

Some Basics of Molecular Biology

Hubert Rehrauer



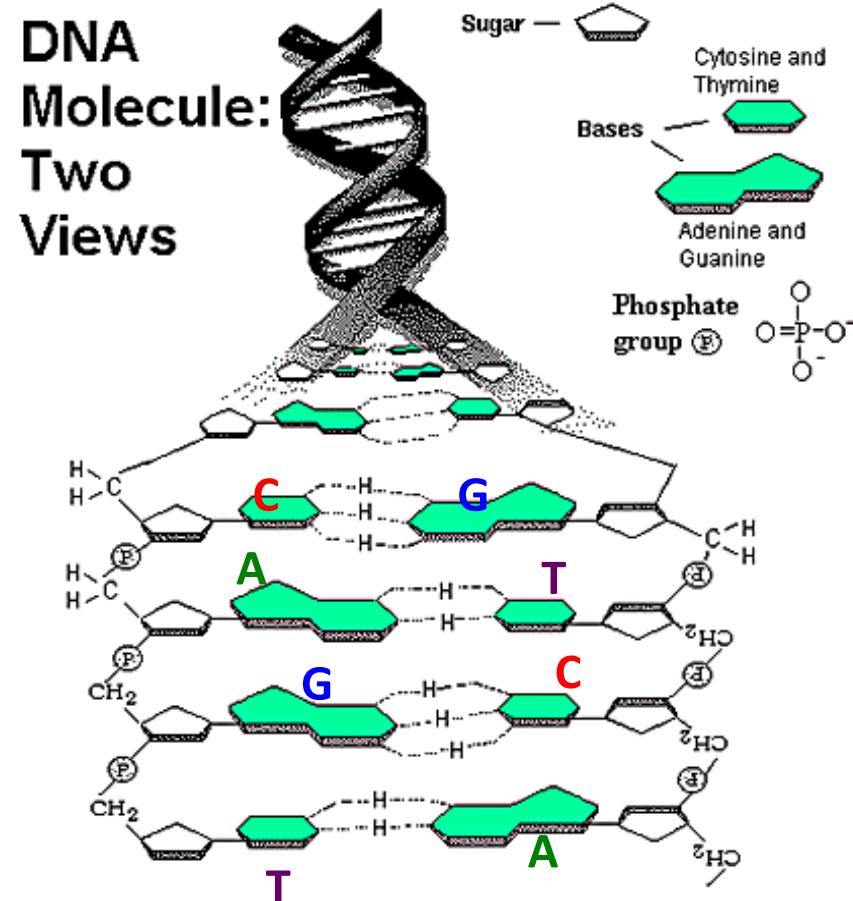
University of
Zurich^{UZH}

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

What is DNA?

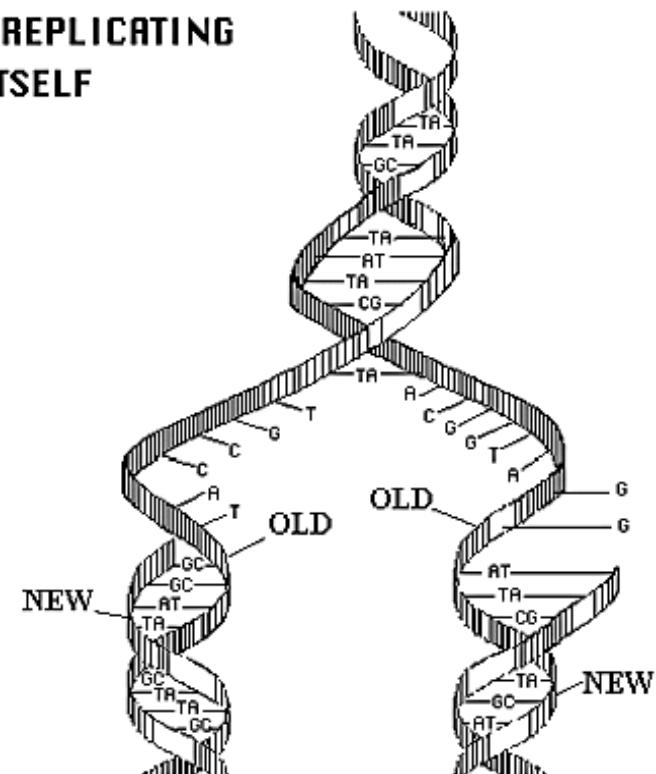
- A long backbone of sugars with nucleotides attached
 - Adenine (A)
 - Guanine (G)
 - Cytosine (C)
 - Thymine (T)
- It can form a self-complementary **double helix**
- In living organisms, the DNA is the carrier of the hereditary information, it is the **source code** of life



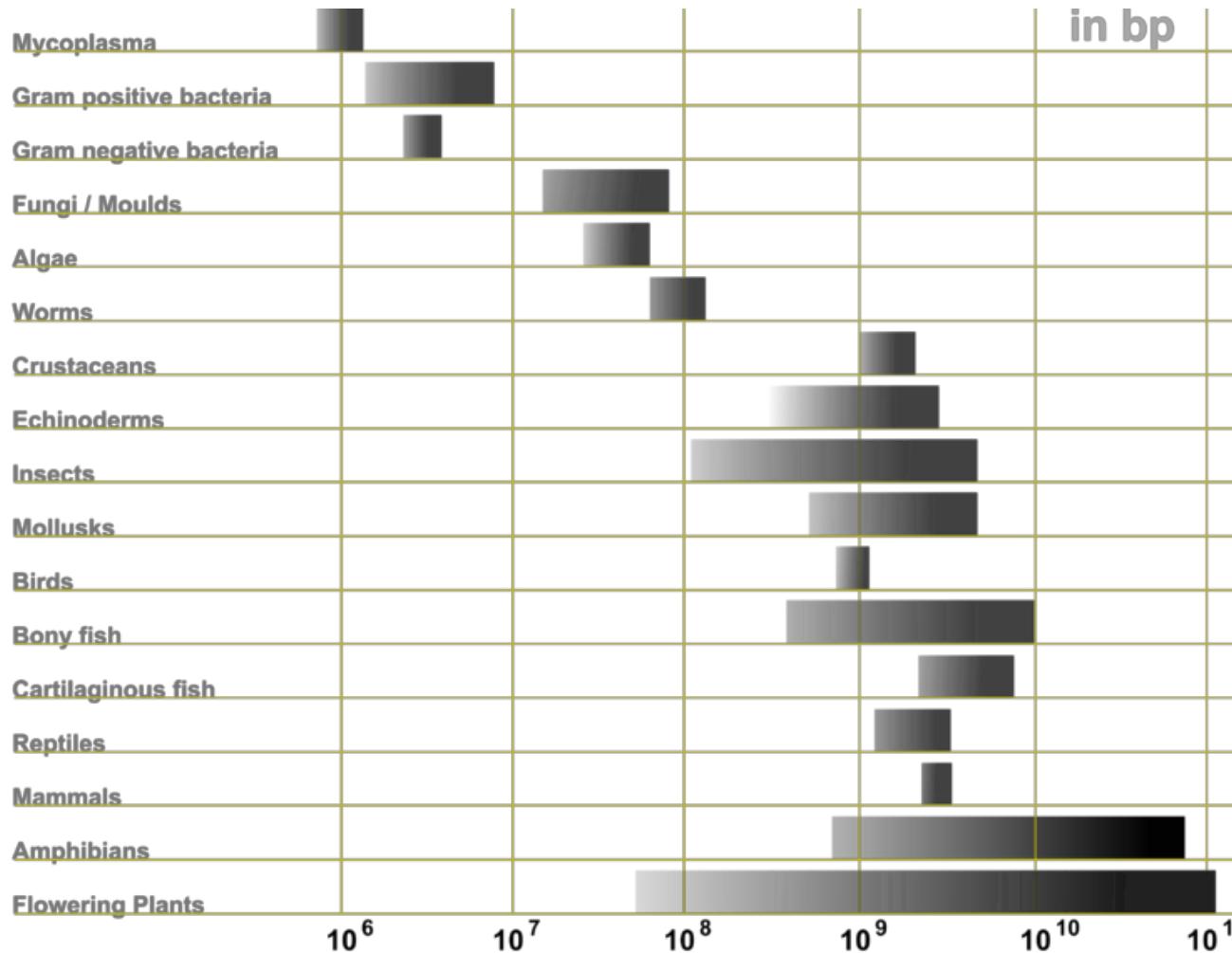
DNA replication

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

DNA REPLICATING
ITSELF



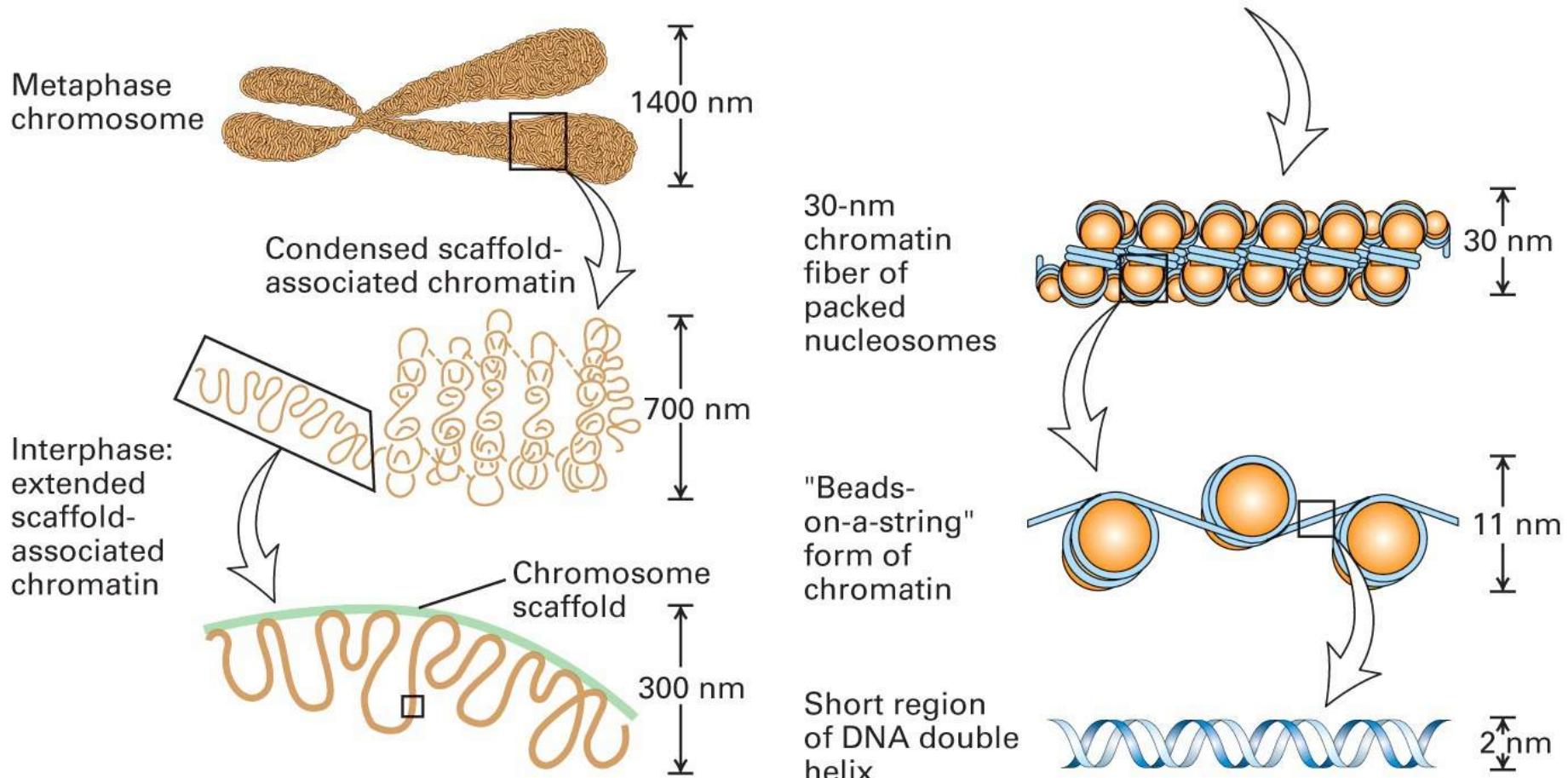
Genome Sizes

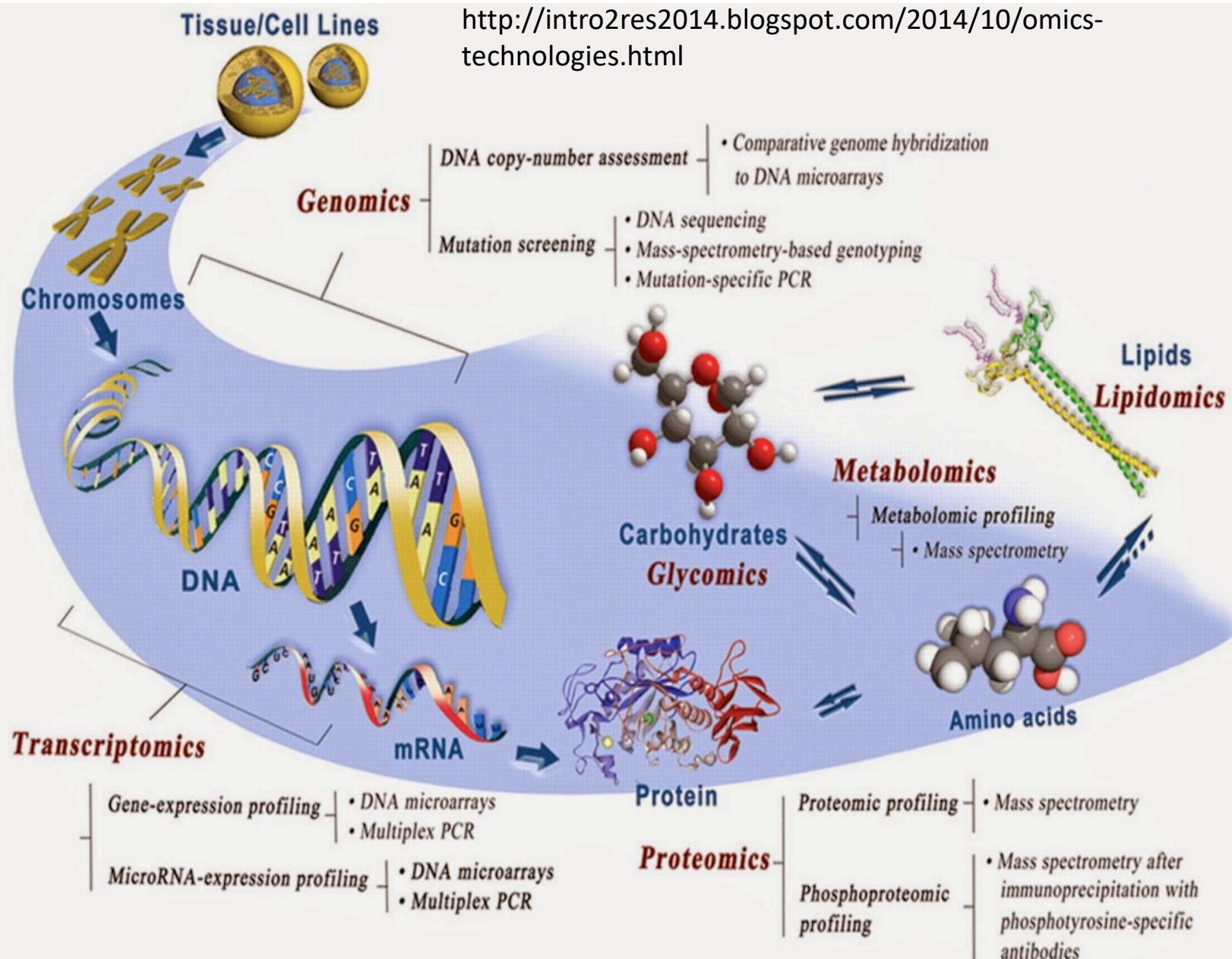


The size of the human genome is 3.2 billion base pairs. The length of this DNA string is approx. 2m.

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

DNA Superstructure

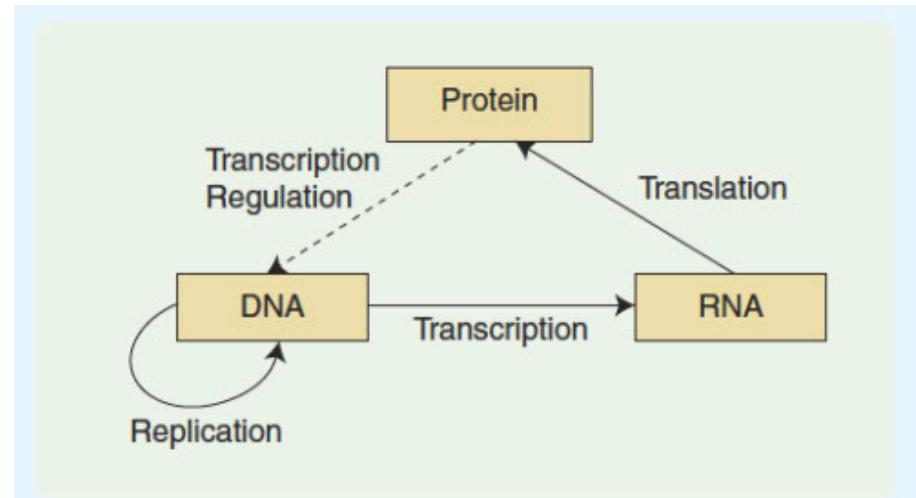




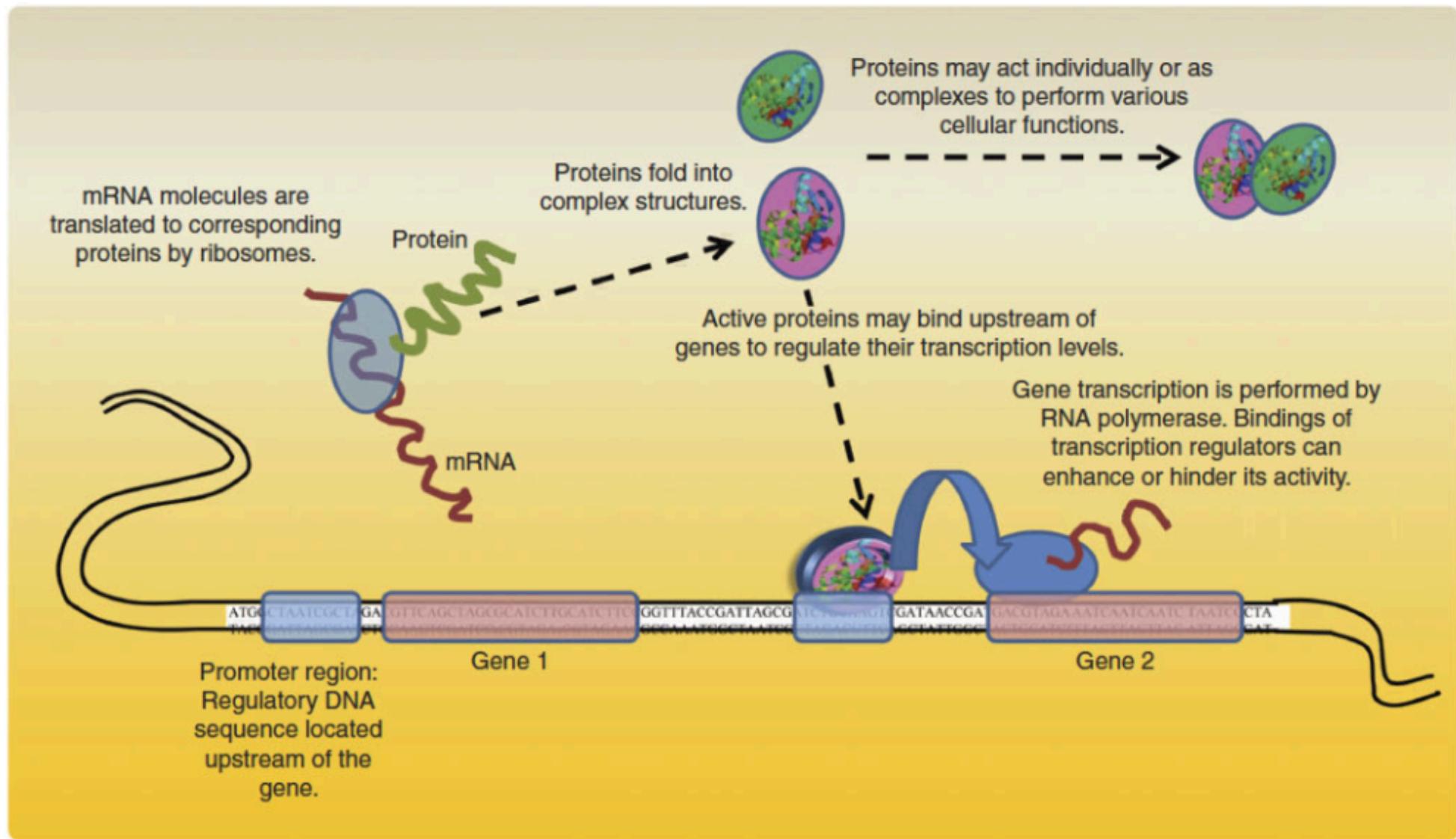
<http://intro2res2014.blogspot.com/2014/10/omics-technologies.html>

Genes

- A gene is a region of DNA that controls a hereditary characteristic
- Usually a gene is transcribed into a messenger RNA which is then translated into a protein.
- In humans genes constitute only ~3% of the human genome



The Central Dogma



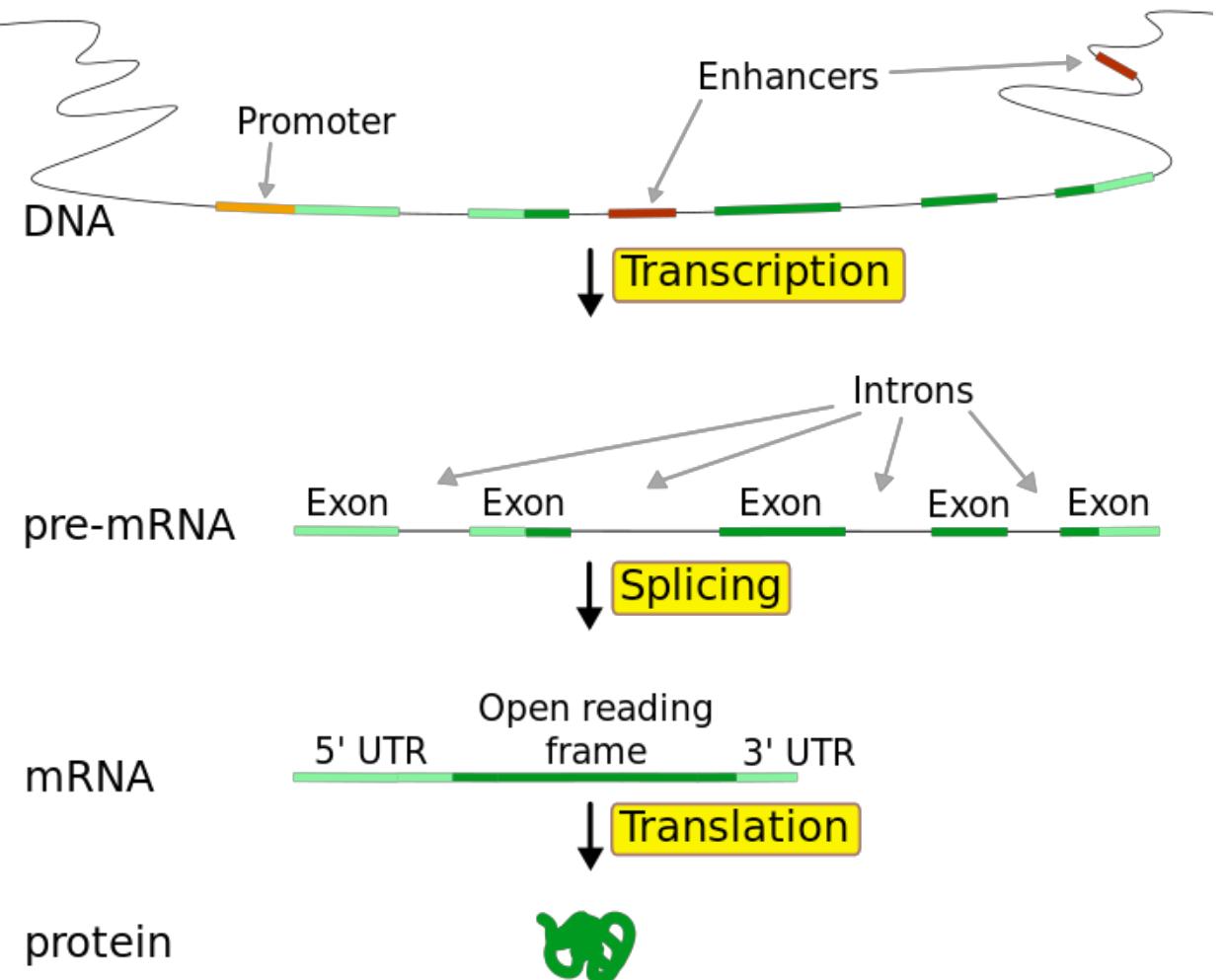
Transcription

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.

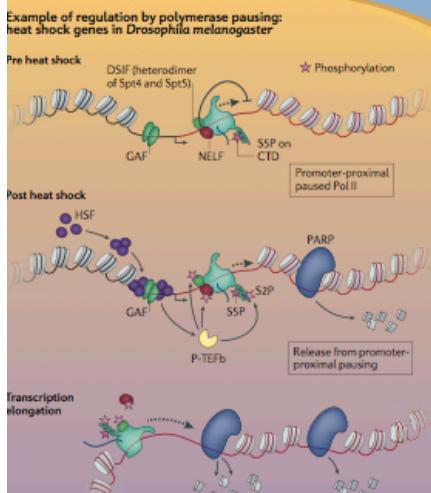
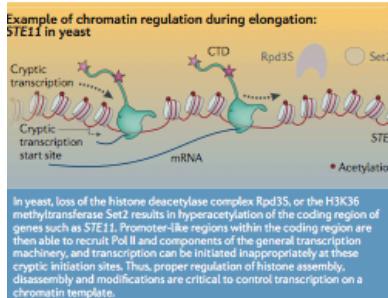


Chromatin remodelling and the transcription cycle

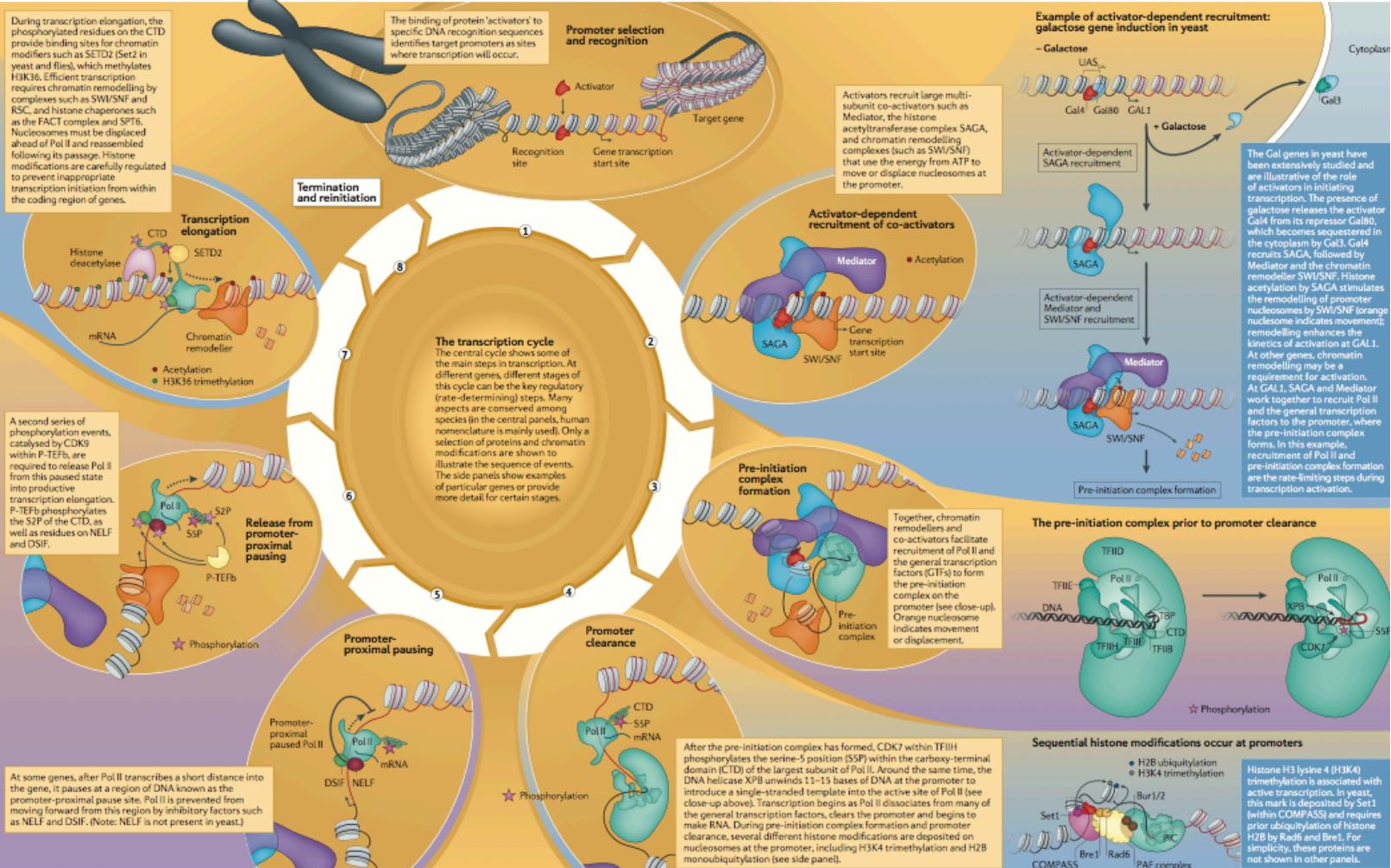
Vikki M. Weake and Jerry L. Workman

Transcription by RNA polymerase II (Pol II) occurs in the context of chromatin within a eukaryotic cell. Chromatin is generally inhibitory to transcription, so a variety of mechanisms are required to activate transcription from a nucleosomal template. One of the first steps is that large co-activator complexes interact with small activator proteins to identify gene promoters that are ready to be transcribed. Nucleosome remodelling complexes that use energy from ATP to move or displace

nucleosomes from DNA facilitate the recruitment and assembly of these complexes on the promoter and enable rapid gene activation. Even during transcription elongation, nucleosomes must be removed for efficient passage of the polymerase. Furthermore, these same nucleosomes must be reassembled rapidly and modified appropriately following passage of the polymerase to prevent inappropriate initiation of transcription from promoter-like elements within the coding region.

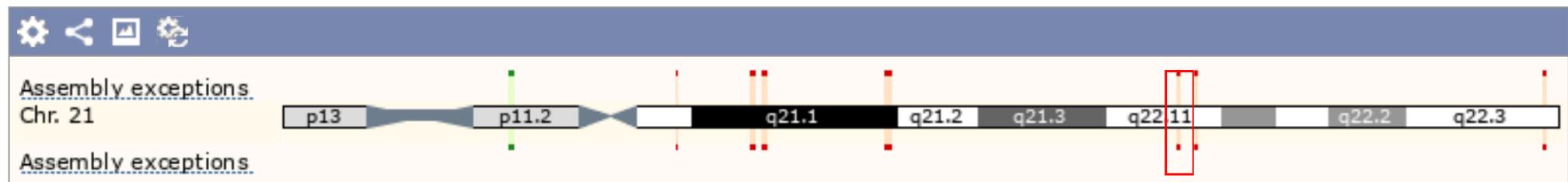


Heat shock genes in *Drosophila melanogaster* are rate-limited during early elongation. Prior to heat shock, GAF, co-activators and the GTFs are bound at Hsp70 and Pol II is present at the promoter-proximal pause site, where it sits in a poised state ready to resume productive elongation. Heat shock induces trimerization of the transcription factor HSF, which then binds to the promoter-proximal pause site. Pol II is prevented from moving forward from this region by inhibitory factors such as NELF and DSIF, as well as serine 2 of the CTD, resulting in release of Pol II into productive transcription elongation. PARP catalyses formation of ADP-ribose polymers, and along with HSF and GAF is required for nucleosome loss at Hsp70 following heat shock. Nucleosome loss precedes the passage of Pol II and facilitates gene activation.

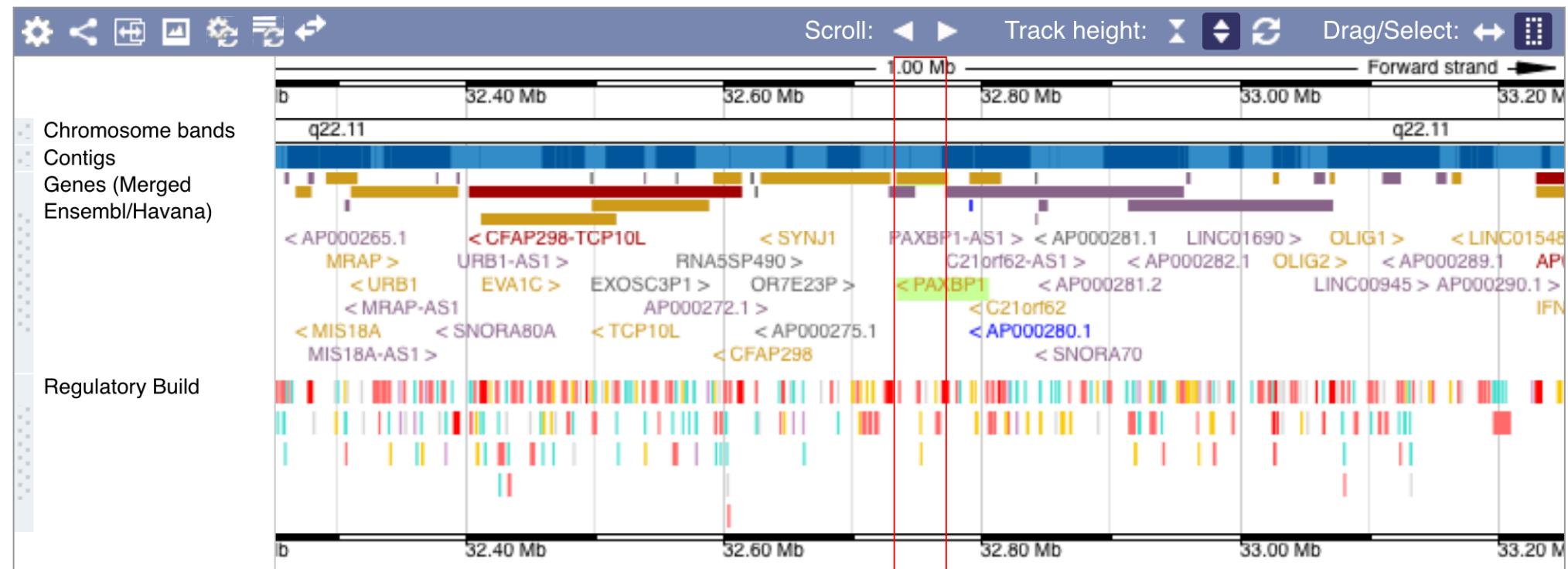


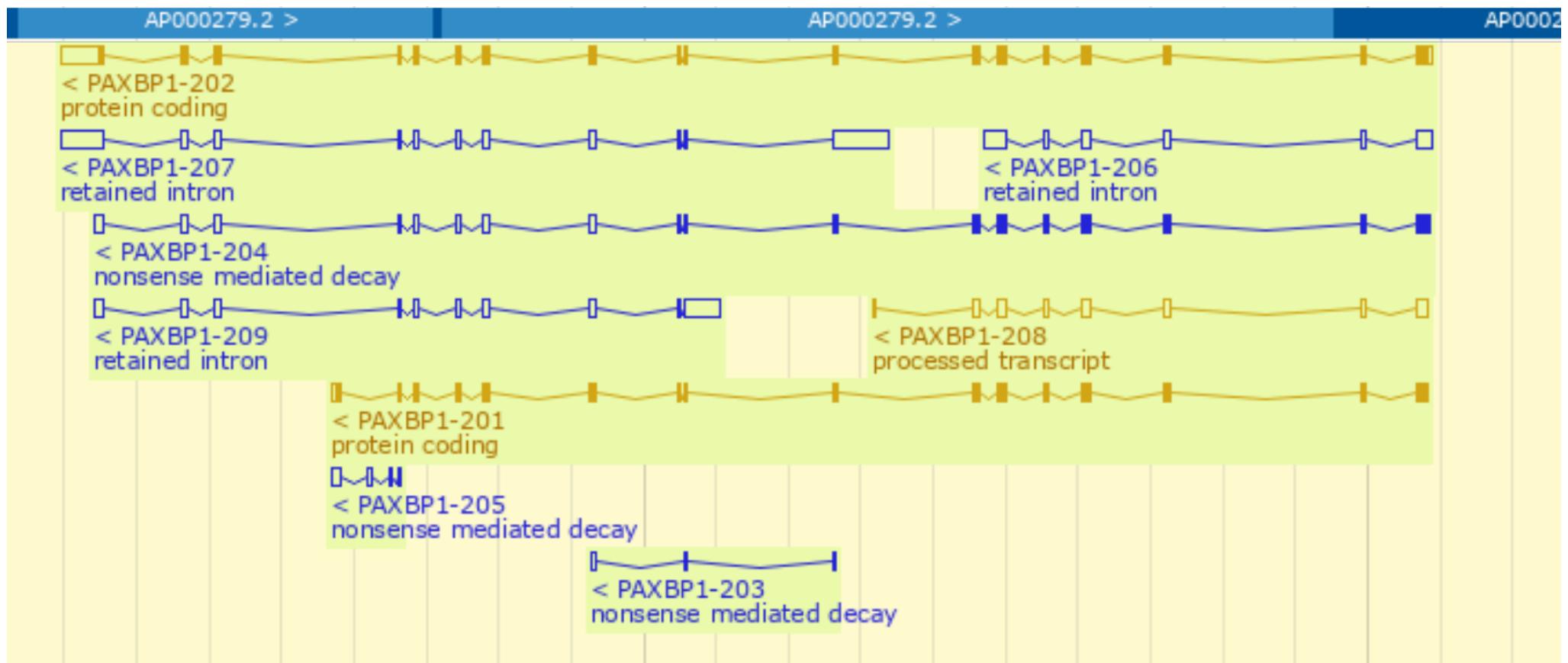
Gene: PAXBP1

Chromosome 21: 32,733,142-32,772,549



Region in detail ?





Protein Coding
 Ensembl protein coding
 merged Ensembl/Havana

CTCF
 Promoter Flank

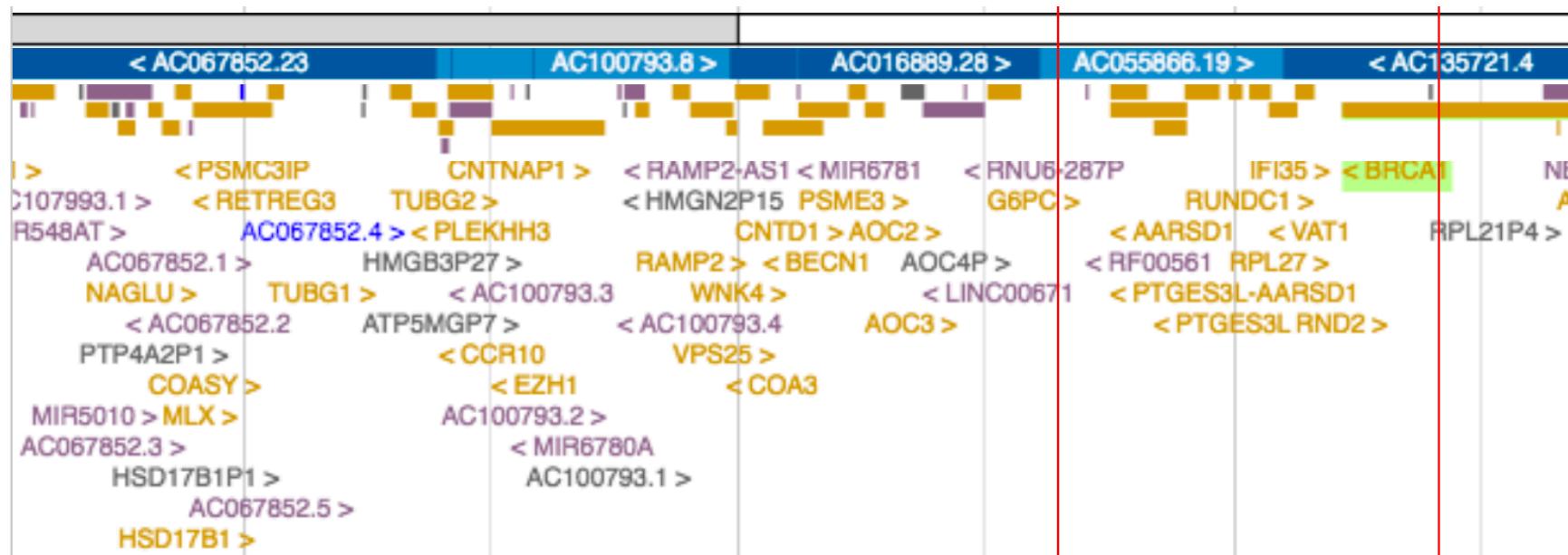
Non-Protein Coding
 processed transcript
 pseudogene
 RNA gene

Enhancer
 Transcription Factor Binding Site

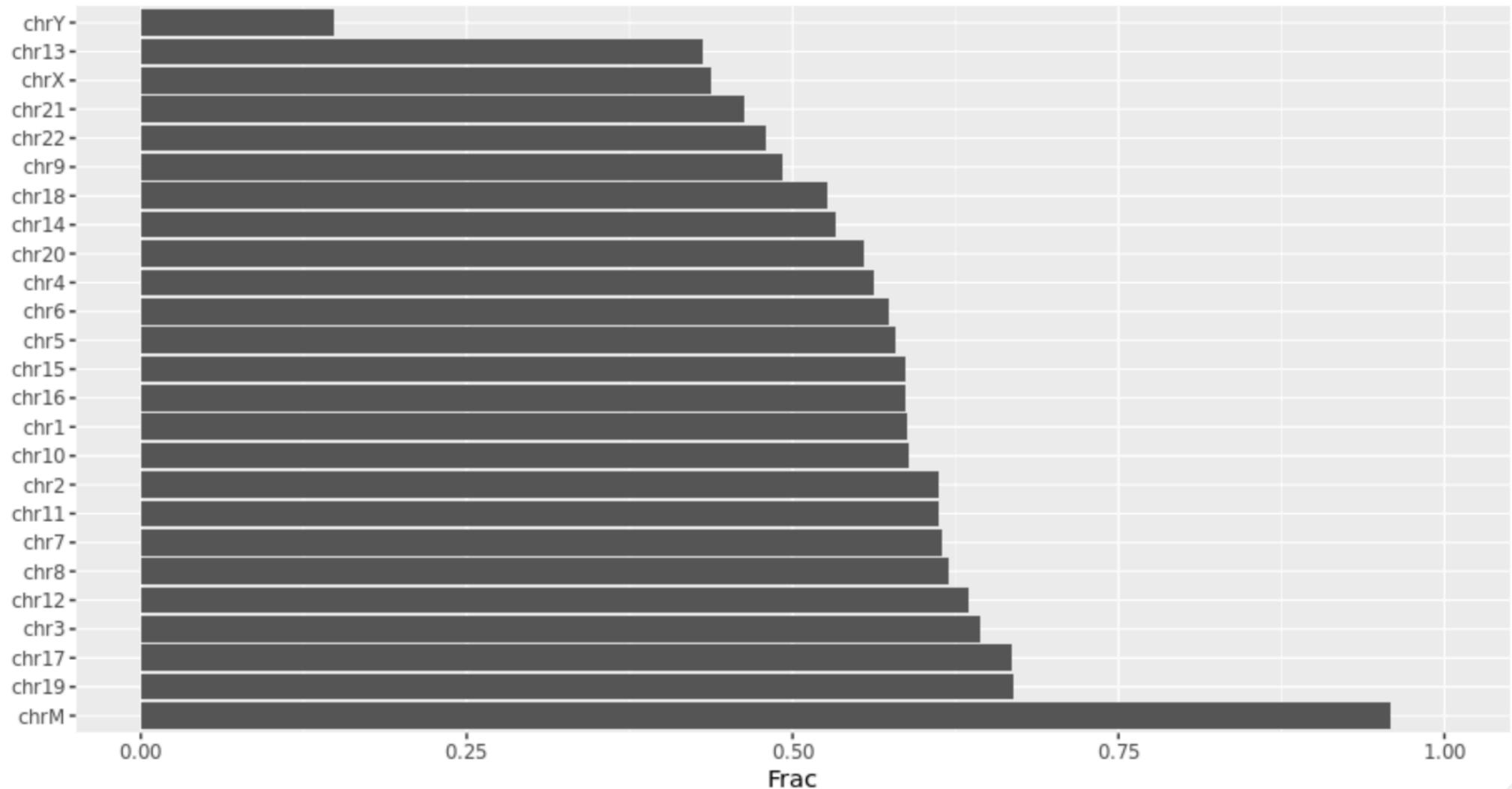
PAXBP1

Promoter

Chromosome bands
 Contigs
 Genes
 (Comprehensive set
 from GENCODE 31)



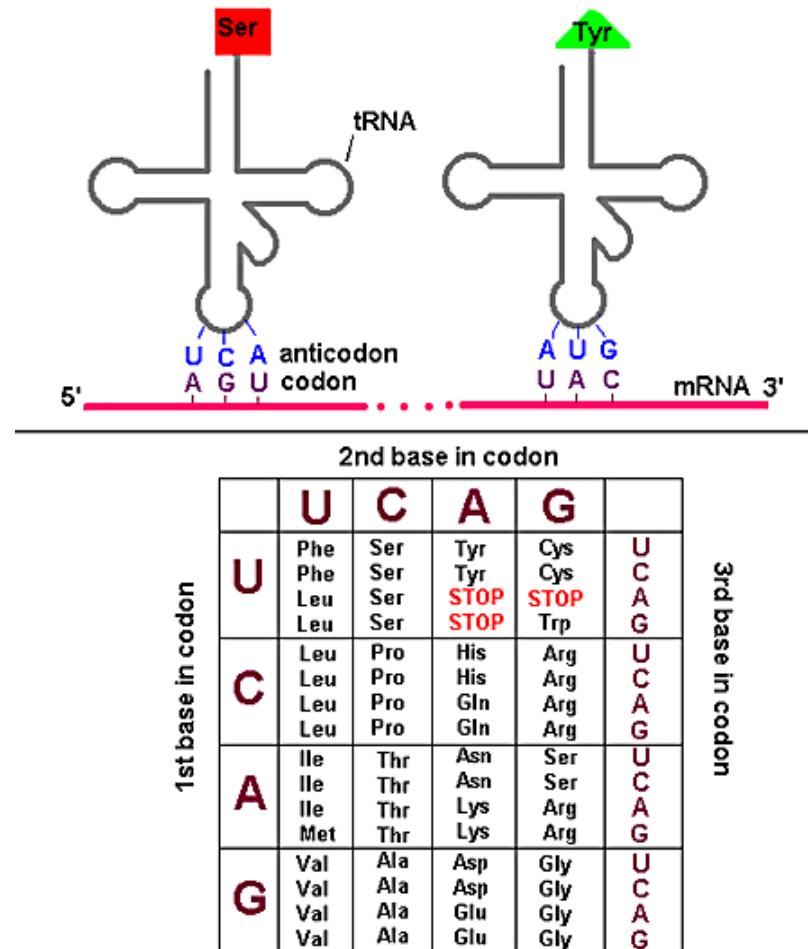
Fraction of Chromosomes that can be transcribed



Translation: The Genetic Code

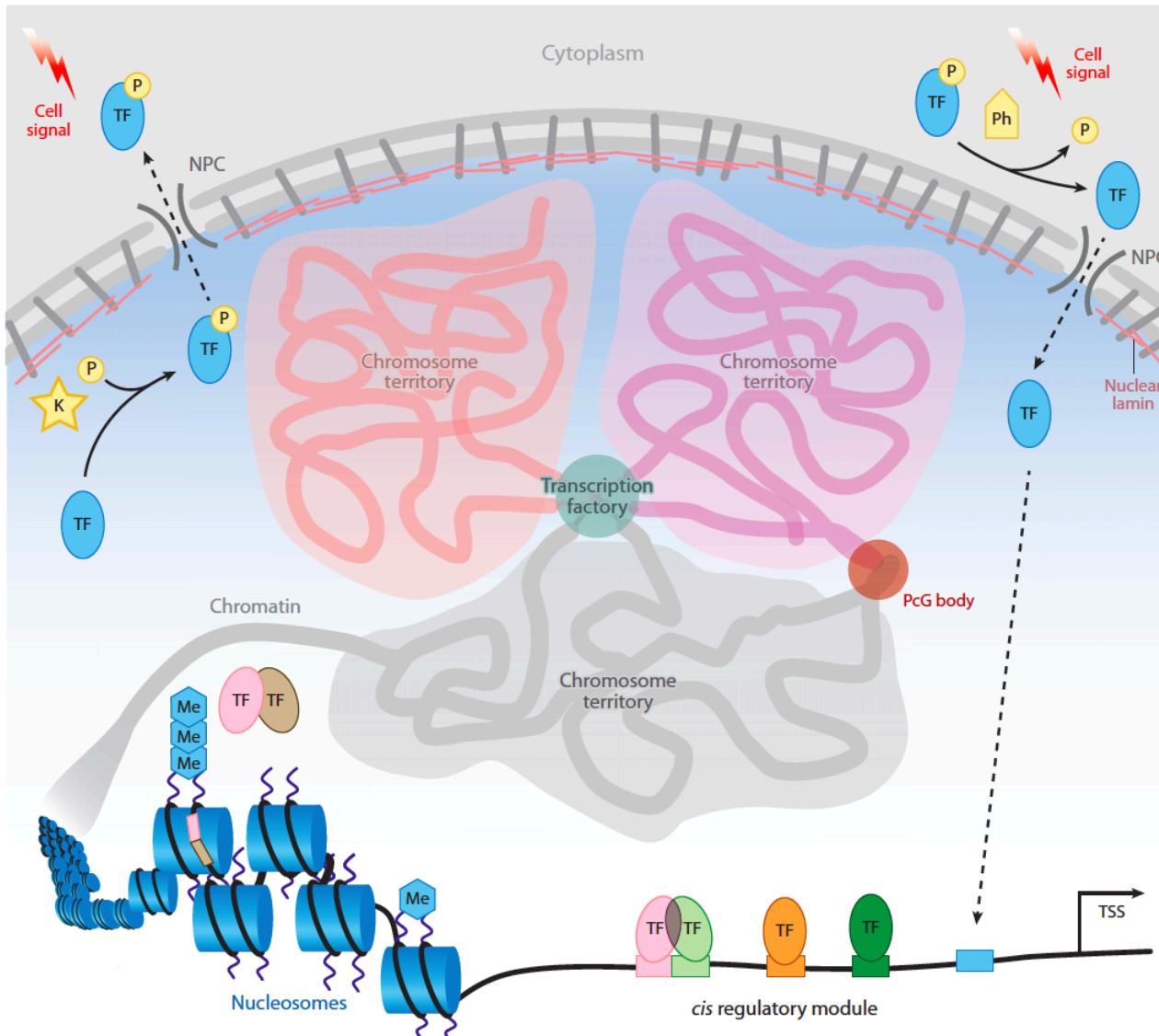
The translation process generates a protein based on the information in the messenger RNA

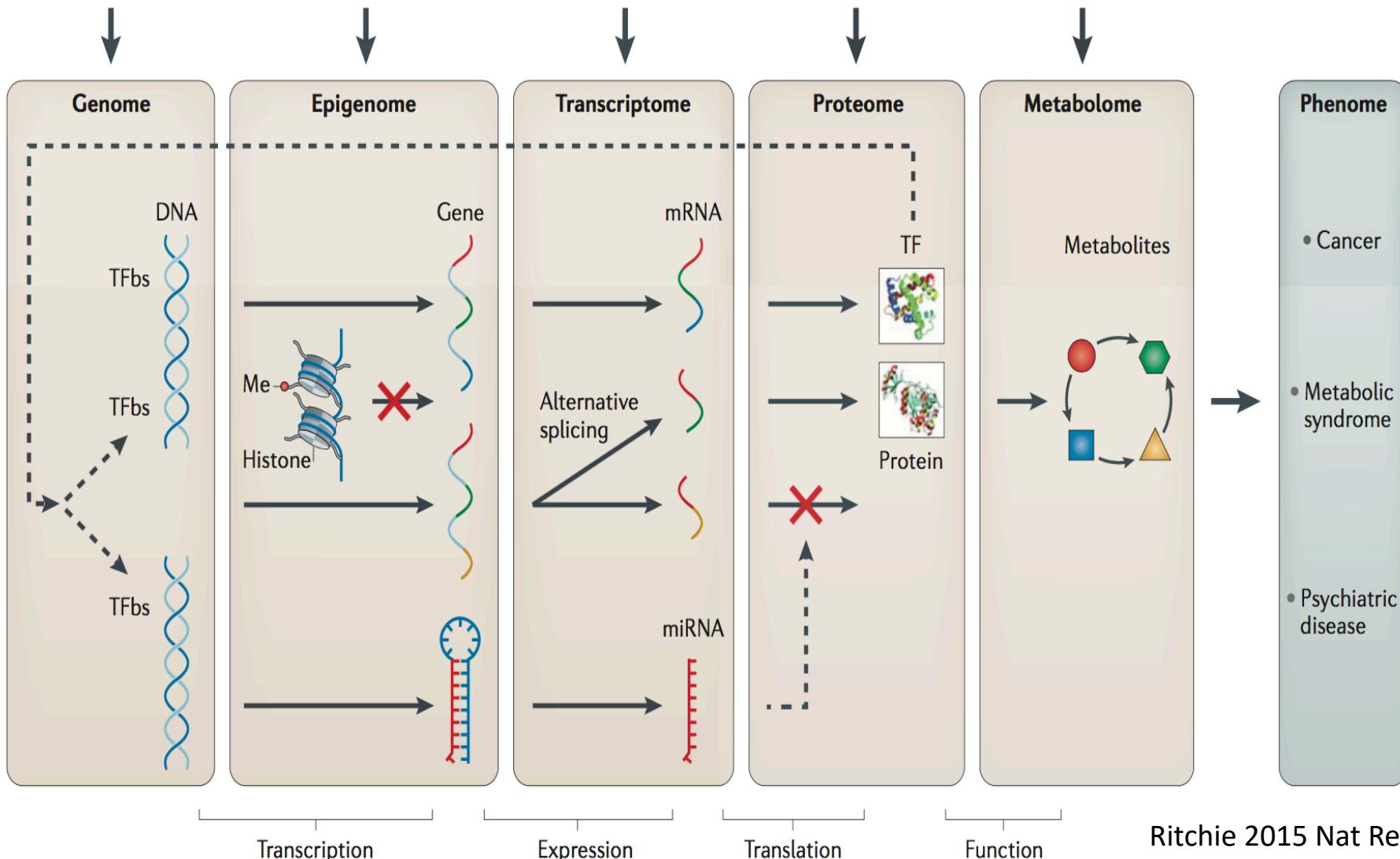
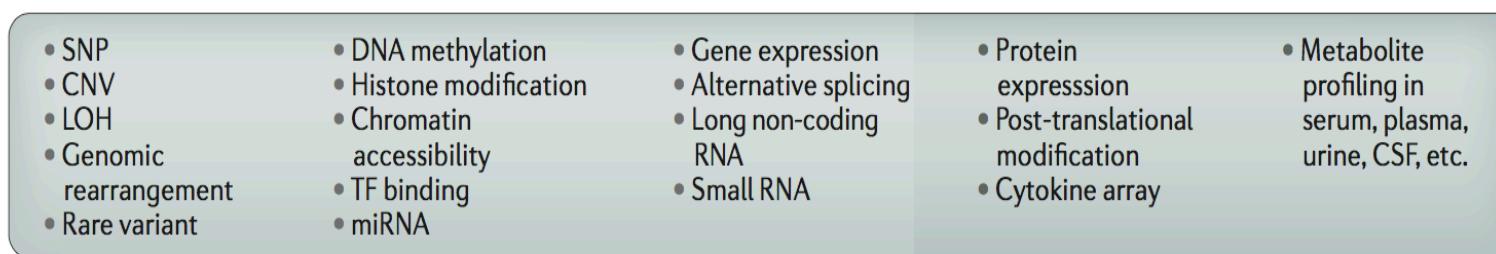
- A protein is a linear polymer of amino acids linked together by peptide bonds.
- Proteins are the main functional chemicals in the cell, carrying out many functions, for example catalysis of the reactions involved in metabolism.
- Proteins have a complex spatial structure



The Genetic Code

Transcriptional Regulation

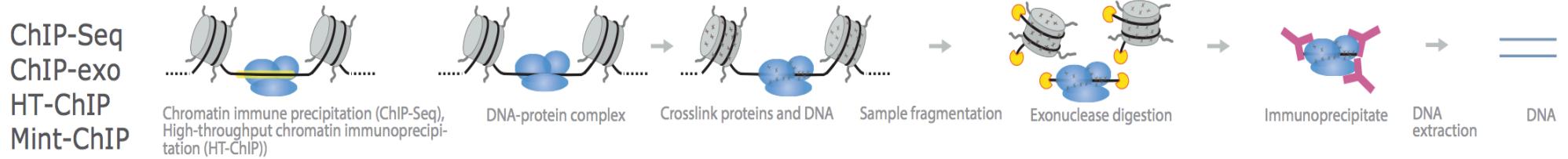




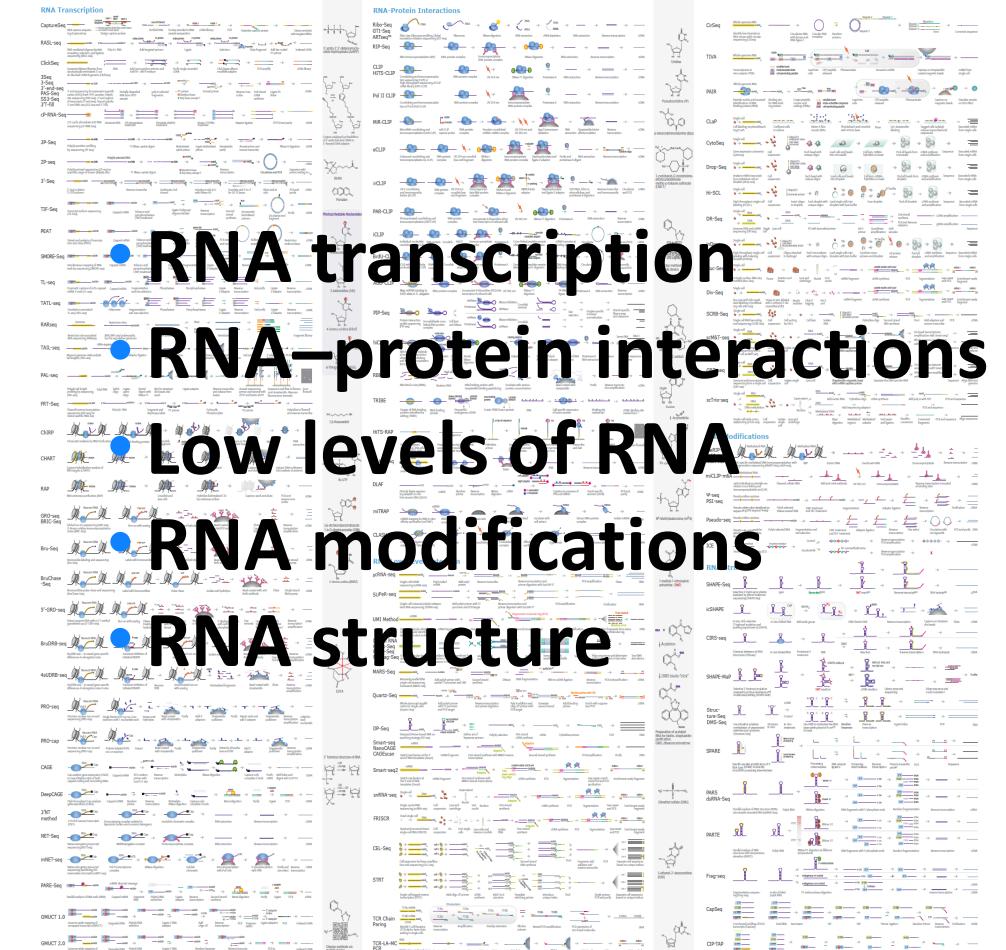
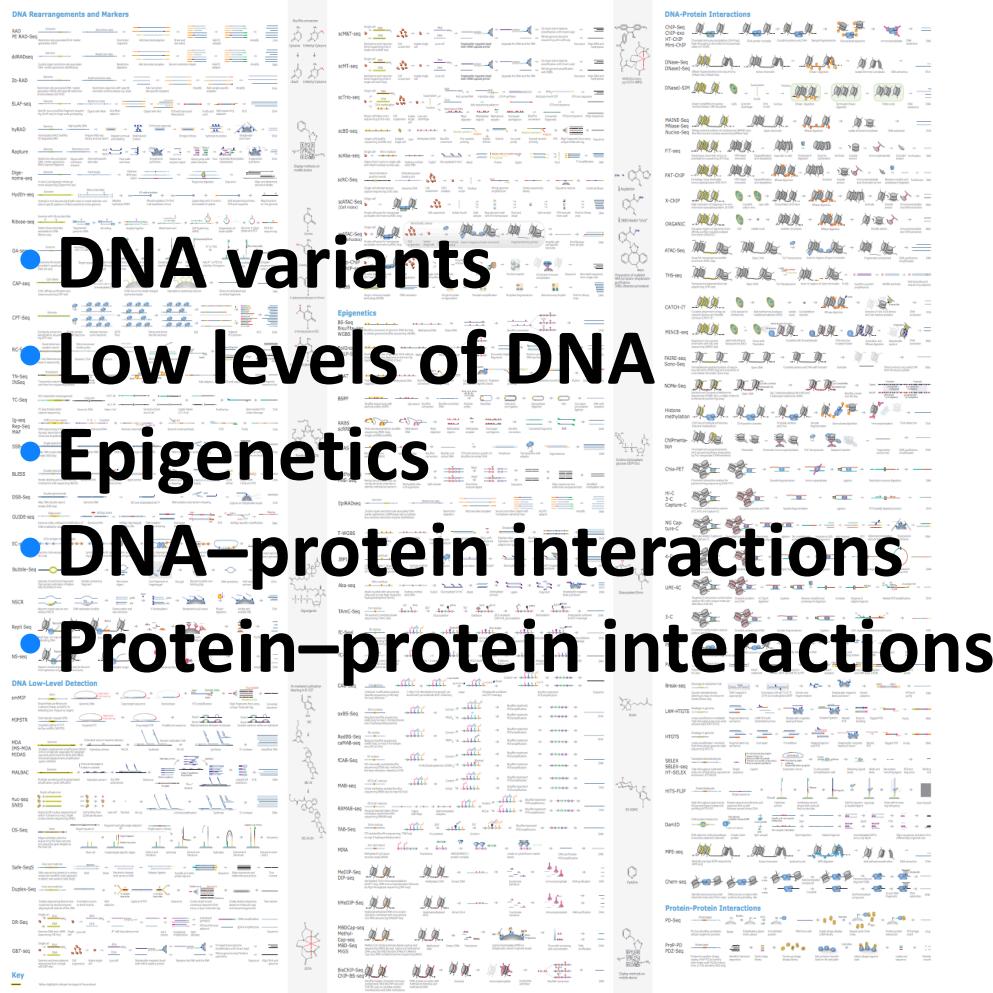
NGS Protocols

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

DNA-Protein Interactions



- The preparation determines how sequenced reads have to be interpreted



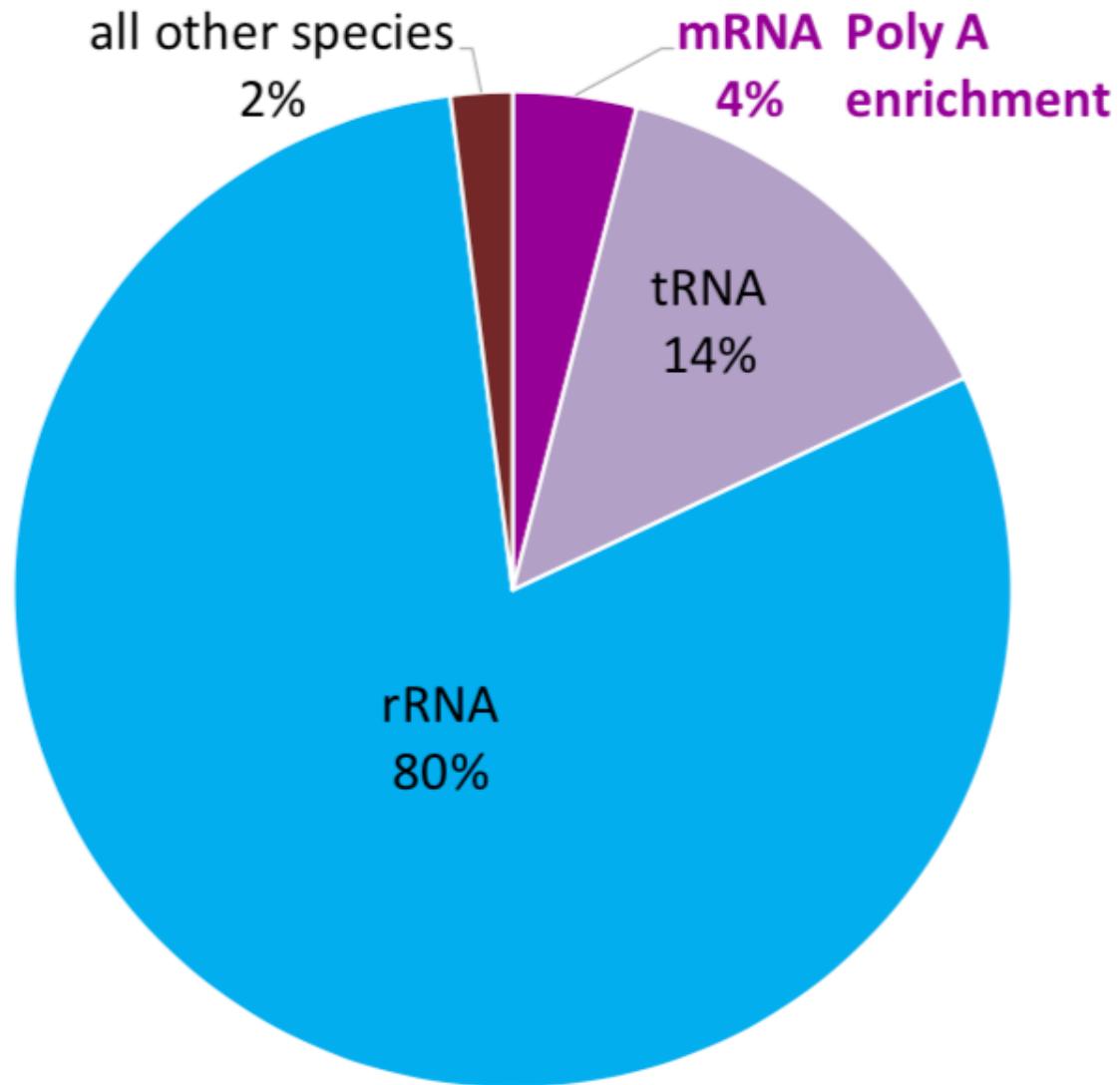
RNA-seq in numbers

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)