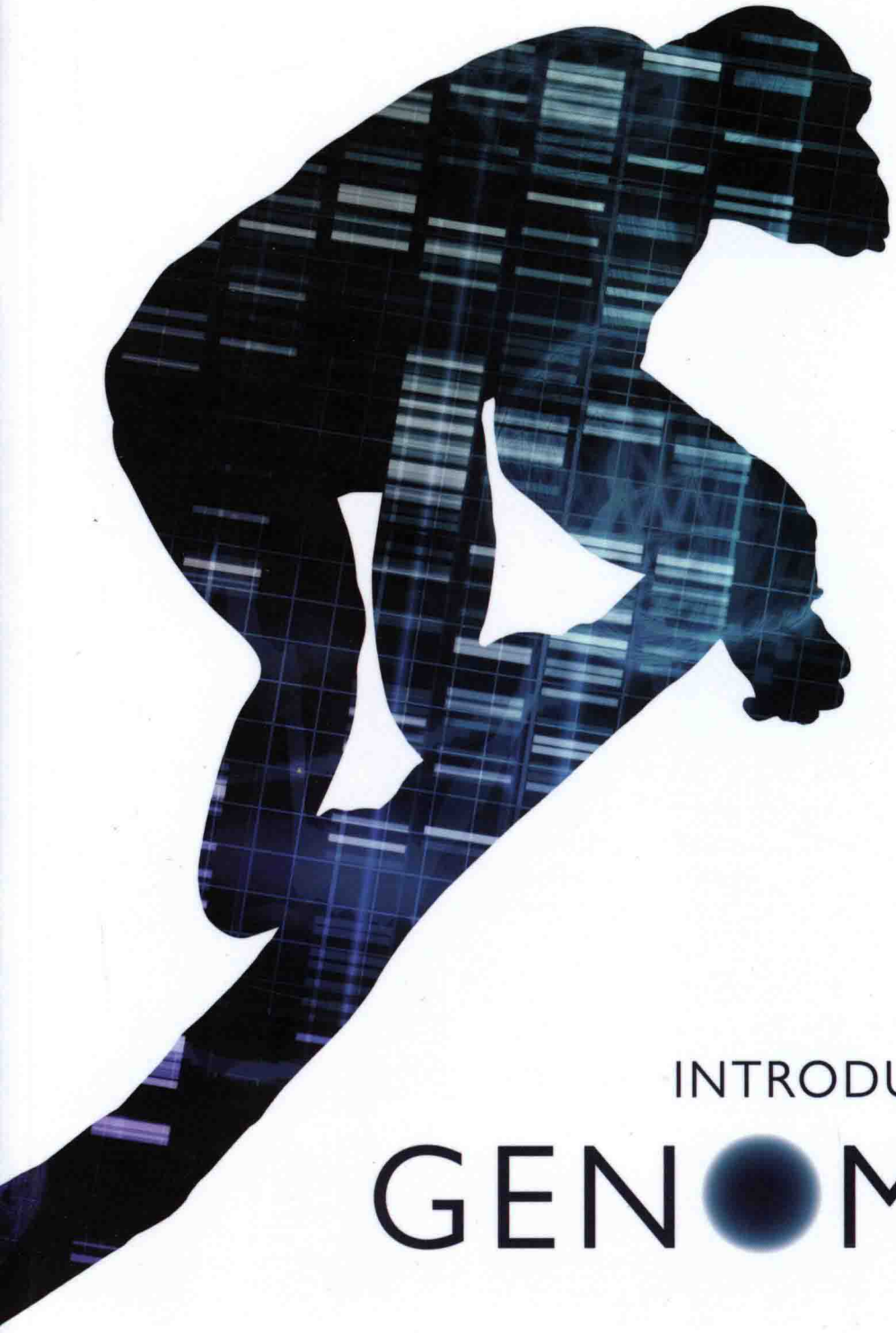
third edition

OXFORD



INTRODUCTION TO GENOMICS

ARTHUR M. LESK

'[*Introduction to Genomics*] is engaging and easy to read . . . I particularly appreciated the "real-world" features such as the coverage of patents.'

Associate Professor David Studholme, University of Exeter

'As in previous editions, the range of ingenious exercises and problems are one of the outstanding features of this textbook.'

Professor Adrian Slater, De Montfort University

'Still on top of the handful of genomics texts.'

Professor Michael Shiaris, University of Massachusetts Boston

This fully updated and restructured new edition takes account of new developments and technologies to present a logical and coherent overview of genome science today. Covering the similarities and differences between organisms; how different organisms evolved; how the genome is constructed and how it operates; and what our understanding of genomics means in terms of our future health and wellbeing, *Introduction to Genomics* is the most up-to-date and complete textbook for students approaching the subject for the first time.

Key features:

- The author's widely-praised writing style leads you through a conceptually challenging subject in a clear and lucid way.
- Numerous learning features throughout help you master the subject by encouraging and supporting independent study.
- Broad and fascinating range of examples ground the study of genomics in the 'real world', inspiring you to apply your knowledge of the subject more widely.

New to this edition:

- Two new chapters, focused on human genomics and applications to health and disease, lay out the concrete outcomes of the human genome project and the most recent advances in cancer genomics.
- An expanded introductory chapter 'sets the scene' for your first encounter with the subject.
- A new 'Looking forward' feature links the topics covered in each chapter to help you make connections between seemingly different topics.
- A glossary of key terms defines all the concepts and abbreviations you need to know.

online resource centre
www.oxfordtextbooks.co.uk/orc/leskgenomics3e/

The Online Resource Centre to accompany *Introduction to Genomics* features:

For students

- Imaginative 'weblems' (web-based problems) for each chapter give you practice with the tools required for further study and research in the field.
- Hints and answers to end-of-chapter exercises and problems support your self-directed learning.
- Guided tour of websites and major archival databanks in genomics offers a wealth of resources to stimulate your own research.
- Journal club: links to related research articles on topics covered in the book are paired with engaging questions to improve your interpretation of the primary literature.
- Rotating figures allow you to visualize complex structures.

For registered adopters

- Figures from the book available to download.

Arthur M. Lesk is Professor of Biochemistry and Molecular Biology at The Pennsylvania State University, USA.

OXFORD
UNIVERSITY PRESS

www.oup.com

ISBN 978-0-19-875483-1

9 780198 754831

LESK INTRODUCTION TO GENOMICS

third edition

OXFORD

INTRODUCTION TO GENOMICS

THIRD EDITION

ARTHUR M. LESK
The Pennsylvania State University

Immured the whole of Life
Within a magic Prison
– Emily Dickinson

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

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

First edition 2007
Second edition 2012
Impression: 1

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

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2016963303

ISBN 978–0–19–875483–1

Printed in Great Britain by
Bell & Bain Ltd., Glasgow

INTRODUCTION TO GENOMICS

For Victor and Valerie

PREFACE

Of all the claims on our curiosity, we want most to understand ourselves. What are we? How have we come to be what we are? What lies in our future? Many features of our lives depend on accidents of history. The time and place of our birth largely determine what language we first learn to speak, whether we are likely to be well fed and well educated, and receive adequate medical care. Many aspects of our future depend on events outside ourselves and beyond our control.

Yet, within us, there are constraints on our lives that brook relatively little argument. In some respects, we are at the mercy of our genomes. Under normal circumstances, all of our basic anatomy and physiology, and eye colour, height, intelligence, and basic personality traits, are ingrained in our DNA sequences. This is not to say that our genomes dictate our lives. Some constraints are tight—for instance, eye colour—but our genetic endowment also confers on us a remarkable robustness.

This robustness also is a product of evolution. Within the last century, lifestyles have changed with a rapidity hitherto unknown (except for the instants of asteroid impacts). We can meet and survive brutal stresses. Our talents have many opportunities to nurture themselves and to develop in novel ways. These are gifts of our genomic endowment: What genes control is the response of an organism to its environment.

The human genome is only one of the many complete genome sequences known. Taken together, genome sequences from organisms distributed widely among the branches of the tree of life give us a sense, only hinted at before, of the very great unity in detail of all life on Earth. This recognition has changed our perceptions, much as the first pictures of the Earth from space engendered a unified view of our planet.

Superimposed on this underlying unity is great variety. We ask: What is special about us? What do we share with our parents and siblings and how do we differ from them? What do we share with all other human beings and what makes us different from the other members of our species? What are the sources of our differences from our closest extant non-human relatives, the chimpanzees? What do we have in common and how do we diverge from other species of primates? Of mammals? Of vertebrates? Of eukaryotes? Of all other living things?

The complete sequences of human and other genomes reveal the underlying text of this story. We are beginning to understand how our lives shape themselves under the influence of our genes, epigenetic modifications, and our surroundings and life histories.

We are also beginning to be able to intervene. Genetic engineering of microorganisms is an established technique. Genetically-modified plants and animals exist, and are the subjects of lively debate. To override the genes for hair colour is trivial. Changes in lifestyle or behaviour can—to some extent—avoid or postpone development of diseases to which we are genetically at risk. Gene therapy offers the promise of rectifying some inborn defects. Novel gene editing techniques based on the CRISPR/Cas system offer revolutionary power in reshaping individuals, whole species, and ecosystems.

Fast, inexpensive sequencing has transformed genomics. The landmark goal, the $US1000 human genome, has been achieved. It is very likely that within the lifetime of many readers, human genome sequencing will be nearly universal. Already, hundreds of thousands of human individuals have had their full genomes sequenced, and many more are on the way. Many people have had sequences determined for individual genes. For example, mutations in *BRCA1*, *BRCA2*, and *PALB2* suggest an increased likelihood of developing breast or ovarian cancer. Many people have had these regions sequenced to provide information about their personal risk. This can be more than informative: some individuals—most famously the actress Angelina Jolie—have opted for prophylactic surgery.

The spectacular progress in high-throughput sequencing, and the resulting explosive growth of the data produced—in quality, quantity, and type—have altered the entire landscape of genomics itself, and its influence has invaded surrounding fields. No area of biology has been left unscathed.

Study of human disease through sequencing continues to be a major effort. The possibility of applications to improve human health were the major motivation for support of the Human Genome Project, and continue to be the major elicitor of largesse for pursuing its consequences. Identification of genes responsible for particular diseases permits testing, genetic counselling, and risk assessment and avoidance. Cancer genomics—the comparison of sequences from normal and tumour cells from single patients—has become a heavy industry.

Understanding the relationships between genes and disease will allow more precise diagnosis and warnings of increased risk of disease in patients and their offspring. For many important diseases, a patient can expect more precise diagnosis and prognosis, and more precise recommendations for treatment. Design of treatment based on a patient's genome, called pharmacogenomics, is already part of clinical practice. Genomes of other organisms also have implications for human health, especially those of pathogenic organisms that have developed, or are threatening to develop, antibiotic resistance. We study the biology of viruses and bacteria to take advantage of their vulnerabilities, and the biology of humans to ward off the consequences of our own.

This century will see a revolution in healthcare development and delivery. Walls between 'blue sky' research and clinical practice are tumbling down. It is possible that a reader of this book will discover a cure for a disease that would otherwise kill him or her. It is extremely likely that Szent-Györgyi's quip, 'Cancer supports more people than it kills', will come true. One hopes that this will happen because the research establishment succeeds in developing therapeutic or preventative measures against tumours, rather than by merely imitating their uncontrolled growth.

Other applications of genomics include improvement of crops and domesticated animals, enhancing food production, and support of conservation efforts dedicated to preserving endangered species. Development of alternative energy sources is a challenge to both physics and biology.

Beyond these applications, genomics offers us a profound understanding of fundamental principles of biology. The history of life is a pageant for which we are beginning to delineate the choreography. On the personal level, within this history, genome exegesis will fundamentally alter our perception of ourselves: what it means to be human.

In this book, I have tried to present a balanced view of the background of the subject, the technical developments that have so greatly increased the data flow, the current state of our knowledge and understanding of the data, and applications to medicine and other fields. With power derived from knowledge goes commitment to act wisely. We have responsibilities, to ourselves, to other people, to other species, and to ecosystems ranging up to the entire biosphere.

Ethical, legal, and social issues have been a prominent component of the human genome project. Most technical questions in genomics, as in other scientific subjects, have objectively correct answers. We do not know all the answers, but they are out there for us to discover. Ethical, legal, and social issues are different. Many choices are possible. Their selection is not the privilege of scientists in individual laboratories, but of society as a whole. Scientists do have a responsibility to contribute to the informed public discussion that is essential for wise decisions.

One aspect of the first edition that I liked was its concision. Unfortunately, this has had to be sacrificed to the stampeding progress of the field. A major problem encountered in writing about genomics is the need to pick and choose from the many riches of the subject. The list of subjects that cannot be left out is too long and threatens to reduce the treatment of each to superficiality. There is also a serious organizational challenge: many phenomena must be approached from several different points of view. A reader may be relieved to conclude that a topic has been beaten thoroughly into submission in one chapter, only to encounter it again, alive and kicking, in a later section, in a different context.

The speed at which the field is moving causes other problems. An author is often pleased with a draft of a section only to find the carefully described conclusions overturned in next week's journals. Yet, there is a great pleasure in seeing nature's secrets emerging before one's eyes.

Another casualty of rapid progress is the frequent dismissal of attention to history and biography. We are fantastically interested in the development of the sea urchin and the fruit fly, but not in the development of molecular genomics. This is too bad: those who do not learn from the successes of history will find

it harder to emulate them. Intellectual struggles that occupied entire careers leave behind only the terse conclusions, often without any appreciation of the experiments that established the facts, much less of the alternative hypotheses tested and rejected. We remember the breakthroughs, but not the frustrations that their achievement required. The force of the scientists' personalities, and their foibles, are forgotten. Making use of other people's results isn't the same thing as creating new ones: 'To imitate the *Iliad* is not to imitate Homer'.

Genomics is an interdisciplinary subject. The phenomena we want to explain are biological. But many fields contribute to the methods and the intellectual approaches that we bring to bear on the data. Physicists, mathematicians, computer scientists, engineers, chemists, clinical practitioners and researchers have all joined in the enterprise. This book will appeal, at least in part, to all of them. However, the central point of view remains focused on the biology.

More specifically, the focus is on human biology. In fact, on the biology of humans who are curious about other species, albeit primarily for what the other species tell us about ourselves. This choice naturally reflects the potential readership of this book. (If bacteria or fruit flies could read, genomics textbooks would look very different.)

In the new edition, extended coverage is given both to clinical applications to humans and to the applications of genome sequences to working out of evolutionary relationships in microorganisms, plants, and animals. Non-clinical applications to human biology and history are sufficient to justify a chapter of their own. The genomics of plant and animal domestications not only shed light on human—as well as plant and animal—history, but emphasize the reciprocal interactions between humans and the rest of the biosphere.

This book assumes that the reader already has some acquaintance with modern molecular biology, and builds on and develops this background, as a self-contained presentation. It is suitable as a textbook for undergraduates or starting postgraduate students.

Progress is happening so fast as to make unavoidable a feeling of frustration in aiming at a moving target. The hope is that the third edition has erected for the reader a sound framework, both intellectual and factual, that will make it possible, when encountering subsequent developments, to see where and how they fit in.

Exercises and problems at the end of each chapter, and 'weblems' on the Online Resource Centre, test and consolidate understanding, and provide opportunities to practise skills and explore additional subjects. Exercises are short and straightforward applications of material in the text. Answers to exercises appear on the website associated with the book. Problems, also, make use of no information not contained in the text, but require lengthier answers or, in some cases, calculations. The third category, 'weblems', requires access to the Internet. Weblems are designed to give readers practice with the tools required for further study and research in the field. Some of these are suitable for use as practical or laboratory assignments, or even as class projects.

Key terms are highlighted in the text, and are defined in the Glossary at the end of the book.

Chapter 1, *Introduction and Background*, sets the scene, and introduces all of the major players: DNA and protein sequences and structures, genomes and proteomes, databases and information retrieval, and bioinformatics and the Internet. Subsequent chapters develop these topics in detail. Chapter 1 sketches the framework in which the pieces fit together and sets genomics in its context among the biomedical, physical, and computational sciences.

The message is clear: genomics is the hub of biology. Whereas genome sequences are determined from individuals, to appreciate life as a whole requires extending our point of view spatially, to populations and interacting populations; and temporally, to consider life as a phenomenon with a history. We can study the characteristics of life in the present, we can determine what came before, and we can—at least to some extent—extrapolate to the future. The 'central dogma' and the genetic code underlie the implementation of the genome, in terms of the synthesis of RNAs and proteins. Absent from Crick's original statement of the central dogma is the crucial role of regulation in making cells stable and robust, two characteristics essential for survival.

Chapter 2, *The Human Genome Project: Achievements and Applications*, focuses on the human genome and the applications of human genome sequences. It reports the current state of the data, although that gives an inadequate sense of their explosive growth. Clinical applications are maturing from hype to serious promise to actual clinical practice.

Putting our species in context involves, most narrowly, comparing our genomes with those of our closest relatives, including Neanderthal man and the chimpanzee, our nearest extant relative. More general applications to anthropology appear in Chapter 9.

Applications of genome sequences in personal identification are well known. These include determinations of paternity, and, less frequently, maternity. Crime-scene investigation has proved the guilt or innocence of many suspects. It is the stuff of popular entertainment.

Although much genome sequence investigation is carried out under clinical or forensic organization, sequencing has gone public with 'pop' applications provided by personal genealogy companies. There are even 'dating sites' that use the correlation between major histocompatibility complex haplotype and mate selection to offer to identify mutually attractive individuals.

Many of these applications involve ethical, legal, and social issues. Different jurisdictions have established different guidelines or regulations.

Chapter 3, *Mapping, Sequencing, Annotation, and Databases*, describes how genomics has emerged from classical genetics and molecular biology. The first nucleic acid sequencing, by groups led by F. Sanger and W. Gilbert, in the 1970s, was a breakthrough comparable to the discovery of the double helix of DNA. The challenges of sequencing stimulated spectacular improvements in technology. First came automation of the Sanger method. The original sequences of the human genome were accomplished by batteries of automated Sanger sequencers. Subsequently, a series of 'new generations' of novel approaches have achieved the landmark goal, the $US1000 human genome. Sequencing power is very widely distributed. There are major specialized institutions, such as the Beijing Genomics Institute (now in Shenzhen). BGI sequencers generate 10 terabytes of raw data per day. (You do the math: that's over 2 human genomes per minute!) Smaller installations are common in universities, hospitals, and companies.

Where do all the publicly available data go? Chapter 3 also introduces the databanks that archive, curate, and distribute the data, and some of the information-retrieval tools that make them accessible to scientific enquiry.

Chapter 4, *Evolution and Genomic Change*, treats relationships. Life has been shaped by evolution, primarily acting through natural selection. T. Dobzhansky famously said, 'Nothing in biology makes sense except in the light of evolution'.[1]

Genomics allows us to trace many aspects of this process, if not always to make sense of them.

Of course, most of evolution took place in the past. We cannot observe it directly. However, we can observe and draw inferences from its products. These

[1] I would add thermodynamics to the list of things except in the light of which nothing in biology makes sense.

products are, for the most part, the genomes—and phenotypes—of extant organisms, plus sporadic data from recently extinct species. In addition, many evolutionary events have left their traces in contemporary genomes.

Chapter 4 explores some of the tools that scientists have developed to analyse sequence data for what it can reveal about evolution. Prominent among these are methods for sequence alignment and calibration of the results, and the computation of phylogenetic trees.

We are entering an era when evolution, even in natural populations, may be under the direct control of molecular biologists wielding tools based on CRISPR/Cas and gene drive. Because most of the material in Chapter 4 is decades or even centuries old, one might think that these sections might be the least-likely parts of this book to need significant revision during the coming decade or so. This may be a dangerous assumption. We shall see.

Chapter 5, *Genomes of Prokaryotes and Viruses*, surveys the genomes of viruses, bacteria, and archaea in more detail. Taxonomy and phylogeny present problems because of extensive horizontal gene transfer. Indeed, horizontal gene transfer challenges the whole idea of a hierarchy of biological classification. In the past, many bacteria have been cloned and studied in isolation, especially those responsible for disease. However, a new field, metagenomics, deals with the entire complement of living things in an environmental sample, allowing us to address questions about interspecies interaction in the 'real world'. Sources include ocean water, soil, and the human body.

Chapter 6 surveys *Genomes of Eukaryotes*. It starts with yeast, which is about as simple as a eukaryote can get. Selected plant, invertebrate, and chordate genomes illuminate the many profound common features of eukaryotic genomes; and the very great variety of structures, biochemistry, and lifestyles that are compatible with the underlying similarities. There are examples of recovery and sequencing of DNA from extinct organisms.

The goal of Chapter 7, *Comparative Genomics*, is to harvest some conclusions from the surveys of viral, prokaryotic, and eukaryotic genomes presented in the preceding chapters. We begin by comparing the different modes of genome organization with which living things have experimented. If the great variety of living things has arisen through evolution by natural selection—the change in allele frequency in a population through differential reproductive success among different variants—how does the variation arise? We must consider the nature and extent of the variability in the genomes within individual populations, and the mechanisms that generate this variability. It will emerge that, especially in prokaryotes, horizontal gene transfer is a very important component of the mechanism.

Chapter 8, *The Impact of Genome Sequences on Health and Disease*, describes clinical applications of genome sequencing. Sometimes, when particular genes are known to correlate with disease or risk of disease, specific regions in a patient's genome are sequenced. More and more, we will see complete genome sequencing in clinical contexts.

Applications include improved diagnosis and prognosis of the causes of presenting syndromes, genetic counselling of parents with family histories of a dangerous genetic condition, and 'pharmacogenomics': the tailoring of treatment to the individual patient, based on DNA sequence information. 'Gene therapy', the replacement of defective genes with correct ones, has already had some successes and, with the development of CRISPR/Cas technology of genome editing, will undoubtedly grow in applicability. (Guidelines emerging from a 2015 Washington conference urged a moratorium on application of CRISPR to humans. In 2017, the US National Academies of Sciences and Medicine recommended allowing human germline editing under certain stringent conditions. But the pressure is too great for deterrence to survive, even in the US.)

Chapter 9, *Genomics and Anthropology*, develops additional applications to the study of our own species. Although clinical applications are undoubtedly the most important, genomics has important contributions to make to human palaeontology and anthropology. The ability to extract DNA from extinct species, including Neanderthals, sheds light on our early evolution. Events in our history, including migrations and domestication of crops and animals, have left their traces in DNA sequences.

With Chapter 10, *Transcriptomics*, we move on from the (relatively) static genome to the selectivity and dynamics of expression patterns. Following the

central dogma, there are at least two stages: transcription, in which DNA makes RNA (Chapter 10); and translation, in which RNA makes protein (Chapter 11).

Measurements of cellular RNAs describe the transcription patterns of regions of the genome. These patterns vary in response to changes in the environment: the *lac* operon of *Escherichia coli* is a classic example. Expression patterns vary among different tissues, different physiological states, and different developmental stages. Thus, the human embryo and neonate synthesizes embryonic and then foetal haemoglobin, switching to expression of adult haemoglobin at about 6 months post-birth.

The inventory of mRNAs in the cell—a component of the transcriptome—is naturally of interest for what it can tell us about the distribution of cellular proteins. Certainly protein synthesis does involve many RNAs, including messenger RNAs, the RNA of the ribosome, transfer RNAs, and the RNAs of the spliceosome, that removes introns from pre-messenger RNAs. However, other RNAs have a variety of roles, including catalytic activity. Catalytically active RNAs are called *ribozymes*. (In both the ribosome and the spliceosome, the catalytic activity is in the RNA, not the protein component.) Other RNAs, such as microRNAs, short interfering RNAs, and silencing RNAs, control gene expression by interacting with messenger RNAs. The RNA transcripts of CRISPR sequences, bound to CRISPR-associated nucleases, identify viral invaders. There is good reason to believe that many other RNA functions remain to be discovered.

Chapter 11, *Proteomics*, describes briefly the principles of protein structure and the high-throughput data streams that provide information about sets of proteins in cells, including methods for predicting protein structure from amino-acid sequence.[2] Proteomics is an essential complement to genomics. (A colleague once entitled a keynote lecture: 'Genes are from Venus, proteins are from Mars.') In fact, one of the major messages of Chapters 10 and 11 is how much the genome *doesn't* tell us about cellular proteins. Nevertheless, the interactions and relationships between the genome and proteome are intimate, both during cellular activity and in the longer term in evolution.

Chapters 12 and 13 introduce systems biology, a relatively new field that presents a description of biological organization based on networks. Systems biology deals with the connectivity of biological networks, and analyses and models the mechanisms that control traffic patterns through them.

A reasonable definition of life would include the criterion that a living thing executes controlled manipulations of matter, energy, and information. Chapters 12 and 13 treat the cellular networks that carry out and regulate these tasks.

Chapter 12, *Metabolomics*, deals with the metabolic networks that manipulate matter and energy. Cells require a source of energy, from nutrients or from light, and use it to drive metabolites through series of metabolic pathways. The flows through the paths in this network must be kept under control. To achieve this, another network, a logical one, keeps cellular activities organized. Cells contain parallel sets of networks based on *physical* and *logical* interactions among molecules. Each network also has *static* and *dynamic* aspects.

Components present in a cell at any instant are subject to direct controls, for instance feedback inhibition by an end product to shut down a metabolic pathway. But transcription is subject to very extensive oversight: turning genes on and off. Chapter 13, *Systems Biology*, treats the logical component of cellular networks that controls transcription and metabolism.

The genius of classical biochemistry was to take cells apart and show that the components could function in isolation. Now our job now is to put things back together.

[2] For a more thorough treatment of proteins see Liljas, A., Liljas, L., Piskur, J., Lindblom, G., Nissen, P., & Kjeldgaard, M. (2009). *Textbook of Structural Biology*. World Scientific, Singapore; or Lesk, A. (2016). *Introduction to Protein Architecture: Structure, Function, and Genomics*, 3rd ed. Oxford University Press, Oxford.

RECOMMENDED READING

Where else might the interested reader turn? This book is designed as a companion volume to three others: *Introduction to Protein Architecture: The Structural Biology of Proteins*; *Introduction to Protein Science: Architecture, Function, and Genomics*; and *Introduction to Bioinformatics* (all published by Oxford University Press). Of course, there are many fine books and articles by many authors, some of which are listed as recommended reading at the ends of the chapters. The goal is that each reader will come to recognize his or her own interests, and be equipped to follow them up.

For the recommendations for additional reading at the ends of chapters, I have limited myself to books, and to articles in the scientific literature. However, if one wants an introduction to a topic, there are many quite good lectures on the Internet, which might well be useful. Or, if one wants up-to-date details about a topic, there are blogs. Indeed, blog entries are often more useful than a full scientific paper. With a full paper there is the problem of extracting a take-home message from a mass of detail, whereas the blog entries often contain one or two concise and 'to-the-point' paragraphs.

Nevertheless, I am reluctant to include these in the recommendations. This is partly out of a commitment to the refereed scientific literature. But also: (1) there is no control over whether a recommended website will remain available. The scientific literature at least has a permanent presence. (2) Blogs can contain a mixture of some contributions that are very useful and others that are not; and it may be difficult to distinguish. But it is undeniable that the Internet contains very many useful sources of information outside of the printed scientific literature.

Many applications of genomics to healthcare are discussed in the book. However, nothing here should be taken as offering medical advice to anyone about any condition.

INTRODUCTION TO GENOMICS ON THE WEB

Results and research in genomics make use of the Internet, both for storage and distribution of data, and for methods of analysis. Readers will need to become familiar with websites in genomics, and to develop skills in using them. Many useful sites are mentioned in the book. The author's *Introduction to Bioinformatics* offers a pedagogical approach to computational aspects of genomics. However, clearly, the place to learn about the Internet is on the Internet itself.

To this end, an Online Resource Centre at www.oxfordtextbooks.co.uk/orc/leskgenomics3e/ accompanies this book. This contains web-based problems ('weblems') and material from the book—figures and 'movies' of the pictures of structures, answers to exercises, hints for solving problems, and guides to useful websites.

ACKNOWLEDGEMENTS

I thank S. Ades, G.F. Anderson, E. Axelsson, M.M. Babu, S.L. Baldauf, P. Berman, B. de Bono, D.A. Bryant, C. Cirelli, A. Cornish-Bowden, R. Diskin, R.B. Eckhardt, N.V. Fedoroff, J.G. Ferry, R. Flegg, J.R. Fresco, S. Girirajan, N. Goldman, J. Gough, D. Grove, D.J. Halitsky, R. Hardison, E. Holmes, H. Klein, E. Koc, A.S. Konagurthu, T. Kouzarides, H.A. Lawson, E.L. Lesk, M.E. Lesk, V.E. Lesk, V.I. Lesk, D.A. Lomas, B. Luisi, J.A. Lumadue, P. Maas, K. Makova, W.B. Miller, C. Mitchell, J. Moult, E. Nacheva, A. Nekrutenko, G. Otto, A. Pastore, D. Perry, C. Praul, K. Reed, G.D. Rose, J. Rossjohn, S. Schuster, R.L. Serrano, B. Shapiro, J. Shi, J. Tamames, A. Tramontano, A.A. Travers, A. Valencia, G. Vriend, L. Waits, R. Wayne, J.C. Whisstock, A.S. Wilkins and V.E. Womble, and E.B. Ziff for helpful advice.

I thank the staff of Oxford University Press for their essential and superb contributions to this book.

The author and publisher would also like to thank those who gave their time and expertise to review the book throughout the writing process. Your help was invaluable:

- Richard Bingham, Department of Biological Sciences, University of Huddersfield, UK
- Joanna Kelley, School of Biological Sciences, Washington State University, USA
- Hans Michael Kohn, Institute of Biosciences and Bioengineering, Rice University, USA
- Emma Laing, Department of Microbial and Cellular Sciences, University of Surrey, UK
- Frances Pearl, School of Life Sciences, University of Sussex, UK
- Michael Shiaris, College of Science and Mathematics, University of Massachusetts Boston, USA
- Adrian Slater, Faculty of Health and Life Sciences, De Montfort University, UK
- David Studholme, College of Life and Environmental Sciences, University of Exeter, UK