Ministry of Public Health of Ukraine
Poltava State Medical University
Department of biological and bioorganic chemistry

Regulation of gene exspression. Mutations. DNA reparation.

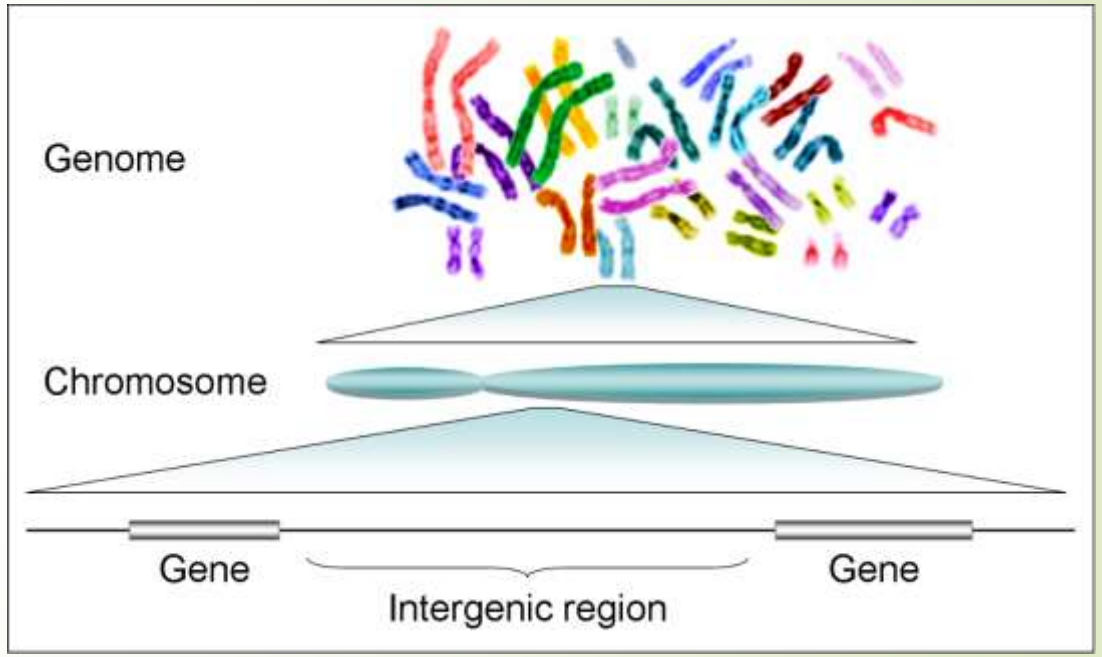
Assoc. Prof. Bilets M.V.

Lecture plan

- The genes expression.
- Regulation of genes expression in prokaryotic cells.
- Regulation of genes expression in eukaryotic cells.
- Mechanisms of mutations.
- DNA repair.

The human genome

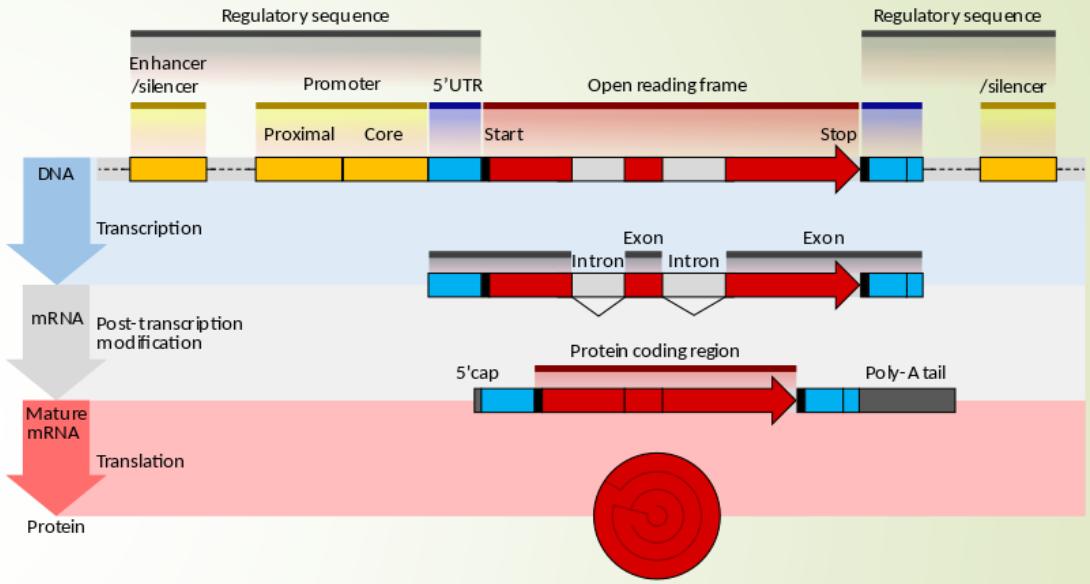
- The **genome is all genetic material of an organism**. It consists of DNA (or RNA in RNA viruses). The genome includes both the genes (the coding regions) and the noncoding DNA, as well as mitochondrial DNA.
- The human genome is a complete set of nucleic acid sequences for humans, encoded as DNA within the 23 chromosome pairs in cell nuclei and in a small DNA molecule found within individual mitochondria. Human genomes include both protein-coding DNA genes and noncoding DNA. Haploid human genomes, which are contained in germ cells (the egg and sperm gamete cells created in the meiosis phase of sexual reproduction before fertilization creates a zygote) consist of three billion DNA base pairs, while diploid genomes (found in somatic cells) have twice the DNA content.
- The haploid human genome (23 chromosomes) is about 3 billion base pairs long and contains around 30,000 genes. Since every base pair can be coded by 2 bits, this is about 750 megabytes of data. An individual somatic (diploid) cell contains twice this amount, that is, about 6 billion base pairs. Men have fewer than women because the Y chromosome is about 57 million base pairs whereas the X is about 156 million.



https://ru.m.wikipedia.org/wiki/Файл:Human_genome_to_genes.png

Gene structure

- Much of gene structure is broadly similar between eukaryotes and prokaryotes. Key differences in gene structure between eukaryotes and prokaryotes reflect their divergent transcription and translation machinery.
- The structures of both eukaryotic and prokaryotic genes involve several nested sequence elements. Each element has a specific function in the multi-step process of gene expression. The sequences and lengths of these elements vary, but the same general functions are present in most genes. Although DNA is a double-stranded molecule, typically only one of the strands encodes information that the RNA polymerase reads to produce protein-coding mRNA or non-coding RNA. This 'sense' or 'coding' strand, runs in the 5' to 3' direction where the numbers refer to the carbon atoms of the backbone's ribose sugar. The open reading frame (ORF) of a gene is therefore usually represented as an arrow indicating the direction in which the sense strand is read.
- Regulatory sequences are located at the extremities of genes. These sequence regions can either be next to the transcribed region (the promoter) or separated by many kilobases (enhancers and silencers). The promoter is located at the 5' end of the gene and is composed of a core promoter sequence and a proximal promoter sequence. The core promoter marks the start site for transcription by binding RNA polymerase and other proteins necessary for copying DNA to RNA. The proximal promoter region binds transcription factors that modify the affinity of the core promoter for RNA polymerase. Genes may be regulated by multiple enhancer and silencer sequences that further modify the activity of promoters by binding activator or repressor proteins. Enhancers and silencers may be distantly located from the gene, many thousands of base pairs away. The binding of different transcription factors, therefore, regulates the rate of transcription initiation at different times and in different cells.

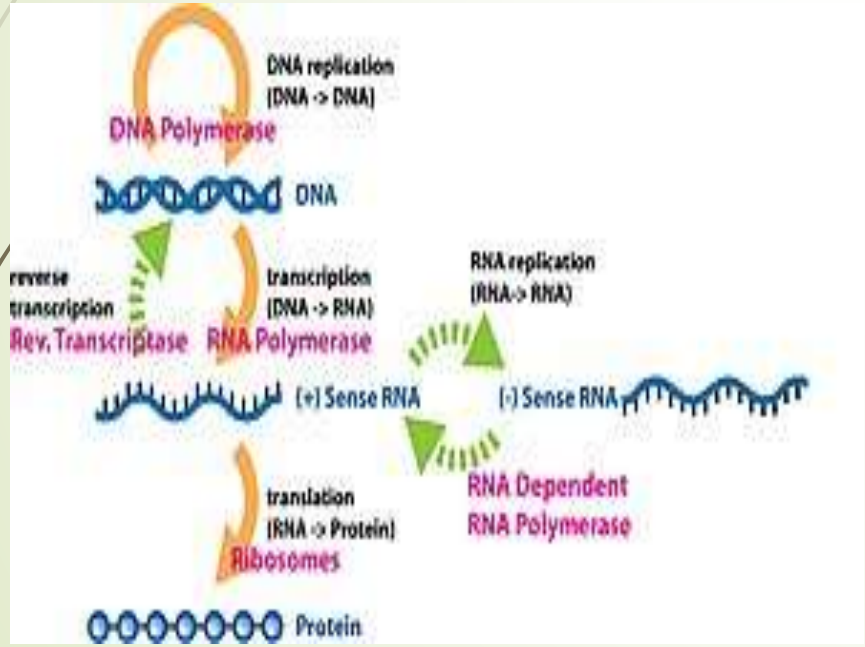


The structure of a eukaryotic protein-coding gene. Regulatory sequence controls when and where expression occurs for the protein coding region (red). Promoter and enhancer regions (yellow) regulate the transcription of the gene into a pre-mRNA which is modified to add a 5' cap and poly-A tail (grey) and remove introns. The mRNA 5' and 3' untranslated regions (blue) regulate translation into the final protein product.

https://en.wikipedia.org/wiki/Gene_structure

Gene expression

- Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product that enable to produce protein as the end product. These products are often proteins, but in non-protein-coding genes such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. Gene expression is summarized in the central dogma of molecular biology:



https://en.wikipedia.org/wiki/Gene_expression

- All steps in the gene expression process may be modulated (regulated), including the **transcription, RNA splicing, translation, and post-translational modification of a protein.** Regulation of gene expression gives control over the timing, location, and amount of a given gene product (protein or ncRNA) present in a cell and can have a profound effect on the cellular structure and function. cellular differentiation, development, morphogenesis and the versatility and adaptability of any organism.

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms	
Prokaryotic organisms	Eukaryotic organisms
Lack nucleus	Contain nucleus
DNA is found in the cytoplasm	DNA is confined to the nuclear compartment
RNA transcription and protein formation occur almost simultaneously	RNA transcription occurs prior to protein formation, and it takes place in the nucleus. Translation of RNA to protein occurs in the cytoplasm.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, nuclear shuttling, post-transcriptional, translational, and post-translational)

Regulation of genes expression in procariotic cells.

- The DNA of prokaryotes is organized into a circular chromosome, supercoiled within the nucleoid region of the cell cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called **operons**.
- In prokaryotic cells, there are three types of regulatory molecules that can affect the expression of operons: **repressors, activators, and inducers**. **Repressors and activators** are proteins produced in the cell. Both repressors and activators regulate gene expression by binding to specific DNA sites adjacent to the genes they control.
- Activators bind to the promoter site, while repressors bind to operator regions.
- Repressors prevent transcription of a gene in response to an external stimulus, whereas activators increase the transcription of a gene in response to an external stimulus.
- Inducers are small molecules that may be produced by the cell or that are in the cell's environment. Inducers either activate or repress transcription depending on the needs of the cell and the availability of substrate.

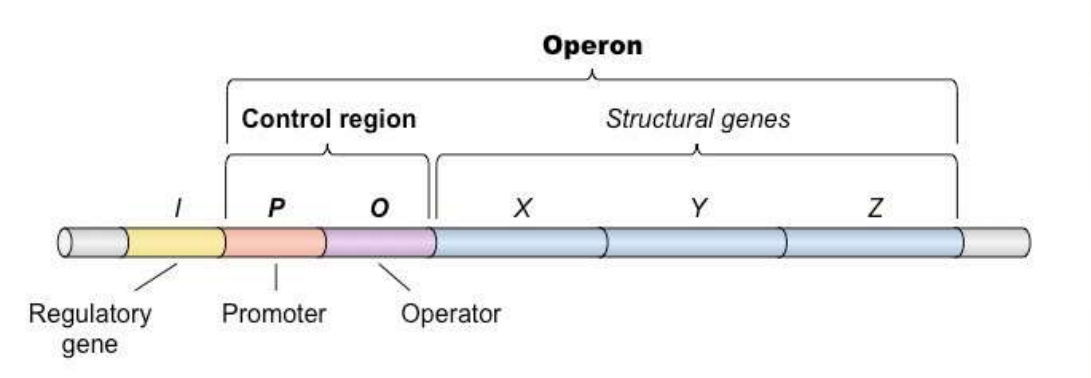
- Prokaryotes commonly control transcription

Constitutive genes are always expressed. Tend to be vital for basic cell functions (often called housekeeping genes)

Regulated genes can be inducible or repressible. Inducible genes are normally off, but can be turned on when substrate is present.

- The lac operon is inducible.
- The lac operon is activated by the CAP (catabolite activator protein), which binds to the promoter to stabilize RNA polymerase binding. CAP is itself activated by cAMP, whose concentration rises as the concentration of glucose falls. However, the lac operon also requires the presence of lactose for transcription to occur. Lactose inactivates the lac repressor, and prevents the repressor protein from binding to the lac operator. With the repressor inactivated, transcription may proceed. Therefore glucose must be absent and lactose must be present for effective transcription of the lac operon.

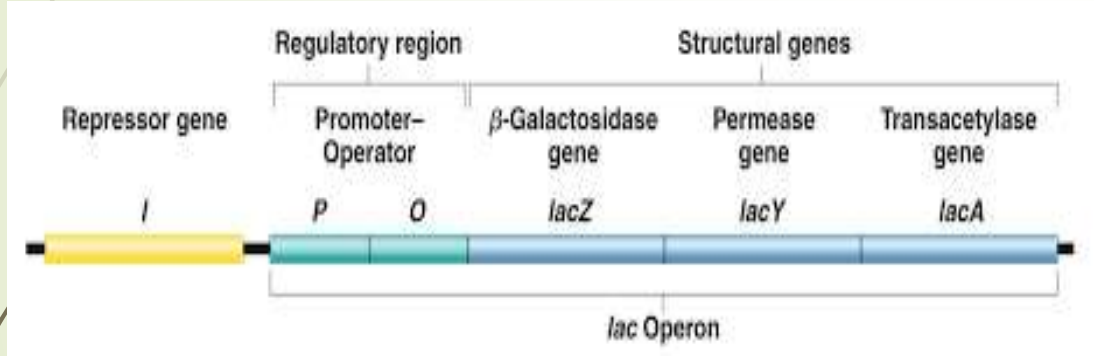
Operon

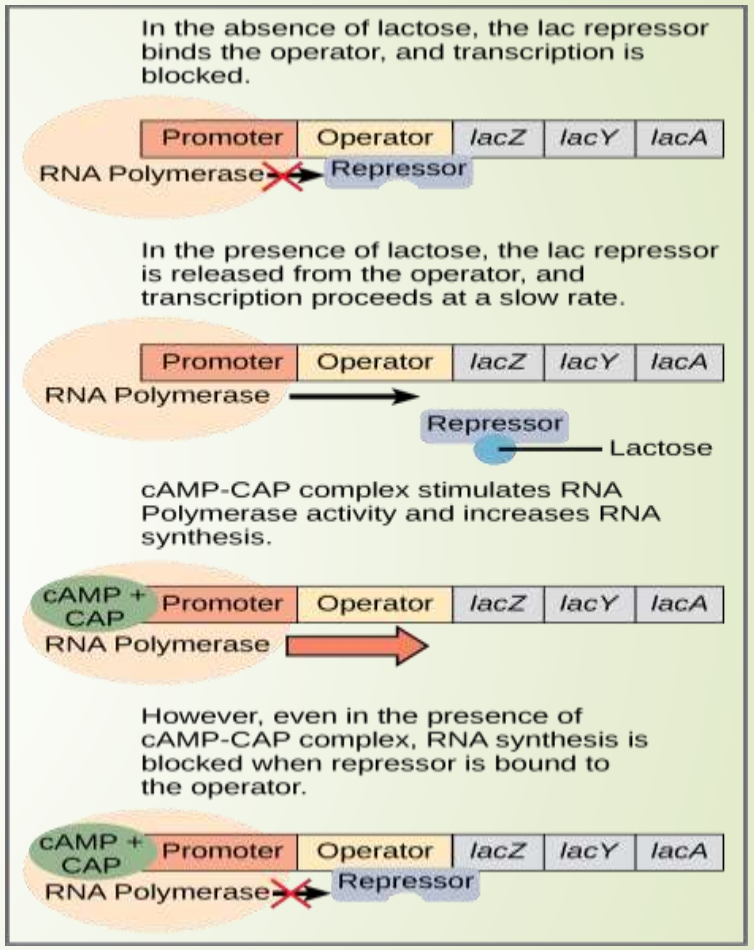


- There are three basic components to an operon
- *Promoter* – Upstream sequence to which RNA polymerase binds
- *Operator* – Segment of DNA to which a repressor protein binds (inhibits transcription by obstructing RNA polymerase)
- *Structural genes* – Genes that are collectively regulated by the operon

https://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/72-transcription-and-gene/operons.html

lac operon



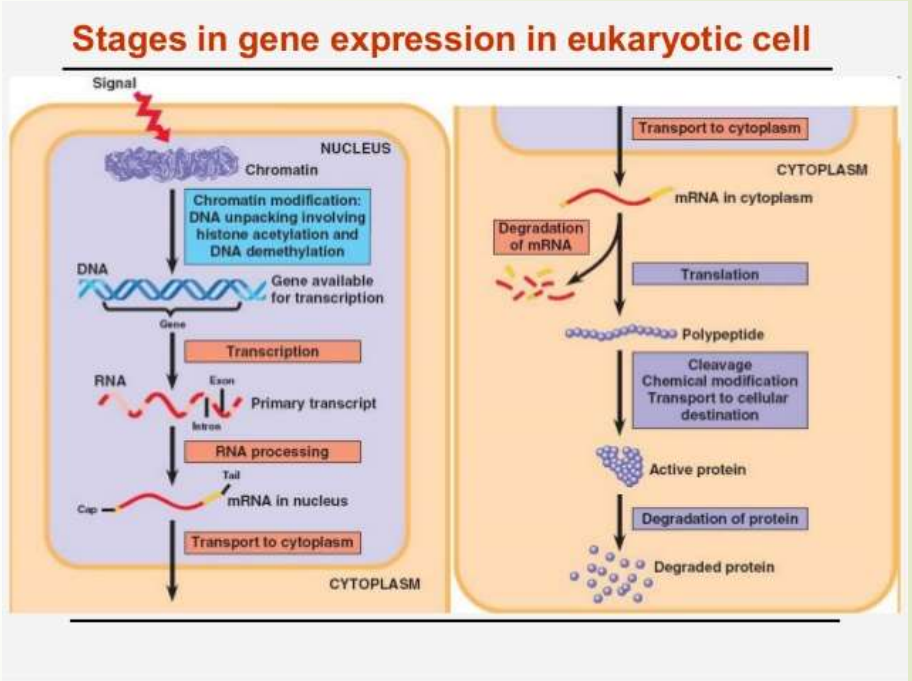


http://kocw-n.xcache.kinxcdn.com/data/document/2019/gachon/nammyeongjin0109/5.pdf

https://opentextbc.ca/biology2eopenstax/chapter/prokaryotic-gene-regulation/#:~:text=The%20regulation%20of%20gene%20expression,prokaryotic%20transcription%3A%20repressors%20and%20activators.&text=In%20the%20trp%20operon%2C%2

Regulation of genes expression in eukaryotic cells.

- The latest estimates are that a human cell, a eukaryotic cell, contains some 21,000 genes.
- Some of these are expressed in all cells all the time. These so-called **housekeeping genes** are responsible for the routine metabolic functions (e.g. respiration) common to all cells.
- Some are expressed as a cell enters a particular pathway of differentiation.
- Some are expressed all the time in only those cells that have differentiated in a particular way. For example, a plasma cell expresses continuously the genes for the antibody it synthesizes.
- Some are expressed only as conditions around and in the cell change. For example, the arrival of a hormone may turn on (or off) certain genes in that cell.



https://www.slideshare.net/NoeMendez9/eukaryotic-gene-regulation-models-by-np-mendez

Regulation of gene expression (chromatin changes)

1. Chromatin changes

- DNA of Eukaryotic cells is packaged in chromatin.
- Heterochromatin is highly condensed - transcriptional enzymes can not reach the DNA
- Genes within highly packed heterochromatin are usually not expressed
- **Acetylation / deacetylation** of histones
- **Methylation** [cytosin] - inactive DNA is highly methylated

1. Chromatin changes

- **DNA methylation**
  - Essential for long-term inactivation of genes during cell differentiation
- **Gene imprinting in mammals**
  - Methylation constantly turns off the maternal/ paternal allele of a gene in early development
  - certain genes are expressed in a parent-of-origin-specific manner
- **Epigenetic inheritance**

Histone Modifications

- In **histone acetylation**, acetyl groups are attached to positively charged lysines in histone tail
  - Acetylation promotes initiation of transcription. Deacetylation does not
- This loosens chromatin structure, thereby promoting the initiation of transcription
- The addition of methyl groups (methylation) can condense chromatin
- The addition of phosphate groups (phosphorylation) next to a methylated amino acid can loosen chromatin

Histone Acetylation

- HISTONES in transcriptionally active genes are often ACETYLATED.
- Acetylation is the modification of lysine residues in histones.
  - Reduces positive charge, weakens the interaction with DNA.
  - Makes DNA more accessible to RNA polymerase II
- Enzymes that ACETYLATE HISTONES are recruited to actively transcribed genes.
- Enzymes that remove acetyl groups from histones are recruited to methylated DNA.

https://www.slideshare.net/NoeMendez9/eukaryotic-gene-regulation-models-by-np-mendez

Regulation of genes expression in eukaryotic cells (regulation of RNA transcription).

One of the most important mechanisms for the regulation of gene expression is the regulation of RNA transcription:

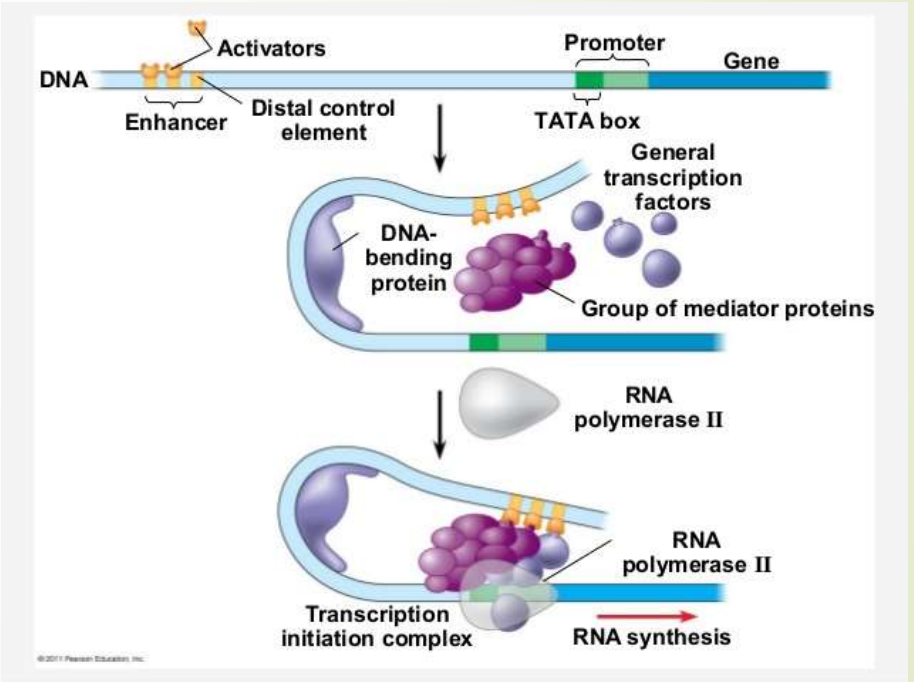
Protein-coding genes have:
exons - coding the polypeptide sequences;

introns - noncoding the polypeptide sequences that will be removed from the mRNA before it is translated;

promoter - is a sequence of DNA to which proteins bind that initiate transcription. Promoters are located near the transcription start sites of genes, upstream on the DNA (towards the 5' region of the sense strand). Promoters can be about 100–1000 base pairs long;

enhancer - is a short (50–1500 bp) region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur.These proteins are usually referred to as transcription factors. They can be located up to 1 Mbp (1,000,000 bp) away from the gene, upstream or downstream from the start site. There are hundreds of thousands of enhancers in the human genome;

silencer - is a DNA sequence capable of binding transcription regulation factors, called repressors. When a repressor protein binds to the silencer region of DNA, RNA polymerase is prevented from transcribing the DNA sequence into RNA. With transcription blocked, the translation of RNA into proteins is impossible. Thus, silencers prevent genes from being expressed as proteins.



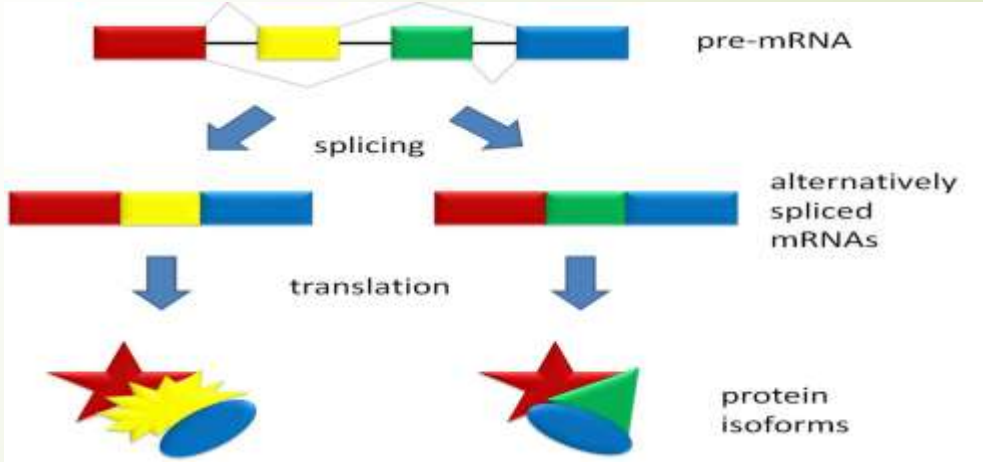
https://www.slideshare.net/NoeMendez9/eukaryotic-gene-regulation-models-by-np-mendez

Regulation of genes expression in eukaryotic cells (regulation of RNA processing).

- **RNA processing or post-transcriptional modification** is a set of biological processes common to most eukaryotic cells by which an RNA primary transcript is chemically altered following transcription from a gene to produce a mature, functional RNA molecule. For example, processing of mRNA includes three major steps : the addition of a 5' cap, the addition of a 3' polyadenylated tail, and RNA splicing.

- **Alternative splicing**, is a process during gene expression that allows a single gene to code for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final, processed messenger RNA (mRNA) produced from that gene. Consequently, the proteins translated from alternatively spliced mRNAs will contain differences in their amino acid sequence and, often, in their biological functions. The alternative splicing allows the human genome to direct the synthesis of many more proteins than would be expected from its 20,000 protein-coding genes.

- Alternative splicing occurs as a normal phenomenon in eukaryotes, where it greatly increases the biodiversity of proteins that can be encoded by the genome; in humans, ~95% of multi-exonic genes are alternatively spliced. There are numerous modes of alternative splicing observed, of which the most common is exon skipping. In this mode, a particular exon may be included in mRNAs under some conditions or in particular tissues, and omitted from the mRNA in others.

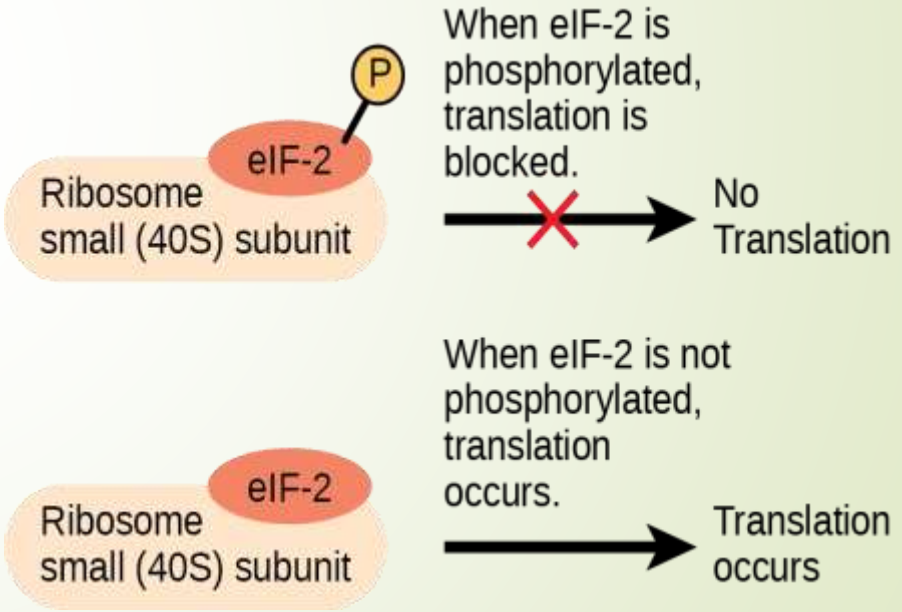


https://simple.wikipedia.org/wiki/Alternative_splicing

Regulation of genes expression in eukaryotic cells (regulation of translation)

- Translation is controlled by proteins that bind and initiate the process.
- In eukaryotes, translation is initiated by binding the initiating met-tRNAi to the 40S ribosome. This tRNA is brought to the 40S ribosome by a protein initiation factor, eukaryotic initiation factor-2 (eIF-2).
- The eIF-2 protein binds to the high-energy molecule guanosine triphosphate (GTP).
- The tRNA-eIF2-GTP complex then binds to the 40S ribosome. A second complex forms on the mRNA.
- Several different initiation factors recognize the 5′ cap of the mRNA and proteins bound to the poly-A tail of the same mRNA, forming the mRNA into a loop.
- The cap-binding protein eIF4F brings the mRNA complex together with the 40S ribosome complex.
- The ribosome then scans along the mRNA until it finds a start codon AUG.
- When the anticodon of the initiator tRNA and the start codon are aligned, the GTP is hydrolyzed, the initiation factors are released, and the large 60S ribosomal subunit binds to form the translation complex.

- The binding of eIF-2 to the RNA is controlled by phosphorylation. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded. When eIF-2 remains unphosphorylated, the initiation complex can form normally and translation can proceed.



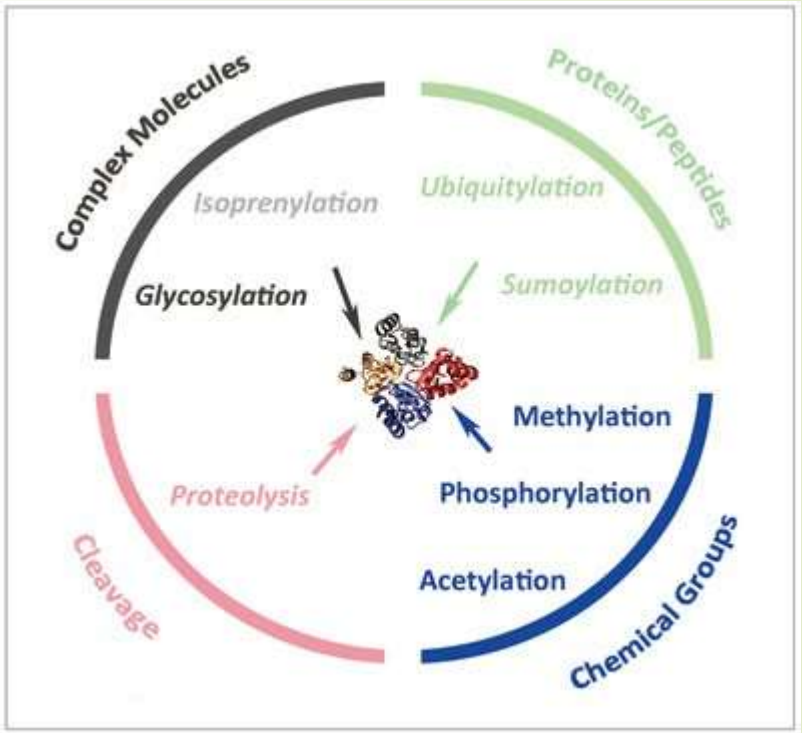
https://opentextbc.ca/biology2eopenstax/chapter/eukaryotic-translational-and-post-translational-gene-regulation/

Regulation of genes expression in eukaryotic cells (regulation of translation and posttranslation modification)

- **Post-translational modification (PTM)** refers to the covalent and generally enzymatic modification of proteins following protein biosynthesis. hormones.
- Post-translational modifications can occur on the amino acid side chains or at the protein's C- or N- termini.
- They can extend the chemical repertoire of the 20 standard amino acids by modifying an existing functional group or introducing a new:

- phosphorylation of amino acids is a very common mechanism for regulating the activity of enzymes and is the most common post-translational modification;
- glycosylation - carbohydrate molecules attached to amino acids;
- lipidation - aAttachment of lipid molecules.

- Other forms of post-translational modification consist of cleaving peptide bonds, as in processing a propeptide to a mature form or removing the initiator methionine residue.
- The formation of disulfide bonds from cysteine residues may also be referred to as a post-translational modification. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds.

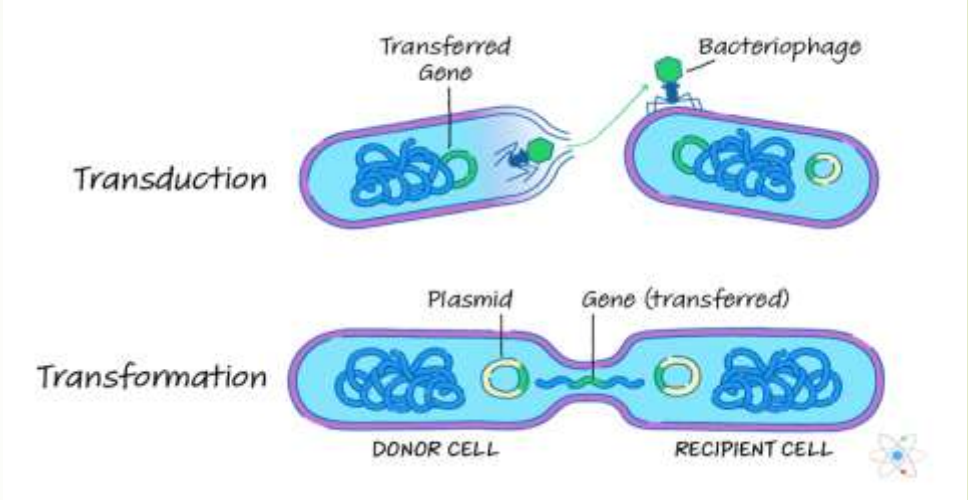


https://www.nature.com/articles/cr2013151

Genetic recombination

- Genetic recombination refers to the rearrangement of DNA sequences by the breakage and rejoining of chromosomes or chromosome segments. Genetic recombination is a programmed feature of meiosis in most sexual organisms, where it ensures the proper segregation of chromosomes.
- Genetic recombination is often used as a general term that includes many types of DNA rearrangements and underlying molecular processes. Meiotic recombination is an example of a reaction that involves DNA sequences that are paired and homologous over very extended lengths. This type of process is termed general, legitimate, or homologous recombination. Recombination of this type is reciprocal, because each participating chromosome receives information comparable to what it donates to the other partner.
- **Genetic recombination in prokaryotes:**

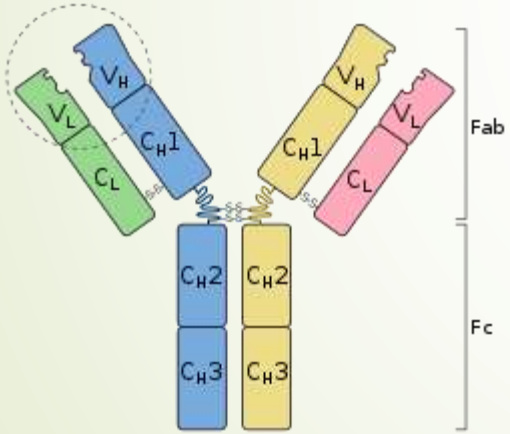
- **Transduction** is the transfer of DNA from one bacterium to another through the action of viruses.
- **Transformation** is the genetic alteration of a cell resulting from the direct uptake and incorporation of exogenous genetic material from its surroundings through the cell membrane.
- **Conjugation** is the bacterial equivalent of sex. It involves physical contact between two cells, possibly via a bridging structure called a pilus. Donor cells must contain a small DNA segment called the F-plasmid, which the recipient must lack. The donor cell provides a single strand of DNA from the F-plasmid and transfers it to the recipient.



https://sciencing.com/transformation-transduction-conjugation-gene-transfer-in-prokaryotes-13717688.html

Recombination of immunoglobulins

- An antibody unit consists of four polypeptide chains; two identical heavy chains and two identical light chains connected by disulfide bonds. Each chain is a series of domains: somewhat similar sequences of about 110 amino acids each. Light chains consist of one variable domain VL and one constant domain CL, while heavy chains contain one variable domain VH and three to four constant domains CH1, CH2.



https://en.wikipedia.org/wiki/Antibody

In mammals there are two types of immunoglobulin light chain, which are called lambda (λ) and kappa (κ). Tthere are five types of heavy chains - α, γ, δ, ε and μ, corresponding to the five isotypes of the antibody classes - IgA, IgG, IgD, IgE and IgM. Antibodies of each isotype differ from other functions and structural features. The colossal variability of antibodies is provided by rearrangements of loci encoding heavy and light chains during V (D) J recombination.

Recombination of immunoglobulins

The **recombination of immunoglobulins, also known as V (D) J** recombination, involves the generation of a unique immunoglobulin variable region.

The variable region of each immunoglobulin heavy or light chain is encoded in several pieces — known as gene segments (subgenes).

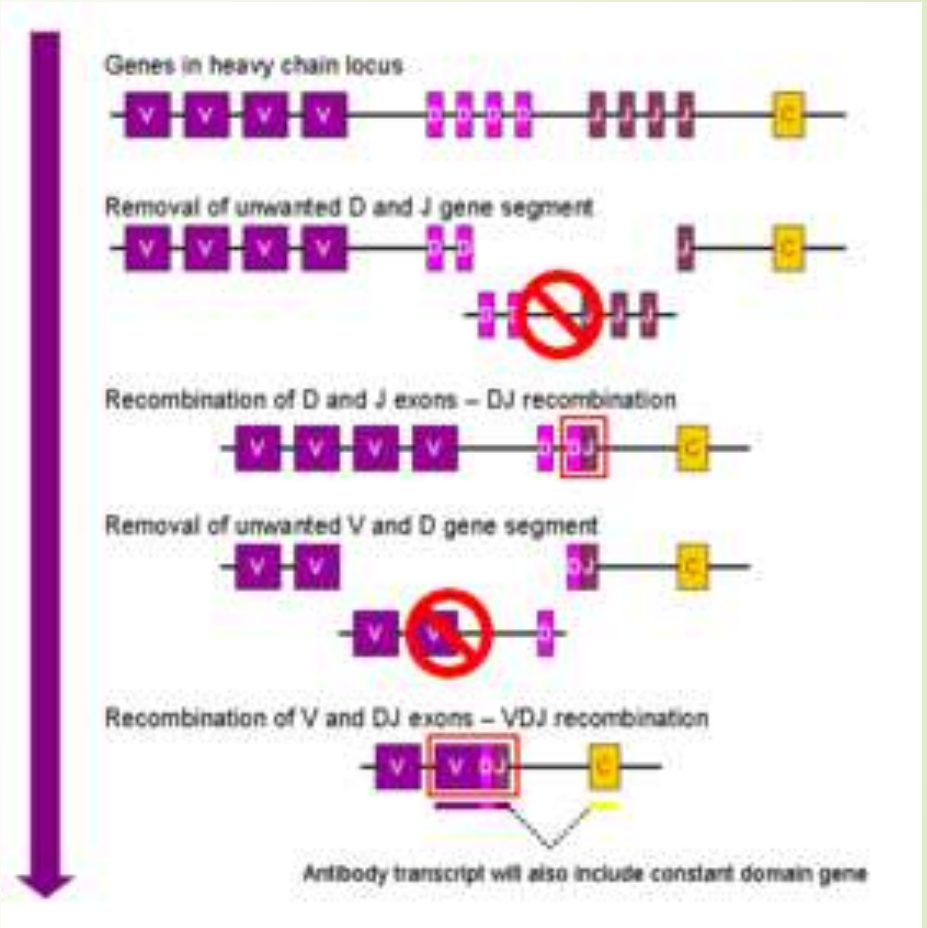
These segments are called variable (V), diversity (D) and joining (J) segments.

V, D and J segments are found in Ig heavy chains, but only V and J segments are found in Ig light chains.

Multiple copies of the V, D and J gene segments exist, and are tandemly arranged in the genomes of mammals.

In the bone marrow, each developing B cell will assemble an immunoglobulin variable region by randomly selecting and combining one V, one D and one J gene segment (or one V and one J segment in the light chain).

As there are multiple copies of each type of gene segment, and different combinations of gene segments can be used to generate each immunoglobulin variable region, this process generates a huge number of antibodies, each with different paratopes, and thus different antigen specificities.



Simplified overview of V(D)J recombination of immunoglobulin heavy chains

https://en.wikipedia.org/wiki/Antibody

Mutations

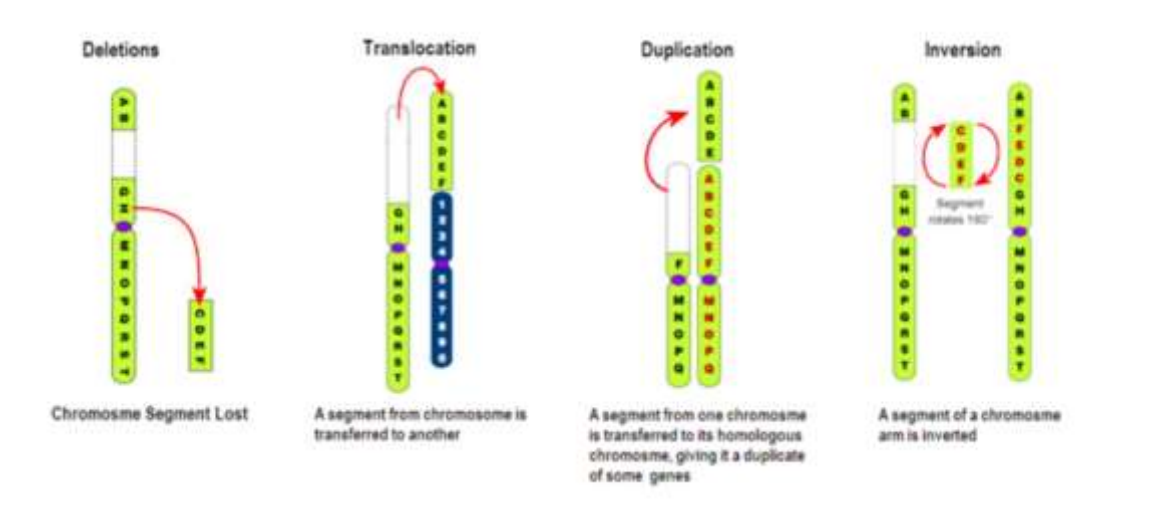
- **A mutation** is a change in a DNA sequence. Mutations can result from DNA copying mistakes made during cell division, exposure to ionizing radiation, exposure to chemicals called mutagens, or infection by viruses.
- Types of mutations:

- genomic
- chromosomal
- gene

Genomic mutations - change in the number of chromosomes.
Aneuploidy (heteroploidy) occurs in humans - a change in the number of chromosomes that is not a multiple of the haploid set. for example a human cell having 45 or 47 chromosomes instead of the usual 46. The most frequent aneuploidy in humans is **trisomy 16** and fetuses affected with the full version of this chromosome abnormality do not survive to term, although it is possible for surviving individuals to have the mosaic form, where trisomy 16 exists in some cells but not all. The most common aneuploidy that infants can survive with is trisomy 21, which is found in **Down syndrome**, affecting 1 in 800 births. Trisomy 18 (**Edwards syndrome**) affects 1 in 6,000 births, and trisomy 13 (**Patau syndrome**) affects 1 in 10,000 births.
Turner syndrome (**TS**), also known **45,X**, or **45,X0**, is a genetic condition in which a female is partly or completely missing an X chromosome. **Klinefelter syndrome** (**KS**), also known as **47,XXY** is the set of symptoms that result from two or more X chromosomes in males.

Chromosomal mutations

- Chromosomal mutations **(chromosome structure mutations)** are alterations that affect whole chromosomes and whole genes rather than just individual nucleotides.
- These mutations result from errors in cell division that cause a section of a chromosome to break off, be duplicated or move onto another chromosome.



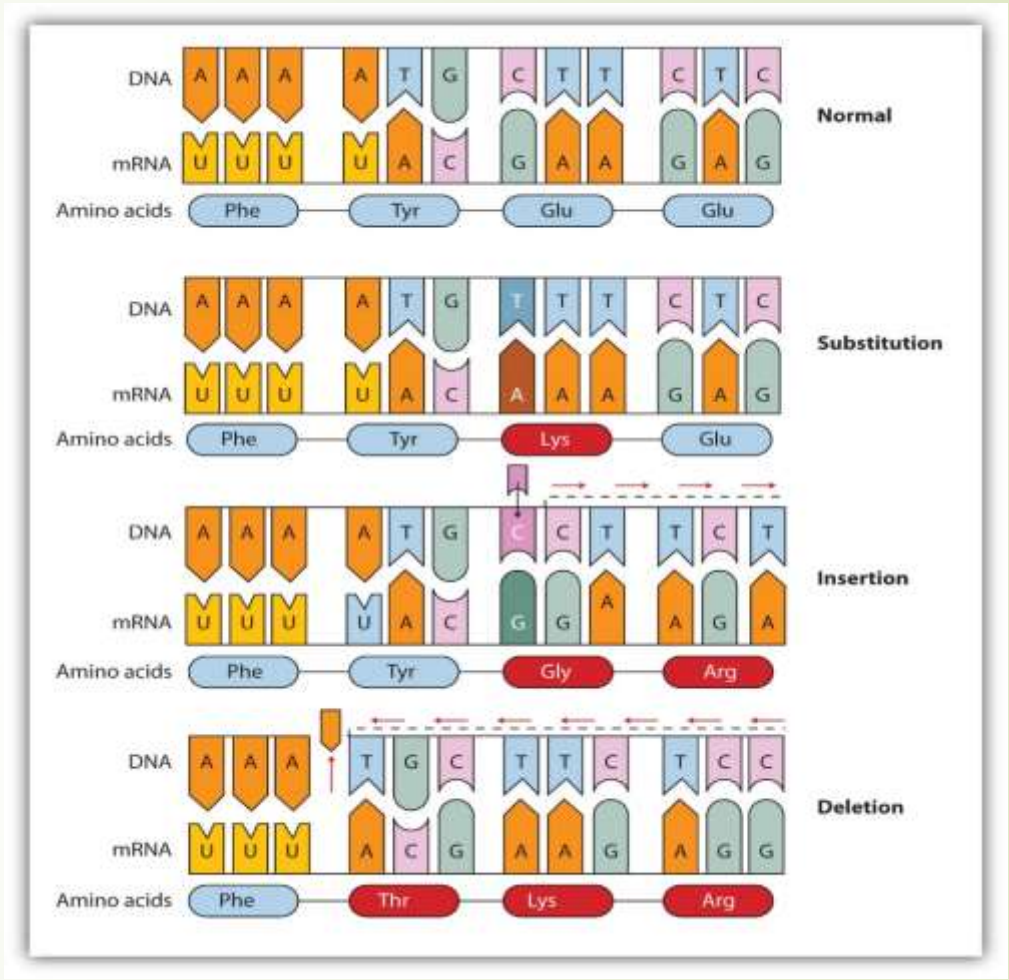
https://socratic.org/questions/what-are-four-types-of-chromosomal-mutations

- Types of chromosome structure mutations:
- **deletion** is where a section of a chromosome is removed.
- **translocation** is where a section of a chromosome is added to another chromosome that is not its homologous partner.
- **inversion** is where a section of a chromosome is reversed.
- **duplication** occurs when a section of a chromosome is added from its homologous partner.
- Chromosome mutations are often lethal as the chromosome structure is altered.

Gene mutations

- A gene mutation is a permanent alteration in the DNA sequence that makes up a gene.
- There are three types of DNA Mutations: **base substitutions,**
- **Base substitutions:**

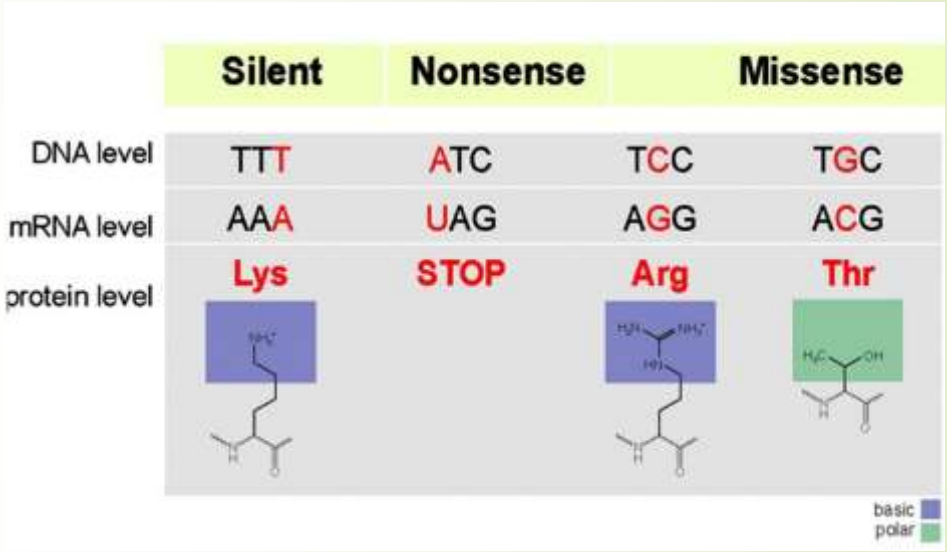
-Transition: this occurs when a purine is substituted with another purine or when a pyrimidine is substituted with another pyrimidine.
-Transversion: when a purine is substituted for a pyrimidine or a pyrimidine replaces a purine.
-Deletions - resulting in a frameshift, results when one or more base pairs are lost from the DNA. If one or two bases are deleted the translational frame is altered resulting in a garbled message and nonfunctional product. A deletion of three or more bases leave the reading frame intact. A deletion of one or more codons results in a protein missing one or more amino acids. This may be deleterious or not.
-Insertions - addition of base. It may lead to frameshifts depending on whether or not multiples of three base pairs are inserted. Combinations of insertions and deletions leading to a variety of outcomes are also possible.



https://saylordotorg.github.io/text_the-basics-of-general-organic-and-biological-chemistry/s22-05-mutations-and-genetic-diseases.html

Gene mutations

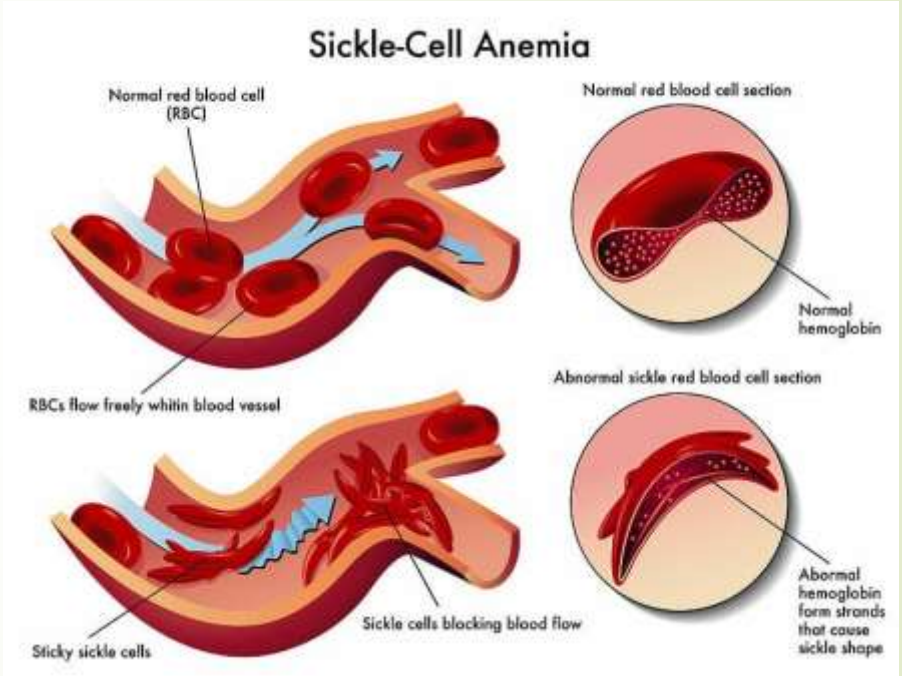
- Single base **substitutions** are called point mutations.
- Point mutations that occur in DNA sequences encoding proteins are either **silent, missense or nonsense.**
- **Silent mutation** if the mRNA codon codes for the same amino acid.
- **Missense mutation** if the mRNA codon codes for a different amino acid.
- **Nonsense mutation** if the mRNA codon becomes a stop codon (UAA, UAG, UGA).



https://www.scienceabc.com/pure-sciences/what-is-mutation-definition-different-types-biology-genetic-missense-nonsense.html

Sickle-cell anemia

- **Sickle-cell anemia** is caused by a point mutation in the β-globin chain of hemoglobin, replacing the amino acid **glutamic acid** with the less polar amino acid **valine** at the sixth position of the β chain. The mutation occurs in exon 1 and changes the nucleic acid sequence from GAG to GTG.



https://biologydictionary.net/missense-mutation/

Causes of Mutations

- **Errors in DNA Replication**
- On very, very rare occasions DNA polymerase will incorporate a noncomplementary base into the daughter strand. During the next round of replication the missincorporated base would lead to a mutation. This, however, is very rare as the exonuclease functions as a proofreading mechanism recognizing mismatched base pairs and excising them.
- **Errors in DNA Recombination**
- DNA often rearranges itself by a process called recombination which proceeds via a variety of mechanisms. Occasionally DNA is lost during replication leading to a mutation.
- **Chemical Damage to DNA**
- Many chemical mutagens, some exogenous, some man-made, some environmental, are capable of damaging DNA. Many chemotherapeutic drugs and intercalating agent drugs function by damaging DNA.
- **Radiation**
- Gamma rays, X-rays, even UV light can interact with compounds in the cell generating free radicals which cause chemical damage to DNA.

Types of mutagens

- **A mutagen** is a physical or chemical agent that changes the genetic material.
- **Types of mutagens**:
- **Physical:**

- Ionizing radiations such as X-rays, gamma ras and alpha particles cause DNA breakage and other damages.

- Ultraviolet radiations with wavelength above 260 nm are absorbed strongly by bases, producing pyrimidine dimers, which can cause error in replication if left uncorrected.

- Radioactive decay, such as ^{14}C in DNA which decays into nitrogen.

- **Biological**

- Transposon, a section of DNA that undergoes autonomous fragment relocation/multiplication. Its insertion into chromosomal DNA disrupts functional elements of the genes.

- Virus – Virus DNA may be inserted into the genome and disrupts genetic function.

- Bacteria – some bacteria such as *Helicobacter pylori* cause inflammation during which oxidative species are produced, causing DNA damage and reducing efficiency of DNA repair systems, thereby increasing mutation.

- **Chemical**

- Reactive oxygen species (ROS) – These may be superoxide, hydroxyl radicals and hydrogen peroxide, and large number of these highly reactive species are generated by normal cellular processes, for example as a by-products of mitochondrial electron transport, or lipid peroxidation. A number of mutagens may also generate these ROS. These ROS may result in the production of many base adducts, as well as DNA strand breaks and crosslinks.

- Deaminating agents, for example nitrous acid which can cause transition mutations by converting cytosine to uracil.

- Polycyclic aromatic hydrocarbon (PAH), when activated to diol-epoxides can bind to DNA and form adducts.

- Alkylating agents such as ethylnitrosourea. The compounds transfer methyl or ethyl group to bases or the backbone phosphate groups. Guanine when alkylated may be mispaired with thymine. Some may cause DNA crosslinking and breakages. Nitrosamines are an important group of mutagens found in tobacco, and may also be formed in smoked meats and fish via the interaction of amines in food with nitrites added as preservatives.

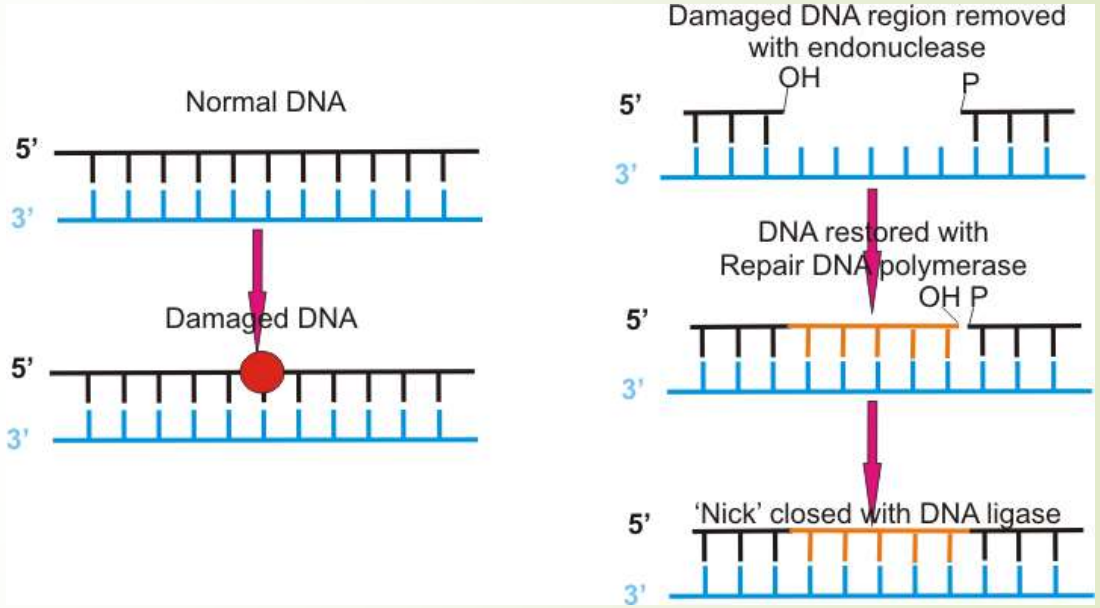
-Alkaloid from plants, such as those from Vinca species, may be converted by metabolic processes into the active mutagen or carcinogen.

DNA Repair

- **DNA repair** is a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome.

Each of the repair systems includes the following components:

- DNA helicase - an enzyme that "recognizes" chemically altered regions in the chain and breaks the chain near the damage;
- DNAase (deoxyribonuclease) is an enzyme that "cuts" 1 DNA strand (nucleotide sequence) at the phosphodiester bond and removes the damaged area: exonuclease works for terminal nucleotides 3` or 5`;
- endonuclease works for nucleotides other than terminal ones;
- DNA polymerase - an enzyme that synthesizes the corresponding section of the DNA chain instead of the deleted one; DNA ligase is an enzyme that closes the last bond in the polymer chain and thereby restores its continuity.

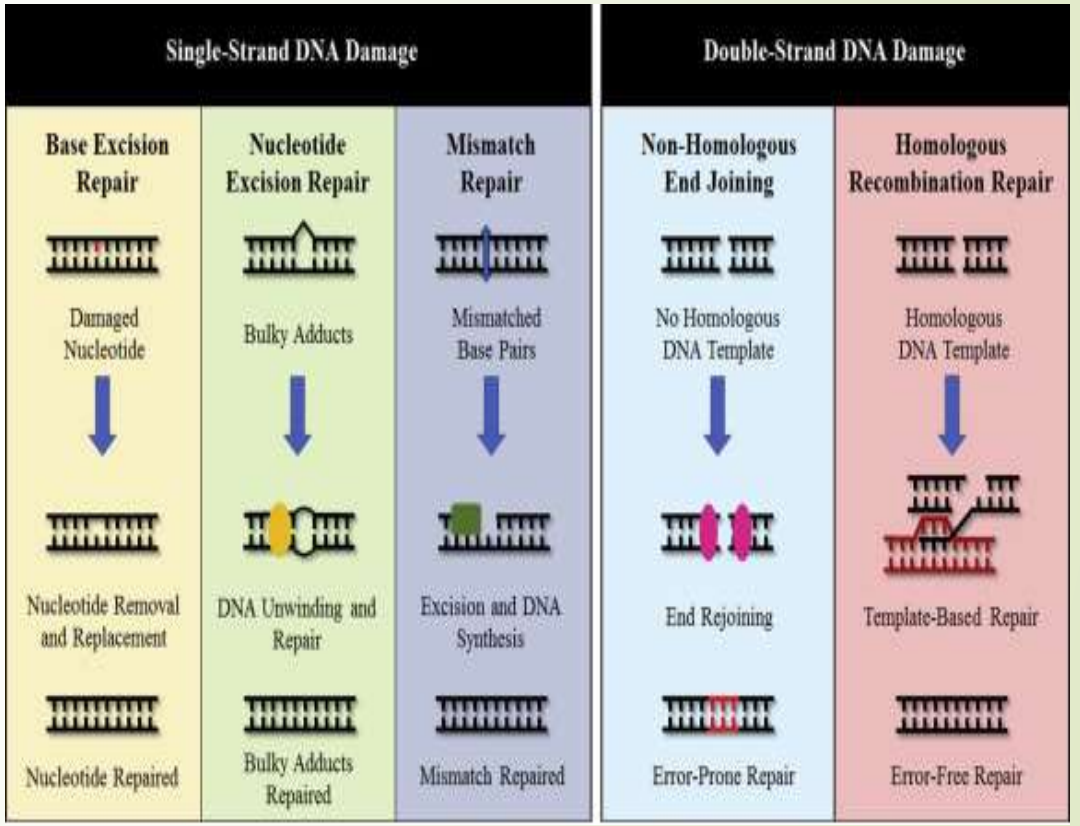


https://uk.m.wikipedia.org/wiki/Файл:Dna_repair.png

DNA Repair

- There are three major DNA repairing mechanisms: base excision, nucleotide excision and mismatch repair.

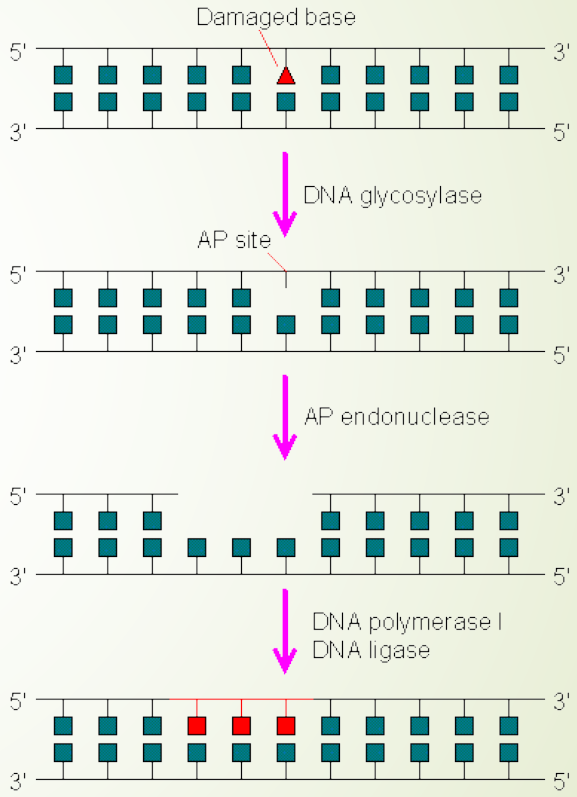
Repair System	Enzymes/proteins	Repair System	Enzymes/proteins
Base excision	DNA glycosylase	Mismatch	Dam methylase
	AP endonuclease		MutS, MutL, MutH
	DNA polymerase I		Exonuclease
	DNA ligase		DNA helicase II
Nucleotide excision	Uvr-A, Uvr-B, Uvr-C		SSB protein
	DNA polymerase I		DNA polymerase III
	DNA ligase		DNA ligase



https://www.clinicaloncologyonline.net/article/S0936-6555(19)30086-X/abstract

Base excision

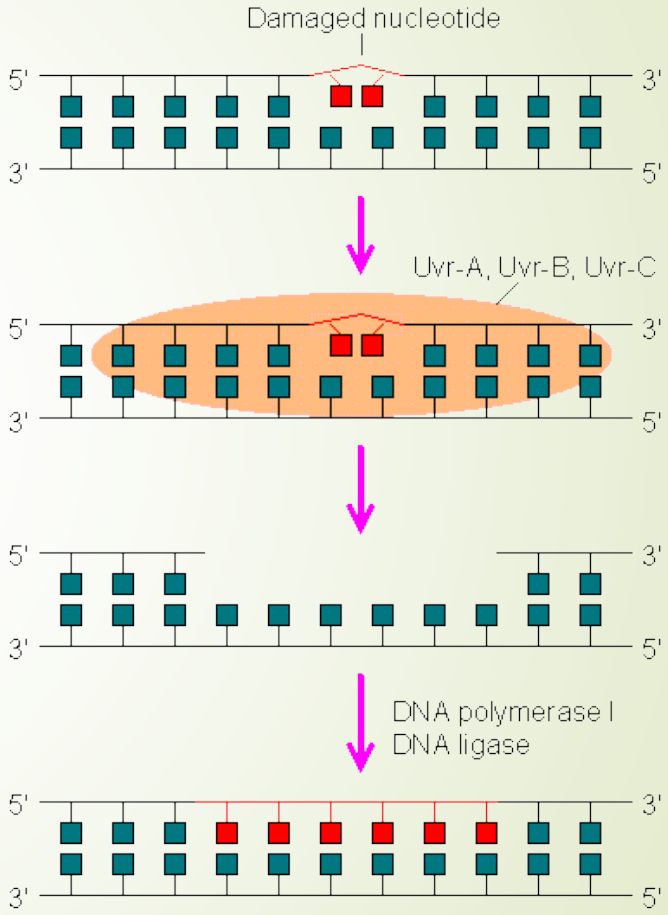
- DNA's bases may be modified by deamination or alkylation. The position of the modified (damaged) base is called the "abasic site" or "**AP site**". In *E.coli*, the DNA glycosylase can recognize the AP site and remove its base. Then, the AP endonuclease removes the AP site and neighboring nucleotides. The gap is filled by DNA polymerase I and DNA ligase.



https://www.web-books.com/MoBio/Free/Ch7G.htm

Nucleotide excision

- In *E. coli*, proteins UvrA, UvrB, and UvrC are involved in removing the damaged nucleotides (e.g., the dimer induced by UV light). The gap is then filled by DNA polymerase I and DNA ligase. In yeast, the proteins similar to Uvr's are named **RADxx** ("RAD" stands for "radiation"), such as RAD3, RAD10. etc.

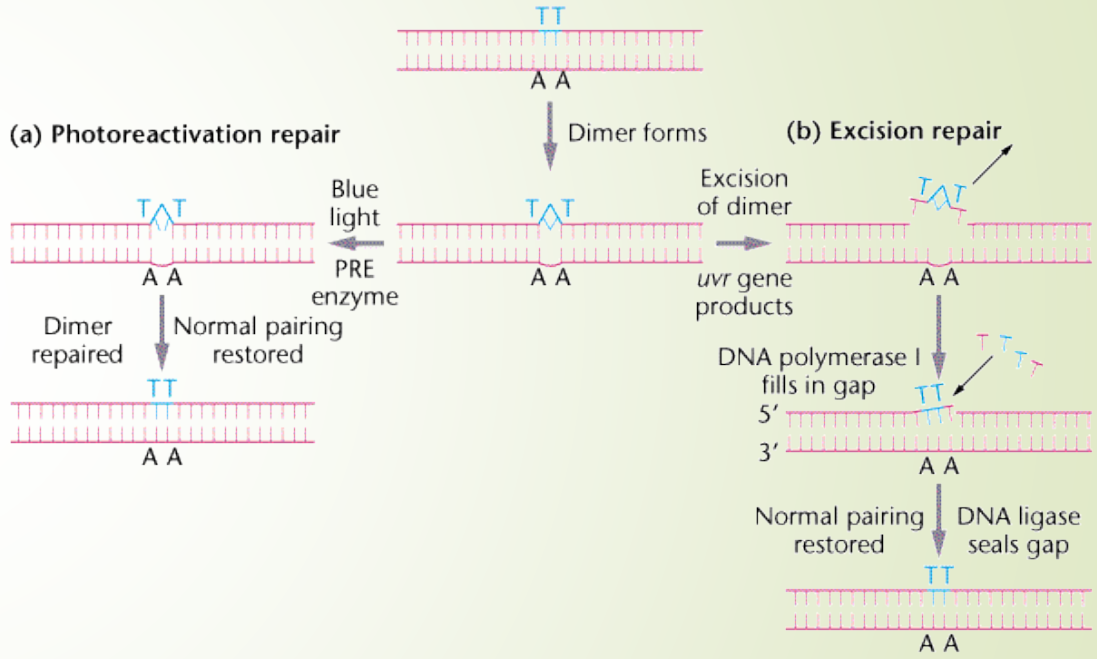


https://www.web-books.com/MoBio/Free/Ch7G.htm

Nucleotide Excision Repair

- Damage to DNA by UV radiation has the potential to cause cell death due to the formation of dimers, 6-4 PP & CPD's, and mutations that disable the DNA from replicating. This potential for cell death outlines the importance and need for an efficient repair system that is able to either reverse the effects or replace the damaged DNA. The two major repair systems used throughout the animal kingdom are either Nucleotide Excision Repair or Photoreactivation.

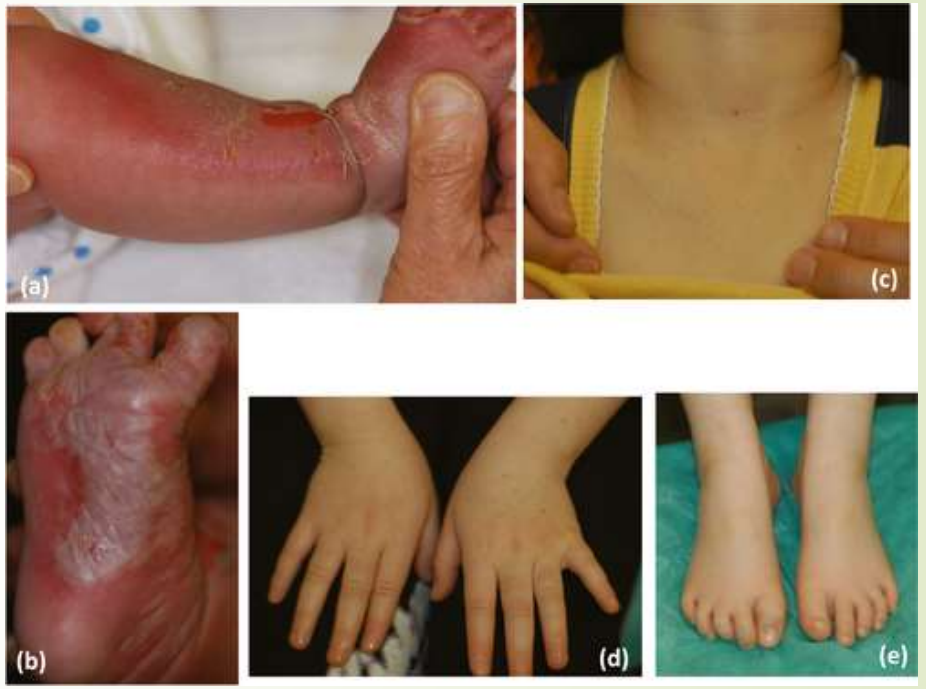
Nucleotide Excision Repair requires various proteins and also other enzymes, in order to remove the damaged strand, replace with correct bases, and seal a new strand into place. Photoreactivation however only requires one enzyme, photolyase, which doesnt remove the bases and require DNA polymerase, but simply reverses the effects with the use of visible light. This ability to reverse mutations makes Photoreactivation the much more efficient repair process. By utilising energy from visible light, the photoreactivation repairing system is very efficient in repairing effects from UV radiation caused by the sun, as the damage and also the energy needed is provided by the sun.



https://www.web-books.com/MoBio/Free/Ch7G.htm

Xeroderma pigmentosum

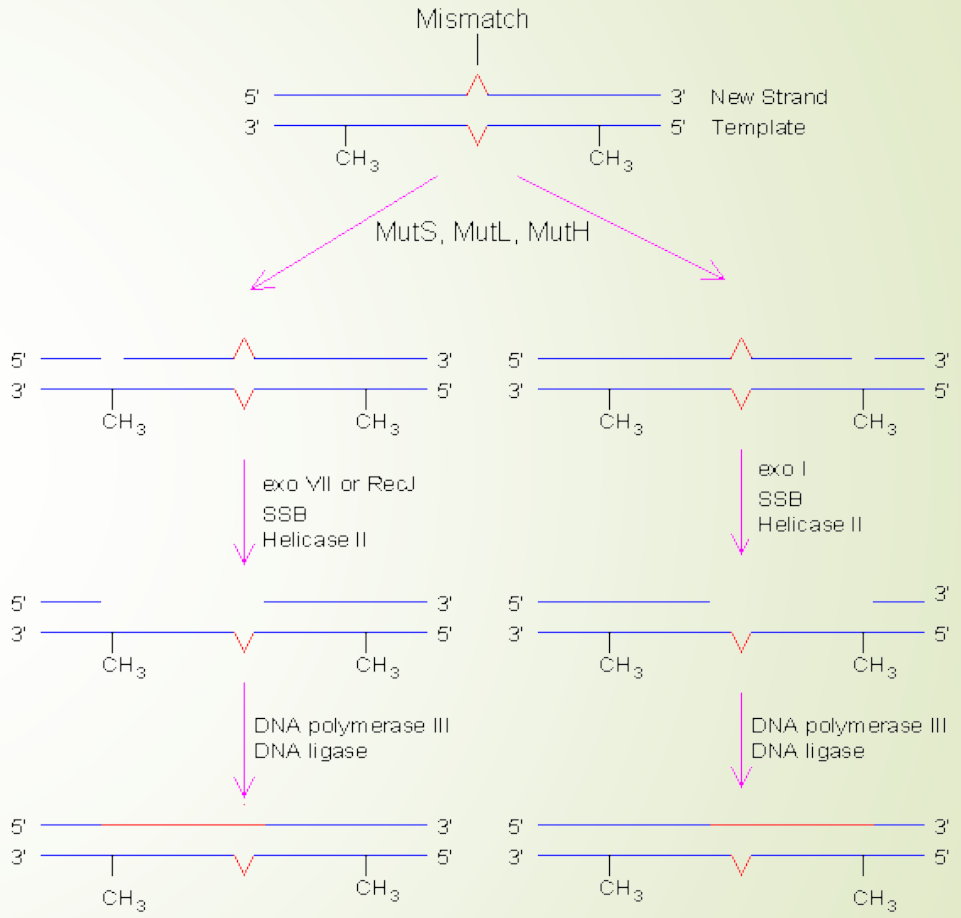
- Xeroderma pigmentosum (XP) is a genetic disorder in which there is a decreased ability to repair DNA damage such as that caused by ultraviolet (UV) light.
- Symptoms may include a severe sunburn after only a few minutes in the sun, freckling in sun exposed areas, dry skin and changes in skin pigmentation.
- Nervous system problems, such as hearing loss, poor coordination, loss of intellectual function and seizures, may also occur.
- Complications include a high risk of skin cancer, with about half having skin cancer by age 10 without preventive efforts, and cataracts.
- There may be a higher risk of other cancers such as brain cancers.



https://onlinelibrary.wiley.com/doi/full/10.1111/php.13052

Mismatch repair

- To repair mismatched bases, the system has to know which base is the correct one. In *E. coli*, this is achieved by a special methylase called the "Dam methylase", which can methylate all adenines that occur within (5')GATC sequences. Immediately after DNA replication, the template strand has been methylated, but the newly synthesized strand is not methylated yet. Thus, the template strand and the new strand can be distinguished.
- The repairing process begins with the protein MutS which binds to mismatched base pairs. Then, MutL is recruited to the complex and activates MutH which binds to GATC sequences. Activation of MutH cleaves the unmethylated strand at the GATC site. Subsequently, the segment from the cleavage site to the mismatch is removed by exonuclease (with assistance from helicase II and SSB proteins). If the cleavage occurs on the 3' side of the mismatch, this step is carried out by exonuclease I (which degrades a single strand only in the 3' to 5' direction). If the cleavage occurs on the 5' side of the mismatch, exonuclease VII or RecJ is used to degrade the single stranded DNA. The gap is filled by DNA polymerase III and DNA ligase.
- The distance between the GATC site and the mismatch could be as long as 1,000 base pairs. Therefore, mismatch repair is very expensive and inefficient.
- Mismatch repair in eukaryotes may be similar to that in *E. coli*. Homologs of MutS and MutL have been identified in yeast, mammals, and other eukaryotes. MSH1 to MSH5 are homologous to MutS; MLH1, PMS1 and PMS2 are homologous to MutL. Mutations of MSH2, PMS1 and PMS2 are related to colon cancer.
- In eukaryotes, the mechanism to distinguish the template strand from the new strand is still unclear.



https://www.web-books.com/MoBio/Free/Ch7G.htm

Sources of information

- https://ru.m.wikipedia.org/wiki/Файл:Human_genome_to_genes.png
- https://en.wikipedia.org/wiki/Gene_structure
- https://en.wikipedia.org/wiki/Gene_expression
- https://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/72-transcription-and-gene/operons.html
- https://opentextbc.ca/biology2eopenstax/chapter/prokaryotic-gene-regulation/#:~:text=The%20regulation%20of%20gene%20expression,prokaryotic%20transcription%3A%20repressors%20and%20activators.&text=In%20the%20trp%20operon%2C%2
- http://kocw-n.xcache.kinxcdn.com/data/document/2019/gachon/nammyeongjin0109/5.pdf
- https://www.slideshare.net/NoeMendez9/eukaryotic-gene-regulation-models-by-np-mendez
- https://simple.wikipedia.org/wiki/Alternative_splicing
- https://opentextbc.ca/biology2eopenstax/chapter/eukaryotic-translational-and-post-translational-gene-regulation
- https://www.nature.com/articles/cr2013151
- https://sciencing.com/transformation-transduction-conjugation-gene-transfer-in-prokaryotes-13717688.html
- https://en.wikipedia.org/wiki/Antibody
- https://socratic.org/questions/what-are-four-types-of-chromosomal-mutations
- https://saylordotorg.github.io/text_the-basics-of-general-organic-and-biological-chemistry/s22-05-mutations-and-genetic-diseases.html
- https://www.scienceabc.com/pure-sciences/what-is-mutation-definition-different-types-biology-genetic-missense-nonsense.html
- https://biologydictionary.net/missense-mutation
- https://www.clinicaloncologyonline.net/article/S0936-6555(19)30086-X/abstract
- https://onlinelibrary.wiley.com/doi/full/10.1111/php.13052
- https://www.web-books.com/MoBio/Free/Ch7G.htm

- Biological and Bioorganic Chemistry: in 2 books. Book 2. Biological Chemistry /Yu.Gubsky, I.V. Nizhenkovska, M.M. Korda et al. ; edited by Yu.Gubsky, I.V.Nizhenkovska. – Kyiv:AUS Medicine Publishing, 2020.- 544 p.ISBN 978-617-505-785-8
- Halkerston I.D.K. Biochemistry: 2nd edition. The National medical series for independent study / Halkerston I.D.K. - 1988. - 522 p.
- Harper`s Biochemistry. R.K.Murray, D.K.Granner, P.A.Mayes, V.W.Rodwell. Prentice-Hall International Inc., 2010. – 1134 p.
- Gubsky Yu. Biological chemistry: textbook. – Vinnytsia: Nova Knyha, 2017. – 488 p.
- Koolman J. Color Atlas of Biochemistry / J.Koolman, K.-H. Rom. – Stuttgart. New York. – Thieme Verlag. — 1996. – 435 p.
- Lehninger A. Principles of Biochemistry / Lehninger A. – New York. – W.H.Freeman and Company. – 2005. – 1010 p.
- Pamela C.Champe Lippincott’s Illustrated Reviews: Biochemistry, 3rd Edition / Pamela C.Champe and Richard A.Harvey. – Baltimore, Lippincott Williams & Wilkins, MD ©, 2005. – 534p.