Preface

A good teacher aims to *un*cover a subject, not *cover* it. So said a wise former teacher of mine. In revising *Essential Genetics and Genomics, Seventh Edition*, I've tried to heed this advice. To *un*cover a subject means to expose, exhibit, unveil. To help *un*cover genetics, this new edition is:

- **Streamlined,** with *emphasis on concepts* illuminated by vivid example and stripped of extraneous detail
- **Focused,** with *Learning Objectives* stated explicitly at the beginning of each chapter
- **Skills oriented**, with *Stop & Think* problems inserted at strategic points in the text to enhance the reading experience and encourage higher-order, analytical thinking.

The brevity of the text meets the needs of the shorter, less comprehensive introductory course of one semester or quarter. The choice of topics is intended to help students master the following scientific competencies:

- Understand the basic processes of gene transmission, mutation, expression, and regulation.
- Analyze genetic processes using numerical relationships including ratios and proportions.
- Learn to formulate genetic hypotheses in a statistical framework, work out their consequences, and test the results against observed data.
- Develop basic skills in problem solving, including single-concept exercises, those requiring the application of several concepts in logical order, and numerical problems requiring some arithmetic for solution.
- Interpret genetic data and make valid inferences to reveal the underlying causes.
- Acquire an appreciation of current trends in genetics, as well as the social and historical context in which genetics has developed.

Scientific competency is the desired learning outcome of any course in a curriculum in STEM (science, technology, engineering, mathematics). Genetics is an excellent subject for achieving scientific competency. It is integrative over a broad territory, ranging from molecular biology to evolutionary genetics. It is also quantitative, using concepts from probability, statistics, and computational biology. Fortunately, students come to a course in genetics highly motivated because of media reports about the human genome and genetic risk factors for disease, as well as many social and ethical controversies related to genetics such as direct-to-consumer and over-the-counter genetic testing, genetic privacy, cloning, stem-cell research, and genetically modified organisms. The challenges for the instructor are to sustain this motivation and to help students acquire the skills and habits of thought that constitute scientific competency.

What's New in the *Seventh Edition*?

This seventh edition has been completely revised and updated. Each chapter has been thoroughly reworked. Important new methods and findings have been added, including **synthetic bacteria**, **higher-order chromatin structure, high-throughput genomic sequencing methods**, **personalized medicine**, and **CRISPR/Cas9 gene editing**. As new material has been added, an equal or greater amount of nonessential or outdated material has been deleted. Several of the chapters have been reorganized to allow smooth integration of the new material. The entire text has been condensed, clarified, and updated.

Major revisions and additions in the *Seventh Edition* include the following:

- Chapter 1 includes a new section emphasizing that most common traits are actually **complex traits** affected by multiple interacting genetic factors as well as environment. This principle includes most **common diseases**, which are influenced by multiple genetic risk factors and lifestyle choices. The section on genomes and proteomes has been updated to include **Syn3.0**, the first living, multiplying bacterial cell whose genome was created entirely by chemical synthesis.

- Chapter 2 includes a new section emphasizing that **genes affect traits at multiple levels** (molecular, cellular, developmental, morphological, and behavioral), and that in many cases the traits that are affected appear to be unrelated until the underlying biology is understood.
- Chapter 3 now includes discussion of the **epigenetic specification of the kinetochore**, in which a specialized histone (CENPA) replaces histone 3 in **centromeric nucleosomes** that helps recruit kinetochore-associated proteins leading to the assembly of the mature kinetochore to which spindle fibers attach.
- Chapter 4 has a much shortened and sharper discussion of the **principal types of genetic variation** with emphasis on single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and simple tandem repeats (STRs). The update also includes a summary of the results of **genome-wide association studies (GWAS)** to detect genetic risk factors for common disorders and genetic factors affecting complex traits.
- Chapter 5 has been significantly shortened and streamlined.
- Chapter 6 includes a completely reorganized and simplified discussion of DNA replication updated to include the **trombone model of replication** showing how synthesis of the leading and lagging strands is coordinated. The section on **massively parallel sequencing** has been expanded and moved to Chapter 10.
- Chapter 7 contains a clearer description of how bacterial cells are brought together in conjugation, and unnecessary detail on genetic fine structure of the rII gene in bacteriophage T4 has been eliminated.
- Chapter 8 makes good use of the **Stop & Think** feature to reinforce fundamental concepts and processes of transcription and translation.
- Chapter 9 includes a major new section on how chromatin is organized into higher-order structures composed of **topologically associating domains (TADs)**, **insulators**, and **compartments**. The section on **RNA interference** and **long noncoding RNAs** has been updated, and the material on programmed DNA rearrangements has been removed because it is less generally applicable than once thought to be.
- A completely revised and reorganized Chapter 10 includes a summary of the latest **high-throughput DNA sequencing technology** including **reversible terminators**, **ion-torrent sequencing**, **single-molecule sequencing**, and **nanopore sequencing**. It also includes a major new section on **personalized medicine (precision medicine)** as well as **direct-to-consumer genetic services** and **over-the-counter genetic testing kits**. This section points out the potential values of these approaches but also emphasizes their associated **ethical, legal, and social implications**. Finally, Chapter 10 contains a wholly new section on **CRISPR-Cas9 technology for genetic engineering** showing how CRISPR-Cas9 functions as a sort of immune system in bacteria, how the CRISPR-Cas9 molecules are used to create targeted knockout mutations, and how they are used in DNA editing to change the genome sequence in a predetermined manner. Methods of CRISPR-Cas9 use in insects, mice, and plants are also discussed.
- Chapter 11 contains a slightly expanded section on the use of **epistasis in the analysis of switch-regulation pathways**, and some unnecessary details have been omitted such as minutiae of genetic control of yeast mating type.
- New to Chapter 12 is a major new section on estimates of the **rate of base-substitution mutation in humans** as determined by genome sequencing of parental and offspring genomes. The mutation rate increases steadily with father's age but not with mother's age, and we discuss why this finding is completely consistent with the reproductive biology of males and females. I've also deleted the section on the "ClB method" for detecting mutations in *Drosophila*, as this is mainly of historical interest.
- Chapter 13 puts greater emphasis on the connection between genetic control of the **cell cycle and cancer**, and it has been extensively revised, many of the illustrations simplified, and dispensable details eliminated.
- Chapter 14 contains a shorter, streamlined section on **molecular phylogenetics**.
- What's new in Chapter 15 is a major new section on **genome-wide association studies (GWAS)** focusing on the usually **large number of genes affecting complex traits** and their typically **small individual effects**. Each genetic risk factor for a disease usually increases disease risk by only a small amount. This generalization underlines the importance of proper interpretation of direct-to-consumer (DTC) and over-the-counter (OTC) genetic testing. The chapter also includes a new discussion of **physiological epistasis and statistical epistasis**

and explains why genes can exhibit a great deal of physiological epistasis at the molecular, cellular, and organismal levels without showing any substantial statistical epistasis at the population level. A classic method for estimating the number of genes affecting quantitative traits, based on differences between means of inbred lines and the genetic variance, has been deleted because it is obsolete and usually results in absurdly small estimates.

Chapter Organization

Each chapter begins with a set of **Learning Objectives** to orient students toward the knowledge and skills they should focus on. Explicit **Learning Objectives** help students to:

- *Identify* what they should know or be able to do as a result of their study.
- *Focus* on the knowledge they should have acquired from studying the chapter.
- *Guide* the student to identify key concepts and use them at a variety of learning levels including comprehension, application, analysis, and synthesis.
- *Highlight the skills* they should acquire through practice problems of various types.

Each chapter has an opening paragraph that gives an overview of what is to come, illustrates the subject with engaging examples, and shows how the material is connected to genetics as a whole. The section and subsection **Headings** are in the form of complete sentences that encapsulate the main message. The text makes liberal use of **Numbered** and **Bulleted Lists** to aid students in organizing their learning, as well as **Key Concepts** set apart from the main text to emphasize important principles. A feature called **Stop & Think** recognizes that assessments in real time are critical to reinforce understanding. Each chapter also includes the **Human Connection.** This special feature highlights a research paper in human genetics that reports a key experiment or raises important social, ethical, or legal issues. Each **Human Connection** has a brief introduction of its own, explaining the importance of the experiment and the context in which it was carried out. At the end of each chapter is a complete **Chapter Summary** in the form of bullet points highlighting the most important concepts.

Each chapter also includes several different types and levels of **Problems**, including concept, synthesis, and discussion questions in the form of **Issues and Ideas**, a guide to problem solving called **Solutions: Step By Step**, and application and analysis problems designated **Concepts In Action**.

LEARNING OBJECTIVES

- To understand how genetic information is stored in the base sequence of DNA. For a given sequence of bases in a transcribed strand of protein-coding DNA, you will specify the sequence of bases in the corresponding region of messenger RNA and the sequence of amino acids in the protein. For a mutation in which a specified base is replaced with another, you will deduce the resulting mRNA and protein sequence.
- To realize that enzymes work in sequence in a metabolic pathway. Given a linear metabolic pathway for an essential nutrient, you will deduce which intermediates will restore the ability to grow mutant strains that are defective for any of the enzymes in the pathway. Conversely, using data that specify which intermediates in a linear metabolic pathway restore the ability of mutants to grow, you will infer the order of the enzymes and intermediates in the pathway.
- To learn that genetic complementation is the operational definition of a gene. Given data on the complementation or lack of complementation among all pairs of a set of mutations affecting a biological process, you will sort the mutations into complementation groups, each corresponding to a different gene.

At the end of the book are **Answers** to even-numbered problems, a complete **Glossary** of key terms in genetics, and a compilation of frequently used **Word Roots** that will help students to understand key genetic terms and make them part of their vocabulary. Answers to odd-numbered problems will be available to instructors.

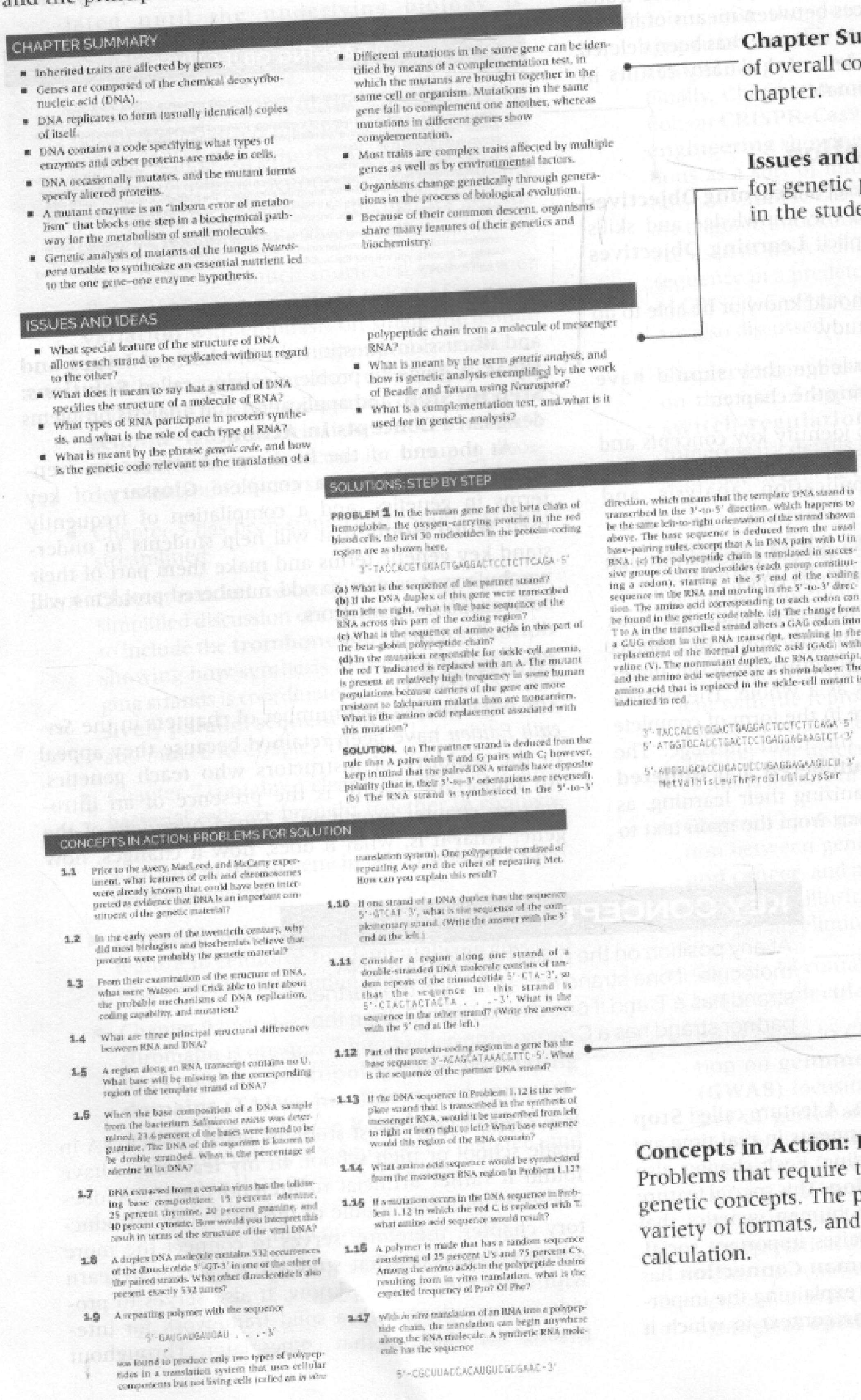
Contents

The organization and number of chapters in the *Seventh Edition* have been retained because they appeal to the majority of instructors who teach genetics. An important feature is the presence of an introductory chapter providing a broad overview of the gene: what it is, what it does, how it changes, how it evolves. Today, most students learn about DNA in grade school or high school. In my teaching, I have found it rather artificial to pretend that DNA does not exist until the middle of the term. The introductory chapter, therefore, serves to connect the more advanced concepts that students are about to learn with what they already know. It also serves to provide each student with a solid framework for integrating the material that comes later. Throughout

KEY CONCEPT

At any position on the paired strands of a DNA molecule, if one strand has an A, then the partner strand has a T; and if one strand has a G, then the partner strand has a C.

each chapter, there is a balance between challenge and motivation, between observation and theory, and between principle and concrete example. Molecular and classical genetics are integrated throughout, and the principles of human genetics are interwoven with the entire fabric of the book. On the other hand, the book is also liberally supplied with examples from animals and plants, especially model organisms. Several points related to organization and coverage should be noted:



CHAPTER SUMMARY

- Inherited traits are affected by genes.
- Genes are composed of the chemical deoxyribonucleic acid (DNA).
- DNA replicates to form (usually identical) copies of itself.
- DNA contains a code specifying what types of enzymes and other proteins are made in cells.
- DNA occasionally mutates, and the mutant forms specify altered proteins.
- A mutant enzyme is an "inborn error of metabolism" that blocks one step in a biochemical pathway for the metabolism of small molecules.
- Genetic analysis of mutants of the fungus *Neurospora* unable to synthesize an essential nutrient led to the one gene–one enzyme hypothesis.
- Different mutations in the same gene can be identified by means of a complementation test, in which the mutants are brought together in the same cell or organism. Mutations in the same gene fail to complement one another, whereas mutations in different genes show complementation.
- Most traits are complex traits affected by multiple genes as well as by environmental factors.
- Organisms change genetically through generations in the process of biological evolution.
- Because of their common descent, organisms share many features of their genetics and biochemistry.

ISSUES AND IDEAS

- What special feature of the structure of DNA allows each strand to be replicated without regard to the other?
- What does it mean to say that a strand of DNA specifies the structure of a molecule of RNA?
- What types of RNA participate in protein synthesis, and what is the role of each type of RNA?
- What is meant by the phrase *genetic code*, and how is the genetic code relevant to the translation of a polypeptide chain from a molecule of messenger RNA?
- What is meant by the term *genetic analysis*, and how is genetic analysis exemplified by the work of Beadle and Tatum using *Neurospora*?
- What is a complementation test, and what is it used for in genetic analysis?

SOLUTIONS: STEP BY STEP

PROBLEM 1 In the human gene for the beta chain of hemoglobin, the oxygen-carrying protein in the red blood cells, the first 30 nucleotides in the protein-coding region are as shown here.

3'-TACCACGTGGACTGAGGACTCCTCTTCAGA-5'

(a) What is the sequence of the partner strand? **(b)** If the DNA duplex of this gene were transcribed from left to right, what is the base sequence of the RNA across this part of the coding region? **(c)** What is the sequence of amino acids in this part of the beta-globin polypeptide chain? **(d)** In the mutation responsible for sickle-cell anemia, the red T indicated is replaced with an A. The mutant is present at relatively high frequency in some human populations because carriers of the gene are more resistant to falciparum malaria than are noncarriers. What is the amino acid replacement associated with this mutation?

SOLUTION. (a) The partner strand is deduced from the rule that A pairs with T and G pairs with C; however, keep in mind that the paired DNA strands have opposite polarity (that is, their 5'-to-3' orientations are reversed). (b) The RNA strand is synthesized in the 5'-to-3' direction, which means that the template DNA strand is transcribed in the 3'-to-5' direction, which happens to be the same left-to-right orientation of the strand shown above. The base sequence is deduced from the usual base-pairing rules, except that A in DNA pairs with U in RNA. (c) The polypeptide chain is translated in successive groups of three nucleotides (each group constituting a codon), starting at the 5' end of the coding sequence in the RNA and moving in the 5'-to-3' direction. The amino acid corresponding to each codon can be found in the genetic code table. (d) The change from T to A in the transcribed strand alters a GAG codon into a GUG codon in the RNA transcript, resulting in the replacement of the normal glutamic acid (GAG) with valine (V). The nonmutant duplex, the RNA transcript, and the amino acid sequence are as shown below. The amino acid that is replaced in the sickle-cell mutant is indicated in red.

3'-TACCACGTGGACTGAGGACTCCTCTTCAGA-5'
5'-ATGGTGCACCTGACTCCTGAGGAGAAGTCT-3'

5'-AUGGUGCACCUGACUCCUGAGGAGAAGUCU-3'
MetValHisLeuThrProGluGluLysSer

CONCEPTS IN ACTION: PROBLEMS FOR SOLUTION

1.1 Prior to the Avery, MacLeod, and McCarty experiment, what features of cells and chromosomes were already known that could have been interpreted as evidence that DNA is an important constituent of the genetic material?

1.2 In the early years of the twentieth century, why did most biologists and biochemists believe that proteins were probably the genetic material?

1.3 From their examination of the structure of DNA, what were Watson and Crick able to infer about the probable mechanisms of DNA replication, coding capability, and mutation?

1.4 What are three principal structural differences between RNA and DNA?

1.5 A region along an RNA transcript contains no U. What base will be missing in the corresponding region of the template strand of DNA?

1.6 When the base composition of a DNA sample from the bacterium *Salinicoccus roseus* was determined, 23.6 percent of the bases were found to be guanine. The DNA of this organism is known to be double stranded. What is the percentage of adenine in its DNA?

1.7 DNA extracted from a certain virus has the following base composition: 15 percent adenine, 25 percent thymine, 20 percent guanine, and 40 percent cytosine. How would you interpret this result in terms of the structure of the viral DNA?

1.8 A duplex DNA molecule contains 532 occurrences of the dinucleotide 5'-GT-3' in one or the other of the paired strands. What other dinucleotide is also present exactly 532 times?

1.9 A repeating polymer with the sequence

5'-GAUGAUGAUGAU . . . -3'

was found to produce only two types of polypeptides in a translation system that uses cellular components but not living cells (called an *in vitro* translation system). One polypeptide consisted of repeating Asp and the other of repeating Met. How can you explain this result?

1.10 If one strand of a DNA duplex has the sequence 5'-GTCAT-3', what is the sequence of the complementary strand. (Write the answer with the 5' end at the left.)

1.11 Consider a region along one strand of a double-stranded DNA molecule consists of tandem repeats of the trinucleotide 5'-CTA-3', so that the sequence in this strand is 5'-CTACTACTACTA . . . -3'. What is the sequence in the other strand? (Write the answer with the 5' end at the left.)

1.12 Part of the protein-coding region in a gene has the base sequence 3'-ACAGCATAAACGTTC-5'. What is the sequence of the partner DNA strand?

1.13 If the DNA sequence in Problem 1.12 is the template strand that is transcribed in the synthesis of messenger RNA, would it be transcribed from left to right or from right to left? What base sequence would this region of the RNA contain?

1.14 What amino acid sequence would be synthesized from the messenger RNA region in Problem 1.12?

1.15 If a mutation occurs in the DNA sequence in Problem 1.12 in which the red C is replaced with T, what amino acid sequence would result?

1.16 A polymer is made that has a random sequence consisting of 25 percent U's and 75 percent C's. Among the amino acids in the polypeptide chains resulting from in vitro translation, what is the expected frequency of Pro? Of Phe?

1.17 With in vitro translation of an RNA into a polypeptide chain, the translation can begin anywhere along the RNA molecule. A synthetic RNA molecule has the sequence

5'-CGCUUACCACAUGUCGCGAAC-3'

Chapter Summary: Summary of overall concepts discussed in chapter.

Issues and Ideas: Questions asking for genetic principles to be restated in the student's own words.

Solutions: Step by Step: A section that demonstrates problems worked in full, explaining step by step a path of logical reasoning that can be followed to analyze the problem.

Concepts in Action: Problems for Solution: Problems that require the student to reason using genetic concepts. The problems make use of a variety of formats, and many require some numerical calculation.

Chapter 1 is an overview of genetics designed to bring students with disparate backgrounds to a common level of understanding. This chapter enables classical genetics, molecular genetics, evolutionary genetics, and genomics to be integrated throughout the rest of the book. Included in Chapter 1 are the basic concepts of genetics: genes as regions of DNA that function through transcription and translation, that change by mutation, and that affect organisms through inborn errors of metabolism. Chapter 1 also explains that most traits are actually complex traits affected by multiple genetic and environmental factors, and it introduces genomics and proteomics.

Chapters 2 through 5 are the core of Mendelian genetics, including segregation and independent assortment, the chromosome theory of heredity, mitosis and meiosis, linkage and chromosome mapping, tetrad analysis in fungi, and chromosome mechanics. An important principle of genetics, too often ignored or given inadequate treatment, is that of the complementation test and how complementation differs from segregation or other genetic principles. Chapter 4 expands on the use of molecular markers in genetics, because these are the principal types of genetic markers in use today.

Chapter 6 deals with DNA, including the details of DNA structure and replication. It also discusses how basic research that revealed the molecular mechanisms of DNA replication ultimately led to such important practical applications as DNA hybridization analysis, DNA sequencing, and the polymerase chain reaction. These examples illustrate the value of basic research in leading, often quite unpredictably, to practical applications.

Chapter 7 deals with the principles of genetics in prokaryotes, beginning with the genetics of mobile DNA, plasmids, and integrons, and their relationships to the evolution of multiple antibiotic resistance. There is a thorough discussion of mechanisms of genetic recombination in microbes, including transformation, conjugation, and transduction, as well as a discussion of temperate and virulent bacteriophages.

Chapters 8 through 12 focus on molecular genetics in the strict sense. Chapter 8 examines the details of gene expression, including transcription, RNA processing, and translation.

Chapter 9 is an integrative chapter that deals with genetic mechanisms of regulation, with examples of mechanisms of gene regulation in prokaryotes as well as eukaryotes. Broader aspects of gene regulation that are topics of much current research, such as higher-order chromatin organization, imprinting, and RNAi are included.

Chapter 10 deals with high-throughput genome sequencing and its implications for personalized medicine and the ethical, legal, and social implications of this technology. It also includes basic methods of recombinant DNA, and there is a major new section of CRISPR/Cas9 in DNA editing and its application to genetic engineering.

Chapter 11 examines the genetic control of development with emphasis on models in *C. elegans*, *D. melanogaster*, and *A. thaliana*.

Chapter 12 focuses on mechanisms of mutation and DNA repair, including chemical mutagens.

Chapter 13 stresses cancer from the standpoint of the genetic control of the cell cycle, with emphasis on the checkpoints that, in normal cells, result either in inhibition of cell division or in programmed cell death (apoptosis). Cancer results from a series of successive mutations, usually in somatic cells, which overcome the normal checkpoints that control cellular proliferation.

Chapters 14 and 15 deal with molecular evolution and population genetics. The discussion includes gene trees and species trees and the population genetics of the CCR5 receptor mutation that confers resistance to infection by HIV. It also includes DNA typing in criminal investigations, paternity testing, the effects of inbreeding, and the evolutionary mechanisms that drive changes in allele frequency. The approach to quantitative genetics includes a discussion of how particular genes influencing quantitative traits (QTLs, or quantitative-trait loci) may be identified and mapped by linkage analysis. There is also a section on what has been learned from genome-wide association studies of complex traits in humans, including the identification of QTLs through genetic mapping or studies of candidate genes.