**An allele** is one of two or more versions of DNA sequence (a single base or a segment of bases) at a given genomic location. An individual inherits two alleles, one from each parent, for any given genomic location where such variation exists. If the two alleles are the same, the individual is homozygous for that allele. If the alleles are different, the individual is heterozygous.

**Aneuploidy** is an abnormality in the number of chromosomes in a cell due to loss or duplication. In humans, aneuploidy would be any number of chromosomes other than the usual 46.

**A codon** is a DNA or RNA sequence of three nucleotides (a trinucleotide) that forms a unit of genetic information encoding a particular amino acid.

**An anticodon** is a trinucleotide sequence located at one end of a transfer RNA (tRNA) molecule, which is complementary to a corresponding codon in a messenger RNA (mRNA) sequence. Each time an amino acid is added to a growing polypeptide during protein synthesis, a tRNA anticodon pairs with its complementary codon on the mRNA molecule, ensuring that the appropriate amino acid is inserted into the polypeptide.

**A base pair** consists of two complementary DNA nucleotide bases that pair together to form a “rung of the DNA ladder.” DNA is made of two linked strands that wind around each other to resemble a twisted ladder — a shape known as a double helix. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are held together by hydrogen bonds between pairs of bases: adenine pairs with thymine, and cytosine pairs with guanine.

**Bioinformatics,** as related to genetics and genomics, is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences. Scientists and clinicians use databases that organize and index such biological information to increase our understanding of health and disease and, in certain cases, as part of medical care.

**cDNA** (short for copy DNA; also called complementary DNA) is synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase. While DNA is composed of both coding and non-coding sequences, cDNA contains only coding sequences. Scientists often synthesize and use cDNA as a tool in gene cloning and other research experiments.

A **centimorgan** (abbreviated cM) is a unit of measure for the frequency of genetic recombination. One centimorgan is equal to a 1% chance that two markers on a chromosome will become separated from one another due to a recombination event during meiosis (which occurs during the formation of egg and sperm cells). On average, one centimorgan corresponds to roughly 1 million base pairs in the human genome.

**Central Dogma.** The central dogma of molecular biology is a theory first proposed by Francis Crick in 1958. It states that genetic information flows only in one direction, from DNA to RNA to protein. Scientists have since discovered several exceptions to the theory.

**Centromere.** The centromere appears as a constricted region of a chromosome and plays a key role in helping the cell divide up its DNA during division (mitosis and meiosis). Specifically, it is the region where the cell’s spindle fibers attach. Following attachment of the spindle fibers to the centromere, the two identical sister chromatids that make up the replicated chromosome are pulled to opposite sides of the dividing cell, such that the two resulting daughter cells end up with identical DNA.

**Cromatid.** A chromatid is one of the two identical halves of a chromosome that has been replicated in preparation for cell division. The two “sister” chromatids are joined at a constricted region of the chromosome called the centromere. During cell division, spindle fibers attach to the centromere and pull each of the sister chromatids to opposite sides of the cell. Soon after, the cell divides in two, resulting in daughter cells with identical DNA.

**Chromatin.** Chromatin refers to a mixture of DNA and proteins that form the chromosomes found in the cells of humans and other higher organisms. Many of the proteins — namely, histones — package the massive amount of DNA in a genome into a highly compact form that can fit in the cell nucleus.

**Chromosome.** Chromosomes are threadlike structures made of protein and a single molecule of DNA that serve to carry the genomic information from cell to cell. In plants and animals (including humans), chromosomes reside in the nucleus of cells. Humans have 22 pairs of numbered chromosomes (autosomes) and one pair of sex chromosomes (XX or XY), for a total of 46. Each pair contains two chromosomes, one coming from each parent, which means that children inherit half of their chromosomes from their mother and half from their father. Chromosomes can be seen through a microscope when the nucleus dissolves during cell division.

**CRISPR** (short for “clustered regularly interspaced short palindromic repeats”) is a technology that research scientists use to selectively modify the DNA of living organisms. CRISPR was adapted for use in the laboratory from naturally occurring genome editing systems found in bacteria.

**Crossing Over.** Crossing over, as related to genetics and genomics, refers to the exchange of DNA between paired homologous chromosomes (one from each parent) that occurs during the development of egg and sperm cells (meiosis). This process results in new combinations of alleles in the gametes (egg or sperm) formed, which ensures genomic variation in any offspring produced.

**Deoxyribonucleic acid** (abbreviated DNA) is the molecule that carries genetic information for the development and functioning of an organism. DNA is made of two linked strands that wind around each other to resemble a twisted ladder — a shape known as a double helix. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are connected by chemical bonds between the bases: adenine bonds with thymine, and cytosine bonds with guanine. The sequence of the bases along DNA’s backbone encodes biological information, such as the instructions for making a protein or RNA molecule.

**Deletion.** A deletion, as related to genomics, is a type of mutation that involves the loss of one or more nucleotides from a segment of DNA. A deletion can involve the loss of any number of nucleotides, from a single nucleotide to an entire piece of a chromosome.

**Diploid** is a term that refers to the presence of two complete sets of chromosomes in an organism’s cells, with each parent contributing a chromosome to each pair. Humans are diploid, and most of the body’s cells contain 23 chromosomes pairs. Human gametes (egg and sperm cells), however, contain a single set of chromosomes and are said to be haploid.

**DNA replication** is the process by which the genome’s DNA is copied in cells. Before a cell divides, it must first copy (or replicate) its entire genome so that each resulting daughter cell ends up with its own complete genome.

**DNA sequencing** refers to the general laboratory technique for determining the exact sequence of nucleotides, or bases, in a DNA molecule. The sequence of the bases (often referred to by the first letters of their chemical names: A, T, C, and G) encodes the biological information that cells use to develop and operate. Establishing the sequence of DNA is key to understanding the function of genes and other parts of the genome. There are now several different methods available for DNA sequencing, each with its own characteristics, and the development of additional methods represents an active area of genomics research.

**Dominant Traits and Alleles.** Dominant, as related to genetics, refers to the relationship between an observed trait and the two inherited versions of a gene related to that trait. Individuals inherit two versions of each gene, known as alleles, from each parent. In the case of a dominant trait, only one copy of the dominant allele is required to express the trait. The effect of the other allele (the recessive allele) is masked by the dominant allele. Typically, an individual who carries two copies of a dominant allele exhibits the same trait as those who carry only one copy. This contrasts to a recessive trait, which requires that both alleles be present to express the trait.

**Electrophoresis** is a laboratory technique used to separate DNA, RNA or protein molecules based on their size and electrical charge. An electric current is used to move the molecules through a gel or other matrix. Pores in the gel or matrix work like a sieve, allowing smaller molecules to move faster than larger molecules. To determine the size of the molecules in a sample, standards of known sizes are separated on the same gel and then compared to the sample.

**Epigenetics** (also sometimes called epigenomics) is a field of study focused on changes in DNA that do not involve alterations to the underlying sequence. The DNA letters and the proteins that interact with DNA can have chemical modifications that change the degrees to which genes are turned on and off. Certain epigenetic modifications may be passed on from parent cell to daughter cell during cell division or from one generation to the next. The collection of all epigenetic changes in a genome is called an epigenome.

**Exon**. An exon is a region of the genome that ends up within an mRNA molecule. Some exons are coding, in that they contain information for making a protein, whereas others are non-coding. Genes in the genome consist of exons and introns.

**Gene.** The gene is considered the basic unit of inheritance. Genes are passed from parents to offspring and contain the information needed to specify physical and biological traits. Most genes code for specific proteins, or segments of proteins, which have differing functions within the body. Humans have approximately 20,000 protein-coding genes.

**Genetic code** refers to the instructions contained in a gene that tell a cell how to make a specific protein. Each gene’s code uses the four nucleotide bases of DNA: adenine (A), cytosine (C), guanine (G) and thymine (T) — in various ways to spell out three-letter “codons” that specify which amino acid is needed at each position within a protein.

**Genetic testing** is the use of a laboratory test to examine an individual’s DNA for variations, typically performed in the context of medical care, ancestry studies or forensics. In a medical setting, the results of a genetic test can be used to confirm or rule out a suspected genetic disease. Results may also be used to determine the likelihood of parents passing on a genetic mutation to their offspring. Genetic testing may be performed prenatally or after birth. Genetic testing is also used to study the genomes of tumors in cancer cases.

**Genetics** is the branch of biology concerned with the study of inheritance, including the interplay of genes, DNA variation and their interactions with environmental factors.

**Genome.** The genome is the entire set of DNA instructions found in a cell. In humans, the genome consists of 23 pairs of chromosomes located in the cell’s nucleus, as well as a small chromosome in the cell’s mitochondria. A genome contains all the information needed for an individual to develop and function.

<https://www.genome.gov/genetics-glossary>