

# DNA, genes and chromosomes

## Learning objectives

By the end of this learning material you would have learnt about the components of a DNA and the process of DNA replication, gene types and sequencing and the structural properties of a chromosome.

## DNA

**DNA (or deoxyribonucleic acid)** is the molecule that carries the genetic information in all cellular forms of life and some viruses. It belongs to a class of molecules called the nucleic acids, which are polynucleotides - that is, long chains of nucleotides.

Each nucleotide consists of three components:

- a nitrogenous base: cytosine (C), guanine (G), adenine (A) or thymine (T)
- a five-carbon sugar molecule (deoxyribose in the case of DNA)
- a phosphate molecule

The backbone of the polynucleotide is a chain of sugar and phosphate molecules. Each of the sugar groups in this sugar-phosphate backbone is linked to one of the four nitrogenous bases.

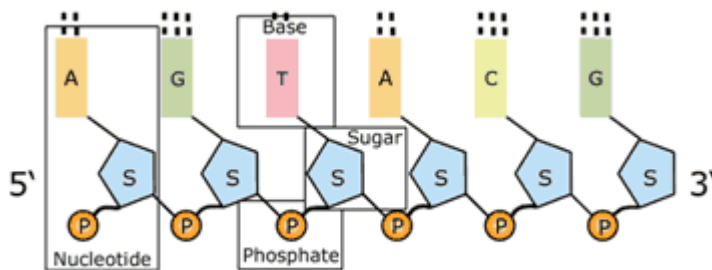


Image adapted from: National Human Genome Research Institute.

DNA's ability to store - and transmit - information lies in the fact that it consists of two polynucleotide strands that twist around each other to form a double-stranded helix. The bases link across the two strands in a specific manner using hydrogen bonds: cytosine (C) pairs with guanine (G), and adenine (A) pairs with thymine (T).

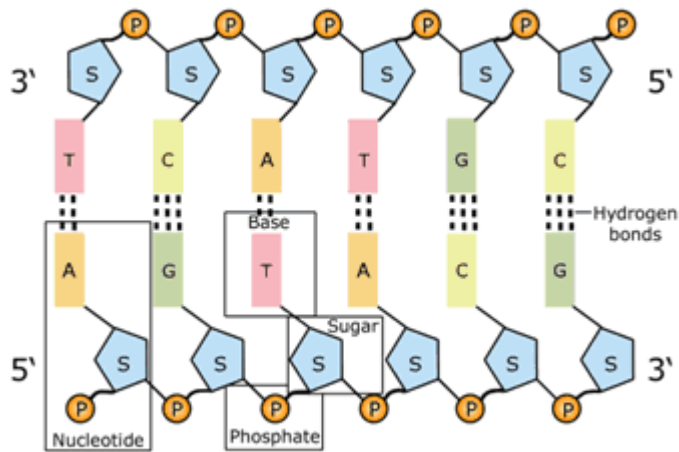


Image adapted from: National Human Genome Research Institute.

The double helix of the complete DNA molecule resembles a spiral staircase, with two sugar phosphate backbones and the paired bases in the centre of the helix. This structure explains two of the most important properties of the molecule. First, it can be copied or 'replicated', as each strand can act as a template for the generation of the complementary strand. Second, it can store information in the linear sequence of the nucleotides along each strand.

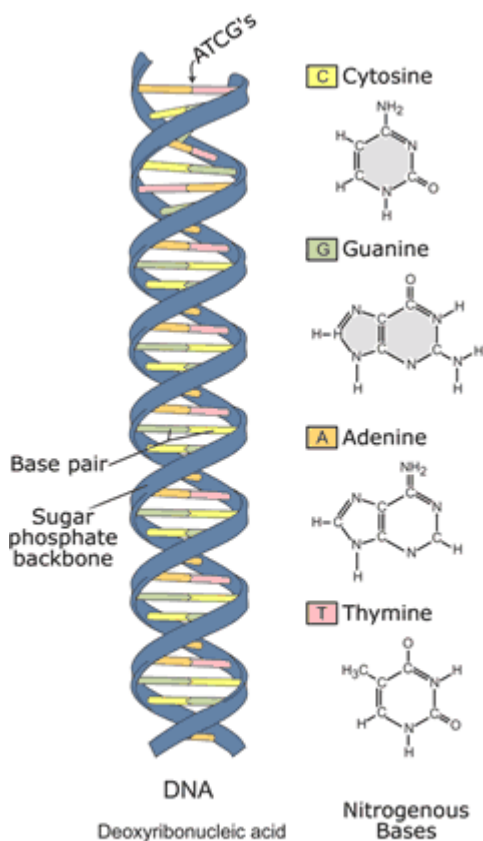
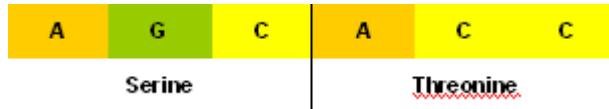


Image adapted from: National Human Genome Research Institute.

It is the order of the bases along a single strand that constitutes the genetic code. The four-letter 'alphabet' of A, T, G and C forms 'words' of three letters called codons. Individual codons code for specific amino acids. A gene is a sequence of nucleotides along a DNA

strand - with 'start' and 'stop' codons and other regulatory elements - that specifies a sequence of amino acids that are linked together to form a protein.

So, for example, the codon AGC codes for the amino acid serine, and the codon ACC codes for the amino acid threonine.



There are a two points to note about the genetic code:

- It is **universal**. All life on Earth uses the same code (with a few minor exceptions).
- It is **degenerate**. Each amino acid can be coded for by more than one codon. For example, AGC and ACC both code for the amino acid serine.

A codon table sets out how the triplet codons code for specific amino acids.

		Second base of codon				
		U	C	A	G	
First base of codon	U	UUU Phenylalanine phe UUC phe UUA Leucine leu UUG leu	UCU Serine ser UCC Serine ser UCA Serine ser UCG Serine ser	UAU Tyrosine tyr UAC Tyrosine tyr UAA STOP codon UAG STOP codon	UGU Cysteine cys UGC Cysteine cys UGA STOP codon UGG Tryptophan trp	U C A G
	C	CUU Leucine leu CUC Leucine leu CUA Leucine leu CUG Leucine leu	CCU Proline pro CCC Proline pro CCA Proline pro CCG Proline pro	CAU Histidine his CAC Histidine his CAA Glutamine gin CAG Glutamine gin	CGU Arginine arg CGC Arginine arg CGA Arginine arg CGG Arginine arg	U C A G
	A	AUU Isoleucine ile AUC Isoleucine ile AUA Isoleucine ile AUG Methionine met (start codon)	ACU Threonine thr ACC Threonine thr ACA Threonine thr ACG Threonine thr	AAU Asparagine asn AAC Asparagine asn AAA Lysine lys AAG Lysine lys	AGU Serine ser AGC Serine ser AGA Arginine arg AGG Arginine arg	U C A G
	G	GUU Valine val GUC Valine val GUA Valine val GUG Valine val	GCU Alanine ala GCC Alanine ala GCA Alanine ala GCG Alanine ala	GAU Aspartic acid asp GAC Aspartic acid asp GAA Glutamic acid glu GAG Glutamic acid glu	GGU Glycine gly GGC Glycine gly GGA Glycine gly GGG Glycine gly	U C A G

© Clinical Tools, Inc.

## DNA replication

The enzyme helicase breaks the hydrogen bonds holding the two strands together, and both strands can then act as templates for the production of the opposite strand. The process is catalysed by the enzyme DNA polymerase, and includes a proofreading mechanism.

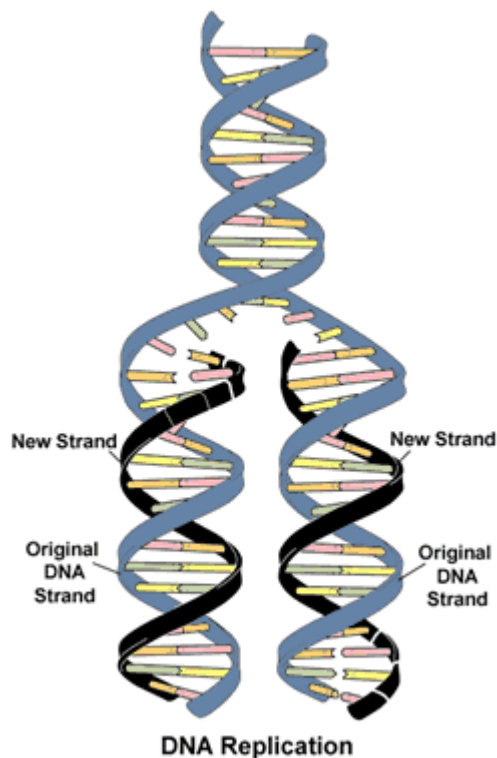


Image adapted from: National Human Genome Research Institute.

## Genes

The **gene** is the basic physical and functional unit of heredity. It consists of a specific sequence of nucleotides at a given position on a given chromosome that codes for a specific protein (or, in some cases, an RNA molecule).

Genes consist of three types of nucleotide sequence:

- coding regions, called **exons**, which specify a sequence of amino acids
- non-coding regions, called **introns**, which do not specify amino acids
- regulatory sequences, which play a role in determining when and where the protein is made (and how much is made)

A human being has 20,000 to 25,000 genes located on 46 chromosomes (23 pairs). These genes are known, collectively, as the human genome.

## Chromosomes

### Eukaryotic chromosomes

The label **eukaryote** is taken from the Greek for 'true nucleus', and eukaryotes (all organisms except viruses, Eubacteria and Archaea) are defined by the possession of a nucleus and other membrane-bound cell organelles.

The nucleus of each cell in our bodies contains approximately 1.8 metres of DNA in total, although each strand is less than one millionth of a centimetre thick. This DNA is tightly packed into structures called **chromosomes**, which consist of long chains of DNA and

associated proteins. In eukaryotes, DNA molecules are tightly wound around proteins - called **histone proteins** - which provide structural support and play a role in controlling the activities of the genes. A strand 150 to 200 nucleotides long is wrapped twice around a core of eight histone proteins to form a structure called a **nucleosome**. The histone octamer at the centre of the nucleosome is formed from two units each of histones H2A, H2B, H3, and H4. The chains of histones are coiled in turn to form a **solenoid**, which is stabilised by the histone H1. Further coiling of the solenoids forms the structure of the chromosome proper.

Each chromosome has a **p arm** and a **q arm**. The p arm (from the French word 'petit', meaning small) is the short arm, and the q arm (the next letter in the alphabet) is the long arm. In their replicated form, each chromosome consists of two **chromatids**.

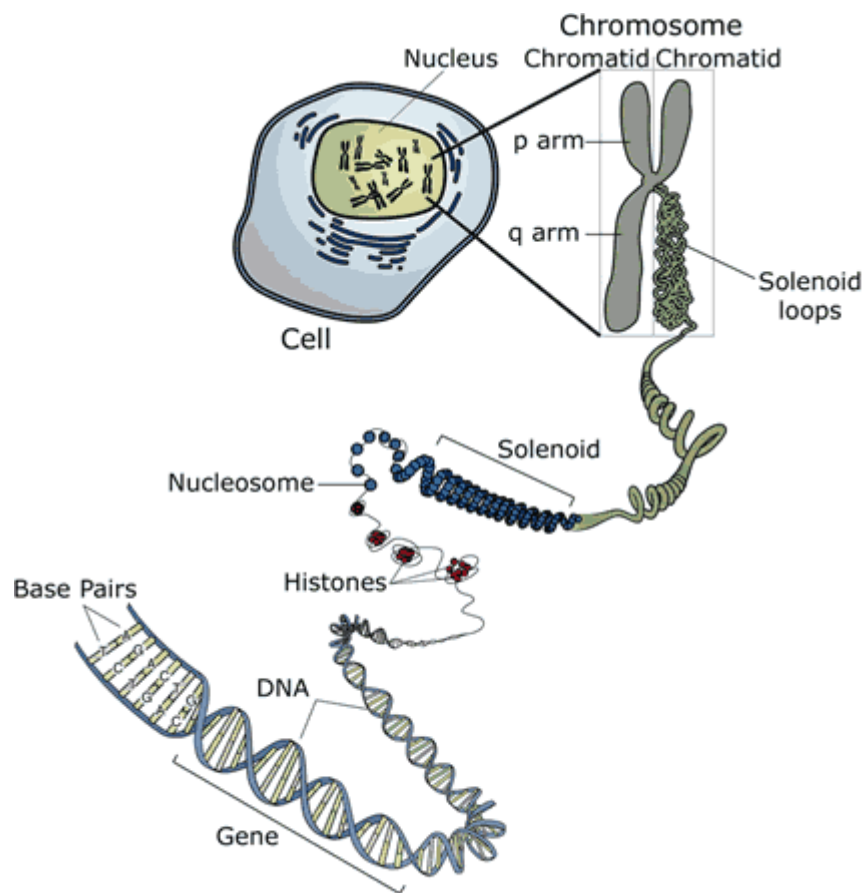
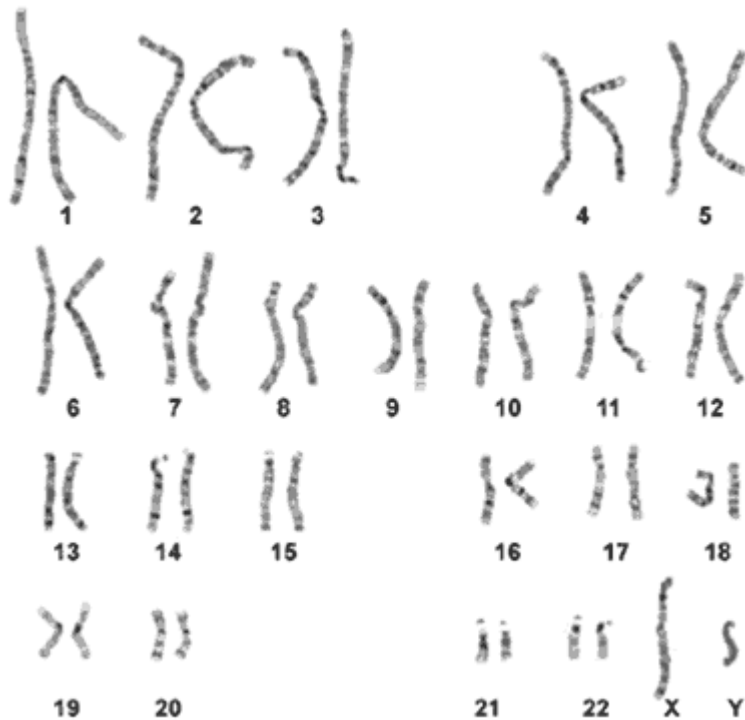


Image adapted from: National Human Genome Research Institute.

The chromosomes - and the DNA they contain - are copied as part of the cell cycle, and passed to daughter cells through the processes of mitosis and meiosis. Human beings have 46 chromosomes, consisting of 22 pairs of **autosomes** and a pair of **sex chromosomes**: two X sex chromosomes for females (XX) and an X and Y sex chromosome for males (XY). One member of each pair of chromosomes comes from the mother (through the egg cell); one member of each pair comes from the father (through the sperm cell).

A photograph of the chromosomes in a cell is known as a **karyotype**. The autosomes are numbered 1-22 in decreasing size order.



© Clinical Tools, Inc.

### Prokaryotic chromosomes

The **prokaryotes** (Greek for 'before nucleus' - including Eubacteria and Archaea) lack a discrete nucleus, and the chromosomes of prokaryotic cells are not enclosed by a separate membrane.

Most bacteria contain a single, circular chromosome. (There are exceptions: some bacteria - for example, the genus *Streptomyces* - possess linear chromosomes, and *Vibrio cholerae*, the causative agent of cholera, has two circular chromosomes.) The chromosome - together with ribosomes and proteins associated with gene expression - is located in a region of the cell cytoplasm known as the **nucleoid**.

The genomes of prokaryotes are compact compared with those of eukaryotes, as they lack introns, and the genes tend to be expressed in groups known as **operons**. The circular chromosome of the bacterium *Escherichia coli* consists of a DNA molecule approximately 4.6 million nucleotides long.

In addition to the main chromosome, bacteria are also characterised by the presence of extra-chromosomal genetic elements called **plasmids**. These relatively small circular DNA molecules usually contain genes that are not essential to growth or reproduction.



This work is licensed under a [Creative Commons Licence](https://creativecommons.org/licenses/by-nc-sa/4.0/)