

# Overview of Genetics

By **David N. Finegold**, MD, University of Pittsburgh

Last full review/revision Sep 2019 | Content last modified Sep 2019

A gene, the basic unit of heredity, is a segment of DNA containing all the information necessary to synthesize a polypeptide (protein). Protein synthesis, folding, and tertiary and quaternary structure ultimately determine much of the body's structure and function.

## Structure

---

Humans have about 20,000 to 23,000 genes depending on how a gene is defined. Genes are contained in chromosomes in the cell nucleus and mitochondria. In humans, somatic (nongerm) cell nuclei normally have 46 chromosomes in 23 pairs. Each pair consists of one chromosome from the mother and one from the father. Twenty-two of the pairs, chromosome numbers 1 to 22, the autosomes, are normally homologous (identical in size, shape, and position and number of genes). The 23rd pair, the sex chromosomes (X and Y), determines a person's sex as well as containing other functional genes. Women have 2 X chromosomes (which are homologous) in somatic cell nuclei; men have one X and one Y chromosome (which are heterologous).

The X chromosome carries genes responsible for many hereditary traits; the smaller Y chromosome carries genes that initiate male sex differentiation, as well as a few other genes. Because the X chromosome has many more genes than the Y chromosome, many X chromosome genes in males are not paired; in order to maintain a balance of genetic material between men and women, one of the X chromosomes in women is randomly inactivated ([lyonization](#)). A karyotype illustrates the full set of chromosomes in a person's cells.

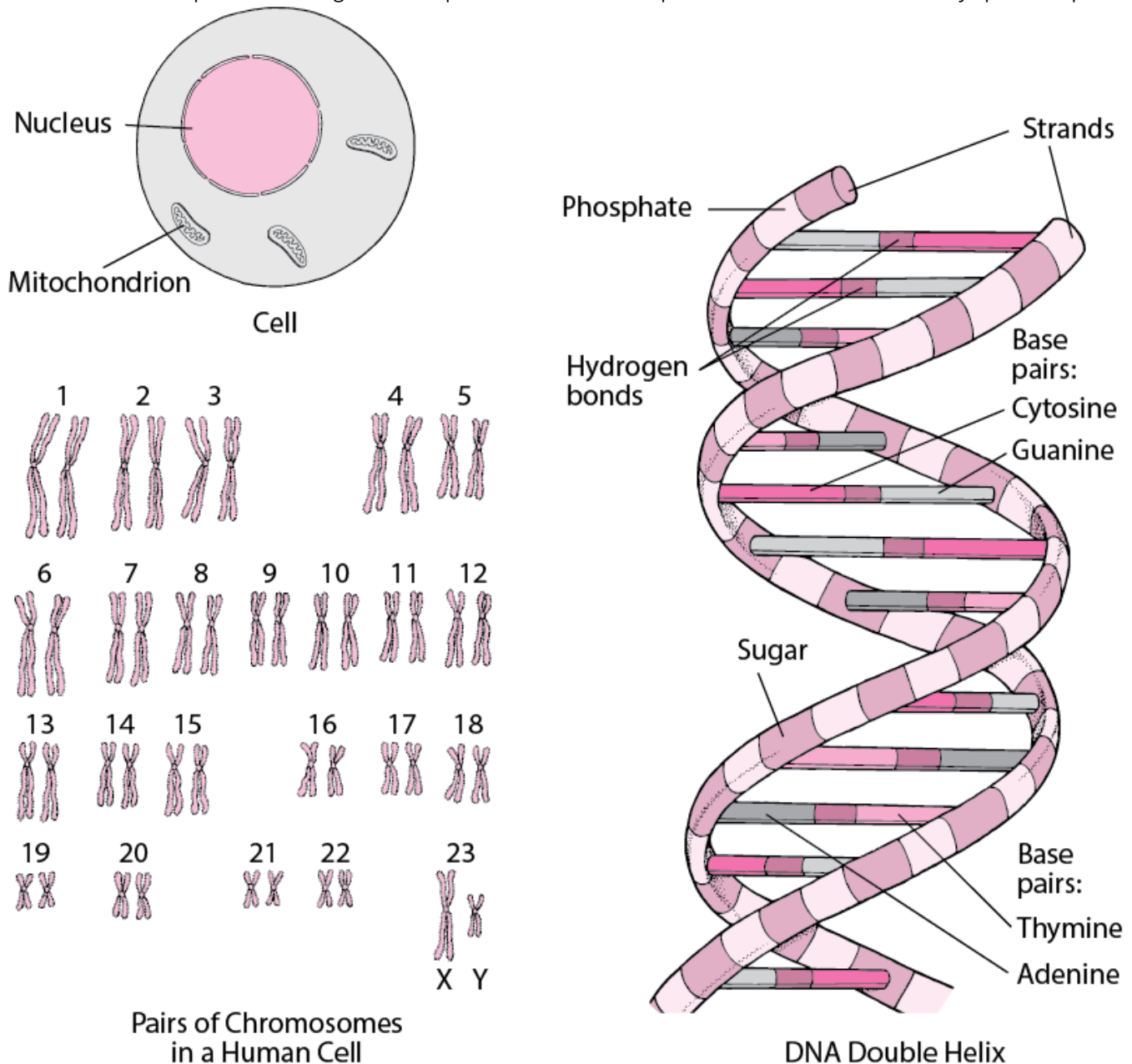
Germ cells (egg and sperm) divide through meiosis, which reduces the number of chromosomes to 23—half the number in somatic cells. In meiosis, the genetic information inherited from a person's mother and father is recombined through crossing over (exchange between homologous chromosomes). When an egg is fertilized by a sperm at conception, the normal number of 46 chromosomes is reconstituted.

Genes are arranged linearly along the DNA of chromosomes. Each gene has a specific location (locus), which is typically the same on each of the 2 homologous chromosomes. The genes that occupy the same locus on each chromosome of a pair (one inherited from the mother and one from the father) are called alleles. Each gene consists of a specific DNA sequence; 2 alleles may have slight differences or the same DNA sequences. Having a pair of identical alleles for a particular gene is homozygosity; having a pair of nonidentical alleles is heterozygosity. Some genes occur in multiple copies that may be next to each other or in different locations in the same or different chromosomes.

## Structure of DNA

DNA is the cell's genetic material, contained in chromosomes within the cell nucleus and mitochondria. Somatic (nongerm) cell nuclei normally have 46 chromosomes in 23 pairs. A chromosome contains many genes. A gene, the basic unit of heredity, is a segment of DNA containing all the information necessary to synthesize a polypeptide (protein).

The DNA molecule is a long, coiled double helix of two strands composed of sugar (deoxyribose) and phosphate molecules, that are connected by pairs of four nucleotides (bases). The nucleotides are paired: adenine (A) with thymine (T) and guanine (G) with cytosine (C). Each pair of nucleotides is held together by a hydrogen bond. The code in DNA is written in triplets containing 3 of the 4 possible nucleotides. Specific amino acids are coded by specific triplets.



## Gene Function

Genes consist of DNA. The length of the gene determines the length of the protein the gene codes for. DNA is a double helix in which nucleotides (bases) are paired:

Adenine (A) is paired with thymine (T)

Guanine (G) is paired with cytosine (C)

DNA is transcribed during protein synthesis, in which one strand of DNA is used as a template against which messenger RNA (mRNA) is synthesized. RNA has the same base pairs as DNA, except that uracil (U) replaces thymine (T). Parts of mRNA travel from the nucleus to the cytoplasm and then to the ribosome, where protein synthesis occurs. Transfer RNA (tRNA) brings each amino acid back to the ribosome where it is added to the growing polypeptide chain in a sequence determined by the mRNA. As a chain of amino acids is assembled, it folds upon itself to create a complex 3-dimensional structure under the influence of nearby chaperone molecules.

The code in DNA is written in triplets containing 3 of the 4 possible nucleotides. Specific amino acids are coded by specific triplets. Because there are 4 nucleotides, the number of possible triplets is  $4^3$  (64). Because there are only 20 amino acids, there are redundant (extra) triplet combinations. Some triplets code for the same amino acids as other triplets. Other triplets may code for elements such as instructions to start or stop protein synthesis and the order in which to combine and assemble amino acids.

Genes consist of exons and introns. Exons code for amino acid components of the final protein. Introns contain other information that affects control and speed of protein production. Exons and introns together are transcribed onto mRNA, but the segments transcribed from introns are later spliced out. Many factors regulate transcription, including antisense RNA, which is synthesized from the DNA strand that is not transcribed into mRNA. In addition to DNA, chromosomes contain histones and other proteins that affect gene expression (which proteins and how many proteins are synthesized from a given gene).

**Genotype** refers to a specific genetic composition and sequence; it determines which proteins are coded for production.

**Genome** refers to the entire composition of a set of haploid chromosomes (single strand), including the genes they contain.

**Phenotype** refers to the entire physical, biochemical, and physiologic makeup of a person—ie, how the cell (and thus the body) functions. Phenotype is determined by the types and amounts of proteins actually synthesized, ie, how the genes are actually expressed. Specific genotypes may or may not correlate well with phenotype.

**Expression** refers to the process in which the information encoded in a gene is used to control the assembly of a molecule (usually protein or RNA). Gene expression depends on multiple factors such as whether a trait is dominant or recessive, the penetrance and expressivity of the gene (see [Factors Affecting Gene Expression](#)), degree of tissue differentiation (determined by tissue type and age), environmental factors, whether expression is sex-limited or subject to chromosomal inactivation or genomic imprinting, and other unknown factors.

## Epigenetic factors

Factors that affect gene expression without changing the genome sequence are epigenetic factors.

Knowledge of the many biochemical mechanisms that mediate gene expression is growing rapidly. One mechanism is variability in intron splicing (also called alternative splicing). Because introns are spliced out, the exons may also be spliced out, and then the exons can be assembled in many combinations, resulting in many different mRNAs capable of coding for similar but different proteins. The number of proteins that can be synthesized by humans is > 100,000 even though the human genome has only about 20,000+ genes.

Other mechanisms mediating gene expression include DNA methylation and histone interactions involving methylation and acetylation. DNA methylation tends to silence gene function. Histone proteins resemble spools around which DNA coils. Histone modifications such as methylation can increase or decrease the quantity of proteins synthesized from a particular gene. Histone acetylation is associated with decreased gene expression. The strand of DNA that is not transcribed to form mRNA may also be used as a template for synthesis of RNA that controls transcription of the opposite strand.

Another important mechanism involves microRNAs (miRNAs). MiRNAs are short, hairpin-derived (hairpin refers to the shape the RNA sequences assume as they bind together) RNAs that repress target gene expression after transcription. They may be involved in regulation of as many as 60% of transcribed proteins.

## Traits and Inheritance Patterns

A trait may be as simple as eye color or as complex as susceptibility to diabetes. Expression of a trait may involve one gene or many genes. Some [single-gene defects](#) cause abnormalities in multiple tissues, an effect called pleiotropy. For example, [osteogenesis imperfecta](#) (a connective tissue disorder that often results from abnormalities in a single collagen gene) may cause fragile bones, deafness, blue-colored sclerae, dysplastic teeth, hypermobile joints, and heart valve abnormalities.

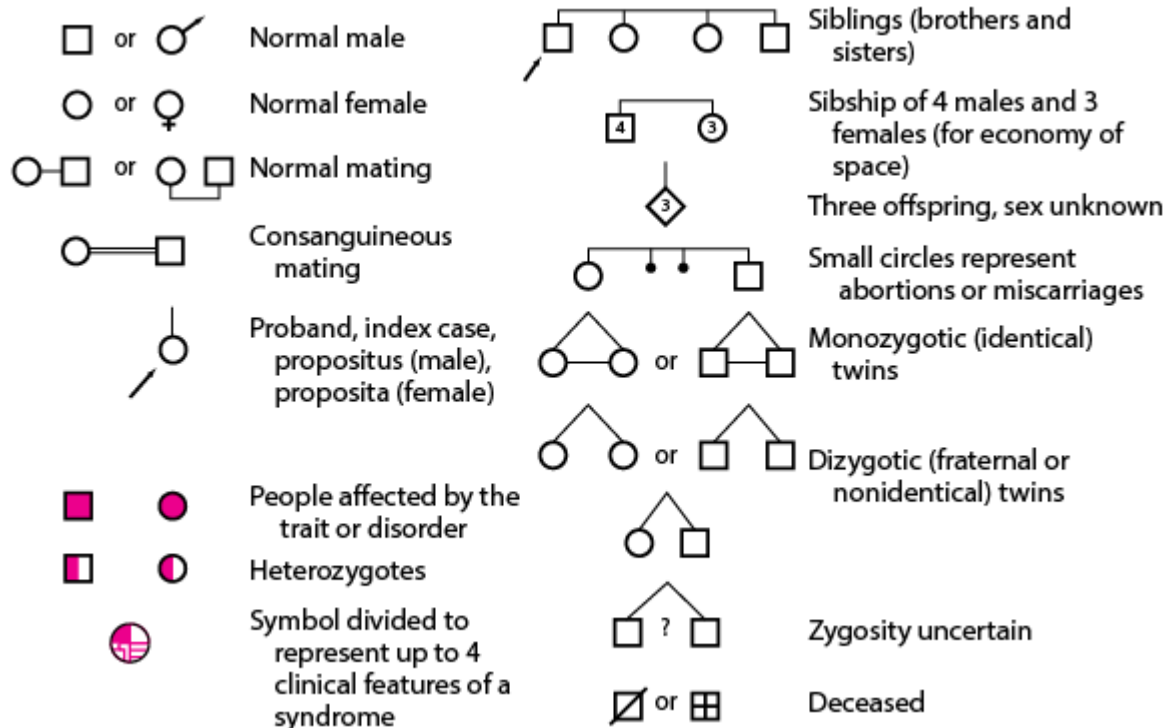
## Construction of a family pedigree

The family pedigree (family tree) is used to diagram inheritance patterns. Pedigrees are commonly used in genetic counseling. The pedigree uses conventional symbols to represent family members and pertinent health information about

them (see Figure: [Symbols for constructing a family pedigree](#)). Some familial disorders with identical phenotypes have multiple patterns of inheritance.

## Symbols for constructing a family pedigree

In the pedigree, symbols for each generation in the family are placed in a row and numbered with Roman numerals, starting with the older generation at the top and ending with the most recent at the bottom (see Figure: [Autosomal dominant inheritance](#), see Figure: [Autosomal recessive inheritance](#), see Figure: [X-linked dominant inheritance](#), and see Figure: [X-linked recessive inheritance](#)). Within each generation, people are numbered from left to right with Arabic numerals. Siblings are usually listed by age, with the oldest on the left. Thus, each member of the pedigree can be identified by 2 numbers (eg, II, 4). A spouse is also assigned an identifying number.



## Key Points

Phenotype is determined by gene expression as well as the genotype.

Mechanisms regulating gene expression are being elucidated and include intron splicing, DNA methylation, histone reactions, and microRNAs.

