

A Brief Guide to Genomics

DNA, Genes, and Genomes

Deoxyribonucleic acid (DNA) is the chemical compound that contains the instructions needed to develop and direct the activities of nearly all living organisms. DNA molecules are made of two twisting, paired strands, often referred to as a double helix.

Each DNA strand is made of four chemical units, called nucleotide bases, which comprise the genetic “alphabet”. The bases are adenine (A), thymine (T), guanine (G), and cytosine (C). Bases on opposite strands pair specifically: an A always pairs with a T; a C always pairs with a G. The order of the As, Ts, Cs, and Gs determines the meaning of the information encoded in that part of the DNA molecule just as the order of letters determines the meaning of a word.

An organism’s complete set of DNA is called its genome. Virtually every single cell in the body contains a complete copy of the approximately 3 billion DNA base pairs, or letters, that make up the human genome.

With its four-letter language, DNA contains the information needed to build the entire human body. A gene traditionally refers to the unit of DNA that carries the instructions for making a specific protein or set of proteins. Each of the estimated 20,000 to 25,000 genes in the human genome codes for an average of three proteins.

Located on 23 pairs of chromosomes packed into the nucleus of a human cell, genes direct the production of proteins with the assistance of enzymes and messenger molecules. Specifically, an enzyme copies the information in a gene’s DNA into a molecule called messenger ribonucleic acid RNA (mRNA). The mRNA travels out of the nucleus and into the cell’s cytoplasm, where the mRNA is read by a tiny molecular machine called a ribosome, and the information is used to link together small molecules called amino acids in the right order to form a specific protein.

Proteins make up body structures like organs and tissue, as well as control chemical reactions and carry signals between cells. If a cell’s DNA is mutated, an abnormal protein may be produced, which can disrupt the body’s usual processes and lead to a disease, such as cancer.

DNA Sequencing

Sequencing simply means determining the exact order of the bases in a strand of DNA. Because bases exist as pairs, and the identity of one of the bases in the pair determines the other member of the pair, researchers do not have to report both bases of the pair.

In the most common type of sequencing used today, called the chain termination method, a DNA strand is treated with a variety of nucleotides, a set of enzymes, and a specific primer to generate a collection of smaller DNA fragments. Four fluorescent tags, each specific for a given base, is part of the mixture. Each of the fragments differs in length by one base and is marked with a fluorescent tag that identifies the last base of the fragment. The fragments are then separated according to size and passed by a detector that reads the fluorescent tag. Then, a computer reconstructs the entire sequence of the long DNA strand by identifying the base at each position from the size of each fragment and the particular fluorescent signal at its end.

At present, this technology only can determine the order of up to 800 base pairs of DNA at a time. So, to assemble the sequence of all the bases in a large piece of DNA, such as a gene, researchers need to read the sequence of overlapping segments. This allows the longer sequence to be assembled from shorter pieces, somewhat like putting together a linear jigsaw puzzle. In this process, each base has to be read not just once, but at least several times in the overlapping segments to ensure accuracy.

Researchers can use DNA sequencing to search for genetic variations and/or mutations that may play a role in the development or progression of a disease. The disease-causing change may be as small as the substitution, deletion, or addition of a single base pair or as large as a deletion of thousands of bases.

The Human Genome Project

The Human Genome Project, which was led at the National Institutes of Health (NIH) by the National Human Genome Research Institute, produced a very high-quality version of the human genome sequence that is freely available in public databases. That international project was successfully completed in April 2003, under budget and more than two years ahead of schedule.

The sequence is not that of one person, but is a composite derived from several individuals. Therefore, it is a “representative” or generic sequence. To ensure anonymity of the DNA donors, more blood samples (nearly 100) were collected from volunteers than were used, and no names were attached to the samples that were analyzed. Thus, not even the donors knew whether their samples were actually used.

The Human Genome Project was designed to generate a resource that could be used for a broad range of biomedical studies. One such use is to look for the genetic variations that increase risk of specific diseases, such as cancer, or to look for the type of genetic mutations frequently seen in cancerous cells. More research can then be done to fully understand how the genome functions and to discover the genetic basis for health and disease.

The International HapMap Project, in which NIH also played a leading role, represents a major step in that direction. In October 2005, the project published a comprehensive map

of human genetic variation that is already speeding the search for genes involved in common, complex diseases, such as heart disease, diabetes, blindness, and cancer.

Another initiative that builds upon the tools and technologies created by the Human Genome Project is The Cancer Genome Atlas pilot project. This three-year pilot, which was launched in December 2005, will develop and test strategies for a comprehensive exploration of the universe of genetic factors involved in cancer.

Implications of Genomics for Medical Science

Virtually every human ailment, except perhaps trauma, has some basis in our genes. Until recently, doctors were able to take the study of genes, or genetics, into consideration only in cases of birth defects and a limited set of other diseases. These were conditions, such as sickle cell anemia, which have very simple, predictable inheritance patterns because each is caused by a change in a single gene.

With the vast trove of data about human DNA generated by the Human Genome Project and the HapMap Project, scientists and clinicians have much more powerful tools to study the role that genetic factors play in much more complex diseases, such as cancer, diabetes, and cardiovascular disease that constitute the majority of health problems in the United States. Genome-based research is already enabling medical researchers to develop more effective diagnostic tools, to better understand the health needs of people based on their individual genetic make-ups, and to design new treatments for disease. Thus, the role of genetics in health care is starting to change profoundly and the first examples of the era of personalized medicine are on the horizon.

It is important to realize, however, that it often takes considerable time, effort, and funding to move discoveries from the scientific laboratory into the medical clinic. Most new drugs based on genome-based research are estimated to be at least 10 to 15 years away. According to biotechnology experts, it usually takes more than a decade for a company to conduct the kinds of clinical studies needed to receive approval from the Food and Drug Administration.

Screening and diagnostic tests, however, are expected to arrive more quickly. Rapid progress is also anticipated in the emerging field of pharmacogenomics, which involves using information about a patient's genetic make-up to better tailor drug therapy to their individual needs.

Clearly, genetics remains just one of several factors that contribute to people's risk of developing most common diseases. Diet, lifestyle, and environmental exposures also come into play for many conditions, including many types of cancer. Still, a deeper understanding of genetics will shed light on more than just hereditary risks by revealing the basic components of cells and, ultimately, explaining how all the various elements work together to affect the human body in both health and disease.