

HUMAN MOLECULAR GENETICS

GS
Garland Science

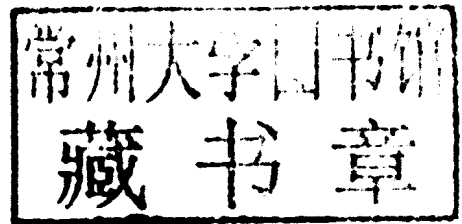
4TH EDITION



TOM STRACHAN AND ANDREW READ

HUMAN MOLECULAR GENETICS

4TH EDITION



TOM STRACHAN AND ANDREW READ



Garland Science

Garland Science

Vice President: Denise Schanck
Senior Editor: Elizabeth Owen
Senior Editorial Assistant: David Borrowdale
Development Editor: Mary Purton
Deputy Production Manager: Simon Hill
Illustrator: Nigel Orme
Typesetting: EJ Publishing Services
Cover Design: Andrew Magee
Copyeditor: Bruce Goatly
Proofreader: Jo Clayton
Indexer: Merrall-Ross International Ltd

Tom Strachan is Scientific Director of the Institute of Human Genetics and Professor of Human Molecular Genetics at Newcastle University, UK, and is a Fellow of the Academy of Medical Sciences and a Fellow of the Royal Society of Edinburgh. Tom's early research interests were in multigene family evolution and interlocus sequence exchange, notably in the HLA and 21-hydroxylase gene clusters. While pursuing the latter, he became interested in medical genetics and disorders of development. His most recent research has focused on developmental control of the cohesin regulators Nipbl and Mau-2.

Tom Strachan and **Andrew Read** were recipients of the European Society of Human Genetics Education Award in 2007.

© 2011 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the copyright holder.

ISBN 978-0-815-34149-9

Library of Congress Cataloging-in-Publication Data

Strachan, T.

Human molecular genetics / Tom Strachan and Andrew Read. -- 4th ed.

p. ; cm.

Rev. ed. of: Human molecular genetics 3 / Tom Strachan and Andrew Read. 3rd ed. c2004.

Includes bibliographical references and indexes.

ISBN 978-0-8153-4149-9

1. Human molecular genetics. I. Read, Andrew P., 1939- II.

Strachan, T. Human molecular genetics 3. III. Title.

[DNLM: 1. Genome, Human. 2. Molecular Biology. QU 470 S894h 2010]

QH431.S787 2010

611'.01816--dc22

2010004605

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 270 Madison Avenue, New York NY 10016, USA, and 2 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

T&F informa

Taylor & Francis Group, an informa business

Visit our website at <http://www.garlandscience.com>

Front Cover Image is adapted from the image created for the 1000 Genomes Project of the National Human Genome Institute, National Institutes of Health, courtesy of Jane Ades.

Back Cover Images Massively parallel pyrosequencing relies on DNA capture beads. Single beads carrying a unique type of single-stranded DNA library fragment are placed in individual wells of a fiber-optic slide into which are deposited smaller beads containing immobilized enzymes needed for pyrophosphate sequencing. Images courtesy of 454 Sequencing © 2010 Roche Diagnostics.

Andrew Read is Emeritus Professor of Human Genetics at Manchester University, UK and a Fellow of the Academy of Medical Sciences. Andrew has been particularly concerned with making the benefits of DNA technology available to people with genetic problems. He established one of the first DNA diagnostic laboratories in the UK over 20 years ago (it is now one of two National Genetics Reference Laboratories), and was founder chairman of the British Society for Human Genetics, the main professional body in this area. His own research is on the molecular pathology of various hereditary syndromes, especially hereditary hearing loss.



Mixed Sources

Product group from well-managed forests, controlled sources and recycled wood or fiber

www.fsc.org Cert no. SW-COC-002985

©1996 Forest Stewardship Council

Preface

The pace of scientific and technical advance in human genetics has not slackened since our third edition appeared in 2004. This has mandated a thorough revision and reorganization of *Human Molecular Genetics* with much of the text being completely rewritten. While only a few of the basic introductory chapters retain their identity from the third edition, the aims of the text remain the same: to provide a framework of principles rather than a list of facts, to provide a bridge between basic textbooks and the research literature, and to communicate our continuing excitement about this very fast-moving area of science.

The ‘finished’ human genome reference sequence was published in 2004 and we are now entering an era where vast DNA sequence datasets will be produced annually. The game changer is the advent of massively parallel DNA sequencing which is already transforming how we approach genetics. Single molecule sequencing will lead to a dramatic reduction in DNA sequencing costs and promises the ability to sequence a human genome in hours. We can confidently expect that the genomes of huge numbers of organisms and individuals will have been completed before the next edition of this book.

Powerful bioinformatics programs are already being pressed into service to compare our genome with that of a burgeoning number of other organisms. Comparative genomics is helping us understand the forces that have shaped the evolution of our genome and that of many model organisms that are so important to research and various biomedical applications. These studies have already been extremely helpful in defining the most highly conserved and presumably important parts of our genome. They are also helping us to identify the fastest changing components of our genome and what it is that makes us unique.

Sequence-based transcriptomic analysis will become a major industry. It will be an important player in our quest to understand human gene function within the context of large projects, such as the ENCODE project that aims to create an encyclopedia of DNA elements of known function. Eventually, as vast datasets are accumulated on gene function, the stage will truly be set for systems biology to develop.

Other large scale projects such as HapMap have been exploring the range of genetic variation across the world’s populations. In disease-related research, genome-wide screening for copy number variants has identified the problems affecting many individual patients and led to the delineation of new microdeletion and microduplication syndromes. Whole exome sequencing is now poised to explain the causes of many rare recessive conditions. In cancer, the first full genome sequences of tumors are starting to reveal the landscape of carcinogenesis in unprecedented detail.

For common complex diseases, however, the picture is less pleasing. A combination of new science (HapMap) and new technology (high-throughput SNP genotyping) has finally allowed researchers to identify genetic susceptibility factors for common diseases, but it has become apparent that the variants revealed by genome-wide association studies explain only a very small part of the overall genetic susceptibility to most complex diseases. We are left with a problem—where is the hidden heritability? Will it be found by large-scale resequencing, or perhaps might it lie in epigenetic effects?

All these developments have affected both the way genetics research is done, and the way we think about our genome. Genetics is more than ever about processing and collating vast amounts of private and public data to extract meaningful patterns. The data have also been forcing us to revise some of our basic ideas about human genetics. Humans are more variable than we thought, with copy number variants accounting for more variable nucleotides than SNPs. We transcribe almost all of our genome, and the old picture of discrete genes thinly scattered across a sea of junk DNA is starting to look untenable. Cells are now known to be awash with a startling variety of noncoding RNAs of unknown function. Perhaps our genome might be primarily an RNA, rather than a protein, machine.

The fourth edition of *Human Molecular Genetics* has therefore been heavily updated in order to maintain its hallmark currency and continue to provide a framework for understanding this exciting and rapidly advancing subject. Coverage of epigenetics, noncoding RNAs, and cell biology, including stem cells, have all been expanded. Greater detail has been provided on the major animal models used in genetic studies and how they are used as models for human disease. The most recent developments in next generation sequencing and comparative genomics have been included. The text closes by looking at the development of therapies to treat human disease. Genetic testing and screening, stem cells and cell therapy, and personalized medicine are all discussed together with a balanced view of the ethical issues surrounding these issues.

We would like to thank the staff at Garland Science who have undertaken the job of converting our drafts into the finished product, Elizabeth Owen, Mary Purton, David Borrowdale, and Simon Hill, and hope readers will appreciate all the work they have put into this. As ever we are grateful to our respective families for their forbearance and support.

TEACHING RESOURCES

The images from the book are downloadable from the web in JPEG and Powerpoint® formats via the Classwire™ course management system. The system also provides access to instructional resources for other Garland Science books. In addition to serving as an online archive of electronic teaching resources, Classwire allows instructors to build customized websites for their classes. Please visit www.classwire.com/garlandscience or email science@garland.com for further information. (Classwire is a trademark of Chalkfree, Inc.)

Acknowledgments

In writing this book we have benefited greatly from the advice of many geneticists and biologists. We are grateful to many colleagues at Newcastle and Manchester Universities who commented on some aspects of the text notably: C. Brooks, K. Bushby, A. Knight, M. Lako, H. Middleton-Price, S. Ramsden, G. Saretzki, N. Thakker, and A. Wallace. We are also appreciative of help and data provided by various staff members at the European Bioinformatics Institute notably: Xose Fernandez, Javier Herrero, Julio Fernandez Banet, Simon White, and Ewan Birney. In addition, we would like to thank the following for their suggestions in preparing this edition.

Alexandra I.F. Blakemore (Imperial College London, UK); Daniel A. Brazeau (University at Buffalo, USA); Carolyn J. Brown (University of British Columbia, Canada); Frederic Chedin (University of California Davis, USA); Edwin Cuppen (University Medical Center, Utrecht, The Netherlands); Ken Dewar (McGill University, Canada); Ian Dunham (European Bioinformatics Institute, UK); T. Mary Fujiwara (McGill University, Canada); Adrian J. Hall (Sheffield Hallam University, UK); Lise Lotte Hansen (Aarhus University, Denmark); Verle Headings (Howard University, USA); Graham Heap (Barts and The London School of Medicine and Dentistry, UK); Mary O. Huff (Bellarmine University, USA); David L. Hurley (East Tennessee State University, USA); David Iles (University of Leeds, UK); Colin A. Johnson (University of Leeds, UK); Bobby P.C. Koeleman (University Medical Center, Utrecht, The Netherlands); Michael R. Lodomery (University of the West of England, UK); Dick Lindhout (University Medical Center, Utrecht, The Netherlands); John Loughlin (Newcastle University, UK); Donald Macleod (University of Edinburgh, UK); Eleanor M. Maine (Syracuse University, USA); Rhayza Maingon (Keele University, UK); Gudrun Moore (Institute of Child Health, UK); Tom Moore (University College Cork, Ireland); Kenneth Morgan (McGill University, Canada); Karen E. Morrison (University of Birmingham, UK); Brenda Murphy (University of Western Ontario, Canada); Roberta Palmour (McGill University, Canada); Nollaig Parfrey (University College Cork, Ireland); Aimee K. Ryan (McGill University, Canada); Jennifer Sanders (Brown University, USA); Sharon Shriver (Pennsylvania State University, USA); John J. Taylor (Newcastle University, UK); Leo P. ten Kate (VU University Medical Center, The Netherlands); Jürgen Tomiuk (University of Tübingen, Germany); Patricia N. Tonin (McGill University, Canada); David A. van Heel (Barts and The London School of Medicine and Dentistry, UK); Malcolm von Schantz (University of Surrey, UK).

Contents

| | | |
|-------------------|--|-----|
| Chapter 1 | Nucleic Acid Structure and Gene Expression | 1 |
| Chapter 2 | Chromosome Structure and Function | 29 |
| Chapter 3 | Genes in Pedigrees and Populations | 61 |
| Chapter 4 | Cells and Cell–Cell Communication | 91 |
| Chapter 5 | Principles of Development | 133 |
| Chapter 6 | Amplifying DNA: Cell-based DNA Cloning and PCR | 163 |
| Chapter 7 | Nucleic Acid Hybridization: Principles and Applications | 191 |
| Chapter 8 | Analyzing the Structure and Expression of Genes and Genomes | 213 |
| Chapter 9 | Organization of the Human Genome | 255 |
| Chapter 10 | Model Organisms, Comparative Genomics, and Evolution | 297 |
| Chapter 11 | Human Gene Expression | 345 |
| Chapter 12 | Studying Gene Function in the Post-Genome Era | 381 |
| Chapter 13 | Human Genetic Variability and Its Consequences | 405 |
| Chapter 14 | Genetic Mapping of Mendelian Characters | 441 |
| Chapter 15 | Mapping Genes Conferring Susceptibility to Complex Diseases | 467 |
| Chapter 16 | Identifying Human Disease Genes and Susceptibility Factors | 497 |
| Chapter 17 | Cancer Genetics | 537 |
| Chapter 18 | Genetic Testing of Individuals | 569 |
| Chapter 19 | Pharmacogenetics, Personalized Medicine, and Population Screening | 605 |
| Chapter 20 | Genetic Manipulation of Animals for Modeling Disease and Investigating Gene Function | 639 |
| Chapter 21 | Genetic Approaches to Treating Disease | 677 |
| | Glossary | 719 |
| | Index | 737 |

Detailed Contents

Chapter 1

Nucleic Acid Structure and Gene Expression 1

1.1 DNA, RNA, AND POLYPEPTIDES 2

Most genetic information flows in the sequence

DNA → RNA → polypeptide 2

Nucleic acids and polypeptides are linear sequences of simple repeat units 3

Nucleic acids 3

Polypeptides 4

The type of chemical bonding determines stability and function 6

1.2 NUCLEIC ACID STRUCTURE AND DNA REPLICATION 7

DNA and RNA structure 7

Replication is semi-conservative and semi-discontinuous 9

DNA polymerases sometimes work in DNA repair and recombination 9

Many viruses have RNA genomes 11

1.3 RNA TRANSCRIPTION AND GENE EXPRESSION 12

Most genes are expressed to make polypeptides 14

Different sets of RNA genes are transcribed by the three eukaryotic RNA polymerases 15

1.4 RNA PROCESSING 16

RNA splicing removes unwanted sequences from the primary transcript 16

Specialized nucleotides are added to the ends of most RNA polymerase II transcripts 17

5' capping 18

3' polyadenylation 18

rRNA and tRNA transcripts undergo extensive processing 19

1.5 TRANSLATION, POST-TRANSLATIONAL PROCESSING, AND PROTEIN STRUCTURE 20

mRNA is decoded to specify polypeptides 20

The genetic code is degenerate and not quite universal 22

Post-translational processing: chemical modification of amino acids and polypeptide cleavage 23

Addition of carbohydrate groups 23

Addition of lipid groups 24

Post-translational cleavage 24

The complex relationship between amino acid sequence and protein structure 25

The α -helix 25

The β -pleated sheet 26

The β -turn 27

Higher-order structures 27

FURTHER READING 27

Chapter 2

Chromosome Structure and Function 29

2.1 PLOIDY AND THE CELL CYCLE 30

2.2 MITOSIS AND MEIOSIS 31

Mitosis is the normal form of cell division 31

Meiosis is a specialized reductive cell division that gives rise to sperm and egg cells 33

Independent assortment 34

Recombination 34

X-Y pairing 35

Mitosis and meiosis have key similarities and differences 36

2.3 STRUCTURE AND FUNCTION OF CHROMOSOMES 36

Chromosomal DNA is coiled hierarchically 36

Interphase chromatin varies in its degree of compaction 38

Each chromosome has its own territory in the interphase nucleus 38

Centromeres have a pivotal role in chromosome movement but have evolved to be very different in different organisms 39

Replication of a mammalian chromosome involves the flexible use of multiple replication origins 40

Telomeres have specialized structures to preserve the ends of linear chromosomes 41

Telomere structure, function, and evolution 41

Telomerase and the chromosome end-replication problem 42

2.4 STUDYING HUMAN CHROMOSOMES 43

Chromosome analysis is easier for mitosis than meiosis 44

Chromosomes are identified by size and staining pattern 44

Chromosome banding 44

| | | | |
|--|----|---|-----|
| Reporting of cytogenetic analyses | 45 | Mosaics | 77 |
| Molecular cytogenetics locates specific DNA sequences on chromosomes | 46 | Chimeras | 78 |
| Chromosome fluorescence <i>in situ</i> hybridization (FISH) | 47 | 3.4 GENETICS OF MULTIFACTORIAL CHARACTERS: THE POLYGENIC THRESHOLD THEORY | 79 |
| Chromosome painting and molecular karyotyping | 47 | In the early twentieth century there was controversy between proponents of Mendelian and quantitative models of inheritance | 79 |
| Comparative genome hybridization (CGH) | 48 | Polygenic theory explains how quantitative traits can be genetically determined | 79 |
| 2.5 CHROMOSOME ABNORMALITIES | 50 | Regression to the mean | 80 |
| Numerical chromosomal abnormalities involve gain or loss of complete chromosomes | 50 | Hidden assumptions | 81 |
| Polyploidy | 50 | Heritability measures the proportion of the overall variance of a character that is due to genetic differences | 81 |
| Aneuploidy | 50 | Misunderstanding heritability | 82 |
| Mixoploidy | 51 | The threshold model extended polygenic theory to cover dichotomous characters | 82 |
| Clinical consequences | 52 | Using threshold theory to understand recurrence risks | 82 |
| A variety of structural chromosomal abnormalities result from misrepair or recombination errors | 53 | Counseling in non-Mendelian conditions is based on empiric risks | 83 |
| Different factors contribute to the clinical consequences of structural chromosome abnormalities | 55 | 3.5 FACTORS AFFECTING GENE FREQUENCIES | 84 |
| Incorrect parental origins of chromosomes can result in aberrant development and disease | 56 | A thought experiment: picking genes from the gene pool | 84 |
| CONCLUSION | 58 | The Hardy–Weinberg distribution relates genotype frequencies to gene frequencies | 84 |
| FURTHER READING | 59 | Using the Hardy–Weinberg relationship in genetic counseling | 84 |
| Chapter 3 | | Inbreeding | 85 |
| Genes in Pedigrees and Populations | 61 | Other causes of departures from the Hardy–Weinberg relationship | 86 |
| 3.1 MONOGENIC VERSUS MULTIFACTORIAL INHERITANCE | 62 | Gene frequencies can vary with time | 87 |
| 3.2 MENDELIAN PEDIGREE PATTERNS | 64 | Estimating mutation rates | 88 |
| There are five basic Mendelian pedigree patterns | 64 | The importance of heterozygote advantage | 88 |
| X-inactivation | 65 | CONCLUSION | 89 |
| Mosaicism due to X-inactivation | 66 | FURTHER READING | 89 |
| Few genes on the Y chromosome | 67 | | |
| Genes in the pseudoautosomal region | 67 | Chapter 4 | |
| Conditions caused by mutations in the mitochondrial DNA | 68 | Cells and Cell–Cell Communication | 91 |
| The mode of inheritance can rarely be defined unambiguously in a single pedigree | 69 | 4.1 CELL STRUCTURE AND DIVERSITY | 92 |
| Getting the right ratios: the problem of bias of ascertainment | 69 | Prokaryotes and eukaryotes represent a fundamental division of cellular life forms | 92 |
| The relation between Mendelian characters and gene sequences | 70 | The extraordinary diversity of cells in the body | 93 |
| Locus heterogeneity | 71 | Germ cells are specialized for reproductive functions | 93 |
| Clinical heterogeneity | 72 | Cells in an individual multicellular organism can differ in DNA content | 96 |
| 3.3 COMPLICATIONS TO THE BASIC MENDELIAN PEDIGREE PATTERNS | 72 | 4.2 CELL ADHESION AND TISSUE FORMATION | 97 |
| A common recessive condition can mimic a dominant pedigree pattern | 72 | Cell junctions regulate the contact between cells | 98 |
| A dominant condition may fail to manifest itself | 73 | Tight junctions | 98 |
| Age-related penetrance in late-onset diseases | 73 | Anchoring cell junctions | 99 |
| Many conditions show variable expression | 74 | Communicating cell junctions | 99 |
| Anticipation | 74 | The extracellular matrix regulates cell behavior as well as acting as a scaffold to support tissues | 99 |
| Imprinting | 75 | Specialized cell types are organized into tissues | 100 |
| Male lethality may complicate X-linked pedigrees | 76 | Epithelium | 100 |
| Inbreeding can complicate pedigree interpretation | 76 | Connective tissue | 101 |
| New mutations and mosaicism complicate pedigree interpretation | 76 | Muscle tissue | 101 |
| | | Nervous tissue | 101 |

| | | | |
|---|------------|---|------------|
| 4.3 PRINCIPLES OF CELL SIGNALING | 102 | Additional recombination and mutation mechanisms contribute to receptor diversity in B cells, but not T cells | 130 |
| Signaling molecules bind to specific receptors in responding cells to trigger altered cell behavior | 102 | The monospecificity of Igs and TCRs is due to allelic exclusion and light chain exclusion | 131 |
| Some signaling molecules bind intracellular receptors that activate target genes directly | 103 | FURTHER READING | 132 |
| Signaling through cell surface receptors often involves kinase cascades | 105 | | |
| Signal transduction pathways often use small intermediate intracellular signaling molecules | 106 | Chapter 5 | |
| Synaptic signaling is a specialized form of cell signaling that does not require the activation of transcription factors | 107 | Principles of Development | 133 |
| 4.4 CELL PROLIFERATION, SENESCENCE, AND PROGRAMMED CELL DEATH | 108 | 5.1 AN OVERVIEW OF DEVELOPMENT | 134 |
| Most of the cells in mature animals are non-dividing cells, but some tissues and cells turn over rapidly | 108 | Animal models of development | 135 |
| Mitogens promote cell proliferation by overcoming braking mechanisms that restrain cell cycle progression in G ₁ | 109 | 5.2 CELL SPECIALIZATION DURING DEVELOPMENT | 136 |
| Cell proliferation limits and the concept of cell senescence | 111 | Cells become specialized through an irreversible series of hierarchical decisions | 136 |
| Large numbers of our cells are naturally programmed to die | 111 | The choice between alternative cell fates often depends on cell position | 137 |
| The importance of programmed cell death | 112 | Sometimes cell fate can be specified by lineage rather than position | 138 |
| Apoptosis is performed by caspases in response to death signals or sustained cell stress | 113 | 5.3 PATTERN FORMATION IN DEVELOPMENT | 139 |
| Extrinsic pathways: signaling through cell surface death receptors | 113 | Emergence of the body plan is dependent on axis specification and polarization | 139 |
| Intrinsic pathways: intracellular responses to cell stress | 113 | Pattern formation often depends on gradients of signaling molecules | 141 |
| 4.5 STEM CELLS AND DIFFERENTIATION | 114 | Homeotic mutations reveal the molecular basis of positional identity | 142 |
| Cell specialization involves a directed series of hierarchical decisions | 114 | 5.4 MORPHOGENESIS | 144 |
| Stem cells are rare self-renewing progenitor cells | 115 | Morphogenesis can be driven by changes in cell shape and size | 144 |
| Tissue stem cells allow specific adult tissues to be replenished | 115 | Major morphogenetic changes in the embryo result from changes in cell affinity | 145 |
| Stem cell niches | 117 | Cell proliferation and apoptosis are important morphogenetic mechanisms | 145 |
| Stem cell renewal versus differentiation | 117 | 5.5 EARLY HUMAN DEVELOPMENT: FERTILIZATION TO GASTRULATION | 146 |
| Embryonic stem cells and embryonic germ cells are pluripotent | 117 | During fertilization the egg is activated to form a unique individual | 146 |
| Origins of cultured embryonic stem cells | 118 | Cleavage partitions the zygote into many smaller cells | 147 |
| Pluripotency tests | 119 | Mammalian eggs are among the smallest in the animal kingdom, and cleavage in mammals is exceptional in several ways | 148 |
| Embryonic germ cells | 119 | Only a small percentage of the cells in the early mammalian embryo give rise to the mature organism | 148 |
| 4.6 IMMUNE SYSTEM CELLS: FUNCTION THROUGH DIVERSITY | 119 | Implantation | 150 |
| The innate immune system provides a rapid response based on general pattern recognition of pathogens | 121 | Gastrulation is a dynamic process by which cells of the epiblast give rise to the three germ layers | 151 |
| The adaptive immune system mounts highly specific immune responses that are enhanced by memory cells | 122 | 5.6 NEURAL DEVELOPMENT | 154 |
| Humoral immunity depends on the activities of soluble antibodies | 123 | The axial mesoderm induces the overlying ectoderm to develop into the nervous system | 155 |
| In cell-mediated immunity, T cells recognize cells containing fragments of foreign proteins | 125 | Pattern formation in the neural tube involves the coordinated expression of genes along two axes | 155 |
| T-cell activation | 127 | Neuronal differentiation involves the combinatorial activity of transcription factors | 157 |
| The unique organization and expression of Ig and TCR genes | 128 | | |

| | |
|--|-----|
| 5.7 GERM-CELL AND SEX DETERMINATION IN MAMMALS | 158 |
| Primordial germ cells are induced in the early mammalian embryo and migrate to the developing gonads | 158 |
| Sex determination involves both intrinsic and positional information | 159 |
| 5.8 CONSERVATION OF DEVELOPMENTAL PATHWAYS | 160 |
| Many human diseases are caused by the failure of normal developmental processes | 160 |
| Developmental processes are often highly conserved but some show considerable species differences | 160 |
| FURTHER READING | 161 |

Chapter 6

Amplifying DNA: Cell-based DNA Cloning and PCR 163

| | |
|--|-----|
| Cell-based DNA cloning | 165 |
| The polymerase chain reaction (PCR) | 165 |
| 6.1 PRINCIPLES OF CELL-BASED DNA CLONING | 165 |
| Manageable pieces of target DNA are joined to vector molecules by using restriction endonucleases and DNA ligase | 166 |
| Basic DNA cloning in bacterial cells uses vectors based on naturally occurring extrachromosomal replicons | 168 |
| Cloning in bacterial cells uses genetically modified plasmid or bacteriophage vectors and modified host cells | 169 |
| Transformation is the key DNA fractionation step in cell-based DNA cloning | 171 |
| Recombinant DNA can be selectively purified after screening and amplifying cell clones with desired target DNA fragments | 172 |
| 6.2 LARGE INSERT CLONING AND CLONING SYSTEMS FOR PRODUCING SINGLE-STRANDED DNA | 172 |
| Early large insert cloning vectors exploited properties of bacteriophage λ | 173 |
| Large DNA fragments can be cloned in bacterial cells by using extrachromosomal low-copy-number replicons | 174 |
| Bacterial artificial chromosome (BAC) and fosmid vectors | 174 |
| Bacteriophage P1 vectors and P1 artificial chromosomes | 175 |
| Yeast artificial chromosomes (YACs) enable cloning of megabase fragments of DNA | 175 |
| Producing single-stranded DNA for use in DNA sequencing and <i>in vitro</i> site-specific mutagenesis | 176 |
| M13 vectors | 176 |
| Phagemid vectors | 177 |
| 6.3 CLONING SYSTEMS DESIGNED FOR GENE EXPRESSION | 178 |
| Large amounts of protein can be produced by expression cloning in bacterial cells | 178 |

| | |
|---|-----|
| In phage display, heterologous proteins are expressed on the surface of phage particles | 179 |
| Eukaryotic gene expression is performed with greater fidelity in eukaryotic cell lines | 180 |
| Transient expression in insect cells by using baculovirus | 181 |
| Transient expression in mammalian cells | 181 |
| Stable expression in mammalian cells | 181 |
| 6.4 CLONING DNA <i>IN VITRO</i> : THE POLYMERASE CHAIN REACTION | 182 |
| PCR can be used to amplify a rare target DNA selectively from within a complex DNA population | 183 |
| The cyclical nature of PCR leads to exponential amplification of the target DNA | 183 |
| Selective amplification of target sequences depends on highly specific binding of primer sequences | 184 |
| PCR is disadvantaged as a DNA cloning method by short lengths and comparatively low yields of product | 185 |
| A wide variety of PCR approaches have been developed for specific applications | 186 |
| Allele-specific PCR | 187 |
| Multiple target amplification and whole genome PCR methods | 187 |
| PCR mutagenesis | 188 |
| Real-time PCR (qPCR) | 189 |
| FURTHER READING | 189 |

Chapter 7

Nucleic Acid Hybridization: Principles and Applications 191

| | |
|--|-----|
| 7.1 PRINCIPLES OF NUCLEIC ACID HYBRIDIZATION | 192 |
| In nucleic acid hybridization a known nucleic acid population interrogates an imperfectly understood nucleic acid population | 192 |
| Probe–target heteroduplexes are easier to identify after capture on a solid support | 193 |
| Denaturation and annealing are affected by temperature, chemical environment, and the extent of hydrogen bonding | 194 |
| Stringent hybridization conditions increase the specificity of duplex formation | 195 |
| The kinetics of DNA reassociation is also dependent on the concentration of DNA | 196 |
| 7.2 LABELING OF NUCLEIC ACIDS AND OLIGONUCLEOTIDES | 197 |
| Different classes of hybridization probe can be prepared from DNA, RNA, and oligonucleotide substrates | 197 |
| Long nucleic acid probes are usually labeled by incorporating labeled nucleotides during strand synthesis | 197 |
| Labeling DNA by nick translation | 198 |
| Random primed DNA labeling | 198 |
| PCR-based strand synthesis labeling | 198 |
| RNA labeling | 198 |

| | | | |
|--|-----|--|-----|
| Radioisotopes can be used to label nucleic acids but are short-lived and can be hazardous | 199 | Single-molecule sequencing | 220 |
| Fluorophores are commonly used in nonisotopic labeling of nucleic acids | 200 | Mircoarray-based DNA capture methods allow efficient resequencing | 223 |
| 7.3 HYBRIDIZATION TO IMMOBILIZED TARGET NUCLEIC ACIDS | 203 | 8.3 GENOME STRUCTURE ANALYSIS AND GENOME PROJECTS | 224 |
| Dot-blot hybridization offers rapid screening and often employs allele-specific oligonucleotide probes | 203 | Framework maps are needed for first time sequencing of complex genomes | 225 |
| Southern and northern blot hybridizations detect size-fractionated DNA and RNA | 204 | The linear order of genomic DNA clones in a contig matches their original subchromosomal locations | 226 |
| Southern blot hybridization | 204 | The Human Genome Project was an international endeavor and biology's first Big Project | 228 |
| Northern blot hybridization | 204 | The first human genetic maps were of low resolution and were constructed with mostly anonymous DNA markers | 228 |
| In an <i>in situ</i> hybridization test, sample DNA or RNA is immobilized within fixed chromosome or cell preparations | 205 | Physical maps of the human genome progressed from marker maps to clone contig maps | 230 |
| Chromosome <i>in situ</i> hybridization | 205 | The final sequencing phase of the Human Genome Project was a race to an early finish | 231 |
| Tissue <i>in situ</i> hybridization | 206 | Genome projects have also been conducted for a variety of model organisms | 234 |
| Hybridization can be used to screen bacterial colonies containing recombinant DNA | 206 | Powerful genome databases and browsers help to store and analyze genome data | 235 |
| 7.4 MICROARRAY-BASED HYBRIDIZATION ASSAYS | 207 | Different computer programs are designed to predict and annotate genes within genome sequences | 235 |
| Microarray hybridization allows highly parallel hybridization assays using thousands of different immobilized probes | 207 | Obtaining accurate estimates for the number of human genes is surprisingly difficult | 238 |
| High-density oligonucleotide microarrays offer enormously powerful tools for analyzing complex RNA and DNA samples | 208 | 8.4 BASIC GENE EXPRESSION ANALYSES | 239 |
| Affymetrix oligonucleotide microarrays | 209 | Principles of expression screening | 239 |
| Illumina oligonucleotide microarrays | 209 | Hybridization-based methods allow semi-quantitative and high-resolution screening of transcripts of individual genes | 240 |
| Microarray hybridization is used mostly in transcript profiling or assaying DNA variation | 209 | Hybridization-based methods for assaying transcript size and abundance | 241 |
| FURTHER READING | 211 | Tissue <i>in situ</i> hybridization | 241 |
| Chapter 8 | | Quantitative PCR methods are widely used for expression screening | 241 |
| Analyzing the Structure and Expression of Genes and Genomes | 213 | Specific antibodies can be used to track proteins expressed by individual genes | 242 |
| 8.1 DNA LIBRARIES | 214 | Protein expression in cultured cells is often analyzed by using different types of fluorescence microscopy | 244 |
| Genomic DNA libraries comprise fragmented copies of all the different DNA molecules in a cell | 214 | 8.5 HIGHLY PARALLEL ANALYSES OF GENE EXPRESSION | 245 |
| cDNA libraries comprise DNA copies of the different RNA molecules in a cell | 214 | DNA and oligonucleotide microarrays permit rapid global transcript profiling | 245 |
| To be useful, DNA libraries need to be conveniently screened and disseminated | 215 | Modern global gene expression profiling increasingly uses sequencing to quantitate transcripts | 248 |
| Library screening | 215 | Global protein expression is often profiled with two-dimensional gel electrophoresis and mass spectrometry | 249 |
| Library amplification and dissemination | 216 | Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) | 249 |
| 8.2 SEQUENCING DNA | 217 | Mass spectrometry | 250 |
| Dideoxy DNA sequencing involves enzymatic DNA synthesis using base-specific chain terminators | 217 | Comparative protein expression analyses have many applications | 252 |
| Automation of dideoxy DNA sequencing increased its efficiency | 218 | FURTHER READING | 253 |
| Iterative pyrosequencing records DNA sequences while DNA molecules are being synthesized | 219 | | |
| Massively parallel DNA sequencing enables the simultaneous sequencing of huge numbers of different DNA fragments | 220 | | |
| Massively parallel sequencing of amplified DNA | 220 | | |

Chapter 9**Organization of the Human Genome 255****9.1 GENERAL ORGANIZATION OF THE HUMAN GENOME 257**

| | |
|--|-----|
| The mitochondrial genome is densely packed with genetic information | 257 |
| Replication of mitochondrial DNA | 257 |
| Mitochondrial genes and their transcription | 257 |
| The mitochondrial genetic code | 258 |
| The human nuclear genome consists of 24 widely different chromosomal DNA molecules | 260 |
| The human genome contains at least 26,000 genes, but the exact gene number is difficult to determine | 261 |

| | |
|---|-----|
| Human genes are unevenly distributed between and within chromosomes | 263 |
| Duplication of DNA segments has resulted in copy-number variation and gene families | 263 |
| Gene duplication mechanisms | 263 |

9.2 PROTEIN-CODING GENES 265

| | |
|--|-----|
| Human protein-coding genes show enormous variation in size and internal organization | 265 |
| Size variation | 265 |
| Repetitive sequences within coding DNA | 266 |
| Different proteins can be specified by overlapping transcription units | 266 |
| Overlapping genes and genes-within-genes | 266 |
| Genes divergently transcribed or co-transcribed from a common promoter | 266 |

| | |
|---|-----|
| Human protein-coding genes often belong to families of genes that may be clustered or dispersed on multiple chromosomes | 268 |
|---|-----|

| | |
|--|-----|
| Different classes of gene family can be recognized according to the extent of sequence and structural similarity of the protein products | 269 |
|--|-----|

| | |
|---|-----|
| Gene duplication events that give rise to multigene families also create pseudogenes and gene fragments | 271 |
|---|-----|

9.3 RNA GENES 274

| | |
|--|-----|
| More than 1000 human genes encode an rRNA or tRNA, mostly within large gene clusters | 277 |
| Ribosomal RNA genes | 277 |
| Transfer RNA genes | 279 |

| | |
|---|-----|
| Dispersed gene families make various small nuclear RNAs that facilitate general gene expression | 280 |
| Spliceosomal small nuclear RNA (snRNA) genes | 281 |
| Non-spliceosomal small nuclear RNA genes | 282 |
| Small nucleolar RNA (snoRNA) genes | 282 |
| Small Cajal body RNA genes | 283 |

| | |
|--|-----|
| Close to 1000 different human microRNAs regulate complex sets of target genes by base pairing to the RNA transcripts | 283 |
|--|-----|

| | |
|--|-----|
| Many thousands of different piRNAs and endogenous siRNAs suppress transposition and regulate gene expression | 285 |
| Piwi-protein-interacting RNA | 285 |
| Endogenous siRNAs | 286 |

| | |
|---|-----|
| More than 3000 human genes synthesize a wide variety of medium-sized to large regulatory RNAs | 287 |
|---|-----|

9.4 HIGHLY REPETITIVE DNA: HETEROCHROMATIN AND TRANSPOSON REPEATS 289

| | |
|---|-----|
| Constitutive heterochromatin is largely defined by long arrays of high-copy-number tandem DNA repeats | 289 |
|---|-----|

| | |
|---|-----|
| Transposon-derived repeats make up more than 40% of the human genome and arose mostly through RNA intermediates | 290 |
|---|-----|

| | |
|------------------------------|-----|
| Human LTR transposons | 291 |
| Human DNA transposon fossils | 291 |

| | |
|--|-----|
| A few human LINE-1 elements are active transposons and enable the transposition of other types of DNA sequence | 291 |
|--|-----|

| | |
|--|-----|
| Alu repeats are the most numerous human DNA elements and originated as copies of 7SL RNA | 292 |
|--|-----|

CONCLUSION 293**FURTHER READING 294****Chapter 10****Model Organisms, Comparative Genomics, and Evolution 297****10.1 MODEL ORGANISMS 298**

| | |
|---|-----|
| Unicellular model organisms aid understanding of basic cell biology and microbial pathogens | 298 |
|---|-----|

| | |
|---|-----|
| Some invertebrate models offer cheap high-throughput genetic screening, and can sometimes model disease | 301 |
|---|-----|

| | |
|--|-----|
| Various fish, frog, and bird models offer accessible routes to the study of vertebrate development | 301 |
|--|-----|

| | |
|--|-----|
| Mammalian models are disadvantaged by practical limitations and ethical concerns | 303 |
|--|-----|

| | |
|--|-----|
| Humans are the ultimate model organism and will probably be a principal one some time soon | 305 |
|--|-----|

10.2 COMPARATIVE GENOMICS 306

| | |
|---|-----|
| Evolutionary constraint and the preservation of function by purifying selection | 306 |
|---|-----|

| | |
|---|-----|
| Rapidly evolving sequences and positive selection | 308 |
|---|-----|

| | |
|---|-----|
| A variety of computer programs allow automated genome sequence alignments | 308 |
|---|-----|

| | |
|--|-----|
| Comparative genomics helps validate predicted genes and identifies novel genes | 309 |
|--|-----|

| | |
|---|-----|
| Comparative genomics reveals a surprisingly large amount of functional noncoding DNA in mammals | 310 |
|---|-----|

| | |
|--|-----|
| Comparative genomics has been particularly important in identifying regulatory sequences | 312 |
|--|-----|

10.3 GENE AND GENOME EVOLUTION 313

| | |
|---|-----|
| Gene complexity is increased by exon duplication and exon shuffling | 314 |
|---|-----|

| | |
|------------------|-----|
| Exon duplication | 314 |
| Exon shuffling | 315 |

| | |
|--|-----|
| Gene duplication can permit increased gene dosage but its major value is to permit functional complexity | 315 |
|--|-----|

| | | | |
|---|-----|---|-----|
| The globin superfamily illustrates divergence in gene regulation and function after gene duplication | 317 | Elongation of the transcript | 349 |
| Two or three major whole genome duplication events have occurred in vertebrate lineages since the split from tunicates | 318 | Termination of transcription | 349 |
| Major chromosome rearrangements have occurred during mammalian genome evolution | 319 | Many other proteins modulate the activity of the basal transcription apparatus | 349 |
| In heteromorphic sex chromosomes, the smaller chromosome is limited to one sex and is mostly non-recombining with few genes | 320 | Sequence-specific DNA-binding proteins can bind close to a promoter or at more remote locations | 350 |
| The pseudoautosomal regions have changed rapidly during evolution | 322 | Co-activators and co-repressors influence promoters without binding to DNA | 352 |
| Human sex chromosomes evolved after a sex-determining locus developed on one autosome, causing it to diverge from its homolog | 323 | 11.2 CHROMATIN CONFORMATION: DNA METHYLATION AND THE HISTONE CODE | 353 |
| Abundant testis-expressed genes on the Y chromosome are mostly maintained by intrachromosomal gene conversion | 324 | Modifications of histones in nucleosomes may comprise a histone code | 353 |
| X-chromosome inactivation developed in response to gene depletion from the Y chromosome | 326 | Open and closed chromatin | 354 |
| 10.4 OUR PLACE IN THE TREE OF LIFE | 326 | ATP-dependent chromatin remodeling complexes | 356 |
| Molecular phylogenetics uses sequence alignments to construct evolutionary trees | 326 | DNA methylation is an important control in gene expression | 357 |
| Evolutionary trees can be constructed in different ways, and their reliability is tested by statistical methods | 328 | Methyl-CpG-binding proteins | 358 |
| The G-value paradox: organism complexity is not simply related to the number of (protein-coding) genes | 329 | DNA methylation in development | 359 |
| Striking lineage-specific gene family expansion often involves environmental genes | 332 | Chromatin states are maintained by several interacting mechanisms | 360 |
| Regulatory DNA sequences and other noncoding DNA have significantly expanded in complex metazoans | 333 | The role of HP1 protein | 361 |
| Mutations in <i>cis</i> -regulatory sequences lead to gene expression differences that underlie morphological divergence | 333 | A role for small RNA molecules | 361 |
| Lineage-specific exons and <i>cis</i> -regulatory elements can originate from transposable elements | 335 | A role for nuclear localization | 361 |
| Gene family expansion and gene loss/inactivation have occurred recently in human lineages, but human-specific genes are very rare | 337 | No single prime cause? | 362 |
| Comparative genomic and phenotype-led studies seek to identify DNA sequences important in defining humans | 339 | The ENCODE project seeks to give a comprehensive overview of transcription and its control | 362 |
| CONCLUSION | 341 | Transcription is far more extensive than previously imagined | 362 |
| FURTHER READING | 342 | Predicting transcription start sites | 364 |
| Chapter 11 | | 11.3 EPIGENETIC MEMORY AND IMPRINTING | 365 |
| Human Gene Expression | 345 | Epigenetic memory depends on DNA methylation, and possibly on the polycomb and trithorax groups of proteins | 365 |
| 11.1 PROMOTERS AND THE PRIMARY TRANSCRIPT | 347 | X-inactivation: an epigenetic change that is heritable from cell to daughter cell, but not from parent to child | 365 |
| Transcription by RNA polymerase II is a multi-step process | 347 | Initiating X-inactivation: the role of XIST | 366 |
| Defining the core promoter and transcription start site | 347 | Escaping X-inactivation | 367 |
| Assembling the basal transcription apparatus | 348 | At imprinted loci, expression depends on the parental origin | 367 |
| | | Prader-Willi and Angelman syndromes are classic examples of imprinting in humans | 368 |
| | | Two questions arise about imprinting: how is it done and why is it done? | 370 |
| | | Paramutations are a type of transgenerational epigenetic change | 370 |
| | | Some genes are expressed from only one allele but independently of parental origin | 371 |
| | | 11.4 ONE GENE—MORE THAN ONE PROTEIN | 372 |
| | | Many genes have more than one promoter | 372 |
| | | Alternative splicing allows one primary transcript to encode multiple protein isoforms | 373 |
| | | RNA editing can change the sequence of the mRNA after transcription | 374 |
| | | 11.5 CONTROL OF GENE EXPRESSION AT THE LEVEL OF TRANSLATION | 375 |
| | | Further controls govern when and where a mRNA is translated | 375 |

| | |
|--|-----|
| The discovery of many small RNAs that regulate gene expression caused a paradigm shift in cell biology | 376 |
| MicroRNAs as regulators of translation | 376 |
| MicroRNAs and cancer | 378 |
| Some unresolved questions | 378 |
| CONCLUSION | 378 |
| FURTHER READING | 379 |

Chapter 12 **Studying Gene Function in the Post-Genome Era** **381**

| | |
|---|-----|
| 12.1 STUDYING GENE FUNCTION: AN OVERVIEW | 382 |
| Gene function can be studied at a variety of different levels | 383 |
| Gene expression studies | 383 |
| Gene inactivation and inhibition of gene expression | 383 |
| Defining molecular partners for gene products | 383 |
| Genomewide analyses aim to integrate analyses of gene function | 384 |
| 12.2 BIOINFORMATIC APPROACHES TO STUDYING GENE FUNCTION | 384 |
| Sequence homology searches can provide valuable clues to gene function | 385 |
| Database searching is often performed with a model of an evolutionarily conserved sequence | 386 |
| Comparison with documented protein domains and motifs can provide additional clues to gene function | 388 |
| 12.3 STUDYING GENE FUNCTION BY SELECTIVE GENE INACTIVATION OR MODIFICATION | 389 |
| Clues to gene function can be inferred from different types of genetic manipulation | 389 |
| RNA interference is the primary method for evaluating gene function in cultured mammalian cells | 390 |
| Global RNAi screens provide a systems-level approach to studying gene function in cells | 392 |
| Inactivation of genes in the germ line provides the most detailed information on gene function | 392 |
| 12.4 PROTEOMICS, PROTEIN–PROTEIN INTERACTIONS, AND PROTEIN–DNA INTERACTIONS | 393 |
| Proteomics is largely concerned with identifying and characterizing proteins at the biochemical and functional levels | 393 |
| Large-scale protein–protein interaction studies seek to define functional protein networks | 395 |
| Yeast two-hybrid screening relies on reconstituting a functional transcription factor | 396 |
| Affinity purification–mass spectrometry is widely used to screen for protein partners of a test protein | 397 |
| Suggested protein–protein interactions are often validated by co-immunoprecipitation or pull-down assays | 398 |

| | |
|--|-----|
| The protein interactome provides an important gateway to systems biology | 399 |
| Defining nucleic acid–protein interactions is critical to understanding how genes function | 401 |
| Mapping protein–DNA interactions <i>in vitro</i> | 401 |
| Mapping protein–DNA interactions <i>in vivo</i> | 402 |
| CONCLUSION | 403 |
| FURTHER READING | 404 |

Chapter 13 **Human Genetic Variability and Its Consequences** **405**

| | |
|--|-----|
| 13.1 TYPES OF VARIATION BETWEEN HUMAN GENOMES | 406 |
| Single nucleotide polymorphisms are numerically the most abundant type of genetic variant | 406 |
| Both interspersed and tandem repeated sequences can show polymorphic variation | 408 |
| Short tandem repeat polymorphisms: the workhorses of family and forensic studies | 408 |
| Large-scale variations in copy number are surprisingly frequent in human genomes | 409 |
| 13.2 DNA DAMAGE AND REPAIR MECHANISMS | 411 |
| DNA in cells requires constant maintenance to repair damage and correct errors | 411 |
| The effects of DNA damage | 413 |
| DNA replication, transcription, recombination, and repair use multiprotein complexes that share components | 415 |
| Defects in DNA repair are the cause of many human diseases | 415 |
| Complementation groups | 416 |
| 13.3 PATHOGENIC DNA VARIANTS | 416 |
| Deciding whether a DNA sequence change is pathogenic can be difficult | 416 |
| Single nucleotide and other small-scale changes are a common type of pathogenic change | 417 |
| Missense mutations | 417 |
| Nonsense mutations | 418 |
| Changes that affect splicing of the primary transcript | 419 |
| Frameshifts | 420 |
| Changes that affect the level of gene expression | 421 |
| Pathogenic synonymous (silent) changes | 422 |
| Variations at short tandem repeats are occasionally pathogenic | 422 |
| Dynamic mutations: a special class of pathogenic microsatellite variants | 423 |
| Variants that affect dosage of one or more genes may be pathogenic | 425 |
| 13.4 MOLECULAR PATHOLOGY: UNDERSTANDING THE EFFECT OF VARIANTS | 428 |
| The biggest distinction in molecular pathology is between loss-of-function and gain-of-function changes | 428 |
| Allelic heterogeneity is a common feature of loss-of-function phenotypes | 429 |

| | |
|--|-----|
| Loss-of-function mutations produce dominant phenotypes when there is haploinsufficiency | 429 |
| Dominant-negative effects occur when a mutated gene product interferes with the function of the normal product | 431 |
| Gain-of-function mutations often affect the way in which a gene or its product reacts to regulatory signals | 432 |
| Diseases caused by gain of function of G-protein-coupled hormone receptors | 433 |
| Allelic homogeneity is not always due to a gain of function | 433 |
| Loss-of-function and gain-of-function mutations in the same gene will cause different phenotypes | 433 |
| 13.5 THE QUEST FOR GENOTYPE-PHENOTYPE CORRELATIONS | 435 |
| The phenotypic effect of loss-of-function mutations depends on the residual level of gene function | 435 |
| Genotype-phenotype correlations are especially poor for conditions caused by mitochondrial mutations | 436 |
| Variability within families is evidence of modifier genes or chance effects | 437 |
| CONCLUSION | 438 |
| FURTHER READING | 438 |

Chapter 14

Genetic Mapping of Mendelian

Characters

| | |
|--|-----|
| 14.1 THE ROLE OF RECOMBINATION IN GENETIC MAPPING | 442 |
| Recombinants are identified by genotyping parents and offspring for pairs of loci | 442 |
| The recombination fraction is a measure of the genetic distance between two loci | 442 |
| Recombination fractions do not exceed 0.5, however great the distance between two loci | 445 |
| Mapping functions define the relationship between recombination fraction and genetic distance | 445 |
| Chiasma counts give an estimate of the total map length | 446 |
| Recombination events are distributed non-randomly along chromosomes, and so genetic map distances may not correspond to physical distances | 446 |
| 14.2 MAPPING A DISEASE LOCUS | 448 |
| Mapping human disease genes depends on genetic markers | 448 |
| For linkage analysis we need informative meioses | 449 |
| Suitable markers need to be spread throughout the genome | 449 |
| Linkage analysis normally uses either fluorescently labeled microsatellites or SNPs as markers | 450 |
| 14.3 TWO-POINT MAPPING | 451 |
| Scoring recombinants in human pedigrees is not always simple | 451 |
| Computerized lod score analysis is the best way to analyze complex pedigrees for linkage between Mendelian characters | 452 |

| | |
|---|-----|
| Lod scores of +3 and -2 are the criteria for linkage and exclusion (for a single test) | 453 |
| For whole genome searches a genomewide threshold of significance must be used | 455 |
| 14.4 MULTIPOINT MAPPING | 455 |
| The CEPH families were used to construct marker framework maps | 455 |
| Multipoint mapping can locate a disease locus on a framework of markers | 456 |
| 14.5 FINE-MAPPING USING EXTENDED PEDIGREES AND ANCESTRAL HAPLOTYPES | 457 |
| Autozygosity mapping can map recessive conditions efficiently in extended inbred families | 457 |
| Identifying shared ancestral chromosome segments allows high-resolution genetic mapping | 459 |
| 14.6 DIFFICULTIES WITH STANDARD LOD SCORE ANALYSIS | 460 |
| Errors in genotyping and misdiagnoses can generate spurious recombinants | 461 |
| Computational difficulties limit the pedigrees that can be analyzed | 462 |
| Locus heterogeneity is always a pitfall in human gene mapping | 462 |
| Pedigree-based mapping has limited resolution | 462 |
| Characters with non-Mendelian inheritance cannot be mapped by the methods described in this chapter | 463 |
| CONCLUSION | 463 |
| FURTHER READING | 464 |

Chapter 15

Mapping Genes Conferring Susceptibility to Complex Diseases

| | |
|---|-----|
| 15.1 FAMILY STUDIES OF COMPLEX DISEASES | 468 |
| The risk ratio (λ) is a measure of familial clustering | 468 |
| Shared family environment is an alternative explanation for familial clustering | 469 |
| Twin studies suffer from many limitations | 469 |
| Separated monozygotic twins | 470 |
| Adoption studies are the gold standard for disentangling genetic and environmental factors | 471 |
| 15.2 SEGREGATION ANALYSIS | 471 |
| Complex segregation analysis estimates the most likely mix of genetic factors in pooled family data | 471 |
| 15.3 LINKAGE ANALYSIS OF COMPLEX CHARACTERS | 473 |
| Standard lod score analysis is usually inappropriate for non-Mendelian characters | 473 |
| Near-Mendelian families | 473 |
| Non-parametric linkage analysis does not require a genetic model | 474 |
| Identity by descent versus identity by state | 474 |
| Affected sib pair analysis | 474 |
| Linkage analysis of complex diseases has several weaknesses | 475 |
| Significance thresholds | 475 |
| Striking lucky | 476 |
| An example: linkage analysis in schizophrenia | 476 |

| | | | |
|---|------------|---|-----|
| 15.4 ASSOCIATION STUDIES AND LINKAGE DISEQUILIBRIUM | 477 | Mouse models have a special role in identifying human disease genes | 503 |
| Associations have many possible causes | 477 | 16.2 THE VALUE OF PATIENTS WITH CHROMOSOMAL ABNORMALITIES | 504 |
| Association is quite distinct from linkage, except where the family and the population merge | 479 | Patients with a balanced chromosomal abnormality and an unexplained phenotype provide valuable clues for research | 504 |
| Association studies depend on linkage disequilibrium | 479 | X-autosome translocations are a special case | 505 |
| The size of shared ancestral chromosome segments | 480 | Rearrangements that appear balanced under the microscope are not always balanced at the molecular level | 506 |
| Studying linkage disequilibrium | 481 | Comparative genomic hybridization allows a systematic search for microdeletions and microduplications | 507 |
| The HapMap project is the definitive study of linkage disequilibrium across the human genome | 481 | Long-range effects are a pitfall in disease gene identification | 508 |
| The use of tag-SNPs | 484 | 16.3 POSITION-INDEPENDENT ROUTES TO IDENTIFYING DISEASE GENES | 509 |
| 15.5 ASSOCIATION STUDIES IN PRACTICE | 484 | A disease gene may be identified through knowing the protein product | 509 |
| Early studies suffered from several systematic weaknesses | 485 | A disease gene may be identified through the function or interactions of its product | 509 |
| The transmission disequilibrium test avoids the problem of matching controls | 485 | A disease gene may be identified through an animal model, even without positional information | 511 |
| Association can be more powerful than linkage studies for detecting weak susceptibility alleles | 486 | A disease gene may be identified by using characteristics of the DNA sequence | 511 |
| Case-control designs are a feasible alternative to the TDT for association studies | 487 | 16.4 TESTING POSITIONAL CANDIDATE GENES | 512 |
| Special populations can offer advantages in association studies | 488 | For Mendelian conditions, a candidate gene is normally screened for mutations in a panel of unrelated affected patients | 512 |
| A new generation of genomewide association (GWA) studies has finally broken the logjam in complex disease research | 488 | Epigenetic changes might cause a disease without changing the DNA sequence | 513 |
| The size of the relative risk | 490 | The gene underlying a disease may not be an obvious one | 514 |
| 15.6 THE LIMITATIONS OF ASSOCIATION STUDIES | 491 | Locus heterogeneity is the rule rather than the exception | 514 |
| The common disease–common variant hypothesis proposes that susceptibility factors have ancient origins | 491 | Further studies are often necessary to confirm that the correct gene has been identified | 515 |
| The mutation–selection hypothesis suggests that a heterogeneous collection of recent mutations accounts for most disease susceptibility | 492 | 16.5 IDENTIFYING CAUSAL VARIANTS FROM ASSOCIATION STUDIES | 515 |
| A complete account of genetic susceptibility will require contributions from both the common disease–common variant and mutation–selection hypotheses | 493 | Identifying causal variants is not simple | 516 |
| CONCLUSION | 494 | Causal variants are identified through a combination of statistical and functional studies | 516 |
| FURTHER READING | 495 | Functional analysis of SNPs in sequences with no known function is particularly difficult | 518 |
| Chapter 16 | | Calpain-10 and type 2 diabetes | 518 |
| Identifying Human Disease Genes and Susceptibility Factors | 497 | Chromosome 8q24 and susceptibility to prostate cancer | 518 |
| 16.1 POSITIONAL CLONING | 498 | 16.6 EIGHT EXAMPLES OF DISEASE GENE IDENTIFICATION | 519 |
| Positional cloning identifies a disease gene from its approximate chromosomal location | 499 | Case study 1: Duchenne muscular dystrophy | 519 |
| The first step in positional cloning is to define the candidate region as tightly as possible | 500 | Case study 2: cystic fibrosis | 520 |
| The second step is to establish a list of genes in the candidate region | 500 | Case study 3: branchio-oto-renal syndrome | 522 |
| The third step is to prioritize genes from the candidate region for mutation testing | 501 | Case study 4: multiple sulfatase deficiency | 522 |
| Appropriate expression | 501 | Case study 5: persistence of intestinal lactase | 523 |
| Appropriate function | 502 | Case study 6: CHARGE syndrome | 525 |
| Homologies and functional relationships | 502 | Case study 7: breast cancer | 526 |
| | | Case study 8: Crohn disease | 528 |