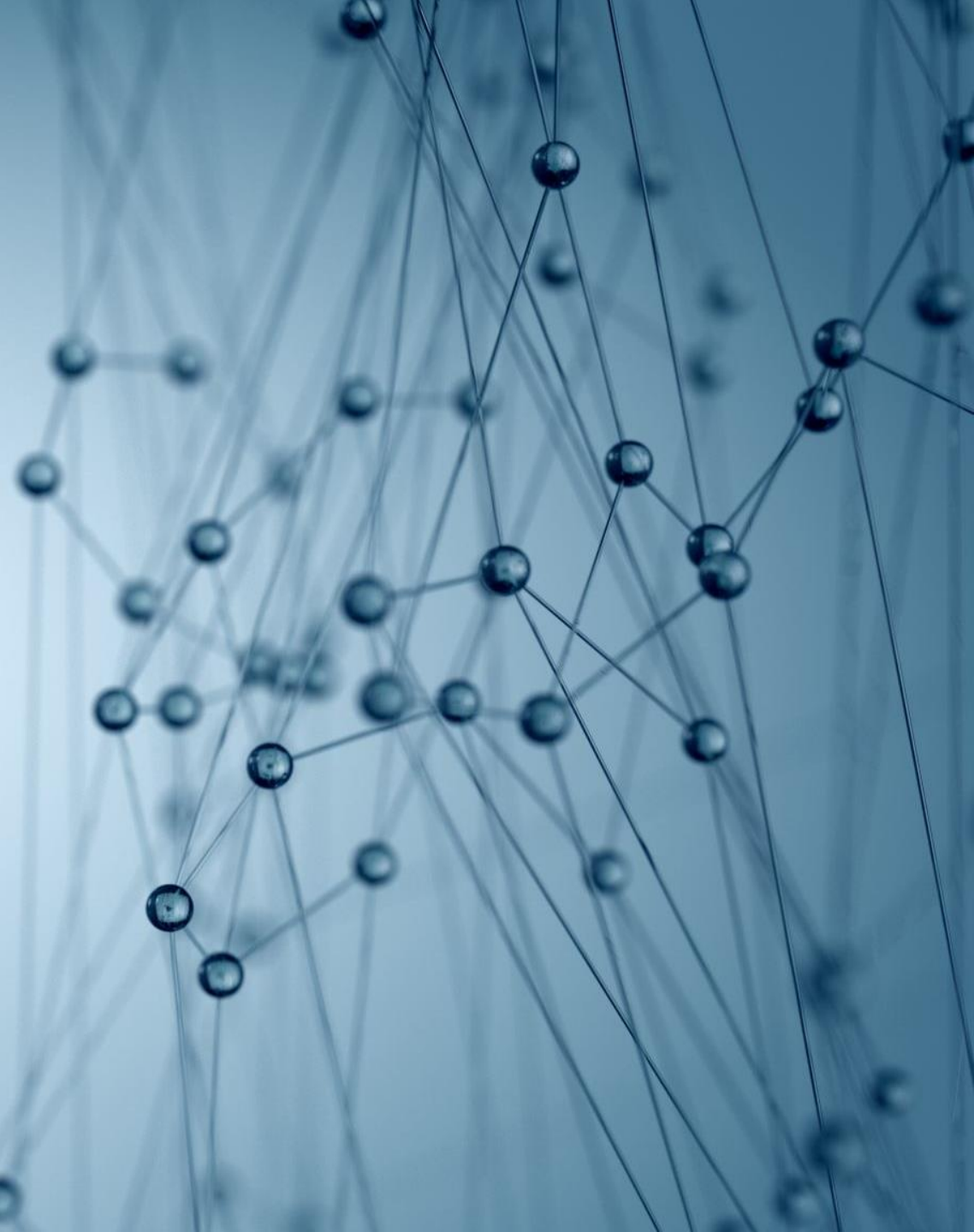


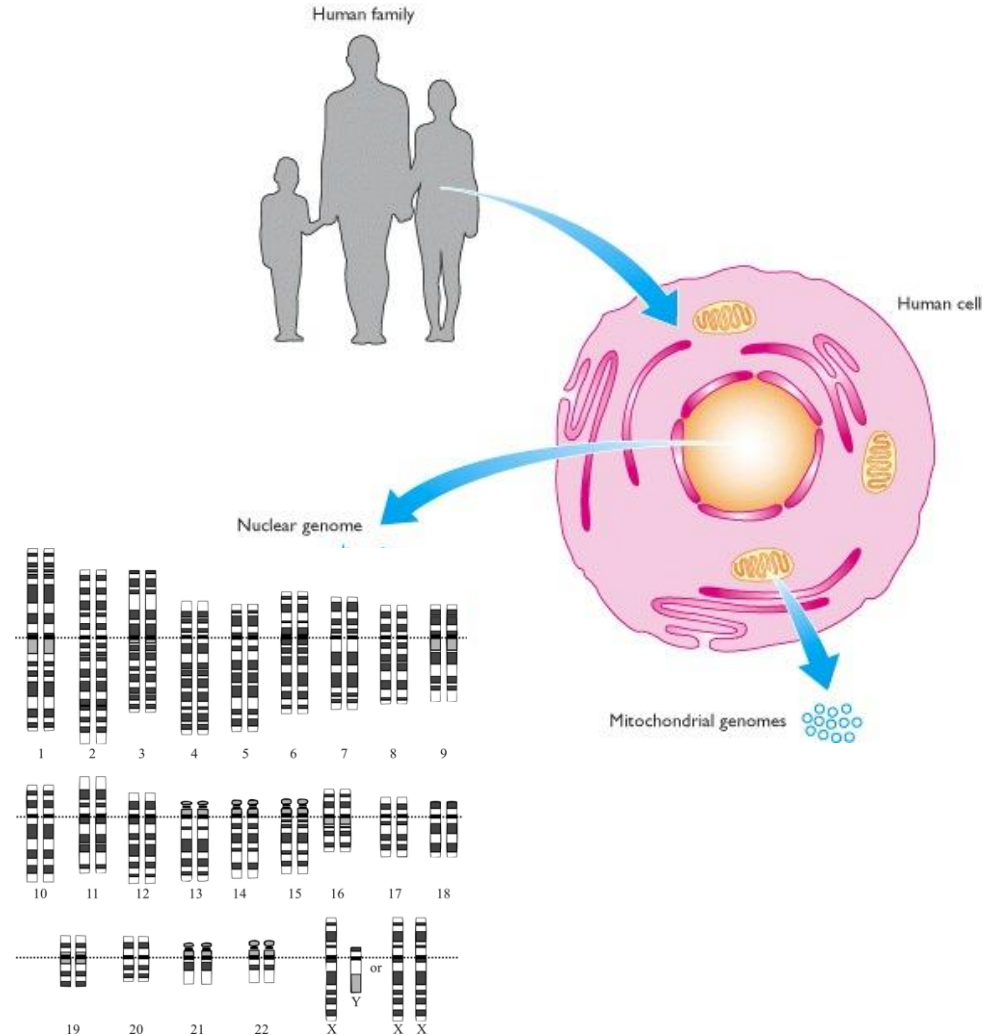
# Structure of genome and its interaction with environmental factors Epigenetics

# Genome structure

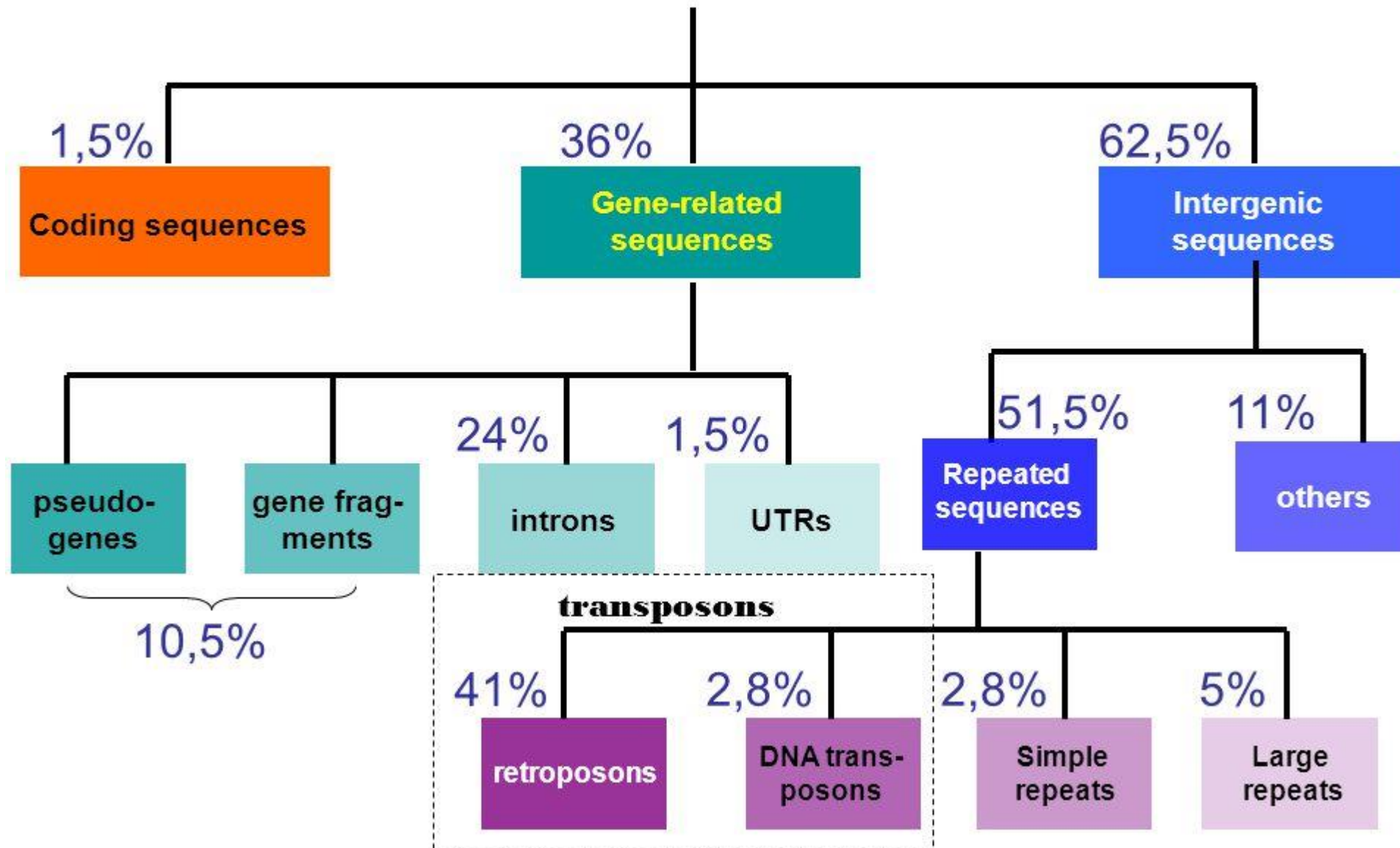


# Human genome

- Humans have two genomes, **nuclear** and **mitochondrial**.
- **Normal diploid cells** contain two copies of the nuclear genome and a much larger but variable number of copies of the mitochondrial genome.
- **23 chromosome pairs** in cell nuclei. In a pair of chromosomes, one chromosome is always inherited from the mother and one from the father.
- Small DNA molecules found within individual mitochondria. Each mitochondrion contains 2-10 mitochondrial DNA copies.



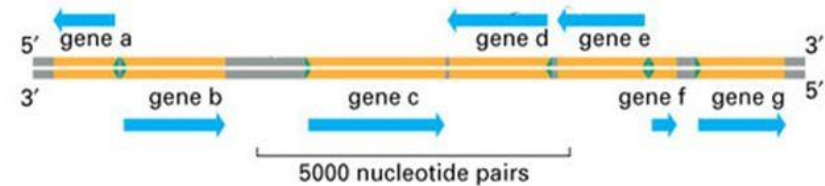
# Human genome



# Structure of protein-coding genes

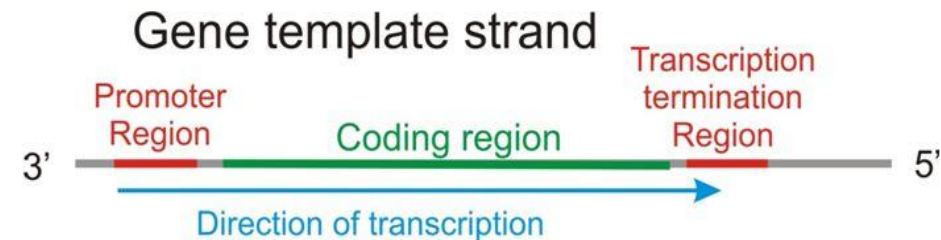
- **each gene** has its own unique position or **locus** on chromosome.
- The **different versions** of the gene are known as **alleles**.

## How are genes arranged on DNA?



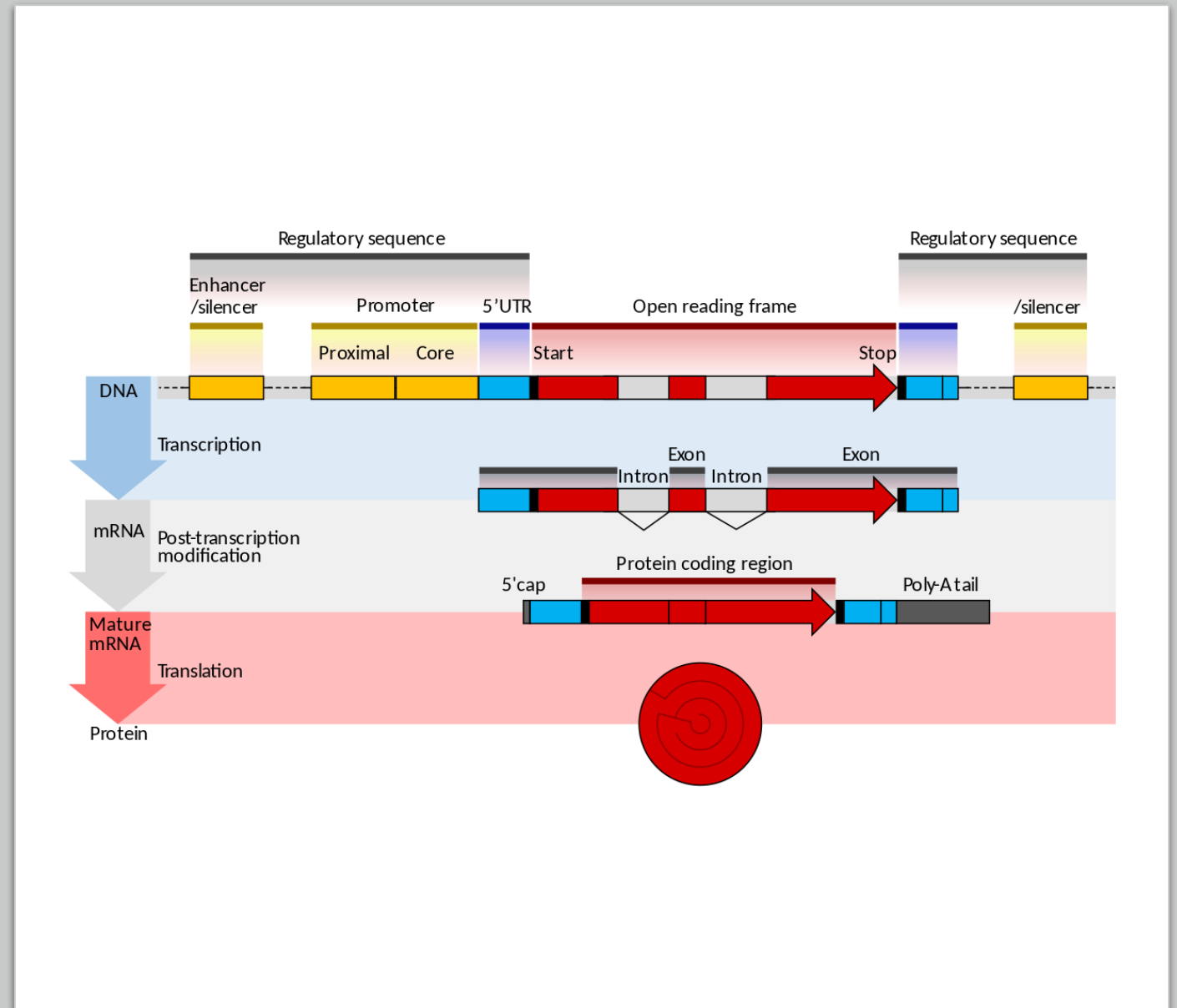
Template strand vs complementary strand

## What are the functional regions of a gene?



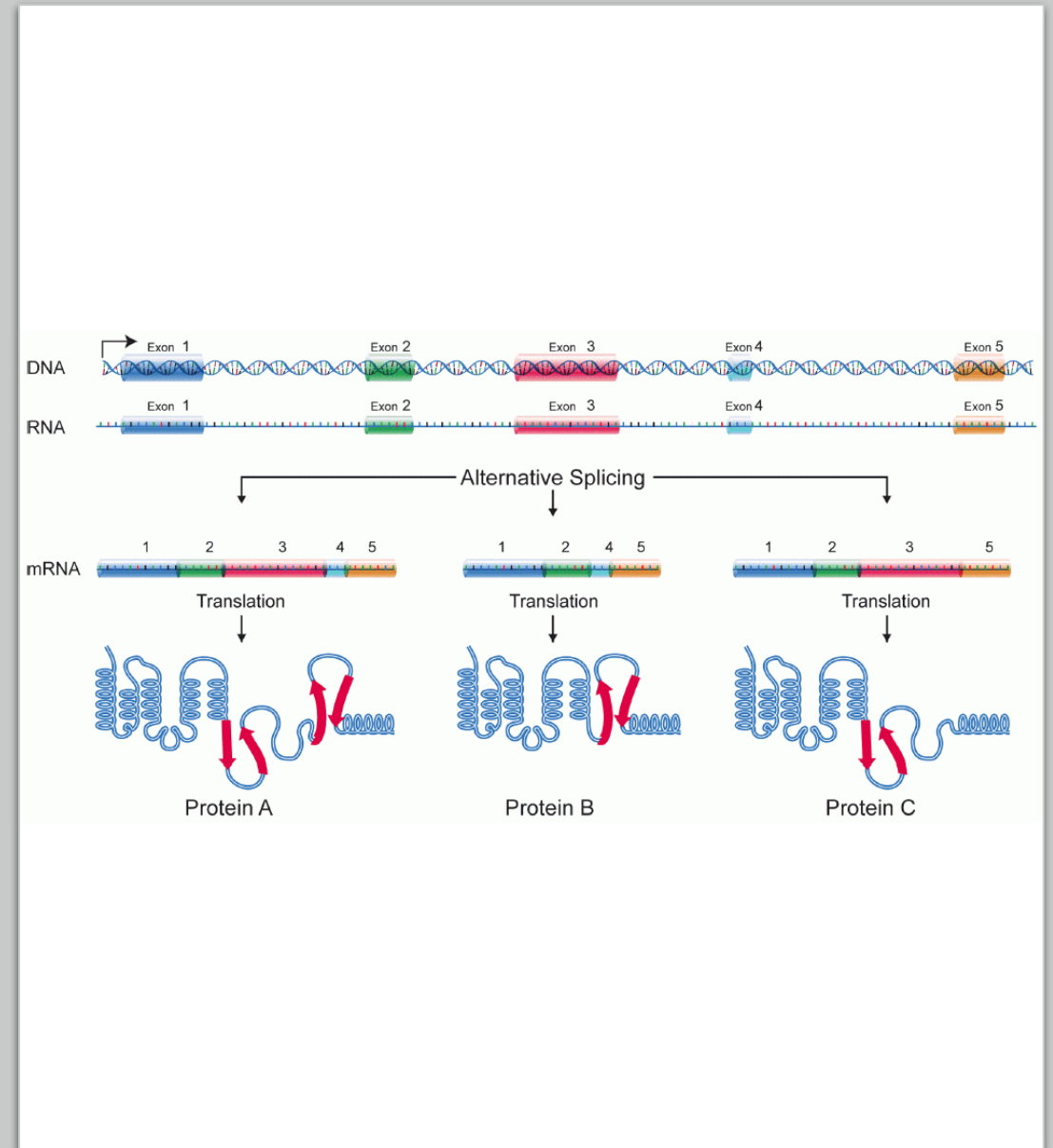
# Structure of protein-coding genes

- **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 remove introns (light grey) and add a **5' cap** and **poly-A tail** (dark grey).
- The mRNA **5' and 3' untranslated regions** (UTR; blue) regulate translation into the final protein product.



# Alternative splicing

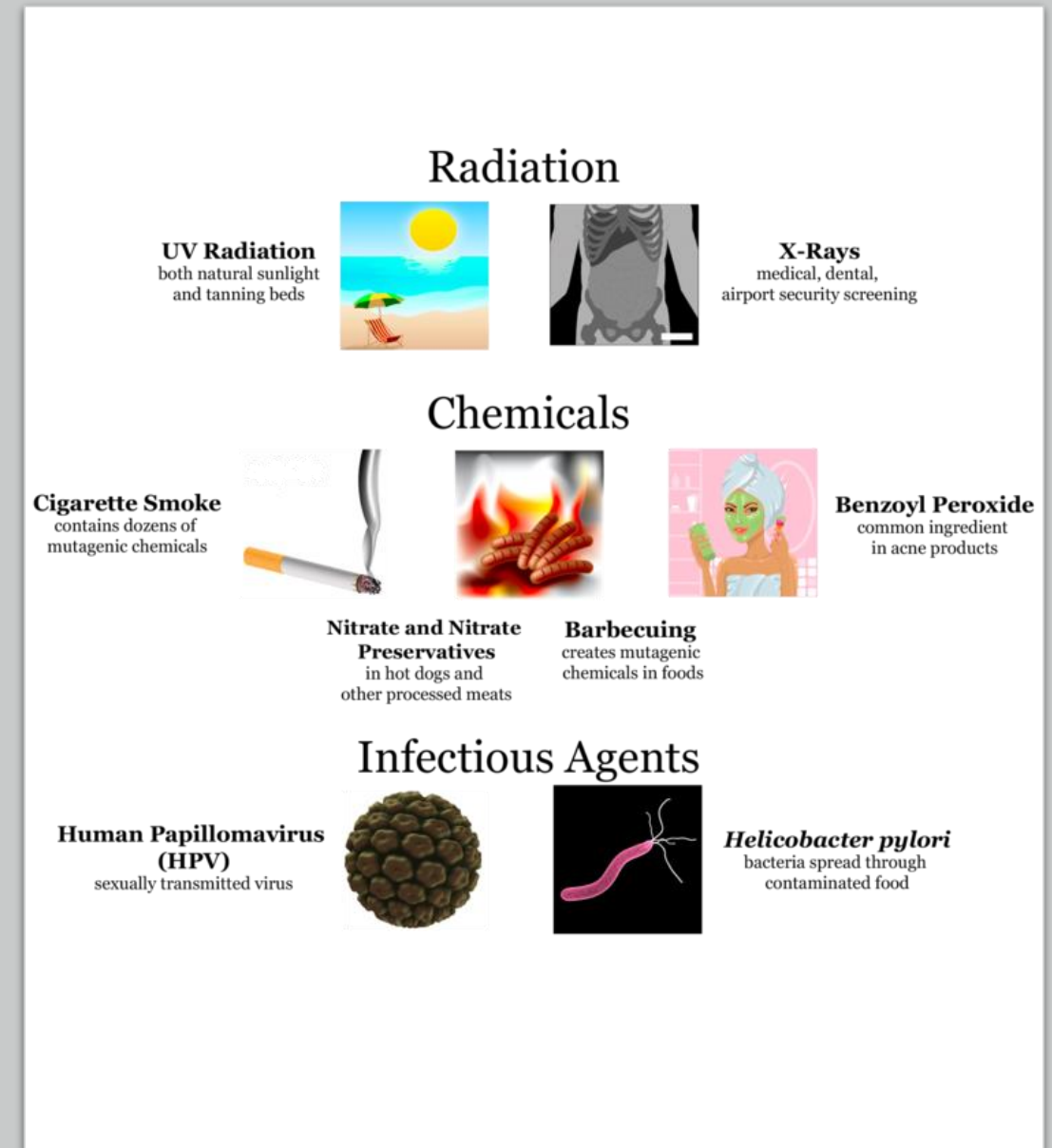
- Alternative splicing is a process that **enables** a messenger RNA (mRNA) to direct **synthesis** of different protein variants (isoforms) that may have different cellular functions or properties.
- It occurs by rearranging the pattern of intron and exon elements that are joined by splicing to alter the mRNA coding sequence.





# Mutations

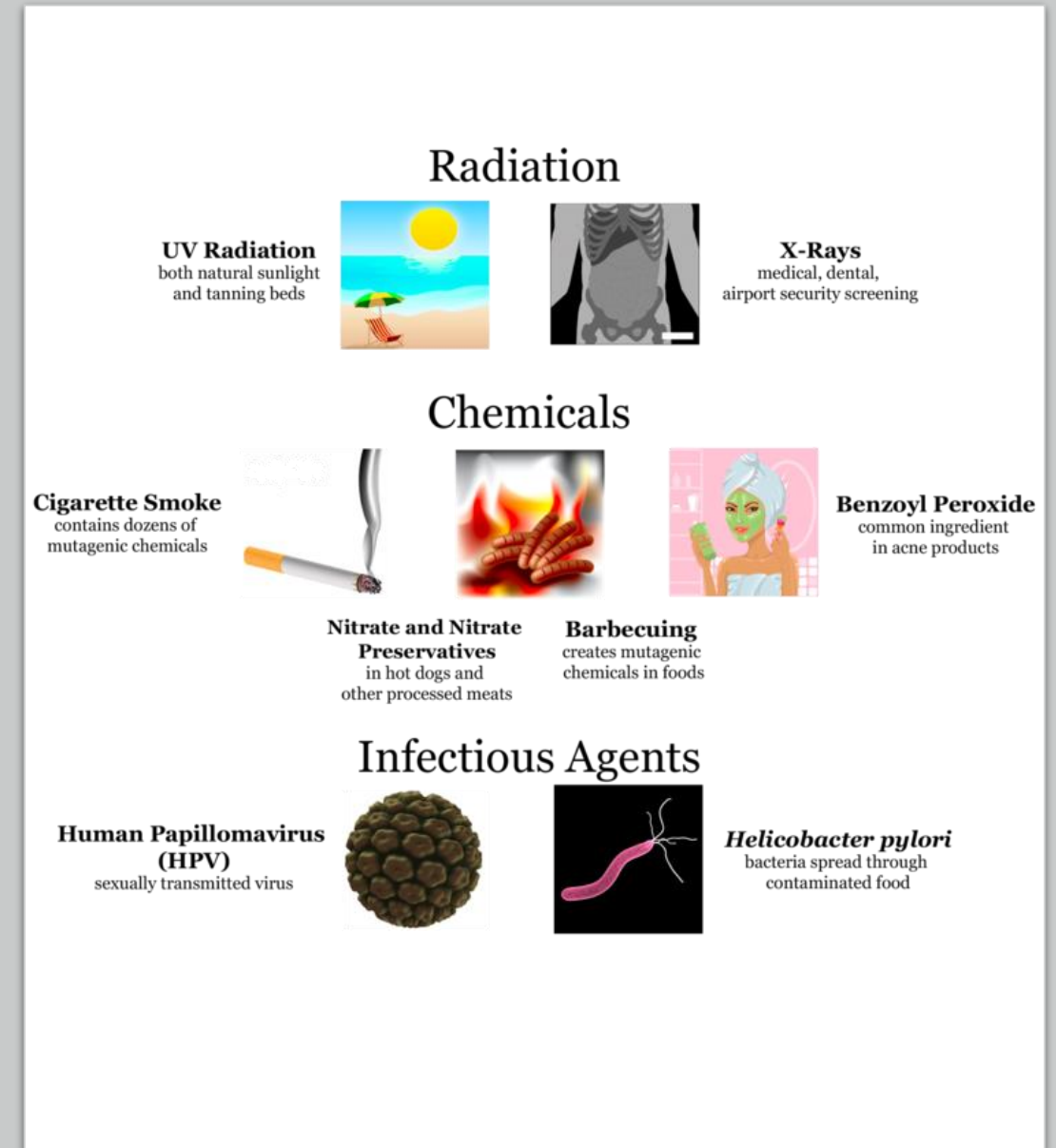
- A genetic mutation is a permanent alteration in the DNA sequence (DNA sequence differs from what is found in most people).
- Mutations range in size; they can affect anywhere from a single DNA building block (base pair) to a large segment of a chromosome that includes multiple genes (point mutations, chromosomal mutations, copy number variation).





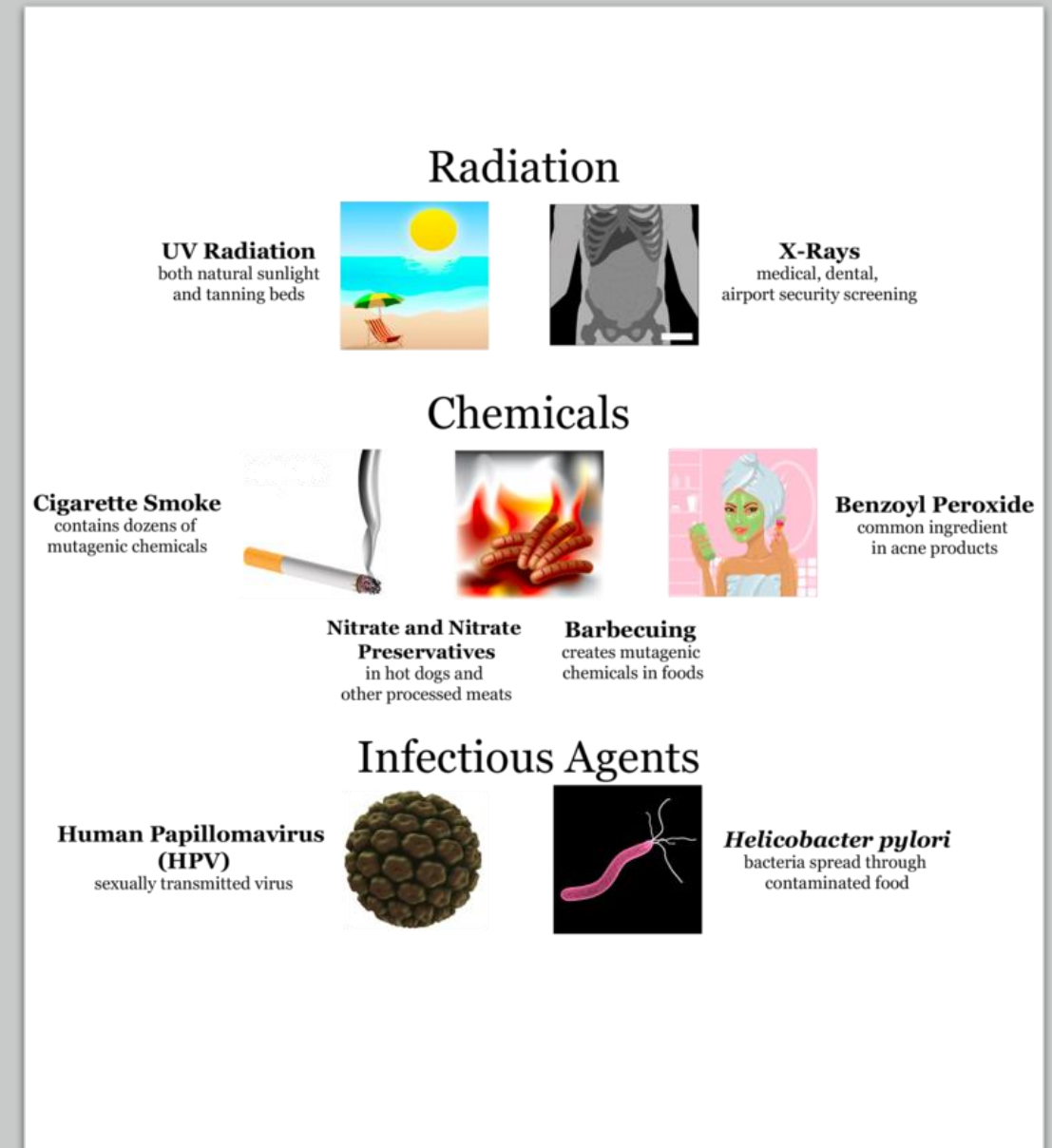
# Mutations

- **Hereditary mutations** are inherited **from a parent** and are present in virtually every cell in the body. These mutations are also called germline mutations because they are present in the parent's egg or sperm cells.
- **Acquired** (or somatic) **mutations** occur at some time **during** a person's **life** and are present only in certain cells, not in every cell in the body. These changes can be caused by environmental factors such as ultraviolet radiation, ionizing radiation, chemicals, or can occur if an error is made as DNA copies itself during cell division. Acquired mutations in somatic cells cannot be passed to the next generation.



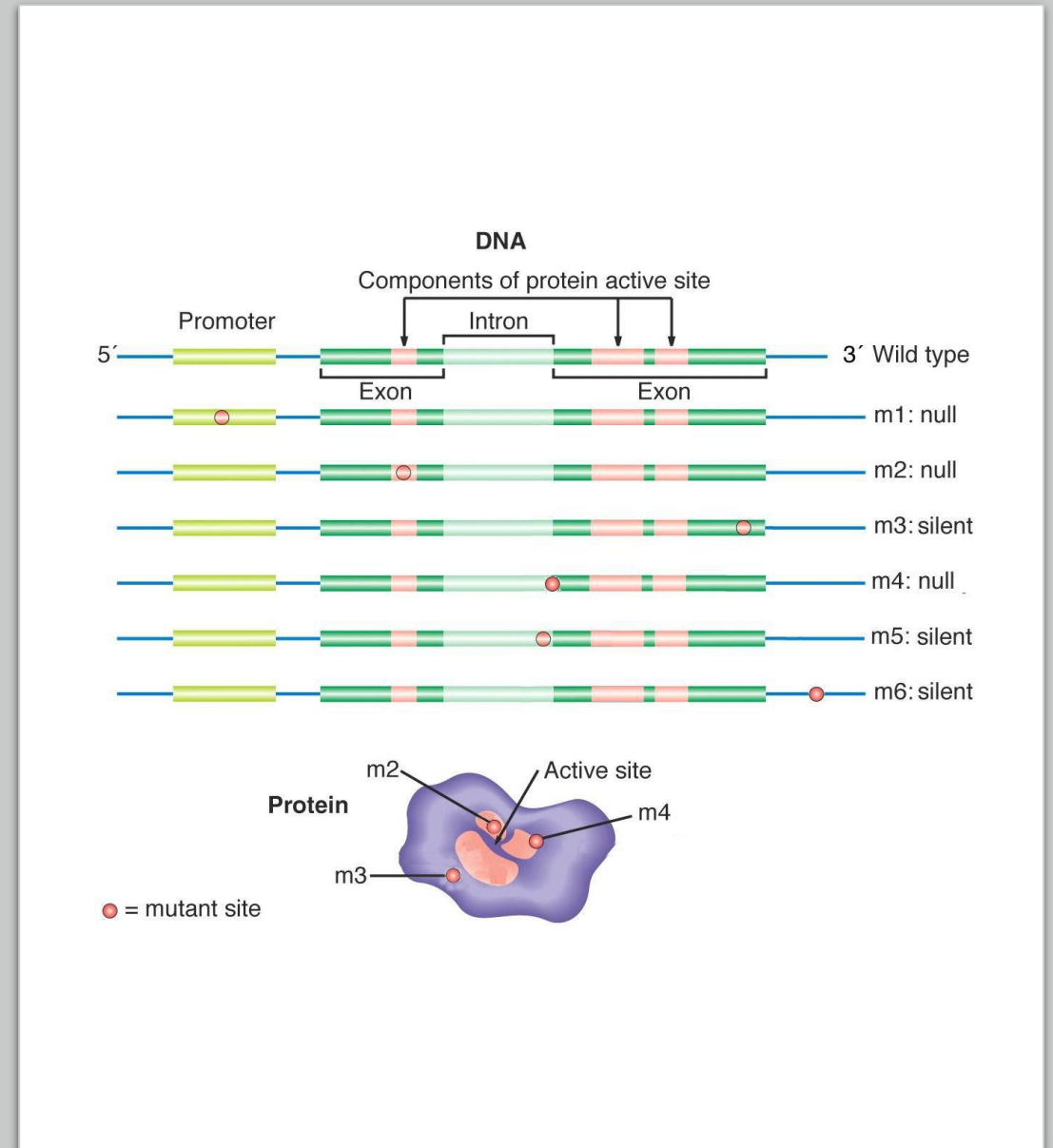
# Mutations

- Mutations can result from many events, including unequal crossing-over during meiosis. In addition, **some areas** of the genome simply seem to be **more prone** to mutation than others. These "*hot spots*" are often a result of the DNA sequence itself being **more accessible** to mutagens. Hot spots **include areas** of the genome **with highly repetitive sequences**, such as trinucleotide repeats, in which a sequence of three nucleotides is repeated many times. During DNA replication, these repeat regions are often altered because the polymerase can "slip" as it disassociates and reassociates with the DNA strand.



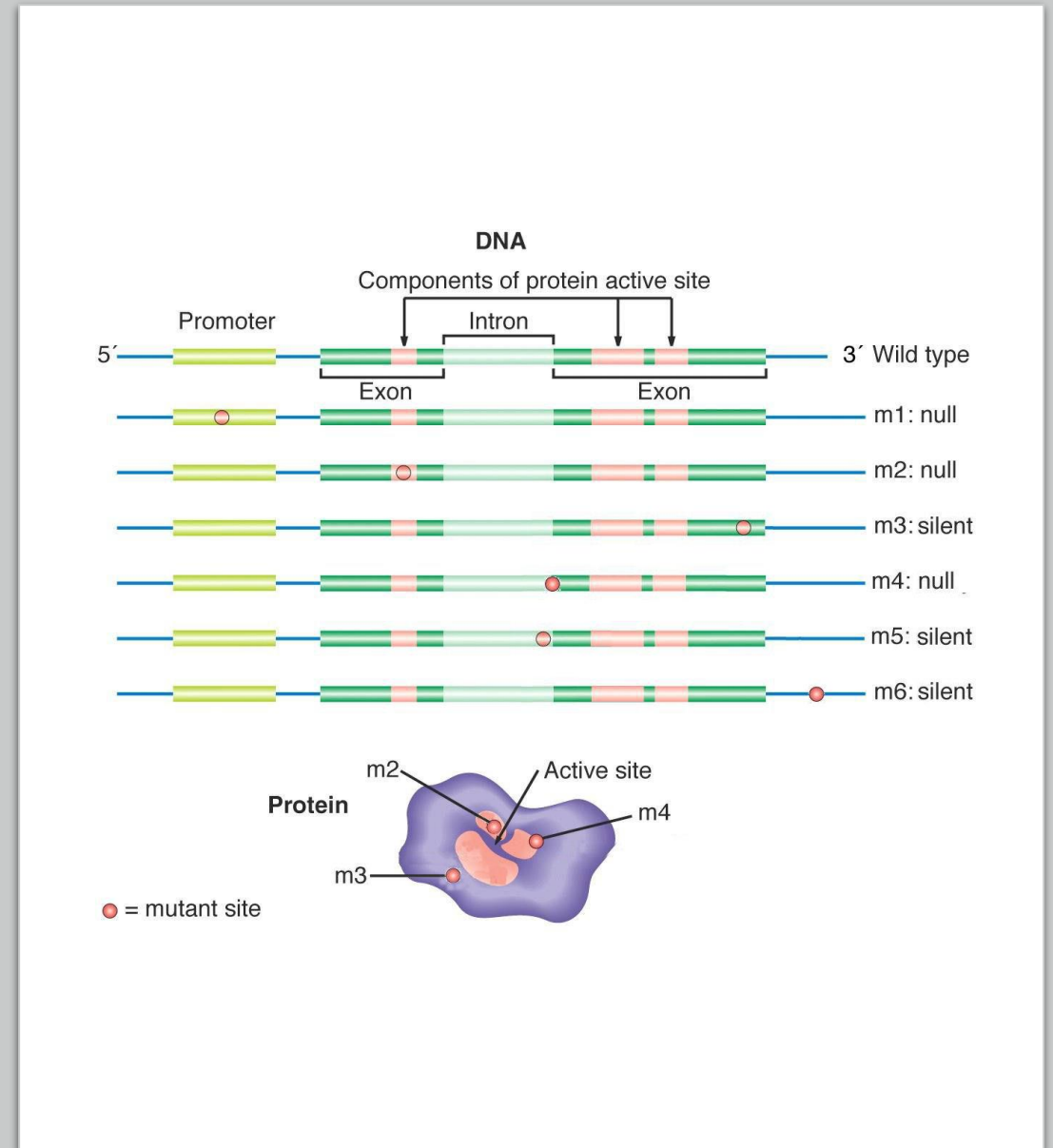
# Consequences of DNA mutations depend on their position in the gene region

- **m1**: mutations **in the promoter region** may **affect gene transcription**. May lead to non-functional (null) alleles.
- **m2**: mutations **in exons**, if they result in the **substitution** of an amino acid **in the active site** or other critical region of the protein, also lead to alleles with changed or null functionality.



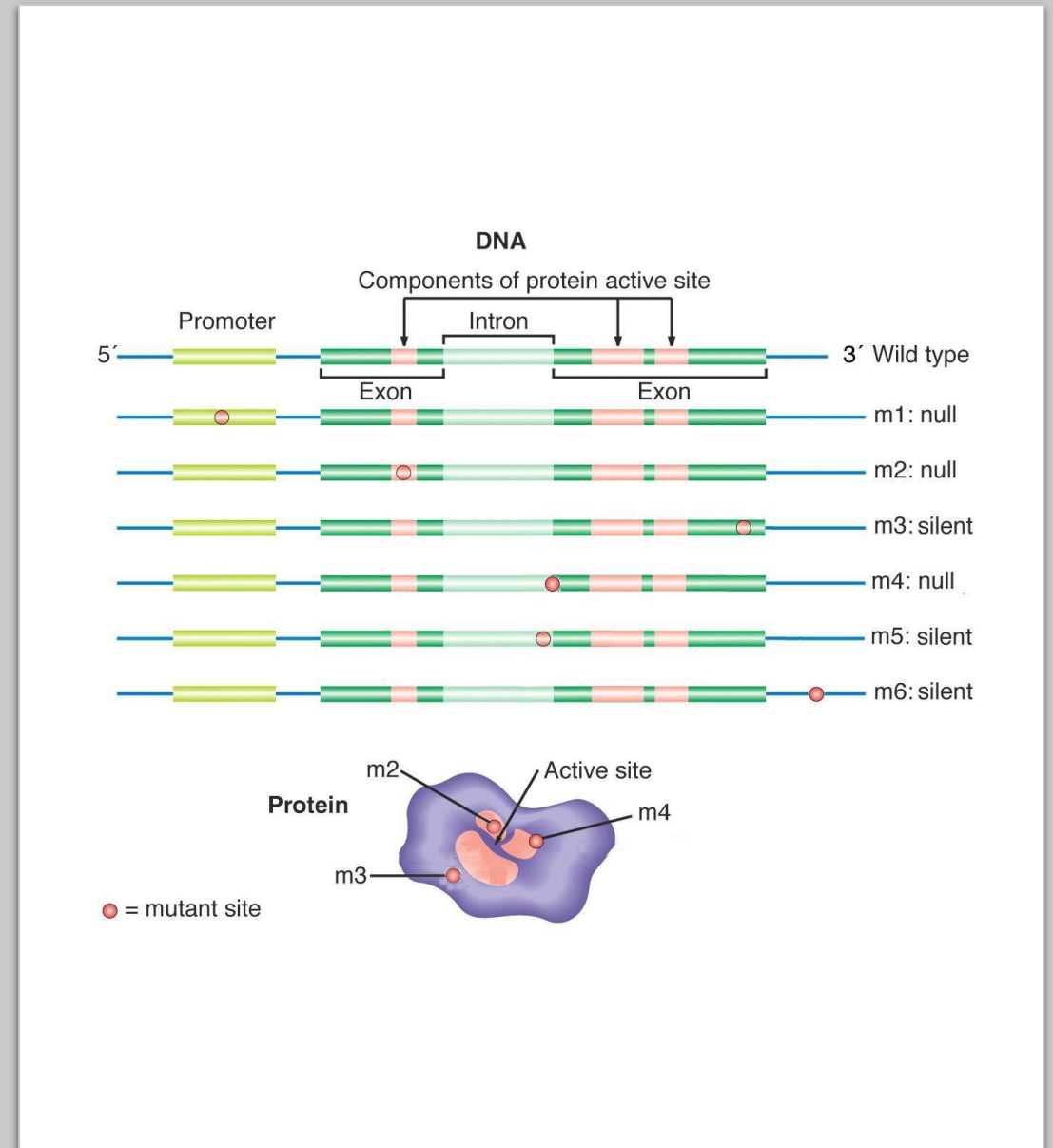
# Consequences of DNA mutations depend on their position in the gene region

- **m3: exon mutations** that result in **changes outside the active sites** or synonymous mutations (do not alter the amino acid encoded by the affected codon due to the degeneracy of the genetic code but change the DNA and RNA sequence) may have little or no effect on gene function. These mutations are called **silent** (if the amino acid is unchanged) or **neutral** (if the change has no effect).



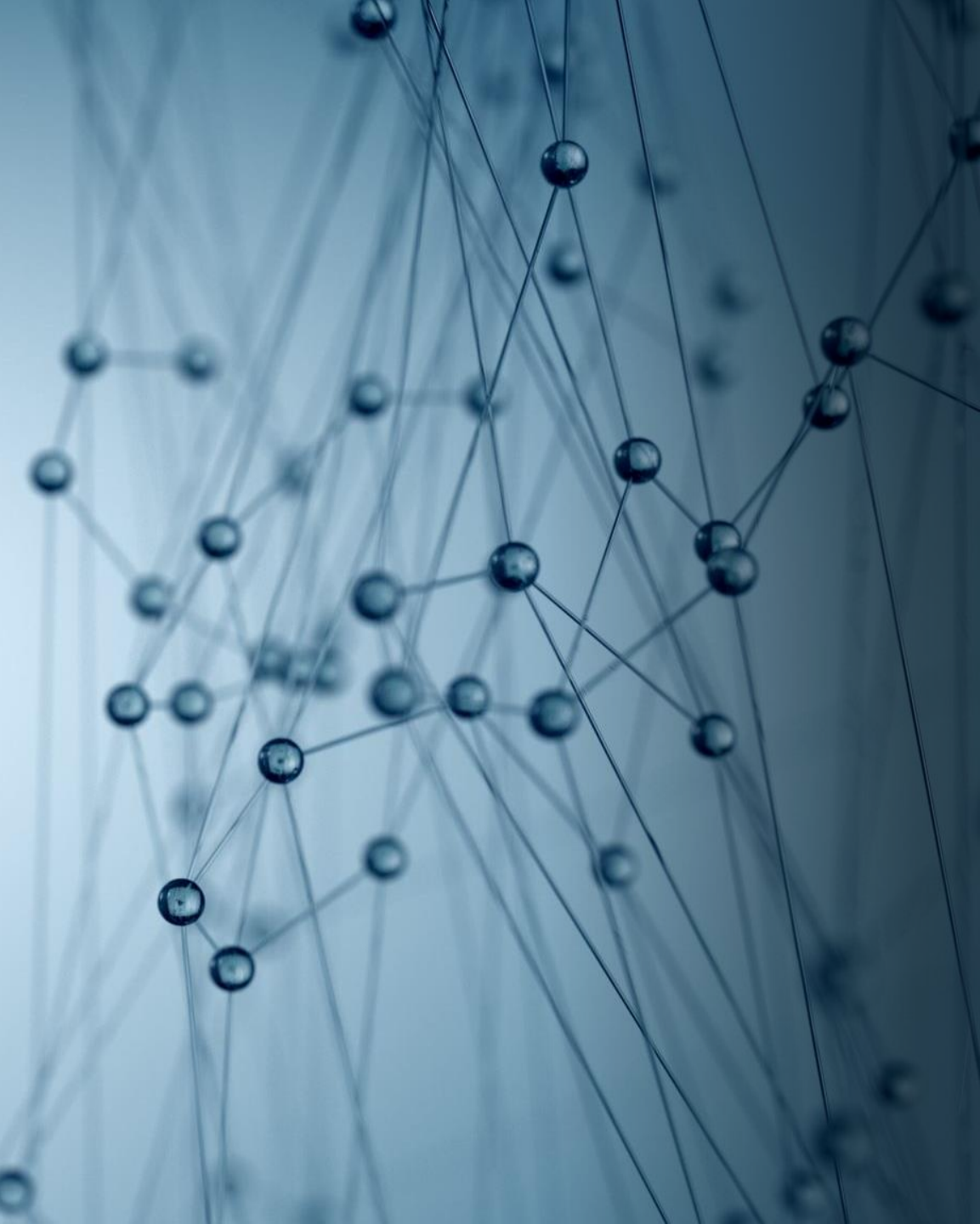
# Consequences of DNA mutations depend on their position in the gene region

- **m4**: mutations at critical positions **near intron / exon junctions** may affect mRNA splicing and lead to the deletion or retention of entire exons, and result in null alleles.
- **m5**: mutations that occur in **non-coding introns**, may have little or no effect on gene function.
- **m6**: mutations that occur in **5' or 3' UTR** may have little or no effect on gene function but may affect the level of translation.



# Epigenetics

*If all cells are created from the same genetic material,  
why are there so many different cell types?*





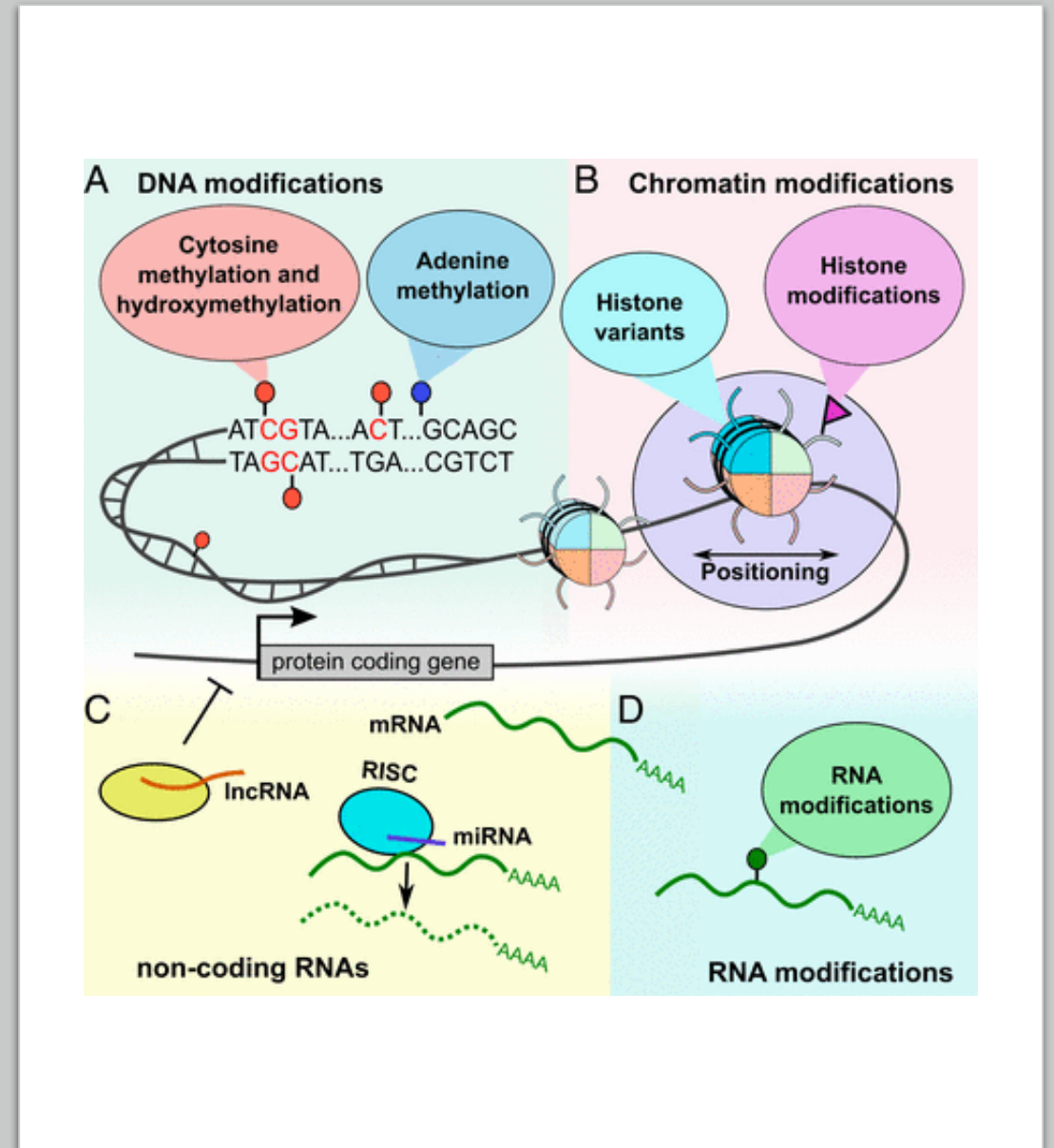
# Epigenetics

- Epigenetics is defined as the study of changes in gene expression that are **not derived from a change in the underlying DNA sequence** i.e. a change in the phenotype without changing the genotype.
- Epigenomics – the study of the complete set of epigenetic alterations.
- Epigenetic code – epigenetic features that maintain different phenotypes in different cells.



# Epigenetic mechanisms

- A. DNA modifications
- B. Chromatin modifications
- C. Non-coding RNAs
- D. RNA modifications

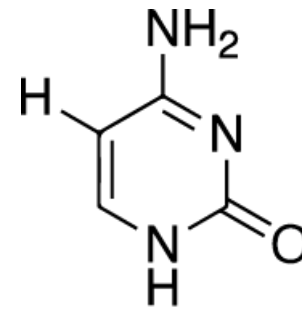


## A. DNA modifications

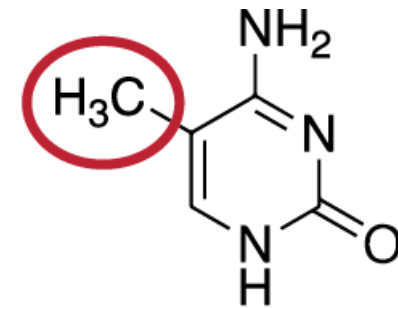
- DNA can be **modified at cytosine** and **adenine** residues by the addition of chemical groups.
- **Cytosines** can be modified by **methylation**, **hydroxymethylation**, **formylation** and **carboxylation**, while **adenines** can be modified by **methylation**.
- In humans, cytosine methylation is the most frequent, occurring often at CpG sites (cytosine followed by a guanine base in the DNA sequence). DNA methylation can also be found at cytosines followed by i.e., adenine, cytosine, or thymine, such non-CpG methylation is an abundant modification in neural tissues and increases during development.

# DNA methylation

- DNA **methylation** is a process by which **methyl groups** are added to the DNA molecule. Methylation can **change the activity** of a DNA segment without changing the sequence.
- In humans, cytosine methylation is the most frequent.
- **Cytosine methylation** is **catalysed** by enzymes known as **DNA methyltransferases (DNMT)** on position 5 (C5) of cytosine residues to form 5-methyl cytosine (5-mC).



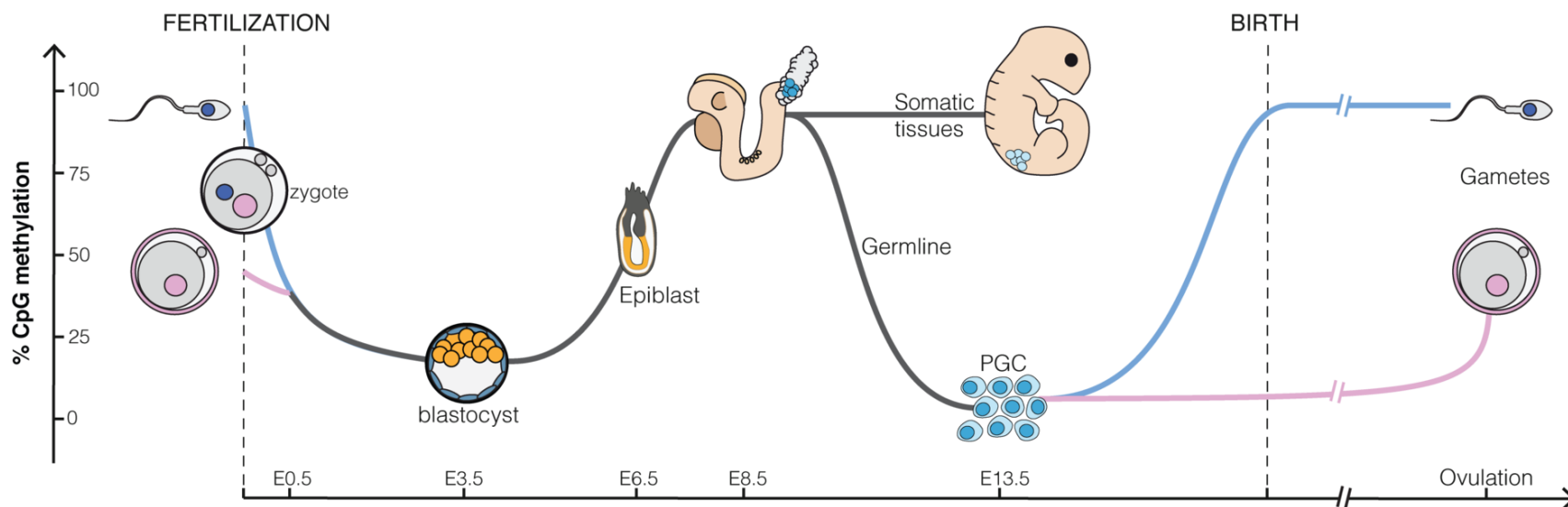
**Cytosine**



**methylated Cytosine**

# DNA methylation

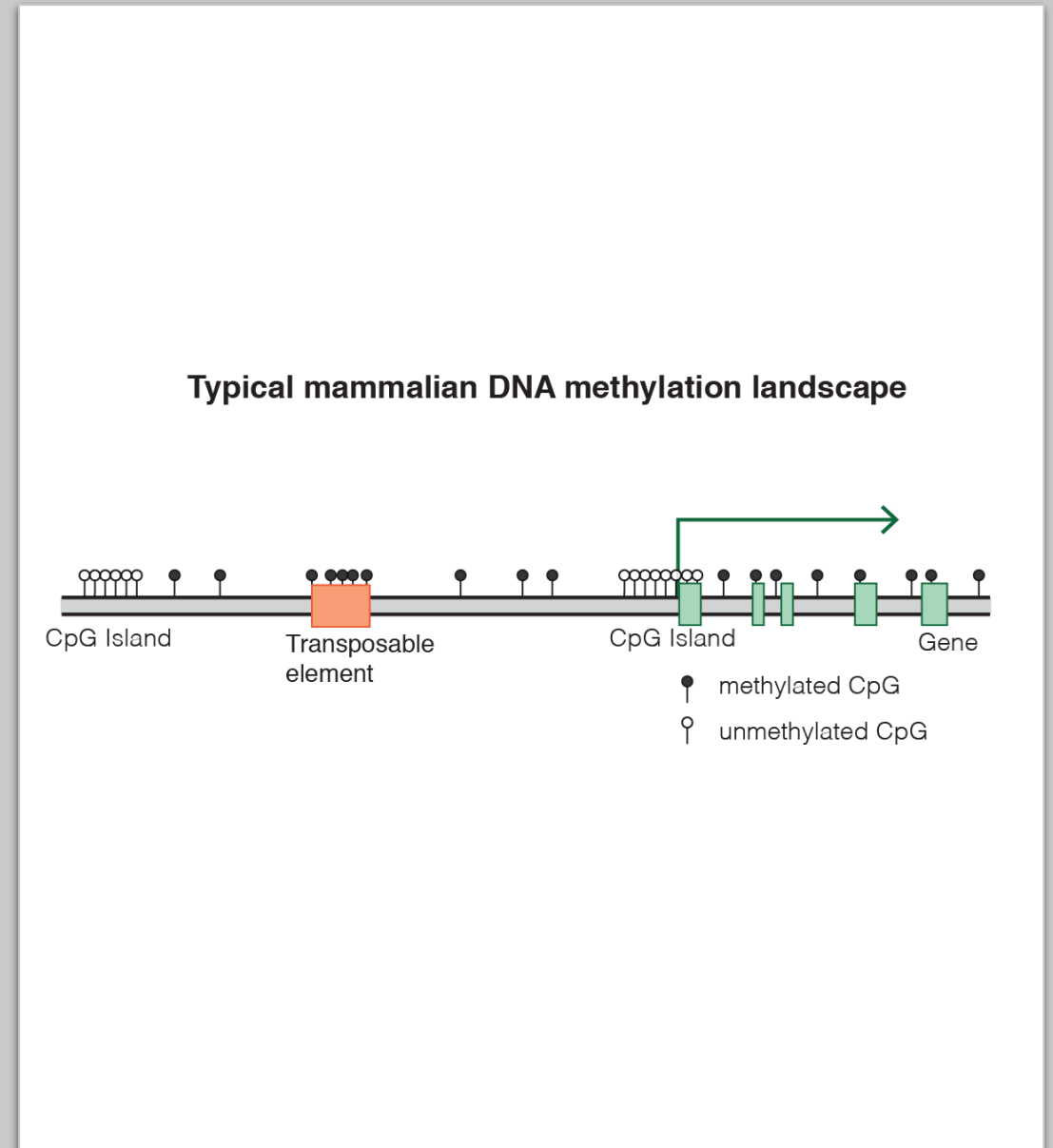
- Maintenance **DNMT1** activity is necessary to preserve DNA methylation after every **cellular DNA replication cycle**.
- Each tissue and cell type has unique DNA methylation profiles.
- DNA methylation patterns change during ontogenesis and are largely erased and then re-established between generations in mammals. Almost all parental methylation patterns are erased during gametogenesis and in early embryogenesis.



*Dynamic of DNA methylation during mouse embryonic development. E3.5-E6, etc., refer to days after fertilization. PGC: primordial germ cells.*

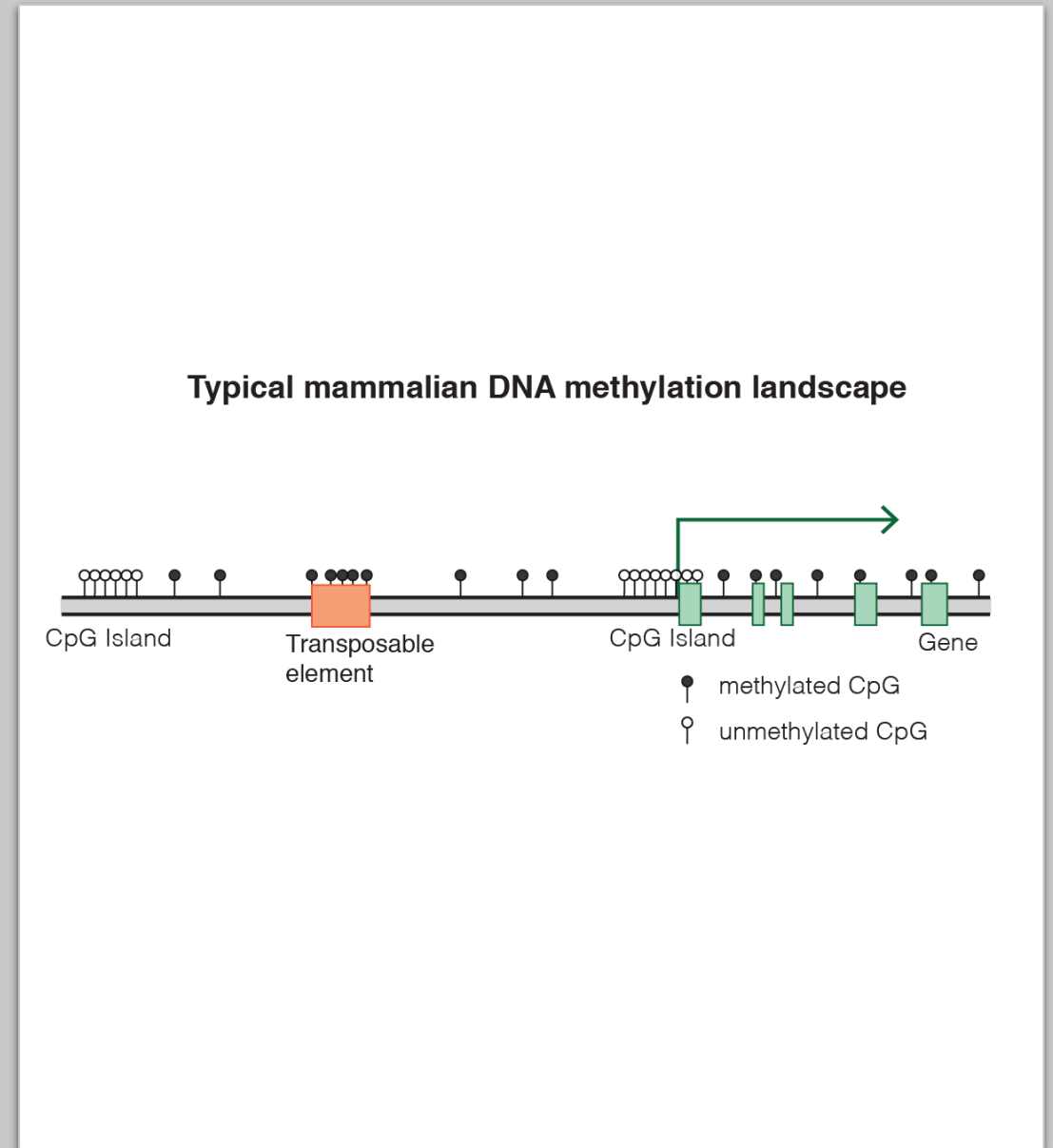
# Consequences of DNA methylation

- Transcription of most genes encoding proteins is initiated at promoters rich in CpG sequences. These sections of CpG-rich DNA are known as **CpG islands**.
- **When** the CpG island in the promoter region of the gene is **methylated**, the **expression** of the gene is **suppressed**. CpG-dense promoters of actively transcribed genes are never methylated.
- The methylation of DNA may physically impede the binding of transcriptional factors.



# Consequences of DNA methylation

- Methylated DNA may be bound by proteins known as methyl-CpG-binding domain proteins (**MBDs**). MBD proteins then recruit chromatin remodeling proteins that can **modify histones**.
- Repression of transposable elements.
- High CpG methylation increases the frequency of spontaneous mutations. (Methylated C residues can spontaneously deaminate to T residues).



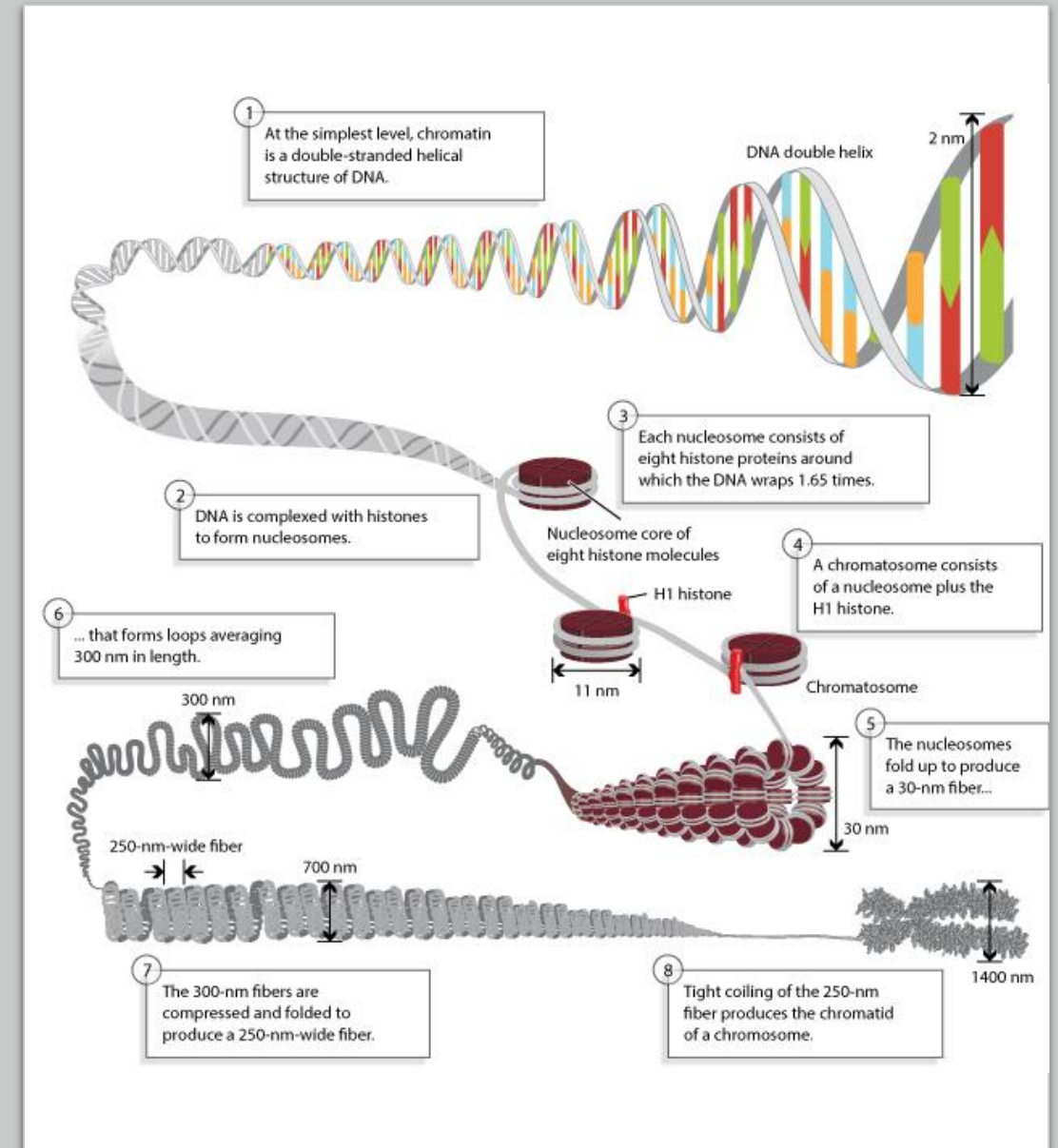
## B. Chromatin modifications

- The **primary protein** components of chromatin are **histones**, which bind to DNA and function as "anchors" around which the DNA strands are wound.
- Regions of chromatin containing **genes** which are **actively transcribed** are less tightly compacted and closely associated with RNA polymerases in a structure known as **euchromatin**.
- Regions containing **transcriptionally inactive genes** (**heterochromatin**) are generally more condensed.



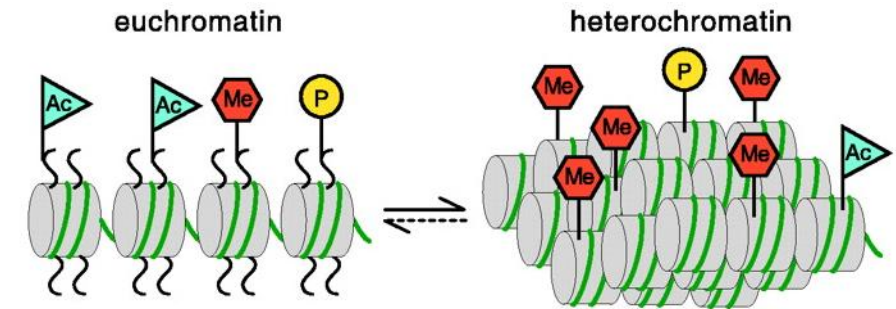
# Chromatin

- Chromatin structure **is not rigid** and can be modified by protein complexes that specify the location, composition, and modification state of nucleosomes, ultimately regulating access to the underlying DNA sequence.
- Canonical histones (H3, H4, H2A, H2B, and linker histone H1) package the newly replicated genome.
  - They can be replaced with histone variants that alter nucleosome structure, stability, dynamics, and DNA accessibility.



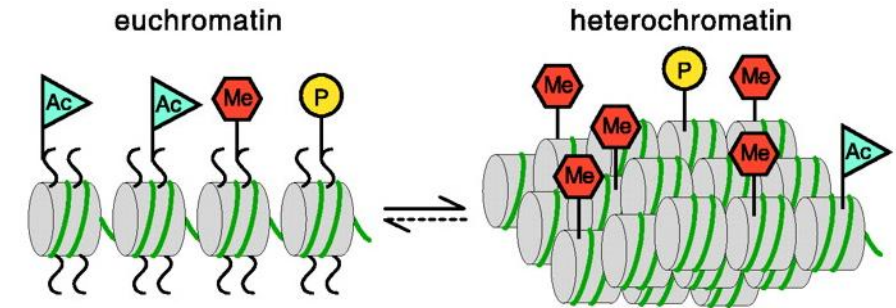
# Histon modifications

- Posttranslational modification of histone proteins alters chromatin function and DNA accessibility.
- **N-termini** of histones (called histone tails) are particularly **highly modified**.
- Histon modifications include acetylation, methylation, ubiquitylation, phosphorylation, sumoylation, ribosylation and citrullination. **Acetylation** is the most highly studied.



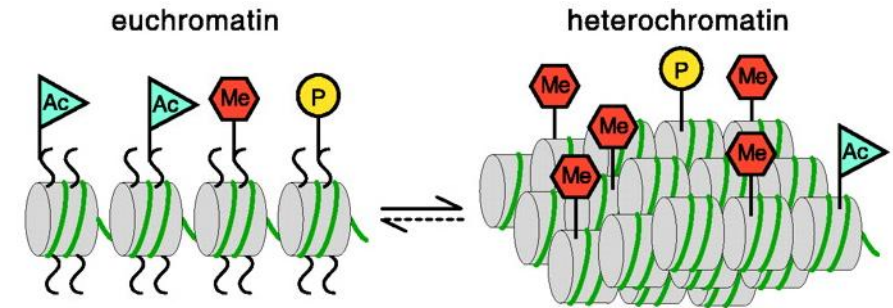
# Histon modifications

- Histone acetylation is performed by **histone acetyltransferases (HATs)** which add an acetyl group to lysine residues in the histone tail which masks the positive charge of the residue, loosening the electrostatic interaction between DNA and histones, causing decondensation of chromatin thereby allowing gene transcription.
- **Histone deacetylases (HDACs)** remove the acetylation mark of histone lysine residues.
  - The HDAC-containing enzyme complex binds to methylated DNA via MeCP1 and MeCP2 binding proteins. Thus, DNA methylation and histone modifications are interconnected processes.



# Histon modifications

- Methylation of lysine 9 of histone H3 has been associated with constitutively transcriptionally silent chromatin (constitutive heterochromatin).
- Phosphorylation (the addition of a phosphate group from ATP) is possible on serine residues in histone tails.



## C. RNA modifications

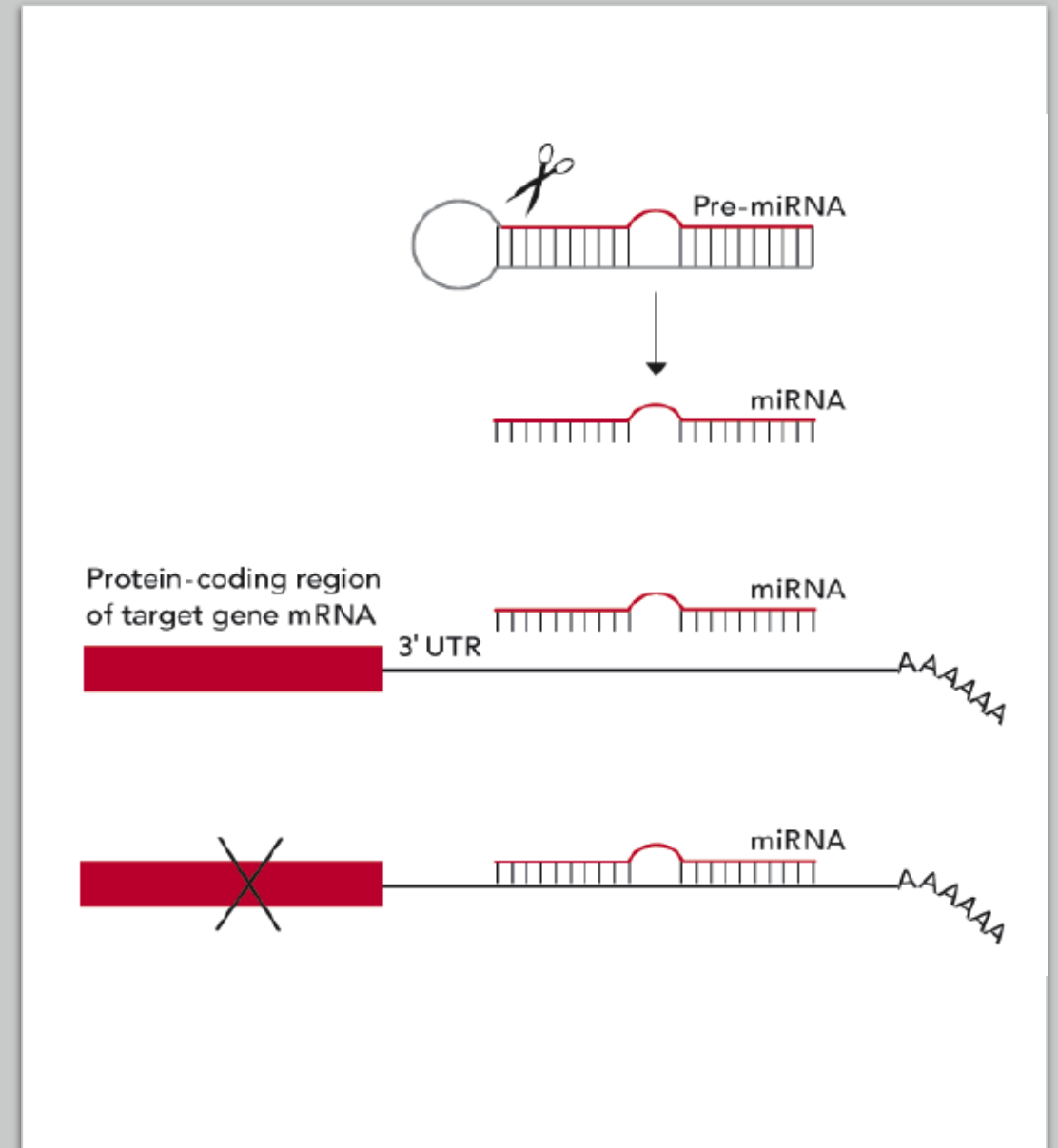
- Epigenetic modifications occur not only in DNA but also in RNA (**the epitranscriptome**).
- mRNA, tRNA, rRNA and non-coding RNAs contain thousands of post-transcriptional chemical modifications.
- N<sup>6</sup>-methyl-adenosine (m6A) modification is the most abundant.
- m6A modification is recognized by families of RNA binding proteins that affect many aspects of mRNA function.
- RNA modifications represent another layer of epigenetic regulation of gene expression, analogous to DNA methylation and histone modification.

## D. Non-coding RNAs

- **RNA interference** is a biological process in which RNA molecules inhibit gene expression or translation, by neutralizing targeted mRNA molecules.

# miRNA

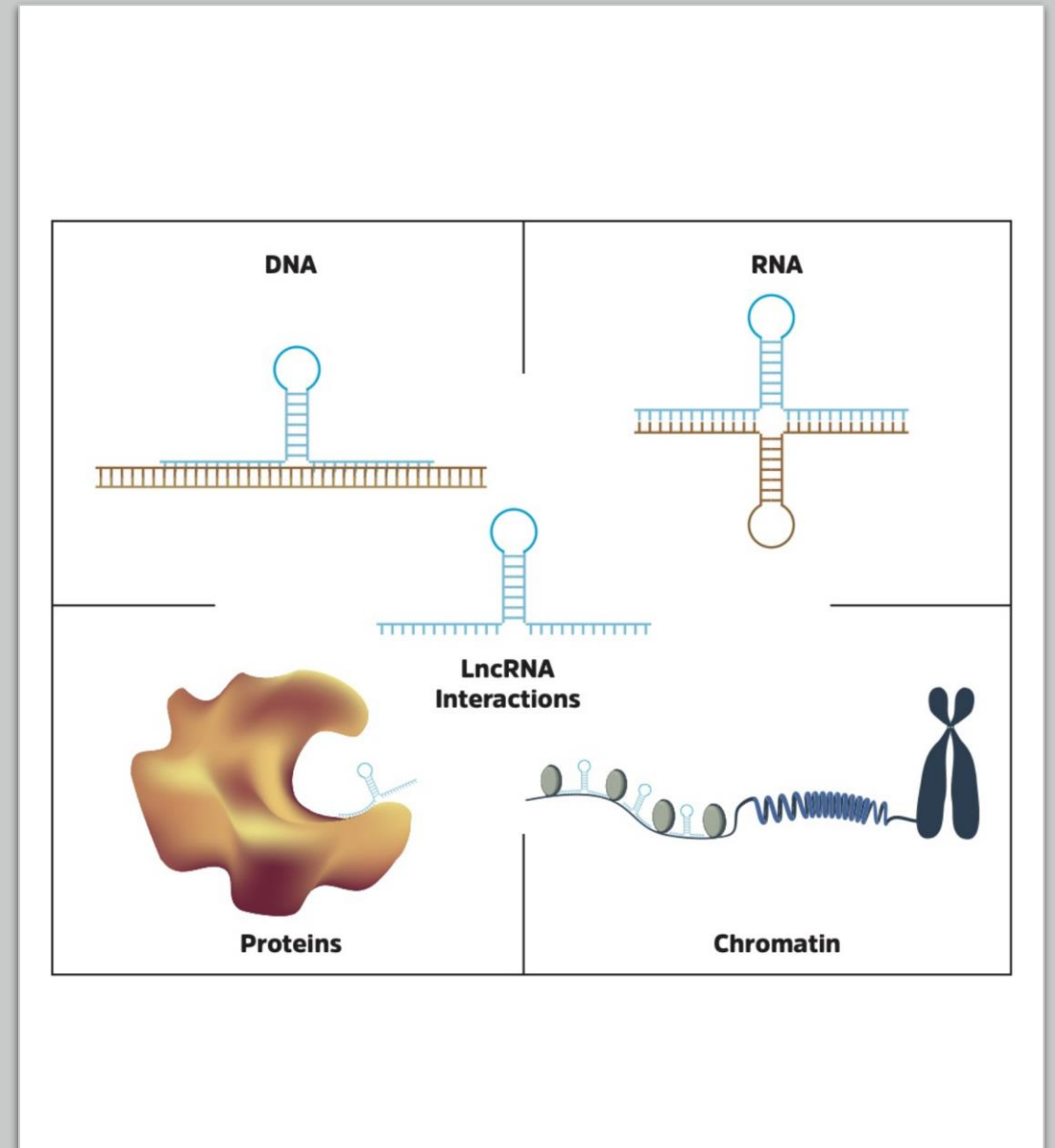
- **MicroRNAs** (miRNAs) are short (~22 nucleotides in length), single-stranded, noncoding RNAs which regulate mRNAs via degradation or inhibition of their translation into proteins. Each miRNA may target about 100 to 200 mRNAs. Many miRNAs are epigenetically regulated. About 50 % of miRNA genes are associated with CpG islands, that may be repressed by epigenetic methylation.





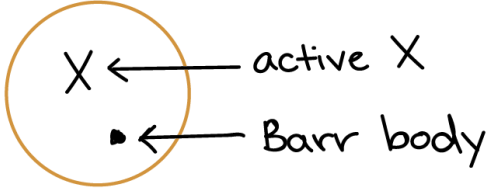
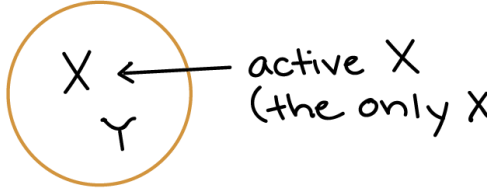
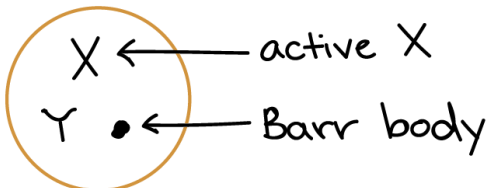
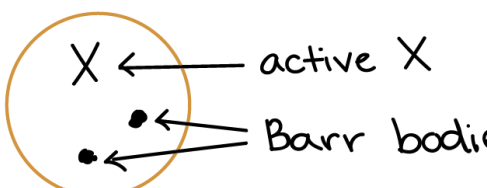
# lncRNA

- **Long noncoding RNAs (lncRNAs)** are a large family (~50,000 in the human genome) of RNAs that are >200 base pairs in length. lncRNAs are abundant in neural tissues. lncRNAs reduce transcription, regulate the processing of mRNAs to modulate the abundance of subtly different transcripts from the same gene (alternative splicing), and influence the activity of miRNAs.



# X chromosome inactivation

- In female somatic cells, one of the two X chromosomes is inactivated to equalize the dose of sex-linked gene products between female and male cells.
- The most iconic of all long non-coding RNAs, the X-inactive specific transcript (Xist), mediates X chromosome silencing.
- Xist RNA was found to physically coat the inactive-X chromosome.

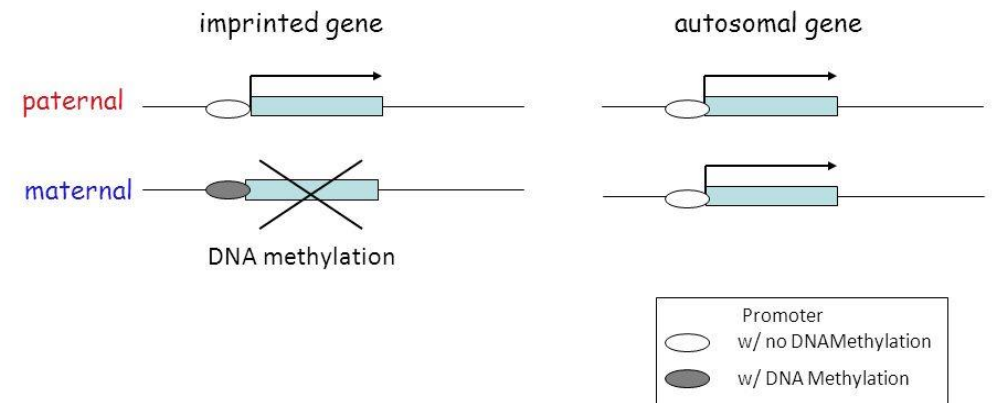
XX female	
XY male	
XXY male (Klinefelter)	
XXX female (triple X)	

# Imprinting

- Father and mother have **different epigenetic patterns** for specific genomic loci in their germ cells.
- **Some loci are expressed only paternally**, this is referred to as maternal imprinting. **Other loci are expressed exclusively maternally**, indicating paternal imprinting. The second allele of the gene is inactive (imprinted).
- This phenomenon is estimated to affect several hundred human genes under physiological conditions.
- In the case of imprinted alleles, the identity of reciprocal crosses found by Gregor Mendel does not apply, because paternal and maternal information are not equivalent.

## Genomic Imprinting

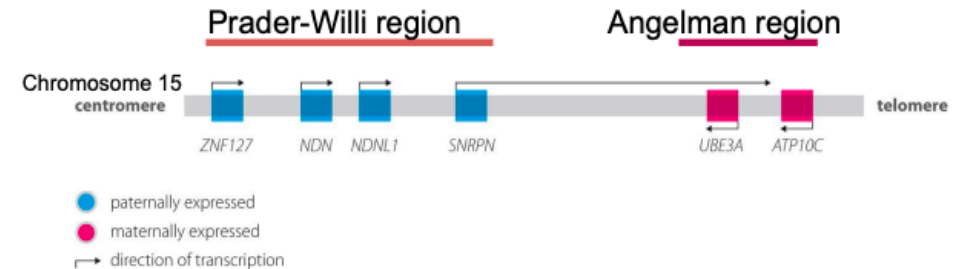
- Preferential expression (or repression) of one parental allele
- Epigenetic modification mechanism (CpG methylation)



# Imprinting

- The **epigenetic imprints** are established **during** male and female **gametogenesis**, passed to the zygote through fertilization, maintained throughout development and adult life, and **erased in primordial germ cells** before the new imprints are set.
- The best-known case of imprinting in human disorders is that of **Angelman syndrome** and **Prader-Willi syndrome**.
  - Both can be produced by the same genetic mutation (chromosome 15q partial deletion). The particular syndrome that will develop depends on whether the mutation is inherited from the child's mother or father.
  - Loss of the maternal contribution is linked to Angelman syndrome and the loss of the paternal contribution is linked to PWS.

## Genes in the Prader-Willi/ Angelman syndrome region



Most deletions span both the Prader-Willi and Angelman regions

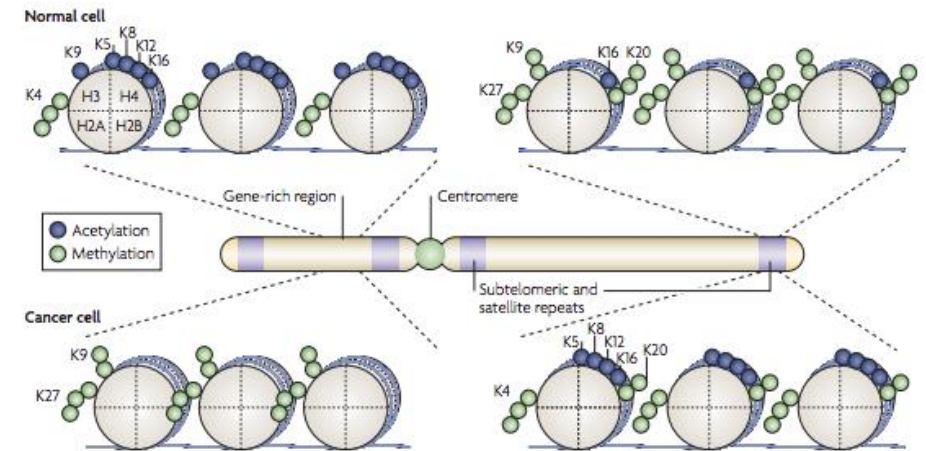
# Epigenetics and disease

- Epigenome is intricately linked to the environment and disease state. Many environmental exposures can induce changes in the epigenome which alters patterns of gene expression without directly altering the underlying DNA sequence to induce disease states.
- The combination of epigenetic marks has been shown to predispose an individual to certain disease states such as diabetes, cancer, cardiovascular disease and obesity.
- For example, the environment you were exposed to in the uterus (poor maternal diet, maternal smoking, high stress levels, pollution etc.) directly affects the epigenome.
- Epigenetic modification act as mediators or risk factors showing the interconnected relationship between the environment and disease.

# Cancer

- A variety of epigenetic mechanisms can be perturbed in different types of cancer.
- Epigenetic alterations of DNA repair genes or cell cycle control genes are very frequent in sporadic (non-germ line) cancers.
- Epigenetic alterations are important in malignant transformation and their manipulation holds great promise for cancer prevention, detection and therapy.

## Example: Histone modification profiles Normal vs Cancer





Thank you for your attention