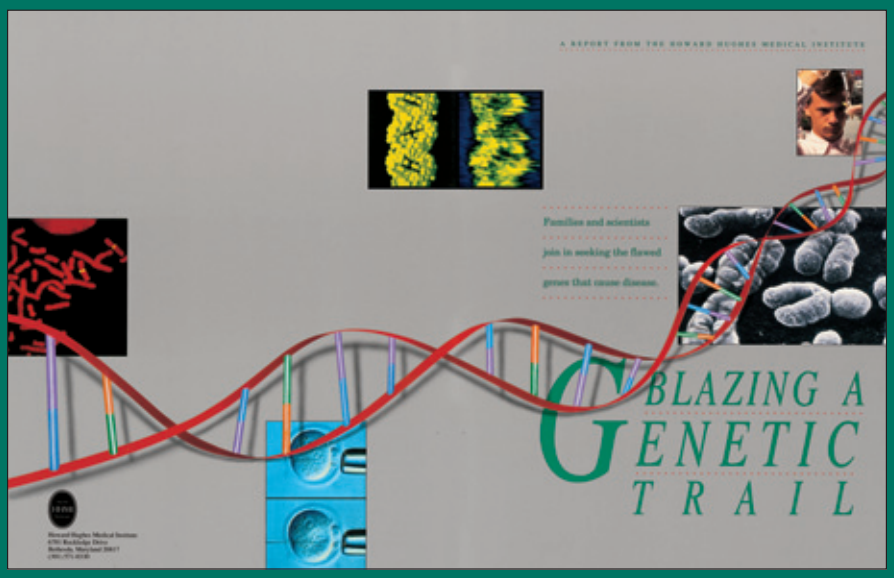
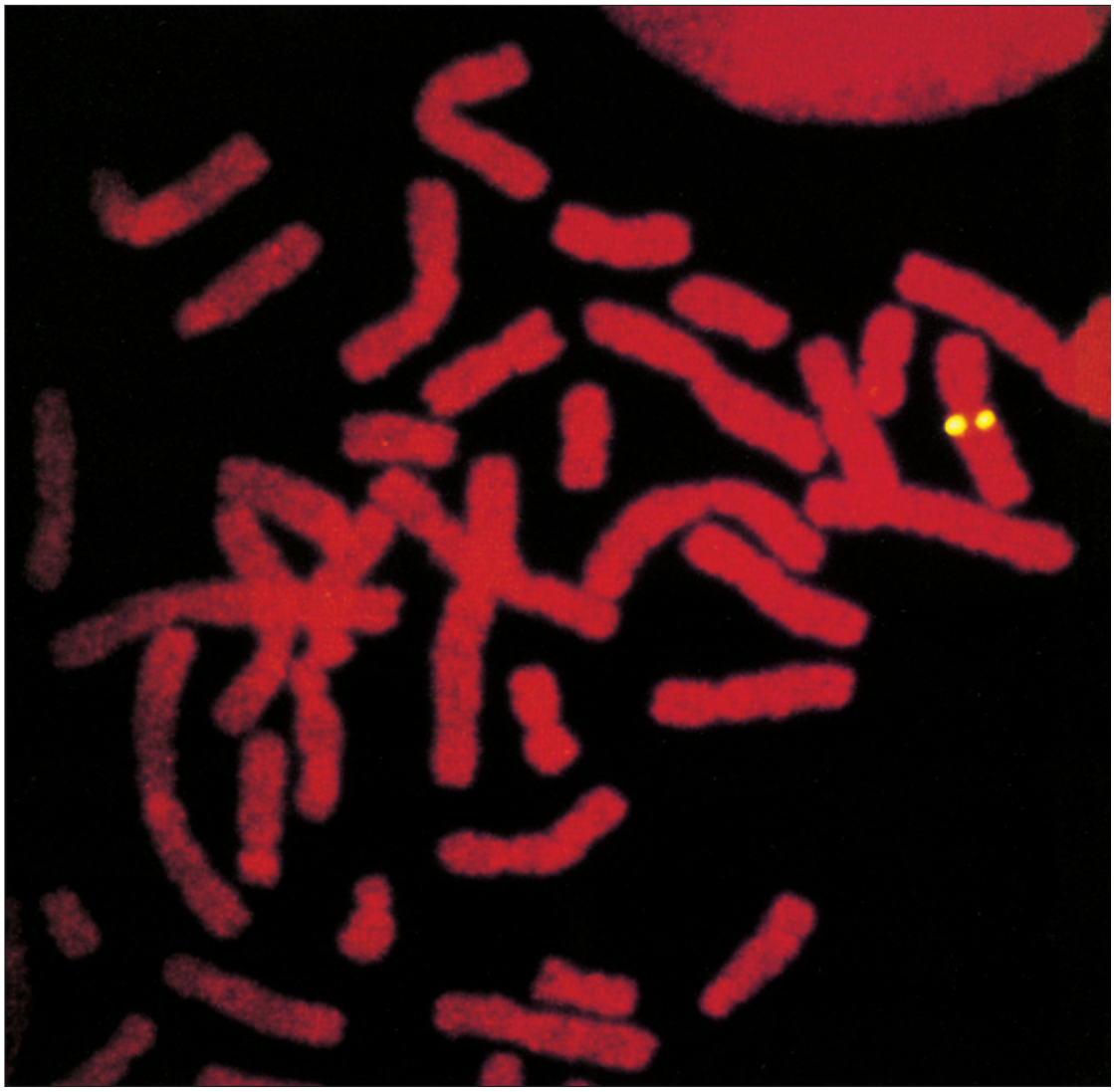
Blazing



A GENETIC TRAIL

Laboratory of David Ward, Yale University



Fluorescent yellow dots show the location of a human muscle gene on a pair of chromosomes in a human cell.

A REPORT FROM THE HOWARD HUGHES MEDICAL INSTITUTE

BLAZING A GENETIC TRAIL

Families and scientists join in seeking the flawed genes that cause disease.

Introduction *by Maya Pines* 4

Stalking a Lethal Gene *by Maya Pines*

A Gifted Young Patient Seeks His Own Genetic Flaw 6

"Walking" and "Jumping" Toward the Gene 12

Why So Many Errors in Our DNA? 17

Who Should Be Tested? 21

How To Conquer a Genetic Disease—a fold-out guide 25

In Search of Large Families *by Beverly Merz* 33

Reading the Human Blueprint 39

Of Mice and Men *by Sandra Blakeslee* 42

The Future of Genetic Research *by Harold M. Schmeck, Jr.* 48

Progress Continues 52

INTRODUCTION

"There's lots of zip in DNA-based biology today," declares James D. Watson, who, with Francis Crick, won a Nobel Prize in 1953 for discovering the structure of DNA. The speed of progress in understanding human genes since that time has been dizzying. Yet it is bound to accelerate even more in the new millennium, when genome projects reveal the sequence of all the genes that specify a human being.

The six feet of colorless, astonishingly thin DNA filaments that lie coiled inside each of the cells in our bodies contain all the genetic instructions necessary for our growth, development, and health. And now, for the first time, scientists are learning what these instructions spell out.

As researchers become fluent in DNA's language, however, they keep finding new copying errors in the precious texts. Every time a human cell divides, six billion base pairs of DNA must be reproduced in the correct sequence. Considering that this happens several billion times a day in an adult, our DNA is remarkably free of errors. But mistakes do occur. In some cases we are born with DNA typos that we inherited from our parents. Other defects arise anew inside our cells. Whatever their source, genetic flaws are common causes of pain, illness, and death.

This report focuses on the struggle to *conquer* inherited diseases. It sketches out a trail that, if completed, will lead to means of prevention or cure. To progress along this trail, scientists must answer three basic questions about each inherited disease. These questions and their answers—most of which depend on the use of recombinant DNA technology—are illustrated in a foldout guide that can also serve as a wall chart.

Easiest to analyze are disorders caused by errors in a single gene, such as cystic fibrosis or Duchenne muscular dystrophy. As shown in the story of Jeff Pinard, a gifted young man who was seeking his own mutation, scientists have now answered most of the early questions about cystic fibrosis: What faulty gene causes the disease? What kind of protein does the normal gene instruct the cell to make—in what quantities, and in what specific places? Does the flawed gene produce too little protein, the wrong kind of protein, or no protein at all? And, finally, what treatment will counteract this error? Scientists are still grappling with the problem of treatment, but they have considerable hope of success.

Although single-gene disorders are relatively rare, their clarity allows researchers to uncover mechanisms that may apply to other, more widespread diseases. For instance, as a result of identifying the gene defect that causes Duchenne muscular dystrophy, a muscle-wasting disease, scientists discovered a previously unknown protein that plays an important role in all muscle function. This gave them a richer view of how muscle cells work, enabled them to diagnose other muscle disorders with greater precision, and suggested new approaches to treatment.

Finding the causes of such complex ailments as heart disease, breast cancer, colon cancer, diabetes, arthritis, or schizophrenia presents much tougher challenges. Yet researchers have already identified a number of DNA errors that predispose people to developing these conditions, as well as some genetic variations that protect against them. Now scientists are zeroing in on the genetic pathways through which the flawed genes interact with other genes, or with the environment, to produce disease. Each pathway offers several possible targets for drugs, so this work raises the hope of developing new, individualized treatments for many widespread ills.

Thanks to the growth of powerful databases, researchers who identify a new gene can usually find a match for it right on their own computers by searching through a wealth of DNA sequences or proteins from animal models, as well as from humans. Additional databases are being set up to help identify the genetic variations that underlie disease and to speed the development of genetic tests and treatments. Every day, it seems, more pieces of the genetic puzzle fall into place.

Maya Pines, *Editor*

Under ultraviolet light in a cramped darkroom, Jeff Pinard wears a mask to protect his eyes as he examines how far DNA fragments (marked with a fluorescent dye) have moved through a gel. The fragments' size is revealed by how fast they move in an electric field.

STALKING A LETHAL GENE

A GIFTED YOUNG PATIENT SEEKS HIS OWN GENETIC FLAW

by Maya Pines

In the summer of 1992, 20-year-old Jeff Pinard set out to find the flaw in his genes that causes him to have cystic fibrosis.

He already knew quite a lot about genetic diseases, especially his own. Cystic fibrosis (CF) is a fatal disorder that clogs the lungs and other organs with a viscous, sticky mucus that interferes with breathing and digestion. It is the most common lethal inherited disease among white children and young adults, attacking about 30,000 Americans. Until recently, most patients died before reaching the age of 30.

Pinard adds fragments of a CF patient's DNA to a dye that will help track these fragments as they move through a porous gel.

But Pinard, a microbiology major at the University of Michigan, was full of hope and could hardly contain his excitement at the thought of working with top scientists at the cutting edge of research on CF.

A year earlier—in one of the triumphs of molecular genetics—a research team headed by Francis Collins of the HHMI unit at the University of Michigan and Lap-Chee Tsui and John Riordan of Toronto's Hospital for Sick Children had discovered the errant gene that is responsible for CF. The researchers also identified the specific mutation—a missing snippet of deoxyribonucleic acid (DNA), the genetic material packed in the chromosomes within the body's cells—involved in most cases of CF.

As many as 1 in 25 Americans of northern European descent—some 10 million people—carry a gene with a CF-causing defect. Because CF is a recessive disorder, people who inherit a defective gene from only one parent remain healthy. Those unlucky enough to inherit a defective gene from both parents develop the disease (see p. 22). But in some patients the specific defects that cause CF differ from the one originally found by the research team. This was the case with Jeff Pinard.

When Pinard was a child, his parents could not understand what ailed him—why he had two serious bouts of pneumonia before he was six years old and why he was in and out of hospitals all the time. No one else in their family had had such problems. They were especially worried about their son's inability to gain weight.

He carefully inserts DNA mixed with dye into a well at the starting end of the gel. A different well is used for each patient.

"I was emaciated," Pinard says. "I weighed only 30 pounds at age six."

They questioned their family doctor, who blamed Pinard's mother for "not feeding me properly," Pinard recalls. "It wasn't until my little sister nearly died of dehydration at one and a half that a different doctor, in Grand Rapids, diagnosed her—she has cystic fibrosis, too—and asked about others in the family." A test that measures salt content in sweat established that Jeff Pinard also had CF.

Thunderstruck—and without any information about what the disease meant—his parents went to the library to look it up. They found "just a little paragraph in an encyclopedia that said, 'children with cystic fibrosis die by the age of six,' " Pinard says. "It was devastating for them. But in spite of what they read, I got better and I lived."

In fact, Pinard is so full of life it is difficult to remember that he constantly faces the threat of a fatal flare-up of his disease. He shares an apartment near campus with another student. He often plays racquetball with his friends. He rides a bicycle, plays the trumpet. He has been singing in choirs since he was in sixth grade, and recently he toured the East Coast for two weeks with the University of Michigan Men's Glee Club, which gave him a chance to see the ocean for the first time.

Most of his waking hours, though, are focused on genetic research. As far back as he can remember, he has always been interested in science, he says. He turned to genetics in high school, where he wrote a 40-page report on the subject and became

increasingly engrossed in it. When he entered the University of Michigan as a freshman and heard about the CF research going on there, he asked if he might do volunteer work in the lab. Soon he proved so useful that the Cystic Fibrosis Foundation gave him a small grant to work in the lab part time during his sophomore year. And last summer he received a larger grant so that he could work full time.

Pinard's presence "has had a fairly profound effect on my lab," said Francis Collins. "He's all enthusiasm, good humor, smarts. Many people in the lab had no direct contact with anyone who had the disease before. Now it's not so easy for them to see their work as just an intellectual exercise. They're much more driven by the urgency of it."

Every time Pinard coughs—often in mid-sentence—he brings a sobering reminder of the fragility of his life and that of other CF patients. At mealtimes, he must take five pills of pancreatic enzymes so he can digest his food. The pills cost $20 a day. Pinard remembers that "I had terrible pains and stomach troubles when I was little—really bad—until they found out I had CF and gave me these pills. Now I take them every single time I eat"—even when he eats just a single M & M.

He suffers from severe bacterial infections in his lungs several times a year. These are treated with antibiotics, but Pinard hates to take the drugs because

then he can't stay out in the sun (they cause photosensitivity) and there is a chance that the antibiotics may affect his liver. While 95 percent of CF patients die from lung complications and heart failure—"the heart works so hard and has so little air coming in," Pinard explains—the other 5 percent die of liver damage. "So it's kind of a tightrope," he says. He adds, "My medications run about $10,000 a year when I'm healthy."

To keep healthy, he must be pounded on the chest for an hour every day to loosen the mucus in his lungs. "There are eleven positions for the pounding," he says, "including the lower rib cage and the back. Since I moved here, I've been doing it myself, but it's not nearly so effective as having someone else do it. My arms get extremely tired, and I can't do my own back." He points out that most other patients with CF have to have their chests pounded two or three times a day. "And in bad cases, they need it four times a day, a full hour each time, just to get rid of that awful stuff—that mucus that lodges there, like a thick web," he says. "You've got to jog it loose."

"I've had many friends die of the disease," he adds quietly. "I used to go to a CF camp in Battle Creek, and I always felt guilty because I was healthy while others had it really bad."

Pinard attributes his relatively good health to the unusual mutation in one of his CF genes. His DNA was among the first to be tested last year, after the team isolated the gene. The researchers extracted his DNA from the nuclei of white cells in his blood and rapidly found the known mutation—called Delta-F508—on one of his CF genes. But his other CF gene contained a yet-unknown, probably milder, defect.

"My little sister seems to have exactly the same type of disease as I do," says Pinard. "It'll be very interesting for me to see how my mutation correlates with the severity of the disease—I can't wait to find out. One thing I hope is that the mutation we have will apply to a lot of other patients. That would cut down on the amount of tests that are needed to identify carriers. If there really are over 100 different CF mutations, it's going to be a bear! It'll be very expensive and a lot of work and harder to find treatments."

Last summer Pinard worked 40 to 50 hours a week in the lab, hunting for his own mutation and those of other atypical patients. He painstakingly analyzed hundreds of DNA fragments but could not locate his own flaw or any other unusual mutation. Then, just as he was examining some DNA that looked interesting, he fell sick and had to be hospitalized for ten days with a severe infection.

When he came out of the hospital, he learned that while he was away the lab had found two new mutations, from two different patients, in the very batch of DNA he had been working with. "It was so frustrating," he says. "I was happy for them, but wished I could have reaped the fruit of all that labor!" Meanwhile the mutation in his own gene remains a mystery. Pinard is still looking for it as he continues working in the lab on a part-time basis during the school year.

"It's really crucial to find the other mutations," says Collins. "Every new mutation tells us something more about the gene: which regions of it are essential for its function, specific details we need in order to develop effective treatments."

Some 85 groups of scientists in over 20 different countries are now collaborating in the search for new CF mutations through a consortium organized by Lap-Chee Tsui in Toronto. They have identified more than 75 new mutations, but most of these affect only a very small number of patients.

The scientists correspond by fax machine and share their findings months before their papers are published in journals. They understand the need for speed. As Collins puts it, "I fervently hope we'll find a cure in time to help someone like Jeff."

"His presence has had a profound effect on my lab. . ."

Postdoctoral fellow Theresa Strong gives Pinard some pointers as they look for new mutations in the CF gene. They are comparing the sequence of bases in patients' DNA with the sequence of the normal gene.

Midwives used to lick the forehead of infants they delivered, believing that if the newborn's skin tasted abnormally salty it was a bad omen—the baby would soon grow sickly and probably die.

Their predictions were often accurate, for excessively salty sweat is a symptom of cystic fibrosis. Until recently, not much else was known about the deadly disease, even though doctors kept their patients alive longer by using antibiotics.

Since CF seemed to run in families, several teams of researchers tried for years to find the gene (or genes) responsible for it. But until 1978, the only way to identify which of the 50,000 to 100,000 human genes caused a disease was to find a physical clue: an abnormal chromosome that could be seen under the microscope, or a protein (the product of a gene) that was defective or missing in blood, urine, or body cells. Cystic fibrosis offered no such clues, and all attempts to find out which of the 23 pairs of chromosomes in human cells harbored the faulty gene hit a dead end.

In 1978, two scientists at the University of California, San Francisco—Y. W. Kan, who is now an HHMI investigator, and Andrée Dozy—discovered a harmless variation in DNA that was inherited together with sickle-cell disease, a blood disorder, within certain families. The variation could be used to detect the disease prenatally in such families. It also pointed the way to the guilty gene.

Since then, researchers have detected so many other variations in DNA that, in any family, a given gene defect is likely to be linked to one. This makes it possible to use what has come to be called "positional cloning." Instead of starting with a protein and then looking for the gene that produced it, scientists do the reverse. They find the general location of the gene for a specific trait by comparing the DNA of people who have the trait with that of close relatives who don't. Eventually, if the researchers examine the right DNA fragments, they zero in on the gene. Then they can deduce what protein the gene codes for and what flaw in the protein is responsible for the disease.

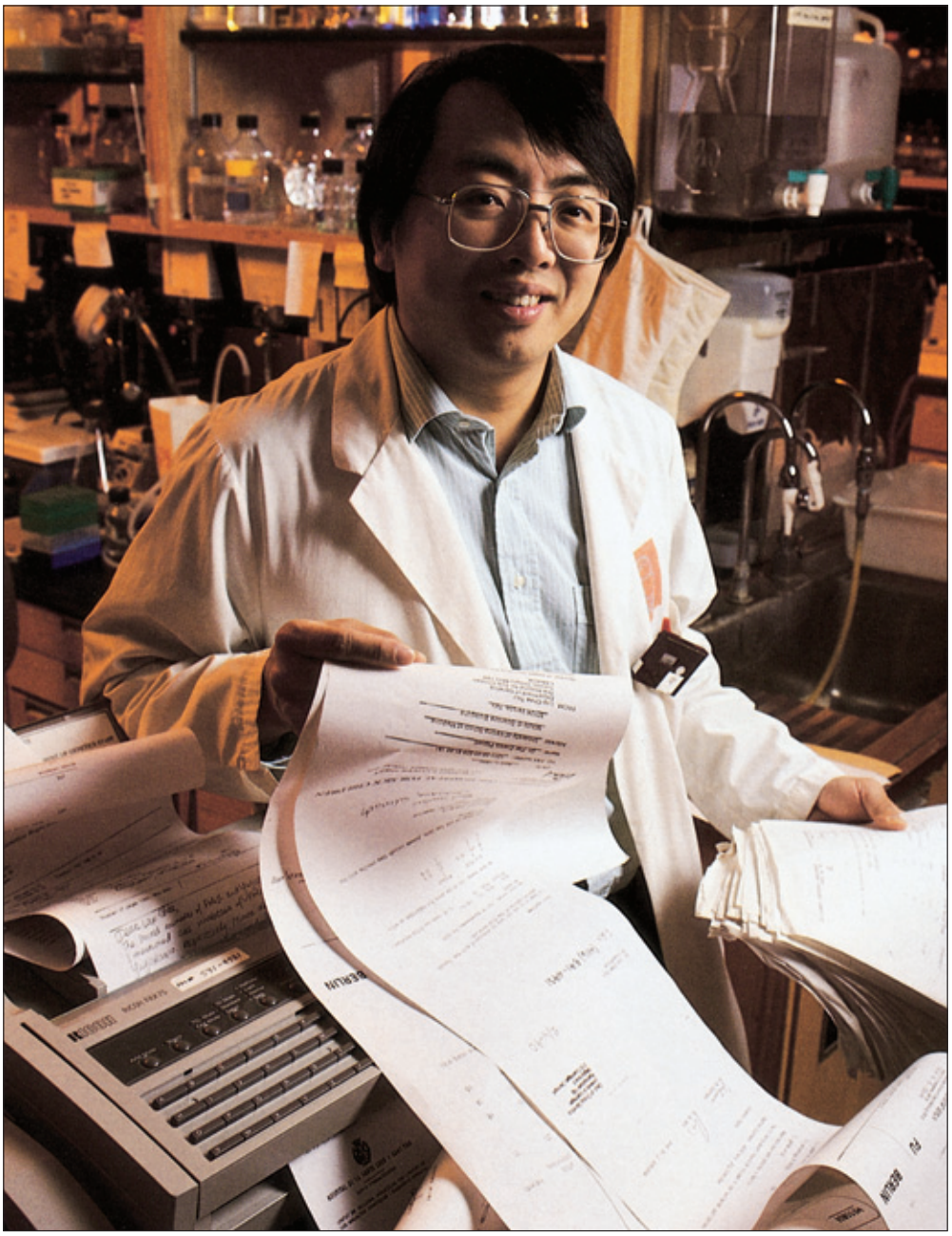
Using this approach, in 1982 Lap-Chee Tsui (pronounced "Choy") of Toronto's Hospital for Sick Children and his colleague Manuel Buchwald started looking for statistical links between CF and various genetic markers—recognizable variations in DNA that serve as molecular landmarks. The more consistently a marker is linked to a disease within a family, the more likely that the disease-causing gene is inherited together with the marker and, therefore, located close to it on the same chromosome.

Tsui contacted a large number of Canadian families whose children had CF and obtained samples of blood from more than 50 families in which at least two children had the fatal disease. He identified many previously uncharted variations in human DNA and looked for evidence that one of these was inherited more frequently by CF patients than by the general population. But for two years he did not find even a hint of linkage between any of these markers and his patients.

So when Tsui noticed a short report in an international journal from a group of researchers in Copenhagen who thought they saw a possible link between CF and one of their markers, he was very excited. The Danish group noted that they needed more families to confirm their link.

"I called them up at once and offered them my families," says Tsui. "By then they

"Walking" and "Jump i



In Toronto, Lap-Chee Tsui analyzes a stream of faxed reports from CF researchers around the world and coordinates the search for new mutations.

ing" Toward the Gene

had already excluded that marker, but they had found another one. They applied our families to it and established a linkage."

Tsui's report aroused the interest of Collaborative Research, Inc., a biotechnology firm in Massachusetts. The company offered Tsui a deal: It would provide him with probes—labeled fragments of single-stranded DNA that search out and identify complementary DNA fragments—for some 200 additional markers that its researchers had developed, if Tsui would provide samples of DNA from the white blood cells of his CF families. Tsui accepted. Within three weeks, he found a definite link between one of the new markers and the patients' DNA.

Soon afterwards, Collaborative Research mapped this marker to chromosome 7. It was a key finding, for it meant that the CF gene, which was inherited along with the marker, was also located somewhere on that chromosome.

Meanwhile, two other research teams temporarily overtook Tsui. Raymond White, a pioneer in gene mapping who had been hunting for the CF gene at the University of Utah's HHMI unit, discovered a close link between CF and *met*, a cancer-causing gene that had been identified by scientists at the National Cancer Institute in Bethesda, Maryland. Robert Williamson of St. Mary's Hospital Medical School in London, one of the first researchers to seek the CF gene by studying affected families, discovered another close marker on chromosome 7 called J3.11. The findings of the three teams were published together in November 1985.

Up to that point, looking for the CF gene had been much like trying to find a particular house without any idea of its address, or even what continent it was on. Now researchers had a place to start. In fact, they could narrow their search to an area that was equivalent to a particular country, rather than a whole continent, since White's and Williamson's markers gave them a general location on chromosome 7: its long arm.

As a result, parents who already had a child with CF and were expecting

A Brief Key To Basic Genetics

A human cell. *Each of the 100 trillion cells in the human body (except red blood cells) contains the entire human genome—all the genetic information necessary to build a human being. This information is encoded in 6 billion base pairs of DNA (3 billion base pairs from each parent). Egg and sperm cells each contain 3 billion base pairs of DNA.*

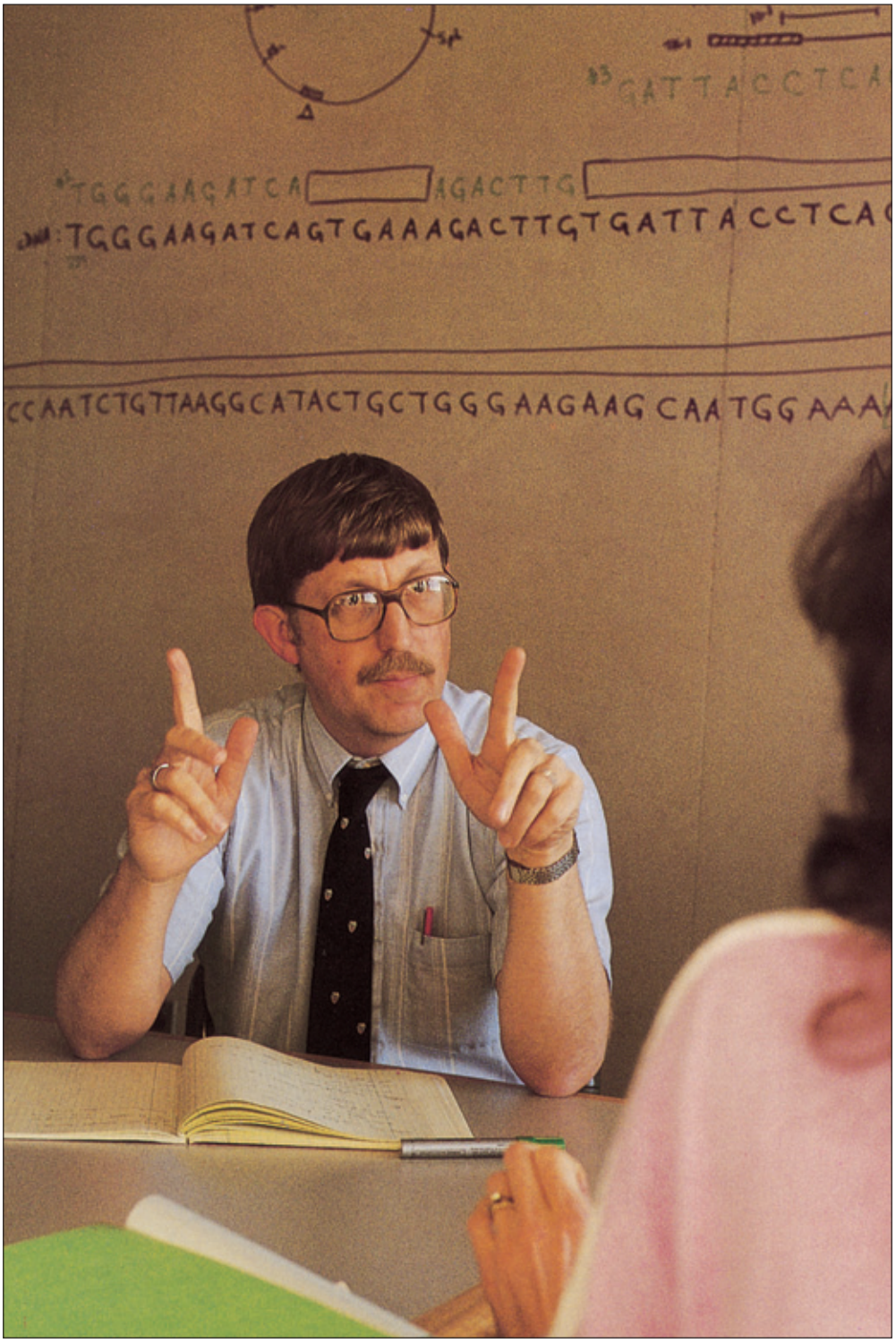
The cell nucleus. *Inside the cell nucleus, 6 feet of DNA are packaged into 23 pairs of chromosomes (one chromosome in each pair coming from each parent).*

A chromosome. *Each of the 46 human chromosomes contains the DNA for thousands of individual genes, the units of heredity.*

A gene. *Each gene is a segment of double-stranded DNA that holds the recipe for making a specific molecule, usually a protein. These recipes are spelled out in varying sequences of the four chemical bases in DNA: adenine (A), thymine (T), guanine (G), and cytosine (C). The bases form interlocking pairs that can fit together in only one way: A pairs with T; G pairs with C.*

A protein. *Proteins, which are made up of amino acids, are the body's workhorses—essential components of all organs and chemical activities. Their function depends on their shapes, which are determined by the 50,000 to 100,000 genes in the cell nucleus.*

Francis Collins explains how a normal CF gene was pieced together in the lab out of smaller, overlapping fragments of DNA. Part of the gene's sequence is shown on the blackboard.



It was a time for ingenuity—and fierce competition.

another child could be told with reasonable accuracy whether the fetus was destined to develop the disease. The verdict depended on whether the DNA extracted from fetal cells carried the same markers as the stricken child's DNA, showing that both parents had contributed a chromosome 7 harboring a CF mutation.

Finding the gene was still a long way off, however. The researchers did not even know what direction to take—was the gene to the left or right of the two markers, or in between? The only thing they could be sure of was that the three bits of DNA—the gene and the two markers—were quite close together on the chromosome. Therefore, determining the positions of the three pieces relative to one another, a statistically difficult task, would require DNA from an extremely large number of families in which some members had the disease while others did not.

Seven independent research teams had been racing to find the CF gene. Now they saw that none of them could do so large a study alone. At a meeting in Toronto, they decided to pool their families (211 in all), their probes, and their data.

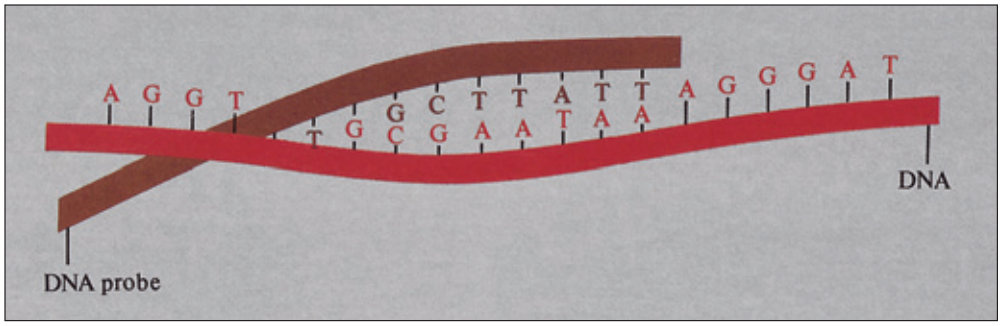
"That was a really lucky break for CF research," says Tsui, "because in December 1986 Jean-Marc Lalouel (of the HHMI unit at Utah) and Ray White found that the two markers flanked the gene."

The search for the CF gene could then focus on a well-defined stretch of DNA—equivalent to an interstate highway, rather than an entire country. The scientists rejoiced. Nevertheless, the piece of DNA between the two markers was still very long—a whopping 1.6 million subunits, or base pairs—long enough to hold perhaps 50 to 100 genes. It was far too great a distance for researchers to "walk" toward the gene.

For geneticists, walking along a chromosome means using partly overlapping DNA fragments to come closer and closer to the target gene, one step at a time, checking each fragment to see whether it is inherited together with the disease. But it is a slow and arduous task. The steps are very small and the path is strewn with roadblocks—repetitive sequences of DNA or other stretches that are difficult to cross. Tsui estimates that walking along 1.6 million bases would take about 18 years for the average lab.

They needed a shortcut. It was a time for imagination and ingenuity—as well as fierce competition. All thoughts of further collaboration vanished overnight as the seven research groups scrambled to be the first to find the CF gene.

At the beginning, Robert Williamson's team in London seemed far ahead of the pack. They had focused on a potentially interesting area 500,000 base pairs from the *met* oncogene, Williamson reported in April 1987, and found a "strong candidate" for the CF gene. "We really thought we had it," he said later. His report was so convincing that Raymond White and several other scientists dropped out of the race. But Williamson's group soon noticed some discrepancies. When they sequenced their candidate gene (that is, determined the sequence of bases in its DNA) and analyzed it to find out what kind of protein it coded for, the protein's profile looked wrong. The symptoms of CF suggested that the guilty protein would lie in or near cell membranes.



A probe—a strip of radioactively labeled DNA (black)—seeks out a complementary fragment of DNA and sticks to it.

Why So Many Errors In Our DNA?

As scientists learn to read the instructions in our genes, they are discovering that much of our DNA is riddled with errors.

Fortunately, most of these errors are harmless. Considering the difficulties involved—the 6 feet of DNA in a human cell consists of 6 billion subunits, or base pairs, coiled and tightly packed into 46 chromosomes, all of which must be duplicated every time a cell divides—our general state of health is something of a miracle.

We each inherit hundreds of genetic mutations from our parents, as they did from their forebears. In addition, billions of mutations occur in our DNA during our lifetime, either through mistakes in copying or through exposure to the environment. Bits of our DNA may be deleted, inserted, broken, or substituted. Most of these changes are harmless, usually because they affect parts of DNA that do not contain a gene's instructions. If they happen to affect genes that control cell growth, however, they may lead to cancer.

Problems arise only when an error in DNA alters a message that tells certain cells to manufacture a particular protein. Such messages are spelled out in varying sequences of the four chemical bases that make up DNA: adenine (A), thymine (T), guanine (G), and cytosine (C).

To stay alive and functioning, the human body requires a daily crop of billions of fresh protein molecules—about 50,000 different kinds of proteins that must be supplied in the right quantities, at the right times, and in the right places. We need hemoglobin to carry oxygen through the bloodstream, antibodies to fight foreign substances, hormones to deal with stress, neurotransmitters to evoke movements, emotions, and thought, and many other proteins to give structure to organs or speed up chemical reactions.

Our cells are kept extremely busy linking together amino acids—the building blocks of proteins—in the right order to produce these diverse proteins. The order is determined by the genes. According to the genetic code, which was deciphered in the 1960s, each triplet of bases in the genes' instructions either calls for a particular amino acid or gives a signal to start or stop making a protein.

An error in just one base can bring the wrong amino acid, altering the protein. And should one or two bases be missing, each succeeding triplet will be read in the wrong combination; such "reading-frame shifts" generally prevent cells from making the protein at all.

Actually the DNA's instructions are not transmitted directly; a copy made of ribonucleic acid (RNA) acts as an intermediary. The original DNA remains safely in the nucleus, somewhat like the printing block in a printing press, while the RNA copy is produced by transcribing just one strand of DNA, which carries the instructions for manufacturing protein.

Reading the DNA of humans and other mammals is complicated by the astonishing fact—discovered a little over a decade ago—that the genes' instructions are split into separate segments of DNA. These instructions must be spliced together before they can be carried out by a cell. Only about 5 percent of the DNA in mammalian genes actually contains the recipe for making a protein. The remaining 95 percent consists of intervening sequences, or "introns," whose function is unknown.

Splicing together the "exons"—the protein-coding sequences—is a very delicate, precise operation that involves snipping out the introns to end up with a much shorter strand of potent RNA. At the exon-intron boundaries are splicing signals, which researchers can now identify. Several genetic diseases have been traced to disrupted splicing.

Much of the recent progress in reading DNA has come from analyses of genetic errors.

James Wilson treats a CF patient's cells, shown growing in petri dishes, by inserting a virus that contains normal copies of the CF gene. The normal gene corrects the genetic defect within 24 hours.

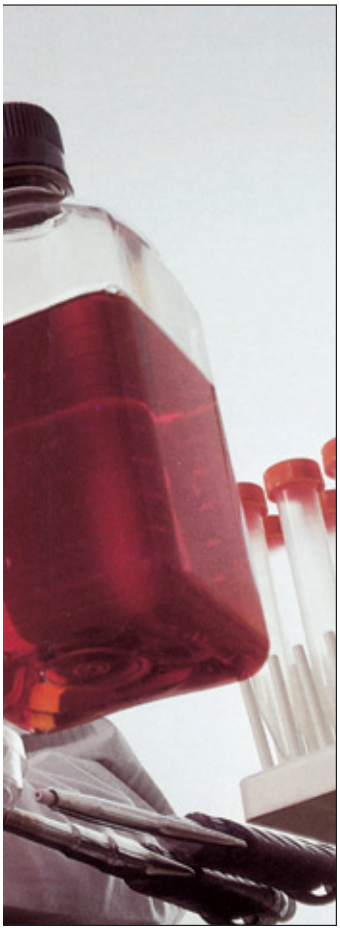
Yet this protein did not resemble other membrane proteins. A few months after his announcement, Williamson regretfully admitted that he had made an error.

In Toronto, meanwhile, Tsui had continued his dogged hunt for the gene. His strategy was to bombard chromosome 7 with a large number of additional markers, which he created by cutting up and analyzing thousands of fragments of normal DNA from chromosome 7 "libraries." If he used enough markers, he reasoned, one of them was bound to be close to the CF gene. He referred to his method as "saturation mapping," or brute force.

When Williamson announced his candidate gene, "we didn't stop," Tsui says. "We figured that the basic defect in CF is really very subtle, since the lungs and pancreas of patients take years to deteriorate. It had to be a very small error—just one or a few base pairs. We thought it would be difficult for Williamson's group to find this error, and perhaps we would succeed."

Eventually he hit on a marker that was closer to the CF gene than any other except for Williamson's. But he still needed to get from this marker to the gene. Walking was still too slow. So he contacted Francis Collins of the University of Michigan, who had devised an imaginative technique for "jumping" along chromosomes.

Jumping is five to ten times faster than walking. It allows researchers to cover 100,000 to 200,000 bases of DNA at one time and simply leap over areas that might otherwise be difficult to cross. The technique involves snipping out a long segment



of the DNA under study—a segment that is labeled at one end—and letting it curl into a circle. This brings the labeled end next to a sequence of DNA that would otherwise be thousands of bases away. The circular segment of DNA is then opened up and used as a sort of bridge over very long stretches of DNA, while its far end is labeled for use as the starting point for another jump.

Joining forces, the researchers combined Collins' jumping technique with Tsui's marker. They walked, they jumped, and they walked some more. They knew they were making real progress. Nevertheless, "it took us one and a half years to find the gene," Tsui recalls, his eyes still shining as he relives the excitement of this period.

Each DNA fragment they used to walk or jump along the chromosome was examined by means of "zoo-blots," which compared the fragment with DNA sequences from animal species. One day they found a match with a sequence from a gene of chickens, mice, and cows. This implied that the fragment was important—only sequences that serve important functions in the body tend to be "conserved" in different species in the course of evolution. But they still had no evidence that it was the CF gene, and other conserved sequences had turned out to be dead ends.

Then another researcher at Toronto's Hospital for Sick Children, John Riordan, clinched it. Riordan, a biochemist, had been culturing and studying the membranes of sweat-gland cells—cells that are clearly involved in CF. Every time Tsui had a new segment of DNA to study, Riordan tested it to see if it corresponded to a genetic message in sweat-gland cells, and finally he found a match with a small part of this last conserved fragment. "That was the beginning of the gene," says Tsui happily. The researchers soon established that a gene containing this fragment was also expressed in other tissues that are specifically affected by CF. Furthermore, the gene looked as if it might code for a membrane protein.

"At that point," says Tsui, "his lab and ours began to sequence like crazy, looking for a difference between DNA from normal and CF cells."

In June 1989, as Collins and Tsui hovered over a fax machine in a freshman dorm room at Yale University, where they were attending a scientific conference, the reports from Tsui's lab brought the news they were waiting for: A small mutation in one particular DNA fragment had been sighted in 70 percent of the chromosomes from CF patients but was absent from normal chromosomes. Collins and Tsui realized they had found the CF gene. But they had to keep quiet about it because, as Collins recalls, "in the next room was one of our major competitors. The dorm walls were pretty thin, so we had to talk in whispers and not yell and scream."

Their reports were published in *Science* on September 8, 1989, to wide acclaim. "The one in 2,000 children born each year with a fatal defect now has a greater chance for a happy future," wrote Daniel E. Koshland, the journal's editor. "Until now, cystic fibrosis could not be studied in animals," he pointed out. The discovery of the gene would make this possible, "thus bringing the day of therapy and cure much closer."

Having the gene at hand would also make it possible to diagnose CF in the unborn, even in families that had no affected members. In addition, it would allow couples to be tested before they started having children, to see if they were carriers of CF (see p. 21).

As the researchers examined their new gene, they found it was made up of 27 segments of DNA that code for parts of a protein (known as "exons"), separated by a similar number of intervening sequences of DNA (known as "introns"). In the majority of patients, the error that caused CF was tiny: three of the gene's 250,000 base pairs were missing. This deletion led to the loss of just one amino acid out of the 1,480 in the protein for which the gene coded instructions. Yet this slight change was enough to radically disrupt the function of patients' lungs, sweat glands, and pancreas.

"Now that the gene has been identified, we have new possibilities for treatment," says Collins. He is encouraged by the fact

"I can't believe it worked!"

that the CF gene is not garbled and has only a small deletion. "The gene can still make a protein," he says. "The protein is there, but it doesn't function. Why?"

That is the question he and a dozen other researchers are now trying to answer. The new protein has been named CFTR, for CF transmembrane conductance regulator. It appears to work like a two-way pump, channeling vital compounds in and out of a cell.

When it functions normally, Collins explains, the protein somehow regulates the transport of chloride and sodium across cell membranes. But in CF this process fails, and the chloride channel stays closed. As a result, water is retained in the cell instead of going out into the airway; this leads to the buildup of thick, dehydrated mucus that characterizes the disease.

"Could we do something to tweak the mutant protein slightly, to make it work a little better?" Collins asks. "I think we'll figure out the basic biological defect in a year or two. I hope we'll find there's a drug that's already been used for something else that can stimulate the protein to function. That would be the nicest outcome. If not, we'll have to develop a new chemical treatment based on our growing understanding of the precise molecular defect."

Collins believes that eventually this will be feasible, and he has a suggestion about how it might be done: Aerosol sprays could be used to bring such treatments right to the patient's lungs, where they are most needed. "That's got to be an attractive delivery system," he says. "It could deliver either the normal CFTR protein or a drug that compensates for the faulty protein."

Looking even further ahead, Collins thinks aerosol sprays might deliver some form of gene therapy. The idea would be "to get the right gene into the right cell, and then get the cell to make the right amount of normal protein," he says.

In September 1990, two teams of researchers announced that they had actually corrected the CF defect—in a dish—by inserting normal genes into cells from CF patients. Michael Welsh, of the HHMI unit at the University of Iowa, and his colleagues "infected" cells taken from a patient's respiratory tract with a virus that had been engineered to act as a vehicle for the normal gene. The cells then began to manufacture the normal CFTR protein, and the chloride channel opened up. At the University of Michigan, HHMI investigator James Wilson and his associates corrected the same CF defect in cells grown from the pancreas of a CF patient. After they ferried the normal gene into these cells on a modified virus, the defective cells started to work normally.

Few scientists believed that success would come so quickly. When their experiment succeeded, Welsh and his co-workers were at first astonished. "There was a great deal of fun and excitement here," Welsh reported in a newspaper interview. "I remember one of my students running out of the lab room saying, 'I can't believe it worked!' "

If gene therapy is tried on CF patients, it may not rely on modified viruses to transport the normal gene because of concern that such viruses might eventually pose some danger. Instead, the normal gene might be packaged inside liposomes, small fat-coated bubbles that pass through cell membranes.

Gene therapy of this sort would raise no ethical issues, since it would not affect the genes of future generations. It would simply be a different kind of chemical treatment.

Many technical hurdles remain before scientists are able "to get the gene into the airways of a living, breathing person—and do it safely," Collins says. For instance, nobody knows whether the heavy mucus in the lungs of CF patients would get in the way of treatment. How many cells would need to be treated for the patients' airways to behave more normally? What is the precise function of the CFTR protein? How long would the inserted genes go on doing their job?

The answers to such questions may come when researchers succeed in developing an animal model of CF and then cure these animals of the disease.

Who Should Be Tested?

Though ten million Americans carry an abnormal copy of the CF gene, hardly any of them know it or realize that they risk having children with the disease.

In 1989, right after the gene was isolated, the relatively small number of clinics that do genetic testing braced themselves for the expected onslaught of people demanding to be tested for CF. But testing was delayed because of the many mutations that can cause the disease.

Screening the general population with a test that identifies only 70 to 75 percent of carriers would be a mistake, a panel at the National Institutes of Health declared in March 1990. The panel pointed out that such screening would detect only about half the couples at risk of producing a child with the disease (couples in which both partners carry a recognizable CF mutation). Many people might be left feeling anxious or uncertain. If one member of a couple tested positive and the other negative, for example, the one who tested negative might actually have an as-yet-unidentified CF mutation. The panel called for pilot programs to determine how best to screen for CF, but recommended postponing more widespread programs until the test can detect at least 90 to 95 percent of carriers.

Since then, much progress has been made. It turns out that by taking into account ethnic origins and including the appropriate mutations in the test, one can greatly increase the test's ability to detect carriers—especially now that some new, moderately prevalent mutations have been uncovered.

"We're very close to being able to identify 85 percent of carriers," Arthur Beaudet, an HHMI investigator at the Baylor College of Medicine in Houston, Texas, said in October 1990. It is easy to test for several mutations simultaneously, he pointed out: "We can get four different mutations on a single test for as little effort as one." A test that identified 85 percent of carriers would detect 72 percent of the couples at risk, he said. Beaudet added that a reasonable goal for the next year or two would be to develop a CF test that can detect 95 percent of carriers, and emphasized that "population-based testing is an opportunity to prevent CF."

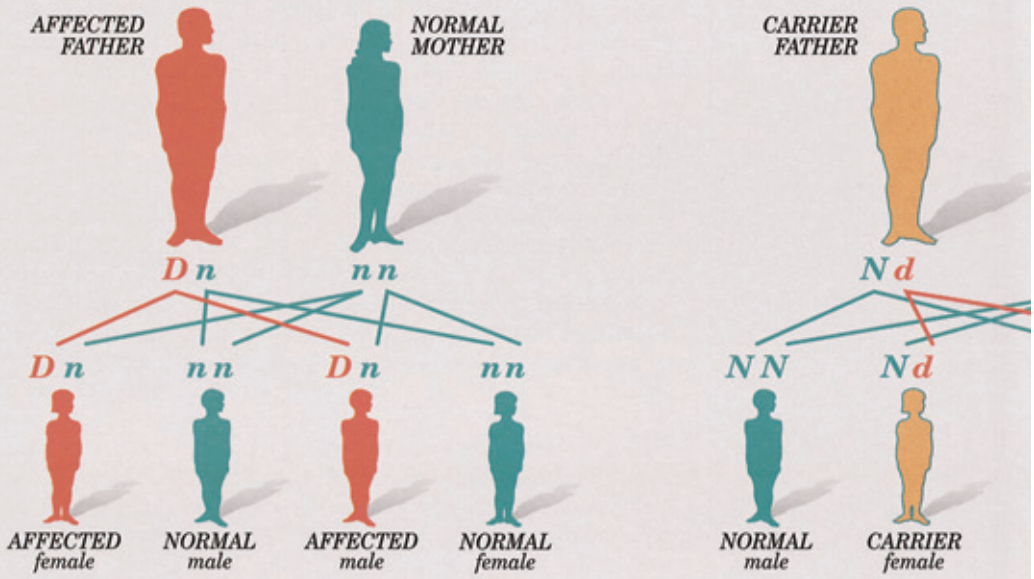
Some physicians are already offering a CF test to their private patients. Furthermore, several pilot projects for population screening have been started in various parts of the world. One of them, carried out by Robert Williamson's lab at St. Mary's Hospital in London, prides itself on its simplicity: Instead of requiring samples of blood, it uses a mouthwash.

"You take a small tube of saline solution, you wash your mouth out with it, and spit it back into a bottle. That's it!" explains Edward Mayall in the Williamson lab. "The great advantage is that you don't need a nurse to take a blood sample," says Eila Watson, the project's research coordinator. "So many people are terrified at the idea of giving a blood sample—they're much more likely to come if they don't have to give it. Plus the ease of handling. From the point of view of people in the lab, it's preferable not to handle blood if possible, so they don't have to worry about AIDS."

The mouthwash test costs less than $30, and the entire test can be done in one day. For children too young to do the mouthwash themselves, the lab uses a mouth scrape. In either case, only a few cells from the lining of the mouth are required. The cells are then heated, which makes them break down and release DNA from their nuclei. Extracting DNA from blood takes much longer.

"We've compared results from the blood test and the mouthwash," says Mayall, "and so far we have found absolutely no difference in the accuracy of the test. Even if the mouthwash sample is contaminated by bacteria or a bit of food, these are unlikely to have the same DNA

How Genetic Disorders Are Inherited



Dominant Disorders: A Fifty-Fifty Chance

The affected parent has a single defective gene (D), which dominates its normal counterpart (n). Each child has a 50 percent risk of inheriting the faulty gene and the disorder.

Recessive Disorders:

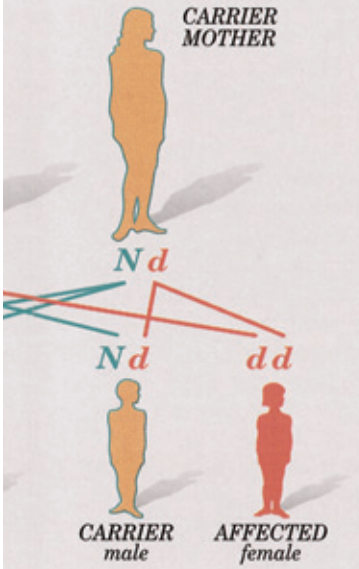
Both parents carry a single defective gene (d) but are protected by the presence of a normal gene (N), which is generally sufficient for normal function. Two defective copies of the

sequence as the mouth cells, so they're no problem."

What really takes time, says Eila Watson, is explaining the test. The project provides genetic counseling in group sessions, eight persons at a time. Those who take the test pay nothing; the entire procedure is covered under the British national health service.

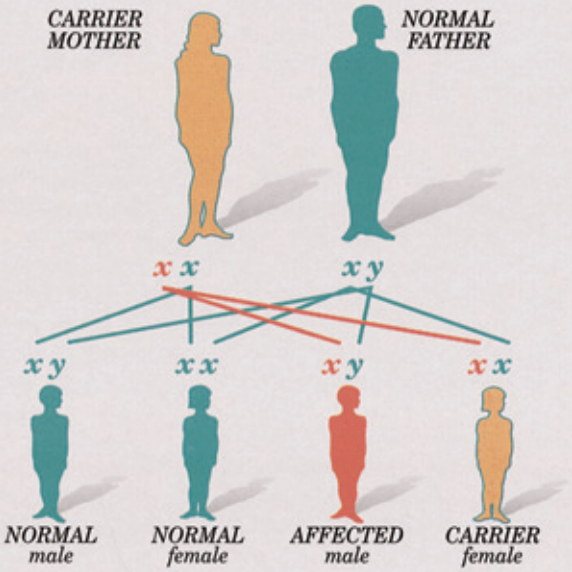
"We want to allow people to make informed decisions before conception," says Watson. This is why one health district that handles some 3,000 births per year is offering the test to everyone within the reproductive age bracket, 16 to 44, both through general practitioners and through family-planning clinics. Whenever a carrier is identified, the test is offered to all members of the carrier's family.

If screening becomes practical on a wide scale, many questions will need to be answered. First, who should be screened? Should everyone be tested, including blacks (whose incidence of CF is only one-



One Chance in Four

gene are required to produce a disorder. Each child has a 50 percent chance of being a carrier like both parents and a 25 percent risk of inheriting the disorder.



X-Linked Disorders: Males Are at Risk

One normal copy (blue x) of a gene on the X chromosome is generally sufficient for normal function. Women who have a defective gene (red x) on one of their two X chromosomes are protected by the normal copy of the same gene on the second chromosome. But men lack this protection, since they have one X and one Y chromosome. Each male child of a mother who carries the defect has a 50 percent risk of inheriting the faulty gene and the disorder. Each female child has a 50 percent chance of being a carrier like her mother.

tenth that of whites) and Asians (who almost never have the disease)? Or should CF screening be limited to whites, as the President's ethics commission recommended in 1983?

Another question is how to prevent insurance companies from obtaining the results of such tests and using them to deny health-care coverage to people with genetic defects. Discrimination of this sort will grow unless new laws ensure that genetic information is confidential, according to Paul Billings, a medical ethicist at Pacific Presbyterian Medical Center in San Francisco.

Finally, at what age should people be tested?

Testing in the schools has some advantages—students are a captive audience mostly in the pre-pregnancy stage—but geneticists fear that confusion about the test's results might lead to stigma. They remember what happened in the early 1970s when the nation embarked on an

energetic program of screening for sickle-cell disease, which affects 1 in 400 American blacks.

At that time, many of the people who learned they were carriers of the sickle-cell defect mistakenly thought they had the disease (which requires inheriting two copies of the defective gene) and might die of it. Some were denied jobs or health insurance. And since no prenatal test was available then—it exists today—some carrier couples were told that the only way to prevent the disease was to avoid having children, advice that led to charges of racism.

When screening programs are not properly introduced or explained, they run the risk of being counterproductive. This danger is best illustrated by an often mentioned but perhaps apocryphal story about a Greek city's experience with screening for thalassemia, a severe blood disorder prevalent among Mediterranean peoples. Although people who carry just one copy of the thalassemia gene defect are perfectly healthy, parents are said to have looked askance at any prospective bride or groom who tested positive for it. Carriers thus became pariahs and ended up marrying each other—exactly the reverse of the desired outcome, since this greatly increased the chance that their children would inherit two copies of the defective gene and develop the disease.

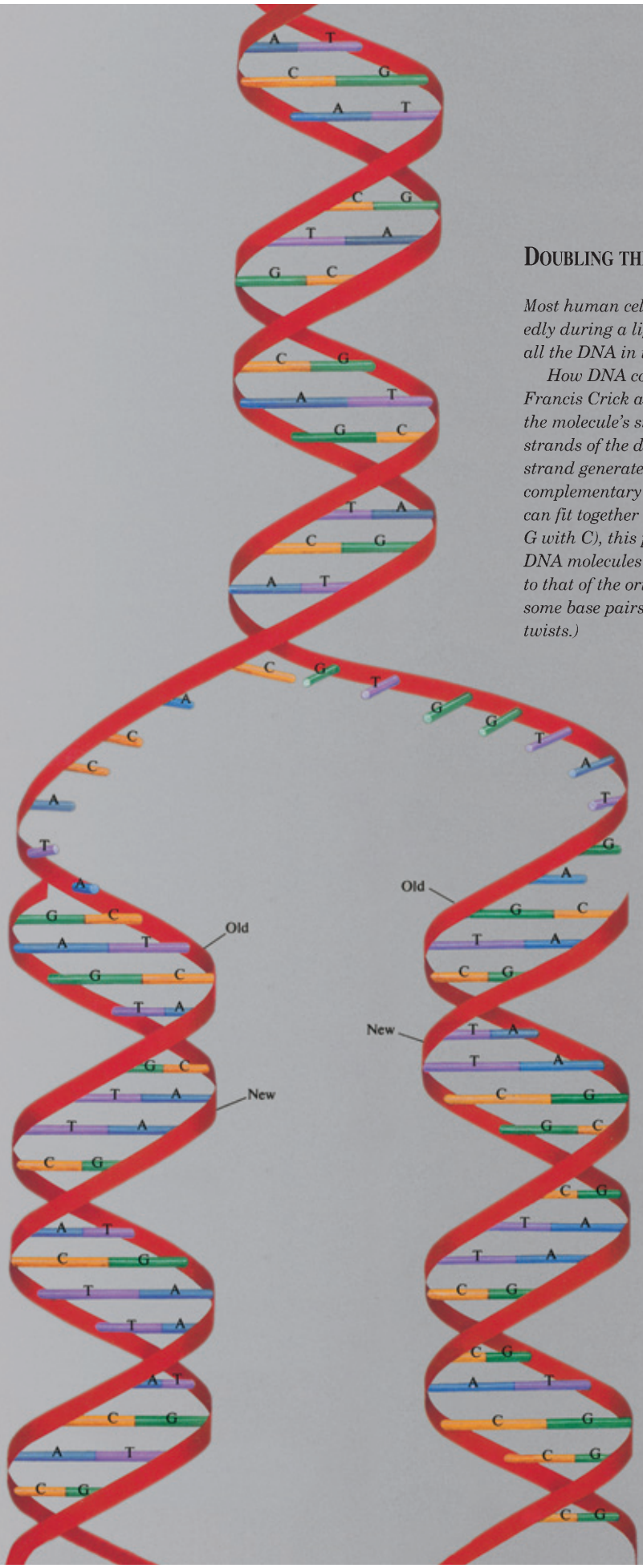
By contrast, in Ferrara, Italy—a city that used to have about 30 new cases of thalassemia every year—screening was combined with extensive education about the program. This was so successful that hardly a child has been born with the disease in nearly a decade.

The most reasonable time to screen for CF might well be when a couple gets a marriage license, suggests Francis Collins. Testing at this stage would reduce the likelihood that the test results would influence a person's choice of marriage partner. But it would miss the large number of unmarried couples who have children. On the other hand, it would be a mistake to wait until the first prenatal visit to the obstetrician because that would limit a couple's options. "Most couples would like to have this information prior to conception," Collins says.

There are, in fact, quite a number of options for people who learn that they carry a CF gene defect and want to make sure their children do not suffer from the disease. If they marry someone who is not a carrier, the problem does not arise. If both partners are carriers, they can ask for prenatal diagnosis and, if the fetus has CF (a one-in-four chance), they may choose to terminate the pregnancy. If they find abortion unacceptable and do not want to take the chance of having an affected fetus, they can opt for artificial insemination with sperm from someone who is not a carrier. They can adopt a baby. Or they can try *in vitro* fertilization. At the Illinois Masonic Hospital in Chicago, for instance, Charles Strom and Yury Verlinsky are developing a way to screen the eggs of women who are carriers, select only eggs that do not carry the CF mutation, and then use those eggs for *in vitro* fertilization with the husband's sperm; once fertilized, the eggs are reimplanted in the woman's body.

Everyone agrees that screening for genetic diseases should be voluntary and confidential and that counselors should not impose their own values. "If a patient who is struggling with a terrible decision asks me what I would do, I don't answer," says Francis Collins. "I must not answer. It is a sacred and unbreakable rule."

What counselors can do is explain the options. It is hard to predict how long CF children who are conceived today will live with good medical care, or how they will respond to new treatments. However, "there is still no cure for CF at this stage," Lap-Chee Tsui points out. "So if it is possible to avoid the disease, one should be given the chance."



Doubling the Double Helix

Most human cells divide and multiply repeatedly during a lifetime. Before each cell division, all the DNA in the nucleus must be duplicated.

How DNA copies itself was revealed when Francis Crick and James Watson deciphered the molecule's structure in 1953: The two strands of the double helix unwind, and each strand generates a mirror image made up of complementary bases. Since DNA base pairs can fit together in only one way (A with T, and G with C), this process results in two daughter DNA molecules whose sequences are identical to that of the original DNA. (In this illustration, some base pairs are hidden behind the DNA twists.)

Human chromosomes

HOW TO CONQUER A GENETIC DISEASE

Nearly 4,000 genetic diseases afflict human beings. Given enough time and effort, scientists can learn to prevent or treat a great many of them. This requires answering three questions—major landmarks on a trail of genetic discoveries:

- Which altered gene causes the disease?
- What protein does this gene normally produce?
- Can the altered protein or gene be fixed or replaced?

Two different strategies may be used. Researchers may find the altered protein first (if it can be detected chemically in tissues that are affected by the disease) and then locate the gene that codes for it. When this is impossible, they use positional cloning: They find the gene first (by zeroing in on the DNA inherited with the disease, or by locating a similar gene in a mouse) and then identify the protein the gene makes.

The trail shown here illustrates positional cloning. This strategy recently led to spectacular progress toward diagnosing or treating cystic fibrosis (CF), Duchenne muscular dystrophy, neurofibromatosis, and other inherited disorders.

© 1991 Howard Hughes Medical Institute

A Brief Key To Basic Genetics

A human cell. *Each of the 100 trillion cells in the human body (except red blood cells) contains the entire human genome—all the genetic information necessary to build a human being. This information is encoded in 6 billion base pairs of DNA (3 billion base pairs from each parent). Egg and sperm cells each contain 3 billion base pairs of DNA.*

The cell nucleus. *Inside the cell nucleus, 6 feet of DNA are packaged into 23 pairs of chromosomes (one chromosome in each pair coming from each parent).*

A chromosome. *Each of the 46 human chromosomes contains the DNA for thousands of individual genes, the units of heredity.*

A gene. *Each gene is a segment of double-stranded DNA that holds the recipe for making a specific molecule, usually a protein. These recipes are spelled out in varying sequences of the four chemical bases in DNA: adenine (A), thymine (T), guanine (G), and cytosine (C). The bases form interlocking pairs that can fit together in only one way: A pairs with T; G pairs with C.*

A protein. *Proteins, which are made up of amino acids, are the body's workhorses—essential components of all organs and chemical activities. Their function depends on their shapes, which are determined by the 50,000 to 100,000 genes in the cell nucleus.*

WHICH GENE IS AT FAULT?

1 A child (blue box) develops a currently incurable genetic disease. Scientists who wish to find a specific treatment or a means of preventing the disease in other children must trace it to its cause: an altered gene.

2 Various clues, such as a visibly missing piece of a chromosome, may reveal the gene's rough location on a chromosome. When there are no such clues, researchers look for "markers" of the disease by comparing the DNA of the affected child to that of parents, relatives, and persons in other families. When placed on a "genetic map," these markers reveal which chromosome carries the altered gene. As more markers are added to the map, the location is narrowed to the space between two known markers.

3 RESULT: Scientists may be able to diagnose the disease prenatally by following the inheritance of markers in an affected family. They may also recognize healthy carriers of the altered gene (light blue boxes). The family shown here has a recessive disease that develops only when a child inherits the altered gene from both parents.

4 "Walking" or "jumping" toward the gene, scientists create a physical map, or a chain of overlapping segments of DNA in the space between the flanking markers. One of these segments must contain the altered gene. In the future, when researchers have covered all the chromosomes with overlapping fragments of DNA, any fragment they want will be available from a computer database.

5 Zeroing in on the altered gene, scientists analyze each segment: Is it different from normal DNA? Finally they find the guilty gene and determine the error in its sequence of bases. The most common error in the CF gene is a deletion of three DNA bases (top right) out of a total 250,000.

A mutant CF gene sequence:

ATCATTGGTGTT

A normal CF gene sequence:

ATCATCTTTGGTGTT

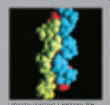
(Some letters are hidden behind the DNA twists.)

WHAT PROTEIN DOES IT MAKE?

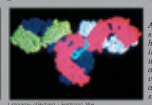
6 RESULT: Scientists can test for the disease directly in patients (blue boxes) and prenatally. They can identify healthy carriers of the altered gene (light blue boxes) in the general population—not just among members of an affected family. They can study the disease process in cultured cells and in animals, with a view to developing new treatments.

7 Knowing the gene's sequence, scientists use the genetic code to determine which amino acids make up the protein. Then they study the protein and find out what it is supposed to do. The tens of thousands of proteins in the body have different shapes and do different jobs, depending on instructions encoded in the genes.

The elongated shape of collagen protein, shown in this molecular model, allows collagen to provide structural support to cells and organs.

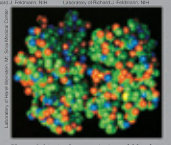


Laboratory of Mr. Ralph J. Dickerson, NIH



Laboratory of Richard J. Feldmann, NIH

Antibodies, such as this human immunoglobulin, recognize and attack viruses, bacteria, and other foreign substances.



Laboratory of Max F. Perutz, MRC, Cambridge, U.K.

Hemoglobin is the protein in red blood cells that delivers oxygen to all parts of the body via the bloodstream.

8 Does the sick child's altered gene produce too little protein, a flawed protein, or no protein at all? Scientists need to understand just how the protein change causes the disease.

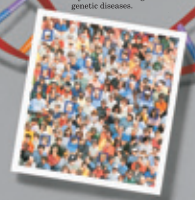
9 RESULT: As the mechanism of the disease becomes clear, scientists can devise new approaches to treatment involving either the protein or the gene. Understanding a relatively rare inherited disorder may also bring important insights into more common and complex diseases.

CAN THE PROTEIN OR GENE BE REPLACED?

10 To make up for the genetic error, scientists may try to replace a missing or ineffective protein with a drug or with the normal protein. Such experiments are usually carried out first in cultured cells in the laboratory, then in animals, and finally in humans.

11 Another option is gene therapy. Some scientists "infect" cells with a virus into which they have inserted normal genes. Others use non-viral methods or even inject DNA directly into cells. Experiments that work in cultured cells are tried in animals and then in humans. For example, a patient's bone-marrow cells may be removed, treated with normal genes, and returned to the patient.

12 RESULT: Treatments are being developed for some genetic diseases. People will always carry genetic alterations, but in the future, prevention and treatment will vastly reduce suffering from genetic diseases.

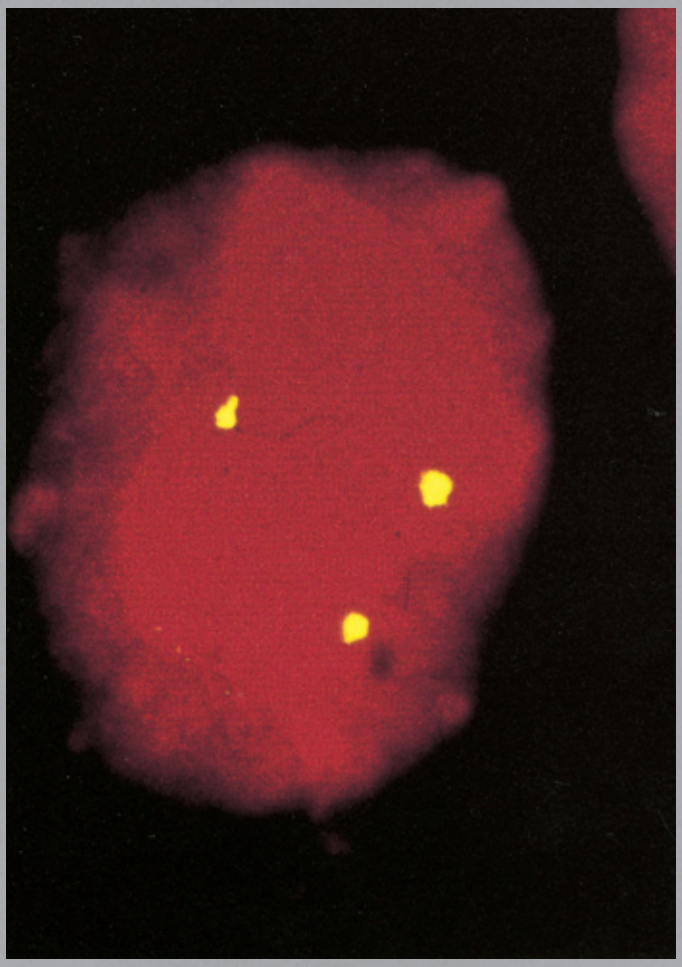


Illustrations: Vital Biomedicals Studio, Inc

THE OTHER KINDS OF GENETIC DISORDERS: ABNORMAL CHROMOSOMES, ABNORMAL CELLS

Some genetic accidents go far beyond inherited errors in a single gene. Large chunks of a chromosome sometimes break off or attach themselves to the wrong place during the formation of egg and sperm cells. An entire chromosome may fail to separate properly, and an egg or sperm may end up with two chromosomes of the same type instead of one. The most frequent cause of Down's syndrome is an extra chromosome 21—three instead of the normal two per cell—as detected in the ominous picture to the right.

A totally different kind of genetic disorder—neither inherited nor inborn—results from accidents in a cell. For example, cancer is a genetic disease that can start in any somatic (body) cell, such as a breast, lung, or blood cell. If radiation, chemicals, or some other agents produce a mutation in the DNA of a somatic cell, the mutation will be transmitted to daughter cells as the cell divides. This may lead to cancer if there are other contributing factors.



Laboratory of David Ward, Yale University

IN SEARCH OF LARGE FAMILIES

Like Southern novelists, geneticists love to chart the destinies of large, interrelated families with inherited disorders. But geneticists don't have the luxury of literary invention. They must find these kindreds in the flesh—and they will go to great lengths to do so.

Close-knit, isolated communities are as valuable to medical geneticists as inbred strains of plants are to botanists, points out Victor McKusick, who started the department of medical genetics at The Johns Hopkins University School of Medicine. Since these family groups—in obedience to religious or political beliefs, or because of geographic obstacles—have not mingled with other populations for generations, the genes of a small colony of founders are often passed on intact, and a variety of traits can be followed through parents, children, cousins, and aunts.

The lure of such families has brought geneticists to unlikely settings. It took Nancy Wexler from the National Institutes of Health in Bethesda, Maryland, to a remote shore of Lake Maracaibo in Venezuela, and it led Janice Egeland from Yale University to the sequestered Amish community in Lancaster County, Pennsylvania. However, any family can be valuable to geneticists if it is large and complete enough. In search of truly large families, Raymond White moved from the University of Massachusetts in Worcester to Salt Lake City, Utah, where families with twelve children are not too rare.

Nancy Wexler's remarkable quest for the cause of Huntington's disease (HD) has become a legend in less than a decade. HD is a genetic time bomb. In 1872 a New York physician, George Huntington, first described the inexorable course of this disease, which usually starts in middle age and slowly destroys its victims both mentally and physically. He noted that when either parent had the disease, "one or more offspring almost invariably suffer

by Beverly Merz

from it, if they live to adult age." Later physicians realized that HD's inheritance followed the pattern of "dominant" traits (see p. 22) first described by the Austrian monk Gregor Mendel in 1866. This meant that any child of a patient stood a fifty-fifty chance of developing the disease.

Having watched the disease erode her mother's body, mind, and spirit over the course of a decade, Wexler joined her father Milton, a Los Angeles psychoanalyst, at the Hereditary Disease Foundation, an organization he had created to do research into the disorder. At one of the foundation's research seminars, Wexler was riveted by a film of a Latin American village in which dozens of people lurched about, seemingly afflicted with the very disorder that held her mother in its grip. When federal funding for HD research finally became available in 1978, Wexler knew where to head with the money—San Luis, the Venezuelan village on Lake Maracaibo in which the film was set.

They hit the jackpot on the third try.

The people in the village were unaccustomed to medical attention and wary of strangers. However, Wexler had a unique bond with them because she, too, was at risk for HD. She managed to charm, wheedle, and cajole them into donating blood samples from which DNA could be extracted, while at the same time sorting out family connections. She eventually compiled the entire pedigree—a family tree showing the presence or absence of a trait—of almost 10,000 people and collected 2,000 samples, which are immortalized in cell cultures.

Wexler collaborated with Harvard molecular biologist James Gusella, Indiana University's Michael Conneally, and the Massachusetts Institute of Technology's David Housman, who agreed to analyze the DNA samples she collected. They had probes to recognize markers on an assortment of chromosomes and hoped that one of these probes might detect a marker that traveled with the gene. Expecting to spend years in this trial-and-error enterprise, they were elated to hit the jackpot on the third try.

Their third probe bound to fragments of DNA from people with HD in a distinctive pattern that differed from that of their healthy relatives. However, the pattern of markers that signaled the disease varied from family to family, making it necessary in some cases to test a large number of relatives to establish which pattern correlated with the presence of HD.

While the researchers had not found the gene itself, they had made it possible to tell, in most cases, which members of an affected family had inherited the fateful DNA patterns and therefore were likely to develop HD in middle age. This prediction could be made years before the treacherous course of the disease became manifest. It allowed people to make informed decisions during their childbearing years. Another new option within these families was prenatal diagnosis. The discovery of the marker also paved the way for finding the gene itself in the future and eventually developing new treatments.

At the time Nancy Wexler first traversed the hemisphere in search of the HD gene, Janice Egeland, who is now at the University of Miami, already had two decades of experience as a gene hunter among the Old Order Amish in the Pennsylvania farmlands. Egeland had approached the Amish in 1959—with some apprehension. She knew that the strict Mennonite sect, which eschewed modern dress, electricity, and automobiles, did not eagerly embrace outsiders. Her quiet perseverance eventually broke down their resistance, and she was taken into their confidence.

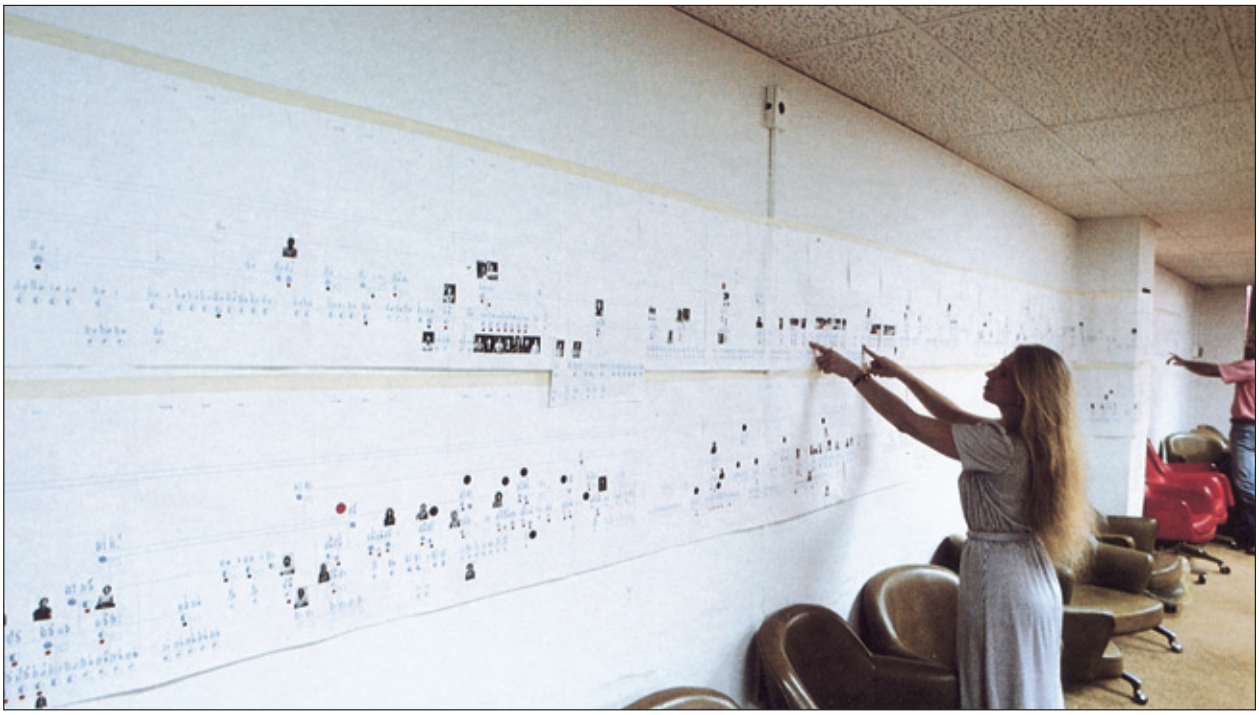
Egeland found that the Amish were as meticulous in their family record-keeping as they were in adhering to their religious convictions. It took her nine years, but she traced every one of the 7,000 living members back to the original 30 pioneer couples who founded the community. This allowed her to see patterns of disease running through families.

With an eye on the records and an ear tuned to local gossip, she also became aware of striking behavioral deviations within the Amish community. Members of certain families were known to have aban-

In a house built on stilts over a Venezuelan lagoon, researchers gather pedigree information from families with many cases of HD.

doned their chores and spent days sitting in a corner, staring at the floor, only to rise a few weeks later and suddenly become loud or boastful. They went on sudden trips, ordered as much as a ton of chocolates for their small grocery store, dressed in a flashy way, or cut their hair in defiance of the community's rules. While these aberrations might have been overlooked in a big city, they were as out of place as neon lights in the staid Amish setting, and Egeland came to recognize them as hallmarks of manic-depressive disorder.

By the late 1970s, Egeland had charted the passage of the disorder through as many as 6 generations of 32 families. Yet she hadn't a clue as to where the gene that



Nancy Wexler examines a section of the Venezuelan family's HD pedigree on a wall at NIH. The family tree now includes nearly 10,000 persons and is over 100 feet long.

governed these mood swings might reside. Then, in 1982, she decided to try positional cloning. She collected blood samples from 81 family members, including 19 who suffered from manic-depressive illness, and sent them to Yale and MIT for testing. Daniela Gerhard, then a postdoctoral fellow in David Housman's lab at MIT (she is now at Washington University in St. Louis), found a marker on chromosome 11 that was present in the DNA of all the family's manic-depressive members but absent in most of their normal relatives. The marker pointed to the guilty gene's location on the tip of that chromosome.

Other groups have failed to substantiate these findings, however. Some have linked manic-depressive disease in families to markers on the X chromosome, while others have implicated chromosome 5 or found no links at all. And in 1989, after a few more members of the Amish families had come down with symptoms of manic-depressive disease, Egeland's own research group published a "re-evaluation" that markedly reduced the likelihood of their marker being linked to the disease.

These results illustrate the difficulty of tracking down the genes for complex diseases, says Raymond White. Complex diseases such as cancer, schizophrenia, and heart disease, as well as manic-depressive disorders, are probably produced by several genes acting together and by different combinations of errors in different families. However, environmental influences such as diet, smoking, chemicals, or viral infections also play a role in triggering these diseases.

In dealing with mental illness, the picture may be clouded even further by uncertainties in diagnosis and a variable age of onset. By contrast, heart disease, despite its complexities, at least provides some solid indicators that can be measured and is therefore much more amenable to study.

In 1985 Michael Brown and Joseph Goldstein of the University of Texas Health Science Center at Dallas won a Nobel Prize for their work on a genetic disorder that can lead to early death from heart attack: famil-

ial hypercholesterolemia, which involves extremely high levels of cholesterol in the blood. The two scientists found that, though the symptoms of this disease are fairly uniform, it is caused by a variety of mutations that alter different functions of the same gene in different families. Some defects prevent special receptors on the surface of cells from binding as they should to low-density lipoprotein (LDL), which carries cholesterol in the blood. Other defects allow binding but prevent the uptake of LDL into cells. Either way, the cells fail to clear cholesterol from the blood. As Goldstein explains, "The more families we looked at, the more defects we continued to find."

Because every family may have a different skeleton hidden in its genetic closet, the quest to locate more large families with documented histories is endless. One of the best places to find such families is Salt Lake City, headquarters of the Church of Jesus Christ of Latter-Day Saints, also known as the Mormon Church. The Mormon Church not only encourages its members to produce large families, but urges them to compile elaborate genealogies.

"Every family has a historian, and every historian can tell you what diseases the members have had," says Raymond White. However he has found that he doesn't need to spend time combing the family historians' records for interesting pedigrees. They come to him.

"From time to time a doctor will call us and say, 'I've been taking care of a family you might be interested in,' " he explains. One family referred to White had familial polyposis of the colon, a condition that predisposes people to colon cancer. This condition was also being studied by Mark Skolnick, a population geneticist who has worked with Mormon genealogical records at the University of Utah since 1974. The classic harbinger of malignancy in people with this trait is the appearance of more than a hundred polyps, usually before the age of thirty. To avert the development of colon cancer, surgeons remove the colon at the first sign of polyps, usually while the patients are still in their twenties.

Pathologists have found that the polyps contain a variety of cells in different stages of alteration, ranging from cells closely resembling normal colon cells to cells that are all but cancerous. Oncologists now believe that this orderly progression to malignancy is caused by an accumulation of genetic mutations in colon cells. One of the principal genes involved has been traced to chromosome 5.

The patients in this family seemed to fit the classic pattern. Over seven generations, 15 of the 115 members developed colon cancer and 12 died of the disease. Yet the disease didn't always follow its predictable course; some members of the family had only a moderate number of polyps and one member who had no polyps at all developed colon cancer and died.

When White and his colleagues Mark Leppert and Jean-Marc Lalouel analyzed DNA from 81 family members, they found that several who appeared healthy actually carried the marker that signaled the gene on chromosome 5. White says that if these people hadn't been members of the family he was studying, they would have escaped identification as carriers of the gene and as people at risk for colon cancer. Now they and their children can be examined periodically to determine whether they are developing polyps, and they may be able to avoid losing their colons if their polyps are removed early. "Without the statistical power of a large family, it is often difficult to determine whether polyps are due to an inherited gene defect or to a sporadic mutation," White says.

Finding the genes for specific diseases is all very well, but Raymond White's main goal is to develop a set of tools for all geneticists—a genetic linkage map. In a landmark paper in 1980, David Botstein of the Massachusetts Institute of Technology, White, Skolnick, and Ronald Davis of Stanford University had proposed making such a map with the aid of DNA variations. White came to Utah with the intent of blanketing the genome with markers that could serve as reference points, or landmarks, for gene mapping. To do so he collected DNA from 46 normal families of

Every family may have a different skeleton hidden in its closet.

three generations—four grandparents, two parents, and at least six children. He then resorted to a shotgun approach, using every available probe to look for markers.

By 1991 he had 500 markers scattered throughout the genome—a nearly complete, continuously linked set of landmarks no more than 20 million base pairs apart and generally much closer. He has sent probes for these markers to geneticists engaged in tracking down genes for inherited diseases, as well as to researchers looking at the sporadic mutations that cause cancer. His markers were instrumental in guiding Francis Collins and Lap-Chee Tsui to the cystic fibrosis gene, in establishing the location of the gene for familial polyposis of the colon, and in directing his own team to the gene for neurofibromatosis (NF), a disfiguring disease that used to be called erroneously "elephant man's disease."

It turns out that the NF gene is one of a family of genes which, when normal, operate as anticancer genes. Proteins produced by these genes may work to block tumor formation in the brain and central nervous system, as well as in the lung, liver, colon, and breast. This discovery may open the door to entirely new drugs to block the development of cancer.

All too often geneticists remain frustrated in their search for the "right" families. Although several teams of researchers are tracking the gene responsible for Alzheimer's disease, for example, there are few large families in which the disease has been documented reliably over a period of time. As a result, there have been conflicting reports. For example, a team led by John Hardy of St. Mary's Hospital, London, announced it had discovered a mutation in a gene on chromosome 21 in members of two unrelated families who have Alzheimer's disease. This agreed with a Harvard team's finding of linkage to markers on chromosome 21, but disagreed with a Duke University finding of linkage to chromosome 19.

One way to speed up research, Nancy Wexler suggests, is for all families to follow the lead of the Amish and the Mormons and chart their own pedigrees. "Virtually everyone knows what a family tree is and how to put one together. And people know what diseases run in their families—not necessarily rare diseases but heart disease, cancer, diabetes." She adds that people with unusual family trees—for example, several members with early heart attacks—as well as those with rare diseases should try to link up with researchers through such groups as the National Organization for Rare Disorders, especially if their families are large.

Finding the key to hereditary disease is a family affair.

Owen Davison, a management consultant in Hershey, Pennsylvania, did just that after his son Richard, a young social worker, died unexpectedly during routine surgery for a broken ankle. Richard was diagnosed as having malignant hyperthermia, an allergic condition in which certain commonly used anesthetics and muscle relaxants trigger a lethal series of events, including muscle contractions, bleeding, an increase in body temperature to 108° F or more, and cardiac arrest. The syndrome affects at least 1 person in 40,000. When Davison learned that malignant hyperthermia was a dominant trait, he wrote letters to some 300 relatives, warning them that they might be susceptible to the disorder and suggesting they might want to have a muscle biopsy to find out. Several members of the family who followed his advice, including Davison's niece Suellen Gallamore, learned that they, too, might succumb if subjected to these anesthetics. They were advised to wear identification bracelets, notifying hospital personnel of their susceptibility, and to take special precautions if scheduled for surgery.

In 1980 Davison, Gallamore, and Henry Rosenberg, an anesthesiologist at Hahnemann University, founded the Malignant Hyperthermia Association of the United States. The group has attracted other families with the disorder, as well as anesthesiologists and physiologists who are studying the bizarre syndrome.

Like the Wexlers, the Davisons have demonstrated that finding the key to hereditary disease is very much a family affair.

Reading The Human Blueprint

"All human disease is genetic in origin," Nobel laureate Paul Berg of Stanford University told a cancer symposium a few years ago.

Berg was exaggerating, but only slightly. It has become increasingly evident that virtually all human afflictions, from cancer to psychological disorders and susceptibility to infection, are rooted in our genes. "What we need to do now is find those genes," says James Watson, who shared a Nobel Prize for deciphering the structure of DNA and who helped launch the Human Genome Project.

Single-gene diseases such as CF are relatively easy targets. Disorders that seem to be caused by the interplay of several genes—hypertension, atherosclerosis, and most forms of cancer and mental illness—are much more difficult to track down. Having a map of the entire human genome will make it possible to identify every gene that contributes to them.

A gene map can also lead researchers to new frontiers in drug development. Once all the genes are identified and their bases are sequenced, it will be possible to produce virtually any human protein—valuable natural pharmaceuticals, such as tissue plasminogen activator, interferon, and erythropoietin—as well as new molecules designed specifically to block disease-producing proteins.

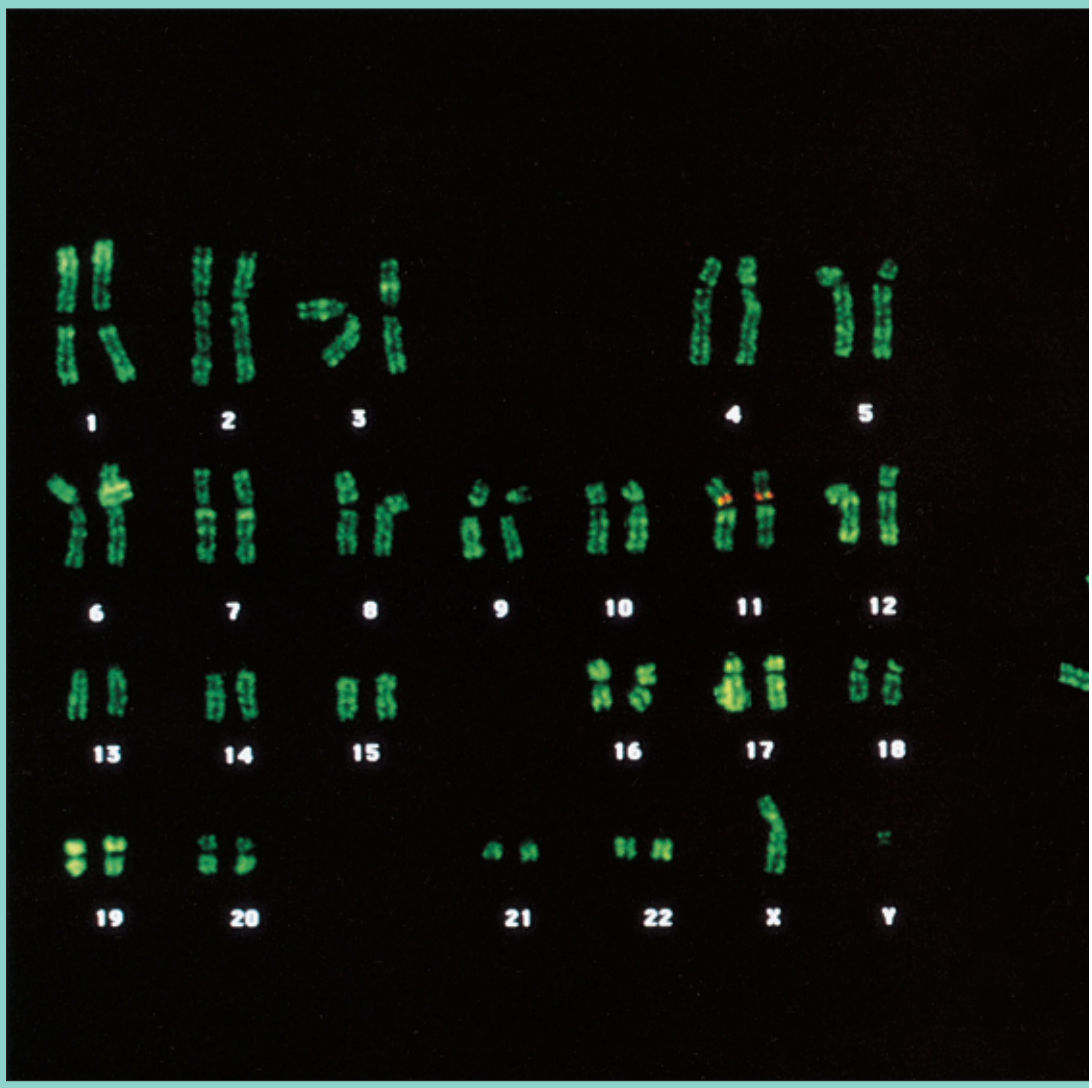
The Human Genome Project officially began in October 1990. But the map of the human genome has been in the making for a good part of the century. It started in 1911, when the gene responsible for red-green color blindness was assigned to the X chromosome, following the observation that this disorder was passed on to sons by mothers who saw colors normally (see p. 23). Some other disorders that affect only males were likewise mapped to the X chromosome on the theory that females, who have two X chromosomes, were protected from these disorders by a normal copy of the gene on their second X chromosome—unlike males, who have one X and one Y chromosome.

The other 22 pairs of chromosomes remained virtually uncharted until the late 1960s. Then biologists fused human and mouse cells to create uneasy hybrid cells that cast off human chromosomes until only one or a few remained. Any recognizable human proteins in these hybrid cells thus had to be produced by genes located on the remaining human chromosomes. This strategy allowed scientists to assign about 100 genes to specific chromosomes.

Map-making really took off in the early 1970s, when geneticists discovered characteristic light and dark stripes or bands across each chromosome after it was stained with a chemical. These bands, which fluoresced under ultraviolet light, provided the chromosomal equivalent of latitudes. They made it easier to identify individual human chromosomes in hybrid cells and served as rough landmarks on the chromosomes, leading to the assignment of some 1,000 genes to specific chromosomes.

Around the same time, recombinant DNA technology began to revolutionize biology by allowing researchers to snip out pieces of DNA and splice them into bacteria, where they could be grown, or cloned, in large quantities. This led to two new mapping strategies. In one, *in situ* hybridization, scientists stop the division of human cells in such a way that each chromosome is clearly visible under a light microscope. Then they use probes to find the location of any DNA fragment on these chromosomes. Originally these probes were radioactively labeled, but chemically-tagged probes that can be made to fluoresce have been found to yield far more accurate and rapid results (see p. 41).

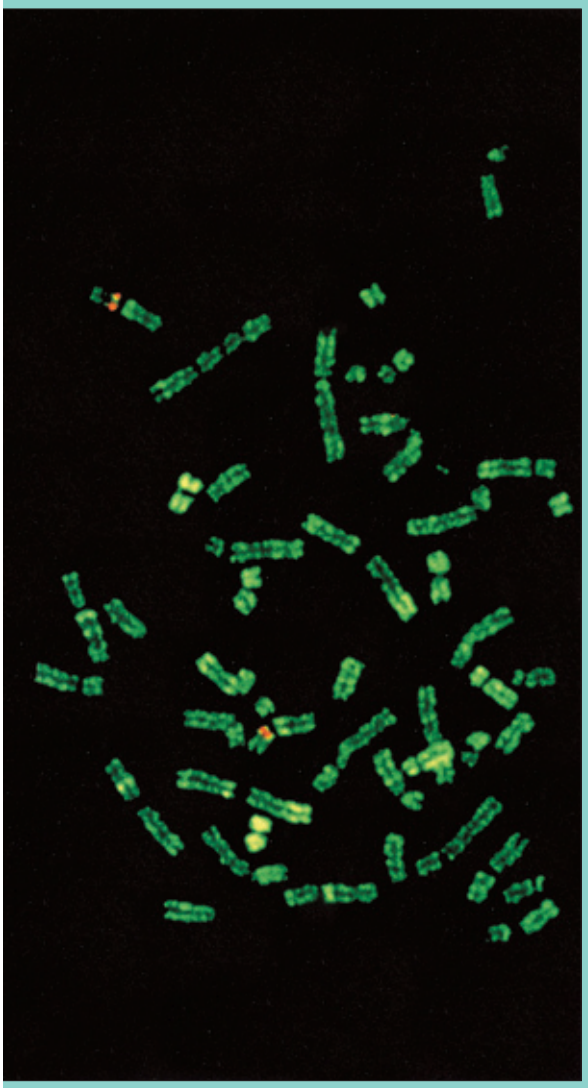
The other strategy is to use DNA variations as markers on the human genome, as proposed by Botstein, White,



Skolnick, and Davis in 1980. This resulted in a flood of new markers and an explosion in the knowledge of genes' chromosomal whereabouts. The number of genes mapped tripled in 10 years. Gene mappers now update the map every day via electronic databases.

In the mid-1970s, when Frederick Sanger at Cambridge University and Walter Gilbert and Allan Maxam at Harvard University developed efficient new methods for determining the order of bases in a strand of DNA, scientists learned to sequence the genes they isolated. Automated sequencing followed in the 1980s. Now, once a new gene has been identified, it is immediately sequenced to understand the nature of the protein it codes for and to identify mutations that are related to disease.

Sequencing the entire genome, however, means sequencing at least 3 billion



Laboratory of David Ward, Yale University

A previously uncharted fragment of human DNA is mapped to chromosome 11 with the aid of probes that glow in red. The 22 pairs of autosomal (non-sex) chromosomes in the genome, plus one X and one Y chromosome, are shown jumbled (right). After being identified by their banding patterns, they are arranged in order (left). Each chromosome appears double because it was caught in the process of division.

base pairs of DNA—one chromosome of each type, or half the total number of chromosomes in a human cell.

Generally the most interesting or accessible genes have been located first, creating a disparity among chromosomal maps. The Human Genome Project is expected to produce a genetic linkage map in which the positions of genes for specific traits and diseases are superimposed on a grid of evenly spaced markers along the chromosomes.

Scientists will be able to determine any gene's location relative to these markers.

In addition, a physical map will show actual distances along the chromosomes in terms of base pairs. The physical map probably will be constructed of long overlapping stretches of DNA cloned in yeast and known as yeast artificial chromosomes (YACs). Developed in 1987 by Maynard Olson, now an HHMI investigator at Washington University in St. Louis, YACs make it possible to clone and store much larger DNA segments than those cloned in bacteria. The technique has reduced the number of DNA pieces that need to be placed in the right order from about 100,000 to 10,000. Olson distributed his YAC library of the entire genome for the use of gene mappers.

Meanwhile, new strategies promise to speed up sequencing significantly. Automated sequencing and advanced computer software are already reducing sequencing time.

If a map of the genome and sets of overlapping clones had been available when researchers set out to find the cystic fibrosis gene, their task would have taken only a fraction of the time and cost, points out Thomas Caskey, of the HHMI unit at the Baylor College of Medicine. "The investigators could have simply reached into the freezer and pulled out two markers flanking the gene. The same would be true for many other diseases. And remember—once we make this map, we will never have to do it again." —*B.M.*

"Once we make this map, we will never have to do it again."

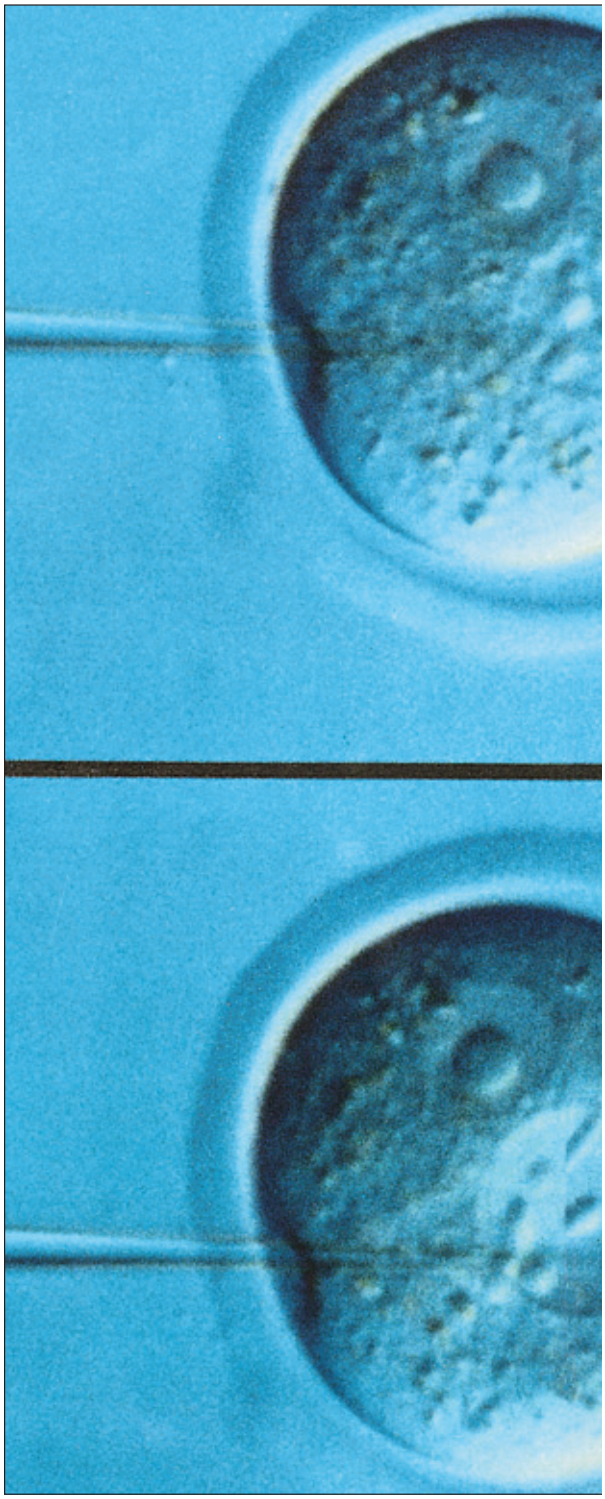
OF MICE AND MEN

by Sandra Blakeslee

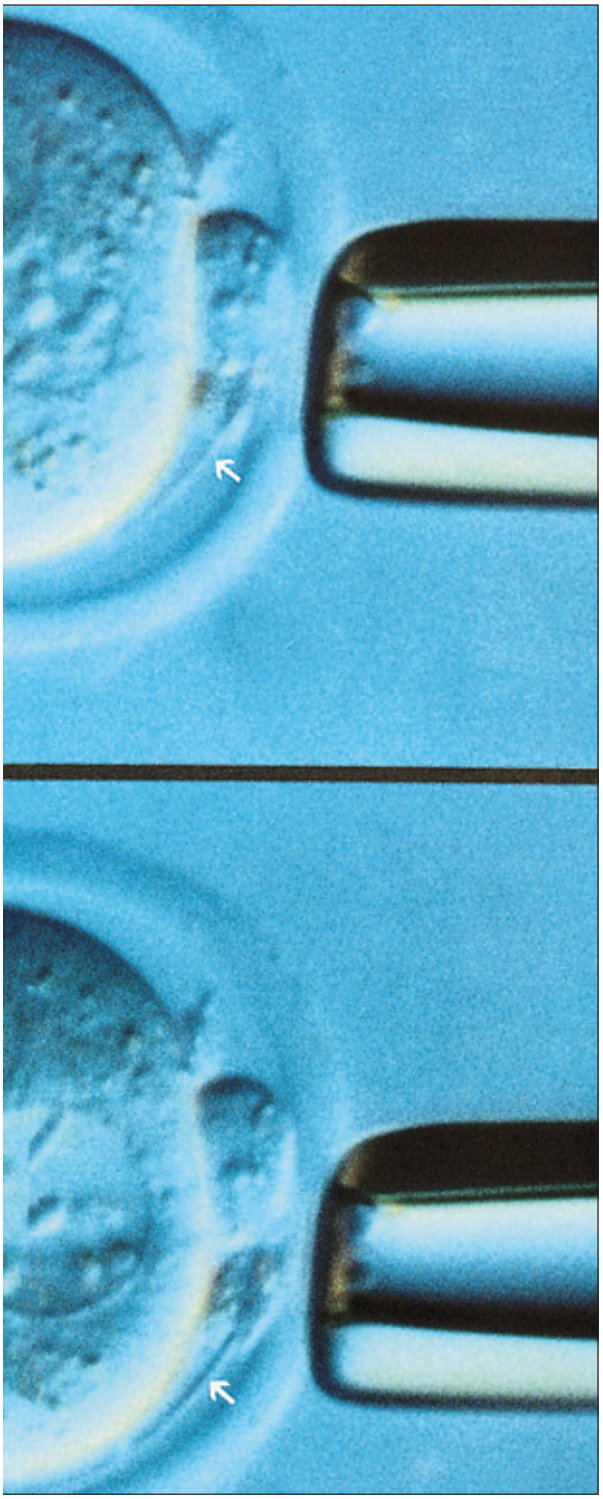
In the summer of 1980, Melvin Bosma, an immunologist at the Fox Chase Cancer Center in Philadelphia, began to examine four mice that were brought to his attention because of the peculiar results of their blood tests. The mice had none of the usual antibodies in their blood. In fact, they seemed to have no immune reactions at all.

The four mice turned out to be littermates, which suggested that their lack of immunity might be a genetic trait passed down from mother and father. Bosma soon realized that he and his coresearcher and wife, Gayle Bosma, had discovered an exceedingly rare and useful spontaneous mutation.

Named *scid* for *s*evere *c*ombined *i*mmuno*d*eficiency, the mutant mouse cannot make T cells or B cells, white blood cells



Laboratory of Ralph Brinster, University of Pennsylvania



that fight off infections and foreign implanted tissues. It is one of the best animal models ever found for studying the basic biology of the immune system as well as several diseases.

The *scid* mouse "was a gift of nature," Melvin Bosma says. The Bosmas belong to a generation of modern scientists who use various strains of mice to study fundamental life processes.

Mice are small, handy, and remarkably fecund: three months after her own birth, a female mouse can produce a dozen new babies. Mice live only two to three years, allowing researchers to follow disease processes from beginning to end in a relatively short time. And their genes are remarkably similar to those of humans, despite an evolutionary distance of 75 million years between the two species.

As a result, researchers have come to realize that they can use mouse genes as a blueprint for finding and studying human genes, including disease genes. They can also use mouse models to test drugs, devise novel therapies, and study the physiology and biochemistry of genetic diseases in ways not possible in humans.

To create a mouse that models a human disorder, researchers inject a solution containing many copies of a defective gene into a fertilized mouse egg held by a large pipette (above). The male pronucleus, the genetic material that came from the sperm, swells up from the injection (below). The remnants of the sperm's tail can be seen (arrows) between the egg and the protective layer surrounding it.

They no longer have to wait for Mother Nature. . .

Most of the mutant mice strains under study arose spontaneously, and many of them are housed at the Jackson Laboratory in Bar Harbor, Maine, which maintains over 450 strains of mice with a wide variety of genetic afflictions and apt names.

The "stargazer" mutant, for example, has a neurological disorder that forces it to throw back its head and look at the sky. "Twitcher," "shiverer," and "quaking" mice have damaged nerve fibers and abnormal gaits. "Dwarf" and "little" are lacking growth hormones.

At least 20 new mutants are identified each year. Using various breeding tricks, the scientists are able to perpetuate the mutant strains in stable colonies.

Some of these mice have versions of common human diseases, such as diabetes, whose two forms—juvenile (Type I) and late-onset (Type II)—afflict millions of people worldwide. According to diabetes expert Edward Leiter of the Jackson Laboratory, these mice give researchers the opportunity to tease apart the genetic and environmental factors that underlie the disease. When fed high-fat diets, for example, some strains of mice become glucose-intolerant (as in Type-II diabetes), while others do not. By comparing the two strains of mice, scientists may be able to find the genes responsible for the different responses to diet.

The discovery of the *scid* mouse has allowed scientists to range farther afield and, for example, produce the first animal model of AIDS. It was generally believed that mice could not be infected with AIDS because many human viruses, such as HIV, which causes AIDS, attack only humans and higher primates. But in 1988, Michael McCune, who was treating AIDS patients at San Francisco General Hospital, had an inspiration: Since *scid* mice do not have a working immune system, they cannot reject human tissue, he reasoned; then why not implant human immune-system tissues into them, to see if that would allow the AIDS virus to infect the mice?

It was a wild idea, he admits. But to the utter astonishment of all his colleagues at Stanford University, where the experiment was carried out, it worked.

McCune now routinely infects his *scid*-human mice with the AIDS virus. The mice are treated with the drug AZT at different intervals after infection to test the advantages of early treatment. Hundreds of other antiviral compounds are also being screened, either alone or in combination, to see if they will stop the fatal march of AIDS.

In similar ways, a variety of fetal and adult human tissues can be implanted into *scid* mice to study cancer, brain disorders, and other conditions, as well as to develop gene therapy. McCune and other scientists have actually grown human lung, pancreas, intestinal, and brain tissues in these animals. The goal of research by Charles Baum of Systemix, a Palo Alto biotechnology firm, and HHMI investigator Irving Weissman of Stanford, is to isolate the human blood stem cell—a master cell that gives rise to all types of human blood and immune cells—which scientists will need in large quantities to make gene therapy become practical.

Researchers still lack animal models for most human genetic diseases. But this is changing rapidly as scientists learn they no longer have to wait for Mother Nature to make their mutants—they can create them to order.

By inserting foreign genes into animal embryos, they can produce "transgenic" animals, whose cells follow the instructions of the interloper genes as well as those of their ancestral genes. The result is an explosion of new information on how genes work in specific cells and how they go about promoting health and disease in both mice and humans.

In 1982, Richard Palmiter, an HHMI investigator at the University of Washington in Seattle, Ralph Brinster of the University of Pennsylvania, and their colleagues injected a modified rat growth-hormone gene into a fertilized mouse egg. The gene normally produces small quantities of growth hormone in the pituitary gland, but the researchers wanted to see what would happen if an animal had unusually high levels of the hormone.

Before injecting the gene into the mouse egg, therefore, they attached it to promoters, regions of DNA that control which tissue expresses a gene. The promoters would redirect the gene's expression to other cell types, such as liver cells, where the gene would be freed from its normal controls and would produce large quantities of hormone. Then they implanted the egg into a mouse foster mother. The mother gave birth to normal-sized pups that grew at an unusually rapid pace to become giant mice, nearly twice the size of their littermates. The picture of one of these super-mice was splashed across newspapers and magazines throughout the world.

This experiment paved the way for the first attempt to cure a genetic disorder—dwarfism—in transgenic animals by gene therapy. The mice "patients" were undersized because they lacked sufficient growth hormone. By inserting a modified growth-hormone gene into them, the researchers and Robert Hammer, who was then at the University of Pennsylvania (he is now a senior associate in the HHMI unit at the University of Texas), corrected the genetic defect. The correction was so good that the mice grew slightly larger than normal.

Since that seminal experiment, geneticists have been striving to find permanent cures for a variety of genetic diseases. But first they have to understand the basic biology of the diseases.

Five research groups—including the team of Palmiter, Brinster, Richard Behringer, then an HHMI associate in Brinster's lab, Tim Townes of the University of Alabama at Birmingham, and their associates—have succeeded in creating various mouse models of sickle-cell disease. A painful and often lethal genetic disease, sickle-cell disease affects about 2 million people around the world, mostly blacks. One in twelve American blacks carries the abnormal gene. More than 50,000 Americans suffer from the disease, which develops in children who inherit the gene from both parents.

Sickle-cell disease was the first genetic disease to be understood at the molecular level—it results from the mutation of a single base in the DNA that codes for hemoglobin, the oxygen-carrying protein in red blood cells—yet there is still no effective treatment for it. Scientists have long been eager for an animal model on which to test new therapies.

To produce such a model, Richard Behringer injected a few hundred copies of the abnormal gene into fertilized mouse eggs. The pups born of these eggs were examined to see which, by chance, had incorporated the human gene into one or more chromosomes. Those with the gene were mated in the hope that the human gene had lodged in their egg or sperm and that the trait would be passed on to the next generation. By breeding these mice with a strain of mice that produced low amounts of mouse hemoglobin, the researchers have created mice with red blood cells that sickle.

However, these mice still have some of their normal mouse hemoglobin, Palmiter points out. Further experiments may yield mice with more abnormal human

Laboratory of Ralph Brinster, University of Pennsylvania

A giant mouse (left) grown from an egg injected with rat growth-hormone genes weighs nearly twice as much as its normal sibling. This experiment was a major step toward creating animal models of human disease.

hemoglobin—mice that may be more like sickle-cell patients. Researchers could then test anti-sickling drugs in the mice and explore methods of gene therapy.

In other laboratories, cancers of the eye, breast, lymph tissues, pancreas, and other organs have been induced in mice by cancer genes and combinations of cancer genes. This work is fueling one of the most important revolutions in twentieth-century medicine—the ultimate understanding of cancer as a genetic disease.

It is now believed that cancer is caused by genetic mutations—most often, by a series of mutations, some of which may be inherited. Certain normal genes involved in cell growth, development, and differentiation can be converted into cancer-causing "oncogenes" by mutation. Other genes that normally prevent the uncontrolled growth of cells—"suppressor" genes—can also produce cancer if they are knocked out by genetic mutations.

Single mutations are generally not sufficient to cause cancer, but they produce changes that predispose cells to malignant growth. Additional mutations in other genes continue the cells' malignant transformation. Thus, cancer is a multi-step process involving the interaction between genes and their environment.

To test the hypothesis, researchers have put a variety of oncogenes into mice, using promoters to direct the genes to specific tissues. In this way, HHMI investigator Philip Leder and his co-workers at Harvard created and patented "onco-mice"—animals that reliably develop breast and lymph cancers. Onco-mice are being used worldwide to test drugs and therapies against those two forms of cancer, says Leder.

Aside from cancer, cardiovascular disease is the biggest killer in developed countries. Several researchers are using animal models in an effort to uncover the genes responsible for high levels of fats and cholesterol in the blood, to develop tests for identifying people with a genetic predisposition to heart disease, and to explore new therapies.

Michael Brown, Joseph Goldstein, Robert Hammer, and their associates at the University of Texas recently created transgenic mice that never develop high levels of cholesterol in their blood, no matter what they eat, because they carry a human gene for receptor proteins that latch onto LDL, the carrier of "bad" cholesterol, and remove it from the bloodstream.

Most researchers make transgenic animals as Palmiter and Brinster do, by injecting foreign genes into fertilized eggs and then implanting the eggs in a surrogate mother. Others link the foreign genes to a virus and inject the viral combination into an embryo at a later stage of development; the virus then integrates itself into the animal's chromosomes, carrying the foreign genes along with it.

Either way, the researchers face a major problem: They cannot specify where in the animal's DNA the foreign gene will become integrated. If a gene is taken up at the wrong spot, it may disrupt a native gene or even cause a lethal mutation; if taken up in the right spot, it may cure a disease. Furthermore, one animal might integrate hundreds of copies of the gene, while another animal, under the same experimental conditions, might integrate but a single copy.

Another problem is that researchers can only add genes to the animal's own genome: They cannot get rid of an animal's unwanted or defective gene. The animal's own gene, equivalent to the inserted foreign gene, still functions in some form and could influence the experiment, as appears to have happened in mouse models of sickle-cell disease.

Recently a new technique has overcome these limitations and opened up a whole world of possibilities for making almost any wished-for animal model. It is called homologous recombination, or more loosely, gene targeting, and it is awesomely precise.

In homologous recombination, a desired gene finds an identical, or homologous, sequence of DNA in the animal's genome and swaps places with it. Several teams of researchers, including that of Oliver

Smithies, who is now at the University of North Carolina, have devised ingenious ways to achieve this precise exchange.

In 1987, Mario Capecchi and Kirk Thomas of the HHMI unit at the University of Utah showed that such strategies could be particularly effective when used with mouse embryonic stem (ES) cells. These cells, which were first isolated from very early mouse embryos by Martin Evans at Cambridge University and Gail Martin at the University of California, San Francisco, in the early 1980s, are unspecialized precursors of other cells. Each one is capable of giving rise to an entire animal.

When Capecchi introduced new genes into ES cells, most of the genes integrated into the chromosomes randomly. But in 1 out of 100 or 1,000 times, depending on the gene, the new gene found its exact counterpart in the mouse genome and integrated itself there, Capecchi says. It was as if the foreign gene had cruised up and down the mouse chromosomes and found its home address.

The challenge, then, was to find those cells in which the genes had landed on their home targets. If they could be identified, a powerful new genetic tool would be available.

Capecchi and his associates have pioneered a double-barreled system for selecting these rare cells. They start out by adding two other genes to the gene that is seeking its homologous site in an ES cell.

One of the added genes (the herpes virus thymidine kinase, or *tk*, gene) makes a cell vulnerable to the antibiotic gancyclovir. It is placed just outside the homologous sequence, so when the inserted DNA hits its precise target the *tk* gene is discarded because its sequence does not match up. The cells in which homologous recombination has occurred, therefore, remain invulnerable to gancyclovir, but those in which the DNA integrated randomly still have the *tk* gene and are killed by the antibiotic.

The second added gene, which confers resistance to the antibiotic neomycin, is placed right in the middle of the homologous sequence. When the cells are bathed with neomycin, those that failed to integrate the homologous gene are weeded out.

The cells that survive both treatments have the homologous gene right on target. These cells can then be used to recreate a mouse. Whole generations of mice can be raised from these cells, permanently carrying the inserted gene in exactly the right spot.

Homologous recombination allows scientists to carry out a new type of research with transgenic mice: knockout experiments. By introducing defective mouse genes into ES cells, researchers knock out the native gene, and the resulting mouse expresses the defective gene.

The technique should prove important for studying human diseases, such as cystic fibrosis, which until now have had no animal models. But there's a rub, says Capecchi. To do knockout experiments, researchers must already have cloned the human disease gene and its mouse homologue. So far, only a few hundred such genes have been deciphered.

Nevertheless, scientists are reveling in their new freedom to manipulate genomes in many ways.

"With gene targeting, you can knock out a gene, replace it, or change it more subtly," says Capecchi. "For example, you can put a gene under the control of a switch so that if you inject a certain drug, the gene is turned on—or turned off." Almost any disease process can now be studied in animals in this way.

Mice remain the favorite subjects for such studies—not only because they are convenient to work with, but because recent advances in mapping both the mouse and human genomes have made comparisons between the two especially fruitful. Scientists who work with transgenic mice confess astonishment and glee at the rapid progress in the field. The ability to add or replace mouse genes at will is leading to a medical revolution, they say, and new experiments are limited only by the human imagination.

Gene targeting is awesomely precise.

THE FUTURE OF GEN

by Harold M. Schmeck, Jr.

The eminent British molecular biologist Sydney Brenner once got a hearty laugh from his audience by describing how some future graduate student will define a mouse—"ATC,GCC,AAG,GGT,GTA,ATA…"

But every year the idea of defining an organism by the sequence of its DNA bases seems a little less farfetched.

Victor McKusick, of The Johns Hopkins University School of Medicine, notes that scientists' growing ability to read and write in the language of the genes has already explained some of the once-mysterious basic concepts of genetics.

The difference between dominant and recessive traits as causes of genetic disease used to be just an abstraction based on a great deal of observation. If a genetic defect expressed itself only in patients who inherited the trait from both parents, it was called recessive; both copies of the gene coding for the trait were presumably defective, resulting in disease. If the trait was dominant, on the other hand, it meant that one defective copy of the gene was sufficient to spell disaster. But why should some disorders require two mistakes, while others resulted from only one? Molecular biology has given a concrete and remarkably simple explanation.

"It now appears that these two categories (recessive and dominant) correspond pretty closely to the two fundamental categories of proteins: enzymatic and structural," McKusick said in a review of genetics research. Recessive disorders tend to result from failures in genes that code for enzymes, the biological catalysts that do much of the body's chemical work. A person who has inherited the defective gene from only one parent often goes disease-free because the normal gene inherited from the other parent produces enough of the enzyme to serve the body's needs. The disorder appears only when the person inherits the same defect from both parents and therefore lacks any working copy of the normal gene.

If the genetic defect affects structural proteins, however—for example, collagen, a key component of connective tissues and bones—only one copy of the faulty gene is usually enough to cause disease. It is easy to see why. A four-engine airplane can still fly even if one of its engines fails, as long as the other engines provide enough power, but a single faulty strut that makes a wing fall off will cause the plane to crash.

The reason some genetic disorders are relatively common while most are extremely rare has also proved to be almost ridiculously obvious. The bigger the gene, the greater the chance that something will go wrong with part of it. In many cases, it seems as simple as that.

ETIC RESEARCH

Sometimes rather subtle differences in the defects of a single gene can make a profound difference in a patient's fate, as Louis Kunkel of the HHMI unit at Harvard University learned after he and his team discovered the gene for Duchenne muscular dystrophy (DMD) in 1986. Major flaws in that huge gene result in the presently incurable DMD, a muscle-wasting disease that leaves young boys wheelchair-bound by age 12 and may kill them by age 20 because the muscles that control breathing fail. By contrast, lesser defects in that same gene produce a much more benign disease, Becker's muscular dystrophy.

A year after discovering this gene, the team identified the protein it codes for—a previously unknown protein, now named dystrophin, which occurs in muscles in such small amounts that it would never have been found by ordinary means. Dystrophin plays a key role in muscle cells and may be involved in many other muscle diseases. Researchers are now analyzing how dystrophin functions, what other proteins it interacts with, and whether it might be replaced to interrupt the course of disease.

Experts see many more insights such as these in the future, as research in molecular genetics opens some of the "black boxes" of biology.

"I think we are going to have an explosion of understanding," says David Valle, of the HHMI unit at The Johns Hopkins University. For example, the causes of mental disorders certainly include environmental factors, but biological psychiatrists believe the genes are whispering an important message, if only it can be heard.

Genetic research will illuminate many disorders of single organs such as the eye, teeth, skin, and cochlea (the hearing apparatus of the ear), Valle believes. The deafness of about two-thirds of patients with serious hearing problems has a genetic basis, he says. Molecular biologists can find genes that are expressed only in the cochlea and therefore are probably important in hearing. Once such genes have been identified, several strategies exist for determining their functions and suggesting treatments.

Valle's current research focuses on a rare genetic disorder of the eye, gyrate atrophy, which leads to blindness through degeneration of the retina. The basic fault is an enzyme defect that causes an abnormal buildup of the amino acid ornithine. Surprisingly, some 35 different mutations in a single gene are able to produce the disease. The excess ornithine is found almost everywhere in the body—blood, urine, tears, spinal fluid—but the serious ill effects are limited almost entirely to the retina. As yet, nobody knows why.

Understanding the genetic cause of the

disease has led to a medical treatment that seems effective: severely restricting the patient's diet to bring the ornithine levels down to nearly normal. Scientists have compared the effects of this treatment on children in whom it was started early and on siblings who did not receive it until an older age. The studies confirm that the dietary restriction minimizes damage to the retina, Valle reports. But the diet is only a stopgap solution. Geneticists are searching for more effective remedies, including possible treatment for the gene defect itself.

"One of the really exciting things about modern molecular genetics is that we now have opportunities to make animal models of these diseases and to study what happens at the tissue level in a direct way," Valle says.

Philip Sharp, director of the Center for Cancer Research at MIT, divides the benefits of genetics research into two categories: those that generate knowledge and those that generate treatment. He sees animal models as extremely important to both. Deliberately produced genetic diseases in animals will have pathologies like those of human diseases. "We will learn how to recognize them, treat them, and analyze them in animals," he says. "That is going to be the forefront of biomedical science, in one area of it at least."

In addition, many aspects of human development will be clarified by work with mice, flies, and worms, he says. Scientists have discovered that genes which are developmentally active in both *Drosophila*, the fruit fly, and *C. elegans*, the nematode worm, have direct counterparts in mammals, although the functions of these genes in humans are not yet entirely clear.

Louis Kunkel and his team found the genetic flaws that cause two forms of muscular dystrophy.

When genes of species that separated from each other many millions of years ago show so much similarity, there is every reason to believe they are related. Many molecular biologists have noticed that nature is quite frugal in preserving devices that have proved biologically effective. As an example, Valle points out that the human enzyme ornithine delta aminotransferase, which is defective in gyrate atrophy, is 54 percent identical to the comparable enzyme that functions in yeast.

"I think one of the real themes of biology is that Mother Nature uses things over and over again once she figures out how to solve a problem," Valle says.

This concept offers scientists a great opportunity, says Eric Lander of the Whitehead Institute. He thinks there is hope of compiling, eventually, a complete thesaurus of protein parts that function in Earth's myriad species. "That would be spectacular," Lander says. "If we had the thesaurus of all the moving parts, then we would understand life in a remarkable way."

Gene mapping and cloning are key to the assembly of the thesaurus, and progress in these areas is clearly accelerating. Howev-

er, most of the 50,000 to 100,000 human genes remain totally unknown, and there is still a long way to go.

To date, most of the progress in understanding the genetics of human disease has involved relatively rare conditions, such as cystic fibrosis or Duchenne muscular dystrophy, that are caused by errors in single genes. But science is also stalking the genes that contribute to heart disease, cancer, diabetes, and mental illness—the big killers and cripplers of mankind.

It may soon be possible to tell some people that they have certain genetic predispositions to a specific major illness and suggest that they tailor their lifestyles accordingly. Similarly, the use of drugs to treat some of the important diseases could be tailored to the genetically varied needs of patients, with benefits for them and for the health care system in general—"Different strokes for different strokes," as one scientist put it.

"Different strokes for different strokes"

On the other hand, some scientists fear that people might be stigmatized or become uninsurable because of genetic traits, such as carrier states, that don't in themselves have any appreciable effect on health.

Genetic research is advancing steadily, often rapidly, on many fronts. It has long been known that some disorders affect males, others affect females primarily, while still others may appear in either sex. But a few years ago researchers discovered that, even in some of the latter disorders, the gravity and sometimes even the nature of disease may depend on which parent provided the faulty gene. This phenomenon is called imprinting. Although it has been detected only in rare human conditions, imprinting is a subject of intense study as researchers look for other examples.

Other scientists have forsaken the genes that reside in cell nuclei and are finding new clues to disease in the genes of what are probably our oldest and most entrenched "parasites"—the mitochondria, tiny, energy-generating organs inside every cell. Mitochondria are thought to be the descendants of ancient bacteria that not only found a home in animal cells, but also adapted so thoroughly that they became indispensable functional parts of those cells.

We inherit mitochondria only from our mothers; sperm leave their mitochondria behind when they enter the egg. Flaws in mitochondrial genes have been found to lead to certain types of blindness and epilepsy and may also contribute to some degenerative disorders, such as dementia, that are associated with aging.

"Mitochondrial DNA gives us a whole new way to think about genetic transmission of diseases," said Douglas Wallace of Emory University, a specialist in those vital intracellular power stations.

In even more fundamental ways, discoveries in genetics have led to novel strategies for treating disease. DNA is mainly the archive of genetic information. Its orders are translated into action by segments of ribonucleic acid (RNA), which serve as the working blueprints for all proteins. Today, chemists are beginning to create valuable new drugs by fabricating "anti-sense" segments of RNA, whose sequence is the exact opposite of an unwanted sequence, to combine with certain existing strands of RNA and thus block the action of specific genes.

The bottom line in any kind of biomedical research lies in the realm of treatment and prevention. The ultimate step in that direction is gene therapy—the deliberate transplantation of genes to treat or even prevent human disease. Many geneticists dismiss gene therapy as a distant prospect; others disagree.

Though gene therapy has not been too successful so far, it adds to the excitement that pervades genetics today. Studies of microbes, plants, animals, and many normal human beings are all contributing to the explosion of new knowledge. In recent years, molecular genetics has given important insights into the origin of life and its evolution, the emergence of humans, and our intimate relatedness to every other species on Earth. We can expect many more advances as geneticists continue to explore the wonder of life.

Progress Continues...

Much of the excitement in genetics today comes from its lively offspring, "genomics." This newcomer specializes in large-scale analyses of all the genetic material in the genomes of organisms ranging from bacteria to mammals. Genomics is expected to provide the functional meaning of newly revealed DNA sequences: What do these genes really do? And that precise knowledge, in turn, heralds a revolution in the diagnosis, monitoring, and treatment of diseases.

Meanwhile, new DNA sequencing techniques using high-speed robots are flourishing. So are ingenious combinations of YACs (yeast artificial chromosomes), BACs (bacterial artificial chromosomes), PACS (fragments of DNA in a vector derived from a bacteriophage known as P1), and MACs (mammalian artificial chromosomes), which supply genetic fodder for the machines that do the sequencing.

A Gifted Young Patient Battles Cystic Fibrosis

Jeff Pinard, the young man with cystic fibrosis who was seeking his own genetic flaw as a college student, is now 29 and doing quite well. In May of 1998 he married Darcy Hartley, who is studying for her master's degree in elementary education. They live in a house of their own in Grand Rapids, Michigan. He does computer work for an electric utility, mostly from home. But he has had some very rough times.

Several bad bouts of illness landed him in the hospital for months and forced him not only to give up his research work, but to drop out of the University of Michigan. After going home to his parents in Grand Rapids, he was hospitalized again because of pancreatic problems that produced excruciating pain. These problems remain unsolved, despite a variety of treatments. However, Pinard feels generally better and, according to his mother, "he's had some periods of time when he has been without pain." How long? "Oh, a couple of weeks at a time ...."

Lap-Chee Tsui of the University of Toronto finally identified and sequenced Pinard's second, milder CF mutation, adding it to the list of more than 850 known mutations in the CF gene. But it is difficult to count CF mutations these days, Tsui says, because "many mutations are now found in atypical diseases, such as male infertility and pancreatitis."

A major study of pancreatitis led by Jonathan Cohn of Duke University Medical Center and published in the *New England Journal of Medicine* recently concluded that many adults who suffer from so-called "idiopathic" pancreatitis (pancreatitis of unknown cause) actually have cystic fibrosis. The authors add that these findings "will change how physicians treat patients with this condition."

A Natural Antibiotic

The major symptom of CF is lung infection. In 1996, Michael Welsh, an HHMI investigator who teaches medicine and physiology at the University of Iowa, discovered why these infections occur—and offered a new approach to treatment that is still being developed. Welsh had a longtime interest in epithelia, the sheets of cells that line the internal and external surfaces of the body, including those lining the airway. When the CFTR gene was identified, he was among the first to examine the role of the protein made by this gene.

"We found out that it's actually a chloride channel, through which salt moves across the membrane," Welsh says. "That was very satisfying, because then you could begin to tie together the physiology—defective epithelia—and the gene product, the chloride channel." Then his team made an intriguing discovery: Normal epithelial tissue can kill a large number of bacteria, while similar tissue from people with CF fails to do so, or even allows the bacteria to multiply.

Welsh guessed that the fluid covering the airway normally contains factors such as defensins, molecules that are part of our innate, nonspecific defense system. He wondered whether these were also present in people with CF. To his surprise, he found natural antibiotic substances both in healthy people and in those with CF. In CF patients, however, the substances' activity was greatly reduced by the abnormally high salt concentration resulting from the defective CFTR channel. When Welsh lowered the salt concentration, even the epithelia of CF patients became able to kill bacteria.

The team's conclusion: Drugs that reduce the salt concentration in airway fluid may help treat or prevent the sometimes fatal lung infections of CF patients. Other antibiotic drugs that resemble defensins may also be developed for this purpose.

As for his experiments with gene transfer, "they remain just that—experiments," says Welsh. "We can deliver the normal CFTR gene, but we cannot deliver it efficiently enough," he explains. "The problem is the delivery. We need to go back to the lab and try to make it work better."

Genetic Screening

"We are slowly moving closer and closer to implementation of genetic screening for CF," says Arthur Beaudet. New recommendations about such screening emerged from an NIH Consensus Development Conference held in the spring of 1997. The consensus panel declared that all expectant couples and those planning to have children should be offered the option to test for gene mutations that cause CF; so should couples in which one or both partners have a family history of the disease. The panel further recommended that insurance cover the cost of these tests.

Several labs around the country now provide such tests. "And we can get 50 different mutations on a single test for as little effort as one," Beaudet says. He adds that the tests have become so sensitive that "somewhat over 90 percent of CF carriers would be correctly identified." Therefore the tests would detect more than 81 percent of the couples at risk. Beaudet believes that newly married couples should be given a set of prepared mouth swabs to take home, so that they can test themselves at their leisure. Further tests would then be necessary only if both members of the pair are carriers of CF.

About 40 specialized centers worldwide offer *in vitro* fertilization to avoid genetic diseases. At the Illinois Masonic Hospital in Chicago, for example, Charles Strom and Yuri Verlinsky screen the eggs of mothers who are carriers of CF before fertilizing them with the husband's sperm. In this analysis they use only the eggs' polar bodies, which would be cast off anyway, Strom explains. If the test indicates the egg is free of the CF mutation, the doctors proceed with fertilization. "More than 16 healthy children have been born to CF carriers with this method," Strom reports. The method has now been extended to a variety of genetic diseases, including hemophilia, thalassemia, and sickle cell anemia.

New Findings About Brain Disorders

There was great rejoicing when Nancy Wexler's quest for the cause of Huntington's disease finally succeeded in 1993, after nearly eight years of effort described as "a nightmare of false leads, confounding data, and backbreaking work." The faulty gene was named *huntingtin.* The guilty mutation turned out to code for an extra-long, repeated stretch of glutamine, an amino acid in huntingtin, the protein made by this gene. But no one knew how the expanded glutamine repeats cause brain neurons to sicken and die.

Several other "triplet-repeat" diseases are known. They all attack some part of the nervous system, and all of them are still mysterious. Scientists have begun to search for clues to the function of huntingtin in the proteins that interact with it.

"The beauty of having the Huntington's disease gene in hand is that we are now able to place it in animals and learn its effects," says Wexler. In 1996, Gillian Bates and her team at Guy's Hospital, London, put fragments of the human HD gene into mice for the first time. The mice developed HD-like symptoms two months after birth and died soon afterward.

Fruit flies were also enlisted in the fight. In 1998, George Jackson of UCLA's Department of Neurology inserted fragments of the HD gene into the large nerve cells in the eye of a fruit fly. He found that, just as in human beings, the cells' fate depended on the number of glutamine repeats in the HD gene's DNA. The eyes of flies whose gene had only two repeats remained normal. Those with 75 repeats were normal for a month, but then began to degenerate slowly. When the flies had 120 repeats, their eyes suffered massive cell destruction.

Wexler is greatly encouraged by these findings. She points out that "the fly eye is a perfect laboratory to test the effects of drugs that will protect the eye and prevent degeneration." And because of many similarities between HD and other neurodegenerative conditions, including Alzheimer's, and Parkinson's diseases, scientists hope that the findings from one of these areas will advance research in the others.

The Viking Genes

Other gene defects uncovered in recent years include mutations predisposing people to such widespread ailments as breast cancer, familial polyposis of the colon, Alzheimer's, and Parkinson's.

Many family groups have helped in these searches. Scientists now look forward to working with the biggest genetic trove of all—the Viking gene pool, which can be found in very pure form among the 170,000 people of Iceland. Some of these families can be traced back for hundreds of years, and their records will soon be available to researchers. As DNA-based biology expands, so does the need for large groups of people with detailed and accurate family trees.