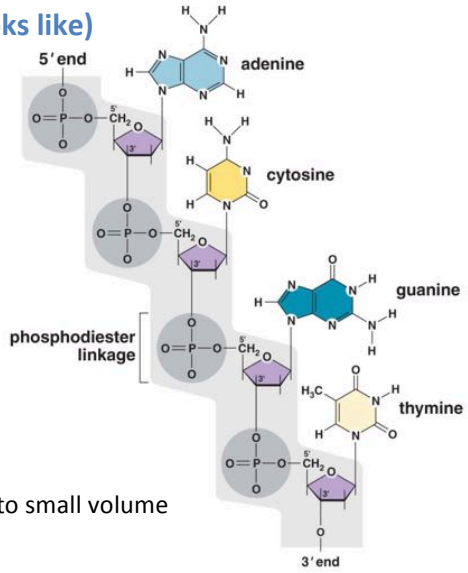
MBLG2071 – Semester 1

L2: Introduction to the structure of the genome

Review chemical composition and structure of DNA (what is looks like)



DNA is:

- biopolymer made up of nucleotides (base, sugar and phosphates) linked together by phosphodiester bonds
- Double stranded → nucleotides form 'complimentary' base pairs
    - A → T
    - G → C

Appreciate the need for DNA compaction

- Eukaryotes have **significantly more bp** than procaryotes
    - (humans = 2m while E. coli = 1.3mm) → need to reduce length into small volume

Understand at molecular level the packaging of DNA into chromosomes

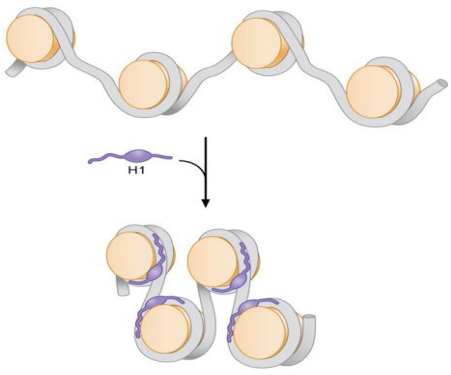
1. Each chromosome has single molecule of DNA
2. DNA wrapped around histone core octamer to form nucleosomes
3. Nucleosomes compacted in 30 nm fibre by histone H1 → further condenses
4. Loops of DNA held together by nuclear scaffold (proteins such as topoisomerases II, SMC etc.)→ further higher order packaging

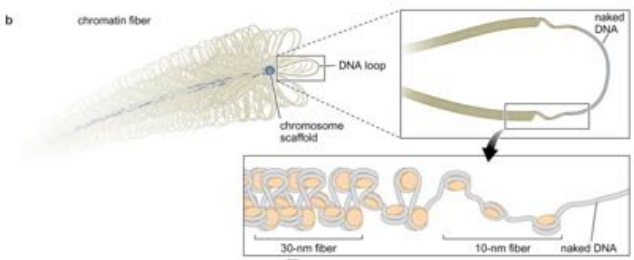
Procaryotes

- DNA organised into 50-100 domains/loops
- **Independently** supercoiled
- RNA and protein contribute to folded structure = "folded genome"

Eucaryotes

- Packaging begins with formation of **nucleosome**
- DNA wound around small proteins called **histones**
- Histones have lots of lysine and arginine residues, making them very positively charged at pH 7 → attracts negatively charged phosphates on DNA

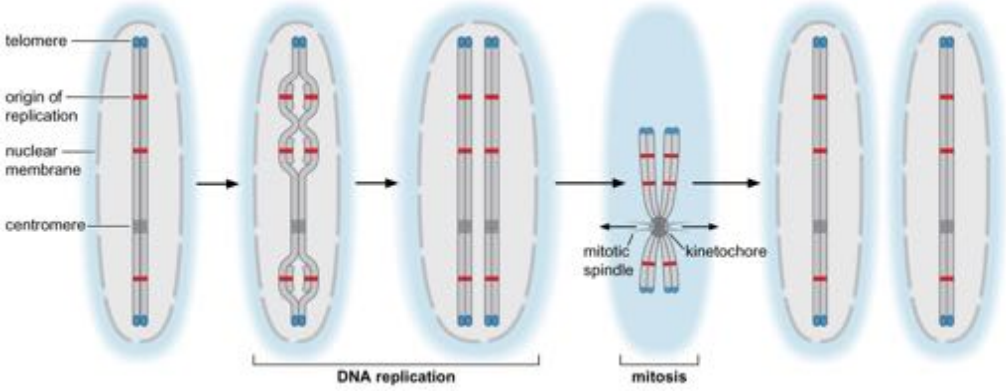




30 nm fibre

Loops of 40-90 kb held together by **nuclear scaffold** (includes topoisomerase II and SMC)

Learn about the features of chromosomes



- **Chromosome** = DNA + histones + non-histone proteins
- Several features allow their maintenance and transmission to next generation

1. Origin of replication

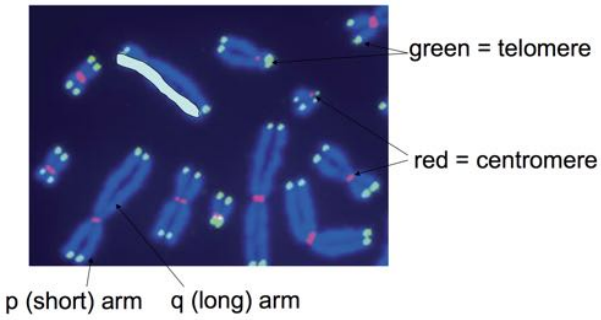
- Specific genomic DNA sequences that direct the initiation of DNA replication
- Eukaryotes have **multiple origins of replication** every 30-40 kb throughout the length of each chromosome (

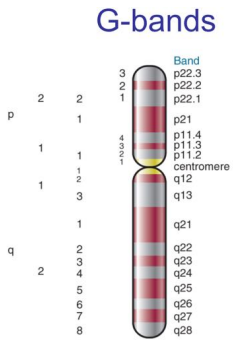
2. Centromere

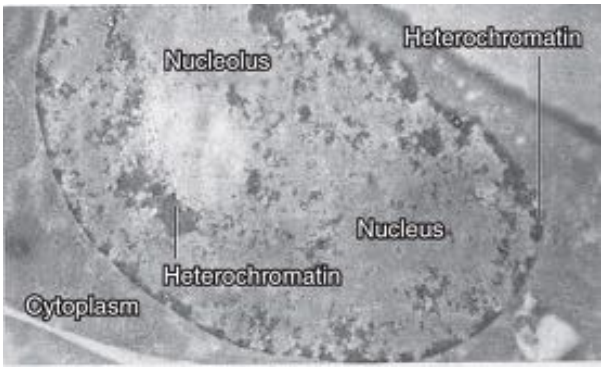
- Segregates chromosomes at mitosis & meiosis
- contain large amounts of repetitive sequences
- are relative markers on a chromosome
- presence of centromere divides chromosome into two parts called arms
- arms are not the same length
    - short arm = p
    - long arm = q

3. Telomere (TTAGGG)

- located at the ends of chromosomes
- simple repeating structure **(TTAGGG)**
- allow cells to recognize ends of chromosomes and not something to be repaired







Learn how DNA packaging relates back to phenotypic function of DNA

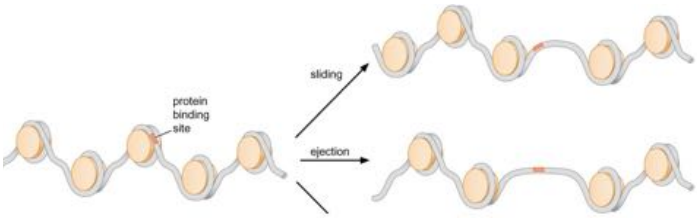
- Even in euchromatic regions, DNA is associated with histones

How do the transcription factors and other regulatory proteins gain access to the DNA?

- DNA association with histone octamer is **dynamic** (non-covalent interaction= constant release + rebinding)
- Access to hidden DNA done through **Chromatin remodelling & Chromatin modification**
    - Independent of DNA sequence

1. Chromatin remodelling – moving nucleosomes

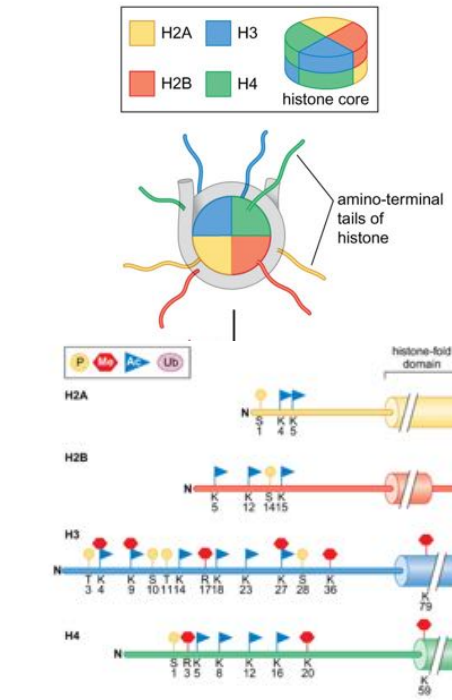
- Performed by large protein complexes = nucleosome
- Remodelling complexes
- Active process ⟶ requires hydrolysis of ATP
- Modification types:
    - sliding
    - displacement



---

2. Chromatin modification – adding or removing chemical groups from histone proteins

- Each histone proteins in the core octamer has an **N-terminal "tail"**
- Tail consists of **20 AAs** that protrude from the nucleosome between the turns of DNA ⟶ can bind to chemical groups to change structure/shape of histone ⟶ affects packaging



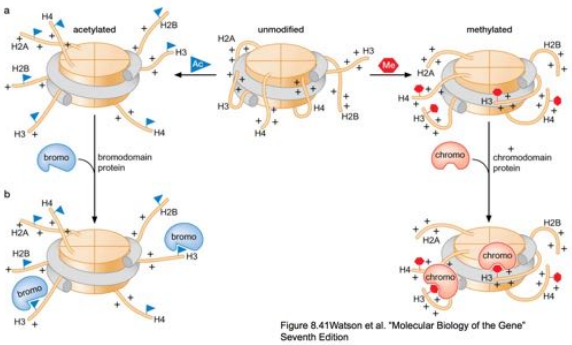


Figure 8.41Watson et al. "Molecular Biology of the Gene" Seventh Edition

- histone acetyl transferases (HATs) add acetyl groups to lysine residues and histone deacetylases (HDACs) remove them

L3: Genomic sequence complexity

Appreciate that genome size is not directly related to an organism's complexity (C-value paradox)

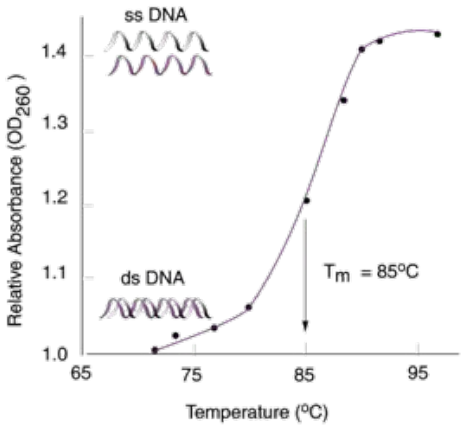
- The C-value describes an organism's genome size
- It is the # of picograms of DNA in a haploid set of that organism's chromosomes → but:
    - huge disparity between C-value and organism complexity

	C-value
Human	3.5
Funnel-web	≈ 5
Atlantic ocean shrimp	15
Salamander	35

Understand how we study the composition of genomes (C_ot plots)

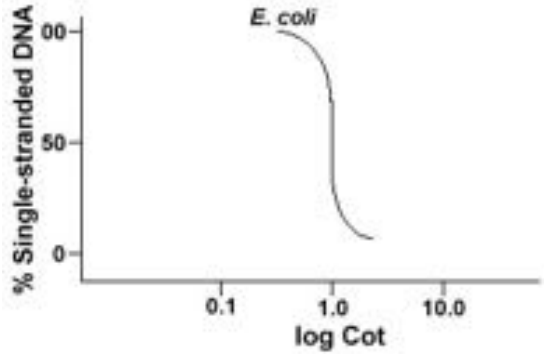
Study composition of genomes to resolve C-value paradox

- dsDNA can be separated by heating (melting)
- monitor using UV spectrophotometry
- aromatic DNA bases absorb at 260 nm
- in dsDNA bases are less accessible → less absorbance
- ssDNA = more exposed bases = greater absorbance



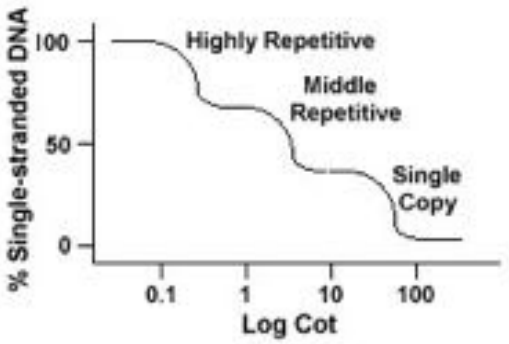
- Extract sample DNA
- Cut into manageable lengths (use restriction endonucleases)
- Melt DNA → ssDNA
- Monitor **re-annealing** (reassociation of dsDNA) by UV absorbance
- plot%SSDNA on y-axis and time

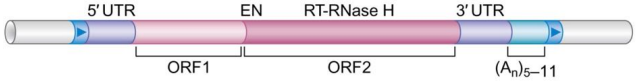
  (initial concentration * time = C_ot) as log scale on x axis = C_ot plot



Learn about the 3 types of sequence that make up the human genome

- Complex DNA sample = human genome
- Tri-phasic during **annealing procedure**
- 3 kinds of sequences
- Repetitive sequences **reanneal more quickly** (higher probability for annealing to occur due to greater numbers)



Sequence Type	Types	Explanation
1. Highly repetitive sequences	**Microsatellites** **(≈ 3% of human genome)**	• tandem arrays of **short repeat units** (usually < 13 bp) • most common are **dinucleotide repeats** (e.g. *gtgtgtgtgt*) • often clustered around centromere (spindle attachment site) and telomeres (ends of chromosomes) • associated with heterochromatin (tightly packaged
2. Middle repetitive sequences	**Multi-copy protein coding gene**	• increases amount of protein made • histones
	tRNA sequences	
	rRNA sequences	
	SINEs *(short interspersed nuclear elements)*	• **Non - autonomous** poly-A retrotransposons • most common is *Alu* repeat (contains recognition sequence of restriction enzyme *Alu*) • ≈300bp – over 1 million copies (10% of genome) 
	LINEs *(long interspersed nuclear elements)*	• **Autonomous** poly-A retrotransposons • make up ≈20% of human genome • most common human LINE is LINE-1 (L1) → 6000 bp sequence → appears 50K x in human genome 
3. Single Copy	**Protein Coding genes**	some sequences present in a few copies eg globins (has 4 copies of beta-globin coding protein)

Appreciate that the # of protein coding genes is not directly related to an organism's complexity **(G-value paradox)**

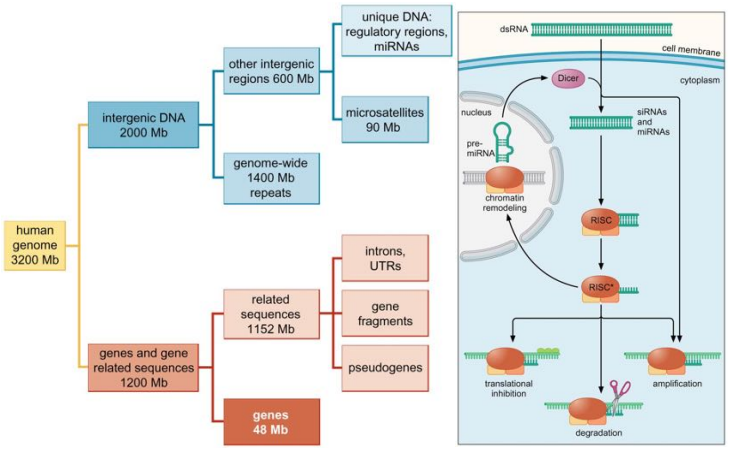
- *G-value paradox:* Disparity between the G-value (number of protein-coding genes) and relative complexity of an organism

G-value paradox explanation

- part of the explanation might be that some organisms make more use of **alternative splicing**
- **non-protein-coding sequences** more important than initially thought

Understand the functional importance of non- protein-coding DNA

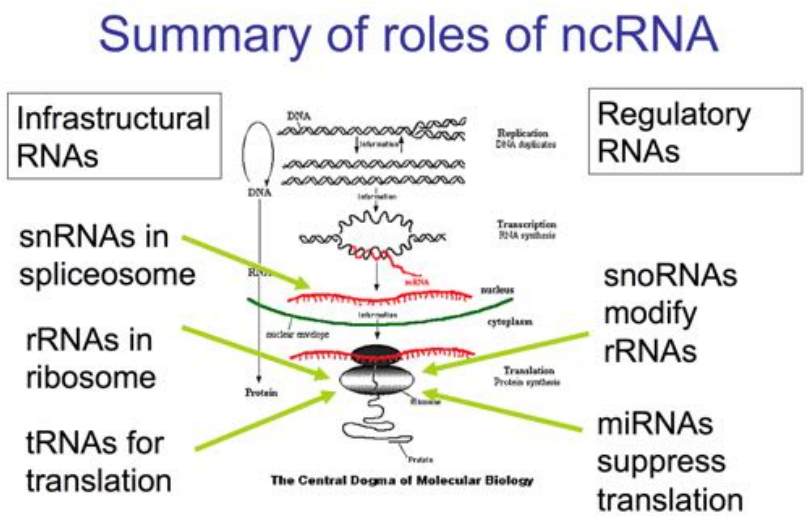
- a lot of non-coding DNA is transcribed = non-coding RNA



What is the function of the non protein-coding RNA transcripts?

- protein-coding sequences make up only 1.5% of genome
- BUT: **60-70% of genome is transcribed but not translated**

Non-coding RNA Transcripts	Function	Examples
Infrastructural RNAs	• Required for splicing and translation • Sequence-specific recognition of RNA substrates and catalytic processes	• rRNA • tRNA • small nuclear RNA
Small regulatory RNAs	• Affect translation of target mRNAs	• miRNAs

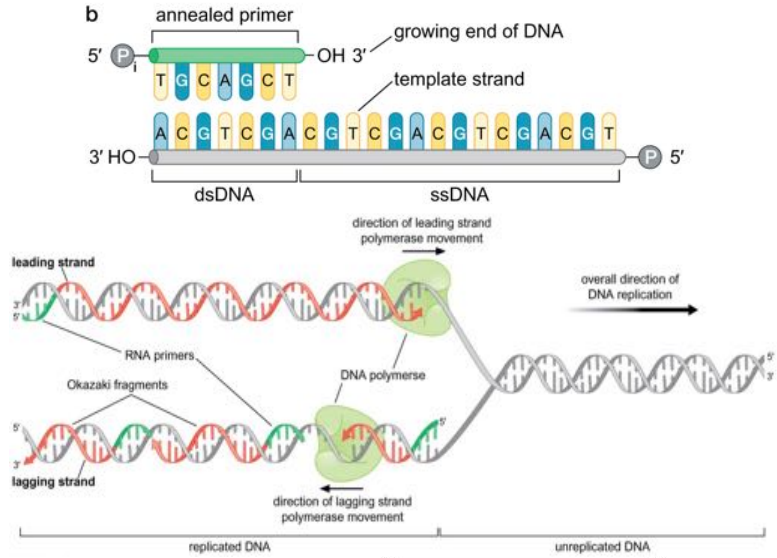


L4: Eukaryotic DNA replication

Review DNA synthesis

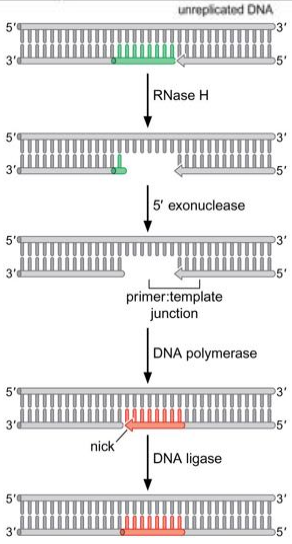
The Replication Fork

- RNA primase generates **primer + 3'OH** for DNA polymerase to begin replication
- Replication forks proceed bidirectionally from multiple origins of replication to replicate both strands of DNA
- **Leading strand** synthesised continuously as the replication fork moves
- **Lagging strand** synthesized discontinuously as rep. fork moves to create okazaki fragments → requires multiple primers

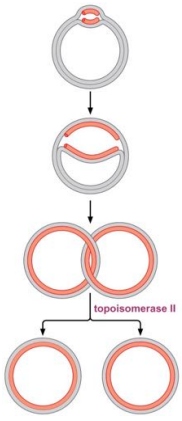


Removal of RNA primers from newly synthesised DNA

1. RNase H degrades the RNA primer
2. 5' exonuclease removes last RNA nucleotide
3. DNA polymerase then fills in gaps
4. DNA ligase makes last phosphodiester bond to join nucleotides together on lagging strand



Appreciate similarities and differences between eukaryotic and prokaryotic DNA replication



	Procaryotes	Eucaryotes
dsDNA	circular	linear
Origin of replication	1, oriC	Many + selective
DNA synthesis machinery @ Rep. fork	DNA pol	**DNA pol α/ primase** **DNA pol δ** and **DNA pol ε**
Single-stranded Binding protein	1 subunit	3 subunits
Completion of replication	Type II topoisomerases separates 2 catenated (linked) daughter DNA molecules	Telomerases elongate ends of chromosomes to prevent degredation