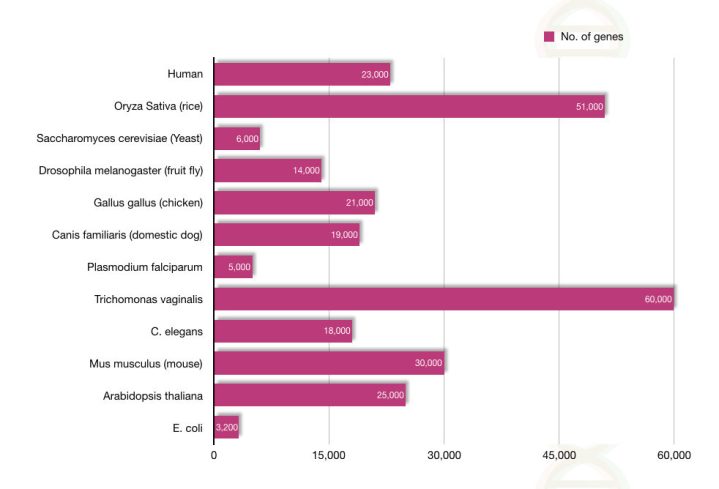
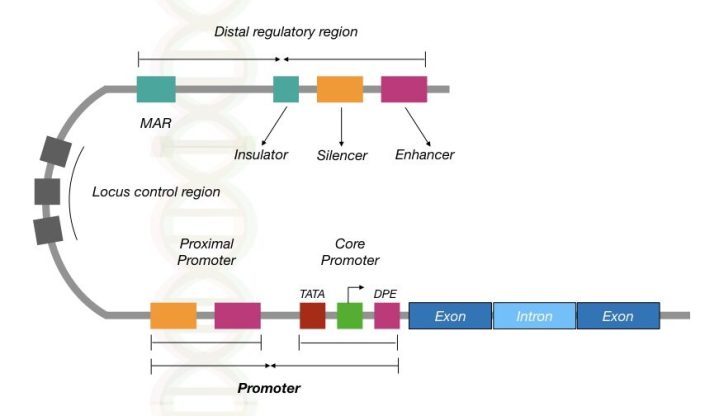
* **Gene and Genome**
* **What is a Gene - Definition, Structure And Function**

# A gene is a functional- hereditary unit made up of nucleotides which forms protein.

* The term gene was coined by Wilhelm Johannsen in 1909.
* Notably, the chemical structure of genes is almost similar in prokaryotes as well as eukaryotes although their location and regulation differ.
* Approximately 22,000 to 26,000 genes are present in the human genome. Some of them are active and some are inactive. Although, the number of genes in different organisms varies.
* Genes are located on chromosomes.
* 
* The number of genes in different organisms.
* Note that the size of the organism is not at all related to the number of genes
* Functionally, some genes can encode proteins while some can’t!. Those genes which can’t manufacture a specific protein helps in the [**regulation of gene expression**](https://geneticeducation.co.in/dna-to-protein-a-brief-overview-of-gene-expression/)**.**
* It’s not of a fixed size. Some genes are larger while some are very small.
* **Structure of gene:**
* Genes are actually DNA strands thus are made up of the nucleotide chain. The chemical structure of a gene comprises nucleotides. A part of DNA- genes are made up of A, T, G and C nucleotides.
  + ***only ~2.2% of human DNA are coding sequences.***
* **In general, the gene structure consists of Three types of elements:**
* **1-Core elements - The core elements or sequences actually take parts in protein formation.** **Exons** are core elements.
* **2-Regulatory elements-  The regulatory elements maintain gene expression**, are located on the extreme ends of a gene. Included: -
* **promoters**, **enhancers** and **silencers**
* **3- The third type of element called maintenance elements- possesses information for DNA repair, modification and replication.**
* **The functional or physical structure of a gene comprises**
* **1-Introns 2- Exons 3- Promotes 4- Enhancer 5- 3- UTRs.**
* ***1-Introns*** are intervening non-coding sequences removed from the final transcript.
* ***2-Exons*** are coding part of a gene which are joined after splicing and constructs the final transcript.
* 

**Goldberg-Hogness box**)

**downstream promoter element-Initiation**

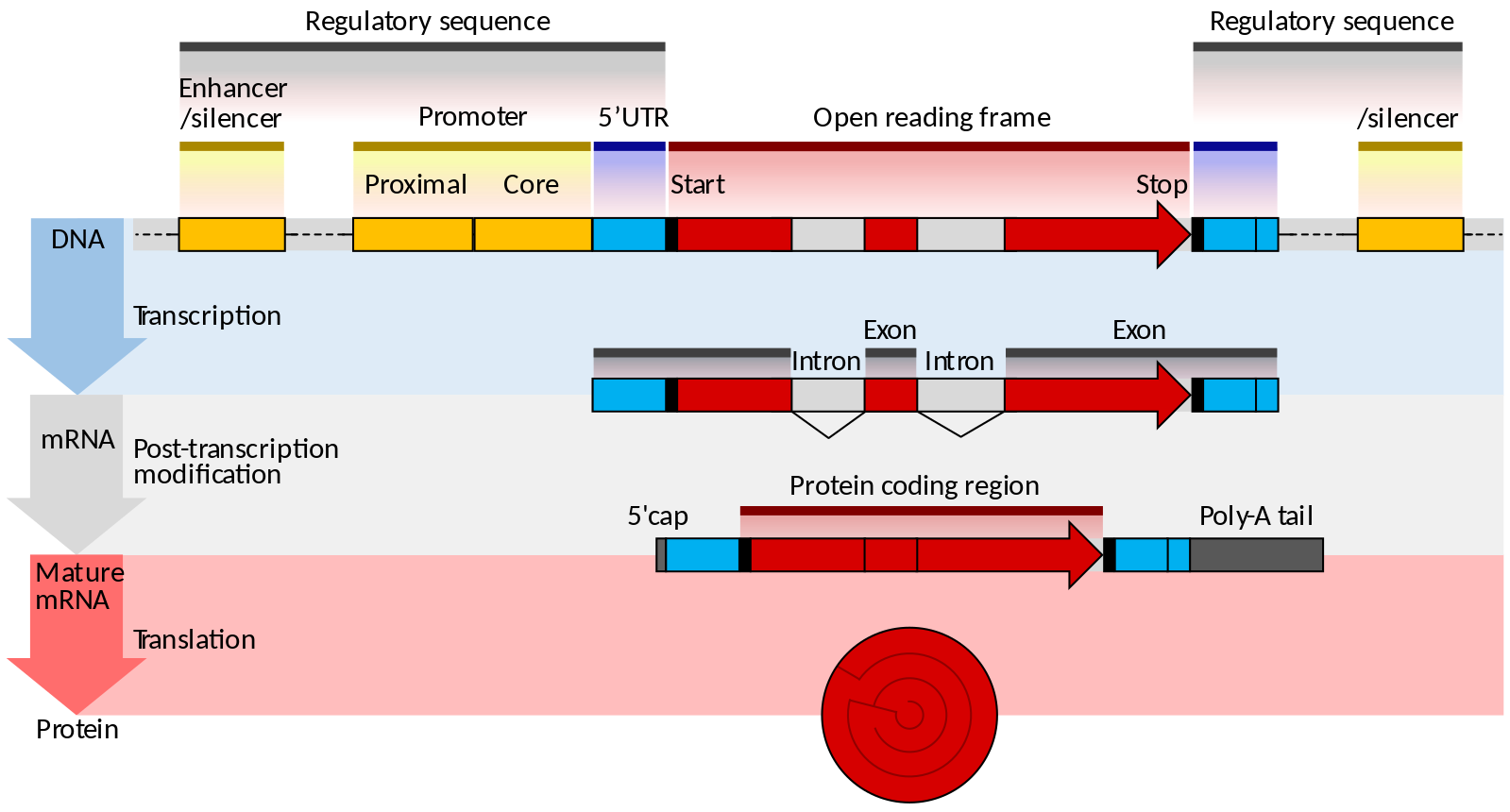
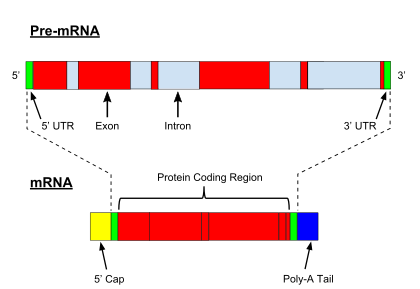
* **Regulate gene expression.**

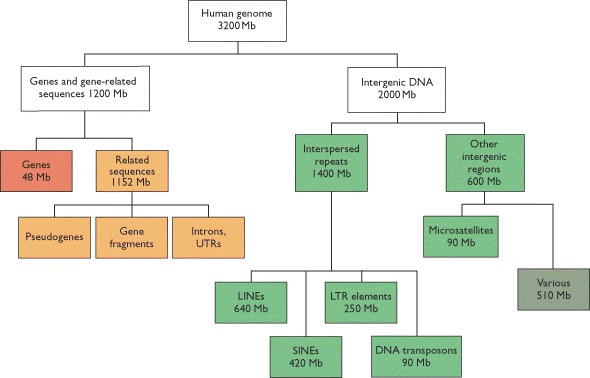
**RNA polymerase bindings**

transcriptional factors

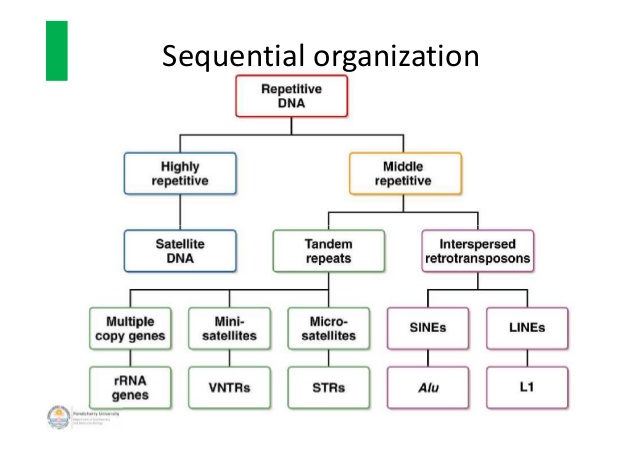
* **Fig(1) The molecular structure of a gene.**
* ***3-Promotes*** are non-coding sequences but facilitates binding sites for enzymes and transcriptional factors to work. The promoter consists of TATA box and CCAAT sequences for enzyme binding.
* The entire promoter region is located on the 5’ end and made up of core promoter and proximal promoter sequences (see the above image).
* **Here, the core promoter facilitates** **RNA polymerase bindings (and other proteins) to start transcription.** While the proximal promoter provides bindings for transcriptional factors.
* **4-The enhancer** induces transcription while the silencer represses it. Collectively, enhancers and silencers located far away from exon, regulate gene expression.
* **5 -The 3’ untranslated regions are non-coding regions** of gene helps in aborting the process of transcription and to form the final transcript.
* Once the RNA polymerase reaches the untranslated region it stops synthesizing RNA and detached or release from the strand.
* **Functions of gene:**
* The main function of a gene is to form or manufacture a protein, however, it’s not the only function. Indeed It’s partially true.
* Some genes can’t form protein, although they transcribe into mRNA. For instance, the [microRNAs](https://geneticeducation.co.in/microrna-mirna-and-gene-regulation/) are the type of tiny ribonucleic acid formed from some genes but it doesn’t undergo protein formation. It helps in gene regulation instead.
* **Now first let’s understand how genes form proteins?**
* From replication, the DNA or genes copied from one cell to two daughter cells. The process of replication is regulated by DNA polymerase.
* After that, the definite region of DNA (a gene) undergo transcription through RNA polymerase.
* As we said, RNA polymerase binds near the promoter region and starts adding nucleotides. The mRNA is constructed from a gene.
* After that, post-transcriptional modifications happen, followed by the migration of mRNA to the cytoplasm.
* At the ribosome, in the cytoplasm, the mRNA translated into the chain of amino acid. That is how the entire mechanism of protein formation occurs in a cell.
* **Here as we are discussing a “gene”, it is very important to mention how different genes work. 2 types of genes are present in our genome:-**
* ***1-Genes that encode single protein***— some genes only encode one particular protein, It has massages only to form a single type of protein product, for example, the hemoglobin gene *HBA* and *HBB*.
* The HBA gene encodes the alpha chain of Hb while the HBB gene encodes the beta-globin chain of the Hb protein.
* ***2-Genes that encode many proteins–*** In conventional genetics, scientists were believed that a single protein formed from one single gene. Nonetheless, the assumption was totally wrong.
* A gene, with many different exons, can create more than one kind of protein products. With the combination of different exonic sequences, various types of amino acid chains are constructed. And that is the beauty of it.
* **-You can’t believe in this!, “ A single fruit fly gene can encode 38,000 types of different proteins.”**
* **Types of genes**
* ***1-Non-protein coding genes–* some** **genes can’t form proteins instead of that they act in gene regulation.**
* Besides this, several other sets of genes based on their function are categorized here:
* **2-Housekeeping genes:** Genes required to perform normal functions of every cell are known as [housekeeping genes](https://geneticeducation.co.in/housekeeping-genes-a-basal-cell-maintenance-kit/). Usually, These genes code protein products for transcription, translation and replication.
  + **3-Inducible genes:** Normally inducible genes remain inactive, it
    - expressed under the influence of extrinsic factors.
    - **4-Developmental genes:** These types of genes helps in the early stage

of growth and development of organisms.

* **5- Luxury genes**  **(Tissue-specific genes):** Unlike the housekeeping genes, tissue-specific genes express only in some type of tissues. And it remains inactive in other types of tissues.  Tissue-specific or organ-specific, which means they are not expressed in all cells. They are not constantly expressed, only when their function is needed.
* **6-Homologous genes:** genes inherited from a common ancestor, share a common function and have sequence similarities are categorized into homologous genes.
* **7-Non-homologous genes:** Genes that aren’t inherited from a common ancestor instead, it is originated due to some evolutionary forces are known as non-homologous genes.
* **8-Autosomal genes:** genes located on autosomal chromosomes are categorized into autosomal genes
* **9-Sex-linked genes:** Genes located on X or Y chromosome (in humans) or on sex chromosomes are classified in this category. These genes are very crucial for reproductive health of a person.
* 
* **You can think of the genome as being split up into two parts:-**
* **1-There's the stuff that codes for proteins. We call it coding DNA**
* **2-Another stuff called - non-coding DNA. (Some people will like to try and refer to this as junk DNA).**
* **Types of non-coding DNA sequences**
* **1-Cis- and trans-regulatory elements**
  + - * ***cis*-acting elements Included :-** Promoters, enhancers, silencers, activators, insulators, locus control regions and MARs- matrix attachment regions.
        + ***trans*-acting elements.**
* While other transcriptional proteins which are formed from some genes are categorized into *trans*-elements. Which control the [transcription](https://en.wikipedia.org/wiki/Transcription_(genetics)) of a distant gene.
* **2—Introns**
* [](https://en.wikipedia.org/wiki/File:Pre-mRNA.svg)
* **Illustration of an unspliced pre-mRNA precursor, with five**[**introns**](https://en.wikipedia.org/wiki/Intron)**and six**[**exons**](https://en.wikipedia.org/wiki/Exon)**(top). After the introns have been removed via splicing, the mature mRNA sequence is ready for translation (bottom).**
* **3-Pseudogenes**
* [Pseudogenes](https://en.wikipedia.org/wiki/Pseudogene) are DNA sequences, related to known [genes](https://en.wikipedia.org/wiki/Gene), that have lost their protein-coding ability or are otherwise no longer [expressed](https://en.wikipedia.org/wiki/Gene_expression) in the cell
* **4-Repeat sequences, transposons and viral elements.**
* [**Transposons**](https://en.wikipedia.org/wiki/Transposon)**and**[**retrotransposons**](https://en.wikipedia.org/wiki/Retrotransposon)**are**[**mobile genetic elements**](https://en.wikipedia.org/wiki/Mobile_genetic_elements). Retrotransposon [repeated sequences](https://en.wikipedia.org/wiki/Repeated_sequence_(DNA)), which include: -
  + - [long interspersed nuclear elements](https://en.wikipedia.org/wiki/Retrotransposon#LINEs) (LINEs) (e.g [L1 sequences](https://en.wikipedia.org/wiki/Alu_sequence))
    - [Short interspersed nuclear elements](https://en.wikipedia.org/wiki/Retrotransposon#SINEs) (SINEs), account for a large proportion of the genomic sequences in many species. (e.g [Alu sequences](https://en.wikipedia.org/wiki/Alu_sequence)), are the most abundant mobile elements in the human genome
* **5-Telomeres**
* [Telomeres](https://en.wikipedia.org/wiki/Telomere) are regions of repetitive DNA at the end of a [chromosome](https://en.wikipedia.org/wiki/Chromosome), which provide protection from chromosomal deterioration during [DNA replication](https://en.wikipedia.org/wiki/DNA_replication). Recent studies have shown that telomeres function to aid in its own stability.
* 6-Junk DNA
* The term "junk DNA" became popular in the 1960s. According to [T. Ryan Gregory](https://en.wikipedia.org/wiki/T._Ryan_Gregory), the nature of junk DNA was first discussed explicitly in 1972 by a genomic biologist, David Comings, who applied the term to all non-coding DNA.



**Mb** (= Mbp) = mega base pairs = 1,000,000 bp.



(20–100 bp)

2-12 bp

Tandem repeats
Microsatellite DNA
• Unit - 2-4 bp (most 2).
• Repeat - on the order of 10-
100 times.
• Location - General...

Interspersed Repetitive DNA
• Interspersed repetitive DNA accounts for 25–40 % of mammalian DNA.
• They are scattered rand...