HUMAN MOLECULAR GENETICS

4TH EDITION

TOM STRACHAN AND ANDREW READ

GS
Garland Science

HUMAN MOLECULAR GENETICS

4TH EDITION

TOM STRACHAN AND ANDREW READ

GS
Garland Science

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

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

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

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

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

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

ISBN 978- 0-815-34149-9

Library of Congress Cataloging-in-Publication Data
Strachan, T.
Human molecular genetics / Tom Strachan and Andrew Read. -- 4th ed.
p. ; cm.
Rev. ed. of: Human molecular genetics 3 / Tom Strachan and Andrew Read. 3rd ed. c2004.
Includes bibliographical references and indexes.
ISBN 978-0-8153-4149-9
1. Human molecular genetics. I. Read, Andrew P., 1939- II. Strachan, T. Human molecular genetics 3. III. Title.
[DNLM: 1. Genome, Human. 2. Molecular Biology. QU 470 S894h 2010]
QH431.S787 2010
611'.01816--dc22
2010004605

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

Printed in the United States of America
15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

T&F informa
Taylor & Francis Group, an informa business

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

Preface

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

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

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

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

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

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

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

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

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

TEACHING RESOURCES

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

Acknowledgments

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

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

Contents

Detailed Contents