1 Your Personal Genome: Googling Your DNA

Not many homeowners can boast having a garage that changed the world. But Susan Wojcicki can. She can look back at her decision in 1998 to rent out her garage at 232 Santa Margarita Avenue in Menlo Park, California, as a world-changing event. Her renters, two graduate students in computer science at nearby Stanford University, needed space to develop a new company around their revolutionary approach to searching the web. Sergey Brin and Larry Page would not occupy her garage for long; they soon needed more spacious headquarters for the company that would launch their combined net worth into the stratospheric level occupied by the likes of Bill Gates and Warren Buffett.

But before they launched Google, Brin and Page devised something they called PageRank, a method to rank a webpage according to how many other pages have links to it, and the number of links each of those other pages in turn has to yet other pages. PageRank provided a much more effective means of searching the Web than that offered by then-available search engines such as AltaVista or Excite. Capitalizing on its vast ability to organize and serve up information on the Web and provide a host of other services—Google Earth, Google Image, Google News, Google Maps, Google Groups, Google Books, and more—Google quickly went beyond PageRank as its instrument for dominating the lucrative search market.

Susan Wojcicki must have realized that her young tenants were on to something: she became one of Google's first employees, going on to develop its online advertising business. But in addition to connecting people to information, Wojcicki also introduced Sergey Brin to her younger sister Anne. That introduction culminated in the May 2007 wedding of Brin and Anne Wojcicki, then both thirty-three, on a private island in the

Bahamas. The wedding couple wore bathing suits, and some of the guests joined them in swimming to the ceremony.

Like her husband, Anne Wojcicki is one half of a business duo that launched a start-up company. But the company founded by Wojcicki and her partner Linda Avey—23andMe—does not deal in Internet search results, digital images, or city maps. Instead, it serves up genetic information. Despite Google's investment of $3.9 million in 23andMe, this start-up, and other similar ventures such as Navigenics, deCODEme, Knome, and Sciona, leverage no recent insight from computer science. They rely on a principle known for over seventy years: the more closely related two individuals are to each other, the more similar are their personal DNA codes. This is the principle that allows geneticists to figure out whether a particular version of a gene makes an individual who carries it more prone to a disease such as diabetes, Parkinson's disease, or cancer than does another version. This is also the guiding principle of our book, as we lead you to an understanding of what DNA is, how its four-letter alphabet spells out the genes that determine the traits of human beings with all our complexity, and how the way those four letters line up six billion times in our personal DNA code influences how we look, how we behave, how we get sick, and how we respond to treatment.

Named after the 23 pairs of chromosomes we carry in every one of our cells, 23andMe has as its mission "to take the genetic revolution to a new level by offering a secure, web-based service where individuals can explore, share and better understand their own genetic information." It can't provide an individual with her complete DNA code, but for $999 it can reveal enough of it (about 1/6000th) to give her a glimpse of her risk for a few diseases. The web-based service the company provides allows you, in effect, to Google your own DNA: instead off typing words into a search engine to scan the world's web pages, you type the DNA letters of a gene to scan the world of your own DNA, searching for the relevant text stored therein.

Today their service is at the cutting edge; in a few years it will seem primitive. This mere glimpse of a customer's DNA will soon be replaced by a complete reading of how the four letters of its alphabet are used 6 billion times to spell out his DNA code. So far, the DNA code of only a few people has been read, at a cost of over $1 million each, prohibitive for all but the wealthiest. But new technologies are rapidly reducing the cost, bringing

this brave new world into view. An added bonus to the acclaim and profits that will accrue to those who solve the issues of cost and speed is the Archon X Prize of $10 million, which will be awarded to the first team that reads the DNA code of one hundred people in ten days at a cost of ten thousand dollars each. It is possible, maybe even likely, that that reward will have been claimed by the time you read these words.

While a reading of our complete DNA code may be still out of reach for you and for us, it was realized recently by James D. Watson. More than half a century ago, on February 28, 1953, Watson and his colleague, Francis Crick, launched an age of genetic discovery with their announcement to the lunchtime patrons of the Eagle Pub in Cambridge, England, that they had "found the secret of life." Their discovery of the structure of DNA—the most important molecule of life, which specifies the form and function of every living thing—made clear how traits are passed down through the generations. Watson and Crick's breakthrough paved the way for an age of discovery that culminated in the announcement on June 25, 2000—not in a pub but at the White House—that the human DNA code had been determined. A few years later Watson himself became one of the first two people to read his own personal DNA code.

After 1953, Watson went on to a celebrated career, directing a laboratory at Harvard University, then a storied scientific institution at Cold Spring Harbor on Long Island, and ultimately the Human Genome Project, which deciphered the human DNA code. But shortly before his own DNA code was determined, Watson's professional life ended amid charges of racism. He was quoted in the October 14, 2007 edition of The Sunday Times that he was "inherently gloomy about the prospect of Africa" because "all our social policies are based on the fact that their intelligence is the same as ours—whereas all the testing says not really." He also said, "There is no firm reason to anticipate that the intellectual capacities of peoples geographically separated in their evolution should prove to have evolved identically. Our wanting to reserve equal powers of reason as some universal heritage of humanity will not be enough to make it so."

As his comments rapidly circled the globe, drawing condemnation from his fellow scientists, the 1962 Nobel Laureate quickly apologized for them. Speaking at a meeting of the Royal Society in London on October 18, 2007, he said, "To all those who have drawn the inference from my words that

Africa, as a continent, is somehow genetically inferior, I can only apologize unreservedly. . . . That is not what I meant. More importantly, there is no scientific basis for such a belief."

But the damage was done, and so was Watson's job. The board of trustees of Cold Spring Harbor Laboratory relieved Watson of his position as chancellor. They wrote, "The comments attributed to Dr. James Watson that first appeared in . . . *The Sunday Times* U.K. are his own personal statements and in no way reflect the mission, goals, or principles of Cold Spring Harbor Laboratory's Board, administration or faculty. . . . The Board of Trustees, administration and faculty vehemently disagree with these statements and are bewildered and saddened if he indeed made such comments."

Watson had steered into the always-dangerous shoals of the genetics of race, and he should not have been surprised that his words sank him. In our penultimate chapter we, too, venture into these treacherous waters. We will show you that there are many more genetic differences *within* racially defined populations such as Africans and Caucasians than *between* these populations. You can see the close resemblance of the DNA codes of these races if you compare the few available sequences. Or, you can wait a few years and see it when you read your entire DNA code.

Just as Google's computers read a digital code composed of 1s and 0s, living creatures read a chemical code of four different units, abbreviated as A, C, G and T. We'll see how strings of these four chemicals get decoded in the production of proteins, the workhorses of the body that enable us to move, see, breathe, think, and reproduce. These four chemical units are strung together 6 billion times (6 followed by 9 zeros, or 6 thousand thousand thousand)—but this number is infinitesimal compared to a "googol"—1 followed by a hundred zeros, or ten thousand trillion trillion trillion trillion trillion trillion trillion trillion. Yet even a googol is barely a speck in comparison to a "googolplex," which is 10 raised to the power of one googol, or 1 followed by 10^{100} zeros. (It would take much more space than the pages in this book to write that number.)

The algorithm Larry Page developed to search the Web originally went by the unbusinesslike name of BackRub. But in late 1997, as he and Sergey Brin contemplated starting a company to exploit their search engine, BackRub had to go in favor of a more fashionable term that would connote

the vastness of what they were trying to organize. Unfortunately, the names they first came up with had already been claimed by other people. Page's officemate Sean Anderson made a number of suggestions, but Page nixed all of them. Anderson eventually offered "Googolplex," a name that suggested the vast amount of information the new search engine could scan. Page liked it, but preferred the shorter "Googol." Computational brilliance they may have possessed, but world-class spelling was not their forte. When Anderson used the new search engine to see if the name was available, he typed in "Google" and found that it was unclaimed. That evening Page registered the domain name Google.com. Only the next day did they learn they had misspelled the term, and discovered that the domain name "Googol.com" had in fact been claimed.

As Google rapidly expanded, Brin and Page focused on maintaining its spirit of adventure and cohesiveness. Employee number 56, who arrived in November 1999, was Charlie Ayers, their executive chef. Ayers provided free, wholesome food to the young Google workforce, maintaining their energy for the ambitious tasks they were tackling. He later recalled to David A. Vise and Mark Malseed, authors of *The Google Story*, "I could feel the energy. They had it. Everyone was so focused and into it, and they all had one goal: to make this company successful. It was 'Look at what we did,' not 'Look at me.'"

An equivalent organizational spirit exists within every one of the trillions of cells in your body. DNA provides the corporate vision and hiring plan, but it's the roughly twenty thousand varieties of proteins that carry out all the necessary activities of the cell. Like Google employees, proteins engage in a team effort that is much greater than the sum of their parts. You'll see as you read on how proteins read the DNA code in the single cell that is the fertilized egg and tell it to divide into two, then four, then eight and so on. Successive generations of cells take on new functions, specializing as heart and blood, brain and nerves, bone and teeth and all the other tissues of the body. The result, a living human being, is more magnificent than any company, no matter how much revenue it generates.

Sergey Brin was born in the Soviet Union in 1973 to two mathematicians. His father, Michael, is now a professor at the University of Maryland, and his mother, Eugenia, works at NASA's Goddard Space Flight Center, in

Washington, D.C. Anti-Semitism in the Soviet Union in the 1970s prevented his parents from advancing very far in their academic careers, so they decided to apply for exit visas, even though doing that meant running the risk of becoming unemployed and ostracized by colleagues. They were fortunate to be some of the last Jews allowed to leave the Soviet Union before it broke up a decade later. Michael and Eugenia Brin left with their young son in 1979, and settled into a new life in Maryland.

Larry Page, born just a few months earlier than Brin, is the son of the late Carl Victor Page, a professor of computer science at Michigan State University who was one of the first to receive a doctorate in this field of study. His mother, Gloria, earned a master's degree in computer science and taught programming at Michigan State. Page benefited from a rich exposure to computers long before most Americans had ever seen one.

Brin and Page seem to have a knack for science and technology, like their parents. They are bright, inquisitive, and creative, like their parents. They have a risk-taking, adventuresome attitude, like their parents. Humans have recognized for thousands of years that offspring resemble their parents, knowledge that they applied early in the course of human civilization to the selective breeding of plants and animals. Yet if you could compare the strings of DNA units in Brin's or Page's personal DNA code to the strings in your own DNA code, you would find that they are 99.9 percent identical. If you lined up your DNA code with that of Tiger Woods, or Madonna, or Barack Obama, George W. Bush, Hillary Rodham Clinton, or Paris Hilton—you would find that these, too, are 99.9 percent identical. Since all these people look and act differently from each other and from you, that 0.1 percent difference between any of these people must play a big role in personal appearance and behavior. And as we'll discuss in subsequent chapters, that 0.1 percent difference also results in some of us getting cancer, or Alzheimer's disease, or having the good fortune to escape these diseases altogether.

We expect that within the next ten years the parents of a newborn will be presented with the code that is written in the strings of letters of her DNA, in addition to their child's footprints and thumbprints and APGAR score. They will be able to Google that code and predict their child's risk for some diseases and behaviors. They will know what to watch for, and in some cases they will be able to intervene to minimize unwanted consequences and maximize desirable outcomes. The child born ten years from

now will have unprecedented self-awareness (genetically speaking), and unheard of self-control (medically speaking).

Clearly, we are on the cusp of a genetic revolution, one that will affect all of us in a personal way, every day. To benefit from this revolution and manage the consequences, everyone needs to understand what is meant by "personal DNA code" and how the small fraction of differences between individuals' long lists of DNA letters makes each of us unique. How does that code determine who we are, what diseases we may get, how we feel, and what we're capable of? How much control does our personal DNA code exert over our fate?

In this book we provide some answers to the questions: What is a "personal DNA code" and what will it tell us about ourselves? How do the genes specified in the human DNA code get assigned to specific cellular functions? How do individuals' differences in this code result in differences in disease risk? The answers require us first to explain what DNA and genes are, how they are inherited, and how they collaborate to allow a fertilized egg to turn into a creature of amazing complexity in nine short months. From there, we can begin to show you why each person's DNA code is a little bit different from everyone else's, and how these variations influence our lives.

2 Genes Are the Instructions for Life: AIDS and the Uncommon Man

Just one small change in one gene might have given the world more books like *Pebble in the Sky*, *The Stars Like Dust*, *The Foundation Trilogy*, and *I, Robot*. To science fiction enthusiasts of a certain age, the publication of a new Isaac Asimov novel or short story was cause for celebration. Before iPods and instant messaging, before YouTube and Facebook, before Xboxes and PlayStations, young fans would curl up under the bedcovers with one of Asimov's intergalactic tales and read late into the night. For a period in the 1960s and 1970s, Asimov's large black glasses and mutton-chop sideburns made him one of the world's most recognizable authors. His books sold in the millions. Millions more might have flown off booksellers' shelves had Asimov inherited a personal DNA code with one small change.

Asimov penned more than just science fiction; he wrote on almost any topic. He explained mathematics and astronomy, chemistry and biology, as well as the Bible and Shakespeare, American history and the Roman empire, along with Gilbert and Sullivan and *Paradise Lost*, and limericks, and Egyptian history. Asimov wrote almost nonstop, averaging about a thousand words a day, every day, for fifty years. "Being a prolific writer has its disadvantages, of course," Asimov commented in one of his three autobiographies. "It complicates the writer's social and family life, for a prolific writer has to be self-absorbed . . . and has not time for anything else." This obsession for writing—and the accompanying unwillingness to do much else—had unwelcome consequences, including the breakup of his first marriage. And "a prolific writer . . . has to *love* his own writing," Asimov noted. He certainly loved writing: he wrote over four hundred books!

Asimov's many popular nonfiction works include *The Chemicals of Life* (1954), and *The Genetic Code* (1963). In the latter work he attempted to

explain to a lay audience what a gene was at a time when even the research biologists working on genetics barely understood what it was.

What is a gene? Surely those research biologists must have figured it out in the forty-seven years that have intervened since Asimov weighed in on the subject. And the answer should be common knowledge to a society accustomed to headlines proclaiming "Scientists identify gene for schizophrenia," or for obesity, or for colon cancer, or for any number of other diseases and conditions.

Apparently not. None of the people we queried who weren't biologists came close to providing a definition of the gene that would pass muster in a high school biology classroom. Many associated the gene with physical traits or emotional characteristics, a reflection of the realization that genes come from our parents and grandparents, and the notion that they explain Johnny's big ears and Mary's quick temper. The gene was occasionally linked to concepts such as DNA or chromosomes, although those terms were fairly fuzzy to those more than two years out of tenth grade biology.

Most commonly, our requests to explain what a gene is were met with the same look that might surface in response to a question about what the World Bank does, or why tort reform is needed, or how tornados start. But unlike those technical details, which we can safely leave to economists, lawyers, and meteorologists, genes are too important to be relegated to the safely obscure. We fail to understand them at our peril, because they influence crucial aspects of our daily life: our health, our lifespan, our mood, our insurability, and our food supply, to name a few.

But before we deal with *what* genes are, let's ask an even simpler question: *where* in your body are genes found? This was the query posed to Americans of varied ethnic and educational backgrounds by Angela D. Lanie, a scientist at the University of Michigan. Nearly a quarter of the respondents said genes are "in the brain"; about one eighth said "in the blood"; a few said in the reproductive system, or heart, or bones, or lymph nodes or various other locations.

Only about one third of Lanie's respondents gave the correct answer: genes are present all over the body—everywhere, in virtually every cell. Genes are found in your skin and your stomach cells, your lung and your liver cells, your brain and your bone cells. They are in every part of your body. And not just in your body: they are in every cell of broccoli and beets, and chicken and cows, and apples and apricots. If you stopped eating

food that contains genes, you'd have made the unpalatable (and unsustainable) choice to dine on not much more than sugar and water.

An understanding of genes is within easy reach for all of us, because the principles that govern the operation of genes are simple. What is a gene? It is a stretch of DNA that contains the instructions for the cell to manufacture a protein. Once the terms "DNA" and "protein" and "cell" have been explained, this definition is surprisingly satisfying to most nonscientists (and to most scientists, too, for that matter).

Let's start with the cell. Imagine your body as an enormous hotel composed of about 100 trillion (1 followed by fourteen zeros) rooms, each a self-contained space enclosed by a set of walls within which sits a bed, dresser, night table, and other furniture. Each cell in your body is a similarly self-contained unit measuring about a fortieth of a millimeter across (about a tenth the width of a human hair), surrounded by a flexible membrane that protects it from the environment. There are compartments in the cell where specific functions are carried out, including maintaining the cell's DNA, burning fuel to provide energy, and transporting material to where it needs to be.

Each room in a hotel has plumbing that connects it to a central water supply, a source of electricity to power its appliances, and heating and cooling units to control its temperature. A central processing system housed on the top floor of the hotel—a phone switchboard and a computer with an Internet connection—allows every room to be in contact with the front desk, and with all the other rooms in the hotel—indeed, with the rest of the world. Likewise, each cell connects to and communicates with adjacent cells and with the rest of the body using chemical and electrical signals.

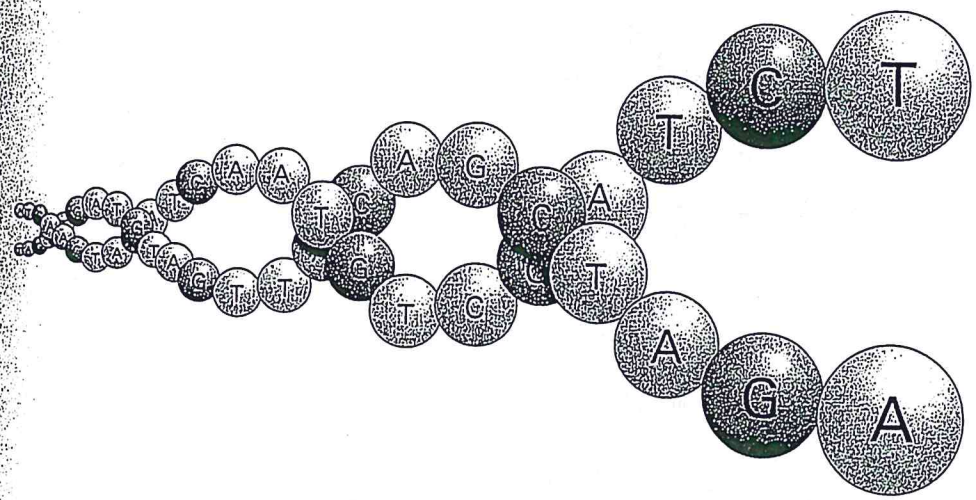
From a distance the many rooms of the hotel look the same, but if we look more closely we see that some have one bed, others two. Some have separate kitchenettes, or sitting areas, or extra bedrooms. Some are so tiny that they barely accommodate one guest; others are suites that accommodate a large family. Most rooms are rectangular, but some special rooms have unusual shapes (like the octagonal honeymoon suite). Cells, too, are specialized to carry out particular functions, differing in their size and shape and internal structures. There are blood cells and brain cells, lung cells and liver cells—cells with long and narrow projections, cells with unusual talents to filter blood or detoxify alcohol.

Rooms are not arranged willy-nilly in a hotel but are arrayed in orderly wings of multiple floors. Cells are also arrayed in the body in an orderly fashion; they organize themselves into successively larger units of tissues, organs, and organ systems. Most of these systems are familiar to us. The collection of cells that make up the mouth, esophagus, stomach, small and large intestine, the liver, pancreas, and gall bladder constitutes the digestive system, which processes food. Cells of the heart, arteries, veins, and blood form the circulatory system, which delivers nutrients and oxygen to the far reaches of the body. Cells that make up the bladder and the colon cooperate to manage a storage system that holds waste until it is ready for disposal. Cells of the brain and spinal cord constitute the nervous system that manages it all.

Cells entrust the instructions for their construction and operation to DNA, a chemical found in all living things. How did we come to know that DNA serves this vital function? In 1944, Oswald T. Avery, a physician and scientist working at the Rockefeller Institute for Medical Research (now The Rockefeller University) and his coworkers Colin MacLeod and Maclyn McCarty reported that they could dramatically change the properties of a cell—in their case a cell of the bacterium that causes pneumonia—by changing only its DNA. They concluded that DNA was the long-sought substance of heredity.

Attributing such importance to DNA was a startling result, because DNA was known to be a molecule consisting of a seemingly endless, monotonous string of only a few very similar subunits. How could such a "stupid molecule" (as some then called it) determine what kind of covering enclosed a bacterial cell (the trait that Avery and his colleagues analyzed), much less perform the amazing feat in more complex creatures of specifying the appearance of limbs and lungs and livers in all the right places and of the right size, and the proper number of teeth and toes, and irises and corneas and retinas that form eyes, and much, much more? Surely, thought many biologists, a more complex molecule was needed to accomplish those amazing feats.

It's easy to see why they thought this way, because DNA is indeed a simple molecule. It is composed of only five atoms: carbon, hydrogen, oxygen, phosphorus, and nitrogen—the organic elements from which all living things are built. DNA is a polymer—a long molecule made up of



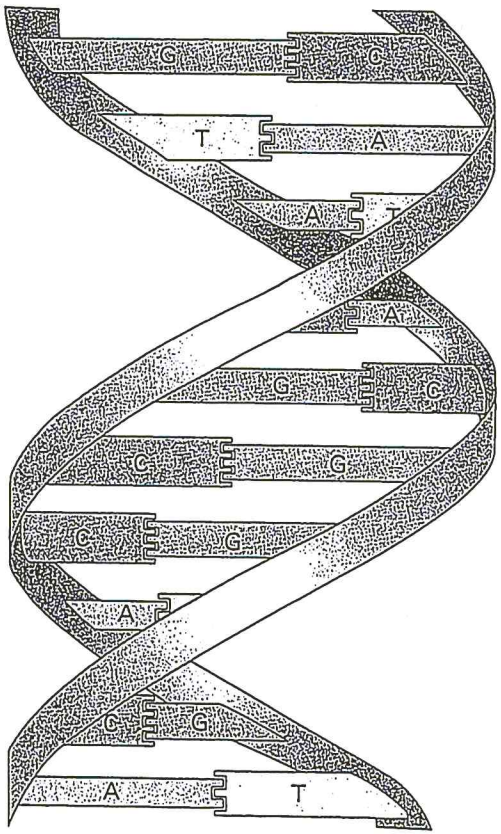
small units linked together, one after the other, like pearls in a necklace. These smaller units are known as "bases," and they come in only four types, commonly called by the first letters of their names: A, C, G and T (adenine, cytosine, guanine, and thymine).

These four bases link one to another to form a very long string—think of an extremely long necklace made up of four different kinds of pearls. Two of these strings of bases wrap around each other—think of a double pearl necklace—to form the iconic double helix structure with two strands of bases spiraling around each other (see figure).

A chromosome is a long, unbroken string of the DNA double helix (along with some packaging material to wrap up the DNA molecule so it fits inside the cell). Each human chromosome is a double string of around 100 million DNA bases. In some creatures the chromosomes may be less than one hundredth that size; in others they can be up to ten times longer.

But DNA turns out to be not so stupid a molecule, because the order of the bases—the exact sequence of the A's, C's, G's, and T's—is the information that specifies the characteristics of an organism. Of all organisms. Of us.

Now comes the most important fact: the sequence of A, C, G, and T bases needs to be specified for only one of the two strands of a DNA molecule because the sequence of bases of one of the DNA strands specifies the sequence of bases of the other. This is a result of the way the two strands of the double-helical structure are held together by interactions between the bases: bases in one strand attract and stick to bases of the other.



But the bases don't interact haphazardly: A and T stick to each other but not to G or C; likewise, G and C stick to each other but not to A or T. Think of 110-volt electrical plugs with two prongs and 220-volt plugs with three prongs. The bases A and T are like a matched pair of a two-pronged plug and a two-prong socket; G and C are matched like a three-pronged plug and a three-prong socket (see figure). (The "plugs" and "sockets" of the double helix are really a kind of chemical bond—a quite weak one—called a "hydrogen bond"; G and C have three of them, A and T have two.) As a consequence of these specific matches, each strand of the helix carries the information to specify its partner strand's type.

When a cell divides to form two cells, it copies the two strands by peeling them apart—unplugging the plugs from their sockets—and then uses each strand as a template on which a new strand is synthesized. The result is two identical copies of the chromosome, because every A attracts a T in the

newly made strand, every T attracts an A, every C attracts a G, every G attracts a C. These A-T and C-G combinations are the "base-pairs" of DNA.

On Saturday morning, February 28, 1953, in Cambridge, England, James Watson and Francis Crick realized how the sequence of bases of one strand of the DNA double helix specified the sequence of bases of the other strand. They saw for the first time how the plugs and sockets fit together. When they went to lunch that day at their favorite haunt—the Eagle Pub, not far from their lab—they left the other patrons dumbfounded with their announcement that they had learned the secret of life.

Indeed, Watson and Crick's revelation of the double helix as the structure of DNA, and the realization that the sequence of bases on one strand specifies the sequence of bases on the other, is one of the most important scientific discoveries ever made. Immediately, and very clearly, it explained a major mystery: How does one cell give rise to two identical cells? By revealing the double-helical structure of DNA, Watson and Crick answered a question that had confounded people since before the time of Aristotle: How do organisms replicate themselves? They do it by using one strand of the DNA double helix to specify the sequence of bases of the other strand, producing two identical DNA molecules from one. For their discovery, Watson and Crick were awarded the Nobel Prize in 1962.

With that discovery Watson and Crick ushered in the age of molecular biology, which provided a detailed understanding of the nature of the gene. It culminated in the Human Genome Project, an international effort to determine the sequence of base-pairs in the DNA of every one of our chromosomes. That goal was achieved on the fiftieth anniversary of Watson and Crick's discovery.

Here is a portion of the sequence of bases of one strand of a human chromosome:

CCTTCCGTTAAACCAATGGGAAACAAGTCCCTGGGAGTGTCCCGCCCT
GGGGTGAGAACTGCGAACCAATAAAAATTGAAACCTGAGCGGTGGCGC
GGCCAGCTGTGGGTGGAGTCACCCCGCGGACTGGACGGGAACCTGG
CGGGGTCAGGTCCCGTCAAGCAGCCTGGCTCATGGCTGTGTGCGGCC
TGGGGAGCCGTCTTGGCCTGGGGAGCCGTCTTGGCCTGCGCGGGTGC
TTCGGCGCCGCCAGGCTCCTGTATCCCCGTTTCCAGAGCCGCGGCCC
TCAGGGCGTGGAAGACG

This string of three hundred bases is but a tiny fraction of the sequence of one strand of human chromosome 12, which, at over 132 million

base-pairs (actually, 132,349,534 base-pairs), is a medium-sized chromosome. The letters of chromosome 12, if written in the size of the letters of this book, would fill fifty-two thousand pages.

Each human cell carries twenty-three pairs of chromosomes, forty-six chromosomes in all: one set of twenty-three came from Mom; the other set of twenty-three came from Dad. Altogether, we have about 6 billion base-pairs of DNA, 3 billion in each of the two sets of chromosomes, in every one of the approximately 100 trillion cells in our bodies. It's a prodigious amount of DNA: laid out end to end, the DNA from one human would reach to the sun and back more than sixty times. These three billion base-pairs constitute the genetic material that is the human genome, the material that directs the production and maintenance of each of us. But don't let these large numbers intimidate you. Remember, as complex as the complete human genome is, and with biologists still at the earliest stages of deciphering the instructions embedded in its sequence of base-pairs, DNA is just a long chemical, and a simple one at that: just strings of A's, C's, G's, and T's.

Asimov may never have taken time out of his writing schedule to exercise or take vacations, but he would always make time for a good meal. At the age of fifty-seven, and perhaps as the direct result of consuming a giant slice of cheesecake, Asimov suffered a heart attack that hospitalized him for three weeks. He continued to experience angina over the next several years, the pain becoming so severe that even walking became a chore. In November 1983, his doctor advised a triple-bypass operation. Given the choice of waiting until after Christmas or having the operation right away, Asimov chose right away. But he worried that might prevent him from attending the annual banquet of the Baker Street Irregulars, his fellow Sherlock Holmes aficionados, which was to be held on January 6. He had prepared a song for the banquet, and although he expected to be there to sing it, he prepared a taped version that he gave to his wife, just in case.

The evening before his operation, Asimov dreamed that he died on the operating table and that consequently his wife had to play the tape for the Baker Street Irregulars, who stood in tears and applauded for, lo, twenty minutes. But Asimov survived, "and my first thought was that now I wouldn't get the kind of applause I would have gotten if I had been dead. 'Oh—[expletive deleted],' I said in disappointment."

Although Asimov's operation was a success, the blood transfusion that he received was contaminated with the human immunodeficiency virus (HIV), because blood was not then routinely tested for its presence. After suffering numerous medical problems in the years after his surgery, Asimov learned in 1990 that he had AIDS. He died in April 1992 from heart and kidney complications, the true cause of death not being revealed until ten years later when his wife published *It's Been a Good Life,* composed of excerpts from his three autobiographies.

Since you're reading a book about genes, and in particular their role in disease, you may well be wondering why the first disease we mention is AIDS. Surely AIDS, which ranks among the most virulent *infectious* diseases that humankind has faced, is not *genetic* in origin, you may be thinking. AIDS is spread by sexual contact, blood transfusions, contaminated needles, and passage of a fetus through the birth canal of an infected mother. But rare is the disease that escapes the influence of our genes. So we can tell you the following quite confidently: If Isaac Asimov had had a mutation in both copies of his *CCR5* gene—a mutation that resulted in the removal of thirty-two base-pairs of DNA—he would not have contracted AIDS. This gene, identified in the 1990s, specifies a protein that sits on the surface of cells of the immune system, looking for a signal that invaders have breached the lines of defense. The HIV virus uses the CCR5 protein as a landing pad, alighting on it before invading the cell. If Asimov had lacked those 32 base-pairs in his *CCR5* gene his immune cells would not have had the HIV landing pad, causing them to be resistant to the virus. Unfortunately, even though the prevalence of this mutation is higher in the Ashkenazi Jewish population to which he belonged than in most other populations, Asimov was not so lucky. As a consequence, the world got many fewer Asimov books than it might have.

How do we find the gene responsible for a trait such as resistance to the AIDS virus? How does a gene specify a protein? What do proteins do? What does it mean to have a mutation in a gene, and why does the prevalence of different mutations vary in populations? Read on, and you'll see that these questions have straightforward answers.

3 Proteins Are the Workhorses of the Cell: Misdiagnosis of a Metabolic Malady

Patricia Stallings had had a tough life. She had spent several years on the skids. Homeless much of the time, she found it difficult to take care of herself, let alone the son she had borne out of wedlock. When accused of child abuse for not adequately caring for the child, she gave him up for adoption.

But by the summer of 1989 Patty's life had turned around. She found a good man in David Stallings, and their marriage gave her the kind of life she could only dream about a few years earlier. With their move into a trim, white frame house in a subdivision overlooking Lake Wauwanoka, not far from St. Louis, the Stallings family—Patty, David, and their newborn son, Ryan, born in April—joined the middle class. "That truly was the happiest time of my life," she later reflected. "Everything was perfect. Everything. A new house, a new baby. I mean, what could be wrong?"

Plenty, Patty would soon discover. One Friday evening early in July 1989, three-month-old Ryan threw up his evening meal. He seemed better the next day, but Sunday morning he again could not keep food in his stomach. When he turned lethargic and his breathing became labored, Patty called St. Louis Children's Hospital and arranged to bring Ryan there. She hurriedly strapped him into the car seat and drove the forty miles north to St. Louis, but in the confusion of city traffic she ended up at Cardinal Glennon Hospital, a few miles short of her intended destination. But it was close enough: being a children's hospital, surely its doctors should know what to do for Ryan, Patty thought.

The physicians ordered the usual workup, and when the lab results came back they were shocked: high levels of ethylene glycol—antifreeze—had been found in Ryan's blood. Because Ryan's symptoms were consistent

with ethylene glycol poisoning, the attending physician suspected Ryan had been poisoned. He notified authorities, and Ryan was promptly placed in protective custody.

Patty was distraught. She knew she wouldn't harm her son, and she couldn't imagine that David would, either. Why had he been taken from them? She visited Ryan as often as possible, always under the watchful eye of a social worker, except on September 1. That day Patty was left alone with Ryan for several minutes while she fed him from a bottle.

Three days later Ryan again became ill, exhibiting the same symptoms that had led to his first hospitalization. Lab tests again revealed high levels of ethylene glycol in his blood, and the lab technicians identified a trace of ethylene glycol in the bottle Patty had used to feed Ryan. A second lab confirmed the presence of antifreeze in Ryan's blood, and a search of the Stallings's home turned up a gallon jug of antifreeze. Perhaps with her past in mind, authorities arrested Patty and charged her with poisoning her child. By the time she arrived at the jail, her five-month-old son was barely clinging to life. She was forbidden to see him, and on September 7, 1989, Ryan died. Patty was charged with first-degree murder. The prosecutor said he would seek the death penalty.

While in jail, grieving the loss of her son, Patty realized that she was pregnant again. She was still in jail in February 1990 when she gave birth to her and David's second son, David, Jr., called D.J. D.J. was immediately placed in foster care. Not only was his incarcerated mother prevented from seeing him, but his father, too, was denied contact with his son, even though David Sr. had been charged with no crime and had no criminal record.

A few weeks later D.J. became ill, with symptoms remarkably similar to those that Ryan had exhibited before he died. D.J. was taken to St. Louis Children's Hospital (the one to which Patty had intended to take Ryan), where he was eventually diagnosed with methylmalonic aciduria (MMA), a rare hereditary disease.

People with MMA can only partially break down the nutrients in milk and other foods. In D.J.'s case the problem was due to a missing protein that goes by the name cobalamin adenosyltransferase. This protein is necessary to carry out one of the steps in the digestive process, and without it, D.J. could only partially metabolize the milk he was fed. Consequently, toxic byproducts accumulated in his bloodstream. But because he was correctly diagnosed very early in his life, his diet could be modified before the toxic metabolites took their toll, so D.J. survived.

Could Ryan have died because his personal DNA code also resulted in a nonfunctional version of the same protein? Had toxic metabolic byproducts due to MMA, rather than antifreeze, killed Ryan? Had Patty spent seven months in jail for the "crime" of transmitting to her son a gene that specified a defective protein?

What are proteins? What do they do? Why does the absence of the protein cobalamin adenosyltransferase cause children to become sick? When we say "protein" here, we're not using the term in the generic sense of a constituent of our food, as when we say that meat and eggs and nuts contain a lot of protein whereas bread is mostly carbohydrate, and butter is basically fat. In the context here we are talking about individual proteins, of which there are roughly twenty thousand different varieties encoded by the twenty thousand genes in the human genome. Just as DNA is a chemical, each of those 20,000 proteins is a distinct chemical, in this case composed of carbon, hydrogen, oxygen, nitrogen, and sulfur atoms. (But certain foods we think of as protein-rich contain a lot of a particular type of protein: eggs are rich in a protein called albumin; milk is full of a protein called casein.)

While DNA gets all the glory, proteins do all the heavy lifting. Proteins are the tiny machines that carry out nearly every cellular process, working in conjunction with other constituents of the cell to keep it alive and carry out its functions. The proteins in these machines are like gears and flywheels and valves: they fit together with exquisite precision and act in synchrony to carry out a specific cellular task.

Proteins determine much of what we see when we look at someone. They provide the texture to our hair and skin, and the color to our blood. But most of what they do is done quietly and invisibly. Some proteins function to copy the DNA when a cell divides, others break down nutrients into digestible bits, and other proteins use those bits of nutrients to synthesize new cellular material. Yet other proteins are sentinels that monitor the environment and transmit what they learn about it to the interior of the cell and to neighboring cells. Many proteins are enzymes—like the cobalamin adenosyltransferase that D.J lacked—biological facilitators that speed up chemical reactions, like those that occur when we digest food.

A human cell may make on the order of ten thousand different proteins, but most people are familiar with only a tiny fraction of them. These well-known proteins include insulin, which modulates the amount of sugar in the bloodstream, and hemoglobin, which captures oxygen in the lungs and ferries it through the bloodstream to the tissues. Antibodies are familiar proteins that serve as our border patrol, making the rounds of the body to defend us against potential attackers such as invading bacteria and viruses.

Less well-known proteins are the targets of virtually all drugs. Lipid-lowering drugs known as statins, prescribed to bring down elevated cholesterol levels, inhibit a protein essential for cholesterol synthesis (its name is 3-hydroxy-3-methyl-glutaryl-CoA reductase); the pain relievers ibuprofen and aspirin target a protein involved in inflammation (cyclooxygenase-2); Prozac relieves depression by inhibiting a protein whose job is to regulate the level of a chemical, serotonin, that relays signals between brain cells. AIDS has become a treatable disease because drugs are available to block the activity of two proteins, a protease and a reverse transcriptase, that are necessary for the virus to reproduce.

Proteins are the workhorses of the cell. If we think of our body as a diverse company whose mission is keeping us alive and happy, genes are management; proteins are the labor force.

When disease strikes, the immediate cause is usually the absence of a normal human protein, as was the case with David Stallings Jr., or a detrimental change in a human protein. Cancer results from the uncontrolled division of cells, which can occur either because a protein that normally puts the brakes on cell division is defective, or because a protein whose job is to promote cell division is hyperactive. One form of diabetes is due to the failure to make enough of the protein insulin; another form is due to defects in proteins responsible for detecting insulin. Neurodegenerative diseases such as Alzheimer's, Parkinson's, or ALS (amyotrophic lateral sclerosis, known as Lou Gehrig's disease) are still poorly understood, but it is clear that aberrant proteins play a role in most of them.

Disease can also be caused by the presence of a toxic foreign protein. Cholera, diphtheria, botulism, and anthrax are caused by poisonous proteins that are released from bacteria that have invaded the body.

Proteins, like DNA, are polymers, long chains of a few different types of simple chemicals—in this case, small molecules called amino acids. The

protein polymer is more complex than the DNA polymer because it consists of twenty different kinds of molecules—twenty different amino acids—rather than just the four types of bases (A, C, G, and T) of DNA. A few of these amino acids, such as tryptophan, have achieved some notoriety as dietary supplements. Others have been implicated in disease, such as the amino acid phenylalanine, which causes severe problems for people with the inherited disease called phenylketonuria. But most amino acids remain well off people's radar screens.

Proteins range widely in size, from just a few to thousands of amino acids linked together. The typical protein is a chain of three hundred to five hundred of the twenty different amino acids. Just as with DNA, it is the *sequence* of these amino acid subunits—their exact order in the protein—that determines a protein's chemical and physical properties.

Sixty years ago biochemists argued over whether each protein has a single unique sequence of amino acids, or is a collection of different amino acid sequences. The English biochemist Fred Sanger was the scientist who settled this argument by determining the order of amino acids in the protein insulin, confirming that their sequence is unique. For that accomplishment Sanger was awarded the Nobel Prize in Chemistry in 1958. (In 1980 he became the only person to win two Nobel Prizes in Chemistry, one of only four people to win two Nobel Prizes in any field.) Despite his remarkable accomplishments, he is known for his modesty and his quiet, unassuming nature. He preferred "to putter about in the laboratory" rather than to be a high-profile globe-trotting scientist.

Sanger determined the sequence of amino acids in insulin because it is a small protein that could be obtained in large amounts because of its medical importance. But small as the insulin protein is, it still has too many amino acids to sequence straight away, so Sanger first chopped it into smaller pieces. He then applied the elegant methods he and his colleagues had developed for identifying the order of amino acids in small fragments of protein. Having established the sequence of amino acids in the protein fragments, he was able to assemble the sequence of amino acids in the whole protein.

The process Sanger used is conceptually simple. Imagine a string of letters whose sequence (their order) is to be solved. The string is chopped up randomly into smaller pieces, and the sequence of the letters in each piece is determined. For example, the pieces may have the following

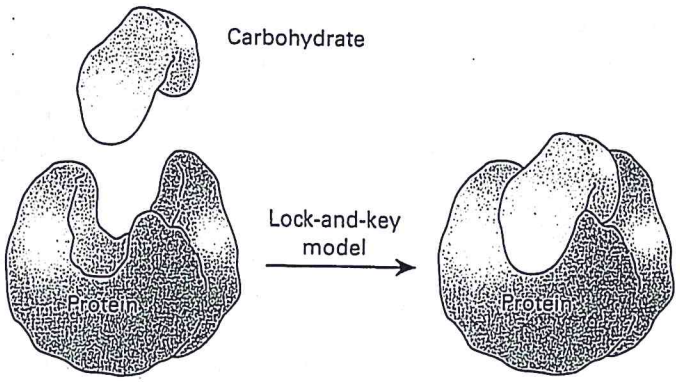
sequences: M-R, L-P-L, R-L-L, L-W, L-L, P-L, W-M-R, L-W-M, L-L-P (each letter is an abbreviation for one of the twenty amino acids). Knowing that these short sequences all come from the same longer sequence, you can line up the fragments:

```
L-W
L-W-M
  W-M-R
     M-R
        R-L-L
          L-L
          L-L-P
            L-P-L
              P-L
```

and see that this sequence must be L-W-M-R-L-L-P-L. This is the order of a stretch of amino acids in insulin. You undoubtedly appreciate that the longer the sequence gets, the tougher the problem becomes. Eventually Sanger was able to work out the order of all fifty-one amino acids in insulin, thus earning himself a trip to Stockholm.

How is it that the sequence of amino acids in insulin instructs cells to take up the sugar glucose from the bloodstream, whereas a different sequence of amino acids of hemoglobin causes it to ferry oxygen around the body? Both proteins are composed of the same twenty amino acids; it's the different order in which the amino acids are strung together that determines each protein's distinct properties. Each of the twenty different amino acids has a different chemical structure, so each has a different shape and different physical properties, which determines how they interact with each other.

The order of the amino acid subunits in a protein chain determines which amino acids interact with each other to cause it to fold up into its own unique three-dimensional shape. Like the ridges on a key, the shape of a protein is the main feature that determines its function and how it contributes to constituting a creature and sustaining life. That's because proteins are designed to fit precisely with other constituents of the cell, much as a key fits into a lock (see figure). Some proteins that play a role in copying DNA have shapes that match particular strings of base-pairs in DNA; some proteins have shapes that enable them to wrap around carbo-



hydrate molecules, which they then cleave into simpler sugars. The protein insulin fits snugly into a pocket of another protein, which then signals that there's too much glucose in the blood. Proteins come in a multitude of shapes and sizes, and those shapes and sizes determine what they can do.

Back in Jefferson County, Missouri, Prosecuting Attorney George B. McElroy III found the evidence against Patty Stallings to be overwhelming. Antifreeze had been found in Ryan's blood on two occasions, by two different diagnostic laboratories, using two different methods of analysis. Those laboratories also found traces of antifreeze in the bottle that Patty used to give Ryan his last meal, and the police found a gallon jug of antifreeze in the Stallings home. Perhaps the most damning evidence against Stallings was the crystals of calcium oxalate found at autopsy in Ryan's brain—a telltale sign of ethylene glycol poisoning.

But since it had been established that D.J. had MMA, a hereditary disease, there was a good chance that Ryan had also had that disease. Could MMA be confused with ethylene glycol poisoning? "Impossible!" said the experts that prosecutor McElroy consulted. They maintained that there was no way MMA could cause high levels of ethylene glycol in the blood. Ryan may have had MMA, they said, but there was no doubt that he had died of antifreeze poisoning. And the Stallings's attorney did not produce any experts to challenge the lab results. The results of the blood tests seemed unimpeachable. It was hard to deny that Ryan Stallings had been poisoned, and Patty was the only person who could have done it. She remained in jail until May 1990, when she was released on bail to await her trial for murder.

How does the sequence of base-pairs in the gene encoding cobalamin adenosyltransferase specify the sequence of amino acids in the cobalamin adenosyltransferase protein? By 1950 it was clear that DNA carries the code for making an organism, and biologists began to focus on how the information specified in the DNA sequence gets converted into proteins. The answer, which was largely worked out by 1965, turned out to be satisfyingly simple.

The DNA sequence does not directly specify a protein sequence. Instead, the process occurs in two steps. First, DNA is copied ("transcribed," in biologists' lingo) into a very similar molecule called RNA (for ribonucleic acid). The RNA is then read ("translated") by a cellular machine to make a protein.

Like DNA, RNA is a polymer consisting of four nucleic acid bases linked together in a long chain. The four kinds of bases in RNA are almost identical to the ones in DNA but not quite: they have one additional oxygen atom. The RNA versions of A, C, G, and T are strung together in the RNA chain in exactly the same order as their counterparts in the DNA that directs production of the RNA copy. The same order of bases is maintained in the RNA because the DNA base-pairing rules (A pairs with T; G pairs with C) also apply to RNA. The protein machine that makes RNA glides down one of the DNA strands "reading" the sequence of bases while synthesizing an RNA copy. When the machine encounters a T, it inserts the RNA version of A in the RNA chain. If the next base it encounters is a G, it inserts the RNA version of C in the growing RNA chain; if an A, it inserts the RNA version of T, and if a C, it inserts the RNA version of G.

RNA differs from DNA in another important way besides the extra oxygen atom: it is single-stranded, a copy of just one of the two strands of DNA. For some genes the RNA-synthesizing machine reads the "Watson" strand of the double helix while making the RNA copy; for other genes it reads the "Crick" strand while making the RNA copy. In either case the product is a single-strand RNA copy of the gene suitable for directing synthesis of a protein.

In the second step of protein production the sequence of bases in the RNA copy of the DNA is translated into a protein sequence by another molecular machine composed of many different proteins. The protein-synthesizing machine reads the RNA bases in groups of three, each succeeding group of three bases in the RNA specifying one of the twenty amino acids at each succeeding position in the protein chain.

Each sequence of three bases, called a base triplet, specifies a particular amino acid. The list of all possible base triplets and the amino acid each of them specifies is the genetic code. It is much like the list of dots and dashes of the Morse code, which specify letters of the alphabet. For example, the triplet AAG specifies—codes for—the amino acid lysine: whenever those three bases appear next to each other in a stretch of RNA, they cause the code-reading machine, traveling down the RNA like a train on its tracks, to insert lysine at the corresponding position of the protein chain. If the next base triplet in the gene is CGA, then the next amino acid added to the growing protein by the code-reading machine is arginine. There are sixty-four possible base triplets (four possible bases at the first position of the triplet times four possible bases at the middle position times four possible bases at the last position of the triplet), but only twenty amino acids, so most of the amino acids are specified by more than one triplet of bases. For example, the amino acid lysine is specified by both the AAA and AAG base triplets. Four of the base triplets have special roles: ATG serves as the signal for the code-reading machine to START making protein, beginning with the amino acid methionine; TAG, TAA, TGA are signals to STOP translating the RNA sequence into protein.

With this knowledge, we can translate into amino acids the beginning of the RNA that gets copied from the DNA sequence of chromosome 12 shown in the last chapter, which encodes the beginning of the cobalamin adenosyltransferase protein. Once the code-reading machine finds the ATG triplet in the RNA template, which tells it to start to synthesize a protein, each succeeding triplet directs the insertion of the next amino acid into the growing protein chain. The beginning of the RNA encoding cobalamin adenosyltransferase directs the synthesis of these nine amino acids in precisely this order at the beginning of the protein: methionine-alanine-valine-cysteine-glycine-leucine-glycine-serine-arginine:

CTGGCGGGGTCAGGTCCCGTCAAGCAGCCTGGCTC ATG GCT GTG TGC GGC CTG GGG AGC CGT ...

methionine alanine valine cysteine glycine leucine glycine serine arginine ...

The code-reading machine marches down the complete RNA template, three bases at a time, using the genetic code to translate each base triplet into an amino acid that gets incorporated into the growing protein chain. Eventually it encounters one of the three triplets that tell it to stop translating the RNA sequence (for this RNA, the code-reading machine would continue for 723 more bases before it encounters a "STOP" triplet, producing a cobalamin adenosyltransferase protein of 250 linked amino acids).

A conceptual framework that may be helpful to understanding the roles of DNA, RNA and protein in the cell has DNA as the wiring diagram for the circuitry of the cell, RNA as the carbon copy of the diagram that gets carried to the fabricators, the genetic code as the legend that reveals what all the squiggly symbols in the wiring diagram mean, and proteins as the switches, batteries, lights, fuses, and other components of the circuits. A mistake in a part of the wiring diagram (a gene) can lead to a defective component (a protein), which can lead to a faulty circuit (disease).

Prosecutor McElroy told the jury: "Don't try to understand why Patricia Stallings poisoned her child by feeding him from a baby bottle laced with antifreeze. The point is she did it. Only she could have done it." After hearing these words, the jury didn't take very long to reach a verdict. A few hours later, on February 1, 1991, the jury foreman, Delmar Fisher, stood before the court and announced the verdict: Patty Stallings was guilty of first-degree murder. A few weeks later Circuit Judge Gary P. Kramer sentenced Patty to life in prison without the possibility of parole. Patty's friends and family sat in the gallery wearing T-shirts bearing the legend "Please help us: Patricia Stallings is innocent."

In fact, help was on the way. Patty's husband, David, had been working hard to get the case more publicity, hoping that someone who was able to help would take an interest in Patty's plight. He managed to get the producers of the TV show *Unsolved Mysteries* interested in the case, and they ran an episode on Patty's predicament in May 1991.

Among those who watched the show was Dr. William Sly, a well-regarded geneticist and pediatrician who was chairman of the Department of Biochemistry at Saint Louis University. As a coauthor of the major textbook on inherited metabolic disorders, Sly well knew how similar are the effects of MMA and ethylene glycol poisoning, and he was very skeptical that Ryan could have suffered from both.

Dr. Sly learned that one of his colleagues, Dr. James Shoemaker, who ran a metabolic testing lab at Saint Louis University, had obtained a small sample of Ryan's blood from one of the labs whose analysis had helped convict Patty. Shoemaker's analysis of the sample also turned up something that looked like ethylene glycol, but only a small amount, nowhere near enough to poison a child. But he saw something else—something that the other two labs had not reported: a large amount of propionic acid.

Shoemaker and Sly knew that propionic acid, which is chemically very similar to ethylene glycol, is a toxic metabolite that accumulates in the blood of people with MMA. Could propionic acid in Ryan's blood have been misidentified as antifreeze? Sly and Shoemaker scrutinized the results from the labs that claimed to have found antifreeze in Ryan's blood, and they were taken aback: the results matched those obtained from a pure sample of propionic acid, and not those of a pure sample of ethylene glycol.

Sly sent a letter to Prosecutor McElroy stating that he was confident Ryan had died from MMA, not from ethylene glycol poisoning. McElroy started to have some misgivings about his case against Patty Stallings, but he was still not convinced of her innocence. What about the ethylene glycol in the bottle Patty used to feed Ryan, and the gallon of antifreeze found in her house? And, most important, how to explain that signature of ethylene glycol poisoning—crystals of calcium oxalate—that the coroner found in Ryan's brain?

The Stallings had fired their first lawyer, and their new lawyer, renowned St. Louis attorney Robert Ritter, asked McElroy: "What would it take to convince you Patty did not poison her son?" The prosecutor said he needed to hear from another expert on metabolic diseases, someone renowned in the field and not associated with the case.

Ritter approached Dr. Piero Rinaldo, a well-respected geneticist on the faculty at Yale University and an expert on inherited metabolic diseases. It didn't take Dr. Rinaldo long to agree with Dr. Sly that both labs that analyzed Ryan's blood misread the results. Their analysis, Rinaldo told *St. Louis Post-Dispatch* reporter Bill Smith, was "totally unacceptable, unbelievable, out of this world. I was astonished. I couldn't believe that somebody would let this go through a criminal trial unchallenged."

Prosecutor McElroy had finally heard enough. On September 19, 1991, two years after Patty was first arrested, after she had mourned the death of her son Ryan, had spent thirteen months in jail, and had never been

allowed to spend time with her new son D.J., Patty was absolved of all charges against her. In front of reporters and television cameras, Prosecutor McElroy apologized to Patty and David for what he had put them through: "Unfortunately, we can't undo the suffering that the Stallingses have endured during this entire ordeal. And I apologize to them, both personally, and for the state of Missouri." The subdued smile on Patty's face belied her bittersweet feelings.

What about the traces of antifreeze found in the bottle Patty used to feed Ryan? The bottle had been washed in a dishwasher and filled with infant formula before testing, and the compound identified as ethylene glycol "could have been anything," Rinaldo concluded. "Their approach was: anything that showed up in a certain window in that chromatogram would automatically be labeled ethylene glycol. This is just . . . unacceptable," he said with a sad and disbelieving shake of his head.

And those crystals of calcium oxalate in Ryan's brain? Dr. Rinaldo concluded that they were a result of the ethanol drip used to treat Ryan's presumed ethylene glycol poisoning, an appropriate treatment for that condition, but completely inappropriate for someone with MMA; Dr. Rinaldo suspected it had, in fact, hastened Ryan's death. Two years later the Stallingses received out-of-court settlements for Ryan's wrongful death, from Cardinal Glennon Children's Hospital and from the laboratories that got his diagnosis wrong. The amount of the settlements was not disclosed, but whatever the amount, the money cannot possibly have compensated Patty and David for the loss of their son and their ordeal.

But Patty *was* lucky in one regard: there was only a 1-in-4 chance that D.J. would inherit from both his parents a version of the gene that encodes a defective cobalamin adenosyltransferase. In chapter 7 we'll find out why this is so. If D.J. had been born healthy, there would have been no clue that Ryan suffered from a hereditary disease, and Patty most likely would have remained in jail. Patty beat the odds and was absolved of Ryan's death; D.J. did not beat the odds.

4 All from a Single Cell: How a Fertilized Egg Develops into a Baby

Nine months after a human egg is fertilized, a baby's lungs fill with air and she bawls out her first lusty cry. Just thirty-eight weeks ago she was a single cell, created by the union of one of her father's sperm and one of her mother's eggs. How did one cell give rise, in that short span of time, to an organized mass of human flesh with limbs and lungs in the correct places, with the proper number of fingers and toes, and with eyes and ears and everything else working properly? It seems like a miracle. While it is marvelous, it's not miraculous: biologists have learned the principles of the process that produces a complex organism from a single cell.

This process has intrigued scientists for a long time. An early theory to explain human development, dating back more than two thousands years, is that of preformation. This theory provided a simple answer: we already contain in our bodies very small but fully formed members of the next generation, who merely grow within the mother until they reach the size of a baby able to survive outside the womb. Many scientists thought they saw this tiny person—which they called a homunculus—when they peered at sperm through the first microscopes in the seventeenth century.

This explanation sidestepped the seemingly intractable issue of how complexity unfolds. But there's a big problem with this theory of little people: the homunculus must carry its own sperm that shelter an even smaller version of the person who will be born in the next generation, and that prehomunculus must have in its sperm an even smaller pre-prehomunculus that is to be born two generations hence, and so on ad infinitum for all future generations of humankind. By the nineteenth century embryologists had come to see this fundamental flaw in the preformation theory. The alternative view was that the egg was formless, and,

after fertilization, goes through a series of transformations that result in a fully formed individual.

But how? Enter now the fruit fly, *Drosophila melanogaster*. The humble fruit fly seems to appear like magic whenever we leave an open bottle of wine on the table or neglect to toss out a banana peel, calling to mind another discredited theory—spontaneous generation—the idea that life forms can spring from nonliving material. *Drosophila* species are cosmopolitan, having hitchhiked from place to place along trade routes, and spread west in North America with the migration of people and their fruits and vegetables and garbage.

The fruit fly also populates thousands of research laboratories, serving as an ideal subject for the investigation of all sorts of biological phenomena. With its small size (a mere 0.1 inches head to tail), short generation time (just a couple of weeks), large litters (hundreds of eggs per mom), and low feeding and housing costs (quite happy to spend their lives in milk bottles feeding on yeast), *Drosophila* has been a fond object of biologists' attention for more than a century. And it is this fly that has yielded many of the secrets of embryonic development.

That a fly would be key to unlocking the path from egg to adult seemed unlikely in the early part of the twentieth century. Tiny *Drosophila* made its name not in developmental biology but in genetics, while larger animals like the frog and the sea urchin were the darlings of embryologists. For a period of about thirty years, beginning around 1910, researchers in the laboratory of Thomas Hunt Morgan—first at Columbia University, later at the California Institute of Technology—made groundbreaking genetic discoveries using the fly. These included showing that genes lie on chromosomes, uncovering the process by which chromosomes exchange pieces of themselves, and figuring out that sex-linked traits are specified by the X chromosome, discoveries that we will discuss shortly, and that garnered a Nobel Prize for Morgan in 1933.

In the 1940s, following its heyday in Morgan's laboratory, *Drosophila* was eclipsed by even smaller creatures as the objects of geneticists' attention. Taking its place in the new field of molecular biology were the bread mold *Neurospora crassa* and the intestinal bacterium *Escherichia coli* and its viruses. Experiments on these rapidly dividing organisms revealed the nature of the gene, the genetic code, the process of protein production, and the principles of gene function.

Beginning in the 1970s, *Drosophila* began its comeback, led by a young German biologist, Christiane Nüsslein-Volhard, who dazzled developmental biologists with her work showing how a single cell turns into a fully formed organism with trillions of cells. In partnership with a young American biologist, Eric Wieschaus, Nüsslein-Volhard tackled a project so audacious in its concept that another geneticist wondered, "Does she have the whole German army working for her?" But it was just Nüsslein-Volhard and Wieschaus, sitting across from each other at a small table in their lab in Heidelberg, Germany for an entire year isolating mutant flies—ones with changes in their DNA sequence that produce deformed embryos—in the hope that learning what goes wrong in each mutant would reveal how the normal flies do it right.

Nüsslein-Volhard and Wieschaus's mutant flies, first described in 1980 in the international scientific journal *Nature*, were crucial to solving the mystery of development, because they led to the identification of the key proteins that decide each cell's fate by turning particular genes on or off. The two biologists analyzed the flies' cells as an investor might analyze a new company to predict whether it is going to be successful: identify the key executives, find out what critical decisions they are making, and observe how the company responds to their strategic mistakes. Nüsslein-Volhard and Wieschaus were shrewd investors: their acumen won them the 1995 Nobel Prize in Physiology or Medicine.

What was striking about Nüsslein-Volhard's approach was its simplicity: it required only a commercially available chemical to cause mutations in the flies, an ordinary microscope for observing the fly embryos, and standard genetic analysis—all of which were available as far back as 1930. Why did no one think to try this approach in the intervening four decades?

Nüsslein-Volhard had been trained as a biochemist; she wrote her doctoral dissertation on her studies of an RNA-synthesizing enzyme from bacteria. She turned to *Drosophila* because she wanted to apply genetics to the problem of development, and found that she "immediately loved working with flies. They fascinated me, and followed me around in my dreams." As a newcomer to the field of developmental biology, Nüsslein-Volhard was unencumbered by the constraints that limited the thinking of other scientists interested in these problems. "I, compared to other people working in this field, came up with ideas. They were blocked in

their minds. Other biologists would say, "'This is not done. We don't do that in our field.' . . . I did things that were completely unconventional."

Nüsslein-Volhard knew that the different types of cells in an organism are different because they deploy (biologists say "express") different sets of genes to produce different kinds of proteins. How did she know that? Isn't it possible that a cell develops into a liver cell rather than a lung cell because it possesses a different set of genes than the lung cell? Might not unspecialized cells of a developing organism lose genetic information as they divide, retaining different sets of genes that determine the type of cell they will eventually become? A cell destined to become a liver cell might retain only those genes that are needed to specify a liver cell; a cell destined to become a lung cell might retain a different set of genes that cause it to become a lung cell. Might it work like that?

This very reasonable idea—that development of one cell type might proceed by loss of genes for all other cell types—was ruled out in the 1960s by the English scientist John Gurdon, who showed that specialized cells possess all the genetic information necessary to specify all the other types of cell in an animal; specialized cells do not lose any genes while assuming their particular identity. Gurdon established this principle with a clever and technically impressive experiment with frogs: he used the genetic information present in a single specialized frog cell to program the development of an entire frog. He was the first person to clone an organism.

Gurdon began by removing the chromosomes from an unfertilized frog egg, literally reaching into the egg with a very thin straw and sucking out its nucleus, the part that contains the DNA wrapped up into chromosomes. He then put into that DNA-denuded egg a nucleus he had similarly extracted from a cell he obtained from a frog's gut. If the specialized gut cell had acquired its identity because it retained only gut cell genes, then its genetic material should not be able to program that egg to develop into a frog, because it would be missing genes necessary for making other types of cells. But Gurdon saw complete, normal frogs develop from some of the eggs he had manipulated. He concluded that specialized cells carry all the genetic information necessary to specify an entire animal.

Because Gurdon did these experiments with frogs, his conclusion was met with some skepticism. Some people said the rules for frog development might be different from the rules for other animals. Frogs, after all, are

cold-blooded—very different from us and our warm-blooded cousins. Thirty years later, Ian Wilmut, a Scottish veterinarian, quieted any remaining doubters when he cloned a sheep. Employing Gurdon's methods, Wilmut replaced the nucleus of a sheep egg with a nucleus taken from a cell of an adult sheep's mammary gland. The reprogrammed egg was placed into a ewe who served as a surrogate mom, and five months later Dolly burst into the world, proving that the mammary gland cell carried all the information necessary to create a fully formed, normal lamb. This experiment has since been successfully repeated with many other kinds of animal using several different kinds of specialized cell as the source of the nucleus that programs the egg, proving beyond a reasonable doubt that virtually every one of our cells carries the same complete set of genes.

If all cells in an organism have the same genes, specialized cells must acquire their particular identity by using only some of those genes. A liver cell is what it is because it uses only a subset of its genes, those that provide the proteins that make a liver cell and carry out its tasks. It does not express genes for making brain or bone cells. Lung cells deploy a different set of genes, which give them their unique characteristics; they do not express genes for making skin or spleen cells. And so on for the hundreds, perhaps thousands, of different types of cells in our bodies. So now the key question becomes: How do cells in the developing embryo come to use some genes but not others and thereby become a specific type of cell, eventually leading to the organized mass of tissues we call an organism?

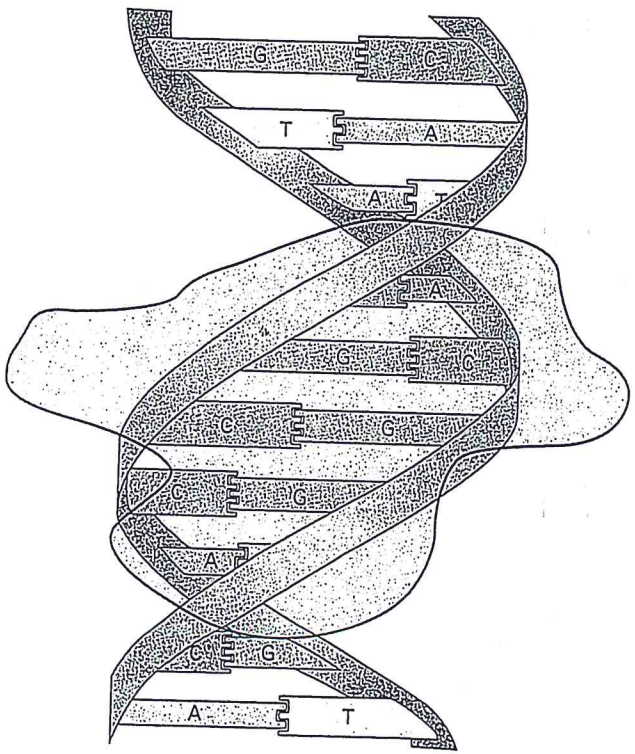
As we discussed in chapters 2 and 3, genes are stretches of DNA that contain the information for making a protein. Genes direct the synthesis of a protein by first being copied into a molecule called RNA, which is very similar to DNA but consists of only one strand of bases rather than the two of the double helix. This first step in the expression of a gene is carried out by a special protein machine in the cell that transcribes the sequence of the DNA into RNA copies that are then translated into proteins, much as medieval monks transcribed sacred texts onto papyrus for translation by their colleagues.

Only some of the genes in each cell are used like this: maybe only ten thousand of the twenty thousand or so genes in each cell get expressed. Each gene has a switch that controls whether it is "on" or "off." If the switch is in the "on" position the gene will spring into action and be transcribed

into RNA; if the switch is in the "off" position the gene will remain at rest. The switches of some genes are in the "on" position only in muscle cells, while the switches of other genes are flipped "on" only in nerve cells.

What determines whether a gene's switch is on or off? The decision is made by a class of proteins that we can think of as the executives: their job is to decide whether certain genes are to be on or off. They do this by recognizing and binding to specific DNA sequences near particular genes and regulating their transcription into RNA. Hence their name: transcription factors.

Each transcription factor recognizes one particular short DNA sequence (usually six to twelve base-pairs in length) that is present near the genes it controls. A remarkable property of transcription factors is that they can find their short recognition sequence among the other three billion base-pairs of DNA in the human genome (see figure). They rapidly search through the genome—much the way Google searches through billions of web pages—until they find their sequence, and then they glom on to it.



Most genes contain recognition sequences for several transcription factors. The sum of the effect of each transcription factor bound to the gene determines the state of the gene's switch. Some transcription factors act to turn transcription on, others strive to turn it off. The transcription factors are like the transistors that constitute the motherboard of a computer, integrating the input they receive and responding with the coordinated output you see on your screen. This integrated circuitry of transcription factors bound near a gene constitutes the switch that turns the gene on or off.

Actually, these switches are more like rheostats that can be turned up or down, the brightness or dimness of the rheostat's setting being determined by the particular combination of transcription factors that are bound to the gene. Since the human genome encodes about fifteen hundred different transcription factors, the number of different combinations of them is huge, so the rheostats can be set to an almost limitless number of levels. And since the settings of the rheostats on all 20,000 genes determine the identity of a cell, the great diversity of cell types in the human body should no longer be a surprise.

Wise investors know that too many executives often spell doom for a company, so we may wonder why successful organisms such as humans have so many transcription factors. But a complex organism has to make many more decisions than even the largest of companies, and we need all those transcription factors to do that. The factors ask questions about what's going on inside and outside the cell: Are there enough nutrients? What are the cells next door up to? Is there a big demand in the rest of the body for things this cell makes? And many, many other important questions.

The transcription factors learn the answers to these questions, integrate that information, and take action by turning on the genes that are needed (and turning off those that are not needed) by a cell that finds itself in that specific situation at that particular time. The diversity of transcription factors allows many questions about cellular fitness to be asked simultaneously and continuously. The answers to those questions comprise a huge amount of data that the transcription factors process in deciding which genes should be active, and thus which proteins will be present at that specific time in that particular cell.

The decision to turn a gene on or off is like the choice an editor must make whether to run a story about a big fire with a banner headline on

page 1 or to go with a more modest mention on an interior page. The editor gets input from several reporters: some at the fire watching a rescue in progress, others at the mayor's press conference hearing what the city's emergency teams are doing, and a few at the hospital listening to the stories of victims. The editor integrates this input and decides the story will run on page 3, but with a large headline. Transcription factors are the cells' editorial staff, collecting information from a press corps of proteins that gather a huge amount of news as they survey the situation.

We can see, then, that cells become different by expressing different sets of genes, which results from each kind of cell having a unique collection of transcription factors. Liver cells have a corps of transcription factors that turn on genes necessary to make a liver cell and turn off genes necessary for making other cell types; lung cells have a different corps of transcription factors that are responsible for turning on the genes lung cells need and for turning off the genes used by other types of cells.

The whole developmental program, from the first division of the fertilized egg to the birth of a fully formed organism consisting of trillions of cells, is largely a diversification of the transcription factor collections in cells as they divide. How does a cell that is destined to contribute to the iris of the eye come to possess *just* the right set of transcription factors to ensure that the genes for making an iris (and *not* genes for making a retina, or lens, or cornea) are expressed at *just* the right levels and at *just* the right time as the embryo develops? In other words, how do different cells come to possess different transcription factors?

Cells assemble their complement of transcription factors by expressing the genes that encode those transcription factors. Each transcription factor is a protein encoded by a different gene, and the subset of those transcription factor genes that a cell expresses determines the collection of transcription factors it contains.

What determines which transcription factor genes get expressed in a cell? The same mechanism that determines the expression of every other gene in the cell: the particular collection of transcription factors it contains.

Oh, oh . . . We seem to have boxed ourselves into a corner: different cells express different genes because they possess different combinations of transcription factors. But they possess different combinations of transcription factors because the genes that encode those transcription factors are acted upon by yet other combinations of transcription factors. A bit circular, isn't it?

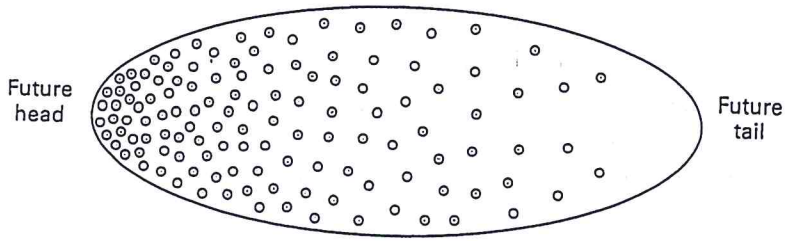
The genes that encode transcription factors, like all genes, have rheostats that govern their output, and, as is the case for all genes, those rheostats are set by transcription factors. So the set of transcription factors that a given cell has at any given moment is the result of which particular transcription factor genes were expressed during the course of development of that cell. This logic makes the developing organism seem like a set of nested Russian dolls: to determine why a set of transcription factors came to be present in a liver cell, you have to look at the transcription factors in the cell that gave rise to the liver cell, and to determine why that particular set of transcription factors came to be present in *that* cell you have to look at its precursor cell, and so on, all the way back to the original fertilized egg.

That is precisely what Nüsslein-Volhard set out to do: go all the way back to the first few cells of the embryo and identify the transcription factors they have that make them different, then learn how the cells produced in successive divisions come to possess different combinations of transcription factors that cause them to express unique sets of genes and thus become increasingly specialized.

The genes she discovered that control this process operate by a few general principles. While these principles are simple, the complex process of development is anything but. We'll illustrate the principles that govern development in the fly; human development operates a bit differently, but the fundamental principles are similar.

One principle is that the fly egg, even before it ever sees a sperm, is already subdivided into specialized areas: one end will give rise to the head, the other end to the tail; one part will become the top of the fly, another part the bottom. The basis of this polarity of the egg is chemical gradients in the egg.

Certain proteins in the egg get synthesized by the mother at one end of the egg—say, the end that will become the head—and their levels diminish as they spread outward toward the other end of the egg. If one of these proteins is a transcription factor, it would be most effective

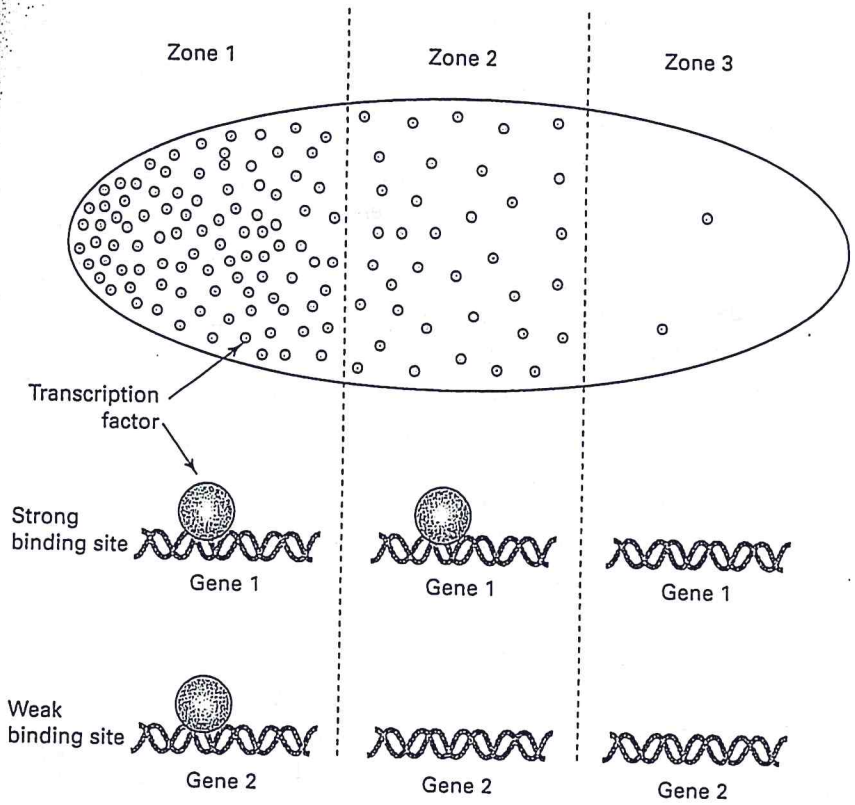


controlling the activity of genes in cells that lie near the end of the embryo destined to become the head, with decreasing effectiveness as its concentration diminishes toward the end of the egg that will form the tail (see figure).

This can be visualized by imagining that you've opened up a can of blue paint in preparation to repaint your kitchen. It sits peacefully in a corner while you gather the brushes and track down the tarp. Just then your teenager zooms in to demonstrate her latest skateboard maneuver, tipping over the can as she glides across the room. The blue paint pours across the floor, a thick puddle in the region nearest the corner where the can stood, thinning out as it spreads across the floor. There is now a gradient of paint that spreads from one end of the kitchen to the other.

A second principle is that different genes respond to different amounts of a transcription factor. One gene might need a high level of a transcription factor to be turned on, a level present only at the region of the egg that will give rise to the head. This may occur because the DNA sequences in the gene that that transcription factor binds to are not very good matches to the sequence it recognizes, so that many copies of the transcription factor are necessary to ensure that some of them recognize and latch on to the partial recognition sequence. Another gene might contain a DNA sequence that is a close match to the sequence recognized by that transcription factor and may therefore require less of the transcription factor to be switched on. As a consequence, that gene will be turned on in cells farther away from the head-forming end of the embryo.

A concentration gradient of a single transcription factor will already define three zones of the fertilized egg: a zone of high concentration at one end of the egg (say, where the head of the fly will form), where the factor turns on genes containing strong and weak recognition sequences



for the transcription factor; a zone of medium concentration near the middle of the egg, where it turns on only genes that have close matches to the recognition sequence (strong binding sites for the transcription factor); and a zone of low concentration near the opposite end of the egg (where the tail of the fly will form), where there is not enough of the transcription factor to turn on either kind of gene (see figure).

A third principle is that cells talk to one another, and these conversations influence which genes get expressed, much as conversations in the hall of a high school influence who is going to the prom with whom. Neighboring cells communicate with each other through proteins they display on their cell surfaces, which act like molecular feelers, or antennae. When these antennae make contact with a neighboring cell, or detect molecules given off by neighboring cells, they send signals into the cell that affect the function of certain transcription factors that result in changes in gene expression.

Among the genes whose expression is affected by these signals are those that encode transcription factors. Since cells in different parts of the developing embryo get different cues from their neighbors, their antennae generate different signals, and thus different cells come to express different sets of transcription factor genes, which eventually cause them to express different genes, which determine the fate of the cell.

These kinds of intercellular conversations are constantly going on in the developing embryo. It's a veritable cacophony. Let's listen in: "Hi neighbor! I've decided to become a cell of the iris, but I can't form the iris all by myself so I'd like you to join me. Hey, you over there! Listen up and get with the program! I'm sending you a signal, so pay attention. And after you receive it make sure you pass it on to your neighbors. We're going to need some of them to become cells for a cornea." By means of these intercellular conversations cells continually refine the set of transcription factor genes they express, ultimately causing them to express the specific set of genes that results in their taking on very specific functions.

Why was it that Christiane Nüsslein-Volhard, rather than some other biologist, had the idea to seek the *Drosophila* mutants whose analysis would reveal these principles? Evelyn Fox Keller points out that as a German scientist, Nüsslein-Volhard was less affected by the gene-centric view of biology typified by the Americans, and was more willing to consider how the other components of the cell participated in the process. Furthermore, as a molecular biologist she was impatient, unlike many developmental biologists; she was accustomed to getting quick results from her experiments. Most critically, she had the imagination to come up with novel ideas.

If everything goes right with the gradients of transcription factors, with the combinatorial interplay of proteins sitting on the DNA, with the intercellular conversations and negotiations, and with the many other things that go into the developmental process, then a complete organism is eventually born, with limbs and lungs in the right places and with the proper number of fingers and toes and with irises and eyelids that work.

Most of the time it does go right—remarkably so, given the complexity of the process. But things can go wrong. A very large percentage of human pregnancies—perhaps 30 to 50 percent—spontaneously abort before the

pregnancy is detected because something goes very wrong soon after fertilization of the egg. Fifteen to 20 percent of known pregnancies also result in miscarriages, most of them probably due to mistakes in the developmental program. And the parents of one out of every twenty-eight babies get the distressing news that their child has a birth defect. But when one considers everything that must go right with the process for a healthy child to be born, it's remarkable that we're here at all, let alone with all of our organs and limbs in their proper places and working correctly. And all from a single cell!

5 When the Gene Is the Cure: Immunodeficiency and Gene Therapy

David Phillip Vetter could not live like this any longer. His doctors knew it; his parents knew it; he knew it. They all agreed he had to risk the bone-marrow transplant. Without it he would have to continue living in the bubble—his sterile isolation chamber—waiting for a cure to be developed for his affliction. Because David suffered from Severe Combined Immunodeficiency (SCID), he had no immune system to fight off even the most timid of invaders. He had already waited for twelve years, and still no cure for his condition was in sight. On October 21, 1983, he received some of his sister's bone marrow. It didn't take. Worse, it gave him cancer. He died February 22, 1984, 15 days after walking out of his bubble for the first time.

The first son of Carol Ann and David Vetter Jr. also began life with no immune system, and died of a massive infection six months after birth. His personal DNA code included an X chromosome, inherited from his mother, that carried a defective copy of the gene called *IL2RG*, which provides the instructions to make a protein required for the immune system to develop properly. Because there was a mutation—a change in the DNA sequence—in the *IL2RG* gene David Joseph inherited from his mother, the gene directed the production of a nonfunctional protein. Without the IL2RG protein, David Joseph's thymus, a small organ near the lungs where immature white blood cells from the bone marrow bivouac before going into battle, could not send off white cells to fight infections.

After their experience with their first son, Carol Ann and David Jr. understood that if their next child were a son, he would also have a 50 percent chance of being born with no immune system. A son has only the single X chromosome he inherits from his mother, his other sex chromosome being the Y chromosome he inherits from his father. So if one of the genes on the X chromosome were defective, he would suffer the consequences

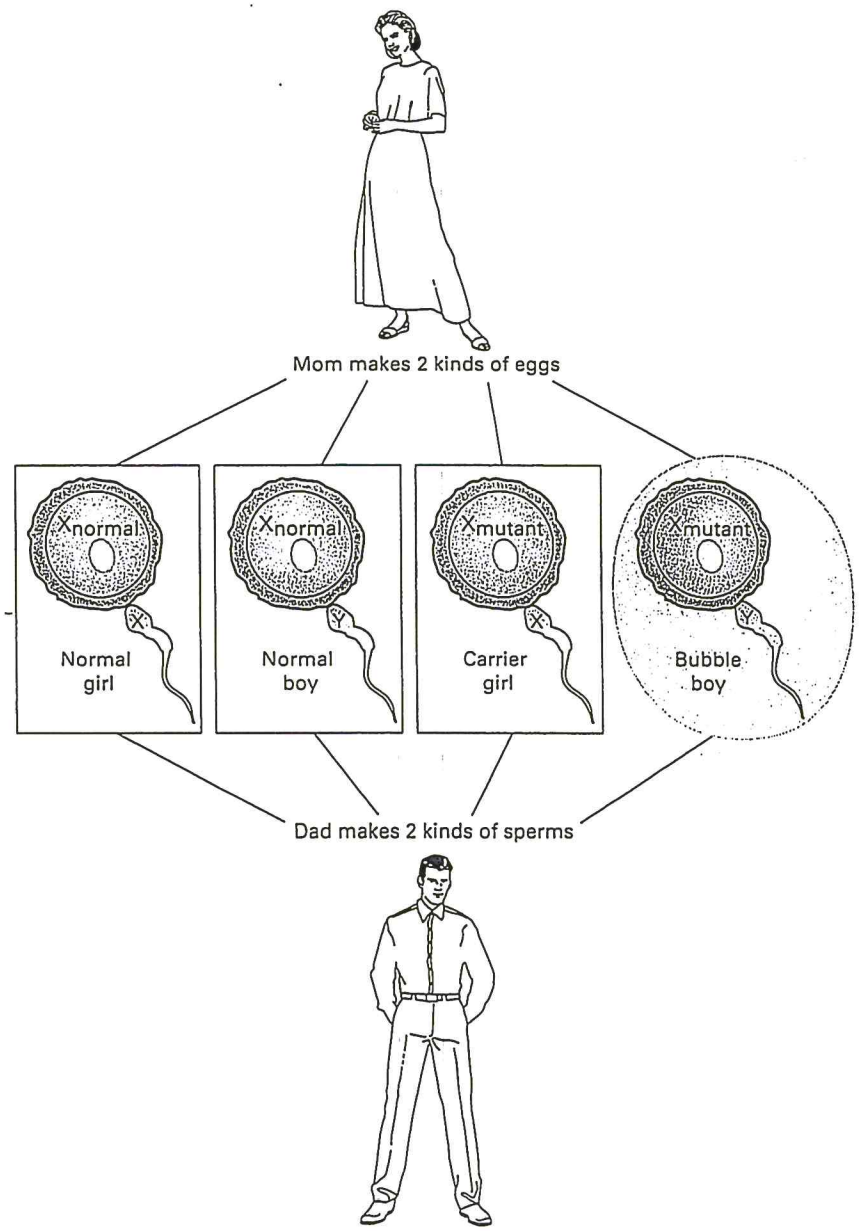
(see figure). A daughter would be safe, because even if Carol Ann gave her the X chromosome with the defective gene on it, her father, David Jr., would provide another X chromosome carrying a good version of the gene. (In chapter 7 we discuss in more detail why one good gene may be all you need.)

Diseases like SCID that are due to a defective gene on the X chromosome are passed to boys only from their carrier mothers, who have a good version of the gene on their other X chromosome. Males with the mutant gene on the X die of the immune disease before they are old enough to reproduce and pass the flawed X chromosome on to their daughters.

But the Vetters were told that even if their next son were unlucky and drew the defective chromosome, he would not necessarily be doomed: The doctors thought they could cure his disease, either with a bone-marrow transplant from his sister or with a cure they thought was just around the corner. They had on their team Dr. Raphael Wilson, an expert in germ-free environments, who would build and maintain the sterile isolation chamber—the "bubble"—that would protect their infant son from the germs that had killed his brother.

A few years before, Wilson had reported stunning success in Germany with a sterile isolator he built for twins with immunodeficiency: after a short time in the bubble their immune systems suddenly, and inexplicably, came to life, and the twins were taken out of the isolator. So the Vetters' doctors were optimistic. Wilson "just swept us along with his enthusiasm. He had the confidence to say, 'We can do this. We can do this,'" said Dr. Mary Ann South, one of the members of the medical team, in a documentary film about the child who became known as the Bubble Boy.

Carol Ann and David Jr. were eager for another child. Although they had a healthy girl, they wanted a boy to carry on the Vetter family name. "Children were very essential to our hope and to our dream of the future. We wanted to have children right away; we wanted to have as many as God would send us," Carol Ann explained in the TV documentary. So they had another child. Happily, one of their dreams came true: it was a boy. Sadly, their other dream did not: the boy did not inherit his mother's functional *IL2RG* gene. Instead, like his older brother, he inherited her X chromosome that carried the defective gene. He was whisked into the isolator within seconds of his birth, and that's where he stayed for twelve years, until it became obvious that no cure was imminent.



David Phillip Vetter, the Bubble Boy, lived a celebrated life that stimulated a hit song by Paul Simon, feature films starring John Travolta and Jake Gyllenhaal, and an episode on *Seinfeld*. Celebrated, but tragic. The journalist Steve McVicker described in a 1997 article in the *Houston Post* how David responded when his friend the psychologist Mary Murphy asked him why he was so angry: "Why am I so angry all the time? Whatever I do depends on what somebody else decides I do. Why school? Why did you make me learn to read? What good will it do? I won't ever be able to do anything anyway. So why? You tell me why!" Murphy had no answer for David Phillip.

Had David Phillip Vetter been born twenty years later he might have chosen to wait a little longer, because in the year 2000 a cure for SCID finally became available. Not a perfect cure, but there can be little doubt David would have jumped at the chance to try it. The cure comes in the form of the good *IL2RG* gene—the gene whose lone copy on David's lone X chromosome didn't work.

If a functional copy of the gene can be delivered to the bone-marrow cells of a SCID patient, those cells begin to make white blood cells competent to fight infections, giving the patient something he wasn't born with: a functional immune system. Treating disease with genes—gene therapy—is the brass ring that David and his parents and his doctors were waiting for. It cured the disease for thirteen boys in France and England. But gene therapy came too late for David.

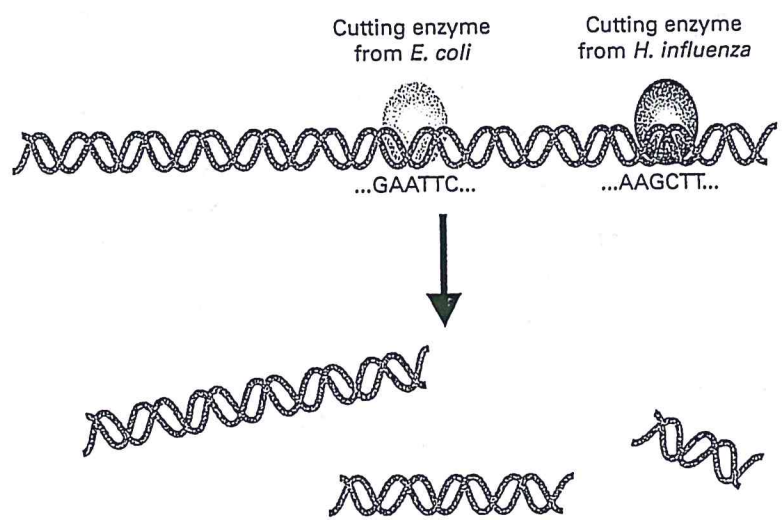
What occurred after David's death in 1984 that made gene therapy a viable treatment for his disease by 2002? Lots. The human genome was mapped, making it possible by 1987 to identify the region of the X chromosome that carries the *IL2RG* gene. These maps, as we'll discuss in chapter 12, show the positions of genes along a chromosome, just as roadmaps show the positions of cities along a highway. By 1993 scientists had isolated the *IL2RG* gene, using methods for isolating genes developed in the 1970s. In the 1990s scientists devised methods to deliver genes to human cells, so by 1999 they could deliver the *IL2RG* gene to the bone-marrow cells of five children with SCID. By 2002 it was clear that most of these children were cured: four have a nearly normal immune system and are enjoying what David longed for: a life outside the bubble.

How, exactly, was all of this done? Once a gene is located on a chromosome, how is it purified and isolated in the test tube? The principle

is simple: the chromosomes are fragmented into small pieces of DNA, and the piece containing a particular gene is fished out of the mixture and copied millions of times, in a process called cloning. It's like making copies of an animal, as was done to clone the sheep Dolly, but in this case multiple identical copies (clones) of the gene are made from a pure template. Each copy is a clone of the original gene that provided the template.

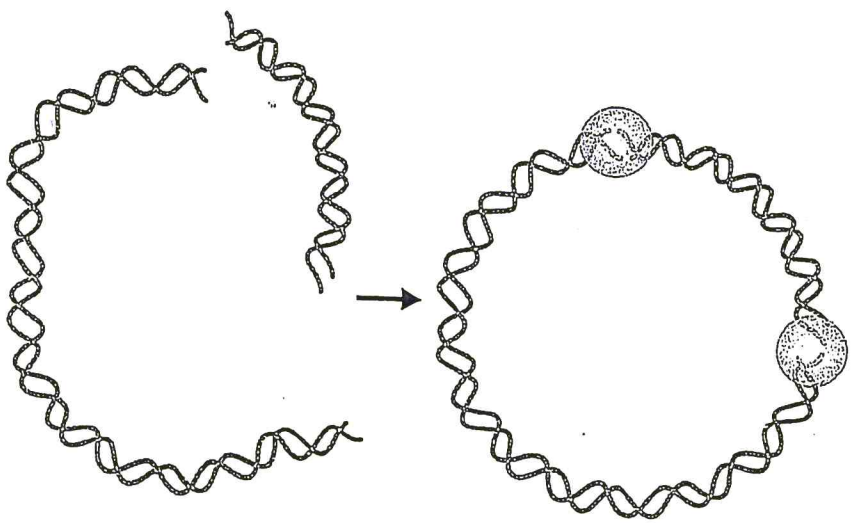
Gene cloning is not much different from what you do when you include a passage from Shakespeare in your wedding announcement. You open your massive compendium of the bard's plays and search through it, page by page, until you find the specific sequence of letters you desire: "Doubt that the stars are fire; Doubt that the sun doth move; Doubt truth to be a liar; But never doubt I love." You extract that passage, insert it into your announcement card, and make many copies of the card to send to friends and family. You have cloned a passage from *Hamlet*, act II, scene ii.

Because of remarkable technical advances of the 1970s, it's now almost as easy to find and copy genes as it is to find and copy passages from a book. The first step is to chop all the chromosomes into small pieces, which can be accomplished by adding to the chromosomes enzymes that cut DNA. Those enzymes don't cut the DNA just anywhere. They recognize specific short sequences of bases in DNA and cut wherever those sequences occur, producing a discrete set of fragments (see figure).



These enzymes are obtained from bacteria, where they provide defense against invaders such as viruses that infect bacteria by cutting up the viral DNA to prevent the viruses from commandeering the bacterial cell. Each species of bacteria has a unique set of these enzymes, and each enzyme cuts DNA at a different sequence of DNA base-pairs. For example, an enzyme in the common gut bacterium *Escherichia coli* cuts DNA wherever it finds the base sequence GAATTC on one of the strands; an enzyme from the bacterium *Hemophilus influenzae*, which causes pneumonia, cuts DNA wherever it finds the base sequence AAGCTT. Over three thousand of these enzymes have been characterized, and most are readily available, allowing gene hunters to divvy up the chromosomes into bite-sized pieces—basically dividing the book that is our genome into paragraphs.

Having cut up the genome, we need to separate all the fragments—there are millions of them—much as we separate the pieces of a jigsaw puzzle before starting to piece them back together. The fragments are first inserted, one by one, into a minichromosome, a small circular piece of DNA derived from a bacterial chromosome, using another kind of enzyme that joins two pieces of DNA, as an ironworker constructing a skyscraper joins two steel beams (see figure). All cells have such a DNA-joining enzyme because they constantly need to unite pieces of DNA to repair the damage that DNA continually incurs.



The minichromosomes are slipped back into bacteria, which act like little copying machines to make copies of the minichromosomes that carry a piece of a human chromosome. Each bacterial cell divides to produce a colony of billions of cells, each carrying a minichromosome containing a particular piece of the human genome. Among those millions of identical-looking colonies of bacteria, we need to identify the one that carries the minichromosome with the fragment that includes the gene we want. How can we do that?

We know the DNA sequence of the *IL2RG* gene because the base sequence—the precise order of A's, C's, G's, and T's—of the human genome has been determined, so the process now is simple: the base-pairing principle worked out by Watson and Crick means that any DNA strand will match up to its complementary partner strand through the specific plugs and sockets on each base. For instance, the sequence AGGCTAAC will match up to TCCGATTG because A pairs with T, and C pairs with G. So we can design and have synthesized a piece of DNA with a sequence of bases that matches up to some sequence in the *IL2RG* gene. Such synthetic DNA molecules can be ordered from providers of scientific supplies for not much more than you'd pay for a T-shirt. Because that piece of DNA will stick to the *IL2RG* gene, it can be used as a "probe" for the gene. We will attach to that probe a molecule that can be easily detected, like a compound that is fluorescent. When we spread some of this fluorescent DNA on top of the millions of bacterial colonies, it will find its mate only in the rare colony of bacterial cells that carries the *IL2RG* gene on its minichromosome, and the probe will literally "light up" that cell. We can then recover the cells of that glowing bacterial colony and retrieve the minichromosome from them, with not much more difficulty than we retrieve copies from the output tray of a Xerox machine.

The same enzymes that enable gene isolation spawned the biotechnology industry, today a worldwide enterprise that generates over $50 billion in yearly revenue. This new industry has provided such drugs as erythropoeitin (Epo) and granulocyte colony stimulating factor (G-CSF) for stimulating the growth of blood cells in patients who have undergone chemotherapy for cancer, and antibodies such as the one that fights breast cancer by inhibiting the Her2 protein, and insulin for diabetics, and Etanercept for treating disorders of the immune system such as arthritis

and psoriasis, and the blood-clotting Factor VIII for hemophiliacs, and many more.

In addition to enzymes that split apart and splice together DNA molecules, there are enzymes that can duplicate DNA sequences to make more copies of them, enzymes that can change one sequence to another, and enzymes that carry out some of the many steps required to manufacture pharmaceuticals. None of these enzymes was sought by biologists to create an industry. Rather, they were discovered—quite fortuitously—in scientists' quest to understand how bacteria fight infection, or how they replicate their DNA, or how they synthesize their proteins, and many other seemingly esoteric questions. Clearly, basic research is a good value.

Soon after he was born in Buckinghamshire, England in August 2003, Alexander Locke was diagnosed with the same disease that killed David Joseph and David Phillip Vetter. Alexander's parents, Carol and Colin Locke, like the Vetters, had no idea their firstborn son was at risk of having SCID. "We realised Alexander had a problem when his tummy button inexplicably failed to heal after birth, despite repeated courses of antibiotics. At four months, he developed a severe viral respiratory infection. He spent his first Christmas in hospital, attached to oxygen lines and antibiotic drips," Colin told Andrea Kon, a reporter for England's *Daily Telegraph*. "He had inherited his defective X-gene from me," said his mother, "and it was hard to accept that it was my 'fault.' I had no idea I carried a 'bad' X-gene."

Alexander was put in an isolator in London's Great Ormond Street Hospital for Children. It was more comfortable than David Phillip Vetter's bubble because the technology had improved in the intervening twenty years: Alexander had an entire room to romp around in. Alexander was protected by an airlock through which all his visitors had to pass in order to have the air around them cleaned and filtered. He was to live in the isolator while he waited for his doctors to identify a perfectly matched bone-marrow donor, something for which David Phillip Vetter waited for in vain for twelve years.

But Alexander spent only eight months in the isolator. Drs. Adrian Thrasher and Bobby Gaspar at Great Ormond Street Hospital were getting ready to test an experimental gene therapy for treatment of SCID, and when Alexander's bone-marrow transplant fell through (because the nearly perfectly matched donor carried a virus that would almost certainly have killed him), Alexander entered the gene therapy trial, along with four other boys with SCID.

Drs. Thrasher and Gaspar had developed a vehicle to deliver to Alexander's bone-marrow cells a good version of his defective gene. The vehicle was a virus that infects human cells. Viruses are ideal for this job because they are basically tiny Trojan horses that carry DNA within their protein coat. The viral DNA contains genes, just as our DNA does, and these genes code for the viral proteins that make up the protective coat and that make copies of the viral DNA. Although the viral DNA is minuscule compared to ours—most viruses have just a handful of genes and often only a few thousand DNA base-pairs—scientists have found places in this DNA where other genes, human genes, can be inserted.

The virus enters a cell and removes its coat, thereby delivering its DNA inside the cell. If it's a normal virus, the cargo is the viral chromosome with its genes that encode proteins to commandeer the cell's machinery to make more virus. Some viruses are aggressive, making many copies of themselves and killing their host cells in the process, releasing more viruses that go off to infect and kill other cells. Other viruses are relatively benign, incorporating their own DNA into a human chromosome while allowing the cell to live, lying in wait to check out the situation before deciding to make more virus. But if some of the viral genes are removed from its chromosome, the virus is disabled: it can deliver its DNA into cells, but that incomplete viral chromosome cannot take over the cell or produce more virus.

Drs. Thrasher and Gaspar spliced the *IL2RG* gene into the chromosome of a disabled virus that infects human cells. They made many copies of the engineered viral chromosome, packaged them into viral coats, and added the viruses to a test tube containing Alexander's bone-marrow cells. The viruses latched on to the marrow cells and quietly slipped into them, taking off their viral coats as they went in.

The viral DNA made its way to the cell's nucleus where it pasted itself along with the *IL2RG* gene into one of the human chromosomes. As the cells grew and divided they passed the viral DNA along with the good *IL2RG* gene on to other bone-marrow cells. The engineered cells were returned to Alexander's bloodstream, where they found their way back to his bone marrow.

Many processes have to go right for gene therapy to work, and scientists are still a long way from having a failure-proof procedure. For gene therapy to work, biologists need to deliver the viruses carrying the therapeutic genes to the appropriate cells, where the good gene can do its job. Discoveries of how cells specialize in certain tasks have led to improvements in this cell targeting. Even if the good gene gets to the right cells, it must get turned on at the right time and at the right level to provide cells with the right amount of the protein they are missing, when it is needed. Furthermore, expression of that gene must persist for long periods of time, so understanding how transcription factors and other proteins determine whether a gene is on or off is invaluable.

Drs. Thrasher and Gaspar waited to see if the engineered bone-marrow cells would give rise to the white blood cells Alexander needed to fight infections. "Alexander was allowed home for the first time in May, aged eight months. It was more complicated than having a newborn," Carol told Andrea Kon. "Every tube and piece of equipment needed sterilising. We had to use a spreadsheet to keep track of his medical regime. He had lost the ability to suck during the first days in intensive care and was being fed through a gastronasal tube. We were administering four drugs four times a day through the tube."

Despite their best efforts, Alexander contracted infections and had to be rushed back to Great Ormond Street Hospital. "The second time doctors fought a nine-hour battle for his life; we were so lucky. It's astounding he didn't suffer any long-term brain damage," Colin exclaimed.

Alexander's new bone-marrow cells grew and began to spawn competent white blood cells. He was allowed to venture out of his airlocked room for longer and longer periods of time. "He loved mixing with other children, although he had never played with a child until six months ago," his mother said. "We feared that his life as a 'bubble baby' might have left developmental or physical scars, but he's caught up in every way. The only treatment he needs now is a prophylactic course of antibiotics once a fortnight. Soon he'll go to primary school. That would have been unthinkable two years ago." The cure that David Phillip Vetter was waiting for finally arrived. It came in the form of a gene.

But good as it seems, the cure is far from perfect. The virus that delivers the *IL2RG* gene to bone-marrow cells of SCID sufferers can also deliver

something deadly: cancer. Four of eight boys in France who were cured of SCID by gene therapy traded it for another disease: leukemia. One has since died from it. The cancer occurred because the piece of viral DNA that carries the *IL2RG* gene, which usually inserts itself benignly into apparently nonfunctional regions of the genome far from any critical genes, landed in these boys' genomes near a gene encoding a protein that accelerates cell growth. The viral DNA caused this gene to be turned on, and the protein it made set those cells on the path to cancer.

The doctors had no way of knowing in advance that the virus would land near this gene, but as soon as they learned that it had, they stopped doing gene therapy with the virus while they searched for a way to prevent it from happening again. Is the prospect of a cure for their disease worth the risk of leukemia for these boys? We suspect that David Phillip Vetter would have said it is.

It is not surprising that the only real successes of gene therapy since David Phillip Vetter's death in 1984 have been with disorders, like SCID, that affect cells that doctors can easily get their hands on. Surgeons are good at harvesting bone-marrow cells from patients, and scientists are adept at growing them and modifying them in the laboratory. And it is easy to get the engineered cells to the place they are needed, because when they are reintroduced into the bloodstream they find their way back to the bone marrow, like salmon returning to their home river to spawn.

Other attempts at gene therapy have not been so successful. Dr. Ronald G. Crystal at the U.S. National Institutes of Health had a great idea for delivering to patients a good copy of the gene that is defective in people with cystic fibrosis. This gene encodes a protein that sits in the membrane of lung cells and allows salt to pass in and out. If the protein is defective and the salt balance is upset, then the layer of mucus that keeps germs out of the lungs becomes thick, providing an attractive breeding ground for infectious bacteria. The inflammation of the airway that results makes breathing difficult. Cystic fibrosis is a disease that usually leads to an early death.

Crystal reasoned that he could deliver the gene to patients' lung cells simply by having them inhale disabled viruses that carry the cystic fibrosis gene. The viruses would be sucked into the lungs, where they would attach

to cells and inject the functional gene. He used a relatively harmless virus that naturally infects the lungs and gives people mild cold symptoms. While the idea seems terrific, it didn't work because the lung cells are protected by an armor of mucus and cilia, little hairs that sweep away foreign particles that enter the lungs, which ended up blocking access of the viruses to the cells.

One gene therapy failure in 1999 was especially tragic. Jesse Gelsinger, at eighteen, suffered from a metabolic disorder caused by the lack of the ornithine transcarbamylase (OTC) enzyme, which is needed to prevent a toxic product of metabolism, ammonia, from accumulating in the blood. The accumulation can lead to brain damage, coma, even death. Jesse had a mild form of the disease, which he was able to keep under control with medication (he took thirty-two pills a day) and a low-protein diet. He knew gene therapy was unlikely to help him, but he was eager to try it because it promised to help those with more severe forms of the disease. Jesse told Sheryl Gay Stolberg, a reporter for the *New York Times*, "What's the worst that can happen to me? I die, and it's for the babies."

At 10:30 a.m. on Monday, September 13, 1999, a large dose of a virus carrying the *OTC* gene was injected into a vein that emptied into Jesse's liver. The plan was that it would deliver the gene to his liver cells, which would then make the enzyme he needed. We'll never know if that happened, because Jesse died four days later, the victim of a massive reaction of his immune system to the virus. His death cast a pall over gene therapy for several years.

Given its few successes and its several failures and tragedies, gene therapy has yet to live up to its much-ballyhooed potential. There are still enormous challenges in getting functional versions of genes into the right cells and, once there, getting them to produce an appropriate amount of the needed proteins for long periods of time.

And gene therapy brings enormous ethical questions. If we can change someone's personal DNA code by replacing a defective *IL2RG* or *OTC* gene in order to cure a disease, we could probably change a gene to make a child taller, or stronger, or have a lower level of cholesterol, or concentrate for longer periods of time. What are the limits on human characteristics that are permissible to alter? And what about making changes to the DNA code that would be passed down to future generations? Although there is general agreement that this type of gene therapy should never be attempted,

human history suggests we must remain vigilant. Scientists are nothing if not persistent and ingenious, and they have no lack of alternative strategies to someday bring gene therapy into standard practice. But the expectant public that has learned of the potential of gene therapy to relieve suffering from diseases, as well as scientists themselves, must be patient. For David Phillip Vetter the wait would have been twenty years for a possible cure; for other patients with other diseases, the wait will be longer. But the cures *will* come. Of that we are confident.