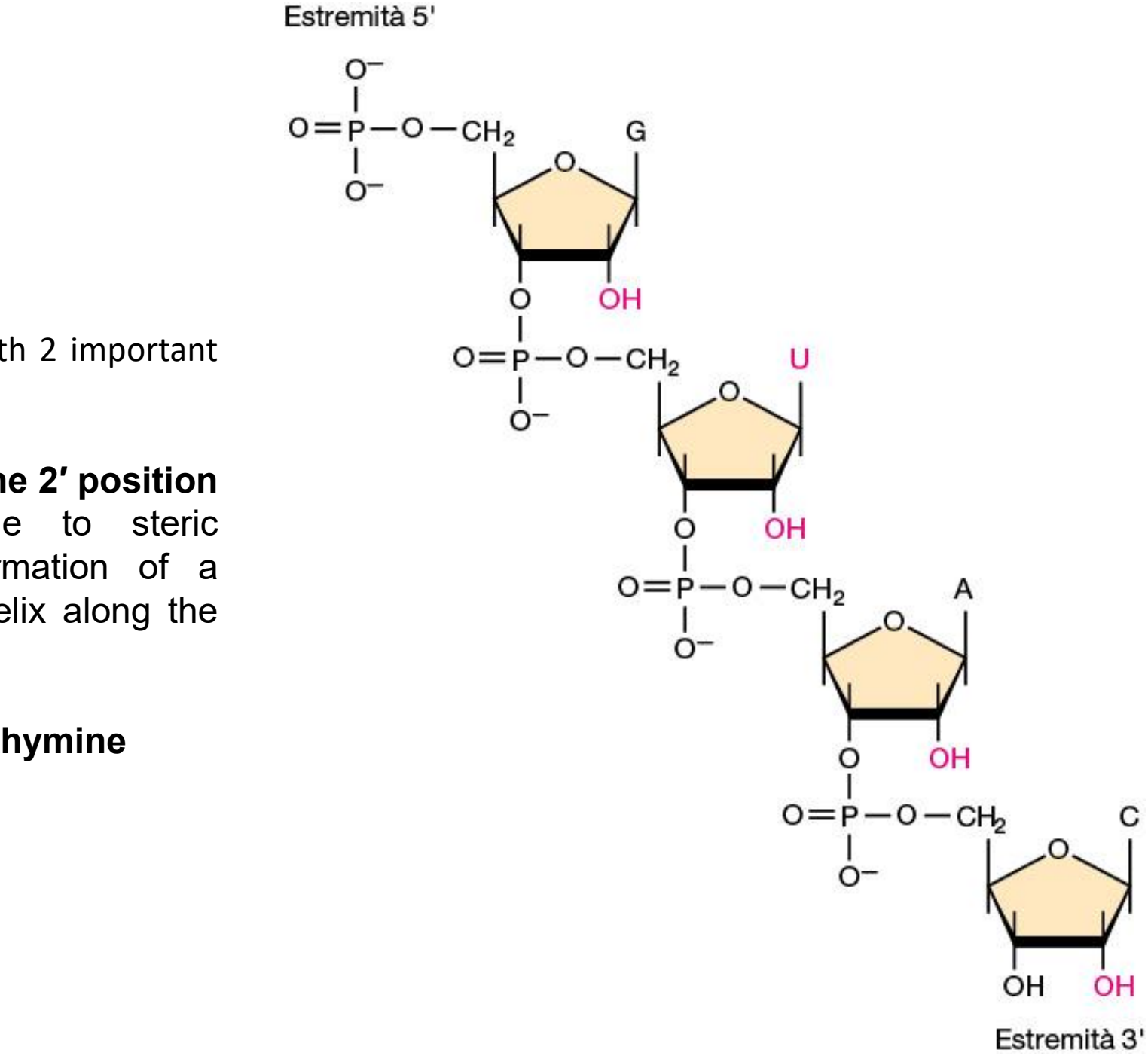
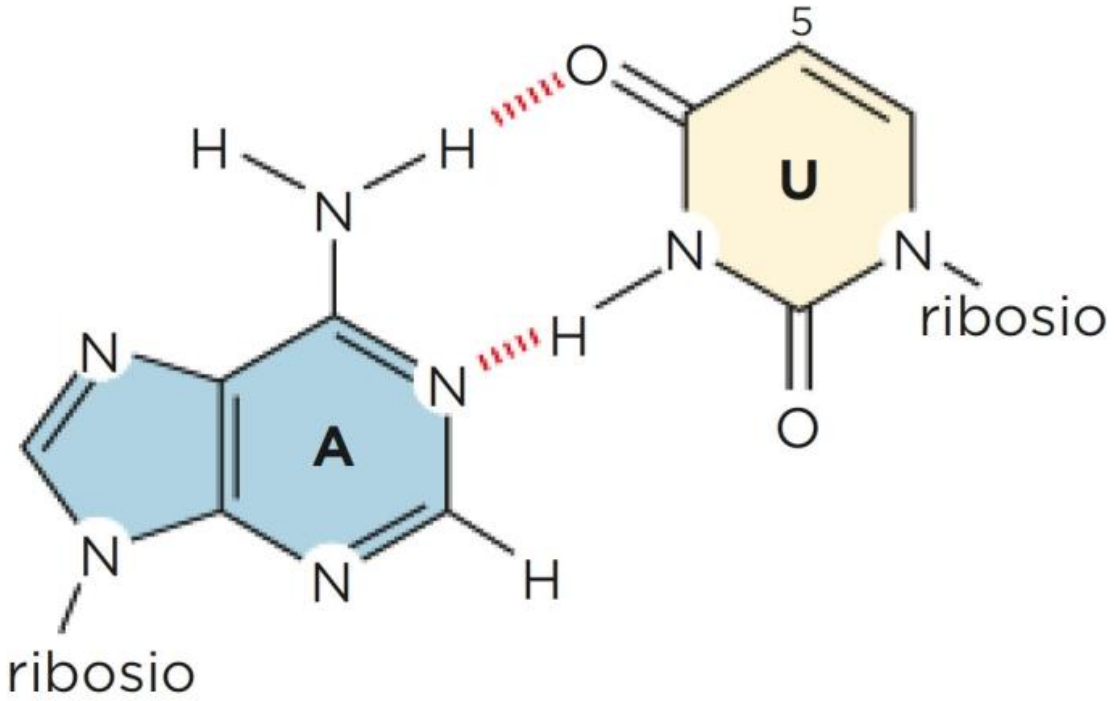
RNA

Similar structure to that of DNA with 2 important exceptions:

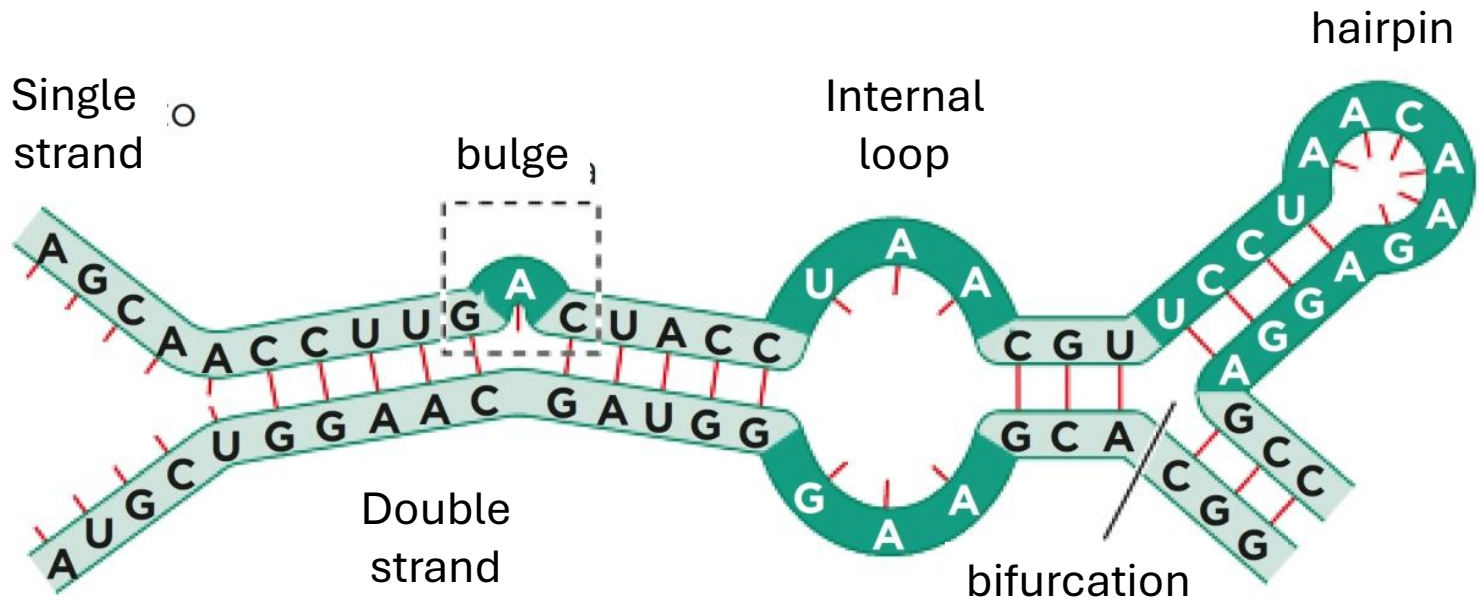
- **presence of an OH group at the 2′ position of the sugar RIBOSE** (due to steric hindrance, it prevents the formation of a continuous and stable double helix along the entire length of a strand)

- **presence of uracil instead of thymine**





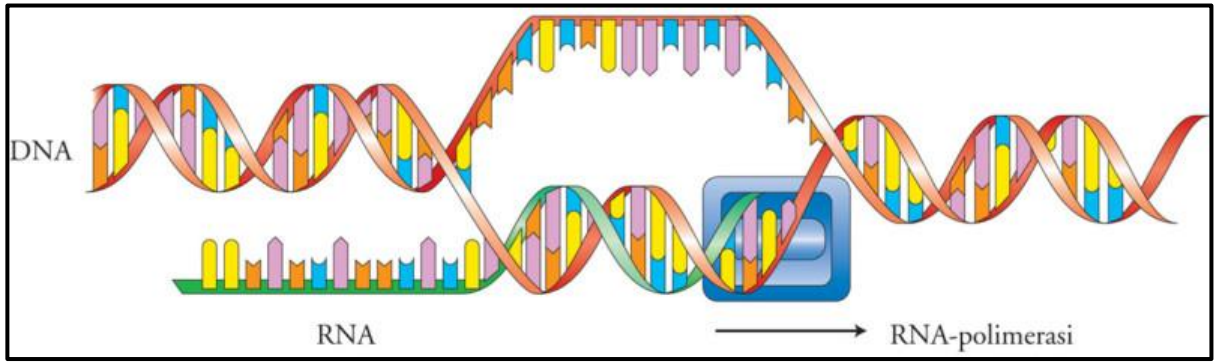
Uracil is still able to form
hydrogen bonds with Adenine



Structural features of RNA. In a molecule of RNA that displays self-complementary bases, stretches of double helix (duplex) can be formed, however they are not long. A **bulge** occurs when unpaired nucleotides protrude from one side of a stem. An **internal loop** is formed when unpaired nucleotides appear on both strands within a stem. A **hairpin** (or stem–loop) arises when a sequence folds back on itself, creating a paired stem region with an unpaired loop at the end. These elements constitute the RNA secondary structure and are fundamental for **RNA folding** (3D shape acquisition) and **function**.
Indeed, because **RNA has greater structural diversity than DNA**, it carries out a **wide range of cellular functions** and serves as much more than just a storage medium for genetic information.

RNA type	Full name	Main function	Relative abundance	Typical length
rRNA	Ribosomal RNA	Structural and catalytic component of ribosomes.	~80–85% of total RNA	120–5,000 nt (18S ~1.9 kb, 28S ~5 kb, 5.8S ~160 nt, 5S ~120 nt)
tRNA	Transfer RNA	Delivers amino acids to ribosomes during translation.	~10–15%	~70–90 nt
mRNA	Messenger RNA	Carries genetic information from DNA to ribosomes for protein synthesis.	~1–5%	~500–10,000 nt (highly variable, avg. ~2 kb)
snRNA	Small nuclear RNA	Involved in pre-mRNA splicing (spliceosome).	<1%	~100–300 nt
snoRNA	Small nucleolar RNA	Guides chemical modifications of rRNA, tRNA, and snRNA.	<1%	~60–300 nt
miRNA	MicroRNA	Regulates gene expression post-transcriptionally.	<0.01%	~20–24 nt
siRNA	Small interfering RNA	RNA interference and gene silencing.	Trace	~20–25 nt
lncRNA	Long non-coding RNA	Regulates gene expression at multiple levels.	~1–2% (variable, cell-type specific)	>200 nt, often 1–10 kb
piRNA	Piwi-interacting RNA	Silences transposons in germ cells.	High in germline cells; negligible in somatic cells	~24–31 nt
circRNA	Circular RNA	Regulatory functions, often act as miRNA sponges.	Low, varies widely across cell types	Highly variable, ~200 nt to >5 kb

RNA is always transcribed from DNA

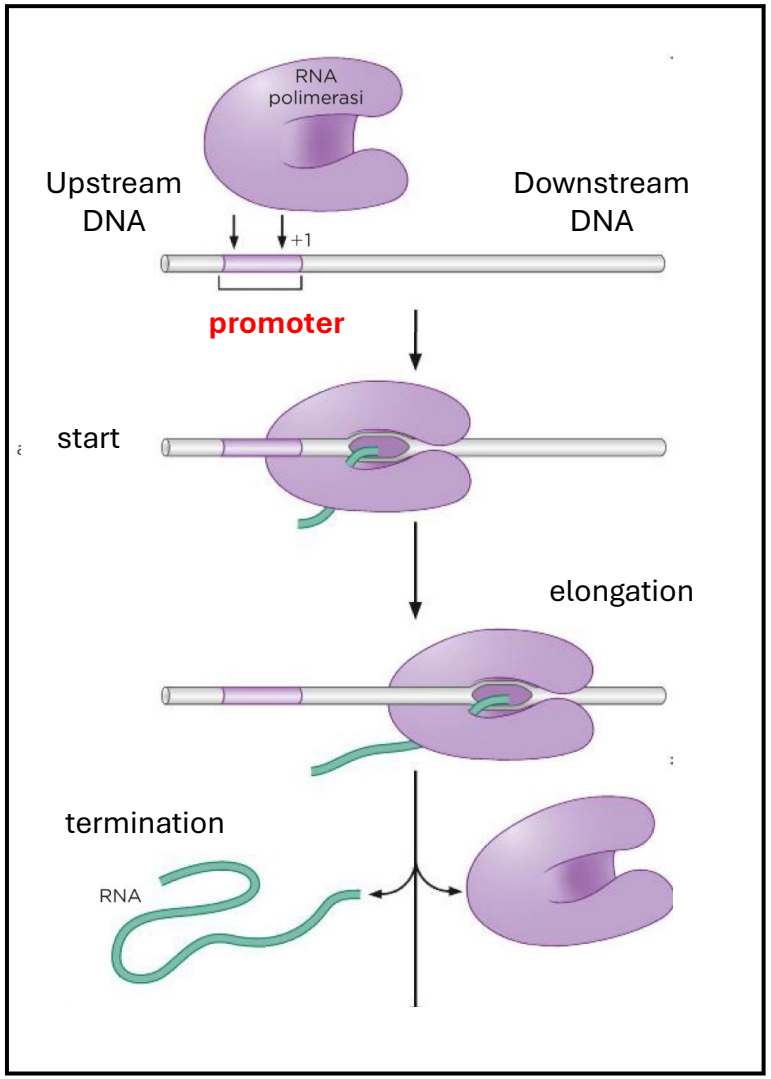


Transcription is the biological process by which a DNA sequence is copied into an RNA molecule by RNA polymerase.

The **RNA polymerase** is similar to DNA polymerase, but it does not require a primer and it elongates NTPs and not dNTPs (different building blocks).

The RNA carries the same information (i.e. sequence) of the DNA.

Mechanism of transcription

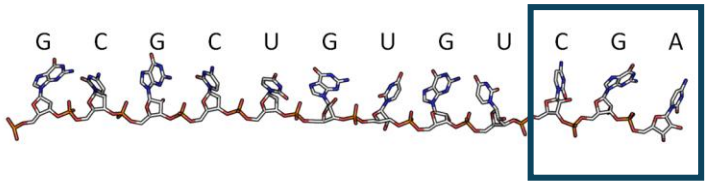


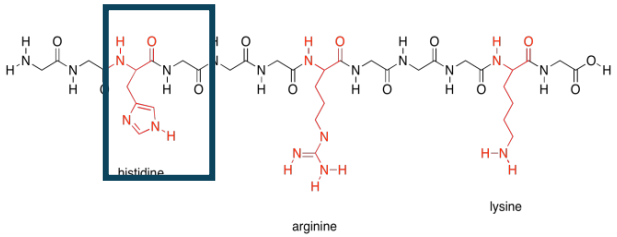
RNA can act as a messenger to encode proteins

DNA/RNA: nucleotide polymers (4 building blocks)

Proteins: amino acid polymers (20 building blocks)

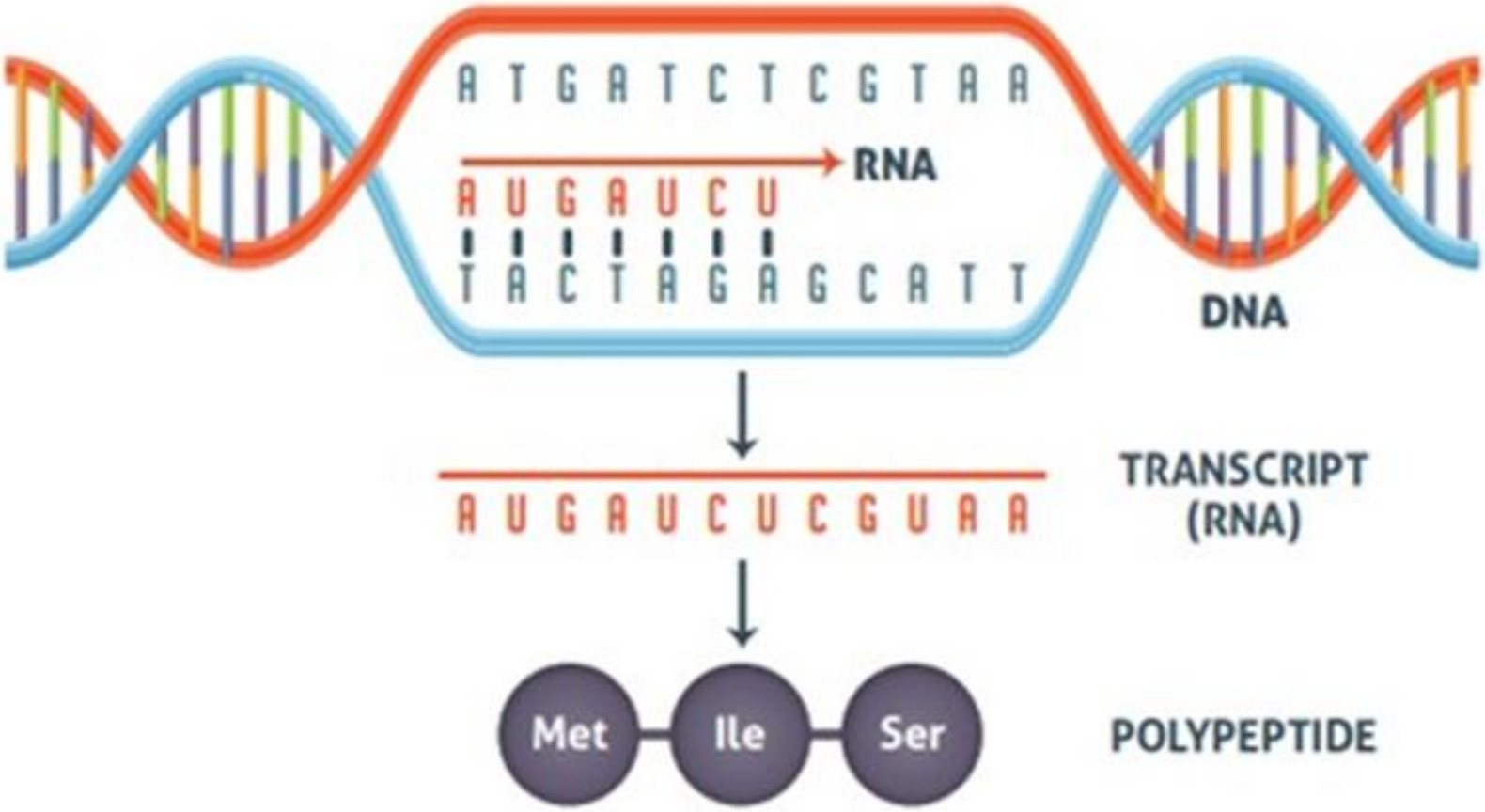
Primary structure





Genetic code triplet

RNA can act as a messenger to encode proteins



The **messenger RNA** (**mRNA**) is a type of RNA that transfer the genetic information from DNA to a protein.

mRNA is the real measure of if and how much a gene is expressed.

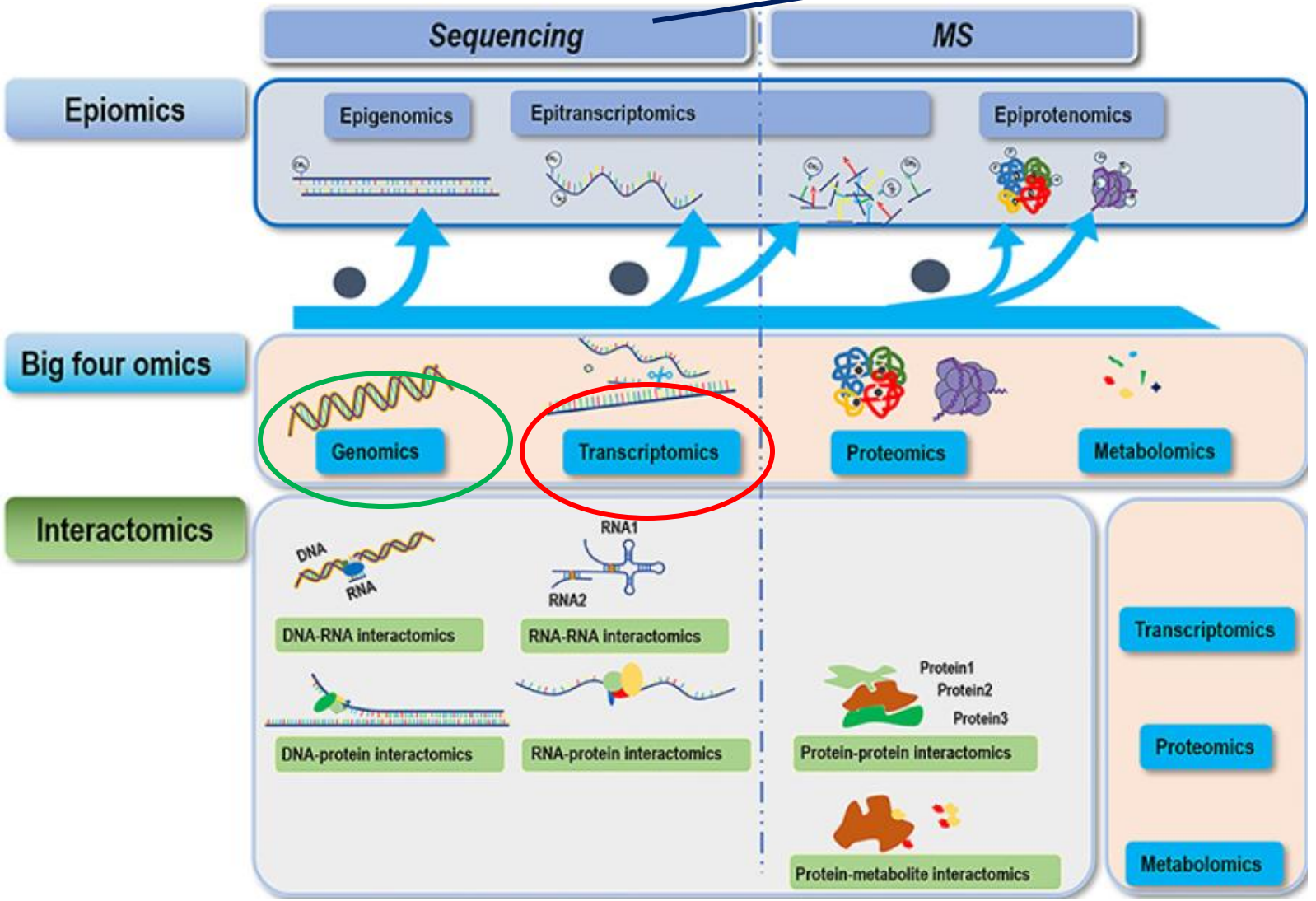
➔ No mRNA: no gene expression

➔ Low mRNA: few gene expression

➔ High mRNA: abundant gene expression

If you want to know what genes are expressed in a particular condition (tissue, pathology, etc) you have to analyze the pool of mRNAs present in your sample, not DNA

Overview on Omics disciplines



Sequence identification is the main advanced technology allowing to investigate the DNA and RNA omics of a given biological system.

- **Genomics:** study of the entire set of DNA molecules of a biological system
- **Transcriptomics:** study of the entire set of transcripts (RNA) of a biological system

DOI: 10.3389/fmed.2022.911861

A transcriptome is a collection of all the gene readouts (i.e. transcripts) present in the cell of a particular tissue. Each gene can give rise to numerous transcripts. Being gene expression tissue specific, the transcriptome is highly variable

Yu, Next Generat Sequenc & Applic 2015, 2:2
DOI: 10.4172/2469-9853.1000e104

Journal of
Next Generation Sequencing & Applications

Editorial Open Access

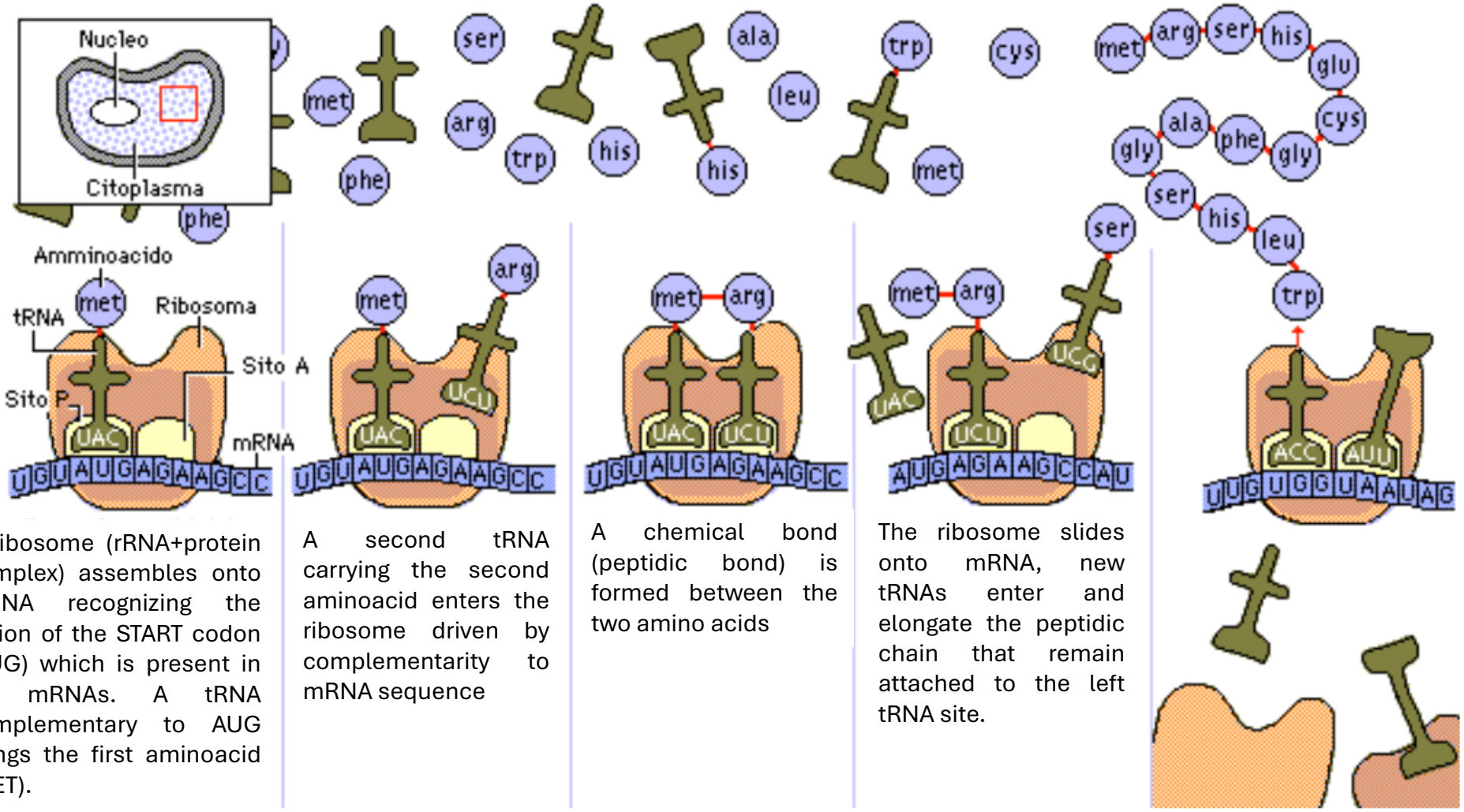
The Human Transcriptomes Project: Is it hard?

Jun Yu*

CAS Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.

The Human Transcriptomes Project is a larger project as compared to the Human Genome Project, where genome sequences are largely defined by population diversity and the number of sequence variations is finite in number and occurrence but transcriptomes are highly variable for each individuals and cells and influenced by genetic, epigenomic, and environmental factors. It is essential to have a common information platform to integrate all molecular data in a cellular context; it is equally important to have a standard protocol for all efforts to produce consistent transcriptomes. Nonetheless, it is not

Translation converts mRNA into protein, thanks to ribosomes and tRNA



A ribosome (rRNA+protein complex) assembles onto mRNA recognizing the region of the START codon (AUG) which is present in all mRNAs. A tRNA complementary to AUG brings the first aminoacid (MET).

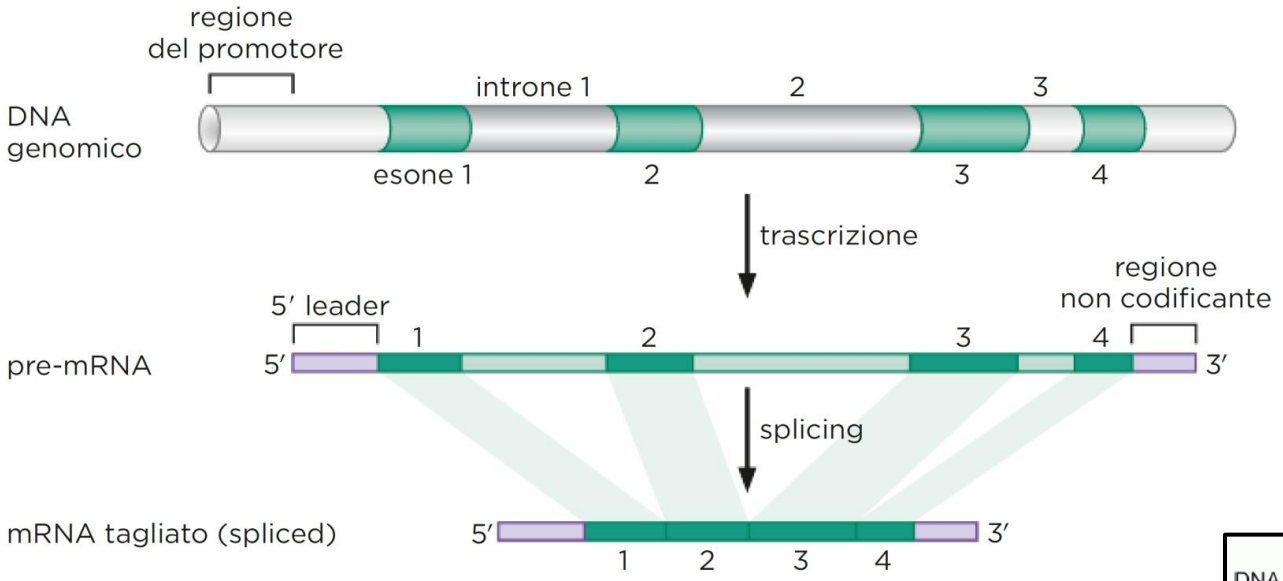
A second tRNA carrying the second aminoacid enters the ribosome driven by complementarity to mRNA sequence

A chemical bond (peptidic bond) is formed between the two amino acids

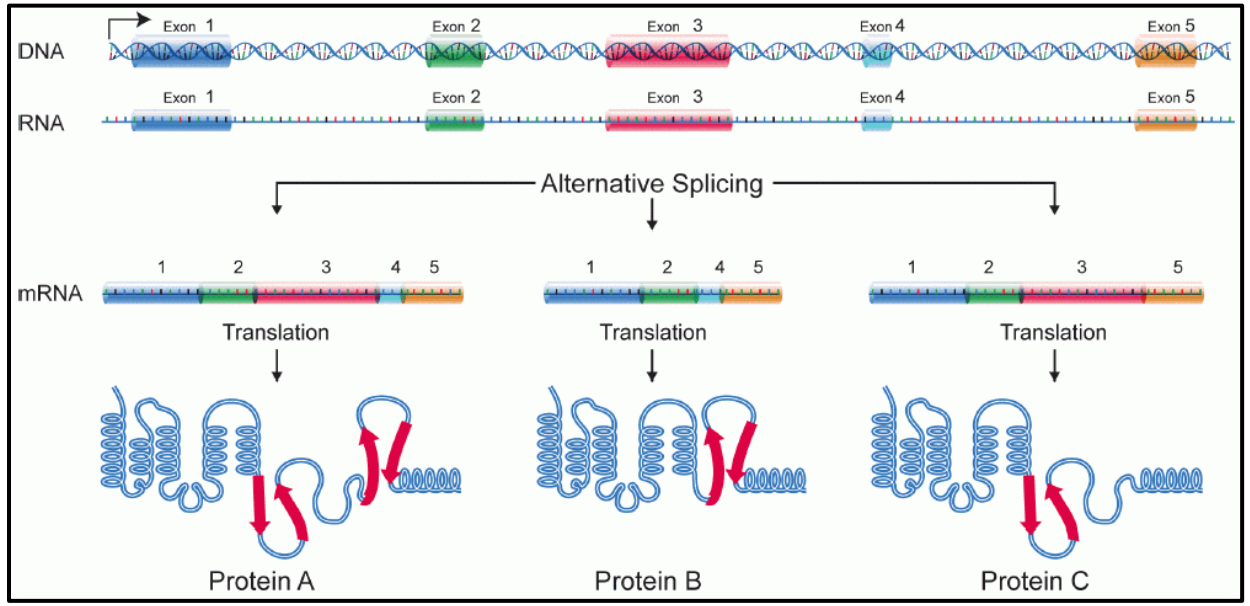
The ribosome slides onto mRNA, new tRNAs enter and elongate the peptidic chain that remain attached to the left tRNA site.

The last codon of the mRNA is hybridized by a tRNA carrying no amino acid. This stops translation and the peptidic chain (protein) is detached from the mRNA.

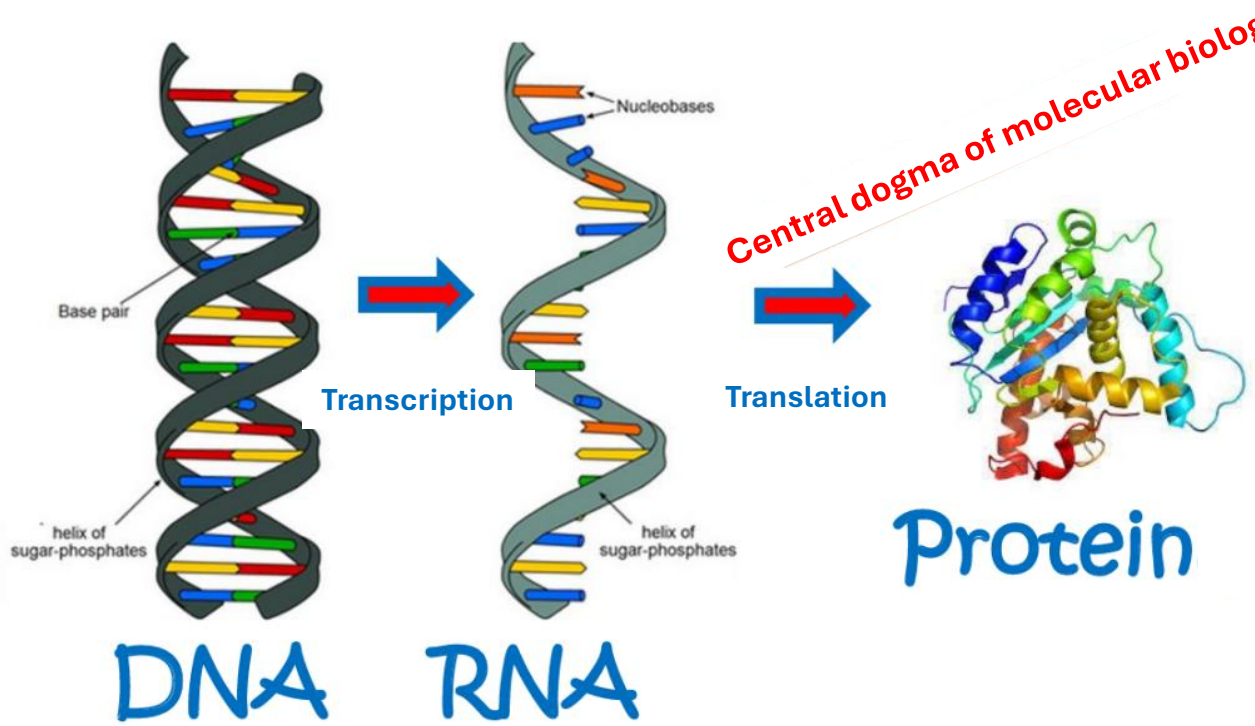
mRNA splicing generates variability



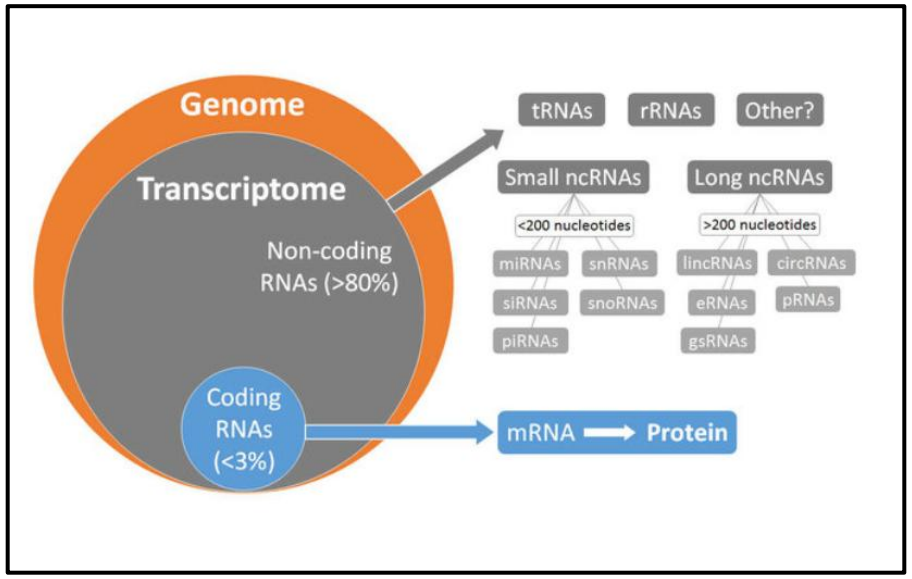
Splicing is the process by which **introns** are removed from pre-mRNA and **exons** (the real coding regions) are joined together, generating mature transcripts; its regulation produces **alternative isoforms**, augmenting the variability of gene expression -- a key focus of transcriptome analysis.



A typical human gene architecture. The gene contains exons, separated by introns. All follow the promoter region, from which transcription starts generating a pre-mRNA, that contains all exons and introns. Splicing removes introns and joins exons generating the mature mRNA, leaving small regulatory sequences at the 5’ and 3’ ends (UTR=untranslated region). Several possible mRNA can be obtained by **alternative splicing**.



Originally, a 1:1:1 stoichiometry between DNA, RNA and proteins was assumed. This was called the **central dogma of molecular biology**: all the genetic information present in DNA is converted into RNA and then into proteins.



From 2005, with the first **RNA-seq** experiments, it has become clear that the number of transcripts encoded by the genome is around 10 times as great as the number of genes.
Most of them are **non-coding RNA**, displaying **effector** and **regulatory** roles.
This has lead to the introduction of the concept of «**junk DNA**».

RNA type	Full name	Main function	Relative abundance	Typical length
rRNA	Ribosomal RNA	Structural and catalytic component of ribosomes.	~80–85% of total RNA	120–5,000 nt (18S ~1.9 kb, 28S ~5 kb, 5.8S ~160 nt, 5S ~120 nt)
tRNA	Transfer RNA	Delivers amino acids to ribosomes during translation.	~10–15%	~70–90 nt
mRNA	Messenger RNA	Carries genetic information from DNA to ribosomes for protein synthesis.	~1–5%	~500–10,000 nt (highly variable, avg. ~2 kb)
snRNA	Small nuclear RNA	Involved in pre-mRNA splicing (spliceosome).	<1%	~100–300 nt
snoRNA	Small nucleolar RNA	Guides chemical modifications of rRNA, tRNA, and snRNA.	<1%	~60–300 nt
miRNA	MicroRNA	Regulates gene expression post-transcriptionally.	<0.01%	~20–24 nt
siRNA	Small interfering RNA	RNA interference and gene silencing.	Trace	~20–25 nt
lncRNA	Long non-coding RNA	Regulates gene expression at multiple levels.	~1–2% (variable, cell-type specific)	>200 nt, often 1–10 kb
piRNA	Piwi-interacting RNA	Silences transposons in germ cells.	High in germline cells; negligible in somatic cells	~24–31 nt
circRNA	Circular RNA	Regulatory functions, often act as miRNA sponges.	Low, varies widely across cell types	Highly variable, ~200 nt to >5 kb

Effectors in translation

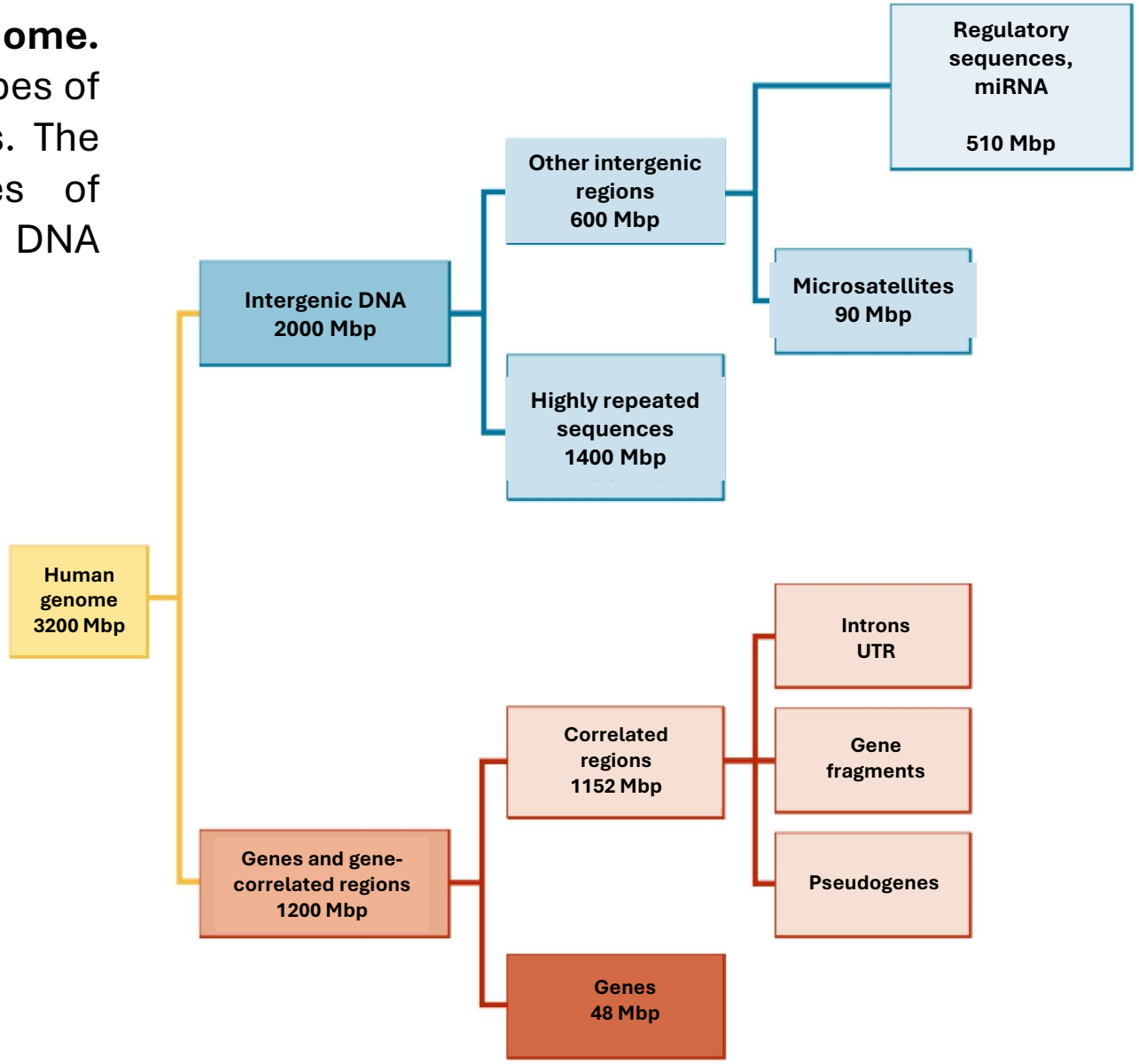
Result of gene expression

Regulators of gene expression by directly or indirectly interacting with mRNAs

Junk DNA is definitely not junk

Organization and content of the human genome. The human genome is composed of many different types of sequences, most of which do not code for proteins. The distribution and quantity of the different types of sequences are shown. Main functions of intergenic DNA are:

1. **Regulation of gene expression**
   - Contains **regulatory sequences** (enhancers, silencers, insulators) that control when, where, and how much a gene is expressed.
   - Includes regions producing **non-coding RNAs** (e.g., miRNAs, lncRNAs) with regulatory roles.
2. **Genome structural organization**
   - Repetitive sequences contribute to **chromatin compaction** and the 3D architecture of the genome.
   - Help define functional chromosomal domains.
3. **Evolution and genetic variability**
   - Repetitive and mobile sequences (transposons, retrotransposons) are a source of **genetic innovation** and evolutionary diversity.
4. **Gene protection**
   - Acts as a **spacer** between genes, reducing the likelihood that mutations or recombination directly damage coding regions.

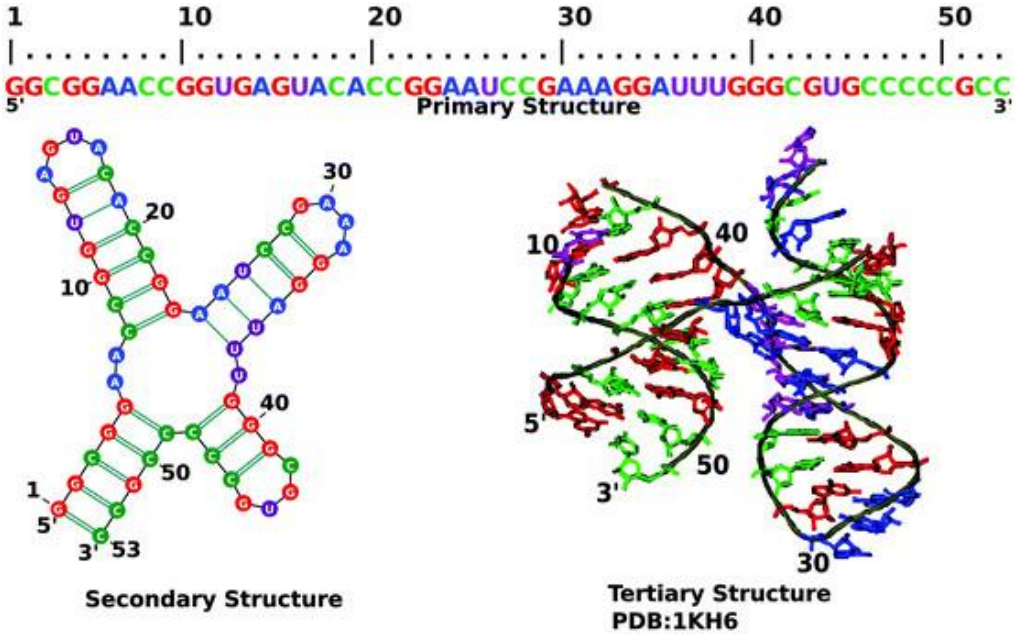


1 Mbp = 1 Mega base-pair= 1 million paired bases
Measure unit equivalent to nucleotides per each strand (nt)

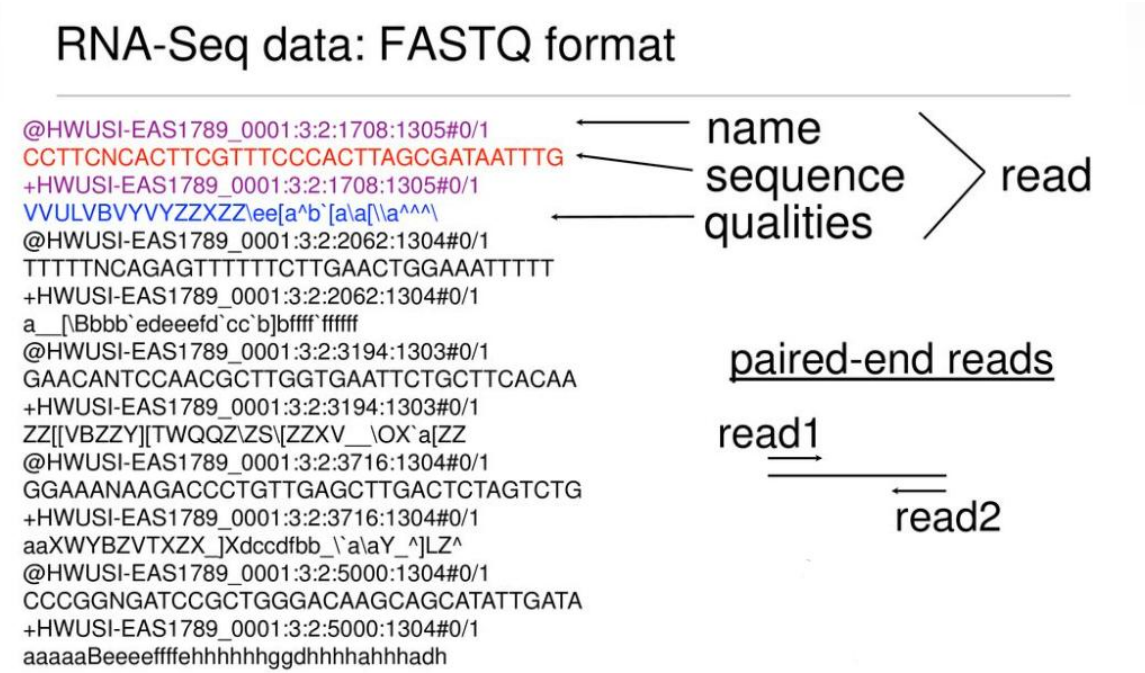
RNA primary sequence

For the bioinformatician…

For the molecular biologist…

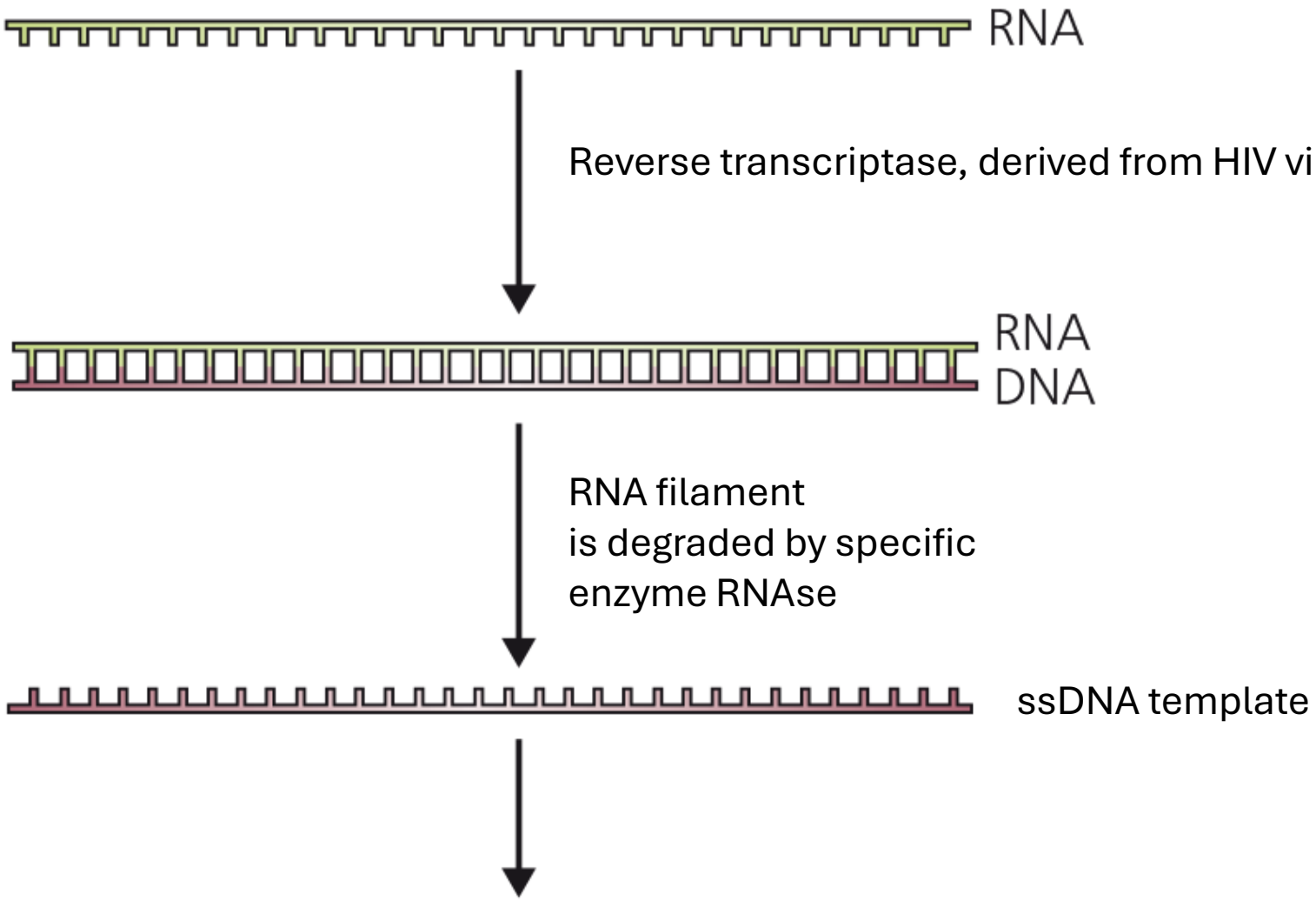


The sequence of RNA is often represented as sequence of DNA. This is because, for most sequencing methods, the RNA sample is first converted to DNA and then amplified by PCR before sequencing (e.g. Illumina, Ion Torrent methods).
More recent sequencing technologies (third generation sequencing) allow direct RNA sequencing without the need to convert it into DNA and amplify it by PCR (e.g. Oxford Nanopore and PacBio sequencing).



How to convert an RNA into a DNA sequence?

With a retrotranscription reaction!

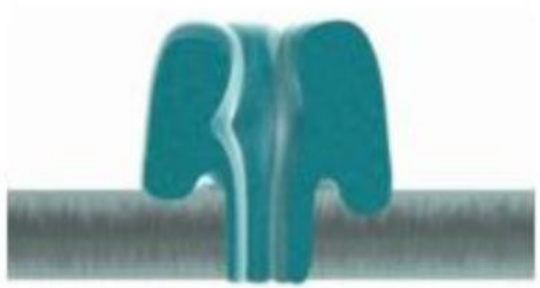


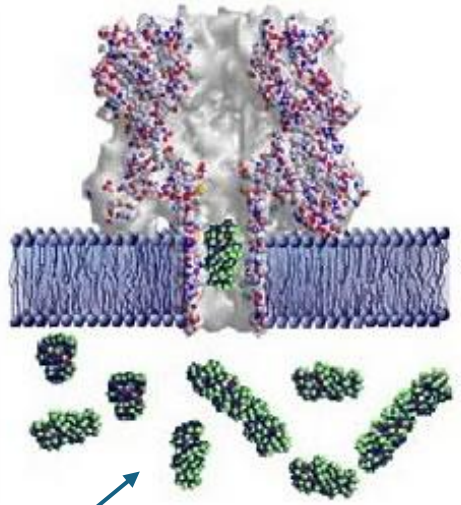
Third generation sequencing

Oxford nanopore technology

A nanopore: a nano-scale hole.

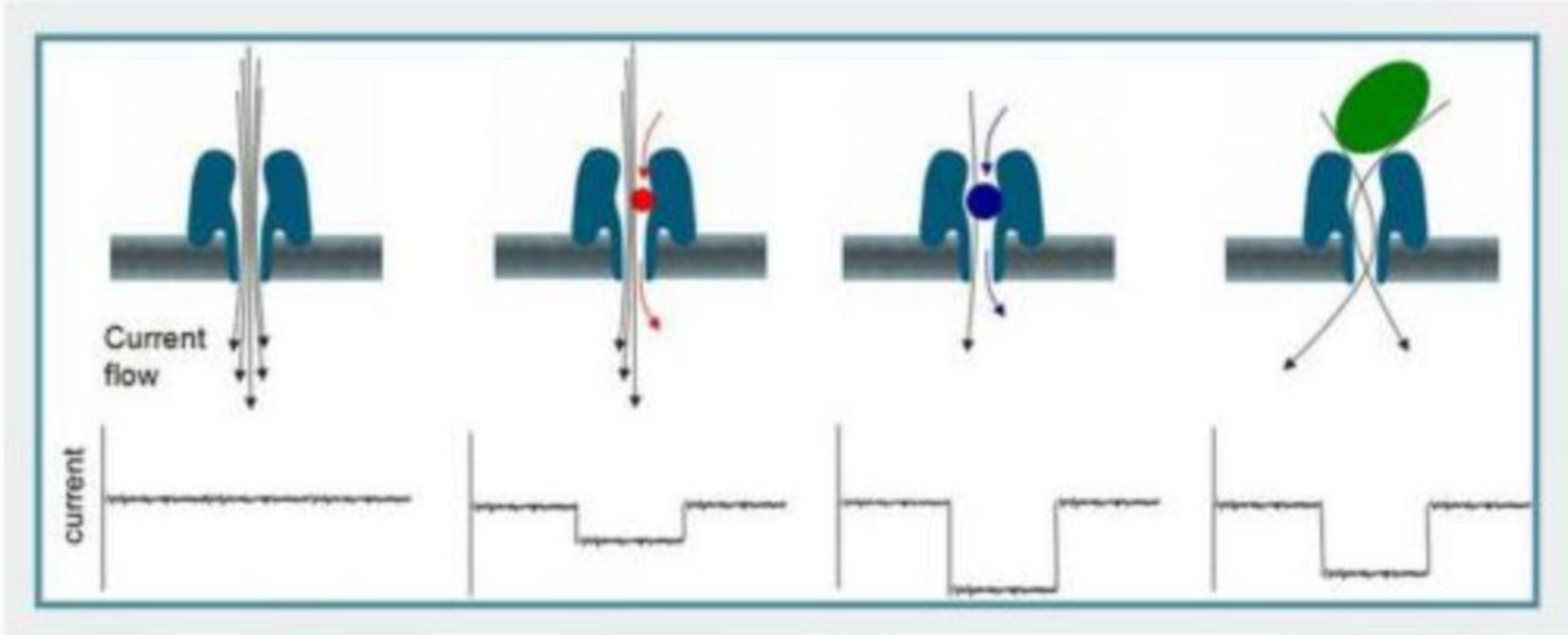
- Biological: a pore-forming protein (e.g. α-Hemolysin) in a membrane (e.g. lipid bilayer)
- Solid-state: in synthetic materials (e.g. silicon nitride or graphene)
- Hybrid: formed by a pore-forming protein set in synthetic material





Nanopore sensing

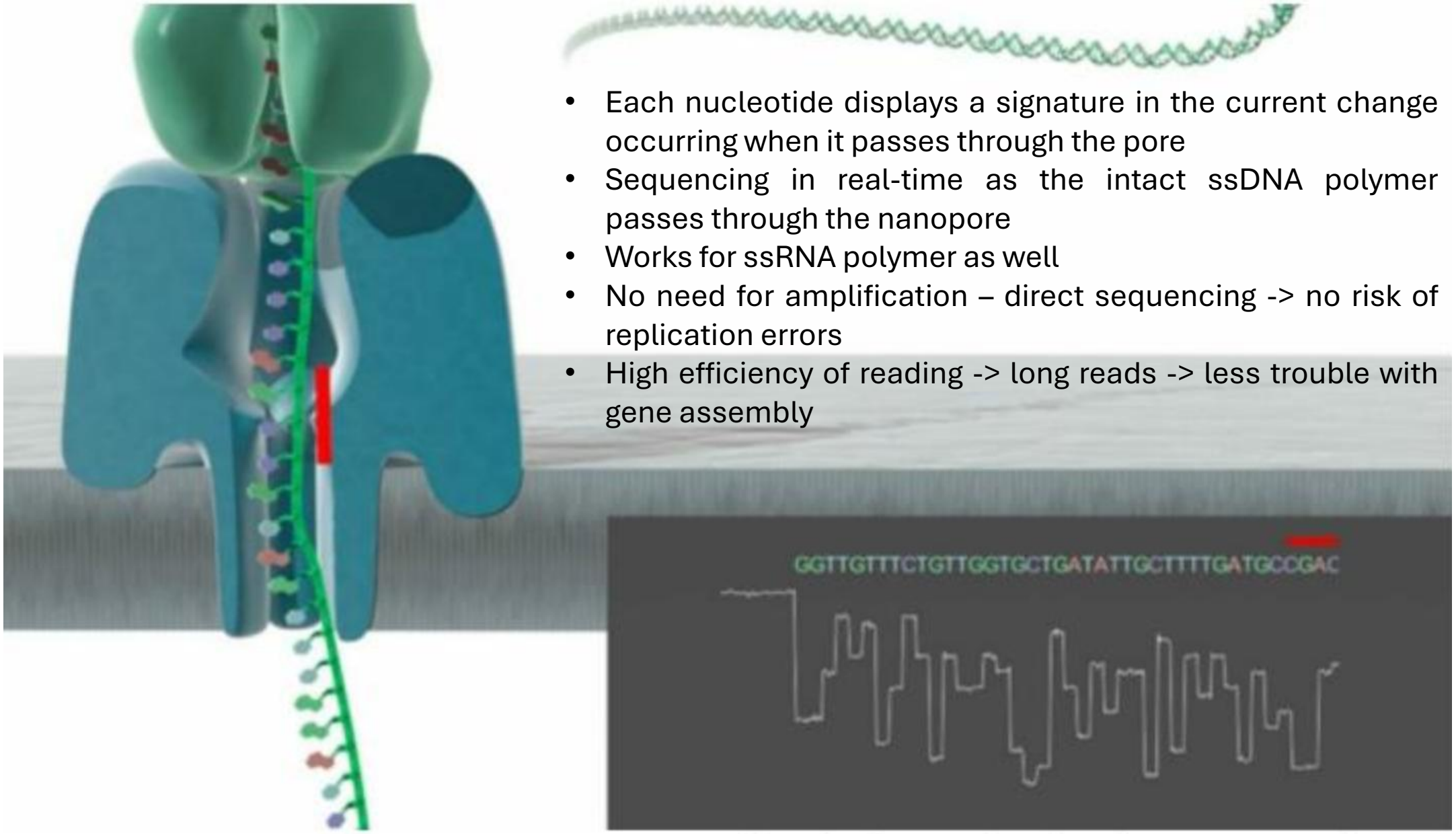
Ionic current passed through membrane by setting a voltage across the membrane.



- Disruption in current detected when analyte passes through the pore or near its aperture.

- Characteristic disruption indentifies the molecule in question.

Strand Sequencing



- Each nucleotide displays a signature in the current change occurring when it passes through the pore
- Sequencing in real-time as the intact ssDNA polymer passes through the nanopore
- Works for ssRNA polymer as well
- No need for amplification – direct sequencing -> no risk of replication errors
- High efficiency of reading -> long reads -> less trouble with gene assembly

Snapshot from movie at http://www.nanoporetech.com