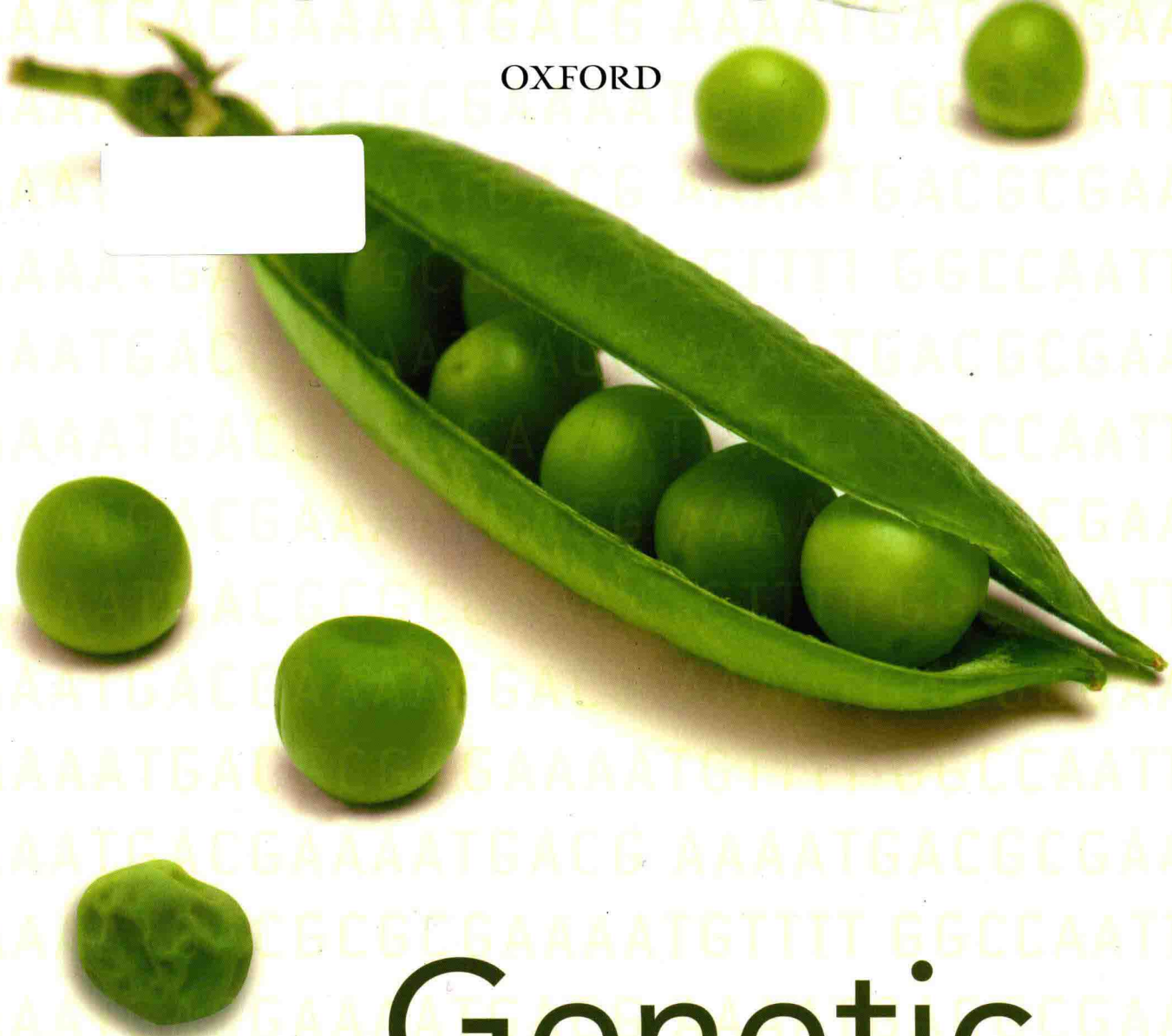
OXFORD



Genetic Analysis

Genes, Genomes, and Networks in Eukaryotes

Philip Meneely

Second Edition

Genetic Analysis

Genes, Genomes, and Networks in Eukaryotes

Philip Meneely

Haverford College

Second Edition

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First edition 2009

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2013950212

ISBN 978-0-19-968126-6 (hbk)
ISBN 978-0-19-965181-8 (pbk)

Printed and bound in China by
C & C Offset Printing Co. Ltd

Genetic Analysis

PREFACE TO THE SECOND EDITION

As with the first edition, the goal of this book is to describe the logic of genetic analysis as it applies in an era when genome sequences have become readily available. We are learning much from having DNA sequences from the genomes of so many different organisms, and we undoubtedly have many lessons yet to be learned. But all that we learn from these new and powerful sequencing methods should be interpreted in light of what was known previously from work in classical genetics and molecular biology. As recounted in the book, the mantra of genetic analysis for decades has been: 'Find a mutant!' It may be tempting to think that the mantra in this new era of genetic analysis is: 'Sequence the genome!' In this book, I show how finding a mutant and sequencing a genome are both essential tools for studying complicated biological problems.

The book describes both of these essential tools, mutants and sequences, as well as many other tools for genetic analysis. However, I focus on the strategies for using these tools and thinking about what the results mean. Thus, the emphasis of the book is not on the tools themselves nor even on what we have learned from these tools, although those are part of the necessary foundation. Geneticists worked successfully and intelligently for many years on many different biological questions before it became possible to find the DNA sequence of even a single gene, let alone an entire genome. The ready access to all of the genome sequence information does not replace this previous work, but rather adds a rich dimension to it. That richness is what I hope to have captured in this book.

Course suitability

I have taught a course entitled advanced genetic analysis for undergraduates at Haverford College for more than 15 years. These students have taken an introductory course in basic genetics and molecular biology, and have a general familiarity with biochemistry and cell biology as well. The book is based on that course, so it is written with that level of preparation in mind. The first edition of this book was entitled *Advanced Genetic Analysis* because it was aimed at advanced undergraduates. We have changed the title for this edition to *Genetic Analysis* in part because some instructors use it for an introductory graduate class rather than an advanced undergraduate class. Although the structure is designed for one semester, this book includes far more topics than I am able to include in my course. Even with that, I am keenly aware of the many topics that I have had to omit to keep the book to a manageable length and within my own areas of knowledge.

I have retained a focus on five model organisms, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus*, with occasional references or examples drawn from other well-studied model organisms. For those who work on a different model organism and feel

slighted that I have not included, for example, fission yeast or zebrafish, I sincerely apologize for not including these organisms. I tend to be syncretic, so I would like to have the ability and the space to talk about many more model organisms than I do. Material on genomic and genetic analysis on these and other model organisms is included in the Online Resource Centre (ORC). Coverage of human genetics has been expanded in this edition, and several chapters (notably Chapters 3, 5, and 9) include many examples from human genetic diversity and disease. The increased emphasis on human genetics is unabashedly parochial; like most of my students, I find our species to be particularly interesting. But it also reflects the maturity of human genetics that these provide some of the best and most interesting examples of fundamental biological concepts that have been studied by genetic approaches.

Organization and changes in the second edition

While the overall approach to the topics is similar to the first edition, the organization has been changed somewhat. There are now 14 chapters in four units. The first unit, ***Genes and genomes***, retains a historical overview of genetic principles in Chapter 1 and an introduction to the five model organisms in Chapter 2, with revisions and updates to each chapter. In particular, reporter genes and microarrays, important tools for the molecular analysis of genes and genomes, are now included in Chapter 1, as is an expanded discussion of genes whose functional product is an RNA molecule. Chapter 3 is entirely new in this edition, and places the genome into the context of chromosome structure; this allows a substantial discussion of epigenetics and a new case study on imprinting. Epigenetics is both an old and an emerging field, and some of its unifying principles remain to be uncovered. The ENCODE-related projects to identify all of the functional elements in a genome are discussed in detail in Chapter 3 and are providing a wealth of new information about gene structures, non-coding RNA genes, histone modifications, and other key principles that affect our understanding of biology at every level of education. I used them as a narrative theme, while also trying to acknowledge some of the limitations of such large-scale projects.

The second unit, ***Genes and mutations***, includes the same chapters as in the first edition, although each with substantial revisions. Chapter 4 combines information on genetic screens and mutant analysis into one revised (and somewhat shortened) chapter. Chapter 5 describes how genes are cloned, with a substantially revised approach based on having genomes completely sequenced. Most significantly, examples of how genes responsible for rare genetic diseases have been identified by exome sequencing are now included and thoroughly discussed. Chapter 6 is an updated chapter on 'reverse genetics', primarily with gene disruptions in yeast and mice; Chapter 7 is an updated chapter on genome-wide mutant screens, including RNAi screens which now are the dominant methods for finding mutant phenotypes in flies and worms. Some material from the first edition has now been moved to the ORC for these chapters, particularly in those situations where the classical genetic approach is no longer widely used.

The third unit, ***Genes and populations***, consists of three entirely new chapters. Chapter 8 is an introduction to genome-wide association studies, with an emphasis on their application to gene identification in humans. It puts these into the context of other positional cloning methods, which have been used in many organisms for decades. Using the background from Chapter 8, Chapter 9 focuses on complex traits, an important but perhaps neglected topic in many undergraduate genetics courses. I was pleased to be able to return to concepts that I learned and appreciated in my graduate courses in quantitative genetics many years ago in writing this chapter. Both of these chapters (as well as Chapter 3 on epigenetics) began to be developed soon after the first edition appeared, and I have used draft versions with students in my courses for the past few years. Chapter 10 continues the discussion from Chapter 8 about the genetic structure of natural populations, and uses three examples to show how natural variation can be used like laboratory-induced mutations in the analysis of a fundamental biological question, in this case evolutionary adaptation. This unit replaces Unit 3 on gene activity from the first edition. Many of the topics on gene activity from those chapters in the first edition have been moved into other chapters in this edition, while other topics such as mosaic analysis and temperature-shift experiments have been rewritten and are available in the ORC.

The fourth unit, ***Genes and pathways***, focuses on gene interactions. Chapter 11 on suppressors and enhancers and Chapter 12 on epistatic pathways are revisions of the corresponding chapters from the first edition. Chapter 13 is a much updated introduction to gene interaction networks and some principles of systems biology. Genome-wide analysis of gene interactions in yeast, referred to as 'the genetic landscape of a cell', is a significant new feature of this chapter and most of the work appeared since the first edition was published. Finally, Chapter 14 on genes and phenotypes is another new chapter that provides some context for thinking about genome-enabled medicine or personal genomics and the challenges of trying to predict the phenotype based on knowledge of the genotype. Because of genome sequencing, genotypes are easier to know now than at any previous time; connecting the genotype to the phenotype is a central link in personalized genomics and genome-enabled medicine. While many of the ideas in the chapter are not new—geneticists have thought about these challenges for many decades—these have been placed into the context of our ability to analyze and sequence individual genomes.

Pedagogical features

The book is designed to be accessible to different levels of expertise by the use of progressively more focused sections of each chapter. First, and most broadly accessible, is the main narrative of the chapter. I have tried to maintain a tone that is conversational and familiar without being overly casual, although some readers and teachers may want greater detail. My grandfather was an entertaining and inveterate story-teller; it was only later that I realized that he sometimes invented episodes and descriptions to make a better narrative. I hope to have retained his story-telling

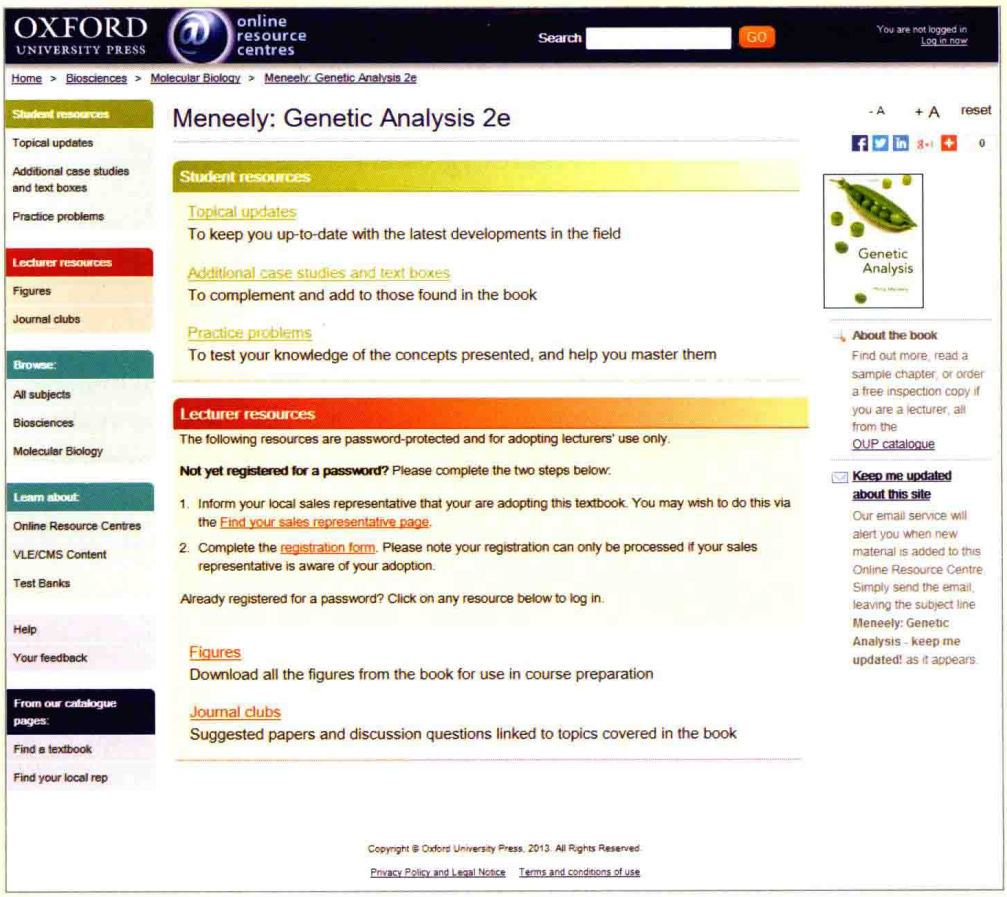
legacy without also including the less-accurate factual elements of his narratives. I am confident that I have not consciously invented any part of the narrative, but further experimentation may prove that I have been guilty of leaving out important information or of misinterpreting what we currently see. For those who want more detail or to find the source of some of the information, each chapter now includes references for additional reading. Specific references in some chapters are also available online. To improve its use by students, the book also includes a glossary and an index. In the text, **glossary** terms are emboldened and colored blue.

The book contains two features to add depth to the main narrative. Most chapters have ***text boxes*** that describe some particular topic in more detail; some chapters have multiple text boxes. My opinion is that some of the most intriguing material in the book resides in these text boxes. Most chapters also have an associated ***case study*** that looks at one particular example in much more detail, using the original papers and data as much as possible. My eventual goal is to make the case studies into a 'course within a course', so that an instructor could cover these or similar case studies in class while leaving the main narrative for the students to read. I am always on the alert for good examples to use as case studies, some of which I expect to make available in the ORC.

Genetics is a rapidly changing field, of course. The ORC will provide updates and corrections, as well as additional information for each chapter. Two features at the ORC are designed specifically to enhance student learning. First, as in the first edition, the ORC includes one or more guided journal clubs for each chapter. In these journal clubs, a student is guided through an original research paper, frequently one associated with the case study, via a series of questions with answers. I often use such journal clubs in my courses to help students learn to read the original papers and analyze the original data, so I encourage instructors to use these as well. The second feature of the ORC to enhance learning is that I have now included some study questions for each chapter. These range from testing definitions and fundamental concepts to more advanced questions. The study questions focus primarily on the main narrative so they are designed to be used together with the chapter to provide that depth of coverage, while the case studies and journal clubs are designed to be used together to provide greater depth for interested or more advanced students.

Online Resource Centre

www.oxfordtextbooks.co.uk/orc/meneely2e/



This book has been written with web usage in mind, so the reader is urged to consult the website associated with the book (www.oxfordtextbooks.co.uk/orc/meneely2e/) for updates and additional resources as they are developed. Key resources on this site include the following:

For registered adopters of the book

- Figures from the book in electronic format, ready to download.
- Journal club: suggested papers and discussion questions linked to topics covered in the book.

For students

- Topical updates: key updates on topics or tools presented in the book to keep you up to date with the latest developments in the field.
- Additional case studies and text boxes to complement and add to those found in the book.
- Practice problems, to test your knowledge of the concepts presented, and help you master them.

Bookmark the site, or register for updates by going to **www.oxfordtextbooks.co.uk/orc/meneely2e/** and clicking on the 'Keep me updated about this site' link.

ACKNOWLEDGMENTS

During the course of the two editions of this book and other projects, Jonathan Crowe has become more than an editor. His insights, wisdom, patience, and good humor have given me more support and guidance than anyone could expect. He is an outstanding editor, who clarifies some of my most tangled paragraphs with a few simple corrections and suggestions. The expertise of the editorial team at Oxford University Press, which Jonathan heads, has produced a much better book than I could have imagined. Thank you.

My colleague Matthew Willmann has made many valuable contributions both for this edition and the previous one. He continued to teach me that plants cannot be thought of only as animals that have chlorophyll, and I have appreciated his many contributions. My colleagues at Haverford College, particularly Rachel Hoang and Mel Santer, have been both supportive and helpful as I think through and discuss different parts of the book.

I have the great privilege to teach genetics to some of the brightest and most industrious undergraduates that any professor could want. This book reflects their contributions and questions throughout. At the risk of overlooking some of the most significant contributions, I want to particularly recognize how Michael Ross and Richard Sarro helped me understand next-generation sequencing, and how Richard, Kulia Wooddell, and I worked together through the Costanzo *et al.* paper that formed the cornerstone of Chapter 13. I could probably cite some student contribution to every chapter, but these stand out. I am fond of telling people that I get paid for grading papers and sitting on committees, but that I happily teach genetics for free.

Foremost among those to thank are the members of my family for their support of this book and their indulgence with my love of genetics. They accept, with a sigh, that any dinner table conversation or walk around the park might turn into a discourse on some principles of genetics. My wife Deb encouraged me to write and accepted that most evenings and weekends involved me working on this book or another project involving genetics. My children have married and my family has grown to the F_2 generation since the first edition. Nothing else in my experience with genetics is as enriching as seeing these next generations.

FIGURE ACKNOWLEDGMENTS

Figure 2.1 (A) Fly: istock.com; Mouse: Photodisc; Worm: National Institutes of Health (B) Cricket: Photodisc; Fly: istock.com; Butterfly: Photodisc.

Figure 2.14 Copyright Cold Spring Harbor Press and used by permission. Also courtesy of Zeynep Altun of Wormatlas.

Figure 2.15 Used by permission of Cold Spring Harbor Press and Zeynep Altun, courtesy of Wormatlas.

Figure 2.21 Courtesy of Min Han, University of Colorado.

Figure 2.22 Copyright Oxford University Press.

Figure 2.23 Copyright Oxford University Press.

Figure 2.25 Courtesy of Peter Carlson and John Sedat, University of California, San Francisco.

Figure 2.26 Courtesy of Matthew Willmann, University of Pennsylvania.

Figure 2.38 Copyright Nature Publishing Group and used by permission; and Photodisc.

Figure C2.1 Copyright Nature Publishing Group and used by permission; and Nancy Nehring/istock.com.

Figure C2.2 Copyright Nature Publishing Group and used by permission.

Figure 3.1 From The ENCODE Project Consortium (2011). A User's Guide to the Encyclopedia of DNA elements (ENCODE). *PLOS Biology* **9**: e1001046.

Figure 3.3 Kenny Louie/CC-BY.

Figure 3.4 From The ENCODE Project Consortium (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **330**: 1787. Copyright AAAS.

Figure 3.8 From Hardison and Taylor (2012). *Nature Reviews Genetics* **13**: 469–83.

Figure 3.22 From Gerstein *et al.* (2010). Integrative Analysis of the *Caenorhabditis elegans* Genome by the modENCODE Project. *Science* **330**: 1775. Copyright AAAS.

Figure 3.24 Wolffe & Hayes (1999). Chromatin disruption and modification. *Nucleic Acids Research* **27**(3): 711–20, by permission of Oxford University Press.

Figures 3.31 and 3.35 From The ENCODE Project Consortium (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**: 1787. Copyright AAAS.

Figure 3.36 From Gerstein *et al.* (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE Project. *Science* **330**: 1775. Copyright AAAS.

Figures 3.37–3.39 From Riddle *et al.* (2011). Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Research* **21**: 147–63. Copyright Cold Spring Harbor Press and used by permission.

Figure B3.1 From Craig *et al.* (2010). *Molecular Biology: Principles of Genome Function.* Copyright Oxford University Press.

Figures B3.3–3.6 From Metzker (2010). Sequencing technologies – the next generation. *Nature Reviews Genetics* **11**:31–46.

Figure 4.6 Courtesy of Maize Genetics and Genomics Database.com.

Figure 5.1 Snail photo: Monika Korzeniec/CC-BY-SA.

Figure 5.15 From Teer and Mullikin (2010). Exome Sequencing: the sweet spot before whole genomes. *Human Molecular Genetics* **19**: R145. With permission from Oxford University Press.

Figures 6.6 and 6.9 Derived from Capecchi (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews Genetics* **6**: 507. Copyright Nature Publishing Group.

Figure C6.2 Copyright AAAS and used by permission.

Figure 7.10 White: R. Daniaud/istock.com and Purple: R. Sanchez/istock.com, © Stephen Vickers, Fotolia.com.

Figure 8.17 From Sutter *et al.* (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**:b112–15. Copyright AAAS.

Figure 9.7 From 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**: 661–78. Copyright Nature Publishing Group.

Figure 10.1 Photograph by Robert C. Wilson, and used by permission.

Figures 10.14–10.16 Photographs taken from Linnen *et al.* (2013) and used by permission.

Figure 10.20 Redrawn from Kamberov *et al.* (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**: 691–702. Copyright Elsevier.

Figure 11.2 From Pasqualone and Huffaker (1994). STU1, a suppressor of beta-tubulin mutation, encodes a novel and essential component of the yeast mitotic spindle. *Journal of Cell Biology* **127**: 1973. Copyright The Rockerfeller University Press.

Figure 11.15 From Kidd *et al.* (1999). Slit is the midline repellent for the Robo receptor in *Drosophila*. *Cell* **96**: 785. Copyright Elsevier Inc.

Figure C11.3 From Pasqualone and Huffaker (1994). STU1, a suppressor of beta-tubulin mutation, encodes a novel and essential component of the yeast mitotic spindle. *Journal of Cell Biology* **127**: 1973. Copyright The Rockerfeller University Press.

Figure C11.5 From Richards *et al.* (2000). Structure–function relationships in yeast tubulins. *Molecular Biology of the Cell* **11**:1887. Copyright The American Society for Cell Biology.

Figure C11.7 From Pasqualone and Huffaker (1994). STU1, a suppressor of beta-tubulin mutation, encodes a novel and essential component of the yeast mitotic spindle. *Journal of Cell Biology* **127**: 1973. Copyright The Rockerfeller University Press.

Figure 12.8 Used by permission of Cold Spring Harbor Press and courtesy of Zeynep Altun from Wormatlas.

Figure 13.1 Copyright Elsevier Limited for Cell Press and used by permission.

Figure 13.12 Extracted and redrawn from ModENCODE.org, and The ENCODE Project Consortium (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**: 1787. Copyright AAAS.

Figures 13.13 and 13.14 Extracted and redrawn from ModENCODE.org, and The ENCODE Project Consortium (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**:1787; and Gerstein *et al.* (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**: 1775. Copyright AAAS.

Figure 13.18 Copyright Nature Publishing Group and used by permission.

Figure 13.19 Copyright Nature Publishing Group and used by permission.

Figures 13.24 and 13.25 From Costanzo *et al.* (2010). The genetic landscape of a cell. *Science* **327**: 425–31 and used with permission from AAAS.

Figure 13.26 Redrawn from Chuang *et al.* (2010). A decade of systems biology. *Annual Review of Cell and Development Biology* **26**: 721–44.

Figure 14.11 Photos taken by Mary Hobbs and Philip Meneely.

BRIEF CONTENTS

DETAILED CONTENTS