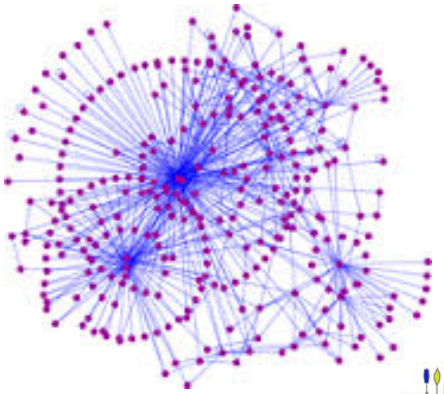
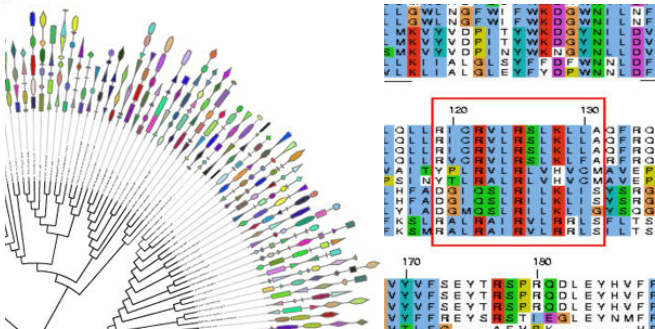
Background

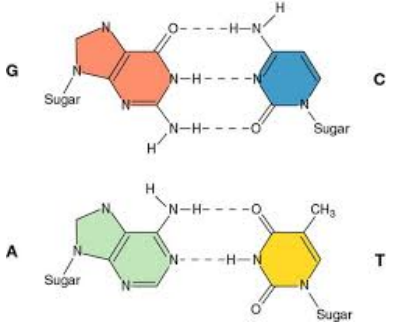
- Education:
  - Snow: AA/AS-Spanish and Chemistry
  - USU: BS-Biochemistry (Research in Microbial Ecology)
  - University of Iowa: PhD-Bacterial Pathogenics and Genetics
  - University of Iowa: Postdoc-Molecular Pathophysiology in Otolaryngology





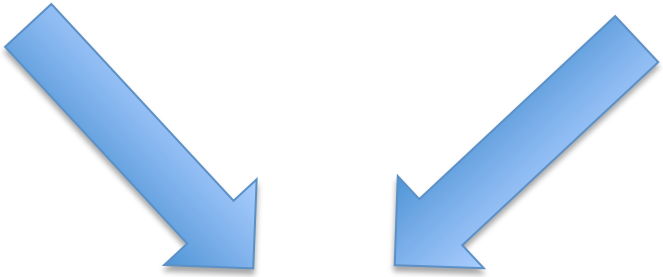
Bioinformatics: Coding for Life with DNA

History of DNA



- Fred Griffith (1928) and Oswald Avery (1943) used two different strains (Rough and Smooth) of *S. pneumoniae* to infect mice

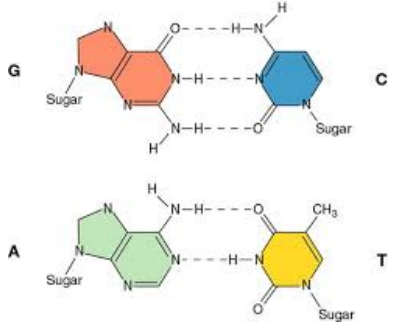
Rough (cannot kill) Smooth (deadly)



Mouse Dies

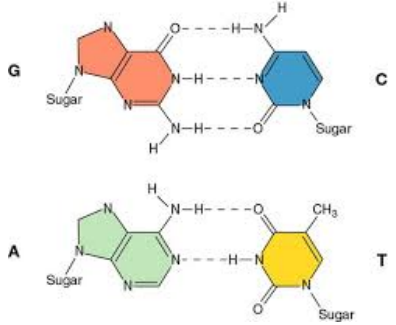
Isolate bacteria from dead mouse and only finds Smooth bacteria. Why?

History of DNA

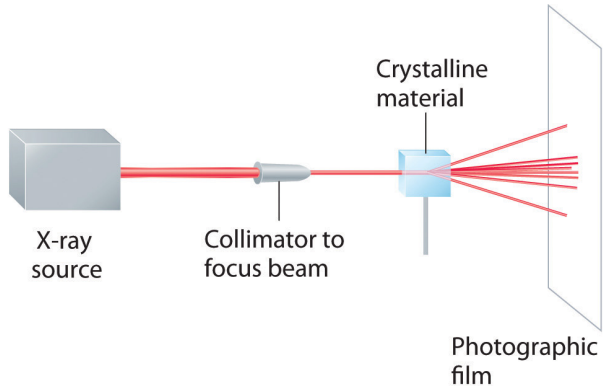


- Edwin Chargaff (1950) identified the chemical composition of DNA and concluded that contained a deoxyribose sugar, phosphate, and one of four bases: adenine, thymine, guanine, and cytosine.
  - The concentration of adenine always equaled thymine
  - The concentration of guanine always equaled cytosine

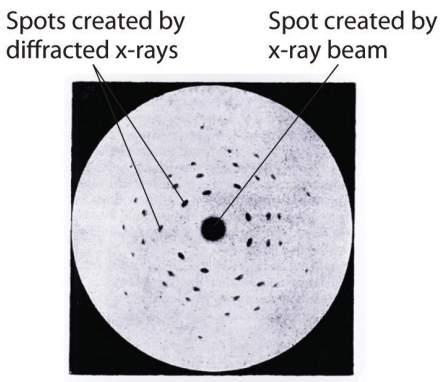
History of DNA



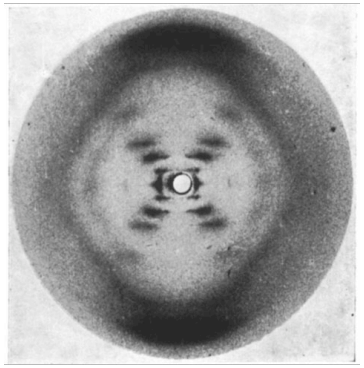
- Alfred Hershey and Martha Chase (1952) identified that DNA is the molecule of heredity
  - That which is passed on from species to species
- Rosalind Franklin (1952) used a technique called x-ray diffraction to study crystallized DNA.



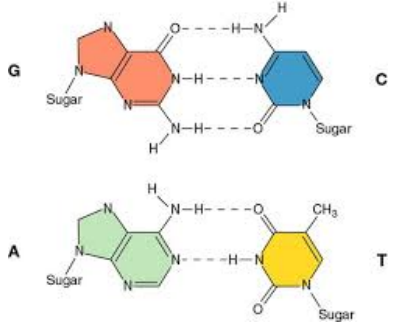
(a) X-ray diffraction



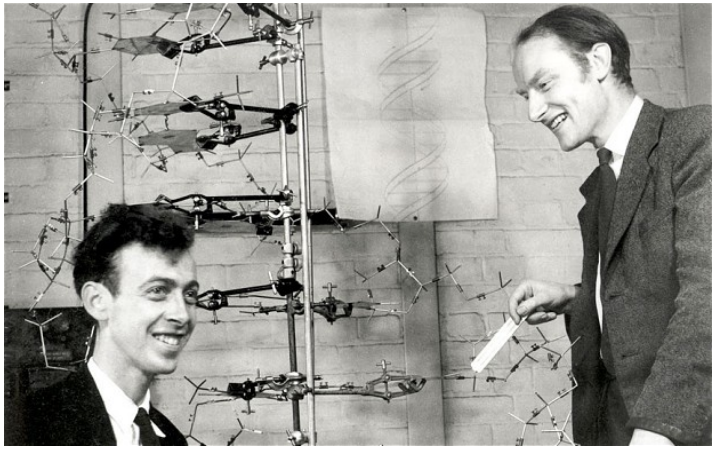
(b) X-ray diffraction pattern captured on photographic film



History of DNA

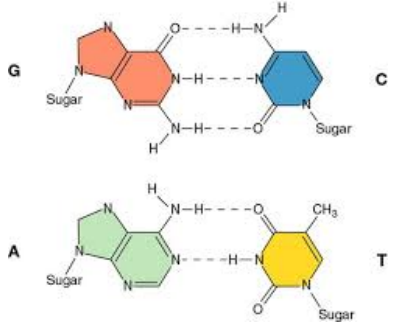


- James Watson and Francis Crick (1953) used Franklin's X-ray image identified the structure of DNA as a double helix.



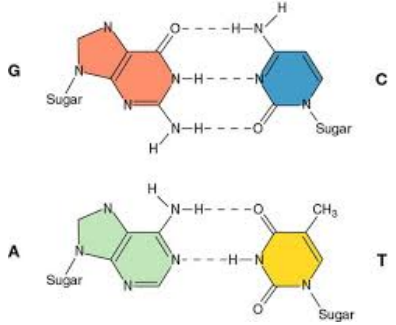
- Frederick Sanger (1975) developed a way to sequence (figure out the order of the bases).

History of DNA



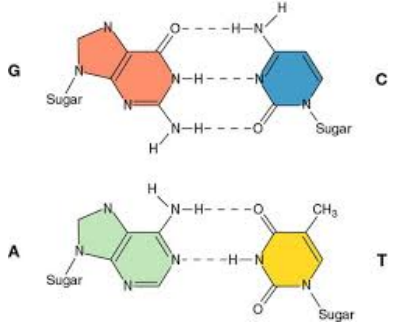
- Gementech (1977-79) made the first genetically modified organism. Microorganisms produced insulin and human growth hormone.
- Kary B. Mullis (1985) published research on PCR (polymerase chain reaction) that is used to make millions of copies of tiny amounts of DNA quickly.

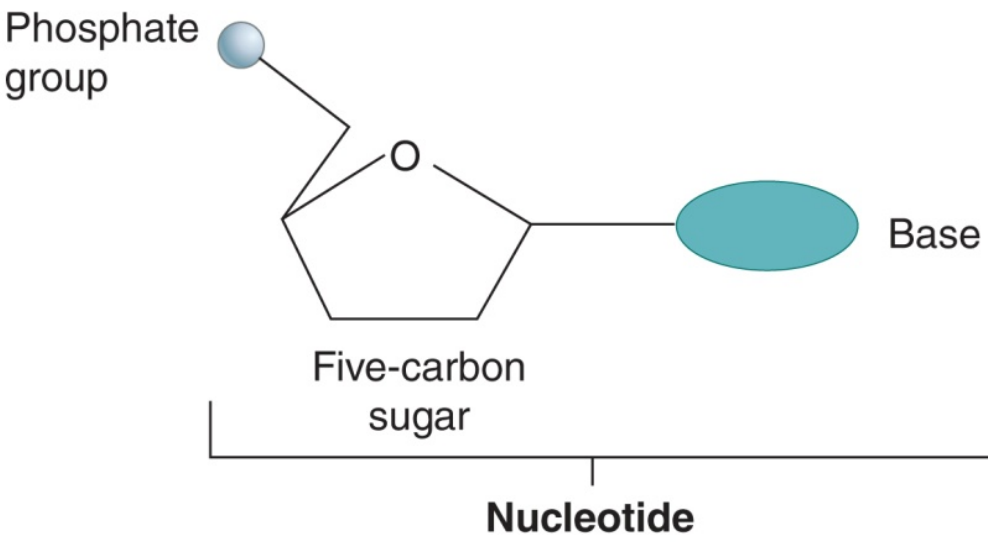
History of DNA

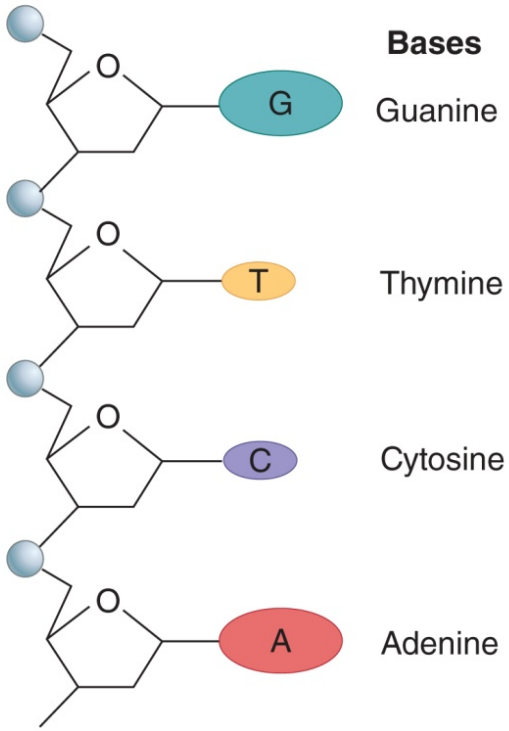


- 1990. NIH (National Institutes of Health) approved gene replacement therapy for an immunodeficiency disease.
- 1993. Calgene Inc. produce the first genetically modified crop. Tomatoes were given resistance to rotting.
- 1996. Dolly the sheep is cloned.
- 2003. Human genome sequenced (13 years)

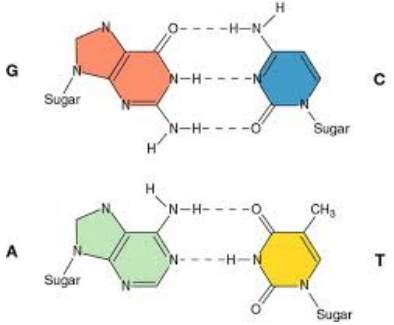
What is DNA?


G
C
A
T
Sugar
Sugar
Sugar
Sugar


Phosphate group
O
Base
Five-carbon sugar
Nucleotide

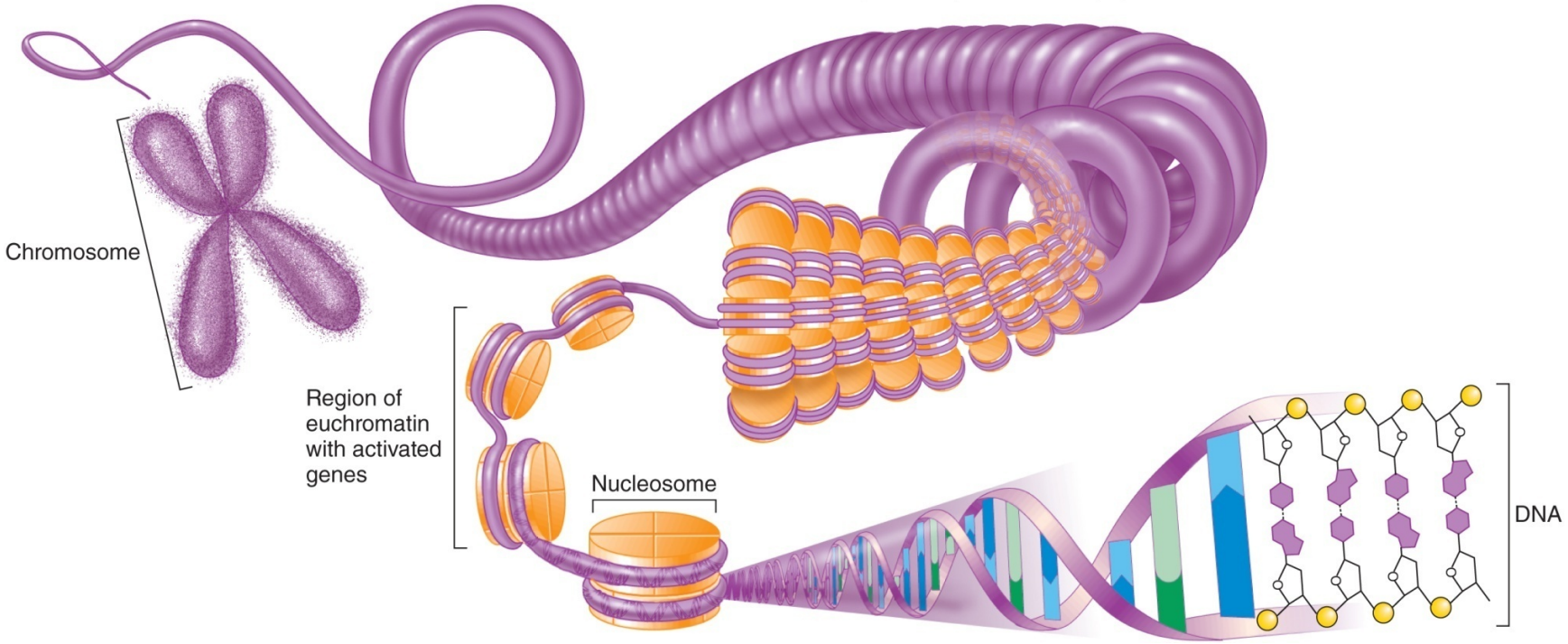

Bases
O
G
Guanine
O
T
Thymine
O
C
Cytosine
O
A
Adenine

What Does DNA Look Like?

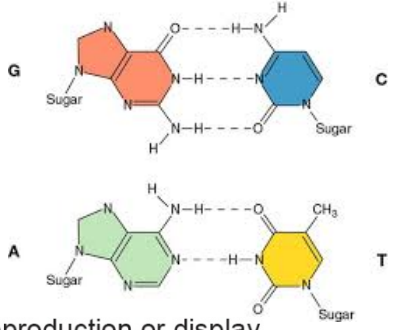


Each cell has about 2 Meters of DNA. Average human has ~100 trillion cells. That is enough DNA to go from the earth to the sun more than 400 times! DNA has a diameter of 2×10^{-9} m.

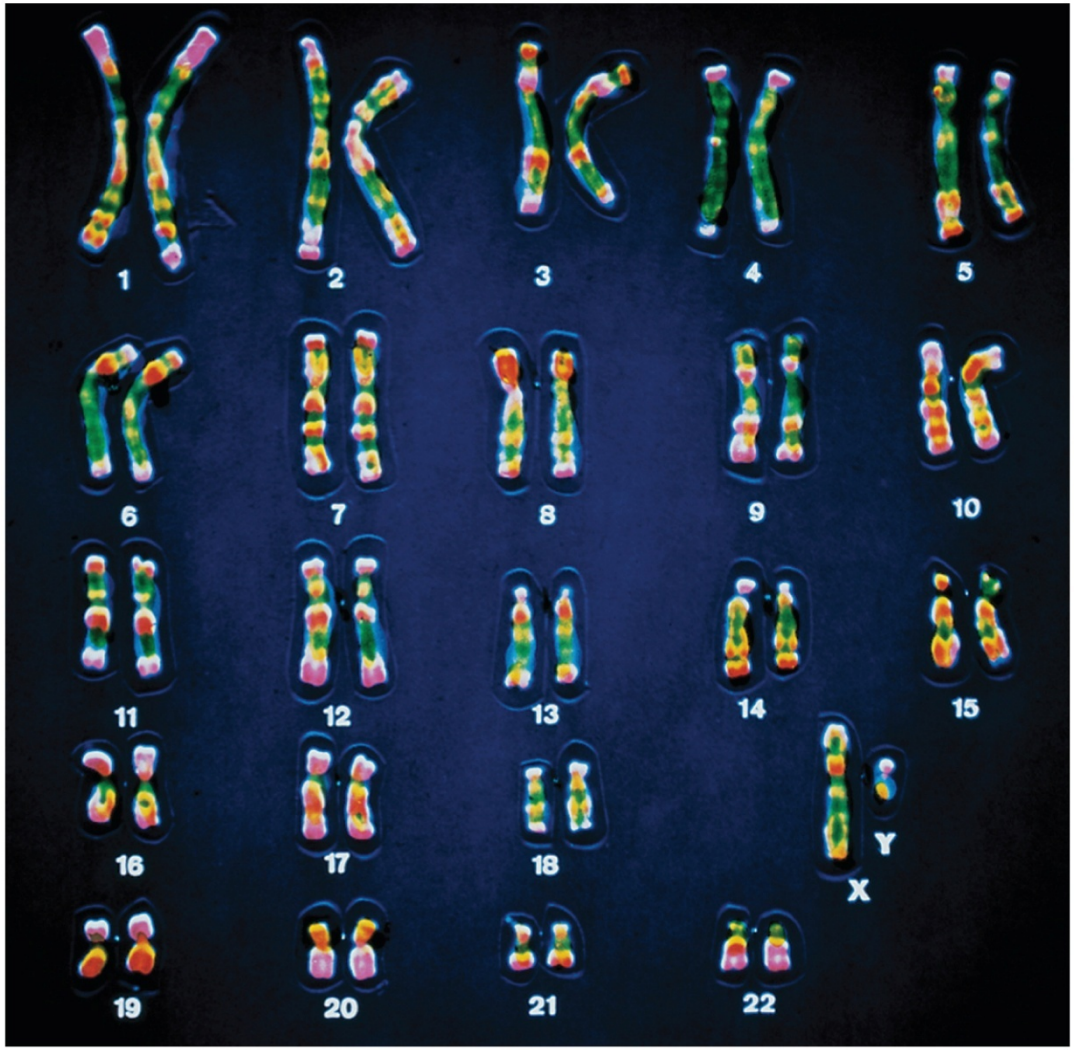
How can you fit the much DNA into each cell?



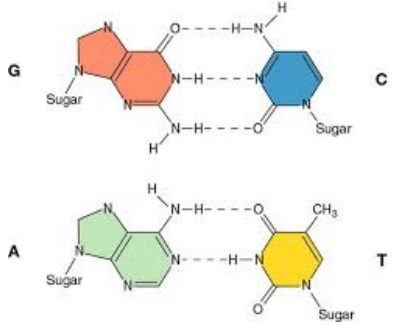
What Does DNA Look Like?



- 23 chromosomes in duplicate
  (46 total in each cell)

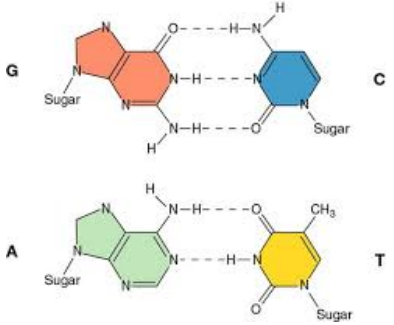


What Does DNA Do?

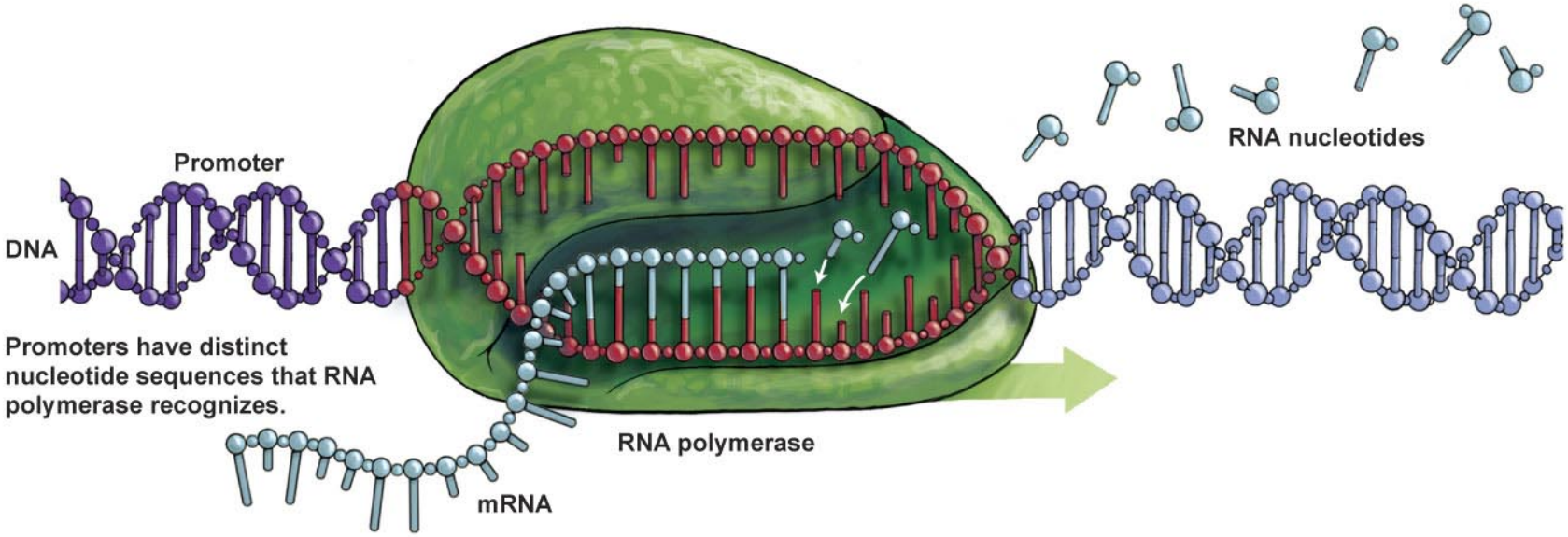


- We have ~3 billion base pairs that comprise our set of 23 chromosomes
- DNA can undergo transcription (RNA), translation (Protein), replication (Mitosis), or Meiosis (only specialized cells, sperm and ova)

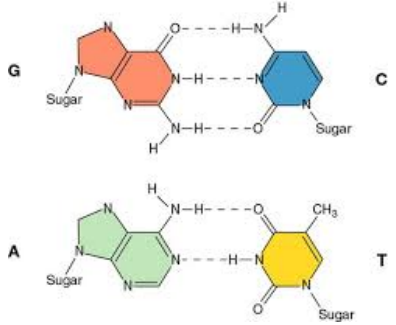
Transcription



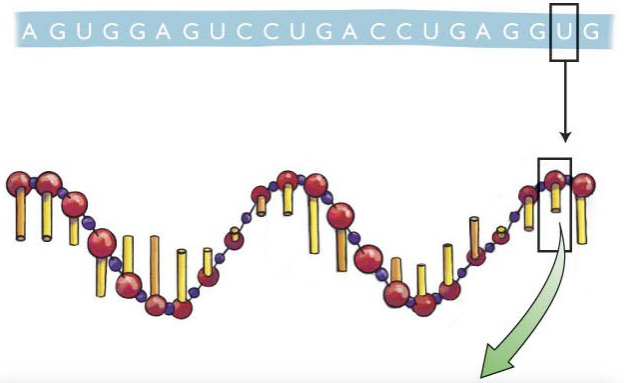
- Transcription:
  - Takes the DNA base pairs to RNA base pairs and then the non-functioning protein segments (introns) are removed

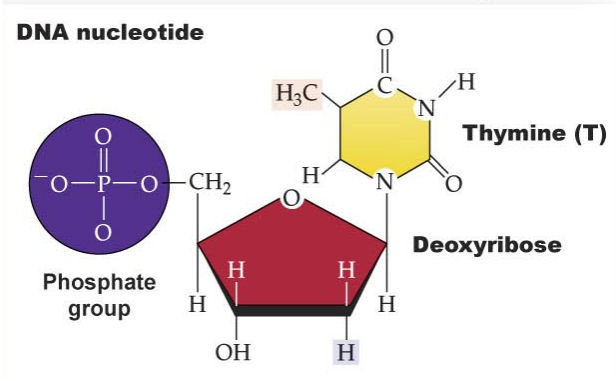


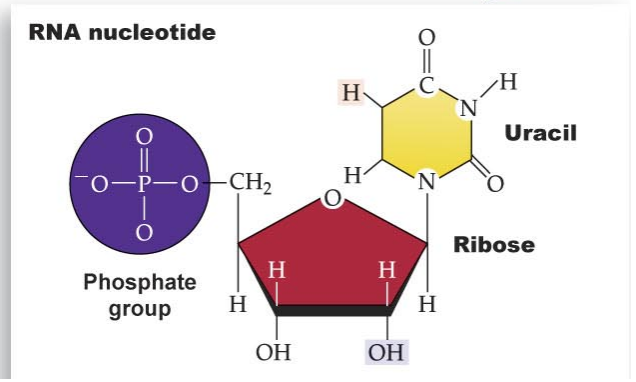
Difference Between DNA and RNA



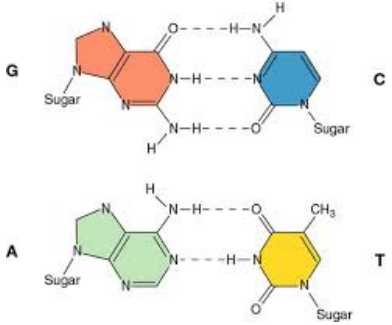
- DNA
  - Double stranded
  - Contains Thymine
  - Deoxyribose
- RNA
  - Single stranded
  - Contains Uracil
  - Ribose

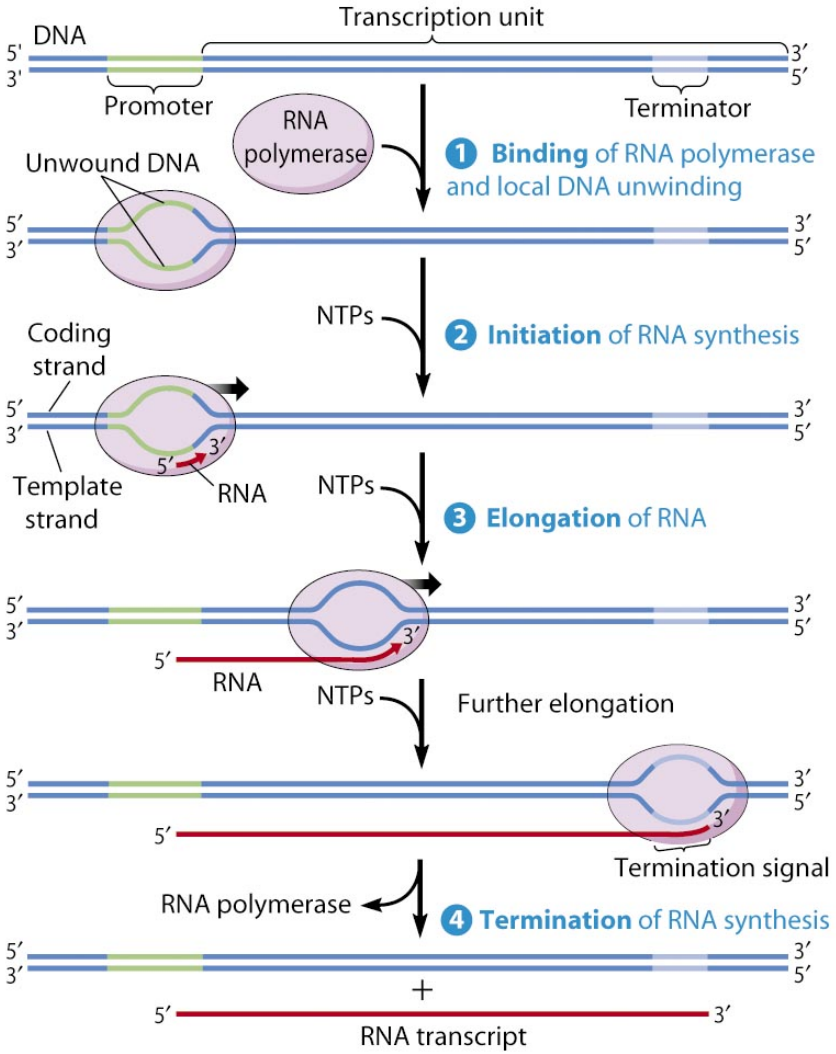




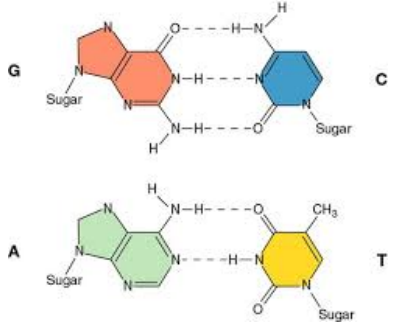


Transcription Detailed

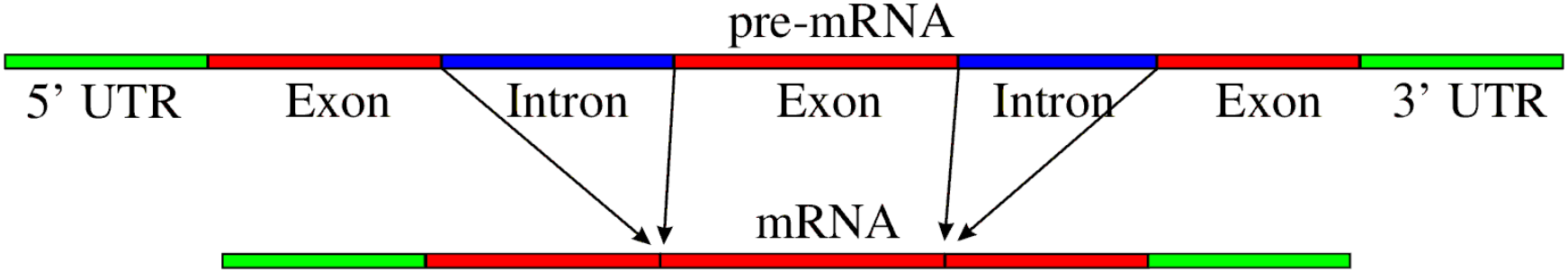




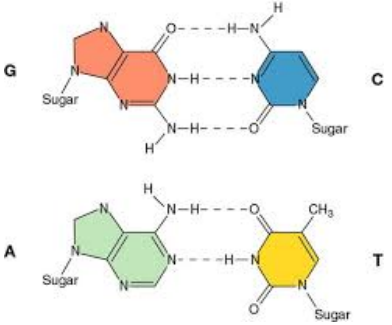
Transcription Detailed

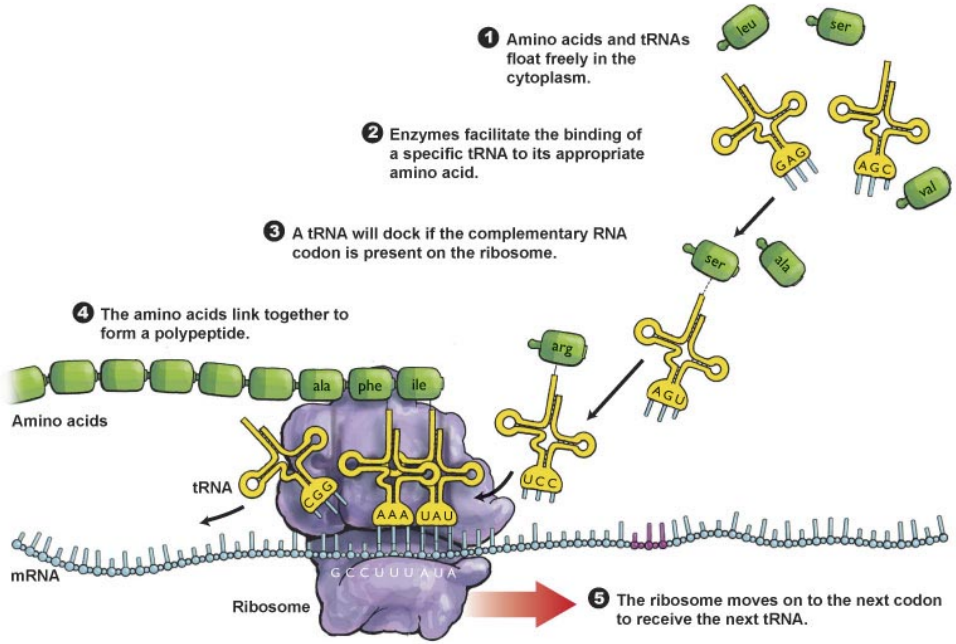


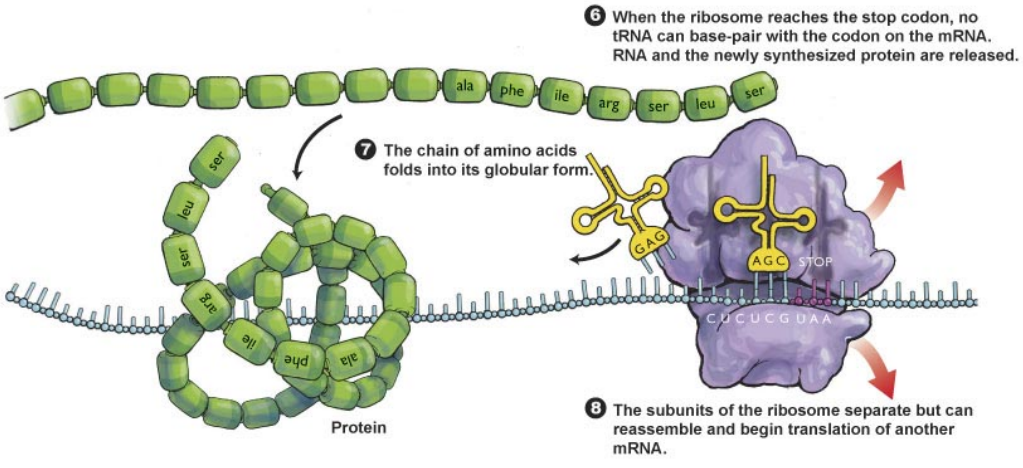
- Pre-messenger RNA (pre-mRNA)
  - Contains:
    - Introns- DNA sequences that don't code for proteins
    - Exons- DNA sequences that are protein building instructions
      - A Gene is consists of introns and exons.
- Introns are cut out.



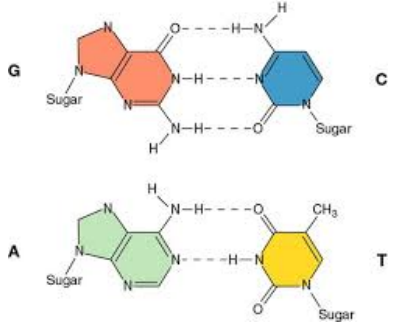
Translation





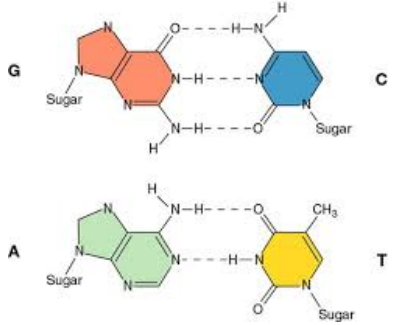


Gene to Proteins



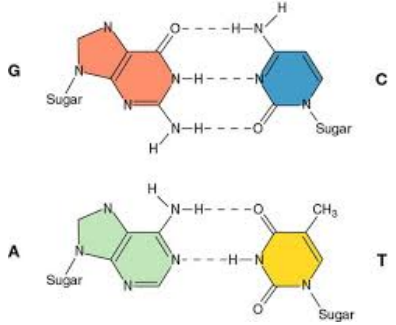
- DNA code
  - Triplets
- Transcription
  - DNA code is transcribed to RNA code as codons (introns cut out), final product mRNA (messenger)
- Translation
  - mRNA is transported to a ribosome
  - tRNA's with an anti-codon and associated Amino Acid recognize the codon of the mRNA and a polypeptide is made

Genetic Code



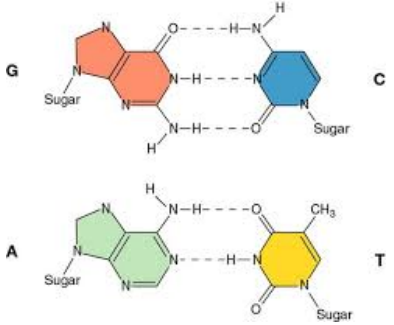
First base	Second base: U	Second base: C	Second base: A	Second base: G	Third base
U	UUU, UUC: Phenyl-alanine (phe) UUA, UUG: Leucine (leu)	UCU, UCC, UCA, UCG: Serine (ser)	UAU, UAC: Tyrosine (tyr) UAA: **Stop codon** UAG: **Stop codon**	UGU, UGC: Cysteine (cys) UGA: **Stop codon** UGG: Tryptophan (trp)	U C A G
C	CUU, CUC, CUA, CUG: Leucine (leu)	CCU, CCC, CCA, CCG: Proline (pro)	CAU, CAC: Histidine (his) CAA, CAG: Glutamine (gln)	CGU, CGC, CGA, CGG: Arginine (arg)	U C A G
A	AUU, AUC, AUA: Isoleucine (ile) AUG: Methionine (met) **Start codon**	ACU, ACC, ACA, ACG: Threonine (thr)	AAU, AAC: Asparagine (asn) AAA, AAG: Lysine (lys)	AGU, AGC: Serine (ser) AGA, AGG: Arginine (arg)	U C A G
G	GUU, GUC, GUA, GUG: Valine (val)	GCU, GCC, GCA, GCG: Alanine (ala)	GAU, GAC: Aspartic acid (asp) GAA, GAG: Glutamic acid (glu)	GGU, GGC, GGA, GGG: Glycine (gly)	U C A G

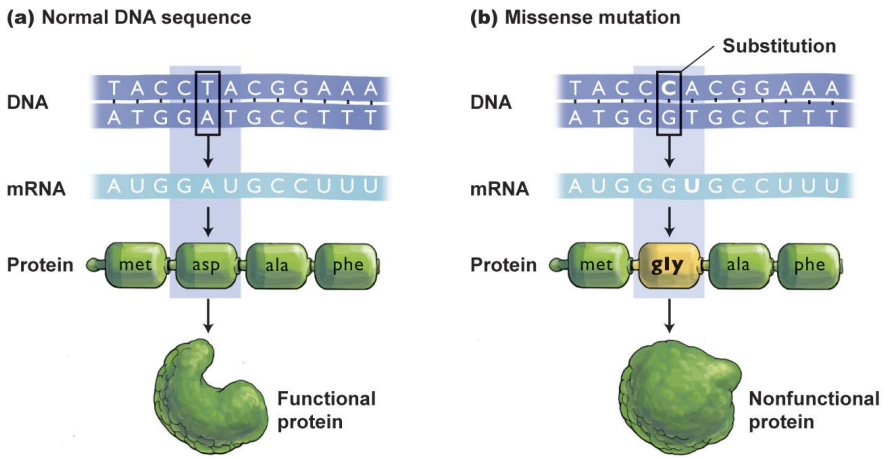
Genetic Code

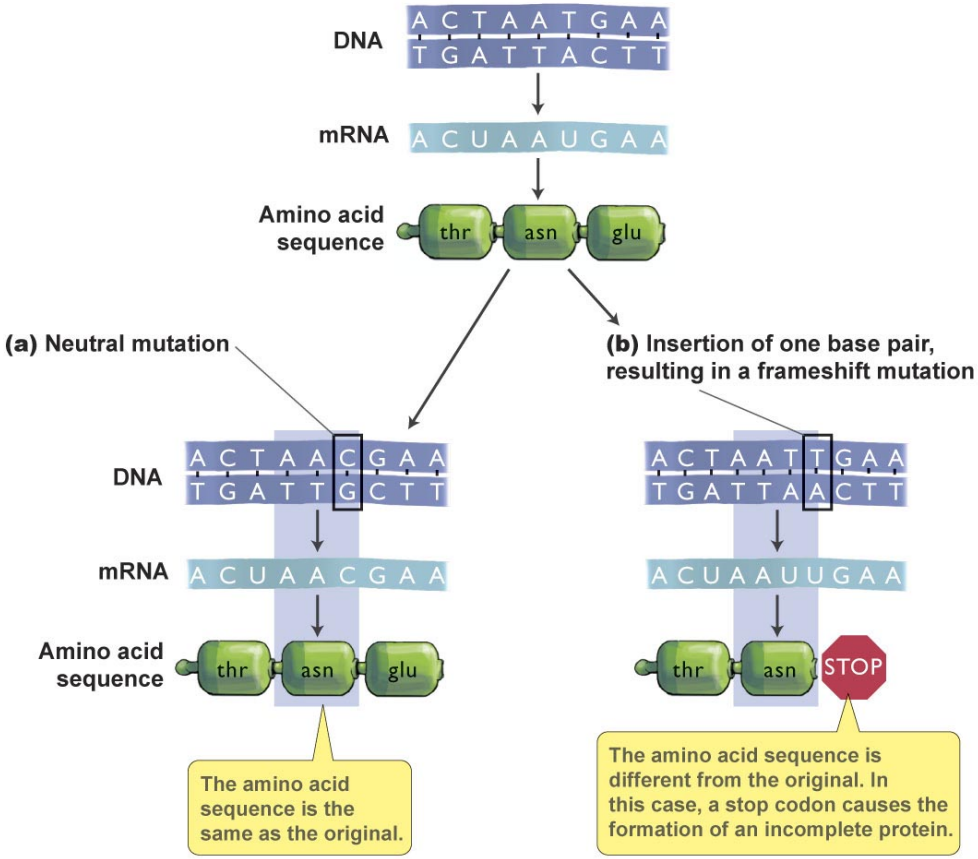


- 64 codons
  - 3 code for stops and 61 code for amino acids
  - A codon never codes for more than one amino acid
  - Code is universal among all living organisms
  - Mutations can result in a non-functional protein or a different protein

Mutations

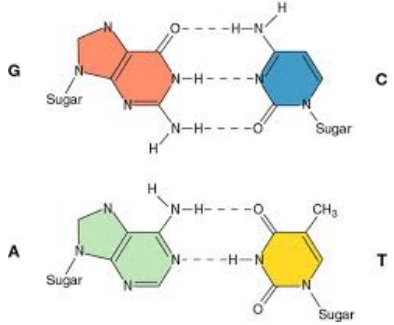


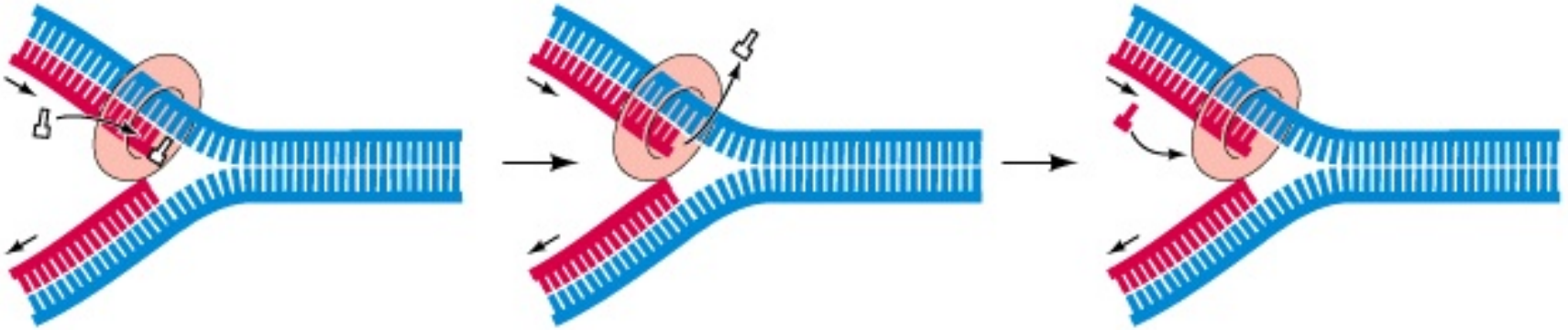


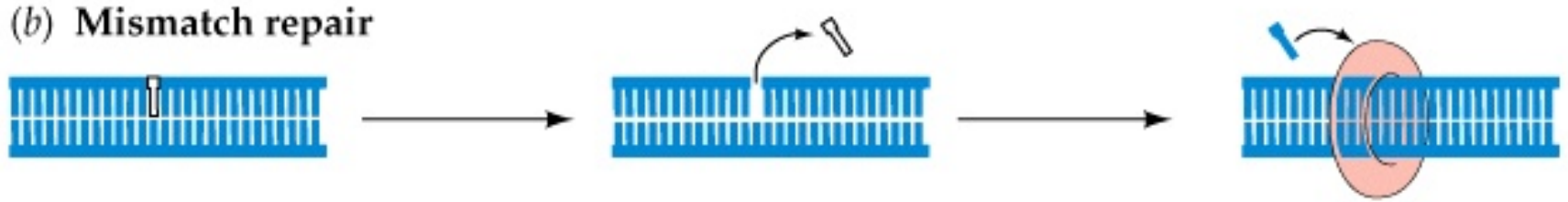


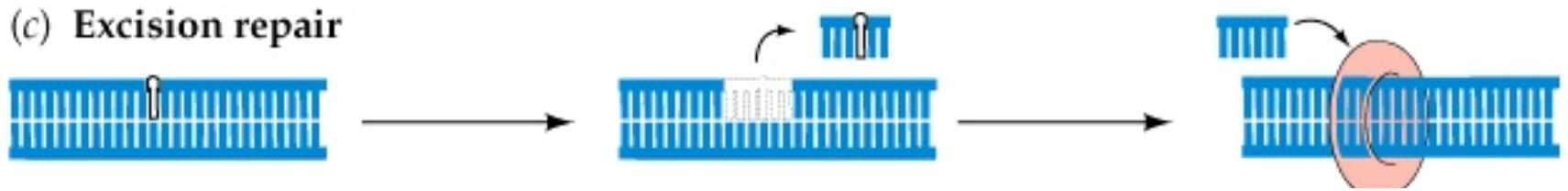
Frameshift: “the cat ate his dog” changes to “the cHa tat ehi sdo g”

Correction Mechanisms

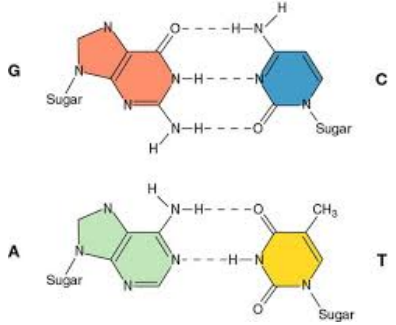








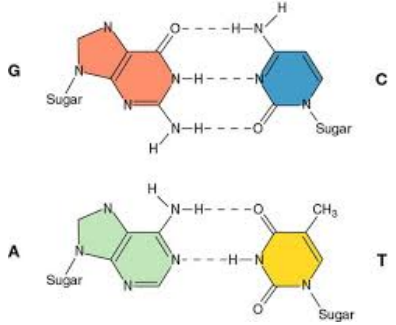
Problems with Mutations



- May mess up the cell cycle
  - Stop
  - Increase replication!
    - Tumors
      - Benign
      - Cancerous

- Can look at your Genome and determine if you have had exposure to radiation, smoke, toxic chemicals by comparing the base pair ratios.

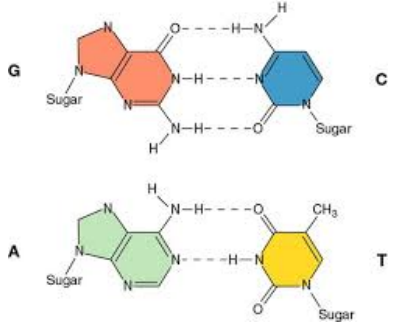
Genome



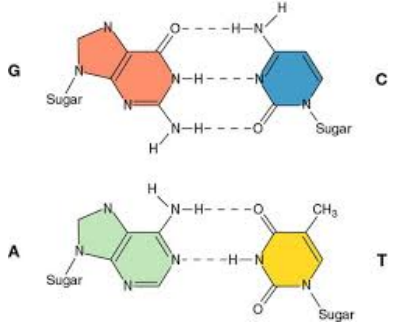
- Consists ~3 billion base pairs
- However, ~3% of these base pairs code for proteins (Exome). What is the other 97% doing?

- We have ~20,000 genes, but we have over 100,000 proteins (Proteome). How?

DNA Extraction

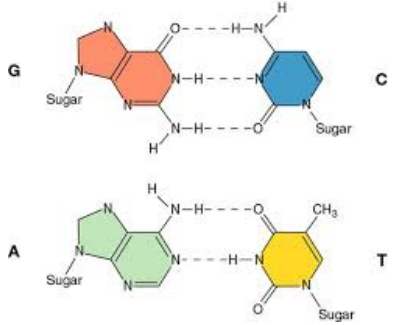


Articles



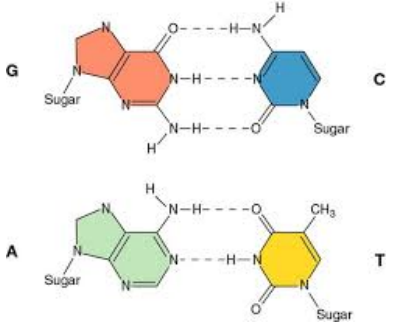
- What did you think?
- Pros?
- Cons?
- Limitations?

Walking Data Storage Units



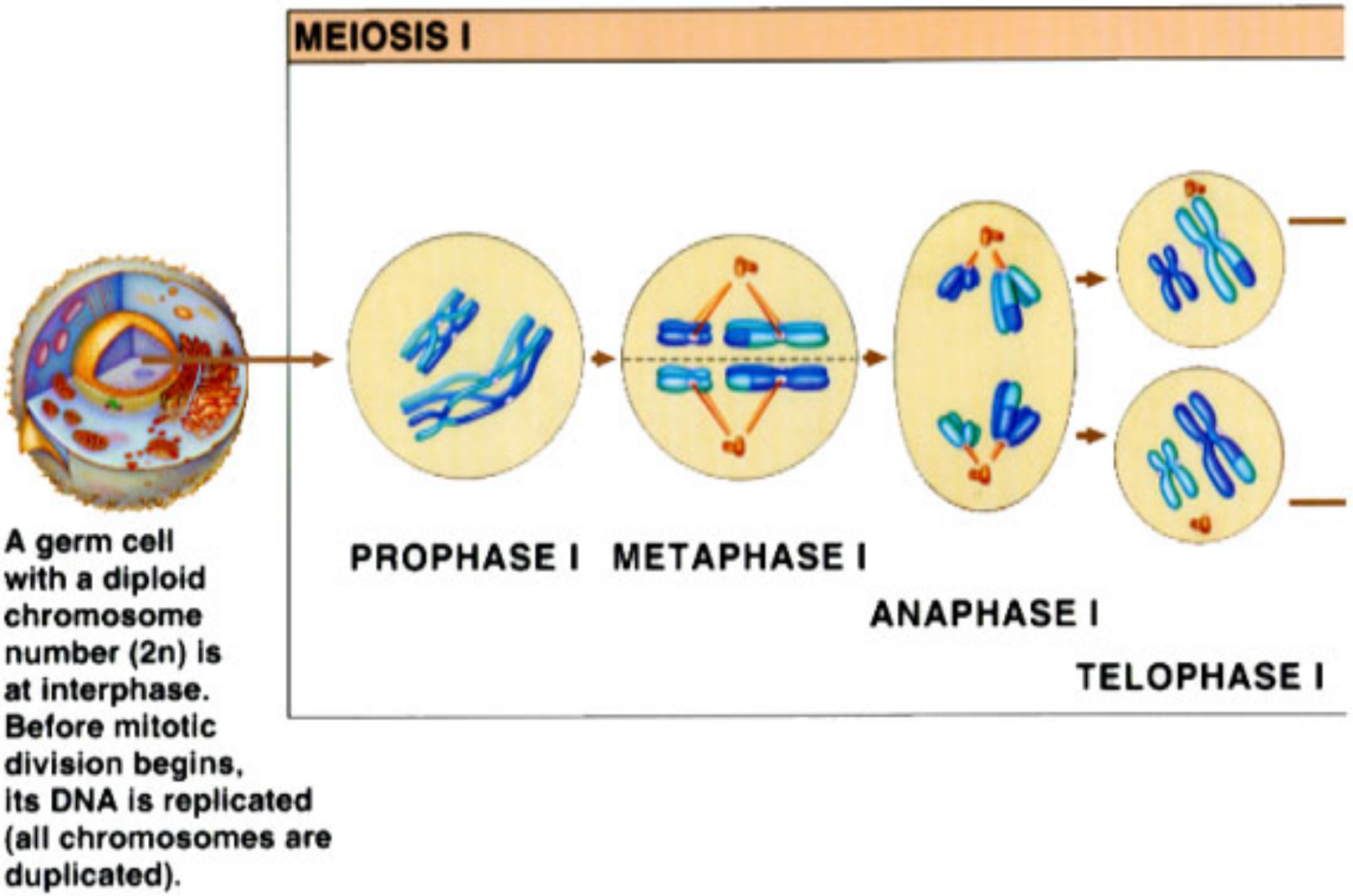
- 3 billion bp in 23 chromosomes
- Each cell has chromosomes duplicated (46)
- $6x10^9$ bp per cell
- A single byte can represent 4 DNA bp's
- $6x10^9$ bp x 1 byte/(4 bp's)= $1.5x10^9$ bytes
  - (1.5 Gb per cell)
- 1.5 Gb x 100 trillion human cells = 150 Zettabytes!
- We also have 1 quadrillion bacterial cells in/on our body . . .

What are the Possibilities?

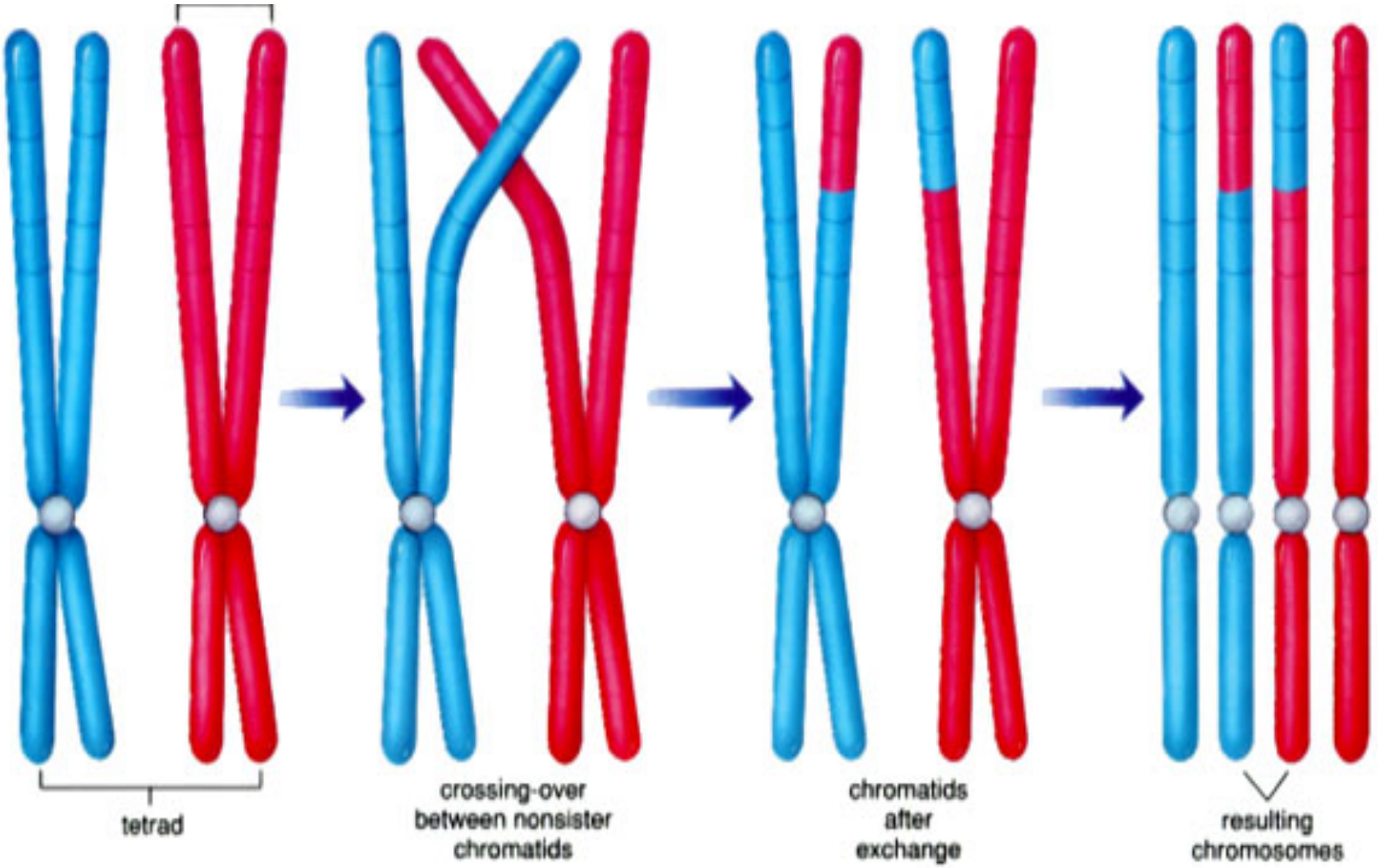


- Storage
- Can make your own databases of information
- Genetic puzzles
- Make proteins glow
- Spider silk in goats milk
- Personalized medicine
- . . .

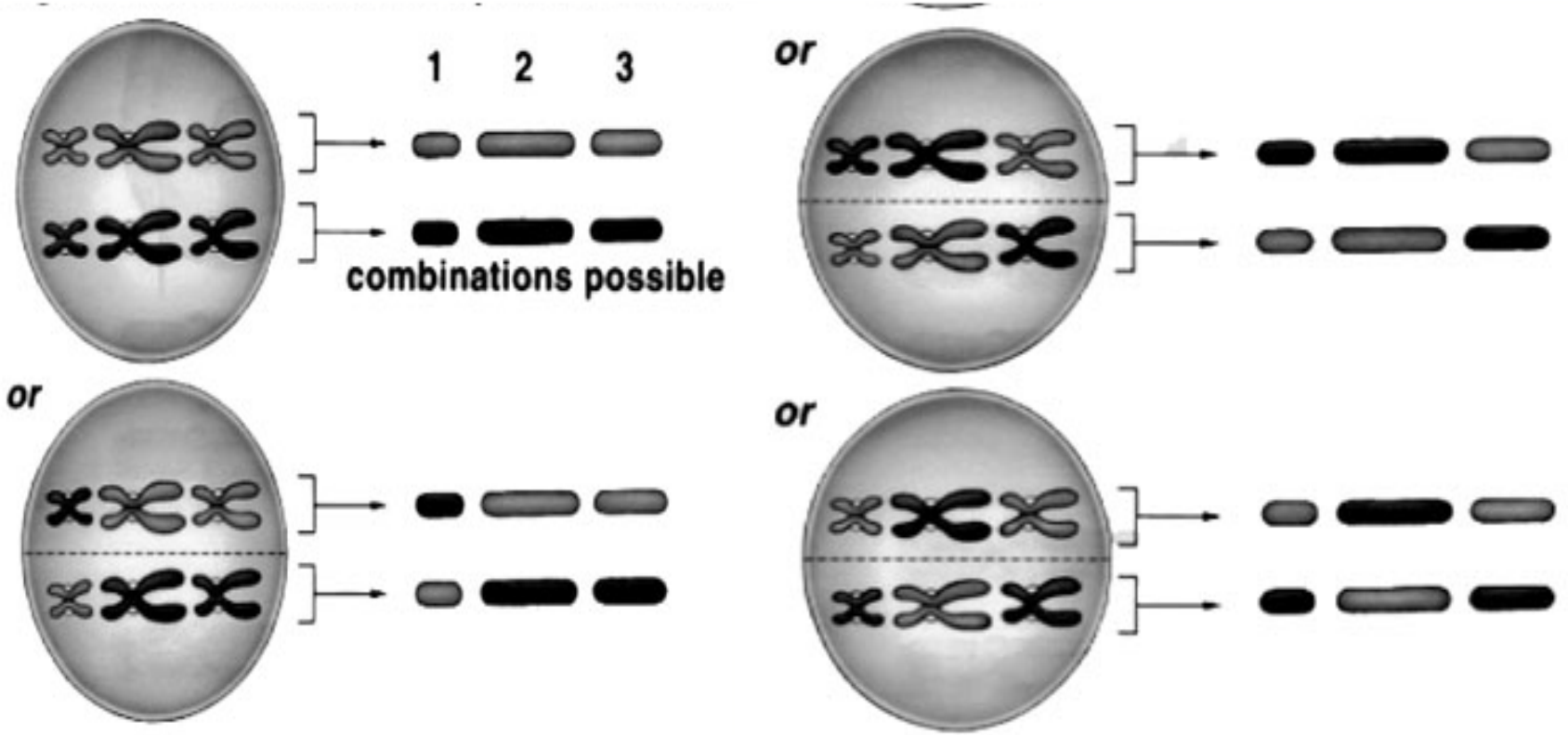
Meiosis I


MEIOSIS I
PROPHASE I
METAPHASE I
ANAPHASE I
TELOPHASE I
A germ cell with a diploid chromosome number (2n) is at interphase. Before mitotic division begins, its DNA is replicated (all chromosomes are duplicated).

Crossing Over

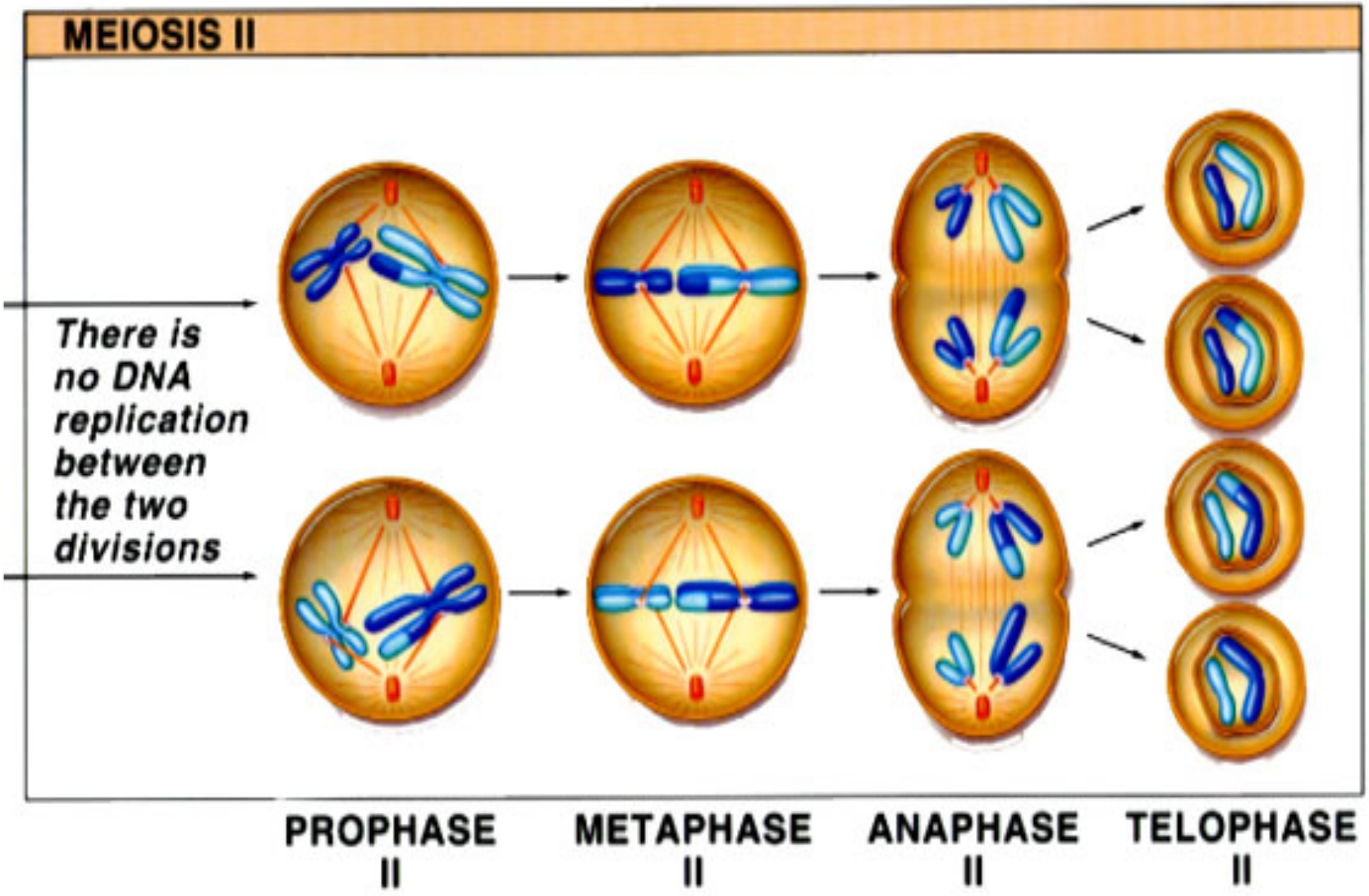


Random Alignment



Possible combinations for just 3 chromosomes

Meiosis II


MEIOSIS II
There is no DNA replication between the two divisions
PROPHASE II
METAPHASE II
ANAPHASE II
TELOPHASE II

All Together-Meiosis