age, and hormonal status. Beginning in the fourth decade, however, resorption exceeds formation, resulting in a steady decline in skeletal mass.

DEVELOPMENTAL DISORDERS OF BONE AND CARTILAGE

Developmental abnormalities of the skeleton often stem from inherited mutations and become apparent during the earliest stages of bone formation. In contrast, acquired diseases usually appear in adulthood. The spectrum of disorders of bone development is broad, and there is no standard approach to their classification. Here we will categorize the major diseases according to their pathogenesis.

Developmental anomalies can result from localized disruption of the migration and condensation of mesenchyme (dysostosis) or global disorganization of bone and/or cartilage (dysplasia). Dysostoses may occur in isolation or as part of more complex syndromes, and are caused by defects in mesenchymal condensation and differentiation into cartilage anlage. The most common forms include complete absence of a bone or entire digit (aplasia), extra bones or digits (supernumerary digit), and abnormal fusion of bones (e.g., syndactyly, craniosynostosis). Genetic alterations that affect genes encoding transcription factors (especially homeobox genes), cytokines, and cytokine receptors are especially common among dysostoses. In contrast, dysplasias arise from mutations in genes that control development or remodeling of the entire skeleton. It is important to note that while the term dysplasia in this context implies abnormal growth, it is not a precursor of neoplasia, as is the case with dysplasias of epithelial cells (Chapter 7).

More than 350 skeletal dysostoses and dysplasias are recognized, most of which are extremely rare. The classification has evolved from clinical and radiographic descriptions to one that includes the causative genetic defects. Table 26.1 lists some of the better characterized developmental abnormalities and their associated defective genes. The relationships between specific mutations and phenotypes are complex; different point mutations in a single gene (e.g., *COL2A1*) can result in distinct phenotypes, while mutations in disparate genes (e.g., *LRP5, RANKL*) can lead to similar phenotypes.

Defects in Nuclear Proteins and Transcription Factors

Defects in nuclear proteins and transcription factors, especially homeobox proteins, result in disorganized mesenchymal condensation and differentiation of osteoblasts and chondrocytes, leading to abnormal bone development.

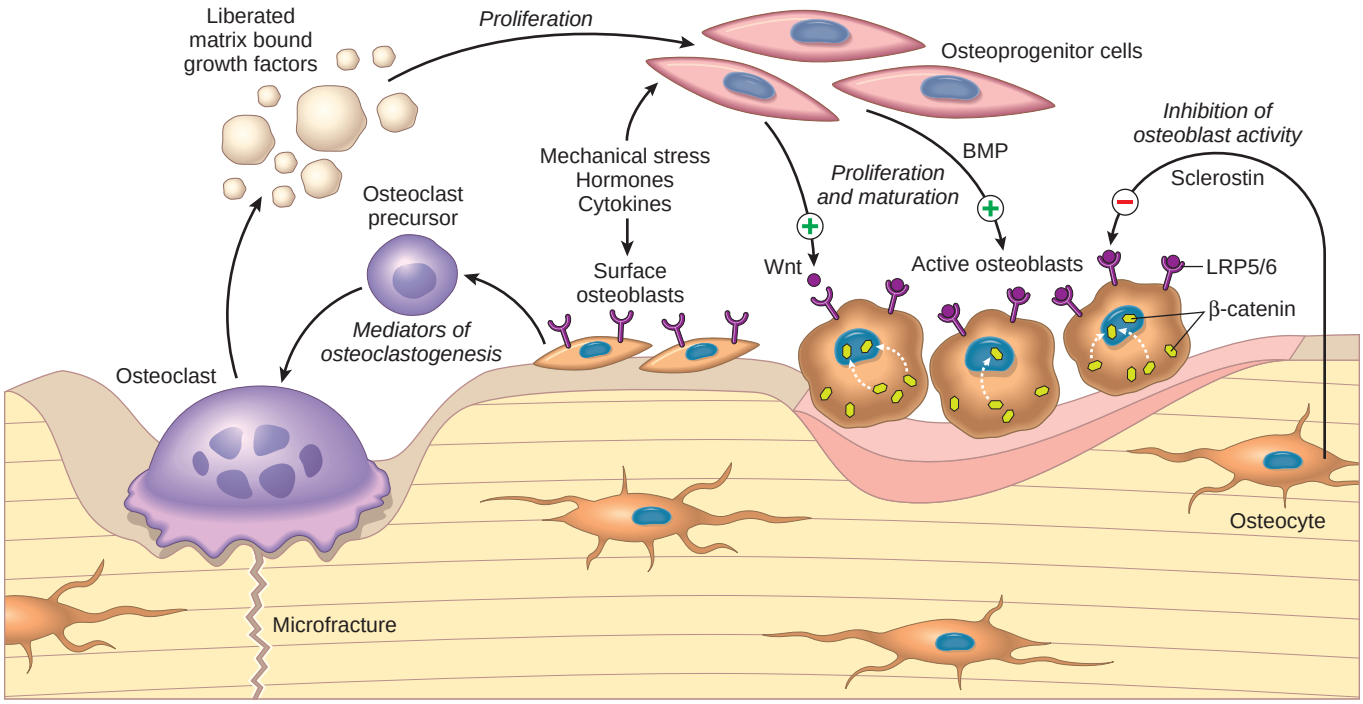


Figure 26.5 Bone cells and their interrelated activities. Hormones, cytokines, growth factors, and signal-transducing molecules are instrumental in bone formation and maturation and allow communication between osteoblasts and osteoclasts. Bone resorption and formation in remodeling are coupled processes controlled by systemic factors and local cytokines, some of which are deposited in the bone matrix. *BMP,* Bone morphogenic protein; *LRP5/6,* LDL receptor–related proteins 5 and 6.

- *Brachydactyly types D and E* are caused by mutations in the homeobox *HOXD13* gene and are characterized by shortening of the terminal phalanges of the thumb and big toe, respectively.
- Loss-of-function mutations in *RUNX2* result in *cleidocranial dysplasia,* an autosomal dominant disorder characterized by patent fontanelles, delayed closure of cranial sutures, Wormian bones (extra bones that occur within a cranial suture), delayed eruption of secondary teeth, primitive clavicles, and short stature.

Defects in Hormones and Signal Transduction Proteins

Achondroplasia is the most common skeletal dysplasia and a major cause of dwarfism. It is an autosomal dominant disorder caused by gain-of-function mutations in the FGF receptor 3 *(FGFR3)* gene, ~90% of which stem from new mutations in the paternal allele. FGF-mediated FGFR3 activation normally inhibits endochondral growth. This effect is exaggerated by *FGFR3* gain-of-function mutations. The retarded cartilage growth results in shortened proximal extremities, an enlarged head with bulging forehead, and depression of the root of the nose despite a trunk of relatively normal length. These skeletal abnormalities are not usually associated with changes in longevity, intelligence, or reproductive status.

Thanatophoric dysplasia is the most common lethal form of dwarfism. It affects ~1 in every 20,000 live births and is caused by *FGFR3* gain-of-function mutations distinct from those that cause achondroplasia, which appear to cause greater increases in FGFR3 signaling than those that produce achondroplasia, and thus a more severe phenotype. Affected individuals have disproportionately short (micromelic) limbs, frontal bossing, relative macrocephaly, a small chest cavity, and a bell-shaped abdomen. The underdeveloped thoracic cavity leads to respiratory insufficiency, and these individuals frequently die at birth or soon after. Histologic examination of the growth plate reveals reduced chondrocyte proliferation and disorganization within the zone of proliferation.

Defects in Extracellular Structural Proteins

Mutations in the major bone and cartilage collagens (types I, II, IX, X, and XI) give rise to highly variable presentations ranging from lethal disease to premature osteoarthritis (OA).

Type I Collagen Diseases (Osteogenesis Imperfecta)

Osteogenesis imperfecta (OI), or brittle bone disease, the most common inherited disorder of connective tissue, is a phenotypically heterogenous disorder caused by deficiencies in type I collagen synthesis. OI principally affects bone, but it also impacts other tissues rich in type I collagen (joints, eyes, ears, skin, and teeth). It is caused by mutations in genes encoding the α1 and α2 chains of type I collagen. Many of the over 800 mutations identified lead to replacement of a glycine residue within the triple-helical domain with another amino acid. Collagen synthesis and extracellular transport require triple-helix formation, and these mutations result in misfolding of collagen polypeptides and defective assembly of higher-order collagen chains. Mutant collagens also interfere with assembly of wild-type collagen chains, that is, they exert a dominant negative effect, thereby explaining the autosomal dominant inheritance pattern.

Table 26.1 Diseases of the Skeleton With Identified Genetic Defects

Disorder	Gene Symbol	Affected Molecule	Clinical Phenotype
Defects in Transcription Factors Producing Abnormalities in Mesenchymal Condensation and Related Cell Differentiation			
Brachydactyly types D and E	*HOXD13*	Transcription factor	Short, broad terminal phalanges of first digits
Camptomelic dysplasia	*SOX9*	Transcription factor	Sex reversal, abnormal skeletal development
Cleidocranial dysplasia	*RUNX2*	Transcription factor	Abnormal clavicles, Wormian bones, supernumerary teeth
Holt-Oram syndrome	*TBX5*	Transcription factor	Congenital abnormalities, forelimb anomalies
Nail-patella syndrome	*LMX1B*	Transcription factor	Hypoplastic nails, hypoplastic or aplastic patellas, dislocated radial head, progressive nephropathy
Waardenburg syndrome types 1 and 3	*PAX3*	Transcription factor	Hearing loss, abnormal pigmentation, craniofacial abnormalities
Defects in Hormones and Signal Transduction Proteins Producing Abnormal Proliferation or Maturation of Osteoblasts, Osteoclasts, or Chondrocytes			
Achondroplasia	*FGFR3*	Receptor	Short stature, rhizomelic shortening of limbs, frontal bossing, midface deficiency
Hypochondroplasia	*FGFR3*	Receptor	Disproportionately short stature, micromelia, relative macrocephaly
Osteopetrosis, autosomal dominant	*LRP5*	Receptor	Increased bone density, hearing loss, skeletal fragility
Osteopetrosis, infantile form	*RANKL*	Receptor ligand	Increased bone density
Osteoporosis-pseudoglioma syndrome	*LRP5*	Receptor	Congenital or infant-onset loss of vision, skeletal fragility
Thanatophoric dysplasia	*FGFR3*	Receptor	Severe limb shortening and bowing, frontal bossing, depressed nasal bridge
Defects in Extracellular Structural Proteins			
Achondrogenesis type 2	*COL2A1*	Type II collagen	Short trunk
Metaphyseal dysplasia, Schmid type	*COL10A1*	Type X collagen	Mildly short stature
Osteogenesis imperfecta types 1 to 4	*COL1A1, COL1A2*	Type I collagen	Bone fragility
Defects in Metabolic Enzymes and Transporters			
Osteopetrosis with renal tubular acidosis	*CA2*	Carbonic anhydrase	Increased bone density, fragility, renal tubular acidosis
Osteopetrosis, late onset type 2	*CLCN7*	Chloride channel	Increased bone density, fragility

Modified from Mundlos S, Olsen BR: Heritable diseases of the skeleton. Part I: Molecular insights into skeletal development—transcription factors and signaling pathways, *FASEB J* 11(2):125–132, 1997; Mundlos S, Olsen BR: Heritable diseases of the skeleton. Part II: Molecular insights into skeletal development—matrix components and their homeostasis, *FASEB J* 11(4):227–233, 1997; Superti-Furga A, Bonafé L, Rimoin DL: Molecular-pathogenetic classification of genetic disorders of the skeleton, *Am J Med Genet* 106(4):262–293, 2001; Krakow D, Rimoin DL: The skeletal dysplasias, *Genet Med* 12(6):327–341, 2010.

The fundamental abnormality in OI is too little bone, resulting in extreme skeletal fragility. Other findings include blue sclerae caused by decreased collagen content, making the sclera translucent and allowing partial visualization of the underlying choroid; hearing loss related to both a sensorineural deficit and impeded conduction due to abnormalities in the bones of the middle and inner ear; and dental imperfections (small, misshapen, and blue-yellow teeth) secondary to dentin deficiency.

OI is separated into four major clinical subtypes of varying severity (Table 26.2). Mutations that result in decreased synthesis of qualitatively normal collagen are associated with mild skeletal abnormalities. More severe or lethal phenotypes are associated with mutant collagens that interfere with triple helix formation. The type 2 variant, which is uniformly fatal in utero or during the perinatal period, is characterized by extraordinary bone fragility and multiple intrauterine fractures (Fig. 26.6). In contrast, individuals with type 1 OI have a normal life span but experience childhood fractures that decrease in frequency following puberty.

Defects in Metabolic Pathways (Enzymes, Ion Channels, and Transporters)

Osteopetrosis

Osteopetrosis comprises a group of rare genetic diseases characterized by reduced bone resorption due to deficient osteoclast development or function, which leads to diffuse, symmetric skeletal sclerosis. Although the term osteopetrosis implies that the bones are stonelike, they are actually brittle and fracture easily. Osteopetrosis is classified into variants based on both the mode of inheritance and the severity of clinical findings.

Pathogenesis

Most of the mutations underlying osteopetrosis interfere with acidification of the osteoclast resorption pit, which is required for the dissolution of calcium hydroxyapatite within the matrix. For example, *Albers-Schönberg disease,* a mild autosomal dominant form of osteopetrosis, is caused by mutation of *CLCN7,* which encodes a proton-chloride

Table 26.2 Subtypes of Osteogenesis Imperfecta

Subtype	Collagen Defect	Inheritance	Major Clinical Features	Prognosis
I	Decreased synthesis of pro-α1(I) chain Abnormal pro-α1(I) or pro-α2(I) chains	Autosomal dominant	Postnatal fractures, blue sclera Normal stature Skeletal fragility Dentinogenesis imperfecta Hearing impairment Joint laxity Blue sclerae	Compatible with survival
II	Abnormally short pro-α1(I) chain Unstable triple helix Abnormal or insufficient pro-α2(I)	Most autosomal recessive Some autosomal dominant New mutations	Death in utero or within days of birth Skeletal deformity with excessive fragility and multiple fractures Blue sclera	Perinatal lethal
III	Altered structure of pro-peptides of pro-α2(I) Impaired formation of triple helix	Autosomal dominant (75%) Autosomal recessive (25%)	Compatible with survival Growth retardation Multiple fractures Progressive kyphoscoliosis Blue sclera at birth that become white Hearing impairment Dentinogenesis imperfecta	Progressive, deforming
IV	Short pro-α2(I) chain Unstable triple helix	Autosomal dominant	Postnatal fractures, normal sclerae Moderate skeletal fragility Short stature Sometimes dentinogenesis imperfecta	Compatible with survival

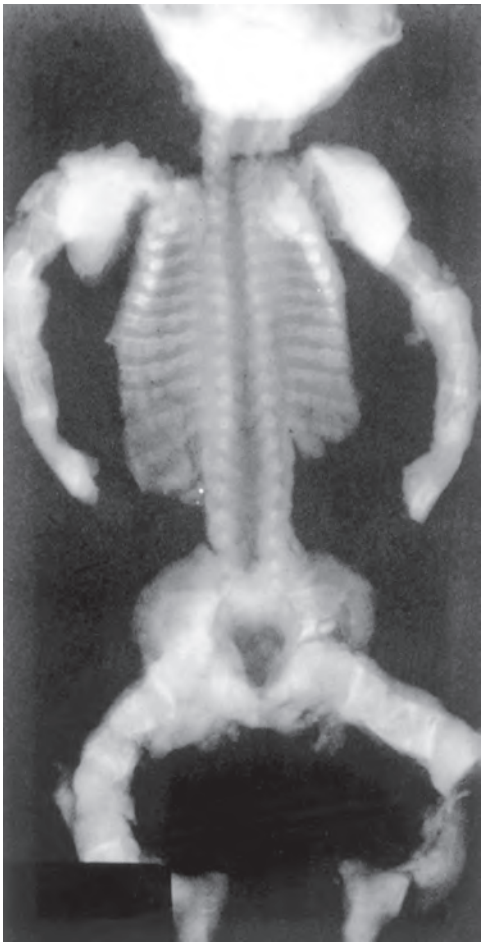


Figure 26.6 Skeletal radiograph of a fetus with lethal type 2 osteogenesis imperfecta. Note the numerous fractures of virtually all bones, resulting in accordion-like shortening of the limbs.

exchanger on the osteoclast surface that is required for resorption pit acidification. Similarly, most cases of autosomal recessive osteopetrosis are caused by a mutation of *TCIRG1,* which encodes a subunit of the osteoclast vacuolar H^+-ATPase that is also necessary for acidification of the resorption pit. Other causes of autosomal recessive osteopetrosis include defects in carbonic anhydrase 2 (CA2), which, like all carbonic anhydrase isozymes, generates protons and bicarbonate from carbon dioxide and water. CA2 facilitates resorption pit acidification by osteoclasts and urinary acidification by renal tubular epithelial cells. Osteopetrosis due to *CA2* mutations is, therefore, accompanied by renal tubular acidosis. Mutations in *IKBKG,* which encodes NEMO, the regulatory subunit of the inhibitor of kappaB kinase (IKK) complex that is involved in NF-κB activation, is a cause of osteopetrosis that is not due to defective acidification. NEMO, which is X-linked, is required for osteoclastogenesis and osteoclast survival, which are regulated by NF-κB. Because NF-κB serves many other functions, *IKBKG* mutations result in a multisystem disorder called X-linked anhidrotic ectodermal dysplasia with immunodeficiency that includes osteopetrosis.

MORPHOLOGY

Due to deficient osteoclast activity, bones involved by osteopetrosis lack a medullary canal, and the ends of long bones are bulbous (Erlenmeyer flask deformity) and misshapen (Fig. 26.7). The neural foramina are small and can compress exiting nerves. The primary spongiosa, which is normally removed during growth, persists and fills the medullary cavity, leaving no room for the hematopoietic marrow and preventing the formation of mature trabeculae (Fig. 26.8). Deposited bone is not remodeled and tends to be woven rather than lamellar. Depending on the underlying genetic defect, the number of osteoclasts may be normal, increased, or decreased.

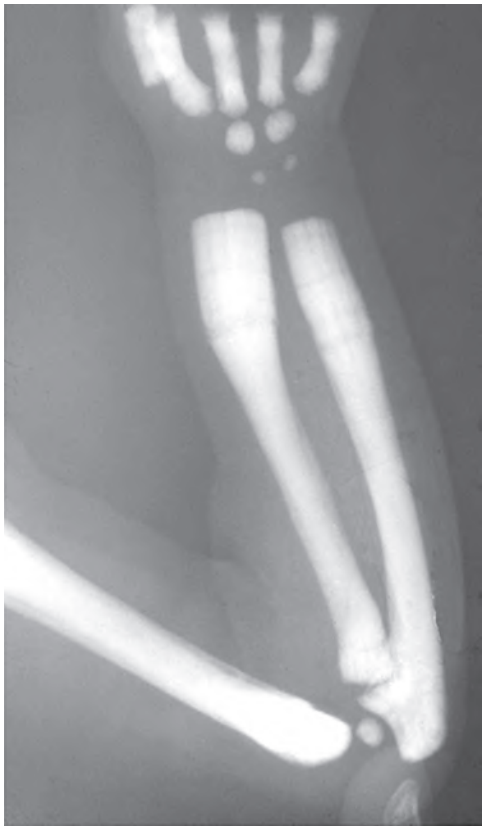


Figure 26.7 Radiograph of the upper extremity in an individual with osteopetrosis. The bones are diffusely sclerotic, and the distal metaphyses of the ulna and radius are poorly formed.

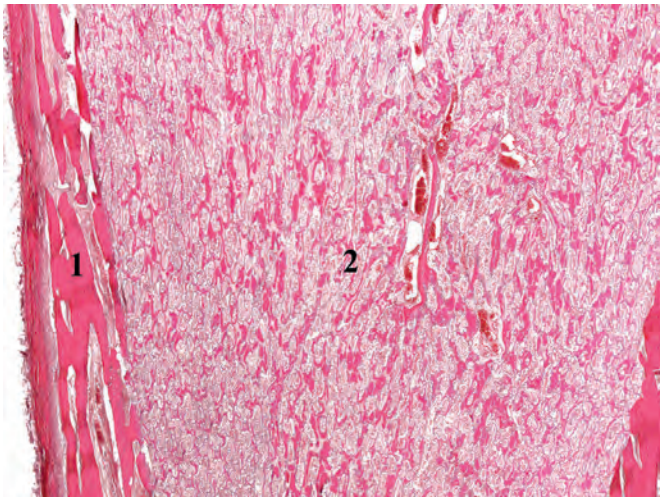


Figure 26.8 Section of proximal tibial diaphysis from a fetus with osteopetrosis. The cortex (1) is present, but the medullary cavity (2) is filled with primary spongiosa, which replaces the hematopoietic elements.

Clinical Features

Severe infantile osteopetrosis is an autosomal recessive disorder that usually becomes evident in utero or soon after birth. Fracture, anemia, and hydrocephaly are often seen, resulting in postpartum mortality. Affected individuals who survive into their infancy have cranial nerve defects (optic atrophy, deafness, and facial paralysis) and repeated—often fatal—infections because of leukopenia due to reduced marrow space and decreased hematopoiesis. The compensatory extramedullary hematopoiesis can lead to prominent hepatosplenomegaly. Autosomal dominant forms are milder and may not be detected until adolescence or adulthood when discovered on x-ray studies, either incidentally or because of repeated fractures.

Osteopetrosis was the first genetic bone disease treated with hematopoietic stem cell transplantation, which is effective because osteoclasts are derived from hematopoietic precursors. The normal osteoclasts produced from donor stem cells reverse many of the skeletal abnormalities.

Diseases Associated With Defects in Degradation of Macromolecules

Mucopolysaccharidoses

The mucopolysaccharidoses, discussed in Chapter 5, are lysosomal storage diseases. They are caused by deficiencies in enzymes, primarily acid hydrolases, that degrade dermatan sulfate, heparan sulfate, and keratan sulfate. Mesenchymal cells, particularly chondrocytes, degrade extracellular matrix mucopolysaccharides. In these diseases, mucopolysaccharides accumulate within chondrocytes and induce apoptosis. Extracellular mucopolysaccharide accumulation leads to structural defects in articular cartilage. Consequently, many of the skeletal manifestations of the mucopolysaccharidoses result from abnormalities in the cartilage anlage, growth plates, costal cartilages, and articular surfaces. Affected individuals are frequently of short stature and have chest wall abnormalities and malformed bones.

KEY CONCEPTS

DEVELOPMENTAL DISORDERS OF BONE AND CARTILAGE

Abnormalities in a single bone or a localized group of bones are called dysostoses and arise from defects in mesenchyme migration and condensation. They manifest as absent, supernumerary, or abnormally fused bones. Global disorganization of bone and/or cartilage is termed dysplasia. Developmental abnormalities can be categorized by the associated genetic defect.

- Transcription factors: Homeobox genes, such as *HOXD13*, are frequently mutated in brachydactyly syndromes.
- Signal transduction molecules: *FGFR3* mutations are responsible for achondroplasia and thanatophoric dysplasia, both of which manifest as dwarfism.
- Structural proteins: Mutations in the genes for type I collagen underlie most types of osteogenesis imperfecta (brittle bone disease), characterized by defective bone formation and skeletal fragility.
- Metabolic enzymes and transporters: Mutations that interfere with osteoclast acidification of the resorption pit or osteoclastogenesis cause osteopetrosis, in which bones are hard but brittle. Depending on the gene mutated, extraskeletal manifestations may also be present.

METABOLIC DISEASES OF BONE

Osteopenia and Osteoporosis

Osteopenia refers to decreased bone mass; osteoporosis is defined as osteopenia that is severe enough to significantly increase the risk of fracture. Radiographically, osteoporosis is considered bone mass at least 2.5 standard deviations below mean peak bone mass in young adults. Osteopenia is 1 to 2.5 standard deviations below the mean. The disorder may be localized to a certain bone or region, as in disuse osteoporosis of a limb, or may involve the entire skeleton as a manifestation of a metabolic bone disease. Generalized osteoporosis may be primary or secondary to a variety of conditions (Table 26.3). Each year, approximately 1 million Americans experience an osteoporosis-related fracture with an estimated cost of more than $14 billion. The following discussion relates largely to these forms of osteoporosis.

Pathogenesis

Peak bone mass is achieved during young adulthood. Its magnitude is determined largely by hereditary factors, especially polymorphisms in the genes that influence bone metabolism (Fig. 26.9). Physical activity, muscle strength, diet, and hormonal state also make important contributions.

Table 26.3 Categories of Generalized Osteoporosis

Primary
Idiopathic
Postmenopausal
Senile
Secondary
Endocrine Disorders
Addison disease
Diabetes, type I
Hyperparathyroidism
Hyperthyroidism
Hypothyroidism
Pituitary tumors
Neoplasia
Carcinomatosis
Multiple myeloma
Gastrointestinal
Hepatic insufficiency
Malabsorption
Malnutrition
Vitamin C, D deficiencies
Drugs
Alcohol
Anticoagulants
Anticonvulsants
Chemotherapy
Corticosteroids
Miscellaneous
Anemia
Homocystinuria
Immobilization
Osteogenesis imperfecta
Pulmonary disease

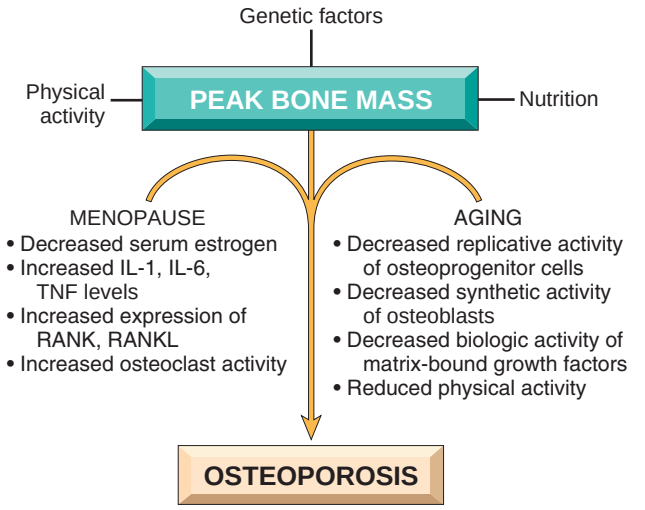


Figure 26.9 Pathophysiology of postmenopausal and senile osteoporosis *(see text).*

Once maximal skeletal mass is attained, bone resorption slightly exceeds formation, resulting in age-related bone loss that averages 0.7% per year. Both sexes are affected equally, but the process is more rapid on average in Caucasians than in those of African descent.

Although numerous factors affect bone mass, the most common forms of osteoporosis are senile and postmenopausal.

- *Age-related changes* include a reduced proliferative and biosynthetic capacity, and attenuated response to growth factors of osteoblasts, resulting in a diminished ability to make bone. This form of osteoporosis, known as *senile osteoporosis,* is categorized as a low-turnover variant.
- *Reduced physical activity* increases the rate of bone loss in experimental animals and humans, because mechanical forces normally stimulate bone remodeling. Bone loss in an immobilized or paralyzed extremity or in astronauts experiencing reduced gravitational forces for prolonged periods, and, conversely, increased bone density in athletes, exemplify the importance of physical forces in bone maintenance. When considering physical activity, the effect of load magnitude on bone density is greater than load repetition, explaining why resistance exercises, such as weight training, increase bone mass more effectively than endurance activities such as bicycling. The decreased physical activity that is associated with normal aging contributes to senile osteoporosis.
- *Genetic factors* such as single-gene defects are rare causes of osteoporosis. However, polymorphisms in certain genes may contribute to variation in peak bone density within populations. The most strongly linked genes identified by genome-wide association studies include *RANK, RANKL,* and *OPG,* all of which encode key osteoclast regulators; the HLA locus (for unknown reasons); and the estrogen receptor gene (discussed later).
- *Calcium nutritional state* contributes to peak bone mass. Adolescent girls (more than boys) tend to have insufficient calcium intake in the diet. If it occurs during a period of rapid bone growth, calcium deficiency reduces peak bone mass and increases the risk of osteoporosis. Relative

deficiencies of calcium and vitamin D and elevated PTH levels may also contribute to development of senile osteoporosis.

- *Hormonal influences.* In the decade after menopause, up to 2% of cortical bone and 9% of cancellous bone may be lost each year. Estrogen deficiency plays the major role in this phenomenon, and close to 40% of postmenopausal women are affected by *postmenopausal osteoporosis*. Although decreased estrogen increases both bone formation and resorption, the latter dominates, resulting in high-turnover osteoporosis. Estrogen loss leads to increased secretion of inflammatory cytokines, such as IL-6, TNF, and IL-1, by innate immune cells in the blood and marrow through unknown mechanisms. The resulting increases and decreases in RANKL and OPG, respectively, stimulate osteoclast recruitment and activity.

MORPHOLOGY

The hallmark of osteoporosis is histologically normal bone that is decreased in quantity. The entire skeleton is affected (Fig. 26.10), but certain bones tend to be more severely impacted. Postmenopausal osteoporosis affects mainly bones or portions of bones that have increased surface area, such as the cancellous compartment of vertebral bodies. The trabecular plates become perforated and thinned, and lose their interconnections (Fig. 26.11), leading to microfractures and eventually to vertebral collapse. In senile osteoporosis, the cortex is thinned by subperiosteal and endosteal resorption, and the Haversian systems are widened.

Clinical Features

The clinical manifestations of osteoporosis depend on which bones are involved. Vertebral fractures that frequently occur in the thoracic and lumbar regions are painful, and, when multiple, can cause significant loss of height and deformities such as lumbar lordosis and kyphoscoliosis. Fractures of the femoral neck, pelvis, or spine lead to immobilization and complications such as pulmonary embolism and pneumonia, resulting in 40,000 to 50,000 deaths per year.

Osteoporosis cannot be reliably detected in plain radiographs until 30% to 40% of the bone mass is lost. The best estimates of bone loss, aside from biopsy (which is rarely performed), are specialized radiographic imaging techniques, such as dual-energy x-ray absorptiometry and quantitative computed tomography, both of which measure bone density.

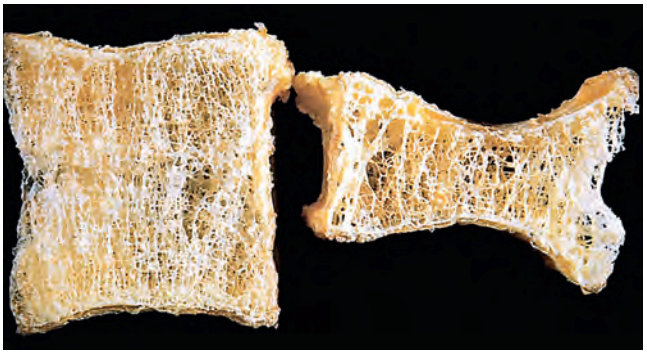


Figure 26.10 Osteoporotic vertebral body *(right)* shortened by compression fractures compared with a normal vertebral body *(left)*. Note that the osteoporotic vertebra has a characteristic loss of horizontal trabeculae and thickened vertical trabeculae.

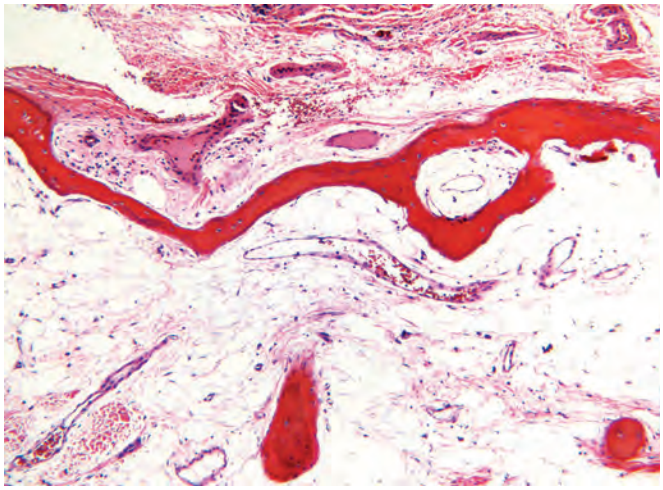


Figure 26.11 In advanced osteoporosis, both the trabecular bone of the medulla *(bottom)* and the cortical bone *(top)* are markedly thinned.

Preventive and therapeutic management of osteoporosis includes exercise, appropriate calcium and vitamin D intake, and pharmacologic agents, most commonly bisphosphonates, which reduce osteoclast activity and induce their apoptosis. Denosumab, an anti-RANKL antibody, has shown promise in treating some forms of postmenopausal osteoporosis. Although menopausal hormone therapy has been used to prevent osteoporosis and associated fractures, complications, particularly deep vein thrombosis and stroke, have prompted search for more selective estrogen receptor modulators.

Osteomalacia and Rickets

Osteomalacia and rickets are manifestations of impaired mineralization of bone matrix. This contrasts with osteoporosis, in which mineralization of bone is normal and bone mass is decreased. Most examples of undermineralized matrix result from abnormal vitamin D metabolism or vitamin D deficiency (detailed in Chapter 9). Rickets refers to the disorder in children, in whom it interferes with the deposition of bone in the growth plates. Osteomalacia is the adult counterpart, in which bone formed during remodeling is undermineralized and predisposed to fractures.

Hyperparathyroidism

Hyperparathyroidism causes increased bone resorption. As discussed in Chapter 24, parathyroid hormone (PTH) has a central role in calcium homeostasis through the following effects:

- *Activation of osteoclasts,* increasing bone resorption and calcium mobilization. PTH mediates the effect indirectly by increasing RANKL expression on osteoblasts.
- *Increasing calcium resorption* by the renal tubules
- *Increasing urinary phosphate excretion*
- *Increasing synthesis of active vitamin D, $1,2_5$(OH)2-D,* by the kidneys, thereby enhancing intestinal calcium absorption and mobilizing bone calcium by inducing RANKL expression on osteoblasts

The net result of PTH action is elevated serum calcium, which normally inhibits PTH production. Excessive or inappropriate PTH release may stem from autonomous parathyroid secretion *(primary hyperparathyroidism)* or may be the result of renal disease *(secondary hyperparathyroidism)* (see Chapter 24). In either setting, hyperparathyroidism produces changes throughout the skeleton as a consequence of unrestrained osteoclast activity. Elevated PTH is responsible for bone changes in primary hyperparathyroidism, but additional factors contribute in secondary hyperparathyroidism. In chronic renal insufficiency, 1,25-$(OH)_2$- vitamin D synthesis is reduced due to decreased α1-hydroxylase activity, which stems from loss of renal function and the suppressive effects of hyperphosphatemia on α_1-hydroxylase. Inadequate vitamin D ultimately limits intestinal calcium absorption. Secondary hyperparathyroidism can also be complicated by metabolic acidosis and aluminum deposition in bone. With time, loss of bone mass increases susceptibility to fractures, bone deformation, and joint problems.

MORPHOLOGY

Symptomatic, untreated primary hyperparathyroidism manifests with three interrelated skeletal abnormalities: osteoporosis, brown tumors, and osteitis fibrosa cystica. The **osteoporosis** is generalized but is most severe in the phalanges, vertebrae, and proximal femur. Hyperparathyroidism most prominently enhances osteoclast activity within cortical bone (subperiosteal and endosteal surfaces), but cancellous bone can also be affected. At these sites, osteoclasts may tunnel into and dissect centrally along the length of the trabeculae, leaving adjacent marrow spaces to be replaced by fibrovascular tissue, producing **dissecting osteitis** (Fig. 26.12). Radiographically, dissecting osteitis is seen as decreased bone density or osteoporosis.

Bone loss in hyperparathyroidism predisposes to microfractures, secondary hemorrhage, macrophage recruitment, and ingrowth of reparative fibrous tissue to create a mass lesion called **brown tumor** (Fig. 26.13). The brown color reflects vascularity, hemorrhage, and hemosiderin deposition. Cystic degeneration of brown tumors is common. The combination of increased bone cell activity, peritrabecular fibrosis, and cystic brown tumors is the hallmark of severe hyperparathyroidism and is known as **generalized osteitis fibrosa cystica (von Recklinghausen disease of bone)**.

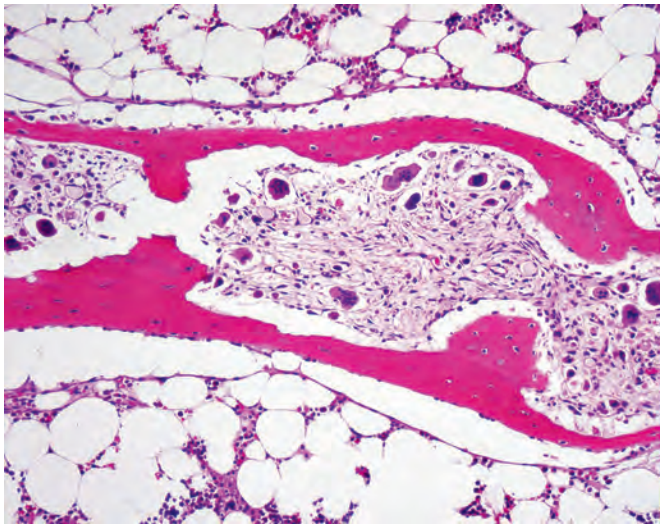


Figure 26.12 Hyperparathyroidism with osteoclasts boring into the center of the trabeculum (dissecting osteitis).

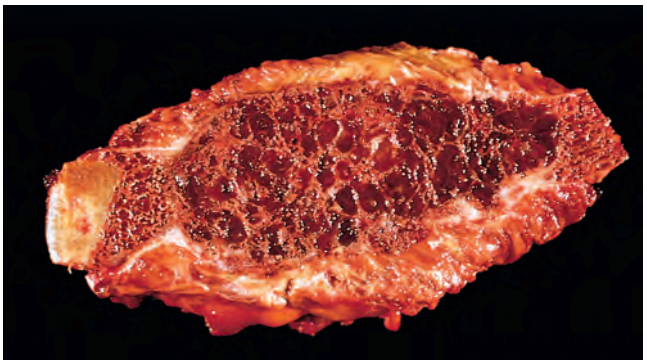


Figure 26.13 Resected rib, harboring an expansile brown tumor adjacent to the costal cartilage.

Osteitis fibrosa cystica is now uncommon because hyperparathyroidism is usually diagnosed on routine blood tests and treated at an early stage. Although secondary hyperparathyroidism can cause similar changes, the process is usually less severe, and skeletal abnormalities tend to be milder. Bony changes regress or disappear completely when hyperparathyroidism is controlled.

Renal Osteodystrophy

The term renal osteodystrophy describes the collective skeletal changes that occur in chronic renal disease, including those associated with dialysis. The manifestations include many of the entities described earlier including (1) osteopenia/osteoporosis, (2) osteomalacia, (3) secondary hyperparathyroidism, and (4) growth retardation. As medical advances have prolonged the lives of individuals with renal disease, the impact of this disease on skeletal homeostasis has assumed greater clinical importance. Histologic bone changes in individuals with end-stage renal failure can be divided into three major types of disorders:

- *High-turnover osteodystrophy* is characterized by increased bone resorption and bone formation, with the former predominating.
- *Low-turnover* or *aplastic disease* is manifested by adynamic bone (little osteoclastic and osteoblastic activity) and, less commonly, osteomalacia.
- *Mixed pattern disease* with areas of high turnover and low turnover.

Pathogenesis

Kidney disease causes skeletal abnormalities through three mechanisms (Fig. 26.14):

- *Tubular dysfunction* most commonly leading to renal tubular acidosis. The associated systemic acidosis dissolves hydroxyapatite, resulting in matrix demineralization and osteomalacia.
- *Secondary hyperparathyroidism,* due to reduced phosphate excretion, chronic hyperphosphatemia, and hypocalcemia.

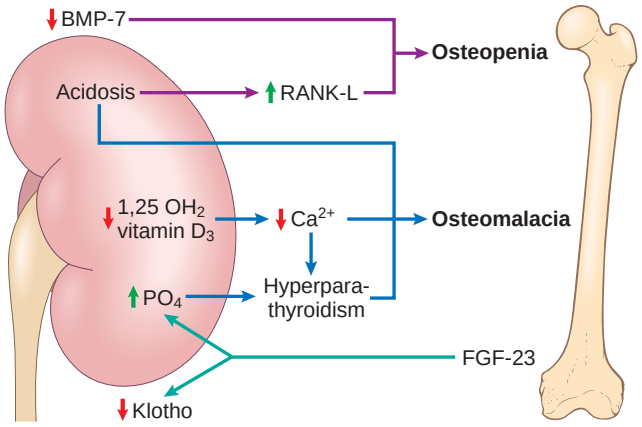


Figure 26.14 Mechanisms of renal osteodystrophy involve homeostasis and endocrine signaling between bone and kidney.

As discussed earlier, the resulting metabolic state is not completely analogous to primary hyperparathyroidism, and bone volume, turnover, and mineralization can vary independently.

- *Decreased biosynthetic function,* including reduced renal vitamin D hydroxylation to generate 1,25-OH_2-vitamin D_3. This results in hypocalcemia that contributes to secondary hyperparathyroidism. A hormonal feedback loop between kidney and bone that regulates calcium and phosphate homeostasis involves the secreted proteins BMP-7 and FGF-23 and the membrane protein Klotho. BMP-7, produced by renal tubular cells, induces osteoblast differentiation and proliferation, whereas FGF-23, made by osteocytes, acts on the kidney to regulate Klotho-dependent renal phosphate homeostasis and vitamin D hydroxylation. Chronic renal failure disrupts this signaling axis and contributes to osteopenia and osteomalacia.

KEY CONCEPTS

METABOLIC DISEASES OF BONE

- Osteopenia and osteoporosis are conditions in which bone is normally mineralized but decreased in quantity. Osteoporosis is defined as bone loss sufficient to increase fracture risk and is associated with significant morbidity and mortality from fractures. Multiple factors including peak bone mass, age, activity, genetics, nutrition, and hormonal influences contribute to its pathogenesis.
- Osteomalacia is defined by the presence of bone that is insufficiently mineralized. In the developing skeleton, this results in rickets.
- Hyperparathyroidism arises from either autonomous or compensatory hypersecretion of PTH and can lead to osteoporosis, brown tumors, and osteitis fibrosa cystica. In high income countries, where early diagnosis is the norm, these manifestations are rare.
- Renal osteodystrophy is marked by a constellation of bone abnormalities (osteopenia, osteomalacia, hyperparathyroidism, and growth retardation) that occur as a consequence of chronic renal failure. The bone changes stem from decreased tubular, glomerular, and hormonal renal functions.

Paget Disease (Osteitis Deformans)

Paget disease is a disorder marked by increased, but disordered and structurally unsound, bone mass. It develops in three sequential phases: (1) an initial osteolytic stage; (2) a mixed osteoclastic-osteoblastic stage; and (3) a burned-out quiescent osteosclerotic stage in which osteoblast activity predominates (Fig. 26.15).

Paget disease often presents in late adulthood and becomes progressively more common with increasing age. An intriguing aspect is the striking geographic variation in its prevalence. It is relatively common in Caucasians in England, France, Austria, regions of Germany, Australia, New Zealand, and the United States, and is rare in native populations of Scandinavia, China, Japan, and Africa. Its

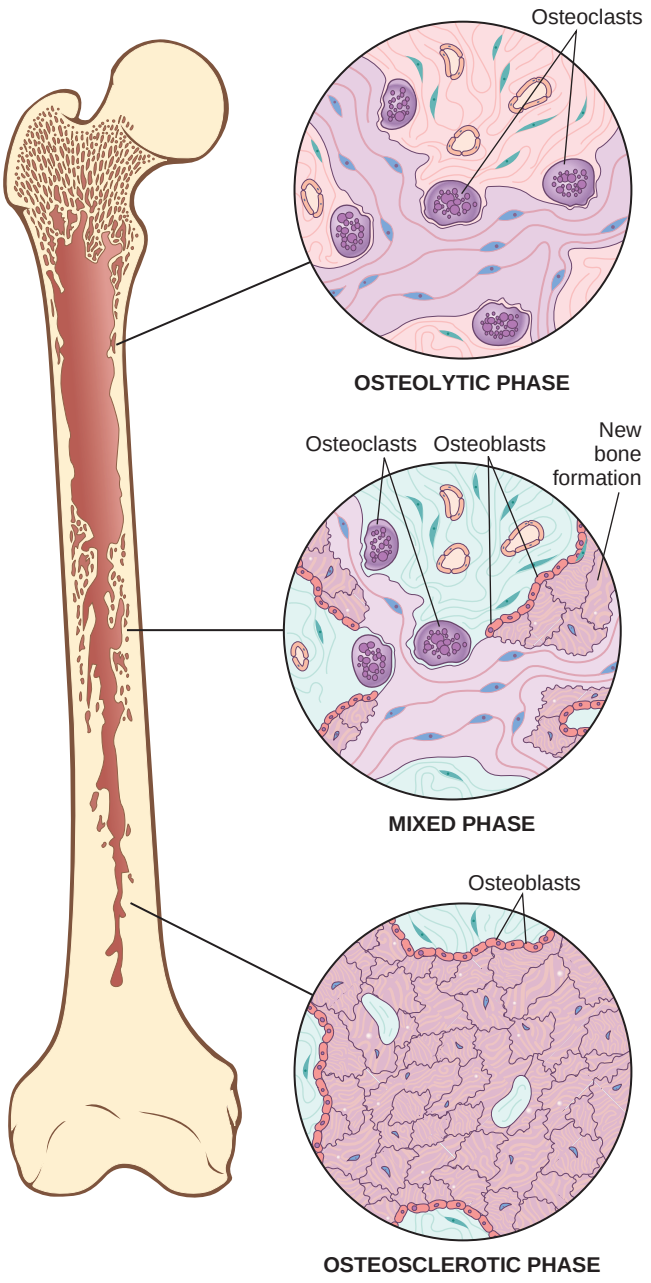


Figure 26.15 Diagrammatic representation of Paget disease of bone demonstrating the three phases of disease evolution.

exact incidence is difficult to determine because many affected individuals are asymptomatic; it is estimated that 1% of the US population older than 40 years of age is affected. The prevalence in England is 2.5% for men and 1.6% for women 55 years of age or older. A decrease in new cases has been observed in some countries over the past 30 years.

Pathogenesis

The cause of Paget disease remains uncertain, but current evidence suggests both genetic and environmental contributions. Approximately 50% of cases of familial Paget disease and 10% of sporadic cases are associated with mutations in the *SQSTM1* gene that increase NF-κB activity and thus enhance osteoclast activity. Activating *RANK* mutations and inactivating *OPG* mutations account for some cases of juvenile Paget disease. In vitro studies suggest that chronic infection of osteoclast precursors by measles or other RNA viruses may also play a role.

MORPHOLOGY

Paget disease shows remarkable histologic variation over time and across sites. Its hallmark is a mosaic pattern of lamellar bone that develops in the sclerotic phase. This jigsaw puzzle–like appearance is produced by unusually prominent **cement lines,** which join haphazardly oriented units of lamellar bone (Fig. 26.16). The features are less specific during other phases of disease, but include waves of osteoclastic activity and numerous resorption pits in the lytic phase. The osteoclasts are abnormally large and have many more than the normal 10 to 12 nuclei; sometimes 100 nuclei are present. Osteoclasts persist in the mixed phase, but many of the bone surfaces are also lined by plump osteoblasts. The marrow adjacent to the bone-forming surface is replaced by loose connective tissue that contains osteoprogenitor cells and numerous blood vessels. The newly formed bone may be woven or lamellar, but eventually it is all remodeled into lamellar bone. As the mosaic pattern takes hold and cell activity decreases, periosseous fibrovascular tissue recedes and is replaced by normal marrow. In the end, the coarsely thickened trabeculae and soft and porous cortices lack structural stability and make the bone vulnerable to fractures.

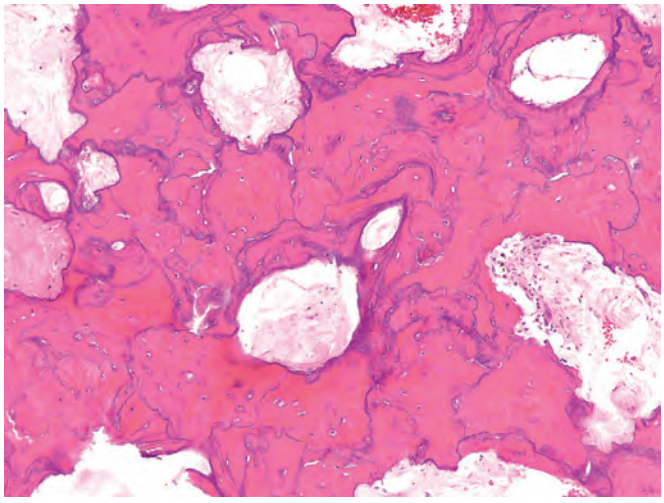


Figure 26.16 Mosaic pattern of lamellar bone pathognomonic in Paget disease.

Clinical Features

Clinical findings are extremely variable and depend on the extent and site of the disease. Paget disease is monostotic in about 15% of cases and polyostotic in the remainder. The axial skeleton or proximal femur is involved in up to 80% of cases. Most cases are asymptomatic and are discovered as an incidental radiographic finding. Pain localized to the affected bone may be present due to microfractures or bone overgrowth that compresses spinal and cranial nerve roots. Enlargement of the craniofacial skeleton may produce leontiasis ossea (lion face) and a cranium so heavy that it is difficult for the person to hold the head erect. The weakened Pagetic bone may lead to invagination of the skull base (platybasia) and compression of the posterior fossa. Weight bearing causes anterior bowing of the femurs and tibiae and distorts the femoral heads, resulting in the development of severe secondary osteoarthritis. Chalk stick–type fractures are common and usually involve the long bones of the lower extremities. Vertebral compression fractures can result in spinal cord injury and kyphosis. Rarely, the hypervascularity of Pagetic bone warms the overlying skin, and in severe polyostotic disease the increased blood flow can act as an arteriovenous shunt, leading to high-output heart failure or exacerbation of underlying cardiac disease.

The most dreaded complication of Paget disease is sarcoma, which occurs in less than 1% of all individuals with Paget disease, and in 5% to 10% of those with severe polyostotic disease. The sarcomas, usually osteosarcoma or fibrosarcoma, arise in Paget lesions in the long bones, pelvis, skull, and spine.

The diagnosis of Paget disease can be made radiographically. Pagetic bone is typically enlarged with thick, coarsened cortices and medulla (Fig. 26.17). Active disease has a wedge-shaped lytic leading edge that may progress along the length of the bone at a rate of 1 cm per year. Many affected individuals have elevated serum alkaline phosphatase levels, but serum calcium and phosphate levels are normal.

In the absence of malignant transformation, Paget disease is usually not serious or life-threatening. Most affected individuals have mild symptoms that are readily suppressed by treatment with calcitonin and bisphosphonates.

FRACTURES

Fractures are defined as loss of bone integrity. They are some of the most common pathologic conditions affecting bone. The following qualifiers describe fracture types and affect treatment:

- *Simple:* the overlying skin is intact.
- *Compound:* the bone communicates with the skin surface.
- *Comminuted:* the bone is fragmented.
- *Displaced:* the ends of the bone at the fracture site are not aligned.
- *Stress:* a slowly developing fracture that follows a period of increased physical activity in which the bone is subjected to repetitive loads.
- *Greenstick:* extending only partially through the bone, common in infants when bones are soft.

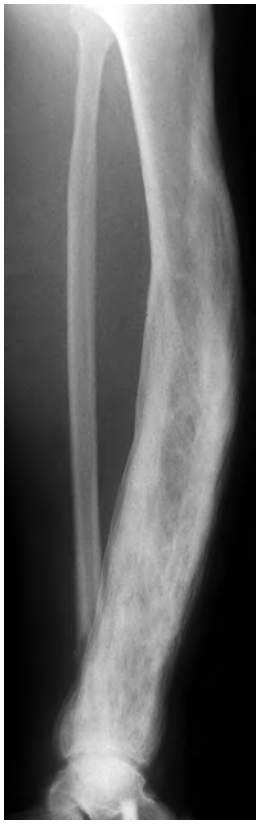


Figure 26.17 Severe Paget disease. The tibia is bowed. The affected portion is enlarged and sclerotic, and it exhibits irregular thickening of both cortical and cancellous bone.

- *Pathologic:* involving bone weakened by an underlying disease process, such as a tumor.

Healing of Fractures

Bone has a remarkable capacity for repair. This process involves regulated expression of a multitude of genes and can be separated into overlapping stages. Immediately after fracture, rupture of blood vessels results in a hematoma that fills and surrounds the area of injury (Fig. 26.18). The clot provides a fibrin mesh that seals the fracture site and provides a framework for the inflammatory cell influx, fibroblast ingrowth, and capillary proliferation that characterize granulation tissue. Release of PDGF, TGF-β, FGF, and other growth factors by degranulated platelets and inflammatory cells activates osteoprogenitor cells in the periosteum, medullary cavity, and surrounding soft tissues to stimulate osteoclastic and osteoblastic activity. Uncalcified tissue known as *soft tissue callus* or *procallus* forms, providing some anchorage but not structural rigidity for weight bearing.

Within 2 weeks of injury, the activated osteoprogenitor cells deposit subperiosteal trabeculae of woven bone oriented perpendicular to the cortical axis and within the medullary cavity. These processes transform the procallus into *bony callus,* which reaches maximal girth at the end of the second or third week and helps stabilize the fracture site. Activated soft tissue mesenchymal cells may also differentiate into chondrocytes that produce fibrocartilage and hyaline cartilage. Endochondral ossification creates a contiguous

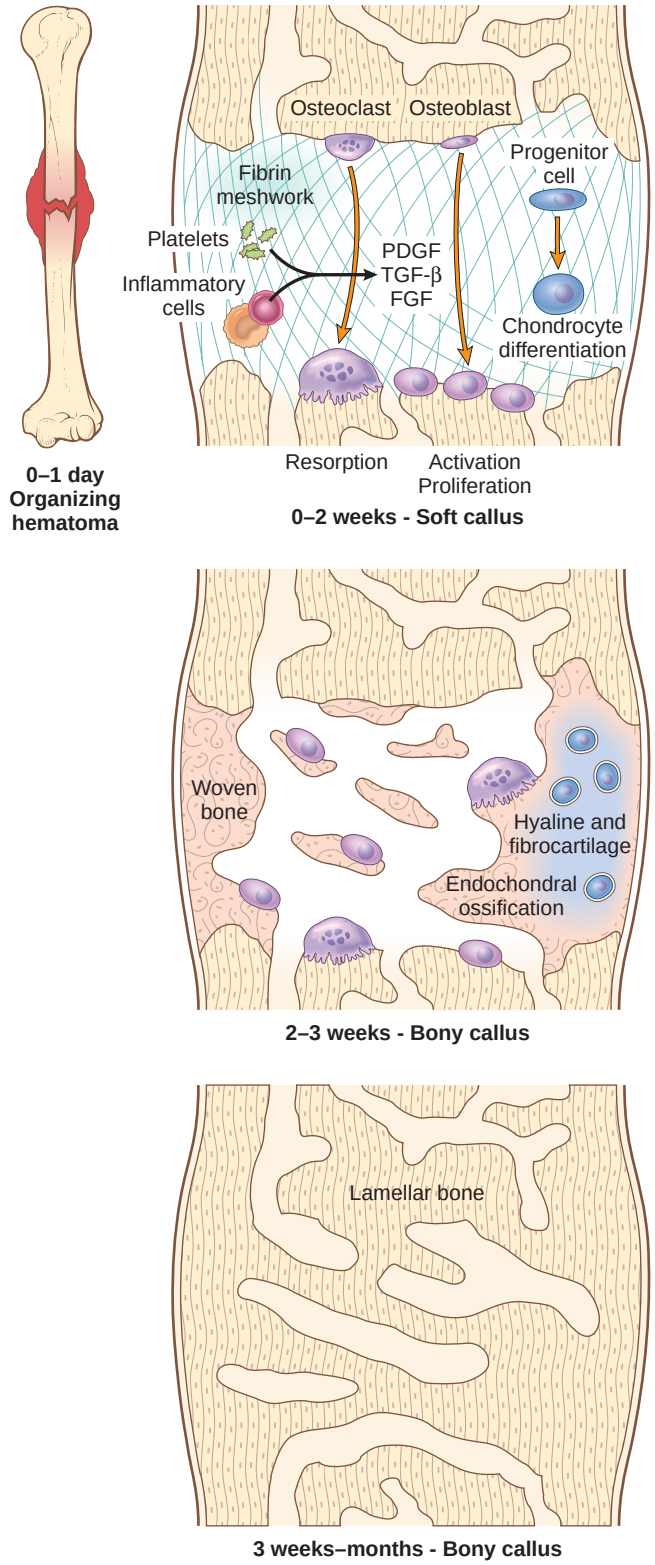


Figure 26.18 The reaction to a fracture begins with an organizing hematoma. Within 2 weeks, the two ends of the bone are bridged by a fibrin meshwork in which osteoclasts, osteoblasts, and chondrocytes differentiate from precursors. These cells produce cartilage and bone matrix, which, with adequate immobilization, remodels into normal lamellar bone.

network of bone and newly deposited bone trabeculae in the medulla and beneath the periosteum. As a result, fractured bone ends are bridged and, with progressive mineralization, the stiffness and strength of the callus increases to allow weight bearing.

In the early stages of callus formation, excess fibrous tissue, cartilage, and woven bone is produced. The portions that are not subjected to physical stress are resorbed as the callus matures, which both reduces the size of the healing bone and recreates lamellar bone. The healing process is completed by restoration of the medullary cavity.

In children and young adults, near-perfect union is the norm, although some deformity typically persists after healing of displaced and comminuted fractures. In older adults, fractures often occur in the background of other bone disorders (e.g., osteoporosis and osteomalacia). In such settings, surgical immobilization is often needed for adequate repair. Other factors may also interfere with healing. Inadequate immobilization, which permits movement of the callus and interferes with normal maturation, can result in *delayed union* or *nonunion*. If nonunion persists, the malformed callus undergoes cystic degeneration and the luminal surface can become lined by synovial-like cells, creating a false joint or *pseudoarthrosis*. Infection of the fracture site, especially common in open fractures, is another serious obstacle to healing, as are malnutrition and skeletal dysplasia.

OSTEONECROSIS (AVASCULAR NECROSIS)

Infarction of bone and marrow is relatively common. It can be limited to the medullary cavity or involve both the medulla and cortex. Fractures and corticosteroid administration are the two most common causes, but many other conditions also predispose to osteonecrosis, including alcohol abuse, bisphosphonate therapy, connective tissue disease, chronic pancreatitis, Gaucher disease, pregnancy, radiation therapy, sickle cell crisis (Chapter 14), tumors, and dysbarism (e.g., decompression sickness).

MORPHOLOGY

Medullary infarcts are geographic in shape and involve both trabecular bone and marrow. Collateral blood flow usually limits cortical involvement. In subchondral infarcts, a triangular or wedge-shaped segment with the subchondral bone plate as its base undergoes necrosis; the overlying articular cartilage remains viable due to nutrients within synovial fluid. Microscopically, dead bone is characterized by empty lacunae surrounded by necrotic adipocytes. The released fatty acids bind calcium and form insoluble calcium soaps. The remaining trabeculae act as scaffolding for deposition of new bone, while osteoclasts resorb necrotic trabeculae. The slow pace of substitution in subchondral infarcts (Fig. 26.19) results in collapse of the necrotic bone, fracture, and sloughing of the articular cartilage.

Clinical Features

Symptoms depend on location and extent of infarction. Typically, subchondral infarcts cause pain that is initially associated with activity but becomes constant as secondary changes supervene. Subchondral infarcts often collapse, resulting in secondary osteoarthritis. Medullary infarcts are usually small and clinically silent.

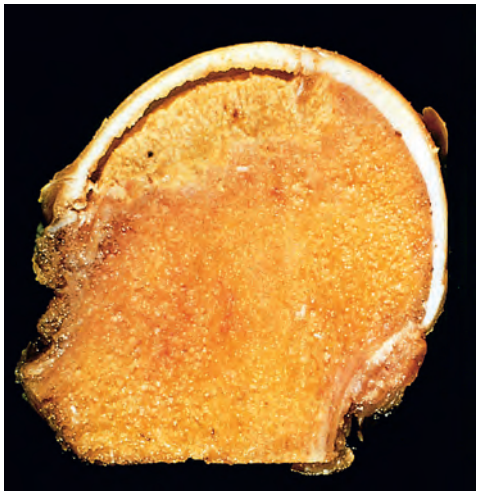


Figure 26.19 Femoral head with a subchondral, wedge-shaped pale area of osteonecrosis. The space between the overlying articular cartilage and bone is caused by trabecular compression fractures without repair.

OSTEOMYELITIS

Osteomyelitis denotes inflammation of bone and marrow, virtually always secondary to infection. Osteomyelitis frequently manifests as a primary solitary focus of disease, but can also be a complication of any systemic infection. All types of organisms, including viruses, parasites, fungi, and bacteria, can produce osteomyelitis, but certain pyogenic bacteria and mycobacteria are the most common culprits. In the United States, unusual infections in immigrants from lower income countries and opportunistic infections in immunosuppressed individuals have made diagnosis and treatment of osteomyelitis challenging.

Pyogenic Osteomyelitis

Pyogenic osteomyelitis is almost always caused by bacterial infection. Organisms may reach the bone by hematogenous spread, extension from a contiguous site, or direct implantation (e.g., in traumatic injury). In otherwise healthy children, osteomyelitis is most often due to hematogenous spread and involves long bones. The initiating bacteremia may stem from seemingly trivial mucosal injuries, such as may occur during defecation or vigorous chewing of hard foods, or from minor skin infections. In adults, osteomyelitis more often occurs as a complication of open fractures, surgical procedures, and infections of the feet in the setting of diabetes. Osteomyelitis is particularly worrisome in infants, in whom epiphyseal infection can spread to a joint, resulting in septic or suppurative arthritis, articular cartilage destruction, and permanent disability.

Staphylococcus aureus is responsible for 80% to 90% of culture-positive pyogenic osteomyelitis. Bacterial cell wall proteins bind to bone matrix components, such as collagen, which facilitates adherence to bone. *Escherichia coli*,

Pseudomonas, and *Klebsiella* are more frequently isolated from individuals with genitourinary tract infections or intravenous drug users. Mixed bacterial infections typically reflect direct spread or inoculation during surgery or in open fractures. *Haemophilus influenzae* and group B streptococci are common in neonates, and sickle cell disease predisposes to *Salmonella* infection. No specific organism is identified in nearly 50% of patients.

The location of bone infections is influenced by the osseous vascular circulation, which varies with age. In neonates, metaphyseal vessels penetrate the growth plate, resulting in frequent infection of the metaphysis, epiphysis, or both. In older children, involvement of the metaphysis is typical. The epiphyses and subchondral regions are more commonly involved in adults, after growth plate closure, in which merger of metaphyseal and epiphyseal vessels provides a route for bacterial spread.

MORPHOLOGY

Changes associated with osteomyelitis vary temporally and by location. In the acute phase, bacteria proliferate and recruit neutrophils to the site. Necrosis of bone cells and marrow ensues within 48 hours. Bacteria and the associated inflammatory response then spread longitudinally and access the Haversian systems to reach the periosteum. Because the periosteum is loosely attached to the cortex in children, subperiosteal abscesses may form and extend for long distances. Associated periosteal lifting further impairs blood supply and contributes to the necrosis. Soft tissue abscesses may also form after periosteal rupture, and these can channel to the skin as draining sinuses. Dead bone, or **sequestrum,** can crumble and release fragments into the sinus tract.

As the inflammatory process evolves, chronic inflammatory cells recruited during the first week release cytokines that stimulate bone resorption, fibrous tissue ingrowth, and peripheral deposition of reactive bone. This new bone can form a living shell, or **involucrum,** around the devitalized, infected bone (Fig. 26.20).

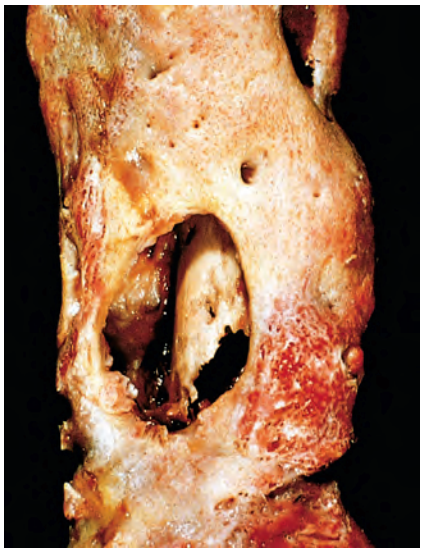


Figure 26.20 Resected femur in a person with draining osteomyelitis. The drainage tract in the subperiosteal shell of viable new bone (involucrum) reveals the inner native necrotic cortex (sequestrum).

Clinical Features

Hematogenous osteomyelitis may present acutely as a systemic illness with malaise, fever, chills, leukocytosis, and intense throbbing pain over the infected bone. In other instances, the presentation is subtle, with only unexplained fever (infants) or localized pain (adults). The characteristic radiographic finding of a lytic focus of bone destruction surrounded by sclerosis strongly suggests osteomyelitis. Blood cultures can be positive before treatment, but biopsy and bone cultures are required to identify the pathogen in most instances. The combination of antibiotics and surgical drainage is usually curative.

Acute osteomyelitis fails to resolve and persists as a chronic infection in 5% to 25% of cases. These are typically associated with delayed diagnosis, extensive bone necrosis, inadequate antibiotic therapy or surgical debridement, or weakened host defenses. Chronic infections may be punctuated by spontaneous flares that may occur after years of dormancy. Other complications of chronic osteomyelitis include pathologic fracture, secondary amyloidosis, endocarditis, and sepsis.

Mycobacterial Osteomyelitis

Mycobacterial osteomyelitis tends to be more destructive and resistant to control than pyogenic osteomyelitis. Mycobacterial osteomyelitis was previously limited to lower income countries, but the incidence has risen worldwide due to immigration and increased numbers of immunocompromised individuals. Risk of mycobacterial osteomyelitis is also increased in those with pulmonary or extrapulmonary tuberculosis, up to 3% of whom have osseous infection. The infection may persist for years before diagnosis and typically presents with localized pain, low-grade fever, chills, or weight loss. Infection is usually solitary in immunocompetent patients but can disseminate in the immunocompromised.

Mycobacteria are usually blood borne and originate from a focus of active visceral disease during the initial stages of primary infection. Bone involvement may occur by direct extension (e.g., from a pulmonary focus into a rib or from tracheobronchial nodes into adjacent vertebrae) or following spread via blood vessels and lymphatics. The histologic findings of granulomatous inflammation and caseous necrosis are typical of tuberculosis (Chapter 8).

Forty percent of mycobacterial osteomyelitis cases involve the spine *(Pott disease)*. Infection breaks through intervertebral discs to involve multiple vertebrae and surrounding soft tissues. Discs and vertebral destruction result in compression fractures that culminate in scoliosis, kyphosis, and neurologic deficits. Tuberculous osteomyelitis may also cause tuberculous arthritis, sinus tract formation, psoas abscess, and amyloidosis.

Skeletal Syphilis

Syphilis *(Treponema pallidum)* and yaws *(Treponema pertenue)* can involve bone. Although syphilis incidence is increasing, bone involvement remains rare because diagnosis and treatment typically occur before this develops.

In congenital syphilis, bone lesions appear in the fifth month of gestation and are fully developed at birth. The spirochetes concentrate at sites of active endochondral ossification and within the periosteum to cause osteochondritis and periostitis, respectively. The characteristic *saber shin* is produced by reactive periosteal bone deposition on the medial and anterior surfaces of the tibia.

In acquired syphilis, bone disease begins early in the tertiary stage, usually 2 to 5 years after the initial infection. Bones of the nose, palate, skull, and extremities, especially the long tubular bones such as the tibia, are involved most frequently.

MORPHOLOGY

Spirochetes can be detected with silver stains or by immunohistochemistry. Edematous granulation tissue with numerous plasma cells and necrotic bone characterizes syphilitic bone disease. **Obliterative endarteritis** typically accompanies caseous necrosis as part of gummas, which may be present in congenital or acquired syphilis (Chapter 8).

BONE TUMORS AND TUMOR-LIKE LESIONS

Primary bone tumors are rare and are vastly outnumbered by metastases and hematopoietic tumors. Nevertheless, the dismal survival of only 50% and disfiguring surgery often required for treatment make management of bone malignancies challenging. Therapy for the ~2400 new bone sarcomas diagnosed each year in the United States aims to optimize survival and maintain function of affected body parts.

Most bone neoplasms have a propensity for the long bones of the extremities. The age groups and anatomic sites affected are typical of specific tumor types. For example, osteosarcoma incidence peaks during adolescence and most frequently involves the knee. In contrast, chondrosarcoma affects the pelvis and proximal extremities of older adults.

Benign bone tumors are often asymptomatic and identified incidentally. Others, however, produce pain, cause a slow-growing mass, or produce a pathologic fracture. Radiologic imaging defines tumor location and can also detect features that narrow the differential diagnosis, but biopsy is necessary for definitive diagnosis in almost all cases.

Bone tumors are classified according to the normal cell types recapitulated or matrix produced. Lesions that do not have normal tissue counterparts are grouped according to clinicopathologic features (Table 26.4). After exclusion of hematopoietic neoplasms, the most common primary bone cancers are osteosarcoma, chondrosarcoma, and Ewing sarcoma.

Bone-Forming Tumors

Tumors in this category produce unmineralized osteoid or mineralized woven bone.

Table 26.4 Classification of Major Nonhematopoietic Primary Bone Tumors

Category and Fraction (%)	Behavior	Tumor Type	Common Locations	Age (years)	Morphology
Cartilage forming (40)	Benign	Osteochondroma	Metaphysis of long bones	10–30	Bony excrescence with cartilage cap
		Chondroma	Small bones of hands and feet	30–50	Circumscribed hyaline cartilage nodule in medulla
		Chondroblastoma	Epiphysis of long bones	10–20	Circumscribed, pericellular calcification
		Chondromyxoid fibroma	Tibia, pelvis	20–30	Collagenous to myxoid matrix, stellate cells
	Malignant	Chondrosarcoma (conventional)	Pelvis, shoulder	40–60	Extends from medulla through cortex into soft tissue, chondrocytes with increased cellularity and atypia
Bone forming (32)	Benign	Osteoid osteoma	Metaphysis of long bones	10–20	Cortical, interlacing microtrabeculae of woven bone
		Osteoblastoma	Vertebral column	10–20	Posterior elements of vertebra, histology similar to osteoid osteoma
	Malignant	Osteosarcoma	Metaphysis of distal femur, proximal tibia	10–20	Extends from medulla to lift periosteum, malignant cells producing woven bone
Unknown origin (19)	Benign	Giant cell tumor	Epiphysis of long bones	20–40	Destroys medulla and cortex, sheets of osteoclasts
		Aneurysmal bone cyst	Proximal tibia, distal femur, vertebra	10–20	Vertebral body, hemorrhagic spaces separated by cellular, fibrous septae
	Malignant	Ewing sarcoma	Diaphysis of long bones	10–20	Sheets of primitive small round cells
		Adamantinoma	Tibia	30–40	Cortical, fibrous bone matrix with epithelial islands
Notochordal (5)	Malignant	Chordoma	Clivus, sacrum	30–60	Destroys medulla and cortex, foamy cells in myxoid matrix

Modified from Unni KK, Inwards CY: *Dahlin's Bone Tumors*, ed 6, Philadelphia, 2010, Lippincott-Williams & Wilkins, p 5; by permission of Mayo Foundation.

Osteoid Osteoma and Osteoblastoma

Osteoid osteoma and osteoblastoma are benign bone-producing tumors with similar histologic features but different sizes, sites of origin, and symptoms; malignant transformation is rare. By definition, osteoid osteomas are less than 2 cm in diameter. They are most common in young men in their teens and 20s and have a predilection for the appendicular skeleton, with about 50% involving the cortex of the femur or tibia. A thick rim of reactive cortical bone may be the only radiographic clue. Despite their small size, osteoid osteomas present with severe nocturnal pain that is probably caused by prostaglandin E_2 (PGE_2) produced by the proliferating osteoblasts and is relieved by aspirin and other NSAIDs. Osteoblastomas are larger than 2 cm, typically involve the posterior spine (laminae and pedicles), cause pain that is unresponsive to aspirin, and do not induce reactive cortical bone. Osteoid osteoma can be treated by radiofrequency ablation, but osteoblastoma usually requires curettage or en bloc excision.

MORPHOLOGY

Osteoid osteoma and osteoblastoma are round-to-oval, well-circumscribed masses of hemorrhagic, gritty tan tissue. Microscopic examination demonstrates randomly interconnected trabeculae of woven bone rimmed by a single layer of prominent osteoblasts (Fig. 26.21) and surrounded by loose connective tissue with many dilated and congested capillaries. The relatively small size, well-defined margins, and benign cytologic features of the neoplastic osteoblasts help distinguish these tumors from osteosarcoma. Reactive bone encircles the actual neoplasm, or nidus, which forms a small radiolucent core that may be centrally mineralized (Fig. 26.22).

Osteosarcoma

Osteosarcoma is the most common primary malignant tumor of bone and accounts for approximately 20% of bone cancers. The age distribution of osteosarcoma is bimodal, with 75% occurring before 20 years of age. A smaller peak occurs in older adults, in whom it is frequently associated with predisposing conditions such as Paget disease, bone infarcts, and prior radiation (sometimes called secondary osteosarcomas). Men are affected slightly more often than women (1.6:1). Although any bone can be involved, tumors usually arise in the metaphyseal region of the long bones; almost 50% are near the knee in the distal femur or proximal tibia.

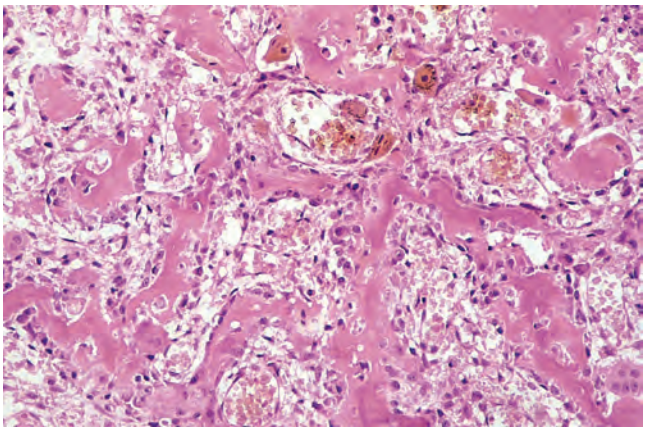


Figure 26.21 Osteoid osteoma composed of haphazardly interconnecting trabeculae of woven bone that are rimmed by prominent osteoblasts. The intertrabecular spaces are filled by vascularized loose connective tissue.

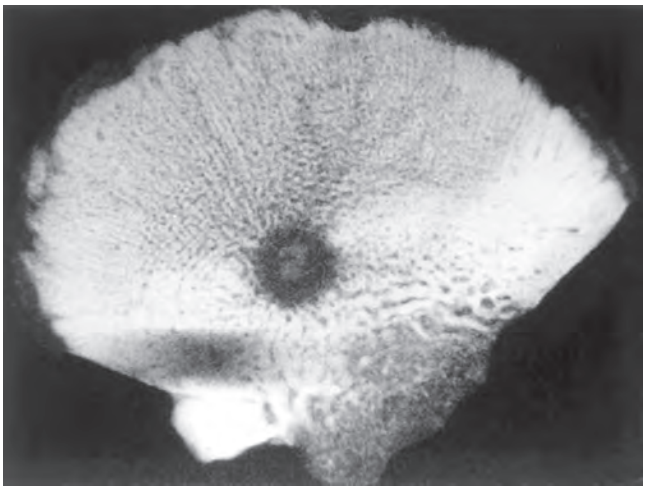


Figure 26.22 Specimen radiograph of intracortical osteoid osteoma. The round radiolucency with central mineralization represents the lesion and is surrounded by abundant reactive bone that has massively thickened the cortex.

Osteosarcomas often present with pain, sometimes due to pathologic fractures. Radiographically, the enlarging tumor forms a destructive, mixed lytic and blastic mass with infiltrative margins (Fig. 26.23). The tumor frequently breaks through the cortex and lifts the periosteum, inducing reactive periosteal bone formation. The triangular shadow between the cortex and raised periosteal ends, known radiographically as a *Codman triangle,* indicates an aggressive tumor.

Pathogenesis

The peak incidence of osteosarcoma is during the adolescent growth spurt. The tumor occurs most frequently in the growth plate of rapidly growing bones, where increased proliferation may predispose to mutations that drive oncogenesis. Approximately 70% of osteosarcomas have acquired genetic abnormalities including chromosomal aberrations. These are usually associated with mutations in well-known tumor suppressor genes and oncogenes (Chapter 7), including the following:

- *RB* mutations are present in up to 70% of sporadic osteosarcomas; germline *RB* mutations confer a 1000-fold increased risk of osteosarcoma.
- *TP53* is mutated in the germline of individuals with Li-Fraumeni syndrome, who have a greatly increased incidence of osteosarcoma. Abnormalities that interfere with p53 function are common in sporadic osteosarcomas.
- *CDKN2A* (also known as *INK4a*), which encodes two tumor suppressors, p16 and p14, is inactivated in many osteosarcomas.
- *MDM2* and *CDK4,* which inhibit p53 and RB function, respectively, are overexpressed in many low-grade osteosarcomas, often through chromosomal amplification of region 12q13–q15.

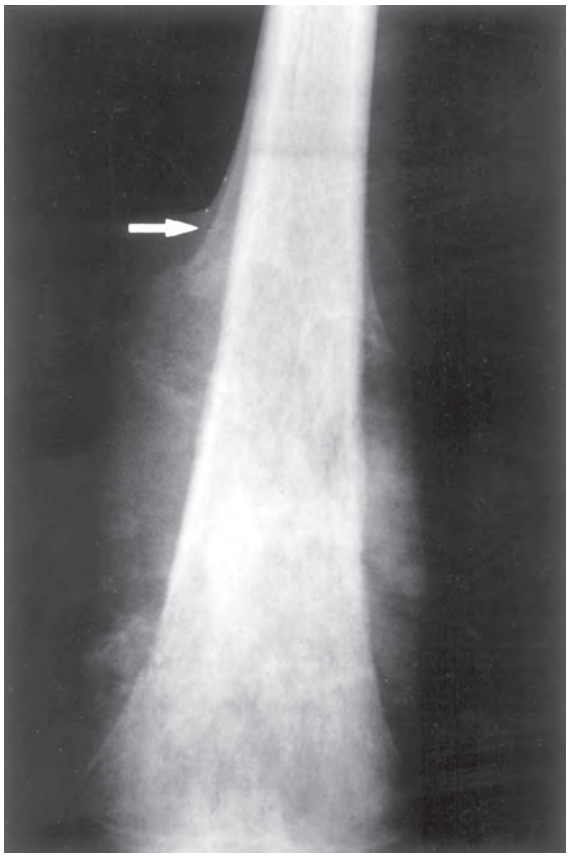


Figure 26.23 Distal femoral osteosarcoma with prominent bone formation extending into the soft tissues. The periosteum, which has been lifted, has laid down a proximal triangular shell of reactive bone known as a Codman triangle *(arrow)*.

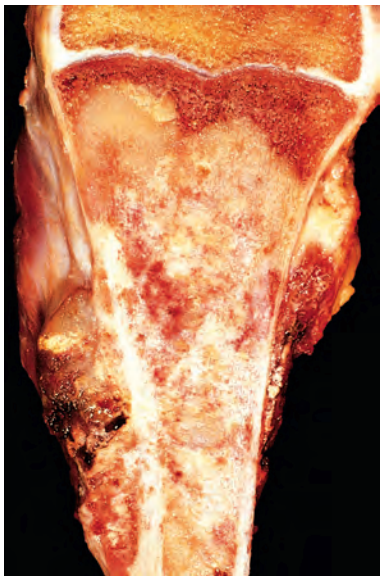


Figure 26.24 Osteosarcoma of the proximal tibia. The tan-white tumor fills most of the medullary cavity of the metaphysis and proximal diaphysis. It has infiltrated through the cortex, lifted the periosteum, and formed soft tissue masses on both sides of the bone.

MORPHOLOGY

Osteosarcomas are bulky, gritty, gray-white tumors that often contain hemorrhage and cystic degeneration (Fig. 26.24). The tumor frequently destroys the surrounding cortices to produce soft tissue masses, spread extensively in the medullary canal, and replace hematopoietic marrow. Tumors infrequently penetrate the epiphyseal plate or enter the joint, where they may grow along tendoligamentous structures or through the attachment site of the joint capsule.

Osteosarcomas demonstrate pleomorphism, large hyperchromatic nuclei, bizarre tumor giant cells, and abundant mitoses including abnormal (e.g., tripolar) forms. Extensive necrosis and intravascular invasion are also common. Diagnosis of osteosarcoma requires the presence of malignant tumor cells producing unmineralized osteoid or mineralized bone (Fig. 26.25), which is typically fine and lacelike, but can also form broad sheets or primitive trabeculae. Neoplastic cells can also produce cartilage; if abundant, such tumors are classified as **chondroblastic osteosarcoma.**

Clinical Features

Based on their known natural history, all osteosarcoma patients are assumed to have occult metastases at the time of diagnosis. As a result, treatment generally includes neoadjuvant chemotherapy, surgery, and postoperative adjuvant chemotherapy. Chemotherapy has greatly improved osteosarcoma prognosis, with 5-year survival reaching 70% in individuals without overt metastases at initial diagnosis. Osteosarcoma metastasizes hematogenously to the lungs, bones, brain, and other sites. The outcome for those with clinically evident metastases, recurrent disease, or secondary osteosarcoma remains guarded, with a 5-year survival rate less than 20%.

Cartilage-Forming Tumors

Cartilaginous tumors account for the majority of benign and malignant primary bone tumors. They are characterized by the formation of hyaline or myxoid cartilage; fibrocartilage and elastic cartilage are rare. Benign cartilage tumors are much more common than malignant lesions.

Osteochondroma

Osteochondroma, or exostosis, is the most common benign bone tumor. It is attached to the skeleton by a bony stalk

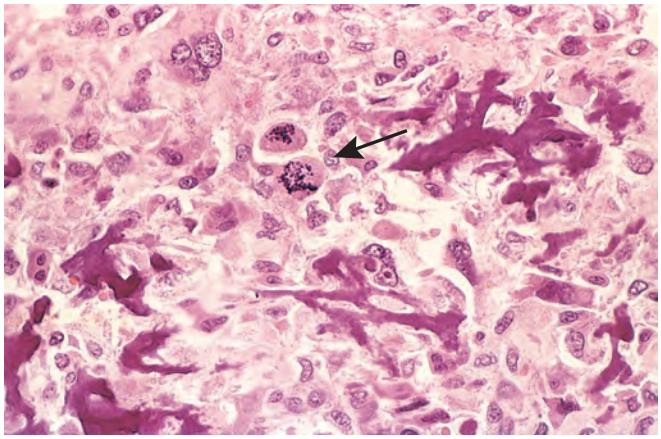


Figure 26.25 Fine, lacelike pattern of neoplastic bone produced by anaplastic malignant tumor cells in an osteosarcoma. Note the abnormal mitotic figures *(arrow)*.

capped by cartilage. About 85% of osteochondromas are solitary and sporadic, with the remainder occurring as part of autosomal dominant *multiple hereditary exostosis syndrome.* Solitary osteochondromas are usually diagnosed in late adolescence and early adulthood, but multiple osteochondromas become apparent during childhood. Men are affected three times more often than women. Osteochondromas develop only in bones of endochondral origin. The most common site is the metaphysis near the growth plate of long tubular bones, especially near the knee, followed by the bones of the pelvis, scapula, and ribs, where they tend to have short stalks. Osteochondromas present as slow-growing masses, which can be painful if they impinge on a nerve or if the stalk is fractured. In many cases, they are detected incidentally. In multiple hereditary exostosis syndrome, the underlying bones may be bowed and shortened, reflecting an associated disturbance in epiphyseal growth.

Pathogenesis

Hereditary exostoses are associated with germline loss-of-function mutations in either the *EXT1* or the *EXT2* gene and subsequent loss of the remaining wild-type allele in chondrocytes of the growth plate. Reduced expression of *EXT1* or *EXT2* has also been observed in sporadic osteochondromas. These genes encode enzymes that synthesize heparan sulfate glycosaminoglycans. The reduced or abnormal glycosaminoglycans may prevent normal diffusion of Indian hedgehog (Ihh), a local regulator of cartilage growth, thereby disrupting Hedgehog signaling and chondrocyte differentiation.

MORPHOLOGY

Osteochondromas are sessile or pedunculated and range in size from 1 to 20 cm. The cap is composed of benign hyaline cartilage (Fig. 26.26) and is covered by perichondrium. The cartilage has the histologic appearance of a disorganized growth plate and undergoes endochondral ossification, with newly made bone forming the inner portion of the head and stalk. The cortex of the stalk merges with the cortex of the host bone, so that the medullary cavity of the osteochondroma is in continuity with that of the bone from which it arises.

Clinical Features

Osteochondromas usually stop expanding at the time of growth plate closure and, when symptomatic, are cured by simple excision. Secondary chondrosarcoma occurs rarely, usually in tumors associated with multiple hereditary exostosis.

Chondroma

Chondroma is a benign tumor of hyaline cartilage that occurs in bones of endochondral origin. Tumors can arise within the medullary cavity, where they are termed *enchondromas,* or on the bone surface, where they are called *juxtacortical chondromas.* Enchondromas are the most common intraosseous cartilage tumor; they are typically solitary metaphyseal lesions of tubular bones of the hands and feet. Radiographically, enchondromas display a circumscribed lucency with central irregular calcifications, a sclerotic rim, and an intact cortex (Fig. 26.27). *Ollier disease* and *Maffucci syndrome* are nonhereditary disorders characterized by multiple enchondromas. Maffucci syndrome is distinguished by the presence of spindle cell hemangiomas and other noncartilage neoplasms.

Enchondromas are most often diagnosed between 20 and 50 years of age. When they involve large bones, enchondromas are usually asymptomatic and detected incidentally.

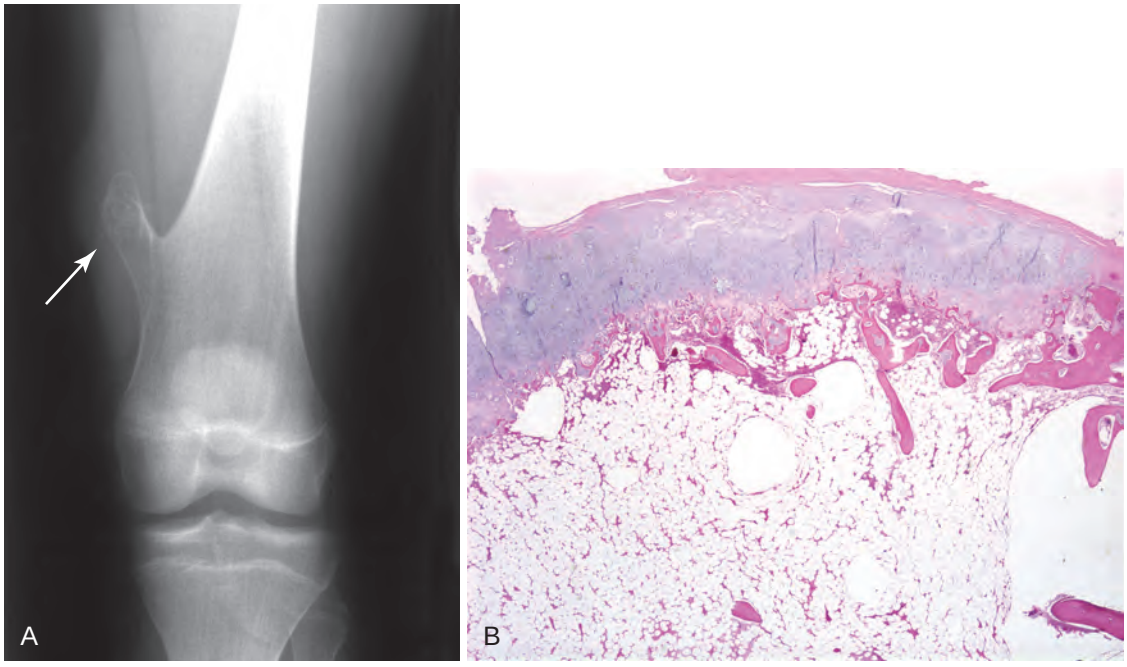


Figure 26.26 Osteochondroma. (A) Radiograph of an osteochondroma arising from the distal femur *(arrow).* (B) The cartilage cap has the histologic appearance of disorganized growth plate–like cartilage.

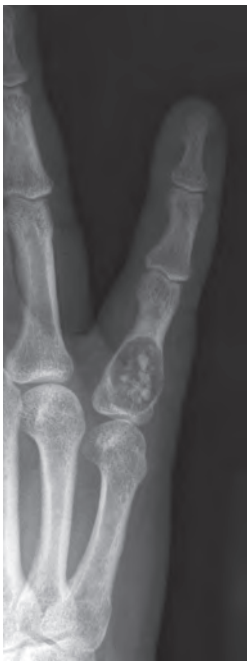


Figure 26.27 Enchondroma of the proximal phalanx. The radiolucent nodule of cartilage with central calcification thins but does not penetrate the cortex.

Occasionally, they can be painful or cause pathologic fractures. In enchondromatosis, the tumors may be numerous and large, producing severe deformities.

Pathogenesis

Heterozygous mutations in the *IDH1* and *IDH2* genes are present within enchondromas. Individuals with enchondroma syndromes are genetically mosaic, harboring *IDH* mutations in a subset of otherwise normal cells throughout their bodies. Similarly, *IDH* mutations are present in only a subset of tumor cells in both syndromic and sporadic enchondromas. This unusual situation may be explained by the functional consequences of *IDH* mutations, which cause the encoded proteins, two isoforms of the enzyme isocitrate dehydrogenase, to acquire a new enzymatic activity that leads to the synthesis of 2-hydroxyglutarate. This "oncometabolite" interferes with regulation of DNA methylation (Chapter 7). It is hypothesized that 2-hydroxyglutarate produced by *IDH*-mutated cells diffuses into neighboring cells with normal *IDH* genes, thereby causing oncogenic epigenetic changes in genetically normal neighbors, a phenomenon referred to as transformation by association.

MORPHOLOGY

Enchondromas are well-circumscribed nodules that are usually smaller than 3 cm, gray-blue, and translucent. Histologically, enchondromas are composed of hyaline cartilage containing cytologically benign chondrocytes (Fig. 26.28). Peripheral endochondral ossification may occur, and the center can calcify and infarct. The enchondromas in Ollier disease and Maffucci syndrome are sometimes more cellular than sporadic enchondromas and can exhibit cytologic atypia, making them difficult to distinguish from chondrosarcomas.

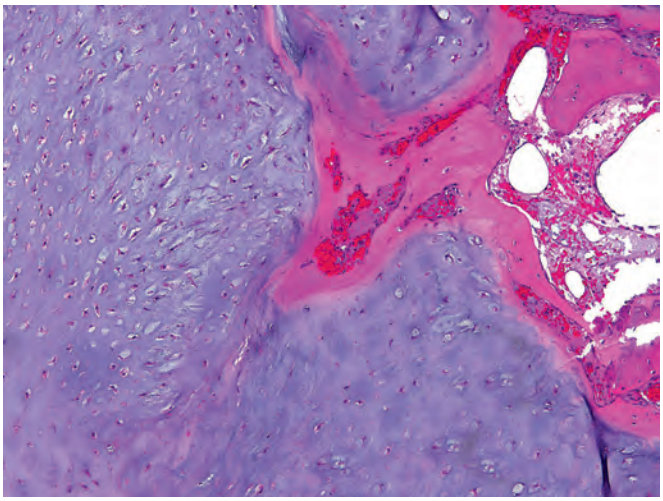


Figure 26.28 Enchondroma composed of a nodule of hyaline cartilage encased by a thin layer of reactive bone.

Clinical Features

The growth potential of chondromas is limited. Treatment depends on the clinical situation and is usually observation or curettage. Solitary chondromas rarely undergo sarcomatous transformation, but those associated with enchondromatosis do so more frequently. Individuals with Maffucci syndrome develop multiple enchondromas and spindle cell hemangiomas and are also at risk of developing other malignancies, including brain gliomas, another type of cancer associated with IDH gene mutations (Chapter 28).

Chondrosarcoma

Chondrosarcomas are malignant cartilage-producing tumors. They are subclassified histologically as conventional, clear cell, dedifferentiated, and mesenchymal variants. After osteosarcoma, chondrosarcoma is the second most common malignant matrix-producing tumor of bone. Individuals with chondrosarcoma are usually in their 40s or older, and men are affected twice as frequently as women; most are of the conventional histologic type. The clear cell and mesenchymal variants occur in younger people, in their teens or 20s. Chondrosarcomas commonly arise in the axial skeleton, especially the pelvis, shoulder, and ribs, but, unlike benign enchondroma, distal extremities are rarely involved. On imaging, the calcified matrix of chondrosarcomas appears as foci of flocculent densities. Slow-growing, low-grade tumors cause reactive cortical thickening, whereas aggressive high-grade neoplasms destroy the cortex and form soft tissue masses. The clear cell variant is unique in that it originates in the epiphyses of long tubular bones. About 15% of conventional chondrosarcomas are secondary and arise from enchondromas or osteochondromas.

Pathogenesis

Although chondrosarcomas are genetically heterogeneous, a few recurrent driver mutations are recognized. Chondrosarcomas arising in multiple osteochondroma syndrome exhibit mutations in the *EXT* genes, while both chondromatosis-related and sporadic chondrosarcomas may have *IDH1* or *IDH2* mutations. Silencing of the *CDKN2A* tumor suppressor

locus by DNA methylation is relatively common in sporadic tumors.

MORPHOLOGY

Conventional chondrosarcomas are large bulky tumors made up of nodules of glistening gray-white, translucent cartilage, but the matrix is often gelatinous or myxoid and can ooze from the cut surface (Fig. 26.29A). Spotty calcifications are typically present, and central necrosis may create cystic spaces. Tumor spreads through the cortex and into surrounding muscle and fat. Histologically, neoplastic cartilage infiltrates the marrow space and surrounds preexisting bony trabeculae (Fig. 26.29B). Chondrosarcomas vary in cellularity, degree of cytologic atypia, and mitotic activity and, on this basis, are graded from 1 to 3 (Fig. 26.29C).

Dedifferentiated chondrosarcoma is a low-grade chondrosarcoma with a second, high-grade component that does not produce cartilage. **Clear cell chondrosarcoma** contains sheets of large, malignant chondrocytes that have abundant clear cytoplasm, numerous osteoclast-type giant cells, and intralesional bone formation. **Mesenchymal chondrosarcoma** is composed of islands of well-differentiated hyaline cartilage surrounded by sheets of primitive appearing small round cells.

Clinical Features

Chondrosarcomas present as painful, progressively enlarging masses. The histologic grade correlates directly with biologic behavior. Most conventional chondrosarcomas are grade 1 tumors, which only rarely metastasize and have 5-year survival rates of 80% to 90%. In contrast, 70% of grade 3 tumors spread hematogenously, especially to the lungs, and 5-year survival is only 43%. Treatment of conventional chondrosarcoma is wide surgical excision, but mesenchymal and dedifferentiated tumors require excision and adjuvant chemotherapy because of their more aggressive clinical course.

Tumors of Unknown Origin

Ewing Sarcoma

Ewing sarcoma is a malignant bone tumor characterized by primitive round cells without obvious differentiation. Ewing sarcomas account for approximately 6% to 10% of primary malignant bone tumors and follow osteosarcoma as the second most common group of bone sarcomas in children. Approximately 80% of patients are younger than 20 years of age. Boys are affected slightly more often than girls, and there is a striking predilection for Caucasians; individuals of African or Asian descent are rarely afflicted.

Ewing sarcoma usually arises in the diaphysis of long tubular bones, especially the femur and the flat bones of the pelvis and presents as a painful, enlarging mass. The affected site is frequently tender, warm, and swollen, and there may be systemic findings that mimic infection, including fever, elevated sedimentation rate, anemia, and leukocytosis. Radiographs show a destructive lytic tumor with permeative, or moth-eaten, margins that extend into surrounding soft tissues. The characteristic periosteal reaction produces layers of reactive bone deposited in an onion-skin fashion.

Pathogenesis

Greater than 90% of Ewing sarcomas contain a balanced translocation involving the *EWSR1* gene on chromosome 22; in a large majority of tumors, the other partner is the *FLI1* gene on chromosome 11, creating an *EWSR1/FLI1* fusion gene. This gene encodes a chimeric EWS/FLI1 protein that binds to chromatin and dysregulates transcription, leading to uncontrolled growth and abnormal differentiation through uncertain mechanisms. The cell of origin is not certain, but mesenchymal stem cells and primitive neuroectodermal cells are most likely.

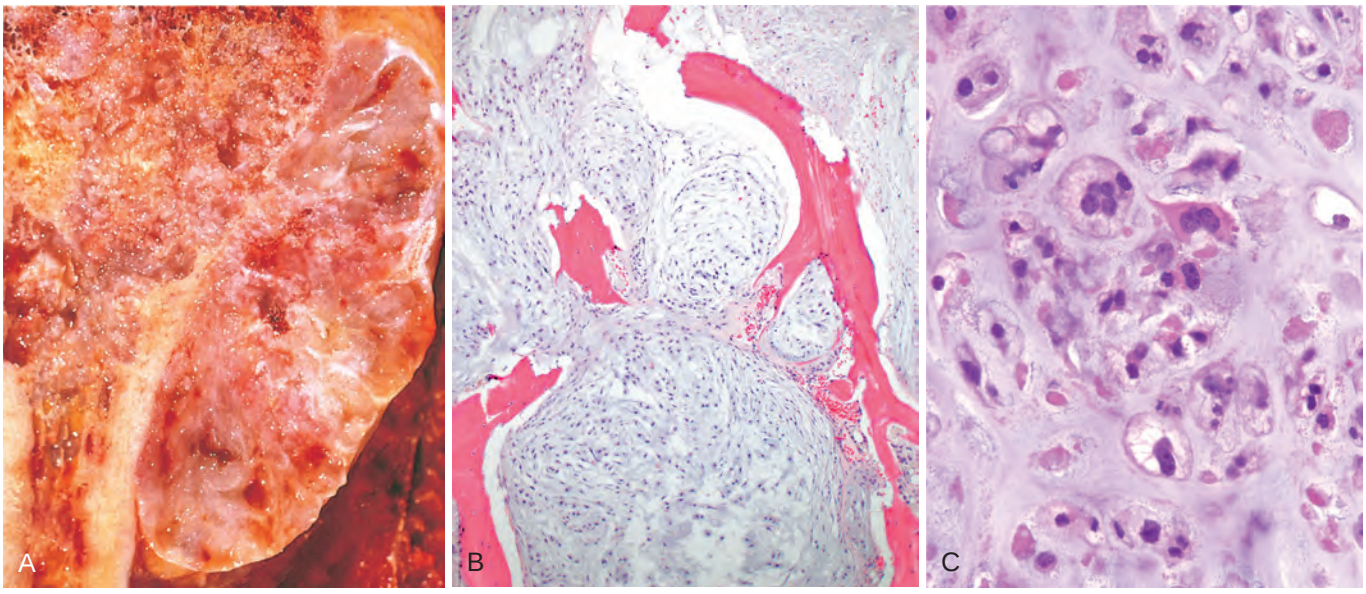


Figure 26.29 Chondrosarcoma. (A) Nodules of hyaline and myxoid cartilage permeate the medullary cavity, grow through the cortex, and form a relatively well-circumscribed soft tissue mass. (B) Conventional chondrosarcoma appearing as a hypercellular cartilaginous mass with entrapped normal bone trabeculae. (C) Anaplastic chondrocytes amid hyaline cartilage matrix in a grade 3 chondrosarcoma.

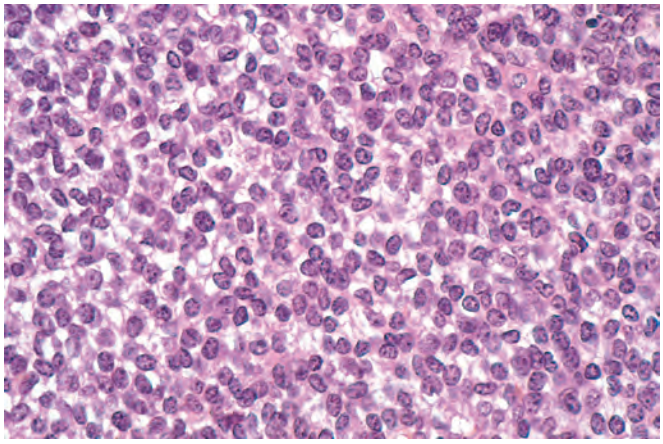


Figure 26.30 Ewing sarcoma composed of sheets of uniform, small round cells with scant clear cytoplasm.

MORPHOLOGY

Ewing sarcoma usually arises in the medullary cavity and invades the cortex, periosteum, and soft tissue. The tumor is soft, tan-white, and frequently contains areas of hemorrhage and necrosis. It is composed of sheets of uniform small, round cells with scant cytoplasm that may be clear due to abundant glycogen (Fig. 26.30). When present, **Homer-Wright rosettes,** which are rounded cell clusters with a central fibrillary core, indicate neuroectodermal differentiation. Fibrous septae may also be present, but there is little stroma. Geographic necrosis may be prominent.

Clinical Features

Although Ewing sarcoma is aggressive, neoadjuvant chemotherapy followed by surgical excision and adjuvant chemotherapy, with or without irradiation, has achieved 75% 5-year survival and 50% long-term cure rates. Chemotherapy-induced necrosis is a positive prognostic indicator.

Giant Cell Tumor

Giant cell tumor is characterized by the presence of numerous multinucleated osteoclast-type giant cells, and is also called *osteoclastoma*. Although giant cell tumors are benign, they can be locally aggressive. This uncommon tumor usually arises in the third through fifth decades of life.

Pathogenesis

Most cells within giant cell tumors are non-neoplastic osteoclasts and their precursors. The neoplastic cells are primitive osteoblast precursors that express high levels of RANKL, which in turn promotes the proliferation of osteoclast precursors and their differentiation into mature osteoclasts. The absence of normal feedback between osteoblasts and osteoclasts results in localized, but highly destructive, bone resorption. The neoplastic cells have acquired mutations in the gene encoding histone 3.3, a chromatin packaging protein; precisely how this leads to tumor formation is unknown.

Giant cell tumors develop within the epiphysis and may extend into the metaphysis. The majority are near the knee, involving the distal femur or proximal tibia, but virtually any bone can be involved. Because they typically arise near joints, giant cell tumors may cause arthritis-like symptoms. They can also present with pathologic fractures.

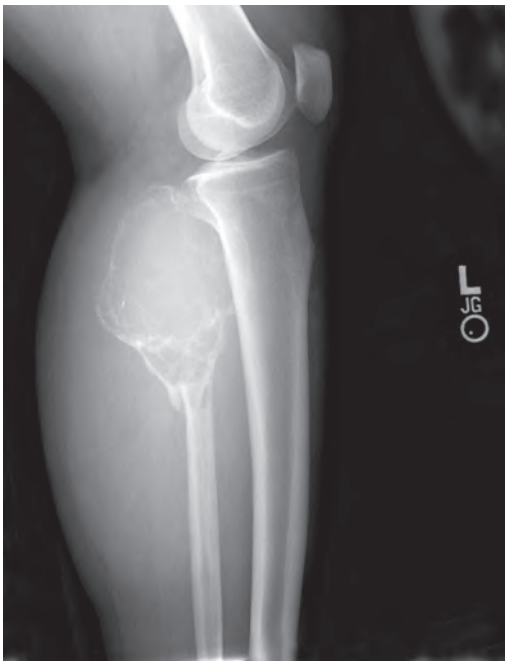


Figure 26.31 Giant cell tumor of the proximal fibula is predominantly lytic and expansile with destruction of the cortex. A pathologic fracture is also present.

MORPHOLOGY

Giant cell tumors often destroy the overlying cortex, resulting in a bulging soft tissue mass delineated by a thin shell of reactive bone (Fig. 26.31). Grossly, the tumors are large, red-brown masses that frequently undergo cystic degeneration. Histologically, the tumors consist of uniform oval mononuclear tumor cells and abundant osteoclast-type giant cells with 100 or more nuclei (Fig. 26.32). Necrosis and mitotic activity may be prominent. Although reactive bone may be present, especially at the periphery, tumor cells do not synthesize bone or cartilage.

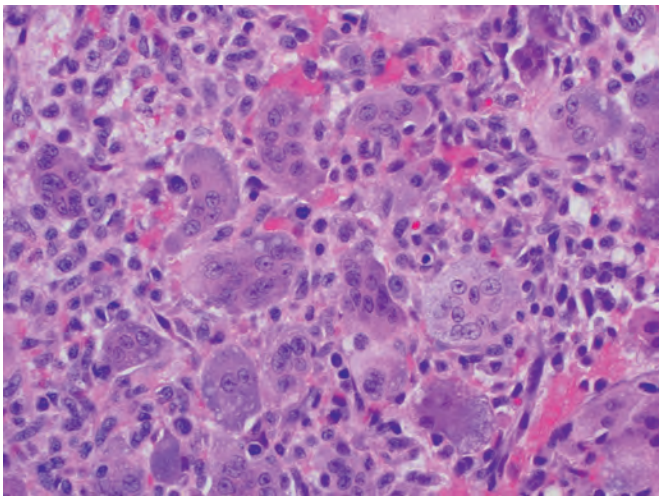


Figure 26.32 Benign giant cell tumor illustrating an abundance of multinucleated giant cells with background mononuclear stromal cells.

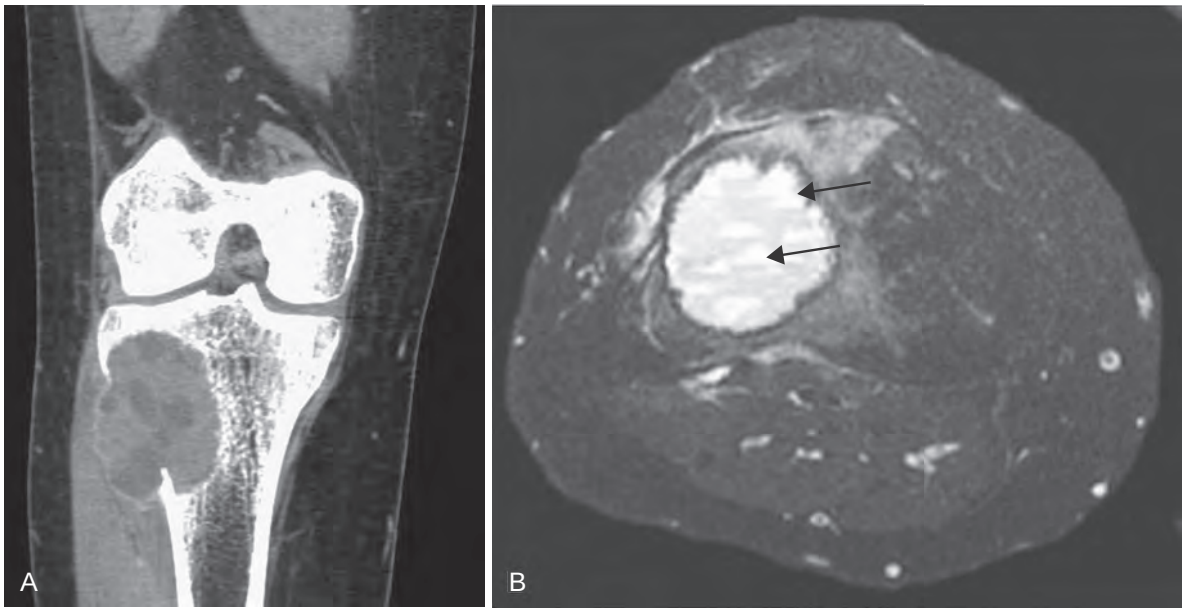


Figure 26.33 (A) Coronal computed axial tomography scan showing an eccentric aneurysmal bone cyst of the tibia. The soft tissue component is delineated by a thin rim of reactive subperiosteal bone. (B) Axial magnetic resonance image demonstrating characteristic fluid levels *(arrows)*.

Clinical Features

Giant cell tumors are treated by curettage, but 40% to 60% recur locally. Although up to 4% metastasize to the lungs, these can regress spontaneously and are seldom fatal. The RANKL inhibitor denosumab has shown promise as an adjuvant therapy.

Aneurysmal Bone Cyst

Aneurysmal bone cyst (ABC) is characterized by multiloculated blood-filled spaces. All age groups are affected, but most cases present in adolescence. ABC develops most frequently in the femur, tibia, and vertebral body posterior elements.

Radiographically, ABC is usually an expansile, well-circumscribed lytic lesion with well-defined margins (Fig. 26.33A). Most show central lysis and a thin sclerotic "eggshell" of reactive bone at the periphery. Supporting trabeculae may create a "soap bubble" appearance on plain films and internal septae with air-fluid levels by computed tomography and magnetic resonance imaging (Fig. 26.33B). These findings are not specific, as similar radiographic (and histologic) changes can also occur as a reaction to trauma and in association with other bone tumors.

Pathogenesis

In nearly 70% of cases of ABC, chromosome 17p13 rearrangements are present within the plump spindle cells, but not in multinucleated giant cells, inflammatory cells, endothelial cells, or osteoblasts. The chromosomal rearrangement results in fusion of the *USP6* coding region with the promoters of genes that are highly expressed in osteoblasts, such as *CDH11*. *USP6* encodes a ubiquitin-specific protease that regulates the activity of the transcription factor NF-κB, which in turn up-regulates expression of proteins, such as matrix metalloproteases, that lead to cystic bone resorption.

MORPHOLOGY

ABCs consist of multiple blood-filled cystic spaces separated by thin, tan-white septa (Fig. 26.34). The septa are composed of plump spindle cells, multinucleated osteoclast-like giant cells, and reactive woven bone lined by osteoblasts. Bone deposition typically follows the contours of fibrous septa. Approximately one-third of cases contain an unusual densely calcified, basophilic metaplastic matrix referred to as **"blue bone."** Necrosis is uncommon unless a pathologic fracture is present.

Clinical Features

ABC presents with localized pain and swelling that may result in a limp, vertebral lesions, and nerve compression.

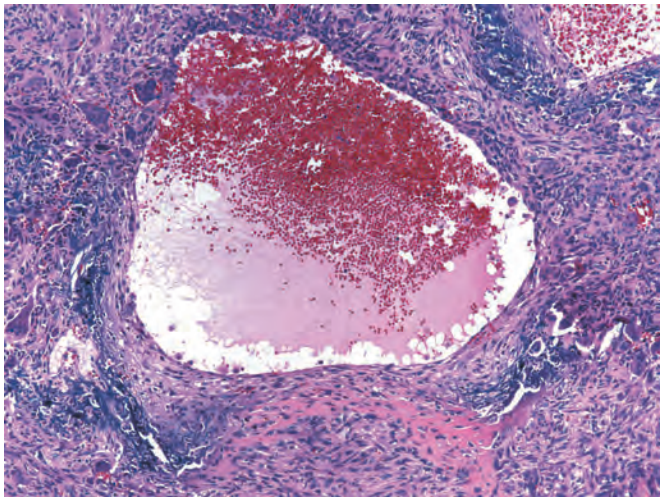


Figure 26.34 Aneurysmal bone cyst with blood-filled cystic space surrounded by a fibrous wall containing proliferating fibroblasts, reactive woven bone, and osteoclast-type giant cells.

Despite being benign, ABC is locally aggressive. Treatment of ABC is by curettage or excision. Recurrence occurs in 10% to 50% of cases.

Lesions Simulating Primary Neoplasms

Fibrous Cortical Defect and Nonossifying Fibroma

Fibrous cortical defects, also known as metaphyseal fibrous defects, are common developmental abnormalities in which fibrous connective tissue replaces bone. These lesions are present in up to 50% of children older than 2 years of age and typically present as an incidental finding in adolescents. The vast majority arise eccentrically in the metaphysis of the distal femur and proximal tibia; almost half are bilateral or multiple. Most are less than 0.5 cm in diameter, but those that grow to 5 or 6 cm are classified as *nonossifying fibromas*.

MORPHOLOGY

Both fibrous cortical defects and nonossifying fibromas are small, sharply demarcated radiolucent masses surrounded by a thin rim of sclerosis (Fig. 26.35). They are grossly yellow-brown and consist histologically of bland fibroblasts that are frequently arranged in a storiform (pinwheel) pattern and macrophages that can take the form of clustered cells with foamy cytoplasm or multinucleated giant cells (Fig. 26.36). Hemosiderin is commonly present.

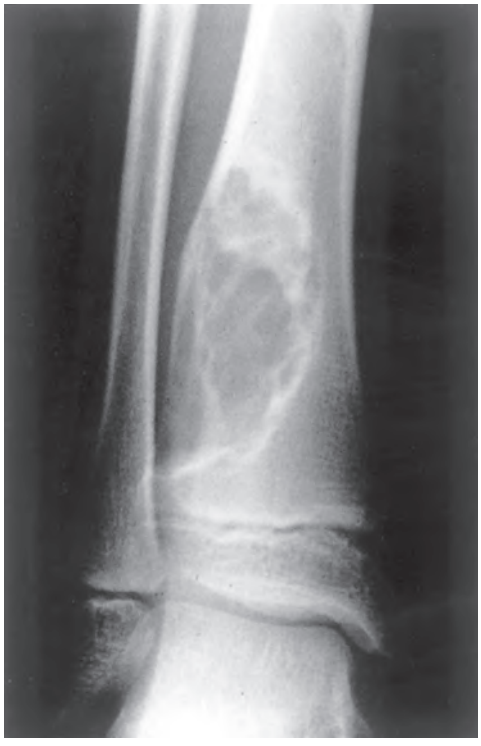


Figure 26.35 Nonossifying fibroma of the distal tibial metaphysis producing an eccentric, lobulated radiolucency surrounded by a sclerotic margin.

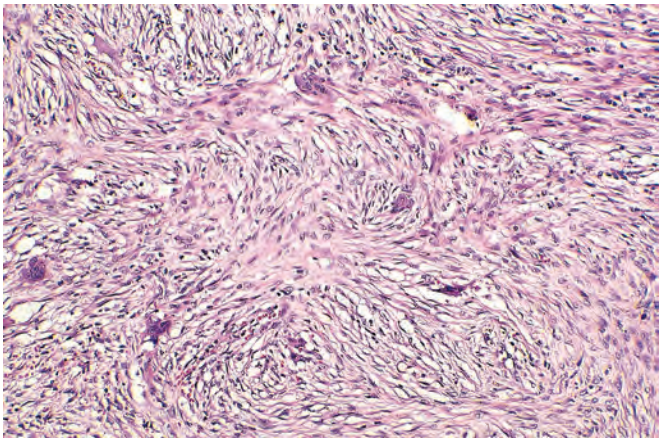


Figure 26.36 Storiform pattern created by benign spindle cells with scattered osteoclast-type giant cells characteristic of a fibrous cortical defect and nonossifying fibroma.

Clinical Features

Fibrous cortical defects are asymptomatic and are detected incidentally on radiographic studies. The radiographic findings are sufficiently specific that biopsy is rarely necessary. Most fibrous cortical defects have limited growth potential and undergo spontaneous resolution over time as they are replaced by normal cortical bone. The few that progressively enlarge into nonossifying fibromas may present with pathologic fracture or require biopsy to exclude other types of tumors. They are treated with curettage and may require bone-grafting for proper healing.

Fibrous Dysplasia

Fibrous dysplasia is a benign tumor that has been likened to localized developmental arrest; all of the components of normal bone are present, but they do not differentiate into mature structures. The lesions arise during skeletal development and appear in several distinctive but sometimes overlapping clinical patterns:

- *Monostotic:* involvement of a single bone
- *Polyostotic:* involvement of multiple bones
- *Mazabraud syndrome:* fibrous dysplasia (usually polyostotic) and soft tissue myxomas
- *McCune-Albright syndrome:* polyostotic disease, associated with café-au-lait skin pigmentations and endocrine abnormalities, especially precocious puberty

Pathogenesis

All forms of fibrous dysplasia result from somatic gain-of-function mutations in *GNAS1*, which encodes the stimulatory α subunit of G_s and is also mutated in pituitary adenomas (Chapter 24). The resulting constitutively active G_s protein promotes cellular proliferation and disturbs osteoblast differentiation. These mutations occur early in embryogenesis, and affected individuals are genetic mosaics. The phenotype depends on (1) the stage of embryogenesis when the mutation is acquired and (2) the fate of the cell harboring the mutation. At one extreme, a mutation during early embryogenesis produces McCune-Albright syndrome. In contrast, a mutation in an osteoblast precursor, during or after formation of the skeleton, results in monostotic fibrous dysplasia.

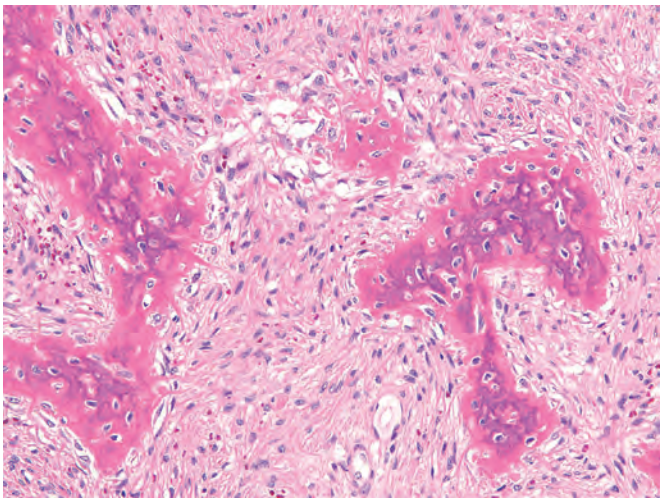


Figure 26.37 Fibrous dysplasia composed of curvilinear trabeculae of woven bone that lack conspicuous osteoblastic rimming and arise in a background of fibrous tissue.

MORPHOLOGY

The lesions of fibrous dysplasia are intramedullary lytic lesions that may expand and cause bowing and cortical thinning. Periosteal reaction usually is absent. Lesional tissue is tan-white and gritty on gross examination and is composed of curvilinear trabeculae of woven bone surrounded by a moderately cellular fibroblastic proliferation without prominent osteoblastic rimming (Fig. 26.37). Nodules of hyaline cartilage with the appearance of disorganized growth plates are present in approximately 20% of cases. Cystic degeneration, hemorrhage, and foamy macrophages are also common.

Clinical Features

Monostotic fibrous dysplasia occurs in early adolescence and often stops enlarging at the time of growth plate closure. The femur, tibia, ribs, jawbones, and calvarium are most commonly affected. The lesion is frequently asymptomatic and usually discovered incidentally but may cause pain, discrepancies in limb length, and pathologic fracture. Growth of the lesions may be reactivated during pregnancy. Symptomatic lesions are treated by curettage, but recurrence is common.

Polyostotic fibrous dysplasia manifests at a slightly earlier age than the monostotic type and may continue to cause problems into adulthood. The femur, skull, and tibia are most frequently affected. Craniofacial involvement is present in 50% of those who have a moderate number of bones affected and in 100% of those with extensive skeletal disease. Involvement of the shoulder and pelvic girdles results in progressive disease and may produce crippling deformities and fractures that require multiple corrective orthopedic surgical procedures. Bisphosphonates can be used to reduce the severity of the bone pain. A rare complication, usually in the setting of polyostotic involvement, is malignant transformation of a lesion into a sarcoma.

Mazabraud syndrome presents with skeletal features of polyostotic fibrous dysplasia in childhood followed by the appearance of intramuscular myxomas in adulthood, often in the same anatomic region as existing fibrous dysplasia. Although benign, myxomas can cause local compression symptoms or limb deformity but are cured by surgical excision.

The most common clinical presentation of McCune-Albright syndrome is precocious sexual development, primarily in girls. The syndrome can include other endocrinopathies such as hyperthyroidism, pituitary adenomas that secrete growth hormone, and primary adrenal hyperplasia. Bone lesions are often unilateral, and skin lesions, when present, are limited to the same side of the body. The cutaneous lesions are classically large macules, dark to café-au-lait in color, with irregular serpiginous borders. Skeletal manifestations are managed as for polyostotic fibrous dysplasia, while the endocrinopathies are treated medically (e.g., with aromatase inhibitors for precocious puberty).

Metastatic Tumors

Metastatic tumors are the most common form of skeletal malignancy, greatly outnumbering primary bone cancers. The pathways of spread to bone include (1) direct extension, (2) lymphatic or hematogenous dissemination, and (3) intraspinal seeding (via the Batson plexus of veins). Any cancer can spread to bone, but in adults more than 75% of skeletal metastases originate from cancers of the prostate, breast, kidney, and lung. In children, neuroblastoma, Wilms tumor, osteosarcoma, Ewing sarcoma, and rhabdomyosarcoma are the most common metastases to bone.

Skeletal metastases are typically multifocal. However, carcinomas of the kidney and thyroid may present with solitary lesions. Most metastases involve the axial skeleton (vertebral column, pelvis, ribs, skull, and sternum) in which the marrow has a rich capillary network.

The radiographic appearance of metastases may be *lytic* (bone destroying), *blastic* (bone forming), or *mixed* lytic and blastic. Some cancers are associated with predominantly one pattern. For example, prostatic adenocarcinoma is predominantly blastic, whereas carcinomas of the kidney, lung, and gastrointestinal tract and malignant melanoma produce lytic lesions. Bidirectional interactions between metastatic cancer cells and native bone cells account for changes in the bone matrix. Tumor cells do not directly resorb bone in lytic lesions, but they secrete substances such as prostaglandins, cytokines, and PTHrP that up-regulate RANKL on osteoblasts and stromal cells, thereby stimulating osteoclast activity. Conversely, tumor cell growth is supported by the release of matrix-bound growth factors (e.g., TGF-β, IGF-1, and FGF) as bone is resorbed. Sclerotic metastases may result from tumor cells secreting WNT proteins that stimulate osteoblastic bone formation.

The presence of bone metastases portends a poor prognosis because it indicates wide dissemination of the cancer. Goals of therapy are symptomatic relief and prevention of further spread. When treated, options include systemic chemotherapy or immunotherapy, localized radiation, and bisphosphonates. Surgery may be necessary to stabilize pathologic fractures, particularly when metastases involve the spine.

KEY CONCEPTS

BONE TUMORS AND TUMOR-LIKE LESIONS

The majority of bone tumors are classified according to the normal cell they resemble or matrix produced; the remainder are grouped by clinicopathologic features. Most primary bone tumors are benign. Metastases, especially adenocarcinomas, are more common than primary bone neoplasms.

Major categories of primary bone tumors include the following:

- Bone forming: Osteoblastoma and osteoid osteoma consist of benign osteoblasts that synthesize osteoid. Osteosarcoma is a tumor of malignant osteoblasts with an aggressive clinical course that predominantly involves adolescents.
- Cartilage forming: Osteochondroma is a polypoid exostosis with a cartilage cap. Syndromic forms are most often associated with mutations in *EXT* genes. Chondromas are benign intramedullary tumors that produce hyaline cartilage and usually arise in the digits. Chondrosarcomas are malignant tumors of cartilage involving the axial skeleton in adults.
- Ewing sarcoma is an aggressive malignant small round cell tumor associated with t(11;22).
- Fibrous cortical defect and fibrous dysplasia are unusual developmental abnormalities; the latter is caused by somatic gain-of-function mutations in the *GNAS1* gene during embryogenesis.

Joints

Joints allow movement while providing mechanical stability. They are classified as solid (nonsynovial) and cavitated (synovial). Solid joints, or synarthroses, provide structural integrity, lack a joint space, and allow only minimal movement. They are grouped according to the type of connective tissue (fibrous tissue or cartilage) that bridges the ends of the bones. Fibrous synarthroses include the cranial sutures and the bonds between roots of teeth and the jawbones. Immobile cartilaginous joints, or synchondroses, include symphyses (manubriosternalis and pubic). Synovial joints, in contrast, have a joint space that allows for a wide range of motion. Situated between the ends of bones formed via endochondral ossification, they are strengthened by a dense fibrous capsule reinforced by ligaments and muscles. The boundary of the joint space consists of the synovial membrane, which is firmly anchored to the underlying capsule and does not cover the articular surface. The synovium is smooth except near the osseous insertion, where it is thrown into numerous villous folds. Synovial membranes are lined by two types of cells arranged in one to four cell deep layers. Type A synoviocytes are specialized macrophages with phagocytic activity. Type B synoviocytes are similar to fibroblasts and synthesize hyaluronic acid and various proteins. The synovial lining lacks a basement membrane, which allows for efficient exchange of nutrients, waste, and gases between blood and synovial fluid. Synovial fluid is therefore a hyaluronic acid–rich plasma filtrate that acts as a viscous lubricant and provides nutrition for the articular hyaline cartilage.

Hyaline cartilage is a unique connective tissue ideally suited to serve as an elastic shock absorber and wear-resistant surface. Hyaline cartilage is composed of water (70%), type II collagen (10%), proteoglycans (8%), and chondrocytes. The collagen fibers enable resistance to tensile stresses and transmit vertical loads; water and proteoglycans limit compression and friction. Hyaline cartilage lacks blood vessels, lymphatics, and nerves. Chondrocytes synthesize and enzymatically digest matrix, with the half-life of the different components ranging from weeks (proteoglycans) to years (type II collagen). Chondrocytes secrete degradative enzymes in inactive forms and enrich the matrix with enzyme inhibitors. Diseases that destroy articular cartilage do so by activating the degradative enzymes and decreasing the production of their inhibitors, leading to matrix breakdown. Cytokines such as IL-1 and TNF are released by chondrocytes, synoviocytes, fibroblasts, and inflammatory cells and trigger degradative processes. Articular cartilage destruction by indigenous cells contributes to pathogenesis of many joint diseases.

ARTHRITIS

Arthritis refers to inflammation of joints. The clinically most important forms of arthritis are osteoarthritis and rheumatoid arthritis, which differ in pathogenesis and clinical and pathologic manifestations (Table 26.5). Other types of arthritis

Table 26.5 Comparative Features of Osteoarthritis and Rheumatoid Arthritis

	Osteoarthritis	Rheumatoid Arthritis
Primary pathogenic abnormality	Mechanical injury to articular cartilage	Autoimmunity
Role of inflammation	May be secondary; inflammatory mediators exacerbate cartilage damage	Primary: cartilage destruction is caused by T cells and antibodies reactive with joint antigens
Joints involved	Primarily weight bearing (knees, hips)	Often begins with small joints of fingers; progression leads to multiple joints involved
Pathology	Cartilage degeneration and fragmentation, bone spurs, subchondral cysts; minimal inflammation	Inflammatory pannus invading and destroying cartilage; severe chronic inflammation; joint fusion (ankylosis)
Serum antibodies	None	Various, including ACPA, rheumatoid factor
Involvement of other organs	No	Yes (lungs, heart, other organs)

ACPA, Anti-citrullinated peptide antibody.

are caused by immune reactions, infections, and crystal deposition.

Osteoarthritis

Osteoarthritis (OA), also called degenerative joint disease, is characterized by cartilage degeneration that results in structural and functional failure of synovial joints. It is the most common type of joint disease. Although the term osteoarthritis implies an inflammatory disease, it is considered to be primarily a degenerative disease of cartilage.

In most instances, OA appears insidiously, without apparent initiating cause, as an aging phenomenon *(idiopathic* or *primary OA)*. In these cases, the disease usually affects few joints (oligoarticular) but may be generalized. In about 5% of cases, OA appears in younger individuals with predisposing conditions, such as a joint deformity, a previous joint injury, or an underlying systemic disease such as diabetes, ochronosis, hemochromatosis, or marked obesity that places joints at risk. In these settings, the disease is called *secondary OA*. Gender has some influence on distribution. The knees and hands are more commonly affected in women and the hips in men. The association with aging is strong; the prevalence of OA increases exponentially beyond 50 years of age, and about 40% of people older than 70 years of age are affected.

Pathogenesis

The lesions of OA stem from degeneration of the articular cartilage and disordered repair. Biomechanical stress is the principal pathogenic mechanism, but genetic factors, including polymorphisms in genes encoding components of the matrix and signaling molecules, may predispose to chondrocyte injury that causes matrix alteration (Fig. 26.38). Chondrocytes proliferate and continuously synthesize proteoglycans, but disease develops when degradation exceeds synthesis. This leads to changes in proteoglycan composition as the disease progresses. Chondrocytes also secrete matrix metalloproteases (MMPs) that degrade the type II collagen network. Cytokines and diffusible factors from chondrocytes and synovial cells, particularly TGF-β (which induces the production of MMPs), TNF, prostaglandins, and nitric oxide, are also implicated in OA, and chronic, low-level inflammation contributes to disease progression. Advanced disease is characterized by chondrocyte loss and severe matrix degradation.

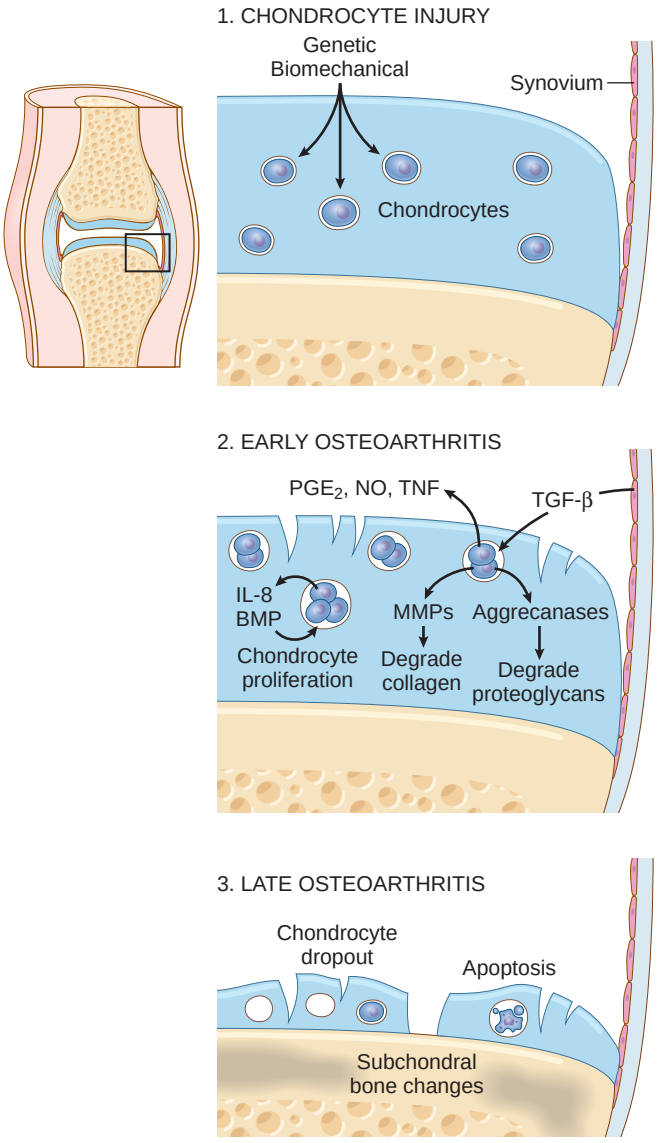


Figure 26.38 Schematic view of osteoarthritis (OA). The process is initiated by biomechanical cartilage injury (1) that can be accelerated in genetically predisposed individuals and results in changes to the extracellular matrix. (2) Although chondrocytes may proliferate and attempt to repair damaged matrix, continued degradation exceeds repair in early OA. (3) Late OA is evidenced by loss of both matrix and chondrocytes with subchondral bone damage.

MORPHOLOGY

Chondrocytes proliferate and form clusters in early stages of OA. Concurrently, matrix water content increases, proteoglycan concentration decreases, and horizontally arranged collagen type II fibers in the superficial zone are cleaved. These processes result in fissures and clefts that create a granular and soft articular surface (Fig. 26.39A). As chondrocytes die, full-thickness portions of cartilage are sloughed into the joint, forming **loose bodies.** Exposed subchondral bone becomes the new articular surface, which is burnished by friction with the opposing surface, giving it the appearance of polished ivory **(bone eburnation)** (Fig. 26.39B). Underlying articular bone undergoes rebuttressing and sclerosis, and develops small fractures, creating gaps that allow synovial fluid to be forced into the subchondral regions. As the loculated fluid collection increases in size, fibrous-walled cysts form. Mushroom-shaped bony outgrowths **(osteophytes)** develop at the margins of the articular surface and are capped by fibrocartilage and hyaline cartilage that gradually ossify. The synovium is usually only mildly congested and fibrotic and contains few inflammatory cells.

Clinical Features

Individuals with primary osteoarthritis are usually asymptomatic until they are in their 50s. If a young person has significant manifestations of OA, a search for some underlying cause should be made. Characteristic symptoms include deep, achy pain that worsens with use, morning stiffness, crepitus, and limitation of range of movement.

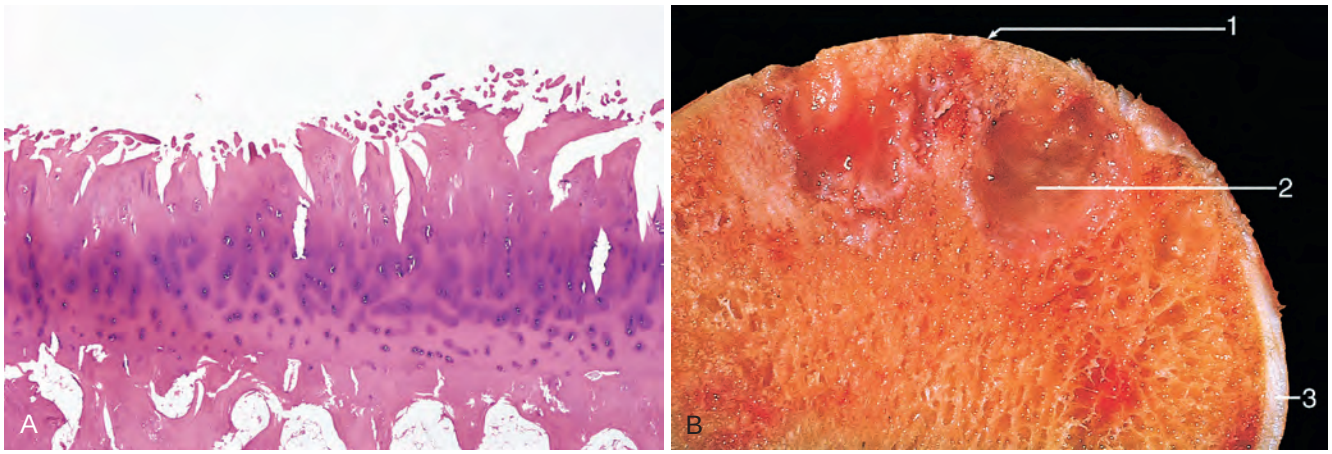


Figure 26.39 Osteoarthritis. (A) Histologic demonstration of the characteristic fibrillation of articular cartilage. (B) Eburnated articular surface exposing subchondral bone (1), subchondral cysts (2), and residual articular cartilage (3).

Impingement on spinal foramina by osteophytes results in cervical and lumbar nerve root compression and radicular pain, muscle spasms, muscle atrophy, and neurologic deficits. Typically, only one or a few joints are involved, except in the uncommon generalized variant. The joints commonly involved include the hips (Fig. 26.40), knees, lower lumbar and cervical vertebrae, proximal and distal interphalangeal joints of the fingers, first carpometacarpal joints, and first tarsometatarsal joints. *Heberden nodes,* prominent osteophytes at the distal interphalangeal joints, are common in women. Wrists, elbows, and shoulders are usually spared. Joint deformities can develop over time, but unlike rheumatoid arthritis (RA; discussed next), fusion does not occur (Fig. 26.41). The level of disease severity detected radiographically does not correlate well with pain and disability. There are no treatments to prevent or halt progression of primary OA, and therapies include pain management, NSAIDs to reduce inflammation, intraarticular corticosteroids, activity modification, and, for severe cases, arthroplasty.

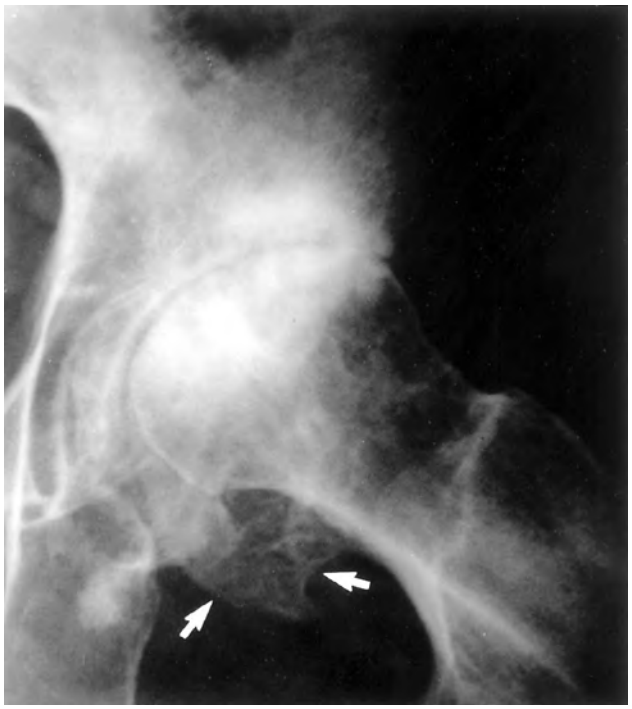


Figure 26.40 Severe osteoarthritis of the hip. The joint space is narrowed, and there is subchondral sclerosis with scattered oval radiolucent cysts and peripheral osteophyte lipping *(arrows).*

Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a chronic autoimmune disorder that principally attacks the joints, producing a nonsuppurative proliferative and inflammatory synovitis. RA often progresses to articular cartilage destruction and, in some cases, joint fusion *(ankylosis)*. Extraarticular lesions may involve skin, heart, blood vessels, and lungs, leading to clinical manifestations that overlap with other autoimmune disorders including systemic lupus erythematosus and scleroderma. The prevalence of RA in the United States is approximately 1%. The incidence peaks in the second to fourth decades. RA is three times more common in women than in men.

Pathogenesis

The autoimmune response in RA is initiated by CD4+ helper T cells. As in other autoimmune diseases, genetic predisposition and environmental factors contribute to the development, progression, and persistence of disease. The pathologic changes in RA are mediated by antibodies against self antigens and inflammation induced by cytokines, predominantly those secreted by CD4+ T cells (Fig. 26.42).

The activated CD4+ T cells release inflammatory mediators that stimulate other inflammatory cells, leading to tissue injury. Although many cytokine mediators can be isolated from inflamed joints, the most important ones include:

- IFN-γ from Th1 cells, which activates macrophages and resident synovial cells.
- IL-17 from Th17 cells that recruits neutrophils and monocytes.
- RANKL, which is expressed on activated T cells and stimulates bone resorption.
- TNF and IL-1 from macrophages stimulate resident synovial cells to secrete proteases that destroy hyaline cartilage.

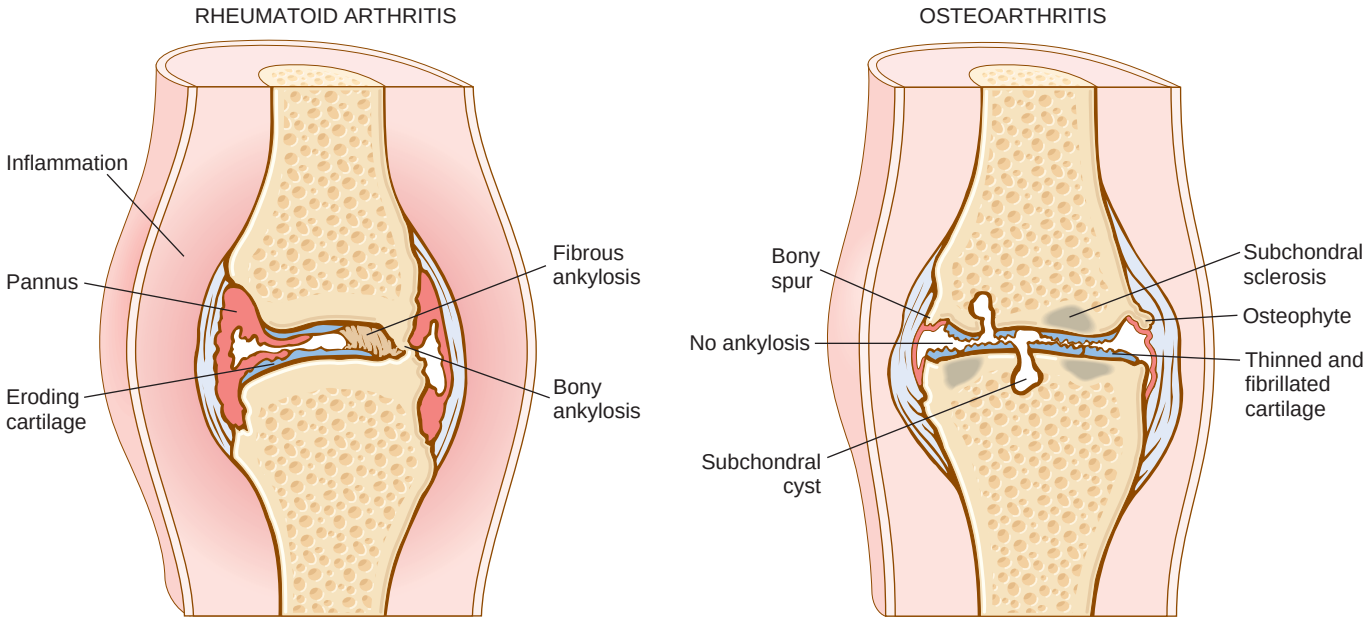


Figure 26.41 Comparison of the morphologic features of rheumatoid arthritis and osteoarthritis.

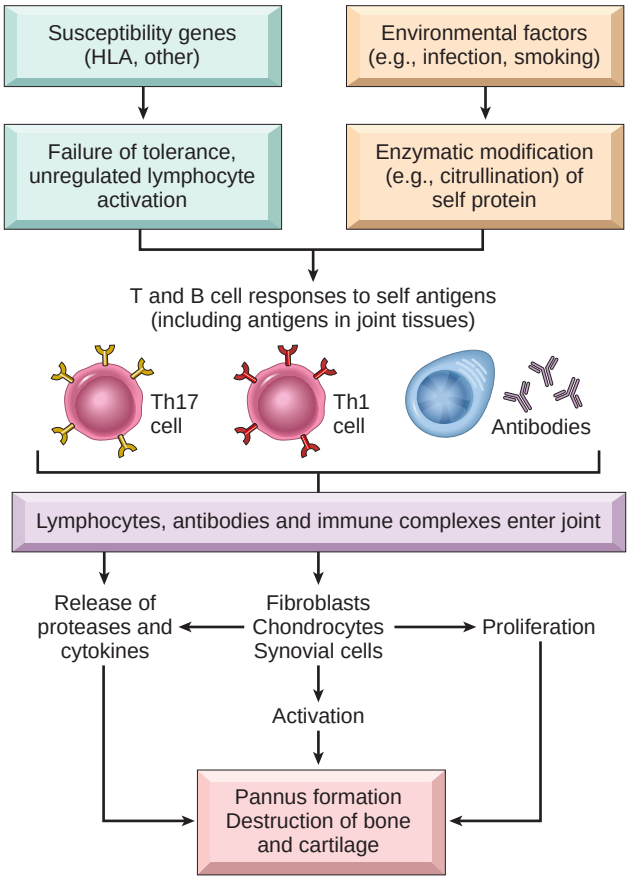


Figure 26.42 Major processes involved in the pathogenesis of rheumatoid arthritis.

Of these, TNF has been most firmly implicated in RA pathogenesis, and anti-TNF biologics have revolutionized its treatment (see later).

In RA, the synovium often contains germinal centers with secondary follicles and abundant plasma cells. Some of these plasma cells secrete antibodies that recognize self antigens. Many of these autoantibodies, which are produced in lymphoid organs as well as synovium, are specific for citrullinated peptides (CCPs) in which arginine residues are post-translationally converted to citrulline. The modified epitopes are found in several proteins found in joints, including fibrinogen, type II collagen, α-enolase, and vimentin. *Anti-citrullinated peptide antibodies* (ACPAs) are diagnostic markers that can be detected in serum of up to 70% of RA patients. Some data suggest that, in combination with T cells reactive to citrullinated proteins, ACPAs drive disease persistence. IgM and IgA autoantibodies that bind to IgG Fc regions are present in 80% of individuals. These autoantibodies, collectively referred to as *rheumatoid factor,* may be deposited in joints as immune complexes, but are not present in all patients and can be detected in some individuals without RA.

It is estimated that 50% of the risk of developing RA is related to inherited genetic susceptibility. The *HLA-DR4* allele is associated with ACPA-positive RA. Evidence suggests that an epitope on a citrullinated protein, vinculin, mimics an epitope on many microbes and can be presented by the class II HLA-DR4 molecule.

Environmental factors that promote autoimmunity are involved, but as in many autoimmune diseases, they have not been clearly defined. For example, infection (including periodontitis) and smoking may promote citrullination of self proteins, creating new epitopes that trigger autoantibody production.

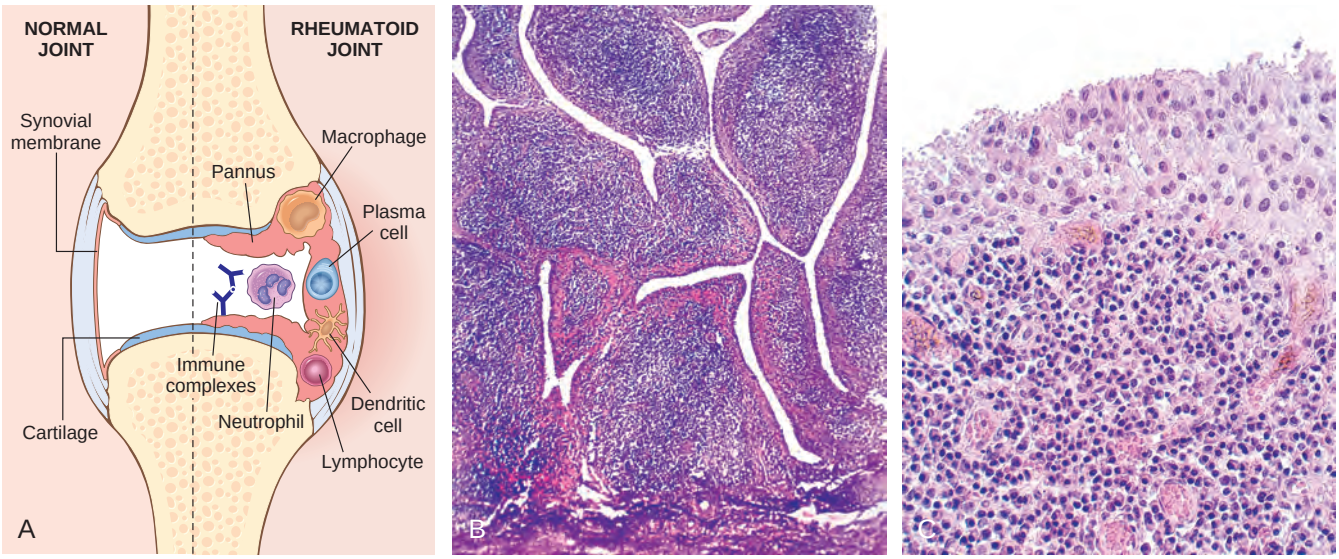


Figure 26.43 Rheumatoid arthritis. (A) Schematic view of the joint lesion. (B) Low magnification reveals marked synovial hypertrophy with formation of villi. (C) At higher magnification, subsynovial tissue containing a dense lymphoid aggregate. (A, Modified from Feldmann M: Development of anti-TNF therapy for rheumatoid arthritis, *Nat Rev Immunol* 2(5):364–371, 2002.)

MORPHOLOGY

RA typically affects small joints of the hands and feet. The synovium becomes grossly edematous, thickened, and hyperplastic, transforming its smooth contour to one covered by delicate and bulbous villi (Fig. 26.43A–B). The characteristic histologic features include (1) synovial cell hyperplasia and proliferation; (2) dense inflammatory infiltrates (frequently forming lymphoid follicles) of CD4+ helper T cells, B cells, plasma cells, dendritic cells, and macrophages (see Fig. 26.43); (3) increased vascularity due to angiogenesis; (4) fibrinopurulent exudate on the synovial and joint surfaces; and (5) osteoclastic activity in subchondral bone, allowing the inflamed synovium to penetrate the bone and cause periarticular erosions and subchondral cysts. Together, these changes produce the **pannus,** a mass of edematous synovium, inflammatory cells, granulation tissue, and fibroblasts that grows over and causes erosion of articular cartilage. In time, after the cartilage has been destroyed, the pannus bridges the apposing bones to form a **fibrous ankylosis,** which eventually ossifies and results in bone fusion, or bony ankylosis.

Rheumatoid nodules are an infrequent manifestation of RA and typically occur in the subcutaneous tissue of the forearm, elbows, occiput, and lumbosacral area. These small masses are firm, nontender, and round to oval. Microscopically, they resemble necrotizing granulomas with a central zone of fibrinoid necrosis surrounded by a prominent rim of activated macrophages and numerous lymphocytes and plasma cells (Fig. 26.44). Severe disease may be associated with **leukocytoclastic vasculitis** (Chapter 11), an acute necrotizing vasculitis of small and large arteries that may involve pleura, pericardium, or lung and evolve into a chronic fibrosing process. Leukocytoclastic vasculitis produces purpura, cutaneous ulcers, and nail bed infarction. Ocular changes such as uveitis and keratoconjunctivitis (similar to Sjögren syndrome, Chapter 6) may also occur.

Clinical Features

RA may be distinguished from other forms of polyarticular inflammatory arthritis serologically by ACPA detection and radiographically by characteristic changes (see later). It begins with malaise, fatigue, and generalized musculoskeletal pain in about half of patients; joint involvement develops after weeks to months. The pattern of joint involvement varies, but it is generally symmetrical and affects small joints before larger ones. Symptoms usually develop in the hands and feet, followed in decreasing frequency by the wrists, ankles, elbows, and knees. In the hands, the metacarpophalangeal and proximal interphalangeal joints are involved, in contrast to OA (see earlier).

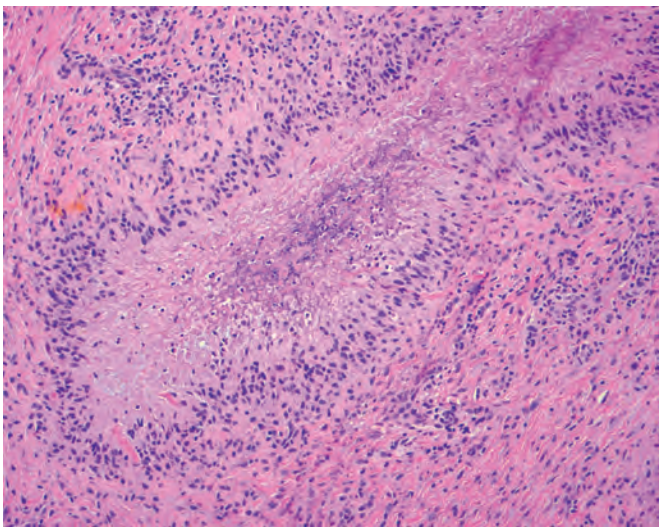


Figure 26.44 Rheumatoid nodule composed of central necrosis rimmed by palisaded macropahges.

Involved joints are swollen, warm, and painful. Unlike OA, morning joint stiffness does not subside with activity. The typical patient has progressive joint enlargement and decreasing range of motion during a chronic waxing and waning course. In a minority of individuals, especially those lacking ACPA and rheumatoid factor, disease may stabilize or even regress. A small number of patients (about 10%) develop acute-onset RA, with severe symptoms and polyarticular involvement that develops over several days.

Inflammation in tendons, ligaments, and occasionally adjacent skeletal muscle frequently accompanies the arthritis and produces characteristic radial deviation of the wrist, ulnar deviation of the fingers, and flexion-hyperextension of the fingers (swan-neck and boutonnière deformities). Radiographic hallmarks include joint effusions and juxta-articular osteopenia with erosions and narrowing of the joint space and loss of articular cartilage (Fig. 26.45).

RA treatment is aimed at relieving pain and inflammation and slowing or arresting joint destruction. Therapies include corticosteroids, other immunosuppressants such as methotrexate, and, most notably, TNF antagonists, which are effective in many patients. Unfortunately, individuals must be maintained on anti-TNF therapy to avoid disease flares.

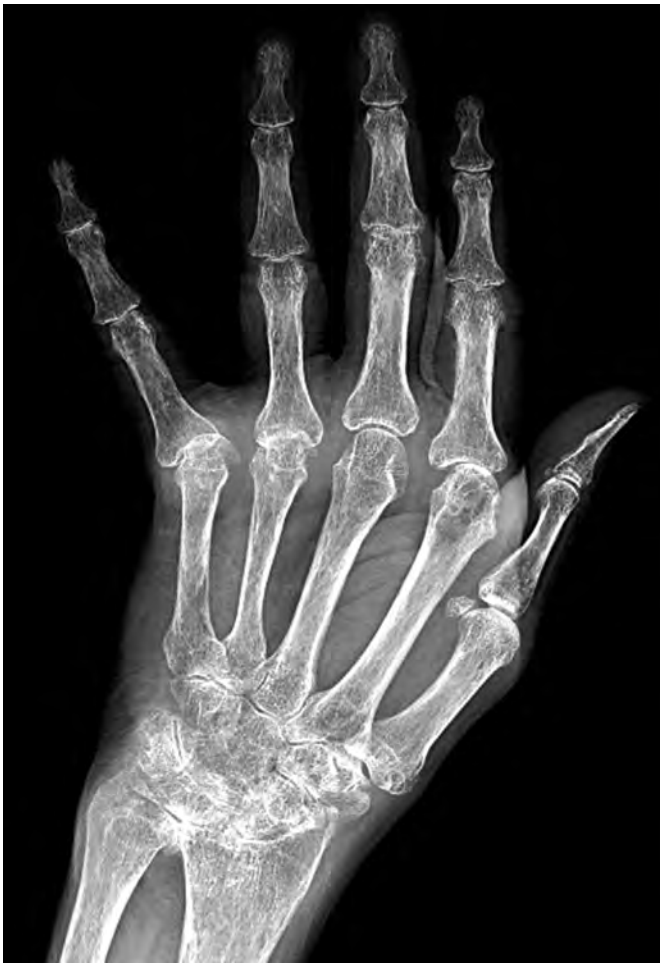


Figure 26.45 Rheumatoid arthritis of the hand. Characteristic features include diffuse osteopenia, periarticular bony erosions, and marked loss of the joint spaces of the carpal, metacarpal, phalangeal, and interphalangeal joints.

This long-term treatment with anti-TNF agents predisposes individuals to infection with opportunistic organisms such as *M. tuberculosis*. Other biologic agents that interfere with T- and B-lymphocyte responses have also been approved for therapeutic use.

Juvenile Idiopathic Arthritis

Juvenile idiopathic arthritis (JIA) is a heterogeneous group of disorders of unknown cause that present with arthritis before 16 years of age and persist for at least 6 weeks. In the United States, 30,000 to 50,000 individuals are affected. Relative to RA, JIA is more commonly associated with oligoarthritis, systemic disease, and involvement of large joints. Antinuclear antibody (ANA) seropositivity is typical, but rheumatoid nodules are usually absent. JIA pathogenesis is undefined, but, it shares features with adult RA, including associations with certain HLA and PTPN22 gene variants and studies that link its pathogenesis to involvement of Th1 and Th17 cells and inflammatory mediators such as IL-1, IL-17, TNF, and IFN-γ.

Current subclassification of JIA is based on clinical (e.g., oligoarticular, systemic) and laboratory (ANA, rheumatoid factor titers) features. Some subgroups (e.g., systemic, polyarticular rheumatoid factor–positive; enthesitis-related, which refers to involvement of sites of ligament and cartilage insertion into bone [entheses]; oligoarticular) are similar, while others (e.g., polyarticular rheumatoid factor–negative; psoriatic) are highly heterogeneous. Treatment of all subgroups is similar to adult RA, and there has been some success using a systemic IL-6 receptor antibody. Long-term JIA prognosis is variable; many patients have chronic disease, but only about 10% develop serious functional disability.

Seronegative Spondyloarthropathies

Spondyloarthropathies are a heterogeneous group of disorders unified by the following features:

- Absence of rheumatoid factor
- Pathologic changes in the ligamentous attachments (i.e. entheses) rather than synovium
- Sacroiliac joint involvement
- Association with HLA-B27
- Bony proliferation leading to ankylosis

The manifestations are immune mediated and triggered by T-cell responses, presumably against undefined microbial antigens, that cross-react with musculoskeletal system components.

Ankylosing Spondylitis

Ankylosing spondylitis causes destruction of articular cartilage and bony ankyloses, especially of the sacroiliac and vertebral apophyseal joints between tuberosities and processes. The disease presents as lower back pain and spinal immobility, usually in the second and third decades of life. Peripheral joints, such as the hips, knees, and shoulders, are involved in at least one-third of cases. Approximately 90% of patients are HLA-B27 positive. The role of HLA-B27 is unknown; it is presumably related to the ability of this MHC variant to present one or more antigens that

somehow trigger the disease, but neither the antigen nor the pathogenic immune cell is known.

Reactive Arthritis

Reactive arthritis is defined by the triad of arthritis, nongonococcal urethritis or cervicitis, and conjunctivitis following an infection. More recently, the definition has been broadened to include mono- or oligoarticular arthritis that occurs days to weeks after genitourinary *(Chlamydia)* or gastrointestinal (*Shigella, Salmonella, Yersinia, Campylobacter,* and *Clostrdioides difficile*) infections. HIV-positive patients may also be affected; the reactive arthritis in these individuals is thought to be caused by pathogens other than HIV. HLA-B27 is common in patients with reactive arthritis. The disease most often presents in young adults, with a global prevalence of approximately 30 per 100,000 adults. Arthritic episodes usually wax and wane for about 6 months, but nearly 50% of patients have recurrent arthritis, tendonitis, and lumbosacral pain. When active disease persists for more than 6 months, the term chronic reactive arthritis is used. Spondyloarthritis can occur in these patients, and spine involvement can be indistinguishable from ankylosing spondylitis.

Patients initially experience an acute-onset asymmetric oligoarthritis, often affecting the lower extremities, especially the knee. Inflammatory low back pain is a common accompanying symptom but is rarely the only symptom. Enthesitis occurs frequently and often involves the insertions of the Achilles tendon and plantar fascia on the calcaneus, resulting in swelling of the heel. *Dactylitis,* or inflammation of the digits, is present in a minority of patients and typically presents as sausage-like digits, joint stiffness, and low back pain. Synovitis of a digital tendon sheath produces "sausage" finger or toe. This may progress to ossification at tendoligamentous insertion sites and development of calcaneal spurs and bony outgrowths in those with chronic reactive arthritis. Extraarticular involvement includes conjunctivitis and uveitis; dysuria, pelvic pain, urethritis, balanitis, cervicitis, mucosal ulcers, skin rashes, and other cutaneous manifestations (e.g., psoriasis-like nail changes); and cardiac valvular disease (e.g., aortic insufficiency).

Infectious Arthritis

Microorganisms of all types can seed joints during hematogenous dissemination. Articular structures can also become infected by direct inoculation or contiguous spread from a soft tissue abscess or focus of osteomyelitis. Because cartilage, unlike bone, has a limited regenerative capacity, the rapid joint destruction that may ensue can result in permanent deformities.

Suppurative Arthritis

Bacterial infections that cause acute suppurative arthritis usually enter the joints from distant sites by hematogenous spread. In neonates, contiguous spread from underlying epiphyseal osteomyelitis is relatively common. *H. influenzae* arthritis predominates in children younger than 2 years of age. *S. aureus* is the main causative agent in older children and adults, while gonococcus is prevalent during late adolescence and young adulthood. Individuals with sickle cell disease are prone to *Salmonella* infection at any age. Except for gonococcal arthritis, which is seen mainly in women, joint infections are equally common in males and females. Individuals with deficiencies of components of the complement membrane attack complex (C5–C9) are especially susceptible to disseminated gonococcal infections and arthritis. Other predisposing conditions include immunodeficiencies (congenital and acquired), debilitating illness, joint trauma, chronic arthritis of any cause, and intravenous drug use.

The classic presentation is the sudden development of an acutely painful and swollen joint with a restricted range of motion. Fever, leukocytosis, and elevated sedimentation rate are common. In disseminated gonococcal infection, the symptoms are often subacute, and the infection usually involves only a single joint, most commonly the knee, hip, shoulder, elbow, wrist, or sternoclavicular joints. Axial joint involvement is most frequent in drug users. Joint aspiration is diagnostic if it yields purulent fluid in which the infectious organism can be identified. Prompt recognition and effective antimicrobial therapy can prevent joint destruction.

Mycobacterial Arthritis

Mycobacterial arthritis is a chronic progressive monoarticular infection caused by *M. tuberculosis*. It primarily occurs in adults as a complication of adjacent osteomyelitis or after hematogenous dissemination from a visceral (usually pulmonary) site of infection. Onset is insidious and associated with gradually increasing pain. Systemic symptoms may or may not be present. Mycobacterial seeding of the joint induces formation of confluent granulomas with caseous necrosis. The synovium may grow as a pannus over the articular cartilage and erode the bone along the joint margins. Chronic disease results in fibrous ankylosis and obliteration of the joint space. Weight-bearing joints, especially the hips, knees, and ankles, in descending order of frequency, are most commonly affected.

Viral Arthritis

Arthritis can occur with a variety of viral infections. The most common are alphavirus, parvovirus B19, rubella, Epstein-Barr virus, and hepatitis B and C viruses. The manifestations range from acute to subacute arthritis. Joint symptoms may be caused by direct infection of the joint by the virus, as in rubella and some alphavirus infections, or by an autoimmune reaction triggered by the infection.

Lyme Arthritis

Lyme arthritis is caused by the spirochete *Borrelia burgdorferi,* which is transmitted by deer ticks of the Ixodes ricinus complex. It is the leading arthropod-borne disease in the United States. It most often occurs in New England, the mid-Atlantic states, and the upper Midwest, but its geographic distribution is expanding. In the classic form, Lyme disease involves multiple organ systems through three clinical phases (Chapter 8). The initial skin infection *(early localized stage)* is followed within days or weeks by dissemination of the organism to other cutaneous sites, cranial nerves, heart, and meninges *(early disseminated stage)*. If left untreated, arthritis *(late disseminated stage)* occurs months after infection.

Currently, arthritis occurs in less than 10% of cases of *Borrelia* infection because most individuals are treated and

cured at an earlier stage. However, if left untreated, up to 80% of individuals develop a migratory arthritis *(Lyme arthritis)* lasting weeks to months. Large joints, especially the knees, shoulders, elbows, and ankles, in descending order of frequency, are usually involved. Initial attacks last for a few weeks to months, and are usually limited to one or two joints at a time. Arthritis may subsequently develop at new sites. Spirochetes are only identified in 25% of those with arthritis, but serologic detection of anti-*Borrelia* antibodies is diagnostic. Histologically, infected synovium exhibits chronic synovitis, synoviocyte hyperplasia, fibrin deposition, mononuclear cell infiltrates (especially CD4+ T cells), and *obliterative endarteritis*. In severe cases, the histopathology mimics rheumatoid arthritis.

Lyme disease treatment relies on antibiotics active against *Borrelia* and results in cure rates of 90%. Although an effective vaccine was briefly available, it was withdrawn from the market as a result of negative media reports, coverage of only the most common North American species, lack of safety data in children, and other concerns.

Chronic, antibiotic-refractory arthritis can develop in the late disseminated stage of Lyme disease. In many of these cases, *Borrelia* cannot be detected in the joint fluid, even by polymerase chain reaction. Some have hypothesized that this antibiotic-refractory arthritis is an autoimmune disease triggered by immune responses to *Borrelia* outer surface protein A. Chronic disease is also associated with nonspecific symptoms such as fatigue and cognitive complaints collectively termed *posttreatment Lyme disease syndrome*.

Crystal-Induced Arthritis

Articular crystal deposits are associated with a variety of acute and chronic joint disorders. Endogenous crystals include monosodium urate *(gout)*, calcium pyrophosphate dihydrate *(pseudogout)*, and basic calcium phosphate. Exogenous crystals, such as the biomaterials used in prosthetic joints, can also induce arthritis as they accumulate with wear. Endogenous and exogenous crystals produce disease by triggering inflammatory reactions that destroy cartilage.

Gout

Gout is marked by transient attacks of acute arthritis initiated by monosodium urate crystals deposited within and around joints. Whether gout is primary or secondary to another underlying disease (Table 26.6), the common feature is excessive uric acid in the tissues and body fluids. In the primary form (90% of cases), gout is the major manifestation of the disease; the cause is unknown.

Pathogenesis

Hyperuricemia (plasma urate above 6.8 mg/dL) is necessary, but not sufficient, for development of gout. Elevated uric acid can result from overproduction, reduced excretion, or both (Table 26.6). Uric acid levels are determined by several factors:

- *Uric acid production.* Urate synthesis is the end product of purine catabolism. Purines are themselves the product of two interlinked pathways: the de novo pathway, in which purine nucleotides are synthesized from nonpurine precursors, and the salvage pathways, in which nucleotides are synthesized from free purine bases in the diet and those generated by purine nucleotide catabolism.
- *Uric acid excretion.* In the kidney, uric acid is filtered by the glomerulus, but is almost completely resorbed by the proximal tubule. The small fraction of total uric acid in the urine is the result of secretion in the distal nephron.

Table 26.6 Classification of Gout

Clinical Category	Uric Acid Production	Uric Acid Excretion
Primary Gout (90%)		
Unknown enzyme defects (85%–90%)	Normal ↑	↓ Normal
Known enzyme defects (e.g., partial HGPRT deficiency)	↑	Normal
Secondary Gout (10%)		
Increased nucleic acid turnover (e.g., leukemia)	↑↑	↑
Chronic renal disease	Normal	↓
Congenital (e.g., Lesch-Nyhan syndrome, HGPRT deficiency)	↑↑	↑

HGPRT, Hypoxanthine guanine phosphoribosyl transferase.

Asymptomatic hyperuricemia appears around puberty in males and after menopause in females. In primary gout, hyperuricemia is usually due to reduced excretion. The mechanism is not known, but genome-wide association studies have identified polymorphisms in genes including *URAT1, GLUT9,* and *KCNQ1,* which are involved in transport and homeostasis of urate and other ions. A small fraction of primary gout cases are caused by uric acid overproduction due to identifiable enzymatic defects. For example, partial deficiency of hypoxanthine guanine phosphoribosyl transferase (HGPRT) interrupts the salvage pathway, so purine metabolites cannot be salvaged and are, instead, degraded into uric acid. Complete absence of HGPRT also results in hyperuricemia, but the significant neurologic manifestations of this condition *(Lesch-Nyhan syndrome)* dominate the clinical picture, which is therefore classified as secondary gout. Secondary gout can also be caused by increased production (e.g., rapid cell lysis during chemotherapy for leukemia) or decreased excretion (chronic renal disease).

Inflammation is triggered by precipitation of monosodium urate crystals into the joints, which result in the production of cytokines that recruit leukocytes (Fig. 26.46). Macrophages and neutrophils phagocytose the urate crystals. This activates the inflammasome (Chapter 3) and leads to the secretion of cytokines, including IL-1, which promote accumulation of more neutrophils and macrophages within the joint. A vicious cycle develops as newly recruited inflammatory cells release cytokines, free radicals, proteases, and arachidonic acid metabolites to recruit even more leukocytes. Complement activation by the alternative pathway may also contribute to leukocyte recruitment. Rupture of phagolysosomes induced by ingested urate crystals can cause further release of proteases and inflammatory mediators. The resulting acute arthritis typically remits spontaneously in days to weeks. Repeated attacks of acute arthritis lead eventually to the formation of *tophi,* aggregates of urate crystals and inflammatory tissue in the inflamed synovial membrane

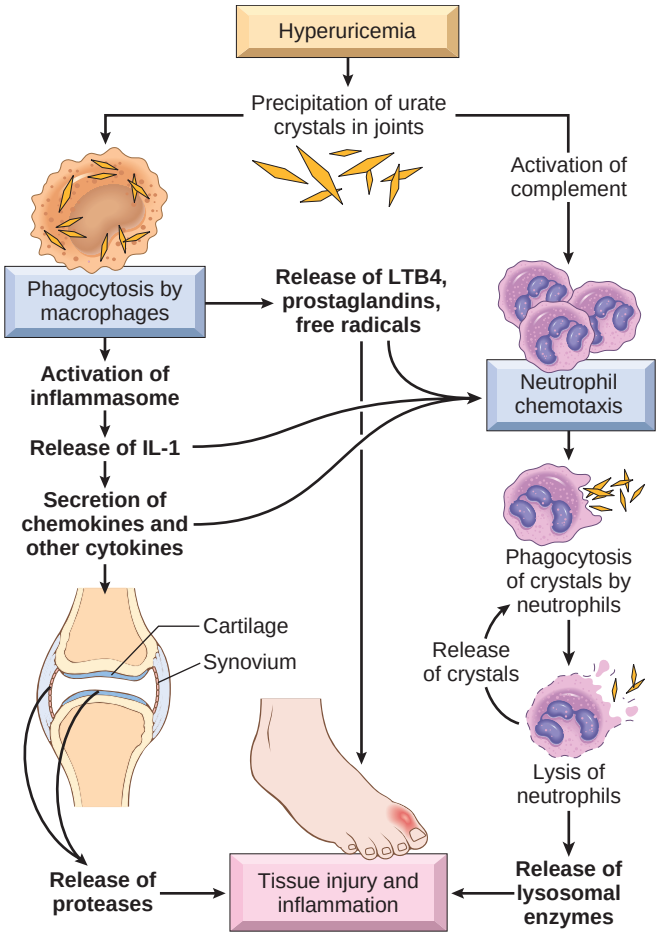


Figure 26.46 Pathogenesis of acute gouty arthritis. Urate crystals are phagocytosed by macrophages and stimulate the production of various inflammatory mediators that elicit the inflammation characteristic of gout. Note that IL-1 stimulates the production of chemokines and other cytokines from a variety of cells. *LTB4*, Leukotriene B4.

and periarticular tissue as well as severe cartilage damage that compromises joint function.

Only about 10% of individuals with hyperuricemia develop gout. Other factors linked to disease include the following:

- Age of the individual and duration of the hyperuricemia, as gout usually appears after 20 to 30 years of hyperuricemia
- Male sex
- Genetic predisposition, including X-linked *HGPRT* abnormalities and primary gout, which has a multigenic pattern of inheritance
- Alcohol consumption
- Obesity
- Drugs (e.g., thiazides) that reduce urate excretion

MORPHOLOGY

The distinctive patterns of gout are (1) acute arthritis, (2) chronic tophaceous arthritis, (3) accumulation of tophi in extraarticular sites, and (4) urate (gouty) nephropathy.

Acute arthritis is characterized by a dense neutrophilic infiltrate that permeates the synovium and synovial fluid. Slender, needle-shaped urate crystals are arranged in small clusters in the synovium and are frequently found in the cytoplasm of neutrophils within aspirated joint fluid (Fig. 26.47A). The edematous and congested synovium also contains scattered lymphocytes, plasma cells, and macrophages. The acute attack remits when the episode of crystallization abates and crystals are solubilized.

Chronic tophaceous arthritis evolves from repetitive precipitation of urate crystals during acute attacks. Urate encrusts the articular surface and forms visible deposits in the synovium, which becomes hyperplastic, fibrotic, and thickened by inflammatory cells. The resulting pannus destroys the underlying cartilage and leads to juxta-articular bone erosions. In severe cases, fibrous or bony ankylosis ensues, resulting in loss of joint function.

Tophi are the pathognomonic hallmark of gout. They are formed by large aggregations of urate crystals surrounded by an intense inflammatory reaction of foreign-body giant cells. (Fig. 26.47B–C). Tophi may appear in the articular cartilage, ligaments, tendons, and bursae. Less frequently they occur in soft tissues (earlobes, fingertips) or kidneys. Superficial tophi can ulcerate through the overlying skin.

Urate nephropathy refers to the renal complications caused by urate crystals or tophi in the renal medullary interstitium or tubules. Complications include uric acid nephrolithiasis and pyelonephritis, particularly when urinary obstruction occurs.

Clinical Features

Primary gout initially presents as acute arthritis with sudden-onset, excruciating joint pain, localized hyperemia and warmth, and, occasionally, mild fever; constitutional symptoms are uncommon. Most first attacks are monoarticular, and 50% occur in the first metatarsophalangeal joint of the big toe. Untreated, acute gouty arthritis may last for hours to weeks, but gradually there is complete resolution followed by a symptom-free interval, known as the intercritical period. Some individuals never have another attack, but most experience a second acute episode within months to a few years. In the absence of appropriate therapy, the attacks recur at shorter intervals and frequently become polyarticular, involving other parts of the foot and less commonly the upper extremity. Chronic tophaceous gout develops about 10 years after the initial acute attack and is characterized by juxta-articular bone erosion caused by osteoclastic bone resorption and loss of the joint space.

Nonpharmacologic treatments for gout include lifestyle modifications, such as weight loss in obese individuals; dietary changes to reduce purine intake, particularly of animal and seafood proteins; reduced alcohol and sugar-sweetened beverage consumption; and regular exercise. Pharmacologic therapies include uricosuric drugs (e.g., probenecid), xanthine oxidase inhibitors (e.g., allopurinol), urate oxidases (uricases), NSAIDs, and colchicine. Finally, gout can be precipitated by serum urate–increasing drugs taken for comorbidities; their replacement with alternatives may be helpful. Generally, gout does not shorten the life span, but it does impact quality of life.

Calcium Pyrophosphate Crystal Deposition Disease (Pseudogout)

Calcium pyrophosphate crystal deposition disease (CPPD), also known as *pseudogout,* usually occurs in individuals older than 50 years of age and becomes more common

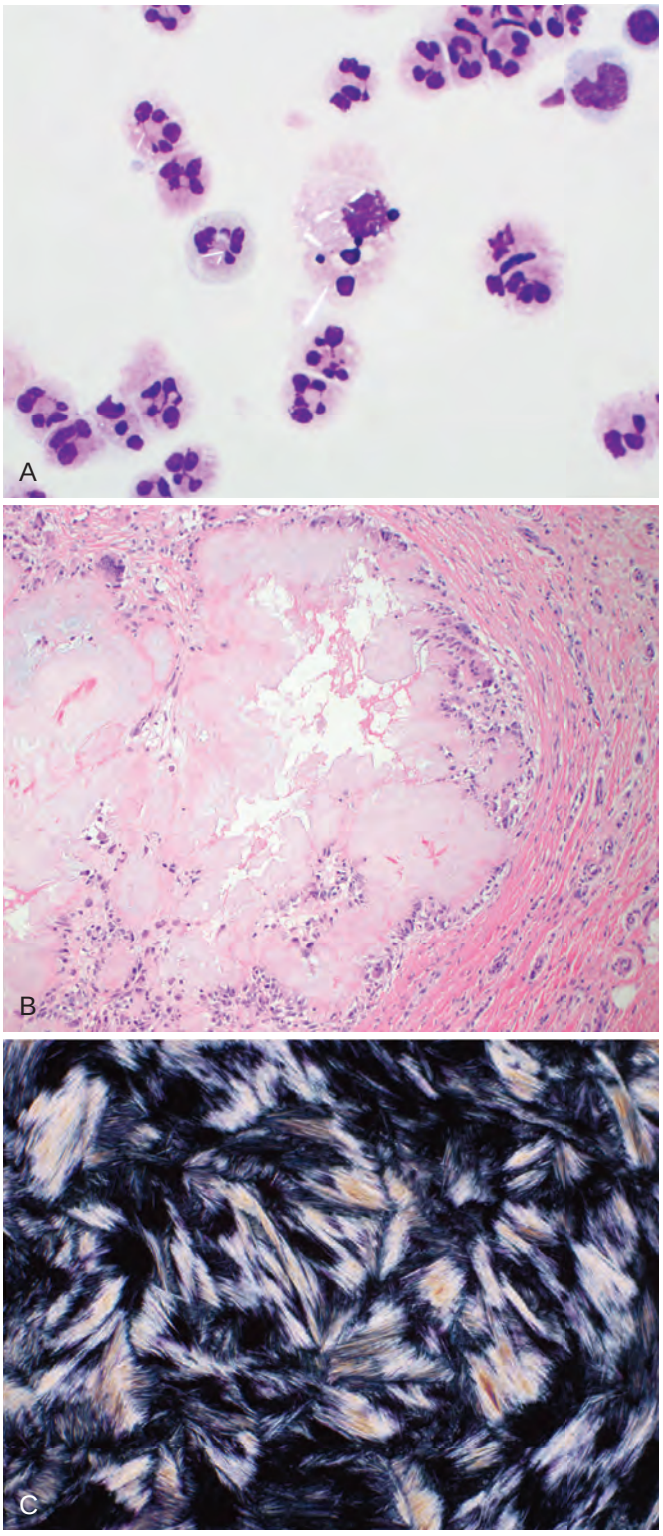


Figure 26.47 Gout. (A) Joint aspirate from acute gout arthritis demonstrates polarizable needle shape urate crystals in neutrophils. (B) Gouty tophus—an aggregate of dissolved urate crystals is surrounded by reactive fibroblasts, mononuclear inflammatory cells, and giant cells. (C) Needle-shaped urate crystals are easily identified under polarized light.

with increasing age; up to 60% of those 85 years of age or older are affected. There is no sex or racial predisposition. CPPD is divided into sporadic (idiopathic), hereditary, and secondary types. An autosomal dominant variant is caused by germline gain-of-function mutations in *ANKH,* an inorganic pyrophosphate transporter, that result in crystal deposition and osteoarthritis at relatively early ages. These mutations are distinct from those associated with craniometaphyseal dysplasia. Various disorders, including previous joint damage, hyperparathyroidism, hemochromatosis, hypomagnesemia, hypothyroidism, ochronosis, and diabetes predispose to secondary CPPD.

Pathogenesis

The basis for crystal formation is not known, but studies suggest that degradation of articular cartilage proteoglycans, which normally inhibit mineralization, allows crystallization around chondrocytes. As in gout, inflammation is caused by activation of the inflammasome in macrophages (see Fig. 26.46).

MORPHOLOGY

The crystals first develop in the articular cartilage, menisci, and intervertebral discs; as the deposits enlarge they may rupture and seed the joint. The crystals form chalky, white friable deposits, which are seen histologically in stained preparations as oval blue-purple aggregates (Fig. 26.48A). Individual crystals are rhomboid, 0.5 to 5 μm in greatest dimension (Fig. 26.48B), and birefringent. Inflammation is usually milder than in gout.

Clinical Features

CPPD is frequently asymptomatic. However, it may produce acute, subacute, or chronic arthritis that can be confused with osteoarthritis or rheumatoid arthritis clinically. Joint involvement may last from several days to weeks and may be monoarticular or polyarticular; the knees, followed by the wrists, elbows, shoulders, and ankles, are most commonly affected. Chondrocalcinosis, radiographically evident calcification of hyaline or fibrocartilage, is often present but is not synonymous with CPPD. Ultimately, approximately 50% of affected individuals experience significant joint damage. Therapy is supportive to minimize symptoms. There is no known treatment that prevents or slows crystal formation.

KEY CONCEPTS

ARTHRITIS

- Osteoarthritis (degenerative joint disease), the most common disease of joints, is a degenerative disorder of articular cartilage in which matrix breakdown exceeds synthesis. Inflammation is minimal and typically secondary, but locally produced cytokines may contribute to progression of joint degeneration.
- Rheumatoid arthritis is a chronic autoimmune inflammatory disease that affects mainly small joints but can be systemic. RA is caused by a cellular and humoral immune response against self antigens, particularly citrullinated proteins. TNF plays a central role, and anti-TNF biologics are effective.
- Seronegative spondyloarthropathies are a heterogeneous group of autoimmune arthritides that preferentially involve the

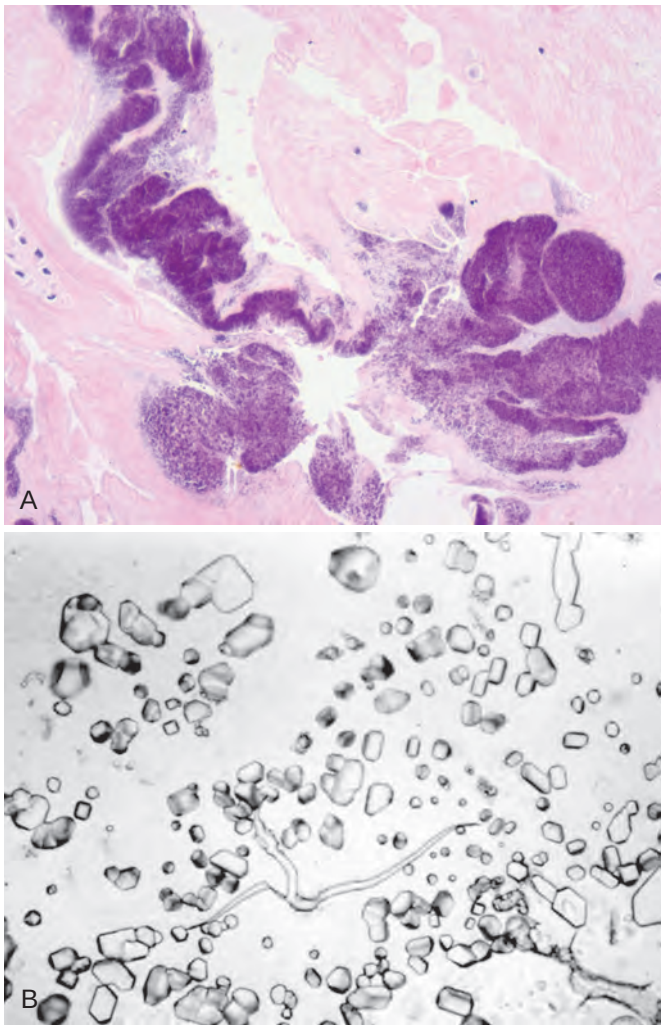


Figure 26.48 Pseudogout. (A) Deposits are present in cartilage and consist of amorphous basophilic material. (B) Smear preparation of calcium pyrophosphate crystals.

sacroiliac, vertebral joints and entheses and are associated with HLA-B27. Reactive arthritis is an uncommon immune-mediated disorder that develops after resolution of genitourinary or gastrointestinal infection.

- Suppurative arthritis describes direct infection of a joint space by bacterial organisms.
- Lyme disease is a systemic infection by *B. burgdorferi* that manifests, in part, as an infectious arthritis, possibly with an autoimmune component in chronic stages.
- Gout and pseudogout result from inflammatory responses to precipitated urate or calcium pyrophosphate, respectively.

JOINT TUMORS AND TUMOR-LIKE CONDITIONS

Reactive tumor-like lesions, such as ganglion cysts, synovial cysts, and osteochondral loose bodies, commonly involve joints and tendon sheaths. They usually result from trauma or degenerative processes and are much more common than neoplasms. Primary neoplasms are rare, usually benign, and tend to recapitulate the cells and tissue types (synovial membrane, fat, blood vessels, fibrous tissue, and cartilage) native to joints and related structures. The rare malignant tumors are discussed with soft tissue tumors.

Ganglion and Synovial Cysts

Ganglion cysts are small, up to 1.5 cm, and nearly always located near a joint capsule or tendon sheath. A common location is around the joints of the wrist, where the cysts appear as firm, fluctuant, pea-sized translucent nodules. Ganglion cysts develop due to cystic or myxoid connective tissue degeneration; hence the cyst wall lacks a cell lining. Lesions may be multilocular and can enlarge through coalescence and degeneration of adjacent connective tissue. The fluid that fills ganglion cysts is similar to synovial fluid, but there is no communication with the joint space. Despite the name, the lesion is unrelated to ganglia of the nervous system.

Herniation of synovium through a joint capsule or massive enlargement of a bursa may produce a *synovial cyst.* A well-recognized example is the popliteal synovial cyst, or *Baker cyst,* associated with rheumatoid arthritis. The cyst lining resembles the synovium, and both cyst and synovium may be hyperplastic. Cyst fluid often contains inflammatory cells and fibrin.

Tenosynovial Giant Cell Tumor

Tenosynovial giant cell tumor is a benign neoplasm that develops in the synovial lining of joints, tendon sheaths, and bursae. Tenosynovial giant cell tumor can be diffuse (previously known as *pigmented villonodular synovitis*) or localized. The diffuse type tends to involve large joints, while the localized type usually occurs as a discrete nodule attached to a tendon sheath, commonly in the hand. Both variants are most often diagnosed in the 20s to 40s and affect the sexes equally.

Pathogenesis

Both diffuse and localized tenosynovial giant cell tumors harbor a reciprocal somatic chromosomal translocation, t(1;2)(p13;q37), resulting in fusion of the type VI collagen α-3 promoter upstream of the coding sequence of the *MCSF* gene. As a result, the tumor cells overexpress M-CSF, which, through autocrine and paracrine effects, stimulates proliferation of macrophages in a manner similar to giant cell tumor of bone (described earlier).

MORPHOLOGY

Tenosynovial giant cell tumors are red-brown to orange-yellow. In diffuse tumors, the normally smooth joint synovium is converted into a tangled mat by red-brown folds, fingerlike projections, and nodules (Fig. 26.49A); localized tumors are well circumscribed. The neoplastic cells, which are a minority of the cells in the mass, are polygonal, moderately sized, and resemble synoviocytes (Fig. 26.49B). In the diffuse variant, they spread along the surface and infiltrate the subsynovial tissue. A solid aggregate may be attached to the synovium by a pedicle. Both diffuse and localized types are densely infiltrated by macrophages that may contain hemosiderin or foamy lipids and can be multinucleated. Older lesions may become fibrotic.

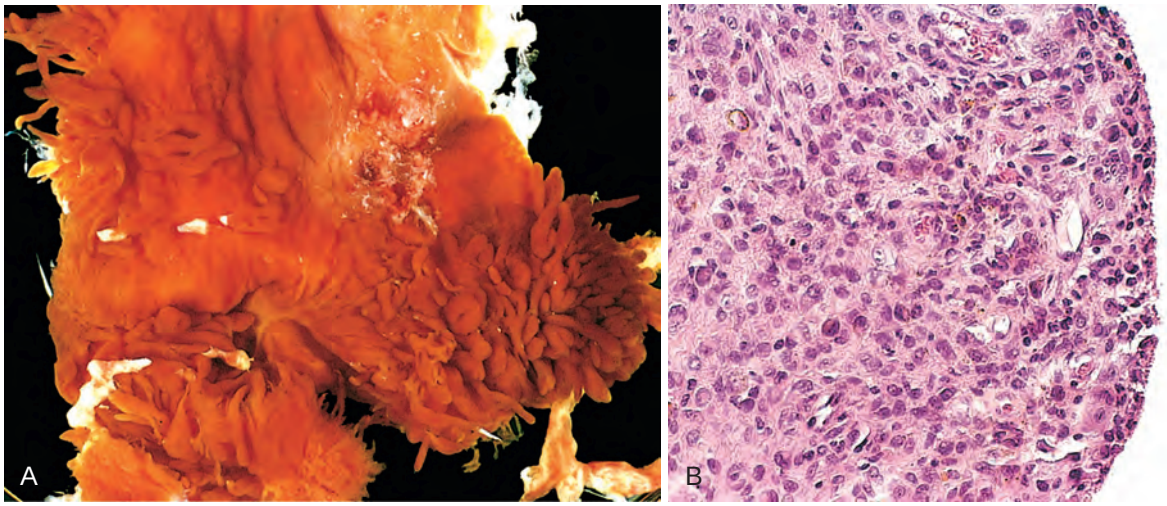


Figure 26.49 Tenosynovial giant cell tumor, diffuse type. (A) Excised synovium with fronds and nodules typical of diffuse tenosynovial giant cell tumor. The color and texture explain the old name of pigmented villonodular synovitis. (B) Sheets of proliferating cells in tenosynovial giant cell tumor bulging the synovial lining.

Clinical Features

Diffuse tenosynovial giant cell tumors present in the knee in 80% of cases. Affected individuals complain of pain, locking, and recurrent swelling, similar to monoarticular arthritis. Sometimes a palpable mass is appreciated. Aggressive tumors erode into adjacent bones and soft tissues, causing confusion with other types of neoplasms. Recurrence is common. Localized tenosynovial giant cell tumors present as a solitary, slow-growing, painless mass that frequently involves the tendon sheaths along the wrists and fingers. Bone erosion and recurrence are less common than in the diffuse type. Surgical excision is the mainstay of treatment. Clinical trials using antagonists of M-CSF signaling have had promising results.

Soft Tissue Tumors

By convention, soft tissue refers to non-epithelial tissue excluding the skeleton, joints, central nervous system, and hematopoietic and lymphoid tissues. Although nonneoplastic conditions can involve soft tissue, they are seldom confined to this compartment, so the area of soft tissue pathology is restricted to neoplasms. With the exception of skeletal muscle neoplasms, benign soft tissue tumors are 100-fold more frequent than their malignant counterparts, the sarcomas. In the United States, the incidence of soft tissue sarcomas is approximately 12,000 per year, which is less than 1% of all cancers. Sarcomas, however, cause 2% of all cancer mortality, reflecting their aggressive behavior and resistance to chemotherapy. Most soft tissue tumors arise in the extremities, particularly the thigh. Approximately 15% arise in children; the incidence increases with age.

Pathogenesis

Most sarcomas are sporadic and have no known predisposing cause. A small minority are associated with germline mutations in tumor suppressor genes that are associated with well-described syndromes (neurofibromatosis 1, Gardner syndrome, Li-Fraumeni syndrome, Osler-Weber Rendu syndrome). Others are linked to known environmental exposures such as radiation, burns, or toxins.

Unlike carcinomas and certain hematologic malignancies, which often arise from well-recognized precursor lesions, sarcoma precursors are undefined. Some sarcomas recapitulate a recognizable mesenchymal lineage (e.g. skeletal muscle), but they are thought to arise from pluripotent mesenchymal stem cells, which acquire somatic "driver" mutations in oncogenes and tumor suppressor genes. The genetics of tumorigenesis are heterogeneous, but some generalizations can be made based on karyotypic complexity:

- *Simple karyotype* (20%): Like many leukemias and lymphomas, sarcomas are occasionally euploid tumors with a single or limited number of chromosomal changes (Table 26.7) that occur early in tumorigenesis and are specific enough to serve as diagnostic markers. Tumors with these features most commonly arise in younger individuals and tend to have a monomorphic microscopic appearance. Examples include Ewing sarcoma (discussed earlier) and synovial sarcoma. In some cases, the oncogenic effect of chromosomal rearrangements is reasonably well understood (see Table 26.7). In others, the mechanisms are unknown. Oncogenic tumor-specific fusion proteins represent potential targets for therapy.
- *Complex karyotype* (80%): These tumors are usually aneuploid or polyploid and demonstrate multiple chromosomal gains and losses, a feature that suggests an underlying abnormality producing genomic instability. Examples include leiomyosarcomas and undifferentiated pleomorphic sarcoma. Such tumors are more common in adults and tend to be microscopically diverse, with a

Table 26.7 Chromosomal Abnormalities in Soft Tissue Tumors

Tumor	Cytogenetic Abnormality	Gene Fusion	Proposed Function
Ewing sarcoma	t(11;22)(q24;q12) t(21;22)(q22;q12)	*EWS–FLI1* *EWS–ERG*	Disordered protein with multiple functions, including aberrant transcription, cell cycle regulation, RNA splicing, and telomerase
Extraskeletal myxoid chondrosarcoma	t(9;22)(q22;q12)	*EWS–CHN*	
Desmoplastic small round-cell tumor	t(11;22)(p13;q12)	*EWS–WT1*	
Clear-cell sarcoma	t(12;22)(q13;q12)	*EWS–ATF1*	
Liposarcoma—myxoid type	t(12;16)(q13;p11)	*FUS–DDIT3*	Arrests adipocytic differentiation
Synovial sarcoma	t(x;18)(p11;q11)	*SS18–SSX1* *SS18–SSX2* *SS18–SSX4*	Chimeric transcription factors, interrupts cell cycle control
Rhabdomyosarcoma—alveolar type	t(2;13)(q35;q14) t(1;13)(p36;q14)	*PAX3–FOXO1* *PAX7–FOXO1*	Chimeric transcription factors, disrupts skeletal muscle differentiation
Dermatofibrosarcoma protuberans	t(17;22)(q22;q15)	*COL1A1–PDGFB*	Promoter driven overexpression of PDGF-β, autocrine stimulation
Alveolar soft-part sarcoma	t(X;17)(p11.2;q25)	*TFE3–ASPL*	Unknown
Infantile fibrosarcoma	t(12;15)(p13;q23)	*ETV6–NTRK3*	Chimeric tyrosine kinase leads to constitutively active Ras/MAPK pathway
Nodular fasciitis	t(22;17)	*MYH9–USP6*	Increased Wnt/β-catenin signaling

range of cell sizes and shapes within a single tumor (i.e., they are pleomorphic).

Classification of soft tissue tumors continues to evolve as new molecular genetic abnormalities are identified. Clinically, soft tissue tumors range from benign, self-limited lesions that require minimal treatment to intermediate-grade, locally aggressive tumors with limited metastatic risk to highly aggressive malignancies with significant metastatic risk and mortality. All highly aggressive malignancies are classified as *sarcomas,* but this term is used less consistently among locally aggressive, nonmetastasizing tumors. Pathologic classification integrates morphology (e.g., muscle differentiation), immunohistochemistry, and molecular diagnostics (Table 26.8). In addition to accurate diagnosis, grade (degree of differentiation) and stage (size and depth) are important prognostic indicators. The next section will consider representative soft tissue tumors.

TUMORS OF ADIPOSE TISSUE

Lipoma

Lipoma, a benign tumor of fat, is the most common soft tissue tumor in adults. The conventional lipoma is the most common subtype, from which rare variants are distinguished according to characteristic morphologic and genetic features.

MORPHOLOGY

The conventional lipoma is a well-encapsulated mass of mature adipocytes. It usually arises in the subcutis of the proximal extremities and trunk during middle adulthood. Infrequently, lipomas are large, intramuscular, and poorly circumscribed. **Lipomatosis** occurs when multifocal lipomas involve a limb. Most lipomas are soft, mobile, painless, and cured by simple excision.

Liposarcoma

Liposarcomas, malignant tumors of adipose tissue, are the most common sarcomas of adulthood. These tumors typically develop in deep soft tissues of the proximal extremities and retroperitoneum of individuals in the sixth and seventh decades of life.

Pathogenesis

The three distinct subtypes of liposarcoma, well-differentiated, myxoid, and pleomorphic, have different genetic aberrations. Well-differentiated liposarcomas harbor amplifications of chromosome region 12q13–q15, which includes the p53 inhibitor *MDM2*. In myxoid liposarcoma, a fusion gene generated by a t(12;16) translocation arrests adipocyte differentiation, leading to unregulated proliferation of primitive cells. Pleomorphic liposarcomas have complex karyotypes without reproducible genetic abnormalities.

MORPHOLOGY

Liposarcomas are histologically divided into **well-differentiated, myxoid,** and **pleomorphic** subtypes. Well-differentiated liposarcomas are composed of mature adipocytes with scattered atypical spindle cells (Fig. 26.50A). Myxoid liposarcoma is characterized by abundant basophilic extracellular matrix, arborizing capillaries, and primitive cells at various stages of adipocyte differentiation resembling fetal fat (Fig. 26.50B). Pleomorphic liposarcomas are characterized by sheets of anaplastic cells with bizarre nuclei admixed with variable numbers of immature adipocytes, termed lipoblasts.

Clinical Features

Liposarcomas recur locally, and often repeatedly, unless adequately excised. Well-differentiated tumors can be indolent, while pleomorphic liposarcomas are aggressive and frequently metastasize; the behavior of myxoid tumors is intermediate between these extremes.

Table 26.8 Clinical Features of Soft Tissue Tumors

Category	Behavior	Tumor Type	Common Locations	Age (years)	Morphology
Adipose	Benign	Lipoma	Superficial extremity, trunk	40–60	Mature adipose tissue
	Malignant	Well-differentiated liposarcoma	Deep extremity, retroperitoneum	50–60	Adipose tissue with scattered atypical spindle cells
		Myxoid liposarcoma	Thigh, leg	30s	Myxoid matrix, "chicken wire" vessels, round cells, lipoblasts
Fibrous	Benign	Nodular fasciitis	Arm, forearm	20–30	Tissue culture growth, extravasated erythrocytes,
		Deep fibromatosis	Abdominal wall	30–40	Dense collagen, long, unidirectional fascicles
Skeletal muscle	Benign	Rhabdomyoma	Head and neck	0–60	Polygonal rhabdomyoblasts, "spider" cells
	Malignant	Alveolar rhabdomyosarcoma	Extremities, sinuses	5–15	Uniform round discohesive cells between septae
		Embryonal rhabdomyosarcoma	Genitourinary tract	1–5	Primitive spindle cells, "strap" cells
Smooth muscle	Benign	Leiomyoma	Extremity	20s	Uniform, plump eosinophilic cells in fascicles
	Malignant	Leiomyosarcoma	Thigh, retroperitoneum	40–60	Pleomorphic eosinophilic cells
Vascular	Benign	Hemangioma	Head and neck	0–10	Circumscribed mass of capillary or venous channels
	Malignant	Angiosarcoma	Skin, deep lower extremity	50–80	Infiltrating capillary channels
Nerve sheath	Benign	Schwannoma	Head and neck	20–50	Encapsulated, fibrillar stroma, nuclear palisading
		Neurofibroma	Wide, cutaneous, subcutis	10–20+	Myxoid, ropy collagen, loose fascicles, mast cells
	Malignant	Malignant peripheral nerve sheath tumor	Extremities, shoulder girdle	20–50	Tight fascicles, atypia, mitotic activity, necrosis
Uncertain histotype	Benign	Solitary fibrous tumor	Pelvis, pleura	20–70	Branching ectatic vessels
	Malignant	Synovial sarcoma	Thigh, leg	15–40	Tight fascicles of uniform basophilic spindle cells, pseudoglandular structures
		Undifferentiated pleomorphic sarcoma	Thigh	40–70	High-grade anaplastic polygonal, round, or spindle cells, bizarre nuclei, atypical mitoses, necrosis

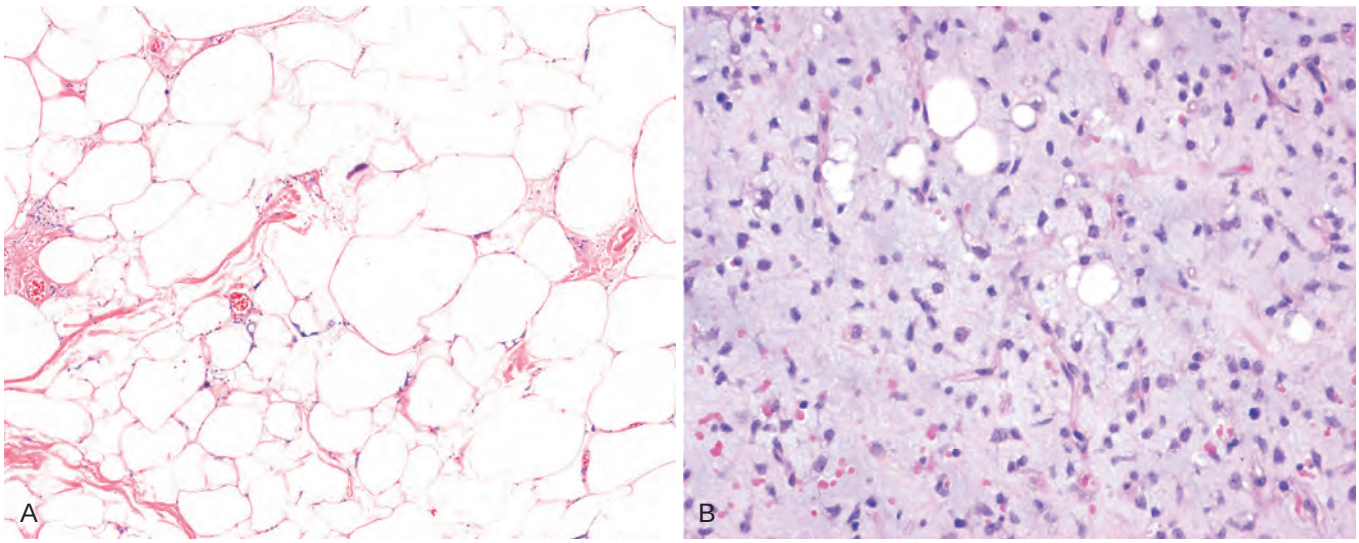


Figure 26.50 Liposarcoma. (A) the well-differentiated subtype consists of mature adipocytes and scattered spindle cells with hyperchromatic nuclei. (B) Myxoid liposarcoma with abundant basophilic ground substance and a rich capillary network with scattered immature adipocytes and more primitive round to stellate cells.

FIBROUS TUMORS

Nodular Fasciitis

Nodular fasciitis is a self-limited fibroblastic and myofibroblastic proliferation that occurs in young adults. The tumors occur most frequently on the forearm, chest, and back and grow rapidly over weeks to months. Due to a history of trauma in 10% to 50% of cases, this lesion was previously considered a reactive lesion. It is, however, a clonal proliferation that harbors a t(17;22) translocation that produces a *MYH9-USP6* fusion. The defect that prevents neoplastic cells from becoming malignant has not been defined. It is, however, interesting to note that aneurysmal bone cyst, a benign neoplasm previously considered reactive, also contains a *USP6* fusion gene, albeit to partners other than *MYH9*. Nodular fasciitis typically regresses spontaneously and, if excised, rarely recurs.

MORPHOLOGY

Nodular fasciitis arises in the deep dermis, subcutis, fascia, or muscle. Grossly the non-encapsulated lesions are well-circumscribed or slightly infiltrative and less than 3 cm in diameter. Histologically, nodular fasciitis is composed of plump, immature-appearing fibroblasts and myofibroblasts containing elongated nuclei with punctate nucleoli. Mitoses are frequent, but atypical forms are notably absent (Fig. 26.51). A gradient, termed zonation, transitions from hypercellular regions with myxoid stroma to hypocellular areas with fibrous stroma. Storiform or fascicular patterns are common in cellular areas. Metaplastic bone, cystic areas, ganglion-like cells, prominent vessels, extravasated red cells, and infiltrating lymphocytes are common.

Fibromatoses

Superficial Fibromatosis

Superficial fibromatosis is a benign growth that can cause local deformity but has an innocuous clinical course. It affects males more frequently than females. These nodular, proliferative lesions are composed of plump spindle cells arranged in poorly defined broad bundles or long, sweeping fascicles surrounded by abundant dense collagen. Clinical subtypes include Dupuytren contracture, Ledderhose disease, and Peyronie disease.

- *Dupuytren contracture,* or *palmar fibromatosis,* is an irregular or nodular thickening of the palmar fascia that can be unilateral or bilateral. Incidence increases with age. Attachment to overlying skin can cause puckering or dimpling and a slowly progressive flexion contracture.
- *Ledderhose disease,* or *plantar fibromatosis,* occurs in boys from under 10 years of age into adolescence. It is unilateral and does not cause contractures but can be associated with palmar and penile fibromatosis.
- *Peyronie disease,* or *penile fibromatosis,* is a palpable induration or mass on the dorsolateral aspect of the penis. Eventually, it may cause abnormal curvature of the shaft and constriction of the urethra.

Palmar and plantar fibromatoses progress in about 50% of cases. The remainder stabilize and do not progress; some resolve spontaneously. Nevertheless, recurrence is common, even after excision.

Deep Fibromatosis (Desmoid Tumors)

Deep fibromatoses, also called desmoid tumors, are large, infiltrative masses that frequently recur but do not metastasize. They arise most frequently in the teens to 30s, predominantly in women. Abdominal fibromatosis generally arises in the musculoaponeurotic structures of the anterior abdominal wall, but tumors can also arise in the limb girdles or mesentery. Deep fibromatoses contain mutations in the *APC* or *CTNNB1* (β-catenin) genes, both of which lead to increased Wnt signaling. The majority of tumors harbor sporadic *CTNNB1* mutations, but individuals with familial adenomatous polyposis, who have germline *APC* mutations, are predisposed to deep fibromatosis (Gardner syndrome, Chapter 17).

MORPHOLOGY

Fibromatoses are gray-white, firm, poorly demarcated masses varying from 1 to 15 cm in diameter. They are rubbery and tough and infiltrate surrounding muscle, nerve, and fat. Bland fibroblasts arranged in long parallel fascicles amid dense collagen are characteristic (Fig. 26.52). The resulting histologic appearance can resemble a scar.

Clinical Features

In addition to possibly being disfiguring or disabling, deep-seated fibromatosis can be painful. Because they are infiltrative, complete excision of deep fibromatosis can be difficult. Recent efforts have concentrated on medical or radiation therapy as alternatives to surgery.

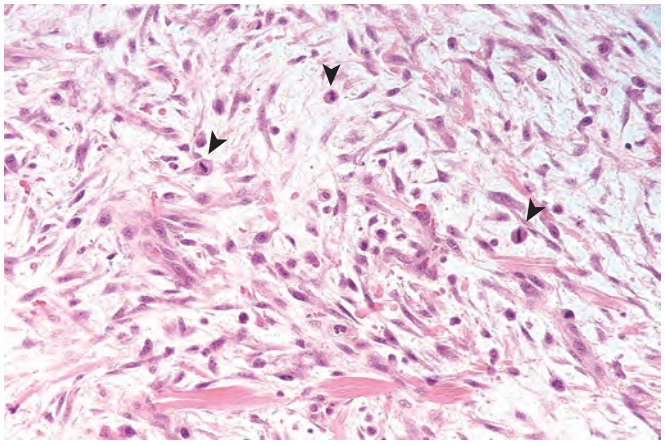


Figure 26.51 Nodular fasciitis with plump, randomly oriented spindle cells surrounded by myxoid stroma. Note the mitotic activity *(arrowheads)* and extravasated red blood cells.

SKELETAL MUSCLE TUMORS

In contrast to tumors of other cell types, skeletal muscle neoplasms are almost all malignant. The benign variant,

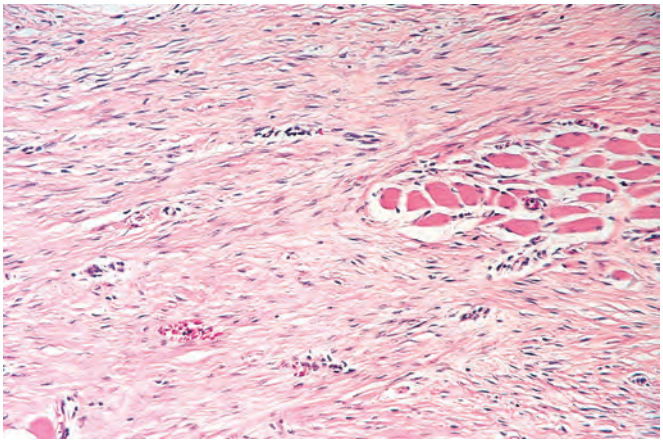


Figure 26.52 Deep fibromatosis infiltrating skeletal muscle.

rhabdomyoma, is frequently associated with tuberous sclerosis (Chapter 28).

Rhabdomyosarcoma

Rhabdomyosarcoma is a malignant mesenchymal tumor with skeletal muscle differentiation. Four subtypes are recognized: alveolar (20%), embryonal (50%), pleomorphic (20%), and spindle cell/sclerosing (10%). Alveolar and embryonal subtypes of rhabdomyosarcoma are the most common soft tissue sarcoms a of childhood and adolescence, usually appearing before 20 years of age. Pleomorphic rhabdomyosarcoma occurs in adults, while the spindle cell/sclerosing type affects all ages. Pediatric rhabdomyosarcomas often arise in the sinuses, head and neck, and genitourinary tract; these locations do not normally contain much skeletal muscle, underscoring the hypothesis that sarcomas do not arise from mature, terminally differentiated cells. Alveolar rhabdomyosarcoma frequently contains fusions of the *FOXO1* gene to either *PAX3* or *PAX7* due to t(2;13) or t(1;13) translocations, respectively. PAX3 is a transcription factor that initiates skeletal muscle differentiation; the chimeric PAX3-FOXO1 fusion protein interferes with differentiation, a mechanism similar to many of transcription factor fusion proteins found in acute leukemia. The remaining subtypes of rhabdomyosarcoma are genetically heterogeneous.

MORPHOLOGY

Embryonal rhabdomyosarcoma is a soft, gray, infiltrative mass. The tumor cells recapitulate skeletal muscle at various stages of differentiation and include sheets of primitive round and spindled cells in myxoid stroma (Fig. 26.53A). Rhabdomyoblasts with visible cross-striations may be present. **Sarcoma botryoides** (Chapter 22) is an embryonal rhabdomyosarcoma variant that develops in the walls of hollow, mucosal-lined structures, such as the nasopharynx, common bile duct, bladder, and vagina.

Alveolar rhabdomyosarcoma is traversed by a network of fibrous septae that divide the cells into clusters or aggregates, creating a crude resemblance to pulmonary alveoli. Cells in the center of the aggregates are only minimally cohesive, while those at the periphery adhere to the septae. Tumor cells are uniform and round with little cytoplasm. Cross-striations are uncommon (Fig. 26.53B).

Pleomorphic rhabdomyosarcoma is composed of large, sometimes multinucleated, bizarre eosinophilic tumor cells that can overlap histologically with other pleomorphic sarcomas. Immunohistochemistry (e.g., myogenin staining) is usually necessary to confirm rhabdomyoblastic differentiation.

Spindle cell/sclerosing rhabdomyosarcomas consist of fusiform cells with vesicular chromatin arranged in long fascicles or in a storiform pattern. Rhabdomyoblasts are occasionally present. Dense, collagenous, sclerotic stroma may be more common in adults.

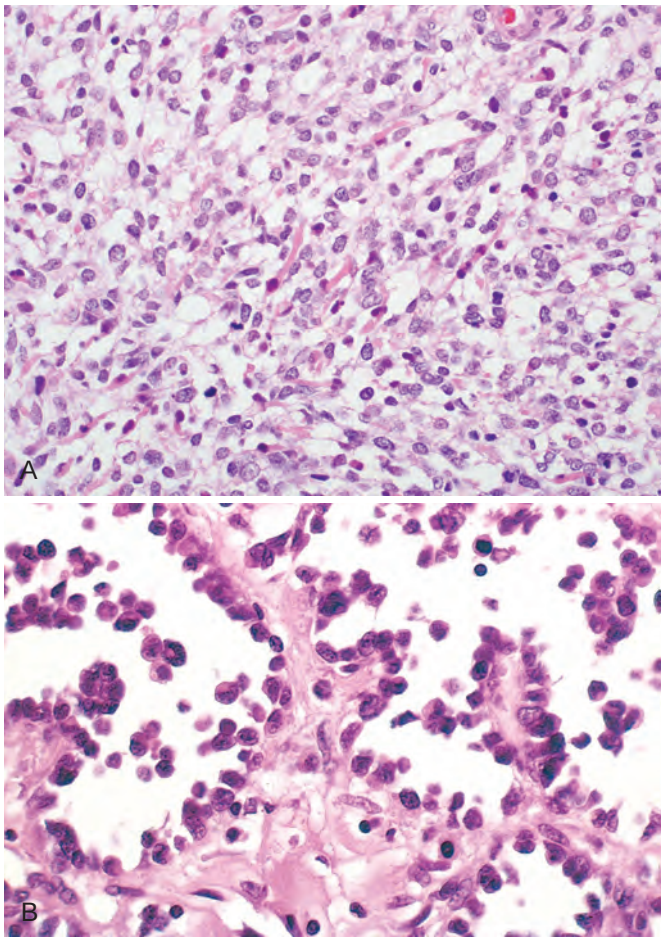


Figure 26.53 Rhabdomyosarcoma. (A) Embryonal subtype composed of malignant cells ranging from primitive and round to densely eosinophilic with skeletal muscle differentiation. (B) Alveolar rhabdomyosarcoma with numerous spaces lined by discohesive, uniform round tumor cells.

Clinical Features

Rhabdomyosarcomas are aggressive and usually treated by surgical resection and chemotherapy, with or without radiation therapy. Histologic type and location are correlated with survival; sarcoma botryoides has the most favorable prognosis, while pleomorphic rhabdomyosarcoma is often fatal.

SMOOTH MUSCLE TUMORS

Leiomyoma

Leiomyomas, benign smooth muscle tumors, often arise in the uterus and are the most common neoplasm in women (Chapter 22). Depending on their number, size, and location, uterine leiomyomas may cause a variety of symptoms including infertility and menorrhagia. *Pilar leiomyomas* arise from cutaneous erector pili muscles and, rarely, develop in deep soft tissues or the gastrointestinal tract. A germline loss-of-function mutation in the fumarate hydratase gene on chromosome 1q42.3 leads to multiple cutaneous leiomyomas, uterine leiomyomas, and renal cell carcinoma. Fumarase hydratase is a Krebs cycle enzyme, and this association is another example of the link between metabolic abnormalities and neoplasia.

Soft tissue leiomyomas are usually 1 to 2 cm and are composed of fascicles of densely eosinophilic spindle cells that often intersect each other at right angles. Tumor cells have blunt-ended, elongated nuclei with minimal atypia and few mitotic figures. Solitary lesions are easily cured. However, multiple tumors may be so numerous that surgical removal is impractical.

Leiomyosarcoma

Leiomyosarcoma, a malignant tumor of smooth muscle, most often develops in the deep soft tissues of the extremities and the retroperitoneum. It accounts for 10% to 20% of soft tissue sarcomas. They occur in adults and afflict women more frequently than men. A particularly deadly form arises from the great vessels, often the inferior vena cava. Leiomyosarcomas have underlying defects in genomic stability leading to complex genotypes.

MORPHOLOGY

Leiomyosarcomas present as painless firm masses. Retroperitoneal tumors may be large and bulky and cause abdominal symptoms. Histologically, they consist of eosinophilic spindle cells with blunt-ended, hyperchromatic nuclei arranged in interweaving fascicles. Immunohistochemical detection of smooth muscle proteins, including smooth muscle actin, desmin, and caldesmon, can aid in diagnosis.

Clinical Features

Leiomyosarcoma treatment depends on tumor size, location, and grade. Superficial or cutaneous leiomyosarcomas are usually small and have a good prognosis, while those involving retroperitoneum are large, frequently unresectable, and often fatal as a result of local extension and metastatic spread, especially to the lungs.

TUMORS OF UNCERTAIN ORIGIN

Although many soft tissue tumors can be assigned to recognizable histologic types, a large proportion do not recapitulate known mesenchymal lineages. This group includes tumors with simple or complex karyotypes; an example of each is described here.

Synovial Sarcoma

The first described cases of synovial sarcoma arose in soft tissues near the knee joint and led to speculation of a relationship with synovium. The proved to be mistaken, as **synovial sarcoma can present in locations that lack synovium, and its morphologic features are inconsistent with an origin from synoviocytes.** Synovial sarcomas account for approximately 10% of all soft tissue sarcomas and typically occur in people in their 20s to 40s. Individuals often present with a deep-seated mass that has been present for several years. A characteristic chromosomal translocation t(x;18)(p11;q11) that produces fusion of the *SS18* gene to one of three *SSX* genes is present in most synovial sarcomas. The fusions encode chimeric proteins that interfere with normal chromatin remodeling, the repositioning of nucleosomes on chromatin that influences gene expression.

MORPHOLOGY

Synovial sarcomas can be monophasic or biphasic. Monophasic synovial sarcomas consist of uniform spindle cells with scant cytoplasm and dense chromatin arranged in short, tightly packed fascicles. The biphasic type also contains glandlike structures composed of cuboidal to columnar epithelioid cells (Fig. 26.54). Immunohistochemistry is helpful because tumor cells, especially in the biphasic type, are positive for epithelial antigens (e.g., keratins), thereby differentiating them from most other sarcomas.

Clinical Features

Synovial sarcomas are treated aggressively with limb-sparing surgery and chemotherapy. The 5-year survival varies from 25% to 62% and is related to tumor stage and patient age. Common sites of metastases are the lung and, unusually for sarcomas, regional lymph nodes.

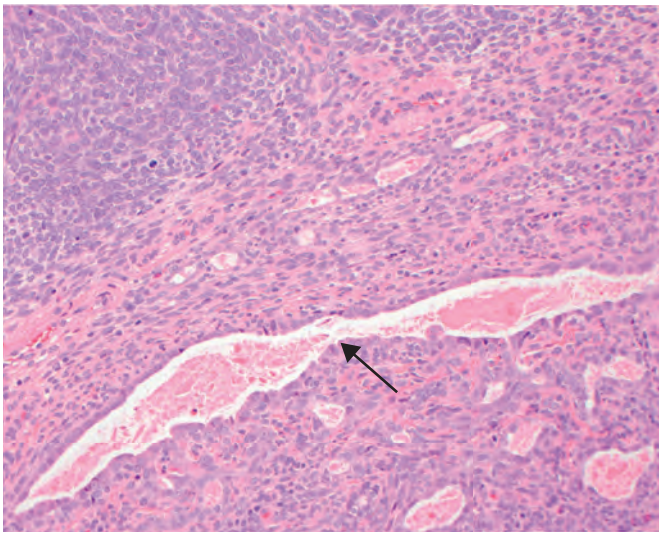


Figure 26.54 Synovial sarcoma revealing the classic biphasic spindle cell and glandlike *(arrow)* histologic appearance.

Undifferentiated Pleomorphic Sarcoma

Undifferentiated pleomorphic sarcomas include malignant mesenchymal tumors with high-grade, pleomorphic cells that cannot be classified into another category by histomorphology, immunophenotype, and genetics. Most arise in the deep soft tissues of the extremities, especially the thigh of middle-aged or older adults. Most undifferentiated pleomorphic sarcomas are aneuploid with multiple structural and numerical chromosomal changes.

MORPHOLOGY

Undifferentiated pleomorphic sarcomas are usually large, gray-white fleshy masses and can grow to 20 cm, depending on the anatomic compartment. Necrosis and hemorrhage common. Microscopically, they are extremely pleomorphic and composed of sheets of large, anaplastic spindled to polygonal cells with hyperchromatic irregular, sometimes bizarre nuclei (Fig. 26.55). Mitotic figures, including atypical asymmetric forms, are abundant. By definition, tumor cells lack both differentiation along recognized lineages and characteristic genetic defects.

Clinical Features

Undifferentiated pleomorphic sarcomas are aggressive malignancies that are treated with surgery and adjuvant chemotherapy and/or radiation. Despite this, prognosis is generally poor. Metastases occur in 30% to 50% of cases.

KEY CONCEPTS

SOFT TISSUE TUMORS

- Soft tissue tumors are malignant mesenchymal lesions that are distinct from tumors of epithelial, skeletal, central nervous system, hematopoietic, or lymphoid tissues.
- Soft tissue tumors likely arise from pluripotent mesenchymal stem cells rather than mature cells.
- Tumors that recapitulate a mature mesenchymal tissue (e.g., smooth muscle) can be subdivided into benign and malignant forms.
- Some tumors are composed of cells for which there is no normal counterpart (e.g., synovial sarcoma and undifferentiated pleomorphic sarcoma).
- Sarcomas with simple karyotypes demonstrate reproducible, chromosomal, and molecular abnormalities that contribute to pathogenesis and have diagnostic utility.
- Most adult sarcomas are genetically heterogeneous with complex karyotypes, histologically pleomorphic, and associated with a poor prognosis.

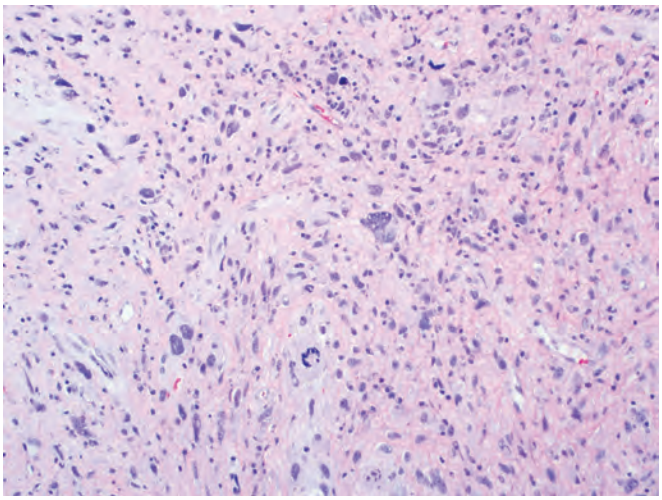


Figure 26.55 Undifferentiated pleomorphic sarcoma revealing anaplastic spindled to polygonal cells.

ACKNOWLEDGMENT

We thank Dr. Andrew Rosenberg for his outstanding contribution to previous editions of this chapter.

SUGGESTED READINGS

Basic Structure and Biology of Bone

Florencio-Silva R, Sasso GR, Sasso-Cerri E et al: Biology of bone tissue: structure, function, and factors that influence bone cells, *Biomed Res Int* 2015:421746, 2015. *[A recent review of bone cell biology].*

Kogianni G, Noble BS: The biology of osteocytes, *Curr Osteoporos Rep* 5:81–86, 2007. *[A good review of the cellular basis of bone remodeling].*

Zaidi M: Skeletal remodeling in health and disease, *Nat Med* 13:791–801, 2007. *[An excellent review of the genetics and pathophysiology of skeletal diseases].*

Skeletal Dysplasias

Krakow D, Rimoin DL: The skeletal dysplasias, *Genet Med* 12:327–341, 2010. *[A comprehensive summary of the dysplasias with a very useful summary table].*

Van Dijk FS, Sillence DO: Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment, *Am J Med Genet A* 164A(6):1470–1481, 2014. *[An overview of the current classification of osteogenesis imperfecta].*

Osteoporosis

Black DM, Rosen CJ: Clinical practice. Postmenopausal osteoporosis, *N Engl J Med* 374(3):254–262, 2016. doi:10.1056/NEJMcp1513724. Review. Erratum in: N Engl J Med.;374(18):1797. *[A recent review of postmenopausal osteoporosis].*

Styrkarsdottir U, Halldorsson BV, Gretarsdottir S et al: Multiple genetic loci for bone mineral density and fractures, *N Engl J Med* 358:2355–2365, 2008. *[A seminal paper elucidating the molecular pathways that underlie osteoporosis].*

Hyperparathyroidism

Silva BC, Bilezikian JP: Parathyroid hormone: anabolic and catabolic actions on the skeleton, *Curr Opin Pharmacol* 22:41–50, 2015. *[A nice summary of the effects of parathyroid hormone on calcium and bone metabolism].*

Paget Disease

Cundy T: Paget's disease of bone, *Metabolism* 80:5–14, 2018. *[A recent update on the pathology, epidemiology, and biology of Paget disease].*

Paul Tuck S, Layfield R, Walker J et al: Adult Paget's disease of bone: a review, *Rheumatology (Oxford)* 56(12):2050–2059, 2017. *[A recent update on the genetic contribution to Paget disease].*

Osteonecrosis

Seamon J, Keller T et al: The pathogenesis of nontraumatic osteonecrosis, *Arthritis* 2012. 601763. *[A summary of recent developments regarding the pathophysiology of femoral head osteonecrosis].*

Osteosarcoma

Gianferante DM, Mirabello L, Savage SA: Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy, *Nat Rev Endocrinol* 13(8):480–491, 2017. *[A concise review of the biology of osteosarcoma]*.

Klein MJ, Siegal GP: Osteosarcoma: anatomic and histologic variants, *Am J Clin Pathol* 125:555–581, 2006. *[A recent update on the classification of osteosarcoma]*.

Chondrogenic Tumors

Bovee JV, Hogendoorn PC, Wunder JS et al: Cartilage tumours and bone development: molecular pathology and possible therapeutic targets, *Nat Rev Cancer* 10:481–488, 2010. *[A good review of the known genetic abnormalities in cartilage tumors]*.

Pansuriya TC, van Eijk R, d'Adamo P et al: Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome, *Nat Genet* 43:1256–1261, 2011. *[A seminal article identifying IDH mutations in inherited cartilage tumors]*.

Wuyts W, Van Hul W: Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes, *Hum Mutat* 15:220–227, 2000. *[A discussion of the genetic basis of multiple osteochondroma syndrome]*.

Ewing Sarcoma

Cidre-Aranaz F, Alonso J: EWS/FLI1 target genes and therapeutic opportunities in Ewing sarcoma, *Front Oncol* 5:162, 2015. *[A review of targets of the EWSR1-FLI1 fusion protein and therapeutic implications]*.

Giant Cell Tumor of Bone

Wu PF, Tang JY, Li KH: RANK pathway in giant cell tumor of bone: pathogenesis and therapeutic aspects, *Tumour Biol* 36(2):495–501, 2015. *[A recent review of RANK-RANKL interactions in the pathogenesis of giant cell tumors]*.

Aneurysmal Bone Cyst

Oliveira AM, Chou MM: The TRE17/USP6 oncogene: a riddle wrapped in a mystery inside an enigma, *Front Biosci (Schol Ed)* 4:321–334, 2012. *[An excellent review of the USP6 clonal genetic rearrangement in various self-limited mesenchymal tumors]*.

Fibrous Dysplasia

Riminucci M, Robey PG, Saggio I et al: Skeletal progenitors and the GNAS gene: fibrous dysplasia of bone read through stem cells, *J Mol Endocrinol* 45:355–364, 2010. *[An excellent discussion of how a mutation can affect skeletal progenitor cells and cause clinical expression of fibrous dysplasia]*.

Osteoarthritis

Goldring MB, Goldring SR: Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis, *Ann N Y Acad Sci* 1192:230–237, 2010. *[A succinct and thoughtful presentation of the role of articular structures in the development of osteoarthritis]*.

Rheumatoid Arthritis and Related Conditions

Deane KD, El-Gabalawy H: Pathogenesis and prevention of rheumatic disease: focus on preclinical RA and SLE, *Nat Rev Rheumatol* 10:212–228, 2014. *[A recent, nicely illustrated update on the pathogenesis of RA and SLE]*.

Scott DL, Wolfe F, Huizinga TW: Rheumatoid arthritis, *Lancet* 376:1094–1108, 2011. *[A review of the pathogenesis and treatment of rheumatoid arthritis]*.

Taurog JD, Chhabra A, Colbert RA et al: Ankylosing spondylitis and axial spondyloarthritis, *N Engl J Med* 374:2563–2574, 2016. *[A recent review of the diagnosis, classification, and treatment of spondyloarthritis]*.

van Heemst J, Jansen DT et al: Crossreactivity to vinculin and microbes provides a molecular basis for HLA-based protection against rheumatoid arthritis, *Nat Commun* 6:6681, 2015. *[A description of the relationship among autoimmunity, HLA alleles, and microbes]*.

Lyme Disease

Steere AC, Strle F, Wormser GP et al: Lyme borreliosis, *Nat Rev Dis Primers* 2:16090, 2016. *[A recent, nicely illustrated, comprehensive review of Lyme disease]*.

Gout and Pseudogout

Dalbeth N, Merriman TR, Stamp LK: Gout, *Lancet* 388(10055):2039–2052, 2016. *[A comprehensive review of clinical, pathologic, and pathophysiologic aspects of gout]*.

Tenosynovial Giant Cell Tumor

Moller E, Mandahl N, Mertens F et al: Molecular identification of COL6A3-CSF1 fusion transcripts in tenosynovial giant cell tumors, *Genes Chromosomes Cancer* 47:21–25, 2008. *[A seminal article identifying clonal chromosomal rearrangement in tenosynovial giant cell tumor]*.

Soft Tissue Sarcomas

Antonescu CR, Dal Cin P: Promiscuous genes involved in recurrent chromosomal translocations in soft tissue tumours, *Pathology* 46:105–112, 2014. *[An excellent review of specificity (or lack thereof) of genetic defects in soft tissue tumors]*.

Barretina J, Taylor BB, Bnaerji S et al: Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy, *Nat Genet* 42:715–721, 2010. *[An extensive study of molecular alterations in soft tissue sarcomas as therapeutic targets]*.

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 27

Peripheral Nerves and Skeletal Muscles

Peter Pytel • Douglas C. Anthony

CHAPTER CONTENTS

Neuromuscular diseases are a complex group of disorders that typically present with weakness, muscle pain, or sensory deficits and may be inherited or acquired. They can be grouped according to anatomy, disease course, and pathogenesis. Physicians keep all these characteristics in mind when evaluating a patient with neuromuscular symptoms. This chapter uses an anatomic approach, grouping neuromuscular disorders into those that preferentially affect the peripheral nerves, the neuromuscular junction, or the skeletal muscles. A discussion of neoplasms that arise from peripheral nerves ends the chapter. Conditions that can produce similar clinical symptoms but are caused by disorders of the central nervous system (CNS) are discussed in Chapter 28.

DISEASES OF PERIPHERAL NERVES

The two components of peripheral nerves involved in impulse transmission are axons and myelin sheaths made by Schwann cells. Injury to either of these components may result in a peripheral neuropathy. Before discussing the pathology of these disorders, a brief review of peripheral nerve structure and function is in order. *Somatic motor function* is carried out by the motor unit, which consists of (1) a lower motor neuron located in the anterior horn of the spinal cord or in the brainstem; (2) an axon that travels in a nerve to a target; (3) the neuromuscular junctions; and (4) multiple innervated myofibers (muscle fibers). *Somatic sensory function* depends on (1) the distal nerve endings, which may contain specialized structures that serve to register specific sensory modalities; (2) a distal axon segment that travels as part of a peripheral nerve to the dorsal root ganglion; and (3) a proximal axon segment that synapses on neurons in the spinal cord or brainstem. *Autonomic nerve fibers,* which transmit all visceral motor and sensory functions, outnumber somatic fibers in the peripheral nervous system, but signs and symptoms related to their involvement are generally not prominent features of peripheral neuropathies, with a few important exceptions (e.g., in some cases of diabetic neuropathy, discussed later).

Specific sensations (pain, temperature, touch) and motor signals are each conveyed by axons that can be distinguished, in part, based on their diameter. Axonal diameters are in turn correlated with the thickness of their myelin sheaths and with their conduction speeds:

- *Thin unmyelinated fibers* mediate autonomic functions as well as certain pain and temperature sensations and have

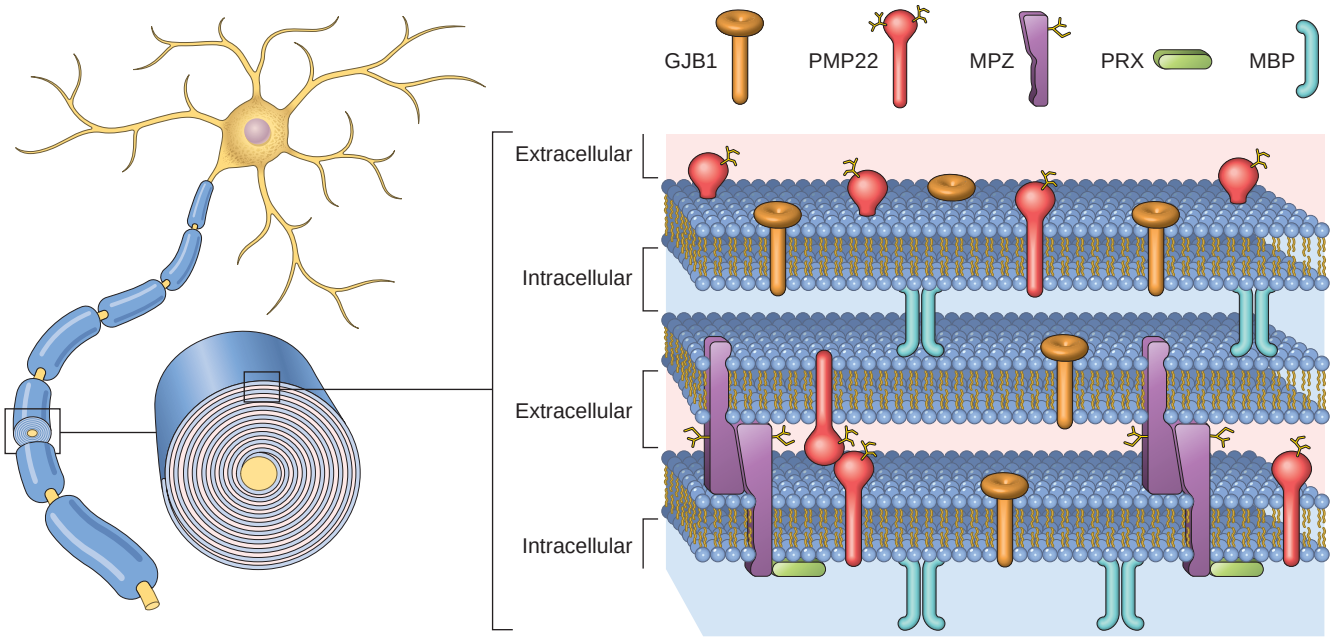


Figure 27.1 Relationship between lipid bilayers and associated proteins in myelin within internodes. Myelin basic protein *(MBP)* is an intracellular protein that has a role in myelin compaction. Mutant forms of myelin protein zero *(MPZ)*, peripheral myelin protein 22 *(PMP22)*, gap junction protein, beta 1 *(GJB1)*, and periaxin *(PRX)* cause some forms of Charcot-Marie-Tooth disease, a hereditary demyelinating neuropathy.

the slowest conduction speeds due to their lack of myelin and small axonal diameter.

- *Large diameter axons* with thick myelin sheaths transmit light touch and motor signals and have fast conduction speeds.

In the case of myelinated axons, individual Schwann cells make exactly one myelin sheath that wraps around a single axon to create a myelinated segment called an *internode*. Internodes are separated by unmyelinated gaps referred to as *nodes of Ranvier,* which are uniformly spaced along the length of the axon. A number of specialized proteins are essential for normal assembly and function of myelin within internodes (Fig. 27.1). Unmyelinated axons are also intimately associated with Schwann cells but in a different arrangement in which one cell surrounds segments of multiple axons.

Most peripheral nerves carry out both motor and sensory functions and thus contain axons of varying diameter and myelin thickness. **The axons are bundled together by three major connective tissue components: the epineurium, which encloses the entire nerve; the perineurium, a multilayered concentric connective tissue sheath that groups subsets of axons into fascicles; and the endoneurium, which surrounds individual nerve fibers.**

General Types of Peripheral Nerve Injury

Axonal Neuropathies

Axons are the primary target of the damage in this large group of peripheral neuropathies (Fig. 27.2). The morphologic hallmarks of axonal neuropathies can be produced experimentally by cutting a peripheral nerve, which results in a prototypical pattern of injury described as *Wallerian degeneration*. Portions of axons that are distal to the point of transection are disconnected from the cell bodies *(perikarya)* and degenerate. Within a day of injury, the distal axons begin to fragment, and the associated myelin sheaths unravel (Fig. 27.3) and disintegrate into spherical structures *(myelin ovoids)*. Macrophages are recruited and participate in the removal of axonal and myelin debris. Regeneration starts at the site of transection with the formation of a growth cone and the outgrowth of new branches from the stump of the proximal axon. Schwann cells and their associated basement membranes guide the sprouting axons, which grow at about 1 mm per day, toward their distal target. Continuous pruning of the sprouting axons removes misguided branches. The Schwann cells create new myelin sheaths around the regenerating axons, but these myelin internodes tend to be thinner and shorter than in the original ones. The regeneration is successful only if the two transected ends remain closely approximated. A failure of the outgrowing axons to find their distal target can produce a "pseudotumor" termed *traumatic neuroma*—a nonneoplastic haphazard proliferation of axonal processes and associated Schwann cells that results in a painful nodule (Fig. 27.4).

The changes observed following experimental nerve transections only partially resemble those seen in various axonal neuropathies. One key difference is that in these neuropathies (unlike nerve transection) damage occurs over an extended period of time. As a result, degenerating and regenerating axons coexist in a single biopsy. With time, damage tends to outpace repair, resulting in progressive loss of axons. In cases of toxic and metabolic insults, axons often degenerate in a length-dependent fashion with the longest axons being most susceptible, resulting in a *"dying-back"* type of pattern of progression. The electrophysiologic hallmark of axonal neuropathies is a reduction in signal amplitude owing to the dropout of axons from affected peripheral nerves, with relative preservation of conduction velocity.

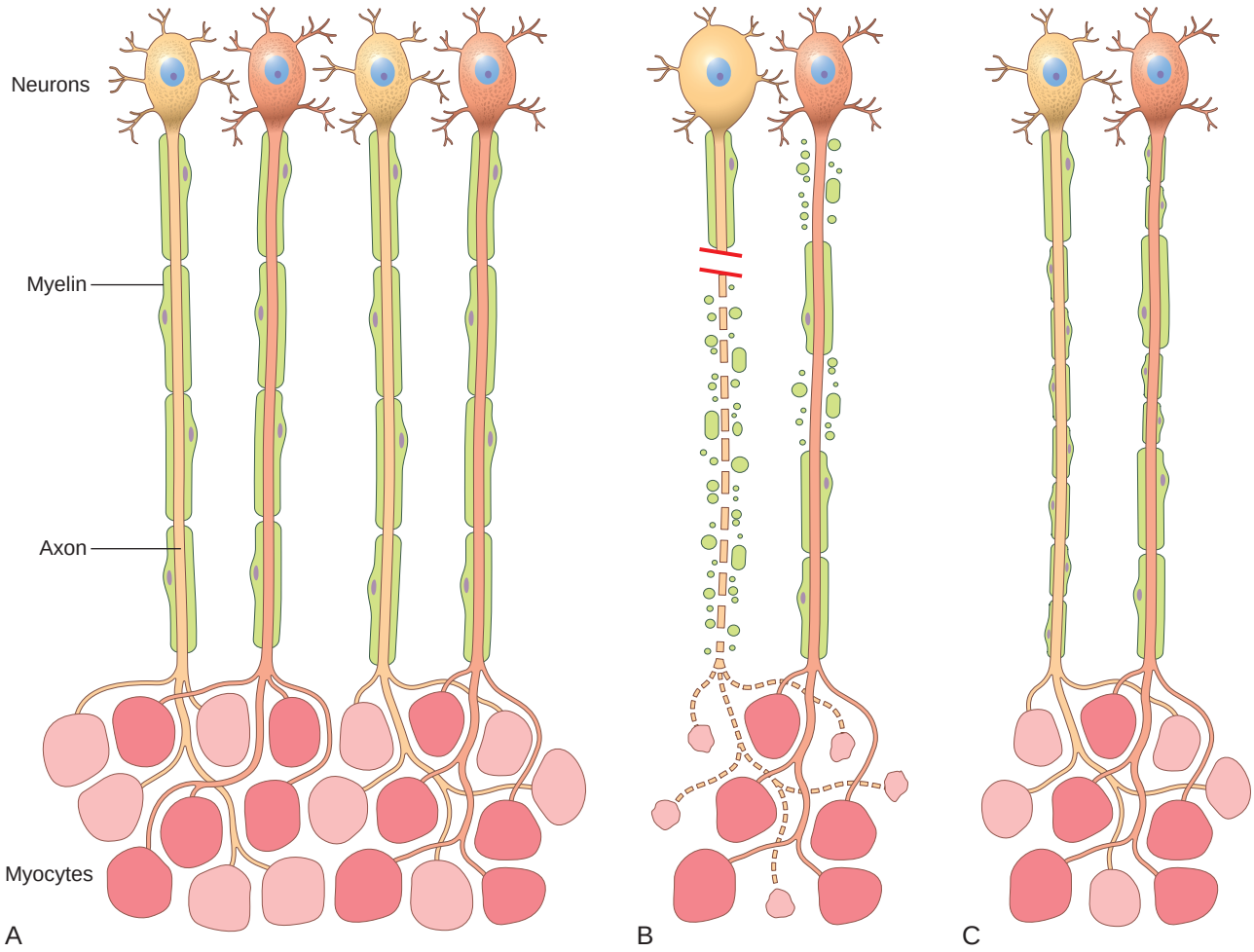


Figure 27.2 Patterns of peripheral nerve damage. (A) In normal motor units, type I and type II myofibers are arranged in a "checkerboard" distribution, and the internodes along the motor axons are uniform in thickness and length. (B) Acute axonal injury *(left axon)* results in degeneration of the distal axon and its associated myelin sheath, with atrophy of denervated myofibers. In contrast, acute demyelinating disease *(right axon)* produces random segmental degeneration of individual myelin internodes, while sparing the axons. (C) Regeneration of axons after injury *(left axon)* allows reinnervation of myofibers. The regenerated axon is myelinated by proliferating Schwann cells, but the new internodes are shorter and the myelin sheaths are thinner than the original ones. Remission of demyelinating disease *(right axon)* allows remyelination to take place, but the new internodes are shorter and have thinner myelin sheaths than flanking normal undamaged internodes. See Fig. 27.7 for comparison with reinnervation.

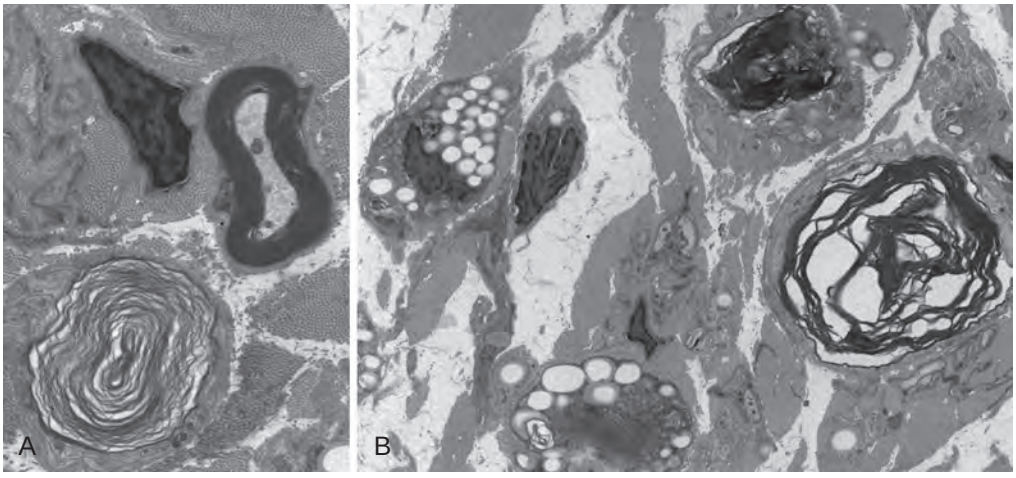


Figure 27.3 Electron micrographs illustrating features of axonal degeneration. (A) Degenerating myelin with loosened myelin layers is seen in the degenerating axon in the lower left corner, to be contrasted with a normal myelin sheath with tightly packed myelin and intact axon in the upper right corner. (B) In addition to an unraveling myelin sheath, several cells contain lipid droplets (seen as vacuoles) derived from degenerating myelin.

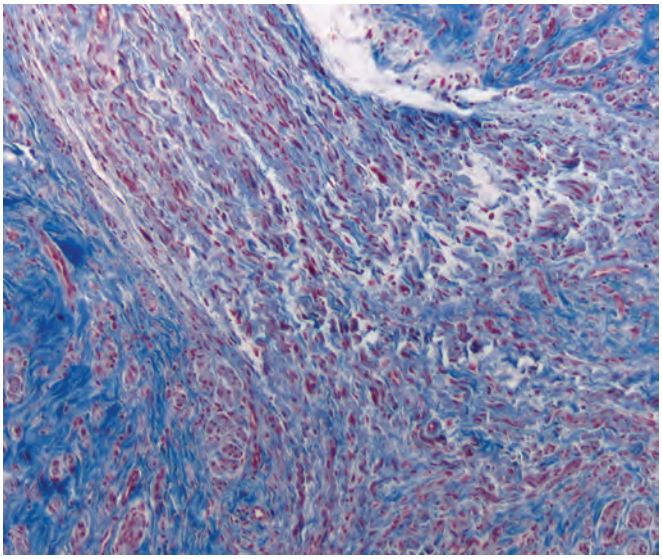


Figure 27.4 Trichrome-stained section of a traumatic neuroma showing the transition from normal nerve containing a parallel arrangement of axons *(upper left corner)* to a haphazard swirl of red-stained axons associated with admixture of Schwann cells and blue-staining connective tissue.

Demyelinating Neuropathies

In these disorders, Schwann cells with their myelin sheaths are the primary targets of damage (see Fig. 27.2), whereas axons are relatively preserved. This definition is similar to that of demyelinating diseases that affect the CNS (Chapter 28). Individual myelin sheaths degenerate in a seemingly random pattern, resulting in discontinuous damage of myelin segments. In response to this damage, Schwann cells or Schwann cell precursors proliferate and initiate repair through the formation of new myelin sheaths, but these again tend to be shorter and thinner than the original ones. The electrophysiologic hallmark of demyelinating neuropathies is slowed nerve conduction velocity, reflective of the loss of myelin.

Neuronopathies

Neuronopathies result from destruction of neurons, leading to secondary degeneration of axonal processes. Infections like herpes zoster and toxins like platinum compounds are examples of insults that may lead to neuronopathies. Because the damage is at the level of the neuronal cell body, peripheral nerve dysfunction caused by neuronopathies is equally likely to affect proximal and distal parts of the body (unlike peripheral axonopathies, which preferentially affect the distal extremities).

Anatomic Patterns of Peripheral Neuropathies

Peripheral neuropathies can be separated into several groups according to the anatomic distribution of involvement and the associated neurologic deficits. This approach can be helpful clinically, since each pattern has a different set of potential underlying causes. These anatomic patterns of injury are as follows:

- *Mononeuropathies* affect a single nerve and result in deficits in a restricted distribution dictated by nerve anatomy. Trauma, entrapment, and infections are common causes of mononeuropathy.
- *Polyneuropathies* are characterized by involvement of multiple nerves, usually in a symmetric fashion. In most cases axons are affected in a length-dependent fashion leading to deficits that start in the feet and ascend with disease progression. The hands often start to show involvement by the time deficits extend to the level of the knee, resulting in a characteristic "stocking and glove" distribution of sensory deficits.
- *Mononeuritis multiplex* describes a disease process that damages individual nerves in a haphazard fashion. An affected patient might have a right wrist drop from involvement of the right radial nerve and a left foot drop from peroneal nerve damage. Vasculitis is a common cause of this pattern of injury.
- *Polyradiculoneuropathies* affect nerve roots as well as peripheral nerves, leading to diffuse symmetric symptoms in proximal and distal parts of the body.

Specific Peripheral Neuropathies

Patients with peripheral neuropathy often complain of numbness, painful "pins and needles" sensations, and weakness, most often in the distal portions of the extremities. Many different types of disease processes can damage peripheral nerves, including inflammatory diseases, infections, metabolic changes, toxic injury, trauma, paraneoplastic disease, and inherited gene defects.

Inflammatory Neuropathies

Guillain-Barré Syndrome (Acute Inflammatory Demyelinating Polyneuropathy)

Guillain-Barré syndrome is an immunologically mediated demyelinating peripheral neuropathy that may lead to life-threatening respiratory paralysis. The overall annual incidence is approximately 1 case per 100,000 persons. The disease is characterized clinically by weakness beginning in the distal limbs that rapidly advances to affect proximal muscle function ("ascending paralysis"). Histologic features are inflammation and demyelination of spinal nerve roots and peripheral nerves (radiculoneuropathy).

Pathogenesis

In most cases, Guillain-Barré syndrome is thought to be an acute-onset immune-mediated demyelinating neuropathy. Approximately two-thirds of cases are preceded by an acute, influenza-like illness from which the affected individual has recovered by the time the neuropathy becomes symptomatic. Infections with *Campylobacter jejuni,* cytomegalovirus, Epstein-Barr virus, and *Mycoplasma pneumoniae* as well as prior vaccination have significant epidemiologic associations with Guillain-Barré syndrome. No infectious agent has been demonstrated in affected nerves, and an immunologic reaction is favored as the underlying cause. A similar inflammatory disease of peripheral nerves can be reproduced in experimental animals by immunization with a peripheral nerve myelin protein. A T-cell–mediated immune response ensues, accompanied by segmental demyelination induced by the actions of activated macrophages. Transfer of these T cells to a naive animal results in comparable lesions.

Moreover, lymphocytes from individuals with Guillain-Barré syndrome have been shown to produce demyelination in tissue cultures of myelinated nerve fibers. Circulating antibodies that cross-react with components of peripheral nerves may also play a role.

MORPHOLOGY

The dominant finding in sections stained with hematoxylin and eosin is **inflammation of peripheral nerves,** manifested as perivenular and endoneurial infiltration by lymphocytes, macrophages, and a few plasma cells. Segmental demyelination affecting peripheral nerves is the most prominent lesion, but damage to axons is also seen, particularly when the disease is severe. Electron microscopy has identified an early effect on myelin sheaths. The cytoplasmic processes of macrophages penetrate the basement membrane of Schwann cells, particularly in the vicinity of the nodes of Ranvier, and extend between the myelin lamellae, stripping the myelin sheath from the axon. Ultimately, the remnants of the myelin sheath are engulfed by the macrophages. Inflammation and demyelination can be widespread in the peripheral nervous system but are typically most prominent proximally, close to the nerve roots.

Clinical Features

The clinical picture is dominated by ascending paralysis and areflexia. Deep tendon reflexes disappear early in the process. Sensory involvement, including loss of pain sensation, is often present but is usually not a prominent feature. Nerve conduction velocities are slowed because of multifocal destruction of myelin segments in many axons within a nerve. Cerebrospinal fluid (CSF) protein levels are elevated due to inflammation and altered permeability of the microcirculation within the spinal roots as they traverse the subarachnoid space. Inflammatory cells, on the other hand, remain confined to the roots; therefore, there is little or no CSF pleocytosis. Many patients spend weeks in the intensive care unit (ICU) before recovering normal function. With improved supportive respiratory care, cardiovascular monitoring, and prophylaxis against deep venous thrombosis, the mortality rate has fallen. Plasmapheresis and intravenous immunoglobulin therapy hasten recovery, apparently because these remove pathogenic antibodies and suppress immune function, respectively. However, 2% to 5% of affected patients die of respiratory paralysis, autonomic instability, cardiac arrest, or related complications, and up to 20% of hospitalized survivors suffer long-term disability.

Chronic Inflammatory Demyelinating Poly(radiculo)neuropathy

This is the most common chronic acquired inflammatory peripheral neuropathy, characterized by symmetric mixed sensorimotor polyneuropathy that persists for 2 months or more. By definition, signs and symptoms must be present for at least 2 months, but often the disease evolves over years, usually with relapses and remissions. Typically there is a symmetric, mixed sensorimotor polyneuropathy, but some patients may present with predominantly sensory or motor impairment. Clinical remissions can often be achieved with intravenous immunoglobulin, or with other immunosuppressive therapies, such as plasmapheresis, glucocorticoids, and cytotoxic agents directed against T cells or B cells. The time course and the response to steroids distinguish chronic inflammatory demyelinating polyradiculoneuropathy from Guillain-Barré syndrome.

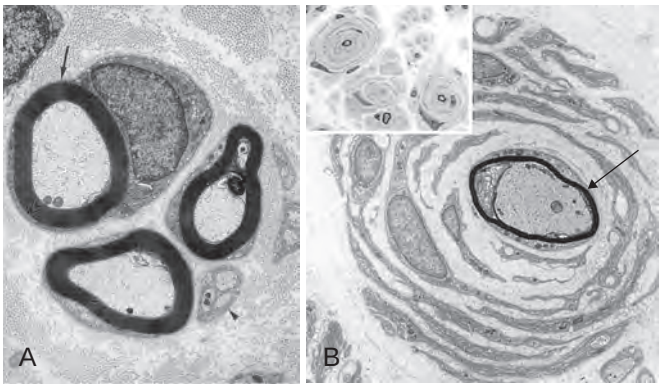


Figure 27.5 Onion bulb neuropathy. (A) shows normal myelinated axons (one that includes the nucleus of the associated Schwann cells is marked with an arrow) and unmyelinated axons *(arrowhead)*. Compared with the normal ultrastructure of axons in a nerve (A), an "onion bulb" (B) is composed of a thinly myelinated axon *(arrow)* surrounded by multiple concentrically arranged Schwann cells. *Inset,* Light microscopic appearance of an onion bulb neuropathy, characterized by "onion bulbs" surrounding axons. (B, Courtesy G. Richard Dickersin, MD, from *Diagnostic Electron Microscopy: A Text Atlas,* New York, 2000, Igaku-Shoin Medical Publishers, p 984.)

Pathogenesis

T cells as well as antibodies are implicated in the inflammatory process. Molecules expressed at the Schwann cell–axon junction and in noncompact areas of myelin appear to be the target of the immune response. Complement-fixing immunoglobulin G (IgG) and IgM can be found on the myelin sheaths, and the deposition of these opsonins leads to recruitment of macrophages that strip myelin from axons. Sural nerve biopsies show evidence of recurrent demyelination and remyelination associated with proliferation of Schwann cells. When excessive, this proliferation leads to the formation of so-called *onion bulbs*—structures in which multiple layers of Schwann cells wrap around an axon like the layers of an onion (Fig. 27.5).

Neuropathy Associated With Systemic Autoimmune Diseases

Systemic autoimmune diseases like rheumatoid arthritis, Sjögren syndrome, or systemic lupus erythematosus may be associated with peripheral neuropathies that often take the form of distal sensory or sensorimotor polyneuropathies. These neuropathies are distinct from vasculitic peripheral neuropathies, which can arise as secondary manifestations of these same diseases.

Neuropathy Associated With Vasculitis

Vasculitis is a noninfectious inflammation of blood vessels that can involve and damage peripheral nerves. About one-third of patients with vasculitis, depending on the type of vasculitis, have peripheral nerve involvement, and neuropathy may be the presenting feature. Vasculitis often presents as mononeuritis multiplex, but mononeuritis and polyneuropathy are also encountered.

When peripheral neuropathy occurs in systemic vasculitis, it is most common in the MPO-ANCA–associated vasculitides (microscopic polyangiitis and eosinophilic granulomatosis with polyangiitis/Churg-Strauss syndrome) and polyarteritis nodosa and much less common in PR3-ANCA–associated granulomatosis with polyangiitis. The most common form of vasculitis associated with peripheral neuropathy is a localized form referred to as *nonsystemic vasculitic neuropathy,* which is not associated with any anti-neutrophil cytoplasmic antibodies (ANCAs). Regardless of the type of vasculitis, peripheral nerves involved by vasculitis typically show patchy axonal degeneration and loss, with some fascicles being more severely affected than others. Perivascular inflammatory infiltrates are often present. Identification of blood vessels with characteristic forms of acute or chronic damage (Chapter 11) helps establish the diagnosis.

Infectious Neuropathies

Many infectious processes affect peripheral nerves. Among these, leprosy, diphtheria, and varicella-zoster cause relatively specific pathologic changes in nerves that are the focus here. Each of these disorders is also discussed in more detail in Chapter 8.

Leprosy (Hansen Disease)

Peripheral nerves are involved in both lepromatous and tuberculoid leprosy (discussed in Chapter 8).

- In *lepromatous leprosy,* Schwann cells are invaded by *Mycobacterium leprae,* which proliferate and eventually infect other cells. There is evidence of segmental demyelination and remyelination and loss of both myelinated and unmyelinated axons. As the infection advances, endoneurial fibrosis and multilayered thickening of the perineurial sheaths occur. Affected individuals develop a symmetric polyneuropathy that is most severe in the relatively cool distal extremities and in the face because lower temperatures favor mycobacterial growth. The infection prominently involves pain fibers, and the resulting loss of sensation contributes to injury, since the patient is rendered unaware of injurious stimuli and damaged tissues. Thus, large traumatic ulcers may develop.
- *Tuberculoid leprosy* is characterized by an active cell-mediated immune response to *M. leprae* that usually manifests as dermal nodules containing granulomatous inflammation. The inflammation injures cutaneous nerves in the vicinity; axons, Schwann cells, and myelin are lost; and there is fibrosis of the perineurium and endoneurium. In tuberculoid leprosy, affected individuals have much more localized nerve involvement.

Lyme Disease

Lyme disease causes various neurologic manifestations in the second and third stages of the disease. These include polyradiculoneuropathy and unilateral or bilateral facial nerve palsies.

HIV/AIDS

Patients infected with human immunodeficiency virus (HIV) develop several patterns of peripheral neuropathy that are poorly understood, but all appear to be related in some way to immune dysregulation. Early-stage HIV infection can be associated with mononeuritis multiplex and demyelinating disorders that may resemble Guillain-Barré syndrome or chronic inflammatory demyelinating polyradiculoneuropathy. More commonly, later stages of HIV infection are associated with a distal sensory neuropathy that is often painful.

Diphtheria

Peripheral nerve dysfunction results from the effects of the diphtheria exotoxin. It produces an acute peripheral neuropathy associated with prominent bulbar and respiratory muscle dysfunction, which can lead to death or long-term disability. The mechanism of action of diphtheria toxin is described in Chapter 8. Diphtheria is most commonly found in low income countries and is a continuing medical problem because of incomplete immunization or waning immunity in adults.

Varicella-Zoster Virus

Varicella-zoster is one of the most common viral infections of the peripheral nervous system. Following chickenpox, a latent infection persists within neurons of sensory ganglia. If the virus is reactivated, sometimes many years later, it may be transported along the sensory nerves to the skin. Here it infects keratinocytes, leading to a **painful, vesicular skin eruption in a distribution that follows sensory dermatomes (shingles).** Most common is the involvement of thoracic or trigeminal nerve dermatomes. The factors underlying reactivation of the virus are not fully understood, but decreased cell-mediated immunity is suspected to play a role. In a small proportion of patients, weakness is also apparent in the same distributions. Affected ganglia show neuronal death, usually accompanied by abundant mononuclear inflammatory cell infiltrates; focal necrosis and hemorrhage may also be found. Peripheral nerves show degeneration of the axons that belong to the dead sensory neurons. Focal destruction of the large motor neurons of the anterior horns or cranial nerve motor nuclei may be seen at the corresponding levels. Intranuclear inclusions generally are not found in the peripheral nervous system.

Metabolic, Hormonal, and Nutritional Neuropathies

Diabetes

Diabetes is the most common cause of peripheral neuropathy. The prevalence of this complication depends on the duration of the disease; up to 50% of patients with diabetes overall and up to 80% of those who have had the disease for more than 15 years have clinical evidence of peripheral neuropathy. Patients with both type 1 and type 2 diabetes are affected (Chapter 24). Several distinct clinicopathologic patterns of diabetes-related peripheral neuropathy are recognized (described later), but the most common by far is an ascending distal symmetric sensorimotor polyneuropathy.

Pathogenesis

The mechanism of diabetic neuropathy is complex and not completely resolved; both metabolic and secondary vascular changes are believed to contribute to the damage of axons and Schwann cells. Hyperglycemia causes the nonenzymatic glycosylation of proteins, lipids, and nucleic acids. The resulting advanced glycosylation end-products (AGEs)

may interfere with normal protein function and activate inflammatory signaling through the receptor for the AGE. Excess glucose within cells is reduced to sorbitol, a process that depletes NADPH and increases intracellular osmolality. These and other metabolic disturbances may predispose peripheral nerves to injury by reactive oxygen species. In addition, the vascular injuries that occur in chronic diabetes due to hyperlipidemia and other metabolic alterations may cause ischemic damage of the nerves.

MORPHOLOGY

In individuals with a distal symmetric sensorimotor neuropathy, the predominant pathologic finding is an axonal neuropathy. Nerve biopsies show reduced numbers of axons. Variable degrees of ongoing axonal damage, marked by degenerating myelin sheaths and regenerative axonal clusters, may be present. Endoneurial arterioles show thickening, hyalinization, and intense periodic acid–Schiff positivity of their walls and extensive reduplication of basement membranes (Fig. 27.6).

Clinical Features

Distal symmetric diabetic polyneuropathy typically presents with sensory symptoms like numbness, loss of pain sensation, difficulty with balance, and paresthesias or dysesthesias. Paresthesias or dysesthesias (abnormal painful sense of touch) are so-called "positive" symptoms—painful sensations that result from abnormal discharges of damaged nerves. Neuropathy leads to considerable morbidity, in particular an increased susceptibility to foot and ankle fractures and chronic skin ulcers, which may eventually lead to amputations.

Another manifestation is dysfunction of the autonomic nervous system; this affects 20% to 40% of individuals with diabetes mellitus, nearly always in association with a distal sensorimotor neuropathy. Diabetic autonomic neuropathy has protean manifestations, including postural hypotension, incomplete emptying of the bladder (resulting in recurrent infections), and sexual dysfunction. Some affected individuals, especially older adults with a long history of diabetes, develop a peripheral neuropathy that manifests with asymmetric presentations, including *mononeuropathy, cranial neuropathy, and radiculoplexus neuropathy*. The last-mentioned is a devastatingly painful acute disorder that presents in the distribution of the brachial or lumbosacral nerve plexus. It is often monophasic and can improve over several months. These asymmetric manifestations may be caused by microvascular disease.

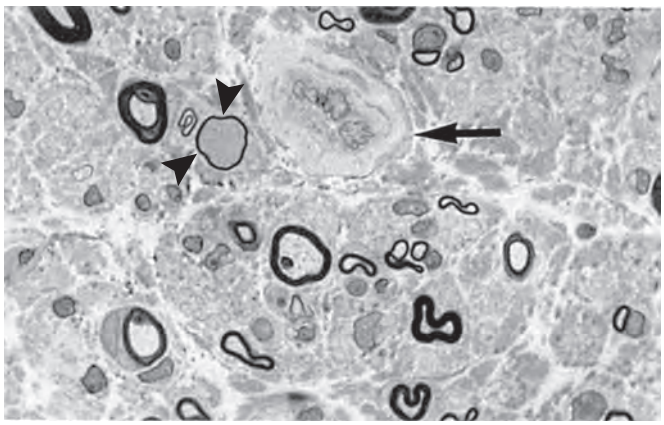


Figure 27.6 Diabetic neuropathy with marked loss of myelinated fibers, a thinly myelinated fiber *(arrowheads)*, and thickening of endoneurial vessel wall *(arrow)*.

Other Metabolic, Hormonal, and Nutritional Neuropathies

A diverse group of metabolic, hormonal, and nutritional disorders are associated with peripheral neuropathy, including the following:

- *Uremic neuropathy.* Most individuals with renal failure have a peripheral neuropathy. Typically this is a distal, symmetric neuropathy that may be asymptomatic or may be associated with muscle cramps, distal dysesthesias, and diminished deep tendon reflexes. In these patients, axonal degeneration is the primary event; occasionally there is secondary demyelination. Regeneration and recovery are common after dialysis.
- *Thyroid dysfunction.* Hypothyroidism can lead to compression mononeuropathies such as carpal tunnel syndrome or cause a distal symmetric predominantly sensory polyneuropathy. In rare cases, hyperthyroidism is associated with a neuropathy resembling Guillain-Barré syndrome.
- *Vitamin B_{12} (cyanocobalamin) deficiency* classically results in subacute combined degeneration with damage to long tracts in the spinal cord (Chapter 28) and also peripheral nerves.
- *Deficiencies of vitamin B_1 (thiamine), vitamin B_6 (pyridoxine), folate, copper, and zinc* all have been associated with peripheral neuropathy.

Toxic Neuropathies

Peripheral neuropathies may appear after exposure to industrial or environmental chemicals, biologic toxins, or therapeutic drugs. Important causes of toxic peripheral nerve damage include alcohol (independent of associated nutritional deficiencies), heavy metals (lead, mercury, arsenic, and thallium), and organic solvents. Various medications can cause toxic nerve damage, but the most notorious are chemotherapeutic agents. These include vinca alkaloids and taxanes, microtubule inhibitors that interfere with axonal transport, and cisplatin, which may cause a neuronopathy.

Neuropathies Associated With Malignancy

Neuropathies associated with cancers may stem from local effects, complications of therapy, paraneoplastic effects, or (in the case of B-cell tumors) tumor-derived immunoglobulins.

- *Direct infiltration or compression of peripheral nerves* by tumors is a common cause of mononeuropathy and may be a presenting symptom of cancer. These neuropathies include brachial plexopathy from neoplasms of the apex of the lung, obturator palsy from pelvic malignant neoplasms, and cranial nerve palsies from intracranial tumors or tumors of the base of the skull. A polyradiculopathy involving the lower extremity may develop when the cauda equina is involved by meningeal carcinomatosis.
- *Paraneoplastic neuropathies*. These can occur at any time during the patient's course, but often precede the diagnosis of the underlying tumor. Sensorimotor neuropathy is the

most common paraneoplastic form, but a pure sensory neuronopathy, chronic inflammatory demyelinating polyneuropathy, plexopathy, and autonomic neuropathy may also be seen. Paraneoplastic sensory neuronopathy is most commonly associated with small cell lung cancer. Antibodies that recognize proteins expressed by cancer cells and normal neurons (for example, anti-Hu antibodies) are often present, but the damage appears to be mediated by a CD8+ cytotoxic T-cell attack on dorsal root ganglion cells. Sensory symptoms usually start distally in an asymmetric and multifocal pattern. Other patients with anti-CV2 autoantibodies (which recognize CRMP5, an intracellular signaling protein) tend to present with a painful axonal asymmetric sensorimotor neuropathy.

- *Neuropathies associated with monoclonal gammopathies*. Neoplastic B cells may secrete monoclonal immunoglobulins or immunoglobulin fragments (so-called paraproteins) that damage nerves. For example, tumors that secrete IgM immunoglobulin may be associated with a demyelinating peripheral neuropathy. In most cases, the pathogenic IgM paraprotein is thought to bind directly to myelin-associated antigens such as myelin-associated glycoprotein. Deposition of IgM can be seen ultrastructurally between the membrane layers of the myelin sheath. IgG or IgA paraproteins may also be associated with peripheral neuropathy. One distinctive presentation is *POEMS syndrome* (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes), in which patients develop a demyelinating neuropathy associated with deposition of paraprotein between noncompacted myelin lamellae. Finally, excess immunoglobulin light chain may deposit as amyloid (Chapter 6), which can lead to peripheral neuropathy due to vascular insufficiency or a direct toxic effect.

Neuropathies Caused by Physical Forces

Peripheral nerves are commonly injured by trauma or entrapment.

- *Lacerations* result from cutting injuries and from sharp fragments of fractured bone, both of which may sever a nerve.
- *Avulsion* of a nerve may occur when tension is applied, often to one of the limbs.
- *Compression neuropathy (entrapment neuropathy)* occurs when a peripheral nerve is chronically subjected to increased pressure, often within an anatomic compartment. Carpal tunnel syndrome, the most common entrapment neuropathy, results from compression of the median nerve at the level of the wrist within the compartment delimited by the transverse carpal ligament. Women are more commonly affected than men, and the problem is frequently bilateral. The disorder may be observed in association with many conditions including tissue edema, pregnancy, inflammatory arthritis, hypothyroidism, amyloidosis (especially related to β_2-microglobulin deposition in individuals on renal dialysis), acromegaly, diabetes mellitus, and excessive repetitive motions of the wrist. Symptoms are limited to dysfunction of the median nerve and typically include numbness and paresthesias of the tips of the thumb and first two digits. Other nerves prone to compression neuropathies include the ulnar nerve at the level of the elbow, the peroneal nerve at the level of the knee, and the radial nerve in the upper arm; the latter occurs from sleeping with the arm in an awkward position ("Saturday night palsy"). Another form of compression neuropathy is found in the foot, affecting the interdigital nerve at intermetatarsal sites. This problem, which occurs more often in women than in men, leads to foot pain (metatarsalgia) and is associated with a histologic lesion called a *Morton neuroma,* which is marked by perineural fibrosis.

Inherited Peripheral Neuropathies

Inherited peripheral neuropathies are a group of genetically diverse disorders with overlapping clinical phenotypes that often present in adults. Because of the delayed onset, the possibility of an inherited neuropathy has to be considered in the differential diagnosis for any patient who presents with a peripheral neuropathy. The major types of inherited peripheral neuropathies include (1) hereditary motor and sensory neuropathies, sometimes known as *Charcot-Marie-Tooth* (CMT) disease, described below, which is a genetically heterogeneous disorder caused by mutations in genes that encode proteins involved in the structure and function of either the peripheral nerve axon or the myelin sheath; (2) hereditary motor neuropathies; (3) hereditary sensory neuropathies, with or without autonomic neuropathy; and (4) other inherited conditions causing neuropathy, including familial amyloidosis and inherited metabolic diseases.

Historically, these diseases were classified based on their inheritance pattern and clinical features. Now there is a continuously growing list of altered genes that are linked to these diseases. The complexity of the genetics of inherited neuropathies is no doubt a reflection of the complicated homeostatic mechanisms that sustain normal peripheral nerve function. There is no simple unifying concept tying all the implicated genes together; some of the implicated genes are illustrated in Fig. 27.1. Below, some of the more common and distinctive types of inherited peripheral neuropathies are described in brief.

Hereditary Motor and Sensory Neuropathies/ Charcot-Marie-Tooth Disease

These are by far the most common inherited peripheral neuropathies, affecting up to 1 in 2500 people. The initial description of these disorders, based on clinical features, was deceptively simple—an inherited disease associated with distal muscle atrophy, sensory loss, and foot deformities. It is now appreciated that this clinical phenotype can result from mutations in more than 50 different genes. Current systems classify hereditary motor and sensory neuropathies based on the mode of inheritance and the pattern of injury (e.g., axonal or demyelinating). Demyelinating forms of CMT disease are associated with morphologic features of demyelination and remyelination including Schwann cell hyperplasia and onion bulb formation, which may be so severe that the involved nerve is palpably enlarged. Listed are the four most common types, which together account for over 90% of patients with CMT disease:

- *CMT1 encompasses a group of autosomal dominant demyelinating neuropathies* that collectively are the most common subtype of hereditary motor and sensory neuropathy. *CMT1A* accounts for some 55% of genetically defined CMT. It is caused by a duplication of a region on

chromosome 17 that includes the peripheral myelin protein 22 *(PMP22)* gene. The disease usually presents in the second decade of life as a slowly progressive distal demyelinating motor and sensory neuropathy. *CMT1B* is also a demyelinating neuropathy caused by mutations in the myelin protein zero gene *(MPZ)* and accounts for about 9% of genetically defined cases of CMT.

- *CMTX* encompasses X-linked forms of CMT disease. *CMT1X* is the most common of these, accounting for 15% of genetically defined cases of CMT. It is linked to mutations in the *GJB1* gene, which encodes connexin32, a gap junction component that is expressed in Schwann cells.
- *CMT2* includes autosomal dominant neuropathies associated with axonal rather than demyelinating injury. *CMT2A* is the most common subtype, accounting for 4% of all CMT disease. It is caused by mutations in the *MFN2* gene, which is required for normal mitochondrial fusion. The phenotype is typically severe, with disease onset in early childhood.

Hereditary Sensory Neuropathies With or Without Autonomic Neuropathy

This is a diverse group of diseases marked by loss of sensation and variable autonomic disturbances, but with preserved motor strength. Loss of pain and temperature sensation is the most common symptom. The inability to sense pain leads to traumatic injury of the hands and feet. These are typically axonal neuropathies.

Familial Amyloid Polyneuropathies

These are hereditary disorders characterized by amyloid deposition within peripheral nerves. Most are caused by germline mutations of the transthyretin gene. The mutated transthyretin protein, which is normally involved in serum binding and transport of thyroid hormone, is prone to deposit as amyloid fibrils in a number of tissues including peripheral nerve (Chapter 6). The clinical presentation is similar to that of hereditary sensory and autonomic neuropathies.

KEY CONCEPTS

PERIPHERAL NEUROPATHIES

- Anatomic patterns include mononeuropathy, mononeuritis multiplex, polyneuropathy, and polyradiculoneuropathy.
- Damage may occur primarily in Schwann cells (demyelinating neuropathy), axons (axonal neuropathy), or central neurons (neuronopathy); mixed patterns of injury occur.
- Inflammatory disease, infections, metabolic changes, toxic injury, trauma, paraneoplastic disorders, and inherited gene defects all can cause peripheral neuropathy.
- Diabetes mellitus is the most common cause of peripheral neuropathy, which most often presents as a distal symmetric neuropathy.
- Guillain-Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy are the major acute and chronic acquired demyelinating peripheral neuropathies, respectively.
- Inherited peripheral neuropathies are genetically and phenotypically diverse disorders that often present in adulthood and may be marked by sensory, motor, or autonomic dysfunction, alone or in combination.

DISEASES OF THE NEUROMUSCULAR JUNCTION

The neuromuscular junction is a complex specialized structure located at the interface of motor nerve axons and skeletal muscle that serves to control muscle contraction. Neuromuscular junctions are found midway along the length of myofibers. Here, the distal ends of peripheral motor nerves branch into small processes that terminate in bulbous synaptic boutons. Upon depolarization, these presynaptic nerve terminals release acetylcholine (ACh) into the synaptic cleft, the space separating the nerve endings from the myofiber membrane (referred to as the *sarcolemma*). The postsynaptic sarcolemma is characterized by complex infoldings and exhibits distinct specializations with localized clustering of ACh receptors. These receptors are responsible for the initiation of signals leading to muscle contraction.

Regardless of cause, patients with disorders that impair the function of neuromuscular junctions complain of painless weakness and fatigue. Autoantibodies that recognize key neuromuscular junction proteins are the most common cause of disrupted neuromuscular transmission, as found in *myasthenia gravis* (literally, "grave weakness"). Understandably, inherited defects in specialized neuromuscular junction proteins are also associated with myasthenic syndromes. Disorders caused by toxins that alter neuromuscular transmission are rarely encountered, but had an important role historically in elucidating how the neuromuscular junction functions.

Antibody-Mediated Diseases of the Neuromuscular Junction

Myasthenia Gravis

Myasthenia gravis is an autoimmune disease that is usually associated with autoantibodies directed against ACh receptors. It has a prevalence of 150 to 200 per 1 million and shows a bimodal age distribution. In young adults, the female-to-male ratio is 2:1, but in older adults there is a male predominance.

Pathogenesis

About 85% of patients have autoantibodies against postsynaptic ACh receptors, while most of the remaining patients have antibodies against the sarcolemmal protein muscle-specific receptor tyrosine kinase. These autoantibodies appear to be pathogenic, as the disease can be passively transferred to animals with serum from affected individuals, and therapeutic maneuvers that decrease autoantibody levels are associated with a reduction in symptoms.

The mechanism of action of the various autoantibodies appears to differ. Anti–ACh receptor antibodies are thought to lead to the aggregation and degradation of the receptors, as well as to damage of the postsynaptic membrane through complement fixation. As a result, postsynaptic membranes show alterations in morphology and are depleted of ACh receptors. This limits the ability of myofibers to respond to ACh. Autoantibodies directed against muscle-specific receptor tyrosine kinase do not fix complement. Instead, these antibodies seem to interfere with the trafficking and

clustering of ACh receptor within the sarcolemmal membrane, the net effect again being decreased ACh receptor function.

There is a strong association between pathogenic anti–ACh receptor autoantibodies and thymic abnormalities. Approximately 10% of patients with myasthenia gravis have a *thymoma,* a tumor of thymic epithelial cells (Chapter 13). An additional 30% of patients (and particularly young patients) have a different thymic abnormality called thymic hyperplasia. This peculiar condition is marked by the appearance of B-cell follicles in the thymus. The thymus normally contains small numbers of myoid cells, stromal cells that express skeletal muscle antigens. It is hypothesized that both thymoma and thymic hyperplasia disrupt normal thymic function in a manner that promotes autoimmunity against ACh receptors expressed on thymic myoid cells. In contrast, thymic abnormalities are usually absent in cases of myasthenia gravis that occur in older patients; the basis for the development of autoantibodies in such cases is unknown.

Clinical Features

Patients with anti–ACh receptor antibodies typically present with fluctuating weakness that worsens with exertion and often over the course of the day. Diplopia and ptosis due to involvement of extraocular muscles are common and distinguish myasthenia gravis from myopathies, in which involvement of extraocular muscles is unusual. In some patients, symptoms are confined to ocular muscles, while others develop generalized weakness that can be so severe as to require mechanical ventilation. Cases with antibodies against muscle-specific receptor tyrosine kinase differ from typical cases by exhibiting more focal muscle involvement (neck, shoulder, facial, respiratory, and bulbar muscles).

Diagnosis is based on clinical history, physical findings, identification of autoantibodies, and electrophysiologic studies. Electrophysiologic studies reveal a decrement in muscle response with repeated stimulation, a characteristic of this disorder. Overall mortality has dropped from over 30% in the 1950s to less than 5% with current therapies. Acetylcholinesterase inhibitors that increase the half-life of ACh are the first line of treatment. Other treatments, such as plasmapheresis and immunosuppressive drugs (e.g., glucocorticoids, cyclosporine, rituximab), can bring symptoms under control by decreasing autoantibody production. Thymectomy is often effective in patients with thymoma, but is of uncertain benefit in those with thymic hyperplasia or lacking thymic abnormalities.

Lambert-Eaton Myasthenic Syndrome

Lambert-Eaton myasthenic syndrome is an autoimmune disorder caused by antibodies that block ACh release by inhibiting a presynaptic calcium channel. In contrast to myasthenia gravis, rapid repetitive stimulation increases muscle response. Muscle strength is augmented after a few seconds of muscle activity. Patients typically present with weakness of their extremities. In about half of cases there is an underlying malignancy, most often neuroendocrine carcinoma of the lung. Symptoms may precede the diagnosis of cancer, sometimes by years. It is thought that the stimulus for autoantibody formation in paraneoplastic cases may be the expression of the same calcium channel in the neoplastic cells. Patients without cancer often have other autoimmune diseases, such as vitiligo or thyroid disease. Treatment consists of drugs that increase ACh release by depolarizing synaptic membranes and immunosuppressive agents, such as those used to treat myasthenia gravis.

Congenital Myasthenic Syndromes

These rare disorders, which most commonly have an autosomal recessive mode of inheritance, are marked by varying degrees of muscle weakness. Causative mutations have been identified in genes encoding several different presynaptic, synaptic, or postsynaptic proteins. The most common of these are loss-of-function mutations in the gene encoding the ε-subunit of the ACh receptor. Another group of mutations affect proteins that are important in normal clustering of ACh receptors on postsynaptic membranes. Many patients with congenital myasthenic syndromes present in the perinatal period with poor muscle tone, external eye muscle weakness, and breathing difficulties, but others have milder forms of the disease and may not come to clinical attention until adolescence or adulthood. The clinical presentation, response to drugs such as acetylcholinesterase inhibitors, and prognosis depend largely on the underlying mutation.

Disorders Caused by Toxins

Botulism is caused by exposure to a neurotoxin (botulinum toxin, popularly known as Botox) that is produced by the anaerobic gram-positive organism *Clostridium botulinum*. Botox acts by blocking the release of ACh from presynaptic neurons (Chapter 8). Curare is the common name for plant-derived muscle relaxants that block ACh receptors, resulting in flaccid paralysis. It was initially discovered and used as poison on arrow tips by indigenous people in the Amazon rain forest. At one time it was used as a muscle relaxant during certain forms of surgery, but has now been supplanted by other related drugs with a similar mechanism of action.

KEY CONCEPTS

DISEASES OF THE NEUROMUSCULAR JUNCTION

- Disorders of neuromuscular junctions present with painless weakness.
- Myasthenia gravis and Lambert-Eaton myasthenic syndrome, the most common forms, are both immune mediated, being caused by antibodies to postsynaptic ACh receptors and presynaptic calcium channels, respectively.
- Myasthenia gravis is often associated with thymic hyperplasia or thymoma, frequently involves ocular muscles, and is marked by fluctuating weakness that worsens with exertion.
- Lambert-Eaton myasthenic syndrome presents with weakness in the extremities that improves with repetitive stimulation and is often a paraneoplastic disorder associated with lung cancer.
- Genetic defects in neuromuscular junction proteins give rise to congenital myasthenic syndromes.
- Biologic toxins such as Botox can block neuromuscular transmission by blocking the release of ACh from presynaptic neurons.

DISEASES OF SKELETAL MUSCLE

Skeletal muscle has unique structural, cellular, and molecular characteristics and accordingly unique patterns of injury and repair. During embryogenesis, skeletal muscle develops through the fusion of mononucleated precursor cells (myoblasts) into multinucleated myotubes. These subsequently mature into myofibers (muscle fibers) of varying length that contain thousands of nuclei. In adult tissues, these myofibers are arranged in fascicles, each associated with a small pool of tissue stem cells referred to as satellite cells, which can contribute to muscle regeneration following injury (outlined later). Myofibers are of two main functional types, type I and type II (Table 27.1), which are admixed in a checkerboard pattern in normal skeletal muscle. Fiber type is determined by signals received from innervating motor neurons, and as a result all fibers that are part of a motor unit are of the same type.

Skeletal Muscle Atrophy

Skeletal muscle atrophy is a common feature of many disorders. Loss of innervation, disuse, cachexia, old age, and primary myopathies all can produce muscle atrophy and, if the atrophy is severe, loss of muscle mass. Although muscle atrophy may occur under many conditions, certain patterns of atrophy are suggestive of specific underlying etiologies:

- *Clusters or groups of atrophic fibers* are seen in neurogenic disease (Fig. 27.7).
- *Perifascicular atrophy* is seen in dermatomyositis (see later).
- *Type II fiber atrophy* with sparing of type I fibers is seen with prolonged corticosteroid therapy or disuse.

Table 27.1 Muscle Fiber Types

	Type I	Type II
Action	Sustained force	Fast movement
Activity type	Aerobic exercise	Anaerobic exercise
Power produced	Low	High
Resistance to fatigue	High	Low
Lipid content	High	Low
Glycogen content	Low	High
Energy metabolism	Low glycolytic capacity, high oxidative capacity	High glycolytic capacity, low oxidative capacity
Mitochondrial density	High	Low
Enzyme activity	NADH-TR, dark staining ATPase at pH 4.3, dark staining ATPase at pH 9.4, light staining	NADH-TR, light staining ATPase at pH 4.3, light staining ATPase at pH 9.4, dark staining
Myosin heavy chain gene expressed	*MYH7*	*MYH2, MYH4, MYH1*
Color	Red (high myoglobin content)	Pale red/tan (low myoglobin content)
Prototype	Soleus (pigeon)	Pectoral (pigeon)

ATPase, Adenosine triphosphatase; *NADH-TR*, nicotinamide adenine dinucleotide, reduced form, tetrazolium reductase.

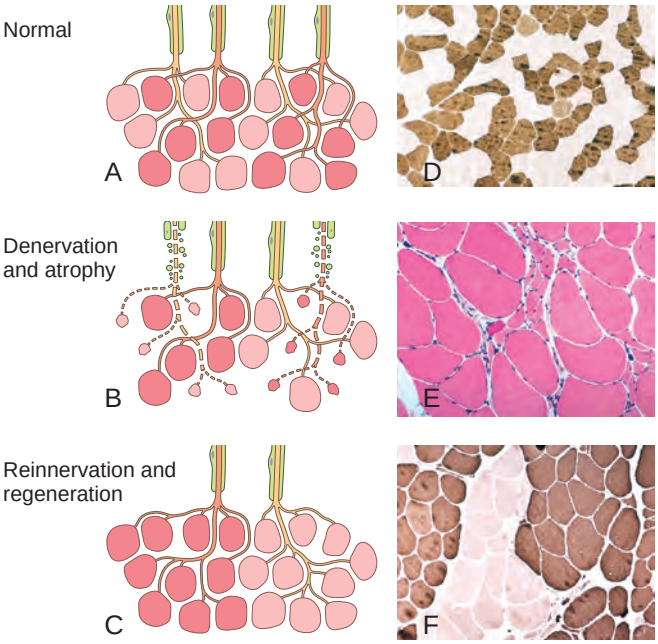


Figure 27.7 (A) This diagrammatic representation of four normal motor units shows a normal checkerboard-type admixture of light- and dark-stained fibers of opposite type. (B) Damage to innervating axons leads to loss of trophic input and atrophy of myofibers. (C) Reinnervation of myofibers can lead to a switch in fiber type and segregation of fibers of like type. As illustrated here, reinnervation is also often associated with an increase in motor unit size, with more myofibers innervated by an individual axon. (D) Normal muscle has a checkerboard-type distribution of type I (light) and type II (dark) fibers on this adenosine triphosphatase reaction (pH 9.4) corresponding to findings in diagram in A. (E) Clustered flattened, "angulated" atrophic fibers *(group atrophy)* are a typical finding associated with disrupted innervation. (F) With ongoing denervation and reinnervation, large clusters of fibers appear that all share the same fiber type *(type grouping)*, corresponding to diagram in C.

Neurogenic and Myopathic Changes in Skeletal Muscle

Disorders of skeletal muscles may be caused by direct injury to myofibers (myopathic injury) or by disruption of muscle innervation (neurogenic injury). Each type is described next.

Neurogenic injuries lead to *fiber type grouping* and *grouped atrophy* (see Fig. 27.7), both of which stem from the disruption of muscle innervation. The key to understanding these abnormalities is to recognize that muscle fiber type is determined by the innervating motor neuron. Following denervation, myofibers undergo atrophy, often assuming a flattened, angulated shape. Reinnervation restores fiber size and shape, but may make a denervated myofiber part of a different motor unit, in turn leading to a change in fiber type. In the face of ongoing axonal or neuronal damage and dropout, residual motor axons may innervate increasingly larger numbers of myofibers, leading to enlargement of motor units, each comprised of a single type of muscle fiber (fiber type grouping). These large motor units are also susceptible to grouped atrophy if the innervating axon is damaged.

In contrast, most primary myopathic injuries are associated with a distinct set of morphologic changes that include the following:

- *Segmental myofiber degeneration and regeneration* is seen when only part of a myofiber undergoes necrosis. Degeneration is associated with release of cytoplasmic enzymes such as creatine kinase into the blood, making these useful markers of muscle damage. The sarcomeres and other components of the damaged myofiber segment are engulfed by macrophages in a process termed myophagocytosis. Removal of the degenerated debris sets the stage for regeneration. Fusion of activated satellite cells to damaged myofibers is an important step for regeneration. Eventually, new sarcomeres are generated, and the continuity of the original myofiber is restored. Regenerating myofibers are rich in RNA and therefore blue (basophilic) in hematoxylin and eosin–stained sections. They have enlarged nuclei with prominent nucleoli that are often randomly distributed in the cytoplasm, instead of occupying their normal subsarcolemmal location. Depending on the nature of the primary insult, atrophic myofibers may also be seen. Regeneration can restore normal muscle following an acute, transient injury, but in chronic disease states regeneration often fails to keep pace with damage. In this setting, muscles often show endomysial fibrosis (collagen deposition), dropout of myofibers, and fatty replacement.
- *Myofiber hypertrophy* can be seen as a physiologic adaptation to exercise or in association with certain chronic myopathic conditions.
- *Cytoplasmic inclusions* in the form of vacuoles, aggregates of proteins, or clustered organelles are characteristic of several primary forms of myopathy.

Inflammatory Myopathies

Historically, dermatomyositis, polymyositis, and inclusion body myositis have been considered the three main primary inflammatory myopathies:

- *Dermatomyositis* as a disease that affects skin as well as skeletal muscles. The histologic hallmark, discussed below, is perifascicular atrophy.
- *Inclusion body myositis,* as a slowly progressive disease associated with distinct inclusions termed "rimmed vacuoles" with some similarities to neurodegenerative diseases of the CNS.
- *Polymyositis,* to some extent a diagnosis of exclusion, as a T cell–mediated autoimmune disease affecting skeletal muscles that lacks the features of dermatomyositis or inclusion body myositis.

In addition to the triad mentioned above, other entities have been described and updated classifications have been proposed, but the old historical triad is still used by many physicians. In a more granular classification system, many cases traditionally viewed as polymyositis are now regarded as immune-mediated necrotizing myopathy (IMNM), or as a connective tissue disease–associated myositis. With that, polymyositis is a less common diagnosis than in the past. The systemic connective tissue disorders that can affect skeletal muscle include systemic lupus erythematosus, systemic sclerosis, and sarcoidosis (Chapter 6). Rarely certain infectious agents can cause inflammation of skeletal muscle (Chapter 8).

Dermatomyositis

Dermatomyositis is a systemic autoimmune disease that typically presents with proximal muscle weakness and skin changes.

Pathogenesis

Dermatomyositis is an immunologic disease in which damage to small blood vessels contributes to muscle injury. The vasculopathic changes can be seen as *telangiectasias* (dilated capillary loops) in the nail folds, eyelids, and gums and as dropout of capillary vessels in skeletal muscle. Biopsies of muscle and skin may show deposition of the complement membrane attack complex (C5b-9) within capillary beds. An inflammatory signature enriched for genes that are upregulated by type I interferons is seen in muscle and in leukocytes. The prominence of this signature appears to correlate with disease activity. Direct immunologic injury to the muscle fibers may also play a role. Various autoantibodies are detected by serologic studies, and B lymphocytes as well as plasma cells are part of the inflammatory infiltrate that is seen in muscles. Certain autoantibodies detected in the circulation tend to be associated with specific clinical features:

- *Anti-Mi2 antibodies* (directed against a helicase implicated in nucleosome remodeling) show a strong association with prominent Gottron papules and heliotrope rash (described later).
- *Anti-Jo1 antibodies* (directed against the enzyme histidyl t-RNA synthetase) are associated with interstitial lung disease, nonerosive arthritis, and a skin rash described as "mechanic's hands." In some classification schemes myopathies with anti-Jo1 antibodies are described separately as "anti-synthetase syndrome associated myositis."
- *Anti-P155/P140 antibodies* (directed against several transcriptional regulators) are associated with paraneoplastic and juvenile cases of dermatomyositis.

A direct link between these autoantibodies and disease pathogenesis has not yet been established.

MORPHOLOGY

Muscle biopsies of affected patients show infiltrates of mononuclear inflammatory cells that tend to be most pronounced in the perimysial connective tissue and around blood vessels. Sometimes there is a distinctive pattern in which myofiber atrophy is accentuated at the edges of the fascicles—*perifascicular atrophy* (Fig. 27.8B). Segmental fiber necrosis and regeneration may also be seen. Immunohistochemical studies may identify an infiltrate rich in CD4+ T-helper cells and the deposition of C5b-9 in capillary vessels. Electron microscopic studies may show tubuloreticular endothelial cell inclusions, a feature of a number of inflammatory disorders that are linked to a type I interferon response.

Clinical Features

Muscle weakness is slow in onset, symmetric, and often accompanied by myalgias. It typically affects the proximal muscles first. As a result, tasks such as getting up from a

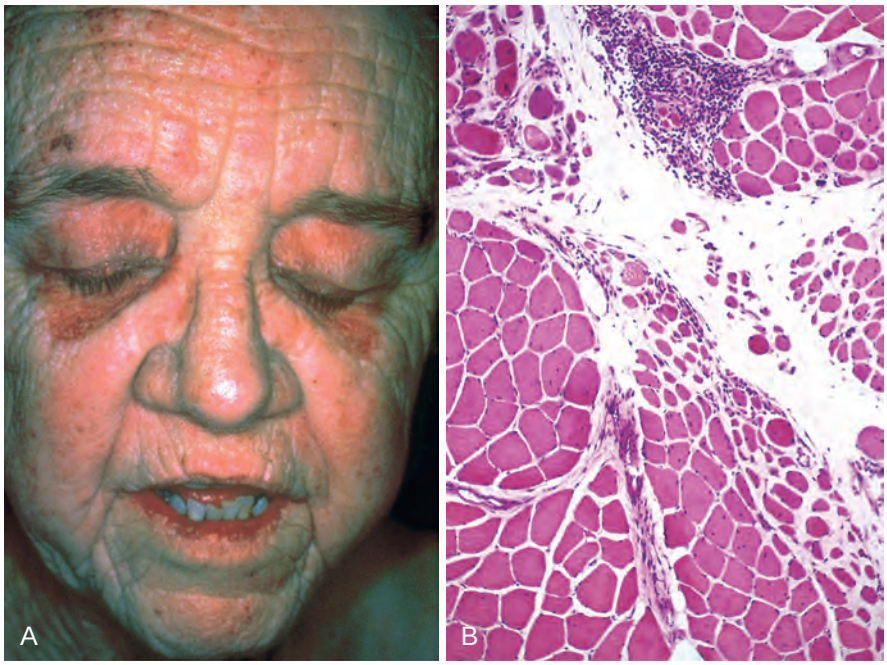


Figure 27.8 (A) Dermatomyositis. Note the heliotrope rash affecting the eyelids. (B) Dermatomyositis. The histologic appearance of muscle shows perifascicular atrophy of muscle fibers and inflammation. (Courtesy Dr. Dennis Burns, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Tex.)

chair and climbing steps become increasingly difficult. Fine movements controlled by distal muscles are affected only late in the disease. Associated myopathic changes on electrophysiologic studies and elevation in serum creatine kinase levels are reflective of muscle damage. Various rashes are described in dermatomyositis, but the most characteristic ones are a lilac-colored discoloration of the upper eyelids *(heliotrope rash)* associated with periorbital edema (Fig. 27.8A) and a scaling erythematous eruption or dusky red patches over the knuckles, elbows, and knees *(Gottron papules)*. Dysphagia resulting from involvement of oropharyngeal and esophageal muscles occurs in one-third of the affected individuals, and another 10% of patients have interstitial lung disease, which can sometimes be rapidly progressive and lead to death. Cardiac involvement is common, but rarely leads to cardiac failure.

Dermatomyositis may occur in adults or in children. The average age of onset of juvenile dermatomyositis is 7 years, whereas adult cases tend to present from the fourth to sixth decade of life. Dermatomyositis is the most common inflammatory myopathy in children. Compared to adult disease, childhood disease is more likely to be associated with calcinosis and lipodystrophy and less likely to be associated with myositis-specific antibodies, cardiac involvement, interstitial lung disease, or an underlying malignancy. As might be expected based on these differences, the overall prognosis is better in children than in adults. From 15% to 24% of adult patients have an associated malignancy, and in such patients, dermatomyositis may be viewed as a paraneoplastic disorder. A dermatomyositis-like picture with myositis and rash has also been described following treatment of cancer with check-point inhibitors (Chapter 7).

Immune-Mediated Necrotizing Myopathy (IMNM)

IMNM represents an autoimmune disease that is often associated with distinct autoantibodies (sometimes also referred to as necrotizing autoimmune myopathy). Affected patients present with subacute muscle weakness that is typically associated with significantly increased creatine kinase levels. Muscle biopsies show fairly prominent myofiber necrosis and regeneration in most cases, while inflammatory cell infiltrates are usually absent or minimal despite the autoimmune nature of the disease. In many patients, IMNM is associated with autoantibodies against HMG-CoA reductase. Their formation is often attributed to prior statin exposure, but some patients develop these autoantibodies without prior use of these medications. In some patients IMNM has been linked to antibodies to the signal recognition particle

Polymyositis

Polymyositis is an adult-onset inflammatory myopathy that shares myalgia and weakness with dermatomyositis but lacks its distinctive cutaneous features and is therefore to some degree a diagnosis of exclusion. Descriptions of other entities have made cases of polymyositis less common as outlined above.

Pathogenesis

The pathogenesis of polymyositis is uncertain, but it is believed to have an immunologic basis. CD8+ cytotoxic T cells are a prominent part of the inflammatory infiltrate in affected muscle, and it is hypothesized that these cells are the mediators of tissue damage. Unlike dermatomyositis, vascular injury does not have a major role in the pathogenesis of polymyositis.

MORPHOLOGY

Mononuclear inflammatory cell infiltrates are present, but in contrast to dermatomyositis, these are usually endomysial in location. Sometimes myofibers with otherwise normal morphology appear to be invaded by mononuclear inflammatory cells, predominantly CD8+ T cells. Degenerating necrotic, regenerating, and atrophic myofibers are typically found in a random or patchy distribution. The perifascicular pattern of atrophy that is characteristic of dermatomyositis is absent.

Inclusion Body Myositis

Inclusion body myositis is a disease of late adulthood that typically affects patients older than 50 years and is the most common inflammatory myopathy in patients older than age 65 years. Most affected individuals present with slowly progressive muscle weakness that tends to be most severe in the quadriceps and the distal upper extremity muscles. Dysphagia from esophageal and pharyngeal muscle involvement is not uncommon. Laboratory studies usually show modestly elevated creatine kinase levels; most myositis-associated autoantibodies are absent, although an antibody to cytosolic 5′-nucleotidase 1A (cN1A) has recently been described. It is present in about half of the patients and is a useful marker of the disease.

MORPHOLOGY

Inclusion body myositis has a number of features that are similar to those found in polymyositis, including:

- Patchy, often endomysial mononuclear inflammatory cell infiltrates rich in CD8+ T cells
- Increased sarcolemmal expression of major histocompatibility complex class I antigens
- Focal invasion of normal-appearing myofibers by inflammatory cells
- Admixed degenerating and regenerating myofibers

Other associated changes, however, are more typical or even specific for inclusion body myositis, as follows:

- Abnormal cytoplasmic inclusions described as "rimmed vacuoles" (Fig. 27.9)
- Tubolofilamentous inclusions in myofibers, seen by electron microscopy
- Cytoplasmic inclusions containing proteins typically associated with neurodegenerative diseases, like beta-amyloid, TDP-43, and ubiquitin
- Endomysial fibrosis and fatty replacement, reflective of a chronic disease course

Whether inclusion body myositis is indeed an inflammatory condition or a degenerative process with secondary inflammatory changes remains an unresolved question. It has certain features in common with polymyositis, as discussed earlier. On the other hand, it shares some features with neurodegenerative diseases, such as the presence of abnormal protein aggregates. Furthermore, there are several familial inclusion body myopathies that are also associated with chronic myopathic changes and rimmed vacuoles. These typically lack any associated inflammation—hence the designation *inclusion body "myopathy"* rather than *"myositis."* Inflammatory myopathies are treated with immunosuppresive drugs such as steroids, azathioprine, and in some cases intravenous immunoglobulins (IVIG). Such treatments have improved the prognosis of dermatomyositis and polymyositis, but inclusion body myositis responds poorly. This is another feature that argues against an inflammatory or immune origin of inclusion body myositis.

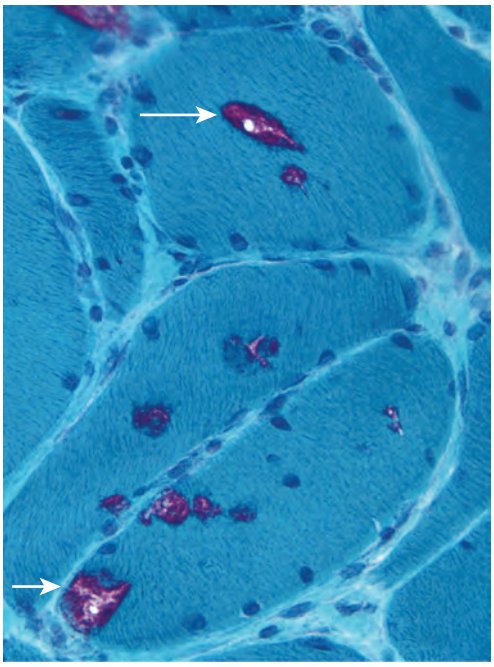


Figure 27.9 Inclusion body myositis, showing myofibers containing rimmed vacuoles—inclusions with reddish granular rimming *(arrows)*. Modified Gomori trichrome stain.

Toxic Myopathies

Toxic myopathies can be caused by prescription or recreational drugs or by certain hormonal imbalances. Among prescription drugs, *statins* are among the leading culprits. Statins are cholesterol-lowering drugs that are widely used to reduce the risks of acute ischemic cardiac events and stroke. **Myopathy is the most common complication of statins** (e.g., atorvastatin, simvastatin, pravastatin). Pure toxic injury is in part linked to dose and statin type and has to be distinguished from IMNM, which is often caused by statin-induced autoantibodies.

Chloroquine and hydroxychloroquine were originally used as antimalarial agents and are currently given as long-term therapy to some patients with systemic autoimmune diseases. These drugs interfere with normal lysosomal function and can cause a drug-induced lysosomal storage myopathy that presents with slowly progressive muscle weakness. The muscle tissue shows myopathic changes including vacuolation that predominantly affects type I fibers. Ultrastructural studies identify aggregates of whorled, lamellar membranous structures, including curvilinear bodies that mimic those seen in ceroid lipofuscinoses (Chapter 28). Cardiac muscle can also be affected by these drugs and can exhibit similar pathologic changes.

ICU myopathy, also known as myosin-deficient myopathy, is a neuromuscular disorder seen in patients during the course of treatment for critical illness (usually in an ICU) especially with corticosteroid therapy. There may be relatively selective degradation of sarcomeric myosin thick

filaments, producing profound weakness that can complicate the clinical course (e.g., by interfering with the weaning of a patient from a mechanical respirator).

Thyroid dysfunction can lead to several types of myopathy. *Thyrotoxic myopathy* presents most commonly as an acute or chronic proximal muscle weakness that may precede other signs of hyperthyroidism. Such patients may also present with exophthalmic ophthalmoplegia, characterized by swelling of the eyelids, edema of the conjunctiva, and diplopia. Hypothyroidism can cause cramping or aching of muscles, and decreased movement. Reflexes may be slowed. Findings in skeletal muscle include fiber atrophy, an increased number of abnormally localized nuclei, glycogen aggregates, and (occasionally) deposition of mucopolysaccharides in connective tissue.

Alcohol can also be myopathic. Most notably, binge drinking may produce an acute toxic syndrome of rhabdomyolysis, myoglobinuria, and renal failure. The affected individual may complain of acute myalgias that are generalized or confined to a single muscle group.

Inherited Diseases of Skeletal Muscle

Inherited mutations are responsible for a diverse collection of disorders marked by defects in skeletal muscle. In some of these disorders, skeletal muscle is the main site of disease, but in others multiple organs are involved. Of the other organs involved, the heart is of particular importance, since cardiac involvement is common and often life limiting.

Historically, inherited myopathies have been subdivided into several broad categories based on inheritance pattern, anatomic pattern of muscle involvement, onset age, clinical course, and underlying pathogenesis. These categories should largely be understood as illustrative of key concepts. On closer inspection, there are exceptions and there is significant overlap between these disease categories. Included in the group of inherited muscular disorders are the following.

Congenital myopathies (Table 27.2) typically present in infancy with muscle defects that tend to be static or to even improve over time. They are often associated with distinct structural abnormalities of the muscle.

Muscular dystrophies, characterized by progressive muscle damage that typically comes to attention after infancy.

Congenital muscular dystrophies, by contrast, tend to present in infancy and are often associated with developmental abnormalities of the CNS as well as progressive muscle damage. Congenital muscular dystrophies include two important groups:

- *Conditions with defects in extracellular matrix surrounding myofibers.* These are exemplified by Ullrich congenital muscular dystrophy (UCMD) and merosin deficiency. In the former the causative mutations involve one of

Table 27.2 Congenital Myopathies

Disease and Inheritance	Gene and Locus	Clinical Findings	Pathologic Findings
Central core disease; autosomal dominant	Ryanodine receptor-1 *(RYR1)* gene; 19q13.2	Early-onset hypotonia and weakness; "floppy infant"; associated skeletal abnormalities like scoliosis, hip dislocation, or foot deformities; some *RYR1* mutations cause central core disease, some cause malignant hyperthermia, and some cause both	Cytoplasmic cores represent demarcated central zones in which the normal arrangement of sarcomeres is disrupted and mitochondria are decreased in number
Nemaline myopathy (NEM)	AD NEM1—α-tropomyosin 3 *(TPM3)* gene; 1q21.3 AR NEM2—nebulin *(NEB)* gene; 2q23.3 AR NEM3—α-actin-1 *(ACTA1)* gene; 1q42.13 AD NEM4—tropomyosin-2 *(TPM2)* gene; 9p13.3 AR NEM5—troponin T1 *(TNNT1)* gene; 19q13.42 AR NEM7—coffilin-2 *(CFL2)* gene; 14q13.1	Childhood weakness; some with more severe weakness, hypotonia at birth ("floppy infant")	Aggregates of spindle-shaped particles *(nemaline rods)*; occur predominantly in type 1 fibers; derived from Z-band material (α-actinin) and best seen on modified Gomori stain or by electron microscopy
Centronuclear (myotubular) myopathy	XL—myotubularin *(MTM1)* gene; Xq28 AD—dynamin-2 *(DNM2)* gene (and others); 19p13.2 AR—amphiphysin-2 *(BIN1)* gene; 2q14.3	Severe congenital hypotonia, "floppy infant," and poor prognosis in X-linked form ("myotubular myopathy") Childhood onset or young adult onset with other variants with weakness and hypotonia	Many fibers contain nuclei in the geometric center of the myofiber; central nuclei are more common in type 1 fibers, which are small in diameter, but can occur in both fiber types
Congenital fiber type disproportion	Selenoprotein 1 *(SELENON)* gene; 1p36.11 α-actin-1 *(ACTA1)* gene; 1q42.13 Tropomyosin 3 *(TPM3)* gene; 1q21.3	Hypotonia, weakness, failure to thrive, facial and respiratory weakness, contractures Wide phenotypic spectrum Mutations in *SELENON* are also associated with protein aggregate myopathy and rigid spine muscular dystrophy; mutations in *ACTA1* are also associated with nemaline myopathy and protein aggregate myopathy; mutations in *TPM3* are also associated with nemaline myopathy	Predominance and atrophy of type 1 fibers (not specific)

AD, Autosomal dominant; *AR,* autosomal recessive; *XL,* X-linked.

three collagen VI alpha genes; in the case of merosin deficiency, the gene encoding merosin is disrupted. UCMD is characterized by hypotonia, proximal contractures, and distal hyperextensibility. A morphologic hallmark is mismatched expression of normally co-localized matrix proteins perlecan and collagen VI.

- *Conditions with abnormalities in receptors for extracellular matrix*. In this group are diseases that disrupt the post-translational modification of alpha-dystroglycan (Fig. 27.10) by *O*-linked glycosylation. Mutations of alpha-dystroglycan itself result in fetal demise, but defects in its post-translational modification result in milder forms of dystroglycan deficiency. Alpha-dystroglycan expression is important for CNS and eye development. Severe cases exhibit features of congenital muscular dystrophy as well as developmental defects of the CNS and eyes that cause seizures, intellectual disability, and blindness. Milder forms may only cause skeletal muscle disease. Some of these mutations are also linked to a presentation described as limb-girdle muscular dystrophy (see later).

The following section focuses on the most common and best understood forms of inherited myopathies.

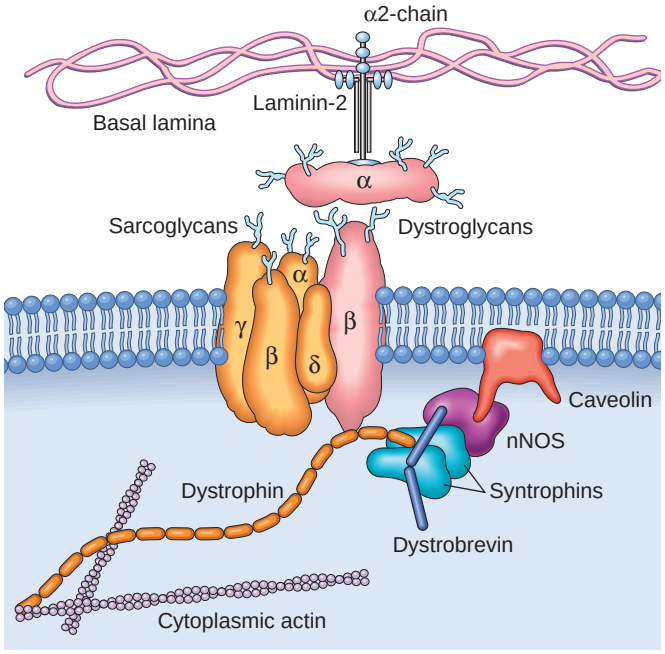


Figure 27.10 Relationship between the cell membrane (sarcolemma) and the sarcolemmal associated proteins. Dystrophin, an intracellular protein, forms an interface between the cytoskeletal proteins and a group of transmembrane proteins, the dystroglycans and the sarcoglycans. These transmembrane proteins have interactions with the extracellular matrix, including the laminin proteins. Dystrophin also interacts with dystrobrevin and the syntrophins, which form a link with neuronal-type nitric oxide synthetase *(nNOS)* and caveolin. Mutations in dystrophin are associated with the X-linked muscular dystrophies; mutations in caveolin and the sarcoglycan proteins are associated with the limb-girdle muscular dystrophies, which can be autosomal dominant or recessive disorders; and mutations in the α_2-laminin (merosin) are associated with autosomal recessive congenital muscular dystrophy. Sugar moieties on the dystroglycans are affected by disruption of enzymes responsible for this post-translational modification. Mutations in genes that encode these glycosylation enzymes can cause myopathy as discussed below. These moieties are depicted as light blue branching structure.

Muscular Dystrophies

Muscular dystrophies include many inherited disorders of skeletal muscle that have in common progressive muscle damage that typically manifests between childhood and adulthood. As mentioned earlier, with the exception of congenital muscular dystrophies, these diseases do not present in infancy. While our focus is on X-linked muscular dystrophies, other forms in which the disease pathogenesis is reasonably well understood are also briefly discussed.

X-Linked Muscular Dystrophy With Dystrophin Mutation Including Duchenne and Becker Muscular Dystrophy

The most common muscular dystrophies are X-linked and stem from mutations that disrupt the function of a large structural protein called dystrophin. As a result, these diseases are sometimes referred to as *dystrophinopathies*. The most common early-onset form is referred to as *Duchenne muscular dystrophy*. It has an incidence of 1 per 3500 live male births and has a severe progressive phenotype. *Becker muscular dystrophy* is a second relatively common dystrophinopathy that is characterized by later disease onset and a milder phenotype. Other rare dystrophinopathies may present with isolated cardiomyopathy, asymptomatic elevations of creatine kinase levels, or exercise intolerance because of myalgias and cramps. As with many X-linked diseases, female carriers of dystrophin mutations may be mildly symptomatic due to unfavorable X-chromosome inactivation.

Pathogenesis

Duchenne and Becker muscular dystrophy are caused by loss-of-function mutations in the dystrophin gene on the X chromosome. *Dystrophin* is one of the largest human genes, spanning 2.3 million base pairs and composed of 79 exons. The encoded protein, dystrophin, is a key component of the dystrophin glycoprotein complex (see Fig. 27.10). This complex spans the plasma membrane and serves as a link between the cytoskeleton inside the myofiber and the basement membrane outside of the cell. By doing so, dystrophin is thought to provide mechanical stability to the myofiber and its cell membrane during muscle contraction. Defects in the complex may lead to small membrane tears that permit influx of calcium, triggering events that result in myofiber degeneration. In addition to its mechanical function, dystrophin may have a role in signaling pathways; for example, its carboxy terminus interacts with nitric oxide synthase, which generates NO.

Identification and characterization of specific *dystrophin* mutations has provided an explanation for some of the phenotypic variation in patients with dystrophinopathies. Duchenne muscular dystrophy is typically associated with deletions or frameshift mutations that result in total absence of dystrophin. In contrast, the mutations in Becker muscular dystrophy typically permit the synthesis of a truncated version of dystrophin, which presumably retains partial function.

MORPHOLOGY

The basic morphologic changes that muscular dystrophies elicit in skeletal muscle tissue may differ in severity but do not

discriminate between different forms of dystrophy. The changes in Duchenne muscular dystrophy serve as an example. This disease is marked by chronic muscle damage that outpaces the capacity for repair (Fig. 27.11). Muscle biopsies in young boys show ongoing damage in the form of segmental myofiber degeneration and regeneration associated with an admixture of atrophic myofibers. The fascicular architecture is preserved at this stage of the disease, and there is usually no inflammation except for the presence of myophagocytosis. As the disease progresses, muscle tissue is replaced by collagen and fat cells ("fatty replacement" or "fatty infiltration"). The remaining myofibers at this point in the course show prominent variation in size, from small atrophic fibers to large hypertrophied fibers. This remodeling distorts the fascicular architecture of the muscle, which becomes markedly abnormal over time. **Immunohistochemical studies for dystrophin show absence of the normal sarcolemmal staining pattern in Duchenne muscular dystrophy and reduced staining in Becker muscular dystrophy.**

Clinical Features

Boys with Duchenne muscular dystrophy appear normal at birth. Very early motor milestones are met, but walking is often delayed. The first indications of muscle weakness are clumsiness and inability to keep up with peers. Weakness begins in the pelvic girdle muscles and then extends to the shoulder girdle. Enlargement of the muscles of the lower leg associated with weakness, termed pseudohypertrophy, is often present. The mean age of wheelchair dependence is around 9.5 years. Patients develop joint contractures, scoliosis, worsening respiratory reserve, and sleep hypoventilation.

Dystrophin is also expressed in the heart and the CNS. Dystrophin deficiency in cardiac muscle often leads to the development of cardiomyopathy and arrhythmias, particularly in older patients. Cognitive impairment and learning disabilities, presumably due to a functional role for dystrophin in the brain, is also common and sometimes produces frank intellectual disability. Despite supportive care, the mean age of death for patients with Duchenne muscular dystrophy is 25 to 30 years of age, with most patients succumbing to respiratory insufficiency, pulmonary infection, or heart failure. This is in contrast to Becker muscular dystrophy, which typically presents in later childhood, adolescence, or adult life; has slower progression; and may have a near-normal life expectancy.

The diagnosis is based on the history, physical examination, and laboratory studies. Serum creatine kinase is markedly elevated during the first decade of life due to ongoing muscle damage, and then falls as the disease progresses and muscle mass is lost. The detection of a dystrophin mutation offers definitive diagnosis.

Treatment of patients with dystrophinopathies is challenging. Current treatment consists primarily of supportive care. Definitive therapy requires restoration of dystrophin levels in skeletal and cardiac muscle fibers. Work in this area is emboldened by the recognition that expression of some dystrophin protein (as in patients with Becker muscular dystrophy) is sufficient to substantially ameliorate the disease phenotype. One approach involves the expression of antisense RNAs that alter RNA splicing so as to cause "skipping" of exons containing deleterious mutations, thus permitting the expression of a truncated, but partially functional, dystrophin protein. A second strategy is exploring the use of drugs that promote ribosomal "read-through" of stop codons, another ploy that may enable the expression of some dystrophin protein. Both of these approaches are mutation-specific and thus need to be tailored to individual

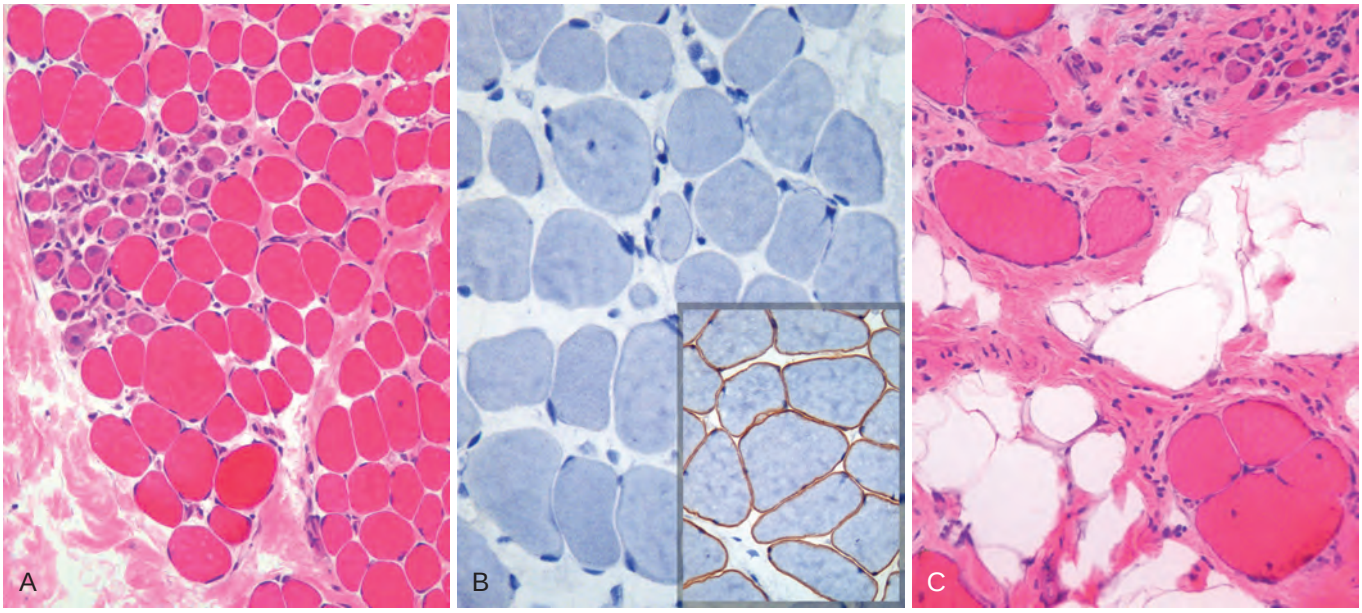


Figure 27.11 Duchenne muscular dystrophy. Histologic images of muscle biopsy specimens from two brothers. (A and B) Specimens from a 3-year-old boy. (C) Specimen from his 9-year-old brother. (A) At a younger age, fascicular muscle architecture is maintained, but myofibers show variation in size. Additionally, there is a cluster of basophilic regenerating myofibers *(left side)* and slight endomysial fibrosis, seen as focal pink-staining connective tissue between myofibers. (B) Immunohistochemical staining shows complete absence of membrane-associated dystrophin, seen as a brown stain in normal muscle *(inset)*. (C) Biopsy specimen from the older brother illustrates disease progression, which is marked by extensive variation in myofiber size, fatty replacement, and endomysial fibrosis.

patients. Gene therapy (introduction of a normal dystrophin gene) is being investigated, but gene delivery to skeletal muscle cells remains a daunting hurdle.

Limb-Girdle Muscular Dystrophy

Limb-girdle muscular dystrophies are a heterogeneous group of at least 8 autosomal dominant and 23 autosomal recessive entities. Their overall incidence is 1 in 25,000 to 50,000 individuals. As indicated by the name, all forms are characterized by muscle weakness that preferentially involves proximal muscle groups. Both age of onset and disease severity are highly variable. The causative mutations involve genes that participate in diverse cellular functions, making it difficult to discern a unifying mechanism of disease pathogenesis. Based on current knowledge, the implicated genes can be grouped according to function as follows:

- *Genes encoding structural components* (sarcoglycans) of the dystrophin glycoprotein complex
- *Genes encoding enzymes that are responsible for glycosylation of α-dystroglycan,* a component of the dystrophin glycoprotein complex (see Fig. 27.10)
- *Genes encoding proteins that associate with the Z-disks of sarcomeres*
- *Genes encoding proteins involved in vesicle trafficking and cell signaling*
- *Genes that seemingly stand alone, such as CAPN3 encoding the protease calpain 3 and lamin A/C* (which is also mutated in some patients with Emery-Dreifuss muscular dystrophy [EMD]; see below)

Myotonic Dystrophy

Myotonic dystrophy is an autosomal dominant multisystem disorder associated with skeletal muscle weakness, cataracts, endocrinopathy, and cardiomyopathy. It affects about 1 in 10,000 individuals. Myotonia, a sustained involuntary contraction of muscles, is a key feature of the disease. Rarely, patients present with "*congenital myotonia,*" marked by severe manifestations in infancy.

Pathogenesis

The disease is caused by expansions of CTG triplet repeats in the 3′-noncoding region of the myotonic dystrophy protein kinase *(DMPK)* gene, but precisely how this genetic aberration produces the disease phenotype is unknown. The correlation between the length of expansion and disease severity is variable compared to some other triplet repeat expansion disorders like Huntington disease (Chapter 28). Experimental studies suggest that the skeletal muscle phenotype stems from a "toxic" gain-of-function caused by the triplet repeat expansion. Specifically, the expanded CUG-repeats in the *DMPK* mRNA transcript appear to bind and sequester proteins like muscleblind-like splicing regulator 1, which have an important role in RNA splicing. The resulting disruption of normal splicing events leads to missplicing of other RNA transcripts, including the transcript for a chloride channel called *CLC1*. It is believed that the resulting deficiency of CLC1 is responsible for the characteristic myotonia. In support of this scenario, one rare form of congenital myotonia is caused by germline loss-of-function mutations in *CLC1,* indicating that CLC1 is required for normal muscle relaxation.

Emery-Dreifuss Muscular Dystrophy (EMD)

EMD is caused by mutations in genes that encode nuclear lamin proteins. Clinically, it is marked by a triad consisting of slowly progressive humeroperoneal weakness; cardiomyopathy associated with conduction defects; and early contractures of the Achilles tendon, spine, and elbows. The X-linked form (EMD1) and the autosomal form (EMD2) are caused by mutations in the genes encoding *emerin* and *lamin A/C,* respectively, both of which localize to the inner face of the nuclear membrane. It is hypothesized that these proteins help maintain the shape and mechanical stability of the nucleus during muscle contraction. They may also influence gene expression by affecting chromatin organization in the nucleus. How defects in these proteins produce the observed phenotypes is unknown.

Fascioscapulohumeral Dystrophy

Fascioscapulohumeral dystrophy is associated with a characteristic pattern of muscle involvement that includes prominent weakness of facial muscles and muscles of the shoulder girdle. It is an autosomal dominant disease affecting about 1 in 20,000 individuals.

Pathogenesis

The pathogenesis of fascioscapulohumeral dystrophy is complex and only partly understood. It results from various mechanisms that culminate in the (over)expression of a retrogene called *DUX4* that is located in a region of subtelomeric repeats on the long arm of chromosome 4. *DUX4* encodes a transcription factor, suggesting that the disease ultimately results from the overexpression of *DUX4* target genes.

Diseases of Lipid or Glycogen Metabolism

Many inborn errors of lipid or glycogen metabolism affect skeletal muscle. These disorders tend to produce one of two general patterns of muscle dysfunction: In some, patients become symptomatic only with exercise or fasting, which may produce severe muscle cramping and pain or even extensive muscle necrosis *(rhabdomyolysis)*. In others, there is slowly progressive damage of muscle, without episodic manifestations. Listed below are some examples:

- *Carnitine palmitoyltransferase II deficiency* is the most common disorder of lipid metabolism to cause episodic muscle damage with exercise or fasting. The defect in this disorder impairs the transport of free fatty acids into mitochondria.
- *Myophosphorylase deficiency (McArdle disease)* is one of the more common glycogen storage diseases affecting skeletal muscle; it also results in episodic muscle damage with exercise.
- *Acid maltase deficiency* results in impaired lysosomal conversion of glycogen to glucose, causing glycogen to accumulate within lysosomes. Severe deficiency results in the generalized glycogenosis of infancy, *Pompe disease* (Chapter 5). Milder deficiency can cause a progressive adult-onset myopathy that preferentially involves the respiratory and truncal muscles. Enzyme replacement therapy has been used to treat some affected patients.

Mitochondrial Myopathies

Mitochondrial diseases are complex systemic conditions that can involve many organ systems, including skeletal muscle. The genetics of these disorders are varied and unusually complex (discussed later), but many of the causative mutations appear to impair the ability of mitochondria to generate adenosine triphosphate (ATP). As a result, these diseases tend to affect skeletal muscles and other tissues rich in cell types with high ATP requirements, particularly cardiac muscle cells and neurons.

Skeletal muscle involvement can manifest as weakness, elevations in serum creatine kinase levels, or rhabdomyolysis. Although the anatomic pattern of muscle weakness is variable, involvement of extraocular eye muscles is common and can be a clue to the diagnosis. Indeed, *chronic progressive external ophthalmoplegia* is a common feature of mitochondrial disorders and may occur as an isolated phenomenon or a part of a multisystem syndrome. The reason that extraocular eye muscles are particularly sensitive to mitochondrial disease is uncertain, but it may be that these muscles have exceptionally high requirements for ATP. In line with this idea, extraocular eye muscles have the most mitochondria per mass of any of the body's muscles.

Mitochondrial proteins and tRNAs may be encoded by either the nuclear genome or the mitochondrial genome (mtDNA). While mutations in nuclear mitochondrial genes follow Mendelian inheritance patterns, mutations in mtDNA are maternally inherited, since all of the mitochondria in the embryo are contributed by the oocyte (Chapter 5). In addition, unlike nuclear DNA, which is present in only two copies and is evenly distributed from a mother cell to daughter cells during cell division, each cell contains thousands of mtDNA copies, which are distributed in a random fashion to daughter cells at the time of cell division. It is believed that disease results only when a certain threshold of mutated mtDNA copies is exceeded within a substantial fraction of "at-risk" cells (e.g., skeletal muscle cells) in a tissue.

MORPHOLOGY

The most consistent pathologic change in skeletal muscle is abnormal aggregates of mitochondria that are seen preferentially in the subsarcolemmal area of affected myofibers, producing an appearance that is referred to as "ragged red fibers" (Fig. 27.12). By electron microscopy, morphologically abnormal mitochondria may be seen. Loss of particular mitochondrial enzyme activities characterizes some mitochondrial diseases and may be appreciated by histochemical staining for cytochrome oxidase. Some mitochondrial diseases lack morphologic changes and can be diagnosed only through enzymatic assays or genetic analyses.

Clinical Features

Due to the complexity of mitochondrial genetics, genotype/phenotype relationships in mitochondrial disorders are not straightforward. For example, a single point mutation in the mitochondrial leucine tRNA gene may produce isolated chronic progressive external ophthalmoplegia in one patient and in another, a much more severe phenotype, *mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes*.

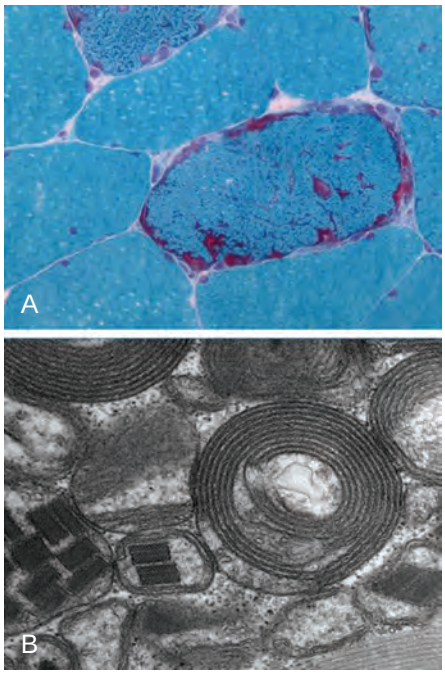


Figure 27.12 (A) Ragged red fiber with increased reddish granular subsarcolemmal staining reflective of abnormal aggregation of mitochondria. (B) Electron micrograph showing morphologically abnormal mitochondria with concentric membranous rings (so-called "phonograph records") and rhomboid paracrystalline inclusions *(lower left side)*.

Similarly, deletions in mtDNA may lead to either isolated ophthalmoplegia or to *Kearns-Sayre syndrome,* characterized by ophthalmoplegia, pigmentary degeneration of the retina, and complete heart block. *Myoclonic epilepsy with ragged red fibers* and *Leber hereditary optic neuropathy* are other examples of mitochondrial disease caused by point mutations in mtDNA. Many mitochondrial disorders, such as subacute necrotizing encephalopathy *(Leigh syndrome),* are remarkably heterogeneous genetically and may be caused by mutations in either mtDNA or the nuclear genome. In the case of Leigh syndrome, causative mutations have been identified in more than 30 different genes, the common feature being that all of the affected genes encode proteins with essential roles in mitochondrial metabolism.

Spinal Muscular Atrophy and Differential Diagnosis of a Hypotonic Infant

Spinal muscular atrophy is a neuropathic disorder in which loss of motor neurons leads to muscle weakness and atrophy. Infants with neurologic or neuromuscular disease may present with generalized hypotonia ("floppy infant"). The differential diagnosis of infantile hypotonia includes primary diseases of skeletal muscle (e.g., congenital myasthenic syndrome, congenital myotonia, congenital myopathies, and congenital muscular dystrophies); abnormalities of the brain (e.g., encephalopathy); and neuronopathies, of which spinal muscular atrophy is a prototypic example.

Spinal muscular atrophy is an autosomal recessive disorder with an incidence of 1 in 6000 births and is caused by loss-of-function mutations in the *SMN1* (survival of motor neuron-1) gene. The function of the gene is uncertain—the

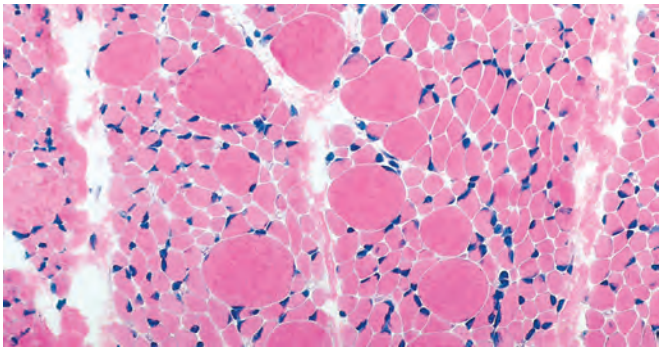


Figure 27.13 Spinal muscular atrophy with only rare hypertrophied myofibers admixed with numerous atrophic rounded myofibers. The larger fibers are those that are innervated and have undergone compensatory hypertrophy.

encoded protein may have a role in RNA splicing—but SMN1 deficiency has a dramatic effect on motor neuron survival, sometimes leading to loss of motor neurons in utero. The resulting denervation of skeletal muscle may lead to characteristic morphologic changes consisting of large zones of severely atrophic myofibers mixed with scattered normal-sized or hypertrophied myofibers, found individually or in small groups (Fig. 27.13). These normal or hypertrophied fibers are those that retain innervation from remaining motor neurons.

Ion Channel Myopathies (Channelopathies)

Channelopathies are a group of inherited diseases caused by mutations affecting the function of ion channel proteins. Most channelopathies are autosomal dominant disorders with variable penetrance. Depending on the channel that is affected, clinical manifestations may include epilepsy, migraine, movement disorders with cerebellar dysfunction, peripheral nerve disease, and muscle disease.

Different ion channel myopathies may cause decreased or increased excitability resulting in hypotonia or hypertonia. Disorders associated with hypotonia can be further subclassified based on whether symptomatic patients have elevated, depressed, or normal serum potassium levels and are called *hyperkalemic, hypokalemic,* or *normokalemic periodic paralysis,* respectively. Examples of gene mutations that are associated with muscle dysfunction are the following:

- *KCNJ2:* Mutations affecting this potassium channel cause *Andersen-Tawil syndrome,* an autosomal disorder associated with periodic paralysis, cardiac arrhythmias, and skeletal abnormalities.
- *SCN4A:* Mutations affecting this sodium channel cause several autosomal disorders with presentations ranging from myotonia to periodic paralysis.
- *CACNA1S:* Missense mutations in this protein, a subunit of a muscle calcium channel, are the most common cause of hypokalemic paralysis.
- *CLC1:* Mutations affecting this chloride channel cause myotonia congenita. As already discussed, *CLC1* expression is decreased in myotonic dystrophy.
- *RYR1:* Mutations in the *RYR1* gene disrupt the function of the ryanodine receptor, which regulates calcium release from the sarcoplasmic reticulum. *RYR1* mutations are linked to a congenital myopathy (central core disease) and to *malignant hyperthermia*. The latter is characterized by a hypermetabolic state (tachycardia, tachypnea, muscle spasms, and later hyperpyrexia) that can be triggered by anesthetics, most commonly halogenated inhalational agents and succinylcholine. Upon exposure to anesthetic, the mutated receptor allows increased efflux of calcium from the sarcoplasmic reticulum, leading to tetany and excessive heat production.

KEY CONCEPTS

DISORDERS OF SKELETAL MUSCLE

- Altered muscle function may stem from neurogenic or primary myopathic processes.
- Myopathic disorders are often marked by degeneration and regeneration of myofibers.
- The three traditional inflammatory myopathies are polymyositis, dermatomyositis, and inclusion body myositis, but other forms like immune-mediated necrotizing myopathy (IMNMs) are now recognized.
  - Inclusion body myositis is a chronic progressive disease of older patients associated with rimmed vacuoles.
  - Dermatomyositis occurs in children and adults, the latter frequently as a paraneoplastic disorder. Immune damage to small blood vessels and perifascicular atrophy are common features.
  - Polymyositis is an adult-onset myopathy caused by CD8+ T cells. This disease is less common than originally thought because many patients are reclassified as having IMNM, systemic connective tissue disease, or even early inclusion body myositis.
- Muscular dystrophies and congenital myopathies result from genetic mutations that disrupt the function of proteins that are important for various aspects of muscle development, function, and regeneration. Some of these diseases present in infancy, others in adulthood. They may be relentlessly progressive or cause relatively static deficits.
- Myopathy can result from toxic injury or be the result of metabolic diseases including those of lipid metabolism, glycogen metabolism, and mitochondria.

PERIPHERAL NERVE SHEATH TUMORS

The vast majority of benign and malignant neoplasms of peripheral nerve sheaths are composed of cells that show evidence of Schwann cell differentiation. These include the three common types: schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (MPNST). Other rare tumors arising from nerves may show evidence of perineurial cell differentiation. There is an abrupt transition between myelination by oligodendrocytes (central myelin) and myelination by Schwann cells (peripheral myelin) that occurs as nerves extend out from the substance of the brain. Thus, peripheral nerve tumors sometimes arise within the dura as well as along the distal course of peripheral nerves.

Peripheral nerve sheath tumors have several unique features. Firstly, their association with relatively common familial tumor syndromes, including neurofibromatosis type 1 (NF1), neurofibromatosis type 2 (NF2), and schwannomatosis.

Secondly, MPNSTs seen in the context of NF1 are thought to arise through malignant transformation of preexisting benign plexiform neurofibromas. Although malignant transformation of a preexisting benign lesion is a common origin for certain carcinomas (e.g., colon cancer), it is unusual in soft tissue tumors. Tumors with skeletal muscle differentiation are discussed in Chapter 26.

Schwannomas

These are benign tumors that exhibit Schwann cell differentiation and often arise directly from peripheral nerves. Schwannomas are a component of NF2, and even sporadic schwannomas are commonly associated with inactivating mutations in the *NF2* gene on chromosome 22. Loss of expression of the *NF2* gene product, merlin, is a consistent finding in all schwannomas. Through interactions affecting the actin cytoskeleton, merlin participates in the regulation of several key signaling pathways that are involved in control of cell shape, cell growth, and the attachment of cells to one another (cell adhesion).

MORPHOLOGY

Schwannomas are well-circumscribed, encapsulated masses that are attached loosely to the associated nerve without invading it. As a result they can often be resected without sacrificing nerve function. Grossly, these tumors form firm, gray masses. Microscopically, they are comprised of an admixture of dense and loose areas referred to as **Antoni A** and **Antoni B** areas, respectively (Fig. 27.14A). The dense eosinophilic Antoni A areas often contain spindle cells arranged into cellular intersecting fascicles. Palisading of nuclei is common. Resulting structures with central "nuclear-free zones" ramified by palisading nuclei are termed **Verocay bodies** (Fig. 27.14B). In the loose, hypocellular **Antoni B** areas the spindle cells are spread apart by a prominent myxoid extracellular matrix that may be associated with microcyst formation. Schwann cells are characterized by the presence of a spindled elongated nucleus with a wavy or buckled shape. Electron microscopy shows basement membrane deposits encasing single cells and collagen fibers. Because the lesion displaces the nerve of origin as it grows, axons are largely excluded from the tumor, although they may become entrapped in the capsule. The Schwann cell origin of these tumors is borne out by their uniform immunoreactivity for S-100. A variety of degenerative changes may be found in schwannomas, including nuclear pleomorphism, xanthomatous change, vascular hyalinization, cystic change, and necrosis. Some large mitotically active schwannomas lacking Antoni B areas can mimic a sarcoma. Schwannomas may recur locally if incompletely resected, but malignant transformation is extremely rare (in contrast to plexiform neurofibromas, discussed later).

Clinical Features

Most schwannomas cause symptoms by local compression of the involved nerve or adjacent structures (e.g., brainstem or spinal cord). Within the cranial vault, most schwannomas occur at the cerebellopontine angle, where they are attached to the vestibular branch of the eighth nerve. Affected individuals often present with tinnitus and hearing loss; the tumor is commonly referred to as an *acoustic neuroma*—a double misnomer, since the tumor neither arises from the acoustic portion of the nerve nor is it a neuroma. Elsewhere within the dura, sensory nerves are preferentially involved, including branches of the trigeminal nerve and dorsal roots. When extradural, schwannomas can arise in association with large nerve trunks or as soft tissue lesions without an identifiable associated nerve. Surgical removal is curative.

Neurofibromas

Neurofibromas are benign nerve sheath tumors that are more heterogeneous in composition than schwannomas. **The neoplastic Schwann cells are admixed with perineurial-like cells, fibroblasts, mast cells, and CD34+ spindle cells**. Neurofibromas may be either sporadic or NF1-associated. Different types of neurofibroma can be distinguished depending on their growth pattern.

- *Superficial cutaneous neurofibromas* often present as pedunculated nodules that can be seen isolated (if sporadic) or multiple (if NF1-associated).
- *Diffuse neurofibromas* often present as a large plaque-like elevation of skin and are typically NF1-associated.
- *Plexiform neurofibromas* can be found in deep or superficial locations in association with nerve roots or large nerves and are uniformly NF1-associated.

Pathogenesis

Only the Schwann cells in neurofibromas show complete loss of the *NF1* gene product, neurofibromin, indicating that these are the neoplastic cells. You will recall from Chapter 7 that neurofibromin is a tumor suppressor that inhibits RAS activity by stimulating the activity of a guanosine triphosphatase (GTPase). Loss of the GTPase activity causes RAS to be trapped in GTP bound active state. Haploinsufficiency for the *NF1* gene in other associated cells may also contribute to the growth of NF1-associated tumors. For example, there is evidence that *NF1*-haploinsufficient mast cells are hypersensitive to KIT ligand produced by Schwann cells and in response secrete factors that stimulate Schwann cell growth. This form of tumor/stromal cell cross-talk may be targetable with inhibitors of the KIT receptor tyrosine kinase. Other studies suggest that plexiform neurofibromas and dermal neurofibromas arise from different neural crest–derived precursor cells. Malignant transformation of neurofibroma to MPNST is usually observed in the plexiform variant but is sometimes also seen in the diffuse type. The overall incidence of MPNST in NF1 patients is about 5% to 10%, but patients with large numbers of plexiform neurofibromas and large deletions in the *NF1* gene are at higher risk.

MORPHOLOGY

Localized cutaneous neurofibroma. These are small, well-delineated but unencapsulated nodular lesions that arise in the dermis and subcutaneous fat. They have relatively low cellularity and contain bland Schwann cells admixed with stromal cells such as mast cells, perineurial cells, CD34+ spindle cells, and fibroblasts. Adnexal structures are sometimes entrapped at the edges of the lesion. The stroma of these tumors contains loose collagen.

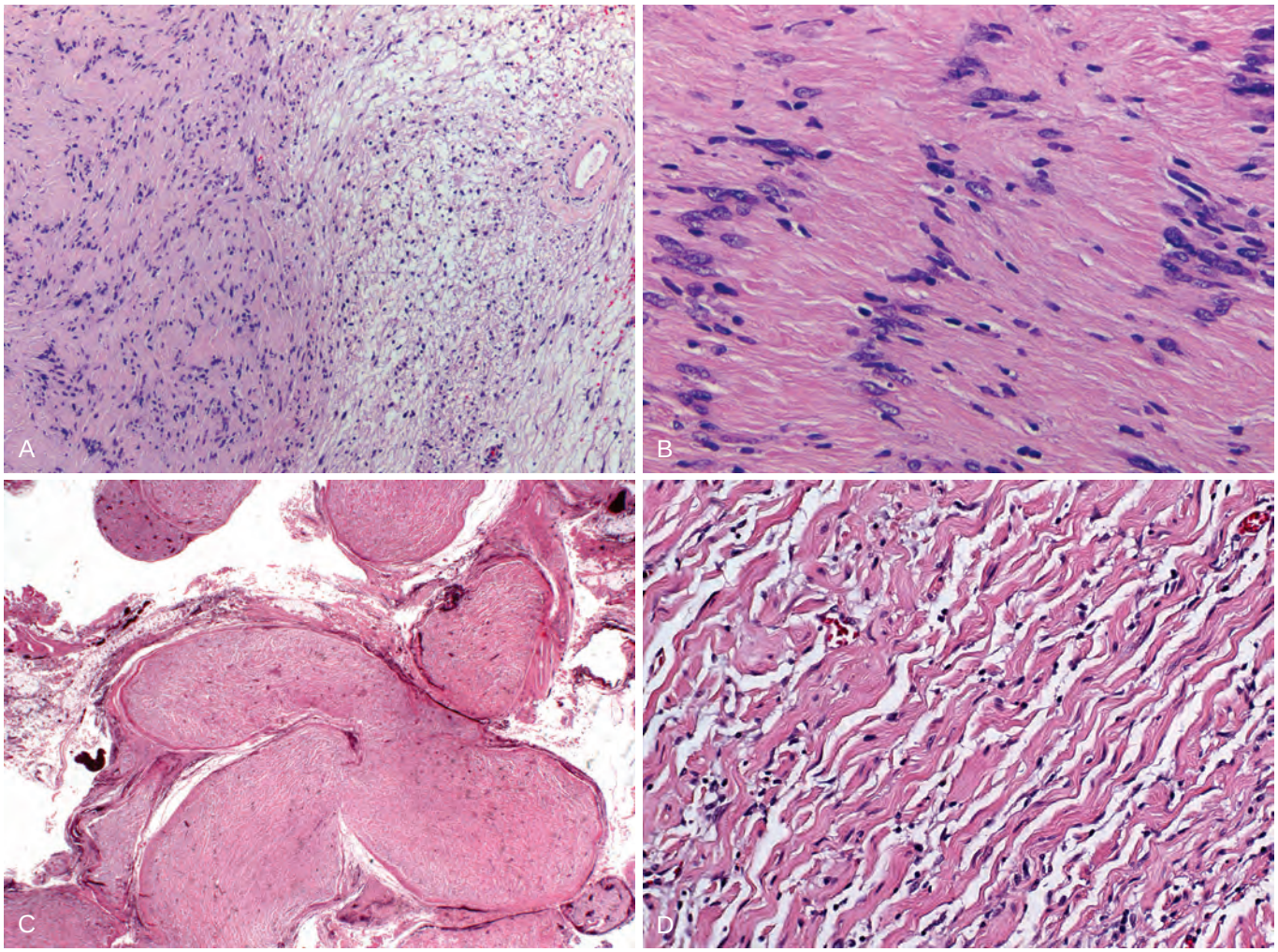


Figure 27.14 Schwannoma and plexiform neurofibroma. (A and B) Schwannoma. (A) Schwannomas often contain dense eosinophilic Antoni A areas *(left)* and loose, pale Antoni B areas *(right)*, as well as hyalinized blood vessels *(right)*. (B) Antoni A area with the tumor cell nuclei aligned in palisading rows leaving anuclear zones and resulting in the formation of structures termed *Verocay bodies*. (C and D) Plexiform neurofibroma. (C) Multiple nerve fascicles are expanded by infiltrating tumor cells. (D) At high power, bland spindle cells are admixed with wavy collagen bundles resembling carrot shavings.

Diffuse neurofibroma. This tumor has morphologic features similar to those seen in localized cutaneous neurofibromas but exhibits a distinctly different growth pattern. The tumor diffusely infiltrates the dermis and subcutaneous connective tissue, entrapping fat and appendage structures and producing a plaque-like appearance. Some of these neurofibromas can grow to large sizes. Focal collections of cells mimicking the appearance of Meissner corpuscles (so-called **pseudo-Meissner corpuscles** or **tactile-like bodies**) are an associated feature.

Plexiform neurofibroma. These tumors grow within and expand nerve fascicles (Fig. 27.14C), entrapping associated axons. The external perineurial layer of the nerve is preserved, giving individual nodules an encapsulated appearance. The expanded, ropy thickening of multiple nerve fascicles results in what is sometimes referred to as a "bag of worms" appearance. The tumor has cellular composition similar to that of other neurofibromas. The extracellular matrix varies from loose and myxoid to more collagenous and fibrous. Often the collagen is seen in bundles likened to "shredded carrot" (Fig. 27.14D).

Malignant Peripheral Nerve Sheath Tumors

Most MPNSTs (approximately 85%) are high-grade tumors, but low-grade variants are recognized. About half arise in NF1 patients and are assumed to result from malignant transformation of a plexiform neurofibroma. Sporadic cases may arise de novo. Most are associated with larger peripheral nerves in the chest, abdomen, pelvis, neck, or limb-girdle. MPNSTs exhibit complex chromosomal aberrations including chromosome gains, losses, and rearrangements. The molecular alterations driving malignant transformation of a neurofibroma to MPNST are still poorly understood.

MORPHOLOGY

The lesions are poorly defined tumor masses that frequently infiltrate along the axis of the parent nerve and invade adjacent soft tissues. A wide range of histologic appearance can be encountered. Typical cases show a fasciculated arrangement of

spindle cells. At low power the tumor often appears "marbleized" due to variations in cellularity. Mitoses, necrosis, and nuclear anaplasia are common. An interesting phenomenon observed in MPNST is described as "divergent differentiation." This term refers to the presence of focal areas that exhibit other lines of differentiation, including glandular, cartilaginous, osseous, or rhabdomyoblastic morphology. A tumor exhibiting the latter is referred to as **Triton tumor.** Due to the poorly differentiated nature of MPNST, the distinction from an undifferentiated sarcoma may not be straightforward. Helpful clues include a diagnosis of NF1 in the affected patient and a clearly demonstrated anatomic relationship to a nerve or to a preexisting neurofibroma.

Neurofibromatosis Type 1 and Type 2

Neurofibromatosis Type 1

This is a common autosomal dominant disorder with a frequency of 1 in 3000. It is a systemic disease associated with nonneoplastic manifestations and with a variety of tumors, including neurofibromas of all types, MPNSTs, gliomas of the optic nerve, other glial tumors and hamartomatous lesions, and pheochromocytomas. Other features include intellectual disability or seizures, skeletal defects, pigmented nodules of the iris *(Lisch nodules)*, and cutaneous hyperpigmented macules *(café au lait spots)*. The disease is caused by loss-of-function mutations in the *NF1* gene, located at 17q11.2, which encodes the tumor suppressor neurofibromin. The neoplastic cells in NF1-related tumors lack neurofibromin due to biallelic defects in the *NF1* gene. As has been mentioned earlier, NF-1 protein has GTPase activity that restrains RAS function. In the absence of NF-1, RAS remains trapped in its active state.

The disease has a high penetrance but variable expressivity. Some patients exhibit only subtle features, while others show disease that is restricted to certain parts of the body, a distribution that is attributable to mosaicism. An unfortunate subset has severe disease. Large chromosomal deletions that span *NF1* and extend to involve adjacent genes tend to be associated with more severe phenotypes.

Neurofibromatosis Type 2

This is an autosomal dominant disorder resulting in a range of tumors, most commonly bilateral eighth nerve schwannomas, multiple meningiomas, and ependymomas of the spinal cord. Many individuals with NF2 also have nonneoplastic lesions, which include nodular ingrowth of Schwann cells into the spinal cord (schwannosis), meningioangiomatosis (a proliferation of meningeal cells and blood vessels that grows into the brain), and glial hamartia (microscopic nodular collections of glial cells at abnormal locations, often in the superficial and deep layers of cerebral cortex). This disorder is much less common than NF1, having a frequency of 1 in 40,000 to 50,000. Certain other rare familial syndromes are also associated with multiple schwannomas, such as schwannomatosis and Carney complex.

The *NF2* gene is located on chromosome 22q12 and is also commonly mutated in sporadic meningiomas and schwannomas. The *NF2* gene product, merlin, is a cytoskeletal protein that participates in the regulation of several key signaling pathways that are involved in control of cell shape, cell growth, and the attachment of cells to one another (cell adhesion). There is some correlation between the type of mutation and clinical symptoms, with nonsense and frameshift mutations causing more severe phenotypes than missense mutations.

KEY CONCEPTS

PERIPHERAL NERVE SHEATH TUMORS

- The three common peripheral nerve sheath tumors—schwannoma, neurofibroma, and MPNST—all likely arise from cells of Schwann cell lineage.
- Schwannomas are encapsulated benign tumors that can be associated with NF2.
- Neurofibromas are benign peripheral nerve sheath tumors sometimes associated with NF1 that can be subtyped as localized cutaneous, diffuse, or plexiform.
- MPNSTs can be de novo sporadic neoplasms or NF1-associated tumors arising through malignant transformation of a (plexiform) neurofibroma.

SUGGESTED READINGS

Neuropathies and Other Non-Neoplastic Disorders of Peripheral Nerves

Barohn RJ, Dimachkie MM: Peripheral neuropathies, *Neurol Clin* 31:343, 2013.

Callaghan BC, Price RS, Chen KS et al: The importance of rare subtypes in diagnosis and treatment of peripheral neuropathy: a review, *JAMA Neurol* 72:2015, 1510.

Collins MP, Hadden RD: The nonsystemic vasculitic neuropathies, *Nat Rev Neurol* 13:301, 2017.

Duchesne M, Mathis S, Richard L et al: Nerve biopsy is still useful in some inherited neuropathies, *J Neuropathol Exp Neurol* 77:88, 2018.

Jennette JC, Falk RJ, Bacon PA et al: 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides, *Arthritis Rheum* 65:1, 2013.

Mathey EK, Park SB, Hughes RAC et al: Chronic inflammatory demyelinating polyradiculoneuropathy: from pathology to phenotype, *J Neurol Neurosurg Psychiatry* 86:973, 2015.

Mathis S, Goizet C, Tazir M et al: Charcot-Marie-Tooth diseases: an update and some new proposals for the classification, *J Med Genet* 52:681, 2015.

Schoser B, Eymard B, Datt J et al: Lambert-Eaton myasthenic syndrome (LEMS): a rare autoimmune presynaptic disorder often associated with cancer, *J Neurol* 264:2017, 1854.

van Schaik IN, van Geloven VB, Hartung H-P et al: Subcutaneous immunoglobulin for maintenance treatment in chronic inflammatory demyelinating polyneuropathy (PATH): a randomized, double-blind, placebo-controlled, phase 3 trial, *Lancet Neurol* 17:35, 2018.

Willison HJ, Jacobs BC, van Doorn PA: Guillain-Barré syndrome, *Lancet* 388:717, 2016.

Muscular Dystrophies and Other Disorders of Muscle

Allenbach Y, Benveniste O, Goebel HH et al: Integrated classification of inflammatory myopathies, *Neuropathol Appl Neurobiol* 43(1):62–81, 2017. *[Review of the histomorphologic classification of inflammatory myopathies].*

Amato AA, Griggs RC: Unicorns, dragons, polymyositis, and other mythological beasts, *Neurology* 61(3):288–289, 2003 *[Opinion paper about the declining importance of polymyositis].*

Bello L, Pegoraro E: Genetic diagnosis as a tool for personalized treatment of Duchenne muscular dystrophy, *Acta Myol* 35(3):122–127, 2016. *[Description of novel treatment strategies for Duchenne muscular dystrophy and implication for diagnostic approach].*

Bönnemann CG, Wang CH, Quijano-Roy S et al: Diagnostic approach to the congenital muscular dystrophies, *Neuromuscul Disord* 24(4):289–311, 2014. *[Committee communication on classification and diagnosis of congenital muscular dystrophies].*

Dalakas MC: Inflammatory muscle diseases, *N Engl J Med* 372(18): 1734–1747, 2015. *[Review on classification and clinical features of inflammatory myopathies].*

Daxinger L, Tapscott SJ, van der Maarel SM: Genetic and epigenetic contributors to FSHD, *Curr Opin Genet Dev* 33:56–61, 2015. *[Review of the biologic mechanisms of facioscapulohumeral dystrophy].*

Dubowitz V, Sewry CA, Oldfors A: *Muscle biopsy: a practical approach*, ed 4, Oxford, 2013, Elsevier. *[Overview of disorders affecting muscle].*

Nigro V, Savarese M: Genetic basis of limb-girdle muscular dystrophies: the 2014 update, *Acta Myol* 33(1):1–12, 2014. *[Summary of the different categories of limb-girdle muscular dystrophies].*

North KN, Wang CH, Clarke N et al: Approach to the diagnosis of congenital myopathies, *Neuromuscul Disord* 24(2):97–116, 2014. *[Committee communication on classification and diagnosis of congenital myopathies].*

Thornton CA, Wang E, Carrell EM: Myotonic dystrophy: approach to therapy, *Curr Opin Genet Dev* 44:135–140, 2017. *[Review of possible treatment strategies for myotonic dystrophy].*

Thornton CA: Myotonic dystrophy, *Neurol Clin* 32(3):705–719, 2014. *[Review of myotonic dystrophy].*

Tumors of Peripheral Nerves

Rodriguez FJ, Folpe AL, Giannini C et al: Pathology of peripheral nerve sheath tumors: diagnostic overview and update on selected diagnostic problems, *Acta Neuropathol* 123:295, 2012. *[Review of the pathologic classification and features of peripheral nerve sheath tumors].*

Schaefer IM, Fletcher CDM: Recent advances in the diagnosis of soft tissue tumours, *Pathology* 50:37, 2018. *[Review of the molecular changes in soft tissue tumors, including nerve sheath tumors].*

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 28

The Central Nervous System

Marta Margeta • Arie Perry

CHAPTER CONTENTS

The principal functional unit of the central nervous system (CNS) is the *neuron*. Neurons of different types and in different locations have distinct properties, including diverse functional roles, patterns of synaptic connections, neurotransmitters used, and metabolic requirements, which vary with electrical activity. A set of neurons, not necessarily clustered together in a region of the brain, may thus show *selective vulnerability* to particular insults because they share one or more of these properties. Because brain functions are anatomically compartmentalized, the pattern of clinical signs and symptoms that follow injury depend as much on the region of brain that is affected as on the pathologic process. Mature neurons are incapable of cell division, so destruction of even a small number of neurons essential for a specific function may produce a neurologic deficit.

In addition to neurons, the CNS contains other cells, such as *astrocytes* and *oligodendrocytes*, which make up the *glia*. Different cellular components of the CNS are affected by distinct unique neurologic disorders and also respond to common insults (e.g., ischemia, infection) in a manner that is distinct from other tissues. We start our discussion of diseases of the CNS with an overview of the patterns of injury of different cells and the reactions of these cells to various insults.

CELLULAR PATHOLOGY OF THE CENTRAL NERVOUS SYSTEM

Neurons and glia of the CNS display a range of functional and morphologic changes following injury. Understanding these patterns can provide information about the mechanism of cellular injury and type of disease.

Reactions of Neurons to Injury

Neuronal injury may be an acute process, often as a consequence of depletion of oxygen or glucose or trauma, or may develop over a period of years, as occurs in degenerative disorders of the brain. Neurons are highly metabolically active and require a continuous supply of oxygen and glucose to meet their metabolic needs. Moreover, since neurons are postmitotic cells that are incapable of proliferation, they must be maintained throughout life. Perhaps because of their lengthy lifespan and high metabolic activity, neurons are unusually susceptible to the accumulation of misfolded proteins, which trigger the unfolded protein response (Chapter 2) and appear to be central to the pathogenesis of many degenerative disorders of the CNS.

MORPHOLOGY

Acute neuronal injury ("red neurons") refers to changes seen following acute CNS hypoxia/ischemia, severe hypoglycemia, and other acute insults, and are the earliest morphologic markers of neuronal cell death (see Fig. 28.15A). "Red neurons" are evident by 6 to 12 hours after an irreversible hypoxic/ischemic insult. Typical features include shrinkage of the cell body, pyknosis of the nucleus, disappearance of the nucleolus, loss of Nissl substance, and intense cytoplasmic eosinophilia.

Subacute and chronic neuronal injury ("degeneration") refers to neuronal death that occurs as a result of a progressive disease of months to years duration, as is seen in slowly evolving neurodegenerative diseases such as amyotrophic lateral sclerosis and Alzheimer disease. Prior to death, neurons often suffer a loss of synapses (sites of interneuronal communication), which may stem from aberrations in synaptic pruning (a process responsible for normal brain development and plasticity). Loss of synapses is followed by cell death (often selectively involving functionally related groups of neurons) and reactive gliosis. At an early stage, the cell loss is difficult to appreciate; the associated reactive glial changes are often the best indicator of neuronal injury. In many of these diseases, the predominant mechanism of cell death appears to be apoptosis.

Axonal reaction is a change observed in the cell body during regeneration of the axon; it is best seen in anterior horn cells of the spinal cord when motor axons are cut or seriously damaged. The increase in protein synthesis that occurs in response to the injury is reflected in enlargement and rounding up of the cell body, peripheral displacement of the nucleus, enlargement of the nucleolus, and dispersion of Nissl substance from the center to the periphery of the cell (central chromatolysis).

Neuronal damage may be associated with a wide range of subcellular alterations in the neuronal organelles and cytoskeleton. **Neuronal inclusions** may occur as a manifestation of aging, and consist of intracytoplasmic accumulations of complex lipids (lipofuscin), proteins, or carbohydrates. Abnormal cytoplasmic accumulations of lipids and other substances also occur with certain inborn errors of metabolism, genetic disorders caused by mutations that lead to the loss of specific enzyme activities (Chapter 5). Viral infection can lead to the appearance of viral inclusions, e.g., in herpetic infection (Cowdry A or B bodies), rabies (Negri body), and cytomegalovirus infection.

Some degenerative diseases of the CNS are associated with neuronal intracytoplasmic inclusions, such as neurofibrillary tangles of Alzheimer disease and Lewy bodies of Parkinson disease; others cause abnormal vacuolization of the perikaryon and neuronal cell processes in the neuropil (Creutzfeldt-Jakob disease).

Wallerian degeneration refers to degeneration of axons after disruption of nerve fibers (see Chapter 27).

Reactions of Astrocytes to Injury

Gliosis is the most important histopathologic marker of CNS injury, regardless of etiology, and is characterized by hypertrophy and hyperplasia of astrocytes. The astrocyte derives its name from its star-shaped appearance. These cells have multipolar, extensively branched cytoplasmic processes that emanate from the cell body and express glial fibrillary acidic protein (GFAP), a cell type–specific intermediate filament. Astrocytes act as metabolic buffers and detoxifiers within the brain. Additionally, through the foot processes (which surround capillaries or extend to the subpial and subependymal zones), they contribute to barrier functions by controlling the flow of macromolecules between the blood, the cerebrospinal fluid (CSF), and the brain.

MORPHOLOGY

In gliosis, the nuclei of astrocytes (which are typically oval with evenly dispersed, pale chromatin) enlarge, become vesicular, and

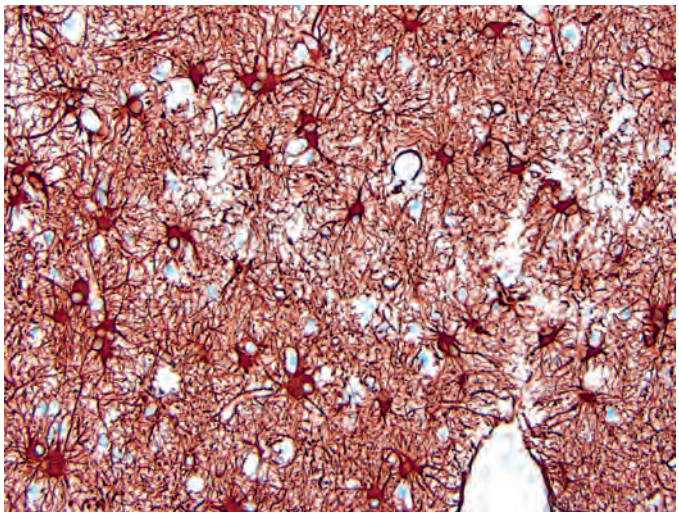


Figure 28.1 Reactive astrocytes. Immunohistochemical staining for glial fibrillary acidic protein *(brown)* highlights the stellate astrocytic processes from which the cell's name is derived.

occasionally develop prominent nucleoli. The previously indistinct cytoplasm becomes bright pink due to increased expression of GFAP, and the cells develop numerous stout, ramifying processes; these cells are called **reactive or gemistocytic astrocytes** (Fig. 28.1). There are at least two subtypes of reactive astrocytes; although morphologically indistinguishable, these cells form in response to different injurious stimuli, have distinct gene expression patterns, and have different effects on neuronal health and survival, with one subtype promoting CNS injury and the other contributing to CNS repair.

Acute astrocytic injury, as occurs in hypoxia, hypoglycemia, and toxic injuries, is manifested by cellular swelling, as in other cells (Chapter 2). The **Alzheimer type II astrocyte** (unrelated to the disease but first described by the same neuroscientist) is a gray matter cell with a large (two to three times normal) nucleus, pale-staining central chromatin, an intranuclear glycogen droplet, and a prominent nuclear membrane and nucleolus. Alzheimer type II astrocytes are mainly seen in individuals with hyperammonemia due to chronic liver disease, Wilson disease, or hereditary metabolic disorders of the urea cycle.

Other types of astrocyte injury lead to the formation of cytoplasmic inclusion bodies. **Rosenthal fibers** are thick, elongated, brightly eosinophilic, irregular structures that occur within astrocytic processes and contain two heat-shock proteins (αB-crystallin and hsp27) as well as ubiquitin. Rosenthal fibers are typically found in regions of long-standing gliosis; they are also characteristic of one type of glial tumor, pilocytic astrocytoma (see Fig. 28.49C). In Alexander disease, a leukodystrophy associated with mutations in the gene encoding GFAP, abundant Rosenthal fibers are found in periventricular, perivascular, and subpial locations. More commonly seen are **corpora amylacea,** or polyglucosan bodies. These are round, faintly basophilic, periodic acid-Schiff (PAS)–positive, concentrically lamellated structures of 5 to 50 μm in diameter that are located wherever there are astrocytic end processes, especially in the subpial and perivascular zones. They consist primarily of glycosaminoglycan polymers, as well as heat-shock proteins and ubiquitin. Corpora amylacea occur in increasing numbers with advancing age and are thought to represent a degenerative change in the astrocyte. **Lafora bodies**, which are seen in the cytoplasm of neurons (as well as hepatocytes, myocytes, and other cells) in a variant of myoclonic epilepsy, have a similar structure and biochemical composition.

Reactions of Microglia to Injury

Microglia are phagocytic cells derived early in embryonic development from the yolk sac or fetal liver that serve as the resident macrophages of the CNS and share many surface markers with bone marrow–derived peripheral monocytes/macrophages. At rest, microglia are tiled (i.e., cover non-overlapping territories) and have highly branched, amoeboid processes. During development, microglia prune unused synaptic connections, most likely through a phagocytic process that may involve the complement system; aberrant reactivation of this developmental process has recently been implicated in a number of different brain diseases, including schizophrenia, encephalitis, Alzheimer disease, and frontotemporal dementia. Microglia respond to injury by (1) proliferating; (2) developing elongated nuclei; (3) forming aggregates around small foci of tissue necrosis *(microglial nodules);* or (4) congregating around cell bodies of dying neurons *(neuronophagia)*. In addition to resident microglia, blood-derived macrophages may also be present in inflammatory foci.

Reactions of Other Glial Cells to Injury

Oligodendrocytes are cells that wrap their cytoplasmic processes around axons and form myelin. Each oligodendrocyte myelinates numerous internodes on multiple axons, in contrast to the myelinating Schwann cell in peripheral nerves, which only myelinates the internode of a single axon. Injury or apoptosis of oligodendrocytes is a feature of acquired demyelinating disorders and leukodystrophies. Oligodendroglial nuclei may harbor viral inclusions in progressive multifocal leukoencephalopathy (PML). Glial cytoplasmic inclusions, primarily composed of α-synuclein, are found in oligodendrocytes in multiple system atrophy (MSA).

Ependymal cells, the ciliated columnar epithelial cells lining the ventricles, do not show specific reaction patterns. When there is inflammation or marked dilation of the ventricular system, disruption of the ependymal lining is paired with proliferation of subependymal astrocytes to produce small irregularities on the ventricular surfaces *(ependymal granulations)*. Certain infectious agents, particularly CMV, may produce extensive ependymal injury, with viral inclusions in ependymal cells. However, neither oligodendrocytes nor ependymal cells mediate significant responses to most forms of injury in the CNS.

KEY CONCEPTS

CELLULAR PATHOLOGY OF THE CENTRAL NERVOUS SYSTEM

- Each cellular component of the nervous system responds to injury in a distinctive way.

- Neuronal injury commonly results in cell death, either by apoptosis or necrosis. Loss of neurons or synapses is often difficult to detect morphologically, but is a major contributor to neurologic dysfunction.
- Astrocytes respond to injury through apparent hypertrophy (increased cytoplasm due to accumulation of the intermediate filament protein GFAP) and hyperplasia; this response can be beneficial or harmful, depending on the type of injurious stimulus and the exact nature of astrocytic response.
- Microglia, the resident monocyte-lineage population of the CNS, proliferate and accumulate in response to injury. Aberrant activation of microglia- and complement-mediated synaptic pruning plays an important role in the pathogenesis of many different neurologic disorders.

CEREBRAL EDEMA, HYDROCEPHALUS, RAISED INTRACRANIAL PRESSURE, AND HERNIATION

The brain and the spinal cord are encased and protected by the rigid skull, dural reflections, and the bony spinal canal. The pressure within the cranial cavity may rise in any one of three commonly observed clinical settings: generalized brain edema, increased CSF volume, and focally expanding mass lesions. Depending on the degree and rapidity of the pressure increase and the nature of the underlying lesion, the consequences range from subtle neurologic deficits to death.

Cerebral Edema

Cerebral edema (more precisely, brain parenchymal edema) is the result of increased fluid leakage from blood vessels and injury to various cells of the CNS. There are two main pathways of edema formation in the brain.

- *Vasogenic edema* is an increase in extracellular fluid caused by blood-brain barrier disruption and increased vascular permeability, allowing fluid to shift from the intravascular compartment to the intercellular spaces of the brain. The paucity of lymphatics greatly impairs the resorption of excess extracellular fluid. Vasogenic edema may be either localized, e.g., adjacent to inflammation or neoplasms, or generalized, as may occur following a global ischemic injury.
- *Cytotoxic edema* is an increase in intracellular fluid secondary to neuronal, glial, or endothelial cell membrane injury, as might be encountered with a generalized hypoxic/ischemic insult or with a metabolic derangement that prevents maintenance of the normal membrane ionic gradients.

In practice, conditions associated with generalized edema often have elements of both vasogenic and cytotoxic edema. In generalized edema, the gyri are flattened, the intervening sulci are narrowed, and the ventricular cavities are compressed. As the brain expands, herniation may occur.

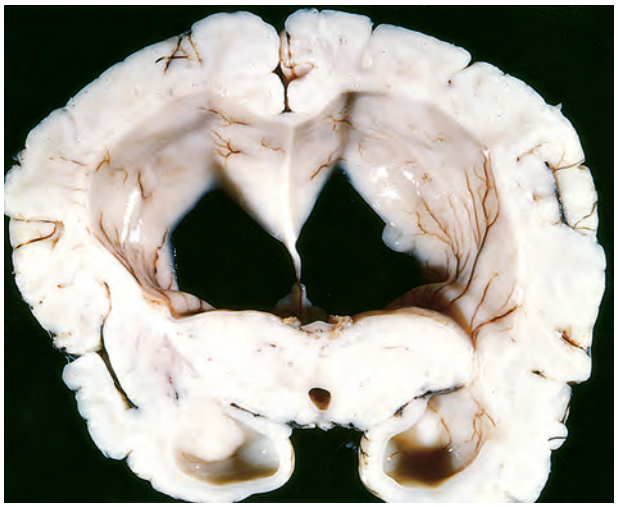


Figure 28.2 Hydrocephalus. Dilated lateral ventricles seen in a coronal section through the midthalamus.

Hydrocephalus

Hydrocephalus is the accumulation of excessive CSF within the ventricular system (Fig. 28.2). The choroid plexus within the ventricular system produces CSF, which normally circulates through the ventricular system and enters the cisterna magna at the base of the brainstem through the foramina of Luschka and Magendie. Subarachnoid CSF bathes the superior cerebral convexities and is absorbed by the arachnoid granulations. Most cases of hydrocephalus are a consequence of impaired flow and resorption of CSF; overproduction is a rare cause that can accompany tumors of the choroid plexus. In all forms, the increased volume of CSF expands the ventricles and can elevate the intracranial pressure.

When hydrocephalus develops in infancy before closure of the cranial sutures, the head enlarges. By contrast, after the sutures close, hydrocephalus is associated with expansion of the ventricles and increased intracranial pressure, without a change in head circumference. If the ventricular system is only focally obstructed, due to a mass in the third ventricle or to aqueductal stenosis, it is called *noncommunicating (obstructive) hydrocephalus*. In *communicating hydrocephalus,* the ventricular system remains in continuity with the subarachnoid space and there is enlargement of the entire ventricular system; causes include overproduction of CSF from a choroid plexus tumor and arachnoid fibrosis following meningitis. The term *hydrocephalus ex vacuo* refers to a compensatory increase in ventricular volume secondary to a loss of brain parenchyma.

Raised Intracranial Pressure and Herniation

Herniation is the displacement of brain tissue past rigid dural folds (the falx and tentorium) or through openings in the skull because of increased intracranial pressure. As the volume of the brain increases, CSF is displaced, leading to increasing pressure within the cranial cavity; herniation occurs when this pressure exceeds the brain's limited capacity to accommodate the increased intracranial pressure. Brain herniation is mostly caused by mass effects, either diffuse

(generalized brain edema) or focal (tumors, abscesses, or hemorrhages). Elevated intracranial pressure may also compress the vasculature and reduce perfusion of the brain, causing ischemic injury and further exacerbating cerebral edema.

MORPHOLOGY

The brain may herniate through different openings, and if the expansion is sufficiently severe, herniation may occur simultaneously in several locations (Fig. 28.3):

- **Subfalcine (cingulate) herniation** occurs when unilateral or asymmetric expansion of a cerebral hemisphere displaces the cingulate gyrus under the falx. This may lead to compression of the anterior cerebral artery and its branches, resulting in secondary infarcts.
- **Transtentorial (uncal, mesial temporal) herniation** occurs when the medial aspect of the temporal lobe is compressed against the free margin of the tentorium. With increasing displacement of the temporal lobe, the third cranial nerve is compromised, resulting in pupillary dilation and impaired ocular movements on the side of the lesion. The posterior cerebral artery may also be compressed, resulting in an infarct of its territory (which includes the primary visual cortex). When the extent of herniation is large enough, the contralateral cerebral peduncle may be compressed, resulting in hemiparesis ipsilateral to the side of the herniation. Progression of transtentorial herniation is often accompanied by secondary hemorrhagic lesions in the midbrain and pons, termed **Duret hemorrhages** (Fig. 28.4). These linear or flame-shaped lesions usually occur in the midline and paramedian regions and are believed to be due to distortion or tearing of penetrating veins and arteries supplying the upper brainstem.
- **Tonsillar herniation** refers to displacement of the cerebellar tonsils through the foramen magnum. This pattern of herniation is life-threatening because it causes brainstem compression and compromises vital respiratory and cardiac centers in the medulla.

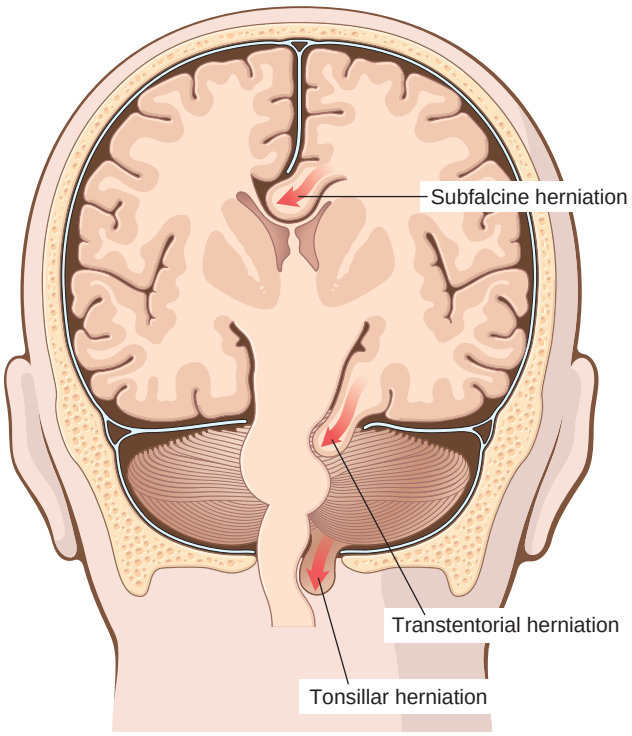


Figure 28.3 Major herniation syndromes of the brain: subfalcine, transtentorial, and tonsillar.

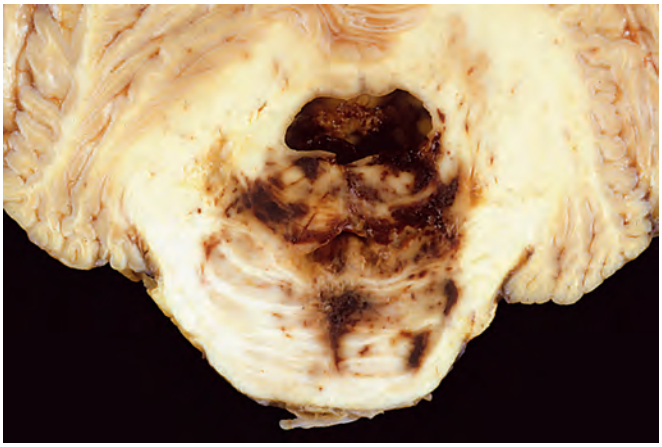


Figure 28.4 Duret hemorrhage. As mass effect displaces the brain downward, there is disruption of the vessels that enter the pons along the midline, leading to hemorrhage.

KEY CONCEPTS

CEREBRAL EDEMA, HYDROCEPHALUS, RAISED INTRACRANIAL PRESSURE, AND HERNIATION

- Cerebral edema is the accumulation of excess fluid within the brain parenchyma. Hydrocephalus is an increase in CSF volume within all or part of the ventricular system.
- Increase in the intracranial volume (due to increased CSF volume, parenchymal edema, hemorrhage, or tumor) raises intracranial pressure.
- Increased intracranial pressure may result in cerebral herniation, decreased perfusion, and secondary infarction of the affected areas.

MALFORMATIONS AND DEVELOPMENTAL DISORDERS

Although the pathogenesis and etiology of many CNS malformations remain unknown, both genetic and environmental influences appear to be involved. Genomic sequencing has begun to uncover a range of genetic variants that are associated with certain malformations. The causal relationship between these genetic alterations and the pathogenesis of the malformations is the subject of active research. Besides genetic factors, many toxic compounds and infectious agents also have teratogenic effects and may cause brain malformations.

Neural Tube Defects

Neural tube defects are midline malformations that involve some combination of neural tissue, meninges, and overlying

bone or soft tissue; collectively, they are the most common CNS malformations. Two distinct pathogenic mechanisms are contributory: (1) failure of neural tube closure, in which secondary mesenchymal tissue defects stem from aberrant skeletal modeling around the malformed tube (e.g., anencephaly and myelomeningocele); and (2) primary bony defects that are caused by abnormal axial mesoderm development and lead to secondary CNS abnormalities (e.g., encephalocele, meningocele, and spina bifida).

MORPHOLOGY

- **Anencephaly** is a malformation of the anterior end of the neural tube that leads to absence of most of the brain and calvarium. Forebrain development is disrupted at approximately 28 days of gestation, and all that remains in its place is the **area cerebrovasculosa,** a flattened remnant of disorganized brain tissue with admixed ependyma, choroid plexus, and meningothelial cells. The posterior fossa structures may be spared, depending on the extent of the skull deficit; descending tracts associated with disrupted structures are, as expected, absent.
- **Myelomeningocele** (or meningomyelocele) refers to extension of CNS tissue through a defect in the vertebral column; the term **meningocele** applies when there is only a meningeal extrusion. Myelomeningoceles occur most commonly in the lumbosacral region (Fig. 28.5). Affected individuals have motor and sensory deficits in the lower extremities as well as disturbances of bowel and bladder control. They are often complicated by superimposed infection of the cord due to defective barrier function of the thin, overlying skin.

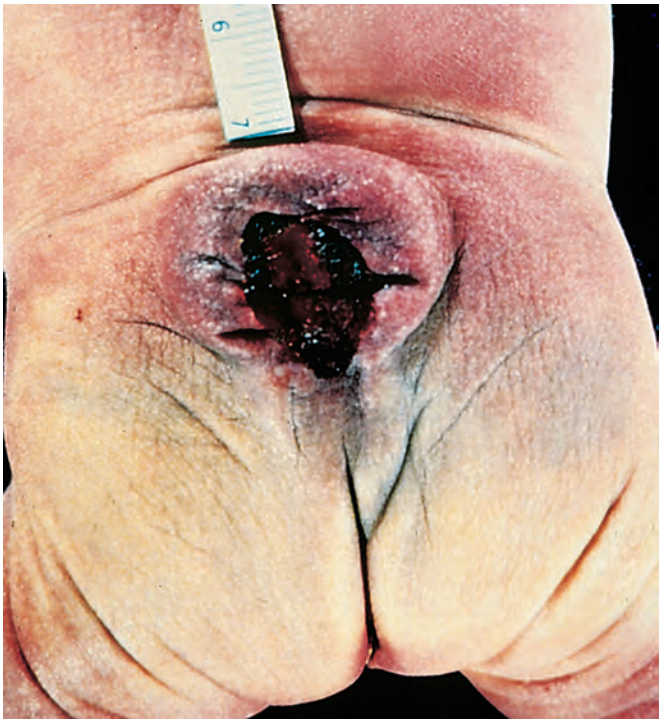


Figure 28.5 Myelomeningocele. Both meninges and spinal cord parenchyma are included in the cystlike structure visible just above the buttocks.

- **Encephalocele** refers to an extrusion of malformed brain tissue through a midline defect in the cranium. It most often occurs in the occiput, although nasofrontal variants involving the orbit, ethmoid, or cribriform plate (sometimes misleadingly referred to as a "nasal glioma") also are seen.
- **Spinal dysraphism** or **spina bifida** (the most common neural tube defects) may be an asymptomatic bony defect (spina bifida occulta) or a severe malformation with a flattened, disorganized segment of spinal cord, associated with an overlying meningeal outpouching.

Clinical Features

The frequency of neural tube defects varies widely among different ethnic groups, with the overall recurrence rate for a neural tube defect in subsequent pregnancies estimated at 4% to 5%. Folate deficiency during the first several weeks of gestation is a well-established risk factor; differences in rates of neural tube defects among populations can be attributed in part to polymorphisms in enzymes involved in folic acid metabolism. Folate supplementation can lower the risk of neural tube defects, but because neural tube closure is normally complete by day 28 of embryonic development (before most pregnancies are recognized), it must be given to women throughout their reproductive years to be fully effective. Precisely how folate deficiency increases the risk is uncertain; effects on DNA methylation (an important epigenetic mode of gene regulation) are suspected.

Forebrain Anomalies

Abnormalities in the generation and migration of neurons result in malformations of the forebrain that may be focal or involve entire structures. The pool of proliferating precursor cells in the developing brain lies in the germinal matrix adjacent to the ventricular system; the total number of neurons is determined by the fraction of proliferating cells that undergo transition into migrating cells with each cell cycle. The migration of neurons from the germinal matrix to the cerebral cortex follows two paths: radial migration, for progenitor cells destined to become excitatory neurons; and tangential migration, for those that will become inhibitory interneurons.

A range of different anatomic malformation patterns has been defined. Recently, it has become clear that many of these patterns are caused by mutations in genes that are required for proper cerebral development. Changes may be seen in the complexity of the brain surface (either too few or too many gyri present), the organization of the brain into normal lobes, the structure of the cerebral cortex, or the distribution of neurons within the brain.

- The volume of brain may be abnormally large *(megalencephaly)* or abnormally small *(microencephaly)*. Microencephaly, by far the more common of the two, is typically accompanied by a small head circumference. It is associated with a number of conditions, including chromosome abnormalities, fetal alcohol syndrome, and viral infection acquired in utero (e.g., with human immunodeficiency virus 1 [HIV-1] or Zika virus). It is postulated that the underlying anomaly is a reduction in the number of

neurons that reach the neocortex, which leads to a simplification of gyral folding, a mechanism supported by experimental results in mouse models.

- *Lissencephaly* is a malformation characterized by reduction in the number of gyri, which in the extreme case may show no gyral pattern *(agyria)*. Two general patterns are observed, a smooth-surface form (type 1), and a rough- or cobblestone-surface form (type 2). In general, type 1 forms are associated with mutations that disrupt mechanisms involved in cell migration, such as mutations in the cytoskeletal "motor" proteins that drive migration of neuroblasts. In contrast, type 2 lissencephaly is most commonly associated with genetic alterations that disrupt the "stop signal" for migration. This signal depends on a set of specifically glycosylated proteins, and mutations in the enzymes that place sugars onto these proteins are the most common causes of this form of lissencephaly.
- *Polymicrogyria* is characterized by numerous small, irregularly formed cerebral convolutions with shallow sulci. The cerebral cortex is composed of four or fewer layers (instead of the normal six layers), with fusion of the molecular layers between gyri. Polymicrogyria can be induced by localized tissue injury toward the end of neuronal migration, although genetically determined forms, which are typically bilateral and symmetric, are also recognized.
- *Neuronal heterotopias* are a group of migrational disorders defined by collections of subcortical neurons in inappropriate locations along the pathway of migration. As might be expected, one location is along the ventricular surface—as though the cells never left their place of birth. Periventricular nodular heterotopias can be caused by mutations in the gene encoding filamin A, an actin-binding protein responsible for assembly of complex meshworks of filaments. This gene is on the X chromosome, and the mutant allele causes male lethality; in females, the process of X inactivation separates neurons into those with a normal allele (in the correct location) and those with the mutant allele (in the heterotopia). Another microtubule-associated protein, doublecortin (DCX), is also encoded by a gene on the X chromosome; mutations in this gene result in lissencephaly in males and in subcortical bandlike heterotopias in females (this parallel layer of gray matter imparts the impression of a "double cortex"). Nodular subcortical heterotopias may also be encountered.
- *Holoprosencephaly* is a spectrum of malformations characterized by incomplete separation of the cerebral hemispheres across the midline. Severe forms manifest midline facial abnormalities, including cyclopia; less severe variants *(arrhinencephaly)* show absence of the olfactory cranial nerves and related structures. Intrauterine diagnosis of severe forms by ultrasonography is now possible. Holoprosencephaly is associated with trisomy 13 as well as other genetic syndromes. Mutations in genes that encode components of the sonic hedgehog signaling pathway may also produce holoprosencephaly.
- *Agenesis of the corpus callosum*, a relatively common malformation, is the absence of the white matter bundles that carry cortical projections from one hemisphere to the other (Fig. 28.6). Radiologic imaging studies show misshapen lateral ventricles ("bat-wing" deformity); on coronal whole-mount sections of the brain, "Probst bundles" of anteroposteriorly oriented white matter can be demonstrated. Agenesis of the corpus callosum is sometimes associated with intellectual disability but may also be found in normal individuals. It may be sporadic or familial and can be present in isolation or in association with a range of other malformations.

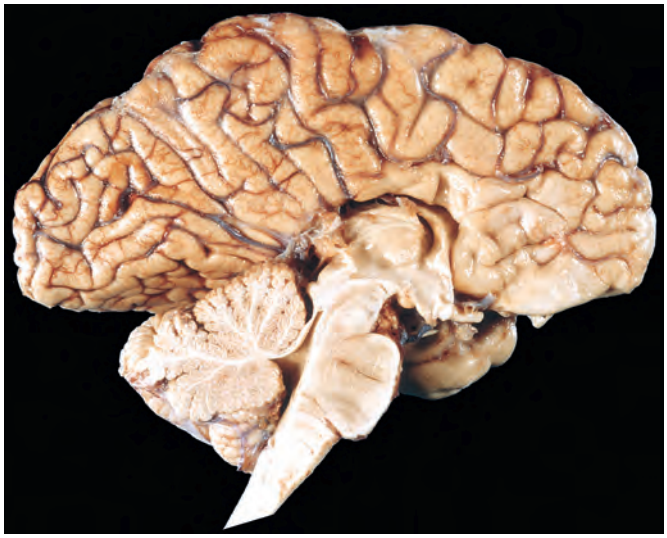


Figure 28.6 Agenesis of the corpus callosum. The midsagittal view of the left hemisphere shows the lack of a corpus callosum and cingulate gyrus.

Posterior Fossa Anomalies

A distinct set of malformations primarily affect the brainstem and the cerebellum, which often show dramatic changes in size and shape. These may be accompanied by morphologic changes in other regions of the brain.

MORPHOLOGY

- **Arnold-Chiari malformation** (Chiari type II malformation) consists of a small posterior fossa, a misshapen midline cerebellum with downward extension of vermis through the foramen magnum (Fig. 28.7), and, almost invariably, hydrocephalus and a lumbar myelomeningocele. Other associated changes may include caudal displacement of the medulla, malformation of the tectum, aqueductal stenosis, cerebral heterotopias, and hydromyelia (see later).
- **Chiari type I malformation** is a less severe disorder in which low-lying cerebellar tonsils extend down into the vertebral canal. It may be a silent abnormality or may become symptomatic because of impaired CSF flow and medullary compression; if present, these symptoms can usually be corrected by neurosurgical intervention.
- **Dandy-Walker malformation** is characterized by an enlarged posterior fossa. The cerebellar vermis is absent or present only in rudimentary form in its anterior portion; in its place is a large midline cyst that is lined by ependyma and is contiguous with leptomeninges on its outer surface. This cyst represents the expanded, roofless fourth ventricle in the absence of a normally formed vermis. Dysplasias of brainstem nuclei are commonly found in association with Dandy-Walker malformation.

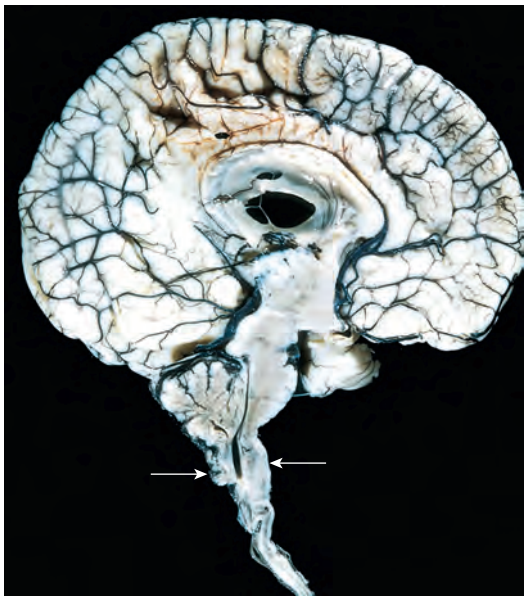


Figure 28.7 Arnold-Chiari malformation. Midsagittal section showing small posterior fossa contents, downward displacement of the cerebellar vermis, and deformity of the medulla (*arrows* indicate the approximate level of the foramen magnum).

- **Joubert syndrome** and related disorders share hypoplasia of the cerebellar vermis with apparent elongation of the superior cerebellar peduncles and an altered shape of the brainstem; together these changes give rise to the "molar tooth sign" on imaging. This group of malformations has been found to be caused by diverse mutations affecting genes that encode components of the primary (nonmotile) cilium.

Syringomyelia and Hydromyelia

These disorders are characterized by expansion of the ependyma-lined central canal of the cord *(hydromyelia)* or by the formation of a fluid-filled cleftlike cavity in the inner portion of the cord *(syringomyelia, syrinx)* that may extend into the brainstem *(syringobulbia)*.

Syringomyelia may be associated with a Chiari malformation; it may also occur in association with intraspinal tumors or following traumatic injury. In general, the histologic appearance is similar in all of these conditions, with destruction of the adjacent gray and white matter, surrounded by dense reactive gliosis. The disease generally becomes manifest in the second or third decade of life. The distinctive symptoms and signs of a syrinx are the isolated loss of pain and temperature sensation in the upper extremities because of disruption of the crossing anterior spinal commissural fibers of the spinal cord.

KEY CONCEPTS

MALFORMATIONS AND DEVELOPMENTAL DISORDERS

- Malformations may be associated with single-gene mutations, larger-scale genetic alterations, or exogenous factors.
- Overall, the earlier in development a malformation occurs, the more severe the morphologic and functional phenotype.
- Neural tube defects are caused by the failure of the tube to close or by skeletal bony defects that lead to secondary tube abnormalities; they range from incidental findings to severe malformations.
- Cortical development depends on proper orchestration of progenitor cell proliferation in the germinal matrix and migration of progenitors upward into the developing cortex. Disruption of these processes can alter the size, shape, and organization of the brain.
- Malformations involving the posterior fossa are typically distinct from those that affect the cerebral hemispheres.

PERINATAL BRAIN INJURY

Brain injury occurring in the perinatal period is an important cause of childhood-onset neurologic disability. Injuries that occur early in gestation may destroy brain tissue without eliciting the reactive changes observed in adult brain and, therefore, may be difficult to distinguish from malformations. Different patterns of injury may occur.

- The term *cerebral palsy* refers to a nonprogressive neurologic motor deficit characterized by combinations of spasticity, dystonia, ataxia/athetosis, and paresis, attributable to brain injury occurring during the prenatal and perinatal periods. Signs and symptoms may not be apparent at birth and only declare themselves later, as development proceeds. Postmortem examinations of children with cerebral palsy have shown a wide range of neuropathologic findings, including destructive lesions traced to remote events that may have caused hemorrhage and infarction.
- In premature infants, there is an increased risk of *germinal matrix hemorrhage,* often near the junction between the developing thalamus and caudate nucleus. Hemorrhages may remain small and localized or extend into the ventricular system and subarachnoid space, sometimes leading to hydrocephalus and death in severe cases.
- Infarcts may occur in the supratentorial periventricular white matter *(periventricular leukomalacia),* especially in premature infants; they take the form of chalky yellow plaques that consist of discrete white matter necrosis, often with dystrophic calcification. Ultimately, the infarcted areas develop into large cystic lesions (Fig. 28.8); when damage is extensive and involves both gray and white matter, the condition is termed *multicystic encephalopathy*.
- In perinatal ischemic lesions of the cerebral cortex, the depths of sulci bear the brunt of injury and result in mushroom-shaped gyri with thinned-out, gliotic stalks *(ulegyria)*. The basal ganglia and thalamus may also suffer ischemic injury, with patchy neuronal loss and reactive gliosis. Later, aberrant and irregular myelinization gives rise to a marble-like appearance of the deep nuclei *(status marmoratus)*. Because the lesions are in the caudate, putamen, and thalamus, movement disorders such as choreoathetosis are common clinical sequelae.

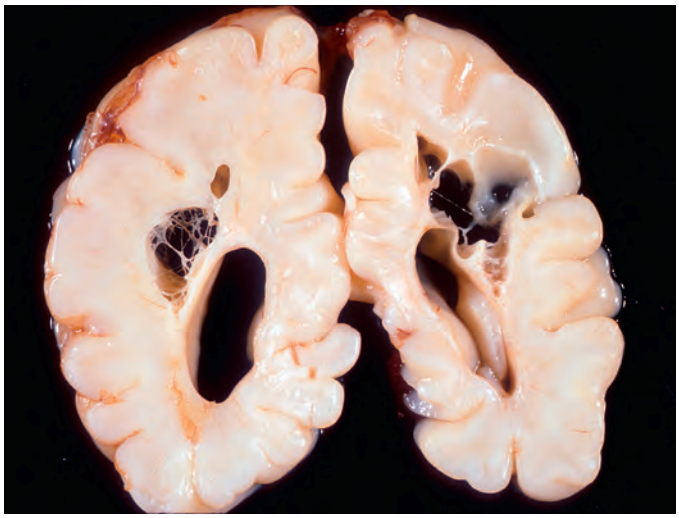


Figure 28.8 Chronic stage of periventricular leukomalacia. Large cystic spaces in the periventricular white matter (seen in both hemispheres of this brain) are the long-term sequalae of a severe prenatal or perinatal ischemic injury.

TRAUMA

The anatomic location of the lesion and the limited capacity of the brain for functional repair are major determinants of the consequences of CNS trauma. Injury of several cubic centimeters of brain parenchyma may be clinically silent (e.g., in the frontal lobe), severely disabling (e.g., in the spinal cord), or fatal (e.g., in the brainstem).

The physical forces associated with head injury may result in skull fractures, parenchymal injury, and vascular injury; all three can coexist. The magnitude and distribution of a traumatic brain lesion depends on the shape of the object causing the trauma, the force of impact, and whether the head is in motion at the time of injury. A blow to the head may be penetrating or blunt; it may cause either an open or a closed injury.

Skull Fractures

A fracture in which bone is displaced into the cranial cavity by a distance greater than the thickness of the bone is called a *displaced skull fracture*. The thickness of the cranial bones varies; therefore, their resistance to fracture differs greatly. Also, the relative incidence of fractures among skull bones is related to the pattern of falls. For example, when an individual falls while awake (such as might occur when stepping off a ladder), the site of impact is often the occipital portion of the skull; in contrast, a fall that follows loss of consciousness (as might follow a syncopal attack) can result in either frontal or occipital impact. Symptoms referable to the lower cranial nerves or the cervicomedullary region, and the presence of orbital or mastoid hematomas distant from the point of impact, raise the suspicion of a basal skull fracture, which typically follows impact to the occiput or sides of the head; CSF leakage from the nose or ear and infection (meningitis) may follow.

Parenchymal Injuries

The structural consequences and clinical manifestations of the parenchymal injury depend on the nature and severity of the causative head trauma and range from mild (concussion) to severe (contusions and lacerations).

Concussion

Concussion is a clinical syndrome of altered consciousness secondary to head injury, typically brought about by a change in the momentum of the head (e.g., following impact with a rigid object). The characteristic clinical picture includes the sudden onset of transient neurologic dysfunction, including loss of consciousness, temporary respiratory arrest, and loss of reflexes. Although neurologic recovery is usually complete, amnesia for the event often persists. The pathogenesis of the disruption of neurologic function is unknown, but likely involves dysregulation of the reticular activating system in the brainstem. Postconcussive neuropsychiatric syndromes, typically associated with repetitive injuries, are well recognized, and there is increasing evidence that significant cognitive impairment can emerge along with distinct pathologic findings termed chronic traumatic encephalopathy (discussed later).

Direct Parenchymal Injury

Contusions and *lacerations* are brain injuries caused by transmission of kinetic energy to the brain. A contusion is analogous to the familiar bruise caused by blunt trauma, and a laceration is an injury caused by penetration of an object and tearing of tissue. As with any other organ, a blow to the surface of the brain, transmitted through the skull, leads to rapid tissue displacement, disruption of vascular channels, and subsequent hemorrhage, tissue injury, and edema. Hemorrhage can extend into the subarachnoid space from these lesions. The crests of gyri are most susceptible because this is where the direct force is greatest. The most common locations for contusions correspond to the most frequent sites of direct impact and to regions of the brain that overlie a rough and irregular inner skull surface, such as the frontal lobes along the orbital ridges and the temporal lobes. Contusions are less frequent over the occipital lobes, brainstem, and cerebellum, but may be seen when there is an adjacent skull fracture (fracture contusions).

A person who suffers a blow to the head may develop a contusion at the point of contact (a *coup* injury) or on the brain surface diametrically opposite to it (a *contrecoup* injury). Their macroscopic and microscopic appearances are indistinguishable, and the distinction between them is based on identification of the point of impact. In general, if the head is immobile at the time of trauma, only a coup injury is found. If the head is mobile, both coup and contrecoup lesions may be found, though the latter predominate and are thought to develop when the brain strikes the opposite inner surface of the skull after sudden deceleration.

Sudden impacts that result in violent posterior or lateral hyperextension of the neck (as occurs when a pedestrian is struck from the rear by a vehicle) may avulse the pons from the medulla or the medulla from the cervical cord, causing instant death.

MORPHOLOGY

When seen on cross-section, contusions are wedge shaped, with the broad base lying along the surface at the point of impact (Fig. 28.9A). The appearance of contusions is similar regardless of the source of the trauma. In the earliest stages, there is edema and hemorrhage, which is often pericapillary. During the next few hours, the extravasation of blood extends throughout the involved tissue, across the width of the cerebral cortex, and into the white matter and subarachnoid space. Morphologic evidence of neuronal injury (pyknosis of the nucleus, eosinophilia of the cytoplasm, and disintegration of the cell) takes 12 to 24 hours to appear, although functional deficits generally occur earlier. Axonal swellings develop along the full length of the damaged neurons. The inflammatory response to the injured tissue follows its usual course, with the appearance of sparse neutrophils followed by abundant macrophages. Old traumatic lesions on the surface of the brain have a characteristic gross appearance. They are depressed, retracted, yellowish brown patches involving the crests of gyri, most commonly those that are located at the sites of contrecoup injuries (inferior frontal cortex, temporal and occipital poles); these lesions, called **plaque jaune** (Fig. 28.9B), can become epileptic foci.

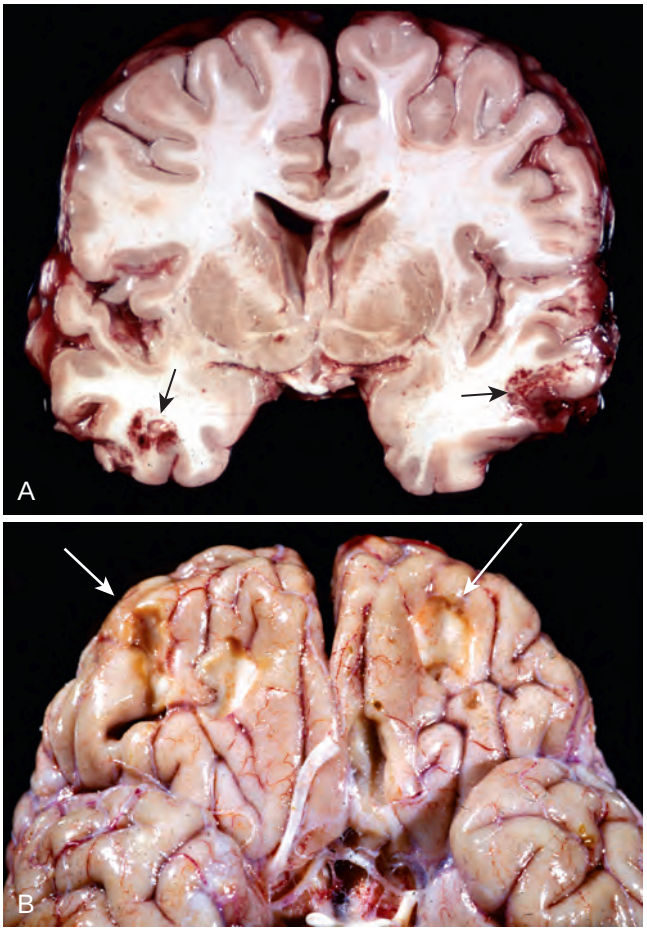


Figure 28.9 (A) Acute contusions are present in both temporal lobes, with areas of hemorrhage and tissue disruption *(arrows)*. (B) Remote contusions *(arrows)* are present on the inferior frontal surface of this brain and have a yellow color (plaque jaune) that reflects hemosiderin accumulation.

More extensive hemorrhagic regions of brain trauma give rise to larger cavitated lesions, which resemble remote infarcts. In old contusions, gliosis and residual hemosiderin-laden macrophages predominate.

Diffuse Axonal Injury

Traumatic brain injuries may also damage the deep white matter regions, cerebral peduncles, superior colliculi, and deep reticular formation in the brainstem. The microscopic findings include axonal swelling, indicative of *diffuse axonal injury,* and focal hemorrhagic lesions. As many as 50% of individuals who develop coma shortly after trauma, even without cerebral contusions, are believed to have diffuse axonal injury. Axons are injured directly by mechanical forces, with subsequent alterations in axoplasmic flow. Such injuries may occur from marked changes in angular acceleration (e.g., from blast injuries), even in the absence of physical impacts involving the skull.

MORPHOLOGY

Diffuse axonal injury is characterized by widespread, often asymmetric axonal swellings that appear within hours of the injury and may persist for much longer. The swelling is best demonstrated with silver impregnation techniques or with immunoperoxidase stains for axonally transported proteins, such as amyloid precursor protein and α-synuclein. Later, increased numbers of microglia are seen in damaged areas of the cerebral cortex, and subsequently there is degeneration of the involved fiber tracts.

Traumatic Vascular Injury

Vascular injury is a frequent component of CNS trauma. It results from disruption of the vessel wall and leads to hemorrhage in different anatomic sites (Table 28.1). Depending on the position of the ruptured vessel, hemorrhage may occur in the epidural, subdural, subarachnoid, and intraparenchymal compartments, sometimes in combination (Fig. 28.10). Epidural and subdural hemorrhages rarely occur outside of the setting of trauma; in some settings, such as coagulopathy or significant cerebral atrophy, subdural hemorrhages can follow even minor trauma. Subarachnoid hemorrhage almost always accompanies parenchymal trauma, but can also develop spontaneously secondary to vascular anomalies (discussed later).

Epidural Hematoma

The dura is fused with the periosteum and is supplied by a number of dural arteries. These arteries, most notably the middle meningeal artery, are vulnerable to traumatic injury. In adults, this most often occurs with temporal skull fractures in which the fracture crosses the course of the vessel. In children, in whom the skull is deformable, a temporary displacement of the skull bones leading to laceration of a vessel can occur in the absence of a skull fracture.

Once a vessel has been torn, the extravasation of blood under arterial pressure can cause the dura to separate from the periosteum, creating a space (Fig. 28.11). The expanding hematoma compresses the underlying brain. When blood accumulates slowly, patients may experience a lucid period

Table 28.1 Patterns of Hemorrhage in the Central Nervous System

Location	Etiology	Additional Features
Epidural space	Trauma	Usually associated with skull fracture (in adults); rapidly evolving neurologic symptoms (often after a short lucid period) that require intervention
Subdural space	Trauma	May follow minor trauma; slowly evolving neurologic symptoms, often with a delay from the time of injury
Subarachnoid space	Trauma	Typically associated with underlying parenchymal injury
	Vascular abnormality (arteriovenous malformation or aneurysm)	Sudden onset of severe headache, often with rapid neurologic deterioration; secondary injury may emerge and is associated with vasospasm
Intraparenchymal space	Trauma (contusions)	Selective involvement of the crests of gyri, where the brain is in contact with the inner surface of the skull (frontal and temporal tips, orbitofrontal surface)
	Ischemia (hemorrhagic conversion of an ischemic infarct)	Petechial hemorrhages in an area of previously ischemic brain, usually following the cortical ribbon
	Cerebral amyloid angiopathy	"Lobar" hemorrhage, involving subcortical white matter and often with extension into the subarachnoid space
	Hypertension	Centered in the deep white matter, thalamus, basal ganglia, or brainstem; may extend into the ventricular system
	Tumors (primary or metastatic)	Associated with high-grade gliomas or certain metastases (melanoma, choriocarcinoma, renal cell carcinoma)

before the onset of neurologic signs. A symptomatic epidural hematoma is a neurosurgical emergency; without prompt diagnosis and drainage, fatal brain herniation may occur within a few hours.

Subdural Hematoma

The dura is composed of two layers—an external collagenous layer and an inner more cellular layer containing fibroblasts. Bridging veins travel from the convexities of the cerebral hemispheres through the subarachnoid space and dura to empty into the dural sinuses. The brain is suspended in CSF, but the venous sinuses are fixed relative to the dura; as a result, traumatic displacement of the brain can tear the veins at the point where they penetrate the dura. The extravasated blood dissects through the two layers of the dura, producing a subdural hematoma. In older individuals with brain atrophy, the bridging veins are stretched, hence the increasing incidence of subdural hematoma with aging. Infants are also very susceptible to subdural hematomas because their bridging veins are thin-walled.

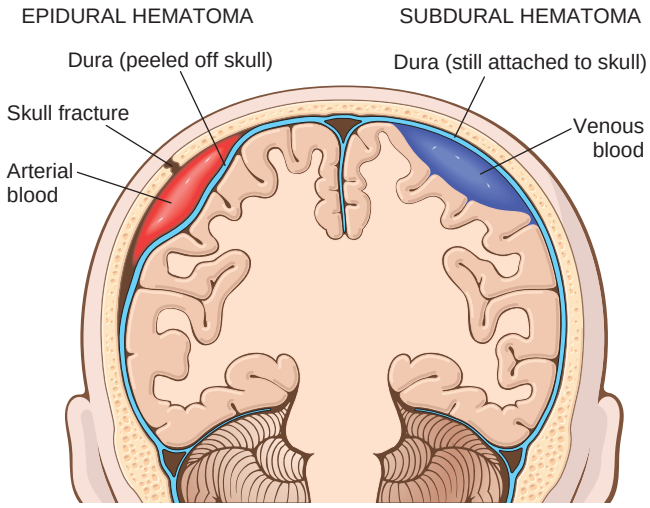


Figure 28.10 Epidural hematoma *(left)* in which rupture of a meningeal artery, usually associated with a skull fracture, leads to accumulation of arterial blood between the dura and the skull. In a subdural hematoma *(right)*, damage to bridging veins between the brain and the superior sagittal sinus leads to the accumulation of blood between the dura and the arachnoid.

MORPHOLOGY

Grossly, **acute subdural hematomas** appear as a collection of freshly clotted blood along the brain surface, without extension into the depths of sulci (Fig. 28.12). The underlying brain is flattened, and the subarachnoid space is often clear. Usually, venous bleeding is self-limited and the resulting hematoma is broken down and organized over time; this most often occurs in the following sequence:

- Lysis of the clot (about 1 week)
- Growth of fibroblasts from the dural surface into the hematoma (2 weeks)

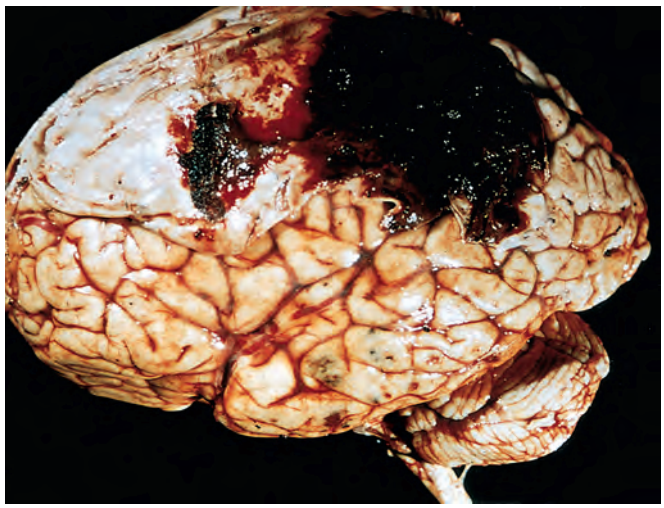


Figure 28.11 Epidural hematoma covering a portion of the dura. Also present are multiple small contusions in the temporal lobe. (Courtesy the late Dr. Raymond D. Adams, Massachusetts General Hospital, Boston, Mass.)

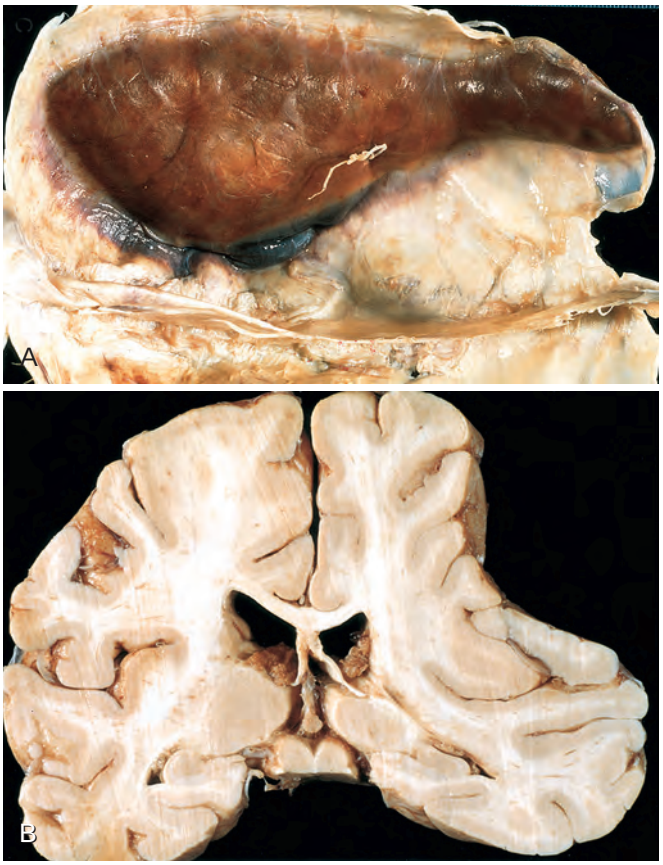


Figure 28.12 (A) Large organizing subdural hematoma attached to the dura. (B) Coronal section of the brain showing atrophy of the hemisphere compressed by the subdural hematoma shown in A.

- Early development of hyalinized connective tissue (1 to 3 months)

Typically, the organized hematoma is firmly attached to the inner surface of the dura by ingrowing fibrous tissue and is free of the underlying arachnoid, which does not contribute to healing. The lesion can eventually retract as the granulation tissue matures until only a thin layer of reactive connective tissue remains ("subdural membranes"). In other cases, however, multiple recurrent episodes of bleeding occur **(chronic subdural hematoma),** presumably from the thin-walled vessels of the granulation tissue.

Clinical Features

Symptomatic subdural hematomas most often manifest within 48 hours of injury. They are most common over the lateral aspects of the cerebral hemispheres and are bilateral in about 10% of cases. Neurologic signs are attributable to the pressure exerted on the adjacent brain. There may be focal signs, but often the clinical manifestations are nonlocalizing and include headache and confusion. Slowly progressive neurologic deterioration is typical, but acute decompensation may also occur. The treatment of subdural hematomas is to remove the blood and associated organizing tissue. The risk of repeat bleeding is greatest in the first few months after the initial hemorrhage.

Sequelae of Brain Trauma

A broad range of neurologic syndromes may become manifest months or years after brain trauma of any cause. These have gained increasing notice in the context of litigation involving issues of compensation for those in the civilian work force, professional athletes, and the military services.

- *Posttraumatic hydrocephalus* is largely due to obstruction of CSF resorption from hemorrhage into the subarachnoid space.
- *Chronic traumatic encephalopathy* (CTE, previously referred to as "dementia pugilistica") is a dementing illness that develops after repeated head trauma. Affected brains are atrophic, with enlarged ventricles, and show accumulation of tau-containing neurofibrillary tangles in a characteristic pattern involving gyral depths and perivascular regions in the frontal and temporal lobe cortices. Although concussion is thought of as having no structural consequences, it is clear that repeated events are antecedents to CTE; however, it remains uncertain which factors determine whether encephalopathy will ultimately develop (the number, frequency, and/or severity of individual traumatic events, or some combination of these).
- Other important sequelae of brain trauma include posttraumatic epilepsy, risk of infection, and psychiatric disorders.

Spinal Cord Injury

The spinal cord is vulnerable to trauma from its skeletal encasement. Most injuries that damage the cord are associated with the transient or permanent displacement of the vertebral column. The level of cord injury determines the extent of the neurologic manifestations: lesions involving the thoracic vertebrae or below can lead to paraplegia; cervical lesions result in quadriplegia; those above C4 can, in addition, lead to respiratory compromise from paralysis of the diaphragm. Damage at the region of impact to descending and ascending white matter tracts isolates the distal spinal cord from the rest of the brain. This interruption, paired with the localized gray matter damage at the level of the impact, is the principal cause of neurologic deficits.

MORPHOLOGY

The histologic features of traumatic injury of the spinal cord are similar to those found at other sites in the CNS. At the level of injury the acute phase consists of hemorrhage, necrosis, and axonal swelling in the surrounding white matter. The lesion tapers above and below the level of injury. In time, central areas of neuronal destruction become cystic and gliotic; cord sections above and below the lesion show secondary ascending and descending Wallerian degeneration, respectively, involving the long white-matter tracts at the site of trauma.

KEY CONCEPTS

Perinatal Brain Injury

- Timing of injury is critical, with earlier events resulting in greater damage and deficits.

- Cerebral palsy is the term used for nonprogressive deficits associated with injury during the prenatal and perinatal periods.

Trauma

- Physical injury to the brain can occur when the inside of the skull comes into forceful contact with the brain.
- In blunt trauma, if the head is mobile there may be brain injury both at the original point of contact (coup injury) and on the opposite side of the brain (contrecoup injury).
- Parenchymal injuries take the form of contusions, with hemorrhage extending into the subarachnoid space.
- Rapid displacement of the head and brain can tear axons (diffuse axonal injury), often immediately causing severe, irreversible neurologic deficits.
- Traumatic tearing of blood vessels leads to epidural or subdural hematoma.

CEREBROVASCULAR DISEASE

Cerebrovascular disease—injury to the brain as a consequence of altered blood flow—can be grouped into ischemic and hemorrhagic etiologies, with tissue infarction the ultimate consequence of both. "Stroke" is the clinical designation applied to these conditions and is defined as neurologic signs and symptoms that can be explained by a vascular mechanism, have an acute onset, and persist beyond 24 hours. (If symptoms disappear within 24 hours, the event is termed "transient ischemic event.") Cerebrovascular disease is the third leading cause of death (after heart disease and cancer) in the United States, and the most prevalent cause of morbidity and mortality from neurologic disease. Stroke stems from two major mechanisms:

- *Ischemia and/or hypoxia* resulting from impairment of blood supply and oxygenation of CNS tissue. This can be either a global or focal process, with the clinical manifestations determined by the region of brain affected. In the brain, embolism is a more common cause of vascular occlusion than thrombosis.
- *Hemorrhage* resulting from rupture of CNS vessels. Common etiologies include hypertension and vascular anomalies (aneurysms and malformations; see Table 28.1).

Hypoxia and Ischemia

Although the brain accounts for only 1% to 2% of body weight, it receives approximately 15% of the resting cardiac output and accounts for 20% of the body's oxygen consumption. Cerebral blood flow remains relatively constant over a wide range of blood pressure and intracranial pressure because of autoregulation of vascular resistance. The brain is strictly dependent on aerobic metabolism to meet its constant energy demands, and it may be deprived of oxygen by either hypoxemia (low blood oxygen content) or ischemia (inadequate blood flow). Inadequate blood flow may result from a reduction in perfusion pressure (as in hypotension), small- or large-vessel obstruction, or both.

When blood flow to a portion of the brain is reduced, the survival of the tissue at risk depends on the presence of collateral circulation, the duration of ischemia, and the magnitude and rapidity of the reduction of flow. These factors determine, in turn, the precise anatomic site and size of the lesion and, consequently, the clinical deficit.

The general biochemical changes in cells resulting from ischemia are discussed in Chapter 2. In addition to processes shared with ischemia in other parts of the body, ischemia in the CNS can result in inappropriate release of excitatory amino acid neurotransmitters such as glutamate, which can damage neurons by allowing excessive influx of calcium ions through N-methyl-D-aspartate (NMDA)-type glutamate receptors; this phenomenon is termed *excitotoxicity*. In the region of transition between necrotic tissue and the normal brain, there is an area of "at-risk" brain, referred to as the *penumbra;* this region can be rescued from cell death in many animal models with a variety of anti-apoptotic interventions, implying that ischemic neurons may die by apoptosis as well as necrosis.

Focal Cerebral Ischemia

Focal cerebral ischemia follows reduction or cessation of blood flow to a localized area of the brain due to partial or complete arterial obstruction. When the ischemia is sustained, infarction follows in the territory of the compromised vessel. The size, location, and shape of the infarct and the extent of tissue damage that results are influenced by the duration of the ischemia and the adequacy of collateral flow. The major source of collateral flow is the circle of Willis (supplemented by external carotid-ophthalmic artery collaterals). Inconstant collateral leptomeningeal vessels from the surface of the brain may also supply the distal branches of the anterior, middle, and posterior cerebral arteries through cortical-leptomeningeal anastomoses; in contrast, there is little if any collateral flow for the deep penetrating vessels of the thalamus, basal ganglia, and deep white matter.

Occlusive vascular disease of severity sufficient to lead to cerebral infarction may be due to embolization from a distant source, in situ thrombosis, or various vasculitides; the pathology of these conditions is also discussed in Chapters 4 and 11.

- *Embolism* to the brain occurs from a variety of sources. Cardiac mural thrombi are among the most common culprits; myocardial infarct, valvular disease, and atrial fibrillation are important predisposing factors. Next in frequency are thromboemboli originating from arteries, most often atheromatous plaques within the carotids. Other sources of emboli include paradoxical thromboemboli, particularly in children with cardiac anomalies; thromboemboli associated with cardiac surgery; and emboli of other types (tumor, fat, or air). The territory supplied by the middle cerebral artery—the direct extension of the internal carotid artery—is most frequently affected by embolic infarction; the incidence is about equal in the two hemispheres. Emboli tend to lodge where blood vessels branch or in areas of preexisting luminal stenosis. "Shower embolization," as in fat embolism, may occur after fractures; affected individuals manifest generalized cerebral dysfunction with disturbances of higher

cortical function and consciousness, often without localizing signs. Widespread hemorrhagic lesions involving the white matter are characteristic of embolization of bone marrow after trauma.

- *Thrombotic occlusion of the cerebral arteries* is most commonly caused by acute change of vulnerable atherosclerotic plaques, as in coronary artery disease (Chapter 12). The most common sites are the carotid bifurcation, the origin of the middle cerebral artery, and either end of the basilar artery. Thrombi cause progressive narrowing of the lumen, may be accompanied by anterograde extension, and may progress to fragmentation and distal embolization. Atherosclerotic cerebrovascular disease is frequently associated with systemic diseases such as hypertension and diabetes.
- *Inflammatory processes* that involve blood vessels may also lead to luminal narrowing, occlusion, and, hence, cerebral infarcts. Although infectious vasculitis of small and large vessels occurs with syphilis and tuberculosis, it is now more common in the setting of immunosuppression and opportunistic infection (e.g., aspergillosis). *Polyarteritis nodosa* and other noninfectious vasculitides may involve cerebral vessels and cause single or multiple infarcts throughout the brain. *Primary angiitis of the CNS* can also develop in the absence of systemic vasculitis.
- Other conditions that may cause thrombosis (and intracranial hemorrhage) include hypercoagulable states, dissecting aneurysm of extracranial arteries in the neck that supply the brain, and drug abuse (amphetamines, heroin, cocaine).

Brain infarcts are subdivided into two broad groups based on the presence of secondary hemorrhage. Because the brain has end-organ circulation with limited collateral supply, occlusive brain infarcts generally start as nonhemorrhagic (pale/anemic; Fig. 28.13A); clinically, these nonhemorrhagic infarcts are called ischemic, a confusing term as every infarct, not just this type, is caused by tissue ischemia. Secondary hemorrhage can occur from ischemia-reperfusion injury following spontaneous or therapeutic dissolution or fragmentation of the intravascular occlusive material. This process (termed *secondary hemorrhagic transformation* and leading to a hemorrhagic infarct) develops if the causative ischemic event lasts long enough to damage small blood vessels in the affected area; the resulting reperfusion hemorrhages are largely petechial in nature, but may be multiple or even confluent (see Fig. 28.13B). The clinical management of patients with nonhemorrhagic and hemorrhagic infarcts differs greatly, although the underlying causes are the same (for instance, thrombolytic therapy is contraindicated in a patient with brain hemorrhage of any etiology).

MORPHOLOGY

Both the gross and microscopic appearance of a **nonhemorrhagic infarct** changes over time. Grossly, there is little change in appearance during the first 6 hours of irreversible injury. By 48 hours, however, the tissue becomes pale, soft, and swollen, and the gray-white matter junction becomes indistinct. From 2 to 10 days, the brain becomes gelatinous and friable, and the previously ill-defined boundary between normal and infarcted tissue becomes more distinct as edema resolves in the viable adjacent tissue. From 10 days to 3 weeks, the tissue liquefies, eventually leaving a fluid-filled cavity that continues to expand until all of the dead tissue has been removed (Fig. 28.14).

Microscopically, the tissue reaction evolves along the following sequence:

- **Acute infarct** (Fig. 28.15A). After the first 6 to 12 hours, neurons in the affected area show eosinophilic neuronal necrosis (increased eosinophilia of the cytoplasm followed by nuclear pyknosis and karyorrhexis; "dead red neurons"); both cytotoxic and vasogenic edema are present. There is loss of the usual tinctorial characteristics of white- and gray-matter structures. Endothelial and glial cells, mainly astrocytes, swell, and myelinated fibers begin to disintegrate. Up to 48 hours, neutrophilic emigration progressively increases and then falls off (but is never as prominent as in myocardial infarction).
- **Subacute (evolving) infarct** (Fig. 28.15B). Phagocytic cells, derived from circulating monocytes and activated microglia, are evident at 48 to 72 hours and become the predominant cell type in the ensuing 2 to 3 weeks. The macrophages become stuffed with the products of myelin breakdown or blood and may persist in the lesion for months to years. Reactive astrocytes

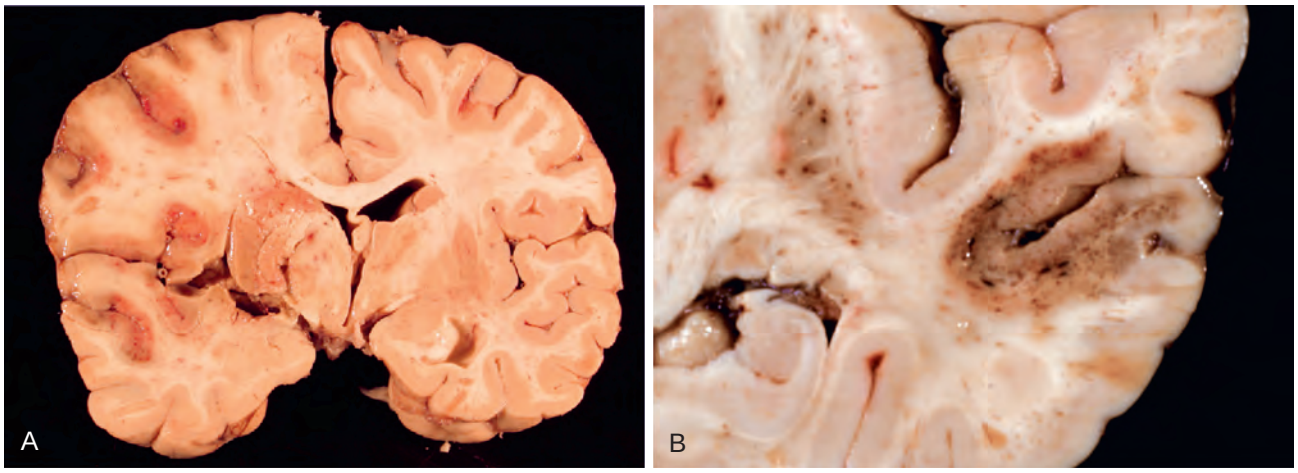


Figure 28.13 (A) An ischemic infarct involves the territory of the middle cerebral artery, including the striatum, on the left side of this brain. (B) A hemorrhagic infarct with punctate hemorrhages, consistent with ischemia-reperfusion injury, is present in the temporal lobe.

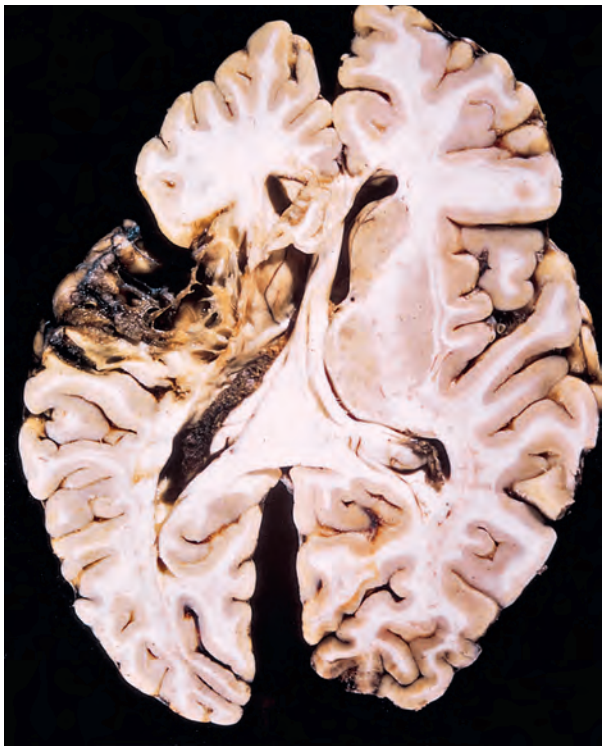


Figure 28.14 Old cystic infarct showing cavitation from loss of brain parenchyma.

and newly formed vessels can be seen at the periphery of the lesion as early as 1 week after the insult. As the process of liquefaction and phagocytosis proceeds, astrocytes at the edges of the lesion progressively enlarge, divide, and develop a prominent network of cytoplasmic extensions.

- **Healed infarct** (Fig. 28.15C). After several months, the astrocytic response recedes, leaving behind a dense meshwork of glial fibers admixed with new capillaries and some perivascular connective tissue. In the cerebral cortex, the cavity is separated from the meninges and subarachnoid space by a gliotic layer of tissue, which is derived from the molecular layer of the cortex. The pia and arachnoid are not affected. Infarcts undergo these reactive and reparative stages from the edges inward; thus, different areas of a lesion may appear chronologically divergent, revealing the natural progression of the response.

The features and temporal evolution of **hemorrhagic infarctions** parallel ischemic infarctions, with the addition of blood extravasation and resorption. In individuals receiving anticoagulant treatment, hemorrhagic infarcts may be associated with extensive intracerebral hematomas. Venous infarcts are often hemorrhagic and may occur after thrombotic occlusion of the superior sagittal sinus or other sinuses, or after occlusion of the deep cerebral veins. Neoplasms, localized infections, and other conditions leading to a hypercoagulable state increase the risk for venous thrombosis.

Lacunar Infarcts

Hypertension affects the deep penetrating arteries and arterioles that supply the basal ganglia and hemispheric white matter as well as the brainstem; these cerebral vessels develop *arteriolosclerosis,* also known as small vessel disease because it affects arteries 40 to 900 μm in diameter, described in more detail later. If this disease process progresses to thrombosis and complete vessel occlusion, the end-result is development of small cavitary infarcts known as *lacunes* or *lacunar infarcts* (Fig. 28.16). These lakelike spaces, arbitrarily defined as less than 15 mm wide, can be single or multiple and involve the putamen, globus pallidus, thalamus, internal capsule, deep white matter, caudate nucleus, and pons, in descending order of frequency. On microscopic

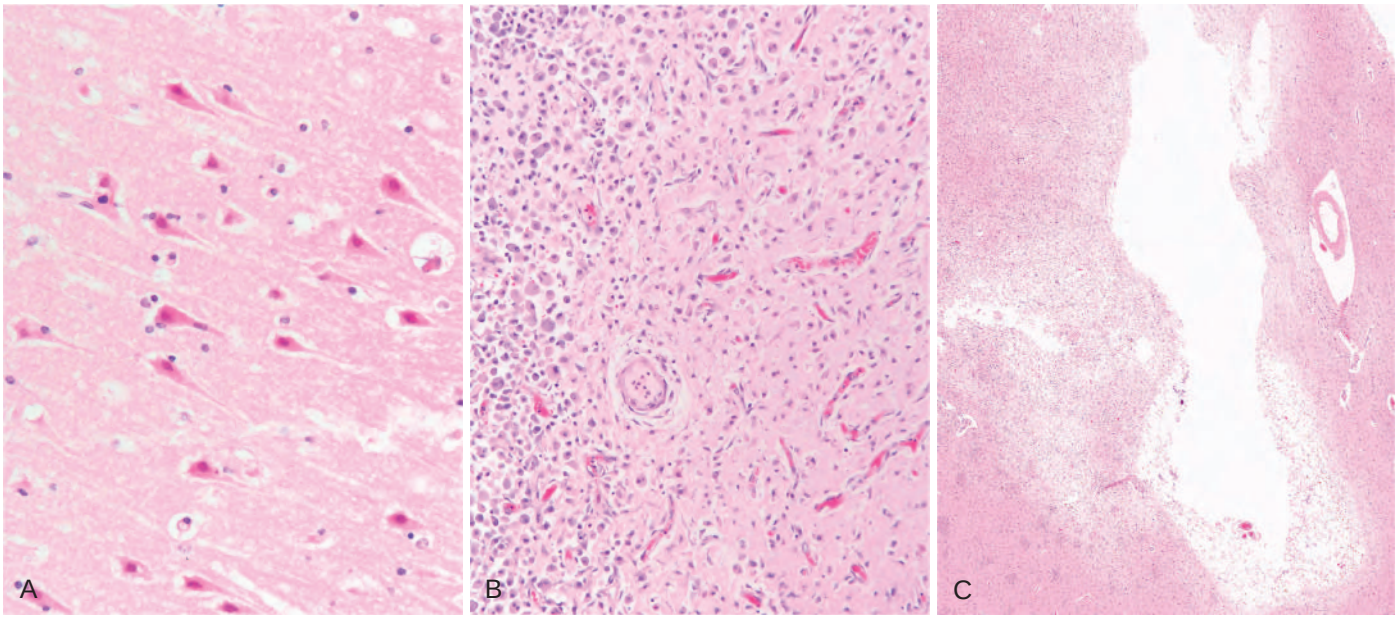


Figure 28.15 Cerebral infarcts. (A) Acute ischemic injury causes diffuse eosinophilia of neurons, which are beginning to shrink. (B) After about 10 days, the lesion is characterized by the presence of foamy macrophages *(best seen on the left)* and adjacent reactive gliosis with neovascularization *(on the right).* (C) Remote small infarct is seen as an area of tissue loss surrounded by residual gliosis.

Figure 28.16 Lacunar infarcts in the caudate and putamen *(arrows)*.

examination, there is tissue loss surrounded by gliosis. Depending on their location in the CNS, lacunar infarcts can be clinically silent or cause severe neurologic impairment. Affected vessels may also be associated with widening of the perivascular spaces without tissue infarction (*état criblé*).

Clinical Features

Deficits produced by infarction are determined by the anatomic distribution of the damage, rather than the underlying cause. Neurologic symptoms referable to the area of injury often develop rapidly, over minutes, and may continue to evolve over hours. There can be improvement in severity of symptoms that is associated with reversal of injury in the ischemic penumbra and resolution of associated local edema. In general, there is often slow improvement during a period of months. Because strokes are frequently associated with cardiovascular disease, many of the genetic and lifestyle risk factors are shared. Early diagnosis is of critical importance, as rapid treatment of nonhemorrhagic strokes with thrombolytic agents can often limit or entirely prevent the development of permanent neurologic deficits.

Global Cerebral Hypoxia/Ischemia

Global cerebral hypoxia or ischemia occurs when there is a generalized reduction of cerebral perfusion (as in cardiac arrest, shock, and severe hypotension) or decreased oxygen carrying capacity of the blood (e.g., in carbon monoxide poisoning). The clinical outcome varies with the severity and length of the insult. In mild cases, there may only be a transient postischemic confusional state, followed by complete recovery. Nevertheless, irreversible CNS damage may occur in individuals who experience global hypoxic and/or ischemic insults *(diffuse hypoxic/ischemic encephalopathy)*. Among CNS cells, there is a hierarchy of sensitivity to hypoxia/ischemia: neurons are the most sensitive, although glial cells (oligodendrocytes and astrocytes) are also vulnerable. The most sensitive neurons in the brain are pyramidal neurons in the hippocampus (especially area CA1, also referred to as Sommer sector), cerebellar Purkinje cells, and pyramidal neurons in the cerebral cortex (especially layers III and V); the molecular mechanisms underlying this selective vulnerability are not understood. With severe global cerebral hypoxia/ischemia, widespread neuronal death occurs, irrespective of regional vulnerability; patients who survive this injury often remain in a persistent vegetative state. Other patients meet the current clinical criteria for "brain death," including evidence of irreversible diffuse cortical injury (isoelectric, or "flat," electroencephalogram), brainstem damage (such as absent reflexes and respiratory drive), and absent cerebral perfusion. When individuals with this pervasive form of injury are maintained on mechanical ventilation, the brain gradually undergoes widespread liquefaction, producing so-called "respirator brain."

Border zone ("watershed") infarcts occur in the regions of the brain or spinal cord that lie at the most distal reaches of the arterial blood supply (i.e., the border zones between arterial territories). In the cerebral hemispheres, the border zone between the anterior and the middle cerebral artery distributions is at greatest risk. Damage to this region, located a few centimeters lateral to the interhemispheric fissure, results in a cortical wedge-shaped infarct that usually shows secondary hemorrhagic transformation (Fig. 28.17) and is often bilateral. Border zone infarcts usually develop after severe hypotensive episodes and are most commonly seen in patients resuscitated after cardiac arrest.

MORPHOLOGY

In the setting of global ischemia, the brain becomes edematous and swollen, producing widening of the gyri and narrowing of the sulci. The cut surface shows poor demarcation between gray and white matter. The microscopic features of irreversible ischemic injury evolve over time and mimic the changes seen in infarcts; the distinction between global and focal ischemic injury is based not on the nature of cellular pathology but on the overall pattern of brain involvement. **Early changes,** occurring 6 to 12 hours

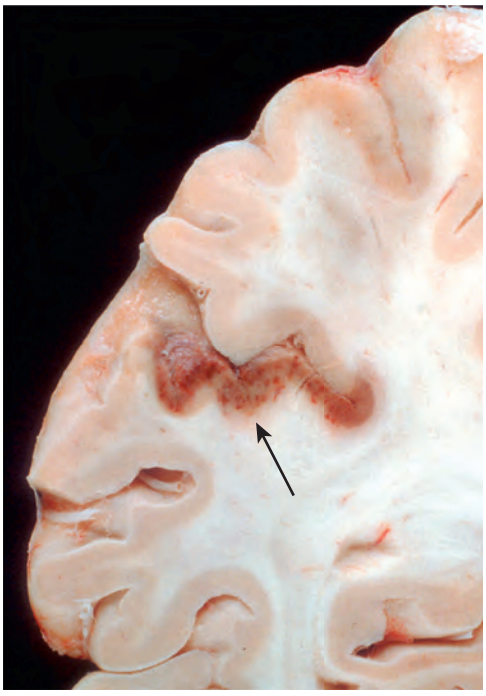


Figure 28.17 Classic watershed infarct with secondary hemorrhagic transformation *(arrow)*; boundary between anterior and middle cerebral artery circulations.

after the insult, are seen in neurons (dead red neurons described earlier); similar acute changes occur somewhat later in astrocytes and oligodendroglia. **Subacute changes,** occurring at 24 hours to 2 weeks, include tissue necrosis, influx of macrophages, vascular proliferation, and reactive gliosis. **Repair,** robust after approximately 2 weeks, is characterized by removal of necrotic tissue, loss of normal CNS architecture, and gliosis. In the cerebral neocortex, the neuronal loss and gliosis are uneven, with preservation of some layers and destruction of others, producing a pattern of injury termed **laminar necrosis.**

Intracranial Hemorrhage

Hemorrhages may occur at any site within the cranium—outside the brain or within it (intraparenchymal). Hemorrhages in the epidural or subdural space are typically associated with trauma and were discussed earlier; hemorrhages within the brain parenchyma and in the subarachnoid space, in contrast, are more often a manifestation of underlying cerebrovascular disease and are discussed in the following sections (see Table 28.1).

Intraparenchymal Hemorrhage

Rupture of a small intraparenchymal vessel can result in a primary hemorrhage within the brain, often associated with sudden onset of neurologic symptoms (stroke); this should not be confused with the secondary hemorrhagic transformation of an occlusive infarct (described earlier). Spontaneous (nontraumatic) intraparenchymal hemorrhages occur most commonly in middle to late adult life, with a peak incidence at about 60 years of age. Hemorrhages in the basal ganglia and thalamus are commonly designated "ganglionic hemorrhages," whereas those that occur in the lobes of the cerebral hemispheres are called "lobar hemorrhages"; the two major causes of these patterns of hemorrhage are hypertension and cerebral amyloid angiopathy, respectively. In addition, other local and systemic factors may cause or contribute to nontraumatic hemorrhage, including systemic coagulation disorders, neoplasms, vasculitis, aneurysms, and vascular malformations.

Hypertension is the risk factor most commonly associated with deep brain parenchymal hemorrhages, accounting for more than 50% of clinically significant hemorrhages and for roughly 15% of deaths among individuals with chronic hypertension. *Hypertensive intraparenchymal hemorrhage* may originate in the putamen (50% to 60% of cases), thalamus, pons, cerebellar hemispheres (rarely), and other regions of the brain (Fig. 28.18A). Hypertension leads to a number of vessel wall abnormalities, including accelerated atherosclerosis in larger arteries, hyaline arteriolosclerosis in smaller arteries, and (in severe cases) proliferative changes and frank necrosis of arterioles. Arteriolar walls affected by hyaline change (Fig. 28.18B) are thickened but more vulnerable to rupture than normal vessels; these changes are most prominent in the basal ganglia and the subcortical white matter. As described earlier, if small arteries affected by hyaline arteriolosclerosis do not rupture but are occluded, the result is lacunar infarction.

Cerebral amyloid angiopathy (CAA) is the risk factor most commonly associated with lobar hemorrhages (Fig. 28.18C). In CAA, amyloidogenic peptides, usually the same ones found in Alzheimer disease (Aβ; see later), are deposited in the walls of medium- and small-caliber meningeal, cortical, and cerebellar vessels; involved vessels are rigid, and as a result fail to collapse during tissue processing and sectioning. Although similar to hyaline arteriolosclerosis on routine hematoxylin and eosin (H&E) stain, the hyaline material in CAA consists of β amyloid rather than collagen (Fig. 28.18D) and is primarily seen in the leptomeningeal and cortical (rather than basal ganglia and white matter) vessels. Amyloid deposition can weaken the vessel wall and lead to hemorrhage; as a result, many individuals with CAA have evidence of numerous small hemorrhages within the brain ("microbleeds") that can be visualized by various imaging methods. As with Alzheimer disease (discussed later), there is a relationship between polymorphisms in the gene that encodes apolipoprotein E (ApoE) and risk of disease; specifically, the presence of either an ε2 or ε4 allele increases the risk of bleeding. Autosomal dominant forms of CAA are associated with certain mutations in the *APP* gene, which encodes the precursor for the Aβ peptides that are prone to deposit as amyloid.

Other forms of hereditary small-vessel diseases of the CNS have been identified. *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy* (CADASIL) is an autosomal dominant disorder caused by mutations in the *NOTCH3* gene that lead to misfolding of the extracellular domain of the NOTCH3 receptor. The disease is characterized clinically by recurrent small vessel strokes (usually infarcts, less often hemorrhages) and dementia. Imaging studies show that the first detectable changes are in white matter, usually presenting around 35 years of age and then progressing further over time. Other forms of the heritable small vessel disease include a disorder associated with mutations in the gene for COL4A1, a component of the vascular basement membrane.

MORPHOLOGY

Acute primary intraparenchymal hemorrhages are characterized by a central core of clotted blood that compresses the adjacent parenchyma; this compression leads to secondary infarction of the affected brain tissue, with anoxic neuronal and glial changes as well as edema. Eventually the edema resolves, hemosiderin- and lipid-laden macrophages appear, and proliferation of reactive astrocytes is seen at the periphery of the lesion; the cellular events then follow the same time course that is observed after cerebral infarction. Old hemorrhages show areas of parenchymal cavitary destruction with a rim of brownish discoloration.

Clinical Features

Intracerebral hemorrhage can be clinically devastating if it involves a large part of the brain or extends into the ventricular system. When hemorrhage affects smaller regions, it is either clinically silent or evolves like an infarct; over weeks or months, there is a gradual removal of the hematoma, sometimes with considerable clinical improvement. Again, the location of the hemorrhage determines the clinical manifestations.

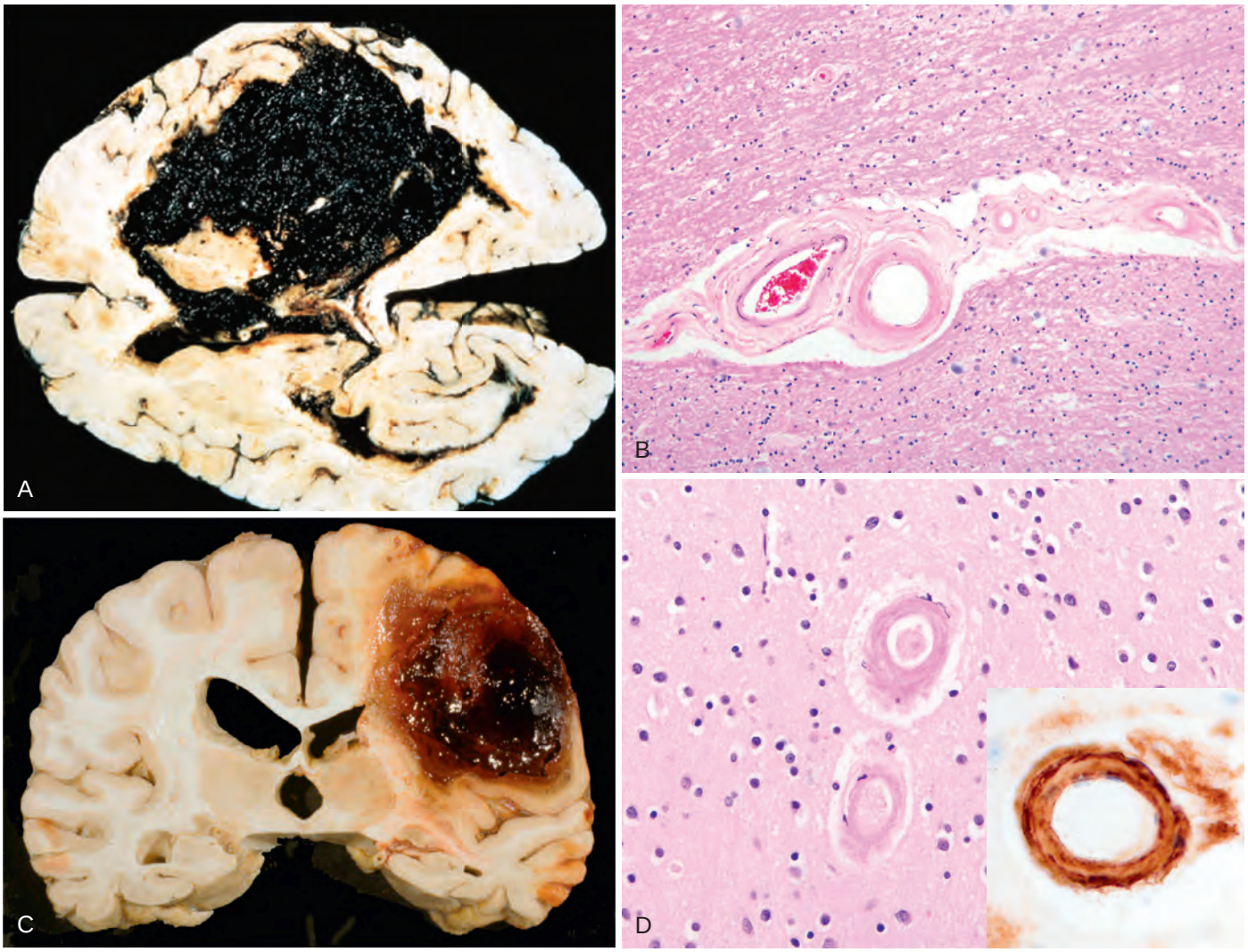


Figure 28.18 (A) Massive hypertensive ganglionic hemorrhage rupturing into a lateral ventricle. (B) Hyaline arteriolosclerosis (fibrosis and thickening of the arteriolar walls) develops in the basal ganglia and subcortical white matter of patients with long-standing hypertension; it is a risk factor for hypertensive hemorrhages as well as lacunar infarcts. (C) Large lobar hemorrhage due to cerebral amyloid angiopathy; it focally dissects into the subarachnoid space. (D) Amyloid deposition in cortical arterioles in cerebral amyloid angiopathy; *inset,* immunohistochemical staining for Aβ highlights the deposited material in the vessel wall. (C, Courtesy Dr. Dimitri Agamanolis, http://neuropathology-web.org.)

Subarachnoid Hemorrhage and Ruptured Saccular Aneurysm

The most frequent cause of spontaneous subarachnoid hemorrhage is rupture of a saccular ("berry") aneurysm in a cerebral artery. (As noted earlier, brain trauma is the most common cause of subarachnoid hemorrhage overall.) Nontraumatic subarachnoid hemorrhage may also result from rupture of a primary intracerebral hemorrhage into the ventricular system, vascular malformation, hematologic disturbances, and tumors.

Saccular aneurysm is the most common type of intracranial aneurysm; other aneurysm types include atherosclerotic (fusiform; mostly of the basilar artery), mycotic, traumatic, and dissecting. These latter three, like saccular aneurysms, are most often found in the anterior circulation, but more often cause cerebral infarction rather than subarachnoid hemorrhage.

Saccular aneurysms are found in about 2% of the population according to recent data from community-based radiologic studies. About 90% of saccular aneurysms are found near major arterial branch points in the anterior circulation (Fig. 28.19); multiple aneurysms exist in 20% to 30% of cases based on autopsy series.

Pathogenesis

Although the etiology of saccular aneurysms remains obscure, the structural abnormality of the involved vessel (absence of smooth muscle and intimal elastic lamina) suggests that they are developmental anomalies. The majority occur sporadically, but genetic factors may be important in their pathogenesis because there is an increased incidence of aneurysms in first-degree relatives of those affected. There is also an increased incidence in individuals with certain Mendelian disorders (e.g., autosomal dominant polycystic kidney disease, Ehlers-Danlos syndrome type

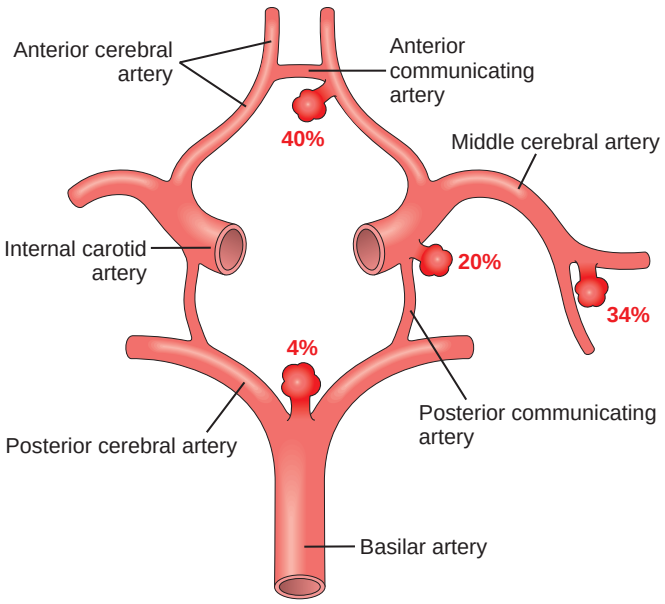


Figure 28.19 Common sites of saccular (berry) aneurysms in the circle of Willis.

IV, neurofibromatosis type 1 [NF1], and Marfan syndrome), fibromuscular dysplasia of extracranial arteries, and coarctation of the aorta. Other predisposing factors include cigarette smoking and hypertension (estimated to be present in about half of affected individuals). Although they are sometimes referred to as "congenital," the aneurysms are not present at birth but develop over time because of an underlying defect in the media of the vessel wall.

MORPHOLOGY

An unruptured saccular aneurysm is a thin-walled outpouching, usually at an arterial branch point along the circle of Willis or a major vessel just beyond. Saccular aneurysms measure from a few millimeters to 2 or 3 cm in diameter and have a bright red, shiny surface and a thin, translucent wall (Fig. 28.20). Atheromatous plaques, calcification, or thrombi may be found in the wall or lumen of the aneurysm. Sometimes there is evidence of prior hemorrhage, in the form of brownish discoloration of the adjacent brain and meninges. The neck of the aneurysm may be wide or narrow. Rupture usually occurs at the apex of the sac and leads to extravasation of blood into the subarachnoid space, the substance of the brain, or both. The arterial wall adjacent to the neck of the aneurysm often shows some intimal thickening and attenuation of the media. Smooth muscle and intimal elastic lamina do not extend into the neck and are absent from the aneurysm sac itself, which is made up of thickened hyalinized intima and a covering of adventitia.

Clinical Features

Rupture of an aneurysm leading to subarachnoid hemorrhage is most frequent in the fifth decade and is slightly more frequent in women. Overall, aneurysms rupture at a rate of 1.3% per year, but the risk is higher for larger aneurysms; for example, aneurysms greater than 10 mm in diameter have a roughly 50% risk of bleeding per year. Rupture may occur at any time, but in about one-third of cases it is associated with acute increases in intracranial pressure, such as with straining at stool or sexual orgasm. Blood under arterial pressure is forced into the subarachnoid space, and affected individuals are stricken with a sudden, excruciating headache ("the worst headache I've ever had") and rapidly lose consciousness. Between 25% and 50% of patients die with the first rupture, but patients who survive often improve and recover consciousness in minutes. Repeat bleeding is common in survivors and unpredictable in timing. With each episode of bleeding, the prognosis is worse.

The clinical consequences of blood in the subarachnoid space can be separated into acute events (occurring within

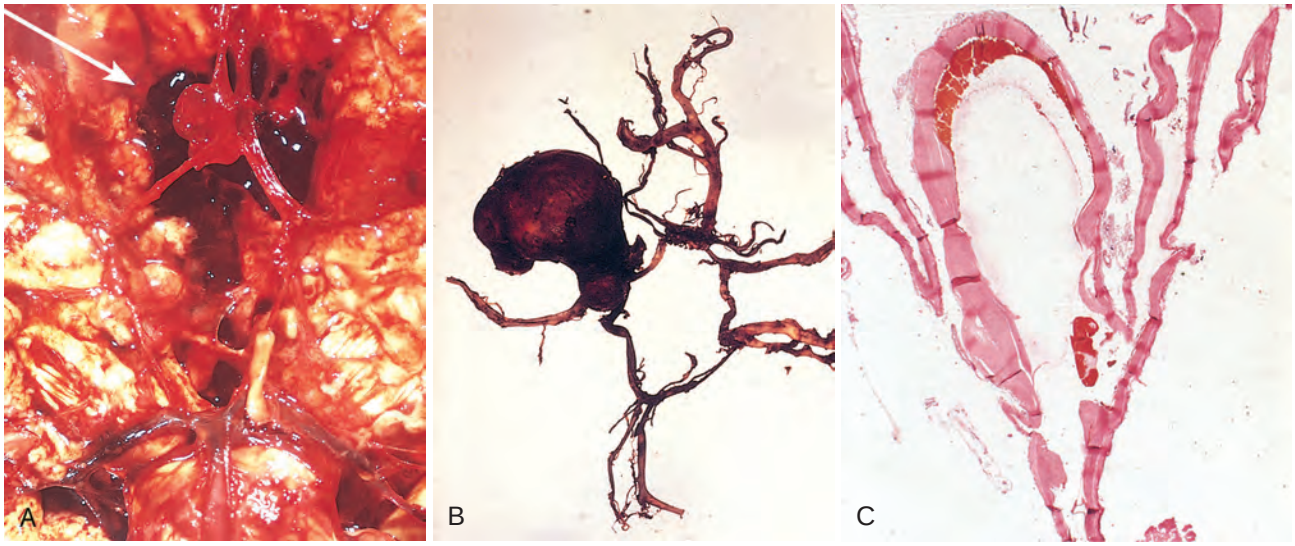


Figure 28.20 (A) View of the base of the brain, dissected to show the circle of Willis with an aneurysm of the anterior cerebral artery *(arrow)*. (B) Dissected circle of Willis to show large aneurysm. (C) Section through a saccular aneurysm showing the hyalinized fibrous vessel wall (hematoxylin and eosin).

hours to days after the hemorrhage) and late sequelae associated with the healing process. In the first few days after a subarachnoid hemorrhage, regardless of the etiology, there is an increased risk of additional ischemic injury from vasospasm affecting vessels bathed in the extravasated blood. This problem is of greatest significance in cases of basal subarachnoid hemorrhage, in which vasospasm can involve major vessels of the circle of Willis. Various mediators have been proposed to have a role in this process, including endothelins, nitric oxide, and arachidonic acid metabolites. In the healing phase of subarachnoid hemorrhage, meningeal fibrosis and scarring occur, sometimes leading to obstruction of CSF flow as well as interruption of the normal pathways of CSF resorption.

Vascular Malformations

Vascular malformations of the brain are classified into four principal groups: arteriovenous malformations, cavernous malformations, capillary telangiectasias, and venous angiomas. Of these, the first two are the types associated with risk of hemorrhage and development of neurologic symptoms. Although lacking morphologic features that typify neoplasms, arteriovenous malformations are frequently associated with activating somatic mutations in the *KRAS* oncogene within the endothelial cells that line the malformed vessels, suggesting that dysregulated RAS signaling has a central role in their pathogenesis.

MORPHOLOGY

Arteriovenous malformations (tangled networks of wormlike vascular channels with high blood flow due to prominent, pulsatile arteriovenous shunting) may involve vessels in the subarachnoid space, the brain, or both (Fig. 28.21). They are composed of greatly enlarged blood vessels separated by gliotic tissue, often with evidence of prior hemorrhage. Some vessels can be recognized as arteries with duplication and fragmentation of the internal elastic lamina, while others show marked thickening or partial replacement of the media by hyalinized connective tissue.

Cavernous malformations consist of distended, loosely organized vascular channels arranged back to back with collagenized walls of variable thickness; there is usually no brain parenchyma between vessels in this type of malformation. They occur most often in the cerebellum, pons, and subcortical regions, in decreasing order of frequency, and are "low-flow" channels that do not participate in arteriovenous shunting. Foci of old hemorrhage, infarction, and calcification frequently surround the abnormal vessels.

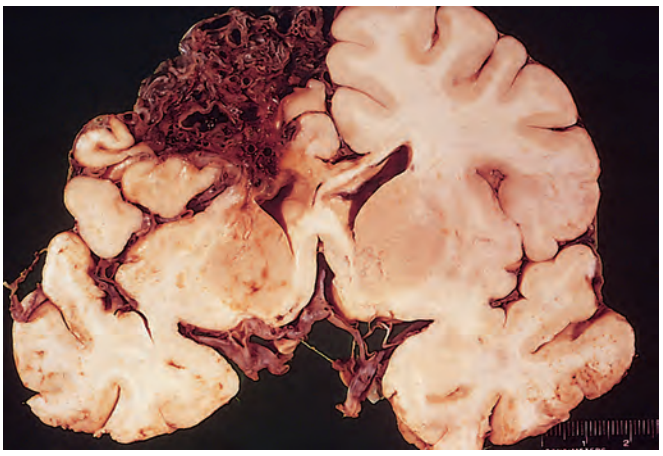


Figure 28.21 Large arteriovenous malformation in the left cerebral hemisphere.

Clinical Features

Arteriovenous malformations are the most common clinically significant vascular malformation. Males are affected twice as frequently as females. The lesion often presents between 10 and 30 years of age as a seizure disorder, an intracerebral hemorrhage, or a subarachnoid hemorrhage. The most common site is the territory of the middle cerebral artery, particularly its posterior branches. Large arteriovenous malformations occurring in the newborn period can lead to congestive heart failure because of shunt effects, especially if the malformation involves the vein of Galen. Cavernous malformations are unique among this group of disorders in that familial forms are relatively common. Multiplicity of lesions is a hallmark of familial cases, which are inherited as a highly penetrant autosomal dominant trait.

Vascular Dementia

Individuals who, over the course of many months and years, suffer multiple, bilateral, gray matter (cortex, thalamus, basal ganglia) and white matter (centrum semiovale) infarcts may develop a distinctive clinical syndrome characterized by dementia, gait abnormalities, and pseudobulbar signs, often with superimposed focal neurologic deficits. The syndrome, generally referred to as *vascular dementia,* is caused by multifocal vascular disease of several types, including (1) cerebral atherosclerosis, (2) vessel thrombosis or embolization from carotid vessels or from the heart, and (3) cerebral arteriolosclerosis from chronic hypertension. When the pattern of injury preferentially involves large areas of the subcortical white matter with myelin and axon loss, the disorder is referred to as *Binswanger disease (subcortical white matter dementia);* this distribution of vascular white-matter injury must be distinguished clinically and radiologically from other diseases that affect the hemispheral white matter. In addition, many individuals with neurodegenerative diseases resulting in cognitive impairment or dementia also have evidence of cerebrovascular disease. The presence of significant cerebrovascular disease increases risk of neurologic impairment for a given level of lesions associated with the degenerative diseases, suggesting that it is an independent contributing factor to disruption of normal brain function.

KEY CONCEPTS

CEREBROVASCULAR DISEASES

- Stroke is the clinical term for acute-onset neurologic deficits that last longer than 24 hours and that are a consequence of vessel occlusion or vessel rupture.
- Cerebral infarction follows loss of blood supply and can be focal, widespread, or restricted to regions with the least robust vascular supply (boundary zones).

- Large artery infarcts are most commonly embolic; with subsequent dissolution of the embolus and reperfusion, a nonhemorrhagic infarct can become hemorrhagic.
- Primary intraparenchymal hemorrhages lead to infarction of the adjacent brain parenchyma and are typically either ganglionic (most commonly due to hypertension) or lobar (most commonly due to CAA).
- Hyaline arteriolosclerosis (small vessel disease), which is caused by a long-standing hypertension, can lead to either lumen occlusion (resulting in a lacunar infarct) or wall rupture (resulting in a ganglionic hemorrhage).
- Spontaneous subarachnoid hemorrhage is usually caused by a structural vascular abnormality, such as an aneurysm or arteriovenous malformation.

INFECTIONS

Infection may damage the nervous system directly through injury of neurons or glia by the infectious agent, or indirectly through microbial toxins, the destructive effects of the inflammatory response, or as a result of immune-mediated mechanisms. There are four principal routes by which microbes enter the nervous system.

- *Hematogenous spread* is the most common; infectious agents ordinarily gain access through the arterial circulation, but retrograde venous spread can occur via anastomoses with veins of the face.
- *Direct implantation* of microorganisms is most often traumatic, but can be sometimes associated with congenital malformations (e.g., meningomyelocele) that provide ready access for microorganisms.
- *Local extension* can originate from infected adjacent structures, such as air sinuses, teeth, skull, or vertebrae.
- Viruses also may be transported along the *peripheral nervous system,* as occurs with rabies and herpes zoster viruses.

General aspects of the pathology of infectious agents are discussed in Chapter 8; distinctive forms of CNS infections are described herein (Table 28.2).

Acute Meningitis

Meningitis is an inflammatory process of the leptomeninges and CSF within the subarachnoid space, usually caused by an infection. *Meningoencephalitis* refers to inflammation of the meninges and brain parenchyma. Although infections are the most common causes of meningitis and meningoencephalitis, this reaction may also occur in response to a nonbacterial irritant introduced into the subarachnoid space *(chemical meningitis)* or in the context of a systemic autoimmune disease. Based on the etiology and clinical evolution of the illness, infectious meningitis is broadly classified into *acute pyogenic* (usually bacterial), *aseptic* (usually acute or subacute viral), and *chronic* (usually tuberculous, spirochetal, or cryptococcal). Each type is accompanied by characteristic changes in the CSF.

Table 28.2 Common Central Nervous System Infections

Type of Infection	Clinical Syndrome	Common Causative Organisms
Bacterial Infections		
Meningitis	Acute pyogenic meningitis	*Escherichia coli* or group B streptococci (infants) *Neisseria meningitidis* (young adults) *Streptococcus pneumoniae* or *Listeria monocytogenes* (older adults)
	Chronic meningitis	*Mycobacterium tuberculosis*
Localized infections	Abscess	Streptococci and staphylococci
	Empyema	Polymicrobial (staphylococci, anaerobic gram-negative)
Viral Infections		
Meningitis	Acute aseptic meningitis	Enteroviruses Influenza species Lymphocytic choriomeningitis virus
Encephalitis	Encephalitic syndromes	Herpes simplex (HSV-1, HSV-2) Cytomegalovirus Human immunodeficiency virus JC polyomavirus (progressive multifocal leukoencephalopathy)
	Arthropod-borne encephalitis	West Nile virus Eastern equine encephalitis virus Western equine encephalitis virus St. Louis encephalitis virus La Crosse encephalitis virus Venezuelan equine encephalitis virus Japanese encephalitis virus Tick-borne encephalitis virus
Brainstem and spinal cord syndromes	Rhombencephalitis	Rabies
	Acute flaccid myelitis/ poliomyelitis	Poliovirus West Nile virus Enterovirus D68
Spirochetes and Fungi		
Meningoencephalitic syndromes	Neurosyphilis	*Treponema pallidum*
	Lyme disease (neuroborreliosis)	*Borrelia burgdorferi*
	Fungal meningitis	*Cryptococcus neoformans* *Candida albicans*
Protozoa and Metazoa		
Encephalitis	Amebic encephalitis	*Balamuthia* and *Acanthamoeba* species *Naegleria* species
Localized infections	Toxoplasmosis	*Toxoplasma gondii*
	Cysticercosis	*Taenia solium*

Acute Pyogenic (Bacterial) Meningitis

Distinctive microorganisms cause acute pyogenic meningitis in various age groups: *Escherichia coli* and the group B streptococci in neonates; *Streptococcus pneumoniae* and *Listeria monocytogenes* in the elderly; and *Neisseria meningitidis* in adolescents and young adults, with clusters of cases raising public health concerns. The introduction of immunization against *Haemophilus influenzae* has markedly reduced the incidence of this infection in the developed world, particularly among infants (who used to be at high risk).

Affected individuals typically show systemic signs of infection superimposed on symptoms related to meningeal irritation and neurologic impairment, including headache, photophobia, irritability, clouding of consciousness, and neck stiffness. A spinal tap yields cloudy or frankly purulent CSF with as many as 90,000 neutrophils per cubic millimeter, increased CSF pressure and increased protein concentration, and markedly reduced glucose content. Untreated pyogenic meningitis can be fatal, while effective treatment with antibiotics markedly reduces mortality. One feared complication is the *Waterhouse-Frederichsen syndrome,* which results from meningitis-associated septicemia and hemorrhagic infarction of the adrenal glands (Chapter 24). It occurs most often with meningococcal and pneumococcal meningitis. In the immunosuppressed individual, purulent meningitis may be caused by several other infectious agents, such as *Klebsiella* or anaerobic organisms, and may have an atypical clinical course and uncharacteristic CSF findings, rendering timely diagnosis more difficult.

MORPHOLOGY

In acute meningitis, an exudate is evident within the leptomeninges over the surface of the brain (Fig. 28.22). The meningeal vessels are engorged and stand out prominently. The anatomic distribution of the exudate varies; in *H. influenzae* meningitis, for example, it is usually basal, whereas in pneumococcal meningitis it is often densest over the cerebral convexities near the sagittal sinus. From the areas of greatest accumulation, tracts of pus follow along blood vessels on the surface of the brain. When the meningitis is fulminant, the inflammation may extend to the ventricles, producing ventriculitis. The ventricles may also be the portal for CSF involvement by blood-borne infections because the choroid plexus lacks a blood-brain barrier.

On microscopic examination, neutrophils fill the subarachnoid space in severely affected areas and are found predominantly around the leptomeningeal blood vessels in less severe cases. Particularly in untreated meningitis, Gram stain reveals variable numbers of bacteria. In fulminant meningitis, the inflammatory cells infiltrate the walls of the leptomeningeal veins and may extend focally into the substance of the brain (cerebritis); secondary vasculitis and venous thrombosis may lead to hemorrhagic cerebral infarction.

Leptomeningeal fibrosis may follow pyogenic meningitis and cause hydrocephalus. Particularly in pneumococcal meningitis, large quantities of the capsular polysaccharide of the organism produce a gelatinous exudate that promotes arachnoid fibrosis, a condition referred to as *chronic adhesive arachnoiditis*.

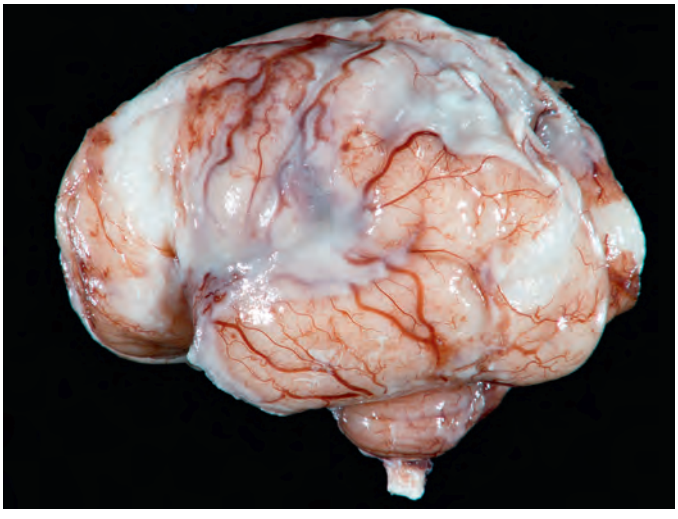


Figure 28.22 Pyogenic meningitis. A thick layer of suppurative exudate covers the brain surface and thickens the leptomeninges.

Acute Aseptic (Viral) Meningitis

Aseptic meningitis is a clinical term used for an absence of organisms by bacterial culture in a patient with manifestations of meningitis, including meningeal irritation, fever, and alterations of consciousness of relatively acute onset. The disease is generally of viral etiology, but may be bacterial, rickettsial, or autoimmune in origin. The clinical course is less fulminant than that of pyogenic meningitis, and the CSF findings also differ; in aseptic meningitis, there is a lymphocytic pleocytosis, the protein elevation is only moderate, and the glucose content is nearly always normal. The viral aseptic meningitides are usually self-limited and are treated symptomatically. Remarkably, the etiologic agent is identified in only a minority of cases; however, this may change through use of more sensitive and specific detection techniques (such as next-generation sequencing), which are under development. When pathogens are identified, enteroviruses are the most common etiology, accounting for 80% of cases. The spectrum of pathogens varies seasonally and geographically. An aseptic meningitis-like picture may also develop subsequent to rupture of an epidermoid cyst into the subarachnoid space or the introduction of a chemical irritant (chemical meningitis). The CSF is sterile in these cases, and there is pleocytosis with neutrophils and an increased protein concentration, but the sugar content is usually normal.

Acute Focal Suppurative Infections

Focal suppurative infections are typically caused by pyogenic bacteria or fungi, and can arise in several different compartments including the brain parenchyma, subdural space, and extradural space.

Brain Abscess

A brain abscess is a localized focus of necrosis of brain tissue with accompanying inflammation, usually caused by a bacterial infection. Brain abscesses may arise by direct implantation of organisms, local extension from adjacent

foci (mastoiditis, paranasal sinusitis), or hematogenous spread (usually from a primary site in the heart, lungs, or bones of the extremities, or secondary to bacteremia from dental procedures). Predisposing conditions include acute bacterial endocarditis, which may give rise to multiple brain abscesses; congenital heart disease with right-to-left shunting and loss of pulmonary filtration of organisms; chronic pulmonary sepsis, as in bronchiectasis; and systemic disease with immunosuppression. Streptococci and staphylococci are the most common offending organisms identified in nonimmunosuppressed patients.

MORPHOLOGY

Abscesses are discrete lesions with central liquefactive necrosis surrounded by brain swelling (Fig. 28.23). At the outer margin of the necrotic lesion, there is exuberant granulation tissue with neovascularization. The newly formed vessels are abnormally permeable, accounting for marked vasogenic edema in the adjacent brain tissue. In well-established lesions, a collagenous capsule is produced by fibroblasts derived from the walls of blood vessels. Outside the fibrous capsule is a zone of reactive gliosis containing numerous gemistocytic astrocytes.

Clinical Features

Cerebral abscesses are destructive lesions, and patients often present with progressive focal neurologic deficits; signs and symptoms related to increased intracranial pressure may also develop. Typically, the CSF has a high white cell count and an increased protein concentration, but the glucose content is normal. The source of infection may be apparent or may be traced to a small distant focus that is not symptomatic. The increased intracranial pressure can lead to fatal herniation; other complications include abscess rupture with ventriculitis or meningitis, and venous sinus thrombosis. With surgery and antibiotic treatment, the otherwise high mortality rate can be reduced to less than 10%.

Subdural Empyema

Bacterial, and rarely fungal, infections of the skull bones or air sinuses can spread to the subdural space, producing a subdural empyema. Although the underlying arachnoid and subarachnoid spaces are usually not affected, a large subdural empyema with mass effect and/or thrombophlebitis of the bridging veins can cause venous occlusion and infarction of the brain. In addition to symptoms referable to the source of the infection, most patients are febrile and have headache and neck stiffness. The CSF profile is similar to that seen in brain abscesses because both are parameningeal infectious processes. If untreated, focal neurologic signs, lethargy, and coma may develop. With prompt diagnosis and treatment, including surgical drainage, resolution and full recovery is possible, the only residuum being a thickened dura.

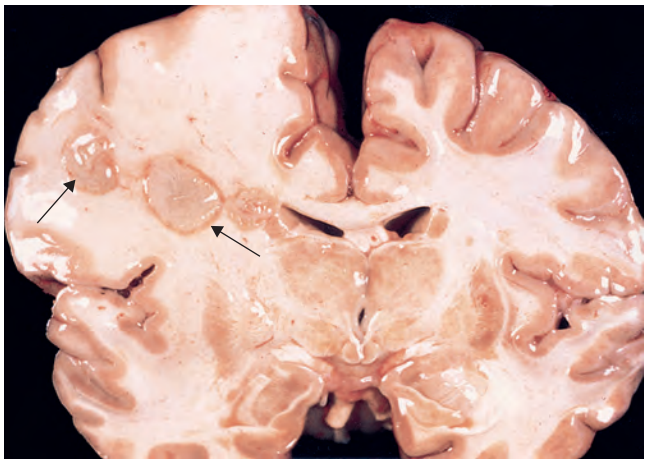


Figure 28.23 Cerebral abscesses *(arrows)*.

Extradural Abscess

Extradural abscess, commonly associated with osteomyelitis, often arises from an adjacent focus of infection, such as sinusitis, or following a surgical procedure. When the process occurs in the spinal epidural space, it may cause spinal cord compression and constitute a neurosurgical emergency.

Chronic Bacterial Meningoencephalitis

Chronic bacterial infection of the meninges and the brain may be caused by *Mycobacterium tuberculosis*, *Treponema pallidum*, and *Borrelia* species.

Tuberculosis

Tuberculosis of the CNS may be part of active disease elsewhere in the body, or appear in isolation following seeding from silent lesions elsewhere, usually the lungs. It may involve the meninges or the brain.

MORPHOLOGY

The most common pattern of tuberculous involvement is a diffuse **meningoencephalitis.** The subarachnoid space contains a gelatinous or fibrinous exudate that characteristically involves the base of the brain, effacing the cisterns and encasing cranial nerves. There may be discrete, white areas of inflammation scattered over the leptomeninges. On microscopic examination, involved areas contain a mixed inflammatory infiltrate with lymphocytes, plasma cells, and macrophages. Florid cases show well-formed granulomas with caseous necrosis and giant cells. Arteries running through the subarachnoid space may show obliterative endarteritis and marked intimal thickening. Organisms can often be seen with acid-fast stains. The infectious process may spread to the choroid plexus and ependymal surface, traveling through the CSF. In long-standing cases, a dense, fibrous adhesive arachnoiditis may develop, most conspicuous around the base of the brain. Hydrocephalus may result.

CNS involvement may also take the form of one or more well-circumscribed intraparenchymal masses **(tuberculomas),** which may be associated with meningitis. A tuberculoma may be as large as several centimeters in diameter, causing significant mass effect. These granulomas usually have central caseous necrosis; calcification may occur in inactive lesions.

Clinical Features

Patients with tuberculous meningitis usually have headache, malaise, mental confusion, and vomiting. The CSF typically shows a pleocytosis made up of mononuclear cells or a

mixture of neutrophils and mononuclear cells, an elevated protein concentration (often strikingly so), and a moderately reduced or normal glucose. The most serious complications of chronic tuberculous meningitis are arachnoid fibrosis producing hydrocephalus and obliterative endarteritis producing arterial occlusion and brain infarction. When the process involves the spinal cord subarachnoid space, nerve roots may also be affected. Tuberculomas produce symptoms typical of space-occupying brain lesions and must be distinguished from CNS tumors.

CNS tuberculosis in patients with acquired immunodeficiency syndrome (AIDS) is pathologically similar, but there may be less host reaction than in immunocompetent individuals.

Neurosyphilis

Neurosyphilis is a manifestation of the tertiary stage of syphilis and occurs in only about 10% of individuals with untreated infection. The major patterns of CNS involvement are meningovascular neurosyphilis, paretic neurosyphilis, and tabes dorsalis. Affected individuals often show an incomplete or mixed picture, most commonly the combination of tabes dorsalis and paretic disease (taboparesis). Because of impaired cell-mediated immunity, individuals infected with HIV are at increased risk for neurosyphilis, particularly acute syphilitic meningitis or meningovascular disease; the rate of disease progression and severity are also accelerated.

MORPHOLOGY

Neurosyphilis presents in several distinct forms.

- **Meningovascular neurosyphilis** is chronic meningitis involving the base of the brain and more variably the cerebral convexities and spinal leptomeninges. In addition, there may be an associated obliterative endarteritis (Heubner arteritis) accompanied by a distinctive perivascular inflammatory reaction rich in plasma cells and lymphocytes. Cerebral gummas (plasma cell-rich mass lesions) may also occur in the meninges and extend into the parenchyma.
- **Paretic neurosyphilis** is caused by invasion of the brain by *T. pallidum* and manifests as insidious but progressive cognitive impairment associated with mood alterations (including delusions of grandeur) that terminate in severe dementia **(general paresis of the insane).** Parenchymal damage of the cerebral cortex is particularly common in the frontal lobe, but also occurs in other areas of the isocortex. The lesions are characterized by loss of neurons, proliferation of microglia, gliosis, and iron deposits; the latter are demonstrable with the Prussian blue stain perivascularly and in the neuropil, and are presumably the sequelae of small bleeds stemming from microvascular damage. The spirochetes can, at times, be demonstrated in tissue sections.
- **Tabes dorsalis** is the result of damage to the sensory axons in the dorsal roots. This causes impaired joint position sense and ataxia (locomotor ataxia); loss of pain sensation, leading to skin and joint damage (Charcot joints); other sensory disturbances, particularly the characteristic "lightning pains"; and absence of deep tendon reflexes. On microscopic examination, there is loss of both axons and myelin in the dorsal roots, with corresponding pallor and atrophy in the dorsal columns of the spinal cord. Organisms are not demonstrable in the cord lesions.

Neuroborreliosis (Lyme Disease)

Lyme disease is caused by the spirochete *Borrelia burgdorferi,* which is transmitted by various species of *Ixodes* tick (Chapter 8). Involvement of the nervous system is referred to as *neuroborreliosis*. Neurologic symptoms are highly variable and include aseptic meningitis, facial nerve palsies and other polyneuropathies, as well as encephalopathy. The rare cases that have come to autopsy have shown a focal proliferation of microglial cells in the brain as well as scattered extracellular organisms.

Viral Meningoencephalitis

Viral encephalitis is a parenchymal infection of the brain almost invariably associated with meningeal inflammation (meningoencephalitis) and sometimes with simultaneous involvement of the spinal cord (encephalomyelitis).

Some viruses have a propensity to infect the nervous system. Such neural tropism takes several forms: some infect specific cell types (e.g., oligodendrocytes), while others preferentially involve particular areas of the brain (e.g., medial temporal lobes or the limbic system). Latency is an important phase of several viral infections of the CNS (e.g., herpes zoster and progressive multifocal leukoencephalopathy). Systemic viral infections in the absence of direct evidence of viral penetration into the CNS may be followed by an immune-mediated disease, such as perivenous demyelination. Viral infection of the fetus may cause congenital malformations, as occurs with rubella and Zika virus. A slowly progressive degenerative disease syndrome may follow many years after a viral illness; an example is postencephalitic parkinsonism after the 1918 viral influenza pandemic.

Arthropod-Borne Viral Encephalitis

Arboviruses are an important cause of epidemic encephalitis, especially in tropical regions of the world, and are capable of causing serious morbidity and high mortality. In the Western hemisphere, the most important arboviruses are Eastern and Western equine, West Nile, Venezuelan, St. Louis, and La Crosse; elsewhere in the world, pathogenic arboviruses include Japanese B (Far East), Murray Valley (Australia and New Guinea), and tick-borne (Russia and Eastern Europe).

All of these viruses have animal hosts and insect vectors. Clinically, affected individuals develop generalized neurologic deficits, such as seizures, confusion, delirium, and stupor or coma, as well as focal signs, such as reflex asymmetry and ocular palsies. Involvement of the spinal cord in West Nile encephalitis can lead to a polio-like syndrome with paralysis. In general, the CSF is usually colorless, with slightly elevated pressure, an elevated protein level, and a normal glucose. Initially the CSF exhibits a neutrophilic pleocytosis, but this rapidly converts to a lymphocytosis.

MORPHOLOGY

The encephalitides caused by various arboviruses produce similar histopathologic changes that differ only in severity and extent. Characteristically, there is a meningoencephalitis marked by the perivascular accumulation of lymphocytes (and sometimes

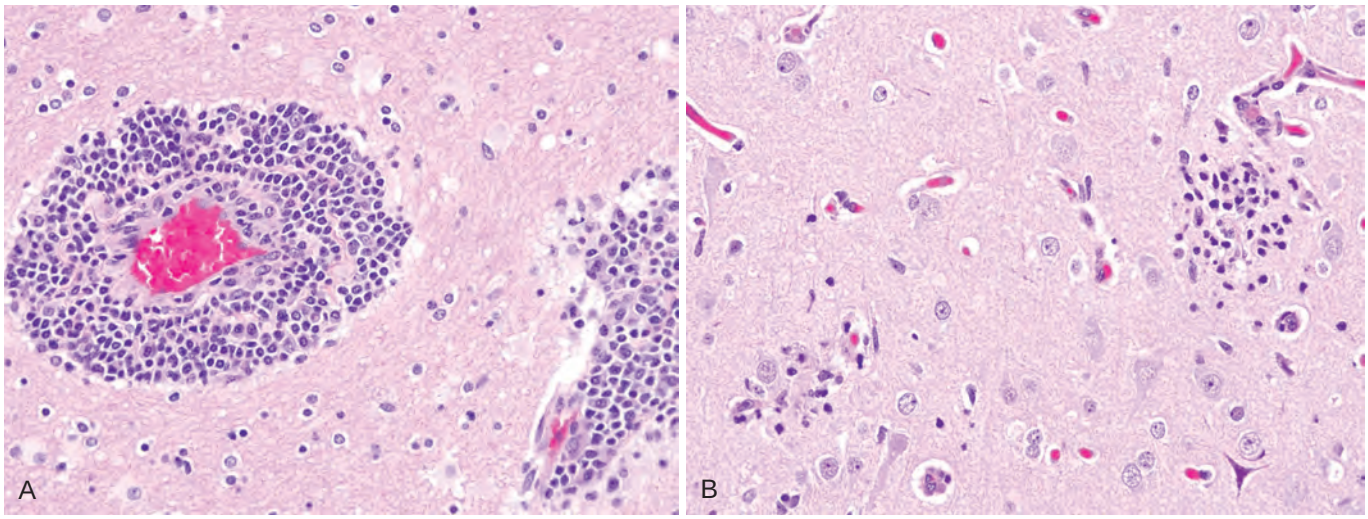


Figure 28.24 Characteristic findings of viral encephalitis include perivascular cuffs of lymphocytes (A) and microglial nodules (B).

neutrophils) (Fig. 28.24A). Multiple foci of necrosis of gray and white matter are found; in particular, there is evidence of single-cell neuronal necrosis with phagocytosis of the debris **(neuronophagia).** Microglial cells form small aggregates, called **microglial nodules** (see Fig. 28.24B). In severe cases, there may be a necrotizing vasculitis with associated focal hemorrhages. Although some viruses declare their presence by formation of intracellular inclusions, in tissue samples the causative virus is most often identified by a combination of ultrastructural, immunohistochemical, and molecular methods.

Herpes Simplex Virus Type 1

Herpes simplex virus type 1 (HSV-1) encephalitis occurs most commonly in children and young adults; only about 10% of affected individuals have a history of prior herpetic infection. The typical presenting symptoms are alterations in mood, memory, and behavior. Antiviral agents now provide effective treatment in many cases, with a significant reduction in the mortality rate. In some individuals, HSV-1 encephalitis follows a subacute course with clinical manifestations (weakness, lethargy, ataxia, seizures) that evolve during a more protracted period (4 to 6 weeks).

MORPHOLOGY

This encephalitis starts in and most severely involves the inferior and medial regions of the temporal lobes and the orbital gyri of the frontal lobes (Fig. 28.25A). The infection is necrotizing and often hemorrhagic in the most severely affected regions. Perivascular inflammatory infiltrates are usually present, and Cowdry type A intranuclear viral inclusions may be found in both neurons and glia (Fig. 28.25B). In individuals with slowly evolving HSV-1 encephalitis, there is more diffuse involvement of the brain.

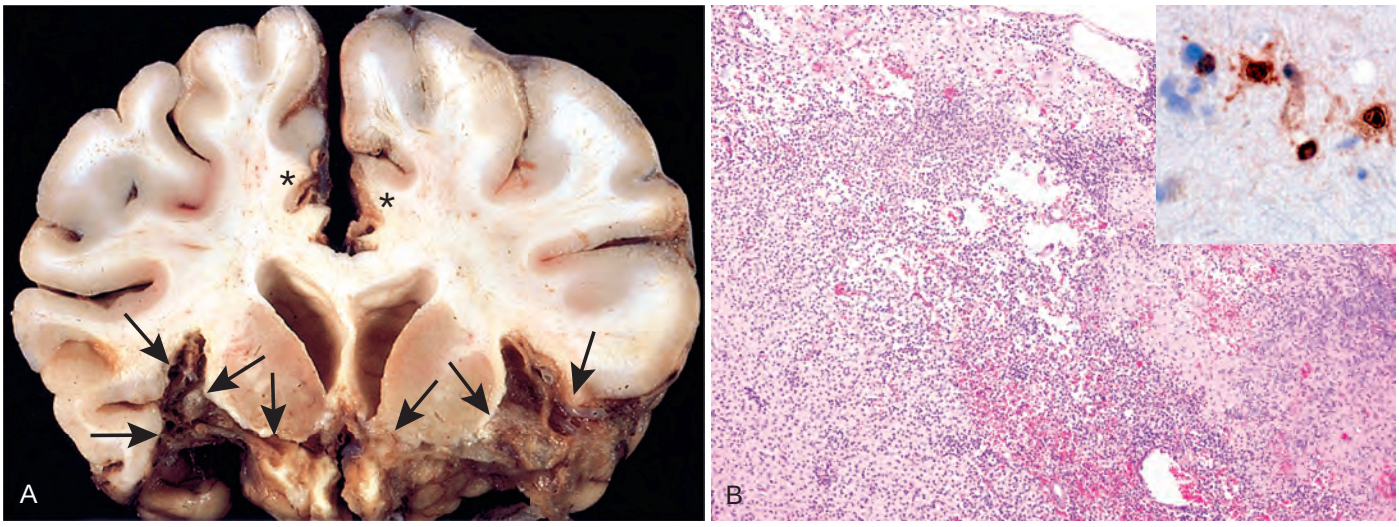


Figure 28.25 (A) Herpes encephalitis showing extensive destruction of inferior frontal and anterior temporal lobes *(arrows)* and the cingulate gyri *(asterisks)*. (B) Necrotizing inflammatory process characterizes acute herpes encephalitis; nuclear viral inclusions can be highlighted by immunostaining *(inset)*. (A, Courtesy Dr. T.W. Smith, University of Massachusetts Medical School, Worcester, Mass.)

Herpes Simplex Virus Type 2

Herpes simplex virus type 2 (HSV-2) can infect the nervous system. In adults it causes meningitis, but as many as 50% of neonates born by vaginal delivery to women with active primary HSV genital infections acquire the infection during passage through the birth canal and develop severe encephalitis. In individuals with active HIV infection, HSV-2 may cause an acute hemorrhagic and necrotizing encephalitis.

Varicella-Zoster Virus

Primary infection with varicella-zoster virus causes one of the childhood exanthems *(chickenpox),* without evidence of neurologic involvement. Following the cutaneous infection, the virus enters a latent phase within sensory neurons of the dorsal root or trigeminal ganglia. Reactivation of infection in adults *(shingles* or *herpes zoster)* usually manifests as a painful, vesicular skin eruption confined to one or several dermatomes. Herpes zoster is typically self-limited, but there may be a postherpetic neuralgia syndrome (particularly after 60 years of age) characterized by persistent pain, sometimes induced by stimuli that are otherwise not painful. Vaccination is now available to prevent these complications of varicella-zoster virus reactivation.

Cytomegalovirus

CMV infection of the nervous system occurs in fetuses and immunosuppressed individuals. The outcome of infection in utero is periventricular necrosis that produces severe brain destruction followed later by microcephaly and periventricular calcification. CMV is a common opportunistic viral pathogen in individuals with AIDS, with CNS involvement also occurring in this setting.

MORPHOLOGY

In the immunosuppressed individual, CMV most commonly causes subacute encephalitis, which may be associated with CMV inclusion-bearing cells (see Fig. 8.12). The infection tends to localize in the subependymal regions of the brain, where it results in a severe hemorrhagic necrotizing ventriculoencephalitis and choroid plexitis. The virus can also attack the lower spinal cord and roots, producing a painful radiculoneuritis. Any cell in the CNS (neurons, glia, ependyma, or endothelium) may be infected. Prominent enlarged cells with intranuclear and intracytoplasmic inclusions can be readily identified by conventional light microscopy; CMV infection is confirmed by immunohistochemistry.

Poliomyelitis

Although paralytic poliomyelitis has been eradicated by vaccination in many parts of the world, there are still a few countries where it remains a serious problem. In addition, a cluster of cases of acute flaccid myelitis (a polio-like syndrome) has recently occurred in the United States and seems to have been caused by a novel strain of enterovirus D68 (which usually causes only a mild respiratory infection). In nonimmunized individuals, poliovirus infection causes a subclinical or mild gastroenteritis, similar to that caused by other members of the picornavirus group of enteroviruses. In a small fraction of the vulnerable population, however, the virus secondarily invades the nervous system.

MORPHOLOGY

Acute cases show mononuclear cell perivascular cuffs and neuronophagia of the **anterior horn motor neurons of the spinal cord.** The inflammatory reaction is usually confined to the anterior horns but may extend into the posterior horns, and the damage is occasionally severe enough to produce cavitation. Poliovirus RNA has been detected in anterior horn cell motor neurons; the cranial motor nuclei are sometimes involved as well. Postmortem examination in long-term survivors of symptomatic poliomyelitis shows loss of neurons and gliosis in the affected anterior horns of the spinal cord, atrophy of the anterior (motor) spinal roots, and neurogenic atrophy of denervated muscle.

Clinical Features

CNS infection manifests initially with meningeal irritation and a CSF picture consistent with aseptic meningitis; the disease may progress no further or may advance to involve the spinal cord. When the disease affects the motor neurons of the spinal cord, it produces a flaccid paralysis associated with muscle wasting and hyporeflexia in the corresponding region of the body—the permanent neurologic residue of poliomyelitis. Because of the destruction of motor neurons, paresis or paralysis follows; when the diaphragm and intercostal muscles are affected, severe respiratory compromise and even death may occur. A myocarditis sometimes complicates the acute infection. *Postpolio syndrome* can develop in patients 25 to 35 years after resolution of the initial illness. It is characterized by progressive weakness associated with decreased muscle mass and pain, and has been attributed to superimposed loss of remaining motor neurons, without any convincing evidence of viral re-activation.

Rabies

Rabies is severe encephalitis transmitted to humans by the bite of a rabid animal, usually a dog or various wild mammals that are natural reservoirs. Exposure to certain species of bats, even without a known bite, can also lead to rabies.

MORPHOLOGY

External examination of the brain shows intense edema and vascular congestion. Microscopically, there is widespread neuronal degeneration and an inflammatory reaction that is most severe in the brainstem; the basal ganglia, spinal cord, and dorsal root ganglia may also be involved. **Negri bodies,** the pathognomonic microscopic finding, are cytoplasmic, round to oval, eosinophilic inclusions found in pyramidal neurons of the hippocampus and Purkinje cells of the cerebellum, sites usually devoid of inflammation (Fig. 28.26). Rabies virus can be detected within Negri bodies by ultrastructural and immunohistochemical methods.

Clinical Features

Because the virus enters the CNS by ascending along the peripheral nerves from the wound site, the incubation period (usually between 1 and 3 months) depends on the distance between the wound and the brain. The disease begins with nonspecific symptoms such as malaise, headache, and fever, but the conjunction of these symptoms with local paresthesias around the wound is diagnostic. As the infection advances,

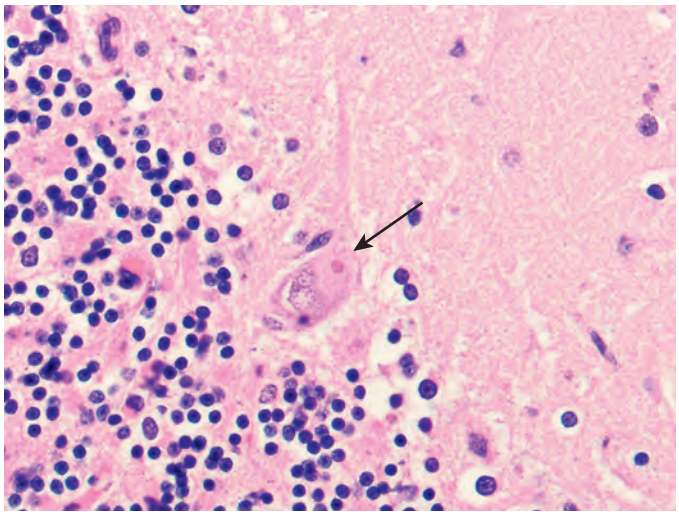


Figure 28.26 The diagnostic histologic finding in rabies is the eosinophilic Negri body, as seen here in a Purkinje cell *(arrow)*.

the affected individual exhibits extraordinary CNS excitability; the slightest touch is painful and produces violent motor responses or even convulsions. Contracture of the pharyngeal musculature on swallowing produces foaming at the mouth, which may create an aversion to swallowing even water (hydrophobia). There are signs of meningeal irritation and, as the disease progresses, flaccid paralysis. Alternating periods of mania and stupor progress to coma and eventually death ensues from respiratory failure.

Human Immunodeficiency Virus

In the period before the availability of effective antiretroviral therapy, neuropathologic changes were demonstrated at postmortem examination in as many as 80% to 90% of cases of AIDS. These changes stem from direct effects of the virus on the nervous system, opportunistic infections, and primary CNS lymphoma, most commonly an EBV-positive B-cell tumor. There has been a decrease in the frequency of these secondary effects of HIV infection thanks to the efficacy of multidrug antiretroviral therapy.

HIV aseptic meningitis occurs within 1 to 2 weeks of seroconversion in about 10% of patients; antibodies to HIV can be demonstrated, and the virus can be isolated from the CSF. The neuropathologic studies of the early, acute phases of HIV invasion of the CNS have shown mild lymphocytic meningitis, perivascular inflammation, and some myelin loss. Among CNS cell types, only microglia express both the CD4 coreceptor and the chemokine receptors (CCR5 or CXCR4) that are required in combination for efficient infection by HIV. During the chronic phase, HIV encephalitis is commonly found when symptomatic individuals come to autopsy.

An *immune reconstitution inflammatory syndrome* (IRIS) has been identified in patients with AIDS after effective treatment; the syndrome is recognized as a paradoxical deterioration after starting therapy, and consists of an exuberant "reconstituted" inflammatory response while on antiretroviral therapy (Chapter 6). In the CNS of patients with opportunistic infections, IRIS has caused paradoxical exacerbation of symptoms; neuropathologic studies confirm intense inflammation with an influx of CD8+ lymphocytes.

MORPHOLOGY

HIV encephalitis is a chronic inflammatory reaction associated with widely distributed **microglial nodules,** often containing macrophage-derived **multinucleated giant cells** (Fig. 28.27); foci of tissue necrosis and reactive gliosis are sometimes seen together with these lesions. Some of the microglial nodules are found near small blood vessels, which show abnormally prominent endothelial cells and perivascular foamy or pigment-laden macrophages. These changes are especially prominent in the subcortical white matter, diencephalon, and brainstem. In some cases, there is also a disorder of white matter characterized by multifocal or diffuse areas of myelin pallor, axonal swelling, and gliosis. HIV can be detected in CD4+ microglia and mononuclear or multinucleated macrophages.

Clinical Features

Cognitive changes, some mild and others florid enough to be termed *HIV-associated dementia,* appear to have persisted in the era of effective anti-HIV treatment regimens; collectively, they are termed *HIV-associated neurocognitive disorders* (HAND). Rather than having a specific pathologic lesion as its correlate, HAND is most closely related to inflammatory activation of microglial and perivascular macrophages, some of which are infected by HIV. A wide range of possible mechanisms for neuronal dysfunction and injury have been proposed, including the actions of inflammatory cytokines, a cascade of toxic effects of HIV-derived proteins, neurotoxic effects of anti-HIV therapy, accelerated aging, and aberrant synaptic pruning; in all probability, most (if not all) of these mechanisms contribute to the pathogenesis of this disorder.

Progressive Multifocal Leukoencephalopathy

PML is an encephalitis caused by the JC polyomavirus; because the virus preferentially infects oligodendrocytes, demyelination is its principal pathologic effect. The disease occurs almost exclusively in immunosuppressed individuals in various clinical settings, including chronic lymphoproliferative or myeloproliferative illnesses, immunosuppressive chemotherapy

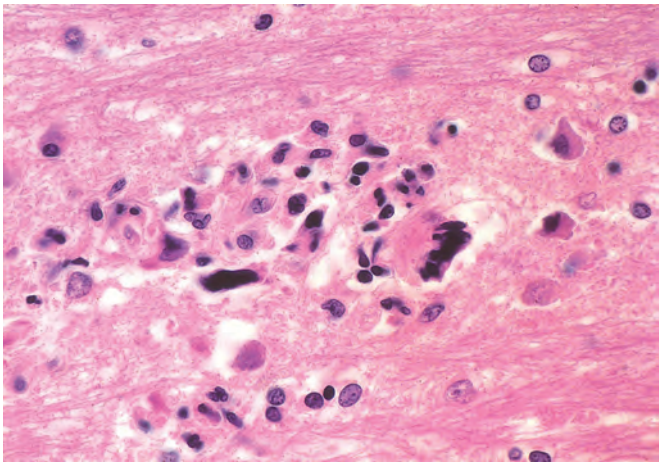


Figure 28.27 HIV encephalitis. Note the microglial nodule and multinucleated giant cells.

(including monoclonal antibody therapy targeting certain integrins), granulomatous diseases, and AIDS.

Although most people have serologic evidence of exposure to JC virus by 14 years of age, primary infection is asymptomatic; PML results from the reactivation of virus in the setting of immunosuppression. Clinically, affected individuals develop focal and relentlessly progressive neurologic symptoms and signs, with imaging studies showing extensive, often multifocal, cerebral or cerebellar white matter lesions.

MORPHOLOGY

The lesions consist of patches of irregular, ill-defined white matter injury that range in size from millimeters to large confluent regions (Fig. 28.28A). Microscopically, a PML lesion shows an area of demyelination, most often in a subcortical location, in the center of which are sheets of lipid-laden macrophages and a reduced number of axons (Fig. 28.28B). Greatly enlarged oligodendrocyte nuclei containing glassy amphophilic viral inclusions (Fig. 28.28B, *inset*), which can be identified by immunohistochemistry, are typically found at the lesion edge. Bizarre giant astrocytes with one to several irregular, hyperchromatic nuclei are intermixed with more typical reactive astrocytes. Infection of granule cell neurons in the cerebellum has been demonstrated in rare instances.

Fungal Meningoencephalitis

Fungal infections of the CNS are encountered primarily in immunocompromised individuals. The brain is usually involved following widespread hematogenous dissemination of fungi; the most frequent offenders are *Candida albicans,* Mucor species, *Aspergillus fumigatus,* and *Cryptococcus neoformans*. In endemic areas, pathogens such as *Histoplasma capsulatum*, *Coccidioides immitis*, and *Blastomyces dermatitidis* may involve the CNS after a primary pulmonary or cutaneous infection; again, this often follows immunosuppression. Although most fungi reach the brain by hematogenous dissemination, direct extension may also occur, particularly in mucormycosis in the setting of diabetes mellitus.

The three main forms of injury in CNS fungal infection are chronic meningitis, vasculitis, and parenchymal invasion. Vasculitis is most frequently seen with *Mucormycosis* and *Aspergillosis,* both of which directly invade blood vessel walls, but it occasionally occurs with other infections such as candidiasis. The resultant vascular thrombosis produces infarction that is often strikingly hemorrhagic.

Parenchymal infection, usually in the form of granulomas or abscesses, can occur with most of the fungi and often coexists with meningitis. The most commonly encountered fungi that invade the brain are *Candida* and *Cryptococcus*. Candidiasis usually produces multiple microabscesses, with or without granuloma formation; in contrast, parenchymal aggregates of *Cryptococcal* organisms are typically found within expanded perivascular (Virchow-Robin) spaces and are associated with minimal to no inflammation or gliosis.

Cryptococcal meningitis, a common opportunistic infection in the setting of AIDS, may be fulminant and fatal in as little as 2 weeks or indolent, evolving over months or years. The CSF may contain few cells but usually has a high concentration of protein. The mucoid-encapsulated yeasts can be visualized in the CSF with special stains or detected indirectly using assays for cryptococcal antigens (see Fig. 8.43).

Most cases of cryptococcal infection in immunosuppressed individuals are caused by *C. neoformans*. Recently, a second species, *C. gattii,* has been recognized and appears more likely to cause disease in immunocompetent individuals. Some case studies suggest that *C. gattii* involvement of the CNS is more likely to take the form of mass lesions ("cryptococcomas") and to present with symptoms related to increased CNS pressure due to mass effects.

Other Infectious Diseases of the Nervous System

Protozoal diseases (including malaria, toxoplasmosis, amebiasis, and trypanosomiasis), rickettsial infections (e.g.,

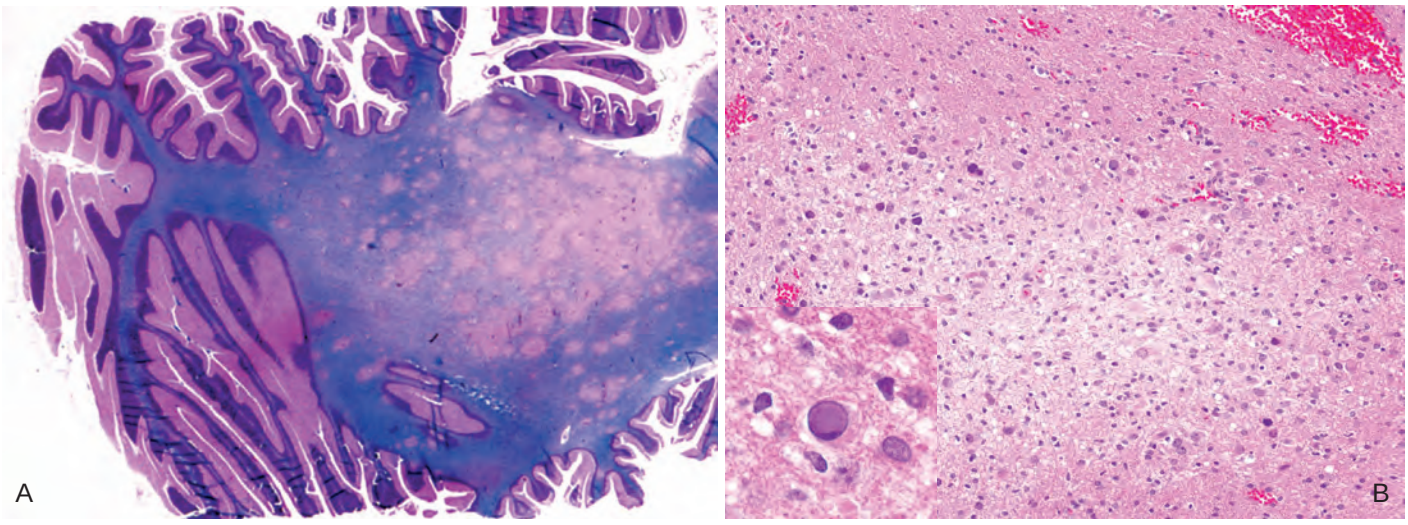


Figure 28.28 (A) Progressive multifocal leukoencephalopathy. Section stained for myelin showing irregular, poorly defined areas of demyelination, which become confluent in places. (B) Microscopically, the lesions consists of areas of demyelination. *Inset,* Enlarged oligodendrocyte nucleus represents the effect of viral infection.

typhus and Rocky Mountain spotted fever), and metazoal diseases (especially cysticercosis and echinococcosis) may also involve the CNS and are discussed in Chapter 8.

- *Cerebral toxoplasmosis* is an opportunistic infection commonly found in the setting of HIV-associated immunosuppression. The clinical symptoms of infection of the brain by *Toxoplasma gondii* are subacute, evolving during a 1- or 2-week period, and may be both focal and diffuse. Computed tomography and magnetic resonance imaging studies may show multiple ring-enhancing lesions; however, this radiographic appearance is not specific, as CNS lymphoma, tuberculosis, and fungal infections produce similar findings. In nonimmunosuppressed hosts, the impact of toxoplasmosis on the brain is most often seen when primary maternal infection occurs early in pregnancy. Such infections often spread to the brain of the developing fetus and cause severe damage in the form of multifocal necrotizing lesions that may calcify. With early diagnosis, cerebral toxoplasmosis is often treatable with antibiotics, which may also be administered empirically if this infection is in the differential diagnosis.
- *Cerebral amebiasis.* A rapidly fatal necrotizing encephalitis results from infection with *Naegleria* species, and a chronic granulomatous meningoencephalitis has been associated with infection with *Acanthamoeba.* The amebae may be difficult to distinguish morphologically from activated macrophages (Fig. 28.29). Methenamine silver or PAS stains are helpful in visualizing the organisms, although definitive identification ultimately depends on immunofluorescence studies, culture, and molecular methods.
- *Cerebral malaria.* Cerebral malaria is a complication of infection by *Plasmodium falciparum,* the species with the highest mortality. Most likely the result of sticking of infected red cells to inflamed vascular endothelium, cerebral involvement by malaria is accompanied by reduced cerebral blood flow and results in ataxia, seizures, and coma in the acute phase, and leads to long-term cognitive deficits in up to 20% of affected children (Chapter 8).

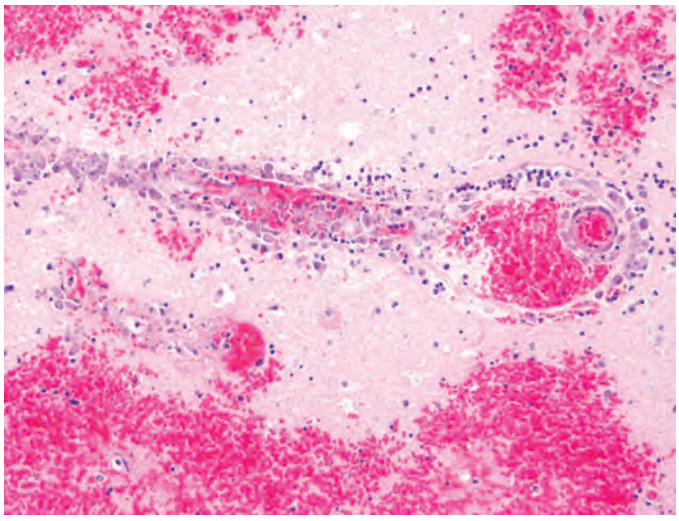


Figure 28.29 Necrotizing amebic meningoencephalitis; in this image, the organisms are largely perivascular. (*Acanthamoeba* resemble macrophages on routine stains, but are slightly smaller and less granular.)

MORPHOLOGY

Toxoplasmosis of the CNS produces brain abscesses, which are found most often in the cerebral cortex (near the gray-white junction) and deep gray nuclei, less often in the cerebellum and brainstem, and rarely in the spinal cord (Fig. 28.30A). Acute lesions exhibit central necrosis, petechial hemorrhages surrounded by acute and chronic inflammation, macrophage infiltration, and vascular proliferation. Both free tachyzoites and encysted bradyzoites may be found at the periphery of the necrotic foci. The organisms are often seen on routine H&E or Giemsa stains, but are more easily recognized by immunohistochemical methods (Fig. 28.30B). The blood vessels near these lesions often show intimal proliferation or frank vasculitis with fibrinoid necrosis and thrombosis.

KEY CONCEPTS

INFECTIONS

- Pathogens from viruses through parasites can infect the brain. Different pathogens use distinct routes to reach the brain and cause different patterns of disease.
- Routes of access of organisms to the brain include: hematogenous spread (e.g., abscess formation in the setting of endocarditis), direct extension (following trauma or with extension from the sinuses), and retrograde transport along nerves (e.g., rabies).
- Bacterial infections may cause meningitis, cerebral abscesses, or a chronic meningoencephalitis. The distribution of pathogens is influenced by various host factors, such as age and level of immune function.
- Viral infections can cause meningitis or meningoencephalitis. Some viruses have characteristic patterns of infection (HSV-1 in the temporal lobes, polio in the anterior horn).
- HIV can directly cause meningoencephalitis, or indirectly affect the brain by increasing the risk of opportunistic infections (toxoplasmosis, CMV) and EBV-positive CNS lymphoma.

DEMYELINATING DISEASES

Demyelinating diseases of the CNS are acquired conditions characterized by preferential damage to myelin with relative preservation of axons. The clinical deficits, at least initially, are due to the effect of myelin loss on the transmission of electrical impulses along axons. The natural history of demyelinating diseases is determined, in part, by the limited capacity of the CNS to regenerate normal myelin and by the degree of secondary damage to axons that occurs as the disease runs its course.

Several pathologic processes can cause loss of myelin. These include immune-mediated destruction of myelin (as in multiple sclerosis [MS]) and infections (as in PML, described earlier). In addition, inherited disorders may affect synthesis or turnover of myelin components; these are termed *leukodystrophies* and are discussed with metabolic disorders.

Multiple Sclerosis (MS)

MS is an autoimmune demyelinating disorder characterized by distinct episodes of neurologic deficits that are

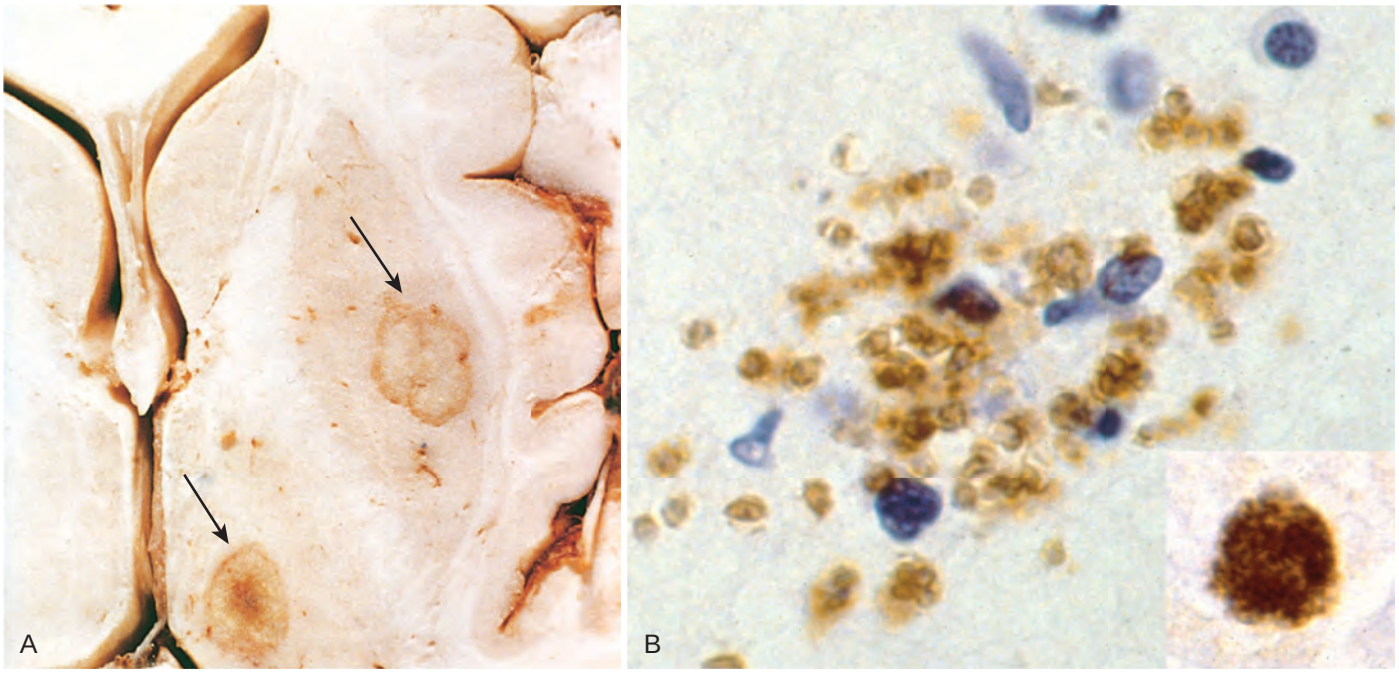


Figure 28.30 (A) Toxoplasma abscesses in the putamen and thalamus *(arrows).* (B) Free tachyzoites demonstrated by immunostaining; *inset, Toxoplasma* pseudocyst with bradyzoites highlighted by immunostaining.

separated in time and are attributable to patchy white matter lesions that are separated in space. It is the most common of the demyelinating disorders, having a prevalence of approximately 1 per 1000 persons in most of the United States and Europe. The disease may become clinically apparent at any age, although onset in childhood or after 50 years of age is relatively rare. Women are affected twice as often as are men. In most individuals with MS, the clinical course takes the form of relapsing and remitting episodes of variable duration (weeks to months to years) marked by neurologic defects, followed by gradual and partial recovery of neurologic function. The frequency of relapses tends to decrease over time, but there is a steady neurologic deterioration in most affected individuals.

Pathogenesis

MS is caused by an autoimmune response directed against components of the myelin sheath. As in other autoimmune disorders, the pathogenesis of this disease involves both genetic and environmental factors (Chapter 6). The incidence of MS is 15-fold higher when the disease is present in a first-degree relative and roughly 150-fold higher with an affected monozygotic twin. Despite a series of well-powered studies, only a portion of the genetic basis of the disease has been explained. There is a strong association with a DR haplotype of the major histocompatibility complex. Genome-wide association studies have identified additional associations with the IL-2 and IL-7 receptor genes, and subsequently with a number of other genes encoding proteins involved in the immune response, including cytokines and their receptors, costimulatory molecules, and cytoplasmic signaling molecules; many of these loci are associated with other autoimmune diseases. These genetic studies have not explained why the clinical course for individuals with MS is so variable.

Environmental factors are also important. For example, there is geographic variation in the prevalence of MS, with a higher number of cases diagnosed away from the equator; it has been proposed that this latitude dependence may be related to a low level of vitamin D (an immune system modulator) in people who are not exposed to sunlight during winter.

Immune mechanisms that underlie the destruction of myelin are the focus of much investigation. The available evidence indicates that the disease is initiated by Th1 and Th17 T cells that react against myelin antigens and secrete cytokines. Th1 cells secrete IFN-γ, which activates macrophages, while Th17 cells promote the recruitment of leukocytes (Chapter 6). The demyelination is caused by these activated leukocytes and their injurious products. The infiltrate in plaques and surrounding regions of the brain consists of T cells (mainly CD4+, some CD8+) and macrophages. How the autoimmune reaction is initiated is not understood; a role of viral infection (e.g., EBV) has been proposed but remains controversial.

Based on the growing understanding of the pathogenesis of MS, therapies are being developed that modulate or inhibit T-cell responses and block the recruitment of T cells to the brain. A potential contribution of humoral immunity has also long been suspected, based on the early observation of oligoclonal bands of immunoglobulin in CSF; the demonstration that treatment with agents that deplete B cells decreases the incidence of demyelinating lesions in patients with MS lends support to this idea.

MORPHOLOGY

MS is a white matter disease that is best appreciated in sections of the brain and spinal cord. In the fresh state, the lesions are

firmer than the surrounding white matter **("sclerosis")** and appear as well circumscribed, somewhat depressed, glassy, gray-tan, irregularly shaped **plaques** (Fig. 28.31A). The area of demyelination often has sharply defined borders, a feature best appreciated with stains for myelin (Fig. 28.31B). The size of lesions varies considerably, from small foci that are only recognizable microscopically to large confluent plaques. Plaques commonly occur adjacent to the lateral ventricles and are also frequent in the corpus callosum, optic nerves and chiasm, brainstem, ascending and descending fiber tracts, cerebellum, and spinal cord. Plaques can extend into gray matter because myelinated fibers are present there as well.

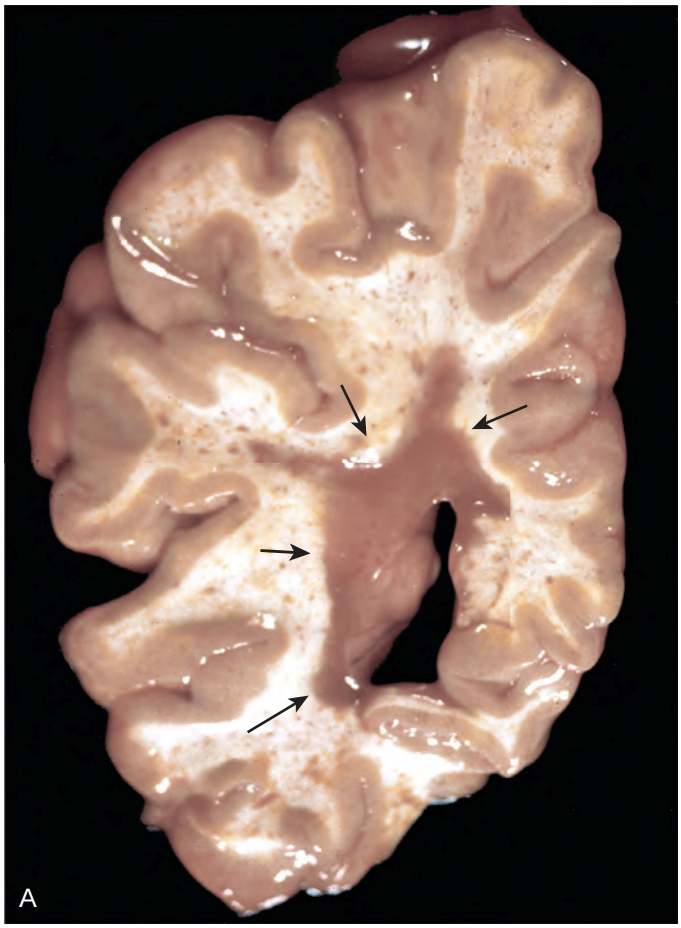
Microscopically, in an **active plaque,** there is ongoing myelin breakdown associated with abundant foamy macrophages (Fig. 28.32A); lymphocytes are also present, mostly as perivascular cuffs, especially at the outer edge of the lesion. Active lesions are often centered on small veins; myelin is usually completely absent (Fig. 28.32B), but axons are relatively preserved (Fig. 28.32C). In time, astrocytes undergo reactive changes. As lesions become quiescent, the inflammatory cells slowly disappear. Within **inactive plaques,** there is no macrophage-rich infiltrate, little to no myelin is found, and there is a reduction in the number of oligodendrocyte nuclei; instead, reactive gliosis is prominent. Axons in old gliotic plaques are usually greatly diminished in number.

In some MS plaques (shadow plaques), the white matter pallor is less severe, and abnormally thinned-out myelin sheaths can be demonstrated, especially at the outer edges. This phenomenon is most commonly interpreted as evidence of partial and incomplete remyelination by surviving oligodendrocytes. Abnormally myelinated fibers have also been observed at the edges of typical plaques. Although these histologic findings suggest a limited potential for remyelination in the CNS, the remaining axons within most MS plaques remain unmyelinated.

Clinical Features

Although MS lesions can occur anywhere in the CNS and consequently may induce a wide range of clinical manifestations, certain patterns of neurologic symptoms and signs are more common. Unilateral visual impairment due to involvement of the optic nerve *(optic neuritis)* is a frequent initial manifestation of MS. However, only a minority of individuals (10% to 50%, depending on the population studied) with an episode of optic neuritis go on to develop MS (which requires multiple episodes to support the diagnosis). Involvement of the brainstem produces cranial nerve signs (ataxia, nystagmus, and internuclear ophthalmoplegia from interruption of the fibers of the medial longitudinal fasciculus). Spinal cord lesions give rise to motor and sensory impairment of trunk and limbs, spasticity, and loss of bladder control.

Examination of the CSF in individuals with MS shows a mildly elevated protein level and in one-third of cases a moderate pleocytosis. IgG levels in the CSF are increased, and oligoclonal IgG bands are usually observed on immunoelectrophoresis; these are indicative of the presence of a small number of activated B-cell clones in the CNS, which are postulated to be self-reactive. Magnetic resonance imaging has taken on a prominent role in assessing disease progression; these studies, along with autopsy and clinical findings, indicate that some plaques may be clinically silent even in otherwise symptomatic patients. Treatment consists of several types of immunosuppressive or immunomodulatory



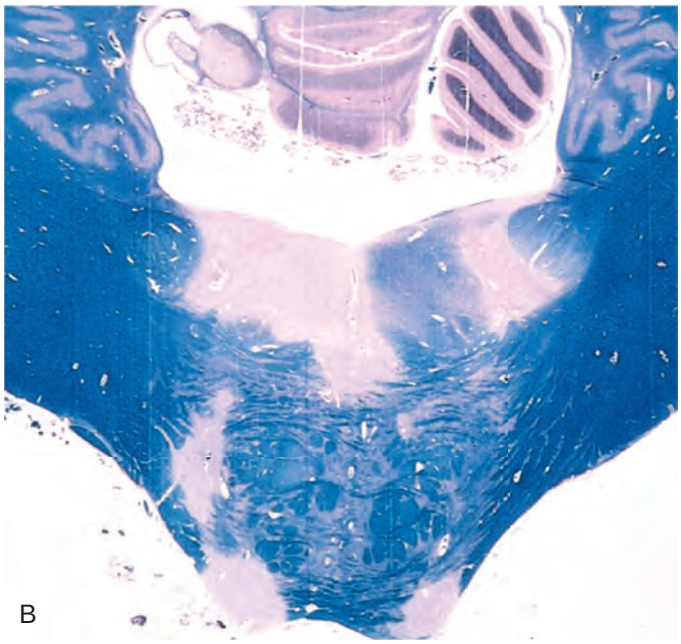


Figure 28.31 Multiple sclerosis. (A) Section of fresh brain shows gray-brown plaque around occipital horn of the lateral ventricle *(arrows)*. (B) Regions of demyelination (MS plaques) around the fourth ventricle lack the normal blue staining of myelin (Luxol fast blue periodic acid–Schiff stain).

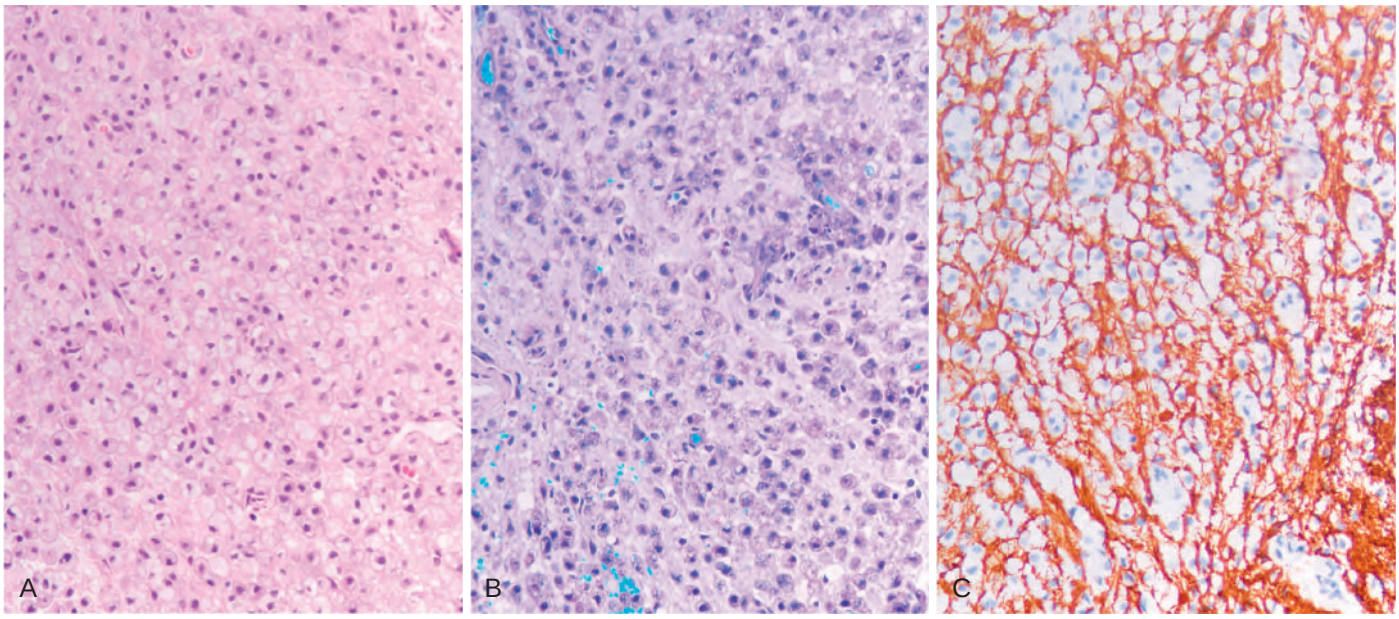


Figure 28.32 Multiple sclerosis. (A) Active demyelinating plaques appear very cellular due to the presence of numerous lipid-laden macrophages. (B) The same lesion stained with the Luxol fast blue periodic acid–Schiff stain shows a complete lack of myelin. (C) Relative preservation of axons is seen on the neurofilament immunostain (*brown*).

agents, which may slow the progression of the disease but are not curative.

Neuromyelitis Optica

Neuromyelitis optica (NMO) is a syndrome with synchronous (or near-synchronous) bilateral optic neuritis and spinal cord demyelination. Once considered a variant of MS, it is clearly a distinct disorder. NMO has an even greater skewing toward women than MS, is more commonly associated with poor recovery from the first attack, and is often characterized by the presence of aquaporin-4 antibodies, which appear to be pathogenic. Areas of demyelination in NMO show loss of aquaporin-4, the major water channel of astrocytes. White cells (often including neutrophils) are common in the CSF. Within the damaged areas of white matter, there is typically necrosis, an inflammatory infiltrate including neutrophils and eosinophils, and vascular immunoglobulin and complement deposition. Acute attacks are treated with glucocorticoids or plasma exchange, while long-term treatments include agents that decrease antibody titers (e.g., by depleting B cells) or that inhibit complement (e.g., inhibitors of the C5–C9 membrane attack complex).

Acute Disseminated Encephalomyelitis and Acute Necrotizing Hemorrhagic Encephalomyelitis

Acute disseminated encephalomyelitis is a diffuse, monophasic demyelinating disease that follows a viral infection or, rarely, a viral immunization. Symptoms typically develop 1 or 2 weeks after the antecedent event and include headache, lethargy, and coma rather than focal findings, as seen in MS; in contrast to MS, all brain lesions appear similar, consistent with the clinically monophasic nature of the disorder. The clinical course is rapid, and as many as 20% of those affected die; the remaining patients usually recover completely.

Acute necrotizing hemorrhagic encephalomyelitis (also known as acute hemorrhagic leukoencephalitis or Weston Hurst disease) is a fulminant syndrome of CNS demyelination, typically affecting young adults and children. The illness is almost invariably preceded by a recent episode of upper respiratory infection, most often of unknown cause. This disorder shows histologic similarities with acute disseminated encephalomyelitis; however, the damage is more severe. The disease is fatal in many patients, with significant deficits present in most survivors.

The lesions of acute disseminated encephalomyelitis are similar to those induced by immunization of animals with myelin components or with early rabies vaccines that had been prepared from brains of infected animals. This has led to the suggestion that acute disseminated encephalomyelitis is an acute autoimmune reaction to myelin and that acute necrotizing hemorrhagic encephalomyelitis is a hyperacute variant, but inciting antigens have yet to be identified.

Central Pontine Myelinolysis

Central pontine myelinolysis is an acute disorder characterized by loss of myelin in the base of the pons and portions of the pontine tegmentum, typically in a roughly symmetric pattern. It most commonly arises 2 to 6 days after rapid correction of hyponatremia, although it can also be associated with other severe electrolyte disturbances or osmolar imbalances, and is clinically known as osmotic demyelination syndrome. It appears that rapid increases in osmolality damage oligodendrocytes through uncertain mechanisms. Inflammation is absent from the lesions, and neurons and axons are well preserved. Because of the synchronous onset of damage, all lesions appear to be at the same stage of myelin loss and reaction. Although originally described in the pons, extra-pontine lesions with similar appearance and apparent etiology may also occur.

Although it can involve most parts of the brain, periventricular and subpial regions are spared, and it is extremely rare for the process to extend below the pontomedullary junction. The clinical presentation is a rapidly evolving quadriplegia, which may be fatal or lead to severe long-term deficits, including the "locked-in" syndrome, in which patients are fully conscious yet unresponsive. It is imperative that hyponatremia be corrected slowly and carefully to prevent this tragic complication.

KEY CONCEPTS

DEMYELINATING DISEASES

- Because of the critical role of myelin in nerve conduction, diseases of myelin can lead to widespread and severe neurologic deficits.
- Demyelinating diseases show evidence of breakdown and destruction of previously normal myelin, often by inflammatory processes. Secondary injury to axons typically emerges over time.
- MS, an autoimmune demyelinating disease primarily affecting young adults, is the most common disorder of myelin. It often pursues a relapsing-remitting course, with eventual progressive accumulation of neurologic deficits that is thought to reflect secondary axon loss.
- Less common forms of immune-mediated demyelination generally follow infections and have an acute, monophasic clinical course.

NEURODEGENERATIVE DISEASES

Neurodegenerative diseases are disorders characterized by the progressive loss of particular groups of neurons, which often have shared functions. Thus, the different diseases tend to involve particular neural systems and have relatively stereotypic presenting signs and symptoms.

The pathologic process that is common across most of the neurodegenerative diseases is the accumulation of protein aggregates (hence the occasional use of the term "proteinopathy"). Protein aggregates may arise because of mutations that alter the affected protein's conformation or disrupt the pathways that are involved in the processing or clearance of an otherwise normal protein. In other situations, there may be a subtle imbalance between protein synthesis and clearance (from genetic, environmental, or stochastic factors) that allows gradual accumulation of proteins.

Regardless of how they arise, the protein aggregates typically are resistant to degradation and show aberrant localization within neurons. The current evidence suggests that large (i.e., microscopically visible) protein aggregates are not toxic to cells; their formation appears to be an adaptive response that enables cells to sequester smaller (oligomeric) aggregates of the same proteins, which are directly toxic to neurons. However, as more and more protein is shunted into the aggregates, the normal function of the protein may also be lost, and this may also contribute to cell injury. Other recent evidence suggests that protein aggregates are capable of behaving like prions (see later); that is, aggregates derived from one cell may be taken up by another and provoke additional protein aggregation. The data supporting this concept are largely derived from experimental animal studies, but studies of patients who died with Alzheimer or Parkinson disease are consistent with the idea that these diseases spread from one site in the brain to another. However, only classic prion diseases have been shown to be truly transmissible.

The protein aggregates are recognized histologically as inclusions, which serve as diagnostic hallmarks. The basis for aggregation varies from one disease to another. It may be directly related to an intrinsic feature of a mutated protein (e.g., expanded polyglutamine repeats in Huntington disease [HD]), an intrinsic feature of a peptide derived from a larger precursor protein (e.g., Aβ in Alzheimer disease [AD]), or an unexplained alteration of a normal cellular protein (e.g., α-synuclein in sporadic Parkinson disease [PD]).

Neurodegenerative diseases vary with respect to the anatomic localization of involved areas and their specific cellular abnormalities (e.g., tangles, plaques, Lewy bodies). Accordingly, they can be classified using two different approaches:

- *Symptomatic/anatomic:* based on the anatomic regions that are most affected, which is typically reflected in the clinical symptoms (e.g., neocortical involvement results in cognitive impairment and dementia)
- *Pathologic:* based on the types of inclusions or abnormal structures observed (e.g., diseases with inclusions containing tau or containing synuclein)

Nevertheless, within the spectrum of degenerative diseases there is remarkable overlap in terms of characteristic neurologic deficits, functional/anatomic distribution of lesions, and cellular pathology (Tables 28.3 and 28.4). For the sake of simplicity, we will follow the time-honored classification based on the original description of these diseases.

Prion Diseases

Prion diseases are rapidly progressive neurodegenerative disorders caused by aggregation and intercellular spread of a misfolded prion protein (PrP); they may be sporadic, familial, or transmitted. Prion diseases include Creutzfeldt-Jakob disease (CJD), Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia, and kuru in humans; scrapie in sheep and goats; mink-transmissible encephalopathy; chronic wasting disease of deer and elk; and bovine spongiform encephalopathy. All of these diseases are characterized morphologically by "spongiform change" caused by intracellular vacuoles in neurons and glia, and clinically by a rapidly progressive dementia.

Pathogenesis

Prion diseases are conceptually important because they exemplify degenerative disorders that are caused by "spreading" of misfolded proteins, a remarkable phenomenon that allows a pathogenic protein to acquire some of the characteristics of an infectious organism. Normal PrP is a 30-kD cytoplasmic protein of unknown function. Disease occurs when PrP undergoes a conformational change from its normal α-helix–containing isoform (PrP^c) to an abnormal β-pleated sheet isoform, usually termed PrP^{sc} (for scrapie) (Fig. 28.33); associated with this conformational change, PrP acquires resistance to digestion with proteases, such as

Table 28.3 Features of the Major Neurodegenerative Diseases

Disease	Clinical Pattern	Inclusions	Genetic Causes
Prion diseases	Rapidly progressive dementia	Kuru plaques and diffuse PrP^{sc} deposits	PrP
Alzheimer disease (AD)	Dementia	Aβ (plaques) and tau (tangles)	APP, PS-1, PS-2, trisomy 21
Frontotemporal lobar degeneration (FTLD)	Behavioral changes, language disturbance	tau TDP-43 FUS	tau TDP-43, progranulin, C9orf72 FUS
Parkinson disease (PD)	Hypokinetic movement disorder with or without dementia	α-synuclein tau α-synuclein or none	α-synuclein (mutations or amplification) LRRK2 DJ-1, PINK1, parkin
Progressive supranuclear palsy (PSP)	Parkinsonism with abnormal eye movements	tau	tau
Corticobasal degeneration (CBD)	Parkinsonism with asymmetric movement disorder	tau	
Multiple system atrophy (MSA)	Parkinsonism, cerebellar ataxia, autonomic failure	α-synuclein	
Huntington disease (HD)	Hyperkinetic movement disorder	Huntington (polyglutamine)	Htt
Spinocerebellar ataxias (SCA-1, 2, 3, 6, 7, 17 and DRPLA)	Cerebellar ataxia	Various proteins (polyglutamine containing)	Multiple loci
Amyotrophic lateral sclerosis (ALS)	Weakness with upper and lower motor neurons signs	SOD1 TDP-43 FUS	SOD1 TDP-43, C9orf72 FUS
Spinal bulbar muscular atrophy (SBMA)	Lower motor neuron weakness, diminished androgen	Androgen receptor (polyglutamine containing)	Androgen receptor

proteinase K. Accumulation of PrP^{sc} in neural tissue seems to be the cause of the pathologic changes in these diseases, but how this material induces the development of cytoplasmic vacuoles and eventual neuronal death is still unknown. Immunostaining for PrP after partial digestion with proteinase K allows detection of PrP^{sc}, which is diagnostic.

The conformational change resulting in PrP^{sc} may occur spontaneously at an extremely low rate (resulting in sporadic cases of CJD) or at a higher rate if various mutations are present in PrP^{c}, such as occurs in familial forms of CJD, Gerstmann-Sträussler-Scheinker syndrome, and fatal familial insomnia. Independent of the means by which it originates, PrP^{sc} then facilitates, in a cooperative fashion, the conversion of other PrP^{c} molecules to PrP^{sc} molecules; it is this propagation of PrP^{sc} that accounts for the transmissible variants of prion diseases, which include iatrogenic CJD, variant CJD, and kuru. The suggestion that, at least within an individual, there may be cell-to-cell spread of disease-associated protein aggregates provides a link between prion diseases and other neurodegenerative disorders such as Alzheimer and Parkinson disease.

Table 28.4 Relationship Between Proteins and Neurodegenerative Diseases

Protein	Diseases With Inclusions
Aβ	Alzheimer disease
tau	Alzheimer disease Frontotemporal lobar degeneration Parkinson disease (with *LRRK2* mutations) Progressive supranuclear palsy Corticobasal degeneration Chronic traumatic encephalopathy
TPD-43	Frontotemporal lobar degeneration Amyotrophic lateral sclerosis
FUS	Frontotemporal lobar degeneration Amyotrophic lateral sclerosis
α-synuclein	Parkinson disease Multiple system atrophy
Polyglutamine aggregates (different protein in each disease)	Huntington disease Some forms of spinocerebellar ataxia Spinal bulbar muscular atrophy

Creutzfeldt-Jakob Disease (CJD)

The most common prion disease, CJD is a rare disorder that manifests clinically as a rapidly progressive dementia. The sporadic form of CJD has an annual incidence of approximately 1 per 1,000,000 people and accounts for about 90% of CJD cases; familial forms are caused by mutations in *PRNP,* the gene that encodes PrP. The disease has a peak incidence in the seventh decade. There are also well-established cases of iatrogenic transmission, notably by corneal or dural transplantation, deep implantation of electrodes in the brain, and administration of contaminated preparations of cadaveric human growth hormone. The onset is marked by subtle changes in memory and behavior followed by a rapidly progressive dementia, often associated with pronounced involuntary jerking muscle contractions on sudden stimulation (startle myoclonus). Signs of cerebellar dysfunction, usually manifested as ataxia, are present in a minority of affected individuals. The disease is uniformly fatal; the average survival is only 7 months after the onset of symptoms. A few patients have lived for several years, and these long-surviving cases show extensive atrophy of involved gray matter.

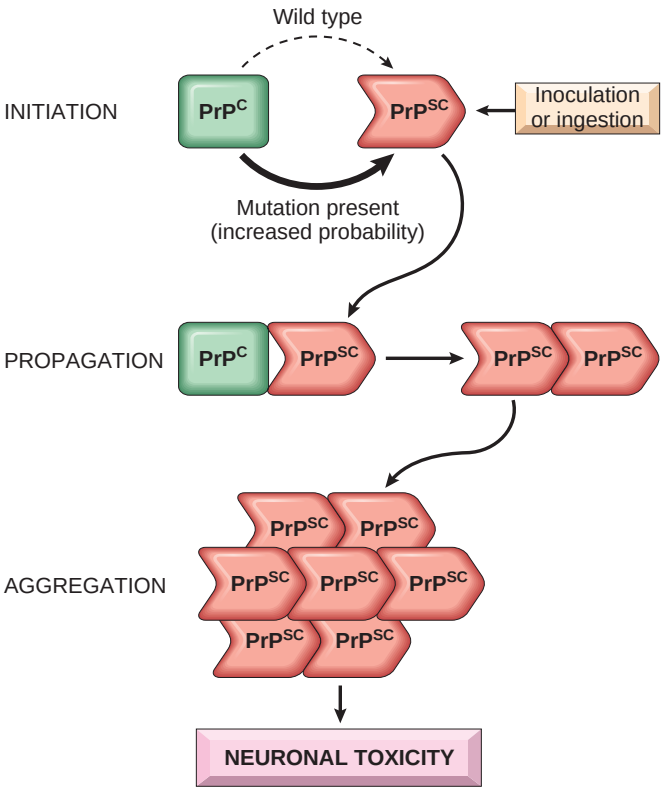


Figure 28.33 Pathogenesis of prion disease. α-helical PrP^c may spontaneously shift to the β-sheet PrP^{sc} conformation, an event that occurs at a much higher rate in familial diseases associated with germline PrP mutations. PrP^{sc} may also be acquired from exogenous sources, such as contaminated food, medical instrumentation, or medicines. Once present, PrP^{sc} converts additional molecules of PrP^c into PrP^{sc} through physical interaction, eventually leading to the formation of pathogenic PrP^{sc} aggregates.

Variant Creutzfeldt-Jakob Disease

Starting in 1995, a new CJD-like illness came to medical attention in the United Kingdom. This illness was different from classical CJD in several important respects: the disease affected young adults, behavioral disorders figured prominently in the early stages of the disease, and the neurologic syndrome progressed more slowly than in individuals with other forms of CJD. Multiple lines of evidence indicated that the new variant of CJD (vCJD) was linked to exposure to bovine spongiform encephalopathy, either through consumption of contaminated foods or via transfusion of blood from patients in the asymptomatic/preclinical stage of vCJD. Recognition of this link led to public health measures that have been effective in limiting the spread of vCJD; the incidence in the United Kingdom peaked in 2000 and dropped precipitously since that time.

MORPHOLOGY

The progression of the dementia in CJD is usually so rapid that there is little if any grossly evident brain atrophy. The pathognomonic finding is a **spongiform** transformation of the cerebral cortex and deep gray matter structures (caudate, putamen); this multifocal process results in the uneven formation of small and apparently empty microscopic vacuoles of varying sizes within the neuropil and sometimes in the perikaryon of neurons (Fig. 28.34A). In advanced cases, there is severe neuronal loss, reactive gliosis, and sometimes expansion of the vacuolated areas into cystlike spaces ("status spongiosus"). Inflammation is notably absent. Electron microscopy shows the vacuoles to be membrane-bound and located within the cytoplasm of neuronal processes. **Kuru plaques** are extracellular deposits of aggregated abnormal PrP. They are Congo red– and PAS-positive and usually occur in the cerebellum (Fig. 28.34B), but are abundant in the cerebral cortex in cases of vCJD (Fig. 28.34C). In all forms of prion disease, immunohistochemical staining demonstrates the presence of proteinase K–resistant PrP^{sc} in tissue.

Alzheimer Disease (AD)

AD is the most common cause of dementia in older adults, with an increasing incidence as a function of age. The disease usually becomes clinically apparent as insidious impairment of higher cognitive functions. As the disease progresses, deficits in memory, visuospatial orientation, judgment, personality, and language gradually emerge; over a course of 5 to 10 years, the affected individual becomes profoundly disabled, mute, and immobile. Patients rarely become symptomatic before 50 years of age; the incidence of the disease increases with age, and the prevalence roughly doubles every 5 years, starting from a level of 1% for the 60- to 64-year-old population and reaching 40% or more for the 85- to 89-year-old cohort. This progressive increase in incidence with increasing age has given rise to major medical, social, and economic concerns in countries with aging populations. About 5% to 10% of cases are familial; these have provided important insight into the pathogenesis of the more common sporadic form of the disease. Although pathologic examination of brain tissue obtained at autopsy remains necessary for definitive diagnosis of AD, the combination of clinical assessment and currrent radiologic methods allows accurate premortem diagnosis in 80% to 90% of cases.

Pathogenesis

The fundamental abnormality in AD is the accumulation of two proteins (Aβ and tau) in specific brain regions, likely as a result of excessive production and defective removal (Fig. 28.35). The two pathologic hallmarks of AD, particularly evident in the end stages of the illness, are *amyloid plaques* and *neurofibrillary tangles*. Plaques are deposits of aggregated Aβ peptides in the neuropil, while tangles are aggregates of the microtubule binding protein tau, which develop intracellularly and then persist extracellularly after neuronal death. Both plaques and tangles appear to contribute to the neural dysfunction, and the interplay between the processes that lead to the accumulation of these two types of abnormal protein aggregates is a critically important aspect of AD pathogenesis that has yet to be fully unraveled.

Several lines of evidence strongly support a model in which **Aβ generation is the critical initiating event for the development of AD**. First, there are diseases in which tau deposits appear, such as frontotemporal lobar degenerations, progressive supranuclear palsy, and corticobasal degeneration (discussed later), but Aβ deposits do not ensue, and full-blown AD does not develop. This suggests that having

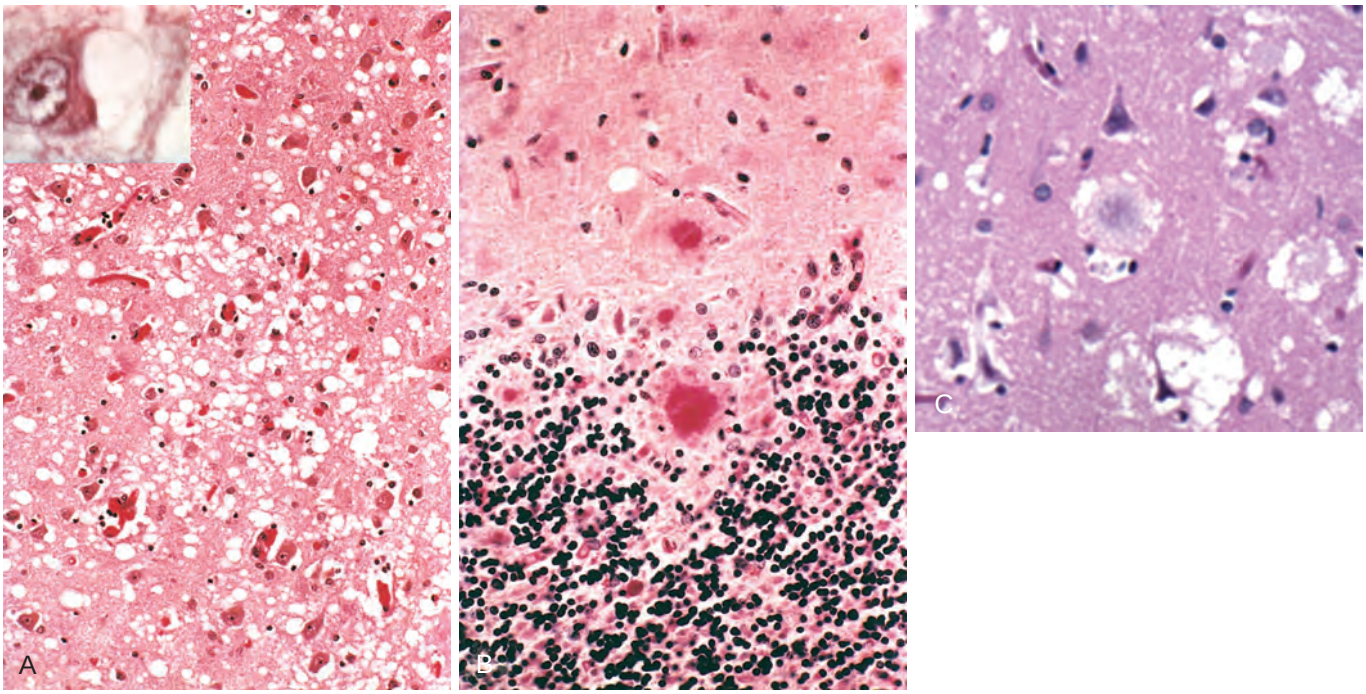


Figure 28.34 Prion disease. (A) Spongiform change in the cerebral cortex. *Inset*, High magnification of neuron with vacuoles. (B) Cerebellar cortex showing kuru plaques (periodic acid–Schiff stain) that consist of aggregated PrP^{sc}. (C) Cortical kuru plaques surrounded by spongiform change in variant Creutzfeldt-Jakob disease.

abnormal deposits of tau in the brain is not a sufficient stimulus to elicit deposition of Aβ. Additionally, multiple lines of genetic evidence (discussed later) point to the likely importance of altered Aβ metabolism. In contrast, mutations in the *MAPT* gene that encodes tau do not give rise to AD but rather cause frontotemporal lobar degeneration (discussed later).

The pathogenesis of AD thus involves Aβ, tau, and several other genetic and host factors, as follows:

- *Role of Aβ.* APP is a cell surface protein with a single transmembrane domain that may function as a receptor, possibly for prion protein (PrP^c) among other ligands. The Aβ portion of the protein extends from the extracellular region into the transmembrane domain (see Fig. 28.35). Processing of APP begins with cleavage in the extracellular domain, followed by an intramembranous cleavage. There are two potential extracellular sites of cleavage, which may be carried out by two different classes of proteases, α-secretase and β-secretase. If the first cut occurs at the α-secretase site, then Aβ is not generated (the non-amyloidogenic pathway); this cleavage event mostly occurs at the cell surface because proteases with α-secretase activity are involved in the shedding of surface proteins. Surface APP may also be endocytosed into vesicles, where it can be cleaved by β-secretase, which cuts at a slightly more N-terminal site within APP (the amyloidogenic pathway). Following cleavage of APP at either of these sites, the γ-secretase complex performs an intramembranous cleavage. Cleavage by α-secretase and β-secretase releases Aβ42, which is prone to aggregation and amyloid formation. The γ-secretase complex—containing presenilin-1 or presenilin-2, nicastrin, PEN2, and APH1—is also responsible for processing of Notch receptors and many other membrane proteins. Once generated, Aβ42 is highly prone to aggregation—first into small oligomers (which appear to be the toxic form responsible for neuronal dysfunction), and eventually into large aggregates and fibrils that can be visualized by microscopy.

  Familial forms of AD support the central role of Aβ generation as a critical step for initiation of AD pathogenesis. The gene encoding APP, on chromosome 21, lies in the Down syndrome region; AD pathology contributes to the cognitive impairment of patients with this chromosomal abnormality, with appearance of histopathologic AD findings in the second and third decades followed by neurologic decline about 20 years later. A similar gene dosage effect is produced by localized chromosome 21 duplications that span the *APP* locus in some patients with familial AD. Point mutations in *APP* are another cause of familial AD. Some mutations lie near the β-secretase and γ-secretase cleavage sites, and others sit in the Aβ sequence and increase its propensity to aggregate. The two loci identified as causes of the majority of early-onset familial AD (*PSEN1* on chromosome 14 and *PSEN2* on chromosome 1) encode presenilin-1 and presenilin-2. These mutations lead to a gain of function, such that the γ-secretase complex generates increased amounts of Aβ, particularly $Aβ_{42}$.
- *Role of tau.* Because neurofibrillary tangles contain the tau protein, there has been much interest in the role of this protein in AD. Tau is a microtubule-associated protein present in axons in association with the microtubular network. With the development of tangles in AD, it shifts to a somatic-dendritic distribution, becomes hyperphosphorylated, and loses the ability to bind to microtubules. The formation of tangles is an important component of AD, and the increased tangle burden in the brain over

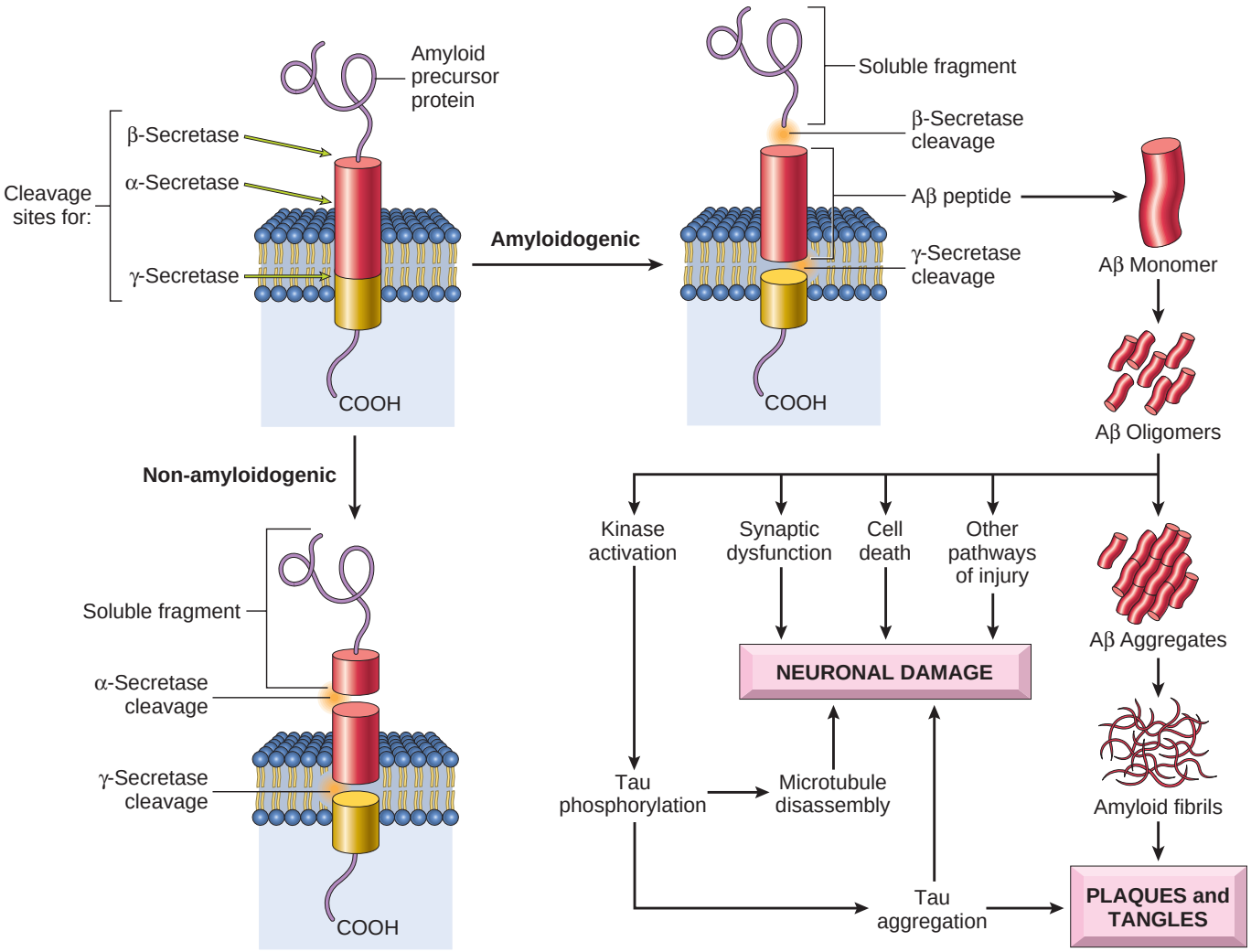


Figure 28.35 Protein aggregation in Alzheimer disease. Amyloid precursor protein cleavage by α-secretase and γ-secretase produces a harmless soluble peptide, whereas amyloid precursor protein cleavage by β-amyloid–converting enzyme and γ-secretase releases Aβ peptides, which form pathogenic aggregates and contribute to the characteristic plaques and tangles of Alzheimer disease.

the course of the illness eventually appears to become independent of the Aβ. How tangles injure neurons remains poorly understood, but two possible pathways have been suggested, which are not mutually exclusive: (1) the aggregates of tau protein elicit a stress response; and (2) the loss of tau protein destabilizes microtubules.

- *Other genetic risk factors.* The genetic locus on chromosome 19 that encodes apolipoprotein E (ApoE) has a strong influence on the risk of developing AD. Three alleles exist (ε2, ε3, and ε4) based on two amino acid polymorphisms. The dosage of the ε4 allele increases the risk of AD and lowers the age of onset of the disease, such that individuals with the ε4 allele are overrepresented in populations of patients with AD. This ApoE isoform promotes Aβ generation and deposition, but also appears to exacerbate Aβ-independent, *tau*-mediated neurodegeneration. Overall, this locus has been estimated to convey about one-fourth of the risk for development of late-onset AD. Genome-wide association studies have identified several other loci that contribute to the risk of AD; the connection between these genetic loci and the pathogenesis of AD remains to be explored.
- *Role of inflammation.* Both small aggregates and larger deposits of Aβ elicit an inflammatory response from microglia and astrocytes. This response probably assists in the clearance of the aggregated peptide, but it may also stimulate the secretion of mediators that cause damage. Additional consequences of the activation of these inflammatory cascades may include alterations in tau phosphorylation, oxidative injury to neurons, and aberrant pruning of synapses.
- *Basis for cognitive impairment.* Although there remains disagreement regarding the best correlate of dementia in individuals with AD, it is clear that the presence of a large burden of plaques and tangles is highly associated with severe cognitive dysfunction. The number of neurofibrillary tangles correlates better with the degree of dementia than does the number of neuritic plaques. Biochemical markers that have been correlated with the degree of dementia include loss of choline acetyltransferase, loss of synaptophysin immunoreactivity, and amyloid burden.
- *Biomarkers.* Among the more important recent developments in the understanding of AD is the discovery of

possible biomarkers, which draw on the understanding of the biologic processes discussed earlier. It is now possible to demonstrate Aβ deposition in the brain through imaging methods that rely on 18F-labeled amyloid-binding compounds; this approach can identify asymptomatic patients who are in the early stages of AD. Additional evidence of neuronal degeneration associated with AD-related pathologic processes includes the presence of increased phosphorylated tau and reduced Aβ in the CSF. Together, these biomarkers have allowed for the identification of preclinical stages of AD, well in advance of the development of dementia or other signs and symptoms; this in turn has enabled the focus of pharmacologic trials to shift toward individuals in the earliest stages of the illness, in whom it is hoped interventions will slow or prevent disease progression and limit disability.

MORPHOLOGY

Grossly, the brain shows variable **cortical atrophy** marked by gyral narrowing and sulcal widening that is most pronounced in the frontal, temporal, and parietal lobes (Fig. 28.36). With significant atrophy, there is compensatory ventricular enlargement (hydrocephalus *ex vacuo*) secondary to reduced brain volume. Structures of the medial temporal lobe, including the hippocampus, entorhinal cortex, and amygdala, are involved early in the disease course and are usually severely atrophied in the later stages.

The major microscopic abnormalities of AD are **neuritic (senile) plaques** and **neurofibrillary tangles.** There is progressive, eventually severe, neuronal loss and reactive gliosis in the same regions that bear the burden of plaques and tangles.

Neuritic plaques are focal, spherical collections of dilated, tortuous, axonal or dendritic processes (dystrophic neurites) often around a central amyloid core, which may be surrounded by a clear halo (Fig. 28.37A). Dystrophic neurites contain tau aggregates that are biochemically similar to neurofibrillary tangles.

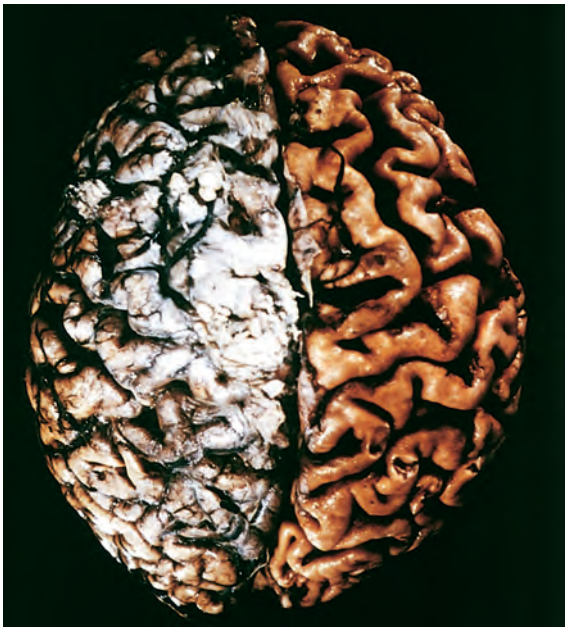


Figure 28.36 Alzheimer disease with cortical atrophy most evident on the right, where meninges have been removed. (Courtesy the late Dr. E.P. Richardson, Jr., Massachusetts General Hospital, Boston, Mass.)

Neuritic plaques range in size from 20 to 200 μm in diameter; microglial cells and reactive astrocytes are present at their periphery. Plaques are found in the hippocampus, amygdala, and neocortex, although there is usually relative sparing of primary motor and sensory cortices (this also applies to neurofibrillary tangles). The amyloid core, which can be stained by Congo red or a beta amyloid immunostain, contains several abnormal proteins. The dominant component of the amyloid plaque core is Aβ, a peptide derived by proteolytic cleavage of APP (Figs. 28.37B and 28.35); other proteins, including components of the complement cascade, proinflammatory cytokines, α_1-antichymotrypsin, and apolipoproteins, are present in lesser abundance. In some cases, there is deposition of Aβ peptides in the absence of the surrounding dystrophic neurites. These lesions, termed **diffuse plaques,** are found mainly in superficial portions of the cerebral cortex, the basal ganglia, and the cerebellar cortex. Based on studies of individuals with trisomy 21, diffuse plaques are believed to be an early stage of plaque development. Although neuritic plaques contain both $A\beta_{40}$ and $A\beta_{42}$, diffuse plaques are predominantly made up of $A\beta_{42}$.

Neurofibrillary tangles are tau-containing bundles of filaments in the cytoplasm of the neurons that displace or encircle the nucleus. In pyramidal neurons, they often have an elongated "flame" shape; in rounder cells, the basket weave of fibers around the nucleus takes on a rounded contour ("globose" tangles). Neurofibrillary tangles are visible as basophilic fibrillary structures with H&E staining (Fig. 28.37C) but are demonstrated much more clearly by silver (Bielschowsky) staining (Fig. 28.37D) and with tau immunohistochemistry (Fig. 28.37E). They are commonly found in cortical neurons, especially in the entorhinal cortex, as well as in other sites such as pyramidal cells of the hippocampus, the amygdala, the basal forebrain, and the raphe nuclei. Neurofibrillary tangles are insoluble and apparently resistant to clearance in vivo, thus remaining visible in tissue sections as "ghost" or "tombstone" tangles long after the death of the parent neuron. Ultrastructurally, neurofibrillary tangles are composed predominantly of paired helical filaments along with some straight filaments that appear to have a similar composition. Aggregated tau is also present in dystrophic neurites that form the outer portions of neuritic plaques and in axons coursing through the affected gray matter as neuropil threads. In contrast to neuritic and diffuse plaques, tangles are found in other neurodegenerative diseases and are thus not specific for AD.

In addition to the diagnostic features of plaques and tangles, several other pathologic findings are seen in the setting of AD. **Cerebral amyloid angiopathy (CAA)** (Fig. 28.18D) is an almost invariable accompaniment of AD; however, it can also be found in brains of individuals without AD. In contrast to the amyloid deposited in neuritic and diffuse plaques, which mainly consists of Aβ42, the amyloid in CAA is predominantly comprised of Aβ40.

Although abundant burdens of plaques and tangles characterize the end stage of AD, in which affected individuals are fully demented, it is clear that these histologic changes appear well in advance of clinical symptoms. To provide a correlation between neuropathologic findings and clinical symptomatology, the most recent recommendations for describing these lesions consider all deposition of Aβ in brain parenchyma to be a form of AD neuropathologic change. The scheme then generates a histopathologic score based on the distribution of Aβ deposits, plaques, and tangles, which is used to predict the likelihood of an individual being cognitively impaired that is based on population studies.

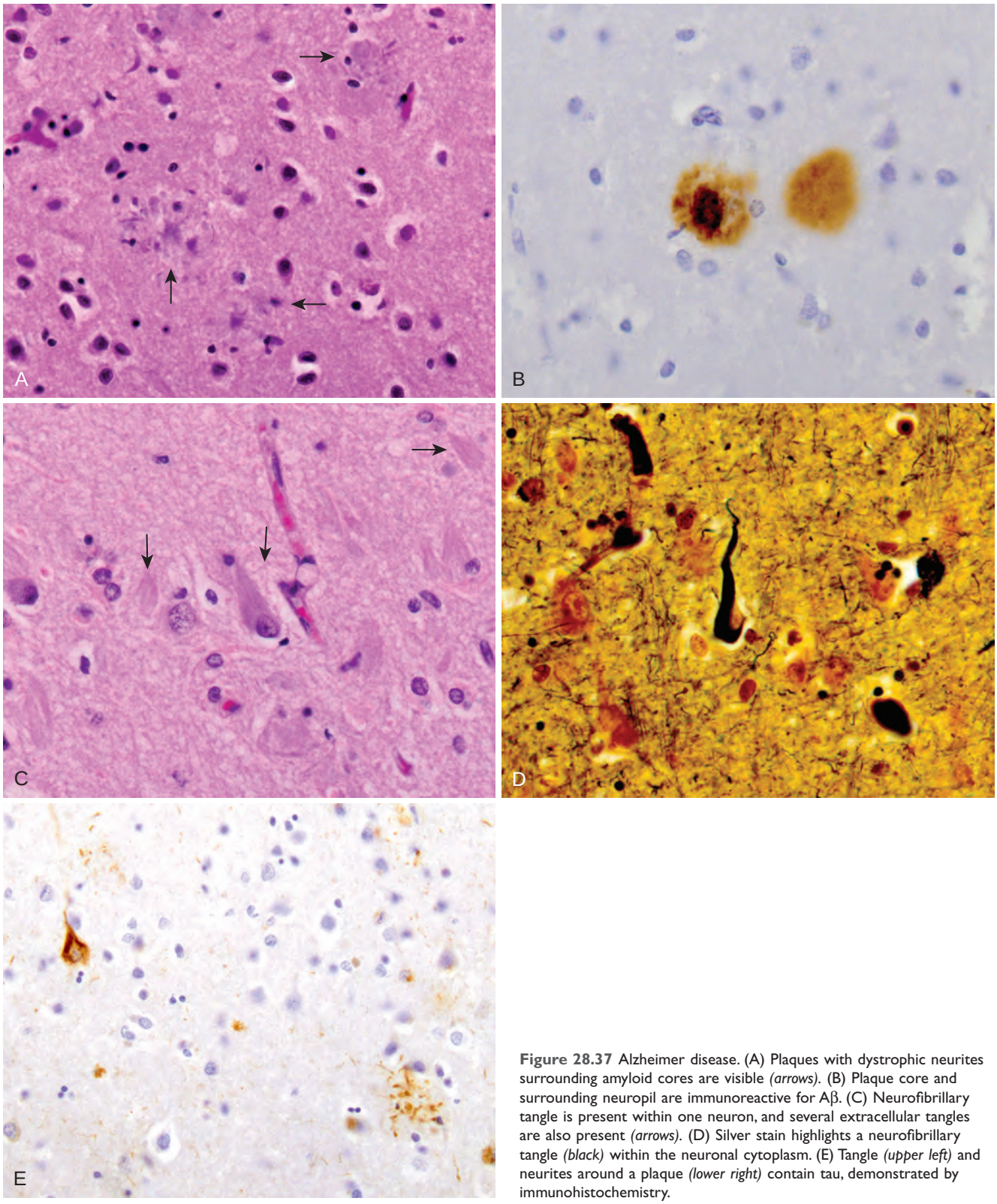


Figure 28.37 Alzheimer disease. (A) Plaques with dystrophic neurites surrounding amyloid cores are visible *(arrows)*. (B) Plaque core and surrounding neuropil are immunoreactive for Aβ. (C) Neurofibrillary tangle is present within one neuron, and several extracellular tangles are also present *(arrows)*. (D) Silver stain highlights a neurofibrillary tangle *(black)* within the neuronal cytoplasm. (E) Tangle *(upper left)* and neurites around a plaque *(lower right)* contain tau, demonstrated by immunohistochemistry.

Clinical Features

The progression of AD is slow but relentless, with a symptomatic course often running more than 10 years. Initial symptoms are forgetfulness and other memory disturbances; with progression, other symptoms emerge, including language deficits, loss of mathematical skills, and loss of learned motor skills. In the final stages, affected individuals may become incontinent, mute, and unable to walk; intercurrent disease, often pneumonia, is usually the terminal event. Current clinical trials are focused on treating subjects in early, preclinical stages of the illness, using strategies that include clearing Aβ from the brain through immunologic approaches, disrupting the generation of Aβ with pharmacologic agents that target secretases, as well as approaches aimed at preventing alterations in tau.

Frontotemporal Lobar Degenerations (FTLDs)

FTLDs are a heterogeneous group of disorders associated with focal degeneration of frontal and/or temporal lobes. They are clinically distinguished from AD by the pattern of involvement, with alterations in personality, behavior, and language (aphasias) preceding memory loss; however, as with other neurodegenerative diseases, postmortem pathologic evaluation is required for definitive diagnosis. Global dementia occurs with progressive disease, and a subset of patients develops extrapyramidal motor loss. Several clinical variants have been described based on whether the behavioral change or aphasias dominate, but they have overlapping features. FTLDs are one of the more common causes of early-onset dementia and occur at the same frequency as AD in those younger than 65 years of age. Commonly referred to in the clinical setting as *frontotemporal dementia* (FTD), the preferred pathologic terminology highlights the anatomic pattern of involvement (lobar degeneration) rather than the clinical symptom (dementia).

Like most neurodegenerative diseases, FTLD is associated with cellular inclusions comprised of specific proteins; the two most common patterns are those with tau inclusions (FTLD-tau) and those with TDP-43 inclusions (FTLD-TDP). Within each of these groups, there are heritable and sporadic forms. There is no fixed relationship between the clinical subtypes of FTLD and the type of neuronal inclusions.

FTLD-tau

These are forms of FTLD in which the affected cortical regions demonstrate progressive neuronal loss and reactive gliosis, along with the presence of cytoplasmic tau inclusions in neurons. Unlike Alzheimer disease, which is characterized by the combination of Aβ and tau deposition, FTLD-tau shows only tau aggregation and accumulation. In some cases FTLD-tau inclusions resemble the tangles seen in AD, and in other forms of the disease there are smooth contoured inclusions (Pick bodies).

Pathogenesis

FTLD-tau may be associated with mutations in *MAPT,* the gene that encodes tau, or it may arise sporadically. As mentioned earlier, tau interacts with microtubules; its binding to microtubules may be regulated by phosphorylation. There is an inverse relationship between the degree of tau phosphorylation and its ability to bind microtubules. Tau, particularly when phosphorylated, has a propensity to aggregate. Interestingly, tau exists as a complex series of isoforms that are encoded by different mRNA splice variants; the balance between these isoforms appears to be critical for normal tau function in neurons, and disturbances in isoform ratio may also provoke tau aggregation.

Two different types of tau mutations have been described. Some missense point mutations effect tau phosphorylation, tipping the balance from microtubule binding to aggregation. Other mutations include point mutations that affect splicing; many of these are intronic and alter the loop-stem structures recognized by the spliceosome. The resulting change in isoform ratio is thought to lead to neuronal dysfunction and, as discussed earlier, may also enhance tau aggregation.

It remains unclear how abnormal tau injures neurons, although there appears to be both a loss-of-function component, as aggregation depletes neurons of tau, and a toxic gain-of-function component due to presence of aberrantly hyperphosphorylated aggregated protein.

MORPHOLOGY

There is atrophy of frontal and temporal lobes to variable extent and severity. The pattern of atrophy can often be predicted from the clinical symptoms. The atrophic regions of cortex are marked by neuronal loss, gliosis, and the presence of tau-containing neurofibrillary tangles (Fig. 28.38A). These tangles may contain a variety of tau isoforms. Nigral degeneration may occur. Inclusions can also be found in glial cells in some forms of the disease.

In **Pick disease** (a subtype of FTD-tau), the brain invariably shows a pronounced and frequently asymmetric atrophy of the frontal and temporal lobes, with conspicuous sparing of the posterior two-thirds of the superior temporal gyrus and only rare involvement of the parietal and occipital lobes. The atrophy can be severe, reducing gyri to a wafer-thin ("knife-edge") appearance. The neuronal loss is most severe in the outer three layers of the cortex. Some of the surviving neurons show a characteristic swelling (Pick cells), and others contain **Pick bodies,** cytoplasmic, round to oval, filamentous inclusions that stain strongly with silver methods (Fig. 28.38B).

FTLD-TDP

Some individuals with clinically diagnosed FTD and macroscopic changes of relatively localized cortical atrophy (the "lobar degeneration" of the term) have inclusions that do not contain tau and instead contain TDP-43, an RNA-binding protein. Individuals with FTLD-TDP typically present with either behavioral problems or language deficits, features similar to FTLD-tau.

Pathogenesis

Mutations in three different genes have been found in the inherited forms of FTLD-TDP.

- The most common familial form of FTLD-TDP is the result of an expansion of a hexanucleotide repeat in the 5′ region of *C9orf72* (a gene encoding a protein of unknown function); the spectrum of disease associated with *C9orf72* expansion also includes amyotrophic lateral sclerosis

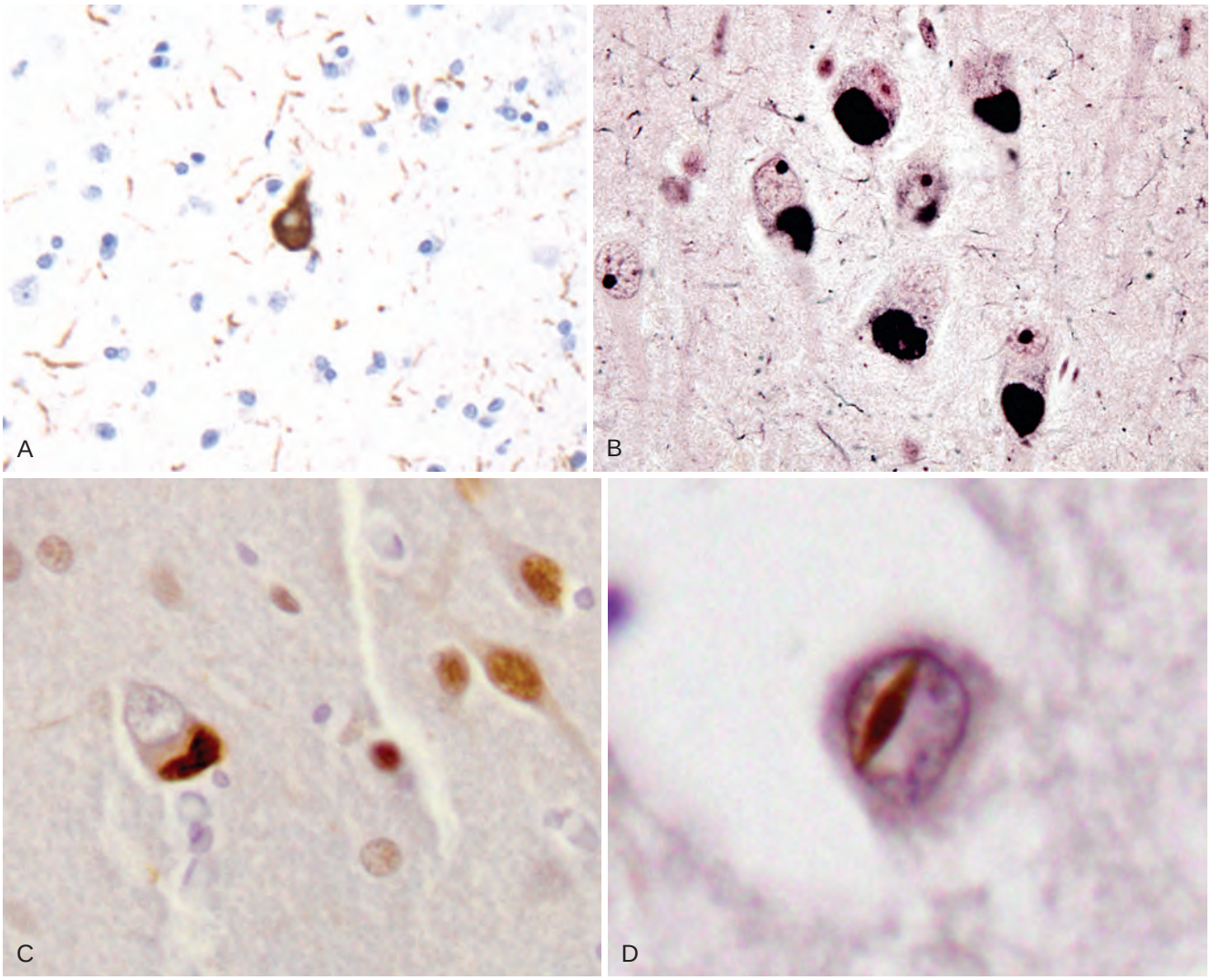


Figure 28.38 Frontotemporal lobar degenerations (FTLDs). (A) FTLD-tau. A tangle is present along with numerous tau-containing neurites. (B) Pick disease. Pick bodies are round, homogeneous neuronal cytoplasmic inclusions that stain intensely with silver stains. (C) FTLD-TDP. Cytoplasmic inclusions containing TDP-43 are seen in association with loss of normal nuclear immunoreactivity. (D) FLTD-TDP. With progranulin mutations, the TDP-43–containing inclusions are commonly intranuclear.

(discussed later). How the repeat expansion results in formation of aggregates of TDP-43 is a mystery at present. Interestingly, the brains from *C9orf72*-associated FTLD and amyotrophic lateral sclerosis cases also contain ubiquitin-positive inclusions that consist of dipeptide repeats generated by aberrant translation of the expanded repeats (in individuals without repeat expansion, this region of the gene contains regulatory sequences and is not translated).

- Mutations in the *TARDBP* gene (encoding TDP-43) are less common and also occur in some familial cases of amyotrophic lateral sclerosis. TDP-43 is an RNA-binding protein with roles in RNA processing and the formation of stress granules, RNA-protein aggregates of uncertain function. Both loss of function and toxic gain of function may contribute to this form of FTD.
- A third form of FTLD-TDP is the result of mutations in the *GRN* gene, which encodes a protein called progranulin; in contrast to *TARDBP* and *C9orf72,* mutations in *GRN* have not been linked to amyotrphic lateral sclerosis. Progranulin mutations cause loss of function, and the disease is believed to stem from deficient progranulin activity. Progranulin is a secreted protein expressed in glia and neurons that is cleaved into multiple small peptides; these peptides have been implicated in regulation of inflammation in the brain, but the link between this activity and the accumulation of TDP-43–containing inclusions in FTLD is currently obscure.

There are forms of FTLD in which there are neither tau- nor TDP-containing inclusions. Although these are infrequent, the underlying genetic causes show overlap with the pathways involved in other forms of FTLD-TPD. Mutations in the *FUS* (fused in sarcoma) gene may cause either FTLD or ALS, and FUS is another RNA-binding protein that may be involved in formation of stress granules.

MORPHOLOGY

The appearance is similar to the other forms of FLTD, with variable atrophy of frontal and temporal lobes accompanied by neuronal loss and gliosis. Normally, TDP-43 is found diffusely in the nucleus but in FTLD-TDP, TDP-43 is found in inclusions (Fig. 28.38C) that may be in the cell body (neuronal cytoplasmic inclusions), in the nucleus (neuronal intranuclear inclusions), or in neurites (TDP-43 threads). In the inclusions, TDP-43 is phosphorylated and ubiquitinated. Inclusions are most abundant in the frontal and temporal cortex, the striatum, and the dentate gyrus of the hippocampus. There is a strong correlation between the presence of needle-like intranuclear inclusions and progranulin mutations (Fig. 28.38D).

There are also forms of FTLD in which there are neither tau- nor TDP-containing inclusions. Although these are infrequent, the underlying genetic causes show connections with the pathways involved in other forms of FTLD-TDP. Mutations in the *FUS* (fused in sarcoma) gene may cause either FTLD or amyotrophic lateral sclerosis, and FUS is another RNA-binding protein that may be involved in formation of stress granules.

Parkinson Disease (PD)

PD is a neurodegenerative disease marked by a hypokinetic movement disorder that is caused by loss of dopaminergic neurons from the substantia nigra. The clinical syndrome of *parkinsonism* combines diminished facial expression (often termed masked facies), stooped posture, slowing of voluntary movement, festinating gait (progressively shortened, accelerated steps), rigidity, and a "pill-rolling" tremor. This type of motor disturbance is seen in a number of conditions that are associated with damage to the nigrostriatal dopaminergic system. Although PD is the most common cause of parkinsonism, similar symptoms can be induced by dopaminergic antagonists or by toxins that selectively damage the dopaminergic system. Also discussed later are other rare diseases that have parkinsonism as part of the clinical presentation.

The clinical diagnosis of PD is based on the presence of the triad of parkinsonism—tremor, rigidity, and bradykinesia—in the absence of a toxic or other known underlying etiology. This clinical impression is confirmed by symptomatic response to L-DOPA replacement therapy. Although the diagnosis of PD is based in large part on the presence of the motor symptoms, which reflect the decreased dopaminergic innervation of the striatum, the disease is not restricted to dopaminergic neurons or to the basal ganglia; in fact, degeneration of the substantia nigra (which results in the motor symptoms) represents a mid-stage in a progressive disease that begins lower in the brainstem and eventually progresses to involve the cerebral cortex, leading to cognitive impairment (see "Dementia with Lewy Bodies," later in this chapter).

The dopaminergic neurons of the substantia nigra project to the striatum, and their degeneration in PD is associated with a reduction in striatal dopamine content. The severity of the motor syndrome is proportional to the dopamine deficiency, which can, at least in part, be corrected by replacement therapy with L-DOPA (the immediate precursor of dopamine). However, treatment does not reverse the structural changes or arrest the progress of the disease; with progression, drug therapy becomes less effective, and symptoms more difficult to manage. Deep brain stimulation has emerged over the past decade as a therapy for the motor symptoms of PD. In addition, the well-characterized neural and biochemical deficits in PD have also provided a rationale for therapeutic trials of neural transplantation and gene therapy.

An acute parkinsonian syndrome and destruction of neurons in the substantia nigra follows exposure to MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), discovered as a contaminant in illicitly synthesized batches of the opioid meperidine. This toxin has been used to generate animal models of PD, which are used to test new therapies. Epidemiologic evidence has also suggested that pesticide exposure is a risk factor for PD, while caffeine and nicotine appear to be protective.

Pathogenesis

PD is associated with protein accumulation and aggregation, mitochondrial abnormalities, and neuronal loss in the substantia nigra and elsewhere in the brain. Although most PD is sporadic, identification of genetic causes helped shed light on its pathogenesis.

- The first mutations identified as a cause of autosomal dominant PD involved *SNCA,* a gene that encodes α-synuclein, an abundant lipid-binding protein normally localized to synapses. This protein is a major component of the *Lewy body,* which is the diagnostic hallmark of PD. Mutations in *SCNA* are rare; they include point mutations and amplifications of the region of chromosome 4q21 that contains the gene. The occurrence of disease caused by increases in gene copy number implies a gene dosage effect and suggests that polymorphisms in the α-synuclein promoter influence the risk of PD. Like Aβ in AD, α-synuclein forms aggregates; of these, small oligomers appear to be the most toxic to neurons. There is also evidence that aggregates can be released from one neuron and taken up by another, suggesting a prion-like pattern of spread within the brain. Consistent with this idea, α-synuclein–containing aggregates (Lewy bodies and Lewy neurites) are first seen in the medulla and then in contiguous areas of the brain, ascending through the brainstem and extending into limbic structures and finally the neocortex. Remarkably, emerging evidence suggests that α-synuclein aggregates are initially formed in the neurons of the enteric nervous system and spread from the gut to the medulla via the vagus nerve; this "gut-brain" hypothesis of PD pathogenesis provides a possible explanation for association of PD with exposure to pesticides and other environmental toxins.
- *Mitochondrial dysfunction* has been implicated as a contributing factor based on autosomal recessive forms of PD that are caused by mutations in genes that encode the proteins DJ-1, PINK1, and parkin. DJ-1 has multiple cellular roles, including acting as a transcriptional regulator, but in settings of oxidative stress, it can relocate to the mitochondria and have cytoprotective effects. PINK1 is a kinase that localizes to the outer membrane of dysfunctional mitochondria, where it recruits and phosphorylates

parkin. This modification activates parkin, an E3 ubiquitin ligase that targets a number of substrates for proteasomal destruction. Under normal circumstances, the combination of PINK1 and parkin results in clearance of dysfunctional mitochondria through mitophagy; this process is impaired by defects in PINK1 or parkin. Intriguingly, levels of mitochondrial complex I, a component of the oxidative phosphorylation cascade, are reduced in the brains of patients with sporadic PD.

- Heterozygous mutations in the lysosomal enzyme glucorebrosidase are the most important risk factor for development of PD, accounting for ~5% of PD cases (homozygous mutations in this enzyme cause Gaucher disease, a lysosomal storage disease described in more detail in Chapter 5); this genetic link suggests that dysfunction of the lysosome-autophagy pathway contributes to PD pathogenesis, most likely through abnormal metabolism of α-synuclein, activation of the unfolded protein response, and increased inflammation.
- Mutations in the gene encoding *LRRK2* (leucine-rich repeat kinase 2) are the most common cause of autosomal dominant PD and are also found in some late-onset, apparently sporadic cases of the disease. LRRK2 is a cytoplasmic kinase; several of the pathogenic mutations increase its kinase activity, suggesting that gains in LRRK2 function—either hyperphosphorylation of normal targets or emergence of novel targets—might contribute to the development of PD. However, the molecular mechanism linking LRRK2 dysfunction and α-synuclein accumulation has not yet been elucidated.

MORPHOLOGY

A characteristic gross finding in PD is **pallor of the substantia nigra** (compare Fig. 28.39A and B) and locus ceruleus, which is due to loss of the pigmented, catecholaminergic neurons in these nuclei. **Lewy bodies** (Fig. 28.39C) are usually found in some of the remaining neurons; these are single or multiple, cytoplasmic, eosinophilic, round to elongated inclusions that often have a dense core surrounded by a pale halo. Ultrastructurally, Lewy bodies are composed of fine filaments, densely packed in the core but loose at the rim; these filaments are composed of α-synuclein. Lewy bodies can also be found in the cholinergic neurons of the basal nucleus of Meynert, as well as in other brainstem nuclei including the locus ceruleus and the dorsal motor nucleus of the vagus. Areas of neuronal loss also typically show gliosis. Lewy neurites are dystrophic processes that contain aggregated α-synuclein.

Dementia With Lewy Bodies

About 10% to 15% of individuals with PD develop dementia, particularly with advancing age. Characteristic features of this disorder include a fluctuating course, visual hallucinations, and prominent frontal signs. Although some affected individuals have pathologic evidence of Alzheimer disease (or, less frequently, other degenerative diseases associated with cognitive changes) in combination with PD, in others the most prominent histologic correlate is the presence of widespread Lewy bodies in neurons of the cortex and brainstem. In some cases, dementia with Lewy bodies represents an advanced stage of PD in which protein aggregates appear to have "spread," possibly through propagation of misfolded proteins, to neurons in the cerebral cortex. In other individuals, however, the dementia is the first presenting symptom, and the motor symptoms appear later, suggesting a descending rather than ascending pattern of disease spread.

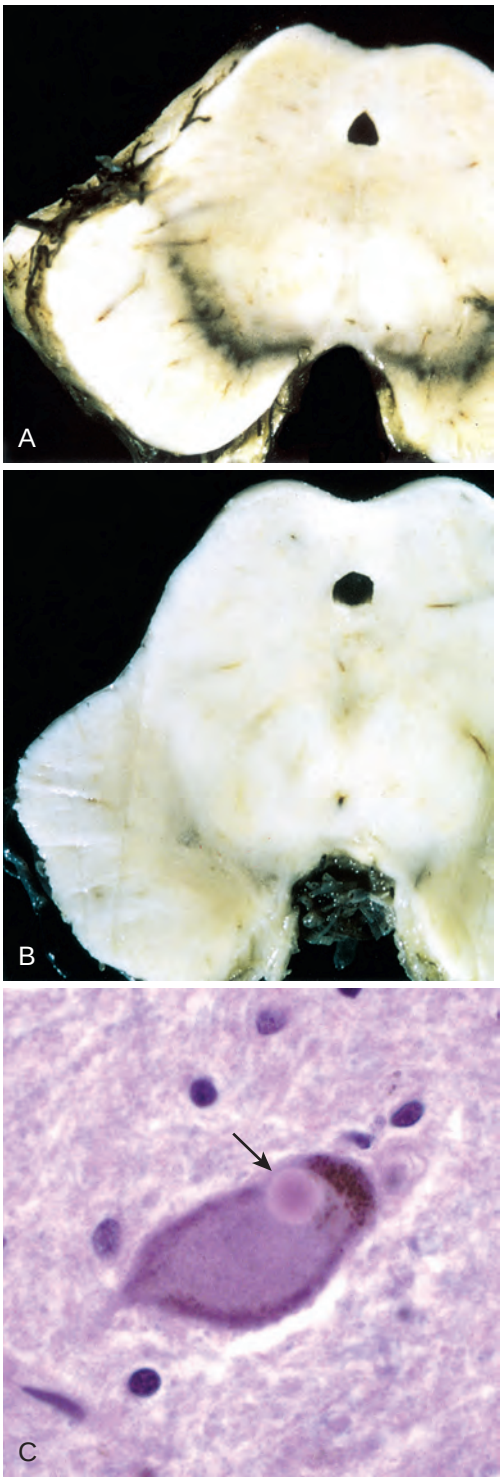


Figure 28.39 Parkinson disease. (A) Normal substantia nigra. (B) Depigmented substantia nigra in idiopathic Parkinson disease. (C) Lewy body in a substantia nigra neuron, staining bright pink *(arrow)*.

Cortical Lewy bodies are less distinctive than those in the brainstem, but are also composed predominantly of α-synuclein. Immunohistochemical staining for α-synuclein also reveals the presence of abnormal neurites that contain aggregated protein—called *Lewy neurites* even though Lewy never saw them! In this setting, the gross pathologic findings typically include depigmentation of the substantia nigra and locus ceruleus, paired with relative preservation of the cortex, hippocampus, and amygdala. The burden of cortical Lewy bodies is usually extremely low, and the mechanism by which this disease wreaks havoc on cognitive functioning is not clear. There is evidence that the burden of oligomeric α-synuclein in cortical neurons is more important in disease causation than the Lewy bodies.

Atypical Parkinsonian Syndromes

As discussed earlier, the clinical syndrome of parkinsonism, with bradykinesia and rigidity, reflects dysfunction of the dopaminergic nigrostriatal pathway. In addition to the forms of PD already discussed, there are a variety of disorders that include parkinsonism as a component of the symptoms. These diseases, in general, are minimally responsive to treatment with L-DOPA; they are also distinguished from PD through the presence of additional signs and symptoms. For these reasons, they are considered to be *atypical parkinsonian syndromes* or *Parkinson-plus syndromes*. These disorders include progressive supranuclear palsy and corticobasal degeneration, both of which are tauopathies, and multisystem atrophy, a synucleinopathy.

Progressive Supranuclear Palsy (PSP)

PSP is a tauopathy in which affected individuals develop progressive truncal rigidity, disequilibrium with frequent falls, and difficulty with voluntary eye movements. Other common symptoms include nuchal dystonia, pseudobulbar palsy, and a mild progressive dementia. The onset is usually between the fifth and seventh decades, with males affected approximately twice as frequently as females. The disease is often fatal within 5 to 7 years of onset.

Although the pathologic hallmark of PSP is the presence of tau-containing inclusions in neurons and glia, causative mutations in the tau gene have been identified in only a few cases. However, risk for sporadic PSP is linked to single nucleotide polymorphisms that map near the tau gene locus. Other risk alleles have been identified through genome-wide association studies, but how they influence development of PSP is unclear.

Corticobasal Degeneration (CBD)

CBD is a progressive tauopathy that is most often characterized by extrapyramidal rigidity, asymmetric motor disturbances (jerking movements of limbs), and impaired higher cortical function (typically in the form of apraxia). As with PSP, cognitive decline may occur, typically later in the illness. The same tau variant linked to PSP is also highly associated with CBD. Overall, CBD and PSP share many clinical and pathologic features; in general, with PSP there is a greater burden of tau-containing lesions in brainstem and deep gray matter, while in CBD the balance is shifted toward involvement of the cerebral cortex.

MORPHOLOGY

In PSP, there is widespread neuronal loss in the globus pallidus, subthalamic nucleus, substantia nigra, colliculi, periaqueductal gray matter, and dentate nucleus of the cerebellum; globose fibrillary tangles are found in these affected regions, in neurons as well as in glia. In CBD, the brain shows cortical atrophy, mainly of the motor, premotor, and anterior parietal lobes. In affected regions of cortex, there is severe loss of neurons, gliosis, and **"ballooned" neurons.** Tau immunoreactivity has been found in astrocytes ("tufted astrocytes"), oligodendrocytes ("coiled bodies"), basal ganglionic neurons, and, variably, cortical neurons. Clusters of tau-positive processes around astrocytes ("astrocytic plaques") and the presence of tau-positive threads in gray and white matter may be the most specific pathologic findings of CBD. The substantia nigra and locus ceruleus show neuronal loss, ballooned neurons, and tangles.

Multiple System Atrophy (MSA)

MSA is a sporadic disorder that affects several functional systems in the brain and is characterized by α-synuclein inclusions in the cytoplasm of oligodendrocytes. In contrast to other degenerative diseases, which primarily affect neurons, the histopathologic changes in MSA are primarily observed in glial cells and are associated with degeneration of white matter tracts. In addition, there is accompanying neuronal degeneration, but typically without the presence of inclusions. The "multiple" in the term *multiple system atrophy* refers to three distinct neuroanatomic circuits that are commonly involved: the striatonigral circuit (leading to parkinsonism), the olivopontocerebellar circuit (leading to ataxia), and the autonomic nervous system including the central elements (leading to autonomic dysfunction, with orthostatic hypotension as a prominent component). In a given individual, one of these components may predominate at the onset of the illness, but typically the other systems become affected as MSA progresses.

Pathogenesis

As in PD, α-synuclein is the major component of the inclusions. MSA is a sporadic disease, and no causative mutations in the gene encoding α-synuclein have been identified; nonetheless, polymorphisms near this gene appear to confer increased risk. The relationship between glial cytoplasmic inclusions and disease is supported by the observation that the burden of inclusions increases as the disease progresses, although inclusions eventually disappear as cells die in the final stages. It appears that glial cytoplasmic inclusions can occur in the absence of neuronal loss, suggesting that they are the primary pathologic event; for example, glial cytoplasmic inclusions are consistently observed in the white matter projecting to and from the motor cortex. The origin of the α-synuclein in oligodendrocytes remains perplexing because this is a neuronal protein associated with synaptic vesicles. Several studies have shown that there is no upregulation of α-synuclein expression in white matter or in oligodendrocytes in MSA. It has been suggested that oligodendrocytes may acquire α-synuclein aggregates secondarily from injured or dying neurons. When α-synuclein is present in oligodendrocytes, they become more sensitive

to oxidative stress and show impaired interaction with the extracellular matrix.

MORPHOLOGY

The pathologic findings in MSA match the clinical presentation in any particular case. In cerebellar forms, there is atrophy of the cerebellum, including the cerebellar peduncles, pons (especially the base, Fig. 28.40A), and medulla (especially the inferior olive); in parkinsonian forms, the atrophy involves the substantia nigra and striatum (especially the putamen). Autonomic symptoms are related to cell loss from the catecholaminergic nuclei of the medulla and the intermediolateral cell column of the spinal cord. Atrophic brain regions show evidence of neuronal loss as well as variable numbers of neuronal cytoplasmic and nuclear inclusions.

The diagnostic glial cytoplasmic inclusions were originally demonstrated in oligodendrocytes with silver impregnation methods and contain α-synuclein as well as ubiquitin (see Fig. 28.40B). Similar inclusions may also be found in the cytoplasm of neurons, in neuronal and glial nuclei, and in axons.

Huntington Disease (HD)

HD is an autosomal dominant disease caused by degeneration of striatal neurons and characterized by a progressive movement disorder and dementia. Jerky, hyperkinetic, sometimes dystonic movements involving all parts of the body (chorea) are characteristic; affected individuals may later develop bradykinesia and rigidity. The disease is relentlessly progressive and uniformly fatal, with an average course of about 15 years.

Pathogenesis

HD is a prototypic polyglutamine trinucleotide repeat expansion disease (Chapter 5). The gene for HD, *HTT,* located on chromosome 4p16.3, encodes a 348-kD protein known as *huntingtin*. In the first exon of the gene, there is a stretch of CAG repeats that encodes a polyglutamine region near the N terminus of the protein. Normal *HTT* genes contain 6 to 35 copies of the repeat; when the number of repeats is increased beyond this level, it is associated with disease. There is an inverse relationship between repeat number and age of onset, such that longer repeats tend to be associated with earlier onset. Repeat expansions occur during spermatogenesis, so that paternal transmission is associated with earlier onset in the next generation, a phenomenon termed *anticipation*. In contrast to many other degenerative diseases, there is no sporadic form of HD. New mutations are uncommon; most apparently sporadic cases are explained by nonpaternity, the death of a parent before the disease develops, or an unaffected father with a mild repeat expansion that enlarges to a pathogenic size during spermatogenesis.

The biologic function of normal huntingtin remains unknown, but it appears that the expansion of the polyglutamine region bestows a toxic gain-of-function on huntingtin. For this reason, various approaches to silencing expression of the mutant allele are being investigated as potential therapies. It is interesting to note that although huntingtin is expressed in all tissues of the body, the deleterious effects of mutant huntingtin occur only in selected parts of the central nervous system.

Although development of intranuclear inclusions containing huntingtin is a pathologic hallmark of HD, this process is not directly involved in cellular injury and, based on in vitro evidence, actually appears to be neuroprotective (among the

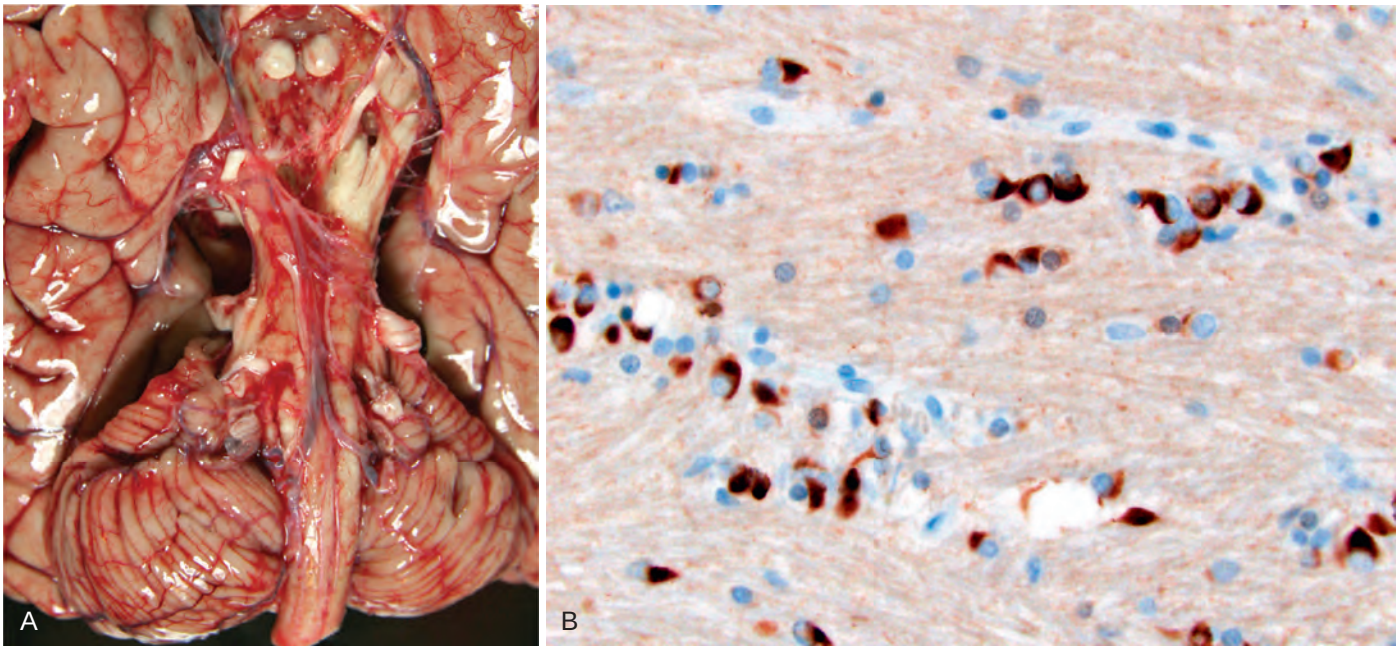


Figure 28.40 Multiple system atrophy (MSA). (A) Severe atrophy of the basis pontis in a case of MSA-C. (B) Inclusions in oligodendrocytes contain α-synuclein.

striatal neurons overexpressing mutant huntingtin, those that form inclusions survive the longest). Although the mechanism for this neuroprotective effect is not clear, it is generally thought that it stems from the sequestration of neurotoxic forms of mutated huntingtin. Transcriptional dysregulation has been implicated in HD, based on the observation that mutant forms of huntingtin bind various transcriptional regulators. Some of the transcription factors that interact with mutant huntingtin are involved in mitochondrial biogenesis and protection against oxidative injury, and their reduced activity may result in increased susceptibility to oxidative stress. Other alterations that may contribute to the pathogenesis of HD include reduced expression of the growth factor brain-derived neurotrophic factor (BDNF), and disruption of proteasomal and autophagic degradation pathways. Interestingly, emerging evidence suggests that aggregated huntingtin can be transferred between cells, another example of possible prion-like spread of a pathogenic protein.

MORPHOLOGY

The brain is small and shows striking **atrophy of the caudate nucleus and the putamen,** components of the dorsal striatum; the globus pallidus may atrophy secondarily, and the lateral and third ventricles are dilated (Fig. 28.41A). Atrophy is frequently also seen in the frontal lobe, less often in the parietal lobe, and occasionally throughout the entire cortex. On microscopic examination, there is profound loss of striatal neurons; the most marked changes are found in the caudate nucleus, especially in the tail and in portions nearer the ventricle. Pathologic changes develop in a medial-to-lateral direction in the caudate and from dorsal to ventral in the putamen. Medium-sized, spiny neurons that use γ-aminobutyric acid as their neurotransmitter, along with enkephalin, dynorphin, and substance P, are especially affected; fibrillary gliosis is more extensive than in the usual reaction to neuronal loss. There is a direct relationship between the degree of degeneration in the striatum and the severity of clinical symptoms. Protein aggregates containing huntingtin can be found in neurons in the striatum and cerebral cortex (see Fig. 28.41, *inset*).

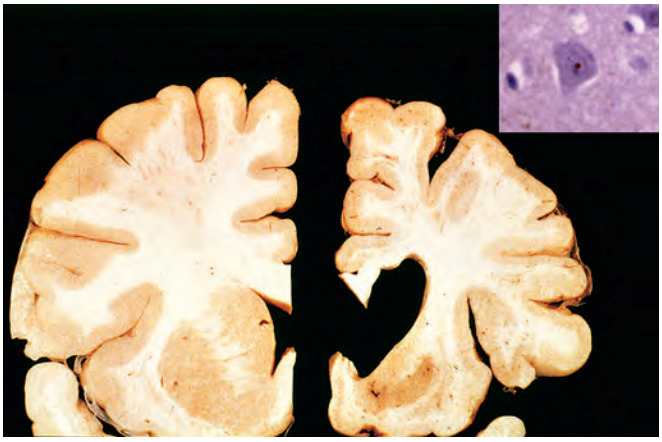


Figure 28.41 Huntington disease. Normal hemisphere on the left compared with the hemisphere with Huntington disease on the right showing atrophy of the striatum and ventricular dilation. *Inset,* Intranuclear inclusions in neurons are highlighted by an immunohistochemical stain for ubiquitin. (Courtesy Dr. J-P Vonsattel, Columbia University, NY.)

Clinical Features

The loss of striatal neurons, which function to dampen motor activity, results in increased motor output, often manifested as choreoathetosis. The cognitive changes associated with the disease are probably related to the neuronal loss in cerebral cortex. The age at onset is most commonly in the fourth and fifth decades and is related to the length of the CAG repeat in the *HTT* gene. Motor symptoms often precede the cognitive impairment. The movement disorder of HD is choreiform, with increased and involuntary jerky movements of all parts of the body; writhing movements of the extremities are typical. Early symptoms of higher cortical dysfunction include forgetfulness as well as cognitive and affective dysfunction, with progression to a severe dementia.

The most common immediate cause of death in HD patients is pneumonia, typically occurring at an advanced stage of the disease. Tragically, HD patients resort to suicide at a rate that is approximately twice that of the general population. Given the ability to screen for disease-causing mutations, one might assume that genetic screening of individuals at risk would be routine; however, this is a situation in which the ability to detect the likelihood of disease has surpassed any possible treatment. In the absence of effective therapy and given the devastating nature of the disease, the ethics of screening is debatable.

Spinocerebellar Degenerations

The shared feature of this group of degenerative diseases is that they cause neuronal loss and dysfunction in the cerebellum, the spinal cord, and peripheral nerves. In other regards, they are quite heterogeneous, with differences in genetics, age at onset, and signs and symptoms. Genetic analysis has impacted the classification of these diseases, but has yet to lead to clearer insight into their pathogenesis or to effective treatments.

The term *spinocerebellar ataxia* (SCA) is applied to a group of autosomal dominant inherited spinocerebellar degenerations. We will also briefly discuss Friedreich ataxia and ataxia-telangiectasia, two of the more common autosomal recessive disorders that are characterized by spinocerebellar degeneration. Finally, there is a small set of hereditary disorders characterized by episodes of ataxia or other symptoms of isolated cerebellar dysfunction, which are mostly associated with mutations in genes for ion channel subunits.

Spinocerebellar Ataxias (SCA)

This group of spinocerebellar degenerations is remarkably heterogeneous; current classifications that rely on a combination of genetic, clinical, and pathologic findings include 35 distinct subtypes. Here we focus on the common varieties of pathogenic mutations that are found in SCA.

Pathogenesis

The genes responsible for over half of the numerous SCAs have been identified. Three distinct types of pathogenic mutations are recognized:

- *Polyglutamine diseases* linked to expansion of a CAG repeat, similar to HD. Seven types of SCA are caused by this type of mutation, and in each the mutated protein

accumulates in neurons as nuclear inclusions, suggesting the involvement of pathogenic mechanisms previously discussed for HD. The group of polyglutamine diseases includes SCA1, SCA2, SCA3 (also known as *Machado-Joseph disease*), SCA6, SCA7 (relatively unique in that it includes visual impairment), SCA17, and dentatorubropallidoluysian atrophy.

- *Expansion of noncoding repeats,* similar to myotonic dystrophy. This mechanism is involved in five forms of SCA (SCA8, SCA10, SCA12, SCA31, and SCA36); in most, the genetic locus underlying the disease has been defined, and each involves a different locus. The pathogenic connection between expansion of these repeats and disease is obscure.
- *Other mutations*. The remaining SCAs are associated with deletions, insertions, and/or point mutations in a variety of genes that encode proteins involved in signal transduction (e.g., receptors, ion channels, kinases).

Friedreich Ataxia

Friedreich ataxia is an autosomal recessive disease characterized by progressive ataxia, spasticity, weakness, sensory neuropathy, and cardiomyopathy. It generally begins in the first decade of life with gait ataxia, followed by hand clumsiness and dysarthria. Deep tendon reflexes are depressed or absent, but an extensor plantar reflex is typically present. Joint position and vibratory sense are impaired, and there is sometimes loss of pain, temperature sensation, and light touch; many affected individuals develop pes cavus and kyphoscoliosis. Most patients become wheelchair-bound within about 5 years of onset, and life expectancy is typically limited to 40 or 50 years of age. The accompanying cardiomyopathy is associated with a high incidence of arrhythmias and congestive heart failure, which contribute to the deaths of most affected individuals. Concomitant diabetes is found in up to 25% of patients.

Friedreich ataxia is caused by expansion of a GAA trinucleotide repeat in the first intron of a gene on chromosome 9q13 that encodes frataxin, a protein found in the mitochondrial inner membrane, where it is involved in assembly of iron-sulfur cluster enzymes of mitochondrial complexes I and II. Affected individuals have extremely low levels of this protein, and the severity of the disease course may correlate better with the loss of frataxin than with the size of the GAA-repeat expansion. With reduced frataxin, there is decreased mitochondrial oxidative phosphorylation (similar to the defect in mitochondrial encephalomyopathies) as well as increased free iron; the presence of free iron within the mitochondria may contribute to oxidative stress. Most cases of Friedreich ataxia are associated with GAA repeat expansion in both alleles, but sometimes one allele has a repeat expansion and the other harbors a point mutation. Both of these patterns of inheritance are consistent with the disease stemming from loss of frataxin function.

Ataxia-Telangiectasia

Ataxia-telangiectasia (Chapter 7) is an autosomal recessive disorder characterized by an ataxic-dyskinetic syndrome beginning in early childhood, with the subsequent development of telangiectasias in the conjunctiva and skin, along with immunodeficiency. The ataxia-telangiectasia mutated *(ATM)* gene on chromosome 11q22-q23 encodes a kinase with a critical role in orchestrating the cellular response to double-stranded DNA breaks. In addition to this critical cellular role, signals mediated through ATM also regulate apoptosis, maintenance of telomeres, mitochondrial homeostasis, response to oxidative stress, and the ubiquitin-proteosomal degradation system. It remains unclear which of these pathways contribute to the degenerative phenotype observed in the setting of loss of ATM protein in neurons. Intriguingly, there are comparable patterns of neurodegeneration associated with other diseases in which repair of single-stranded DNA breaks is disrupted.

Clinical Features

The disease is relentlessly progressive, with death early in the second decade. Common initial symptoms are recurrent sinopulmonary infections and unsteadiness in walking; later on, speech is noted to become dysarthric, and eye movement abnormalities develop. Many affected individuals develop lymphoid neoplasms, which are most often T-cell leukemias.

Amyotrophic Lateral Sclerosis (ALS)

ALS is a progressive disorder in which there is loss of upper motor neurons in the cerebral cortex and lower motor neurons in the spinal cord and brainstem. Loss of these neurons results in denervation of muscles, producing weakness that becomes profound as the disease progresses. The disease has an overall incidence of about 2 per 100,000 people, affects men slightly more frequently than women, and commonly emerges in the fifth decade or later. Sporadic ALS is more common than familial ALS, which may account for up to 20% of cases.

Pathogenesis

Both sporadic and familial ALS are associated with degeneration of upper and lower motor neurons, often in association with toxic protein accumulation. Close to two dozen genetic loci have been identified as causing familial ALS, with nearly all being autosomal dominant disorders.

One of the earliest discovered hereditary forms of ALS has mutations in the gene encoding copper-zinc superoxide dismutase *(SOD1)* on chromosome 21; this variant accounts for about 20% of familial cases. Initially, identification of *SOD1* mutations suggested that neuronal injury in ALS might reflect an impaired capacity to detoxify free radicals, but it is now believed that mutations lead to an adverse gain-of-function phenotype associated with mutant SOD1 protein. It appears that mutated SOD1 protein misfolds and forms aggregates (which can include wild-type protein) that cause cellular injury through a variety of postulated mechanisms, including dysregulation of proteasomal and autophagic catabolism, disruption of axonal transport and mitochondrial function, and sequestration of other proteins within the aggregates. Accumulation of protein aggregates can eventually trigger the unfolded protein response, with subsequent initiation of apoptosis. The overall importance of protein degradation pathways is reinforced by discovery of a range of uncommon mutations in genes implicated in protein degradation that are also associated with familial ALS.

The most common mutation that gives rise to ALS and FTLD simultaneously is an expansion of a hexanucleotide

repeat in the 5′-untranslated region of a transcript of unknown function, *C9orf72*. This mutation is estimated to be the basis of up to 40% of familial ALS and a smaller fraction of what appear to be sporadic cases of ALS. Non–AUG-initiated translation (in all three reading frames) can occur from these expanded repeats, and neuronal deposits of the derived peptides have been found in the setting of the mutation; whether these novel peptide aggregates contribute to cellular injury remains obscure.

Other genetic loci that cause ALS and FTLD encode stress granule proteins with RNA-binding capacity, such as TDP-43 and FUS. The underlying link between alterations in RNA-binding proteins and manifestations of motor neuron disease remains unclear; possibly, nuclear depletion of TDP-43 results in inappropriate processing of some RNAs, while the aggregation of the protein in the cytoplasm activates the unfolded protein response common to many of the proteinopathies.

MORPHOLOGY

The anterior roots of the spinal cord are thin (Fig. 28.42A) due to loss of lower motor neuron axons, and the precentral motor gyrus in the cortex may be atrophic in especially severe cases. There is a reduction in the number of anterior horn neurons throughout the length of the spinal cord, associated with reactive gliosis. Similar findings are seen in the hypoglossal, ambiguus, and motor trigeminal cranial nerve nuclei. Remaining neurons often contain PAS-positive cytoplasmic inclusions called Bunina bodies (which appear to be remnants of autophagic vacuoles) and TDP-43-positive cytoplasmic inclusions (which are seen in sporadic cases and familial cases caused by mutations in *C9orf72* and the TDP-43 gene [*TARDBP*], but are absent in *SOD1*-mutant and *FUS*-mutant cases). Skeletal muscles innervated by the degenerated lower motor neurons show neurogenic atrophy. Loss of the upper motor neurons leads to degeneration of the corticospinal tracts, resulting in volume loss and absence of myelinated fibers, which may be particularly evident at the lower segmental levels (Fig. 28.42B).

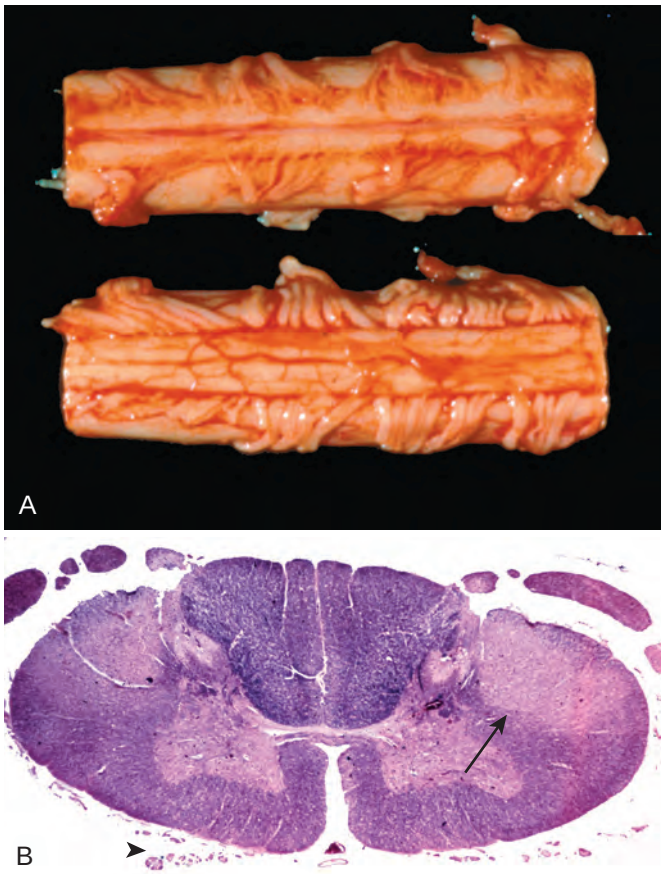


Figure 28.42 Amyotrophic lateral sclerosis. (A) Segment of spinal cord viewed from anterior *(upper)* and posterior *(lower)* surfaces showing attenuation of anterior (motor) roots compared with posterior (sensory) roots. (B) Spinal cord showing loss of myelinated fibers (lack of stain) in corticospinal tracts (best seen on the right side of this specimen; *arrow*) as well as degeneration of anterior roots *(arrowhead)*.

Clinical Features

Early symptoms of ALS include asymmetric weakness of the hands, manifested as dropping of objects and difficulty performing fine-motor tasks, and cramping and spasticity of the arms and legs. As the disease progresses, muscle strength and bulk diminish, and involuntary contractions of individual motor units, termed fasciculations, occur. The disease eventually involves the respiratory muscles, leading to recurrent bouts of pneumonia. Although most affected individuals have a combination of upper and lower motor neuron involvement, other patterns are observed. The term *progressive muscular atrophy* applies to uncommon cases in which lower motor neuron involvement predominates, while *primary lateral sclerosis* refers to cases with mostly upper motor neuron involvement. In some affected individuals, degeneration of the lower brainstem cranial motor nuclei occurs early and progresses rapidly, a pattern referred to as *progressive bulbar palsy* or *bulbar ALS*. In these individuals, abnormalities of swallowing and speech dominate, and the clinical course is inexorable during a 1- or 2-year period. When bulbar involvement is less severe, about half of affected individuals are alive 2 years after diagnosis. The motor neurons innervating extra-ocular muscles are among the last to be involved in ALS; with long survival, usually associated with ventilator support, even this form of motor output fails. Familial cases develop symptoms earlier than most sporadic cases, but the clinical course is comparable.

Although ALS is considered a disease of the motor system, a significant fraction of affected individuals have evidence of more widespread cerebral cortical disease. The clinical presentation of cerebral disease is usually a frontotemporal dementia, with pathologic findings most often matching those of FTLD associated with TDP-43 inclusions. The mechanistic link between these two processes is further strengthened by the presence of TDP-43–containing inclusions in many cases of ALS as well as overlap of genetic alterations in ALS and FTLD.

Other Motor Neuron Diseases

In addition to ALS, motor neurons are the primary target in a few other neurodegenerative diseases with distinct etiology and pathogenesis.

Spinal and Bulbar Muscular Atrophy (Kennedy Disease)

This X-linked polyglutamine repeat-expansion disease is characterized by distal limb amyotrophy and bulbar signs, such as atrophy and fasciculations of the tongue and dysphagia, that are associated with degeneration of lower motor neurons in the spinal cord and brainstem. The expanded repeat occurs in the first exon of the androgen receptor and results in androgen insensitivity, gynecomastia, testicular atrophy, and oligospermia. The basis for the selective motor neuron involvement is unclear, but as in other polyglutamine expansion diseases such as HD and some forms of spinocerebellar atrophy, there are intranuclear inclusions that contain the involved protein. Progression of the disease is slow, and most affected individuals remain ambulatory until late in the disease course. Life span is normal.

Spinal Muscular Atrophy (SMA)

SMA includes a group of genetically linked disorders of childhood characterized by marked loss of lower motor neurons that results in progressive weakness. The most severe form (SMA type I, Werdnig-Hoffmann disease) has onset during the first year of life, and is usually fatal by the age of 2 years. Other forms, with later onset, have more gradual courses; in SMA type III (Kugelberg-Welander disease), motor disability usually emerges during later childhood and adolescence.

The severity of SMA is related to loss of a protein (termed SMN) that is involved in the assembly of the spliceosome. Most SMN protein comes from mRNA transcripts derived from the *SMN1* gene on chromosome 5q. There is an adjacent *SMN2* gene that differs by a few base pairs; these changes alter mRNA splicing and increase the production of SMN protein that is unstable and rapidly degraded. *SMN2* is an example of a gene that shows copy number variation, such that normal individuals may have one or multiple copies of *SMN2* on each chromosome 5. All forms of SMA are caused by loss-of-function mutations in *SMN1* (usually deletions), with the differences in clinical phenotype being determined by the number of *SMN2* gene copies. A novel gene therapy for SMA is now available that uses an adeno-associated virus to express SMN in motor neurons, albeit at a cost of over $2 million. Genetic therapies that alter the splicing of *SMN2*-derived transcripts have also been approved for treatment of this devastating disease.

KEY CONCEPTS

NEURODEGENERATIVE DISEASES

- Neurodegenerative diseases are characterized by progressive neuronal loss involving specific neuronal circuits and brain regions. Most of these diseases are associated with accumulation of abnormal protein aggregates, typically in the form of cellular inclusions. The disease phenotype reflects the patterns of brain involvement rather than the type of inclusion.
- These diseases can be grouped by clinical presentation into dementias, hypokinetic movement disorders, hyperkinetic movement disorders, cerebellar ataxias, and motor neuron diseases.
- Prion diseases may be sporadic, familial, or transmissible. Regardless of the cause, the disease is driven by the conversion of a normal cellular protein (PrP^{c}) into an abnormal conformation (PrP^{sc}), with the acquisition of distinct characteristics including relative resistance to protease digestion, self-propagation, and the ability to spread. The most common clinical presentation is that of rapidly progressive dementia.
- Among dementias, Alzheimer disease (with plaques of Aβ and tangles of tau) is the most common; other predominantly dementing diseases include the various forms of frontotemporal lobar degeneration (usually with either TDP-43– or tau-containing inclusions) and dementia with Lewy bodies (with α-synuclein inclusions).
- Among the hypokinetic movement disorders, Parkinson disease is the most common (again with α-synuclein–containing inclusions); other diseases with parkinsonism include progressive supranuclear palsy and corticobasal degeneration (both forms of tauopathy).
- Amyotrophic lateral sclerosis is the most common form of motor neuron disease, with diverse genetic causes as well as sporadic forms.

GENETIC METABOLIC DISEASES

Genetic disorders that disrupt metabolic processes in neurons and glia typically present early in life and are often relentlessly progressive. Some of these disorders manifest in the immediate postnatal period, while others emerge later in life. Overall, metabolic diseases with earlier onset tend to have a more rapid and aggressive clinical course.

This heterogeneous group of disorders can be organized based on the organelles or the type of cell that is affected, and by the anatomic distribution of disease, as follows:

- *Neuronal storage diseases* are predominantly autosomal recessive disorders caused by the deficiency of an enzyme or trafficking protein that is required for the catabolism of sphingolipids (including the gangliosides), mucopolysaccharides, or mucolipids.
- *Leukodystrophies* are mostly autosomal recessive disorders caused by mutations in genes encoding enzymes involved in myelin synthesis or catabolism.
- *Mitochondrial encephalomyopathies* are disorders of oxidative phosphorylation, often affecting multiple tissues including skeletal muscle (Chapter 27). These disorders may be caused by mutations in the mitochondrial or the nuclear genomes.

Neuronal Storage Diseases

These disorders are characterized by the accumulation of storage material within neurons, typically followed by neuronal death. The neurologic manifestations that result from this neuronal dysfunction or death are most often seizures and generalized loss of neurologic function. Although many of these disorders are associated with deficits of specific enzymes, others appear to be caused by defects in protein or lipid trafficking within neurons. This is a large class of disorders, with genetic heterogeneity even within clinically homogeneous entities. A few diseases in this category (Tay-Sachs and Niemann-Pick diseases, mucopolysaccharidoses, and others) are described in more detail in Chapter 5.

Leukodystrophies

These disorders are caused by mutations of genes whose products are involved in the generation, turnover, or maintenance of myelin. Normal brain function depends as much on the connections between neurons as it does on the integrity of the neurons themselves; therefore, the progressive and cumulative damage to myelinated axons in leukodystrophies has devastating consequences. Several clinical features separate these dysmyelinating leukodystrophies from demyelinating diseases such as MS: the leukodystrophies typically present with an insidious and progressive loss of cerebral function, often at younger ages, and are associated with diffuse and symmetric changes on imaging studies. Although many of the leukodystrophies are caused by single enzyme defects resulting in altered metabolism of myelin-associated lipids, a variety of other genetic alterations can lead to white matter diseases. The following examples are representative of this spectrum of disorders.

- *Krabbe disease* is an autosomal recessive leukodystrophy resulting from a deficiency of galactocerebroside β-galactosidase, the enzyme required for the catabolism of galactocerebroside to ceramide and galactose. As a consequence of the impaired catabolism of galactocerebroside in the brain, an alternative catabolic pathway shunts galactocerebroside to galactosylsphingosine; elevated levels of this compound are toxic to both oligodendrocytes and astrocytes. The clinical course is rapidly progressive, with onset of symptoms (dominated by motor signs such as stiffness and weakness) between 3 and 6 months of age; survival beyond 2 years of age is uncommon. The brain shows loss of myelin and oligodendrocytes (Fig. 28.43); a similar process affects peripheral nerves. Neurons and axons are relatively spared. A unique and diagnostic feature of Krabbe disease is the aggregation of engorged macrophages *(globoid cells)* in the brain parenchyma and around blood vessels (Fig. 28.43, *inset*).
- *Metachromatic leukodystrophy* is an autosomal recessive disease that results from a deficiency of the lysosomal enzyme, arylsulfatase A. This enzyme, present in a variety of tissues, cleaves the sulfate from sulfate-containing lipids (sulfatides) as the first step in their degradation. The enzyme deficiency thus leads to an accumulation of the sulfatides, especially cerebroside sulfate, that are toxic to white matter. The most striking histologic finding is demyelination with resulting gliosis. Macrophages with vacuolated cytoplasm are scattered throughout the white matter. The membrane-bound vacuoles contain complex crystalloid structures composed of sulfatides; when bound to certain dyes such as toluidine blue, sulfatides shift the absorbance spectrum of the dye, a property called *metachromasia*. Similar metachromatic material can be detected in peripheral nerves and in urine, the latter being a sensitive method of establishing the diagnosis.
- *Adrenoleukodystrophy* is an X-linked recessive disease caused by loss-of-function mutations in *ABCD1,* which encodes an ATP-binding cassette transporter protein that is required for transport of fatty acids into peroxisomes. In the typical form, young boys present with behavioral changes and adrenal insufficiency. The loss of ABCD1 leads to an inability to catabolize very-long-chain fatty acids within peroxisomes, resulting in elevated serum levels of these lipids. The symptoms result from progressive CNS, peripheral nerve, and adrenal disease. In the white matter, myelin loss is accompanied by gliosis and extensive lymphocytic infiltration. Adrenal cortical atrophy explains the hypocortisolism.

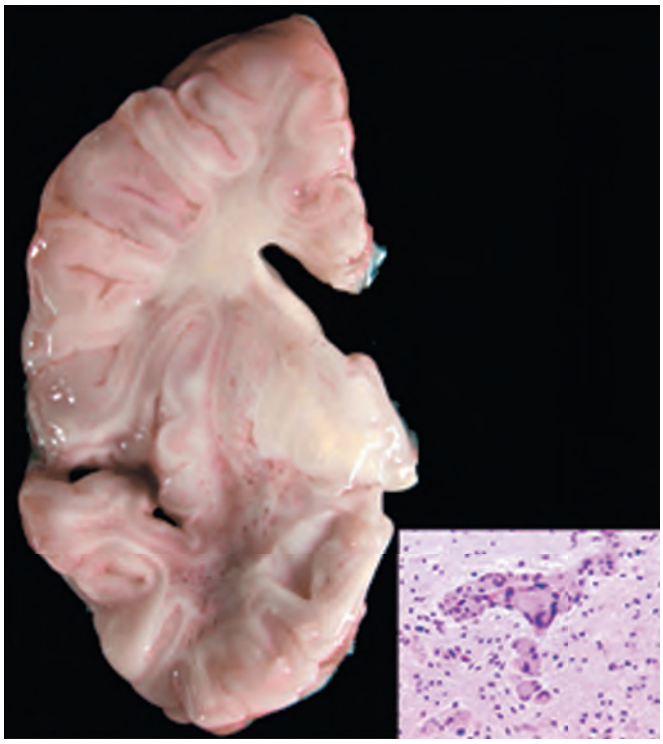


Figure 28.43 Krabbe disease. Much of the white matter is gray/yellow because of the loss of myelin. *Inset,* "Globoid" cells are the hallmark of the disease.

A wide range of other leukodystrophies are recognized, some with defects in lipid metabolism. Several other mechanisms of white matter injury have been identified, including pathogenic mutations in genes that encode proteins required for myelin formation *(Pelizaeus-Merzbacher disease)*, the intermediate filament protein GFAP *(Alexander disease)*, or the translation initiation factor eIF2B *(vanishing white matter leukoencephalopathy)*.

Mitochondrial Encephalomyopathies

Disorders of energy generation can cause a range of neurologic disease, often in association with abnormalities in other tissues. Although many of the inherited disorders of mitochondrial oxidative phosphorylation present as muscle diseases (Chapter 27), the critical dependence of neurons on oxidative phosphorylation for generation of ATP is reflected in frequent CNS involvement.

The mitochondrial genome, which is entirely inherited from the mother, encodes only 13 proteins, 22 tRNAs, and two rRNAs. The remaining proteins that are required for mitochondrial structure and function are encoded by the nuclear genome. Thus, some mitochondrial disorders show maternal transmission because the affected genes lie in the mitochondrial genome, but many others do not. There is a

complex genotype-phenotype relationship in these disorders: the same mutation may manifest as different phenotypes, and the same phenotype may be caused by different mutations.

A critical mitochondrial disease concept is *heteroplasmy*, defined as the presence of both wild type and mutated mitochondrial genomes within different populations of mitochondria in single cells. Because mitochondria are distributed in a more or less random fashion when "mother cells" divide, daughter cells may inherit different proportions of functional and dysfunctional mitochondria, leading to wide cell-to-cell variation in disease expression. In general, mitochondrial disorders in the CNS selectively target neurons; the disruption of energy generation is often reflected in elevated tissue lactate levels, which can be demonstrated by spectroscopic imaging methods. At the histologic level, there can be loss of enzymatic activity of cytochrome C oxidase, which can be assessed with special histochemical stains.

Some of the better recognized mitochondrial encephalopathies include the following entities:

- *Mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes (MELAS)* is characterized by recurrent episodes of acute neurologic dysfunction, cognitive changes, muscle weakness, and lactic acidosis. The strokelike episodes are often associated with reversible deficits that do not correspond to specific vascular territories. Pathologically, areas of infarction are observed, sometimes with vascular proliferation and focal calcification. The most common mutation observed in MELAS is in the gene encoding mitochondrial tRNA-leucine *(MTTL1)*.
- *Myoclonic epilepsy with ragged red fibers (MERRF)* is a maternally transmitted disease in which affected individuals have myoclonus (a seizure disorder) and a myopathy that is characterized by ragged red fibers on muscle biopsy (Chapter 27). Ataxia, associated with neuronal loss from the cerebellar system, is also common. Most cases of MERRF are associated with mutations in mitochondrial tRNA genes.
- *Leigh syndrome* is a disease of infancy characterized by lactic acidemia, arrest of psychomotor development, feeding problems, seizures, extraocular palsies, and weakness with hypotonia; death usually occurs within 1 to 2 years. On histologic examination, there are multifocal regions of destruction of brain tissue associated with a spongiform appearance and vascular proliferation. Brainstem nuclei, the thalamus, and the hypothalamus are typically involved, usually in a symmetric manner. A wide spectrum of nuclear and mitochondrial DNA mutations has been identified.

KEY CONCEPTS

GENETIC METABOLIC DISEASES

- Mutations that disrupt metabolic or synthetic pathways can affect the nervous system. These pathways can involve general cellular processes or those that are relatively specific to the nervous system.
- Diseases with earlier onset are typically more severe in the degree of damage and pace of illness.
- Neuronal storage diseases are usually autosomal recessive disorders. The characteristic finding is accumulation of material within neurons, along with evidence of neuronal death.
- Leukodystrophies are also typically autosomal recessive disorders associated with disruption of the synthesis or turnover of myelin components.
- Mitochondrial encephalomyopathies are a pleiotropic set of disorders that involves neurons and tissues outside of the nervous system. These can be caused by mutations in the nuclear or mitochondrial genomes.

TOXIC AND ACQUIRED METABOLIC DISEASES

Toxic and acquired metabolic diseases are relatively common causes of neurologic illnesses. These diseases are discussed in Chapter 9; only aspects that are relevant to CNS pathology are presented here.

Vitamin Deficiencies

Neural tissue has a very high metabolic demand and is thus mainly affected by deficiencies of B vitamins, which play a key role in cellular metabolism; two distinct clinicopathologic syndromes are recognized.

Thiamine (Vitamin B_1) Deficiency

The classical neurologic disease resulting from thiamine deficiency is called the *Wernicke-Korsakoff syndrome* and consists of two related manifestations. *Wernicke encephalopathy* is characterized by acute psychosis and ophthalmoplegia. These symptoms are reversible when treated with thiamine. However, if unrecognized and untreated, they may be followed by a prolonged and largely irreversible condition *(Korsakoff syndrome)* that is marked by disturbances of short-term memory and confabulation. The syndrome is particularly common in the setting of chronic alcoholism, but it may also be encountered in individuals with thiamine deficiency resulting from gastric disorders, including carcinoma, chronic gastritis, or persistent vomiting.

MORPHOLOGY

In Wernicke encephalopathy, there are foci of hemorrhage and necrosis in the mamillary bodies and the walls of the third and fourth ventricles. Early lesions show dilated capillaries with prominent endothelial cells; subsequently, the capillaries become leaky, producing hemorrhagic areas. With time, there is infiltration of macrophages and development of a cystic space containing hemosiderin-laden macrophages. These chronic lesions predominate in individuals with Korsakoff syndrome. Lesions in the dorsomedial nucleus of the thalamus seem to correlate best with memory disturbance and confabulation.

Vitamin B_{12} Deficiency

Subacute combined degeneration of the spinal cord is caused by deficiency of vitamin B_{12} and is marked by degeneration of the ascending and descending spinal tracts. The lesions result from a defect in myelin formation; the mechanism of this defect is not known. Symptoms may present over a few weeks, initially with bilaterally symmetrical numbness,

tingling, and slight ataxia in the lower extremities, which may progress to include spastic weakness of the lower extremities. Complete paraplegia may occur, usually only later in the course. With prompt vitamin replacement therapy, clinical improvement occurs; however, once complete paraplegia has developed, recovery is poor. On microscopic examination, there is swelling of myelin layers, producing vacuoles in the affected tracts; with time, axons degenerate as well. In the early stages of the disease, the mid-thoracic level of the spinal cord is preferentially affected, from which the process may extend proximally and distally.

Neurologic Sequelae of Metabolic Disturbances

Hypoglycemia

Because the brain requires glucose and oxygen for its energy production, the cellular effects of diminished glucose resemble those of oxygen deprivation, described earlier. Some regions of the brain are more sensitive to hypoglycemia than others. Glucose deprivation initially leads to selective injury to large cortical pyramidal neurons, which, if severe, may result in pseudolaminar necrosis of the cortex, predominantly involving deep layers. The Sommer sector (area CA1) of the hippocampus is also vulnerable, as are cerebellar Purkinje cells, albeit to a lesser extent than with hypoxia. If the level and duration of hypoglycemia are of sufficient severity, there may be widespread injury to many areas of the brain.

Hyperglycemia

Hyperglycemia, most commonly associated with inadequately controlled diabetes mellitus and associated with either ketoacidosis or hyperosmolar coma, does not elicit significant morphologic changes in the brain. The affected individual becomes dehydrated and develops confusion, stupor, and eventually coma. The fluid depletion must be corrected gradually; otherwise, severe cerebral edema may follow.

Hepatic Encephalopathy

The encephalopathy found in the setting of impaired liver function (Chapter 18) is accompanied by a glial response within the CNS; critical mediators appear to include elevated ammonia levels as well as proinflammatory cytokines. Astrocytes with enlarged nuclei and minimal cytoplasm, known as Alzheimer type II cells, appear in the cerebral cortex, basal ganglia, and other subcortical gray matter regions.

Toxic Disorders

Cellular and tissue injury from toxic agents is discussed in Chapter 9. Aspects of several important toxic disorders that are of unique neurologic importance are discussed here.

Carbon Monoxide

Many of the pathologic findings that follow acute carbon monoxide exposure are the result of impaired oxygen-carrying capacity of hemoglobin; in addition, CO can interact with the heme of cytochrome C oxidase, inhibiting tissue utilization of oxygen by blocking electron transport in the mitochondria. Selective injury of the neurons in layers III and V of the cerebral cortex, Sommer sector of the hippocampus, and Purkinje cells is characteristic. A unique pattern of bilateral globus pallidus necrosis may also occur. Demyelination of white matter tracts may be a later event.

Ethanol

The effects of acute ethanol intoxication are reversible, but chronic alcohol abuse is associated with a variety of neurologic sequelae, including Wernicke-Korsakoff syndrome from thiamine deficiency (see earlier). The toxic effects of chronic alcohol intake may be either direct or secondary to nutritional deficits. Cerebellar dysfunction occurs in about 1% of chronic alcoholics, and is associated with a clinical syndrome of truncal ataxia, unsteady gait, and nystagmus. The histologic changes are atrophy and loss of granule cells, predominantly in the superior anterior vermis (Fig. 28.44). In advanced cases, there is loss of Purkinje cells and proliferation of the adjacent astrocytes *(Bergmann gliosis)* between the depleted granular cell layer and the molecular layer of the cerebellum. The fetal alcohol syndrome is discussed in Chapter 10.

Radiation

Exposure of the brain to radiation can occur accidentally or as part of therapeutic regimens for tumors. As discussed in Chapter 9, exposure to very high doses of radiation (>10 Gy) can cause intractable nausea, confusion, convulsions, and rapid onset of coma, followed by death. Delayed effects of radiation can also present with rapidly evolving symptoms, including headaches, nausea, vomiting, and papilledema, which may appear months to years after irradiation.

The pathologic findings consist of large areas of coagulative necrosis, primarily in white matter, with all tissue elements within the area affected. This is accompanied by marked edema and gliosis in the surrounding tissue, along with vascular fibrinoid necrosis and sclerosis. The combination of radiation and methotrexate, administered either concurrently or sequentially, can act synergistically to cause tissue injury; in this setting, the pattern of injury is similar in appearance to that caused by radiation alone. Axons and cell bodies in the vicinity of the radiation-induced lesions

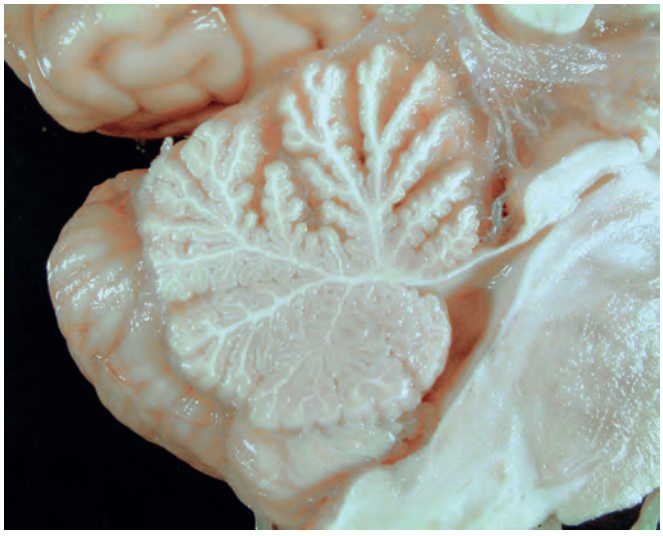


Figure 28.44 Alcoholic cerebellar degeneration. The anterior and superior portion of the vermis *(upper portion of figure)* is atrophic, with widened spaces between the folia.

undergo dystrophic mineralization. Radiation can also induce tumors, which usually develop years after radiation therapy and include sarcomas, gliomas, and meningiomas.

KEY CONCEPTS

TOXIC AND ACQUIRED METABOLIC DISEASES

- Certain vitamin deficiencies result in neurologic disease as well as systemic disorders, with thiamine and vitamin B_{12} deficiencies being the most common.
- The metabolic demand of the CNS makes it highly susceptible to injury from hypoglycemia and carbon monoxide poisoning.
- Chronic alcohol exposure may result in several forms of brain injury; atrophy of the superior anterior cerebellar vermis is particularly characteristic.
- Metabolic disturbances may disrupt brain function, often without detectable morphologic changes.

TUMORS

The annual incidence of tumors of the CNS ranges from 10 to 17 per 100,000 persons for intracranial tumors and 1 to 2 per 100,000 persons for intraspinal tumors; the majority of these are primary tumors. Tumors of the CNS account for nearly 20% of all cancers of childhood. Seventy percent of childhood CNS tumors arise in the posterior fossa; a comparable number of tumors in adults arise within the cerebral hemispheres above the tentorium.

Although pathologists have developed classification schemes that distinguish between benign and malignant lesions on histologic grounds, the clinical course of a patient with a brain tumor is strongly influenced by patterns of growth and location. Thus, even benign tumors and low-grade malignanices can lead to serious clinical deficits and may prove fatal. Because of the capacity of diffuse gliomas to widely infiltrate the brain parenchyma, such tumors are not amenable to complete surgical resection without compromising neurologic function. Also, any CNS neoplasm, regardless of histologic grade or classification, may be lethal if situated in a critical brain region; for example, a benign posterior fossa meningioma may cause cardiorespiratory arrest by compressing vital centers in the medulla. Even the most highly malignant gliomas rarely metastasize outside the CNS. Some malignant pediatric tumors spread through the CSF when they encroach on the subarachnoid space, resulting in cerebrospinal dissemination far from the original tumor site.

Classification of tumors is one of the arts of pathology, drawing on long-recognized histologic and biologic features and newer molecular analyses. Treatment protocols and clinical trials are currently based on the 2016 World Health Organization (WHO) classification, which segregates tumors into one of four grades according to their biologic behavior, ranging from grade I to grade IV. Like most cancers, malignant brain tumors tend to progress and become more aggressive with time due to clonal evolution, a change in behavior that is often reflected in a change to a higher tumor grade (which, in this case, also means a new tumor name).

The major classes of primary brain tumors to be considered here include gliomas, neuronal/glioneuronal tumors, embryonal tumors, and a few less common categories. In addition, meningeal tumors and familial tumor syndromes will be covered.

Gliomas

Gliomas, the most common group of primary brain tumors, include *astrocytomas, oligodendrogliomas*, and *ependymomas*. These tumor types have characteristic histologic features that form the basis for the classification. However, as is true of many hematologic malignancies (Chapter 13), it appears that these tumors are derived from a multipotent progenitor cell that preferentially differentiates down a particular cellular lineage. Many CNS tumors have a predilection for specific anatomic regions within the brain and tend to occur within particular age groups.

Astrocytoma

The two major categories of astrocytic tumors are the diffusely infiltrating astrocytomas (WHO grade II to IV) and the more localized astrocytomas, the most common example of which is the pilocytic astrocytoma, WHO grade I. Astrocytomas may occur from the first decade of life onward and may be found anywhere along the neuroaxis from the cerebral hemispheres to the spinal cord.

Infiltrating Astrocytomas (WHO Grades II to IV)

Infiltrating astrocytoma and glioblastoma (the synonym for "grade IV astrocytoma") account for about 80% of primary brain tumors in adults. Usually found in the cerebral hemispheres, they may also occur in the cerebellum, brainstem, or spinal cord, most often in the fourth through sixth decades. The most common presenting signs and symptoms are seizures, headaches, and focal neurologic deficits related to the anatomic site of involvement. The degree of histologic differentiation of infiltrating astrocytomas correlates well with clinical outcome; tumors range from *diffuse astrocytoma* (grade II) to *anaplastic astrocytoma* (grade III) and *glioblastoma* (grade IV), and are further stratified based on mutations of the isocitrate dehydrogenase genes (*IDH1* or *IDH2*) into *IDH-mutant* and *IDH–wild-type* forms, the former associated with considerably better prognosis than the latter (Table 28.5). There are no WHO grade I infiltrating astrocytomas because they are considered malignant by definition.

Pathogenesis

Glioblastoma tends to occur in two different clinical settings—most commonly as a new onset disease, typically in older individuals (primary glioblastoma), and less frequently in younger patients due to progression of a lower-grade astrocytoma (secondary glioblastoma). Secondary glioblastomas and their lower-grade precursors are associated with driver mutations of *IDH1,* or (less often) its homologue *IDH2* (see Chapter 7 for a discussion of these oncometabolites and their role in tumorigenesis). In contrast, most primary glioblastomas are IDH–wild type, WHO grade IV; they harbor other common genetic alterations, most commonly gains of chromosome 7, loss of chromosome 10, *TERT*-promoter mutations, and *EGFR* gene amplification. These and other less common genetic and genomic alterations contribute to the acquisition of cancer hallmarks (Chapter 7). For example, most gliomas have genetic aberrations that

Table 28.5 Features of Diffuse Gliomas

	Low-Grade or Anaplastic Astrocytoma (WHO Grade II or III)	Low-Grade or Anaplastic Oligodendroglioma (WHO Grade II or III)	Glioblastoma, *IDH*-mut (WHO Grade IV)	Glioblastoma, *IDH*-wt (WHO Grade IV)
IDH status	Mutant	Mutant	Mutant	Wild-type
Other genetics	*TP53*-mut, *ATRX*-mut	1p/19q-codeleted	*TP53*-mut, *ATRX*-mut	+7/–10, *pTERT*-mut, *EGFR*-amp
Primary or secondary	Usually primary	Usually primary	Usually secondary	Usually primary
Typical morphology	Nuclear atypia; mitoses in grade III	Round nuclei, clear haloes; mitoses, MVP, and/or necrosis in grade III	Nuclear atypia, mitoses, MVP, necrosis	Nuclear atypia, mitoses, MVP, necrosis
Prognosis	OS: 5 to 15 years	OS: 10 to 20 years	OS: 2 to 4 years	OS: 6 months to 2 years

amp, Amplification; *mut*, mutant; *MVP*, microvascular proliferation; *OS*, overall survival time after onset of disease (average); *p*, promoter *wt*, wild-type; +, whole-chromosome gain; –, whole-chromosome loss.

lead to evasion of senescence (either telomerase mutations or mutations that lead to alternative lengthening of telomeres); escape from normal growth controls (biallelic *CDKN2A* deletion); activation of growth factor signaling pathways (*EGFR* or *PDGFR* gene amplification); and resistance to apoptosis (*TP53* mutation).

MORPHOLOGY

Diffuse astrocytomas are poorly defined, gray, infiltrative tumors that expand and distort involved brain (Fig. 28.45A). These tumors range in size from a few centimeters to lesions that replace nearly the entire brain. The cut surface of the tumor may be either firm or soft and gelatinous; cystic degeneration may be seen. The tumor may appear well demarcated, but is more commonly ill-defined due to infiltration beyond the perceived margins.

Microscopically, these tumors are hypercellular compared to normal white matter and feature enlarged, elongated or irregular, hyperchromatic nuclei embedded within a fibrillar background (Fig. 28.45B) that is often GFAP-immunoreactive. An immunostain for the IDH1 R132H mutant protein highlights the tumor cells in up to 90% of cases (Fig. 28.45B, *inset*); *IDH1* and *IDH2* sequencing is required to identify less common pathogenic IDH mutations, and is done when the immunostain is negative. Individual tumor cells infiltrate brain tissue some distance away from the main lesion.

Anaplastic astrocytomas have a similar appearance, but are more densely cellular and have readily detectable mitotic activity.

In **glioblastomas**, variation in the appearance of the tumor from region to region is characteristic; some areas are firm and gray-white, while others are soft and yellow due to necrosis or red due to hemorrhage and hypervascularity (Fig. 28.46). The histologic appearance is similar to anaplastic astrocytoma with the additional features of necrosis and/or microvascular proliferation. Necrosis in glioblastoma often occurs in a serpentine pattern with tumor hypercellularity along the edges of the necrotic regions, a histologic pattern referred to as palisading (Fig. 28.47). The microvascular cell proliferation produces tufts of cells that pile

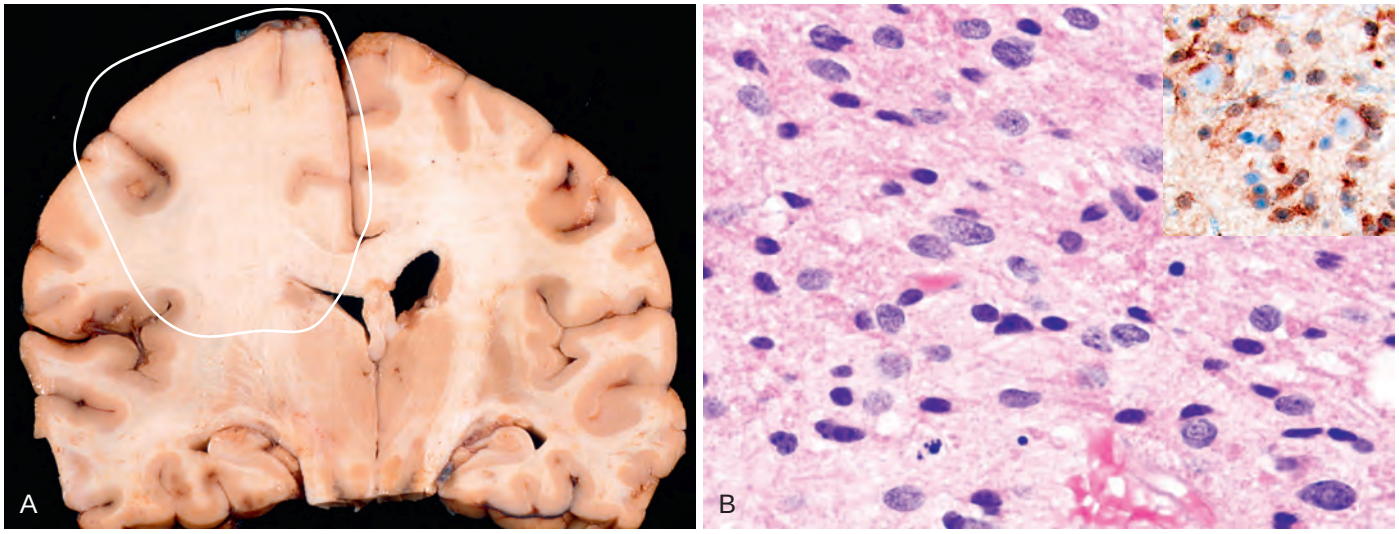


Figure 28.45 Diffuse astrocytoma. (A) On coronal section at autopsy, the left frontal white matter is expanded, and there is blurring of the corticomedullary junction due to infiltrative tumor *(circled region)*. (B) This histologic section from the white matter shows enlarged, irregular, hyperchromatic nuclei that appear embedded within the native fibrillary matrix of the brain; the smaller round and oval nuclei are native oligodendrocytes and reactive astrocytes, respectively. *Inset*, An immunostain for IDH1 R132H is positive for mutant protein in tumor cells, some of which surround the larger immunonegative cortical neurons ("perineuronal satellitosis").

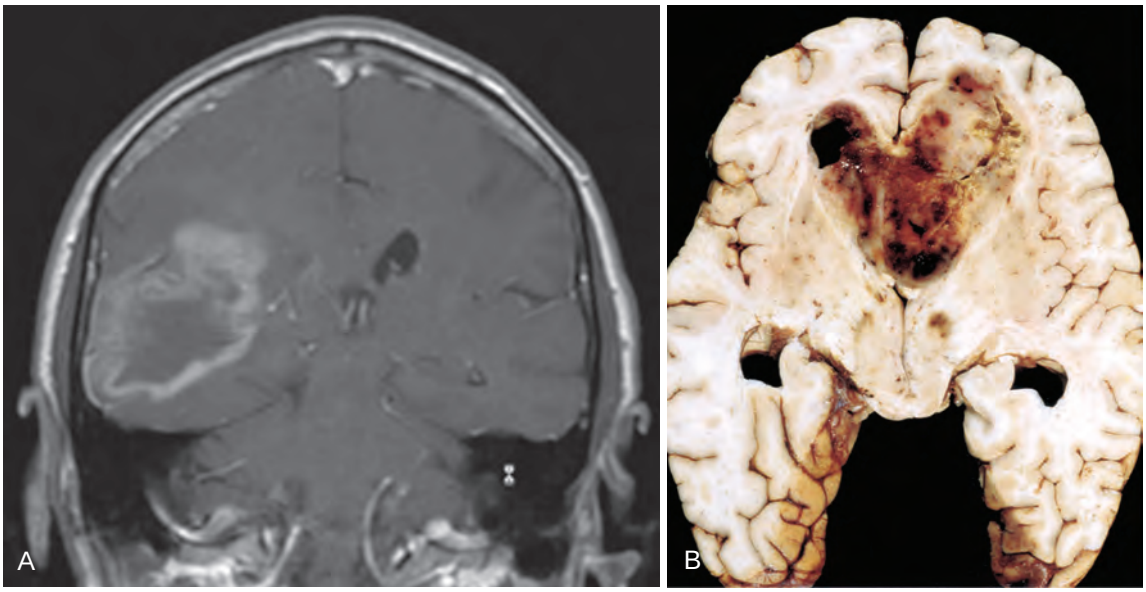


Figure 28.46 (A) Contrast T1-weighted coronal magnetic resonance image shows a large mass in the right temporal lobe with "ring" enhancement. (B) Glioblastoma appearing as a necrotic, hemorrhagic, infiltrating mass.

up and bulge into the lumen of small blood vessels (Fig. 28.47 *inset*); VEGF, produced by malignant astrocytes in response to hypoxia, contributes to this distinctive vascular change. Because histologic features can be extremely variable from one region to another, small biopsy specimens may not be representative, and the tumor may be undergraded.

Clinical Features

The presenting symptoms of infiltrating astrocytomas depend, in part, on the location and growth rate of the tumor. WHO grade II astrocytomas may remain stable or progress slowly; mean survival exceeds 5 years. Eventually, however, clinical deterioration occurs due to the emergence of a more rapidly growing subclone of higher histologic grade. Radiologic studies show a non-enhancing mass and adjacent "edema," which often includes a tumor infiltrate. High-grade astrocytomas have abnormal vessels that are "leaky," with an abnormally permeable blood-brain barrier, and therefore demonstrate contrast enhancement, which is often ringlike in glioblastomas (Fig. 28.46A). The prognosis for individuals with IDH–wild-type glioblastoma (~90% of glioblastomas) is very poor, although the use of newer chemotherapeutic agents has provided some benefit. Epigenetic silencing of the promoter for the gene encoding the DNA repair enzyme MGMT predicts responsiveness to DNA alkylating drugs (e.g., temozolomide), as would be expected given that MGMT is critical for the repair of chemotherapy-induced DNA modifications. With current treatment, which consists of resection followed by radiation therapy and chemotherapy, the mean length of survival after diagnosis has increased to 15 months; 25% of such patients are alive after 2 years. Survival is substantially shorter in older patients and in patients with lower performance status or with large unresectable lesions. As stated earlier, the IDH-mutant glioblastomas are also aggressive but have a better prognosis overall, with an average survival of 2 to 3 years.

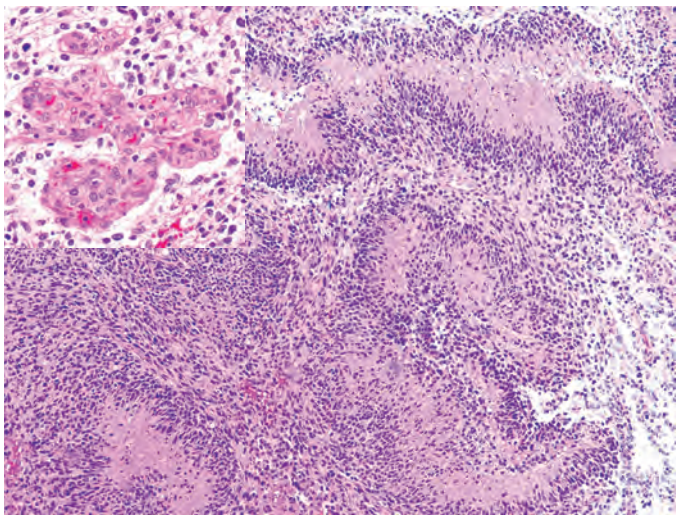


Figure 28.47 Glioblastoma. Serpiginous foci of palisading necrosis (tumor nuclei lined up around the red anucleate zones of necrosis). *Inset,* Microvascular proliferation.

Oligodendroglioma

The other major subtype of infiltrating glioma is comprised of cells that resemble oligodendrocytes. When corrected for tumor grade, the oligodendrogliomas have the best prognosis among glial tumors; as with their astrocytic counterparts, they are now defined using morphologic and genetic features. These tumors constitute 5% to 15% of gliomas and are most common in the fourth and fifth decades. Patients may have had several years of neurologic complaints, often including seizures. The lesions are found mostly in the cerebral hemispheres and have a predilection for white matter.

Pathogenesis

Oligodendroglioma is molecularly defined by an *IDH* mutation in combination with whole-arm chromosomal codeletion

of 1p and 19q. Additional genetic alterations occur with progression to anaplastic oligodendroglioma; the more common of these include losses of chromosome 9p, which result in deletion of the *CDKN2A* tumor suppressor gene. In contrast to high-grade astrocytic tumors, *EGFR* gene amplification is not seen. The vast majority of oligodendrogliomas feature the same *TERT*-promoter mutations as those encountered in glioblastoma.

MORPHOLOGY

Oligodendrogliomas are gelatinous, gray masses, often with cysts, focal hemorrhage, and calcification. The tumors are composed of sheets of regular cells with spherical nuclei containing finely granular chromatin (similar to normal oligodendrocytes) surrounded by a clear halo of vacuolated cytoplasm (Fig. 28.48). The tumor typically contains a delicate network of anastomosing capillaries, resembling "chicken wire." Calcification, present in as many as 90% of these tumors, ranges from microscopic foci to massive depositions. Tumor cells that infiltrate the cerebral cortex often collect around neurons (perineuronal satellitosis). Mitotic activity and proliferation indices are low in low-grade (WHO grade II) oligodendroglioma.

Anaplastic oligodendrogliomas (WHO grade III) are characterized by a higher cell density, nuclear anaplasia, increased mitotic activity, and occasionally necrosis. Such cases may arise de novo, but more often progress from WHO grade II oligodendrogliomas.

Clinical Features

In general, individuals with oligodendrogliomas have a better prognosis than do those with astrocytomas. Current treatment with surgery, chemotherapy, and radiation therapy has yielded average survivals of 15 to 20 and 10 to 15 years for WHO grades II and III tumors, respectively. Progression from low- to higher-grade lesions occurs, typically over a period of 5 or more years.

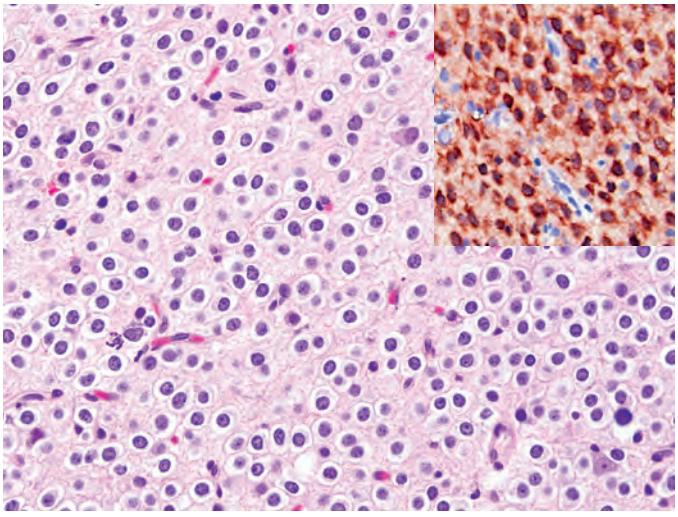


Figure 28.48 Oligodendroglioma. Tumor nuclei are round, with cleared cytoplasm forming "halos" and vasculature composed of thin-walled capillaries. Similar to diffuse astrocytoma (see Fig. 28.46B), tumor cells are usually positive for IDH1 R132H–mutant protein *(inset)*.

Pilocytic Astrocytoma

Pilocytic astrocytomas (WHO grade I) are distinguished from infiltrative astrocytomas by their more discrete borders, unique histopathologic and molecular characteristics, and relatively benign behavior. They typically occur in children and young adults and are usually located in the cerebellum, but they may also appear in the region around the third ventricle, optic nerves, spinal cord, and occasionally in the cerebral hemispheres. The histologic separation of these tumors from other astrocytomas is supported by their relative circumscription, unique histopathologic features, and distinct genetic alterations of the MAP kinase pathway, the most common being a *KIAA1549-BRAF* gene fusion/duplication. Pilocytic astrocytomas grow very slowly and can often be treated by resection alone. Symptomatic recurrence and incompletely resected lesions may be treated with repeat surgery, chemotherapy, or targeted therapies.

MORPHOLOGY

Pilocytic astrocytoma is often radiologically and grossly well demarcated and cystic (Fig. 28.49A), with a contrast-enhancing mural nodule. Histologically, it differs from infiltrative astrocytoma of any grade by showing only limited brain invasion, with an often biphasic architecture combining loose "microcystic" and compact, densely fibrillar areas (Fig. 28.49B). The tumor is composed of bipolar cells with long, thin "hairlike" processes that are GFAP-positive and form a dense fibrillary meshwork; Rosenthal fibers (eosinophilic corkscrew-shaped inclusions) and eosinophilic granular bodies (mulberry-like inclusions) are characteristic findings (Fig. 28.49C). In contrast to diffuse gliomas, the presence of microvascular proliferation or necrosis does not imply an unfavorable prognosis.

Ependymoma

Ependymomas are tumors that most often arise in proximity to the ependyma-lined ventricular system, including the oft-obliterated central canal of the spinal cord. In the first two decades of life, they typically occur near the fourth ventricle and constitute 5% to 10% of the primary brain tumors in this age group; less commonly, they are found in the cerebral hemispheres, where some lack an obvious connection to the ventricles. In adults, the spinal cord is the most common location; tumors in this site are more frequent in the setting of neurofibromatosis type 2 (NF2).

Pathogenesis

Given the association of spinal ependymomas with NF2, it is not surprising that the *NF2* gene on chromosome 22 is commonly mutated in ependymomas arising in the spinal cord but not at other sites. Ependymomas do not share the genetic alterations that are found in infiltrating gliomas. Despite their similar morphologic appearance, ependymomas occurring at different sites (supratentorial, posterior fossa, spinal) tend to be associated with distinct sets of driver mutations.

MORPHOLOGY

Grossly and on MRI, ependymomas are solid (i.e., noninfiltrative) masses. In the posterior fossa, they typically arise from the floor of

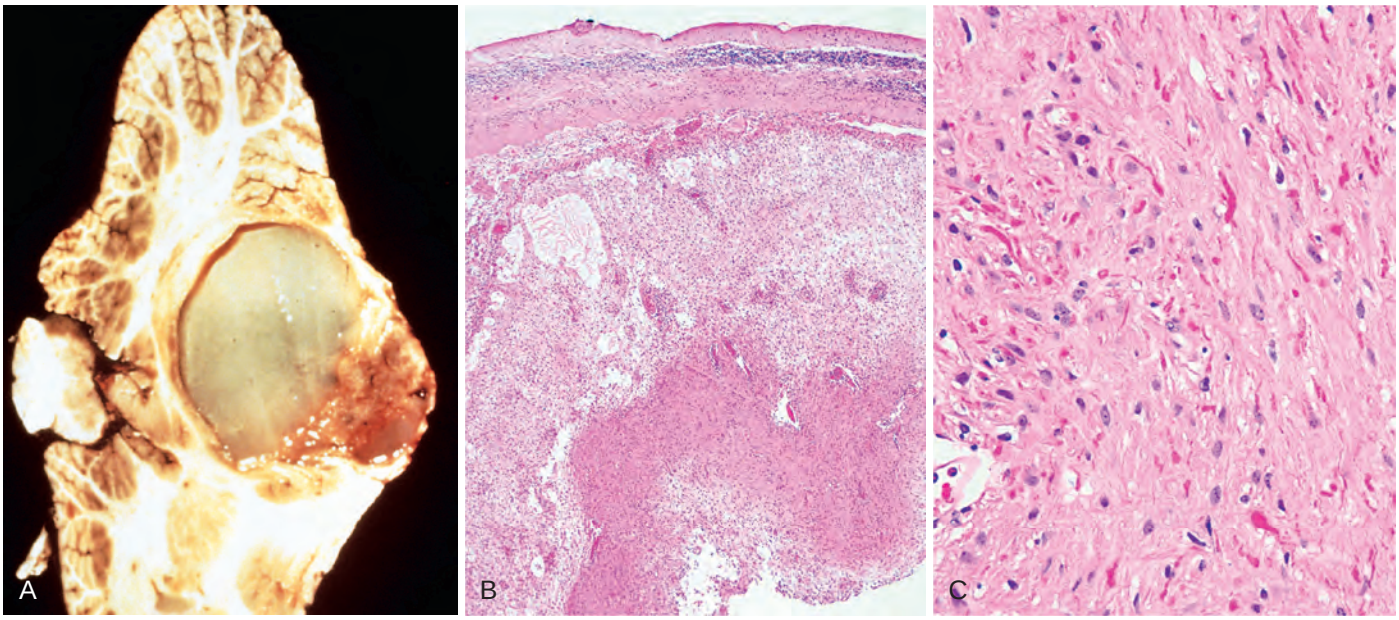


Figure 28.49 Pilocytic astrocytoma. (A) Grossly, this cerebellar tumor forms a mural nodule within a cyst. (B) At low magnification, one can appreciate atrophic cerebellar cortex *(top),* sharp circumscription, and a biphasic pattern with alternating loose *(middle)* and compact *(bottom)* regions of tumor growth. (C) Higher magnification reveals oval to irregular tumor nuclei resembling those of diffuse astrocytoma, but with numerous Rosenthal fibers (brightly eosinophilic corkscrew-shaped inclusions) in the background.

the fourth ventricle (Fig. 28.50A), occasionally growing through the foramina of Magendie or Luschka and extending below the foramen magnum into the cervical spinal canal. Although ependymomas are moderately well demarcated from adjacent brain, the proximity to vital structures often makes complete removal impossible; in the spinal cord, however, total resection is more feasible. Ependymomas are composed of cells with fibrillary processes that contain regular, round to oval nuclei with granular chromatin. Tumor cells may form glandlike round or elongated structures (rosettes, canals) that resemble the embryologic ependymal canal, with long, delicate processes extending into a lumen (Fig. 28.50B); more frequently present are **perivascular pseudorosettes** (Fig. 28.50B) in which there is a perivascular nuclear-free zone consisting of thin ependymal processes radiating toward the central blood vessel. Although most ependymomas are well differentiated and behave as WHO grade II lesions, anaplastic ependymomas (WHO grade III) feature increased cell density, high mitotic rates, areas of palisading necrosis, and/or microvascular proliferation. However, the tumor grade is less predictive of outcome than either the extent of surgical resection or the molecular subtype.

Two other variants of ependymoma merit brief mention. **Myxopapillary ependymomas** (WHO grade I) are distinct lesions that occur in the filum terminale of the spinal cord. Cuboidal tumor cells are arranged in a variably papillary architecture around mucin-rich fibrovascular cores. Prognosis depends on completeness of surgical resection.

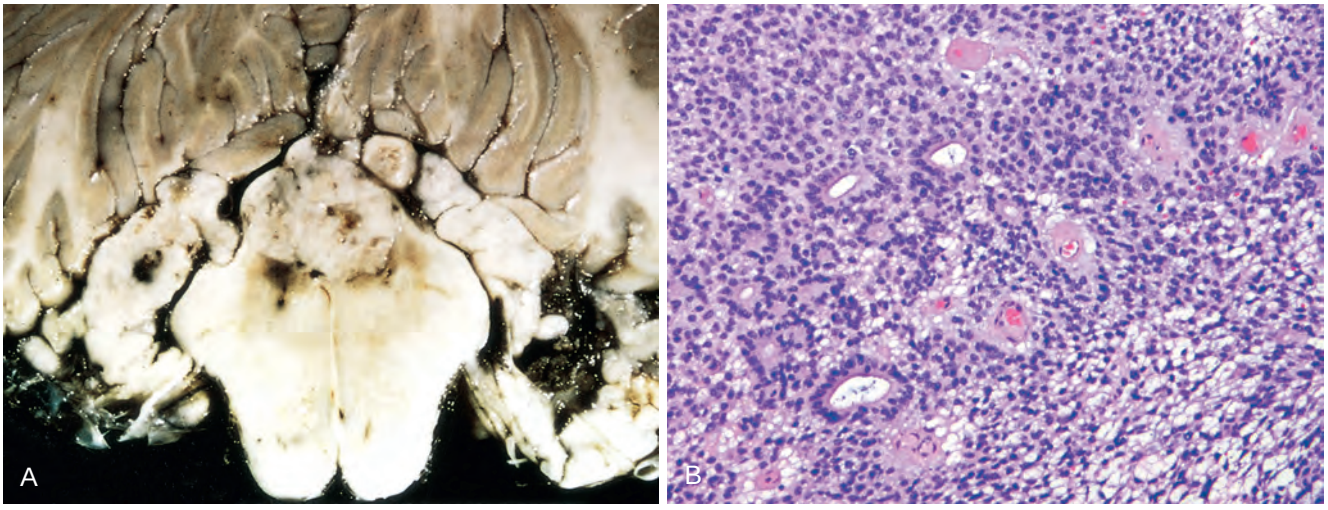


Figure 28.50 Ependymoma. (A) Tumor of the fourth ventricle, distorting, compressing, and infiltrating surrounding structures. (B) The microscopic appearance includes both true rosettes (with a glandlike central lumen) and perivascular pseudorosettes (nuclear-free zone composed of fibrillary processes radiating toward a central blood vessel).

Subependymomas (WHO grade I) are solid, sometimes calcified, slow-growing nodules that are usually found protruding into the lateral or fourth ventricle by growing under the ependyma. They are usually asymptomatic and are incidental findings at autopsy or imaging. However, if they are sufficiently large or strategically located, they can cause obstructive hydrocephalus. Subependymomas have a characteristic microscopic appearance, consisting of clusters of ependymal-appearing nuclei scattered in a dense fibrillar background of glial processes.

Clinical Features

Posterior fossa ependymomas often manifest with hydrocephalus secondary to progressive obstruction of the fourth ventricle. Despite the relationship of ependymomas to the ventricular system, CSF dissemination is uncommon. The clinical outcome for completely resected ependymomas is considerably better than for their subtotally resected counterparts.

Choroid Plexus Tumors

Both benign and malignant choroid plexus tumors are recognized, all of which are rare. *Choroid plexus papillomas* are intraventricular tumors that are most common in children, in whom they tend to occur in the lateral ventricle; in adults, they more frequently involve the fourth ventricle. These papillary growths recapitulate the structure of the normal choroid plexus, wherein epithelioid tumor cells cover fibrovascular stalks. Clinically, choroid plexus papillomas usually present with hydrocephalus due to obstruction of the ventricular system by tumor or overproduction of CSF. The far rarer *choroid plexus carcinomas* resemble adenocarcinoma; these tumors are almost always found in young children, where metastatic carcinoma is not typically encountered.

Neuronal and Glioneuronal Tumors

Far less common than glial tumors are those that exhibit neuronal differentiation. In general, neuronal tumors are more often seen in children and young adults with epilepsy.

- *Gangliogliomas* (WHO grade I) are tumors composed of a mixture of mature neuronal and glial cells. They are typically superficial lesions that present with seizures and are the most common neuronal tumors in the CNS. Most of these tumors are slow growing. When gangliogliomas present with medically refractory epilepsy, surgical resection is usually effective in controlling the seizures. Approximately 20% to 50% of these tumors have an activating mutation in the *BRAF* gene (V600E), which occasionally serves as the basis of targeted therapy.

  Gangliogliomas are most commonly found in the temporal lobe and often have a cystic component. The neoplastic ganglion cells are irregularly clustered and often have random orientation of processes; dysmorphic and/or binucleate forms are found. The glial component of these lesions most often resembles a pilocytic astrocytoma.
- *Dysembryoplastic neuroepithelial tumor* is a rare, benign (WHO Grade I) tumor often associated with epilepsy. It has a good prognosis following surgical resection, with both low recurrence rates and favorable seizure control. These lesions are typically located in the superficial temporal lobe, although other sites may be seen; they usually form multiple discrete, mucin-rich, intracortical nodules of small, oligodendrocyte-like cells arranged in columns around central axonal bundles or capillaries. There are also "floating neurons" that sit in basophilic, mucin-rich pools.

Embryonal Neoplasms

A large category of mostly pediatric tumors is described as embryonal, meaning that they appear primitive or undifferentiated, although a limited degree of neuronal and less commonly glial differentiation is often encountered. The most common embryonal neoplasm is the *medulloblastoma,* which accounts for 20% of pediatric brain tumors.

Medulloblastoma

This malignant embryonal tumor occurs predominantly in children and exclusively in the cerebellum (by definition). Although histologically, molecularly, and prognostically distinct subtypes have been identified, all are considered grade IV tumors. Rapid growth may occlude the flow of CSF, leading to hydrocephalus.

Pathogenesis

Molecular subtypes of medulloblastoma have been identified through genomic studies, revealing alterations of signaling pathways involved in normal cerebellar development, such as the sonic hedgehog–patched (SHH) pathway (which is involved in control of normal proliferation of cerebellar granule cells) and the WNT/β-catenin signaling pathway. Medulloblastomas in which the Wnt pathway is activated are often located around the anterior fourth ventricle or middle cerebellar peduncle, while sonic hedgehog–activated medulloblastomas most often involve the vermis or lateral cerebellum. On the basis of molecular alterations, medulloblastoma can be divided into four or five molecular groups that vary in terms of age of onset, tumor location, and prognosis.

MORPHOLOGY

Medulloblastoma is often well circumscribed, gray, and friable, and may extend to the surface of the cerebellar folia and involve the leptomeninges (Fig. 28.51A). On microscopic examination, the tumor is very densely cellular, with sheets of small primitive-appearing cells (see Fig. 28.51B); individual tumor cells have scant cytoplasm and hyperchromatic nuclei that are frequently elongated or crescent-shaped. Mitoses are abundant, and markers of cellular proliferation, such as Ki-67, are positive in a high percentage of the cells. In the **classic** subtype, tumor cells form **Homer Wright rosettes** (Fig. 28.51B; also seen in neuroblastoma described in Chapter 10) and often express neuronal markers such as synaptophysin; the expression of glial markers (e.g., GFAP) is less common. The **desmoplastic/nodular variant** is characterized by internodular areas of stromal response, marked by collagen and reticulin deposition (i.e., desmoplasia), and "pale islands" or nodules that have more neuropil and show greater neuronal differentiation (Fig. 28.51C). The **large cell/anaplastic variant**

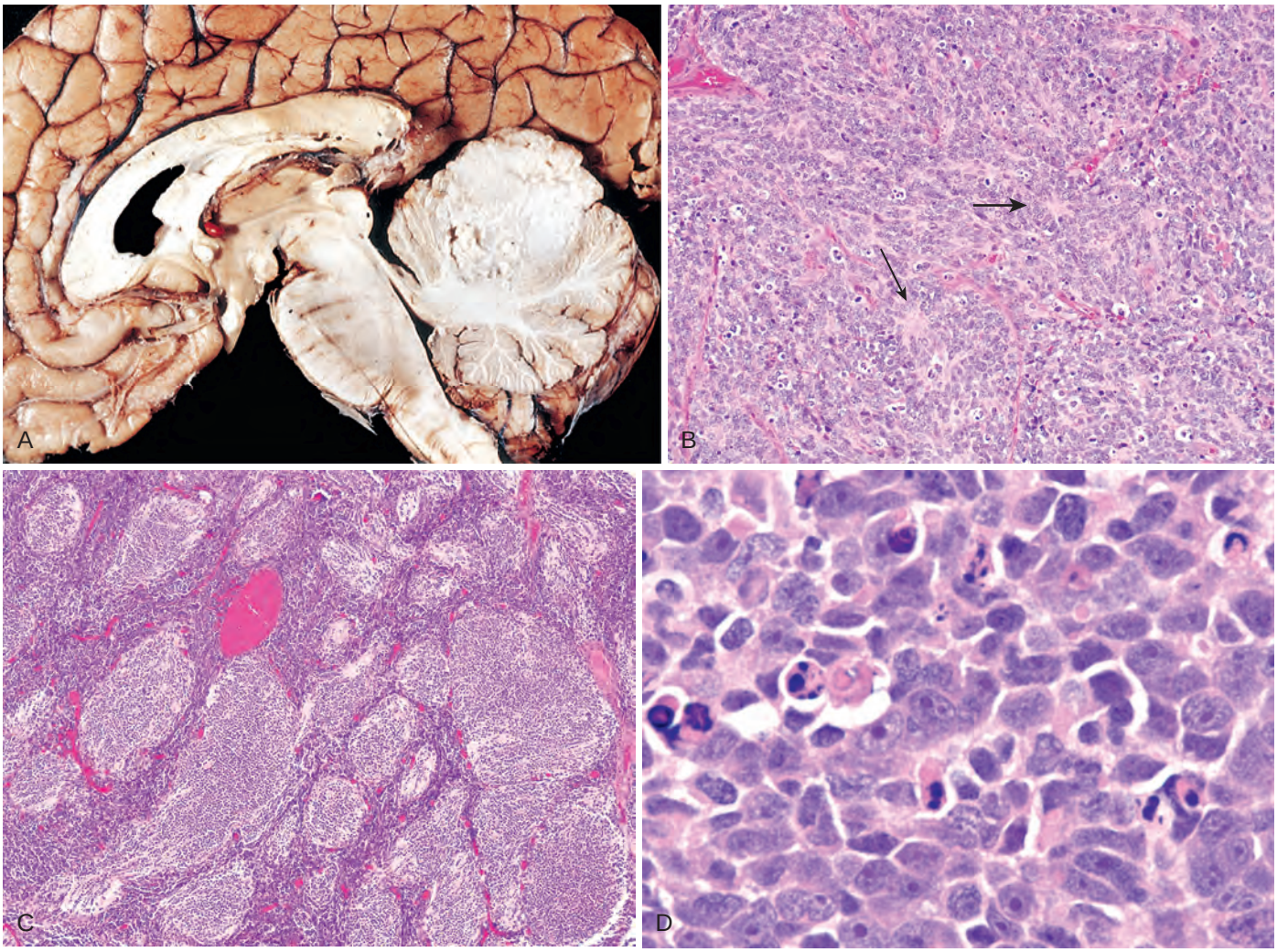


Figure 28.51 Medulloblastoma. (A) Sagittal section of brain showing medulloblastoma replacing part of the superior cerebellar vermis. (B) The microscopic appearance of classic medulloblastoma includes primitive appearing "small blue cells" that form sheets and Homer Wright (neuroblastic) rosettes with central neuropil *(arrows)*. (C) The desmoplastic/nodular variant includes reticulin-rich internodular zones with more primitive appearing cells and "pale nodules" representing centers of partial neuronal differentiation; this variant is nearly always the SHH-activated molecular subtype. (D) The large cell/anaplastic variant features increased cell size, large cherry red nucleoli, and cell wrapping (wherein tumor cells wrap around one another or surround pyknotic nuclei of dying tumor cells).

is characterized by large and irregular vesicular nuclei, prominent nucleoli, cell wrapping (a cell-in-cell appearance), and high mitotic/apoptotic indices (Fig. 28.51D).

Medulloblastomas have a propensity to spread to the subarachnoid space. Dissemination through the CSF is a common complication, giving rise to either extensive growth along the surface likened to "icing," or nodular "drop metastases," which tend to involve the cauda equina.

Clinical Features

Medulloblastomas are highly malignant, and the prognosis for untreated patients is dismal; however, it is exquisitely radiosensitive. The molecular and histologic subtype dramatically influences prognosis, with WNT-activated medulloblastomas associated with nearly 100% survival at 5 years, whereas more aggressive subtypes fall into the 20% to 30% survival range. To account for these differences, ongoing clinical trials are exploring the option of either less aggressive or more aggressive treatment regimens based on subtype. Similarly, targeted therapies are being explored for those with actionable molecular alterations.

Other rare embryonal neoplasms that have unique clinicopathologic and genetic features are now being recognized throughout the CNS; their classification continues to evolve.

Primary CNS Lymphoma

Primary CNS lymphoma accounts for 2% of extranodal lymphomas and 1% of intracranial tumors. It is the most common CNS neoplasm in immunosuppressed individuals, including those with AIDS and immunosuppression after transplantation. In non-immunosuppressed populations, the age spectrum is relatively wide, but the frequency increases after 60 years of age.

The term *primary* emphasizes the distinction between these lesions and secondary involvement of the CNS by

lymphoma arising elsewhere in the body (Chapter 13). Primary brain lymphoma is often multifocal and may also involve the eye, yet involvement outside of the CNS (in lymph nodes or bone marrow) is a rare and late complication. Conversely, lymphoma arising outside the CNS rarely involves the brain parenchyma; in this situation, secondary CNS involvement usually affects the meninges or the CSF, the latter sometimes diagnosed by the presence of malignant cells in a lumbar puncture specimen.

The vast majority of primary CNS tumors are diffuse large B-cell lymphomas; they are aggressive and generally have worse outcomes than tumors of comparable histology that occur at non-CNS sites (Chapter 13). In the setting of immunosuppression, the malignant B cells are usually latently infected by Epstein-Barr virus. In those who are not immunosuppressed, primary CNS large B-cell lymphomas often have amplification and overexpression of the *PDL1* gene, which encodes an important immune checkpoint protein that inhibits T-cell responses (Chapter 7). Trials of immune checkpoint inhibitors are ongoing and have produced excellent responses in some patients.

MORPHOLOGY

Primary CNS lymphomas are frequently multiple, soft gray-white masses in the deep subcortical parts of the brain; periventricular spread is common. The malignant cells infiltrate the parenchyma of the brain and accumulate around blood vessels, express B-cell markers such as CD20, and have high proliferative indices. When tumors arise in the setting of immunosuppression, necrosis is often prominent and the Epstein-Barr virus is detectable by in situ hybridization for EBERs, small nuclear RNAs that are encoded by the viral genome.

Meningiomas

Meningiomas are predominantly benign tumors of adults that arise from the meningothelial cells of the arachnoid and are usually attached to the dura. Meningiomas may be found along any of the external surfaces of the brain as well as within the ventricular system, where they arise from the stromal arachnoid cells of the choroid plexus. Prior radiation therapy to the head and neck, typically decades earlier, is a risk factor for development of meningioma. Other tumors such as metastases, solitary fibrous tumors, and a range of poorly differentiated sarcomas may also grow as dura-based masses.

Pathogenesis

The most common cytogenetic abnormality is loss of chromosome 22, especially the long arm (22q). The deletions include the region of 22q12 that harbors the *NF2* gene, which encodes the protein merlin; as expected, meningiomas are a common lesion in the setting of NF2 (see later). Of sporadic meningiomas, 50% to 60% harbor mutations in the *NF2* gene; these are most commonly seen in convexity tumors, including those that eventually become higher grade. In meningiomas without *NF2* mutations, the most common mutations occur in *TRAF7, KLF4, AKT1,* and *SMO;* most such tumors involve the skull base and have a lower risk of malignant progression. Higher-grade meningiomas more often have chromosomal losses involving several chromosomes, *TERT*-promoter mutations, and homozygous *CDKN2A* gene deletions.

MORPHOLOGY

Meningiomas are usually rubbery, rounded, or bosselated dural masses that compress underlying brain but are easily separated from it (Fig. 28.52A). They may also grow **en plaque,** in which the tumor spreads in a sheetlike fashion along the surface of the dura; this form is commonly associated with hyperostotic changes in the adjacent bone due to invasion, although this by itself does not signify malignancy. The lesions range from firm to finely gritty (sandlike), the latter typically due to numerous psammomatous calcifications ("psammos" is Greek for *sand*).

Most meningiomas have a relatively low risk of recurrence or aggressive growth, and so are considered WHO grade I. Many histologic patterns are observed, with most showing no prognostic

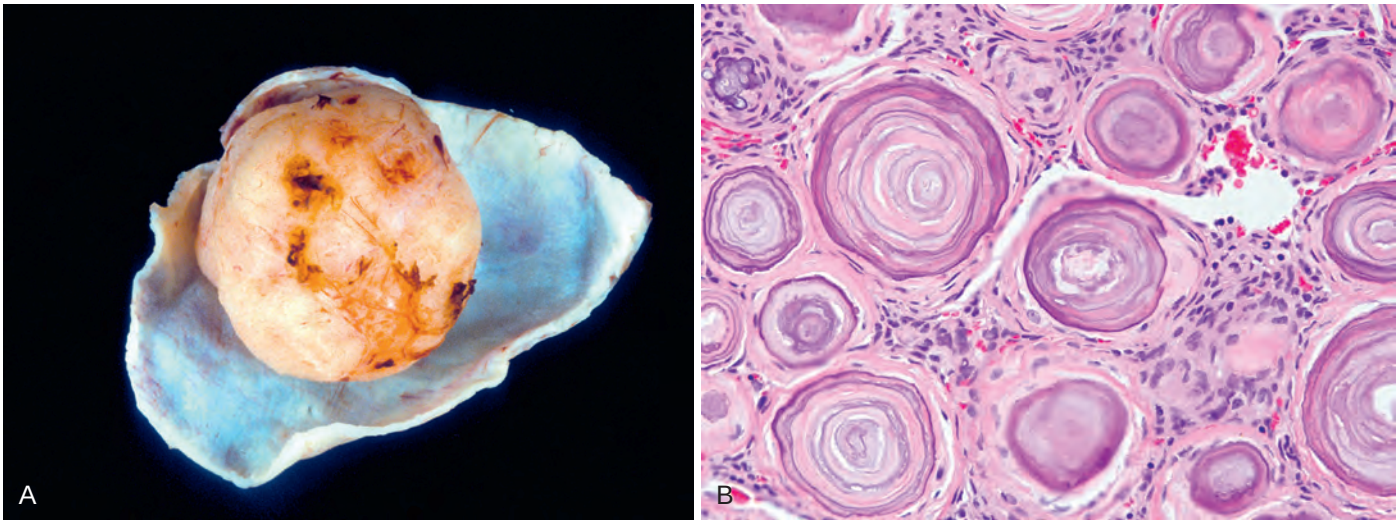


Figure 28.52 (A) Resected meningioma specimen showing rounded contour and dural attachment. (B) Meningioma with a whorled pattern of cell growth and numerous psammoma bodies (calcifications with concentric rings).

significance. The most common include **meningothelial,** containing clusters of epithelioid cells with fuzzy or indiscernible cell membranes; **fibroblastic,** with intersecting fascicles of spindled cells and abundant collagen deposition; **transitional,** with mixed meningothelial and fibroblastic features, often including many whorls (concentric wrapping of tumor cells around one another); and **psammomatous,** with a predominance of psammoma bodies, the latter representing concentric rings of calcification deposited over pre-existing whorls (see Fig. 28.52B).

Atypical meningiomas (WHO grade II) represent about one-fourth of meningiomas and are associated with higher recurrence rates and more aggressive local growth, which may require radiation therapy in addition to surgery. They are distinguished from lower-grade meningiomas by having an increased mitotic index, brain invasion, or other aggressive features. Certain histologic patterns (**clear cell** and **chordoid**) are considered grade II by definition because of their more aggressive behavior.

Anaplastic (malignant) meningioma (WHO grade III) is a highly aggressive malignancy that often resembles a carcinoma or sarcoma, but retains some evidence of meningothelial origin. They account for 1% to 3% of all meningiomas and may arise de novo or from malignant progression of lower-grade meningiomas. Mitotic rates are often markedly elevated. **Papillary** and **rhabdoid** meningiomas are rare WHO grade III variants that also feature high recurrence and mortality rates.

Clinical Features

Meningiomas are usually slow-growing tumors. Patients present either with vague nonlocalizing symptoms or with focal findings referable to compression of underlying brain. Common sites of involvement include the parasagittal aspect of the brain convexity, dura over the lateral convexity, wing of the sphenoid, olfactory groove, sella turcica, and foramen magnum. Meningiomas are uncommon in children and generally show a moderate (3:2) female predominance, although the ratio is 10:1 for spinal meningiomas (which are also commonly psammomatous). Lesions are usually solitary, but when present at multiple sites, especially in association with vestibular schwannoma and/or spinal ependymoma, the possibility of NF2 should be considered. Meningiomas often express a variety of hormone (e.g., progesterone) receptors and may grow more rapidly during pregnancy, only to regress after delivery.

Metastatic Tumors

Metastatic lesions, mostly carcinomas, account for approximately one-fourth to one-half of intracranial tumors in hospitalized patients. The five most common primary sites are lung, breast, skin (melanoma), kidney, and gastrointestinal tract, accounting for about 80% of all metastases combined. Some rare tumors (e.g., choriocarcinoma) have a high likelihood of metastasizing to the brain, whereas other more common tumors (e.g., prostatic adenocarcinoma) almost never do so. The meninges are also a frequent site of involvement by metastatic disease, where they may either form discrete masses or a more disseminated subarachnoid, meningitis-like pattern known as "meningeal carcinomatosis"; the latter is often highly challenging to treat and is associated with both cranial and spinal neuropathies from nerve root compression. Metastatic tumors present clinically as mass lesions and may occasionally be the first manifestation of systemic cancer. In general, localized treatment of solitary brain metastases improves the quality and duration of the patient's life.

MORPHOLOGY

Intraparenchymal metastases form sharply demarcated masses, often at the junction of gray and white matter, usually surrounded by a zone of edema. The boundary between tumor and brain parenchyma is usually well-defined microscopically, although melanoma does not always follow this rule and may be more infiltrative. The tumors typically resemble their parent tumor type, but can be very poorly differentiated in some cases; for the roughly 10% that present as a metastasis of unknown primary, immunohistochemistry is often helpful for further subtyping. Meningeal carcinomatosis, with tumor nodules studding the surface of the brain, spinal cord, and intradural nerve roots, is most commonly associated with carcinomas of the lung and breast.

Paraneoplastic Syndromes

In addition to the direct and localized effects produced by CNS metastases, peripheral neoplasms sometimes produce *paraneoplastic syndromes* that involve the peripheral and/or central nervous systems and can precede the clinical recognition of the primary tumor. A variety of paraneoplastic syndromes has been described; a shared underlying mechanism appears to be the development of an immune response against tumor antigens that cross-reacts with antigens in the central or peripheral nervous system. The spectrum of known circulating antibodies and target antigens continues to expand; illustrative CNS examples are as follows:

- *Subacute cerebellar degeneration* is associated with destruction of Purkinje cells, gliosis, and a chronic inflammatory cell infiltrate. One group of affected patients has a circulating PCA-1 antibody (anti-Yo) that recognizes cerebellar Purkinje cells; this antibody occurs predominantly in women with ovarian, uterine, or breast carcinomas.
- *Limbic encephalitis* is characterized by subacute dementia and shows perivascular inflammatory cuffs, microglial nodules, some neuronal loss, and gliosis, all of which are most evident in the anterior and medial portions of the temporal lobe; the microscopic picture resembles that of a viral encephalitis. A comparable process involving the brainstem can be seen in isolation or together with limbic system involvement. Some affected patients have a circulating ANNA-1 antibody (anti-Hu) that recognizes neuronal nuclei in the central and peripheral nervous systems; ANNA-1 is most commonly associated with small cell carcinoma of the lung. Another group of patients has a circulating antibody that recognizes the NMDA receptor and cross-reacts with hippocampal neurons. Originally identified in women with ovarian teratomas, the same clinical syndrome is now also recognized in a small proportion of patients with sporadic encephalitis. A third group of patients has a circulating VGKC-complex antibody that recognizes the voltage-gated potassium channel; this antibody may also be associated with peripheral neuropathy.

- *Eye movement disorders,* most commonly opsoclonus, may occur by themselves or in association with other evidence of cerebellar and brainstem dysfunction. In children, this is most commonly associated with neuroblastoma and is often accompanied by myoclonus.

The peripheral nervous system can also be affected by the following paraneoplastic syndromes:

- *Subacute sensory neuropathy* may be found in association with limbic encephalitis or in isolation. It is marked by loss of sensory neurons and lymphocytic inflammation in the dorsal root ganglia.
- *Lambert-Eaton myasthenic syndrome* is caused by antibodies against the voltage-gated calcium channel in the presynaptic elements of the neuromuscular junction. This syndrome can also be seen in the absence of malignancy.

For some paraneoplastic syndromes, there is evidence that immunotherapy (removal of circulating antibodies and immunosuppression) and tumor removal lead to clinical improvement. In general, clinical syndromes associated with plasma membrane–reactive antibodies (e.g., antibodies that recognize VGKC or NMDA receptor) respond to immunotherapy better than those associated with intracellular antigens (e.g., ANNA-1 and PCA-1).

Familial Tumor Syndromes

A number of inherited diseases is associated with increased risk of neoplasms (Chapter 7). In several of these (discussed in the following sections), tumors of the nervous system are a prominent aspect of the disease.

Neurofibromatoses and Schwannomatosis

The two forms of neurofibromatosis, NF1 and NF2, are familial autosomal dominant syndromes characterized by tumors of the peripheral and central nervous systems; in contrast, schwannomatosis involves the peripheral nervous system only and may be either familial or sporadic. NF1 is most common, with a frequency of 1 in 3000, and is characterized by neurofibromas of peripheral nerves, gliomas of the optic nerve, pigmented nodules of the iris (Lisch nodules), and cutaneous hyperpigmented macules (café au lait spots). NF2 has a frequency of 1 in 40,000 to 50,000 and is characterized by the occurrence of bilateral vestibular (cranial nerve VIII) schwannomas, multiple meningiomas, and ependymomas of the cervical spinal cord. Schwannomatosis, roughly as common as NF2, is the most recently recognized syndrome; it is defined by multiple nonvestibular schwannomas, either throughout the body or limited to one region. These disorders are discussed in greater detail in Chapter 27.

Tuberous Sclerosis Complex

Tuberous sclerosis complex (TSC) is an autosomal dominant syndrome that occurs at a frequency of approximately 1 in 6000 births. It is characterized by the development of hamartomas and benign neoplasms involving the brain and other tissues; the most frequent clinical manifestations are seizures, autism, and intellectual disability. Hamartomas within the CNS take the form of cortical tubers and subependymal nodules; subependymal giant cell astrocytomas are benign neoplasms that appear to develop from the hamartomatous nodules in the same location. Cortical tubers are often epileptogenic, and surgical resection can be beneficial when medical management of the seizures fails. Elsewhere in the body, renal angiomyolipomas, retinal glial hamartomas, pulmonary lymphangioleiomyomatosis, and cardiac rhabdomyomas develop during childhood and adolescence. Cysts may be found at various sites, including the liver, kidneys, and pancreas. Cutaneous lesions include angiofibromas, localized leathery thickenings (shagreen patches), hypopigmented areas (ash-leaf patches), and subungual fibromas.

One tuberous sclerosis gene *(TSC1)* is found on chromosome 9q34 and encodes a protein known as hamartin; the more commonly mutated gene *(TSC2)* is on 16p13.3 and encodes tuberin. These two proteins form a complex that inhibits the kinase mTOR, a key regulator of protein synthesis and other aspects of anabolic metabolism. Mutations in *TSC1* or *TSC2* disrupt this control and lead to increased and unregulated mTOR activity. Of note, mTOR controls cell size, and the tumor cells associated with TSC, particularly the subependymal giant cell astrocytomas, are remarkable for having voluminous cytoplasm. Treatment is usually symptomatic, including anticonvulsant therapy for control of seizures; however, treatment with mTOR inhibitors has sometimes resulted in dramatic clinical improvement.

MORPHOLOGY

Cortical hamartomas of TSC are firm areas that, in contrast to the softer adjacent cortex, have been likened to potatoes (hence the appellation "tubers"). The tubers are composed of haphazardly arranged neurons that lack the normal laminar organization. In addition, some dysplastic cells have an appearance intermediate between glia and neurons (large vesicular nuclei with nucleoli, resembling neurons, and abundant eosinophilic cytoplasm resembling gemistocytic astrocytes); they often express intermediate filaments of both neuronal (neurofilament) and glial (GFAP) types. Similar hamartomatous features are present in the subependymal nodules, where the large astrocyte-like cells cluster beneath the ventricular surface. These small waxlike masses that bulge into the ventricular system gave rise to the term **candle-gutterings.** Subependymal giant cell astrocytoma is histologically similar, but larger; it often presents with obstructive hydrocephalus.

Von Hippel-Lindau Disease

Individuals with this autosomal dominant disease develop hemangioblastomas of the CNS; cysts that involve the pancreas, liver, and kidneys; renal cell carcinomas; and pheochromocytomas. Hemangioblastomas are most common in the cerebellum and retina, but may also occur in other CNS locations. The disease frequency is 1 in 30,000 to 40,000.

The gene associated with von Hippel-Lindau disease, called *VHL,* is a tumor-suppressor gene located on chromosome 3p25.3; it encodes a protein (VHL) that is a component of a ubiquitin ligase complex that down-regulates hypoxia-induced factor 1 (HIF-1). HIF-1 is a transcription factor that regulates expression of vascular endothelial growth factor (VEGF), erythropoietin, and other growth factors. VEGF leads to excessive vessel growth, contributing to the development of hemangioblastomas. Erythropoietin overexpression leads to polycythemia in about 10% of hemangioblastoma patients.

HIF also regulates the expression of genes that control cellular metabolism and cell growth, activities that likely contribute to tumor formation.

MORPHOLOGY

Hemangioblastomas are highly vascular neoplasms that form a mural nodule associated with a large fluid-filled cyst. The lesion consists of numerous capillary-size or somewhat larger thin-walled vessels with intervening neoplastic cells that have vacuolated, lipid-rich cytoplasm. The neoplastic stromal cells express erythropoietin, VEGF, and inhibin (a diagnostically useful marker), and are believed to be derived from an early mesenchymal progenitor cell capable of differentiating into both endothelial and hemopoietic cells.

Therapy is directed at the symptomatic neoplasms, including resection of the cerebellar hemangioblastomas and laser therapy for retinal hemangioblastomas.

KEY CONCEPTS

TUMORS

- Tumors of the CNS may arise from the brain coverings (meningiomas), the brain parenchyma (gliomas, neuronal tumors, choroid plexus tumors), and other resident CNS cell populations (primary CNS lymphoma); they can also originate elsewhere in the body (metastases).
- Even low-grade or benign tumors can have poor clinical outcomes, depending on where they occur in the brain.
- Distinct types of tumors affect specific brain regions (e.g., cerebellum for medulloblastoma) and specific age populations (medulloblastoma and pilocytic astrocytoma in pediatric age groups; glioblastoma and lymphoma in older patients).
- Glial tumors are broadly classified into astrocytomas, oligodendrogliomas, and ependymomas. Increasing tumor malignancy is associated with more cytologic anaplasia, increased cell density, necrosis, and mitotic activity. Diagnostic, prognostic, and/or predictive genetic alterations have recently been recognized in some of these tumors.
- Metastatic spread of brain tumors to other regions of the body is rare, but the brain is a common recipient of metastatic disease from systemic malignancies. Carcinomas are most common.

ACKNOWLEDGMENT

The contributions of Dr. Matt Frosch to previous versions of this chapter are gratefully acknowledged.

SUGGESTED READINGS

General

Many areas of neuropathology and neurologic diseases are well covered in the following standard texts:

Ellison D, Love S, Chimelli LMC et al: *Neuropathology: a Reference Text of CNS Pathology*, ed 3, London, 2013, Elsevier Mosby.

Louis DN, Ohgaki H, Wiestler OD et al: *WHO Classification of Tumours of the Central Nervous System*, ed 4, Lyon, 2016, International Agency for Research on Cancer.

Love S, Budka H, Ironside JW et al, editors: *Greenfield's Neuropathology*, ed 9, Boca Raton, 2015, CRC Press.

Response to Injury

Li Q, Barres BA: Microglia and macrophages in brain homeostasis and disease, *Nat Rev Immunol* 18:225, 2018.

Liddelow SA, Barres BA: Reactive astrocytes: production, function, and therapeutic potential, *Immunity* 46:957, 2017.

Presumey J, Bialas AR, Carroll MC: Complement system in neural synapse elimination in development and disease, *Adv Immunol* 135:53, 2017.

Malformations and Developmental Diseases

Adle-Biassette H, Golden JA, Harding B: Developmental and perinatal brain diseases, *Handb Clin Neurol* 145:51, 2017.

Molloy AM, Pangilinan F, Brody LC: Genetic risk factors for folate-responsive neural tube defects, *Annu Rev Nutr* 37:269, 2017.

Parrini E, Conti V, Dobyns WB et al: Genetic basis of brain malformations, *Mol Syndromol* 7:220, 2016.

Trauma

Asken BM, Sullan MJ, DeKosky ST et al: Research gaps and controversies in chronic traumatic encephalopathy: a Review, *JAMA Neurol* 74:1255, 2017.

Hay J, Johnson VE, Smith DH et al: Chronic traumatic encephalopathy: the neuropathological legacy of traumatic brain injury, *Annu Rev Pathol* 11:21, 2016.

Cerebrovascular Disease

Banerjee G, Carare R, Cordonnier C et al: The increasing impact of cerebral amyloid angiopathy: essential new insights for clinical practice, *J Neurol Neurosurg Psychiatry* 88:982, 2017.

McAleese KE, Alafuzoff I, Charidimou A et al: Post-mortem assessment in vascular dementia: advances and aspirations, *BMC Med* 14:129, 2016.

Nikolaev SI, Vetiska S, Bonilla X et al: Somatic activating KRAS mutations in arteriovenous malformations of the brain, *N Engl J Med* 378:250, 2018.

Infections

Devakumar D, Bamford A, Ferreira MU et al: Infectious causes of microcephaly: epidemiology, pathogenesis, diagnosis, and management, *Lancet Infect Dis* 18:e1, 2018.

Farhadian S, Patel P, Spudich S: Neurological complications of HIV infection, *Curr Infect Dis Rep* 19:50, 2017.

Kennedy PGE, Quan PL, Lipkin WI: Viral encephalitis of unknown cause: current perspective and recent advances, *Viruses* 9:138, 2017.

Ru W, Tang SJ: HIV-associated synaptic degeneration, *Mol Brain* 10:40, 2017.

Venkatesan A: Epidemiology and outcomes of acute encephalitis, *Curr Opin Neurol* 28:277, 2015.

Demyelinating Diseases

Mallucci G, Peruzzotti-Jametti L, Bernstock JD et al: The role of immune cells, glia and neurons in white and gray matter pathology in multiple sclerosis, *Prog Neurobiol* 127-128:1, 2015.

Pierrot-Deseilligny C, Souberbielle JC: Vitamin D and multiple sclerosis: an update, *Mult Scler Relat Disord* 14:35, 2017.

Pittock SJ, Lucchinetti CF: Neuromyelitis optica and the evolving spectrum of autoimmune aquaporin-4 channelopathies: a decade later, *Ann N Y Acad Sci* 1366:20, 2016.

Prion Diseases

Erana H, Venegas V, Moreno J et al: Prion-like disorders and transmissible spongiform encephalopathies: an overview of the mechanistic features that are shared by the various disease-related misfolded proteins, *Biochem Biophys Res Commun* 483:1125, 2017.

Will RG, Ironside JW: Sporadic and infectious human prion diseases, *Cold Spring Harb Perspect Med* 7:a024364, 2017.

Neurodegenerative Diseases

Cooper-Knock J, Shaw PJ, Kirby J: The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype, *Acta Neuropathol* 127:333, 2014.

Chung WS, Verghese PB, Chakraborty C et al: Novel allele-dependent role for APOE in controlling the rate of synapse pruning by astrocytes, *Proc Natl Acad Sci USA* 113:10186, 2016.

Gegg ME, Schapira AHV: The role of glucocerebrosidase in Parkinson disease pathogenesis, *FEBS J* 2018. Epub ahead of print.

Hernandez DG, Reed X, Singleton AB: Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance, *J Neurochem* 139 Suppl 1:59, 2016.

Klingelhoefer L, Reichmann H: Pathogenesis of Parkinson disease—the gut-brain axis and environmental factors, *Nat Rev Neurol* 11:625, 2015.

Lane CA, Hardy J, Schott JM: Alzheimer's disease, *Eur J Neurol* 25:59, 2018.

Mann DMA, Snowden JS: Frontotemporal lobar degeneration: pathogenesis, pathology and pathways to phenotype, *Brain Pathol* 27:723, 2017.

Rossi S, Cozzolino M, Carri MT: Old versus new mechanisms in the pathogenesis of ALS, *Brain Pathol* 26:276, 2016.

Shi Y, Yamada K, Liddelow SA et al: ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy, *Nature* 549:523, 2017.

Sun YM, Lu C, Wu ZY: Spinocerebellar ataxia: relationship between phenotype and genotype - a review, *Clin Genet* 90:305, 2016.

Ugalde CL, Finkelstein DI, Lawson VA et al: Pathogenic mechanisms of prion protein, amyloid-beta and alpha-synuclein misfolding: the prion concept and neurotoxicity of protein oligomers, *J Neurochem* 139:162, 2016.

Genetic Metabolic Diseases

Craven L, Alston CL, Taylor RW et al: Recent advances in mitochondrial disease, *Annu Rev Genomics Hum Genet* 18:257, 2017.

Lax NZ, Gorman GS, Turnbull DM: Review: central nervous system involvement in mitochondrial disease, *Neuropathol Appl Neurobiol* 43:102, 2017.

van der Knaap MS, Bugiani M: Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms, *Acta Neuropathol* 134:351, 2017.

Tumors

Louis DN, Perry A, Reifenberger G et al: The 2016 World Health Organization classification of tumors of the central nervous system: a summary, *Acta Neuropathol* 131:803, 2016.

Perry A, Brat D: *Practical Surgical Neuropathology*, ed 2, Philadelphia, 2018, Elsevier/Churchill Livingstone.

See TARGETED THERAPY available online at www.studentconsult.com

CHAPTER 29

The Eye

Robert Folberg

CHAPTER CONTENTS

Although this chapter comes at the end of the book, it is not an afterthought. Vision is a major quality-of-life issue. Before the public awareness of acquired immunodeficiency syndrome (AIDS) and Alzheimer disease, the most feared disease among Americans was cancer, and the second most feared disease was blindness. So great is the fear of blindness that even today, people often tell their physicians, "Doctor, I'd rather be dead than be blind!"

In general, diseases that produce loss of vision do not attract as much attention as do many of the life-threatening conditions described in this book. For example, age-related macular degeneration (AMD) is the most common cause of irreversible visual loss in the United States. Most individuals with AMD do not suffer from a total loss of vision—an immersion into darkness. The histopathology is unimpressive: small scars develop in the macula. But consider the effect of these tiny scars in a retired schoolteacher with AMD. The central portion of her or his vision is lost. The faces of spouse or grandchildren are not visible. He or she cannot read a book or newspaper. Once a model of independence, this teacher can no longer drive a car and must be chauffeured everywhere. In short, this person is robbed of the common joys that most of us take for granted.

To study the eye, one needs to comprehend all that has come before. For example, the pathology of the eyelids builds on knowledge of dermatopathology (see Chapter 25), and the pathology of the retina and optic nerve extends what was learned in Chapter 28 about the brain and central nervous system. However, the study of ocular pathology does not merely repeat what has been presented thus far. **The eye provides the only site in which a physician can directly visualize a variety of microcirculatory disturbances ranging from arteriosclerosis to angiogenesis in the clinic.** Although there are conditions that are unique to the eye (e.g., cataract and glaucoma), many ocular conditions share similarities with disease processes elsewhere in the body that are modified by the unique structure and function of the eye (Fig. 29.1).

In recent years, the elucidation of the molecular pathogenesis of disease has been translated rapidly to therapeutic applications in the eye. Many blinding conditions, such as corneal neovascularization, diabetic retinopathy, and certain

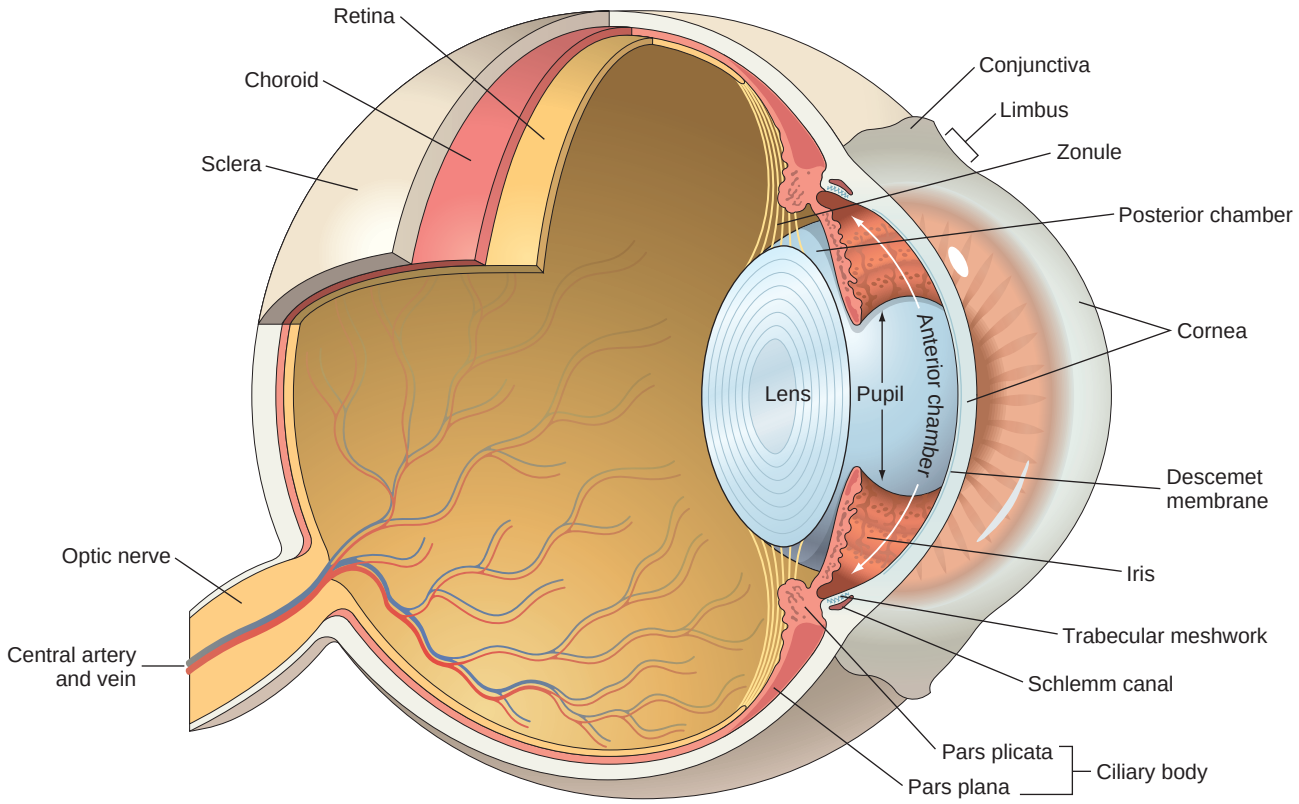


Figure 29.1 Anatomy of the eye.

forms of age-related neovascularization, result from pathologic angiogenesis. Successful treatment of these conditions with vascular endothelial growth factor (VEGF) antagonists has saved vision in patients who might have been blinded just a few years ago.

This chapter is organized on the basis of ocular anatomy. The discussion of each region of the eye begins with anatomic and functional considerations and their impact on the understanding of ocular diseases.

ORBIT

Functional Anatomy and Proptosis

The orbit is a compartment that is closed medially, laterally, and posteriorly. Diseases that increase orbital contents therefore displace the eye forward, a condition known as *proptosis*. Aside from the obvious cosmetic concerns, the proptotic eye might not be covered completely by the eyelids, and the tear film might not be distributed evenly across the cornea. Chronic corneal exposure to air is injurious, leading to pain and predisposing to corneal ulceration and infection. Proptosis may be axial (directly forward) or positional. For example, any enlargement of the lacrimal gland from inflammation (e.g., *sarcoidosis*) or neoplasm (e.g., *lymphoma, pleomorphic adenoma,* or *adenoid cystic carcinoma*) produces a proptosis that displaces the eye inferiorly and medially because the lacrimal gland is positioned superotemporally within the orbit.

Masses contained within the cone formed by the horizontal rectus muscles generate axial proptosis: the eye bulges straight forward. The two most common primary tumors of the optic nerve (a tract of the central nervous system), *glioma* and *meningioma,* produce axial proptosis because the optic nerve is positioned within the muscle cone. The orbital contents are subject to the same disease processes that affect other tissues. Representative inflammatory conditions and neoplasms of the orbit are discussed briefly next.

Thyroid Ophthalmopathy (Graves Disease)

In the chapter on endocrine disorders (Chapter 24) it was noted that axial proptosis is an important clinical manifestation of Graves disease. Proptosis is caused by the accumulation of extracellular matrix proteins and variable degrees of fibrosis in the rectus muscles (Fig. 29.2). The development of thyroid ophthalmopathy may, in some cases, be independent of the status of thyroid function.

Other Orbital Inflammatory Conditions

The floor of the orbit is the roof of the maxillary sinus, and the medial wall of the orbit—the lamina papyracea—separates the orbit from the ethmoidal sinuses. As a result, uncontrolled sinus infection may spread to the orbit either as an acute bacterial infection or as a component of a fungal infection. This occurs most commonly in immunosuppressed individuals, in patients with diabetic ketoacidosis, or, rarely, in persons without any predisposition. Systemic conditions such as *granulomatosis with polyangiitis* (Chapter 11) may present first in the orbit and may be confined there for

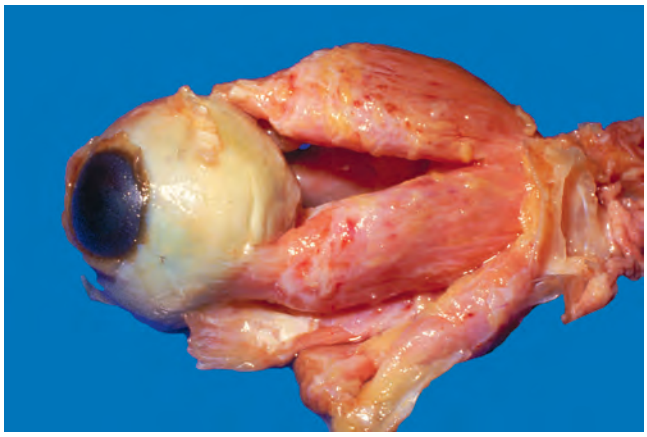


Figure 29.2 The extraocular muscles are greatly distended in this postmortem dissection of tissues from a patient with thyroid (Graves) ophthalmopathy. Note that the tendons of the muscles are spared. (Courtesy Dr. Ralph C. Eagle, Jr., Wills Eye Hospital, Philadelphia, Pa.)

prolonged periods of time, or alternatively, it may involve the orbit secondarily by extension from the sinuses.

Idiopathic orbital inflammation, also known as orbital inflammatory pseudotumor (Fig. 29.3), is another inflammatory condition affecting the orbit. This condition may be unilateral or bilateral and may affect all orbital tissue elements or may be confined to the lacrimal gland *(sclerosing dacryoadenitis),* extraocular muscles *(orbital myositis),* or Tenon capsule, the fascial layer that wraps around the eye *(posterior scleritis).* IgG4-related disease (Chapter 6) should be excluded before declaring an orbital inflammation to be idiopathic.

MORPHOLOGY

Idiopathic orbital inflammation is characterized histologically by chronic inflammation and variable degrees of fibrosis. The inflammatory infiltrate typically includes lymphocytes and plasma cells and occasionally eosinophils. Germinal centers, when present, raise the suspicion of a reactive lymphoid hyperplasia. Vasculitis may be present, suggesting an underlying systemic condition. The presence of necrosis and degenerating collagen along with vasculitis should raise the suspicion of granulomatosis with polyangiitis. Idiopathic orbital inflammation is typically confined to the orbit but may develop concomitantly with sclerosing inflammation in the retroperitoneum, the mediastinum, and the thyroid, especially as a manifestation of IgG4-related disease.

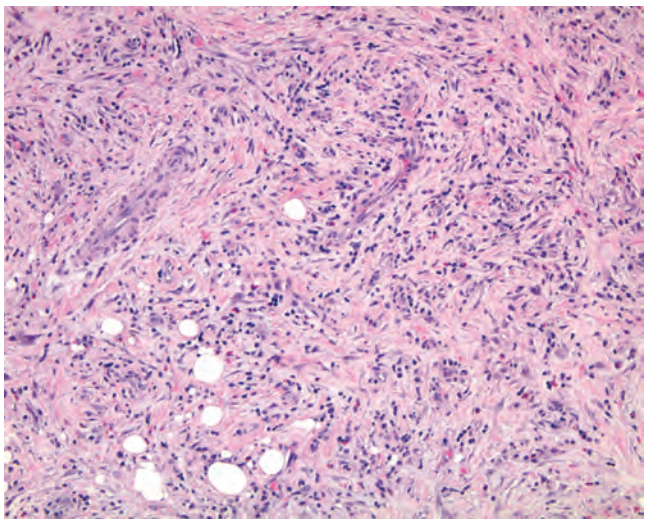


Figure 29.3 In idiopathic orbital inflammation (orbital inflammatory pseudotumor) the orbital fat is replaced by fibrosis. Note the chronic inflammation, accompanied in this case by eosinophils.

Neoplasms

The most frequently encountered primary neoplasms of the orbit are vascular in origin: capillary hemangioma of infancy and early childhood and lymphangioma (both of which are unencapsulated) and encapsulated cavernous hemangioma found typically in adults. These are described in other chapters. Only a handful of orbital masses are encapsulated (e.g., pleomorphic adenoma of the lacrimal gland, dermoid cyst, neurilemmoma), and the recognition of encapsulation on imaging studies allows the surgeon to anticipate pathologic findings.

Non-Hodgkin lymphoma, like idiopathic orbital inflammation, can affect the entire orbit or can be confined to compartments of the orbit such as the lacrimal gland. Orbital lymphomas are classified according to the World Health Organization (WHO) classification system (Chapter 13).

Primary orbital malignancies may arise from any of the orbital tissues and are classified according to the scheme used for the parent tissue. For example, the lacrimal gland may be considered a minor salivary gland, and tumors of the lacrimal gland are classified as salivary gland tumors.

Metastases to the orbit may present with distinctive signs and symptoms that point to the origin of the tumor. For example, metastatic prostatic carcinoma may present clinically like idiopathic orbital inflammation; metastatic neuroblastoma and Wilms tumor—richly vascular neoplasms—may produce characteristic periocular ecchymoses. Neoplasms may also invade from the sinuses into the orbit.

KEY CONCEPTS

- Proptosis results from lesions or pathologic changes in tissue that occupy space in the orbit. The orbit is a compartment that is only open anteriorly and is closed in all other dimensions by bone.
- Inflammation in the orbit may develop by extension of local disease in adjacent tissues (e.g., sinusitis) or as a component of systemic disease (e.g., granulomatosis with polyangiitis).
- The most common primary tumors of the orbit are vascular (e.g., capillary and cavernous hemangiomas).

EYELID

Functional Anatomy

The eyelid is composed of skin externally and mucosa (the conjunctiva) on the surface apposed to the eye (Fig. 29.4). In addition to covering and protecting the eye, elements within

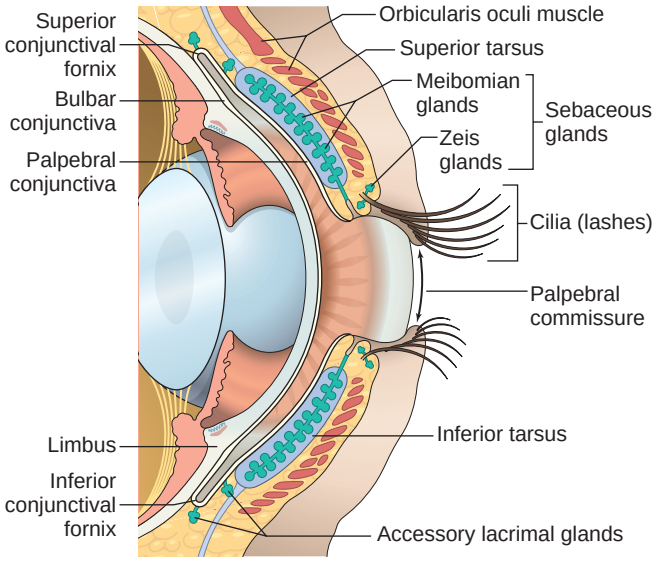


Figure 29.4 Anatomy of the conjunctiva and eyelids.

the eyelid generate critical components of the tear film. If the drainage system of the sebaceous glands is obstructed by chronic inflammation at the eyelid margin *(blepharitis)* or, less commonly, by neoplasm, lipid may extravasate into surrounding tissue and provoke a granulomatous response producing a lipogranuloma, or *chalazion*.

Neoplasms

The most common malignancy of the eyelid is basal cell carcinoma. Surprisingly, primary melanomas of the eyelid skin are extremely rare. Regardless of histogenesis, eyelid neoplasms may distort tissue and prevent the eyelids from closing completely. Because chronic exposure to air damages the cornea, prompt treatment of locally invasive basal cell carcinomas is imperative to preserve vision. Basal cell carcinoma has a distinct predilection for the lower eyelid and the medial canthus.

Sebaceous carcinoma may form a local mass that mimics *chalazion* or may diffusely thicken the eyelid. This neoplasm may also resemble inflammatory processes such as blepharitis or *ocular cicatricial pemphigoid* because of a predilection for intraepithelial spread as occurs in Paget disease of the nipple (Chapter 23) or vulva (Chapter 22). Sebaceous carcinoma tends to spread first to the parotid and submandibular nodes. The overall mortality rate can be as high as 22%. Sebaceous carcinoma of the eyelid is less likely to be associated with Muir-Torre syndrome than sebaceous neoplasms developing elsewhere.

MORPHOLOGY

In moderately differentiated or well-differentiated sebaceous carcinoma, vacuolization of the cytoplasm is present and helps in the diagnosis. This cancer may, however, resemble a variety of other malignancies histologically, including basal cell carcinoma; hence establishing the correct diagnosis can be difficult. Pagetoid spread (Fig. 29.5) may mimic Bowenoid actinic keratosis in the eyelid and carcinoma in situ in the conjunctiva. Sebaceous carcinoma may spread through the conjunctival epithelium and the epidermis to the lacrimal drainage system and the nasopharynx. It may also extend into the lacrimal gland ductules and thereby into the main lacrimal gland.

In individuals with AIDS, *Kaposi sarcoma* may develop in either the eyelid or the conjunctiva. In the eyelid the lesion may appear clinically to have a purple hue because the vascular lesion is embedded in the dermis, but in the thin mucous membrane of the conjunctiva, Kaposi sarcoma appears bright red and may be confused clinically with a subconjunctival hemorrhage.

KEY CONCEPTS

- Basal cell carcinoma is the most common primary malignancy of the eyelid and may be very invasive locally.
- Sebaceous carcinoma of the eyelid, by contrast, may metastasize and is therefore a serious and potentially life-threatening condition.

CONJUNCTIVA

Functional Anatomy

The conjunctiva is divided into zones (see Fig. 29.4), each with distinctive histologic features and responses to disease. The conjunctiva lining the interior of the eyelid, the *palpebral conjunctiva,* is tightly tethered to the tarsus and may respond to inflammation by being thrown into minute papillary folds as may occur in allergic conjunctivitis and bacterial conjunctivitis. The conjunctiva in the *fornix* is a pseudostratified columnar epithelium rich in goblet cells. The fornix also contains accessory lacrimal tissue, and the ductules of the

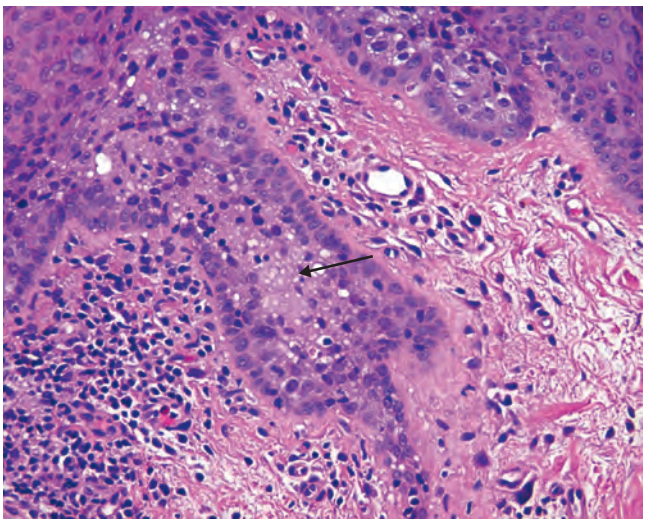


Figure 29.5 Pagetoid spread of sebaceous carcinoma. Neoplastic cells with foamy cytoplasm are present within the epidermis *(arrow)*. Invasive sebaceous carcinoma was identified elsewhere in this biopsy sample.

main lacrimal gland pierce through the conjunctiva in the fornix superiorly and laterally. The lymphoid population of the conjunctiva is most noticeable in the fornix, and *in viral conjunctivitis, lymphoid follicles may enlarge sufficiently to be visualized clinically* by slit-lamp examination. *Granulomas* associated with systemic sarcoidosis may be detected in the conjunctival fornix, and the yield of granulomas from a nondirected conjunctival biopsy in individuals suspected of having sarcoid may be as high as 50%. Primary lymphoma of the conjunctiva (typically indolent marginal zone B-cell lymphoma) is most likely to develop in the fornix. The *bulbar conjunctiva*—the conjunctiva that covers the surface of the eye—is a nonkeratinizing stratified squamous epithelium. The limbus, the intersection between the sclera and cornea, also marks the transition between conjunctival and corneal epithelium (see Fig. 29.1).

The conjunctiva, like the eyelid, is richly invested with lymphatic channels. Malignant neoplasms arising in the eyelid and conjunctiva tend to spread to regional lymph nodes (parotid and submandibular node groups).

Conjunctival Scarring

Many cases of bacterial or viral conjunctivitis cause redness and itching, but most heal without sequelae. However, infection with *Chlamydia trachomatis* (*trachoma*) may produce significant conjunctival scarring. Conjunctival scarring is also seen after exposure of the ocular surface to caustic alkalis or as a sequela to ocular cicatricial *pemphigoid* (Chapter 25). A reduction in the number of goblet cells due to conjunctival scarring leads to a decrease in surface mucin, which is essential for the adherence of the aqueous component of tears to the corneal epithelium. Thus, even if the aqueous component of the tear film is adequate, the affected individual will suffer from a dry eye. More commonly, however, dry eye results from a deficiency in the aqueous component of the tear film generated by the accessory lacrimal glands embedded within the eyelid and fornix.

The conjunctiva may be scarred iatrogenically through reaction to drugs or as a consequence of surgery. In other parts of the body, cancer surgery requires excision of the lesion with a margin of normal tissue to ensure complete removal. However, extensive surgical excision of even diseased conjunctiva can remove a large number of goblet cells or compromise lacrimal gland ductules that traverse the conjunctiva. Thus, removal of a conjunctival neoplasm or a precursor lesion may leave the affected individual with a painful dry eye that can compromise vision. Therefore, surgeons often remove only the invasive components of conjunctival neoplasms and treat the intraepithelial components with tissue-sparing modalities such as cryotherapy or topical chemotherapy delivered as eyedrops.

Pinguecula and Pterygium

Both *pinguecula* and *pterygium* appear as submucosal elevations on the conjunctiva. They result from actinic damage and are therefore located in the sun-exposed regions of the conjunctiva (i.e., in the fissure between both the upper and the lower eyelids—the interpalpebral fissure). Pterygium typically originates in the conjunctiva astride the limbus. It is formed by a submucosal growth *of fibrovascular connective tissue that migrates onto the cornea,* dissecting into the plane occupied normally by the Bowman layer. Pterygium does not cross the pupillary axis and, aside from the possible induction of mild astigmatism, does not pose a threat to vision. Although most pterygia are entirely benign, it is worthwhile submitting the excised tissue for pathologic examination because, on occasion, precursors of actinic-induced neoplasms—squamous cell carcinoma and melanoma—are detected in these lesions. Pinguecula, which, like pterygium, appears astride the limbus, is a small, yellowish submucosal elevation.

Neoplasms

Both squamous neoplasms and melanocytic neoplasms and their precursors tend to develop at the limbus. Conjunctival *squamous cell carcinoma* may be preceded by intraepithelial neoplastic changes analogous to those seen in the evolution of cervical squamous cell carcinoma. In the conjunctiva the spectrum of changes from mild dysplasia through carcinoma in situ is designated as *ocular surface squamous neoplasia*. Squamous papillomas and conjunctival intraepithelial neoplasia may be associated with the presence of human papillomavirus types 16 and 18.

Conjunctival nevi are encountered commonly but seldom invade the cornea or appear in the fornix or over the palpebral conjunctiva. Pigmented lesions in these zones of the conjunctiva most likely represent melanomas or melanoma precursors. Compound nevi of the conjunctiva characteristically contain subepithelial cysts lined by surface epithelium (Fig. 29.6A, B). In late childhood or adolescence, conjunctival nevi may acquire an inflammatory component rich in lymphocytes, plasma cells, and eosinophils. The resultant *inflamed juvenile nevus* is completely benign.

Conjunctival melanomas are unilateral neoplasms, typically affecting fair-complexioned individuals in middle age (Fig. 29.6C, D). Most cases of conjunctival melanoma develop through a phase of intraepithelial growth termed *primary acquired melanosis with atypia* or *conjunctival melanocytic intraepithelial neoplasia (C-MIN)*. Between 50% and 90% of individuals with incompletely treated primary acquired melanosis with atypia develop conjunctival melanoma. *BRAF* V600 mutations may be identified in nearly 40% of conjunctival melanomas. Perhaps the best treatment of conjunctival melanoma is its prevention through extirpation of its precursor lesion. The lesions tend to spread first to the parotid or submandibular lymph nodes. Approximately 25% of conjunctival melanomas prove to be fatal.

KEY CONCEPTS

- Conjunctival scarring, a consequence of a variety of conditions, may result in painful loss of vision by interfering with the delivery and maintenance of the tear film.
- Many conjunctival neoplasms originate at the limbus, the seat of stem cells of the ocular surface.
- Conjunctival malignancies—especially conjunctival melanomas—tend to spread through the rich lymphatics of the conjunctiva to regional lymph nodes.

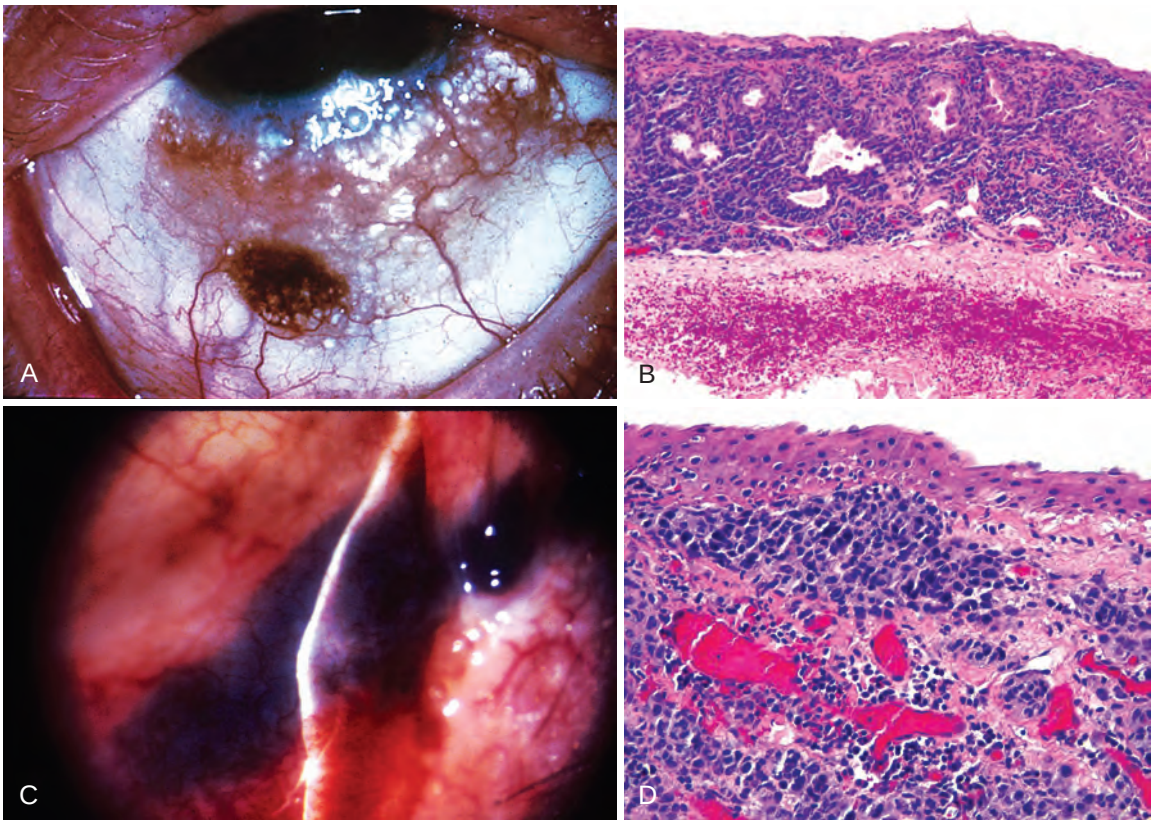


Figure 29.6 (A, B) Cystic compound nevus of the conjunctiva. (C, D) Conjunctival malignant melanoma. In C, note the deflection of the beam of the slit lamp over the surface of the lesion, indicative of invasion. (A, B, From Folberg R, et al: Benign conjunctival melanocytic lesions: clinicopathologic features, *Ophthalmology* 96:436, 1989.)

SCLERA

The sclera consists mainly of collagen and contains few blood vessels and fibroblasts; hence wounds and surgical incisions tend to heal poorly. Immune complex deposits within the sclera, such as in *rheumatoid arthritis,* may produce a necrotizing *scleritis*.

The sclera may appear "blue" in a variety of conditions. Some of these are:

- It may become thin following episodes of *scleritis,* and the normally brown color of the uvea may appear blue clinically because of the optical Tyndall effect.
- *Sclera may be thinned in eyes with exceptionally high intraocular pressure* and because this zone of scleral ectasia is lined by uveal tissue, the resulting lesion, known as a *staphyloma,* also appears blue.
- The sclera may appear blue in *osteogenesis imperfecta.*
- The sclera may appear blue because of a heavily pigmented congenital nevus of the underlying uvea, a condition known as *congenital melanosis oculi*. When accompanied by periocular cutaneous pigmentation, this condition is known as *nevus of Ota*.

CORNEA

Functional Anatomy

The cornea and its overlying tear film—not the lens—make up the major refractive surface of the eye (Fig. 29.7). Parenthetically, *myopia* typically develops because the eye is too long for its refractive power, and *hyperopia* results from an eye that is too short. The popularity of procedures such as laser-assisted in situ keratomileusis (LASIK) to sculpt the cornea and change its refractive properties attests to the importance of corneal shape in contributing to the refractive power of the eye.

Anteriorly, the cornea is covered by *epithelium* that rests on a basement membrane. The *Bowman layer,* situated just beneath the epithelial basement membrane, is acellular and forms an efficient barrier against the penetration of malignant cells from the epithelium into the underlying stroma.

The *corneal stroma* lacks blood vessels and lymphatics, a feature that contributes not only to the transparency of the cornea, but also to high rate of success of corneal transplantation. Indeed, nonimmunologic graft failure

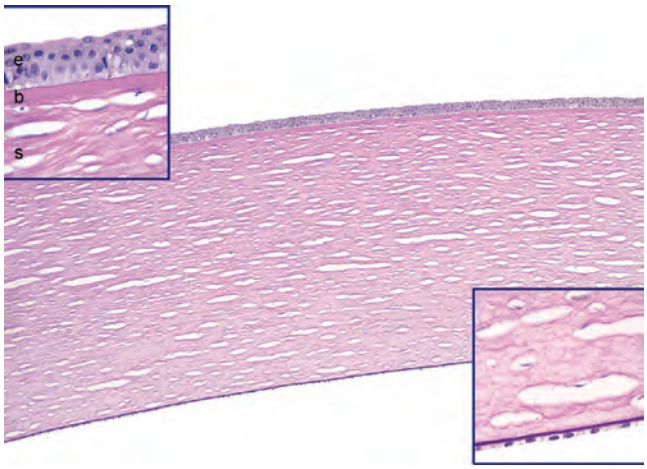


Figure 29.7 Normal corneal microarchitecture. The corneal tissue is stained by periodic acid–Schiff (PAS) to highlight basement membranes. The inset at the upper left is a high magnification of the anterior layers of the cornea: epithelium *(e)*, Bowman layer *(b)*, and stroma *(s)*. A very thin PAS-positive basement membrane separates the epithelium from the Bowman layer. Note that the Bowman layer is acellular. The inset at the lower right is a high magnification of the PAS-positive Descemet membrane and the corneal endothelium. The "holes" in the stroma are artifactitious spaces between parallel collagenous stromal lamellae.

(associated with loss of endothelial cells and subsequent corneal edema) is seen more commonly than is immunologic graft rejection. The risk of corneal graft rejection increases with stromal vascularization and inflammation. A precise alignment of collagen in the corneal stroma also contributes to transparency.

Corneal vascularization may accompany chronic corneal edema, inflammation, and scarring. The application of topical VEGF antagonists affords a promising approach to preventing corneal vascularization. Scarring and edema both disrupt the spatial alignment of stromal collagen and contribute to corneal opacification. Scars may result from trauma or inflammation. Normally, the corneal stroma is in a state of relative deturgescence (dehydration), maintained in large part by active pumping of fluid from the stroma back into the anterior chamber by the corneal endothelium.

The corneal *endothelium* is derived from neural crest and is not related to vascular endothelium. It rests on its basement membrane, Descemet membrane. A decrease in endothelial cells or a malfunction of endothelium results in stromal edema, which may be complicated by bullous separation of the epithelium *(bullous keratopathy)*. *Descemet membrane* increases in thickness with age. It is the site of copper deposition in the Kayser-Fleischer ring of Wilson disease (Chapter 18).

Keratitis and Ulcers

Various pathogens—bacterial, fungal, viral (especially herpes simplex and herpes zoster), and protozoal (*Acanthamoeba*)—can cause corneal ulceration. In all forms of keratitis, dissolution of the corneal stroma may be accelerated by activation of collagenases within corneal epithelium and stromal fibroblasts (also known as keratocytes). Exudate

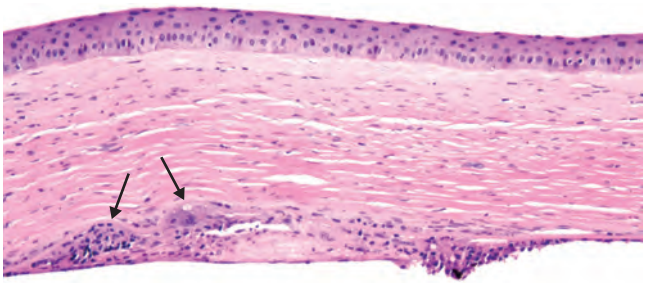


Figure 29.8 Chronic herpes simplex keratitis. The cornea is thin and scarred (note the increased number of fibroblast nuclei). Granulomatous reaction in the Descemet membrane, illustrated in this photomicrograph *(arrows)*, is a histologic hallmark of chronic herpes simplex keratitis.

and cells leaking from iris and ciliary body vessels into the anterior chamber may be visible by slit-lamp examination and may accumulate in sufficient quantity to become visible even by a penlight examination *(hypopyon)*. Although the corneal ulcer may be infectious, the hypopyon seldom contains organisms and is an example par excellence of the vascular response to acute inflammation. The specific forms of keratitis may have certain distinctive features. For example, chronic herpes simplex keratitis may be associated with a granulomatous reaction involving the Descemet membrane (Fig. 29.8).

Corneal Degenerations and Dystrophies

Ophthalmologists have traditionally divided many corneal disorders into degenerations and dystrophies. Corneal degenerations may be either unilateral or bilateral and are typically nonfamilial. By contrast, corneal dystrophies are typically bilateral and are hereditary. Corneal dystrophies may affect selective corneal layers or the changes may be distributed throughout multiple layers.

Band Keratopathies

Two types of band keratopathy serve as examples of corneal degenerations. *Calcific band keratopathy* is characterized by deposition of calcium in the Bowman layer. This condition may complicate chronic uveitis, especially in individuals with chronic juvenile rheumatoid arthritis. *Actinic band keratopathy* develops in individuals who are exposed chronically to high levels of ultraviolet light. In this condition, extensive solar elastosis develops in the superficial layers of corneal collagen in the sun-exposed interpalpebral fissure—hence the horizontally distributed band of pathology. Similar to pinguecula, the sun-damaged collagen of the cornea may take on a yellow hue to the point that this condition is sometimes erroneously called "oil-droplet keratopathy."

Keratoconus

With an incidence of 1 in 2000, *keratoconus* is a fairly common disorder characterized by progressive thinning and ectasia of the cornea without evidence of inflammation or vascularization. Such thinning results in a cornea that has a conical rather than spherical shape. This abnormal shape generates irregular astigmatism that is difficult to correct with spectacles. Rigid contact lenses generate a smooth, spherical

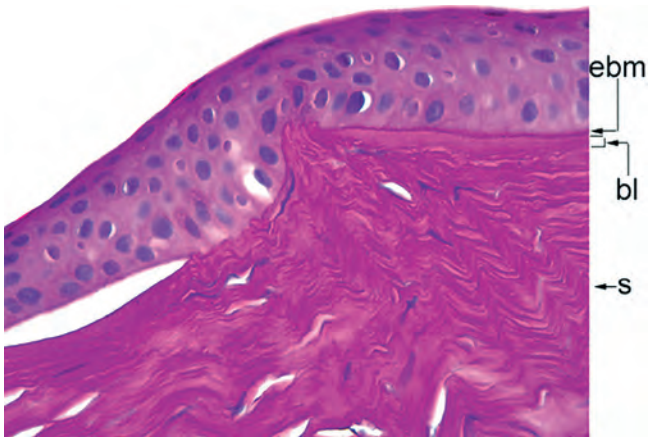


Figure 29.9 Keratoconus. The tissue section is stained by periodic acid–Schiff to highlight the epithelial basement membrane *(ebm)*, which is intact; the Bowman layer *(bl)*, situated between the epithelial basement membrane; and the stroma *(s)*. Following the Bowman layer from the *right side* of the photomicrograph toward the *center*, there is a discontinuity, diagnostic of keratoconus. The epithelial separation just to the *left* of the Bowman layer discontinuity resulted from an episode of corneal hydrops, caused by a break in the Descemet membrane (not shown).

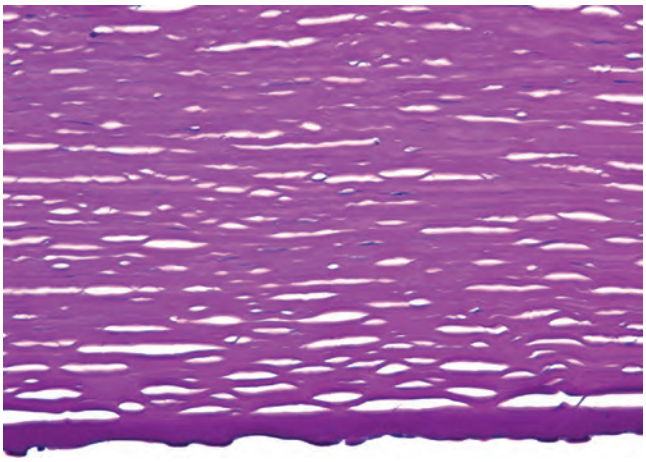


Figure 29.10 Fuchs dystrophy. This tissue section is stained by periodic acid–Schiff to highlight the Descemet membrane, which is thick. Numerous drop-like excrescences—guttata—protrude downward from the Descemet membrane. Endothelial cell nuclei are not seen. Epithelial bullae, not shown in this micrograph, were present, reflecting corneal edema.

surface to the cornea and may provide refractive relief for individuals with keratoconus. Patients whose vision cannot be corrected with spectacles or contact lenses are excellent candidates for corneal transplantation, which has a high degree of success in this condition. Unlike many types of degeneration, keratoconus is typically bilateral. Keratoconus is associated with Down syndrome, Marfan syndrome, and atopic disorders. Its development may stem from a genetic predisposition superimposed by an environmental insult, such as eye rubbing in response to atopic conditions.

MORPHOLOGY

Thinning of the cornea with breaks in the Bowman layer are the histologic hallmarks of keratoconus (Fig. 29.9). In some patients the Descemet membrane may rupture precipitously, allowing the aqueous humor in the anterior chamber to gain access to the corneal stroma. The sudden effusion of aqueous humor through a gap in the Descemet membrane—corneal **hydrops**—may also cause vision to worsen suddenly. An episode of hydrops may be followed by corneal scarring that can also contribute to visual loss. Acute corneal hydrops can complicate Descemet membrane ruptures that develop secondary to extraordinary elevations of intraocular pressure in *infantile glaucoma (Haab striae)* or following the now uncommon obstetric forceps injury to the eye.

Fuchs Endothelial Dystrophy

Fuchs endothelial dystrophy, one of several dystrophies, results from loss of endothelial cells and the resulting edema and thickening of the stroma. It is one of the principal indications for corneal transplantation in the United States. The two major clinical manifestations of Fuchs endothelial dystrophy—*stromal edema and bullous keratopathy*—are both related to a primary loss of endothelial cells. Early in the course of the disease, endothelial cells produce drop-like deposits of abnormal basement membrane material *(guttata)* that can be visualized clinically by slit-lamp examination. With disease progression, there is a decrease in the total number of endothelial cells, and the residual cells are incapable of maintaining stromal deturgescence. Consequently the stroma becomes edematous and thickens; it acquires a ground-glass appearance clinically, and vision is blurred (Fig. 29.10). Because of chronic edema, the stroma may eventually become vascularized. On occasion the number of endothelial cells may decrease following cataract surgery even in individuals who do not have early forms of Fuchs dystrophy. This condition, known as *pseudophakic bullous keratopathy,* is also a common indication for corneal transplantation.

KEY CONCEPTS

- The cornea—not the lens—is the major refractive surface of the eye. Keratoconus is an example of a condition that distorts the contour of the cornea and alters this refractive surface, producing an irregular form of astigmatism.
- The normal cornea is avascular, a feature that contributes to transparency and the low incidence of graft rejection after corneal transplantation.
- Inflammations of the cornea may be accompanied by a noninfectious exudative process in the anterior chamber that may organize to distort anterior segment anatomy and contribute to secondary glaucoma and to cataract.
- Corneal dystrophies are generally inherited and degenerations are typically not inherited. Fuchs dystrophy and pseudophakic bullous keratopathy both produce visual loss through the final common pathway of corneal edema, and both of these conditions are leading indications for corneal transplantation in the United States.

ANTERIOR SEGMENT

Functional Anatomy

The anterior chamber is bounded anteriorly by the cornea, laterally by the trabecular meshwork, and posteriorly by the iris (Fig. 29.11). Aqueous humor, formed by the pars plicata of the ciliary body, enters the posterior chamber, bathes the lens, and circulates through the pupil to gain access to the anterior chamber. The posterior chamber lies behind the iris and in front of the lens.

The lens is a closed epithelial system; the basement membrane of the lens epithelium (known as the lens capsule) totally envelops the lens. Thus, the lens epithelium does not exfoliate like the epidermis or a mucosal epithelium. Instead, the lens epithelium and its derivative fibers accumulate within the confines of the lens capsule, thus "infoliating." With aging, therefore, the size of the lens increases. Neoplasms of the lens have not been described.

Cataract

The term *cataract* describes lenticular opacities that may be congenital or acquired. Systemic diseases (e.g., galactosemia, diabetes mellitus, Wilson disease, and atopic dermatitis), drugs (especially corticosteroids), radiation, trauma, and many intraocular disorders are associated with cataract. Age-related cataract typically results from opacification of the lens nucleus *(nuclear sclerosis)*. The accumulation of urochrome pigment may render the lens nucleus brown, thus distorting the individual's perception of blue color (the predominance of yellow hues in Rembrandt's paintings later in life might have been a consequence of nuclear sclerotic cataracts). Other physical changes in the lens may generate opacities. For example, the lens cortex may liquefy. Migration

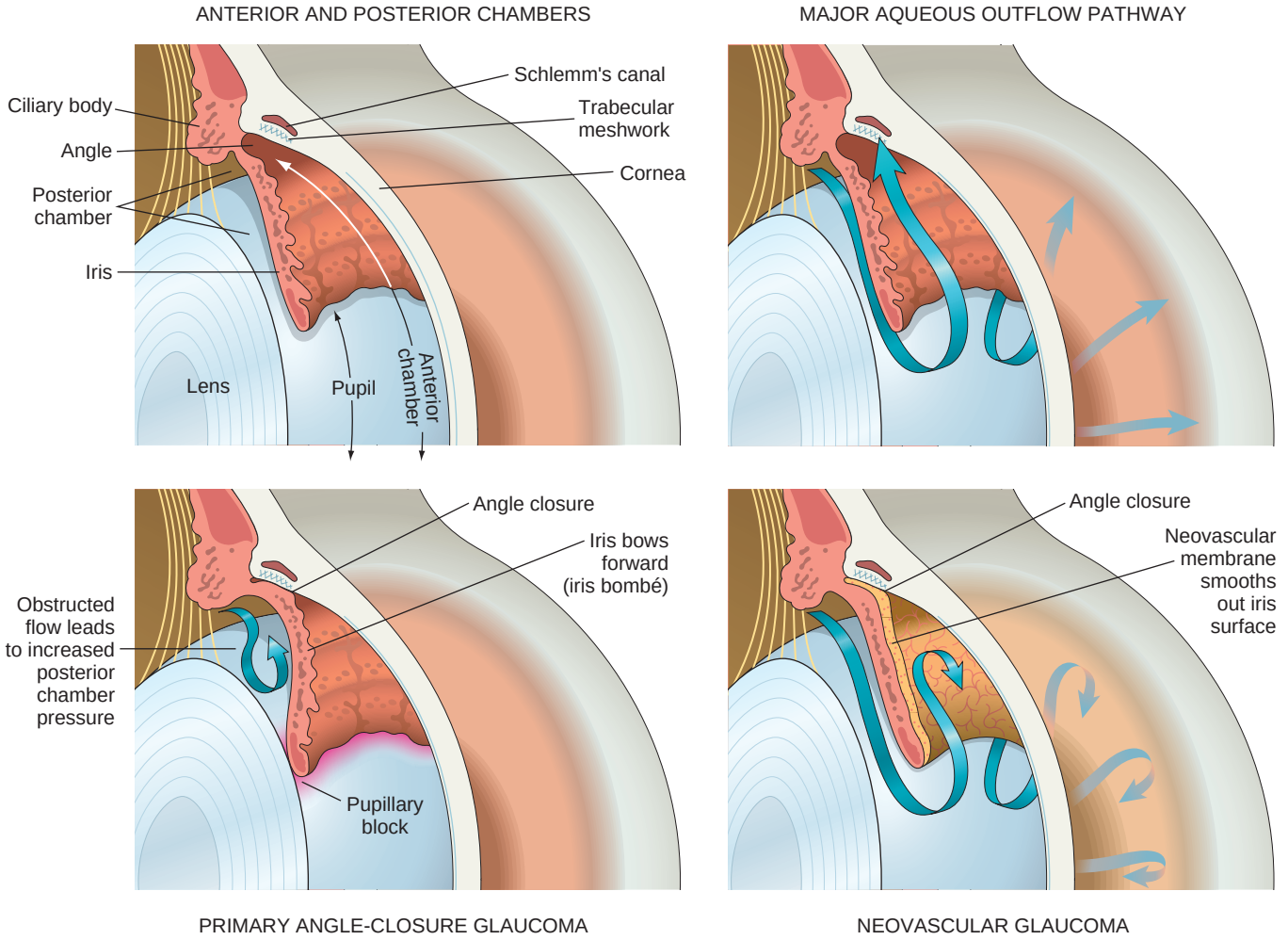


Figure 29.11 *Upper left,* The normal eye. Note that the surface of the iris is highly textured with crypts and folds. *Upper right,* The normal flow of aqueous humor. Aqueous humor, produced in the posterior chamber, flows through the pupil into the anterior chamber. The major pathway for the egress of aqueous humor is through the trabecular meshwork, into the Schlemm canal. Minor outflow pathways (uveoscleral and iris, not depicted) contribute to a limited extent to aqueous outflow. *Lower left,* Primary angle-closure glaucoma. In anatomically predisposed eyes, transient apposition of the iris at the pupillary margin to the lens blocks the passage of aqueous humor from the posterior chamber to the anterior chamber. Pressure builds in the posterior chamber, bowing the iris forward (iris bombé) and occluding the trabecular meshwork. *Lower right,* A neovascular membrane has grown over the surface of the iris, smoothing the iris folds and crypts. Myofibroblasts within the neovascular membrane cause the membrane to contract and to become apposed to the trabecular meshwork (peripheral anterior synechiae). Outflow of aqueous humor is blocked, and the intraocular pressure becomes elevated.

of the lens epithelium posterior to the lens equator may result in *posterior subcapsular cataract* secondary to enlargement of abnormally positioned lens epithelium. The technique that is most commonly used to remove opacified lenses extracts the lens contents, leaving the lens capsule intact. A prosthetic intraocular lens is typically inserted into the eye.

Anterior Segment and Glaucoma

The term *glaucoma* refers to a collection of diseases characterized by distinctive changes in the visual field and in the cup of the optic nerve. Most of the glaucomas are associated with elevated intraocular pressure, although some individuals with normal intraocular pressure may develop characteristic optic nerve and visual field changes (*normal* or *low-tension glaucoma*). The relationship between intraocular pressure and optic nerve damage is discussed later under Optic Nerve.

To understand the pathophysiology of glaucoma it is useful to consider the formation and drainage of aqueous humor. As Fig. 29.11 illustrates, aqueous humor is produced in the ciliary body and passes from the posterior chamber through the pupil into the anterior chamber. Although there are multiple pathways for the egress of fluid from the anterior chamber, most of the aqueous humor drains through the trabecular meshwork, situated in the angle formed by the intersection between the corneal periphery and the anterior surface of the iris. With this background, glaucoma can be classified into two major categories.

- In *open-angle glaucoma* the aqueous humor has complete physical access to the trabecular meshwork, and the elevation in intraocular pressure results from an increased resistance to aqueous outflow in the open angle.
- In *angle-closure glaucoma* the peripheral zone of the iris adheres to the trabecular meshwork and physically impedes the egress of aqueous humor from the eye.

Both open-angle and angle-closure glaucoma can be subclassified into primary and secondary types. In *primary open-angle glaucoma,* the most common form of glaucoma, the angle is open, and few changes are apparent structurally.

There are multiple causes of *secondary open-angle glaucoma*. Pseudoexfoliation glaucoma, perhaps the most common form of secondary open-angle glaucoma, is associated with the deposition of fibrillar material of varying composition throughout the anterior segment. In addition to deposition in the anterior chamber, fibrillar material is deposited around blood vessels in connective tissue and in many visceral organs such as liver, kidney, and gallbladder.

Particulate material such as high-molecular-weight lens proteins produced by phacolysis, senescent red cells after trauma *(ghost cell glaucoma),* iris epithelial pigment granules *(pigmentary glaucoma),* and necrotic tumors *(melanomalytic glaucoma)* can clog the trabecular meshwork in the presence of an open angle. Elevations in the pressure on the surface of the eye (episcleral venous pressure) in the presence of an open angle also contribute to secondary open-angle glaucoma. This type of glaucoma is associated with surface ocular vascular malformations seen in *Sturge-Weber syndrome* or is a consequence of arterialization of the episcleral veins following a spontaneous or traumatic carotid-cavernous fistula.

Primary angle-closure glaucoma typically develops in eyes with shallow anterior chambers, often found in individuals with hyperopia. Transient apposition of the pupillary margin of the iris to the anterior surface of the lens may result in obstruction to the flow of aqueous humor through the pupillary aperture *(pupillary block)*. Continued production of aqueous humor by the ciliary body thus elevates pressure in the posterior chamber and may bow the iris periphery forward *(iris bombé),* apposing it to the trabecular meshwork. These anatomic changes provoke a marked elevation in intraocular pressure (see Fig. 29.11). Since the crystalline lens is avascular and the lens epithelium receives its nutrition from the aqueous humor, unremitting elevation in intraocular pressure in primary angle-closure glaucoma can damage the lens epithelium. This leads to minute anterior subcapsular opacities that are visible by slit-lamp examination. Although the affected individual might have a normal complement of healthy corneal endothelial cells, sustained elevated intraocular pressure can produce corneal edema and bullous keratopathy.

There are many causes of *secondary angle-closure glaucoma*. Contraction of various types of pathologic membranes that form over the surface of the iris can draw the iris over the trabecular meshwork, occluding aqueous outflow. For example, chronic retinal ischemia is associated with the upregulation of VEGF and other proangiogenic factors. The appearance of VEGF in the aqueous humor is thought to induce the development of thin, clinically transparent fibrovascular membranes over the surface of the iris. Contraction of myofibroblastic elements in these membranes leads to occlusion of the trabecular meshwork by the iris: *neovascular glaucoma* (see Fig. 29.11). Necrotic tumors, especially retinoblastomas, can also induce iris neovascularization and glaucoma. Secondary angle-closure glaucoma may be caused by other mechanisms as well; for example, tumors in the ciliary body can mechanically compress the iris onto the trabecular meshwork, closing off the major pathway of aqueous outflow.

Endophthalmitis and Panophthalmitis

In intraocular inflammation, vessels in the ciliary body and iris become leaky, allowing cells and exudate to accumulate in the anterior chamber. These changes can be visualized with a slit lamp; at times the inflammatory cells may adhere to the corneal endothelium, forming clinically visible *keratic precipitates*. The size and shape of these precipitates can provide clues to the underlying cause of the inflammation. For example, aggregates of macrophages on the endothelium in sarcoid produce characteristic "mutton-fat" keratic precipitates.

Just as pleural exudate in acute bronchopneumonia can lead to adhesions between the visceral and parietal pleura, the presence of exudate in the anterior chamber can lead to the formation of adhesions between the iris and the trabecular meshwork or cornea *(anterior synechiae)* or between the iris and the anterior surface of the lens *(posterior synechiae)*. Anterior synechiae can lead to elevation in intraocular pressure, which may lead to optic nerve damage. Prolonged contact between the iris and the anterior surface of the lens can deprive lens epithelium of contact with aqueous humor and can induce fibrous metaplasia of the lens epithelium:

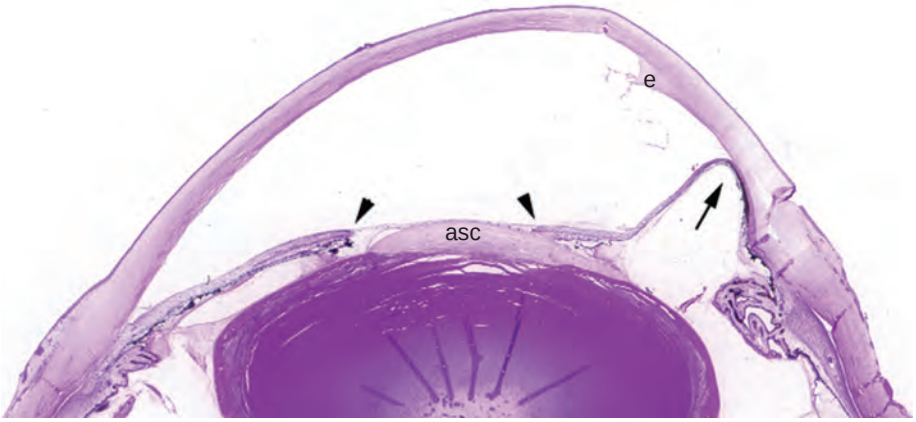


Figure 29.12 Sequelae of anterior segment inflammation. This eye was removed for complications of chronic corneal inflammation (not visible at this magnification). The exudate *(e)* present in the anterior chamber would have been visualized with a slit lamp as an optical "flare." The iris is adherent focally to the cornea, obstructing the trabecular meshwork (anterior synechia, *arrow*), and to the lens (posterior synechiae, *arrowheads*). An anterior subcapsular cataract *(asc)* has formed. The radial folds in the lens are artifacts.

anterior subcapsular cataract (Fig. 29.12). The pharmacologic induction of pupillary dilation and cycloplegia in individuals with intraocular inflammation is intended in part to prevent the formation of synechiae and their sequelae.

Although inflammation confined to the anterior segment is technically intraocular inflammation, the term *endophthalmitis* is reserved for inflammation within the vitreous humor. The retina lines the vitreous cavity, and suppurative inflammation in the vitreous humor is poorly tolerated by the retina; a few hours may be sufficient to cause irreversible retinal injury. Endophthalmitis is classified as *exogenous* (originating in the environment and gaining access to the interior of the eye through a wound) or *endogenous* (delivered to the eye hematogenously). The term *panophthalmitis* is applied to inflammation within the eye that involves the retina, choroid, and sclera and extends into the orbit (Fig. 29.13).

KEY CONCEPTS

- The term cataract describes opacities of the lens that may be congenital or acquired.
- The term glaucoma describes a group of conditions characterized by distinctive changes in the visual field and the size and shape of the optic nerve cup and usually by an elevation in intraocular pressure.
- Glaucoma may develop in the context of either an open or closed angle. Open-angle and angle-closure glaucomas are further subclassified into primary and secondary types.
- Endophthalmitis is a term used to describe inflammation of the interior of the eye involving the vitreous humor, and panophthalmitis is the term used to describe inflammation of the interior of the eye that also extends into the uvea and sclera.
- Endophthalmitis may originate from infection within the body (endogenous endophthalmitis complicating generalized sepsis) or as a complication of corneal infection or a wound, an accidental injury, or a surgical procedure (exogenous endophthalmitis).
- In panophthalmitis, inflammation extends from the interior of the eye into the ocular coats: the retina, choroid, and the sclera.

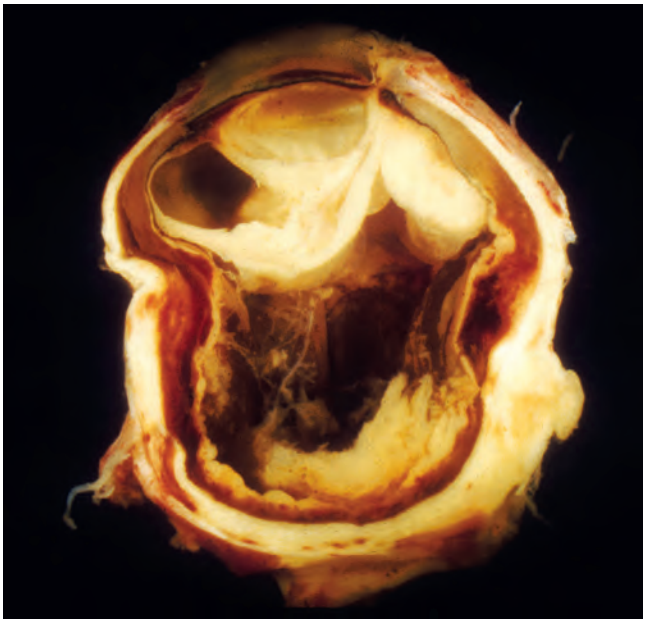


Figure 29.13 Exogenous panophthalmitis. This eye was removed after a foreign body injury. Note the suppurative inflammation behind the lens that is drawn up to the right of the lens to the cornea, the site of the wound. The central portion of the vitreous humor was extracted surgically (by vitrectomy). Note the adhesions to the surface of the eye at the 8 o'clock position, indicating that the intraocular inflammation has spread through the sclera into the orbit: panophthalmitis. (From Folberg R: The eye. In Spencer WH, editor: *Ophthalmic Pathology—An Atlas and Textbook,* ed 4, Philadelphia, 1985, WB Saunders.)

UVEA

Together with the iris, the choroid and ciliary body constitute the uvea. The choroid is among the most richly vascularized sites in the body.

Uveitis

The term *uveitis* can be applied to any type of inflammation in one or more of the tissues that compose the uvea. Thus, the iritis that develops after blunt trauma to the eye or that accompanies a corneal ulcer is technically a form of uveitis. However, in clinical practice the term *uveitis* is restricted to a diverse group of chronic diseases that may be either components of a systemic process or localized to the eye. Uveal inflammation may be manifest principally in the anterior segment (e.g., in *juvenile rheumatoid arthritis*) or may affect both the anterior and the posterior segments. The complications of chronic anterior segment inflammation were discussed earlier; the remainder of this discussion therefore focuses on the effects of uveal inflammation on the posterior segment of the eye. As will be described briefly, uveitis is frequently accompanied by retinal pathology. Uveitis may be caused by infectious agents (e.g., *Pneumocystis carinii*), may be idiopathic (e.g., sarcoidosis), or may be autoimmune in origin (sympathetic ophthalmia). Examples are described later.

Granulomatous uveitis is a common complication of sarcoidosis (Chapter 15). In the anterior segment it gives rise to an exudate that evolves into "mutton-fat" keratic precipitates described earlier. In the posterior segment, sarcoid may involve the choroid and retina. Thus, granulomas may be seen in the choroid. Retinal pathology is characterized by perivascular inflammation; this is responsible for the well-known ophthalmoscopic sign of "candle wax drippings." Conjunctival biopsy can be used to detect granulomatous inflammation and confirm the diagnosis of ocular sarcoid.

Numerous infectious processes can affect the choroid or the retina. Inflammation in one compartment is typically associated with inflammation in the other. Retinal *toxoplasmosis* is usually accompanied by uveitis and even scleritis. Individuals with AIDS may develop cytomegalovirus retinitis and uveal infection such as *Pneumocystis* or mycobacterial choroiditis.

Sympathetic ophthalmia is an example of noninfectious uveitis limited to the eye. This condition is characterized by bilateral granulomatous inflammation typically affecting all components of the uvea: a panuveitis. Sympathetic ophthalmia, which blinded young Louis Braille, may complicate a penetrating injury of the eye. In the injured eye, retinal antigens sequestered from the immune system may gain access to lymphatics in the conjunctiva and thus set up a delayed hypersensitivity reaction that affects not only the injured eye but also the contralateral, noninjured eye. The condition may develop from 2 weeks to many years after injury. Enucleation of a blind eye (which can be the sympathizing eye rather than the directly injured eye) may yield diagnostic findings. It is characterized by diffuse granulomatous inflammation of the uvea (choroid, ciliary body, and iris). Plasma cells are typically absent, but eosinophils may be identified in the infiltrate (Fig. 29.14). Sympathetic ophthalmia is treated by the administration of systemic immunosuppressive agents.

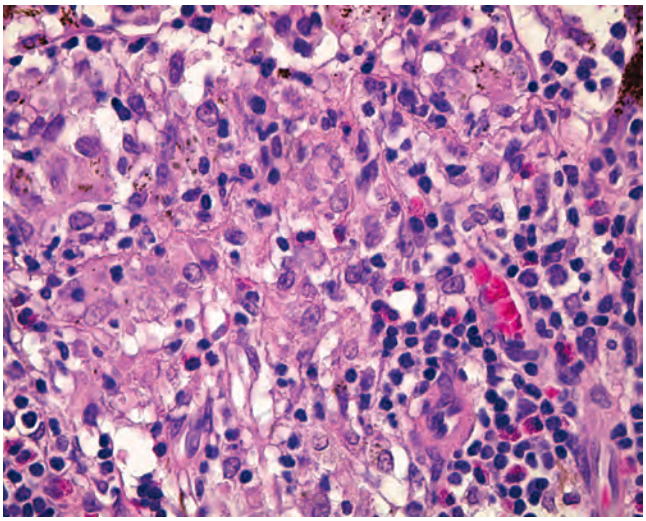


Figure 29.14 Sympathetic ophthalmia. The granulomatous inflammation depicted here was identified diffusely throughout the uvea. The uveal granulomas may contain melanin pigment and may be accompanied by eosinophils.

Neoplasms

The most common intraocular malignancy of adults is metastasis to the uvea, typically to the choroid. The occurrence of metastases to the eye is associated with an extremely short survival, and treatment of ocular metastases, usually by radiotherapy, is palliative.

Uveal Nevi and Melanomas

Uveal melanoma is the most common primary intraocular malignancy of adults. In the United States, these tumors account for approximately 5% of melanomas and have an age-adjusted incidence of 5.1 per 1 million per year. Uveal nevi, especially choroidal nevi, are rather common, affecting an estimated 5% of the Caucasian population.

Epidemiology and Pathogenesis

Unlike cutaneous melanoma, the occurrence of uveal melanoma has remained stable over many years, and there is no clear link between exposure to ultraviolet light and risk. In line with this observation, sequencing of tumor genomes has revealed that the molecular pathogenesis of uveal melanoma is distinct from that of cutaneous melanoma. Unlike cutaneous and conjunctival melanomas, *BRAF* mutations do not appear to play a role in the pathogenesis of uveal melanomas. The most important oncogenes in uveal melanoma are *GNAQ* and *GNA11*, both of which encode G protein–coupled receptors. Roughly 85% of uveal melanomas harbor a gain-of-function mutation in one of these genes that activates pathways that promote proliferation, such as the mitogen-activated protein kinase (MAPK) pathway (Chapter 7). Notably, uveal nevi are also associated with *GNAQ* and *GNA11* mutations, yet rarely transform to melanoma, indicating that other genetic events are required for the development of uveal melanoma. One such event is loss of chromosome 3, which appears to be selected for because it leads to deletion of *BAP1*, a tumor suppressor gene on chromosome 3 that encodes a deubiquitinating enzyme. BAP1 is a component of protein complexes that place repressive marks on chromatin that lead to gene silencing; thus, uveal melanoma has joined the increasing list of cancers in which epigenetic alterations appear to have a central role in tumor pathogenesis (Chapter

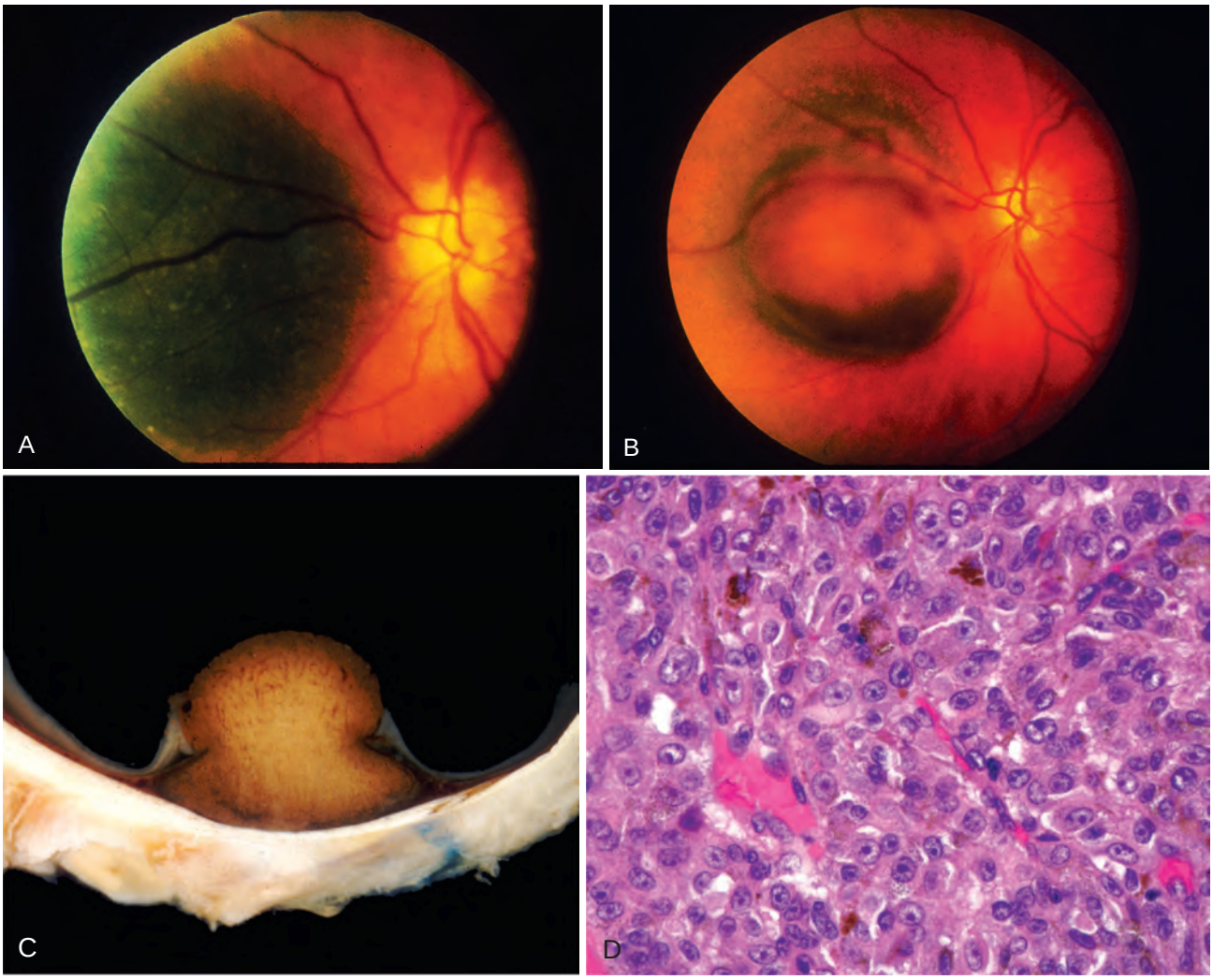


Figure 29.15 Uveal melanoma. (A) Fundus photograph from an individual with a relatively flat pigmented lesion of the choroid near the optic disc. (B) Fundus photograph of the same individual several years later; the tumor has grown and has ruptured through the Bruch membrane. (C) Gross photograph of a choroidal melanoma that has ruptured the Bruch membrane. The overlying retina is detached. (D) Epithelioid melanoma cells associated with an adverse outcome. (A–C, From Folberg R: *Pathology of the Eye—An Interactive CD-ROM Program,* Philadelphia, 1996, Mosby.)

7). Germline mutations in *BAP1* may predispose patients to the development of uveal melanomas, renal cell carcinomas, and malignant mesothelioma among several others.

MORPHOLOGY

Histologically, uveal melanomas may contain two types of cells, spindle and epithelioid, in various proportions (Fig. 29.15). **Spindle cells** are fusiform in shape, whereas **epithelioid cells** are spherical and have greater cytologic atypicality. Like cutaneous melanomas, large numbers of tumor-infiltrating lymphocytes may be seen in some cases. An unusual feature that is commonly seen is the presence of looping slit-like spaces lined by laminin that surround packets of tumor cells. These spaces (which are not blood vessels) connect to blood vessels and serve as extravascular conduits for the transport of plasma and possibly blood. In vitro studies and examination of human tissues suggest that these unusual growth patterns are promoted by tumor cells through a process termed **vasculogenic mimicry.**

Uveal melanomas, with very rare exception, spread exclusively by a hematogenous route (the only exception being the rare case of melanoma that spreads through the sclera and invades the conjunctiva, thereby gaining access to conjunctival lymphatics). Most uveal melanomas spread first to the liver, an excellent example of a tumor-specific tropism for a particular metastatic site.

Clinical Features

Most uveal melanomas are incidental findings or present with visual symptoms, which may be related to retinal detachment or glaucoma. The prognosis of choroidal and ciliary body melanomas is related to (1) size (in contrast to cutaneous melanoma, the lateral extent of the tumor rather than tumor depth is the size dimension related to adverse outcome); (2) cell type (tumors containing epithelioid cells have a worse prognosis than do those containing exclusively spindle cells); (3) and proliferative index. Cytogenetic profiles, especially monosomy 3, and gene expression profiling may

be helpful in stratifying patients into categories with differing risks of developing metastatic disease.

There seems to be no difference in survival between tumors treated by removal of the eye (enucleation) and those receiving eye-sparing radiotherapy, which is the treatment of choice. Melanomas situated exclusively in the iris tend to follow a relatively indolent course, whereas melanomas of the ciliary body and choroid are more aggressive.

Although the 5-year survival rate is approximately 80%, the cumulative melanoma mortality rate is 40% at 10 years, increasing 1% per year thereafter. Metastases may appear "out of the blue" many years after treatment, making uveal melanoma a prime candidate for the investigation of the phenomenon of tumor dormancy. Targeted therapies such as MAPK inhibitors have shown some encouraging responses in clinical trials, but currently there is no effective treatment for metastatic uveal melanoma.

KEY CONCEPTS

- Uveitis is restricted to a diverse group of chronic diseases that may be either components of a systemic process or localized to the eye.
- Sarcoid is an example of a systemic condition that may produce granulomatous uveitis, and sympathetic ophthalmia may produce bilateral granulomatous inflammation as a possible consequence of penetrating injury to one eye.
- The most common intraocular tumor of adults is metastasis to the eye.
- The most common primary intraocular tumor of adults is uveal melanoma.
- Uveal melanoma disseminates hematogenously, and the first evidence of metastasis is typically detected in the liver.
- Uveal melanoma shows marked differences in epidemiologic risk factors and driver mutations compared to cutaneous melanoma.

RETINA AND VITREOUS

Functional Anatomy

The neurosensory retina, like the optic nerve, is an embryologic derivative of the diencephalon. The retina therefore responds to injury by means of gliosis. As in the brain, there are no lymphatics. The architecture of the retina accounts for the ophthalmoscopic appearance of a variety of ocular disorders. Hemorrhages in the nerve fiber layer of the retina are oriented horizontally and appear as streaks or "flames"; the external retinal layers are oriented perpendicular to the retinal surface, and hemorrhages in these outer layers appear as dots (the tips of cylinders). Exudates tend to accumulate in the outer plexiform layer of the retina, especially in the macula (Fig. 29.16).

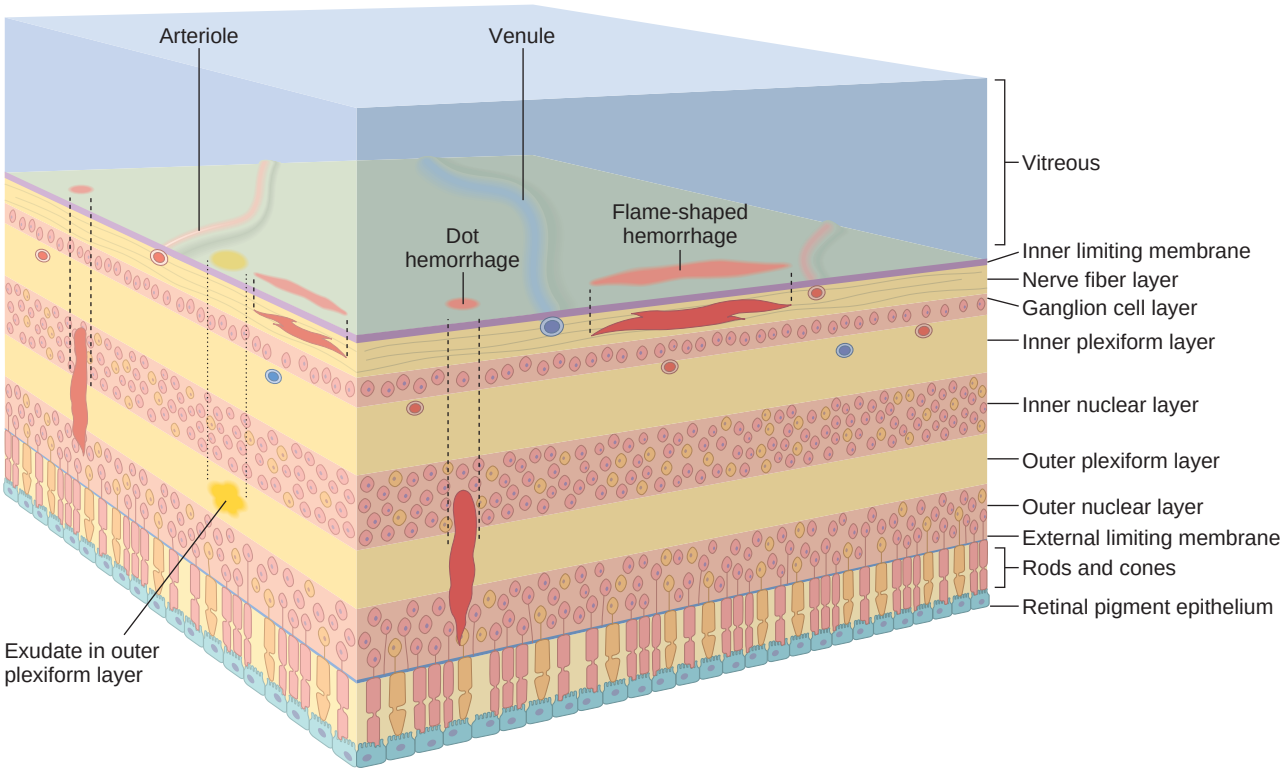


Figure 29.16 Clinicopathologic correlations of retinal hemorrhages and exudates. The location of the hemorrhage within the retina determines its appearance by ophthalmoscopy. The retinal nerve fiber layer is oriented parallel to the internal limiting membrane, and hemorrhages of this layer appear to be flame-shaped ophthalmoscopically. The deeper retinal layers are oriented perpendicular to the internal limiting membrane, and hemorrhages in this location appear as cross-sections of a cylinder, or "dot" hemorrhages. Exudates that originate from leaky retinal vessels accumulate in the outer plexiform layer.

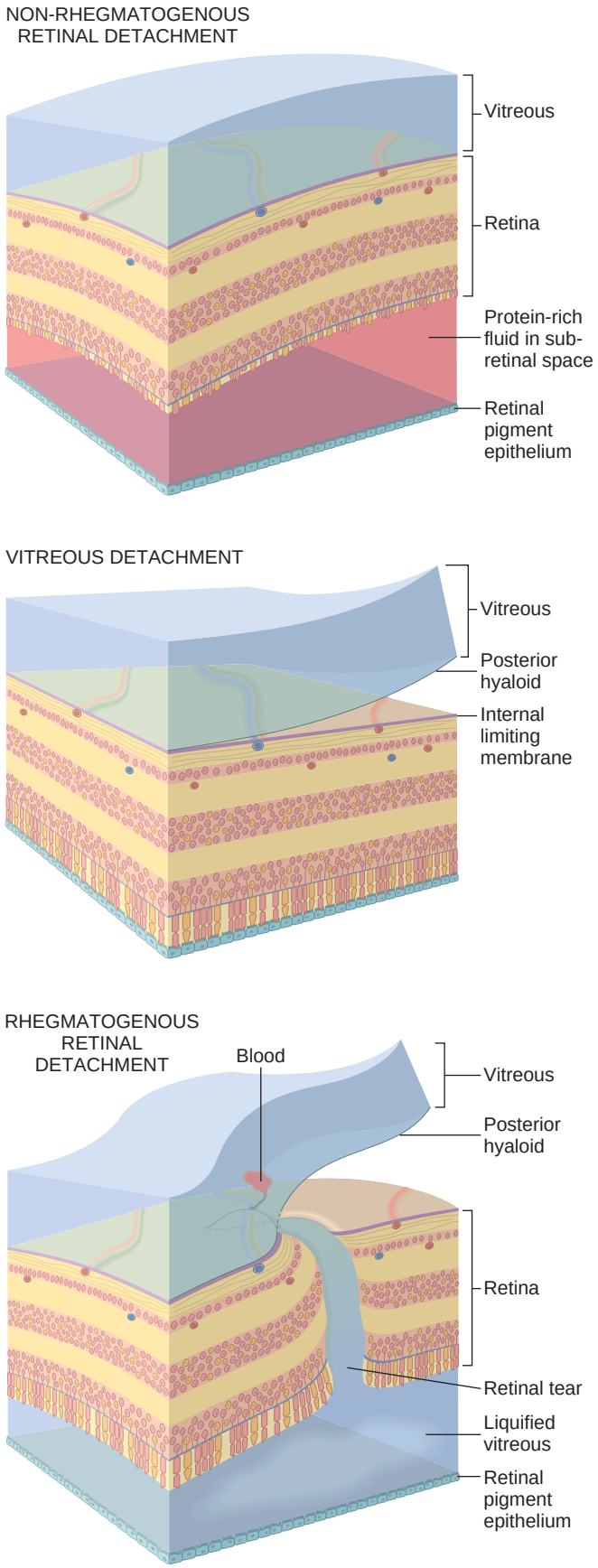


Figure 29.17 Retinal detachment is defined as the separation of the neurosensory retina from the retinal pigment epithelium (RPE). Retinal detachments are classified broadly into nonrhegmatogenous (without a retinal break) and rhegmatogenous (with a retinal break) types. *Top,* In nonrhegmatogenous retinal detachment the subretinal space is filled with protein-rich exudate. Note that the outer segments of the photoreceptors are missing (see Fig. 29.16 for orientation of layers). This indicates a chronic retinal detachment, a finding that can be seen in both nonrhegmatogenous and rhegmatogenous detachments. *Middle,* Posterior vitreous detachment involves the separation of the posterior hyaloid from the internal limiting membrane of the retina and is a normal occurrence in the aging eye. *Bottom,* If during a posterior vitreous detachment the posterior hyaloid does not separate cleanly from the internal limiting membrane of the retina, the vitreous humor will exert traction on the retina, which will be torn at this point. Liquefied vitreous humor seeps through the retinal defect, and the retina is separated from the RPE. The photoreceptor outer segments are intact, illustrating an acute detachment.

The retinal pigment epithelium (RPE), like the retina, is derived embryologically from the primary optic vesicle, an outpouching of the brain. Separation of the neurosensory retina from the RPE defines a *retinal detachment*. The RPE has an important role in the maintenance of the outer segments of the photoreceptors. Disturbances in the RPE–photoreceptor interface are implicated in hereditary retinal degenerations such as *retinitis pigmentosa*.

The adult vitreous humor is avascular. Incomplete regression of fetal vasculature running through the vitreous humor can produce significant pathology as a retrolental mass *(persistent fetal vasculature)*. The vitreous humor can be opacified by hemorrhage from trauma or retinal neovascularization. With age the vitreous humor may liquefy and collapse, creating the visual sensation of "floaters." Also, with aging, the posterior face of the vitreous humor—the posterior hyaloid—may separate from the neurosensory retina *(posterior vitreous detachment)*. The relationship between the posterior hyaloid and the neurosensory retina has a key role in the pathogenesis of retinal neovascularization and in some forms of retinal detachment.

Retinal Detachment

Retinal detachment (separation of the neurosensory retina from the RPE) is broadly classified by etiology based on the presence or absence of a break in the retina. *Rhegmatogenous retinal detachment* is associated with a full-thickness retinal defect. Retinal tears may develop after the vitreous collapses structurally, and the posterior hyaloid exerts traction on points of abnormally strong adhesion to the retinal internal limiting membrane. Liquefied vitreous humor then seeps through the tear and gains access to the potential space between the neurosensory retina and the RPE (Fig. 29.17). Reattachment of the retina to the RPE generally requires relief of vitreous traction through indenting of the sclera by surgical procedures. This can be accomplished by the application of strips of silicon to the surface of the eye (scleral buckling) and possibly by removal of vitreous material (vitrectomy). Rhegmatogenous retinal detachment

may be complicated by *proliferative vitreoretinopathy*, the formation of epiretinal or subretinal membranes by retinal glial cells or RPE cells.

Nonrhegmatogenous retinal detachment (retinal detachment without retinal break) may complicate retinal vascular disorders associated with significant exudation and any condition that damages the RPE and permits fluid to leak from the choroidal circulation under the retina. Retinal detachments associated with choroidal tumors and malignant hypertension are examples of nonrhegmatogenous retinal detachment.

Retinal Vascular Disease

Hypertension

Normally, the thin walls of retinal arterioles permit a direct visualization of the circulating blood by ophthalmoscopy. In retinal arteriolosclerosis the thickened arteriolar wall changes the ophthalmic perception of circulating blood: vessels may appear narrowed, and the color of the blood column may change from bright red to copper and to silver depending on the degree of vascular wall thickness (Fig. 29.18A). Retinal arterioles and veins share a common adventitial sheath. Therefore, in pronounced retinal arteriolosclerosis the arteriole may compress the vein at points where both vessels cross (Fig. 29.18B). Venous stasis distal to arteriolar-venous crossing may precipitate occlusions of the retinal vein branches.

In malignant hypertension, vessels in the retina and choroid may be damaged. Damage to choroidal vessels may produce focal choroidal infarcts, seen clinically as *Elschnig spots*. Damage to the choriocapillaris, the internal layer of the choroidal vasculature, may, in turn, damage the overlying RPE and permit the exudate to accumulate in the potential space between the neurosensory retina and the RPE, thereby producing a retinal detachment. Exudate from damaged retinal arterioles typically accumulates in the outer plexiform layer of the retina (see Fig. 29.18A). The ophthalmoscopic finding of a macular star—a spoke-like arrangement of exudate in the macula in malignant hypertension—results from exudate accumulating in the outer plexiform layer of the macula that is oriented obliquely instead of perpendicular to the retinal surface.

Occlusion of retinal arterioles may produce infarcts of the nerve fiber layer of the retina (axons of the retinal ganglion cell layer populate the nerve fiber layer). Axoplasmic transport in the nerve fiber layer is interrupted at the point of axonal damage, and accumulation of mitochondria at the swollen ends of damaged axons creates the histologic illusion of cells *(cytoid bodies)*. Collections of cytoid bodies populate the nerve fiber layer infarct, seen ophthalmoscopically as "cotton-wool spots" (Fig. 29.19). Although nerve fiber layer infarcts are described here in the context of hypertension, they may be detected in a variety of retinal occlusive vasculopathies. For example, retinal nerve fiber layer infarcts may develop in individuals with AIDS due to a retinal vasculopathy that is similar to the brain vasculopathy that may develop in this condition.

Diabetes Mellitus

The eye is profoundly affected by diabetes mellitus. The effects of hyperglycemia on the lens and iris have already

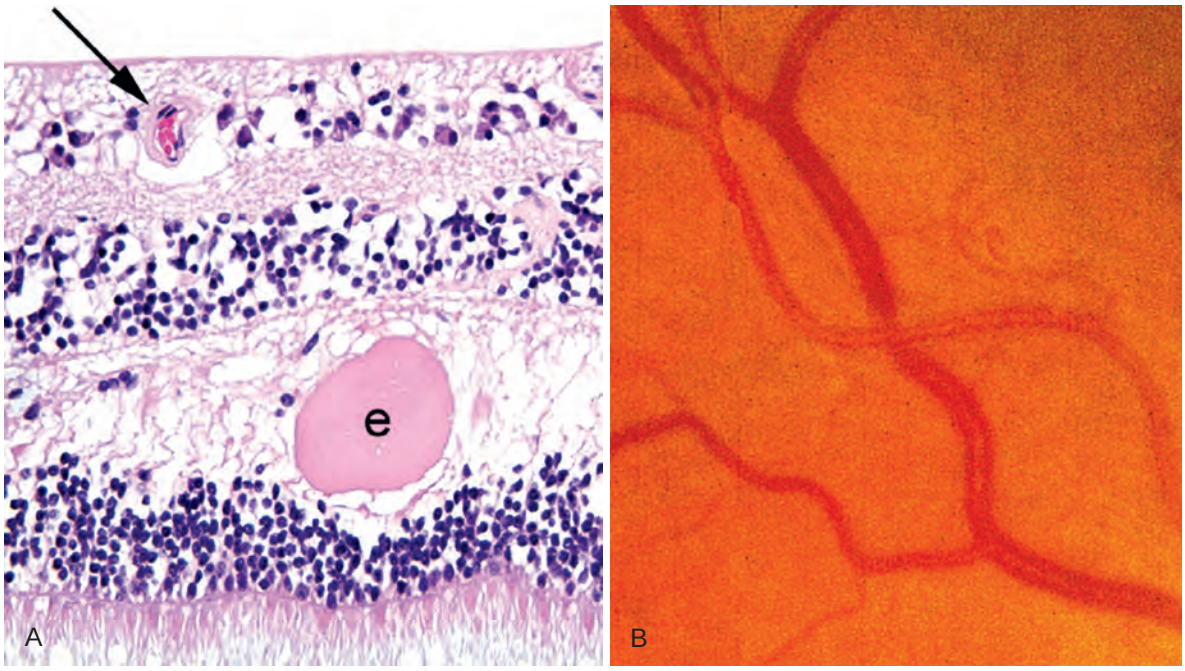


Figure 29.18 The retina in hypertension. (A) The wall of the retinal arteriole *(arrow)* is thick. Note the exudate *(e)* in the retinal outer plexiform layer. (B) The fundus in hypertension. The diameter of the arterioles is reduced, and the color of the blood column appears to be less saturated (copper wire–like). If the wall of the vessel were thicker still, the degree of red color would diminish such that the vessels might appear clinically to have a "silver wire" appearance. In this fundus photograph, note that the vein is compressed where the sclerotic arteriole crosses over it. (B, Courtesy Dr. Thomas A. Weingeist, Department of Ophthalmology and Visual Science, University of Iowa, Iowa City, Ia.)

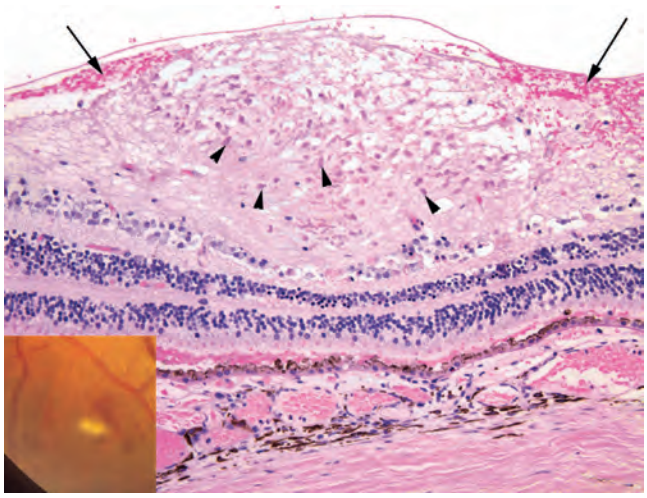


Figure 29.19 Nerve fiber layer infarct. A "cotton-wool spot" is illustrated in the inset, adjacent to a flame-shaped (nerve fiber layer) hemorrhage. The histology of a cotton-wool spot—an infarct of the nerve fiber layer of the retina—is illustrated in the photomicrograph. A focal swelling of the nerve fiber layer is occupied by numerous red to pink cytoid bodies *(arrowheads)*. Hemorrhage *(arrows)* surrounding the nerve fiber layer infarct as illustrated here is a variable and inconsistent finding. (Fundus photograph, Courtesy Dr. Thomas A. Weingeist, Department of Ophthalmology and Visual Science, University of Iowa, Iowa City, Ia.)

been mentioned. Thickening of the basement membrane of the epithelium of the pars plicata of the ciliary body is a reliable histologic marker of diabetes mellitus in the eye (Fig. 29.20) and is reminiscent of similar changes in the glomerular mesangium. This discussion focuses on the retinal microangiopathy associated with diabetes mellitus, a prototype for the consideration of other retinal microangiopathies.

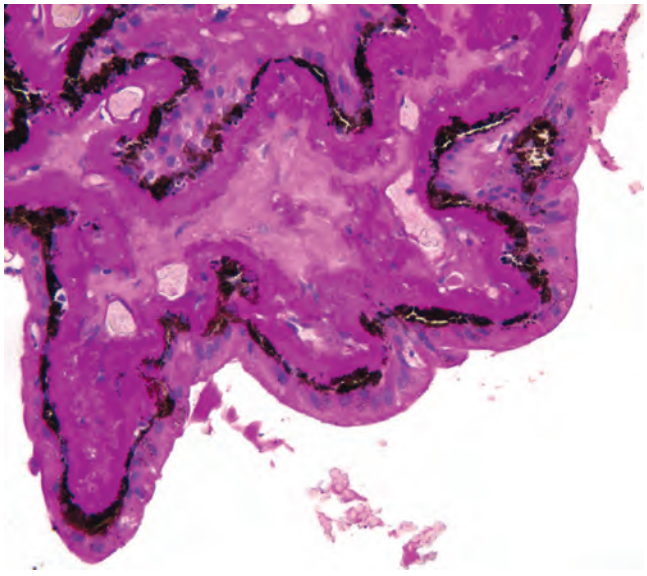


Figure 29.20 The ciliary body in chronic diabetes mellitus, periodic acid–Schiff stain. Note the massive thickening of the basement membrane of the ciliary body epithelia, reminiscent of changes in the mesangium of the renal glomerulus.

MORPHOLOGY

The retinal vasculopathy of diabetes mellitus can be classified into **nonproliferative** and **proliferative diabetic retinopathy.**

Nonproliferative diabetic retinopathy includes a spectrum of changes resulting from structural and functional abnormalities of retinal vessels (i.e., confined beneath the internal limiting membrane of the retina). As with diabetic microangiopathy in general, the **basement membrane of retinal blood vessels is thickened.** In addition, the number of pericytes relative to endothelial cells diminishes. **Microaneurysms** are an important manifestation of diabetic microangiopathy. They are typically smaller than the resolution of direct ophthalmoscopes, and findings customarily described as microaneurysms by ophthalmoscopy may in fact be retinal microhemorrhages. Structural changes in the retinal microcirculation have been associated with a physiologic breakdown in the blood-retinal barrier. Recall that VEGF was initially called vascular permeability factor. Thus, the retinal microcirculation in diabetics may be exceptionally leaky, giving rise to **macular edema,** a common cause of visual loss in these patients. The vascular changes may also produce **exudates** that accumulate in the outer plexiform layer. Although the retinal microcirculation is often hyperpermeable, it is also subject to the effects of micro-occlusion. Both vascular incompetence and vascular micro-occlusions can be visualized clinically after intravenous injection of fluorescein. Nonperfusion of the retina due to the microcirculatory change described earlier is associated with upregulation of VEGF and intraretinal angiogenesis (located beneath the internal limiting membrane of the retina).

Proliferative diabetic retinopathy is defined by the appearance of new vessels sprouting on the surface of either the optic nerve head or the surface of the retina (Fig. 29.21C). The term "retinal neovascularization" is applied only when the newly formed vessels breach the internal limiting membrane of the retina. The quantity and location of retinal neovascularization guide the ophthalmologist in the treatment of proliferative diabetic retinopathy. The web of newly formed vessels is referred to as a neovascular membrane. It is composed of angiogenic vessels with or without a substantial supportive fibrous or glial stroma (Fig. 29.21B).

If the vitreous humor has not detached and the posterior hyaloid is intact, neovascular membranes extend along the potential plane between the retinal internal limiting membrane and the posterior hyaloid. If vitreous humor later separates from the internal limiting membrane of the retina **(posterior vitreous detachment),** there may be massive hemorrhage from the disrupted neovascular membrane. In addition, scarring associated with the organization of the retinal neovascular membrane may wrinkle the retina, disrupting the orientation of retinal photoreceptors and producing visual distortion, and may exert traction on the retina, separating it from the RPE (retinal detachment). **Traction retinal detachment** may begin as a nonrhegmatogenous detachment, but severe traction may tear the retina, producing a traction rhegmatogenous detachment.

Retinal neovascularization may be accompanied by the development of a neovascular membrane on the iris surface, presumably secondary to increased levels of VEGF in the aqueous humor. Contraction of the iris neovascular membrane may lead to adhesions between the iris and trabecular meshwork (anterior synechiae), thus occluding a major pathway for aqueous outflow and thereby contributing to elevation of the intraocular pressure **(neovascular glaucoma).**

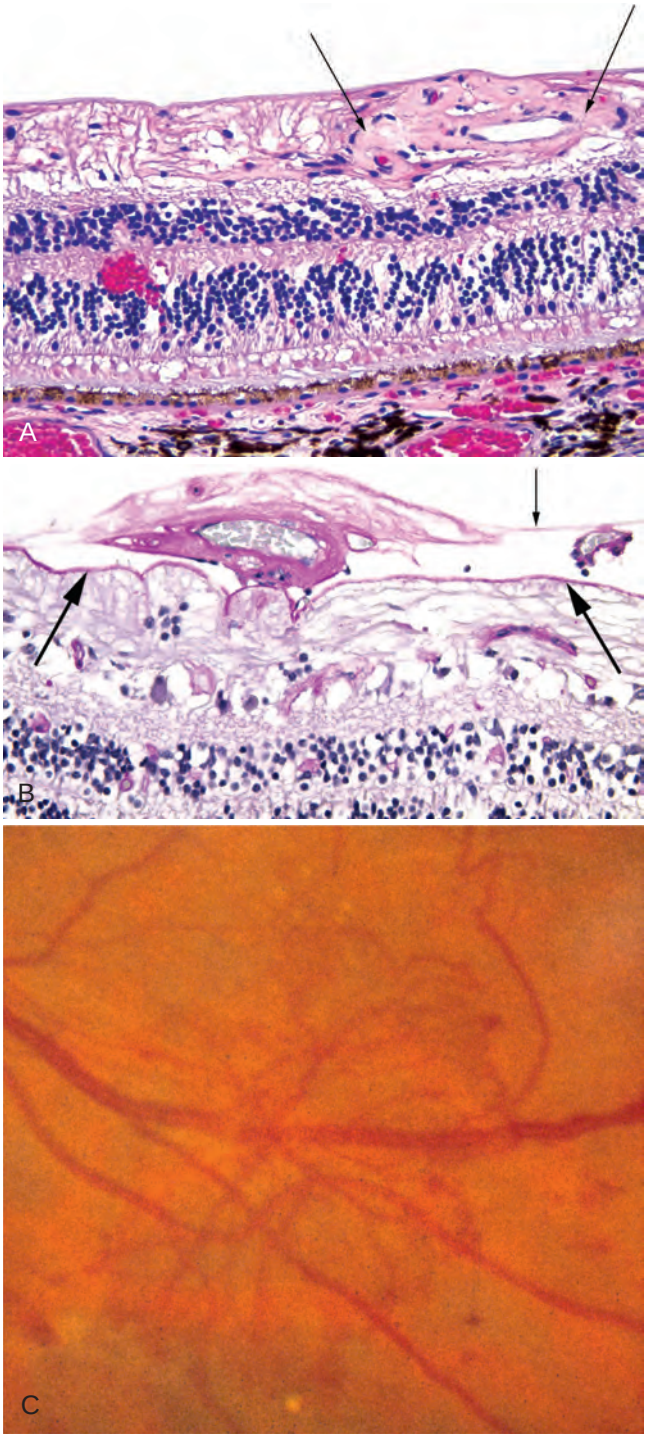


Figure 29.21 The retina in diabetes mellitus (see Fig. 29.16 for a schematic of retinal structure). (A) A tangle of abnormal vessels lies just beneath the internal limiting membrane of the retina on the right half of the photomicrograph *(between arrows)*. This is an example of intraretinal angiogenesis known as intraretinal microangiopathy (IRMA). Note the retinal hemorrhage in the outer plexiform layer in the left half. The ganglion cell layer and the nerve fiber layer—the axons of the ganglion cells—are absent. The rarefied space beneath internal limiting membrane to the left of the focus of IRMA consists largely of elements of retinal glial (Müller) cells. Absence of the ganglion cell and nerve fiber layers is a hallmark of glaucoma. The chronic diabetes mellitus in this individual was complicated by iris neovascularization and secondary angle-closure glaucoma (neovascular glaucoma). (B) In this section stained by periodic acid–Schiff, the internal limiting membrane is indicated by the *thick arrows*, and the posterior hyaloid of the vitreous is indicated by the *thin arrow*. In the potential space between these two landmarks, the vessels to the left of the *thin arrow* are invested with a fibrous-glial stroma and would appear ophthalmoscopically as a white neovascular membrane. The thin-walled vessel to the right of the *thin arrow* is not invested with connective tissue. A posterior vitreous detachment in an eye such as this might exert traction on these new vessels and precipitate a massive vitreous hemorrhage. (C) Ophthalmoscopic view of retinal neovascularization (known clinically as neovascularization "elsewhere" in contrast to neovascularization of the optic disc) creating a neovascular membrane.

Ablating nonperfused retina by laser photocoagulation or cryopexy triggers regression of both retinal and iris neovascularization, emphasizing the central role that retinal hypoxia has in these disorders. More recently, the injection of VEGF inhibitors into the vitreous has been used to treat diabetic macular edema and retinal neovascularization, a successful example of how knowledge of the molecular pathogenesis of a condition may evolve into a successful therapeutic strategy.

Retinopathy of Prematurity (Retrolental Fibroplasia)

At term, the temporal (lateral) aspect of the retinal periphery is incompletely vascularized, whereas the medial aspect is vascularized. In premature or low-birth-weight infants treated with oxygen, immature retinal vessels in the temporal retinal periphery constrict, rendering the retinal tissue distal to this zone ischemic. Retinal ischemia may result in upregulation of proangiogenic factors such as VEGF and lead to retinal angiogenesis. Contraction of the resulting peripheral retinal neovascular membrane may "drag" the temporal aspect of the retina toward the peripheral zone, displacing the macula (situated temporal to the optic nerve) laterally. Neovascular membrane contraction may create sufficient force to cause retinal detachment. The use of VEGF inhibition in this condition is under investigation.

Sickle Retinopathy, Retinal Vasculitis, and Radiation Retinopathy

Retinopathy affecting individuals with sickle hemoglobinopathies (Chapter 14) has been divided into two types that roughly parallel those used for diabetic retinopathy: nonproliferative (intraretinal angiopathic changes) and proliferative (retinal neovascularization). The final common pathway in both types is vascular occlusion. Low oxygen tension within the blood vessels in the retinal periphery results in red cell sickling and microvascular occlusions. In the nonproliferative form (which occurs in individuals with hemoglobin SS and SC genotypes), vascular occlusions are thought to contribute to preretinal, intraretinal, and subretinal hemorrhages. The resolution of these hemorrhages may give rise to several ophthalmoscopically visible changes, known as *salmon patches, iridescent spots,* and *black sunburst lesions.*

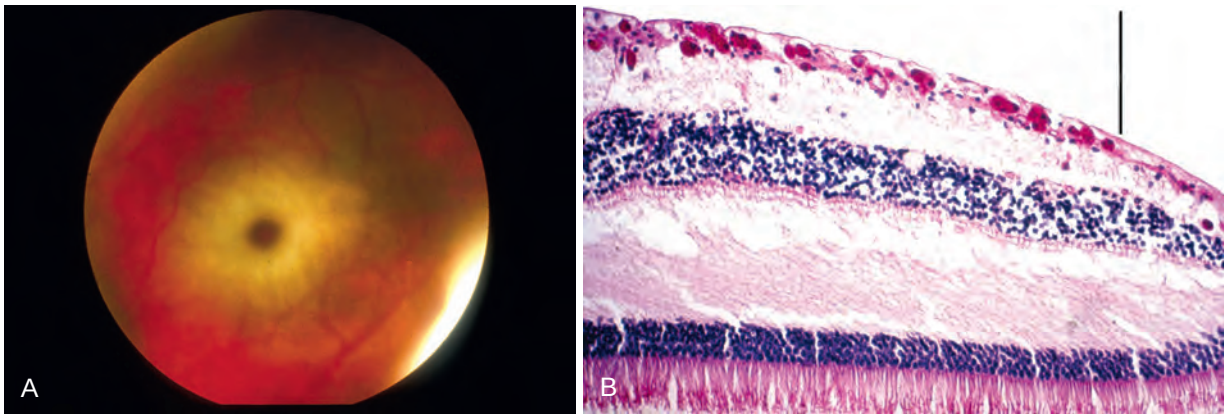


Figure 29.22 Cherry-red spot in Tay-Sachs disease. (A) Fundus photograph of the cherry-red spot in Tay-Sachs disease. (B) Photomicrograph of the macula in an individual with Tay-Sachs disease, stained with periodic acid–Schiff to highlight the accumulation of ganglioside material in the retinal ganglion cells. The presence of ganglion cells filled with gangliosides outside the fovea blocks the transmission of the normal orange-red color of the choroid, but absence of ganglion cells within the fovea (to the right of the *vertical bar*) permits the normal orange-red color to be visualized, accounting for the so-called cherry-red spot. (A, Courtesy Dr. Thomas A. Weingeist, Department of Ophthalmology and Visual Science, University of Iowa, Iowa City, Ia.; B, from the teaching collection of the Armed Forces Institute of Pathology.)

Organization of pre-retinal hemorrhage may result in retinal traction and retinal detachment. Vascular occlusions may also contribute to angiogenesis secondary to upregulation of both VEGF and basic fibroblast growth factor. This can give rise to areas of florid neovascularization in the periphery of the retina, described clinically as "sea-fans."

Neovascularization also occurs in a variety of other clinical settings such as peripheral retinal vasculitis and in irradiation used to treat intraocular tumors. The feature common to these conditions is damage to retinal vessels, producing zones of retinal ischemia that trigger retinal angiogenesis and its complications hemorrhage and traction, which, in turn, may cause detachment.

Retinal Artery and Vein Occlusions

The central retinal artery or its branches can be occluded by disorders that affect the vessels in general. For example, the lumen of the central retinal artery can be narrowed significantly by atherosclerosis, thus predisposing to thrombosis. Emboli to the central retinal artery can originate from thrombi in the heart or from ulcerated atheromatous plaques in the carotid arteries. Fragments of atherosclerotic plaques can lodge within the retinal circulation *(Hollenhorst plaques)*. Total occlusion of a branch of retinal artery can produce a segmental infarct of the retina. With sudden cessation of blood supply, the retina (an embryologic derivative of brain tissue) swells acutely and becomes optically opaque. By ophthalmoscopy the fundus in the affected area appears white instead of red or orange because the retinal opacity blocks the view of the richly vascular choroid.

Total occlusion of the central retinal artery can produce a diffuse infarct of the retina. Following an acute occlusion, the retina appears relatively opaque by ophthalmoscopy. The fovea and foveola are physiologically thin; therefore, the normal orange-red of the choroid is not only visible but also highlighted by the surrounding opaque retina—the origin of the *cherry-red spot* of the central retinal artery occlusion. Cherry-red spots can also be seen in rare storage diseases such as *Tay-Sachs* and *Niemann-Pick* diseases because of the structural organization of the retina. The storage material accumulates in retinal ganglion cells: the ganglion cell layer of the macula surrounding the fovea is thick, but there are no ganglion cells in the center of the macula, the fovea. Thus, the fovea is relatively transparent to the underlying choroidal vasculature but is rimmed by relatively opaque retina, the result of storage material accumulating in the perifoveal macular ganglion cells (Fig. 29.22).

Retinal vein occlusion may occur with or without ischemia. In ischemic retinal vein occlusion, VEGF and other proangiogenic factors are upregulated in the retina, leading to neovascularization of the retina and surface of the optic nerve head as well as neovascularization of the iris and subsequent angle-closure glaucoma. Nonischemic retinal vein occlusion may be complicated by hemorrhages, exudates, and macular edema.

Age-Related Macular Degeneration (AMD)

AMD results from damage to the macula, which is required for central vision. It occurs in two forms, dry and wet, that are distinguished by the presence of neoangiogenesis in the wet form and its absence in the dry form. From the name of this disorder, it is clear that advancing age is a risk factor. The cumulative incidence of AMD in individuals 75 years of age and older is 8%, and with increasing longevity AMD is becoming a major health problem.

Atrophic or "dry" AMD is characterized ophthalmoscopically by diffuse or discrete deposits in the Bruch membrane *(drusen)* and geographic atrophy of the RPE. Loss of vision may be severe in these individuals. The oral consumption of zinc and of vitamins with antioxidant properties may be associated with a slowing of the progression to AMD. There is currently no effective treatment for "dry" or atrophic AMD; a regenerative approach to replacing diseased RPE cells with stem cells is under investigation.

Neovascular or "wet" AMD is characterized by *choroidal neovascularization*, defined by the presence of vessels that presumably originate from the choriocapillaris and penetrate through the Bruch membrane beneath the RPE (Fig. 29.23). This neovascular membrane may also penetrate the RPE and

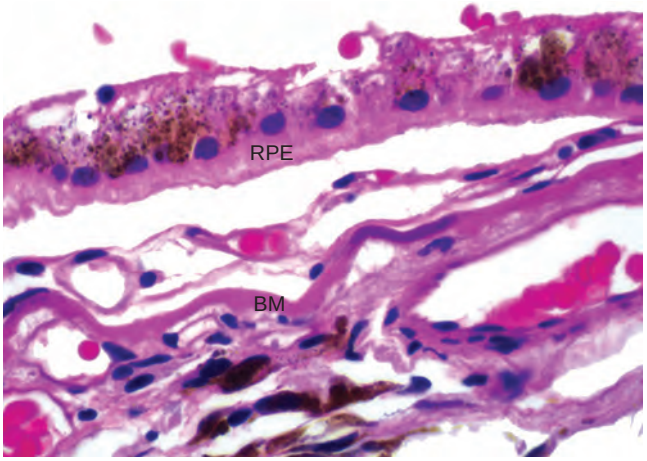


Figure 29.23 "Wet" age-related macular degeneration. A neovascular membrane is positioned between the retinal pigment epithelium *(RPE)* and Bruch membrane *(BM)*. Note the blue discoloration of BM to the right of the label, indicating focal calcification.

become situated directly beneath the neurosensory retina. The vessels in this membrane may leak, and the exuded blood may be organized by RPE cells into macular scars. Occasionally, these vessels are the source of hemorrhage, leading to the localized suffusion of blood that may be mistaken clinically for an intraocular neoplasm, or give rise to diffuse vitreous hemorrhage. Currently the mainstay of treatment for neovascular AMD is the injection of VEGF antagonists into the vitreous of the affected eye. Photodynamic therapy is also being evaluated for effectiveness.

Choroidal neovascular membranes can develop in diverse conditions that are unrelated to age, such as pathologic myopia (Fuchs spot), following disruption of the Bruch membrane (due to trauma or other causes), or an immunologic response to systemic histoplasmosis (presumed ocular histoplasmosis syndrome).

To understand the pathogenesis of AMD it is important to appreciate the existence of a structural and functional unit composed of the RPE, Bruch membrane (which contains the basement membrane of the RPE), and the innermost layer of the choroidal vasculature, the choriocapillaris. Disturbance in any component of this "unit" affects the health of the overlying photoreceptors, producing visual loss.

Attention is now focused on the roles of several genes, especially *CFH* (complement factor H) and other complement regulatory genes in the pathogenesis of this condition. The complement regulatory gene variants that are associated with AMD all appear to decrease their function, implying that AMD may stem from an excess of complement activity. Environmental exposures such as cigarette smoking may also increase risk of AMD, especially in genetically predisposed individuals.

Other Retinal Degenerations

Retinitis Pigmentosa

Retinitis pigmentosa is an inherited condition resulting from mutations that affect rods and cones or RPE. It can cause varying degrees of visual impairment including, in some cases, total blindness. The term *"retinitis" pigmentosa* is an unfortunate relic of the time when these disorders were incorrectly presumed to be inflammatory. They may be inherited as X-linked recessive, autosomal recessive, or autosomal dominant (the age of onset correlates with the inheritance pattern, with autosomal dominant retinitis pigmentosa appearing later in life). Retinitis pigmentosa may be part of a syndrome, such as the *Bardet-Biedl syndrome, Usher syndrome,* or *Refsum disease,* or may develop in isolation (nonsyndromic retinitis pigmentosa).

Retinitis pigmentosa is linked to mutations affecting *RHO, USH2A, RGPR,* and *EYS* genes; they regulate the functions of either the photoreceptor cells or the RPE. *Typically, both rods and cones are lost to apoptosis,* though in varying proportions. Loss of rods may lead to early night blindness and constricted visual fields. As cones are lost, central visual acuity may be affected. Clinically, retinal atrophy is accompanied by constriction of retinal vessels and optic nerve head atrophy ("waxy pallor" of the optic disc) and the accumulation of retinal pigment around blood vessels, thus accounting for the "pigmentosa" in the disease name. The electroretinogram reveals abnormalities characteristic of this disease.

Retinitis

A variety of pathogens can contribute to the development of infectious retinitis. For example, *Candida* may disseminate to the retina hematogenously, especially in the setting of intravenous drug abuse or in systemic candidemia from other causes. Hematogenous dissemination of pathogens to the retina typically results in multiple retinal abscesses. As was mentioned previously, cytomegalovirus retinitis is an important cause of visual morbidity in immunocompromised individuals, especially those with AIDS.

Retinal Neoplasms

Retinoblastoma

Retinoblastoma is the most common primary intraocular malignancy of children. The molecular genetics of retinoblastoma are discussed in detail in Chapter 7. Although the name retinoblastoma might suggest origin from a primitive retinal cell that is capable of differentiation into both glial and neuronal cells, it is now clear that the cell of origin of retinoblastoma is a neuronal progenitor. Recall that in approximately 40% of cases, retinoblastoma occurs in individuals who inherit a germline mutation of one *RB* allele. Retinoblastoma arises when the retinal progenitor suffers a second, somatic mutation and the *RB* gene function is lost. In the sporadic cases, both *RB* alleles are lost by somatic mutations. Retinoblastomas arising in those with germline mutations are often bilateral. In addition, they may be associated with pinealoblastoma ("trilateral" retinoblastoma), which is associated with a dismal outcome.

MORPHOLOGY

The pathology of retinoblastoma, both hereditary and sporadic types, is identical. Tumors may contain both undifferentiated and

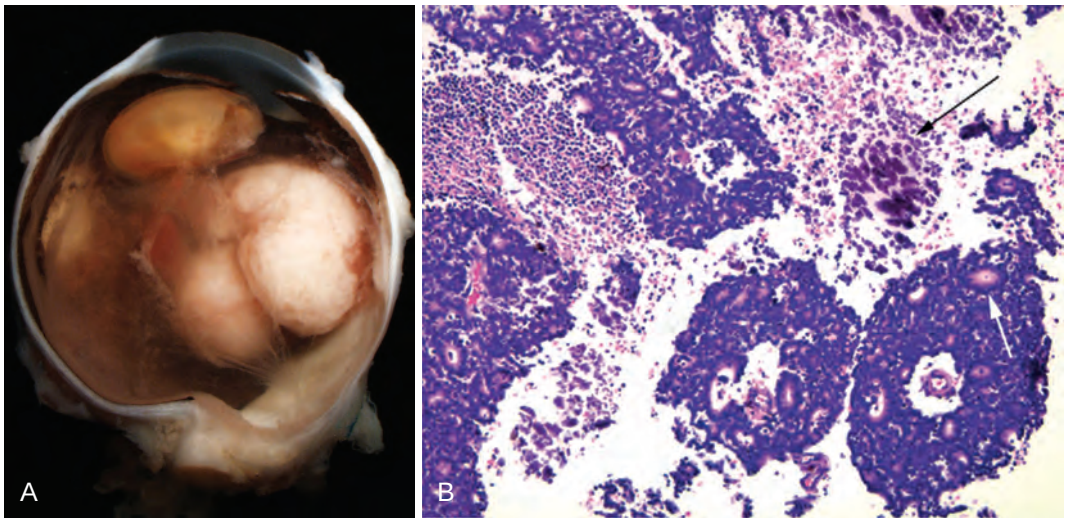


Figure 29.24 Retinoblastoma. (A) Gross photograph of retinoblastoma. (B) Tumor cells appear viable when in proximity to blood vessels, but necrosis is seen as the distance from the vessel increases. Dystrophic calcification *(black arrow)* is present in the zones of tumor necrosis. Flexner-Wintersteiner rosettes—arrangements of a single layer of tumor cells around an apparent "lumen"—are seen throughout the tumor, and one such rosette is indicated by the *white arrow*.

differentiated elements. The former appear as collections of small, round cells with hyperchromatic nuclei. In well-differentiated tumors there are **Flexner-Wintersteiner rosettes** and fleurettes reflecting photoreceptor differentiation. It should be noted, however, that the degree of tumor differentiation does not appear to be associated with the prognosis. As seen in Fig. 29.24, viable tumor cells are found encircling tumor blood vessels with zones of necrosis typically found in relatively avascular areas, illustrating the dependence of retinoblastoma on its blood supply. Focal zones of dystrophic calcification are characteristic of retinoblastoma.

In an effort to preserve vision and eradicate the tumor, many ophthalmic oncologists now attempt to reduce tumor burden by administration of chemotherapy, including selective delivery of the drug to the eye through the ophthalmic artery; after chemoreduction, tumors may be obliterated by laser treatment or cryopexy. Retinoblastoma tends to spread to the brain and bone marrow and seldom disseminates to the lungs. Prognosis is adversely affected by extraocular extension and invasion along the optic nerve and by choroidal invasion. A variant of retinoblastoma—retinocytoma or retinoma—has been reported and appears to be a premalignant lesion. The appearance of retinoblastoma in one eye and retinocytoma in the other eye is characteristic of heritable retinoblastoma.

Retinal Lymphoma

Primary retinal lymphoma is an aggressive tumor that characteristically involves the two retinal layers derived from brain, the neurosensory retina and the RPE. Primary intraocular lymphoma tends to occur in older individuals and may mimic uveitis clinically. Most are diffuse large B-cell lymphomas (Chapter 13). Spread to the brain commonly occurs via the optic nerve. The diagnosis depends on a demonstration of lymphoma cells in vitreous aspirates.

KEY CONCEPTS

- Retinal detachment, a separation of the neurosensory retina from the RPE, may be the consequence of a break in the retina (rhegmatogenous retinal detachment) or may develop without a retinal break because of pathology within or beneath the retina (nonrhegmatogenous retinal detachment).
- The clinical appearance of the retina by ophthalmoscopy can be linked to specific pathologic changes: the change in caliber and color of retinal blood vessels may reflect varying degrees of arteriolosclerosis, and the location of hemorrhages and exudates in the retina is related to their locations within the retinal layers.
- Several major causes of blindness result from pathologic intraocular angiogenesis including proliferative diabetic retinopathy and exudative (wet) age-related macular degeneration, among many other conditions. VEGF antagonists may prevent visual loss in many of these conditions.
- Retinoblastoma is the most common primary intraocular tumor of children.
- Primary retinal lymphoma is an aggressive tumor that often involves the brain as well.

OPTIC NERVE

As a sensory tract of the central nervous system, the optic nerve is surrounded by meninges, and cerebrospinal fluid circulates around the nerve. The pathology of the optic nerve is similar to the pathology of the brain. For example, the most common primary neoplasms of the optic nerve are glioma (typically pilocytic astrocytomas) and meningioma.

Anterior Ischemic Optic Neuropathy

There are striking similarities between stroke and a condition known in ophthalmic terminology as anterior ischemic

optic neuropathy (AION). As used clinically, the term AION includes a spectrum of injuries to the optic nerve varying from ischemia to infarction. Thus, transient partial interruptions in blood flow to the optic nerve can produce episodes of transient loss of vision, whereas total interruption in blood flow can give rise to an optic nerve infarct which may be segmental or total. Zones of relative ischemia may surround segmental infarcts of the optic nerve. Optic nerve function in these poorly perfused but not infarcted zones may recover. The optic nerve does not regenerate, and visual loss from infarction is permanent.

Interruption in the blood supply to the optic nerve can result from inflammation of the vessels that supply the optic nerve or from embolic or thrombotic events. Bilateral total infarcts of the optic nerve resulting in total blindness have been reported in temporal arteritis, adding urgency to the treatment of this condition with high doses of corticosteroids.

Papilledema

Edema of the head of the optic nerve may develop as a consequence of compression of the nerve (as in a primary neoplasm of the optic nerve when swelling of the nerve head produces unilateral disc edema) or from elevations of cerebrospinal fluid pressure surrounding the nerve (resulting typically in bilateral disc edema). The concentric increase in pressure encircling the nerve contributes to venous stasis and also interferes with axoplasmic transport, leading to nerve head swelling. Swelling of the optic nerve head in elevated intracranial pressure is bilateral and is commonly termed *papilledema.* Typically, acute papilledema from increased intracranial pressure is not associated with visual loss. Ophthalmoscopically, the optic nerve head is swollen and hyperemic; by contrast, the optic nerve head in the relatively acute phases of anterior ischemic optic neuropathy appears swollen and pale because of decreased nerve perfusion (Fig. 29.25). In papilledema secondary to increased intracranial pressure, the optic nerve may remain congested for a prolonged period of time.

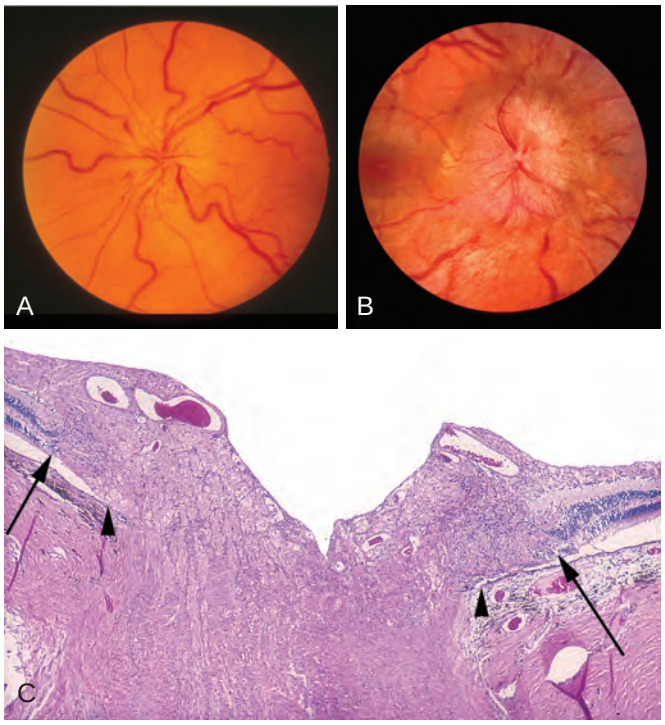


Figure 29.25 The optic nerve in anterior ischemic optic neuropathy (AION) and papilledema. (A) In the acute phases of AION the optic nerve may be swollen, but it is relatively pale because of decreased perfusion. (B) In papilledema secondary to increased intracranial pressure, the optic nerve is typically swollen and hyperemic. (C) Normally, the termination of Bruch membrane *(arrowheads)* is aligned with the beginning of the neurosensory retina, as indicated by the presence of stratified nuclei *(arrows),* but in papilledema the optic nerve is swollen, and the retina is displaced laterally. This is the histologic explanation for the blurred margins of the optic nerve head seen clinically in this condition. (A and B, Courtesy Dr. Sohan S. Hayreh, Department of Ophthalmology and Visual Science, University of Iowa, Iowa City, Ia.; C, from the teaching collection of the Armed Forces Institute of Pathology.)

Glaucomatous Optic Nerve Damage

As discussed, the majority of individuals with glaucoma have elevated intraocular pressure. However, there is a small group that develops the visual field and optic nerve changes typical of glaucoma with normal intraocular pressure—so-called normal-tension glaucoma. Conversely, some individuals with elevated intraocular pressure who are followed over long periods of time never develop visual field changes or optic nerve cupping. Therefore, it is clear that there is a spectrum of neuronal susceptibility to the effects of elevated intraocular pressure. Considerable research is now directed toward understanding mechanisms by which the optic nerve axons may be protected from injury.

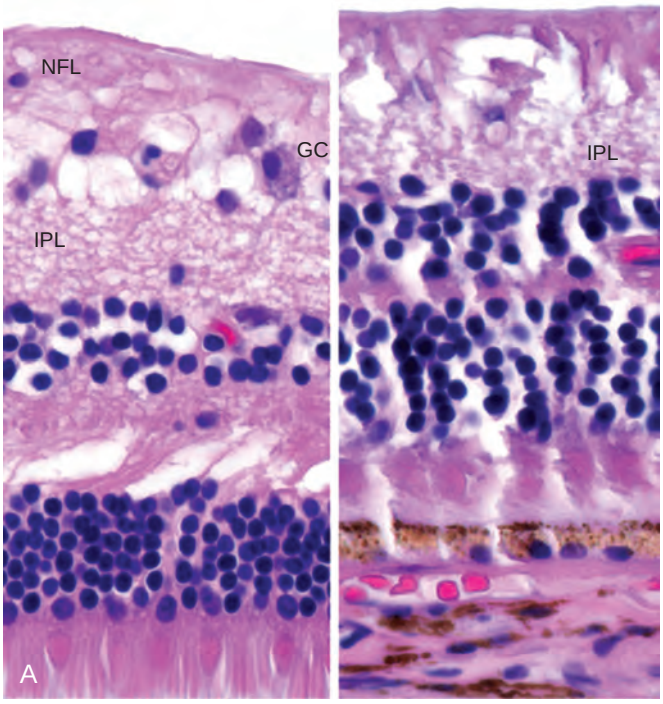
MORPHOLOGY

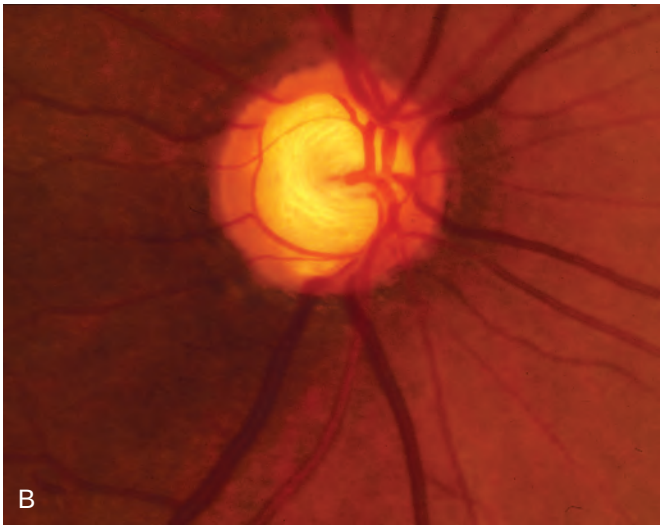
Characteristically, there is a diffuse loss of ganglion cells and thinning of the retinal nerve fiber layer (Fig. 29.26), which can be measured by optical coherence tomography. In advanced cases, the optic nerve is both cupped and atrophic, a combination unique to glaucoma. Elevated intraocular pressure in infants and children can lead to diffuse enlargement of the eye **(buphthalmos)** or enlargement of the cornea **(megalocornea).** After the eye reaches its adult size, prolonged elevation of intraocular pressure can lead to focal thinning of the sclera, and uveal tissue may line ectatic sclera **(staphyloma).**

Other Optic Neuropathies

Optic neuropathy may be inherited or may be secondary to nutritional deficiencies or toxins such as methanol. Individuals may suffer severe visual compromise. If the nerve fibers that originate from the macula are affected, then central visual acuity is lost.

Leber hereditary optic neuropathy results from inheritance of mitochondrial gene mutations (Chapter 5). Since neuronal health is dependent on axoplasmic transport of mitochondria, mitochondrial dysfunctions give rise to neurologic disorders including optic neuropathy. Leber optic neuropathy shows maternal inheritance pattern typical of mitochondrial gene mutations. Males are affected far more commonly than





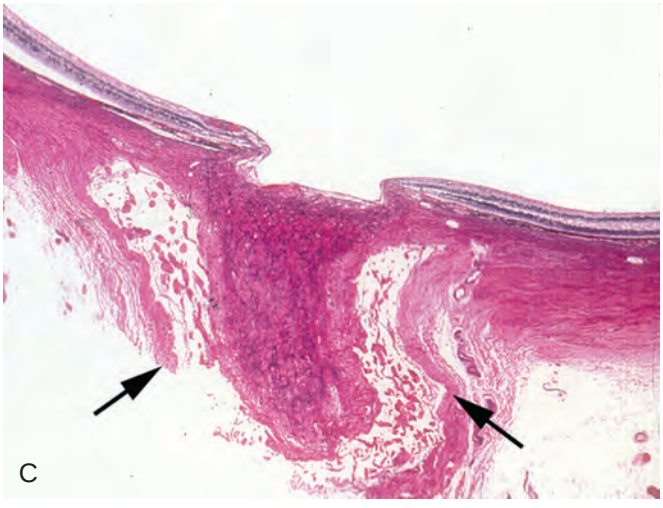


Figure 29.26 The retina and optic nerve in glaucoma. (A) *Left panel,* normal retina; *right panel,* the retina in long-standing glaucoma (same magnification). The full thickness of the glaucomatous retina is captured *(right),* a reflection of the thinning of the retina in glaucoma. In the glaucomatous retina, the areas corresponding to the nerve fiber layer *(NFL)* and ganglion cell layer *(GC)* are atrophic; the inner plexiform layer *(IPL)* is labeled for reference. Note also that the outer nuclear layer of the glaucomatous retina is aligned with the inner nuclear layer of the normal retina due to the thinning of the retina in glaucoma. See Fig. 29.16 for orientation. (B) Glaucomatous optic nerve cupping results in part from loss of retinal ganglion cells, the axons of which populate the optic nerve. (C) The *arrows* point to the dura of the optic nerve. Notice the wide subdural space, a result of atrophy of the optic nerve. There is a striking degree of cupping on the surface of the nerve as a consequence of long-standing glaucoma.

females. The usual age of onset is between 10 and 30 years. It begins with clouding of vision that may progress to total loss of vision.

Optic Neuritis

Many unrelated conditions have historically been grouped under the heading of optic neuritis. Unfortunately, the term itself suggests optic nerve inflammation, which might not accurately describe the pathophysiologic changes. In common clinical usage the term *optic neuritis* is used to describe a loss of vision secondary to demyelinization of the optic nerve. One of the most important causes of optic neuritis is multiple sclerosis (Chapter 28). Indeed, optic neuritis may be the first manifestation of this disease. The 10-year risk of developing multiple sclerosis after the first attack of optic neuritis increases if the affected person has concomitant evidence of brain lesions as detected by magnetic resonance imaging. Individuals with a single episode of optic nerve demyelinization may recover vision and remain disease-free.

KEY CONCEPTS

- The term "anterior ischemic optic neuropathy" refers to a spectrum of ischemic injuries to the optic nerve varying from transient ischemia to infarction.
- Bilateral swelling of the optic nerve head known as papilledema may develop as a consequence of elevated cerebrospinal fluid pressure and stasis of axoplasmic transport within the optic nerve. Unilateral optic nerve head swelling may result from compression of the optic nerve such as in primary tumors of the nerve.
- In chronic glaucoma, the optic nerve may atrophy and the cup on the surface of the nerve may enlarge and deepen.
- Optic neuropathy may be inherited (as in Leber hereditary optic neuropathy) or may result from nutritional deficiencies or toxins such as methanol.

THE END-STAGE EYE: PHTHISIS BULBI

Trauma, intraocular inflammation, chronic retinal detachment, and many other conditions can give rise to an eye that is both small (atrophic) and internally disorganized: phthisis bulbi. Congenitally small eyes—hypoplastic or microphthalmic eyes—are generally not disorganized internally. Phthisical eyes typically show the following changes: the presence of exudate or blood between the ciliary body and sclera and the choroid and sclera *(ciliochoroidal effusion);* the presence of a membrane extending across the eye from one aspect of the ciliary body to the other *(cyclitic membrane);* chronic retinal detachment; optic nerve atrophy; the presence of intraocular bone, which is thought by many to originate from osseous metaplasia of the RPE; and a thickened sclera, especially posteriorly. Ciliochoroidal effusion is typically associated with the physiologic state of low intraocular pressure *(hypotony)*. The normal pull of the extraocular muscles on a hypotonous eye may render the appearance of the eye as square rather than round.

SUGGESTED READINGS

Orbit

Andrew NH, Coupland SE, Pirbhai A et al: Lymphoid hyperplasia of the orbit and ocular adnexa: a clinical pathologic review, *Surv Ophthalmol* 61:778, 2016. *[Concise and comprehensive review of lymphoid lesions of the orbit including IgG4-related disease and nonspecific orbital inflammation].*

Smith DJ, Hagedüs L: Graves' disease, *N Engl J Med* 375:1552, 2016. *[Excellent overview of current understanding of the pathogenesis of thyroid eye disease and discussion of the pathologic basis of therapeutic strategies].*

Eyelid and Conjunctiva

Hanovar SG, Manjandavida FP: Tumors of the ocular surface, *Indian J Ophthalmol* 63:187, 2015. *[Another well-illustrated, concise, and comprehensive review].*

Pe'er J: Pathology of eyelid tumors, *Indian J Ophthalmol* 65:177, 2016. *[Well-illustrated, concise, and comprehensive review].*

Cornea

Chang JH, Garg NK, Lunde E et al: Corneal neovascularization: an anti-VEGF therapy review, *Surv Ophthalmol* 57:415, 2012. *[Review of the pathologic basis for the application of vascular endothelial growth factor antagonists to prevent and treat corneal neovascularization].*

Mas-Tur V, MacGregor C, Jayaswal R et al: A review of keratoconus: diagnosis, pathophysiology and genetics, *Surv Ophthalmol* 2017. *[Keratoconus is a common indication for corneal transplant].*

Vedana G, Villareal G Jr, Jun AS: Fuchs endothelial corneal dystrophy: current perspectives, *Clin Ophthalmol* 10:321, 2016. *[Fuchs endothelial dystrophy is a major indication for corneal transplant. This article reviews the clinical features, genetics, and pathophysiology of this condition].*

Glaucoma

Jonas JB, Aung T, Bourne RR et al: Glaucoma, *Lancet* 390:2183, 2017.

Uvea

Onken MD, Worley LA, Char DH et al: Collaborative ocular oncology group report number 1: prognostic validation of a multi-gene prognostic assay in uveal melanoma, *Ophthalmology* 119:1596, 2012. *[This article describes a multicenter study to test the validity of a gene expression profile for prognostication of patients with primary uveal melanoma].*

Van Raamsdonk CD, Bezrookove V, Green G et al: Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi, *Nature* 457:599, 2009.

Wiesner T, Obenauf AC, Murali R et al: Germline mutations in BAP1 predispose to melanocytic tumors, *Nat Genet* 43:1018, 2011.

Yonekawa Y, Kim IK: Epidemiology and management of uveal melanoma, *Hematol Oncol Clin North Am* 26:1169, 2012.

Retina

Antonetti DA, Klein R, Gardner TW: Mechanisms of disease: diabetic retinopathy, *N Engl J Med* 366:1227, 2012. *[This is an outstanding and comprehensive review of the pathogenesis of diabetic retinopathy].*

Chan CC, Rubenstein JL, Coupland SE et al: Primary vitreoretinal lymphoma: a report from an international primary central nervous system lymphoma collaborative group symposium, *Oncologist* 16:1589, 2011. *[This is an excellent and comprehensive description of primary retinal lymphoma including the pathology of this condition].*

Daiger SP, Bowne SJ, Sullivan LS: Perspective on genes and mutations causing retinitis pigmentosa, *Arch Ophthalmol* 125:151, 2007. *[This is a review of the molecular genetics of the various conditions that are clustered under the rubric of "retinitis pigmentosa".]*

Dimaras H, Kimani K, Dimba EOA et al: Retinoblastoma, *Lancet* 379:1436, 2012. *[This is a comprehensive review of retinoblastoma, including the clinical manifestations, molecular pathogenesis, pathology, and treatment strategies].*

Dimras H, Corson TW, Cobrinik D et al: Retinoblastoma, *Nat Rev Dis Primers* 2015.

Duh EJ, Sun JK, Stitt AW: Diabetic retinopathy: current understanding, mechanisms, and treatment strategies, *JCI Insight* 2(14):e93751, 2017.

Giacalone JC, Wiley LA, Burnight ER et al: Concise review: patient-specific stem cells to interrogate inherited eye disease, *Stem Cells Transl Med* 5:132, 2016.

Lim LS, Mitchell P, Seddon JM et al: Age-related macular degeneration, *Lancet* 379:1728, 2012. *[This is an excellent and comprehensive review of age-related macular degeneration including the pathogenesis of this condition and the pathologic basis of molecular therapy].*

Rivera JC, Sapieha P, Joyal JS et al: *Neonatology* 100:343, 2011. *[This is a comprehensive review of the pathogenesis of retinopathy of prematurity].*

Optic Nerve

Arnold AC: Pathogenesis of nonarteritic anterior ischemic optic neuropathy, *J Neuroophthalmol* 23:157, 2003.

Bernstein SL, Johnson MA, Miller NR: Nonarteritic anterior ischemic optic neuropathy (NAION) and its experimental models, *Prog Retin Eye Res* 30:167, 2011. *[Although this paper deals with nonarteritic anterior ischemic optic neuropathy, the pathogenesis of optic nerve damage is described in detail].*

Chang EE, Goldberg JL: Glaucoma 2.0: neuroprotection, neuroregeneration, neuroenhancement, *Ophthalmology* 119:979, 2012. *[Most attention in the past has been directed to addressing elevations in intraocular pressure in the pathogenesis and treatment of glaucoma. This review article highlights a critical re-examination of the pathogenesis of optic nerve disease: the prevention of damage through neuroprotection and the means to evoke adaptive responses within the nerve to damage].*

Newman NJ: Treatment of hereditary optic neuropathies, *Nat Rev Neurol* 8:545, 2012. *[This comprehensive review focuses on Leber hereditary optic neuropathy and approaches to understanding the sequelae of mitochondrial dysfunction].*

Pau D, Al Dubidi N, Yalamanchili S et al: Optic neuritis, *Eye (Lond)* 25:833, 2011. *[This review provides an update on the association (or lack thereof) between optic neuritis and multiple sclerosis].*

Index

Page number followed by *t, f,* or *b* indicates table, figure, or box, respectively.