Introduction to Molecular Biology

Part 1

Discovery of cells

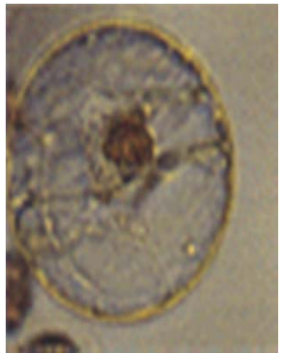
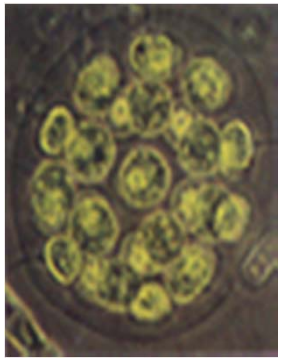
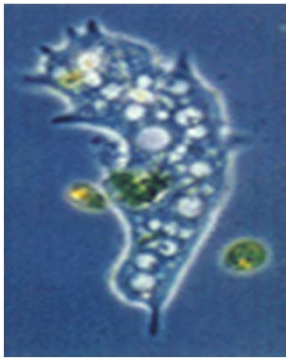
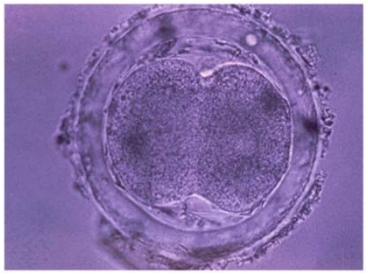
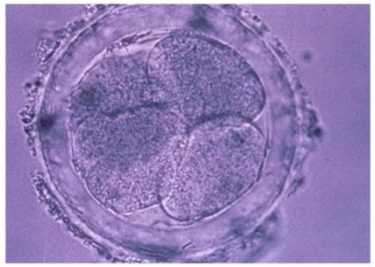
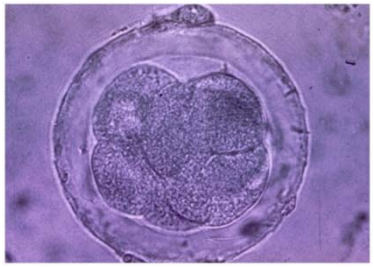
- Microscopic biology began in 1665
- Robert Hooke (1635-1703) discovered organisms are made up of cells

Cells

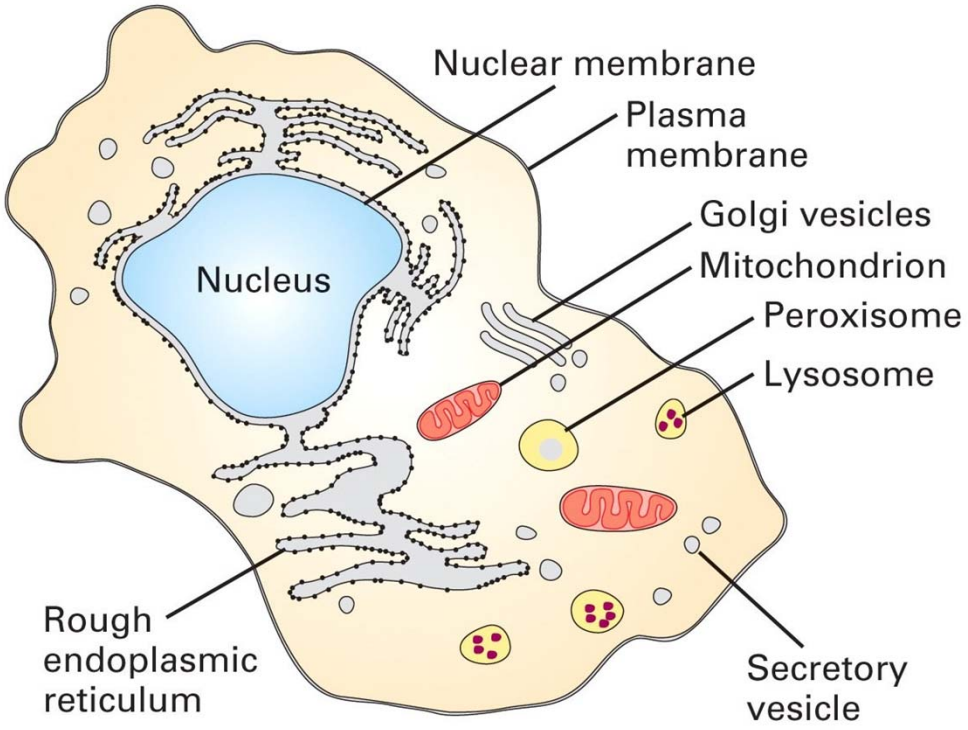
- Fundamental working units of all living systems
- Every organism is composed of one or two different types of cells
  - Prokaryotic cells
  - Eukaryotic cells
- Prokaryotes and Eukaryotes are descended from the same primitive cell
- All extant cells are the result of 3.5 billion years of evolution

Cells



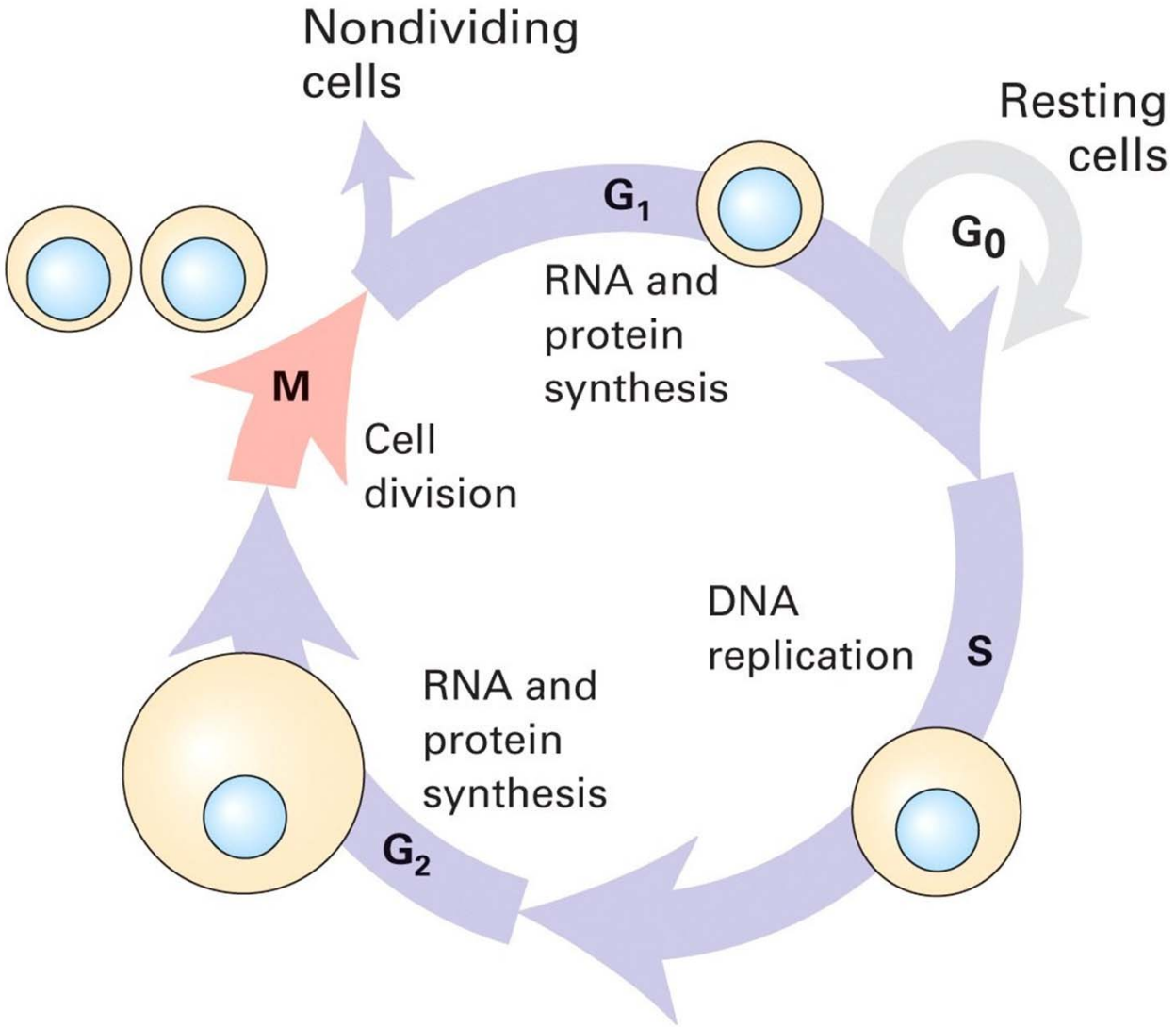







Cells



- A cell is a smallest structural unit of an organism that is capable of independent functioning
- All cells have some common features

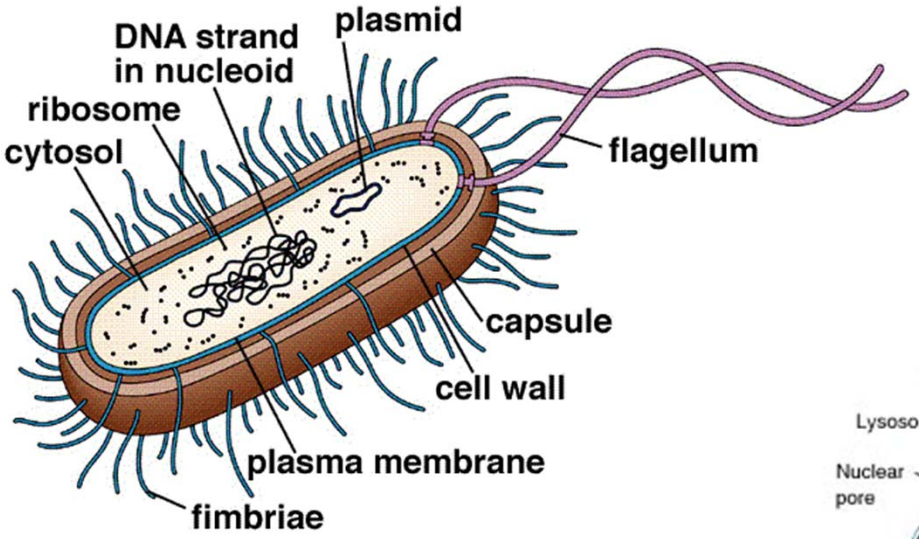
Cell Cycles: Born, Eat, Replicate, Die

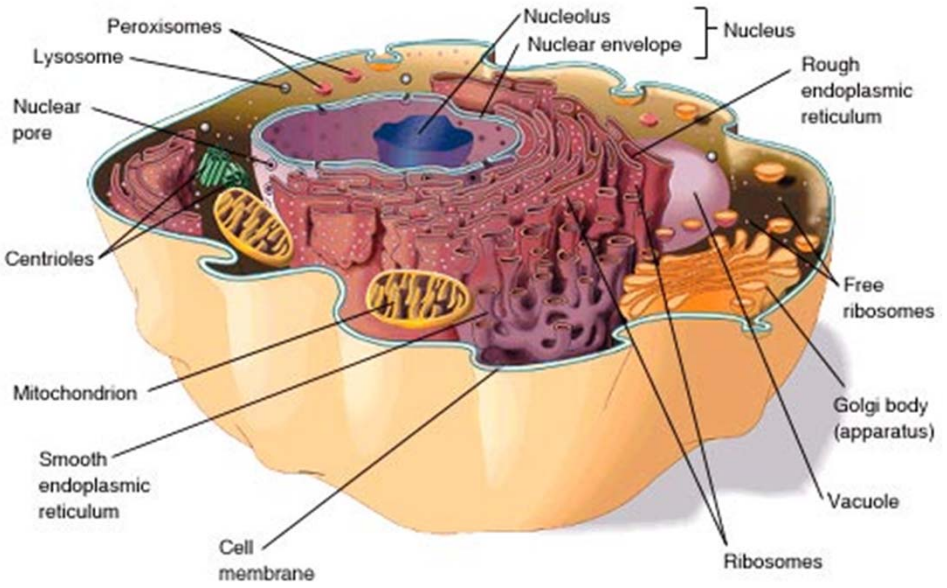


Cell: Contents…

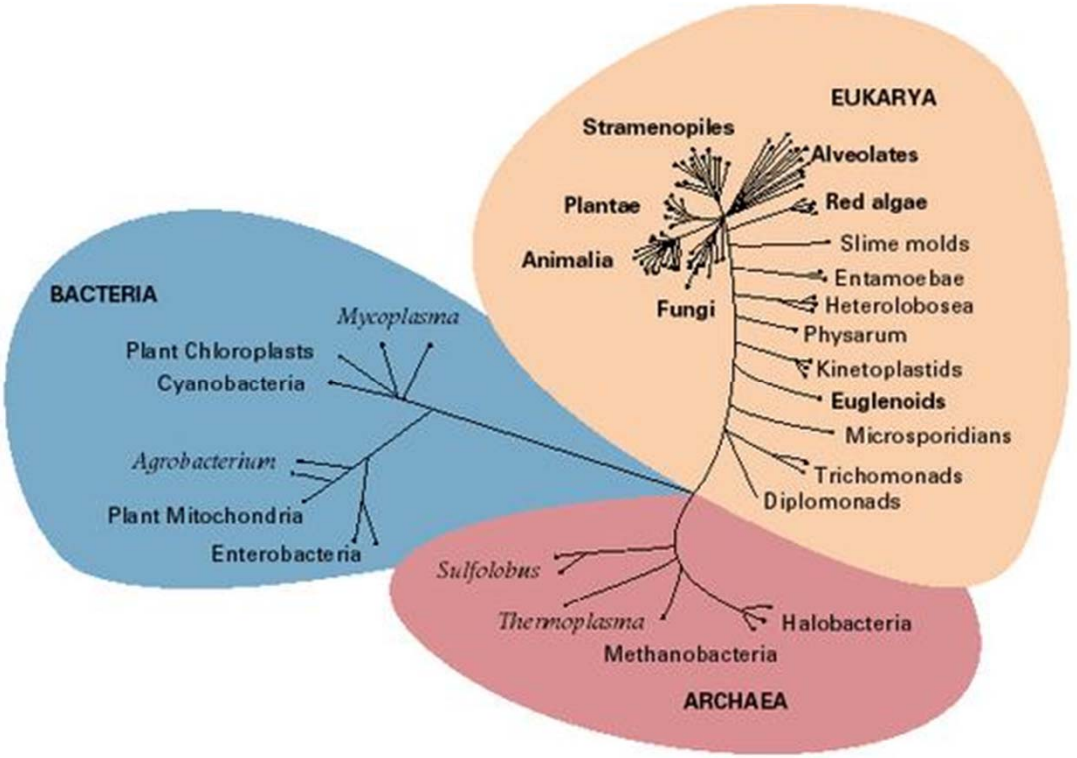
- Chemical Composition (by weight)
  - 70% water
  - 7% small molecules
    - Salts, lipids, amino acids, nucleotides
  - 23% macromolecules
    - Proteins, polysaccharides, lipids
- Biochemical (Metabolic) Pathways
- Translation of mRNA into Proteins

Cells: Prokaryotes & Eukaryotes

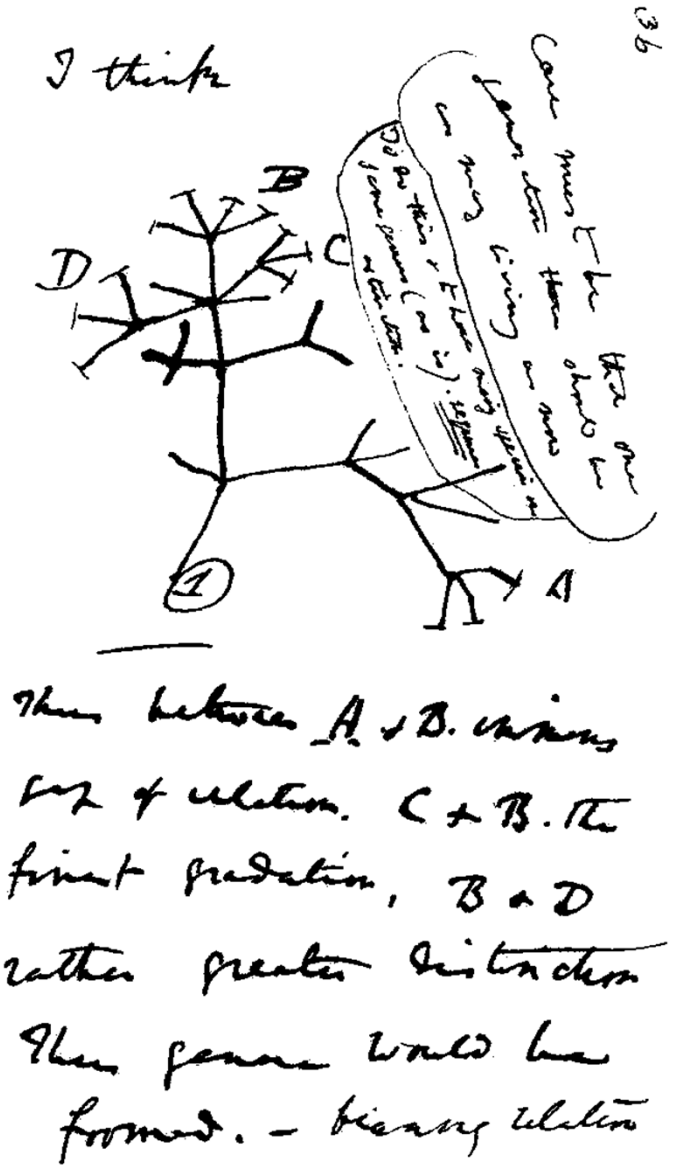




Prokaryotes & Eukaryotes



Charles Darwin: Tree of Life



"I think case must be that one generation should have as many living as now. To do this and to have as many species in same genus (as is) requires extinction . Thus between A + B the immense gap of relation. C + B the finest gradation. B+D rather greater distinction. Thus genera would be formed. Bearing relation" (next page begins) "to ancient types with several extinct forms"

-: Charles Darwin, 1837

Prokaryotes & Eukaryotes

Prokaryotes

- Single Cell
- No nucleus
- No organelles
- One piece of Circular DNA
- No mRNA post-transcriptional modification

Eukaryotes

- Single or multi cell
- Nucleus
- Organnelles
- Chromosomes
- Exons/Introns splicing

Prokaryotes & Eukaryotes Structural Differences

Prokaryotes

- Eubacteria (blue green algae) and archaebacteria
- Only one type of membrane Plasma membrane forms the boundary of the cell
- The smallest cells known are bacteria (E. Coli cell, $3 * 10^6$protein molecules, 1000-2000 polypeptide species)

Eukaryotes

- Plants, animals, Protista, fungi
- Complex systems of internal membranes forms organelle and compartments
- Volume of cell is several hundred times larger (Hela cell, $5 * 10^9$ protein molecules, 5000-10000 polypeptide species

Prokaryotes & Eukaryotes Chromosomal Differences

Prokaryotes

- The genome of E. Coli contains $4 * 10^6$base pairs
- >90% of DNA encode protein
- Lacks a membrane bound nucleus
  Circular DNA
  and supercoiled domain
- Histones are unknown

Eukatryotes

- The genome of yeast contains $1.35 * 10^7$base pairs
- A small fraction of the DNA encodes protein (many repeats of non-coding sequences)
- All chromosomes are contained in a membrane bound nucleus (DNA is divided between one or more chromosomes)
- A set of five histones
  DNA packaging and gene expression regulation

Signaling Pathways Control Gene Activity

- Instead of having brains, cells make decisions through complex networks of chemical reactions, called pathways
  - Synthesize new materials
  - Break other materials down for spare parts
  - Signal to eat or die

Cells: Information & Machinery

- Cells store all information to replicate
  - Human genome is around $3 * 10^9$base pairs long
  - Almost every cell in a human body contains same set of genes
  - But not all genes are used/expressed by all cells
- Machinery
  - Collect and manufacture components
  - Carry out replication
  - Kick-start its new offspring

  (A cell is like a car factory!)

The Human Genome Project

- **1986** Leroy Hood: Developed automated sequencing mechanism
- **1986** Human Genome Initiative announced
- **1990** The 15 year Human Genome project is launched by congress
- **1995** John Craig Venter: First bactierial genomes sequenced
- **1997** E. Coli sequenced
- **1996** First eukaryotic genome-yeast-sequenced

The Human Genome Project

- **2000** J. Craig Ventnor (Celera) and Francis Collins (IHGSC) announce sequencing the complete human genome
- **2001** International Human Genome Sequencing: First draft of the sequence of the human genome published
- False start…until **2003**

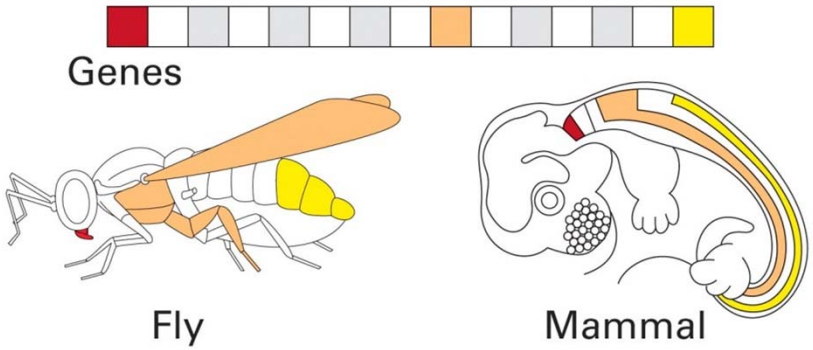
Organization of Life: Overview

- Nucleus = library
- Chromosomes = bookshelves
- Genes = books
- Almost every cell in an organism contains the same libraries and the same set of books
- Books represent all the information (DNA) that every cell in the body needs so it can grow and carry out its various functions

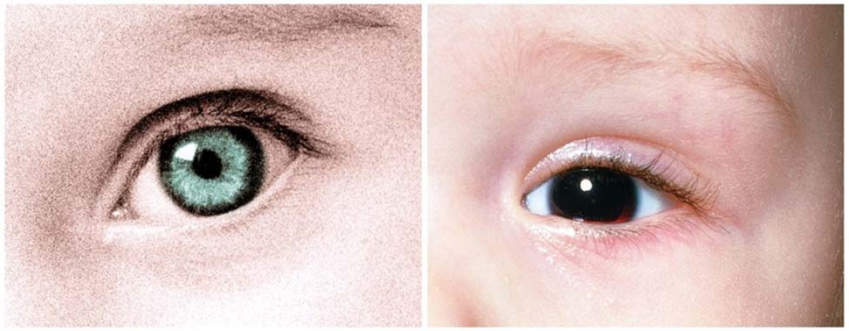
Terminology

- **Genome:** An organism's genetic material
- **Gene:** Discrete unit of hereditary information located on the chromosomes and consisting of DNA
- **Genotype:** The genetic makeup of an organism
- **Phenotype:** The physical expressed traits of an organism
- **Nucleic Acid:** Biological molecules (DNA & RNA) that allow organisms to reproduce

Genotype & Phenotype



- Genes are inherited and are expressed
  - **genotype** (genetic makeup)
  - **phenotype** (physical expression)



- On the left, is the eye's phenotypes of green and black eye genes.

Genotype & Phenotype

- Genes are like recipes (genotype)
- Think of a recipe for a cake…
- Only partly guarantee the end result (phenotype)
- Environment plays a crucial role

Terminology…

- The **genome** is an organism's complete DNA.
  A bacteria contains about 600,000 DNA base pairs
  Human and mouse genomes have 3 billion base pairs.
- Human genome has 24 distinct chromosomes.
  Each chromosome contains many **genes.**
- **Genes** are the basic functional units of heredity.
  Specific sequences of DNA bases that encode instructions on how to make **proteins.**
- **Proteins** make up the cellular structure.
  Large complex molecules made up of smaller subunits called **amino acids.**

Life: 3 Critical Molecules

- **DNA**

  Holds information on how cell works

- **RNA**

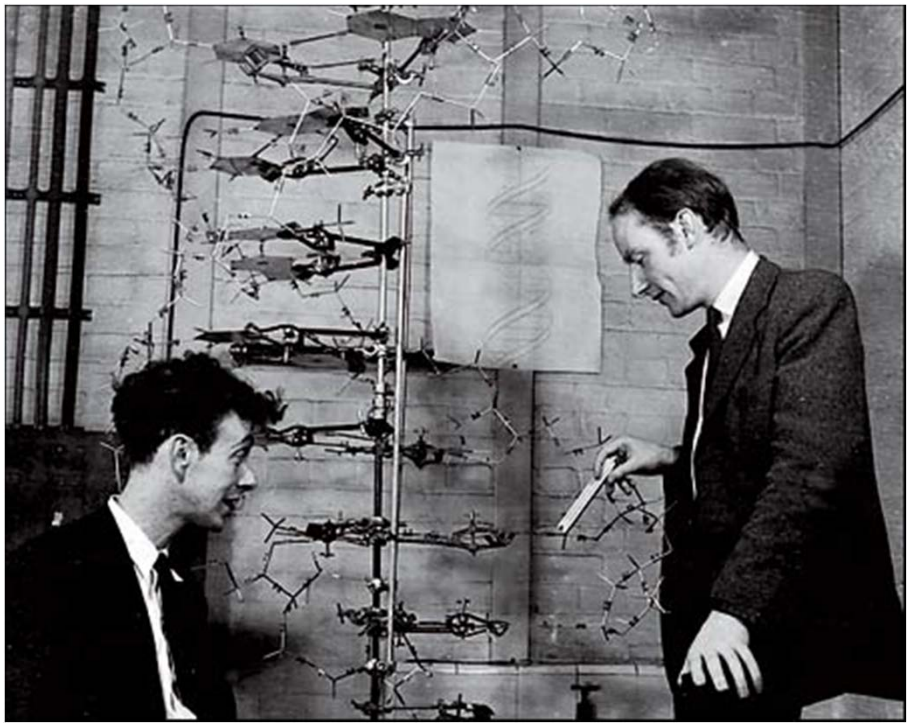
  Acts to transfer short pieces of information to different parts of a cell
  Provides templates to synthesize proteins

- **Proteins**

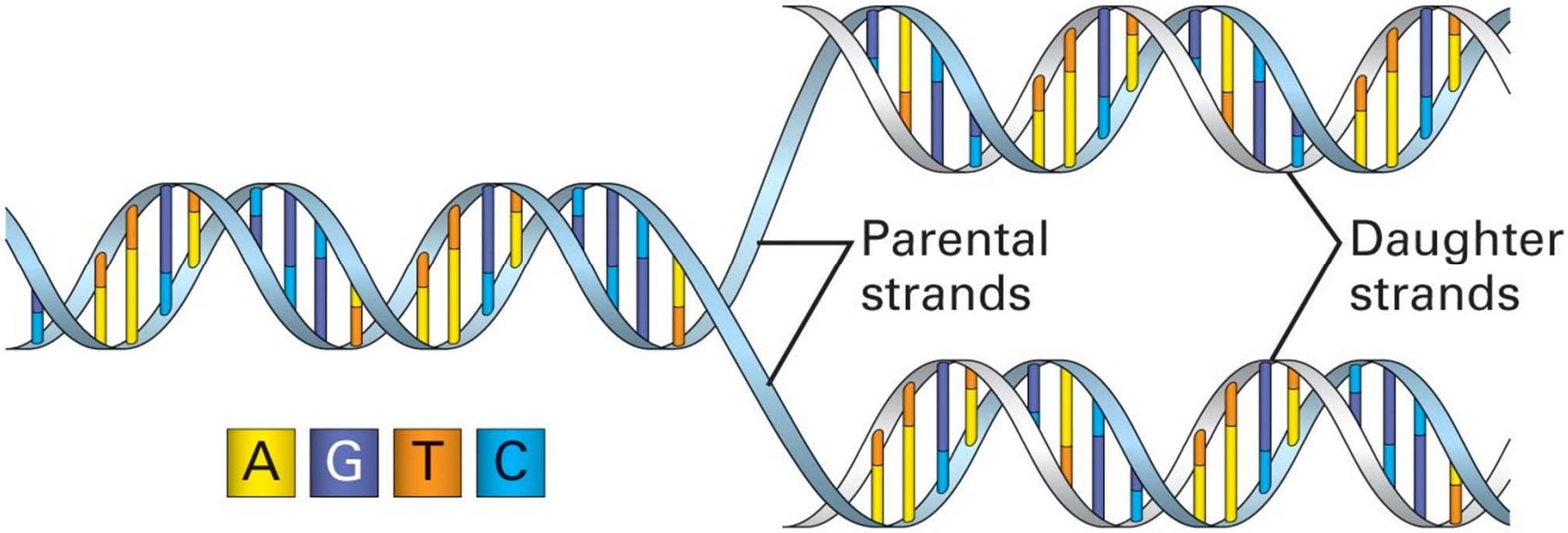
  Form enzymes that send signals to other cells and regulate gene activity
  Form body's major components (e.g. hair, skin, etc.)

DNA: The Double Helix

- **1952-1953** James D. Watson and Francis H. C. Crick deduced the double helical structure of DNA

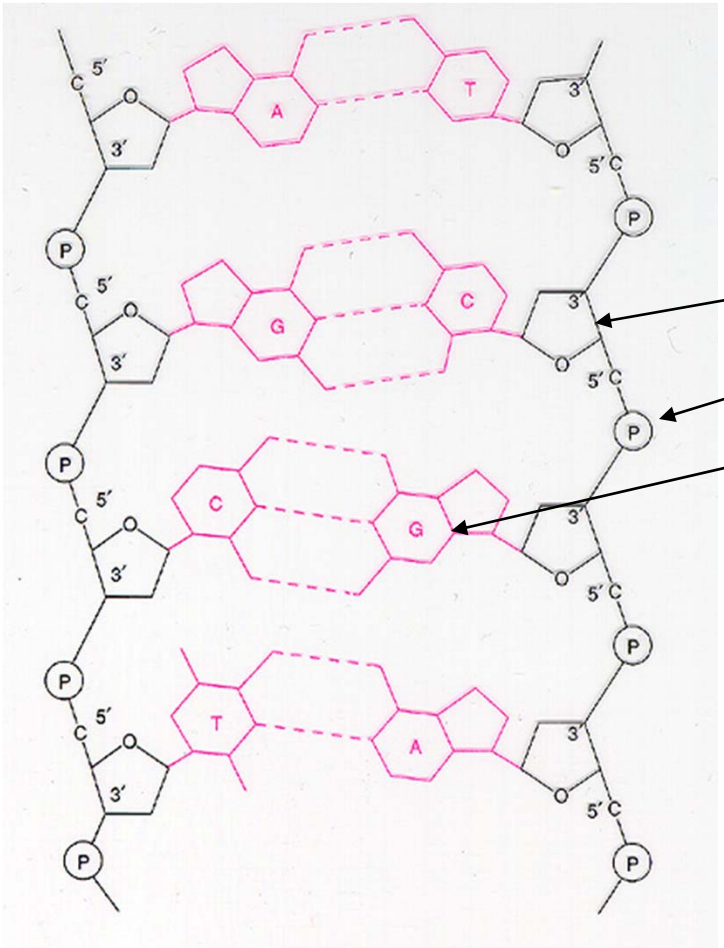


DNA: The Code of Life



- The structure and the four genomic letters code for all living organisms
- Adenine, Guanine, Thymine, and Cytosine which pair A-T and C-G on complimentary strands.

DNA: The Code of Life



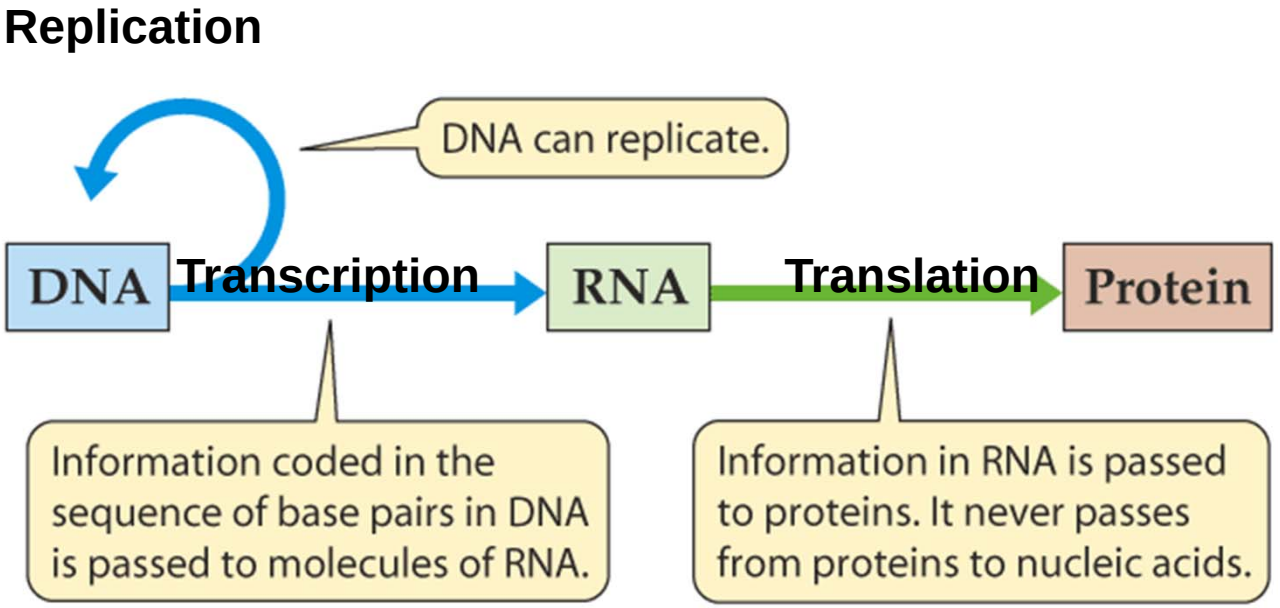
- DNA has a double helix structure which composed of
  - sugar molecule
  - phosphate group
  - and a base (A,C,G,T)

- DNA always reads from 5’ end to 3’ end for transcription replication

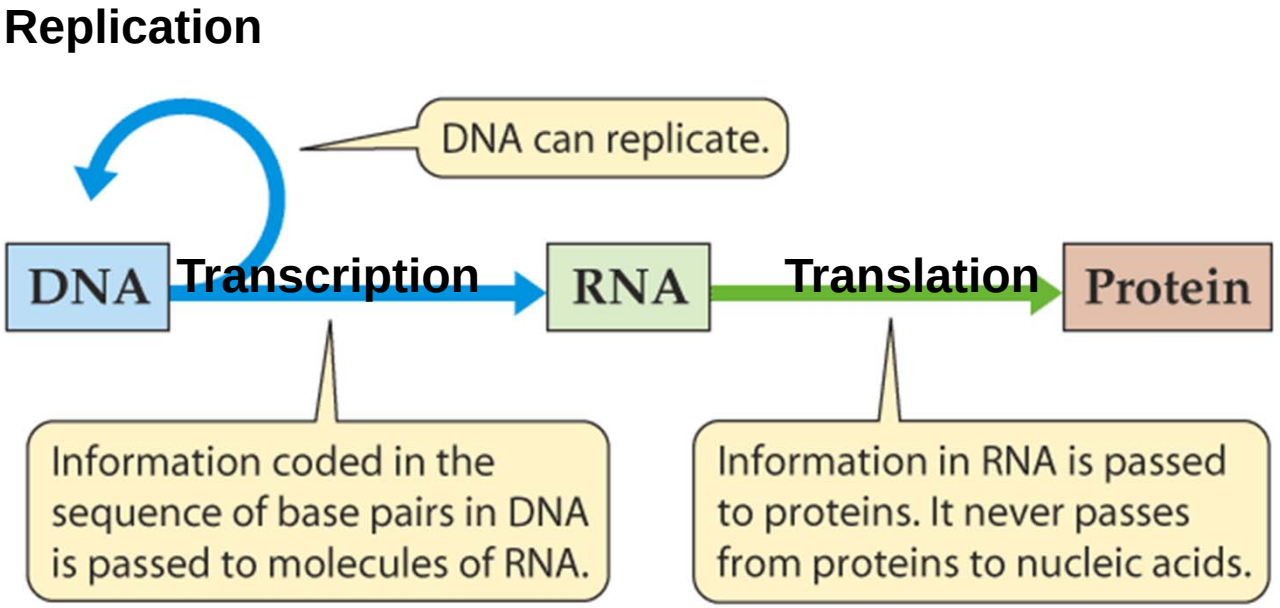
  5’ ATTTAGGCC 3’

  3’ TAAATCCGG 5’

DNA &RNA: Flow of Information

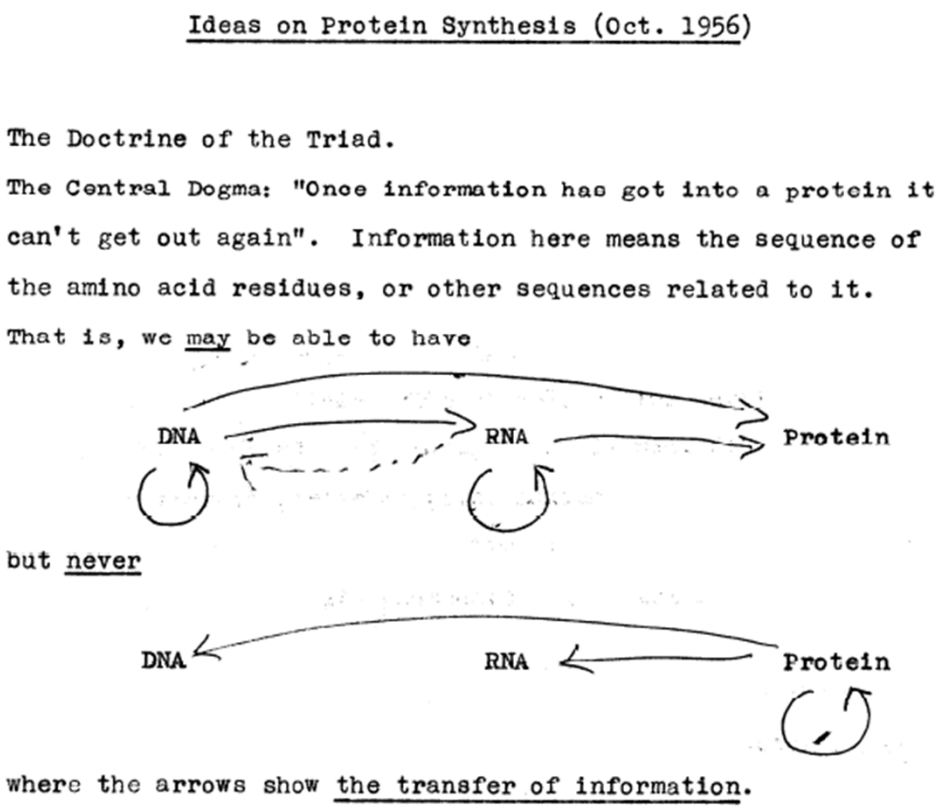


DNA & RNA: Flow of Information



aka "The Central Dogma"!!

Francis Crick



Ideas on Protein Synthesis (Oct. 1956)

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we may be able to have

DNA → RNA → Protein

but never

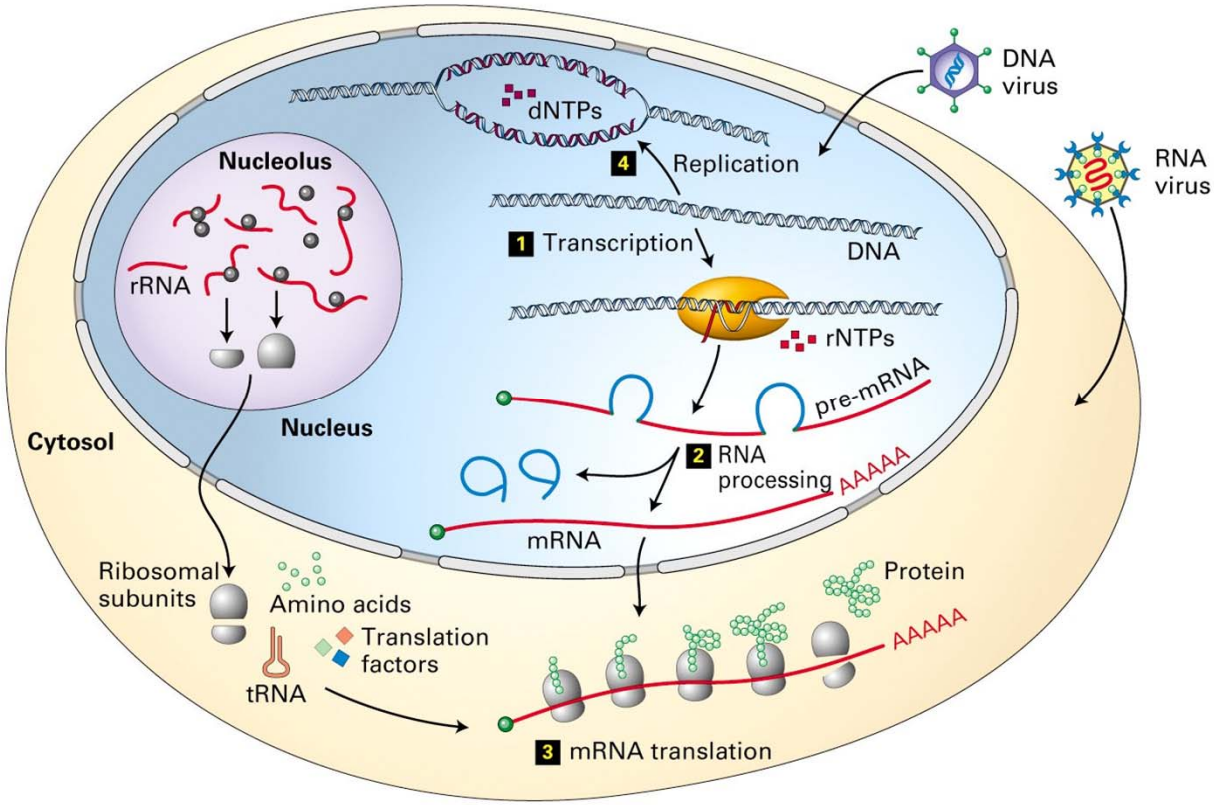
DNA ← RNA ← Protein

where the arrows show the transfer of information.

"The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein of nucleic acid."

-: Francis Crick, *Central Dogma of Molecular Biology*, Nature, Volume 227, August 1970.

DNA to RNA to Protein

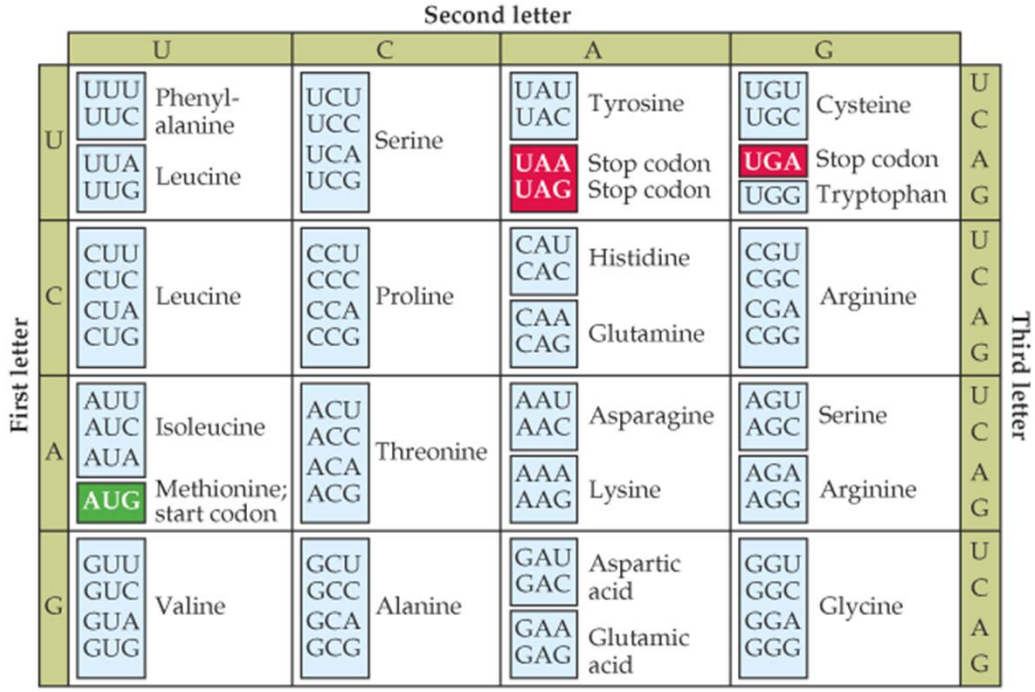


A gene is expressed in two steps

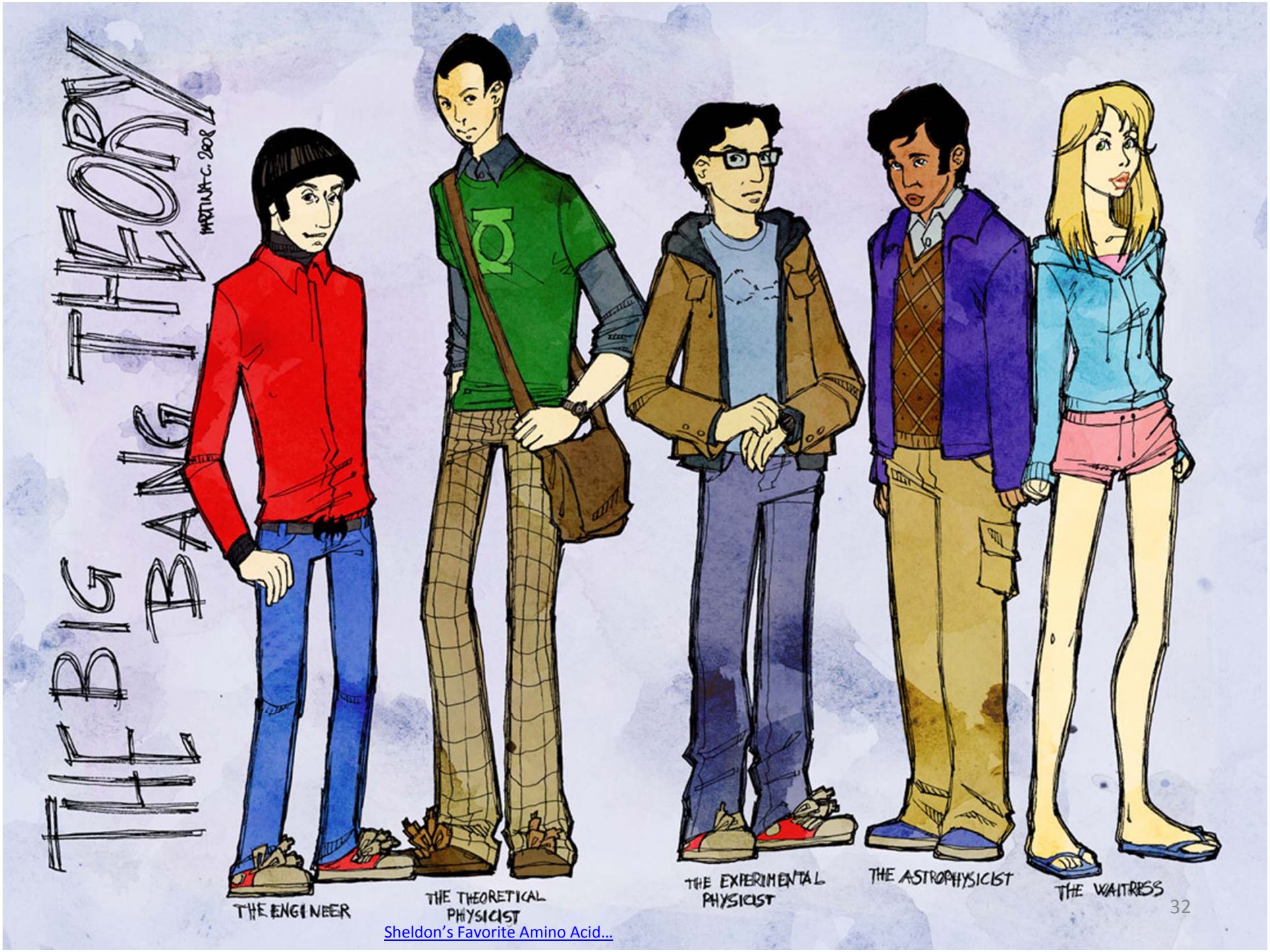
1. **Transcription:** RNA Synthesis
2. **Translation:** Protein Synthesis

The Code Book

- DNA, RNA, and Proteins are examples of strings written in either the four-letter nucleotide of DNA and RNA (A C G T/U)
- or the twenty-letter amino acid of proteins. Each amino acid is coded by 3 nucleotides called codons

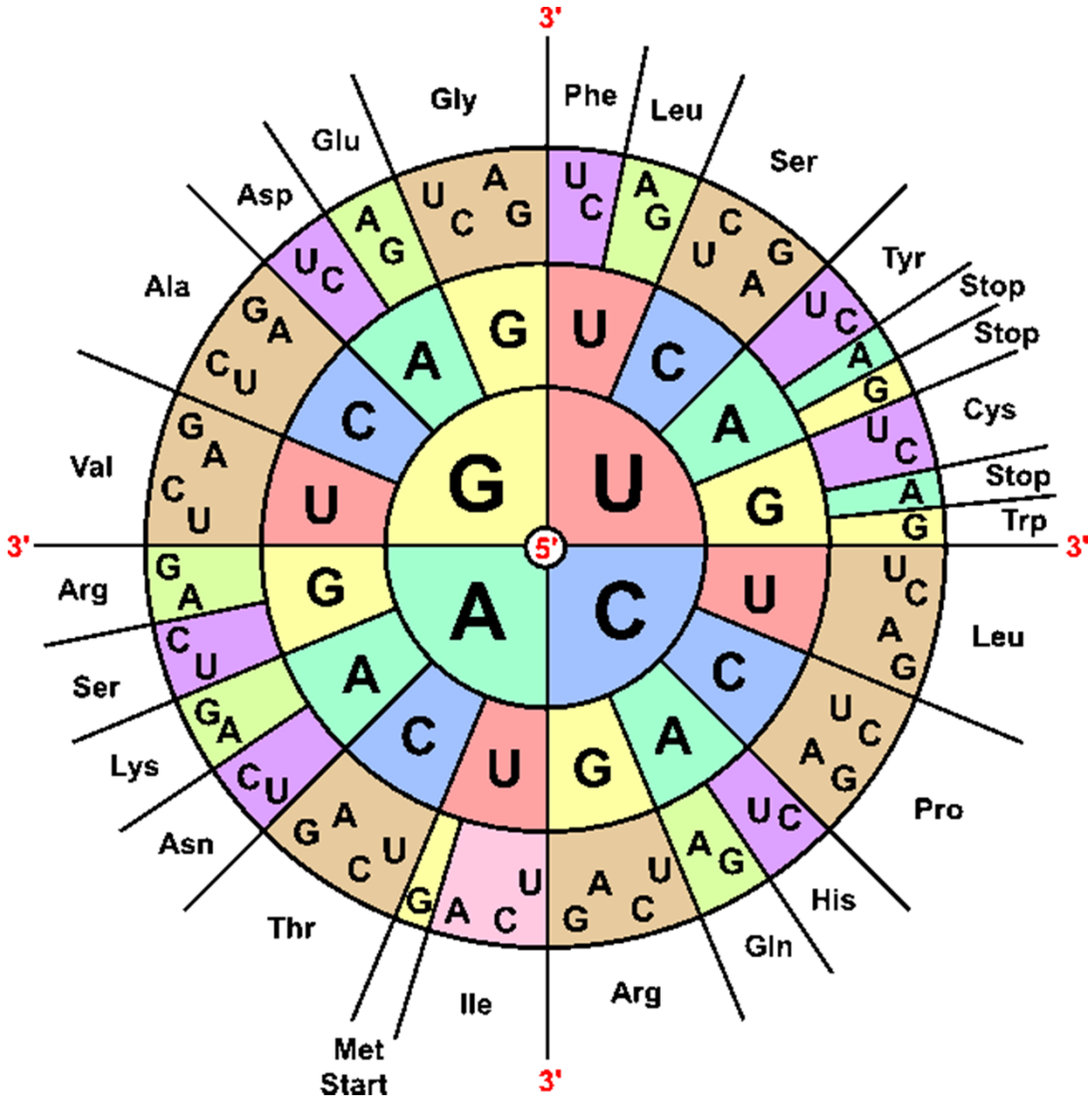


First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phenyl-alanine	UCU Serine	UAU Tyrosine	UGU Cysteine	U
	UUC Phenyl-alanine	UCC Serine	UAC Tyrosine	UGC Cysteine	C
	UUA Leucine	UCA Serine	UAA Stop codon	UGA Stop codon	A
	UUG Leucine	UCG Serine	UAG Stop codon	UGG Tryptophan	G
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine	U
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine	C
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine	A
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine	G
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	U
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine	C
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine	A
	AUG Methionine; start codon	ACG Threonine	AAG Lysine	AGG Arginine	G
G	GUU Valine	GCU Alanine	GAU Aspartic acid	GGU Glycine	U
	GUC Valine	GCC Alanine	GAC Aspartic acid	GGC Glycine	C
	GUA Valine	GCA Alanine	GAA Glutamic acid	GGA Glycine	A
	GUG Valine	GCG Alanine	GAG Glutamic acid	GGG Glycine	G



[Sheldon's Favorite Amino Acid…]

The Code Book



DNA & RNA

- DNA = Deoxyribonucleic acid
- RNA = Ribonucleic acid
- They are almost the same…
- There is no T base in RNA
- A similar base U takes its place
- An oxygen atom is added to the sugar component of RNA

References

- Neil C. Jones and Pavel A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT Press 2004.
- Adapted from slides posted at the web site of the above book.
- Francis Crick, *Central Dogma of Molecular Biology*, Nature, Volume 227, August 1970.
- Luciano Floridi, *Information: A Very Short Introduction*, Oxford 2010.