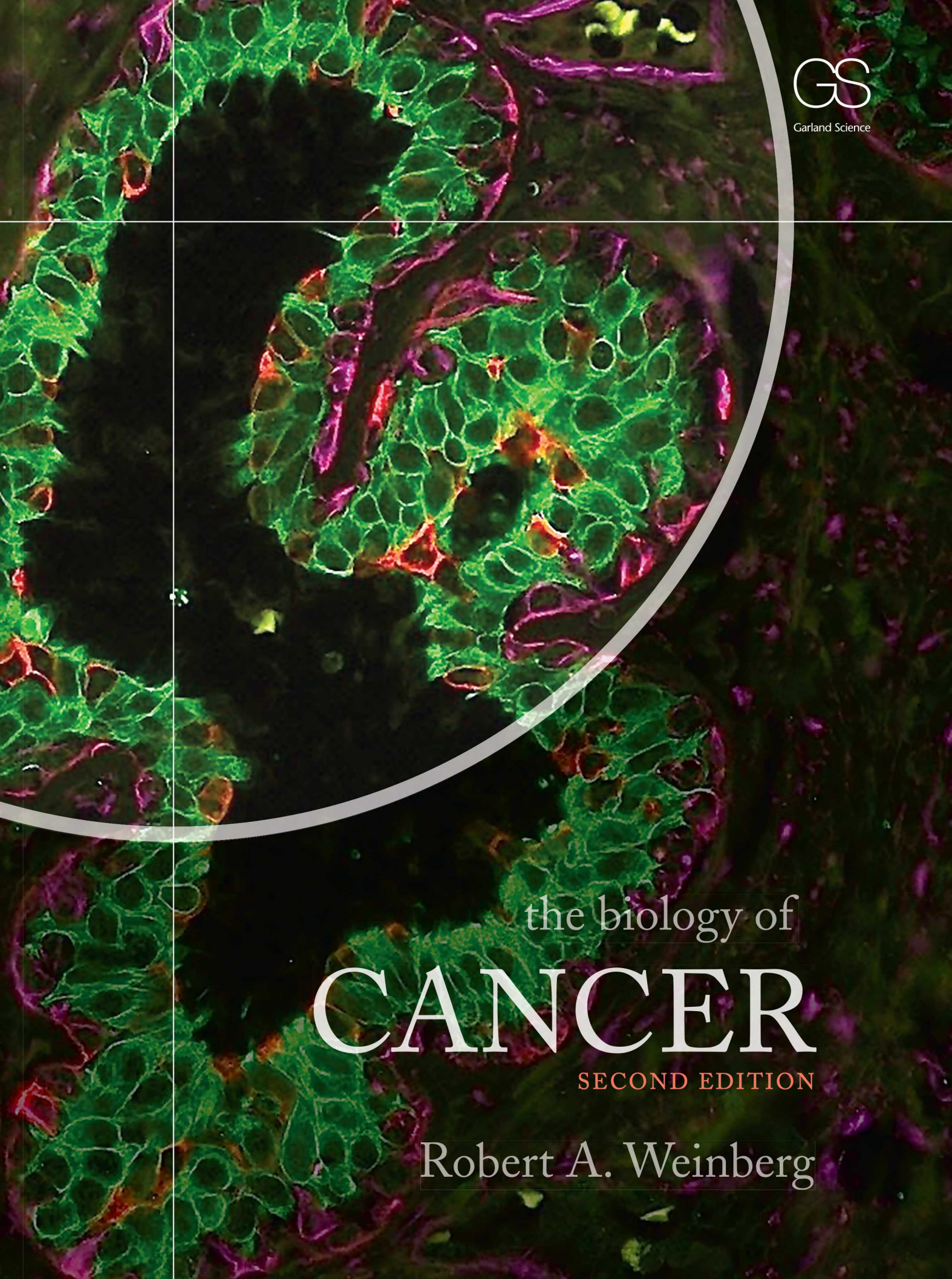

GS
Garland Science
the biology of
CANCER
SECOND EDITION
Robert A. Weinberg

This page intentionally left blank
to match pagination of print book

the biology of CANCER

SECOND EDITION

This page intentionally left blank
to match pagination of print book

the biology of CANCER

SECOND EDITION

Robert A. Weinberg

Garland Science
Taylor & Francis Group
NEW YORK AND LONDON

Garland Science
Vice President: Denise Schanck
Assistant Editor: Allie Bochicchio
Production Editor and Layout: EJ Publishing Services
Text Editor: Elizabeth Zayatz
Copy Editor: Richard K. Mickey
Proofreader: Sally Huish
Illustrator: Nigel Orme
Designer: Matthew McClements, Blink Studio, Ltd.
Permissions Coordinator: Becky Hainz-Baxter
Indexer: Bill Johncocks
Director of Digital Publishing: Michael Morales
Editorial Assistant: Lamia Harik

ISBNs: 978-0-8153-4219-9 (hardcover); 978-0-8153-4220-5 (softcover).

Library of Congress Cataloging-in-Publication Data

Weinberg, Robert A. (Robert Allan), 1942-
The biology of cancer. -- Second edition.
pages cm
Includes bibliographical references.
ISBN 978-0-8153-4219-9 (hardback) -- ISBN 978-0-8153-4220-5 (pbk.) 1. Cancer--Molecular aspects. 2. Cancer--Genetic aspects. 3. Cancer cells. I. Title.
RC268.4.W45 2014
616.99'4--dc23
2013012335

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 711 Third Avenue, New York, NY 10017, USA, and 3 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

Garland Science
Taylor & Francis Group

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

About the Author

Robert A. Weinberg is a founding member of the Whitehead Institute for Biomedical Research. He is the Daniel K. Ludwig Professor for Cancer Research and the American Cancer Society Research Professor at the Massachusetts Institute of Technology (MIT). Dr. Weinberg is an internationally recognized authority on the genetic basis of human cancer and was awarded the U.S. National Medal of Science in 1997.

Front Cover

A micrograph section of a human *in situ* ductal carcinoma with α-smooth muscle actin stained in *pink*, cytokeratins 5 and 6 in *red-orange*, and cytokeratins 8 and 18 in *green*. (Courtesy of Werner Böcker and Igor B. Buchwalow of the Institute for Hematopathology, Hamburg, Germany.)

Dedication

I dedicate this second edition, as the first one, to my dear wife, Amy Shulman Weinberg, who endured long hours of inattention, hearing from me repeatedly the claim that the writing of this edition was almost complete, when in fact years of work lay ahead. She deserved much better! With much love.

This page intentionally left blank
to match pagination of print book

Preface

Compared with other areas of biological research, the science of molecular oncology is a recent arrival; its beginning can be traced with some precision to a milestone discovery in 1975. In that year, the laboratory of Harold Varmus and J. Michael Bishop in San Francisco, California demonstrated that normal cell genomes carry a gene—they called it a proto-oncogene—that has the potential, following alteration, to incite cancer. Before that time, we knew essentially nothing about the molecular mechanisms underlying cancer formation; since that time an abundance of information has accumulated that now reveals in outline and fine detail how normal cells become transformed into tumor cells, and how these neoplastic cells collaborate to form life-threatening tumors.

The scientific literature on cancer pathogenesis has grown explosively and today encompasses millions of research publications. So much information would seem to be a pure blessing. After all, knowing more is always better than knowing less. In truth, it represents an embarrassment of riches. By now, we seem to know too much, making it difficult to conceptualize cancer research as a single coherent body of science rather than a patchwork quilt of discoveries that bear only a vague relationship with one another.

This book is written in a far more positive frame of mind, which holds that this patchwork quilt is indeed a manifestation of a body of science that has some simple, underlying principles that unify these diverse discoveries. Cancer research is indeed a field with conceptual integrity, much like other areas of biomedical research and even sciences like physics and chemistry, and the bewildering diversity of the cancer research literature can indeed be understood through these underlying principles.

Prior to the pioneering findings of 1975, we knew almost nothing about the molecular and cellular mechanisms that create tumors. There were some intriguing clues lying around: We knew that carcinogenic agents often, but not always, operate as mutagens; this suggested that mutant genes are involved in some fashion in programming the abnormal proliferation of cancer cells. We knew that the development of cancer is often a long, protracted process. And we knew that individual cancer cells extracted from tumors behave very differently than their counterparts in normal tissues.

Now, almost four decades later, we understand how mutant genes govern the diverse traits of cancer cells and how the traits of these individual cells determine the behavior of tumors. Many of these advances can be traced to the stunning improvements in experimental tools. The techniques of genetic analysis, which were quite primitive at the beginning of this period, have advanced to the stage where we can sequence entire tumor cell genomes in several days. (This is in sharp contrast to the state of affairs in 1975, when the sequencing of oligonucleotides represented a formidable task!) Given the critical role of genotype in determining phenotype, we now understand, as least in outline, why cancer cells behave the way that they do. On the one hand, the molecular differences among individual cancers suggest hundreds of distinct types of human cancer. On the other, molecular and biochemical analyses reveal that this bewildering diversity really manifests a small number of underlying common biochemical traits and molecular processes.

Amusingly, much of this unification was preordained by decisions made 600 million years ago. Once the laws and mechanisms of organismic development were established, they governed all that followed, including the behavior of both normal and neoplastic cells. Modern cancer researchers continue to benefit from this rigid adherence to the fundamental, evolutionarily conserved rules of life. As is evident repeatedly throughout this book, much of what we understand about cancer cells, and thus about the disease of cancer, has been learned by studying the cells of worms and fruit flies and frogs. These laws and principles are invoked repeatedly to explain the complex behaviors of human tumors. By providing context and perspective, they can be used to help us understand all types of human cancer.

While these basic principles are now in clear view, critical details continue to elude us. This explains why modern cancer research is still in active ferment, and why new, fascinating discoveries are being reported every month. While they create new perspectives, they do not threaten the solidity of the enduring truths, which this book attempts to lay out. These principles were already apparent seven years ago when the first edition of this book appeared and, reassuringly, their credibility has not been undermined by all that has followed.

In part, this book has been written as a recruiting pamphlet, as new generations of researchers are needed to move cancer research forward. They are so important because the lessons about cancer's origins, laid out extensively in this book, have not yet been successfully applied to make major inroads into the prevention and cure of this disease. This represents the major frustration of contemporary cancer research: the lessons of disease causation have rarely been followed, as day follows night, by the development of definitive cures.

And yes, there are still major questions that remain murky and poorly resolved. We still do not understand how cancer cells create the metastases that are responsible for 90% of cancer-associated mortality. We understand rather little of the role of the immune system in preventing cancer development. And while we know much about the individual signaling molecules operating inside individual human cells, we lack a clear understanding of how the complex signaling circuitry formed by these molecules makes the life-and-death decisions that determine the fate of individual cells within our body. Those decisions ultimately determine whether or not one of our cells begins the journey down the long road leading to cancerous proliferation and, finally, to a life-threatening tumor.

Contemporary cancer research has enriched numerous other areas of modern biomedical research. Consequently, much of what you will learn from this book will be useful in understanding many aspects of immunology, neurobiology, developmental biology, and a dozen other biomedical research fields. Enjoy the ride!

Robert A. Weinberg
Cambridge, Massachusetts
March 2013

A Note to the Reader

The second edition of this book is organized, like the first, into 16 chapters of quite different lengths. The conceptual structure that was established in the first edition still seemed to be highly appropriate for the second, and so it was retained. What has changed are the contents of these chapters: some have changed substantially since their first appearance seven years ago, while others—largely early chapters—have changed little. The unchanging nature of the latter is actually reassuring, since these chapters deal with early conceptual foundations of current molecular oncology; it would be most unsettling if these foundational chapters had undergone radical revision, which would indicate that the earlier edition was a castle built on sand, with little that could be embraced as well-established, unchanging certainties.

The chapters are meant to be read in the order that they appear, in that each builds on the ideas that have been presented in the chapters before it. The first chapter is a condensed refresher course for undergraduate biology majors and pre-doctoral students; it lays out many of the background concepts that are assumed in the subsequent chapters.

The driving force of these two editions has been a belief that modern cancer research represents a conceptually coherent field of science that can be presented as a clear, logical progression. Embedded in these discussions is an anticipation that much of this information will one day prove useful in devising novel diagnostic and therapeutic strategies that can be deployed in oncology clinics. Some experiments are described in detail to indicate the logic supporting many of these concepts. You will find numerous schematic drawings, often coupled with micrographs, that will help you to appreciate how experimental results have been assembled, piece-by-piece, generating the syntheses that underlie molecular oncology.

Scattered about the text are "Sidebars," which consist of commentaries that represent detours from the main thrust of the discussion. Often these Sidebars contain anecdotes or elaborate on ideas presented in the main text. Read them if you are interested, or skip over them if you find them too distracting. They are presented to provide additional interest—a bit of extra seasoning in the rich stew of ideas that constitutes contemporary research in this area. The same can be said about the "Supplementary Sidebars," which have been relegated to the DVD-ROM that accompanies this book. These also elaborate upon topics that are laid out in the main text and are cross-referenced throughout the book. Space constraints dictated that the Supplementary Sidebars could not be included in the hardcopy version of the textbook.

Throughout the main text you will find extensive cross-references whenever topics under discussion have been introduced or described elsewhere. Many of these have been inserted in the event that you read the chapters in an order different from their presentation here. These cross-references should not provoke you to continually leaf through other chapters in order to track down cited sections or figures. If you feel that you will benefit from earlier introductions to a topic, use these cross-references; otherwise, ignore them.

Each chapter ends with a forward-looking summary entitled "Synopsis and Prospects." This section synthesizes the main concepts of the chapter and often addresses

ideas that remain matters of contention. It also considers where research might go in the future. This overview is extended by a list of key concepts and a set of questions. Some of the questions are deliberately challenging and we hope they will provoke you to think more deeply about many of the issues and concepts developed. Finally, most chapters have an extensive list of articles from research journals. These will be useful if you wish to explore a particular topic in detail. Almost all of the cited references are review articles, and many contain detailed discussions of various subfields of research as well as recent findings. In addition, there are occasional references to older publications that will clarify how certain lines of research developed.

Perhaps the most important goal of this book is to enable you to move beyond the textbook and jump directly into the primary research literature. This explains why some of the text is directed toward teaching the elaborate, specialized vocabulary of the cancer research literature, and many of its terms are defined in the glossary. Boldface type has been used throughout to highlight key terms that you should understand. Cancer research, like most areas of contemporary biomedical research, is plagued by numerous abbreviations and acronyms that pepper the text of many published reports. The book provides a key to deciphering this alphabet soup by defining these acronyms. You will find a list of such abbreviations in the back.

Also contained in the book is a newly compiled List of Key Techniques. This list will assist you in locating techniques and experimental strategies used in contemporary cancer research.

The DVD-ROM that accompanies the book also contains a PowerPoint® presentation for each chapter, as well as a companion folder that contains individual JPEG files of the book images including figures, tables, and micrographs. In addition, you will find on this disc a variety of media for students and instructors: movies and audio recordings. There is a selection of movies that will aid in understanding some of the processes discussed; these movies are referenced on the first page of the corresponding chapter in a blue box. The movies are available in QuickTime and WMV formats, and can be used on a computer or transferred to a mobile device. The author has also recorded mini-lectures on the following topics for students and instructors: Mutations and the Origin of Cancer, Growth Factors, p53 and Apoptosis, Metastasis, Immunology and Cancer, and Cancer Therapies. These are available in MP3 format and, like the movies, are easy to transfer to other devices. These media items, as well as future media updates, are available to students and instructors at: http://www.garlandscience.com. On the website, qualified instructors will be able to access a newly created Question Bank. The questions are written to test various levels of understanding within each chapter. The instructor's website also offers access to instructional resources from all of the Garland Science textbooks. For access to instructor's resources please contact your Garland Science sales representative or e-mail science@garland.com.

The poster entitled "The Pathways of Human Cancer" summarizes many of the intracellular signaling pathways implicated in tumor development. This poster has been produced and updated for the Second Edition by Cell Signaling Technology.

Because this book describes an area of research in which new and exciting findings are being announced all the time, some of the details and interpretations presented here may become outdated (or, equally likely, proven to be wrong) once this book is in print. Still, the primary concepts presented here will remain, as they rest on solid foundations of experimental results.

The author and the publisher would greatly appreciate your feedback. Every effort has been made to minimize errors. Nonetheless, you may find them here and there, and it would be of great benefit if you took the trouble to communicate them. Even more importantly, much of the science described herein will require reinterpretation in coming years as new discoveries are made. Please email us at science@garland.com with your suggestions, which will be considered for incorporation into future editions.

Acknowledgments

The science described in this book is the opus of a large, highly interactive research community stretching across the globe. Its members have moved forward our understanding of cancer immeasurably over the past generation. The colleagues listed below have helped the author in countless ways, large and small, by providing sound advice, referring me to critical scientific literature, analyzing complex and occasionally contentious scientific issues, and reviewing individual chapters and providing much-appreciated critiques. Their scientific expertise and their insights into pedagogical clarity have proven to be invaluable. Their help extends and complements the help of an equally large roster of colleagues who helped with the preparation of the first edition. These individuals are representatives of a community, whose members are, virtually without exception, ready and pleased to provide a helping hand to those who request it. I am most grateful to them. Not listed below are the many colleagues who generously provided high quality versions of their published images; they are acknowledged through the literature citations in the figure legends. I would like to thank the following for their suggestions in preparing this edition, as well as those who helped with the first edition. (Those who helped on this second edition are listed immediately, while those who helped with the first edition follow.)

Second edition Eric Abbate, Janis Abkowitz, Julian Adams, Peter Adams, Gemma Alderton, Lourdes Aleman, Kari Alitalo, C. David Allis, Claudia Andl, Annika Antonsson, Paula Apsell, Steven Artandi, Carlos Arteaga, Avi Ashkenazi, Duncan Baird, Amy Baldwin, Frances Balkwill, Allan Balmain, David Bartel, Josep Baselga, Stephen Baylin, Philip Beachy, Robert Beckman, Jürgen Behrens, Roderick Beijersbergen, George Bell, Robert Benezra, Thomas Benjamin, Michael Berger, Arnold Berk, René Bernards, Rameen Beroukhim, Donald Berry, Timothy Bestor, Mariann Bienz, Brian Bierie, Leon Bignold, Walter Birchmeier, Oliver Bischof, John Bixby, Jenny Black, Elizabeth Blackburn, Maria Blasco, Matthew Blatnik, Günter Blobel, Julian Blow, Bruce Boman, Gareth Bond, Katherine Borden, Lubor Borsig, Piet Borst, Blaise Bossy, Michael Botchan, Nancy Boudreau, Henry Bourne, Marina Bousquet, Thomas Brabletz, Barbara Brandhuber, Ulrich Brandt, James Brenton, Marta Briarava, Cathrin Brisken, Jacqueline Bromberg, Myles Brown, Patrick Brown, Thijn Brummelkamp, Ferdinando Bruno, Richard Bucala, Janet Butel, Eliezer Calo, Eleanor Cameron, Ian Campbell, Judith Campbell, Judith Campisi, Lewis Cantley, Yihai Cao, Mario Capecchi, Robert Carlson, Peter Carmeliet, Kermit Carraway, Oriol Casanovas, Tom Cech, Howard Cedar, Ann Chambers, Eric Chang, Mark Chao, Iain Cheeseman, Herbert Chen, Jen-Tsan Chi, Lewis Chodosh, Gerhard Christofori, Inhee Chung, Karen Cichowski, Daniela Cimini, Tim Clackson, Lena Claesson-Welsh, Michele Clamp, Trevor Clancy, Rachael Clark, Bayard Clarkson, James Cleaver, Don Cleveland, David Cobrinik, John Coffin, Philip Cohen, Robert Cohen, Michael Cole, Hilary Coller, Kathleen Collins, Duane Compton, John Condeelis, Simon Cook, Christopher Counter, Sara Courtneidge, Lisa Coussens, Charles Craik, James Darnell, Mark Davis, George Daley, Titia de Lange, Pierre De Meyts, Hugues de Thé, Rik Derynck, Mark Dewhirst, James DeCaprio, Mark Depristo, Channing Der, Tom DiCesare, John Dick, Daniel DiMaio, Charles Dimitroff, Nadya Dimitrova, Charles Dinarello, Joseph DiPaolo, Peter Dirks, Vishwa Dixit, Lawrence Donehower, Philip Donoghue, Martin Dorf, David Dornan, Gian Paolo Dotto, Steven Dowdy, James Downing, Harry Drabkin, Brian Druker, Crislyn D'Souza-Schorey, Eric Duell, Patricia Duffner, Michel DuPage, Robert Duronio, Michael Dyer, Nick Dyson, Connie Eaves, Michael Eck, Mikala Egeblad, Charles Eigenbrot, Steve Elledge, Robert Eisenman, Susan Erster, Manel Esteller, Mark Ewen, Patrick Eyers, Doriano Fabbro, Reinhard Fässler, Mark Featherstone, David Felser, Karen Ferrante, Soldano Ferrone, Isaiah Fidler, Barbara Fingleton, Zvi Fishelson, Ignacio Flores, Antonio Foji, David Foster, A. Raymond Frackelton jr., Hervé Wolf Fridman, Peter Friedl, Kenji Fukasawa, Priscilla A. Furth, Vladimir Gabai, Brenda Gallie, Jerome Galon, Sanjiv Sam Gambhir, Levi Garraway, Yan Geng, Bruce Gelb, Richard Gelber, Frank Gertler, Gad Getz, Edward Giovannucci, Michael Gnant, Sumita Gokhale, Leslie Gold, Alfred Goldberg, Richard Goldsby, Jesus Gomez-Navarro, David Gordon, Eyal Gottlieb, Stephen Grant, Alexander Greenhough, Christoph Kahlert, Florian Greten, Jay Grisolano, Athur Grollman, Bernd Groner, Wenjun Guo, Piyush Gupta, Daniel Haber, William Hahn, Kevin Haigis, Marcia Haigis, William Hait, Thanos Halazonetis, John Haley, Stephen Hall, Douglas Hanahan, Steven Hanks, J. Marie Hardwick, Iswar Hariharan, Ed Harlow, Masanori Hatakeyama, Georgia Hatzivassiliou, Lin He, Matthias Hebrok, Stephen Hecht,

Kristian Helin, Samuel Hellman, Michael Hemann, Linda Hendershot, Meenhard Herlyn, Julian Heuberger, Philip Hinds, Susan Hilsenbeck, Michelle Hirsch, Andreas Hochwagen, H. Robert Horvitz, Susan Horwitz, Peter Howley, Ralph Hruban, Peggy Hsu, David Huang, Paul Huang, Robert Huber, Honor Hugo, Tony Hunter, Richard Hynes, Tan Ince, Yoko Irie, Mark Israel, Jean-Pierre Issa, Yoshiaki Ito, Michael Ittmann, Shalev Itzkovitz, Tyler Jacks, Stephen Jackson, Rudolf Jaenisch, Rakesh Jain, Katherine Janeway, Ahmedin Jemal, Harry Jenq, Kim Jensen, Josef Jiricny, Claudio Joazeiro, Bruce Johnson, Candace Johnson, David Jones, Peter Jones, Nik Joshi, Johanna Joyce, William Kaelin, Kong Jie Kah, Nada Kalaany, Raghu Kalluri, Lawrence Kane, Antoine Karnoub, John Katzenellenbogen, Khandan Keyomarsi, Katherine Janeway, William Kaelin jr., Andrius Kazlauskas, Joseph Kelleher, Elliott Kieff, Nicole King, Christian Klein, Pamela Klein, Frederick Koerner, Richard Kolesnick, Anthony Komaroff, Konstantinos Konstantopoulos, Jordan Krall, Igor Kramnik, Wilhelm Krek, Guido Kroemer, Eve Kruger, Genevieve Kruger, Madhu Kumar, Charlotte Kuperwasser, Thomas Kupper, Bruno Kyewski, Sunil Lakhani, Eric Lander, Lewis Lanier, Peter Lansdorp, David Largaespada, Michael Lawrence, Emma Lees, Jacqueline Lees, Robert Lefkowitz, Mark Lemmon, Stanley Lemon, Arnold Levine, Beth Levine, Ronald Levy, Ephrat Levy-Lahad, Kate Liddell, Stuart Linn, Marta Lipinski, Joe Lipsick, Edison Liu, David Livingston, Harvey Lodish, Lawrence Loeb, Jay Loeffler, David Louis, Julie-Aurore Losman, Scott Lowe, Haihui Lu, Kunxin Luo, Mathieu Lupien, Li Ma, Elisabeth Mack, Alexander MacKerell jr., Ben Major, Tak Mak, Shiva Malek, Scott Manalis, Sridhar Mani, Matthias Mann, Alberto Mantovani, Richard Marais, Jean-Christophe Marine, Sanford Markowitz, Ronen Marmorstein, Lawrence Marnett, Chris Marshall, G. Steven Martin, Joan Massagué, Lynn Matrisian, Massimilano Mazzone, Sandra McAllister, Grant McArthur, David McClay, Donald McDonald, David Glenn McFadden, Wallace McKeehan, Margaret McLaughlin-Drubin, Anthony Means, René Medema, Cornelis Melief, Craig Mermel, Marek Michalak, Brian Miller, Nicholas Mitsiades, Sibylle Mittnacht, Holger Moch, Ute Moll, Deborah Morrsion, Aristides Moustakis, Gregory Mundy, Cornelius Murre, Ruth Muschel, Senthil Muthuswamy, Jeffrey Myers, Harikrishna Nakshatri, Inke Näthke, Geoffrey Neale, Ben Neel, Joel Neilson, M. Angela Nieto, Irene Ng, Ingo Nindl, Larry Norton, Roel Nusse, Shuji Ogino, Kenneth Olive, Andre Oliveira, Gilbert Omenn, Tamer Onder, Moshe Oren, Barbara Osborne, Liliana Ossowski, David Page, Klaus Pantel, David Panzarella, William Pao, Jongsun Park, Paul Parren, Ramon Parsons, Dhavalkumar Patel, Mathias Pawlak, Tony Pawson, Daniel Peeper, Mark Peifer, David Pellman, Tim Perera, Charles Perou, Mary Ellen Perry, Manuel Perucho, Richard Pestell, Julian Peto, Richard Peto, Stefano Piccolo, Jackie Pierce, Eli Pikarsky, Hidde Ploegh, Nikolaus Pfanner, Kristy Pluchino, Heike Pohla, Paul Polakis, Michael Pollak, John Potter, Carol Prives, Lajos Pusztai, Xuebin Qin, Priyamvada Rai, Terence Rabbitts, Anjana Rao, Julia Rastelli, David Raulet, John Rebers, Roger Reddel, Peter Reddien, Danny Reinberg, Michael Retsky, Jeremy Rich, Andrea Richardson, Tim Richmond, Gail Risbridger, Paul Robbins, James Roberts, Leonardo Rodriguez, Veronica Rodriguez, Mark Rolfe, Michael Rosenblatt, David Rosenthal, Theodora Ross, Yolanda Roth, David Rowitch, Brigitte Royer-Pokora, Anil Rustgi, David Sabatini, Erik Sahai, Jesse Salk, Leona Samson, Yardena Samuels, Bengt Samuelsson, Christopher Sansam, Richard Santen, Van Savage, Andrew Sharrocks, Brian Schaffhausen, Pepper Schedin, Christina Scheel, Rachel Schiff, Joseph Schlessinger, Ulrich Schopfer, Hubert Schorle, Deborah Schrag, Brenda Schulman, Wolfgang Schulz, Bert Schutte, Hans Schreiber, Robert Schreiber, Martin Schwartz, Ralph Scully, John Sedivy, Helmut Seitz, Manuel Serrano, Jeffrey Settleman, Kevin Shannon, Phillip Sharp, Norman Sharpless, Jerry Shay, Stephen Sherwin, Yigong Shi, Tsukasa Shibuye, Ben-Zion Shilo, Piotr Sicinski, Daniel Silver, Arun Singh, Michail Sitkovsky, George Sledge, Jr., Mark Sliwkowski, David I. Smith, Eric Snyder, Pierre Sonveaux, Jean-Charles Soria, Ben Stanger, Sheila Stewart, Charles Stiles, Jayne Stommel, Shannon Stott, Jenny Stow, Michael Stratton, Ravid Straussman, Jonathan Strosberg, Charles Streuli, Herman Suit, Peter Sun, Thomas Sutter, Kathy Svoboda, Alejandro Sweet-Cordero, Mario Sznol, Clifford Tabin, Wai Leong Tam, Hsin-Hsiung Tai, Makoto Taketo, Wai Leong Tam, Filemon Tan, Michael Tangrea, Masae Tatematsu, Steven Teitelbaum, Sabine Tejpar, Adam Telerman, Jennifer Temel, David Tenenbaum, Mine Tezal, Jean Paul Thiery, Craig Thompson, Michael Thun, Thea Tlsty, Rune Toftgård, Nicholas Tonks, James Trager, Donald L. Trump, Scott Valastyan, Linda van Aelst, Benoit van den Eynde, Matthew Vander Heiden, Maarten van Lohuizen, Eugene van Scott, Peter Vaupel, Laura van't Veer, George Vassiliou, Inder Verma, Gabriel Victora, Christoph Viebahn, Danijela Vignjevic, Bert Vogelstein, Robert Vonderheide, Daniel von Hoff, Dorien Voskuil, Karen Vousden, Geoffrey Wahl, Lynne Waldman, Herbert Waldmann, Graham Walker, Rongfu Wang, Patricia Watson, Bill Weis, Stephen Weiss, Irv Weissman, Danny Welch, H. Gilbert Welch, Zena Werb, Marius Wernig, Bengt Westermark, John Westwick, Eileen White, Forest White, Max Wicha, Walter Willett, Catherine Wilson, Owen Witte, Alfred Wittinghofer, Norman Wolmark, Sopit Wongkham, Richard Wood, Nicholas Wright, Xu Wu, David Wynford-Thomas, Michael Yaffe, Jing Yang, James Yao, Yosef Yarden, Robert Yauch, Xin Ye, Sam Yoon, Richard Youle, Richard Young, Patrick Zarrinkar, Ann Zauber, Jiri Zavadil, Lin Zhang, Alicia Zhou, Ulrike Ziebold, Kai Zinn, Johannes Zuber, James Zwiebel.

Special thanks to **Makoto Mark Taketo** of Kyoto University and **Richard A. Goldsby** of Amherst College.

First edition Joan Abbott, Eike-Gert Achilles, Jerry Adams, Kari Alitalo, James Allison, David Alpers, Fred Alt, Carl Anderson, Andrew Aprikyan, Jon Aster, Laura Attardi, Frank Austen, Joseph Avruch, Sunil Badve, William Baird, Frances Balkwill, Allan Balmain, Alan Barge, J. Carl Barrett, David Bartel, Renato Baserga, Richard Bates, Philip Beachy, Camille Bedrosian, Anna Belkina, Robert Benezra, Thomas Benjamin, Yinon Ben-Neriah, Ittai Ben-Porath, Bradford Berk, René Bernards, Anton Berns, Kenneth Berns, Monica Bessler, Neil Bhowmick, Marianne Bienz, Line Bjørge, Harald von Boehmer, Gareth Bond, Thierry Boon, Dorin-Bogdan Borza, Chris Boshoff, Noël Bouck, Thomas Brabletz, Douglas Brash, Cathrin Brisken, Garrett Brodeur, Patrick Brown, Richard Bucala, Patricia Buffler, Tony Burgess, Suzanne Bursaux, Randall Burt, Stephen Bustin, Janet Butel, Lisa Butterfield, Blake Cady, John Cairns, Judith Campisi, Harvey Cantor, Robert Cardiff, Peter Carroll, Arlindo Castelanho, Bruce Chabner, Ann Chambers, Howard Chang, Andrew Chess, Ann Cheung, Lynda Chin, Francis Chisari, Yunje Cho, Margaret Chou, Karen Cichowski, Michael Clarke, Hans Clevers, Brent Cochran, Robert Coffey, John Coffin, Samuel Cohen, Graham Colditz, Kathleen Collins, Dave Comb, John Condeelis, Suzanne Cory, Christopher Counter, Sara Courtneidge, Sandra Cowan-Jacob, John Crispino, John Crissman, Carlo Croce, Tim Crook, Christopher Crum, Marcia Cruz-Correa, Gerald Cunha, George Daley, Riccardo Dalla-Favera, Alan D'Andrea, Chi Dang, Douglas Daniels, James Darnell, Jr., Robert Darnell, Galina Deichman, Titia de Lange, Hugues de Thé, Chuxia Deng, Edward Dennis, Lucas Dennis, Ronald DePinho, Theodora Devereaux, Tom DiCesare, Jules Dienstag, John DiGiovanni, Peter Dirks, Ethan Dmitrovsky, Daniel Donoghue, John Doorbar, G. Paolo Dotto, William Dove, Julian Downward, Glenn Dranoff, Thaddeus Dryja, Raymond DuBois, Nick Duesbery, Michel DuPage, Harold Dvorak, Nicholas Dyson, Michael Eck, Walter Eckhart, Argiris Efstratiadis, Robert Eisenman, Klaus Elenius, Steven Elledge, Elissa Epel, John Eppig, Raymond Erikson, James Eshleman, John Essigmann, Gerard Evan, Mark Ewen, Guowei Fang, Juli Feigon, Andrew Feinberg, Stephan Feller, Bruce Fenton, Stephen Fesik, Isaiah Fidler, Gerald Fink, Alain Fischer, Zvi Fishelson, David Fisher, Richard Fisher, Richard Flavell, Riccardo Fodde, M. Judah Folkman, David Foster, Uta Francke, Emil Frei, Errol Friedberg, Peter Friedl, Stephen Friend, Jonas Frisen, Elaine Fuchs, Margaret Fuller, Yuen Kai (Teddy) Fung, Kyle Furge, Amar Gajjar, Joseph Gall, Donald Ganem, Judy Garber, Frank Gertler, Charlene Gilbert, Richard Gilbertson, Robert Gillies, Doron Ginsberg, Edward Giovannucci, Inna Gitelman, Steve Goff, Lois Gold, Alfred Goldberg, Mitchell Goldfarb, Richard Goldsby, Joseph Goldstein, Susanne Gollin, Mehra Golshan, Todd Golub, Jeffrey Gordon, Michael Gordon, Siamon Gordon, Martin Gorovsky, Arko Gorter, Joe Gray, Douglas Green, Yoram Groner, John Groopman, Steven Grossman, Wei Gu, David Guertin, Piyush Gupta, Barry Gusterson, Daniel Haber, James Haber, William Hahn, Kevin Haigis, Senitiroh Hakomori, Alan Hall, Dina Gould Halme, Douglas Hanahan, Philip Hanawalt, Adrian Harris, Curtis Harris, Lyndsay Harris, Stephen Harrison, Kimberly Hartwell, Leland Hartwell, Harald zur Hausen, Carol Heckman, Ruth Heimann, Samuel Hellman, Brian Hemmings, Lothar Hennighausen, Meenhard Herlyn, Glenn Herrick, Avram Hershko, Douglas Heuman, Richard Hodes, Jan Hoeijmakers, Robert Hoffman, Robert Hoover, David Hopwood, Gabriel Hortobagyi, H. Robert Horvitz, Marshall Horwitz, Alan Houghton, Peter Howley, Robert Huber, Tim Hunt, Tony Hunter, Stephen Hursting, Nancy Hynes, Richard Hynes, Antonio Iavarone, J. Dirk Iglehart, Tan Ince, Max Ingman, Mark Israel, Kurt Isselbacher, Tyler Jacks, Rudolf Jaenisch, Rakesh Jain, Bruce Johnson, David Jones, Richard Jones, William Kaelin, Jr., Raghu Kalluri, Alexander Kamb, Barton Kamen, Manolis Kamvysselis, Yibin Kang, Philip Kantoff, Paul Kantrowitz, Jan Karlsreder, Michael Kastan, Michael Kauffman, William Kaufmann, Robert Kerbel, Scott Kern, Khandan Keyomarsi, Marc Kirschner, Christoph Klein, George Klein, Yoel Kloog, Alfred Knudson, Frederick Koerner, Anthony Komaroff, Kenneth Korach, Alan Korman, Eva Kramarova, Jackie Kraveka, Wilhelm Krek, Charlotte Kuperwasser, James Kyranos, Carole LaBonne, Peter Laird, Sergio Lamprecht, Eric Lander, Laura Landweber, Lewis Lanier, Andrew Lassar, Robert Latek, Lester Lau, Derek Le Roith, Chung Lee, Keng Boon Lee, Richard Lee, Jacqueline Lees, Rudolf Leibel, Mark Lemmon, Christoph Lengauer, Jack Lenz, Gabriel Leung, Arnold Levine, Beth Levine, Jay Levy, Ronald Levy, Fran Lewitter, Frederick Li, Siming Li, Frank Lieberman, Elaine Lin, Joachim Lingner, Martin Lipkin, Joe Lipsick, David Livingston, Harvey Lodish, Lawrence Loeb, Edward Loechler, Michael Lotze, Lawrence Lum, Vicky Lundblad, David MacPherson, Sendurai Mani, Alberto Mantovani, Sandy Markowitz, Larry Marnett, G. Steven Martin, Seamus Martin, Joan Massagué, Patrice Mathevet, Paul Matsudaira, Andrea McClatchey, Frank McCormick, Patricia McManus, Mark McMenamin, U. Thomas Meier, Matthew Meyerson, George Miller, Nathan Miselis, Randall Moon, David Morgan, Rebecca Morris, Simon Conway Morris, Robert Moschel, Bernard Moss, Paul Mueller, Anja Mueller-Homey, William A. Muller, Gregory Mundy, Karl Münger, Lance Munn, Ruth Muschel, Lee Nadler, David G. Nathan, Jeremy Nathans, Sergei Nedospasov, Benjamin Neel, David Neuhaus, Donald Newmeyer, Leonard Norkin, Lloyd Old, Kenneth Olive, Tamer Onder, Moshe Oren, Terry Orr-Weaver, Barbara Osborne, Michele Pagano, David Page, Asit Parikh, Chris Parker, William Paul, Amanda Paulovich, Tony Pawson, Mark Peifer, David Pellman, David Phillips, Jacqueline Pierce, Malcolm Pike, John Pintar, Maricarmen Planas-Silva, Roland Pochet, Daniel Podolsky, Beatriz Pogo, Roberto Polakiewicz, Jeffrey Pollard, Nicolae Popescu, Christoph Poremba, Richmond Prehn, Carol Prives, Vito Quaranta, Peter Rabinovitch, Al Rabson, Priyamvada Rai, Klaus Rajewsky, Sridhar Ramaswamy, Anapoorni Rangarajan, Jeffrey Ravetch, Ilaria Rebay, John Reed, Steven Reed, Alan Rein, Ee Chee Ren, Elizabeth Repasky, Jeremy Rich, Andrea Richardson, Dave Richardson, Darrell Rigel, James Roberts, Diane Rodi, Clifford Rosen, Jeffrey Rosen, Neal Rosen, Naomi Rosenberg, Michael Rosenblatt, Theodora Ross, Martine Roussel, Steve Rozen, Jeffrey Ruben, José Russo, David Sabatini, Julien Sage, Ronit Sarid, Edward Sausville, Charles Sawyers, David Scadden, David Schatz, Christina Scheel, Joseph Schlessinger, Anja Schmidt, Stuart Schnitt, Robert Schoen, Robert Schreiber, Edward Scolnick, Ralph Scully, Harold

Seifried, William Sessa, Jeffrey Settleman, Fergus Shanahan, Jerry Shay, James Sherley, Charles Sherr, Ethan Shevach, Chiaho Shih, Frank Sicheri, Peter Sicinski, Sandy Simon, Dinah Singer, Arthur Skarin, Jonathan Skipper, Judy Small, Gilbert Smith, Lauren Sompayrac, Holger Sondermann, Gail Sonenshein, Deborah Spector, Michael Sporn, Eric Stanbridge, E. Richard Stanley, Louis Staudt, Philipp Steiner, Ralph Steinman, Gunther Stent, Sheila Stewart, Charles Stiles, Jonathan Stoye, Michael Stratton, Bill Sugden, Takashi Sugimura, John Sullivan, Nevin Summers, Calum Sutherland, Clifford Tabin, John Tainer, Jussi Taipale, Shinichiro Takahashi, Martin Tallman, Steven Tannenbaum, Susan Taylor, Margaret Tempero, Masaaki Terada, Satvir Tevethia, Jean Paul Thiery, William Thilly, David Thorley-Lawson, Jay Tischfield, Robertus Tollenaar, Stephen Tomlinson, Dimitrios Trichopoulos, Elaine Trujillo, James Umen, Alex van der Eb, Wim van Egmond, Diana van Heemst, Laura van't Veer, Harold Varmus, Alexander Varshavsky, Anna Velcich, Ashok Venkitaraman, Björn Vennström, Inder Verma, Shelia Violette, Bert Vogelstein, Peter Vogt, Olga Volpert, Evan Vosburgh, Geoffrey Wahl, Graham Walker, Gernot Walter, Jack Wands, Elizabeth Ward, Jonathan Warner, Randolph Watnick, I. Bernard Weinstein, Robin Weiss, Irving Weissman, Danny Welch, H. Gilbert Welch, Zena Werb, Forest White, Michael White, Raymond White, Max Wicha, Walter Willet, Owen Witte, Richard Wood, Andrew Wyllie, John Wysolmerski, Michael Yaffe, Yukiko Yamashita, George Yancopoulos, Jing Yang, Moshe Yaniv, Chun-Nan Yeh, Richard Youle, Richard Young, Stuart Yuspa, Claudio Zanon, David Zaridze, Patrick Zarrinkar, Bruce Zetter, Drazen Zimonjic, Leonard Zon, Weiping Zou

Readers: Through their careful reading of the text, these graduate students provided extraordinarily useful feedback in improving many sections of this book and in clarifying sections that were, in their original versions, poorly written and confusing.

Jamie Weyandt (Duke University), Matthew Crowe (Duke University), Venice Calinisan Chiueh (University of California, Berkeley), Yvette Soignier (University of California, Berkeley)

Question Bank: Jamie Weyandt also produced the accompanying question bank available to qualified adopters on the instructor resource site.

Whitehead Institute/MIT: Christine Hickey was responsible over several years' time in helping to organize the extensive files that constituted each chapter. Her help was truly extraordinary.

Dave Richardson of the Whitehead Institute library helped on countless occasions to retrieve papers from obscure corners of the vast scientific literature, doing so with lightning speed!

Garland: While this book has a single recognized author, it really is the work of many hands. The prose was edited by Elizabeth Zayatz and Richard K. Mickey, two editors who are nothing less than superb. To the extent that this book is clear and readable, much of this is a reflection of their dedication to clarity, precision of language, graceful syntax, and the use of images that truly serve to enlighten rather than confound. I have been most fortunate to have two such extraordinary people looking over my shoulder at every step of the writing process. And, to be sure, I have learned much from them. I cannot praise them enough!

Many of the figures are the work of Nigel Orme, an illustrator of great talent, whose sense of design and dedication to precision and detail are, once again, nothing less than extraordinary.

Garland Science determined the structure and design and provided unfaltering support and encouragement through every step of the process that was required to bring this project to fruition. Denise Schanck gave guidance and cheered me on every step of the way. Unfailingly gracious, she is, in every sense, a superb publisher, whose instincts for design and standards of quality publishing are a model. All textbook authors should be as fortunate as I have been to have someone of her qualities at the helm!

The editorial and logistical support required to organize and assemble a book of this complexity was provided first by Janete Scobie and then over a longer period by Allie Bochicchio, both of whom are multitalented and exemplars of ever-cheerful competence, thoroughness, and helpfulness. Without the organizational skills of these two in the Garland office, this text would have emerged as an incoherent jumble.

The truly Herculean task of procuring permissions for publication of the myriad figures fell on the shoulders of Becky Hainz-Baxter. This remains a daunting task, even in this age of Internet and email. Without her help, it would have been impossible to share with the reader many of the images that have created the field of modern cancer research.

The layout is a tribute to the talents of Emma Jeffcock, once again an exemplar of competence, who has an unerring instinct for how to make images and the pages that hold them accessible and welcoming to the reader; she also provided much-valued editorial help that resulted in many improvements of the prose.

The electronic media associated with this book are the work of Michael Morales, whose ability to organize clear and effective visual presentations are indicated by the electronic files that are carried in the accompanying DVD-ROM. He and his editorial assistant, Lamia Harik, are recognized and thanked for their dedication to detail, thoroughness, and their great talent in providing accessible images that inform the reader and complement the written text.

Additional, highly valuable input into the organization and design were provided by Adam Sendroff, Alain Mentha, and Lucy Brodie.

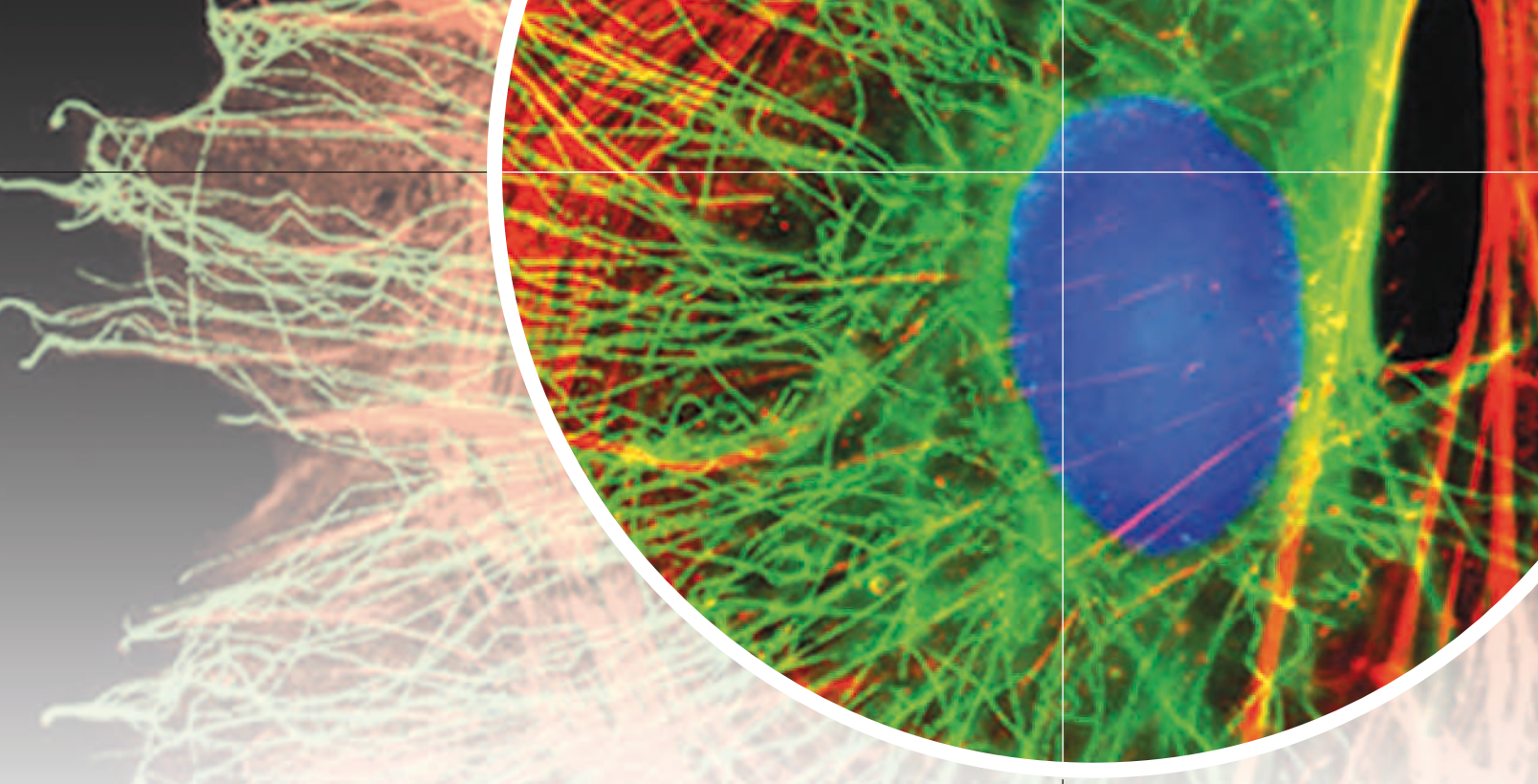
Together, the Garland team, as cited above, represents a unique collection of gifted people whose respective talents are truly peerless and, to say so a second time, individuals who are unfailingly gracious and helpful. Other textbook authors should be as fortunate as I have been in receiving the support that I have enjoyed in the preparation of this second edition!

Contents

List of Key Techniques

 Can be found on the DVD-ROM accompanying the book.

Detailed Contents



Chapter 1

The Biology and Genetics of Cells and Organisms

> Protoplasm, simple or nucleated, is the formal basis of all life... Thus it becomes clear that all living powers are cognate, and that all living forms are fundamentally of one character. The researches of the chemist have revealed a no less striking uniformity of material composition in living matter.
>
> Thomas Henry Huxley, evolutionary biologist, 1868

> Anything found to be true of *E. coli* must also be true of elephants.
>
> Jacques Monod, pioneer molecular biologist, 1954

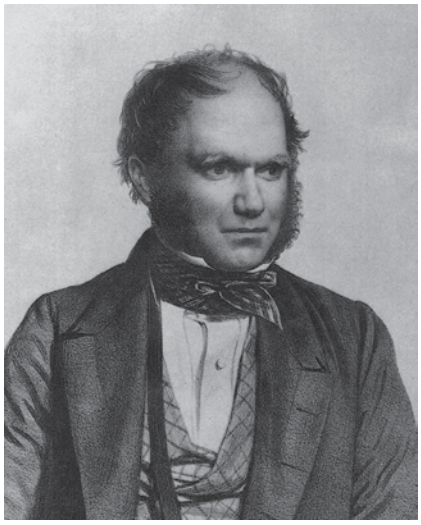
The biological revolution of the twentieth century totally reshaped all fields of biomedical study, cancer research being only one of them. The fruits of this revolution were revelations of both the outlines and the minute details of genetics and heredity, of how cells grow and divide, how they assemble to form tissues, and how the tissues develop under the control of specific genes. Everything that follows in this text draws directly or indirectly on this new knowledge.

This revolution, which began in mid-century and was triggered by Watson and Crick's discovery of the DNA double helix, continues to this day. Indeed, we are still too close to this breakthrough to properly understand its true importance and its long-term ramifications. The discipline of molecular biology, which grew from this discovery, delivered solutions to the most profound problem of twentieth-century biology—how does the genetic constitution of a cell or organism determine its appearance and function?

Without this molecular foundation, modern cancer research, like many other biological disciplines, would have remained a descriptive science that cataloged diverse biological phenomena without being able to explain the mechanics of how they occur.

Movies in this chapter

1.1 Replication I
1.2 Replication II
1.3 Translation I
1.4 Transcription


(A)


(B)

Figure 1.1 Darwin and Mendel
(A) Charles Darwin's 1859 publication of *On the Origin of Species by Means of Natural Selection* exerted a profound effect on thinking about the origin of life, the evolution of organismic complexity, and the relatedness of species. (B) Darwin's theory of evolution lacked a genetic rationale until the work of Gregor Mendel. The synthesis of Darwinian evolution and Mendelian genetics is the foundation for much of modern biological thinking. (A, from the Grace K. Babson Collection, the Henry E. Huntington Library, San Marino, California. Reproduced by permission of The Huntington Library, San Marino, California. B, courtesy of the Mendelianum Museum Moraviae, Brno, Czech Republic.)

Today, our understanding of how cancers arise is being continually enriched by discoveries in diverse fields of biological research, most of which draw on the sciences of molecular biology and genetics. Perhaps unexpectedly, many of our insights into the origins of malignant disease are not coming from the laboratory benches of cancer researchers. Instead, the study of diverse organisms, ranging from yeast to worms to flies, provides us with much of the intellectual capital that fuels the forward thrust of the rapidly moving field of cancer research.

Those who fired up this biological revolution stood on the shoulders of nineteenth-century giants, specifically, Darwin and Mendel (**Figure 1.1**). Without the concepts established by these two, which influence all aspects of modern biological thinking, molecular biology and contemporary cancer research would be inconceivable. So, throughout this chapter, we frequently make reference to evolutionary processes as proposed by Charles Darwin and genetic systems as conceived by Gregor Mendel.

1.1 Mendel establishes the basic rules of genetics

Many of the basic rules of genetics that govern how genes are passed from one complex organism to the next were discovered in the 1860s by Gregor Mendel and have come to us basically unchanged. Mendel's work, which tracked the breeding of pea plants, was soon forgotten, only to be rediscovered independently by three researchers in 1900. During the decade that followed, it became clear that these rules—we now call them Mendelian genetics—apply to virtually all sexual organisms, including **metazoa** (multicellular animals), as well as **metaphyta** (multicellular plants).

Mendel's most fundamental insight came from his realization that genetic information is passed in particulate form from an organism to its offspring. This implied that the entire repertoire of an organism's genetic information—its genome, in today's terminology—is organized as a collection of discrete, separable information packets, now called genes. Only in recent years have we begun to know with any precision how many distinct genes are present in the genomes of mammals; many current analyses of the human genome—the best studied of these—place the number in the range of 21,000, somewhat more than the 14,500 genes identified in the genome of the fruit fly, *Drosophila melanogaster*.

Mendel's work also implied that the constitution of an organism, including its physical and chemical makeup, could be divided into a series of discrete, separable entities. Mendel went further by showing that distinct anatomical parts are controlled by distinct genes. He found that the heritable material controlling the smoothness of peas behaved independently of the material governing plant height or flower color. In

	Seed shape	Seed color	Flower color	Flower position	Pod shape	Pod color	Plant height
One form of trait (dominant)	round	yellow	violet-red	axial	inflated	green	tall
A second form of trait (recessive)	wrinkled	green	white	terminal	pinched	yellow	short

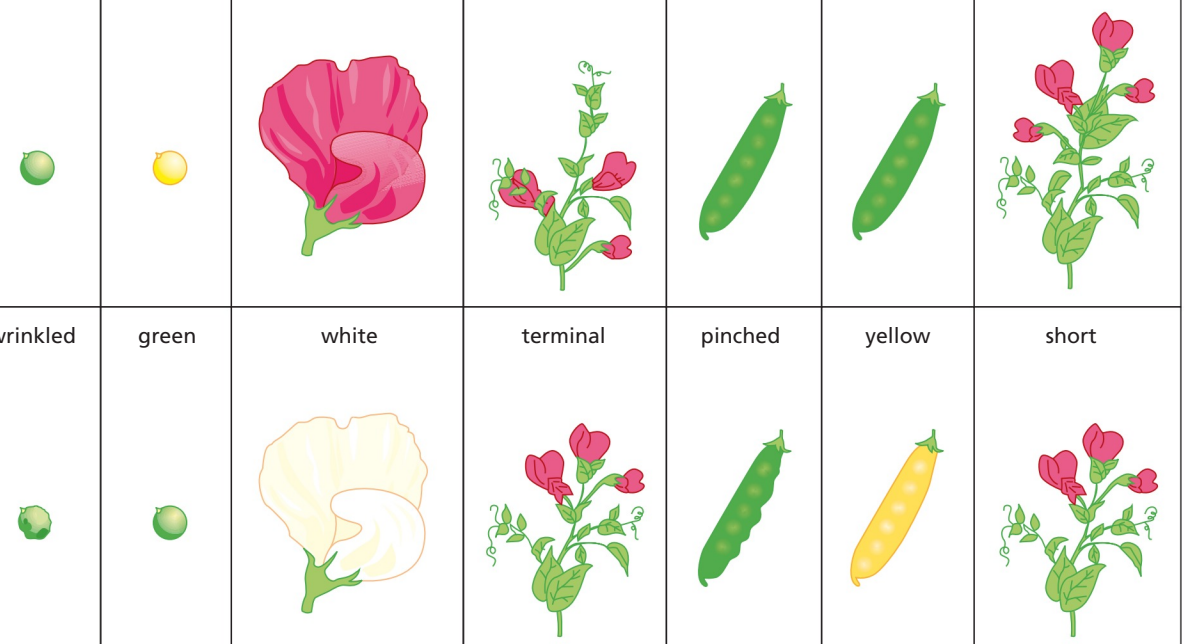


Figure 1.2 A particulate theory of inheritance One of Gregor Mendel's principal insights was that the genetic content of an organism consists of discrete parcels of information, each responsible for a distinct observable trait. Shown are the seven pea-plant traits that Mendel studied through breeding experiments. Each trait had two observable (phenotypic) manifestations, which we now know to be specified by the alternative versions of genes that we call alleles. When the two alternative alleles coexisted within a single plant, the "dominant" trait (*above*) was always observed while the "recessive" trait (*below*) was never observed. (Courtesy of J. Postlethwait and J. Hopson.)

effect, each observable trait of an individual might be traceable to a separate gene that served as its blueprint. Thus, Mendel's research implied that the genetic constitution of an organism (its **genotype**) could be divided into hundreds, perhaps thousands of discrete information packets; in parallel, its observable, outward appearance (its **phenotype**) could be subdivided into a large number of discrete physical or chemical traits (**Figure 1.2**).

Mendel's thinking launched a century-long research project among geneticists, who applied his principles to studying thousands of traits in a variety of experimental animals, including flies (*Drosophila melanogaster*), worms (*Caenorhabditis elegans*), and mice (*Mus musculus*). In the mid-twentieth century, geneticists also began to apply Mendelian principles to study the genetic behavior of single-celled organisms, such as the bacterium *Escherichia coli* and baker's yeast, *Saccharomyces cerevisiae*. The principle of genotype governing phenotype was directly transferable to these simpler organisms and their genetic systems.

While Mendelian genetics represents the foundation of contemporary genetics, it has been adapted and extended in myriad ways since its embodiments of 1865 and 1900. For example, the fact that single-celled organisms often reproduce asexually, that is, without mating, created the need for adaptations of Mendel's original rules. Moreover, the notion that each attribute of an organism could be traced to instructions carried in a single gene was realized to be simplistic. The great majority of observable traits of an organism are traceable to the cooperative interactions of a number of genes. Conversely, almost all the genes carried in the genome of a complex organism play roles in the development and maintenance of multiple organs, tissues, and physiologic processes.

Mendelian genetics revealed for the first time that genetic information is carried redundantly in the genomes of complex plants and animals. Mendel deduced that there were two copies of a gene for flower color and two for pea shape. Today we know that this twofold redundancy applies to the entire genome with the exception of the genes carried in the sex chromosomes. Hence, the genomes of higher organisms are termed **diploid**.

Mendel's observations also indicated that the two copies of a gene could convey different, possibly conflicting information. Thus, one gene copy might specify rough-surfaced and the other smooth-surfaced peas. In the twentieth century, these different versions of a gene came to be called **alleles**. An organism may carry two identical alleles of a gene, in which case, with respect to this gene, it is said to be **homozygous**. Conversely, the presence of two different alleles of a gene in an organism's genome renders this organism **heterozygous** with respect to this gene.

Because the two alleles of a gene may carry conflicting instructions, our views of how genotype determines phenotype become more complicated. Mendel found that in many instances, the voice of one allele may dominate over that of the other in deciding the ultimate appearance of a trait. For example, a pea genome may be heterozygous for the gene that determines the shape of peas, carrying one round and one wrinkled allele. However, the pea plant carrying this pair of alleles will invariably produce round peas. This indicates that the round allele is **dominant**, and that it will invariably overrule its **recessive** counterpart allele (wrinkled) in determining phenotype (see Figure 1.2). (Strictly speaking, using proper genetic parlance, we would say that the phenotype encoded by one allele of a gene is dominant with respect to the phenotype encoded by another allele, the latter phenotype being recessive.)

In fact, classifying alleles as being either dominant or recessive oversimplifies biological realities. The alleles of some genes may be **co-dominant**, in that an expressed phenotype may represent a blend of the actions of the two alleles. Equally common are examples of **incomplete penetrance**, in which case a dominant allele may be present but its phenotype is not manifested because of the actions of other genes within the organism's genome. Therefore, the dominance of an allele is gauged by its interactions with other allelic versions of its gene, rather than its ability to dictate phenotype.

With such distinctions in mind, we note that the development of tumors also provides us with examples of dominance and recessiveness. For instance, one class of alleles that predispose cells to develop cancer encode defective versions of enzymes involved in DNA repair and thus in the maintenance of genomic integrity (discussed again in Chapter 12). These defective alleles are relatively rare in the general population and function recessively. Consequently, their presence in the genomes of many **heterozygotes** (of a wild-type/mutant genotype) is not apparent. However, two heterozygotes carrying recessive defective alleles of the same DNA repair gene may mate. One-fourth of the offspring of such mating pairs, on average, will inherit two defective alleles, exhibit a specific DNA repair defect in their cells, and develop certain types of cancer at greatly increased rates (**Figure 1.3**).

1.2 Mendelian genetics helps to explain Darwinian evolution

In the early twentieth century, it was not apparent how the distinct allelic versions of a gene arise. At first, this variability in information content seemed to have been present in the collective gene pool of a species from its earliest evolutionary beginnings. This perception changed only later, beginning in the 1920s and 1930s, when it became apparent that genetic information is corruptible; the information content in genetic texts, like that in all texts, can be altered. **Mutations** were found to be responsible for changing the information content of a gene, thereby converting one allele into another or creating a new allele from one previously widespread within a species. An allele that is present in the great majority of individuals within a species is usually termed **wild type**, the term implying that such an allele, being naturally present in large numbers of apparently healthy organisms, is compatible with normal structure and function.

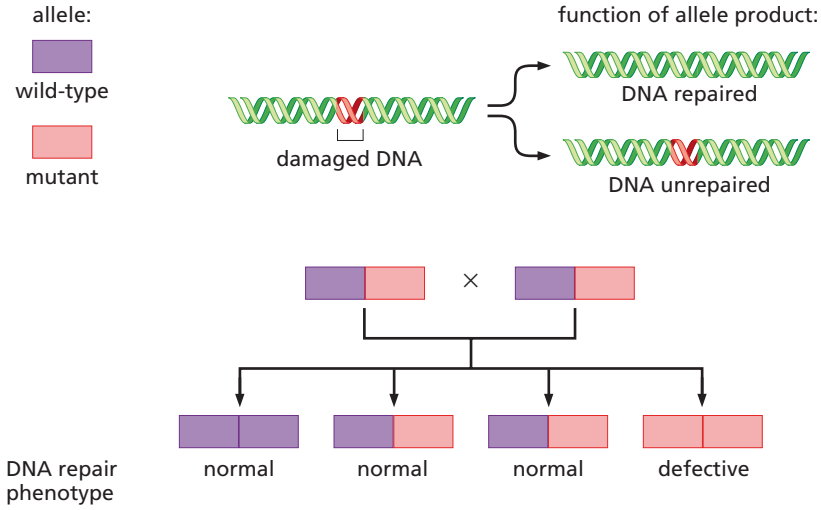


Figure 1.3 Discrepancy between genotype and phenotype The phenotype of an individual often does not indicate genotype. For example, individuals who are phenotypically normal for a trait may nevertheless, at the level of genotype, carry one wild-type (normal) and one mutant (defective) allele of the gene that specifies this trait; this mutant allele will be recessive to the wild-type allele, the latter being dominant. Such individuals are heterozygotes with respect to this gene. In the example shown here, two individuals mate, both of whom are phenotypically normal but heterozygous for a gene specifying a DNA repair function. On average, of their four children, three will be phenotypically normal and their cells will exhibit normal DNA repair function: one of these children will receive two wild-type alleles (be a homozygote) and two will be heterozygotes like their parents. A fourth child, however, will receive two mutant alleles (i.e., be a homozygote) and will be phenotypically mutant, in that this child's cells will lack the DNA repair function specified by this gene. Individuals whose cells lack proper DNA repair function are often cancer-prone, as described in Chapter 12.

Mutations alter genomes continually throughout the evolutionary life span of a species, which usually extends over millions of years. They strike the genome and its constituent genes randomly. Mutations provide a species with a method for continually tinkering with its genome, for trying out new versions of genes that offer the prospect of novel, possibly improved phenotypes. The result of the continuing mutations on the genome is a progressive increase during the evolutionary history of a species in the genetic diversity of its members. Thus, the collection of alleles present in the genomes of all members of a species—the **gene pool** of this species—becomes progressively more heterogeneous as the species grows older.

This means that older species carry more distinct alleles in their genomes than younger ones. Humans, belonging to a relatively young species (<150,000 years old), have one-third as many alleles and genetic diversity as chimpanzees, allowing us to infer that they have been around as a species three times longer than we have.

The continuing diversification of alleles in a species' genome, occurring over millions of years, is countered to some extent by the forces of natural selection that Charles Darwin first described. Some alleles of a gene may confer more advantageous phenotypes than others, so individuals carrying these alleles have a greater probability of leaving numerous descendants than do those members of the same species that lack them. Consequently, natural selection results in a continual discarding of many of the alleles that have been generated by random mutations. In the long run, all things being equal, disadvantageous alleles are lost from the pool of alleles carried by the members of a species, advantageous alleles increase in number, and the overall fitness of the species improves incrementally.

Now, more than a century after Mendel was rediscovered and Mendelian genetics revived, we have come to realize that the great bulk of the genetic information in our own genome—indeed, in the genomes of all mammals—does not seem to specify phenotype and is often not associated with specific genes. Reflecting the discovery in 1944 that genetic information is encoded in DNA molecules, these "noncoding" stretches in the genome are often called **junk DNA** (**Figure 1.4**). Only about 1.5% of a mammal's genomic DNA carries sequence information that encodes the structures of proteins. Recent sequence comparisons of human, mouse, and dog genomes suggest that another ~2% encodes important information regulating gene expression and mediating other, still-poorly understood functions.

Because mutations act randomly on a genome, altering true genes and junk DNA indiscriminately, the great majority of mutations alter genetic information—nucleotide sequences in the DNA—that have no effect on cellular or organismic phenotype. These mutations remain silent phenotypically and are said, from the point of view of natural selection, to be **neutral mutations**, being neither advantageous nor

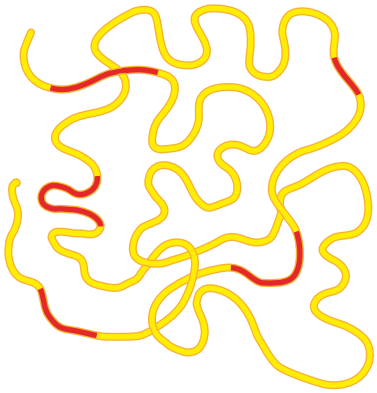


Figure 1.4 Biologically important sequences in the human genome The human genome can be characterized as a collection of relatively small islands of biologically important sequences (~3.5% of the total genome; *red*) floating amid a sea of "junk" DNA (*yellow*). The proportion of sequences carrying biological information has been greatly exaggerated for the sake of illustration. (With the passage of time, genes that appear to play important roles in cell and organismic physiology and specify certain noncoding RNA species have been localized to these intergenic regions; hence the blanket classification of all genomic sequences localized between a human cell's ~21,000 protein-coding genes as useless junk is simplistic.)

disadvantageous (Figure 1.5). Since the alleles created by these mutations are silent, their existence could not be discerned by early geneticists whose work depended on gauging phenotypes. However, with the advent of DNA sequencing techniques, it became apparent that hundreds of thousands, even a million functionally silent

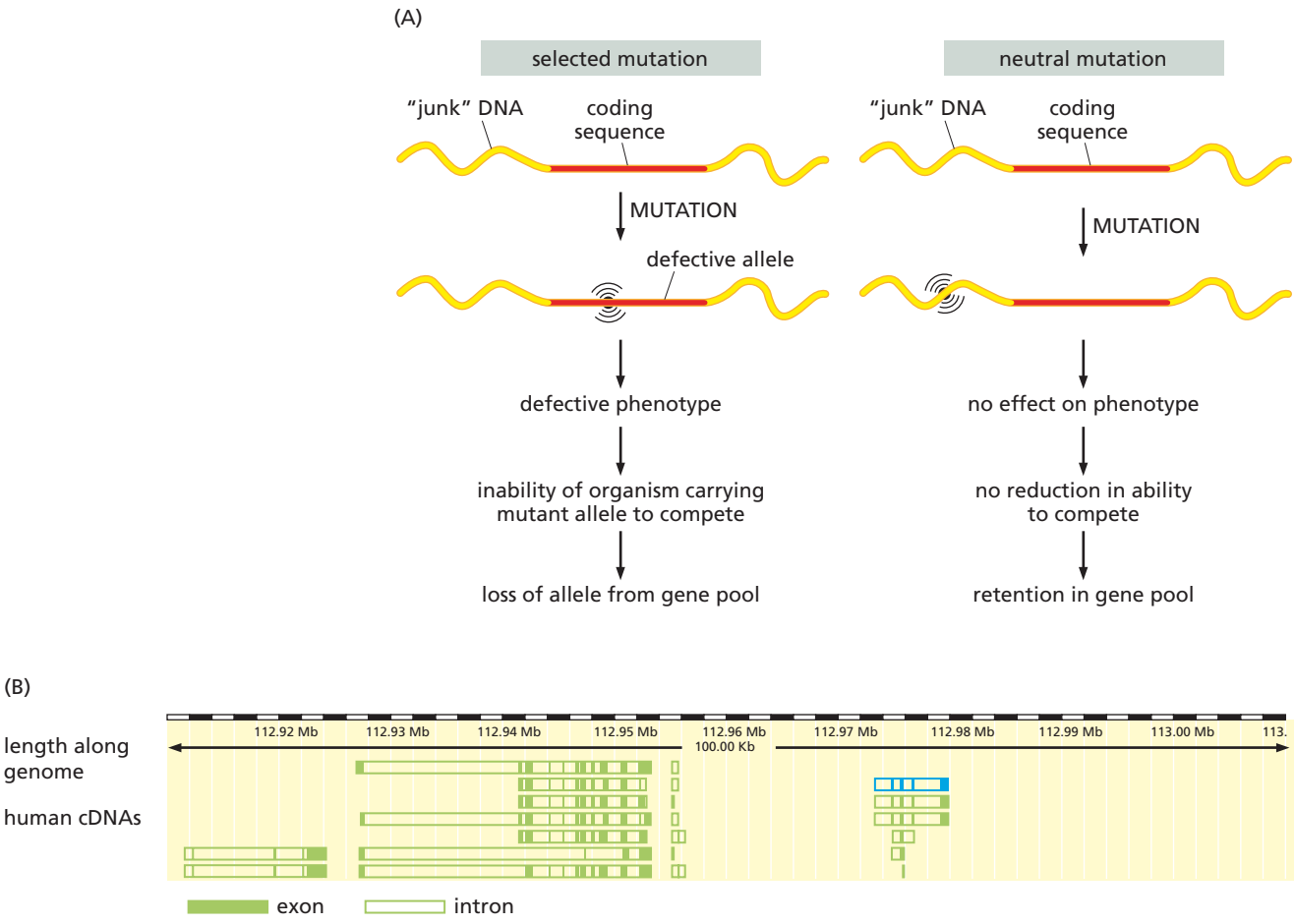


Figure 1.5 Neutral mutations and evolution (A) The coding sequences (*red*) of most genes were optimized in the distant evolutionary past. Hence, many mutations affecting amino acid sequence and thus protein structure (*left*) create alleles that compromise the organism's ability to survive. For this reason, these mutant alleles are likely to be eliminated from the species' gene pool. In contrast, mutations striking "junk" DNA (*yellow*) have no effect on phenotype and are therefore often preserved in the species' gene pool (*right*). This explains why, over extended periods of evolutionary time, coding DNA sequences change slowly, while noncoding DNA sequences change far more rapidly. (B) Depicted is a physical map of a randomly chosen 0.1-megabase segment of human Chromosome 1 (from base pair 112,912,286 to base pair 113,012,285) containing four genes. Each consists of a few islands (*solid rectangles*) that are known or likely to specify segments of mRNA molecules (i.e., exons) and large stretches of intervening sequences (i.e., introns) that do not appear to specify biological information (see Figure 1.16). The large stretches of DNA sequence between genes have not been associated with any biological function. (B, courtesy of The Wellcome Trust Sanger Institute.)

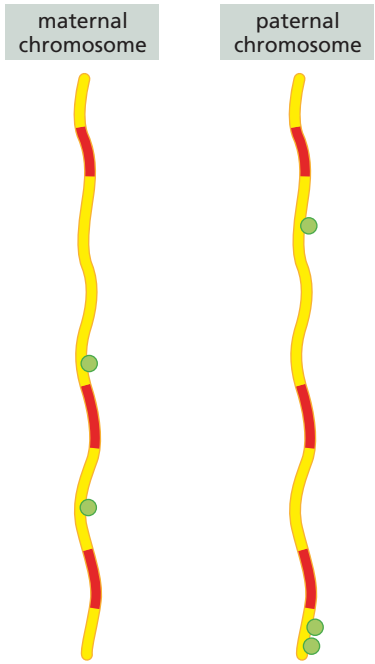


Figure 1.6 Polymorphic diversity in the human gene pool Because the great majority of human genomic DNA does not encode biologically important information (*yellow*), it has evolved relatively rapidly and has accumulated many subtle differences in sequences—polymorphisms—that are phenotypically silent (see Figure 1.5). Such polymorphisms are transmitted like Mendelian alleles, but their presence in a genome can be ascertained only by molecular techniques such as DNA sequencing. The dots (*green*) indicate where the sequence on this chromosome differs from the sequence that is most common in the human gene pool. For example, the prevalent sequence in one stretch may be TAACTGG, while the variant sequence TTACTGG may be carried by a minority of humans and constitute a polymorphism. The presence of a polymorphism in one chromosome but not the other represents a region of heterozygosity, even though a nearby gene (*red*) may be present in the identical allelic version on both chromosomes and therefore be in a homozygous configuration.

mutations can be found scattered throughout the genomes of organisms such as humans. The genome of each human carries its own unique array of these functionally silent genetic alterations. The term *polymorphism* was originally used to describe variations in shape and form that distinguish normal individuals within a species from each other. These days, geneticists use the term **genetic polymorphisms** to describe the inter-individual, functionally silent differences in DNA sequence that make each human genome unique (**Figure 1.6**).

During the course of evolution, the approximately 3.5% of the genome that does encode biological function behaves much differently from the junk DNA. Junk DNA sequences suffer mutations that have no effect on the viability of an organism. Consequently, countless mutations in the noncoding sequences of a species' genome survive in its gene pool and accumulate progressively during its evolutionary history. In contrast, mutations affecting the coding sequences usually lead to loss of function and, as a consequence, loss of organismic viability; hence, these mutations are weeded out of the gene pool by the hand of natural selection, explaining why genetic sequences that do specify biological phenotypes generally change very slowly over long evolutionary time periods (**Sidebar 1.1**).

1.3 Mendelian genetics governs how both genes and chromosomes behave

In the first decade of the twentieth century, Mendel's rules of genetics were found to have a striking parallel in the behavior of the chromosomes that were then being visualized under the light microscope. Both Mendel's genes and the chromosomes were found to be present in pairs. Soon it became clear that an identical set of chromosomes is present in almost all the cells of a complex organism. This chromosomal array, often termed the **karyotype**, was found to be duplicated each time a cell went through a cycle of growth and division.

The parallels between the behaviors of genes and chromosomes led to the speculation, soon validated in hundreds of different ways, that the mysterious information packets called genes were carried by the chromosomes. Each chromosome was realized to carry its own unique set of genes in a linear array. Today, we know that as many as several thousand genes may be arrayed along a mammalian chromosome. (Human Chromosome 1—the largest of the set—holds at least 3148 distinct genes.) Indeed, the length of a chromosome, as viewed under the microscope, is roughly proportional to the number of genes that it carries.

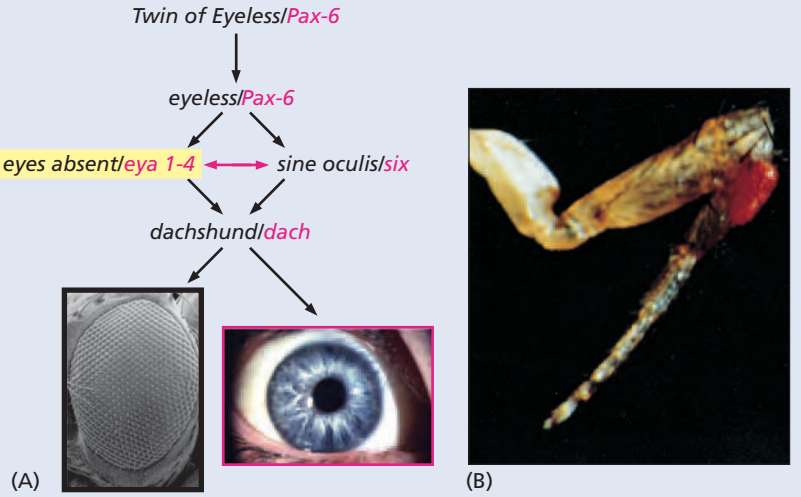
Each gene was found to be localized to a specific site along the length of a specific chromosome. This site is often termed a genetic **locus**. Much effort was expended by geneticists throughout the twentieth century to map the sites of genes—genetic loci—along the chromosomes of a species (**Figure 1.8**).

Sidebar 1.1 Evolutionary forces dictate that certain genes are highly conserved Many genes encode cellular traits that are essential for the continued viability of the cell. These genes, like all others in the genome, are susceptible to the ever-tinkering hand of mutation, which is continually creating new gene sequences by altering existing ones. Natural selection tests these novel sequences and determines whether they specify phenotypes that are more advantageous than the preexisting ones.

Almost invariably, the sequences in genes required for cell and therefore organismic viability were already optimized hundreds of millions of years ago. Consequently, almost all subsequently occurring changes in the sequence information of these genes would have been deleterious and would have compromised the viability of the cell and, in turn, the organism. These mutant alleles were soon lost, because the mutant organisms carrying them failed to leave descendants. This dynamic explains why the sequences of many genes have been highly conserved over vast evolutionary time periods. Stated more accurately, the structures of their encoded proteins have been highly conserved.

In fact, the great majority of the proteins that are present in our own cells and are required for cell viability were first developed during the evolution of single-cell **eukaryotes**. This is indicated by numerous observations showing that many of our proteins have clearly recognizable counterparts in single-cell eukaryotes, such as baker's yeast. Another large repertoire of highly conserved genes and proteins is traceable to the appearance of the first multicellular animals (metazoa); these genes enabled the development of distinct organs and of organismic physiology. Hence, another large group of our own genes and proteins is present in counterpart form in worms and flies (Figure 1.7).

By the time the ancestor of all mammals first appeared more than 150 million years ago, virtually all the biochemical and molecular features present in contemporary mammals had already been developed. The fact that they have changed little in the intervening time points to their optimization long before the appearance of the various mammalian orders. This explains why the embryogenesis, physiology, and biochemistry of all mammals is very similar, indeed, so similar that lessons learned through the study of laboratory mice are almost always transferable to an understanding of human biology.

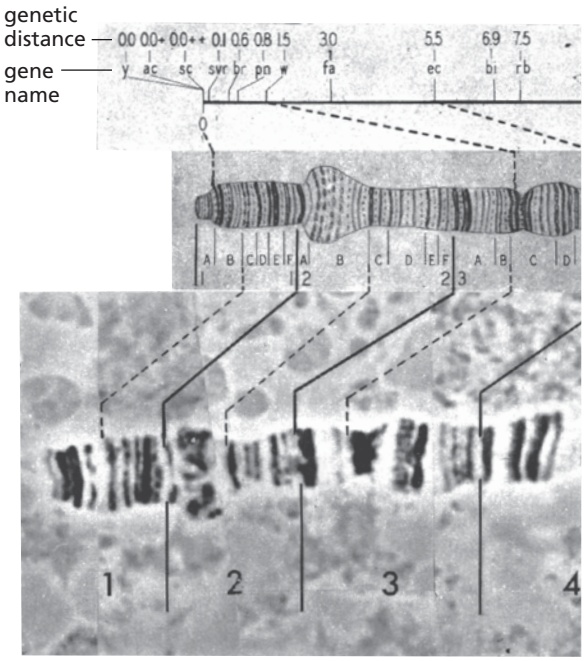


human	FGLARAFGIPIRVYTHEVVTLWYRSPEVLLGS
S. pombe	FGLARSFGVPLRNYTHEIVTLWYRAPEVLLGS
S. cerevisiae	FGLARAFGVPLRAYTHEIVTLWYRAPEVLLGG
human	ARYSTPVDIWSIGTIFAELATKLPLFHGDSEI
S. pombe	RHYSTGVDIWSVGCIFAENIRRSPLFPGDSEI
S. cerevisiae	KQYSTGVDTWSIGCIFAEHCNRLPIFSGDSEI
human	DQLFRIPRALGTPNNEVWPEVESLQDYKNTFP
S. pombe	DEIFKIPQVLGTPNEEVWPGVTLLQDYKSTFP
S. cerevisiae	DQIFKIPRVLGTPNEAIWPDIVYLPDFKPSFP

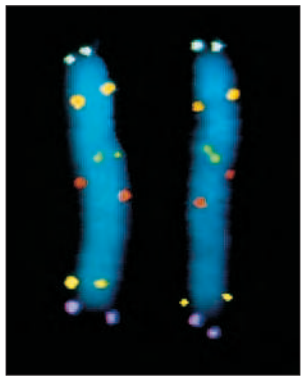
(C)

Figure 1.7 Extraordinary conservation of gene function
The last common ancestor of flies and mammals lived more than 600 million years ago. Moreover, fly (i.e., arthropod) eyes and mammalian eyes show totally different architectures. Nevertheless, the genes that orchestrate their development (*eyeless* in the fly, *Pax-6/small eye* in the mouse) are interchangeable—the gene from one organism can replace the corresponding mutant gene from the other and restore wild-type function. (A) Thus, the genes encoding components of the signal transduction cascades that operate downstream of these master regulators to trigger eye development (*black* for flies, *pink* for mice) are also highly conserved and interchangeable. (B) The expression of the mouse *Pax-6/small eye* gene, like the *Drosophila eyeless* gene, in an inappropriate (*ectopic*) location in a fly embryo results in the fly developing a fly eye on its leg, demonstrating the interchangeability of the two genes. (C) The conservation of genetic function over vast evolutionary distances is often manifested in the amino acid sequences of homologous proteins. Here, the amino acid sequence of a human protein is given together with the sequences of the corresponding proteins from two yeast species, *S. pombe* and *S. cerevisiae*. (A, courtesy of I. Rebay. B, courtesy of Walter Gehring. C, adapted from B. Alberts et al., Essential Cell Biology, 3rd edition New York: Garland Science, 2010.)

The diploid genetic state that reigns in most cells throughout the body was found to be violated in the *germ cells*, sperm and egg. These cells carry only a single copy of each chromosome and gene and thus are said to be **haploid**. During the formation of germ cells in the testes and ovaries, each pair of chromosomes is separated and one of the pair (and thus associated genes) is chosen at random for incorporation into the sperm or egg. When sperm and egg combine subsequently during fertilization,



(A)



(B)

Figure 1.8 Localization of genes along chromosomes (A) The physical structure of *Drosophila* chromosomes was mapped by using the fly's salivary gland chromosomes, which exhibit banding patterns resulting from alternating light (sparse) and dark (condensed) chromosomal regions (*bottom*). Independently, genetic crosses yielded linear maps (*top*) of various genetic loci arrayed along the chromosomes. These loci were then aligned with physical banding maps, like the one shown here for the beginning of the left arm of *Drosophila* chromosome 1. (B) The availability of DNA probes that hybridize specifically to various genes now makes it possible to localize genes along a chromosome by tagging each probe with a specific fluorescent dye or combination of dyes. Shown are six genes that were localized to various sites along human Chromosome 5 by using fluorescence *in situ* hybridization (FISH) during metaphase. (There are two dots for each gene because chromosomes are present in duplicate form during metaphase of mitosis.) (A, from M. Singer and P. Berg, Genes and Genomes. Mill Valley, CA: University Science Books, 1991, as taken from C.B. Bridges, *J. Hered.* 26:60, 1935. B, courtesy of David C. Ward.)

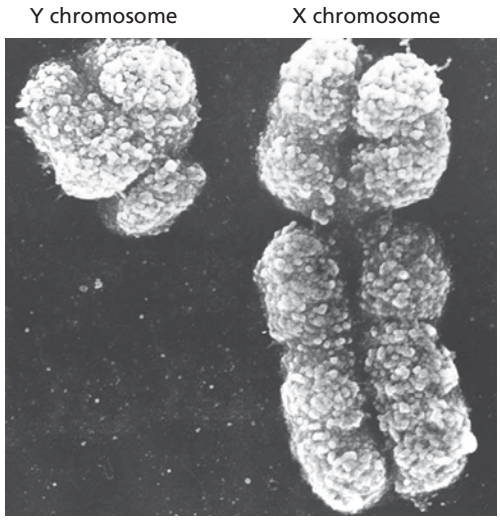
the two haploid genomes fuse to yield the new diploid genome of the fertilized egg. All cells in the organism descend directly from this diploid cell and, if all goes well, inherit precise replicas of its diploid genome. In a large multicellular organism like the human, this means that a complete copy of the genome is present in almost all of the approximately 3×10^{13} cells throughout the body!

With the realization that genes reside in chromosomes, and that a complete set of chromosomes is present in almost all cell types in the body, came yet another conclusion that was rarely noted: genes create the phenotypes of an organism through their ability to act locally by influencing the behavior of its individual cells. The alternative—that a single set of genes residing at some unique anatomical site in the organism controls the entire organism's development and physiology—was now discredited.

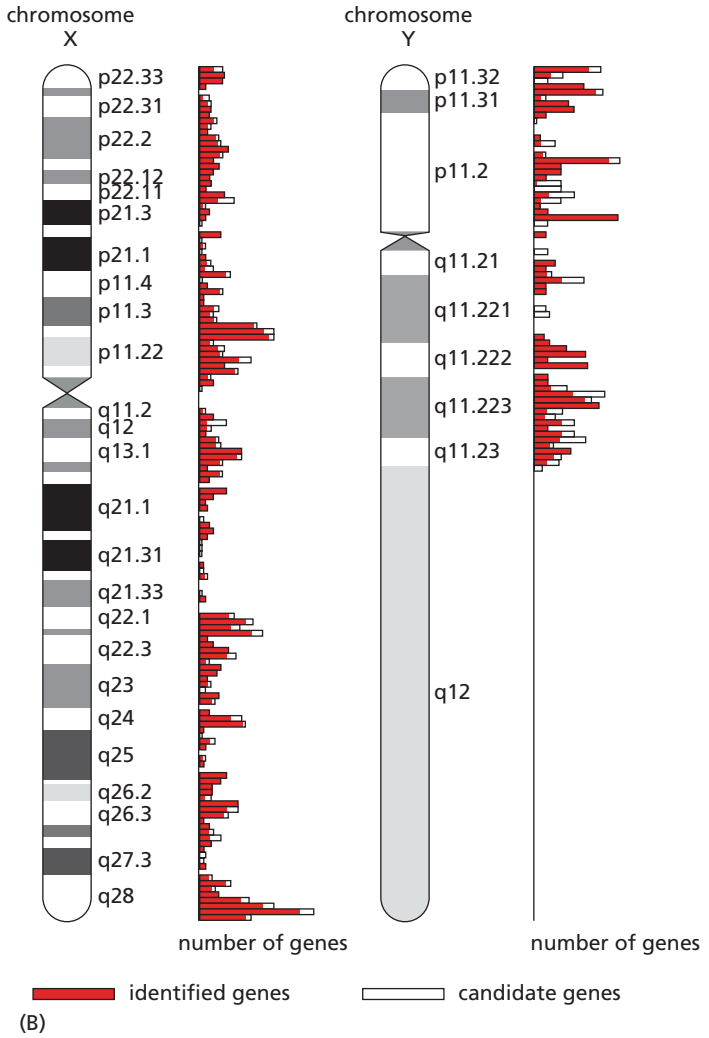
The rule of paired, similarly appearing chromosomes was found to be violated by some of the sex chromosomes. In the cells of female placental mammals, there are two similarly appearing X chromosomes, and these behave like the **autosomes** (the nonsex chromosomes). But in males, an X chromosome is paired with a Y chromosome, which is smaller and carries a much smaller repertoire of genes. In humans, the X chromosome is thought to carry about 900 genes, compared with the 78 distinct genes on the Y chromosome, which, because of redundancy, specify only 27 distinct proteins (**Figure 1.9**).

This asymmetry in the configuration of the sex chromosomes puts males at a biological disadvantage. Many of the 900 or so genes on the X chromosome are vital to normal organismic development and function. The twofold redundancy created by the paired X chromosomes guarantees more robust biology. If a gene copy on one of the X chromosomes is defective (that is, a nonfunctional mutant allele), chances are that the second copy of the gene on the other X chromosome can continue to carry out the task of the gene, ensuring normal biological function. Males lack this genetic fail-safe system in their sex chromosomes. One of the more benign consequences of this is color blindness, which strikes males frequently and females infrequently, due to the localization on the X chromosome of the genes encoding the color-sensing proteins of the retina.

This disparity between the genders is mitigated somewhat by the mechanism of X-inactivation. Early in embryogenesis, one of the two X chromosomes is randomly



(A)



(B)

Figure 1.9 Physical maps of human sex chromosomes (A) Shown is a scanning electron micrograph of human X and Y chromosomes. Like the 22 autosomes (nonsex chromosomes), they have been sequenced. (B) This has allowed the *cytologic* maps of these chromosomes (determined by microscopically examining stained chromosomes at the metaphase of mitosis) to be matched with their DNA sequence. Note that the short arm of a human chromosome is the "p" arm, while the long arm is the "q" arm. Each chromosome has been divided into regions on the basis of the observed banding pattern, and distinct genes have been assigned on the basis of the sequence analyses (*histograms to right of each chromosome*). Identified genes are filled bars (*red*), while sequences that appear to encode still-to-be-identified genes are in open bars; in most chromosomal regions the latter represent a small minority. The human Y chromosome is ~57 megabases (Mb) long, compared with the X chromosome's ~155 Mb. (A, courtesy of Indigo® Instruments. B, courtesy of The Wellcome Trust Sanger Institute. Ensembl genome browser http://www.ensembl.org.)

inactivated in each of the cells of a female embryo. This inactivation silences almost all of the genes on this chromosome and causes it to shrink into a small particle termed the **Barr body**. Subsequently, all descendants of that cell will inherit this pattern of chromosomal inactivation and will therefore continue to carry the same inactivated X chromosome. Accordingly, the female advantage of carrying redundant copies of X chromosome–associated genes is only a partial one (Supplementary Sidebar 1.1).

Color blindness reveals the virtues of having two redundant gene copies around to ensure that biological function is maintained. If one copy is lost through mutational inactivation, the surviving gene copy is often capable of specifying a wild-type phenotype. Such functional redundancy operates for the great majority of genes carried by the autosomes. As we will see later, this dynamic plays an important role in cancer development, since virtually all of the genes that operate to prevent runaway proliferation of cells are present in two redundant copies, both of which must be inactivated in a cell before their growth-suppressing functions are lost and malignant cell proliferation can occur.

1.4 Chromosomes are altered in most types of cancer cells

Individual genes are far too small to be seen with a light microscope, and subtle mutations within a gene are smaller still. Consequently, the great majority of the mutations that play a part in cancer cannot be visualized through microscopy. However, the examination of chromosomes through the light microscope can give evidence of

large-scale alterations of the cell genome. Indeed, such alterations were noted as early as 1892, specifically in cancer cells.

Today, we know that cancer cells often exhibit aberrantly structured chromosomes of various sorts, the loss of entire chromosomes, the presence of extra copies of others, and the fusion of the arm of one chromosome with part of another. These changes in overall chromosomal configuration expand our conception of how mutations can affect the genome: since alterations of overall chromosomal structure and number also constitute types of genetic change, these changes must be considered to be the consequences of mutations (**Sidebar 1.2**). And importantly, the abnormal chromosomes seen initially in cancer cells provided the first clue that these cells might be genetically aberrant, that is, that they were mutants (see Figure 1.11).

The normal configuration of chromosomes is often termed the **euploid** karyotypic state. Euploidy implies that each of the autosomes is present in normally structured pairs and that the X and Y chromosomes are present in the numbers appropriate for the sex of the individual carrying them. Deviation from the euploid karyotype—the state termed **aneuploidy**—is seen, as mentioned above, in many cancer cells. Often this aneuploidy is merely a consequence of the general chaos that reigns within a cancer cell. However, this connection between aneuploidy and malignant cell proliferation also hints at a theme that we will return to repeatedly in this book: the acquisition of extra copies of one chromosome or the loss of another can create a genetic configuration that somehow benefits the cancer cell and its agenda of runaway proliferation.

1.5 Mutations causing cancer occur in both the germ line and the soma

Mutations alter the information content of genes, and the resulting mutant alleles of a gene can be passed from parent to offspring. This transmission from one generation to the next, made possible by the germ cells (sperm and egg), is said to occur via the **germ line** (**Figure 1.10**). Importantly, the germ-line transmission of a recently created mutant allele from one organism to its offspring can occur only if a precondition has been met: the responsible mutation must strike a gene carried in the genome of sperm or egg or in the genome of one of the cell types that are immediate precursors of the sperm or egg within the gonads. Mutations affecting the genomes of cells everywhere else in the body—which constitute the **soma**—may well affect the particular cells in which such mutations strike but will have no prospect of being transmitted to the offspring of an organism. Such **somatic mutations** cannot become incorporated into the vehicles of generation-to-generation genetic transmission—the chromosomes of sperm or eggs.

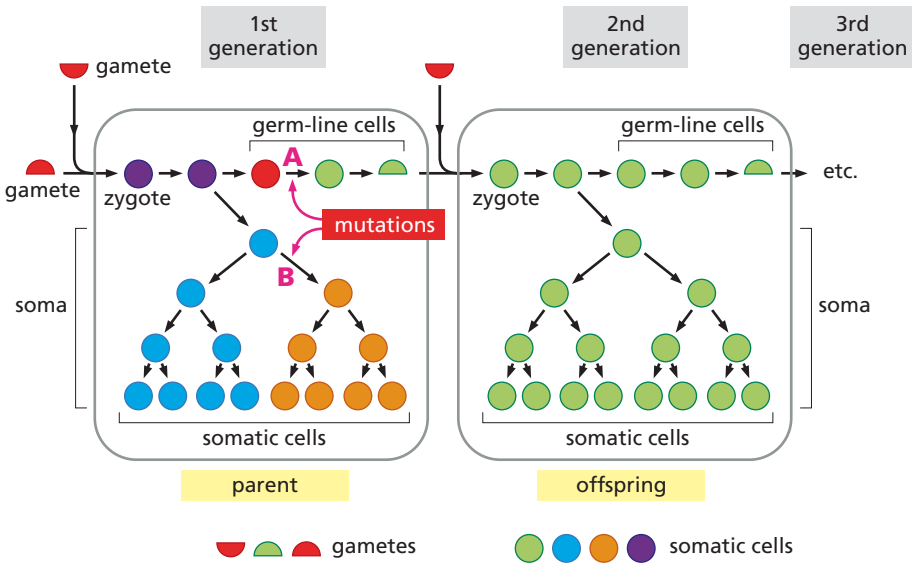


Figure 1.10 Germ-line versus somatic mutations Mutation A, which occurs in the genome of a germ-line cell in the gonads, can be passed from parent (*above left*) to offspring via gametes—sperm or egg (*half circles*). Once incorporated into the fertilized egg (zygote), the mutant alleles can then be transmitted to all of the cells in the body of the offspring (*middle*) outside of the gonads, i.e., its soma, as well as being transmitted via germ-line cells and gametes to a third generation (*not shown*). However, mutation B (*left*), which strikes the genome of a somatic cell in the parent, can be passed only to the lineal descendants of that mutant cell within the body of the parent and cannot be transmitted to offspring. (Adapted from B. Alberts et al., Essential Cell Biology, 3rd ed. New York: Garland Science, 2010.)

Sidebar 1.2 Cancer cells are often aneuploid The presence of abnormally structured chromosomes and changes in chromosome number provided the first clue, early in the twentieth century, that changes in cell genotype often accompany and perhaps cause the uncontrolled proliferation of malignant cells. These deviations from the normal euploid karyotype can be placed into a number of categories. Chromosomes that seem to be structurally normal may accumulate in extra copies, leading to three, four, or even more copies of these chromosomes per cancer cell nucleus (**Figure 1.11**); such deviations from normal chromosome number are manifestations of *aneuploidy*.

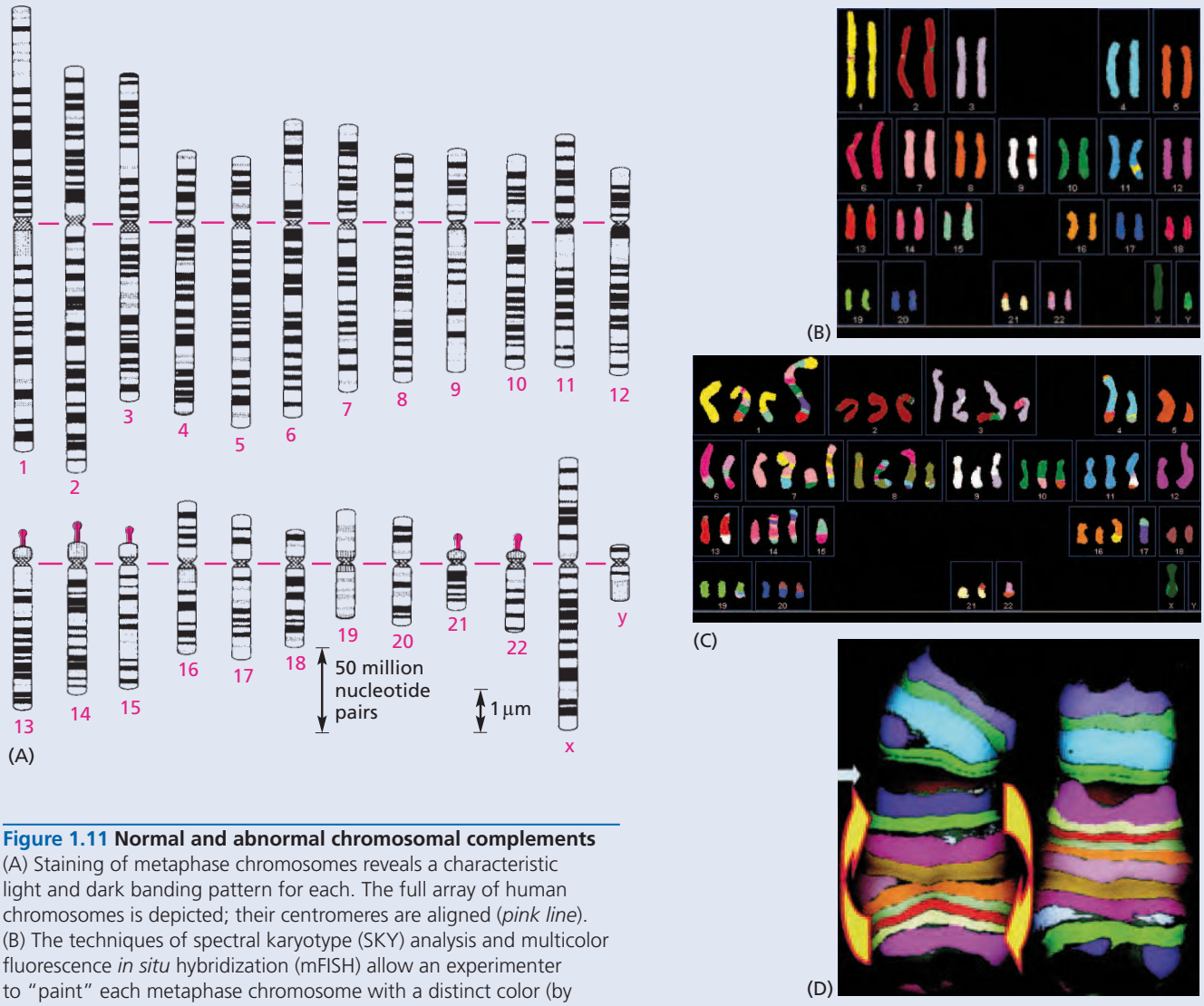
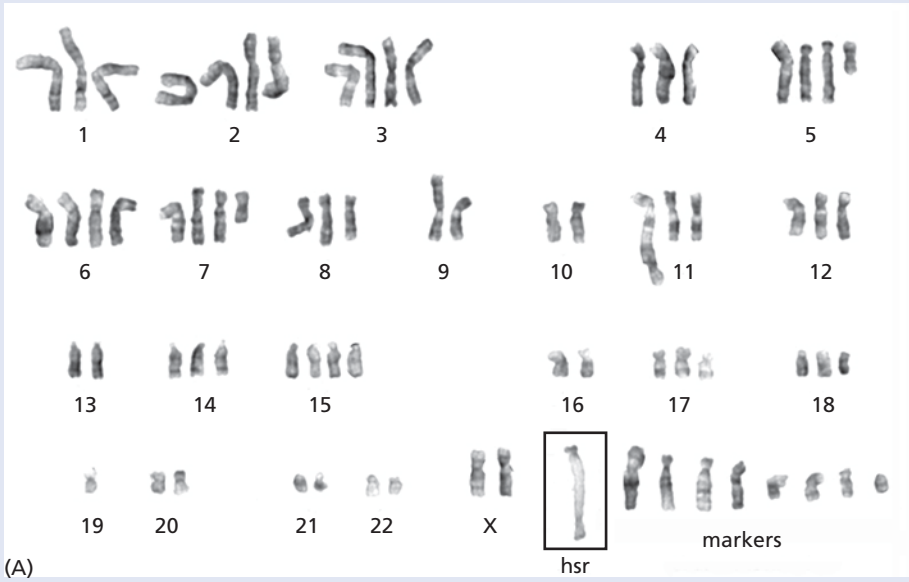
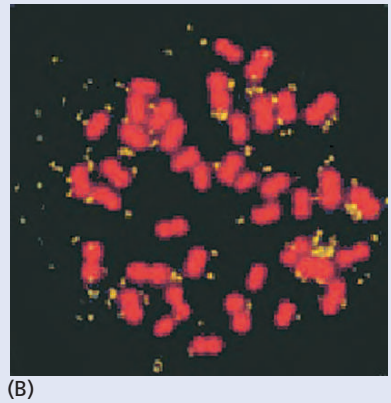


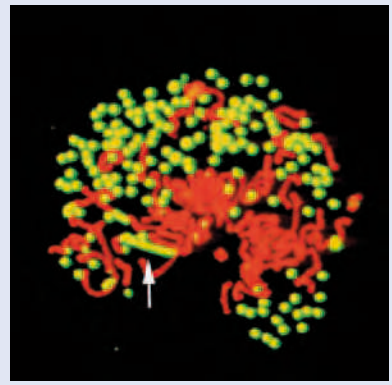
Figure 1.11 Normal and abnormal chromosomal complements (A) Staining of metaphase chromosomes reveals a characteristic light and dark banding pattern for each. The full array of human chromosomes is depicted; their centromeres are aligned (*pink line*). (B) The techniques of spectral karyotype (SKY) analysis and multicolor fluorescence *in situ* hybridization (mFISH) allow an experimenter to "paint" each metaphase chromosome with a distinct color (by hybridizing chromosome-specific DNA probes labeled with various fluorescing dyes to the chromosomes). The actual colors in images such as these are generated by computer. The diploid karyotype of a normal human male cell is presented. (The small regions in certain chromosomes that differ from the bulk of these chromosomes represent hybridization artifacts.) (C) The aneuploid karyotype of a human pancreatic cancer cell, in which some chromosomes are present in inappropriate numbers and in which numerous translocations (exchanges of segments between chromosomes) are apparent. (D) Here, mFISH was used to label intrachromosomal subregions with specific fluorescent dyes, revealing that a large portion of an arm of normal human Chromosome 5 (*right*) has been inverted (*left*) in cells of a worker who had been exposed to plutonium in the nuclear weapons industry of the former Soviet Union. (A, adapted from U. Francke, *Cytogenet. Cell Genet.* 31:24–32, 1981. B and C, courtesy of M. Grigorova, J.M. Staines and P.A.W. Edwards. D, from M.P. Hande et al., *Am. J. Hum. Genet.* 72:1162–1170, 2003.)

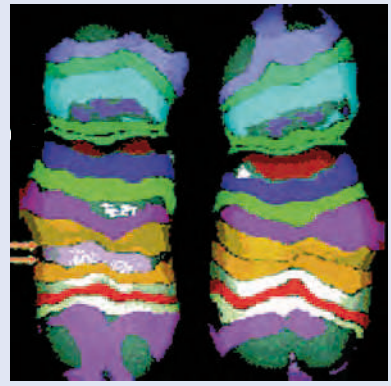
Alternatively, chromosomes may undergo changes in their structure. A segment may be broken off one chromosomal arm and become fused to the arm of another chromosome, resulting in a chromosomal **translocation** (Figure 1.11C). Moreover, chromosomal segments may be exchanged between chromosomes from different chromosome pairs, resulting in **reciprocal translocations**. A chromosomal segment may also become inverted, which may affect the regulation of genes that are located near the breakage-and-fusion points (Figure 1.11D).

A segment of a chromosome may be copied many times over, and the resulting extra copies may be fused head-to-tail in long arrays within a chromosomal segment that is termed an HSR (**homogeneously staining region**; **Figure 1.12A**). A segment may also be cleaved out of a chromosome, replicate









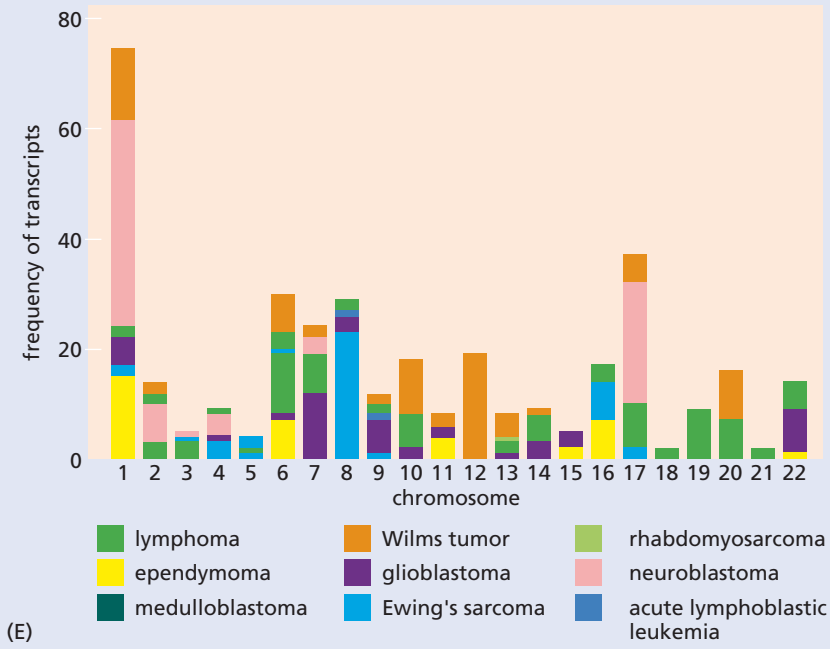


Figure 1.12 Increases and decreases in copy number of chromosomal segments (A) The amplification in the copy number of the *myc* oncogene (see Section 8.9) in a human neuroendocrinal tumor has caused an entire stretch of chromosome to stain *white* (rectangle), creating a homogeneously staining region (HSR). (B) Double-minute chromosomes (DMs) derive from chromosomal segments that have broken loose from their original sites and have been replicated repeatedly as extrachromosomal genetic elements; like normal chromatids, these structures are doubled during metaphase of mitosis. FISH reveals the presence of amplified copies of the *HER2/neu* oncogene borne on DMs (*yellow dots*) in a mouse breast cancer cell. (C) Occasionally, an amplified gene may be found both in an HSR (nested within a chromosome) and in DMs. Here, analysis of COLO320 cells reveals multiple copies of the *myc* oncogene (*yellow*), amid the chromosomes (*red*). One HSR is indicated by the arrow, while many dozens of DMs are apparent. (D) The use of multicolor FISH (mFISH) revealed that a segment within normal human Chromosome 5 (*paired arrows, left*) has been deleted (an interstitial deletion, *right*) following extensive exposure to radiation from plutonium. (E) A survey of nine different types of pediatric cancer indicates that each cancer type has characteristic gene amplification and deletion patterns with corresponding changes in the expression of the altered genes. For example, neuroblastomas (*pink*) often have changes in the copy numbers of genes on chromosomes 1 and 17 and corresponding changes in the levels of the transcripts expressed by these genes. (A, from J.-M. Wen et al., *Cancer Genet. Cytogenet.* 135:91–95, 2002. B, from C. Montagna et al., *Oncogene* 21:890–898, 2002. C, from N. Shimizu et al., *J. Cell Biol.* 140:1307–1320, 1998. D, from M.P. Hande et al., *Am. J. Hum. Genet.* 72:1162–1170, 2003. E, from G. Neale et al., *Clin. Cancer Res.* 14:4572–4583, 2008.)

as an autonomous, extrachromosomal entity, and increase to many copies per nucleus, resulting in the appearance of subchromosomal fragments termed DMs (**double minutes**; Figure 1.12B). These latter two changes cause increases in the copy number of genes carried in such segments, resulting in **gene amplification**. Sometimes, both types of amplification coexist in the same cell (Figure 1.12C). Gene amplification can favor the growth of cancer cells by increasing the copy number of growth-promoting genes.

On some occasions, certain growth-inhibiting genes may be discarded by cancer cells during their development. For example, when a segment in the middle of a chromosomal arm is discarded and the flanking chromosomal regions are joined, this results in an **interstitial deletion** (Figure 1.12D).

These descriptions of copy-number changes in genes, involving both amplifications and deletions, might suggest widespread chaos in the genomes of cancer cells, with gene amplifications and deletions occurring randomly. However, as the karyotypes and genomes of human tumors have been examined more intensively, it has become clear that certain regions of the genome tend to be lost characteristically in certain tumor types but not in others (Figure 1.12E). This suggests a theme that we will pursue in great detail throughout this book—that the gains and losses of particular genes favor the proliferation of specific types of tumors. This indicates that different tumor types undergo different genetic changes as they develop progressively from the precursor cells in normal tissues.

Somatic mutations are of central importance to the process of cancer formation. As described repeatedly throughout this book, a somatic mutation can affect the behavior of the cell in which it occurs and, through repeated rounds of cell growth and division, can be passed on to all descendant cells within a tissue. These direct descendants of a single progenitor cell, which may ultimately number in the millions or even billions, are said to constitute a cell **clone**, in that all members of this group of cells trace their ancestry directly back to the single cell in which the mutation originally occurred.

An elaborate repair apparatus within each cell continuously monitors the cell's genome and, with great efficiency, eradicates mutant sequences, replacing them with appropriate wild-type sequences. We will examine this repair apparatus in depth in Chapter 12. This apparatus maintains genomic integrity by minimizing the number of mutations that strike the genome and are then perpetuated by transmission to descendant cells. One stunning indication of the efficiency of genome repair comes from the successes of organismic cloning: the ability to generate an entire organism from the nucleus of a differentiated cell (prepared from an adult) indicates that this adult cell genome is essentially a faithful replica of the genome of a fertilized egg, which existed many years and many cell generations earlier (Supplementary Sidebar 1.2).

However, no system of damage detection and repair is infallible. Some mistakes in genetic sequence survive its scrutiny, become fixed in the cell genome, are copied into new DNA molecules, and are then passed on as mutations to progeny cells. In this sense, many of the mutations that accumulate in the genome represent the consequences of occasional oversights made by the repair apparatus. Yet others are the results of catastrophic damage to the genome that exceeds the capacities of the repair apparatus.

1.6 Genotype embodied in DNA sequences creates phenotype through proteins

The genes studied in Mendelian genetics are essentially mathematical abstractions. Mendelian genetics explains their transmission, but it sheds no light on how genes create cellular and organismic phenotypes. Phenotypic attributes can range from complex, genetically templated behavioral traits to the **morphology** (shape, form) of cells and subcellular organelles to the biochemistry of cell metabolism. This mystery of how genotype creates phenotype represented the major problem of twentieth-century biology. Indeed, attempts at forging a connection between these two became the obsession of many molecular biologists during the second half of the twentieth century and continue as such into the twenty-first, if only because we still possess an incomplete understanding of how genotype influences phenotype.

Molecular biology has provided the basic conceptual scaffold for understanding the connection between genotype and phenotype. In 1944, DNA was proven to be the

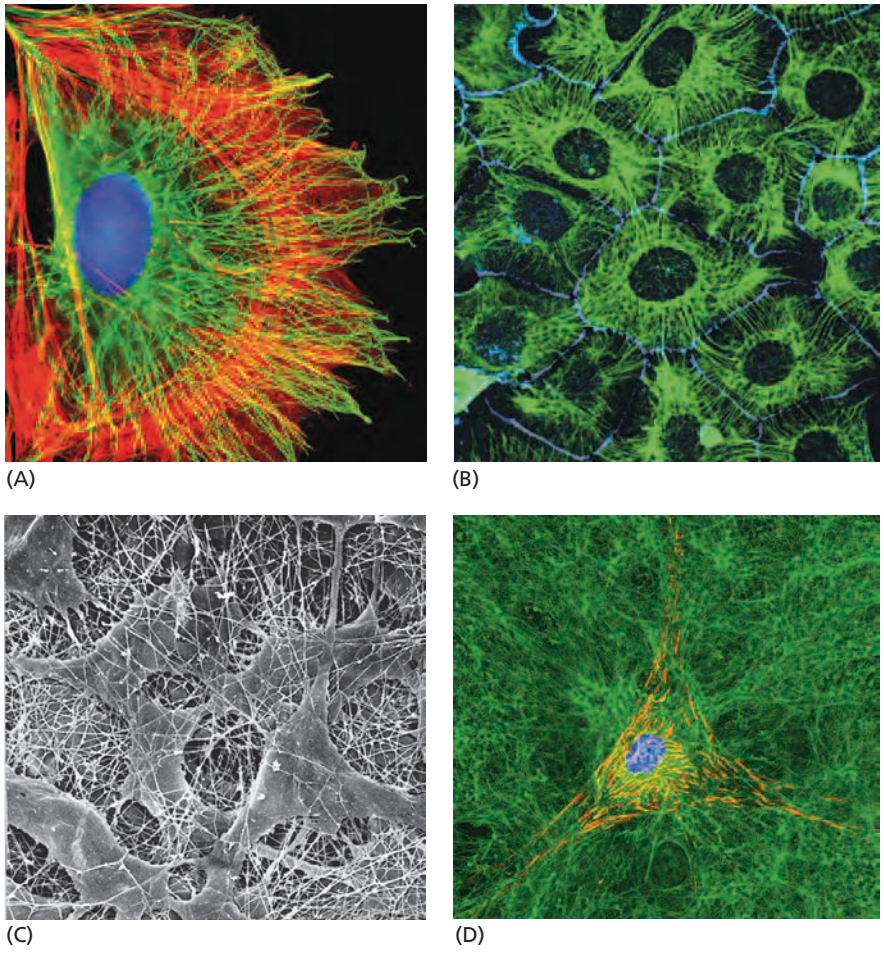


Figure 1.13 Intracellular and extracellular scaffolding The cytoskeleton is assembled from complex networks of intermediate filaments, actin microfilaments, and microtubules. Together, they generate the shape of a cell and enable its motion. (A) In this cultured cell, microfilaments composed of actin (*orange*) form bundles that lie parallel to the cell surface while microtubules composed of tubulin (*green*) radiate outward from the nucleus (*blue*). Both types of fibers are involved in the formation of protrusions from the cell surface. (B) Here, an important intermediate filament of epithelial cells—keratin—is detected using an anti-keratin-specific antibody (*green*). The boundaries of cells are labeled with a second antibody that reacts with a plasma membrane protein (*blue*). (C) Cells secrete a diverse array of proteins that are assembled into the extracellular matrix (ECM). A scanning electron micrograph reveals the complex meshwork of collagen fibers, glycoproteins, hyaluronan, and proteoglycans, in which fibroblasts (connective tissue cells) are embedded. (D) A cell of the NIH 3T3 cell line, which is used extensively in cancer cell biology, is shown amid an ECM network of fibronectin fibers (*green*). The points of cellular attachment to the fibronectin are mediated by integrin receptors on the cell surface (*orange, yellow*). (A, courtesy of Albert Tousson, High-Resolution Imaging Facility, University of Alabama at Birmingham. B, courtesy of Kathleen Green and Evangeline Amargo. C, courtesy of T. Nishida. D, from E. Cukierman et al., *Curr. Opin. Cell Biol.* 14:633–639, 2002.)

chemical entity in which the genetic information of cells is carried. Nine years later, Watson and Crick elucidated the double-helical structure of DNA. A dozen years after that, in 1964, it became clear that the sequences in the bases of the DNA double helix determine precisely the sequence of amino acids in proteins. The unique structure and function of each type of protein in the cell is determined by its sequence of amino acids. Therefore, the specification of amino acid sequence, which is accomplished by base sequences in the DNA, provides almost all the information that is required to construct a protein.

Once synthesized within cells, proteins proceed to create phenotype, doing so in a variety of ways. Proteins can assemble within the cell to create the components of the **cytoarchitecture**, or more specifically, the **cytoskeleton** (**Figure 1.13A and B**). When secreted into the space between cells, such proteins form the **extracellular matrix** (ECM); it ties cells together, enabling them to form complex tissues (Figure 1.13C and D). As we will see later, the structure of the ECM is often disturbed by malignant cancer cells, enabling them to migrate to sites within a tissue and organism that are usually forbidden to them.

Many proteins function as enzymes that catalyze the thousands of biochemical reactions that together are termed **intermediary metabolism**; without the active intervention of enzymes, few of these reactions would occur spontaneously. Proteins can also contract and create cellular movement (**motility**; **Figure 1.14**) as well as muscle contraction. Cellular motility plays a role in cancer development by allowing cancer cells to spread through tissues and migrate to distant organs.

And most important for the process of cancer formation, proteins can convey signals between cells, thereby enabling complex tissues to maintain the appropriate numbers of constituent cell types. Within individual cells, certain proteins receive signals

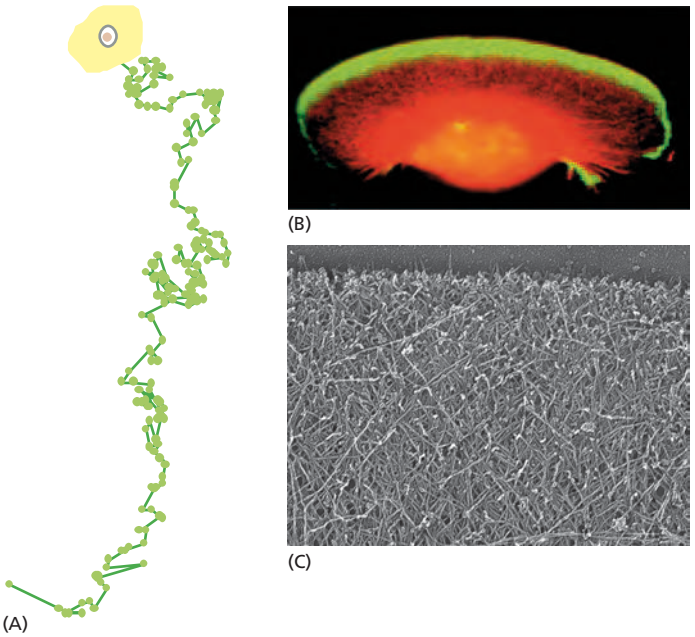


Figure 1.14 Cell motility (A) The movement of individual cells in a culture dish can be plotted at intervals and scored electronically. This image traces the movement of a human vascular endothelial cell (the cell type that forms the lining of blood vessels) toward two attractants located at the bottom—vascular endothelial growth factor (VEGF) and sphingosine-1-phosphate (S1P). Such locomotion is presumed to be critical to the formation of new blood vessels within a tumor. Each point represents a position plotted at 10-minute intervals. This motility is made possible by complex networks of proteins that form the cells' cytoskeletons. (B) The advancing cell is a fish keratocyte; its leading edge (*green*) is pushed forward by an actin filament network, such as the one shown in C. (C) Seen here is the network of actin filaments that is assembled at the leading edge of a motile cell. (A, courtesy of C. Furman and F. Gertler. B and C, from T. Svitkina and G. Borisy, *J. Cell Biol*. 145:1009–1026, 1999. © The Rockefeller University Press.)

from an extracellular source, process these signals, and pass them on to other proteins within the cell; such signal-processing functions, often termed intracellular signal **transduction**, are also central to the creation of cancers, since many of the abnormal-growth phenotypes of cancer cells are the result of aberrantly functioning intracellular signal-transducing molecules.

The functional versatility of proteins makes it apparent that almost all aspects of cell and organismic phenotype can be created by their actions. Once we realize this, we can depict genotype and phenotype in the simplest of molecular terms: genotype resides in the sequences of bases in DNA, while phenotype derives from the actions of proteins. (In fact, this depiction is simplistic, because it ignores the important role of RNA molecules as intermediaries between DNA sequences and protein structure and the recently discovered abilities of some RNA molecules to function as enzymes and others to act as regulators of the expression of certain genes.)

In the complex **eukaryotic** (nucleated) cells of animals, as in the simpler **prokaryotic** cells of bacteria, DNA sequences are copied into RNA molecules in the process termed **transcription**; a gene that is being transcribed is said to be actively **expressed**, while a gene that is not being transcribed is often considered to be **repressed**. In its simplest version, the transcription of a gene yields an RNA molecule of length comparable to the gene itself. Once synthesized, the base sequences in the RNA molecule are **translated** by the protein-synthesizing factories in the cell, its **ribosomes**, into a sequence of amino acids. The resulting macromolecule, which may be hundreds, even thousands of amino acids long, folds up into a unique three-dimensional configuration and becomes a functional protein (**Figure 1.15**).

Post-translational modification of the initially synthesized protein may result in the covalent attachment of certain chemical groups to specific amino acid residues in the protein chain; included among these modifications are, notably, phosphates, complex sugar chains, and methyl, acetyl, and lipid groups (**Sidebar 1.3**). Thus, the extracellular domains of most cell-surface proteins and almost all secreted proteins are glycosylated, having one or more covalently attached sugar side chains; proteins of the Ras family, which are located in the cytoplasm and play important roles in cancer development, contain lipid groups attached to their carboxy termini. An equally important post-translational modification involves the cleavage of one protein by a second protein termed a **protease**, which has the ability to cut amino acid chains at certain sites. Accordingly, the final, mature form of a protein chain may include far fewer amino acid residues than were present in the initially synthesized protein. Following their synthesis, many proteins are dispatched to specific sites within the cell or are exported

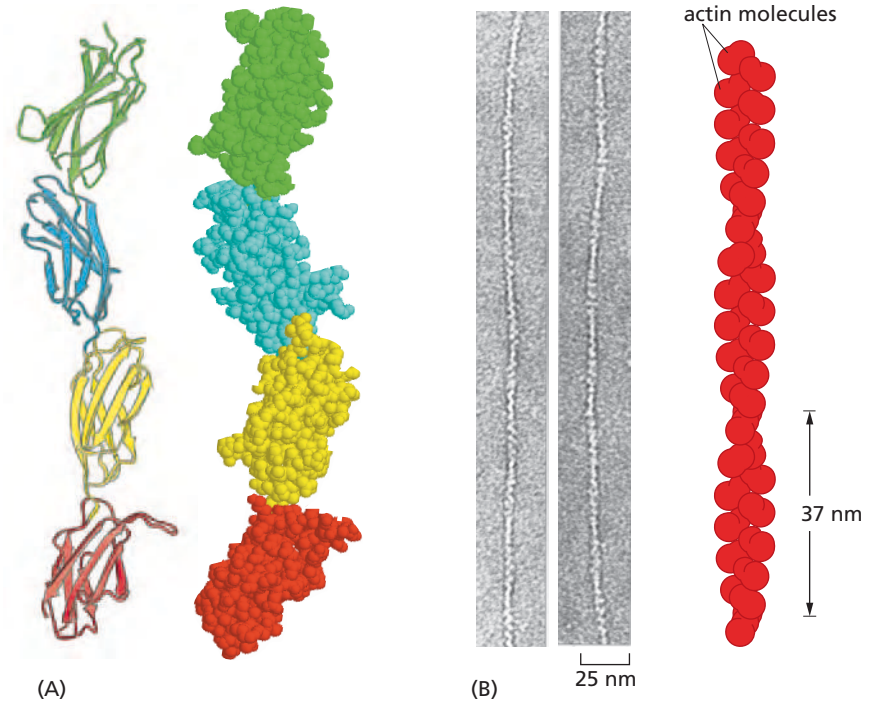


Figure 1.15 Structures of proteins and multiprotein assemblies (A) The three-dimensional structure of part of fibronectin, an important extracellular matrix protein (see Figure 1.13D), is depicted as a ribbon diagram (*left*), which illustrates the path taken by its amino acid chain; alternatively, the space-filling model (*right*) shows the positions of the individual atoms. One portion of fibronectin is composed of four distinct, similarly structured domains, which are shown here with different colors. (B) The actin fibers (*left*), which constitute an important component of the cytoskeleton (see Figures 1.13 and 1.14), are composed of assemblies of individual protein molecules, each of which is illustrated here as a distinct two-lobed body (*right*). (A, adapted from D.J. Leahy, I. Aukhil and H.P. Erickson, *Cell* 84:155–164, 1996. B, left, courtesy of Roger Craig; right, from B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)

from the cell through the process of secretion; these alternative destinations are specified in the newly synthesized proteins by short amino acid (oligopeptide) sequences that function, much like postal addresses, to ensure the diversion of these proteins to specific intracellular sites.

In eukaryotic cells—the main subject of this book—the synthesis of RNA is itself a complex process. An RNA molecule transcribed from its parent gene may initially be almost as long as that gene. However, while it is being elongated, segments of the RNA

Sidebar 1.3 How many distinct proteins can be found in the human body? While some have ventured to provide estimates of the total number of human genes (a bit more than 21,000), it is difficult to extrapolate from this number to the total number of distinct proteins encoded in the human genome. The simplest estimate comes from the assumption that each gene encodes the structure of a single protein. But this assumption is naive, because it ignores the fact that the pre-mRNA transcript deriving from a single gene may be subjected to several *alternative splicing* patterns, yielding multiple, distinctly structured mRNAs, many of which may in turn encode distinct proteins (see Figure 1.16). Thus, in some cells, splicing may include certain exons in the final mRNA molecule made from a gene, while in other cells, these exons may be absent. Such alternative splicing patterns can generate mRNAs having greatly differing structures and protein-encoding sequences. In one, admittedly extreme case, a single *Drosophila* gene has been found to be capable of generating 38,016 distinct mRNAs and thus proteins through various alternative splices of its pre-mRNA; genes having similarly complex alternative splicing patterns are likely to reside in our own genome.

An additional dimension of complexity derives from the post-translational modifications of proteins. The proteins that are exported to the cell surface or released in soluble form into the extracellular space are usually modified by the attachment of complex trees of sugar molecules during the process of **glycosylation**. Intracellular proteins often undergo other types of chemical modifications. Proteins involved in transducing the signals that govern cell proliferation often undergo **phosphorylation** through the covalent attachment of phosphate groups to serine, threonine, or tyrosine amino acid residues. Many of these phosphorylations affect some aspect of the functioning of these proteins. Similarly, the histone proteins that wrap around DNA and control its access by the RNA polymerases that synthesize hnRNA are subject to methylation, acetylation, and phosphorylation, as well as more complex post-translational modifications.

The polypeptide chains that form proteins may also undergo cleavage at specific sites following their initial assembly, often yielding small proteins showing functions that were not apparent in the uncleaved precursor proteins. Later, we will describe how certain signals may be transmitted through the cell via a cascade of the protein-cleaving enzymes termed proteases. In these cases, protein A may cleave protein B, activating its previously latent protease activity; thus activated, protein B may cleave protein C, and so forth. Taken together, alternative splicing and post-translational modifications of proteins generate vastly more distinct protein molecules than are apparent from counting the number of genes in the human genome.

molecule, some very small and others enormous, will be cleaved out of the growing RNA molecule. These segments, termed **introns**, are soon discarded and consequently have no impact on the subsequent coding ability of the RNA molecule (**Figure 1.16**).

Flanking each intron are two retained sequences, the **exons**, which are fused together during this process of **splicing**. The initially synthesized RNA molecule and its derivatives found at various stages of splicing, together with nuclear RNA transcripts being processed from other genes, collectively constitute the **hnRNA** (**heterogeneous**

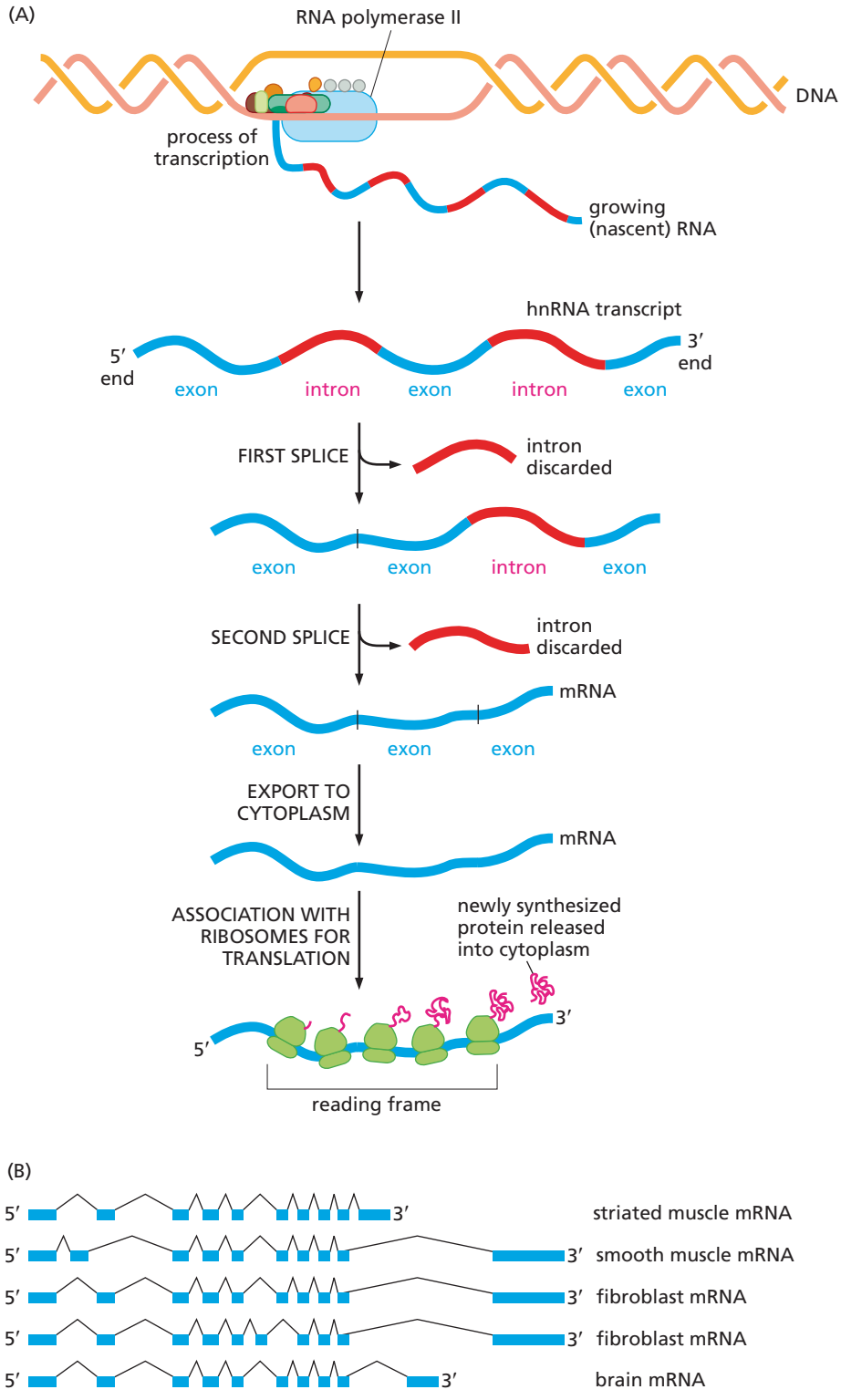


Figure 1.16 Processing of pre-mRNA
(A) By synthesizing a complementary RNA copy of one of the two DNA strands of a gene, RNA polymerase II creates a molecule of heterogeneous nuclear RNA (hnRNA) (*red and blue*). Those hnRNA molecules that are processed into mRNAs are termed pre-mRNA. The progressive removal of the introns (*red*) leads to a processed mRNA containing only exons (*blue*). (B) A given pre-mRNA molecule may be spliced in a number of alternative ways, yielding distinct mRNAs that may encode distinct protein molecules. Illustrated here are the tissue-specific alternative splicing patterns of the α-tropomyosin pre-mRNA molecule, whose mRNA products specify important components of cell (and thus muscle) contractility. In this case, the introns are indicated as *black carets* while the exons are indicated as *blue rectangles*. (B, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)

nuclear RNA). The end product of these post-transcriptional modifications may be an RNA molecule that is only a small fraction of the length of its initially synthesized, hnRNA precursor. This final, mature RNA molecule is likely to be exported into the cytoplasm, where, as an **mRNA** (**messenger RNA**) molecule, it serves as the template on which ribosomes assemble the amino acids that form the proteins. (The term **pre-mRNAs** is often used to designate those hnRNAs that are known precursors of cytoplasmic mRNAs.) Some mature mRNAs may be less than 1% of the length of their pre-mRNA precursor. The complexity of post-transcriptional modification of RNA and post-translational modification of proteins yields an enormous array of distinct protein species within the cell (see Sidebar 1.3).

Of note, an initially transcribed pre-mRNA may be processed through **alternative splicing** into a series of distinct mRNA molecules that retain different combinations of exons (see Figure 1.16B). Indeed, the pre-mRNAs arising from more than 95% of the genes in our genome are subject to alternative splicing. The resulting alternatively spliced mRNAs may carry altered reading frames, explaining, for example, the distinct isoforms of certain proteins that are found in cancer cells but not in their normal counterparts. Alternatively, these splicing events may affect untranslated regions of mRNAs, such as those targeted by microRNAs (miRNAs; Section 1.10); these interactions with miRNAs can alter the function of an mRNA, by regulating either its translation or its stability. Interestingly, a protein that specifies an alternative splicing pattern of pre-mRNAs has been reported, when expressed in excessively high levels in cells, to favor their **transformation** (conversion) from a normal to a cancerous growth state. Such an effect is surprising, since one might imagine that proteins that regulate splicing would mediate the processing of many or all pre-mRNAs within the cell rather than affecting only a subset of genes involved in a specific cell-biological function, such as cell transformation. Moreover, a 2008 survey of alternatively spliced mRNAs found 41 that showed a distinct pattern of alternative splicing in human breast cancer cells compared with normal mammary cells; indeed, these alternatively spliced mRNAs could be used as diagnostic markers of the cancerous state of these cells. Even more dramatic, in 2010 as many as 1000 pre-mRNAs were found to undergo alternative splicing as cells passed through an epithelial–mesenchymal transition (EMT), an important transdifferentiation step that carcinoma cells utilize to acquire traits of high-grade malignancy, as will be discussed in Chapter 14.

1.7 Gene expression patterns also control phenotype

The 21,000 or so genes in the mammalian genome, acting combinatorially within individual cells, are able to create the extraordinarily complex organismic phenotypes of the mammalian body. A central goal of twenty-first-century biology is to relate the functioning of this large repertoire of genes to organismic physiology, developmental biology, and disease development. The complexity of this problem is illustrated by the fact that there are at least several hundred distinct cell types within the mammalian body, each with its own behavior, its own distinct metabolism, and its own physiology.

This complexity is acquired during the process of organismic development, and its study is the purview of developmental biologists. They wrestle with a problem that is inherent in the organization of all multicellular organisms. All of the cells in the body of an animal are the lineal descendants of a fertilized egg. Moreover, almost all of these cells carry genomes that are reasonably accurate copies of the genome that was initially present in this fertilized egg (see Supplementary Sidebar 1.2). The fact that cells throughout the body are phenotypically quite distinct from one another (e.g., a skin cell versus a brain cell) while being genetically identical creates this central problem of developmental biology: how do these various cell types acquire distinct phenotypes if they all carry identical genetic templates? The answer, documented in thousands of ways over the past three decades, lies in the selective reading of the genome by different cell types (**Figure 1.17**).

As cells in the early embryo pass through repeated cycles of growth and division, the cells located in different parts of the embryo begin to assume distinct phenotypes, this being the process of **differentiation**. Differentiating cells become committed to form

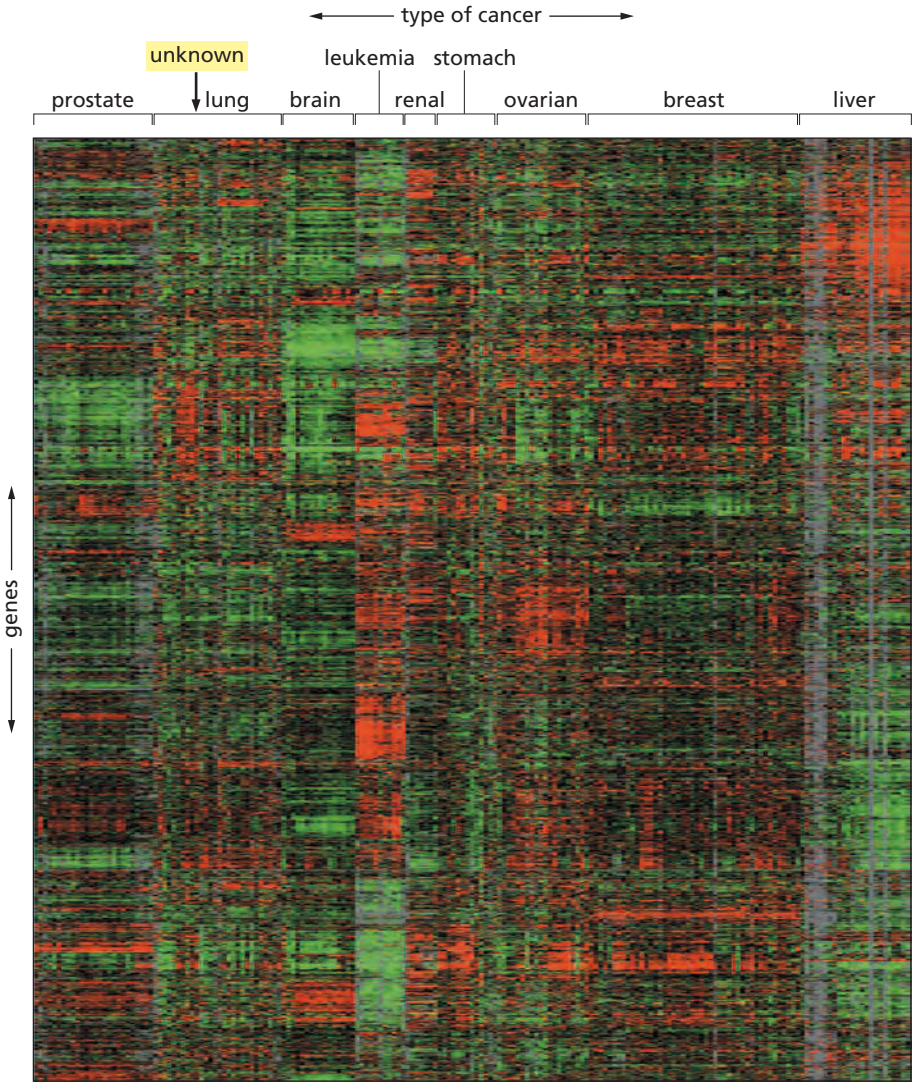


Figure 1.17 Global surveys of gene expression arrays Gene expression microarrays make it possible to survey the expression levels of thousands of genes within a given type of cell. In this image, higher-than-average levels of expression are indicated as *red* pixels, while lower-than-average levels are indicated by *green* pixels. Average-level expression is indicated by *black* pixels. The mRNAs from 142 different human tumors (*arrayed left to right*) were analyzed. In each case, the expression levels of 1800 human genes were measured (*top to bottom*). Each class of tumors has its characteristic spectrum of expressed genes. In this case, a tumor of unknown type (*yellow label*) was judged to be a lung cancer because its pattern of gene expression was similar to those of a series of already-identified lung cancers. (Courtesy of P.O. Brown, D. Botstein and The Stanford Expression Collaboration.)

one type of tissue rather than another, for example, gut as opposed to nervous system. All the while, they retain the same set of genes. This discrepancy leads to a simplifying conclusion: sooner or later, differentiation must be understood in terms of the sets of genes that are expressed (that is, transcribed) in some cells but not in others.

By being expressed in a particular cell type, a suite of genes dictates the synthesis of a cohort of proteins and RNA molecules that collaborate to create a specific cell phenotype. Accordingly, the phenotype of each kind of differentiated cell in the body should, in principle, be understandable in terms of the specific subset of genes that is expressed in that cell type.

The genes within mammalian cells can be grouped into two broad functional classes—the **housekeeping** and the **tissue-specific** genes. Many genes encode proteins that are required universally to maintain viability of all cell types throughout the body or to carry out certain biological functions common to all cell types. These commonly expressed genes are classified as housekeeping genes. Within a given differentiated cell type, housekeeping genes represent the great majority of expressed genes.

A minority of genes within a differentiated cell—the tissue-specific genes—are dedicated to the production of proteins and thus phenotypes that are associated specifically with this cell type. It may be, for example, that 3000–5000 housekeeping genes are expressed by the cell while far fewer than 1000 tissue-specific genes are responsible for the distinguishing, differentiated characteristics of the cell. By implication,

in each type of differentiated cell, a significant proportion of the 21,000 or so genes in the genome are unexpressed, since they are not required either for the cell's specific differentiation program or for general housekeeping purposes.

1.8 Histone modification and transcription factors control gene expression

The foregoing description of differentiation makes it clear that large groups of genes must be coordinately expressed while other genes must be repressed in order for cells to display complex, tissue-specific phenotypes. Such coordination of expression is the job of **transcription factors** (TFs; **Figure 1.18**). Many of these proteins bind to specific DNA sequences in the control region of each gene and determine whether or not the gene will be transcribed. The specific stretch of nucleotide sequence to which the TFs bind, often called a **sequence motif**, is usually quite short, typically 5–10 nucleotides long. In ways that are still incompletely understood at the molecular level, some TFs provide the RNA polymerase enzyme (RNA polymerase II in the case of pre-mRNAs) with access to a gene. Yet other TFs may block such access and thereby ensure that a gene is transcriptionally repressed.

Transcription factors can exercise great power, since a single type of TF can simultaneously affect the expression of a large cohort of downstream responder genes, each of which carries the recognition sequence that allows this TF to bind its promoter (see Figure 1.18). This ability of a single TF (or a single gene that specifies this TF) to elicit multiple changes within a cell or organism is often termed **pleiotropy**. In the case of cancer cells, a single malfunctioning, pleiotropically acting TF may simultaneously orchestrate the expression of a large cohort of responder genes that together proceed to create major components of the cancer cell phenotype. One enumeration of the genes in the human genome that are likely to encode TFs listed 1445 distinct genes (about 7% of the genes carried in the human genome). Not included in this list were variant versions of these proteins arising through alternative splicing of pre-mRNAs.

The transcription of most genes is dependent upon the actions of several distinct TFs that must sit down together, each at its appropriate sequence site (that is, **enhancer**) in or near the gene promoter, and collaborate to activate gene expression. This means that the expression of a gene is most often the result of the combinatorial actions of several TFs. Therefore, the coordinated expression of multiple genes within a cell, often called its gene **expression program**, is dependent on the actions of multiple TFs acting in combination on large numbers of gene promoters.

Figure 1.18 implies that modulation of gene expression is achieved by controlling initiation of transcription by RNA polymerase II (pol II) and that transcription proceeds in one direction. In fact, for many genes, possibly the majority, pol II molecules sit

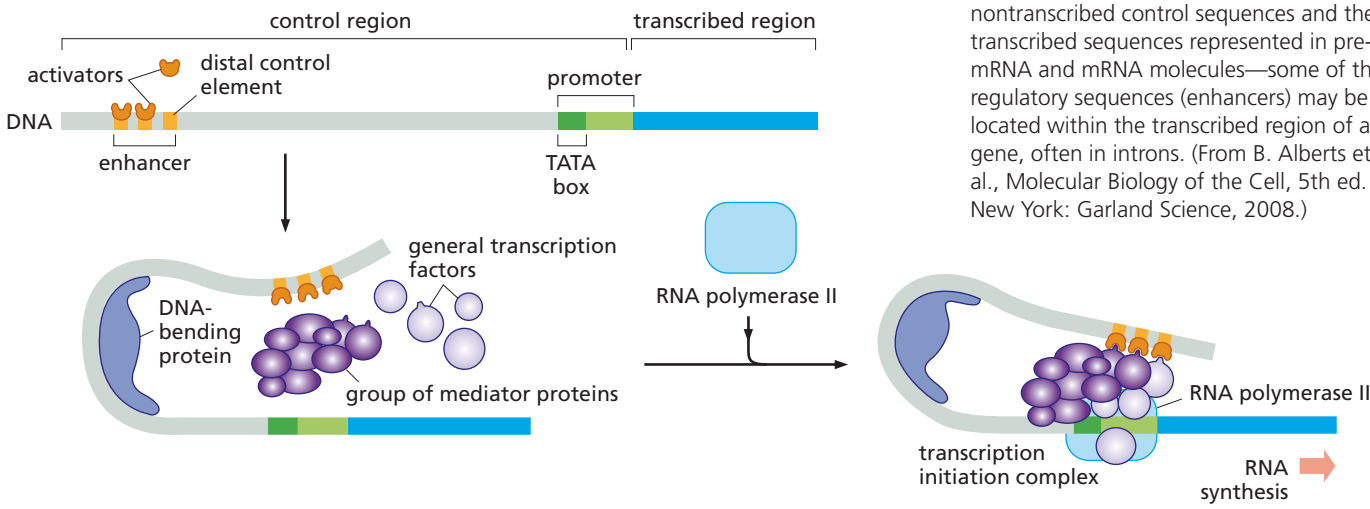


Figure 1.18 Regulation of gene expression The control region of a gene includes specific segments of DNA to which gene regulatory proteins known as transcription factors (TFs) bind, often as multiprotein complexes; in this case TFs, functioning as activators (*light brown*), bind to *enhancer* sequences (*orange*) located some distance upstream of the promoter. In addition, the *promoter* of the gene (*dark, light green*) contains sequences to which RNA polymerase II (pol II) can bind, together with associated general transcription factors. The bound TFs, interacting with the transcription initiation complex via *mediator* proteins, influence the structure of chromatin (notably the histone proteins that package DNA; see Figures 1.19 and 1.20), creating a localized chromatin environment that enables pol II to produce an RNA transcript (*orange-red arrow*). (The general TFs are involved in initiating the transcription of many genes throughout the genome, while the specialized ones regulate the expression of subsets of genes.) Although in general a gene can be separated into two functionally significant regions—the nontranscribed control sequences and the transcribed sequences represented in pre-mRNA and mRNA molecules—some of the regulatory sequences (enhancers) may be located within the transcribed region of a gene, often in introns. (From B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)

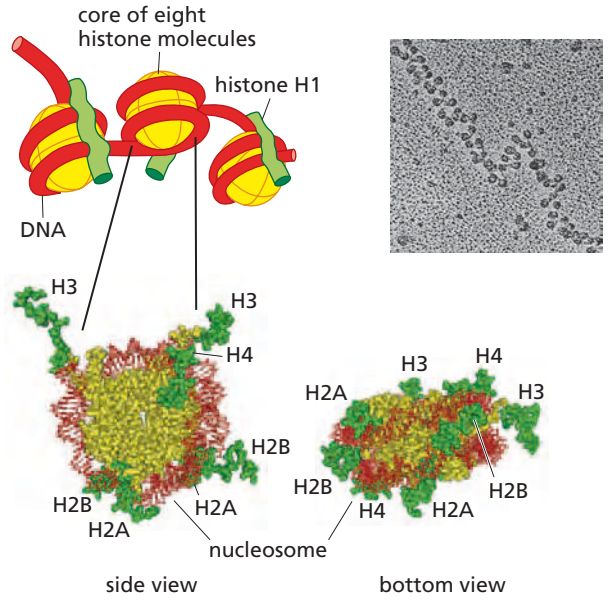


Figure 1.19 Organization of chromatin structure Examination of chromatin under the electron microscope (*above right*) reveals that DNA is associated with small globes of proteins termed nucleosomes, giving the appearance of beads on a string. The DNA double-helix (*above left, red*) is wrapped ~1.7 times around each nucleosome, which consists of a core (*yellow*) formed as an octamer of four different histone molecules (each present in two copies); often an additional histone, H1 (*green*), is located on the outside. X-ray crystallography has revealed (*below*) that the core of the nucleosome (*yellow*) is disc-shaped and that the N-terminal tail (*green*) of each of the four histones extends beyond this core. (Upper schematic, from W.K. Purves et al., Life: The Science of Biology, 5th ed. Sunderland, MA: Sinauer, 1998. Lower schematic from B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008. Micrograph from F. Thoma, T. Koller and A. Klug, *J. Cell Biol.* 83:403–427, 1979.)

down on the promoter of a gene and proceed to transcribe the DNA in both directions. After extending nascent RNA transcripts for 60–80 nucleotides, pol II halts—the process termed **transcriptional pausing.** A subset of the stalled polymerase complexes that have initiated in the appropriate transcriptional direction are then induced by physiologic signals to resume elongation, resulting in full-length pre-mRNA transcripts, while other pol II complexes remain stalled and never resume transcription. The factors that permit stalled pol II to proceed with elongation of transcripts are incompletely understood but would seem to be as important as the conventionally defined TFs in regulating gene expression. One important cancer-causing protein, termed Myc, has been found to act as an anti-pausing protein whose actions permit thousands of cellular genes to be fully transcribed.

Figure 1.18 also implies that both TFs and RNA polymerase interact only with DNA. In fact, in eukaryotic cells, DNA is packaged in a complex mixture of proteins that, together with the DNA, form the **chromatin** (**Figure 1.19**). These chromatin proteins are responsible for controlling the interactions of TFs and RNA polymerases with DNA and therefore play critical roles in governing gene expression.

The core of chromatin is formed by DNA bound to nucleosomes, the latter being octamers consisting of two copies of each of four distinct histone species (H2A, H2B, H3, and H4) with a fifth histone species—H1—bound to some but not all nucleosome octamers. This basic organization of chromatin structure, which resembles beads on a string, is found throughout the chromosomes.

The globular core of the nucleosome represents the basic scaffold of chromatin that is modified in two ways. First, some of the standard histones, such as histones H2A and H3, may be replaced in a minority of nucleosomes by variant forms, for example, histones H2AZ and H3.3 (specified by genes distinct from those encoding the standard histones). Indeed, a number of such variant histones can be found scattered here and there throughout the chromatin; their precise contributions to the regulation of chromatin structure and transcription remain poorly understood.

Second, chromatin structure and transcription is strongly affected by post-translation modifications of the standard four histones. These modifications do not directly alter the globular core of the nucleosome. Instead, they affect the N-terminal tails of the core histones (**Figure 1.20A**), which extend outward from the globular core and undergo a variety of covalent modifications, prominent among these being methylation, acetylation, phosphorylation, and ubiquitylation. For example, one type of histone phosphorylation is associated with the condensation of chromatin that occurs during mitosis and the related global shutdown of gene expression. At other times in the cell cycle, acetylation of core histones is generally associated with active gene

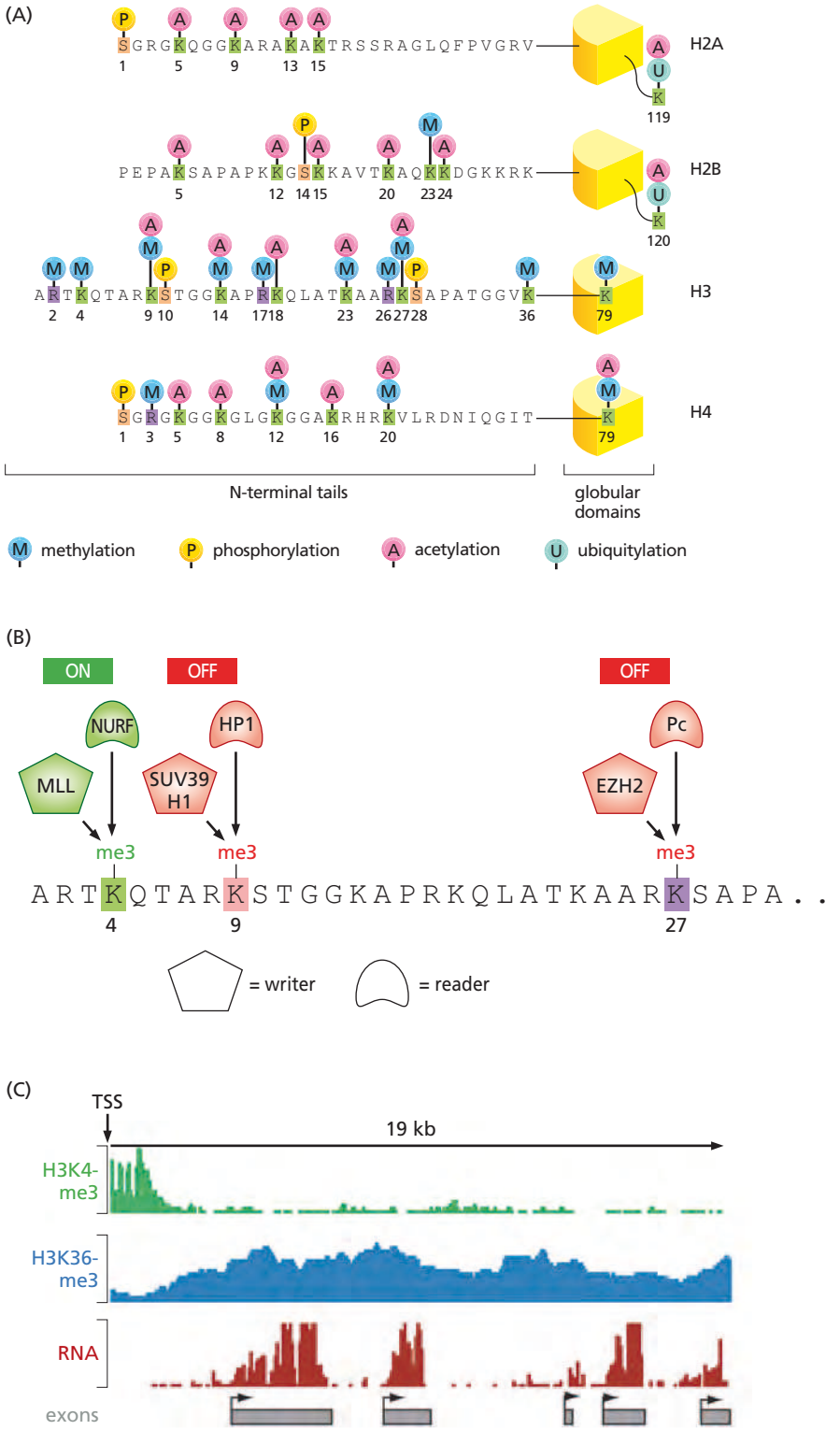


Figure 1.20 Post-translational modification of histone tails (A) Each of these N-terminal histone tails can be modified by the covalent attachment of a variety of chemical groups, most commonly methyl, acetyl, phosphate, and ubiquitin groups. These modifications are attached by histone "writers," which thereby alter the structure and the function of the chromatin, and are removed by histone "erasers." (B) One example of histone modification is provided by three of the lysine (K) residues in the amino-terminal domain of histone H3. (Amino terminus is at *left*; numbers below each K indicate residue number.) Each of these can be trimethylated (indicated by "me3") through the actions of histone methyltransferase writers (HMTs). Trimethylation at the K4 residue is carried out by the MLL1 HMT; the resulting methyl mark is recognized by a NURF (nucleosome remodeling factor) "reader" complex, which contributes to gene activation (*green*). Conversely, trimethylation of the K9 and K27 residues by the SUV39H1 and EZH2 HMT writers, respectively, results in gene repression (*red*). The methylation marks made by the latter two HMTs are recognized by the HP1 and Pc readers, respectively. Once bound, the HP1 reader can trigger the formation of heterochromatin and thereby block transcription. Not shown are other methyltransferase writers that make mono- and dimethyl marks, and histone demethylase erasers that remove the marks made by HMTs on these residues. (C) The locations of various modified histones can be mapped across a gene by using an antibody that specifically immunoprecipitates a modified histone species followed by DNA sequencing of the precipitate. In this fashion, the locations of the nucleosomes containing trimethylated lysine 4 of histone H3 (H3K4me3, *green*) and H3K36me3 (*blue*) have been mapped relative to the transcription start site (TSS) of this gene. Correlations like these indicate that nucleosomes containing H3K4me3 are associated with TSSs, while those containing H3K36me3 are found along the lengths of actively transcribed genes. When the RNA molecules are analyzed (*red*), those that map to known exons of the gene are found in greater abundance, consistent with their long lifetime relative to the short lifetimes of rapidly degraded intron sequences. The function of the gene studied here is not known. (A, from H. Santos-Rosa and C. Caldas, *Eur. J. Cancer* 41:2381–2402, 2005. B, from S.B. Hake, A. Xiao and C.D. Allis, *Brit. J. Cancer* 90:761–769, 2007. C, from M. Guttman et al., *Nature* 458:223–227, 2009.)

expression, while methylation is generally correlated with gene repression. However, as is seen in Figure 1.20B, which presents only one example of a bewildering variety of histone modifications, methylation of histone H3 is correlated with both gene repression and expression, depending on the position of the affected lysine residue.

Rapidly growing evidence indicates that these various histone modifications are functionally important in permitting or preventing transcription by RNA polymerases of specific regions of chromosomal DNA (see Figure 1.20C). Moreover, the modification

state of chromatin can be passed from mother to daughter cells through mechanisms that are still unresolved. This area of research is in great flux: as many as 60 distinct histone-modifying enzymes have been discovered, whose roles in transcriptional regulation and cell biology are largely obscure, and there are likely an even larger number of proteins that form complexes with these enzymes and direct them toward distinct substrates within the chromatin. As more effective sequencing techniques are applied to cancer cell genomes, mutant alleles of the genes encoding these enzymes are being uncovered with ever-increasing frequency.

1.9 Heritable gene expression is controlled through additional mechanisms

The descriptions above of the mechanisms controlling gene expression provide only a partial explanation of how gene expression programs that are established in one human cell are transmitted to its lineal descendants. For example, the specific gene expression program of a fibroblast grown in culture will continue to be expressed by its lineal descendants 10 and 20 cell generations later. Since decisions to express or repress a gene within a fibroblast are not imprinted in the gene's DNA sequence, this implies alternative means of maintaining such decisions in a stable fashion and transmitting them faithfully from one cell generation to the next via biochemical mechanisms that mediate **epigenetic** inheritance.

In addition to the transmission of histone modifications described above, the other key mechanism that enables epigenetic inheritance of gene expression depends on covalent modification of DNA, specifically by DNA methyltransferases—enzymes that attach methyl groups directly to cytosine bases of CpG dinucleotides in the DNA double helix. (The designation CpG indicates that the sequence is a cytidine positioned 5′ immediately before a guanosine.) The affected CpG dinucleotides are often located near transcriptional promoters, and the resulting methylation generally causes repression of nearby genes. The biochemical mechanism of maintenance methylation is well understood: maintenance DNA methyltransferase enzymes recognize **hemi-methylated** segments of recently replicated DNA and proceed to methylate any unmethylated CpG dinucleotides that are complementary to already methylated CpGs in the other DNA strand (**Figure 1.21**).

The mechanism(s) that lead to *de novo* methylation of previously unmethylated CpGs are still elusive. However, recent research reveals how the reverse process occurs: The Tet (ten eleven translocation) enzymes oxidize the methyl group of 5-methyl-cytidine to hydoxymethyl, formyl, and carboxy groups. The altered nucleotides may then be excised by DNA repair enzymes (Chapter 12) and replaced by cytidine; alternatively, when DNA bearing an oxidized cytidine is replicated, the maintenance methylase may fail to methylate the complementary strand. This research has not yet identified how the Tet enzymes are controlled.

The methyl CpG groups do not, on their own, directly block transcription. Instead, they appear to affect the structure of the chromatin proteins that are responsible for packaging chromosomal DNA and presenting it to RNA polymerases for transcription,

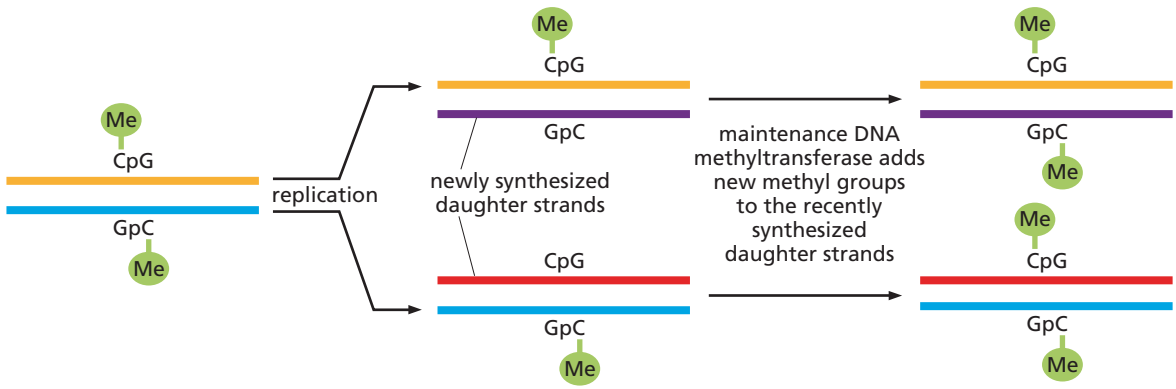


Figure 1.21 Maintenance of DNA methylation following replication When a DNA double helix that is methylated (*green* Me groups, *left*) at complementary CpG sites undergoes replication, the newly synthesized daughter helices will initially lack methyl groups attached to CpGs in the recently synthesized daughter strands (*purple, red*) and will therefore be *hemi-methylated*. Shortly after their synthesis, however, a maintenance DNA methyltransferase will detect the hemi-methylated DNA and attach methyl groups (*green*) to these CpGs, thereby regenerating the same configuration of methyl groups that existed in the parental helix prior to replication. CpG sites that are unmethylated in the parental helix (*not shown*) will be ignored by the maintenance methyltransferase and will therefore remain so in the newly synthesized strands.

as described above. In particular, methyl-CpG–binding proteins associate specifically with methylated dinucleotides and influence the structure of the nearby chromatin in still-poorly understood ways. There is also evidence that the modification of certain histones can operate in the opposite direction to influence the state of DNA methylation.

1.10 Unconventional RNA molecules also affect the expression of genes

The Central Dogma of molecular biology, developed in the decade after the 1953 discovery of the DNA double helix, proposed that information flows in cells from DNA via mRNA to proteins. In addition, non-informational RNA molecules—ribosomal and transfer RNAs—were implicated as components of the translational machinery, and small nuclear RNAs were found to play key roles in the splicing and maturation of pre-mRNAs. In the 1980s, the view of RNA's functions was expanded through the discovery that certain RNA species can act as enzymes, thereby taking their place alongside proteins as catalysts of certain biochemical reactions.

The 1990s revealed an entirely new type of RNA molecule that functions to control either the levels of certain mRNAs in the cytoplasm, the efficiency of translating these mRNAs, or both. These **microRNAs** (miRNAs) are only 21 to 25 nucleotides long and are generated as cleavage products of far larger nuclear RNA precursors. As outlined in **Figure 1.22**, the post-transcriptional processing of a primary miRNA transcript results in the formation in the cytoplasm of a miRNA that is part of a RISC (RNA-induced silencing complex) nucleoprotein. This complex associates with a spectrum of mRNA

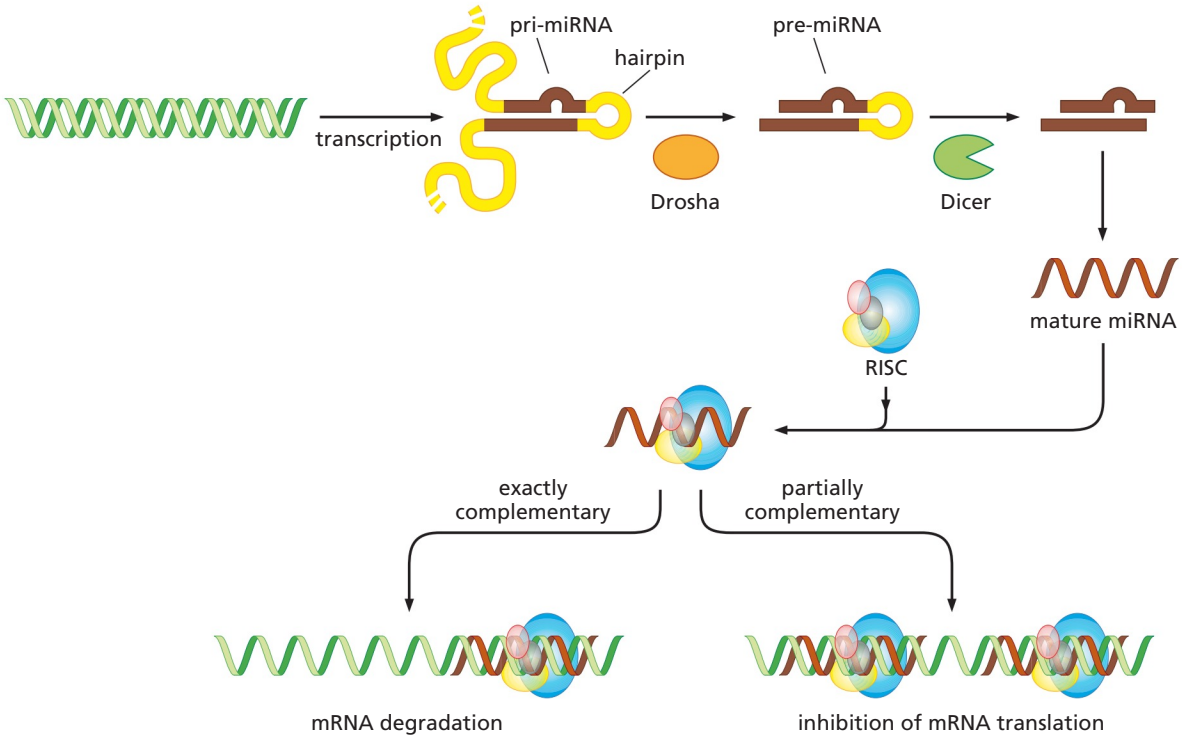


Figure 1.22 MicroRNAs and gene regulation A primary microRNA (pri-miRNA) is transcribed from a gene, and an enzyme complex involving the Drosha protein excises a small segment of the pri-miRNA that has formed a double-stranded RNA hairpin because of the self-complementarity of nucleotide sequences. The resulting pre-miRNA is exported to the cytoplasm, where it is further processed by the Dicer enzyme to generate a mature miRNA of 21 to 25 nucleotides. This miRNA binds to a nucleoprotein complex termed RISC (RNA-induced silencing complex) and associates with mRNAs in the cytoplasm with which it has precise or partial sequence complementarity, resulting in either degradation of the mRNA or inhibition of its translation. Several dozen miRNAs have been found to regulate various steps of tumor formation, either favoring or blocking critical steps of this process. Loss of the Dicer enzyme has been associated with cancer progression, and analyses of miRNA expression patterns, much like expression array analyses of mRNAs (see Figure 1.17), have proved useful in classifying various types of cancer. (Courtesy of P.A. Sharp.)

targets that contain, usually in their untranslated region, a sequence that is partially or completely complementary to the miRNA in the complex. Such association can result in either the inhibition of translation of the mRNA or its degradation, or both.

More than 650 distinct miRNA species have been found in human cells, and this roster continues to grow. Although it is unclear how many of these miRNAs are actually involved in regulating the translation and stability of mRNAs, those that do affect mRNA function are thought to regulate expression of at least one-third of all genes in the human genome. Moreover, a single miRNA species can target and thus regulate the expression of dozens of distinct mRNA species, enabling it to act pleiotropically on a variety of cellular processes.

The potential importance of miRNAs in regulating gene expression is suggested by one survey of mRNAs and corresponding proteins in a group of 76 lung cancers. Only about 20% of the genes studied showed a close correlation between mRNA expression and protein expression levels. Hence, in the remaining 80%, the rate of protein synthesis (which can be strongly influenced by miRNAs) and the post-translational lifetime of proteins (see Supplementary Sidebar 7.4) strongly influenced actual protein levels. Since proteins, rather than mRNA, are responsible for creating cell phenotypes, this also reveals the limitations of studying mRNA levels as indicators of gene activity.

Let-7, an miRNA expressed by the *C. elegans* worm, was one of two initially characterized miRNAs. It was found to suppress expression of the *ras* gene in worms and later in mammals. As we will read later (Chapters 4 through 6), the Ras proteins play critical roles in the development of many types of common human cancers. Since this pioneering work, the overexpression or loss of more than a dozen miRNA species has been associated with the formation of a variety of human cancers and the acquisition by tumors of malignant traits. The list of these miRNAs, which have garnered the term "oncoMiRs," continues to lengthen (see Supplementary Sidebar 1.3). In addition, loss of the Dicer processing enzyme (see Figure 1.22), which is involved in creating mature miRNA, has been found to facilitate the formation of tumors in mice, doing so through still-unknown mechanisms. Interestingly, inheritance of a variant of the *K-ras* gene, which causes a single nucleotide change in the 3′ untranslated region (3′ UTR) of its mRNA, prevents recognition by *Let-7* and is associated with higher levels of the growth-promoting K-Ras protein and as much as a twofold increased risk of certain forms of lung and ovarian cancers.

A decade after the discovery of microRNAs, yet another unusual class of RNAs appeared on the scene: a diverse assay of lncRNA molecules (long non-coding RNAs) were found in the nucleus and cytoplasm to be involved in still-poorly understood ways in regulating gene expression. The discovery of these came from the realization that 4 to 9% of the human genome is transcribed into relatively long (>200 nucleotide) RNA molecules that have no identifiable protein-coding sequences and thus no readily ascertainable functions. Some lncRNAs are polyadenylated while others are not. The few lncRNAs that have been characterized seem to function by associating with proteins that are involved in one fashion or another in regulating transcription, often by serving as scaffolds to hold certain chromatin-modifying proteins together. There may be several thousand distinct lncRNA species encoded by the human genome and they are increasingly viewed as key molecular components of the cell's regulatory machinery.

The role of lncRNAs in cancer development is only beginning to be uncovered. For example, elevated expression of the *HOTAIR* lncRNA has been found to be correlated with metastatic behavior of human breast and colorectal carcinomas. More importantly, forced expression of *HOTAIR* in carcinoma cells causes localization of a transcription-repressing protein complex, termed PRC2, to certain chromosomal sites, altered methylation of histone H3 lysine 27 (see Figure 1.20), and increased cancer invasiveness and metastasis.

The actions of miRNAs and lncRNAs provide a glimpse of the complexity of gene expression and its regulation in mammalian cells. Thus, after the transcription of a gene is permitted, a number of mechanisms may then intervene to control the

accumulation of its ultimate product—a protein that does the actual work of the gene. Among these mechanisms are (1) post-transcriptional processing of pre-mRNA transcripts, including alternative splicing patterns; (2) stabilization or degradation of the mRNA product; (3) regulation of mRNA translation; and (4) post-translational modification, stabilization, or degradation of the protein product. These mechanisms reinforce the notion, cited above, that the rate of transcription of a gene often provides little insight into the levels of its protein product within a cell. Hence, as we will see, distinct patterns of mRNA expression may help us to distinguish various neoplastic cells from one another but, on their own, tell us rather little about how these cells are likely to behave.

1.11 Metazoa are formed from components conserved over vast evolutionary time periods

These descriptions of cell biology, genetics, and evolution are informed in part by our knowledge of the history of life on Earth. Metazoa probably arose only once during the evolution of life on this planet, perhaps 700 million years ago. Once the principal mechanisms governing their genetics, biochemistry, and embryonic development were developed, these mechanisms remained largely unchanged in the descendant organisms up to the present (**Figure 1.23**; see also Figure 1.7). This sharing of conserved traits among various animal phyla has profound consequences for cancer research, since many lessons learned from the study of more primitive but genetically tractable organisms, such as flies and worms, have proven to be directly transferable to our understanding of how mammalian tissues, including those of humans, develop and function.

Upon surveying the diverse organisms grouped within the mammalian class, one finds that the differences in biochemistry and cell biology are minimal. For this reason, throughout this book we will move effortlessly back and forth between mouse biology and human biology, treating them as if they are essentially identical. On occasion, where species-specific differences are important, these will be pointed out.

The complex signaling circuits operating within cells seem to be organized in virtually identical fashion in all types of mammals. Even more stunning is the interchangeability of the component parts. It is rare that a human protein cannot function in place of its counterpart *orthologous* protein (**Sidebar 1.4**) in mouse cells. In the case of many types of proteins, this conservation of both function and structure is so profound that proteins can be swapped between organisms that are separated by far greater evolutionary distances. A striking example of this, noted earlier (see Figure 1.7), is provided by the gene and thus protein that specifies eye formation in mammals and in flies. Extending even further back in our evolutionary history are the histones and the mechanisms of chromatin remodeling discussed earlier. In fact, the counterparts of many molecules and biochemical mechanisms that operate in mammalian cells are already apparent in protozoa.

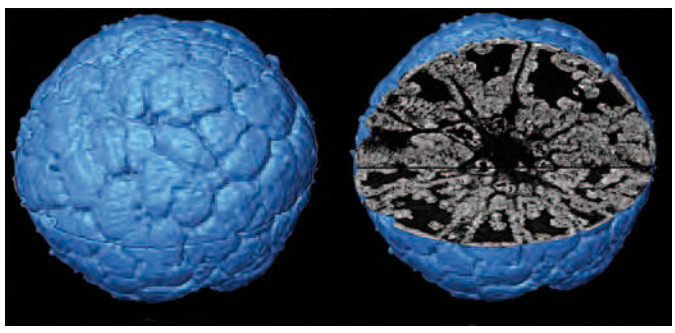


Figure 1.23 Visual evidence of the conservation of metazoan biological traits A stunning visual demonstration that contemporary metazoa develop through pathways that have changed little since the Cambrian era has come from the use of synchrotron-generated X-rays to visualize microscopic fossils at sub-micron resolution, yielding this image of an early Cambrian (~530 million years ago) blastula related either to the modern cnidarian or arthropod phylum. Its resemblance to the blastulas of contemporary metazoa indicates that, in addition to conserved molecular and biochemical mechanisms, certain features of embryonic development have changed relatively little since the emergence of modern metazoan phyla during the Cambrian era. Both the surface (*left*) and the interior cleavage pattern (*right*) are shown. (From P.C.J. Donoghue et al., *Nature* 442:680–683, 2006.)

Sidebar 1.4 Orthologs and homologs All higher vertebrates (birds and mammals) seem to have comparable numbers of genes—in the range of 21,000. Moreover, almost every gene present in the bird genome seems to have a closely related counterpart in the human genome. The correspondence between mouse and human genes is even stronger, given the closer evolutionary relatedness of these two mammalian species.

Within the genome of any single species, there are genes that are clearly related to one another in their information content and in the related structures of the proteins they specify. Such genes form a **gene family**. For example, the group of genes in the human genome encoding globins constitutes such a group. It is clear that these related genes arose at some point in the evolutionary past through repeated cycles of the process in which an existing gene is duplicated followed by the divergence of the two duplicated nucleotide sequences from one another (**Figure 1.24**). More directly related to cancer development are the more than 500 protein kinases encoded by the human genome. Kinases attach phosphate groups to their protein substrates, and almost all of these enzymes are specified by members of a single gene family that underwent hundreds of cycles of gene duplication and divergence during the course of evolution (see Supplementary Figure 16.5).

Genes that are related to one another within a single species' genome or genes that are related to one another in the genomes of two distinct species are said to be **homologous** to one another. Often the precise counterpart of a gene in a human can be found in the genome of another species. These two closely related genes are said to be **orthologs** of one another. Thus, the precise counterpart—the ortholog—of the c-*myc* gene in humans is the c-*myc* gene in chickens. To the extent that there are other *myc*-like genes harbored by the human genome (that is, N-*myc* and L-*myc*), the latter are members of the same gene family as c-*myc* but are not orthologs of one another or of the c-*myc* gene in chickens.

Throughout this book we will often refer to genes without making reference to the species from which they were isolated. This is done consciously, since in the great majority of cases, the functioning of a mouse gene (and encoded protein) is indistinguishable from that of its human or chicken ortholog.

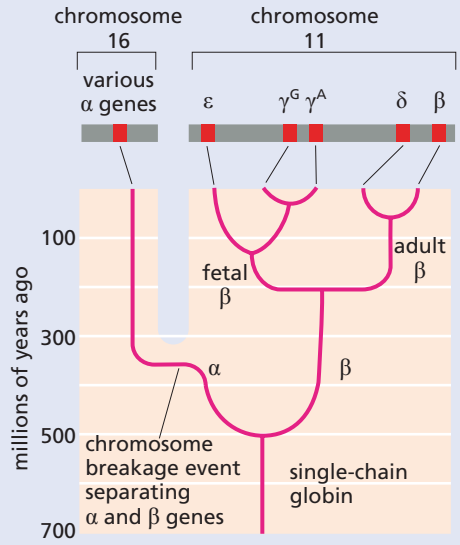


Figure 1.24 Evolutionary development of gene families The evolution of organismic complexity has been enabled, in part, by the development of increasingly specialized proteins. New proteins are "invented" largely through a process of gene duplication followed by diverging evolution of the two resulting genes. Repeated cycles of such gene duplications followed by divergence have led to the development of large numbers of multi-gene families. During vertebrate evolution, an ancestral globin gene, shown here, which encoded the protein component of hemoglobin, was duplicated repeatedly, leading to the large number of distinct globin genes in the modern mammalian genome that are present on two human chromosomes. Because these globins have distinct amino acid sequences, each can serve a specific physiologic function. (From B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)

1.12 Gene cloning techniques revolutionized the study of normal and malignant cells

Until the mid-1970s, the molecular analysis of mammalian genes was confined largely to the genomes of DNA tumor viruses, indeed the viruses described later in Chapter 3. These viruses have relatively simple genomes that accumulate to a high copy number (that is, number of molecules) per cell. This made it possible for biologists to readily purify and study the detailed structure and functioning of viral genes that operate much like the genes of the host cells in which these viruses multiplied. In contrast, molecular analysis of cellular genes was essentially impossible, since there are so many of them (tens of thousands per haploid genome) and they are embedded in a genome of daunting complexity (~3.2 billion base pairs of DNA per haploid cellular genome).

All this changed with the advent of gene cloning. Thereafter, cellular genomes could be fragmented and used to create the collections of DNA fragments known as genomic **libraries**. Various DNA hybridization techniques could then be used to identify the genomic fragments within these libraries that were of special interest to the experimenter, in particular the DNA fragment that carried part or all of a gene under study. The retrieval of such a fragment from the library and the amplification of this

retrieved fragment into millions of identical copies yielded a purified, **cloned** fragment of DNA and thus a cloned gene (see Supplementary Sidebar 1.4). Yet other techniques were used to generate DNA copies of the mRNAs that are synthesized in the nucleus and exported to the cytoplasm, where they serve as the templates for protein synthesis. Discovery of the enzyme **reverse transcriptase** (RT; see Figure 3.18) was of central importance here. Use of this enzyme made it possible to synthesize ***in vitro*** (that is, in the test tube) **complementary DNA** copies of mRNA molecules. These DNA molecules, termed cDNAs, carry the sequence information that is present in an mRNA molecule after the process of splicing has removed all introns. While we will refer frequently throughout this book to DNA clones of the genomic (that is, chromosomal) versions of genes and to cDNAs generated from the mRNA transcripts of such genes, space limitations preclude any detailed descriptions of the cloning procedures *per se*.

For cancer researchers, gene cloning arrived just at the right time. As we will see in the next chapters, research in the 1970s diminished the candidacy of tumor viruses as the cause of most human cancers. As these viruses moved off center stage, cellular genes took their place as the most important agents responsible for the formation of human tumors. Study of these genes would have been impossible without the newly developed gene cloning technology, which became widely available in the late 1970s, just when it was needed by the community of scientists intent on finding the root causes of cancer.

Additional reading

Amaral PP, Diner ME, Mercer TR & Mattick JS (2008) The eukaryotic genome as an RNA machine. *Science* 319, 1787–1789.

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* 116, 281–297.

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.

Bhaumik SR, Smith E & Shilatifard A (2007) Covalent modifications of histones during development and disease pathogenesis. *Nat. Struct. Mol. Biol.* 14, 1008–1016.

Cedar H & Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10, 295–304.

Chi P, Wang GG & Allis CD (2010) Covalent histone modifications: miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* 10, 457–469.

Deng S, Calin GA, Croce CM et al. (2008) Mechanisms of microRNA deregulation in human cancer. *Cell Cycle* 7, 2643–2646.

Esquela-Kerscher A & Slack FJ (2006) Oncomirs: microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269.

Esteller M (2008) Epigenetics in cancer. *N. Engl. J. Med.* 358, 1148–1159.

Fernandez AF, Assenov Y & Martin-Subero J (2011) A DNA methylation fingerprint of 1,628 human samples. *Genome Res.* 22, 407–419.

Frese KK & Tuveson DA (2007) Maximizing mouse cancer models. *Nat. Rev. Cancer* 7, 654–658.

Füllgrabe J, Kavanagh E & Joseph B (2011) Histone onco-modifications. *Oncogene* 30, 3391–3403.

Gancz D & Fishelson Z (2009) Cancer resistance to complement-dependent cytotoxicity (CDC): Problem-oriented research and development. *Mol. Immunol.* 46, 2794–2800.

Kim E, Goren A & Ast G (2008) Insights into the connection between cancer and alternative splicing. *Trends Genet.* 24, 7–10.

Lander ES (2011) Initial impact of sequencing the human genome. *Nature* 470, 187–197.

Melo SA & Esteller M (2010) Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett.* 13, 2087–2099.

Nabel CS, Manning SA & Kohli RM (2012) The curious chemical biology of cytosine: deamination, methylation, and oxidation of genomic potential. *ACS Chem. Biol.* 7, 20–23.

Ponting CP, Oliver PL & Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641.

Rodriguez-Paredes M & Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nature Med.* 17, 330–339.

Sparman A & van Lohuizen M (2006) Polycomb silencers control cell fate, development, and cancer. *Nat. Rev. Cancer* 6, 846–856.

Ting AH, McGarvey KM & Baylin SB (2006) The cancer epigenome—components and functional correlates. *Genes Dev.* 20, 3215–3231.

Ventura A & Jacks T (2009). MicroRNAs and cancer: short RNAs go a long way. *Cell* 136, 586–591.

Wang GG, Allis CD & Chi P (2007) Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol. Med.* 13, 363–372.

This page intentionally left blank
to match pagination of print book