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

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 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 features read a chemical code of four different units, abbreviated as A, C, 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 these 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 began 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

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!

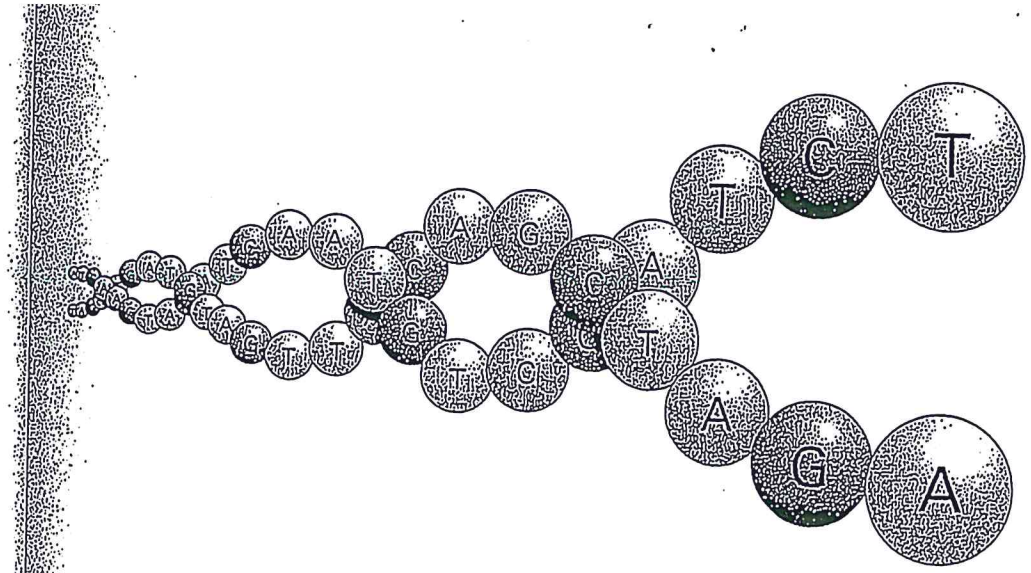
Asimov's many popular nonfiction works include *The Chemicals of Life* (1954), and *The Genetic Code* (1963). In the latter work he attempted to

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.

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.

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.

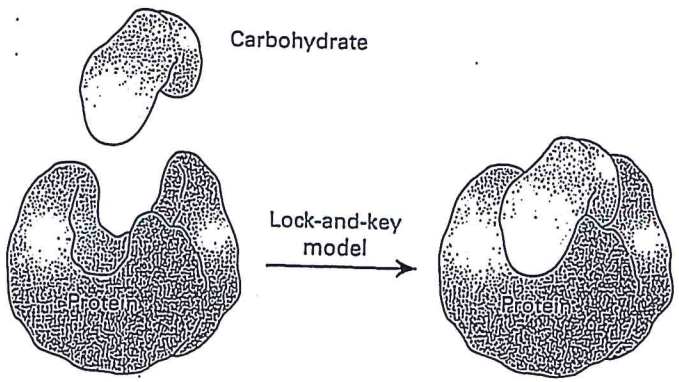
sequences: M-R, L-P-L, R-L-L, L-W, L-L, P-L, W-M-R, L-W-M, L-L-P (each letter is an abbreviation for one of the twenty amino acids). Knowing that these short sequences all come from the same longer sequence, you can line up the fragments:

```
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.

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

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.

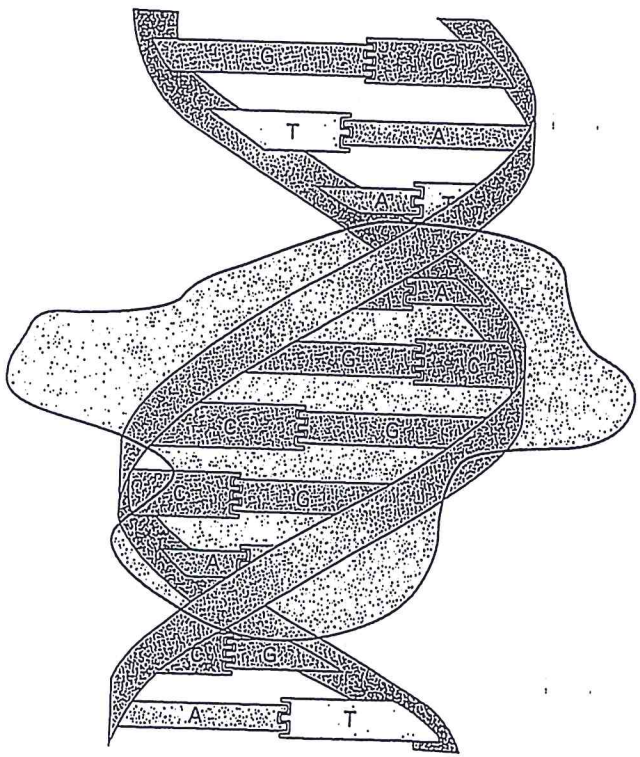
What was striking about Nüsslein-Volhard's approach was its simplicity: it required only a commercially available chemical to cause mutations in the flies, an ordinary microscope for observing the fly embryos, and standard genetic analysis—all of which were available as far back as 1930. Why did no one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

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.



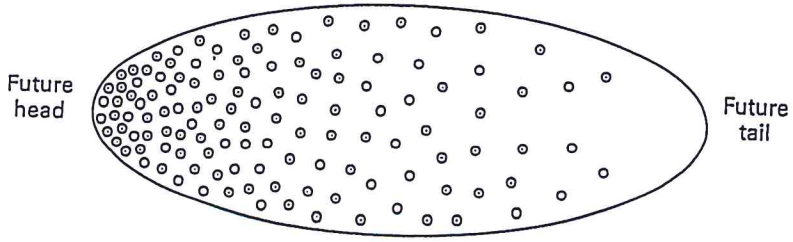
Most genes contain recognition sequences for several transcription factors. The sum of the effect of each transcription factor bound to the gene determines the state of the gene's switch. Some transcription factors act to turn transcription on, others strive to turn it off. The transcription factors 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

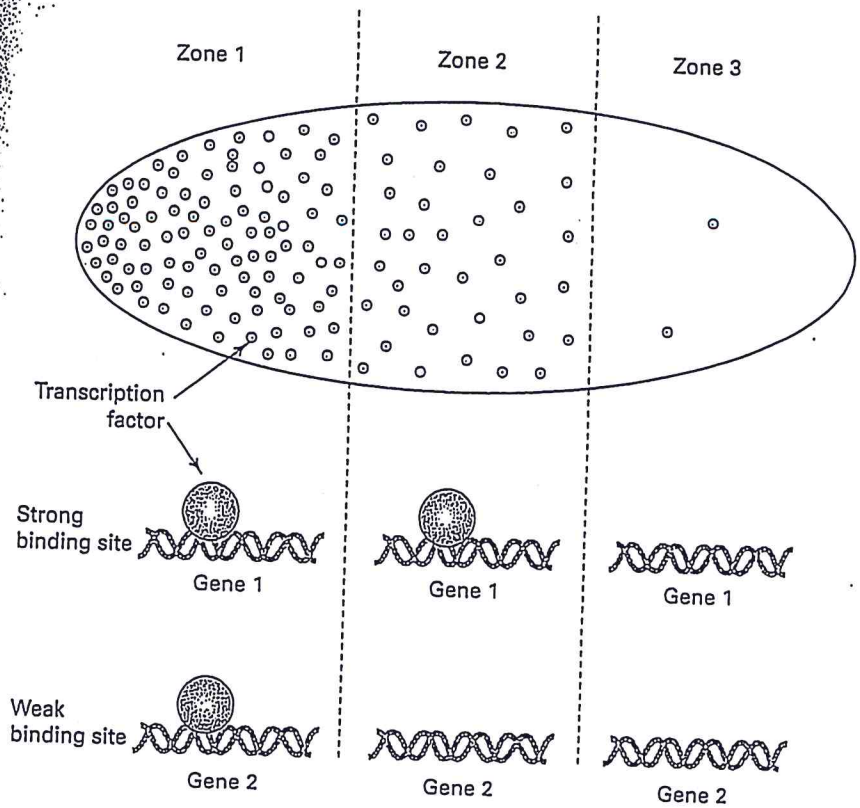


controlling the activity of genes in cells that lie near the end of the embryo destined to become the head, with decreasing effectiveness as its concentration diminishes toward the end of the egg that will form the tail (see figure).

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.

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

David Phillip Vetter, the Bubble Boy, lived a celebrated life that stimu-ated a hit song by Paul Simon, feature films starring John Travolta and ake Gyllenhaal, and an episode on *Seinfeld*. Celebrated, but tragic. The ournalist Steve McVicker described in a 1997 article in the *Houston Post* now 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.

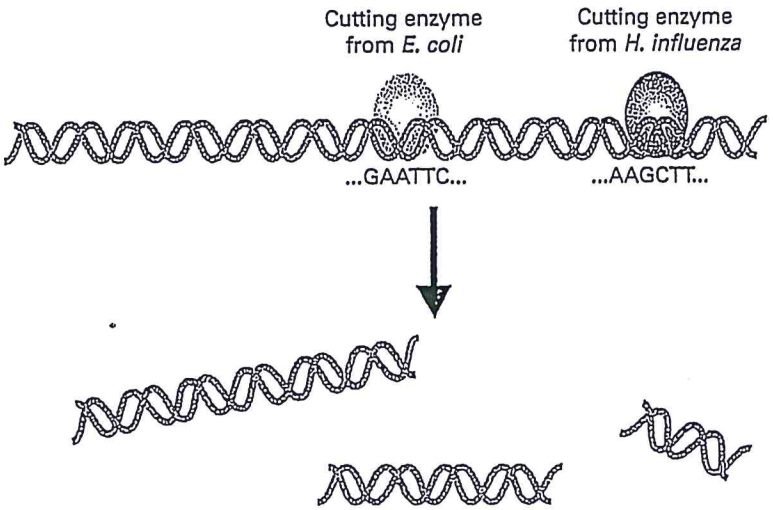
If a functional copy of the gene can be delivered to the bone-marrow cells of a SCID patient, those cells begin to make white blood cells competent to fight infections, giving the patient something he wasn't born with: a functional immune system. Treating disease with genes—gene therapy—is the brass 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).



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.

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.