Genetic Basics

National Institutes of Health
National Institute of General Medical Sciences

What Is NIGMS?

The National Institute of General Medical Sciences (NIGMS) supports basic biomedical research that is not targeted to specific diseases. NIGMS funds studies on genes, proteins, and cells, as well as on fundamental processes like how cells communicate, how our bodies use energy, and how we respond to medicines. The results of this research increase our understanding of life and lay the foundation for advances in disease diagnosis, treatment, and prevention. NIGMS also supports research training programs that produce the next generation of biomedical scientists, and it has special programs to encourage underrepresented minorities to pursue biomedical research careers. NIGMS supported the research of most of the scientists mentioned in this brochure.

Genetic Basics

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute of General Medical Sciences

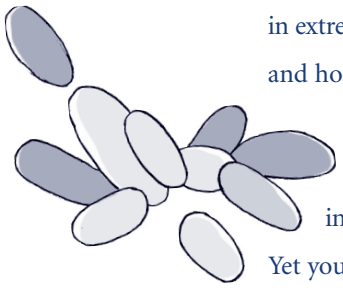
NIH Publication No. 01-662
May 2001
www.nigms.nih.gov

Written by Tabitha M. Powledge under contract 263-MD-817448

Produced by the Office of Communications and Public Liaison
National Institute of General Medical Sciences, National Institutes of Health

Contents

A Science Called Genetics

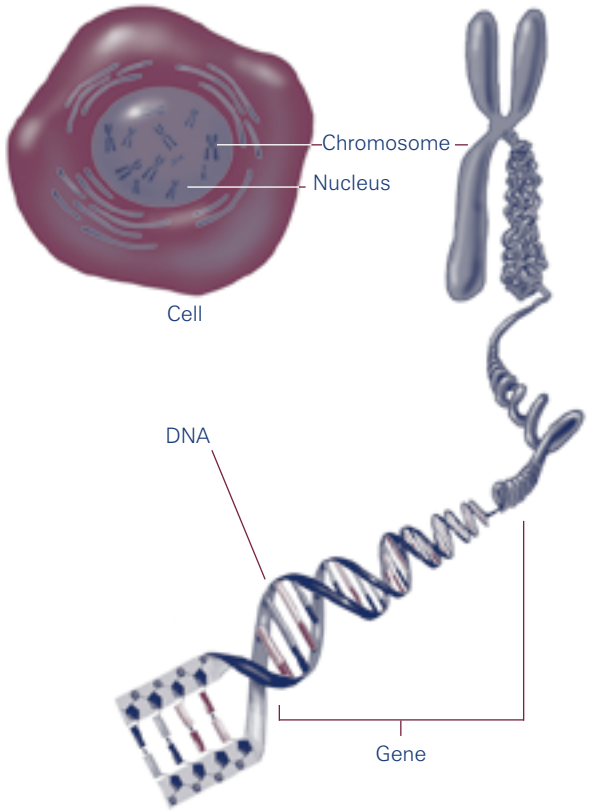


Consider just three of Earth's inhabitants: the bright yellow daffodil that greets the spring, the tiny organism called *Archaea* that lives in extreme environments such as boiling hot springs and hot water vents on the ocean floor, and you. Even a science-fiction writer inventing a story set on a distant planet could hardly imagine three more different forms of life. Yet you, the daffodil, and *Archaea* are related. Indeed, all of the Earth's billions of living things are kin to each other.

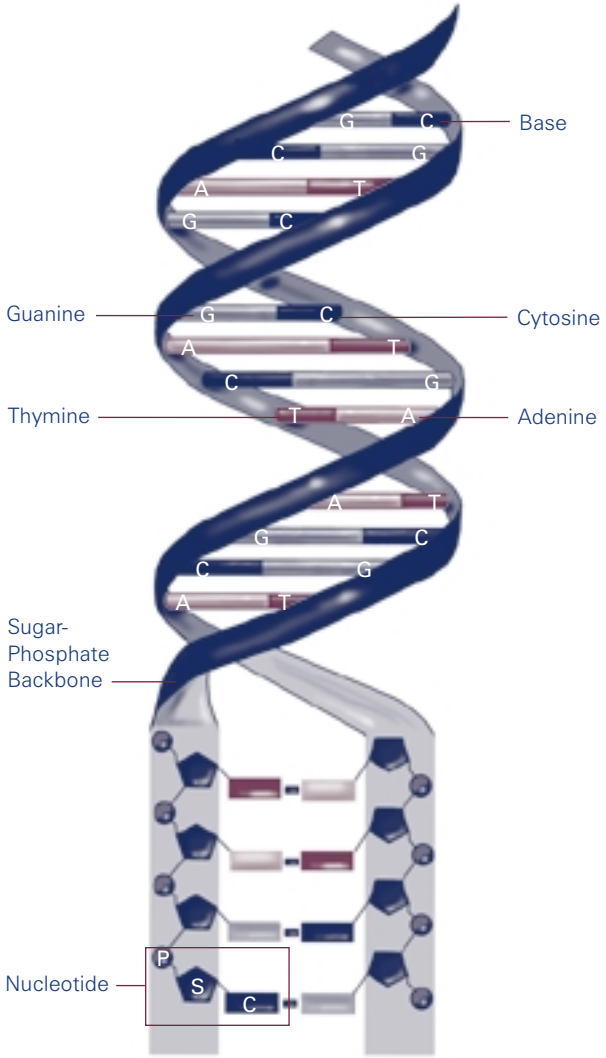
How did we and our very distant cousins come to look so different and develop so many different ways of getting along in the world? A century ago, researchers began to answer that question with the help of a science called genetics. When genetics first started, scientists looked at one gene—or a few genes—at a time. Now, it's possible to look at all of the genes in a living creature at once. This new, "scaled up" genetics is called genomics.

Today's genetics and genomics investigate how a cell's genetic material affects what goes on inside it. Chemical reactions within cells are ultimately what determine an organism's physical characteristics. These reactions are governed in part by genes and in part by the environment. Scientists have only begun to grasp the near-unimaginable intricacy of the complex dance of genes and the environment that results in a daffodil, a hot springs life form—or you.

In most organisms, the genetic material that affects what goes on inside cells is deoxyribonucleic acid, DNA for short. DNA is rather like a vast library stored on structures called chromosomes inside the cell nucleus. You can think of a gene as one book in that library and of a chromosome as a bookcase that holds thousands of books.



▲ Relationship among the cell, the nucleus, a chromosome, DNA, and a gene. Note that a gene would actually be a much longer stretch of DNA than what is shown here.



▲ DNA consists of two long, twisted chains made up of nucleotides. Each nucleotide contains one base, one phosphate molecule, and the sugar molecule deoxyribose. The bases in DNA nucleotides are adenine, thymine, cytosine, and guanine.

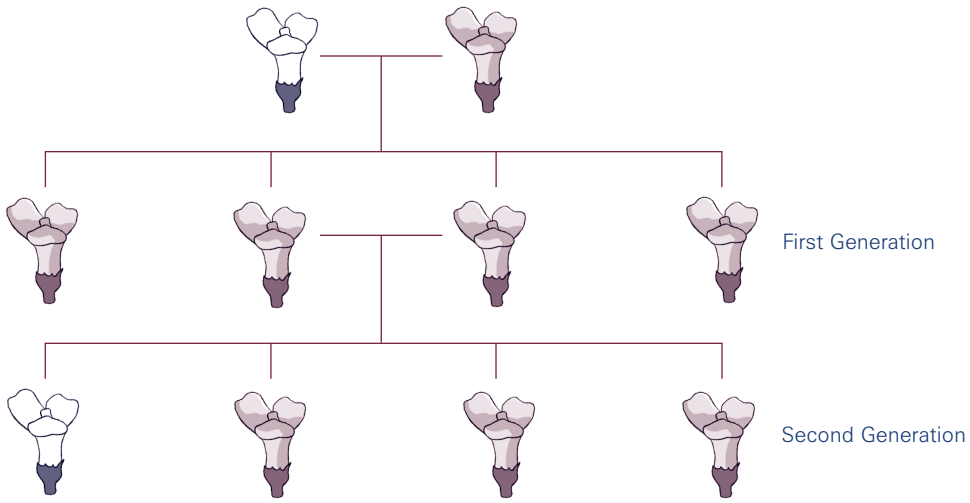
But these genetic library books are written in code. The code contains instructions that tell cells what to do. The DNA code is written in an alphabet of just four chemical "letters" known as bases. Bases are part of larger structures, called nucleotides, that form the building blocks of DNA. Even though there are just four bases—adenine, thymine, cytosine, and guanine, abbreviated A, T, C, and G—they can be strung together in billions of ways. That means billions of different coded instructions can be sent to cells. And if billions of these instructions are possible, that begins to explain how you can be so very different from a daffodil and an *Archaea*, and yet still be related to them.

How Genes Work

People have long known that living things inherit traits from their parents. That common-sense observation led to agriculture, the purposeful breeding and cultivation of animals and plants for desirable characteristics, which began 10,000 or more years ago. But exactly how traits are passed to the next generation was a mystery until the beginning of the 20th century.

In 1900, three European scientists independently found an obscure research paper that had been published nearly 35 years before. Written by Gregor Mendel, an Austrian scientist who was also a monk, it described a series of experiments he had carried out on ordinary garden pea plants.

Mendel had studied how pea plants inherited seven different, easy-to-see traits (for example, white or purple flower color and smooth or wrinkled peas). Mendel counted many generations of pea offspring and discovered that these traits are inherited in orderly, predictable ratios. When he cross-bred purple-flowered pea plants with white-flowered ones, the next generation had only purple flowers. But the white-flower trait was hidden somewhere in the peas of that generation, because when those plants were bred to each other, their offspring displayed the two flower colors again. Furthermore, the second-generation plants displayed the colors in specific ratios: On average, 75 percent of the



▲ Mendel found that his peas inherited individual traits such as flower color in a particular way. When he bred purple-flowered pea plants with white-flowered ones, the next generation had only purple flowers. But in the generation after that, white flowers reappeared. He realized that each plant must carry two "factors" (we now call them genes) for flower color, one from each parent. Breeding a pure purple-flowered plant with a white-flowered one would generate plants with a white factor and a purple factor, but the purple factor was dominant over the white factor, and so all the flowers in the first generation appeared purple. In the next generation, white-flowered plants reappeared because, statistically, one in four of the plants would inherit two white factors.

plants had purple flowers and 25 percent of the plants had white flowers. And those same ratios persisted, generation after generation.

Mendel concluded that the reproductive cells of his pea plants contained discrete "factors," each of which specified a particular trait, such as white flowers. The factors also passed from parent to offspring in a mathematically orderly way. After the 20th century scientists unearthed Mendel's paper, the "factors" were named genes.

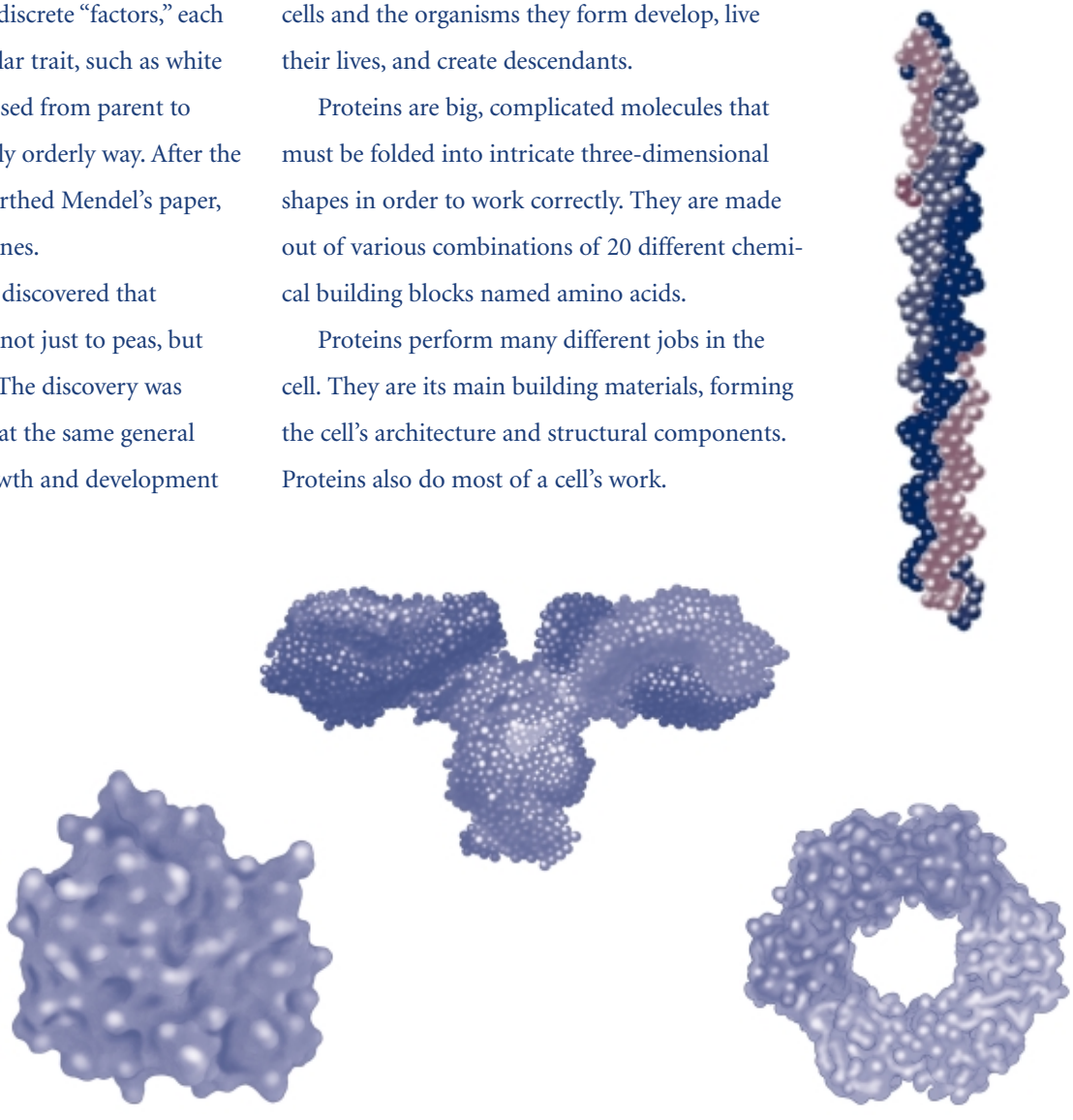
Early geneticists quickly discovered that Mendelian genetics applied not just to peas, but also to poultry and people. The discovery was momentous. It suggested that the same general principles governed the growth and development of all life on Earth.

From Genes to Proteins

So genes do their work by influencing what goes on inside cells. How do they exert that influence? They do it through proteins. Thanks to proteins, cells and the organisms they form develop, live their lives, and create descendants.

Proteins are big, complicated molecules that must be folded into intricate three-dimensional shapes in order to work correctly. They are made out of various combinations of 20 different chemical building blocks named amino acids.

Proteins perform many different jobs in the cell. They are its main building materials, forming the cell's architecture and structural components. Proteins also do most of a cell's work.



▲ Different protein shapes.

Some proteins, known as enzymes, carry out the thousands of chemical reactions that go on in a cell. Enzymes help make other molecules, including DNA. Enzymes also break food down and deliver and consume the energy that powers the cell. Other kinds of proteins, called regulatory proteins, preside over the many interactions that determine how and when genes do their work and are copied. Regulatory proteins also supervise enzymes and the give-and-take between cells and their environment.

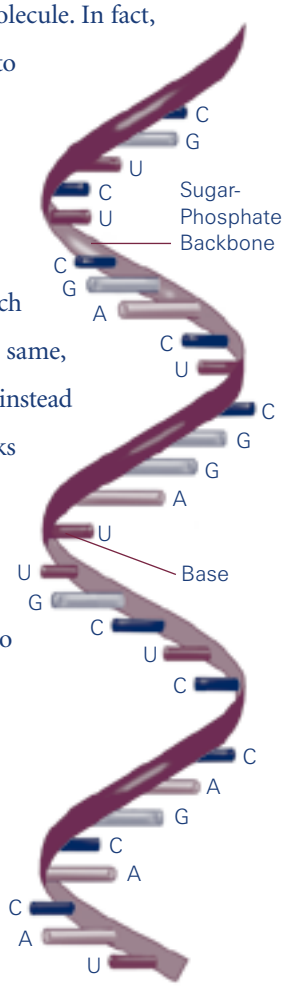
To perform its many functions, a cell constantly needs new copies of proteins. Although proteins do lots of jobs well, they cannot make copies of themselves. To make more proteins, cells use the manufacturing instructions coded in DNA.

The DNA code of a gene—the sequence of its "letters" A, T, C, and G—spells out the precise order in which the amino acids must be strung together to form a particular protein. Sometimes there is a mistake in those instructions, a kind of typographical error. This mistake is called a mutation. A mutation is simply a change in the DNA sequence. Such a change can cause a gene to work incorrectly, or even not work at all. The result is an abnormal protein, or perhaps no protein. But not all mutations are harmful. Some have no effect, and other mutations produce new versions of proteins that may give a survival advantage to the organisms that possess them. Over time, these types of mutations drive the evolution of new life forms.

Remarkable RNA

How does DNA make proteins? It doesn't. DNA is just a collection of instruction manuals. The instructions are carried out by ribonucleic acid (RNA).

RNA is a remarkable molecule. In fact, many scientists have come to believe that RNA appeared on the Earth long before DNA, meaning that RNA is actually DNA's ancestor. RNA is chemically very much like DNA—its bases are the same, except that it has uracil (U) instead of thymine—but RNA looks quite different. DNA is a rigid, ladderlike molecule that is very stable. RNA is flexible; it can twist itself into a variety of complicated three-dimensional shapes. RNA is also unstable. Cells constantly break RNA down and replace it.

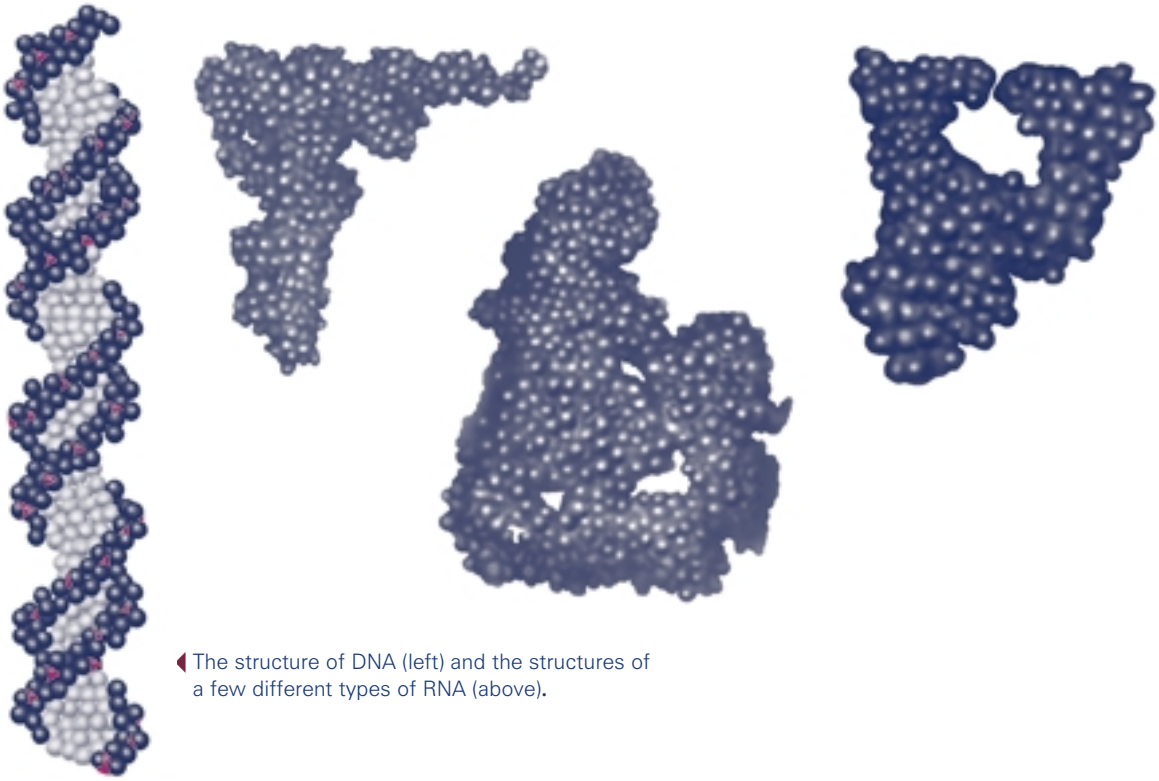


▲ Ribonucleic acid (RNA). RNA has the bases adenine (A), cytosine (C), guanine (G), and uracil (U) instead of the thymine that occurs in DNA.

This means that cells can change their patterns of protein synthesis very quickly in response to what's going on around them.

Genes make their proteins in two major steps. The first is transcription, where the information coded in DNA is copied into a molecule of RNA whose bases are complementary to those of the DNA. ("Complementary" means that the RNA has a U where the DNA has an A, an A where the DNA has a T, a G where the DNA has a C, and a C where the DNA has a G.) The second is translation, where the information now encoded in RNA is deciphered (translated) into instructions for making a protein. Proteins are then manufactured in cell structures known as ribosomes. The manufacturing process occurs in the cytoplasm, which is everything in the cell outside of the nucleus.

Several types of RNA play key roles in protein production. Messenger RNA (mRNA) is what gets translated into protein. It is literally a messenger, bringing information from the DNA in the nucleus to the ribosomes in the cytoplasm. Ribosomal RNA (rRNA) helps build the ribosomes that make proteins. Transfer RNA (tRNA) carries amino acids to a protein under construction. Newly made RNAs are usually incomplete molecules that must be processed before they are ready to leave the nucleus for the cytoplasm and begin working.



The structure of DNA (left) and the structures of a few different types of RNA (above).

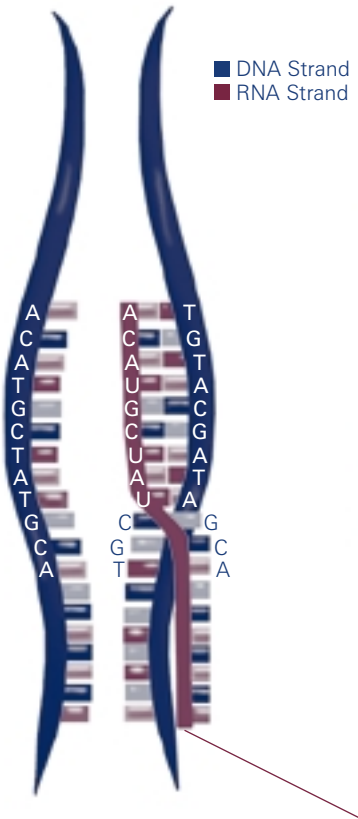
Controlling Genes

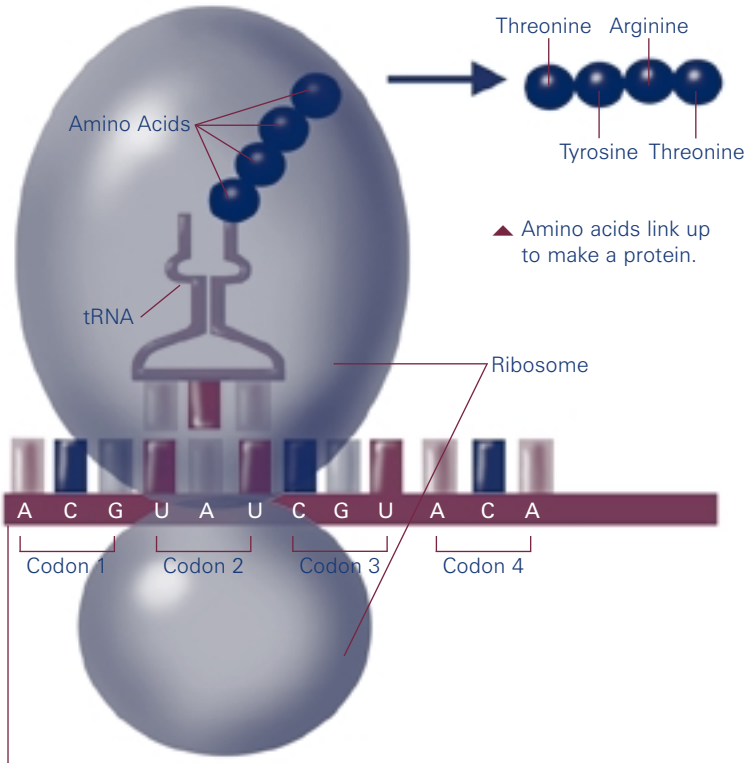
Every cell in an organism contains the same set of instructions encoded in DNA. How, then, can a brain cell be so different from a heart cell, and perform an entirely different job? These different cell types and their different tasks are possible because each cell "turns on," or expresses, only a subset of its total genes, the subset appropriate for running that particular cell at that particular moment. Everything a cell or organism does relates to genes that are turned on or off at any one time.

What turns a gene on—what allows it to provide the instructions for making a protein—is the cell's transcription apparatus. It consists of an enzyme called RNA polymerase plus a set of helper proteins called accessory factors. RNA polymerase makes an RNA copy, basically a working blueprint of a gene, which is then translated into a protein.

DNA

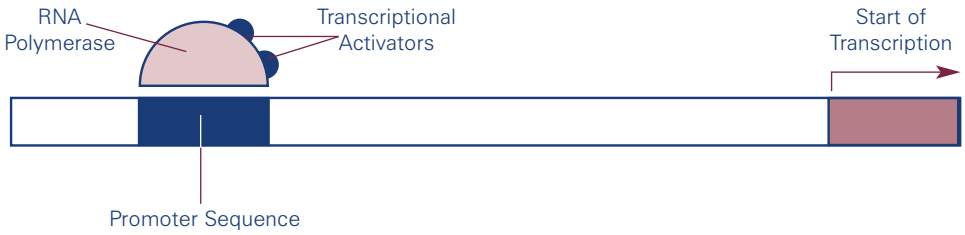
▲ The RNA polymerase II holoenzyme (not shown) transcribes DNA to make messenger RNA (mRNA). The mRNA sequence is complementary to the DNA sequence.





▲ Amino acids link up to make a protein.

▲ On ribosomes, transfer RNA (tRNA) helps convert mRNA into protein.



Initiation of transcription by RNA polymerase.

How does the cell know which working blueprints to turn on and which to turn off? It knows this through the action of proteins called transcriptional activators that attach themselves to the beginning of a gene, in a region known as the promoter. The transcriptional activators, in turn, recruit other helper proteins (called the transcription apparatus) to complete the job of gene activation.

Until 1994, scientists didn't know exactly what this transcription apparatus was. Then, Richard Young and his colleagues at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, discovered a previously unknown gene-reading machine called the RNA polymerase holoenzyme. Gene regulation turned out to be a collaboration between transcriptional activators and this holoenzyme.

It's a collaboration because the RNA polymerase holoenzyme contains nearly 100 protein components that recognize the presence of a transcriptional activator protein and decide whether or not to make a working blueprint—RNA—from the associated gene.

In addition to revealing details of gene transcription, study of the RNA polymerase holoenzyme may end up having direct application to human disease. Researchers have discovered that abnormalities in some of the RNA polymerase holoenzyme's components are linked to a variety of disorders, including one type of mental retardation and several cancers, among them breast cancer.

"How does the cell know to turn on these 1,000 genes and turn off those 820? We just don't know that," Young says. "We don't know globally how regulation occurs because we don't have a description of the set of genes in the entire genome that are on or off at any one time." (A genome is all of an organism's genetic material.)

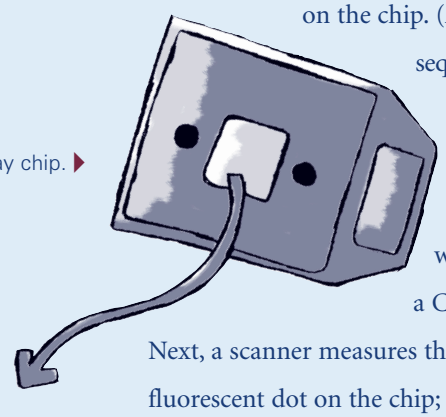
Young has set out to answer these questions. That puts him in the vanguard of the next giant step in genetics: the ability to take a true snapshot of everything a cell is up to at a single moment in time.

The Tools of Genetics: Gene Chips and Microarrays

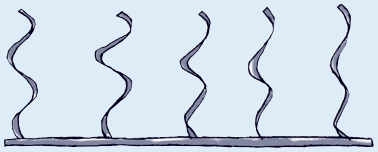
The revolutionary new tool underlying a snapshot of gene expression in a cell is the microarray, sometimes called the gene chip or the DNA chip. Microarrays consist of large numbers of molecules (often, but not always, DNA) distributed in rows in a very small space. The arrays are laid out by robots that can position gene fragments so precisely that more than 10,000 of them can fit on a piece of glass or plastic that is smaller than an ordinary microscope slide.

Pieces of DNA that have been tagged with fluorescent molecules are then placed on the chip, where they bind to their complementary DNA sequences among the fragments that are already on the chip. (A complementary sequence would have a T where the tagged DNA has an A, an A where it has a T, a G where it has a C, and a C where it has a G.) Next, a scanner measures the brightness of each fluorescent dot on the chip; fluorescence indicates that the gene is turned on. The pattern of gene activity is then analyzed by computer. The result is a freeze-frame moment in the life of a cell showing which genes are turned on and which are turned off.

With some life forms, scientists can make an array that includes DNA for all of its genes. These are organisms such as yeast, whose genomes have been fully sequenced—the precise order of nucleotides in all of their DNA is known. "We can ask what working blueprints, what RNA molecules, have been made from the entire population of genes. We can even count them. There's 10 from this gene, 1 from that gene, there's 200 from this other gene," Richard Young explains. "In doing so, we can create a description for what genes are on, what genes are off, and if a gene is on, how much working blueprint is it making? That's pretty remarkable."



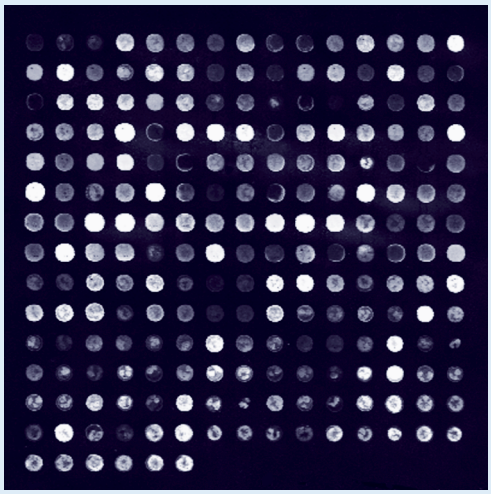
Microarray chip.



DNA fragments are attached to glass or plastic, then fluorescently tagged molecules are washed over the fragments.

Some molecules bind to their complementary sequence. These molecules can be identified because they glow under fluorescent light.

The resulting pattern indicates which genes are active.



DNA Array Facility, Fred Hutchinson Cancer Research Center

With microarrays, Young is amassing descriptions of the degree to which genes are on or off in particular cells under a variety of conditions. Young is also using the technique to discover what human genes do when their cells are infected by disease-causing organisms. But his biggest ambition is to use arrays to put together a map of the complete regulatory circuitry in the yeast *Saccharomyces cerevisiae* (**sack-are-oh-MY-sees sare-a-VEE-see-ay**), an organism that biomedical scientists often use for genetic studies. This is the same yeast that bakers use to make bread.

"We are taking advantage of what we learned about the transcription apparatus, the [RNA polymerase] holoenzyme, where we know many of the components. We want to expand that to understand the entire regulatory circuitry of a living cell," Young recounts. The plan is to move on from studying the behavior of individual genes to studying an entire genome at work. How are cells able to respond rapidly to different environments? How can they alter their gene expression programs to use resources more efficiently and out-reproduce their neighbors? "You can see that only if you examine the behavior of all genes simultaneously and under a variety of different environmental conditions."

Already, Young and his colleagues have discovered that changing the surroundings of a cell—moving it from a nutrient-poor to a nutrient-rich environment, for example—swiftly remodels the expression pattern of its genes. "A big piece, perhaps a third, of the entire genome can be turned on or turned off just because the cell was exposed to a new environment," he says.

The map Young envisions would describe everything from a change in the environment outside the cell to the regulatory pathway that brings news of the change to various proteins—and ultimately to the genes whose expression changes as a consequence. He hopes the regulatory map for yeast will generate insights into how genes behave in other organisms.

"The extent to which we can take this map we are developing with yeast and use it as a foundation for developing similar maps for humans is unclear at this point," Young acknowledges. He points out, however, that scientists have already established that about half of the yeast genome seems to be highly conserved—meaning that the same or very similar genes can be found in more complicated creatures, including people.

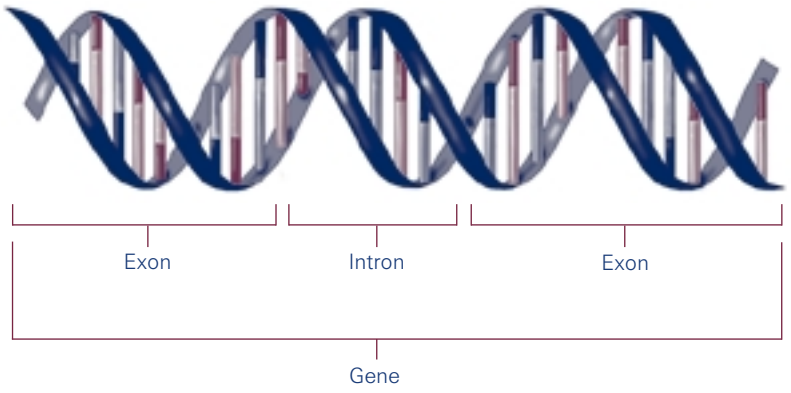
"Extra" DNA in Genes and RNA Splicing

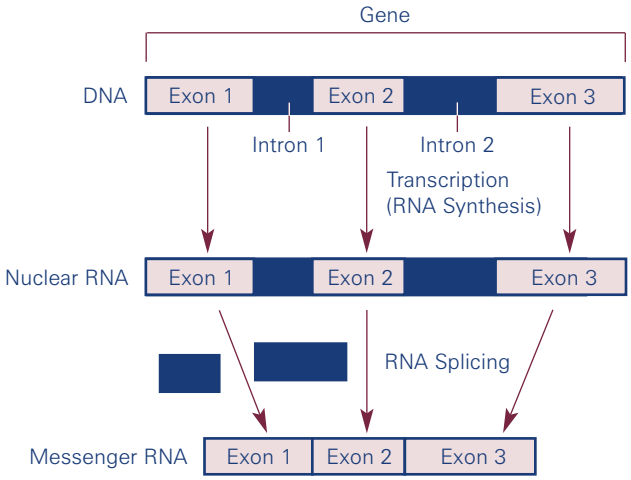
Here's an amazing fact: In cells with an organized nucleus (eukaryotes, which include "higher" organisms, meaning everything from yeast to humans), there is lots of noncoding DNA in the middle of genes. The coding sequences of individual genes—called "exons"—are split up by long stretches of the noncoding DNA. For this reason, scientists call this DNA "intervening sequences," or "introns" for short. The gene for the protein that is abnormal in boys with muscular dystrophy, for example, is divided by introns into 79 exons.

If a gene's RNA transcript is to make a protein that works properly, the intron RNAs must be removed from it first. Then the exon RNAs must be spliced together to make a complete coded message. "This seems like a crazy way to do business," molecular biologist Christine Guthrie points out. In her lab at the University of California, San Francisco, Guthrie and her colleagues have labored for two decades to figure out how this very odd process works and how it came to be.

Not only must intron RNAs be removed, they must be removed extremely accurately. An error in splicing even a single nucleotide in a gene's code will throw the whole sequence out of kilter. The result is usually an abnormal protein—or no protein at all. A form of the brain-destroying Alzheimer disease, for example, is due to this kind of splicing error.

So Guthrie and her colleagues want to discover the mechanism for removing intron RNA and find out how its accuracy is controlled. "A dream goal would be to try to figure out how to improve that accuracy, and thereby eventually have an impact on many different kinds of diseases," she says.





▲ Genes are often interrupted by stretches of DNA that do not contain instructions for making a protein. These stretches are called introns, and they must be removed before the RNA transcript of a gene is used to make a protein. The DNA segments that do contain protein-making instructions are known as exons.

Guthrie studies the splicing process in the same organism that Richard Young is using, yeast. Yeast is just a single cell, but its DNA has introns, although they are fewer and simpler in structure than human introns. In yeast, Guthrie can try to identify which genes are required for splicing by finding variants that mangle splicing.

The splicing machinery is a large structure called the spliceosome. It is made of RNA and proteins, and it has a complicated and changeable structure. For this reason, it is hard to isolate a complete, stable complex that contains all of the individual components of the spliceosome in order to study it further.

"Our current working idea is that the reason [splicing] is so complex and dynamic is that these stages in the assembly are opportunities to determine whether an intron [RNA] has been recognized correctly or not," Guthrie explains. She and her colleagues hypothesize that each step along the pathway presents an opportunity for proofreading, checking over and over again to make sure that the exon splicing has been done correctly.

To further complicate matters, splicing is not always straightforward. A great many genes can be spliced in more than one way. One exon RNA can be substituted for another, and sometimes an exon RNA can be omitted entirely.

Why does this matter? Because alternative splicing generates a different messenger RNA and therefore eventually a different protein. Sometimes these different proteins are made in the same cell, and sometimes they are made in different cells. Alternative splicing begins to explain how one gene can perform more than one job.

How Ribosomes Make Proteins

Harry Noller and his colleagues at the University of California, Santa Cruz have been asking one key question for years: How does the ribosome translate the genetic code into proteins?

Ribosomes are among the biggest and most intricate structures in the cell. The ribosomes of bacteria contain not only huge amounts of RNA, but also more than 50 different proteins. Human ribosomes are full of even larger amounts of RNA and between 70 and 80 different proteins. Protein synthesis is very fast and very accurate. Every second, ribosomes incorporate about 15 amino acids into the growing protein.

Noller and other researchers have found that the ribosome does several key jobs in translating the genetic code of messenger RNA into proteins. As the messenger RNA threads through the ribosome, the ribosome "reads" the sequence and helps recognize the correct transfer RNA to match the code. The ribosome also acts as an enzyme, linking amino acids into a growing protein chain.

For many years, researchers believed that these functions were carried out by proteins in the ribosome—even though, in 1972, Noller published evidence that the functions are actually performed by the ribosomal RNA. Noller's evidence was ignored because at that time it was thought that RNA could not act as an enzyme. Then, in the mid-1980s, Sidney Altman of Yale University in New Haven, Connecticut and Thomas Cech of the University of Colorado at Boulder each discovered that RNA can catalyze chemical reactions. For this discovery, Cech and Altman shared the Nobel Prize in 1989.

UUU phenylalanine	UCU serine	UAU tyrosine	UGU cysteine
UUC phenylalanine	UCC serine	UAC tyrosine	UGC cysteine
UUA leucine	UCA serine	UAA stop	UGA stop
UUG leucine	UCG serine	UAG stop	UGG tryptophan
CUU leucine	CCU proline	CAU histidine	CGU arginine
CUC leucine	CCC proline	CAC histidine	CGC arginine
CUA leucine	CCA proline	CAA glutamine	CGA arginine
CUG leucine	CCG proline	CAG glutamine	CGG arginine
AUU isoleucine	ACU threonine	AAU asparagine	AGU serine
AUC isoleucine	ACC threonine	AAC asparagine	AGC serine
AUA isoleucine	ACA threonine	AAA lysine	AGA arginine
AUG methionine (start)	ACG threonine	AAG lysine	AGG arginine
GUU valine	GCU alanine	GAU aspartic acid	GGU glycine
GUC valine	GCC alanine	GAC aspartic acid	GGC glycine
GUA valine	GCA alanine	GAA glutamic acid	GGA glycine
GUG valine	GCG alanine	GAG glutamic acid	GGG glycine

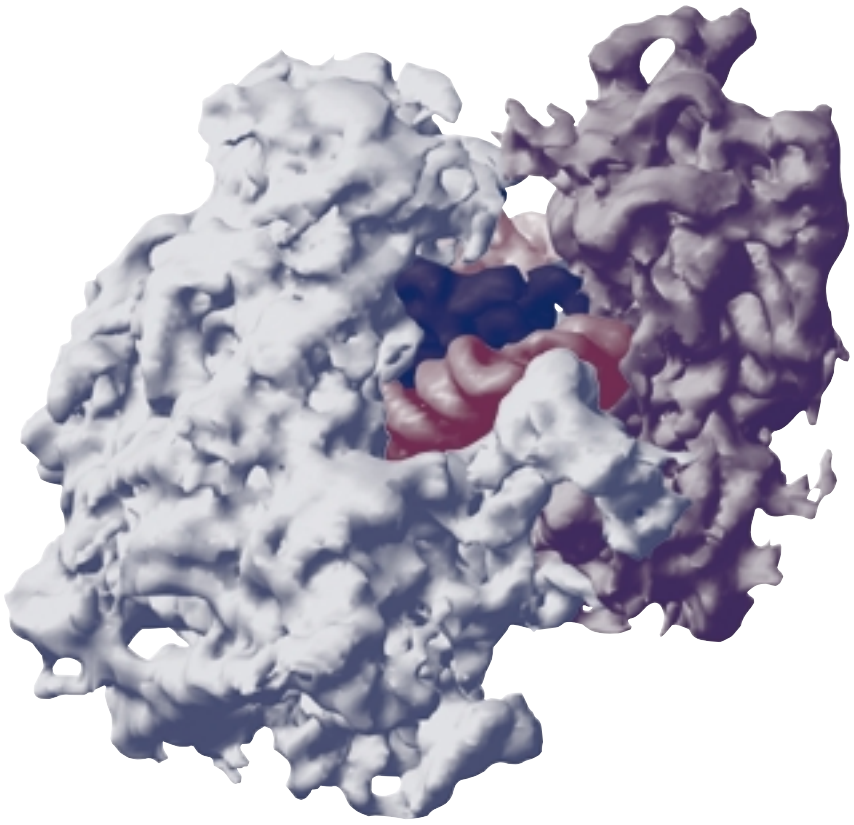
▲ The genetic code. Each triplet of nucleotides in RNA (a codon) codes for one amino acid in a protein, except for three—the "stops"—which signify the end of a protein chain. One amino acid, methionine, can also act as a signal to start protein production.

Fast-forward to 1999, when Noller and his colleagues made images of the actual structure of a bacterial ribosome, the result of decades of work. The images demonstrate how different parts of the ribosome interact with one another and how the ribosome interacts with molecules involved in protein synthesis. The functional centers of the ribosome are RNA, and the proteins are peripheral. "We can now say that the fundamental mechanism of translation is based on RNA," Noller declares.

Now Noller and his colleagues are at work figuring out the ribosome structure in more detail. They want to produce a model of each piece of every molecule in the ribosomal complex. They are also trying to determine the structure of the ribosome throughout protein synthesis.

Of course, it's interesting to learn how proteins are made and to marvel at what science has told us about how complicated—yet how extraordinarily accurate—it all is. It's astonishing to gaze at an image of how the moving parts of an unimaginably tiny structure work together to make the proteins that keep us—and every other living thing—alive and functioning.

But there are also very practical reasons for learning everything there is to know about the ribosome. Will we find new ways to cure infectious disease in the future? The ribosome may help us answer that question "yes."



▲ The structure of the ribosome, showing the large and small subunits with transfer RNAs nestled in the middle.

Ribosome structure courtesy of Jamie Cate, Marat Yusupov, Gulnara Yusupova, Thomas Earnest, and Harry Noller. Graphic courtesy of Albion Baucom, University of California, Santa Cruz.

Why? Because a great many of the antibiotics doctors use against infections target bacterial ribosomes, preventing these disease-causing organisms from making the proteins they need to survive. Erythromycin, neomycin, tetracycline, and hundreds of other antibiotics all work by attacking the ribosomes of bacteria.

A terrible problem facing modern medicine is that bacteria have learned how to outwit many antibiotics. One way they do it is by changing components of their ribosomes so that the ribosomes no longer interact with the antibiotic. They also employ enzymes to change the antibiotic so that it no longer binds to the ribosome. Some bacteria have developed more than a dozen ways of resisting antibiotics.

As a result, doctors are having more and more difficulty curing bacterial diseases. In fact, diseases that had been considered conquered 20 years ago, such as tuberculosis, are now coming back with a vengeance because of drug-resistant bacteria. Even organisms that pose few problems to healthy people can cause very serious diseases in a weakened hospital patient when antibiotics are no longer effective against the organisms. "Something as simple as a pimple, a little superficial infection, could potentially be lethal," Noller points out.

That means, he says, that scientists are going to have to find new antibiotics—or design them. "It is theoretically possible that we can determine the ribosome binding sites for known antibiotics, understand how the bacteria are developing resistance to these, and then design new antibiotics—or derivatives of the previous ones—that will outwit the bacterial defense mechanisms," he explains. "That is of course way in the future still, but it is now not a fantasy."

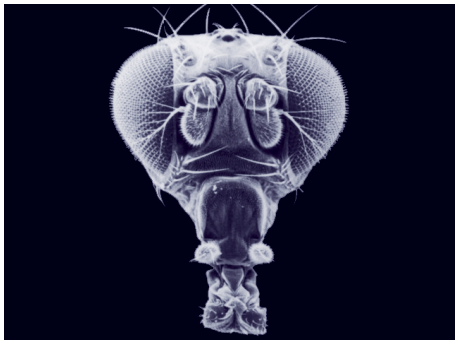
How Genes Control Development

One of the most important jobs genes do is to control how embryos develop. Scientists discovered a hugely important set of genes involved in development by studying strange malformations in fruit flies. The most famous such abnormality is the fruit fly with a leg growing out of its head instead of the usual antenna. "It's a perfectly normal leg. It's just in the wrong place," says Thomas C. Kaufman of Indiana University in Bloomington.

In this abnormality and many others, something goes wrong with the genetic program that directs embryonic cells down specific developmental pathways. In the antenna-into-leg example, it is as if the cells growing from the fly's head, which normally would become an antenna, mistakenly believe that they are in the fly's thorax, and therefore ought to grow into a leg. And so they do.

This discovery told scientists that genes can act as switches. These genes are master controllers that provide each body part with a kind of identification card. If a gene that normally instructs cells to become an antenna is disrupted, it can order the cells to become a leg instead.

Scientists determined that several different genes, each with a common sequence element, provide these anatomical instructions. Kaufman identified and described one of these genes, which became known as the *Antennapedia* (**an-TEN-ah-PEE-dee-yah**) gene. Antennapedia means "antenna feet." Flies with a mutation in the *Antennapedia* gene have a leg where an antenna should be.



FlyBase; R. Turner

▲ Normal fruit fly head.

FlyBase; R. Turner

▲ Fruit fly head showing the effects of the *Antennapedia* gene. This fly has legs where its antennae should be.

Kaufman then began analyzing the molecular structure of the *Antennapedia* gene. In the early 1980s, he and his colleagues made a discovery that has been fundamental to later studies, not just of development but also of evolution. (At about the same time, the discovery was made independently in Switzerland.) The researchers found a short sequence of DNA, now called the homeobox, that is present not only in *Antennapedia* but in the several genes adjacent to it, as well as in other genes with apparently different functions.

Geneticists get pretty excited when they find identical DNA sequences in the genes of different organisms. It usually means that this stretch of genetic material does something so important and useful that evolution uses the sequence over and over and permits very few changes in its structure. Researchers quickly discovered that the homeobox sequence element was not confined to the fruit fly. Nearly identical versions of the homeobox turned up in almost every living thing they examined, no matter how distantly related—first in a frog, then in worms, beetles, chickens, mice, and even yeast and plants. And, of course, the homeobox is found in people.

Hundreds of homeobox-containing genes have been identified, and many of them have turned out to be involved in early development. For example, abnormalities in the cluster of genes that lead to a fruit fly with a leg where its antenna should be can lead, in people, to extra fingers or toes. Homeobox genes demonstrate that people and flies are relatives. Distant relatives, of course. But both people and flies are designed and constructed by similar genes to fit neatly into the characteristic body plan of each organism.

Scientists believe the first homeobox gene, which arose very early in the history of life on Earth, worked in simple ways. But now, some 500 million-plus years later, homeobox genes have become remarkably versatile. They adapt easily to many ways of managing the fates of cells and the body patterns of extremely different kinds of creatures.

The Tools of Genetics: Recombinant DNA

Early in the 1970s, scientists demonstrated that they could transfer genetic material—and genetic traits—from one organism to another. These experiments changed everything. This simple, mind-boggling fact—that genes from one creature can be inserted into another, make themselves at home, and go to work as usual—shook the life sciences to their core. The discovery underlies most of the extraordinary accomplishments of the past three decades of genetics research.

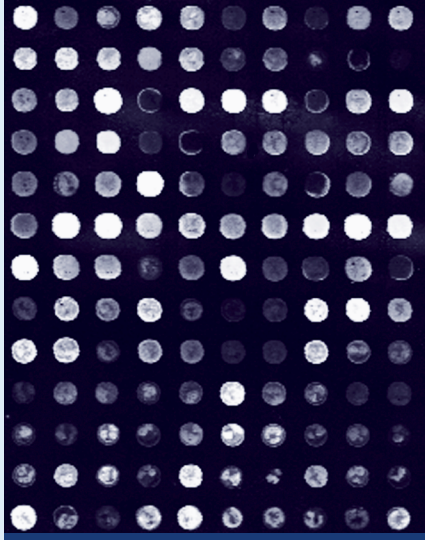
In addition to providing startling evidence of the similarities between life forms, the experiment also showed a way to make many copies of—to "clone"—any gene. Making a lot of copies of a gene is necessary in order to have enough to examine and identify it. In fact, the term gene cloning has come to mean not just gene copying, but gene discovery—the identification of a gene that does a specific job. For example, scientists recently cloned the gene that makes Mendel's peas smooth or wrinkled.

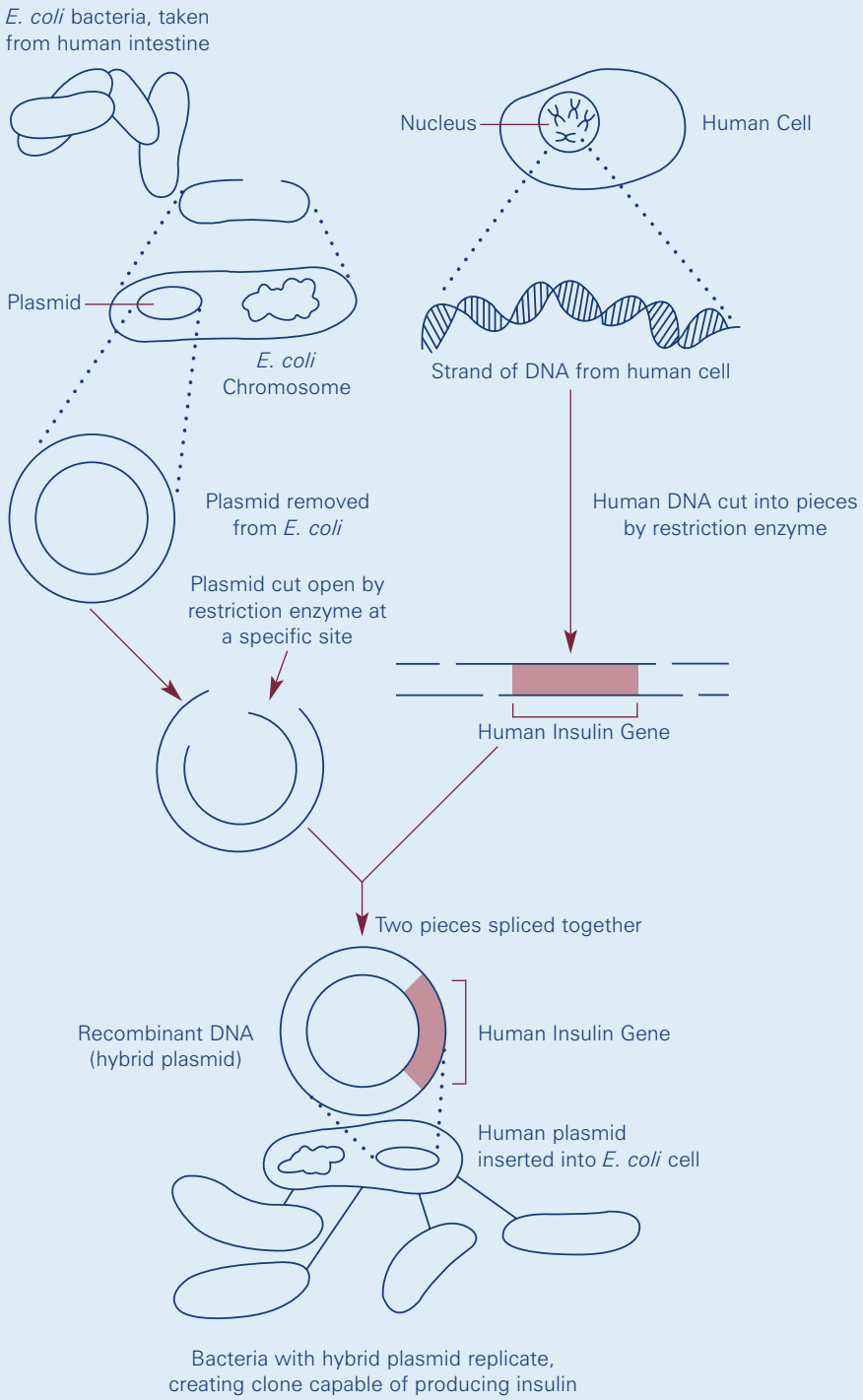
How is this gene transfer made? Here's one method. Suppose a scientist wants to make lots of human insulin. The first step is to transfer the human gene for insulin into the bacterium *Escherichia coli* **(ess-shuh-RICK-ee-uh KOH-lie)**. *E. coli* is an organism often used in genetics research; some forms are normal inhabitants of the human digestive tract.

Then, the scientist would cut the insulin gene out of a piece of human DNA using a special enzyme called a restriction endonuclease. There are scores of these enzymes. Each one cuts DNA at a different sequence, so it is possible to be very precise about DNA cutting by selecting the restriction endonuclease that cuts at the desired sequence.

Next, the scientist would splice (paste) the human gene into a special kind of bacterial DNA called a plasmid. The splicing is done with another enzyme, called DNA ligase. The result: recombinant DNA, a sort of cut-and-pasted circle of human and bacterial DNA.

Finally, the scientist would transfer the recombinant DNA into *E. coli*. *E. coli* will then obligingly divide and go on dividing. In a very short time, there will be millions of *E. coli*. Each one will be carrying a working gene that is fully capable of producing human insulin.





▲ Recombinant DNA. To splice a human gene (in this case, the one for insulin) into a plasmid, scientists take the plasmid out of an *E. coli* bacterium, break the plasmid open at a specific site by means of a restriction enzyme, and splice in insulin-making human DNA. The resulting hybrid plasmid can be inserted into another *E. coli* bacterium, where it multiplies along with the bacterium, thus producing large quantities of insulin.

Got It?

What is a gene?

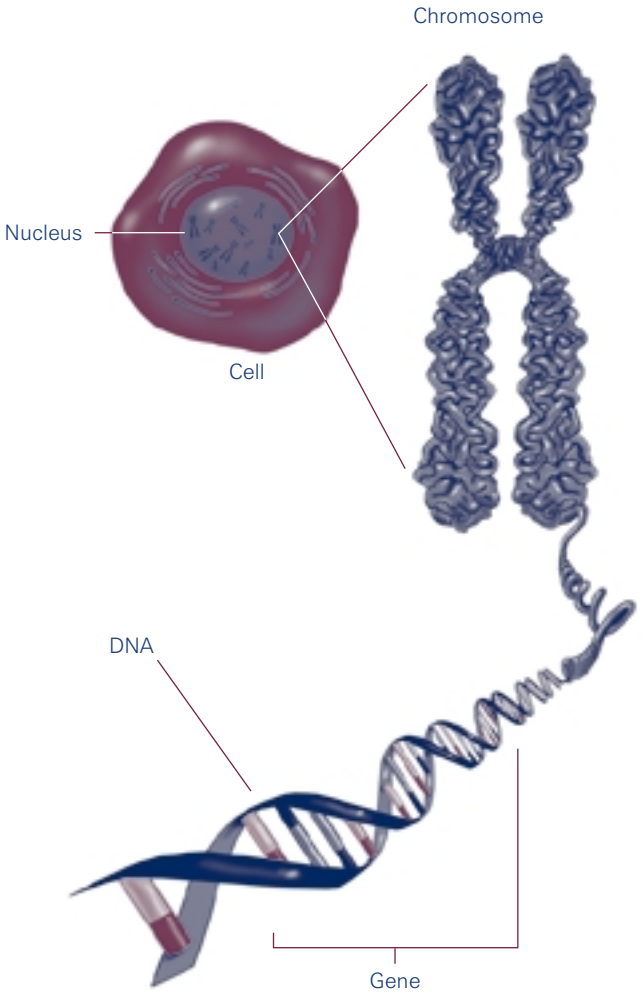
What are mutations? Are they good or bad, or both?

Why is intron RNA spliced out of messenger RNA?

If every cell of an organism contains the same set of genes, why are some of the cells so different from others?

Strange But True: Exceptions to Mendel's Rules

Mendel's observations about how inheritance works in pea plants are the foundation on which 20th century genetics was built. In the first third of the 20th century, scientists discovered an exception to Mendelian genetics involving how genes on the human sex chromosomes, X and Y, are inherited. (Remember that chromosomes are structures in the nucleus that contain an organism's genes. In humans, the X and Y chromosomes are involved in sex determination: Normal human females have two X chromosomes in each cell, while normal human males have one X and one Y.)



▲ Chromosomes are found in the cell nucleus and contain an organism's genes.

Children inherit one copy of most genes from their mothers and another copy from their fathers. But genes on the Y chromosome are different: They are passed directly from father to son. Mothers are not involved at all, since women do not have Y chromosomes. Genes on the X chromosome are also different. Boys inherit only one copy of each "X-linked" gene, and it comes from their mothers.

It turns out that there are lots of genes on the human X chromosome, including genes that cause the most common types of color blindness and muscular dystrophy. Boys are therefore much more likely to inherit these disorders than girls are, because boys do not have a second X chromosome with a gene that could compensate for one that is not working properly on the other X chromosome. For this reason, genetic counselors working with couples tend to be concerned if people in the woman's family, but not the man's family, have X-linked diseases like muscular dystrophy.

In the last few decades, scientists have uncovered more startling exceptions and complications to Mendelian genetics. These discoveries have astounded scientists and left them shaking their heads at how explorations in genetics are becoming ever more intricate—and ever more fascinating.

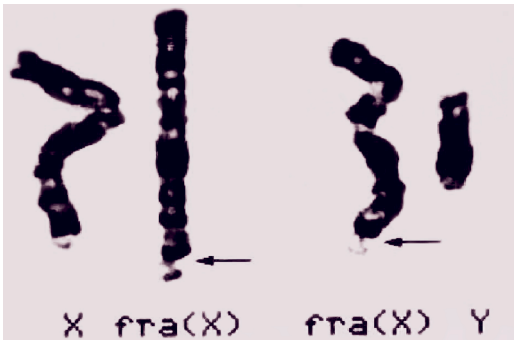
The Genetics of Anticipation

One well-studied example is fragile X syndrome, which causes mental retardation. (The name comes from an unusual narrow place on the X chromosome that can be seen in a microscope; it is called a fragile site.)

Fragile X has several unusual features. One of the oddest is that the risk of a child being affected depends on more than whether a parent has passed along a fragile X chromosome. The risk actually increases as the chromosome is passed down through the generations. A male with a fragile X chromosome is not always retarded, but the grandsons of such a man run a 40 percent risk of retardation, and the risk for his great-grandsons is 50 percent.

Scientists identified the gene that causes fragile X syndrome in 1991 and named it FMR-1. The molecular defect that causes the syndrome is not a conventional mutation, in which nucleotides are switched around or dropped. Instead, it is a kind of stutter in the DNA, a string of repeats of a particular sequence composed of just three nucleotides, CGG. Some people have only one such "triplet repeat," a sequence that reads CGGCGG. Others have more than a thousand.

When scientists studied the FMR-1 triplet repeats, they found a new kind of disease-causing mutation. People who did not have the fragile X chromosome had from 6 to 52 repeats, with an average of about 29. People who had the fragile X chromosome but were not mentally retarded had from 50 repeats to more than 200 repeats. Those who had the fragile X chromosome and were mentally retarded often had 1,000 repeats, or even more. In addition, researchers found that chromosomes carrying more than 52 CGG repeats were so unstable that the number of repeats could increase when the chromosomes were passed down from a parent to a child. One mother with 66 repeats, for example, had a first child with 80 repeats, a second child with 73, and a third with 110. The higher the parent's repeat number, the more likely his or her children are to possess more than 230 CGG repeats. People with fewer than 230 repeats generally are not retarded, while people with more repeats usually are.



National Fragile X Foundation

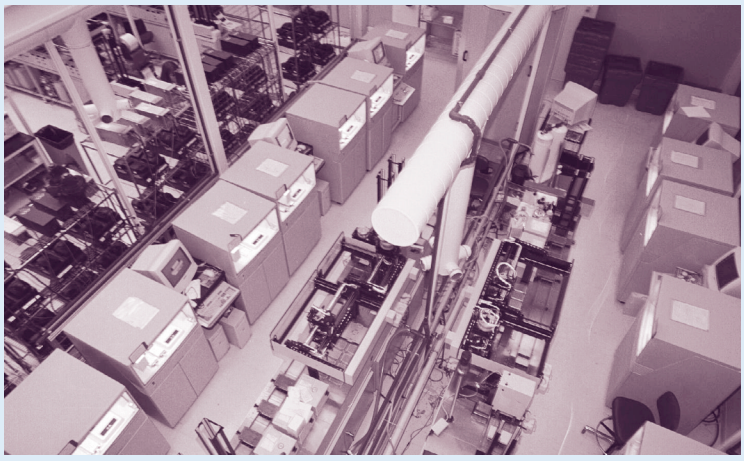
▲ The sex chromosomes of a female and a male with Fragile X syndrome. The two X chromosomes of the female are on the left and the X and Y chromosomes of the male are on the right. The arrows point to the characteristic "fragile" site, which looks as if it is ready to break.

Amazed, scientists went looking for other examples of diseases associated with triplet repeat expansions. Triplet repeat expansions turned out to be the explanation for "anticipation," a puzzling phenomenon first described in the neuromuscular disease myotonic dystrophy: Symptoms of the disease showed up earlier and were more severe in each generation. Some other disorders that have been traced to triplet repeat expansions display anticipation too, such as Huntington disease. The list of triplet repeat diseases keeps on growing. So far it numbers eight, and all of the disorders affect the nervous system.

The Tools of Genetics: Mapping and Sequencing the Human Genome

In the 1980s, geneticists realized that they had the tools—and the need—to learn the complete layout of the human genome. They wanted to know not only where every gene was situated and what its nucleotide sequence was, but also the complete sequence of the entire genome's 3 billion nucleotides.

With that information in hand, scientists reasoned, it would eventually be possible to learn exactly what job each gene performs and exactly how genes contribute to human disease. But learning a lot about how human genes worked would be impossible without first knowing what and where the genes were. Finding out those things would be a foundation for building a real understanding of the human body.



Bethany Versoy

▲ Sequencing center at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts.

Since about 1990, thousands of scientists in labs all over the world have been involved in systematic efforts to decipher human DNA. Many of these scientists are part of the federally sponsored Human Genome Project, while other genome scientists are working at private companies. The scientists are developing maps of human genes showing just where each one is—which chromosome it's on and precisely where it is on that chromosome. They have developed technologies for finding genes; technologies for the fast, automated determination of DNA sequences (a process known as sequencing); and technologies for storing and analyzing the increasing flood of data streaming in from labs everywhere. Researchers are also studying how genes differ slightly between people. The social and ethical issues arising from the increasing use of genetic information in medicine are being explored, too. These issues include the privacy and fair use of genetic information, as well as the impact of genetic testing on individuals, families, and society.

Scientists completed the first draft of the human genome sequence in 2000. Complementing this effort are genome investigations for many other organisms. These nonhuman maps and sequences help scientists figure out what various genes are doing in the organisms and help them identify similar genes in humans. Some of these projects have already been completed, including mapping and sequencing the genomes of four organisms commonly used in genetics research: the roundworm *Caenorhabditis elegans* (**SEE-no-rabb-DYE-tis EL-eh-ganz**), the fruit fly *Drosophila melanogaster* (**dro-SOFF-ill-ah mah-LAN-oh-gas-ter**), the yeast *S. cerevisiae*, and the plant *Arabidopsis thaliana* (**a-RAB-ih-DOP-sis THA-lee-AH-nah**).

The Battle of the Sexes

Another exception to Mendel's picture of inheritance is a startling phenomenon called imprinting. With most genes, both the mother's and father's copies work exactly the same way in their offspring. For some mammalian genes, however, only the mother's or the father's copy is expressed. During the process that generates eggs and sperm, imprinted genes are marked somehow. This marking allows the resulting embryo to distinguish whether a gene copy came from Mom or Dad, and to shut one of the copies down.

One example is insulin-like growth factor 2 (Igf2), a gene critical for the growth of the mammalian fetus. Only the father's copy works. "Although you inherit a perfectly good copy from your mother, that copy is silent for your entire life," notes Shirley Tilghman, a molecular biologist at Princeton University in New Jersey. This selective silencing of Igf2 and many other genes has proved true in all mammals examined so far, but not in birds. This suggests that imprinting appeared sometime between 300 million and 150 million years ago, when mammals and birds became separate branches on the evolutionary tree.

For the past few years, Tilghman and her colleagues have been asking: When did imprinting evolve? Why? How does it work? "From a genetic perspective, it's a really silly thing for an organism to do. Why would you inactivate a perfectly good copy of a gene?" Tilghman asks. For one thing, imprinting puts an organism at risk because there's no backup copy, as there is with most other genes.

Imprinting also seems to violate the idea that a trait evolves because an organism works better if it possesses the trait. Tilghman and many other scientists have come to believe that imprinting evolved not because it was useful to a particular organism, but because where their offspring are concerned, mothers and fathers are at war.

Why war? Because mothers and fathers have competing interests.

It is in a father's interest for his embryos to get bigger faster, because that will improve his offspring's chances of survival after birth. The better a creature's chance of surviving infancy, the better its chance of becoming an adult, mating, and passing its genes on to the next generation.

Mothers have a different agenda. Of course they want strong babies, but a female is likely to be pregnant several times. She needs to divide her resources equally among a number of embryos in different pregnancies. It would therefore be to her advantage to control the growth of any particular embryo.

Researchers have discovered dozens of imprinted genes in mammals since the first came to light in 1991. Sure enough, imprinting controls some of the most important genes that determine embryonic and fetal growth and allocation of maternal resources. Mutations in these genes cause serious growth disorders in mice and humans.

Shirley Tilghman

▲ This family portrait illustrates the impact of imprinted genes on the fetal growth of mice. The smallest mouse (on the left) has a mutation in the paternally expressed insulin-like growth factor 2 gene. The largest two mice (on the right) have a mutation in a maternally expressed gene called H19. The mice in the middle are normal-sized and have mutations in both genes, which cancel each other out.

In addition, scientists have found that imprinted genes are involved in cancer. The stretch of DNA Tilghman studies codes for six or seven imprinted genes, and two of them seem to be involved in tumor formation. Insulin-like growth factor 2 has been implicated in several cancers, including liver cancer and kidney cancer. Another imprinted gene appears to be involved in a disease called Beckwith-Wiedemann syndrome, which is associated with a high incidence of childhood tumors. Imprinted genes could encourage the growth of cancers in much the same way that they encourage the growth of fetuses, Tilghman says.

The Other Human Genome

About a billion and a half years ago, bacteria figured out how to use oxygen to produce the energy they needed for life. Around the same time, a brand-new type of life form arose. It was just a single cell, but it carried its genetic material around in a kind of membrane-enclosed bag we now call a nucleus. This primordial eukaryote gulped down some of the oxygen-using bacteria and found itself with plenty of energy.

It was the beginning of a beautiful relationship. Today, nearly all plant and animal cells contain offspring of those symbiotic energy producers. They are called mitochondria. Mitochondria are the cell's power plants, supplying the energy to carry out all of the cell's jobs.

Mendel knew nothing of mitochondria because they were discovered late in the 19th century. Scientists puzzled out their energy-producing talents in the first third of the 20th century. But it was the 1960s before researchers discovered that mitochondria contain their very own genomes. This is not too surprising when you remember that mitochondria are descended from free-living organisms.

The mitochondria of some organisms contain a lot of DNA, and their genes turn out most of the proteins the mitochondria need. Human mitochondrial DNA (mtDNA) is not very abundant, accounting for less than 1 percent of the total DNA in a human cell. The DNA contains only about

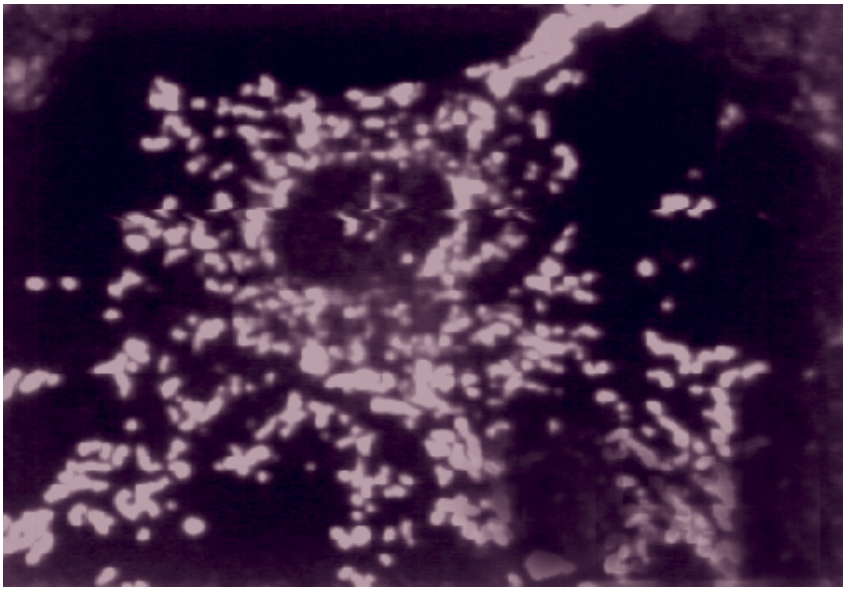
three dozen genes. That's enough to make a few of the proteins that the mitochondrion needs, as well as its own ribosomal RNAs. The rest of the human mitochondrion's genetic machinery has been turned over to the nucleus—including the machinery that controls the transcription and translation of mtDNA. So the energy-producing capabilities of human mitochondria depend on the interaction of hundreds of genes in both the nucleus and the mitochondria.

Douglas Wallace of Emory University in Atlanta was beginning his scientific career shortly after mtDNA was discovered. He reasoned that any structure that provided 90 percent of a cell's energy must be important, and that any structure that contained DNA could have mutations, which meant disease. "From the very beginning, my goal was to try and find traits that ultimately might have disease implications," he says.

First, he and his colleagues showed that mtDNA could encode proteins, so it must contain genes. He also uncovered the most startling single fact about human mtDNA: In both sexes, it is inherited only from mothers. Both egg and sperm contain mitochondria, of course, because both need them for energy. But after fertilization, sperm mtDNA disappears. So forget Mendel; we get all our mtDNA from our mothers, and our mitochondrial defects, too. Men with mitochondrial diseases do not transmit them to their children.

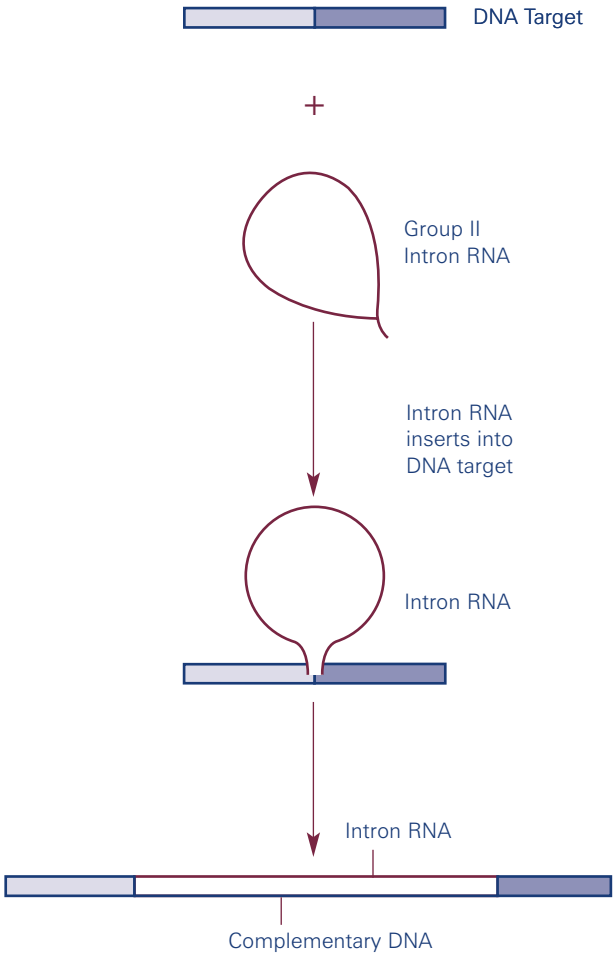
As Wallace foresaw, mitochondrial defects are anything but trivial. They lead to a variety of serious, degenerative diseases. Wallace and his colleagues discovered the first mutation in mtDNA that leads to a disease: Leber optic atrophy, which causes sudden blindness. They also identified a group of diseases in which nerves and muscles degenerate and muscles accumulate large numbers of abnormal mitochondria.

Now that mitochondrial disease is an accepted notion, Wallace and his colleagues are working to develop therapies. One approach might involve transferring "good" mitochondria into cells that have "bad" mitochondria.



Alison Davis

▲ The mitochondria in this cell are lit up with a fluorescent dye.



▲ Group II intron splicing into DNA.

Adapted from an illustration provided by
Alan Lambowitz and Huatao Guo

Jumping Genes

Another oddity has turned up as scientists discovered the intricacies of genetics: Genes can jump around in the genome. The amazing fact that genetic material is not always stationary was discovered in the 1940s by the plant geneticist Barbara McClintock, who was studying corn at Cold Spring Harbor Laboratory on Long Island, New York. Her discovery was so amazing, in fact, that other scientists thought it couldn't possibly be true, so her reports were largely ignored. Eventually, however, the existence of "jumping genes," also known as "transposons," was confirmed by others. In 1983, McClintock was awarded the Nobel Prize for discovering transposons.

Transposons are now often called mobile genetic elements in order to take account of another amazing fact: Introns can jump, too. Alan Lambowitz has been studying these mobile introns for some years. In 1995 he and his colleagues at Ohio State University in Columbus discovered that some

introns, known as group II introns, not only move around and insert themselves into genes, they do it by recognizing certain DNA sequences and slipping into genes only at those points.

After analyzing the way introns recognize their particular insertion sites, Lambowitz and his colleagues did a remarkable thing. By modifying the intron, they were able to coax it to insert into desired target sites on DNA. It suddenly seemed possible that researchers could control the placement of genetic material within a genome precisely.

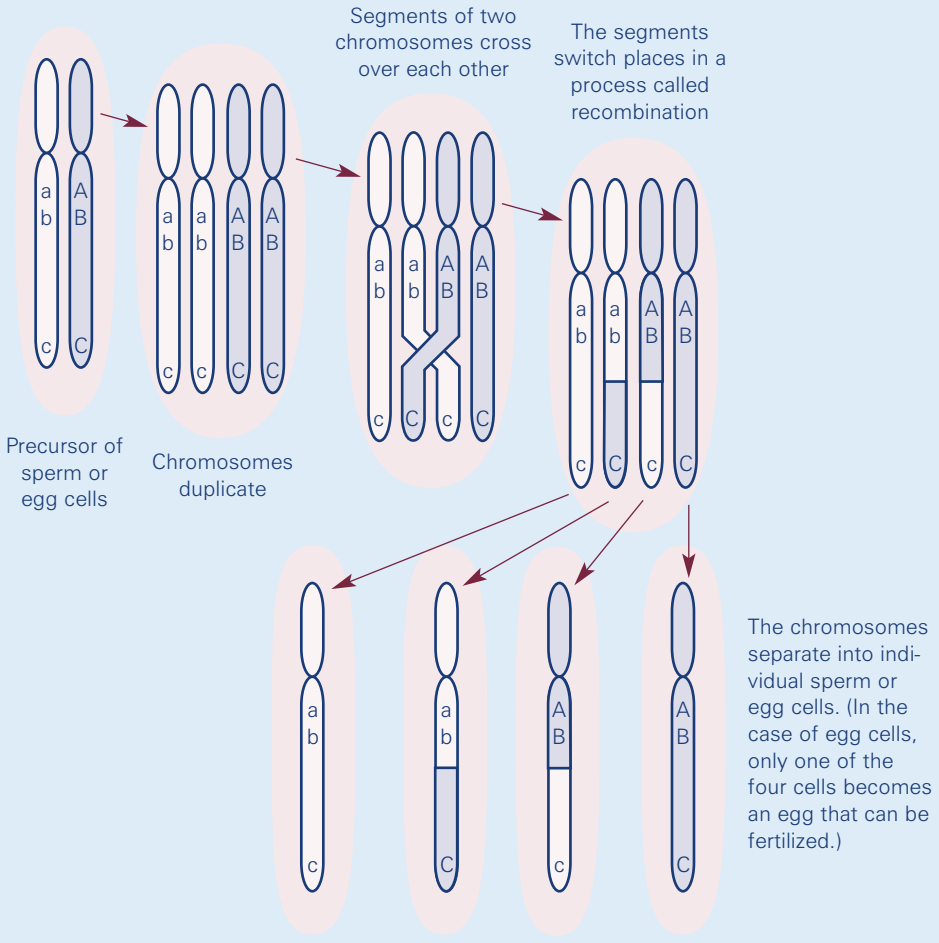
The potential therapeutic applications of being able to hook any gene to an intron and then point the two of them at any spot in the genome are enormous. First, there is the hope of using the technique in gene therapy. Gene therapy is an effort to cure disease by changing a patient's genes. There are, of course, enormous technical hurdles in attempting to insert DNA into the cells of a living human being. One of them has been that, even when desirable DNA has been successfully transported into the cell, it tends to insert randomly into the genome. If it does not insert in the proper place, a gene may not work correctly, or it may not work at all. Gene therapy researchers think they might have better luck if they could control the insertion points. The studies by Lambowitz and his colleagues suggest that pinpoint control over gene insertion might someday be possible.

The work also promises to be useful for creating mice and other genetically engineered organisms—such as fruit flies, worms, and plants—that can serve as disease models and help scientists figure out causes and cures. Lambowitz, who has since moved to the University of Texas at Austin, is hoping that the method can be used to destroy viruses, as well. These include AIDS, herpes, hepatitis B, and human papilloma, which plays a role in cervical cancer. He is also trying to develop introns targeted to cancer-causing genes, which could disrupt and inactivate them.

The Tools of Genetics: Designer Mice

Mice with genes from other organisms are an important tool for today's genetics research. Making these so-called "transgenic" mice involves a technique called gene targeting. The method uses homologous recombination, the normal process of DNA shuffling that occurs during the cell division that makes egg and sperm cells, which is called meiosis. Recombination creates new DNA mixes in each egg and sperm—which is why, unless you have an identical twin, you are genetically unique.

During homologous recombination, strands of DNA containing identical (homologous) nucleotide sequences line up side by side and exchange bits of genetic material. In experiments where he was injecting DNA from another organism into mouse cells, Mario Capecchi of the University of Utah in Salt Lake City discovered three surprising facts: The DNA found its way into chromosomes, more than one DNA molecule could be inserted at the same site, and all of the DNA was oriented properly.



▲ Homologous recombination during meiosis.

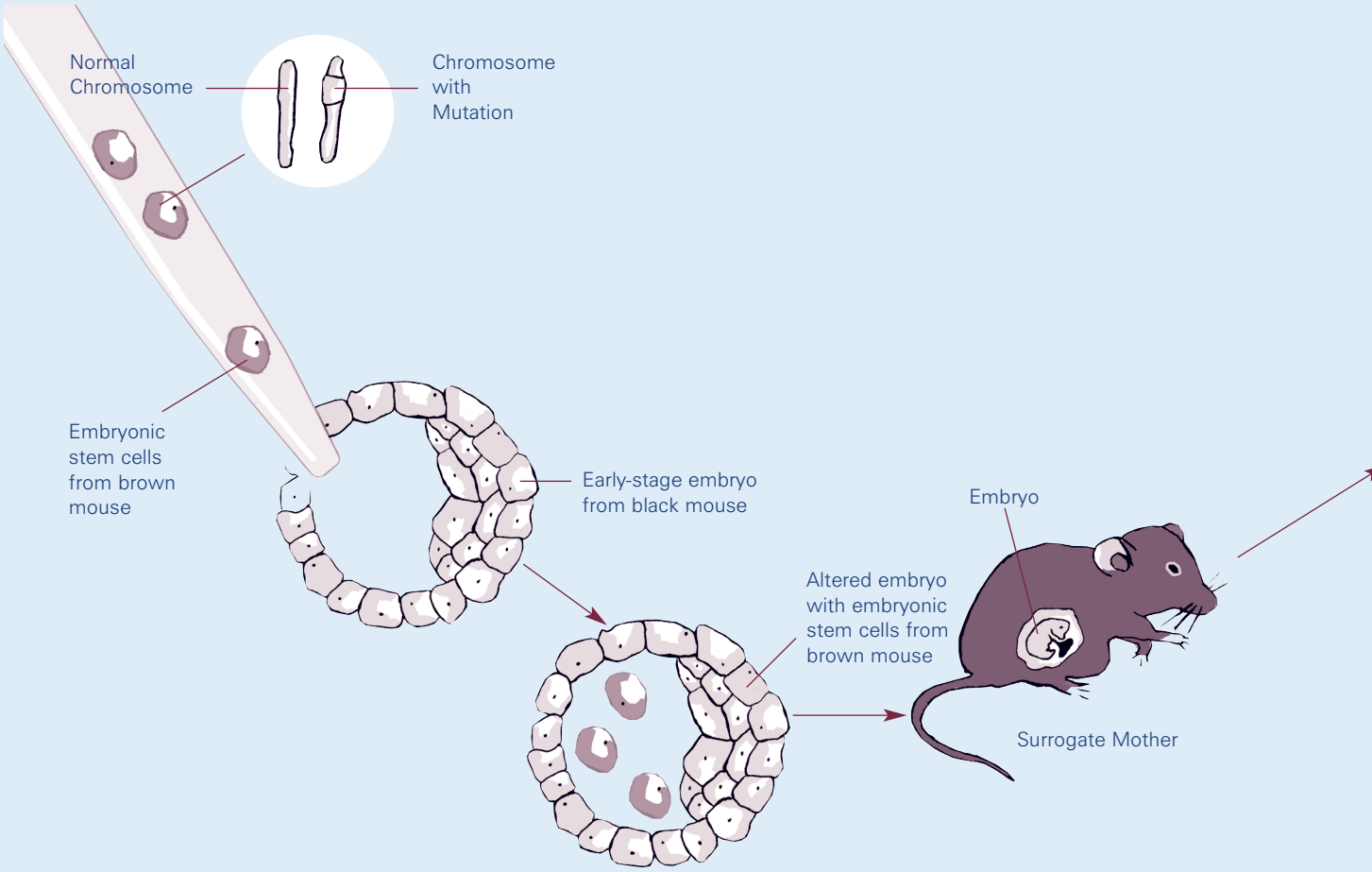
This indicated to Capecchi and his colleagues that the mechanism behind this behavior was homologous recombination. They were using body cells, not cells that were on their way to becoming eggs or sperm, and at that time, homologous recombination was believed to occur only in future egg or sperm cells.

Capecchi recalls that he realized right away that scientists might be able to manipulate this process to insert the DNA of their choice into the mouse genome. Ways of making this transfer have since been devised by Capecchi and his colleagues, and refined by a number of other researchers.

Capecchi's genetic engineering is done in mouse embryonic stem cells. An embryonic stem cell is at the earliest stage of development and has not yet begun to specialize—so much so that it is still capable of growing into every cell type. Most "foreign" DNA transferred into stem cells inserts into chromosomes at random. But very occasionally, the foreign gene links up with its corresponding mouse gene and makes itself at home there. The researchers have invented ways to separate the few cells in which the gene is in the right place from the thousands in which it isn't. Those few become "starter cells" that are grown into brand-new transgenic mice—mice containing a gene from another organism.

This technique can generate "knockout" lab mice, which are enormously valuable for disease research. To make a knockout mouse, scientists transfer a defective version of a gene they want to study into stem cells. The defective gene "knocks out" the normal gene, and scientists can examine the effects of the disabled gene on the resulting young mouse. Using gene targeting, researchers can transfer human disease genes into embryonic stem cells to make mouse models of many human ailments. They not only can learn about a disease in a mammal that is genetically very similar to people, they also can develop possible treatments and test them with no risk to human patients.

Capecchi says there are two main reasons for making model animals. One is the direct effects on treating human disease. He has, for example, made mouse models of one of the most common human genetic diseases, cystic fibrosis. Cystic fibrosis is caused by an inherited mutation in a particular gene, and about 75 percent of cystic fibrosis cases are due to a specific mutation in that gene. The other 25 percent are due to a huge assortment of different mutations—more than 100 at last count. The fact that so many mutations cause the disease explains why some cystic fibrosis patients do much better than others and why

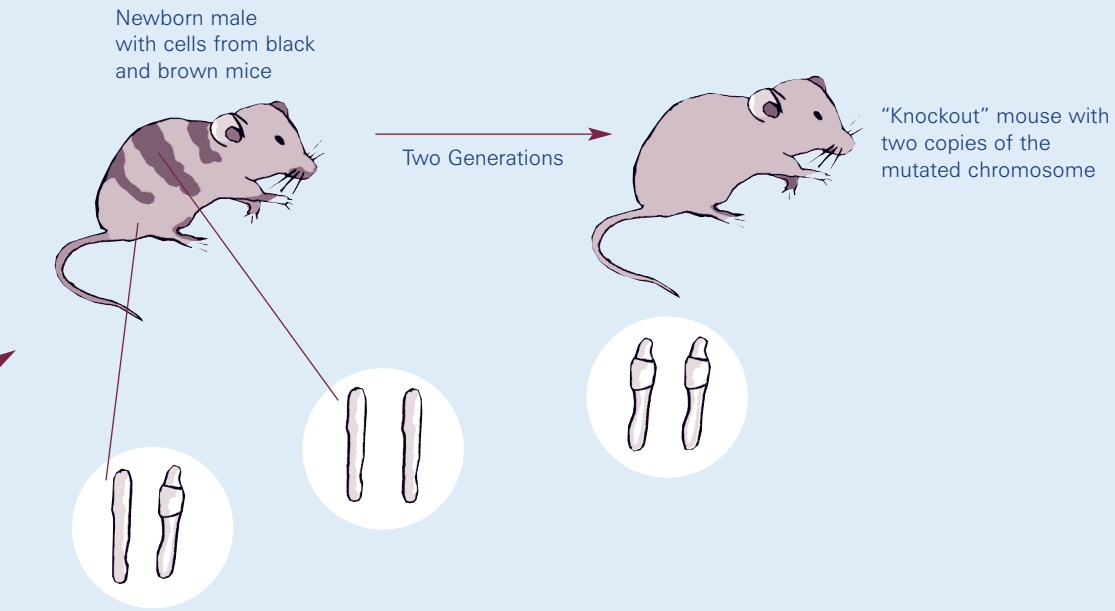


▲ How "knockout mice" are made.

Adapted from an illustration by Jared Schneidman Design

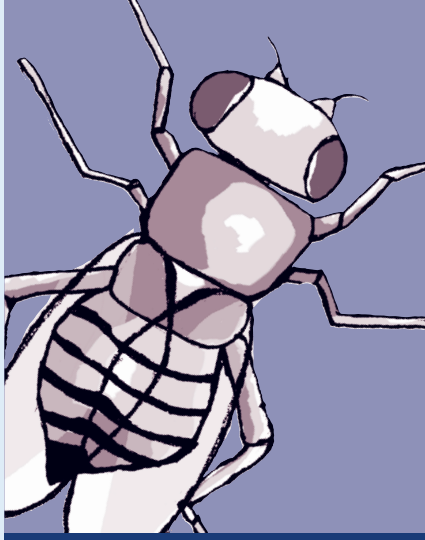
certain symptoms are so much worse in some patients than in others: The various mutations have different effects. But this diversity makes cystic fibrosis enormously hard to study and treat.

Mouse models permit study of all of these mutations. This helps researchers figure out whether a particular mutation causes, for example, more serious problems in the lung or the pancreas. "So by creating a series of very specific mutations



in the mouse, we can study each of the [mutations that cause cystic fibrosis] separately, or combine them in different ways and see whether we can [duplicate] what we see in human patients," Capecchi explains.

But Capecchi expects that we'll ultimately derive the most benefit from mouse models indirectly and over the long term. "If we understand mammalian biology in much greater detail than we do today, we will actually understand medicine much better. Right now, what we do is make a series of drugs and try them all out. It's trial and error. Often, you have no real idea about what the drug is doing," he points out. "In the long run, the more we understand the real biology of the symptoms, the better medicine is going to be. That's where the real contribution is. But that's much longer range."



Got It?

Why are X-linked conditions much more common in boys than in girls?

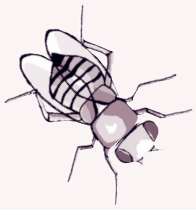
What are triplet repeats? What is their significance for human health?

Why do mitochondria have their own DNA?

What are mobile genetic elements, what do they do, and how are they important?

Why do scientists use knockout mice in genetics research?

Living Laboratories



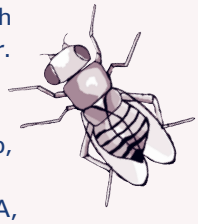
Fruit flies? Tiny worms? Yeast? Mice? What's going on here? Why do life scientists do research on these creatures?

These organisms, and many others, serve as models—living laboratories where researchers can make discoveries and test ideas. Model organisms—living things as different as bread mold and zebrafish—permit scientists to investigate questions they would not be able to study in any other way, in living systems that are, relatively speaking, simple, inexpensive, and easy to work with.

Model organisms are indispensable to science because living creatures that on the surface seem very different from each other—a mouse and a fruit fly, for example—actually resemble each other in body chemistry. Even organisms that seem nothing at all like people—ordinary bread yeast, for example—can give scientists clues to the workings of the human body. How? Because all living things consume food and turn it into similar chemicals that enable them to survive and reproduce. Their biochemistry is similar because their genes are similar. This means that a process discovered in a tiny, transparent worm can also be found—and studied, and clarified—in fruit flies and people, too.

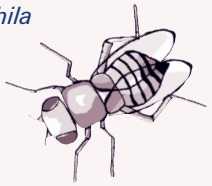
Each organism has characteristics that suit it to a particular sort of research. Scientists have poked into many corners of the animal and plant kingdoms in search of the right organisms to help them answer specific research questions. Not all model organisms are easy to raise and handle and inexpensive to feed and house, but many of them are. Cost and convenience are usually an important part of the decision.

Drosophila melanogaster: The Fruit Fly



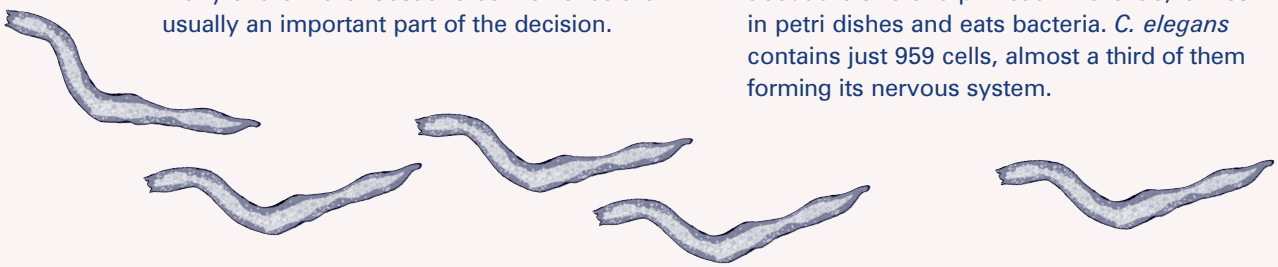
Take fruit flies, for example. The most commonly used species in research is named *Drosophila melanogaster*. A geneticist's fruit fly is pretty much the same as the ones that flit around the fruit bowl. In the lab, flies are exposed to chemicals or radiation, which damage their DNA, and are then permitted to mate. Scientists search among the offspring for flies with abnormalities. Abnormal flies are mated to produce more offspring with the abnormality, then studied to find the mutant gene that is causing it.

Fruit flies have been a favorite experimental organism among geneticists since early in the 20th century. Hundreds of them can live in a pint-sized milk bottle or even a vial, and they reproduce so quickly and so often that keeping track of a particular gene as it passes through several *Drosophila* generations requires only a tiny part of a human lifespan. What's more, after almost a century of investigation so much is known about fruit fly genetics—including the complete sequence of the *Drosophila* genome—that researchers can easily build on earlier studies.



Caenorhabditis elegans: The Roundworm

Caenorhabditis elegans—*C. elegans* for short—is a lot smaller than its name. This harmless roundworm, which lives in soil, is about the size of a pinhead. In the lab, it lives in petri dishes and eats bacteria. *C. elegans* contains just 959 cells, almost a third of them forming its nervous system.



The worm is particularly prized by biologists because it is transparent, so what goes on in its tiny body is in plain view in a microscope. "It is like looking at one of those watches where you can see the gears work. You can see right into its body. You can watch the food enter the digestive system," says Cynthia Kenyon of the University of California, San Francisco. "When we study cell migration, we can just look at cells and watch them move from one region to another."

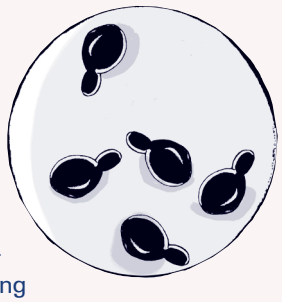
Scientists recently sequenced all of the genes in *C. elegans*. For such a small, simple animal, the worm turned out to possess a lot of genes—more than 19,000.

Deciphering the complete gene sequence for *C. elegans* was a huge milestone for biology. For one thing, it was the first animal genome to be sequenced completely. But even more important, a vast number of the genes in *C. elegans* turn out to be very similar to genes in other organisms. This includes genes of our own species, *Homo sapiens*, which is why a tiny worm can be a great model organism for scientists who want to find out more about how our bodies work and how we develop disease.

Saccharomyces cerevisiae: Yeast

There are hundreds of different kinds of yeast, but *Saccharomyces cerevisiae*, the one scientists use most often, is a staple of human life outside the lab, too. It is the yeast bakers use for bread and brewers use to make beer. Another yeast often used in research is *Schizosaccharomyces pombe* **(SKIZ-o-sack-are-o-MY-sees POM-bay)**. The two types of yeast may look alike to you, but scientists say they are only distantly related. Because it is not as common a model organism as bread yeast, scientists know much less about *S. pombe*.

Yeast is actually a fungus. It is not a mammal, of course, but it is still a eukaryote—a "higher" organism with an organized nucleus. It also grows fast, it's cheap to feed, it's safe to handle, and its genes are easy to work with and change for study. Much has been learned about mammalian genes by inserting them into yeast and then studying how they work and what proteins they make. Scientists have sequenced the genome of *S. cerevisiae*, as well.

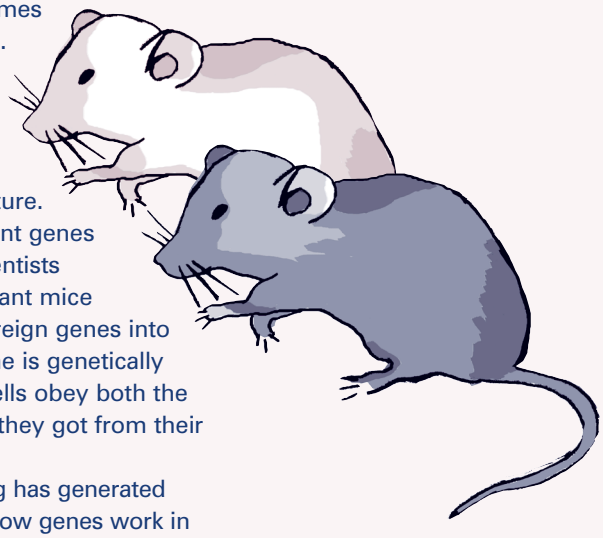


Mus musculus: The Mouse

The evolutionary lines that led eventually to mice and to human beings split off from each other 75 million years ago, back in the dinosaur age. But we are both mammals, and scientists say we share an astonishing 85 percent of our genes. So researchers can use mouse genes to find and study human genes, including those that cause disease. Scientists can also use mice to test drugs, devise new treatments, and study mammalian physiology and biochemistry—in sickness and in health—in ways not possible in humans.

In addition, mice can have diseases that are very similar—sometimes identical—to human diseases. Those mice are exceptionally valuable for research.

Until recently, mice with mutant genes that produce disease were accidents of nature. But mice with particular mutant genes are no longer accidental. Scientists can now make their own mutant mice to order. They put specific foreign genes into mouse embryos. The outcome is genetically engineered animals whose cells obey both the foreign genes and the genes they got from their mouse parents.



Mouse genetic engineering has generated a flood of information about how genes work in specific cells and how they contribute to health or disease—not just in mice, but in people, too.

CHAPTER 3

What Is Basic Research, and Why Do It?

The way scientists like Alan Lambowitz choose what to investigate is pretty typical. Scientists often pick out a topic to study because they want to solve puzzles, to learn the answers to very general questions that can then be added to the immense library of human knowledge. This kind of research is known as basic or untargeted research. Most of the scientists we've been discussing do basic research. They want to piece together the answers to very broad questions like: How do cells work?

The other kind of research is applied or targeted research. Applied research is designed specifically to find solutions to some practical problem. Clinical research on particular diseases is an example. It asks questions like: Does this new drug work well in people with colon cancer?

In the United States, much scientific research is supported by the Federal Government—in other words, by taxpayers. A significant portion of this support goes to basic research. Is that a good approach? Should taxpayers be paying for research that isn't directly targeted to curing disease?

There are three good replies to these questions:

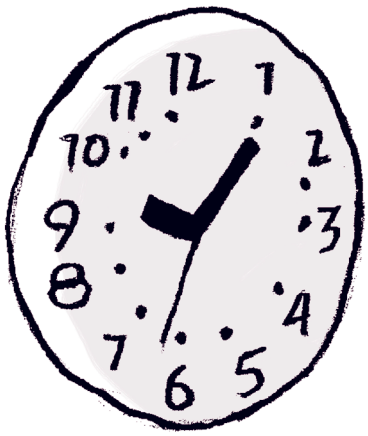
- Explaining underlying mechanisms has broad applications to many areas of science.
- Understanding normal processes helps us understand what goes wrong in disease.
- The findings are often utterly unexpected, but suggest fruitful new directions for research.

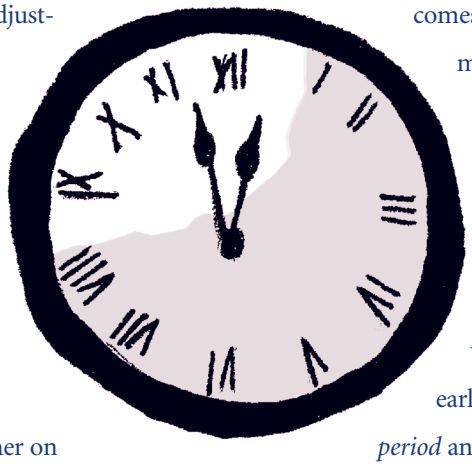
In short, basic research and applied research are closely related. Answers to the very broad question about how cells work will help answer the very specific question about the best way to medically treat a type of cancer.

Living Clocks

Research on biological clocks is a great example of the principle that explaining underlying mechanisms has broad applications to many areas of science. All living things possess these clocks, which govern the regular rhythms of life: waking, sleeping, eating, reproducing, and even seasonal rhythms such as birds flying south for the winter.

Biological clocks are important in physical and mental health. Some medicines and surgical treatments appear to work best at certain times of day. Some forms of insomnia and manic-depressive illness result from biological clock malfunction. Many people must work at night or other unusual times but have difficulty adjusting to their schedules. And anyone who has crossed the country or the ocean by plane has probably suffered from that traveler's misery called jet lag, where the body is forced to adapt quickly to a new time zone.



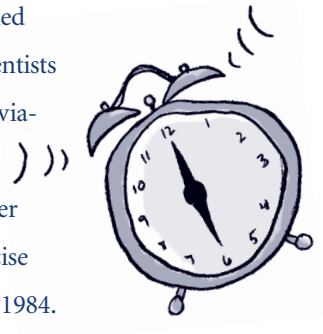


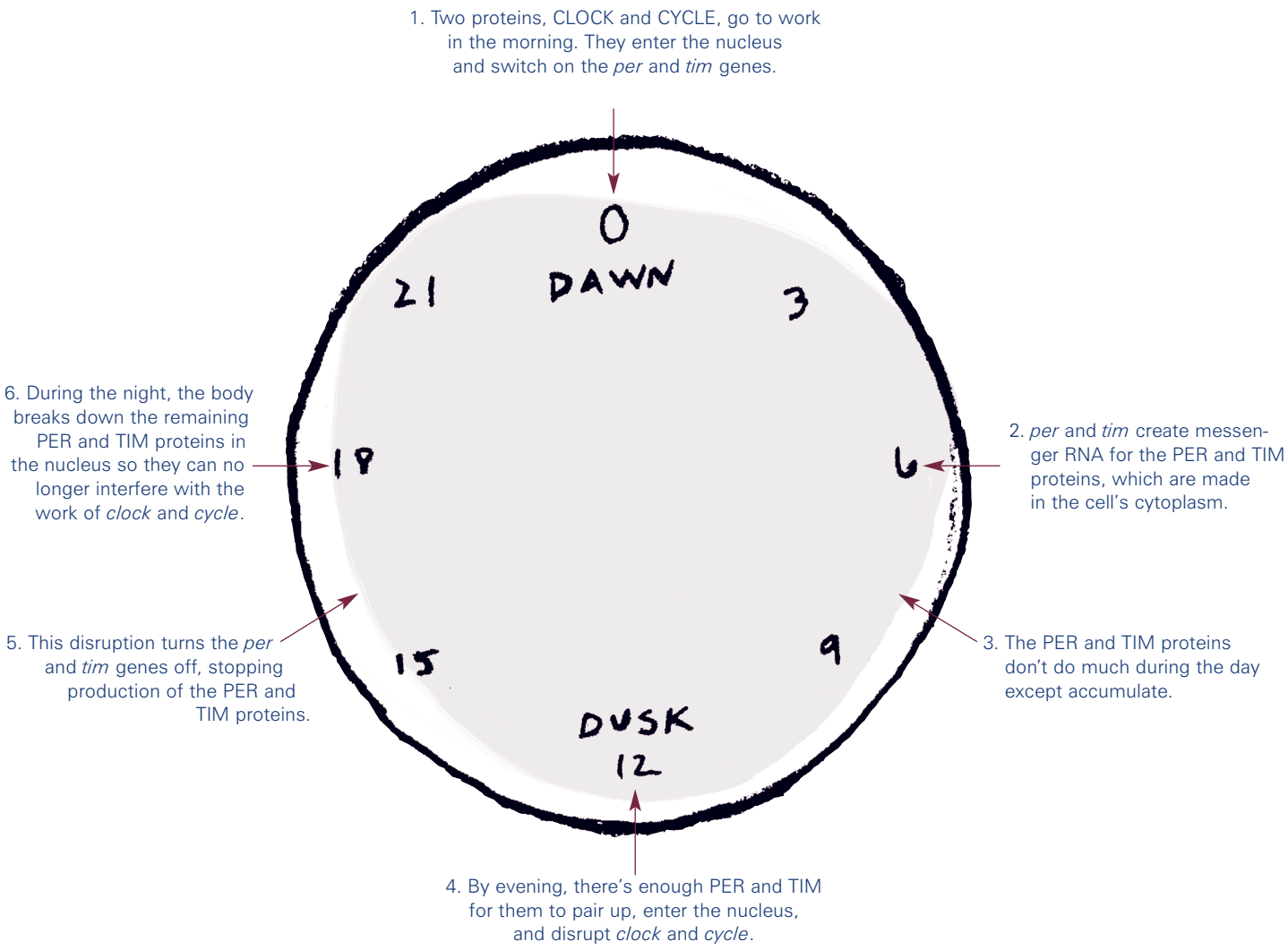
The biological clock is a small group of genes that switch each other on and off in a regular cycle. The switches control the expression of the clock genes; that is, how and when each gene produces its characteristic protein. Thus, the clock genes control each other indirectly through their protein products.

This oscillating pattern of proteins switching each other on and off is one of nature's favorite ways of getting things done. Such oscillation is extremely common in living things. It is called a feedback loop. A feedback loop is a self-regulating, closed control system in which one event causes other events, which then feed back to change the original event. The biological clock's proteins turn each other on and off over a period of about 24 hours, accounting for our physiological "day."

Scientists call this 24-hour oscillation a circadian **(sir-CADE-ee-an)** rhythm. ("Circadian" comes from the Latin words meaning "approximately a day.") All living things—plants, animals, and bacteria—possess a circadian rhythm.

The first clock gene was discovered, in fruit flies, early in the 1970s. It is called *period* and nicknamed *per*. (Scientists often write gene names in italics and the abbreviations for protein names in all capital letters.) Michael Young and his colleagues at Rockefeller University in New York City identified its precise location in the genome and cloned the gene in 1984.





▲ How scientists think the fruit fly clock works.

Scientists knew other genes were part of the clock too, but it took 10 hard, discouraging years of work before Young found the next one, *timeless*, or *tim* for short.

Clock genes work like this: All day long the *per* and *tim* genes in the cell nucleus make the messenger RNA that directs the construction of the PER and TIM proteins. In order to complete the cycle, the proteins must get back into the nucleus so they can turn the *per* and *tim* genes off.

Neither protein can do this by itself. The two must link together to re-enter the nucleus. As the PER and TIM proteins accumulate in the cell, they begin to bump into one another and stick together. Linked in that way, the two proteins form a sort of key that unlocks the nucleus and permits them to enter it.

The *per* and *tim* genes form half of the fruit fly clock's feedback loop. The other half of the loop is two other genes whose protein products enter the

nucleus. Michael Rosbash and Jeffrey Hall, who investigate fruit fly biological clocks at Brandeis University in Waltham, Massachusetts, found the two new genes in 1998 and named them *clock* and *cycle*.

At least one other gene is crucial to the circadian clock. *Double-time*, a gene Young discovered in 1998, explains why the clock's feedback loop stretches over 24 hours. The double-time protein tags the PER protein for destruction, which slows down PER's accumulation in the cell. That's why it takes several hours for the TIM protein to find enough PER for pairing. The pairing also protects PER against double-time's assaults.

As with most other genetic traits, nongenetic factors are also essential to the body's timekeeping. Circadian clocks are self-starting and will tick on until death, but they are not very accurate. Left to themselves, they tend to run either fast or slow. So circadian clocks must be reset every day to stay on a precise 24-hour schedule.

What sets body clocks? Light, mostly. Circadian rhythms track the sun and stay on time as long as the regular alteration of day and night can adjust them. How? Researchers discovered the fruit fly's method in 1998: Light destroys the fly's TIM protein. As a result, there isn't enough of the protein to begin the essential PER-TIM pairing until nightfall.

Rosbash and Hall identified a fruit fly gene involved in this process in 1998. It is called *cry*, and it makes a protein called cryptochrome that is sensitive to the blue part of incoming light, which is most common at dawn and at dusk. Cryptochromes are found in both plants and animals, and *cry* genes involved in the mouse circadian clock were discovered in 1998. Hall says he expects that other light-sensitive genes will be discovered in the circadian clock.

Jay Dunlap and his colleagues at Dartmouth Medical School in Hanover, New Hampshire, study the clock feedback loop in *Neurospora* (**nurr-OSS-por-ah**), a kind of bread mold. His group was the first to show that light resets the clock. The researchers also found two genes essential to *Neurospora's* clock that are particularly intriguing because their proteins regulate response to light—and yet they also work in the dark.

In mammals, the body's master clock is a group of about 10,000 cells in a tiny sliver of brain located behind the eyes, called the suprachiasmatic nucleus or SCN. Scientists now know that the fruit fly clock is very similar to the SCN clock in laboratory mice. Scientists have not yet figured out the clock in people, but because both humans and mice are mammals and the mouse and fruit fly clocks are alike, they are expecting the human clock to work pretty much like the fruit fly clock.

Programmed Cell Death

A second reason for doing basic research is that understanding normal processes helps us understand what goes wrong in disease. There are countless examples. One of the best is cell death.

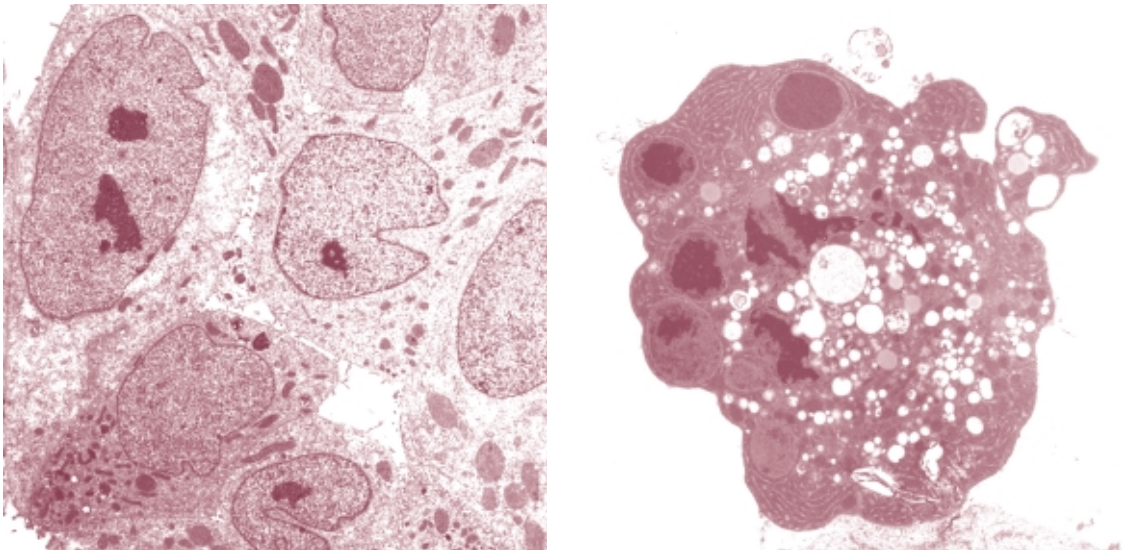
Cells contain the seeds of their own destruction. The seeds are proteins that kill from within, commanding other molecules to demolish a cell by smashing its internal structure. This programmed cell death, known as apoptosis (**a-poe-TOE-sis** or **a-pop-TOE-sis**) is quite different from another kind of cell death, necrosis. In necrosis, cells die because they have been dealt a fatal blow from outside. Apoptosis, by contrast, is a completely normal process in which cells perish in an orderly, highly controlled manner.

Odd as it may seem, programmed cell death is essential to life. It is both a sculpting mechanism and a control mechanism. Cells die in the developing embryo as a natural step toward building a new organism. In adults, cells die during normal tissue turnover and as part of the immune response.

A regular schedule for apoptosis is so important that when it goes awry the results can be devastating. Apoptosis triggered at the wrong time and place can cause the cell loss accompanying two of our most devastating degenerative brain diseases, Alzheimer and Parkinson. On the other hand, if apoptosis fails to occur when it should—for example, after a cell's DNA has been badly damaged—the reverse can happen: out-of-control cell growth and cancer.

One approach to dealing with such diseases would be to find ways of turning apoptosis off and on. Hermann Steller of the Massachusetts Institute of Technology in Cambridge investigates apoptosis with that thought very much in mind. His near-term goal, however, is firmly rooted in basic science. He seeks to lay bare the enormously complicated mechanisms that lead to programmed cell death. "We are trying to understand how cells kill themselves and how cells make the decision to live or die. That decision is influenced by many different signaling systems," he explains.

His organism of choice for studying apoptosis is the fruit fly. Steller and his colleagues first discovered a striking gene alteration that prevents normal cell death in embryos. The researchers found that the genetic change had deleted three genes, which they named *reaper, grim,* and *hid*.



▲ An intact prostate cancer cell (left) compared to a prostate cancer cell undergoing apoptosis (right). The white blobs in the cell on the right are a hallmark of apoptosis.

Electron micrographs by Robert Munn, University of California, Davis. Courtesy of Ingrid Wertz, University of California, Davis.

These three genes can substitute for each other to some extent, and they cluster together as a group in the fruit fly genome. Why have the genes stayed together? Steller believes it's because the same DNA regulates the production of all three.

Reaper, grim, and *hid* are different from other components of the programmed cell death pathway. Unlike protein-destroying enzymes called caspases that are present in inactive form in cells at all times, *reaper, grim,* and *hid* are turned on only when a cell decides to commit suicide. Their job, apparently, is to activate the caspases.

Steller and his colleagues discovered in 1994 that *reaper* is expressed in nearly every fruit fly cell that will die during normal development. It can also be activated in response to just about any harmful stimulus that can induce apoptosis—for example, radiation, defects in cell division, or other kinds of injury or stress. Furthermore, when the reaper protein is put into cells that are supposed to live, the cells die.

"That was very surprising and also very informative. It suggested to us that perhaps *reaper* is sort of a meeting point for different signaling pathways," Steller reports. "It is like a messenger for death, or like a car key that turns on the death engine."

The reaper protein has similar effects in mammalian cells, but with some interesting differences. "Some cells get killed by reaper and other cells are relatively resistant. We have some speculation on why that may be, but we don't quite fully understand it yet."

Steller and his colleagues have already demonstrated that interfering with cell death may someday prove to be a practical approach to fighting disease. In people who have a condition called retinitis pigmentosa, cells in the retina of the eye degenerate, eventually leading to blindness. In 1998, the scientists prevented blindness in a fruit fly version of this human disease simply by preventing apoptosis. "We showed that the retina cells continue to function if we can keep them alive. They are not perfect—they are a little impaired—but they provide rather good vision."

Steller points out that this is not a way to cure retinitis pigmentosa in humans because the flies had been genetically modified with a protein that inhibits caspase. "It's not really practical to think of doing exactly what we did in the human eye. But if we had a drug that would, very selectively, identify the signals that the cells used to turn on the death machinery that leads to retinitis pigmentosa, then we have a possible way to keep people with retinitis pigmentosa from going blind. And that kind of logic can be extrapolated to other disease situations where a lot of cells die by apoptosis."

An Unexpected Discovery About Chromosome Tips

A third reason to ask the most basic questions about the processes of life is that sometimes what you find out is utterly unexpected, but it suggests fruitful new directions for research.

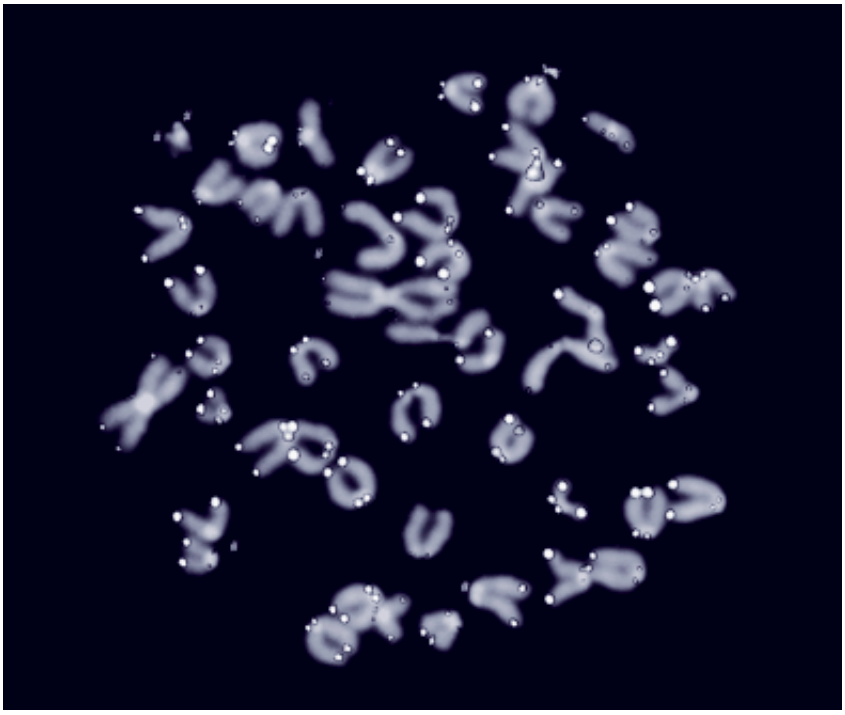
This happened to Elizabeth Blackburn of the University of California, San Francisco. She wanted to understand some of the basic events that go on inside our cells. "And because the fundamentals are pretty similar from one organism to another, you just choose the best experimental system," she says. The system she chose was *Tetrahymena* (**tet-rah-HY-meh-nah**), a single-celled organism that lives in ponds. The tiny, pear-shaped creatures are covered with hairlike cilia that they use to propel themselves through the water as they devour bacteria and fungi. For her, *Tetrahymena* was the best organism because it has a lot of the cellular component that she wanted to study: chromosomes.

In the 1970s, scientists like Blackburn were very curious about the end caps on the tips of chromosomes. Called telomeres (**TEE-low-meers**), the end caps seemed to keep the chromosomes, and the cell that they were in, stable. Chromosomes without these special end caps stick to each other and cause cells to divide abnormally. Blackburn likes to compare telomeres to the hard little tips at the ends of shoelaces. Shoelaces with no tips fray and unravel. Chromosomes without telomeres fray and unravel, too.

Her research was perfectly timed. Methods for sequencing DNA were just being developed. Blackburn found that *Tetrahymena's* telomeres contain an unusual arrangement of DNA: the nucleotide sequence TTGGGG, repeated over and over. (The repeats averaged out to about 50 per telomere.) Since then, scientists have discovered that the telomeres of almost all organisms contain repeated short segments heavy on Ts and Gs. Human and mouse telomeres, for example, contain the sequence TTAGGG, which can be repeated many times—in humans, from one to a few thousand copies per telomere.

The number of those repeats varies enormously, not just from organism to organism but in different cells of the same organism, and even in the same cell over time. This variation struck scientists as extremely strange. "The sequences didn't just sit there, they changed in different ways," Blackburn explains. "That led us to think perhaps there was some enzyme that adds DNA to the ends of previously existing DNAs."

Blackburn and her then-graduate student Carol Greider decided to look for such an enzyme. "By the end of 1984, we had seen enzymatic activity in the test tube that had the properties of this mythical enzyme," Blackburn says. "We did not stumble over it. We went into it with the idea that there would be such an enzyme because of the way telomeres behaved in cells. And then we deliberately went out and looked for it." The discovery of the enzyme, from then on known as telomerase (**tell-AH-mer-ase**), was a landmark in genetics and cell biology, and has earned Blackburn and Greider many honors.



▲ Chromosomes (in pale blue) have been prepared so that their telomeres appear white.

Digital image by Peter M. Lansdorp, BC Cancer Research Centre. Reproduced from *The Journal of Cell Biology*, 1997, Vol. 139, p. 311 by copyright permission of The Rockefeller University Press.

Why all the fuss about telomerase? Telomeres get worn down—in other words, they get shorter—as cells pass through division after division. Most normal cells stop dividing when telomeres wear down to a certain point. Eventually, the cells die. But telomerase can counteract that tendency to shorten. It adds to the lifespan of cells by adding DNA to telomeres and protecting them, making the telomeres stable once again.

So the discovery of telomerase triggered new ideas and thousands of new studies. It seemed as if the enzyme might be important in cancer and aging. Researchers were hoping to find ways to turn telomerase on or off so that cells would continue to divide, to combat aging, or so that cells would stop dividing, to combat cancer.

Blackburn continues to be at the forefront of telomere research. In the spring of 1999, she showed that telomerase can extend the lifespan of human cells without lengthening telomeres.

This seemingly amazing finding explained a puzzle in cancer research. Scientists had expected to find telomerase in cancer cells, encouraging them to divide and grow. But the research results had been inconsistent and confusing. Blackburn explains, "Often you see that telomerase is on, but not always. And telomere lengths are just all over the place."

Blackburn and her colleagues figured out why. In a test tube, they added the gene that turns telomerase on to a group of human cells that were on their way to becoming cancerous. Normally, those cells would have died, because cells contain fail-safe mechanisms that destroy them when they are damaged. But switching their telomerase on saved them, kept them dividing, and even reduced the frequency of abnormal chromosomes.

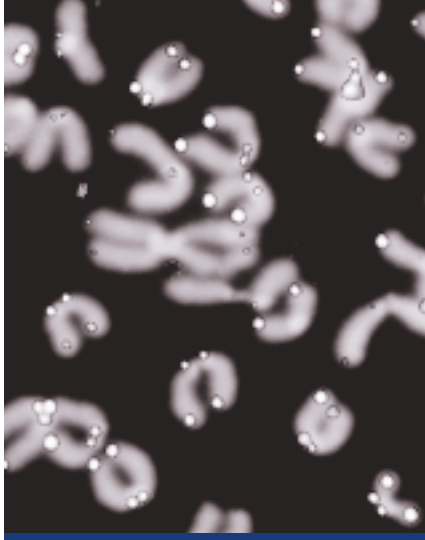
Most astonishing, the cells survived and kept on dividing as their telomeres got shorter and shorter—even when they were shorter than telomeres in cells that had stopped dividing. What telomerase does is push even short telomeres into the capped state, which protects the ends. If telomeres are very long they can be capped without telomerase. But as they get shorter, they need telomerase to keep them capped.

So in human disease, telomerase can have two opposite effects. Blackburn often refers to them as the proverbial "good and bad guys," Dr. Jekyll and Mr. Hyde. "When telomerase is Mr. Hyde, it allows cells to proliferate [multiply] that shouldn't be allowed to proliferate," Blackburn explains. Because the cells have already gone a few steps on the road to cancer, the cells' genomes become more and more unstable. "Normally, what happens when cells undergo a lot of genomic instability is that they do themselves in, they crash and burn.

"But a rare cell, about one in a few million, crawls out and survives. When we put telomerase into these cells they keep on proliferating. So now we have produced a huge population of cells that should have crashed and burned. We have increased the chances that these cells will progress to cancer. So telomerase has had a bad effect because it has promoted cancer-causing events. Telomerase is letting cells that normally would self-destruct keep on dividing."

When telomerase is the good guy, Dr. Jekyll, it protects cells from a certain type of genomic instability. In normal cells that are not on their way to becoming cancer, having telomerase turned on is usually good, because it protects the ends of chromosomes. "But if you have cells that are part way down the cancer path already, I'd keep telomerase a million miles away, because it will allow dangerous cells to proliferate. And yet both outcomes are due to exactly the same property of telomerase: protecting the ends of chromosomes."

Scientists are hoping to be able to manipulate telomerase action to treat disease. They are looking for ways to switch the enzyme off to keep cancer cells from multiplying. In some circumstances, however, they want to be able to turn telomerase on so that cells will continue to multiply—cells from a bone marrow transplant, for instance. Getting cells to keep multiplying might also prove useful in warding off or reversing certain kinds of aging processes, although no one knows for sure yet whether telomerase is one of the driving forces of aging.



Got It?

What is a biological clock? How is studying clock genes in other organisms relevant to human health?

Why is it important for human health to understand what telomerase does?

Why do scientists use animal models to study biological processes, rather than simply studying people?

Why do basic research if you want to learn more about diagnosing, treating, and preventing diseases?

CHAPTER 4

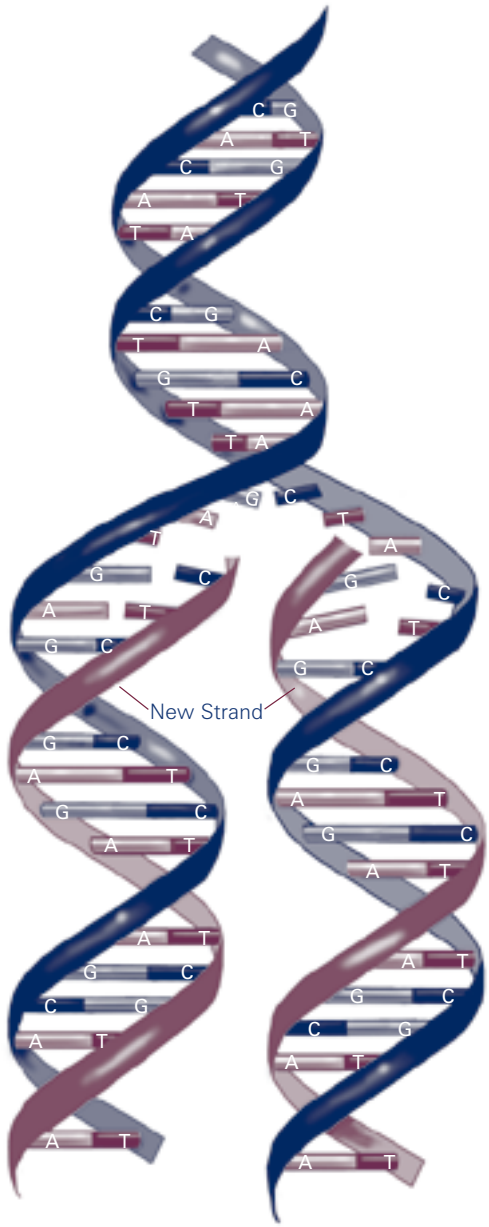
Genes and Disease

DNA Copying and Cancer

One of a cell's jobs is to make more cells exactly like itself. It does this by splitting in two. But before it divides, it must copy its DNA so there will be a complete genome to pass on to each of its daughter cells.

Duplicating DNA is no easy job. For one thing, there's a lot of it. If you could pull on one end of the DNA in a single human cell's nucleus and spread it out, it would be 2 meters long. Some experts have estimated that stretching out all the DNA from just one person would create a thin filament reaching all the way to the sun—and back. To fit into a nucleus, DNA is coiled, folded, and packed in tightly. DNA copies itself, or replicates, by unwinding its helical spiral and separating into two single strands. Each strand is a pattern, or template, for making an exact copy that becomes a new strand.

In their lab at the Massachusetts Institute of Technology, Stephen Bell and his colleagues are trying to understand the first steps in DNA replication. That's important because these first steps are controlled by the cell as it decides whether to divide or not. In particular, the scientists would like to find out how a cell's replication machinery knows where on the DNA to begin copying. Replication must go on at many sites simultaneously, because if it started at only one site on DNA, it would take the cell a month to copy a single human chromosome.



▲ DNA replication. Each strand of the original molecule acts as a template for the synthesis of a new, complementary DNA strand.

The normal replication time for an entire set of human chromosomes is between 4 and 8 hours.

Bell and his colleagues study a group of six proteins in yeast cells called the origin recognition complex, ORC for short, which Bell discovered when he was a postdoctoral fellow in Bruce Stillman's lab at Cold Spring Harbor Laboratory. This complex plays a central role in picking the sites on DNA where replication begins. ORC appears to be what marks the sites, and they remain marked throughout the cell's existence. "Whether they are dividing or not dividing, ORC sits on these sites and seems to be waiting to tell the cell where to start replicating its DNA when the time is right," Bell says.

There is only one kind of yeast ORC, but there are lots of copies of it, and they hook up with particular stretches of the organism's chromosomes. Bell has identified many of those sites. They all contain a particular sequence of DNA that appears at various points along the chromosome and that ORC always recognizes. Those are the "start" sites for replication.

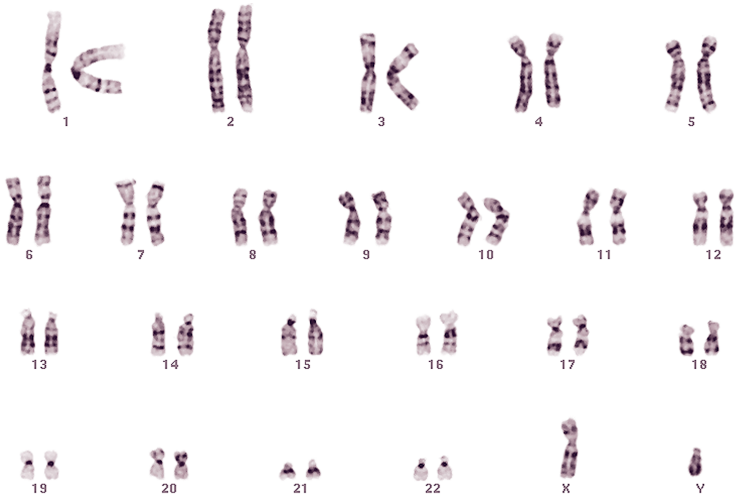
Having discovered where ORC binds, Bell is now trying to find out what it does when it gets there. One of his discoveries is that ORC recruits other proteins to the origin site, and that it is those proteins, not ORC, that duplicate the DNA. "You can think of ORC as a landmark on the chromosome," Bell says. "At the appropriate time, other proteins come to that landmark to initiate replication."

The details of how DNA gets replicated are central to a dreaded human disease: cancer. Cancer is cell division gone out of control. That's one reason why most chemotherapy is designed to disrupt the DNA replication process, in an attempt to halt that growth. Unfortunately, chemotherapy attacks all cells that are growing and dividing. This is why it affects the immune system and causes hair loss, since—like cancer cells—immune cells and hair cells divide often.

It can be difficult to distinguish between cancer cells and normal cells that are supposed to divide frequently. Understanding replication, Bell points out, could be a key to confining a drug's attack to cancer cells only.

Chromosomes and Birth Defects

DNA is organized into individual packages, remarkable structures called chromosomes. Each chromosome consists of a single huge molecule of DNA studded with genes, like beads on a string, as well as some accessory proteins. The number of chromosomes is usually the same in all individuals of a particular species. Humans possess 46 chromosomes in each body cell, 23 pairs, one member of every pair from each parent. Chromosomes are classified and numbered according to size. Human chromosomes are numbered from 1 to 22. The remaining pair are the sex chromosomes, two Xs or an X and a Y.



▲ Chromosomes of a normal human male.

Courtesy of Cytogenetics Laboratory, Brigham and Women's Hospital

Body cells are called "diploid" because they have two sets of chromosomes. The diploid mouse cell has 40 chromosomes (20 pairs) and a diploid fruit fly cell has 8 (4 pairs). Eggs and sperm are known as "haploid" cells. Each haploid cell has only one set of chromosomes (in humans, 23 chromosomes; in mice, 20; and in fruit flies, 4), so that at fertilization a haploid egg cell will combine with a haploid sperm cell to form a diploid cell with the right number of chromosomes.

Before a body cell divides, it must duplicate its chromosomes so that it can pass a complete set on to each of its daughter cells. The job of making sure that each daughter cell gets the correct number and kind of chromosomes during cell division does not always go smoothly. Sometimes a cell ends up with too many or too few chromosomes.

In humans, abnormalities of chromosome number usually occur very soon after fertilization and are usually lethal. Scientists suspect that this sort of abnormality is responsible for a large proportion of miscarriages. But some types of chromosome abnormalities are not fatal, and the result is a baby with the wrong number of chromosomes.

These babies almost always have health problems, often quite serious ones. Many babies with abnormalities of chromosome number are mentally retarded. In fact, the most common form of mental retardation, known as Down syndrome, is due to an extra chromosome. A person with Down syndrome has three copies of chromosome 21 instead of the usual two.

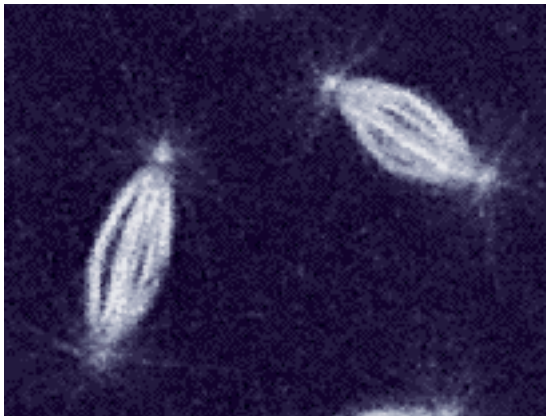
So scientists would like to understand more about how chromosomes behave. One such scientist is Sharyn Endow of Duke University in Durham, North Carolina. For some years, she has studied the forces that underlie the movements of the chromosomes during cell division.

Every cell contains complex transportation systems that ferry cell components from place to place. The tiny biological transportation systems operate on energy supplied by proteins that are known as molecular motors. The motors move around in cells on minute filaments, called microtubules, rather like trains move on tracks. During cell division (mitosis), chromosomes attach themselves to a bundle of microtubules called the mitotic spindle. The chromosomes move to opposite ends of the spindle so that when the cell splits, each half will contain a complete set of chromosomes.

The microtubule motor proteins that power this movement consist of families of related molecules. One family is known as kinesins. Endow and her colleagues discovered Ncd, one important member of the kinesin family, in the fruit fly *Drosophila melanogaster*. They also discovered that the Ncd motor binds to spindles and spindle poles and may help chromosomes attach to the spindle.

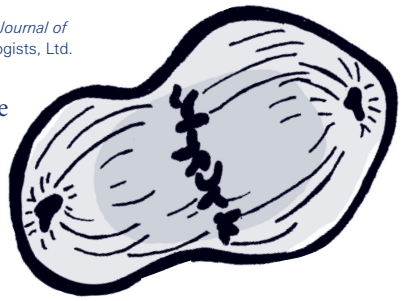
Ncd is a fruit fly protein, but similar microtubule motor proteins operate in all animals, including humans. However, the Ncd motor is unusual. All of the other kinesins shuttle cargo in one direction along the microtubules. Ncd goes the other way; it runs in reverse. Yet when scientists looked at the motors in detail, they seemed to be almost identical.

Endow and her colleagues decided to find out how Ncd works by taking it apart and putting it back together. They studied the structure of the protein and identified several of its parts, which



Sharyn Endow

▲ Mitotic spindles pull apart the chromosomes in a fruit fly embryo. The drawing below shows the chromosomes aligned in the middle of the cell and attached to the mitotic spindle prior to the separation of each pair of chromosomes.



they and other researchers call the motor, the neck, and the stalk. The motor, of course, powers the protein. But what do the neck and the stalk do?

To find out, the scientists created a hybrid protein with a motor from another kinesin and a neck and stalk from Ncd. This hybrid protein runs in reverse, just like Ncd. This meant, they decided, that the neck and stalk must determine the direction the motor runs. They demonstrated that they were right by altering the gene that produces the neck, which kept the motor from working properly. The motor then ran forward, but very slowly, rather than in reverse like the unchanged hybrid protein.

"For the first time, we have been able to identify a component of a motor protein that is responsible

for determining its direction of movement and may help coordinate motor movement," Endow says.

Endow and her colleagues photographed the chromosomes and molecular motors in action as the chromosomes separated in fruit fly eggs and embryos, producing some of the first detailed moving pictures of chromosomes being parceled out into individual fruit fly eggs. They also made movies of what happens to chromosomes in early embryos when Ncd is not working properly. These movies can be viewed on the World Wide Web at http://microbiology.duke.edu/labs/endow/moviepage.html.

Endow has found that other kinesins, in yeast and in a plant, are "reverse" molecular motors. She thinks that both the yeast and plant motor proteins and Ncd belong to a family of reverse kinesin microtubule motors that are probably found in all eukaryotes. The reverse motors, she suspects, may help attach components of the cell division apparatus to one another and also help chromosomes move to the cell poles by sliding microtubules in that direction.

From Fly Lungs to Human Cancer

How does an animal encode in its genes the program for making a complex, three-dimensional structure like an organ?

The human lung, for example, is basically a branching network of tubes. But there are millions of these tubular branches in each lung, and each tube must be just the right size so that smaller tubes always sprout from bigger ones. How is this branching pattern controlled during embryonic development so that it gives rise to a network of ever-smaller tubes that transport oxygen to the bloodstream efficiently?

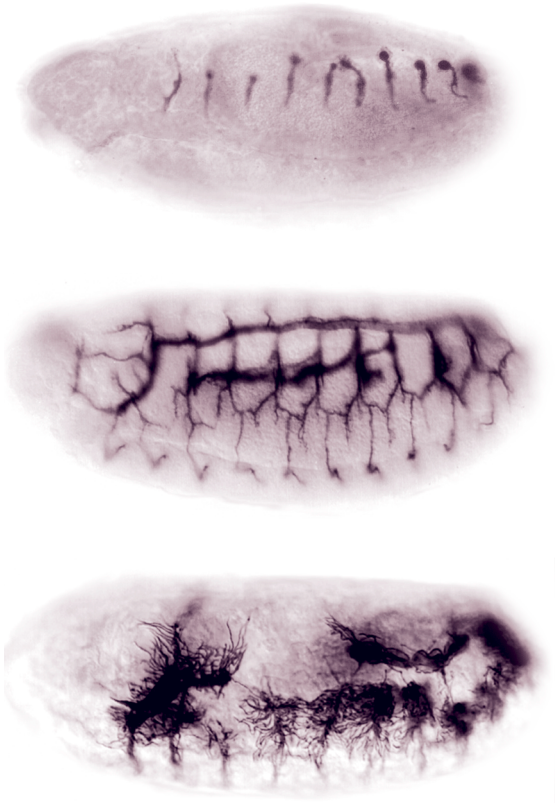
Scientists have known for a long time that the program does not generate branches randomly, because the lung's network of tubes is quite similar from one person to the next. Since there is a standard design for the human lung, that design must be in our DNA instruction manual.

For a decade, Mark Krasnow of Stanford University School of Medicine in California has been trying to figure out how living creatures build their branching organs from standard designs encoded in DNA. What are the genes that make the proteins that carry out this elaborate branching program? When do the genes get turned on in development, and once they have turned on, how do they control the events in cells that sprout tubular structures?

Krasnow began his investigation with the fruit fly, which he chose because he thought he could make progress quickly. The fruit fly has no lungs, but it does have airways that transport oxygen to every part of its body. And these airways (called the trachea) branch in a consistent pattern, just like the lungs of mammals. There are some 10,000 of these branches, but the pattern is less complicated and the branches are not nearly as numerous as they are in mammals.

The fruit fly tracheal system arises in the larva from 20 sacs, each composed of about 80 cells. Each sac sprouts successively finer branches that grow into a treelike form. The branching pattern is remarkable. It occurs because cells move around and change shape, not because they increase in number.

At the first level of branching, groups of cells organize themselves into tubes. The second level of branching occurs several hours later, as smaller branches sprout from the ends of the first branches. These secondary branches are made of individual cells that roll up to form tubes. Then they develop into dozens of terminal branches, the third (and final) level of branching. (Human lungs have 20 levels of branching.) Each sac eventually creates about 500 branches, and some of these fuse with branches from neighboring sacs to form a network of some 10,000 tracheal tubes.



Three fruit fly embryos with different *branchless* gene expression. The middle embryo has normal *branchless* genes and its trachea is developing correctly. The top embryo is a *branchless* mutant in which very little tracheal branching has taken place. The bottom embryo possesses *branchless* genes that are switched on in most of the embryo's cells instead of just in the cells where they are supposed to work. As a result, the embryo is jammed with massive networks of fine branches instead of the normal tracheal branching pattern.

Courtesy of Mark Krasnow

Krasnow and his colleagues used the standard strategy for fruit fly geneticists, which is to search among thousands of flies with genetic alterations looking for ones with specific abnormalities—in this case, abnormalities in the airways. They then selected for further study those flies that had defects in the branching pattern.

More than 50 genes are now known to be involved in tracheal branching. The scientists showed that different gene mutations blocked the

process at different stages of branching. In some cases there was no branching at all; instead of a trachea, the mutant flies possessed only unbranched sacs of cells. One of those genes they named *branchless*, and another group of researchers named a related gene *breathless*. In other mutant flies, there was sprouting of the initial branches but all subsequent branches were blocked. In still others, primary and secondary branches sprouted normally but the fine terminal branches were absent. The existence of these variations showed that separate genes are required for each of the different stages of branching.

Krasnow began to make significant progress in revealing how the newly discovered genes created three-dimensional branching patterns when he started to identify the proteins the genes encoded. The key to branching turns out to be *branchless*.

The *branchless* gene makes a protein called fibroblast growth factor (FGF). The FGFs are an important family of signaling molecules. All animals have them. Even though *branchless* is a fruit fly gene, the protein it makes looks very much like the FGFs found in mammals. That means that the reverse is likely to be true, too: Mammals (in fact, all vertebrates) probably have genes that look and act like *branchless*. In fruit flies, the FGF protein not only makes new branches sprout, it also changes the receiving cells so that during the next stage of branching they interpret the same signal in a slightly different way. This leads to smaller branches on the next level.

An FGF gene in the mouse, the FGF10 gene, has recently been shown by other labs to play a key role in branching of the mouse lung. FGF10 turns out to have a similar structure as *branchless*, and it plays a similar role. The FGF10 gene is also turned on in clusters of cells that surround the embryonic mouse lung and appears, like *branchless*, to direct the branch-sprouting pattern. Mice that have no FGF10 gene do not grow lungs. There is a human FGF10 gene too, but it hasn't been completely studied yet. Says Krasnow, "I think the assumption by everyone is that it is going to be very nearly the same."

Krasnow expects that research done in his lab will eventually help people with lung diseases. "By understanding the genetic program for branching in sufficient detail, we should, hopefully, be able at some point in the future to trigger that program, start the program up again at any time or place in development that we want," he says. Doctors might, for example, be able to generate a new lung, when the existing lung has been damaged by disease, by simply turning on in the diseased adult lung the genetic program that usually generates a lung only in a fetus.

But Krasnow has a broader purpose in mind, too. He thinks what is discovered about branching

patterns in the airways may well shed light on other branching patterns in the body—especially the system of arteries and veins that ferry life-giving oxygen from our lungs to every part of our bodies and haul away the waste products for disposal.

The fruit fly tracheal system does double duty, combining the jobs performed separately by the lungs and the circulatory system in mammals. As it turns out, the sprouting of the fine terminal branches of the fruit fly tracheal system is very much like the vascular system in mammals, which sends capillaries to all the internal tissues of the body to supply them with oxygen and nutrients. This third level of sprouting is not regulated by a fixed developmental program that generates consistent patterns of branching like the first two levels. Terminal branches are highly variable, and their sprouting is regulated by the oxygen needs of the target tissues. Tissues that need more oxygen (and therefore more branches) arrange to get it by increasing expression of the *branchless* FGF, which triggers the sprouting of additional terminal branches.

Krasnow suspects that the vascular system in mammals probably develops in a similar way because the fine blood capillaries, too, vary according to the demands of the tissues they supply. That variability may turn out to be relevant to human disease. Human tumors, for example, must develop a new blood supply in order to grow. Krasnow has shown that to be true in fruit flies, too. Fruit fly tumors grow so fast that they outstrip the available oxygen. This triggers expression of FGF, which causes the sprouting of more branches.

Krasnow hopes that learning more about the vascular branching process will eventually help scientists learn how to turn off a tumor's blood supply and so starve it to death. Or, vice-versa, scientists could learn to turn the branching process back on to create a new blood supply to nourish a faltering heart.

Got It?

How does a cell know where to start replicating its DNA?

What is the connection between DNA replication and cancer?

Why is it important to human health to understand how chromosomes move?

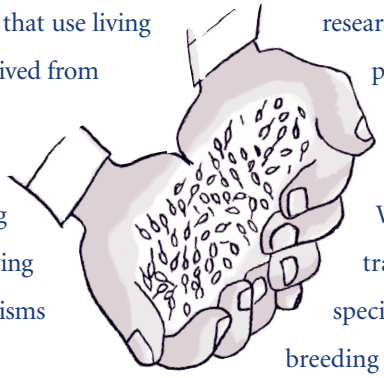
How did the study of genes that affect the development of fruit fly lungs lead to an idea for a way to control tumors in people?

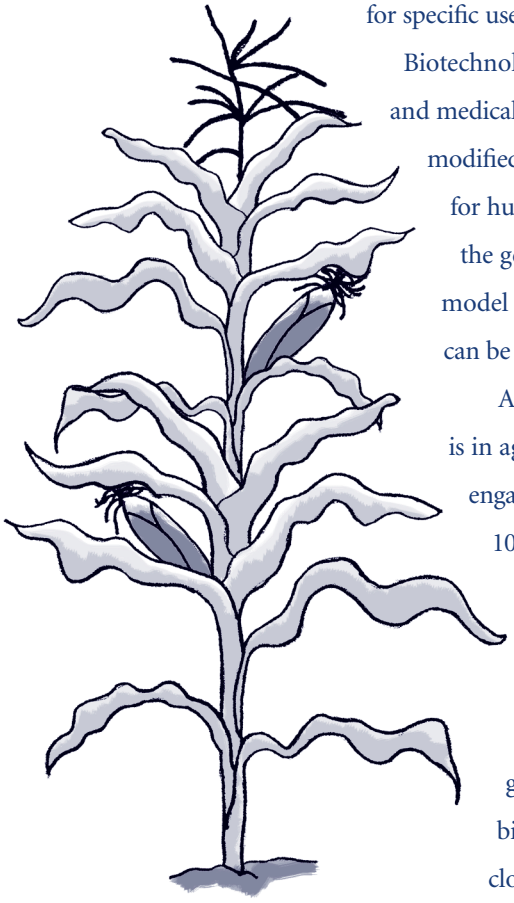
CHAPTER 5

Genetics in the 21st Century

The New Biotechnology

The applications of genetics research are often lumped under the term biotechnology, especially if they lead to products for human use. Biotechnology is usually applied, rather than basic, biological science. It involves techniques that use living organisms—or substances derived from those organisms—for various practical purposes. The term can mean making or modifying a biological product or improving plants, animals, or microorganisms for specific uses.





Biotechnology is often used in medicine and medical research. Getting a vat of specially modified bacteria to produce a medication for human use is biotechnology. So are the gene transfer techniques for making model organisms, like knockout mice, that can be used in disease research.

A major application of biotechnology is in agriculture. In a sense, humanity has engaged in agricultural biotechnology for 10,000 years or more. Many traditional farming practices, from plant breeding to animal husbandry, can be classified as biotechnology.

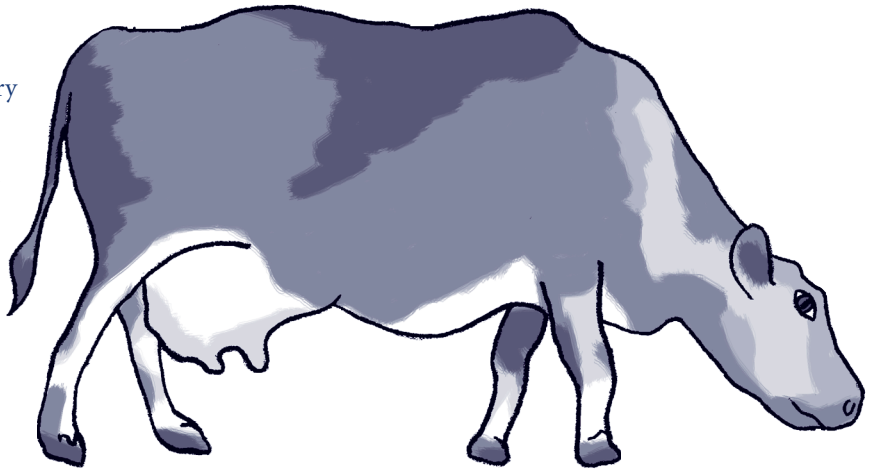
But today, the term biotechnology generally means the use of molecular biology, recombinant DNA technology, cloning, and other recent scientific approaches to produce plants and animals with new traits.

In some cases, this means transferring genetic material from one kind of organism into another. Just as scientists create transgenic mice for research, they also create transgenic crop plants and animals for people to eat.

This is quite different from the traditional breeding of plants and animals. Why? Because it involves the precise transfer of a single known gene with a specific practical end in mind. Traditional breeding is a lottery, the random recombination of entire genomes in the hope that some of the new combinations will have desirable characteristics.

A portion of the corn, soybeans, and cotton grown in the United States comes from seeds that have been genetically modified to resist viruses and other plant pests. Some argue that genetic modifications like these are the only hope for pest-ravaged crops, such as bananas, that are essential to the economies of poor countries. Others would like to invent edible plants that contain medicine, serve as a form of vaccination, or deliver extra nutrients—such as the recently developed rice that is endowed with vitamin A.

But opposition from farmers, consumers, and others has clouded the future for agricultural biotechnology. Some have objected to the development of plants that carry resistance to herbicides, partly out of concern that the trait might jump to



weeds, making them impossible to destroy. Others have expressed worry that pollen from modified plants could transfer foreign traits to their wild relatives, or that genetic modification will harm insects that benefit humanity. However, the U.S. Environmental Protection Agency has stated that there is no evidence to indicate that biotech crops have any unreasonable adverse effects on non-targeted wildlife, plants, or beneficial insects.

Another problem agricultural biotechnology has encountered is that shoppers may be unwilling to pay a premium for desirable traits. This was the fate that befell the Flavr Savr™ tomato. This genetically modified tomato was slow to soften and rot, which meant that it could be picked ripe. The Flavr Savr was marketed in 1994. But difficulties with packaging and distribution, plus consumer resistance to paying more, forced its maker to stop producing it in 1996.

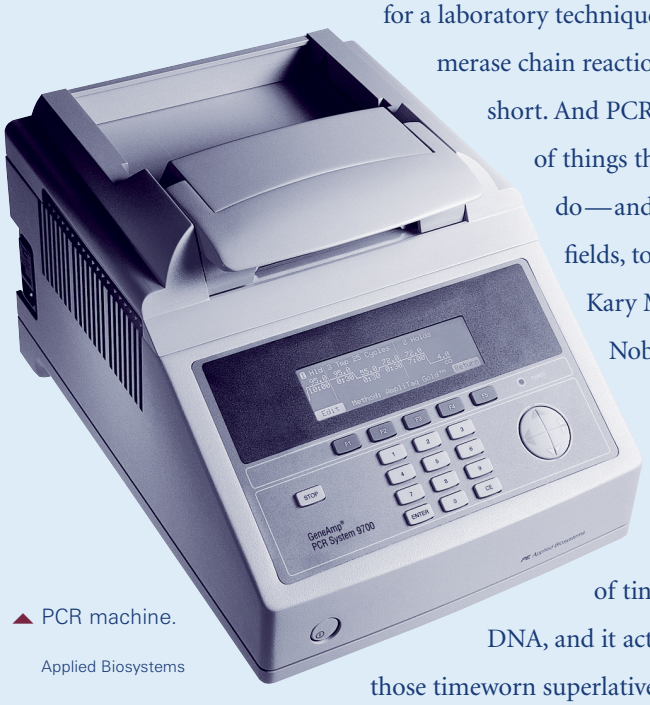
Non-agricultural uses for biotechnology are less controversial. For decades, pharmaceutical companies have made use of living factories, especially bacteria, to produce drugs. The bacteria grow in huge vats called bioreactors—in much the same way as yeast cells are grown to produce wine from grapes and beer from grain. The mammary gland is a kind of bioreactor too, designed solely for the production of proteins suspended in fluid—we call it milk. So another new idea is "pharming": Scientists are experimenting with genetically modifying livestock such as cows and goats to produce particular proteins in the females' milk, such as Factor VIII, a blood clotting protein needed by people with hemophilia. These molecules made in living factories could be used for treating disease. Plans include drugs that prevent blood clots, therapeutic antibodies, and proteins for treating lung disorders.

Scientists also have high hopes for bioprospecting. That's what people call the search for naturally occurring microorganisms that can be harnessed for various tasks, including breaking down garbage in overstuffed landfills, mopping up oil spills, and turning sewage into fertilizer.

The Tools of Genetics: Unlimited DNA

Without bioprospecting, a number of research advances would never have happened. The amazing truth is that a microbe discovered in 1966 in a Yellowstone National Park hot spring is an essential ingredient for one of the most important research tools ever invented.

Thermus aquaticus (**THUR-mus ah-KWA-tih-kus**) is a bacterium that makes a heat-resistant enzyme, which is why it can thrive in hot springs. The enzyme, Taq (**TACK**) polymerase, is essential for a laboratory technique called the polymerase chain reaction, or PCR for short. And PCR is essential to lots of things that life scientists do—and to many other fields, too. PCR's inventor, Kary Mullis, won the Nobel Prize in 1993.



▲ PCR machine.

Applied Biosystems

PCR is a quick, easy method for generating unlimited copies of tiny amounts of DNA, and it actually deserves those timeworn superlatives like "revolutionary" and "breakthrough." PCR can help detect changes in genes, so it is the basis for much of the research discussed in this brochure. It also underlies diagnostic techniques like testing individuals for genes that cause breast cancer. PCR is a key element of "genetic fingerprinting," which has helped open the jailhouse doors for prisoners who relied on it to prove that they were innocent of the crimes that put them there. It has helped convict criminals, as well.

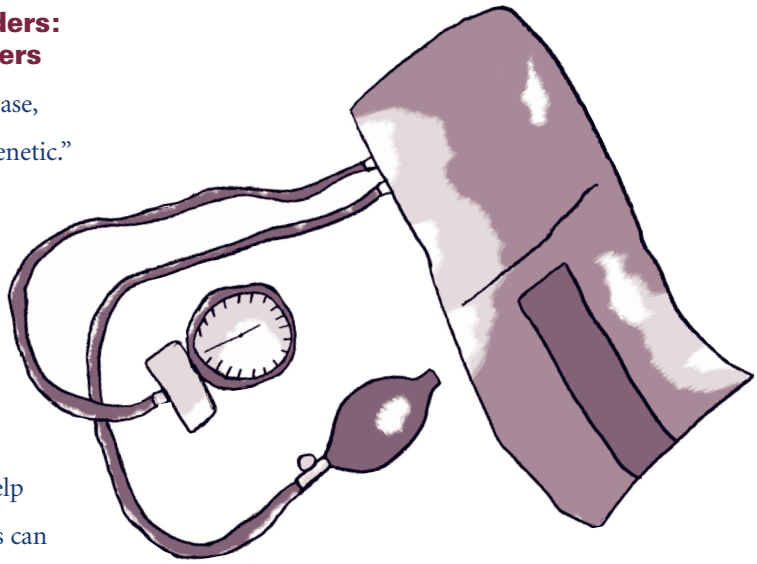
National Park Service

▲ A Yellowstone Park hot spring.

PCR can help track down infectious organisms and diagnose mystery diseases. It underlies modern DNA sequencing methods. It has revolutionized archaeology by helping to analyze even highly damaged ancient DNA, which can reveal new and sometimes unsuspected information about past people and cultures. It is an essential tool for evolutionary biology, helping to trace back the origins of a particular life form—including humans.

PCR has done for genetic material what the invention of the printing press did for written material. It makes copying easy, inexpensive, and available to scientists everywhere.

The Genetics of Complex Disorders: Lessons from Mice and Computers



Genes are involved in nearly all human disease, but that does *not* mean that all disease is "genetic." In fact, diseases caused by mutations in a single gene, such as sickle cell disease or cystic fibrosis, are not very common. Scientists have learned, however, that people vary greatly in their susceptibility to disease, even infectious disease, partly because of their genes. Certain genes can help people who have them resist disease. Others can make people especially vulnerable. Scientists expect that understanding genes will also shed light on the roles nongenetic factors play in resisting disease or succumbing to it.

Common diseases that kill millions of people every year and make millions more miserable are caused by a combination of genetic and environmental factors. These diseases are known as complex disorders. They include most cancers, heart disease, mental disorders, asthma, arthritis, diabetes, and many others. Scientists have turned their attention to these diseases more and more because they have become convinced that the best way to defeat them is to understand the complicated ways they develop. This knowledge will reveal the best methods of attack.

One of these complex disorders is hypertension, a fancy word for blood pressure that is too high. High blood pressure is often called the "silent killer" because it has no obvious symptoms. It damages organs like the kidneys, brain, and heart, and it can lead to death from heart or kidney failure or stroke. Like many other complex disorders, it is exceedingly common: An estimated one in four adult Americans has high blood pressure. Figuring out the causes of high blood pressure is a very high priority for biomedical researchers.

Oliver Smithies of the University of North Carolina, Chapel Hill, is one of these researchers. He and his colleagues are using knockout mice to study whether small genetic changes have measurable effects on blood pressure. (Along with Mario Capecchi, Smithies pioneered techniques for knocking out genes in mice.) He is taking this approach because scientists are beginning to suspect that high blood pressure is due to combinations of small genetic variations, many of which are normal variations that are not harmful by themselves.

John R. Hagaman, University of North Carolina

▲ How do you measure a mouse's blood pressure? With a device very similar to the one used in people—except that of course it is much smaller. And instead of placing the cuff around a person's arm, you slide it up over the mouse's tail.

It's an immense task. In the mouse, 50 or more genes may figure in the control of blood pressure, and at least that many are probably involved in human blood pressure control, too.

Knockout mice are extremely useful tools for figuring out what a particular gene does. However, cases of hypertension due to a disabled gene are very rare in humans. Rather, Smithies believes that hypertension in most individuals is the result of small changes in the amount of the protein the gene produces. So he and his colleagues developed a specific approach to gene targeting that generates animals in which the amount of a gene product is varied. They call it gene titration.

With gene titration, instead of shutting down a particular mouse gene completely, the researchers have it make less or more of its protein. The scientists can play with different combinations of these alterations and measure their effects on mouse blood pressure.

The control of blood pressure—involving normal variations in many genes acting in numerous combinations—is believed to be a good model for many complex disorders. If that turns out to be true, in the next few decades, model animals produced by gene titration are likely to contribute in a major way to understanding a variety of complex disorders and suggesting tools to control them.

Smithies began by making groups of mice that produced different amounts of a protein, called angiotensinogen (AGT), that is known to be involved in blood pressure. He varied the levels of AGT over a relatively modest range in individual mice, from half normal to normal to twice normal.

Then he and his colleagues measured the blood pressure of these animals. The results demonstrated that the amount of AGT does indeed cause differences in blood pressure. Mice with less than the usual amount of AGT had lower blood pressures, and those with more AGT had higher blood pressures. The effects were small, though; on average, systolic blood pressure varied by 8 millimeters of mercury. (Systolic pressure—the top number in a blood pressure measurement—indicates the force of blood on artery walls when the heart beats. Normal systolic pressure in the mouse is the same as it is in people, about 120 millimeters of mercury.)

"That was the first clue we got that small differences in the genetic material can cause small differences in blood pressure," Smithies recalls.

"We've now done that same sort of experiment with a number of other genes. We have altered the amount of product that these genes make and watched what happens to blood pressure." In some cases, the effect was the opposite of what had happened with AGT. One protein made in the heart, for example, reduces blood pressure when it's plentiful and raises blood pressure when the supply is low.

One of these experiments, however, gave a puzzling result. This study focused on a protein called angiotensin-converting enzyme (ACE). Some blood pressure drugs block the action of ACE and are known as ACE inhibitors.

When the researchers measured the blood pressure of mice that produced half-normal, normal, and twice-normal amounts of ACE, they were astonished to find that the genetic differences among the mice seemed to have absolutely no effect on their blood pressure. Why are ACE inhibitors so good at lowering human blood pressure by blocking ACE when varying the amounts of ACE genetically did not affect mouse blood pressure at all?

Smithies and his colleagues approached that question with a computer simulation of the pathway that controls blood pressure. The pathway starts with AGT. The liver makes the AGT protein, which is converted to a small molecule called angiotensin I by an enzyme in the kidney. ACE then converts angiotensin I to angiotensin II, which is, of course, why it's called angiotensin-converting enzyme. Angiotensin II is always present in the blood, and the more of it you have, the higher your blood pressure.

When the researchers varied the amount of AGT in the computer, to their delight the amount of angiotensin II varied over the sort of range that they had seen in the experiment. Then they varied the ACE over a similar modest range, and sure enough that also replicated their mouse results—there was no effect on angiotensin II. In the next simulation they reduced ACE drastically, the way the drug does, and angiotensin II dropped too.

"We were able to replicate in the computer what we had seen in our experiments. In some ways you can say that it is related to the dose of the ACE inhibitor," Smithies declares. "At low doses, which is what the genetic experiment [manipulations deliver], there's no effect on blood pressure," he says. ACE inhibitors in larger doses, however, reduce ACE so dramatically that blood pressure goes down.

So Smithies' current thinking—which he says is not proven but is a good hypothesis—is that blood pressure differences between most people are the result of a lot of little things but no one big one. And he thinks that the differences are not the same in all people.

"I think this is likely to be the explanation for a lot of the common complicated diseases that have genetic factors. The diseases are so common that if there were only a few genes involved, we would have found them already. But our hypothesis is that there are rather a lot of genes, each responsible for rather small differences. So it's quite hard to find them," Smithies says.

The Tools of Genetics: Informatics and Databases

For most of its history, biology managed to amass its data mostly with the help of plain old arithmetic. Gregor Mendel took the first steps in modern genetics simply by counting the different kinds of offspring produced by his peas. By contrast, today's genetics research is creating a flood of data, and new technologies are needed to manage it.

Gene-sequencing machines can read hundreds of thousands of nucleotides a day. The information in GenBank (a widely used database for DNA sequences) nearly doubles every year. It is said that a genetics laboratory can generate 100 gigabytes of data a day, every day—about 20,000 times the volume of data in the complete works of Shakespeare or J.S. Bach.

Image on computer screen courtesy of Tom Slezak, Lawrence Livermore National Laboratory

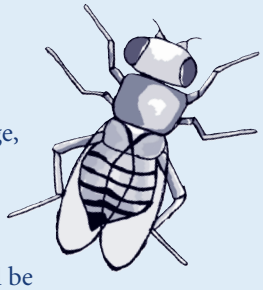
How to make sense of it all? The only way is with computers and software that can store the data and permit researchers to organize, search, and analyze it. In fact, many of today's challenges in biology, from gene analysis to drug discovery, are really challenges in information technology. This is not so odd when you remember that DNA is itself a kind of information technology, and that an organism's genes are an instruction manual, written in a shorthand we call the genetic code.

The result is a new biological specialty known as bioinformatics. "Informatics" just means information science, the field of study that develops hardware and software to handle enormous amounts of data.

FlyBase

By the late 1980s, the accumulating data collected on the fruit fly *Drosophila melanogaster* was so enormous—and so central to biology—that researchers decided they needed an electronic library for storing it. The project called on the talents of several participating groups of *Drosophila* researchers so that it could benefit from various points of view.

The library, soon known as FlyBase (http://flybase.org), was one of the earliest biological databases, and it remains one of the most important. It is a huge, comprehensive, international electronic database for information on the genetics and

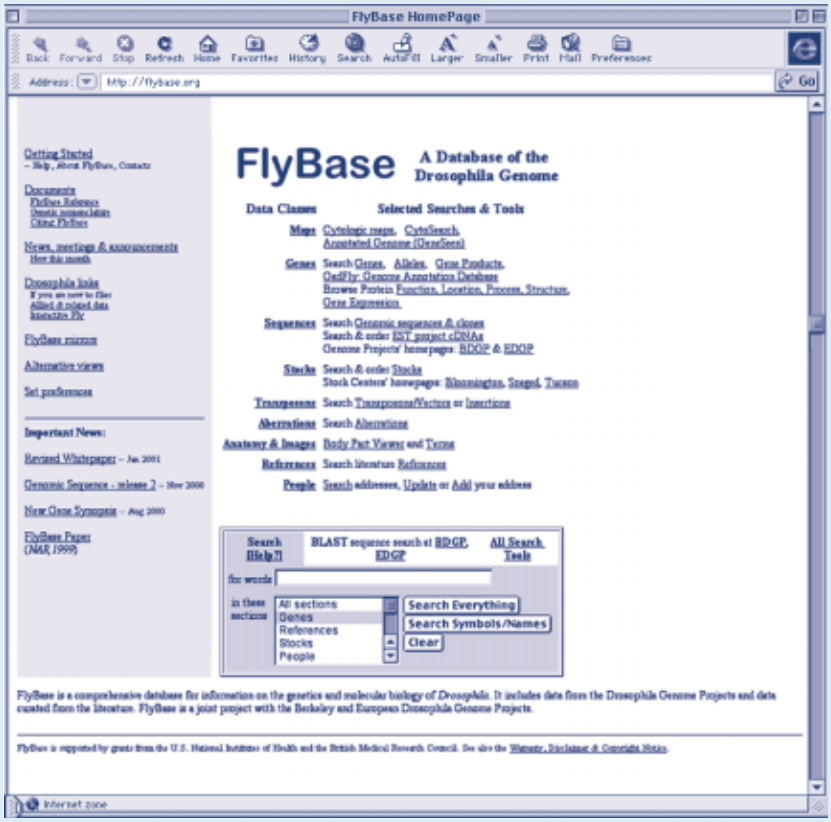


molecular biology of *Drosophila*, run by scientists for scientists. The information is amassed from nearly a century's worth of published scientific literature on fruit flies, and also from a recently completed project to map and sequence the fruit fly genome.

Two main groups of scientists use FlyBase. One, of course, is *Drosophila* researchers themselves. They typically query FlyBase to find out if a gene they have just encountered has been previously discovered by other scientists and to unearth clues to where it lies in the genome and what it does. In addition, FlyBase gives fly researchers access to their basic research material: flies. A researcher designing a genetic experiment can use the database to identify suitable strains of flies and then place an order from stock centers.

But FlyBase is also useful to scientists who study other organisms, like mice or humans. A researcher may run across a new mammalian gene and consult FlyBase to see if fruit flies possess a similar gene and if the database contains hints about what the gene does. "This approach works a lot of the time because the functions of these genes are highly conserved during evolution [they are similar among different organisms]," says William Gelbart of Harvard University in Cambridge, Massachusetts, a member of the FlyBase Consortium.

Ultimately, Gelbart hopes, scientists will be able to devise "power queries" that can reach out simultaneously to many different data sites on the Web, drawing together information from all of them. "The problem is, how do you provide tools that allow researchers to answer a question when they are not quite sure where to look for the answer and maybe not even quite sure how to structure that question? That is a very great challenge."



Human Variation and Disease

Kenneth K. Kidd of Yale University tracks human genetic variation. Many genes come in slightly different forms, variations that are called polymorphisms. Most polymorphisms do not significantly affect the way a gene works, but sometimes, a gene variation impairs a bodily function, and the result is a disease.

Scientists have learned a lot about normal variation from studying abnormal variation. They can examine families to track how individual pieces of DNA are transmitted from parent to child.

"This allows us to compare whether you have the same copies of a particular gene as some of your relatives do, and correlate that with whether or not you or your relatives have the same disease or disorder or particular inherited trait," Kidd explains. Researchers can find out, for example, whether you have exactly the same form of the insulin gene that your first cousin has. Both your gene and your cousin's could have been inherited from your mother's mother. Your grandmother passed the gene to your mother, who passed it to you. And your grandmother transmitted it also to your mother's brother, and from him to your cousin.

Scientists can now apply some of those ideas to whole populations as well as to families. According to Kidd, if you have the same form of a normal variation as another person in that population, you may share other genes as well.

Scientists have used that statistical notion to conduct a special kind of research called an association study. An association study compares a group of people with a particular disorder with a group of people who don't have the disorder, looking for genetic differences between them that might be related to the disorder. Normal genetic variation is a tool for their investigation. If the people with the disorder, on average, have a significantly different frequency of a polymorphism of a particular gene than people without the disorder, then perhaps the gene may be involved in the disease.

But there is a problem with that research approach. It assumes that both groups of people come from the same "gene pool," or ancestral population. "But normal variation varies among

human populations. Everyone has known this for millennia," Kidd points out, adding that scientists who use this association type of study first need to know what the normal level of variation is. In other words, two groups of people—a group with a disease and a group that doesn't have the disease—may indeed have different forms of a gene. But it's possible that the gene has nothing to do with the disease. The gene could be different in the two groups just because their ancestors are different.

Genetic diversity of the sort that can scuttle an association study is particularly common in the United States, where groups of research subjects have inherited their genes from many different populations around the world. So Kidd and his colleagues are putting a lot of effort into finding out what the range of human genetic variation actually is.

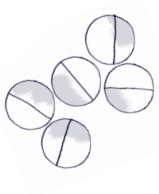
The Tools of Genetics: Genetic Testing

In 1999, *New York Times* health columnist Jane Brody wrote:

"The ability to test for genetic abnormalities that greatly increase a woman's risk of developing breast or ovarian cancer has created new and potentially lifesaving options. But it has also raised a host of new concerns for women with a family history of these cancers, including medical insurance, employment discrimination, emotional distress and strains on personal or family relations." (*New York Times*, August 17, 1999)

Brody identified one of the knottiest dilemmas to emerge from the new era of genetic information. How people should make use of information about their own genes is far from clear. For one thing, two genes associated with inherited forms of breast cancer, BRCA1 and BRCA2, are responsible for at most 1 in 10 cases of the disease. Since 9 out of 10 breast cancers are not inherited, genetic testing is irrelevant for the vast majority of women.

But even for those women with a strong family history of breast and ovarian cancer, deciding to have the test is not a simple choice. Genetic counseling can be helpful in thinking through a decision to be tested. But there is no one-size-fits-all choice about testing, whether for cancer genes or for any other genes that increase a person's risk of developing a disease.



Medicines and Your Genes

Human genetic variation not only underlies most human disease, it has an effect on the body's responses to disease and injury and on its responses to medicines. A particular medicine can work better in some people than in others. One reason may be that, because their genes differ, their bodies handle the medicine differently. Doctors first realized this in the 1950s, when some patients had bad—sometimes fatal—reactions to an anesthetic used in surgery. Investigation revealed that those who reacted badly possessed a genetic variation in the enzyme that should have helped break down and dispose of the anesthetic. The variant enzyme caused them no trouble at all until they needed surgery. In the operating room, the anesthetic failed to work as expected. A normal human genetic polymorphism suddenly became a life-threatening genetic abnormality.

One of the most exciting outcomes of genome sequencing research will be the ability to use genetic information to predict how individual people will respond to medicines. This field of research is known as pharmacogenetics or pharmacogenomics.

Sometimes, the body's reaction to a medicine is governed by a single gene. Most of the time, however, the body handles a medicine in an intricate series of steps governed by many different genes, as well as by factors like a person's diet and environment. Eventually, scientists expect that doctors will be able to decide which medicine to use and even how high the dose should be on the basis of a patient's inherited ability to respond. This will mean far more precise, and effective, therapies. It will also mean fewer side effects and adverse reactions.

Pharmacogenetics is just one example of how genetics research is fueling great advances in the fight against disease. As scientists probe even deeper into the mysteries of how genes work, they will continue to provide knowledge that gives us more power over disease and greater control of our health.

Help Wanted

Opportunities to be part of genetics research have never been greater—or more exciting. In addition to their studies of the human genome, scientists are gathering information about the genes of many other living things, from microbes that cause disease to model organisms like mice and *Drosophila*, livestock, and crop plants. This information resides in immense databases, but it all needs analysis by thousands and thousands of human brains. In addition to identifying genes, scientists must figure out what the genes do and—even more complicated—how they do it.

So the "Help Wanted" signs are up all over the world. What kind of help is needed? Three big categories are laboratory scientists, doctors to do research and treat patients, and genetic counselors to aid people in understanding information about their genes. The door is also wide open for people with expertise in mathematics, engineering, computer science, and physics. One exploding area that is desperately short of qualified workers is bioinformatics, the field of biology that develops hardware and software to store and analyze the huge amounts of data being generated by life scientists.

Many careers in genetics require advanced degrees such as the Ph.D. or M.D. But people with master's or bachelor's degrees are also needed to fill thousands of challenging jobs. For more information on genetics careers, see the Web sites at http://ns1.faseb.org/genetics/gsa/careers/bro-menu.htm and http://www.ornl.gov/hgmis/education/careers.html.

Laboratory Scientist
Bioinformatics Consultant
Research Associate
Medical Geneticist
Genetic Counselor
Scientist-Administrator
Lab Technician

Got It?

What is biotechnology, and what are some of its uses?

Why was PCR a major breakthrough for scientists?

What strategies do scientists use to study complex disorders?

How can genetic studies help doctors prescribe medicines?

Additional Resources

Books

Gonick L, Wheelis M.
The Cartoon Guide to Genetics.
HarperPerennial Library, 1991.

Henig RM.
The Monk in the Garden: The Lost and Found Genius of Gregor Mendel, the Father of Genetics.
Houghton-Mifflin Co., 2000.

McInerney J.
Basic Genetics.
Kendall/Hunt Publishing Co., 1998.

Ridley M.
Genome: The Autobiography of a Species in 23 Chapters.
HarperCollins, 2000.

Web Sites

General Sites

A Hypermedia Glossary of Genetic Terms
http://www.edv.agrar.tu-muenchen.de/~schlind/genglos.html
Contains hundreds of terms and definitions.

DNA Learning Center
http://vector.cshl.org
The site, from the Cold Spring Harbor Laboratory on Long Island, presents a primer on genetics by tracing its historical development. The site also contains other information on genetics and resources for students and teachers.

Genetic Science Learning Center
http://gslc.genetics.utah.edu
Basic information on genetics for general audiences.

Glossary of Genetic Terms
http://www.nhgri.nih.gov/DIR/VIP/Glossary/pub_glossary.cgi
This "talking glossary" contains simple definitions of genetics terms, plus brief audio clips from researchers discussing many of the terms in their own words.

Human Genome Project Information
http://www.ornl.gov/hgmis
http://www.nhgri.nih.gov/HGP
Overviews of the Human Genome Project, ethical issues in genetics, and educational resources.

Sites on Special Topics

Homeobox Genes and Birth Defects
http://www.sfn.org/briefings/hox_genes.html

Inside the Cell
http://www.nigms.nih.gov/news/science_ed/life.html
A brief, simple description of "the fundamental unit of life," the structure where genes do their work.

MIT Biology Hypertextbook, Mendelian Genetics Chapter
http://esg-www.mit.edu:8001/esgbio/mg/mgdir.html
An introduction to Mendelian genetics.

MIT Biology Hypertextbook, Recombinant DNA Chapter
http://esg-www.mit.edu:8001/esgbio/rdna/rdnadir.html
A primer on recombinant DNA.

Mitochondrial Resources on the Web
http://www.gen.emory.edu/MITOMAP/mtDNAsites.html
Basic information and links about mitochondria and mitochondrial genetics.

National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov
A bioinformatics resource.

The WWW Virtual Library: Model Organisms
http://ceolas.org/VL/mo
Contains descriptions of most model organisms now in research use and links to sites specializing in each one. Included are the yeast, roundworm, fruit fly, zebrafish, and mouse.

The WWW Virtual Library of Cell Biology: Apoptosis (Programmed Cell Death)
http://vlib.org/Science/Cell_Biology/apoptosis.shtml
Information on apoptosis and links to specialized apoptosis sites.

Why Do Basic Research?
http://www.nigms.nih.gov/news/science_ed/whydo.html
The basics on basic research.

Glossary

Amino acids | The building blocks of proteins. There are 20 amino acids, each of which is coded for by three adjacent nucleotides in a DNA sequence.

Anticipation | The disease process in which symptoms show up earlier and are increasingly severe in each generation.

Apoptosis | Programmed cell death, a normal process in which cells perish in an orderly, highly controlled manner so as to sculpt and control an organism's development.

Base | Part of a nucleotide (a building block of DNA and RNA). In DNA, the bases are adenine (abbreviated A), thymine (T), cytosine (C), and guanine (G). RNA contains uracil (U) instead of thymine.

Bioinformatics | The field of biology specializing in developing hardware and software to store and analyze the huge amounts of data being generated by life scientists.

Biotechnology | The industrial use of living organisms or biological methods derived through basic research; examples range from genetic engineering to making cheese or bread.

Caenorhabditis elegans | Also called *C. elegans*. A tiny roundworm often used as a model organism for genetics research.

Cell | The basic subunit of any living organism; the simplest unit that can exist as an independent living system.

Chromosome | A cellular structure containing genes. Chromosomes are composed of DNA and proteins. Humans have 23 pairs of chromosomes in each body cell, one of each pair from the mother and the other from the father.

Clone | In genetics, the process of making many copies of a gene. The term also refers to the identification of a gene.

DNA | Abbreviation for deoxyribonucleic acid, the molecule that contains the genetic code for all life forms except for a few viruses. It consists of two long, twisted chains made up of nucleotides. Each nucleotide contains one base, one phosphate molecule, and the sugar molecule deoxyribose. The bases in DNA nucleotides are adenine, thymine, guanine, and cytosine.

DNA chip | See *microarray*.

Drosophila melanogaster | A fruit fly that is often used as a model organism for genetics research.

Enzyme | A substance (usually a protein) that speeds up, or catalyzes, a chemical reaction without being permanently altered or consumed.

Eukaryote | An organism whose cells have a membrane-bound nucleus.

Exon | A DNA sequence in a gene that codes for a gene product.

Gene | A segment of the DNA molecule that contains information for making a protein or, sometimes, an RNA molecule.

Gene chip | See *microarray*.

Gene expression | The process by which the instructions in genes are converted to messenger RNA, which directs protein synthesis.

Gene mapping | Determining the locations of genes relative to one another on the chromosomes.

Genetic code | The language in which DNA's instructions are written. It consists of triplets of nucleotides, with each triplet corresponding to one amino acid in a protein or to a signal to start or stop protein production.

Genetics | The scientific study of genes and heredity—of how particular qualities or traits are transmitted from parents to offspring.

Genome | All of an organism's genetic material.

Genomics | A "scaled-up" version of genetics research in which scientists can look at all of the genes in a living creature at the same time.

Imprinting | The phenomenon in which a gene may be expressed differently in an offspring depending on whether it was inherited from the father or the mother.

Intron | A DNA sequence, or the RNA sequence transcribed from it, that interrupts the sequences coding for a gene product (exons).

Knockout | A cell or model organism in which one or more genes have been "knocked out," or inactivated, in order to study what the gene does. Model organism knockouts, especially mice, are useful for studying human disease.

Meiosis | The type of cell division that makes egg and sperm cells.

Microarray | Sometimes called a gene chip or a DNA chip. Microarrays consist of large numbers of molecules (often, but not always, DNA) distributed in rows in a very small space. Microarrays permit scientists to study gene expression by providing a snapshot of all the genes that are active in a cell at a particular time.

Mitochondrion | The cell's power plant, supplying the energy to carry out all of the cell's jobs. Each cell contains up to a thousand mitochondria. The structures are descended from free-living bacteria, and so they contain their own small genomes, called mitochondrial DNA.

Mitochondrial DNA | DNA found in mitochondria. Abbreviated mtDNA. Some human diseases have been traced to defects in mtDNA.

Mitosis | The type of cell division that makes new body cells.

Mobile genetic elements | See *transposons*.

Mutation | A change in a DNA sequence.

Nucleotide | A building block of DNA or RNA. It includes one base, one phosphate molecule, and one sugar molecule (deoxyribose in DNA, ribose in RNA).

Nucleus | The structure in the eukaryotic cell containing the genetic material.

PCR | The polymerase chain reaction, a quick and easy method for generating unlimited copies of any fragment of DNA.

Polymerase chain reaction | See *PCR*.

Polymorphism | A variant form of a gene. Most polymorphisms are harmless and are part of normal human genetic variation.

Protein | A molecule or complex of molecules consisting of subunits called amino acids. Proteins are the cell's main building materials, and they do most of a cell's work.

Recombinant DNA | Hybrid DNA produced in the laboratory by joining pieces of DNA from different sources.

Recombination | Part of the process of cell division, in which chromosomes pair up and exchange genetic material. The result is different combinations of genes in the offspring.

Replication | The process by which DNA copies itself in order to make a new genome to pass on to a daughter cell.

Ribosome | The cell structure in which proteins are manufactured. Most cells contain thousands of ribosomes.

RNA | Abbreviation for ribonucleic acid, the molecule that carries out DNA's instructions for making proteins. It consists of one long chain made up of nucleotides. Each nucleotide contains one base, one phosphate molecule, and the sugar molecule ribose. The bases in RNA nucleotides are adenine, uracil, guanine, and cytosine. There are three main types of RNA: messenger RNA, transfer RNA, and ribosomal RNA.

Sequencing | Sometimes called DNA sequencing or gene sequencing. Discovering the exact order of the building blocks (see *nucleotide*) of a particular piece of DNA or an entire genome.

Spliceosome | The cell machinery that splices exons together so that they can make proteins.

Stem cells | In embryos, the cells at the earliest stage of development. They have not yet begun to specialize and so they can grow into any kind of cell.

Transcription | The first major step in gene expression, in which the information coded in DNA is transcribed (copied) into a molecule of RNA.

Translation | The second major step in gene expression, in which the information encoded in RNA is deciphered (translated) into instructions for making a protein or for starting or stopping protein synthesis.

Transposons | "Jumping genes," genes that move around in the genome.

Triplet repeat | A kind of stutter in the DNA, a string of repeats of a particular sequence composed of just three nucleotides, CGG. Triplet repeats are responsible for several disorders of the nervous system.

Yeast | A single-celled, eukaryotic organism. Some forms of yeast, including the brewer's yeast *Saccharomyces cerevisiae*, are popular experimental organisms.

Discrimination Prohibited

Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the grounds of race, color, national origin, handicap, or age, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity (or, on the basis of sex, with respect to any education program or activity) receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts, and Executive Order 11246 states that no federally funded contractor may discriminate against any employee or applicant for employment because of race, color, religion, sex, or national origin. Therefore, the programs of the National Institute of General Medical Sciences must be operated in compliance with these laws and Executive Orders.

Accessibility

This publication can be made available in formats that are more accessible to people with disabilities. To request this material in a different format, contact the NIGMS Office of Communications and Public Liaison at 301-496-7301, TDD 301-402-6327, or write to the office at the following address:

45 Center Drive MSC 6200, Room 1AS.25,
Bethesda, MD 20892-6200.

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute of General Medical Sciences

NIH Publication No. 01-662
May 2001
www.nigms.nih.gov

Feedback Encouraged!

We hope you enjoyed reading *Genetic Basics.* We'd like to hear whether you found it interesting and understandable. In this booklet, we tried to explain genetics, convey the excitement of biomedical research, and provide a glimpse of what it's like to be a scientist. How well do you think we did this? We welcome these and any other comments you have about *Genetic Basics.*

To order hard copies of this free booklet, please call (301) 496-7301 or send e-mail to pub_info@nigms.nih.gov. You may order *Genetic Basics* and other free NIGMS publications online at www.nigms.nih.gov/news/publist.html.

Your Name (optional):

Your E-mail Address (optional):

Subject: Genetic Basics Comments/Suggestions

Your comments/suggestions about this publication:

Submit Your Comments Reset

** *If you have trouble submitting this form, send your comments to* dieffena@nigms.nih.gov