Dr. Harsha Dangare

Associate professor Pathology, MIMER medical college, Talegaon Dabhade.

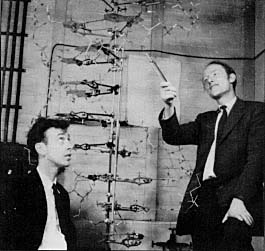
Mobile. No. 7769990125

Email- [dr.harshahj@gmail.com](mailto:dr.harshahj@gmail.com)

Chapter Name- Basics of human genome

**Deoxyribonucleic acid (DNA) is molecule which stores genetic information, f**ound in the nucleus of the cell. It is packed in a nucleus in a coiled manner, DNA in one human nucleus, if stretched will be around 2 meters long. James Watson and Francis Crick explained the molecular structure of DNA in 1953 using X-ray diffraction data collected by Rosalind Franklin and Maurice Wilkins, and model building techniques advocated by Linus Pauling [ 1, 2 ] .

Watson and Crick proposed the double helix: a twisted, spiral ladder structure consisting of two long chains wound around each other and held together by hydrogen bonds.

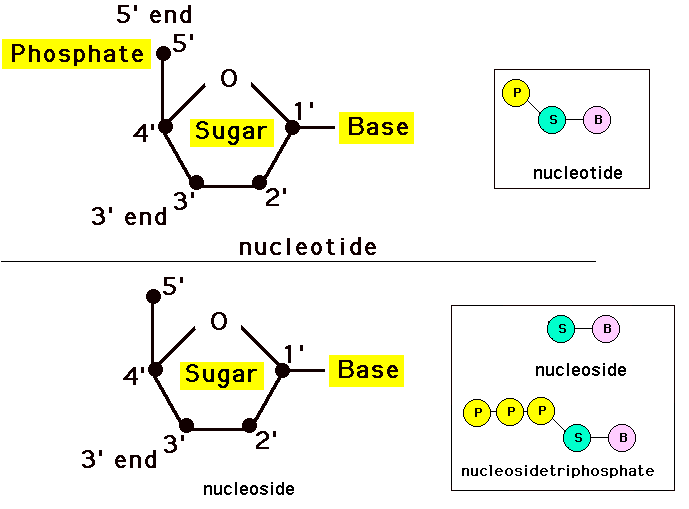


JAMES

WATSON

FRANCIS

CRICK

DNA is composed of repeating units—the nucleotides. Each nucleotide consists of a deoxyribose sugar, a phosphate group, and one of four nitrogen-containing bases: adenine (A), guanine (G), cytosine (C), or thymine (T). Adenine and guanine are purines with a double-ring structure, whereas cytosine and thymine are smaller pyrimidine molecules with a single ring structure. Two nitrogenous bases positioned side by side on the inside of the double helix form one rung of the molecular ladder. The sugar and phosphate groups form the backbone or outer structure of the helix. The fifth (5 C) carbon of one deoxyribose molecule and the third (3 C ) carbon of the next deoxyribose are joined by a covalent phosphate linkage. This gives each strand of the helix a chemical orientation with the two strands running opposite or antiparallel to one another.

* 4 nitrogenous bases are as pyrimidines- single carbon nitrogen ring ex. cytosine(C), thymine(T) / uracil (U) purines- double carbon nitrogen rings i.e. adenine(A), guanine(G).

Fig1.2: Nucleotide

* **UNIT** OF LENGTH OF DNA IS BASE PAIR [ bp ] .
* 1KILOBASE [ Kb ] = 1000 bp
* 1MEGABASE [ Mb ] = 1,000,000 bp(1million )
* TOTAL NUMBER OF BASE PAIRS IN A HAPLOID SET OF HUMAN CHROMOSOMES IS 3 X 10 9

Biochemical analyses performed by Erwin Chargaff showed that the concentrations of guanine and cytosine were always equal, as were the concentrations of adenine and thymine.

This finding became known as Chargaff’s rule. Watson and Crick postulated that in order to fulfill Chargaff’s rule and to maintain a uniform shape to the DNA molecule, there must be

a specific complementary pairing of the bases: adenine must always pair with thymine, and guanine must always pair with cytosine. Each strand of DNA, therefore, contains a nucleotide

sequence that is complementary to its partner. The linkage of these complementary nitrogenous base pairs holds the antiparallel strands of DNA together. Two hydrogen bonds

link the adenine and thymine pairs, whereas three hydrogen bonds link the guanine and cytosine pairs (Fig. 1.3 ). The complementarity of DNA strands is what allows the molecule

to replicate faithfully. The sequence of bases is critical for DNA function because genetic information is determined by the order of the bases along the DNA molecule.

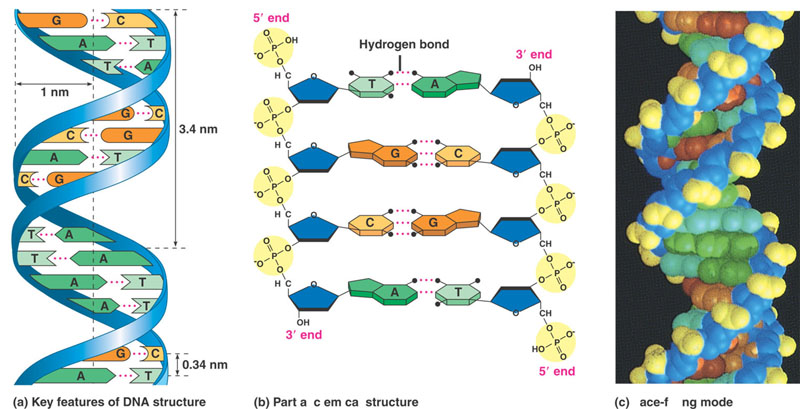
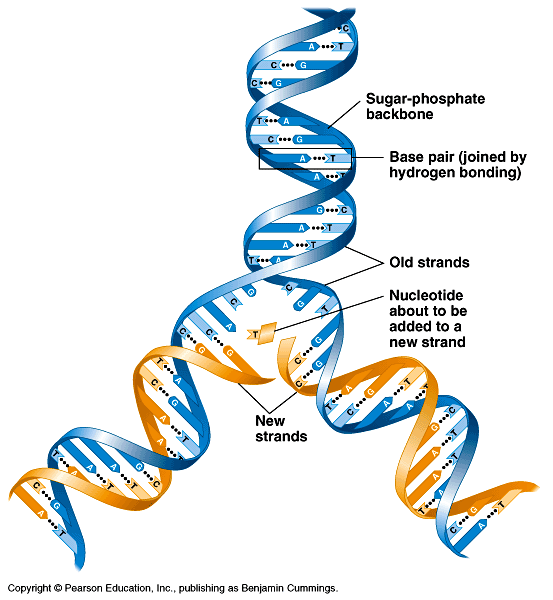


Fig 1.3: DNA double helix

Because of the base-pairing rules, DNA can replicate itself and preserve the sequence of bases. DNA is classified into three general categories: unique sequence, highly repetitive sequence DNA (>105 copies), and middle repetitive sequence DNA (102–104 copies).

Unique sequence or single-copy DNA is the most common class of DNA, comprising about 75% of the human genome [ 13 ] . This DNA consists of nucleotide sequences that are

represented only once in a haploid set. Genes that code for proteins are single-copy DNA. Repetitive or repeated sequence DNA makes up the remaining 25% of the genome and is classified according to the number of repeats and whether the repeats are tandem or interspersed among unique sequence DNA.

 Fig. The DNA double helix and its replication

**GENES**

A segment of DNA that encodes (contains the plans for) a protein.

Gene is the basic physical and functional unit of inheritance. It act as instructions to make molecules called proteins. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. It contains the code for the amino acid sequence of a polypeptide chain and the regulatory sequences necessary for expression. Total number of genes in humans (~30,000). Each gene can form more than one type of proteins. Separated by intergenic DNA – about 70% of the genome.

It is capable of producing some element of biological function, e.g.,

- a readily observable trait, like skin color,

- a cellular property, e.g., length of the cell cycle,

- a molecular biological property, like the three dimensional shape of a protein

Every person has two copies of each gene, one inherited from each parent. Most genes are same in all people, but a small number of genes (< 1%of total) are slightly different between people. Each member of a pair of genes is an allele. Alleles may be identical, or may differ from each other.

* **Allele is an alternative form of a gene (one member of a pair) that is located at a specific position on a specific chromosome**
* **Homozygous** –both alleles at a locus are the same. (In the ABO system, an AA complement represents homozygosity.)
* **Heterozygous –**the two alleles at a locus are different. (In the ABO system, an AO complement represents heterozygosity.)

Different forms of a gene can create proteins with different amino acid sequences. These proteins may have different structures and functions. In this way, different phenotypes stem from different forms of a gene. An example – For the gene that determines one of the polypeptide chains in hemoglobin, one allele determines the standard form of the polypeptide and another allele determines a form of the polypeptide that causes sickle cell anemia.

**Parts of gene**

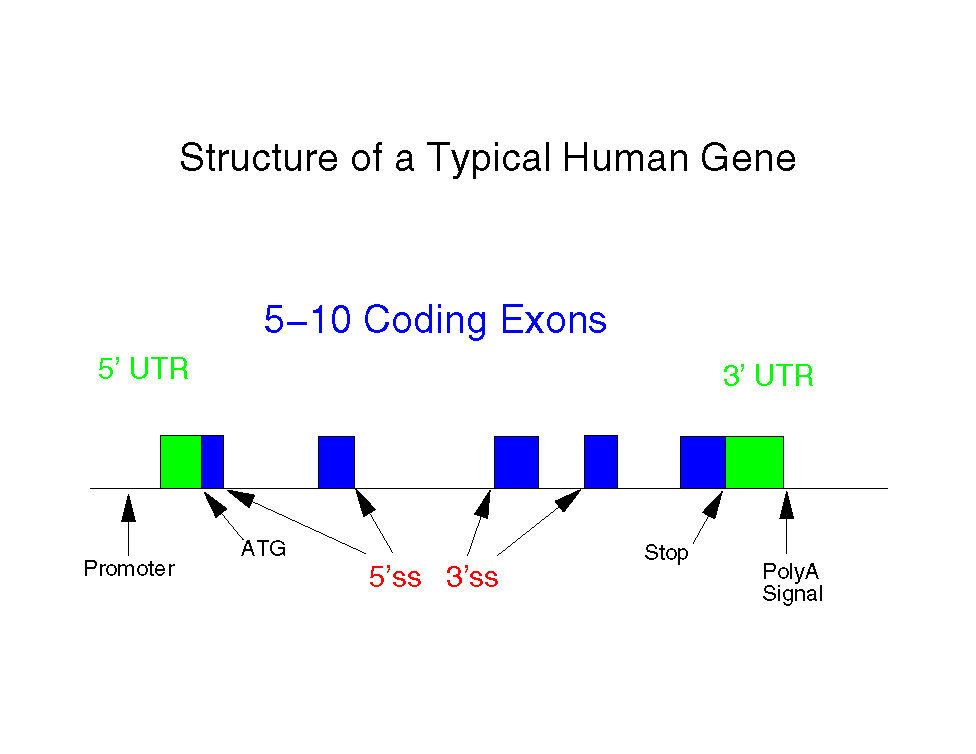
Genes are made of parts represented in the mRNA (exons) and parts that are transcribed but not present in the mRNA (introns). Coding sequences as well as adjacent nucleotide sequences required for proper expression of a gene Fundamental components

Figure .4 Parts of gene

a) Promoter region "upstream" of initiation site

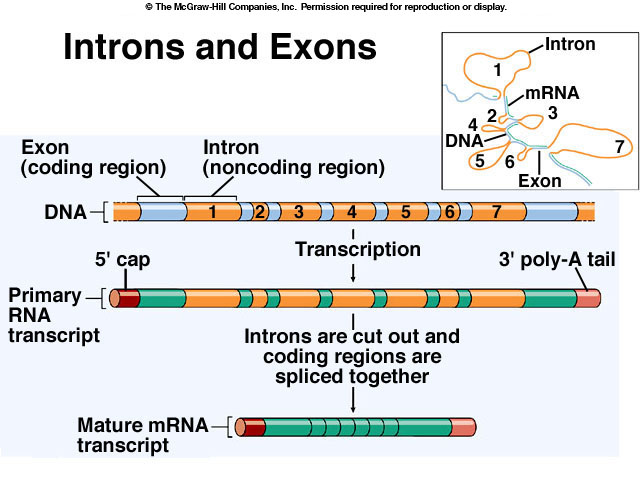
b) Initiation site for transcription

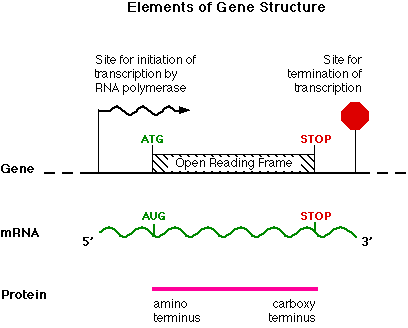
c) Coding region (exon) of structural gene

d) Termination region--halts polymerase from transcribing

e) Introns - non coding sequences

* 5’ end – cap
* 3’ end – polyadenylate tail

****

**a) Promoter region "upstream" of initiation site. It is**

**i) Necessary binding site for RNA polymerase to accomplish translation**

**ii) Bears recognition sequences for enzyme (e.g., TTTA)**

**b) Initiation site for transcription**

**--yields ribosomal binding site in mRNA**

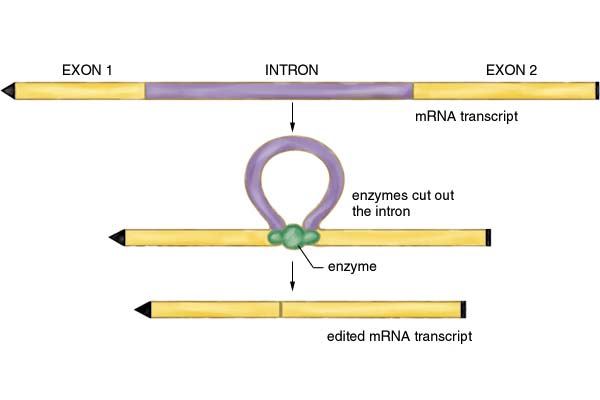
**c) Coding region (exon) of structural gene**

**i) Composed of codons (triplets) of nucleotides**

**ii) Begins with start codon (e.g., TAA)**

**iii) Ends with stop codon**

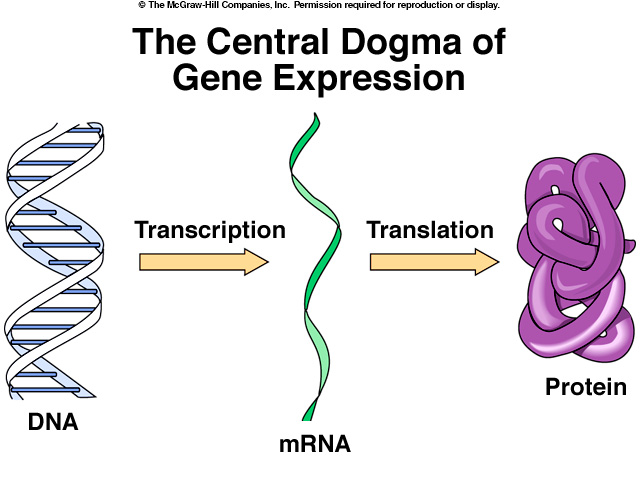
**iv) Codons complementary to mRNA codons-- >amino acids in ultimate protein chain**

****

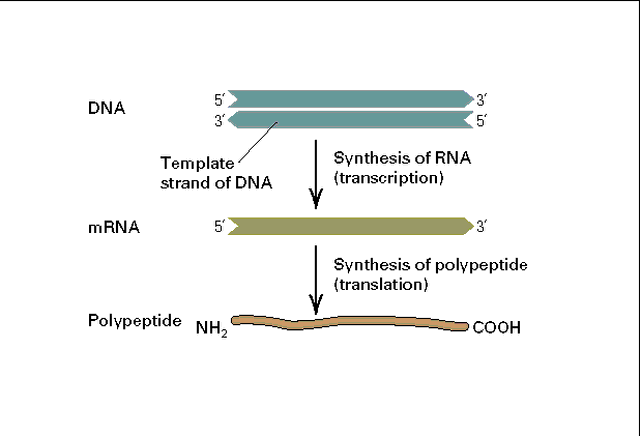
**Fig.1.6 Introns are Removed and Exons are Spliced to Make mRNA**

**Molecular Genetics - From DNA to Phenotype**

**DNA→ RNA → Protein**

****

**Fig. Central Dogma -- DNA→ RNA → Protein**

****

**Transcription - Making an RNA Copy of DNA**

**The decision to transcribe or not transcribe a gene is the most important step in the control of gene expression. It is controlled by promotor region methylation and acetylation**

**GENE expression and regulation**

**It is the phenotypic manifestation of genes by the processes of transcription and translation. Gene expression is a complex process and highly regulated. Regulation occurs at several points during transcription and translational processes.**

**Gene regulation is the process of turning genes on or off.**

Regulation of gene expression controls the amount and appearance agenda of a gene's functional product. All steps of gene expression can be modulated. Regulation of gene expression is the basis for cell differentiation, diversity, and adaptation

1. Cis-action factors are short regulatory sequences located within the promoter or in the vicinity of a gene's structural portion. Cis-sequences facilitate the transcription of adjacent polypeptide-encoding sequences.
2. *Trans-action* factors bind to the cis-acting sequences to control gene expression
3. Enhancer is a short region of DNA that upregulates transcription levels of genes. Enhancer sequences are active when bound to *trans-action* factors.

Response element is a short sequence of DNA within the promoter of a gene that can bind to a specific hormone receptor complex and regulate transcription of genes subject to that hormone.

**Signal Transduction**

A signal transduction pathway is a sequence of enzymes and second messengers by which a receptor communicates with the cell nucleus. The signal transduction pathway "translates" the receptor ligand message at the surface into a cellular response in the nucleus ex.*- abl* and *ras* are signal transducers

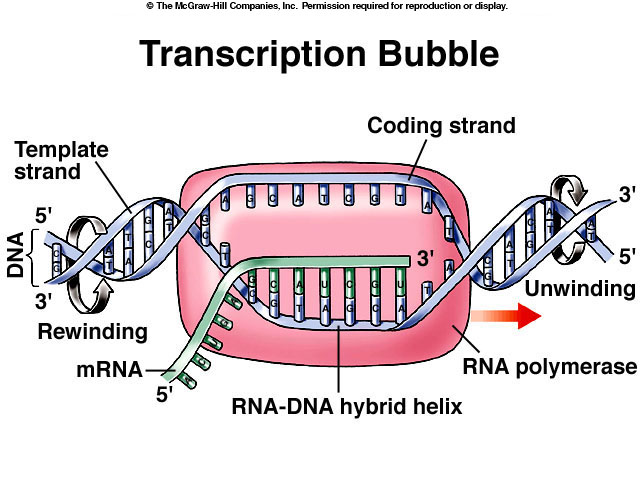
Transcription factors

A transcription factor is a molecule that initiates transcription of DNA in the eukaryotic nucleus. Transcription factors interact with promoter or enhancer sequences either by binding directly to DNA or by interacting with other DNA-bound proteins ex.  *myc* is a transcription factor that activates expression of many genes by binding to consensus sequences

Programmed cell death regulators are molecules that prevent apoptosis . Activation of these regulators leads to overgrowth of abnormal cells

*Ex. bcl-2* is a programmed cell death regulator that governs mitochondrial outer membrane

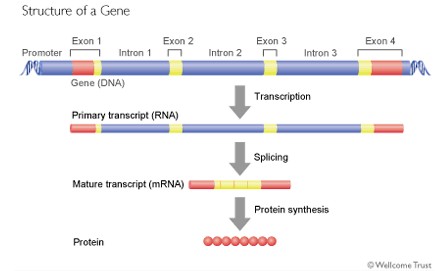
permeability and suppresses apoptosis

****

Transcription starts and stops at distinct sites in a gene. Splicing of one gene can occur in more than one way (alternative splicing). Some genes produce many related proteins by alternative splicing. This means that the count of our genes (~35,000) seriously underestimates the count of our different proteins.

**Translation** is the process that uses the information in mRNA to direct the synthesis of proteins. Translation (protein synthesis) requires:

1. mRNA (provides information about what protein to make).
2. A ribosome (the protein synthesizing “factory”).
3. tRNA (an adaptor that couples mRNA and amino acids).

****

**Overview**

A chromosome is an enormous macromolecule into which somatic DNA is packaged in eukaryotic cells. Three billion base pairs of nucleotides (a complete set of DNA) are

divided among 46 chromosomes, each containing many genes, regulatory elements, and intervening nucleotide sequences

Chromosomes are found only in the eukaryotic nucleus and can be seen only during nuclear division. During most of the life cycle, the genetic material occupies areas of nuclei in the form of chromatin, and individual chromosomes cannot be distinguished.

In eukaryotes, the basic function of the chromosome is to package and compress the DNA, exposing specific genes for transcription during certain phases of the cell life span.

**Chromatin**

Chromatin is the form of genetic material existing during interphase of eukaryotic cells and is made up of DNA and protein. Chromatin can be seen with the light microscope after

staining with nuclear stains

• Chromatin is a packaged state of DNA in a small volume to strengthen the DNA, to allow mitosis and meiosis, and to serve as a mechanism for expression control.

• Chromatin functions as a gene regulator. The changes in chromatin structure are effected mainly by methylation (DNA and proteins) and acetylation (proteins).

Euchromatin is a loosely packed form of chromatin that is involved in active transcription or regulation and is lightly stained by nuclear stains. Heterochromatin is a darkly stained and tightly packed form of DNA. Its major biologic characteristic is that it is not transcribed

• Chromatin composition- Histones are the major chromatin binding proteins.

They act as spools around which DNA winds

• Histones plays role in gene regulation

• Histones H2A, H2B, H3, and H4 form octamers (two of each) with a cylindrical shape.

There are two fundamental types of chromatin in eukaryotic cells: euchromatin and heterochromatin. Euchromatin is loosely organized, extended, and uncoiled. This chromatin

contains active, early replicating genes, and stains lightly with GTG-banding techniques.

There are two special types of heterochromatin that warrant special mention: facultative heterochromatin and constitutive heterochromatin. Both are genetically inactive, late

replicating during the synthesis (S) phase of mitosis, and are highly contracted.

Fig1.9: Nucleosomes are the fundamental repeating subunits of all eukaryotic chromatin. They package DNA into chromosomes inside the cell nucleus and control gene expression. The DNA winding around the nucleosome core particle consists of about 146 bp of dsDNA wrapped in 1.65 left- handed super helical turns around complexes of the four histone proteins known as the histone octamers. The DNA hanging between two nucleosome cores is typically 55-bp long and is known as linker DNA

When DNA winds 1.65 times around a histone octamer a nucleosome results

• Each nucleosome contains 146 bp

• The nucleosomes are stacked and further coiled into a 30-nm fiber, which makes up the chromosome residing in the cell nucleus



• Linker DNA is the DNA hanging between two nucleosomes, typically 55-bp long

• Histone-DNA interaction regulates gene expression. Acetylation of histone modulates gene

expression and leads to transcription activation. The extent of interaction between histone and

DNA is affected by the degree of histone acetylation

• Histone acetylation is the process in which charged lysine side chains are acetylated, leading to reduced affinity between histone and DNA. After acetylation, RNA polymerase and transcription factors have better access to the promoter.

Chromosome structure - Each chromosome has two short arms (p), two long arms (q), one centromere, and four telomeres

• The centromere is the constricted region of a chromosome, which has a special sequence and structure for attachment to the spindle filament during M-phase and for separation of chromosomes during mitosis.

• The centromere divides the chromosome into four arms

• The two equal short arms are designated "p"*(petite)*

• The two equal long arms are designated "q" (follows p in the Latin alphabet)



Chromosome grouping

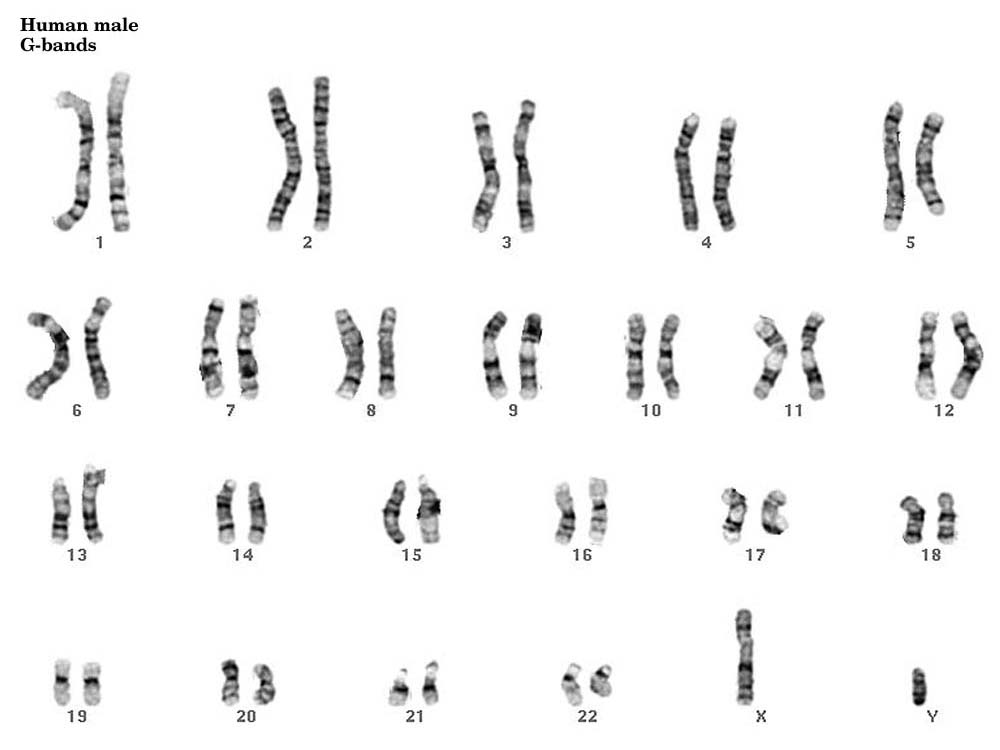
Chromosomes are numbered and grouped according to their morphologic characteristics

Chromosomes are numbered according to their relative sizes from largest to smallest. The position of the centromere determines chromosome grouping

• Group A have nearly equal p and q arms whereas group E have the centromere almost at the

Telomere

Chromosome identification is confirmed by the banding pattern unique to each chromosome (Figure



• Chromosome number

Human cells contain 46 chromosomes (23 from each parent) including 22 pairs of autosomes and one pair of sex chromosomes.