Organisation of Prokaryotic and Eukaryotic Genomes I

1. Introduction

An organism's genome holds the blueprint of all the information necessary for it to survive and reproduce. You have learnt previously in the topic of DNA and Genomics how genes are made of DNA, where a specific DNA sequence determines the function of a gene. You have also understood the central dogma of life involving the processes of transcription of DNA into RNA, and translation of mRNA into proteins. Each of these steps within the central dogma is controlled by specific sequence elements within the gene, making them functional and regulated in their expression.

Much of the gene structure of eukaryotes and prokaryotes is broadly similar. However, there are key differences in the genome organisation between eukaryotes and prokaryotes which reflect their divergent transcription and translation machinery. It is important to understand the structure/ organisation of both the prokaryotic and eukaryotic genomes as a foundation to understanding gene expression, regulation and function, as well as appreciate the difference between prokaryotic and eukaryotic genomes.

This lecture series encompass 3 parts. In Part 1, you will first learn the differences in the organisation of both prokaryotic and eukaryotic genomes. Part 2 explains how both genomes are controlled/ regulated, and Part 3 covers the topic of cancer.

2. Learning Outcomes

2 (d) Describe the structure and organisation of viral, prokaryotic and eukaryotic genomes (including DNA/RNA, single-/double-stranded, number of nucleotides, packing of DNA, linearity/circularity and presence/absence of introns) (Some parts covered in JC1 Biology).

2 (h) Describe the structure and function of non-coding DNA in **eukaryotes** (i.e. portions that do not encode protein or RNA, including introns, centromeres, telomeres, promoters, enhancers and silencers) (knowledge of transposons, satellite DNA, pseudo-genes and duplication of segments is not required)

3. References

Campbell, N.A. and Reece, J.B. (2008). Biology, 8th edition. Pearson.

4. Organisation of Lecture Content

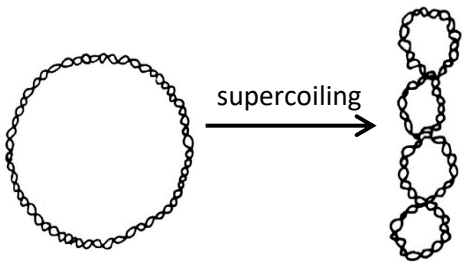
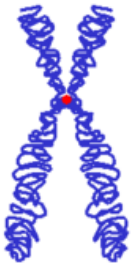
5. Comparison of the Structure and Organization of Prokaryotic and Eukaryotic Genomes

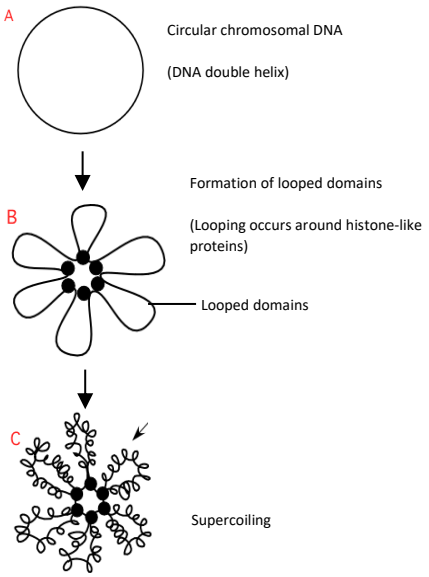
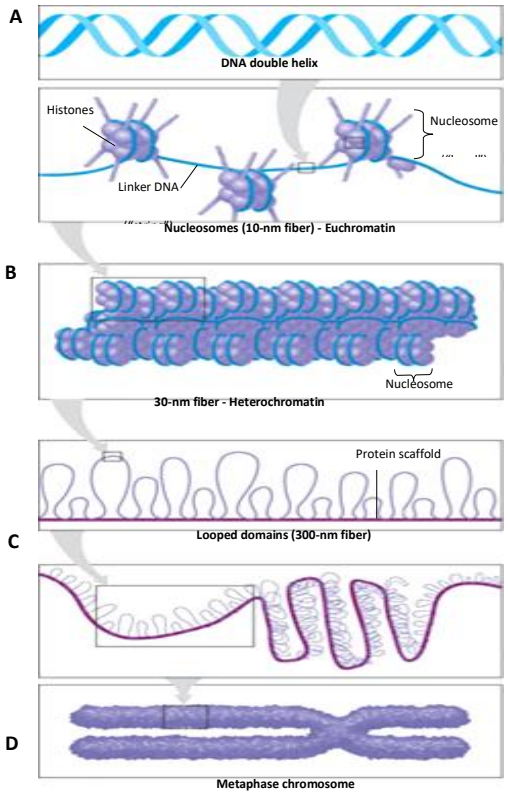
The term **Genome** refers to the complete set of genetic material in a particular cellular component.

The tables 1 and 2 below compare the key structures and organisation of both the prokaryotic and eukaryotic genomes.

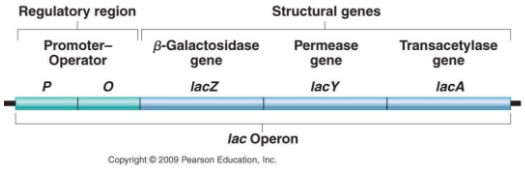
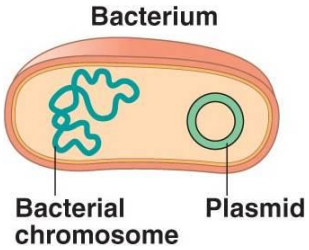
Recall that you have learnt some information about the Eukaryotic Genome in the topic of DNA and Genomics, as well as the Prokaryotic Genome in the topic of Genetics of Bacteria (covered under LO2d).

Table 1: Structure of Genome

Feature	Prokaryotic genome	Eukaryotic genome
Location in cell	**Nucleoid** region, non membrane-bound (prokaryotic cells lack nuclear envelope and nucleus)	**Nucleus**, surrounded by nuclear envelope
Size	10^5-10^7 base pairs	10^7-10^{11} base pairs
Number of genes	4,500	25,000
Molecule	Double Helix DNA	
Appearance	**One** chromosome, referred to as **monoploid** Generally **a single, circular** molecule 	**More than one** chromosome, usually in **diploid or higher ploidy levels** (two or more sets of chromosomes) **Multiple, linear** molecules 
Number of origins of replication (per chromosome)	**One**	**Multiple** Importance: To increase the efficiency of replication due to larger genome size
Presence of Telomeres	**No telomeres** in prokaryotic chromosomes as DNA is circular	**Telomeres present** at **both ends** of the linear chromosome

Presense of Centromeres	**No centromere** present	**Centromere present**
Association with proteins	Naked DNA, insignificant amounts	Yes – large amounts e.g. **histones (octamer)**, scaffold proteins
Level of DNA packing/coiling	Lower degree of condensation:  (A) Unfolded chromosome from *E. coli* has a diameter of 430µm. (B) DNA is folded into **chromosomal looped domains by protein-DNA associations**. Six domains are shown, but actual number is about 50. (C) **Supercoiling** cause further compaction, such that it fills an area of about 1 µm.	Higher degree of condensation:  (A) DNA double helix is associated with proteins called **histones**. DNA molecules are **negatively- charged**, histones are **positively-charged** (due to high proportion of positively-charged amino acids Lys and Arg). DNA thus is held around histones by **electrostatic interactions (ionic bonds)**. Most of DNA is wound around **octamers** of 8 histone proteins (**2 molecules of each histone: H2A, H2B, H2 and H4**) to form **nucleosomes**. The width of a nucleosome is approximately **10 nm** (thus also known as the 10nm fibre). Remainder of DNA, called **linker DNA**, joins adjacent nucleosomes.

		Note: The nucleosome is usually considered the **basic unit** of chromosomal packing. (B) The 10-nm fibre coils around itself to form a 30 nm **chromatin fiber** (or **solenoid**), with the help of a fifth type of histone, **histone H1.** (C) The 30-nm chromatin fibre forms loops called **looped domains (a 300-nm fibre)** when associated with **scaffold proteins** (non- histone proteins). (D) **Supercoiling** present. The loops further coil and fold to produce characteristic **metaphase chromosome** in a dividing cell.

Table 2: Organisation of Genome		
Feature	**Prokaryotic genome**	**Eukaryotic genome**
Location of Functionally-related Genes on Chromosomes	Genes that encode proteins for the **same metabolic pathway** are **grouped** together in a **single operon**. Eg. Lac Operon  	Genes that encode functionally related proteins are usually **located on different chromosomes**.
Non-coding regions (between and within genes)	Not common – typically less than 15%	Common – about 98%
Control by **Promoter**	Promoter present A **single promoter** controls the structural genes grouped in **an operon**.	Promoter present **Each gene** is under the control of its own **individual promoter** (as well as termination sequence and other control sequences).
Presence of **introns**	Typically **absent**	Introns **present** Introns are **interspersed between** exons
Presence of **enhancers or silencers**	Rarely present	Present
Extrachromosomal DNA (Plasmids)	May be present  	**Plasmids absent** (however mitochondria and chloroplast have their own DNA

6. Organization of Eukaryotic Genome

Notes to self

The genome of an organism comprises both **coding and non-coding regions** of DNA.

6.A Non-coding DNA Sequences

(a) **Non-coding regions of DNA** include any part of a genome that **does not code for proteins or RNA products** (i.e. rRNA, tRNA etc).

(b) Initially, non-coding DNA sequences were referred to as "junk DNA".

However, since we know that a **gene** is the **entire** nucleotide sequence **necessary** for the synthesis of a polypeptide chain (functional protein) or RNA product, there must be **important roles** (e.g. regulatory functions) that non-coding DNA sequences play in the eukaryotic cell. Also, non-coding DNA sequences have been preserved in genomes over hundreds of generations. This suggests that there might be significant roles that they play that have yet to be determined.

Hence, the term "junk DNA" is inaccurate and you should **avoid** using it. Instead, use the term "non-coding DNA".

(c) In eukaryotes, a **large proportion** of the genome is made up on **non-coding sequences**. In contrast, only a **small proportion** of the genome consists of **coding regions or genes**.

Figure 1 shows the distribution of coding and non-coding DNA sequences in humans.

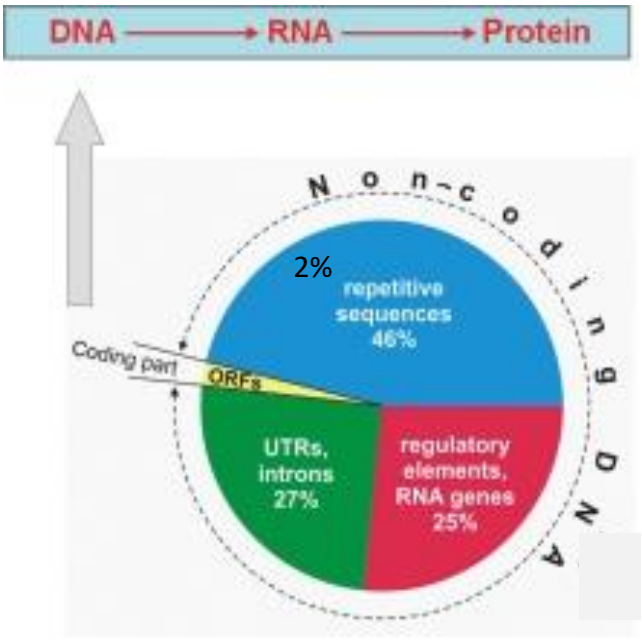


Figure 1. Pie chart showing the distribution of coding and non-coding DNA sequences in humans. Only 2% of the DNA sequences are coding and can be transcribed and translated into polypeptide chains, eventually fold into functional proteins.

Notes to self

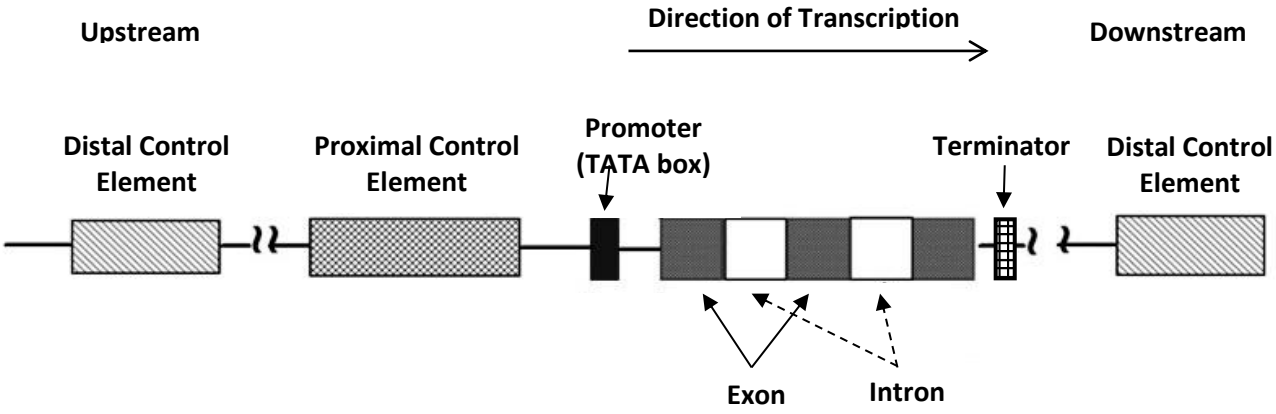
(d) Most of the non-coding sequences consists of **tandem repeated sequences** i.e. a short sequence of nucleotides that is repeated end-to-end multiple times within the genome.

E.g. **TCAT**TCAT**TCAT**TCAT**TCAT**

6.B Typical Features of the Eukaryotic Gene

(a) A gene is the **entire specific nucleotide sequence encoding the synthesis of a polypeptide chain (functional protein) or RNA product**.

(b) Figure 2 shows the typical features present in the Eukaryotic gene. Take note that there are various important non-coding sequences reflected in the figure. They play important roles in regulating the expression of the gene. These will be covered in the next segment.



CODING SEQUENCES

- **Exons**

NON-CODING SEQUENCES

- **Promoter** sequence
- **Terminator** sequence
- Proximal and Distal **Control Elements**
- **Introns**

Figure 2. Typical features of a eukaryotic gene

6.C Non-coding DNA Sequences with Important Functions

Notes to self

(a) Non-coding sequences comprise **promoter, introns, terminator and distal and proximal control elements (**also referred to as **enhancers** and **silencers**).

(b) Generally, control elements are **specific non-coding DNA sequences** that function to either:

- **Switch on and off transcription** (promoter and terminator respectively)
- **Increase and decrease the rate of transcription** (enhancer and silencer respectively).

Note terminology:

The direction in which transciption takes place is referred to as **downstream**. The opposite direction is referred to as **upstream** (refer to Figure 2).

6.Ci Promoters

Structure:

- Promoters are **specific non-coding DNA sequences** that are located **just upstream of the transcription start site of a gene.**
    - Many eukaryotic promoters contain the **sequence TATAAA**, also known as the **TATA box**.

Function:

- The promoter is the **site where transcription of a gene is initiated** as it serves as a **binding site for proteins** that include:
    - **RNA polymerase** that catalyses the transcription of the primary RNA transcript
    - **General/basal transcription factors** that are essential for **initiating transcription** of a gene at the **basal rate** and **recruiting RNA polymerase** to the promoter
- RNA polymerase, together with the general/basal transcription factors assemble at the promoter to form the **transcription initiation complex**.

Notes to self

- Certain **critical elements** or **short sequences** within the promoter **determine the strength of the promoter**. The strength of the promoter in turn determines the **binding efficiency of RNA polymerase** and hence the **frequency of transcription**.

- The greater the binding efficiency between the RNA polymerase and the promoter, the higher the frequency of transcription.

6.Cii Introns

Structure:

- Introns are **non-coding sequences of DNA interspersed between the coding regions** known as exons. Exons encode functional proteins or RNAs.
- Introns are transcribed together with exons to form the **primary mRNA transcript (pre-mRNA).**
- **Splice sites** are found at **both ends** of **each intron** and serve as **signals for RNA splicing**.
- The length and number of introns per gene vary widely between species, and within genes from the same species.

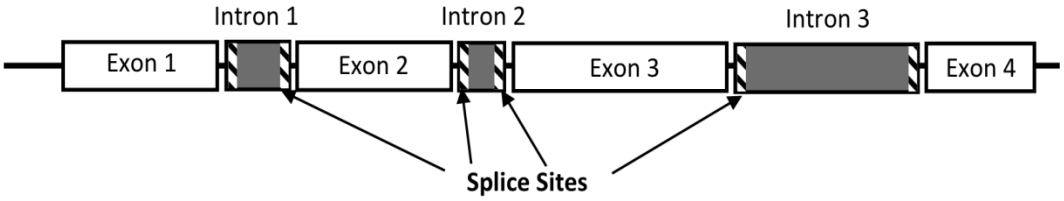


Figure 3. Introns (shaded) intersperse exons (unshaded). Splice sites (striped) flank introns

Function:

- Introns have **no involvement in the translation** of an mRNA. They thus need to be **excised via RNA splicing**.
- **RNA splicing** is a process where **introns are excised** (cut out) from the primary mRNA transcript and the **flanking exons joined together**, to form the **mature mRNA**.

Notes to self

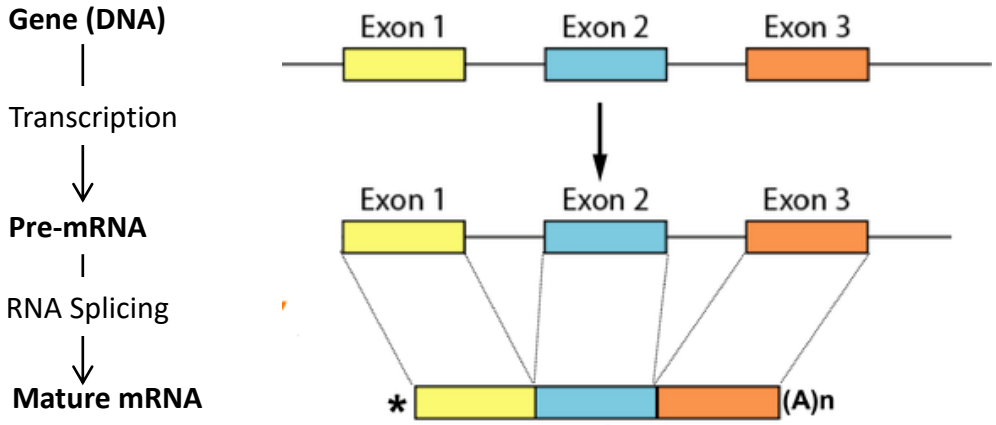


Figure 4. Outline of RNA Splicing. Both the introns and exons are transcribed to form the pre-mRNA. Following this, RNA Splicing, along with further processing processes, give rise to a mature mRNA with a continuous coding sequence (free from introns).

- The presence of introns within a gene **allows the particular gene to potentially encode several different polypeptides**.
    - The **different exons** in a gene **code for the different domains** of the encoded protein.
    - These domains each possess a **specific structure and hence function**.
    - For example, a membrane-bound receptor may contain two domains, one, which forms the binding site of the receptor, and a second, which allows the receptor to bind to the cell surface membrane.
- During the post-transcriptional modification of the primary RNA transcript, introns are cut out from the RNA transcript and the exons are joined together in a process known as **RNA splicing** *(details of this process will be covered later under Control of Prokaryotic and Eukaryotic Genome)*.

- Under different cellular conditions, **a single pre-mRNA can produce different mature mRNA**, depending on which **combinations of exons** are spliced together. This means that **one gene can code for more than one type of polypeptide**. This is known as **alternative RNA splicing**.

- An **advantage** of alternative splicing is that it **enables a larger number of proteins to be produced relative to the number of genes present**. Alternative splicing produces **different protein isoforms from one gene**. Isoforms are alternative forms of the same protein. This is illustrated in Figure 5 below.

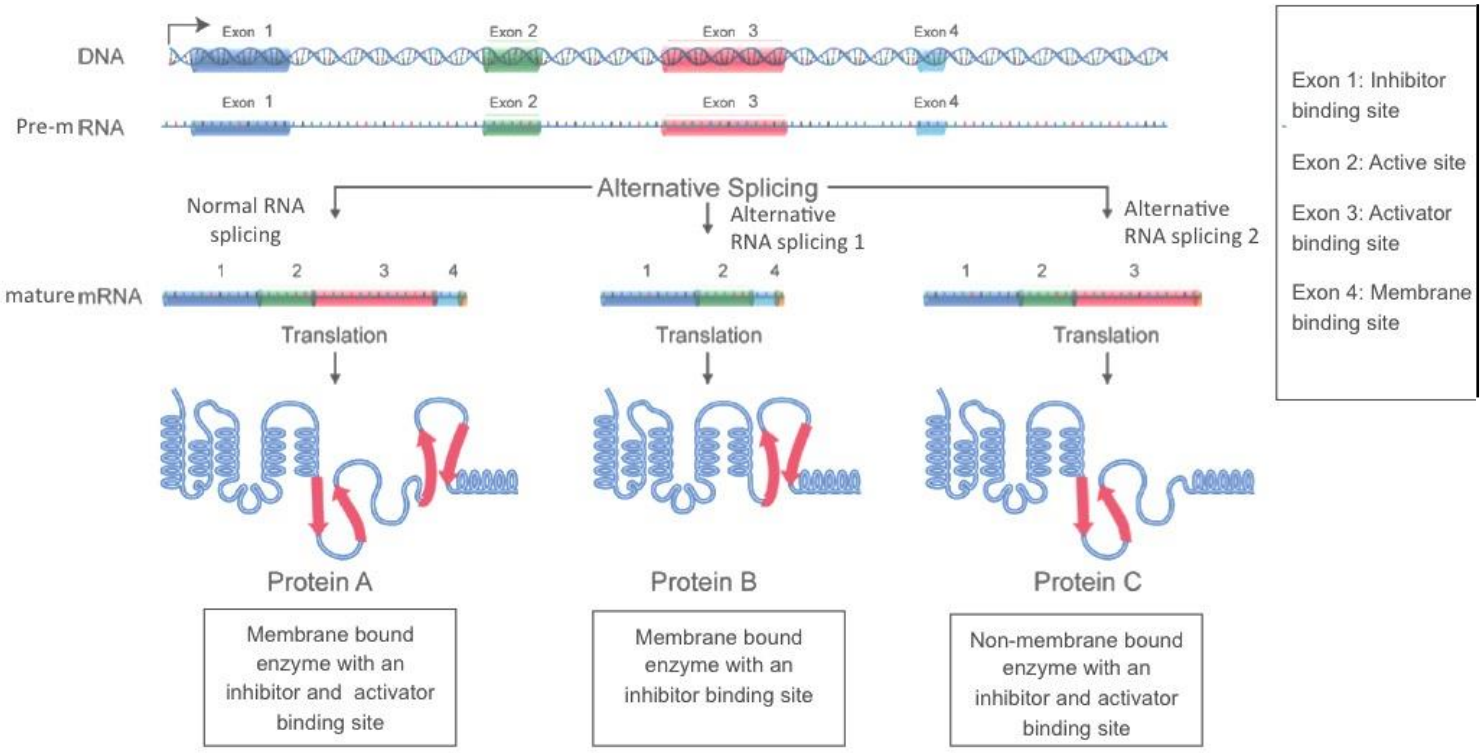


Figure 5. Diagram showing how alternative splicing may generate various protein isoforms from a single mRNA.

Distribution:

- Introns are found **only in eukaryotes.**

Notes to self

6.Ciii Terminators

Structure:

- A terminator is a **specific sequence of non-coding DNA that signals the stop of transcription**.

Notes to self

Function:

- The **transcribed terminator sequence** on the **RNA product** serves as a termination signal, causing RNA polymerase to release the pre-mRNA and detach from the DNA template.

6.Civ Enhancers and Silencers

- **Enhancers** and **Silencers** are short **non-coding DNA sequences** which are **usually located far away** (up to thousands of nucleotides upstream or downstream) from the promoter. They can also be found within an intron or near the gene to be controlled.
- They are regulatory or control elements.

1. **Enhancers**

- Are *regions of **DNA*** which are bound by **specific transcription factors** (proteins) called **activators**.
- When activators bind, they **promote the assembly of the transcription initiation complex**.
- This leads to an **increased rate of gene transcription**.

2. **Silencers**

- Are *regions of **DNA*** which are bound by **specific transcription factors** (proteins) called **repressors**.
- When repressors bind, they **inhibit the assembly of the transcription initiation complex**.
- This leads to a **decreased rate of gene transcription.**

- Note: The function of control elements will be explored in detail in Part 2 of this lecture series (Control of Prokaryotic and Eukaryotic Genome).

Notes to self

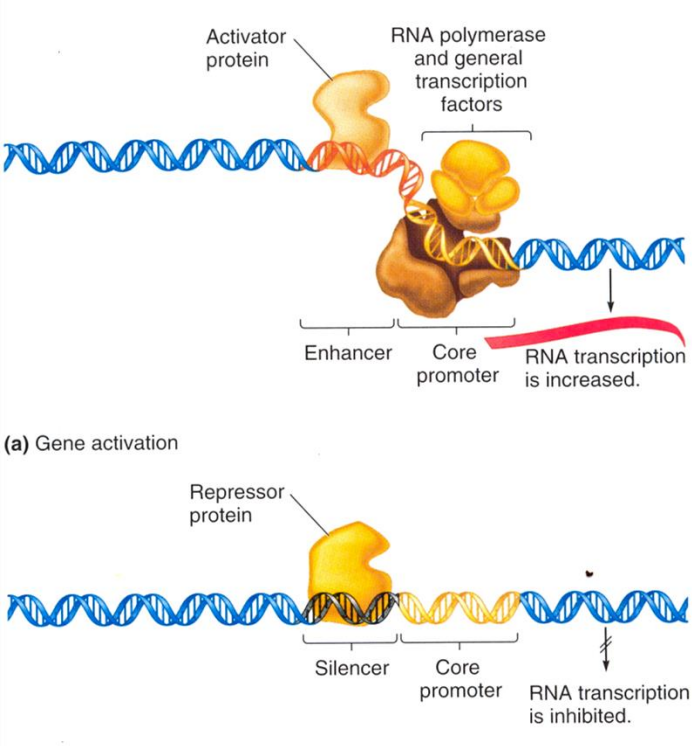


Figure 6. Diagram showing an overview of the transcriptional regulation at (a) the enhancer when activator protein binds, as well as at (b) the silencer, when repressor protein binds.

6.4 Important Features of the Eukaryotic Chromosome made up on non-coding DNA

- There are three essential features of all eukaryotic chromosomes in order for them to be functional, maintain their structural integrity, and for them to be passed on to subsequent generations:
  - Origins of replication
  - Centromeres
  - Telomeres

- Both centromeres and telomeres contain non-coding, tandem repetitive DNA sequences.

6.4i Centromeres

Notes to self

Structure:

- Centromeres typically consist of **large arrays of non-coding repetitive DNA** where the sequence within individual **tandem** repeat-unit is **similar but not identical**.
- Centromeres can **vary in sequence**, and centromeric DNA normally exists in a **heterochromatin state**.
- They are identified during the **metaphase** stage as v**isible constricted regions on chromosomes** where spindle fibres attach to during nuclear division.
- Centromeres can be located anywhere along the length of a chromosome.
- Each sister chromatid has its own centromeric DNA sequences.

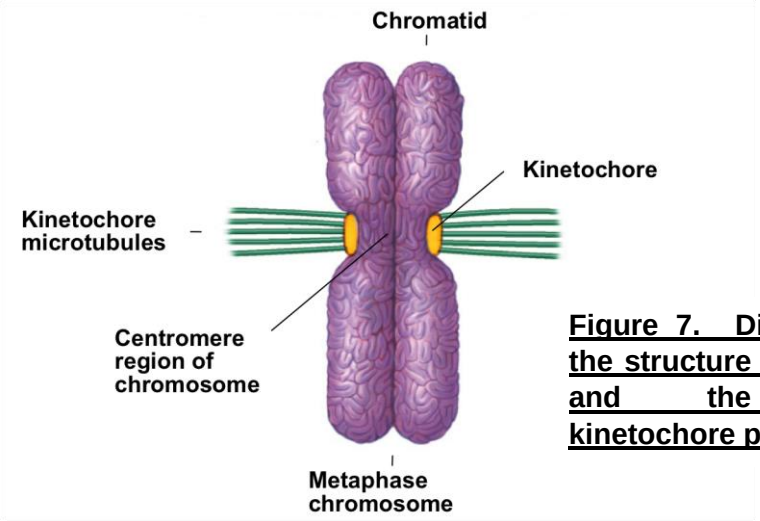


Figure 7. Diagram showing the structure of a centromere and the associated kinetochore proteins.

Functions:

- They facilitate chromatin organisation by enabling **sister chromatids adhesion** during mitosis (prophase and metaphase).
- Being the **site of kinetochore assembly**, centromeres allow the **attachment of spindle fibres** at **kinetochore found at centromere** of sister chromatids.
    - Spindle fibres attached to kinetochore proteins are required to pull chromosomes to opposite poles during anaphase of mitosis and anaphase I and II of meiosis.

- This in turn allows **equal separation of chromatids** to **opposite poles** of cell during **anaphase of mitosis** or **anaphase II of meiosis**, when **centromeres divide.**

Notes to self

Note: Without centromeres, improper chromosomal alignment and segregation will result in **aneuploidy** e.g. Trisomy 21/ Down's Syndrome.

Distribution:

- Centromeres are found **only in eukaryotes** as they are found in **linear chromosomes only**.

6.4ii Telomeres

Structure:

- **Telomeres** are nucleotide sequences found at **both ends** of **linear eukaryotic chromosomes**.
- Telomeres are **non-coding** regions of DNA made up of a series of short **tandem repeat sequences**.
- Telomeres have a **single stranded region of DNA** at their **3'** ends known as the **3' overhang**. This region of DNA does **not** have a complementary strand.
    - The 3' single-stranded end **loops back** and displaces the same sequence in the upstream region of the telomere and binds to the complementary sequence of the other strand. This process is brought about by special telomere-binding proteins.

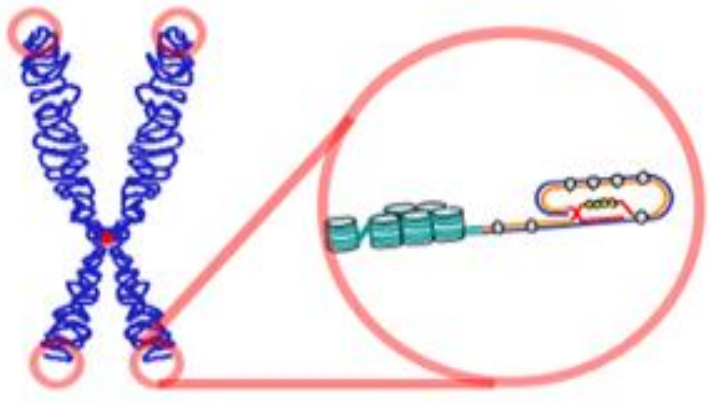


Figure. 8. A T-loop that later associates with nearby double-stranded telomeric DNA to form a displacement or D-loop. The D-loop is effectively a "knot" at the telomeric end that maintains chromosomal stability.

Function:

Notes to self

- **Role 1:**

Telomeres ensure genes are not lost/eroded with each round of DNA replication due to the end replication problem. This prevents loss of vital genetic information.

- What is the **end-replication problem**? (Refer to Figure 9)

  - The end replication problem occurs during the replication of **linear** eukaryotic chromosomes.
  - During DNA replication, **DNA polymerase requires a free 3'OH** of a pre-existing strand to **add free nucleotides**.
  - An **RNA primer** is synthesised to provide a free 3'OH end for the addition of free nucleotides.

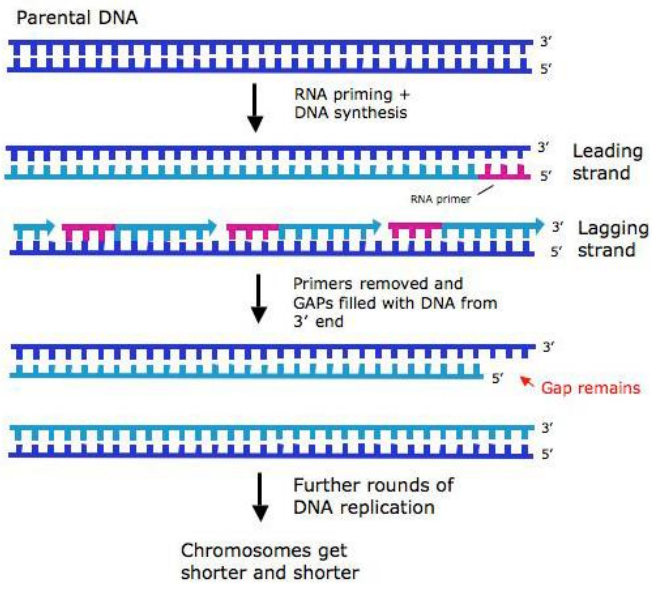


Figure 9: End replication problem

  - However, there is **no exisiting 3'OH group** after the **RNA primer is removed and cannot be replaced** with nucleotides at the **5' end of newly synthesized daughter strand**.
  - Thus complete replication of the 5' ends of daugther strands cannot occur, and as a result, a **3' overhang** is created at the end of chromosome.
  - After repeated rounds of replication and cell division, as the chromosome **shortens**, essential genes would be **eroded** from the ends of chromosomes.

- Due to the end-replication problem, the ends of chromosomes shorten with every round of DNA replication.
- Since telomeres are non-coding, **shortening of chromosomal ends leads to shortening of the telomeres** without any deleterious effects.
- The genes within the chromosome will thus **not be eroded** with each round to DNA replication**, preventing loss of vital genetic information**.
  - Eventually, telomeres in cells which have divided many times tend to be shorter.
  - Note: The shortening of telomeres also **indirectly prevents the**

development of cancer. This is because cells may undergo **apoptosis (programmed cell death)** when their telomeres have shortened to a **critical length**. Apoptosis thus **prevents accumulation of mutations** (e.g. eroded genes and unstable DNA), thus preventing cancer development.

Notes to self

- **Role 2:**
  Telomeres protect and stabilise the terminal ends of chromosomes

- The **single stranded 3' overhang** at the ends of linear chromosomes poses some problems. Without telomeres, chromosome ends r**esemble broken chromosomes**, and lead to the following:

  (i) The single stranded 3' single-stranded overhang of one terminal end of a chromosome may **anneal to a complementary single-stranded region of the terminal end of another chromosome**. This causes joining of different chromosomes.

  (ii) Such ends are **similar to DNA damage** formed due to double stranded breaks, and **sends signals to trigger cell arrest and cell death (apoptosis)**.

- Thus by forming the T loop (Figure 8), the telomeres **stabilize the ends** of the chromosomes by

  - **Preventing them from fusing with other chromosomes**, and
  - Prevent DNA repair machinery from recognizing the ends of chromosomes as DNA breaks, **protecting the chromosome**, hence **preventing apoptosis**.

- **Role 3:**
  Telomeres allow their own extension, by providing an attachment point for the correct positioning of the enzyme telomerase.

- In cells such as **stem cells and immune cells** that need to **divide repeatedly, an enzyme telomerase is expressed to maintain telomere length.**
- Telomerase catalyses the **lengthening of telomeres**.
- Hence, in cells that express telomerase, the end-replication problem is **overcomed as telomeres can be lengthened** with each round of division.
  Note: End replication is **NOT prevented**.

Telomerase

Notes to self

- Telomerase is a **ribonucleoprotein** (a complex of RNA and proteins) that functions as a **reverse transcriptase** (a protein that synthesises DNA using an RNA template).
- In humans, telomerase contains a **single RNA molecule, which provides the template sequence -AAUCCC-** to guide the synthesis and insertion of the telomeric DNA sequence -TTAGGG-.
- **A** protein component called **hTERT** (**human telomere reverse transcriptase**) provides the **catalytic action** of telomerase in humans.

The action of telomerase is summarised on the next page (Figure 10).

Catalytic Action of Telomerase

Notes to self

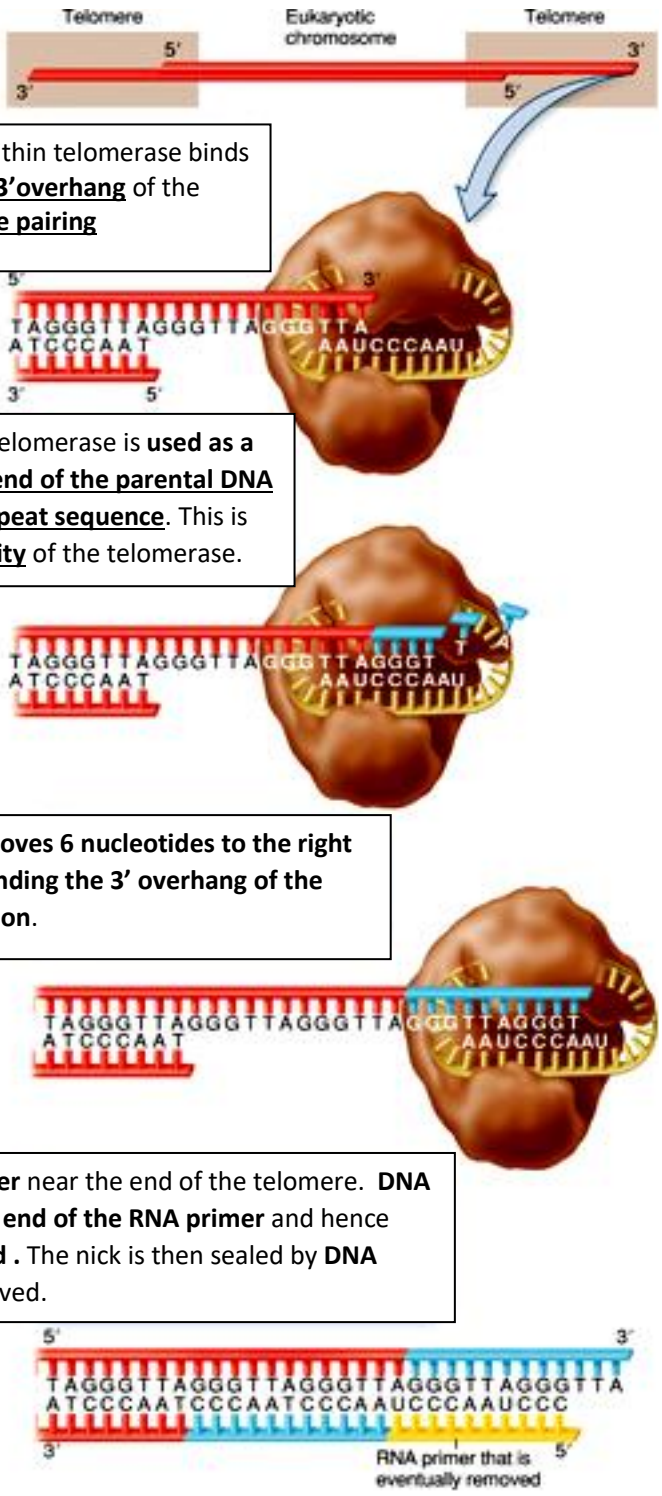


Figure 10: Action of Telomerase

Distribution:

- Telomeres are found **only in eukaryotes,** at both ends of **linear chromosomes**.