به نام خدا

Applications of Genetics to Cardiovascular Medicine

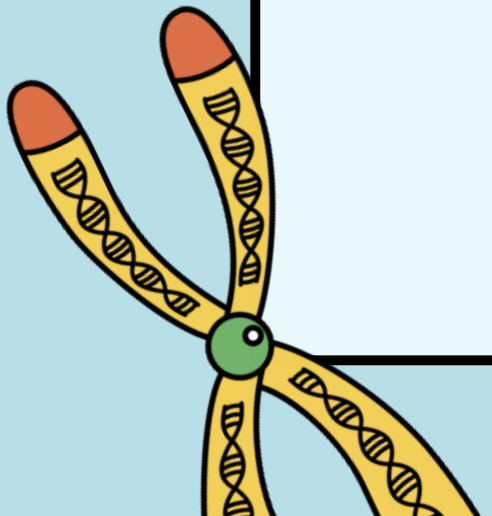
Presented by:

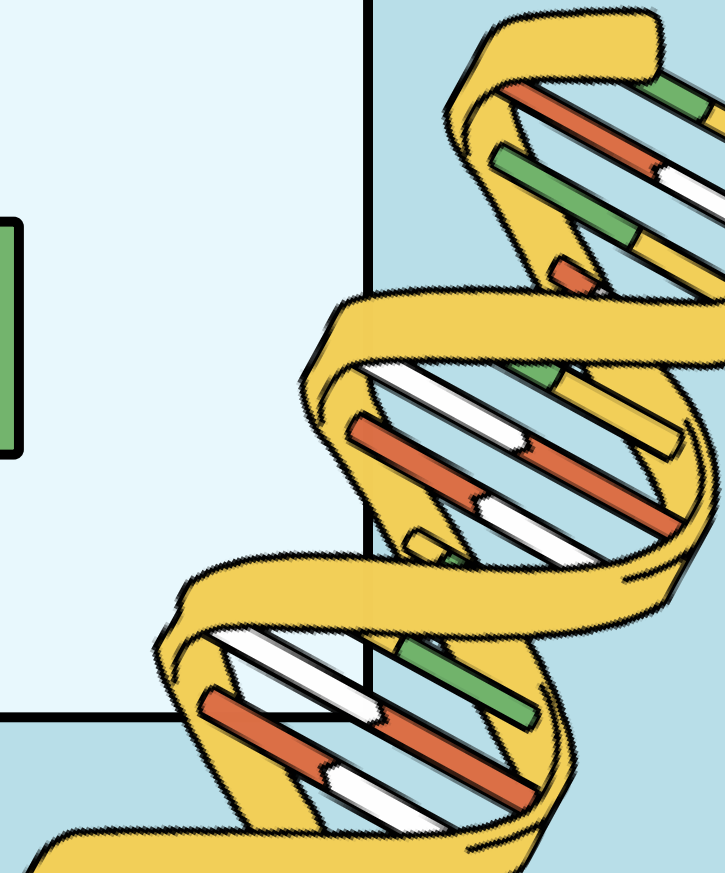
Dr Miri-Moghaddam

Farzane Vafaeie

Structure of DNA

An Introduction to the Function and Characteristics of DNA



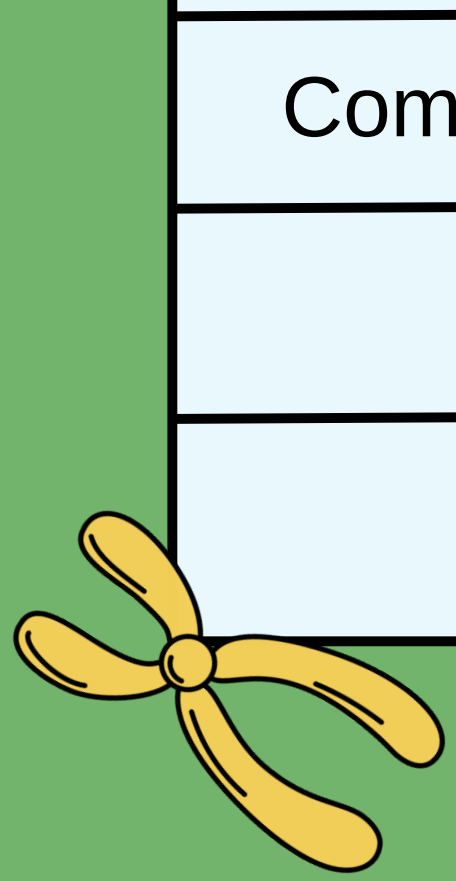


Lesson Outline

Composition, Structure, Replication

Chromosome Structure

Types of DNA Sequence

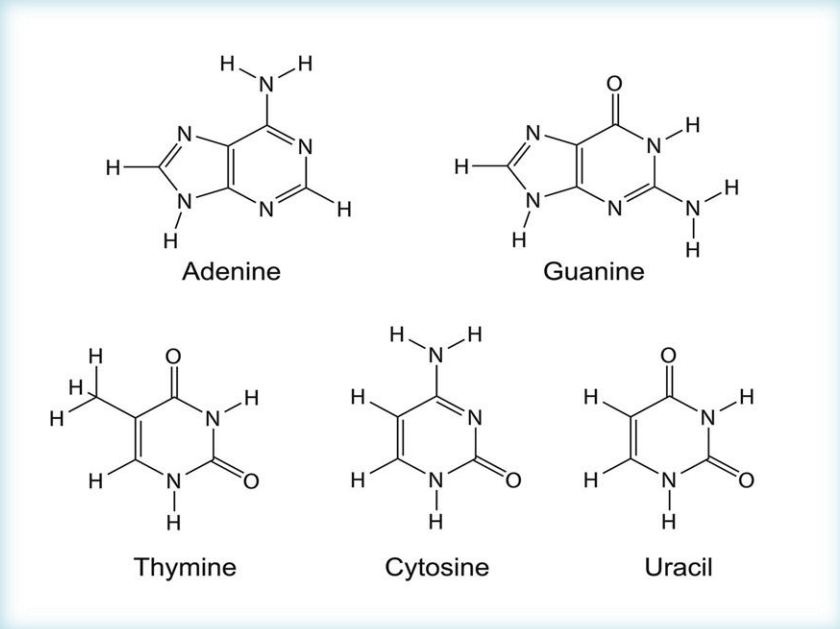


Composition

DNA is a polymer of nucleotide.

Each nucleotide is composed of a **nitrogenous base**, a **sugar molecule**, and a **phosphate molecule**.

The **nitrogenous** bases fall into two types, **purines** and **pyrimidines**. The purines include adenine (A) and guanine (G); the pyrimidines include cytosine (C), thymine (T), and uracil (U).



Structure

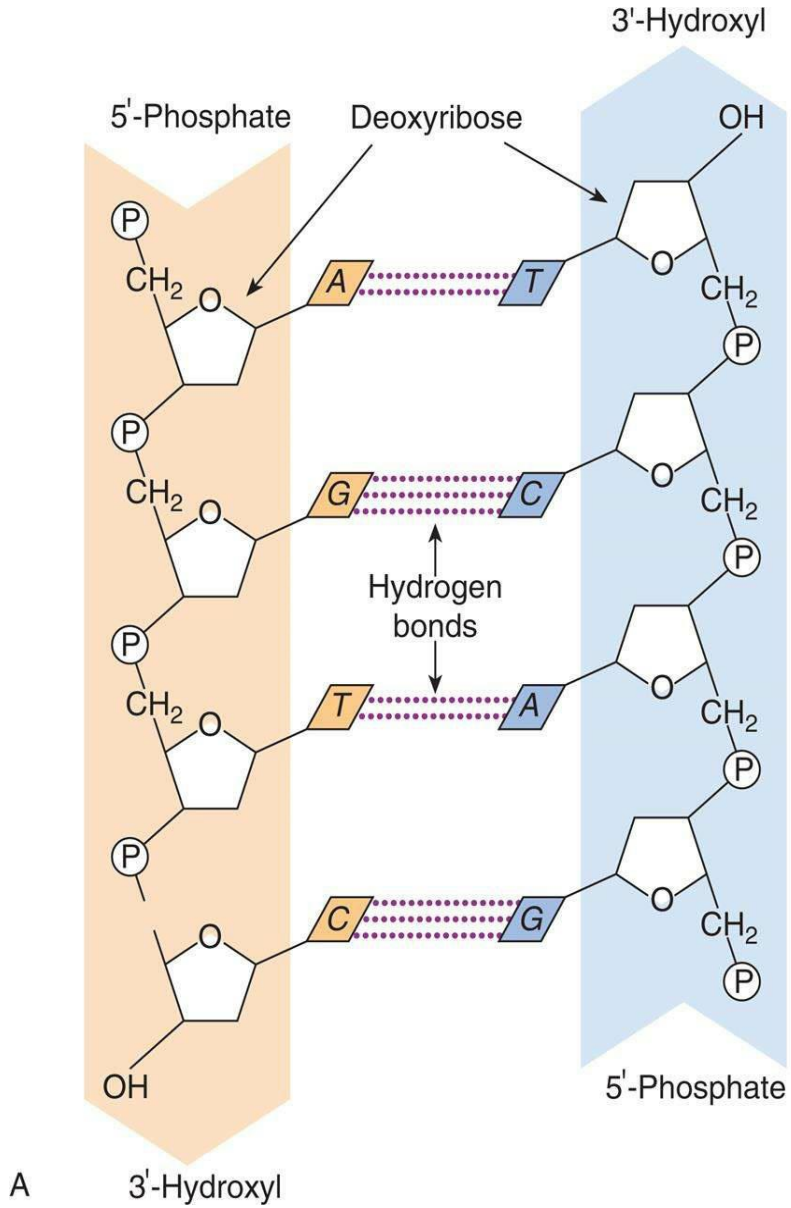
composed of two chains of nucleotides arranged in a double helix

Phosphodiester bonds between the 3′ and 5′ carbons of adjacent sugars

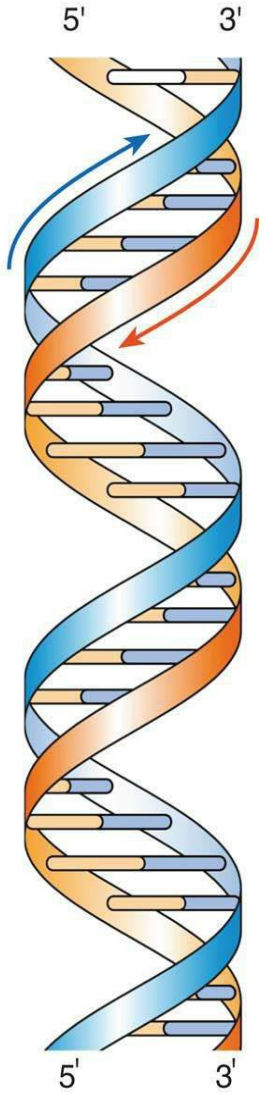
Two chains being held together by hydrogen bonds between the nitrogenous bases

They have opposite orientations and are said to be antiparallel

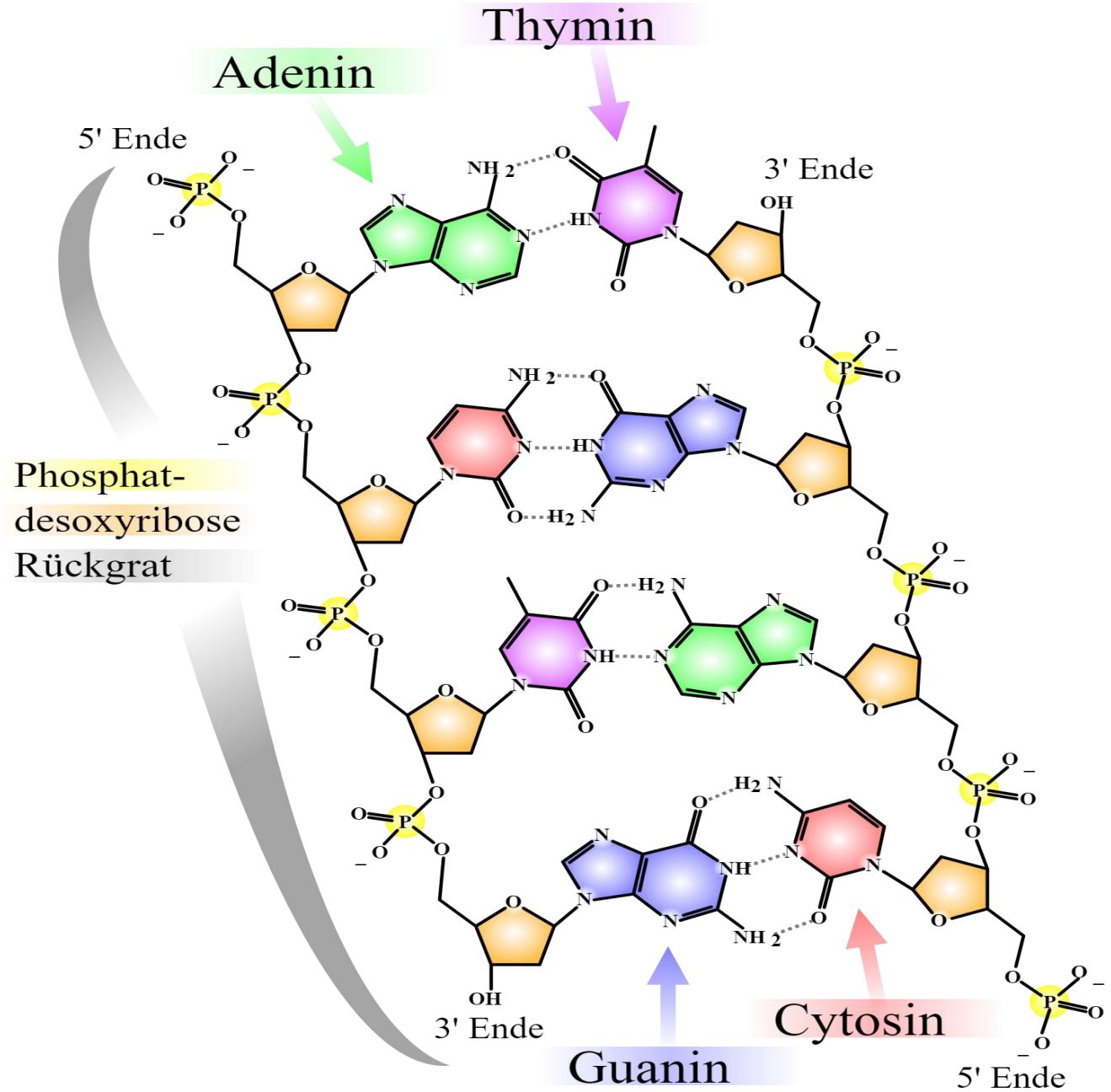
A purine in one chain always pairs with a pyrimidine in the other chain



A



B

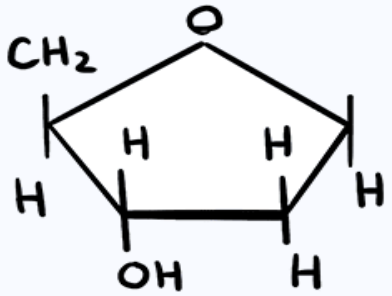

Thymin
Adenin
5' Ende
3' Ende
Phosphat-
desoxyribose
Rückgrat
3' Ende
Guanin
Cytosin
5' Ende

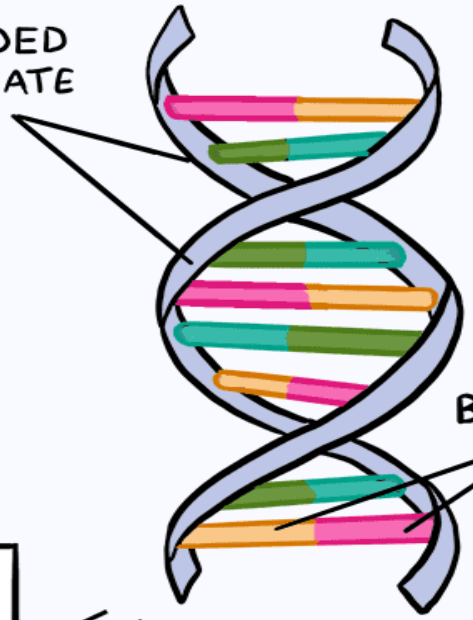
DNA vs. RNA

DEOXYRIBONUCLEIC ACID

DOUBLE-STRANDED
SUGAR* PHOSPHATE

* DEOXYRIBOSE





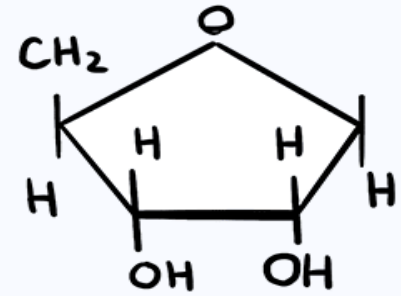
BASE PAIR

RIBONUCLEIC ACID

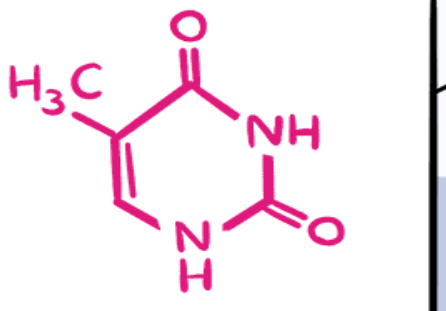
USUALLY SINGLE-STRANDED
SUGAR* PHOSPHATE

* RIBOSE

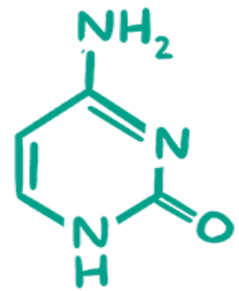
SINGLE
NUCLEOBASE



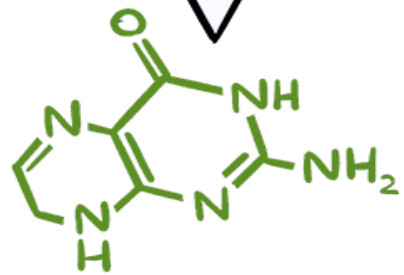
NUCLEOBASES



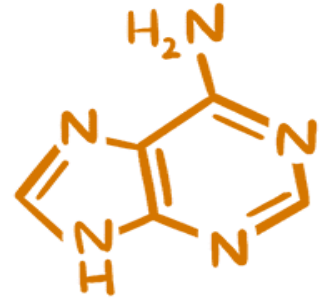
THYMINE



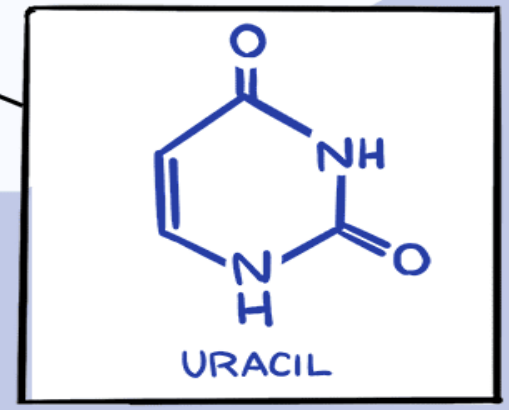
CYTOSINE



GUANINE



ADENINE



URACIL

ThoughtCo.

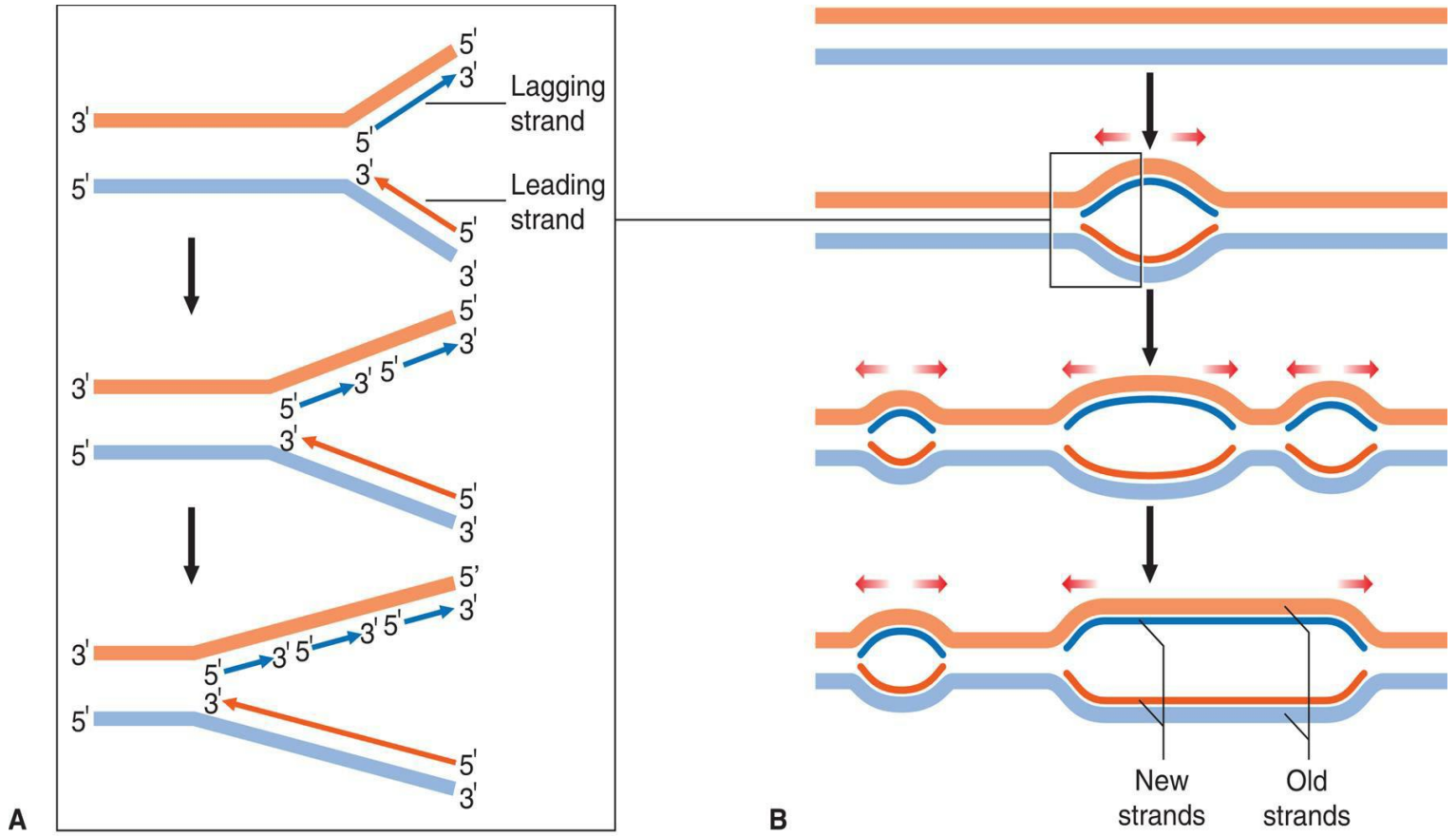
Replication

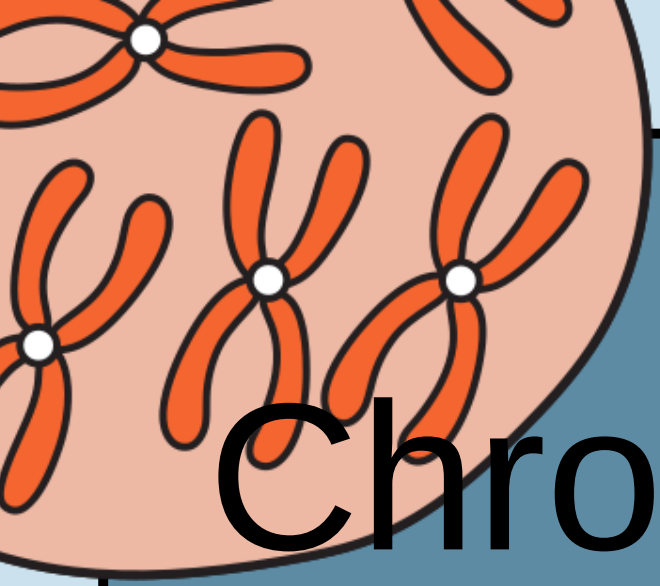
semiconservative

DNA syntesis occurs in the 5′ to 3′ direction.

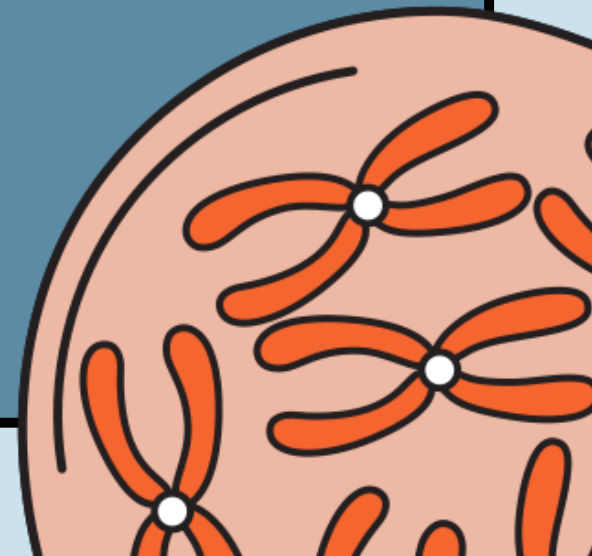
Okazaki fragments

DNA replication takes in the S phase of the cell cycle

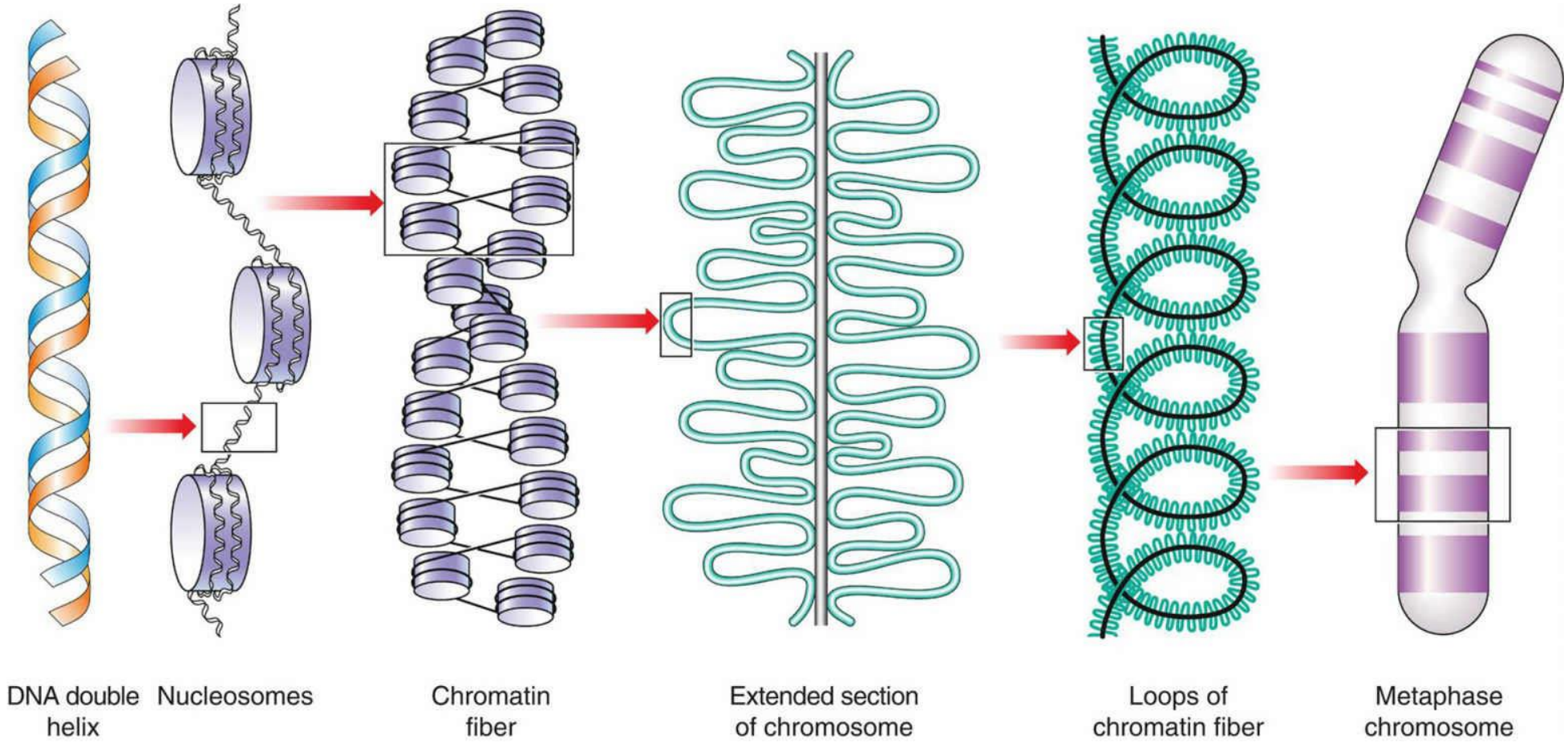




Chromosome structure



The packaging of DNA into chromosomes involves several orders of DNA coiling and folding.



Types of DNA Sequence

- 60% to 70% of the human genome consists of single or low-copy number DNA sequences.
- 30% to 40%, consists of either moderately or ***highly repetitive*** DNA sequences that are not transcribed.
- This latter portion consists of mainly satellite DNA and interspersed DNA sequences

Box 2.1

Types of DNA Sequences

Nuclear (~3×10^9 base pairs)

Genes (~20,000)
Unique, single-copy
Multigene families
Classic gene families
Gene superfamilies

Extragenic DNA (unique/low copy number or moderate/highly repetitive)

Tandem repeat
Satellite
Minisatellite
Telomeric
Hypervariable
Microsatellite
Interspersed

Short interspersed nuclear elements
Long interspersed nuclear elements

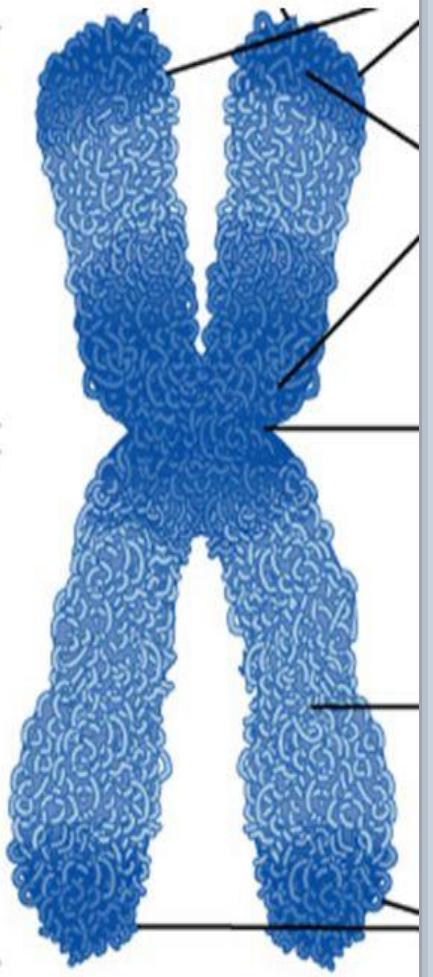
Mitochondrial (16.6 kilobases, 37 genes)

Two ribosomal RNA genes
22 transfer RNA genes

- It is estimated that there are around 21,000 **protein-coding genes** in the nuclear genome.
- The distribution of these genes varies greatly between chr. regions.
- Heterochromatic and centromeric regions are mostly **non-coding**,
- The highest gene density observed in **subtelomeric** regions.

Chromosome Parts:

- Heterochromatin:
  - More condensed
  - Silenced genes (methylated)
  - Gene poor (high AT content)
  - Stains darker
- Euchromatin:
  - Less condensed
  - Gene expressing
  - Gene rich (higher GC content)
  - Stains lighter

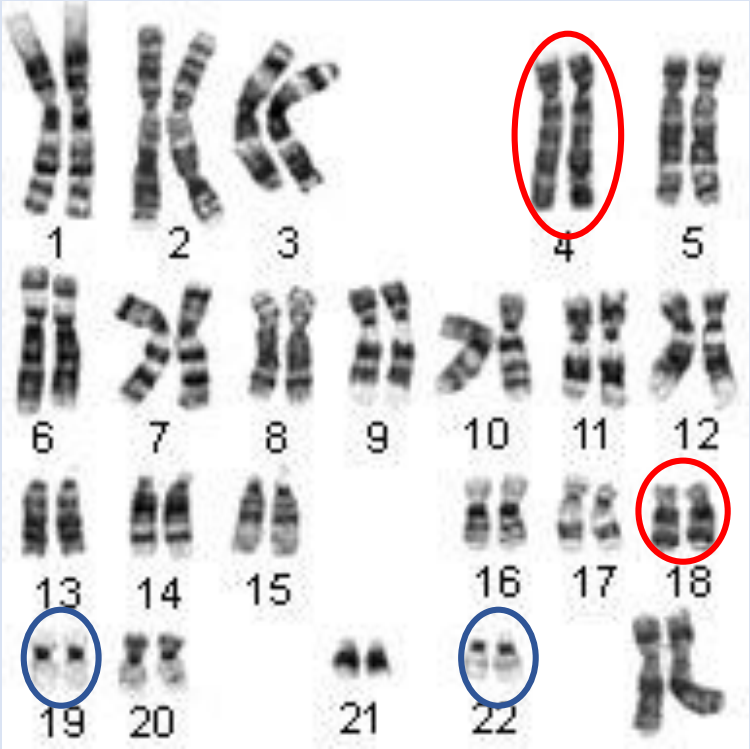


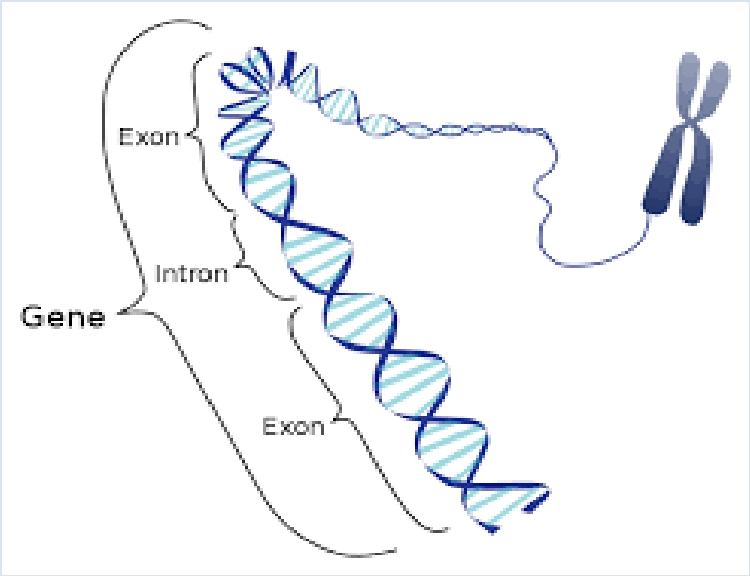
- Chr.s 19 and 22 are gene-rich, whereas 4 and 18 are relatively gene-poor.

- The size of genes also shows great variability.

- **Small genes** with single exons to the **TTN** gene, which encodes the largest known protein in the human body

- It has not only the largest number of exons (363) in any known gene but also the single largest exon [17,106 bp]



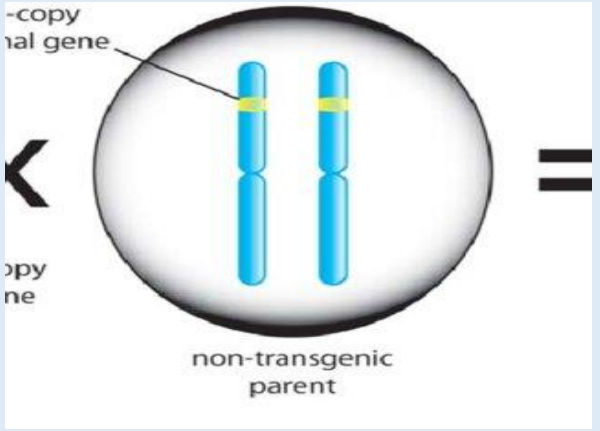


Nuclear Genes

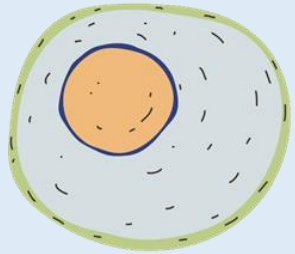
Unique Single-Copy Genes:

Most human genes are unique single-copy genes coding for polypeptides that are involved in or carry out a variety of cellular functions.

- These include enzymes, hormones, receptors, structural and regulatory proteins.



Nuclear Genes



Multigene Families:

- **Classic gene families** — high degree of sequence homology
- **Gene superfamilies** — limited sequence homology

- Many genes have **similar functions**, having arisen through **gene duplication** events with subsequent evolutionary divergence, making up what are known as multigene families.

- Some are found physically *close together in clusters*, for example, the **α- and β-globin** gene clusters on chr.s 16 and 11, whereas others are widely dispersed throughout the genome, occurring on different chr.s, such as the **HOX homeobox** gene family.

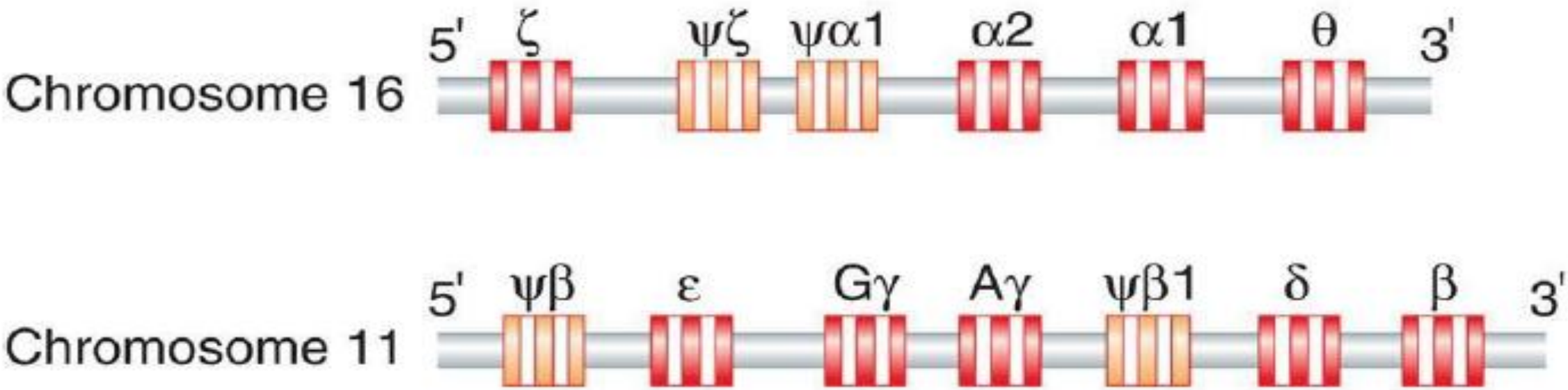


FIG. 2.5 Representation of the α- and β-globin regions on chromosomes 16 and 11.

Gene Structure

Most human genes contain **introns**, but the number and size of both introns and **exons** are extremely variable. Individual introns can be far larger than the coding sequences, and some have been found to contain coding sequences for other genes (i.e., genes occurring within genes).

Genes in humans do not usually overlap, being separated from each other by an average of 30 kb.

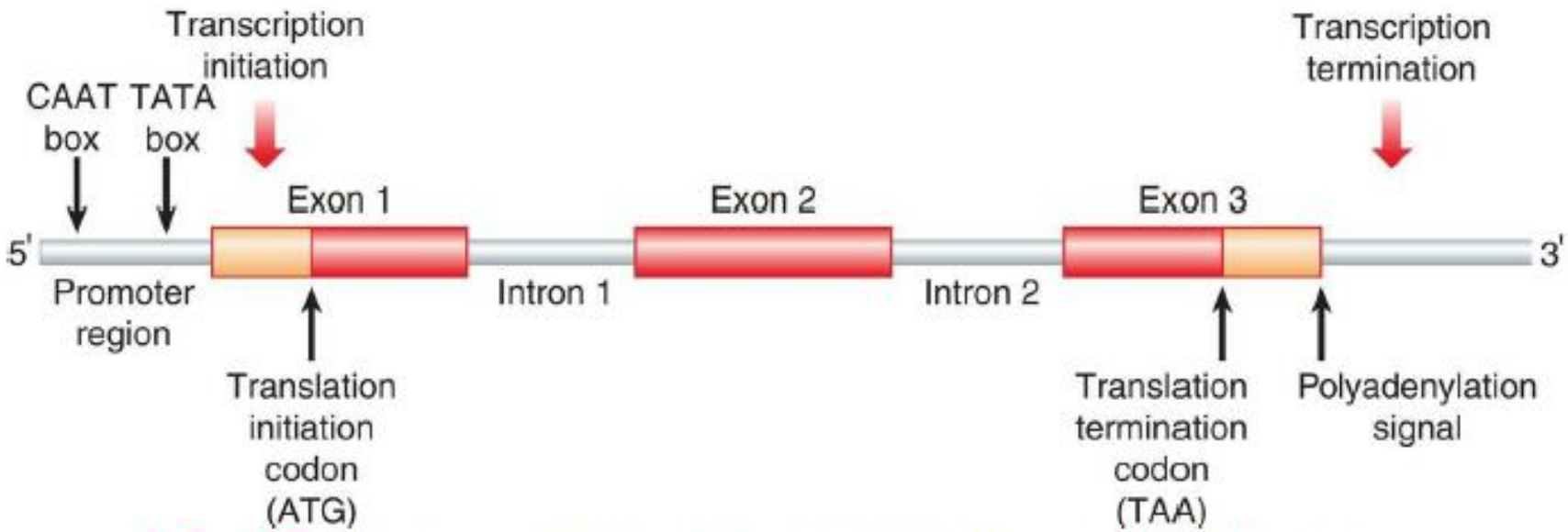
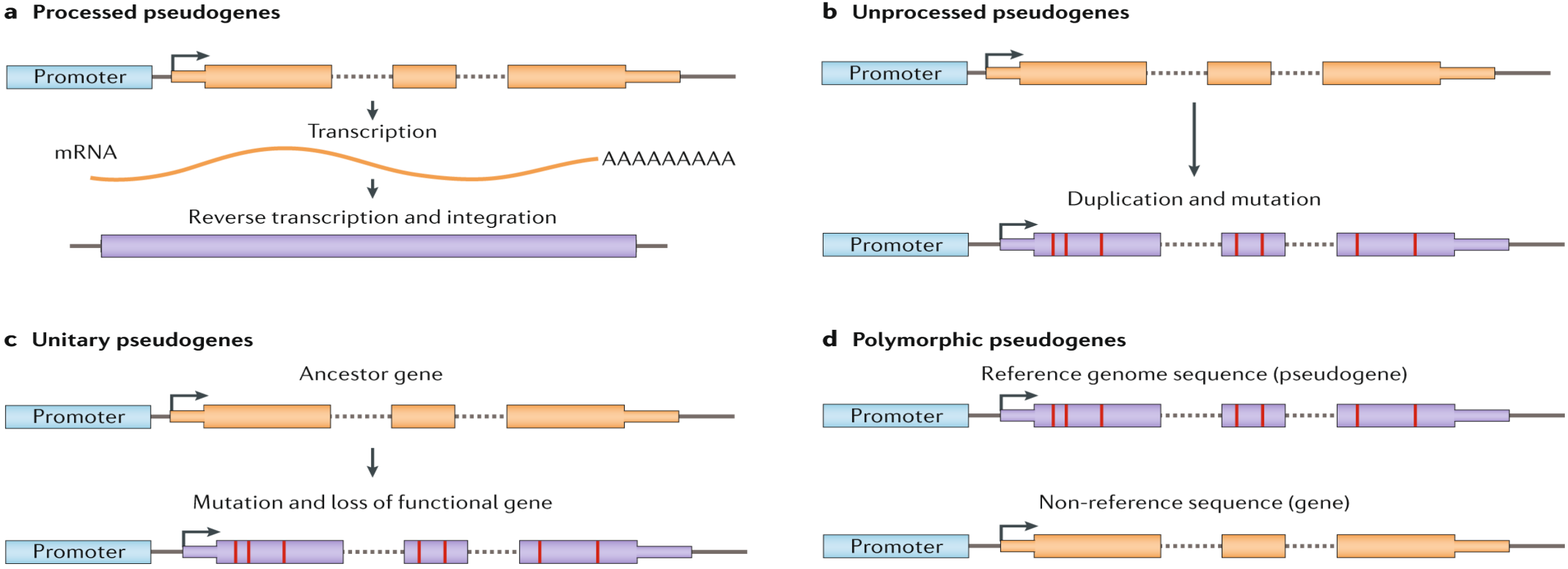


FIG. 2.6 Representation of a typical human structural gene.

Pseudogenes

Particularly fascinating is the occurrence of genes that closely resemble known structural genes but which, in general, are not functionally expressed: so-called pseudogenes.

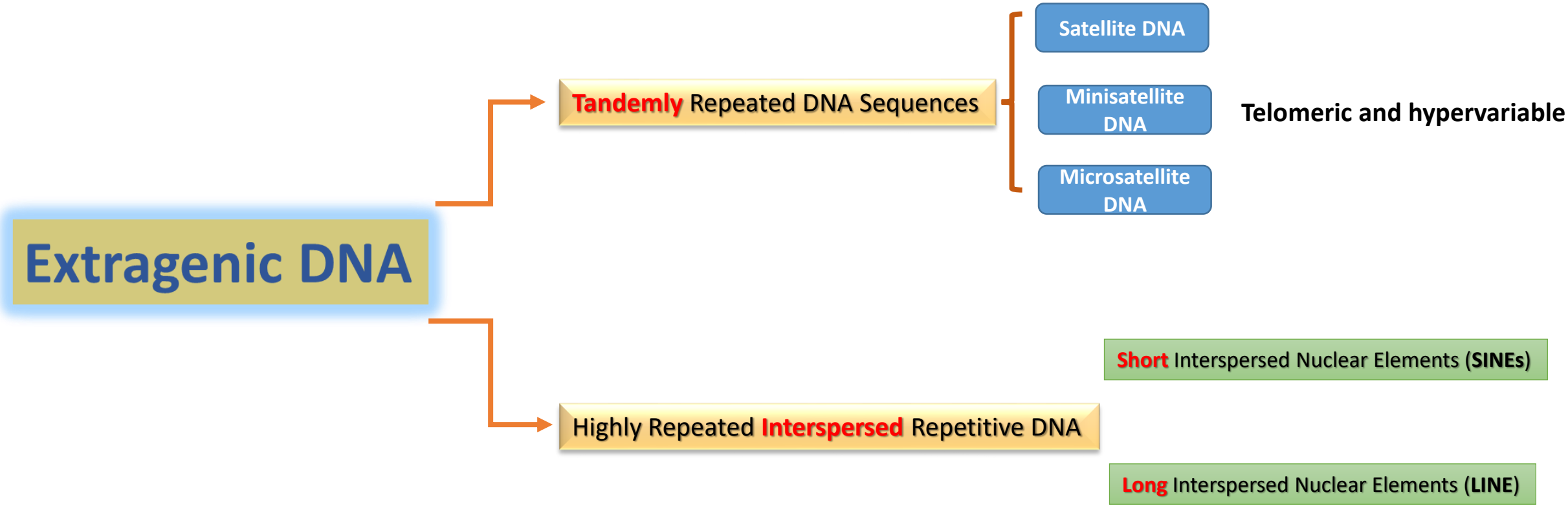


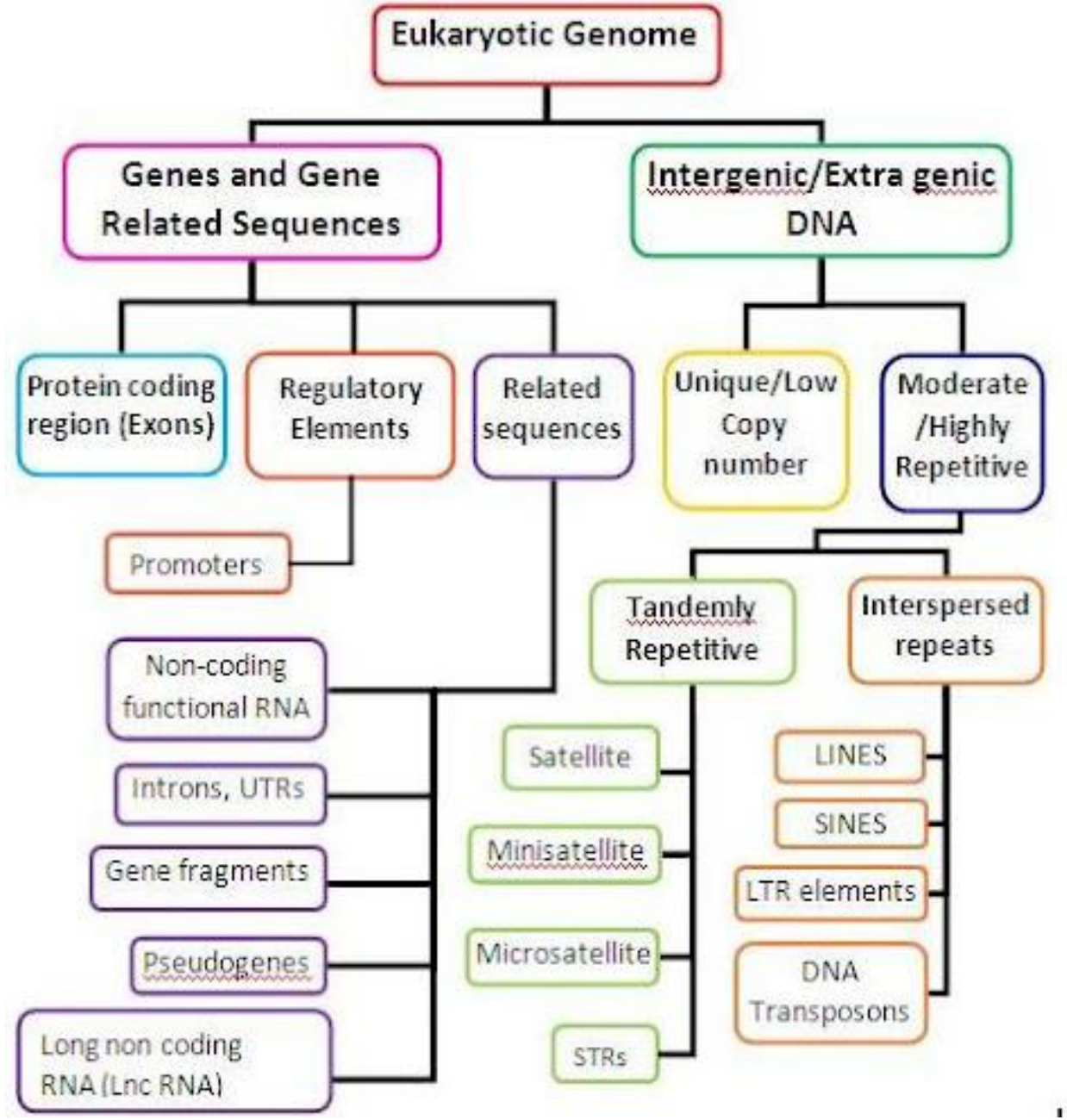
Extra genic DNA

The estimated 20,000 unique single-copy genes that encode proteins represent less than 2% of the human genome.

- The remainder of the human genome is made up of repetitive DNA sequences that are predominantly transcriptionally inactive.

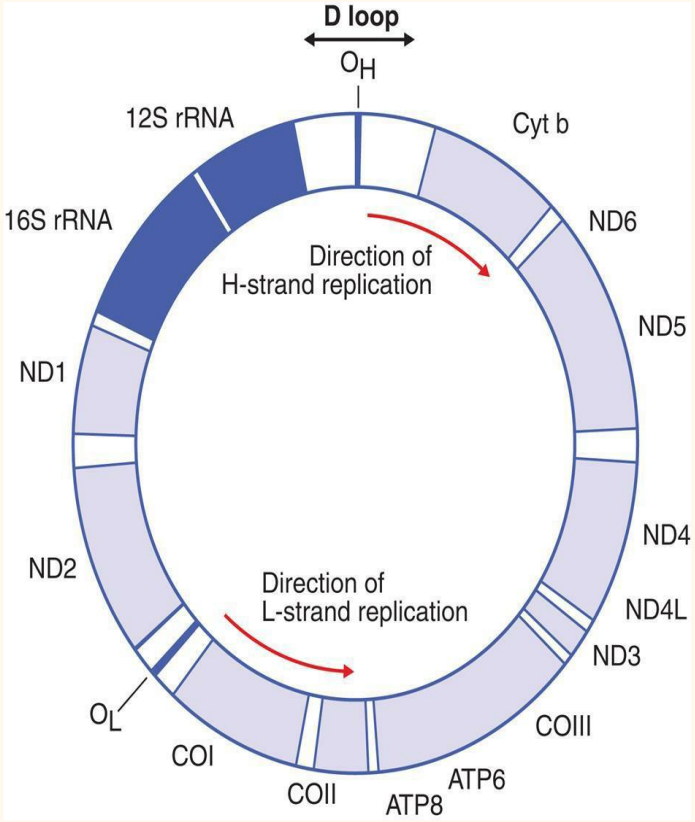
- This has been described as **junk DNA**, but some regions show evolutionary conservation and play a critical role in the regulation of temporal and spatial gene expression.


Extragenic DNA
Tandemly Repeated DNA Sequences
Satellite DNA
Minisatellite DNA
Microsatellite DNA
Telomeric and hypervariable
Highly Repeated Interspersed Repetitive DNA
Short Interspersed Nuclear Elements (SINEs)
Long Interspersed Nuclear Elements (LINE)


Eukaryotic Genome
Genes and Gene Related Sequences
Intergenic/Extra genic DNA
Protein coding region (Exons)
Regulatory Elements
Related sequences
Unique/Low Copy number
Moderate /Highly Repetitive
Promoters
Tandemly Repetitive
Interspersed repeats
Non-coding functional RNA
Introns, UTRs
Gene fragments
Pseudogenes
Long non coding RNA (Lnc RNA)
Satellite
Minisatellite
Microsatellite
STRs
LINES
SINES
LTR elements
DNA Transposons

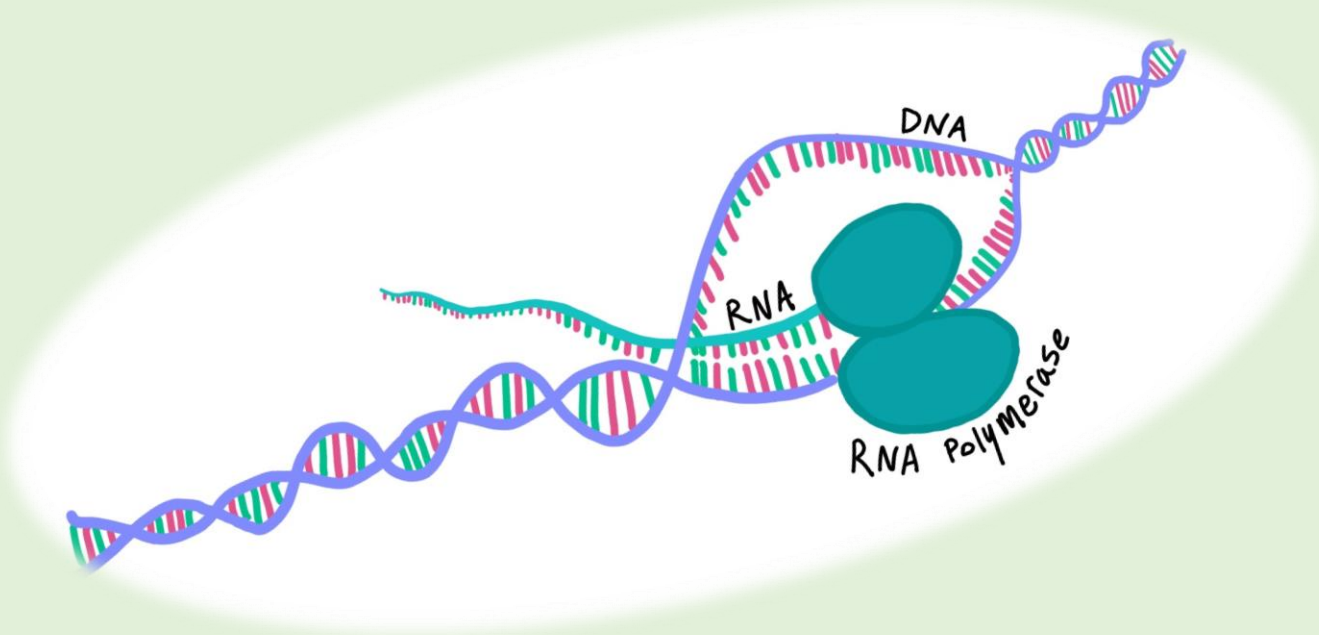
Mitochondrial DNA

In addition to nuclear DNA, the several thousand mitochondria of each cell possess their 16.6-kb circular double-stranded DNA, mitochondrial DNA (mtDNA)

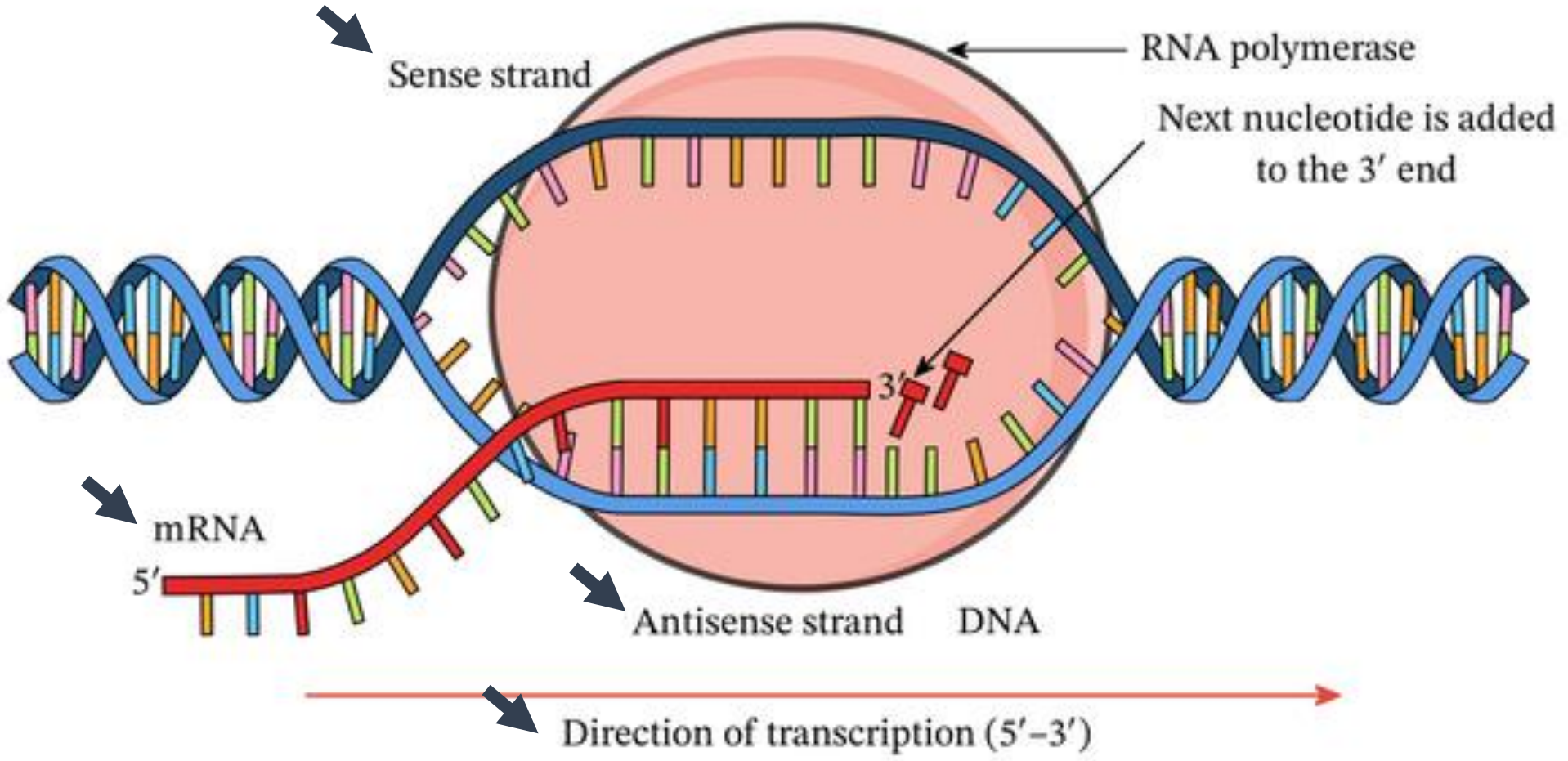


The mitochondrial genome is very compact, containing little repetitive DNA, and codes for 37 genes, which include two rRNA, 22 tRNAs and 13 protein subunits for enzymes, such as cytochrome b and cytochrome oxidase, which are involved in the energy-producing oxidative phosphorylation pathways.

Transcription

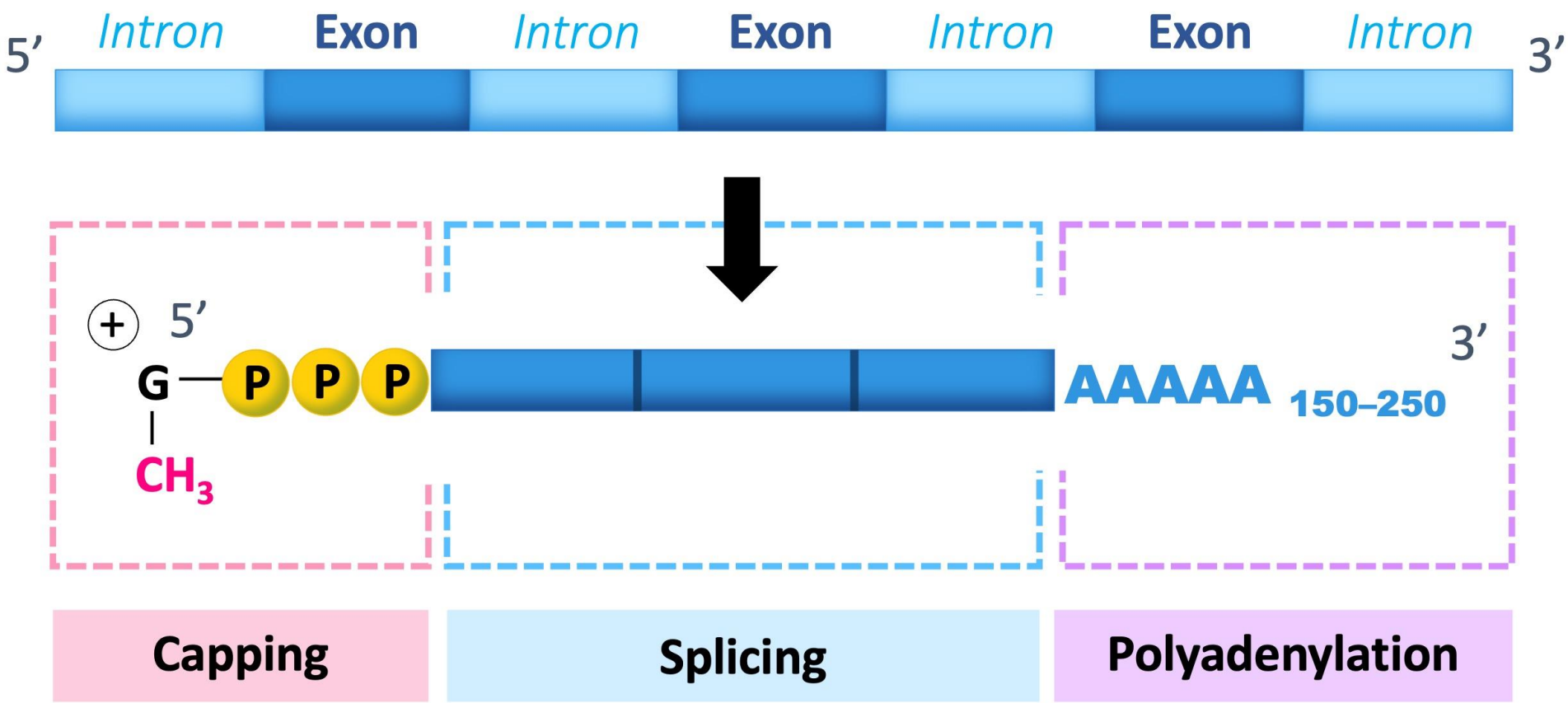


The process whereby genetic information is transmitted from **DNA** to **RNA** is called **transcription.**



RNA Processing

Before the primary mRNA molecule leaves the nucleus it undergoes a number of **modifications**, or what is known as RNA processing. This involves **splicing, capping, and polyadenylation**.

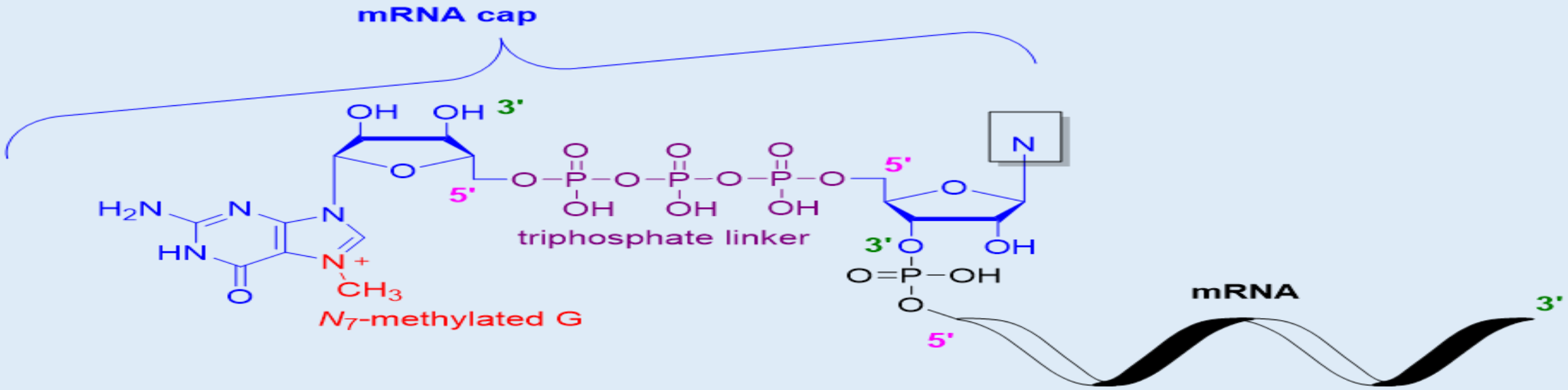


Capping

The **5’ cap** is thought to facilitate the **transport of the mRNA to the cytoplasm** and **attachment to the ribosomes**, as well as to **protect the RNA transcript from degradation by endogenous cellular exonucleases**.

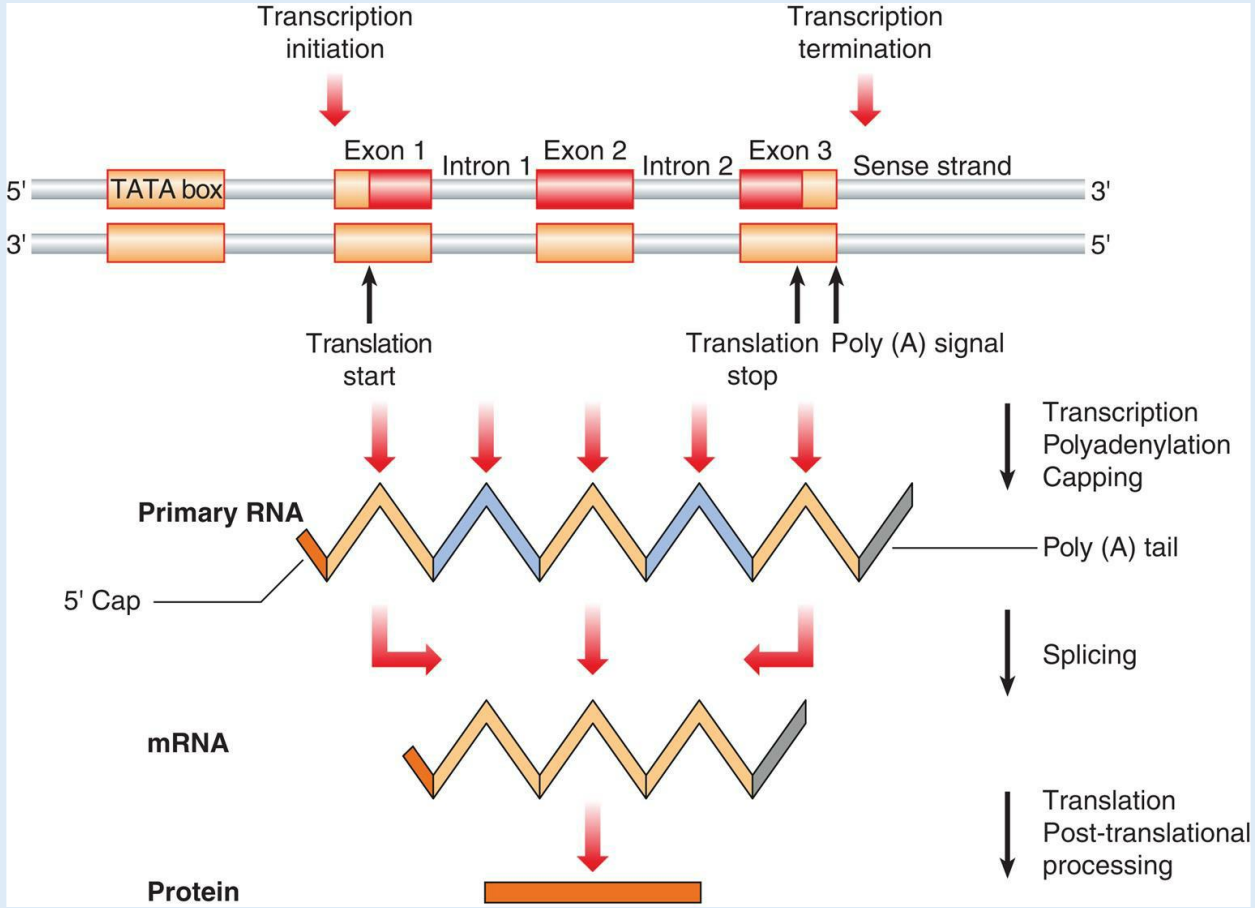
After **20 to 30** nucleotides have been transcribed, the nascent mRNA is modified by the addition of a **G nucleotide** to the 5’ end of the molecule by an unusual 5’ to 5’ triphosphate linkage.

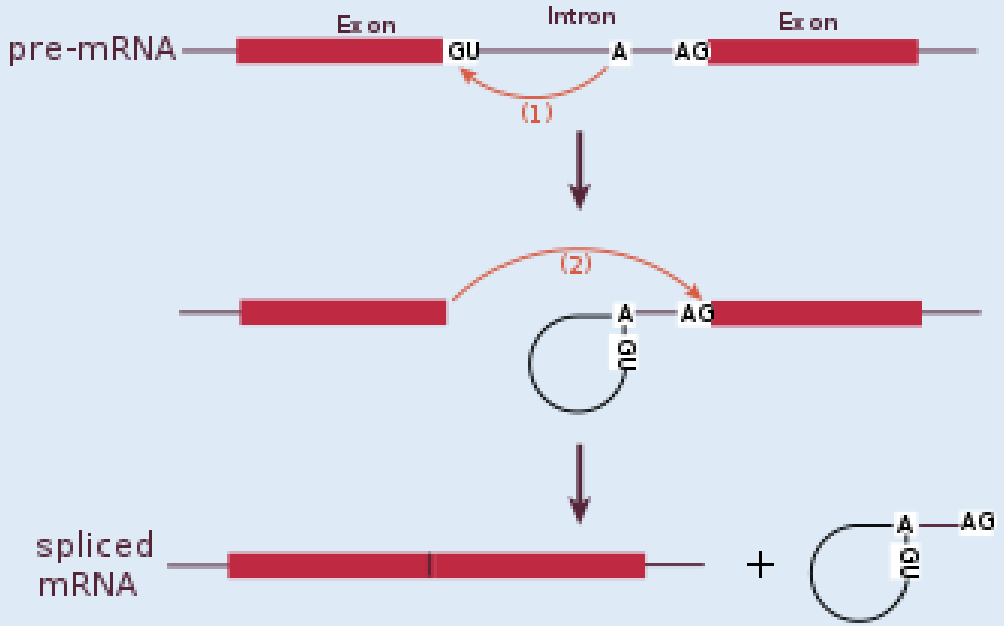
A methyltransferase enzyme then methylates the N7 position of the G, giving the final 5’ cap.



mRNA Splicing

The **non-coding introns** in the precursor mRNA are excised, and the non-contiguous coding **exons** are spliced together to form a shorter mature mRNA **before its transportation to the ribosomes in the cytoplasm for translation**.

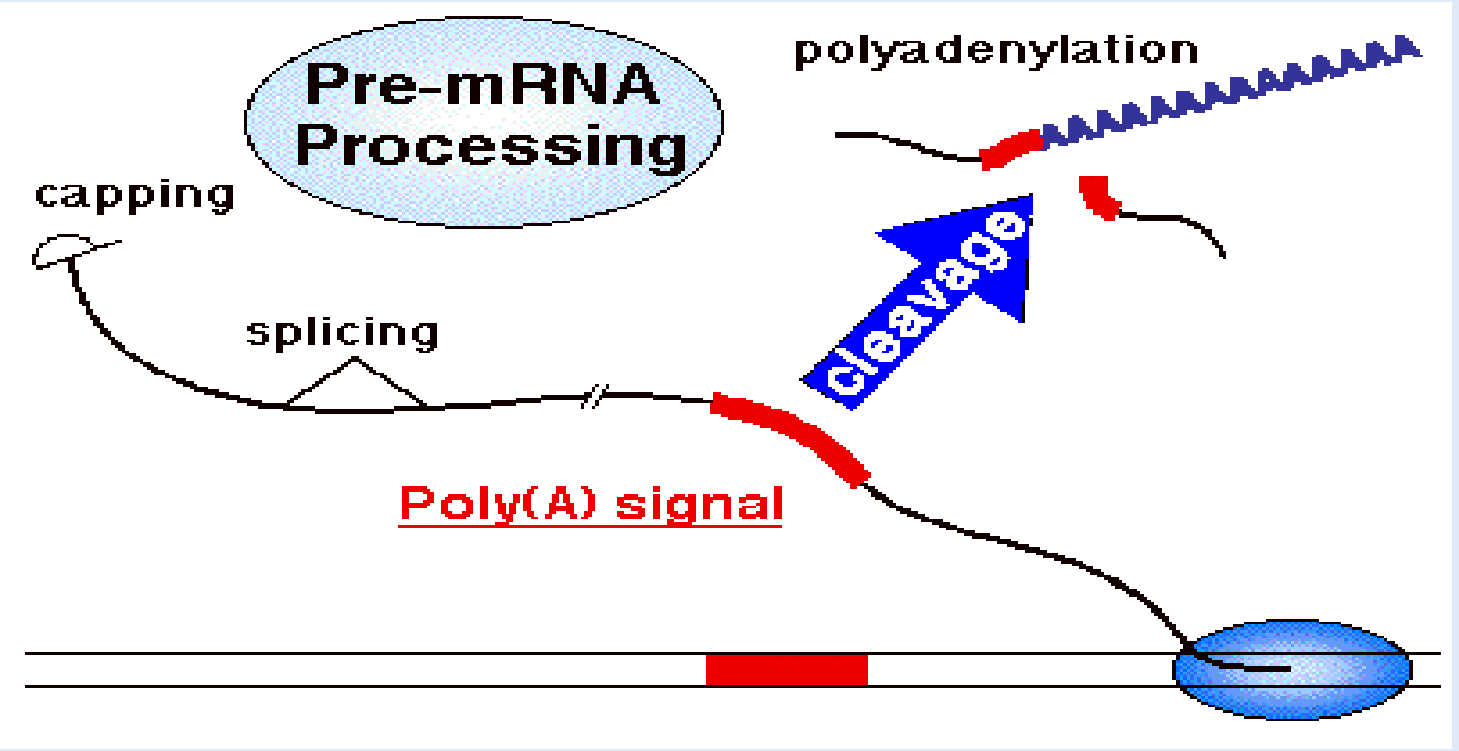




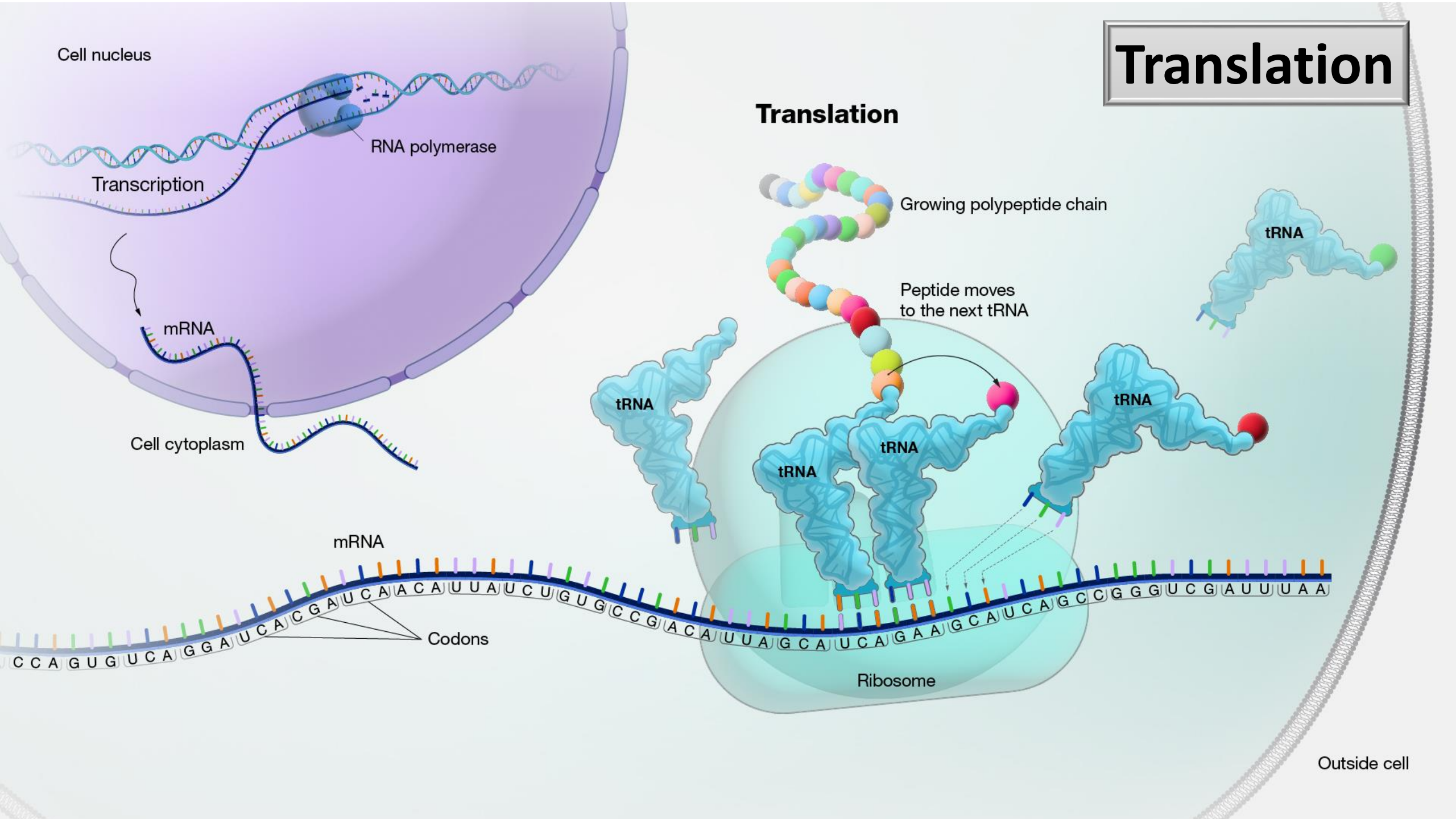
Polyadenylation

- Transcription continues until specific nucleotide sequences are transcribed that cause the mRNA to be cleaved and RNA polymerase II to be released from the DNA template.

- Approximately 200 adenylate residues—the so-called poly(A) tail—are added to the mRNA, which **facilitates nuclear export** and **translation**.



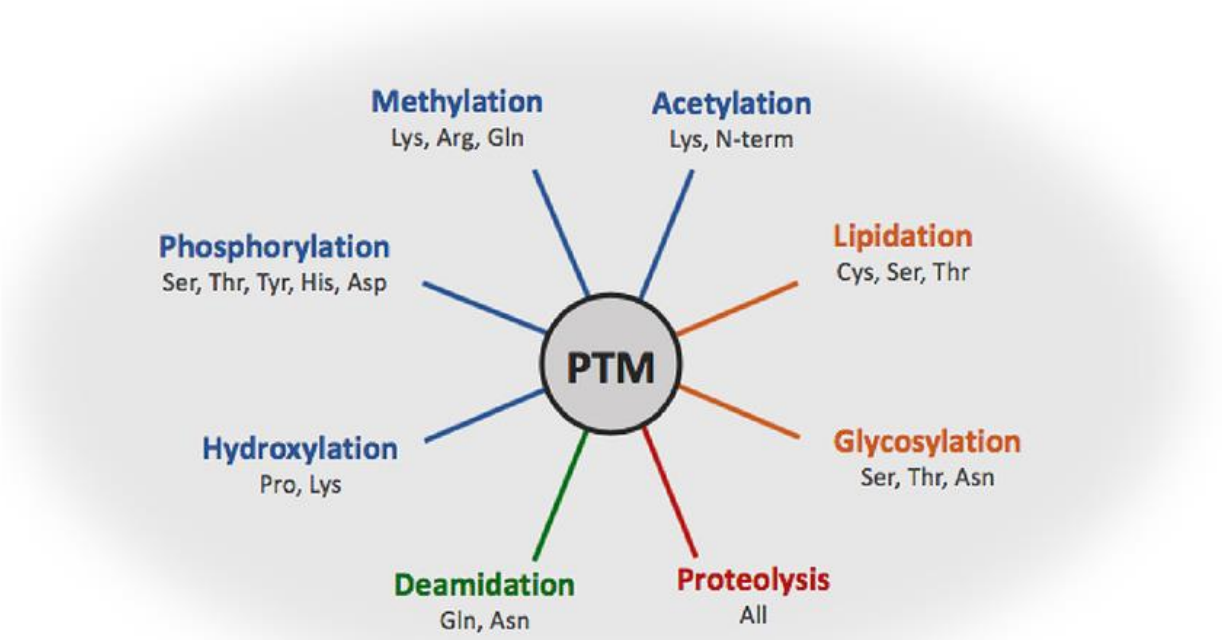
Translation

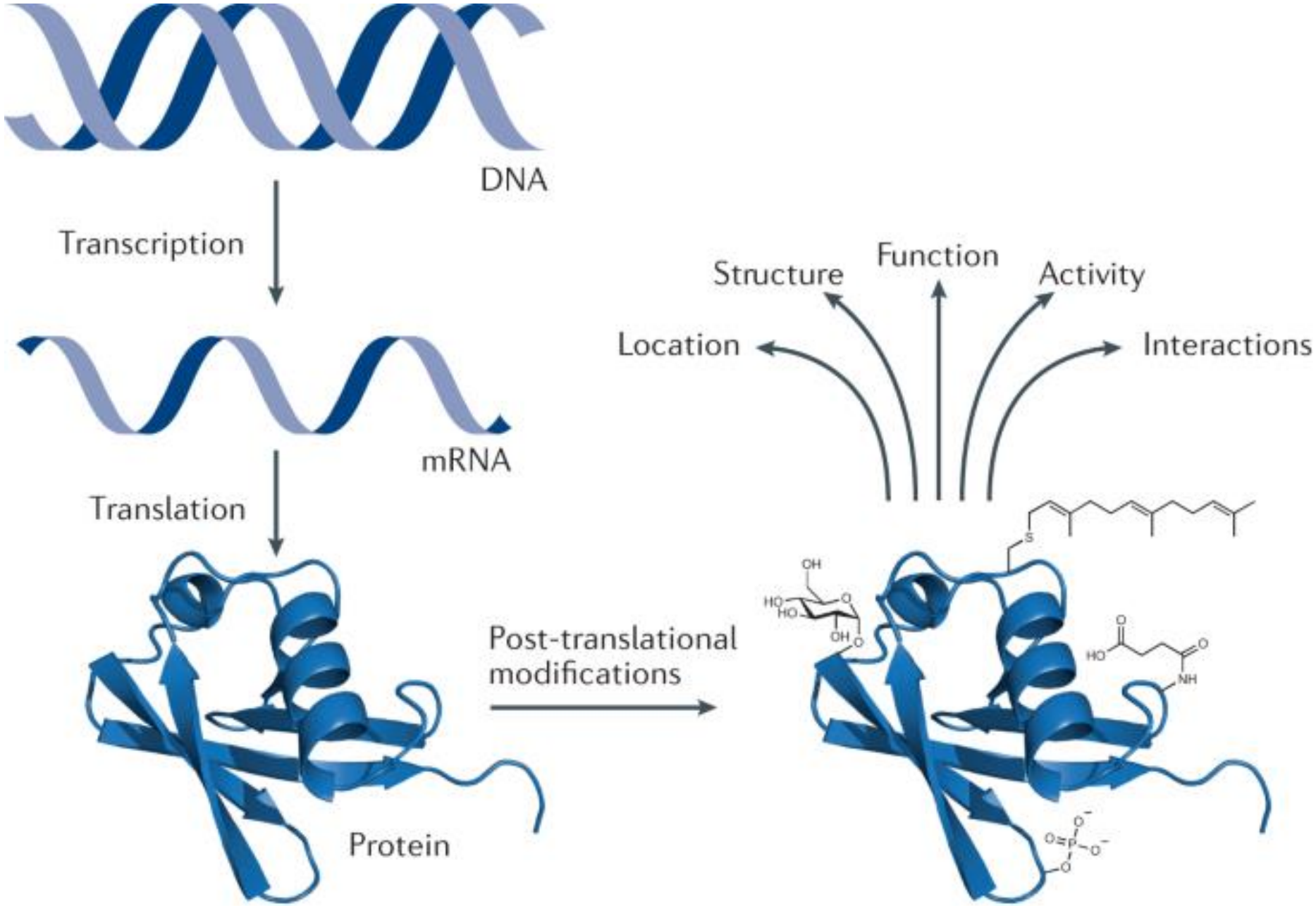


Post-translational Modification

Many proteins**, before they attain their normal structure or functional activity**, undergo **post-translational modification**, which can include

- **chemical modification of amino-acid side chains (e.g., hydroxylation, methylation)**
- **addition of carbohydrate or lipid moieties (e.g., glycosylation)**
- **proteolytic cleavage of polypeptides (e.g., the conversion of proinsulin to insulin)**.



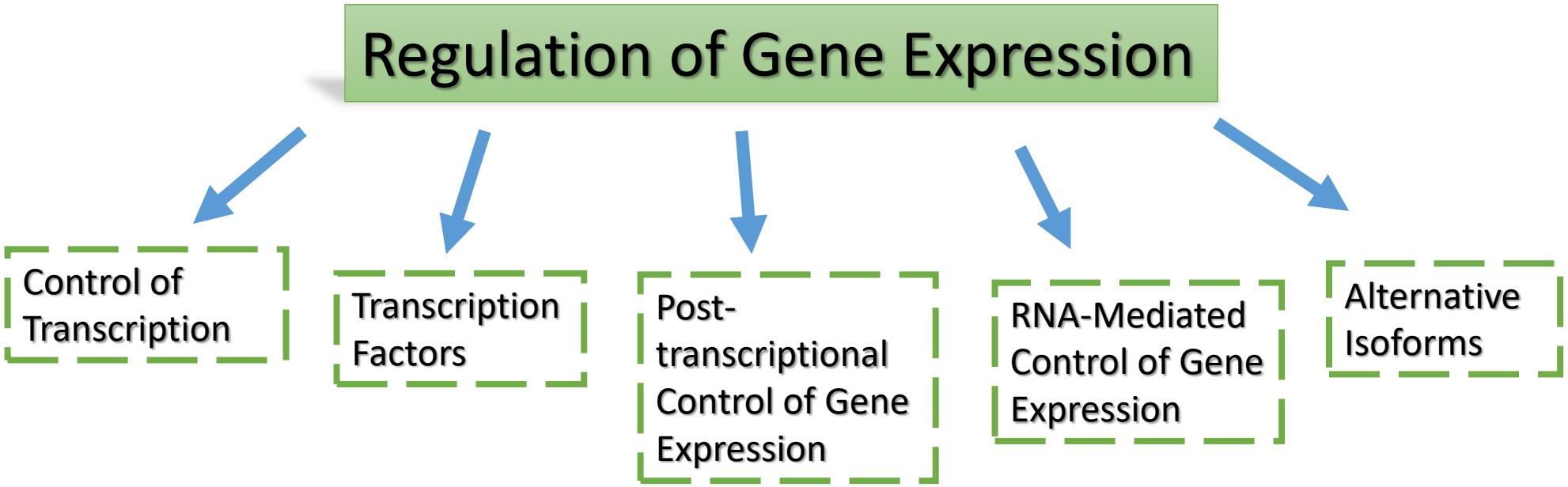

DNA
Transcription
mRNA
Translation
Protein
Post-translational
modifications
Structure
Function
Activity
Location
Interactions

Regulation of Gene Expression

- Many cellular processes, and therefore the genes that are expressed, are common to all cells, for example ribosomal, chromosomal, and cytoskeleton proteins, constituting what are called the **housekeeping** genes.

- Some cells express large quantities of a specific protein in certain tissues or at specific times in development, such as hemoglobin in red blood cells.

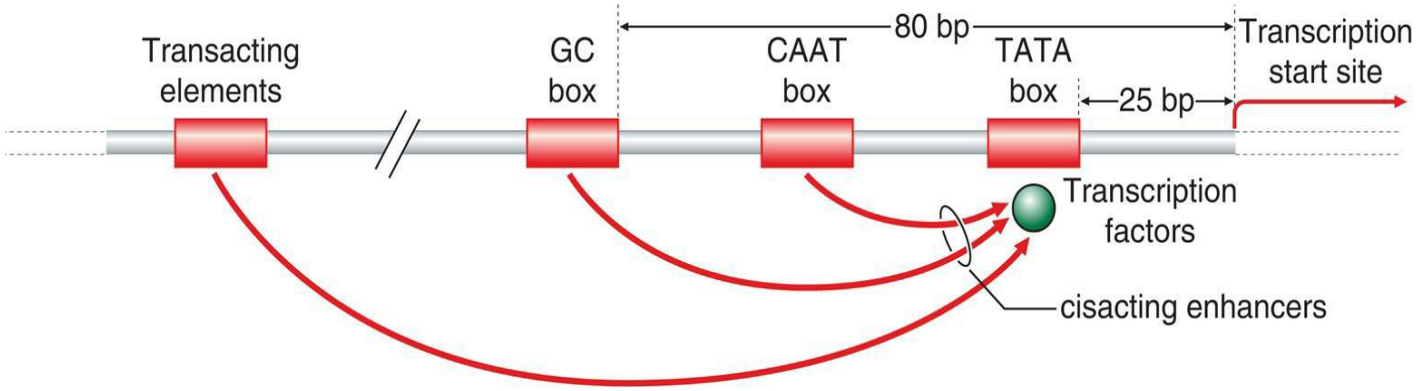
- This differential control of gene expression can occur at a variety of stages.


Regulation of Gene Expression
Control of Transcription
Transcription Factors
Post-transcriptional Control of Gene Expression
RNA-Mediated Control of Gene Expression
Alternative Isoforms

Control of Transcription

- This occurs through a number of different mechanisms that include signaling molecules that bind to regulatory sequences in the DNA known as response elements,
- **Intracellular receptors** known as **hormone nuclear receptors** and **receptors** for specific ligands on the cell surface involved in the process of signal transduction.

All of these mechanisms ultimately affect transcription through the binding of the general **transcription factors** to short specific DNA **promoter elements** located within 200 bp 5' or upstream of most eukaryotic genes in the so-called core promoter region that leads to activation of RNA polymerase.

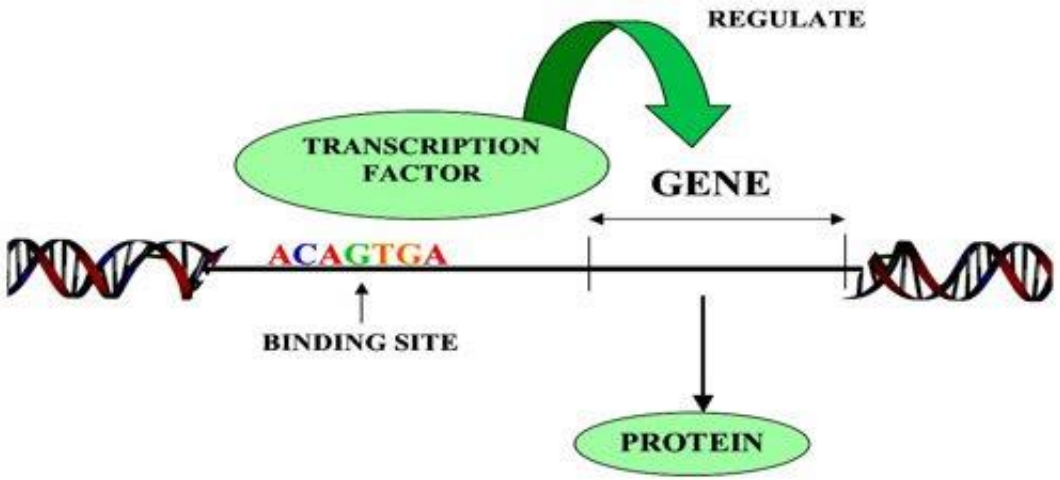


Transcription Factors

- Several genes encode proteins involved in the regulation of gene expression.

- These proteins bind short nucleotide sequences, usually mediated through helical protein motifs, and are known as transcription factors.

- These gene regulatory proteins have a **transcriptional activation domain** and a **DNA-binding domain**.
- There are four types of DNA-binding domains, the most common being the **helix–turn–helix**, made up of two α helices connected by a short chain of amino acids that make up the “turn.”
- The three other types are the **zinc finger**, **leucine zipper,** and **helix–loop–helix** motifs, so named as a result of specific structural features.



Post-transcriptional Control of Gene Expression

Regulation of the expression of most genes occurs at the level of transcription, but can also occur at the levels of :

- ❖RNA processing
- ❖RNA transport
- ❖mRNA degradation
- mRNA translation.

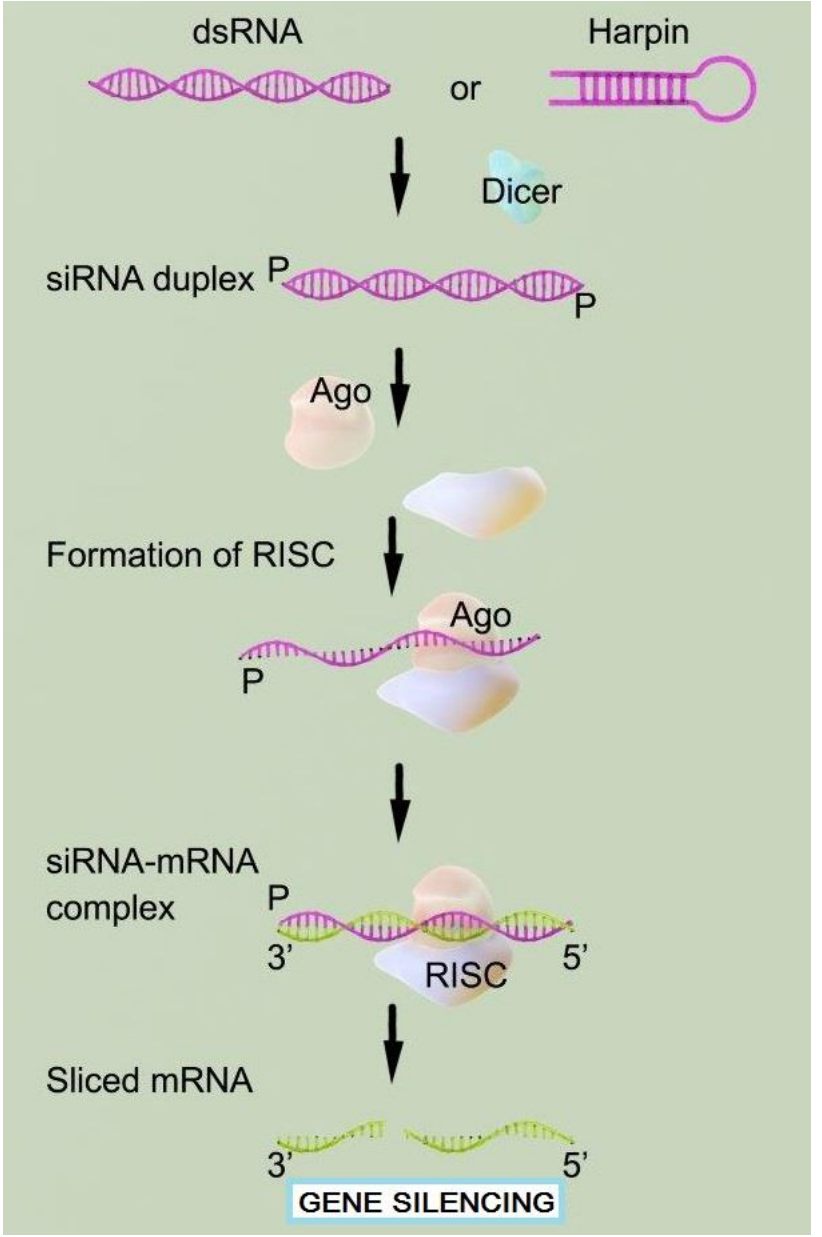
For example, the G to A variant at position 20,210 in the 3’ untranslated region of the prothrombin gene increases the stability of the mRNA transcript, resulting in higher plasma prothrombin levels.

RNA-Mediated Control of Gene Expression

Small interfering RNAs (siRNAs) were discovered in 1998, and are the effector molecules of the RNA interference pathway.

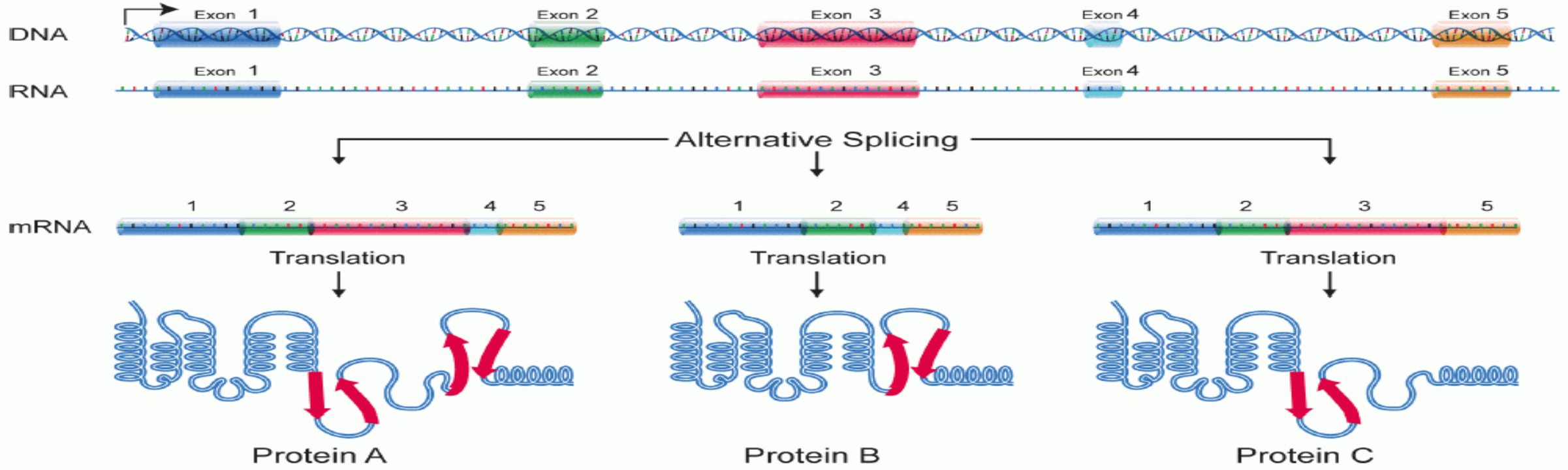
- These short double-stranded RNAs (21–23 nucleotides) **bind to mRNAs** in a sequence-specific manner and result in their degradation via **a ribonuclease-containing RNA-induced silencing complex**.

- MicroRNAs (miRNAs) also bind to mRNAs in a sequence-specific manner.
- They can either cause **endonucleolytic** cleavage of the mRNA or **act by blocking translation.**



Alternative Isoforms

- Most (~95%) human genes undergo alternative splicing and **therefore encode more than one protein**.
- Alternative polyadenylation generates further diversity.
- Some genes have **more than one promoter**, and these alternative promoters may result in tissue-specific isoforms.
- **Alternative splicing of exons** is also seen with individual exons present in only some isoforms.
- The extent of alternative splicing in humans may be inferred from the finding that the human genome includes only approximately **20,000** genes, far fewer than the original prediction of more than **100,000**.



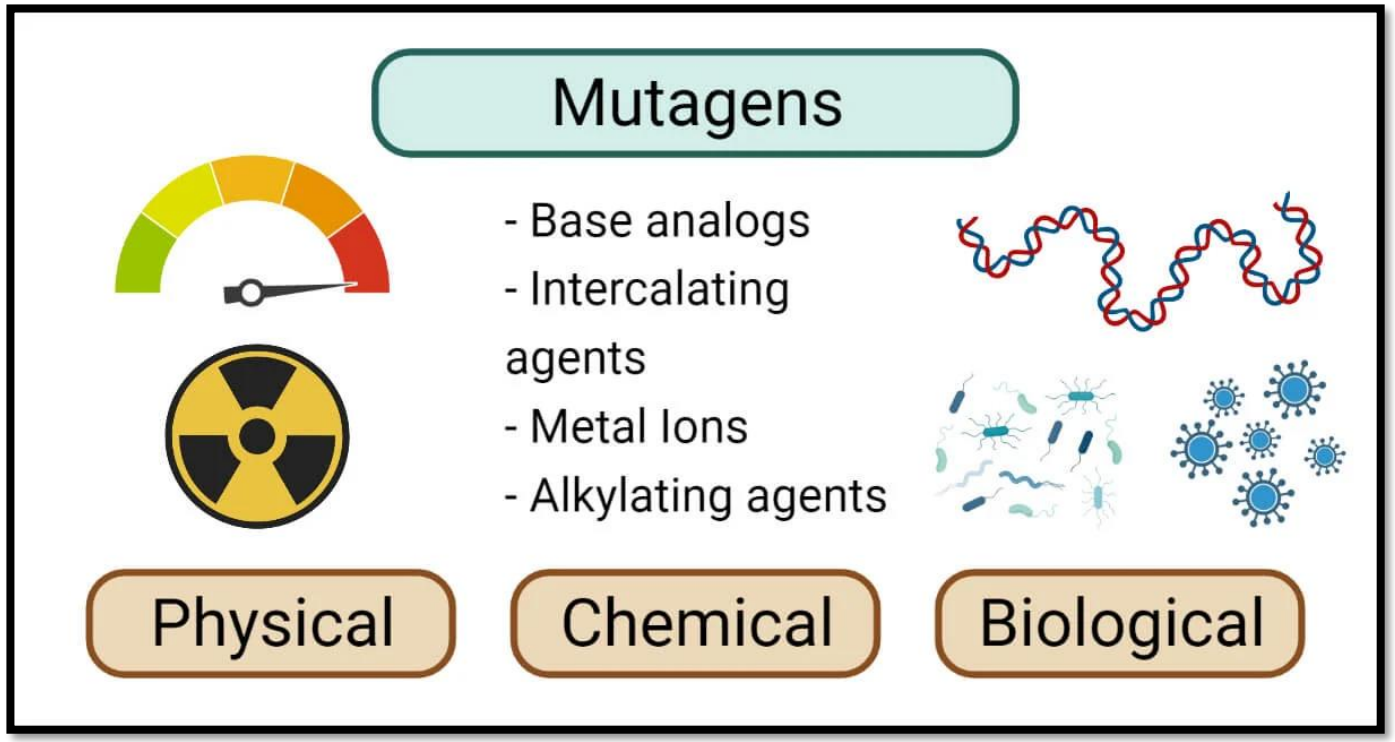
Mutations

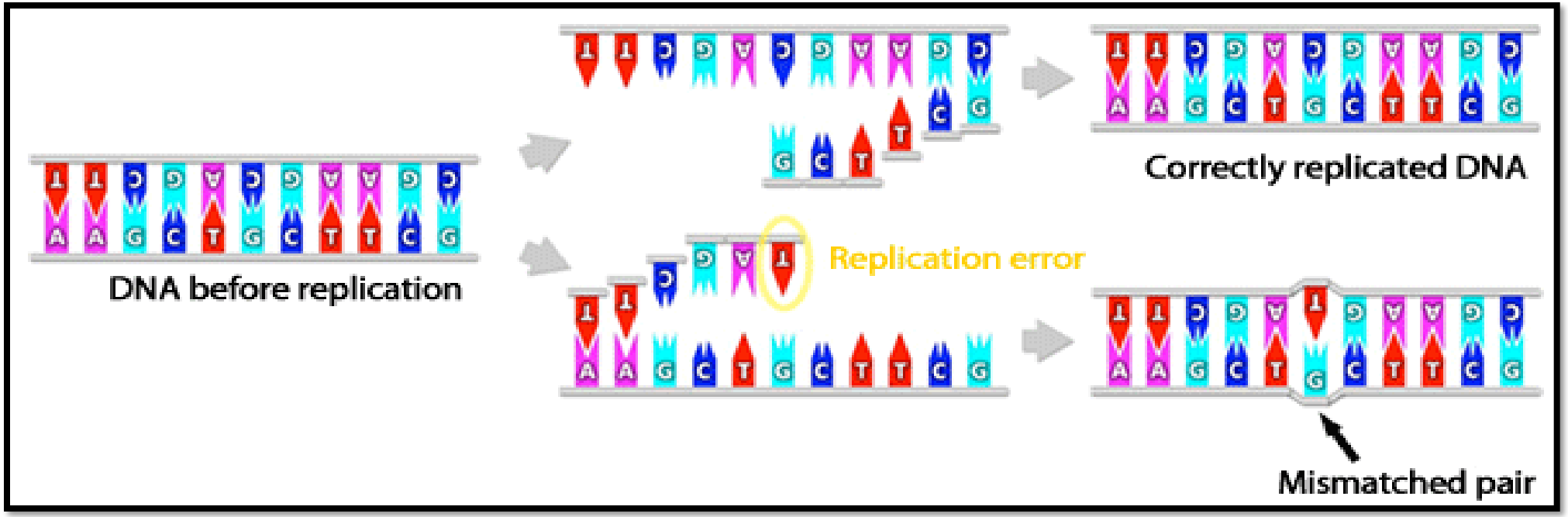
A mutation is defined as a heritable alteration or change in the genetic material

Mutations drive evolution, but can also be pathogenic.

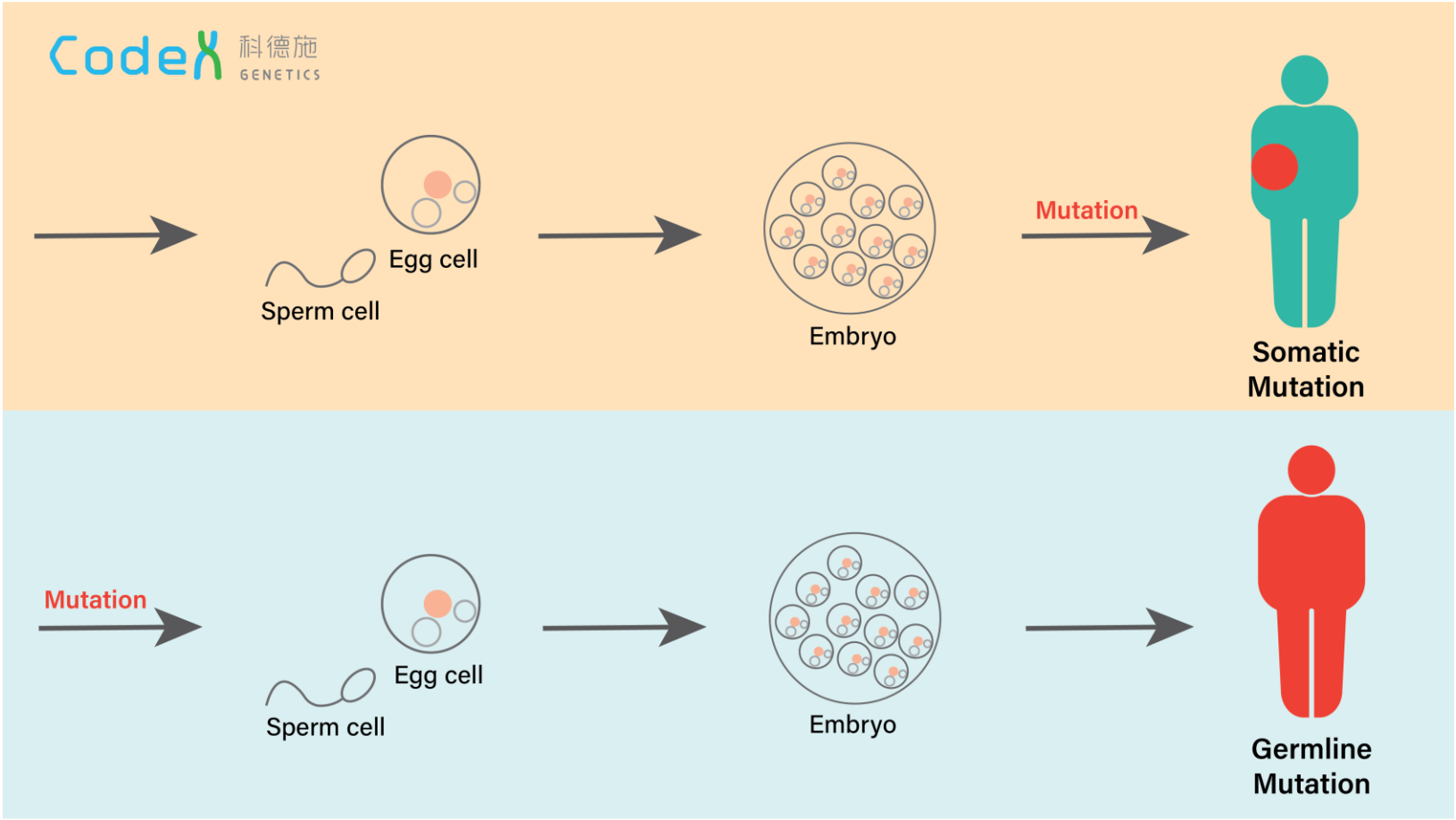
Sequence variants with no obvious effect on phenotype may be termed ***polymorphisms***.

Mutations can arise through exposure to **mutagenic agents** , but the vast majority occur spontaneously through errors in **DNA replication and repair**. Sequence variants with no obvious effect on phenotype may be termed polymorphisms.



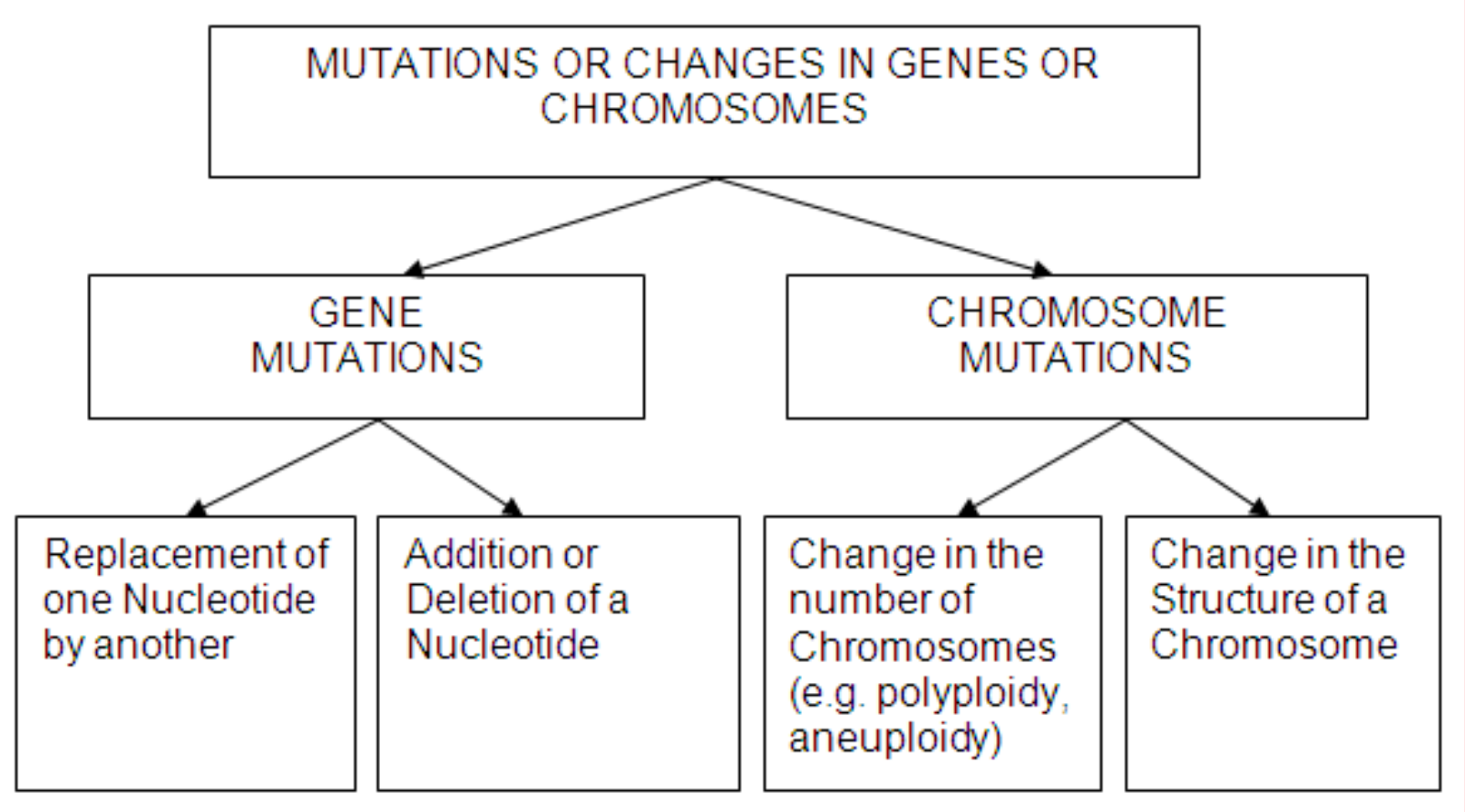


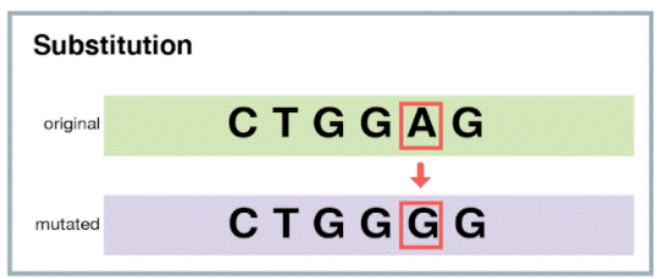
Somatic mutations may cause adult-onset disease, such as **cancer**, but cannot be transmitted to offspring. A mutation in **gonadal tissue** or a **gamete** can be transmitted to future generations, unless it affects fertility or survival into adulthood.



Types of Mutation

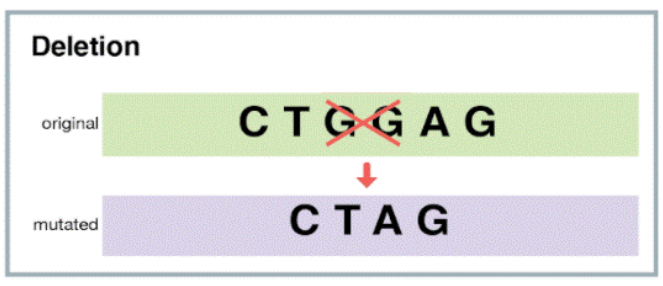
Mutations can range from **single base** substitutions, through insertions and deletions of single or multiple bases, to loss or gain of **entire chromosomes.**



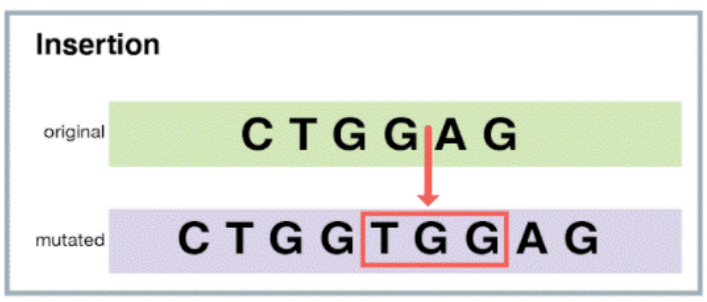


A substitution is the replacement of a single nucleotide by another.
This is the most common type of mutation.

Transition
Transversion



A deletion involves the loss of one or more nucleotides.
Frameshift
Deletion of one or more residues



An insertion involves the addition of one or more nucleotides to a gene.
Frameshift
insertion of one or more residues

Main classes, groups and types of mutations and effects on protein products

Class	Group	Type	Effect on Protein Product
Substitution	Synonymous	Silent[a]	Same amino acid
	Non-synonymous	Missense[a]	Altered amino acid—may affect protein function or stability
		Nonsense[a]	Stop codon—loss of function or expression because of degradation of mRNA
		Splice site	Aberrant splicing—exon skipping or intron retention
		Promoter	Altered gene expression
		Enhancer	Altered gene expression
Deletion	Multiple of three (codon)		In-frame deletion of one or more amino acid(s)—may affect protein function or stability
	Not multiple of three	Frameshift	Likely to result in premature termination with loss of function or expression
	Large deletion	Partial gene deletion	May result in premature termination with loss of function or expression
		Whole gene deletion	Loss of expression
Insertion	Multiple of three (codon)		In-frame insertion of one or more amino acid(s)—may affect protein function or stability
	Not multiple of three	Frameshift	Likely to result in premature termination with loss of function or expression
	Large insertion	Partial gene duplication	May result in premature termination with loss of function or expression
		Whole gene duplication	May have an effect because of increased gene dosage
	Expansion of trinucleotide repeat	Dynamic mutation	Altered gene expression or altered protein stability or function

Structural Effects of Mutations on the Protein

Synonymous or Silent Mutations

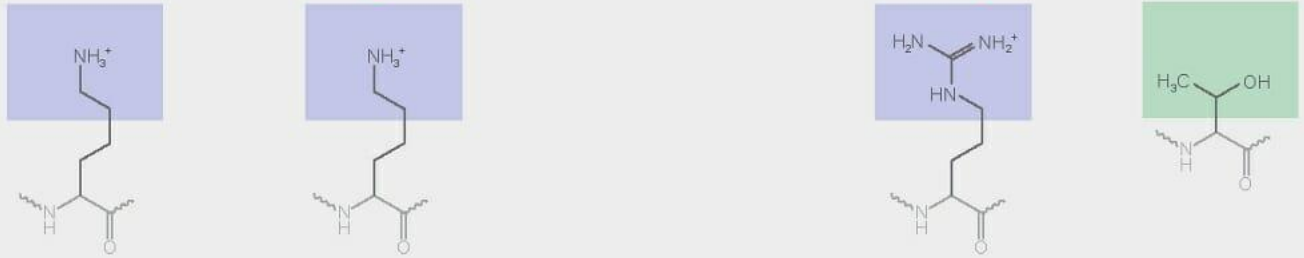
Non-synonymous Mutations

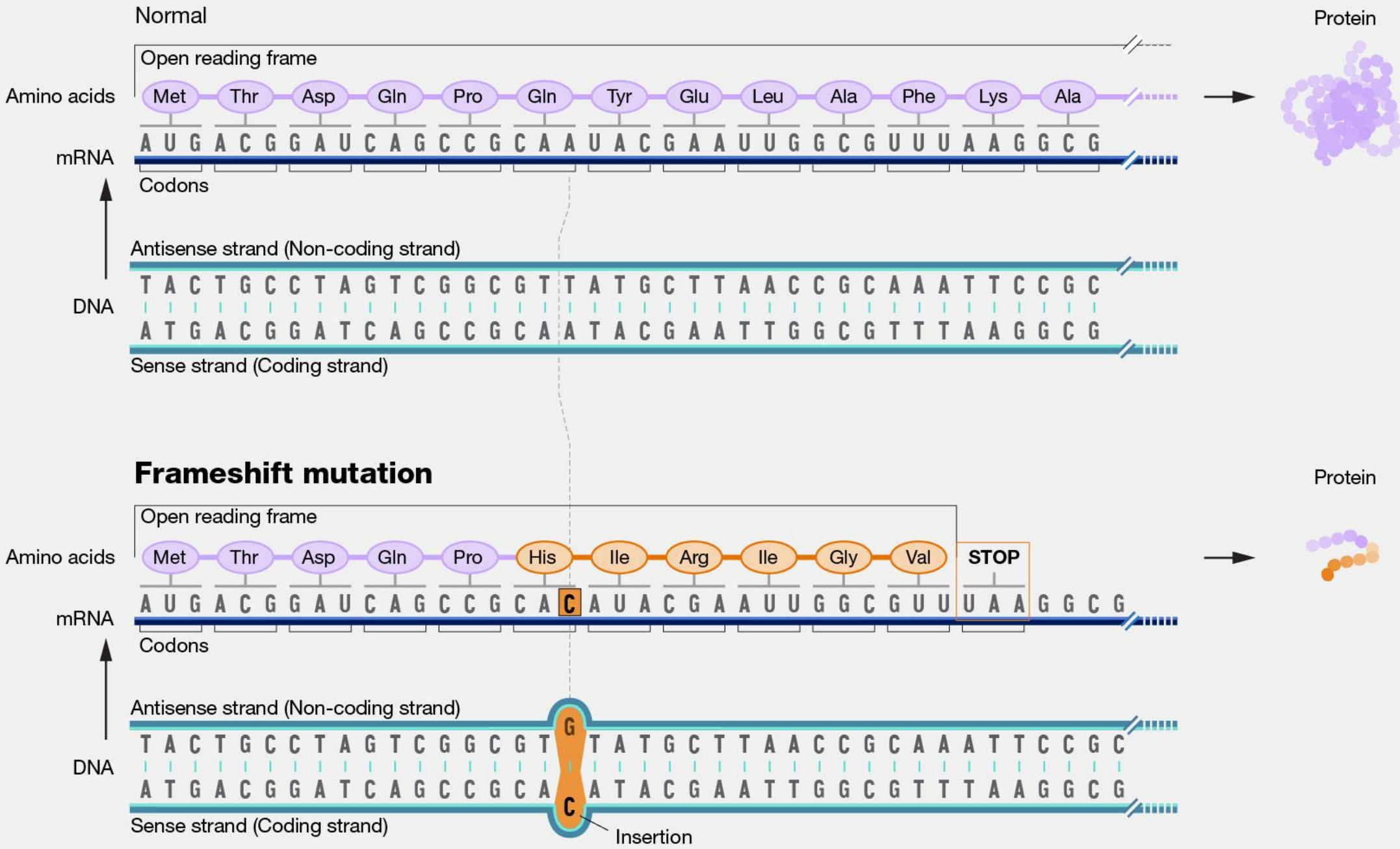
Missense

Nonsense

Frameshift

	No mutation	Point mutations			
		Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr




Normal
Protein
Open reading frame
Amino acids
Met Thr Asp Gln Pro Gln Tyr Glu Leu Ala Phe Lys Ala
mRNA
A U G A C G G A U C A G C C G C A A U A C G A A U U G G C G U U U A A G G C G
Codons
Antisense strand (Non-coding strand)
DNA
T A C T G C C T A G T C G G C G T T A T G C T T A A C C G C A A A T T C C G C
A T G A C G G A T C A G C C G C A A T A C G A A T T G G C G T T T A A G G C G
Sense strand (Coding strand)
Frameshift mutation
Protein
Open reading frame
Amino acids
Met Thr Asp Gln Pro His Ile Arg Ile Gly Val STOP
mRNA
A U G A C G G A U C A G C C G C A C A U A C G A A U U G G C G U U U A A G G C G
Codons
Antisense strand (Non-coding strand)
DNA
T A C T G C C T A G T C G G C G T G T A T G C T T A A C C G C A A A T T C C G C
A T G A C G G A T C A G C C G C A C A T A C G A A T T G G C G T T T A A G G C G
Sense strand (Coding strand)
Insertion

dynamic mutations

A number of single-gene disorders have subsequently been shown to be associated with triplet repeat expansions.

Disease (Gene)	Repeat Sequence	Normal Range (Repeats)	Pathogenic Range (Repeats)	Repeat Location
Huntington disease (*HTT*)	CAG	9–35	36–100	Coding
Myotonic dystrophy type 1 (*DMPK*)	CTG	5–35	50–4000	3′ UTR
Myotonic dystrophy type 2 (*CNBP*)	CCTG	11–26	75–>11000	Intron 1
Fragile X site A (*FMR1*)	CGG	10–50	200–2000	5′ UTR
Kennedy disease (*AR*)	CAG	13–30	40–62	Coding
Spinocerebellar ataxia 1 (*ATXN1*)	CAG	6–36	39–80	Coding
Spinocerebellar ataxia 2 (*ATXN2*)	CAG	13–31	32–79	Coding
Machado–Joseph disease/Spinocerebellar ataxia 3 (*ATXN3*)	CAG	14–44	52–86	Coding
Spinocerebellar ataxia 6 (*CACNA1A*)	CAG	4–18	19–33	Coding
Spinocerebellar ataxia 7 (*ATXN7*)	CAG	7–17	38–220	Coding
Spinocerebellar ataxia 8 (*ATXN8*)	CTG	15–50	71–1300	3′ UTR
Spinocerebellar ataxia 10 (*ATXN10*)	ATTCT	10–29	400–4500	Intron 9
Spinocerebellar ataxia 12 (*PPP2R2B*)	CAG	7–32	51–78	5′ UTR
Spinocerebellar ataxia 17 (*TBP*)	CAG	25–44	47–63	Coding
Dentatorubral-pallidoluysian atrophy	CAG	7–23	53–88	Coding

Mutation nomenclature: examples of CFTR gene mutations

Type of Mutation	Nucleotide (Ref Seq NM_000492.3)	Protein Designation	Consequence Description
Missense	c.350G>A	p.Arg117His	Arginine to histidine
Nonsense	c.1624G>T	p.Gly542*	Glycine to stop
Splicing	c.489 + 1G>T		Splice donor site mutation
Deletion [1 base pair (bp)]	c.948delT	p.Phe316Leufs*12	Frameshift mutation
Deletion (3 bp)	c.1521_1523delCTT	p.Phe508del	In-frame deletion of phenylalanine
Insertion (1 bp)	c.3767dupC	p.Leu1258Phefs*7	Frameshift mutation

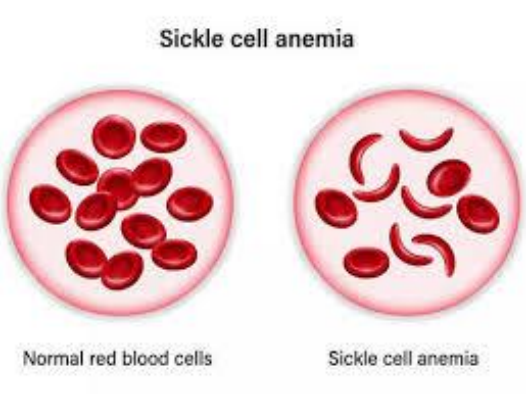
Mutations can be designated according to the genomic or complementary DNA (mRNA) sequence and are prefixed by "g." or "c.," respectively. The first base of the start codon (ATG) is c.1.

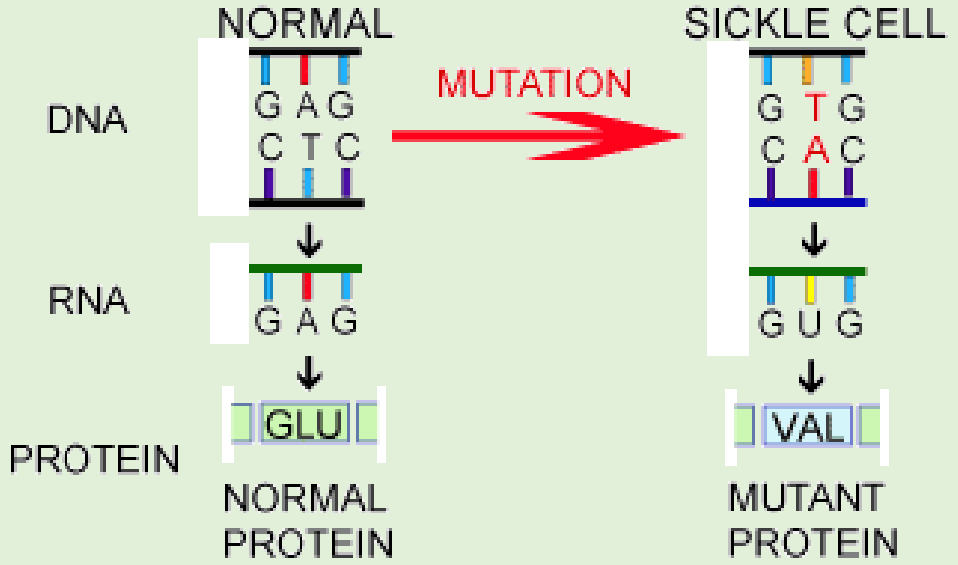
Frequency of different types of mutation

Type of Mutation	Percentage of Total
Missense or nonsense	56
Splicing	9
Regulatory	2
Small deletions, insertions or indels[a]	23
Gross deletions or insertions	9
Other (complex rearrangements or repeat variations)	<1

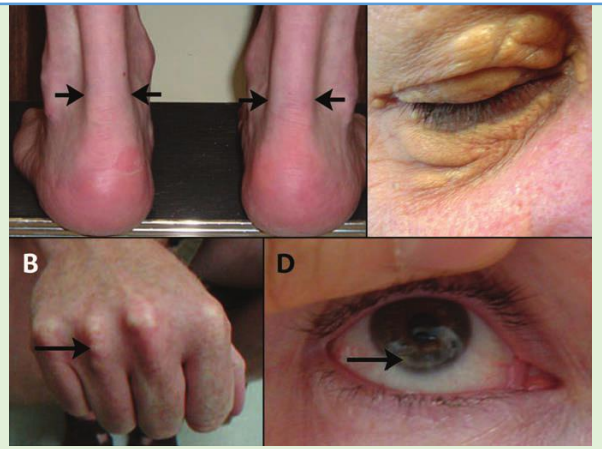
Example of Missense mutation

Sickle cell anemia





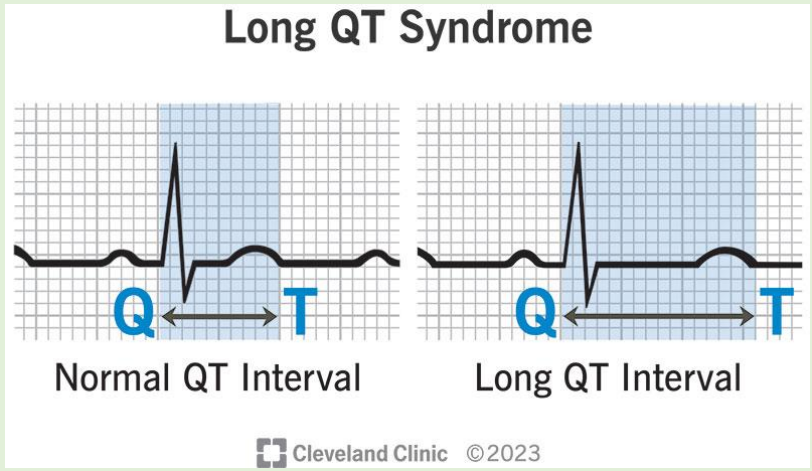
Familial hypercholesterolemia



c.681G>A (p.Gly227Arg)

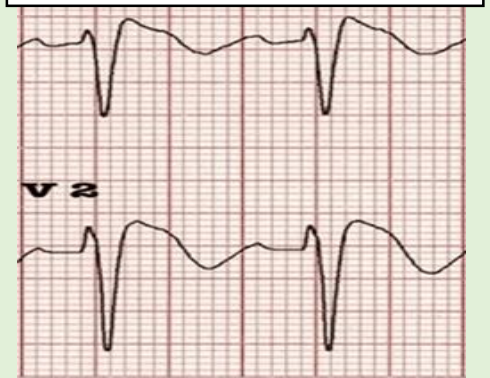
The most common missense mutation in the LDLR

Example of Nonsense mutation



The R518X mutation in the KCNQ1 gene

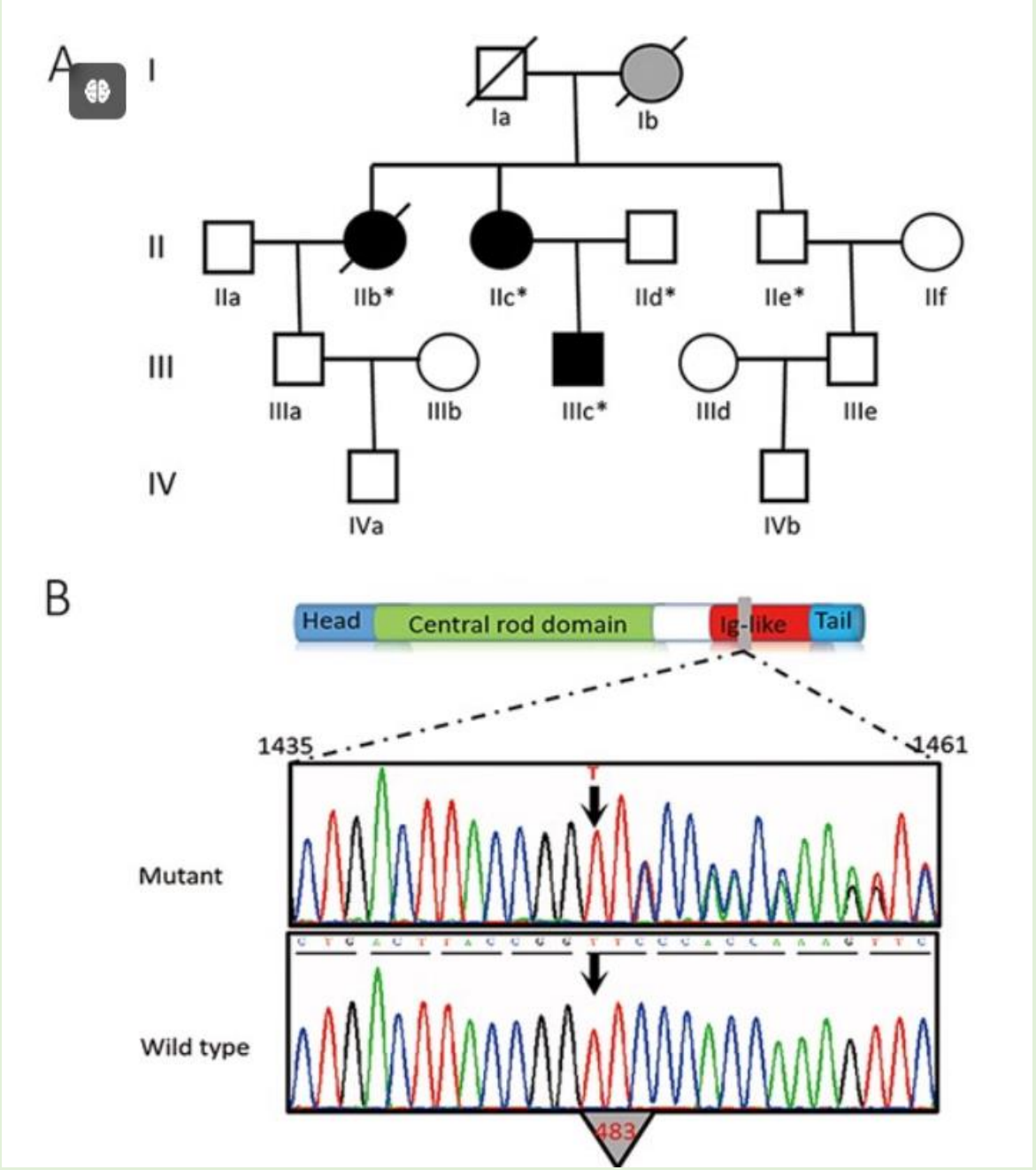
Brugada syndrome



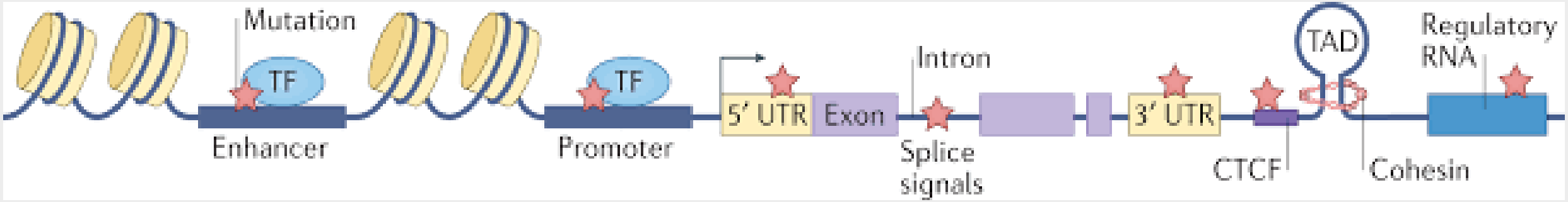
The W156X mutation in the SCN5A gene

Example of Frameshift mutation

LMNA frameshift mutation, p.P485Tfs*67, from a patient with early-onset atrial disease.

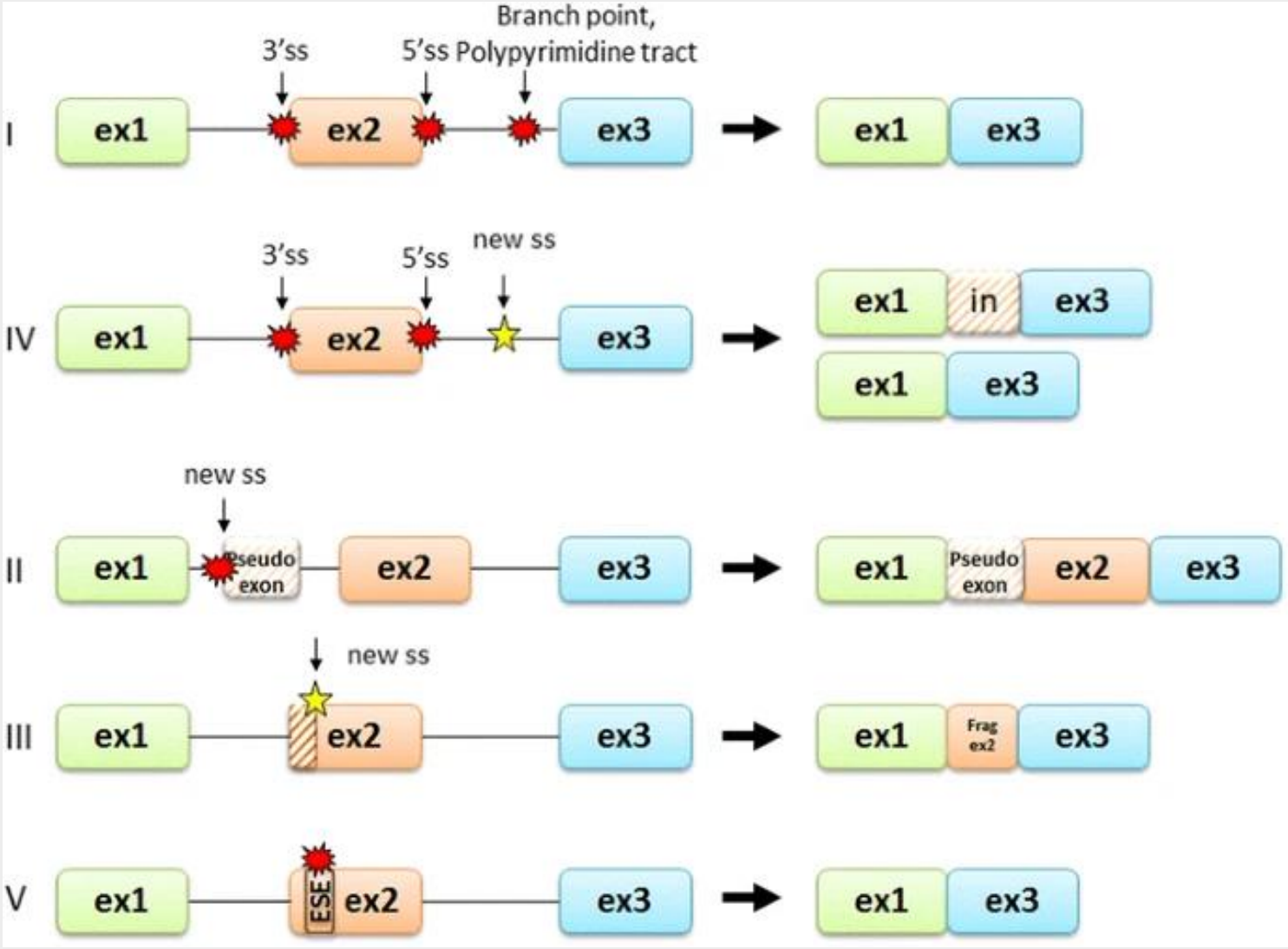


Mutations in Non-coding DNA



Splicing Mutations

Mutations of the highly conserved splice donor (GT) and splice acceptor (AG) sites usually result in aberrant splicing



Functional Effects of Mutations on the Protein

Loss-of-Function Mutations

Loss-of-function mutations can result in either reduced activity or complete loss of the gene product.

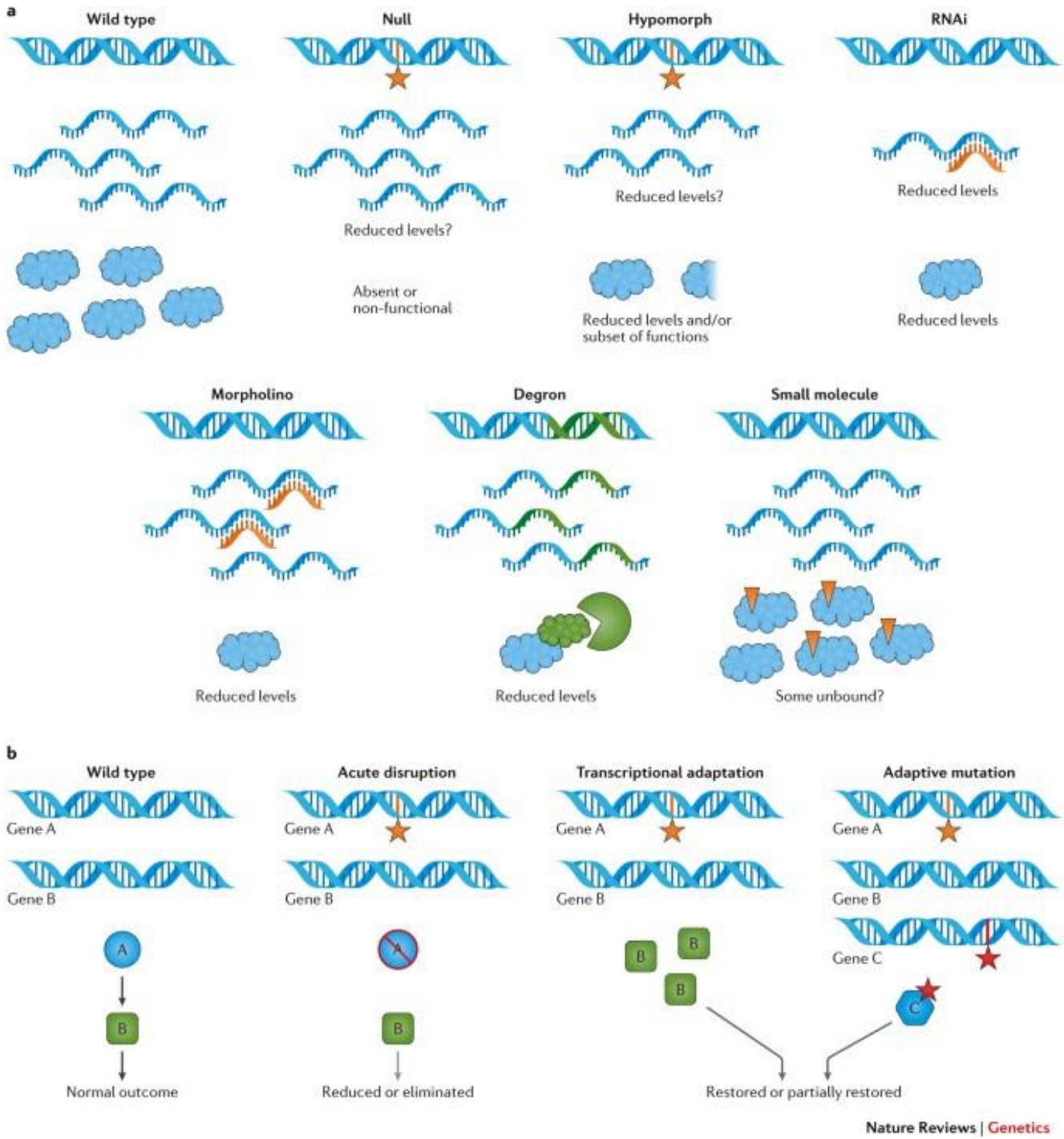
- Haploinsufficiency

Gain-of-Function Mutations

result increased levels of gene expression or the development of a new function(s) of the gene product.

Dominant-Negative Mutations

Dominant-negative mutations are particularly common in proteins that are dimers or multimers


a
Wild type
Null
Hypomorph
RNAi
Reduced levels?
Reduced levels?
Reduced levels
Absent or
non-functional
Reduced levels and/or
subset of functions
Reduced levels
Morpholino
Degron
Small molecule
Reduced levels
Reduced levels
Some unbound?
b
Wild type
Acute disruption
Transcriptional adaptation
Adaptive mutation
Gene A
Gene B
Gene A
Gene B
Gene A
Gene B
Gene A
Gene B
Gene C
A
B
A
B
B
B
B
C
Normal outcome
Reduced or eliminated
Restored or partially restored

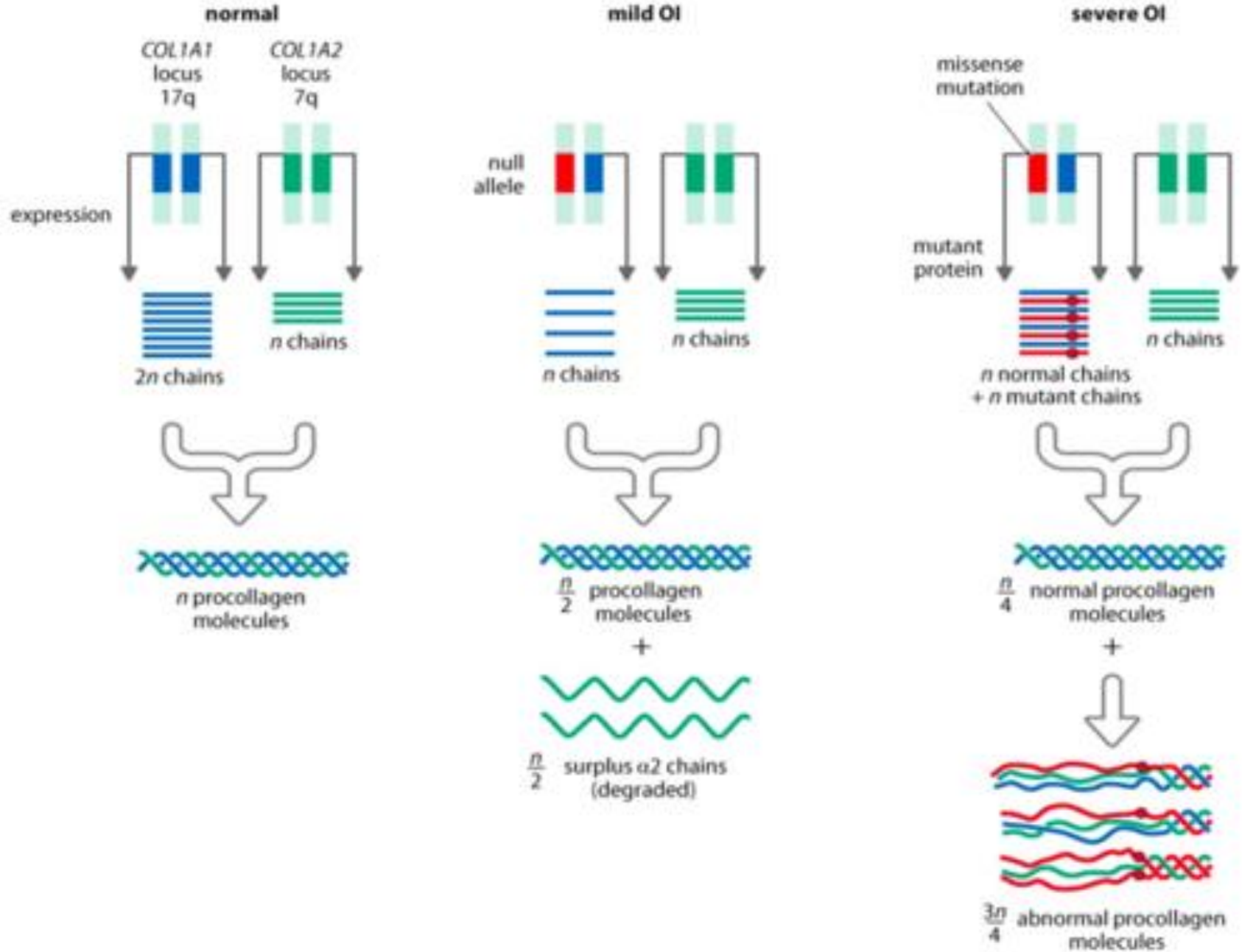


Figure 7.17b Genetics and Genomics in Medicine (© Garland Science 2015)

Mutations and Mutagenesis

Naturally occurring mutations are referred to as **spontaneous mutations** and are thought to arise through chance errors in chromosomal division or DNA replication.

Radiation

X-rays, γ-rays and neutrons have great penetrating power, but α particles can penetrate soft tissues to a depth of only a fraction of a millimeter, and β particles only up to a few millimeters.

Dosimetry

The gonad dose of radiation is often expressed as the amount received in 30 years.

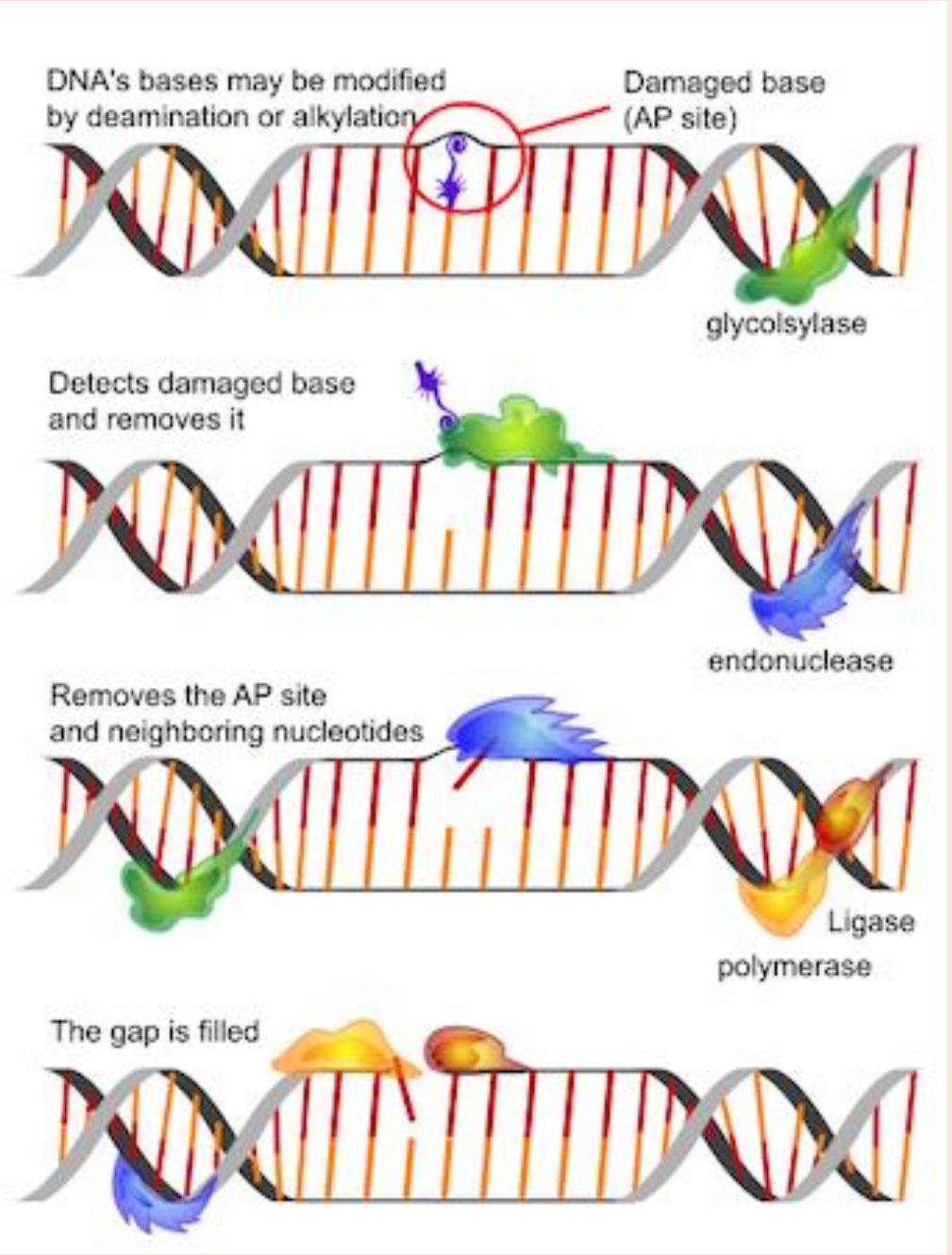
Table 2.6 Approximate average doses of ionising radiation from various sources to the gonads of the general population

Source of Radiation	Average Dose per Year (mSv)	Average Dose per 30 Years (mSv)
Natural		
Cosmic radiation	0.25	7.5
External γ radiation[a]	1.50	45.0
Internal γ radiation	0.30	9.0
Artificial		
Medical radiology	0.30	9.0
Radioactive fallout	0.01	0.3
Occupational and miscellaneous	0.04	1.2
Total	2.40	72.0

Chemical Mutagens

In humans, chemical mutagenesis may be **more important** than radiation in producing genetic damage.

Experiments have shown that certain chemicals, such as **mustard gas, formaldehyde, benzene, some basic dyes and food additives,** are mutagenic in animals. Exposure to environmental chemicals may result in the formation of DNA adducts, chromosome breaks or aneuploidy



DNA Repair

DNA mutations, if left unrepaired, would have serious consequences or both the individual and subsequent generations. The stability of DNA is dependent upon continuous DNA repair by a number of different mechanisms.

Table 2.7 DNA repair pathways, genes, and associated disorders

Type of DNA Repair	Mechanism	Genes	Disorders
Base excision repair (BER)	Removal of abnormal bases	*MYH*	Colorectal cancer
Nucleotide excision repair (NER)	Removal of thymine dimers and large chemical adducts	*XP*	Xeroderma pigmentosum
Postreplication repair	Removal of double-strand breaks by homologous recombination or non-homologous end-joining	*NBS*	Nijmegen breakage syndrome
		BLM	Bloom syndrome
		BRCA1/2	Breast cancer
Mismatch repair (MMR)	Corrects mismatched bases caused by mistakes in DNA replication	*MSH* and *MLH*	Colorectal cancer (HNPCC)

THANK YOU