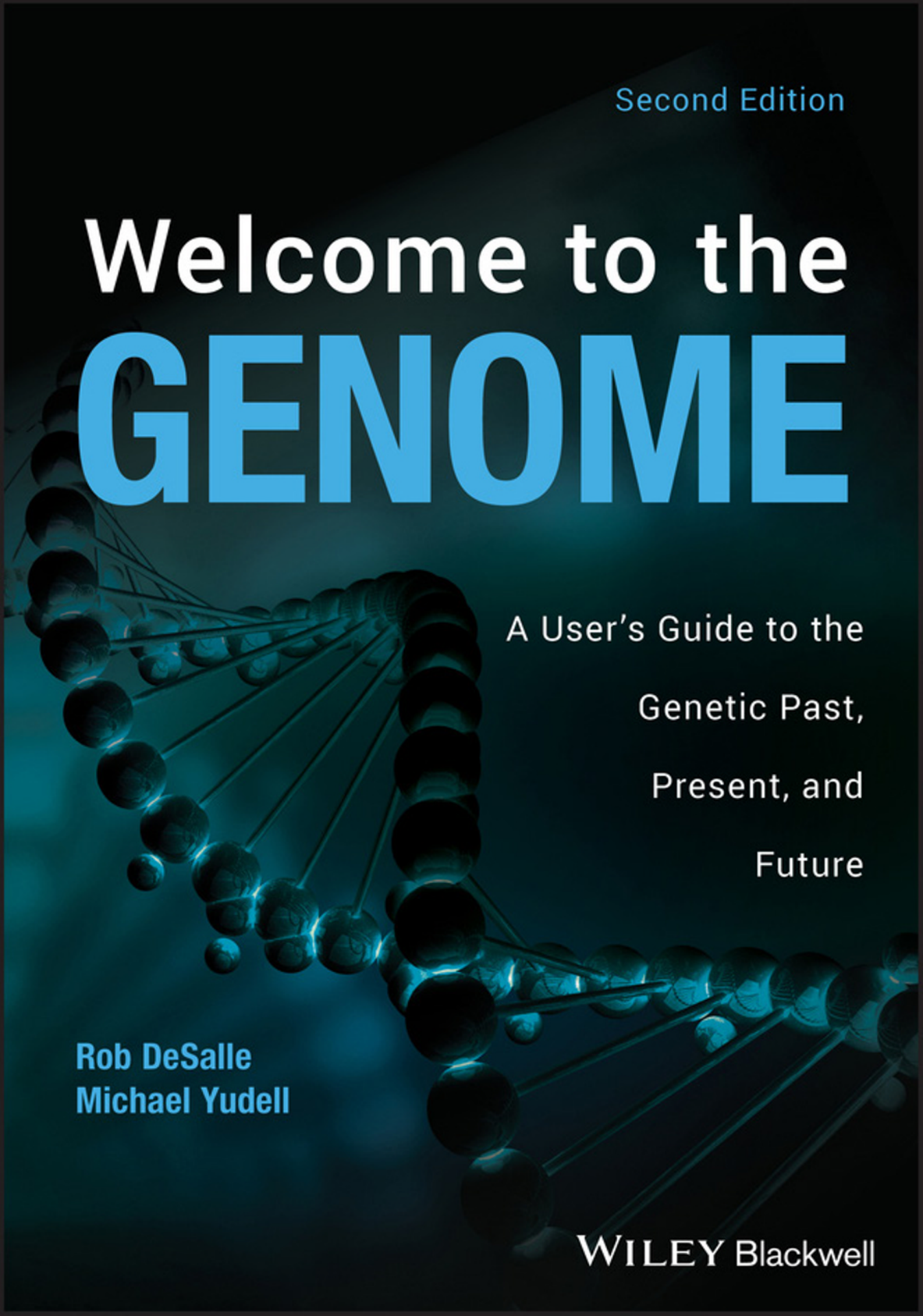
Second Edition

Welcome to the GENOME

A User's Guide to the Genetic Past, Present, and Future



Rob DeSalle
Michael Yudell

WILEY Blackwell

WELCOME TO THE GENOME

WELCOME TO THE GENOME

A User's Guide to the Genetic Past, Present, and Future

SECOND EDITION

Rob DeSalle
Institute for Comparative Genomics
American Museum of Natural History
New York

Michael Yudell
Department of Community Health and Prevention
Dornsife School of Public Health
Drexel University
Philadelphia

WILEY Blackwell

This second edition first published 2020

Edition History
John Wiley & Sons 1e, 2005

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
Boschstr. 12, 69469 Weinheim, Germany

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: DeSalle, Rob, author. | Yudell, Michael, author.
Title: Welcome to the genome : a user's guide to the genetic past, present, and future / by Rob DeSalle, Michael Yudell.
Description: Second edition. | Hoboken, NJ : Wiley-Blackwell, 2020. | Includes bibliographical references and index.
Identifiers: LCCN 2019053496 (print) | LCCN 2019053497 (ebook) | ISBN 9781118107652 (paperback) | ISBN 9781118750315 (adobe pdf) | ISBN 9781118750308 (epub)
Subjects: MESH: Genome, Human | Genomics | Popular Work
Classification: LCC QH447 (print) | LCC QH447 (ebook) | NLM QU 460 | DDC 611/.01816–dc23
LC record available at https://lccn.loc.gov/2019053496
LC ebook record available at https://lccn.loc.gov/2019053497

Cover Design: Wiley
Cover Image: © Mina De La O/Getty Images

Set in 10/13pt Palatino by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

We would like to dedicate this book to the Cullman and Korein families,
who over the past two decades have generously supported and continue to support
the expansion of modern genomics at the American Museum of Natural History.

CONTENTS

FOREWARD: STILL, THE GENOMIC REVOLUTION

In 2001, the American Museum of Natural History (AMNH) produced a daring and innovative exhibition called the Genomic Revolution. It transcended the usual subject matter and issues that the AMNH addresses with its exhibitions. The show used cutting-edge interactives and design and covered many up-to-date subjects in genomics. Development of the show was as dynamic as the field itself, with revisions to data on genome size and other scientific content required within days before the show opened. We were fortunate that the show also anticipated the announcement of the first draft of the human genome that same year. The production of the show was part of the AMNH's desire to enter the twenty-first century with a genomics bang as we also established an institute for comparative genomics and the Ambrose Monell frozen tissue facility at the museum – two very rare and still successful enterprises for a natural history museum.

The show was very popular despite the occurrence of the 9/11 attack and its aftermath during its run, drawing over half a million visitors. Since its run at the AMNH it was installed in natural history or science museums in Toronto, Canada, Sao Paulo, Brazil, Mexico City, Mexico, and several American cities. As a companion to the show, my colleagues Rob DeSalle and Michael Yudell produced a very popular volume entitled *Welcome to the Genome* that tells the story of genetics, genomics, and the sequencing of the human genome. These two authors were involved in the development of the exhibition, one as its curator and the other as a content advisor. Their treatment of the topic spanned discussion of Mendel, Darwin, development of molecular biology, and the DNA sequencing technology of the time, as well as delving into the ethical conundrums caused by the capacity to decipher the so-called stuff of life – DNA. Rob, as a curator at the AMNH, has led the museum's effort to broaden its scope of research and education and demonstrate our capacity to connect with dynamic fields like genomics and with applications like human health.

Now, nearly 20 years later, Rob and Michael have produced a second edition, demonstrating that many things have changed since their first version, and at the same time some remain the same. The sequencing technology has developed to the point where the $1000 human genome is possible and where

we are now getting sequences of the microbes that live in and on the body called the microbiome. The tree of life, a major interest here at the AMNH that Rob and Mike delve into in great detail in this book, has also benefited from the development of genomic technology, which has improved our understanding of human evolution, both deep in the past and more recent. It is now possible to sequence the genomes of Neanderthals and long dead *Homo sapiens* specimens, giving science an unprecedented view of our past history. Ethical issues have also emerged as the technology gets faster and cheaper. Development of new gene editing technologies opens the way for germline editing of humans and raises many questions about the application of this technology. This new edition by Rob and Michael will provide the kind of modern, current, and thought-provoking information that is in keeping with one of the most important scientific revolutions of our time.

As a natural history museum, the AMNH strives to educate the public about science in the natural world and to teach the public about how science is accomplished, who scientists are, and what the science might mean to their future. This edition of *Welcome to the Genome* is a wonderful extension of that initial effort in education that the AMNH launched nearly two decades ago.

Michael J. Novacek
Provost of Science
American Museum of Natural History

ACKNOWLEDGMENTS

Second editions test the foundations of longstanding collaboration, especially when a book's research and writing goes on far longer that either collaborator had predicted (or, for that matter, hoped). But here we are. *Welcome to the Genome* (the second edition) is a book way too many years in the making. The slow pace at which we wrote, however, actually turned out to be a good thing. The science evolved. As did our own perspectives on genomics. And if you look at the tone of the two editions of the book, our evolution on the genomic revolution is clear. We went from "Hey, this is awesome and it will change the world" (first edition) to "The science and technology of genomics are amazing, especially as they impact basic science. But, as far as the genome's impact on our collective health and wellbeing, well, way too much hype" (second edition).

This second edition, built on the foundation of the first, owes so much to the many librarians and archivists who guided us through both editions. We also remain thankful to the many members of our team who made this book possible. At the American Museum of Natural History, Maron Waxman, formerly Special Publications Director, who is now enjoying retirement, was a dedicated editor, staunch advocate, and good friend, and remains the reason why we wrote this book in the first place.

The staff of the American Museum of Natural History's photo studio, including Denis Finnin and Craig Chesek, helped us put together the images seen in both the first and second editions. Some of these images began as components of the Museum's exhibition "The Genomic Revolution," an exhibit that ran at the Museum and then nationally from 2002 to 2010. We are grateful to the Museum's exhibition team, including Vice President for Exhibitions, David Harvey, and exhibition designer, Tim Nessen.

We remain grateful for the inspiration and contributions from the writers and researchers in the Museum's Exhibition Department, who played an important role shaping this book, especially the first edition. Lauri Halderman, Karen de Seve, and Martin Schwabacher were the writers for "The Genomic Revolution." Some of the passages and case studies in both editions were originally researched and written by the exhibition's writing team. We especially relied on their approach and writing in Chapter 10, which looks at genetically modified organisms. And we must also thank and recognize Dr. Yael Wyner for

her efforts in guiding the exhibition's content and thus for her intellectual contributions to this book.

The entire staff of the Museum's library and archives deserves our special thanks for helping us with this project. Finally, we want to thank Museum President, Ellen Futter, and Museum Provost, Michael Novacek, whose continued support and commitment to public education about the genome project helped get both volumes of this book off the ground.

Two colleagues read the manuscript in its entirety. Dr. Holly Tabor, Associate Professor of Medicine at Stanford University, and David Randle from the American Museum of Natural History offered detailed comments on the book. Suzanne Grossman, my former research assistant here at Drexel, spent almost a year working with Rob and I completing the book, editing the text, gaining photo and image permissions, and making sure we got the damn thing done. We are so grateful to Suzanne! Thank you! We are grateful for the additional research assistance provided along the way by Katia Duey and Jaime Earnest at Drexel.

Finally, we are very grateful for the patience and ongoing support of Antony Sami, Priya Subbrayal, and the entire team at John Wiley & Sons, who helped shepherd this second edition to completion.

INTRODUCTION: WELCOME BACK TO THE GENOME

Every one of the trillions of cells in your body contains DNA—from the blood cells that course through your veins to the nerve cells in your brain to the hair follicle cells that line your scalp. The tightly coiled DNA in a single cell, 6 feet long and just one molecule wide when unraveled, packs more than 3 billion bits of information. This complete set of information is your genome. The approximately 20,000 genes in your genome (a figure that has been revised down significantly since we wrote the first edition of this book more than a decade ago), interacting with each other and with your environment, help shape the development of a new human being and are constantly at work instructing our bodies to create new cells, digest food, fend off disease, and store thoughts. Genes and DNA capture our imagination because of their impact on why we are the way we are. But how much control do genes and DNA really have over our bodies and our behavior? And to what extent will our changing understanding of the human genome change who we are and how we see the world? Are our genes our destiny? Are our genomes our fate?

Such questions captured our imagination in the midst of the genomic revolution—the international multi-billion-dollar effort to sequence, interpret, and exploit the human genetic code. It was believed that a map of our genome would offer boundless potential to sequence, interpret, and then exploit the information contained in the genetic code. The excitement over the potential to improve our health—to stave off disease, to apply genomic tools to feed the world's growing population, to save species on the brink of extinction—captured the imaginations of scientists around the globe in the opening decade of the genomic revolution.

In June 2000, scientists triumphantly announced they had sequenced the human genome. (1) By sequencing those 3.2 billion units of our DNA, researchers sparked a firestorm of discovery and ushered in a new age. At a White House ceremony to announce the completion of a draft sequence of the human genome, President Bill Clinton called the genome God's handiwork. "Today," Clinton stated, "we are learning the language in which God created life." (2) Clinton's vision of the genome was one that mixed a metaphor of scientific

advancement with a divine spirit. This image of the human genetic code is a fairly common one. The genome has also been called the book of life, biology's Rosetta Stone, humanity's instruction manual, and biology's Holy Grail. Each of these metaphors conveys a slightly different meaning, and each suggests a subtly different aspect of the genome. Not so hidden in these metaphors is the hope that biology will provide clear-cut answers to long-asked questions regarding the nature of the human soul, the power of science to heal and rebuild the human body, and the role of nature in human social behavior. The genome will indeed provide some answers to these questions, but not the simple answers that many of these metaphors suggest.

The media both anticipated and echoed the hyperbole of that White House ceremony. Headlines like "Long Held Beliefs Are Challenged by New Human Genome Analysis" (*New York Times*, 2001), "Reading the Book of Life: Genome Shows Evolution Has an Eye for Hyperbole" (*New York Times*, 2001), and "Double Helix Is Starting To Make Its Mark In Medicine" (*Financial Times*, 2003) underscored the genome's complexity, allure, and promise. (3) But did we oversell the genome? Has the early allure and promise translated into meaningful scientific results?

Looking back at the sequencing stage of the genomic revolution (1988–2001), it is important to consider whether we were even asking the right questions in the first place about what the genome could tell us about the relationship between our genes and our health. It isn't simply that we spoke in hyperbolic terms about the secrets the human genome would reveal (which we ourselves were somewhat guilty of in the first edition of this book). It was that the language used and the metaphors employed to describe genomics limited our ability to capitalize on the work being done in laboratories around the globe. In other words, the way we described genomics circumscribed how we carried out genomics' research. Think of some of the most popular metaphors used—the book of life, a genetic code, life's blueprint. They suggested, in their simplest terms, that our genomes contained information to read, or as some have suggested, the blueprint from which humanity and other species are built. Some observers have argued that this blueprint approach both reflected and reinforced the type of reductionist thinking that was commonplace in the early years of genomic sequencing (4) and that has its roots in the founding of the field of genetics a century ago.

Today, the language of the genome is changing, and so too is our scientific understanding of the information it contains. A book of life that can reveal the essence of what it means to be human (or any other species, for that matter) has given way to new metaphors that reflect (and perhaps limit) the current science, which seems to value complexity over simplicity.

Let's start with the term post-genomics—a widely used term meant to signify the post-sequencing era we currently inhabit during which science is working to make sense of billions of bits of sequenced genomic information. By calling it the *post*-genome era, we are implying a break from the discovery phase of the genomic revolution to an era in which gathered information is analyzed.

Some have speculated that the post-genome genome is less a linear string of genes that produce traits (alone and in concert with one another) than an organic and dynamic mechanism that responds to both biological and environmental stimuli to produce the proteins that regulate the life of an organism. (5) It is in the complexity of the post-genomic genome where natural and social scientists will untangle the complicated relationship between organism, genes, and environments that the challenges and surprises of life await discovery.

One thing that the old reductionist model has over the new models of genomic complexity is clarity. It would have been a difficult task to sell the genome—at the height of its popularity in the 1990s—as a complex mechanism that regulates life. Indeed, the reductionist model has had its utility in discovering simple, mostly Mendelian, genetic traits. But as we have come to understand genomes as biological systems rather than blueprints or Rosetta Stones, the genomic sciences have come to rely more and more on fields like computer science and bioengineering to make sense of the post-genome.

Genomics is a synthesis of many disparate fields, including biology, public health, engineering, computer science, and mathematics. What makes genomics even more distinctive is that the social sciences and humanities are an integral component of the genomic revolution. Philosophers, ethicists, and historians are helping to lay the foundation of the genomic revolution by pushing for and playing a role in the creation of policies and laws that will guide the integration of genomics into scientific practice and health care. Participants in the genomic revolution, as well as the biologists and others who preceded them, will, we believe, be thought of much in the same way that Newton is remembered for his role in the birth of calculus and physics or the way in which Darwin is remembered as the progenitor of modern biology. However, because genomics is an evolving science that encompasses so many different disciplines, it is hard to find one person who embodies the entire field. Indeed, it will be a group of genomic scientists who will be recorded in history books as pioneers.

The arrival of the genomic age was the culmination of efforts of over a century of science. From the work of Gregor Mendel in the mid-nineteenth century (it was Mendel who formalized the rules of heredity and hypothesized that something like genes must underpin heredity), to the announcement of the discovery of the structure of DNA in 1953 by James Watson and Francis Crick, to the genetic sequencing technologies developed by biologists like Frederick Sanger and Leroy Hood in the closing decades of the twentieth century, the path to genomics has been arduous but has yielded the richest source of biological data we have ever known. This age of discovery is where our journey in this book begins—the first four chapters look at the historical moments in biology over the past 100 or so years that made the sequencing of genomes possible. These chapters will be particularly rewarding to readers with an interest in the science behind genomics, but you do not need to comprehend everything in these chapters to appreciate the material in the rest of the book. Don't get hung up on some of the nitty-gritty science. Utilize the figures to help make sense of difficult concepts, and don't be afraid to look up technical sounding words.

The remainder of the book looks at the interplay of how scientists are coming to make sense of genomic information and how they are applying this information to genomic technologies in evolutionary biology, health-related fields, and agriculture. Chapters 5 through 7 look at how the discovery and exploration of the human genome is yielding to the more practical task of sorting through the scientific and social meaning of all of the data being generated by genomics, particularly in the context of ethics and how we understand and define ourselves as humans, especially given the long history of using genetics to divide and harm ourselves. The choices, social proscriptions, and laws that we develop now around genomic technologies will be an essential part of ensuring the success of genomic technologies in the future. Challenges include creating policies that will help integrate genomics technologies into contemporary medicine and public health practice, and defining the roles and responsibilities of scientists, health care professionals, ethicists, clergy, and lawmakers in the development of these policies. Also, how can we best ensure the safety of genomic technologies? The remaining chapters of the book look at how advances in genomic science—from evolutionary thinking to agriculture biology—are altering scientific practice and impacting our lives. For example, new tools such as clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9 (CRISPR/Cas9) technology have been developed that allow for the direct editing of genomes and may usher in a new age of gene therapy (7) with many of the caveats we initially formulated in the first edition of this book. In the first edition of this book, we suggested that "it will still take years, if not decades before genomic medicine will significantly enhance current practice, let alone replace it." But CRISPR/Cas9 gene editing technology has the potential to bring us directly into the realm of directed gene therapy, in both human and non-human species.

We have set out to write a book that readers with little or no prior knowledge of biology can pick up and enjoy, gaining along the way a deeper understanding of the phenomenon that has become known as genomics. Genomics should not be treated lightly, however, and we hope to reward your interest with more than a nominal exploration of this still-burgeoning science. Indeed, one can pick up any number of magazine or newspaper articles for that. This book offers something more—something useful to you, the consumer—by elucidating today's genomic information and tomorrow's genomic medicines and technologies. It is the latter that will, in various ways, greatly affect our lives. Although we may not directly benefit from incredible genetic discoveries, children born today come into the world with the promise that genomics will have a significant impact on their lives, and for their children the effect will be exponentially greater, continuing likewise through the generations.

For us, though, the consequences of genomics will be no less significant. Although we will benefit from early generations of genome-driven therapeutics, we also face the critical task of struggling with the consequences of these potentially disruptive technologies. We are charged with making sense of the genome's social, cultural, and economic implications, and with successfully

implementing genome technologies. Although lives will be improved and even saved by genomic drugs, our generation's legacy will be much more than the scientific and medical discoveries it leaves to the twenty-first century. Our legacy will also be social—meeting the challenge of making genomics technologically feasible—and at the same time humane, just, and ethical. This will be no easy task, particularly from our current vantage point: At present we as a society still remain largely unprepared for the arrival of the genomic revolution. This book was written with these challenges in mind, and with the hope that we can be a part of the continued effort to make the genome truly public.

At the American Museum of Natural History (AMNH), we have worked toward integrating genomics into Museum scientific practice and into our exhibits. Way back in the fall of 2000, as part of its mission to bring cutting-edge science to the public, the Museum held a 2-day conference examining the social and scientific implications of the genome. *Sequencing the Human Genome: New Frontiers in Science and Technology* was the first major public forum to examine the implications of genomics after the release of the draft sequence of the human genome. That is where much of the thinking about this book began. Renowned scientists, including two Nobel Laureates, bioethicists, historians, biotechnology entrepreneurs, and others participated in a variety of lectures and panel discussions. This effort was followed in spring 2001 with the opening of the exhibition "The Genomic Revolution," the largest and most comprehensive popular examination of the genome to date. Efforts continue through the Museum's education programs and by expanding the reach of "The Genomic Revolution," which has traveled to a nearly a dozen sites around the United States in the past decade. In addition, in 2008 the AMNH renovated its Hall of Human Biology (renamed the Spitzer Hall of Human Origins). This renovation project changed the focus of the hall from strictly paleo-anthropological subject matter to include genomics and genetics of primates and humans specifically. Exhibit material on genomes in this permanent hall includes information on how genomes (including the Neanderthal genome) are sequenced, the similarity of primate genomes, how Neanderthal genomes compare with sapiens' genomes, and how genetic information can be interpreted to give us an idea of the movement of humans across the planet.

For well over a century the Museum's halls, replete with fossils, models, and dioramas, have been home to a diversity of exhibitions that, with few exceptions, have centered on objects—exactly the fossils and dioramas that fill the Museum's galleries. These object-driven exhibits utilize the charisma of a specimen to engage the visitor. An ancient *Barosaurus* standing on its hind legs, towering 40 feet in the air does just that in the main rotunda of the Museum every day. Once a visual connection to a specimen is made, the conceptual aspects of an exhibit can be presented. In the case of the *Barosaurus*, the Museum can discuss a wide range of such dinosaur-related topics as predation, evolution, and extinction. The specimen draws in the visitor, but precisely because of that charismatic attraction he or she leaves with a much deeper understanding of dinosaurs.

The Genomic Revolution approached the art of exhibition-making and museum education in a much different fashion. Instead of relying on the allure of an object, the genomic revolution itself, in its abstract and complicated splendor, is what attracted the visitor. The physical specimens were secondary to theories, ideas, and scientific premises. The challenge for the exhibition team lay in translating these difficult concepts into dynamic and decipherable objects that illustrate the genome. To meet this task a team of Museum scientists, experts in the field, and exhibition specialists grappled with the problems for well over a year before delivering "The Genomic Revolution." Over the past decade the AMNH has undertaken production of an additional two human biology genome-oriented exhibitions. "Brain: The Inside Story," which opened in 2012, focused on the new brain research of the twenty-first century at both the imaging level and the genome level. "The Secret World Inside You," which opened in 2015, focused on the human micro-biome, a genome-enabled research area of human health. Both of these exhibitions used the lessons learned from "The Genomic Revolution" to clearly deliver essential information about human health to the general public. In addition to the exhibitions, the AMNH has expanded its research purview to include the science of genomics and informatics. For instance, in 2015, the AMNH in collaboration with other New York City scientists announced the sequencing of the genome of *Cimex lectularius*, the bedbug. The dynamics of genome evolution of this insect pest and its distribution in the New York City subway system was examined in this uniquely AMNH study. The striking success of these exhibitions and the importance of genomic research at the AMNH, starting with "The Genomic Revolution," suggests to us that charisma is not necessarily object based, and for our purposes here, that was encouraging.

For this book, a dinosaur example is again useful. Looking at the *Titanosaurus* skeleton that stretches the length of the Dinosaur Hall Orientation Center (it's actually so big that the designers of the mounted skeleton replica had to arrange its head to stick menacingly out of the entrance to the hall), our imagination takes us to a prehistoric era when dinosaurs ruled. But for the genome our imaginations are used in a much different way. Genes, neurons, and microbes are, in essence, invisible to us. Imagining molecular processes may be of use to a geneticist or biochemist, but for the rest of us picturing the activities of nucleic acids, DNA, and genes is a challenging, if not futile, exercise.

The charisma of the genome lies instead in its possibilities, not simply in what a molecule of DNA can do, but in what DNA can do for us—its potential to better the human condition and to alter our environment in ways once only dreamed of. Therein lies the public's fascination with the genome and with other biotechnologies.

Despite popular and sometimes scientific opinion to the contrary, genes are not the determinative force that many contend or hope they are. Claims of genetic control over intelligence, sexuality, and aggression have come and gone and will come and go again. However, although genes unquestionably contribute to behavioral and medical outcomes, they generally do not govern how we

Figure I.1 The 40-foot *Barosaurus* welcomes visitors every day to the American Museum of Natural History in New York City. This amazing specimen immediately draws visitors into the lives of dinosaurs. *Credit: American Museum of Natural History*

Figure I.2 This artist's conception of a DNA double helix was displayed in the exhibit "The Genomic Revolution." *Credit: Denis Finnin, American Museum of Natural History*

behave or entirely control what diseases we contract or develop. There is a tendency to confuse genetic destiny and genetic potential—a confusion that lies in our changing understanding of gene function. For nearly a century the dominant paradigm in human heredity theory boasted that traits were inherited via single genes (or loci). Scientific support for a one-gene, one-trait approach in genetics was, in fact, borne out by many of the genetic discoveries of the twentieth century. It was easy to show, for example, that certain traits are directly inherited through the mechanism of a single gene. Devastating diseases such as sickle-cell anemia, Huntington disease, and Tay–Sachs disease could all be pinpointed to a single locus. Ultimately, this approach has been fruitful only in the simplest cases of inheritance. The inheritance of these types of diseases is rare, probably accounting for "no more than 5% of known disease." (8) Yet, this single-gene, single-trait approach still holds considerable sway—even more than a decade into the post-genomic era—among the general public. This despite science's failure to genetically understand common and stubborn diseases such as cancer, heart disease, and diabetes, all of which claim many lives each year, and all of which have complex etiologies that are both genetic and environmental. If genetics in the twentieth century was about the search for origins of human traits gene by gene, then twenty-first-century genomics is about the transition away from single-gene thinking and toward thinking about organisms as complex biological systems that are always interacting with our environments.

Genomic technologies are opening up new ways of thinking about the mechanisms of our heredity, disease, and evolutionary history on this planet. For instance, the post-genomic world has altered our view of the microbial world, and scientists are coming to understand how the microbiome impacts our health. In the field of microbial infectious disease, germ theory has dominated the way we look at and treat infectious disorders for over a century. But post-genomic technology allows us to look at and characterize the communities of microscopic organisms involved in the functioning of our bodies in new ways, thus revolutionizing the germ theory of infectious disease. This new germ theory/microbiome paradigm is, in many ways, a radical change. Where we once thought of interactions with microbes as arms races against specific pathogens (using antibiotics to defeat microorganisms that threatened us), we are beginning to view our health more as a détente between our cells and the cells of communities of microbes.

Another area of paradigm shift in health-related research in the post-genome world involves assessing genetic variability between and among human populations. The initial paradigm in using the genome to study genetic disorders was to use a group-based approach (usually using race) called Genome Wide Association Studies (GWAS) to look for correlation of disease with genetic polymorphism. This race-based approach has slowly given way to individualized approaches to health and a new hope for personalized medicine. Coincidentally, the sequencing of large numbers of *Homo sapiens* for projects like the 1000 genomes project have also led to a broader understanding, and in some cases confusions, about the relationships of human population groups to each other and a better understanding of the involvement of genetics in our conception of race.

Genes are not destiny, and such an assertion undermines the astonishing complexity and possibility that are our genes. But if the role of genes in our lives is not this simple, then why read any further? After all, you are reading a "user's guide" to your genes and may have been expecting us to tout the wonders of our genetic code. We are enchanted by the genome and its potential to change our lives in so many ways, but there is so much more to genes and the genomic revolution than the divine-like control and global panacea that is often ascribed to them. By reading this book you will learn about the myths and realities of the genome, and in doing so prepare yourself to be an educated participant in the changes to come. We must remember that the sequence of the human genome is only a first step, and that despite the promises ahead, genomics is still in its infancy. It is likely that we cannot even envision some of what is to come, our imaginations lacking the technological and biological prowess to see a future beyond science fiction. Educating ourselves about the genome will no doubt improve our visionary skills and empower us to be participants in these amazing times. Putting the genome to work raises questions and dilemmas for us as individuals, families, nations, and even as a species. We need to make decisions about our health, our food, our stewardship of the natural world, and our responsibilities to the next generation.

Welcome back to the genome.

REFERENCES

1. Richard Saltus. 2000. "Decoding of Genome Declared," *Boston Globe* (June 27, 2000): p.A1.
2. Nicholas Wade. 2000. "Reading the Book of Life," *New York Times* (June 27, 2000): p.A1.
3. Nicolas Wade. 2001. "Long Held Beliefs are Challenged by New Human Genome Analysis," *NYTimes* (February 21, 2001): p.A20, https://www.nytimes.com/2001/02/12/us/long-held-beliefs-are-challenged-by-new-human-genome-analysis.html; Natalie Angier. 2001. "Reading the Book of Life: Genome Shows Evolution Has an Eye for Hyperbole," *NYTimes* (February 13, 2001): p.F1, https://www.nytimes.com/2001/02/13/science/reading-the-book-of-life-genome-shows-evolution-has-an-eye-for-hyperbole.html; Geoff Dyer, David Firn, and Victoria Griffith, 2003. "Double Helix is Starting to Make its Mark in Medicine," *Financial Times* (July 4, 2003): p.20.
4. Hallam Stevens. 2015. "Networks' Representations and Tools in Postgenomics," in *Postgenomics: Perspectives on Biology After the Genome*. Sarah S. Richardson and Hallam Stevens, eds. Durham, NC: Duke University Press, p.105.
5. Evelyn Fox Keller. 2015. "The Postgenomic Genome," in *Postgenomics: Perspectives on Biology After the Genome*. Sarah S. Richardson and Hallam Stevens, eds. Durham, NC: Duke University Press, p.25.
6. Leif Bertilsson. 1995. "Geographical/Interracial Differences in Polymorphic Drug Oxidation. Current State of Knowledge of Cytochromes P450 (CYP) 2D6 and 2C19." *Clinical Pharmacokinetics* 29: pp.192–209.
7. Terence R. Flotte. 2015. "Therapeutic Germ Line Alteration: Has CRISPR/Cas9 Technology Forced the Question? *Human Gene Therapy* 26: pp.245–246.
8. Kenneth Olden and Samuel Wilson. 2000. "Environmental Health and Genomics: Visions and Implications." *Nature Reviews Genetics* 1: pp.149–153.

1

From Mendel to Molecules

Since the nineteenth century, scientists have been working to unravel the biological basis of inheritance. With Gregor Mendel's mid-nineteenth-century discovery of the basic mechanisms of heredity, genetics was born, and humanity took its first small steps toward deciphering the genetic code. No longer would heredity solely be the domain of philosophers and farmers. Indeed, Mendel's discoveries set the stage for major advances in genetics in the twentieth century and help put in motion the series of discoveries that led to the development of the sequencing of human and nonhuman genomes. This age of discovery, from Mendel to genome sequencing, is the subject of the first four chapters of this book. Chapter 1 covers some basic biology and tells the story of the evolution of genetics by examining some of the most significant discoveries in the field—discoveries that enabled the development of genomics. Chapter 2 looks specifically at the evolution of genetic and genomic sequencing technologies. Chapter 3 examines the human genome itself and the ways in which we are exploring and exploiting it now and in the future. And, finally, Chapter 4 looks at the sequencing and genome analysis tools of the post-genomic era also called next generation sequencing or (NGS).

Without any further ado, may we present to you the human genome!

This photo (Figure 1.1), also known as a karyotype, shows the 46 human chromosomes, the physical structures in the nuclei of your cells that carry almost the entire complement of your genetic material, also known as your genome. But don't let this two-dimensional representation of the genome fool you into believing in its simplicity. Almost 20 years ago biologist Richard Lewontin called DNA a "triple-helix" to explain how genes function, and how they interact with each other and the environment. This triple helix is largely inseparable, and genetics doesn't make sense unless taking these effects into account.

We could also have introduced you to your genome with a slew of the DNA sequence units—As, Ts, Gs, and Cs—in a string, or we could have shown you a picture of DNA in a test tube or even a picture of a nucleus of one of your cells where the DNA would be visible as dark stringy stuff. There are many ways to visualize the genome and this is part of its beauty.

Welcome to the Genome: A User's Guide to the Genetic Past, Present, and Future, Second Edition.
Rob DeSalle and Michael Yudell.

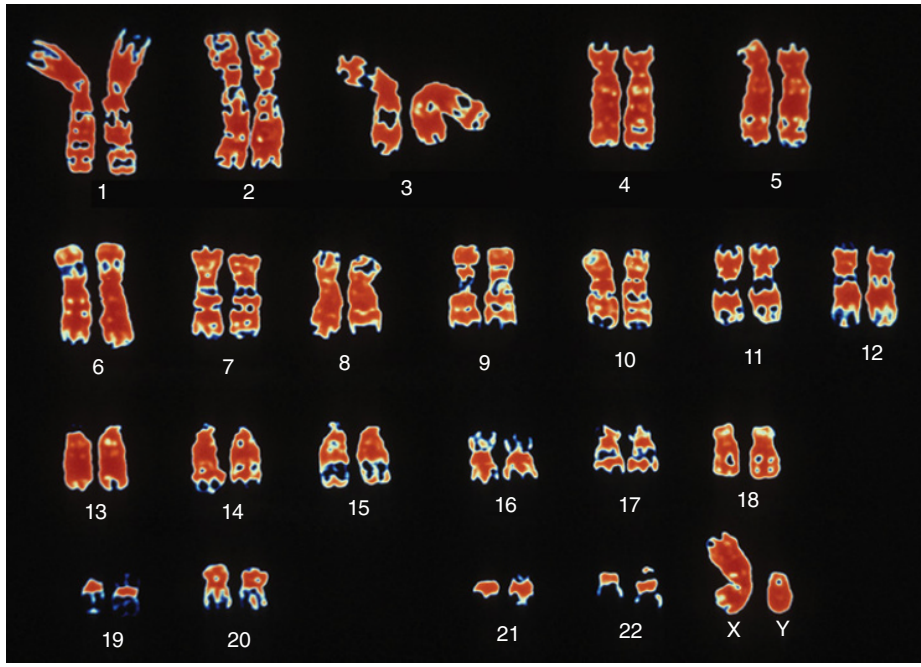


Figure 1.1 This picture, known as a karyotype, is a photograph of all 46 human chromosomes. With an X and a Y chromosome, this is a male's karyotype. A female's karyotype would show two X chromosomes. *Credit: Photo Researchers*

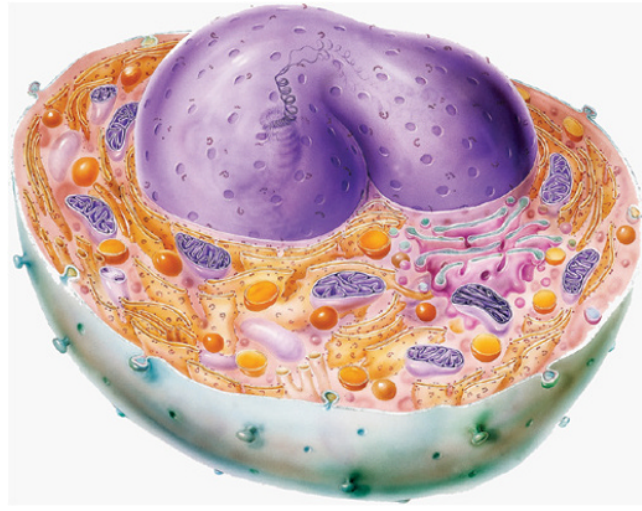


Figure 1.2 The nucleus of every human cell (the large purple mass inside the cell) contains DNA. Mitochondria, organelles in cells that produce energy (the smaller purple objects within the cell), also contain some DNA. *Credit: Wiley*

Still, to understand function, we do need to learn about basic form. And a karyotype, despite its limitations as a representation of the genome, illustrates that in almost all the cells in the human body there are 22 pairs of chromosomes and two sex-determining chromosomes. The double helices that make up your chromosomes are composed of deoxyribonucleic acid, also known as DNA, on which are found approximately 20,000 genes. These cells are called somatic cells, and they are found in almost all nonreproductive tissue.

Humans also have cells with 23 nonpaired chromosomes. In these cells, each chromosome is made up of a single double helix of DNA that contains approximately 20,000 genes. These cells are called germ cells and are the sperm and egg cells produced for reproduction. These germ cells carry a single genome's worth of DNA or more than 3 billion bases worth of nucleic acids.

Chromosomes are somewhat like genetic scaffolding—they hold in place the long, linearly arranged sequences of the nucleotides or base pairs that make up our genetic code. There are four different nucleotides that make up this code—adenine, thymine, guanine, and cytosine. These four nucleotides are commonly abbreviated as A, T, G, and C. Found along that scaffolding are our genes, which are made from DNA, the most basic building block of life. These genes code for proteins, which are the structural and machine-like molecules that make up our bodies, physiology, our mental state. Through the Human Genome Project scientists are not simply learning the order of this DNA sequence, but are also beginning to locate and study the genes that lie on our chromosomes. But not all DNA contains genes.

On average 3 billion base pairs exist in the collection of the chromosomes your mother transmitted to you. Add to that the chromosomes given to you by your father gave you and in your cells there are around 6 billion bases, a complete diploid human genome. There are long stretches of DNA between genes known as intergenic or noncoding regions. And even within genes some DNA may not code for proteins. These areas, when they are found within genes, are called introns. While these genomic regions were once believed to have no products and/or no function, scientists now understand that both introns and intergenic regions play a role in regulating DNA function. The Encyclopedia of DNA Elements or ENCODE Project estimates, for example, that while only 2.94% of the entire human genome is protein coding, 80.4% of genome sequences might govern the regulation of genes. (1) Unlike the human genome and all other eukaryotic genomes, however, bacterial genomes do not have introns and have very short intergenic regions. Curiously though, the archaea, a third major domain of life (in addition to eukaryotes and bacteria) do have introns, but not necessarily the same kind of introns as eukaryotes.

Let's begin our tour of the human genome with a very basic lesson in genetic terminology. For example, what exactly is genetics, and how is it different from genomics? Genetics is the study of the mechanisms of heredity. The distinction between genetics and genomics is one of scale. Geneticists may study single or multiple human traits. In genomics, an organism's entire

collection of genes, or at least many of them, is examined to see how entire networks of genes influence various traits. A genome is the entire set of an organism's genetic material. The fundamental goal of the Human Genome Project was to sequence all of the DNA in the human genome. Sequencing a genome, whether human or nonhuman, simply means deciphering the linear arrangement of the DNA that makes up that genome. In eukaryotes (plants, animals, fungi, and single-celled organisms called protists), the vast majority of the genetic material is found in the cell's nucleus. The Human Genome Project has been primarily interested in the more than 3 billion base pairs of nuclear DNA. A tiny amount of DNA is also found in the mitochondria, a cellular structure responsible for the production of energy within a cell. Whereas the human nuclear genome contains more than 3 billion base pairs of DNA and approximately 20,000 genes (that's nearly 10,000 genes fewer than when the first edition of this book was published in 2005), the reference human mitochondrial genome contains only 16,568 bases and 37 genes. (2) Like bacteria, mitochondrial DNA, or mtDNA, has short intergenic regions and its genes do not contain introns. Another interesting characteristic of mtDNA is that it is always maternally inherited. This has made mtDNA very helpful to track female human evolutionary phenomena. These discoveries were made possible, in part, by sequencing mtDNA.

What about heredity? In the most basic sense we should think about heredity as the transmission of traits from one generation to the next. When we talk about heredity in this book we refer to the ways in which traits are passed between generations via genes. The term heredity is also sometimes used to describe the transmission of cultural traits. Such traits are shared through a variety of means including laws, parental guidance, and social institutions. Unlike genetics, however, there are no physical laws governing the nature of this type of transmission.

What are genes? Genes are regions of DNA and are the basic units of inheritance in all living organisms. These words, genes and DNA, are too often used interchangeably. Both genes and DNA are components of heredity, but we identify genes by examining regions of DNA. In other words, DNA is the basic molecular ingredient of life, whereas genes are discrete components of that molecular brew.

If you look at any family you'll see both shared and unique traits. Family members typically look alike, sharing many features such as eye color and nose shape, but they may also have very different body types and be susceptible to different diseases. This diversity is possible for two reasons. The first reason is that genes come in multiple forms. These alternative forms are known as alleles, and in sexual reproduction they are the staple of organismal diversity. According to the laws of genetics, siblings can inherit different traits from the same biological parents because there is an assortment of alleles that can be randomly passed along. The second reason is that the environment can exert a significant influence on the expression of genes. For example, an individual