

# Gene Structure and Function

A gene can be defined as a segment of DNA that contains information coding for the amino acid sequence of a polypeptide chain and any other regulatory sequences which are essential for its proper expression. By proper expression, we mean for the production of a normal mRNA molecule, in to correct amount, in the correct place, and at the correct time during the cell cycle or during development. A molecular definition could be that a gene is a sequence of DNA that is essential for the production of a functional product (be it a polypeptide or an RNA molecule)

In humans (and in most eukaryotic genomes), most of the genes are discontinuous. This means that they are interrupted by one or more non-coding regions. Amount of genes in humans is approximately 25000 genes arranged on 23 chromosomes.

In cells, not all the genes are active which means that not all of the are expressed. This implies that cells are differentiated (they acquire specialised functions). Promoter regions regulate the expression of genes. These regions are present at the 5' end of each gene and they include sequences which are responsible for the initiation of transcription. Only DNA sequences that are presented by a promoter are transcribed. There are also some regulatory elements. These include enhancers, silencers and locus control regions. They are not necessarily adjacent to the gene. They may lie a significant distance away from the coding portion of a gene. Both promoters and regulatory elements can still be a site of mutation in genetic disease that can interfere with the normal expression of the gene. At the 3' end of the gene, there is an untranslated region which codes a signal for the addition of a sequence of adenosine residues.

Genes are arranged linearly along the chromosomes. Each gene had a specific, defined position called a gene locus. Most genes are spread out randomly along the chromosomes. However, some genes are organised into groups or clusters. There are 2 types:

1. Operons
2. Multigene Families

## Pseudogenes

They are DNA sequences that resemble known genes but they do not have any function. There are many thousands of pseudogenes and they can be divided in 2 categories:

1. Non-processed: They represent genes which used to be active but are now inactivated due to mutations in regulatory or coding sequences.
2. Processed: These genes have been formed by retrotransposition and not by any mutation. The process of retrotransposition includes transcription, generating a DNA copy of the mRNA (reverse transcription with the use of RNA-dependent DNA Polymerase), and finally integration of this DNA copy into the genome. Since this DNA comes from mRNA (which has already been processed), it contains no introns and is not necessarily found on the same chromosome as that of their progenitor cell.

## Non-coding genes

In the human genome, not all genes encode for proteins. Even though their function is not yet fully understood, they seem to be of regulatory and structural significance to genes and cytoplasm/nucleus respectively. For instance, chromosome 11, in addition to its 1300 protein-coding genes, has about 200 non-coding RNA genes. These genes do not have a protein as a final product but an RNA molecule.

## Bibliography

NUSSBAUM, Robert L. *Genetics in Medicine : Thompson & Thompson*. 7th Edition edition. 2007. ISBN 978-1-4160-3080-5.

PASSARGE, Eberhard. *Color Atlas of Genetics*. 3rd Edition edition. 2007. ISBN 978-3-13-100363-8.