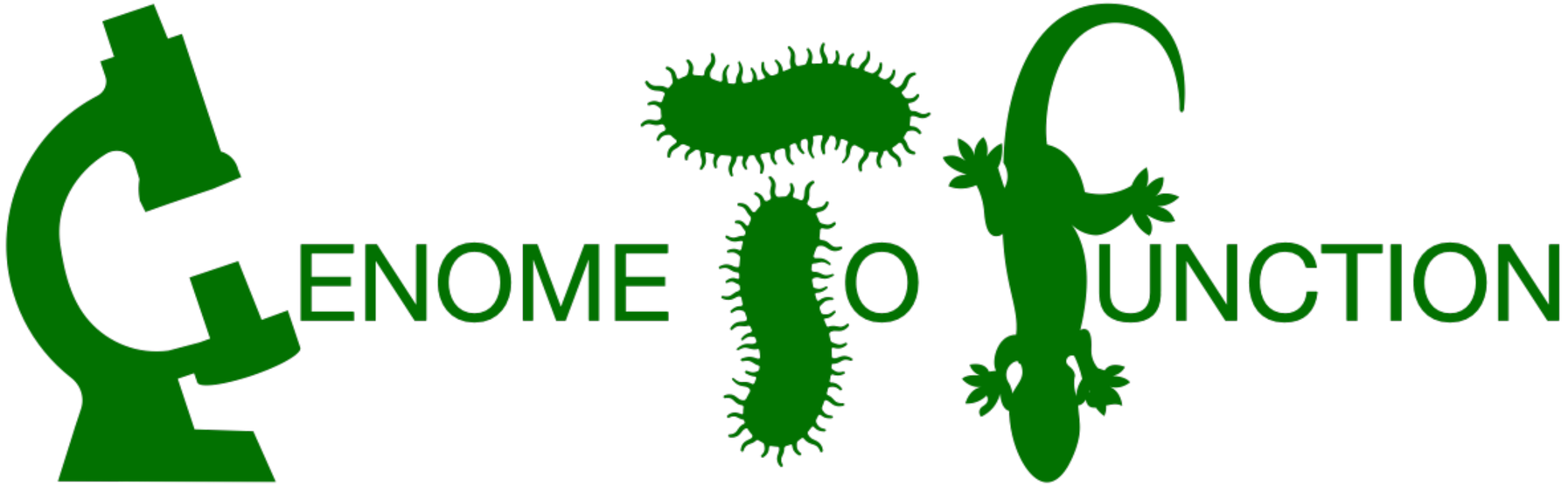
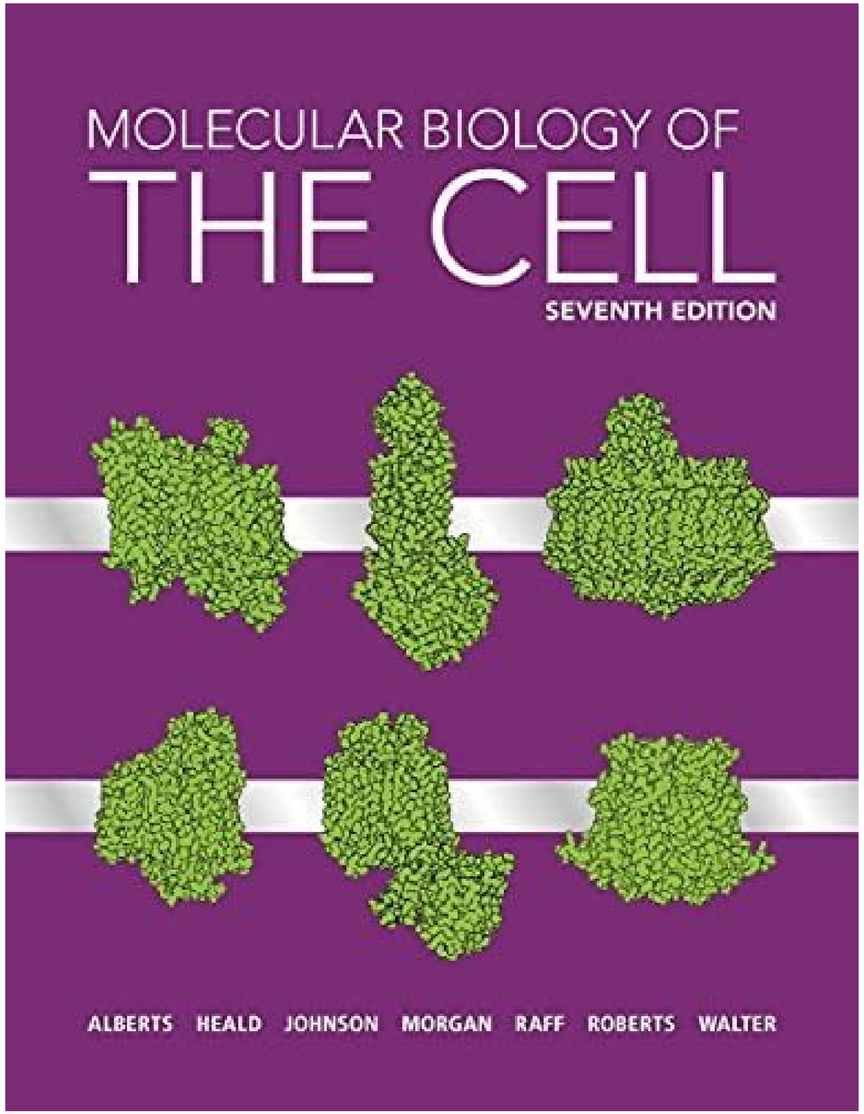


Decoding & Understanding Genomes

Supplement

Recommended Reading



Molecular Biology of the Cell

Seventh Edition

by Bruce Alberts, Rebecca Heald, Alexander Johnson, David Morgan, Martin Raff, Keith Roberts & Peter Walter

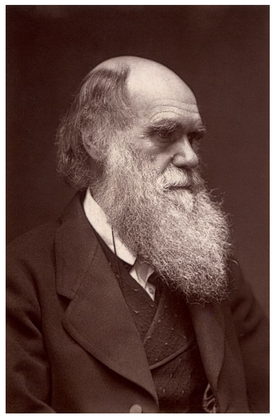
Available in the EPFL library

Link

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Whistle-stop tour of genetics

1859 - Charles Darwin publishes The Origin of Species

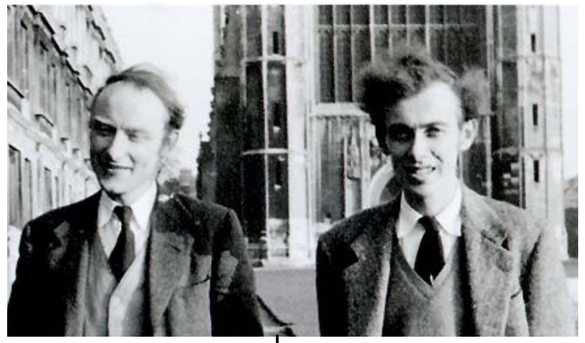


1866 - Gregor Mendel discovers the basic principles of genetics

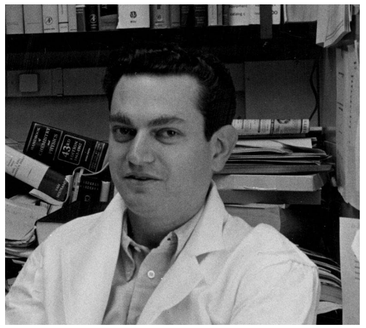
1869 - Friedrich Miescher identifies DNA

1944 - Oswald Avery identifies DNA as the 'transforming principle'

1952 - Rosalind Franklin images crystallised DNA



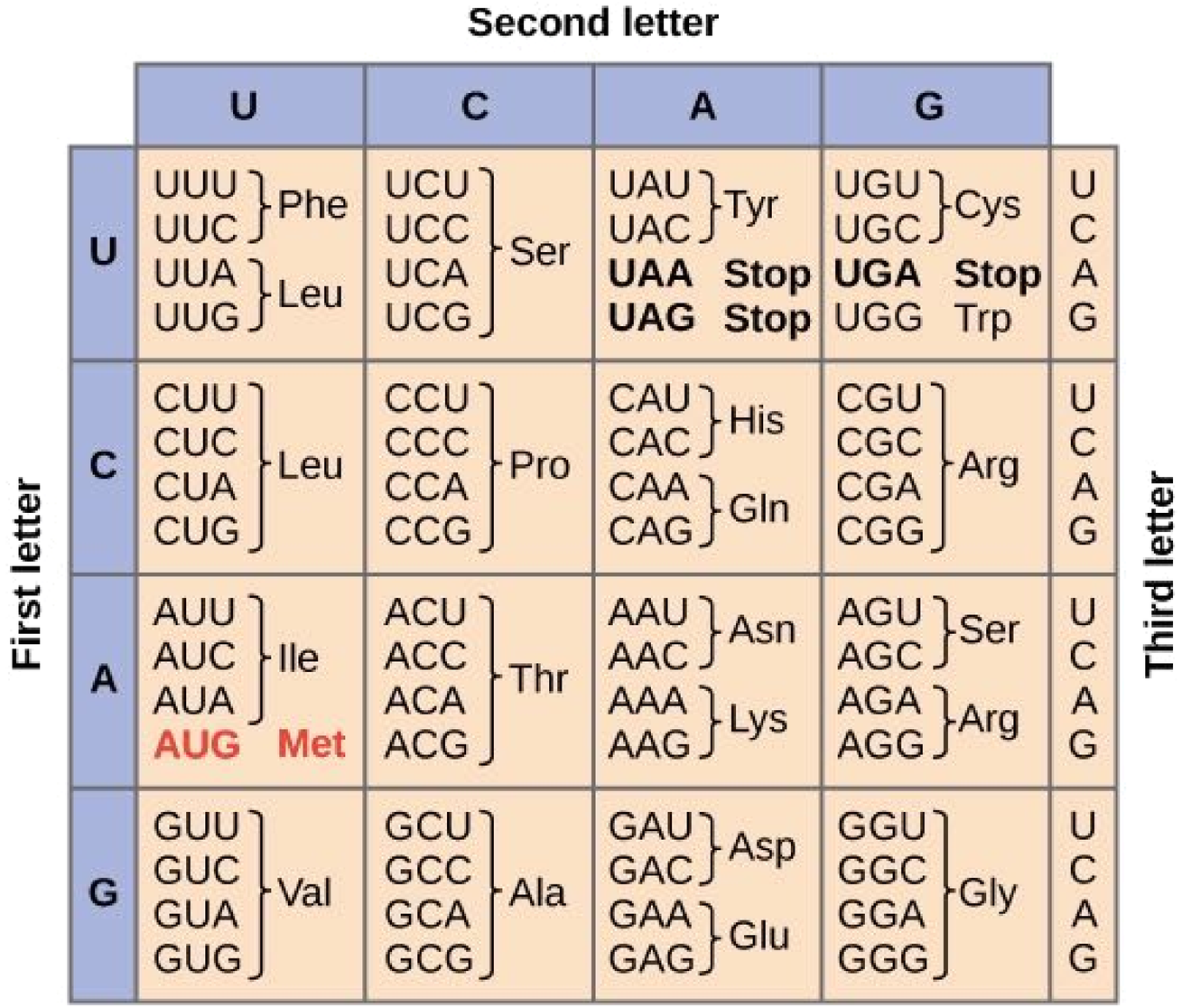
1953 - James Watson and Francis Crick determine DNA is a double helix



1965 - Marshall Nirenberg breaks the genetic code

1977 - Frederick Sanger develops rapid DNA sequencing tecchnique

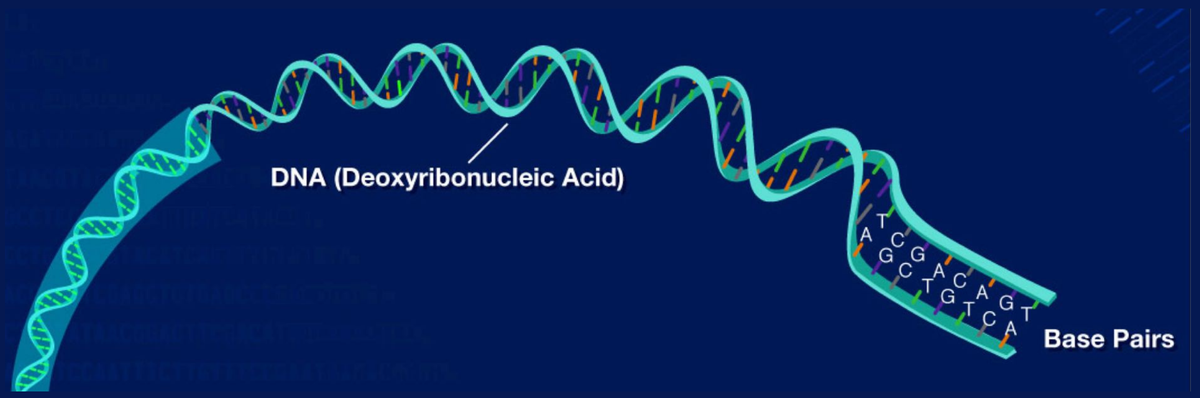
The Genetic code



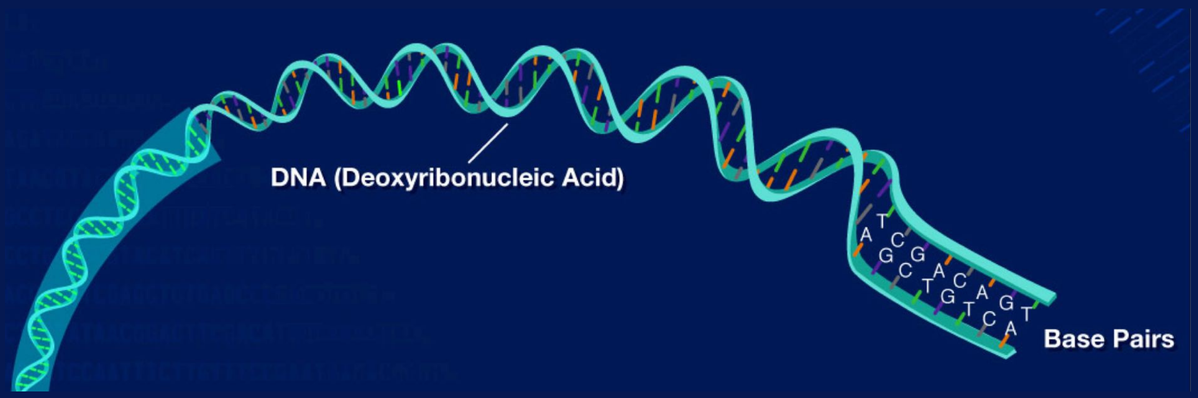
What is a Genome?

- A **complete set** of an organisms DNA is called its genome
- Human **diploid** genome is 6.37/6.27 (female/male) Gigabase pairs (Gbp)
- Weighs **6.41 picograms** and end-to-end would be **205 cm** long
- Full copy in the majority of cells in the body

How is DNA packed into the cell nucleus?

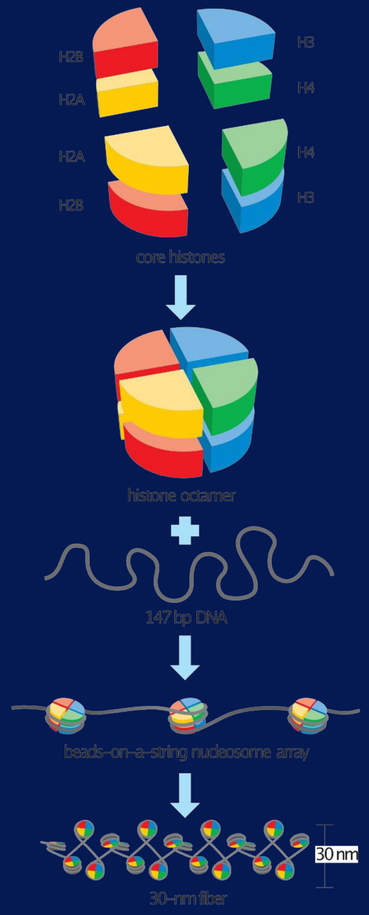


How is DNA packed into the cell nucleus?

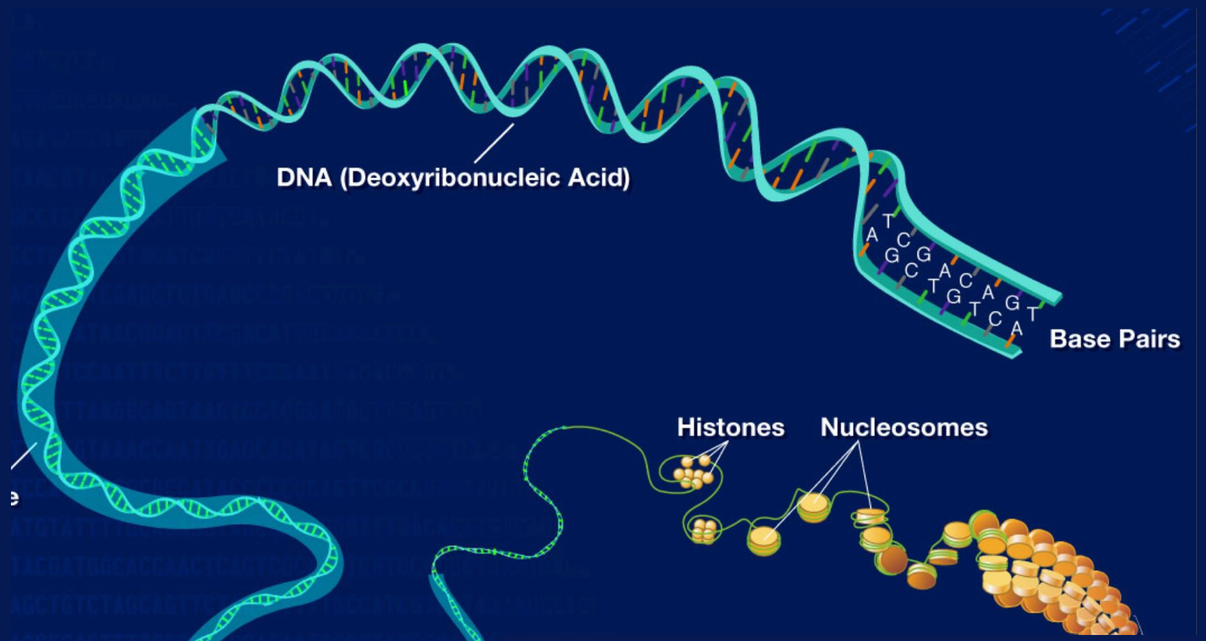


Histone proteins form octomeric complexes which act as spools around which DNA can wind to create structural units called **nucleosomes**

Histones can also modify DNA molecules e.g. through methylation (epigenetic modifications)

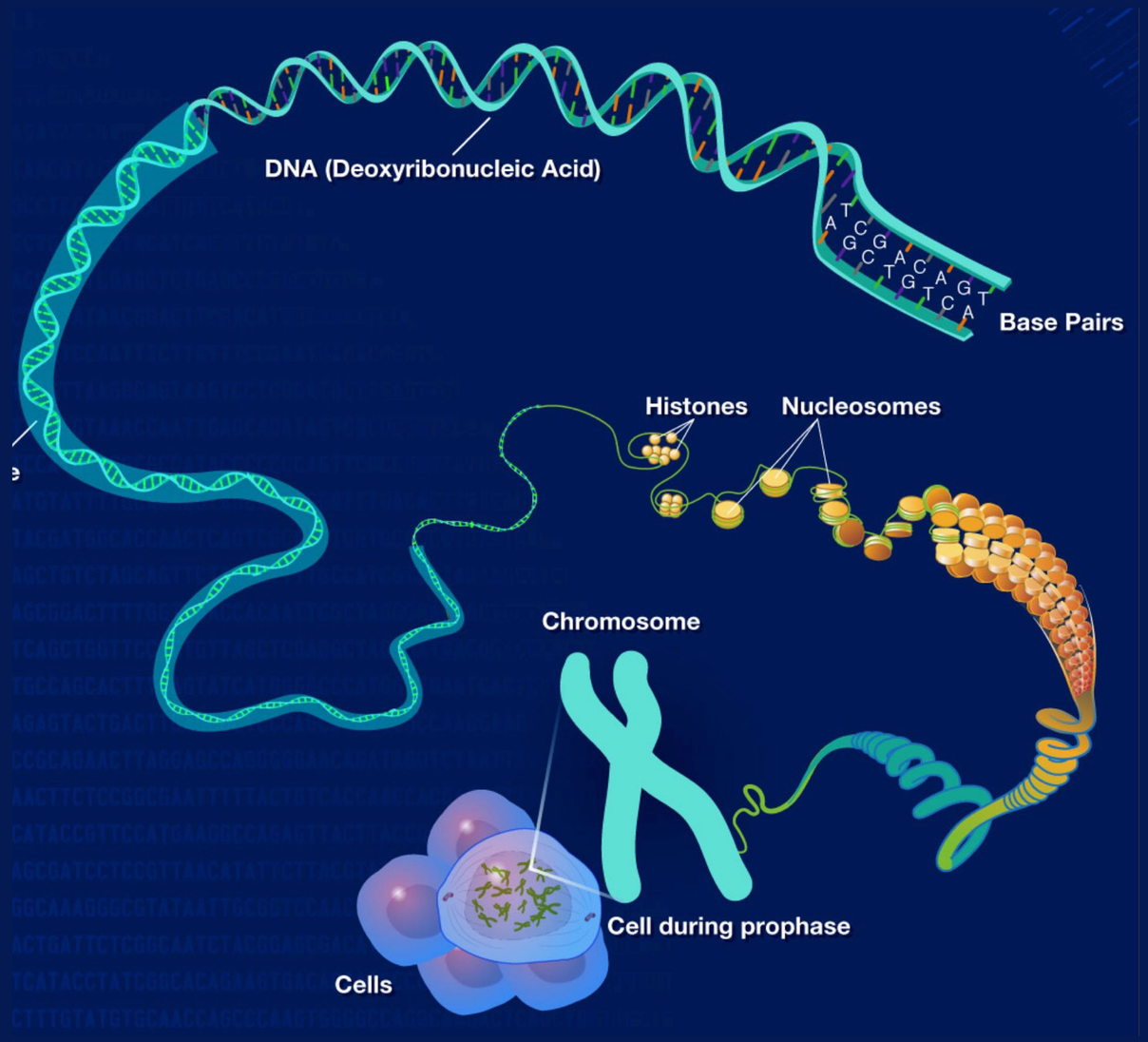


How is DNA packed into the cell nucleus?



This complex of DNA and histone proteins is called **Chromatin**

How is DNA packed into the cell nucleus?

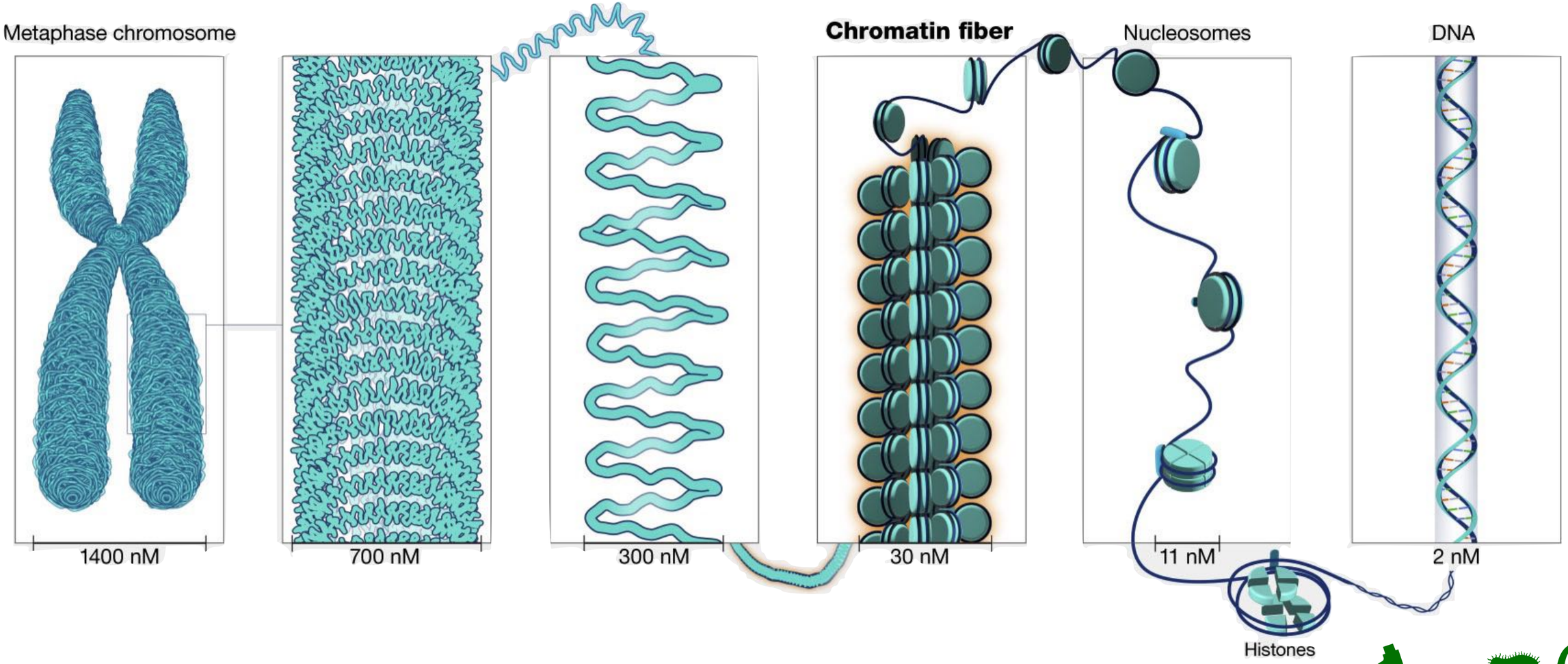


Chromosomes are formed from single DNA molecule in a chromatin complex

Chromatin becomes densely packed during **prophase** of the cell cycle which makes chromosomes visible under the microscope

During other phases of the cell cycle such as **interphase**, chromatin is much less densely packed and chromosomes are much harder to visualise.

How is DNA packed into the cell nucleus?



What is a gene and how many are there?

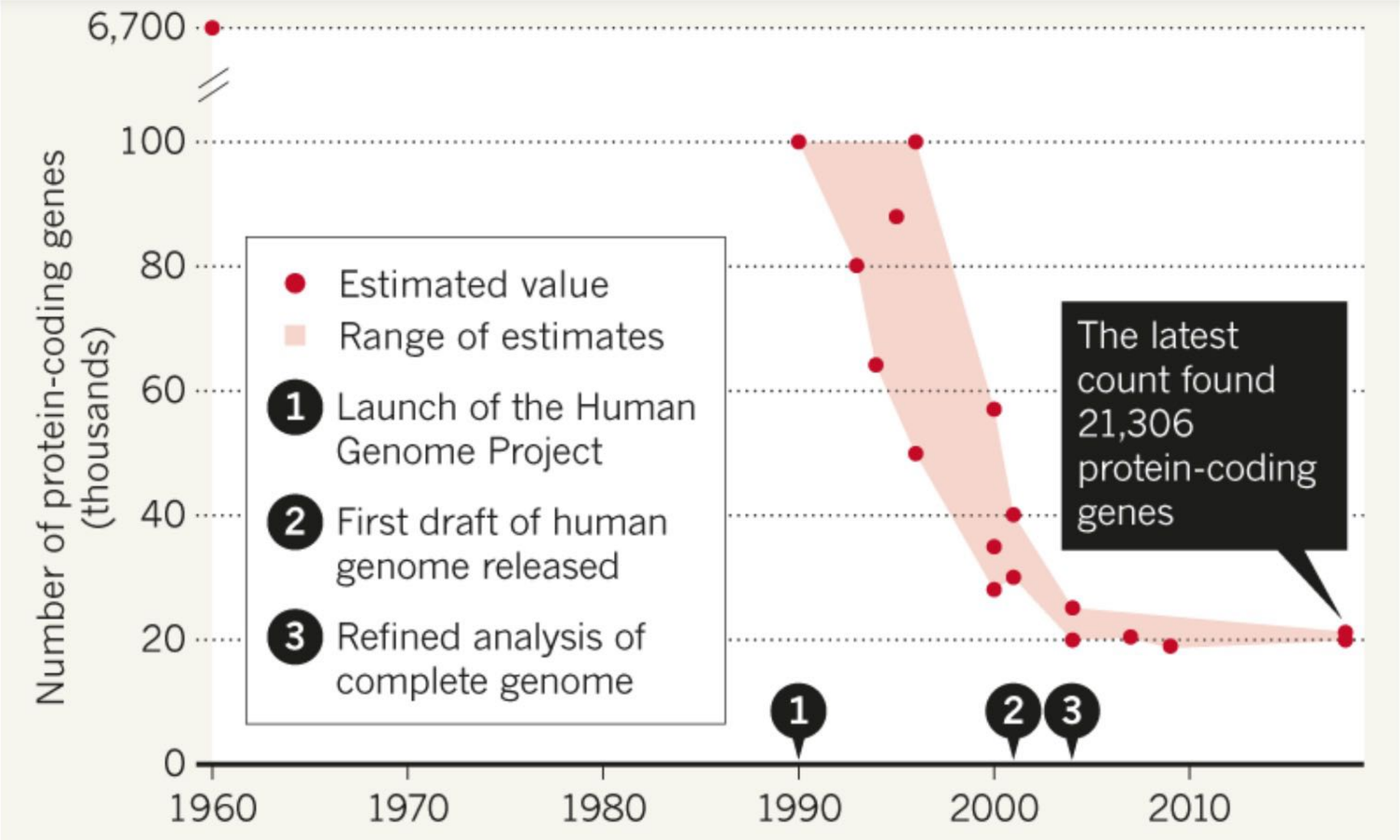
A gene is a section of DNA that encodes a function

Only a small percentage of genomes are translated into functional protein products **(<2%** in case of mouse genome, *FANTOM Consortium,* 2005)

In the human genome it has been estimated that **86,245 different protein products** could be produced from **19,969** protein coding genes by alternative splicing (Nurk et al. 2022).

Due to introns and other regulatory non-coding sequences, ~40% of human genome may be required to produce proteins.

What is a gene and how many are there?



https://doi.org/10.1038/d41586-018-05462-w

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

What is a gene and how many are there?

Table 1: Genome Size and Number of Protein-Coding Genes for a Select Handful of Species

Species and Common Name	Estimated Total Size of Genome (bp)*	Estimated Number of Protein-Encoding Genes*
Saccharomyces cerevisiae (unicellular budding yeast)	12 million	6,000
Trichomonas vaginalis	160 million	60,000
Plasmodium falciparum (unicellular malaria parasite)	23 million	5,000
Caenorhabditis elegans (nematode)	95.5 million	18,000
Drosophila melanogaster (fruit fly)	170 million	14,000
Arabidopsis thaliana (mustard; thale cress)	125 million	25,000
Oryza sativa (rice)	470 million	51,000
Gallus gallus (chicken)	1 billion	20,000–23,000
Canis familiaris (domestic dog)	2.4 billion	19,000
Mus musculus (laboratory mouse)	2.5 billion	30,000
Homo sapiens (human)	2.9 billion	20,000–25,000

Pray, L. (2008) Eukaryotic genome complexity. *Nature Education* 1(1):96

What is a gene and how many are there?

Non-coding genes: Produce transfer RNAs (**tRNAs**) and ribosomal RNAs (**rRNAs**) [essential for protein production], small nuclear RNAs (**snRNAs**) [essential for RNA splicing], small nucleolar RNAs (**sno RNAs**), microRNAs (**miRNAs**), short interfering RNAs (**siRNAs**), PIWI-interacting RNAs (**piRNAs**) and long noncoding RNAs (**lncRNAs**)

Total number is of non-coding genes is estimated at **43’525** (Nurk et al. 2022).

~ 6% of genome (mostly for rRNA genes)

What about the rest?

Centromeres: [Required for chromosome segregation during cell division] repetitive DNA sequences ~ 6.2 % of the genome

Telomeres: [Required for chromosome stability] repetitive DNA sequences, short in humans 5-15Kb

Scaffold attachment regions [required for chromatin structure] 0.3% of genome

‘nonfunctional’ DNA: Pseudogenes, Repeat sequences, transposons and viral elements

Are Human protein coding genes different?

Are Human protein coding genes different?

Only **18 protein coding genes** seem to be present in humans but not other primates, though humans have specific variants of many genes

Most genes in humans are found in other species e.g. ~90% of human genes have a homolog in mouse and 65% a homolog in Drosophila

It is not the protein products of a genome (mostly) but the regulation gene expression pattern and timing (e.g. during development) that provides species uniqueness.

What about differences between the sexes?

Biological females usually have 2 X chromosomes, while biological males have 1 X and 1Y chromosome.

The Y chromosome has ~55 genes (including SRY, a key sex determining factor) while X chromosome has ~ 900. Most of the Y is heterochromatin.

Most genes on the X chromosome **are not involved** in sex determination

In human females, one X chromosome is inactivated (put into a heterochromatic state) to avoid **gene dosage effects**

In other animals (e.g. Drosophila), X chromosome genes produce ~2x the amount proteins in males compared to females to avoid **gene dosage effects**

What does my genome tell about me?

Some genetic risk factors with **strong effects** can be identified (e.g. BRCA alleles and breast cancer).

But most ‘gene’ associations are **weak and effects are hard to predict**.

Gene to phenotype associations are **complex and penetrance** can vary.