

9.3a Overview of Basic Molecular Biology

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Overview of Basic Molecular Biology

The cell, DNA, and RNA

Central Dogma

Protein structure and function

Mendelian heredity

Human genome

Genetic variation

Epigenetics

Technologies advancing bioinformatics

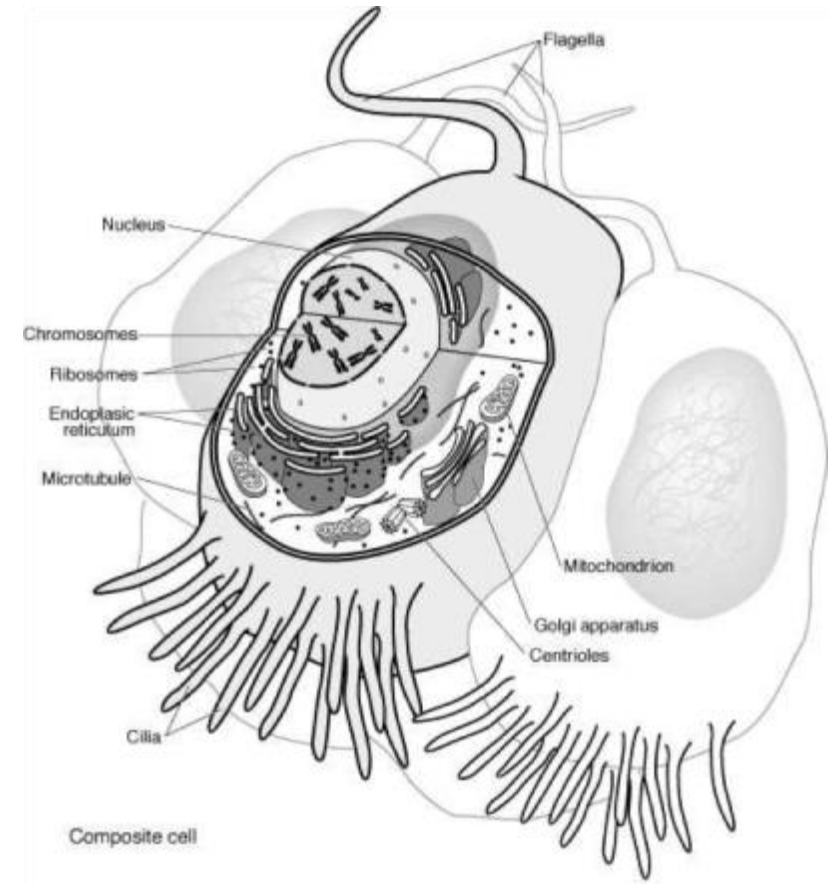
Basic molecular biology

All organisms composed of cells

Plant, fungus, and animal cells have nucleus
whereas bacteria and viruses do not

In multi-cellular organisms, cells differentiate to
perform specific functions, e.g.,

- nerve, muscle, bone, blood, etc.

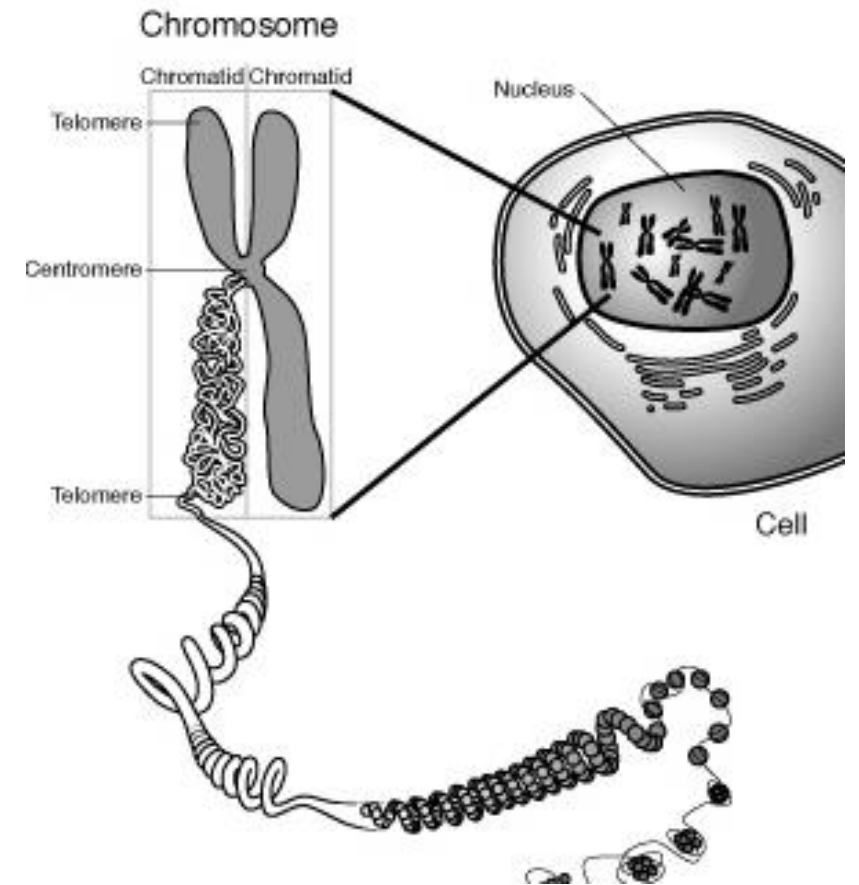


Cell function regulated by nucleus

Nucleus contains chromosomes, which are composed of deoxyribonucleic acid (DNA) and contain genes

Humans have 23 pairs of chromosomes

- 44 autosomes and two sex chromosomes, X and Y, with XX for females and XY for males



“Location” on chromosomes

Designated by

- Chromosome number – sorted by largest to smallest
- Arm – from centromere, p (petit) or q (queue)
- Band – based on order out from centromere, with refinements made by adding digits

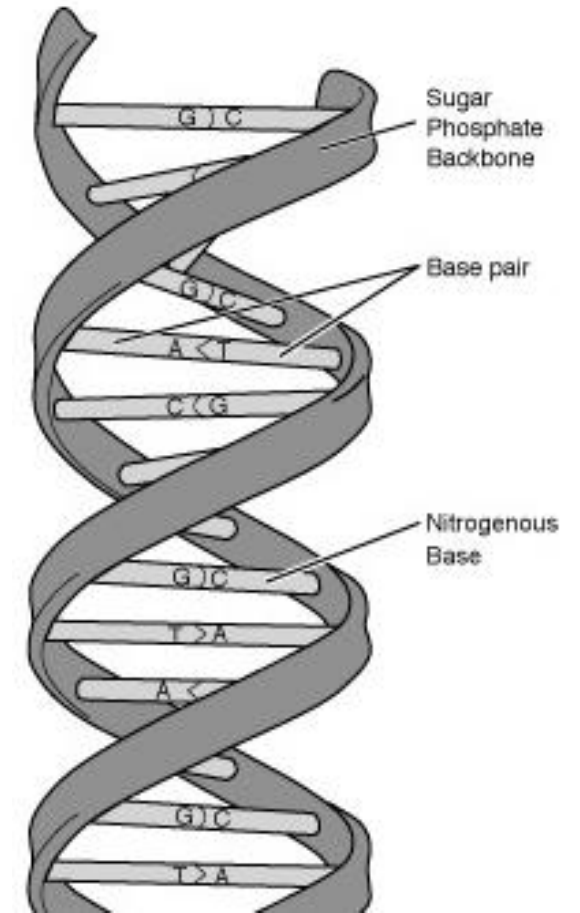
DNA

Double-stranded DNA looks like a ladder that has been twisted

- Sides of ladder are deoxyribose sugar and phosphate backbone
- Rungs of ladder are made of pairs of nitrogenous bases

All cells in an organism have same DNA

- Except those having mutations
- Differentiation is regulated by expression

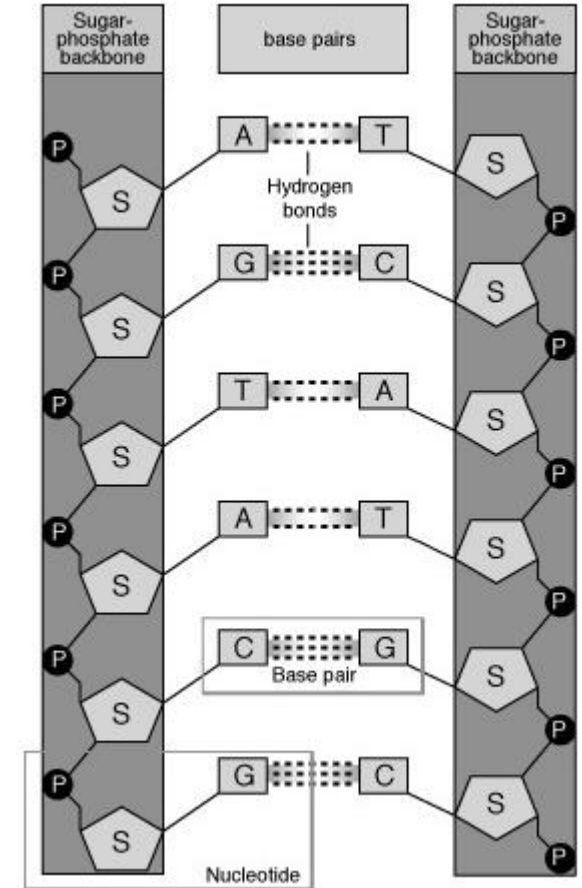


The “information” of DNA

There are four “bases”

- A = adenine
- C = cytosine
- T = thymine
 - In RNA, the base U = uracil used instead of T
- G = guanine

A always pairs with T (or U in RNA) and G always pairs with C

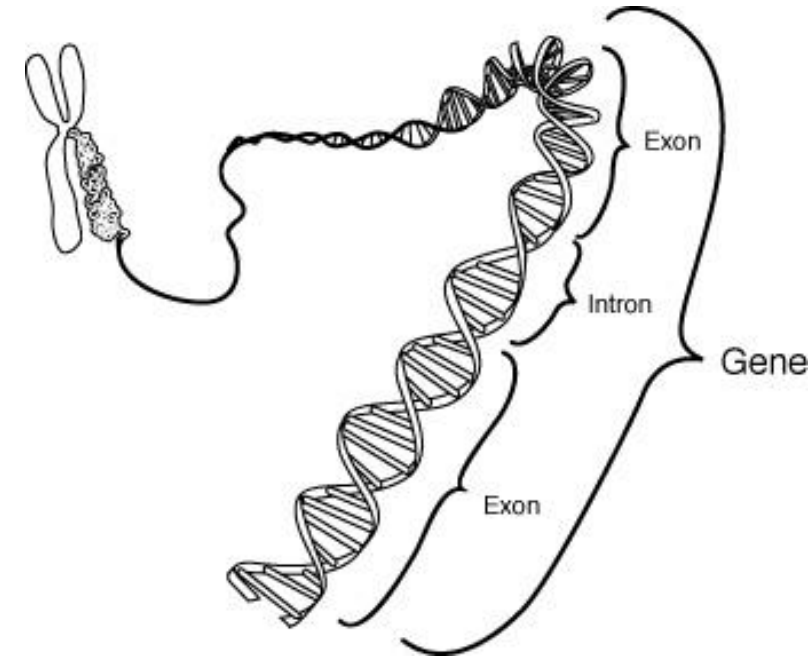


Genes

Sequences of DNA make up genes, which are functional and physical units of heredity passed from parent to offspring

Contain information for making specific proteins or other products

Exons code for genes, introns are non-coding segments in between



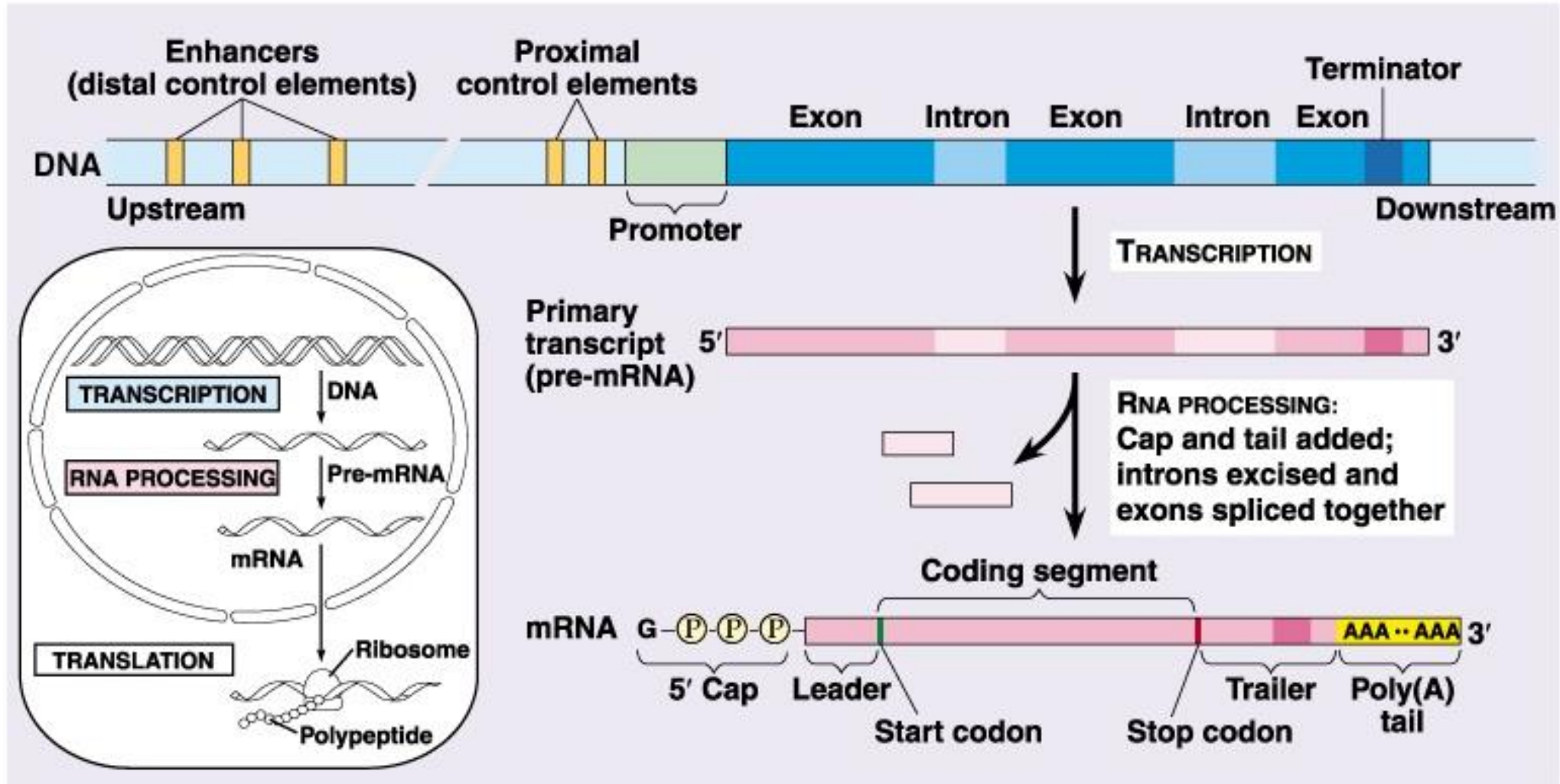
Gene expression

Gene expression is a highly specific process in which a gene is switched on at a certain time and begins production of its protein

A Liver	B Lung	C Muscle	D Spleen	Gene index	Gene description
0.14014	0.00011	0.00000	0.06852	Hs.75442	albumin (ALB)
●	●	●	●		
A > B	B < A	C < A	D < A		
A > C	B < D	C < D	D > B		
A > D			D > C		
0.00057	0.00000	0.00000	0.14466	Hs.272812	hemoglobin, gamma G (HBG2)
●	●	●	●		
A < D	B < D	C < D	D > A		
			D > B		
			D > C		
0.00000	0.00000	0.04179	0.00000	Hs.931	myosin, heavy polypeptide 2, skeletal muscle, adult (MYH2)
●	●	●	●		
A < C	B < C	C > A	D < C		
		C > B			
		C > D			



How the genotype becomes the phenotype



Transcription from DNA to mRNA

Starts from promoter region of DNA

- A protein called a transcription factor binds

Region from promoter to terminator is open reading frame (ORF)

Both exons and introns transcribed but mRNA from introns spliced out before leaving nucleus

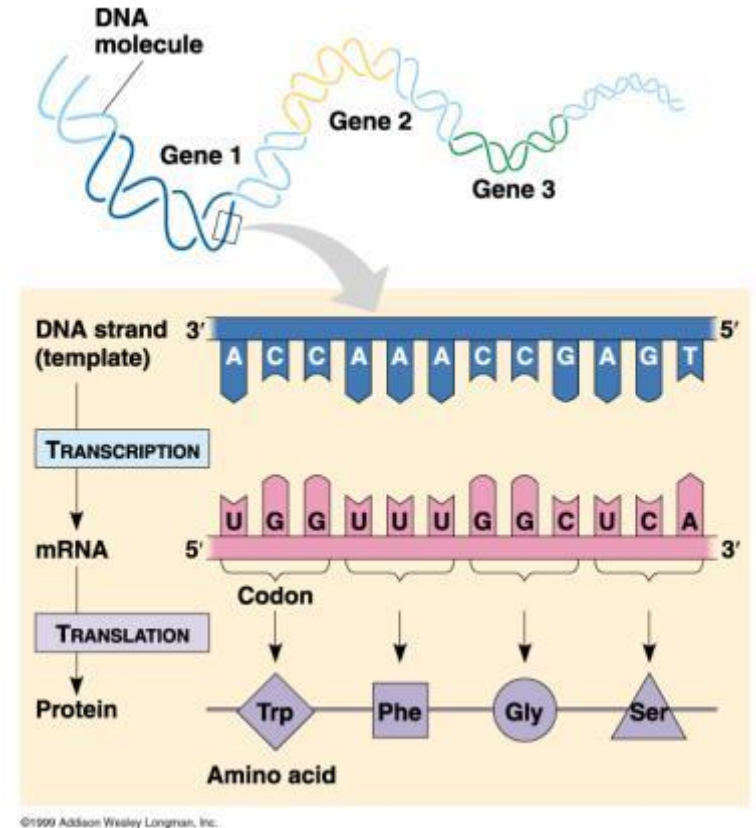
- But some genes have alternative splicing, i.e., different exons included in transcript

Translation from mRNA to protein

A codon is a set of three bases in a DNA or RNA sequence that specify a single amino acid

Each triplet codon codes for a specific amino acid

- Amino acids are also called peptides
- A chain of amino acids is called a polypeptide



Translation from mRNA to protein (cont.)

Each codon specifies which amino acid is needed at every step in making a protein

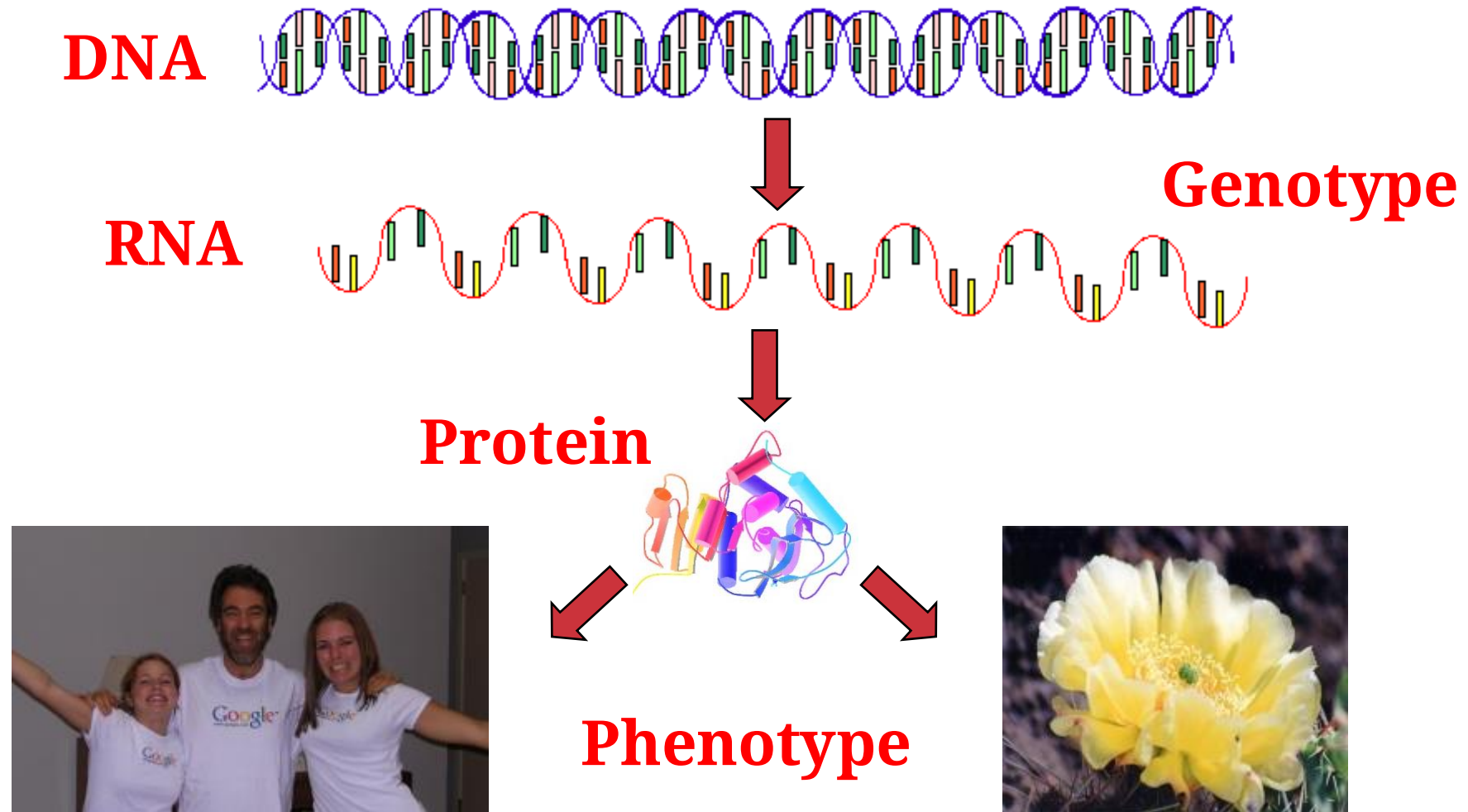
- Three codons are stop codons that indicate the end of the amino acid chain
- Start codon also codes for Met
 - Met usually removed when it is a start codon

		SECOND BASE				
		U	C	A	G	
FIRST BASE (5' end)	U	UUU Phe UUC UUA Leu UUG	UCU UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
	C	CUU CUC CUA CUG	CCU CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC CGA CGG	U C A G
	A	AUU AUC Ile AUA AUG Met or start	ACU ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU GUC GUA GUG	GCU GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC GGA GGG	U C A G

©1998 Addison Wesley Longman, Inc.



After understood, whole process named the “Central Dogma”



Alas, if only the central dogma were so simple...

New research has blown it out of the water (next segment) ... but a preview here

There is no simple DNA to gene correlation, e.g.,

- Proteins can arise from different “reading frames,” some of which overlap with each other
- Others emerge from “alternative splicing” of mRNA
- Regulator portions can be distant from genes

Many proteins undergo “post-translational modification,” e.g., modified after translation from DNA

Proteins and their structure

The end product of transcription and translation

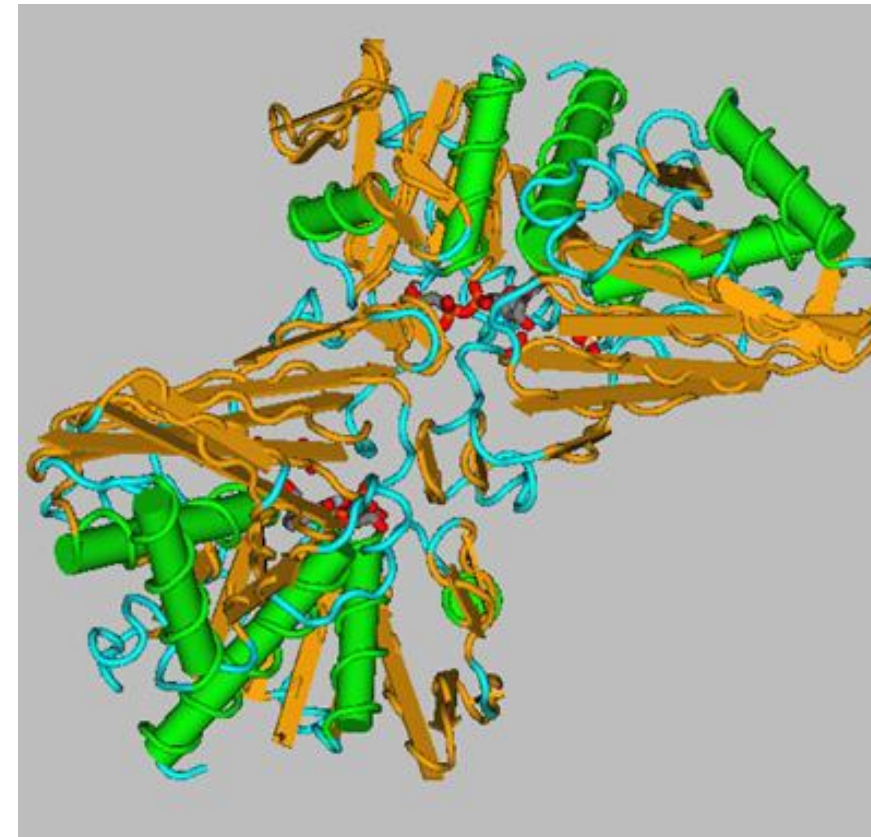
Large complex molecules made up of one or more chains of amino acids

There is not necessarily a one-to-one correspondence between genes and proteins

- One protein may arise from several genes
- One gene may code for more than one (or more than part of one) protein

Polypeptides fold into three-dimensional (3-D) structures

Proteins may be made up of more than one sub-units of polypeptide structures



Protein functions are many

Structural – e.g., collagen is structure of tissues

Storage – e.g., albumin stores calcium

Transport – e.g., hemoglobin transports oxygen

Hormonal – e.g., insulin regulates blood glucose

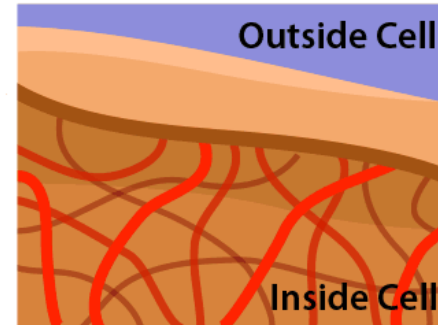
Receptor – e.g., insulin receptor binds insulin

Contractile – e.g., myosin contracts muscle

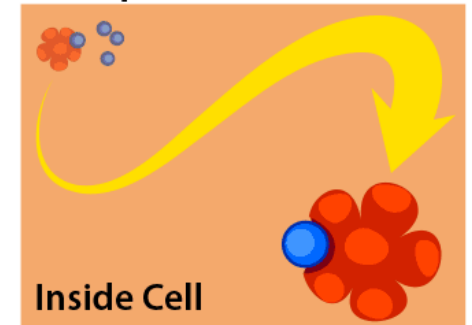
Defensive – e.g., antibodies fight “intruders”

Enzymatic – e.g., polymerase catalyzes reactions

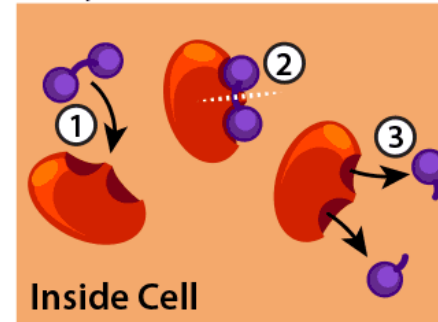
Structure Protein



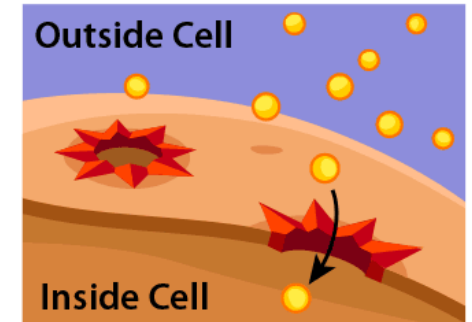
Transport Protein



Enzymes



Channels



Mendelian heredity

Since chromosomes are paired, an organism has two genes for most traits

Mendel discovered recessive and dominant heredity patterns, but many are more complex

- In dominant inheritance, just one gene leads to trait
- In recessive inheritance, must have two genes for trait to occur

Example

- Red blood cells (RBCs) have A or B protein molecules on their surfaces
 - When no proteins are present, type is O
- Each individual inherits one gene from parent, with inheritance pattern as shown in table
- Other blood surface proteins as well, e.g., Rh+/-

Parents	A	B	O
A	A	AB	A
B	AB	B	B
O	A	B	O

Many traits are determined by multiple genes