Topic 1 – MOLECULAR ANATOMY OF CHROMOSOMES, GENES, AND GENOMES

Historical perspectives:

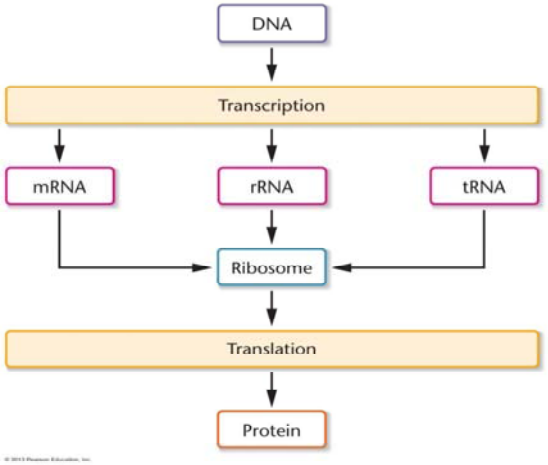
- In 1868: Friedrich Meischer isolated nuclei, determined that DNA is acidic
- 1910: Pheobus Levene found that DNA is made out of 4 nucleotides
- 1940: Erwin Chargaff supported Leven's findings, discovered that C=G, A=T
- 1944: Oswald Avery and Colin Macleod found that DNA is the molecule of heredity (not proteins)
- 1953: Watson and Crick discovered the DNA double helix structure

Genetics as an investigative tool

- Differentiating between individuals -> look at the genes found in DNA.
- The **central dogma** of genetics is fundamental to defining this difference:
    - DNA is transcribed into different forms of RNA (mRNA, tRNA and rRNA)
    - mRNA is translated with the help of tRNA and rRNA into proteins
    - Proteins cause phenotypes in organisms

Characteristics of genetic material:

- **Storage** of information:
    - Genes coding for proteins
- **Replication**:
    - The ability to pass on information
- **Expression** of information:
    - Translation to proteins
- **Variation** by mutation:
    - Necessary for adaptation

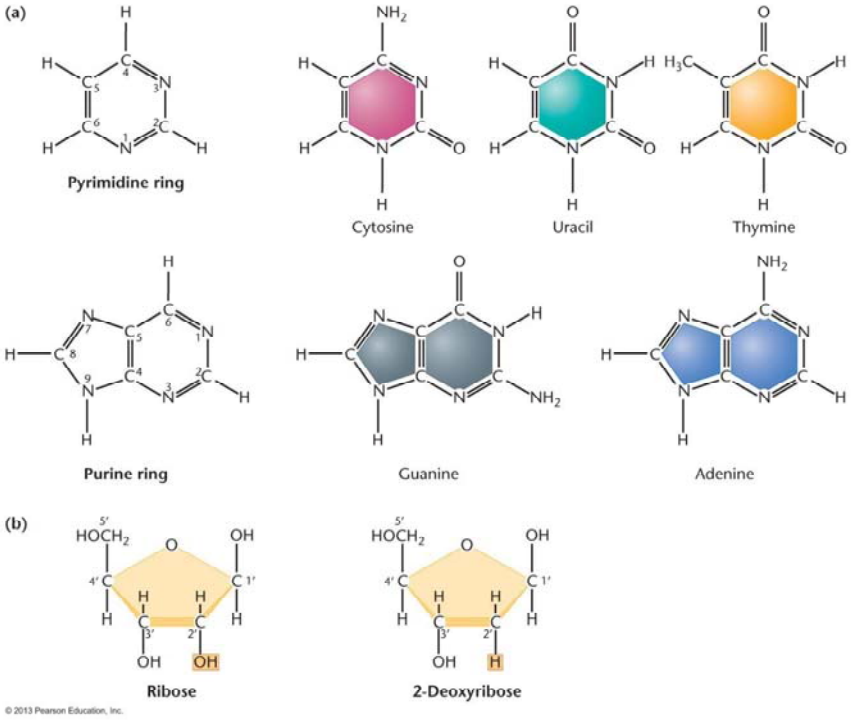


Forwards and reverse genetics are two ways used to identify the function of certain genes:

- **Forward genetics:**
    - What is the genetic basis of this phenotype/trait?
    - Phenotype is observed
    - Mutations are induced to change this phenotype, and the gene responsible for this phenotype change can be identified by tracking the mutation.
    - Phenotype → Induce mutation → DNA sequence changes, gene responsible is found
- **Reverse genetics:**
    - Which phenotype results from this particular gene?
    - A gene is observed
    - Mutations are induced to change it. If there is a change in a certain phenotype, we know that the gene that we mutated plays a role in the phenotype
    - Putative gene → Induce mutation → Gene function

Nucleotides are the building blocks of nucleic acids:

- **Purine** =nucleotide with a double ring (Adenine and Guanine)
- **Pyrimidine** = single ring (Cytosine, Uracil and Thymine)
- **Ribose** is found in RNA, and **Deoxyribose** is found in DNA

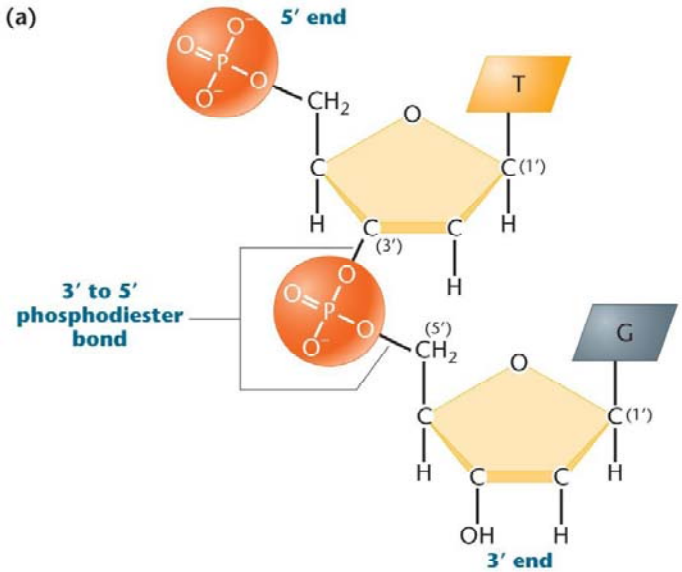


Ribose: **OH** on C2
Deoxyribose: **H** on C2

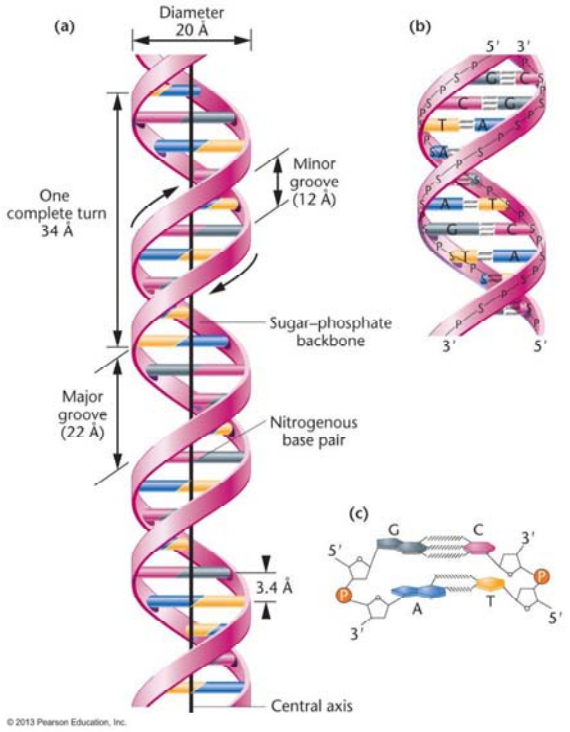
- **Nucleoside** = base + ribose sugar
- **Nucleotide** = base + ribose sugar + phosphate group (Nucleoside + phosphate group)

- DNA and RNA are polynucleotides, with the phosphate group already attached to the chain added to the 5' carbon of the added sugar, and the new phosphate group added to the 3' end carbon on the sugar.

- Huge amount of variation possible: a 1000 nucleotide chain has 4^{1000} possible arrangements!



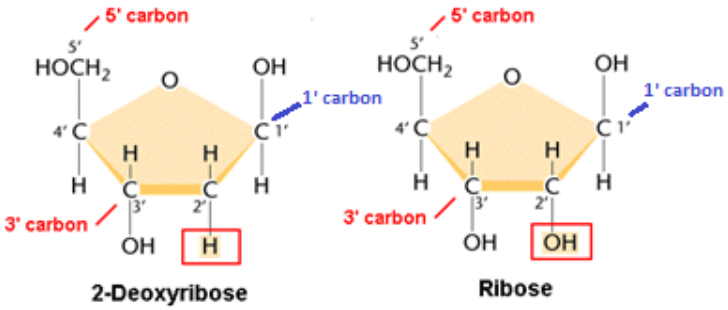
DNA structure: Double Helix Model:



- There are 2 coiled polynucleotide chains, around a central axis. It is a **right handed double helix**
- The chains are **anti-parallel** (One is 3'-5' and the other is 5'-3')
- The nucleotide base pairings are **A-T** and **C-G**
- **Each turn** of the helix is **10 base pairs** long, or **34 Angstrom** (A)
- Base pairs **inside the helix** are perpendicular to the central axis and are **3.4 angstrom apart** from one another (Makes sense, 10 of these measure 34 A)
- Nucleotides are bound via **hydrogen bonds**. Hydrogen bonds are weak, but the sheer number of bonds make for strong overall association
- C-G are **triple bonded** while A-T are **double bounded**, causing the denaturation of C-G bonds to require more heat (higher melting point)
- **Other pairings may occur**, because resonance in the base pairs may occur and cause more/less H-bonding sites – this is one way that we can induce **mutations**!

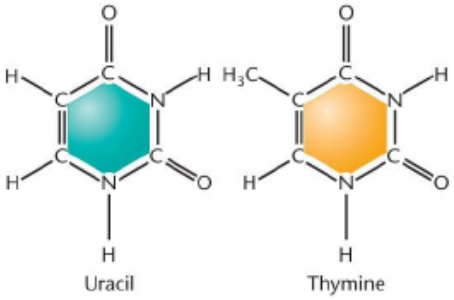
RNA Structure is Chemically Similar to DNA, but Single Stranded

- **Ribose** replaces deoxyribose in **RNA**
- Ribose is a **five membered ring**, sugar molecule
- In ribose, a hydroxyl **(-OH**) molecule is attached to the **2'C**
    - In **deoxyribose**, a **hydrogen** is attached to the 2'C instead of the OH



- **Phosphate groups** attach to the **5'C** and **3'C** of the ribose sugar
- The **nitrogenous bases** attach to the **1'C** of the ribose sugar
- Nitrogenous bases in RNA are the same as those DNA, with one exception
    - **thymine** is replaced by **uracil**

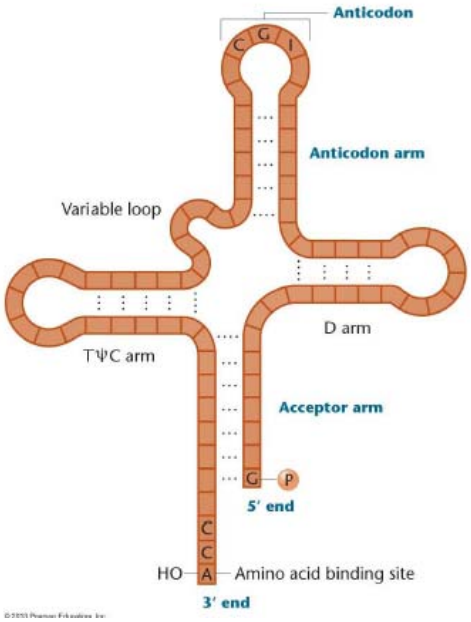
- Thymine is not present in RNA because the **enzymes that translate the mRNA cannot read the thymine**
- Thymine and uracil only differ by the presence of a **methyl group** on the **5'C** of **thymine**



- Base pairing: **A – U** and **G – C**
- Examples of RNA molecules:
  - **Messenger RNA (mRNA)**
    - Created during transcription and leaves the nucleus once transcribed
    - Used as a template during translation – required to make polypeptides
  - **Transfer RNA (tRNA)**
    - Used during translation – brings the required amino acid to the matching codon in the mRNA sequence
  - **Ribosomal RNA (rRNA)**
    - Major part of the ribosomes involved in translation
  - snRNA – small nuclear RNA
  - siRNA – small interfering RNA (bind mRNA and inhibit expression)
  - miRNA – micro RNA
    - snRNA, siRNA, miRNA are involved in gene expression

- All RNA are created from one of the two strands of DNA
  - This occurs during transcription (DNA is **read 3'-5'**, RNA is **created 5'-3'**)

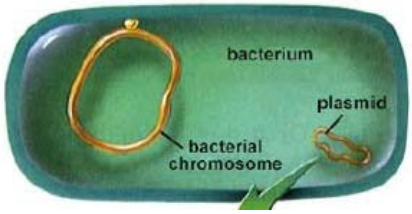
- **Transfer RNA (tRNA):**
  - **Recruits amino acids** to ribosomes during translation for the synthesis of proteins
  - **Single stranded RNA molecule**
  - **Base pairing** exists within the molecule
    - Has **double stranded regions**
    - Pairing gives the molecule a **conformation**
  - the **3' end binds the amino acid**
  - the anticodon binds to the mRNA
  - the **anticodon matches the codon** in the mRNA strand
    - the codon specifies the specific amino acid the tRNA will recruit
    - each tRNA can only recruit for its specific AA (anticodon is constant)



- **Ribosomal RNA (rRNA):**
  - **Most RNA** exists as rRNA (~80%)
  - **Part of the ribosome** that is involved in the translation of mRNA
  - Responsible for **peptidyl transferase** – forming peptide links between adjacent AA's using tRNA's during translation
  - Many different forms – differ between prokaryote and eukaryote organisms

Variations in the Organization of DNA

- **Viral**
  - **Variable genetic material** – RNA or DNA, single stranded or double stranded
  - Often linear in shape, but can be circular
- **Bacterial**
  - Main genome - **circular, double stranded DNA**
  - Many bacteria also contain circular **plasmids**
    - Often acquired from the environment/other bacteria
    - Can be replicated and passed on
    - Can integrate into the main genome or pick up new DNA
    - Location of lots of useful adaptive genes (drug resistance, etc…)

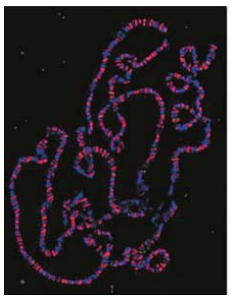


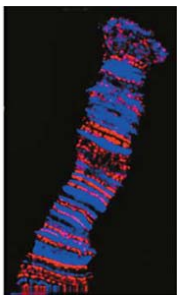
Eukaryote DNA

- Eukaryotic cells hold **linear, double stranded DNA** and the full genome is most likely spread across multiple **chromosomes**
- **Banding patterns are specific** to each chromosome – individual bands are often called **chromomeres**

Special Cases of Eukaryotic DNA

- **Polytene Chromosomes**
  - Polytene chromosomes **continue to replicate**, but **never separate**
  - Gene **expression can be very high** (multiple copies of the same gene)
    - **"Puffs"** -> Uncoiled DNA that is being transcribed at very high levels
    - **Centromere** keeps the chromosomes together

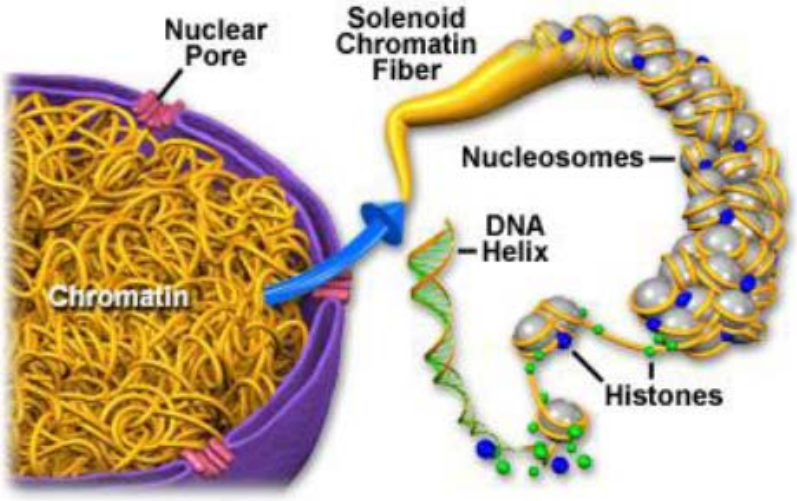




- **Lampbrush Chromosomes**
  - Found in **vertebrate oocytes**
  - DNA becomes **uncoiled during meiosis**, when **high levels of expression** are needed for metabolic purposes associated with meiosis

Typical Organization of Eukaryote DNA

- **Chromatin** form:
  - loose arrangement of DNA
  - structures such as **histones** and nucleosomes are important in condensing and compacting the chromatin
  - relaxed state of DNA – found during **interphase**
  - **Packing**: histones → nucleosomes → solenoids → chromatin → chromosome
  - the chromatin is **condensed during cell division**
  - the **telomeres** and **centromere** region of the chromosomes consist of **repeated DNA** sequences (heterochromatin)
    - **protection** so that the DNA is not harmed during cell division
  - Remember – chromosomes in the "X" form are already duplicated! In this form, there are two copies of the same chromosome (before crossing)

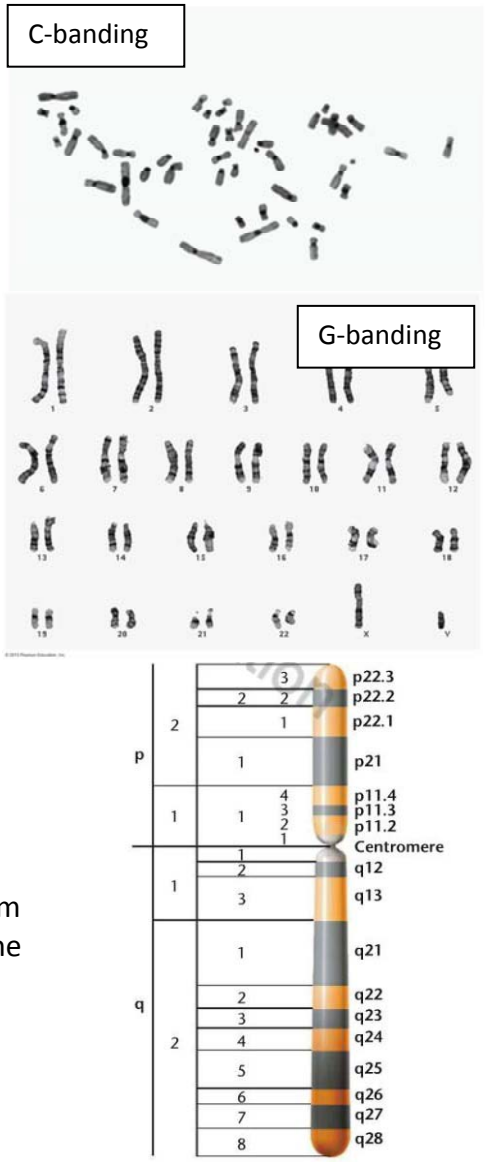


Euchromatin versus Heterochromatin

- **Euchromatin** – chromatin containing genes to be expressed
  - Loosely packed chromatin – easily accessible for transcription
- **Heterochromatin** – inactive or inaccessible regions of the DNA
  - Genes not being expressed
  - Highly compact areas of chromatin – telomere and centromere
- **Barr body** – inactive X chromosome
  - Has a tendency to stick to the nucleus wall
  - Bar bodies are easily stained because they are condensed areas of chromatin
  - The number of Barr bodies in each cell is = **nX-1**, where n is = # X chromosomes
    - Normal male has 0 per cell, normal female has 1 per cell

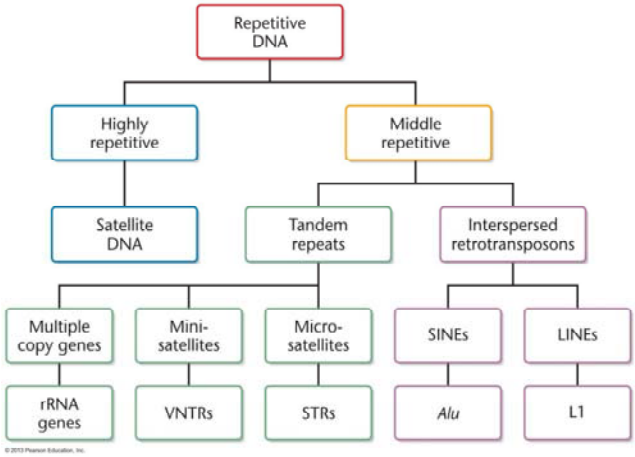
Molecular Anatomy of Chromosomes

- **C-banding**:
  - chromosomes are isolated from the nuclei and treated with **Giemsa** staining
  - The Giemsa stain only stains the **centromeres**
  - This allows for the simple counting of individual chromosomes
- **G-banding**:
  - DNA samples are treated with detergents and digestive enzymes – gets rid of bound proteins
  - The **Giemsa** stain stains certain bands
  - Pairs of **homologous** chromosomes have the same banding pattern.
  - The same genes are in the same order on the homologous chromosomes, but are not necessarily the same alleles.
  - Used when studying modification and/or translocation of a gene.
- **Individual chromosomes:**
  - **Long** arm is called **q arm**
  - **Short** arm is called **p arm**
  - Specific banding patterns allow for a uniform nomenclature to identify specific sites on the chromosome.



Organization of DNA Sequences

- **Repetitive DNA:**
  - Allows for differentiation between species and individuals within the species via the makeup of **markers**
  - Markers can be tracked to monitor crossing between chromosomes
- **Non-coding DNA regions:**
  - DNA in the genome that is there, but does not code for anything
  - Mutations in these regions do not affect the cell, and can be used as **genetic markers**



Highly Repetitive DNA

- **Satellite DNA**
    - Large arrays of **tandemly repeating**, **non-coding** DNA sequences
    - Main component of **centromeres** and **heterochromatin**
    - Prokaryotes do not have satellite DNA
    - The tandem repeats result in **different density** from bulk DNA, and this density difference can be used to separate the satellite DNA from the bulk DNA
        - the amount of satellite DNA in a sample can be easily measured!

Middle Repetitive DNA

- **Tandem Repeats –** repetition of nucleotide sequences side-by-side
    - **VNTRs**
        - Variable Numbers of Tandem Repeats
        - within and between genes – anywhere in the genome
        - many clusters of 15-100 nucleotides
        - can be genetic markers for individualization (different number of repeats at each location in each individual)
        - also called minisatellites
    - **STRs**
        - Short Tandem Repeats
        - within and between genes – anywhere in the genome
        - di-, tri-, tetra-, or pentanucleotides (< 5 base pairs being repeated)
        - can be genetic markers for individualization (different number of repeats at each location in each individual)
        - Found in telomeric DNA sequences (protects chromosome from degradation)
        - Also called microsatellites

Repetitive Transposed Sequences

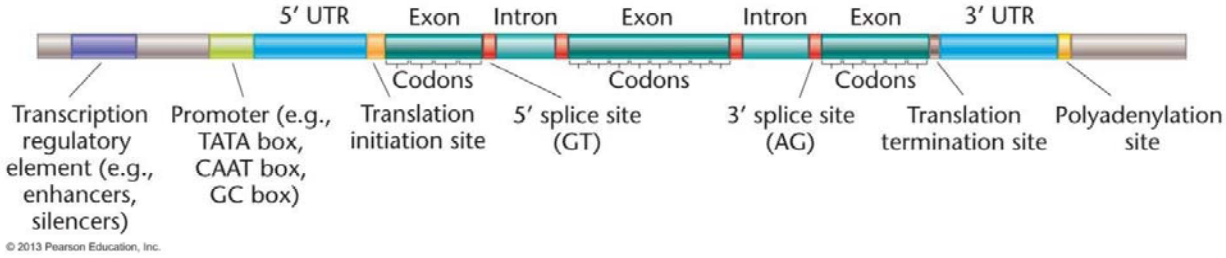
- **Transposable** -> can move from one part of the chromatin to another
- Integrated viral DNA in our cells
- Sequences that are **interspersed individually** throughout the genome (not tandemly repeated)
- Can be used as genetic markers (individualism)

- **SINEs**
    - Short INterspersed Elements
    - **Less than 500 base pairs** long and may be present 500,000 times in the human genome (~13%)
    - **Alu family** – presence of DNA sequences recognized by the restriction endonuclease Alu I
    - Does **NOT** recruit reverse transcriptase; relies on other methods of transposition
    - "Junk DNA"

- **LINEs**
    - Long INnterspersed Elements
    - **~6000 base pairs** long and may be present 800,000 times in the human genome (21%)
    - **L1 family –** LINEs are produced via **retrotransposition**
        - Entire sequence is transcribed into RNA – part of the RNA is translated into a **reverse transcriptase enzyme**
        - The reverse transcriptase enzyme uses the other RNA piece as a template and produces DNA which then integrates at a new position in the genome

Organization of Gene Sequences

- Keep in mind that mutations can happen anywhere in the genome!
    - Point mutations in the **introns** shouldn't matter (they are spliced out)
    - A mutation in an **exon** may change a codon -> changes a AA in the protein
    - A mutation in the **regulatory portion** may increase, decrease, or have no effect on the rate of transcription – it depends on which part of the sequence changes
    - A mutation in the **promoter** may inhibit the ability of polymerase to bind – it may permanently silence the gene (or not)
    - Mutations in the **translation signaling sites** may mess up the protein that is produced (formation/destruction of a start/stop codon)
    - **The take home message** – mutations can change everything or change nothing. It depends on where the mutation happens and what is changed!



- **Regulatory Elements**
    - Control gene expression
        - Enhancers can bind and help recruit polymerase
        - Silencers can block the binding of polymerase
- **Promoter**
    - Includes the TATA box, the main site at which the polymerase enzyme binds
- **Translation Initiation/Termination Sites**
    - Includes the START/STOP codons – the beginning/end of the polypeptide chain
- **Introns/Exons**
    - Exons are coding regions that will stay
    - Introns are spliced out
- **Polyadenylation site**
    - Poly-A tail will be added at this site for protection as the RNA migrates