13 A Family Affair: Mapping a Gene for ALS

Only two years after proclaiming himself "the luckiest man alive" at his tearful retirement from baseball at Yankee Stadium in 1939, Lou Gehrig succumbed to amyotrophic lateral sclerosis (ALS). Lou Gehrig's disease usually strikes in middle age, as it did Gehrig: the diagnosis of ALS that turned out to be a death sentence was handed to him on his thirty-sixth birthday. This was about a month after the first baseman took himself out of a game against the Detroit Tigers because he was no longer able to perform at the level he, his teammates, and fans had come to expect. That game ended his remarkable streak of 2,130 consecutive games played, a record that would stand for fifty-six years.

Gehrig could no longer perform at his accustomed level because he was losing control over his muscles. ALS causes its victims to slowly waste away: the nerves are increasingly unable to send signals to the muscles, first causing clumsiness but eventually progressing to near paralysis. Muscles far from the spinal cord are the first to be affected. A wave of paralysis then marches relentlessly up the arms and legs until it reaches the diaphragm, ultimately causing the victim to be unable to breathe, usually three to five years after the disease is diagnosed. Lou Gehrig's disease is indeed an awful one, for which there is no effective treatment and no cure.

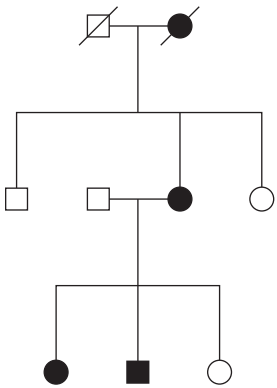
Andrew Mattingly Jackson was somewhat luckier than Lou Gehrig, if you can call someone who inherits a neuromuscular disease lucky, because he contracted a rare juvenile form of ALS known as type 4 ALS, which is much milder than Lou Gehrig's disease, though still serious. Jackson was a descendant of Thomas Mattingly, one of the first British colonists of the New World and the founder of a Maryland-based clan, more than seventy of whose members have suffered from this type of ALS. Jackson began wondering in his early teens why he couldn't run very fast and wasn't

more of an athlete, but he didn't think much more of it because in 1945, at eighteen, he was fit enough to pass a physical and join the U.S. Navy, and he made it through boot camp. But he couldn't ignore his diminished physical abilities for long, because one day two years into his service he found himself unable to climb the ladders onto his ship after a day of strenuous work. Like Lou Gehrig, Jackson took himself off duty, and was quickly diagnosed with ALS.

But unlike Lou Gehrig, Jackson continued to function as well as most people, and needed only a cane to get around until he was well into his fifties, when, as his strength waned, he started to use a walker. In fact, his disease was so mild that at one point physicians at Johns Hopkins University reversed his ALS diagnosis and told him that he had the much milder Charcot-Marie-Tooth disease. People with ALS4 can look forward to an increasingly difficult, but not foreshortened, life. Jackson worked for Black and Decker for forty years and, like Gehrig, retired only when he felt he could no longer perform at the highest level. At seventy-six, Jackson could still stand, albeit with the aid of a power wheelchair; could still type on his computer, albeit with the aid of a device to support his arm; and was still actively involved in his church. Unlike Lou Gehrig, Jackson inherited his disease.

Jackson traced his disease back to his great-great-great-great-great-grandparents Thomas and Elizabeth Mattingly, who left England and landed on the shores of Maryland in 1634. Jackson took great interest in his disease, and compiled an impressively complete pedigree of his clan—a record of his ancestors—spanning 350 years and eleven generations. A small section of the Mattingly-Jackson pedigree (see figure) shows three generations of the Maryland branch of the Mattingly family. The branch of the family that inhabits the western states—including, coincidentally, another great Yankee first baseman, Don Mattingly—is not afflicted with ALS. Perhaps when those Mattinglys trekked westward, the ones with the disease were unable to undertake or survive the strenuous journey.

To make sense of the Mattingly family pedigree we need to tell you the conventions used by geneticists: circles indicate females; squares indicate males; a horizontal line connecting a circle and square indicates a mating; the children that resulted appear at the ends of the vertical lines below the mating, from left to right in the order of their birth; a diagonal slash



through a symbol indicates someone no longer alive. Most important: Filled-in squares and circles indicate individuals with the disease.

What pattern of inheritance does this type of ALS follow? The mutation must be dominant, because those who have the disease pass it on to approximately half of their children, and only a single parent must have the mutation to result in children with the disease. The mutation is not on one of the sex chromosomes (X and Y) because the disease afflicts both males and females. Family members alive today continue to pass on to their children the mutation that Thomas or Elizabeth Mattingly brought with them to the Maryland shore three centuries ago.

Phillip Chance, a researcher at the University of Pennsylvania, learned of Jackson's extensive pedigree and realized he could use it to identify the defective gene that Thomas or Elizabeth Mattingly passed on to so many of their descendants. The inheritance of this defective gene is called Mendelian, because it follows Mendel's two simple laws of inheritance: (1) we have two copies of each gene (though the word "gene" wasn't coined until 25 years after Mendel's death), because our chromosomes come in pairs; (2) the chromosomes where genes lie (though Mendel had no idea what lay on chromosomes) assort randomly into sperm and eggs. Because the pattern of inheritance of ALS in the Mattingly family is straightforward, the pedigree that shows how the mutation is handed down from one generation to the next can be used to track down the defective gene responsible for the disease.

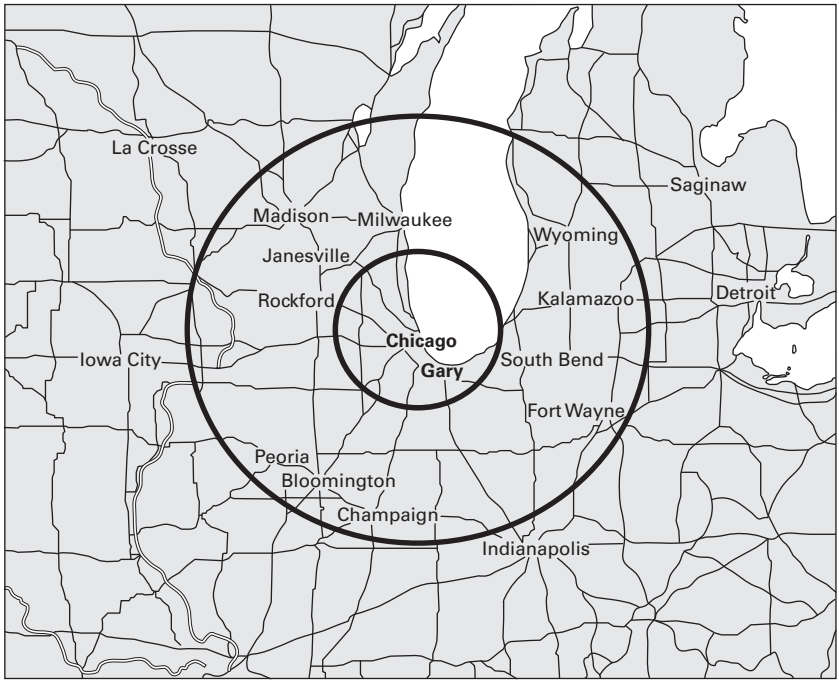
At a 1994 reunion of the Mattingly clan on Solomon's Island in the Chesapeake Bay, Andrew Mattingly Jackson saw to it that, between eating burgers, downing drinks, and heaving horseshoes, all family members had some of their blood drawn. He had the 107 blood samples delivered to Chance, by then at the University of Washington in Seattle, who extracted DNA from each sample and analyzed it to locate the defective gene.

How did Chance do this? While the process is laborious, the principle is straightforward. It's sort of like looking for a missing person. Imagine that Jimmy Bradford, a ten-year-old boy living in Marshfield, Wisconsin, is reported kidnapped by a gang of thugs who sped off in a late-model blue Chevy SUV, their current whereabouts unknown. The first thing the police do is confirm that Jimmy is not just hiding out at one of his favorite haunts: the home of his best friend, Curtis; the fishing hole he often frequents with his eccentric uncle Blake; or the game arcade in the nearby shopping mall that he visits almost daily. Finding no sign of him in any of those places, the police promptly initiate their search. They establish a five-state "Amber alert" with television and radio appeals to call a phone hotline. The next day they receive an anonymous tip: the car and boy have been spotted "about a hundred miles from Chicago," but the caller abruptly hangs up. The agent in charge interprets "about a hundred miles" to be somewhere between 50 and 150, and draws two circles on a map (see figure).

This is still a lot of terrain to cover, but it's much less ground to search than would have been the case if the kidnappers had driven for twenty-four hours in who knows what direction. The officers are pleased with their progress in locating Jimmy.

The police issue pleas to residents throughout this doughnut-shaped area to be on the lookout for the car or its occupants. Another caller to the hotline reports that she spotted a boy that fits Jimmy's description riding with two men in a blue SUV in Peoria, Illinois "not twenty minutes ago." The police quickly draw a new ring with a twenty-mile radius around Peoria and cordon off all the roads leading out of the city. With no means of escape, the thugs have to hunker down. Their SUV is spotted on Miller Lane, a quiet street in Peoria with twenty-five houses on each side. Officers go door to door, and at 106 Miller Lane they find Jimmy safe and sound and arrest his captors.

Phillip Chance and his colleagues hunted for the juvenile-onset ALS gene in a similar way. They scanned the chromosomes of the 107 family



members who donated blood, looking for clues that would lead them to the defective gene. The clues they were hoping to find are particular sequences of DNA base-pairs that vary in the human population and that tend to be found in people with the disease, but not in family members free of the disease. None of these DNA sequence variants causes the disease, but some are located close enough to the disease-causing gene to provide a clue that it is nearby, much like the blue SUV on Miller Lane indicated to the police that Jimmy was nearby.

What are these DNA sequence variants and how do they point to the defective gene? They are simply small sequence differences between individuals, a reflection of that 0.1 percent difference between any two individuals' personal DNA codes. These DNA sequence variants serve as markers along the chromosomes. They are basically of two types. The first type is a single-base-pair difference, where one family member has, say, an A at a given location on one of the strands of a chromosome and another family member has, say, a G there. The second type occurs when there are slight differences in the number of base-pairs at a given location

in the genome, where some family members have a few more and others have a few less base-pairs at that place on a chromosome. Most of these sequence variants lie in the 98 percent of the genome that doesn't code for a protein.

These DNA sequence differences arose by chance in the human population, most of them by about fifty thousand years ago, and were passed down through the generations to those living today (discussed in more detail in chapter 18). Most of the sequence variants are quite benign—only a small fraction of them have a consequence for anyone's health or appearance or anything else. Yet they serve a useful function as markers that allow geneticists to find their way along chromosomes. They are like signposts on a highway that announce the traveler's location. One of these signposts might tell scientists traveling down DNA: "There are 1,237,482 base-pairs to the end of this chromosome."

Identifying these DNA sequence variants used to be a formidable and expensive task, but no longer, because of impressive technological developments. It is now relatively easy and reasonably cheap to detect thousands, in fact, hundreds of thousands, of these DNA signposts in hundreds or even thousands of individuals. Many millions of these DNA sequence variants among individuals have been cataloged in the three-billion-base-pair human genome, so, like the billboards that infest many of our highways, there is a wealth of signposts along each chromosome.

How do these DNA signposts provide clues to the location of genes? One of these sequence variants, of course, is the one that causes ALS. The Mattingly family members with the disease have in their personal DNA code one particular sequence of base-pairs in the gene responsible for their ALS, which makes the gene defective. Those without the disease have a slightly different sequence at those same positions in the gene, which makes it function normally.

If the ALS mutation happened to be one of the DNA signposts that Chance scored in the Mattingly family members, he would have been able to go straight to the gene. But so far no gene hunter has been that lucky, and none are likely to be that lucky in the future. Gene hunters such as Chance don't even try to go straight to the sequence difference that actually causes a disease. Trying to find that one change in three billion base-pairs without clues would be like looking for the proverbial needle in a

very large haystack; the police would have an easier time looking for a ten-year-old boy in a five-state region without any clues.

Instead, Chance looked for those DNA signposts, clues that would point him to the neighborhood of the gene. His logic was as follows. The DNA a child receives from his parents is a mix of half the DNA from each of them; this DNA is one quarter the DNA that each of the four grandparents had; one eighth of the DNA that each of the eight great-grandparents had, and so on. So the further back in generational time you go there are progressively fewer of the DNA sequence variants of each ancestor in someone alive today.

Chance knew that the region of the chromosome that contains the gene responsible for Jackson's disease must be identical in *everyone* in the family with the disease, all through the generations, all the way back to Thomas and Elizabeth Mattingly, the founders of the clan. He knew this because everyone in the pedigree who has this rare mutation and therefore the disease undoubtedly got it from his or her parent, all the way back to Thomas and Elizabeth.

Here's the key point: Some of the DNA surrounding the mutant gene is inherited along with that gene by all ALS victims through each generation. This is because the relatively few generations that have ensued since the two Mattinglys came to the United States provided relatively few opportunities for the reshuffling hatchet to separate the nearby DNA from the disease gene and cause it to be rejoined to DNA from a relative without the disease. So all the DNA sequence variants among individuals in this nearby DNA are passed through the generations *along with* the sequence difference—the mutation—that actually causes the disease. These particular DNA sequence variants are linked to that mutation, just as the specific form of the *TAS* gene that is inherited can provide information about the linked *MET* gene. Sequence variants that are linked to the disease-causing mutation thus provide clues that the gene is nearby. They travel with the mutation, just like the blue SUV accompanied Jimmy and provided a clue that he and his kidnappers were nearby.

Find those DNA sequence variants, and you'll find yourself in the neighborhood of the gene. The closer one of these DNA signposts is to the disease gene, the more often it will accompany the mutation that causes the disease—the less likely it is that the reshuffling hatchet separated it

from the disease-causing mutation—and the better it will be at signaling that the disease gene is nearby.

But, just like the police first looked for Jimmy in his favorite haunts, before Chance and his colleagues began the substantial task of scanning all the chromosomes for DNA signposts associated with the disease, they first looked for them in a few places they had reason to suspect the gene might be found. They looked in the regions of chromosomes 21 and 2 that contain the *ALS1* and *ALS2* genes responsible for other forms of ALS. They looked in the regions of chromosomes 5 and 7 that contain genes responsible for a similar disease: spinal muscular atrophy. Not finding signposts pointing to the Mattingly gene in any of those places, they broadened their search and began to look for it in earnest throughout the genome.

Because they had no way of knowing on which chromosome the gene might reside, they initially cast a wide net, looking for clues throughout the genome. Since the location along the chromosomes of all these DNA signposts is known, the researchers were able to judiciously choose for their initial search a set of 150 DNA signposts that are scattered more or less evenly along each chromosome. As they got closer they could choose additional signposts in increasingly smaller regions of the genome as they tightened their noose around the culprit gene. None of these DNA signposts were likely to have anything to do with the ALS disease, but Chance knew that a few of them were bound to be situated close enough to the gene with the mutation that causes ALS to be passed along with it to the next generation.

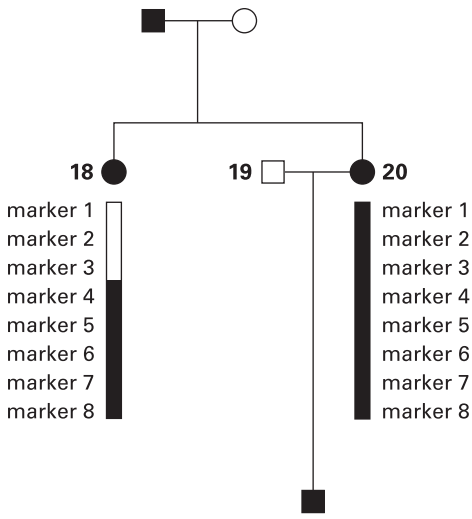
They scored all 107 individuals for the DNA base-pair present at each of these 150 different sites, spaced at intervals of approximately ten million base-pairs across the chromosomes, looking for those positions where a particular base-pair tended to occur in people with the disease but was less frequent in people who showed no signs of the disease. Two such DNA signposts went by the names D9S158 and D9S915. Although these names may seem uninformative, to Chance and his colleagues they were equivalent to the information in that first phone call to the police about Jimmy, because they pointed them to chromosome 9 as the hideout of the disease-causing gene.

They then went over chromosome 9 with the finer-toothed comb of additional DNA signposts, looking for more clues of where the gene lies. Each time they found a DNA signpost that appeared mostly in people with

the disease, it pointed them to a smaller region of the chromosome where the gene lurked. By 1998 they had limited the region where it could be to about five million base-pairs. They had tracked their prey to Peoria.

In case you're curious as to just how this gene-hunting works, we'll show you a little bit of Chance's data from his analysis of the Mattingly clan's pedigree, and work through the same logic that he and his colleagues used. If you find this whole business of walking along a chromosome too arcane, you can safely hop over this section and learn how the story ends.

The black squares and circles in the figure represent affected males and females, and the bars below two of these symbols represent a small piece of chromosome 9 surrounding the disease gene. The identities of the DNA base-pairs at different sites along the chromosome (we indicate eight of them in the figure) were determined by the researchers. For simplicity we call these sites "marker 1" through "marker 8," but Chance and coworkers knew them as D9S1831, D9S67, and so forth. The DNA base-pairs present at each site were compared to the base-pairs present at these sites in each individual's parents and four grandparents, if samples of all of them were available. Because these sites are variable from person to person, this information allowed Chance and his fellow researchers to infer the origin of each piece of the chromosome. That is, they could deduce which chromosome was maternal and which was paternal in origin; for the paternal



chromosome, which parts had come from the paternal grandmother and which from the paternal grandfather, and similarly for the maternal chromosome.

Two daughters, who are numbers 18 and 20 in this pedigree, are the children of a man affected with ALS. Because this father had the disease and Chance had been homing in on the relevant place in the genome where the gene must lie, the researcher knew that the man's chromosome 9 in this region harbors the mutant version of the sought-after gene. The father gave his younger daughter, number 20, this whole chunk of chromosome 9 (indicated as a solid black bar), including the ALS gene, and she contracted the disease, as indicated by the filled-in circle.

Sister 18 also inherited the disease gene from her father and she, too, suffered from ALS. But instead of inheriting this entire region of chromosome 9 intact, this daughter received a reshuffled version. In the sperm from her father that fertilized the egg that became her, the DNA on this chromosome reshuffled between markers 3 and 4. Everything south of marker 3 came from the paternal chromosome carrying the disease region; everything north of marker 4 came from the paternal chromosome with the good version of the gene.

From these results Chance could rule out the region, indicated in white for sister 18, encompassing markers 1, 2, and 3, as a possible location for the ALS gene. The ALS gene must reside south of marker 3. Continuing this process and searching for places where other reshuffling events had occurred, he could progressively delimit the piece of chromosome 9 where the gene could reside.

Chance and his colleagues continued this process over the next few years, looking for clues in increasingly smaller regions of the chromosome, until in 2002 they had it cornered between DNA sequence markers D9S149 and D9S1198, within a 500,000-base-pair region of chromosome 9. They were now on the very street that harbors the gene. And it was a small street with relatively few houses: this region of chromosome 9 contains only nineteen genes. They simply went from gene to gene, looking for a mutation—a DNA sequence difference—in one of them that is present only in people with the disease.

After two years of searching they found the mutation and captured the gene at last in 2004. The entire process took Chance and his colleagues

about ten years. They could now tell Andrew Mattingly Jackson and the rest of his extended family that the mutation one of his great-great-great-great-great-grandparents brought to these shores so long ago inactivates a protein that likely works to untangle molecules of DNA or the related molecule RNA.

You may be surprised to learn that more than two thousand genes responsible for inherited human conditions have been identified in this way, and more than sixty thousand different mutations have been catalogued in them. Why do we need to know which genes cause what disease? Does it matter? Will it help us prevent or treat the disease? The answer today is an inconclusive "maybe," but we expect that it will soon be an unequivocal "yes!" for many genes.

In the case of Andrew Mattingly Jackson's ALS, the bald truth is that identification of the responsible gene in his personal DNA code has so far provided little insight into the disease. The *ALS4* gene encodes a protein called senataxin, which is similar to other proteins that are known to unwind strands of RNA (ribonucleic acid) molecules. It's not clear which RNA molecules senataxin might work on, or why a defect in unwinding RNA could lead to neurological defects. Much more will need to be learned about senataxin before Phillip Chance's success in finding the gene can be translated into real help for families with ALS4. The same can be said for most of the other two thousand human disease–causing genes that have been identified. Although the road to discovery of disease genes gets easier to travel every day, the path to understanding the disease and developing treatments and, hopefully, cures remains arduous.

There are, however, some cases where knowledge of the gene has provided real insight into the disease. Identification of the gene responsible for cystic fibrosis revealed that it encodes a protein that pumps salt molecules out of cells, which explains the buildup of liquid in the lungs of CF patients. The nature of the CF protein suggested new ways to treat the disease, some of which are currently being developed.

The identification of genes involved in certain cancers enabled the development of some highly successful drugs that destroy cancer-causing proteins. Indeed, drug development increasingly is based on our ever-growing understanding of the fundamental biological processes that go awry and cause disease, an understanding that we often gain when disease genes are identified.

But still you may wonder: Why would a medical geneticist spend a good part of ten years of his life to track down a gene responsible for a minute number of ALS cases each year? The answer is that the genes implicated in rare genetic forms of a disease may reveal something about the much more common forms of the disease that are not inherited. Indeed, this has proved to be the case time and again. Genes found to lead to rare inherited cases of Alzheimer's disease, diabetes, cancer, heart disease, hypertension, and other illnesses have told us much about what goes awry in the common cases. Phillip Chance is hopeful that the *ALS4* gene will eventually provide similar insight into the more frequent forms of adult-onset ALS.

But even if fundamental insight into the disease does not prove to be forthcoming, pinpointing the gene and the specific mutations in it responsible for the disease enables doctors to determine whether one of us has the mutation. With that knowledge, genetic counselors can then inform us of our risk of developing the disease, counsel us on how to reduce that risk, and tell us the chances that we'll pass the defective gene on to our children. In the future, knowing the defective genes that cause disease may provide a cure through gene therapy: giving to cells a functional, "good" gene. The gene-finding business has gotten a lot easier for researchers because of the Human Genome Project, and that's good news for all of us.