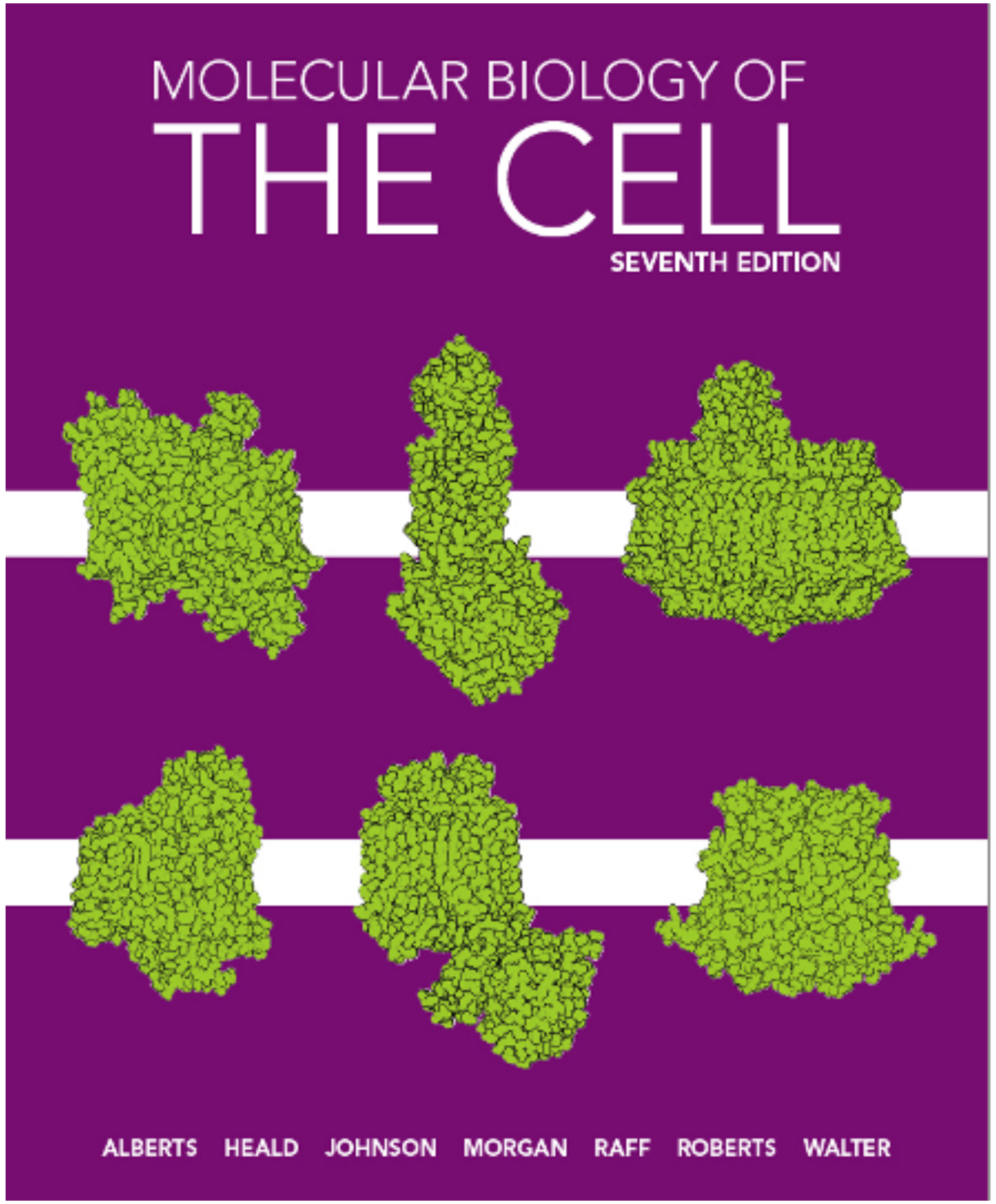
Cellular and Molecular Biology I

BIO-205-2

Camille Goemans

Important information

- How to study for this class?
- ! the content of the videos is extra - no need to study it
- Nice extra videos (on Moodle)
- Question: how is the heterochromatin replicated?

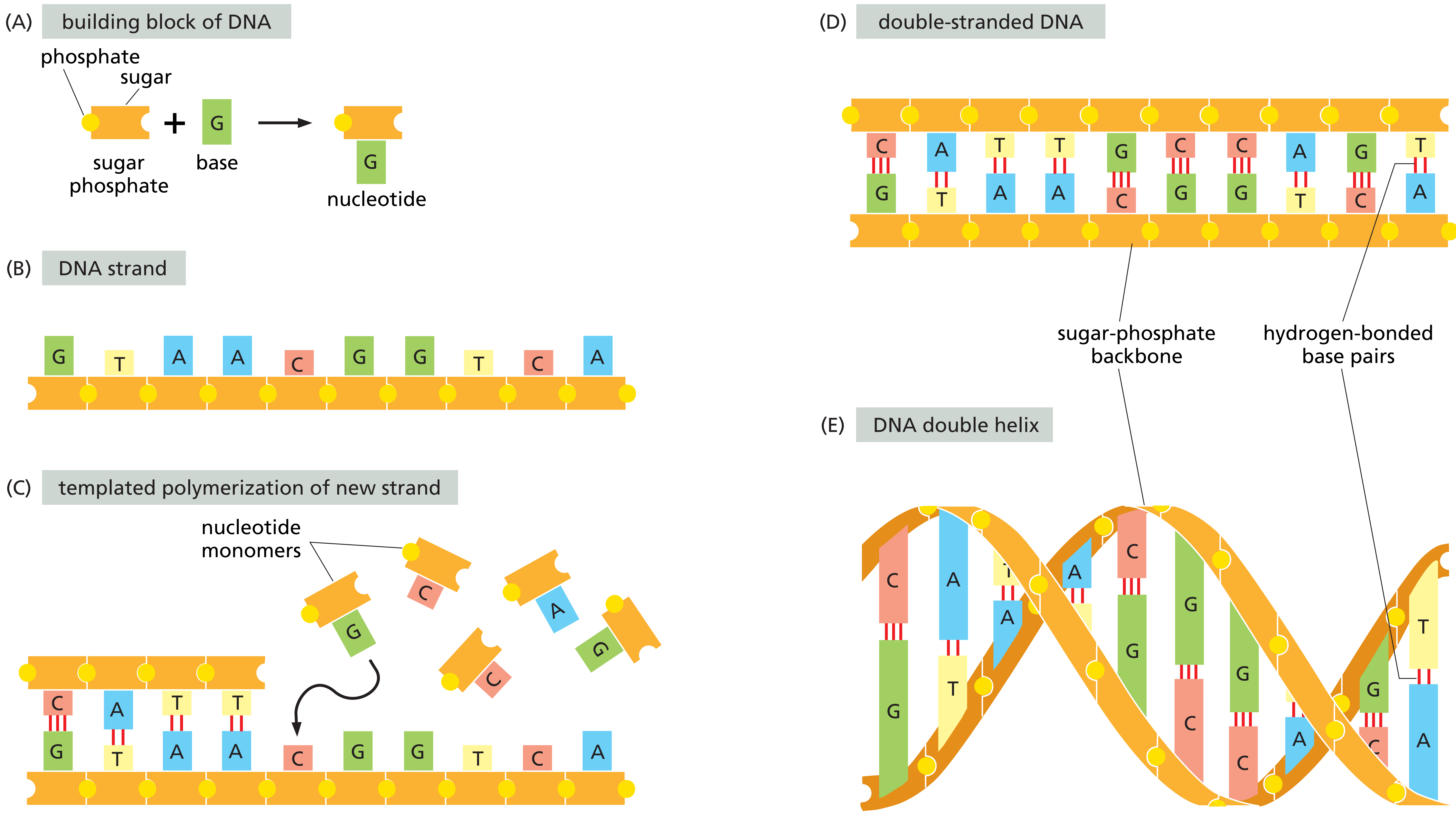

MOLECULAR BIOLOGY OF
THE CELL
SEVENTH EDITION

ALBERTS HEALD JOHNSON MORGAN RAFF ROBERTS WALTER

Chapter 4

DNA, Chromosomes, and Genomes

Quick recap



(A) building block of DNA

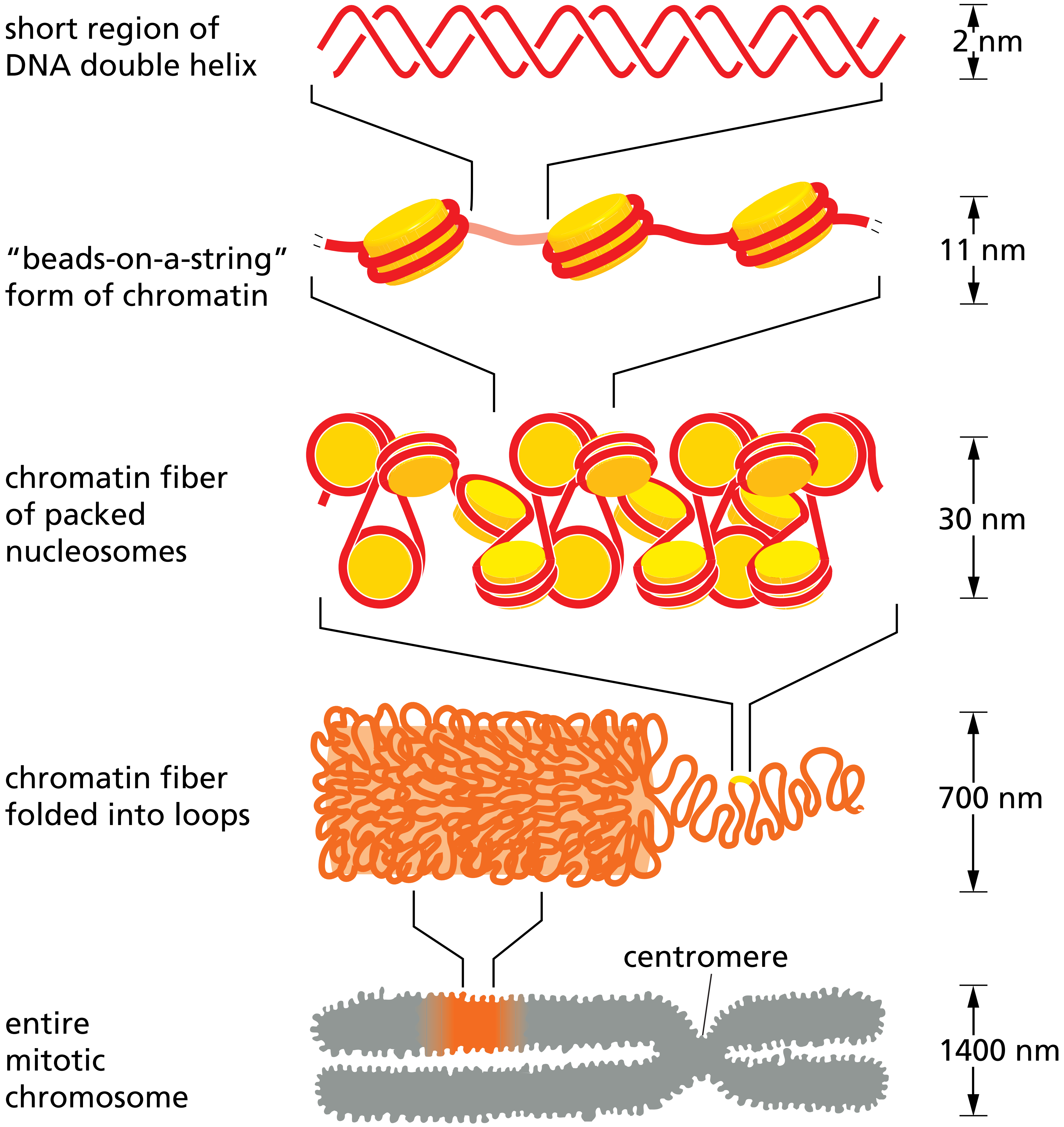
(B) DNA strand

(C) templated polymerization of new strand

(D) double-stranded DNA

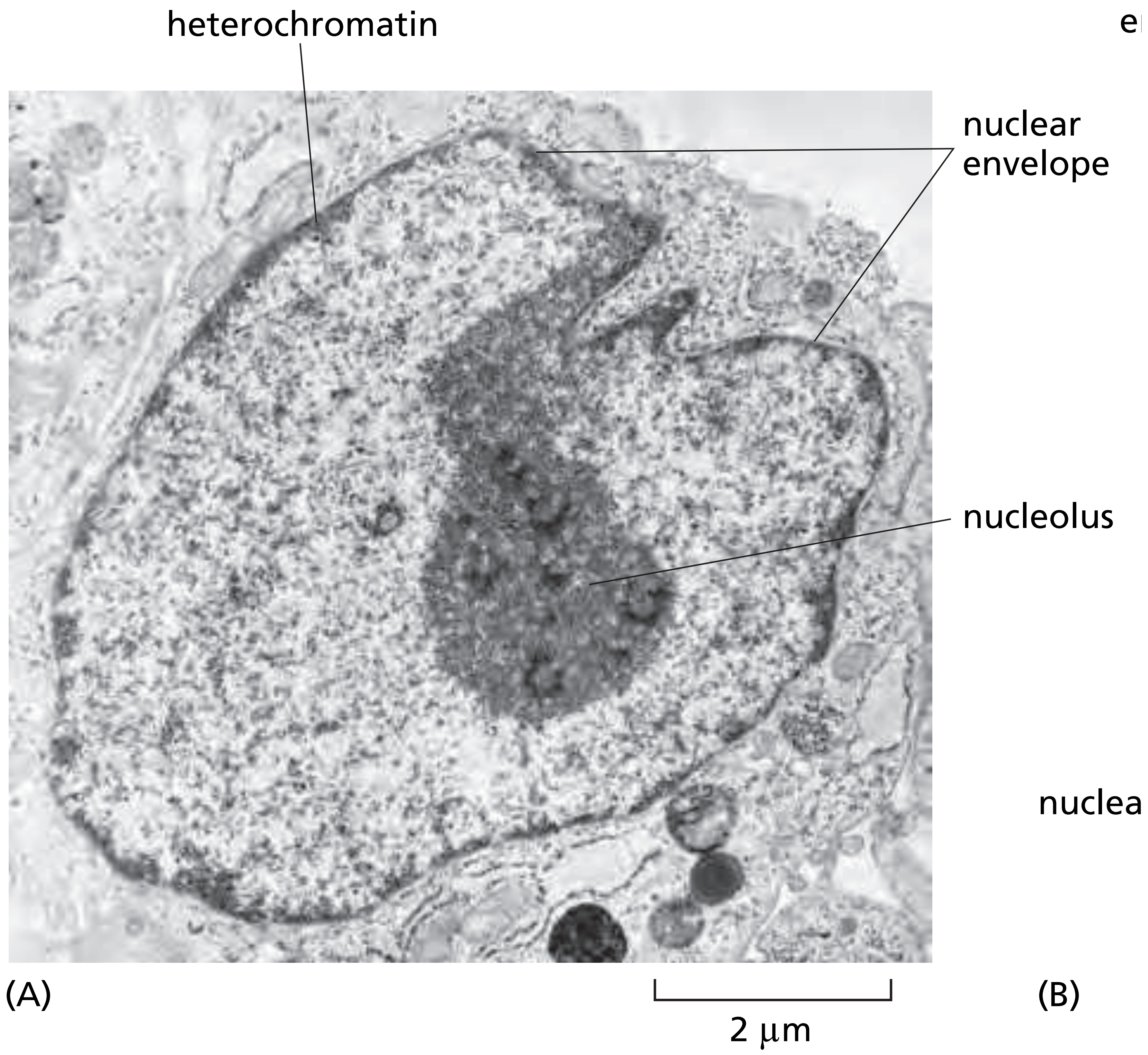
(E) DNA double helix

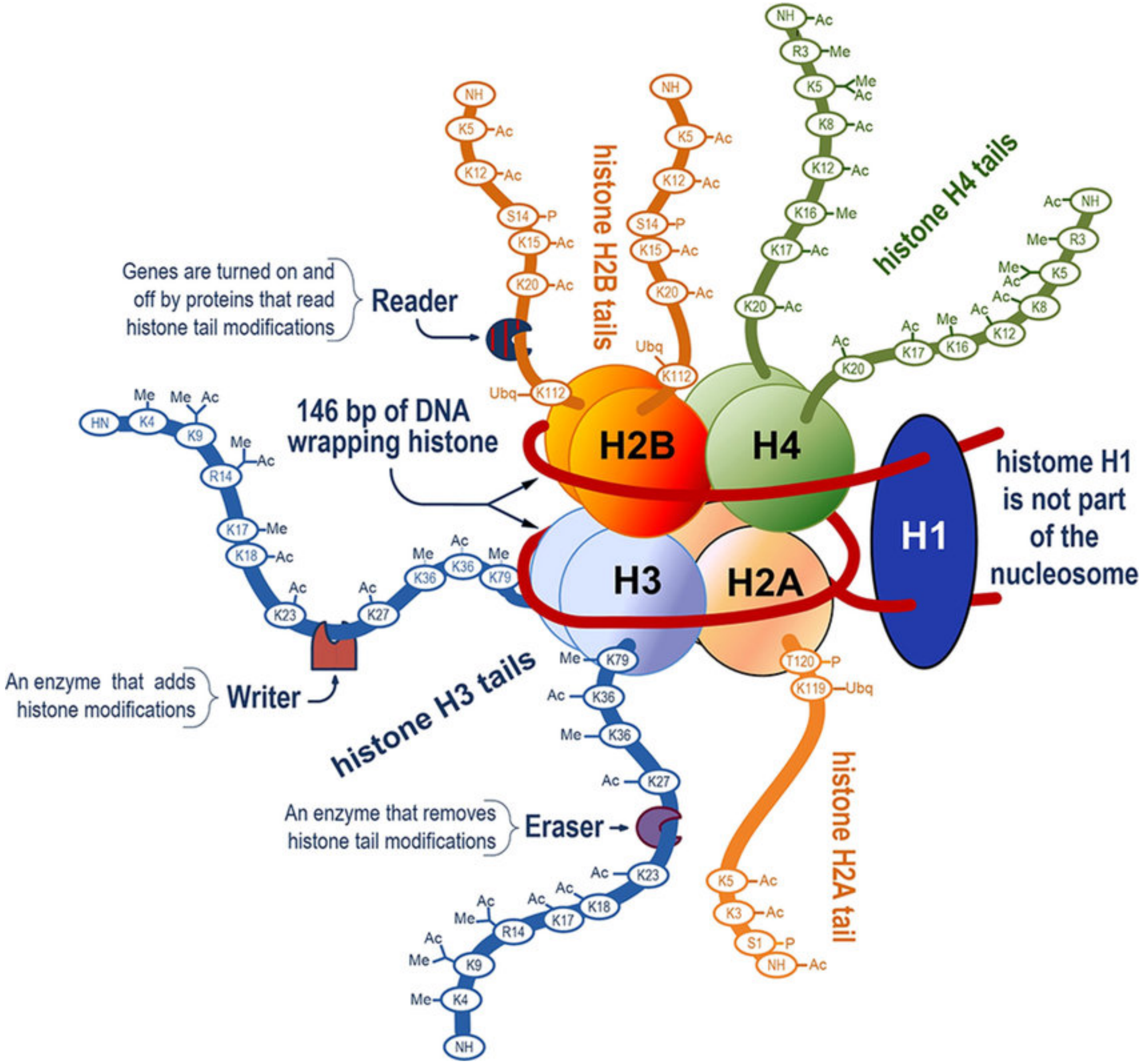
Quick recap



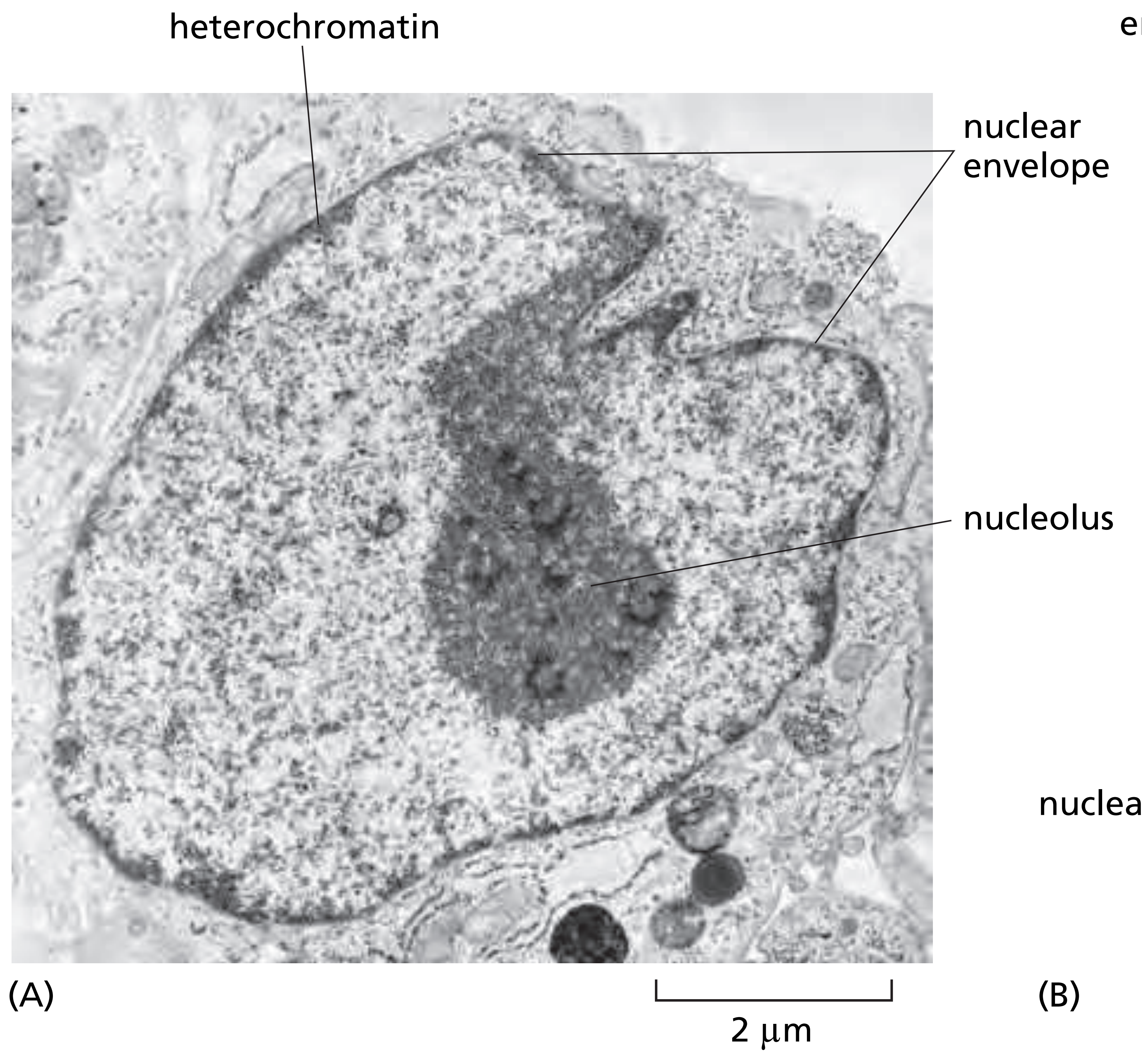
NET RESULT: EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 10,000-FOLD SHORTER THAN ITS FULLY EXTENDED LENGTH

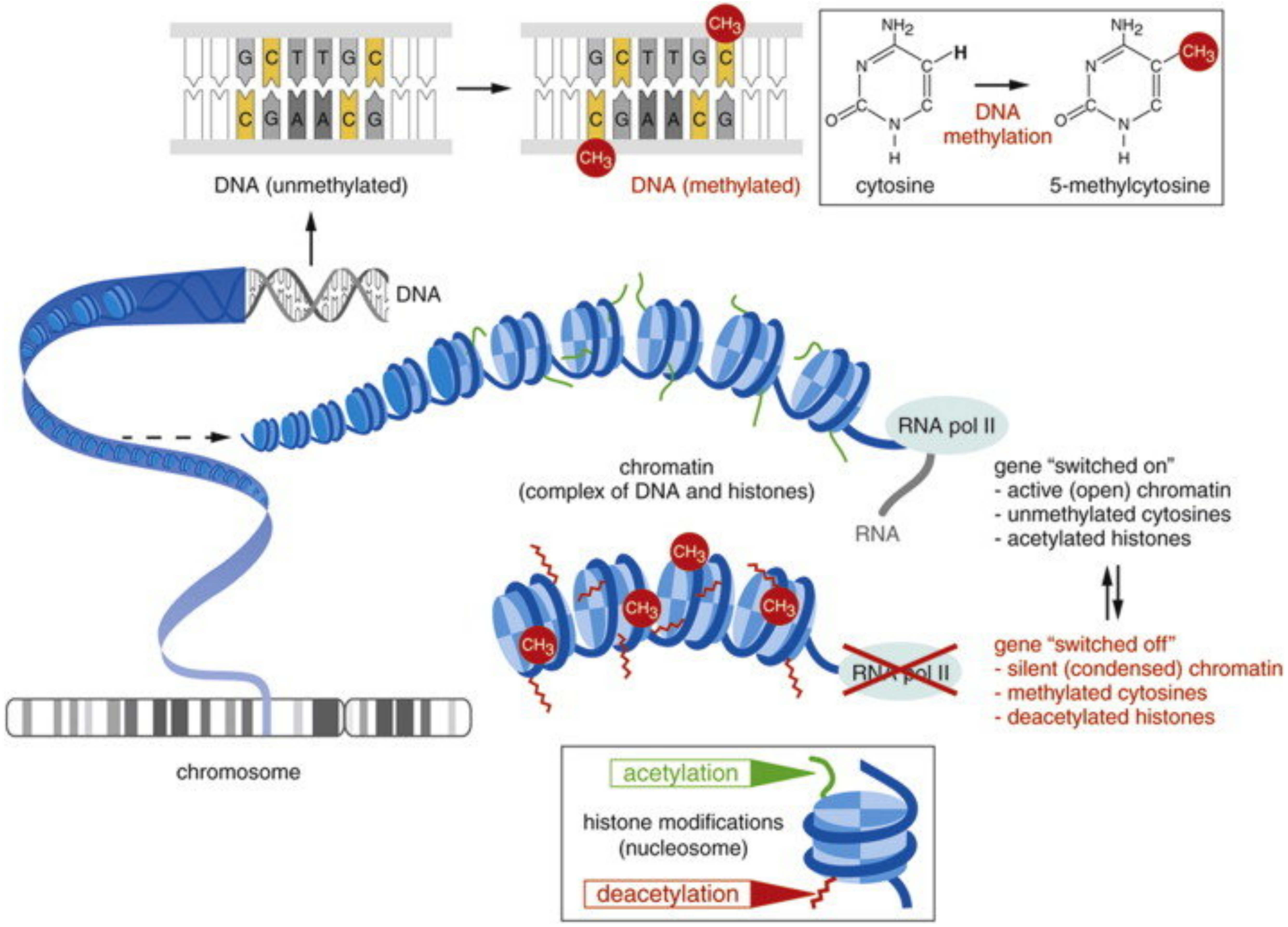
Quick recap


heterochromatin
e
nuclear envelope
nucleolus
nuclea
(A)
2 μm
(B)


Genes are turned on and off by proteins that read histone tail modifications
Reader
146 bp of DNA wrapping histone
An enzyme that adds histone modifications
Writer
An enzyme that removes histone tail modifications
Eraser
histone H2B tails
histone H4 tails
histone H3 tails
histone H2A tail
H2B
H4
H3
H2A
H1
histome H1 is not part of the nucleosome
NH
K5
K12
S14
K15
K20
K112
Ubq
Ac
Me
P
R3
K8
K16
K17
HN
K4
K9
R14
K18
K23
K27
K36
K79
T120
K119
K3
S1

Quick recap


heterochromatin
e
nuclear envelope
nucleolus
nuclea
(A)
2 µm
(B)

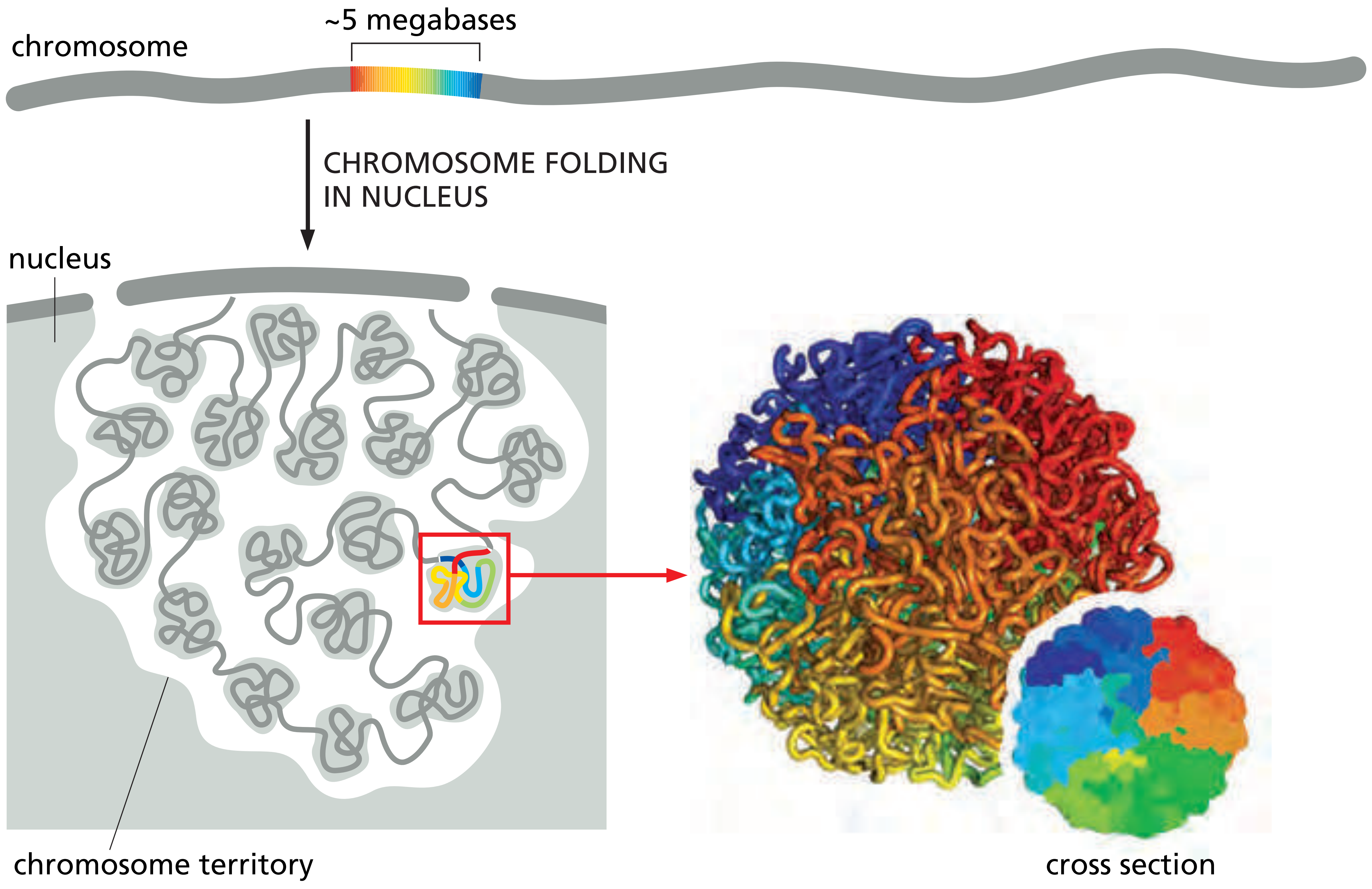

G C T T G C
C G A A C G
DNA (unmethylated)
CH3
CH3
DNA (methylated)
NH2
H
N
O
N
H
cytosine
DNA methylation
CH3
5-methylcytosine
DNA
chromatin
(complex of DNA and histones)
RNA pol II
RNA
gene "switched on"
- active (open) chromatin
- unmethylated cytosines
- acetylated histones
CH3
CH3
CH3
CH3
RNA pol II
gene "switched off"
- silent (condensed) chromatin
- methylated cytosines
- deacetylated histones
chromosome
acetylation
histone modifications
(nucleosome)
deacetylation

Plan

- Quick recap
- <u>The global structure of chromosomes</u>
- Analysis of chromatin organization
- How do genomes evolve

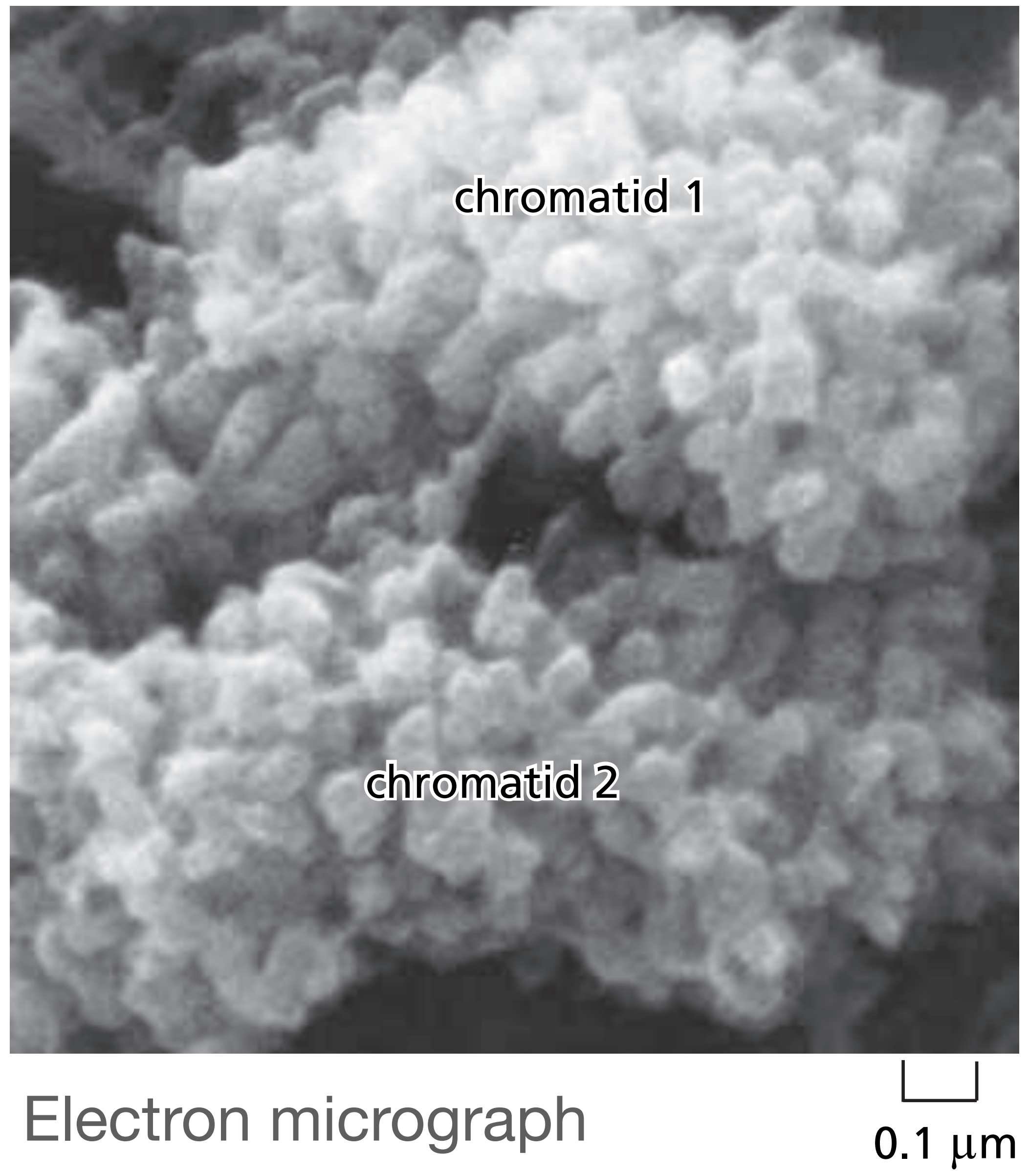
The global structure of chromosomes

- Human DNA is about 2 meters long, yet it fits inside a nucleus just a few micrometers wide.
- To function properly, DNA must be **compacted** without tangling while still allowing regions to be accessed for gene expression and regulation.
- Chromosomes are not folded **randomly** in the nucleus
- Most regions of our chromosomes are arranged into a conformation = **fractal globule** - dense packing that maintains the ability to fold and unfold



The global structure of chromosomes

- **Mitotic chromosomes** are especially highly condensed (visible by light microscopy) - gene expression is shut down
- Two DNA molecules produced during interphase = 2 sister chromatids held together at the centromere
- Final level of chromosome packaging



Electron micrograph

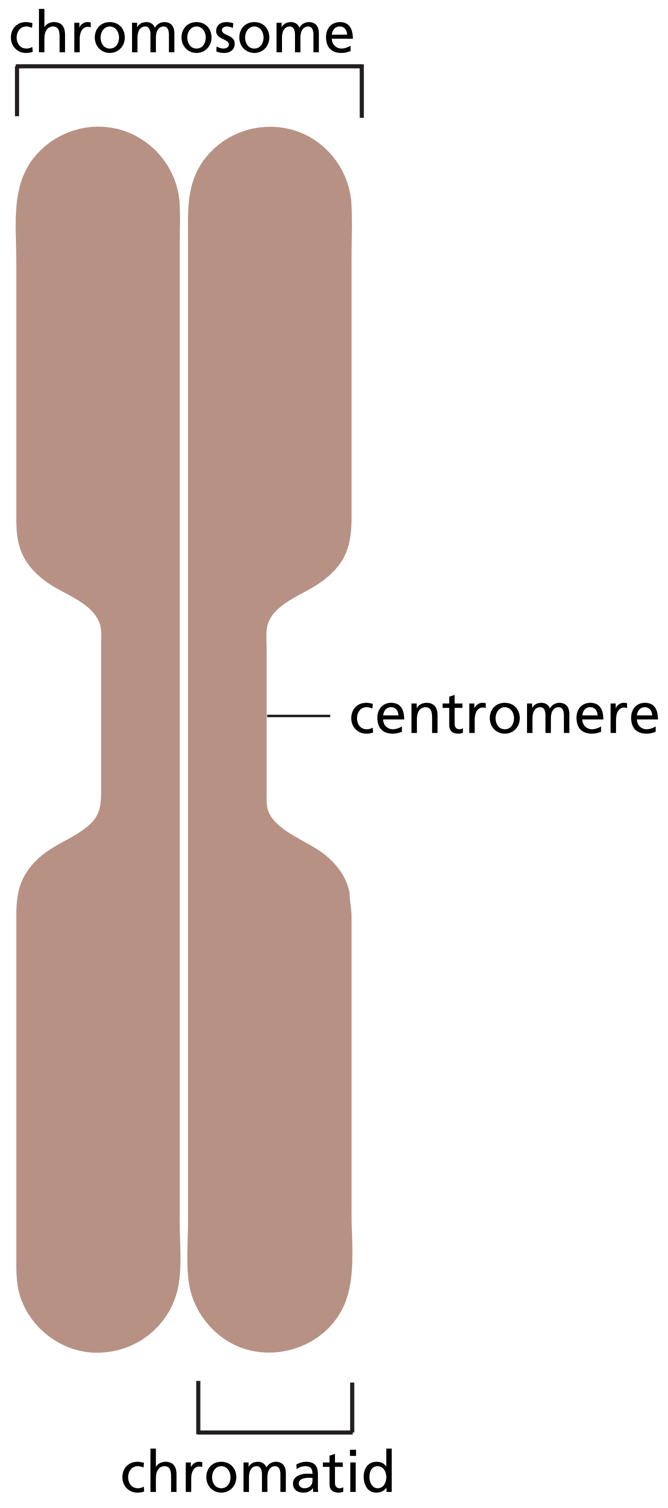
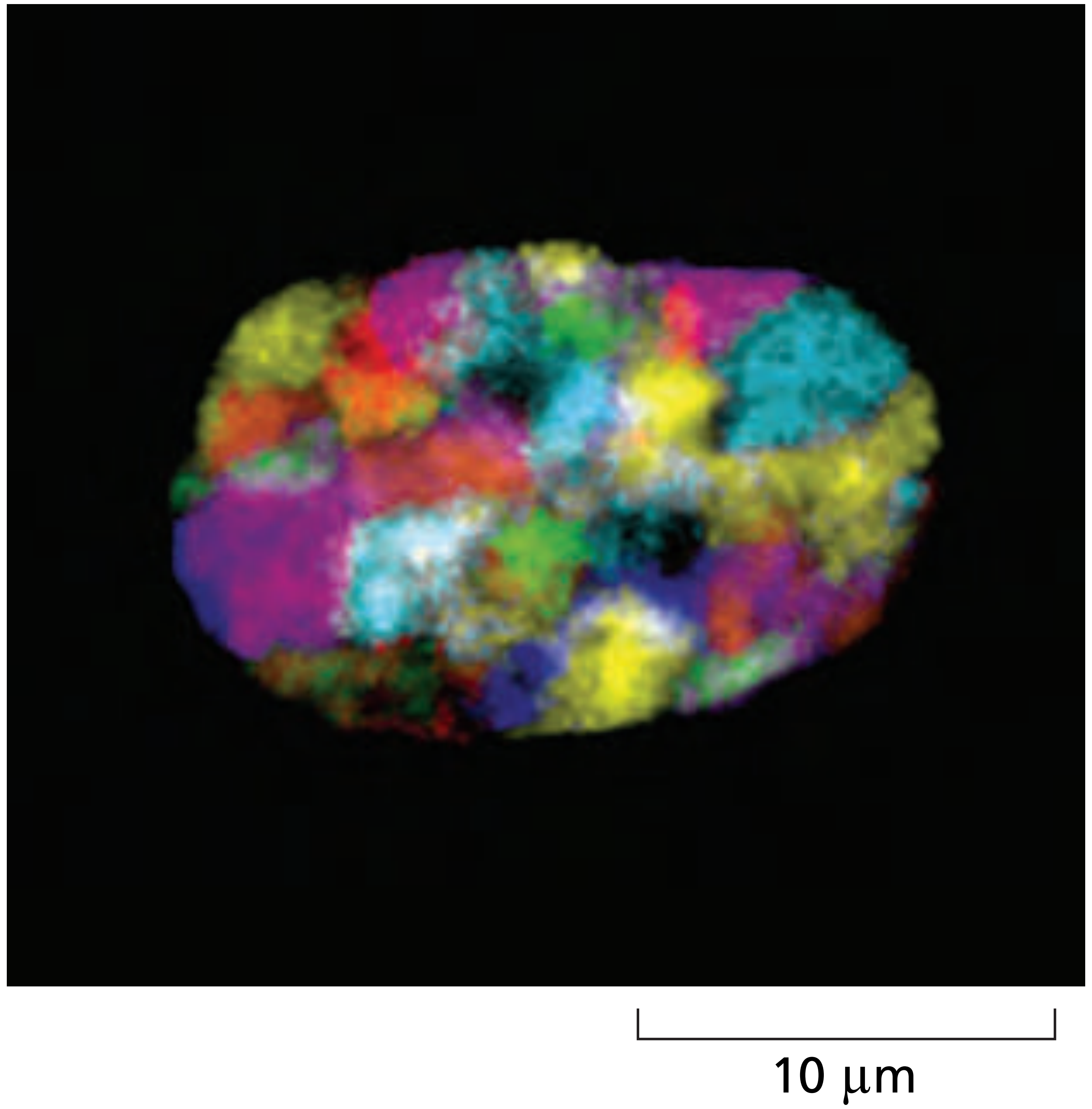
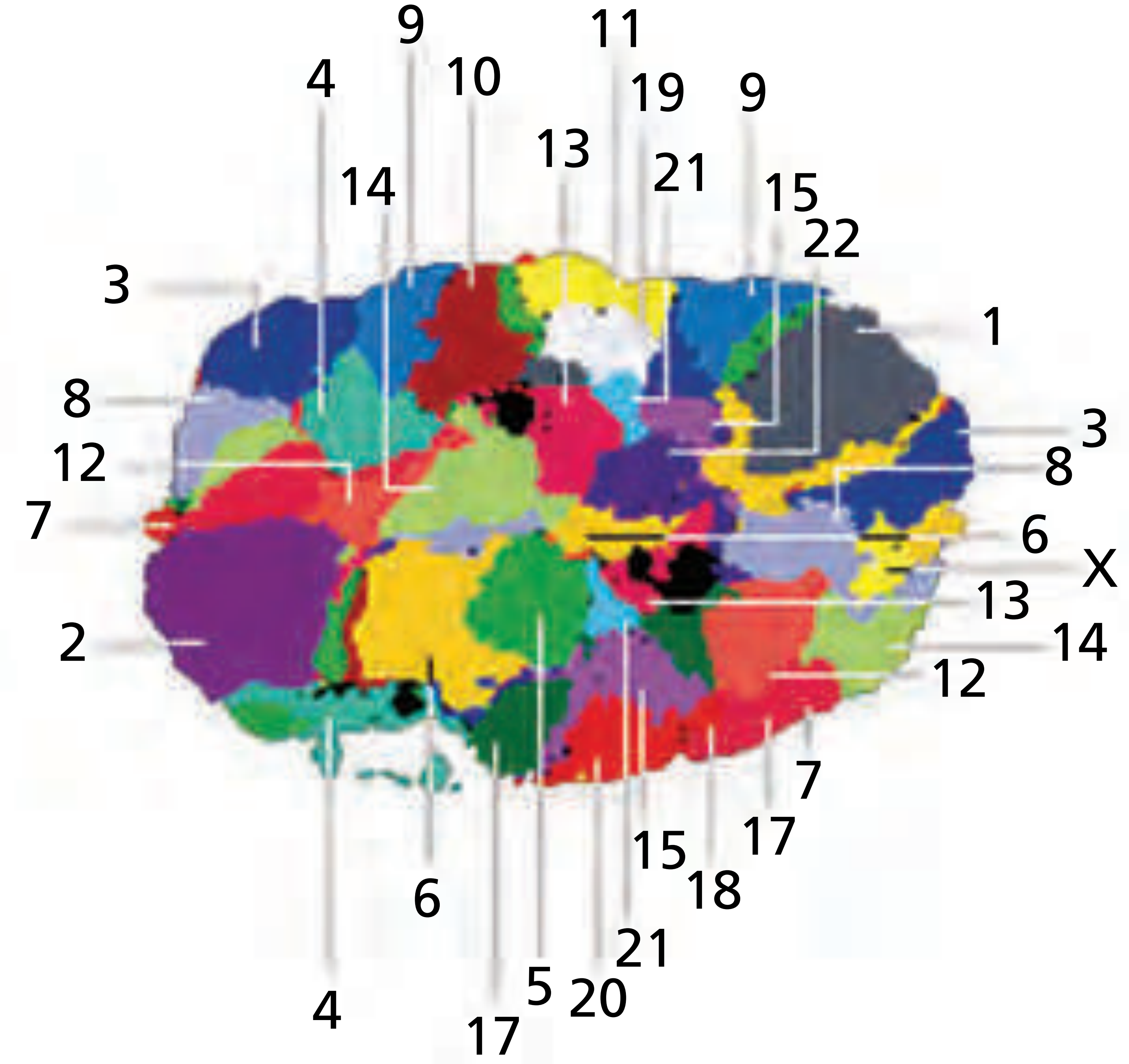


Figure 4–59 A typical mitotic chromosome at metaphase. Each sister chromatid contains one of two identical sister DNA molecules generated earlier in the cell cycle by DNA replication (see also Figure 17–21).

Chromosomes occupy specific territories

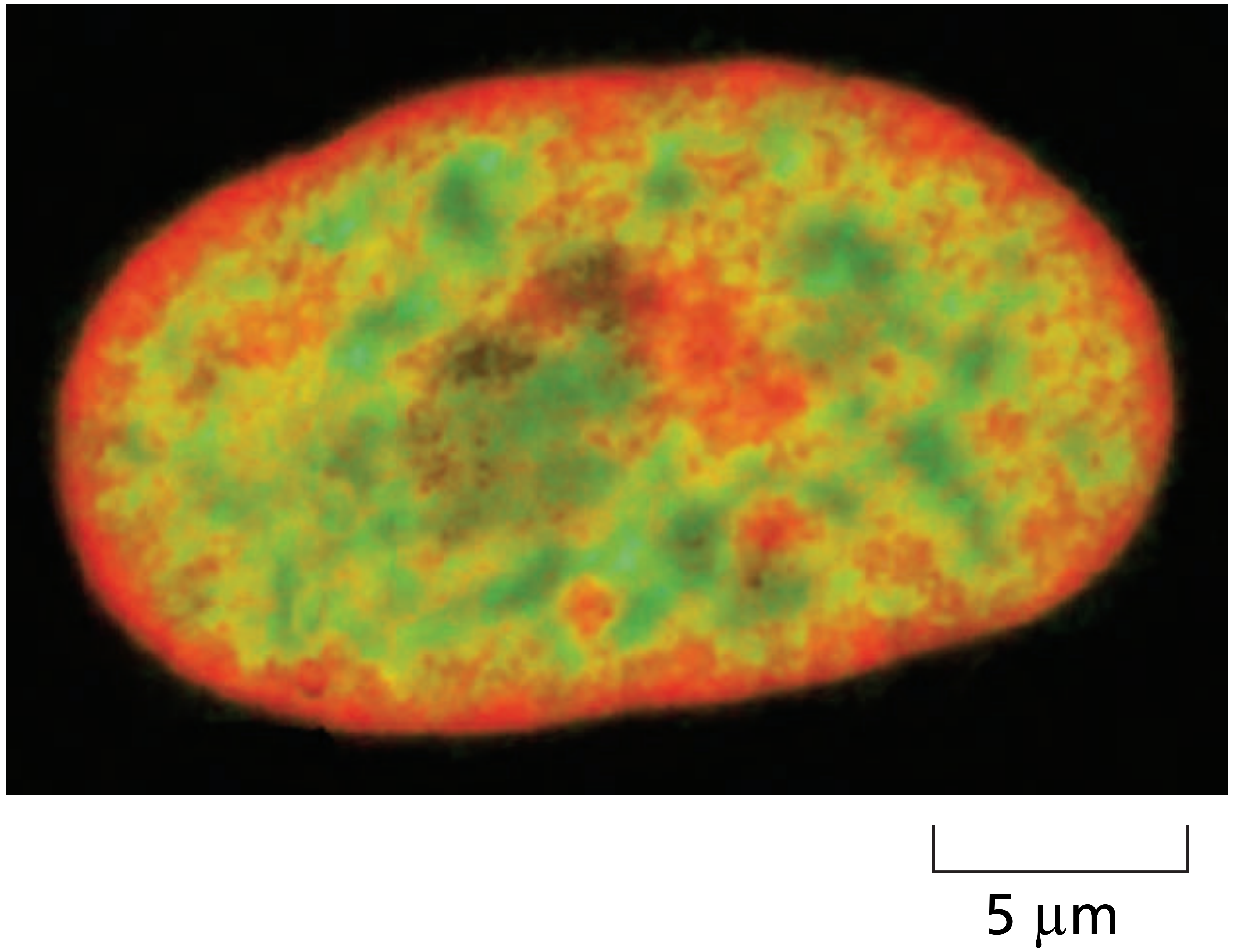
- Each of our 46 interphase chromosomes tend to occupy its own discrete territory



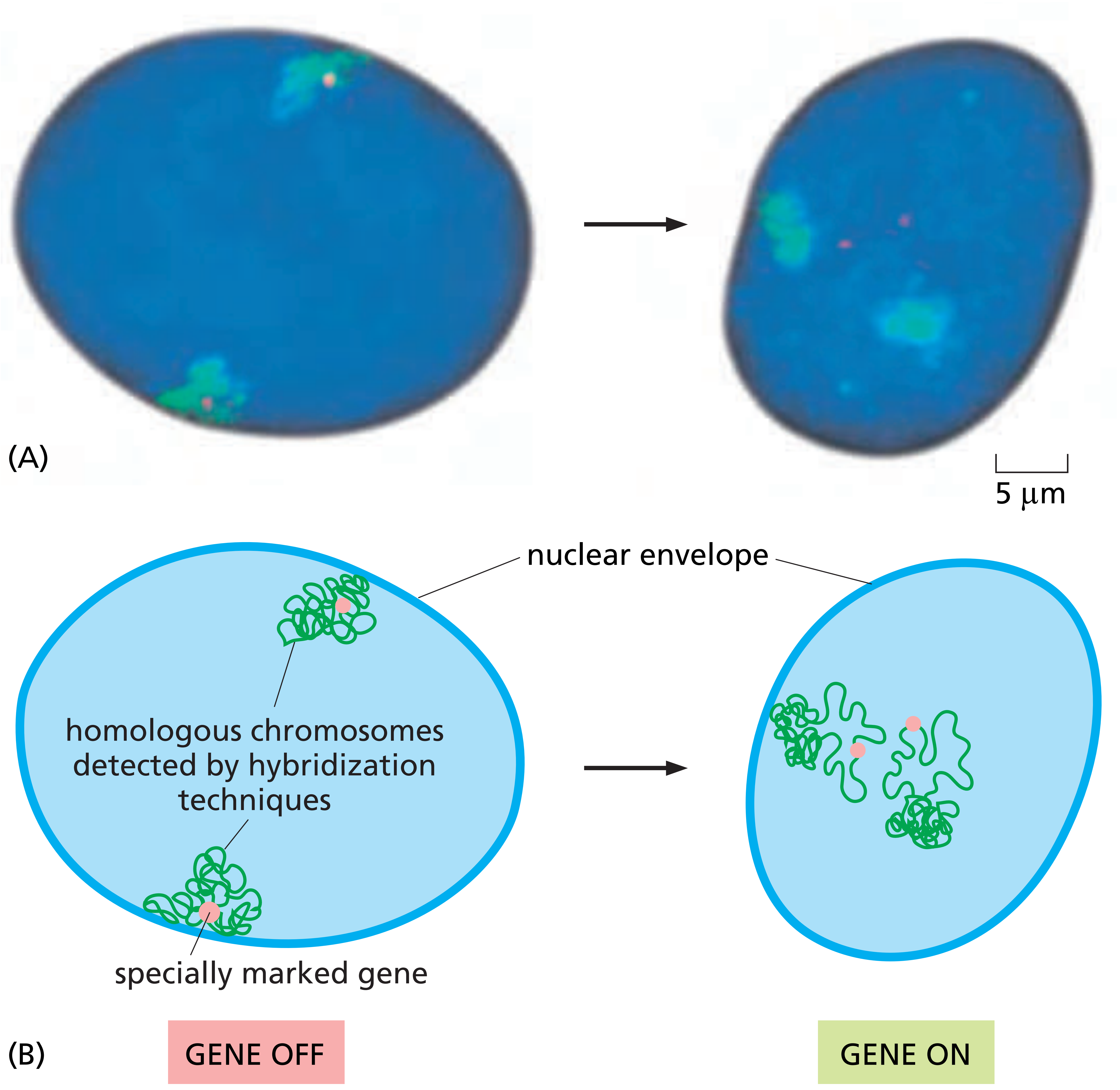


Chromatin is organised within the nucleus

- Heterochromatic regions are associated with nuclear lamina (red), whereas gene-rich regions are more central



Chromatin position can change with gene expression



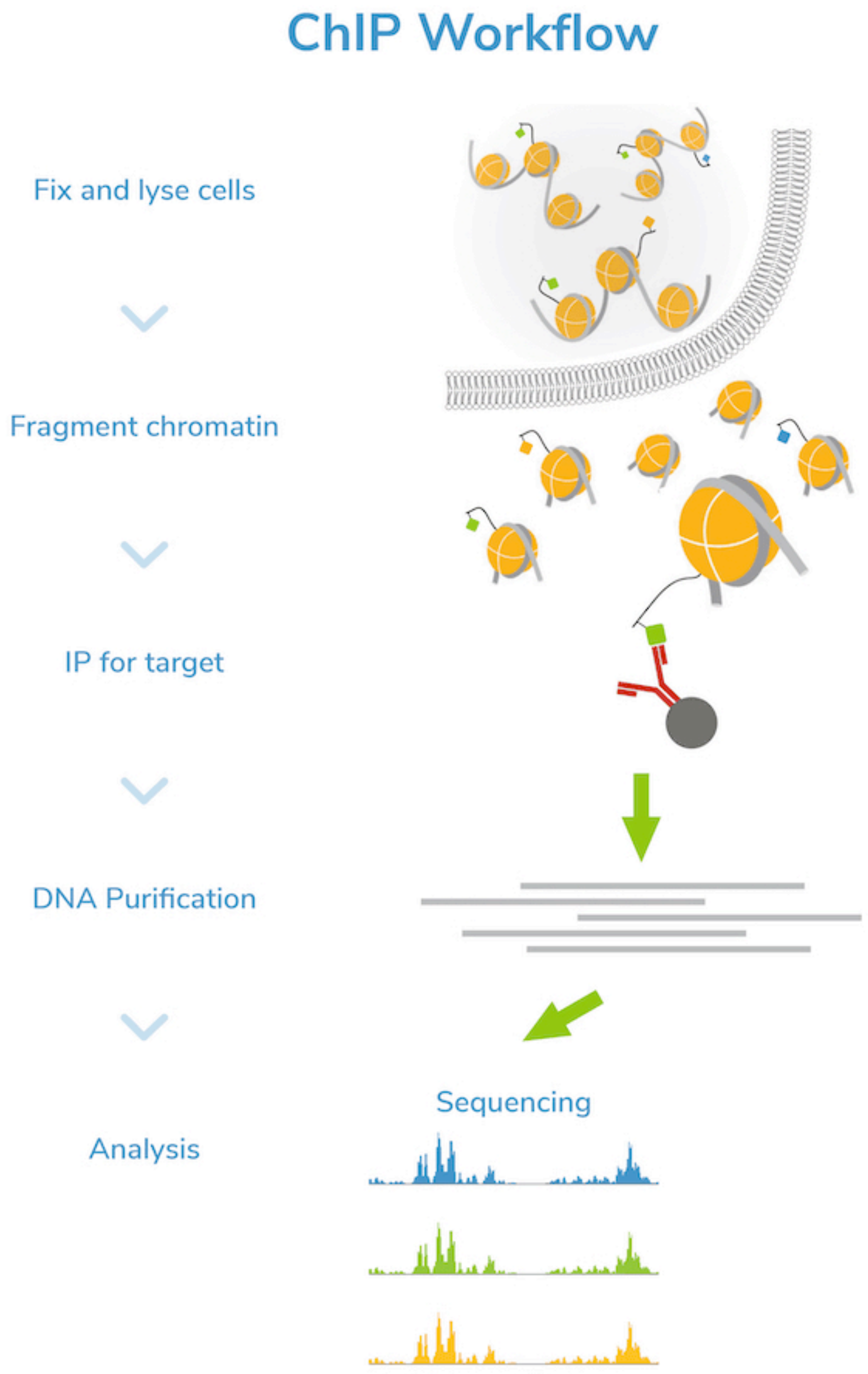
Plan

- Quick recap
- The global structure of chromosomes
- <u>Analysis of chromatin organization</u>
- How do genomes evolve

How to identify all the DNA regions for which histones have a given modification?

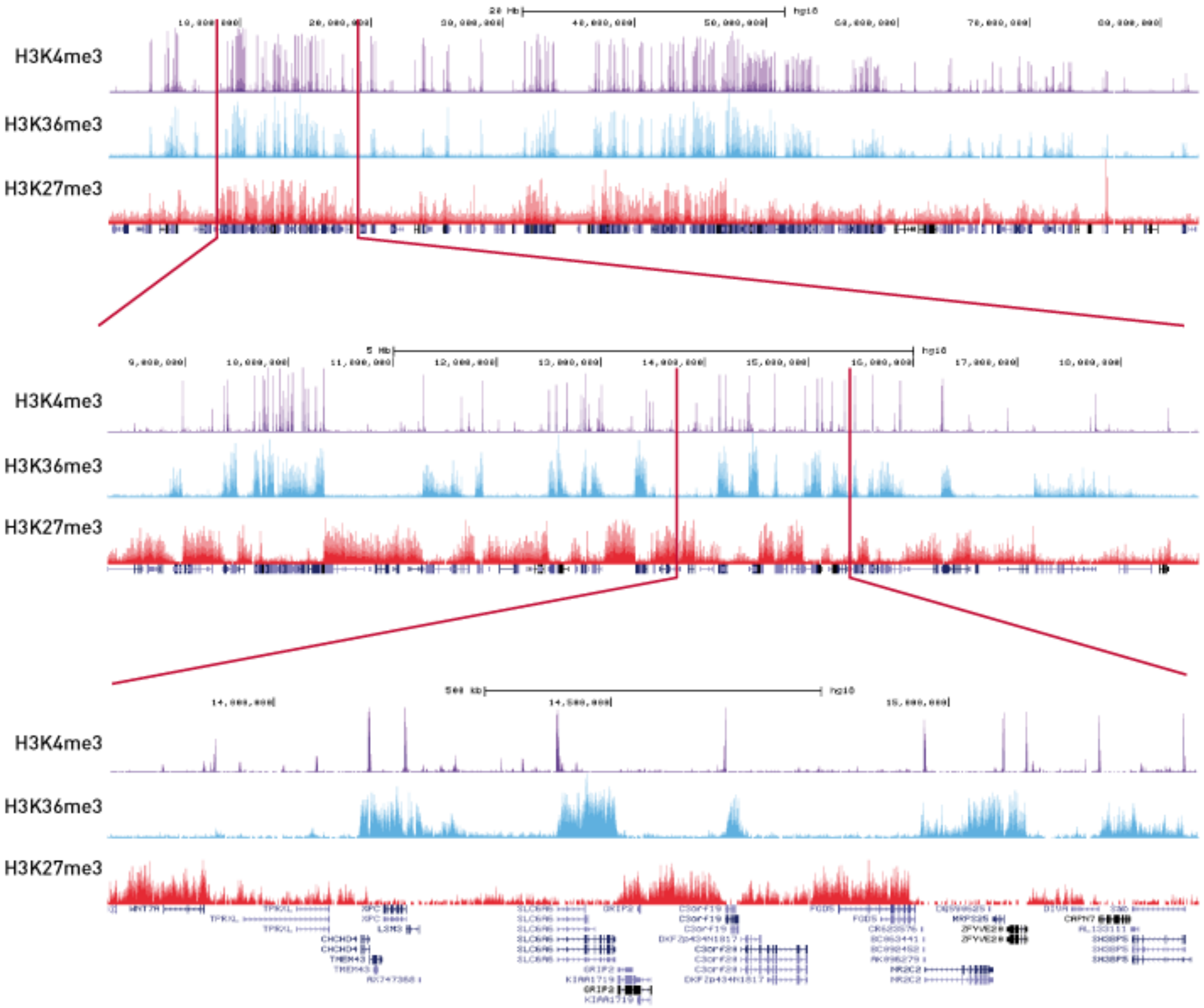
Chromatin Immune-precipitation (ChiP)

- Allows to determine **histone modifications** and their **location** in the genome
- **Crosslink** histones and DNA
- **Fragment** chromatin in small pieces
- Select a **marker** and use the corresponding **antibody**
- **Immunoprecipitation (IP)** to isolate the fragment of interest
- **Purify** the DNA (remove proteins)
- **Sequence** the DNA



Chromatin Immune-precipitation (ChiP)

- Results

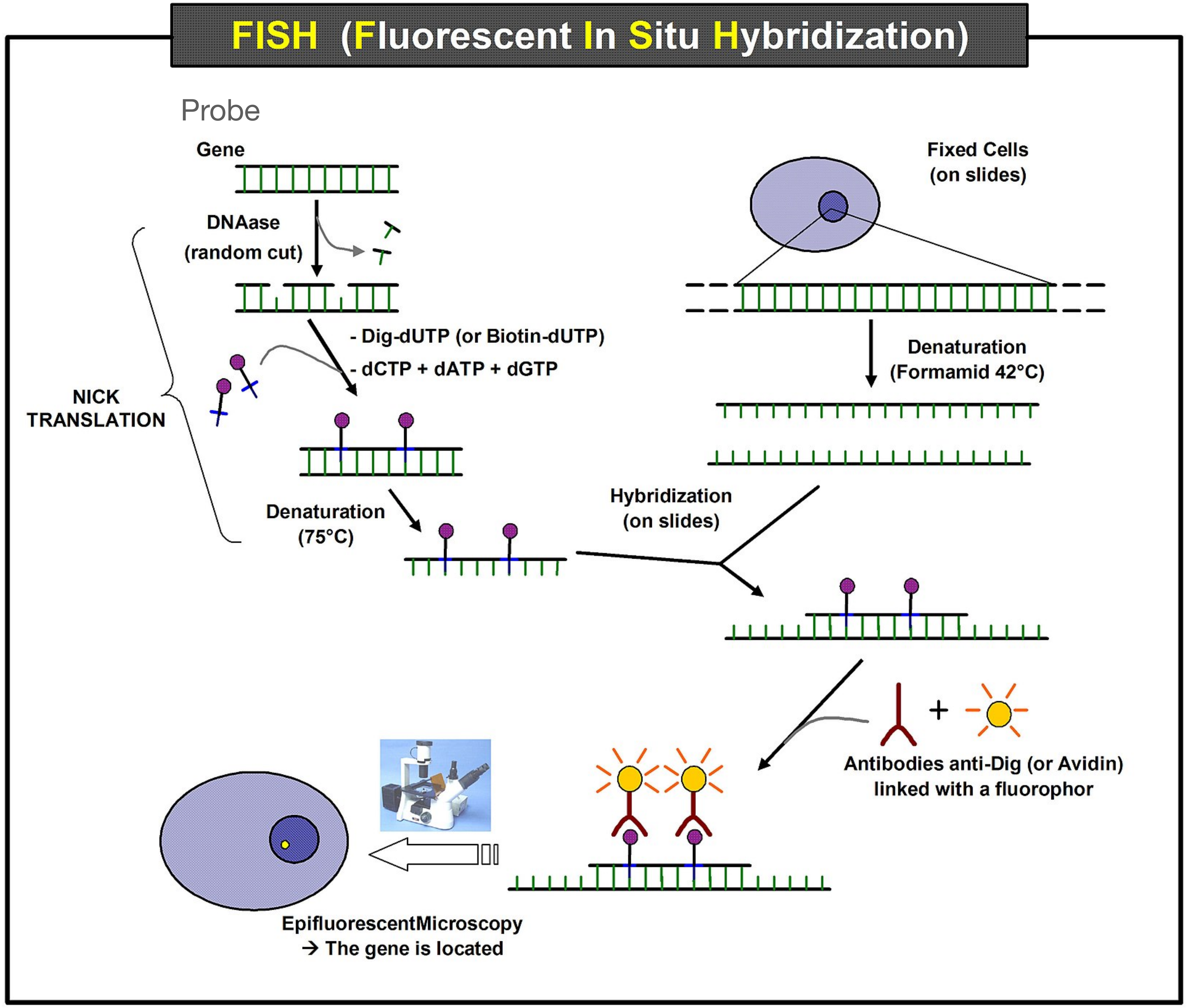


DNA region

How to detect specific sequences of DNA *in situ*?

Chromosome painting using FISH

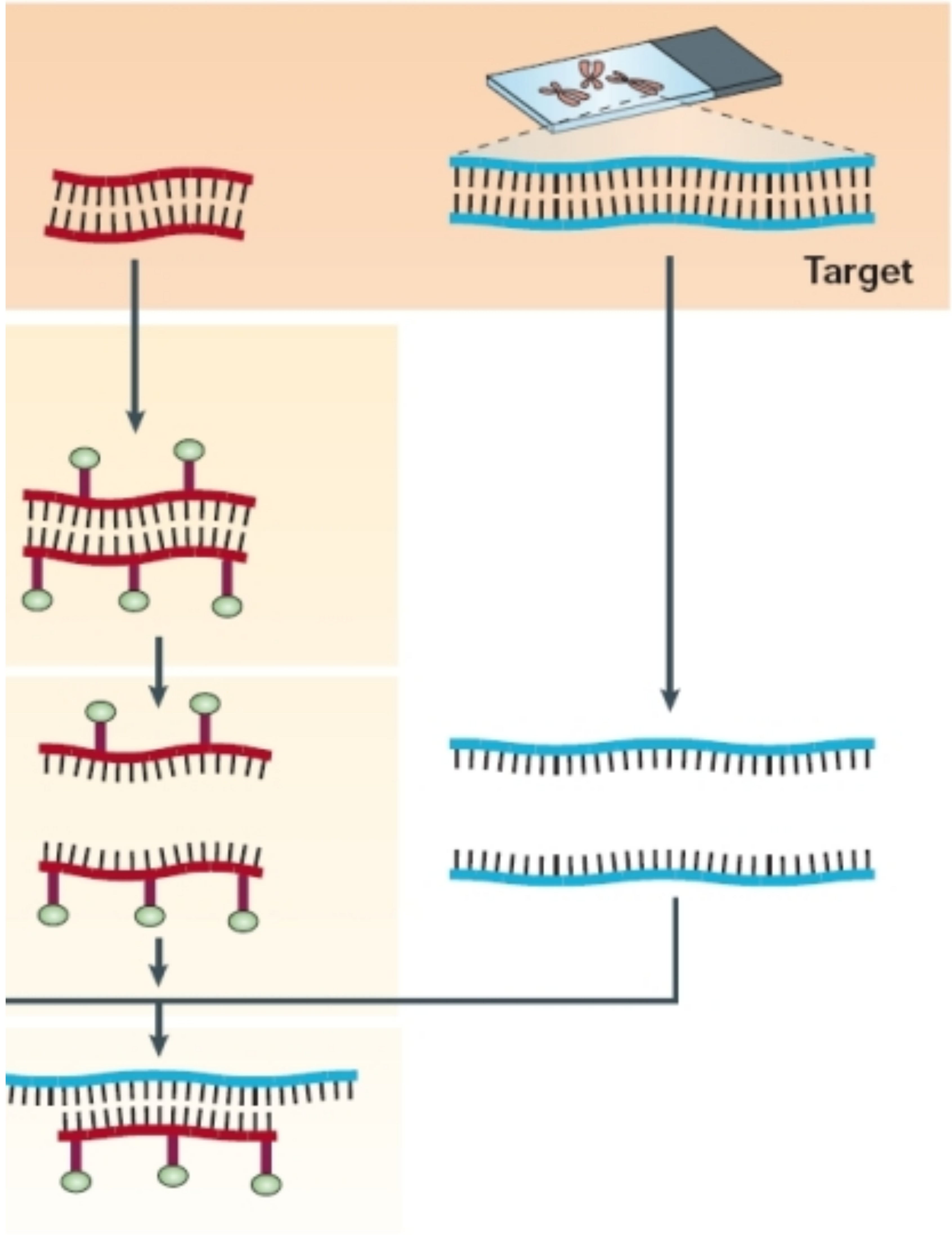
- Detect **specific regions of DNA** in the nucleus
- Cell **fixation**
- **Hybridization** of specific probes labeled with specific fluorescent markers
- **Analysis** of the cells



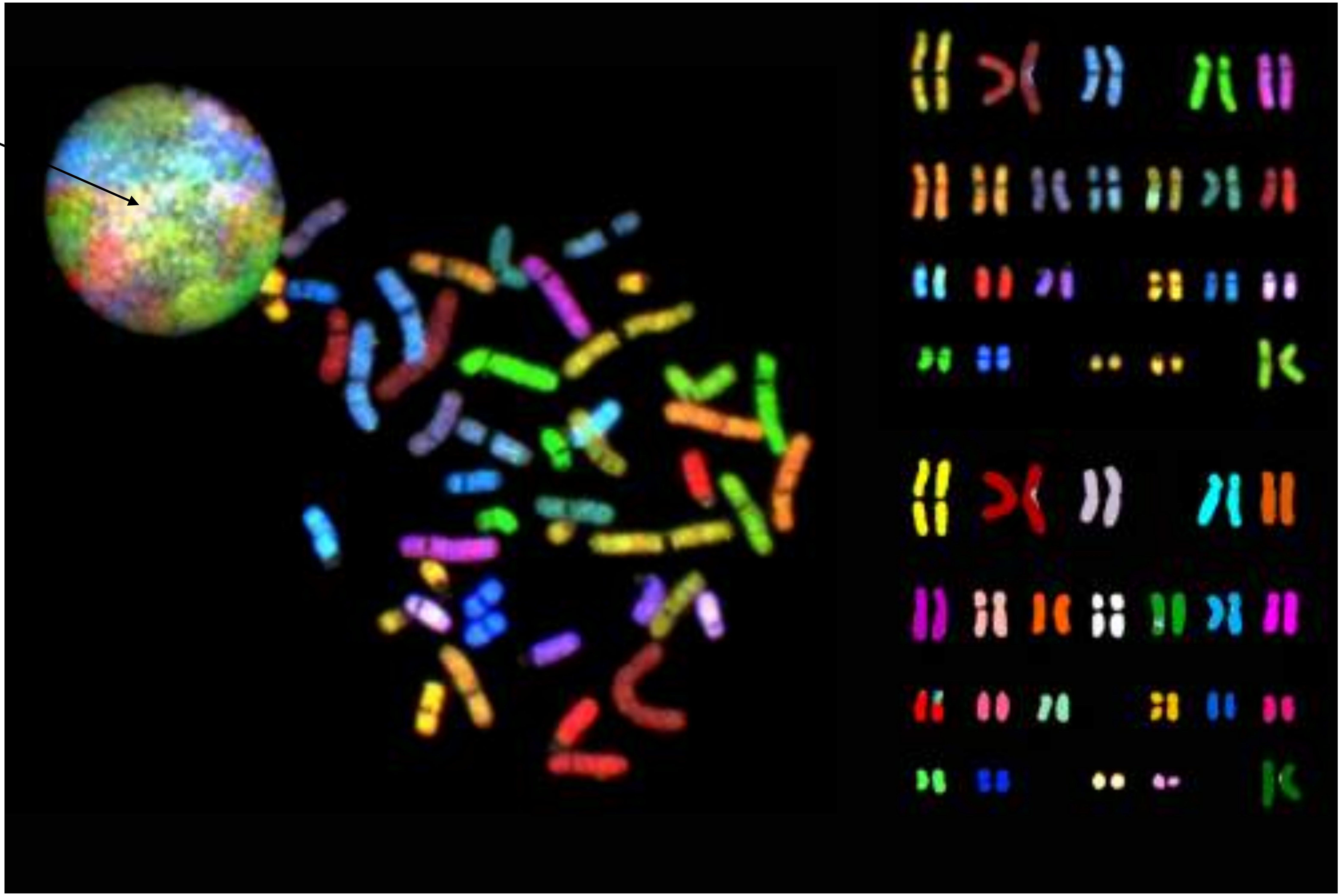
https://en.wikipedia.org/wiki/Fluorescence_in_situ_hybridization

Chromosome painting using FISH

- **Probe** = Single strand synthetic DNA oligonucleotide
- Sequence is **complementary** to the DNA sequence of interest
- Length is variable
- Coupled to **fluorescent marker**
- There are **chromosome-specific** probe libraries



Chromosome painting using FISH



Chromosome positioning (interphase)

Cytogenic analysis (metaphase)

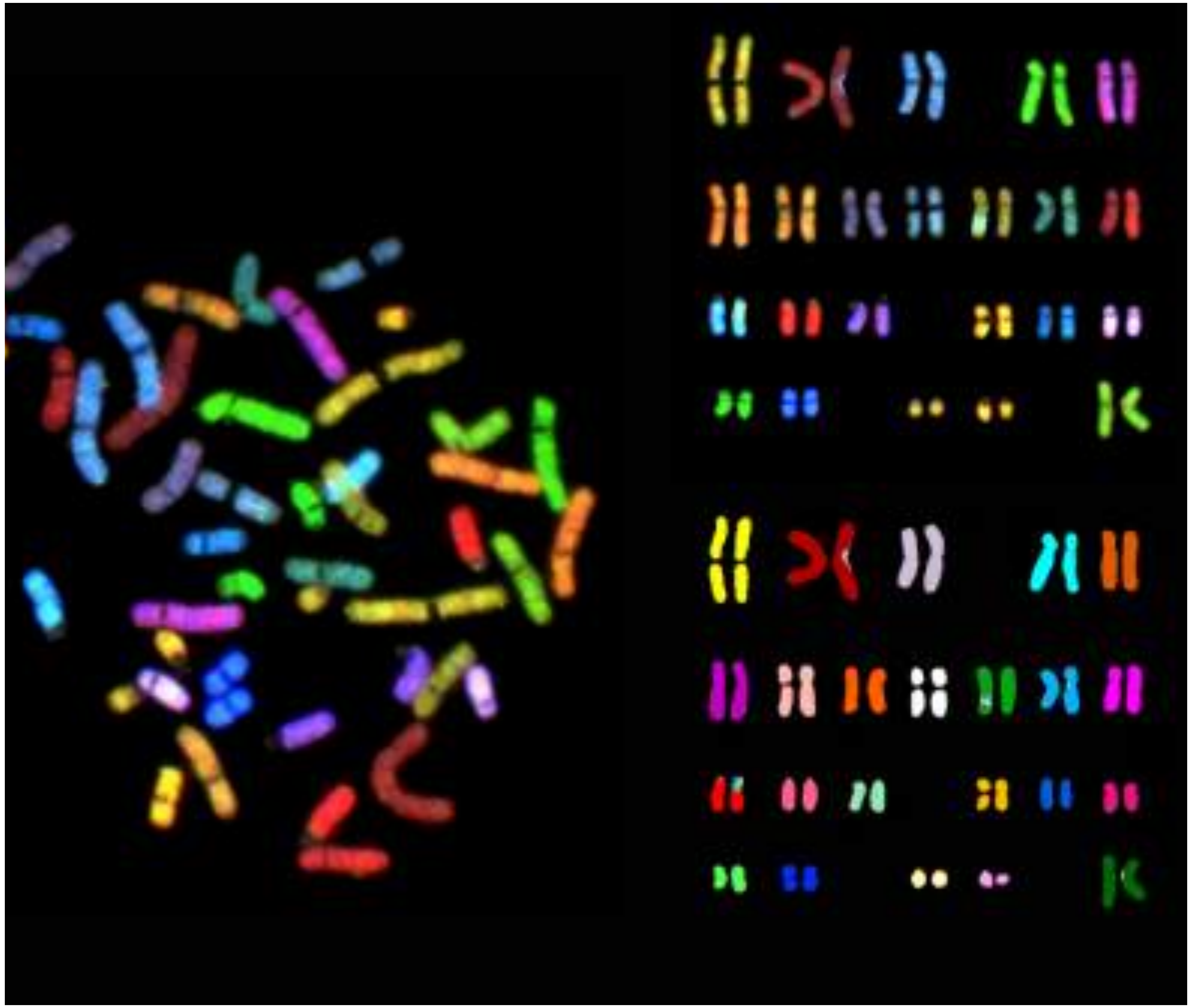
How to detect problems/changes in chromosomes?

Cytogenic analysis

- Allows to look for **changes in chromosomes** including broken, missing, re-arranged or extra chromosomes
- **Fix/Crosslink** the cells
- **Hypotonic solution** to spread the metaphase
- **Hybridization** of specific probes labeled with specific fluorescent markers
- Analysis of the cells in **metaphase**

Cytogenic analysis

- Results

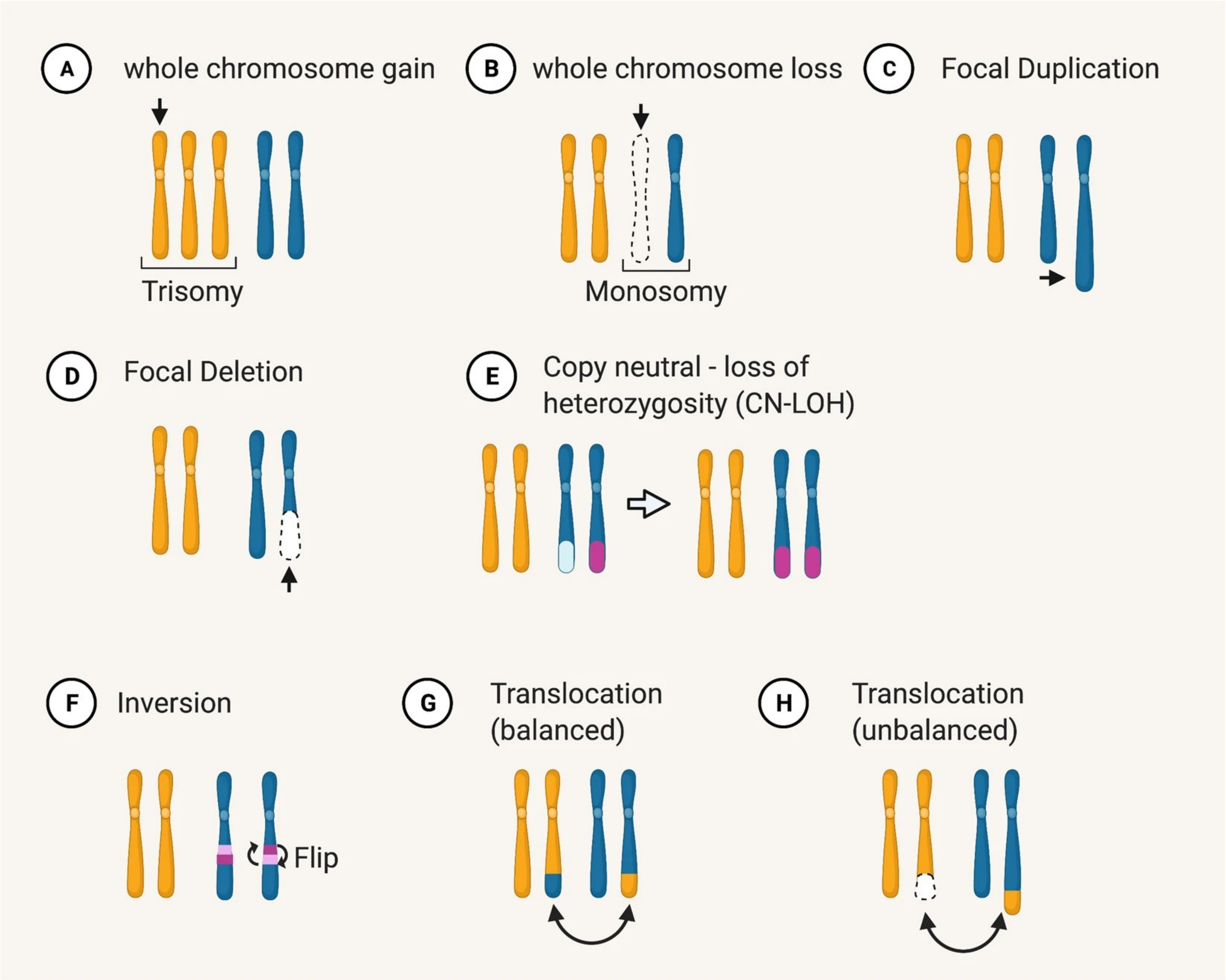


Spread chromosome in the metaphase

Paired chromosome to identify cytogenetic alterations

Chromosome alterations

- Affect multiple genes, large DNA fragments
- **Deletion** = loss of chromosomal regions (heterozygous or homozygous)
- **Duplications** = one or more additional copies of a chromosomal region
- **Translocation** = chromosomal rearrangement (balanced vs. unbalanced)

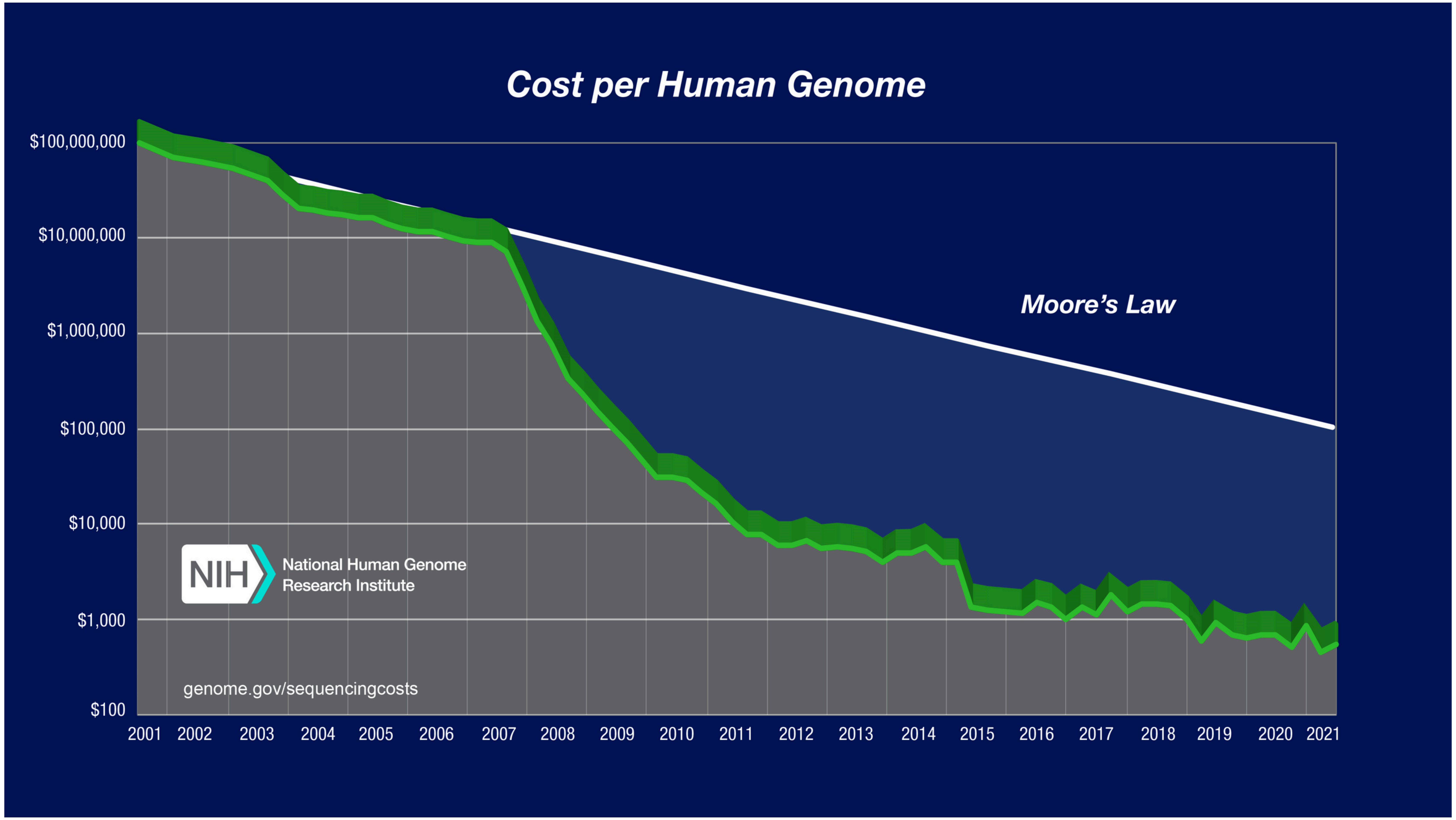


Liu, X., Kamatani, Y. & Terao, C. Genetics of autosomal mosaic chromosomal alteration (mCA). *J Hum Genet* **66**, 879–885 (2021). https://doi.org/10.1038/s10038-021-00964-4

Plan

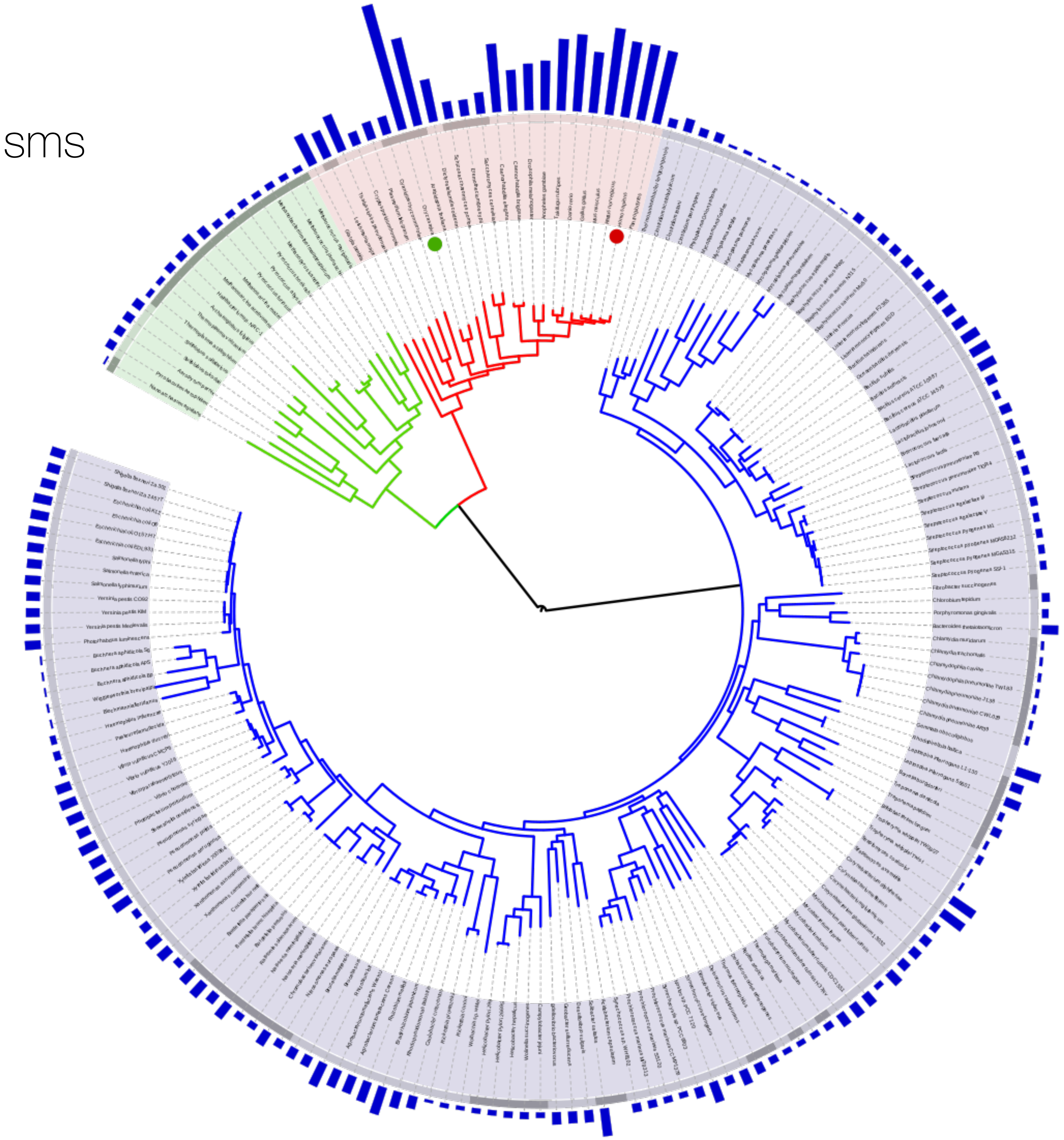
- Quick recap
- The global structure of chromosomes
- Analysis of chromatin organization
- How do genomes evolve?

Genome sequencing



Genome sequencing and comparison

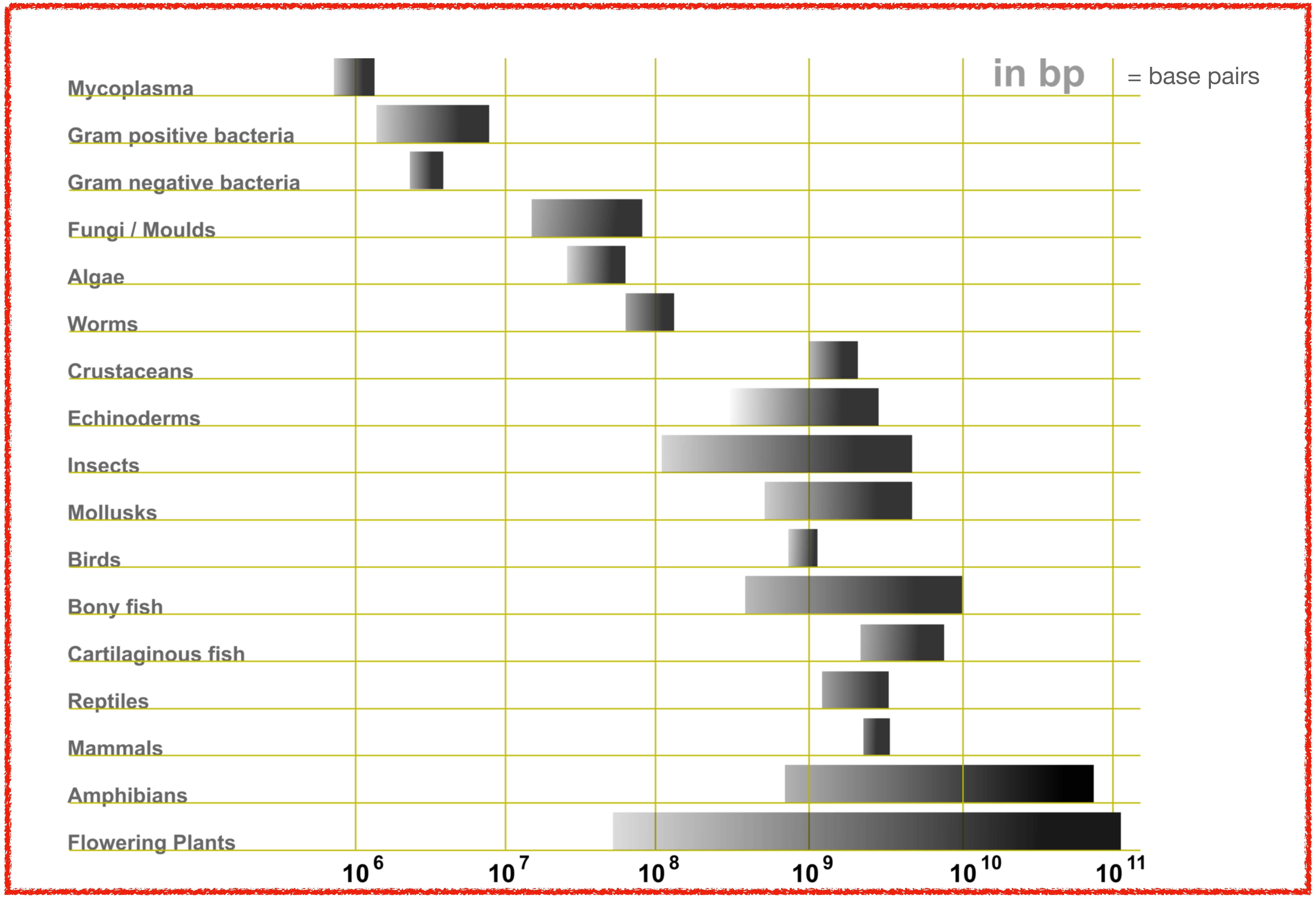
- Complete sequencing of the genomes of thousands of organisms
- Life’s chemistry is shared across all organisms



- Green = Archea
- Red = Eukaryotes
- Blue = Bacteria
- Bars represent genome sizes

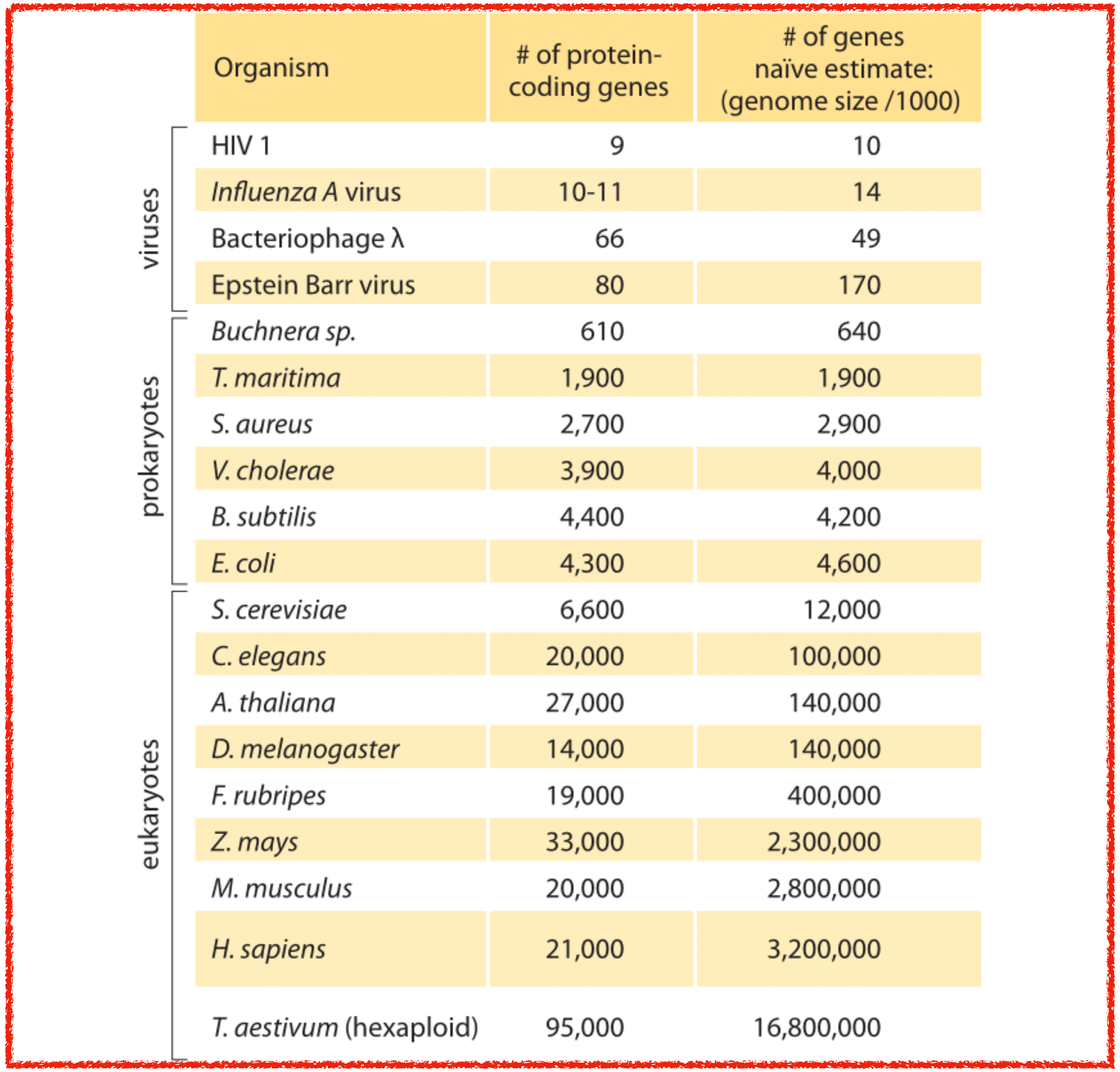
https://en.wikipedia.org/wiki/Genome_size

Genome comparison



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

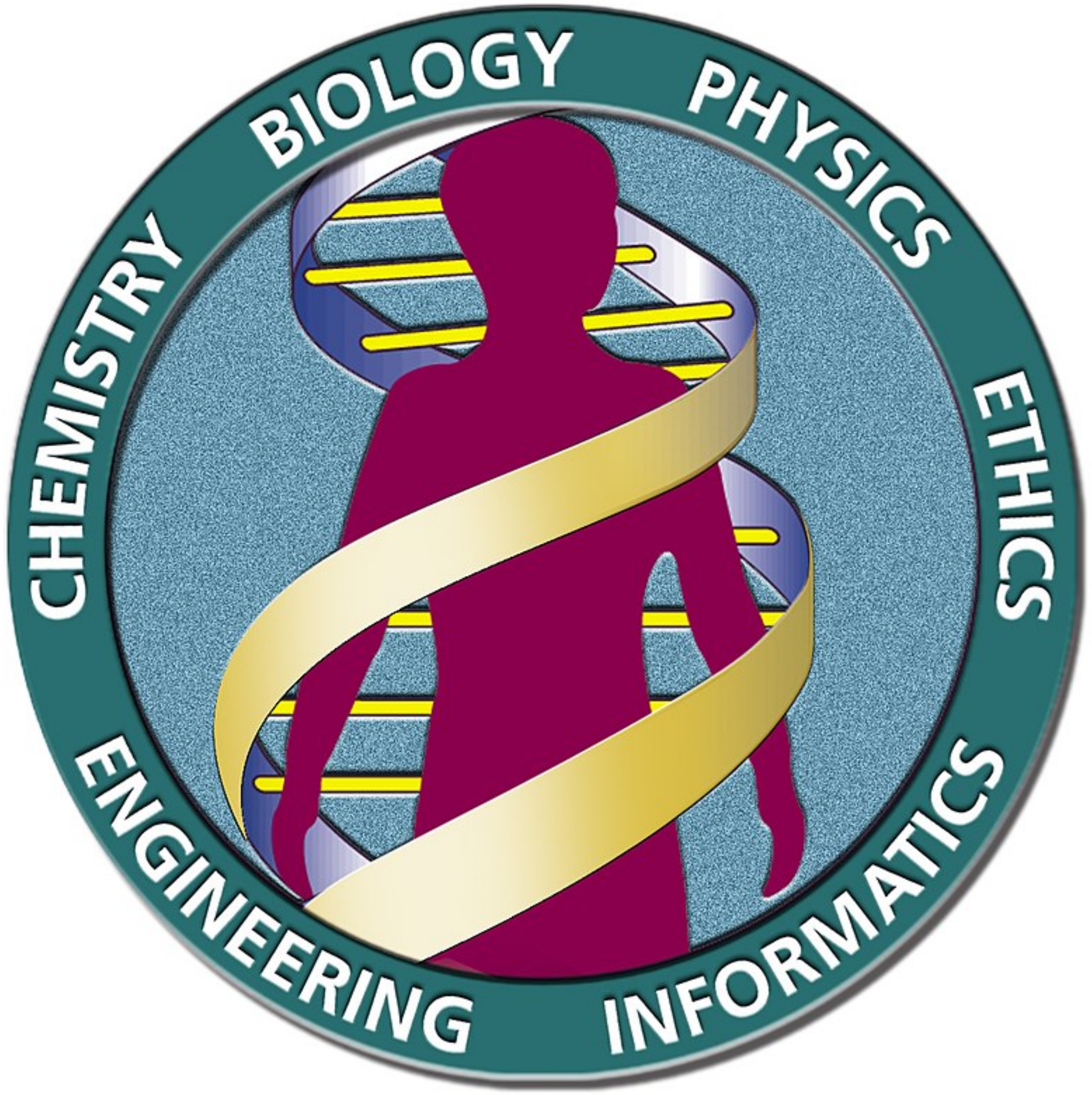
Genome comparison



	Organism	# of protein-coding genes	# of genes naïve estimate: (genome size /1000)
viruses	HIV 1	9	10
	Influenza A virus	10-11	14
	Bacteriophage λ	66	49
	Epstein Barr virus	80	170
prokaryotes	*Buchnera sp.*	610	640
	T. maritima	1,900	1,900
	S. aureus	2,700	2,900
	V. cholerae	3,900	4,000
	B. subtilis	4,400	4,200
	E. coli	4,300	4,600
eukaryotes	*S. cerevisiae*	6,600	12,000
	C. elegans	20,000	100,000
	A. thaliana	27,000	140,000
	D. melanogaster	14,000	140,000
	F. rubripes	19,000	400,000
	Z. mays	33,000	2,300,000
	M. musculus	20,000	2,800,000
	H. sapiens	21,000	3,200,000
	T. aestivum (hexaploid)	95,000	16,800,000

https://book.bionumbers.org/how-many-genes-are-in-a-genome/

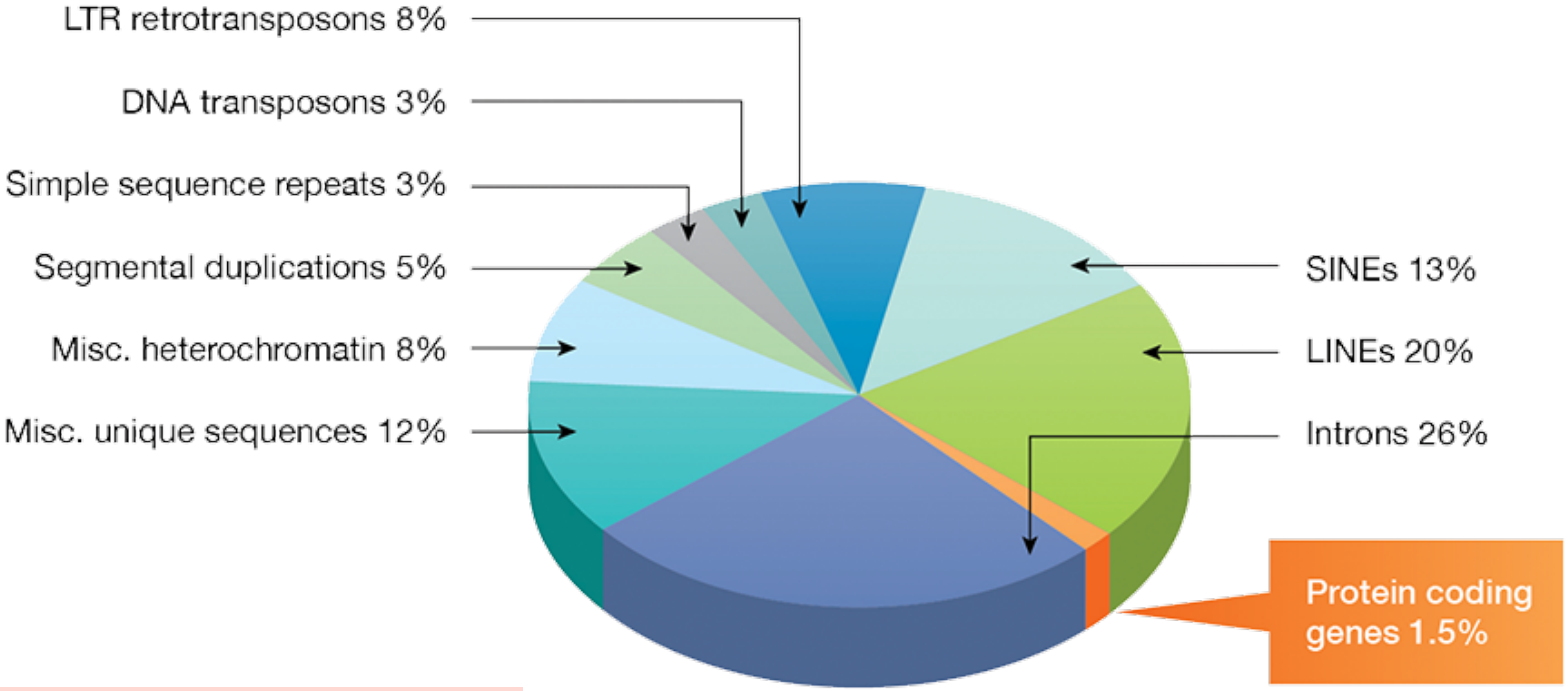
Human Genome Project



- started in 1990 and finished in 2003

Human Genome Project

- Sequence of the human genome: 3.1 billion nucleotide pairs
- 90% of it is probably unimportant
- The part of genomes that code for proteins (exons) are lost in a sea of DNA
- Not important to know the numbers



Eukaryotic genes include alternate coding and non-coding sequences
Coding sequences are **exons**
Non-coding sequences are **introns**

Human Genome Project

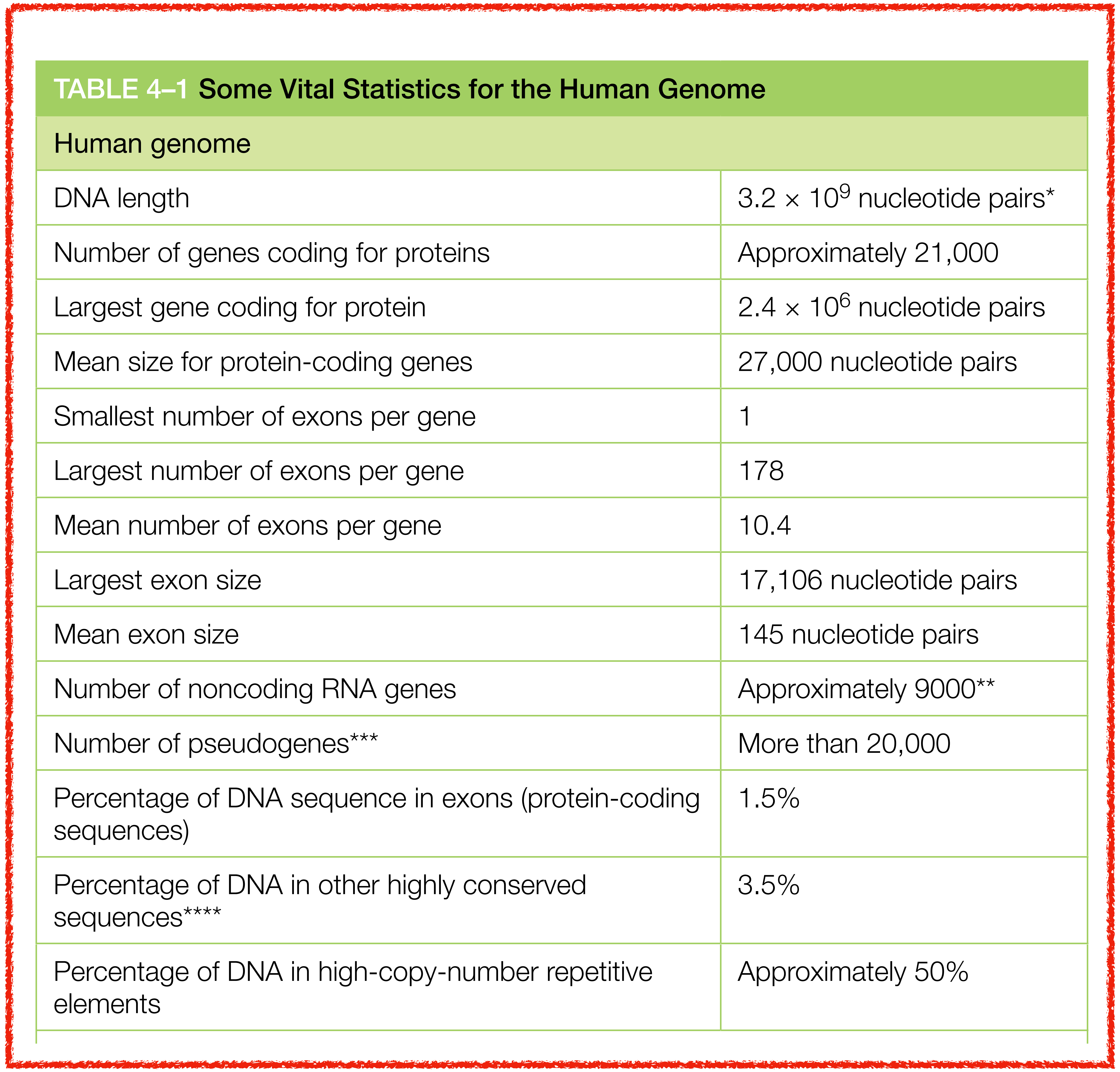


TABLE 4–1 Some Vital Statistics for the Human Genome

Human genome	
DNA length	3.2×10^9 nucleotide pairs*
Number of genes coding for proteins	Approximately 21,000
Largest gene coding for protein	2.4×10^6 nucleotide pairs
Mean size for protein-coding genes	27,000 nucleotide pairs
Smallest number of exons per gene	1
Largest number of exons per gene	178
Mean number of exons per gene	10.4
Largest exon size	17,106 nucleotide pairs
Mean exon size	145 nucleotide pairs
Number of noncoding RNA genes	Approximately 9000**
Number of pseudogenes***	More than 20,000
Percentage of DNA sequence in exons (protein-coding sequences)	1.5%
Percentage of DNA in other highly conserved sequences****	3.5%
Percentage of DNA in high-copy-number repetitive elements	Approximately 50%

How do we know all of this?

- By **comparing genomes**: Gene comparisons reveal **functional DNA sequences** by their conservation through evolution
- These include **functionally important exons**, **RNA molecules** and **regulatory DNA sequences** (and DNA sequences with function unknown so far)
- Non-conserved DNA regions are less likely to have a **critical function**
- These comparative DNA sequencing analyses have shown that **4.5% of the human genome consists of multispecies conserved sequences (with only 1/4 encoding proteins)**

Homologous genes are genetic sequences inherited from a **common ancestor**. They are similar in sequence and can perform similar functions. They can be in a single **organism** or different ones.

How do we know all of this?

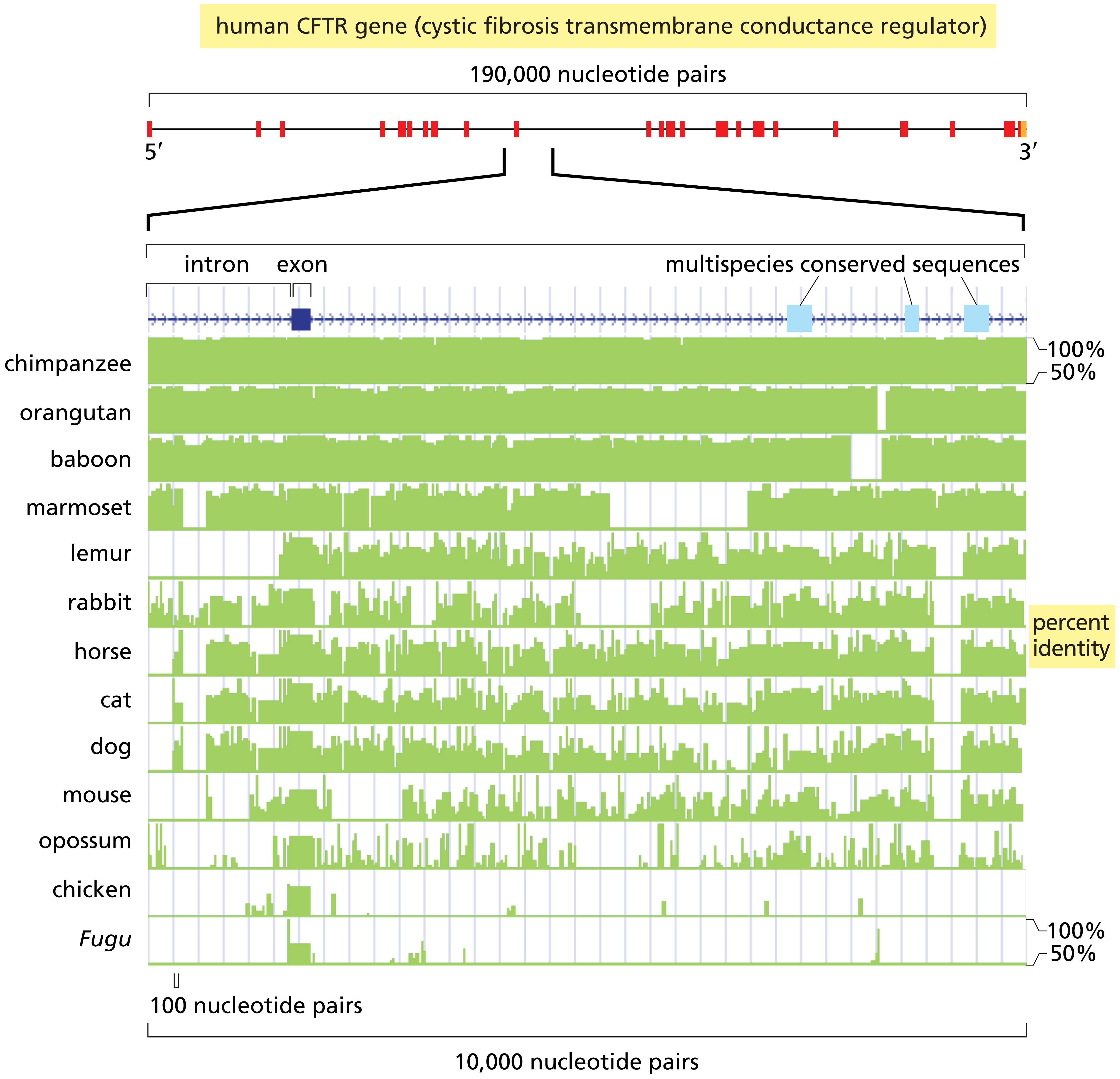
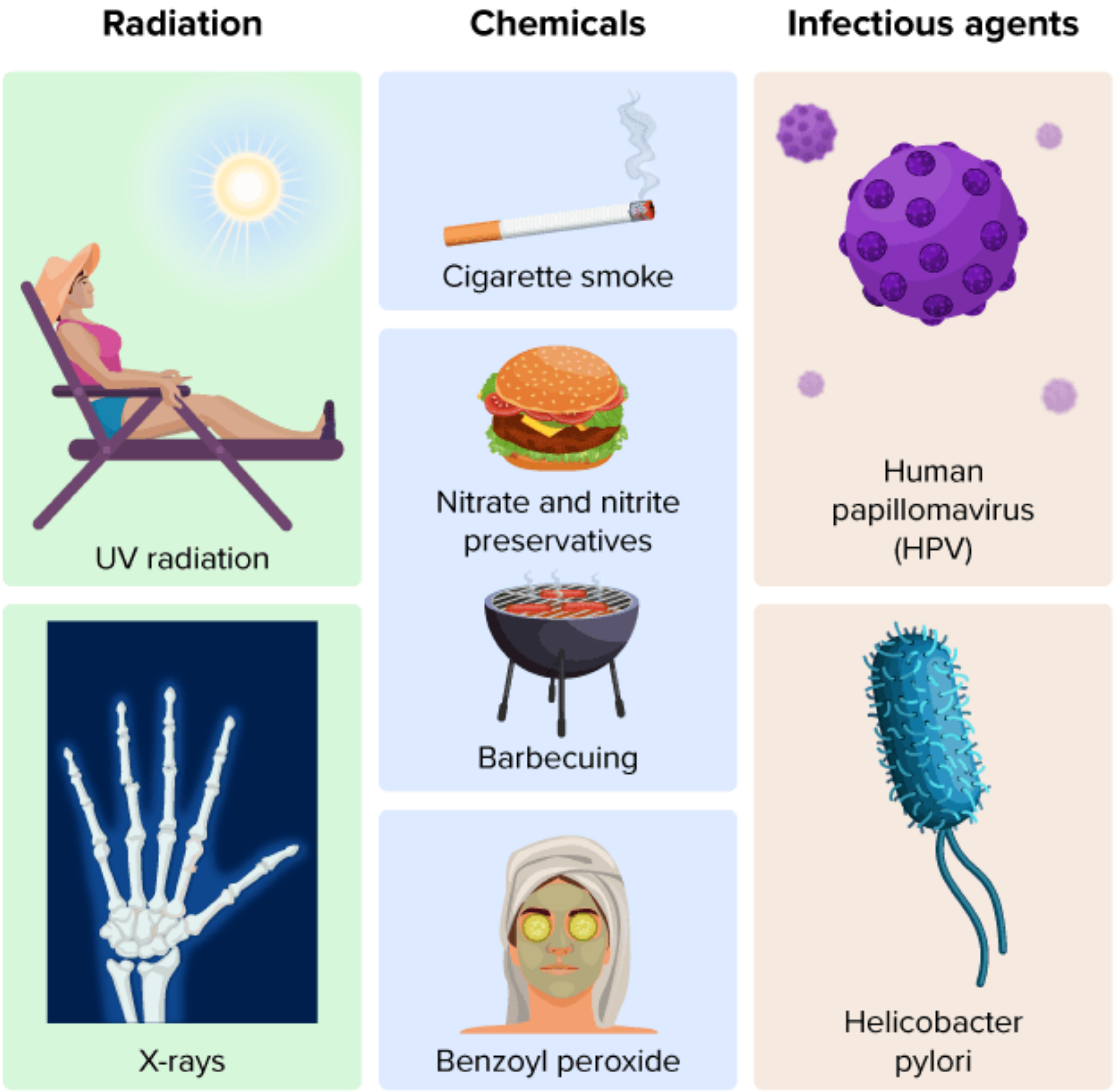


Figure 4–73 The detection of multispecies conserved sequences. In this example, genome sequences for each of the organisms shown have been compared with the indicated region of the human CFTR (cystic fibrosis transmembrane conductance regulator) gene; this region contains one exon plus a large amount of intronic DNA. For each organism, the percent identity with human for each 25-nucleotide block is plotted in *green*. In addition, a computational algorithm has been used to detect the sequences within this region that are most highly conserved when the sequences from all of the organisms are taken into account. Besides the exon (*dark blue* on the line at the top of the figure), the positions of three other blocks of multispecies conserved sequences are indicated (*pale blue*). The function of most such sequences in the human genome is not known. (Courtesy of Eric D. Green.)

How do genomes actually evolve?

- Evolution depends on **accident and mistakes** followed by **non-random survival**
- **Failures** in the mechanisms by which genomes are **copied or repaired**
- When errors (mutations) happen in **germ cells**, they are passed on to the next generation
- Errors are **“rare”** events: ~ 1 in 10^8 per generation (implying that each gamete has in average 30 mutations)

How do genomes actually evolve?



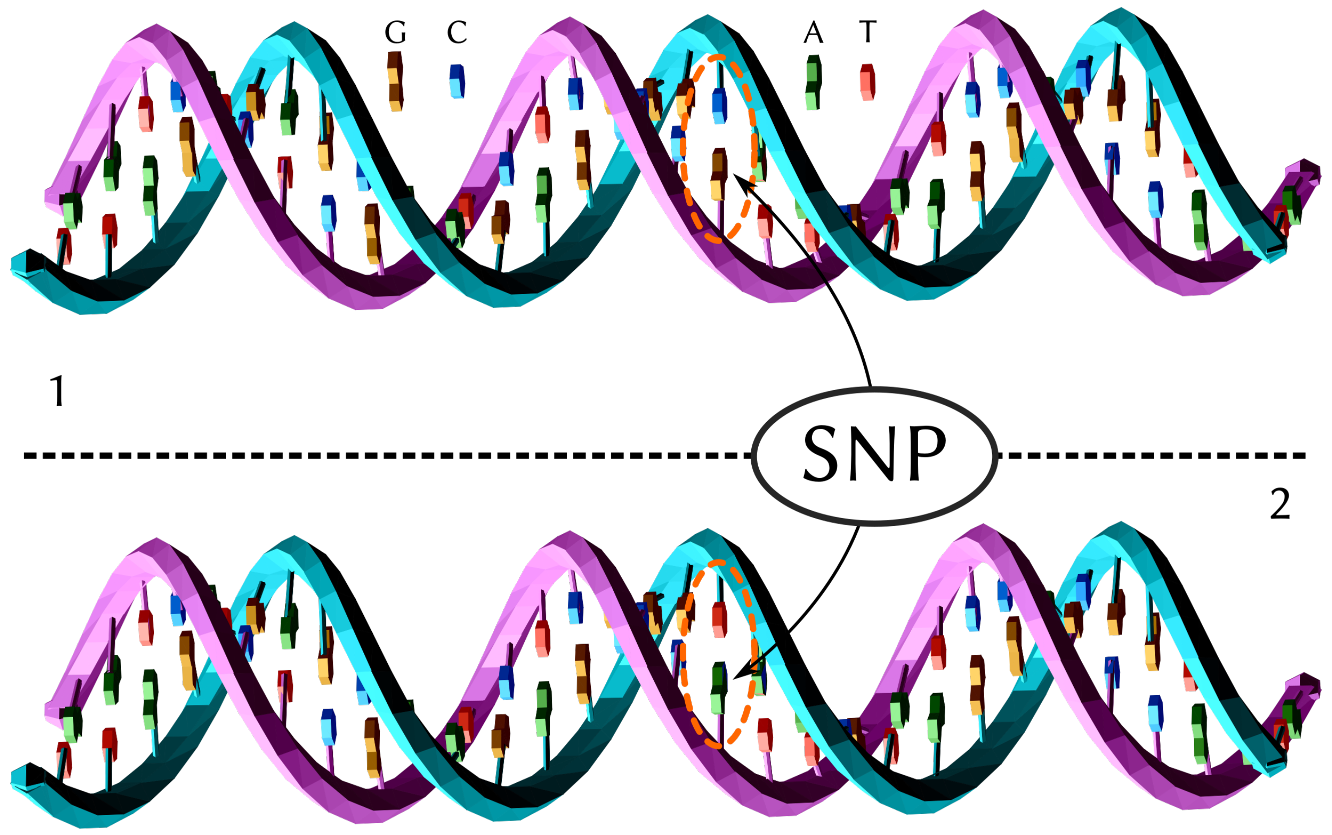
How do genomes actually evolve?

What are the different **types of mutations**?

- **<u>Simple, local changes - point mutations</u>**
- Large-scale genome rearrangements - **deletions, duplications, inversion, translocations**
- In addition, important role of **mobile genetic elements**

Single nucleotide polymorphism (SNP)

- = substitution of **one nucleotide** at a **given position** in the genome
- In some definitions, we call **SNP** a mutation that is present in a **fraction of the population** (e.g. >1%)



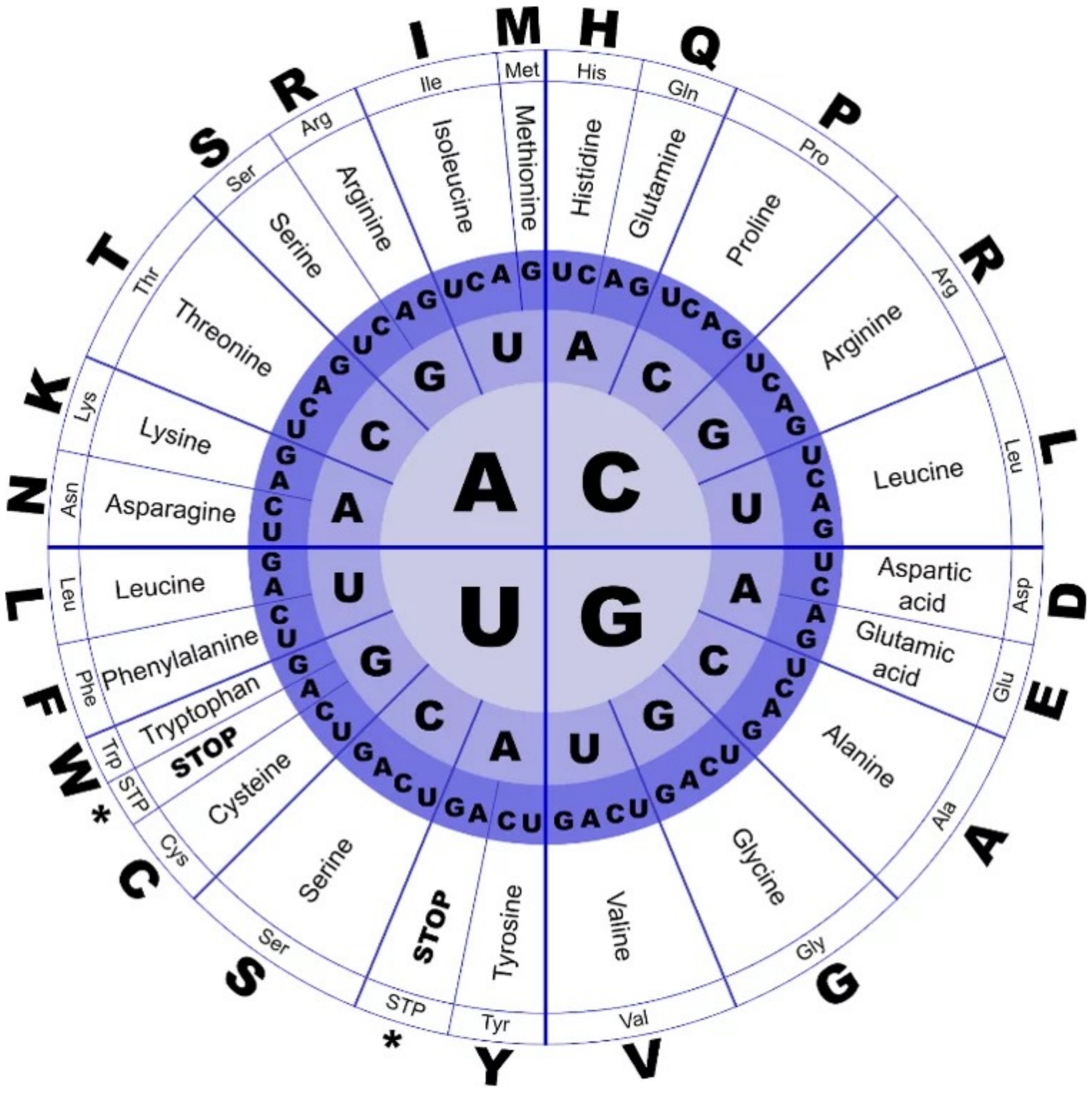
in 90% of the population

in 10% of the population

- The two possible variations of this SNP (G or A) are called **alleles**
- When occuring in genes, SNPs can be **synonymous** (no change in AA) or **non-synonymous** (AA is changed)

https://en.wikipedia.org/wiki/Single-nucleotide_polymorphism#/media/File:Dna-SNP.svg

Single nucleotide polymorphism (SNP)



Single nucleotide polymorphism (SNP)

1. **SNPs (Single Nucleotide Polymorphisms)**

- Definition: A SNP is a single base-pair variation in the DNA sequence among individuals.
  Example: At a specific location, one person might have A, another G.
- Frequency: For it to be considered a SNP, the variant must occur in at least 1% of the population.
- Abundance: SNPs are the most common type of genetic variation — millions are scattered across the human genome.
- Relevance:
  - Many SNPs are neutral (no effect).
  - Some are associated with disease risk, drug response, or physical traits.
  - They can act as markers to track genetic regions linked to traits.

Single nucleotide polymorphism (SNP)

2. GWAS (Genome-Wide Association Studies)

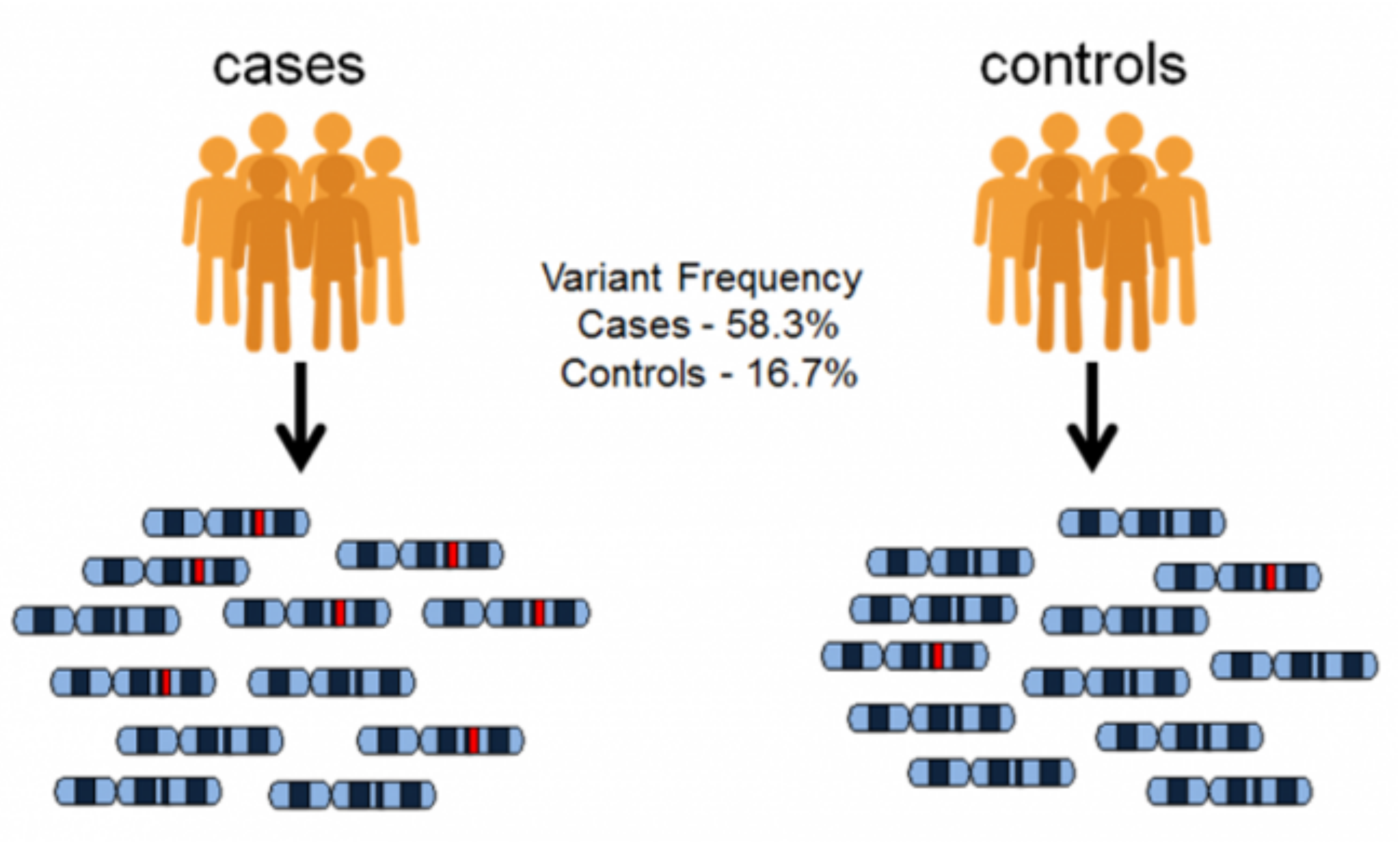
- Definition: A GWAS is a large-scale study that scans the genome for SNPs statistically associated with a trait (e.g., height, diabetes, schizophrenia).
- Method:
  1. Collect DNA from thousands of people (cases vs. controls, or continuous trait values).
  2. Genotype them at hundreds of thousands to millions of SNPs.
  3. Use statistical tests to find SNPs that occur more frequently in people with the trait.
- Output: A list of SNPs (or genomic regions) associated with the trait.

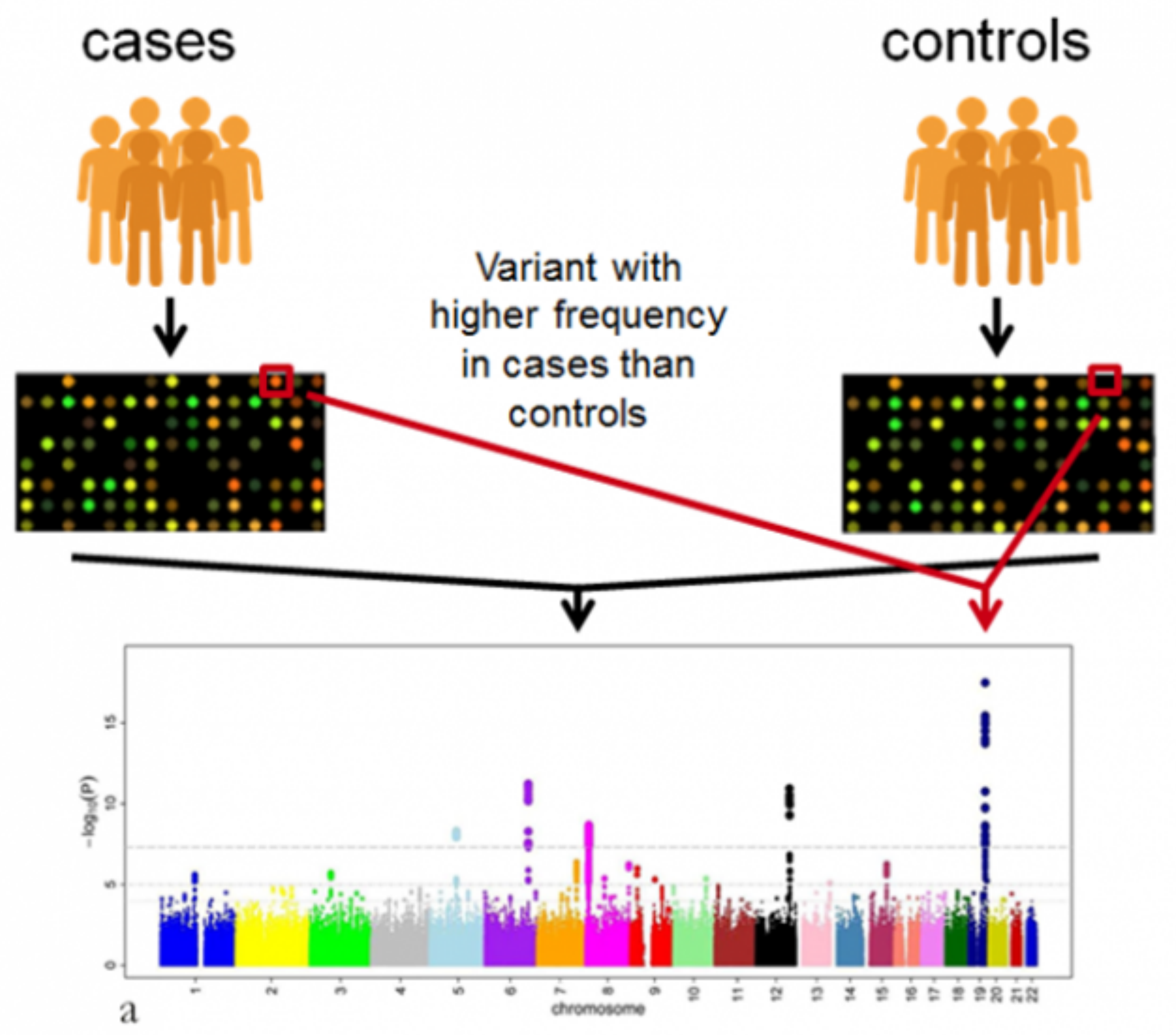
3. Importance of SNPs in GWAS

- SNPs serve as markers, not necessarily causal variants themselves.

Single nucleotide polymorphism (SNP)

- **SNPs** sometimes help to explain the **susceptibility** to a wide-range of **diseases** across the population
- Genome wide association studies (**GWAS**) are hypothesis-free methods for identifying associations between **genetic regions** (loci) and **traits** (including diseases).
- Typical GWAS studies collect data to find out the common variants in a number of individuals, both with and without a common trait (e.g. a disease), across the genome, using genome wide **SNP arrays**





https://www.ebi.ac.uk/training/

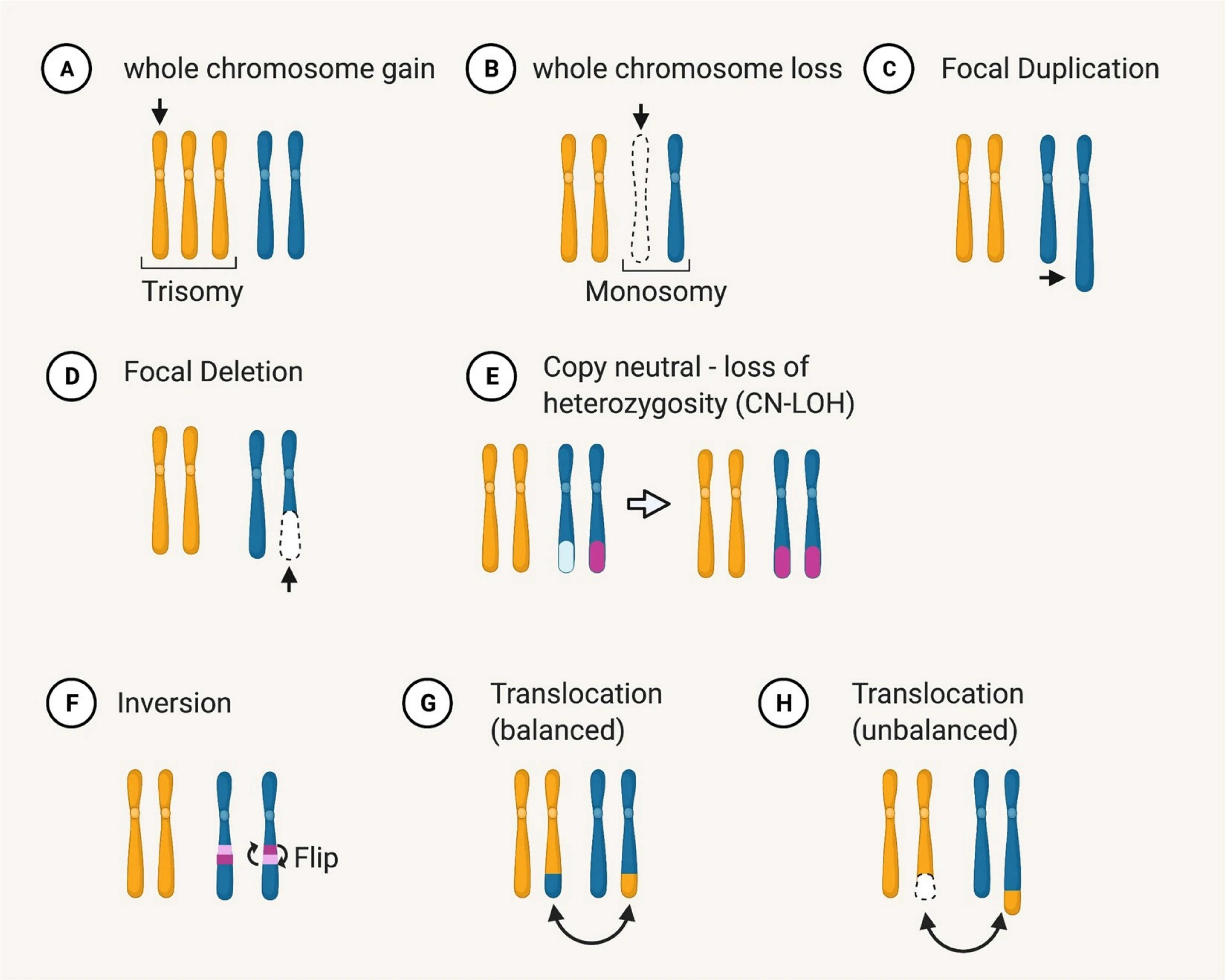
How do genomes actually evolve?

What are the different **types of mutations**?

- Simple, local changes - **point mutations**
- **Large-scale genome rearrangements - deletions, duplications, inversion, translocations**
- In addition, important role of **mobile genetic elements**

Chromosome alterations

- Affect multiple genes, large DNA fragments
- **Deletion** = loss of chromosomal regions (heterozygous or homozygous)
- **Duplications** = one or more additional copies of a chromosomal region
- **Translocation** = chromosomal rearrangement



Liu, X., Kamatani, Y. & Terao, C. Genetics of autosomal mosaic chromosomal alteration (mCA). *J Hum Genet* **66**, 879–885 (2021). https://doi.org/10.1038/s10038-021-00964-4

How do genomes actually evolve?

What are the different **types of mutations**?

- Simple, local changes - **point mutations**
- Large-scale genome rearrangements - **deletions, duplications, inversion, translocations**
- **In addition, important role of mobile genetic elements**

Have a nice day!