

Some Basics of Molecular Biology

Hubert Rehrauer



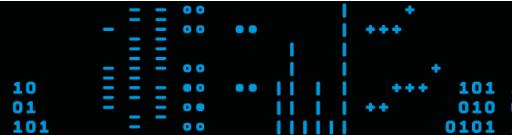
University of
Zurich^{UZH}

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

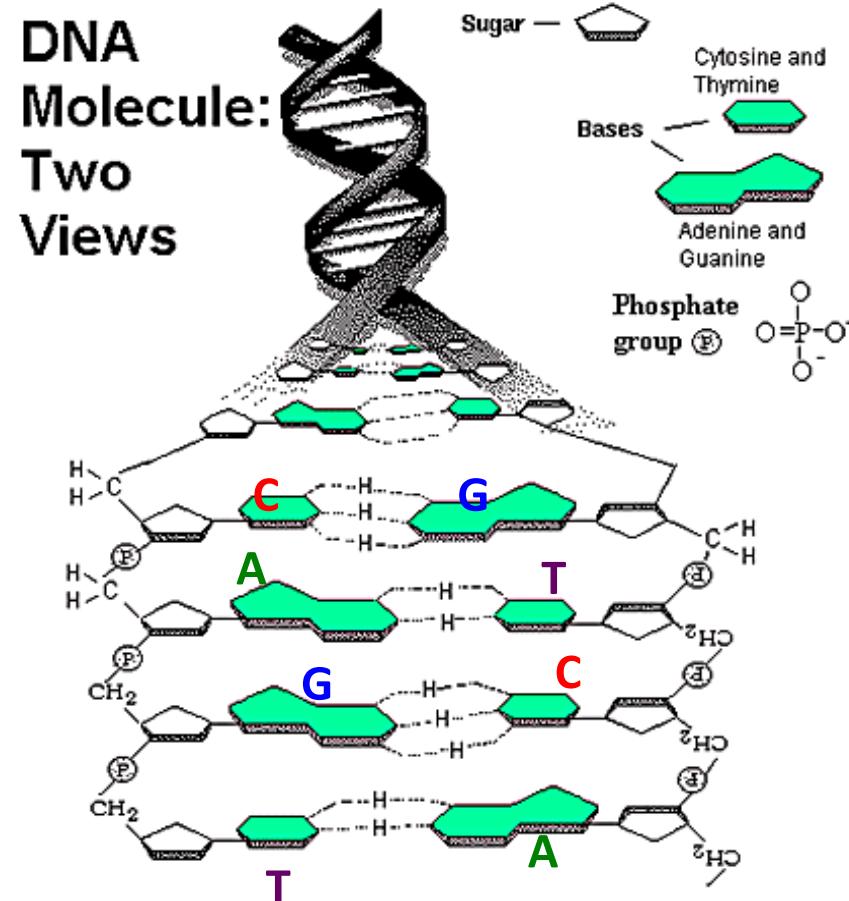
Messages

- DNA / genome size / evolution
- genes – concept
- gene regulation
- RNA species; junk
- RNA numbers
- RNA degradation



What is DNA?

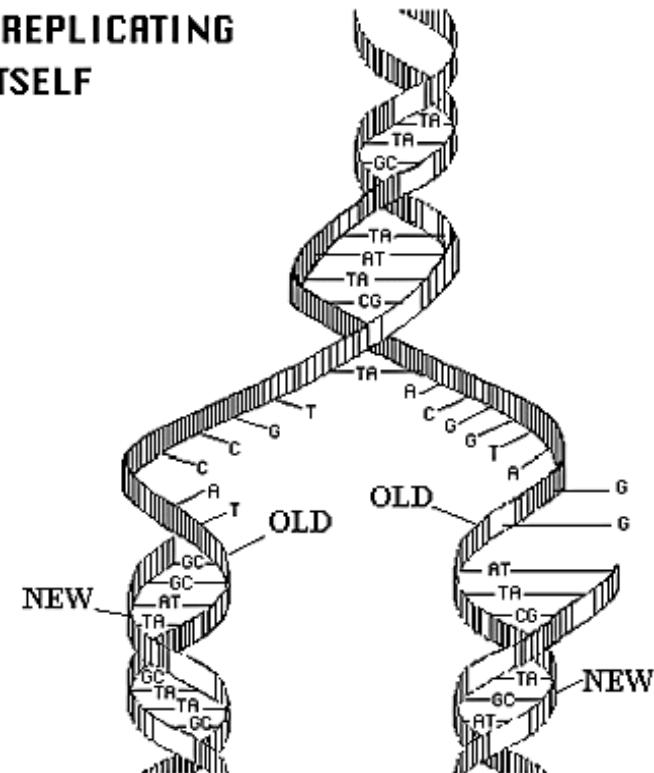
- A long backbone of sugars with nucleotides attached
 - Adenine (A)
 - Guanine (G)
 - Cytosine (C)
 - Thymine (T)
- It can form a self-complementary **double helix**
- In living organisms, the DNA is the carrier of the hereditary information, it is the **source code** of life



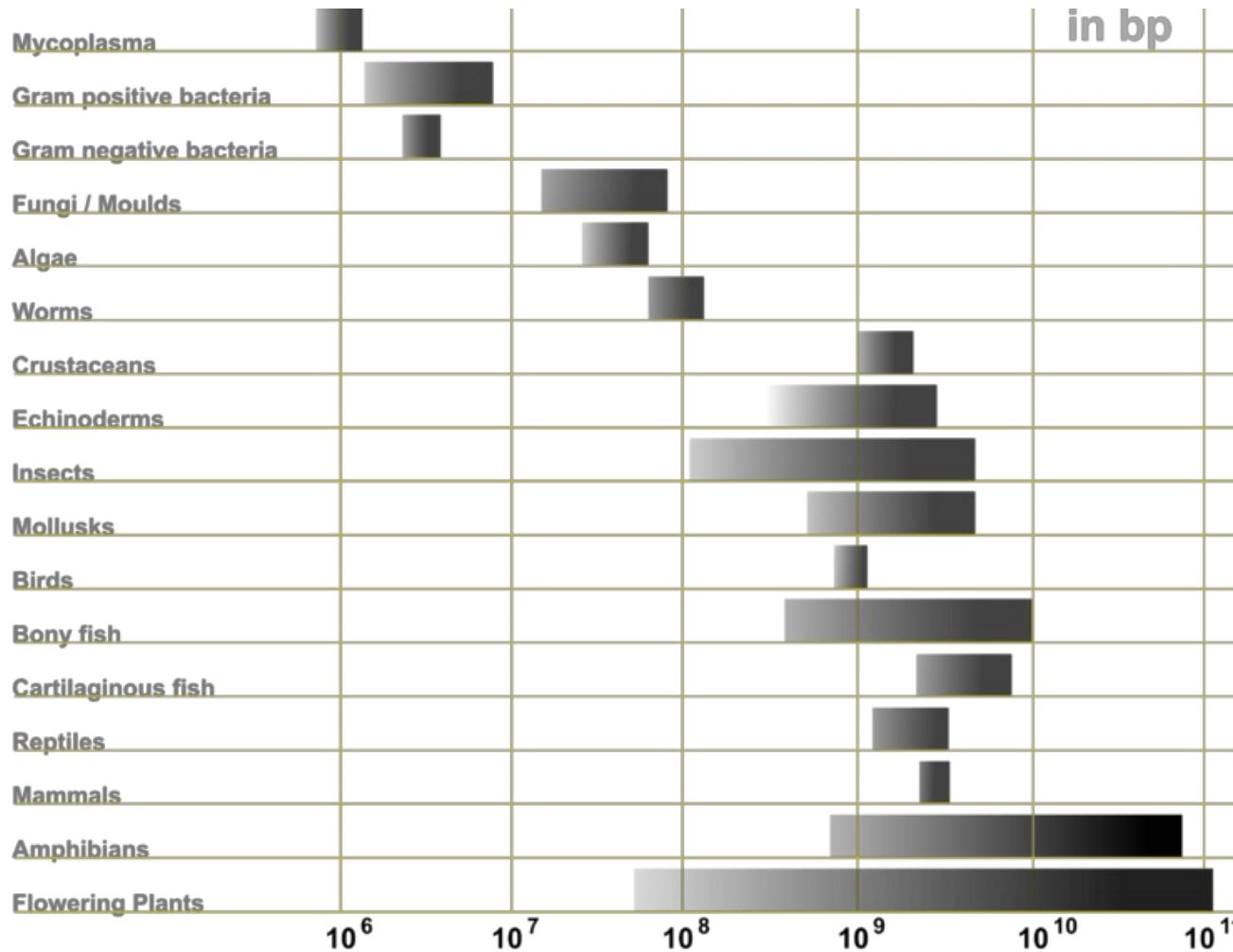
DNA replication

- The helix becomes unzipped and each strand acts as a template for a new complementary strand of DNA

DNA REPLICATING ITSELF



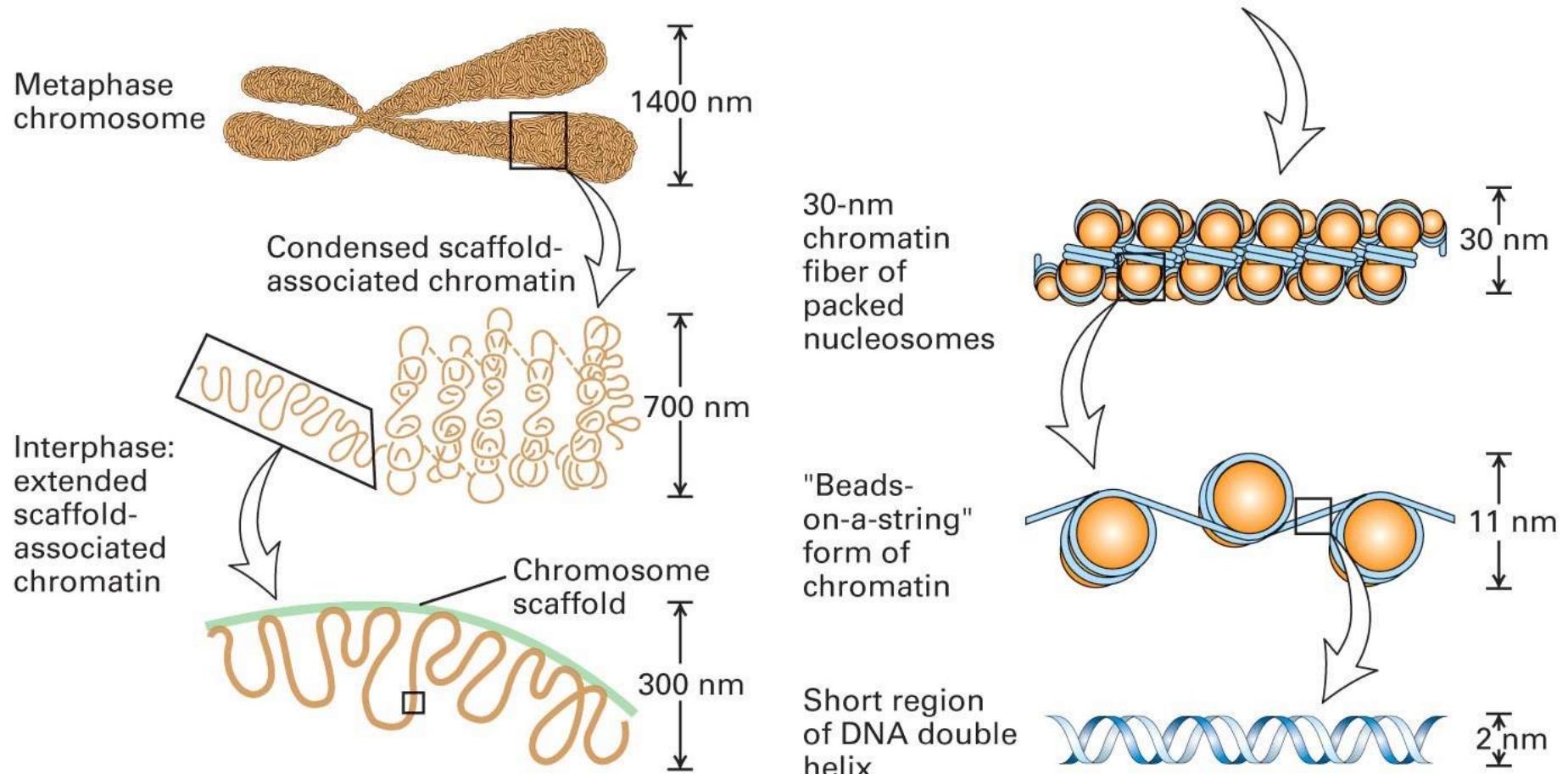
Genome Sizes



The size of the human genome is 3.2 billion base pairs. The length of this DNA string is approx. 2m. http://en.wikipedia.org/wiki/Genome_size

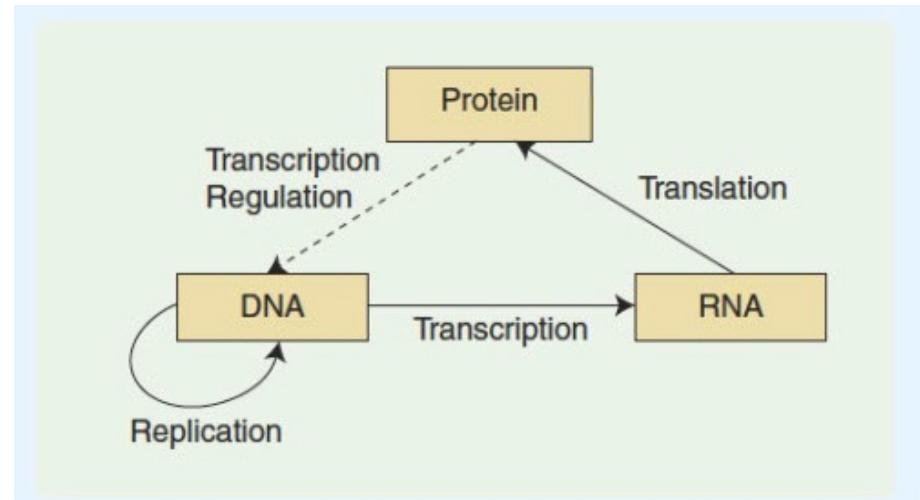
http://en.wikipedia.org/wiki/Genome_size

DNA Superstructure

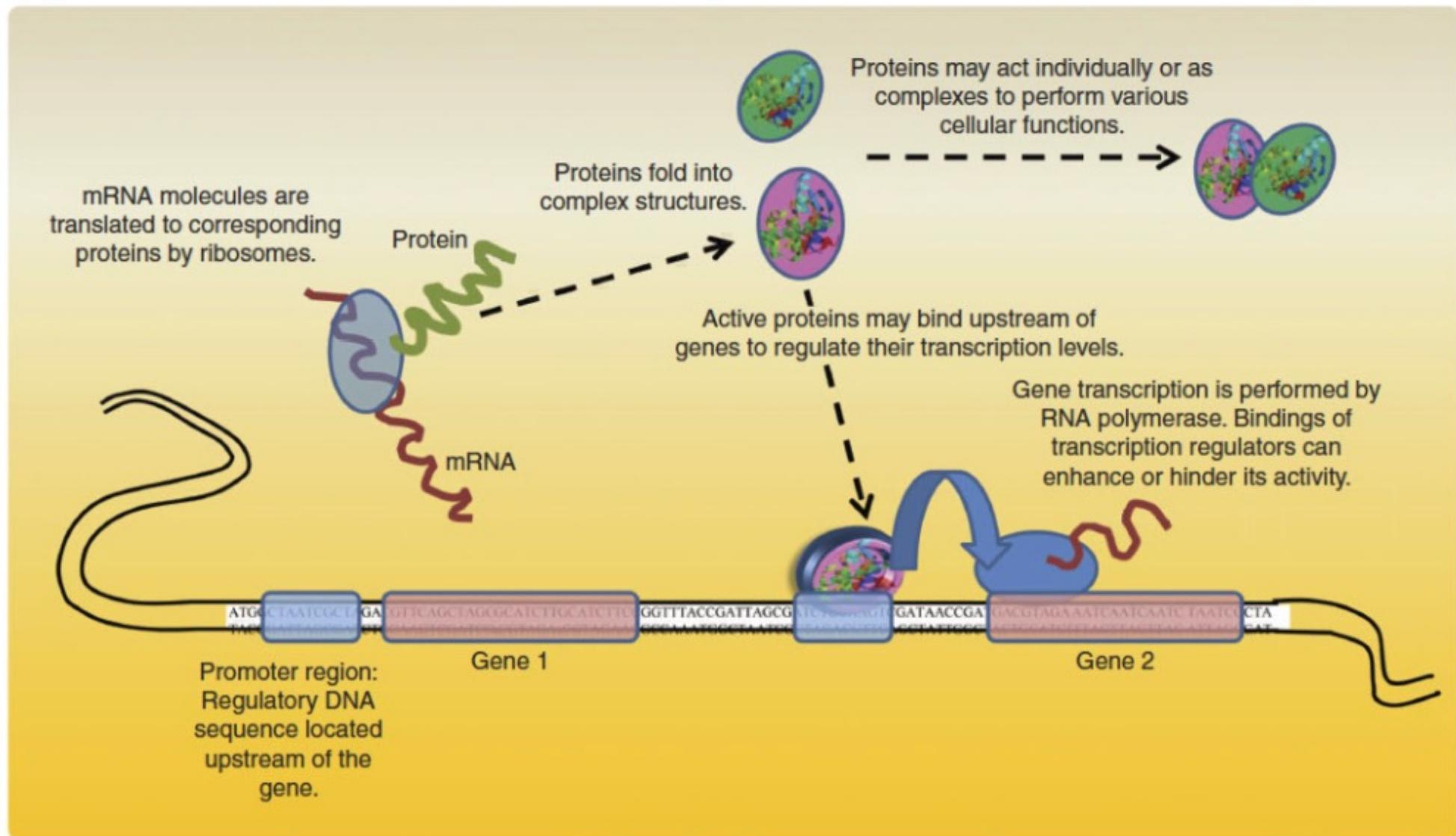


Genes

- A gene is a region of DNA (a locus) that controls a hereditary characteristic
 - protein coding genes are transcribed into messenger RNA which is then translated into protein.
 - In humans, protein coding genes constitute only ~2% of the human genome (precise number still under investigation)

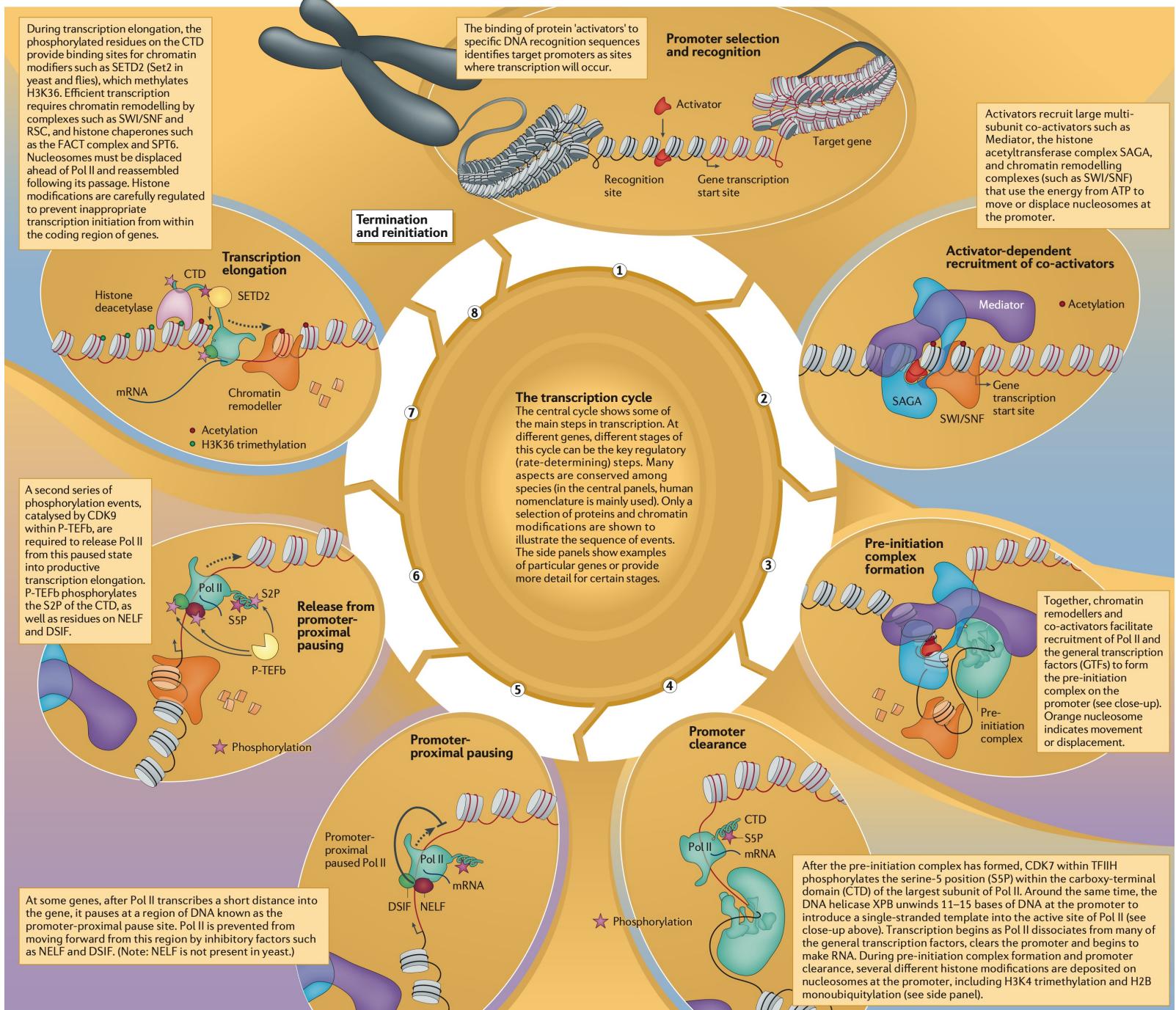


The Central Dogma





Transcription Cycle



Genome Sizes and Transcriptome Complexity

- larger genomes and complex organisms have emerged by **evolution**
- "**Nothing in Biology Makes Sense Except in the Light of Evolution**" --
Theodosius Dobzhansky, 1973
- genome sizes vary much (in some cases even for closely related species)
- genome sizes seems not strictly correlated with organisms
→ other functional elements vary: non-coding genes, repeats,
- number of protein coding genes is relatively constant across species
- C. elegans worm has only ~1000 cells and about the same number of genes like humans
→ alternative splicing

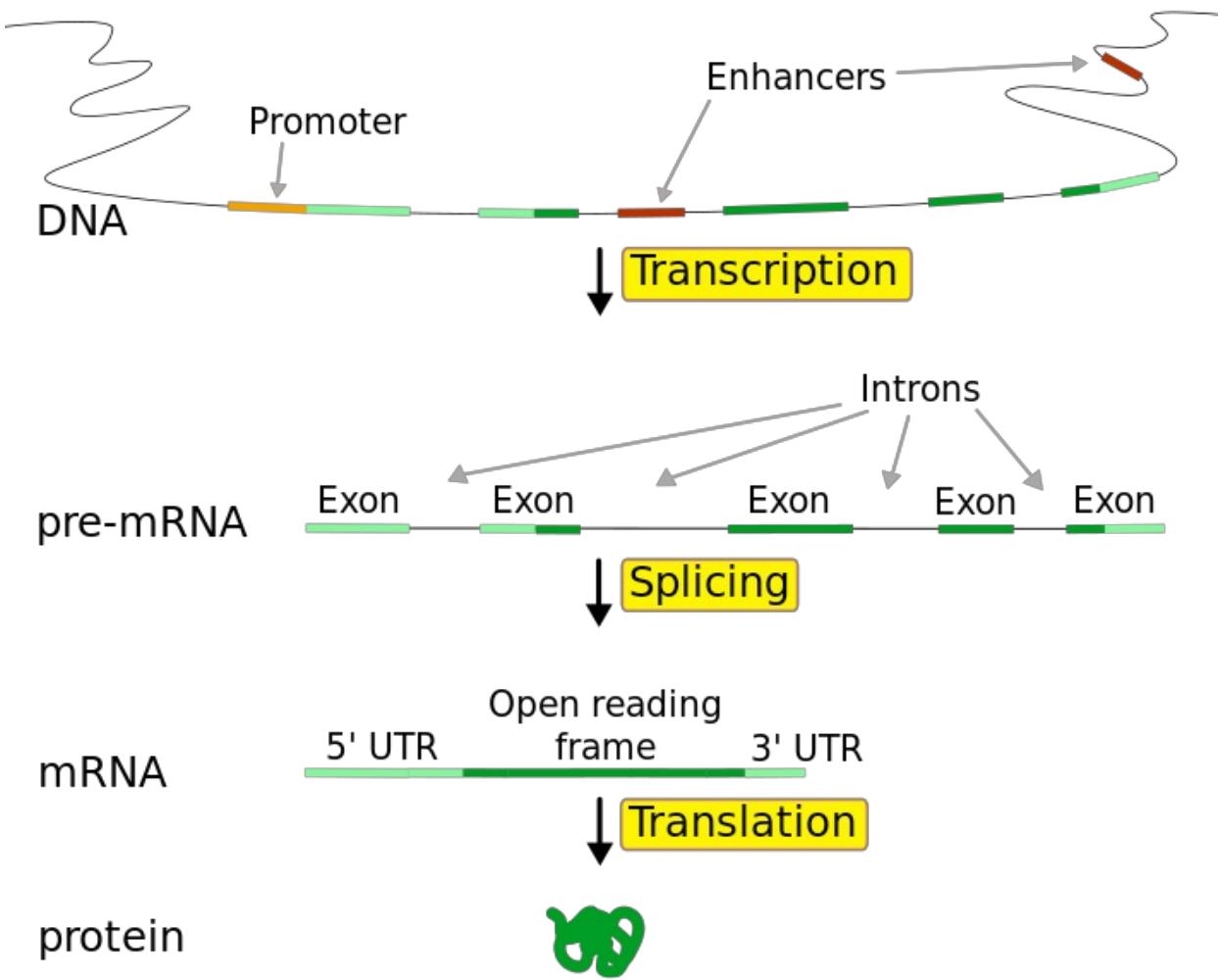
Transcription

The transcription process generates a messenger RNA molecule from a gene region.

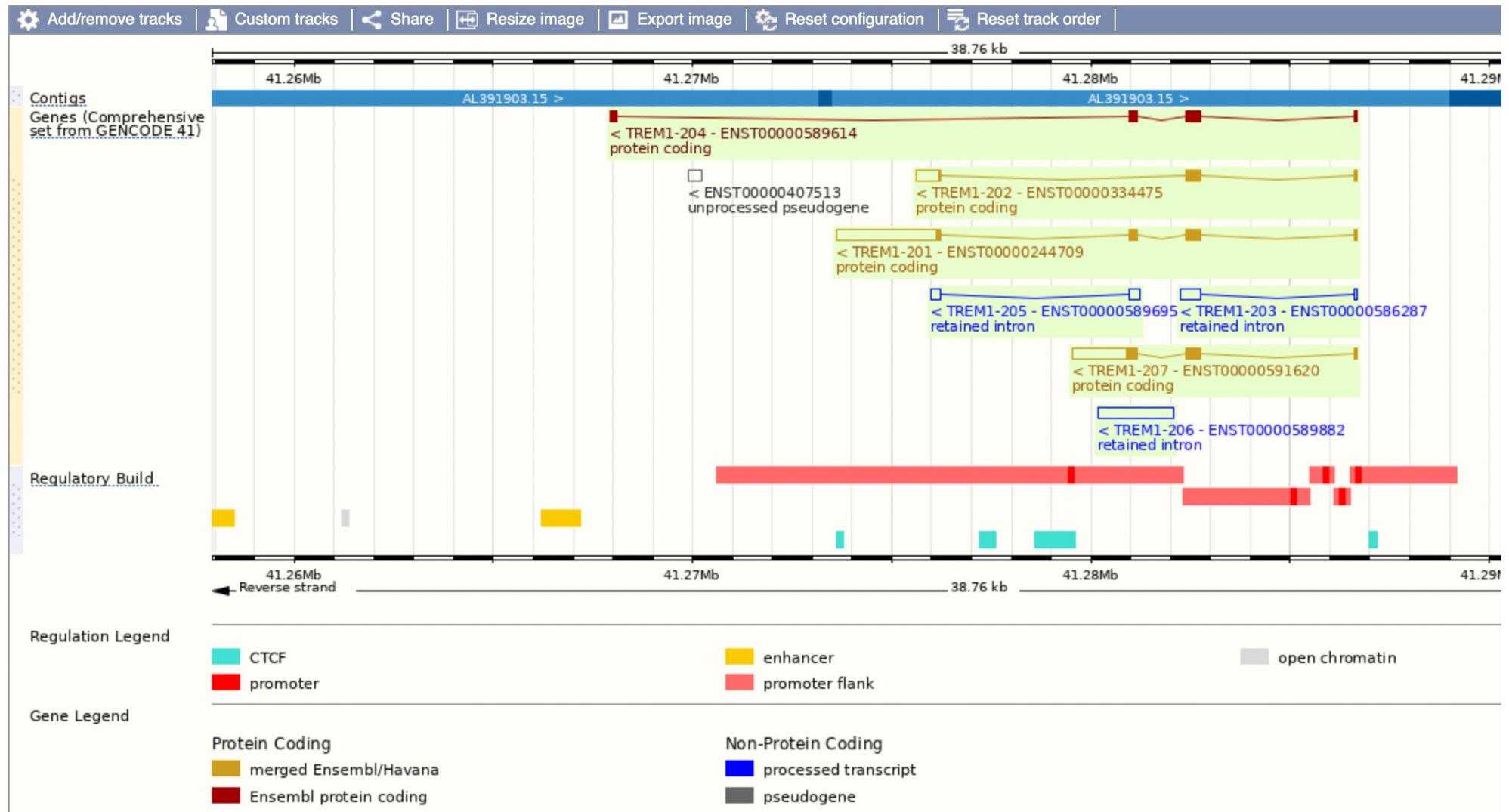
RNA is like DNA but

- the sugar-phosphate is different: ribose instead of deoxyribose
 - In all places where the DNA has a T the RNA has a U (uracil)

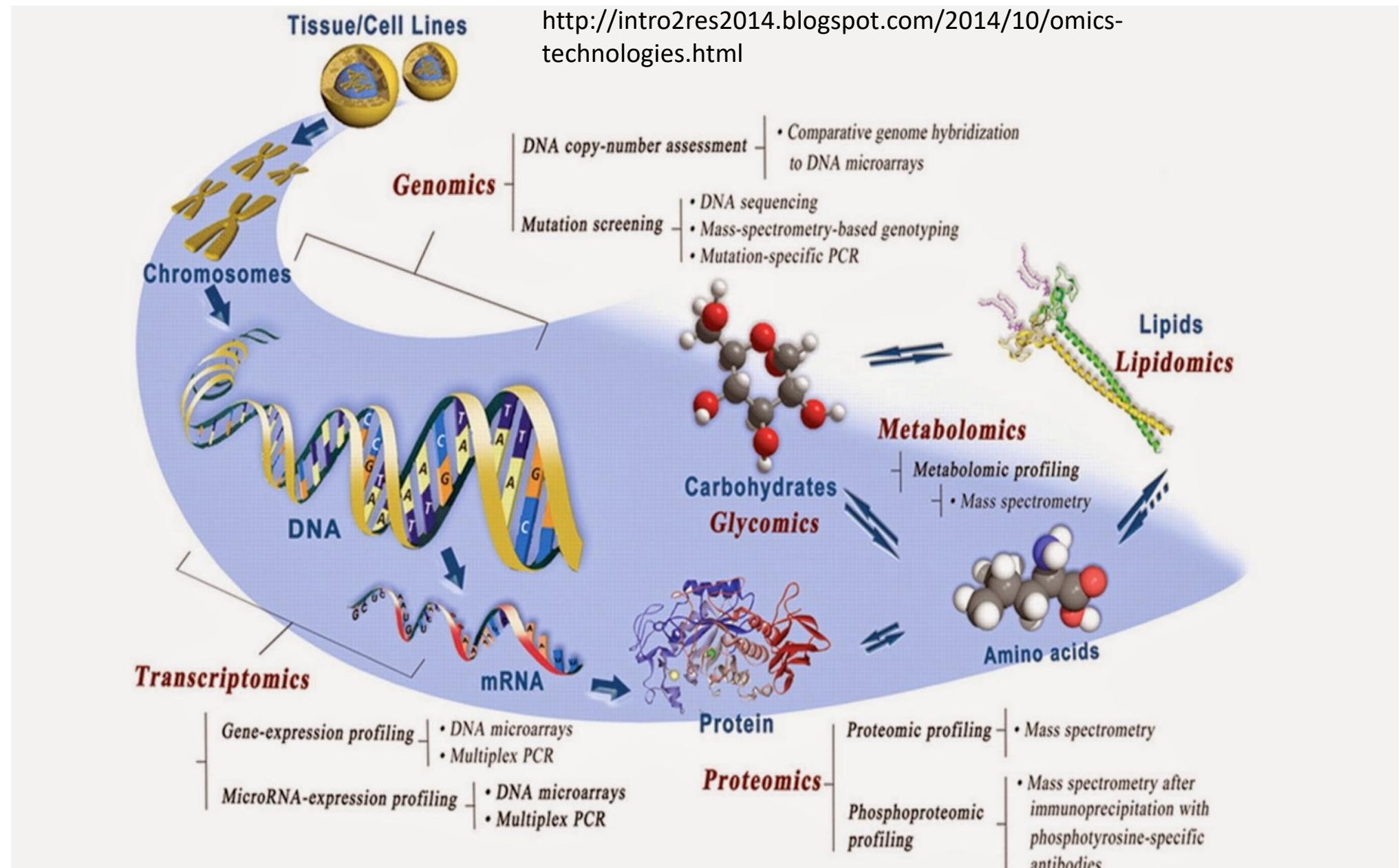
In higher organisms the protein coding sequences (exons) are interspersed by non-coding sequences (introns) which are spliced out.



Gene Structure



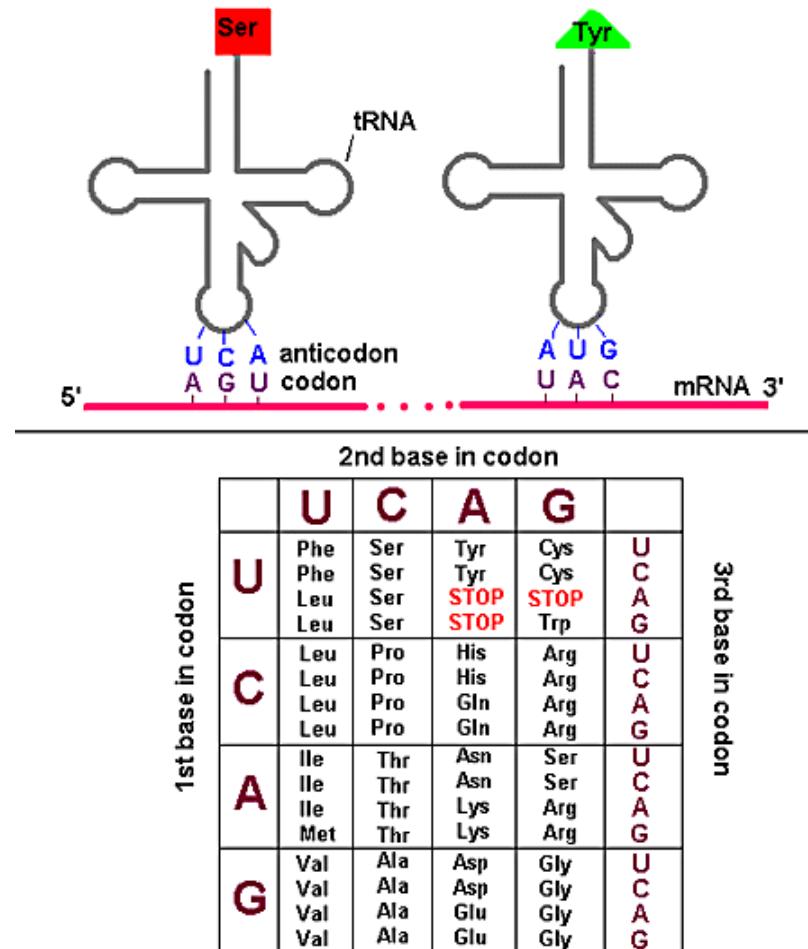
- introns are much longer than exons
- humans: typical gene length(exonic): 1000 – 2000 bases



Translation: The Genetic Code

The translation process generates a protein based on the information in the messenger RNA

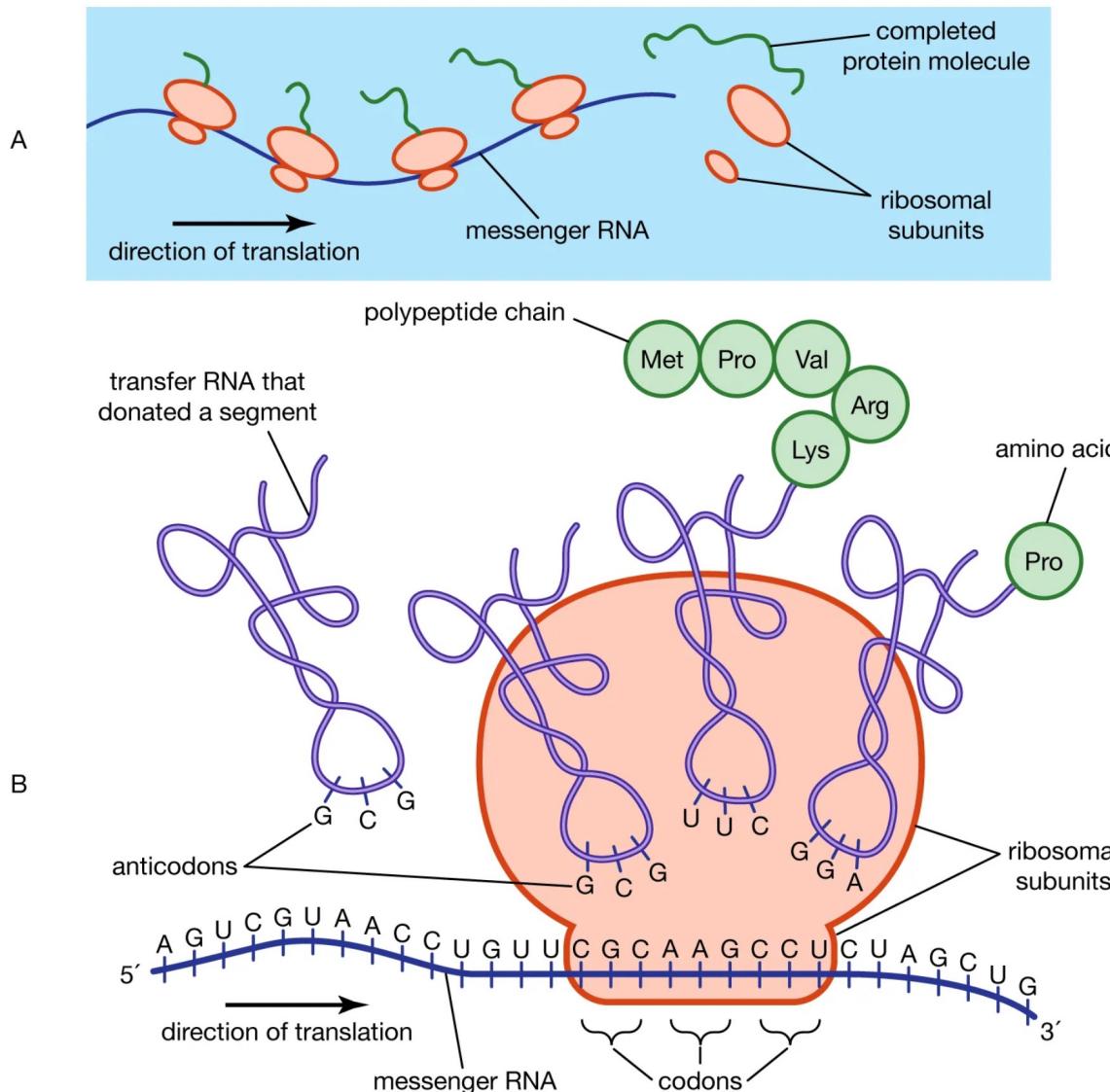
- A protein is a linear polymer of amino acids linked together by peptide bonds.
- Proteins are the main functional chemicals in the cell, carrying out many functions, for example catalysis of the reactions involved in metabolism.
- Proteins have a complex spatial structure

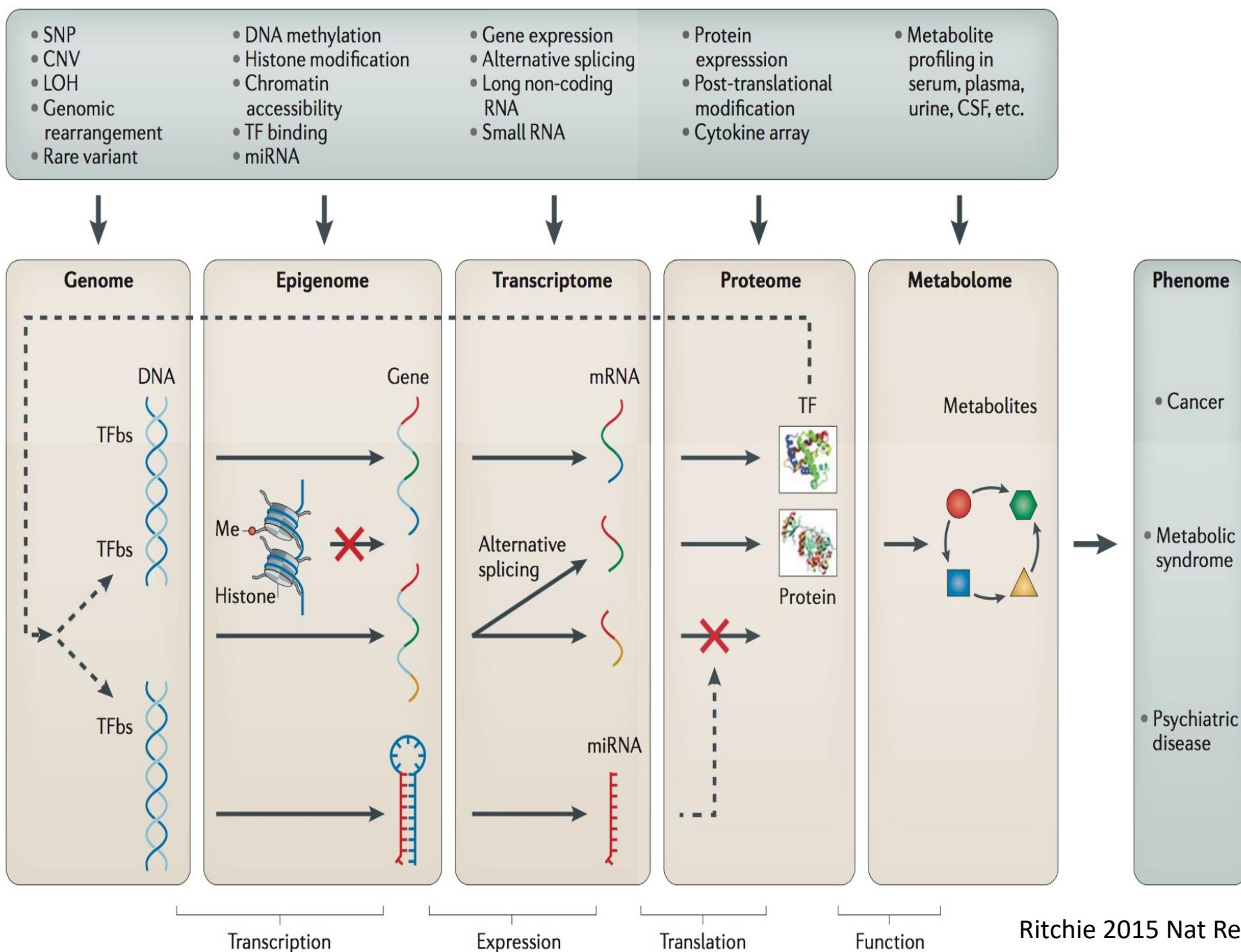


The Genetic Code



Translation

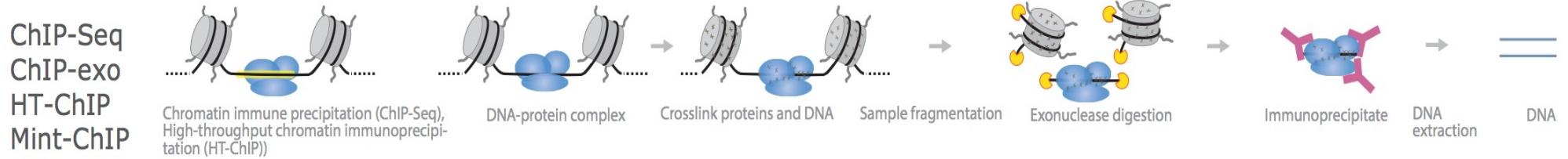




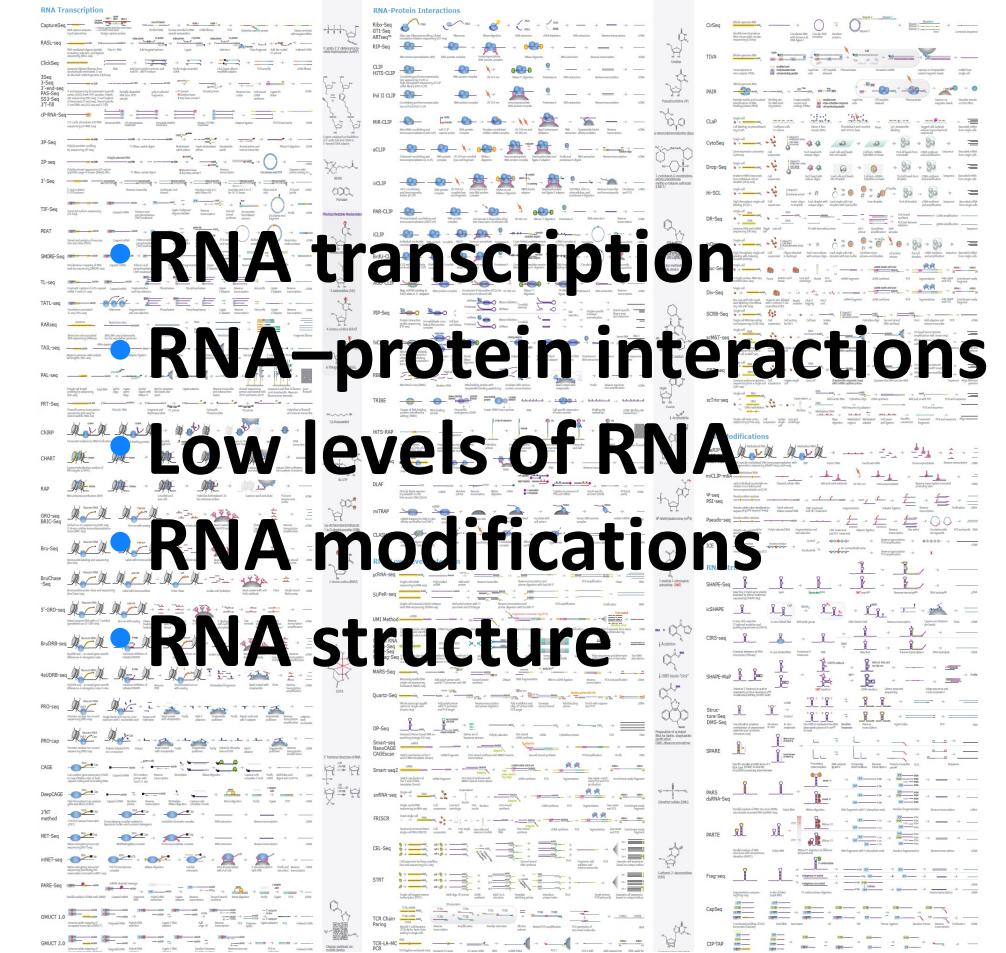
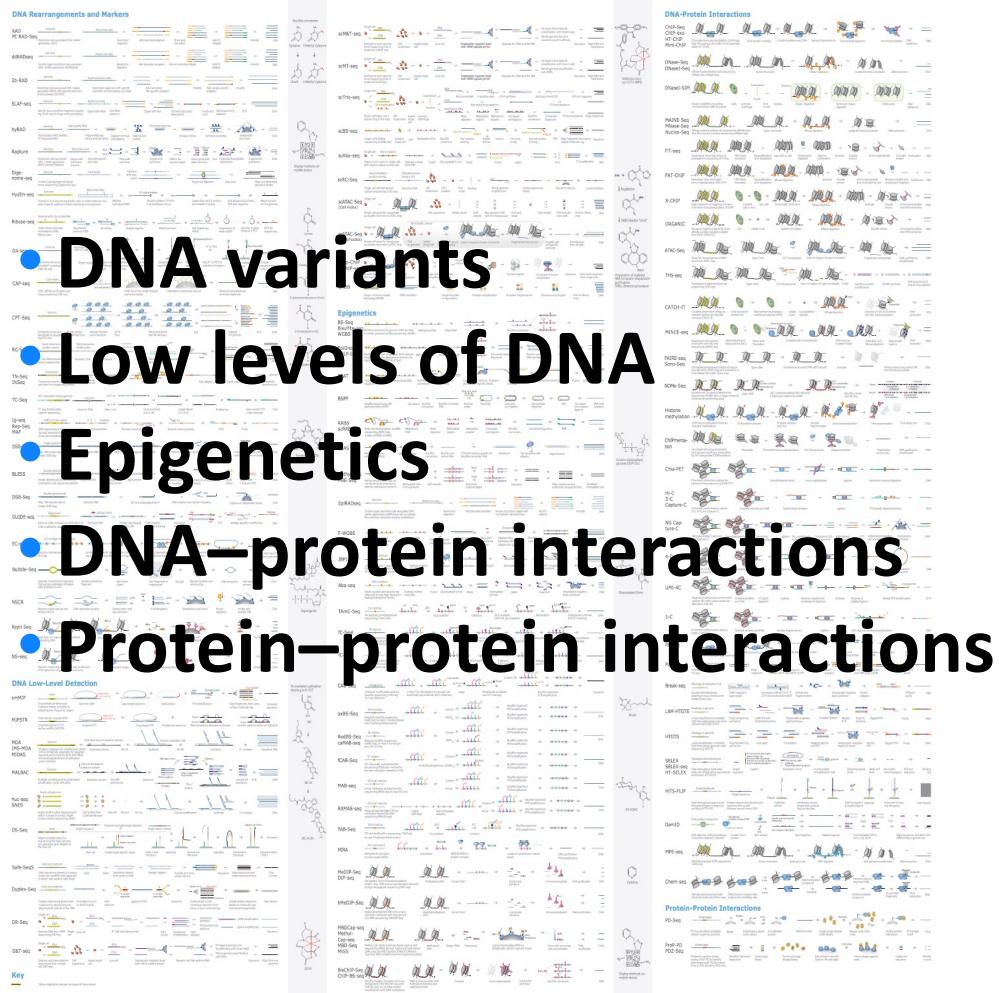
NGS Protocols

- Example: Preparation of DNA for a ChIP-seq experiment

DNA-Protein Interactions



- The preparation determines how sequenced reads have to be interpreted



NGS Reads

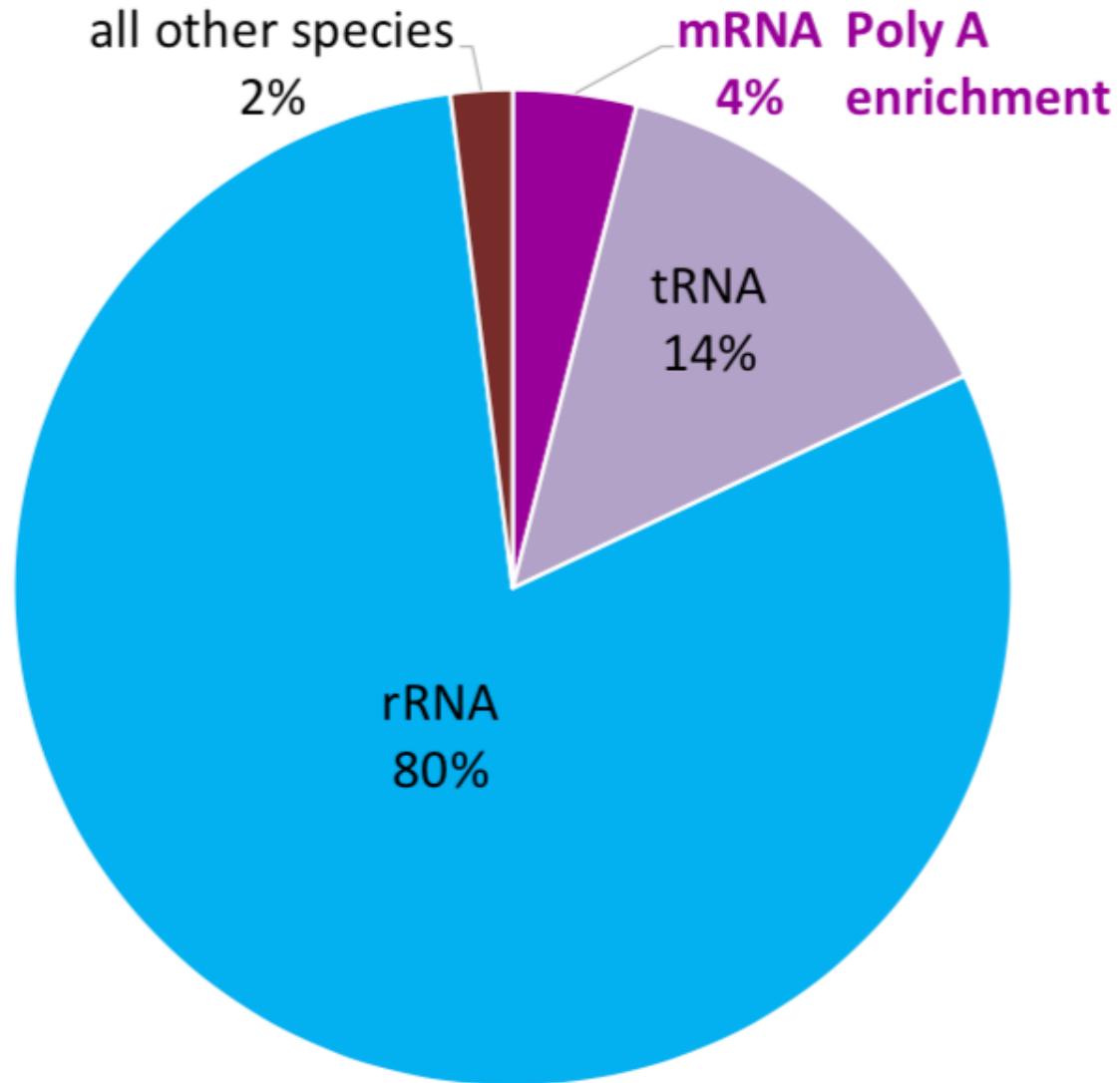


- Extract nucleotide material from specimens
- Select molecules
 - this is the crucial step that determines what information we can learn later from the data
- Add adapters/primers and make the molecules compatible with the sequencer
- Sequencing: Identify the nucleotide sequences

RNA-seq in numbers

- Typical numbers for human cells
- In biology everything is possible!
Be prepared to encounter cells that have 10x less or 10x more.
- 250'000 – 500'000 mRNA molecules per cell
- ~14% of total RNA is in nucleus
- RNA:DNA in nucleus 1:2
- typical mRNA molecule length ~1900nt

Prevalence of RNA species



- transfer RNA (tRNA)
length: 76-90 nt
 - ribosomal RNA (rRNA),
different subunits
 - 5S: 121 nt
 - 18S: 1869 nt
 - 28S: 5070 nt

mRNA abundance

Abundance class	Copies/cell	Number of different messages/cell	Abundance of each message
Low	5–15	11,000	<0.004%
Intermediate	200–400	500	<0.1%
High	12,000	<10	3%

- A highly expressed gene typically does not contribute more than 3% of all transcripts

RNA-seq experiment of bulk tissue

- ~ 1 Mio cells averaged
- 100 ng of total RNA
- > =100 billion molecules as starting material
- Steps:
 - fragmentation
 - amplification
 - sequencing (sampled readout of 20 Mio processed molecules)