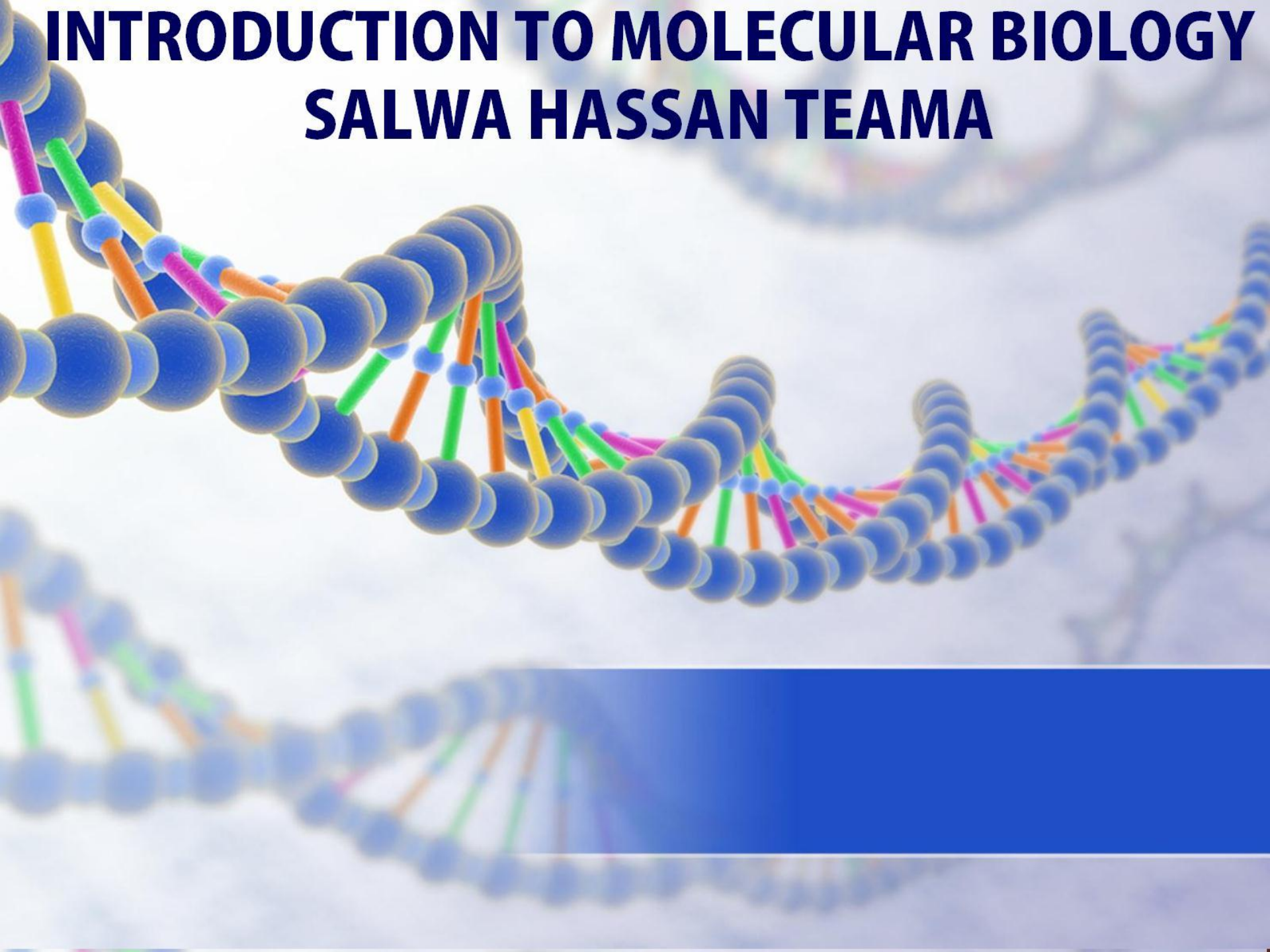
INTRODUCTION TO MOLECULAR BIOLOGY

SALWA HASSAN TEAMA



MOLECULAR BIOLOGY

SALWA HASSAN TEAMA M.D.

Contents

Target Audience

- Physician assistant;
- Postgraduate in clinical specialties;
- Medical students;
- Medical technologist;
- Beginners and
- For every laboratory worker and everyone passion for learning.

Molecular Biology

- Molecular biology; the branch of biology that study gene structure and function at the molecular level.

- The Molecular biology field overlaps with other areas, particularly genetics and biochemistry.

- The Molecular biology allows the laboratory to be predictive in nature; events that occur in the future.

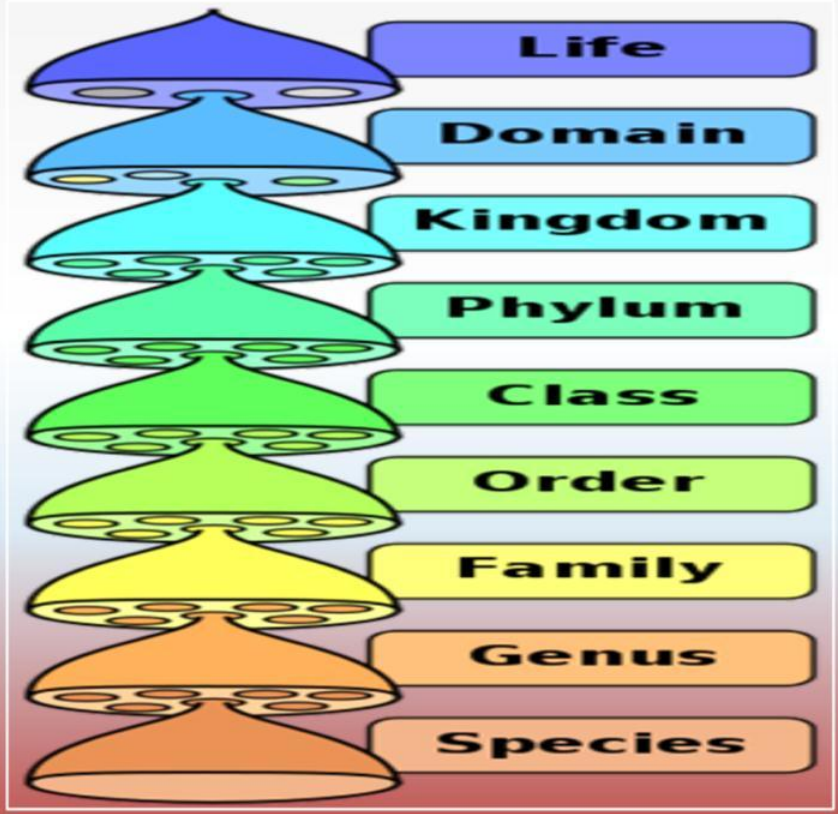
The Genome Database

The genome database is organized in six major organism groups:

- Eukaryotes,
- Bacteria,
- Archaea,
- Viruses,
- Viroids and
- Plasmids.

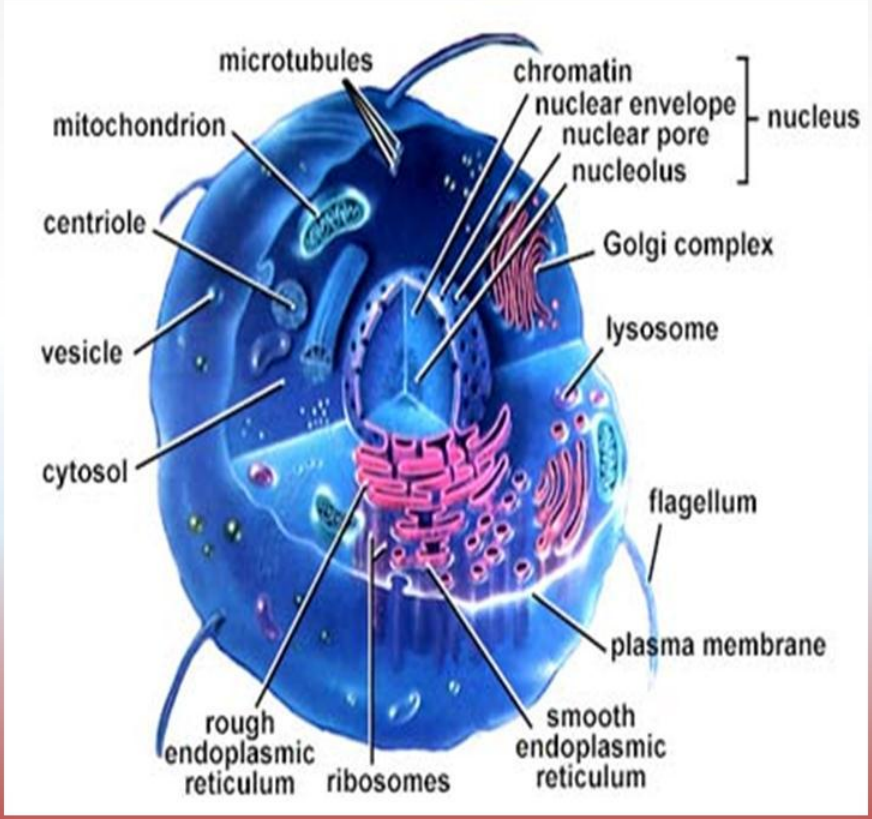
Three Domain of Life

- Eukaryotic
- Prokaryotic
- Archaea



Eukaryotic Cell

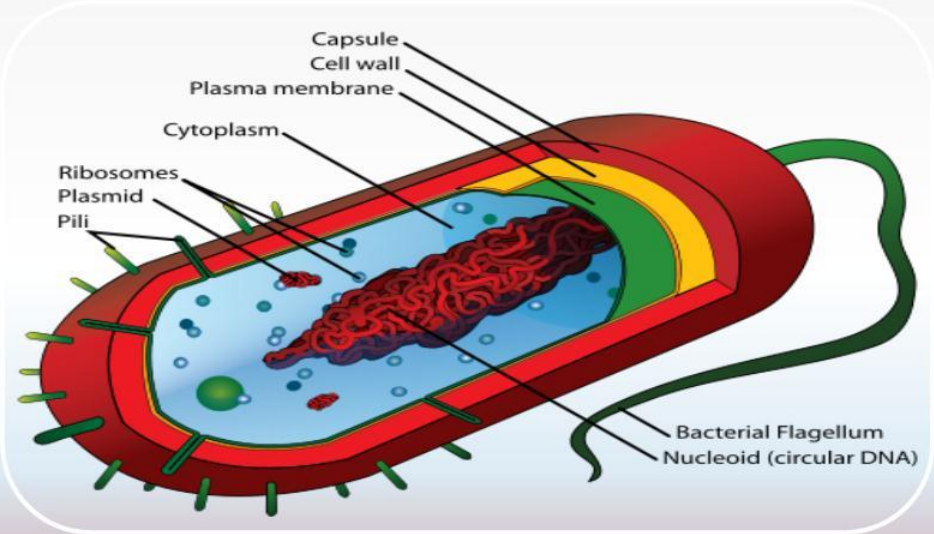
Eukaryotes are generally more advanced than prokaryotes



Eukaryotic Cell

- Eukaryotic cells are found in animals, plants, fungi and protists cell;
- Cell with a true nucleus, where the genetic material is surrounded by a membrane;
- Eukaryotic genome is more complex than that of prokaryotes and distributed among multiple chromosomes;
- Eukaryotic DNA is linear;
- Eukaryotic DNA is complexed with proteins called histones;
- Numerous membrane-bound organelles;
- Complex internal structure;
- Cell division by mitosis.

Prokaryotic Cell



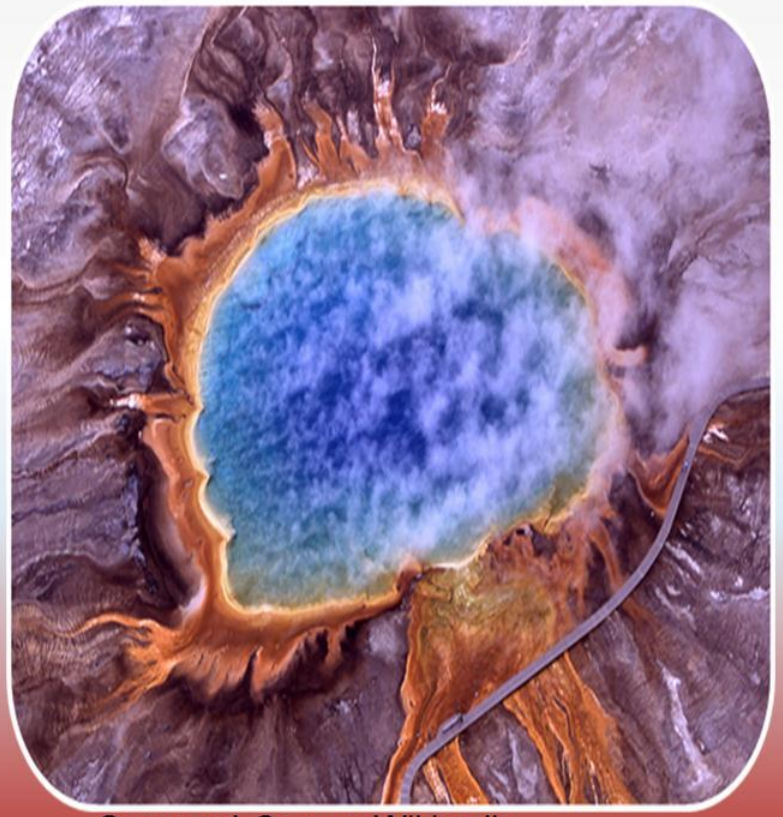
Mariana Ruiz Villarreal, LadyofHats. Source: Wikipedia

Prokaryotic Cell

- Unicellular organisms, found in all environments. These include bacteria and archaea;
- Without a nucleus; no nuclear membrane (genetic material dispersed throughout cytoplasm;
- No membrane-bound organelles;
- Cell contains only one circular DNA molecule contained in the cytoplasm;
- DNA is naked (no histone);
- Simple internal structure; and
- Cell division by simple binary fission.

Archaea

Archaea is prokaryotes; organisms without nucleus but some aspects of their molecular biology are more similar to those of eukaryotes.

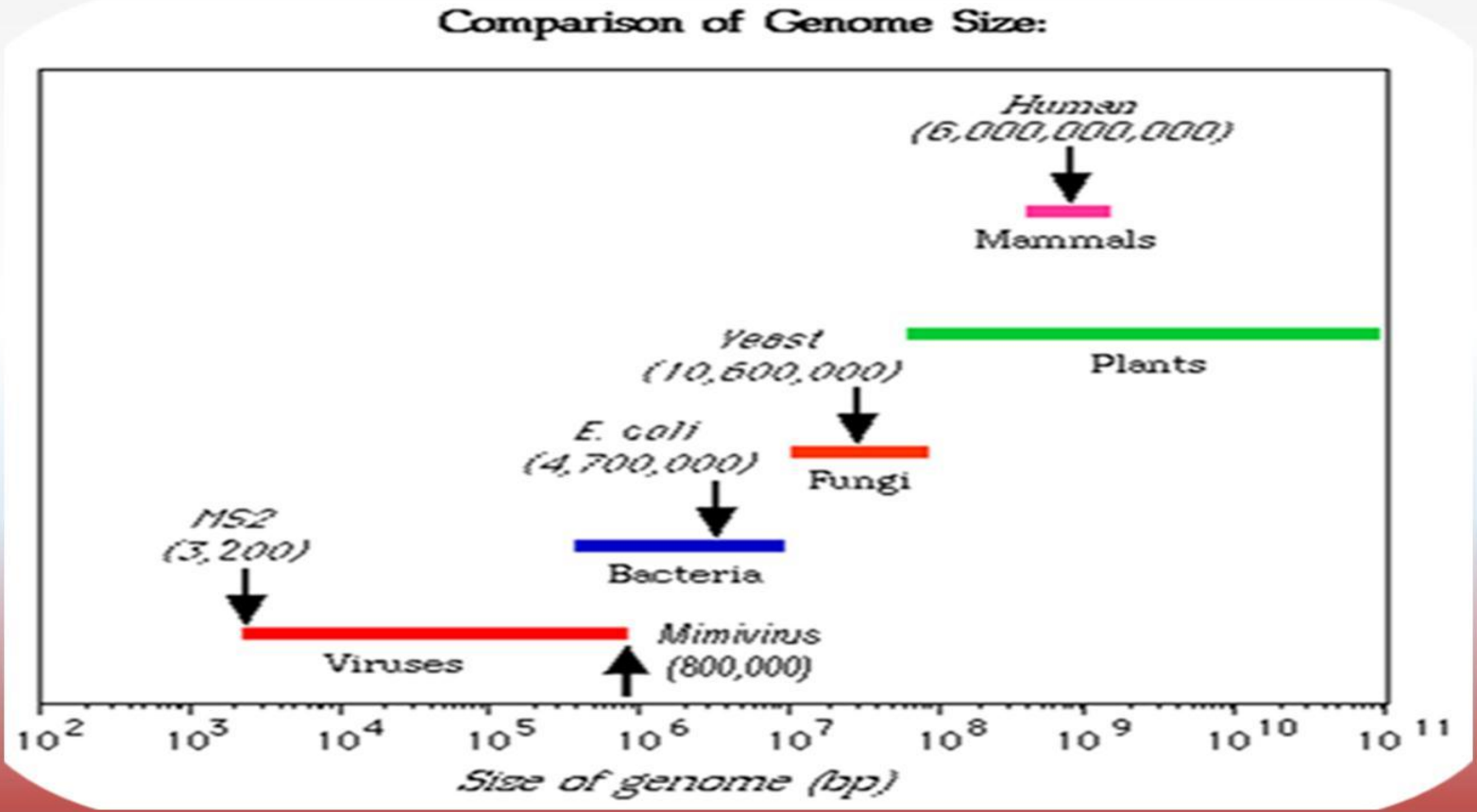


Saperaud. Source: Wikipedia

The Genome

- Totality of genetic information of an organism.
- Encoded in the DNA (for some viruses, RNA).

The Genome Size



Species/ Number of Chromosomes

Species	Number of chromosomes
Human	46
Mouse	40
Rat	42
Fruit flies	8
Bacteria	1

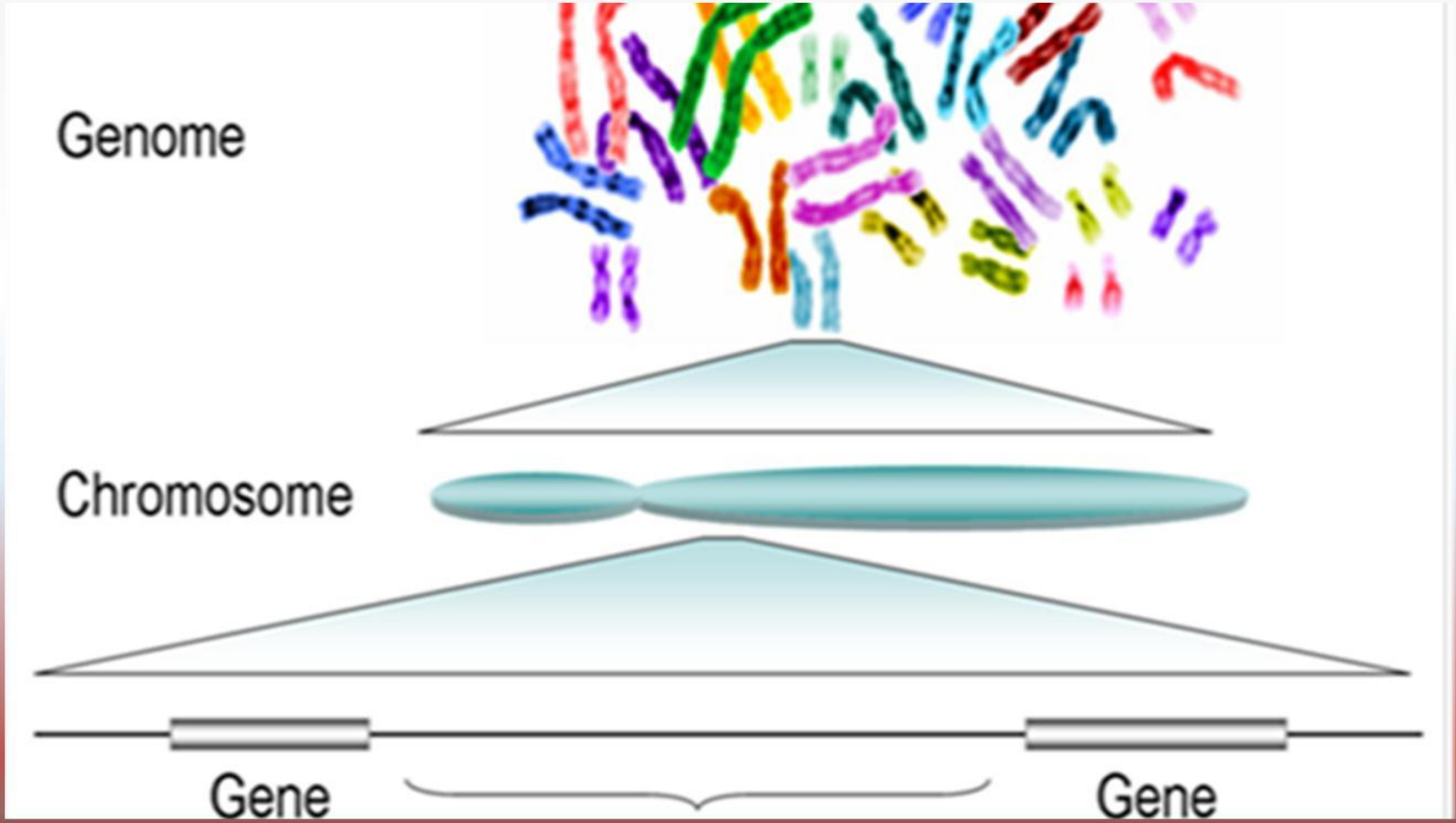
Human Genome

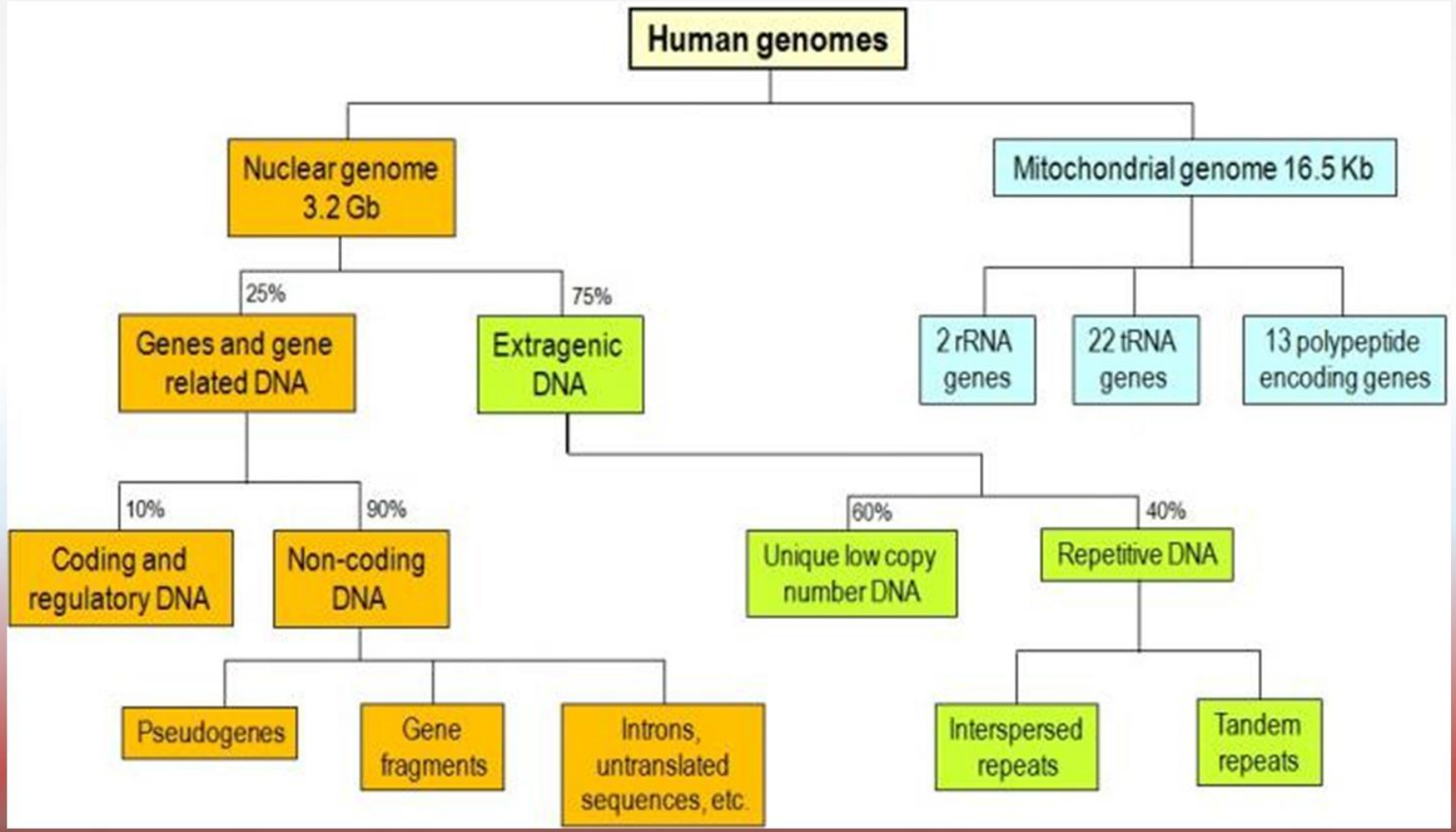
Human Genome; Arranged on multiple chromosomes; twenty three pairs of chromosomes;

- Twenty two pairs (autosomes).
- One pair (sex chromosome) (xx) (female) or (xy) (male).

Humans have 23 pairs of chromosome in every cell (except mature red blood cells..); Gametes or sex cells (sperm and eggs) have half the normal complement of chromosomes.

Human Genome





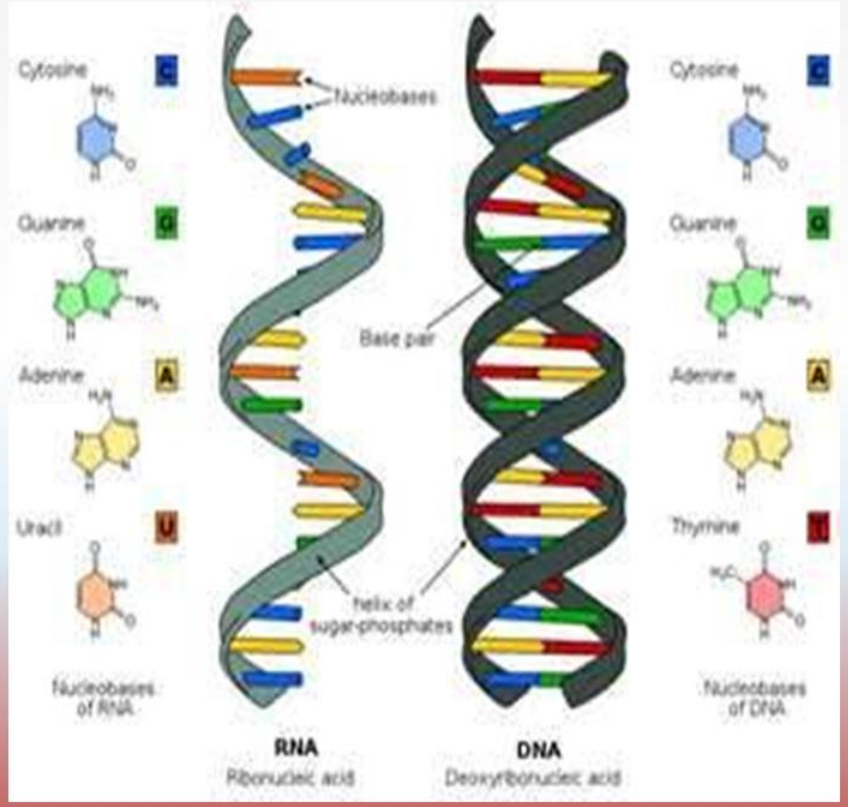
. Source: http://geneticssuite.net/node/33

Chromosome/ Amount of DNA

Chromosome	Amount of DNA (Mb)	Chromosome	Amount of DNA (Mb)
1	263	13	114
2	255	14	109
3	214	15	106
4	203	16	98
5	194	17	92
6	183	18	85
7	171	19	67
8	155	20	72
9	145	21	50
10	144	22	56
11	144	X	164
12			

General Structure of Nucleic Acid

DNA and **RNA** are long chain polymers of small chemical compound called nucleotides.



Nucleotides

Nucleotides; ring shaped structures composed of:

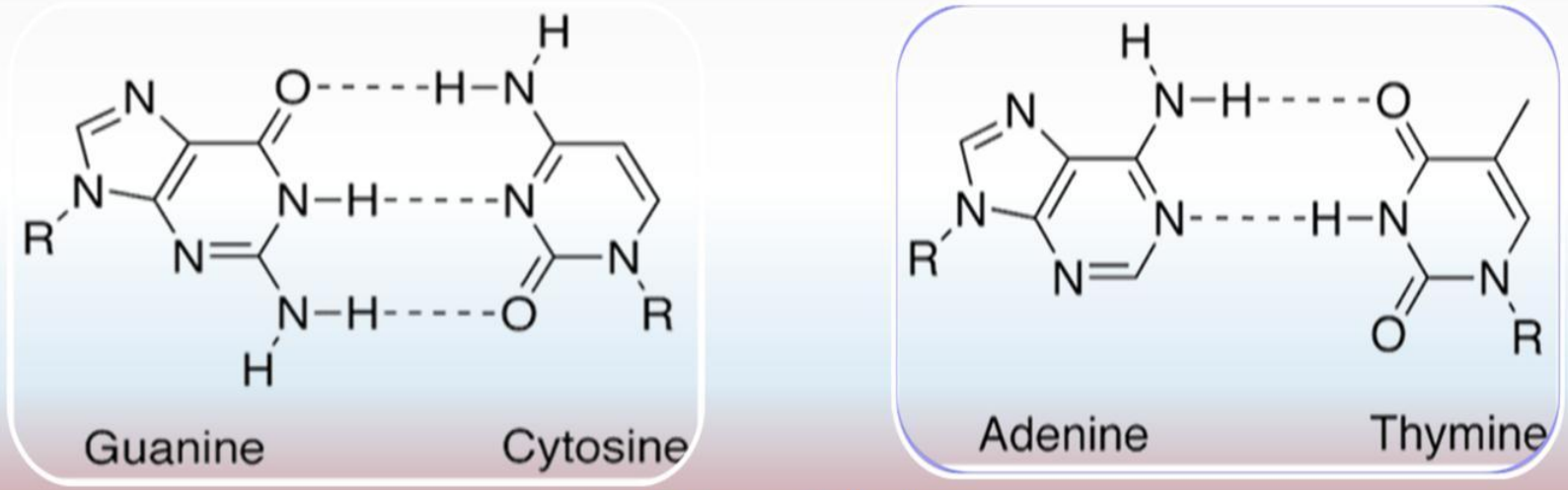
- Nitrogenous base; these bases are classified based on their chemical structures into two groups:
  - Purine; double ringed structure (Adenine and Guanine).
  - Pyrimidine; single ring structures (cytosine and thymine).
- Sugar
- Phosphate group

Nucleotides

- DNA: Four different types of nucleotides differ in nitrogenous base:
  - A is for adenine;
  - G is for guanine;
  - C is for cytosine and
  - T is for thymine.

- RNA: thymine base replaced by uracil base.

Nucleotides



Jypx35. Source: Wikipedia

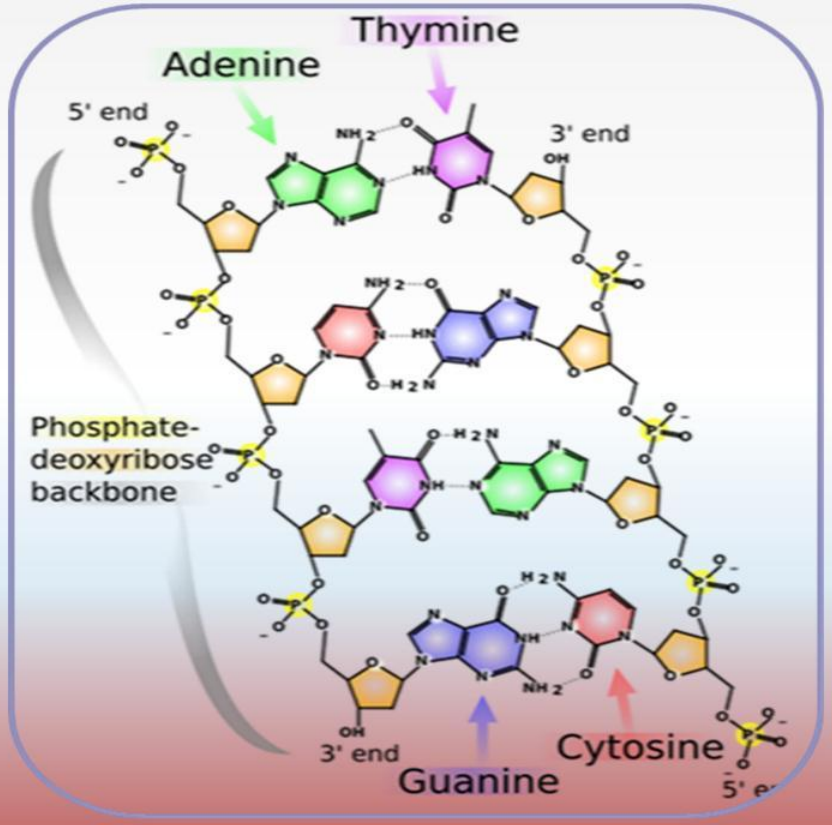
The DNA

- **Deoxyribonucleic Acid (DNA);** the genetic material of all cellular organisms and most viruses.
- **DNA**; the gigantic molecule which is used to encode genetic information for all life on Earth.
- A human cell contains about 2 meters of **DNA**. **DNA** in the body could stretch to the sun and back almost 100 times. So it is tightly packed.
- **DNA** responsible for preserving, copying and transmitting information within cells and from generation to generation.

DNA Double Helix

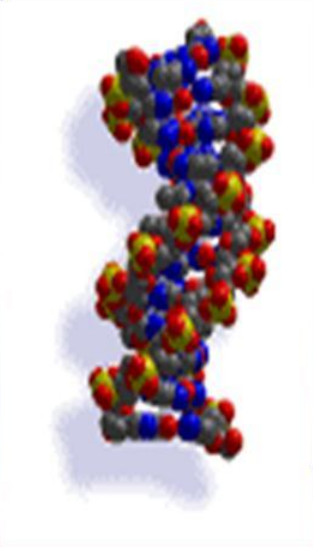
- Linked as a twisted ladder.
- The curving sides of the ladder represent the sugar-phosphate backbone of the two DNA strands; the rungs are the base pairs.
- Possess antiparallel polarity.
- Stabilized by hydrogen bonds between the bases.

DNA Double Helix



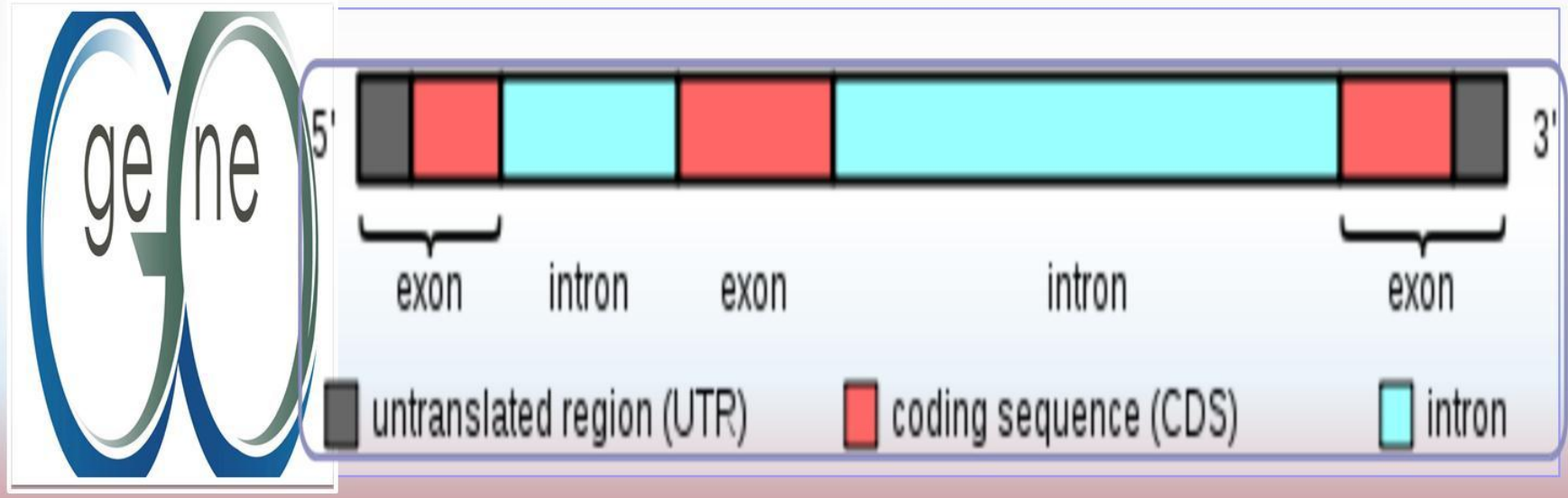
Madprime. Source: Wikipedia

DNA CONTENT

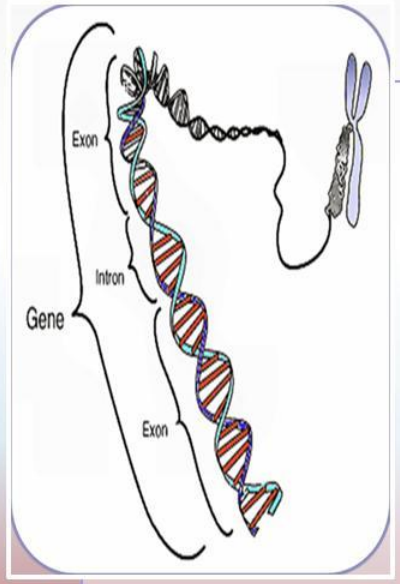


The molecule of DNA in a human chromosome ranges in size from 50 x 10^6 nucleotide pairs in the smallest chromosome up to 250 x 10^6 nucleotide pairs in the largest .

GENE STRUCTURE



THE GENE



The actual number of genes contained in the human genome is not known. However, it has been estimated that the human genome contains about 30.000 essential genes.

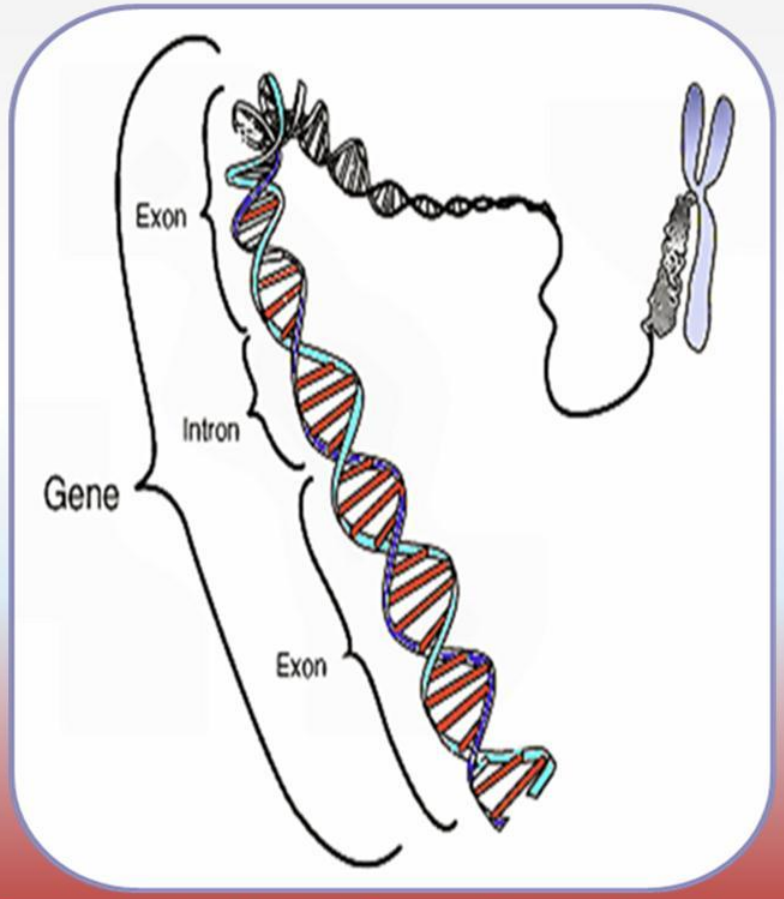
THE GENE

The 30.000 or more protein encoding genes are scattered among 3 billion DNA base pairs 6 x 10^9 (diploid genome) among chromosomes.

The Gene

- **The gene;** it is a segment within a very long strand of DNA.
- **Genes** are the basic units of hereditary.
- **Genes** located on chromosome on its place or locus.
- **Allele;** a variant of the DNA sequence at a given locus. Each allele inherited from a different parent.

The Gene

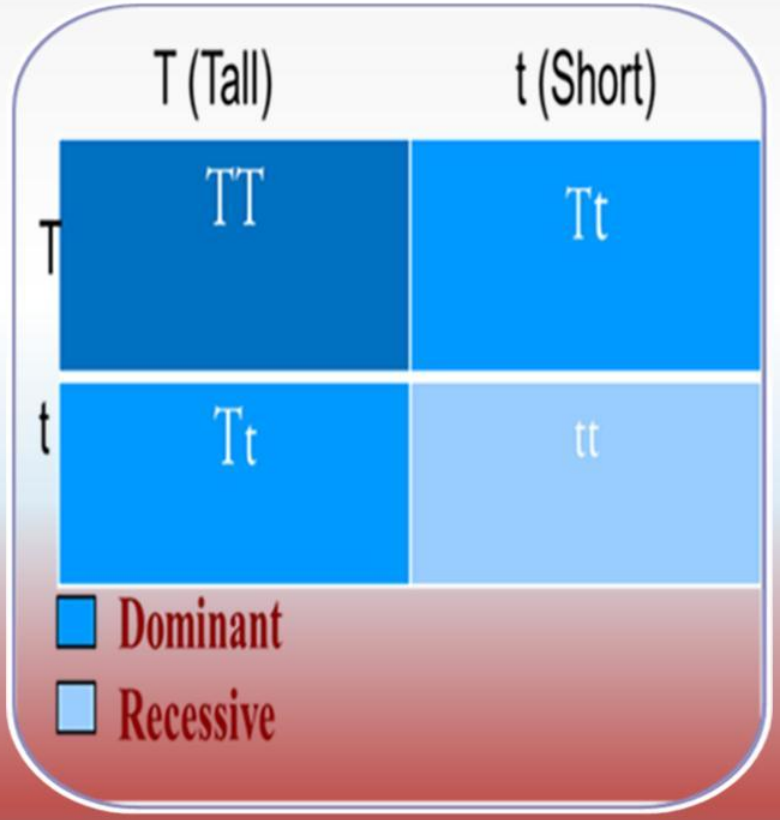


Source: National Human Genome Research Institute.

Dominant and Recessive

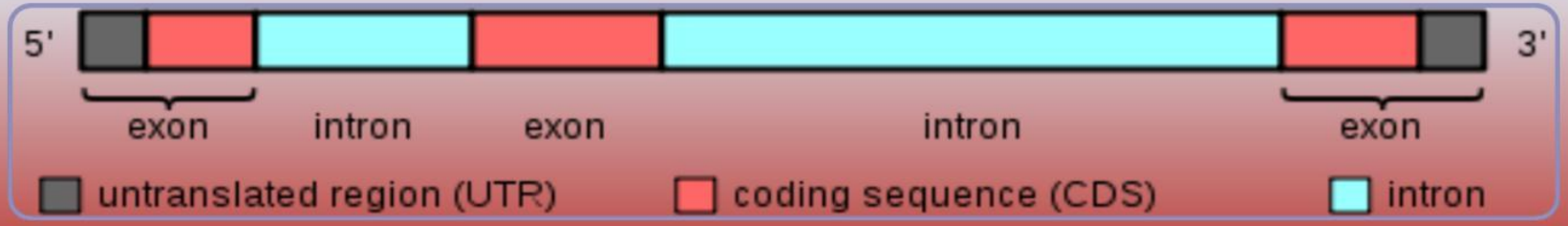
- **Dominant**
  - The one pair of allele that masks the effect of the other when present in the same cell.
- **Recessive**
  - The one pair of allele that is masked by the other when present in the same cell and capable of producing its characteristics phenotype in the organism only when two alleles is present and identical.

Dominant and Recessive



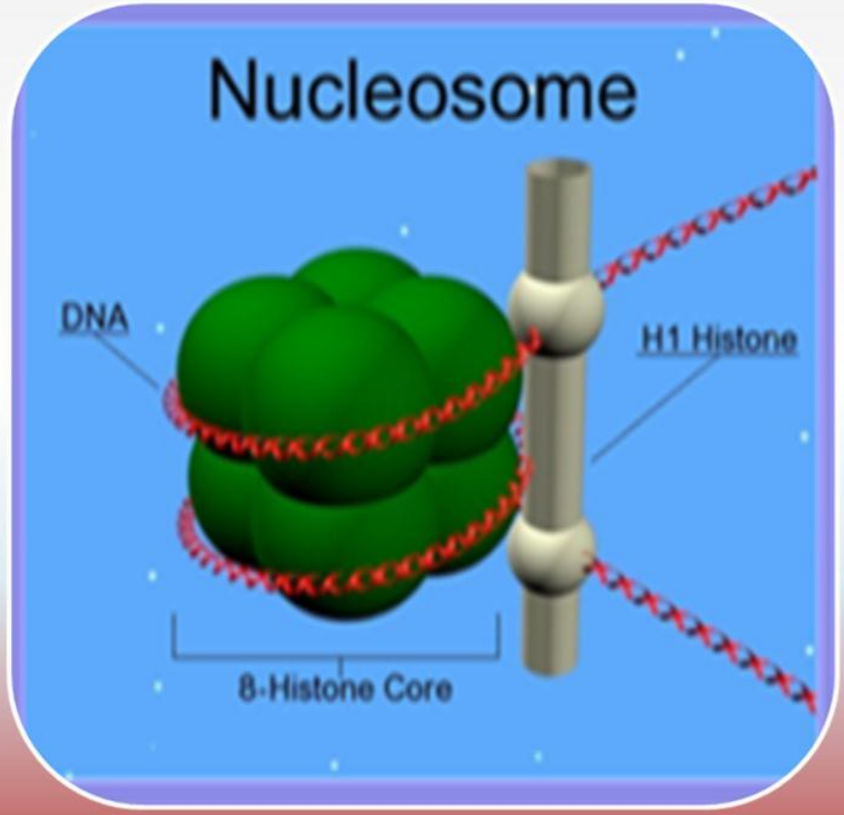
Gene Structure

- Most of the genes consist of; short coding sequences or exons are interrupted by a longer intervening noncoding sequence or introns; although a few genes in the human genome have no introns.

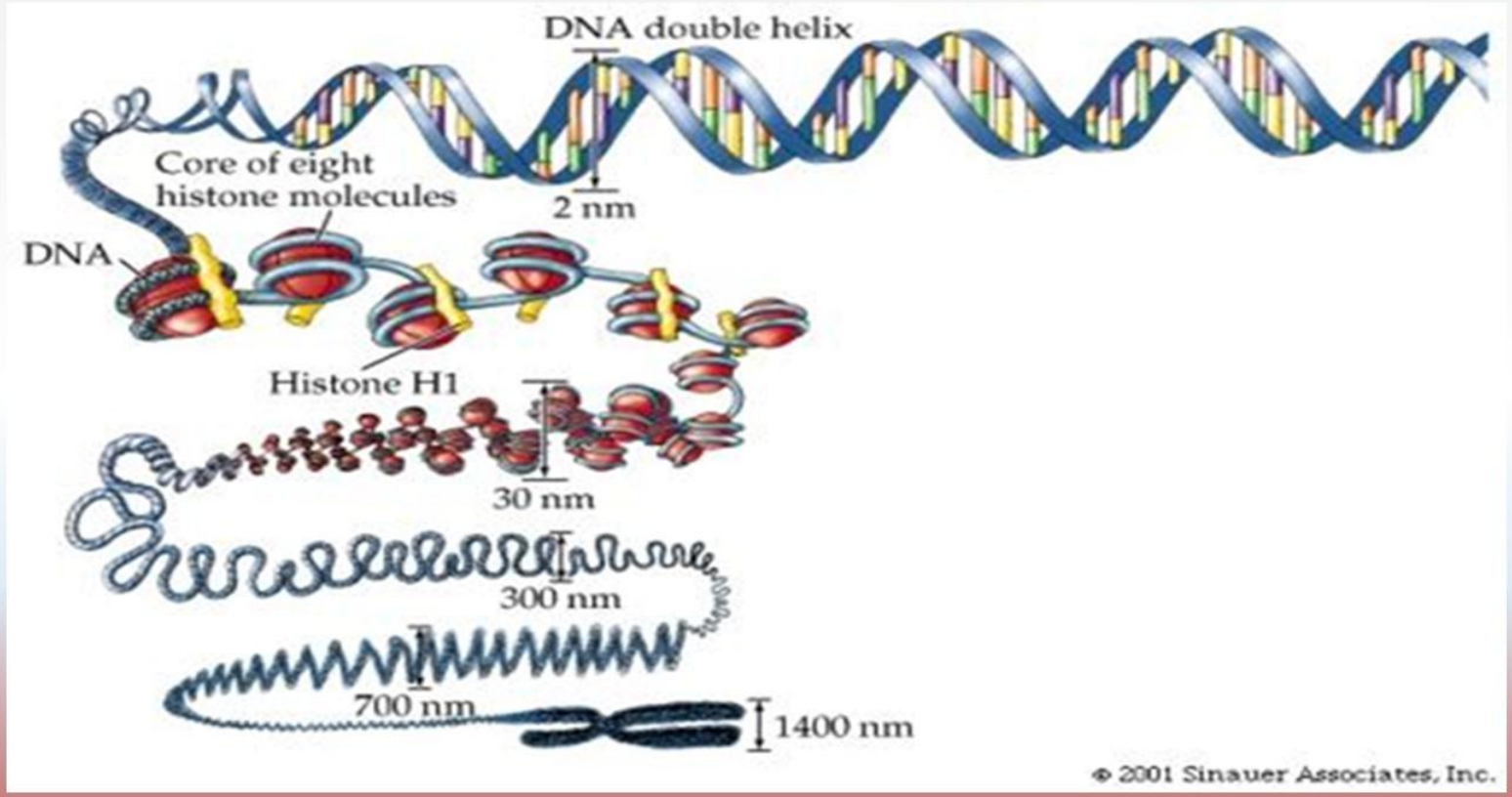


DNA Organization

DNA molecules complexed with other proteins, especially basic proteins called histones to form a substance known as chromatin.



DNA Organization



The RNA

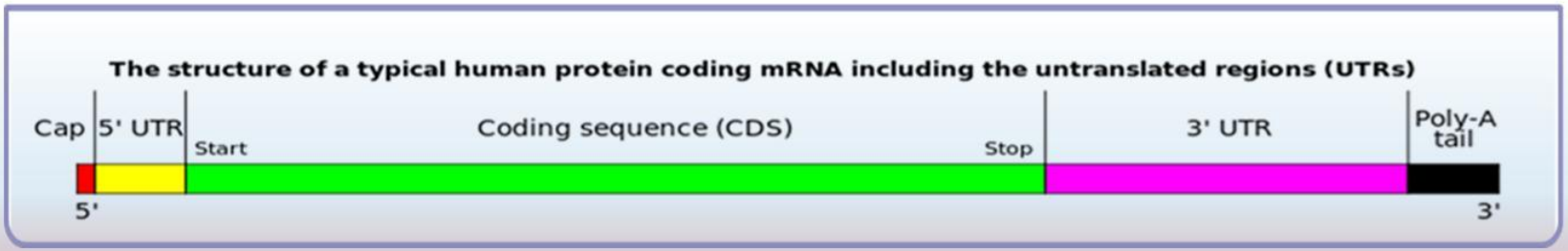
- Three major classes of RNA: messenger (mRNA), transfer (tRNA) and ribosomal (rRNA). Minor classes of RNA include small nuclear RNA; small nucleolar RNA;……

- RNA is a single stranded; the pyrimidine base **uracil** (U) replaces thymine and ribose sugar replaces deoxyribose.

Messenger RNA/ mRNA

- Transcripts of structural genes.
- Encode all the information necessary for the synthesis of a polypeptide of protein.
- The 5' terminus is capped by 7 methyguanosine triphosphate.
- Synthesis of the poly (A) tail involves cleavage of its 3' end and then the addition of about 200 adenine residues.
- Intermediate carrier of genetic information; deliver genetic information to the cytoplasm.

mRNA

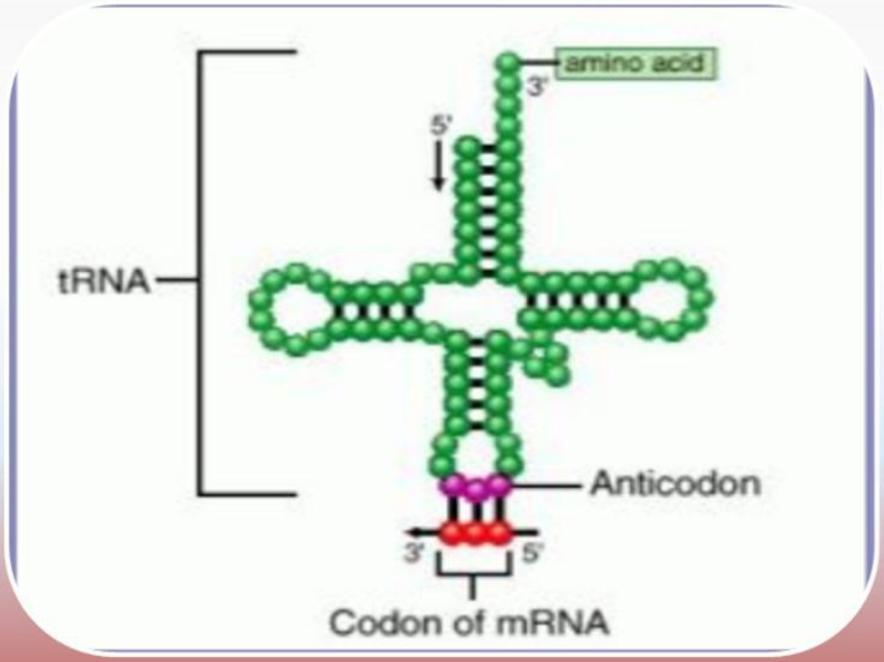

The structure of a typical human protein coding mRNA including the untranslated regions (UTRs)
Cap
5' UTR
Start
Coding sequence (CDS)
Stop
3' UTR
Poly-A tail
5'
3'

Transfer RNA/ tRNA

- All the tRNAs share a common secondary structure resembles a cloverleaf: They have four base- paired stems defining three stem-loops (the D loop, anticodon loop, and T loop) and the acceptor stem.

- tRNA carry correct amino acids to their position along the mRNA template to be added to the growing polypeptide chain.

tRNA



Ribosomal RNA/ rRNA

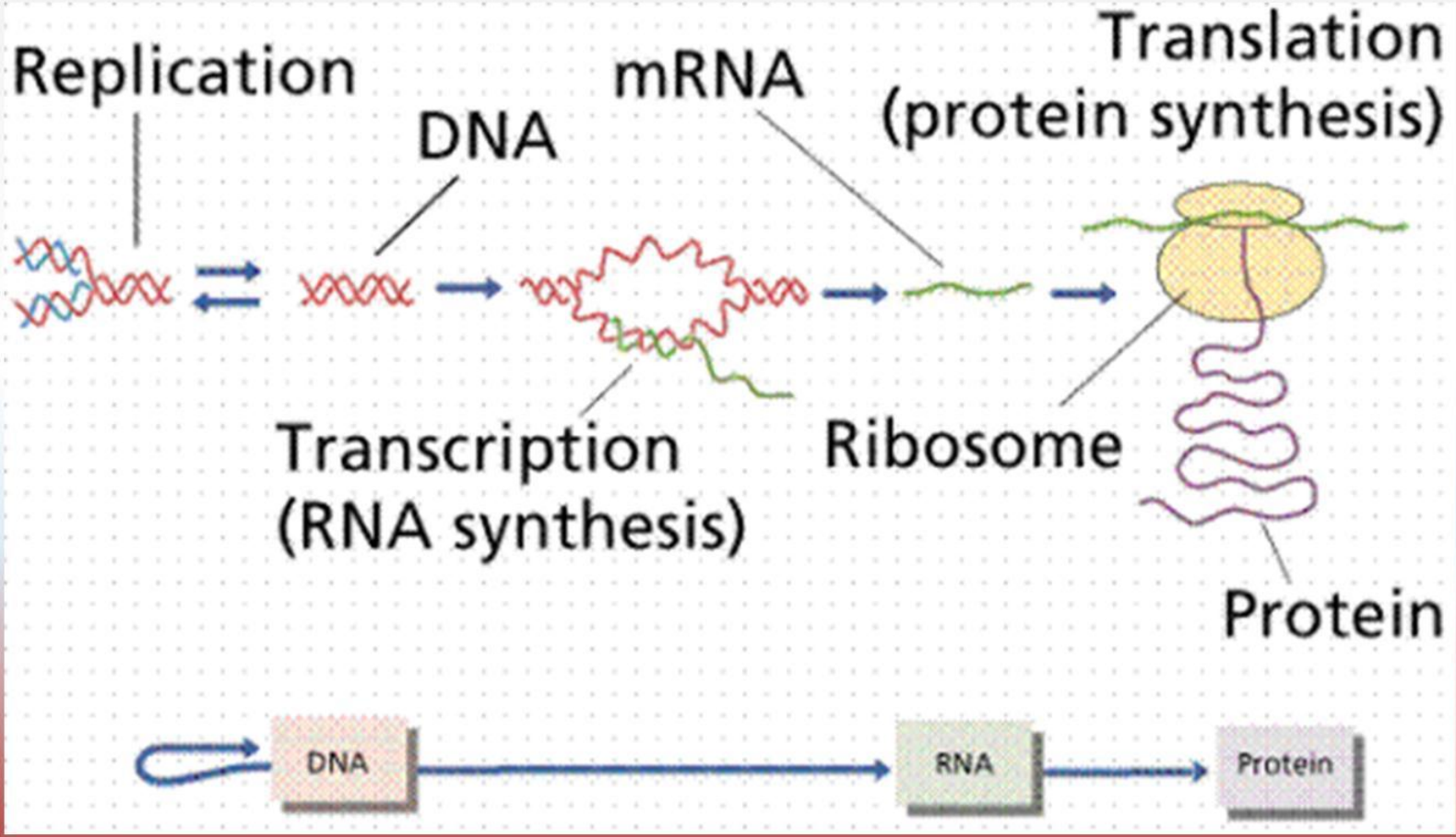
- The central component of the ribosome.
- Ribosome; factory for protein synthesis; composed of ribosomal RNA and ribosomal proteins (known as a Ribonucleoproteinor RNP).
- rRNA provides a mechanism for decoding mRNA into amino acids.

The Central Dogma of Molecular Biology

- DNA molecules serve as templates for either complementary DNA strands during the process of replication or complementary RNA during the process of transcription.

- RNA molecules serve as a template for ordering amino acids by ribosomes during protein synthesis.

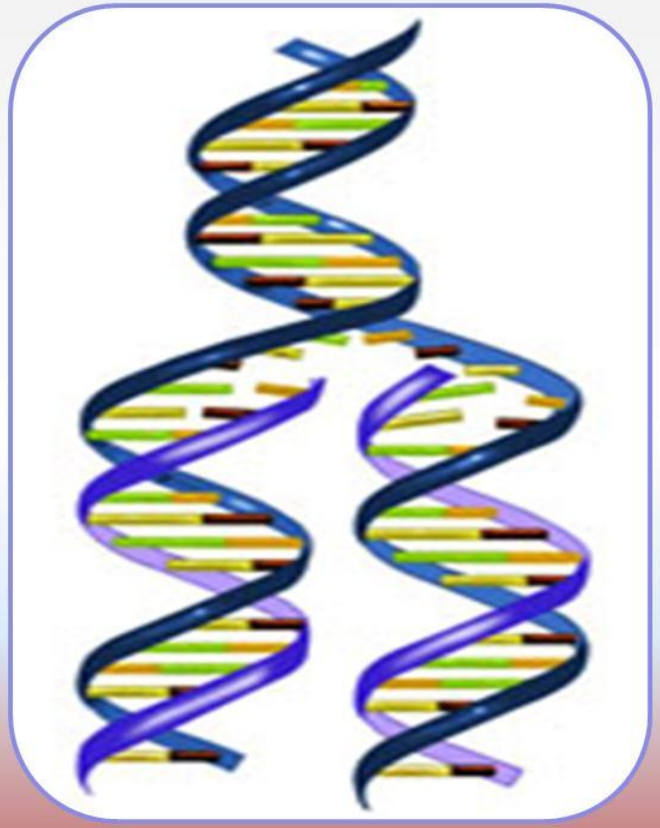
The Central Dogma of Molecular Biology



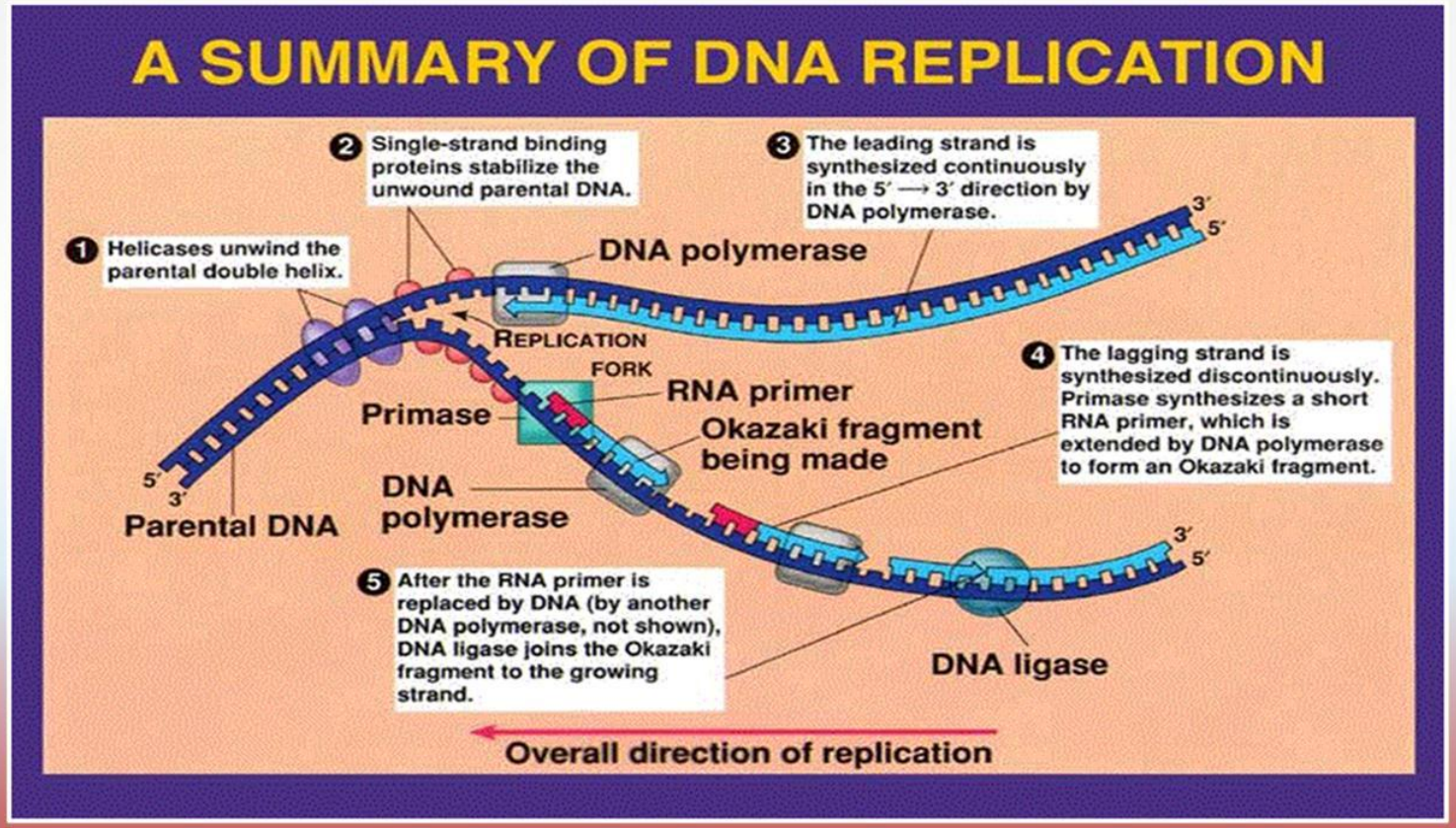
DNA Replication

- The DNA duplication.
- The transfer the genetic information from a parent to a daughter cell.
- The DNA base sequences are precisely copied.

DNA Replication



Source: National Institute of General Medical Sciences.



George Rice. Montana State University.
Source: http://serc.carleton.edu/microbelife/research_methods/genomics/replication.html

Post-Replicative Modification of DNA

- Methylation; one of the major post- replicative reactions.

- Site of methylation of eukaryotic DNA is always on cytosine residues in CG dinucleotide.

- DNA methylation plays an important role for epigenetic gene regulation in development and disease.

Gene Expression

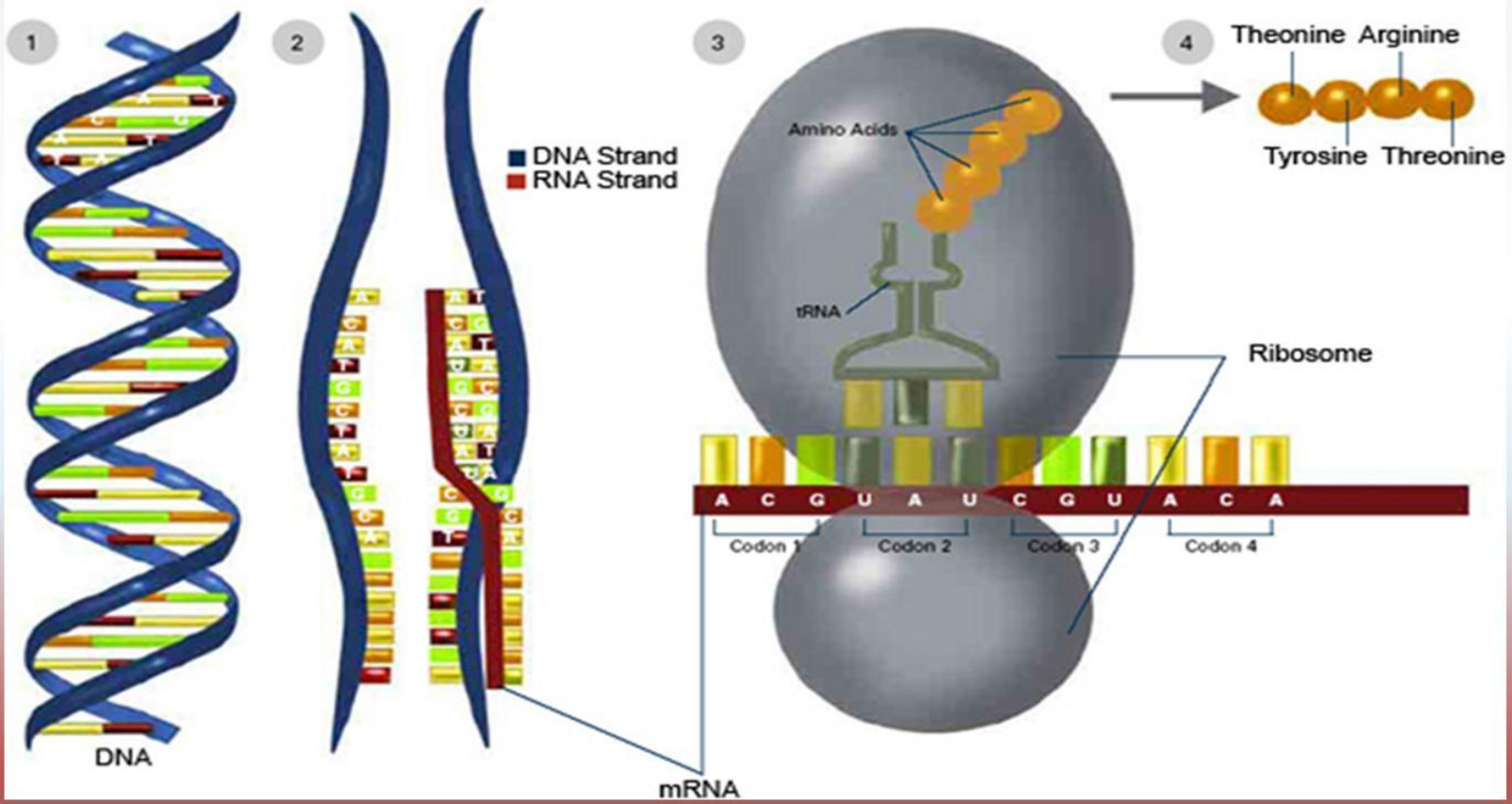
Transcription

RNA polymerase makes a copy of information in the gene (complementary RNA) complementary to one strands of DNA.

Translation

Occurs on **ribosomes**, messenger RNA decoded or translated to determine the sequence of amino acid in the protein being synthesized.

From DNA to Protein



Ashcraft. Source: http://creationwiki.org/File:Gene_expression.PNG

Ribosomes

- Factory for protein synthesis.
- Composed of ribosomal RNA and ribosomal proteins (known as a Ribonucleoprotein or RNP).
- Translate (mRNA) to build polypeptide chains using amino acids delivered by (tRNA).

The Protein

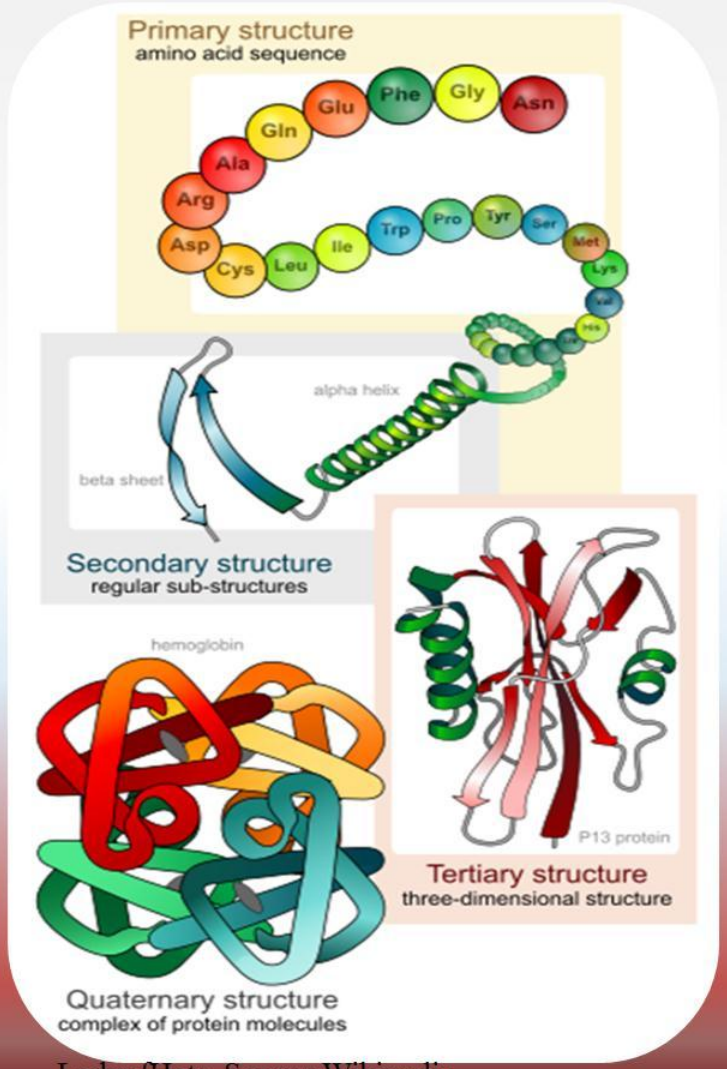
- **Proteins** are chain like polymers of a few or many thousands of amino acids.

- **Amino acids:** (3-nucleotide RNA sequences) (codon).

Four levels of Protein Structure

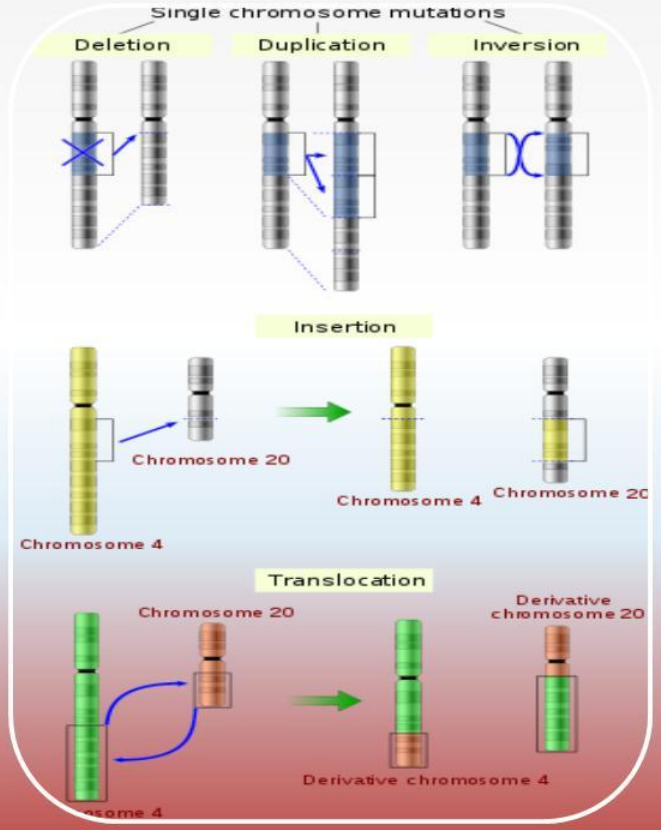
- **Primary protein structure:** Sequence of a chain of amino acid.
- **Secondary protein structure:** A chain of amino acids linked by hydrogen bonds.
- **Tertiary protein structure:** It occurs when certain attraction occurs between alpha helices and pleated sheets.
- **Quaternary protein structure:** Protein containing more than one amino acid chains.

Four levels of Protein Structure



LadyofHats. Source:Wikipedia

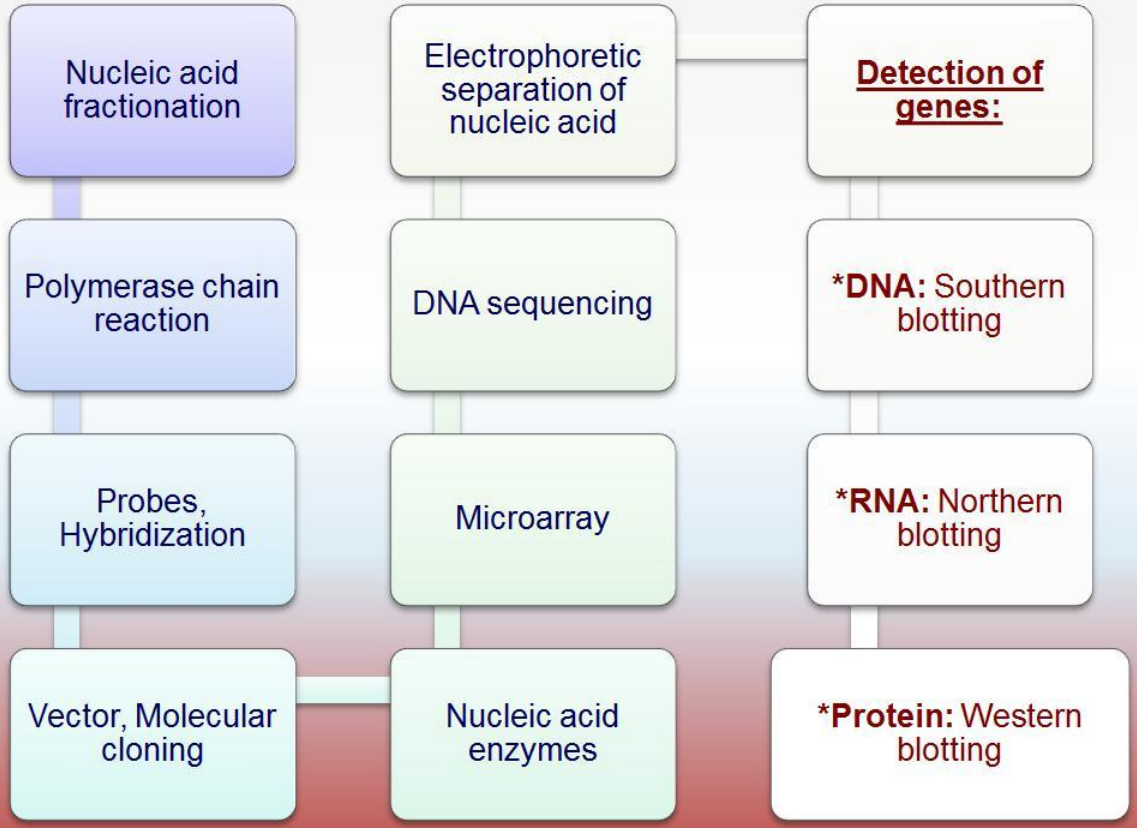
Genetic Mutation



Yassine Mrabet . Source:Wikipedia

A mutation is a change in the DNA sequence or arrangement of DNA.

Common Tools of Molecular Biology



Human Genome Project

- HGP is an international project aiming for:
  - Sequencing and localization of the base sequence that makes up human DNA.
  - Store this information in databases.
  - Mapping of human genome requires a set of landmarks; some of this landmarkers are genes but many more are nameless stretches of DNA such as RFLPs, VNTRs, STSs.

Human Genome Project

- 1990, American geneticists started an ambitious quest to map and sequence the entire human genome.
- 1999, the final draft of human chromosome 22.
- 2000, the final draft of human chromosome 21.
- 2001, working draft of the whole human genome.
- 2004, the finished sequence of the euchromatic part of human genome.

Human Genome Project

Chromosome X

Functional Genomics / Transcriptomics/ Proteomics

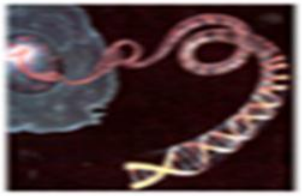
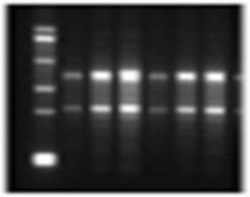
- **Functional Genomics**
  - The study of expression of large number of genes.
- **Transcriptomics**
  - The study of transcriptomes (all the transcripts an organism makes at any given time).
- **Proteomics**
  - The study of proteomes (the set of expressed proteins in a given type of cells or an organism at a given time under defined conditions).

from genotype to phenotype

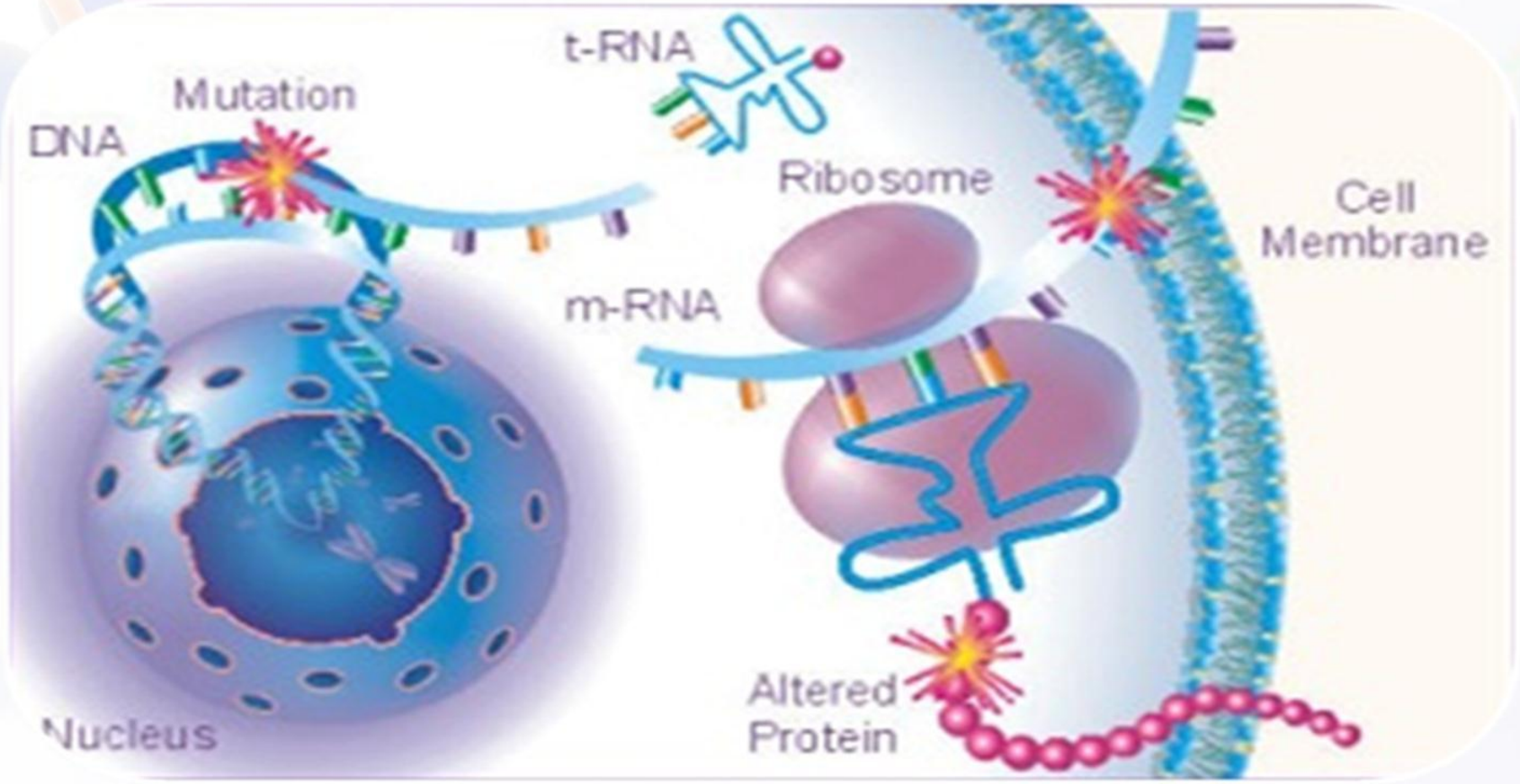
DNA => pre-mRNA => mRNA => => => => protein

↓ transcription ↓ splicing ↓ translation ↓ post-translational modifications

genome transcriptome proteome

Source: European Bioinformatics Institute. http://www.ebi.ac.uk/microarray/biology_intro.html

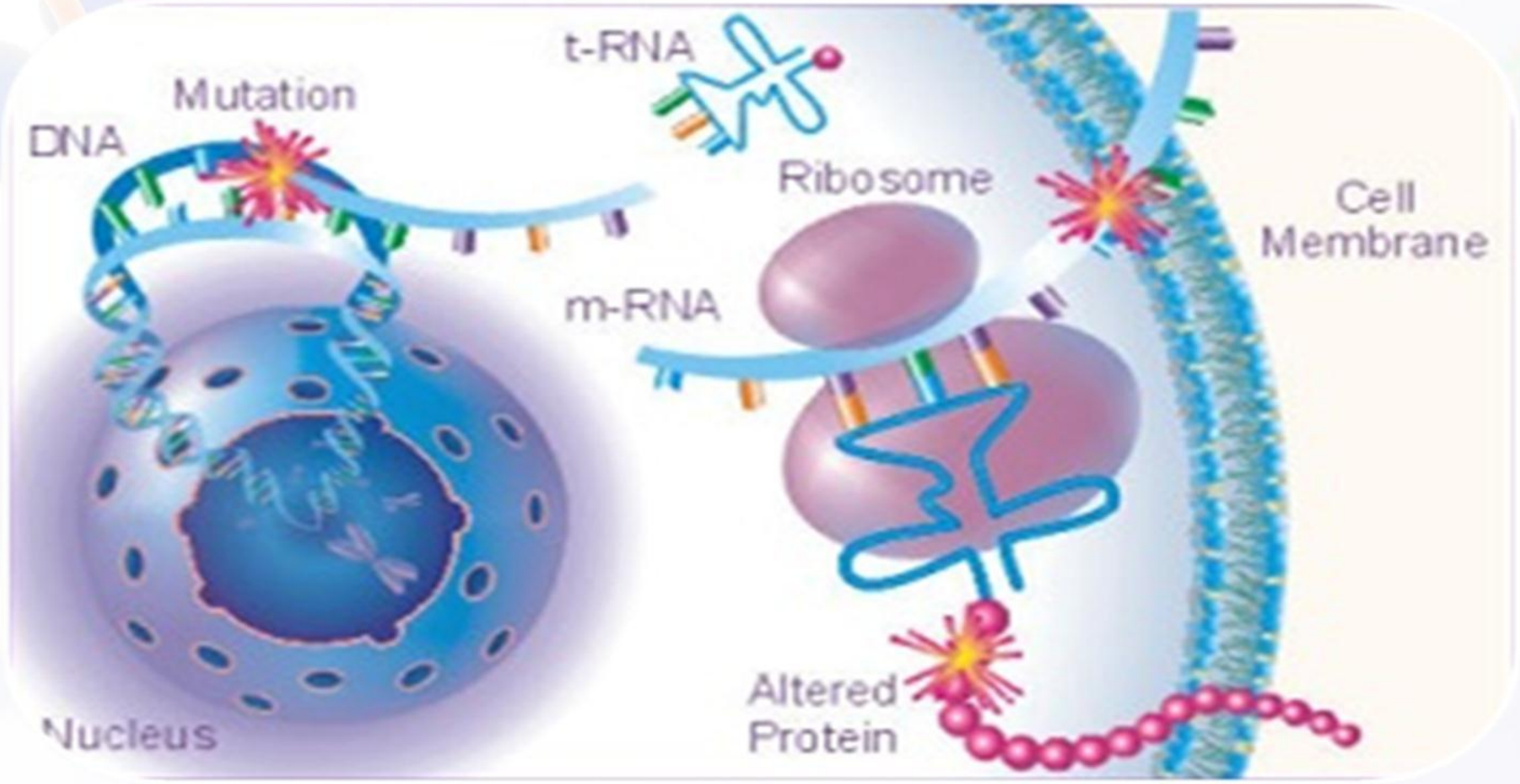

t-RNA
Mutation
DNA
Ribosome
Cell
Membrane
m-RNA
Nucleus
Altered
Protein

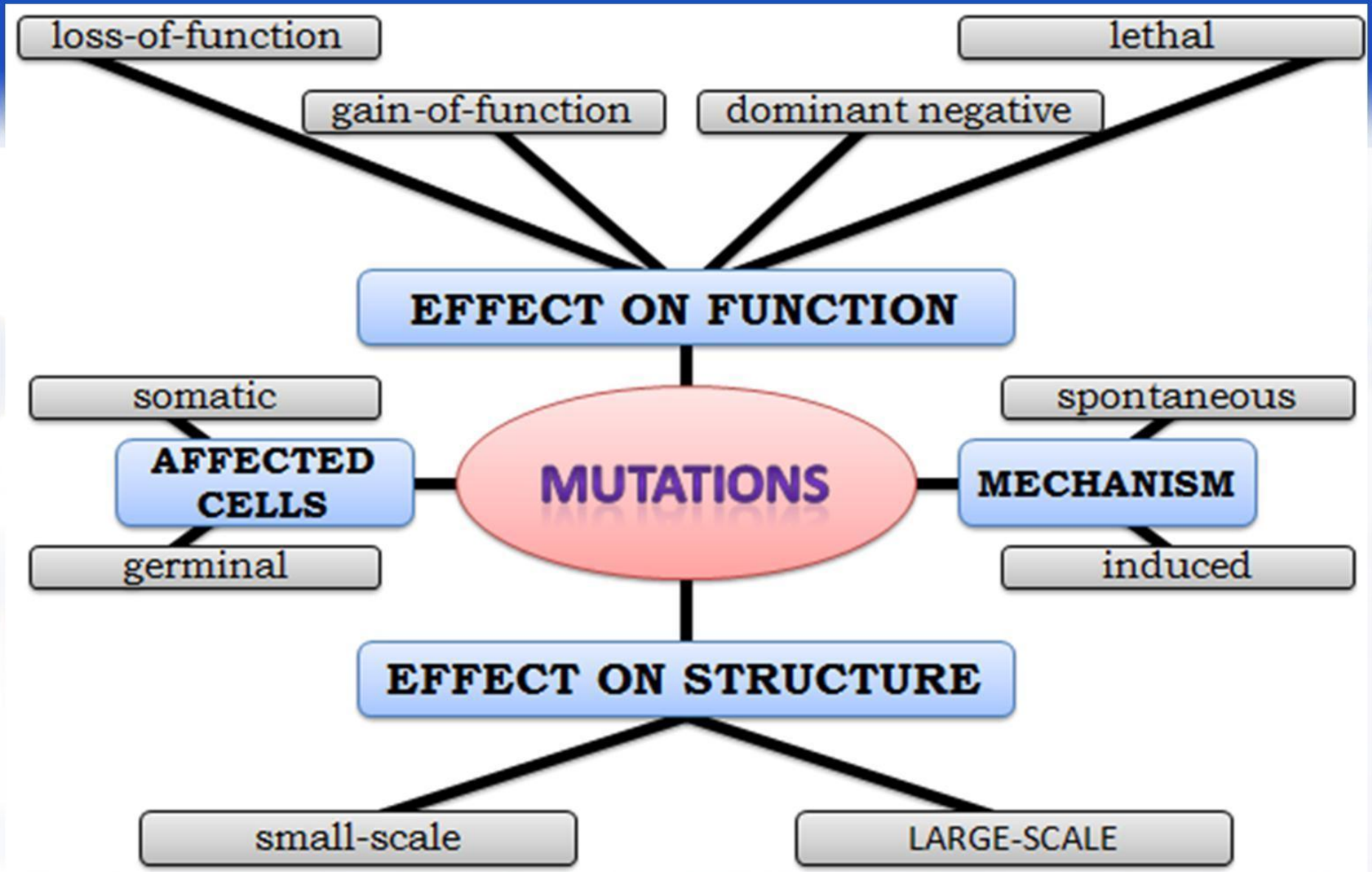
MUTATION

Mutation is a change in genetic material.

Mutation is a change in the DNA sequence or chromosomal mutation or arrangement of DNA.

Mutation are the result of error during DNA replication process/ error during DNA repair. Some types of mutations are known to be caused by certain chemicals and ionizing radiation UV.


t-RNA
Mutation
DNA
Ribosome
Cell
Membrane
m-RNA
Nucleus
Altered
Protein


loss-of-function
lethal
gain-of-function
dominant negative
EFFECT ON FUNCTION
somatic
spontaneous
AFFECTED CELLS
MUTATIONS
MECHANISM
germinal
induced
EFFECT ON STRUCTURE
small-scale
LARGE-SCALE

Spontaneous mutation

- **Spontaneous mutation** that arise naturally and not as a result of exposure to mutagens

Induced mutations

- **Induced mutations** caused by mutagens; Radiation, Chemicals, or Viruses,…

Origin of spontaneous mutation

Tautomerism: A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base pairing during replication.

Depurination: Loss of a purine base (A or G) to form an apurinic site (AP site).

Deamination: Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.

Slipped strand mispairing: renaturation in a different spot ("slipping") after denaturation of the new strand from the template during replication. This can lead to insertions or deletions.

Induced mutation

Dr./Salwa Teama

Mutagens	
Nitrous acid and nitrosoguanidine	Chemical modifications of purine and pyrimidine bases that alter their hydrogen-bonding properties.
Base analogs	During DNA replication, they can be incorporated into the DNA instead of the natural base.
Intercalating agents	During DNA replication, these compounds can insert or intercalate between adjacent base pairs e.g. ethidium bromide.
Ultraviolet Radiation	UV light is absorbed by DNA and causes adjacent thymine bases on the same DNA strand to covalently bond together, forming what are called thymine- Dimmer.
Ionizing Radiation	It ionizes water and other molecules to form radicals (molecular fragments with unpaired electrons) that can break DNA strands and alter purine and pyrimidine bases.

Small scale mutation

- Point mutation
- Insertion/Deletion

Large scale mutation

- Amplifications
- Deletions of large chromosomal regions, leading to loss of the genes within those regions.
- Mutations whose effect is to juxtapose previously separate pieces of DNA, potentially bringing together separate genes to form functionally distinct fusion genes(e.g.bcr-abl). These include:
  - Chromosomal translocations
  - Interstitial deletions
  - Chromosomal inversions: reversing the orientation of a chromosomal segment.
  - Loss of heterozygosity: loss of one allele, either by a deletion or a recombination event, in an organism that previously had two different alleles.

Small scale mutation

- Point mutation
- Insertion/Deletion

Large scale mutation

- Amplifications
- Deletions of large chromosomal regions, leading to loss of the genes within those regions.
- Mutations whose effect is to juxtapose previously separate pieces of DNA, potentially bringing together separate genes to form functionally distinct fusion genes(e.g.bcr-abl). These include:
  - Chromosomal translocations
  - Interstitial deletions
  - Chromosomal inversions: reversing the orientation of a chromosomal segment.
  - Loss of heterozygosity: loss of one allele, either by a deletion or a recombination event, in an organism that previously had two different alleles.

Germline mutation

- A heritable change in the DNA.
- Occurred in a germ cell and incorporated in every cell of the body.
- Can be transmitted to the next generation.
- Germline mutations play a key role in inherited genetic diseases.

Somatic mutation

- Alterations in DNA that occur after conception.
- Occur in any of the cells of the body except germ cell.
- Can not transmitted to the next generation.
- These alterations can (but do not always) cause cancer or other diseases.

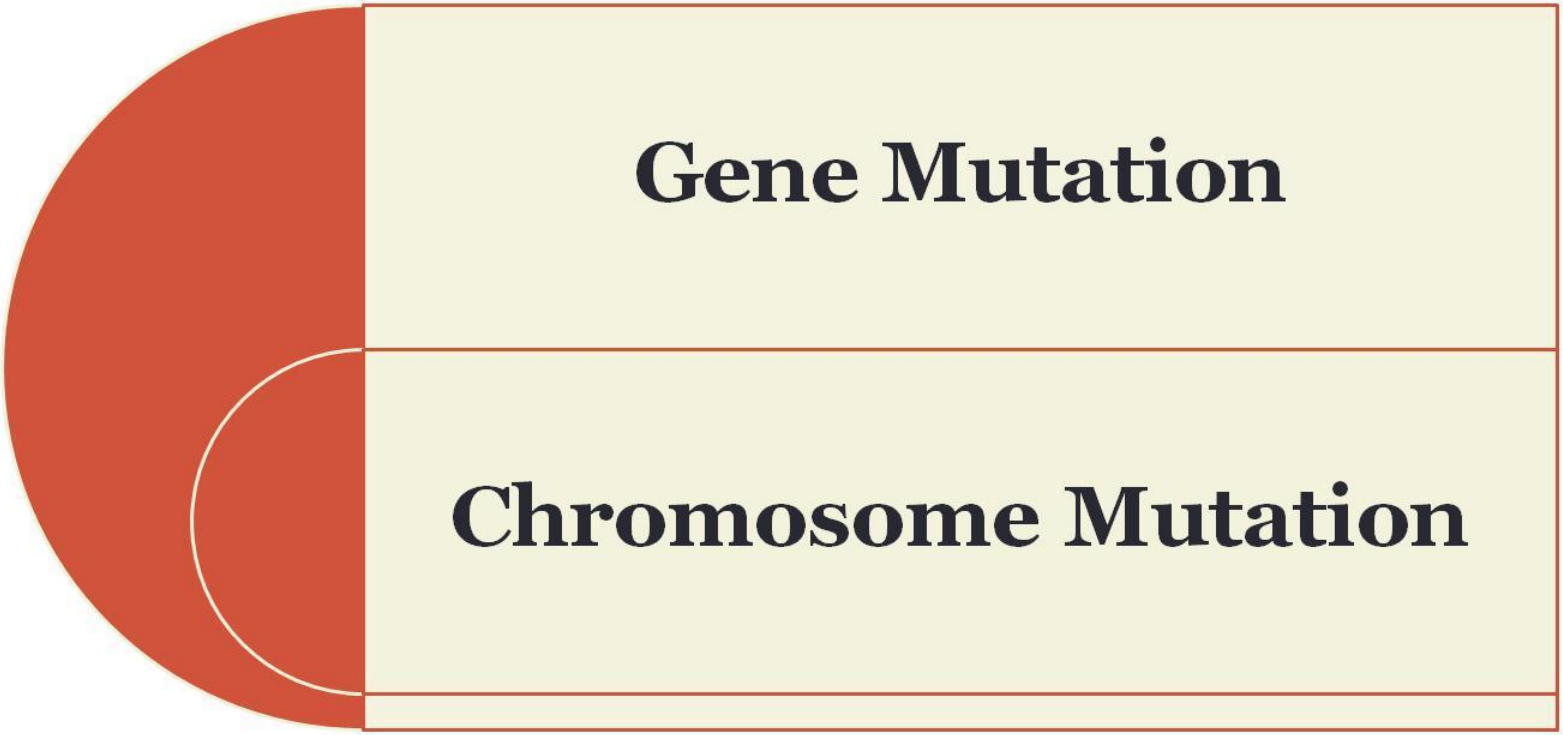
Gain-of-function mutations

- Change the gene product such that it gains a new and abnormal function.
- These mutations usually have dominant phenotypes.
- Often called a neomorphic mutation.

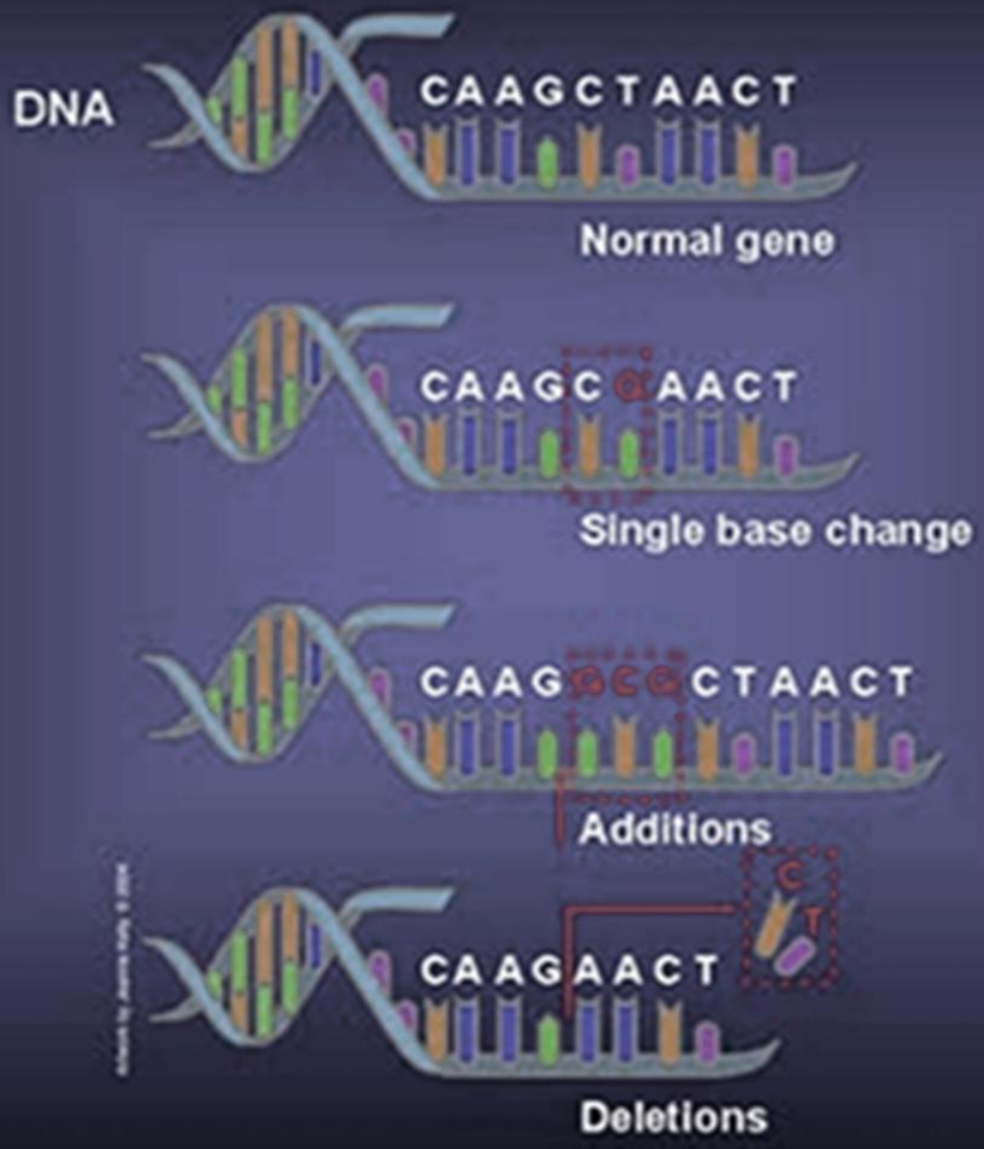
Loss-of-function mutations

- The gene product having less or no function.
- These mutations usually have recessive phenotypes. Exceptions are when the organism is haploid, or when the reduced dosage of a normal gene product is not enough for a normal phenotype (haploinsufficiency).
- When the allele has a complete loss of function (null allele) it is often called an amorphic mutation.

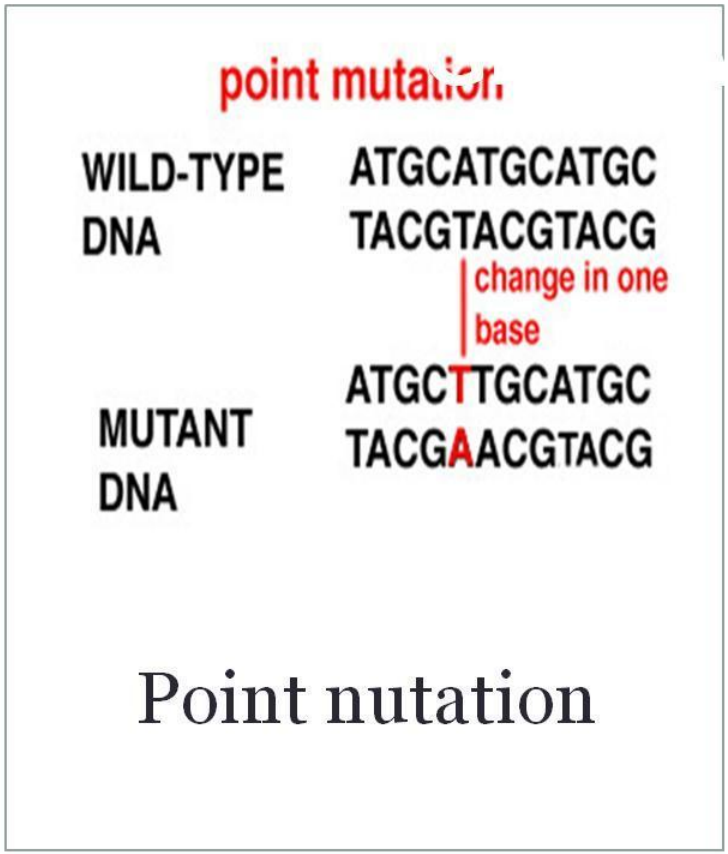
Types of Mutation



DNA Mutation



GENE MUTATION



Point nutation

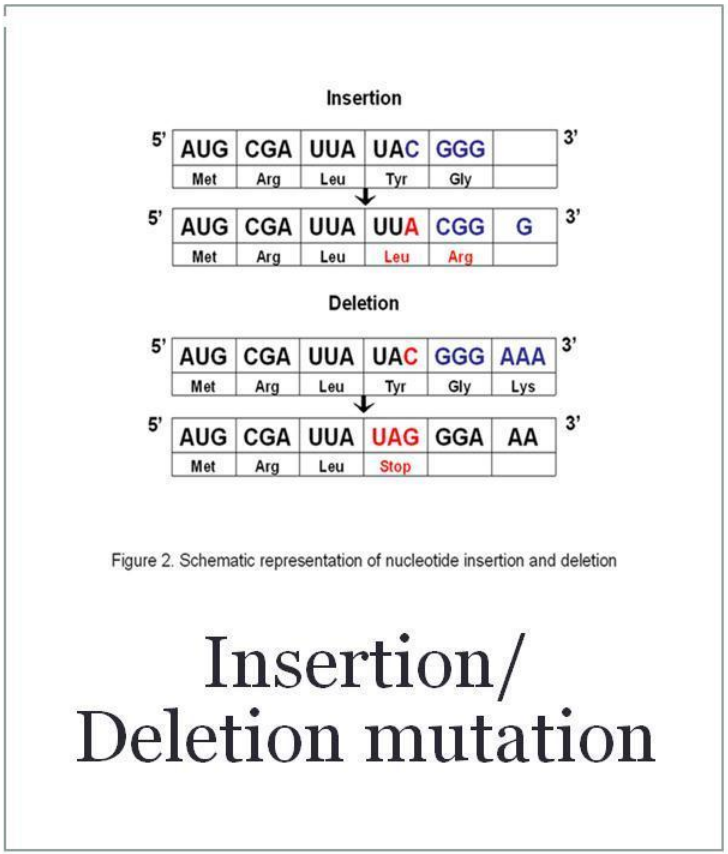


Figure 2. Schematic representation of nucleotide insertion and deletion

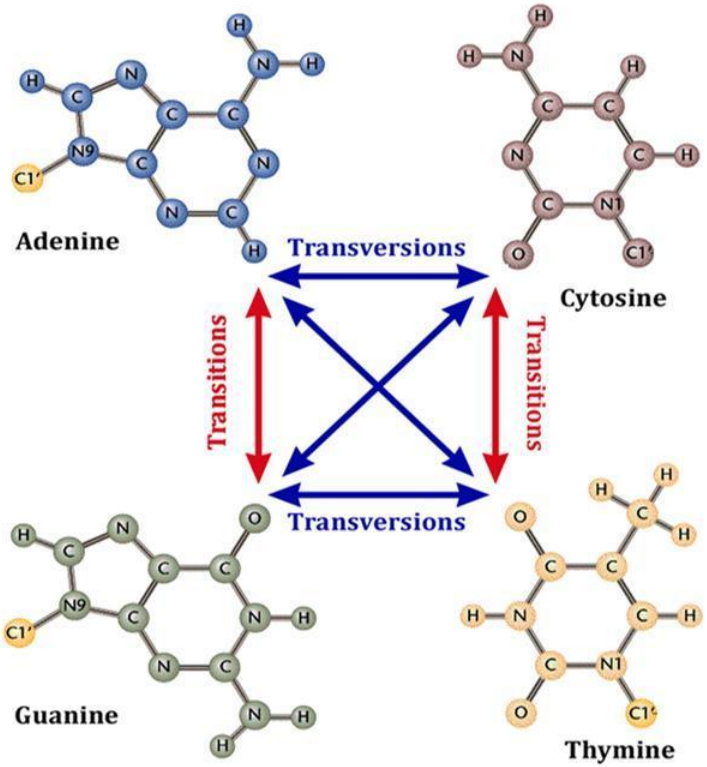
Insertion/
Deletion mutation

Point Mutation

Base substitutions are those mutations in which one base pair is replaced by another.

Transition: The replacement of a base by the other base of the same chemical category (purine/ purine; pyrimidine/ pyrimidine).

Transversion: The replacement of a base of one chemical category by a base of the other (pyrimidine/ purine;purine/ pyrimidine).



Effect of Point Mutation

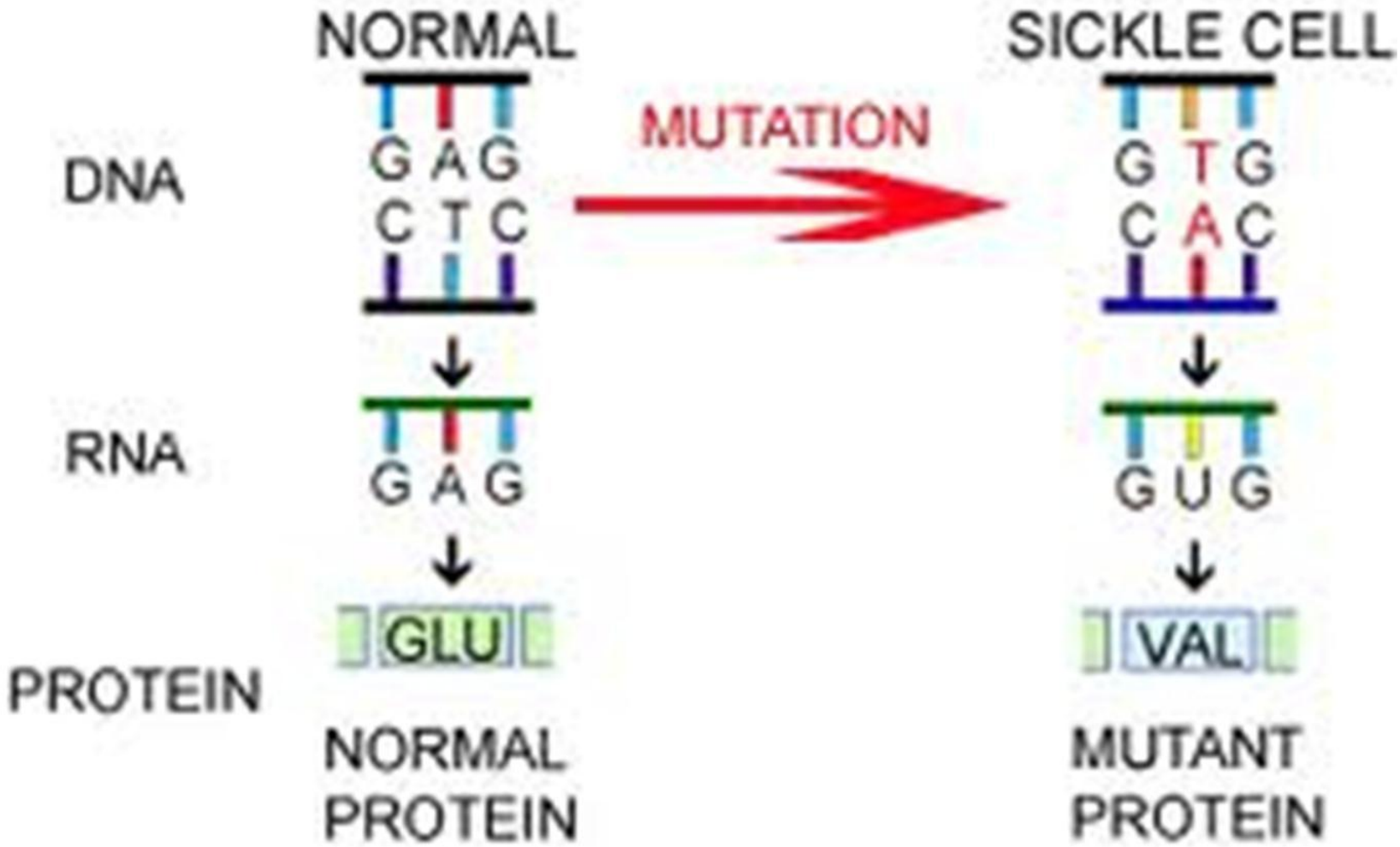
Silent mutation: Single substitution mutation when the change in the DNA base sequence results in a new codon still coding for the same amino acid.

Missense mutation: One triplet codon altered, results in one wrong codon and one wrong amino acid.

- Acceptable missense: This occur when single base change results in replacement of one A.A by another with rather same function.
- Non acceptable missense: This occur when single base change result in the replacement of one amino acid with another with completely different function.

Nonsense mutation: Change a codon that specifies an amino acid into a termination codon.

Sickle Cell Disease



Insertion/Deletion

Insertion/deletion can disrupt the grouping of the codons, resulting in a completely different translation.

Deletions: Remove information from the gene. A deletion could be as small as a single base or as large as the gene itself.

Insertions: Occur when extra DNA is added into an existing gene.

Multiple of 3 (codon) / Not multiple of 3

Insertion/Deletion

Multiple of 3 (codon) / Not multiple of 3

- **Multiple of 3 (codon)**

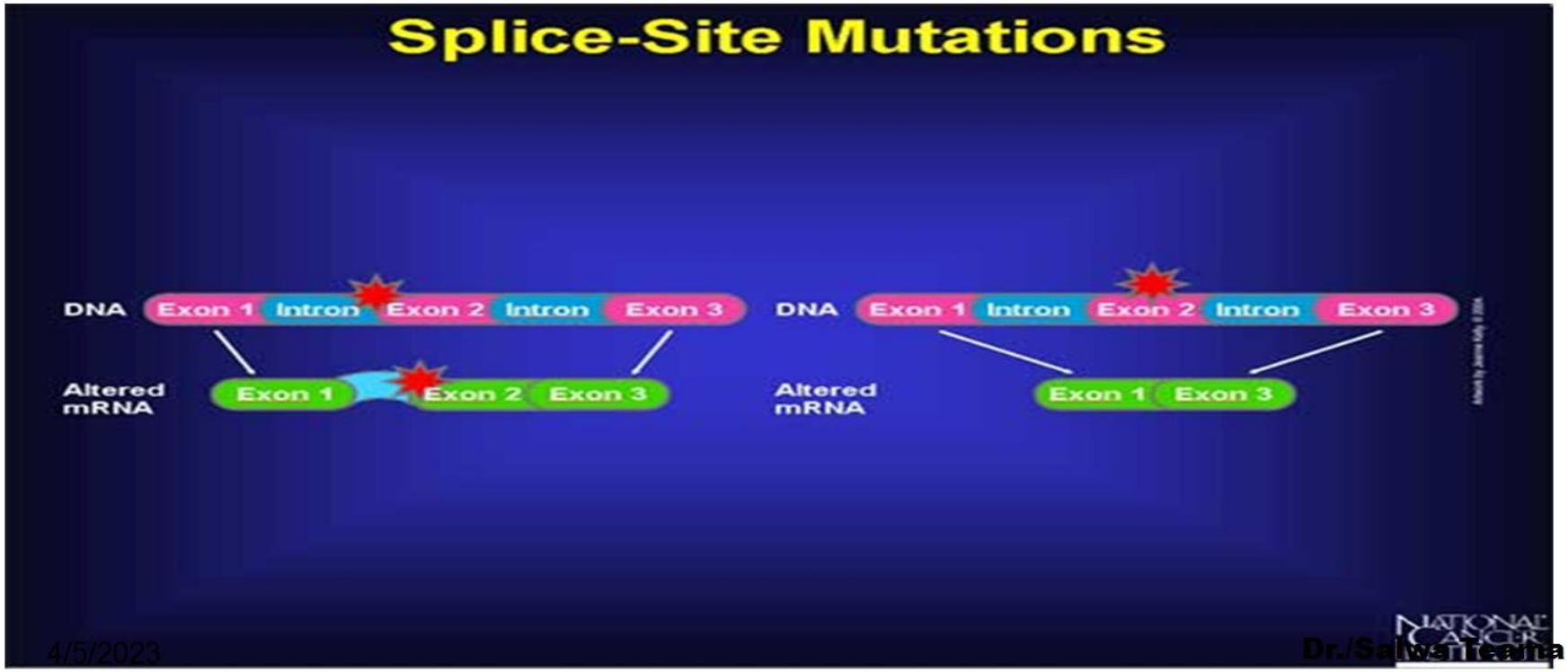
  Multiple of 3 (codon) causes loss or gain of codons and subsequently amino acids in translated product.

- **Not multiple of 3**

  Not multiple of 3; Altered reading frame or fram-shift, altered amino acid sequence

RNA PROCESSING MUTATION

RNA processing mutation destroy consensus splice sites, cap sites, polyadenylation sites, ..Abnormal splicing often leads to frameshift mutations and premature stop codon.



Chromosomal Mutation

Chromosomal aberration: change in chromosome structure

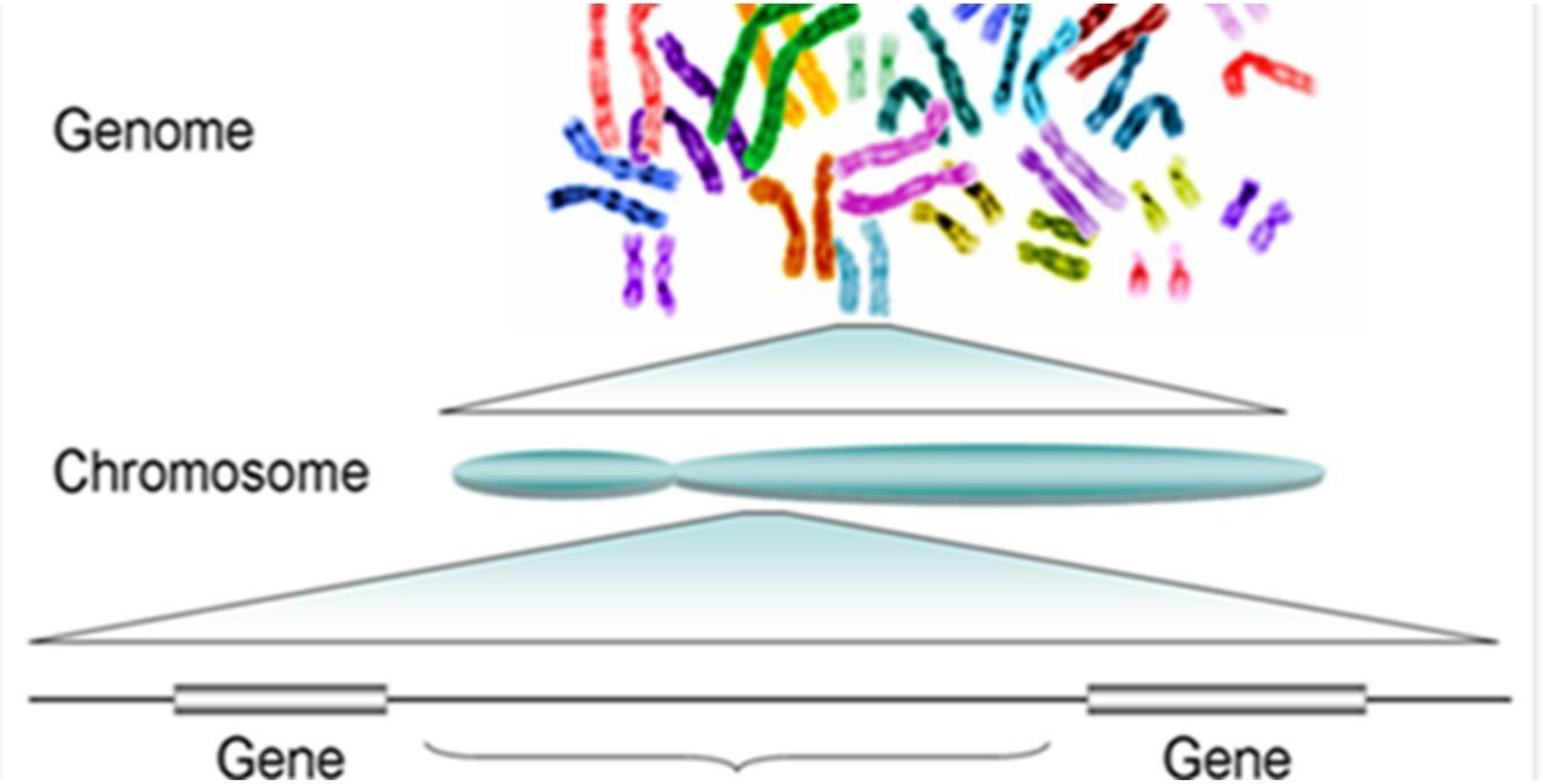
Change in chromosome number

Aberrations of Chromosome Structure	
Type	Description
Deletion	A loss of a portion of a chromosome.
Duplication	The presence of an extra chromosomal segment.
Inversion	A fragmentation of a chromosome by two breaks, followed by a reconstitution of the segment between the break with inversion (ABCdefGHIJK) might become (ABCfedGHIJK).
Insertion	A rare event involving three chromosomal breaks and which a segment is removed from one chromosome and then inserted into a broken region of a nonhomologous chromosome.
Translocation	The exchange of a segment of a chromosome is transferred

Heteroploidy can be mainly of two types:

- Aneuploidy : Gain or loss of one or more chromosome
- Euploidy : Gain or loss of genome

Variations in the chromosome number



Variations in the chromosome number

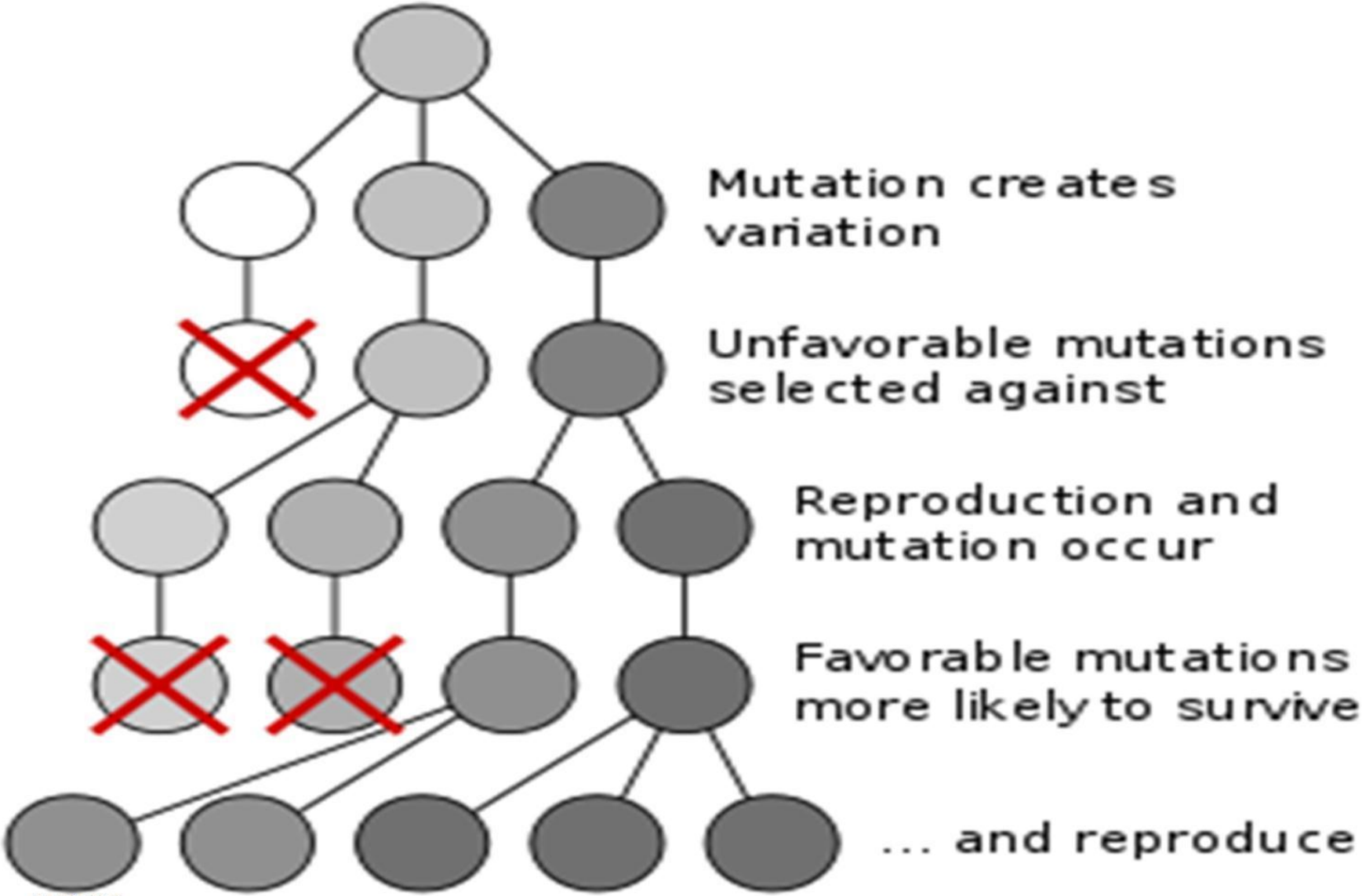
Aneuploidy:

- Hypoploidy: Monosomy, nullisomy
- Hyperploidy: Trisomy, Tetrasomy

Euploidy:

- Haploidy
- Polyplidy
  - Autopolyplidy
  - Allopolyplidy

Type of mutation	Result and example(s)
Forward mutations	
Single-nucleotide-pair (base-pair) substitutions	
At DNA level	
Transition	Purine replaced by a different purine, or pyrimidine replaced by a different pyrimidine: A·T ⟶ C·G C·G ⟶ A·T C·G ⟶ T·A T·A ⟶ C·G
Transversion	Purine replaced by a pyrimidine, or pyrimidine replaced by a purine: A·T ⟶ C·G A·T ⟶ T·A G·C ⟶ T·A G·C ⟶ C·G T·A ⟶ G·C T·A ⟶ A·T C·G ⟶ A·T C·G ⟶ G·C
At protein level	
Silent mutation	Triplet codes for same amino acid: AGG ⟶ CGG both code for Arg
Missense mutation	Codon specifies a different amino acid.
Synonymous missense mutation	Codon specifies chemically similar amino acid: AAA ⟶ AGA changing basic Lys to basic Arg (does not alter protein function in many cases)
Nonsynonymous missense mutation	Codon specifies chemically dissimilar amino acid: UUU ⟶ UCU Hydrophobic Phenylalanine → Polar Serine
Nonsense mutation	Codon signals chain termination: CAG ⟶ UAG changing from a codon for Gln to an amber termination codon
Single-nucleotide-pair addition or deletion: frameshift mutation Addition or deletion of several to many nucleotide pairs	Any addition or deletion of base pairs that is not a multiple of 3 changes the reading frame in DNA segments that code for proteins, resulting in different amino acids from that point on and frequently chain termination.
Reverse mutations	AAA (Lys) (Wild-type) —forward→ GAA (Glu) (Mutant) —reverse→ AAA (Lys) (Wild-type)


Mutation creates variation
Unfavorable mutations selected against
Reproduction and mutation occur
Favorable mutations more likely to survive
... and reproduce

Application of Molecular Biology

- Research
- Diagnosis
- Transplantation
- Paternity
- Forensic analysis
- Gene therapy
- Drug Design
- ……

References and Further Reading

- Ali Khalifa. Applied molecular biology; eds: (Fathi Tash and Sanna Eissa). 109 pages. Egypt. University Book Center. 2002. Available in paper copy from the publisher.
- Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Molecular Biology of the cell. 1392 pages. Garland Science; 5 edition (November 16, 2007).ISBN. 9780815341055. Available in paper copy from the publisher.
- Daniel H. Farkas. DNA Simplified: The Hitchhiker's Guide to DNA. 110 pages. Washington, DC: AACC Press, 1996, ISBN 0-915274-84-1. Available in paper copy from the publisher.
- Daniel P. Stites, Abba T. Terr. Basic Human Immunology: 336 Pages. Appleton & Lange (November 1990). ISBN. 0838505430. Available in paper copy from the publisher.
- Innis, David H. Gelfand, John J. Sninsky. PCR Applications: Protocols for Functional Genomics: 566 pages. Academic Press; 1 edition (May 17, 1999). ISBN:0123721865. Available in paper copy from the publisher.
- Robert F. Mueller, Ian D. Young. **Emery's Elements** of **Medical Genetics**: Publisher: **Churchill Livingstone**. **1995**. ISBN. *044307125X.* Available in paper copy from the publisher.
- **Robert F. Weaver. Molecular Biology. 600 Pages. Fourth Edition. McGraw-Hill International Edition. ISBN 978-0-07-110216-2.** Available in paper copy from the publisher.
- William B. Coleman, Gregory J. Tsongalis. Molecular Diagnostics. For the Clinical Laboratorian: 592 pages. Humana Press; 4th Printing. edition (August 15, 2005). ISBN 1588293564. Available in paper copy from the publisher.


THANK YOU
Dr./Salwa Teama