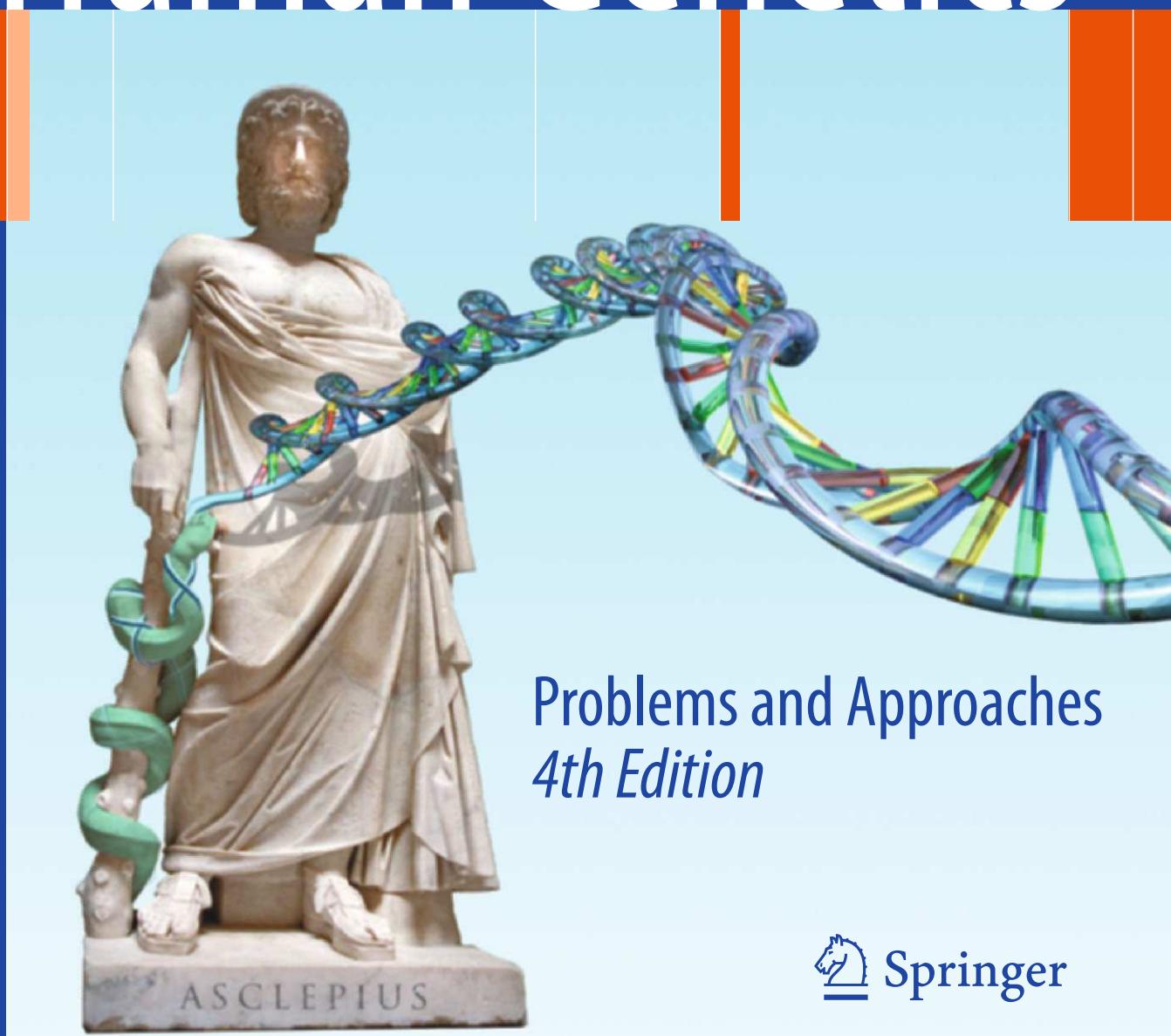


M.R. Speicher
S.E. Antonarakis
A.G. Motulsky
Editors

Vogel and Motulsky's Human Genetics



Problems and Approaches
4th Edition

Speicher • Antonarakis • Motulsky

VOGEL AND MOTULSKY'S HUMAN GENETICS

Problems and Approaches

Fourth Edition

Speicher Antonarakis Motulsky

VOGEL AND MOTULSKY'S HUMAN GENETICS

Problems and Approaches

Fourth, Completely Revised Edition

With 343 Figures and 76 Tables



Editors

Michael R. Speicher

Institute of Human Genetics
Medical University of Graz
Harrachgasse 21/8
8010 Graz
Austria
michael.speicher@medunigraz.at

Stylianos E. Antonarakis

Department of Genetic Medicine and Development
University of Geneva Medical School
and University Hospitals of Geneva
1 rue Michel-Servet
1211 Geneva
Switzerland
Stylianos.Antonarakis@unige.ch

Arno G. Motulsky

Professor Emeritus
Departments of Medicine (Division of Medical Genetics)
and Genome Sciences
University of Washington
School of Medicine
Box 355065
1705 N.E. Pacific St.
Seattle, WA 98195-5065
USA
agmot@u.washington.edu

ISBN 978-3-540-37653-8 e-ISBN 978-3-540-37654-5

DOI 10.1007/978-3-540-37654-5

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009931325

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: Estudio Calamar, Spain

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

In memory of Friedrich Vogel

The editors

*To my wife, Irene, and our children,
Alexander and Julia.*

Michael R. Speicher

*To my parents, my wife, Grigoria, and
our children.*

Stylianos E. Antonarakis

*To the memory of my wife, Gretel, and
to my children.*

Arno G. Motulsky

Preface

The first edition of *Human Genetics, Problems and Approaches*, was published in 1970 by human geneticists Friedrich Vogel and Arno Motulsky as sole authors. The aim was broad coverage and in-depth analysis of both medical and human genetics with an emphasis on problems and approaches with occasional historical comments. This point of view was fully explained in an introductory chapter of the three previous editions (1970, 1976, 1997). The book acquired an excellent reputation as an advanced text of human genetics and has been translated into Italian, Japanese, Russian, Chinese, and Portuguese. Our general aims for the fourth edition remain similar and together with novel developments are now set out in the Introduction of this new fourth edition.

Around 2004/2005, both Friedrich Vogel and Arno Motulsky, as well as the publishers, felt that the book should be continued with a new fourth edition in the same spirit and coverage as earlier editions, but should now include additional expert authors. After some delay and the death of Friedrich Vogel in the summer of 2006, a new editorial team consisting of Michael R. Speicher of the Medical University of Graz, Austria, Stylianos E. Antonarakis of the University of Geneva Medical School, and Arno G. Motulsky of the University of Washington School of Medicine, was constituted for the fourth edition of the Vogel/Motulsky book in the spirit of the original work.

An outline of the fourth edition's contents was developed and various internationally known geneticists, including the new editors, were selected to write the individual chapters. The resultant titles are listed in the table of contents. Most chapters are entirely new, while only three chapters (1, 5, and 6) utilize the third edition with appropriate, up-to-date revisions. Entirely new chapters include the description of the human genome, epigenetics, pharmacogenetics, genetic epidemiology, human evolution, genetics of mental retardation, autism, alcoholism and other addictions, consanguinity and related matters, gene therapy, cloning and genetic aspects of global health. Multiple chapters of various animal models used in the study of human and medical genetics are novel as are Databases and Genome Browsers as well as Databases Used in Human and Medical Genetics. The final content of the book was the result of many e-mails and conference calls. This new, updated, and totally revised edition does not contain some important and historically interesting chapters on certain topics. These can be found in the third edition of the book published in 1997, which was exclusively authored by F. Vogel and A.G. Motulsky. These topics were: enzymes in Chap. 7 (pp. 258–299); mutation rates in Chap. 9 (pp. 393–413); and mutation induction by ionizing radiation and chemicals in Chap. 11 (pp. 457–493).

The staff of Springer was most helpful in giving us extensive and firm aid in getting the book finished. We thank particularly Doris M. Binzenhöfer-Walker for her work during the early stage of the project and Isabella Athanassiou for her efficient help later. Andrea Pillman was a strict task mistress, who encouraged us to finish the book expeditiously.

The editors of the fourth edition want to express their intellectual indebtedness to Friedrich Vogel for having conceived of and played a major role in the first three editions of this book. Arno Motulsky particularly misses his discussions with Friedrich on human genetics and its role in science and medicine.

The cover illustration portrays a marble statue of Asclepius, the Greek god of healing, grasping a serpent-encircled staff as a symbol of medicine. The double helix of DNA joined to Asclepius symbolizes the applications of basic genetics to medicine.

4th edition
June 28, 2009

Michael R. Speicher, Graz
Stylianos E. Antonarakis, Geneva
Arno G. Motulsky, Seattle

Preface to the First Edition

Human genetics provides a theoretical framework for understanding the biology of the human species. It is a rapidly growing branch of science. New insights into the biochemical basis of heredity and the development of human cytogenetics in the 1950s heightened interest in this field. The number of research workers and clinicians who define themselves as full-time or part-time human and medical geneticists has increased sharply, and detailed well-founded knowledge has augmented exponentially. Many scientists and physicians are confronted with genetic problems and use concepts and methodology of human genetics in research and diagnosis. Methods developed in many different fields of the biologic, chemical, medical, and statistical sciences are being utilized toward the solution of genetic problems. The increasing number and sophistication of well-defined and elegantly solved problems helps to refine an extensive framework of genetic theory. These new conceptual insights in their turn lead to solutions of new questions. To mention only one example, the structure of hemoglobin genes has been elucidated using methods derived from protein chemistry and DNA technology. It is an exciting experience to participate in these developments!

Moreover, scientific progress in genetics has practical implications for human well-being. Improved knowledge of the genetic cause of an increasing number of human diseases helps to refine diagnosis, to find new therapeutic approaches, and above all, to prevent genetic diseases. So far, human genetics has had less of an impact on the behavioral and social sciences. It is possible that genetic differences involved in shaping personality structure, cognitive faculties, and possibly human social behavior may be at least as important as genetic variation affecting health and disease. The data, however, are less clear and more controversial. These problems are discussed in detail in the text. The rapid progress of human genetics in recent decades has attracted – and is still attracting – an increasing number of students and scientists from other fields. Various elementary textbooks, more advanced monographs of various branches of the field, and the original journal literature are the usual sources of introduction to human genetics. What seems to be lacking, however, is a fairly thorough and up-to-date treatise on the conceptual basis of the entire field of human genetics and its practical applications. Often, the absence of a broadly based background in the field leads to misunderstanding of its scope, unclear goals for research, improper selection of methods, and imbalanced theoretical discussions. Human genetics is based on a powerful theory, but this implicit conceptual foundation should be made explicit. This goal is the purpose of this book. It certainly is a formidable and possibly even too audacious task for two sole authors. However, both of us have been active in the field for more than 25 years.

We have worked on various problems and with a variety of methods. Since the early years of our careers, we have met occasionally, followed each other's writings, and were often surprised by the similarity of our opinions and judgments despite quite different early medical and scientific backgrounds. Moreover, our knowledge of the literature turned out to be in part overlapping and in part complementary. Since we are working in different continents, AGM had a better knowledge of concepts and results in the USA, while FV knew more of the continental European literature. Moreover, both of us have extensive experience as editors of journals in human genetics and one (FV) published a fairly comprehensive textbook in Germany some time ago (*Lehrbuch der allgemeinen Humangenetik*, Springer 1961), parts of which were still useful for the new book. We finally decided to take the risk, and, by writing an "advanced" text, to expose our deficiencies of knowledge, shortcomings of understanding, and biases of judgment.

A text endeavoring to expose the conceptual framework of human genetics cannot be dogmatic and has to be critical. Moreover, we could not confine ourselves to hard facts and well-proved statements. The cloud of conjectures and hypotheses surrounding a rapidly growing science had to be depicted. By doing so, we face the risk of being disproved by further results.

A number of colleagues helped by reading parts of the manuscript on which they had expert knowledge and by making useful suggestions: W. Buselmaier, U. Ehling, G. Flatz, W. Fuhrmann, S. Gartler, Eloise Giblett, P. Propping, Laureen Resnick, and Traute M. Schroeder. They should not be held responsible for possible errors. J. Krüger was of supreme help in the statistical parts. Our secretaries, Mrs. Adelheid Fengler and Mrs. Gabriele Bauer in Heidelberg, Mrs. Sylvia Waggoner in Seattle, and Mrs. Helena Smith in Stanford gave invaluable aid. The figures were drawn by Edda Schalt and Marianne Lebküchner. Miriam Gallaher and Susan Peters did an expert job of copy editing. The authors are especially grateful to Dr. Heinz Götze and Dr. Konrad F. Springer, of Springer Publishing Company, for the excellent production. The work could not have been achieved had the two authors not been invited to stay at the Center for Advanced Study in the Behavioral Sciences at Stanford (California) for the academic year of 1976/1977. The grant for AGM was kindly provided by the Kaiser Family Foundation, while the Spencer Foundation donated the grant for FV.

The cover of this book shows the mythical first human couple, Adam and Eve, as imagined by Albrecht Dürer (1504). They present themselves in the full beauty of their bodies, ennobled by the genius and skill of a great artist. The drawing should remind us of the uniqueness and dignity of the human individual. Human genetics can help us to understand humanity better and to make human life happier. This science is a cardinal example of Alexander Pope's statement. "The proper study of mankind is man."

Spring 1979

Friedrich Vogel, Heidelberg
Arno G. Motulsky, Seattle

Contents Overview

Introduction	1
1 History of Human Genetics.....	13
Arno G. Motulsky	
2 Human Genome Sequence and Variation	31
Stylianos E. Antonarakis	
3 Chromosomes.....	55
Michael R. Speicher	
4 From Genes to Genomics to Proteomics	139
Michael R. Speicher	
5 Formal Genetics of Humans: Modes of Inheritance	165
Arno G. Motulsky	
6 Linkage Analysis for Monogenic Traits.....	211
Arno G. Motulsky and Michael Dean	
7 Oligogenic Disease	243
Jon F. Robinson and Nicholas Katsanis	
8 Formal Genetics of Humans: Multifactorial Inheritance and Common Diseases	263
Andrew G. Clark	
8.1 Lessons from the Genome-Wide Association Studies for Complex Multifactorial Disorders and Traits.....	287
Jacques S. Beckmann and Stylianos E. Antonarakis	
9 Epigenetics	299
Bernhard Horsthemke	
10 Human Gene Mutation: Mechanisms and Consequences.....	319
Stylianos E. Antonarakis and David N. Cooper	
11 Human Hemoglobin	365
George P. Patrinos and Stylianos E. Antonarakis	

12	Human Genetics of Infectious Diseases.....	403
	Alexandre Alcaïs, Laurent Abel, and Jean-Laurent Casanova	
13	Gene Action: Developmental Genetics	417
	Stefan Mundlos	
14	Cancer Genetics.....	451
	Ian Tomlinson	
15	The Role of the Epigenome in Human Cancers.....	471
	Romulo Martin Brena and Joseph F. Costello	
16	Population Genetic Principles and Human Populations.....	487
	Emmanouil T. Dermitzakis	
17	Consanguinity, Genetic Drift, and Genetic Diseases in Populations with Reduced Numbers of Founders	507
	Alan H. Bittles	
18	Human Evolution	529
	Michael Hofreiter	
19	Comparative Genomics.....	557
	Ross C. Hardison	
20	Genetics and Genomics of Human Population Structure	589
	Sohini Ramachandran, Hua Tang, Ryan N. Gutenkunst, and Carlos D. Bustamante	
21	Genetic Epidemiology.....	617
	Sophia S. Wang , Terri H. Beaty, and Muin J. Khoury	
22	Pharmacogenetics	635
	Nicole M. Walley, Paola Nicoletti, and David B. Goldstein	
23	Behavioral Genetics.....	649
23.1	The Genetics of Personality.....	651
	Jonathan Flint and Saffron Willis-Owen	
23.2	Mental Retardation and Intellectual Disability	663
	David L. Nelson	
23.3	Genetic Factors in Alzheimer Disease and Dementia	681
	Thomas D. Bird	
23.4	Genetics of Autism.....	699
	Brett S. Abrahams and Daniel H. Geschwind	
23.5	The Genetics of Alcoholism and Other Addictive Disorders	715
	David Goldman and Francesca Ducci	
23.6	Behavioral Aspects of Chromosomal Variants	743
	Michael R. Speicher	

23.7	Genetics of Schizophrenia and Bipolar Affective Disorder	759
	Markus M. Nöthen, Sven Cichon, Christine Schmael, and Marcella Rietschel	
24	Model Organisms for Human Disorders	777
	Michael R. Speicher	
24.1	Mouse as a Model for Human Disease	779
	Antonio Baldini	
24.2	<i>Caenorhabditis elegans</i>, A Simple Worm: Bridging the Gap Between Traditional and Systems-Level Biology.....	787
	Morgan Tucker and Min Han	
24.3	<i>Drosophila</i> as a Model for Human Disease	795
	Ruth Johnson and Ross Cagan	
24.4	Human Genetics and the Canine System.....	813
	Heidi G. Parker and Elaine A. Ostrander	
24.5	Fish as a Model for Human Disease.....	827
	Siew Hong Lam and Zhiyuan Gong	
25	Genetic Counseling and Prenatal Diagnosis.....	845
	Tiemo Grimm and Klaus Zerres	
26	Gene Therapy	867
	Vivian W. Choi and R. Jude Samulski	
27	Cloning in Research and Treatment of Human Genetic Disease	875
	Ian Wilmut, Jane Taylor, Paul de Sousa, Richard Anderson, and Christopher Shaw	
28	Genetic Medicine and Global Health	885
	David J. Weatherall	
29	Genetic Databases	903
	Introductory Note by Stylianos E. Antonarakis	
29.1	Databases and Genome Browsers.....	905
	Rachel A. Harte, Donna Karolchik, Robert M. Kuhn, W. James Kent, and David Haussler	
29.2	Ensembl Genome Browser.....	923
	Xosé M. Fernández and Ewan Birney	
29.3	Databases in Human and Medical Genetics.....	941
	Roberta A. Pagon, Ada Hamosh, Johan den Dunnen, Helen V. Firth, Donna R. Maglott, Stephen T. Sherry, Michael Feolo, David Cooper, and Peter Stenson	
	Subject Index	961

Contents

Introduction	1
1 History of Human Genetics.....	13
Arno G. Motulsky	
1.1 The Greeks (see Stubbe [83])	14
1.2 Scientists Before Mendel and Galton	15
1.3 Galton's Work.....	16
1.4 Mendel's Work.....	17
1.5 Application to Humans: Garrod's Inborn Errors of Metabolism	18
1.6 Visible Transmitters of Genetic Information: Early Work on Chromosomes.....	20
1.7 Early Achievements in Human Genetics	20
1.7.1 AB0 and Rh Blood Groups.....	20
1.7.2 Hardy-Weinberg Law.....	21
1.7.3 Developments Between 1910 and 1930.....	21
1.8 Human Genetics, the Eugenics Movement, and Politics	21
1.8.1 United Kingdom and United States	21
1.8.2 Germany.....	22
1.8.3 Soviet Union/Russia (see Harper, Chap. 16 in [38])	23
1.8.4 Human Behavior Genetics	23
1.9 Development of Medical Genetics (1950–the Present)	23
1.9.1 Genetic Epidemiology	23
1.9.2 Biochemical Methods	24
1.9.3 Genetic and Biochemical Individuality	24
1.9.4 Cytogenetics, Somatic Cell Genetics, Prenatal Diagnosis.....	24
1.9.5 DNA Technology in Medical Genetics.....	25
1.9.6 The “Industrialization” of Discoveries and the Team Efforts.....	26
1.9.7 Unsolved Problems.....	26
References.....	27
2 Human Genome Sequence and Variation	31
Stylianos E. Antonarakis	
2.1 The Human Genome.....	31
2.1.1 Functional Elements	33

2.1.1.1	Protein-Coding Genes.....	33
2.1.1.2	Noncoding, RNA-Only Genes	36
2.1.1.3	Regions of Transcription Regulation	38
2.1.1.4	Conserved Elements Not Included in the Above Categories.....	39
2.1.2	Repetitive Elements	40
2.1.2.1	Segmental Duplications	41
2.1.2.2	Special Genomic Structures Containing Selected Repeats.....	41
2.1.3	Mitochondrial Genome.....	43
2.2	Genomic Variability.....	43
2.2.1	Single Nucleotide Polymorphisms	43
2.2.2	Short Sequence Repeats.....	46
2.2.3	Insertion/Deletion Polymorphisms (Indels).....	46
2.2.4	Copy Number Variants	47
2.2.5	Inversions.....	47
2.2.6	Mixed Polymorphisms.....	47
2.2.7	Genome Variation as a Laboratory Tool to Understand the Genome.....	48
	References.....	48
3	Chromosomes.....	55
	Michael R. Speicher	
3.1	History and Development of Human Cytogenetics	56
3.1.1	First Observations on Human Mitotic Chromosomes	56
3.1.2	An Old Error is Corrected and a New Era Begins	57
3.1.3	Birth of Human Cytogenetics 1956–1963	58
3.1.4	Introduction of Banding Technologies from the Late 1960s to the Present	58
3.1.5	The Birth of Molecular Cytogenetics in the Late 1960s.....	59
3.1.6	Molecular Cytogenetics or Fluorescence In Situ Hybridization 1980 to Date	59
3.1.7	Array Technologies: New Dimensions in Resolution from 1997 to Date	59
3.2	Organization of Genetic Material in Human Chromosomes	60
3.2.1	Heterochromatin and Euchromatin.....	60
3.2.2	From DNA Thread to Chromosome Structure	60
3.2.2.1	DNA Condensation.....	60
3.2.2.2	Histone Modifications Organize Chromosomal Subdomains.....	62
3.2.2.3	Chromatin Diseases	63
3.2.3	Centromeres and Kinetochores.....	63
3.2.3.1	Function of Centromeres	63
3.2.3.2	Structure of Centromeres.....	63
3.2.3.3	Centromeres and Human Diseases	65
3.2.4	Chromosome Bands.....	65
3.2.4.1	G- and R-bands	65
3.2.4.2	In Silico-Generated Bands	68
3.2.5	Telomeres.....	68
3.2.5.1	Structure of Telomeres (Loop and Proteins).....	68
3.2.5.2	The Telomere Replication Problem	69

3.2.5.3	Telomerase	70
3.2.5.4	Telomeres and Human Diseases	71
3.3	Cell Cycle and Mitosis.....	71
3.3.1	Cell Cycle: Interphase G ₁ -G ₂ and G ₀	71
3.3.2	Cell Cycle: Mitosis	72
3.3.2.1	Mitosis and Cytokinesis.....	72
3.3.2.2	Prophase.....	73
3.3.2.3	Prometaphase.....	73
3.3.2.4	Metaphase	74
3.3.2.5	Anaphase.....	74
3.3.2.6	Telophase	75
3.3.2.7	Cytokinesis	75
3.3.3	Cell Cycle Checkpoints	75
3.3.3.1	Presently Known Checkpoints.....	75
3.3.3.2	The Spindle-Assembly Checkpoint in Detail	75
3.3.3.3	“Cohesinopathies,” Cornelia de Lange and Robert Syndrome	77
3.3.3.4	Possible Additional Checkpoints.....	77
3.3.3.5	Checkpoint Failures and Human Diseases.....	77
3.3.4	Cell Cycle Coordinators	77
3.4	Chromosome Analysis Methods.....	78
3.4.1	Banding Techniques.....	78
3.4.1.1	Preparation of Mitotic Metaphase Chromosomes	78
3.4.1.2	G-Bands	79
3.4.1.3	R-Bands	79
3.4.1.4	C-Bands	79
3.4.1.5	Ag-NOR Bands.....	80
3.4.1.6	DA/DAPI Staining	80
3.4.2	Karyotype Description.....	80
3.4.2.1	Normal Human Karyotype in Mitotic Metaphase Chromosomes	80
3.4.2.2	Description of Normal Human Karyotypes and Karyotypes with Numerical and Structural Changes.....	81
3.4.3	Fragile Sites	82
3.4.3.1	Fluorescence In Situ Hybridization	83
3.4.3.2	FISH to Metaphase Spreads.....	84
3.4.3.3	Interphase FISH (Interphase Cytogenetics).....	85
3.4.3.4	Fiber FISH	85
3.4.3.5	Detection of Copy Number Changes in the Genome	86
3.4.3.6	Other Array-Based Methods.....	92
3.5	Meiosis.....	93
3.5.1	Biological Function of Meiosis	93
3.5.2	Meiotic Divisions.....	93
3.5.3	Meiotic Recombination Hotspots	95
3.5.4	Differences Between Male and Female Meiosis	96
3.5.4.1	Meiosis in the Human Male.....	96
3.5.4.2	Meiosis in the Human Female	96
3.5.4.3	A Dynamic Oocyte Pool?	96
3.5.4.4	Maternal Age and Aneuploidy.....	97

3.5.5	Nonallelic Homologous Recombination During Meiosis Can Cause Microdeletion/Microduplication Syndromes.....	98
3.5.6	Molecular Mechanisms Involved in Meiosis	99
3.6	Human Chromosome Pathology in Postnatal Diagnostics	99
3.6.1	Syndromes Attributable to Numeric Anomalies of Autosomes	99
3.6.1.1	Mechanisms Creating Anomalies in Chromosome Numbers (Numerical Chromosome Mutations)	99
3.6.1.2	Down Syndrome	101
3.6.1.3	Other Autosomal Trisomies	104
3.6.1.4	Triploidy	104
3.6.1.5	Mosaics	105
3.6.2	Syndromes Attributable to Structural Anomalies of Autosomes	106
3.6.2.1	Karyotypes and Clinical Syndromes	106
3.6.2.2	Small Deletions, Structural Rearrangements, and Monogenic Disorders: Genomic Disorders and Contiguous Gene Syndromes.....	113
3.6.3	Sex Chromosomes	115
3.6.3.1	First Observations	115
3.6.3.2	Dosage Compensation for Mammalian X Chromosomes	117
3.6.3.3	X Chromosomal Aneuploidies in Humans	120
3.6.3.4	Chromosomal Aneuploidies in Humans	125
3.6.4	Chromosome Aberrations and Spontaneous Miscarriage.....	125
3.6.4.1	Incidence of Prenatal Zygote Loss in Humans	125
3.6.4.2	Aneuploidy in Oocytes and Embryos	125
3.7	Chromosome Instability/Breakage Syndromes	126
3.7.1	Fanconi Anemia.....	126
3.7.2	Instabilities Caused by Mutations of Proteins of the RecQ Family of Helicases: Bloom syndrome, Werner syndrome, and Rothmund-Thomson syndrome	127
3.7.2.1	Bloom Syndrome	127
3.7.2.2	Rothmund–Thomson Syndrome	128
3.7.3	The Ataxia-Telangiectasia Group	128
3.7.3.1	Ataxia-Telangiectasia.....	128
3.7.3.2	Nijmegen Breakage Syndrome	129
3.7.4	Immunodeficiency, Centromeric Region Instability, and Facial Anomalies Syndrome	129
3.7.5	Roberts Syndrome/SC Phocomelia.....	129
3.7.6	Mosaic Variegated Aneuploidy.....	129
	References.....	130
4	From Genes to Genomics to Proteomics.....	139
	Michael R. Speicher	
4.1	Single-Gene Approaches	140
4.1.1	What Is a Gene?.....	140
4.1.2	Mutations	141
4.1.3	Silent Mutations and Phenotypic Consequences	142
4.1.4	Mutation Detection by Sanger Sequencing	143

4.1.5	Next-Generation Sequencing.....	144
4.1.6	The Importance of Monogenic Mendelian Disorders.....	144
4.2	Gene Regulation	145
4.2.1	Genetic Regulation	145
4.2.2	Epigenetic Regulation.....	146
4.2.3	Regulatory Transcripts of Small RNAs	147
4.3	“-omics” Sciences.....	147
4.4	Genomics	148
4.4.1	Genomes of Organisms.....	148
4.4.2	Array and Other Technologies.....	148
4.4.3	Next-Generation Sequencing.....	148
4.4.3.1	Roche’s (454) GS FLX Genome Analyzer.....	149
4.4.3.2	Illumina’s Solexa IG Sequencer	149
4.4.3.3	Applied Biosystem’s SOLiD System.....	149
4.4.4	Third-Generation Sequencing.....	152
4.4.5	Personalized Genomics.....	152
4.4.6	Gene Function.....	153
4.5	Transcriptomics	154
4.5.1	Capturing the Cellular Transcriptome, Expression Arrays, and SAGE	154
4.5.2	Regulatory Networks	154
4.5.3	Outlier Profile Analysis	155
4.5.4	High-Throughput Long- and Short-Read Transcriptome Sequencing	156
4.5.5	Disease Classification	156
4.5.6	Tools for Prognosis Estimation.....	156
4.6	Proteomics	156
4.6.1	From Low-Throughput to High-Throughput Techniques.....	157
4.6.2	Mass Spectrometer-Based Methods.....	157
4.6.3	Antibody Array-Based Methods.....	158
4.6.4	Proteomic Strategies	158
4.6.5	Proteomics for Screening and Diagnosis of Disease (Diagnostic and Prognostic Biomarkers).....	158
4.7	Conclusions.....	159
	References.....	159
5	Formal Genetics of Humans: Modes of Inheritance	165
	Arno G. Motulsky	
5.1	Mendel’s Modes of Inheritance and Their Application to Humans	166
5.1.1	Codominant Mode of Inheritance.....	166
5.1.2	Autosomal Dominant Mode of Inheritance	167
5.1.2.1	Late Manifestation, Incomplete Penetrance, and Variable Expressivity	169
5.1.2.2	Effect of Homozygosity on Manifestation of Abnormal Dominant Genes	171
5.1.3	Autosomal-Recessive Mode of Inheritance.....	172
5.1.3.1	Pseudodominance in Autosomal Recessive Inheritance.....	173

5.1.3.2	Compound Heterozygotes	174
5.1.4	X-Linked Modes of Inheritance.....	175
5.1.4.1	X-Linked Recessive Mode of Inheritance	175
5.1.4.2	X-Linked Dominant Mode of Inheritance	177
5.1.4.3	X-Linked Dominant Inheritance with Lethality of the Male Hemizygotes [90]	177
5.1.4.4	Genes on the Y Chromosome	179
5.1.5	“Lethal” Factors [32]	179
5.1.5.1	Animal Models	179
5.1.5.2	Lethals in Humans	180
5.1.6	Modifying Genes	180
5.1.6.1	Modifying Genes in the AB0 Blood Group System.....	180
5.1.6.2	Modifying Genes in Cystic Fibrosis	181
5.1.6.3	Sex-Limiting Modifying Genes	181
5.1.6.4	Modification by the Other Allele	181
5.1.6.5	Modification by Variation in Related Genes.....	182
5.1.6.6	Modification by a DNA Polymorphism Within the Same Gene	182
5.1.7	Anticipation	182
5.1.8	Total Number of Conditions with Simple Modes of Inheritance Known so far in Humans	184
5.1.8.1	Difference in the Relative Frequencies of Dominant and Recessive Conditions in Humans and Animals?.....	185
5.1.9	Uniparental Disomy and Genomic Imprinting	185
5.1.9.1	Phenotypic Consequences of UPD	186
5.1.9.2	Human Disorders Involving UPD.....	187
5.1.10	Diseases Due to Mutations in the Mitochondrial Genome.....	188
5.1.10.1	Leber Optical Atrophy	188
5.1.10.2	Deletions	190
5.1.10.3	Diseases of Advanced Age	190
5.1.10.4	Interaction Between Nuclear and Mitochondrial Genomes.....	190
5.1.11	Unusual, “Near Mendelian” Modes of Inheritance	190
5.1.11.1	Digenic Inheritance.....	190
5.1.11.2	Triallelic Inheritance.....	193
5.1.12	Multifactorial Inheritance	194
5.2	Hardy–Weinberg Law and Its Applications.....	194
5.2.1	Formal Basis	194
5.2.1.1	Derivations from the Hardy–Weinberg Law	194
5.2.2	Hardy–Weinberg Expectations Establish the Genetic Basis of AB0 Blood Group Alleles	195
5.2.2.1	Multiple Allelisms	195
5.2.2.2	Genetics of the AB0 Blood Groups	195
5.2.2.3	Meaning of a Hardy–Weinberg Equilibrium	197
5.2.3	Gene Frequencies.....	198
5.2.3.1	One Gene Pair: Only Two Phenotypes Known.....	198
5.3	Statistical Methods in Formal Genetics: Analysis of Segregation Ratios.....	198
5.3.1	Segregation Ratios as Probabilities.....	198
5.3.2	Simple Probability Problems in Human Genetics	200

5.3.2.1	Independent Sampling and Prediction in Genetic Counseling	200
5.3.2.2	Differentiation Between Different Modes of Inheritance.....	200
5.3.3	Testing for Segregation Ratios Without Ascertainment Bias: Codominant Inheritance.....	201
5.3.3.1	Dominance.....	201
5.3.4	Testing for Segregation Ratios: Rare Traits.....	202
5.3.4.1	Principal Biases	202
5.3.4.2	Methods for Correcting Bias	203
5.3.5	Discrimination of Genetic Entities Genetic Heterogeneity	204
5.3.5.1	Genetic Analysis of Muscular Dystrophy as an Example	204
5.3.5.2	Multivariate Statistics	205
5.3.6	Conditions without Simple Modes of Inheritance	205
5.3.6.1	Empirical Risk Figures	205
5.3.6.2	Selecting and Examining Probands and Their Families	205
5.3.6.3	Statistical Evaluation, Age Correction.....	206
5.3.6.4	Example	206
5.3.6.5	Selection of Probands for Genome-Wide Association Studies.....	207
5.3.6.6	Theoretical Risk Figures Derived from Heritability Estimates?....	207
5.4	Conclusions.....	207
	References.....	207
6	Linkage Analysis for Monogenic Traits.....	211
	Arno G. Motulsky and Michael Dean	
6.1	Linkage: Localization of Genes on Chromosomes.....	211
6.1.1	Classic Approaches in Experimental Genetics: Breeding Experiments and Giant Chromosomes	212
6.1.1.1	Linkage and Association.....	213
6.1.2	Linkage Analysis in Humans.....	213
6.1.2.1	Direct Observation of Pedigrees	213
6.1.2.2	Statistical Analysis.....	215
6.1.2.3	The Use of LOD Scores.....	216
6.1.2.4	Recombination Probabilities and Map Distances	217
6.1.2.5	The Sib Pair Method	217
6.1.2.6	Results for Autosomal Linkage, Sex Difference, and Parental Age.....	219
6.1.2.7	Information from Chromosome Morphology	220
6.1.3	Linkage Analysis in Humans: Cell Hybridization and DNA Techniques	220
6.1.3.1	First Observations on Cell Fusion	220
6.1.3.2	First Observation of Chromosome Loss in Human–Mouse Cell Hybrids and First Assignment of a Gene Locus	221
6.1.3.3	Other Sources of Information for Gene Localization	222
6.1.3.4	DNA Polymorphisms and Gene Assignment	222
6.1.3.5	Gene Symbols to Be Used	223
6.1.3.6	Linkage of X-Linked Gene Loci.....	223
6.1.3.7	Genetic and Physical Map of the Homologous Segment of X and Y Chromosomes	223

6.1.3.8	The Y Chromosome.....	224
6.1.3.9	DNA Variants in Linkage	224
6.1.3.10	Practical Application of Results from Linkage Studies.....	224
6.1.4	Biology and Statistics of Positional Cloning	224
6.2	Gene Loci Located Close to Each Other and Having Related Functions	225
6.2.1	Some Phenomena Observed in Experimental Genetics.....	225
6.2.1.1	Closely Linked Loci May Show a Cis-Trans Effect	225
6.2.1.2	Explanation in Terms of Molecular Biology	225
6.2.1.3	A Number of Genes May Be Closely Linked.....	225
6.2.2	Some Observations in the Human Linkage Map	225
6.2.2.1	Types of Gene Clusters That Have Been Observed.....	225
6.2.2.2	Clusters Not Observed So Far	226
6.2.3	Why Do Gene Clusters Exist?	226
6.2.3.1	They Are Traces of Evolutionary History.....	226
6.2.3.2	Duplication and Clustering May Be Used for Improvement of Function.....	226
6.2.4	Blood Groups: Rh Complex (111700), Linkage Disequilibrium	226
6.2.4.1	History	227
6.2.4.2	Fisher's Hypothesis of Two Closely Linked Loci.....	227
6.2.4.3	Confirmation and Tentative Interpretation of the Sequential Order	228
6.2.4.4	Molecular Basis	228
6.2.4.5	Linkage Disequilibrium	229
6.2.5	Major Histocompatibility Complex [105, 111]	229
6.2.5.1	History	229
6.2.5.2	Main Components of the MHC on Chromosome 6.....	231
6.2.5.3	Complement Components	232
6.2.5.4	Significance of HLA in Transplantation	232
6.2.5.5	Linkage Disequilibrium	232
6.2.5.6	The Normal Function of the System.....	234
6.2.6	Unequal Crossing Over.....	235
6.2.6.1	Discovery of Unequal Crossing Over	235
6.2.6.2	Unequal Crossing Over in Human Genetics.....	235
6.2.6.3	First Event.....	236
6.2.6.4	Consequences of Unequal Crossing Over	237
6.2.6.5	Intrachromosomal Unequal Crossing Over	237
6.3	Conclusions.....	238
	References.....	238
7	Oligogenic Disease	243
	Jon F. Robinson and Nicholas Katsanis	
7.1	The Limitations of Mendelian Concepts	244
7.2	PKU and Hyperphenylalaninemia: Genetic Heterogeneity	245
7.3	Cystic Fibrosis: Genetic Modifiers	246
7.4	Lessons Learned from Established Oligogenic Disorders.....	246
7.4.1	Bardet-Biedl Syndrome.....	247

7.4.2	Determining Oligogenicity	247
7.4.3	Cortisone Reductase Deficiency	248
7.4.4	Hemochromatosis	248
7.4.5	Hirschsprung Disease	249
7.5	Establishing Oligogenicity: Concepts and Methods.....	252
7.5.1	Heritability	252
7.5.2	Mouse Models of Oligogenic Inheritance: Familial Adenomatous Polyposis.....	252
7.5.3	Multigenic Models.....	253
7.5.4	Linkage Analysis	253
7.6	Molecular Mechanisms of Oligogenic Disorders.....	254
7.7	Modular or Systems Biology	257
7.8	Conclusions..... References.....	258 259
8	Formal Genetics of Humans: Multifactorial Inheritance and Common Diseases	263
	Andrew G. Clark	
8.1	Genetic Analysis of Complex Traits.....	264
8.1.1	Variation in Phenotypic Traits	264
8.1.2	Familial Resemblance and Heritability.....	264
8.1.3	The Special Case of Twins.....	267
8.1.4	Embedding a Single Measured Gene Influencing a Continuous Trait	269
8.1.5	A Model for Variance Partitioning.....	269
8.1.6	Relating the Model to Data.....	270
8.1.7	Mendelian Diseases Are Not Simple.....	271
8.2	Genetic Polymorphism and Disease	271
8.2.1	Finding Genes Underlying a Complex Trait.....	271
8.2.2	Limitations of Pedigree Analysis.....	271
8.2.3	A Prevailing Model: Common Disease Common Variants	272
8.2.4	Affected Sib-pairs	273
8.2.5	Transmission Disequilibrium Test	273
8.2.6	Full-Genome Association Testing	274
8.3	LD Mapping and Genome-Wide Association Studies.....	275
8.3.1	Theory and How It Works: HapMap and Genome-Wide LD	275
8.3.2	Technology: The Fantastic Drop in Genotyping Costs.....	276
8.3.3	Case-Control Studies	277
8.3.4	Statistical Inference with Genome-Wide Studies	277
8.3.5	Replication and Validation.....	278
8.3.6	Age-Related Macular Degeneration and Complement Factor H	279
8.4	Admixture Mapping and Population Stratification.....	279
8.4.1	How to Quantify Admixture	279
8.4.2	Using Admixture for Mapping	280
8.4.3	The Perils of Population Stratification.....	280
8.4.4	How to Correct for Hidden Population Stratification	281

8.5	Complications	282
8.5.1	Genotype by Environment Interaction.....	282
8.5.2	Epistasis	284
8.6	Missing Heritability: Why is so Little Variance Explained by GWAS Results?	284
8.7	Concluding Remarks	285
	References.....	285
8.1	Lessons from the Genome-Wide Association Studies for Complex Multifactorial Disorders and Traits.....	287
	Jacques S. Beckmann and Stylianos E. Antonarakis	
8.1.1	Lessons from Current GWAS	292
8.1.2	Genomic Topography of Trait-Associated Variants.....	292
8.1.3	How Important Is the Identified Genetic Contribution to the Variance of the Traits Studied?	292
8.1.4	Predictive Power and Clinical Utility of the Trait-Associated Variants	293
8.1.5	Pathophysiological Dissection of Complex Traits.....	294
8.1.6	Concluding Remarks	294
	References.....	295
9	Epigenetics	299
	Bernhard Horsthemke	
9.1	History, Definition, and Scope	300
9.2	Chromatin-Marking Systems.....	300
9.2.1	DNA Methylation	301
9.2.1.1	DNA Methyltransferases	302
9.2.1.2	Methyl-Cytosine-Binding Proteins.....	303
9.2.2	Histone Modification	303
9.2.2.1	Histone Acetylation	304
9.2.2.2	Histone Methylation	304
9.2.2.3	Histone Phosphorylation and Other Histone Modifications	305
9.2.3	Chromatin Remodeling.....	305
9.2.4	Synergistic Relations Between the Different Chromatin-Marking Systems.....	305
9.3	Specific Epigenetic Phenomena.....	306
9.3.1	Genomic Imprinting.....	307
9.3.2	X Inactivation.....	310
9.3.3	Allelic Exclusion in the Olfactory System	312
9.4	Epigenetic Variation and Disease	312
9.4.1	Obligatory Epigenetic Variation	313
9.4.2	Pure Epigenetic Variation	313
9.4.3	Facilitated Epigenetic Variation.....	315
9.5	Transgenerational Epigenetic Inheritance and Evolution.....	315
	References.....	316

10	Human Gene Mutation: Mechanisms and Consequences	319
	Stylianios E. Antonarakis and David N. Cooper	
10.1	Introduction.....	319
10.2	Neutral Variation/DNA Polymorphisms.....	320
10.3	Disease-Causing Mutations	321
10.3.1	The Nature of Mutation	321
10.3.1.1	Nucleotide Substitutions.....	322
10.3.1.2	Micro-Deletions and Micro-Insertions	323
10.3.1.3	Expansion/Copy Number Variation of Trinucleotide (and Other) Repeat Sequences.....	325
10.3.1.4	Gross Deletions.....	327
10.3.1.5	Large Insertions (Via Retrotransposition).....	329
10.3.1.6	Large Insertion of Repetitive and Other Elements	331
10.3.1.7	Inversions	331
10.3.1.8	Duplications.....	332
10.3.1.9	Gene Conversion.....	332
10.3.1.10	Insertion-Deletions (Indels).....	333
10.3.1.11	Other Complex Defects	333
10.3.1.12	Molecular Misreading.....	333
10.3.1.13	Germline Epimutations.....	334
10.3.1.14	Frequency of Disease-Producing Mutations.....	334
10.3.1.15	Chromosomal Distribution of Human Disease Genes	335
10.3.1.16	Mutation Nomenclature	335
10.3.1.17	Mutations in Gene Evolution.....	335
10.3.2	Consequences of Mutations.....	336
10.3.2.1	Mutations Affecting the Amino Acid Sequence of the Predicted Protein, but not Gene Expression	336
10.3.2.2	Mutations Affecting Gene Expression.....	337
10.3.2.3	Transcription (Promoter) Mutations	337
10.3.2.4	mRNA Splicing Mutants	338
10.3.2.5	RNA Cleavage-Polyadenylation Mutants	342
10.3.2.6	Mutations in miRNA-Binding Sites	342
10.3.2.7	Cap Site Mutations	343
10.3.2.8	Mutations in 5' Untranslated Regions	343
10.3.2.9	Mutations in 3' Regulatory Regions	343
10.3.2.10	Translational Initiation Mutations	343
10.3.2.11	Termination Codon Mutations	344
10.3.2.12	Frameshift Mutations.....	344
10.3.2.13	Nonsense Mutations.....	344
10.3.2.14	Unstable Protein Mutants	345
10.3.2.15	Mutations in Remote Promoter Elements/Locus Control Regions	345
10.3.2.16	Cellular Consequences of Trinucleotide Repeat Expansions	346
10.3.2.17	Mutations Producing Inappropriate Gene Expression	346
10.3.2.18	Position Effect in Human Disorders	346
10.3.2.19	Position Effect by an Antisense RNA.....	347
10.3.2.20	Abnormal Proteins Due to Fusion of Two Different Genes	347

10.3.2.21	Mutations in Genes Involved in Mismatch Repair Associated with Genomic Instability in the Soma.....	348
10.3.2.22	Mosaicism.....	348
10.3.2.23	Sex Differences in Mutation Rates	348
10.3.2.24	Concepts of Dominance and Recessiveness in Relation to the Underlying Mutations	349
10.4	General Principles of Genotype-Phenotype Correlations.....	349
10.5	Why Study Mutation?.....	351
	References.....	351
11	Human Hemoglobin	365
	George P. Patrinos and Stylianos E. Antonarakis	
11.1	Introduction.....	366
11.2	Historical Perspectives.....	366
11.3	Genetics of Hemoglobins	368
11.3.1	Hemoglobin Molecules.....	368
11.3.2	Hemoglobin Genes	369
11.3.3	Regulatory Elements.....	370
11.3.4	Molecular Control of Globin Gene Switching.....	374
11.3.5	DNA Polymorphisms at the Globin Genes.....	377
11.4	Molecular Evolution of the Human Globin Genes	377
11.5	Molecular Etiology of Hemoglobinopathies	380
11.5.1	Thalassemias and Related Conditions	380
11.5.2	β -Thalassemia.....	380
11.5.3	Dominantly Inherited β -Thalassemia	382
11.5.4	$\delta\beta$ -Thalassemias and Hereditary Persistence of Fetal Hemoglobin	383
11.5.5	α -Thalassemia	383
11.5.6	Other Mutation Types Leading to Hemoglobinopathies	384
11.5.7	Hemoglobin Variants	385
11.6	X-Linked Inherited and Acquired α -Thalassemia.....	387
11.6.1	Regulatory SNPs and Antisense RNA Transcription Resulting in α -Thalassemia	388
11.6.2	β -Thalassemia Attributable to Transcription Factor Mutations....	388
11.7	Population Genetics of Hemoglobin Genes.....	388
11.8	Diagnosis of Hemoglobinopathies.....	389
11.8.1	Carrier Screening	389
11.8.2	Hematological and Biochemical Methods.....	391
11.8.3	Molecular Diagnostics of Hemoglobinopathies	391
11.8.4	Prenatal and Preimplantation Genetic Diagnosis of Hemoglobinopathies.....	392
11.8.5	Genetic Counseling.....	393
11.9	HbVar Database for Hemoglobin Variants and Thalassemia Mutations	393
11.10	Therapeutic Approaches for the Thalassemias	394
11.10.1	Hematopoietic Stem Cell Transplantation.....	394
11.10.2	Pharmacological Reactivation of Fetal Hemoglobin	394
11.10.3	Pharmacogenomics and Therapeutics of Hemoglobinopathies	395

11.10.4	Gene Therapy.....	395
	References.....	396
12	Human Genetics of Infectious Diseases	403
	Alexandre Alcaïs, Laurent Abel, and Jean-Laurent Casanova	
12.1	Introduction.....	404
12.2	Mendelian Predisposition to Multiple Infections	406
12.3	Mendelian Predisposition to Single Infections	407
12.4	Mendelian Resistance	408
12.5	Major Genes.....	409
12.6	Multigenic Predisposition.....	410
12.7	Concluding Remarks	411
	References.....	412
13	Gene Action: Developmental Genetics	417
	Stefan Mundlos	
13.1	Genetics of Embryonal Development.....	417
13.1.1	Basic Mechanisms of Development.....	418
13.1.2	Mechanisms of Morphogenesis	419
13.2	The Stages of Development.....	422
13.3	Formation of the Central Nervous System	425
13.4	The Somites	431
13.5	The Brachial Arches	433
13.6	Development of the Limbs.....	435
13.7	Development of the Circulatory System.....	438
13.8	Development of the Kidney	440
13.9	Skeletal Development	442
13.10	Abnormal Development: Definitions and Mechanisms.....	445
13.11	Malformations.....	446
13.12	Disruptions.....	447
13.13	Deformations	448
13.14	Dysplasias	448
13.15	Terminology of Congenital Defects.....	449
	References.....	449
14	Cancer Genetics	451
	Ian Tomlinson	
14.1	Inherited Risk.....	452
14.1.1	Introduction.....	452
14.1.2	Identification of Mendelian Cancer Susceptibility Genes	452
14.1.3	Unidentified Mendelian Cancer Susceptibility Genes	457
14.1.4	Germline Epimutations	458
14.1.5	A Heterozygote Phenotype in Mendelian Recessive Tumor Syndromes?.....	458
14.1.6	Some Cancer Genes are Involved in Both Dominant and Recessive Syndromes.....	458

14.1.7	Phenotypic Variation, Penetrance, and Rare Cancers in Mendelian Syndromes	459
14.1.8	Predisposition Alleles Specific to Ethnic Groups	460
14.1.9	Germinal Mutations can Determine Somatic Genetic Pathways	460
14.1.10	Non-Mendelian Genetic Cancer Susceptibility	460
14.1.11	Concluding Remarks	462
14.2	Somatic Cancer Genetics	462
14.2.1	Mutations Cause Cancer	462
14.2.2	Chromosomal-Scale Mutations, Copy Number Changes, and Loss of Heterozygosity	463
14.2.3	Activating Mutations and Oncogenes; Inactivating Mutations and Tumor Suppressor Genes	464
14.2.4	Epimutations	465
14.2.5	How Many Mutations are Needed to Make a Cancer?	465
14.2.6	The Role of Genomic Instability and Molecular Phenotypes.....	466
14.2.7	The Conundrum of Tissue Specificity	467
14.2.8	Clinicopathological Associations, Response and Prognosis	467
14.2.9	Posttherapy Changes.....	468
14.2.10	Concluding Remarks	468
	References.....	468
15	The Role of the Epigenome in Human Cancers	471
	Romulo Martin Brena and Joseph F. Costello	
15.1	Introduction.....	472
15.2	DNA Methylation in Development and Cellular Homeostasis.....	472
15.3	DNA Methylation is Disrupted in Human Primary Tumors.....	473
15.4	Genome-Wide DNA Methylation Analyses	474
15.5	Gene Silencing Vs. Gene Mutation	476
15.6	Discovery of Cancer Genes via Methylome Analysis	477
15.7	Computational Analysis of the Methylome.....	478
15.8	Histone Modifications and Chromatin Remodeling in Cancer	478
15.9	Epigenome–Genome Interactions in Human Cancer and Mouse Models: Gene Silencing Vs. Gene Mutation	480
15.10	Epigenetics and Response to Cancer Therapy.....	480
	References.....	481
16	Population Genetic Principles and Human Populations	487
	Emmanouil T. Dermitzakis	
16.1	Introduction.....	487
16.2	Processes Shaping Natural Variation	488
16.2.1	Mutation and Polymorphism	488
16.2.2	Historical Perspective of DNA Polymorphisms and Their Use as Genetic Markers.....	489
16.2.3	Demography	490

16.2.3.1	Effective Population Size.....	490
16.2.3.2	Genetic Drift and Isolated Populations.....	491
16.2.3.3	Migration	491
16.2.3.4	Inbreeding/Nonrandom Mating	492
16.2.4	Recombination.....	492
16.2.5	Natural Selection	492
16.3	Patterns of Genetic Variation.....	494
16.3.1	Hardy–Weinberg Equilibrium.....	494
16.3.2	Coalescent Theory	495
16.3.3	Population Differentiation	495
16.3.4	Patterns of Single-Nucleotide Variation in the Human Genome ...	496
16.3.5	Patterns of Structural Variation in the Human Genome	498
16.3.6	Haplotype Diversity—Linkage Disequilibrium.....	499
16.3.7	Detecting Natural Selection.....	500
16.3.8	Historical Perspective of Population Genetic Studies	501
16.4	Current Themes	501
16.4.1	International Efforts to Detect and Describe Sequence Variation	501
16.4.2	Prospects for Mapping Functional and Disease Variation	502
16.5	Summary.....	504
	References.....	504
17	Consanguinity, Genetic Drift, and Genetic Diseases in Populations with Reduced Numbers of Founders	507
	Alan H. Bittles	
17.1	Genetic Variation in Human Populations	508
17.1.1	Random Mating and Assortative Mating.....	509
17.1.2	Genetic Drift and Founder Effects.....	509
17.2	Consanguineous Matings.....	509
17.2.1	Coefficient of Relationship and Coefficient of Inbreeding.....	510
17.2.2	Global Prevalence of Consanguinity	510
17.2.3	Specific Types of Consanguineous Marriage.....	514
17.2.4	The Influence of Religion on Consanguineous Marriage	514
17.2.5	Civil Legislation on Consanguineous Marriage	515
17.2.6	Social and Economic Factors Associated with Consanguinity	515
17.3	Inbreeding and Fertility	516
17.3.1	Genetically Determined Factors Influencing Human Mate Choice.....	516
17.3.2	Inbreeding and Fetal Loss Rates.....	516
17.3.3	Comparative Fertility in Consanguineous and Nonconsanguineous Couples	517
17.4	Inbreeding and Inherited Disease	518
17.4.1	Consanguinity and Deaths in Infancy and Childhood	518
17.4.2	Consanguinity and Childhood Morbidity	518
17.4.3	Consanguinity and Adult Mortality and Morbidity	519
17.5	Incest.....	519
17.5.1	Mortality and Morbidity Estimates for Incestuous Matings.....	520

17.6	Genetic Load Theory and Its Application in Consanguinity Studies.....	520
17.7	Genomic Approaches to Measuring Inbreeding at Individual and Community Levels	521
17.8	The Influence of Endogamy and Consanguinity in Human Populations.....	521
17.8.1	The Finnish Disease Heritage	521
17.8.2	Inter- and Intra-population Differentiation in India.....	523
17.8.3	Consanguinity and the Distribution of Disease Alleles in Israeli Arab Communities.....	523
17.9	Evaluating Risk in Consanguineous Relationships	524
17.10	Concluding Comments	525
	References.....	525
18	Human Evolution	529
	Michael Hofreiter	
18.1	Historical Overview	530
18.1.1	Before and Around Darwin	530
18.1.2	Sarich and Wilson	530
18.1.3	From a Straight Line to a Bush of Hominid Species and Beyond.	531
18.2	The Fossil Record	532
18.2.1	Palaeoanthropology	532
18.2.2	Neanderthals and Ancient DNA	534
18.3	The Genetics of Human Evolution	536
18.3.1	The Genomes of Humans and their Relatives.....	536
18.3.2	Diversity Within the Human Genome	538
18.3.3	Positive and Negative Selection in the Human Genome	541
18.4	Recent Events in Human Evolution.....	545
18.4.1	Out of Africa into New-Found Lands.....	545
18.4.2	Domestication	549
18.4.3	Modern Human Population Structure.....	550
	Glossary	551
	References.....	551
19	Comparative Genomics	557
	Ross C. Hardison	
19.1	Goals, Impact, and Basic Approaches of Comparative Genomics	557
19.1.1	How Biological Sequences Change Over Time.....	558
19.1.2	Purifying Selection	559
19.1.3	Models of Neutral DNA	560
19.1.4	Adaptive Evolution	562
19.2	Alignments of Biological Sequences and Their Interpretation.....	563
19.2.1	Global and Local Alignments	563
19.2.2	Aligning Protein Sequences.....	563
19.2.3	Aligning Large Genome Sequences	563
19.3	Assessment of Conserved Function from Alignments	565

19.3.1	Phylogenetic Depth of Alignments.....	561
19.3.2	Portion of the Human Genome Under Constraint	568
19.3.3	Identifying Specific Sequences Under Constraint	569
19.4	Evolution Within Protein-Coding Genes	570
19.4.1	Comparative Genomics in Gene Finding.....	571
19.4.2	Sets of Related Genes	572
19.4.3	Rates of Sequence Change in Different Parts of Genes.....	574
19.4.4	Evolution and Function in Protein-Coding Exons.....	574
19.4.5	Fast-Changing Genes That Code for Proteins	575
19.4.6	Recent Adaptive Selection in Humans	576
19.4.7	Human Disease-Related Genes.....	578
19.5	Evolution in Regions That Do Not Code for Proteins or mRNA	579
19.5.1	Ultraconserved Elements	579
19.5.2	Evolution Within Noncoding Genes.....	580
19.5.3	Evolution and Function in Gene Regulatory Sequences	581
19.5.4	Prediction and Tests of Gene Regulatory Sequences.....	582
19.6	Resources for Comparative Genomics.....	583
19.6.1	Genome Browsers and Data Marts	583
19.6.2	Genome Analysis Workspaces.....	583
19.7	Concluding Remarks	584
	References.....	585
20	Genetics and Genomics of Human Population Structure.....	590
	Sohini Ramachandran, Hua Tang, Ryan N. Gutenkunst, and Carlos D. Bustamante	
20.1	Introduction.....	590
20.1.1	Evolutionary Forces Shaping Human Genetic Variation	590
20.2	Quantifying Population Structure	592
20.2.1	F_{ST} and Genetic Distance	592
20.2.2	Model-Based Clustering Algorithms	593
20.2.3	Characterizing Locus-Specific Ancestry	594
20.3	Global Patterns of Human Population Structure.....	595
20.3.1	The Apportionment of Human Diversity	595
20.3.2	The History and Geography of Human Genes	596
20.3.3	Genetic Structure of Human Populations	598
20.3.4	A Haplotype Map of the Human Genome	600
20.4	The Genetic Structure of Human Populations Within Continents and Countries.....	602
20.4.1	Genetic Differentiation in Eurasia	603
20.4.2	Genetic Variation in Native American Populations	605
20.4.3	The Genetic Structure of African Populations.....	606
20.5	Recent Genetic Admixture	606
20.5.1	Populations of the Americas	606
20.5.2	Admixture Around the World	608
20.6	Quantitative Modeling of Human Genomic Diversity.....	609
20.6.1	Demographic History.....	609
20.6.2	Quantitative Models of Selection.....	610
	References.....	613

21	Genetic Epidemiology	617
	Sophia S. Wang, Terri H. Beaty, and Muin J. Khoury	
21.1	Introduction.....	618
21.2	Scope and Strategies of Genetic Epidemiology in the Twenty-First Century	619
21.3	Fundamentals of Gene Discovery: Family Studies.....	619
21.4	Fundamentals of Gene Discovery: Population Studies and GWAS	623
21.5	Beyond Gene Discovery: Epidemiologic Assessment of Genes in Population Health.....	626
21.6	Beyond Gene Discovery: Epidemiologic Assessment of Genetic Information in Medicine and Public Health.....	629
21.7	Policy, Ethical and Practice Considerations: the Emergence of Public Health Genomics	631
	References.....	632
22	Pharmacogenetics	635
	Nicole M. Walley, Paola Nicoletti, and David B. Goldstein	
22.1	Introduction.....	636
22.1.1	The Goal of Pharmacogenetics.....	636
22.1.2	Why Pharmacogenetic Studies Are Necessary	636
22.1.1.1	Efficacy	636
22.1.1.2	Safety	636
22.1.1.3	Dose	637
22.2	Important Pharmacogenetic Discoveries to Date.....	637
22.2.1	CYP2D6 Polymorphism and Pharmacogenetics	637
22.2.2	Classic/Paradigmatic Studies	638
22.2.2.1	Isoniazid, Azathioprine, Mercaptopurine and Other Poor Metabolizer Phenotypes	638
22.2.2.2	Efficacy of Asthma Treatment	638
22.2.2.3	Abacavir Hypersensitivity	638
22.2.2.4	Tranilast and Clinical Pharmacogenetic Investigations	638
22.2.2.5	Cancer Pharmacogenetics and Somatic/Acquired Polymorphisms	639
22.3	Pharmacogenetic Methodology: Clinical Practice.....	639
22.3.1	Phenotype	639
22.3.2	Sample and Patient Recruitment.....	640
22.4	Pharmacogenetic Methodology: Genomics	640
22.4.1	Candidate Gene Studies	640
22.4.2	Whole-Genome SNP Analyzes.....	642
22.4.3	Rare Variants and Whole-Genome Sequencing.....	643
22.5	Challenges in Pharmacogenetics	643
22.5.1	Polygenic Inheritance	643
22.5.2	Pharmacogenetics in the Clinic	643
22.5.3	Pharmacogenetics Across Ethnic Groups	644
22.5.4	Drug Classes of Urgent Interest.....	645
2.6	Conclusions.....	646
	References.....	646

23	Behavioral Genetics	649
	Introductory Note by Michael R. Speicher	
23.1	The Genetics of Personality	651
	Jonathan Flint and Saffron Willis-Owen	
23.1.1	Neuroticism.....	653
23.1.1.1	Genetic Association Studies	653
23.1.1.2	Gene by Environment Effects.....	654
23.1.1.3	Linkage Studies	656
23.1.1.4	Genome-Wide Association Studies	656
23.1.2	Extraversion	657
23.1.3	Implications for Future Research.....	657
	References.....	658
23.2	Mental Retardation and Intellectual Disability	663
	David L. Nelson	
23.2.1	Introduction.....	663
23.2.2	Definition of ID.....	664
23.2.3	History of ID	664
23.2.4	Frequency of ID	664
23.2.5	Gene Dosage and ID	665
23.2.6	Down Syndrome	666
23.2.7	Recurrent Deletions and Duplications and ID	667
23.2.8	Prader-Willi and Angelman Syndromes	668
23.2.9	Future Directions in Genome Rearrangements and ID.....	668
23.2.10	Single-Gene Mutations and ID	668
23.2.10.1	Autosomal Recessive	668
23.2.10.2	Autosomal Dominant.....	669
23.2.11	X-Linked ID.....	670
23.2.11.1	Fragile X Syndrome.....	671
23.2.11.2	Rett Syndrome	675
23.2.12	Future Directions	676
	References.....	676
23.3	Genetic Factors in Alzheimer Disease and Dementia	681
	Thomas D. Bird	
23.3.1	Clinical Manifestations of Alzheimer Disease	681
23.3.1.1	Establishing the Diagnosis.....	682
23.3.1.2	Prevalence	682
23.3.2	Causes	683
23.3.2.1	Environmental.....	683
23.3.2.2	Heritable Causes [6, 7]	683
23.3.2.3	Unknown.....	685
23.3.2.4	Molecular Genetic Testing.....	686
23.3.2.5	Early-Onset Familial AD	688
23.3.3	Genetic Counseling.....	688
23.3.3.1	Mode of Inheritance.....	688

23.3.3.2	Risk to Family Members: EOAD	688
23.3.3.3	Related Genetic Counseling Issues.....	688
23.3.4	Management	689
23.3.4.1	Treatment of Manifestations	689
23.3.4.2	Therapies Under Investigation	689
23.3.4.3	Other	689
23.3.5	Other Causes of Dementia	690
23.3.5.1	Frontotemporal Dementia.....	690
23.3.5.2	Familial Prion Disorders	691
23.3.5.3	CADASIL	691
	References.....	692
23.4	Genetics of Autism	699
	Brett S. Abrahams and Daniel H. Geschwind	
23.4.1	Background.....	699
23.4.2	Cytogenetic Findings	701
23.4.3	Linkage	703
23.4.4	Syndromic ASDs	703
23.4.5	Re-sequencing.....	704
23.4.6	Copy Number Variation	705
23.4.7	Common Variations	707
23.4.8	Towards Convergence	707
	References.....	710
23.5	The Genetics of Alcoholism and Other Addictive Disorders	715
	David Goldman and Francesca Ducci	
23.5.1	Introduction.....	716
23.5.2	Definition of Substance Use Disorders and Other Addictions	716
23.5.3	Epidemiology and Societal Impact of Addiction.....	717
23.5.4	Genetics: Family and Twin Studies	718
23.5.4.1	The Heritability of Addictions.....	719
23.5.4.2	Do Genetic Factors Moderating Risk Differ in Men and Women?	719
23.5.4.3	Developmental Dependence of Genes and Environment in Risk.....	720
23.5.4.4	What Is the Nature of the Inheritance of Addictions?	720
23.5.4.5	Agent-Specific and Nonspecific Genetic and Environmental Factors	723
23.5.4.6	Are Genetic and Environmental Risk Factors Independent of Each Other?	724
23.5.4.7	Gene by Environment Correlation	726
23.5.4.8	Gene by Environment Interaction.....	727
23.5.5	Progress Through Intermediate Phenotypes	728
23.5.6	Finding the Specific Genes Underlying Vulnerability to Addiction	728
23.5.6.1	Candidate Genes	729

23.5.6.2	Genetic Mapping	734
23.5.7	Treatment of Addictions	735
23.5.8	Conclusion	736
	References.....	736
23.6	Behavioral Aspects of Chromosomal Variants	743
	Michael R. Speicher	
23.6.1	Introduction: Human Chromosome Aberrations and Behavior, Possibilities, and Limitations.....	744
23.6.2	Numeric Autosomal Aberrations	744
23.6.2.1	Down Syndrome	744
23.6.3	Copy Number Variations Associated with Behavioral Disorders.....	745
23.6.3.1	Autistic Spectrum Disorder	745
23.6.3.2	Schizophrenia	745
23.6.3.3	Bipolar Disorders.....	746
23.6.4	Aberrations of the X Chromosome.....	746
23.6.4.1	Turner Syndrome	746
23.6.4.2	Klinefelter Syndrome.....	746
23.6.4.3	Triple-X Syndrome	748
23.6.5	Aberrations of the Y Chromosome.....	748
23.6.5.1	XYY Syndrome	748
23.6.5.2	Higher Prevalence Among “Criminals”	748
23.6.5.3	Intellectual Dysfunction or Simply Stature?.....	749
23.6.5.4	Behavioral Aspects of XYY Men.....	749
23.6.5.5	Association of Criminal Behavior and Lowered Intelligence in XYY Men	750
23.6.5.6	Social and Therapeutic Consequences.....	750
23.6.5.7	XXYY Syndrome	751
23.6.6	Other Chromosomal Variants	751
23.6.6.1	22q11.2 Deletion Syndrome	751
23.6.6.2	Smith–Magenis Syndrome.....	752
23.6.6.3	Prader-Willi and Angelman Syndrome	753
23.6.6.4	Williams–Beuren Syndrome	753
23.6.6.5	Cri-du-chat Syndrome	753
23.6.6.6	Wolf–Hirschhorn Syndrome	754
	References.....	754
23.7	Genetics of Schizophrenia and Bipolar Affective Disorder	759
	Markus M. Nöthen, Sven Cichon, Christine Schmael, and Marcella Rietschel	
23.7.1	Schizophrenia	759
23.7.1.1	Prevalence	760
23.7.1.2	Environmental Risk Factors.....	760
23.7.1.3	Formal Genetic Studies.....	761
23.7.1.3.1	Family Studies	761
23.7.1.3.2	Twin Studies	761

23.7.1.3.3	Adoption Studies	761
23.7.1.4	Gene–Environment Interaction.....	761
23.7.1.5	The Evolutionary Paradox of Schizophrenia.....	762
23.7.1.6	Molecular Genetic Studies.....	762
23.7.1.6.1	Linkage Studies	762
23.7.1.6.2	Candidate Gene Studies.....	763
23.7.1.6.3	Genome-Wide Association Studies	763
23.7.1.6.4	Submicroscopic Chromosomal Aberrations	764
23.7.1.7	Endophenotypes.....	764
23.7.2	Bipolar Disorder	765
23.7.2.1	Prevalence	765
23.7.2.2	Environmental Risk Factors.....	765
23.7.2.3	Formal Genetic Studies.....	766
23.7.2.3.1	Family Studies	766
23.7.2.3.2	Twin Studies	766
23.7.2.3.3	Adoption Studies	766
23.7.2.4	Molecular Genetic Studies.....	766
23.7.2.4.1	Linkage Studies	767
23.7.2.4.2	Candidate Gene Studies	767
23.7.2.4.3	Genome-Wide Association Studies	767
23.7.2.5	Endophenotypes.....	768
23.7.3	Schizophrenia and Bipolar Disorder: Approaches Beyond a Diagnostic Dichotomy	768
23.7.4	Outlook	768
	References.....	770
24	Model Organisms for Human Disorders	777
	Introductory Note by Michael R. Speicher	
	References.....	777
24.1	Mouse as a Model for Human Disease.....	779
	Antonio Baldini	
24.1.1	Introduction.....	779
24.1.2	Gene and Genome Manipulation Strategies	780
24.1.2.1	Knockout.....	780
24.1.2.2	Knockin and the Use of Site-Specific Recombinases.....	780
24.1.2.3	Conditional Mutations	781
24.1.2.4	Multigene Deletions and Duplications	782
24.1.2.5	Transgenics	782
24.1.3	Generating Genetically Accurate Models.....	783
24.1.3.1	Genetic Diseases Caused by Loss of Function Mutations.....	783
24.1.3.2	Gene Dosage Mutations.....	783
24.1.3.3	Gain of Function Mutations.....	784
24.1.3.4	Segmental Aneuploidies	784
24.1.4	Future Perspectives	784
	References.....	784

24.2	<i>Caenorhabditis elegans</i>, A Simple Worm: Bridging the Gap Between Traditional and Systems-Level Biology	787
	Morgan Tucker and Min Han	
24.2.1	A Primer.....	787
24.2.1.1	A Short History.....	787
24.2.1.2	The Worm, Its Life Cycle, and Its Cultivation.....	788
24.2.2	The How and Why of Screening.....	789
24.2.2.1	Genetic Screens: A Traditional Single-Gene Approach	789
24.2.2.2	RNAi Screens: A High-Throughput Approach	790
24.2.2.3	Compound Screens: Identifying Therapeutics.....	791
24.2.3	Beyond the Simple Screen.....	791
24.2.3.1	A Systems Approach: Protein Interaction Maps	791
24.2.3.2	Exploring Complex Traits: Aging	792
24.2.3.3	Modeling Human Disorders: Alzheimer's Disease	793
24.2.4	Conclusions and Perspectives	793
	References.....	794
24.3	<i>Drosophila</i> as a Model for Human Disease	795
	Ruth Johnson and Ross Cagan	
24.3.1	Why Flies?	795
24.3.2	Cancer	796
24.3.2.1	Cell Cycle	797
24.3.2.2	Cell Death	797
24.3.2.3	Hyperplasia and Neoplasia	798
24.3.2.4	Models of Metastasis	799
24.3.2.5	Models of Specific Cancers	801
24.3.3	Neurodegenerative Diseases	802
24.3.3.1	Parkinson's Disease	803
24.3.3.2	Alzheimer's Disease	803
24.3.3.3	Triplet-Repeat Diseases	804
24.3.4	Heart Disease	804
24.3.5	Diabetes and Metabolic Diseases	805
24.3.5.1	Body Size.....	806
24.3.5.2	Models of Diabetes	806
24.3.6	Addiction	807
24.3.6.1	The Genetics of Addiction: Sensitivity and Tolerance	808
24.3.7	Sleep Disorders.....	808
24.3.8	Conclusions.....	808
	References.....	809
24.4	Human Genetics and the Canine System.....	813
	Heidi G. Parker and Elaine A. Ostrander	
	Abbreviations.....	814
24.4.1	Introduction to the Canine System	814
24.4.1.1	Origins of the Domestic Dog	814
24.4.1.2	Dog Breeds	814

24.4.1.3	Variation Between Breeds.....	814
24.4.1.4	Lack of Variation Within Breeds	815
24.4.1.5	Benefits of Mapping in a Breed-Based System	816
24.4.2	Navigating the Canine Genome	816
24.4.2.1	Maps	816
24.4.2.2	Sequence	816
24.4.3	Canine Disease Gene Studies	816
24.4.3.1	Canine Disease Mirrors Human Disease	817
24.4.3.2	Canine Disease and Mechanisms of Mutation.....	818
24.4.3.2.1	SINE Insertions.....	818
24.4.3.2.2	Simple Repeats	818
24.4.3.2.3	Single Base Mutations	818
24.4.4	Genome Structure in the Domestic Dog	819
24.4.4.1	Linkage Disequilibrium	820
24.4.4.2	Haplotype Structure	820
24.4.4.3	Single Nucleotide Polymorphisms	820
24.4.5	Population Structure in the Domestic Dog	820
24.4.5.1	Canine Breed Clusters Facilitate Mapping Efforts	820
24.4.5.2	Combining Breeds to Improve Mapping	821
24.4.5.3	Homozygosity and Population Bottlenecks.....	821
24.4.6	Mapping Multigenic Traits in the Dog	822
24.4.6.1	Quantitative Trait Loci.....	822
24.4.6.2	Establishing a Cohort.....	822
24.4.6.3	Complex Disease	823
24.4.7	Conclusion	823
	References.....	823
24.5	Fish as a Model for Human Disease	827
	Siew Hong Lam and Zhiyuan Gong	
24.5.1	Introduction.....	828
24.5.1.1	Brief Historical Background and Current Status	828
24.5.1.2	Why Use Fish to Model Human Disorders?.....	829
24.5.2	Strategies for Modeling Human Disorders in Fish	830
24.5.2.1	Forward Genetics	830
24.5.2.1.1	Chemical Mutagenesis.....	830
24.5.2.1.2	Insertional Mutagenesis	831
24.5.2.2	Reverse Genetics.....	831
24.5.2.2.1	Morpholinos.....	832
24.5.2.2.2	Reverse Genetic Screening by TILLING	832
24.5.2.2.3	Transgenesis.....	832
24.5.2.3	Physical Manipulation: Chemical Treatment, Environmental Stressor, and Infection.....	833
24.5.3	Fish Models of Human Disorders.....	834
24.5.3.1	Blood Disorders	834
24.5.3.2	Heart Disorders	837
24.5.3.3	Cancers	838
24.5.4	Modeling Human Disorders in Fish for Drug Discovery	840
	References.....	841

25	Genetic Counseling and Prenatal Diagnosis.....	845
	Tiemo Grimm and Klaus Zerres	
25.1	Genetic Counseling.....	845
25.1.1	Origins and Goals of Genetic Counseling	845
25.1.1.1	Definition of Genetic Counseling	846
25.1.1.2	Origins of Genetic Counseling	846
25.1.1.3	Indications for Genetic Counseling	846
25.1.2	Genetic Diagnosis.....	848
25.1.2.1	Molecular Diagnosis.....	848
25.1.2.2	Heterozygote Detection	850
25.1.3	Recurrence Risk.....	851
25.1.3.1	Recurrence Risk for Multifactorial Conditions	851
25.1.3.2	Reproductive Options and Alternatives	851
25.1.3.3	Predictive Testing.....	851
25.1.4	Communication and Support	852
25.1.5	Directive Vs. Nondirective Genetic Counseling.....	852
25.1.6	Assessment of Genetic Counseling and Psychosocial Aspects	853
25.2	Prenatal Diagnosis	854
25.2.1	Indications for Prenatal Diagnosis.....	854
25.2.1.1	Risk Assessment for Chromosomal Disorders	854
25.2.1.2	Previous Aneuploidy.....	855
25.2.1.3	Parental Chromosomal Rearrangements.....	855
25.2.1.4	Family History of a Monogenic Disorder.....	856
25.2.1.5	Neural Tube Defects	856
25.2.1.6	Psychological Indications	856
25.2.2	Techniques of Prenatal Diagnosis.....	856
25.2.2.1	Investigations Prior to Implantation.....	856
25.2.2.2	Noninvasive Diagnosis During Pregnancy	858
25.2.2.3	Invasive Diagnostics During Pregnancy	859
25.2.2.4	Special Problems with Prenatal Diagnosis	860
25.3.2.5	Consultation and Aftercare	861
25.3.2.6	Aftercare	862
25.3.2.7	Psychological Aspects and Outlook	862
25.3	Conclusions.....	863
25.4	Appendix Example of a Bayesian Table.....	863
	References.....	864
26	Gene Therapy	867
	Vivian W. Choi and R. Jude Samulski	
26.1	Advent of Gene Therapy.....	868
26.2	Diseases Considered Suitable for Gene Therapy.....	868
26.3	Gene Therapy Vectors.....	868
26.3.1	Viral Vectors	868
26.3.1.1	Adenovirus.....	868
26.3.1.2	Retrovirus.....	869
26.3.1.3	Adeno-associated Virus	869
26.3.1.4	Herpes Simplex Virus	869

26.3.2	Nonviral Vectors	869
26.3.2.1	Naked DNA	870
26.3.2.2	Liposomes.....	870
26.3.2.3	Polymers	870
26.4	Factors Affecting the Success of Gene Therapy.....	870
26.4.1	Choice of Gene Delivery Vector.....	870
26.4.2	Route of Administration	870
26.4.3	Integrating or Nonintegrating Vectors	871
26.4.4	Therapeutic Means.....	871
26.4.4.1	Gene Complementation	871
26.4.4.2	Gene Knockdown	871
26.4.4.3	Gene Correction.....	871
26.4.5	Regulation of Gene Expression	871
26.5	Safety Issues of Gene Therapy	872
26.6	Difficulties in Achieving Successful Gene Therapy.....	872
26.7	Conclusions.....	873
	References.....	873

27 Cloning in Research and Treatment of Human Genetic Disease 875

Ian Wilmut, Jane Taylor, Paul de Sousa, Richard Anderson,
and Christopher Shaw

27.1	Introduction.....	875
27.2	The Value of Cell Lines of Specific Genotype	876
27.2.1	Studies of Inherited Disease	876
27.2.2	Cells for Therapy	877
27.3	Somatic Cell Nuclear Transfer	878
27.3.1	Procedure for Nuclear Transfer	878
27.3.2	Present Successes and Limitations	878
27.4	Cells from Cloned Embryos	879
27.4.1	Cells from Cloned Mouse Embryos	879
27.4.2	Derivation of Cells from Cloned Human Embryos	879
27.4.3	Interspecies Nuclear Transfer	880
27.5	Direct Reprogramming of Somatic Cells	880
27.6	Looking to the Future	881
	References.....	881

28 Genetic Medicine and Global Health..... 885

David J. Weatherall

28.1	Introduction.....	886
28.2	Environmental Factors and Genetic Disease	886
28.2.1	Poverty and the Epidemiological Transition.....	886
28.2.2	Why Are Genetic Disease and Congenital Malformation Commoner in the Developing Countries?.....	887
28.2.2.1	Selection	887
28.2.2.2	Consanguineous Marriage	887
28.2.2.3	Parental Age.....	887

28.2.2.4	Population Migration	887
28.2.2.5	Poverty and Dysfunctional Healthcare Systems	888
28.3	Global Burden of Genetic Disease and Congenital Malformation	888
28.4	Monogenic Disease.....	889
28.4.1	Inherited Disorders of Hemoglobin.....	889
28.4.1.1	Global Distribution	889
28.4.1.2	Frequency.....	890
28.4.1.3	Population Genetics and Dynamics	892
28.4.1.4	Clinical Load and Cost–Benefit Issues of Control and Management Posed by the Inherited Hemoglobin Disorders.....	892
28.4.2	Other Monogenic Diseases in the Developing Countries	894
28.5	Communicable Disease	895
28.5.1	Infectious Agents	895
28.5.2	Vectors	896
28.5.3	Varying Susceptibility.....	896
28.5.4	Pharmacogenetics and Treatment	896
28.6	Genetic Components of Other Common Diseases: A Global View	897
28.7	Global Control of Genetic Disease	898
28.7.1	Transcultural, Ethical, and Counseling Issues	898
28.7.1.1	Ethnic Differences in Interpreting the Nature of Disease.....	898
28.7.1.2	Gender Issues.....	898
28.7.1.3	Patient Discrimination	898
28.7.1.4	Informed Consent	898
28.7.1.5	Lack of Regulatory or Ethical Bodies.....	899
28.7.1.6	Biobanks and Biopiracy.....	899
28.7.2	Genetic Services in Developing Countries	899
28.7.3	Organizing International Help for Developing Genetic Programs in Poorer Countries	900
	References.....	900
29	Genetic Databases	903
	Introductory Note by Stylianos E. Antonarakis	
29.1	Databases and Genome Browsers	905
	Rachel A. Harte , Donna Karolchik , Robert M. Kuhn , W.James Kent , and David Haussler	
29.1.1	Historical Background	906
29.1.2	Database Organization.....	907
29.1.3	Genome Annotations	907
29.1.3.1	Overview of Tracks in the Genome Browser.....	907
29.1.3.1.1	Mapping and Sequencing Tracks.....	910
29.1.3.1.2	Phenotype and Disease Associations.....	911
29.1.3.1.3	Gene and Gene Prediction Tracks and mRNA and EST Tracks.....	911
29.1.3.1.4	Expression and Regulation	911

29.1.3.1.5	Variation and Repeats	911
29.1.3.2	Comparative Genomics Tracks.....	912
29.1.3.2.1	Chains, Nets, and Conservation.....	912
29.1.3.2.2	Cross-Species Protein Alignments	912
29.1.3.3	UCSC Genes Set.....	912
29.1.4	Displaying and Sharing Data Using Custom Annotation Tracks.....	913
29.1.5	Example Analysis Using the Genome Browser.....	913
29.1.6	Using BLAT for Genome-Wide Alignments	916
29.1.7	Table Browser	916
29.1.7.1	Overview.....	916
29.1.7.2	Example Using the Table Browser	917
29.1.8	Tools.....	917
29.1.8.1	Introduction.....	917
29.1.8.2	In Silico PCR	918
29.1.8.3	Lifting Coordinates Between Assemblies	918
29.1.8.4	Gene Sorter	918
29.1.8.5	Proteome Browser.....	918
29.1.8.6	Genome Graphs	919
29.1.9	Further Information	919
29.1.10	Future Directions	919
	References.....	920
29.2	Ensembl Genome Browser.....	923
	Xosé M. Fernández and Ewan Birney	
29.2.1	Genomes Galore	924
29.2.1.1	Genomes in Context.....	924
29.2.1.2	ENCODE: Shifting the Paradigm.....	926
29.2.2	Ensembl Annotation	926
29.2.3	Region in Detail: Introduction	927
29.2.3.1	Region in Detail: Features	927
29.2.3.2	Region in Detail: Repeats, Decorations, and Export	927
29.2.4	DAS Sources.....	927
29.2.5	Comparative Genomics.....	929
29.2.6	GeneView.....	930
29.2.7	Variation.....	932
29.2.7.1	Variation Image.....	932
29.2.7.2	Comparison Image.....	932
29.2.8	Ensembl: An Example	932
29.2.9	BioMart Overview	933
29.2.10	Customizing Ensembl.....	934
29.2.10.1	Displaying User Data on Ensembl.....	935
29.2.11	Archive.....	935
29.2.12	Pre! Site.....	935
29.2.13	Further Information	935
29.2.14	Outlook	935
	References.....	936

29.3	Databases in Human and Medical Genetics	941
	Roberta A. Pagon, Ada Hamosh, Johan den Dunnen, Helen V. Firth, Donna R. Maglott, Stephen T. Sherry, Michael Feolo, David Cooper, and Peter Stenson	
29.3.1	GeneTests.....	942
29.3.1.1	Name and URL.....	942
29.3.1.2	Background.....	942
29.3.1.3	Purpose and Target Audiences.....	943
29.3.1.4	Location	943
29.3.1.5	Funding and Governance	943
29.3.1.6	Contents	943
29.3.1.7	Search Mechanisms and Search Results.....	944
29.3.1.8	Data Maintenance	945
29.3.1.9	Usage	945
29.3.1.10	Future Issues	945
29.3.2	Online Mendelian Inheritance in Man.....	945
29.3.2.1	Name and URL.....	945
29.3.2.2	Background.....	945
29.3.2.3	Purpose and Target Audiences.....	945
29.3.2.4	Location	946
29.3.2.5	Funding and Governance	946
29.3.2.6	Content.....	946
29.3.2.7	Search Mechanisms and Search Results.....	947
29.3.2.8	Data Maintenance	947
29.3.2.9	Usage	947
29.3.2.10	Future Issues	947
29.3.3	HGVS, Locus-Specific Databases	947
29.3.3.1	Name and URL.....	948
29.3.3.2	Background.....	948
29.3.3.3	Purpose and Target Audiences.....	948
29.3.3.4	Location	948
29.3.3.5	Funding and Governance	949
29.3.3.6	Content.....	949
29.3.3.7	Search Mechanisms and Search Results.....	949
29.3.3.8	Data Maintenance	949
29.3.3.9	Usage	949
29.3.3.10	Future Issues	949
29.3.4	Decipher.....	950
29.3.4.1	Name and URL.....	950
29.3.4.2	Background.....	950
29.3.4.3	Purpose and Target Audiences.....	950
29.3.4.4	Location	950
29.3.4.5	Funding and Governance	950
29.3.4.6	Contents	950
29.3.4.7	Search Mechanisms and Search Results.....	951
29.3.4.8	Data Maintenance	951
29.3.4.9	Usage	952
29.3.4.10	Future Issues	953

29.3.5	Entrez Gene	953
29.3.5.1	Name and URL.....	953
29.3.5.2	Background.....	953
29.3.5.3	Purpose and Target Audiences.....	953
29.3.5.4	Location	953
29.3.5.5	Funding and Governance	953
29.3.5.6	Contents	954
29.3.5.7	Search Mechanisms and Search Results.....	954
29.3.5.8	Data Maintenance	954
29.3.5.9	Usage	954
29.3.5.10	Future Issues	954
29.3.6	Database of Genotype and Phenotype	954
29.3.6.1	Name and URL	954
29.3.6.2	Background.....	954
29.3.6.3	Purpose and Target Audiences.....	955
29.3.6.4	Location	955
29.3.6.5	Funding and Governance	955
29.3.6.6	Contents	955
29.3.6.7	Search Mechanisms and Search Results.....	955
29.3.6.8	Data Maintenance	956
29.3.6.9	Usage	956
29.3.6.10	Future Issues	956
29.3.7	The Human Gene Mutation Database.....	956
29.3.7.1	Name and URL	956
29.3.7.2	Background.....	956
29.3.7.3	Purpose and Target Audiences.....	956
29.3.7.4	Location	957
29.3.7.5	Funding and Governance	957
29.3.7.6	Contents	957
29.3.7.7	Search Mechanisms and Search Results.....	958
29.3.7.8	Data Maintenance	958
29.3.7.9	Usage	958
29.3.7.10	Future Issues	958
	References.....	959
	Subject Index	961

Contributors

Laurent Abel

Laboratory of Human Genetics of Infectious Diseases, Necker Branch,
Institut National de la Santé et de la Recherche Médicale, U550,
Necker Medical School, Paris, France
laurent.abel@inserm.fr

Brett S. Abrahams

Program in Neurogenetics and Department of Neurology
Semel Institute for Neuroscience and Behavior at the David Geffen School of Medicine
University of California at Los Angeles, Los Angeles, CA 90095-1769, USA
brett.abrahams@gmail.com

Alexandre Alcaïs

Laboratory of Human Genetics of Infectious Diseases, Necker Branch,
Institut National de la Santé et de la Recherche Médicale, U550,
Necker Medical School, Paris, France
alexandre.alcais@inserm.fr

Richard Anderson

Centre for Reproductive Biology, The Queen's Medical Research Institute,
47 Little France Crescent, Edinburgh, EH16 4TJ, UK

Stylianos E. Antonarakis

Department of Genetic Medicine and Development,
University of Geneva Medical School and University Hospitals of Geneva
1 rue Michel-Servet, 1211 Geneva
Switzerland
Stylianos.Antonarakis@unige.ch

Antonio Baldini

Institute of Genetics and Biophysics, National Research Council,
University of Naples Federico II, and the Telethon Institute of Genetics
and Medicine, Via Pietro Castellino 111, Naples 80131, Italy
baldini@cnr.igb.it

Terri H. Beaty

Department of Epidemiology, The Johns Hopkins Bloomberg
School of Public Health, 615 North Wolfe Street, Baltimore,
MD 21205, USA
tbeaty@jhsph.edu

Jacques S. Beckmann

Service and Department of Medical Genetics, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
Jacques.Beckmann@chur.ch

Thomas D. Bird

University of Washington, Geriatric Research Education and Clinical Center (182), VA Puget Sound Health Care System, 1660 S Columbian Way, Seattle, WA 98108, USA
tomnroz@u.washington.edu

Ewan Birney

EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

Alan H. Bittles

Edith Cowan University and Murdoch University, Perth, Perth, WA, Australia; Centre for Comparative Genomics, Murdoch University, South Street, Perth, WA 6150, Australia
abittles@ccg.murdoch.edu.au

Romulo M. Brena

Department of Surgery, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA 91101, USA

Carlos D. Bustamante

Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA
cdb28@cornell.edu

Ross Cagan

Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, Annenberg Building 25-40, New York, NY 10029, USA
Ross.Cagan@mssm.edu

Jean-Laurent Casanova

Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA
Jean-Laurent.Casanova@mail.rockefeller.edu

Vivian W. Choi

Department of Pharmacology, Gene Therapy Center, University of North Carolina, Chapel Hill, NC 27599, USA
Vivian_Choi@alumni.unc.edu

Sven Cichon

Institute of Human Genetics and Department of Genomics, Life & Brain Center, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany
sven.cichon@uni-bonn.de

Andrew G. Clark

Department of Molecular Biology and Genetics, Cornell University,
Ithaca, NY 14853, USA
ac347@cornell.edu

David N. Cooper

Institute of Medical Genetics, School of Medicine,
Cardiff University, Heath Park, Cardiff, CF14 4XN, UK
cooperdn@cardiff.ac.uk

Joseph F. Costello

Department of Neurological Surgery, University of California San Francisco,
Diller Family Comprehensive Cancer Center, San Francisco, CA 94143, USA
jcostello@cc.ucsf.edu

Michael Dean

National Cancer Institute, Frederick, MD 21702, USA
deanm@mail.nih.gov

Emmanouil T. Dermitzakis

Department of Genetic Medicine and Development,
University of Geneva Medical School, 1 Rue Michel-Servet, Geneva 1211,
Switzerland
emmanouil.dermitzakis@unige.ch

Francesca Ducci

Division of Psychological Medicine,
P063 Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK
Francesca.Ducci@kcl.ac.uk

Johan T. den Dunnen

Leiden University Medical Center, Human and Clinical Genetics,
S4-030, PO Box 9600, 2333RC Leiden, The Netherlands
ddunnen@HumGen.nl

Michael Feolo

National Center for Biotechnology Information, Bldg 45,
4AN.12B, Bethesda, MD 20894, USA
feolo@ncbi.nlm.nih.gov

Xosé M. Fernández

EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge, CB10 1SD, UK
xose@ebi.ac.uk

Helen V. Firth

Consultant Clinical Geneticist, Addenbrooke's Hospital, Box 134,
Cambridge, CB2 2QQ, UK
helen.firth@addenbrookes.nhs.uk

Jonathan Flint

Wellcome Trust Centre for Human Genetics, Roosevelt Drive,
Oxford, OX3 7BN, UK
jf@well.ox.ac.uk

Daniel H. Geschwind

Programs in Neurogenetics and Neurobehavioural Genetics,
Neurology Department, and Semel Institute for Neuroscience and Behavior,
David Geffen School of Medicine, University of California at Los Angeles,
Los Angeles, CA 90095-1769, USA
dhg@ucla.edu

David Goldman

Laboratory of Neurogenetics, NIAAA, NIH, 5625 Fishers Lane,
Room 3S32, MSC 9412, Rockville, MD 20852, USA
davidgoldman@mail.nih.gov

David B. Goldstein

Center for Human Genome Variation, Institute for Genome Sciences and Policy,
Duke University, Durham, NC 27708, USA
d.goldstein@duke.edu

Zhiyuan Gong

Department of Biological Sciences, 14 Science Drive 4,
National University of Singapore, Singapore 117543
dbsgzy@nus.edu.sg

Tiemo Grimm

Institute of Human Genetics,
Division of Medical Genetics, University of Würzburg, Biozentrum,
Am Hubland, 97074 Würzburg, Germany
tgrimm@biozentrum.uni-wuerzburg.de

Ryan N. Gutenkunst

Theoretical Biology and Biophysics, and Center for Nonlinear Studies,
Los Alamos National Laboratory,
Los Alamos, NM, USA
ryang@lanl.gov

Ada Hamosh

Institute of Genetic Medicine, Johns Hopkins University School of Medicine,
Blalock 1012D, 600 N. Wolfe St., Baltimore, MD 21287-4922, USA
ahamosh@mail.jhmi.edu

Min Han

Department of Molecular, Cellular, and Developmental Biology,
Howard Hughes Medical Institute, Campus Box 0347,
University of Colorado at Boulder, Boulder, CO 80309, USA
mhan@colorado.edu

Ross C. Hardison

Department of Biochemistry and Molecular Biology,
Center for Comparative Genomics and Bioinformatics,
304 Wartik Laboratory, The Pennsylvania State University,
University Park, PA 16802, USA
rch8@psu.edu

Rachel A. Harte

Center for Biomolecular Science and Engineering, Engineering 2,
University of California Santa Cruz, 1156 High Street, Santa Cruz,
CA 95064, USA
hartera@soe.ucsc.edu

David Haussler

Center for Biomolecular Science and Engineering, Engineering 2,
University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
haussler@soe.ucsc.edu

Michael Hofreiter

Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6,
04103 Leipzig, Germany
michi@palaeo.eu

Bernhard Horsthemke

Institut für Humangenetik, Universitätsklinikum Essen,
Hufelandstrasse 55, 45122 Essen, Germany
bernhard.horsthemke@uni-due.de

Ruth Johnson

Department of Molecular, Cell and Developmental Biology,
Mount Sinai School of Medicine, 1 Gustave L. Levy Place,
Annenberg Building 18-92, New York,
NY 10029, USA
ruth.johnson@mssm.edu

Donna Karolchik

Center for Biomolecular Science and Engineering, Engineering 2,
University of California Santa Cruz, 1156 High Street, Santa Cruz,
CA 95064, USA
donna.k@soe.ucsc.edu

Nicholas Katsanis

Department of Cell Biology, Duke University, 466 A Nanaline Duke Bldg,
Durham, NC 27710, USA
katsanis@jhmi.edu

W. James Kent

Center for Biomolecular Science and Engineering, Engineering 2,
University of California Santa Cruz, 1156 High Street, Santa Cruz,
CA 95064, USA
kent@soe.ucsc.edu

Muin J. Khoury

Office of Public Health Genomics, Centers for Disease Control
and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
muk1@cdc.gov

Robert M. Kuhn

Center for Biomolecular Science and Engineering, Engineering 2,
University of California Santa Cruz, 1156 High Street, Santa Cruz,
CA 95064, USA
kuhn@soe.ucsc.edu

Siew Hong Lam

Department of Biological Sciences, 14 Science Drive 4,
National University of Singapore, Singapore 117543
dbslsh@nus.edu.sg

Donna R. Maglott

Natl Ctr Biotechnology Info, 45 Center Dr, MSC 6510,
Bethesda, MD 20892-6510, USA
maglott@ncbi.nlm.nih.gov

Arno G. Motulsky

Departments of Medicine (Division of Medical Genetics)
and Genome Sciences, University of Washington School of Medicine
Box 355065, 1705 N.E. Pacific St., Seattle, WA 98195-5065, USA
agmot@u.washington.edu

Stefan Mundlos

Institute for Medical Genetics,
Charité, Universitätsmedizin Berlin, Augustenburger Platz 1,
13353 Berlin, Germany; Forschungsgruppe Development & Disease,
Max Planck Institute for Molecular Genetics, Ihnestraße 73,
14195 Berlin, Germany
stefan.mundlos@charite.de

David L. Nelson

Molecular and Human Genetics, Baylor College of Medicine,
One Baylor Plaza, Rm. 902E, Houston, TX 77030, USA
nelson@bcm.tmc.edu

Paola Nicoletti

Center for Human Genome Variation, Institute for Genome Sciences and Policy,
Duke University, Durham, NC, USA; Department of Clinical and Preclinical
Pharmacology, University of Florence, Florence, Italy

Markus M. Nöthen

Institute of Human Genetics and Department of Genomics,
Life & Brain Center, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn,
Germany
markus.noethen@uni-bonn.de

Elaine A. Ostrander

National Human Genome Research Institute, National Institutes of Health,
50 South Drive, MSC 8000, Building 50, Room 5351, Bethesda,
MD 20892-8000, USA
eostrand@mail.nih.gov

Roberta A. Pagon

University of Washington, GeneTests, 9725 Third Ave NE, Suite 602,
Seattle, WA 98115-2024, USA
bpagon@u.washington.edu

Heidi G. Parker

National Human Genome Research Institute, National Institutes of Health,
50 South Drive, MSC 8000, Building 50, Room 5334, Bethesda,
MD 20892-8000, USA
hgarker@mail.nih.gov

George P. Patrinos

Department of Pharmacy, School of Health Sciences, University of Patras,
University Campus, Rion, 26504 Patras, Greece;
MGC-Department of Cell Biology and Genetics, Faculty of Medicine and Health
Sciences, Erasmus University Medical Center, P.O. Box 2040, 3000 CA,
Rotterdam, The Netherlands
g.patrinos@erasmusmc.nl;
g.patrinos@upatras.gr

Sohini Ramachandran

Society of Fellows, Harvard University, Cambridge, MA, USA;
Department of Organismic and Evolutionary Biology, Harvard University,
Cambridge, MA, USA
sramach@fas.harvard.edu

Marcella Rietschel

Department of Genetic Epidemiology in Psychiatry,
Central Institute of Mental Health Mannheim,
University of Heidelberg J5, 68159 Mannheim,
Germany
Marcella.Rietschel@zi-mannheim.de

Jon F. Robinson

McKusick-Nathans Institute of Genetic Medicine,
Department of Molecular Biology and Genetics,
Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

R. Jude Samulski

Department of Pharmacology,
Gene Therapy Center, University of North Carolina,
Chapel Hill, NC 27599, USA
rjs@med.unc.edu

Christine Schmael

Department of Genetic Epidemiology in Psychiatry,
Central Institute of Mental Health Mannheim,
University of Heidelberg J5, 68159 Mannheim,
Germany
c.schmael@gmx.de

Christopher Shaw

Department of Neurology, King's College London School of Medicine,
London, UK

Stephen T. Sherry

Building 38 – Natl Library of Medicine, B1E09B, 38 Center Dr, Bethesda,
MD 20892-6510, USA
sherry@ncbi.nlm.nih.gov

Paul de Sousa

Scottish Centre for Regenerative Medicine, Chancellor' s Building,
University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, UK

Michael R. Speicher

Institute of Human Genetics, Medical University of Graz,
Harrachgasse 21/8, 8010 Graz, Austria
michael.speicher@medunigraz.at

Peter Stenson

Institute of Medical Genetics, School of Medicine,
Cardiff University, Heath Park, Cardiff,
CF14 4XN, UK
stensonpd@cardiff.ac.uk

Hua Tang

Department of Genetics, Stanford Medical School, Stanford, CA, USA
huatang@stanford.edu

Jane Taylor

Scottish Centre for Regenerative Medicine, Chancellor' s Building,
University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, UK

Ian Tomlinson

Molecular and Population Genetics,
Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK
iant@well.ox.ac.uk

Morgan Tucker

Howard Hughes Medical Institute, Department of Molecular,
Cellular, and Developmental Biology, Campus Box 0347,
University of Colorado at Boulder, Boulder, CO 80309, USA
mtucker@colorado.edu

Nicole M. Walley

Department of Biology, Duke University, Durham, NC, USA; Center for Human
Genome Variation, Institute for Genome Sciences and Policy, Duke University,
Durham, NC, USA
nicole.walley@duke.edu

Sophia S. Wang

Division of Cancer Epidemiology and Genetics, National Cancer Institute,
National Institutes of Health, 6120 Executive Boulevard, Room 5104, Rockville,
MD 20852, USA
wangso@mail.nih.gov

David J. Weatherall

University of Oxford,
Weatherall Institute of Molecular Medicine, John Radcliffe Hospital,
Headington, Oxford, OX3 9DS, UK
liz.rose@imm.ox.ac.uk

Saffron Willis-Owen

Molecular Genetics, National Heart and Lung Institute, Guy Scadding Building,
Dovehouse Street, London, SW3 6LY, UK
s.willis-owen@imperial.ac.uk

Ian Wilmut

Scottish Centre for Regenerative Medicine, Chancellor' s Building,
University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, UK
Ian.Wilmut@ed.ac.uk

Klaus Zerres

Institute of Human Genetics,
Medical Faculty, RWTH University
Pauwelsstr. 30, 52074 Aachen, Germany
kzerres@ukaachen.de

Introduction

Human Genetics as Fundamental and Applied Science

Human genetics is both a fundamental and an applied science. As a fundamental science, it is part of genetics – the branch of science that examines the laws of storage, transmission, and realization of information for development and function of living organisms. Within this framework, human genetics concerns itself with the most interesting organism – the human being. This concern with our own species makes us scrutinize scientific results in human genetics not only for their theoretical significance but also for their practical value for human welfare. Thus, human genetics is also an applied science. Its value for human welfare is bound to have repercussions for theoretical research as well, since it influences the selection of problems by human geneticists, their training, and the financing of their research. Because of its continued theoretical and practical interest, human genetics offers fascination and human fulfillment unparalleled by work in fields that are either primarily theoretical or entirely practical in subject matter.

Science of Genetics

Genetics is based on a powerful and penetrating theory. The profundity of a theory depends on the depth of the problems that it sets out to solve and can be characterized by three attributes: the occurrence of high-level constructs, the presence of a mechanism, and high explanatory power [1]. In genetics, the high-level “construct” is the gene as a unit of storage, transmission, and realization of information. Since the rediscovery of Mendel’s laws in 1900, genetic mechanisms

have been worked out step by step to the molecular level – deciphering of the genetic code, analysis of transcription and translation, the function of gene-determined proteins, the fine structure of the genetic material, and DNA sequences outside of genes. The problems of regulation of gene activity in the development and function of organisms are currently a principal goal of fundamental research. So far, the explanatory power of the theory has not nearly been exhausted.

How Does a Science Develop?

Kuhn (1962) [10] described the historical development of a science as follows: In the early, protoscientific stage, there is substantial competition among various attempts at theoretical foundation and empirical verification. Basic observations suggest a set of problems that, however, is not yet visualized clearly. Then, one “paradigm” unifies a group within the scientific community in the pursuit of a common goal, at the same time bringing into sharper focus one or a few aspects of the problem field, and suggesting a way for their solution. If the paradigm turns out to be successful, it is accepted by an increasing part of the scientific community, which now works under its guidance, exploring its possibilities, extending its range of application, and developing it into a scientific theory.

This concept of a paradigm has three main connotations:

1. It points to a piece of scientific work that serves as an “exemplar,” suggesting ways in which a certain problem should be approached.
2. It delimits a group of scientists who try to explore this approach, expand its applicability, deepen its

- theoretical basis by exploration of basic mechanisms, and enhance its explanatory power.
3. Finally, while an elaborate theory must not – and, in most cases, does not – exist when a paradigm is initiated, its germ is already there, and a successful paradigm culminates in the elaboration of this theory.

This process of developing a science within the framework of a paradigm has been described by Kuhn as “normal science.” The basic theory is taken increasingly for granted. It would be sterile at this stage to doubt and reexamine its very cornerstones; instead, it is applied to a variety of problems, expanded in a way that is comparable to puzzle solving. From time to time, however, results occur that, at first glance, defy explanation. First, this leads to attempts at accommodating such results within the theoretical framework by additional ad hoc hypotheses. These attempts are often successful; sometimes, however, they fail. If in such a situation an alternative paradigm is brought forward that explains most of the phenomena accounted for by the old theory as well as the new, hitherto unexplained phenomena, a scientific “revolution” may occur. The new paradigm gains support from an increasing majority of the scientific community, it soon develops into a new – more explanatory – theory, and the process of normal science begins anew.

This portrayal of scientific development has been criticized by some philosophers of science [11]. The concept of “normal” science as outlined above does not appeal to some theorists. Working within the framework of a given set of concepts has been denounced as dull, boring, and in any case not as science should be. According to these philosophers, scientists ought to live in a state of permanent revolution, constantly questioning the basic foundations of their field, always eager to put them to critical tests and, if possible, to refute them [15–18]. Many scientists actively involved in research, on the other hand, have readily accepted Kuhn’s view; he has apparently helped them to recognize some important aspects in the development of their own fields.

Central Theory of Genetics Looked at as a Paradigm

While Kuhn’s concepts were developed on the basis of the history of the physical sciences, his description well fits the development of genetics. Up to the second half of the nineteenth century, the phenomena of heredity eluded

analysis. Obviously, children were sometimes – but by no means always – similar to their parents; some diseases were shown to run in families; it was possible to improve crops and domestic animals by selective breeding. Even low-level laws were discovered, for example Nasse’s law that hemophilia affects only boys but is transmitted by their mothers and sisters (Chap. 5, Sect. 5.1.4). However, a convincing overall theory was missing, and attempts at developing such a theory were unsuccessful. In this situation, Mendel, in his work *Versuche über Pflanzenhybriden* (1865) [12] first improved a procedure; he complemented the breeding experiment by counting the offspring. He then interpreted the results in terms of the random combination of basic units; by assuming these basic units, he founded the gene concept – the nuclear concept underlying genetic theory (Chap. 1, Sect. 1.4).

Since the rediscovery of his work in 1900, Mendel’s insight has served as a paradigm in all three connotations: it provided an exemplar as to how breeding experiments should be designed and evaluated, it resulted in the establishment of a scientific community of geneticists, and it led to the development of a deep and fertile scientific theory. A special problem that has not been answered satisfactorily, in our opinion, concerns the question of why acceptance of Mendel’s paradigm had to wait for as long as 35 years after these experiments were published. It would be too simplistic to blame academic arrogance and shortsightedness of contemporary biologists who did not want to accept the work of a “nonacademic” outsider, even if this factor may indeed have been one of the components for this neglect. We believe rather that the many new biological discoveries in the 35 years following Mendel’s discovery were of such a revolutionary nature as to qualify as a scientific crisis in the Kuhnian sense and therefore required a completely new approach.

Soon after the rediscovery of Mendel’s laws in 1900, however, an initially small, but quickly growing group of scientists gathered who developed genetics in an interplay between theory and experiment and launched the major scientific revolution of the twentieth century in the field of biology.

Human Genetics and the Genetic Revolution

Meanwhile, the biological revolution of the nineteenth century – evolutionary theory – had been accepted by the scientific community. One major consequence was

the realization that human beings had evolved from other, more “primitive” primates, that humans are part of the animal kingdom, and that the laws of heredity which had been found to apply for all other living beings are also valid for our species. Hence, Mendel’s laws were soon applied to traits that were found in human pedigrees – primarily hereditary anomalies and diseases. Analyzing the mode of inheritance of alkaptonuria – a recessive disease – Garrod (1902) [5] clearly recognized the cardinal principle of gene action: genetic factors specify chemical reactions (Chap. 1, Sect. 1.5). This insight also required 30 years before being incorporated into the body of “normal” science.

Elucidation of inheritance in humans did not begin with Mendel’s paradigm. Many relevant observations had been reported before, especially on various diseases. Moreover, another paradigm had been founded by F. Galton in his work on *Hereditary Talent and Character* (1865) [6] and in later works: to derive conclusions as to inheritance of certain traits such as high performance, intelligence, and stature, one should measure these traits as accurately as possible and then compare the measurements between individuals of known degree of relationship (for example, parents and children, sibs, or twins) using statistical methods. This approach did not contain the potential for elucidating the mechanisms of heredity. On the other hand, it seemed to be much more generally applicable to human characteristics than Mendelian analysis; pedigree analysis in terms of Mendel’s laws was hampered by the fact that most human traits simply could not be classified as alternate characteristics, as could round and shrunken peas. Human characteristics are usually graded and show no alternative distribution in the population. Moreover, the phenotypes are obviously determined not only by the genetic constitution but by external, environmental influences as well – the result of an interaction between “nature and nurture” (Galton). Therefore, naive attempts at applying Mendel’s laws to such traits were doomed to failure. For traits that are regarded as important, such as intelligence and personality, but also for many diseases and mental retardation, there was only the choice between research along the lines suggested by Galton or no research at all. Investigations on genetic mechanisms would have to await elucidations of the genetics of other, more accessible organisms. Under these circumstances, scientists chose to follow Galton. This choice had not only theoretical reasons; it was strongly influenced by the desire to help individuals and families by

calculating risk figures for certain diseases, thereby creating a sound basis for genetic counseling. More important, however, was the concern of some scientists about the biological future of the human species, which they saw threatened by deterioration due to relaxation of natural selection. The motives for their research were largely eugenic: it seemed to provide a rational foundation for measures to curb reproduction of certain groups who were at high risk of being diseased or otherwise unfit.

History of Human Genetics: A Contest Between Two Paradigms

The two paradigms – Mendel’s gene concept and Galton’s biometric approach – have developed side by side from 1900 up to the present; many present-day controversies, especially in the field of behavior genetics but also those concerning strategies in the genetic elucidation of common diseases, are immediately understandable when the history of human genetics is conceived as a contest between these two paradigms. This does not mean that the two paradigms are mutually exclusive; in fact, correlations between relatives as demonstrated by biometric analysis were interpreted in terms of gene action by Fisher in 1918 [4]. Some human geneticists have worked during some part of their career within the framework of the one paradigm, and during another within the framework of the other paradigm. By and large, however, the two streams of research have few interconnections and may even become further polarized because of highly specialized training for each group, epitomized by the biochemical and molecular genetic laboratories for the one and the computer for the other group.

In the first decades of the last century the biometric paradigm of Galton appeared to be very successful. Genetic variability within the human population was believed to be established for normal traits such as stature or intelligence as well as for a wide variety of pathologic conditions such as mental deficiency and psychosis, epilepsy, and common diseases such as diabetes, allergies, and even tuberculosis. Mendelian analysis, on the other hand, seemed to be confined to rare hereditary diseases; the ever repeated attempts at expanding Mendelian explanation into the fields of normal, physical characteristics and common diseases

were usually undertaken without critical assessment of the inescapable limitations of Mendelian analysis. The first major breakthrough of Mendelian genetics was the establishment of the three-allele hypothesis for the AB0 blood groups by Bernstein in the 1920s [2] (Chap. 5, Sect. 5.2.2). Further progress, however, had to await the development of genetic theory by work on other organisms such as *Drosophila*, bacteria, and viruses, especially bacteriophages.

The advent of molecular biology in the late 1940s and 1950s had a strong influence on human genetics and, indeed, brought the final breakthrough of Mendel's paradigm. A major landmark was the discovery by Pauling et al. in 1949 [14] that sickle cell anemia is caused by an abnormal hemoglobin molecule.

The foundation of human chromosome research in the late 1950s and early 1960s (Chap. 3, Sect. 3.1) came as a second, important step. At present, most investigations in human genetics have become a part of mainstream research within the framework of genetic theory. The human species, regarded by most early experimental geneticists as a poor tool for genetic research, is now displaying definite advantages for attacking basic problems. Some of these advantages are the large size of available populations, the great number and variety of known mutants and chromosome anomalies, and the unparalleled detailed knowledge of human physiology and biochemistry in health and disease. The improved understanding of human genome structure and its variability (Chap. 2) by the completion of the human genome project, by new sequencing and array technologies, and by efforts to identify all functional elements in the human genome sequence (Chap. 4), further facilitates both basic and applied research in human genetics.

One would expect that such breakthroughs have led to the establishment of Mendel's paradigm as the only leading paradigm in human genetics. This, however, is not the case. In spite of the fact that genetic theory is now pervading many fields that seemed to be closed to it, the paradigm of Galton – biometric analysis – has attained an unsurpassed level of formal sophistication over the past decades. The availability of software tools has greatly facilitated the development and application of biometric techniques. Moreover, in some fields, such as behavior genetics, the application of genetic theory – Mendel's paradigm – is still hampered by severe difficulties (Chap. 23), and here biometric

methods have dominated for a long time. In the same field, however, they are most severely criticized and subject to controversial discussions about ethical issues and possible discrimination.

Progress in Human Genetics and Practical Application

The achievements of molecular biology and chromosome research have not only altered human genetics as a pure science, but have also brought marked progress in its application for human welfare. At the beginning, this progress did not appear very conspicuous; the diagnosis of hereditary diseases was improved, and many, hitherto unexplained malformations were accounted for by chromosome aberrations. The first practical success came in the early 1950s when the knowledge of enzyme defects in phenylketonuria (Chap. 1) and galactosemia led to successful preventive therapy by a specific diet. However, a breakthrough on a much larger scale was achieved when the methods of prenatal diagnosis for chromosome aberrations and for some metabolic defects were introduced in the late 1960s and early 1970s (Chap. 25, Sect. 25.2). Suddenly, genetic counseling could now be based not only on probability statements but, in an increasing number of cases, on certainty of individual diagnoses. This scientific development coincided with a growing awareness in large parts of the human population that unlimited human reproduction must not be accepted as a natural law but can – and should – be regulated in a rational way. Introduction of oral contraceptive agents signaled this awareness. The chance to avoid the births of severely handicapped children is now accepted by a rapidly increasing proportion of the population. At the same time, better knowledge of pathophysiological pathways is improving the chances for individual therapy of hereditary diseases, including the promise somatic gene therapy by introduction of genes into cells of functional tissues (see Chap. 27). Applications of human genetics as a practical tool to prevent suffering and disease have found wide resonance and have now one of the most rewarding approaches in preventive medicine. In many countries, the politically responsible bodies have already created, or are now creating the institutions for widespread application of the new tools.

Effects of Practical Applications on Research

These practical applications have led to a marked increase in the number of research workers and the amount of work within the past decades. From the beginning of the twentieth century up to the early 1950s, human genetics had been the interest of a mere handful of scientists for most of whom it was not even a full-time occupation. Many of the pioneers were trained and worked much of their lifetime as physicians in special fields of medicine, such as Waardenburg and Franceschetti in ophthalmology, and Siemens in dermatology. Others were interested in theoretical problems of population genetics and evolution and chose problems in human genetics as the field of application for their theoretical concepts, most notably J.B.S. Haldane and R.A. Fisher. Still others had their point of departure in physical anthropology. This heterogeneous group of scientists did not form a coherent scientific community. For a long time, there was almost no formal infrastructure for the development of a scientific specialty. There were almost no special departments, journals, and international conferences. This lack of focus resulted in a marked heterogeneity in quality and content of scientific contributions.

All this has changed. Departments and units of human and medical genetics are now the standard in many countries; universities and medical schools offer special curricula, many journals and other publications exist, and numerous congresses and conferences are being held. Human genetics is now an active and vigorous field which continues to grow exponentially.

Dangers of Widespread Practical Application for Scientific Development

This development, however, satisfactory as it is, has also a number of potentially undesirable consequences:

- (a) Research is promoted primarily in the fields of immediate practical usefulness related to hereditary diseases; fields of less immediate practical importance may be neglected.
- (b) Initially the contact with fundamental research in molecular genetics and cell biology was not intensive enough. This may have led to a slowdown in

the transfer of scientific concepts and experimental approaches from these fields. Fortunately, this has changed with the advent of recombinant DNA techniques and many other methods. The speed with which results of basic research are being transferred into practical application has increased significantly.

- (c) As in other sciences, certain topics may evolve to a mainstream research where vast human and financial resources are being invested, drawing it off other areas, which are then neglected in spite of their great importance. For example, at present the immense activities to unravel complex disorders by high-throughput assays have resulted in a decreased interest in studying monogenic Mendelian disorders although their detailed analyses may provide invaluable insights into consequences of mutations and their associated pathophysiology (see Chap. 4, Sect. 4.1).
- (d) Much medical research applies established methods to answer straightforward questions. Many studies collect data with new techniques. Individual results are often not of great import, but the ensemble of such data are the essential building blocks for the future progress of normal science. Much of such work is being carried out in human and medical genetics and is quite essential for many medical and anthropological applications. However, there is continued need in human genetics to develop testable hypotheses and try to test their consequences from all viewpoints.

Human geneticists must not neglect the further development of genetic theory. Basic research is needed in fields in which the immediate practical application of results is not possible but might in the long run be at least as important for the future of the human species as current applications in diagnostic and preventive medicine.

Advantages of Practical Application for Research

The needs of medical diagnosis and counseling have also given strong incentives to basic research. Many phenomena that basic research tries to explain would simply be unknown had they not been uncovered by study of diseases. We would be ignorant regarding the

role of sex chromosomes in sex determination had there not been patients with sex chromosomal anomalies. Phenomena such as spontaneously enhanced chromosome instability in Fanconi's anemia or Bloom's syndrome with all its consequences for somatic mutation and cancer formation (Chap. 3, Sect. 3.7) were discovered accidentally in the process of examining certain patients for diagnostic reasons. Genetic analysis of the "supergene" determining the major histocompatibility complex in humans contributes much to our fundamental understanding of how the genetic material above the level of a single gene locus is structured, and how the high genetic variability within the human population can be maintained (Chap. 6, Sect. 6.2.5). However, research in this field would certainly be much less active had there not been the incentive of improving the chances of organ transplantation.

Whether we like it or not, society pays increasing amounts of money for research in human genetics because we want to have practical benefits. Hence, to promote basic research, we must promote widespread practical applications. To guarantee progress in practical application for the future as well – and not only in the field of medicine – basic research needs to be supported. This is also the only way to attract good research workers and to maintain – or even improve – scientific standards. This paradox creates priority problems for all those concerned with research planning.

Human Genetics and the Sociology of Science

The discussion above should have demonstrated that human genetics – as all other sciences – has not developed in a sociological vacuum, following only the inherent logical laws of growth of theory and experimental testing. Human genetics is the work of social groups of human beings who are subject to the laws of group psychology and are influenced by the society at large in their attitudes toward research and their selection of problems. Unfortunately, sociological investigations of group formation and structure in human genetics have not been carried out. Another group active in the foundation of molecular biology, that which introduced the bacteriophages of *Escherichia coli* into the analysis of genetic information, has been studied extensively [3].

We know from this and from other examples that, during a phase in which a new paradigm is being founded, the group that shares this paradigm establishes close within-group contacts. The normal channels of information exchange such as scientific journals and congresses are superseded by more informal information transfer through telephone calls, e-mail communications, pre-prints, and personal visits. Within the group, influential personalities serve as intellectual and/or organizational leaders. Outside contacts, on the other hand, are often loose. When the acute phase of the scientific revolution is over, the bonds within the group are loosened, and information is again exchanged largely by normal channels of publication.

Similar developments can be observed in the field of human genetics. For example, in Chap. 6 (Sect. 6.2.5) we sketch the groups active in the elucidation of the major histocompatibility complex and in the assignment of gene loci to chromosome segments (Sect. 6.1).

Of similar influence on population genetics has been the first "big science" research project in human genetics – the Atomic Bomb Casualty Commission (ABCC, now the Radiation Effects Research Foundation, RERF; www.rerf.or.jp) project that was launched in the late 1940s in Japan by American and Japanese research workers to examine the genetic consequences of the atomic bombs in Hiroshima and Nagasaki (Chap. 10). In later years, this project led, for example, to comprehensive studies of the genetic effects of parental consanguinity. The second endeavor of this type is the "Human Genome Project" – the attempt at analyzing and sequencing the entire human genome by coordinated international cooperation (see Chap. 44). Today many research efforts are being conducted and can only be accomplished in large, international consortia, as for example the ENCYclopedia Of DNA Elements (ENCODE) project (www.genome.gov) or the Functional Annotation of the Mammalian Genome (FANTOM) project (fantom.gsc.riken.jp).

Many, if not most of the more interesting developments in the field were not initiated by investigators who would declare themselves human geneticists, or who worked in human genetics departments. They were launched by research workers from other fields such as general cytogenetics, cell biology, molecular biology, biochemistry, and immunology, but also from clinical specialties such as pediatrics, hematology, ophthalmology, and psychiatry. A common theme running through many recent developments has been the

application of nongenetic techniques from many different fields such as biochemistry and immunology to genetic concepts. On the other hand, techniques originally developed for solving genetic problems, especially for molecular studies of DNA, are being introduced at a rapidly increasing rate into other fields of research, for example in both medical research and practical medicine. In fact, most recent progress in human genetics comes from such interdisciplinary approaches. The number of research workers in the field has increased rapidly. Most did not start as human geneticists but as molecular biologists, medical specialists, biochemists, statisticians, general cytogeneticists, etc. They were drawn into human genetics in the course of their research. This very variety of backgrounds makes discussions among human geneticists stimulating and is one of the intellectual assets of the present state of our field. However, such diversity is also a liability as it may lead to an overrating of one's small specialty at the expense of a loss of an overview of the whole field [8]. With increasing complexity of research methods, specialization within human genetics has become inevitable. However, this brings with it the danger that the outlook of the scientist narrows, whole fields are neglected, and promising research opportunities remain unexploited.

Human Genetics in Relation to Other Fields of Science and Medicine

The rapid development of human genetics during recent decades has created many interactions with other fields of science and medicine. Apart from general and molecular genetics and cytogenetics, these interactions are especially close with cell biology, biochemistry, immunology, and – with many clinical specialties. Until recently, on the other hand, there have been few if any connections with physiology. One reason for this failure to establish fruitful interactions may be a difference in the basic approach: genetic analysis attempts to trace the causes of a trait to its most elementary components. Geneticists know in principle that the phenotype is produced by a complex net of interactions between various genes, but they are interested more in the components than in the exact mechanism of such interactions. At present, genetic analysis has reached the level of gene structure and the genetic

code; a final goal would be to explain the properties of this code in terms of quantum physics. A malevolent observer might compare the geneticist with a man who, to understand a book, burns it and analyzes the ashes chemically.

The physiologist, on the other hand, tries to read the book. However, he often presupposes that every copy of the book should be exactly identical; variation is regarded as a nuisance. To put it differently, physiology is concerned not with the elements themselves but with their mode of interaction in complicated functional systems (see Mohr [13]). Physiologists are more concerned with the integration of interacting systems than with the analysis of their components. The analysis of regulation of gene activities by feedback mechanisms, for example, the Jacob-Monod model in bacteria, and some approaches in developmental genetics of higher organisms have taught geneticists the usefulness of thinking in terms of systems. On the other hand, methods for molecular analysis of DNA have been introduced into physiology at an increasing scale. Genes for receptors and their components, for example for neurotransmitters, and genes for channel proteins are being localized in the genome and analyzed at the molecular level. Hence, the gulf between physiology and genetics is now being bridged. With the increasing interest of human geneticists in the genetic basis of common diseases and individual genetic variation in response to influences such as nutrition and stress, genetic concepts are increasingly influencing the many branches of medicine that, in the past, have profited relatively little from genetic theory. Molecular biology is developing increasingly into a common basis for many branches of science, and most biomedical scientists are nowadays becoming better acquainted with the principles of genetics. A field of molecular medicine is emerging.

Fields of Human and Medical Genetics

The field of human genetics is large, and its borders are indistinct. The development of different techniques and methods has led to the development of many fields of subspecialization. Many of these overlap and are not mutually exclusive. The field of *human molecular genetics* has its emphasis in the identification and analysis of genes at the DNA level. Methods such as DNA digestion by restriction endonucleases, Southern blotting,

polymerase chain reaction (PCR), sequencing and many others are being applied. *Human biochemical genetics* deals with the biochemistry of nucleic acids, proteins, and enzymes in normal and mutant individuals. Laboratory methods of the biochemist are being used (e.g., chromatography; enzyme assays). *Human cytogenetics* deals with the study of human chromosomes in health and disease. *Immunogenetics* concerns itself largely with the genetics of blood groups, tissue antigens such as the HLA types, and other components of the immune system. *Formal genetics* studies segregation and linkage relationships of Mendelian genes and investigates more complex types of inheritance by statistical techniques.

Clinical genetics deals with diagnosis, prognosis, and to some extent treatment of various genetic diseases. Diagnosis requires knowledge of etiological heterogeneity and acquaintance with many disease syndromes. *Genetic counseling* is an important area of clinical genetics and requires skills in diagnosis, risk assessment, and interpersonal communication. *Population genetics* deals with the behavior of genes in large groups and concerns the evolutionary forces of drift, migration, mutation, and selection in human populations. The structure and gene pool of human populations are studied by considering gene frequencies of marker genes. In recent years population geneticists have become interested in the epidemiology of complex genetic disease that require biometric techniques for their studies. *Behavioral genetics* is a science that studies the hereditary factors underlying behavior in health and disease. Behavior geneticists attempt to work out the genetic factors determining personality and cognitive skills in human beings. The genetics of mental retardation and various psychiatric diseases are also considered. The field of sociobiology tries to explain social behavior by using biological and evolutionary concepts.

Somatic cell genetics is the branch of human genetics that studies the transmission of genes at the cellular level. Cell hybridization between different species is an important tool for the cartography of human genes. *Developmental genetics* studies genetic mechanisms of normal and abnormal development. This field employs to a large extent model organisms and has a strong emphasis on animal experimentation. *Reproductive genetics* is the branch of genetics that studies details of gamete and early embryo formation by genetic techniques. This area is closely related to reproductive physiology. Due to the growing application of assisted

reproductive technologies in couples with infertility disorders this field has recently grown significantly. *Pharmacogenetics* deals with genetic factors governing the disposal and kinetics of drugs in the organism. Special interest in human pharmacogenetics relates to adverse drug reactions. *Ecogenetics* is an extension of pharmacogenetics and deals with the role of genetic variability affecting the response to environmental agents.

Clinical genetics has grown very rapidly in recent years because of the many practical applications of diagnosis and counseling, intrauterine diagnosis, and screening for genetic disease. Most research in human genetics is currently carried out in clinical genetics, cytogenetics, molecular and biochemical genetics, somatic cell genetics, and immunogenetics under medical auspices. Research in formal and population genetics has benefited enormously from the increasing knowledge about genome structure and its variation and the availability of new, cheaper high-throughput sequencing approaches.

Future of Human Genetics

Research methods in science are becoming ever more complicated and expensive, and human genetics is no exception. As a necessary consequence mastering of these methods increasingly requires specialization in a narrow field. Purchase of big instruments creates financial difficulties. Hence, the selection of research problems is often directed not by the intrinsic scientific interest in the problems or the conviction that they could, in principle, be solved, but by the availability of research methods, skilled coworkers, and instruments. Many research projects require large patient cohorts and complex, genome-wide analyses, tasks of a magnitude that can only be performed within international consortia. Such efforts are greatly facilitated by web-based databases (Sects. 29.1–29.3) which provide an easy means for distributing results to the genetic community. Furthermore, such databases ensure that new evolving information can easily be utilized by other persons in the field. For example, data on copy number variation in the human genome and possible consequences for the phenotype are now rapidly assessable in databases (Sects. 29.2 and 29.3) and are thus available for genetic counselors who can use this knowledge to provide their patients with detailed up-to-date information.

However, the tendency toward specialization will inevitably continue, and it is possible that, in this process, important parts of human genetics will be resolved into fields mainly defined by research methods, such as biochemistry, chromosome research, immunology, molecular biology [see 12], or into certain clinical areas. For example, hereditary metabolic diseases or syndromes associated with dysmorphic features and developmental delay are often studied and treated by pediatricians with little genetic training. Several departments of neurology have established their own neurogenetics branches, which are often independent from the respective department of human genetics. However, despite this tendency toward subspecialization, it is important to note that a laboratory performing genetic diagnostic procedures needs trained and experienced personnel, up-to-date equipment, and has to fulfill internationally defined quality standards, which are regulated by law in many countries. Therefore, it is probably not cost-effective to perform genetic diagnostics in small laboratories that offer only a few tests. Therefore, large laboratories performing all important human genetics diagnostic procedures may evolve to organizational structures in which human genetics remains united.

Survival of an established field of science has no value in itself. If a field dies because its concepts and accomplishments have been accepted and are being successfully integrated into other fields, little is lost. In human genetics, however, this state has not been reached yet and it may never get to this point. Many concepts of molecular biology, often in combination with “classical” methods such as linkage analysis, are now being applied to humans. A few decades ago human genetics was a medical field mainly dealing with rare syndromes and prenatal diagnostics. This picture has completely changed as the genetic contributions to common diseases are increasingly being unraveled. For example, genetic counseling is now an integral part of care in families with hereditary cancer diseases (Chap. 14) or neurologic disorders. In addition, data evolving from genome-wide association studies (GWAS) have identified numerous new loci in the genome that may change the susceptibility for diseases or phenotypic features. The effect of these loci may often be only moderate (Sect. 8.1), however, the evolving knowledge may further increase requests for genetic counseling. In future, genetic counseling provided by professionals in the field may have to com-

pete with “direct-to-consumer genetic testing” over the Internet that is already offered by several companies. Such developments are accompanied by growing options for predictive genetic diagnosis, which require standardized procedures for both the counseling session and the molecular genetic testing and which often involve difficult ethical issues. Thus, the tasks in human genetics have changed tremendously over the past decades and new challenges are constantly arising in this rapidly evolving field. Newly evolving technologies, such as whole-genome sequencing (see below), will further expand the future of human genetics. In fact, it can be predicted that human genetics will change medicine, as it has the potential to identify persons with an increased risk for certain diseases and it may provide information about treatment options. These aspects are now often referred to as “personalized medicine” (Chap. 4, Sect. 4.4) and they will likely dominate medicine in upcoming years.

Unsolved and Intriguing Problems

With the rapid increase in knowledge over recent years new and often unexpected problems have arisen. At a time when hereditary traits were defined by their modes of inheritance, the relationship between genotype and phenotype appeared relatively simple. This straightforward relationship seemed correct when some hereditary diseases were shown to be caused by enzyme defects, and when hemoglobin variants turned out to be due to amino acid replacements caused by base substitutions. With increasing knowledge of the human genome, however, many hereditary traits with phenotypes that had been considered identical turned out to be heterogeneous. These were caused either by mutations in different genes or by different mutations within the same genes. However, even mutations that are identical by the strictest molecular criteria sometimes have striking phenotypic differences. Analysis of such genotype–phenotype relationships by the study of genetic and environmental modifiers poses intriguing future problems in human genetics.

The establishment of genotype–phenotype relationships was recently further complicated by two new findings. The first finding represents the unanticipated variation within the human genome (Chap. 2). Future

research will have to elucidate how copy number variants (CNVs) contribute to human phenotypic diversity and disease susceptibility. CNVs are also of interest for a better understanding of the evolution of the genome, as they provide the raw material for gene duplication and gene family expansion. However, in addition to numerical variation there are extensive structural variations, such as inversions or insertions. Their impact on gene function remains to be elucidated. The second finding was the characterization of functional elements by the ENCODE consortium. To date, only 1% of the human genome has been analyzed by various high-throughput experimental and computational techniques; however, the findings revealed an unexpected number and complexity of the RNA transcripts that the genome produces. These findings have challenged traditional views about regulatory elements in the genome and added new insights into the complexity of human genetics, revealing that our understanding of the genome is still far from being complete. In order to address this, the National Human Genome Research Institute (NHGRI) launched two complementary programs in 2007: an expansion of the human ENCODE project to the whole genome (<http://www.genome.gov/ENCODE>) and the model organism ENCODE (modENCODE) project to generate a comprehensive annotation of the functional elements in the *Caenorhabditis elegans* and *Drosophila melanogaster* genomes (<http://www.modencode.org>; <http://www.genome.gov/modENCODE>). These efforts will likely contribute to a better understanding of genome complexity and gene regulation.

At present our understanding of somatic genome variability is very incomplete. Current concepts suggest that erroneous DNA repair and incomplete restoration of chromatin after damage may be resolved and may produce mutations and epimutations. Both mutations and epimutations have been shown to accumulate with age and such an increased burden of mutations and/or epimutations in aged tissues may increase cancer risk and adversely affect gene transcriptional regulation. This may in turn result in a progressive decline in organ function, a phenomenon frequently observed in aging. With the demographic trend of prolonged life expectancy, a better understanding of somatic genome variability and the stability of the genome may grow in importance.

Other problems may arise from new technologies, such as next-generation or third-generation whole-

genome sequencing (Chap. 4, Sect. 4.4), which will make sequencing of entire genomes possible and affordable within in a short period of time. These possibilities will require new bioinformatic tools and interpretation of sequencing results will greatly depend on whether we understand better the aforementioned relevance of structural and copy number variation and whether we can make sense of the various transcriptionally active regions in the genome. If we succeed, there is no doubt that whole-genome sequencing will change human genetics tremendously. They will, for example, contribute to a better understanding of modifier genes in monogenic diseases and thus explain the frequently observed phenotypic variability. Furthermore, they will contribute significantly to further propel research on complex diseases. However, although the new possibilities of human genetics are fascinating they raise at the same time new ethical issues. For example, in prenatal diagnostic settings tests can now be offered not only for devastating diseases but also for common phenotypic traits. Thus, the consequences of the new technologies and new insights do not have consequences only for human geneticists but also for the entire society.

Possible Function of a Textbook

In his book on *The Structure of Scientific Revolutions*, Kuhn in 1962 [2] described the function of textbooks not very flatteringly: they are “pedagogic vehicles for the perpetuation of normal science” that create the impressions as if science would grow in a simple, cumulative manner. They tend to distort the true history of the field by only mentioning those contributions in the past that can be visualized as direct forerunners of present-day achievements. “They inevitably disguise not only the role but the very existence of . . . revolutions . . .”

Below we shall proceed in the same way: we shall describe present-day problems in human genetics as we see them. The result is a largely affirmative picture of normal science in a phase of rapid growth and success. Anomalies and discrepancies may exist, but we often do not identify them because we share the “blind spots” with most other members of our paradigm group. The “anticipation” phenomenon in diseases

such as myotonic dystrophy is one example (Chap. 5, Sect. 5.1.7). This disease tends to manifest more severely and earlier in life with each generation. Obviously, this observation did not appear to be compatible with simple mendelism. Therefore, it was explained away by sophisticated statistical arguments which we cited in earlier editions of this book. In the meantime, however, anticipation has been shown to be a real phenomenon, caused by a novel molecular mechanism. What we can do is to alert the reader that human genetics, as all other branches of science, is by no way a completed and closed complex of theory and results that only needs to be supplemented in a straightforward way and without major changes in conceptualization. Our field has not developed – and will not develop in the future – as a self-contained system. Rather, human genetics, as all other sciences, is an undertaking of human beings – social groups and single outsiders – who are motivated by a mixture of goals such as search for truth, ambition, desire to be acknowledged by one's peer group, the urge to convince the society at large to allocate resources in their field – but also the wish to help people and to do something useful for human society.

Therefore, we shall emphasize the history and development of problems and approaches. Occasionally, we shall ask the reader to step back, reflecting with us as to why a certain development occurred at the time it did, why another development did not occur earlier, or why a certain branch of human genetics did not take the direction that one would have expected logically. Inevitably, this implies much more criticism than is usually found in textbooks. Such criticism will – at least partially – be subjective, reflecting the personal stance of the authors. Our goal is to convince the reader that a critical attitude improves one's grasp of the problems and their possible solutions – it is not our intention to convince him that we are always right.

We would have liked to give more information on the ways in which sociological conditions within the field and – still more important – the developments in the society at large have influenced the development of human genetics, and the ways in which thinking on these problems has in turn influenced the societies. The eugenics movement in the United States and the *Rassenhygiene* ideology in Germany have had a strong – and sometimes devastating – influence on human beings as well as on the social structure of

society at large. Too little systematic research has been carried out, however, to justify a more extended discussion than that presented in Chap. 1 (Sect. 1.8) [17]. Much more historical research along these lines is all the more urgent, as many of the ethical problems – inherent, for example, in the sterilization laws of many countries during the first decades of the twentieth century – are now recurring with full force in connection with prenatal diagnosis, selective abortion and the possibility of germinal gene therapy (Chaps. 25 and 26). Scientists and physicians working in human genetics were actively involved in and sanctioned ethically abhorrent measures in the past such as killing severely malformed newborns and mentally defectives in Nazi Germany – and how will future generations judge our own activities? These are intriguing questions. They show the Janus face of human genetics: it is a fundamental science – guided by a fertile theory and full of fascinating problems. It is also an applied science, and its applications are bound to have a strong impact on society, leading to novel and difficult philosophical, social, and ethical problems.

References

1. Bunge M (1967) Scientific research, vols 1, 2. Springer, Berlin Heidelberg New York
2. Bernstein F (1924) Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen. *Klin Wochenschr* 3:1495–1497
3. Cairns J, Stent GS, Watson JD (1966) Phage and the origin of molecular biology. Cold Spring Harbor Laboratory, New York
4. Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 52:399–433
5. Garrod AE (1902) The incidence of alcaptonuria: a study in chemical individuality. *Lancet* 2:1616–1620
6. Galton F (1865) Hereditary talent and character. *Macmillans Magazine* 12:157
7. Harwood J (1993) Styles of scientific thought: the German genetics community 1900–1933. University of Chicago Press, Chicago
8. Hirschhorn K (1996) Human genetics: a discipline at risk for fragmentation. *Am J Hum Genet* 58:1–6 William Allan Award Address 1995
9. Kevles DJ (1985) In the name of eugenics. Genetics and the uses of human heredity. Knopf, New York
10. Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago
11. Lakatos I, Musgrave A (1970) Criticism and the growth of knowledge. Cambridge University Press, New York

12. Mendel GJ (1865) Versuche über Pflanzenhybriden. *Verhandlungen des Naturforschenden Vereins*, Brünn
13. Mohr H (1977) The structure and significance of science. Springer, Berlin Heidelberg New York
14. Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia: a molecular disease. *Science* 110:543
15. Popper KR (1963) Conjectures and refutations. Rutledge and Kegan Paul, London
16. Popper KR (1970) Normal science and its dangers. In: Lakatos I, Musgrave A (eds) *Criticism and the growth of knowledge*. Cambridge University Press, New York, pp 51–58
17. Popper KR (1971) *Logik der Forschung*, 4th edn. Mohr, Tübingen Original 1934
18. Watkins JWN (1970) Against “normal science”. In: Lakatos I, Musgrave A (eds) *Criticism and the growth of knowledge*. Cambridge University Press, New York, pp 25–38

History of Human Genetics*

1

Arno G. Motulsky

Revised by A.G. Motulsky, S.E. Antonarakis, and M.R. Speicher

Abstract Theories and studies in human genetics have a long history. Observations on the inheritance of physical traits in humans can even be found in ancient Greek literature. In the eighteenth and nineteenth centuries observations were published on the inheritance of numerous diseases, including empirical rules on modes of inheritance. The history of human genetics as a theory-based science began in 1865, when Mendel published his Experiments on Plant Hybrids and Galton his studies on Hereditary Talent and Character. A very important step in the development of human genetics and its application to medicine came with Garrod's demonstration of a Mendelian mode of inheritance in alkaptonuria and other inborn errors of metabolism (1902). Further milestones were Pauling's elucidation of sickle cell anemia as a "molecular disease" (1949), the discovery of genetic enzyme defects as the causes of metabolic disease (1950s, 1960s), the determination that there are 46 chromosomes in humans (1956), the development of prenatal diagnosis by amniocentesis (1968–1969) for the detection of chromosomal defects such as Down syndrome, and the large-scale introduction of molecular methods during the last 25 years. Concepts appropriated from human genetics have often influenced social attitudes and introduced the eugenics movement. Abuses have occurred, such as legally mandated sterilization, initially in the United States and later more extensively in Nazi Germany, where the killing of mentally impaired patients was followed by the genocide of Jews and Romani (Gypsy) people.

Contents

1.1	The Greeks (see Stubbe [83])	14	1.5	Application to Humans: Garrod's Inborn Errors of Metabolism	18
1.2	Scientists Before Mendel and Galton	15	1.6	Visible Transmitters of Genetic Information: Early Work on Chromosomes.....	20
1.3	Galton's Work	16	1.7	Early Achievements in Human Genetics	20
1.4	Mendel's Work.....	17	1.7.1	AB0 and Rh Blood Groups	20
			1.7.2	Hardy-Weinberg Law	21

A.G. Motulsky (✉)

Departments of Medicine (Division of Medical Genetics)
and Genome Sciences

University of Washington, School of Medicine, Box 355065,
1705 N.E. Pacific St. Seattle, WA 98195-5065, USA
e-mail: agmot@u.washington.edu

*This chapter is a revised and updated version of the third edition's history chapter originally authored by Friedrich Vogel (died summer 2006) and Arno Motulsky.

1	1.7.3	Developments Between 1910 and 1930	21	1.9.2	Biochemical Methods.....	24
	1.8	Human Genetics, the Eugenics Movement, and Politics	21	1.9.3	Genetic and Biochemical Individuality	24
	1.8.1	United Kingdom and United States.....	21	1.9.4	Cytogenetics, Somatic Cell Genetics, Prenatal Diagnosis, Clinical Genetics	24
	1.8.2	Germany	22	1.9.5	DNA Technology in Medical Genetics.....	25
	1.8.3	Soviet Union/Russia (see Harper, Chap. 16 in [38]).....	23	1.9.6	The “Industrialization” of Discoveries and the Team Efforts	26
	1.8.4	Human Behavior Genetics.....	23	1.9.7	Unsolved Problems.....	26
	1.9	Development of Medical Genetics (1950—the Present)	23		References	27
	1.9.1	Genetic Epidemiology.....	23			

The history of human genetics is particularly interesting since, unlike in many other natural sciences, concepts of human genetics have often influenced social and political events. At the same time, the development of human genetics as a science has been influenced by various political forces. Human genetics because of its concern with the causes of human variability has found it difficult to either remain a pure science or one of strictly medical application. Concerns regarding the heritability of IQ and the existence of inherited patterns of behavior again have brought the field into public view. A consideration of the history of human genetics with some attention to the interaction of the field with societal forces is therefore of interest. We will concentrate our attention on historical events of particular interest for human genetics and refer to landmarks in general genetics only insofar as they are essential for the understanding of the evolution of human genetics.

Recently, an excellent history of medical genetics was published by the medical geneticist Peter Harper in 2008 [38]. This highly readable book with many photographs presents critical assessments of various developments in the field since its beginnings. Many tables document major discoveries and a detailed timeline of both human and medical genetics presents important developments ranging from early discoveries to recent findings. This book is currently the only major comprehensive text devoted to the history of human/medical genetics.

A 30-page “History of Medical Genetics” by Victor McKusick was published as Chap. 1 in Emery and Rimoin’s *Principles and Practice of Medical Genetics*, 5th edition, 2007 [59]. This remarkably comprehensive chapter emphasizing clinical aspects starts with a brief description of pre-Mendelian concepts and ends with a broadly conceived assessment of current and future trends of medical genetics.

1.1 The Greeks (see Stubbe [83])

Prescientific knowledge regarding inherited differences between humans has probably existed since ancient times. Early Greek physicians and philosophers not only reported such observations but also developed some theoretical concepts and even proposed “eugenic” measures.

In the texts that are commonly ascribed to Hippocrates, the following sentence can be found:

Of the semen, however, I assert that it is secreted by the whole body – by the solid as well as by the smooth parts, and by the entire humid matters of the body ... The semen is produced by the whole body, healthy by healthy parts, sick by sick parts. Hence, when as a rule, baldheaded beget baldheaded, blue-eyed beget blue-eyed, and squinting, squinting; and when for other maladies, the same law prevails, what should hinder that longheaded are begotten by longheaded?

This remarkable sentence not only contains observations on the inheritance of normal and pathological traits but also a theory that explains inheritance on the assumption that the information carrier, the semen, is produced by all parts of the body, healthy, and sick. This theory became known later as the “pangenesis” theory. Anaxagoras, the Athenian philosopher (500–428 B.C.), had similar views (see Capelle [15]).

A comprehensive theory of inheritance was developed by Aristotle (see [6]). He also believed in a qualitatively different contribution by the male and the female principles to procreation. The male gives the impulse to movement whereas the female contributes the matter, as the carpenter who constructs a bed out of wood. When the male impact is stronger, a son is born who, at the same time, is more like his father, when the

female, a daughter, resembling the mother. This is the reason why sons are usually similar to their fathers and daughters are similar to their mothers.

Barthelmess (our translation) [6] writes: “Reading the texts from this culture, one gets the overall impression that the Greeks in their most mature minds came closer to the theoretical problems than to the phenomena of heredity.” Aristotle’s assertion even provides an early example of how observation can be misled by a preconceived theoretical concept. Sons are not more similar to their fathers, nor daughters to their mothers.

Plato, in the *Statesman (Politikos)* [71], explained in detail the task of carefully selecting spouses to produce children who will develop into bodily and ethically eminent personalities. He wrote:

They do not act on any sound or self-consistent principle. See how they pursue the immediate satisfaction of their desire by hailing with delight those who are like themselves and by disliking those who are different. Thus they assign far too great an importance to their own likes and dislikes.

The moderate natures look for a partner like themselves, and so far as they can, they choose their wives from women of this quiet type. When they have daughters to bestow in marriage, once again they look for this type of character in the prospective husband. The courageous class does just the same thing and looks for others of the same type. All this goes on, though both types should be doing exactly the opposite ...

Because if a courageous character is reproduced for many generations without any admixture of the moderate type, the natural course of development is that at first it becomes superlatively powerful but in the end it breaks out into sheer fury and madness ...

But the character which is too full of modest reticence and untinged by valor and audacity, if reproduced after its kind for many generations, becomes too dull to respond to the challenges of life and in the end becomes quite incapable of acting at all.

In the *Republic* [70], Plato not only requires for the “guards” (one of the highest categories in the social hierarchy of his utopia) that women should be common property; children, should be educated publicly but the “best” of both sexes should beget children who are to be educated with care. The children of the “inferior,” on the other hand, are to be abandoned. Democritus, on the other hand, writes: “More people become able by exercise than by their natural predisposition.” Here (as in other places), the nature–nurture problem appears already.

1.2 Scientists Before Mendel and Galton

The literature of the Middle Ages contains few allusions to heredity. The new attitude of looking at natural phenomena from an empirical point of view created modern science and distinguishes modern humans from those in earlier periods. This approach succeeded first in investigation of the inorganic world and only later in biology. In the work *De Morbis Hereditariis* by the Spanish physician Mercado (1605) [66], the influence of Aristotle is still overwhelming, but there are some hints of a beginning emancipation of reasoning. One example is his contention that both parents, not only the father, contribute a seed to the future child. Malpighi (1628–1694) [83, p 77] proposed the hypothesis of “preformation,” which implies that in the ovum the whole organism is preformed in complete shape, only to grow later. Even after the discovery of sperm (Leeuwenhoek et al. 1677) [3, pp 72–73], the preformation hypothesis was not abandoned altogether, but it was believed by some that the individual is preformed in the sperm, only being nurtured by the mother. The long struggle between the “ovists” and the “spermatists” was brought to an end only when C.F. Wolff [99] attacked both sides and stressed the necessity of further empirical research. Shortly thereafter experimental research on heredity in plants was carried out by Gärtner (1772–1850) [33] and Kölreuter (1733–1806) [48]. Their work prepared the ground for Mendel’s experiments [60].

The medical literature of the eighteenth and early nineteenth centuries contains reports showing that those capable of clear observation were able to recognize correctly some phenomena relating to the inheritance of diseases. Maupertuis [57], for example, published in 1753 an account of a family with polydactyly in four generations and demonstrated that the trait could be equally transmitted by father or by mother. He further showed, by probability calculation, that chance alone could not account for the familial concentration of the trait. Probably the most remarkable example, however, was Joseph Adams (1756–1818) (see [1,23,62,64]), a British apothecary who, in 1814, published a book with the title *A Treatise on the Supposed Hereditary Properties of Diseases* [1]. The following findings are remarkable:

- (a) Adams differentiated clearly between “familial” (i.e., recessive) and “hereditary” (i.e., dominant) conditions.
- (b) He knew that in familial diseases the parents are frequently near relatives.

- (c) Hereditary diseases need not be present at birth; they may manifest themselves at various ages.
- (d) Some disease predispositions lead to a manifest disease only under the additional influence of environmental factors. The progeny, however, is endangered even when the predisposed do not become ill themselves.
- (e) Intrafamilial correlations as to age of onset of a disease can be used in genetic counseling.
- (f) Clinically identical diseases may have different genetic bases.
- (g) A higher frequency of familial diseases in isolated populations may be caused by inbreeding.
- (h) Reproduction among persons with hereditary diseases is reduced. Hence, these diseases would disappear in the course of time, if they did not appear from time to time among children of healthy parents (i.e., new mutations!).

Adams' attitude toward "negative" eugenic measures was critical. He proposed the establishment of registries for families with inherited diseases. Weiss [96] recently pointed out that Adams in the same book also hinted at the existence of evolution stressing the concept of adaptive selection saying that environments such as climate put constraints on people: "By these means a race is gradually reared with constitutions best calculated for the climate" [1].

C.F. Nasse, a German professor of medicine, correctly recognized in 1820 one of the most important formal characteristics of the X-linked recessive mode of inheritance in hemophilia and presented a typical comprehensive pedigree [83, p 180]. He wrote (our translation):

All reports on families, in which a hereditary tendency towards bleeding was found, are in agreement that the bleeders are persons of male sex only in every case. All are explicit on this point. The women from those families transmit this tendency from their fathers to their children, even when they are married to husbands from other families who are not afflicted with this tendency. This tendency never manifests itself in these women. . . .

Nasse also observed that some of the sons of these women remain completely free of the bleeding tendency.

The medical literature of the nineteenth century shows many more examples of observations, and attempts to generalize and to find rules for the influence of heredity on disease can be found. The once very influential concept of "degeneration" should be mentioned. Some features that older authors described

as "signs of degeneration" in the external appearance of mentally deficient patients are now known to be characteristic of autosomal chromosomal aberrations or various types of mental retardation.

In the work of most of the nineteenth century authors, true facts and wrong concepts were inextricably mixed, and there were few if any criteria for getting at the truth. This state of affairs was typical for the plight of a science in its prescientific state. Human genetics had no dominant paradigm. The field as a science was to start with two paradigms in 1865: biometry, which was introduced by Galton, and Mendelism, introduced by Mendel with his pea experiments. The biometric paradigm was influential in the early decades of the twentieth century, and some examples and explanations in this book utilize its framework. With the advent of molecular biology and insight into gene action, the pure biometric approach in genetics is on the decline. Nevertheless, many new applications in behavioral or social genetics, where gene action cannot yet be studied, rely on this paradigm and its modern elaborations. The laws that Mendel derived from his experiments, on the other hand, have been of almost unlimited fruitfulness and analytic power. The gene concept emerging from these experiments has become the central concept of all of genetics, including human genetics. Its possibilities have not been exhausted.

1.3 Galton's Work

In 1865, F. Galton published two short papers with the title "Hereditary Talent and Character." He wrote [29]:

The power of man over animal life, in producing whatever varieties of form he pleases, is enormously great. It would seem as though the physical structure of future generations was almost as plastic as clay, under the control of the breeder's will. It is my desire to show, more pointedly than – so far as I am aware – has been attempted before, that mental qualities are equally under control.

A remarkable misapprehension appears to be current as to the fact of the transmission of talent by inheritance. It is commonly asserted that the children of eminent men are stupid; that, where great power of intellect seems to have been inherited, it has descended through the mother's side; and that one son commonly runs away with the talent of the whole family.

He then stresses how little we know about the laws of heredity in man and mentions some reasons, such as

long generation time, that make this study very difficult. However, he considers the conclusion to be justified that physical features of humans are transmissible because resemblances between parents and offspring are obvious. Breeding experiments with animals, however, had not been carried out at that time, and direct proof of hereditary transmission was therefore lacking even in animals. In humans, “we have … good reason to believe that every special talent or character depends on a variety of obscure conditions, the analysis of which has never yet been seriously attempted.” For these reasons, he concluded that single observations must be misleading, and only a statistical approach can be adequate.

Galton evaluated collections of biographies of outstanding men as to how frequently persons included in these works were related to each other. The figures were much higher than would be expected on the basis of random distribution.

Galton himself was fully aware of the obvious sources of error of such biological conclusions. He stressed that “when a parent has achieved great eminence, his son will be placed in a more favorable position for advancement, than if he had been the son of an ordinary person. Social position is an especially important aid to success in statesmanship and generalship . . .”

“In order to test the value of hereditary influence with greater precision, we should therefore extract from our biographical list the names of those that have achieved distinction in the more open fields of science and literature.” Here and in the law, which in his opinion was “the most open to fair competition,” he found an equally high percentage of close relatives reaching eminence. This was especially obvious with Lord Chancellors, the most distinguished lawyers of Great Britain.

Galton concluded that high talent and eminent achievement are strongly influenced by heredity. Having stressed the social obstacles that inhibit marriage and reproduction of the talented and successful, he proceeded to describe a utopic society,

In which a system of competitive examination for girls, as well as for youths, had been so developed as to embrace every important quality of mind and body, and where a considerable sum was yearly allotted. … to the endowment of such marriages as promised to yield children who would grow into eminent servants of the State. We may picture to ourselves an annual ceremony in that Utopia or Laputa, in which the Senior Trustee of the Endowment Fund would address ten deeply-blushing young men, all of twenty-five years old, in the following terms. . . .

In short, they were informed that the commission of the endowment fund had found them to be the best, had selected for each of them a suitable mate, would give them a substantial dowry, and promised to pay for the education of their children.

This short communication already shows human genetics as both a pure and an applied science: on the one hand, the introduction of statistical methods subjects general impressions to scientific scrutiny, thereby creating a new paradigm and turning prescience into science. Later, Galton and his student K. Pearson proceeded along these lines and founded biometric genetics. On the other hand, however, the philosophical motive of scientific work in this field is clearly shown: the object of research is an important aspect of human behavior. The prime motive is the age-old inscription on the Apollo temple at Delphi (“know yourself”).

Hence, with Galton, research in human genetics began with strong eugenic intentions. Later, with increasing methodological precision and increasing analytic success, such investigations were removed from this prime philosophical motive. This motive helps to understand the second aspect of Galton’s work: the utopian idea to improve the quality of the human species by conscious breeding. During the Nazi era in Germany (1933–1945) we saw how cruel the perverted consequences of such an idea may become (Sect. 1.8.2). The question first posed by Galton remains, even more than ever, of pressing importance: What will be the biological future of mankind?

1.4 Mendel’s Work

The other leading paradigm was provided by Mendel in his work *Experiments in Plant Hybridization*, which was presented on 8 February and 8 March 1865 before the *Naturforschender Verein* (Natural Science Association) in Brünn (now Brno, Czech Republic) and subsequently published in its proceedings [60]. It has frequently been told how this work went largely unnoticed for 35 years and was rediscovered independently by Correns, Tschermak, and de Vries in 1900 (see [16, 84, 20]). From then on, Mendel’s insights triggered the development of modern genetics, including human genetics. A book by Stern and Sherwood [82], which reprints these and a variety of other articles regarding Mendel’s paper, is most helpful to assess the impact of this classic work.

Mendel was stimulated to carry out his experiments by observations on ornamental plants, in which he had tried to breed new color variants by artificial insemination. Here he had been struck by certain regularities. He selected the pea for further experimentation. He crossed varieties with differences in single characters such as color (yellow or green) or form of seed (round or angular wrinkled) and counted all alternate types in the offspring of the first generation crosses and of crosses in later generations. Based on combinatorial reasoning, he gave a theoretical interpretation: the results pointed to free combination of specific sorts of egg and pollen cells. In fact, this concept may have occurred to Mendel before he carried out his studies. He may have verified and illustrated his findings by his “best” results, since agreement between the published figures and their expectation from the theoretical segregation ratios is too perfect from a statistical point of view (Fisher [27]). The interpretation of this discrepancy remains controversial [82, 90]. In any case, there is no question that Mendel’s findings were correct.

Mendel discovered three laws: the law of uniformity, which states that after crossing of two homozygotes of different alleles the progeny of the first filial generation (F_1) are all identical and heterozygous; the law of segregation, which postulated 1:2:1 segregation in intercrosses of heterozygotes and 1:1 segregation in backcrosses of heterozygotes with homozygotes; and the law of independence, which states that different segregating traits are transmitted independently.

What is so extraordinary in Mendel’s contribution that sets it apart from numerous other attempts in the nineteenth century to solve the problem of heredity? Three points are most important:

1. He simplified the experimental approach by selecting characters with clear alternative distributions, examining them one by one, and proceeding only then to more complicated combinations.
2. Evaluating his results, he did not content himself with qualitative statements but counted the different types. This led him to the statistical law governing these phenomena.
3. He suggested the correct biological interpretation for this statistical law: The germ cells represent the constant forms that can be deduced from these experiments.

With this conclusion Mendel founded the concept of the gene, which has proved so fertile ever since. The history

of genetics since 1900 is dominated by analysis of the gene. What had first been a formal concept derived from statistical evidence has emerged as the base pair sequence of DNA, which contains the information for protein synthesis and for life in all its forms.

1.5 Application to Humans: Garrod’s Inborn Errors of Metabolism

The first step of this development is described in this historical introduction: A. Garrod’s [30] paper on “The Incidence of Alkaptonuria: A Study in Chemical Individuality.” There are two reasons for giving special attention to this paper. For the first time, Mendel’s gene concept was applied to a human character, and Mendel’s paradigm was introduced into research on humans. Additionally, this work contains many new ideas set out in a most lucid way. Garrod was a physician and in later life became the successor of Osler in the most prestigious chair of medicine at Oxford [8]. His seminal contribution to human genetics remained unappreciated during his lifetime. Biologists paid little attention to the work of a physician. Their interest was concentrated more on the formal aspects of genetics rather than on gene action. The medical world did not understand the importance of his observations for medicine. Garrod first mentioned the isolation of homogentisic acid from the urine of patients with alkaptonuria and stated the most important result of the investigations carried out so far:

As far as our knowledge goes, an individual is either frankly alkapturic or conforms to the normal type, that is to say, excretes several grammes of homogentisic acid per diem or none at all. Its appearance in traces, or in gradually increasing or diminishing quantities, has never yet been observed....

As a second important feature “the peculiarity is in the great majority of instances congenital....” Thirdly: “The abnormality is apt to make its appearance in two or more brothers and sisters whose parents are normal and among whose forefathers there is no record of its having occurred.” Fourthly, in six of ten reported families the parents were first cousins, whereas the incidence of first-cousin marriages in contemporary England was estimated to be not higher than 3%. On the other hand, however, children with alkaptonuria are observed in a very small fraction only of all first-cousin marriages.

There is no reason to suppose that mere consanguinity of parents can originate such a condition as alkaptonuria in their offspring, and we must rather seek an explanation in some peculiarity of the parents, which may remain latent for generations, but which has the best chance of asserting itself in the offspring of the union of two members of a family in which it is transmitted.

Then, Garrod mentioned the law of heredity discovered by Mendel, which “offers a reasonable account of such phenomena” that are compatible with a recessive mode of inheritance as pointed out by Bateson [37]. He cited another remark of Bateson and Saunders (Report to the Evolution Committee of the Royal Society) [7] with whom he had discussed his data:

We note that the mating of first cousins gives exactly the conditions most likely to enable a rare, and usually recessive, character to show itself. If the bearer of such a gamete mates with individuals not bearing it the character will hardly ever be seen; but first cousins will frequently be the bearers of similar gametes, which may in such unions meet each other and thus lead to the manifestation of the peculiar recessive characters in the zygote.

After having cited critically some opinions on the possible causes of alkaptonuria, Garrod proceeded:

The view that alkaptonuria is a “sport” or an alternative mode of metabolism will obviously gain considerably in weight if it can be shown that it is not an isolated example of such a chemical abnormality, but that there are other conditions which may reasonably be placed in the same category.

Having mentioned albinism and cystinuria as possible examples, he went on: “May it not well be that there are other such chemical abnormalities which are attended by no obvious peculiarities [as the three mentioned above] and which could only be revealed by chemical analysis?” And further:

If it be, indeed, the case that in alkaptonuria and the other conditions mentioned we are dealing with individualities of metabolism and not with the results of morbid processes the thought naturally presents itself that these are merely extreme examples of variations of chemical behavior which are probably everywhere present in minor degrees and that just as no two individuals of a species are absolutely identical in bodily structure neither are their chemical processes carried out on exactly the same lines.

He suggested that differential responses toward drugs and infective agents could be the result of such chemical individualities. The paper presents the following new insights:

- (a) Whether a person has alkaptonuria or not is a matter of a clear alternative – there are no transitory

forms. This is indeed a condition for straightforward recognition of simple modes of inheritance.

The condition is observed in some sibs and not in parents.

The unaffected parents are frequently first cousins.

This is explained by the hypothesis of a recessive mode of inheritance according to Mendel. The significance of first-cousin marriages is stressed especially for rare conditions; this may be a precursor to population genetics.

- (b) Apart from alkaptonuria several other similar “sports” such as albinism and cystinuria may exist. This makes alkaptonuria the paradigm for the “inborn errors of metabolism.” In 1909 Garrod published his classic monograph on this topic [31].
- (c) These sports may be extreme and therefore conspicuous examples of a principle with *much more widespread applicability*. Lesser chemical differences between human beings are so frequent that no human being is identical chemically to anyone else.

From these concepts Garrod drew more far-reaching conclusions, which are often overlooked. In a book published in 1931 [32] and reprinted with a lengthy introduction by Scriver and Childs [80], Garrod suggested that hereditary susceptibilities or diatheses are a predisposing factor for most common diseases and not merely for the rare inborn errors of metabolism. These concepts were precursors of current work to delineate the specific genes involved in the etiology of common disease. A valuable biography of Garrod was published by A. Bearn [8], who was a pioneer of human biochemical genetics in the 1950s and later.

Throughout this book the principle of a genetically determined individuality will govern our discussions. Garrod’s contribution may be contrasted with that of Adams [23, 62, 64]. Apart from the “familial” occurrence of some hereditary diseases, Adams observed a number of phenomena that were not noted by Garrod, such as the late onset of some diseases, the intrafamilial correlation of age of onset, and the genetic predisposition leading to manifest illness only under certain environmental conditions. However, Adams did not have Mendel’s paradigm. Therefore, his efforts could not lead to the development of an explanatory theory and coherent field of science. Garrod did have this paradigm and used it, creating a new area of research: human biochemical genetics.

1.6 Visible Transmitters of Genetic Information: Early Work on Chromosomes

Galton's biometric analysis and Mendel's hybridization experiments both started with visible phenotypic differences between individuals. The gene concept was derived from the phenotypic outcome of certain crossings. At the time when Mendel carried out his experiments nothing was known about a possible substantial bearing of genetic information in the germ cells. During the decades to follow, however, up to the end of the nineteenth century, chromosomes were identified, and mitosis and meiosis were analyzed. These processes were found to be highly regular and so obviously suited for orderly distribution of genetic information that in 1900 the parallelism of Mendelian segregation and chromosomal distribution during meiosis was realized, and chromosomes were identified as bearers of the genetic information [18].

Many research workers contributed to the development of cytogenetics [5,6]. O. Hertwig [41] first observed animal fertilization and established the continuity of cell nuclei: *omnis nucleus e nucleo*. Flemming (1880–1882) discovered the separation of sister chromatids in mitosis [83, p 247]; van Beneden (1883) [85] established the equal and regular distribution of chromosomes to the daughter nuclei. Boveri (1888) [5] found evidence for the individuality of each pair of chromosomes. Waldeyer (1888) (see [18]) coined the term "chromosome."

Meanwhile, Naegeli (1885) [77] had developed the concept of "idioplasma," which contains – to use a modern term – the "information" for the development of the next generation [67]. W. Roux [77] seems to have been the first to set out by logical deduction which properties a carrier of genetic information was expected to have. He also concluded that the behavior of cell nuclei during division would perfectly fulfill these requirements. The most important specific property of meiotic divisions, the ordered reduction of genetic material, was first recognized by Weismann.

These results and speculations set the stage for the identification of chromosomes as carriers of the genetic information, which followed shortly after the rediscovery of Mendel's laws and apparently independently by different authors [16, 20, 84].

Chromosome studies and genetic analysis have remained intimately connected in cytogenetics ever

since. Most basic facts were discovered and concepts developed using plants and insects as the principal experimental tools. The fruit fly *Drosophila* played a particularly important role.

The development of human cytogenetics was delayed until 1956 when the correct number of human chromosomes was established as 46 by use of rather simple methods. It should be stressed that this delay could not be explained by the introduction of new cytological methods at that time. In fact, this discovery could have been made many years earlier. The delay was probably related to the lack of interest in human genetics by most laboratory-oriented medical scientists. Human genetics did not exist as a scientific discipline in medical schools since the field was not felt to be a basic science fundamental to medicine. Hereditary diseases were considered as oddities that could not be studied by the methodology of medical science as exemplified by the techniques of anatomy, biochemistry, physiology, microbiology, pathology, and pharmacology. Thus, most geneticists worked in biology departments of universities, colleges, or in agricultural stations. They were usually not attuned to problems of human biology and pathology, and there was little interest to study the human chromosomes. The discovery of trisomy 21 as the cause of Down syndrome and the realization that many problems of sex differentiation owe their origin to sex chromosomal abnormalities established the central role of cytogenetics in medicine. Further details in the development of cytogenetics are described in Chap. 3.

1.7 Early Achievements in Human Genetics

1.7.1 AB0 and Rh Blood Groups

The discovery of the AB0 blood group system by Landsteiner in 1900 [50] and the proof that these blood types are inherited (von Dungern and Hirschfeld [87]) was an outstanding example of Mendelian inheritance applied to a human character. Bernstein in 1924 [11] demonstrated that A, B, and O blood group characters are due to multiple alleles at one locus. The combined efforts of Wiener, Levine, and Landsteiner 25–30 years later led to discovery of the Rh factor and established that hemolytic disease of the newborn owes its origin

to immunological maternal–fetal incompatibility. The stage was set for the demonstration in the 1960s that Rh hemolytic disease of the newborn can be prevented by administration of anti-Rh antibodies to mothers at risk [73,100].

1.7.2 Hardy-Weinberg Law

Hardy [36], a British mathematician, and Weinberg [92], a German physician, at about the same time (1908) set out the fundamental theorem of population genetics, which explains why a dominant gene does not increase in frequency from generation to generation. Hardy published his contribution in the United States in *Science*. He felt that this work would be considered as too trivial by his mathematics colleagues to be published in the United Kingdom. Weinberg was a practicing physician who made many contributions to formal genetics. He developed a variety of methods in twin research [91] and first elaborated methods to correct for biased ascertainment in recessive inheritance [93].

1.7.3 Developments Between 1910 and 1930

The years between 1910 and 1930 saw no major new paradigmatic discoveries in human genetics. Most of the data in formal genetics (such as linkage, nondisjunction, mutation rate) as well as the mapping of chromosomes were achieved by study of the fruit fly, largely in the United States. Many scientists tried to apply the burgeoning insights of genetics to humans. British scientists exemplified by Haldane excelled in the elaboration of a variety of statistical techniques required to deal with biased human data. The same period saw the development of the basic principles of population genetics by Haldane, Fisher, and Penrose [69] in England and by Wright in the United States. This body of knowledge became the foundation of population genetics and is still used by workers in that field. In 1918, Fisher was able to resolve the bitter controversies in England between the Mendelians, on the one hand, and followers of Galton (such as Pearson) on the other, by pointing out that correlations between relatives in metric traits can be explained by the combined action of many individual genes [26]. Novel

steps in the development of medical genetics during this period were the establishment of empirical risk figures for schizophrenia and affective disorders by the Munich school of psychiatric genetics.

1.8 Human Genetics, the Eugenics Movement, and Politics

1.8.1 United Kingdom and United States

The first decade of the century saw the development of eugenics in Europe and in the United States [2,19,21,45, 55,76]. Many biological scientists were impressed by their interpretation of an apparently all-pervasive influence of genetic factors on most normal physical and mental traits as well as on mental retardation, mental disease, alcoholism, criminality, and various other sociopathies. They became convinced that the human species should be concerned with encouragement of breeding between persons with desirable traits (positive eugenics) and discourage the sick, mentally retarded, and disabled from procreation (negative eugenics).

A recent reprint of Davenport's 1911 book, *Heredity in Relation to Eugenics*, is accompanied by thoughtful reflections from contemporary geneticists on Davenport's eugenic concepts and recommendations almost one hundred years later [98]. Various eugenic study units were established in the United States (Eugenics Record Office at Cold Spring Harbor) and the United Kingdom. Much of the scientific work published by these institutions was of poor quality. Particularly, many different kinds of human traits such as "violent temper" and "wandering trait" were forced into Mendelian straightjackets. Most serious geneticists became disenchanted and privately disassociated themselves from this work. For various reasons, including those of friendship and collegiality with the eugenicists, the scientific geneticists did not register their disagreement in public. Thus, the propagandists of eugenics continued their work with enthusiasm, and the field acquired a much better reputation among some of the public than it deserved. Thus, many college courses on eugenics were introduced in the United States.

These trends had several important political consequences. Eugenics sterilization laws were passed in many states in the United States, which made it possible to sterilize a variety of persons for traits such as

criminality for which no good scientific basis of inheritance existed. The attitude that led to the introduction of these laws is epitomized by United States Supreme Court Justice Holmes' statement that "three generations of imbeciles are enough."

Eugenic influences also played an important role in the passing of restrictive immigration laws in the United States. Using a variety of arguments the proponents of eugenics claimed to show that Americans of northwestern European origin were more useful citizens than those of southern European origin or those from Asia. Since such differences were claimed to be genetic in origin, immigration from southern and eastern European countries and from Asia was sharply curtailed. Similar trends were also operative in the United Kingdom. While solid work in human genetics was carried out by a few statistical geneticists, there was also much eugenic propaganda, including that by the distinguished statistician Pearson, the successor to Galton's academic chair in London.

Kevles [46] has published a wide-ranging and insightful history of eugenics and human genetics in the Anglo-Saxon countries. His book is a most carefully researched and exhaustive study of the uses and abuses of eugenic concepts.

who were concerned about the biological future of mankind. Thus, socialists publicized such views in Germany [34]. In 1931, two years before Hitler's coming into power, the German Society of Racial Hygiene added eugenics to its name. However, all efforts in this area soon became identified with the Nazi ideology.

Prominent German human geneticists identified themselves with the use of human genetics in the service of the Nazi state. Recognized scientists, such as Fischer, F. Lenz, Rüdin, and von Verschuer, accepted Nazi leadership and Nazi philosophy. While most of the propaganda for the new racial hygiene was not formulated by scientists but by representatives of the Nazi party, men such as Fischer and von Verschuer [95] participated in spreading Nazi race ideology. Jews were declared foreign genetic material to be removed from the German *Volk*. A eugenic sterilization law was already passed in 1933 that made forced sterilization obligatory for a variety of illnesses thought to be genetic in origin [74]. Heredity courts were established to deal with interpretation of the sterilization law. This law was hailed by some eugenicists in the United States even at the end of the 1930s [47]. Sterilization laws for eugenic indications were also passed in some Scandinavian countries around the same time but allowed voluntary (in contrast to forced) sterilization [74].

The exact role of the German human geneticists in the increasing radicalization and excesses of the application of Nazi philosophy has been assessed [65, 74, 95]; von Verschuer's role in sponsoring twin and other genetic research by his former assistant Mengele in the Auschwitz concentration and extermination camp is clear. We have no record that any voices were raised by these men in protest against "mercy killings" of the mentally retarded and newborn children with severe congenital defects nor against the mass killings of Jews. Evidence suggests that von Verschuer must have had some idea of such events, since he had continued contact with Mengele when the mass killings at Auschwitz were at their height. The "final solution" to the "Jewish problem" resulted in the murder of about 6 million Jews in the early 1940s [75]. While there is no record that human geneticists favored this type of "solution," their provision of so-called "scientific" evidence for a justification of Nazi antisemitism helped to create a climate in which these mass murders became possible [88]. This episode is one of the most macabre and tragic chapters in the history of man's inhumanity to man in the name of pseudoscientific nationalism. Yet, despite

1.8.2 Germany

In Germany [9, 10, 34, 94, 95] eugenics took the name of *Rassenhygiene* from a book of that title published in 1895 by Ploetz [72]. The *Rassenhygiene* movement became associated with mystical concepts of race, Nordic superiority, and the fear of degeneration of the human race in general and that of the German *Volk* in particular by alcoholism, syphilis, and increased reproduction of the feeble-minded or persons from the lower social strata. Often representatives of this movement became associated with a dangerous type of sociopolitical prejudice: antisemitism. They warned the public against contamination of German "blood" by Jewish influences. Most followers of the racial hygiene concept were nationalistic and opposed the development of an open society that allows individual freedom and democratic participation. They shared this attitude with a significant segment of the educated classes in Germany. General eugenic ideas divorced from racism and other nationalist notions were often espoused by intellectuals

their racist publications, several such “scientists” (including von Verschuer) were given academic positions in post-World War II West Germany.

1.8.3 Soviet Union/Russia (see Harper, Chap. 16 in [38])

Eugenics was initiated in the Soviet Union [21,34] in the 1920s by the establishment of eugenics departments, a eugenic society, and a eugenics journal. Eugenic ideals soon clashed with the official doctrine of Marxism-Leninism, however, and these efforts were abandoned by the late 1920s. Scientists who had become identified with eugenics left the field to work with plants and animals.

Remarkable work in early human cytogenetics was carried out between 1931 and 1936, such as using hypotonic solutions for spreading of chromosomes, analysis of cultured embryonic cells, chromosome analysis of human oocytes, and cytogenetic studies of leukemia and other cancers [3,4]. These studies were published in international journals and later taken up by American and European scientists some 20 years later. Would the critical chromosome-related discoveries of the 1950s have been made by Russian scientists if such work on human genetics had not been terminated by Soviet antigenetic policies? [38]

Interest in the medical application of human genetics nevertheless persisted. A large institute of medical genetics, with 200 physicians, was established in Moscow during the 1920s. Its director, the physician S.G. Levit, made notable contributions [54], but was executed in 1938 (Chap. 16 in [38]), and human genetics was officially declared a Nazi science. The later ascendancy of Lysenko [45] stifled all work in genetics, including that of human genetics, and no work whatever was carried out in this field until the early 1960s, after Lysenko’s domination ceased (pp. 435–450 in [38]). The reintroduction of human genetics into the Soviet Union occurred by way of medical genetics. A textbook of medical genetics was published by Efroimson in 1964 [22]. A new institute of medical genetics was established in 1969 under the directorship of the cytogeneticist Bochkov, who had been trained by the well-known *Drosophila* geneticist, Timofeeff-Ressovsky [38]. Work in many areas of medical genetics, similar to that carried out elsewhere, is now done in Russia.

1.8.4 Human Behavior Genetics

Vigorous discussion continues regarding the role of genetic determinants in behavior, IQ, and personality. Some observers entirely deny genetic influences on normal behavior or social characteristics such as personality and intellect. This attitude toward genetics is shared by some psychologists and social scientists and even a few geneticists who are concerned about the possible future political and social misuse of studies in human behavioral genetics that claim to show genetic determinants of intelligence and social behavior.

We do not agree with those who deny any genetic influence on behavior or social traits in humans. However, we also caution against a too ready acceptance of results from comparison of twins and other relatives, which claim high heritabilities for many of these traits. Genetic data and pseudodata may be seriously misused by political bodies. However, as biologists and physicians impressed by biological variation under genetic control, we would be surprised if the brain did not also show significant variation in structure and function. Such variation is expected to affect intellect, personality, and behavior, and usually will interact with environmental factors. The extent to which genetic variation contributes to such traits, and especially the biological nature of such variation, will have to await further studies.

1.9 Development of Medical Genetics (1950–the Present)

1.9.1 Genetic Epidemiology

In the 1940s and 1950s a number of institutions pioneered in research on epidemiology of genetic diseases. T. Kemp’s institute in Copenhagen, J.V. Neel’s department in Ann Arbor, Michigan, and A.C. Stevenson’s in Northern Ireland and later in Oxford contributed much to our knowledge on prevalence, modes of inheritance, heterogeneity, and mutation rates of various hereditary diseases. Recent years have seen a renaissance in this area, with special attention to analysis of common complex diseases (see Chap. 8.1). Utilization of new laboratory methods, including DNA techniques, together with more powerful methods of association studies, and the

search for rare mutations and structural chromosome changes, provide powerful new approaches in this area.

1.9.2 Biochemical Methods

The years after World War II brought a rapid expansion in the field of human genetics by the development of biochemical, molecular, and cytological methods. Human genetics, which had been the concern largely of statistically oriented scientists, now entered the mainstream of medical research. The demonstration by Pauling et al. [68] that sickle cell anemia is a molecular disease was a key event in this area. The hemoglobins allowed detailed study of the consequences of mutation. The genetic code was found to be valid for organisms as far apart as viruses and humans. Many detectable mutations were found to be single amino acid substitutions, but deletions of various sorts and frameshift mutations similar to those discovered in micro-organisms were discovered. The nucleotide sequences of the hemoglobin genes were worked out using techniques developed in biochemistry and molecular genetics. Many inborn errors of metabolism were shown to originate in various enzyme deficiencies, often caused by a genetic mutation that changes enzyme structure. Methemoglobinemia due to diaphorase deficiency and glycogen storage disease were the first enzyme defects to be demonstrated.

1.9.3 Genetic and Biochemical Individuality

Work on hemoglobin and variants of the enzyme glucose-6-phosphate-dehydrogenase and other enzymes helped to establish the concept of extensive mutational variation. Biochemical individuality explained some drug reactions and led to the development of the field of pharmacogenetics [61, 86, 63, 35]. Marked biochemical heterogeneity of human enzymes and proteins was shown [39]. The uniqueness of humans, which is apparent by the physiognomic singularity of each human being, was shown to apply at the biochemical and immunological level as well. Here, as in several other fields (such as the hemoglobin variants and the mechanism of sex determination), studies in

humans led the way to generally valid biological rules. The significance of polymorphism for the population structure (including that of humans) is being widely studied by population geneticists. The hypothesis that some expressed polymorphisms are the genetic substrate against which the environment acts to determine susceptibility and resistance to common disease led to the development of the field of ecogenetics [13,17]. The histocompatibility gene complex has become an important paradigm for the understanding of why several genes with related function occur in closely linked clusters. This locus appears to be of great importance to understand susceptibility to autoimmune diseases. An enormous amount of apparently unexpressed genetic variation has been demonstrated at the DNA and chromosomal level.

1.9.4 Cytogenetics, Somatic Cell Genetics, Prenatal Diagnosis, Clinical Genetics

After cytogenetic techniques became available, they were applied to detect many types of birth defects and intersex states. A specific type of malignancy, chronic myelogenous leukemia, was shown to be caused by a unique chromosomal translocation [78]. Banding techniques developed by Caspersson in 1969 made it possible to visualize each human chromosome and gave cytogenetic methods added powers of resolution.

Soon, biochemical and cytogenetic techniques were combined in somatic cell genetics. Specific enzyme defects were identified in single cells grown in tissue cultures. The development of methods to hybridize human with mouse cells by Henry Harris and Watkins [40] and Ephrussi and Weiss [25] soon allowed the assignment of many genes to specific chromosomes and the construction of a human linkage map.

The developments in somatic cell genetics led to the introduction of prenatal diagnosis in the late 1960s, when amniocentesis at the beginning of the second trimester of pregnancy was developed. This allowed tissue cultures of amniotic cells of fetal origin, permitting both cytogenetic and biochemical characterization of fetal genotypes, assignment of sex, and the diagnosis of a variety of disorders in utero. In the early 1980s chorion villus biopsy – a procedure done during the first trimester of pregnancy – was introduced, and is

being widely used. The discovery that neural tube defects are associated with increases in α -fetoprotein of the amniotic fluid permits intrauterine diagnosis of an important group of birth defects [14]. Ultrasound methods to visualize the placenta and to diagnose fetal abnormalities added to the diagnostic armamentarium. This noninvasive method allows phenotypic diagnosis of a variety of fetal defects more frequently.

Clinical Genetics. The field of clinical genetics was initiated in the 1970s [58] and has been growing rapidly. Many medical schools and hospitals are establishing special clinics in which genetic diseases can be diagnosed and genetic counseling provided. The heterogeneity of genetic disease has been increasingly recognized. Genetic counseling – often by specially trained genetic counselors – is now intensified to provide patients and their families with information on the natural history of the disease, recurrence risks, and reproductive options. Screening programs of the entire newborn population for diseases such as phenylketonuria are being introduced in many countries, and other screening programs such as those to detect carriers of Tay-Sachs disease and other conditions more common among Ashkenazi Jews have undergone extensive trials [81].

With the advent of novel biochemical and DNA techniques (Chap. 4), basic work in human genetics is now performed increasingly by biochemists, cell biologists, molecular biologists, and others, who do not necessarily have training in human genetics. However, human genetics is identified with medical genetics in many of its activities. The scientific developments of the past decades are thus being widely applied in practical medicine.

1.9.5 DNA Technology in Medical Genetics

Advances in molecular genetics and DNA technology are being applied rapidly to practical problems of medical genetics. Since understanding of the hemoglobin genes was more advanced than that of other genetic systems, the initial applications related to the diagnosis of hemoglobinopathies (Chap. 11). Several methods are now being utilized. Inherited variation in DNA sequence that is phenotypically silent was found to be common, supplying a vast number of DNA polymorphisms for study. Just as everyone's physiognomy is unique, each person (except for identical twins) has a unique DNA pattern. DNA

variants are being used in family or association studies as genetic markers to detect the presence of closely linked genes causing diseases. Direct detection of genetic disease has been achieved by utilizing nucleotide probes that are homologous to the mutations that are searched for. The polymerase chain reaction, together with rapidly increasing knowledge on human DNA sequences, has opened up new opportunities for direct diagnosis at the DNA level. Occasionally, a specific restriction enzyme may detect the mutational lesion. Different DNA mutations at the same locus frequently cause an identical phenotypic disease. This finding makes direct DNA diagnosis without family study difficult unless the specific mutation that causes the disease is known.

Completion of the human gene map and human gene sequence was achieved at the beginning of this century. Several hundred DNA markers and SNPs that are spaced over all chromosomes provide the necessary landmarks for detection of the genes for monogenic diseases and are beginning to hint at the contribution of specific genes to common diseases.

Using normal DNA carried by innocuous viruses to treat patients with genetic diseases carried by defective DNA has been under study for the last 15 years (Chap. 26). Such gene transfer aims to repair affected somatic cells (somatic gene therapy). Human studies have been done but no definitive cures have been reported. However, acute leukemia developed in several children treated for hereditary antibody syndrome presumably due to activation of oncogenes. Germinal gene therapy, i.e., insertion of normal genes into defective germ cells (or fertilized eggs) for treatment of human genetic disease, has never been carried out and is not considered ready for safe study. Such an approach is highly controversial, and is even prohibited by law in some countries.

McKusick [59] described a variety of paradigm shifts in the study of human and medical genetics in recent years. These included an emphasis from structural to functional genomics, from map-based to sequence-based gene discovery, from monogenic disease diagnosis to detection of common disorder susceptibility, from the search for etiology to exploration of mechanisms, from an emphasis on single genes to approaches on systems pathways and gene families, from genomics to proteomics and from "old-fashioned" medical genetics to "genetic medicine," implying that genes may be involved in all diseases. McKusick (p. 28 in [59]) further pointed out that human genetics

in recent years has been “medicalized,” “subspecialized,” “professionalized,” “molecularized,” “commercialized,” and even “consumerized.”

1.9.6 The “Industrialization” of Discoveries and Team Efforts

The technological advances, the enormous amount of data generated, the size of the genomes, the impressive variability of individual genomes, the necessary specialized expertise in several disciplines, and the revolution in communication technologies all resulted in the organization and execution of mega-projects related to human genetics in the last 15 years in order to achieve results freely available to the community that provide genome-wide answers to the objectives. These projects, mostly international and funded by different funding agents, often included more than 50 different laboratories and 200 scientists. This paradigm shift is similar to the evolution of experimentation in physics, and underscores the importance of international cooperation in genomic discoveries. In addition, it is remarkable that most of the funding was provided by public sources. The completion of the human genome sequence was the first example of such international projects [44,49]. Other examples include the sequence of the genomes of other organisms and comparative genome analysis [89], the identification of the common genomic variation in a number of human population groups (HapMap project [28,42,43]), the ENCODE project to identify the functional elements in the human genome and that of selected model organisms [12], and the genome-wide association studies to identify common risk variants for the common complex phenotypes [56,79,97] (Chap. 8.1). More recently, the 1000 Genomes Project (<http://www.1000genomes.org>) and other related efforts aim to identify all genetic variation in the genomes of individuals. The major challenge in the future is to provide causative links between genomic variants and phenotypic variation.

1.9.7 Unsolved Problems

Human genetics had been most successful by being able to guide work that was made possible by the development of techniques from various areas of

biology using Mendelian concepts. Important basic frontiers that are still being extended concern problems of gene regulation, especially during embryonic development, control of the immune system and of brain function. Human genetics is likely to contribute to these problems by imaginative use of the study of genetic variation and disease applying novel concepts and techniques. In medical genetics, the problem of common diseases including many birth defects requires study of the specific genes and their interactions involved in such diseases. Insights into the mechanisms of gene action during the aging process remain to be elucidated.

As shown by the many advances in description of genomic anatomy (see Chap. 2) where function is not yet fully understood, there is much need for research in both basic and translational approaches in order to elucidate the role of genomic biology and post-genomic interactions in health and disease. The remarkable similarity of humans and other mammals (and even of more primitive organisms) in both gene number and gene function had not been entirely expected, demonstrating that both new concepts and technical methods will be required to understand and utilize our current and future knowledge for applications in prevention and treatment of disease.

At first glance, the history of human genetics over the past 50 years reads like a succession of victories. The reader could conclude that human geneticists of the last generation pursued noble science to the benefit of mankind. However, how will posterity judge current efforts to make use of our science for the benefit of mankind as we understand it? Will the ethical distinction between selective abortion of a fetus with Down syndrome and infanticide of severely malformed newborns be recognized by our descendants? Are we again moving down the “slippery slope”?

Issues such as selective termination of pregnancy due to disadvantageous genomic variation need to be re-discussed and re-debated due to the ability to diagnose genomic variants with low-penetrance phenotypic consequences. As the dividing line between “severe phenotype” alleles and “low burden” alleles becomes blurred and individualized, consensus criteria and compromised solutions are fluid and constantly revised. Genetic medicine gradually becomes a central preoccupation of health professionals, the patients and their families, and presymptomatic healthy clients.

References

1. Adams J (1814) A treatise on the hereditary properties of disease. J Callow, London, p 33
2. Allen GE (1975) Genetics, eugenics and class struggle. *Genetics* 79:29–45
3. Andres AH, Jiv BV (1936) Somatic chromosome complex of the human embryo. *Cytologia* 7:317–388
4. Andres AH, Shiwago PI (1933) Karyologische Studien an myeloischer Leukämie des Menschen. *Folia Haematol* 49:1–20
5. Baltzer F (1967) Theodor Boveri: Life and work of a great biologist (1862–1915). Berkeley and Los Angeles, University of California Press
6. Barthelmess A (1952) Vererbungswissenschaft. Alber, Freiburg
7. Bateson W, Saunders E (1902) Experimental studies in the physiology of heredity. Reports to the Evolution Committee. R Soc 1:133–134
8. Bearn AG (1993) Archibald Garrod and the individuality of man. Clarendon, Oxford
9. Becker PE (1988) Zur Geschichte der Rassenhygiene (Wege ins Dritte Reich, vol 1). Thieme, Stuttgart
10. Becker PE (1990) SozialDarwinismus, Rassismus, Antisemitismus und völkischer Gedanke (Wege ins Dritte Reich, vol 2). Thieme, Stuttgart
11. Bernstein F (1924) Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen. *Klin Wochenschr* 3:1495–1497
12. Birney E et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
13. Brewer GJ (1971) Annotation: human ecology and expanding role for the human geneticist. *Am J Hum Genet* 23:92–94
14. Brock DJH (1977) Biochemical and cytological methods in the diagnosis of neural tube defects. *Prog Med Genet* 2:1–40
15. Capelle W (1953) Die Vorsokratiker. Kröner, Stuttgart
16. Correns C (1900) G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Ber Dtsch Bot Ges* 18:158–168
17. Costa LG and Eaton DL (eds) (2006) Gene-environmental interaction: Fundamentals of ecogenetics. J Wiley and Sons, Hoboken, New Jersey
18. Cremer T (1986) Von der Zellenlehre zur Chromosomentheorie. Springer, Berlin Heidelberg New York
19. Davenport CB (1911) Heredity in relation to eugenics. Henry Holt, New York
20. De Vries H (1889) Intracellulare Pangenesis. Fisher, Jena
21. Dunn LC (1962) Cross currents in the history of human genetics. *Am J Hum Genet* 14:1–13
22. Efroimson VP (1964) Vvedenie v medicinskuju genetiku (Introduction to medical genetics). Gos Izd Med Lit, Moscow
23. Emery AEH (1989) Joseph Adams (1756–1818). *J Med Genet* 26:116–118
24. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:636–640
25. Ephrussi B, Weiss MC (1965) Interspecific hybridization of somatic cells. *Proc Natl Acad Sci USA* 53:1040
26. Fisher RA (1918) The correlation between relatives on the supposition of mendelian inheritance. *Trans R Soc Edinb* 52:399–433
27. Fisher RA (1936) Has Mendel's work been rediscovered? *Ann Sci* 1:115–137
28. Frazer KA et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
29. Galton F (1865) Hereditary talent and character. Macmillans Mag 12:157
30. Garrod AE (1902) The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 2:1616–1620
31. Garrod AE (1909) Inborn errors of metabolism. Frowde, Oxford University Press, London
32. Garrod AE (1931) The inborn factors of disease. Clarendon Press, Oxford
33. Gärtner CF (1849) Versuche und Beobachtungen über die Bastardierung im Pflanzenreich. Hering, Stuttgart
34. Graham LR (1977) Political ideology and genetic theory: Russia and Germany in the 1920's. *Hastings Cent Rep* 7:30–39
35. Gurwitz D, Motulsky AG (2007) Drug reactions, enzymes, and biochemical genetics: 50 years later. *Pharmacogenomics* 8:1479–1484
36. Hardy GH (1908) Mendelian proportions in a mixed population. *Science* 28:49–50
37. Harper PS (2005) William Bateson, Human genetics and medicine. *Human Genet* 118:141–151
38. Harper PS (2008) A short history of medical genetics. Oxford University Press, Oxford 557 pp
39. Harris H (1969) Enzyme and protein polymorphism in human populations. *Br Med Bull* 25:5
40. Harris H, Watkins JF (1965) Hybrid cells from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature* 205:640
41. Hertwig O (1875) Beiträge zur Kenntnis der Bildung, Befruchtung und Theilung des tierischen Eies. I Abh Morph Jb 1:347–434
42. International HapMap Consortium (2003) T. I. H. The International HapMap Project. *Nature* 426:789–796
43. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
44. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945
45. Joravsky D (1970) The Lysenko affair. Harvard University Press, Boston
46. Kevles DJ (1985) In the name of eugenics. Genetics and the uses of human heredity, Knopf, New York
47. Köhl S (1994) The Nazi connection. Eugenics, American racism, and German national socialism. Oxford University Press, Oxford
48. Kölreuter JG (1761–1766) Vorläufige Nachricht von einigen das Geschlecht der Pflanzen betreffenden Versuchen und Beobachtungen, nebst Fortsetzung 1,2 und 3. Ostwalds Klassiker der Exakten Wissenschaften, no 41. Engelmann, Leipzig
49. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921

50. Landsteiner K (1900) Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutsersums und der Lymphe. Zentralbl Bakteriol 27:357–362
51. Landsteiner K, Wiener AS (1940) An agglutinable factor in human blood recognized by immune sera for rhesus blood. Proc Soc Exp Biol 43:223
52. Levine P, Burnham L, Katzin EM, Vogel P (1941) The role of isoimmunization in the pathogenesis of erythroblastosis fetalis. Am J Obstet Gynecol 42:925–937
53. Levine P, Stetson RE (1939) An unusual case of intragroup agglutination. JAMA 113:126–127
54. Levit SG (1936) The problem of dominance in man. J Genet 33:411–434
55. Ludmerer K (1972) Genetics and American society. Johns Hopkins University Press, Baltimore
56. Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118:1590–1605
57. Maupertuis PLM (1753) *Vénus Physique. The Earthly Venus* (trans: Brangier Boas S) (1966). Johnson Reprint Corporation, New York
58. McKusick VA (1975) The growth and development of human genetics as a clinical discipline. Am J Hum Genet 27: 261–273
59. McKusick VA (2007) History of medical genetics. In: Emery AEH, Rimoin DL (eds) Principles and practice of medical genetics, 5th edn. Churchill Livingstone, Philadelphia, Pennsylvania, pp 3–32
60. Mendel GJ (1865) Versuche über Pflanzenhybriden. Verhandlungen des Naturforschenden Vereins, Brünn
61. Motulsky AG (1957) Drug reactions, enzymes and biochemical genetics. JAMA 165:835–837
62. Motulsky AG (1959) Joseph Adams (1756–1818). A forgotten founder of medical genetics. Arch Intern Med 104: 490–496
63. Motulsky AG (2002) From pharmacogenetics and eugenics to pharmacogenomics. Medicina nei Secoli Arte e Scienza (Journal of History of Medicine) 14:683–705
64. Motulsky AG (2002) The work of Joseph Adams and Archibald Garrod: possible examples of prematurity in human genetics. In: Hook EB (ed) Prematurity and scientific discovery. University of California Press, Berkeley
65. Müller-Hill B (1988) Murderous science: Elimination by scientific selection of Jews, Gypsies, and others, Germany 1933–1945. Oxford University Press, Oxford English trans of Tödliche Wissenschaft, Rowohlt, Hamburg, 1984
66. Musto DF (1961) The theory of hereditary disease of Luis Mercado, chief physician to the Spanish Hapsburgs. Bull Hist Med 35:346–373
67. Nägeli C (1884) Mechanisch-physiologische Theorie der Abstammungslehre. R Oldenbourg, München and Leipzig
68. Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia: a molecular disease. Science 110:543
69. Penrose LS (1967) The influence of the English tradition in human genetics. In: Crow JF, Neel JV (eds) Proceedings of the Third International Congress of Human Genetics. The Johns Hopkins University Press, Baltimore, pp 13–25
70. Plato (360 BCE) The republic
71. Plato (360 BCE) The statesman
72. Ploetz A (1895) Die Tüchtigkeit unserer Rasse und der Schutz der Schwachen: ein Versuch über Rassenhygiene und ihr Verhältnis zu den humanen Idealen, besonders zum Sozialismus. Fischer, Berlin
73. Pollack W, Gorman JG, Freda VJ (1969) Prevention of Rh hemolytic disease. Prog Hematol 6:121–147
74. Proctor R (1988) Racial hygiene: Medicine under the Nazis. Harvard University Press, Cambridge, Massachusetts, pp 95–117
75. Reitlinger G (1961) The final solution. Barnes, New York
76. Rosenberg CE (1976) No other gods. On science and American social thought Johns Hopkins University Press, Baltimore
77. Roux W (1883) Über die Bedeutung der Kerntheilungsfiguren. W Engelmann, Leipzig
78. Rowley JD (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 243: 290–293
79. Saxena R et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336
80. Scriver CR, Childs B (1989) Garrod's inborn factors in disease. Oxford University Press, Oxford
81. Scriver CR, Beaudet AL, Sly WS, Valle D (eds) (2005) The metabolic and molecular bases of inherited disease, 8th edn. McGraw-Hill, New York
82. Stern C, Sherwood E (eds) (1966) The origin of genetics: a Mendel source book. WH Freeman, San Francisco
83. Stubbe H (1972) History of genetics, from prehistoric times to the rediscovery of Mendel's laws. MIT Press, Cambridge, Massachusetts, pp 12–50
84. Tschermak E (1900) Über künstliche Kreuzung bei *Pisum sativum*. Ber Dtsch Bot Ges 18:232–239
85. Van Beneden E (1883) Recherches sur la maturation de l'oeuf et la fecondation. Arch Biol IV:265
86. Vogel F (1959) Moderne Probleme der Humangenetik. Ergeb Inn Med Kinderheilk 12:52–125
87. Von Dungern E, Hirschfeld L (1911) On the group-specific structures of the blood. III. Z Immunitats 8:526–562
88. Von Verschuer O (1937) Was kann der Historiker, der Genealoge und der Statistiker zur Erforschung des biologischen Problems der Judenfrage beitragen? Festschrift Judenfrage 2:216–222
89. Waterston, R. H. et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562
90. Weiling F (1994) Johann Gregor Mendel. Forscher in der Kontroverse. V Med Genet 6:35–50
91. Weinberg W (1901) Beiträge zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen. Arch Gesamte Physiol 88:346–430
92. Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64:368–382
93. Weinberg W (1912) Weitere Beiträge zur Theorie der Vererbung. IV. Über Methode und Fehlerquellen der

- Untersuchung auf Mendelsche Zahlen beim Menschen. Arch Rass Ges Biol 9:165–174
94. Weindling PF (1989) Health, race and German politics between national unification and nazism 1870–1945. Cambridge University Press, Cambridge
95. Weingart P, Kroll J, Bayertz K (1988) Rasse. Blut und Gene, Suhrkamp, Frankfurt
96. Weiss KM (2008) Joseph Adams in the judgment of Paris. Evol Anthropol 17:245–249
97. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14, 000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678
98. Witkowski SA, Inglis JR (eds) (2008) Davenport's dream. 21st century reflections on heredity and eugenics. Cold Spring Harbor Laboratory Press, New York
99. Wolff CF (1759) *Theoria Generationis*, Halle
100. Zimmerman D (1973) Rh. The intimate history of a disease and its conquest. Macmillan, New York, p 371

Stylianios E. Antonarakis

Abstract The knowledge of the content of the individual human genomes has become a *sine qua non* for the understanding of the relationship between genotypic and phenotypic variability. The genome sequence and the ongoing functional annotation require both comparative genome analysis among different species and experimental validation. Extensive common and rare genomic variability exists that strongly influences genome function among individuals, partially determining disease susceptibility.

Contents

2.1	The Human Genome	31
2.1.1	Functional Elements.....	33
2.1.2	Repetitive Elements	40
2.1.3	Mitochondrial Genome	43
2.2	Genomic Variability	43
2.2.1	Single Nucleotide Polymorphisms.....	43
2.2.2	Short Sequence Repeats.....	46
2.2.3	Insertion/Deletion Polymorphisms (Indels).....	46
2.2.4	Copy Number Variants.....	47
2.2.5	Inversions	47
2.2.6	Mixed Polymorphisms	47
2.2.7	Genome Variation as a Laboratory Tool to Understand the Genome.....	48
	References	48

first needed to know the entire nucleotide sequence of the human genome. Thus an international collaborative project has been undertaken named “The Human Genome Project” to determine the nucleotide sequence of the human genome. The project was initiated on 1 October 1990 and was essentially completed in 2004. The potential medical benefits from the knowledge of the human genome sequence were the major rationale behind the funding of this international project. In addition, the involvement and contributions of the biotechnology company Celera may have provided the necessary competition for the timely completion of the project. The last (third) edition of this book was published in 1997 before the knowledge of the human genome sequence; thus, this fourth (“postgenome”) edition of the book proudly begins with the discussion of “genome anatomy,” as the genomic sequence was named by Victor McKusick.

The goals of the different phases of the Human Genome Project were to: (1) determine the linkage map of the human genome [1, 60]; (2) construct a physical map of the genome by means of cloning all fragments and arrange them in the correct order [32, 69]; (3) determine the nucleotide sequence of the genome; and (4) provide an initial exploration of the variation among human genomes.

As of October 2004 about 93% of the human genome (which corresponds to 99% of the euchromatic portion of the genome) had been sequenced to an accuracy of better than one error in 100,000 nucleotides

2.1 The Human Genome

In order to be able to understand the biological importance of the genetic information in health and disease (assign a particular phenotype to a genome variant) we

S.E. Antonarakis (✉)

Department of Genetic Medicine and Development, University of Geneva Medical School, and University Hospitals of Geneva, 1 rue Michel-Servet, 1211 Geneva, Switzerland
e-mail: Stylianios.Antonarakis@unige.ch

[3, 84, 137]. The DNA that was utilized for sequencing from the public effort came from a number of anonymous donors [84], while that from the industrial effort came from five subjects of which one is eponymous, Dr. J.C. Venter [85, 137]. The methodology used was also different between the two participants: the public effort sequenced cloned DNA fragments that had been previously mapped, while that of Celera sequenced both ends of unmapped cloned fragments and subsequently assembled them in continuous genomic sequences. Detailed descriptions of the genome content per chromosome have been published; the first “completed” chromosome published was chromosome 22 in 1999, chromosome 21 was published in 2000, and all other chromosomes followed in the next 6 years [38, 39, 45, 46, 55, 57, 62, 63, 66, 67, 70, 91, 97, 98, 101, 102, 111, 120, 121, 125, 131, 152, 153]. Figure 2.1 shows the parts of the genome (mainly the heterochromatic fraction) that have not yet been sequenced: the pericentromeric regions, the secondary constrictions of 1q, 9q, 16q, the short arms of acrocentric chromosomes (13p, 14p, 15p, 21p, 22p), and the distal Yq chromosome.

The total number of nucleotides of the finished sequence is 2,858,018,193 while the total estimated length that includes the current gaps is ~3,080,419,480 nucleotides (see Table 2.1, taken from the last hg18 assembly of the human genome <http://genome.ucsc.edu/goldenPath/stats.html#hg18>). The length of the human chromosomes ranges from ~46 Mb to ~247 Mb. The average GC content of the human genome is 41%. This varies considerably among the different chromosomes and within the different bands of each chromosome. Chromosomal bands positive for Giemsa staining have lower average GC content of 37%, while

Table 2.1 Taken from <http://genome.ucsc.edu/goldenPath/stats.html#hg18>, showing the number of nucleotides per chromosome in the reference genome. Chromosome “M” is the DNA of the mitochondrial genome (see Sect. 2.1.3)

NCBI Build 36.1, Mar. 2006 Assembly (hg18)

Chr Name	Assembled Size (inc. Gaps)	Sequenced Size	Total Gap Size	Non-Euch. Gap Size
1	247249719	224999719	22250000	20240000
2	242951149	237712649	5238500	4200000
3	199501827	194704827	4979000	4490000
4	191273063	187297063	3976000	3010000
5	180857866	177702766	3155100	3083000
6	170899992	167273992	3626000	3008000
7	158821424	154952424	3869000	3184000
8	146274826	142612826	3662000	3000000
9	140273252	120143252	20130000	18000000
10	135374737	131624737	3750000	2380000
11	134452384	131130853	3321531	3257000
12	132349534	130303534	2046000	1471000
13	114142980	95559980	18583000	17933000
14	106368585	88290585	18078000	18078000
15	100338915	81341915	18997000	18260000
16	88627254	76684754	9942500	9805000
17	78774742	77800220	974522	220000
18	76117153	74656155	1460998	1363998
19	63811651	55785651	8026000	8016000
20	62435964	59505253	2930711	1773661
21	46944323	34171998	12772325	12769767
22	49691432	34851332	14840100	14430000
X	154913754	151058754	3855000	3000000
Y	57772954	25652954	32120000	30500000
M	16571	16571	0	0
<hr/>				
Overall				
Chrom	3080436051	2858034764	222401287	205472426

in Giemsa-negative bands the average GC content is 45%. Interestingly, Giemsa-negative bands are gene-rich regions of DNA (see Chap. 3, Sect. 3.2.4).

Figure 2.2 shows the current status of the “completion” of the human genome sequence [3]. Red bars above the chromosomes represent the sequence gaps. The DNA content of the red blocks (heterochromatin) is still unknown. Heterochromatic regions of chromo-

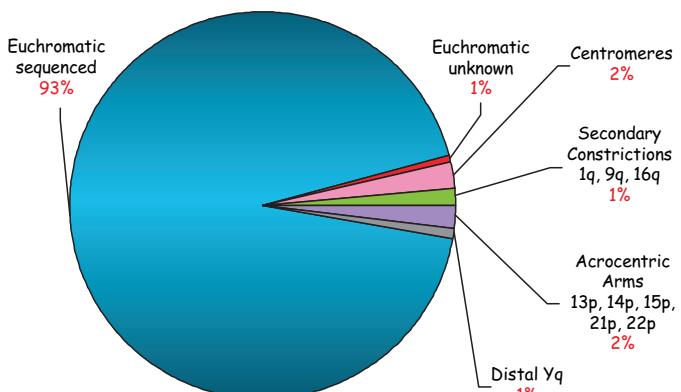


Fig. 2.1 Pie chart of the fractions of the genomes sequenced (blue) and not sequenced (non-blue)

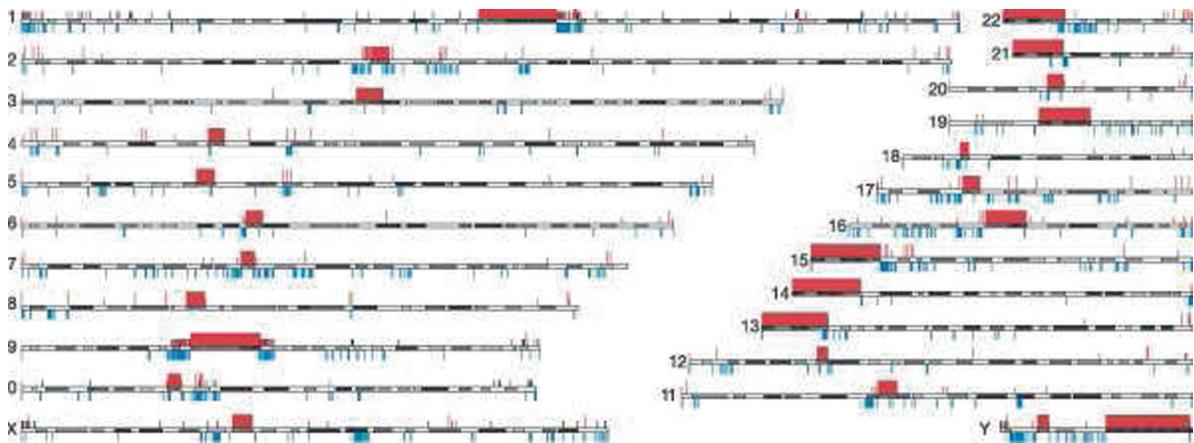


Fig. 2.2 Schematic representation of the completion of the human genome per chromosome. *Red regions* represent areas not sequenced; *blue regions* below the chromosomal line represent gaps in the sequences. The major blocks of unknown sequence include the short arms of acrocentric chromosomes, the pericentromeric sequences, and the large heterochromatic regions (From [3])

somes are those that remain highly condensed throughout the cell cycle (see Chap. 3, Sect. 3.2.1); it is thought that transcription is limited in these regions that contain a considerable number of repetitive elements that renders the assembly of their sequence almost impossible.

The sequence of the human genome is freely and publicly available on the following genome browsers, which also contain many additional annotations (see also Chap. 29):

- (a) <http://genome.ucsc.edu/>
- (b) <http://www.ensembl.org/>
- (c) <http://www.ncbi.nlm.nih.gov/genome/guide/human/>

Representative pages of two of these browsers are shown in Fig. 2.3.

There is now a considerable effort internationally to identify all the functional elements of the human genome. A collaborative project called ENCODE (ENyclopedia Of DNA Elements) is currently in progress with the ambitious objective to identify all functional elements of the human genome [2, 19].

The genome of modern humans, as a result of the evolutionary process, has similarities with the genomes of other species. The order of genomic elements has been conserved in patches within different species such that we could recognize today regions of synteny in different species, i.e., regions that contain orthologous genes and other conserved functional elements. Figure 2.4 shows a synteny map of conserved genomic segments in human and mouse.

The current classification of the functional elements of the genome contains:

1. Protein-coding genes
2. Noncoding, RNA-only genes
3. Regions of transcription regulation
4. Conserved elements not included in the above categories

2.1.1 Functional Elements

2.1.1.1 Protein-Coding Genes

The total number of protein-coding genes is a moving target, since this number depends on the functional annotation of the genome, the comparative analysis with the genomes of other species, and the experimental validation. The so-called CCDS set (consensus coding sequence) is built by consensus among the European Bioinformatics Institute (<http://www.ebi.ac.uk/>), the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>), the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/>), and the University of California, Santa Cruz (UCSC; <http://www.cbse.ucsc.edu/>). At the last update (5 July 2009; genome build 36.3) CCDS contains 17,052 genes. This is the minimum set of protein-coding genes included in all genomic databases. The reference sequence (RefSeq) collection of genes of the NCBI contains 20,366 protein-coding gene entries (<http://www.ncbi.nlm.nih.gov/>

2

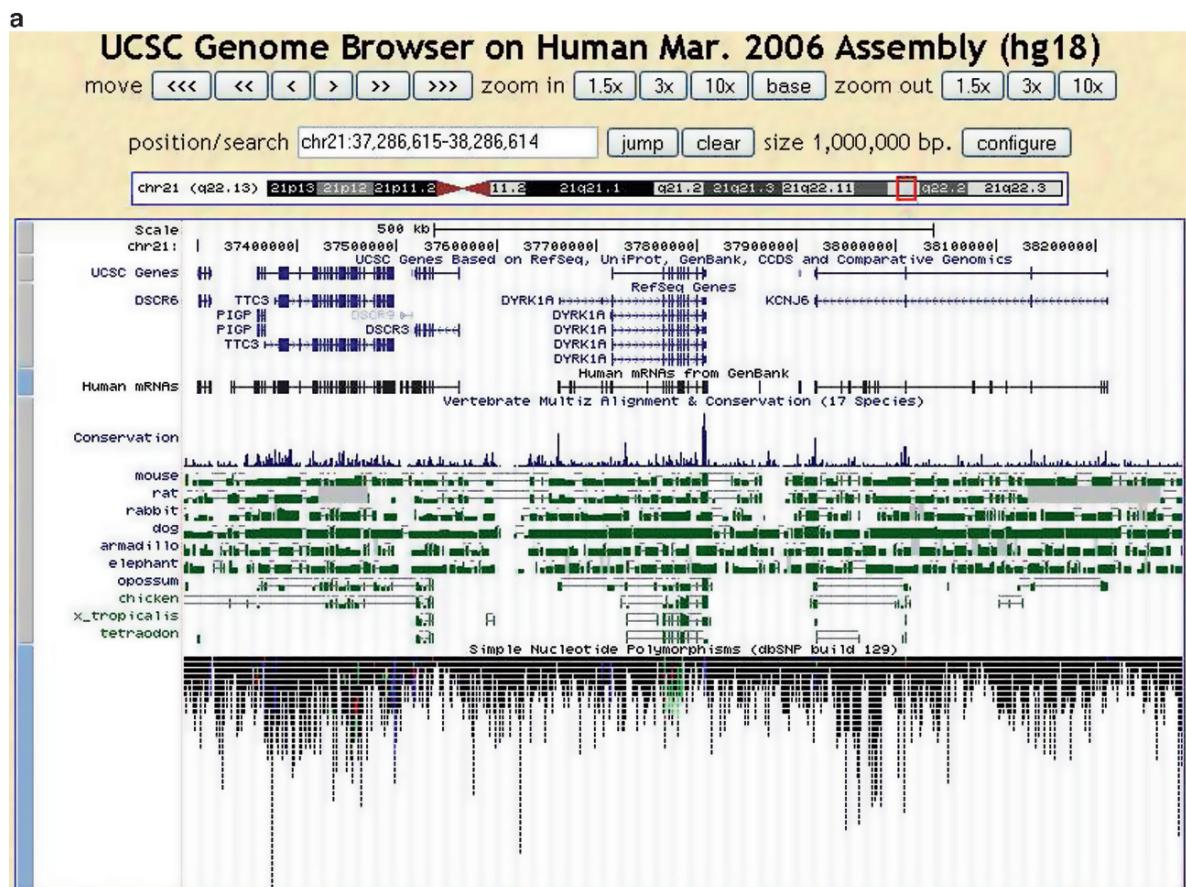


Fig. 2.3 (a) Screenshot of the UCSC genome browser (<http://genome.ucsc.edu/>) for a 1-Mb region of chromosome 21 (21: 37,286,615–38,286,614). Among the many features that could be displayed, the figure shows genes, sequence conservation in 17 species, and single nucleotide polymorphisms (SNPs) that map in this 1-Mb region. The tracks shown from top to bottom

include: a scale for the genomic region, the exact location in nucleotides, schematic representation of genes included in the UCSC database, the mRNAs from GenBank, the conservation in the species shown, and the location of SNPs. The color of some SNPs corresponds to synonymous and nonsynonymous substitutions.

RefSeq/); the UCSC collection of genes contains 23,008 entries (<http://genome.ucsc.edu/>); the Ensembl browser contains 21,416 entries (23 June 2009; build 36; http://www.ensembl.org/Homo_sapiens/Info/StatsTable). The total number of annotated exons listed in the Ensembl database is 297,252 (23 June 2009; build 36). The discrepancy among the databases reflects the ongoing and unfinished annotation of the genome.

Table 2.2 lists the number of protein-coding and other genes in humans taken from different databases.

The human genes are not equally distributed in the chromosomes. In general, Giemsa pale bands are gene rich, and this results in unequal numbers of genes per size unit for the different chromosomes. Figure 2.5 from [84] displays the gene density per megabase for

each chromosome and the correlation with CpG-rich islands.

Chromosomes 22, 17, and 19 are unusually gene-rich, while chromosomes 13, 18, and X are relatively gene-poor (interestingly, trisomies for chromosomes 13 and 18 are among the few human trisomies at birth). The average number of exons per gene is nine, and the average exon size is 122 nucleotides. Thus, the total number of annotated exons range from 210,000 to 300,000 (depending on the database), and the total exonic genome size is up to 78 Mb.

The mapping position of the genes can be seen in the genome browsers, and their names can be found in the gene nomenclature Web site, which contains 28,182 entries (<http://www.genenames.org/>; 30 June 2009).

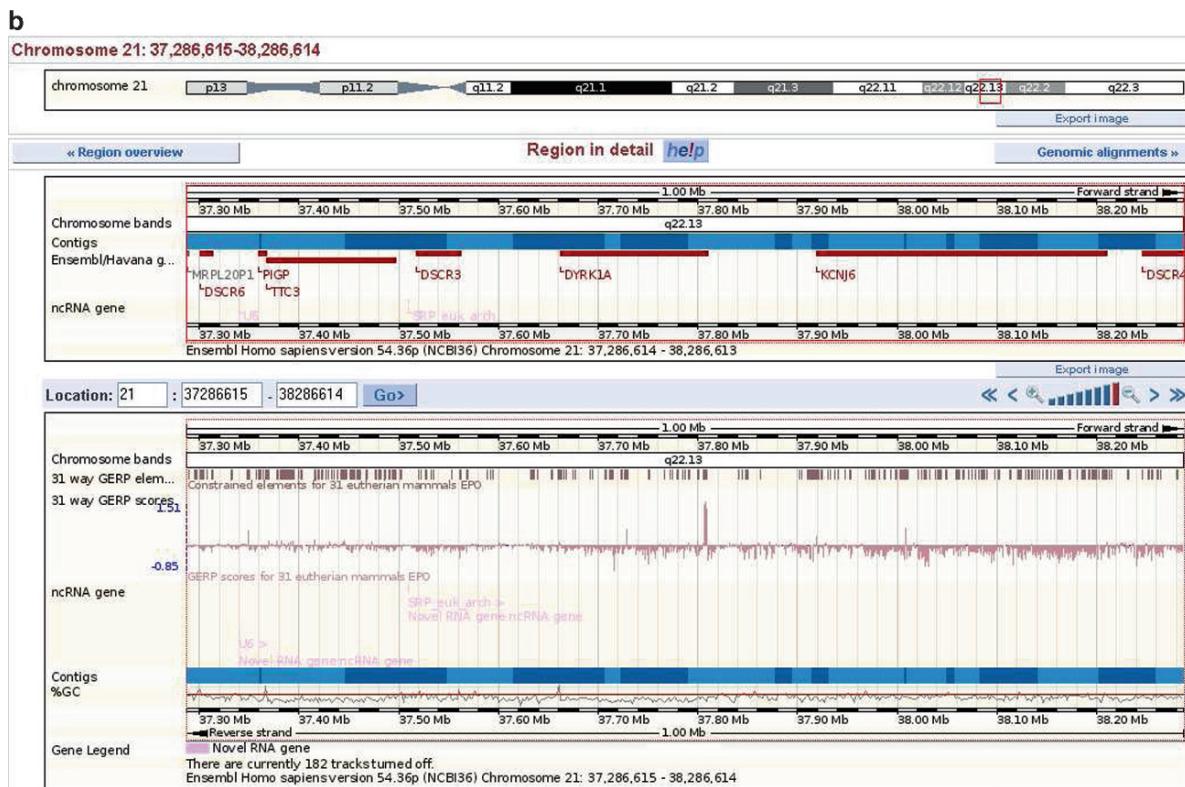


Fig. 2.3 (continued) (b) Screenshot of the Ensembl genome browser (<http://www.ensembl.org>) for a 1-Mb region of chromosome 21 (21: 37,286,615–38,286,614). Among the many features that could be displayed, the figure shows genes (Ensembl/Havana gene track), noncoding RNAs (ncRNA gene

track), sequence conservation in 31 species (31-way GERP track), and GC content in this 1-Mb region. The different browsers have similarities and differences, and some features could only be displayed in one browser (for details see Chaps. 29.1 and 29.2)

A single gene may have different isoforms due to alternative splicing of exons, alternative utilization of the first exon, and alternative 5' and 3' untranslated regions. There are on average 1.4–2.3 transcripts per gene according the different databases (Table 2.2); this is likely an underestimate since, in the pilot ENCODE 1% of the genome that has been extensively studied, there are 5.7 transcripts per gene [19, 61]. The average number of exons per gene, depending on the database, ranges from 7.7 to 10.9.

The size of genes and number of exons vary enormously. The average genomic size of genes (according to the current annotation) is 27 kb. There are, however, small genes that occupy less than 1 kb, and large genes that extend to more than 2,400 kb of genomic space. There are intronless genes (e.g., histones) and others with more than 360 introns (e.g., titin).

The initial results of the ENCODE and other similar projects provided evidence for additional exons to the

annotated genes; these exons could be hundreds of kilobases away (usually 5') to the annotated gene elements [19, 40, 44]. In addition, there is evidence for chimeric transcripts that join two “independent” genes [103]. The investigation of these complicated transcripts is ongoing, and the functional significance of them is unknown.

Protein-coding genes can be grouped in families according to their similarity with other genes. These families of genes are the result of the evolutionary processes that shaped up the genomes of the human and other species. The members of the gene families could be organized in a single cluster or multiple clusters, or could be dispersed in the genome. Examples of gene families include the globin, immunoglobulin, histones, and olfactory receptors gene families. Furthermore, genes encode proteins with diverse but recognizable domains. The database Pfam (<http://pfam.sanger.ac.uk/>, <http://www.uniprot.org/>) is a comprehensive collection of protein domains and families [48]; the current release

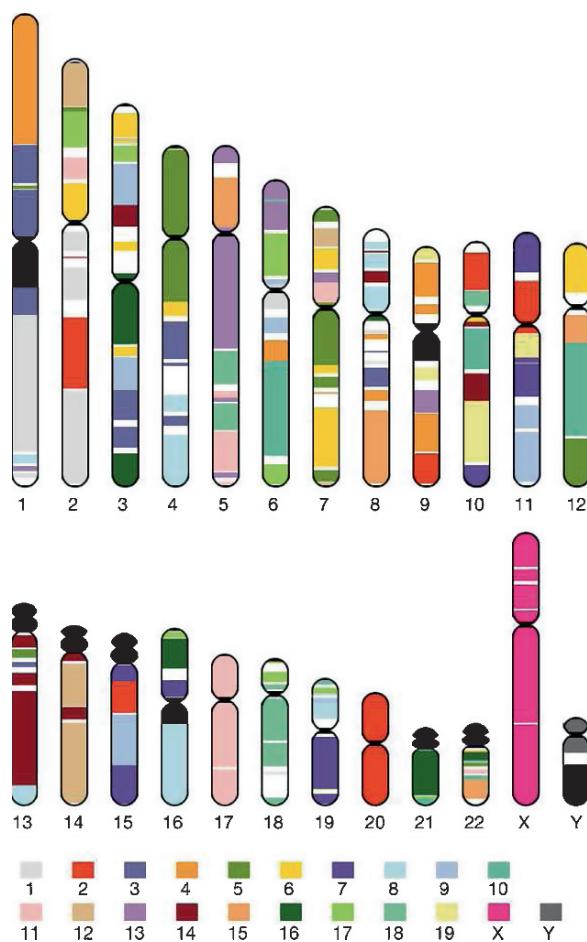


Fig. 2.4 Schematic representation of the observed genomic segments between the human and mouse genomes. The color code of the human chromosomes corresponds to the different mouse chromosomes shown on the bottom. For example, human chromosome 20 is all homologous to mouse chromosome 2; human chromosome 21 is homologous to mouse chromosomes 16, 17, and 10. Centromeric, and heterochromatic regions, and acrocentric p-arms are in black. (From [84])

of Pfam (23.0) contains 10,340 protein families. For example, the WD40 domain family (PF00400) includes 609 human genes, while the homeobox domain family

(PF00046) has 430 genes. The identification of domains helps in the prediction of the function and structure of a protein.

Pseudogenes are “dead” nonfunctional genes. These sequences that could be transcribed and spliced contain mutations that render them inactive. Pseudogenes could be generated by several mechanisms that include:

1. Gene duplication events in which one of the duplicated copies accumulates inactivating mutations; alternatively, the duplicated genes may be truncated. These pseudogenes are also called nonprocessed pseudogenes.
2. Transposition events in which a copy of cDNA is reinserted into the genome. These pseudogenes, also called “processed,” are not functional, usually because they lack regulatory elements that promote transcription. In addition, inactivating mutations also occur in processed pseudogenes.

The current estimated number of human pseudogenes (according to one of the databases <http://www.pseudogene.org/human/index.php>) [151] is 12,534 (~8,000 are processed and ~4,000 duplicated pseudogenes; build 36); while according to the Ensembl browser the number is 9,899 (build 36; 23 June 2009). These pseudogenes belong to 1,790 families; e.g., the immunoglobulin gene family has 1,151 genes and 335 pseudogenes, while the protein kinase gene family has 1,159 genes and 159 pseudogenes (<http://pseudofam.pseudogene.org/pages/psfam/overview.jsf>).

The total number of human genes is not dramatically different from that of other “less” complex organisms. Figure 2.6 depicts the current estimate of the protein-coding gene number for selected species.

2.1.1.2 Noncoding, RNA-Only Genes

Besides the protein-coding genes, there is a growing number of additional genes (transcripts) that produce an

Table 2.2 Human gene, exon, and transcript counts from various databases

Database (June 2009)	Protein-coding genes	RNA-only genes	Total genes	Total number of transcripts	Total number of exons	Average exons per gene	Average transcripts per gene
CCDS	17,052			45,428			2.7
Ensembl	21,416	5,732	27,148	62,877	297,252	10.9	2.3
UCSC	23,008	9,155	32,163	66,802	246,775	7.7	2.1
RefSeq	20,366	2,044	22,410	31,957	211,546	9.4	1.4

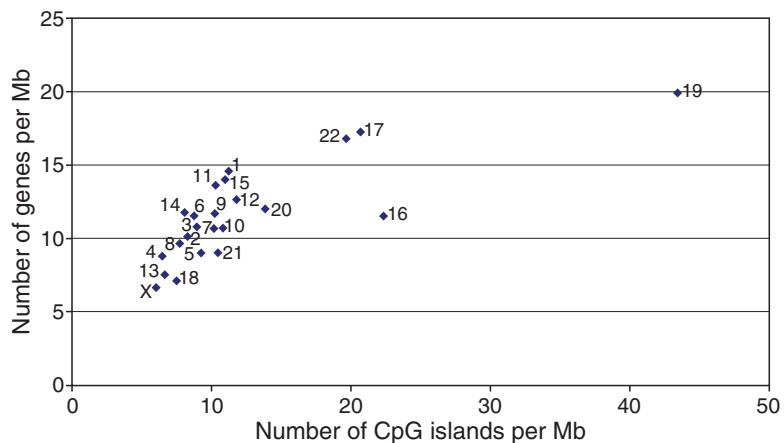


Fig. 2.5 Gene density per chromosome, and correlation with CpG-rich islands of the genome. Chromosome 19 for, example, has the highest gene content and the highest CpG island content. (From [84])

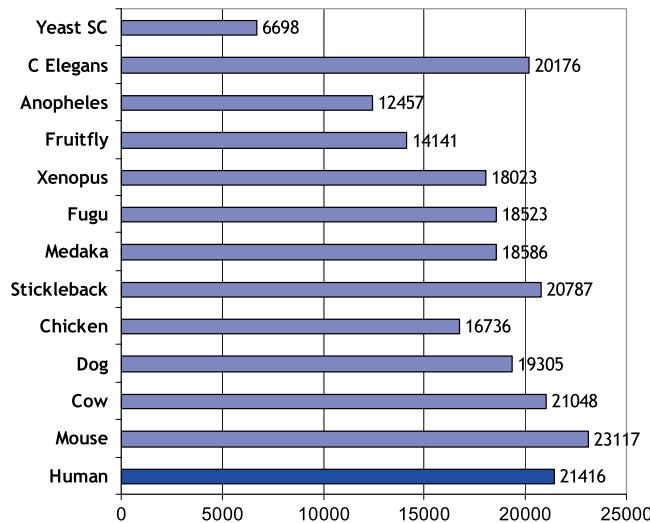


Fig. 2.6 Histogram of the current (5 July 2009) estimate of the number of protein-coding genes in different selected species from the Ensembl browser. These numbers are subject to change

RNA that is not translated to protein (see the databases <http://biobases.ibch.poznan.pl/ncRNA/>, <http://www.ncbi.nlm.nih.gov/frnadb/search.html>, <http://www.sanger.ac.uk/Software/Rfam/> and [56]). Table 2.2 contains the current number of these genes, which ranges from 2,044 in RefSeq to 9,155 in the UCSC browser.

The different classes of RNA-only genes are briefly discussed below:

Ribosomal RNA (rRNA) Genes [53, 82, 84]: ~650–900. These are genes organized in tandemly arranged

clusters in the short arms of the five acrocentric chromosomes (13, 14, 15, 21, and 22). The transcripts for 28 S, 5.8 S, and 18 S rRNAs are included in one transcription unit, repeated 30–50 times per chromosome. These tandemly arranged genes are continuously subjected to concerted evolution, which results in homogeneous sequences due to unequal homologous exchanges. The transcripts for the 5 S rRNAs are also tandemly arranged, and the majority map to chromosome 1qter. There exist also several pseudogenes

for all classes. The total number of these genes is polymorphic in different individuals. The best estimates of the number of rRNA genes are:

28 S (components of the large cytoplasmic ribosomal subunit)	~150–200
5.8 S (components of the large cytoplasmic ribosomal subunit)	~150–200
5 S (components of the large cytoplasmic ribosomal subunit)	~200–300
18 S (components of the small cytoplasmic ribosomal subunit)	~150–200

Transfer RNA (tRNA): ~500 (49 Types). At the last count there are 497 transfer RNA genes (usually 74–95 nucleotides long) encoded by the nucleus and transcribed by RNA polymerase III (additional tRNAs are encoded by the mitochondria genome). There are also 324 tRNA pseudogenes [84]. The tRNA nuclear genes form 49 groups for the 61 different sense codons. Although the tRNA genes are dispersed throughout the genome, more than 50% of these map to either chromosomes 1 or 6; remarkably 25% of tRNAs map to a 4-Mb region of chromosome 6.

Small Nuclear RNA (snRNA) [84, 87, 105]: ~100. These are heterogeneous small RNAs. A notable fraction of these are the spliceosome [139] RNA genes many of which are uridine-rich; the U1 group contains 16 genes, while U2 contains six, U4 4, U6 44, and the other subclasses are represented by one member. Some of these genes are clustered, and there is also a large number of pseudogenes (more than 100 for the U6 class).

Small Nucleolar RNA (snoRNA): ~200. This is a large class of RNA genes that process and modify the tRNAs and snRNAs [135, 147]. There are two main families: C/D box snoRNAs that are involved in specific methylations of other RNAs; and H/ACA snoRNAs, mostly involved in site-specific pseudouridylations. Initially, there were 69 recognized in the first family and 15 in the second [84]; however, the total number is probably larger. A cluster of snoRNAs maps to chromosome 15q in the Prader–Willi syndrome region (at least 80 copies); deletions of which are involved in the pathogenesis of this syndrome [26, 117]. Another cluster of snoRNAs maps to chromosome 14q32 (~40 copies). The majority of snoRNAs map to introns of protein-coding genes and can be transcribed by RNA polymerase II or III.

Micro RNAs (miRNA): (706 Entries on 26 June 2009). These are single-stranded RNA molecules of

about 21–23 nt in length that regulate the expression of other genes. miRNAs are encoded by RNA genes that are transcribed from DNA but not translated into protein; instead they are processed from primary transcripts known as pri-miRNA to short stem-loop structures called pre-miRNA and finally to functional miRNA. Mature miRNA molecules are complementary to regions in one or more messenger RNA (mRNA) molecules, which they target for degradation. A database of the known and putative miRNAs, and their potential targets, can be found in <http://microrna.sanger.ac.uk/>. miRNAs have been shown to be involved in human disorders.

Large Intervening Noncoding RNAs (LincRNAs): ~1,600. This new class has been recently identified using trimethylation of Lys4 of histone H3 as a genomic mark to observe RNA PolII transcripts at their promoter, and trimethylation of Lys36 of histone H3 marks along the length of the transcribed region [95] to identify the spectrum of PolII transcripts. Approximately 1,600 such LincRNA transcripts have been found across four mouse cell types (embryonic stem cells, embryonic fibroblasts, lung fibroblasts, and neural precursor cells) [59]. Among the “exons” of these LincRNAs, approximately half are conserved in mammalian genomes, and are thus present in human. Since this class was described in 2009, further work is needed for its characterization and validation, as well as the potential overlap of its members with the other classes.

Other Noncoding RNAs [7, 75, 113, 126, 136]: ~1,500. The field of noncoding RNA series is constantly expanding. Some of these RNA genes include molecules with known function such as the telomerase RNA, the 7SL signal recognition particle RNA, and the XIST long transcript involved on the X-inactivation [23]. There are also numerous antisense noncoding RNAs, and the current effort to annotate the genome suggests that a substantial fraction of the transcripts are noncoding RNAs.

2.1.1.3 Regions of Transcription Regulation

The genome certainly contains information for the regulation of transcription. The current list of these regulatory elements includes promoters, enhancers, silencers, and locus control regions [92]. These elements are usually found in *cis* to the transcriptional

unit, but there is growing evidence that there is also *trans* regulation of transcription. The discovery of the regulatory elements, their functional interrelationship, and their spatiotemporal specificity provides a considerable challenge. A systematic effort during the pilot ENCODE project has provided initial experimental evidence for genomic regions with enriched binding of transcription factors [19, 80, 86, 133]. A total of 1,393 regulatory genomic clusters were, for example, identified in the pilot ENCODE regions; remarkably only ~25% of these map to previously known regulatory regions and only ~60% of these regions overlap with evolutionarily constrained regions. These results suggest that many novel regulatory regions will be recognized in the years to come, and also that there exist regions of transcriptional regulation that are not conserved and thus novel for different clades and species. The use of model organisms facilitates the experimental validation of regulatory elements, and there are systematic efforts underway for the exploration of conserved elements ([106] and <http://enhancer.lbl.gov/>).

2.1.1.4 Conserved Elements Not Included in the Above Categories

Since it is assumed that functional DNA elements are conserved while nonfunctional DNA diverges rapidly,

it is expected that all other conserved elements are of interest and should be studied for potential pathogenic variability. How much of the human genome is evolutionarily conserved? The answer to this question depends on the species compared and the time of their common ancestor. Comparative genome analysis between human and mouse, for example, is particularly instructive, since the time of the common ancestor between these two species is estimated to be ~75 million years ago, and thus the conserved elements are likely to be functional. Approximately 5% of the human genome is conserved compared to mouse [145] (and to several other mammalian genomes). Of this, ~1–2% are the coding regions of protein-coding genes, and ~3% are conserved non-coding DNA sequences (CNCs; Fig. 2.7) [41, 42]. The function of the majority of CNCs is unknown. Please note that this 5% conserved fraction between human and mouse is an underestimate of the functional fraction of the human genome, which is likely to be bigger and to contain additional sequences not conserved with the mouse.

The ENCODE pilot project [19, 90], with data from 1% of the human genome and sequences from the orthologous genomic regions from 28 additional species, also estimated that the constrained portion of the human genome is at least ~4.9%; remarkably, 40% of this genomic space is unannotated and thus of unknown function (Fig. 2.8).

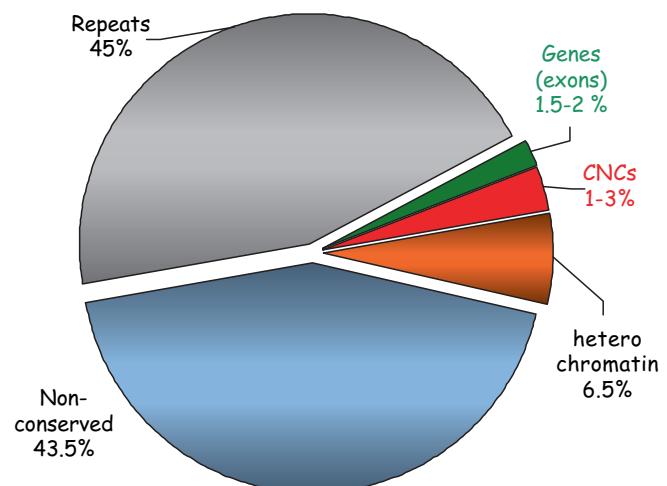


Fig. 2.7 The pie-chart depicts the different fractions of the genome. CNCs, conserved noncoding sequences

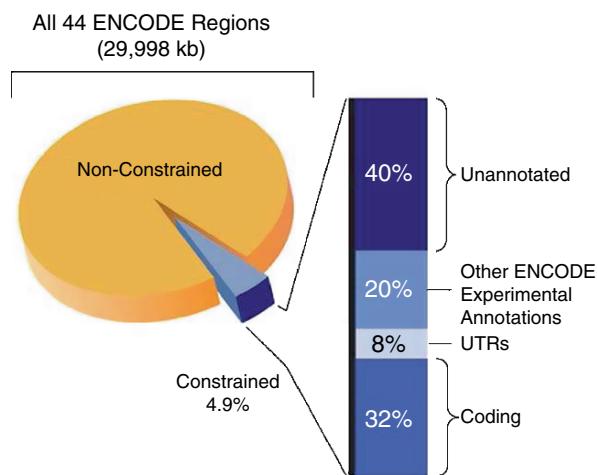


Fig. 2.8 The fractions of different genomic annotations among the 4.9% of constrained sequences in the human genome. Data from the pilot ENCODE project; figure taken from [19]. *UTR*, untranslated region; other ENCODE experimental annotations refers to the fraction of the genome that has been identified using a variety of experimental techniques for transcription, histone modifications, chromatin structure, sequence specific factors, and DNA replication. More information on these experiments is included in Table 1 of [19]

2.1.2 Repetitive Elements

The function of the majority of the human genome is unknown. Remarkably, ~45% of the genome is composed of repetitive elements, and another ~43% is not conserved and does not belong to the functional categories mentioned above. The different interspersed repeats of the human genome are shown in the Fig. 2.9 (from [84]):

- LINEs (long interspersed nuclear elements [76, 77]) are autonomous transposable elements, mostly truncated nonfunctional insertions (average size of 900 bp). More than 20% of the human genome is polluted by LINEs. Transposable elements are mobile DNA sequences which can migrate to different regions of the genome. Autonomous are those that are capable of transposing by themselves. A small fraction of LINEs (~100) are still capable of transposing. The full LINE element is 6.1 kb long, has an internal PolII promoter, and encodes two open reading frames, an endonuclease, and a reverse transcriptase. Upon insertion a target site duplication of 7–20 bp is formed. There are a few subclasses of LINEs according to their consensus sequence. The subfamily LINE1 is the only one capable of autonomous retrotransposition (copy itself and pasting copies back into the genome in multiple places). These LINEs enable transposition of SINEs (defined below), processed pseudogenes, and retrogenes [76, 77]. LINE retrotransposition has been implicated in human disorders [78]. LINEs are more abundant in G-dark bands of human chromosomes.
- SINEs (short interspersed nuclear elements [18]) mainly include the Alu repeats, which are the most abundant repeats in the human genome, occurring on average in every 3 kb. Thus, 13% of the genome is polluted by Alu sequences and other SINEs. They are inactive elements originated from copies of tRNA or from signal recognition particle (SRP; 7SL) RNA. The full-length element is about 280 nt long and consists of two tandem repeats each ~120 nt followed by polyA.

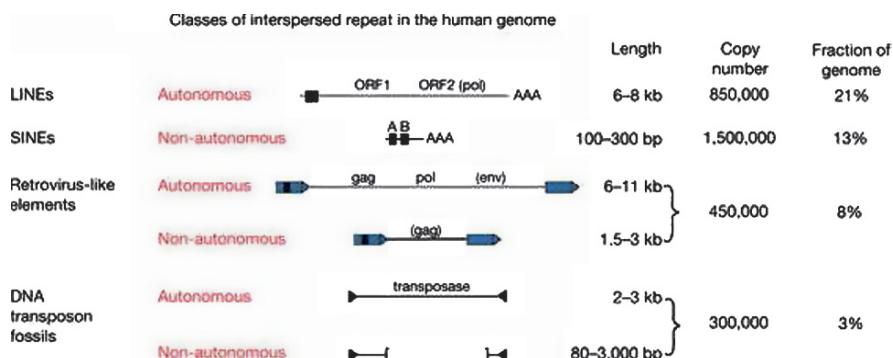


Fig. 2.9 Depicts some basic characteristics of the classes of interspersed repeats in the human genome. For more explanations, see text. (From [84])

Alu sequences are transcriptionally inactive, and are GC-rich. SINEs can retrotranspose in a non-autologous way, since they use the LINE machinery for transposition. Because of their abundance, they could mediate deletion events in the genome that result in human disorders [37]. SINEs are more abundant in G-light bands of human chromosomes (see Sect. 3.2.4).

- Retrovirus like (LTR transposons) are elements flanked by long terminal repeats. Those that contain all the essential genes are theoretically capable of transposition, but that has not happened in the last several million years. Collectively they account for 8% of the genome. Most are known as HERV (human endogenous retroviral sequences) and are transposition defective. Transcription from the HERV genes may modulate the transcriptional activity of nearby protein-coding genes [22].
- DNA transposon fossils [127] have terminal inverted repeats and are no longer active; they include two main families, MER1 and MER2, and comprise 3% of the genome.

More update information about repeats can be found in <http://www.girinst.org/server/RepBase/>.

2.1.2.1 Segmental Duplications

Approximately 5.2 % of the human genome consists of segmental duplications or duplicons, i.e., regions of more than 1 kb, with greater than 90% identity, that are present more than once in the genome. Segmental duplications are either intrachromosomal (on the same chromosome, 3.9%), or interchromosomal (on different chromosomes, 2.3%; Fig. 2.10). Most of the “duplicons” are in the pericentromeric regions.

Figure 2.11 shows the distribution of intrachromosomal duplicons in the human genome [16, 118]. These duplications are important in evolution and as risk factors for genomic rearrangements that cause human disorders because of unequal crossing-over in meiosis (pathogenic microdeletions and microduplications). Some examples of these include cases of α-thalassemia [65] on chromosome 16p, Charcot-Marie–Tooth syndrome [104] on chromosome 17p, and velo-cardiac-facial syndrome [96] on chromosome 22q, Williams–Beuren syndrome [107] on

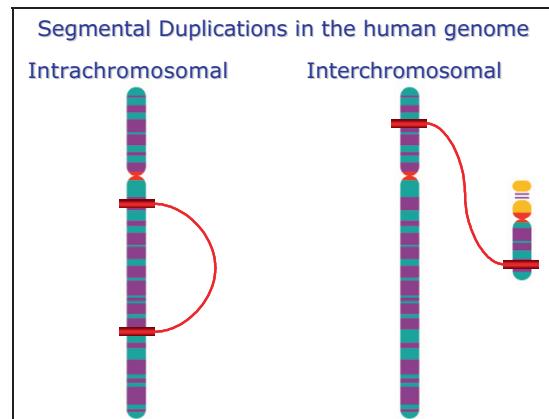


Fig. 2.10 Schematic representation of intra- and inter-chromosomal segmental duplications. The repeat element is shown in red, and there is a *connecting line* indicating the highly homologous sequences

chromosome 7q, and Smith–Magenis syndrome [29] on chromosome 17p.

2.1.2.2 Special Genomic Structures Containing Selected Repeats

2.1.2.2.1 Human Centromeres

Human centromeres consist of hundreds of kilobases of repetitive DNA, some chromosome specific and some nonspecific [114, 122, 124]. Actually, most of the remaining sequence gaps in the human genome are mapped near and around centromeres. The structure of human centromeres is unknown, but the major repeat component of human centromeric DNA is an α-satellite or alphoid sequence [30] (a tandem repeat unit of 171 bp that contains binding sites for CENP-B, a centromeric-binding protein; see also Chap. 3, Sect. 3.2.3). Figure 2.12 shows an example of the structure of two human centromeres [3].

2.1.2.2.2 Human Telomeres

Human telomeres [109] consist of tandem repeats of a sequence $(TTAGGG)_n$ that spans about 3–20 kb, beyond which at the centromeric side there are about 100–300 kb of subtelomeric-associated repeats [3] before any unique sequence is present.

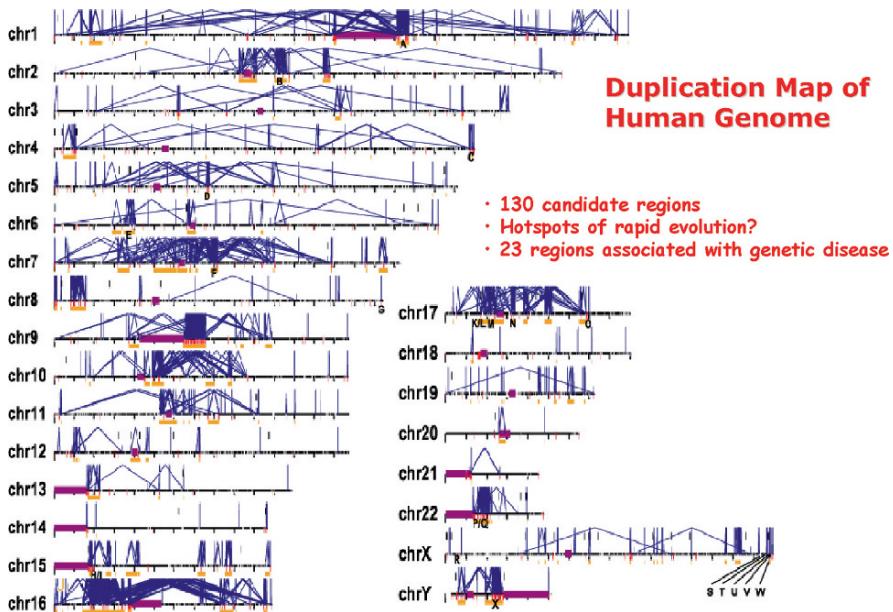


Fig. 2.11 Schematic representation of the intrachromosomal segmental duplications (from [16]). In each chromosome a *blue line* links a duplication pair. For example, on chromosome 21 there is

only one dupilon shown; in contrast, on chromosome 22 there is a considerable number of duplications. *Richly blue areas* are considered susceptible to microduplication/microdeletion syndromes

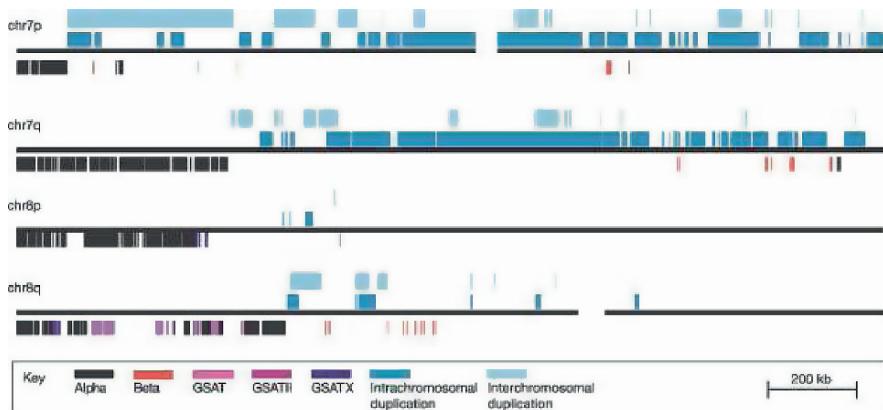


Fig. 2.12 Examples of sequence organization of two human centromeres (chromosomes 7 and 8, (from [3])). Alphoid repeats are the major component of this special chromosomal structure;

in addition, several other repetitive elements border the alphoid sequences. The length of these regions is also polymorphic in different individuals

Figure 2.13 schematically shows the sequence organization of six human subtelomeric regions.

2.1.2.2.3 Short Arms of Human Acrocentric Chromosomes

The finished sequence of the human genome does not include the short arms of acrocentric chromosomes (13p, 14p, 15p, 21p, and 22p). Cytogenetic data show

that the p arms contain large heterochromatic regions of polymorphic length [35, 138]. Molecular analysis revealed that they are composed mainly of satellite and other repeat families, including satellites I (AT-rich repeat of a monomer of 25–48 bp [73]), II (monomer repeat 5 bp [68]), III (monomer repeat also 5 bp [31]), β-satellite (a tandem repeat unit of 68 bp of the Sau3A family [94, 146]), and repeats ChAB4 [36], 724 [83], and D4Z4-like [89]. These repeats have a complex pattern and are often organized in subfamilies shared

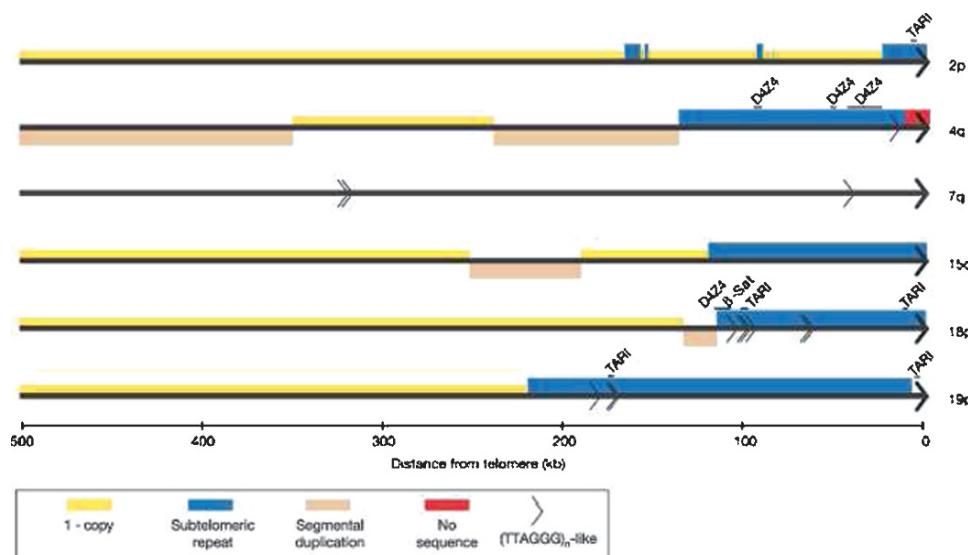


Fig. 2.13 Examples of the sequence organization of six human telomeres (chromosomes 2pter, 4qter, 7qter, 15qter, 18pter, and 19pter, taken from [3]). The arrows represent the TTAGGG repeat, while the blue regions depict the subtelomeric repeats that mainly consist of TAR1 (telomere

associated repeat 1 family [24]), D4Z4 (a 3.3-kb tandem repeat, each copy of which contains two homeoboxes and two repetitive sequences, LSau and hhsmp3 [64]) and β -satellite sequences (a tandem repeat unit of 68 bp of the Sau3A family [94])

between different acrocentric chromosomes. The p arms encode the ribosomal (RNR) gene [53, 82] but may also encode other genes [88, 130]. Currently there is an initiative to sequence the short arm of chromosome 21 and thus extrapolate on the structure of the additional p arms of the other acrocentrics [88].

The most common chromosomal rearrangements in humans are Robertsonian translocations (~1 in 1,000 births), which involve exchanges between acrocentric p arms. Three to five percent of these translocations are associated with phenotypic abnormalities [143].

2.1.3 Mitochondrial Genome

In human cells there is also the mitochondrial genome, which is 16,568 nucleotides long and encodes for 13 protein-coding genes, 22 tRNAs, one 23 S rRNA, and one 16 S rRNA ([140–142]; <http://www.mitomap.org>). The mitochondria genome-encoded genes are all essential for oxidative phosphorylation and energy generation in the cell. Each cell has hundreds of mitochondria and thousands (10^3 – 10^4) of mitochondrial DNA (mtDNA) copies. Human mtDNA has a mutation rate ~20 times higher than nuclear DNA. The inheri-

tance of mtDNA is exclusively maternal (the oocyte contains 10^5 mtDNA copies). Several human phenotypes are due to pathogenic mutations in the mitochondrial genome [140] (Fig. 2.14).

2.2 Genomic Variability

The human genome is polymorphic, i.e., there are many DNA sequence variants among different individuals. These variants are the molecular basis of the genetic individuality of each member of our species. In addition, this genetic variability is the molecular substrate of the evolutionary process. Finally, this variability causes disease phenotypes or predispositions to common complex or multifactorial phenotypes and traits.

2.2.1 Single Nucleotide Polymorphisms

The majority of the DNA variants are single nucleotide substitutions commonly known as SNPs (single nucleotide polymorphisms). The first SNPs were identified in 1978 in the laboratory of Y.W. Kan 3' to the β -globin

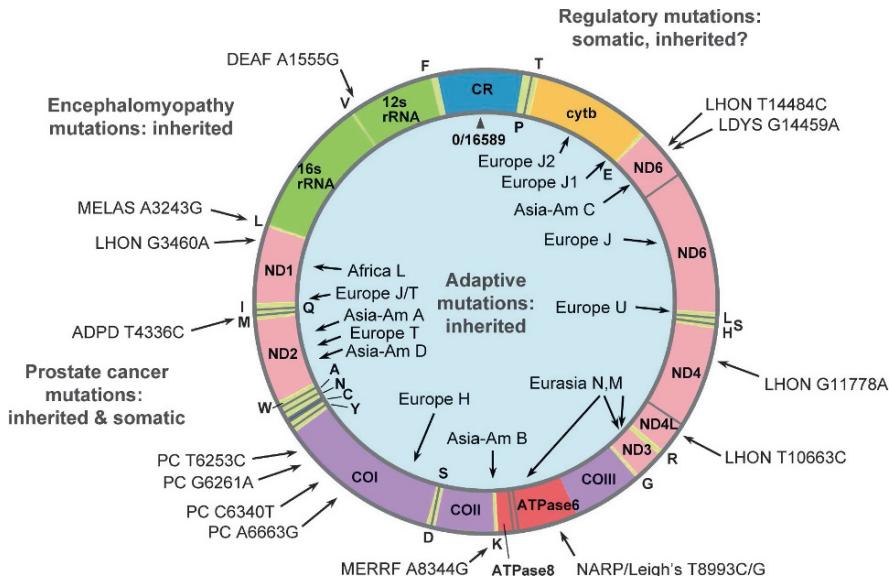


Fig. 2.14 Schematic representation of the circular mtDNA, its genes, its clinical relevant mutations, and certain polymorphic markers. Letters within the ring depict the genes encoded. Letters on the outside indicate amino

acids of the tRNA genes. *CR*, the control of replication region that contains promoters for the heavy and light strands. *Arrows outside* show the location of pathogenic mutations. (From [142])

Genome Variation

SNP

Single Nucleotide Polymorphism

Allele 1

Allele 2

Fig. 2.15 Schematic representation of a single nucleotide polymorphism. Allele 1 has a C in the sequence, while allele 2 contains a T in the same position.

gene [74] (at the time, these DNA polymorphisms were detected by restriction endonuclease digestion of DNAs and were called RFLPs, restriction fragment length polymorphisms). These polymorphic sites have two alternative alleles. In the example shown in Fig. 2.15, the depicted SNP has two alleles in the population: the blue C allele and the red T allele. The frequency of each allele could vary in different populations.

There is on average one SNP in ~1,000 nucleotides between two randomly chosen chromosomes in the population. Many of these SNPs are quite

common. A common SNP is that in which the minor allele frequency (MAF) is more than 5%. On average two haploid genomes differ in ~3,000,000 SNPs. In addition, there is a large number of rare (MAF < 1%) or near-rare (MAF between 1% and 5%) SNP variants that could be identified by the genome sequencing of various individuals. The majority of heterozygous SNPs in the DNA of a given individual are relatively common in the population; on the other hand, most of the SNPs discovered in a population are more likely to be rare. The NCBI SNP database contains 25 million common and rare SNPs (<http://www.ncbi.nlm.nih.gov/>

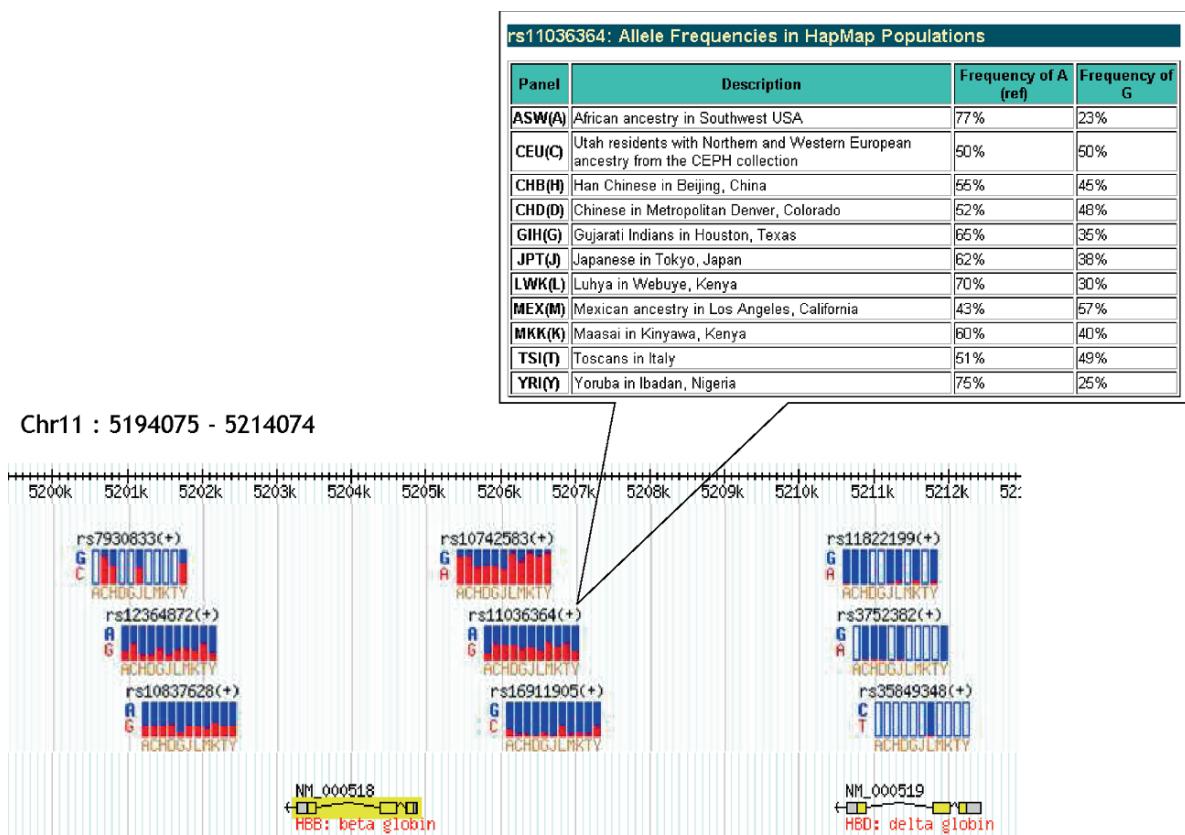


Fig. 2.16 The genomic region of Chr11: 5,194,075–5,214,074 is shown. For each of the nine SNPs shown in the bottom, the frequency of the two alternative alleles is shown in different populations. For example, for SNP rs11036364 that maps between the

HBB and HBD genes, the allele frequencies are shown in the callout. The four original populations of the HapMap project were EUR, YRI, JPT, and CHB, while the other populations were added in a later stage. Modified from <http://www.hapmap.org/>

SNP/snp_summary.cgi; version 130; July 2009; Fig. 2.16). Of those, ~301,000 are in the protein-coding regions of genes, and ~188,000 result in amino acid substitutions (nonsynonymous substitutions). An international project known as HapMap (<http://www.hapmap.org/>) [6, 34, 50] has completed the genotyping of ~4,000,000 common SNPs in individuals of different geo-ethnic origins (4,030,774 SNPs in 140 Europeans; 3,984,356 in 60 Yoruba Africans; 4,052,423 in 45 Japanese and 45 Chinese; <http://www.hapmap.org/downloads/index.html.en>). Additional samples from further populations have been added recently.

The information content of SNPs (and polymorphic variation in general) is usually measured by the number of heterozygotes in the population (homozygotes are individuals that contain the same variant in both alleles; heterozygotes are individuals that contain two different variants in their alleles). The number of heterozygotes is a function of MAF based on Hardy-

Weinberg principles (see Chap. 10). The pattern of DNA polymorphisms in a single chromosome is called haplotype (a contraction of “haploid genotype”; allelic composition of an individual chromosome). In the example shown in Fig. 2.17 the haplotype of polymorphic sites for the paternal (blue) chromosome is CGAACATC while for the maternally inherited red chromosome it is GACGAT.

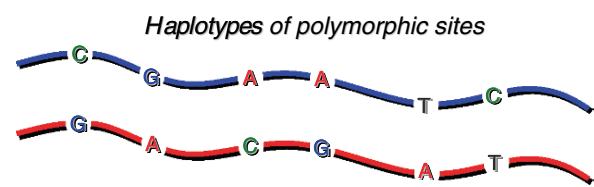


Fig. 2.17 Schematic representation of haplotype of polymorphic variants in a segment of the genome. The parental origin is shown as the *blue* (paternally-inherited) and *red* (maternally-inherited) lines. SNPs are shown as *letters interrupting the lines*. The haplotype is defined as the combination of SNP alleles per haploid genome

2

2.2.2 *Short Sequence Repeats*

Short sequence repeats (SSRs) are polymorphic variations due to a different number of short sequence repeat units, first described by Wyman and White [150] (then called VNTRs, variable number of tandem repeats[99]), and further elucidated by Jeffreys [72]. Most common are the dinucleotide repeats (described after the introduction of polymerase chain reaction amplification), but SSRs could be tri-, tetra-, or penta- repeats (often called microsatellites where the repeat unit $n=1-15$ nucleotides). SSRs with longer repeat units ($n=15-500$ nucleotides) are often termed minisatellites. These sequences comprise ~3% of the genome and there is ~1 SSR per 5 kb [84]. The most frequent dinucleotide SSR is the (GT)_n with an occurrence in the genome of ~28 times per megabase, followed by the (AT)_n SSR with ~19 times per megabase. The most common trinucleotide SSR is the (TAA)_n that occurs approximately four times per megabase. The major advantage of SSRs (or microsatellites) is that there are more than two alleles per polymorphic site, and a large fraction of the human population is heterozygous for each SSR. Therefore, SSRs are extremely useful in linkage mapping and subsequent positional cloning for monogenic disorders [12, 17, 33] and other marking studies.

of the genome including the development of genomic linkage maps [43, 144]. In addition, SSRs are extensively used in forensic studies [15]. Figure 2.18 shows an example of an SSR with three alleles in the population.

2.2.3 Insertion/Deletion Polymorphisms (Indels)

This variation is due to the presence or absence of certain sequences. These sequences could be a few nucleotides, but they could also be transposons or interspersed repeats such as LINE or SINE elements [18, 112, 149]; alternatively, they could be pseudogenes [8] or other elements. Note that this category of variants is not completely separate from the next one; the arbitrary distinction is just the size of the variation in terms of base pairs. There are usually biallelic polymorphisms, which are not as common as SNPs but are useful for evolutionary studies and for the understanding of the dynamic structure of the human genome. In the example shown in Fig. 2.19, the blue sequence was inserted in the DNA and created a variant with two alleles: the blue allele 1 with insertion and the black allele 2 without.

Genome Variation

SSR

Short Sequence Repeats

Allele 1

Allele 2

Allele 3

Fig. 2.18 An example of a dinucleotide SSR with three alleles in the population: the *blue* allele with $(CA)_{13}$ repeats, the *red* allele with $(CA)_{16}$, and the *green* allele with $(CA)_7$.

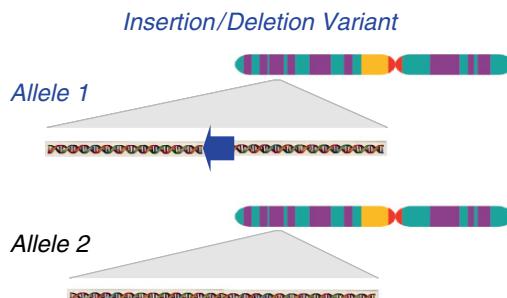


Fig. 2.19 Schematic representation of a polymorphic locus due to insertion/deletion of a genomic element, shown as a *blue arrow*

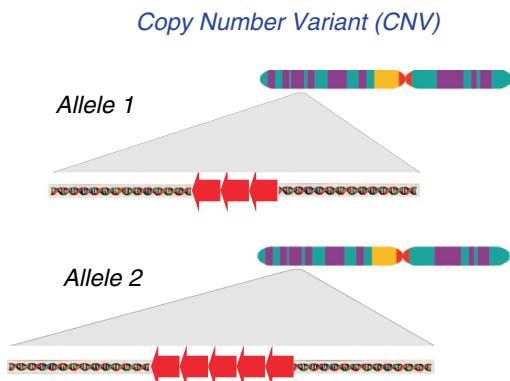


Fig. 2.20 Schematic representation of copy number variation in the human genome. For explanation, see text. Allele 1 in the population contains three copies of a sequence (*red arrowheads*), while allele 2 contains five copies

2.2.4 Copy Number Variants

Copy number variant (CNV) refers to large-scale structural variation of our genome in which there are large tandem repeats of 50 kb to 5 Mb long that are present in a variable number of copies. This type of polymorphic variant includes large-scale duplications and deletions [123] (see also Chap. 3, Sect. 3.4.4). These have been known since studies of the α -globin genes in humans [54]. In the Fig. 2.20 example, allele 1 contains three copies and allele 2 five copies of a large repeat. The phenotypic consequences of some of these variants that may contain entire genes is unknown. A CNV map of the human genome in 270 individuals has revealed a total of 1,440 such CNV regions which cover some 360 Mb (~12% of the genome [79, 108]). More recent estimates using more accurate methods for precise mapping of the size of CNVs suggest that ~6% of the genome contains CNVs. A list of these variants can be found at <http://projects.tcag.ca/variation/>. The extent of CNV in the human genome is certainly underestimated since there are numerous additional CNVs of less than 50 kb. The current methodology for the detection of CNVs is using comparative genomic hybridization (CGH) on DNA microarrays [25]. A further improvement of this method will allow us to detect small CNVs. The most detailed currently available CNV map of the human genome was recently established by the Genome Structural Variation Consortium. This consortium conducted a CNV project to identify common CNVs greater than 500 bp in size in 20 female CEU (European ancestry) and 20 female YRI (African ancestry) samples of the HapMap project. By employing CGH arrays that tile across the

assayable portion of the genome with ~42 million probes from the company NimbleGen, this consortium could map 8,599 copy number variant events. Parts of these data have been provisionally released to the scientific community and can be viewed at <http://www.sanger.ac.uk/humgen/cnv/42mio/>.

2.2.5 Inversions

Large DNA segments could have different orientation in the genomes of different individuals. These inversion polymorphisms (Fig. 2.21) predispose for additional genomic alterations [9]. An example of a common inversion polymorphism involves a 900-kb segment of chromosome 17q21.31, which is present in 20% of European alleles but it is almost absent or very rare in other populations [129]. These variants are difficult to identify and most of them have been detected by sequencing the ends of specific DNA fragments and comparing them with the reference sequence [79, 134].

2.2.6 Mixed Polymorphisms

There are combinations of repeat size variants and single nucleotide variants. Figure 2.22 depicts such an example; the repeat units of an SSR contain a SNP and, thus, even alleles with the same repeat number

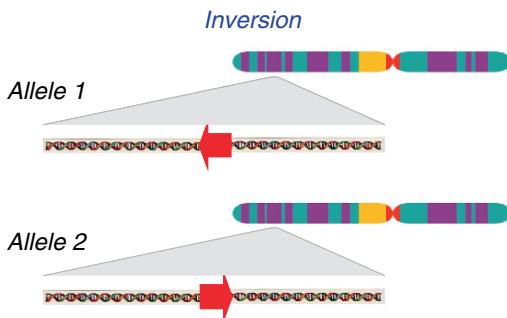


Fig. 2.21 Schematic representation of a polymorphic inversion shown as a *red arrowhead*

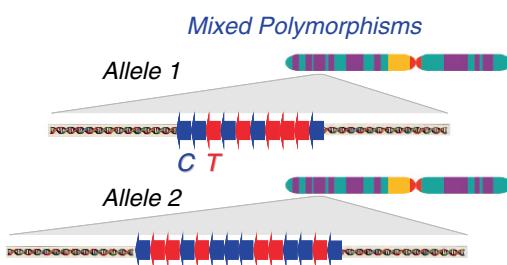


Fig. 2.22 Schematic representation of a highly polymorphic region of the genome with a mixed polymorphism that includes SNPs in the copies of CNVs or SSRs. The copies of the repeat are shown as *arrowheads*; the blue/red color of the repeats designates the SNP in them (blue for C and red for T)

could be distinguished based on their exact DNA sequence [71]. These highly polymorphic systems could serve as “recognition barcodes” in humans.

2.2.7 Genome Variation as a Laboratory Tool to Understand the Genome

DNA variants, besides their functional importance in health and disease, are very useful in human genetics research because they serve as genomic markers for a variety of studies. Some of the uses of DNA variants are to:

1. Create linkage (genetic) maps of human chromosomes [1, 148]. This has allowed the initial mapping of the human genome and it was a prerequisite for the sequence assembly.
2. Map the genomic location of monogenic phenotypes to human chromosomes by linkage analysis

[58, 81]. A large number of such phenotypes have been mapped to small genomic intervals because of the genotyping of members of affected families. Positional cloning of pathogenic mutations was subsequently possible.

3. Map the genomic location of polygenic phenotypes to human chromosomes by genomewide linkage and association studies [4, 20, 119].
4. Allow fetal diagnosis and carrier testing by linkage analysis of the cosegregation of a polymorphic marker and the phenotype of interest [10, 21].
5. Perform paternity and forensic studies [52]. A whole field was developed mainly with the use microsatellite SSR variants [49, 51].
6. Study genome evolution and origin of pathogenic mutations [115, 116].
7. Study the recombination rate and properties of the human genome [28, 93].
8. Study the instability of the genome in tumor tissues [5].
9. Identify loss-of-heterozygosity in human tumors [27, 47].
10. Study uniparental disomy and thus help with understanding genomic imprinting [100, 128].
11. Study parental and meiotic origin, and decipher the mechanisms of nondisjunction [11, 13, 14].
12. Study population history and substructure [110, 132].

The chapters that follow include further discussions on different aspects (including evolution, phenotypic consequences, and disease susceptibility) related to the most precious human genome variability.

Acknowledgments I thank the members of the laboratory, past and present, for discussions, ideas, debates, and data. I also thank my mentors and my students for the learning process.

References

1. No authors listed (1992) A comprehensive genetic linkage map of the human genome. NIH/CEPH Collaborative Mapping Group. *Science* 258:67–86
2. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:636–640
3. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945
4. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven

- common diseases and 3,000 shared controls. *Nature* 447: 661–678
5. Aaltonen LA, Peltomäki P, Leach FS, Sistonen P, Pylkkänen L et al (1993) Clues to pathogenesis of familial colorectal cancer. *Science* 260:812–816
 6. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
 7. Amaral PP, Dinger ME, Mercer TR, Mattick JS (2008) The eukaryotic genome as an RNA machine. *Science* 319:1787–1789
 8. Anagnou NP, O'Brien SJ, Shimada T, Nash WG, Chen MJ, Nienhuis AW (1984) Chromosomal organization of the human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. *Proc Natl Acad Sci USA* 81:5170–5174
 9. Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P et al (2009) Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet* 18:2555–2566
 10. Antonarakis SE (1989) Diagnosis of genetic disorders at the DNA level. *N Engl J Med* 320:153–163
 11. Antonarakis SE (1991) Parental origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. Down Syndrome Collaborative Group. *N Engl J Med* 324:872–876
 12. Antonarakis SE (1994) Genome linkage scanning: systematic or intelligent? *Nat Genet* 8:211–212
 13. Antonarakis SE, Avramopoulos D, Blouin JL, Talbot CC Jr, Schinzel AA (1993) Mitotic errors in somatic cells cause trisomy 21 in about 4.5% of cases and are not associated with advanced maternal age. *Nat Genet* 3:146–150
 14. Antonarakis SE, Petersen MB, McInnis MG, Adelsberger PA, Schinzel AA et al (1992) The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms. *Am J Hum Genet* 50:544–550
 15. Armour JA, Jeffreys AJ (1992) Biology and applications of human minisatellite loci. *Curr Opin Genet Dev* 2:850–856
 16. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV et al (2002) Recent segmental duplications in the human genome. *Science* 297:1003–1007
 17. Ballabio A (1993) The rise and fall of positional cloning. *Nat Genet* 3:277–279
 18. Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370–379
 19. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
 20. Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS et al (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 20:70–73
 21. Boehm CD, Antonarakis SE, Phillips JA 3rd, Stetten G, Kazazian HH Jr (1983) Prenatal diagnosis using DNA polymorphisms. Report on 95 pregnancies at risk for sickle-cell disease or beta-thalassemia. *N Engl J Med* 308:1054–1058
 22. Brady T, Lee YN, Ronen K, Malani N, Berry CC et al (2009) Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev* 23:633–642
 23. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y et al (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527–542
 24. Brown WR, MacKinnon PJ, Villasante A, Spurr N, Buckle VJ, Dobson MJ (1990) Structure and polymorphism of human telomere-associated DNA. *Cell* 63:119–132
 25. Carter NP (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 39:S16–S21
 26. Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet* 11:1527–1538
 27. Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R et al (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305:779–784
 28. Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* 36:1239–1258
 29. Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q et al (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Genet* 17:154–163
 30. Choo K, Vissel B, Nagy A, Earle E, Kalitsis P (1991) A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res* 19:1179–1182
 31. Choo KH, Earle E, McQuillan C (1990) A homologous subfamily of satellite III DNA on human chromosomes 14 and 22. *Nucleic Acids Res* 18:5641–5648
 32. Cohen D, Chumakov I, Weissenbach J (1993) A first-generation physical map of the human genome. *Nature* 366:698–701
 33. Collins FS (1990) Identifying human disease genes by positional cloning. *Harvey Lect* 86:149–164
 34. International HapMap Consortium (2003) The International HapMap Project. *Nature* 426:789–796
 35. Craig-Holmes AP, Shaw MW (1971) Polymorphism of human constitutive heterochromatin. *Science* 174: 702–704
 36. Cserpan I, Katona R, Praznovszky T, Novak E, Rozsavolgyi M et al (2002) The chAB4 and NF1-related long-range multisequence DNA families are contiguous in the centromeric heterochromatin of several human chromosomes. *Nucleic Acids Res* 30:2899–2905
 37. Deininger PL, Batzer MA (1999) Alu repeats and human disease. *Mol Genet Metab* 67:183–193
 38. Deloukas P, Earthrow ME, Graham DV, Rubenfield M, French L et al (2004) The DNA sequence and comparative analysis of human chromosome 10. *Nature* 429:375–381
 39. Deloukas P, Matthews LH, Ashurst J, Burton J, Gilbert JG et al (2001) The DNA sequence and comparative analysis of human chromosome 20. *Nature* 414:865–871
 40. Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R et al (2007) Prominent use of distal 5' transcription start

- sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res* 17:746–759
41. Dermitzakis ET, Reymond A, Antonarakis SE (2005) Conserved non-genic sequences—an unexpected feature of mammalian genomes. *Nat Rev Genet* 6:151–157
 42. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C et al (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420:578–582
 43. Dib C, Faure S, Fizames C, Samson D, Drouot N et al (1996) A comprehensive genetic map of the human genome based on 5, 264 microsatellites. *Nature* 380:152–154
 44. Djebali S, Kapranov P, Foissac S, Lagarde J, Reymond A et al (2008) Efficient targeted transcript discovery via array-based normalization of RACE libraries. *Nat Methods* 5:629–635
 45. Dunham A, Matthews LH, Burton J, Ashurst JL, Howe KL et al (2004) The DNA sequence and analysis of human chromosome 13. *Nature* 428:522–528
 46. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR et al (1999) The DNA sequence of human chromosome 22. *Nature* 402:489–495
 47. Fearon ER, Vogelstein B, Feinberg AP (1984) Somatic deletion and duplication of genes on chromosome 11 in Wilms' tumours. *Nature* 309:176–178
 48. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ et al (2008) The Pfam protein families database. *Nucleic Acids Res* 36:D281–D288
 49. Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P et al (1998) Jefferson fathered slave's last child. *Nature* 396:27–28
 50. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
 51. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N et al (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130–135
 52. Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. *Nature* 318:577–579
 53. Gonzalez IL, Sylvester JE (2001) Human rDNA: evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics* 73: 255–263
 54. Goossens M, Dozy AM, Embury SH, Zachariades Z, Hadjiminas MG et al (1980) Triplicated alpha-globin loci in humans. *Proc Natl Acad Sci USA* 77:518–521
 55. Gregory SG, Barlow KF, McLay KE, Kaul R, Swarbreck D et al (2006) The DNA sequence and biological annotation of human chromosome 1. *Nature* 441:315–321
 56. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Res* 31:439–441
 57. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J et al (2004) The DNA sequence and biology of human chromosome 19. *Nature* 428:529–535
 58. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:234–236
 59. Guttman M, Amit I, Garber M, French C, Lin MF et al (2009) Chromatin signature reveals over a thousand highly conserved large noncoding RNAs in mammals. *Nature* 458:223–227
 60. Gyapay G, Morissette J, Vignal A, Dib C, Fizames C et al (1994) The 1993–94 Genethon human genetic linkage map. *Nat Genet* 7:246–339
 61. Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol* 7([Suppl 1]:S4):1–9
 62. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T et al (2000) The DNA sequence of human chromosome 21. *Nature* 405:311–319
 63. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C et al (2003) The DNA sequence and analysis of human chromosome 14. *Nature* 421:601–607
 64. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ et al (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3:1287–1295
 65. Higgs DR, Weatherall DJ (2009) The alpha thalassaemias. *Cell Mol Life Sci* 66:1154–1162
 66. Hillier LW, Fulton RS, Fulton LA, Graves TA, Pepin KH et al (2003) The DNA sequence of human chromosome 7. *Nature* 424:157–164
 67. Hillier LW, Graves TA, Fulton RS, Fulton LA, Pepin KH et al (2005) Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature* 434:724–731
 68. Hollis M, Hindley J (1988) Satellite II DNA of human lymphocytes: tandem repeats of a simple sequence element. *Nucleic Acids Res* 16:363
 69. Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB et al (1995) An STS-based map of the human genome. *Science* 270:1945–1954
 70. Humphray SJ, Oliver K, Hunt AR, Plumb RW, Loveland JE et al (2004) DNA sequence and analysis of human chromosome 9. *Nature* 429:369–374
 71. Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354:204–209
 72. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67–73
 73. Kalitsis P, Earle E, Vissel B, Shaffer LG, Choo KH (1993) A chromosome 13-specific human satellite I DNA subfamily with minor presence on chromosome 21: further studies on Robertsonian translocations. *Genomics* 16:104–112
 74. Kan YW, Dozy AM (1978) Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 75:5631–5635
 75. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M et al (2005) Antisense transcription in the mammalian transcriptome. *Science* 309:1564–1566
 76. Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. *Science* 303:1626–1632
 77. Kazazian HH Jr, Goodier JL (2002) LINE drive. retrotransposition and genome instability. *Cell* 110:277–280
 78. Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332:164–166