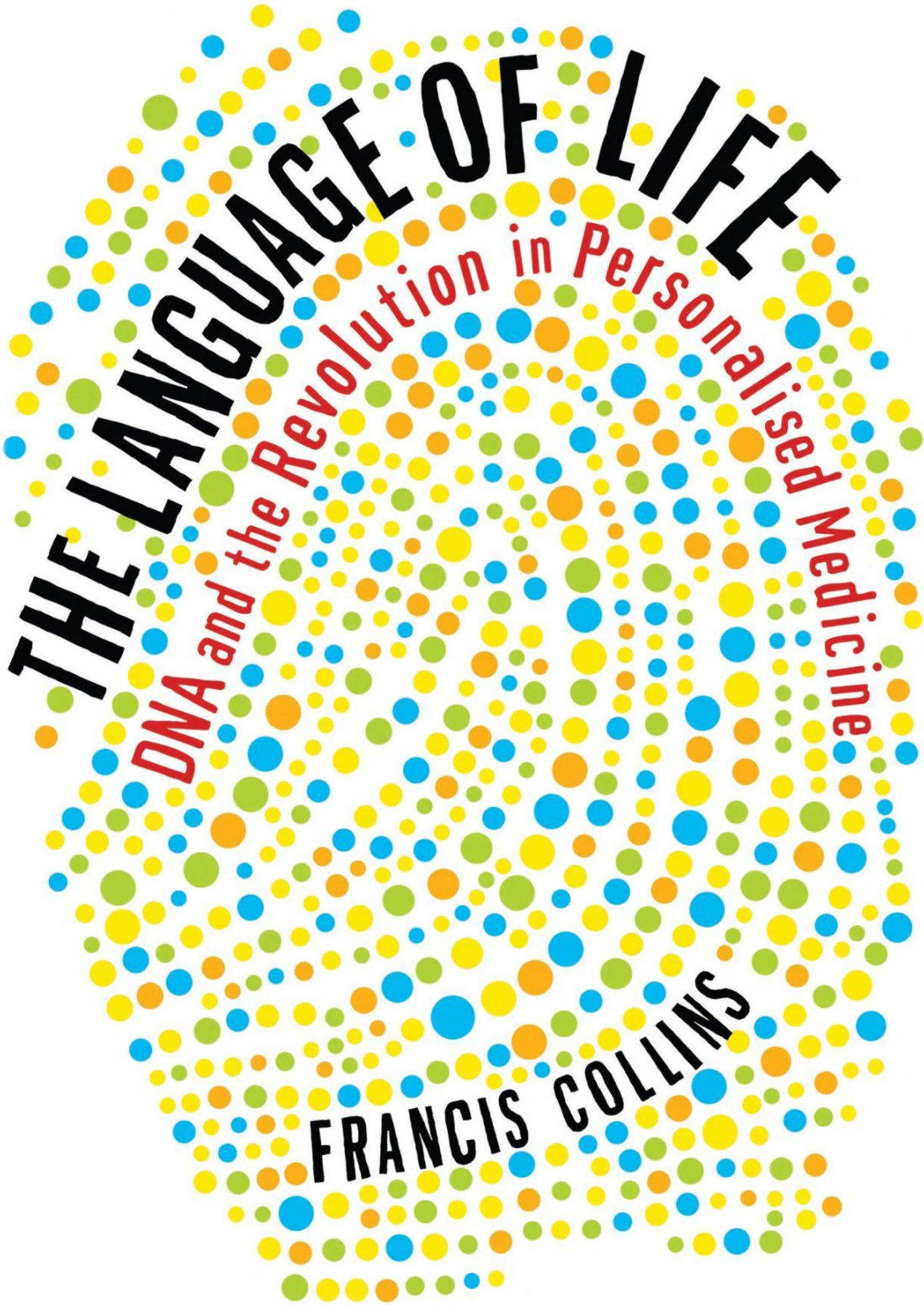
'His groundbreaking work has changed the very ways we consider our health and examine disease' Barack Obama



The Language of Life

DNA and the Revolution in Personalized Medicine

Francis Collins

PROFILE BOOKS

First published in Great Britain in 2010 by
PROFILE BOOKS LTD
3A Exmouth House
Pine Street
London EC1R 0JH
www.profilebooks.com

First published in the United States of America in 2010 by
HarperCollins Publishers

10 9 8 7 6 5 4 3 2 1

Designed by Joy O'Meara

Printed and bound in Great Britain by
Clays, Bungay, Suffolk

A CIP catalogue record for this book is available from the British Library.

ISBN 978 1 84668 353 4

The paper this book is printed on is certified by the © 1996 Forest Stewardship Council A.C. (FSC). It is ancient-forest friendly. The printer holds FSC chain of custody SGS-COC-2061

Contents

Acknowledgments

I am deeply grateful to the many selfless individuals and families who allowed me to include their personal stories in this book, providing the reader with important insights into the realities of personalized medicine. That list includes Blake Aldhaus and his mother, Anita; Tracy Beck; Sam Berns and his parents, Scott Berns and Leslie Gordon; Sergey Brin and Anne Wojcicki; McKenzie Christeson and her parents, Scott and Kris Wood; Bill Elder; Marvin Frazier; Doris Goldman; Jeffrey Gulcher; Wayne Joseph; Judy Orem; Kate Robbins; Anabel Stenzel and Isabel Stenzel Byrnes; Dale Turner; and several others who chose not to have their names used.

I am also greatly indebted to a long list of busy experts who took the time to read a draft of parts or all of the manuscript and provided valuable feedback on its content. That includes Melissa Ashlock, Barbara Biesecker, Stephen Chanock, Marc Chevalier, Brandon Collins, Tony Fauci, Greg Feero, David Ginsburg, Alan Guttmacher, Richard Hodes, Kathy Hudson, Tom Insel, Mark Kay, Teri Manolio, Judy Mosedale, Sharon Terry, Larry Thompson, Eric Topol, and Dick Weinshilboum. Their collective suggestions made this book substantially more accurate and more accessible than it otherwise would have been, but I take full responsibility for any errors that remain.

Thanks go to Judy Hutchinson, who carefully transcribed the dictated first drafts of each chapter; to my wife, Diane Baker, who provided valuable input on content and keyed in all of the initial edits;

and to biomedical illustrator Darryl Leja, who ably and elegantly rendered all of the artwork. Finally, I want to thank my publishing agent, Gail Ross, for her unflagging support; and my editor, Bruce Nichols, for believing in the importance of this book even before I did, for encouraging me to make it a reality, and for providing wise perspective all along the way.

INTRODUCTION

We're Not in Kansas Anymore

The tearful young man was on the phone with his uncle. "My mother is dying. She's in a coma, and I don't think she'll make it through the night." Surrounded by the drone of whirring centrifuges and students' conversation about the previous night's lab party, Dr. Robert James moved to a quiet spot where he could speak privately to his distraught nephew.

"I'm so sorry, Brad," he said. "Your mother truly fought a valiant battle against ovarian cancer, and she rallied so many times when it seemed that all was lost. But it sounds as if this is really the end. What can I do to help?"

"Well," said his nephew, "my sister and I have been worrying about whether her cancer might be hereditary, given all of the other women on her mother's side who suffered from breast cancer or ovarian cancer. You once told us that someday there might be a test to determine whether one or both of us had inherited her cancer risk. If that's true, is it too late to pursue this?"

Dr. James explained how to proceed. A blood sample from his dying sister-in-law was shipped to Dr. James's lab the next day, DNA was prepared, and the sample was carefully stored in the freezer. He

thought it was unlikely this would ever be useful, but at least it was something to do.

Five years later, Brad's sister Katherine contacted Dr. James, explaining that she had been reading articles in the popular press about the discovery of genes involved in hereditary breast and ovarian cancer. Katherine had been having yearly mammograms, even though she was only in her thirties, but she was particularly concerned that there were no good screening tests available for detecting early ovarian cancer. Her mother had originally been diagnosed with this cancer at 52, and Katherine thought every day about her own potential for developing it.

Dr. James confirmed that the discovery of genes known as *BRCA1* and *BRCA2* might well make it possible to be more precise about the risk of cancer in the family, if it turned out that Katherine's mother carried a mutation in one of these genes. Concerned about the risk of losing her health insurance if she tested positive, Katherine wanted to know whether there was some other way to get the information. Her uncle told her about a clinical research study in a nearby city that allowed testing under an assumed name, and Katherine decided to proceed. After genetic counseling about the risks of knowing or not knowing this information, Katherine requested that the DNA sample on her mother, carefully stored for several years in the freezer in Dr. James's laboratory, be forwarded to a testing facility.

A few weeks later Katherine called Dr. James to report that a significant *BRCA1* mutation had been found in her mother's DNA. Katherine faced a 50 percent risk of having inherited that misspelling, in which case her lifetime risk of breast cancer would be approximately 80 percent, and that of ovarian cancer about 50 percent. Katherine was deeply concerned about herself, but even more so about her six-year-old daughter.

She spent two weeks waiting for her own results, and it seemed

like an eternity. She tried to imagine what she would do with a positive result. Would she approach a surgeon about removing her ovaries? Would she even contemplate removing both breasts and undergoing surgical reconstruction, as many women with mutations in *BRCA1* or *BRCA2* have done? What would she tell her daughter, and at what age should her daughter be tested? Some days she was certain the test would be positive—after all, everyone remarked how much she looked like her mother. On other days, she remembered that such information was irrelevant to the possibility of her carrying this specific genetic glitch, and she was more hopeful.

The fateful day arrived when a phone call from the genetic counselor invited Katherine to come to the clinic to hear the results. With her heart in her mouth, she sat across the desk as the counselor opened the file and then broke into a smile. "Katherine," she said, "I have good news. You have not inherited the *BRCA1* mutation carried by your mother. Your risk of breast and ovarian cancer is no greater than that of the average woman of your age, and your daughter likewise carries no special risks for these diseases."

Overjoyed, Katherine called her uncle to share this happy moment. But both of them confessed to remaining uneasy about other maternal relatives in Canada and Europe, and about Katherine's brother, Brad, who had chosen not to be tested. Although males with mutations in *BRCA1* and *BRCA2* face only a slightly increased risk of cancer of the prostate, pancreas, and male breast, their daughters may still be at high risk of breast and ovarian cancer if they've inherited the mutation. Brad's young daughter now became the remaining member of this nuclear family with a potential genetic cloud over her.

Dr. James is a physician who has devoted his professional life to research on molecular genetics, so it was ironic that his own family turned out to be affected by one of the more dramatic discoveries in hereditary disease of the past decade.

But then it happened again. This time, it was his father-in-law, Fred, now in his late seventies, who contacted Dr. James about a medical evaluation. Fred had noticed some discomfort in his legs and a deterioration in his golf game and, after an initial evaluation by his primary physician, had been referred to a neurologist.

Fred was calling to say that the neurologist had detected some slowing of nerve conduction in his legs, and was suggesting that Fred should be tested for an uncommon genetic condition known as Charcot-Marie-Tooth disease, named for the three French investigators who originally identified it. Dr. James was initially appalled at the idea of such testing, since Charcot-Marie-Tooth disease was generally associated with progressive weakness in the legs beginning in the twenties and thirties. Thinking that a genetic test for this condition in an elderly man would be essentially a waste of time and money, Dr. James nonetheless did not voice an objection to this plan, since he didn't want to interfere with his father-in-law's medical evaluation. To his amazement and consternation, the test was positive. After more study of the problem and discussion with the experts, it began to make more sense. Until DNA testing was made available, Charcot-Marie-Tooth disease had been purely a clinical diagnosis. So of course the cases that were discussed in textbooks and medical journals tended to be those with a more severe course. Now that the gene had been identified and could be spotted by a specific molecular test, it was becoming apparent that a milder disease, including the remarkably late onset presented by Fred, was more common than had been appreciated.

This time the diagnosis struck even closer to home. Charcot-Marie-Tooth disease is a dominant condition, and this means that the child of an affected individual has a 50 percent chance of inheriting the abnormal gene and also being affected. Thus Dr. James's wife, Dawn, as well as her brother and sister, might be significantly

affected by this discovery in the future. In fact, it wasn't just a matter of the future; it was also about the past and the present. Dawn's sister, Laura, had long struggled with what had been assumed to be a congenital problem with her feet and ankles. Blamed on "club feet," but never definitively diagnosed, this problem now appeared likely to be a consequence of a particularly early manifestation of the same genetic disease that had appeared so late in her father. Here was a chance to provide a definitive diagnosis. Yet Laura decided not to be tested. She was not convinced that the information would change anything, and she was a cynic about health care. She had had several frustrating experiences over the years with orthopedic interventions that were supposed to help her chronic foot problems but didn't really provide much relief. She respected her brother-in-law, Dr. James, but not the system.

For her part, Dawn considered the possibility of testing, even though she had no symptoms of this disease and was now in her mid-fifties. She ultimately decided to embrace the ambiguity of the situation, rather than obtain a definitive answer, as she was not sure how a positive test result would change her outlook. Dr. James was somewhat puzzled, but he supported her decision. After all, she was healthy and happy. By contrast, he wished he could change her sister's mind. Shouldn't Laura know why she had suffered so long?

Robert James happens to be an M.D. and a geneticist. How surprising is it that he faced two situations involving genetic testing and risk in his own family? Actually, not very. The National Organization for Rare Diseases (NORD) estimates that there are at least 6,000 rare (so-called orphan) diseases, defined as conditions that affect fewer than 200,000 people in the United States. Collectively, 25 million Americans are affected by one of these conditions. If you include their families and friends, then few of us have not been touched in some way by one of these conditions. Many of them are caused by genes

these are early days for making accurate predictions, I still decided it was time to find out. I conferred with my adult daughters, since this kind of testing might also reveal things about them, and they encouraged me to "go for it."

Family medical history is of course a critically important guide. I am blessed with a remarkably healthy group of close relatives—both my parents lived to age ninety-eight, and my three brothers (all older than me) are all athletic and in excellent health. So my own likelihood of future illness is hard to discern from my pedigree. But could there be risks lurking in my DNA that have not shown themselves?

Besides this curiosity about my own genome, I was also interested to find out how these direct-to-consumer companies conduct business and report results. Is their laboratory work accurate? How do they convert a DNA result into a prediction about risk? And how good are they at conveying that information in a fashion that empowers rather than confuses the consumer?

I decided to submit a DNA sample to each of the three companies offering comprehensive DNA analysis. (There are quite a number of other companies—some credible, some not—that are more focused on specific tests for specific purposes.) I decided not to use my own name, as I didn't want these companies to treat me any differently than they would a typical customer.

The costs of the test were substantially different: 23andMe charged just $399, whereas deCODE cost $985, and Navigenics charged $2,499 (but offered telephone genetic counseling as an added feature). The DNA sampling processes were easy: spitting into a special tube for 23andMe and Navigenics, and scraping my cheek for deCODE. Each company promised confidentiality through assignment of passwords to their Web site. And while there were some interesting differences, the lists of conditions tested were heavily overlapping (see Appendix E).

23andMe was the first to report results, in just two weeks. deCODE weighed in a couple of weeks later, and Navigenics reported after seven weeks (but strangely they had not actually completed the analysis, and 7 of the 25 conditions being tested for still had some results pending). As much as I knew the significant limitations of the tests to make precise predictions, I still found it both exciting and a bit unnerving to enter my password and begin to review my own results. Each Web site was reasonably well designed to help me understand the results and to put my own risk in context of the average person. Of the three, I found the 23andMe Web site to be most user-friendly.

To assess genetic risk, all three companies base their work on the same publications in the scientific literature. So in many instances they tested exactly the same variants in my DNA. I looked closely at the details to see if any of the actual lab results were discordant. To my relief, I couldn't find a single example where that was so. So the actual DNA analysis is apparently of very high quality.

What did I learn? For most common diseases, I was happy to see that my risk scored as average or below average. But there were some significant exceptions. All three companies agreed that my risk for type 2 (adult-onset) diabetes was elevated. Though the precise risk estimate varied slightly, my risk came in at about 29 percent, somewhat higher than the average person (23 percent). My risk of age-related macular degeneration, a common cause of blindness in the elderly, and which had taken my aunt's eyesight in her eighties, was also substantially higher than that of the average person. And the chance that I would be affected by a particular type of glaucoma was also elevated, though the companies disagreed about the absolute risk.

Of course this was all statistical information—there was no proof that I would definitely get any of these diseases, and the predictions didn't take into account my family medical history at all. But despite

my being aware of all the shortcomings of these tests, the information had an immediate effect on my view of the future. As a physician, I had known for years about a long list of general recommendations for maintaining good health, but I hadn't necessarily followed them. Now, with these specific threats, I found I was more attentive. Even though the predicted 29 percent risk of diabetes was marginally higher than the 23 percent baseline, and even though my negative family history and absence of obesity no doubt reduced my risk even further, I resolved to go ahead with a long-postponed plan to contact a personal trainer and work harder at a diet and exercise program, knowing that this was the best prevention for whatever diabetes risk still remained. I looked up the most recent research articles on macular degeneration, and concluded that the evidence supporting the protective effect of omega-3 fatty acids was solid enough that it would be a good idea for me to include more fish in my diet. And given the glaucoma risk, I resolved to be sure to have my eyes checked each year, including measurement of intraocular pressure. Were these all things I should have been doing anyway? Perhaps. But we are constantly bombarded by all kinds of generic health advice—eat fish! take a daily aspirin! drink red wine! exercise!—and it's hard if not impossible to remember to do all these things. Despite all of the limitations of the data, the disclosure of this personalized genetic information provided a motivator for specific actions.

There was one test result I thought seriously about just not looking at—the one for Alzheimer's disease risk. This is one of the strongest genetic risk factors yet identified, capable of increasing one's risk by as much as eightfold. And at the present state of medical research, there is nothing you can do about it, other than use the information to try to plan for the future. There's no convincing evidence that diet or medication will delay or prevent the onset of Alzheimer's disease in a susceptible person. Despite my negative family history for Alzheimer's

disease, I felt my heart rate go up as I decided to click on the button and reveal the result. The answer was a relief—my lifetime risk of Alzheimer's disease comes out lower than average, at just 3.5 percent.

A few other results caught my eye. 23andMe and deCODE reported on my ability to metabolize a commonly used drug for blood clots, called coumadin. I have never taken that drug, but my mother was on it for several years and proved to be unusually sensitive, so that her dose had to be adjusted downward to avoid toxicity. Sure enough, the 23andMe report predicted that I would also have "increased sensitivity." Oddly, deCODE looked at exactly the same variants in my DNA, got the same results, but predicted I would need an "average dose." This was a good reminder of the immature state of making predictions from these DNA results. These companies are all looking at the same scientific evidence, but regrettably they haven't achieved consensus on interpretation. They should get together urgently to do this, or the public may start to become confused and potentially disillusioned.

This discordance between the results from the three companies was most apparent for the prostate cancer risk prediction. My father had this disease late in life, and so when my 23andMe results arrived, I was relieved to see a prediction of lower than average risk. But then deCODE disagreed, saying my risk was slightly elevated. Navigenics upped the ante substantially, placing me at a 40 percent higher risk than the average male (24 percent compared to a baseline of 17 percent). What on earth was going on here? To sort this out, I had to drill down into the details of the lab studies—and I discovered the explanation. 23andMe had tested for just 5 variants known to confer prostate cancer risk; deCODE had tested for 13; and Navigenics had tested for 9. There was considerable overlap between the DNA markers tested, but no company had actually tested for the complete set of 16. Having all the results in front of me, I could calculate that risk,

and it came rather close to the Navigenics prediction. So the reassurance I had first obtained from 23andMe was short-lived—here was another condition that I should pay close attention to.

There's a really important lesson here—the field is moving so quickly that any genetic risk predictions based on today's understanding will need to be revised in the context of new discoveries tomorrow. That applies not just to prostate cancer but to all of the rest of my risk predictions—what is possible now is only a blurry picture of reality. As genetic tests get better, and other critical information such as family medical history and current medical status get more effectively integrated with the DNA results, the picture will come increasingly into focus. So anyone embarking on this adventure should be prepared to revisit the risk estimates on a regular basis as new knowledge is obtained.

As the most expensive of the three options, and the one most focused on medical applications, Navigenics also offered the chance to consult with a genetic counselor about my results. I spoke to one of their counselors on the phone, playing my role as an interested consumer without much scientific training. The counselor was careful to say she was not dispensing medical advice, but after going over my DNA results she strongly recommended that I see a physician about my prostate cancer risk. I expressed concern that my physician might not know what to make of these genetic tests, and she indicated that lots of physicians were now calling Navigenics for advice. I asked whether these DNA-based predictions might change in the future, and she correctly pointed out that new information was being derived every day, and Navigenics would keep me informed by e-mail as the predictions became more refined. Oddly, however, she implied that most of the remaining genetic risk factors for common disease will have been discovered in the next two or three years; as a scientist working in this field, that seems quite unlikely to me.

manent part of your electronic medical record, and will be utilized by health care professionals to make a wide variety of decisions about drug prescriptions, diagnostics, and disease prevention. If you fall ill, the therapeutic options waiting for you, many derived from new understanding of the human genome, will be both more effective and less toxic than the treatments available just a few years ago. Many of these therapies will be in pill form, but some will be gene therapies, in which the gene itself becomes the drug. Some will even be cell therapies, based upon the ability to take your own skin cells or blood cells and transform them into cells you might need in, say, your pancreas for diabetes, or your brain for Parkinson's disease.

This book is a report from the front lines of a revolution. It is also a user's manual of what you need to know to benefit your own health, and that of your family. There are things you can do right now—starting with a family medical history—to prepare. But first, you have to be ready to embrace this new world.

For centuries, we considered ourselves to be healthy until symptoms of illness arose. Once diagnosed, correctly or not, we received standardized treatments. In accordance with this view, the human body was generally ignored until something went wrong.

Today, we have discovered that everyone is born with dozens of genetic glitches. There are no perfect human specimens. But not all our glitches are the same, so one treatment often does not fit all sufferers of a given disease. Not just our medicine but our fundamental attitude toward the human body is changing.

A lot has been written—often breathlessly—about the DNA revolution. But this book aims to be about the facts. This is a book about hope, not hype. The accelerating ability to read the language of life is allowing a completely new view of health and disease. If you are interested in living life to the fullest, it is time to harness your double helix for health and learn what this paradigm shift is all about.

The Language of Life

CHAPTER ONE

The Future Has Already Happened

Scientists aren't generally prone to effusiveness. We are privately excited about our work, but in public we often, and rightly, emphasize skepticism and caution. But there are exceptional moments where skepticism is set aside, electricity fills the room, and a scientist with palpable passion and flashing eyes describes unabashedly a change in the landscape that will have lasting significance.

Just five months into the new millennium, I had that experience. Together with more than 2,000 of my colleagues, laboring in 20 centers in 6 countries, we had succeeded in reading almost 90 percent of the letters of the human *DNA* instruction book, otherwise known as the human *genome*. After much anticipation, and many tumultuous moments, the achievement of an almost impossibly audacious goal that had motivated all of us for a decade was now essentially assured.

The public announcement of the complete draft of the genome would follow one month later at the White House. But on this Saturday in May 2000, it fell to me, as the "field marshal" of the international Human Genome Project, to deliver the keynote address at the annual gathering of the genome science community, held at Cold Spring Harbor Laboratory on Long Island. This was the private,

treatments following the lumpectomy, and then was faced with deciding whether to proceed with chemotherapy to reduce the risk of a future recurrence. She consulted with no fewer than three oncologists, all of whom recommended aggressive chemotherapy because of her relatively young age. Karen struggled with that recommendation, but given the unanimity of medical opinion, she resigned herself to proceeding with chemotherapy. She began to explore what kind of wig she might need to purchase.

To her great surprise, she received a phone call from her brother, who was not a medical professional but had seen a story on television about a new test that could be done on the actual breast cancer tissue to make a more precise prediction about the likelihood of recurrence. This test was based on an analysis of which genes were turned on or off in the tumor. Validation on thousands of cases had shown that this "gene expression" analysis gave a more precise picture of that tumor's likely aggressiveness than the traditional approach of looking through a microscope at the appearance of the cells.

Karen consulted with her surgeon, who had heard of the test, though he had not had much experience with it. He agreed to send off her tumor sample to the testing laboratory. Just four days before she was due to start chemotherapy, the test came back, indicating a very low recurrence score. One of her three oncologists was skeptical, but the other two were convinced that this information provided justification for pursuing hormonal therapy alone. Karen decided to adopt that approach. Four years later, she remains completely free of any signs of disease. This particular test represents one of the first fruits of the genomic revolution, and Karen is one of its pioneers.

Karen's case exemplifies a new approach to medicine that will soon affect virtually all facets of health care. No longer satisfied with empirical or superficial explanations of disease, scientists are peering

into the molecular basis of cancer, heart disease, diabetes, Alzheimer's disease, schizophrenia, autism, and virtually all other conditions, peeling off the layers of the onion and finding that many accepted principles of medicine and biology require substantial revision. Fundamental gaps in our understanding about the human body are now being filled in. Hereditary factors for nearly all diseases are now being pinpointed as specific glitches in DNA, and these are appearing in great numbers following the completion of the Human Genome Project. As a result, healthy individuals are increasingly able to discover some of their body's inner secrets and take appropriate action. The potential for individual prediction is beginning to spill out to the general public, offering the opportunity to take more control of your fate.

Those who develop a disease, like Karen, are now offered molecular tools to predict the course, or even to decide that therapy isn't necessary. And the range of therapeutic options is expanding, as knowledge about the human genome provides new targets for the development of powerful treatments. None of this is happening overnight, and ultimate success will depend upon the visionary investment of energy, talent, and financial resources by scientists, governments, universities, philanthropic foundations, biotechnology and pharmaceutical companies, and the general public. But without question, man's knowledge of man is undergoing the greatest revolution since Leonardo.

DNA IS THE LANGUAGE OF LIFE

The discoveries of the past decade, little known to most of the public, have completely overturned much of what used to be taught in high school biology. If you thought the DNA molecule comprised thou-

sands of genes but far more "junk DNA," think again. If you thought the human genome must be the most complex version of DNA on earth, think again.

For the purposes of this book, there is no need to learn every detail of DNA structure (for some of that, see Appendix B). This book is about applications, not engineering. But to understand those applications, it is important to learn some of the principles and some of the vocabulary.

Bacteria have DNA. Yeast have DNA. So do porcupines, peaches, and people. It is the universal language of all living things. We are in a truly historic era, when this language from many different species is being revealed for the first time. All of the DNA of an organism is called its *genome*, and the size of the genome is commonly expressed as the number of *base pairs* it contains. Think of the twisted helix of DNA as a ladder. The rungs of the ladder consist of pairs of four chemicals, called bases, abbreviated *A, C, T, G*. As shown in Figure 1.1, DNA is a long ladder. Its backbone is a monotonous string of sugars and phosphates. The information content resides in those chemical bases arranged within the interior, where A always pairs with T, and C always pairs with G. The simplest free-living single-cell organisms, such as bacteria, generally pack all their information into a genome of a few million base pairs. Fancier multicellular organisms with more complex body plans require larger genomes to specify those functions. Our own genome stacks up as 3.1 billion rungs of the DNA ladder. Most other mammals have genomes of about that size, give or take a billion or so, but many amphibians have genomes substantially larger than ours, and a very simple plant called the whisk fern, lacking flowers, fruit, or even leaves, has a genome 100 times larger than our own!

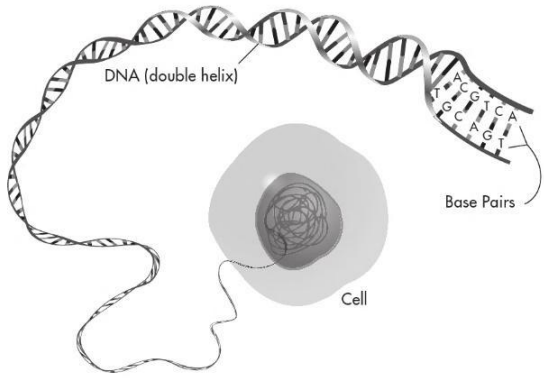


Figure 1.1: The double-helical DNA information molecule, the "instruction book" of all living things, here shown spilling out of the nucleus of a cell. The information content of DNA is specified by the order of the chemical bases (A, C, G, or T). Each of the two strands carries the complete information, since A always pairs with T, and C always pairs with G.

A *gene* is a segment of the DNA ladder that carries a packet of functional information. The shortest genes are only a few hundred base pairs in length; the longest, the Duchenne muscular dystrophy gene, stretches to more than 2 million rungs of the ladder. The best-understood genes are those that code for *protein*. This process involves first making an *RNA* copy of the DNA; that RNA is then transported to the ribosome "protein factories" in the cytoplasm, where the letters of the RNA code are translated into the amino acids used by proteins (Figure 1.2). This translation is carried out using a triplet code word; for example, AAA in the RNA codes for the amino acid lysine, and AGA codes for arginine. Mistakes in the DNA will lead to mistakes in the RNA, and that can result in garbling of the protein (Figure 1.3).

One major surprise from the Human Genome Project was the discovery that human DNA contained only 20,000 protein-coding genes. We expected a lot more than that! Even the lowly roundworm has about 19,000 genes. Some observers were actually upset by this

apparent downgrading of our importance, but we assume our genome must be fancier in some other way. After all, we are the only species that has sequenced our own genome!

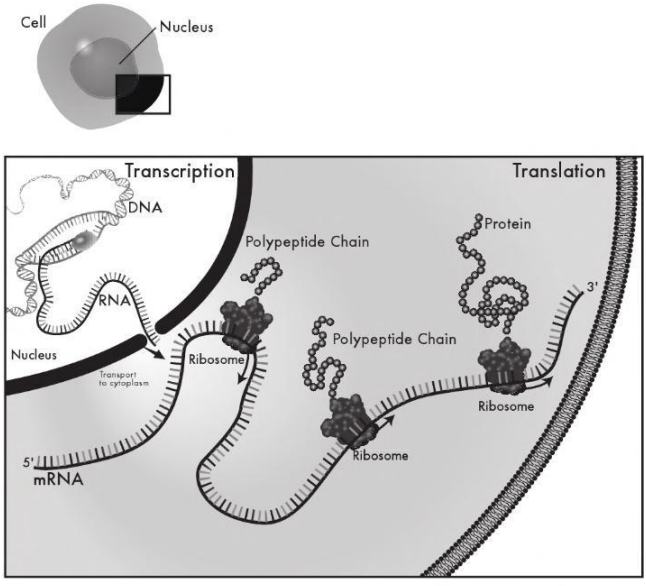


Figure 1.2: The "basic dogma of molecular biology": DNA codes for messenger RNA (called transcription, takes place in the nucleus), which then codes for protein (called translation, carried out by ribosomes in the cytoplasm).

Normal	DNA	HAS	ALL	YOU	CAN	ASK	FOR
Missense	DNA	HAS	ALL	LOU	CAN	ASK	FOR
Nonsense	DNA	HAS	ALL	YOU	STOP		
Frameshift	DNA	HAS	ALY	OUC	ANA	SKF	OR

Figure 1.3: Translation from RNA to protein occurs using a word length of three letters. Mistakes in the DNA genome lead to mistakes in messenger RNA, which then lead to mistakes of various types in the protein.

Most genes are internally interrupted in a puzzling way by long stretches of DNA information that are removed, or "spliced out," in

liver cell different from a brain cell or a muscle cell. These different programs are carried out by various proteins that bind to the DNA, and switch on or off the genes nearby.

Each time a cell divides, the entire genome has to be copied. But mistakes can creep in. Occasionally these can cause a cell to grow more rapidly than it should, and that can even lead to cancer. The environment can play a role here, in that carcinogens such as radiation and cigarette smoke increase the mistake rate of DNA copying.

BEYOND THE HUMAN GENOME PROJECT

Since 2003, rapid progress has been made by building upon the foundation provided by the original sequence of the human genome. Much more hard work is needed, since our ability to interpret the more than 3 billion letters of our own language of life is still rudimentary, and requires many other sources of information in order to make sense of this vast sea of data. With the cost of *DNA sequencing* continuing to drop at a breathtaking rate, it has become possible to determine the complete genome sequence of a vast array of other organisms, including hundreds of microbes and dozens of invertebrates and vertebrates. For some of these organisms, such as mice, rats, and dogs, having the genome sequence is valuable in its own right, since substantial research communities are devoted to understanding their biology. But all of these sequences inform us about the human genome, too. After all, if a particular segment of human DNA shows strong connections with other mammals, or even with species that lie farther away in the evolutionary tree, this particular segment must have performed an important function that does not tolerate much variation over evolutionary time.

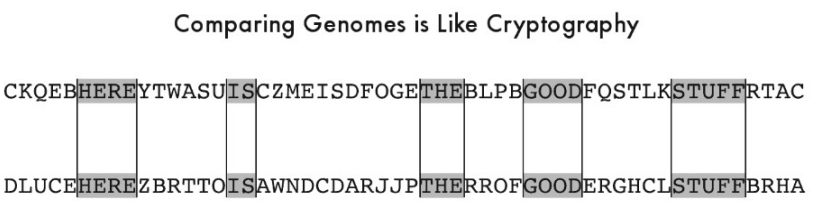


Figure 1.4: Comparing genomes from different organisms is a powerful method to identify the parts that are most functionally important, as these will have been the most constrained during the process of evolution.

Figure 1.4 shows in simplified form how this kind of comparison can reveal genome features that may otherwise be hidden in a sea of gibberish.

Remarkable strides have also been made in characterizing human genetic variation. That of course is critical—if we hope to discover the hereditary factors that influence virtually all diseases, we need a complete understanding of the 0.4 percent of the genome that differs between individuals.

As the cost of DNA sequencing has continued to plummet, the potential of determining the complete instruction book of individuals for medical purposes has become increasingly more realistic. Only five years after the completion of the first human genome sequence, a project has been mounted to sequence 1,000 or more human genomes, drawn from individuals from all over the world. This will provide the most detailed view yet of genetic variation.

Many other large-scale research projects are now aimed more directly at determining genome function. The Encyclopedia of DNA Elements (ENCODE) is a project that involves dozens of laboratories working together to identify all the functional elements of the genome (the "parts list") and to determine how those work together to turn genes on or off in particular tissues.

Other projects are studying model organisms, including a project

that aims to inactivate (or knock out) each of the genes in the laboratory mouse. Since more than 95 percent of mouse genes have clear matches to genes in the human genome, this powerful resource will help determine the function of thousands of mouse and human genes, one at a time.

The consequence of all this progress is that a new science has appeared at the very center of biology and medicine: you could call it DNA cryptography. We've intercepted a highly elaborate message of critical importance for the future of the human species. It is written in a strange and seemingly impenetrable code, disarmingly simple in its use of just four letters, but complex enough that decades will be required before a combination of human ingenuity, laboratory investigation, and elaborate analysis on the most powerful supercomputers will reveal the full secrets of the code. But what an amazing adventure this is!

HOW DOES ALL THIS RELATE TO PERSONALIZED MEDICINE?

We are all individuals; in matters of health and disease, we bring our own genomes, our environmental exposures, and our choices to the table. And most of us live and make those choices with plenty of mixed messages and motivations. We know we should exercise regularly, and eat healthy foods, but we don't always manage to do so. We have information about risks and health all around us, but still we sometimes throw caution to the winds. Young people especially live for the moment and worry little about the future. It's older people, starting especially at parenthood, who tend to be more circumspect.

Because you have picked up this book, I assume that you are interested in learning how to improve your chances of staying healthy.

What if I told you that the single most important source of information about your future health, and your risk of illness, and that of your parents and your children, was readily available, provided a window into your genome, was free, and required only about an hour or so to collect? Most of us ignore this powerful tool, even as we take care to fasten our seat belts, avoid potentially dangerous food, and try to make time for exercise. It's our family health history.

Every medical student is taught to record a family history as part of the evaluation of a new patient, but the actual purpose of the history is not always made clear, and all too often the process is rushed and cursory. How many of our medical records contain the utterly useless notation "noncontributory" in the section on family history? (You may never have even seen your records. Trust me; it's probably there.) This is a truly wasted opportunity.

Family health history turns out to be the strongest of all currently measurable risk factors for many common conditions, incorporating as it does information about both heredity and shared environment. Having a parent or sibling with cardiovascular disease doubles your risk. Having two or more of these "first-degree relatives" with heart disease, if they developed the disease before age 55, multiplies your risk fivefold.

Having a first-degree relative with colon, prostate, or breast cancer increases your risk two- to threefold. Similar risks are found if you have close relatives with diabetes, asthma, and osteoporosis. Surely this is the kind of information you and your doctor should know and incorporate into your own health care, in precise detail. Yet all too often the information is not collected, or is simply ignored.

In an attempt to rectify this situation, my colleagues and I joined with the then United States surgeon general Dr. Richard Carmona in starting a family health history initiative in 2004. This Web-based resource (http://familyhistory.hhs.gov) makes it easy for people to col-

lect their own family health histories conveniently, in their own homes. It helps encourage people to call or e-mail relatives to obtain missing information. With the use of a Web tool that is privacy-protected and freely available from the U.S. Department of Health and Human Services, each family health history can be entered into a standard form, which produces the kind of "pedigree" that health care providers need, and can be readily integrated into the electronic health record.

Hundreds of thousands of individuals have taken advantage of this opportunity, and that number is growing daily. (At the end of this chapter you can find detailed instructions about how to carry out this process for yourself and your family, and I strongly urge you to do this.)

It is profoundly unfortunate that our medical care system has largely failed to encourage this kind of data collection. A recent survey by the Centers for Disease Control and Prevention (CDC) suggests that less than 30 percent of Americans have actively collected health information from our relatives, though 96 percent believe this information is important.

Of course, family health history has limitations. Many individuals are stricken by common diseases such as cancer, diabetes, heart attack, or Alzheimer's disease despite an absence of any relevant family history. And, of course, adopted individuals often lack access to this information.

THE NEW PARADIGM

The revolution that now promises to transform our physical *and* mental lives is the opportunity to combine this knowledge of family history with a survey of your entire DNA instruction book, and identify the specific glitches hiding in your life script. And let's be clear

The phone call Doris Goldman received on that morning in 1979 was a mother's worst nightmare. Her 20-year-old son, Jack, a handsome, athletic college student, had been found dead in his sleeping bag in Wyoming. Extensive postmortem examination, including studies of all possible drugs and toxins, revealed absolutely no cause. Two years later Doris's daughter Sharon, just 19 years old, suffered a cardiac arrest. Though she was resuscitated, she had significant brain damage and struggled mightily over the next few years to recover. She went on to attend community college, get married, and have a son. But then the unspeakable nightmare recurred: Sharon was found dead one morning at the age of 29. Again, no cause could be identified.

Some mothers in this situation might have sunk into depression, anger, or blame. Not Doris. She collected medical histories and electrocardiograms (EKGs) from her large extended family. She learned of another cousin who had died at age 45 in her sleep. A review of the EKGs by cardiologists Doris had recruited began to reveal a possible answer. Specifically, a component of the electrical conduction pattern that the EKG detects, known as the QT interval, was prolonged in quite a number of family members.

This condition, known as "long QT syndrome," had been described in a few families in the medical literature and was, in fact, associated with fainting spells and sudden death, as it predisposed its victims to a potentially fatal heart rhythm called ventricular fibrillation. Careful review of EKGs that had been done in the past on Sharon, read as normal at the time, revealed subtle but convincing evidence that she too had this condition.

The differences between a normal QT interval and one that could be dangerous are quite small, and so in Doris's family it was not entirely possible to identify who was at risk by reviewing the EKGs. But in 1996, tools arising from the Human Genome Project made

it possible for specific genes for long QT syndrome to be identified. Doris's family turned out to have a mutation in a gene, *HERG*, which is normally involved in sodium transport across the cardiac muscle cell membrane. With a specific genetic test now available, no fewer than 37 members of Doris's family were found to have this mutation, and to be at risk for the same sudden death as Jack and Sharon. Though she herself had never had a blackout spell, Doris found that she, too, was on the list of mutation carriers, as were her surviving daughter and Sharon's young son.

This sounds like a grim scenario, but it was not a hopeless one. In this instance the value of the information was profound. Research studies have convincingly demonstrated that individuals with mutations linked to long QT syndrome can have their risk greatly reduced by lifelong treatment with a class of cardiac drugs called beta-blockers. Members of Doris's extended pedigree have now been treated, and there have been no further deaths. All of the affected family members also have automatic external defibrillators in their homes (if they can afford this device), and they make sure that family members are trained to perform resuscitation if the need should arise. Some have even had automatic defibrillators implanted in their chest.

Long QT syndrome is not a disorder that most members of the public, or even most health professionals, have heard of. Yet it turns out to be a critically important condition to recognize. The availability of DNA testing has made it clear that as many as one in 4,000 individuals in the U.S. may be at risk. With some families, death seems to occur during sleep. In others, death strikes at times of exertion or strong emotion. Perhaps the most dramatic example of this is the story of a family who lost two sisters on the same day. It was Super Bowl Sunday. One sister was shoveling snow around her home in Virginia. She suddenly fell dead. As word spread through the family, a second

sister, distraught at the news, collapsed and could not be resuscitated despite the efforts of emergency personnel. It turned out that these two sisters, and six of their siblings, carried mutations in one of the genes for long QT syndrome, as did a number of their children. The tragedy on Super Bowl Sunday provided this family with a chance to find out about this potentially devastating condition, and probably saved the lives of many others, even as they continue to grieve for their lost loved ones.

Several lessons can be derived from the dramatic stories of these families. First of all, knowing your family history can save your life. In the midst of tragedy, investigating the causes of unexplained deaths has led to a chance at a full life for relatives who might otherwise have faced a similar fate.

Second, health professionals don't always know the answers. In these families with long QT syndrome, the sudden death of a young person did not immediately set off the right lightbulb in the minds of the physicians. Motivated individuals in families can make all the difference.

Third, DNA testing, although not always as clear-cut as in these cases, can in the proper circumstances provide much-needed answers, and can even provide powerful predictions of risk for other family members.

Fourth, even for long QT syndrome, a highly heritable condition, it is clear that the environment has a strong role, since cardiac arrest is most likely to occur in particular settings. Another critical environmental influence on long QT syndrome comes from over-the-counter or prescription drugs, many of which can increase the likelihood of fainting or sudden death and must be avoided by individuals with this condition.

Fifth, none of us should ever be fatalistic about a serious condi-

tion, even though it may be written into the DNA of every one of our cells. We will not be in a position to alter our own genomes for a very long time to come, but other medical interventions can have profound benefit.

Finally, one more point should be made about this condition and its broader implications for the rest of us. Although, as noted above, only about one in 4,000 individuals is affected by the long QT syndrome in this highly heritable way, studies of hundreds of individuals have shown that there is considerable variability in the length of the QT interval among otherwise normal individuals. Furthermore, those in the upper end of this distribution face about a threefold increased risk of sudden death, even though they do not actually have long QT syndrome. Recently, variations in several genes have been identified as playing a role in "normal" variations in the QT interval. Though measurement of the QT interval or the genes contributing to it in otherwise normal individuals has not yet emerged as part of personalized medicine, this would not be an unreasonable addition to the information to be collected on individuals in the future, especially given the potential for prevention.

Few individuals have heard of long QT syndrome. Fewer still have encountered this diagnosis in a family member, or have undergone a DNA test for the condition. But as we shall see, not all genetic conditions are rare. If you have children or grandchildren who are less than 35 years old, chances are that they have had a genetic test. If you are a woman with children under 30, there is a fair chance you have had one yourself, although you may not have been fully aware of it. In many ways, personalized medicine is already here.

WHAT YOU CAN DO NOW TO JOIN THE PERSONALIZED MEDICINE REVOLUTION

Take advantage of the U.S. surgeon general's Family Health History Initiative and the tool "My Family Health Portrait." Go to http://familyhistory.hhs.gov/ and learn how you can collect medical information from your family to construct a standard medical pedigree. Once you have put this all together, send copies to all family members. Take your own copy to your next visit with your health care provider, and use this as a means to start a conversation about what your own personal risks for future illness might be, and what you can do about them.

grate into this field, instead of going into neuroscience or immunology. But let's try to get a few of these principles clearly spelled out, as a bit of "Genetics 101" will inform all the future information about personalized medicine in this book. (More details are in Appendix B.)

Principle 1: We humans are *diploid*. That means each of us carries two copies of almost all the genes in our instruction book, one inherited from our mother and one from our father. Genes are carried on *chromosomes*, which can actually be seen under a microscope when the cell is about to divide. Figure 2.1 shows human chromosomes from a normal male, arranged in order to demonstrate the pairs. It is apparent that chromosomes come in different sizes and different banding patterns, but they are all paired except for the *X chromosome* and the *Y chromosome* in a male. A female has two X chromosomes instead.

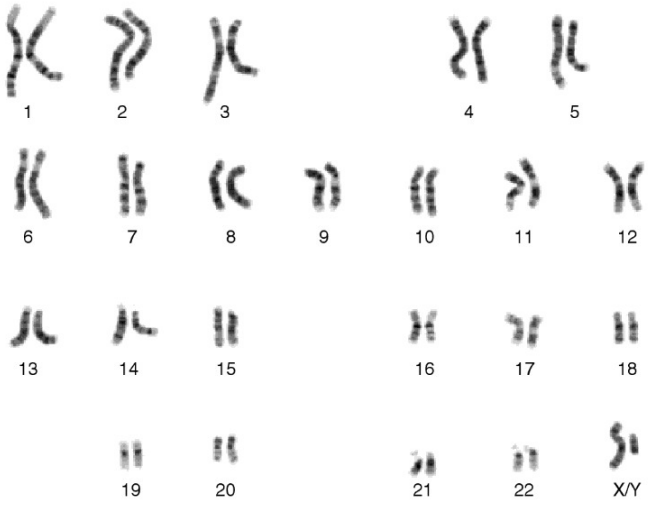


Figure 2.1: The chromosomes of a single cell from a normal human male. A female would have two X chromosomes instead of an X and a Y.

Principle 2: In a recessive disease like cystic fibrosis, *both* copies of the responsible gene must contain misspellings for the disease to occur. As shown in Figure 2.2, that can happen only if each parent carries one misspelled copy and passes it on to the child. The parents in this situation are known as *carriers*, and in the case of a recessive disease they are generally completely normal and unaware of their status. Each child of carrier parents has a one-in-four chance of being affected. I discovered from DNA testing that I am a carrier for alpha-1-antitrypsin deficiency and for hemochromatosis. But neither has affected my own health.

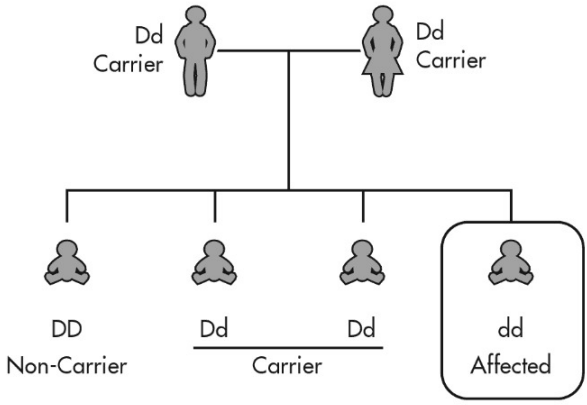


Figure 2.2: Recessive inheritance, as occurs in cystic fibrosis and sickle-cell anemia. "D" is the normal copy of the gene, "d" is the abnormal copy.

Principle 3: In dominant inheritance, an affected individual has one normal copy and one misspelled copy of the gene, and that is sufficient to cause the disease to appear. As shown in Figure 2.3, with this kind of inheritance a disease often appears in subsequent generations, as the child of an affected individual has a 50 percent chance of inheriting the misspelled gene, and also being affected. Well-known examples of dominant genetic diseases include Huntington's disease and neuro-

fibromatosis (sometimes erroneously referred to as "Elephant Man disease"—the Elephant Man actually had a different condition). Another example of a dominant condition is long QT syndrome, referred to in Chapter 1.

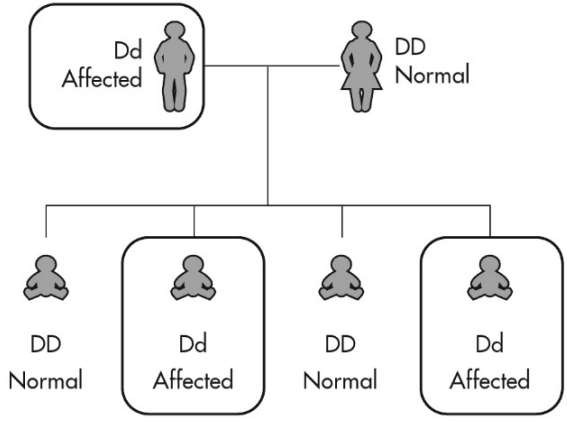


Figure 2.3: Dominant inheritance, as occurs in Huntington's disease.

Principle 4: The inheritance of most genetic conditions is not this simple. Most misspelled genes cause a *predisposition* to, but not a *predetermination* of, disease. Geneticists sometimes refer to this as "incomplete penetrance." Simply put, that means someone who carries a particular gene capable of conferring a risk of disease does not always experience the consequences. The *BRCA1* gene in my own family is an example of incomplete penetrance. Specifically, women who carry a *BRCA1* mutation have about an 80 percent lifetime risk of developing breast cancer and a 50 percent risk of developing ovarian cancer. But this means that some women with these mutations never develop cancer at all. The penetrance is even lower in males, who despite *BRCA1* mutations face only a modest risk of cancers of the pancreas, prostate, and male breast.

Principle 5: Although virtually all common diseases—such as diabetes, heart disease, and cancer—have hereditary components, there are multiple genetic risk factors that contribute to these conditions. We call these diseases *polygenic*. The power of each individual genetic risk factor is generally quite low, and so illness is likely to occur only with a combination of several of them, along with appropriate environmental stimuli. I reported on some of my own genetic risk factors in the introduction, and I will have much more to say about this situation in subsequent chapters.

But now let's return to Anabel and Isabel, and focus on cystic fibrosis as a cardinal example of a disorder caused by mutations in a single gene that is yielding up many of its secrets as a consequence of the genome revolution. In 1972, when Anabel and Isabel were born, not much was known about cystic fibrosis, other than its recessive inheritance and the fact that it affected numerous organ systems of the body. It was known to involve the pancreas (the name "cystic fibrosis" refers to the formation of cysts and fibrous scars in the pancreas of affected individuals), resulting in an inability to secrete digestive enzymes. If these enzymes were not added to the person's diet, profound malnutrition would result. The intestines were known to be involved in some cases—as with Anabel, who required emergency surgery shortly after birth for a serious blockage. Males with CF who lived to adulthood were also noted to be infertile. Most significantly, however, the lungs were known to be seriously affected. Thick, sticky secretions accumulated, followed by recurrent infections, destruction of lung tissue, and all too often an early death.

Years ago, mothers of children with CF noted that when the children were kissed, their skin tasted salty. As a result, a somewhat bi-

zarre diagnostic test was developed for cystic fibrosis: the measurement of chloride levels in sweat. Salty sweat suggested that there might be some problem in the transport of salt and water, and that this same problem might perhaps affect the lungs, the intestines, and the pancreatic ducts, but it was not until the 1980s that a definite connection was demonstrated. Even then the information was insufficient to identify the responsible gene.

My laboratory played a central role in the identification of the genetic mutation in CF. But it took many long years of torturous work, because of the lack of information about the human genome at that time. Families with several affected children were asked to participate in research, in order to try to map the gene to a particular location in the genome. The principle of the method was simple. Since this is a recessive disease, affected siblings must share identical DNA containing the CF gene on both their maternal and their paternal chromosomes, whereas elsewhere in the genome they can be expected to share only 50 percent of their DNA. From studying a very large number of such families, it ultimately became clear that a long stretch of DNA on chromosome 7 must contain the CF gene. But the remaining task was daunting: the region of DNA involved was approximately 2 million base pairs, and the methods for dealing with such large stretches of DNA in the 1980s were very slow and imperfect.

Teaming up in 1987 with an investigator from Toronto, Dr. Lap-Chee Tsui, my lab trolled through this large, uncharted DNA territory, searching for any subtle mutation that might distinguish CF patients from unaffected individuals. After many false starts, and many occasions when the hopes of our research teams were dashed by the next day's data, the answer finally emerged. I can recall the exact moment when we were sure we had it. Lap-Chee and I were attending a meeting at Yale University, and had set up a fax machine in

discovered (about half of CF patients) can differ widely in the severity of their lung disease. Why should that be?

One contributor to this variation turns out to be other genes in the genome that serve as "modifiers." Most genes have some degree of normal variation. Those normal variations in other pathways can play a role in affecting the severity of a genetic disease like CF. Already several such modifiers have been identified for CF.

Another important modifier of severity is the environment. A recent study demonstrated that exposure to secondhand smoke can play a significant role in the progression of CF lung disease. And of course the dramatic improvement in survival in CF (Figure 2.5) cannot be due to changes in the gene pool in a period of just 50 years. In this case, the development of better medical interventions has provided beneficial environmental influences on the disease. These include the availability of pancreatic enzyme capsules to improve nutrition, the use of vigorous chest physical therapy to clear the sticky secretions that otherwise lead to lung infections, the use of aggressive antibiotics to keep infections at bay, the development of an aerosolized enzyme therapy that can digest the sticky DNA in the lung secretions and make them easier to clear, and the use of saltwater mists to assist in keeping the airways clean.

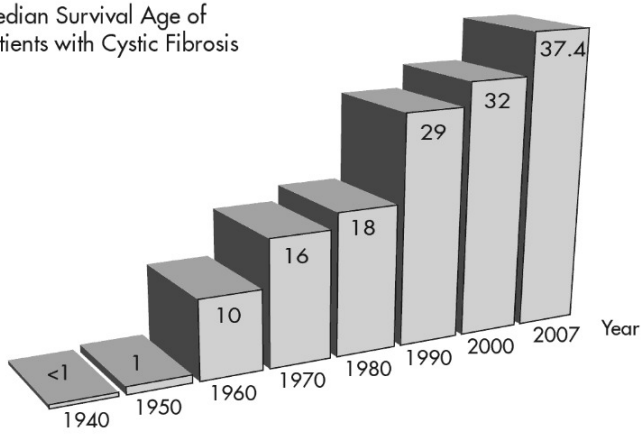


Figure 2.5: Medical research has led to dramatic improvement in survival for individuals with cystic fibrosis.

Most dramatically, when all these efforts fail, double lung transplantation has been lifesaving for hundreds of CF patients such as Anabel and Isabel, though the availability of organs continues to be a major challenge, and the risks of rejection are substantial. In fact, Anabel has already gone through an episode of rejection and has undergone a second lung transplant.

FROM GENE TO CURE?

It is one thing to learn the specific mutation behind a disease. It is quite another to overcome it. Many people scoff that therapy for rare genetic diseases will never be practical. Consider sickle-cell anemia. This was the first recessive genetic disease identified. It is common among individuals whose ancestors lived in areas where malaria has historically been widespread: around the Mediterranean, in Africa, and in Southeast Asia. Sickle-cell carriers—those with one, not two, copies of the sickle mutation—are better able to survive childhood malaria than those who do not have this mutation. Those individuals carrying two copies of the sickle mutation, however, have a serious blood disorder associated with frequent painful crises and a serious limitation in life span.

The genetic mutation in sickle-cell disease, located in one of the genes for hemoglobin, has been known for 50 years, yet this information has led to relatively little in the way of novel therapeutic approaches. So why should I claim that genetic medicine is at a tipping point now? For one thing, the pace of medical progress is not linear. Fifty years of slow progress does not imply that the next 50 (or even the next 10) will be similarly slow. Most researchers actually predict substantial advances in the treatment of sickle-cell anemia in the com-