From the Blueprint to You

NATIONAL HUMAN
GENOME RESEARCH INSTITUTE

www.genome.gov

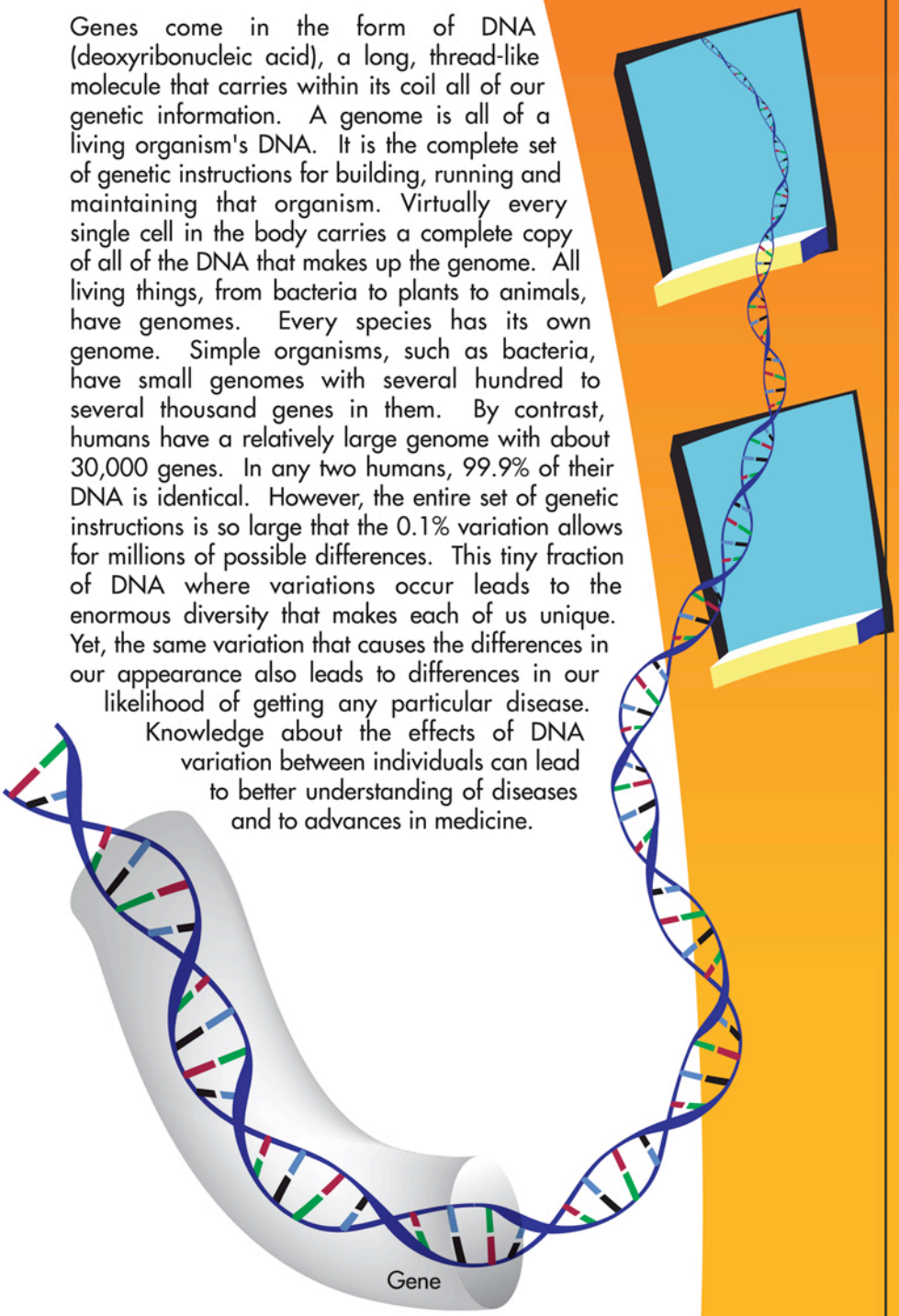
NIH Publication No: 03-5377
April 2003

A Brief Guide to Genetics.

What makes you different from everyone else?

All humans are basically the same; that is, we are all members of the same species. Yet we are each also unique, with different traits that allow us to stand out as individuals. Some people are short, others tall. There are a variety of eye, skin and hair colors. These physical similarities and differences are due to similarities and differences in our genetic instructions. Our own set of genetic instructions, our "genes," determines our particular traits, inherited from our parents.

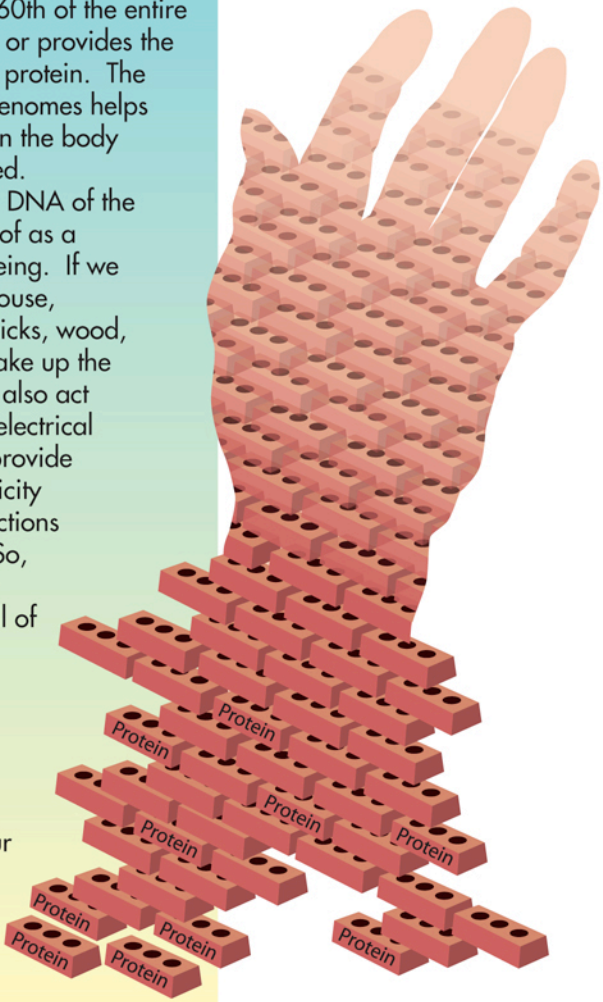
Genes come in the form of DNA (deoxyribonucleic acid), a long, thread-like molecule that carries within its coil all of our genetic information. A genome is all of a living organism's DNA. It is the complete set of genetic instructions for building, running and maintaining that organism. Virtually every single cell in the body carries a complete copy of all of the DNA that makes up the genome. All living things, from bacteria to plants to animals, have genomes. Every species has its own genome. Simple organisms, such as bacteria, have small genomes with several hundred to several thousand genes in them. By contrast, humans have a relatively large genome with about 30,000 genes. In any two humans, 99.9% of their DNA is identical. However, the entire set of genetic instructions is so large that the 0.1% variation allows for millions of possible differences. This tiny fraction of DNA where variations occur leads to the enormous diversity that makes each of us unique. Yet, the same variation that causes the differences in our appearance also leads to differences in our likelihood of getting any particular disease. Knowledge about the effects of DNA variation between individuals can lead to better understanding of diseases and to advances in medicine.

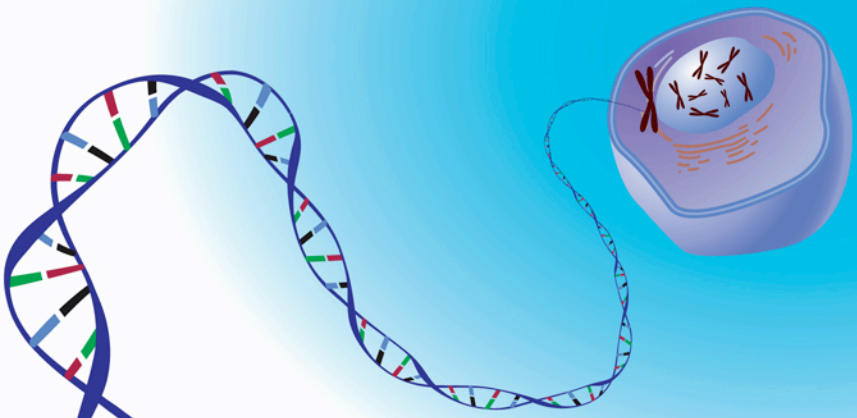


DNA as the Instruction Book.

How does DNA work as the instructions for human traits? The cells of our bodies are made of different kinds of molecules, such as water, minerals, proteins, sugars, fats and DNA. Of those, proteins are particularly important because they are the fundamental components of the body that determine how all of the molecules are organized and how they act. Thus, proteins play a key role in the way we look and in the way we grow. DNA acts as a molecular code for making these proteins. The DNA in each gene provides the instructions for making one protein, or sometimes, a few related proteins. However, only about 1/60th of the entire genome directly codes – or provides the instructions - for making protein. The rest of the DNA in our genomes helps direct when and where in the body each gene should be used.

Taken together, all of the DNA of the genome can be thought of as a blueprint for a human being. If we think of our body as a house, proteins would be the bricks, wood, cement and nails that make up the basic building. Proteins also act as the lights, plumbing, electrical wiring, vents, etc., that provide the running water, electricity and other necessary functions for living in the house. So, just as a house blueprint shows us the layout of all of the parts (the boards, bricks, wires, etc.), the genome is a set of instructions from which we can determine the layout of all the proteins used to build and run our body.





DNA is composed of an ordered series of four chemical structures called nucleotide "bases": adenine, thymine, cytosine and guanine, which are abbreviated A, T, C and G. These bases are lined up one after another along the length of a DNA strand. The sequence of these bases acts as a code that can be deciphered to reveal our genetic instructions. A gene is made up of a specific stretch of DNA, typically several thousand bases long. [For more information on "The Genetic Code," see Figure A.]

How does our body read the genetic instructions and use it to make a protein? DNA can't be converted into protein directly, but instead, sends a message describing the gene's instruction, to a protein-making machine. Each particular gene can be "transcribed," or copied, into a related molecule called mRNA (messenger ribonucleic acid) and is then transported to a molecular, protein-making machine called a ribosome. The job of the ribosome is to read the mRNA copy of the gene and assemble the appropriate protein. [For more information on "Making a Protein," see Figure B].

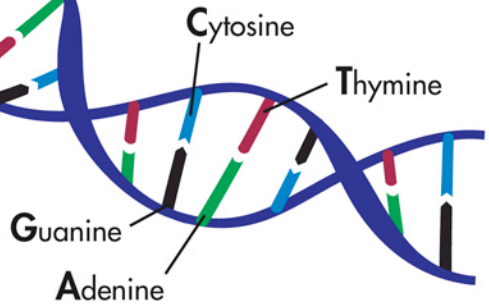


Figure A: The Genetic Code

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr stop stop	Cys Cys stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Genes are made of a series of four nucleotide bases (A, C, G, and T) that line up, one after another, as a nucleotide chain. For a gene to be used to make a protein, the DNA must first be copied into a message (a related molecule called mRNA), which can then be sent out of the cell nucleus to a ribosome, where proteins are assembled. This mRNA molecule is much like DNA, except that its chemical structure is formed using ribose instead of deoxyribose, and one of the bases, thymine (T), is substituted with uracil (U). The order, or sequence, of the nucleotides in mRNA determine how to assemble a protein. Amino acids are the building blocks of a protein. That is, individual amino acids are assembled one after another into what is called a "polypeptide" chain. This polypeptide is then folded and arranged into a protein. Each amino acid that is added to a growing polypeptide chain is indicated by a specific set of three nucleotide bases within the mRNA. Each possible "triplet" set of nucleotide bases is called a codon.

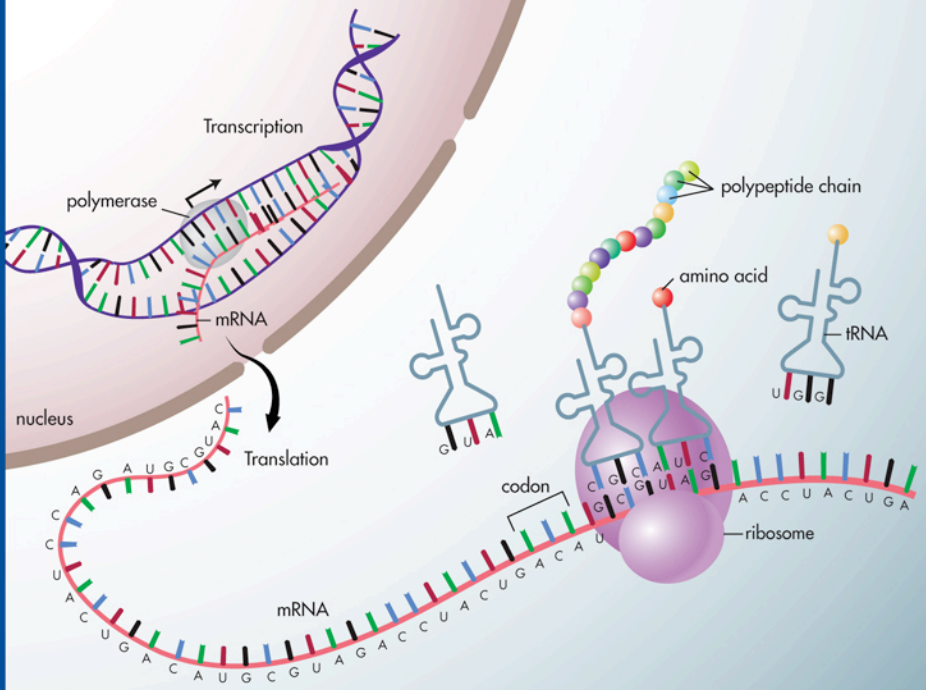
The particular amino acid specified by each codon sequence on mRNA is indicated in the above table. The first base of a codon is indicated on the left side of the table, the second base on top, and the third base is indicated on the right side.

Since RNA is constructed from four types of nucleotides, there are 64 possible triplet sequences or codons (4x4x4). Three of these possible codons specify the termination of the polypeptide chain. They are called "stop codons." That leaves 61 codons to specify only 20 different amino acids. Therefore, most of the amino acids are represented by more than one codon. The genetic code is thus said to be "degenerate."

The Key for the above table:
Ala: Alanine Cys: Cysteine Asp: Aspartic acid Glu: Glutamic acid
Phe: Phenylalanine Gly: Glycine His: Histidine Ile: Isoleucine
Lys: Lysine Leu: Leucine Met: Methionine Asn: Asparagine
Pro: Proline Gln: Glutamine Arg: Arginine Ser: Serine
Thr: Threonine Val: Valine Trp: Tryptophane Tyr: Tyrosisne

A = adenine G = guanine C = cytosine T = thymine U = uracil

Figure B: Making a Protein

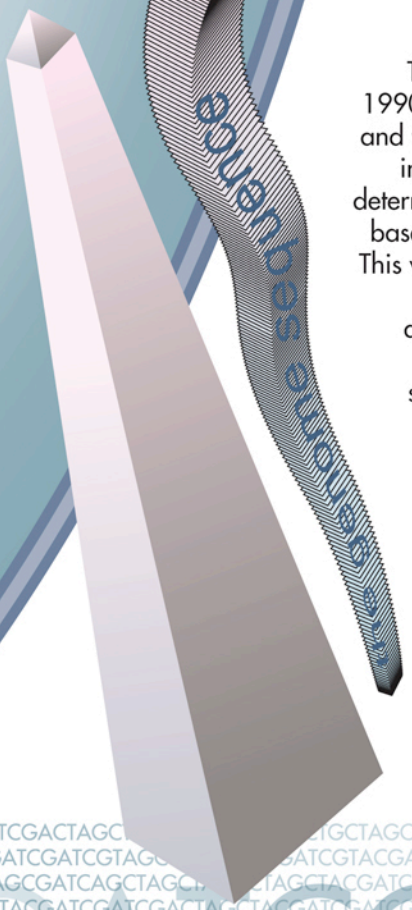


Each nucleotide base (A, C, T, G) along one strand of the double-stranded DNA chain has a complementary base across from it on the other strand. Adenine (A) always attaches with its complementary partner thymine (T). Cytosine (C) always attaches with guanine (G). When the information of a gene is used to make a protein it is first "transcribed" (copied) to a molecule of messenger RNA. The complementary DNA strands "unzip" to expose the coded gene, and a molecular machine known as polymerase makes a complementary strand of mRNA. The mRNA molecules then leave the cell nucleus and move to a ribosome, where codons forming the genetic code specify the particular amino acids that are needed to make the individual protein. The mRNA associated with a ribosome calls for a particular amino acid as determined by the "genetic code" (see Figure A). Each amino acid is brought to the ribosome by another special kind of RNA called transfer RNA (tRNA). These tRNAs are specific for the particular amino acid they carry and recognize the codons along the mRNA. As each amino acid is brought to the ribosome by tRNA and added to a growing polypeptide chain, the ribosome moves further along the mRNA chain to the next codon until the entire sequence is completed. The complete polypeptide chain can then be folded and assembled into a functional protein.

Human Genome Project

ATCGCTATGCTCTA

What is the Human Genome Project?



The Human Genome Project (HGP) began in 1990 when the U.S. National Institutes of Health and the Department of Energy joined forces with international partners in a concerted effort to determine the correct sequence of all three billion bases of DNA within the entire human genome. This was a daunting task. To illustrate the size of the genome, if we printed the single letter abbreviation of each base (A, C, T, G) of the entire genome sequence in a standard print size, we would end up with a stack of paper about as tall as the Washington Monument. An important part of the HGP was that it required the immediate deposit of all DNA sequence information into public databases on the World Wide Web so that anyone – including doctors, scientists, and pharmaceutical and biotechnology companies – could access it for free.

Achieving the goals of the HGP would have been impossible without major advances in technology. Over the course of the 13 years since the project began, the cost of sequencing has dropped from $10 per nucleotide base (A, T, C, or G) to less than nine cents. Thanks to the development of new, low-cost, rapid processing, DNA sequencing machines, which determine the specific order of nucleotide bases in the genome, what would have taken months to sequence now takes seconds. It was also necessary to develop major advances in laboratory tools, complex databases and analytical software, and take advantage of vast improvements in computer processing speeds. Today, there are a large number of resources that search, compare and analyze the human genome, available to the public at no cost. (You can access the human genome from any computer by going to www.genome.gov, clicking on The Human Genome Project and going to the Genome Hub.)

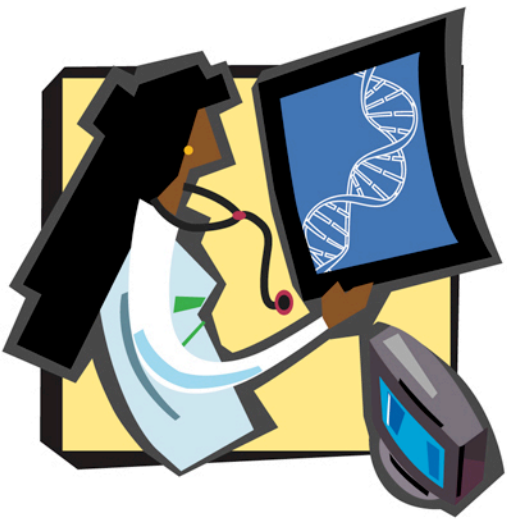
Ethical, Legal and Social Implications

With the powerful new tools of genomics, society needs to look carefully at the ethical, legal and social implications (referred to as "ELSI") that may arise from this science. So, the Human Genome Project included the establishment of an ELSI program to study these issues and to play a central role in encouraging society to use appropriately the knowledge gained from genomic research. How should this new genetic information be interpreted and used? Who should have access to it? How can people be protected from the harm that might result from its improper disclosure or use? How will the study of genomics affect society's concepts of race and ethnicity? Consideration of ELSI issues such as these will help develop the public policy options that include the consideration of the philosophical, theological and ethical consequences of understanding our own DNA blueprint. In this way, sensitive areas can be identified and solutions developed before scientific information is integrated into health care practice.

Implications for the Future: The Genomic Era

How does our knowledge of the human genome help scientists and doctors make advances in health science and medicine?

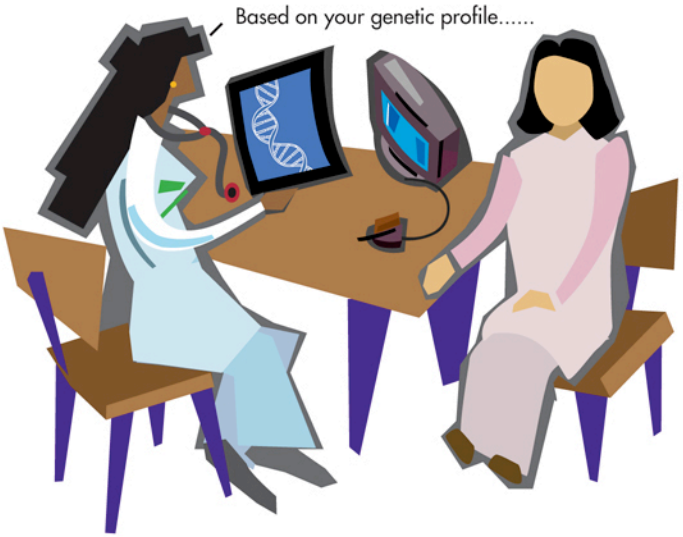
In the past, doctors and scientists did not have the benefit of a human genetic blueprint to help them better understand sickness and develop appropriate treatments. If a house needs repair or maintenance, mechanics and engineers can consult the blueprint when analyzing a problem and avoid unnecessary work, or, more importantly, avoid worsening the problem. Similarly, the blueprint of the human body provided through the Human Genome Project will help analyze problems when something goes wrong with a person (that is, when someone develops a disease). In the future, when a doctor is treating someone who is sick, he or she will be able to consult the patient's genetic blueprint in order to determine what variation of genes that patient has and prescribe a particular treatment that he or she knows is most likely to be effective for that individual. This will also help doctors avoid prescribing a drug that could cause a serious side effect.



With the completion of the Human Genome Project, we are now entering the Genomic Era.

The Human Genome Project has successfully determined the 99.9% of the genome that we all share in common. However, we said at the start that we are all unique. This wonderful diversity is also reflected in the different diseases that afflict us as individuals. So, one of the next challenges of the "Genomic Era" is to learn all of the possible genetic variations that can occur within the entire list of genes that malfunction to cause disease. This will help lead to new prevention strategies and treatments. Doctors will begin to think of diseases in terms of their causes, rather than only their symptoms. It has been said that "an ounce of prevention is worth a pound of cure." The Genomic Era promises great improvements in the prevention of illness, partly by identifying individuals at high risk of developing a disease. This will make medicine more effective and precise because doctors will be able to take the necessary measures to prevent illness, rather than waiting until symptoms occur.

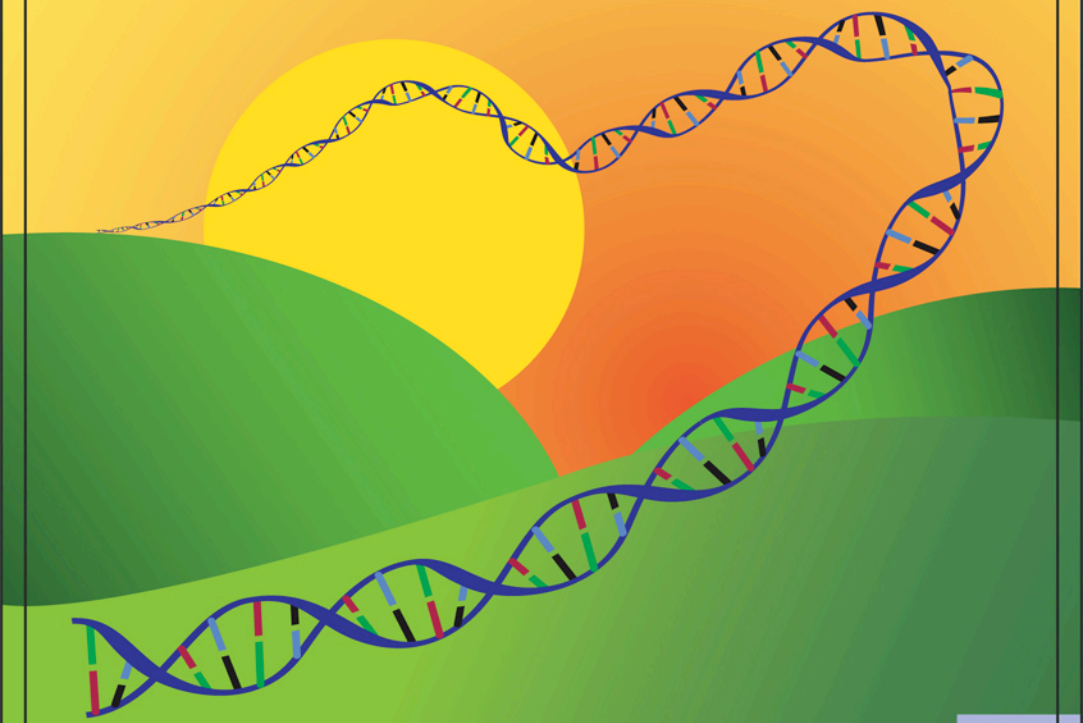
The environment is critically important too. Currently, relatively little is known about the genetic basis of some common diseases such as, cancer, diabetes and Alzheimer's disease, where risk factors may be associated also with diet, cigarette smoking or other environmental factors. Future genomic research will enhance our understanding of which genes interact with which environmental factors to cause these diseases. As our understanding of the genetic component of these and other diseases increases we will be better able to identify individual risks and to develop new, more efficient drugs. Drugs based on a precise understanding of the causes of disease will be more effective and less likely to cause side effects.



There are many other potential benefits of the Human Genome Project involving human health. There are also other benefits that don't directly involve human health, but will advance our knowledge of the world in which we live. As we look to the future of health sciences we see a bright new horizon.

"For the first time in history, humankind can read its genome - its Book of Life. This book is unlike any other, for, in reading it, we will uncover an ever-expanding view of ourselves."

- Francis S. Collins
Director, NHGRI



Glossary of Terms:

Codon: A single unit of the genetic code that is made up of three (triplet) nucleotide bases in a DNA or RNA molecule specifying a single amino acid.

DNA (deoxyribonucleic acid): The molecule that encodes genetic information. DNA is a double-stranded molecule made of two twisting, paired strands held together by weak bonds between base pairs of nucleotides.

ELSI: The Ethical, Legal, and Social Implications involved in genomics.

Gene: The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position within the genome that encodes a specific functional product (i.e., a protein or RNA molecule).

Genetic Code: The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. A gene's DNA sequence can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence.

Genome: All the genetic material of a particular organism; its size is generally given as its total number of base pairs or as its total number of genes.

Genomic Era: The new era in genetic research featuring rapid acquisition and integration of increasingly advanced genetic information resulting from the progress and completion of the Human Genome Project.

Human Genome Project: Research and technology development effort aimed at mapping and sequencing the entire genome of human beings.

mRNA: A molecule that can move from the nucleus to the cytoplasm of cells that serves as the crucial connecting message between information contained in the gene and protein synthesis. The structure of RNA is similar to that of DNA. The mRNA molecule serves as a template for the specific amino acid sequence of a protein.

Nucleotide bases: The basic subunits of DNA or RNA. Thousands of nucleotides are linked to form a DNA or RNA molecule. The four nucleotides in DNA contain the bases adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

Protein: A large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene that codes for the protein. Proteins are required for the structure, function and regulation of the body's cells, tissues and organs, and each protein has unique functions. Examples are hormones, enzymes, and antibodies.

Ribosome: A cytoplasmic organelle that serves as the molecular machine on which polypeptide synthesis from mRNA occurs.

Sequencing: Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule.

Transcription: The synthesis of an mRNA copy from a sequence of DNA (a gene), the first step in gene expression.

Translation: The process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids.

tRNA: A class of RNA that recognizes the triplet nucleotide coding sequences of mRNA and carries the appropriate amino acid to the ribosomes, where proteins are assembled according to the genetic code carried by mRNA.

Prepared by:
The National Human Genome Research Institute
National Institutes of Health
U.S. Department of Health and Human Services

April 2003

For More Information:

Mail:
National Institutes of Health
National Human Genome Research Institute
31 Center Drive, Room 4B09 MSC2152
Bethesda, MD 20892-2152

Phone:
(301) 402-0911

Website:
www.genome.gov

Related Websites:
U.S. Department of Energy Genome Programs: www.ornl.gov/hgmis
Medicine and the New Genetics: www.ornl.gov/hgmis/medicine/medicine.html
Ethical, Legal, and Social Implications: www.nhgri.nih.gov/About_NHGRI/Der/Elsi/
FAQs about genetics: www.genome.gov/page.cfm?pageID=10001191