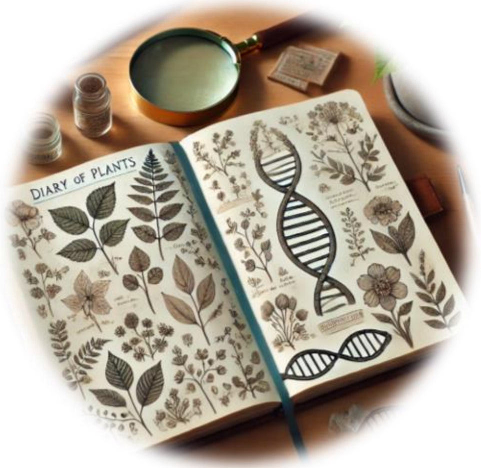
07.05.2025

INSTITUTE
OF PLANT
SCIENCES

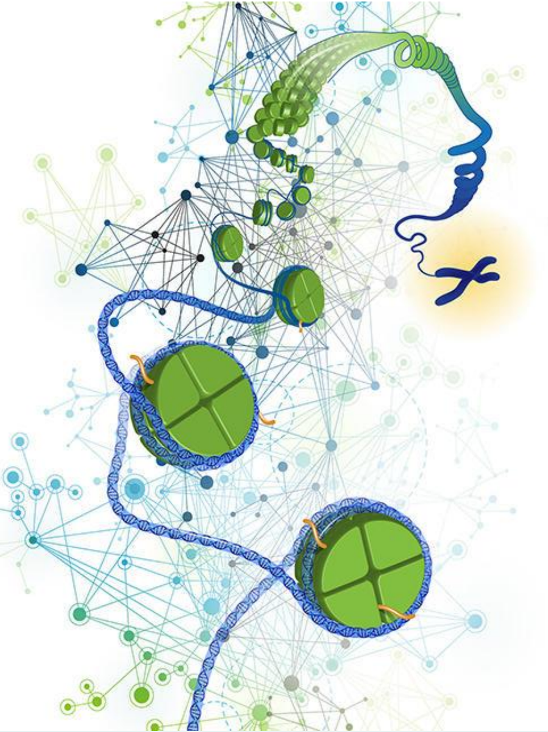
Sant'Anna
School of Advanced Studies – Pisa

# Genomics & Transcriptomics:
## Peaking into the Diary of Plants



Svenja Mager
svenja.mager@santannapisa.it

# ABOUT THIS LECTURE



## Genomics

- ☐ What is “THE GENOME“
- ☐ Structure of Genomes
- ☐ Genomic Features

## Transcriptomics

- ☐ What is a transcriptome
- ☐ Why is it important for living beings
- ☐ Coding and non-coding genes
- ☐ How to study gene expression

# What are Genomics

**An interdisciplinary field of biology studying the many aspects of genomes**

- **Structural genomics:** the study of genome structure and organization
- **Functional genomics:** the study of gene expression (transcriptomics) and proteins (proteomics) and their functions and interactions
- **Comparative genomics:** studying the genome structure and function across different species and the evolution of genomes
- **Metagenomics:** studies environmental samples containing genetic material from several individuals and species
- **Pangenomics:** determines the entirety of genes and genomes withing a certain species

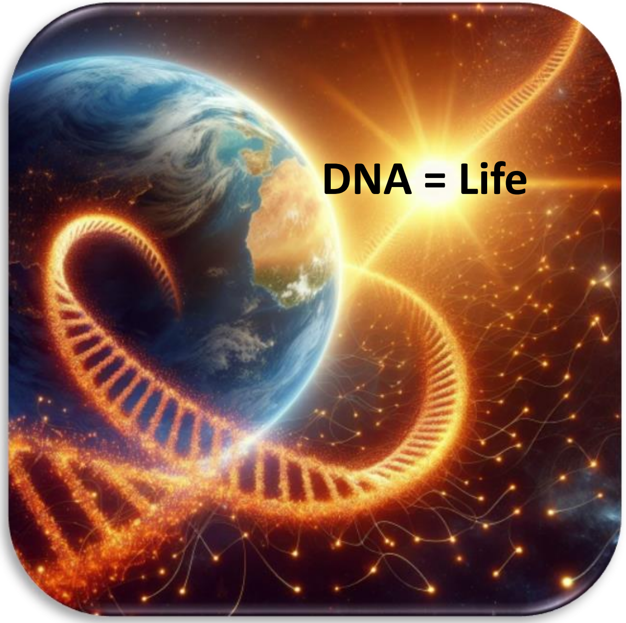
… and others ☺

# What is a genome anyway?

The entirety of genetic material in a **cell**

Sometimes more broadly referred to the genetic make-up of a whole organism

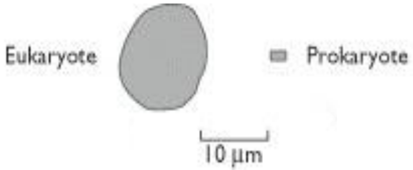
~ 37 000 000 000 000 Cells

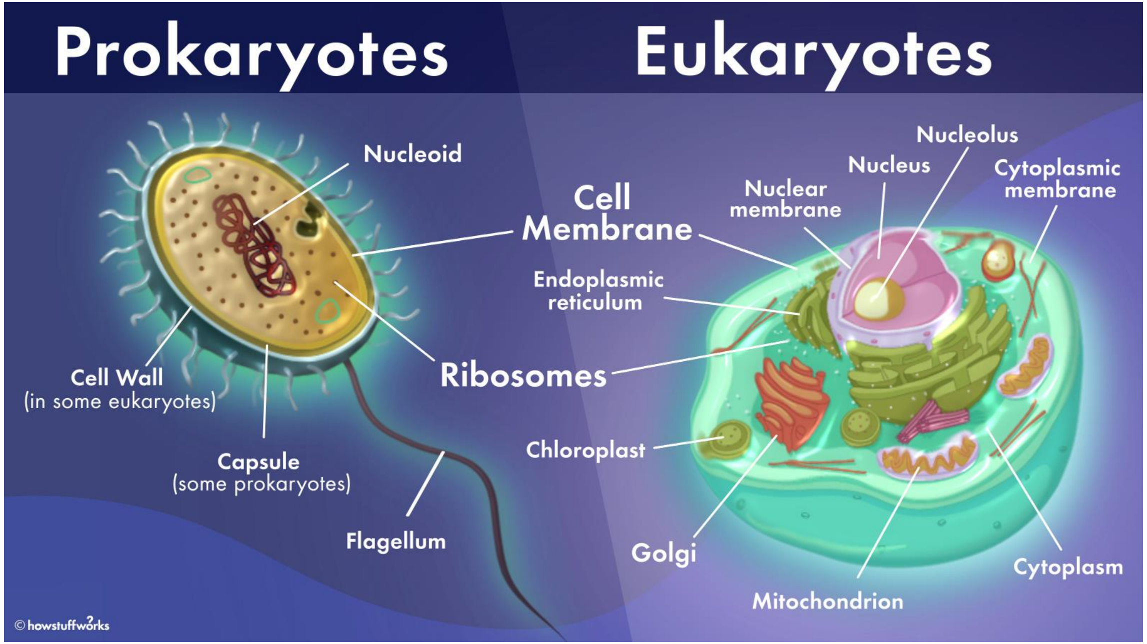


The basis in terms of functionality of an organism

# Physical organization of the genome

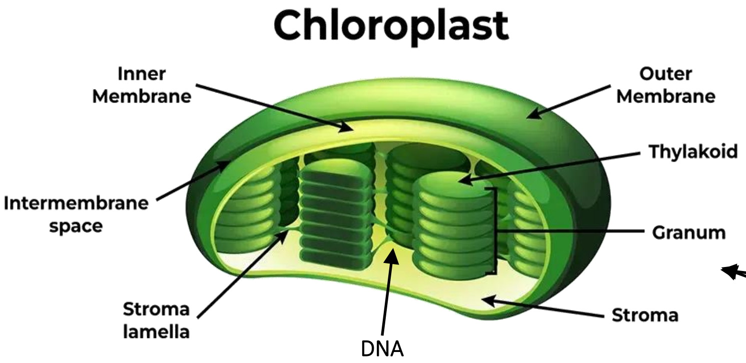
Different in prokaryotes and eukaryotes

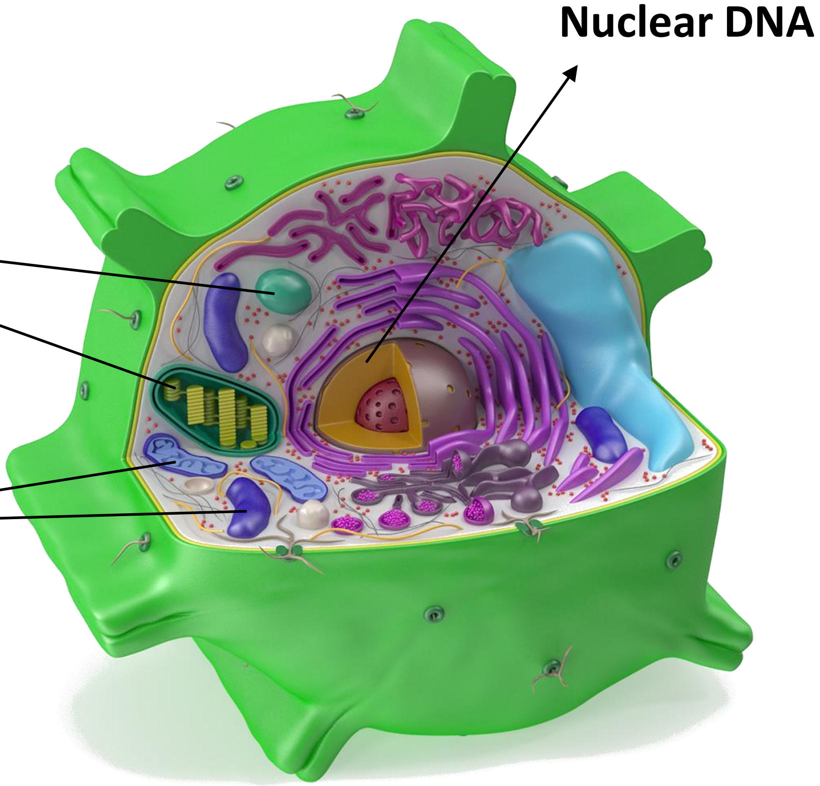


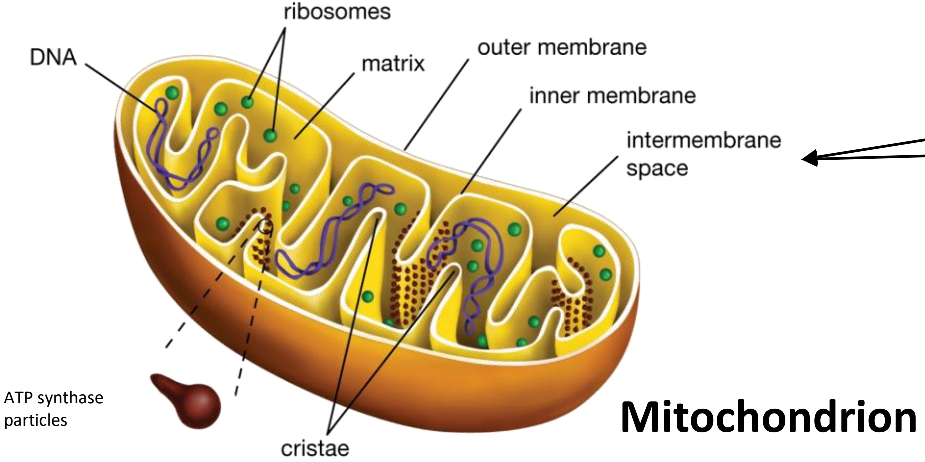


There can be more than one genome in a cell…

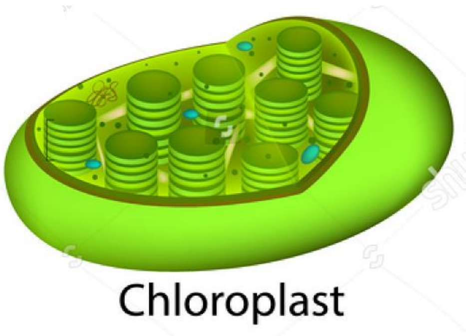
# Genomes in eukaryotic cells

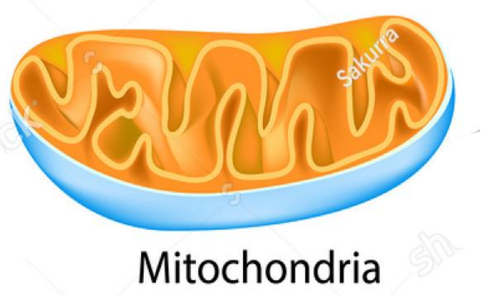






# Chloroplasts and Mitochondria



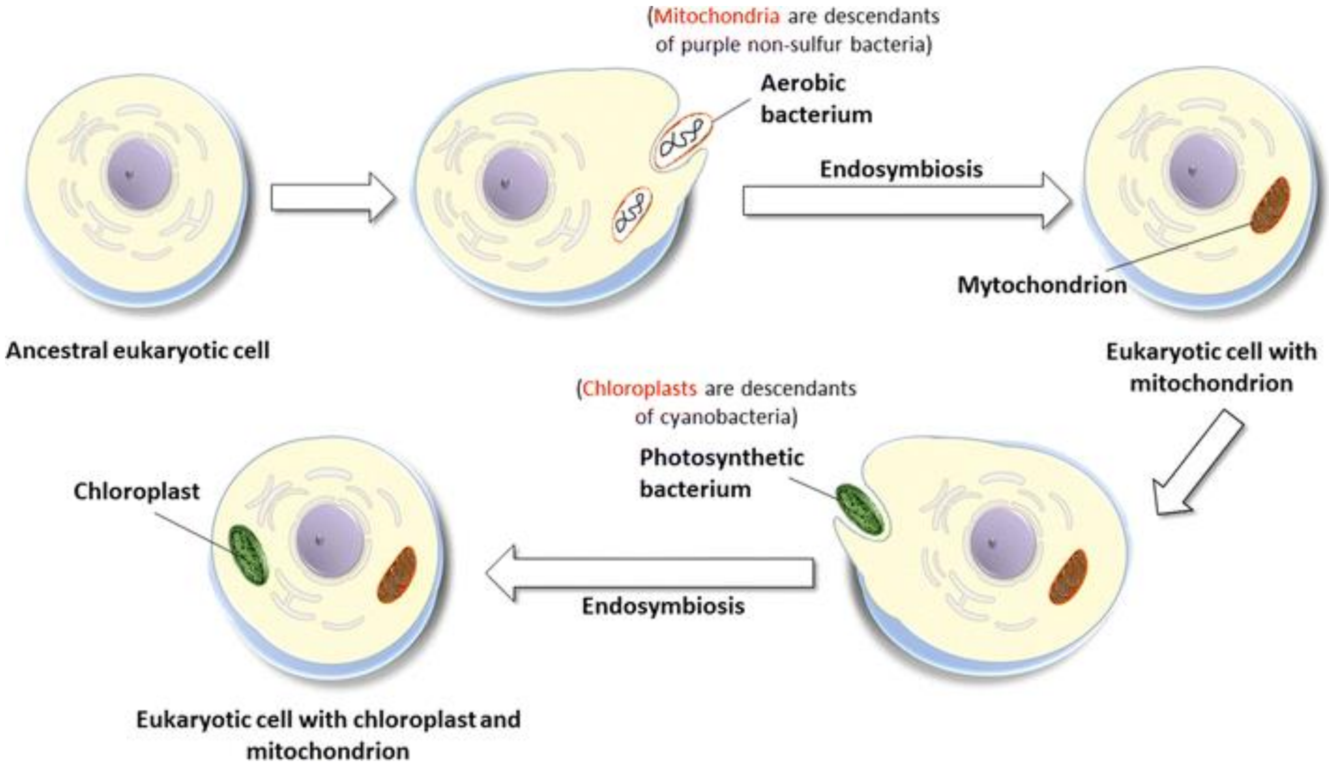


**Photosynthesis:**
Converts the sun's light energy to energy-rich glucose and oxygen (making us breath, yayyy ☺)

**Cellular respiration:**
Breaks down glucose (from food or photosynthesis) to release energy (ATP)

# Chloroplasts and Mitochondria

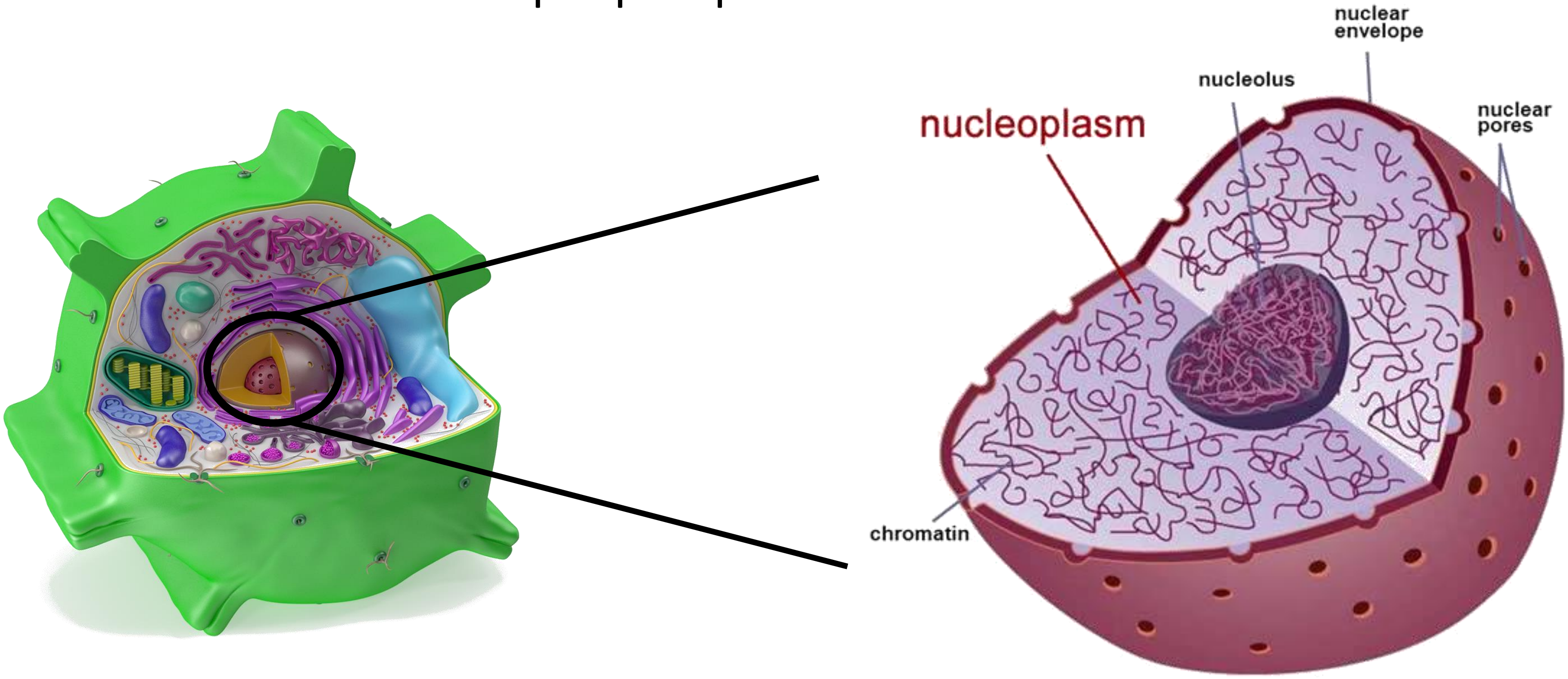
## Mitochodria and Chloroplasts may have originated as endosymbionts



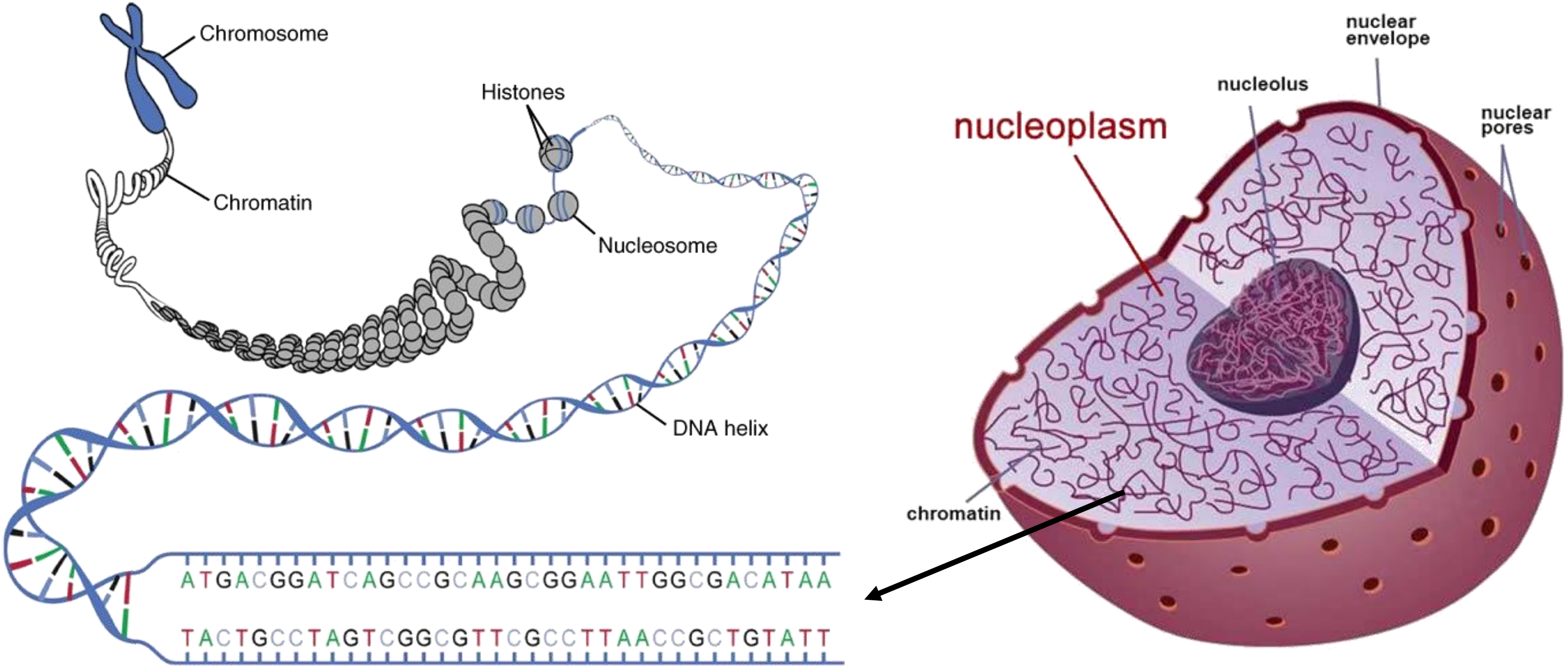
- Gene expression processes in organelles are similar to equivalent processes in bacteria
- Organelle genes are more similar to bacterial genes than to those in the host cell

# Nuclear DNA

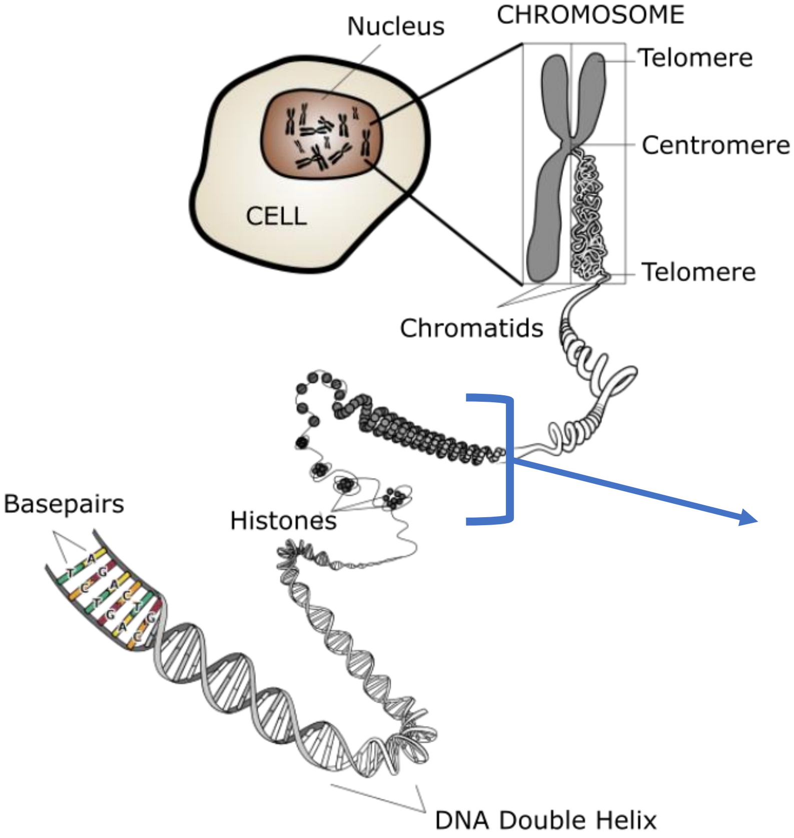
Often referred to when people speak of «the Genome»

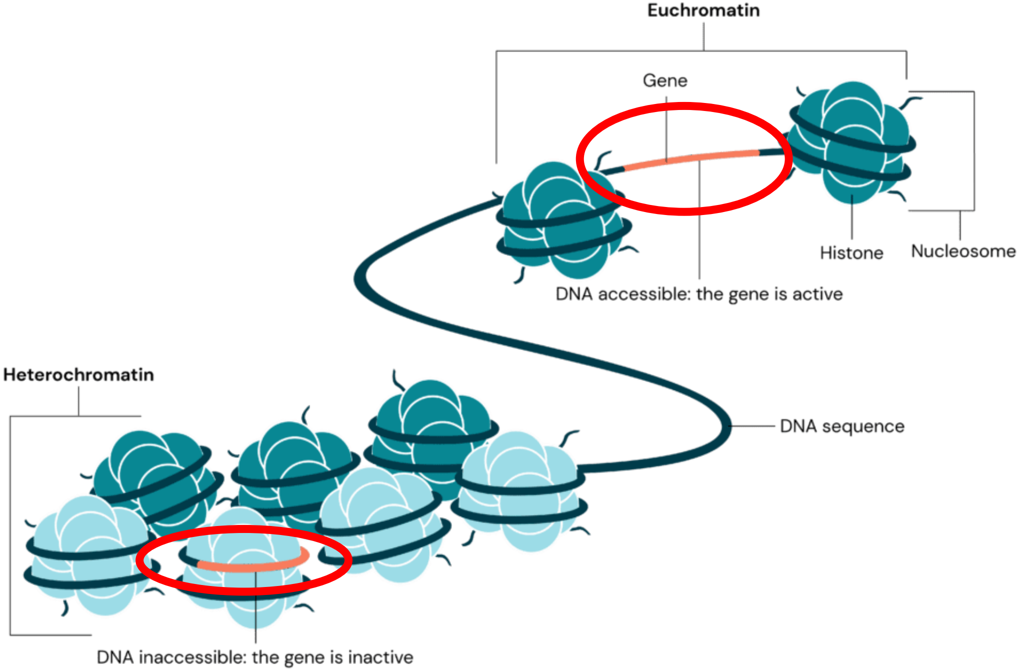


# Nuclear DNA


Chromosome
Histones
Chromatin
Nucleosome
DNA helix
ATGACGGATCAGCCGCAAGCGGAATTGGCGACATAA
TACTGCCTAGTCGGCGTTCGCCTTAACCGCTGTATT
nuclear envelope
nucleolus
nucleoplasm
nuclear pores
chromatin

# Nuclear DNA

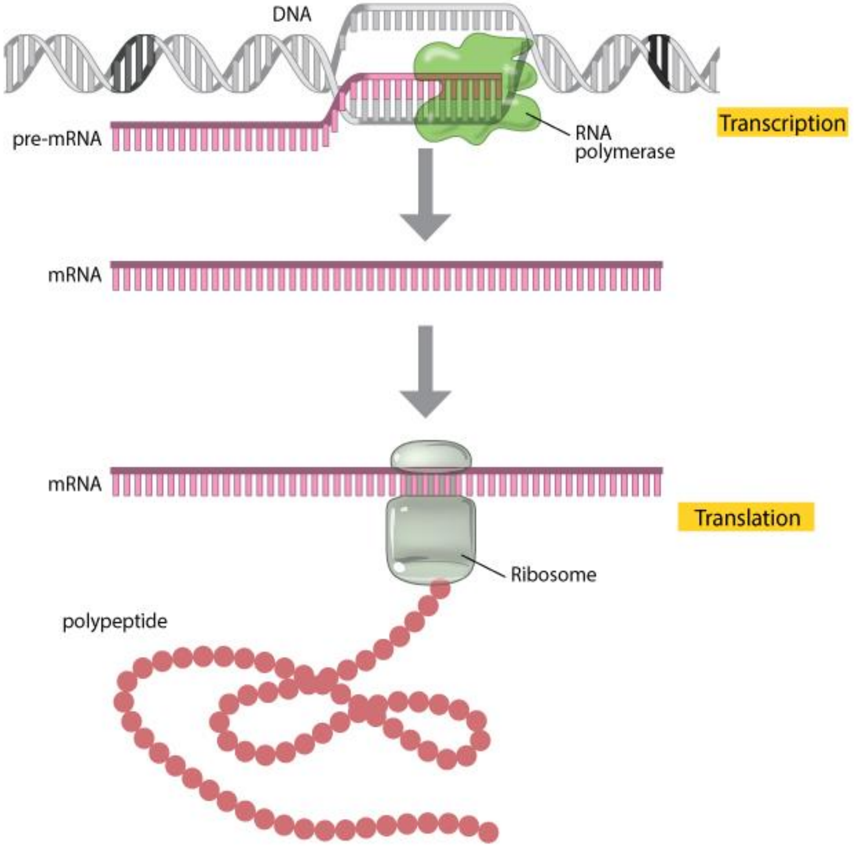

Nucleus
CHROMOSOME
Telomere
Centromere
CELL
Telomere
Chromatids
Basepairs
Histones
DNA Double Helix


Euchromatin
Gene
Histone
Nucleosome
DNA accessible: the gene is active
Heterochromatin
DNA sequence
DNA inaccessible: the gene is inactive

# Genomic Components - Coding genes

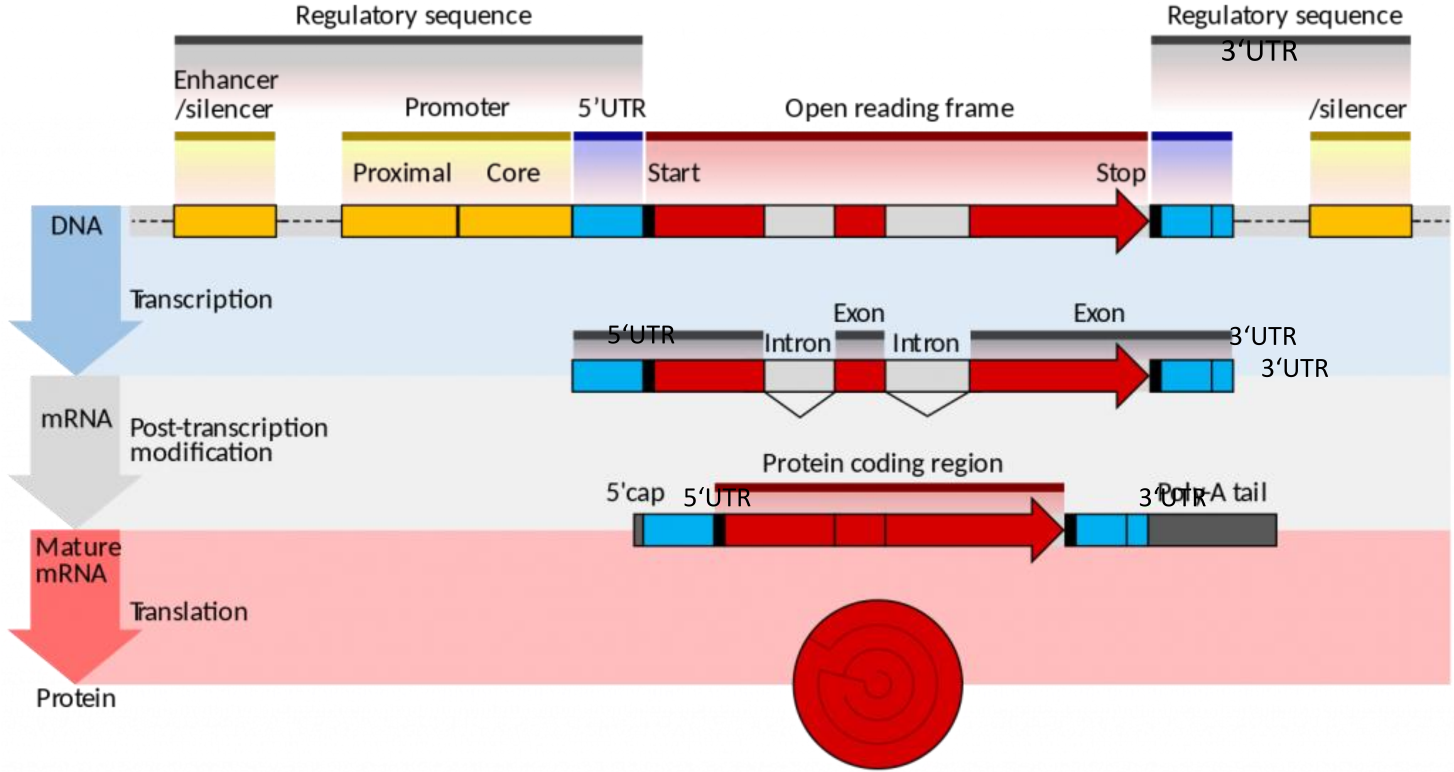
A **coding gene** is a DNA sequence that contains information that is later translated to **proteins** (through the intermediate of RNA)

**Proteins** are molecules composed of amino acids required for the structure, function, and regulation of the body's tissues and organs



# Genomic Components - Coding genes

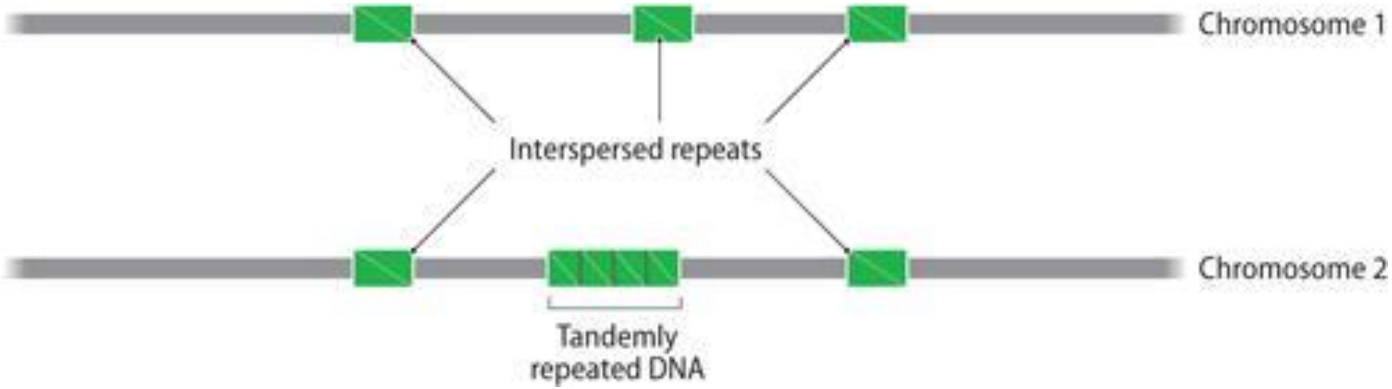
In eukaryotes, the entire gene is transcribed in mRNA but exons are the only components ending up in proteins



# Genomic Components - Repetitive DNA

**Two main types of repetitive DNA:**

- **Tandem repeats:** DNA sequence motifs repeated directly adjacent to each other in the genome
- **Interspersed repeats:** identical or similar DNA sequences found in different locations scattered across the genome



# Genomic Components - Repetitive DNA

## Tandem repeats

- **Microsatellites**: sequence motifs of 1-10 nucleotide length that are repeated 5-50 times
  (e.g. GTC GTC GTC GTC GTC or TATATATATATATATATATATATA )

- **Minisatellites**: sequence motifs 11-100 nucleotide length repeated 2 to several hundred times

# Genomic Components - Repetitive DNA

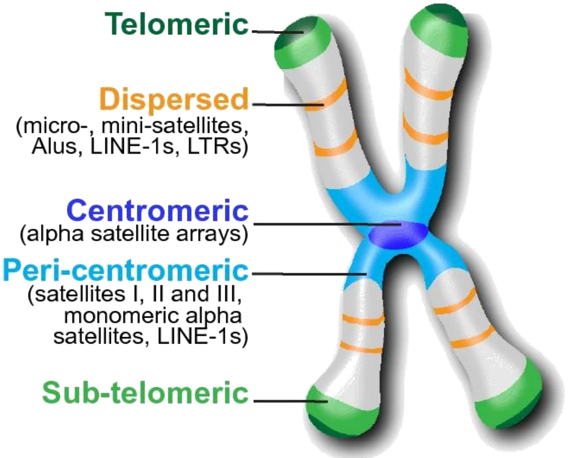
## Tandem repeats

**Functions:**

- When located in promoter or regulatory regions influence gene regulation
- In telomeres and centromeres aid in chromosome integrity and stability and chromosome pairing
- Form recombination hotspots

**Usage:**

- Genetic Markers: In non-coding regions selection pressure might be absent and allow more frequent mutations → useful to study diversity and evolution in a population/species



# Genomic Components - Repetitive DNA

## Interspersed repeats: Transposable elements (TEs)

Transposable elements are the main group of interspersed repeats

Sequences of DNA that move (or jump) from one location in the genome to another

Typically, between 100 to 10,000 base pairs in length, but sometimes far larger

In practice, TE families are usually defined using the “80-80-80” rule, which specifies that insertions are members of the same family if they are longer than 80 base pairs, and share at least 80% sequence identity over 80% of their length

# Genomic Components - Repetitive DNA

## Interspersed repeats: Transposable elements (TEs)

Rightly so?

– Yes and No!

Infamously known as junk DNA (Ohno, 1972), selfish DNA (Dawkins , 1976), parasitic DNA (Orgel and Crick, 1980

TEs are mutagens that can damage cells in several ways

- Extensive source of mutations and polymorphisms
- Modification of gene structure and regulatory regions by TE insertion
  - TE jumping into a gene can disrupt the functioning of the gene
  - TE being cut out of a gene might disrupt the functioning of the gene
- Accumulations of the same sequence can hinder precise chromosomal pairing → unequal crossovers
- Expression of damaging proteins

*So shouldn't I get rid of TEs???*

# Genomic Components - Repetitive DNA

## Interspersed repeats: Transposable elements (TEs)

**TEs are strictly regulated!**

→ They have evolved self-regulatory mechanisms controlling their own copy numbers.
→ A variety of host factors are employed to control TE expression (small RNAs, chromatin and DNA modification pathways , sequence-specific repressors)

**And: TEs have important benefits, too!**

Transposons can drive evolution and play an essential part in gene expression regulation

# Genomic Components - Repetitive DNA

## Interspersed repeats: Transposable elements (TEs)

### Evolution

It is very rare that mutations, indels or duplications have positive effects, but it does happen, driving adaption and providing a selective advantage → Evolution ☺

TEs Contribute to exon shuffling, a mechanism for forming new genes, by taking adjacent bp with them when jumping

### Gene Expression Regulation

Many TEs contain enhancer, promoter, or silencer sequences that influence nearby genes.

Might contribute to stress responses but TEs and stress is a complex system, not clear yet

# Genomic Components - Repetitive DNA

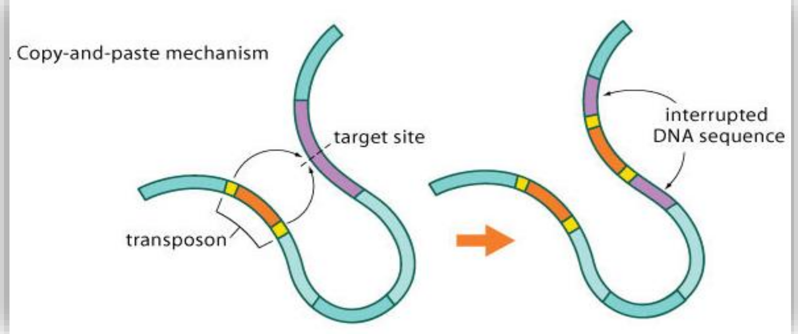
## Interspersed repeats: Transposable elements (TEs)

**Two main types**

**Class I: Retrotransposons**

“copy-and-paste mechanism”

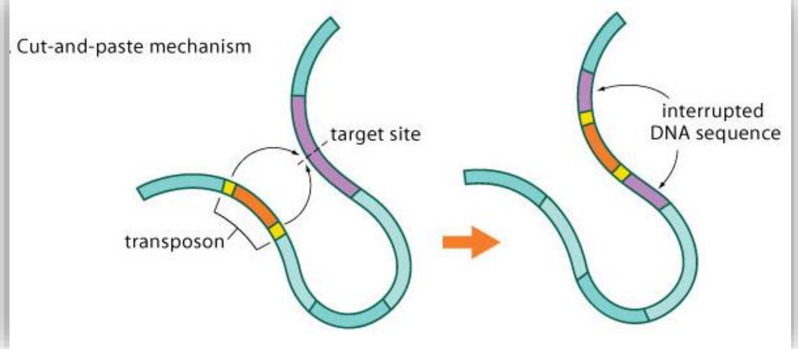
→ they copy themselves in RNA, then back to DNA, then are inserted back in the genome



**Class II: DNA transposons**

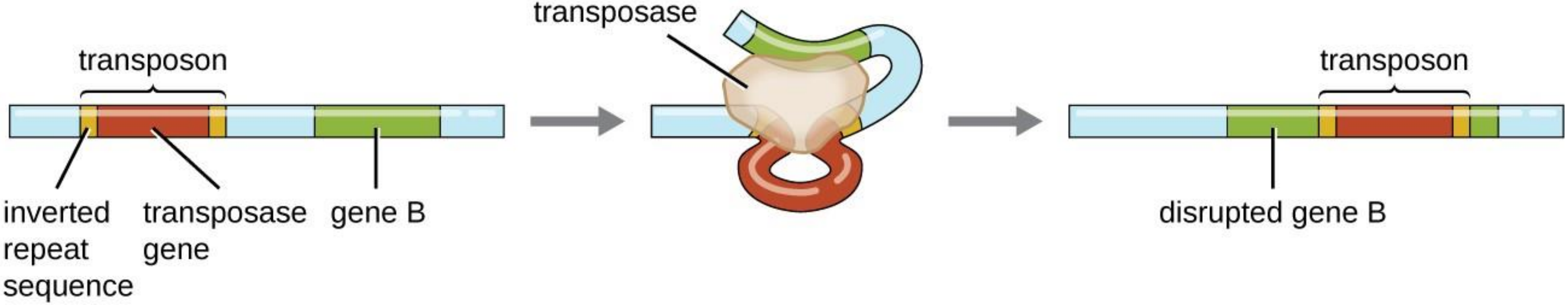
“cut-and-paste mechanism”

→ they are excised from DNA and reintegrated in another location



First discovered in maize by Barbara McClintock, Nobel Prize for Physiology and Medicine 1983

# Transposons can <u>disrupt</u> gene function



① A typical transposon encodes the enzyme transposase, surrounded by inverted repeat sequences.

② Transposase facilitates recombination between inverted repeats. Transposon is cut from its original location and inserted into a new location.

# Transposons can disrupt gene function

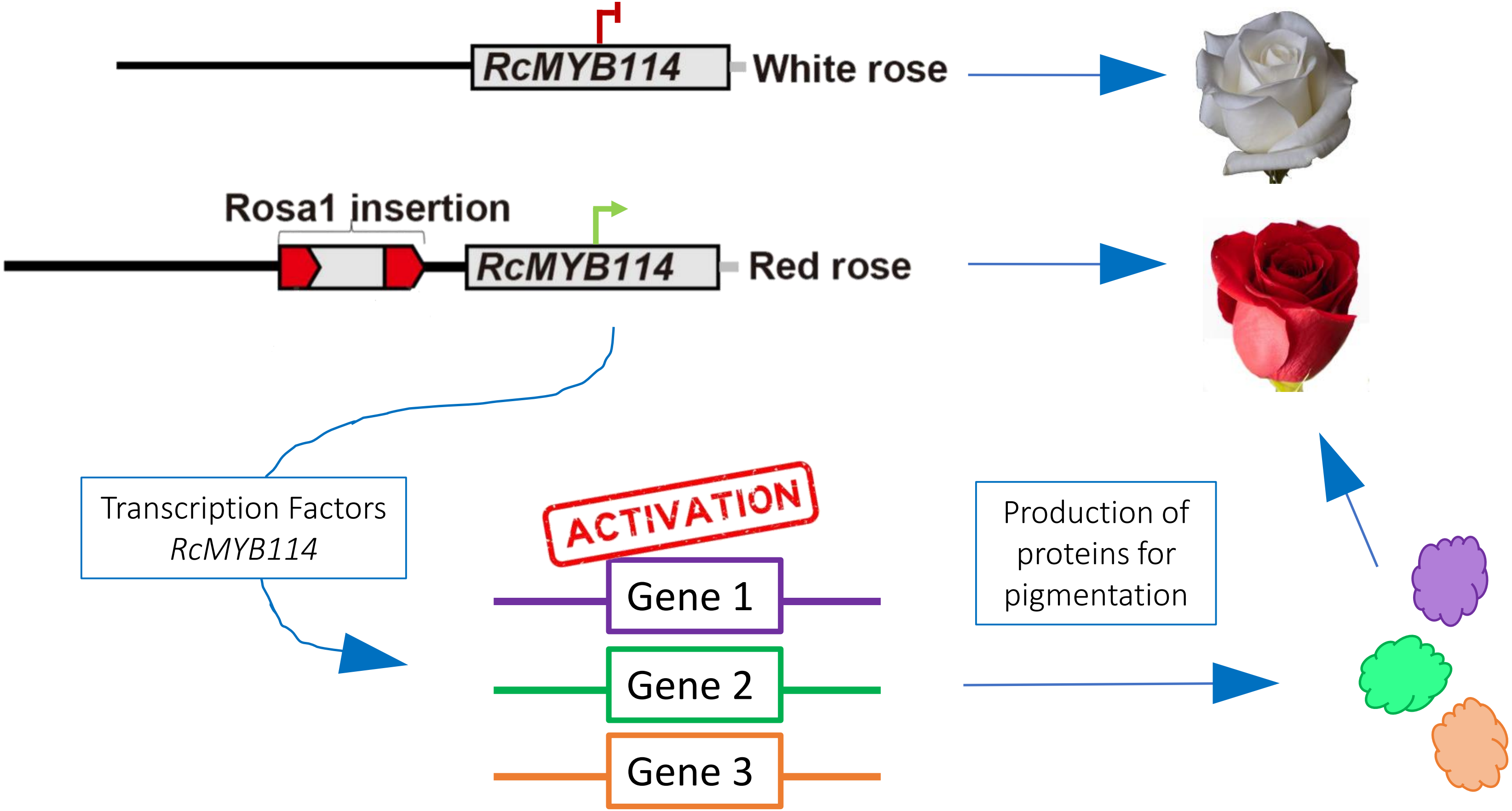
**This can lead to visible effects:**

If a pigment gene in maize is disrupted **in some cells** by a transposable element, the kernels in a grain can have beautiful patterns

Depending on the position of the TE insertion and the duration of the interruption, kernels are more or less pigmented


© W.P. Armstrong 2000

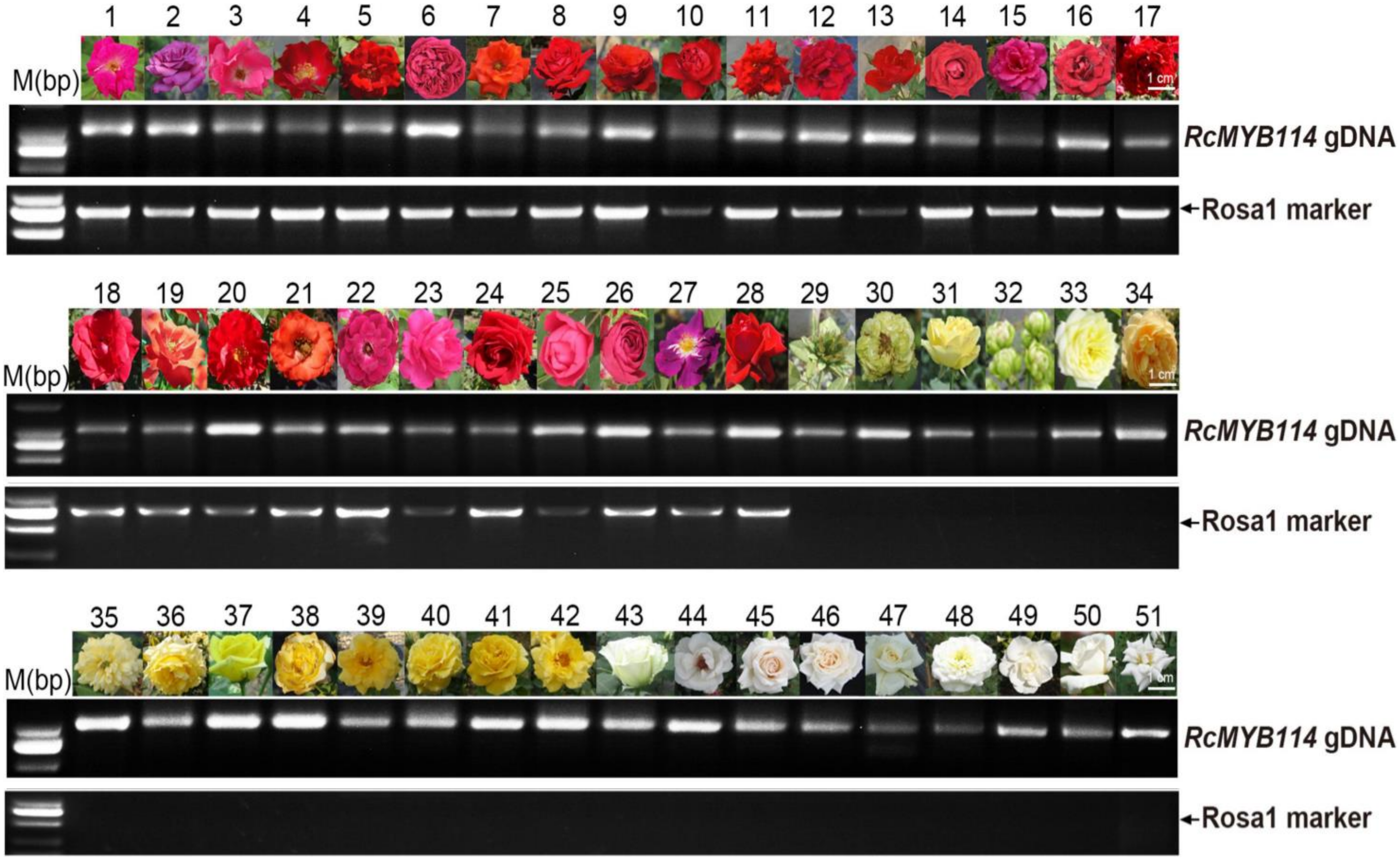
# Transposons can enhance gene function



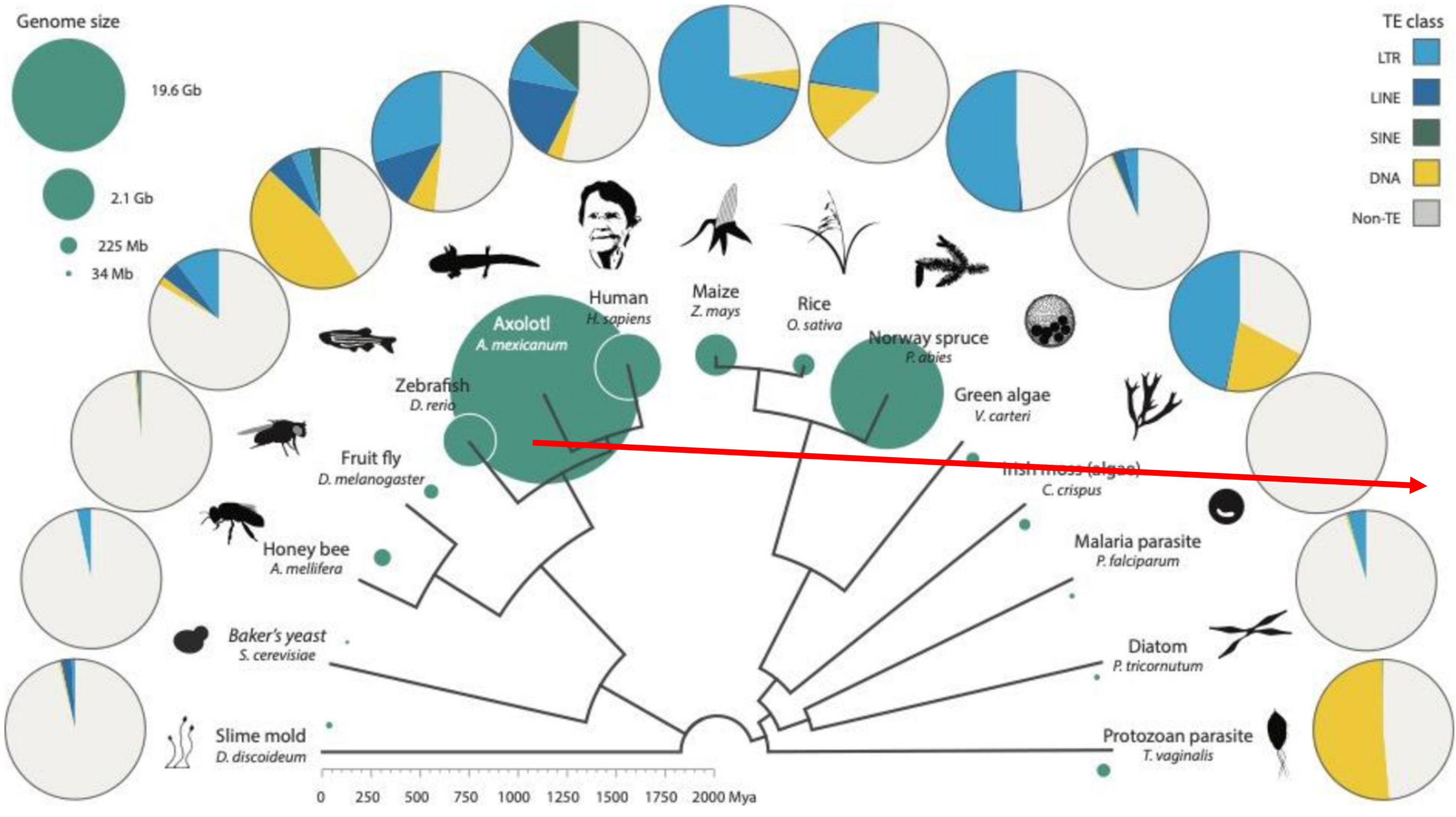
Transcription Factors
*RcMYB114*

Production of proteins for pigmentation

# Transposons can <u>enhance</u> gene function

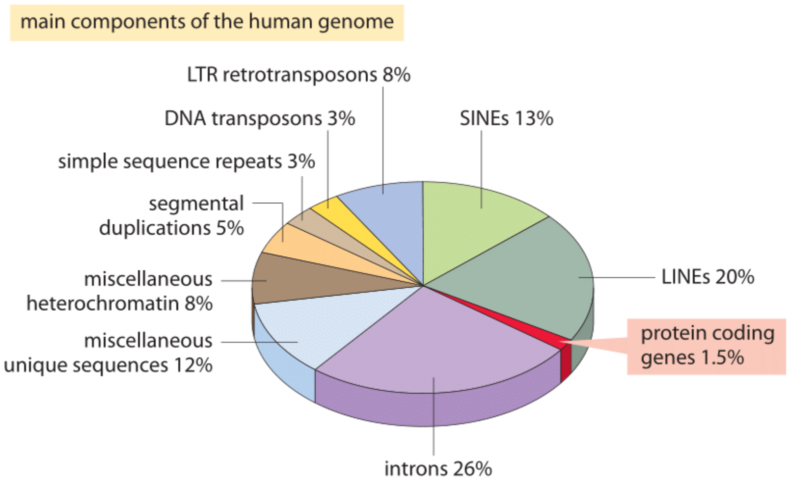


# Genomic Components - Repetitive DNA



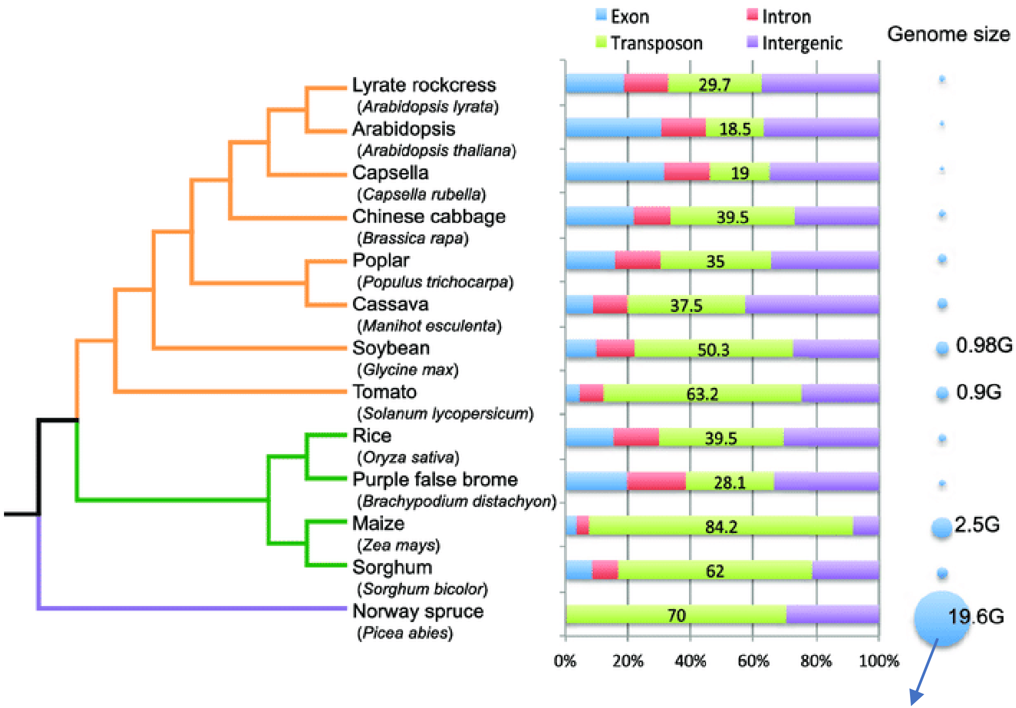
**Axolotl - 32GB genome**
**Largest ever sequenced!**
About the same number of genes as humans

# Let’s get back to genes…



main components of the human genome

3.2 Gb,
~20 000 genes

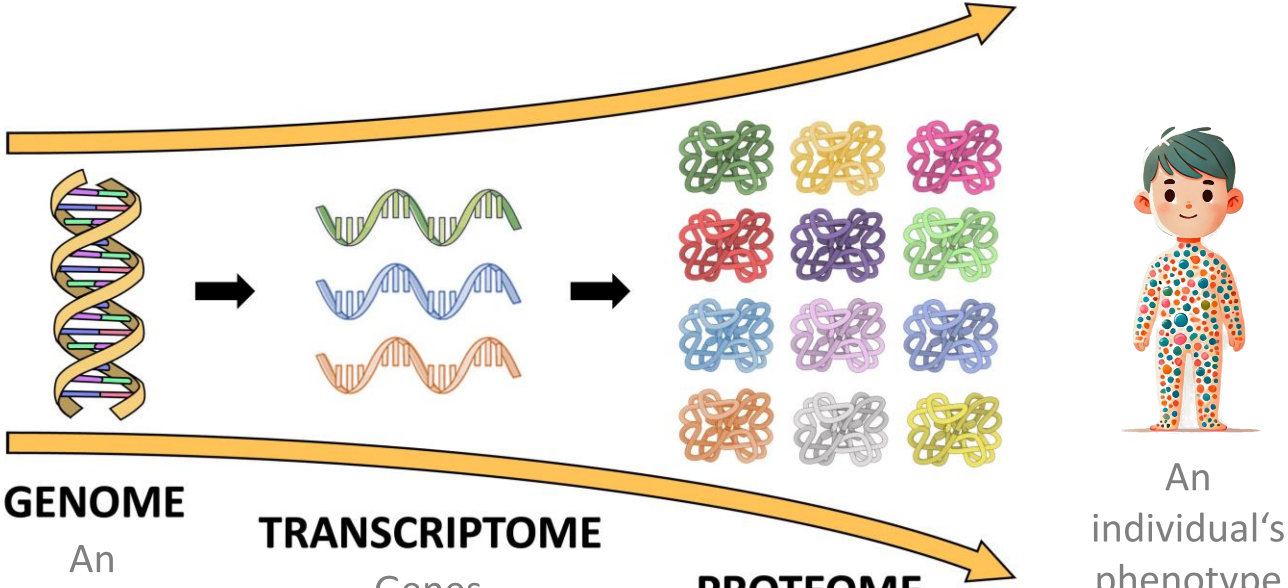


About the same number of genes as humans

# … and finally the transcriptome

**Transcriptome:**
The complete set of transcribed genes at a certain time point in a cell/tissue/organism

**Proteome:**
The complete set of proteins present at a certain time point in a cell/tissue/organism



**GENOME**
An individual‘s genotype

**TRANSCRIPTOME**
Genes transcribed to mRNAs

**PROTEOME**
mRNAs translated to proteins

An individual‘s phenotype

# The genome is “static” - more or less...

## The genome is the same in all cells and tissues of an organism

Exceptions include mutations and mitotic recombination.

→ if heritable, they are a characteristic of evolution

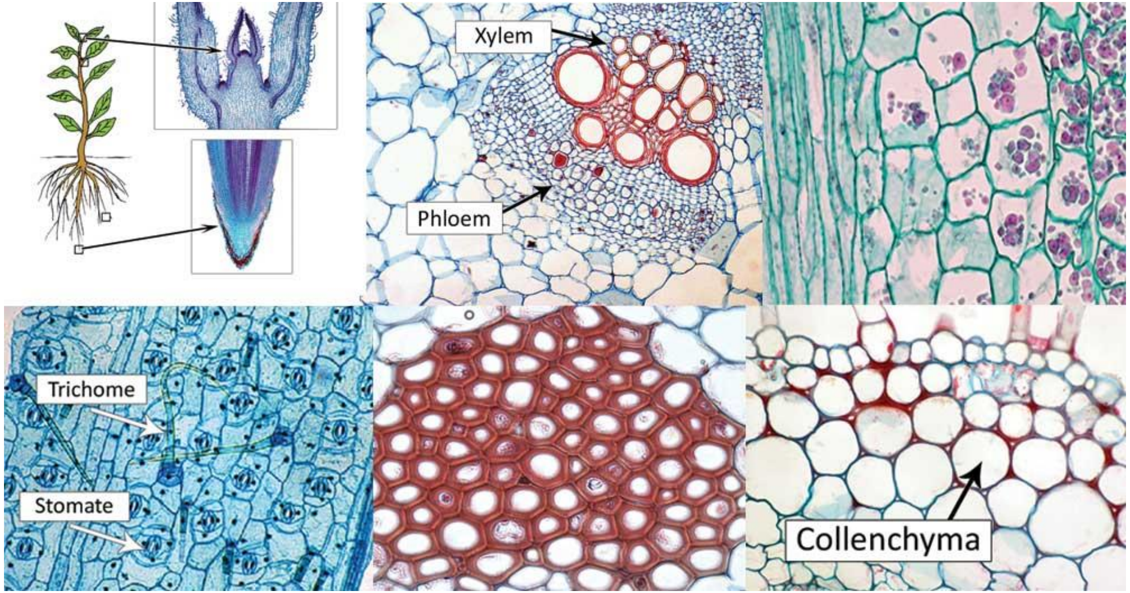
## ...and the transcriptome?

The transcriptome varies between cells and tissues!

# The Transcriptome (and Proteome) vary

## Example: Plants

**Typical Plant =**
**Roots, Stem, Leaves**



**Several cell types ….**

- Parenchyma cells
- Collenchyma cells
- Sclerenchyma cells
- Xylem cells
- Phloem cells
- Meristematic cells
- Epidermal cells

**… building tissues**

- Dermal tissue
- Vascular tissue
- Ground tissue

**… and organs**

- Roots
- Stems
- Leaves

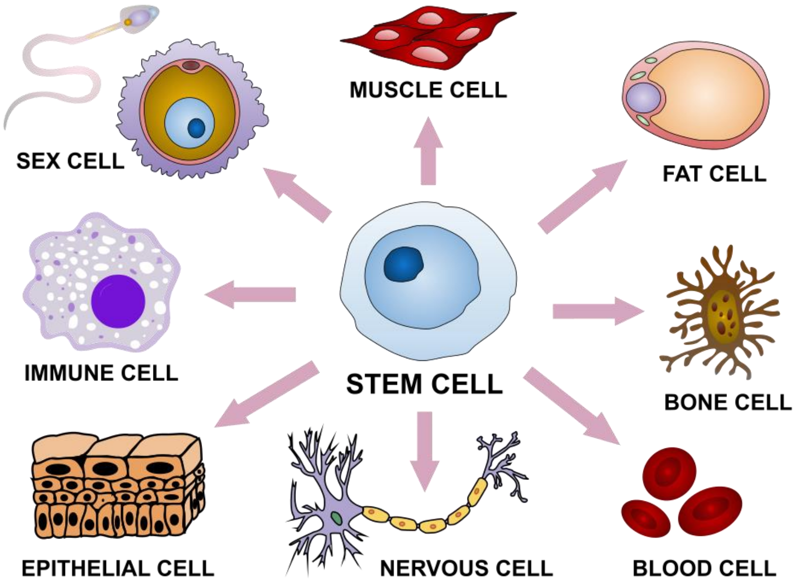
# How do cells know what to be?

"Who am I?
What to do with my life?"

## Cell differentiation

Differentiation changes a cell's size, shape, membrane potential, metabolic activity, and responsiveness to signals.

**No change in DNA sequence involved.**
**Since each cell possesses the same genome, cell types must be regulated differently**



Dolly was cloned
from a mature cell

# How do cells know what to be?

"Who am I?
What to do with my life?"

## Cell differentiation

**Highly complex system of gene expression regulation networks**

Regulation by:

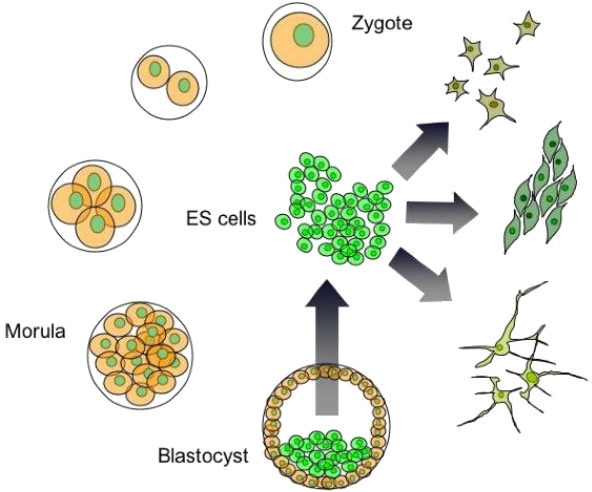
- **cell-extrinsic factors**

→ proteins secreted by other cells, temperature, oxygen etc.

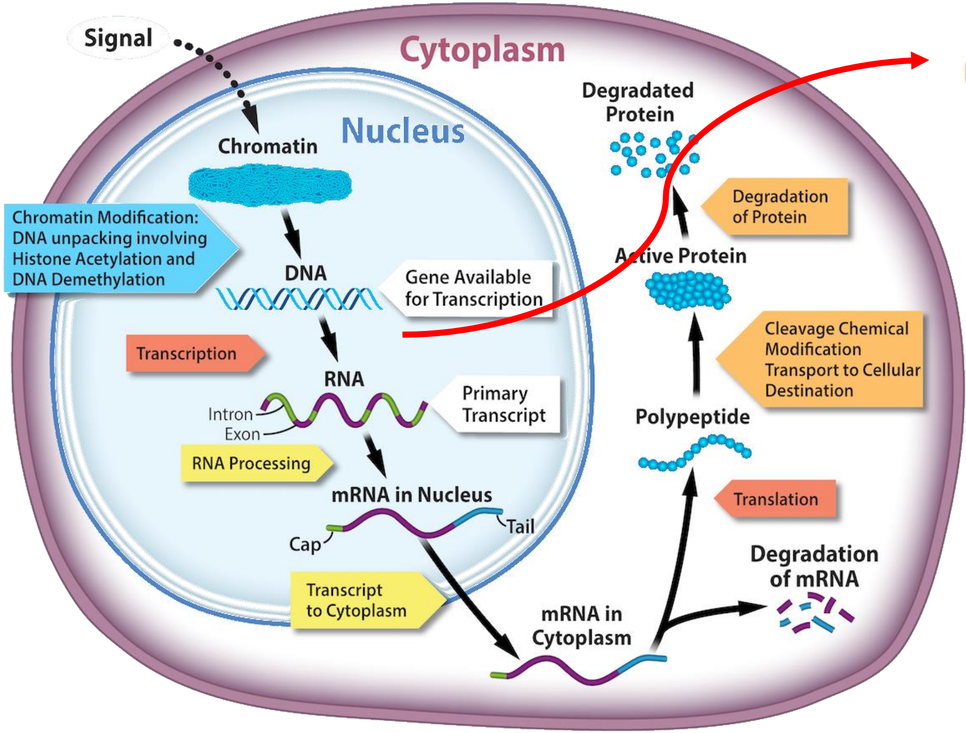
and

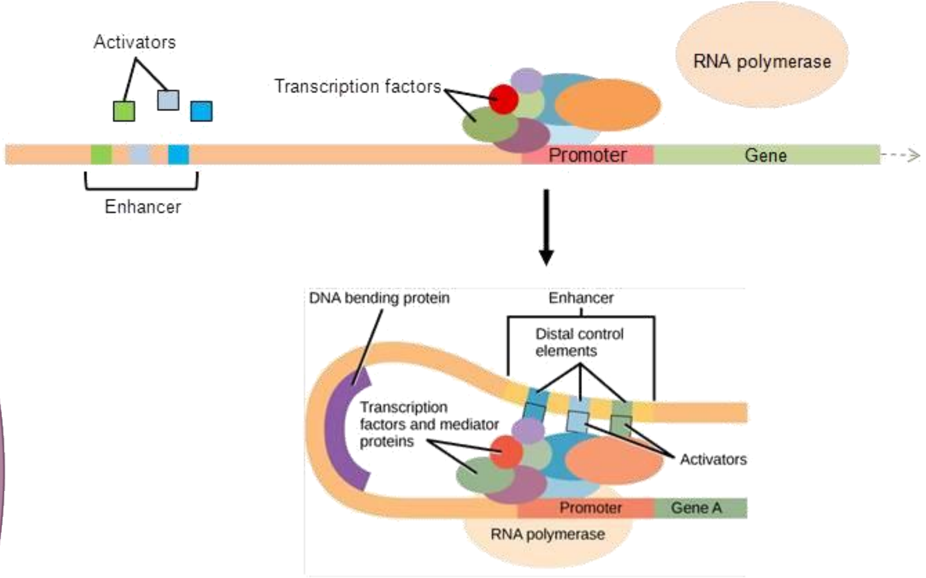
- **cell-intrinsic factors**

→ transcription factors, chemical modification of DNA and histones, chromatin remodeling (epigenetics)



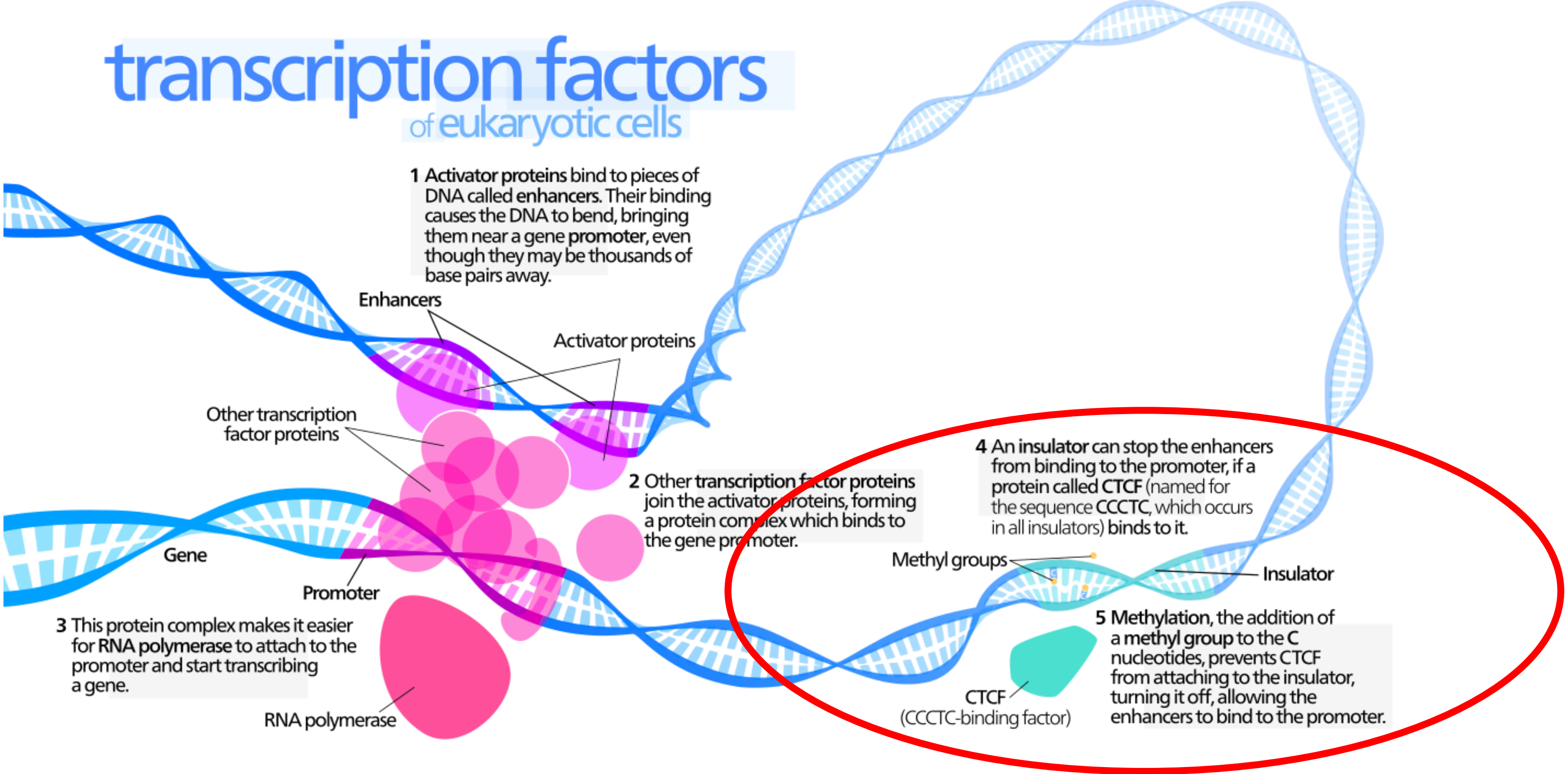
# Gene Expression Regulation





But there are also insulators and inhibitors…

# Gene Expression Regulation



## transcription factors of eukaryotic cells

**1 Activator proteins** bind to pieces of DNA called **enhancers**. Their binding causes the DNA to bend, bringing them near a gene **promoter**, even though they may be thousands of base pairs away.

**2** Other **transcription factor proteins** join the activator proteins, forming a protein complex which binds to the gene promoter.

**3** This protein complex makes it easier for **RNA polymerase** to attach to the promoter and start transcribing a gene.

**4** An **insulator** can stop the enhancers from binding to the promoter, if a **protein called CTCF** (named for the sequence **CCCTC**, which occurs in all insulators) **binds to it.**

**5 Methylation**, the addition of a **methyl group** to the C nucleotides, prevents CTCF from attaching to the insulator, turning it off, allowing the enhancers to bind to the promoter.

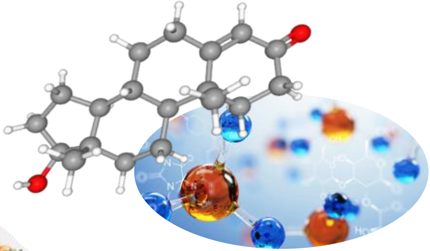
# Cell differentiation - and what else?

## Phenotypic Plasticity

Gene expression regulation processes are not only responsible for the general morphology (**phenotype**) of an organism but also aids in **Phenotypic Plasticity**

→ the ability of an individual’s **genotype** to express different phenotypes in adaptation to varying environmental conditions

Most widespread (but not only) in basically immobile organisms that cannot move away from adverse conditions → plants



Changes induced by phenotypic plasticity include

- Morphological, Physiological, behavioral

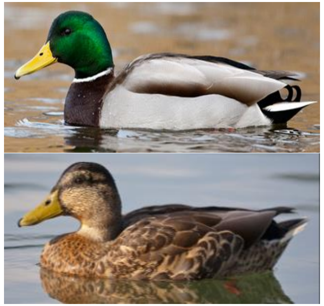
Environmental stimulants can be:

- Seasonal changes, Chemicals (e.g. hormones), Diet

# Phenotypic Plasticity Examples

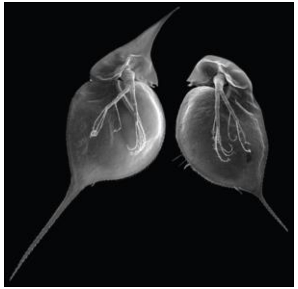
## Seasonal stimulants

Arctic Fox **c**amouflage in Summer and Winter



Male Mallard during mating season and rest of the year

## Chemical cues



Water fleas produce spines and helmets when sensing predator-released chemicals

## Diet

Food determines a bee's fate!

- Royal Jelly diet (secretion from nurse bee glands) →

- 'Beebread' diet (pollen and honey) →

# How can we study all this? OMICS!

All
**Genes**

Genomics

-TATCGATATTGGAAAGG-
-ATTTCAACTCGGGAAAA-
-ATTCGATCTCGGGAAAA-
-ATTCGATCTCGGGAAAA-
-ATTCGATCTCGGGAAAA-

Genome sequence

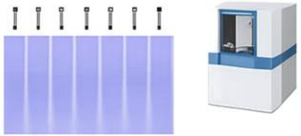
All
**mRNAs**

Transcriptomics

Microarray
RNAseq

All
**Proteins**

Proteomics



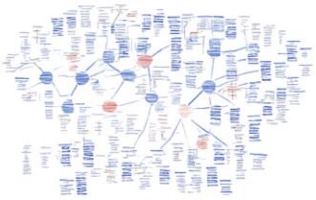
Protein profile

All
**Metabolites**

Metabolomics

Hormonomics

Ionomics



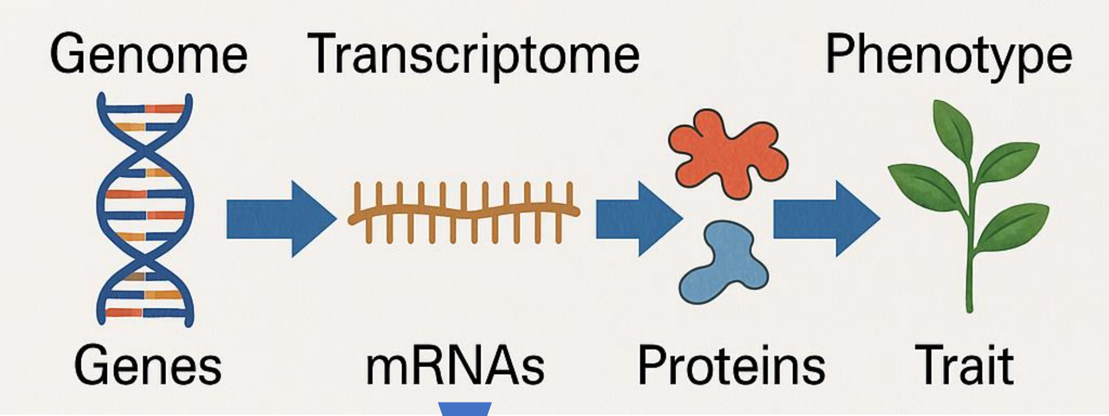
Metabolic map

All
**Phenotypes**

Phenomics

# Let's start with just one: Transcriptomics

**Transcriptome:**
The complete set of **transcribed genes** at a certain time point in a cell/tissue/organism



and **non-coding RNAs!**

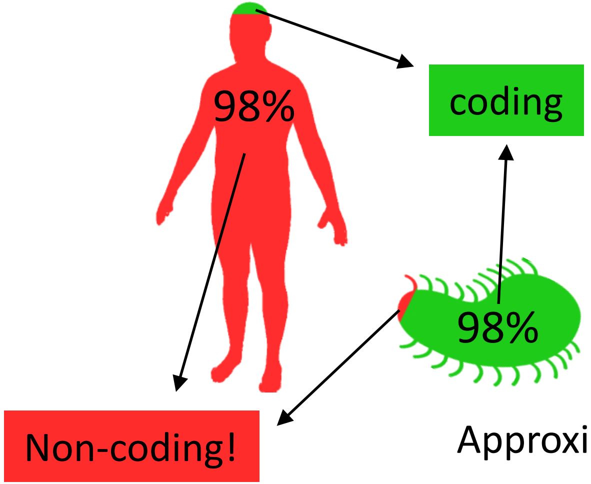
# Transcriptomics - coding and non-coding RNA

Transcriptomics studies often focus on **coding DNA** (translated to proteins), but:

→ only about 2-3% of eukaryote genomes is coding DNA!

The rest was considered “Junk” in the 60’s and 70’s

But: 80-90% of eukaryote genome transcribed at some point → **non-coding DNA**

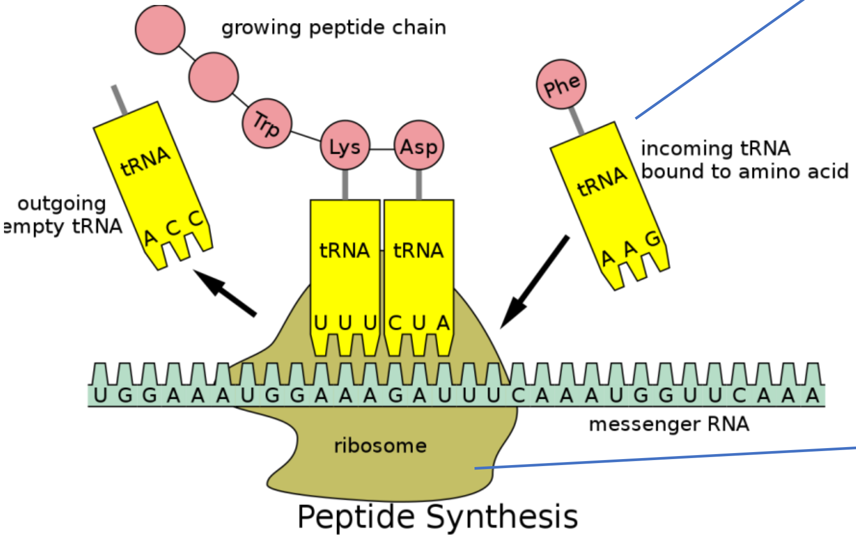


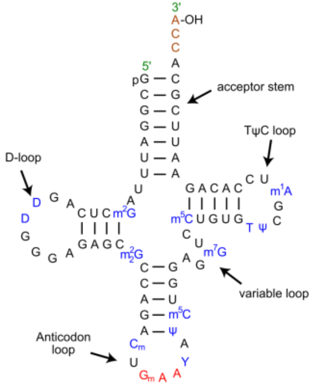
Approximation, ranges among species

# House-keeping non-coding RNAs

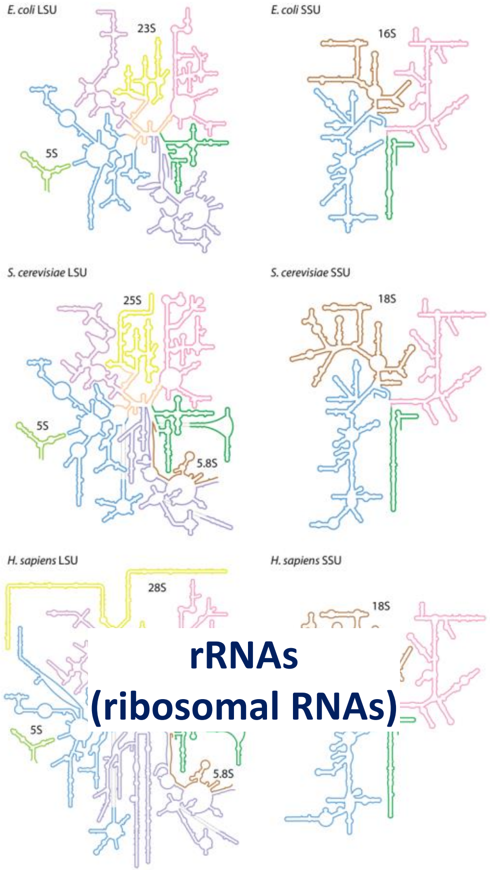
Housekeeping: required all the time for the maintenance of basic cellular function

**TRANSLATION**





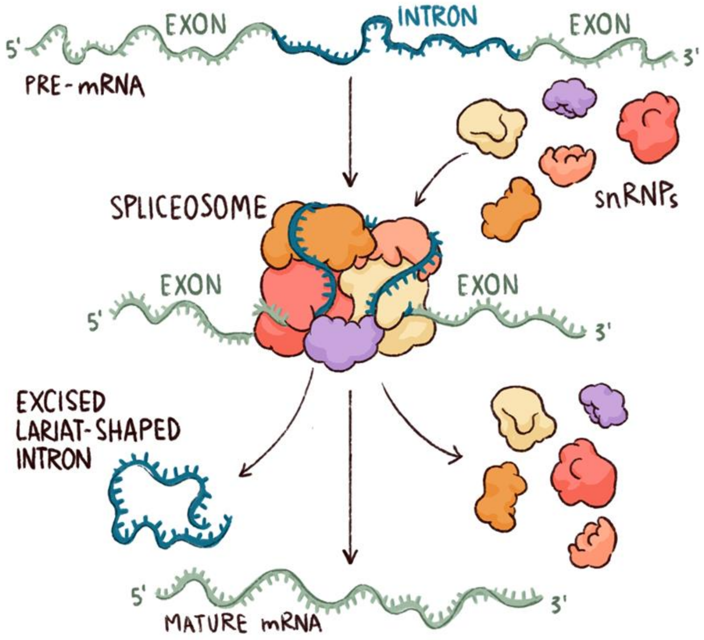
**tRNAs**
**(transfer RNAs)**



**rRNAs**
**(ribosomal RNAs)**

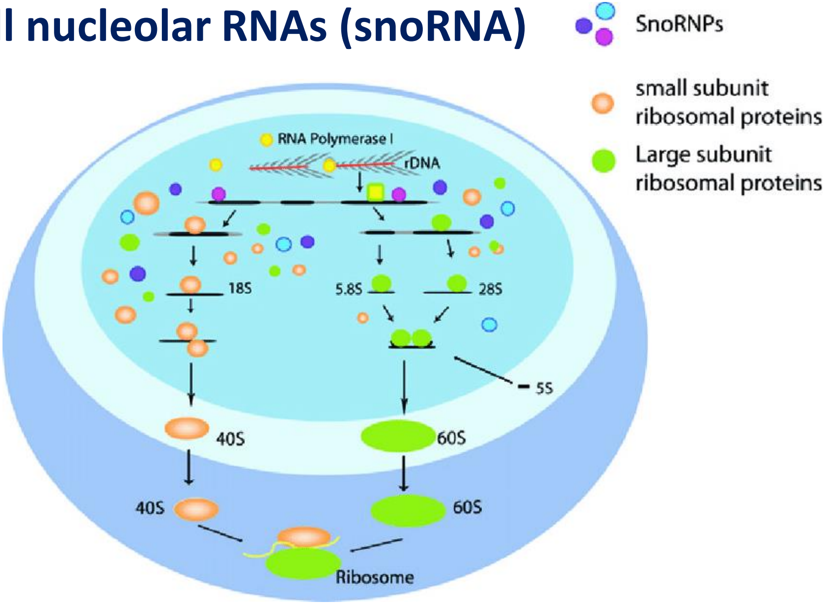
# House-keeping non-coding RNAs

## Small nuclear RNA (snRNA)



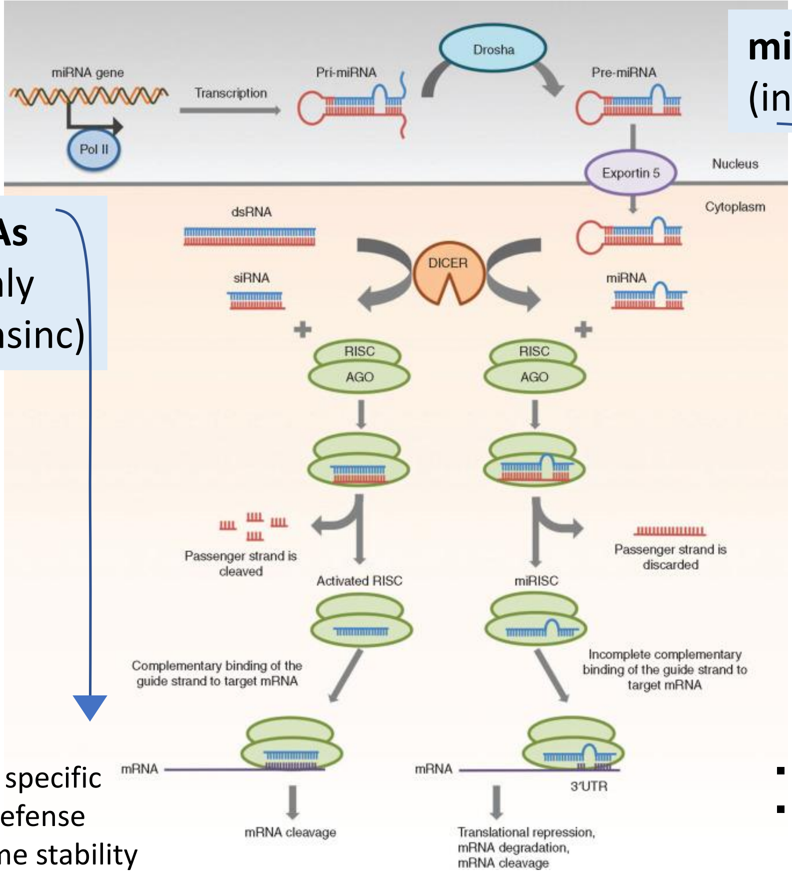
snRNA aid in Splicing

## Small nucleolar RNAs (snoRNA)



snoRNA stabilize the rRNA structure during ribosome biogenesis

# Regulatory non-coding RNAs → miRNAs and siRNAs



**siRNAs** (mainly extrinsinc)

- Highly specific
- Viral defense
- Genome stability

**miRNAs** (intrinsic)

- Multiple targets
- Gene expression regulation

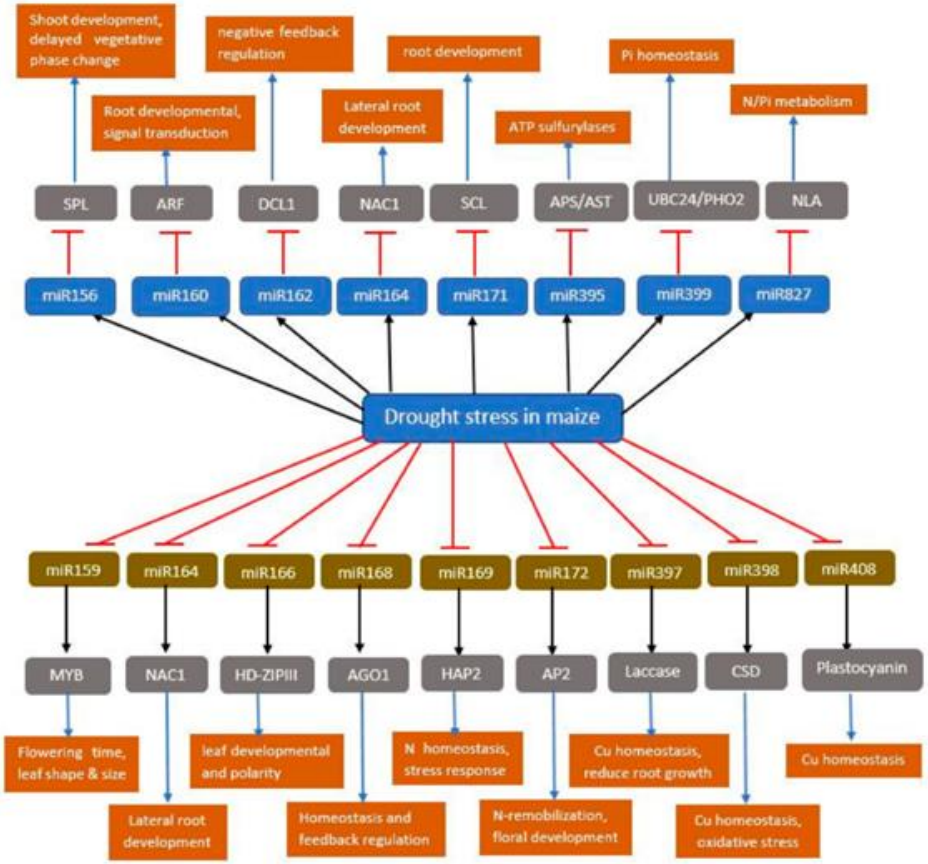
## RNA Interference

Micro RNAs (miRNAs) and
small interfering RNAs (siRNAs)
suppress gene expression by:

- Degradation of mRNA
- Inhibition of translation
- Heterochromatin formation (epigenetics)

… and even more non-coding RNAs, involved in gene expression regulation

# ncRNAs are part of a complex regulatory system



**Example:**
**ncRNA in plant stress response**

Up- and downregulation of miRNAs for drought stress response in maize

# Nice and well, but now how are RNAs investigated?

## Transcriptomics Technologies - Beginnings

**First attempts to capture and investigate transcripts date back to the 1970’s and 1980’s.**

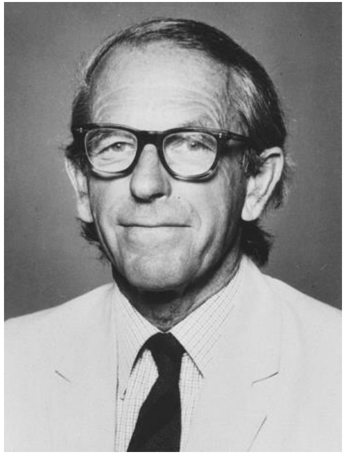
**Among others including**

- Sanger Sequencing
- Expressed Sequence Tags (EST)
- Serial Analysis of Gene Expression (SAGE)
- Northern Blotting
- Reverse transcriptase quantitative PCR
- … …

→ In the beginning, investigating transcripts was laborious, expensive and usually captured only a small number of genes

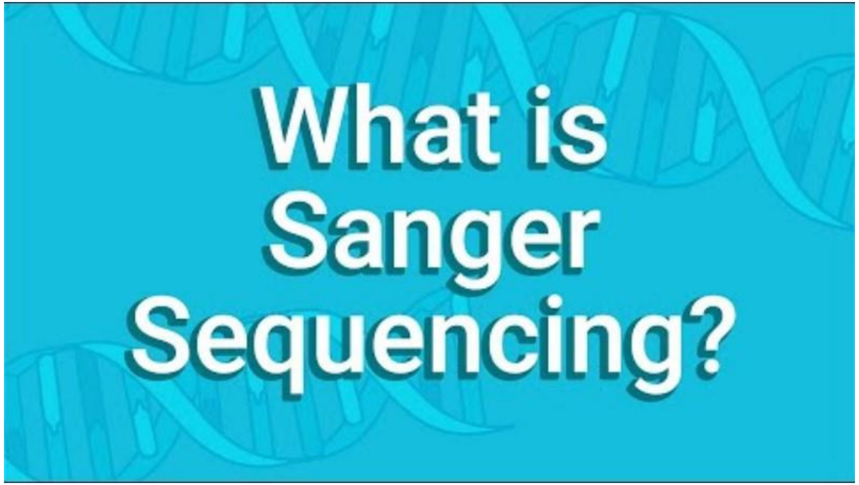
# Transcriptomics Technologies - Beginnings

## Sanger Sequencing



**Frederick Sanger**
British Biochemist

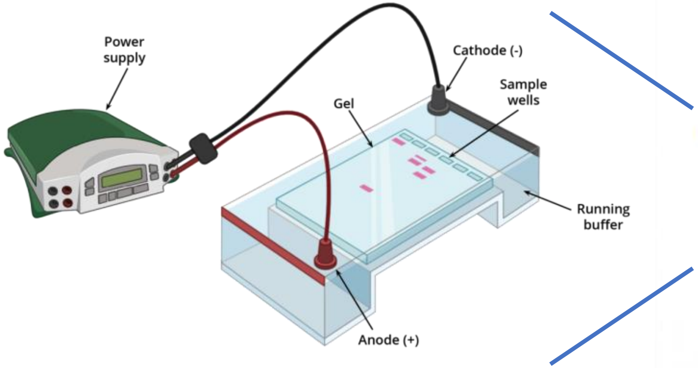
Nobel prize 1980 for first-ever DNA sequencing technique

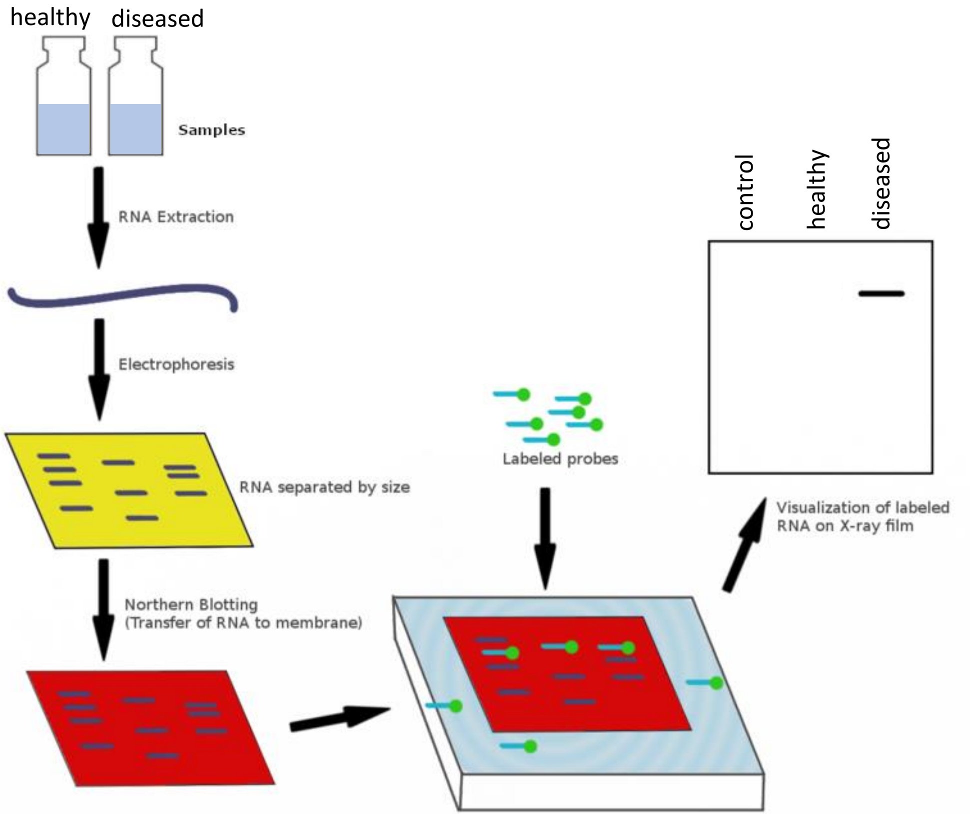


# Transcriptomics Technologies - Beginnings

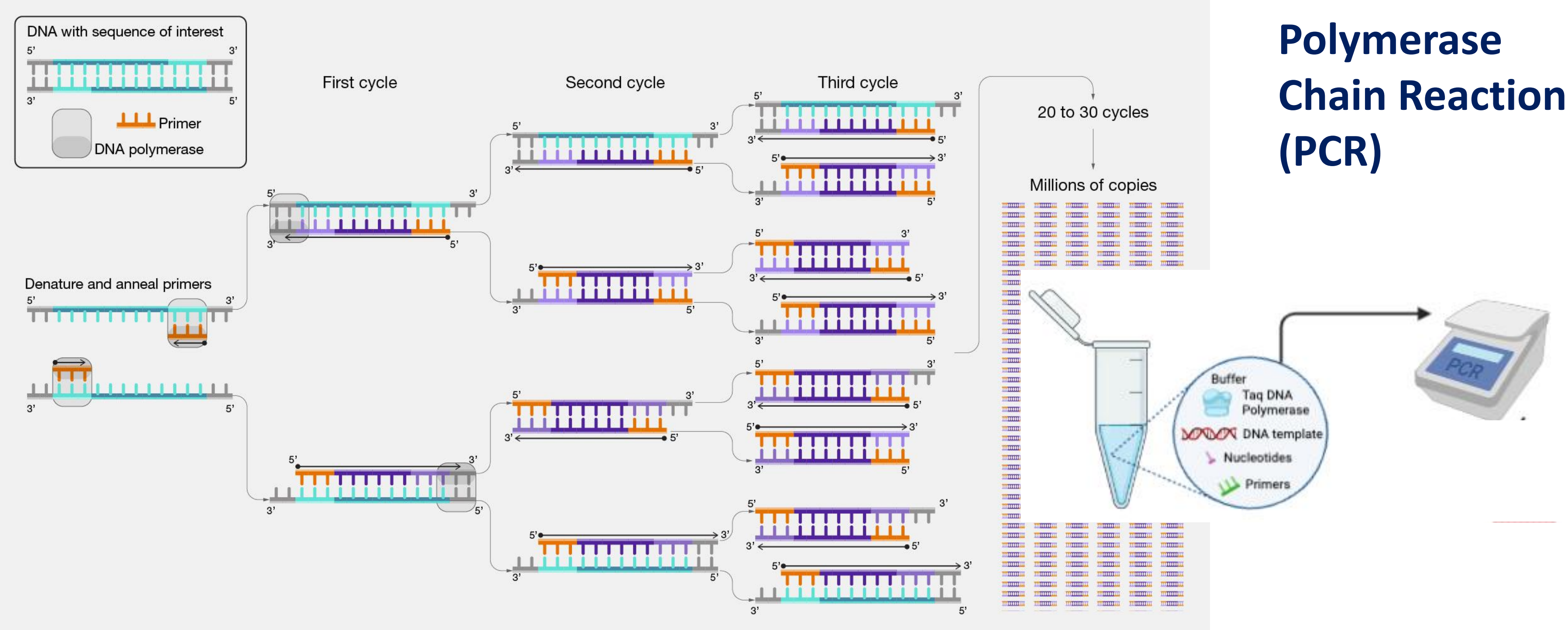
## Northern Blotting

allows to investigate one
or a small number of genes





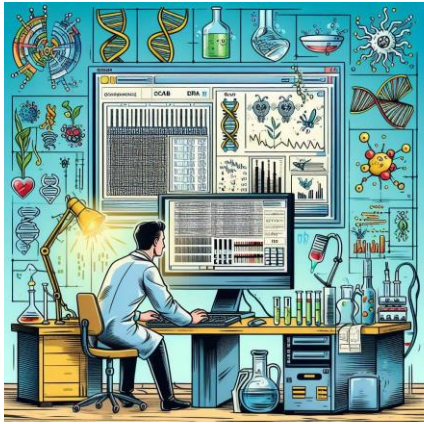
# Transcriptomics Technologies - Beginnings



## Polymerase Chain Reaction (PCR)

# Transcriptomics Technologies - Nowadays

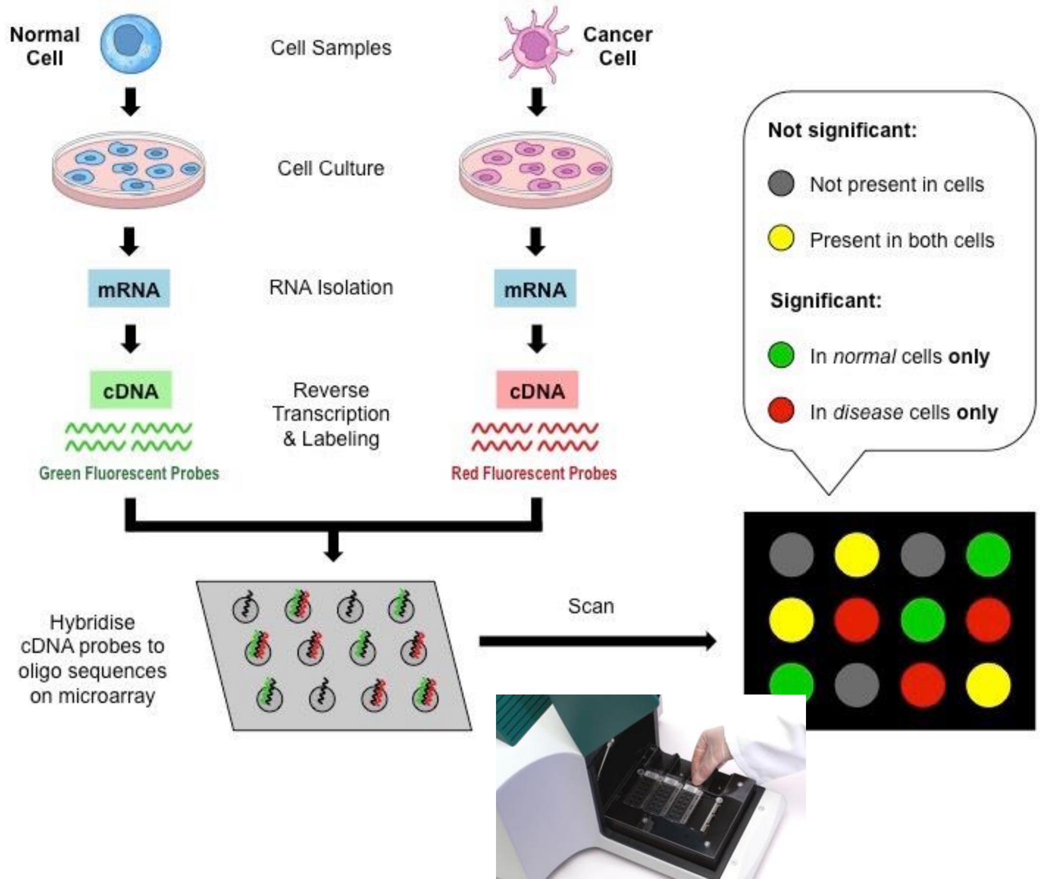
These first techniques are still in use, they are sensitive, specific (with gene-specific primers) and rapid
**BUT they only capture a tiny subsection of a transcriptome**



Two dominant contemporary techniques:
**Microarrays and RNA-Sequencing**
Developed in the 1990's and 2000's, respectively

**Allow to capture a complete picture of the transcriptome**

# Transcriptomics Technologies - Microarrays



**Normal Cell**

Cell Samples

**Cancer Cell**

Cell Culture

**mRNA**

RNA Isolation

**mRNA**

**cDNA**

Reverse Transcription & Labeling

**cDNA**

Green Fluorescent Probes

Red Fluorescent Probes

Hybridise cDNA probes to oligo sequences on microarray

Scan

**Not significant:**

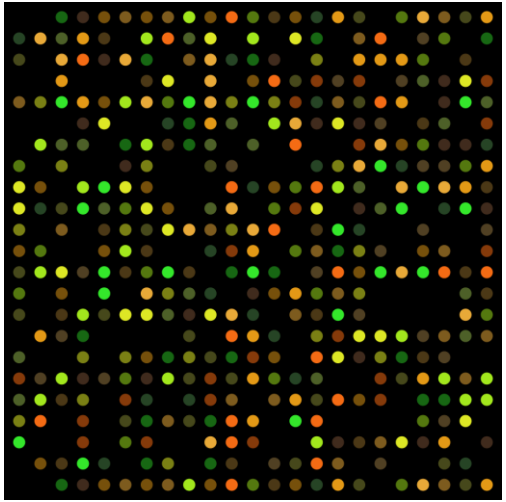
● Not present in cells

● Present in both cells

**Significant:**

● In *normal* cells **only**

● In *disease* cells **only**



# Transcriptomics Technologies - RNA-Sequencing

Number of publications referring to different transcriptomics methods over the last three decades

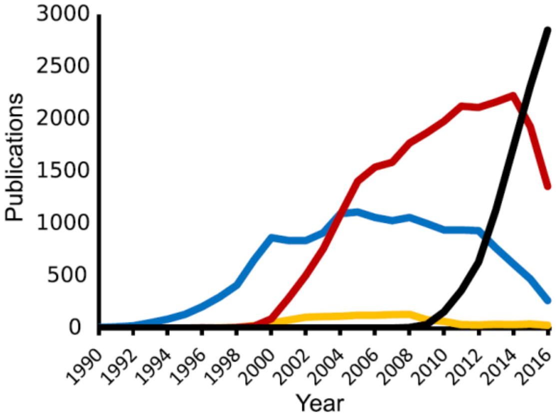


**Fig 1. Transcriptomics method use over time.** Published papers since 1990, referring to RNA sequencing (black), RNA microarray (red), expressed sequence tag (blue), and serial/cap analysis of gene expression (yellow)[12].

**First descriptions of RNA Sequencing in 2006 and 2008, overtook microarrays as dominant technique in 2015.**

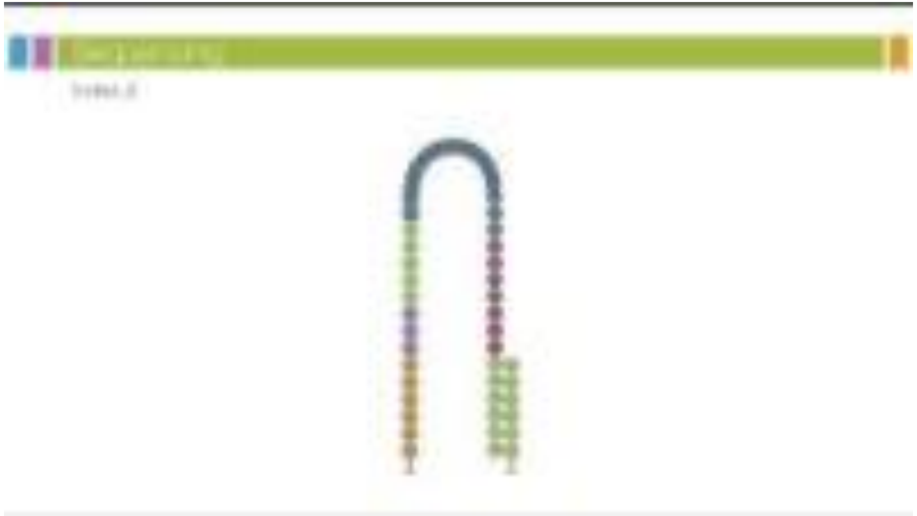
**Advantages:**

- Very high throughput
- Allows detection and quantification of transcripts
- Analyzing of non-coding RNAs
- Alignment to reference and *de novo* assembly possible
- Information about alternative splicing events

# Transcriptomics Technologies - RNA-Sequencing

**Illumina sequencing by synthesis**

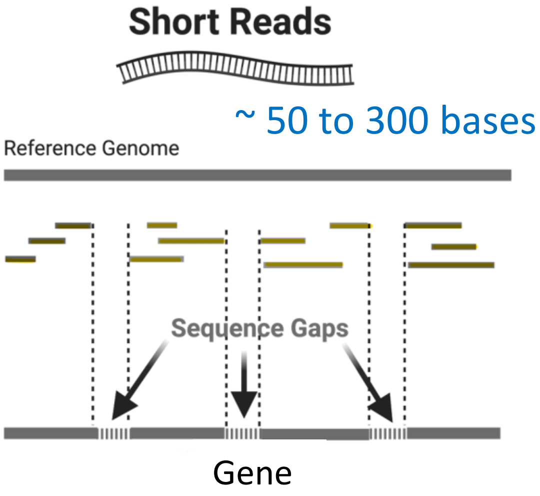
For a long time the absolute gold-standard in sequencing

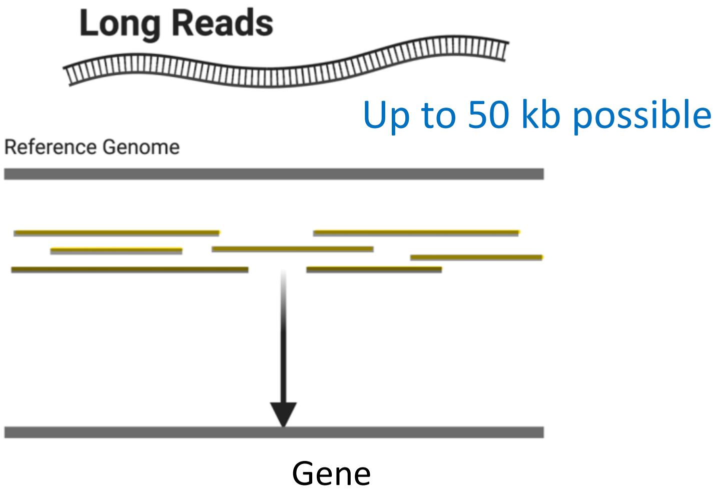


# Transcriptomics Technologies - RNA-Sequencing

## Short- vs Long-read Sequencing

Tremendous advancement for transcript detection **via long-read sequencing**
→ One read = one transcript (full-length cDNA) has become reality





# Transcriptomics Technologies - RNA-Sequencing

For comparison: long-read sequencing nanopore technique

# THANK YOU!

Questions?