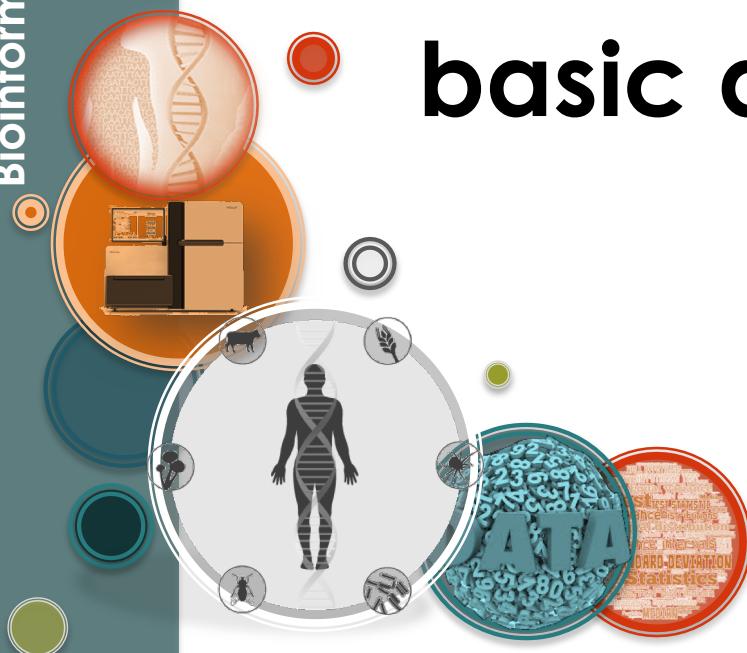
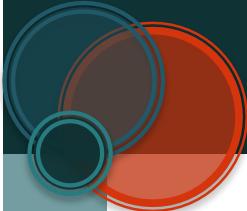


Biology: basic and novel key concepts





Overview

Part 1

Introduction

Part 2

The DNA molecule

Part 3

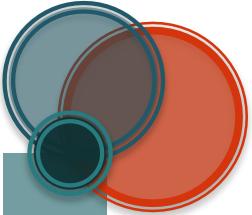
RNA: from central dogma to new functional properties

Part 4

Proteins

Part 5

Point mutations and their consequences



Part 1

Introduction

Biology: key actors

What are the key actors in biology?

DNA, RNA, Proteins

DNA



RNA



Protein



Code for the
hereditary
information

Central
dogma/ novel
multiple
functions

multiple
functions in a
cell

DNA, RNA, Proteins: the key to diversity

Why do we need knowledge about DNA, RNA & Proteins ?

Explain what determines the complexity of living organisms, etc...

Humans as an example:



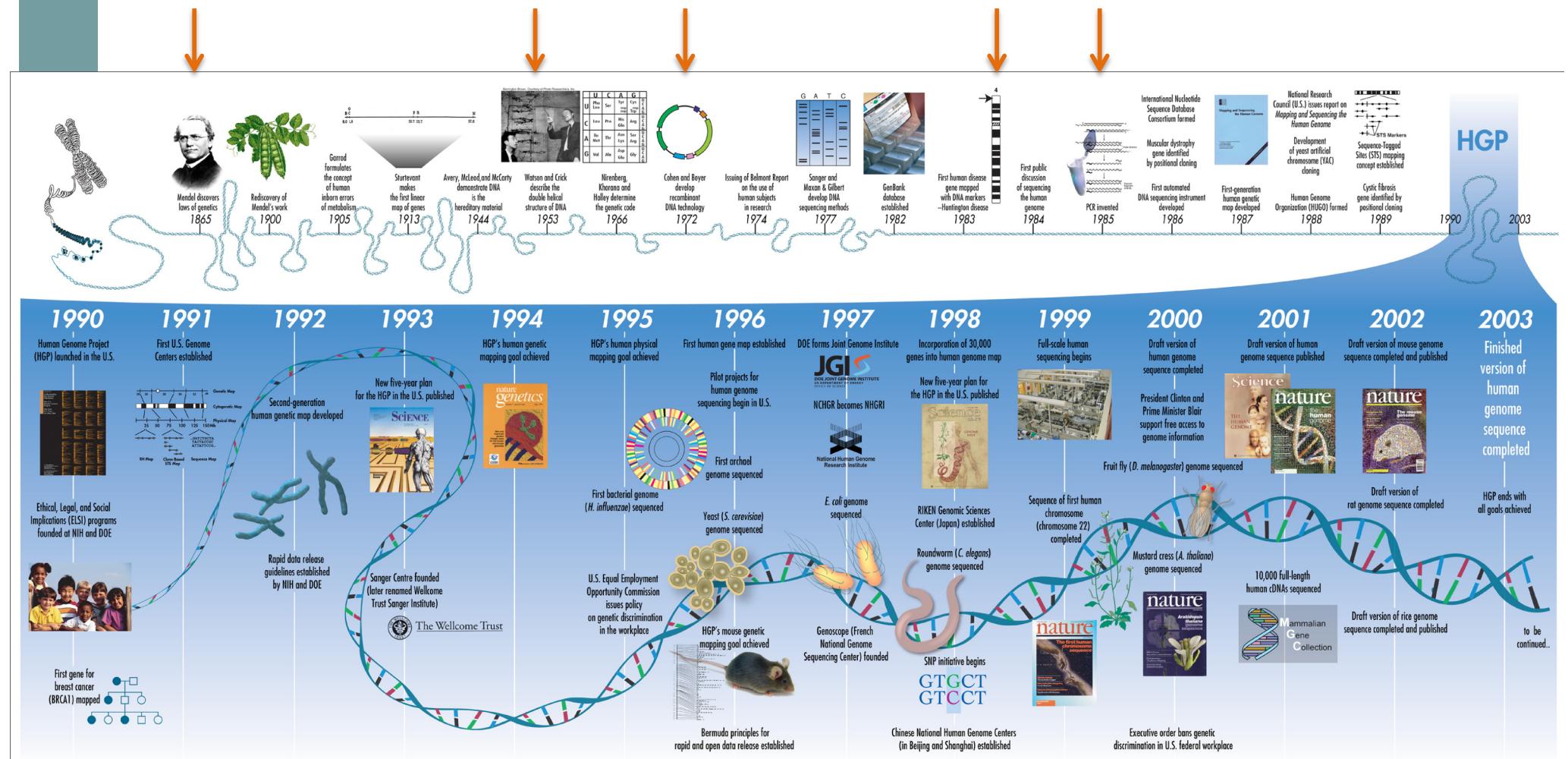
No two individuals are similar:

- What determines their different characteristics ?
- What determines their different susceptibility to diseases?

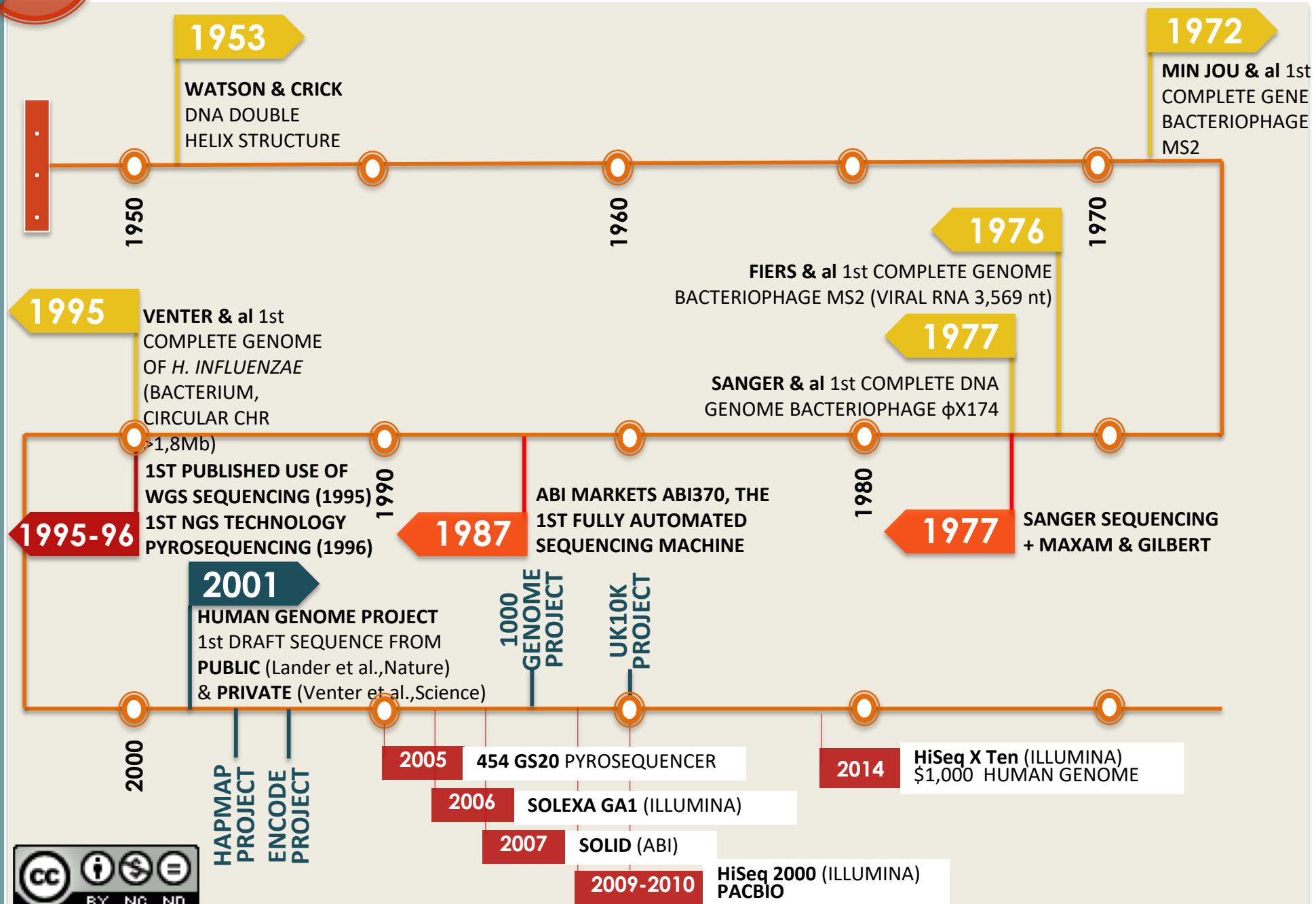
**DNA
RNA
Proteins**

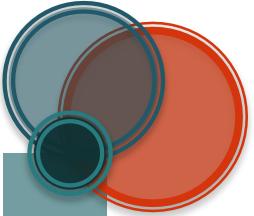
DNA knowledge: a constant evolution

- Discovery of genetics laws (Mendel)
- Discovery of the double helix structure of DNA
- Technologies: recombinant DNA, PCR
- Relationship between genes and diseases ...



DNA knowledge: a constant evolution





Part 2

The DNA molecule

DNA

The DNA molecule

DNA (or deoxyribonucleic acid) is the molecule that carries the genetic information in all cellular forms of life and some viruses.

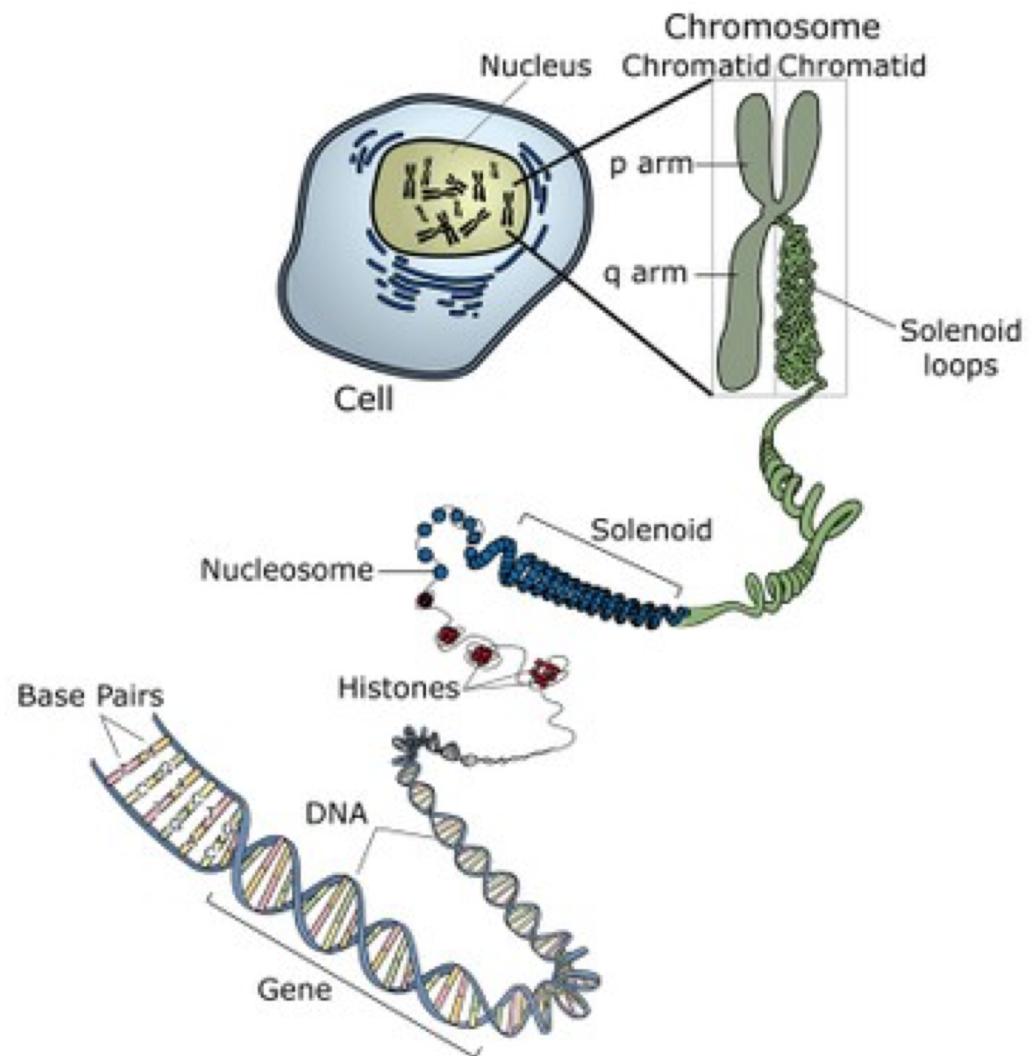


Image adapted from: National Human Genome Research Institute.

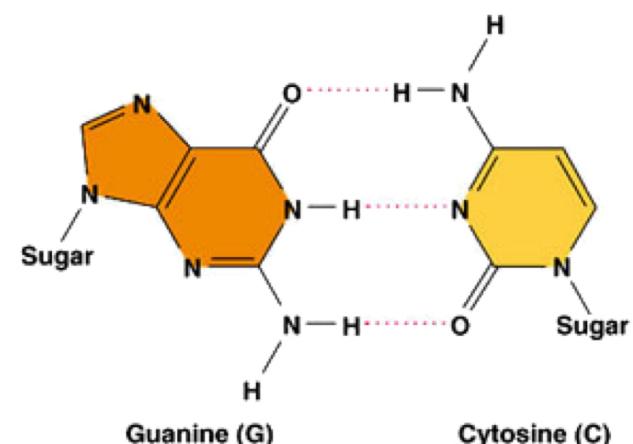
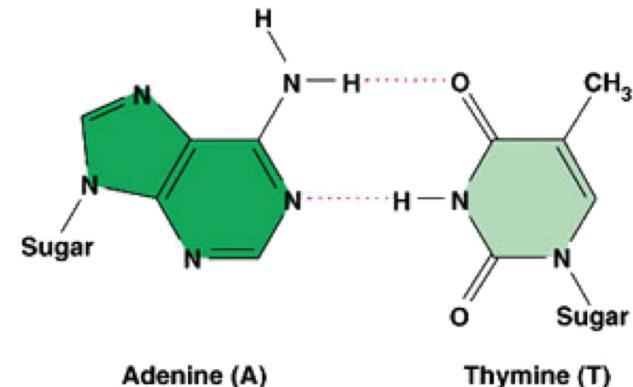
DNA

The DNA molecule

The DNA molecule belongs to a class of molecules called the **nucleic acids**, which are polynucleotides - that is, long chains of nucleotides.

Each nucleotide consists of three components:

- a nitrogenous base: cytosine (C), guanine (G), adenine (A) or thymine (T)
 - a five-carbon sugar molecule (deoxyribose in the case of DNA)
 - a phosphate molecule



DNA

The DNA molecule

In the DNA molecule, **the backbone** of the polynucleotide is **a chain of sugar and phosphate molecules**.

Each of the sugar groups in this sugar-phosphate backbone is **linked to one of the four nitrogenous bases**.

Directionality along the backbone : 5' (phosphate) → 3' (OH)

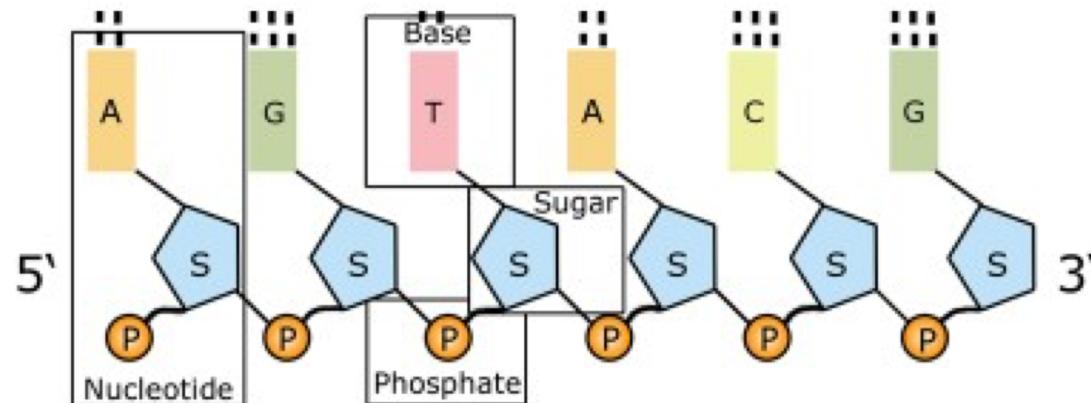


Image adapted from: National Human Genome Research Institute.

The DNA molecule

Base-pairing and strand interactions

- A, G are long (double ring purines)
- C,T are short (single ring pyrimidines)
- need one long and one short nucleotide per pair
- C-G have three hydrogen bonds (slightly stronger matching)
- A-T have two hydrogen bonds (slightly weaker matching)

The DNA molecule

Complementarity in base-pairing : The bases link across the two strands in a specific manner using hydrogen bonds: cytosine (C) pairs with guanine (G), and adenine (A) pairs with thymine (T).

Purines: A, G

Pyrimidines: T, C (U instead of T in RNA)

A different number of hydrogen bonds links the bases

Antiparallel orientation:

1 strand is 5' → 3'

1 strand is 3' → 5'

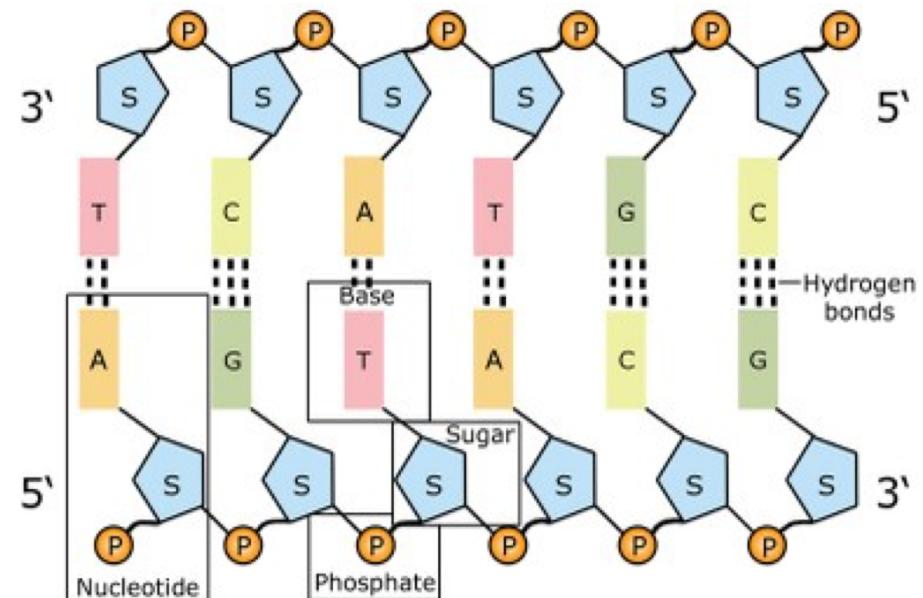
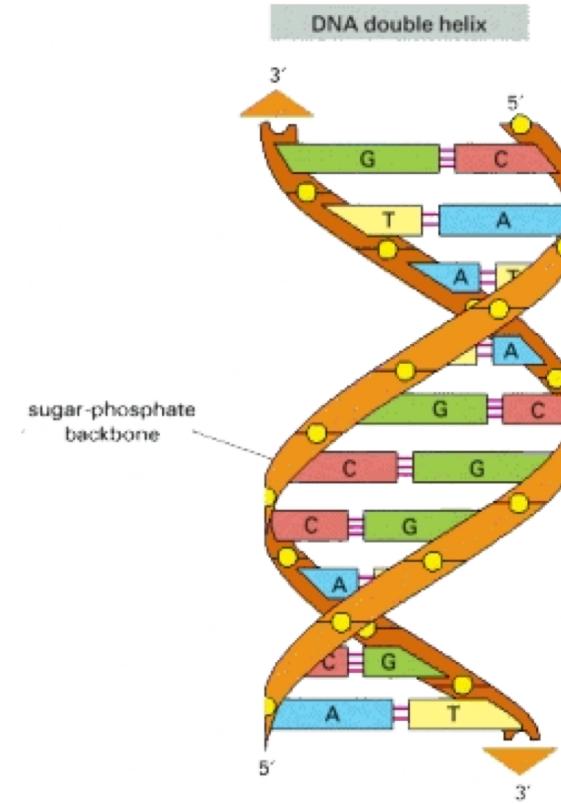


Image adapted from: National Human Genome Research Institute.

The DNA molecule

Helical shape

- 10.4 nucleotides per turn
- diameter = 2 nm
- both major and minor grooves
- called B-DNA.
- The helix twist and diameter can also change under dehydrating conditions and methylation to A-DNA and Z-DNA

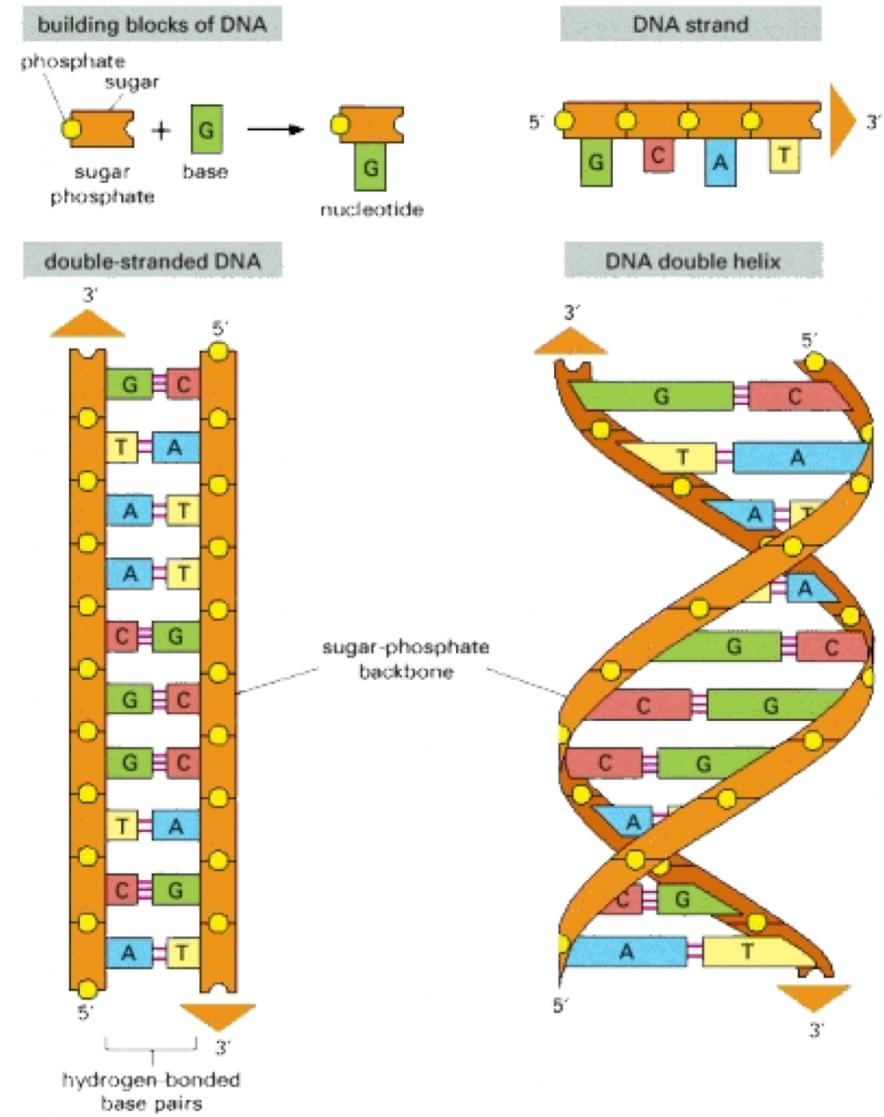


DNA

The DNA molecule

In brief:

- The **nucleotide** (P-sugar+base) is the minimal building block of DNA.
- Hydrogen bonding modes presented here follow the classical **Watson-Crick base pairing**
- The DNA is **double-stranded** and organized in a **double-helix**
- It is the **order of the bases** along a single strand that constitutes the genetic code.



The DNA molecule organization

- Information in DNA sequence is the **genome**
- The chromosomes of prokaryotic cells are not enclosed by a separate membrane. The chromosome - together with ribosomes and proteins associated with gene expression - is located in a region of the cell cytoplasm known as the **nucleoid**.
- In bacteria, genome usually **circular**. Most bacteria contain a single, circular chromosome. (examples of exceptions: Streptomyces possess linear chromosomes, and Vibrio cholerae has two circular chromosomes)
- The genomes of prokaryotes are compact compared with those of eukaryotes, as they lack introns, and the genes tend to be expressed in groups known as **operons**.
- In addition to the main chromosome, bacteria are also characterised by the presence of extra-chromosomal genetic elements called **plasmids**.

The DNA molecule organization

- The genome in eukaryotes is organized into **chromosomes**
 - each chromosome is a separate DNA molecule
 - Chromosomes extend and replicate during interphase portion of the cell cycle?
 - chromosomes are condensed, visible with light during cell division (M phase)
 - Each chromosome has a **p arm** and a **q arm**. In their replicated form, each chromosome consists of two **chromatids**.
 - Specific sequences have particular roles in chromosomes:
 - **Centromers** (*center “pinch point” of a chromosome that allows one copy of each to be pulled apart into two daughter cells during division*)
 - **Telomeres** (*specialized sequences at the chromosomes end that facilitate replication*)
 - **Replication origins** (*multiple locations where the replication machinery first binds*)

The DNA molecule organization

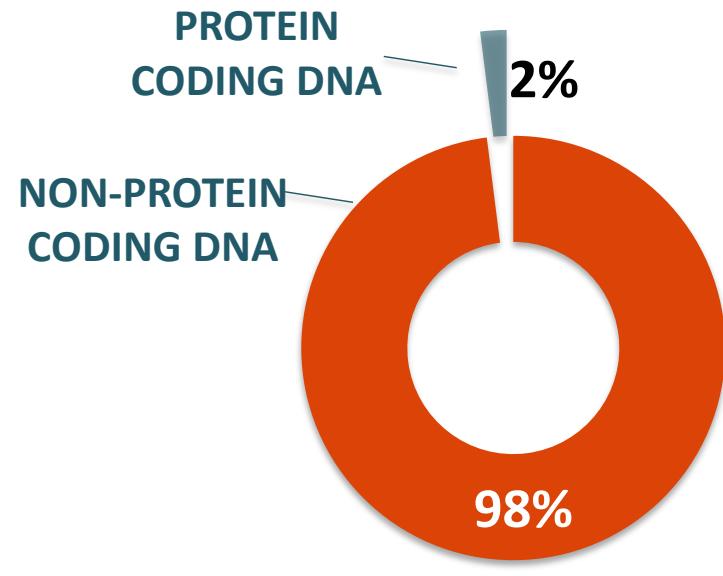
- The **genome** contains **coding** and **non-coding** regions
- The complexity of non-coding regions is higher in higher eukaryotes
- The coding regions contains genes (about 20,000 in the human genome)
- Genes can code for RNAs or for proteins as a final product
- Central dogma in molecular biology:

Remember the central dogma of biology:

DNA (genetic information in genes) → RNA (copies of genes) → proteins
(functional molecules)

Completely changed after the HGP !!!

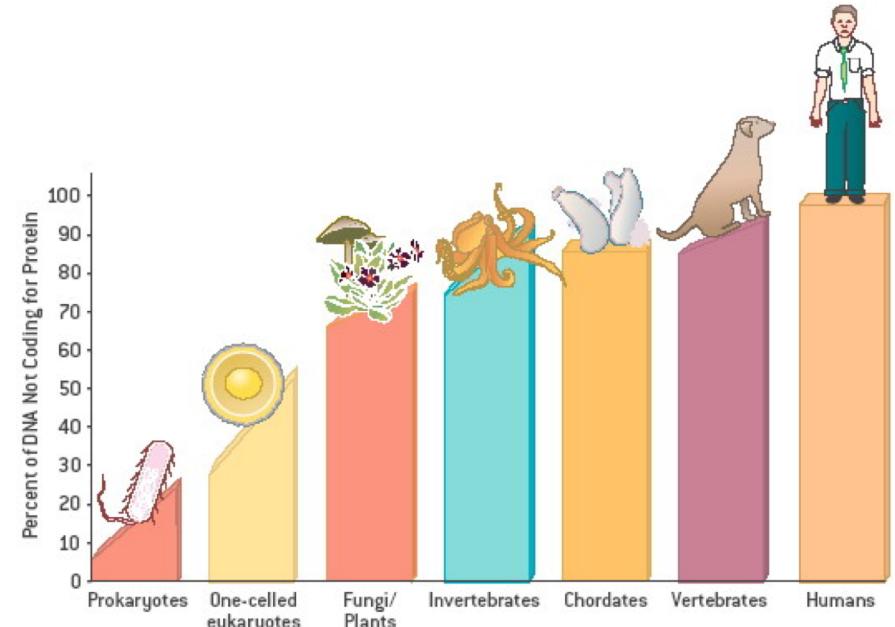
Coding vs non-coding



$\approx 20,000$

The human genome contains only about 20,000 protein-coding genes : sequence alone is not enough to explain the whole complexity !

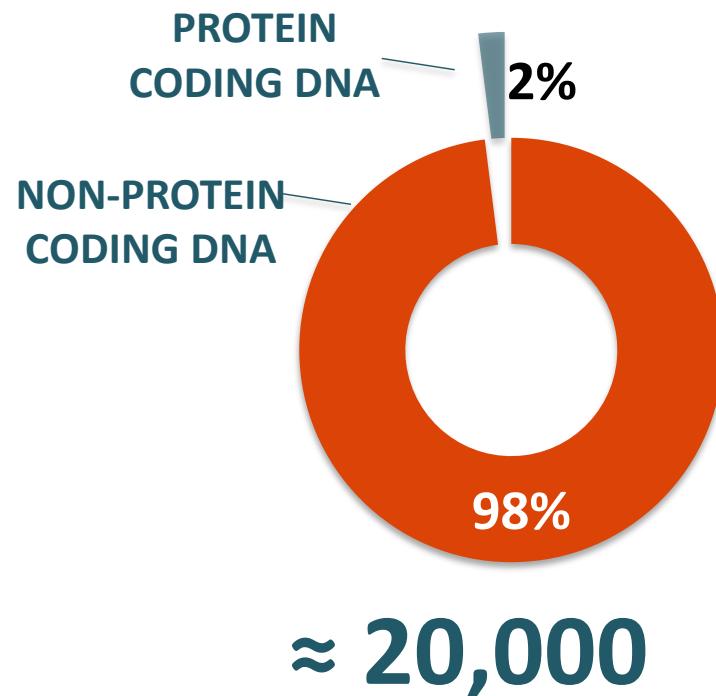
The proportion of non-coding DNA increases with organism complexity



(Mattick, 2011)

<http://www.yourgenome.org/stories/how-is-the-completed-human-genome-sequence-being-used>

Coding vs non-coding



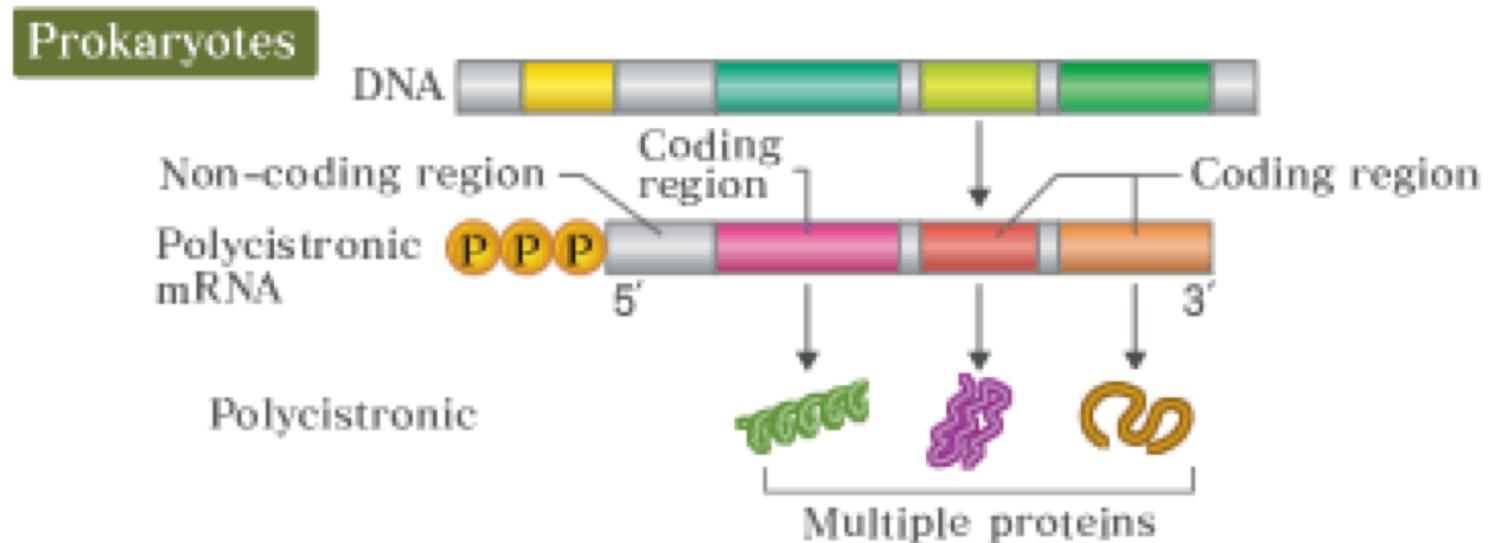
The human genome contains only about 20,000 protein-coding genes : sequence alone is not enough to explain the whole complexity !

Genes

- Basic physical and functional unit of heredity.
- Genes consist of three types of nucleotide sequence:
 - coding regions, called exons, which specify a sequence of amino acids
 - non-coding regions, called introns, which do not specify amino acids
 - regulatory sequences: determine when and where the protein is made + how much

DNA

Genes



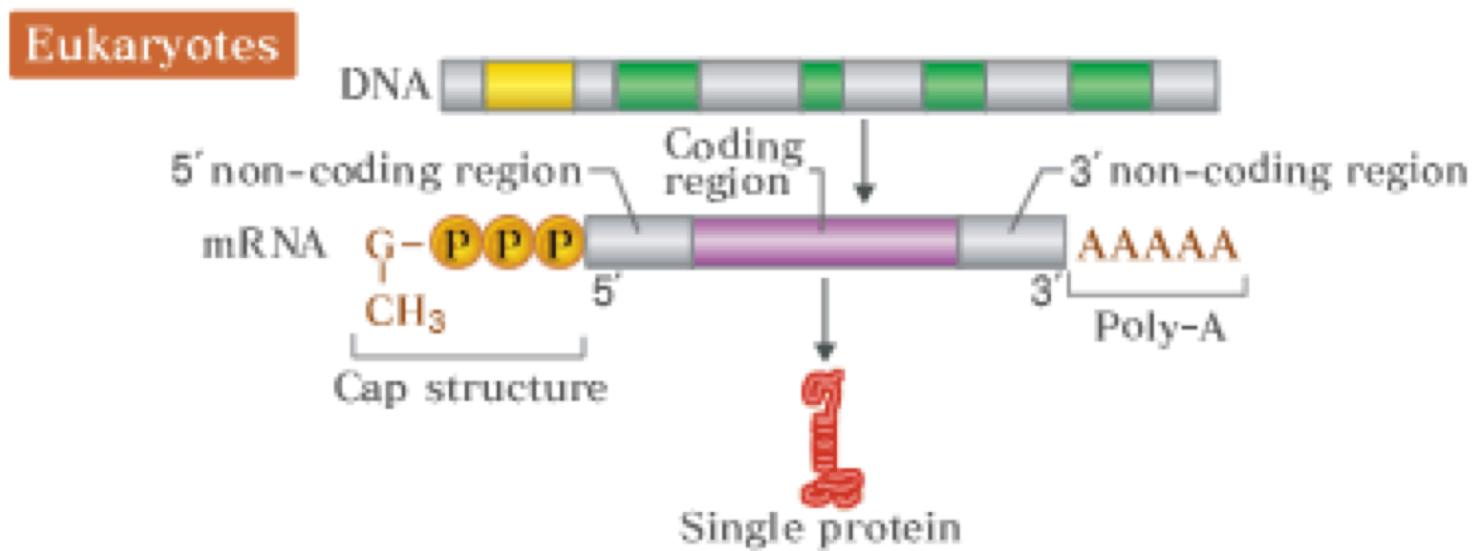
Prokaryotes

"cistron" = gene

"operon" = a gene unit controlled for gene expression by the same regulating region
(e.g. Lactose, Histidine : if present in the medium, the prokaryotic cell suppress all genes related to its synthesis by the genes in operon).

DNA

Genes



Eukaryotes

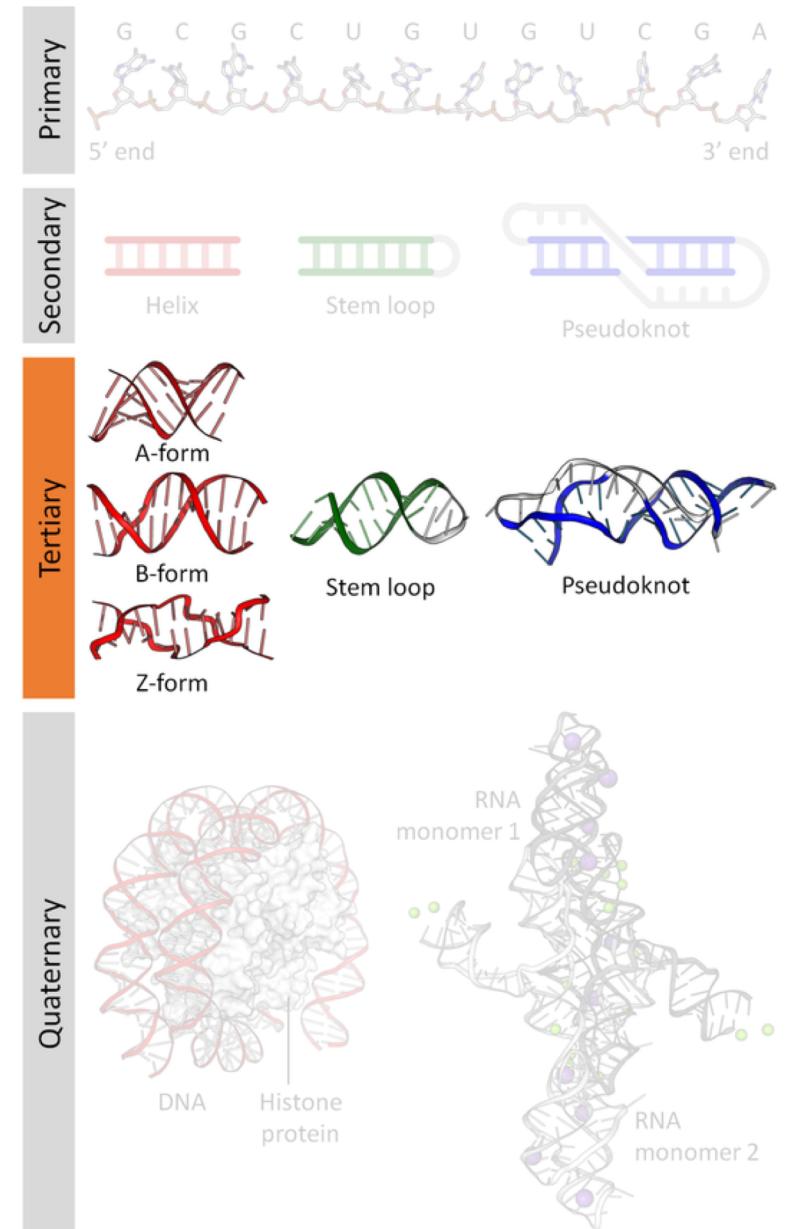
In general: Eukaryotes do not have operons and hence no polycistronic RNA is generated

→ Exceptions !!!! (*Leishmania*)

DNA

The DNA molecule regulation

- Primary, Secondary, Tertiary and Quaternary structures
- The **genome** can contain **methylated regions**
- **Organization in the nucleus**



The DNA molecule is in an active reshaping

The information contained in the DNA as a specific occurrence of sequences or a genetic code is frequently referred to as a "blueprint" because it contains the instructions a cell requires in order to sustain itself.

Evidence demonstrate that informations stored at the DNA level is the basis for the production of various molecules, including DNA itself, RNA and protein.

- DNA replication
- DNA transcription
- DNA translation

The DNA molecule is in an active reshaping

DNA replication

DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule. This process occurs in **all living organisms** and is the basis for **biological inheritance**.

The enzyme **helicase** breaks the hydrogen bonds holding the two strands together, and both strands can then act as templates for the production of the opposite strand. The process is catalysed by the enzyme **DNA polymerase**, and includes a proofreading mechanism.

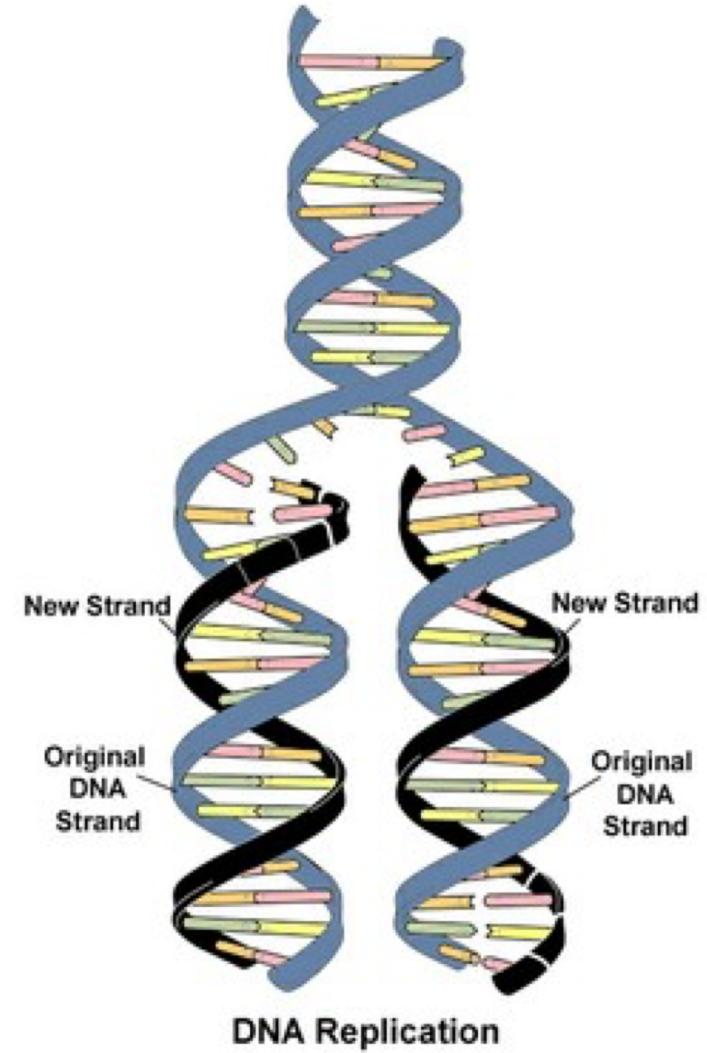


Image adapted from: National Human Genome Research Institute.

The DNA molecule is in an active reshaping

DNA replication

Error-correction machinery

- mutations occur 1 in 10^{10} nucleotides copied → evolution, cancer
- DNA polymerase **proofreads** to make sure correct nucleotide is added → if not, it excises and goes back to add the correct one
- **Mismatch repair** machinery fixes incorrectly added nucleotides not found by DNA polymerase → detects nicks in newly created strand

Damage to DNA continuously occurs

- Homologous recombination uses similar sequences in nearby strands in order to fill in excised damaged DNA
- also the basis of heredity

The DNA molecule is in an active reshaping

DNA replication

Eukaryotes

- Within eukaryotes, DNA replication is controlled within the context of the cell cycle and occurs at the **S phase** (synthetic phase).
- Several checkpoints (cyclins) control the progression of the cell through its cell cycle. The G₁/S checkpoints controls entry in replication, or remaining in the G₀ stage.
- Replication of chloroplast and mitochondrial genomes occurs independently of the cell cycle, through the process of D-loop replication.

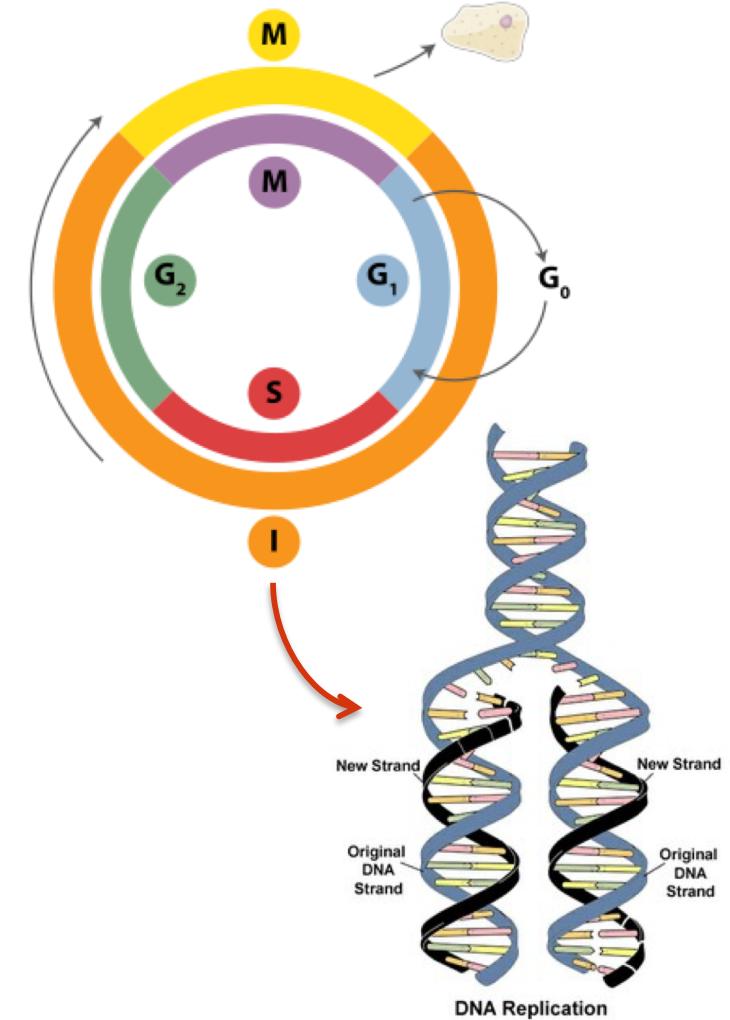


Image adapted from: National Human Genome Research Institute.

The DNA molecule is in an active reshaping

DNA replication

Eukaryotes

- In vertebrate cells, replication sites concentrate into positions called **replication foci**.
- Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome.

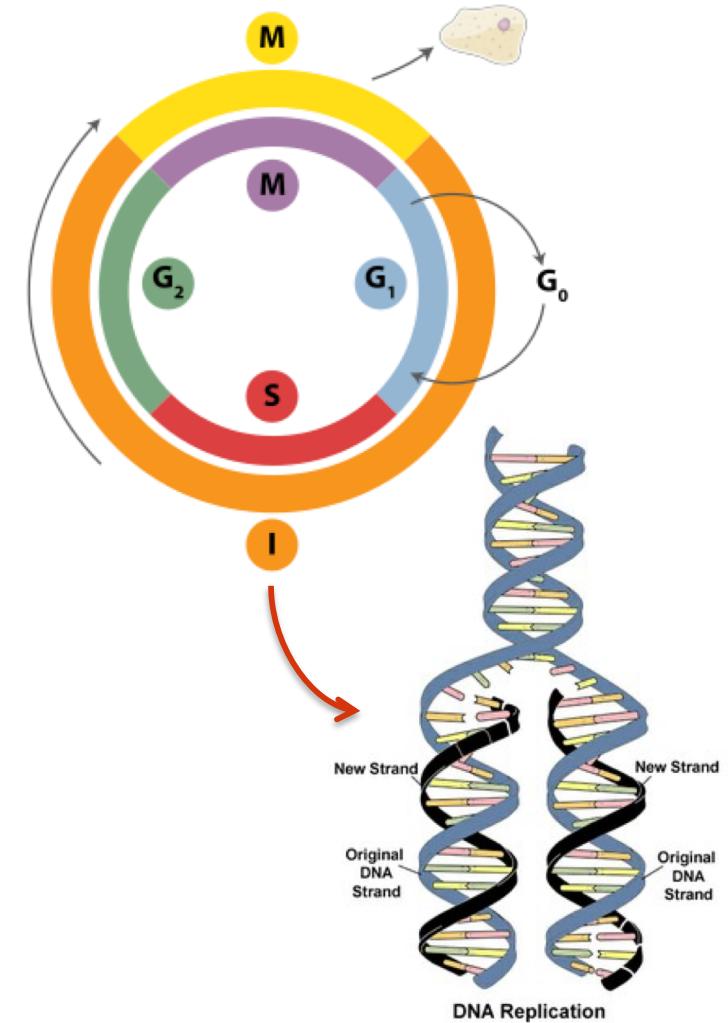


Image adapted from: National Human Genome Research Institute.

The DNA molecule is in an active reshaping

DNA replication

Eukaryotes

- Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes, but ends at the telomere region of repetitive DNA close to the ends.
- This shortens the telomere of the daughter DNA strand. Shortening of the telomeres is a normal process in **somatic cells**. As a result, cells can only divide a certain number of times before the DNA loss prevents further division (known as the Hayflick limit).

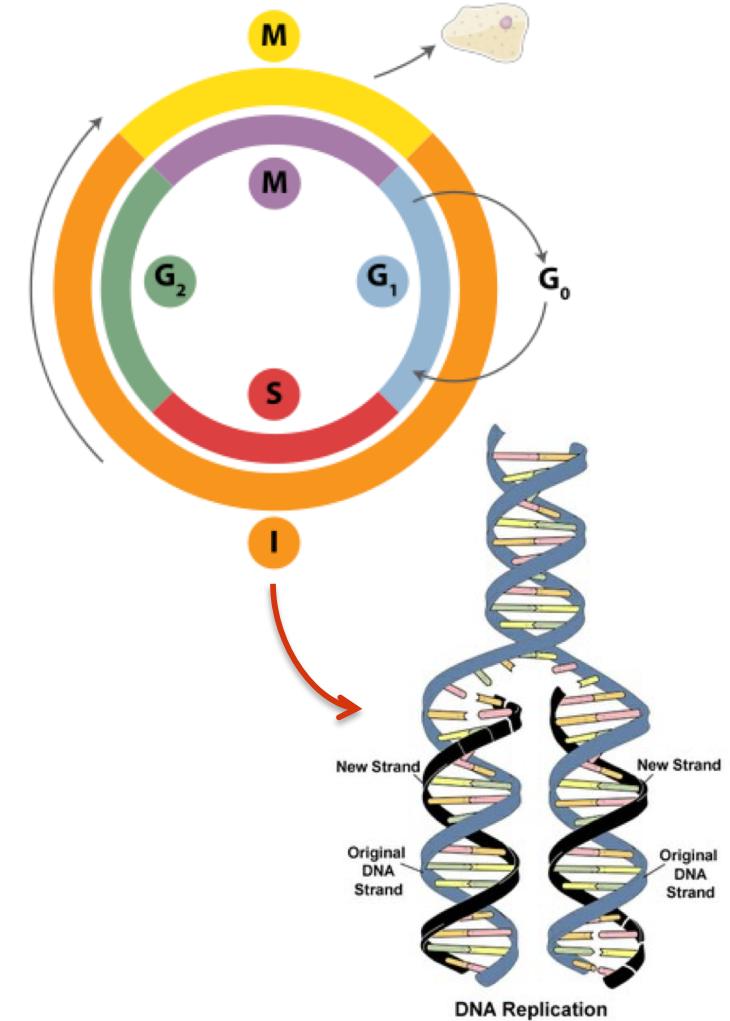


Image adapted from: National Human Genome Research Institute.

The DNA molecule is in an active reshaping

DNA replication

Eukaryotes

- Within the **germ cell line**, which passes DNA to the next generation, **telomerase** extends the repetitive sequences of the telomere region to prevent degradation.
- **Telomerase** can become mistakenly active in **somatic cells**, sometimes leading to cancer formation. Increased telomerase activity is one of the hallmarks of cancer.

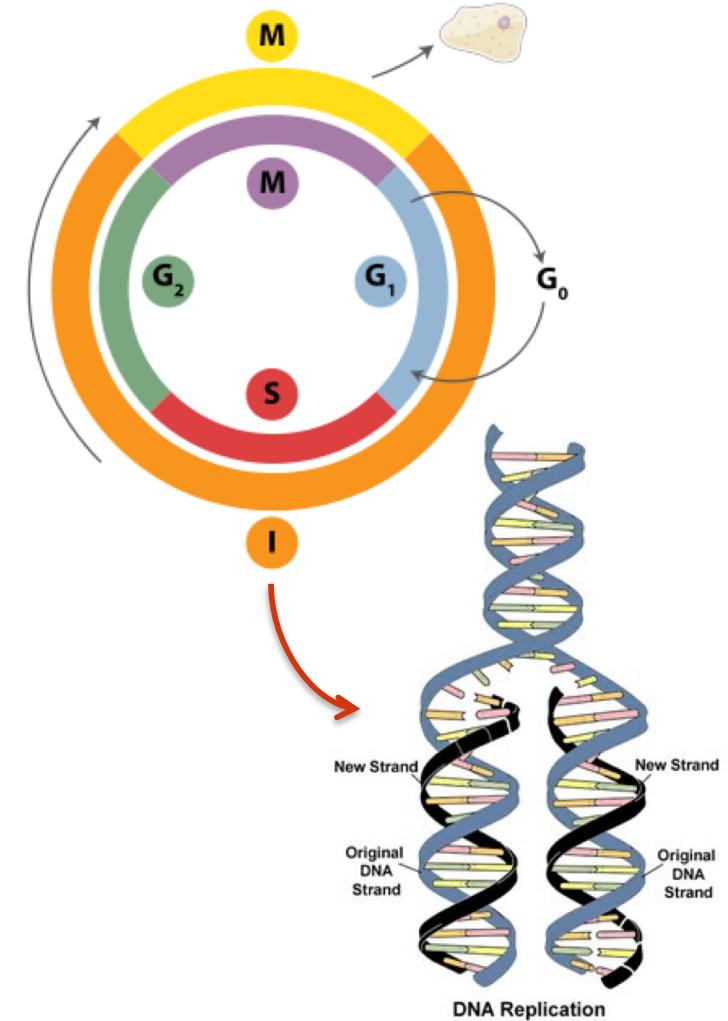


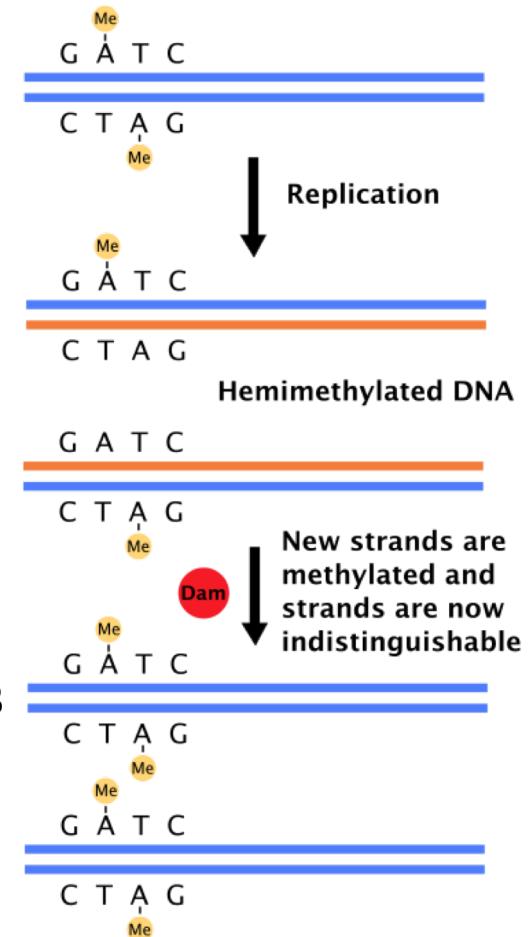
Image adapted from: National Human Genome Research Institute.

The DNA molecule is in an active reshaping

DNA replication

Prokaryotes

- Most bacteria do not go through a well-defined cell cycle but instead continuously copy their DNA → during rapid growth : concurrent occurrence of multiple rounds of replication.
- In *E. coli*, DNA replication is regulated through several mechanisms, including:
 - the **hemimethylation** (*E. coli* methylates GATC) and **sequestering of the origin sequence**, (SeqA recognizes hemiCH3 and binds to this origin)
 - the levels of protein **DnaA**. (DNAa required for replication initiation poorly recognizes hemiCH3 and prevents immediate reinitiation).



The DNA molecule is in an active reshaping

DNA transcription

In transcription, a portion of the double-stranded DNA template gives rise to a single-stranded RNA molecule.

In some cases, the RNA molecule itself is a "finished product" that serves some important function within the cell.

Transcription involves the synthesis of rRNA from DNA using RNA polymerase

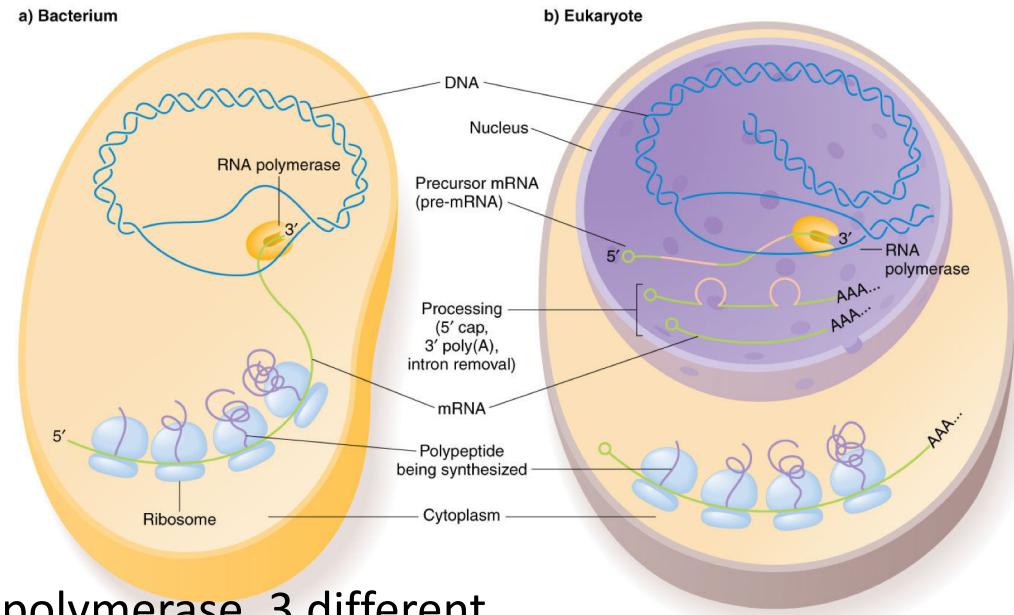
RNA polymerase must unpair and unwind DNA as it is reading it

- much less accurate than replication → errors of 1 in 10^{14}
- protein synthesis can tolerate more errors
- multiple RNAs can be sequenced from the same gene at the same time

DNA

The DNA molecule is in an active reshaping

DNA transcription



Whereas bacteria have only one RNA polymerase, 3 different types of RNA polymerase exist in eukaryotic cells.

- **RNA pol I** → genes that encode most of the rRNAs
- **RNA pol II** → mRNAs, which serve as the templates for production of protein molecules.
- **RNA pol III** → genes for one small rRNA, plus the tRNAs that play a key role in the translation process, as well as other small regulatory RNA molecules.

The DNA molecule is in an active reshaping

DNA transcription

Prokaryotes

- **RNA polymerase** binds to specific regions of the DNA called promoters, specific nucleotide sequences
- **Promoters** orient polymerase in a specific direction
- RNA polymerase binds to the promoter with the help of an accessory protein, called a **sigma factor**
- Synthesis stops at a **terminator sequence**, typically of poly A-T stretches of DNA

The DNA molecule is in an active reshaping

DNA transcription

Eukaryotes

- Different kinds of RNA polymerases
- RNA **polymerase** requires **transcription factors (TFs)** : helper proteins → TFs : assemble to form a large transcription complex (TFIID binds to TATA box...)
- **Chromatin-remodeling proteins** are involved to make DNA accessible from the wound histone structure
- **RNA chemically modified** with a methylated guanine at the 5' end and a poly-A sequence at the 3' end → necessary for translation
- RNA is processed after synthesis → **splicing** to remove noncoding regions called introns → **important : promotes diversity**

The DNA molecule is in an active reshaping

DNA transcription

RNA has some distinctions from DNA

- ribose rather than deoxyribose sugar (differs in an OH group)
- uracil instead of thymine (loss of a methyl group)
- **wobble base pairing** : pairing between two nucleotides in RNA molecules that does not follow Watson-Crick base pair rules : **(G-U), (I-U), (I-A), (I-C)** → fundamental in RNA secondary structure and critical for a proper translation
- single-stranded, and typically folds into unique shapes, like proteins
- less chemically stable

The DNA molecule is in an active reshaping

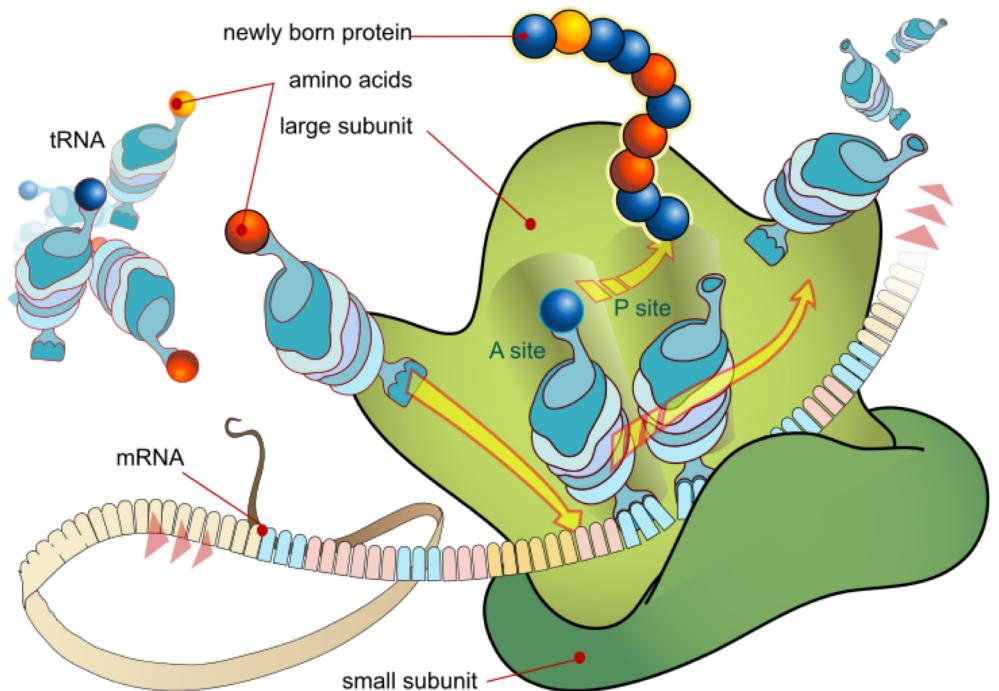
DNA translation

- **Translation** is the process in which ribosomes in the cytoplasm synthesize proteins after the transcription of DNA to RNA.
- In translation, mRNA is decoded by a ribosome, outside the nucleus, to produce a specific amino acid chain, or **polypeptide** → later folds into an active protein → performs its functions in the cell.

The DNA molecule is in an active reshaping

DNA translation

- Translation proceeds in 3 phases:
 - **Initiation:** ribosome assembles around the target mRNA + first tRNA (complementary tRNA anticodon) attached at the mRNA start codon (Met).
 - **Elongation:** tRNA transfers aa to the tRNA of next codon → ribosome *translocates* to the next mRNA codon to continue the process, creating a chain.
 - **Termination:** stop codon reached → the ribosome releases the polypeptide.



DNA

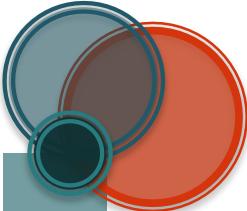
The DNA molecule is in an active reshaping

DNA translation

The genetic code is:

- **universal**: all living organisms use this code (few minor exceptions).
- **degenerate**. Each amino acid can be coded for by more than one codon. For example, AGA and AGG both code for the amino acid arginine.

		Second base of codon								
		U	C	A	G					
First base of codon	U	UUU Phenylalanine UUC phe UUA Leucine UUG leu	UCU Serine UCC ser UCA UCG	UAU Tyrosine UAC tyr UAA STOP codon UAG	UGU Cysteine UGC cys UGA STOP codon UGG Tryptophan					U C A G
	C	CUU Leucine CUC leu CUA CUG	CCU Proline CCC pro CCA CCG	CAU Histidine CAC his CAA Glutamine CAG gin	CGU Arginine CGC arg CGA CGG					U C A G
	A	AUU Isoleucine AUC ile AUA AUG Methionine met (start codon)	ACU Threonine ACC thr ACA ACG	AAU Asparagine AAC asn AAA AAG	AGU Serine AGC ser AGA Arginine AGG arg					U C A G
	G	GUU Valine GUC val GUA GUG	GCU Alanine GCC ala GCA GCG	GAU Aspartic acid GAC asp GAA Glutamic acid GAG glu	GGU Glycine GGC gly GGA GGG					U C A G
Third base of codon										



Part 3

RNA

from central dogma
to new functional properties

Central dogmas

Proposed by Francis Crick 1958

- DNA holds the coded hereditary information in the nucleus
- This code is expressed at the ribosome during protein synthesis in the cytoplasm
- The protein produced by the genetic information is what is influenced by natural selection
- If a protein is modified it cannot influence the gene that codes for it. Therefore only one way flow of information was thought to occur:

DNA → RNA → Protein

RNA

Central dogmas

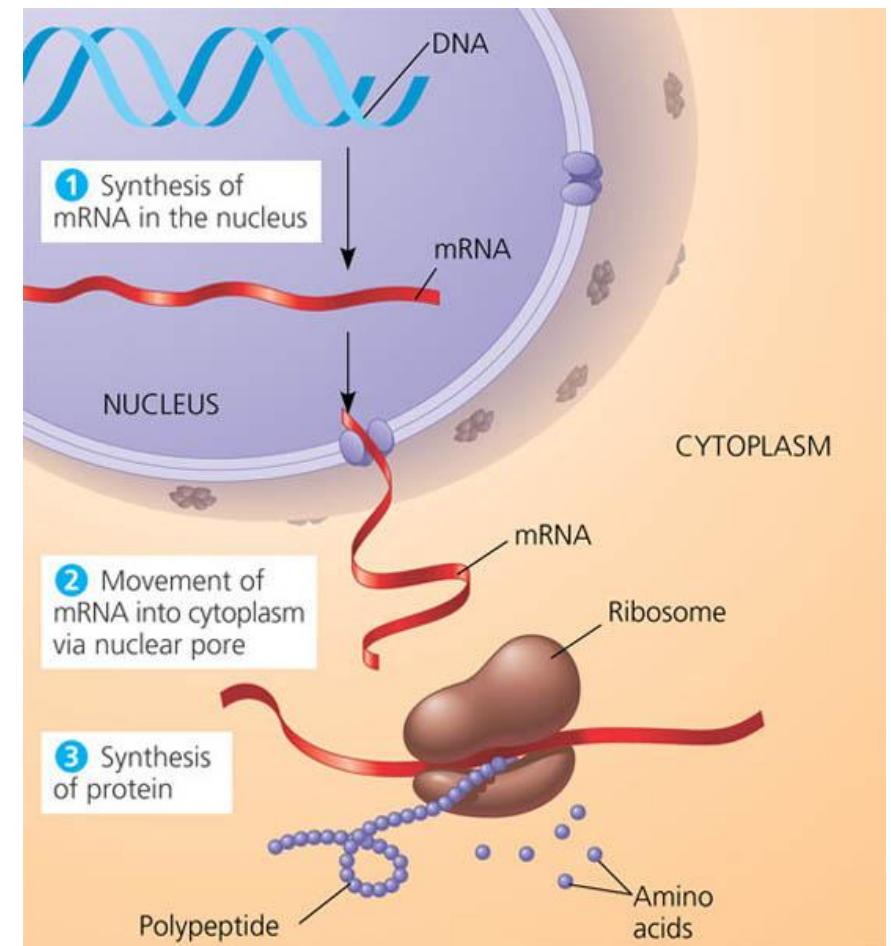
Proposed by Francis Crick 1958

- What is a gene? (definition until less than 10 years ago)

Gene

Is defined as a DNA portion whose corresponding mRNA encodes a protein.

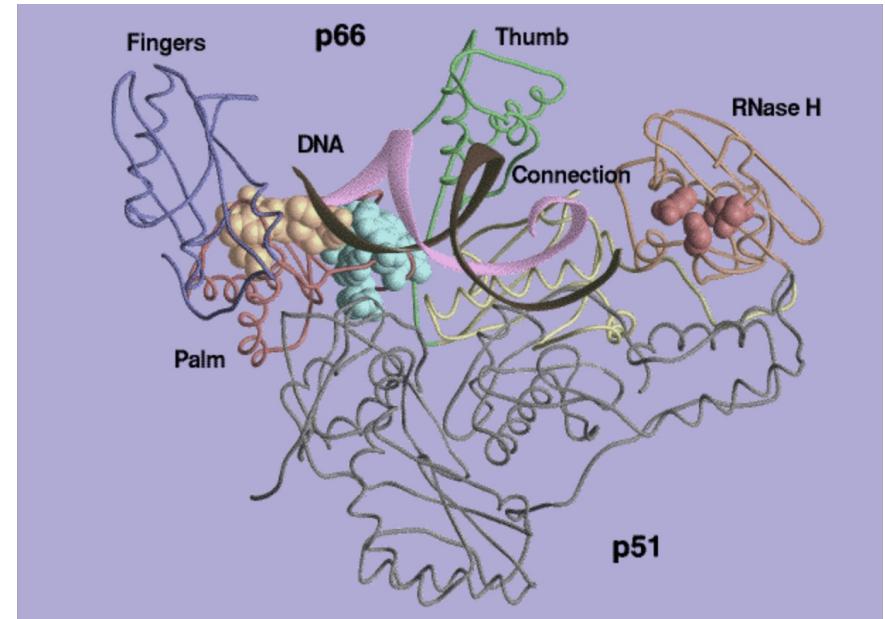
= RNA only considered to be a “**bridge**” in the transfer of biological information between DNA and proteins



Central dogmas

Proposed by Francis Crick 1958

- Retro viruses (e.g. HIV) carry RNA as their genetic information
- When they invade their host cell they convert their RNA into a DNA copy using reverse transcriptase
- Thus the central dogma has been modified:
DNA RNA→Protein
- Has helped to explain an important paradox in the evolution of life.



RNA

New era

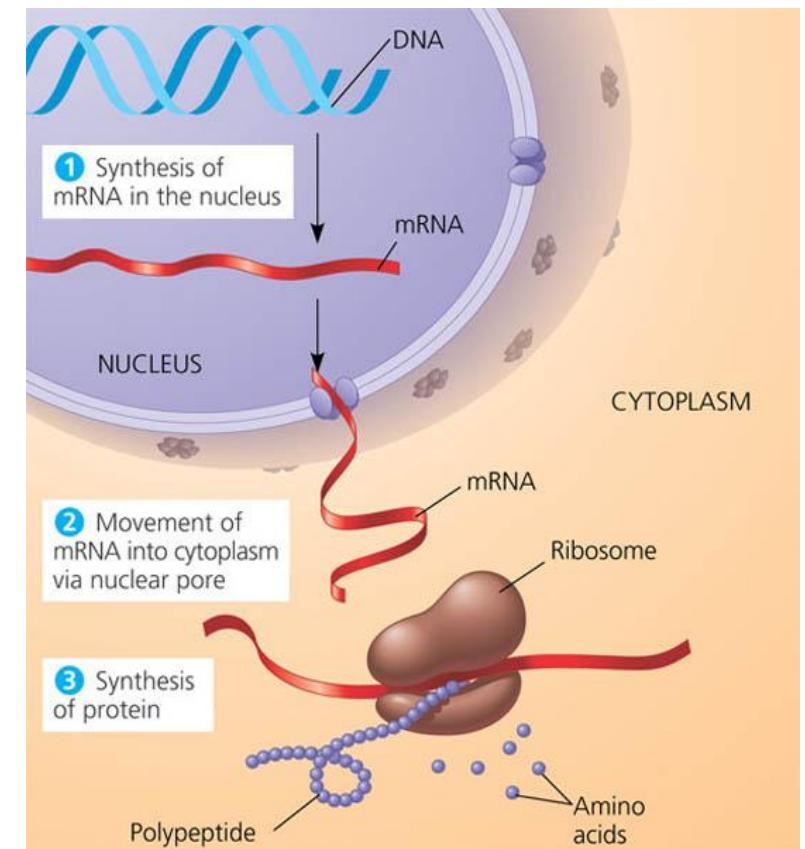
RNA is no longer considered to be a “bridge” in the transfer of biological information between DNA and proteins

Transcriptome consist of

- Coding RNA (mRNA (2–4%)
- Previously known non-coding RNAs
 - ribosomal RNA (80–90%, rRNA)
 - transfer RNA (5–15%, tRNA)
- Small fraction of of intragenic (i.e., intronic) and intergenic non-coding RNA (1%, ncRNA) with **undefined regulatory functions.**

Intragenic & Intergenic sequences

- Are enriched in repetitive elements
- have long been considered genetically inert
- mainly composed of “junk” DNA.



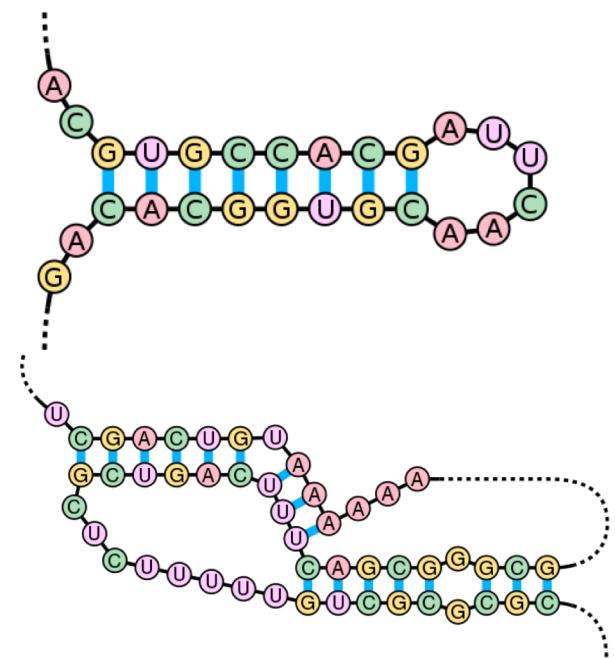
(Costa et al., 2010)

RNA

New era : understanding mechanisms of gene expression

For many RNA molecules, the secondary structure is highly important to the correct function of the RNA — often more so than the actual sequence.

- stem-loop secondary structure of tRNAs
- A **pseudoknot** is a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem.
- RNA secondary structure applies
→ in RNA splicing in certain species.
→ To aid in the analysis of non-coding RNA



New era : understanding mechanisms of gene expression

- Cells can adapt to environment and development by reconstructing their transcriptional networks to regulate diverse cellular processes without altering the underlying DNA sequences.
- Important alterations, namely **epigenetic changes**, occur during cell division, differentiation and cell death.
- Epigenetic changes are governed by various types of determinants, including:
 - DNA methylation patterns
 - histone posttranslational modification signatures
 - histone variants
 - chromatin remodeling
 - **chromosome conformation characteristics**
 - **non-coding RNAs (ncRNAs)**

New era : understanding mechanisms of gene expression

Chromosome conformation characteristics

- Chromosomal interactions can contribute to the silencing and/or activation of genes within the 3D organization of the nuclear architecture.
- The genome forms extensive and dynamic physical interactions in the form of chromatin loops and bridges.
 - brings distal elements of the chromosome into close physical proximity
 - potential consequences for gene expression and/or propagation of the genome.
- Thanks to advances in detection technology: it is now possible to examine these interactions at the molecular level.

New era : understanding mechanisms of gene expression

Chromosome conformation characteristics

- Protein complexes can co-occupy different promoters among different cell types
 - generate cell type-specific DNA loops
 - affect differential gene expression (Kagey et al., 2010)
- CTCF (binding factor) and cohesin, acting as chromatin looping proteins, are responsible primarily for constructing physical contacts, especially short-range loops, between promoters and enhancers in cell type-specific transcription (Tark-Dame et al., 2014).

trans-acting factors and cis-acting elements play vital roles in regulating nearby-gene expression and maintaining genome topology.

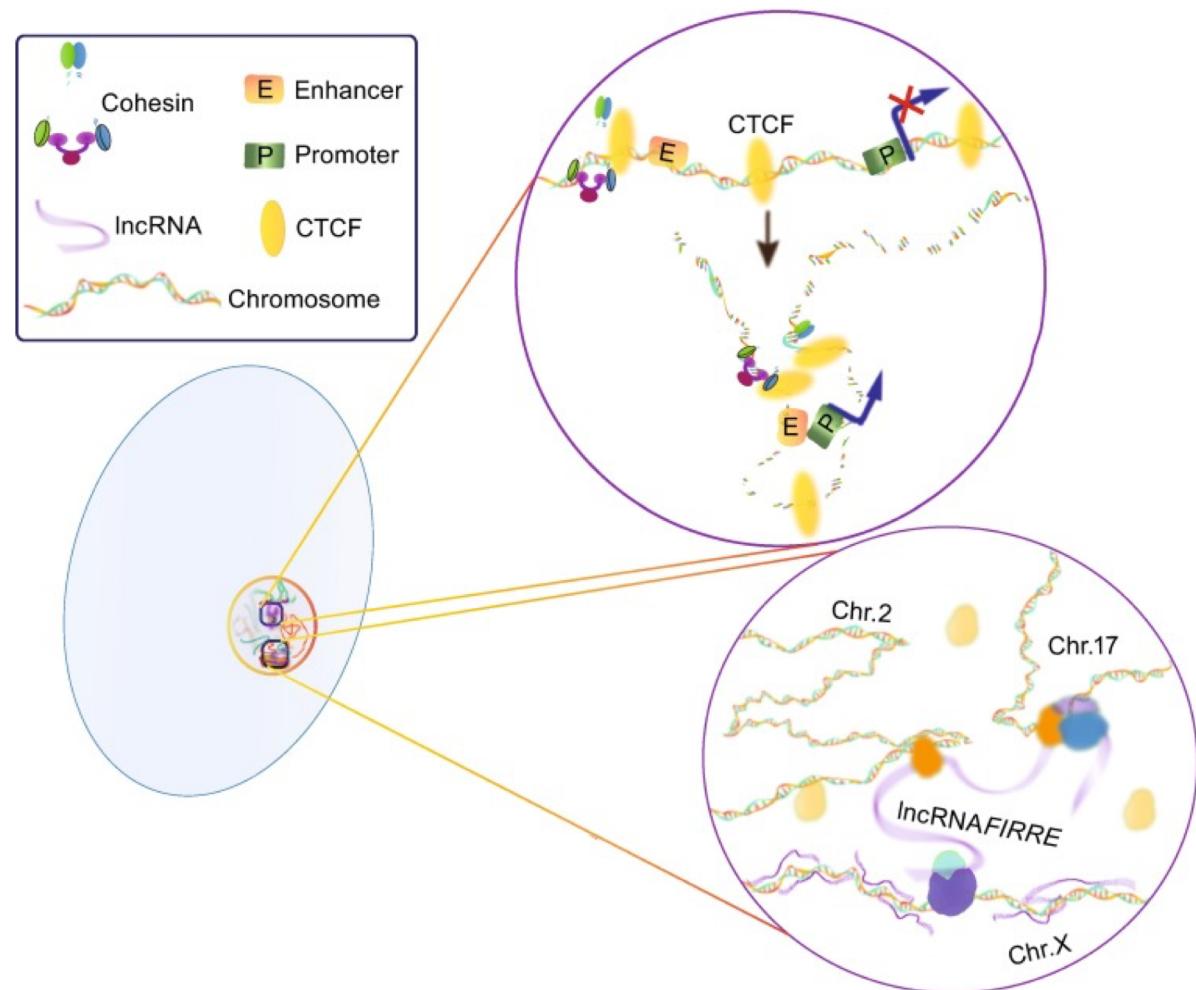
(Cao et al., 2015. Three-dimensional regulation of transcription)

New era : understanding mechanisms of gene expression

Complex genome topology

A simplified example of how CTCF/cohesion and lncRNAs respectively participate in construction of three-dimensional chromosome network.

(The lncRNA network was modified from Hacisuleyman et al, 2014)



(Cao et al., 2015. Three-dimensional regulation of transcription)

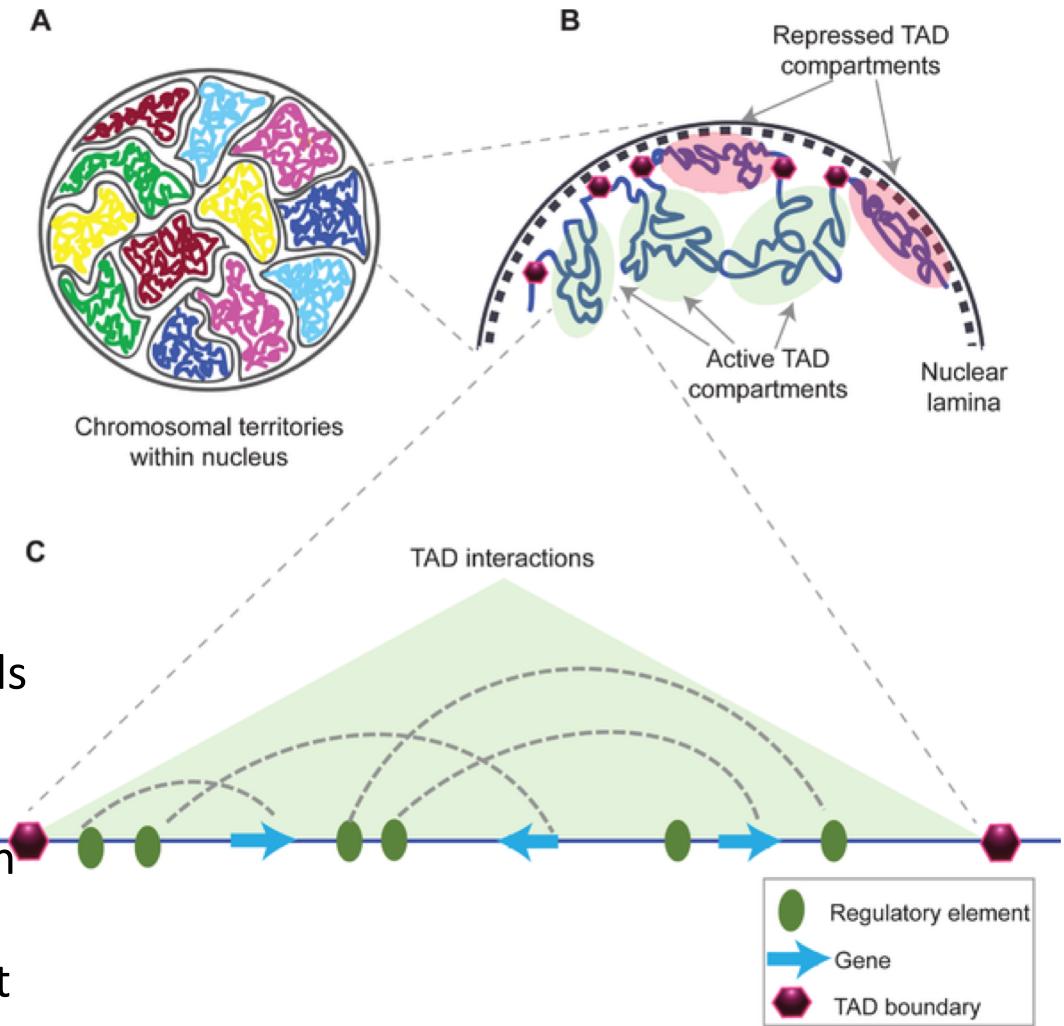
New era : understanding mechanisms of gene expression

Complex genome topology

A **topologically associating domain** (TAD) is a self-interacting genomic region = 3D chromosome structures (animals, plants, fungi, bacteria).

TADs can range in size from thousands to millions of DNA bases.

Functions not fully understood, but in some cases, **disrupting TADs** leads to disease because changes occurring at the 3D organization of the chromosome **disrupts gene regulation**.



New era : understanding mechanisms of gene expression

non-coding RNAs (ncRNAs)

- alternative splicing and posttranslational modifications make a significant contribution to the diversity and functionality of the proteome.
 - HGP and subsequent high-throughput genomic technologies have demonstrated that only $\approx 2\%$ of the human genome encodes for proteins.
- $\approx 98\%$ of genome exists as non-protein-coding sequence
- Important contribution of ncRNAs to developmental complexity and environmental adaptation (Mattick, 2011). It became clear that probably do not solely rely on protein-mediated three-dimensional transcription.

New era : understanding mechanisms of gene expression

non-coding RNAs (ncRNAs)

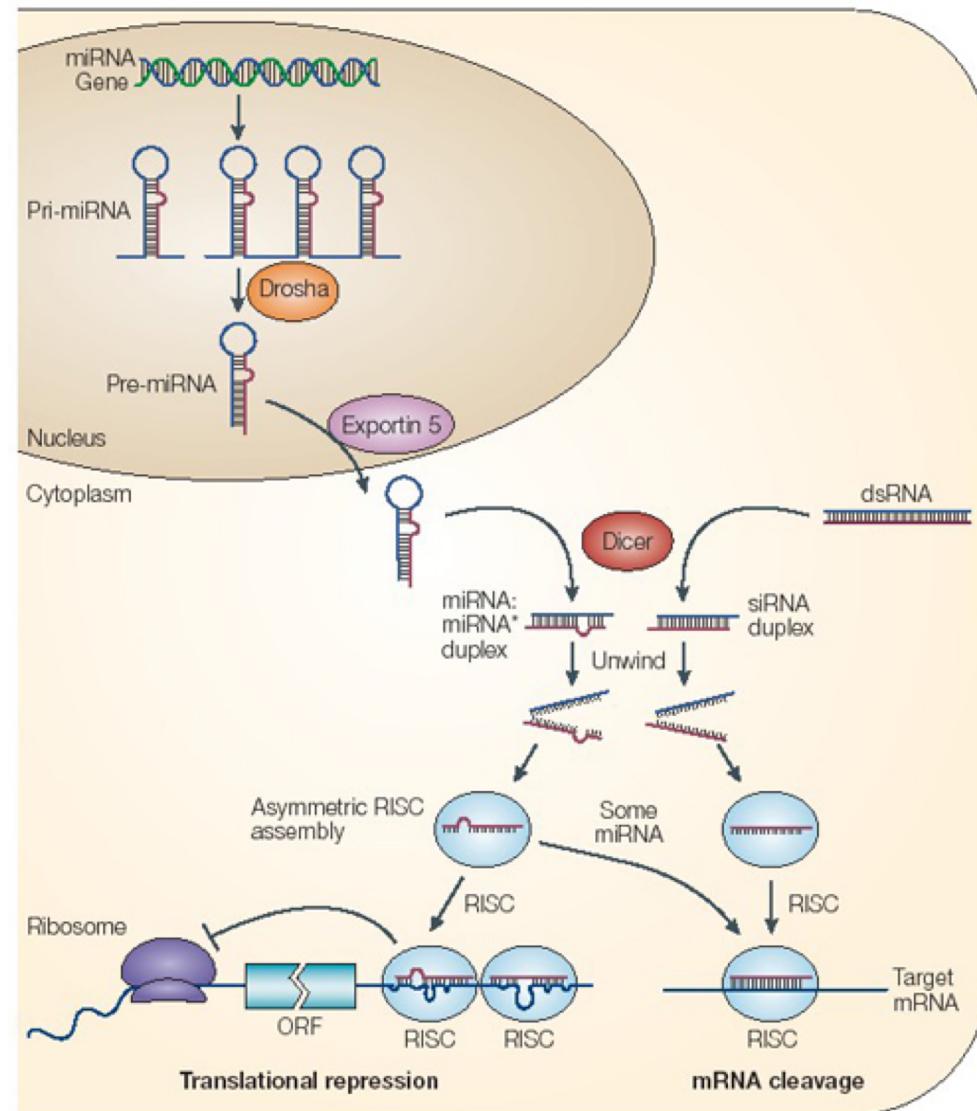
Decisive role in most, if not all, aspects of gene regulation (epigenetic processes...)

- **small regulatory ncRNA** (<200 nucleotides in length) with major roles in many cellular processes : small interfering RNA (siRNAs), microRNAs (miRNAs)...
- **long non-coding RNAs** (lncRNAs, >200 nucleotides in length) essential regulators in diverse cellular progresses including:
 - regulation of gene transcription (Orom et al., 2010; Sun et al., 2013),
 - dosage compensation (Ilik et al., 2013; Maenner et al., 2013)
 - genomic imprinting (Lee and Bartolomei, 2013; Simon et al., 2013),
 - DNA damage and nuclear organization (Wan et al., 2013)

(Cao et al., 2015. Three-dimensional regulation of transcription)

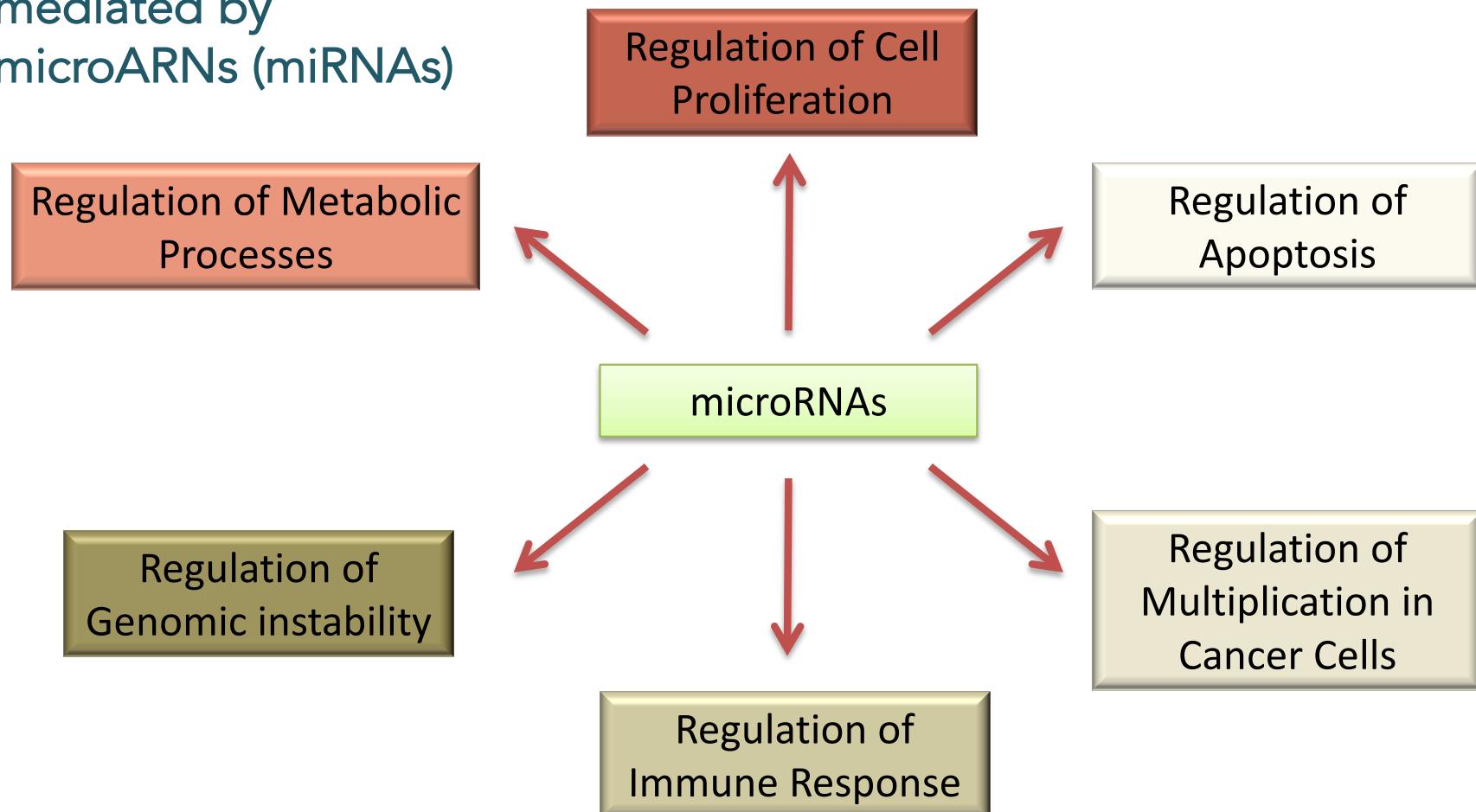
New era : understanding mechanisms of gene expression

Example of regulation
mediated by
microARNs (miRNAs)



New era : understanding mechanisms of gene expression

Example of regulation
mediated by
microARNs (miRNAs)

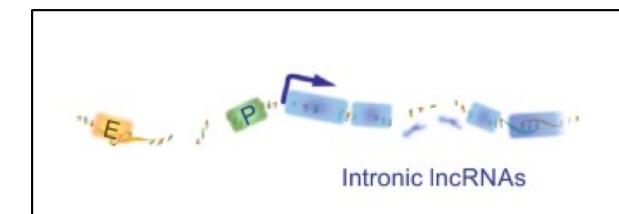
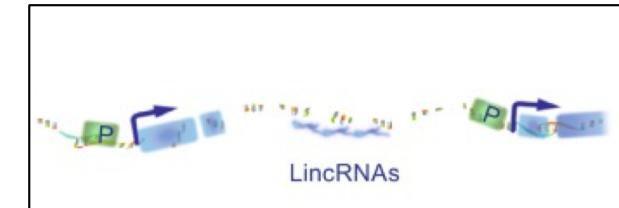
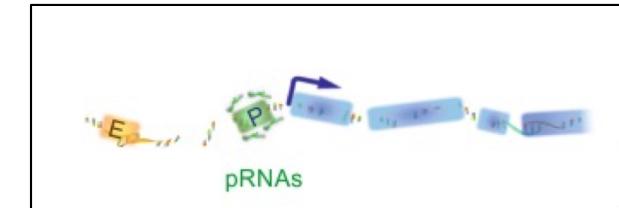
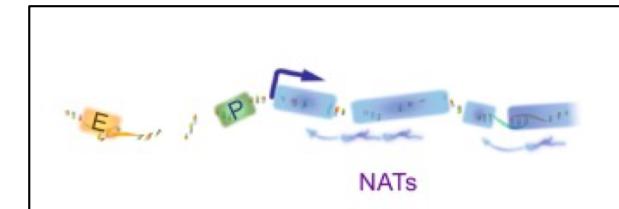


New era : understanding mechanisms of gene expression

lncRNAs : classification

Based on the genomic localization and context, lncRNAs can be classified as

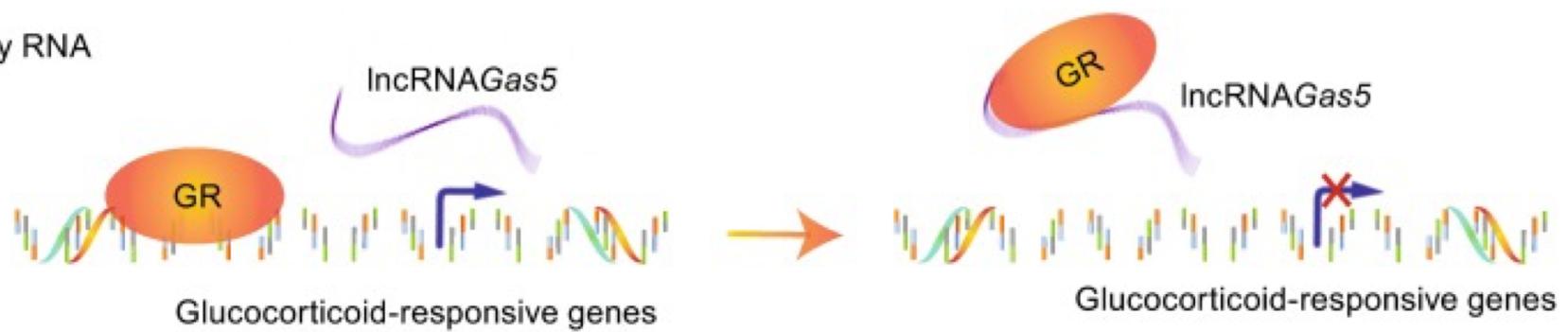
- **eRNAs** bidirectional and nonpolyadenylated transcripts transcribed from enhancers
- **pRNAs** from intragenic promoters
- **NATs** from the opposite strand of either protein or non-protein coding genes
- **lincRNAs** from regions intervening protein-coding loci
- **intronic lncRNAs** from specific introns of protein-coding genes



New era : understanding mechanisms of gene expression

lncRNAs : Regulatory models

A Decoy RNA

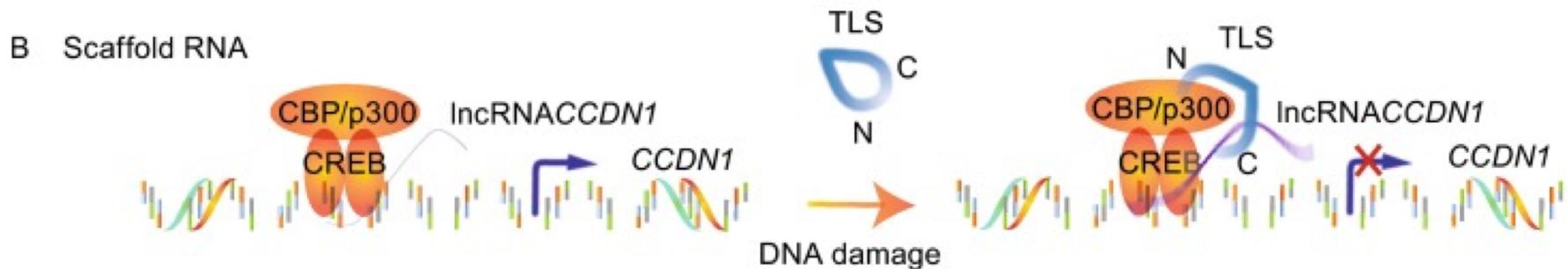


LncRNAs can bind to transcription factors or some other proteins away from chromatin and prevent them from binding to their proper regulatory targets.

Example: *lncRNAGas5*, transcribed from exon 7 of *Gas5* gene, directly interacts with the DNA-binding domain (DBD) of the glucocorticoid receptor (GR) → competes with DNA GREs for binding to the GR DBD.

New era : understanding mechanisms of gene expression

lncRNAs : Regulatory models



lncRNAs have a high structural flexibility → well-suited to assemble diverse combinations of regulatory proteins through specific secondary structures
→ enhance the protein-protein interactions.

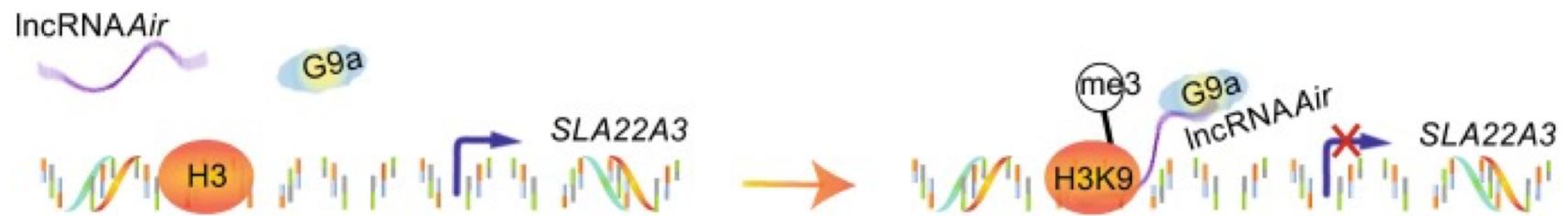
Example: *lncRNACCND1*. It is transcribed from the promoter region of the cyclin D (*CCND1*) gene. Upon DNA damage → acts as molecular 'ligand' for RNA-binding protein TLS → repression of *CCND1* transcription.

(Cao et al., 2015. Three-dimensional regulation of transcription)

New era : understanding mechanisms of gene expression

lncRNAs : Regulatory models

C Guide RNA



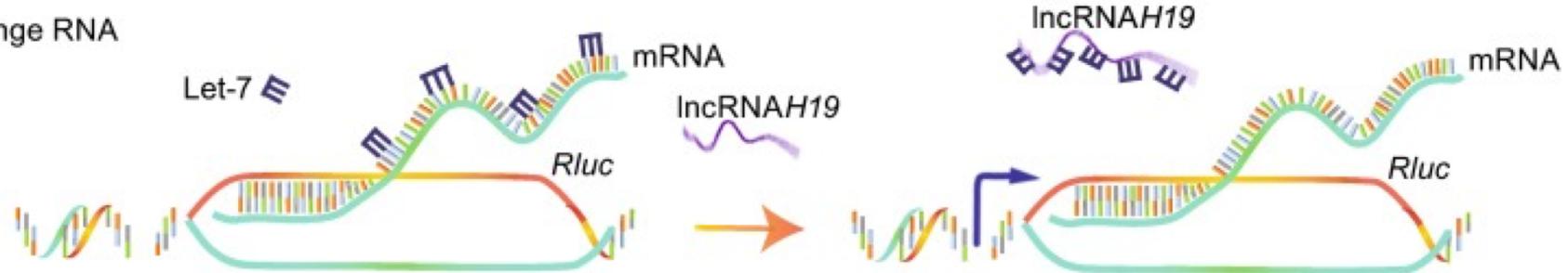
LncRNAs can function in *cis* on nearby gene or in *trans* on distally located genes through recruiting chromatin-modifying enzymes to targets.

Example: *IncRNAAir*, transcribed from an antisense promoter located in intron 2 of *Igf2r* → silences transcription of the distal *Slc22a3* gene via a specific chromosome interaction between *Air* and the *Slc22a3* promoter.

New era : understanding mechanisms of gene expression

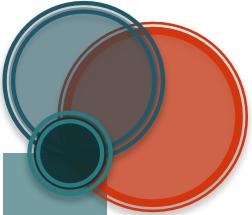
lncRNAs : Regulatory models

D Sponge RNA



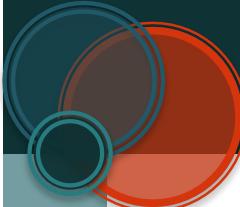
LncRNAs could competitively inhibit the ability of miRNAs to interact with their target mRNA.

Example: *IncRNAH19*, harbors both canonical and noncanonical binding sites for the let-7 family of micro-RNAs → acts as a molecular sponge to specifically sequester endogenous let-7, preventing it from inhibiting *Rluc* expression



Part 4

Proteins



Proteins

Definition

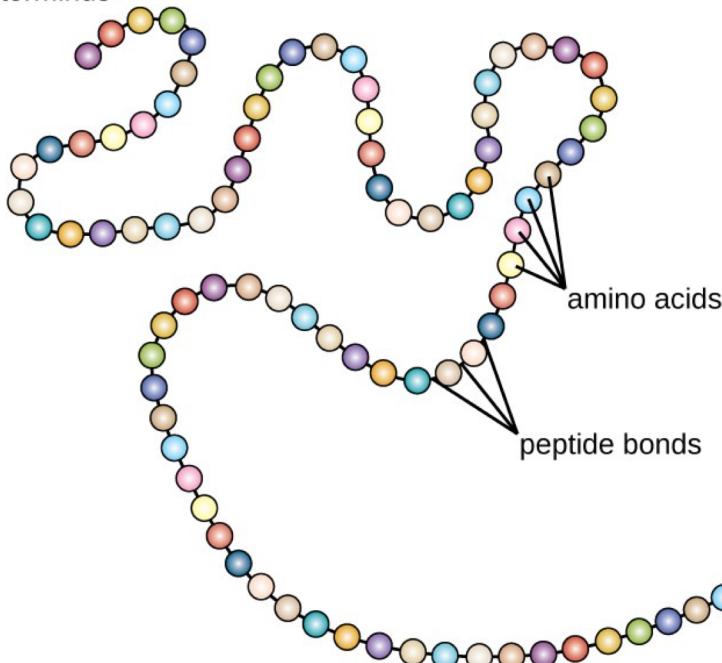
- **Proteins** consist of one or more long chains of amino acid residues → large biomolecules, or macromolecules.
- Perform a vast array of functions within organisms:
 - metabolic reactions
 - DNA replication
 - response to stimuli
 - transporting molecules from one location to another...
- Proteins are assembled from amino acids using information encoded in genes. Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of the gene encoding this protein.
- In general, the genetic code specifies 20 standard amino acids; however, in certain organisms the genetic code can include selenocysteine and—in certain archaea—pyrrolysine.

Proteins

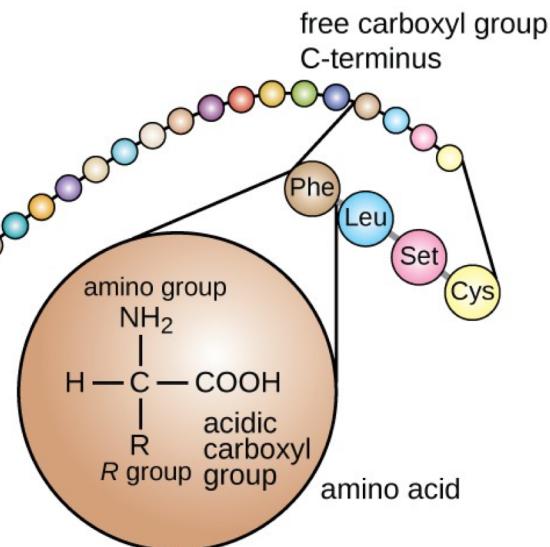
Primary structure

- A linear chain of amino acid residues is called a **polypeptide**. A protein contains at least one long polypeptide.

free amino group,
N-terminus

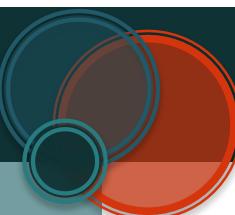


The primary protein structure
is the chain of amino acids
that makes up the protein.



Important

- **NB: Short polypeptides**, containing less than 20–30 residues, are rarely considered to be proteins and are commonly called **peptides**, or sometimes **oligopeptides**. The individual amino acid residues are bonded together by **peptide bonds** and adjacent amino acid residues.



Proteins

Synthesis

Peptides (from Gr.: πεπτός, peptós "digested"; derived from πέσσειν, péssein "to digest") are natural biological or artificially manufactured short chains of amino acid monomers linked by peptide (amide) bonds.

A polypeptide is a long, continuous peptide chain (dipeptides, tripeptides...).

Peptides fall under the broad chemical classes of biological **oligomers** and **polymers**, alongside **nucleic acids**, **oligosaccharides** and **polysaccharides**,

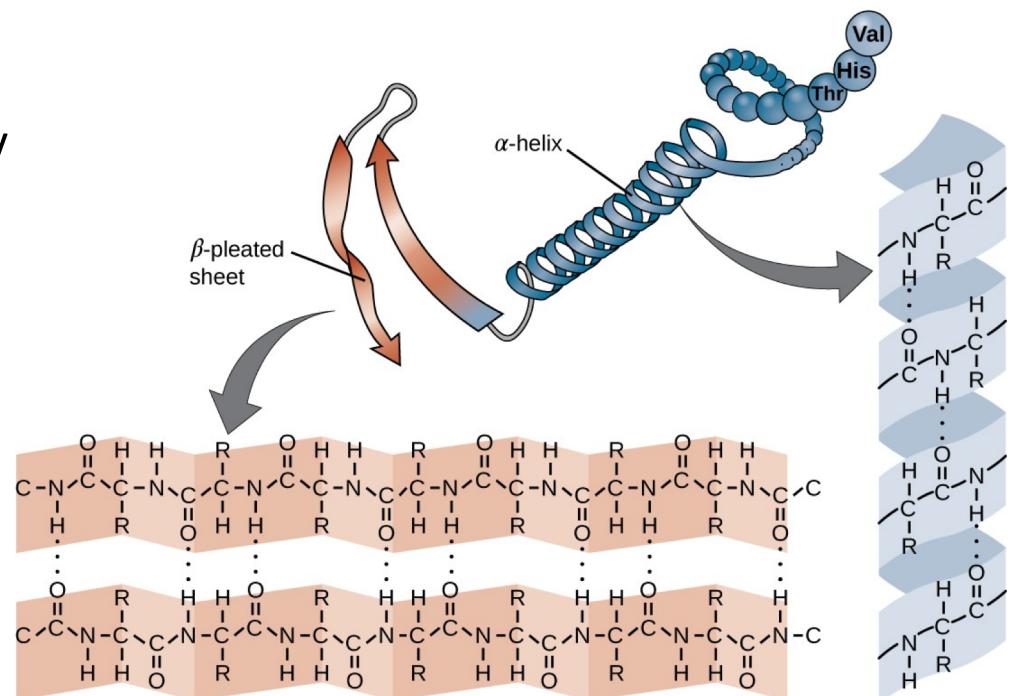
Oligomers : (oligo:"a few", mer:"parts") is a molecular complex that consists of a few monomer units, in contrast to a polymer, where the number of monomers is, in principle, unlimited (dimers, trimers, tetramers...)

Polymers: (poly:"many", mer:"parts"), macromolecule, composed of many repeated subunits = broad range of properties, both synthetic and natural.

Proteins

Secondary Structure

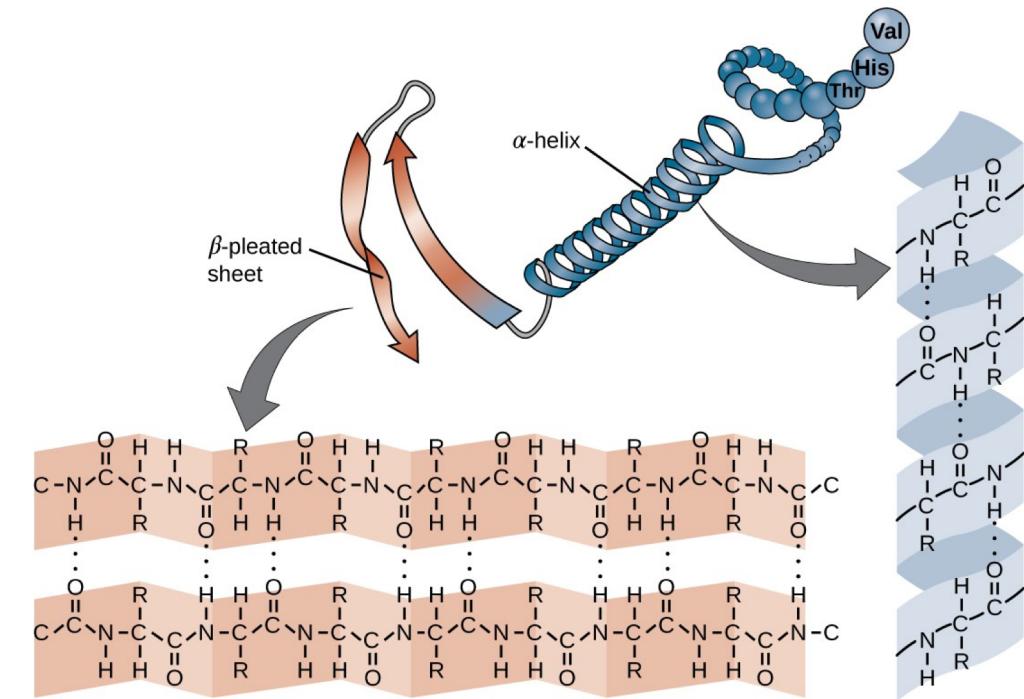
- The chain of amino acids that defines a protein's primary structure is not rigid.
 - In long chains hydrogen bonding may occur between amine and carbonyl functional groups within the peptide backbone
- localized folding of the polypeptide chain into **helices** and **sheets**. These shapes constitute a protein's **secondary structure**.



Proteins

Secondary Structure

- The most common secondary structures are the α -helix and β -pleated sheet.



- α -helix** structure: **H-bonds** between the **O atom** in a **carbonyl group** of one aa and the **H atom** of the **amino group** that is just four amino acid units further.
- β -pleated sheet**, the pleats are formed by similar **H bonds** between continuous sequences of **carbonyl** and **amino groups** that are further separated on the backbone of the polypeptide chain.

Proteins

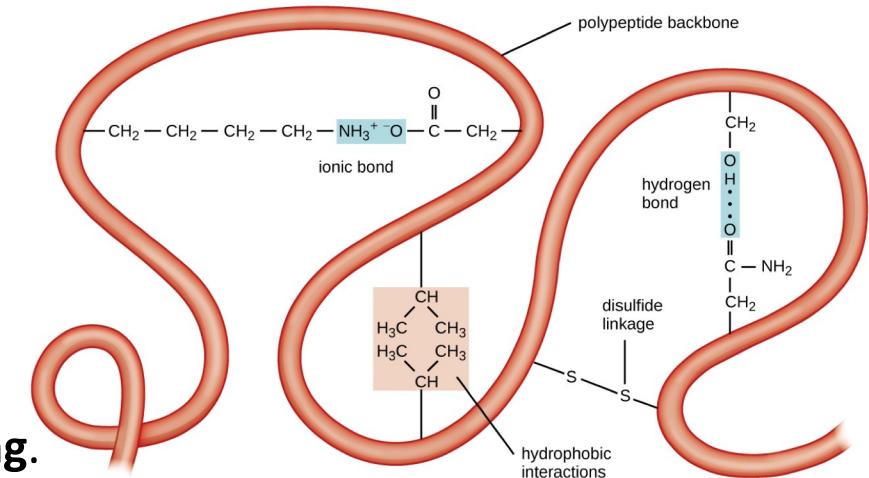
Tertiary Structure

- the **tertiary structure** is the large-scale 3D shape of a single polypeptide chain. Tertiary structure is determined by **interactions between amino acid residues that are far apart in the chain**. A variety of interactions give rise to protein tertiary structure, such as:

- **disulfide bridges** bonds ($-SH$)
- **hydrogen bonds**
- **ionic bonds**
- **hydrophobic interactions** between nonpolar side chains.

→ These processes determine **protein folding**.

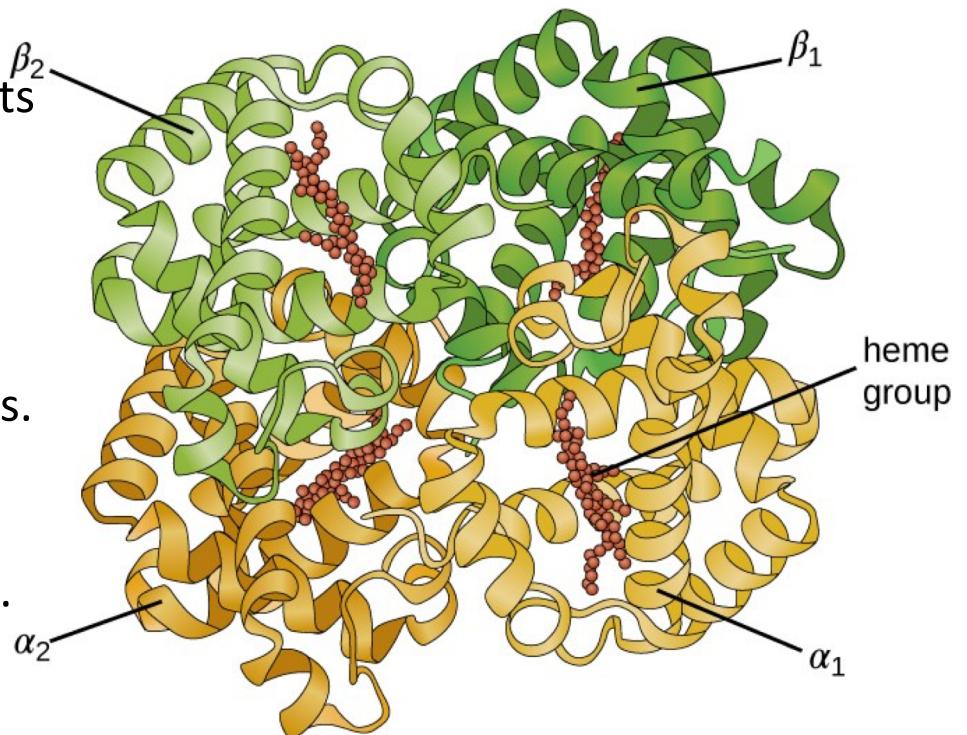
- All these interactions, weak and strong, combine to determine the final **3D shape of the protein** and its **function**. Folded proteins that are fully functional in their normal biological role are said to possess a **native structure**.



Proteins

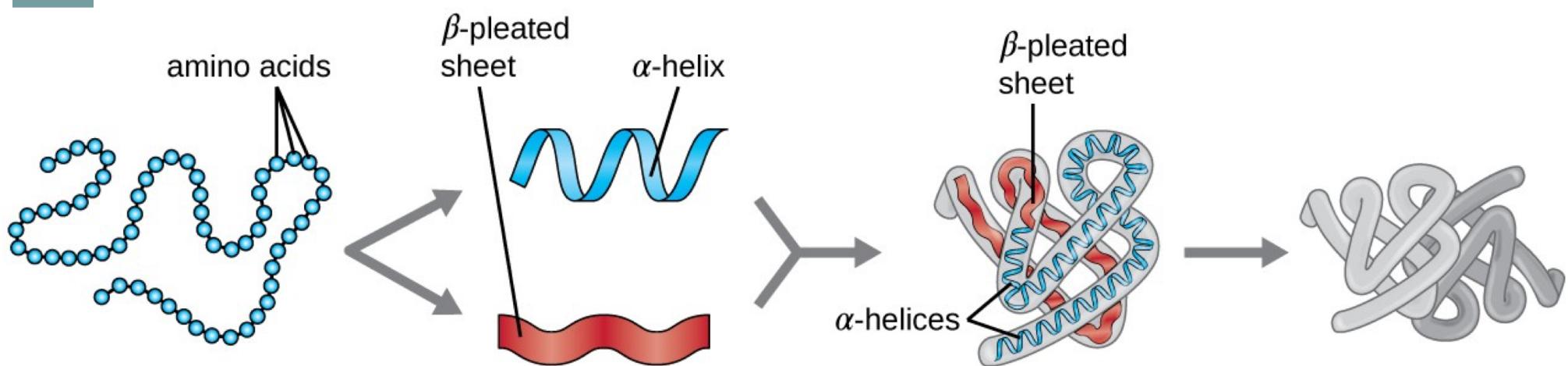
Quaternary Structure

- Some proteins are assemblies of several separate polypeptides, also known as protein subunits. These proteins function adequately only when all subunits are present and appropriately configured.
- The interactions that hold these subunits together constitute the **quaternary structure** of the protein.
- The overall quaternary structure is stabilized by relatively weak interactions.
- Example of Hemoglobin: quaternary structure of 4 globular protein subunits.



Proteins

In Brief



Primary Protein Structure
Sequence of a chain of amino acids

Secondary Protein Structure
Local folding of the polypeptide chain into helices or sheets

Tertiary Protein Structure
three-dimensional folding pattern of a protein due to side chain interactions

Quaternary Protein Structure
protein consisting of more than one amino acid chain

Post-translational modifications (PTMs)

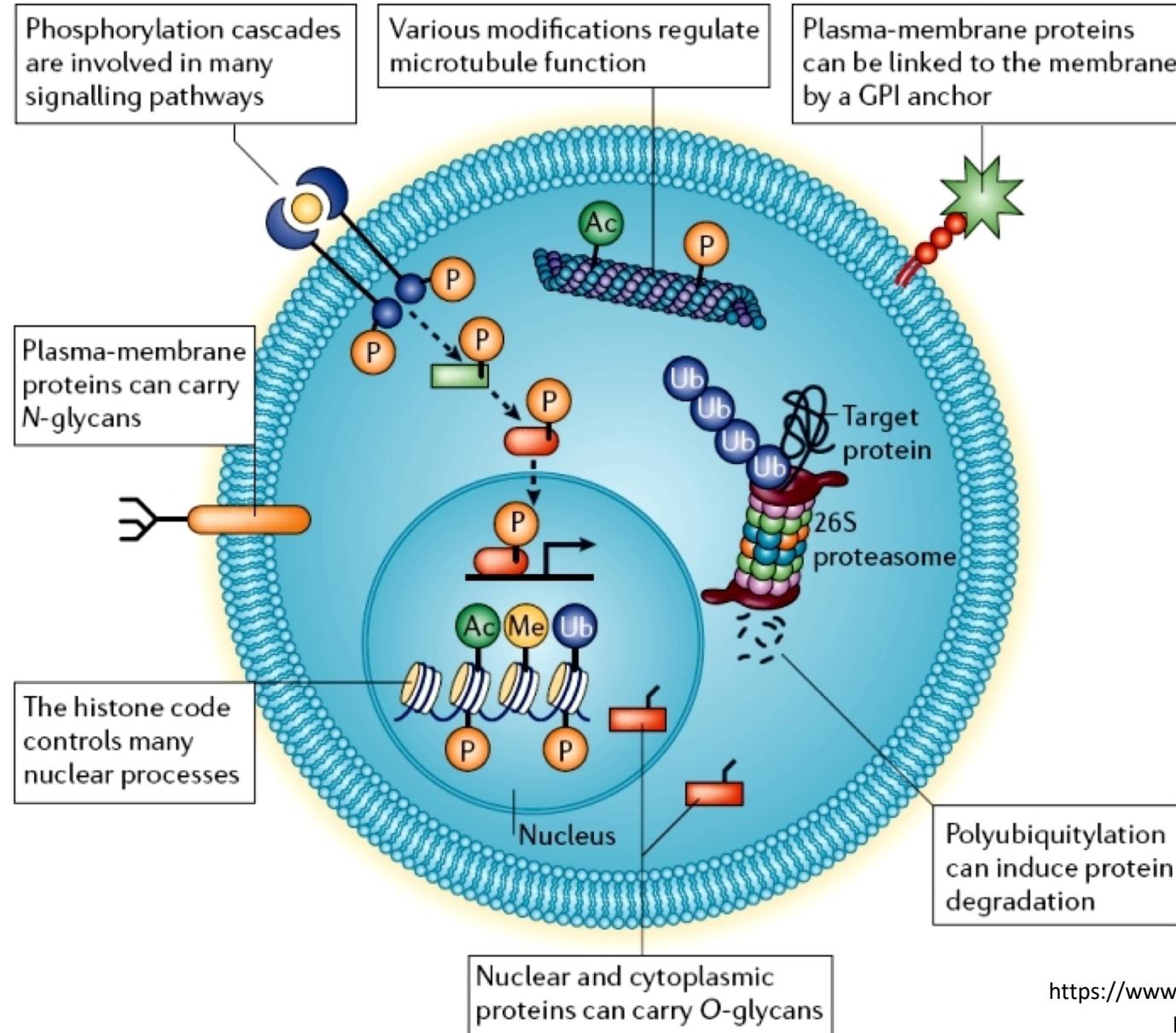
- Shortly after or even during synthesis, the residues in a protein are often chemically modified by post-translational modification
- **Post-translational modifications** alters:
 - the physical and chemical properties
 - Folding
 - Stability
 - Activity
 - Function of the proteins.
- Proteins can also work together to achieve a particular function, and they often associate to form stable **protein complexes**.
- Sometimes proteins have **non-peptide groups attached**, which can be called prosthetic groups or cofactors = **conjugated proteins**.

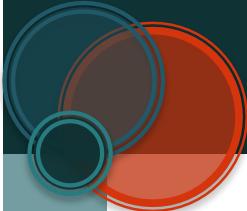
Conjugated Proteins

- Important class of proteins is the conjugated proteins that have a non-protein portion.
- If the conjugated protein has a carbohydrate attached, it is called a **glycoprotein**.
- If it has a lipid attached, it is called a **lipoprotein**.
- These proteins are important components of membranes.

Proteins

Post-translational modifications (PTMs)





Proteins

Abundance in cells

- average-sized bacteria ≈ 2 million proteins per cell (e.g. *E. coli*, *S. aureus*).
- Smaller bacteria contain fewer molecules ≈ 50,000 to 1 million (e.g. Mycoplasma, spirochetes).
- Eukaryotic cells are larger and thus contain much more protein:
 - ≈ 50 million proteins in yeast cells
 - ≈ 1 to 3 billion in human cells.
- The concentration of individual protein copies ranges from a few molecules per cell up to 20 million
→ depends on the number of genes expressed (cell type, external stimuli...)

The Codon Bias

Definition

Codon bias is the tendency of genomes to prefer one particular codon for an amino acid over all the others.

Evolutionary importance of codon bias

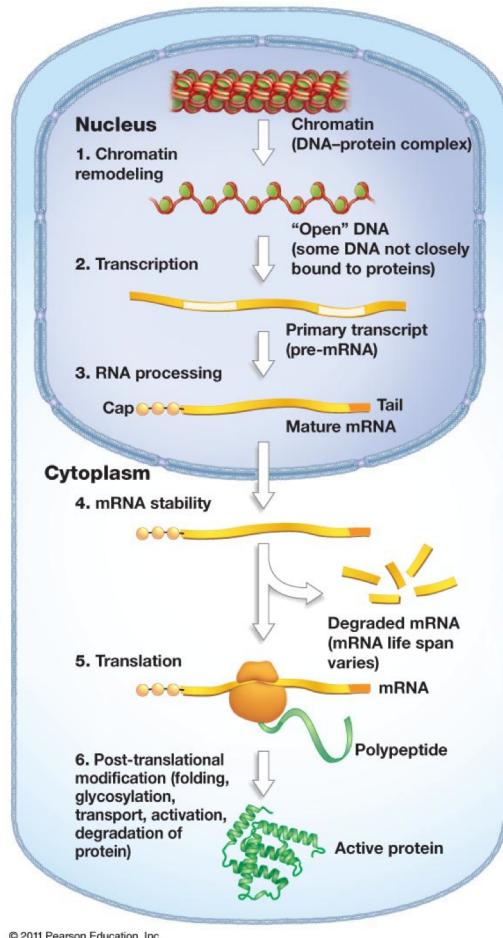
Measure of the Codon Adaptation Index (CAI) : the geometric weight of a specific codon over the whole geometric weight of the genome

→ Comparison of CAI among different organisms:

- codons with the highest CAI's represented the preferred codon, or the codon that receives the bias, and these were codons rich in Guanine and Cytosine.
- GC-rich genomes : GC bases seems evolutionarily superior.

Overview of regulation processes

Methods used by eukaryotes for gene regulation A regulation with several coding levels

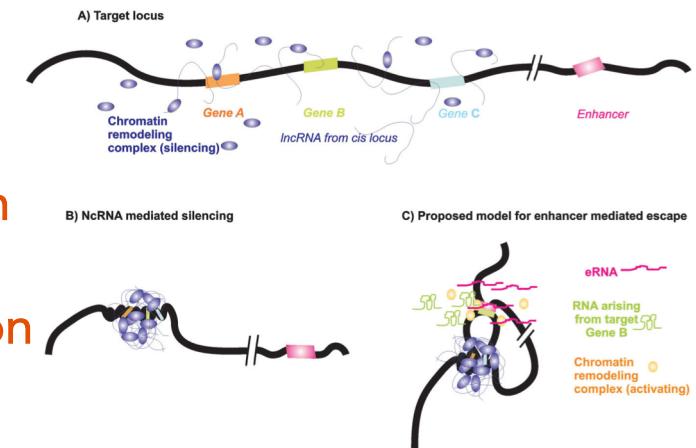


Acetylation
Ubiquitination
Methylation
Phosphorylation
Degradation
(...)

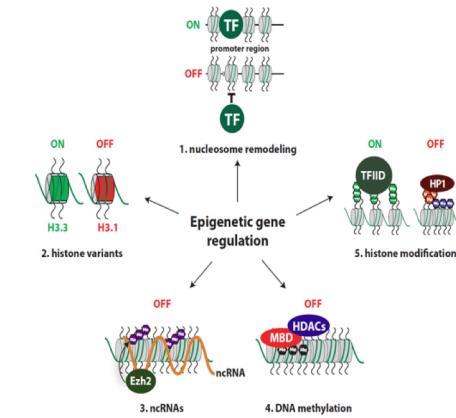
+

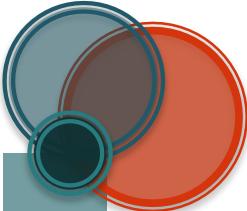
Response to stimuli
Transcription initiation

lncRNA mediated silencing



Epigenetic regulation



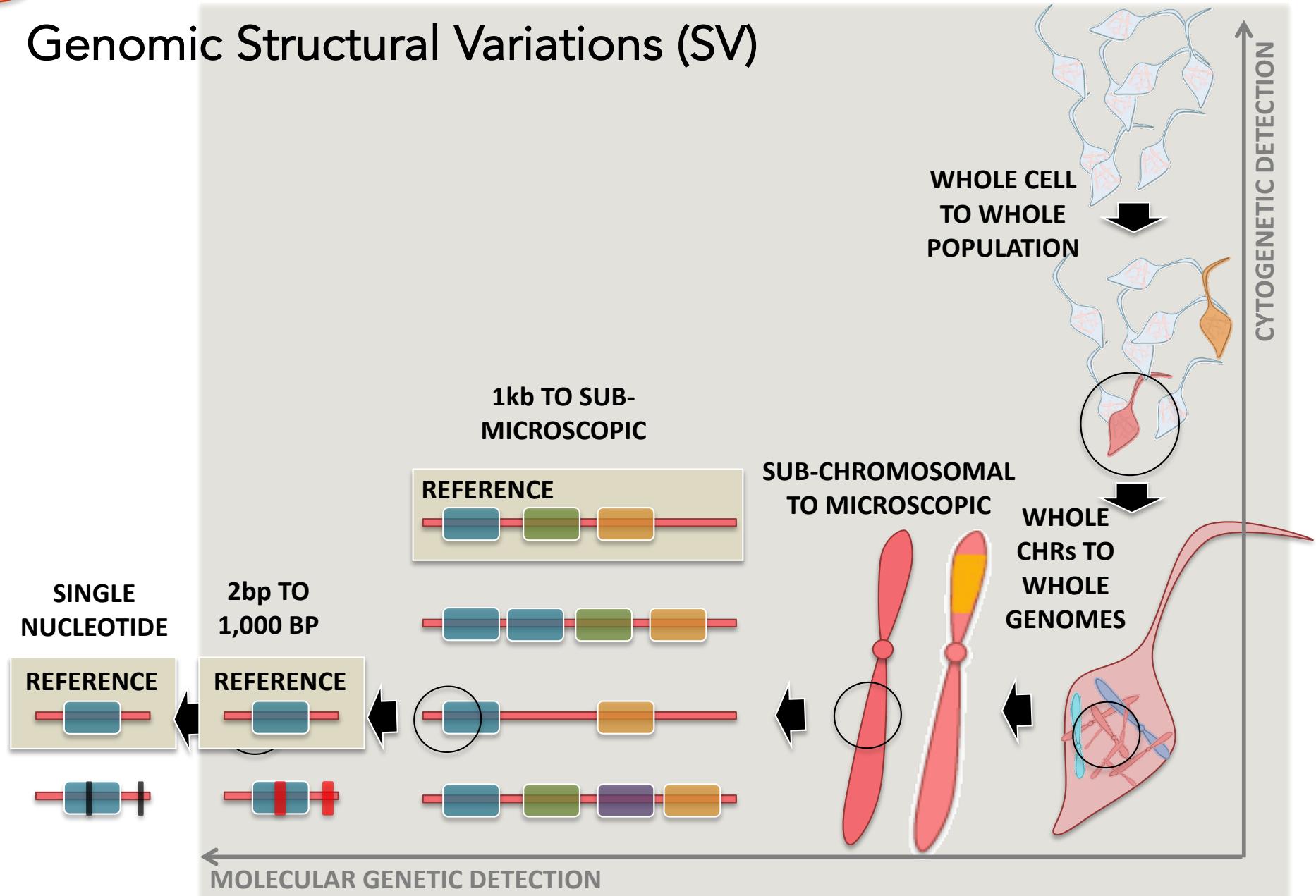


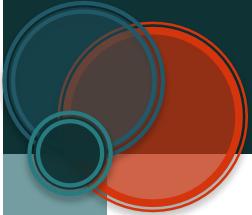
Part 5

Point mutations and their consequences

Point mutations and their consequences

Genomic Structural Variations (SV)





Point mutations and their consequences

Classification

- **Chromosomal mutations:**

Affects entire chromosomes or entire portions of it. Sometimes called segmental mutations

- **Point mutations:**

Affects 1 or few base-pairs. Consists of

- **Substitutions** : change of 1 or few nucleotides in an existing sequence
- **Insertions**: addition of 1 or few nucleotides in an existing sequence
- **Deletions**: loss of 1 or few nucleotides in an existing sequence

Point mutations and their consequences

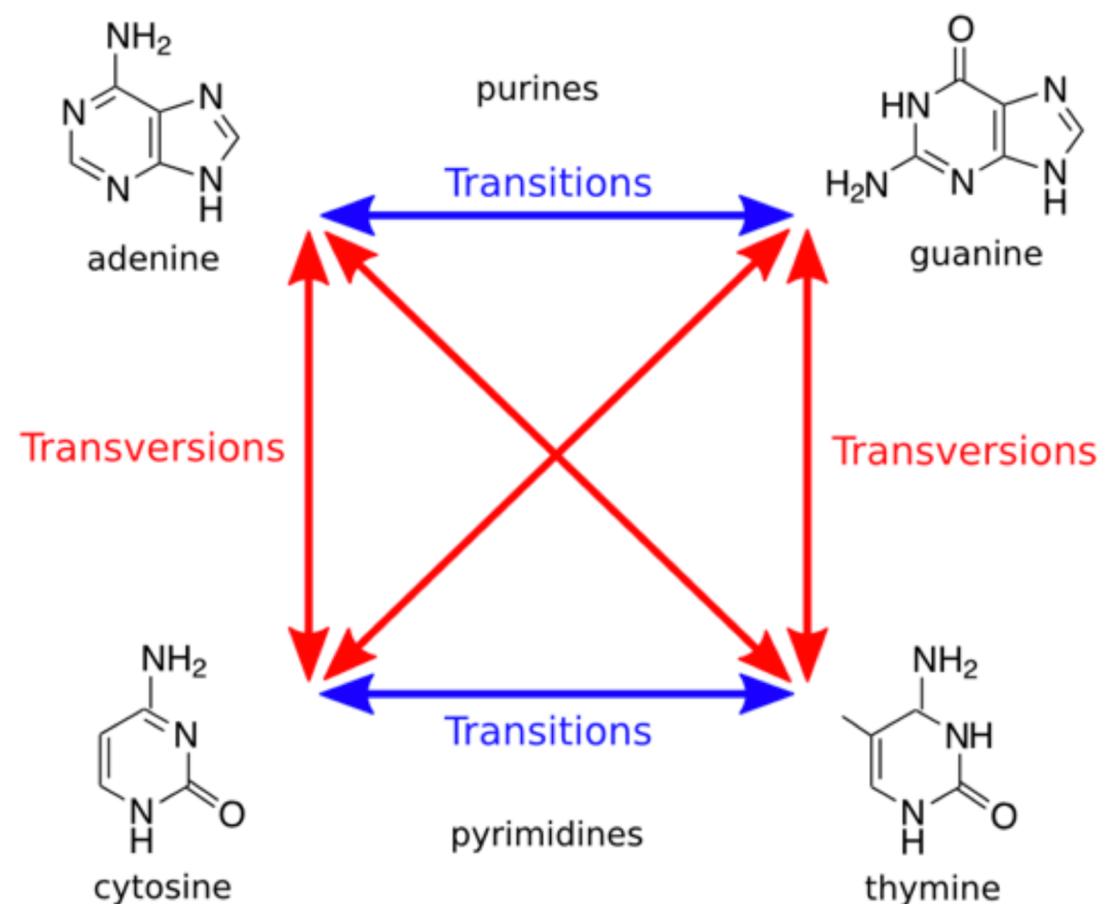
Substitutions

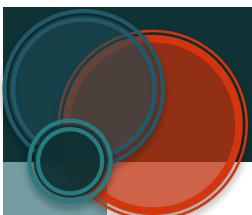
- **Transition**

Substitution of a
(Pyrimidine → Pyrimidine)
or (Purine → Purine)

- **Transversion**

Substitution of a
(Pyrimidine → Purine)
or (Purine → Pyrimidine)





Point mutations and their consequences

Substitutions

Synonymous mutations

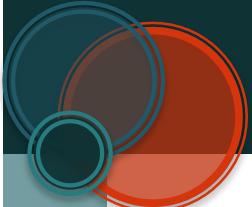
A synonymous codon is one that, if mutated, would still code for the same amino acid.

This is possible because the genetic code is "degenerate".

In most cases, if the last nucleotide in the codon is changed it generally codes for the same amino acid

A point mutation causing a synonymous substitution

Type of structure	Before	Change	After	Result
Codon in a DNA sequence	TTT	harmless mutation, ^[c] Synonymous substitution	TTC	
↓ codes for	↓ codes for		↓ codes for	
Amino acid in a Protein	Phenylalanine	no change	Phenylalanine	Normal protein, normal function



Point mutations and their consequences

Substitutions

Non-Synonymous mutations

A base pair is changed, and the codon codes for a different amino acid.

If the mutation is for an amino acid with similar qualities, such as hydrophobic or polar tendencies, then there is a strong possibility that the resulting protein will resemble much of the same structure.

A point mutation causing a nonsynonymous substitution

Type of structure	Before	Change	After	Result
Codon in a DNA sequence	GAG	Missense mutation; Nonsynonymous substitution	GTG	
↓ codes for	↓ codes for		↓ codes for	
Amino acid in a Protein	Glutamic acid	structural change	Valine	Altered protein may or may not cause harm (e.g. disease) or give new advantage

Point mutations and their consequences

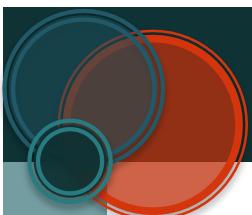
Substitutions in the coding genome

Point mutations				
No mutation	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

Chemical structures:

- lysine (Lys):** A basic amino acid with a long hydrocarbon chain ending in an amino group (NH_2^+). It is shown in blue.
- arginine (Arg):** A basic amino acid similar to lysine, also with a long hydrocarbon chain and an amino group ($\text{HN}=\text{NH}_2^+$). It is shown in blue.
- threonine (Thr):** A polar amino acid with a short hydrocarbon chain ending in a hydroxyl group ($\text{H}_2\text{C}-\text{CH(OH)}-\text{CH}_3$). It is shown in green.

basic
polar



Point mutations and their consequences

Substitutions in the non-coding genome

- Sequencing the coding region, or exome, of individuals suspected of having a genetic disorder identifies 25–50% of disease-associated mutations. This leaves the genetic etiology of many cases undetermined → **role in non-coding?**
- **Functional regulatory elements** (enhancers...) in non-coding regions of the genome are associated with congenital anomalies

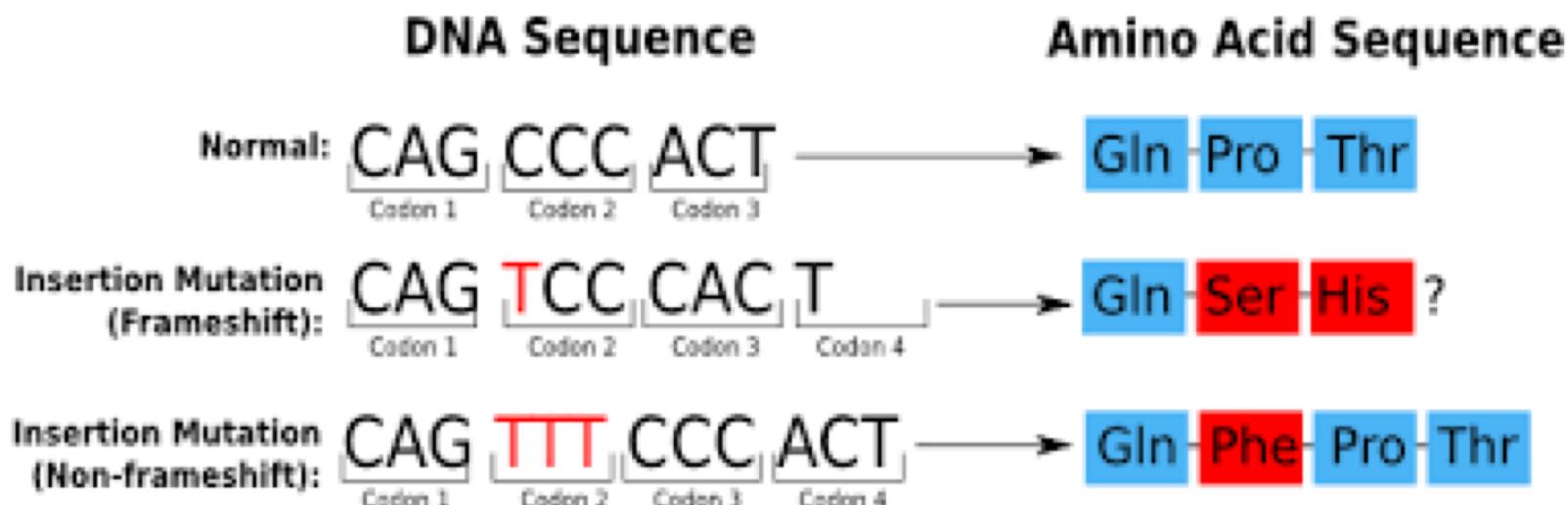
Examples of genetic conditions caused by mutations **outside the exome**

<u>Genetic Disease/Condition</u>	<u>Mutation</u>
Preaxial polydactyly 2 (PPD2)	SHH enhancer
Pancreatic agenesis	PTF1A enhancer
Pierre Robin Sequence (PRS)	SOX9 enhancer
Hirshsprung disease	RET enhancer
Isolated congenital heart defect	TBX5 enhancer
F-syndrome, Polydactyly, Brachydactyly	Disruption of TAD boundary elements near WNT6, IHH, EPHA4, and PAX3

Point mutations and their consequences

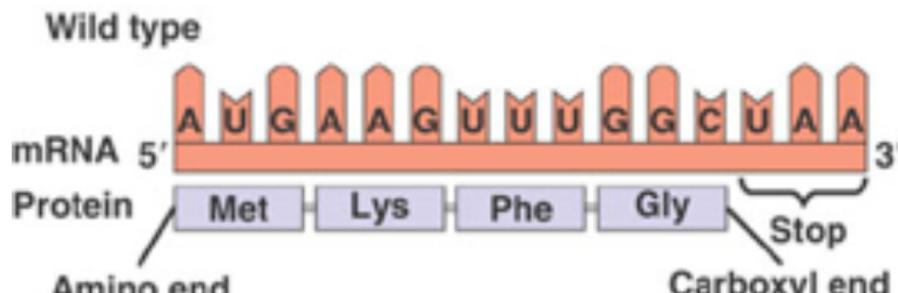
Insertion/Deletion (InDels)

- In-frame mutation if the successive order of codons is maintained
- Frameshift mutation if the sequence is changed or a premature stop codon appears.



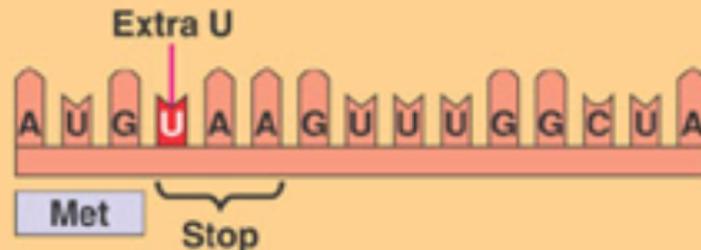
Point mutations and their consequences

Insertion/Deletion (InDels)

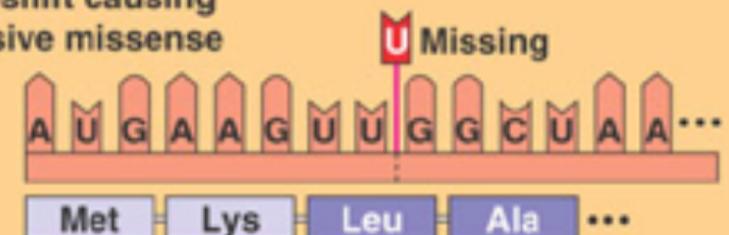


Base-pair insertion or deletion

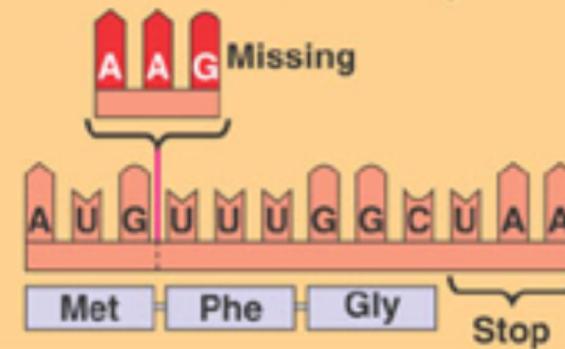
Frameshift causing immediate nonsense



Frameshift causing extensive missense

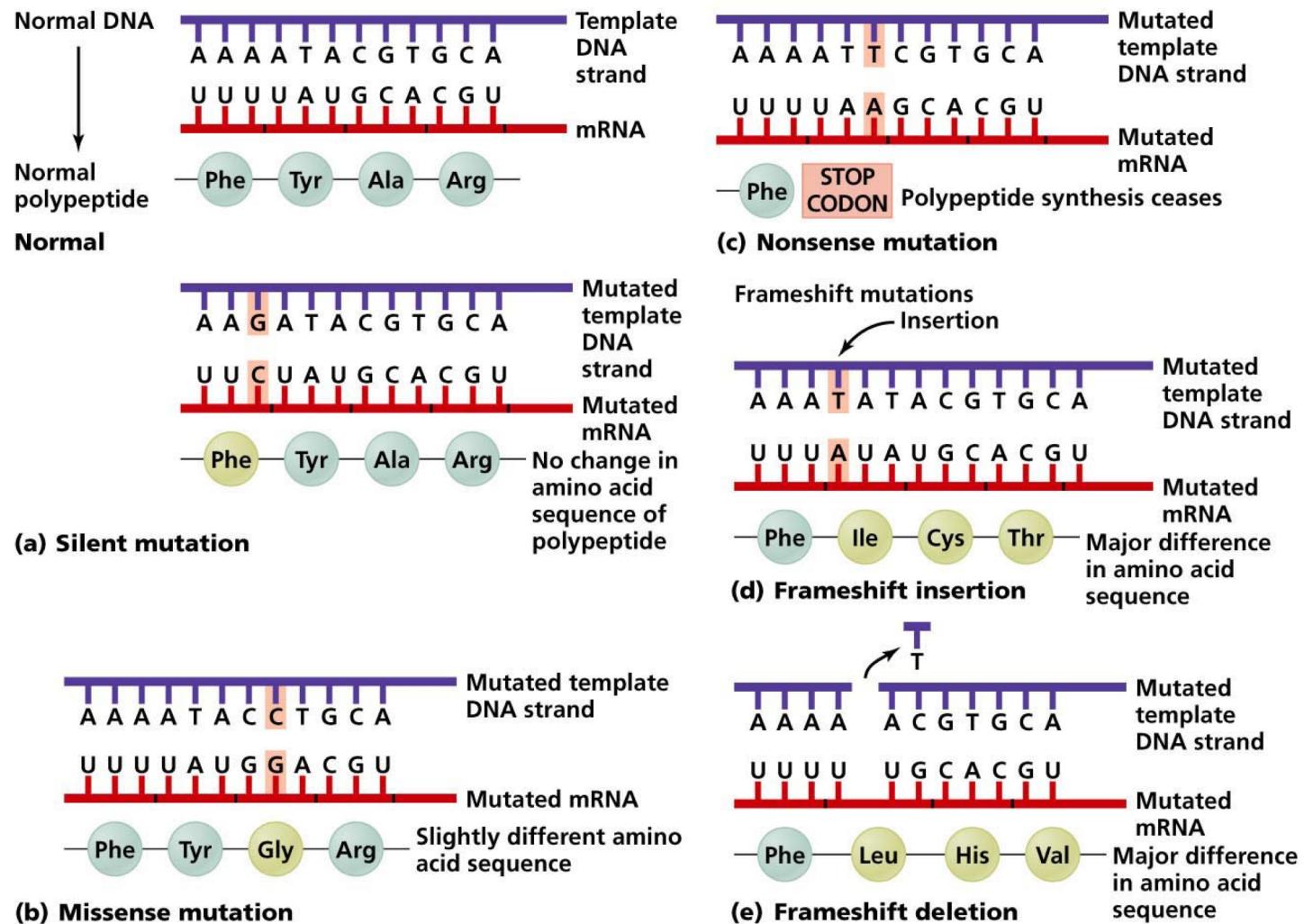


Insertion or deletion of 3 nucleotides:
no frameshift but extra or missing amino acid

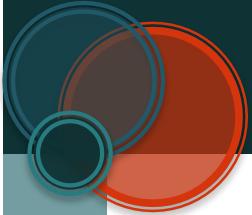


Point mutations and their consequences

Summary



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings.



Point mutations and their consequences

Patterns of Codon usage and Synonymous mutations

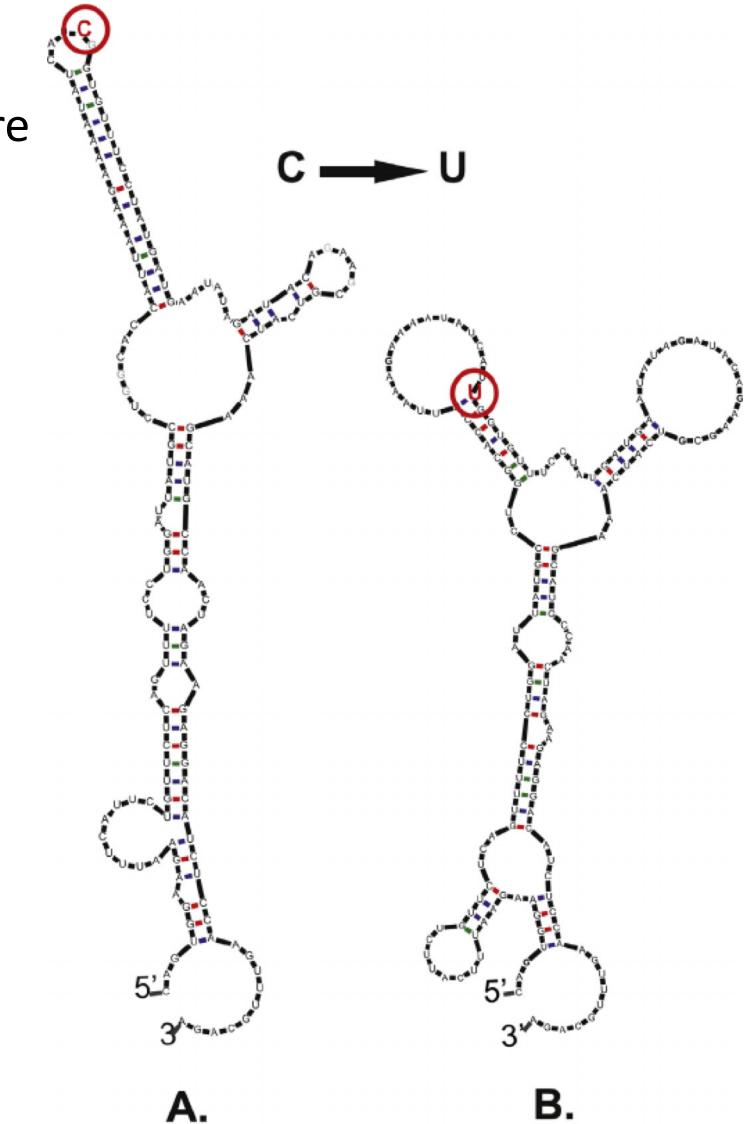
- Codon usage can vary dramatically even within a single gene.
- **Synonymous mutations at specific sites** may experience selection because they disrupt motifs recognized by TFs or by PTMs.
- Codon choice that promotes proper **nucleosome positioning** is selectively advantageous in eukaryotes, especially in 5' regions (Warnecke et al., 2008).
- In mammals, synonymous mutations near an **intron–exon boundary** can create spurious splice sites or disrupt splicing-control elements, causing **disease**.
- This phenomenon helps to explain the reduced rate of synonymous substitutions and SNP density near splicing control elements.

Point mutations and their consequences

Patterns of Codon usage and Synonymous mutations

- Synonymous mutations can affect mRNA structure

Consequences of a synonymous single nucleotide change on the predicted structure of the mRNA (mfold).

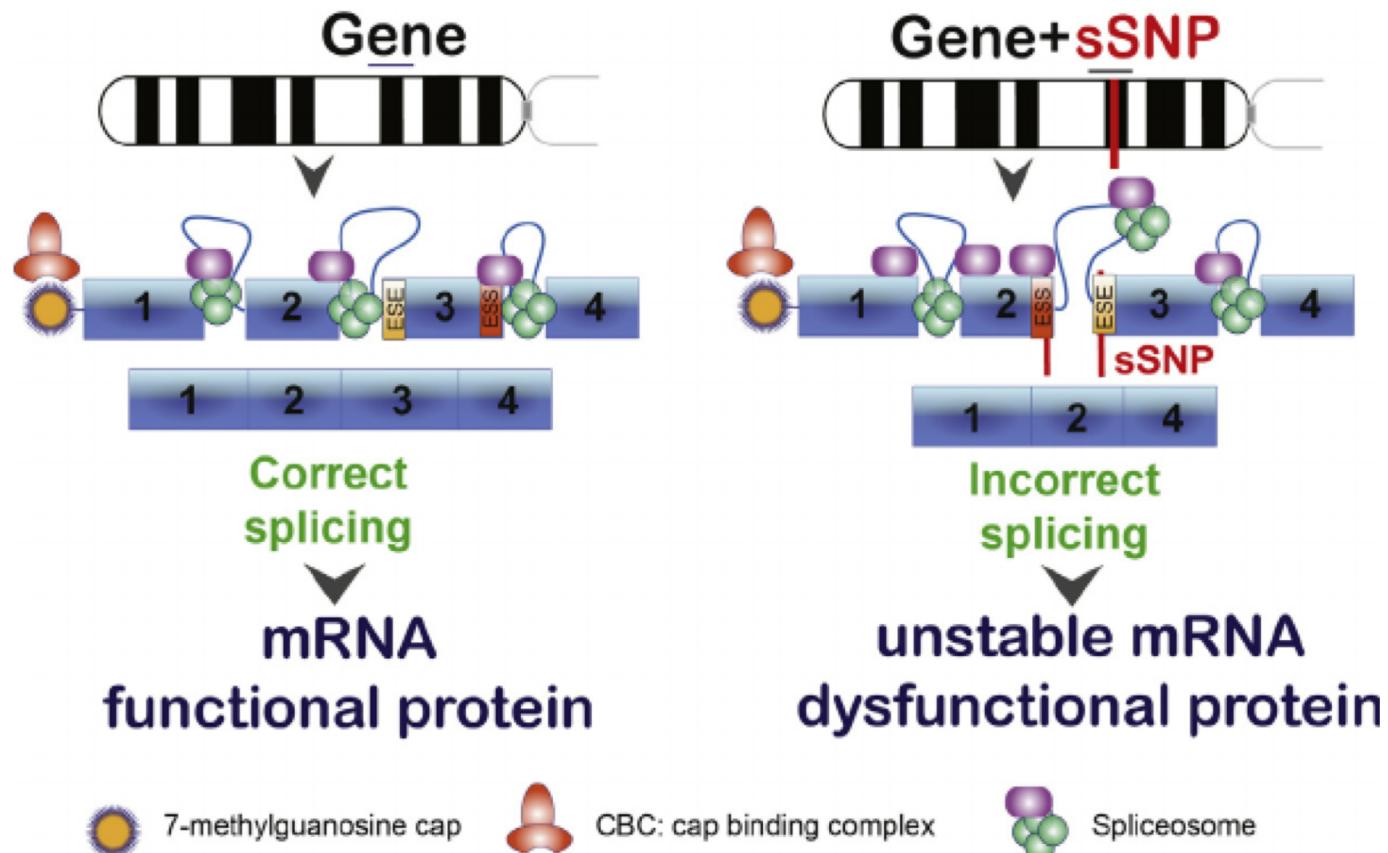


(Bali et al., 2015)

Point mutations and their consequences

Patterns of Codon usage and Synonymous mutations

- Synonymous mutations can affect exonic splice regulatory sites
- The occurrences of synonymous codons vary between species and within genes of the same genome, known as codon usage bias.



Point mutations and their consequences

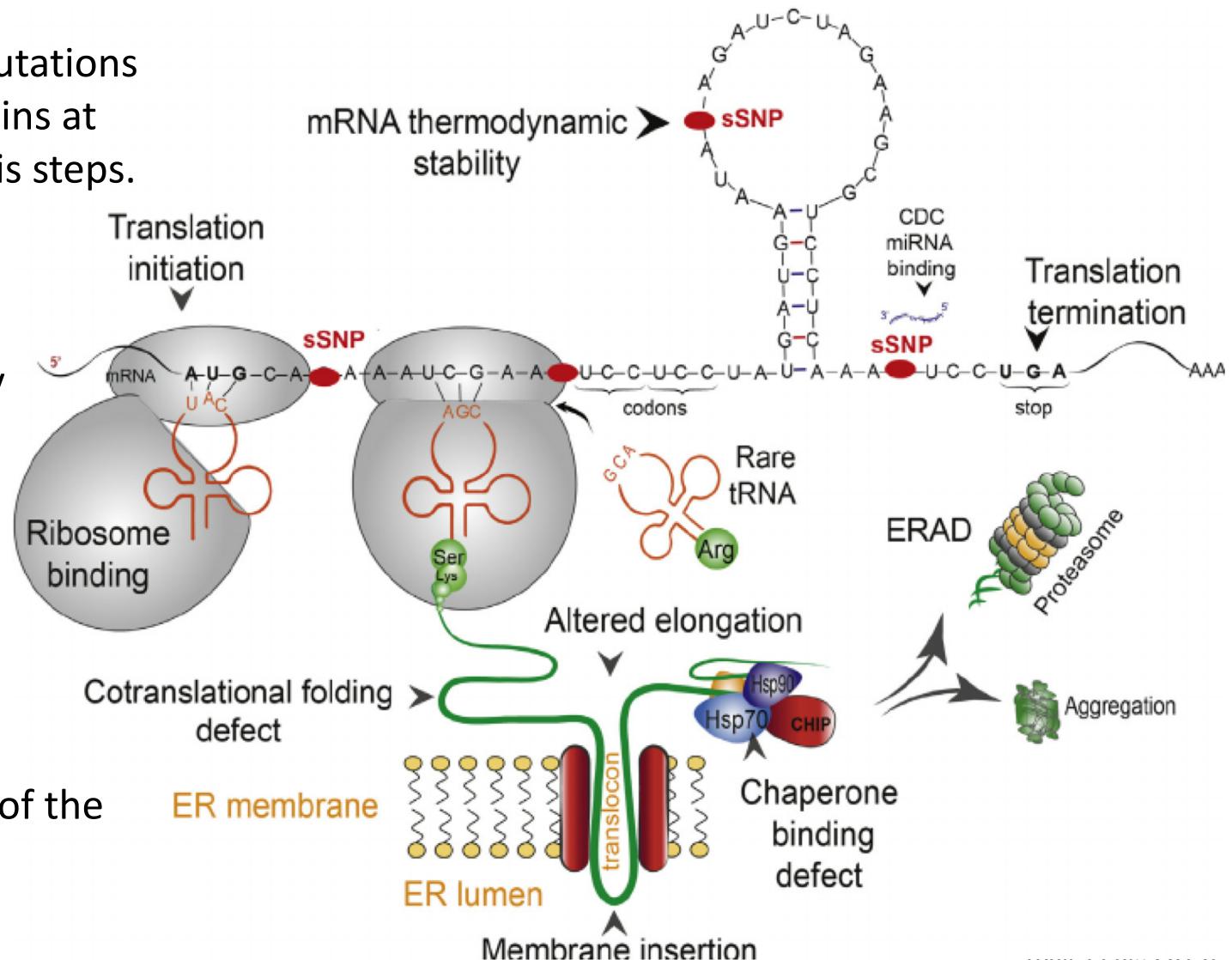
Patterns of Codon usage and Synonymous mutations

- Synonymous mutations can affect proteins at various synthesis steps.

Can alter:

- translation initiation efficiency
- translation elongation rate
- ribosomal pause rhythm
- cotranslational folding

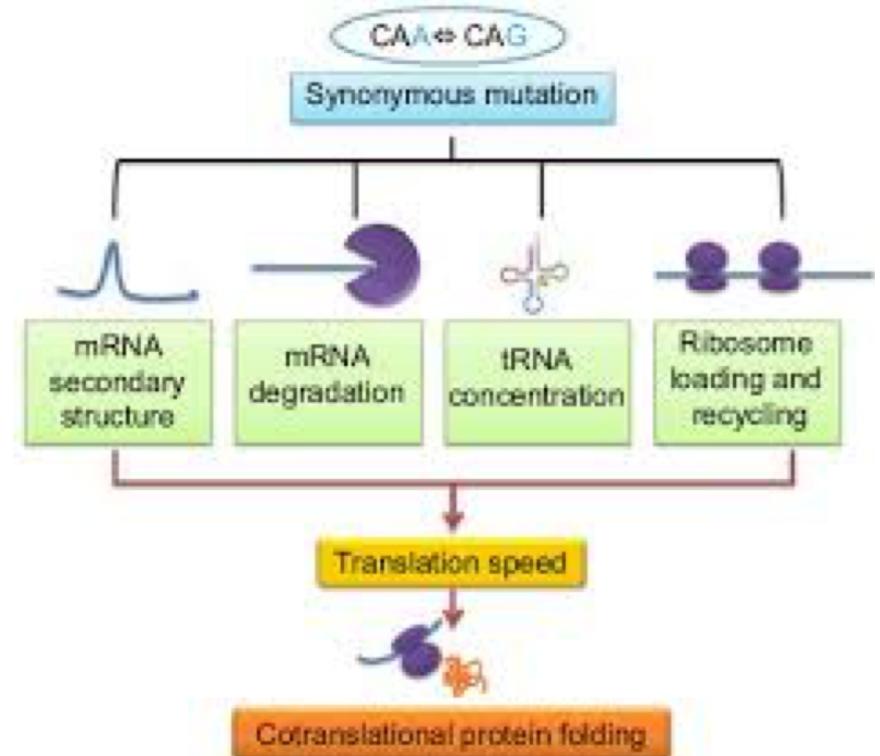
→ the overall fate of the protein.

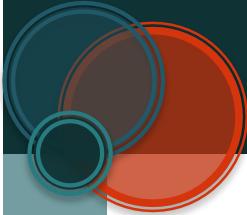


Point mutations and their consequences

Patterns of Codon usage and Synonymous mutations

- Synonymous Codons can affect gene expression and protein structure
 - Ribosome profiling experiments → **ribosome occupancy time** is not the same for different synonymous codons.
 - Conclusion:
 - Synonymous codons influence differently the **speed of translation elongation**
 - This guides further cotranslational **folding kinetics** of a protein.
 - It is now realized that the position of each codon in a coding sequence is important.





Concluding remarks

Complexity and definitions in constant evolution

- Complex regulatory picture since HGP and associated projects
 - Dogmas in constant revision
 - Novel key findings at the level of DNA, RNA and Proteins
- Keep updated !

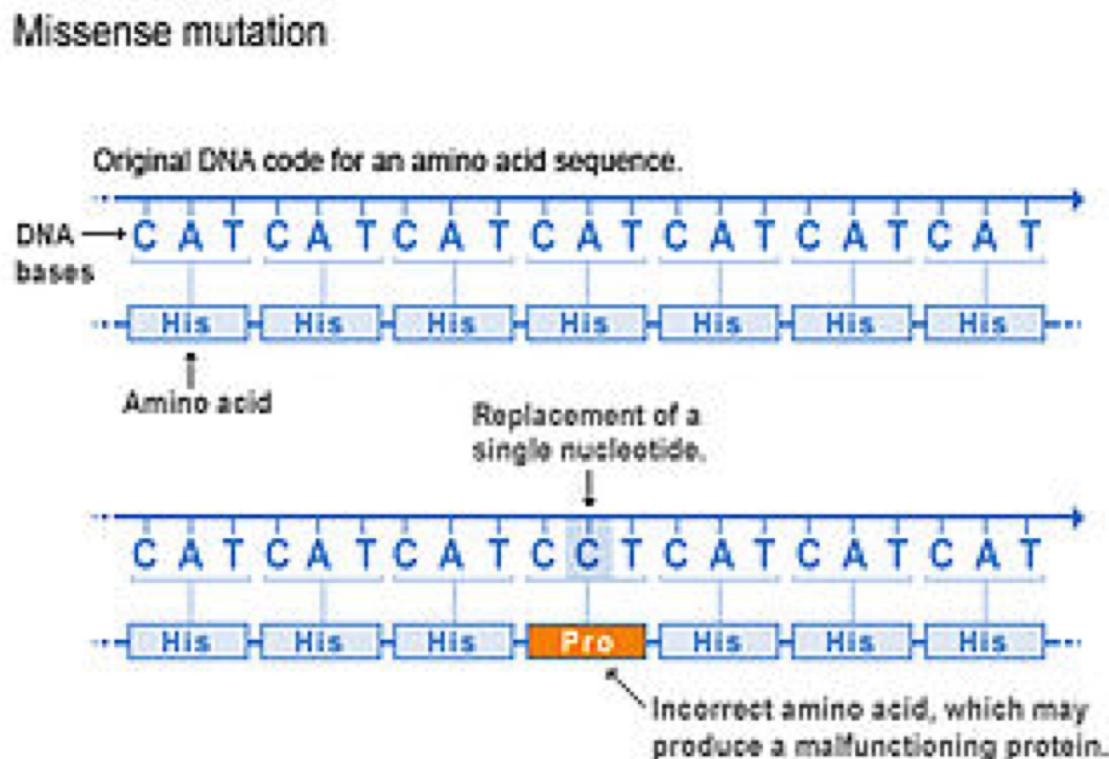
Remember: Dialectic method

Use Socrates method

- The method invented by Socrates is the dialectic: the art of asking oriented questions to reach the truth.
- Dialectics is not knowledge but the method by which it can be achieved.
- Socrates' method aims to help each one find the truth he is seeking. Aware of being imperfect, **Socrates constantly seeks to improve himself by constantly revisiting his knowledge.**
- It does not rise on the pedestal of the one who knows, it defies daily its own ignorance to be able to advance.



In genetics, a missense mutation is a point mutation in which a single nucleotide change results in a codon that codes for a different amino acid.[1] It is a type of nonsynonymous substitution. Missense mutations can render the resulting protein nonfunctional,[2] and such mutations are responsible for human diseases such as Epidermolysis bullosa, sickle-cell disease, and SOD1 mediated ALS



Not all missense mutations lead to appreciable protein changes.

An amino acid may be replaced by an amino acid of very similar chemical properties, in which case, the protein may still function normally; this is termed a neutral, "quiet", "silent" or conservative mutation.

Alternatively, the amino acid substitution could occur in a region of the protein which does not significantly affect the protein secondary structure or function. When an amino acid may be encoded by more than one codon (so-called "degenerate coding") a mutation in a codon may not produce any change in translation; this would be a silent mutation (a type of synonymous substitution, which is not always silent) and not a missense mutation.

Another type of nonsynonymous substitution is a nonsense mutation in which a codon is changed to a premature stop codon that results in truncation of the resulting protein.

Point mutations – (1) substitutions

Point mutations



In coding region



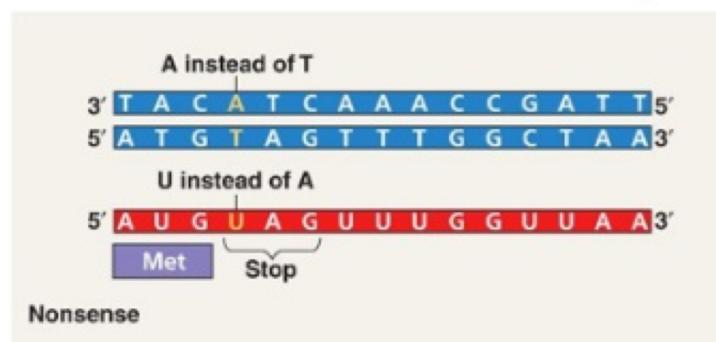
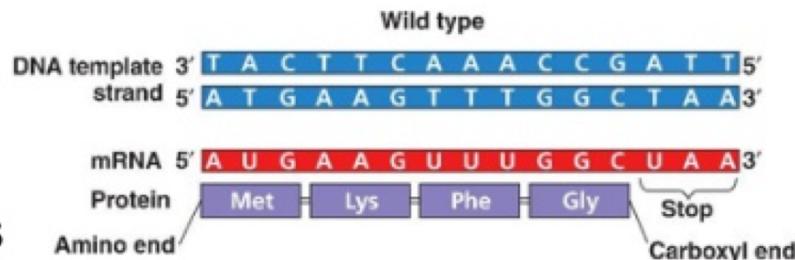
Non-synonymous mutations



Nonsense mutation



Change an amino acid into a stop codon



A Nonsynonymous / Synonymous substitution

TCCGAT	ATATGG	CAA	CCC	GAC	AAA		
S	D	I	W	Q	P	D	K

TCAGAT	CTATGG	CAG	CCC	CAC	AAA		
S	D	L	W	Q	P	R	K

B Radical / Conservative substitution

ATT	GAC	TATTCC	TGT	TGGTTT	GAA	CCAGGC	AGA			
I	D-	Y	S	C ^N	W	F	E-	P	G	R ⁺

ATT	CAC	TACTCC	GGT	TGGTTC	GCA	CCAGGA	AAA			
I	R ⁺	Y	S	G ^N	W	F	A ^N	P	G	K ⁺

- + positive
- negative
- N neutral

La méthode inventée par Socrate est la dialectique : l'art de faire dialoguer deux discours apparemment contradictoires pour accéder à une vérité supérieure. Grâce à un jeu progressif de questions, Socrate fait tomber les fausses connaissances de l'interlocuteur.

La dialectique n'est pas le savoir mais la méthode qui permet d'y parvenir.

Désintéressée et dénuée d'égoïsme, la méthode de Socrate vise à aider chacun à trouver sa propre voie et loi d'action. Conscient d'être imparfait, Socrate cherche sans cesse à se perfectionner. Il ne se hisse pas sur le piédestal de celui qui sait, il défie quotidiennement sa propre ignorance pour pouvoir avancer. Ce faisant, il assume la contradiction apparente qui consiste à mettre en pratique un savoir ou une technique qu'il ne domine pas encore entièrement. Son attitude de vie relève d'une philosophie du risque assumé. Il ne tombe pas dans le piège psychologique d'agir seulement lorsqu'il croit tout savoir et que le risque est nul, ce qui empêche précisément nombre de personnes, paralysées par le syndrome de la perfection, de passer à l'action. Sans confrontation, nul ne peut apprendre ni se perfectionner en quoi que ce soit. Socrate comprend qu'on ne peut jamais être totalement préparé à l'action dans la théorie et que, dans la réalité, ce qui permet d'être prêt est la décision d'agir tout en étant conscient de sa propre imperfection, en apprenant de ce qu'on expérimente, qu'on gagne ou qu'on perde. Le succès selon Socrate réside dans le fait d'avoir le courage et l'intelligence de se confronter à ses peurs et à ses doutes, en les dépassant.

<http://www.sagesse-marseille.com/lhomme-sage/philosophie-dans-la-vie/la-methode-socratique.html>

La relation établie par le dialogue est celle de deux individus qui communiquent entre eux à travers deux consciences pratiquant l'investigation et entrant en relation pour parvenir à