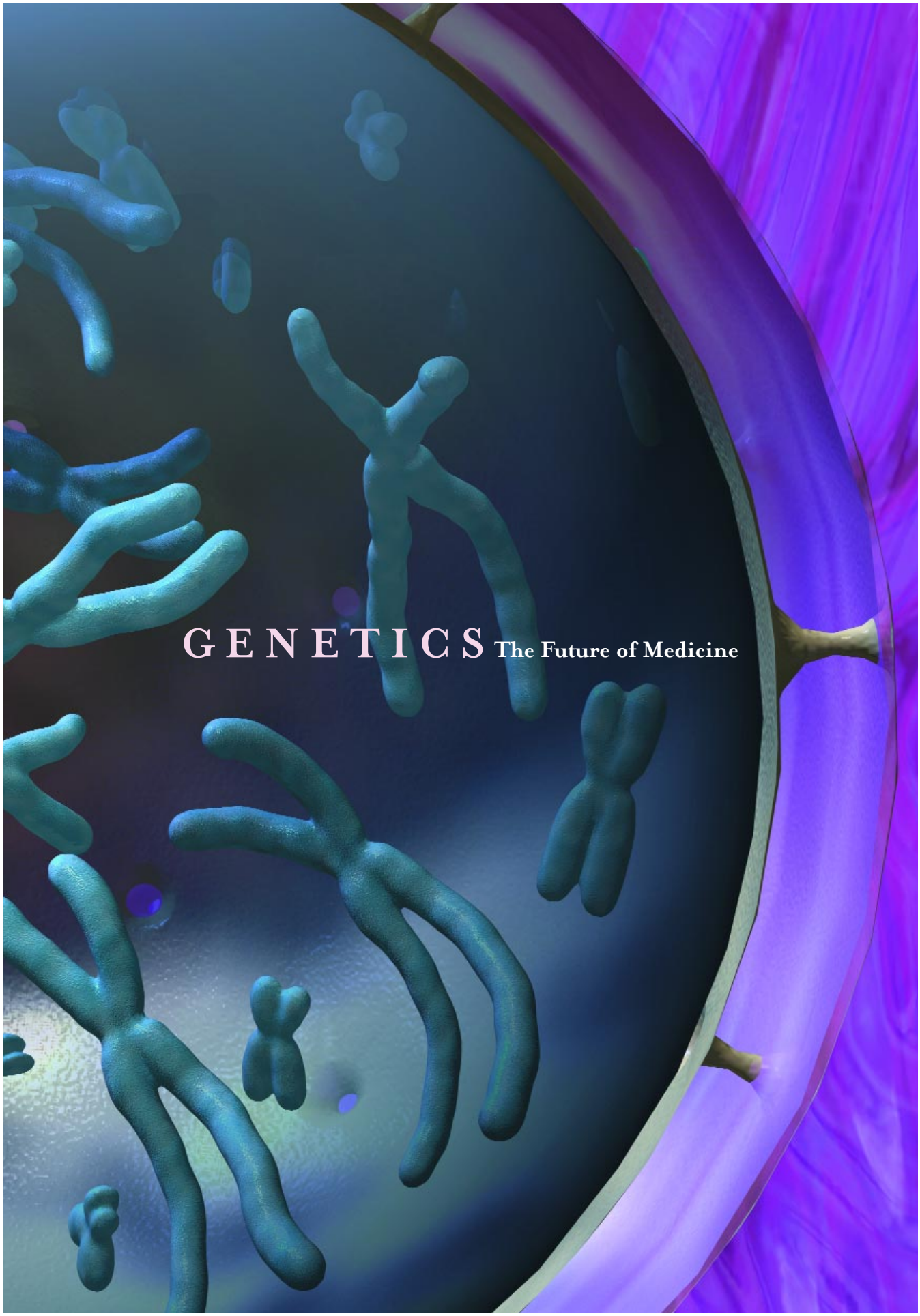

GENETICS The Future of Medicine

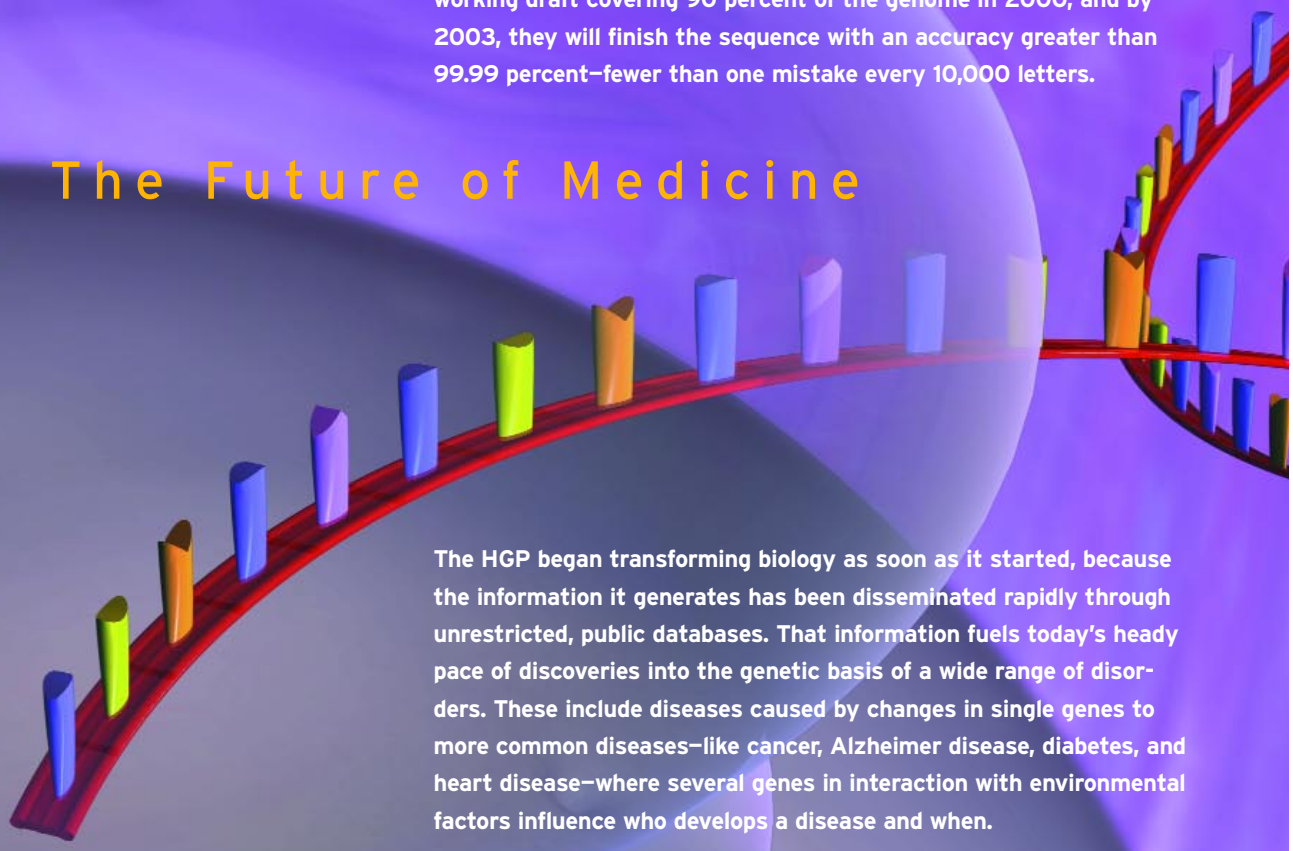
The Human Genome Project

Exploring Our Molecular Selves

DNA contains instructions for everything our cells do, from conception until death. Studying the human genome—all the DNA in our cells—allows us to explore fundamental details about ourselves. The Human Genome Project, the international quest to understand the genomes of humans and other organisms, will shed light on a wide range of basic questions, like how many genes we have, how cells work, how living things evolved, how single cells develop into complex creatures, and what exactly happens when we become ill. Besides answering innumerable questions about our molecular selves, a deeper understanding of the fundamental mechanisms of life promises to lead to an era of molecular medicine, with precise new ways to prevent, diagnose and treat disease.

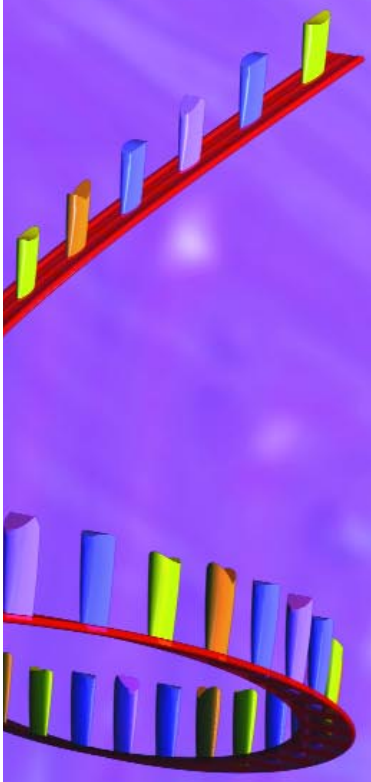
The Human Genome Project (HGP) began in the United States in 1990, when the National Institutes of Health and the Department of Energy joined forces with international partners to decipher the massive amount of information contained in our genomes. The HGP began with a set of ambitious goals but has exceeded nearly all of its targets. Frequently ahead of schedule, HGP scientists have produced an increasingly detailed series of maps that help geneticists navigate through human DNA. They have mapped and sequenced the genomes of important experimental organisms. They completed a working draft covering 90 percent of the genome in 2000, and by 2003, they will finish the sequence with an accuracy greater than 99.99 percent—fewer than one mistake every 10,000 letters.

The Future of Medicine



The HGP began transforming biology as soon as it started, because the information it generates has been disseminated rapidly through unrestricted, public databases. That information fuels today's heady pace of discoveries into the genetic basis of a wide range of disorders. These include diseases caused by changes in single genes to more common diseases—like cancer, Alzheimer disease, diabetes, and heart disease—where several genes in interaction with environmental factors influence who develops a disease and when.

Goals of the Human Genome Project



Map and sequence the human genome

- Build genetic and physical maps spanning the human genome.
- Determine the sequence of the estimated 3 billion letters of human DNA, to greater than 99.99 percent accuracy.
- Chart variations in DNA spelling among human beings.
- Map all the human genes.
- Begin to label the functions of genes and other parts of the genome.

Map and sequence the genomes of important model organisms

(The approximate number of letters, or base pairs, in each species' genome is given in parentheses.)

- The bacterium *Escherichia coli* (4.6 million)
- The yeast *Saccharomyces cerevisiae* (12 million)
- The roundworm *Caenorhabditis elegans* (97 million)
- The fruit fly *Drosophila melanogaster* (165 million)
- The mouse *Mus musculus* (3 billion)
- Other organisms (rat, zebrafish, chimpanzee, dog) will follow

Collect and distribute data

- Distribute genomic information and the tools for using it to the research community.
  - Release within 24 hours all sequence data that spans more than 2000 base pairs.
  - Create and run databases.
  - Develop software for large-scale DNA analysis.
  - Develop tools for comparing and interpreting genomic information.
- Share information with the wider public.

Study the ethical, legal, and social implications of genetic research

Train researchers

Develop technologies

- Make large-scale sequencing faster and cheaper.
- Develop technologies for finding sequence variations.
- Develop ways to study functions of genes on a genomic scale.

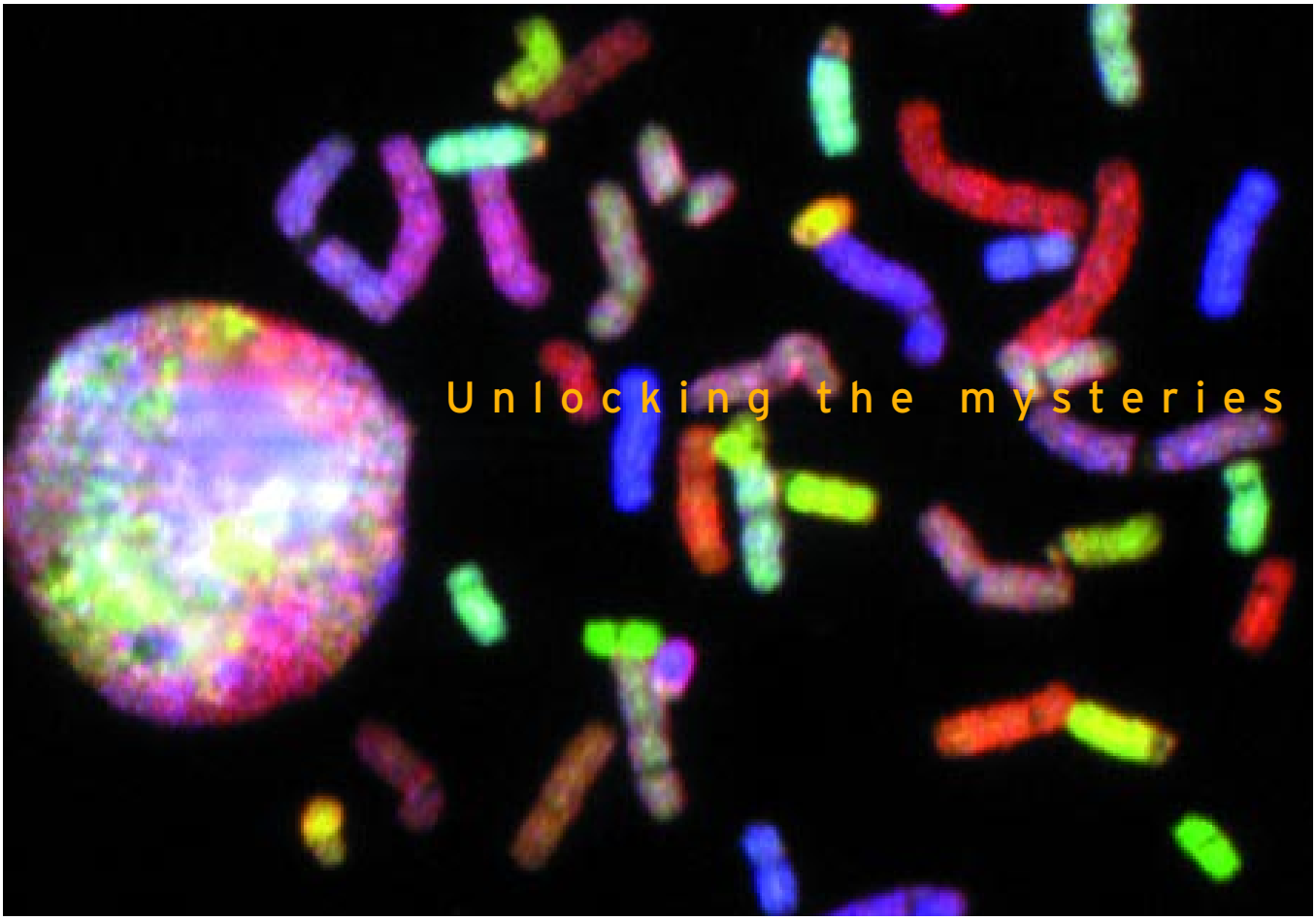
Transfer technologies to the private sector

Genes are made of DNA, a long, thread-like molecule. Almost all human cells contain 23 pairs of chromosomes; each chromosome contains a molecule of DNA with hundreds to thousands of genes arrayed along it. Genes usually code for proteins, the diverse molecules that perform a wide variety of specialized tasks. For example, proteins transmit messages between cells, fight infections, turn genes on or off, sense light and scents and flavors, and form structures, such as tendons and hair.

The instructions for making proteins are written with a four-letter alphabet—A, G, C, and T—where each letter represents one of the four chemical units strung together in DNA.

A single misspelling in the DNA sequence can make a protein malfunction, which, in turn, may cause disease.

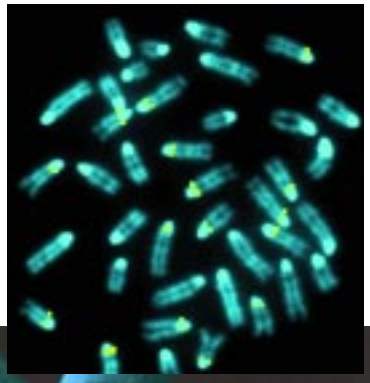
Alterations in our genes are responsible for an estimated 5000 clearly hereditary diseases, like Huntington disease, cystic fibrosis, and sickle cell anemia. The spellings of many other genes influence the development of common illnesses that arise through the interaction of genes with the environment.



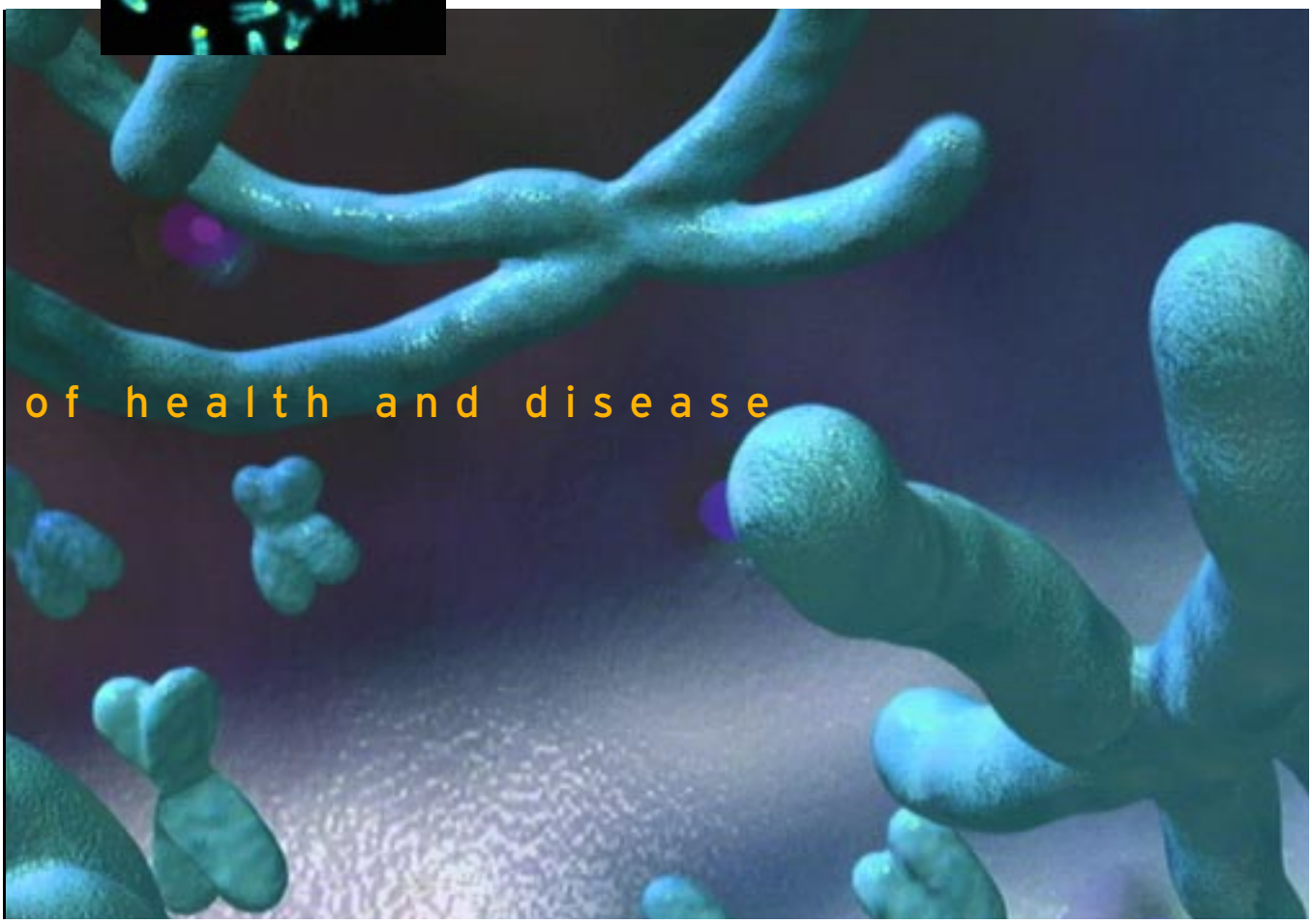
Gene Discovery

Connecting a gene with a disease was a slow, arduous, painstaking, and frequently imprecise process before the advent of the HGP. In 1989, geneticists had tracked down only four genes associated with disease by sorting through heredity. By 1998, the same list included more than 100 genes. Consider two gene hunts, eight years apart: in 1989, scientists found the gene for cystic fibrosis after a 9-year search; eight years later, a gene for Parkinson disease was mapped in only 9 days, and precisely described within 9 months.

With more and more DNA sequence deposited in electronic databases, researchers spend less time collecting data with their own experiments and more time analyzing the wealth of data available to them. They can electronically scan long stretches of DNA to find genes in the sequence that may be responsible for a particular disease. Those are called candidate genes. If a candidate gene actually does play a role in a disease, it should be spelled differently in people with the disease compared to those without it; the alteration in spelling somehow disrupts the normal function of the gene product. For example, rare cases of early-onset Parkinson disease can result from a change in just one DNA letter, which in turn, changes one of the 140 amino acids that make up a key protein.



The gigabytes of DNA sequence data flowing from the HGP and the progressively more detailed catalog of human sequence variations are helping scientists study increasingly complex genetic questions. Instead of restricting their studies to conditions caused by mutations in single genes, scientists can now study the genetic basis for complex diseases, like diabetes and Alzheimer disease, that involve several genes.

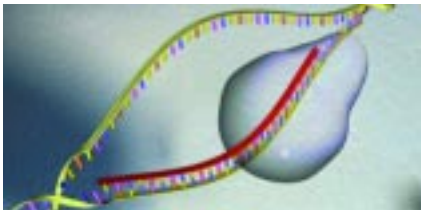


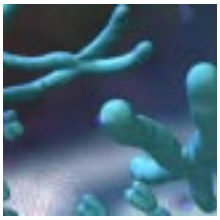
Basic Genetics

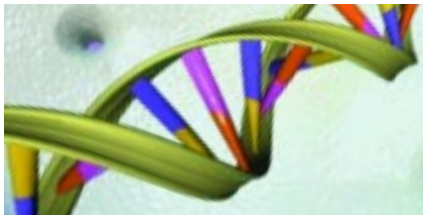
Human cell
Almost all of the 100 trillion cells in the human body contain a copy of the entire human genome, the complete set of genetic instructions necessary to build a human being.

Cell nucleus
The nucleus is a separate compartment in the cell that contains 6 feet of DNA packed into 23 pairs of chromosomes. We each inherit one set of 23 chromosomes from our mother, and another set from our father. Egg and sperm cells carry single sets of 23 chromosomes.

Gene
Each gene contains a segment of DNA, typically several thousand base pairs long, that is copied into a molecule of RNA. Usually, the information in RNA is translated to make a protein.







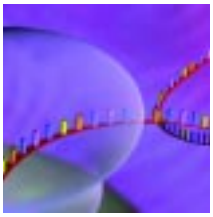
DNA
DNA, or deoxyribonucleic acid, is a long molecule made of two twisting, paired strands. Each strand is made of four chemical units, called nucleotide bases, strung together in a precise order, just as letters string together to make specific words. The bases are adenine (A), guanine (G), cytosine (C), and thymine (T). The bases on opposite strands pair specifically; an A always pairs with a T, and a C always with a G. Each such pair is called a base pair of DNA.

Chromosome
Each of the human chromosomes contains hundreds to thousands of genes, the major functional units of DNA.

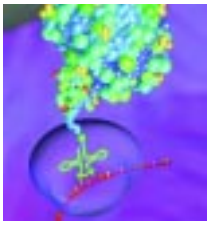
Basic Genetics

Basic G

RNA
RNA, or ribonucleic acid, is chemically similar to DNA, except it is single-stranded, not double-stranded; it contains the base uracil (U) instead of thymine (T); it can migrate out of the nucleus. The sequences of most RNA molecules are translated to make proteins.

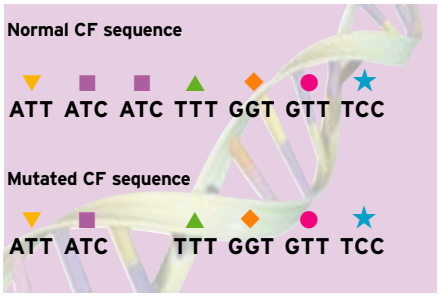


Protein
Proteins make up essential parts of tissues and guide chemical reactions in living things. They are made of 20 different building blocks called amino acids. The DNA sequence of a gene determines the amino acid sequence of the protein that gene encodes. The amino acid sequence of the protein is, in turn, responsible for the protein's shape and function.



Genome
A genome is all the DNA– the complete genetic inheritance–in an organism. The human genome is contained in 23 pairs of chromosomes housed in the nucleus and the small circle of DNA present in mitochondria, the organelles that process energy. The number of genes in the approximately 3 billion base pairs of human DNA is still not known, but is probably between 35,000 and 100,000.

SNP
Pronounced "snip," SNPs are single-nucleotide polymorphisms or one-letter variations in the DNA sequence. SNPs contribute to differences among individuals. The majority have no effect, others cause subtle differences in countless characteristics, like appearance, while some affect the risk for certain diseases.



Mutation
Mutations are changes in DNA spelling that can prevent proteins from functioning normally and cause health problems.

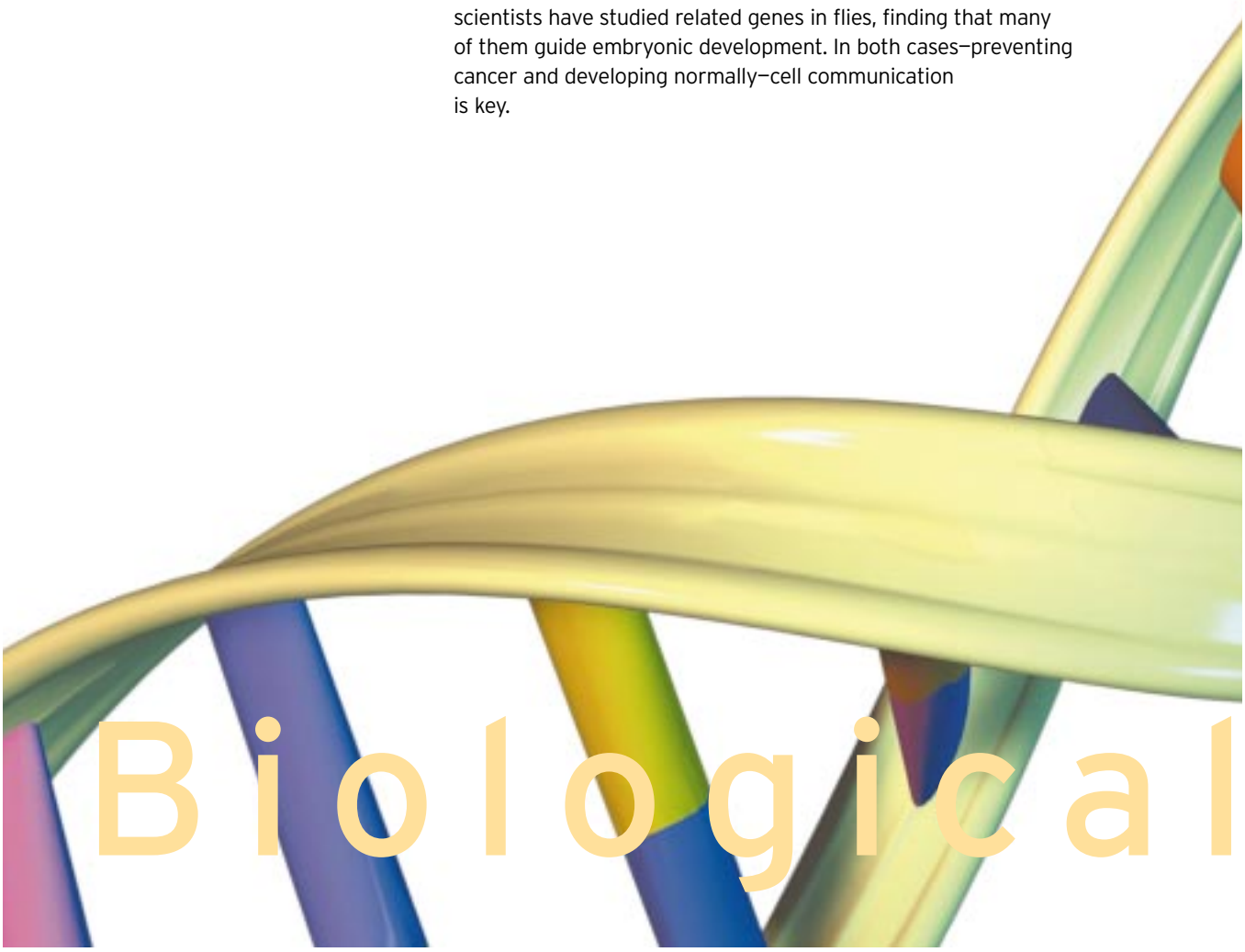
Three DNA bases are deleted in the mutated sequence, resulting in the deletion of an amino acid (phenylalanine) from the CF protein. People with the abnormal protein develop cystic fibrosis.

Understanding Biological Function

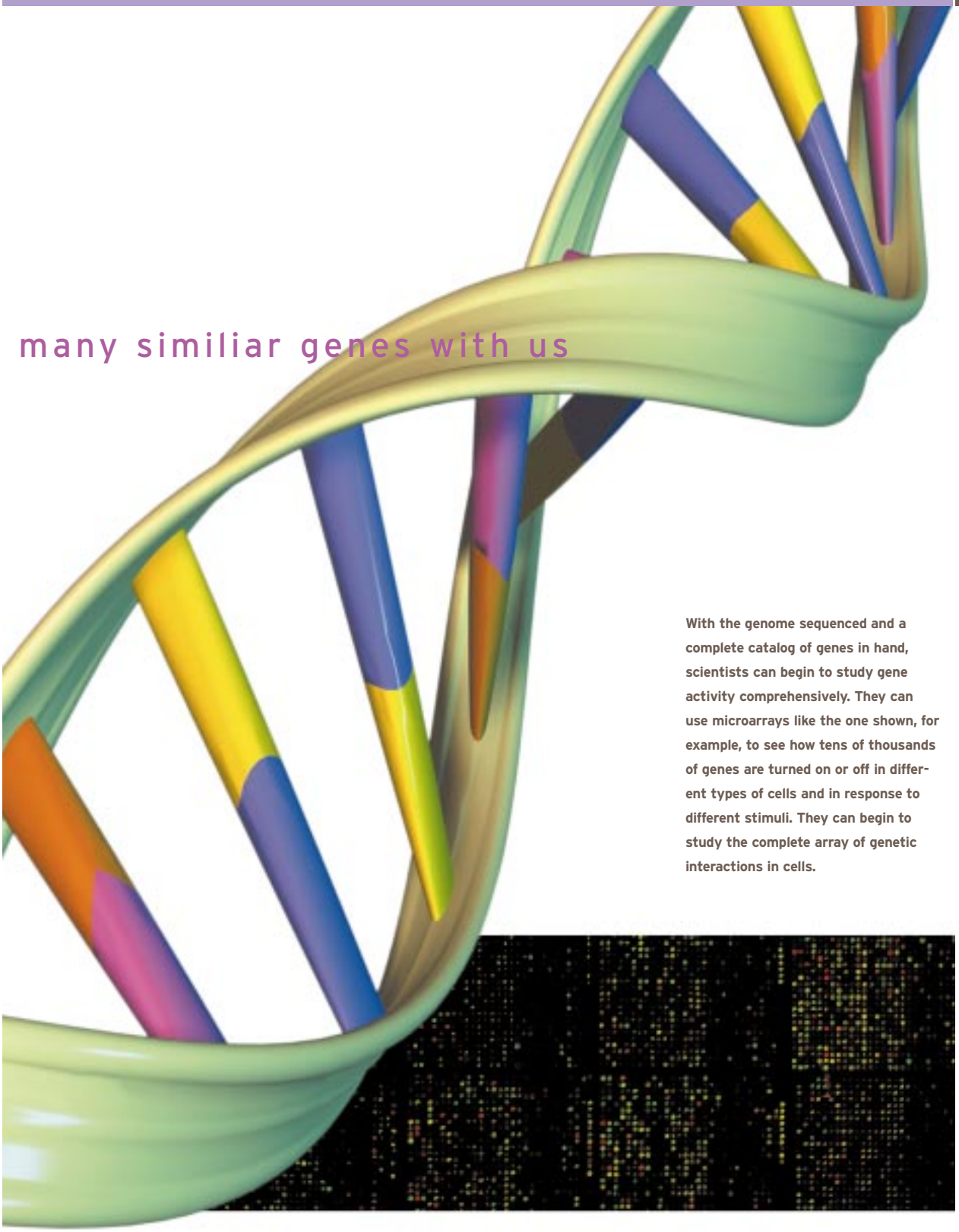
Knowing the DNA sequence of a gene reveals the basic structure of the protein that gene encodes. Scientists can sometimes deduce the 3-dimensional shape and function of the protein as well. Often, they can classify the protein because of similarities to other proteins. For instance, when scientists discovered the gene for cystic fibrosis, the sequence immediately suggested that the CF protein is a gatekeeper embedded in the membrane that surrounds a cell. The sequence also implied that the protein specifically allows salt to pass through the membrane. This fit nicely with the idea that a problem with the transport of salt and water might cause CF and explain why mucus tends to dry up in the lungs of people with the disease.

even roundworms and fruit flies share

Experimental animals play an important role in helping scientists understand the biological function of genes. Human genes have relatives in the genomes of other animals. Even species as seemingly different from us as yeast, roundworms, or fruit flies share many similar genes. In fact, comparing DNA from different species and finding stretches where the sequence is conserved can highlight particularly important features. Often, insights about human diseases come when a newly discovered human disease gene has a close relative in another species such as the mouse or even the fruit fly—species where the role of that gene can be studied and placed in context. For example, the role of some human genes in cancer is understood better than otherwise possible because scientists have studied related genes in flies, finding that many of them guide embryonic development. In both cases—preventing cancer and developing normally—cell communication is key.



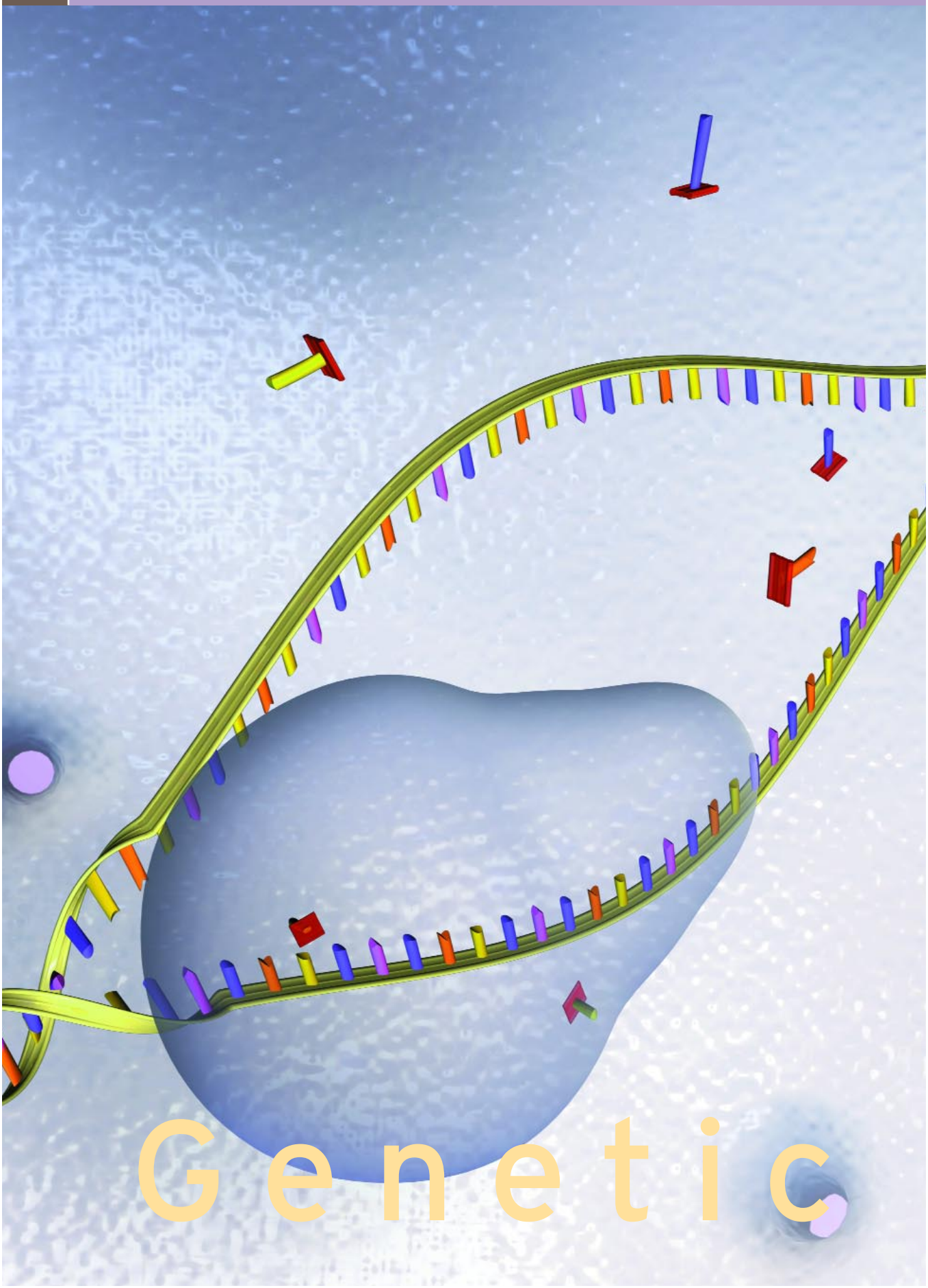
Biological



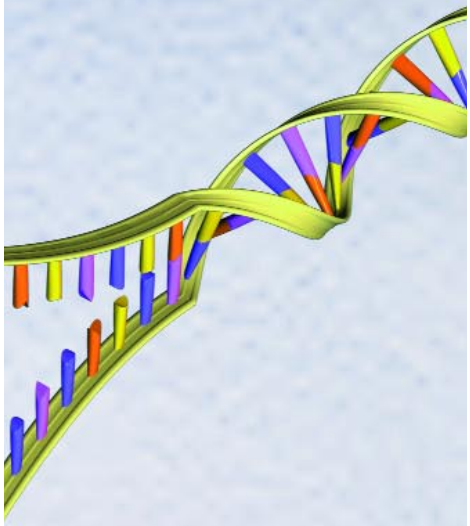
many similiar genes with us

With the genome sequenced and a complete catalog of genes in hand, scientists can begin to study gene activity comprehensively. They can use microarrays like the one shown, for example, to see how tens of thousands of genes are turned on or off in different types of cells and in response to different stimuli. They can begin to study the complete array of genetic interactions in cells.

Function


Genetic

Genetic Testing and Gene-based Medicine



Examining how a particular gene is spelled in an individual can serve quite a few uses:

Diagnosis Genetic analysis now can classify some conditions, like colon cancer and skin cancer, into finer categories. This is important since classifying diseases more precisely can suggest more appropriate treatments. The same approach will soon be possible for heart disease, schizophrenia, and many other medical conditions, as the genetic underpinnings for these diseases become more completely understood.

Pharmacogenomics is a new word that scientists and drug developers use. It describes the idea of tailoring drugs for patients, whose individual response can be predicted by genetic fingerprinting. For example, cancer patients facing chemotherapy may experience fewer side effects and improve their prognoses by first getting a genetic fingerprint of their tumor. This fingerprint can reveal which chemotherapy choices are most likely to be effective. Better understanding of genetics promises a future of precise, customized medical treatments.

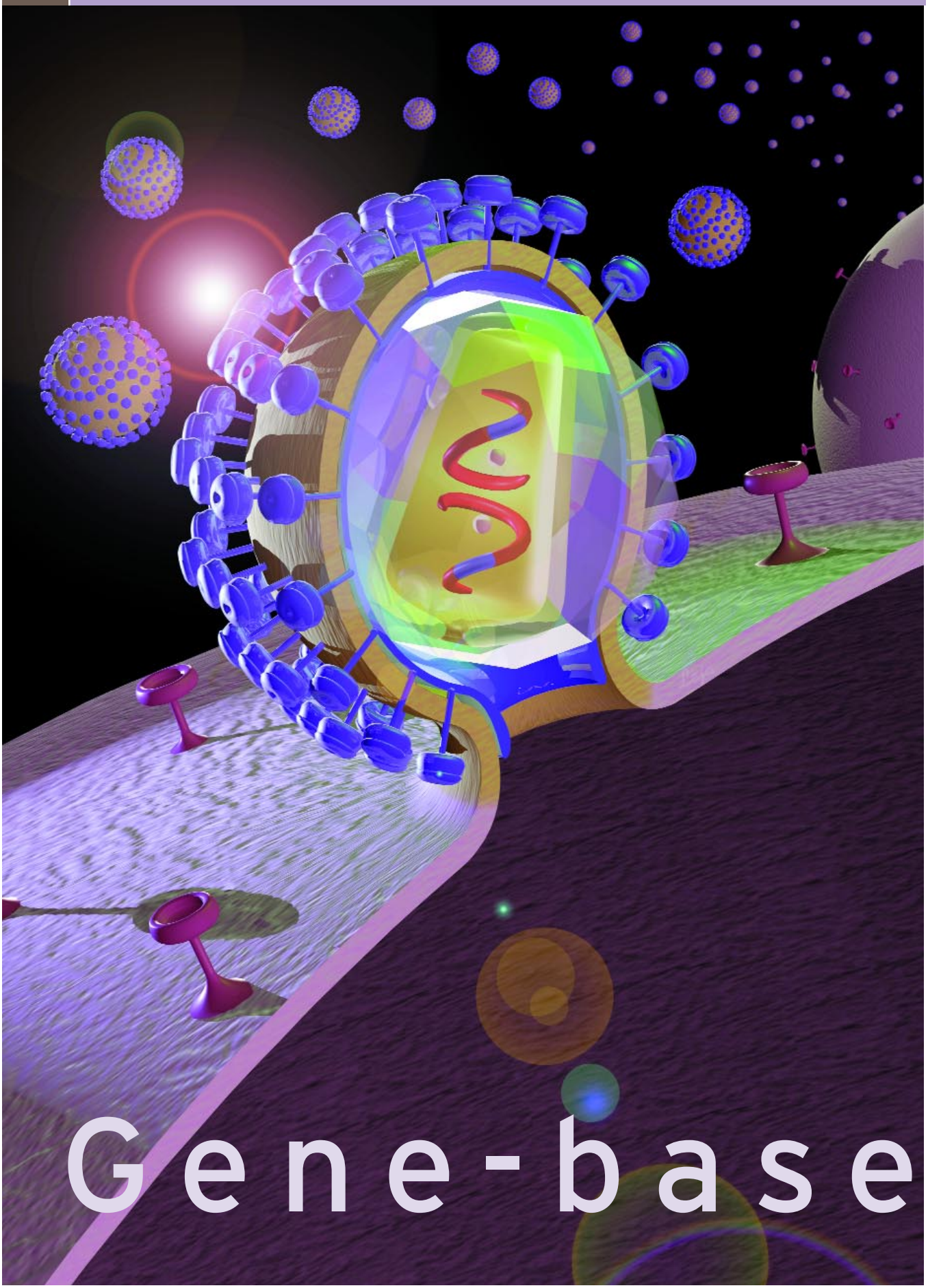
Prognosis Diagnosing ailments more precisely will lead to more reliable predictions about the course of a disease. For example, a genetic work-up can inform a patient with high cholesterol levels how damaging that condition is likely to be. And doctors treating prostate cancer will be able to predict how aggressive a tumor will be. For many diseases, such genetic information will help patients and doctors weigh the risks and benefits of different treatments.

Prevention Once scientists figure out what DNA sequence changes in a gene can cause disease, healthy people can be tested to see whether they risk developing conditions like heart disease, diabetes, or prostate cancer later in life. In many cases, this advance warning can be a cue to start a vigilant screening program, to take preventive medicines, or to make diet or lifestyle changes that might prevent the disease altogether.

For example, those at risk for colon cancer could undergo frequent colonoscopies; those with hereditary hemochromatosis, a common disorder of iron metabolism, could donate blood periodically to remove excess iron and prevent damage to the body. Some women at risk for breast cancer could benefit from tamoxifen; a young person at risk for developing lung cancer may become particularly motivated to quit smoking; those with familial hypercholesterolemia could begin treatment to lower their cholesterol levels and prevent heart attacks and strokes.

Unfortunately, our ability to predict a disease sometimes precedes our ability to prevent or treat it. For example, a genetic test has been available for Huntington disease for years, but no treatment is available yet. As a result, only a minority of people at risk have chosen to be tested.

Testing



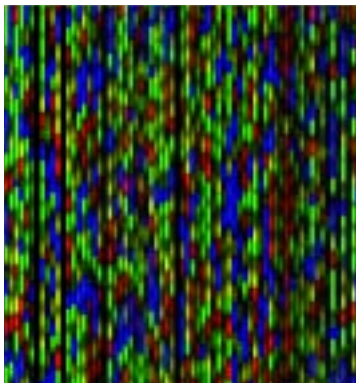
Gene-base

Newborn screening A particular form of predictive testing, newborn screening can sometimes help a great deal. For example, babies in the United States and a few other countries are routinely screened for phenylketonuria (PKU), a metabolic disorder that prevents the breakdown of phenylalanine, one of the building blocks of proteins and a component of the artificial sweetener Aspartame. Excess phenylalanine in the body is toxic to the nervous system. In the past, children with the condition became severely mentally retarded, but the screening program identifies children with the enzyme deficiency, allowing them to grow normally on a diet that strictly avoids phenylalanine.

Carrier screening For some genetic conditions, people who will never be ill themselves can pass a disease to their children. Some couples choose to be tested for this risk before they marry, especially in communities where a feared childhood disease is particularly common. For example, carrier testing for Tay-Sachs disease, which kills young children and is particularly common in some Jewish and Canadian populations, has been available and widely used for years.

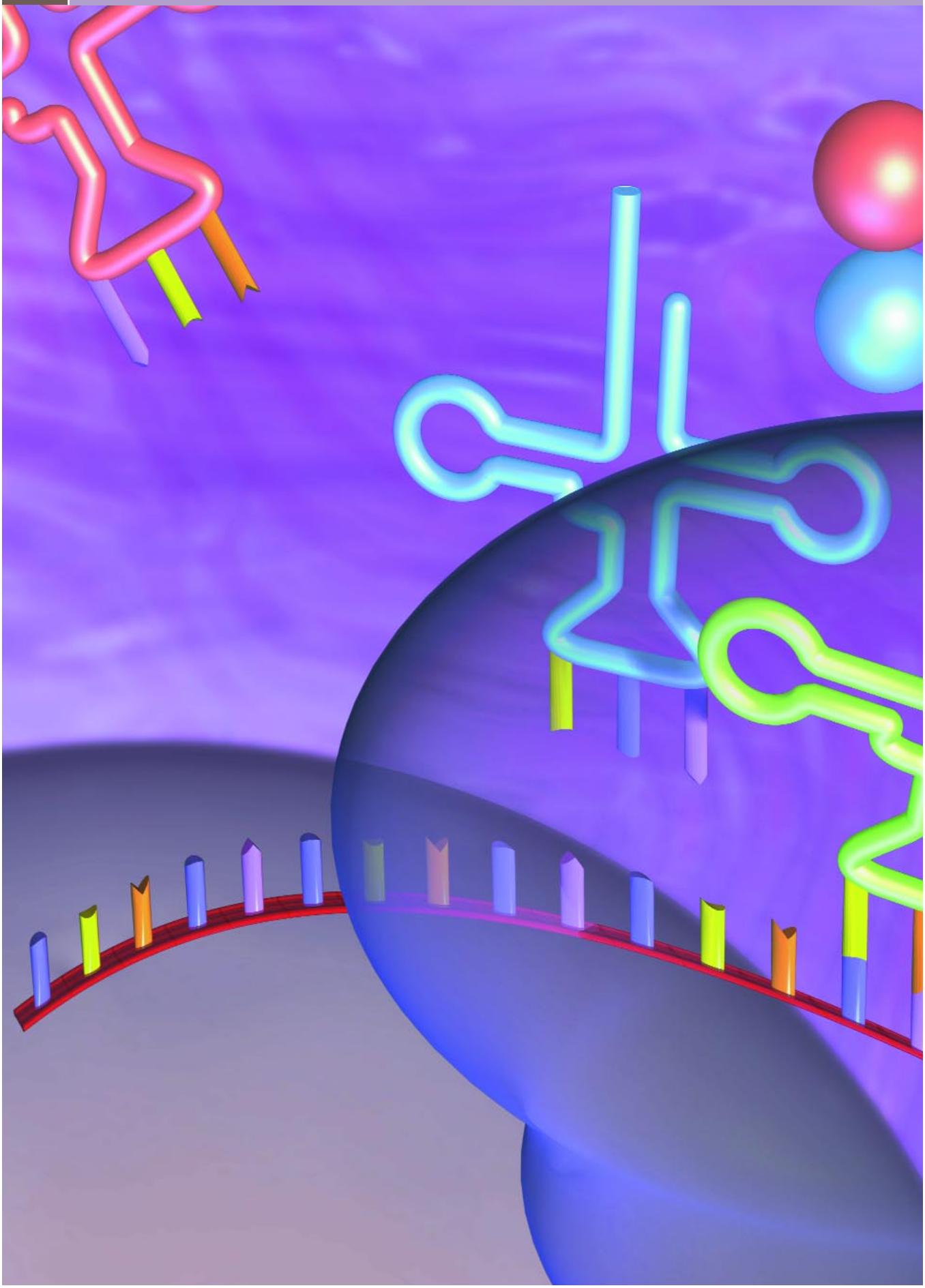
Gene therapy Replacing a misspelled gene with a functional gene has long been an appealing idea. Small groups of patients have undergone gene therapy in clinical trials for more than a decade, but this remains an experimental treatment. Eventually, it likely will become a common treatment for some conditions.

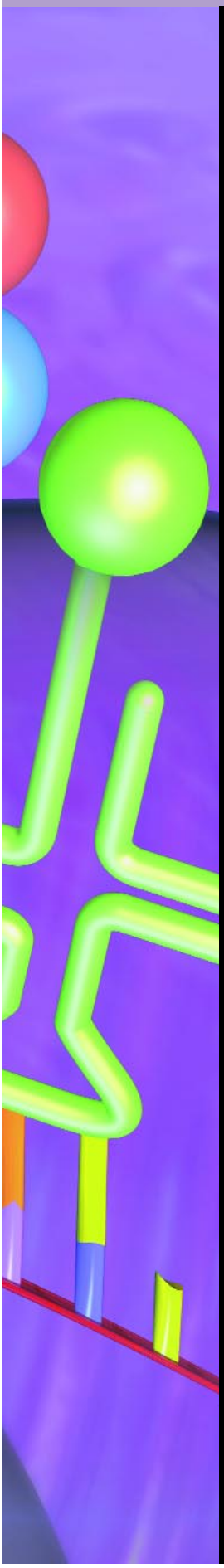
Gene-based therapy Great medical benefit likely will derive from drug design that's guided by an understanding of how genes work and what exactly happens at the molecular level to cause disease. For example, the causes of adult-onset diabetes and the resulting complications remain difficult to decipher and, so, to treat. But researchers are optimistic that a more precise understanding of the underlying causes will lead to better therapies. In many cases, instead of trying to replace a gene, it will be more effective and simpler to replace the protein the gene would give rise to. Alternatively, it may be possible to administer a small molecule that interacts with the protein—as many drugs do—and changes its behavior.



One of the first examples of such a rationally-designed drug targets the genetic flaw that causes chronic myelogenous leukemia, a form of leukemia that mostly affects adults. An unusual joining of chromosomes 9 and 22 produces an abnormal protein that spurs the uncontrolled growth of white blood cells. Scientists have designed a drug that specifically attaches to the abnormal protein and blocks its activity. In preliminary tests, blood counts returned to normal in all patients treated with the drug. And, compared with other forms of cancer treatment, the patients experienced very mild side effects.

Instead of having to rely on chance and screening thousands of molecules to find an effective drug, which is how most drugs we use today were found, scientists will begin the process of drug discovery with a clearer notion of what they're looking for. And because rationally designed drugs are more likely to act very specifically, they will be less likely to have damaging side effects.



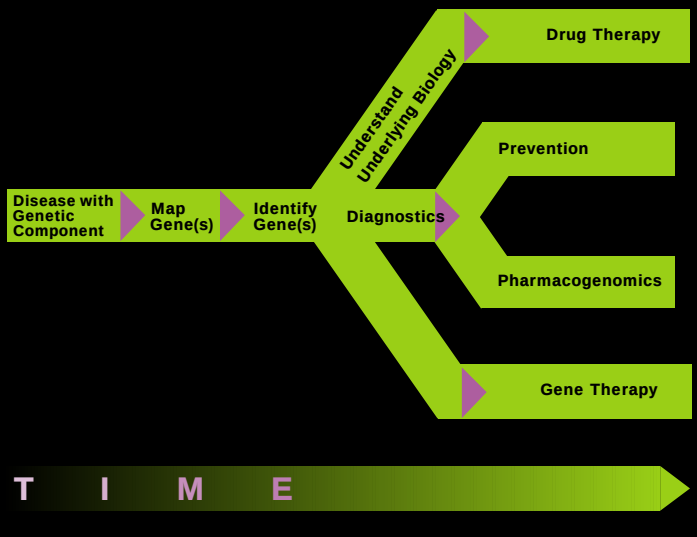


Molecular biology has long held out the promise of transforming medicine from a matter of serendipity to a rational pursuit grounded in a fundamental understanding of the mechanisms of life. Genomics will hasten the advance of molecular biology into the practice of medicine.

As the molecular foundations of diseases become clearer, we may be able to prevent them in many cases and in other cases, design accurate, individualized treatments for them. Genetic tests will routinely predict individual susceptibility to disease. Diagnoses of many conditions will be much more thorough and specific than now. New drugs, derived from a detailed molecular understanding of common illnesses like diabetes and high blood pressure, will target molecules logically. Drugs like those for cancer will routinely be matched to a patient's likely response. Decades from now, many potential diseases may be cured at the molecular level before they arise.

All these changes aren't likely to come quickly. It will take a long time to understand the human genome. But access to genome sequence will increasingly shape the practice of health care over the coming decades, as well as shed light on many of the mysteries of biology.

Development of Genetic Medicine



Written by Karin Jegalian

Produced by
National Human Genome Research Institute
National Institutes of Health
www.nhgri.nih.gov

With co-sponsorship by
Office of Biological and Environmental Research
U.S. Department of Energy
www.ornl.gov/hgmis

Howard Hughes Medical Institute
www.hhmi.org

Pharmaceutical Research and
Manufacturers of America
hhtp://genomics.phrma.org/

Nature
International Weekly Journal of Science
www.nature.com/genomics

Science
Global Weekly of Research
www.scienceonline.org

American Society of Human Genetics
www.faseb.org/genetics/ashg/ashgmenu.htm

NIH Publication No. 00-4873