ESSENTIALS

ESSENTIAL MEDICAL GENETICS

EDWARD S. TOBIAS | MICHAEL CONNOR
MALCOLM FERGUSON-SMITH

6TH EDITION

with Wiley DESKTOP EDITION

with website

WILEY-BLACKWELL

Essential Medical Genetics

Edward S. Tobias
BSc MBChB PhD FRCP
Senior Clinical Lecturer in Medical Genetics
University of Glasgow
and Honorary Consultant in Medical Genetics
West of Scotland Regional Genetics Service
Institute of Medical Genetics
Glasgow

Michael Connor
MD, DSc, FRCP
Professor of Medical Genetics
University of Glasgow
and Director of the West of Scotland
Regional Genetics Service
Institute of Medical Genetics
Glasgow

Malcolm Ferguson-Smith
MBChB, FRCPath, FRCP, FRSE, FRS
Emeritus Professor of Pathology
University of Cambridge
and formerly Director of the East Anglia Regional Genetics Service
Addenbrookes Hospital
Cambridge

Sixth edition

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Blackwell Publishing was acquired by John Wiley & Sons in February 2001. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical and Medical business to form Wiley-Blackwell.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

Library of Congress Cataloguing-in-Publication Data
Tobias, Edward.
Essential medical genetics / Edward Tobias, Michael Connor, Malcolm Ferguson-Smith. – 6th ed.
p. ; cm.
Rev. ed. of : Essential medical genetics / Michael Connor, Malcolm Ferguson-Smith. 5th ed. 1997.
Includes bibliographical references and index.
ISBN 978-1-4051-6974-5 (pbk. : alk. paper) 1. Medical genetics. I. Connor, J. M. (James Michael), 1951- II. Ferguson-Smith, M. A. (Malcolm Andrew) III. Connor, J. M. (James Michael), 1951- Essential medical genetics. IV. Title.
[DNLM: 1. Genetics, Medical. QZ 50]
RB155.C66 2011
616′.042–dc22
2010031705

ISBN: 9781405169745

A catalogue record for this book is available from the British Library.
Set in 10/12 pt and Adobe Garamond by Toppan Best-set Premedia Limited

Printed in Singapore

1 2011

Contents

Companion website

This book has a companion website:

www.wiley.com/go/tobias

with:

- Regularly updated links to genetic databases and analysis tools
- Updated information relating to the book's content
- Additional self-assessment questions and answers
- Figures from the book in Powerpoint format

Preface

This book has been written for those to whom an understanding of modern medical genetics is important in their current or future practice as clinicians, scientists, counsellors and teachers. It is based on the authors' personal experience in both clinical and laboratory aspects of busy regional genetics services over a period of many years. This period has seen the emergence of modern cytogenetics and molecular genetics alongside the development of medical genetics from a purely academic discipline into a clinical specialty of relevance to every branch of medicine. As in our undergraduate and postgraduate education programmes, we emphasize the central role of the chromosome and the human genome in understanding the molecular mechanisms involved in the pathogenesis of genetic disease. Within the term genetic disease, we include not only the classic Mendelian and chromosomal disorders but also the commoner disorders of adulthood with a genetic predisposition and somatic cell genetic disorders, such as cancer.

For this sixth edition, the text has been extensively updated throughout. The structure of the book has, where appropriate, been reorganised, in order to provide a clear description of the essential principles of the scientific basis and clinical application of modern medical genetics. Where appropriate, we have included descriptions of genetic conditions that have been carefully selected as examples of the important principles being described. Since the last edition of this book, several important and exciting new advances have been made in the field of medical genetics, and we have incorporated information about them into the book. Such advances include, for example, the completion of the sequencing of the human genome (with the generation of huge quantities of publicly accessible data), the identification of new classes of RNA molecules, the development of a number of invaluable new molecular genetic and cytogenetic laboratory techniques, the further development of preimplantation genetic diagnosis, and improved methods for antenatal and neonatal screening.

A very significant additional advance has been the development and enormous expansion of many invaluable online clinical and molecular genetic databases. These databases have greatly facilitated the medical genetics work of most clinicians and scientists. The optimal use of several important databases is, however, in many cases far from straightforward. Consequently, retrieving specific information or data from them can take a great deal of time and effort for users who do not access them frequently. The final chapter of this book is therefore devoted to providing guidance on the most efficient use of these databases, together with clear illustrated advice explaining how to find different types of information via the internet as quickly as possible. It is hoped that this guidance, which to our knowledge is currently unavailable elsewhere, will make this process much more straightforward for the reader.

We have also provided an accompanying website (accessed via www.wiley.com/go/tobias) that we will regularly update in order to provide the reader with a way of easily accessing the very latest clinical and molecular genetic information relating to the thousands of genetic conditions, in addition to patient information and support organizations, the identified genes, and gene-testing laboratories worldwide. The links are grouped on the website in a very similar manner to the way in which they are categorised within the final chapter of this book, in order to make it as easy as possible for readers to find relevant information quickly.

Although we have made every effort to ensure that the information contained within this book is accurate at the time of going to press, we look to the continued generosity of our readers in helping to correct any misconceptions or omissions. We would be happy to receive any comments, or recommendations for improvements, at essentialmedgen@gmail.com.

The role of genetic counselling, prenatal diagnosis, carrier detection and other forms of genetic screening in the prevention of genetic disease is now well established and this is reflected in the increasing provision of genetic services throughout the world. It is hoped that our book will be useful to those in training for this important task.

E.S.T, J.M.C. and M.A.F-S.

Acknowledgements

We wish to thank all of the many people who have influenced the production of this book. These include, particularly, our colleagues and students at the Institute of Medical Genetics in Glasgow and at the Cambridge University Centre for Medical Genetics. We also wish to acknowledge the invaluable contributions made by Professor Carolyn Brown (Life Sciences Centre, Vancouver, Canada), Professor Mark Jobling (University of Leicester, UK) and Dr Zofia Esden-Tempska (Medical University of Gdansk, Poland).

The authors are indebted to the editorial and production team at Wiley-Blackwell, including Martin Sugden, Hayley Salter, Laura Murphy, Elizabeth Bishop and Elizabeth Johnston, in addition to the freelance project manager, Anne Bassett.

E.S.T. would like to express his enormous gratitude to his wife, family and friends for their continuous support and understanding while he worked on the manuscript.

We are most obliged to Professor Tom Ellenberger (Washington University School of Medicine, St Louis, Missouri, USA) for his generous permission to use the front cover image, which depicts the interaction between human DNA ligase I and DNA.

We are very grateful to the patients and their families, and to the following, for permission to reproduce these figures:

Fig. 4.2: Alexander Fletcher;
Fig. 4.4: Joăn Lavinha;
Figs. 4.5, 4.8 and 4.9: Gillian Stevens;
Figs. 4.6 and 4.7: Maria Jackson and Leah Marks;
Fig. 4.10: Jim Kelly;
Figs. 4.11 and 7.22: Jayne Duncan;
Fig. 4.12, 13.5 and 16.2: Alexander Cooke;
Fig. 4.14: Julia El-Sayed Moustafa;
Fig. 4.15: Paul Debenham (Cellmark Diagnostics);
Figs 5.2–5.5, 6.17b, 7.6, 7.8, and 9.2: Elizabeth Boyd;
Fig. 5.8: Nigel Carter;
Fig. 5.13: The Editor, *Birth Defects Original Article Series*;
Fig. 5.14: The Editior, *Annales de Genetique*;
Fig. 5.15: Peter Pearson;
Figs 6.2, 6.3, 6.9 and 7.9: The Editor, *Excerpta Medica*;
Figs 6.8 and 7.4(d): Anne Chandley;
Fig. 6.16: John Tolmie;
Fig. 6.18c: Lionel Willatt;
Figs 7.4(b) and 7.4(c): The Editor, *Journal of Medical Genetics*;
Fig. 7.15: Maj Hulten and N. Saadallah;
Figs 7.16 and 7.17: The Editor, *Cytogenetics and Cell Genetics*;
Fig. 8.6: Brenda Gibson;
Figs 8.12 and 18.4: Douglas Wilcox;
Figs. 7.2, 7.21 and 7.32: Catherine McConnell;
Fig. 7.19: Aspasia Divane;
Fig. 7.20: Diana Johnson and BMJ Publishing Group Ltd.;
Fig. 7.30: Evelyn Schröck and Thomas Ried;
Figs. 11.4 and 11.5: Gary Stix and Nature Publishing Group;
Fig. 12.4, 15.5 and 18.20: Margo Whiteford;
Figs. 12.8 and 7.23–26: Norma Morrison;
Figs. 13.7 and 13.8: Janet Stewart;
Fig. 13.10: Springer, Heidelberg;
Fig. 14.1 and 14.2: Inga Prokopenko and Elsevier;
Fig. 14.3: Bart Dermaut and Elsevier;
Fig. 15.7: Peter Cackett and Nature Publishing Group;
Fig. 16.5: Bernhard Horsthemke, Joseph Wagstaff and American Journal of Medical genetics;
Figs. 17.1–17.4: Jenny Crossley and David Aitken;
Fig. 17.5: Joan Mackenzie and Arlene Brown;
Fig. 18.16: WE Tidyman, KA Rauen and Cambridge Journals;
Fig. 18.22: Marie-France Portnoi and Elsevier; and
Figs. 19.45–19.48: Michael Baraitser.

We would also like to thank the curators of the following websites for permission to reproduce screenshots: National Center for Biotechnology Information (NCBI), Ensembl (Wellcome Trust Sanger Institute), GeneCards (Weizmann Institute of Science), University of California Santa Cruz (UCSC) Genome Browser, UK Genetic Testing Network (UKGTN), European Directory of DNA Diagnostic Laboratories (EDDNAL), Primer3Plus, RCSB Protein Data Bank (PDB) and The Phenomizer.

The authors and publisher have made every effort to seek the permission of all copyright holders for the reproduction of copyright material. If any have been overlooked inadvertently, the publisher will be pleased to make the necessary amendments at the earliest opportunity.

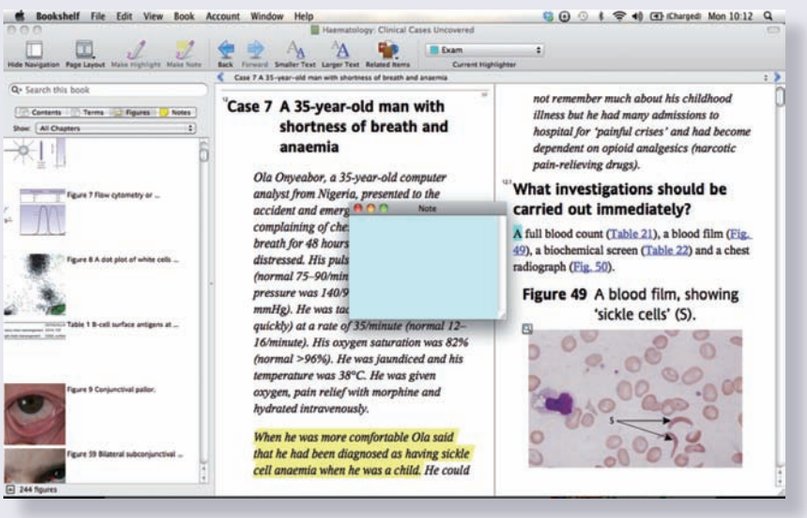
How to get the best out of your textbook

Welcome to the new edition of *Essential Medical Genetics*. Over the next two pages you will be shown how to make the most of the learning features included in the textbook.

An interactive textbook ▶

For the first time, your textbook gives you free access to a Wiley Desktop Edition – a digital, interactive version of this textbook. Your Wiley Desktop Edition allows you to:

Search: Save time by finding terms and topics instantly in your book, your notes, even your whole library (once you've downloaded more textbooks)
Note and Highlight: Colour code highlights and make digital notes right in the text so you can find them quickly and easily
Organize: Keep books, notes and class materials organized in folders inside the application
Share: Exchange notes and highlights with friends, classmates and study groups
Upgrade: Your textbook can be transferred when you need to change or upgrade computers
Link: Link directly from the page of your interactive textbook to all of the material contained on the companion website.



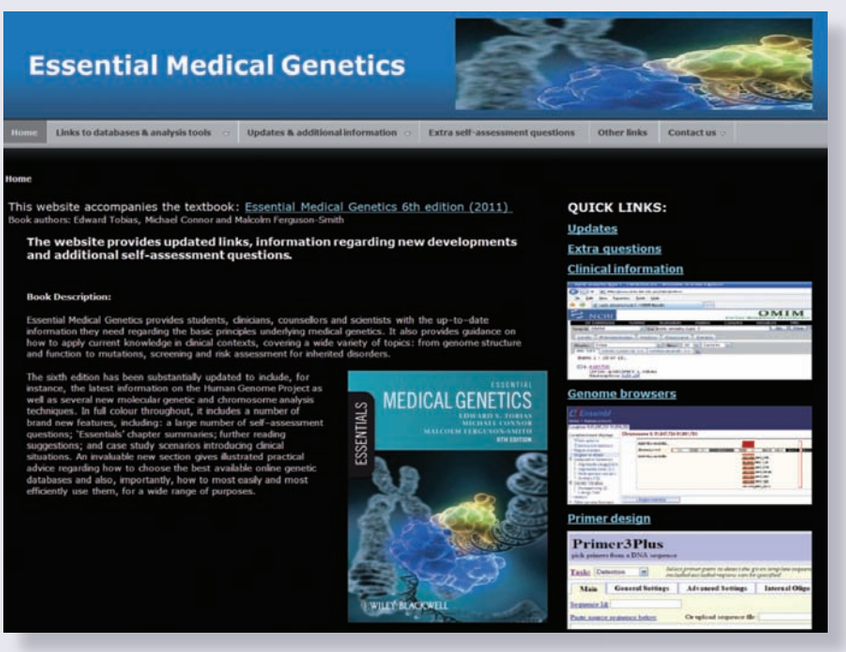
Simply find your unique Wiley Desktop Edition product code and carefully scratch away the top coating on the label on the front cover of this textbook and visit:
http://www.vitalsource.com/software/bookshelf/downloads/ to get started.

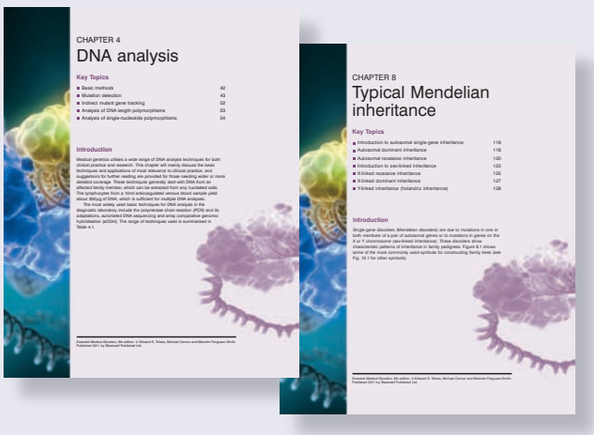
A companion website ▶

Your textbook is also accompanied by a FREE companion website that contains:

- Regularly updated links to genetic databases and analysis tools
- Updated information relating to the book's content
- Additional self-assessment questions and answers
- Figures from the book in Powerpoint format.

Log on to **www.wiley.com/go/tobias** to find out more.

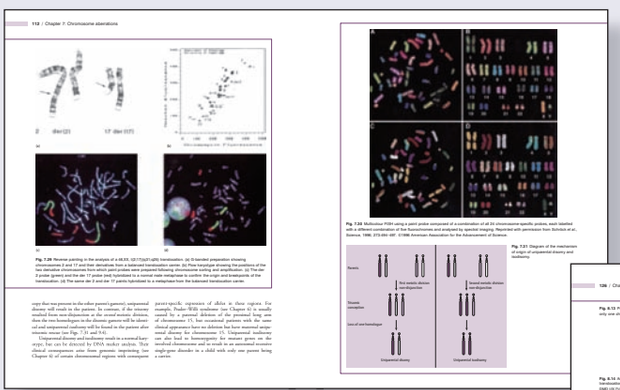




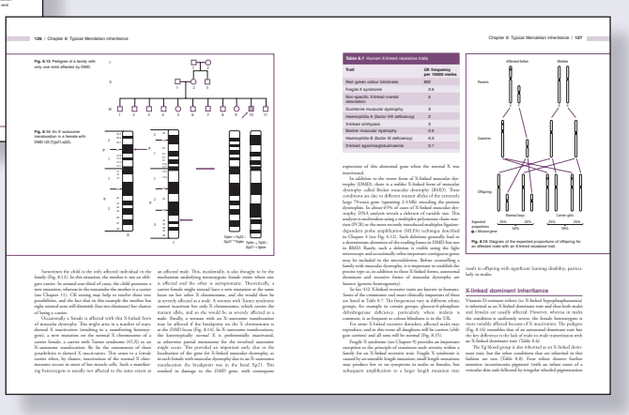
Features contained within your textbook

◀ Every chapter has its own chapter-opening page that offers a list of key topics contained within the chapter.

Throughout your textbook you will find this icon which points you to the online databases and resources found on the companion website. You can also access the website by clicking on this icon in your Desktop Edition.



◀ Your textbook is full of useful photographs, illustrations and tables. The Desktop Edition version of your textbook will allow you to copy and paste any photograph or illustration into assignments, presentations and your own notes.
▼

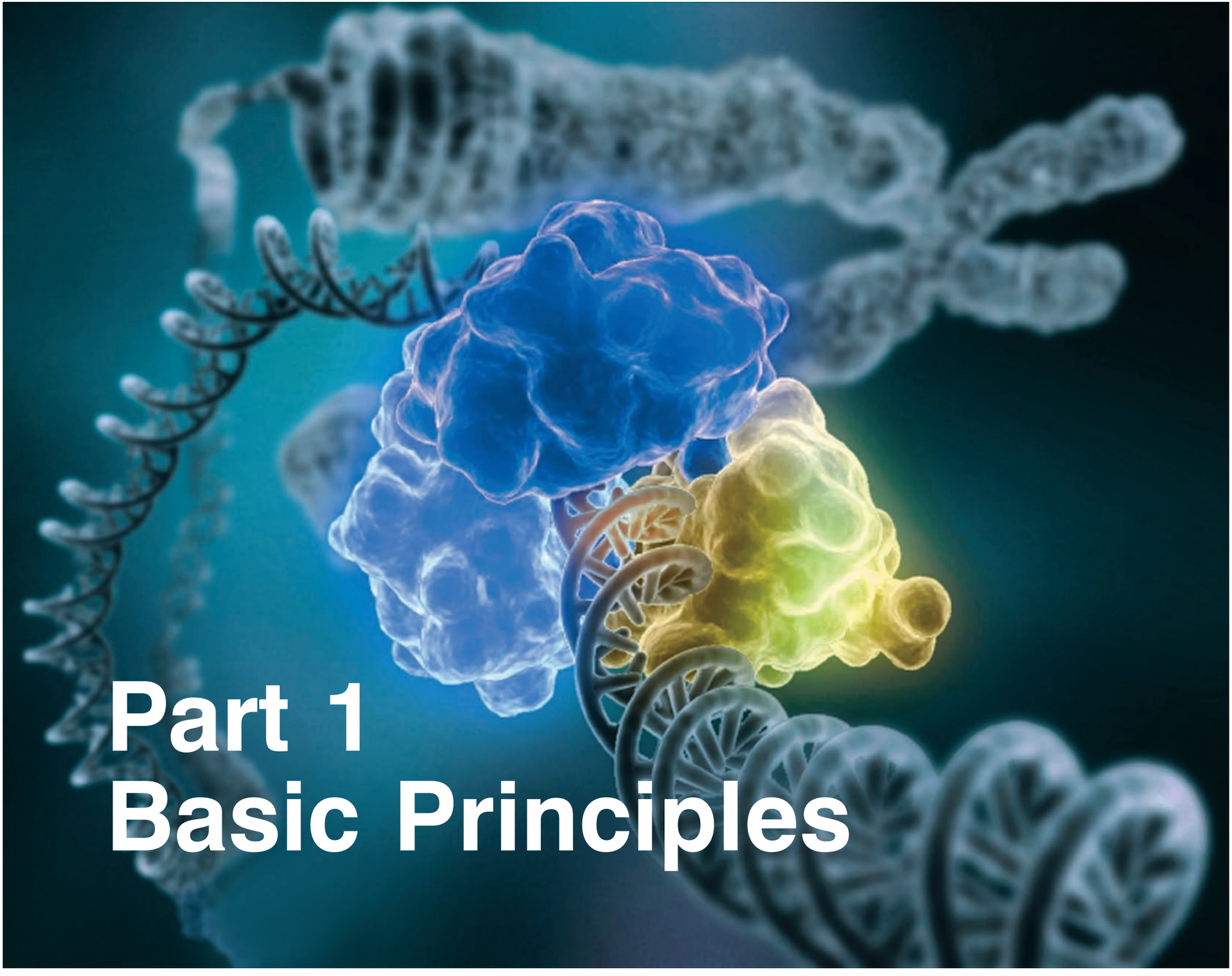


SUMMARY

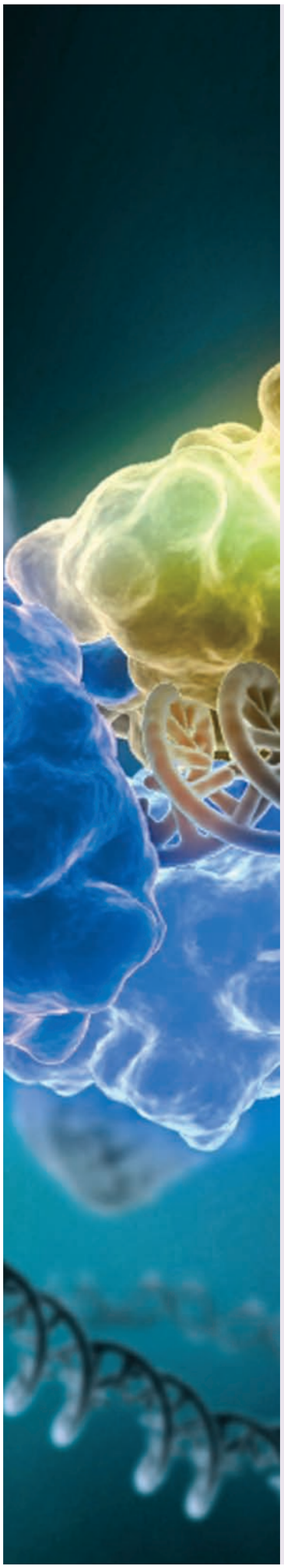
- Multifactorial inheritance implies a contribution of both genetic and environmental factors.
- Twin concordance and family correlation studies can provide support for the multifactorial inheritance of a trait. The observed frequencies in relatives provide the empiric risks upon which genetic counselling for multifactorial disorders is based.
- Multifactorial traits that are continuous (such as height) have a continuously graded distribution, while those that are discontinuous (i.e. with individuals being either affected or unaffected) are present only when a certain threshold of genetic factors is reached.
- For twins, placental membranes that are monochorionic indicate monozygosity, whereas dichorionic membranes represent either monozygous or dizygous twins. Zygosity is determined most reliably by DNA fingerprinting.
- Monozygotic twins are identical genetically (i.e. at the DNA level), whereas dizygotic twins exhibit the same degree of genetic similarity as siblings.
- Genome-wide analyses of the genetic determinants of multifactorial traits may now be undertaken by association studies of the frequencies of each of hundreds of thousands of SNPs in cases and controls.

▲ Every chapter ends with a summary which can be used for both study and revision purposes.

We hope you enjoy using your new textbook. Good luck with your studies!



Part 1
Basic Principles



CHAPTER 1

Medical genetics in perspective

Key Topics

Introduction

Medical genetics is the science of human biological variation as it relates to health and disease. Although people have long been aware that individuals differ, that children tend to resemble their parents and that certain diseases tend to run in families, the scientific basis for these observations was only discovered during the past 140 years. The clinical applications of this knowledge are even more recent, with most progress confined to the past 50 years (see Table 1.1). In particular, the rapid sequencing of the entire human genome, completed in 2003, has greatly accelerated the process of gene mapping for genetic conditions and a vast quantity of valuable and continuously updated information has become readily accessible via the internet (as described in detail in Part 3 and on this book's accompanying website at www.wiley.com/go/tobias).

Table 1.1 Some important landmarks in the development of medical genetics

Year	Landmark	Key figure(s)
1839	Cell theory	Schleiden and Schwann
1859	Theory of evolution	Darwin
1865	Particulate inheritance	Mendel
1882	Chromosomes observed	Flemming
1902	Biochemical variation	Garrod
1903	Chromosomes carry genes	Sutton, Boveri
1910	First US genetic clinic	Davenport
1911	First human gene assignment	Wilson
1944	Role of DNA	Avery
1953	DNA structure	Watson, Crick, Franklin and Wilkins
1956	Amino acid sequence of sickle haemoglobin (HbS)	Ingram
1956	46 chromosomes in humans	Tjio and Levan
1959	First human chromosomal abnormality	Lejeune
1960	Prenatal sexing	Riis and Fuchs
1960	Chromosome analysis on blood	Moorhead
1961	Biochemical screening	Guthrie
1961	X chromosome inactivation	Lyon
1961	Genetic code	Nirenberg
1964	Antenatal ultrasound	Donald
1966	First prenatal chromosomal analysis	Breg and Steel
1966	First print edition of Mendelian Inheritance in Man (MIM)	McKusick
1967	First autosomal assignment	Weiss and Green
1970	Prevention of Rhesus isoimmunisation	Clarke
1970	Chromosome banding	Caspersson and Zech
1975	DNA sequencing	Sanger, Maxam and Gilbert
1976	First DNA diagnosis	Kan
1977	First human gene cloned	Shine
1977	Somatostatin made by genetic engineering	Itakura
1979	*In vitro* fertilisation	Edwards and Steptoe
1979	Insulin produced by genetic engineering	Goeddel
1982	First genetic engineering product marketed (Humulin)	Many contributors
1985	DNA fingerprinting	Jeffreys
1986	Polymerase chain reaction (PCR)	Mullis
1987	Linkage map of human chromosomes developed	Many contributors
1987	Online Mendelian Inheritance in Man (OMIM) first available	McKusick
1990	First treatment by supplementation gene therapy	Rosenberg, Anderson, Blaese
1990	First version of London Dysmorphology Database	Baraitser and Winter
1990	First clinical use of preimplantation genetic diagnosis (PGD)	Handyside, Winston and others
1991	First version of London Neurogenetics Database	Baraitser and Winter
1993	First physical map of the human genome	Many contributors

Table 1.1 *continued*

Year	Landmark	Key figure(s)
2000	First draft of the human genome sequence	Many contributors
2003	Completion of human genome sequencing (99.999%)	HGSC and Celera
2006	Preimplantation genetic haplotyping (PGH) announced	Renwick, Abbs and others
2007	Human genome SNP map (3.1 million SNPs) reported	International HapMap Consortium
2007	Completion of DNA sequencing of personal genomes	Watson and Venter
2008	Launch of project to sequence the genomes of over 1000 individuals from 20 different populations worldwide	International 1000 Genomes Project
2010	Publication of catalogue of human genetic variation (believed to be 95% complete)	International 1000 Genomes Project

HGSC: Human Genome Sequencing Consortium; OMIM: Online Mendelian Inheritance in Man; SNP: single nucleotide polymorphism.

Scientific basis of medical genetics

Mendel's contribution

Prior to Mendel, parental characteristics were believed to blend in the offspring. While this was acceptable for continuous traits such as height or skin pigmentation, it was clearly difficult to account for the family patterns of discontinuous traits such as haemophilia or albinism. Mendel studied clearly defined pairs of contrasting characters in the offspring of the garden pea (*Pisum sativum*). These peas were, for example, either round or wrinkled and were either yellow or green. Pure-bred strains for each of these characteristics were available but when cross-bred (the first filial or F_1 progeny) were all round or yellow. If F_1 progeny were bred then each characteristic was re-observed in a ratio of approximately 3 round to 1 wrinkled or 3 yellow to 1 green (in the second filial or F_2 progeny). Mendel concluded that inheritance of these characteristics must be particulate with pairs of hereditary elements (now called genes). In these two examples, one characteristic (or trait) was dominant to the other (i.e. all the F_1 showed it). The fact that both characteristics were observed in the F_2 progeny entailed *segregation of each pair of genes with one member to one gamete and one to another gamete* (Mendel's first law).

Figures 1.1 and 1.2 illustrate these experiments with upper-case letters used for the dominant characteristic and lower-case letters used for the masked (or recessive) characteristic. If both members of the pair of genes are identical, this is termed homozygous (for the dominant or recessive trait), whereas a heterozygote has one gene of each type.

In his next series of experiments Mendel crossed pure-bred strains with two characteristics, e.g. pure-bred round/yellow with pure-bred wrinkled/green. The F_1 generation showed only the two dominant characteristics – in this case round/yellow. The F_2 showed four combinations: the original two, namely round/yellow and wrinkled/green, in a ratio of approximately 9:1 and two new combinations – wrinkled/yellow and round/green in a ratio of approximately 3:3 (Fig. 1.3).

In these experiments, there was thus no tendency for the genes arising from one parent to stay together in the offspring. In other words, *members of different gene pairs assort to gametes independently of one another* (Mendel's second law).

Although Mendel presented and published his work in 1865, after cultivating and studying around 28,000 pea plants, the significance of his discoveries was not realised until the early 1900s when three plant breeders, De Vries, Correns and Tschermak, confirmed his findings.

Chromosomal basis of inheritance

In 1839, Schleiden and Schwann established the concept of cells as the fundamental living units. Hereditary transmission through the sperm and egg was known by 1860, and in 1868, Haeckel, noting that the sperm was largely nuclear material, postulated that the nucleus was responsible for heredity. Flemming identified chromosomes within the nucleus in 1882, and in 1903 Sutton and Boveri independently realised that the behaviour of chromosomes during the production of gametes paralleled the behaviour of Mendel's hereditary elements. Thus, the chromosomes were discovered to carry the genes. However, at that time, although the chromosomes were known to consist of protein and nucleic acid, it was not clear which component was the hereditary material.

Chemical basis of inheritance

Pneumococci are of two genetically distinct strains: rough or non-encapsulated (non-virulent) and smooth or encapsulated (virulent). In 1928, Griffith added heat-killed smooth bacteria to live rough bacteria and found that some of the rough pneumococci were transformed to the smooth, virulent type. Avery, MacLeod and McCarty repeated this experiment in 1944 and showed that nucleic acid was the transforming agent. Thus, nucleic acid was shown to carry hereditary information. This stimulated intense interest in the composition of nucleic acids,

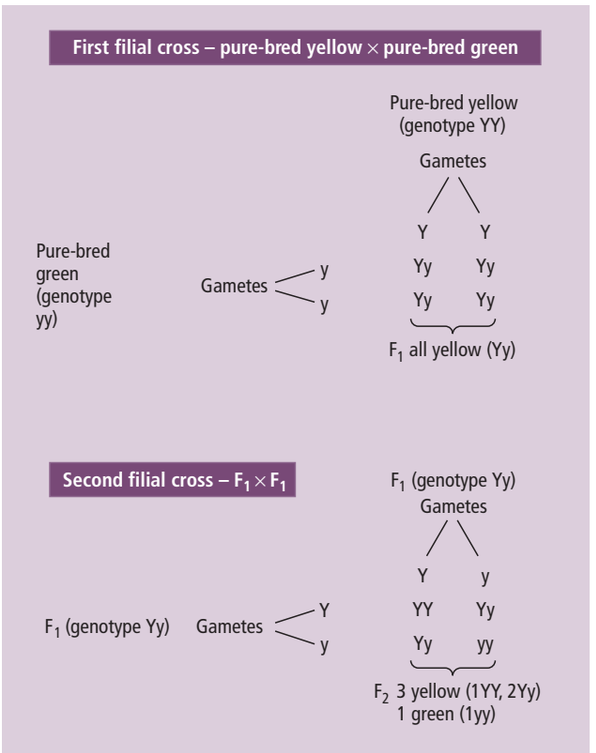


Fig. 1.1 Example of Mendel's breeding experiments for a single trait (yellow or green peas).

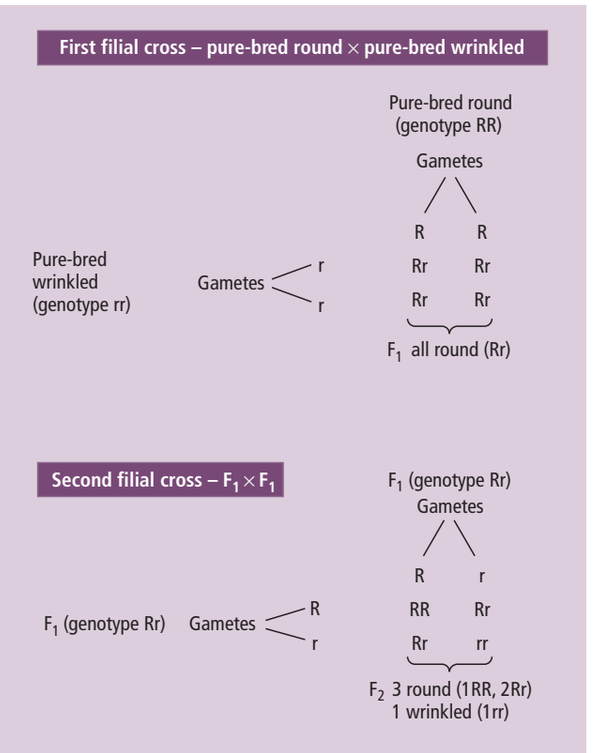
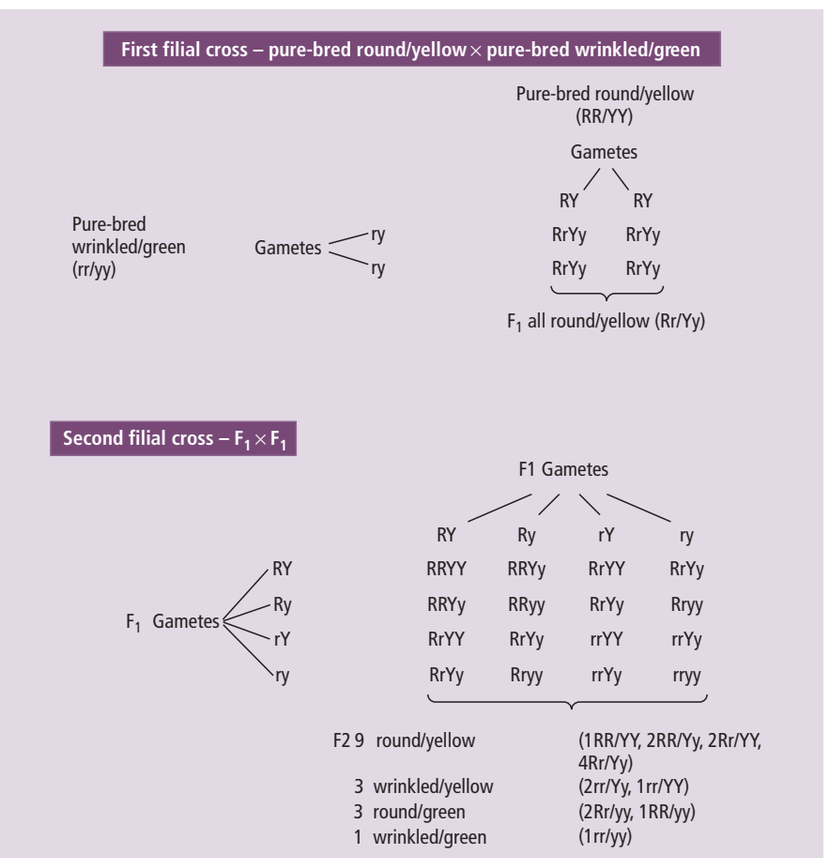


Fig. 1.2 Example of Mendel's breeding experiments for a single trait (round or wrinkled peas).

Fig. 1.3 Example of Mendel's breeding experiments for two traits (yellow or green and round or wrinkled peas).



which culminated in the discovery, by Watson, Crick, Franklin and Wilkins, of the double-helical structure for deoxyribonucleic acid (DNA) in 1953.

Chromosomal disorders

By 1890, it was known that one human chromosome (the X chromosome) did not always have a partner, and in 1905 Wilson and Stevens extended this observation by establishing the pattern of human sex chromosomes. At this time, it was believed that there were 47 chromosomes, including one X chromosome, in each male somatic cell and 48 chromosomes, including two X chromosomes, in each female cell. In 1923, the small Y chromosome was identified, and both sexes were thought to have 48 chromosomes. Tjio and Levan refuted this in 1956 when they showed the normal human chromosome number to be 46. In 1959, the first chromosomal disease in humans, trisomy 21, was discovered by Lejeune and colleagues, and by 1970, over 20 different human chromosomal disorders were known. The development of chromosomal banding in 1970 markedly increased the ability to resolve small chromosomal aberrations, and so by 1990 more than 600 different chromosome abnormalities had been described, in addition to many normal variants. This number has increased further with the development of improved techniques including various fluorescence *in situ* hybridisation (FISH) methods and comparative genomic hybridisation (CGH). In fact, the increased resolution of the more recently developed techniques such as array CGH (see Chapter 7), has led to greater difficulties in differentiating between the increasingly numerous normal and abnormal chromosomal variants. This, in turn, has necessitated the development of international databases of such submicroscopic variants such as DECIPHER (Fig. 1.4), based at the Sanger Institute (http://decipher.sanger.ac.uk/), and the Database of Genomic Variants at Toronto (http://projects.tcag.ca/variation).

Mitochondrial disorders

Mitochondria have their own chromosomes and these are passed on from a mother to all of her children but not from the father. These chromosomes are different in several respects from their nuclear counterparts. For instance, they contain only 37 genes, a high and variable number of DNA copies per cell, very little non-coding DNA and no introns (see Chapter 5). Mutations in genes on these mitochondrial chromosomes can cause disease and this was first shown in 1988 for a maternally inherited type of blindness (Leber optic neuropathy). Since then, it has been shown that many different mitochondrial mutations, including point mutations, deletions and duplications, alone or in combination, can result in a variety of different disorders. Moreover, the relationship between genotype and phenotype is not straightforward, in part due to heteroplasmy, the tendency for a mitochondrial mutation to be present in only a proportion of the cell's mitochondrial genome copies (see Chapter 10).

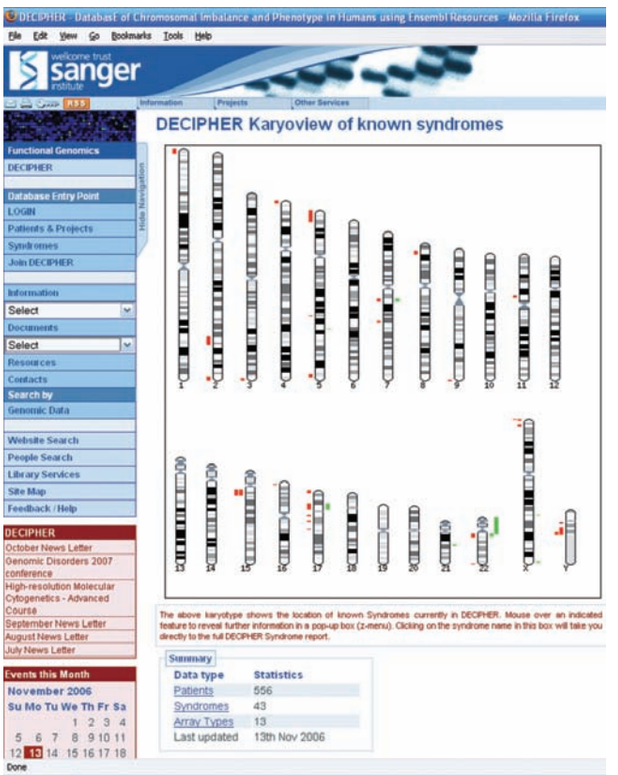


Fig. 1.4 Diagram displayed on the DECIPHER website (at http://decipher.sanger.ac.uk/syndromes) indicating chromosomal loci associated with known clinical syndromes. Reproduced with permission from the Wellcome Trust Sanger Institute. Flicek et al. (2010) Nucleic Acids Res 38 (Database issue):D557–62.

Single-gene disorders

In 1902, Garrod presented his studies on alkaptonuria, a rare condition in which patients have urine that darkens on standing and arthritis. He found three of 11 sets of parents of affected patients to be blood relatives and, in collaboration with Bateson, proposed that this was a Mendelian recessive trait with affected persons homozygous for the underactive gene. This was the first disease to be interpreted as a single-gene trait. Garrod also conceived the idea that patients with alkaptonuria and other inborn errors of metabolism really represented one extreme of human biochemical variation and that other less clinically significant variations were to be expected.

There followed numerous descriptions of distinct human single-gene traits and at the present time more than 7,000 human single-gene traits are known (Table 1.2). In 1949, Pauling suspected an abnormal haemoglobin to be the cause of sickle-cell anaemia, and this was confirmed by Ingram in 1956, who found an altered haemoglobin polypeptide sequence. This was the first demonstration in any organism that a mutation in a structural gene could produce an altered amino acid sequence. In 1959, only two abnormal haemoglobins were known; now the number exceeds 450. In 1948, Gibson

Table 1.2 Human genes and single-gene traits (see McKusick, 2007, and the OMIM database)

	1966	1975	1986	1994	2010
Autosomal dominant	837	1,218	2,201	4,458	19,007 (6,469)
Autosomal recessive	531	947	1,420	1,730	autosomal*
X-linked	119	171	286	412	1,131 (515)
Y-linked	–	–	–	19	59 (11)
Mitochondrial	–	–	–	59	65 (30)
Total	1,487	2,336	3,907	6,678	20,262 (7,025)

*The distinction between autosomal dominant and autosomal recessive traits was not maintained in the Mendelian Inheritance in Man (MIM) catalogue after May 1994 for several reasons. These included: the distinction being only relative (with, for instance, a deficiency state in an otherwise 'autosomal recessive' condition being detectable in a heterozygote with a sufficiently sensitive detection system); and for several conditions, the occurrence of both autosomal dominant and recessive forms that result from the same gene, depending on which specific mutations are present. Figures correct on 22 November 2010. In parenthesis are the total numbers of OMIM entries that have phenotypic information.

demonstrated the first enzyme defect in an autosomal recessive condition (NADH-dependent methaemoglobin reductase in methaemoglobinaemia). The specific biochemical abnormalities in over 400 inborn errors of metabolism have now been determined, but the polypeptide product is still unknown in many human single-gene disorders. Study of these rare, and not so rare, single-gene disorders has provided valuable insights into normal physiological mechanisms; for example, our knowledge of the normal metabolic pathways has been derived largely from the study of inborn errors of metabolism.

Huge progress has been made in the assignment of genes to individual chromosomes, in mapping the genes' precise locations and, more recently, in identifying their entire nucleotide sequences. The first human gene assignment was made by Wilson, who identified the X-linked trait for colour blindness in 1911 and assigned the gene to the X chromosome. Other X-linked traits rapidly followed, while the first autosomal gene to be assigned was thymidine kinase to chromosome 17 in 1967. By 1987, a complete linkage map of all human chromosomes had been developed and this was followed in 1993 by the first physical map. These were essential steps towards the final goal of the Human Genome Project. The Human Genome Project, initiated in 1990, aimed to map and sequence all human genes by the year 2005. Rapid technological advances, particularly the development of high-throughput automated fluorescence-based DNA sequencing (see Chapter 4), in addition to competition between the publicly funded (International Human Gene Sequencing Consortium) and private company (Celera) schemes, led to the early completion of the human genome sequence in 2003 (see Chapter 2). This sequence information, together with an enormous body of associated data, has been made publicly available via internet databases. The information available includes associations with human diseases, gene mapping data, cross-species comparisons, expression patterns and predicted protein features (Fig. 1.5). These and other valuable databases are described in Part 3, and a user's guide is provided online (at www.wiley.com/go/tobias).

Multifactorial (part-genetic) disorders

Galton studied continuous human characteristics such as intelligence and physique, which did not seem to conform to Mendel's laws of inheritance, and an intense debate ensued, with the supporters of Mendel on the one hand and those of Galton on the other. Finally, a statistician, Fisher, reconciled the two sides by showing that such inheritance could be explained by multiple pairs of genes, each with a small but additive effect. Discontinuous traits with multifactorial inheritance, such as congenital malformations, were explained by introducing the concept of a threshold effect for the disorder: manifestation only occurred when the combined genetic and environmental liability passed the threshold. Many human characteristics are determined in this fashion. Usually factors in the environment interact with the genetic background.

Although the genetic contribution to multifactorial disorders is now well accepted, the number and nature of the genes involved and their mechanisms of interaction between each other and environmental factors are largely unknown. This is the current focus of a great deal of research and progress has been made in identifying the genetic contribution for several of these conditions including insulin-dependent diabetes mellitus, rheumatoid arthritis, dementia due to Alzheimer's disease, premature vascular disease, schizophrenia, Parkinson disease, atopic dermatitis and asthma.

Somatic cell genetic (cumulative genetic) disorders

All cancers result from the accumulation of genetic mutations. Usually these mutations only occur after conception and are thus confined to certain somatic cells, but in a small but clinically important proportion, an initial key mutation is inherited. Boveri first advanced the idea that chromosomal changes caused cancer, and early support for this idea came from the demonstration in 1973 of a specific chromosomal translocation (the Philadelphia chromosome) in a type of leukaemia. Subsequently, a large number of both specific and non-specific chromosomal changes have been found in a wide variety of cancers. In turn,

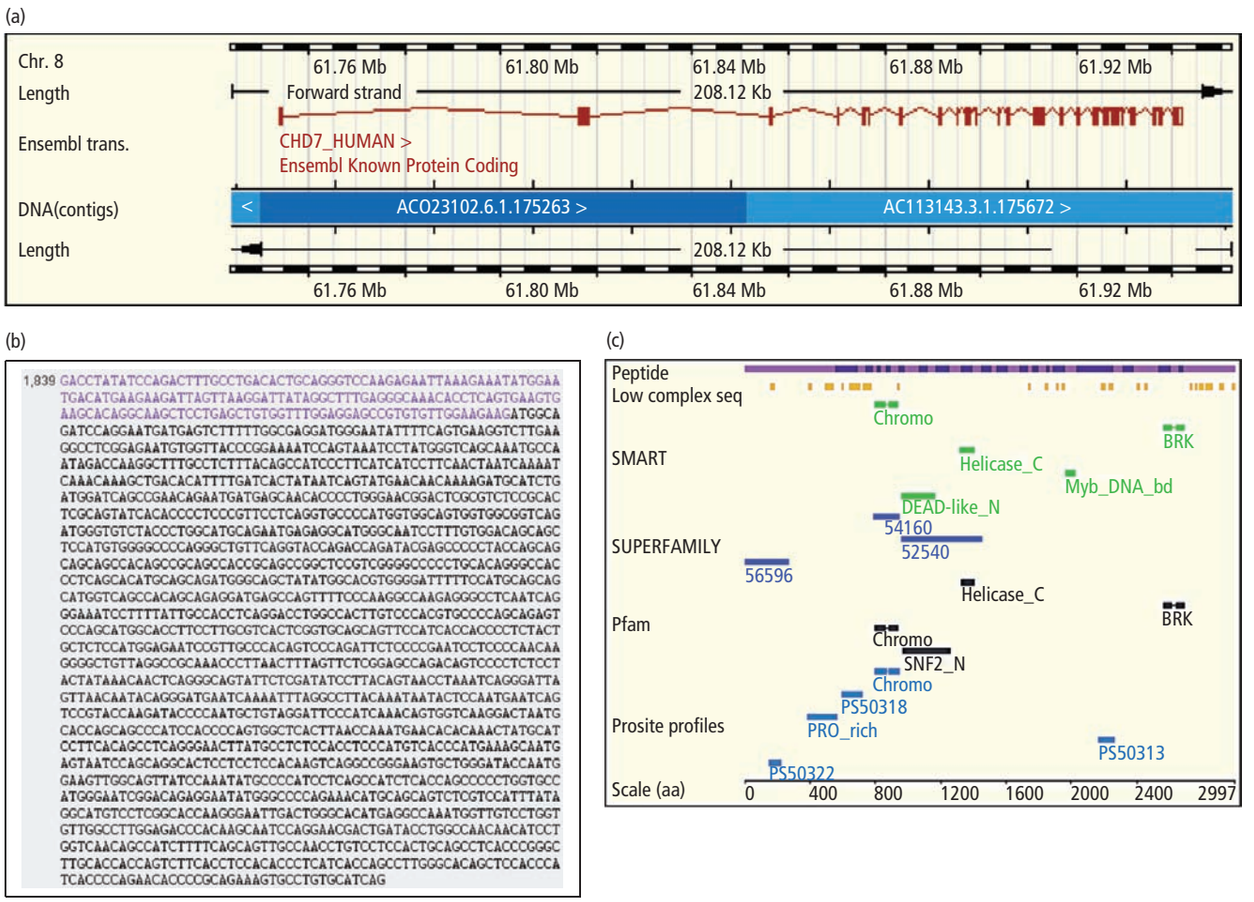


Fig. 1.5 (a) Transcript structure of the 38-exon CHARGE association gene, *CHD7*, on human chromosome 8. (b) DNA sequence of the first coding exon (containing the start codon). The DNA sequence displayed in purple is the untranslated region of this exon, immediately preceding the ATG start codon. (c) Protein features of CHD7, as predicted by the different computer programs (e.g. SMART) shown on the left. Reproduced with permission from the Ensembl database at the Wellcome Trust Sanger Institute. Flicek et al. (2010) Ensembl's 10th year. Nucleic Acids Res 38 (Database issue):D557–62. See Chapter 19.

these changes were clues to specific genes that were key determinants of progression to cancer. Many of these genes have now been cloned and this has resulted in an improved understanding of the molecular basis of cancer and provided the clinician with a means of detection of presymptomatic carriers of cancer-predisposing genes. In addition, it is now recognised that changes in the DNA sequence occurring within somatic cells play an important role in ageing and in certain mosaic disorders such as McCune–Albright syndrome, which results from post-zygotic somatic activating mutations in the *GNAS1* gene. They also may be responsible for the exacerbation of symptoms with age in some inherited disorders such as myotonic dystrophy, in which there is somatic expansion of the inherited mutation (see Chapter 16), and mitochondrial disorders (see Chapter 10).

Clinical applications of medical genetics

Genetically determined disease has become an increasingly important part of ill health in the community now that most infections can be controlled and now that modern medical and nursing care can save many affected infants who previously would have succumbed shortly after birth. This has led to an increased demand for informed genetic counselling and for screening tests both for carrier detection and to identify pregnancies at risk.

Genetic assessment and management

Davenport began to give genetic advice as early as 1910 in the USA, and the first British genetic counselling clinic was established in 1946 at Great Ormond Street, London. Public demand has since caused a proliferation of genetic counselling centres so that there are now more than 40 in the UK and more than 450 in the USA. The scope for genetic counselling has, in fact, in recent years expanded dramatically with the increasingly available data on human genetic disorders (e.g. their mechanism of inheritance in addition to their associated genes and markers) and the increasing availability of mutation analysis. Clinical geneticists play an increasingly important role

in the clinical assessment and genetic testing of patients with genetic conditions and their at-risk relatives. Furthermore, geneticists are now much more involved in the management of patient follow-up, often coordinating several other specialties and initiating patient participation in multicentre clinical studies. These include trials of clinical screening methods and of new therapeutic strategies.

In addition to an accurate assessment of the risks in a family, the clinical geneticist also needs to discuss reproductive options. Important advances in this respect have been made with regard to prenatal diagnosis with the option of selective termination, and this has been a major factor in increasing the demand for genetic counselling. Prenatal diagnosis and now, in certain cases, preimplantation diagnosis (see below), offer reassurance for couples at high risk of serious genetic disorders and allow many couples, who were previously deterred by the risk, the possibility of having healthy children.

Genetic amniocentesis was first attempted in 1966 and the first prenatally detected chromosome abnormality was trisomy 21 in 1969. Chromosome analysis following amniocentesis is now a routine component of obstetric care, and over 200 different types of abnormality have been detected. Amniocentesis or earlier chorionic villus sampling can also be used to detect biochemical alterations in inborn errors of metabolism. This was first used in 1968 for a pregnancy at risk of Lesch–Nyhan syndrome and has since been used for successful prenatal diagnosis in over 150 inborn errors of metabolism. Prenatal diagnosis can also be performed by DNA analysis of fetal samples. This approach was first used in 1976 for a pregnancy at risk of α-thalassaemia and has now been used in over 200 single-gene disorders, and for many of these, including cystic fibrosis, the fragile X syndrome and Duchenne muscular dystrophy, it has become the main method of prenatal diagnosis.

Preimplantation diagnosis (PGD), first used clinically (for sex determination) in 1990, is a more recently established technique that permits the testing of embryos at a very early stage following *in vitro* fertilisation (IVF), prior to implantation in the uterus. In this procedure, a single cell or blastomere is removed by suction, apparently harmlessly, from the embryo. This is usually carried out at the five- to ten-cell stage, at approximately 3 days post-fertilisation. Using the polymerase chain reaction (PCR) or FISH, it is then possible to determine the fetal sex in cases of sex-linked disease or to detect a specific mutation or chromosomal abnormality (also see Chapter 12).

A more recent extension of the PGD technology is the technique known as preimplantation genetic haplotyping (PGH), which was announced in 2006 (see Renwick *et al.*, 2006 in Further reading). In this technique, as in PGD, a cell is extracted from an embryo following IVF. In PGH, however, the DNA undergoes testing for a set of DNA markers closely linked to the disease gene without requiring the prior identification of the precise causative mutation. This can be performed by carrying out simultaneous or multiplex PCRs of several DNA markers, using fluorescence to detect and differentiate the products. The possible future possibilities and likely limitations of PGD are discussed in an interesting opinion article published very recently in *Nature* (see Handyside, 2010).

The prenatal tests that detect chromosomal, biochemical or DNA alterations cannot, however, detect many of the major congenital malformations. The alternative approach of fetal visualisation has been necessary for these. High-resolution ultrasound scanning was first used to make a diagnosis of fetal abnormality (anencephaly) in 1972 and since then over 400 different types of abnormality have been detected. The clinical benefits of the more recently developed three-dimensional ultrasound techniques over standard two-dimensional ultrasound fetal imaging are not yet clear and three-dimensional ultrasound is not currently in routine clinical use during pregnancy in the UK.

Treatment and prevention of genetic disease

A great deal of research has been undertaken into the possibility of effective treatment of genetic diseases. In 1990, the first attempts at human supplementation gene therapy for a single-gene disorder (adenosine deaminase deficiency) were performed. Since then, different gene therapy methods have been devised, depending on the nature of the mutation, and several hundred gene therapy trials are now underway. Unfortunately, the development of a safe, effective, non-immunogenic, well-regulated system that permits the efficient delivery of the therapeutic DNA to sufficient numbers of target cells continues to present a significant challenge.

Although cures for genetic diseases continue to remain elusive, there are now many genetic conditions for which a precise diagnosis leads to significant benefits in terms of clinical management. In some conditions, for example, the almost complete prevention or reversal of the phenotypic effects of a genotype is achievable. This is the case, for instance, with regular venesection for haemochromatosis, with dietary treatment of phenylketonuria (PKU) and medium-chain acyl-CoA dehydrogenase (MCAD) deficiency and with modern enzyme replacement therapy for Gaucher's disease and Fabry's disease. In other cases, appropriate surveillance for clinical complications to permit their early treatment can be instituted. For example, as described in more detail in Chapter 13, screening can permit the early removal of pre-cancerous neoplastic lesions in hereditary cancer syndromes such as familial adenomatous polyposis (FAP), MYH polyposis, hereditary non-polyposis colorectal cancer (HNPCC) and familial breast cancer. In addition, in many other familial conditions, a genetic diagnosis facilitates the detection and early treatment of other complications such as diabetes and heart block in myotonic dystrophy; scoliosis, optic glioma and hypertension in neurofibromatosis type 1 (NF1); and aortic dilatation in Marfan syndrome. Moreover, as mentioned above, following their genetic diagnosis, patients are increasingly enrolled by clinical geneticists in large multicentre trials of new clinical screening and therapeutic methods. Such trials currently include, for instance, biochemical and ultrasound ovarian

screening for women at high risk of developing ovarian cancer and the Mirena intra-uterine device for women with mismatch repair gene mutations who are at risk of endometrial cancer.

The majority of couples are not aware that they are at risk of having offspring with a genetic condition until they have an affected child. This has led to an increased emphasis on prenatal screening, for example by fetal ultrasound examination and by measurement of maternal serum α-fetoprotein and other analytes to detect pregnancies at increased risk of neural tube defects and chromosomal abnormalities. For example, the efficiency of prenatal screening has increased to a point where approximately 85–90% of cases of fetal Down syndrome can be detected by 10–13 weeks' gestation for a false positive rate of 3.5%. Maternal age alone is no longer a suitable indication for prenatal diagnosis and far fewer amniocenteses are now required (see Chapter 17). Neonatal screening was introduced in 1961 for PKU and other conditions where early diagnosis and therapy will permit normal development, such as congenital hypothyroidism. More recently, neonatal screening for cystic fibrosis has been introduced, and it is likely that in the future there will be continued development of population screening, as well as prenatal, neonatal and preconceptional screening, which should lead to a reduced frequency of several genetic diseases.

SUMMARY

- The scientific basis of medical genetics began to be elucidated in 1865 when Mendel published his laws of segregation and independent assortment. These were confirmed around 40 years later.
- Chromosomes were identified in 1882, the hereditary information was shown in 1944 to consist of nucleic acid and the double-helical structure of DNA was discovered in 1953.
- The first single-gene trait, alkaptonuria, was identified in 1902 as a Mendelian recessive condition. Numerous other genes associated with Mendelian traits have been discovered since.
- Extremely rapid advances have been made in gene mapping and automated sequencing, facilitating the early completion of the human genome sequence in 2003.
- Prenatal diagnosis and screening are important adjuncts to genetic counselling as they allow couples at risk of fetal abnormality the confidence to plan for future healthy children.
- PGD is an IVF-based technique that can permit the detection of genetic abnormalities in certain cases, before implantation of an embryo.
- An enormous quantity of human molecular genetic information is now freely available on the internet. Ways of accessing this information are presented in Chapter 19 and online at (www.wiley.com/go/tobias).

FURTHER READING

Bejjani BA, Shaffer LG (2006) Targeted array CGH. *J Mol Diagn* **8**:537–9.

Handyside A (2010) Let parents decide. *Nature* **464**:978–9.

McKusick VA (2007) Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**:588–604.

Ogilvie CM, Braude PR, Scriven PN (2005) Preimplantation genetic diagnosis – an overview. *J Histochem Cytochem* **53**:255–60.

Renwick PJ, Trussler J, Ostad-Saffari E, Fassihi H, Black C, Braude P, Ogilvie CM, Abbs S (2006) Proof of principle and first cases using preimplantation genetic haplotyping – a paradigm shift for embryo diagnosis. *Reprod Biomed Online* **13**:110–9.

WEBSITES

European Society for Human Reproduction and Embryology (ESHRE):
http://www.eshre.com

Human Fertilisation and Embryology Authority (HFEA):
http://www.hfea.gov.uk

OMIM (Online Mendelian Inheritance in Man):
http://www.ncbi.nlm.nih.gov/omim/

Preimplantation Genetics Diagnosis International Society (PGDIS), which is monitoring PGD activity worldwide:
http://www.pgdis.org/

Self-assessment

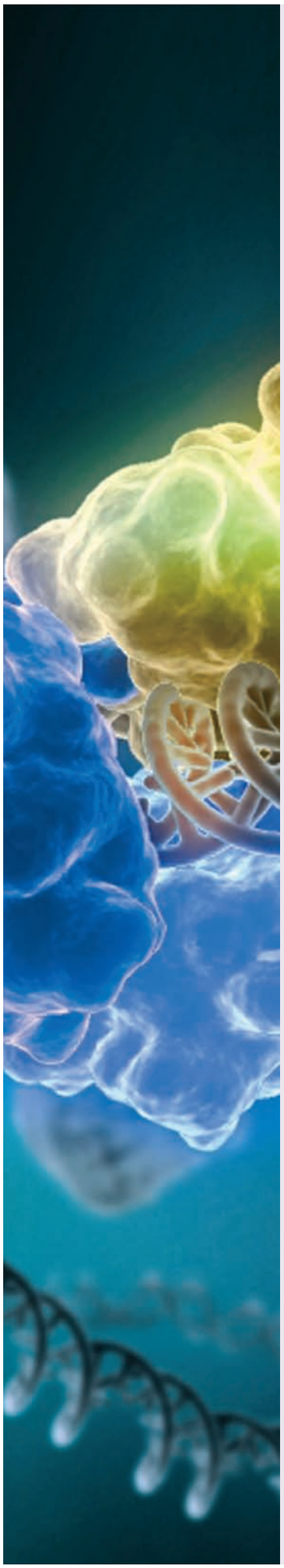
1. Which of the following is not a typical feature of mitochondrial inheritance?
A. Maternal transmission
B. Heteroplasmy
C. More introns in mitochondrial genes than in nuclear genes
D. The presence of fewer than 40 genes in the mitochondrial genome
E. Lack of a straightforward genotype–phenotype relationship

2. In preimplantation genetic diagnosis (PGD), which of the following does not take place?
A. *In vitro* fertilisation
B. Testing of each of the cells of the embryo for the specific mutation
C. Fetal sex determination of embryos in sex-linked disease
D. The use of the polymerase chain reaction (PCR) to detect a specific mutation or haplotype
E. The use of fluorescence *in situ* hybridisation (FISH) to detect an unbalanced chromosome abnormality

3. Which one of the following conditions is not usually regarded as multifactorial?
A. Rheumatoid arthritis
B. Insulin-dependent diabetes mellitus
C. McCune–Albright syndrome
D. Asthma
E. Parkinson disease

4. Which of the following is not useful in connection with the following genetic conditions?
A. Venesection for iron overload in haemochromatosis
B. Regular blood pressure check in neurofibromatosis type 1 (NF1)
C. Neonatal screening for hypothyroidism and phenylketonuria (PKU)
D. Dietary treatment for PKU
E. Enzyme replacement therapy for familial adenomatous polyposis (FAP)

5. Which of the following pairings between individuals and a genetics landmark is incorrect?
A. Mendel and the independent assortment of different gene pairs to gametes
B. Flemming and the identification of chromosomes within the nucleus
C. The discovery of the helical structure of DNA and Watson, Crick, Franklin and Wilkins
D. The first identification of a chromosomal abnormality and Jeffreys
E. PCR and Mullis



CHAPTER 2

The human genome

Key Topics

Introduction

Our knowledge and understanding of the structure and function of the human genome have been vastly augmented by the data generated by the Human Genome Project, completed in 2003. Although, prior to this achievement, the general location of many genes on the chromosomes and their positions relative to each other had been determined (i.e. by 'gene mapping'), the full nucleotide sequence of the chromosomes elucidated by the Human Genome Project provided far more detailed and reliable information. How this was achieved, the insights gained from the data and the uncertainties that remain are outlined within this chapter.

Structure and organisation of the genome

The human nuclear genome contains approximately 3280 million base pairs (bp). In contrast, the much smaller mitochondrial genome (discussed in Chapter 10), which was sequenced in 1981, contains only 16,569 bp and 37 genes. The size of the coding region of a human gene contained in the nucleus is approximately 1000–3500 bp, and there are currently only 30,073 identified genes (21,598 protein-encoding genes and 8,475 RNA genes – see the Ensembl website in Further reading for the latest update. In fact, only 1.1% of the genome is actually protein-coding DNA. Another 4% at least, is, however, also important, consisting of gene-regulatory sequences and RNA genes. A large proportion of the non-coding DNA, around 20% of the genome, consists of introns and untranslated regions of genes in addition to other non-coding gene-related sequences such as pseudogenes. The majority of the non-coding DNA, however, around 75% of the genome, is extragenic, and much of this DNA (55% of the genome) consists of repeated sequences. The majority of this repetitive sequence is derived from transposable elements or transposons, sequences that insert additional copies of themselves randomly throughout the genome and constituting around 45% of the total DNA. These repetitive sequences permit, through the process of recombination (crossing over between two homologous DNA molecules), the rearrangement of parts of the genome, over time modifying the properties of existing genes and even creating new genes. Intriguingly, the proportion of repetitive sequence within the human genome (>50%) is significantly higher than in other organisms, with the corresponding figure being only 3% in the fly and 7% in the worm.

The genes are now known to be clustered in randomly distributed areas within the genome with long regions of non-coding DNA between these gene-dense regions. In general, the gene-rich areas tend to have a higher guanine and cytosine (G + C) content than the gene-poor regions and they tend to appear negative or pale on Giemsa chromosome staining (see Chapter 5).

The clustering of genes encoding structural proteins in part reflects ancestral small duplications with subsequent divergence of function, facilitating evolution by natural selection where the resulting new gene can provide a selective advantage. In this process, some genes become non-functional gene copies termed pseudogenes (e.g. those within the β-globin cluster), some retain similar functions (e.g. the red–green colour vision genes) and some develop novel functions as a result of small sequence changes or exon shuffling. In contrast, the loci for genes of sequential steps in a metabolic pathway tend to be scattered, as are the loci for subunits of complex proteins and the loci for mitochondrial and soluble forms of the same enzyme.

Gene identification

In the past, if a gene's protein product was known, the gene could be cloned by functional cloning. The protein was isolated and the partial sequence of its amino acids determined. This then allowed the synthesis of a corresponding series of oligonucleotide probes based on the genetic code (see Table 3.2) which could be used to identify the complementary gene from a DNA library.

If the gene's protein product was unknown, the gene could be cloned by positional cloning. The first step was to chromosomally map the gene and then to identify candidate genes from that region. The correct candidate was identified by mutational analysis in patients with the disease trait. This procedure has now been greatly facilitated by the availability of accurate mapping and sequence data resulting from the Human Genome Project.

Recently, many genes have been identified by the automated DNA sequencing of the genome as part of the Human Genome Project followed by gene prediction analyses in which genes are recognised by the computerised detection of typical gene features such as transcriptional and translational initiation and termination sequences. The probable functions of these genes can often also be predicted, by automated homology searches in which similarities are found between the sequences of newly identified genes and those of genes, proteins or protein domains already listed in the databases. Nevertheless, the functions and disease associations of many recently identified genes remain to be ascertained.

The Human Genome Project

How it was carried out

The Human Genome Project was commenced in 1990, with the aims of identifying and sequencing all the genes in the human genome within 15 years and making the data publicly available. It was initially coordinated by the US Department of Energy (directed by Ari Patrinos) and the US National Institutes of Health (directed by Francis Collins). The Wellcome Trust Sanger Institute at Hinxton in the UK also became a major partner, ultimately sequencing around one-third of the genome (chromosomes 1, 6, 9, 10, 11, 13, 20, 22 and X), under the direction of Sir John Sulston (Nobel laureate, 2002). In fact, a Human Genome Sequencing Consortium comprising a total of 16 institutions in the USA, Europe, China and Japan was required to carry out the enormous sequencing task. In addition, three institutions provided the necessary complex computational analysis: the National Center for Biotechnology Information (NCBI) at the National Institutes of Health, USA; the European Bioinformatics Institute (EBI) in Cambridge, UK; and the University of California, Santa Cruz (UCSC), USA. The strategy used was a 'hierarchical shotgun method' in which the regions of chromosomes submitted for fragmentation ('shotgunning') and sequencing were large stretches of DNA whose location in the genome had already been determined and which were contained in so-called bacterial artificial chromosomes (BACs).

In September 1999, Craig Venter's private company, Celera, also began to sequence the genome, but using a different

strategy known as the 'whole genome shotgun approach'. This involved initially breaking up the entire genome (rather than BAC clone inserts) into millions of small fragments, sequencing these pieces in no particular order and subsequently reassembling the chromosome sequence by a massive computer analysis on the basis of sequence overlaps. Although the whole shotgun method did not necessitate the prior construction of a map of large fragments covering the genome, there were other challenges in the assembly phase. The public and private projects both used similar fluorescence-based automated sequencing technology, based on the dideoxy sequencing strategy originally devised by the double Nobel laureate Fred Sanger and colleagues, many years previously (see Chapter 4). The even faster recent sequencing technologies now provide the opportunity to compare many individual human genomes and to determine the extent and significance of genetic variation among people and between different ethnic groups (see the review by Tucker *et al.*, 2009, in Further reading).

Total gene numbers

The number of genes on each chromosome varies greatly, with the largest chromosome, chromosome 1, containing the most (2706 genes) and the Y chromosome the fewest (104 genes). The precise total number of genes varies according to the methods used to identify sequences as genes and by the subtypes of genes that are included in the totals. For instance, as mentioned above, in addition to at least 21,598 protein-coding genes, there are at least 8,475 genes that code for RNA molecules but do not encode polypeptides. These RNA genes currently include at least 727 ribosomal RNA (rRNA) and 131 transfer RNA (tRNA) genes. A surprising number of other RNA genes are also now known to be present, although, due to the difficulty in precisely identifying these genes within the genome, the total number is probably still not completely determined.

Recently described RNA gene classes

While their physiological roles are not yet as clearly understood as those of messenger RNAs (mRNAs), rRNAs and tRNAs, a number of intriguing additional RNA molecules are generally believed to be involved in the regulation of gene expression. They include, for instance, at least 903 small cytoplasmic RNA (scRNA) genes, 1048 microRNA genes and 2019 genes that encode small nuclear RNAs (snRNAs) (see Table 2.1). The snRNAs include RNAs that participate in splicing and a subclass of 1173 small nucleolar RNA (snoRNA) genes. These snRNAs are now known to direct the formation and chemical modification (by methylation and pseudouridylation) of other RNAs such as precursor rRNAs. Remarkably, many snoRNAs are processed from the spliced-out introns of other genes rather than being transcribed from separate genes (see Kiss, 2006, in Further reading). In contrast, small cytoplasmic RNAs are usually found in association with cytoplasmic proteins in complexes termed small cytoplasmic ribonucleoproteins (scRNPs), of which an example is the so-called signal recognition particle. MicroRNAs (termed miRNAs in 2001) are short single-stranded RNA molecules of 21–23 nucleotides that regulate the expression of other genes by binding to mRNAs (particularly the 3′ untranslated region, or 3′ UTR, in humans) and causing the degradation of the latter or blocking their translation into proteins. In recent years, there has been enormous scientific interest in these molecules, which do not encode proteins themselves, and a large number of reports of miRNA expression profiles (patterns) or 'signatures' that may be characteristic of specific tissues. In this respect, the tissues that have been particularly frequently studied are those that have been affected by

Table 2.1 Types of RNA

Type	Location	Comments
Messenger RNA (mRNA)	Nucleus and cytoplasm	Variable size, base sequence complementary to transcribed DNA strand, about 4% of total cellular RNA, half-life 7–24 h
Transfer RNA (tRNA)	Cytoplasm	Hairpin-loop shape, 49 cytoplasmic (and 22 mitochondrial) types, amino acid specific, about 10% of total cellular RNA with tens to hundreds of copies of the genes for each tRNA species
Ribosomal RNA (rRNA)	Ribosomes	40–50% of total cellular RNA, synthesised and stored in the nucleolus and nucleoli
Heterogeneous RNA (hnRNA)	Nucleus	High-molecular-weight mRNA precursors; 40–50% of total cellular RNA
Small nuclear RNA (snRNA)	Nucleus	Several types (e.g. U1–U12), involved mainly in RNA splicing
Small nucleolar RNA (snoRNA)	Nucleolus	At least 340 types, involved in chemical modification of rRNA molecules
Small cytoplasmic RNA (scRNA)	Cytoplasm	Form complexes (e.g. signal recognition particle) with cytoplasmic proteins
MicroRNAs (miRNA)	Cytoplasm	Very small (21–23 nucleotides) antisense regulators of other genes. Formed from a long precursor hairpin RNA by the enzyme DICER. Bind to mRNAs and can prevent their translation or induce their degradation. At least 1048 human miRNAs recognised.

conditions such as cancer of various types. It is believed that such miRNA profiles (like mRNA profiles) could serve, in the future, as useful biomarkers of specific phenotypes and may thus be able to provide improved diagnostic and prognostic information to clinicians. Furthermore, the possible pharmacological targeting of specific miRNAs is now being explored (see review by Ferracin *et al.*, 2010, in Further reading).

Uses of the Human Genome Project data and ways of accessing it

An important benefit of the Human Genome Project is the ability to use the electronically compiled genome data to identify genes of interest at particular locations in the genome. This could include, for instance, those genes located around an identified translocation breakpoint, within a microdeletion or microduplication region (e.g. following array comparative genomic hybridisation or aCGH), or those residing at a locus resulting from a linkage study. Such a locus may be defined as a cytogenetic band. Alternatively, it may be a region spanned by a specific probe (such as a recombinant plasmid BAC probe), delineated by known DNA markers (such as microsatellites, see Chapter 3), or defined by precise nucleotide positions as counted from the end of the short arm of the chromosome. Accessing the genome data can be achieved by using one of the well-established genome browsers, such as Ensembl (Figs 2.1 and 2.2) or UCSC (Fig. 2.3), further details of which are given in Chapter 19 (and updated web-links are provided online at www.wiley.com/go/tobias). The genome databases can also be interrogated using BLAST (Basic Local Alignment Search Tool), which will find the site in the genome of any entered stretch of DNA (or protein) sequence. Detailed information about the sequence is available, such as its precise chromosome location, whether it is within an exon or intron or part of a repeat, if it is part of a known gene or gene

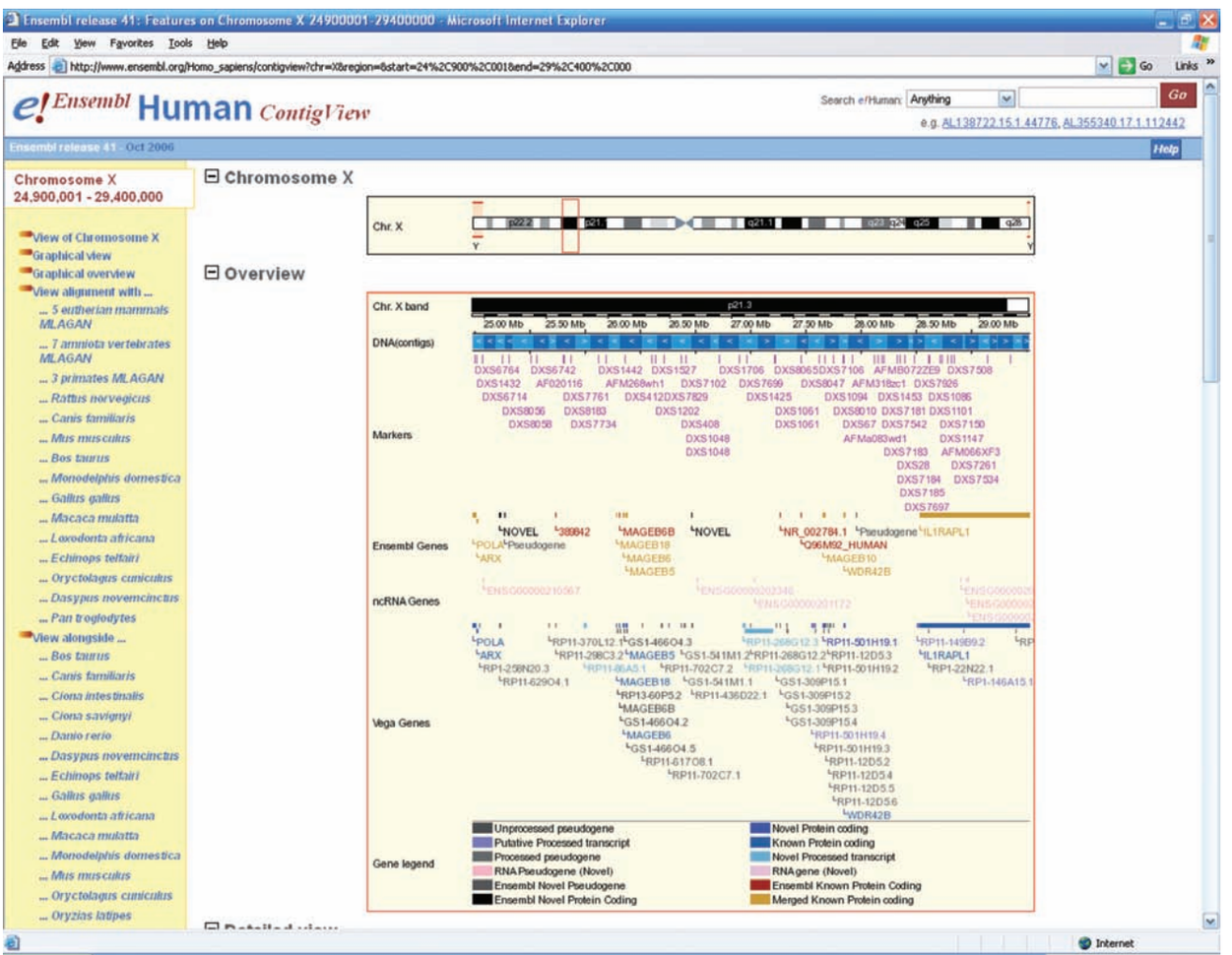


Fig. 2.1 A *Homo sapiens* genome browser display page at Ensembl. This can be reached via the search page at http://www.ensembl.org/Homo_sapiens/index.html. The same region as that shown in the UCSC genome browser example in Fig. 2.3 is displayed. This can be revealed by typing the nucleotide boundaries of the region directly into the sequence position boxes in the *H. sapiens* browser window, shown in Fig. 2.2. Reproduced with permission from the Wellcome Trust Sanger Institute. Flicek et al. (2010) Ensembl's 10th year. Nucleic Acids Res 38 (Database issue):D557–62.

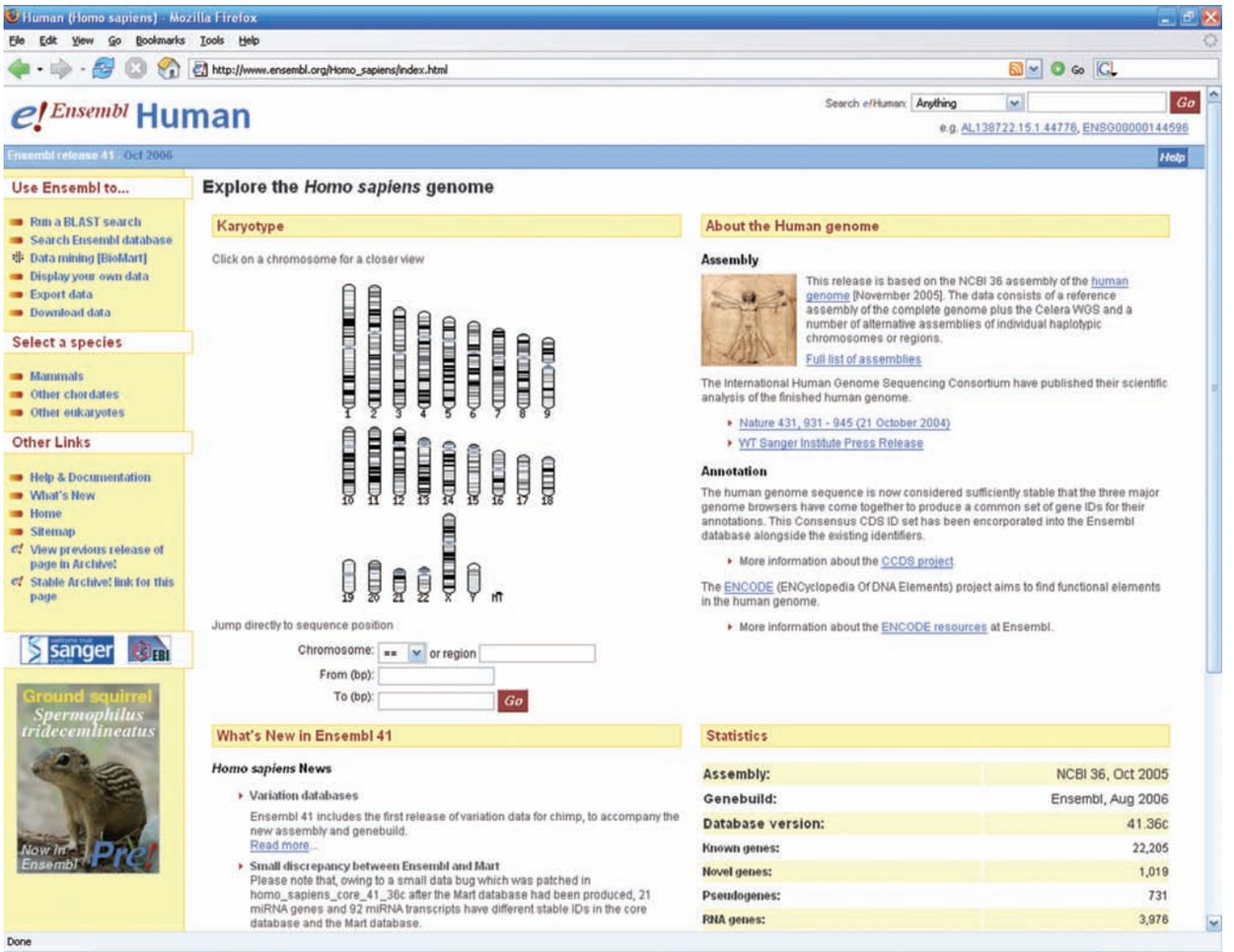


Fig. 2.2 The *H. sapiens* browser window of Ensembl. The latest version of this page can be accessed at http://www.ensembl.org/Homo_sapiens/index.html. Reproduced with permission from the Wellcome Trust Sanger Institute. Flicek et al. (2010) Ensembl's 10th year. Nucleic Acids Res 38 (Database issue):D557–62.

family and whether or not it is conserved in other organisms (see chapter 19 and the accompanying website for further details).

There are many clinical applications of the genome data. For instance it is often helpful to use the DNA sequence of a particular gene to act as an initial reference sequence when carrying out mutation screening, and in order to identify the intron/exon boundaries, regulatory elements and untranslated regions of a gene. The data also permit the rapid identification of genetic markers within or adjacent to genes of interest for human family linkage studies. Moreover, a searchable database of DNA reference sequences is invaluable when designing an oligonucleotide PCR primer that will anneal to a sequence of interest without binding to any other sequence in the genome. Databases (discussed in Chapter 19) also exist that provide continuously updated information regarding gene-related human disease information together with the publications that reported all of these findings.

There are many other uses of the data, including several research-related applications. For instance, the databases facilitate the rapid identification of previously unknown members of recognised gene families, the expression patterns in different stages of embryonic development and in different tissues, and studies of inter-individual sequence variation, including single nucleotide polymorphisms, or SNPs. In addition, cross-species comparisons of gene sequences can be made, allowing identification of evolutionarily conserved, functionally important regions of a gene or protein.

Remaining uncertainties

Despite the abundance of data resulting from the sequencing of the human genome (in addition to that of several other organisms) and from the complex post-sequencing analyses that are currently underway, there remain several areas of uncertainty. These include, firstly, the total gene count. Much of the uncertainty relating to the precise total number of functional genes is a result of the necessary extensive use of *in silico* sequence comparisons and analyses by complex gene