Topic 1 – Molecular Anatomy of Chromosomes, Genes, and Genomes

Historical perspectives:

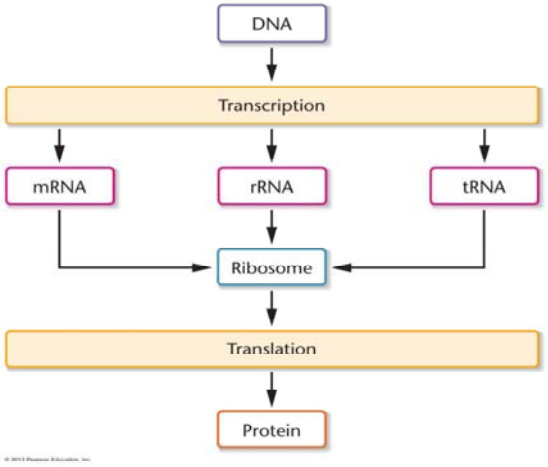
- In 1868: Friedrich Meischer isolated nuclei, determined that DNA is acidic
- 1910: Pheobus Levene found that DNA is made out of 4 nucleotides
- 1940: Erwin Chargaff supported Leven's findings, discovered that C=G, A=T
- 1944: Oswald Avery and Colin Macleod found that DNA is the molecule of heredity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

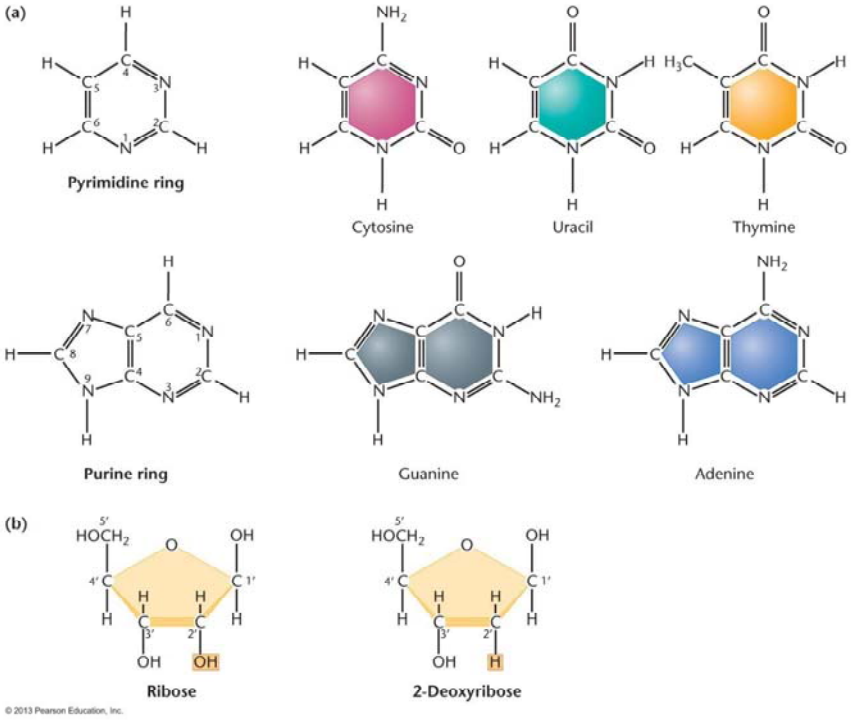


Forwards and reverse genetics are two ways used to identify the function of certain genes:

- **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

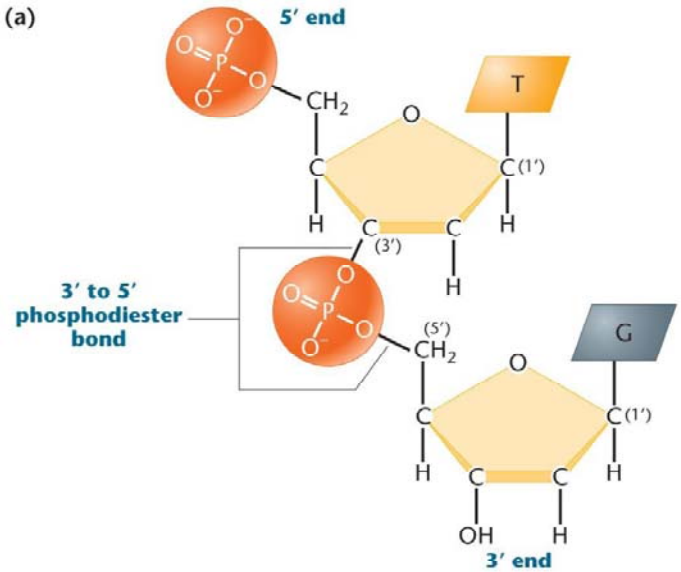


Ribose: **OH** on C2
Deoxyribose: **H** on C2

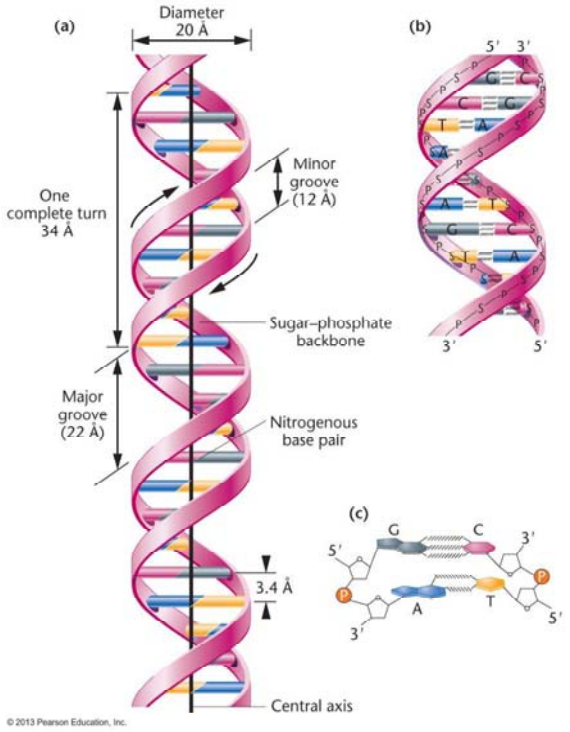
- **Nucleoside** = base + ribose sugar
- **Nucleotide** = base + ribose sugar + phosphate group (Nucleoside + phosphate group)

- 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!



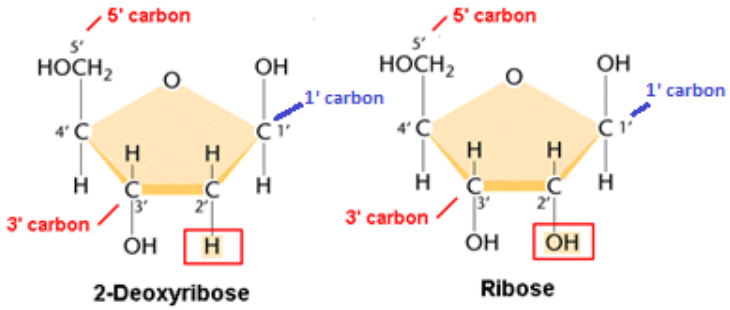
DNA structure: Double Helix Model:



- 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**!

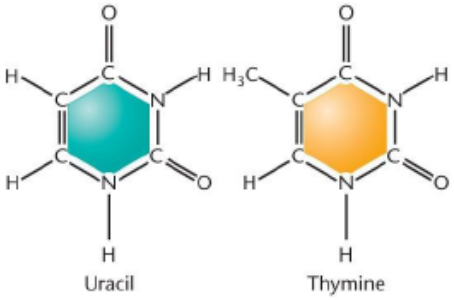
RNA Structure is Chemically Similar to DNA, but Single Stranded

- **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**

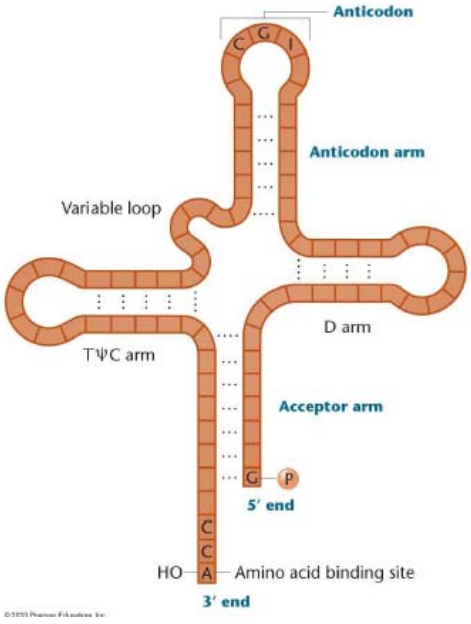
- Thymine is not present in RNA because the **enzymes that translate the mRNA cannot read the thymine**
- Thymine and uracil only differ by the presence of a **methyl group** on the **5'C** of **thymine**



- 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

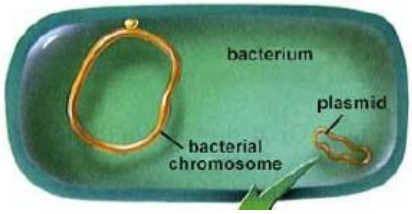
- **Transfer RNA (tRNA):**
  - **Recruits amino acids** to ribosomes during translation for the synthesis of proteins
  - **Single stranded RNA molecule**
  - **Base pairing** exists within the molecule
    - Has **double stranded regions**
    - Pairing gives the molecule a **conformation**
  - the **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…)

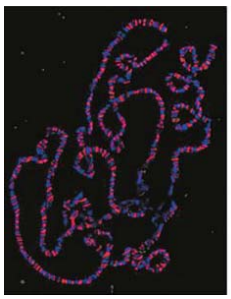


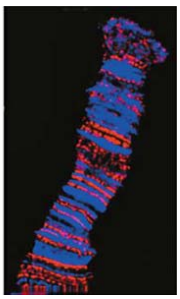
Eukaryote DNA

- 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

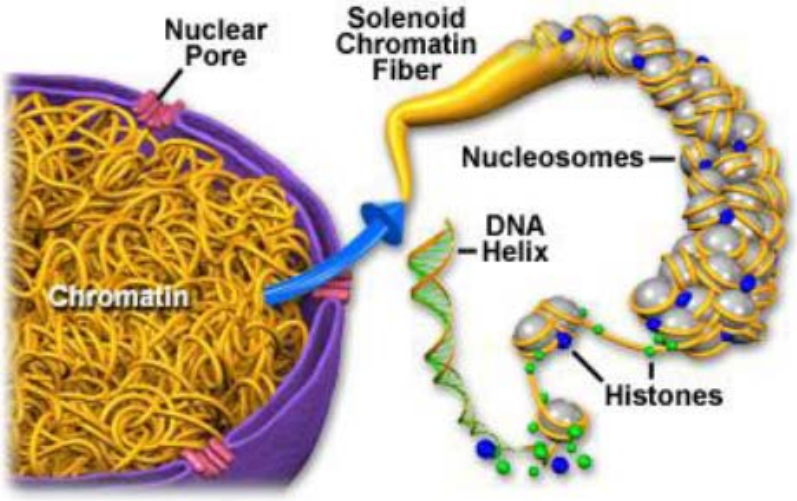




- **Lampbrush Chromosomes**
  - Found in **vertebrate oocytes**
  - DNA becomes **uncoiled during meiosis**, when **high levels of expression** are needed for metabolic purposes associated with meiosis

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)

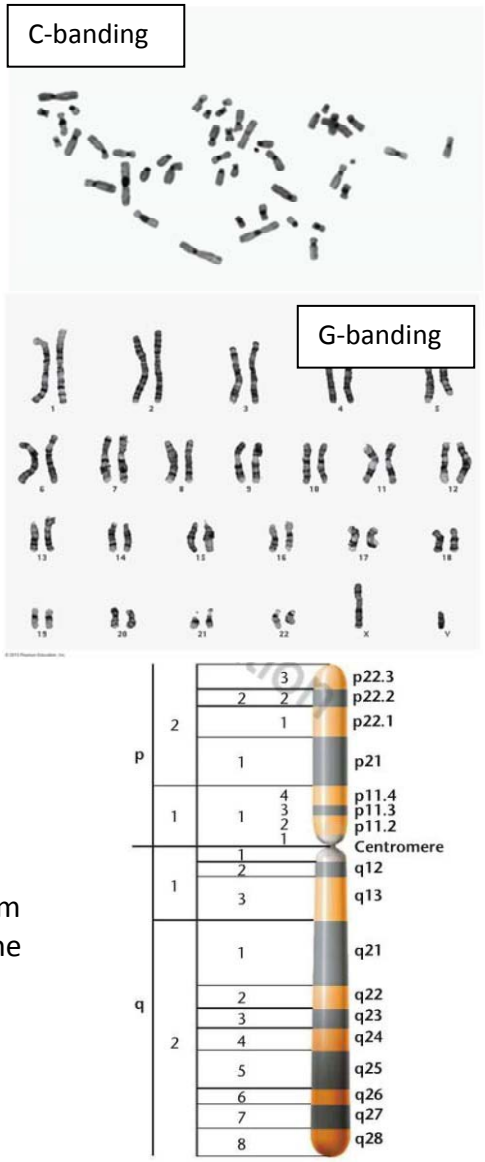


Euchromatin versus Heterochromatin

- **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

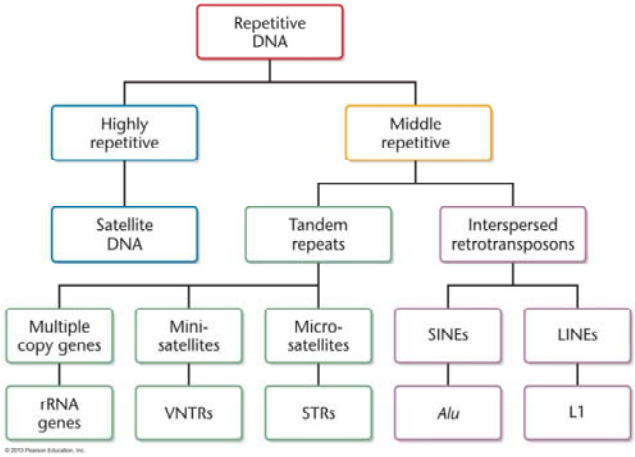
Molecular Anatomy of Chromosomes

- **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

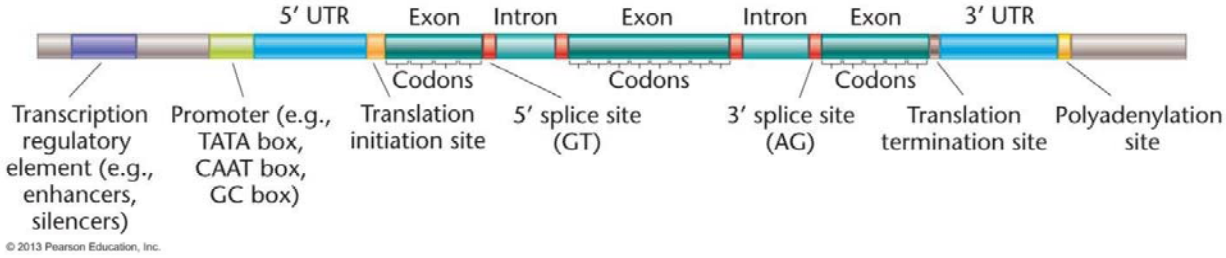
- **Transposable** -> can move from one part of the chromatin to another
- Integrated viral DNA in our cells
- Sequences that are **interspersed individually** throughout the genome (not tandemly repeated)
- Can be used as genetic markers (individualism)

- **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 an 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

Topic 2: DNA Transcription, Translation and Mutations

Forward Genetics

1. We know about the phenotype
2. Induce mutations in a certain gene
3. Screen genetic variants
4. Confirm transmission – did the phenotype change?
5. Links observed phenotypes with associated gene sequences

Reverse Genetics

1. We know about the gene
2. Gene function (confirm sequence)
3. Screen genetic variants
4. Induce mutations – mutate the gene that we know about
5. Does the phenotype change?
6. Links genes with associated phenotypes

Types of Gene Regulation

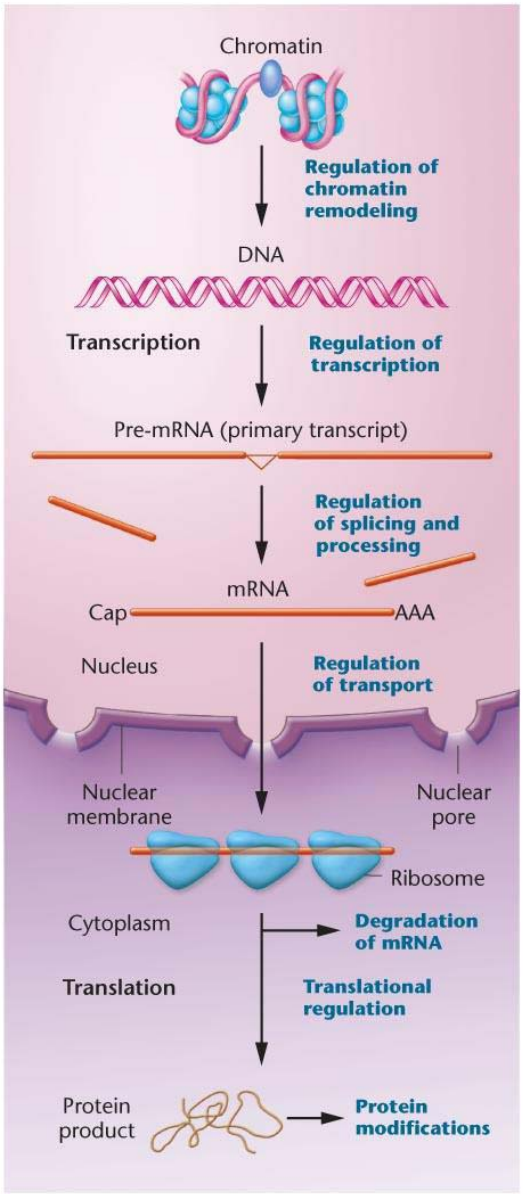
- **Spatial** (different areas in the body have different expressed genes therefore changing the type of cells (heart cells vs skin cells vs bone cells))
- **Temporal** (development, regulation over time (growing diseases or cancers)).
- **Conditional** Self-changed (working out, pregnancy, etc)

Control of Gene Expression

Transcription

Initiation

- RNA polymerase will not bind to the DNA fragment without the promoter elements. (primer, promoter box sequence, start codon, initiator)
- Once on, the DNA will open up (two strands separate) DNA Helicase splits the two strands of DNA (I wish I were helicase so I could unzip your genes)
- Then RNA polymerase begins the transcription process



Elongation

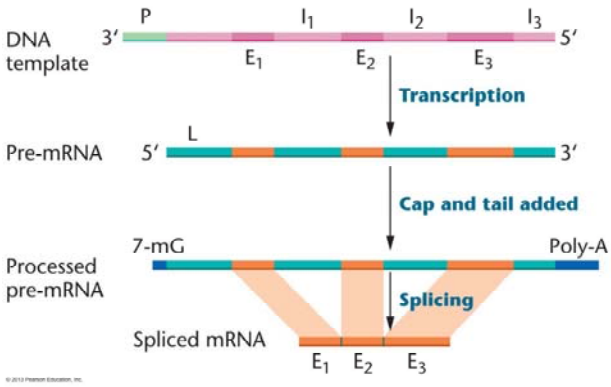
- Nucleotides added in the **5' to 3' direction** (polymerase adds to the 3' hydroxyl group)
- **DNA template is 3' to 5'**
- Nucleotides will be added until it gets to the termination site

Termination

- Once RNA polymerase ends transcription, it is released
- The produced mRNA is free to travel through the nuclear membrane into the cytoplasm to be translated by ribosomes

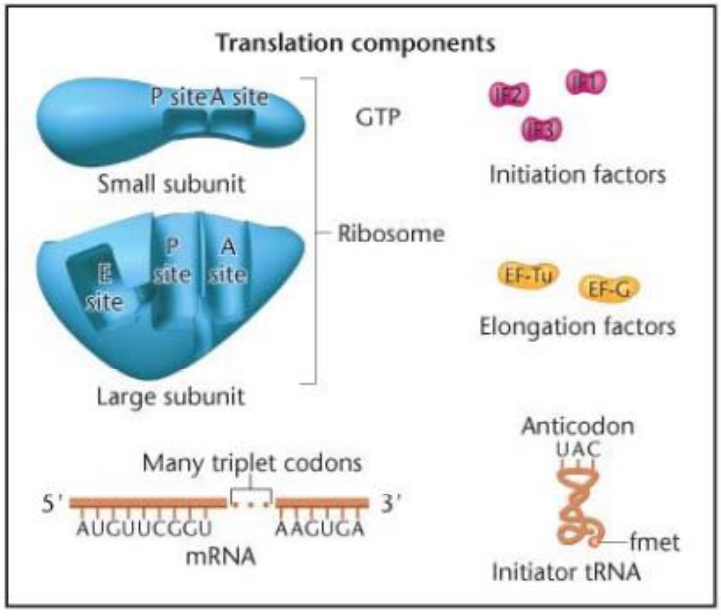
Post-transcriptional modification

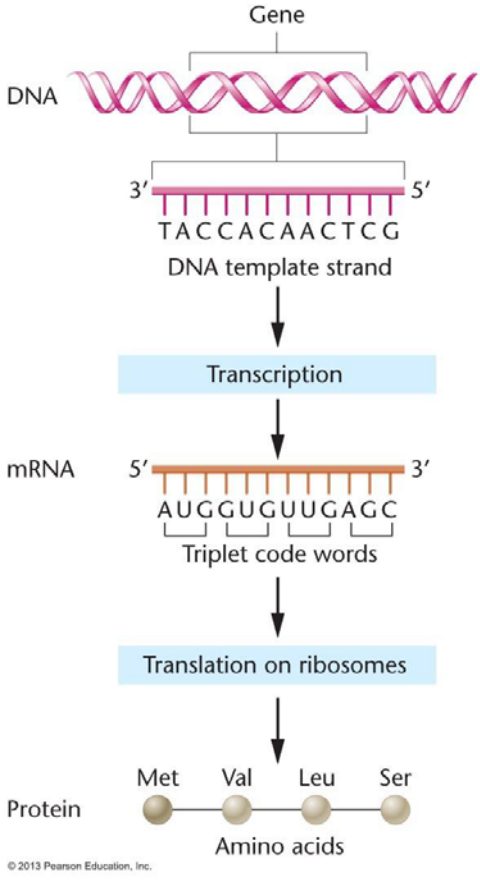
- Extremely important step between transcription and translation
  - Cap and Poly-A tail are added to protect against degradation
  - Introns (non-coding regions) are spliced out of the RNA strand
  - Exons (coding regions) are pieced together



Translation Needs

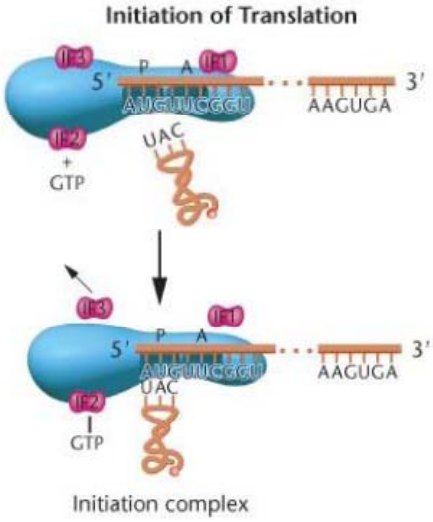
- RNA
- Ribosome (two subunits – small & large)
- **Large Unit: A site, P site, E site**
- **Small unit : A site, P site**
  - **Activator (A), Peptide addition (P), Exit (E)**
- tRNA: brings AA to ribosome by binding to specific anti-codon





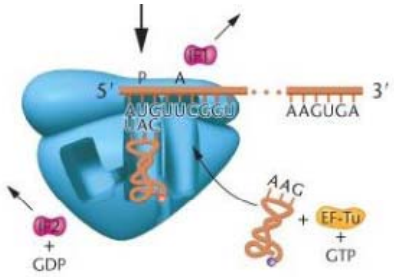
Translation

Initiation



- **Three initiation factor proteins** (IF 1,2,3) bind to the small subunit of the ribosome
- Methionine-bearing tRNA binds to the mRNA
- Large ribosomal subunit binds to this complex, encasing the mRNA strand
- Initiation factors (IF's) are released
- Elongation begins

Elongation

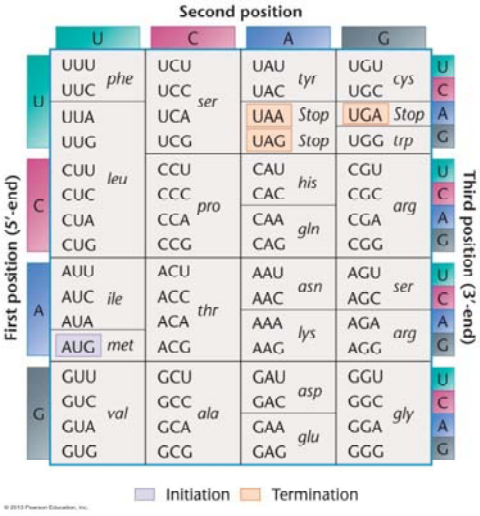


- Ribosome moves along RNA 5' -> 3' (**translocation**)
- tRNA carrying the corresponding second codon binds at the A site
- **Peptidyl transferase** forms peptide bonds between the amino acids
- Ribosome then translocates again so that the first tRNA is in the E site and the **A site opens up** for further elongation of the polypeptide chain

Termination

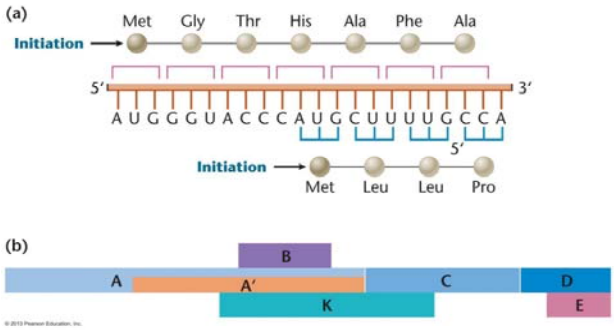
- Once a **stop codon** is reached, a **release factor** binds and facilitates the release of the mRNA from the ribosome
    - There are no tRNAs programmed to recognize the stop codons
- Ribosome dissociates, translation ends

The Genetic Code



- Essentially universal (**unambiguous**)
- **Codons** (neucleiotide triplets) predict protein sequence
- **Many codons** for the **same Amino Acid - DEGENERATE**
    - 61 codons, 20 amino acids
- Initiation (**AUG** -methionine) and Termination codons (**UAA, UAG, UGA**; no tRNA anti codon)
- Non-overlapping but reading can start at different locations
    - ie: more than on methionine codon on the RNA

Different Initiation Points Create Overlapping Genes



- There is often more than one start codon on a gene sequence
- Depending on where translation begins (where the ribosome binds), the ribosome may read in a different open reading frame
    - This results in a totally different protein product!
    - The open reading frame spans from the start to stop codon on the mRNA
- The ribosome could also bind at a later start codon on the mRNA strand and result in a shorter DNA strand – shorter peptide chain, different protein
- **Many different proteins may be produced from a single template strand!**

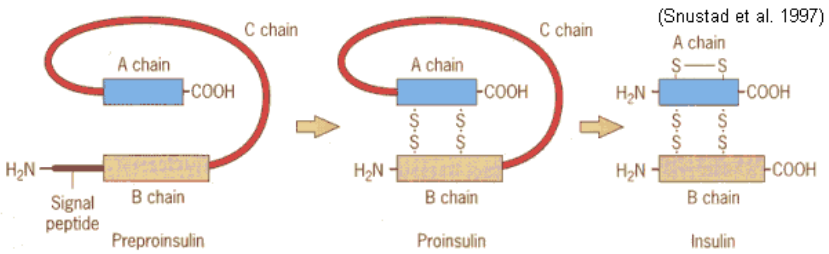
Features of tRNA

- Refer to page 4 for a picture
- Transfer RNA recruits specific amino acids for translation
    - Each tRNA has a unique, constant anticodon

Post-translational Modifications

- **Initial protein product** after translation is **not necessarily functional**.
    - Protein must find its **correct conformation** and orientation.
- **N-terminus** is the first end to exit the ribosome – basically the first part of the protein
    - It generally has a signal protein attached to it to direct it to the proper destination
    - Removed or modified
- Individual **amino acids** can be modified
    - May be an important step to form the proper conformation
- **Carbohydrate sections** could be added (ex. antigens in RBCs)
- Polypeptide chains may be **trimmed**
- **Signal sequences** removed (ex. lamprey growth)
- Polypeptides can be complexed with **metals** (ex. iron atoms in hemoglobin)

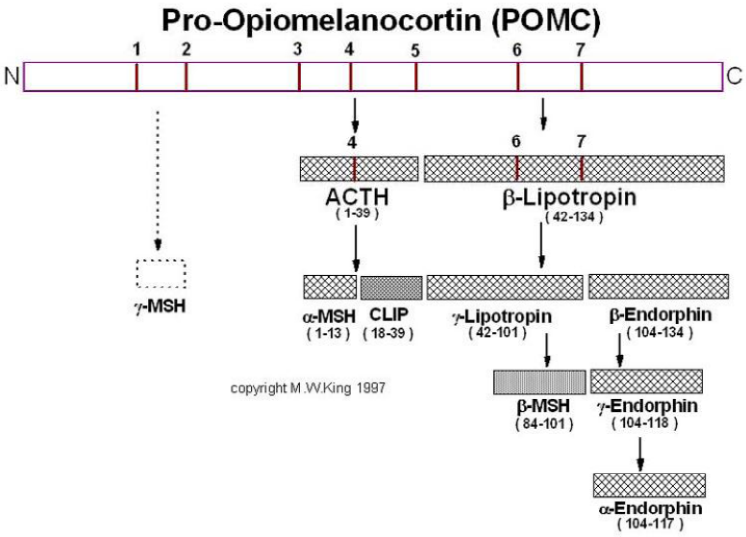
Cleaving of Polypeptide Chains - Insulin



Polypeptide chain connecting the A and B peptide chains is cleaved to form the functional protein

Cleaving of Polypeptide Chains – Multiple proteins from one Amino Acid chain

- After translation, a polypeptide chain may be **cut in different ways** to form **various different proteins**
- As seen in the diagram, 7 cut sites are used by enzymes to create 10 different proteins – all from one polypeptide chain!

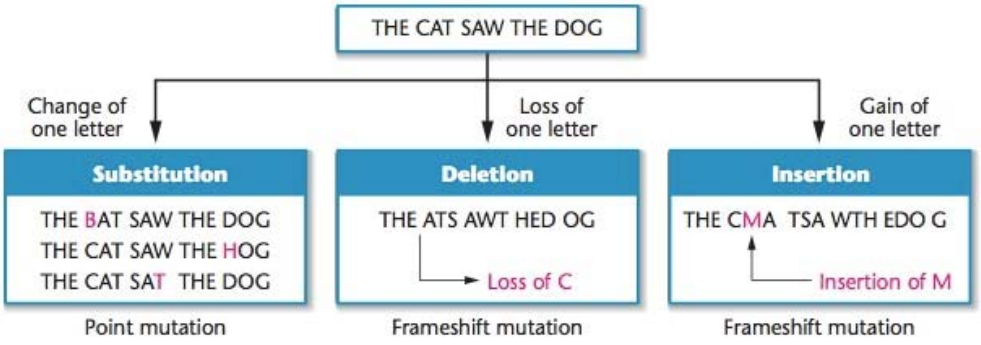


Mutations in DNA and their Consequences

- Can be induced or spontaneous, but are **always random**
- Mutations **only occur in DNA**
- Mutations to individual base pairs are called **point mutations**
- Base **insertion or deletion** is usually the **most serious** mutation

Types of mutations at the DNA level	Results at the molecular level
No mutation	**Wild type** Thr Lys Arg Gly Codon 1 Codon 2 Codon 3 Codon 4 A C A A A G A G A G G T **Codons specify wild-type protein.**
Transition or transversion	**Synonymous mutation** Thr Lys Arg Gly A C A A A G A G C G G T **Altered codon specifies the same amino acid.**
	Missense mutation (conservative) Thr Lys Lys Gly A C A A A G A A A G G T **Altered codon specifies a chemically similar amino acid.**
	Missense mutation (nonconservative) Thr Lys Ile Gly A C A A A G A T A G G T **Altered codon specifies a chemically dissimilar amino acid.**
	Nonsense mutation Thr STOP A C A T A G A G A G G T **Altered codon signals chain termination.**
Indel **Base insertion**	**Frameshift mutation** Thr Glu Glu Arg ··· A C A G A A G A G A G G T ···
Base deletion	**Frameshift mutation** Thr Arg Glu Val ··· A C A A G A G A G G T ··· X

- Frame shift mutations are the most damaging mutations
    - Deletion or addition of a base pair in the middle of the reading frame will completely change the codon sequence and will often render the protein useless



- However, in general, proteins are troopers – they are often able to function even with a few mutations in the template
    - Function of the protein may be affected (weaker affinity, activity, etc.)

Classification of Mutations and their Causes

- Mutations are often described by cell type and chromosome type affected

- **Cell types**
    - **Somatic cells** - all cells except germ-line cells
        - Mutations occurring in somatic cells will not be passed onto the offspring
    - **Germ-line cells** - stems cells and cells in gonads (i.e. gametes).
        - Mutations occurring in germ line cells will potentially carry into the offspring

- **Chromosome types**
    - A karyotype defines the number and appearance of chromosomes in a eukaryote
    - The human karyotype contains 23 pairs of chromosomes
        - **Autosomal chromosomes** – all chromosomes except sex-linked (22 pairs)
        - **X-linked chromosomes** – sex-linked chromosomes (1 pair)
    - Whether or not mutations are passed onto offspring **depends on the type of cell**
        - All cells have the same DNA, but **only germ-line cell DNA is passed on**
        - This means that any mutations in somatic cell DNA will not be passed on
    - X-linked mutations are mutations to the X or Y chromosomes (sex-linked)
        - Since the X chromosome is much larger than the Y, there are more X-linked traits than Y-linked traits
        - However, since males only carry one X chromosome and one Y chromosome, they are more likely to express mutations (no back-ups!)
        - Females have two X chromosomes, so they have a chromosomal back-up

Spontaneous vs. Induced mutations

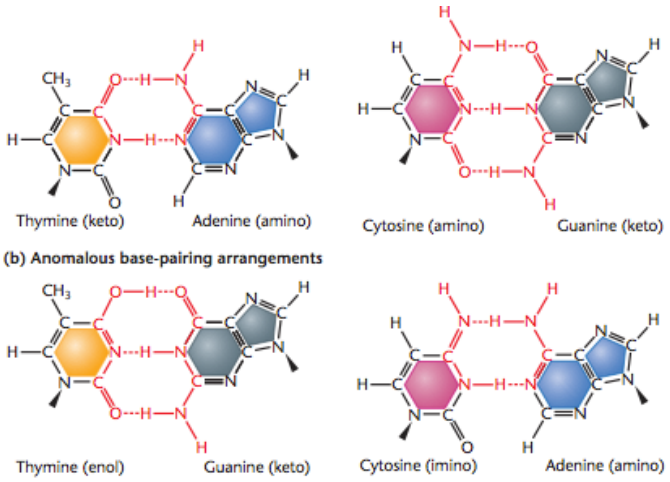
- **Spontaneous** – Occur naturally at random during cellular processes (DNA replication)
- **Induced** – Mutations that researchers deliberately cause through usage of mutagens (chemicals, radiation…)

Spontaneous Mutations

- Cells have mechanisms to correct mutations, but they are often missed and mutations will spread to other cells through replication
- During cell division, DNA is replicated, and copies of both strands go to different cells.
    - If one of the bases change right before replication, or if there's change in base at a location when replication happens, a mutation may not be corrected, and the mutated gene will become widespread
    - During replication, polymerase uses only one strand as a template – there is no way to check if this template is perfectly homologous to the other strand
- **Oxidative mutations** occur when DNA is mutated via the presence of reactive oxidants
- **Replication slippage mutations** occur when a polymerase encounters a tandem repeat of base pairs (ex. GGGGGGGG) – a base pair may be missed/added
- **Depurination and deamination** changes the chemical nature of the molecules (basically unwanted chemical reactions that happen to the base pairs)
- **DNA replication errors** – resonance that occurs in base pairs may increase or decrease the number of hydrogen bonds that can be formed by a base (aka tautomeric shifts)

Tautomeric Shifts

- Resonance of a base pair results in improper bonding
- Improper base pair matching may occur and will not be corrected
- As seen in this image, anomalous base pairing occurs when thymine resonates and is able to form an extra H-bond – in this case, it improperly binds to guanine
- When DNA replication occurs, the two DNA strands are separated and their individual complimentary strands are added – the incorrect base pair now appears to be a regular base pair and a point mutation results



Depurination & Deamination

- **Depurination** is the loss of nitrogenous base from a purine
    - This leaves an apurinic site; if this site is not repaired, DNA polymerase will fill the gap with a random nucleotide during replication
- **Deamination** is the conversion of an amino group in cytosine or adenine into a keto group
    - Cytosine becomes uracil and adenine becomes hypoxanthine

Induced Mutations

- Induced mutations increase the rate of mutation above the natural background
- In order to induce mutations, mutagens are needed to cause damage to the DNA
- Mutagens can be pollutants, toxins and radio-active waves

Base Analogs

- Base analogs are a category of mutagenic chemicals
- These chemicals can take the place of purines or pyrimidines during DNA replication

Alkylating Agents

- Category of mutagenic chemicals
- These are naturally occurring chemicals inside humans
    - They can alter the structure of DNA, causing mutations
- Donate alkyl groups to the keto or amino groups in the nucleotide bases

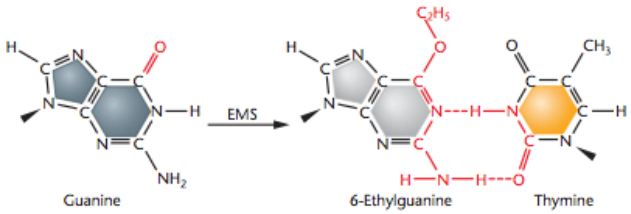


FIGURE 15–6 Conversion of guanine to 6-ethylguanine by the alkylating agent ethylmethane sulfonate (EMS). The 6-ethylguanine base-pairs with thymine.

Radiation

- **Ultraviolet Light**
    - Short wavelengths of light are able to distort pyrimidines by creating dimmers
    - These dimers disrupt the DNA sequence and disrupts replication

- **Ionizing Radiation**
    - Consists of higher energy and shorter length waves like gamma rays and x-rays
    - Rays ionize the base pairs and create point mutations
    - These mutations can cause chromosome aberrations (deletions, translocations, and fragmentations of the DNA)