

A super quick introduction to molecular biology and genomics

Héctor Corrada Bravo

Dept. of Computer Science

Center for Bioinformatics and Computational Biology

University of Maryland

University of Maryland, Fall 2015

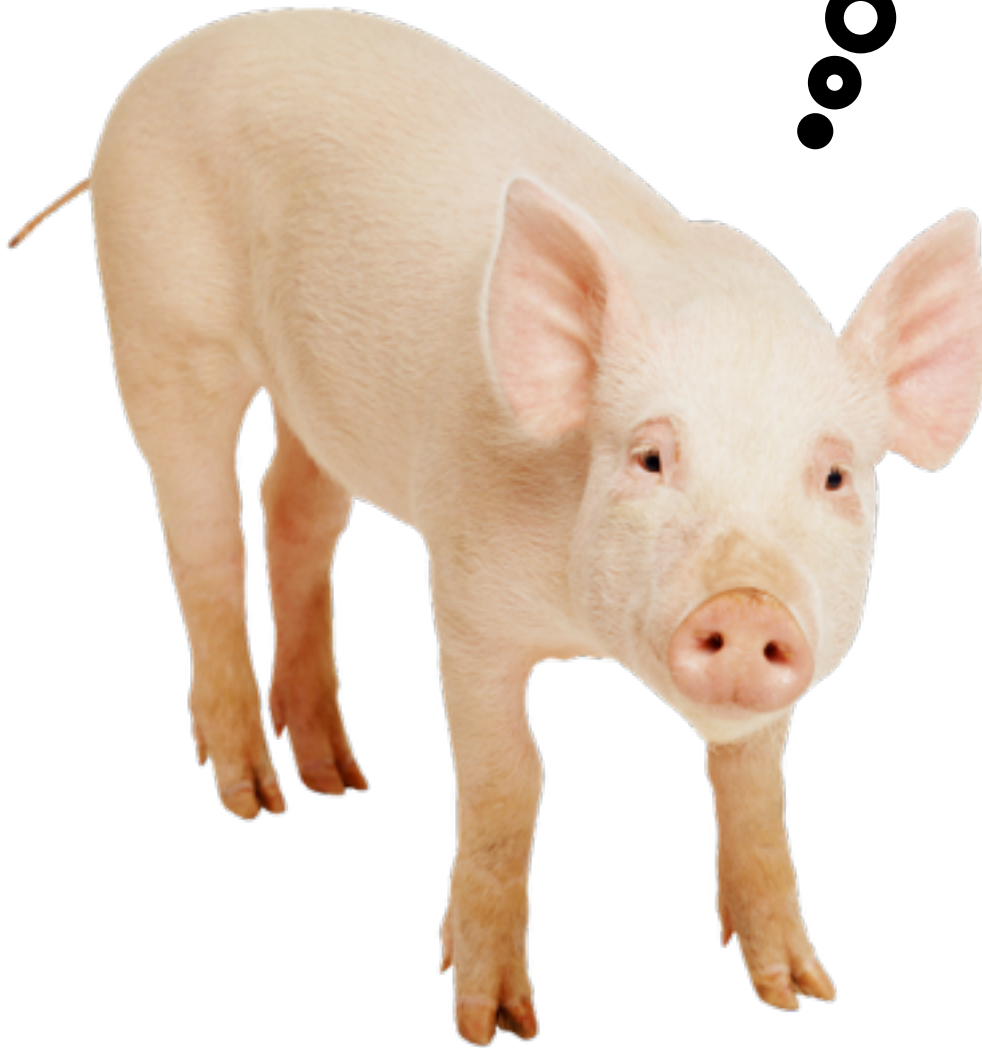
Key terms

- Genotype/Phenotype
- Cell
- Proteins
- Evolution: inheritance, selection, variation
- DNA/RNA
- Chromosome
- Gene
- Genome
- Replication
- Transcription
- Exon/Intron
- Translation
- Codon
- Central Dogma
- Gene Expression
- Regulation
- Epigenetics

Why are my children
such pigs?



Why *am* I such a pig?



*Phenotype, cells,
metabolism, protein*

Proteins

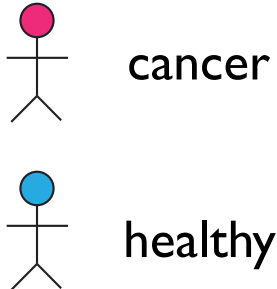
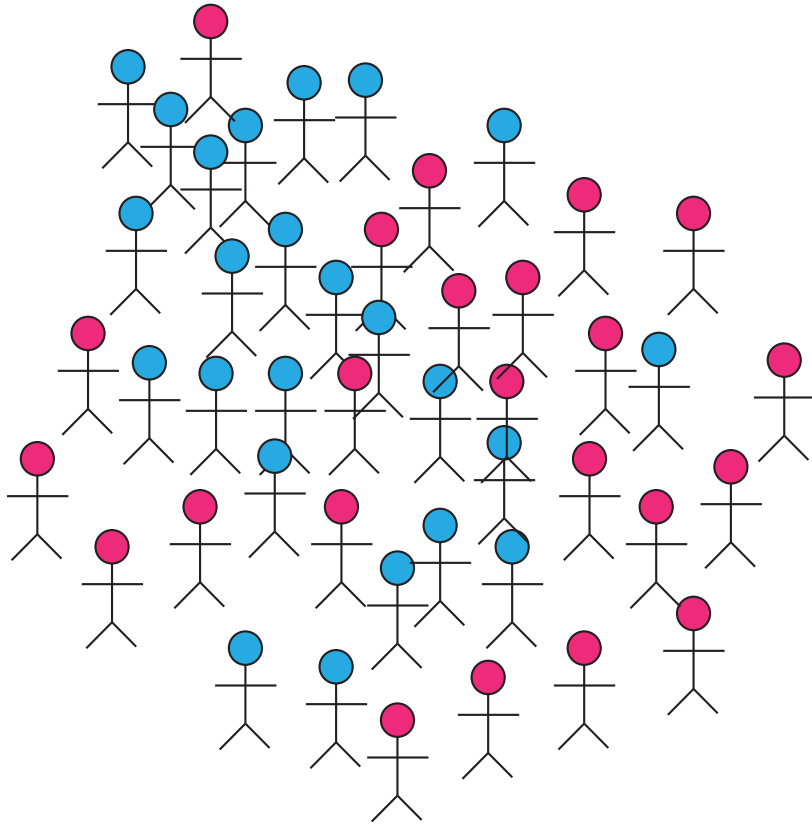
- *phenotype*: characteristics (traits) of an organism
- characteristics due to cellular structures and activities
 - mostly carried out by *proteins*
- Examples:

<i>alpha-keratin</i>	<i>component of hair</i>
<i>insulin</i>	<i>regulates blood glucose level</i>
<i>actin & myosin</i>	<i>muscle contraction</i>
<i>hemoglobin</i>	<i>oxygen transport</i>
<i>DNA polymerase</i>	<i>synthesis of DNA</i>
<i>DNA glycosylases</i>	<i>DNA repair</i>
<i>matrix metalloproteinase</i>	<i>extra-cellular matrix degradation</i>

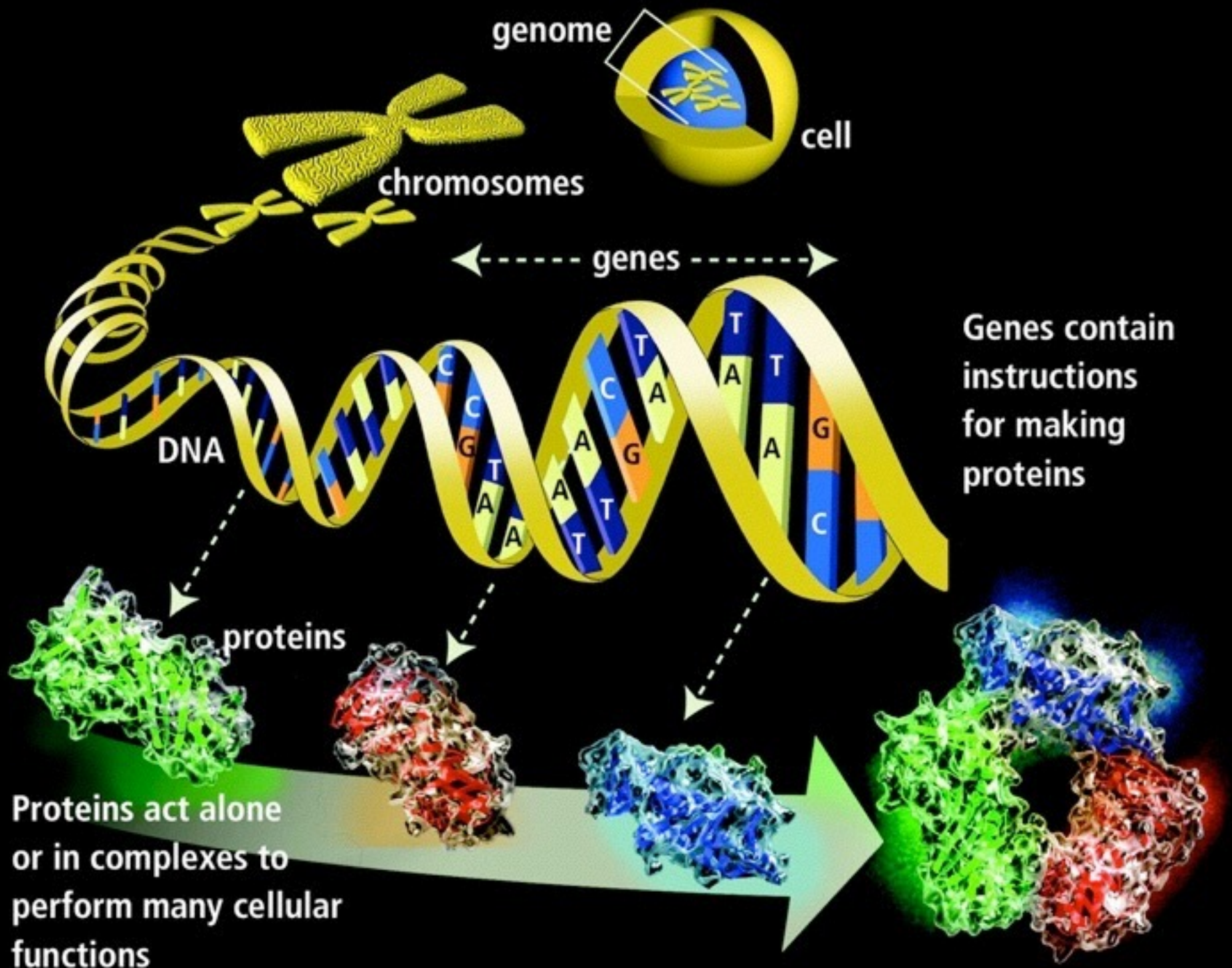
Genetics

- *gene*: in classical genetics it was an abstract concept
 - a unit of inheritance passed from parent to offspring
 - specify proteins
- *genome* refers to the complete set of *genes*
- *genotype*: genetic characteristics of an individual

What is Genomics?



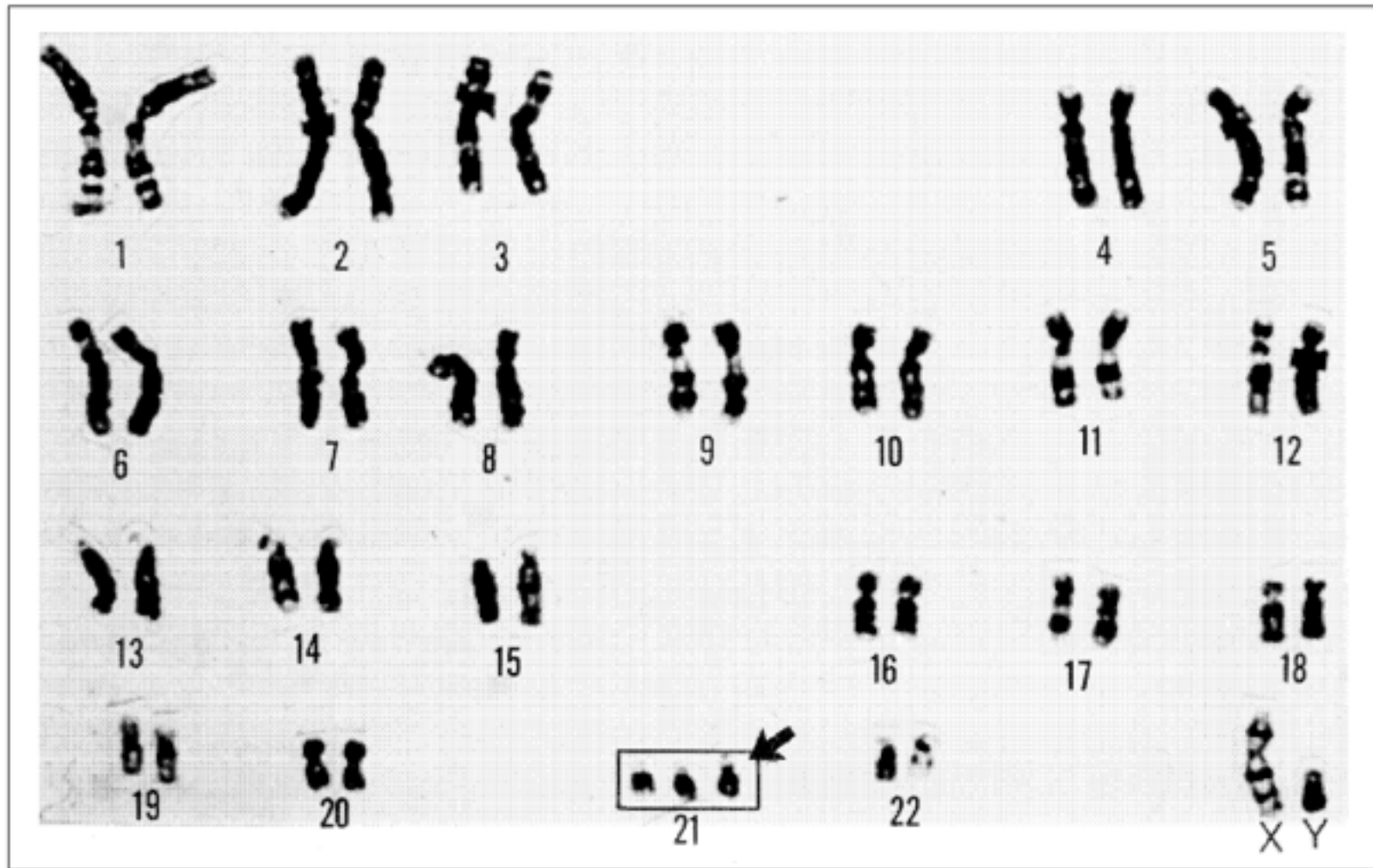
- Study the **molecular** basis of *variation* in development and disease
- Using **high-throughput** experimental methods
 - algorithms
 - ML
 - data management
 - modeling



What is Genomics?

- Each cell contains a complete copy of an organism's **genome**, or blueprint for all cellular structures and activities.
- The genome is distributed along **chromosomes**, which are made of compressed and entwined **DNA**.
- Cells are of many different types (e.g. blood, skin, nerve cells), but all can be traced back to a single cell, the fertilized egg.

Chromosomes



These are actually human. And for a down syndrome patient

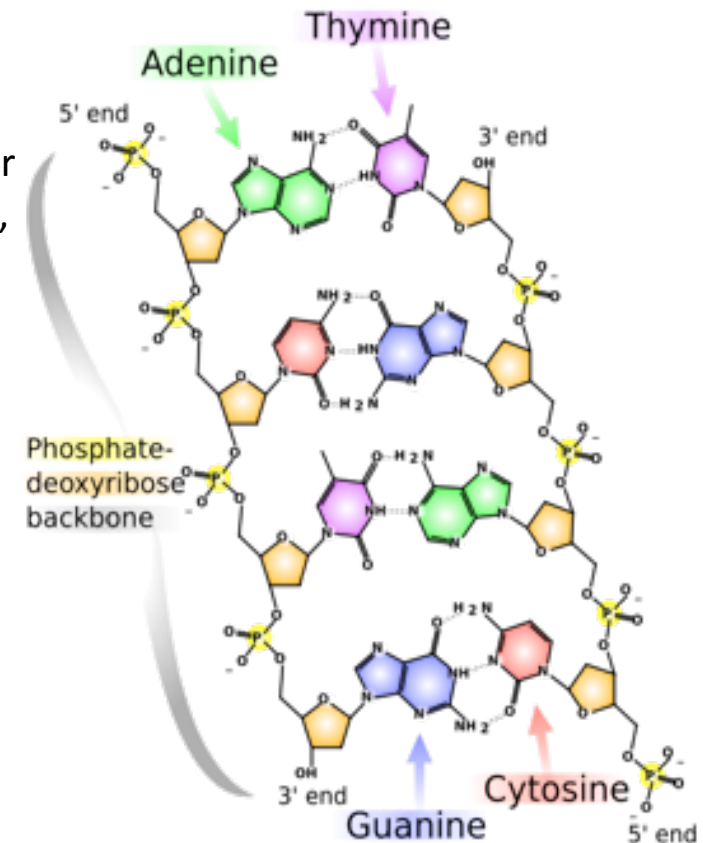
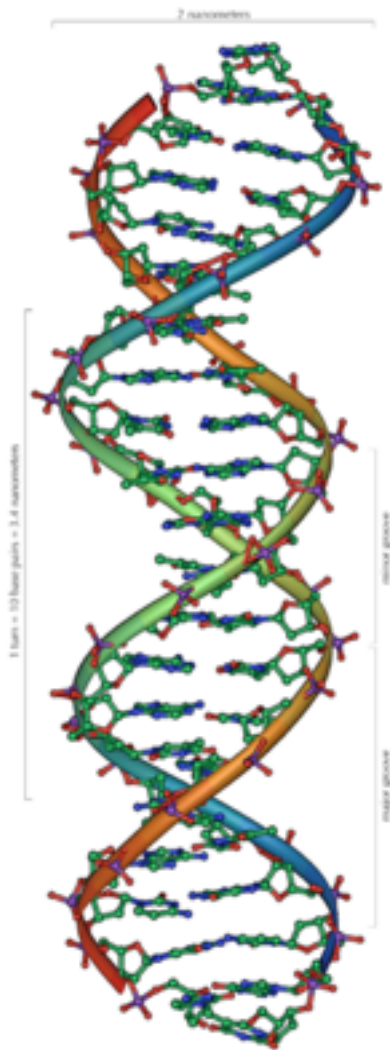
DNA

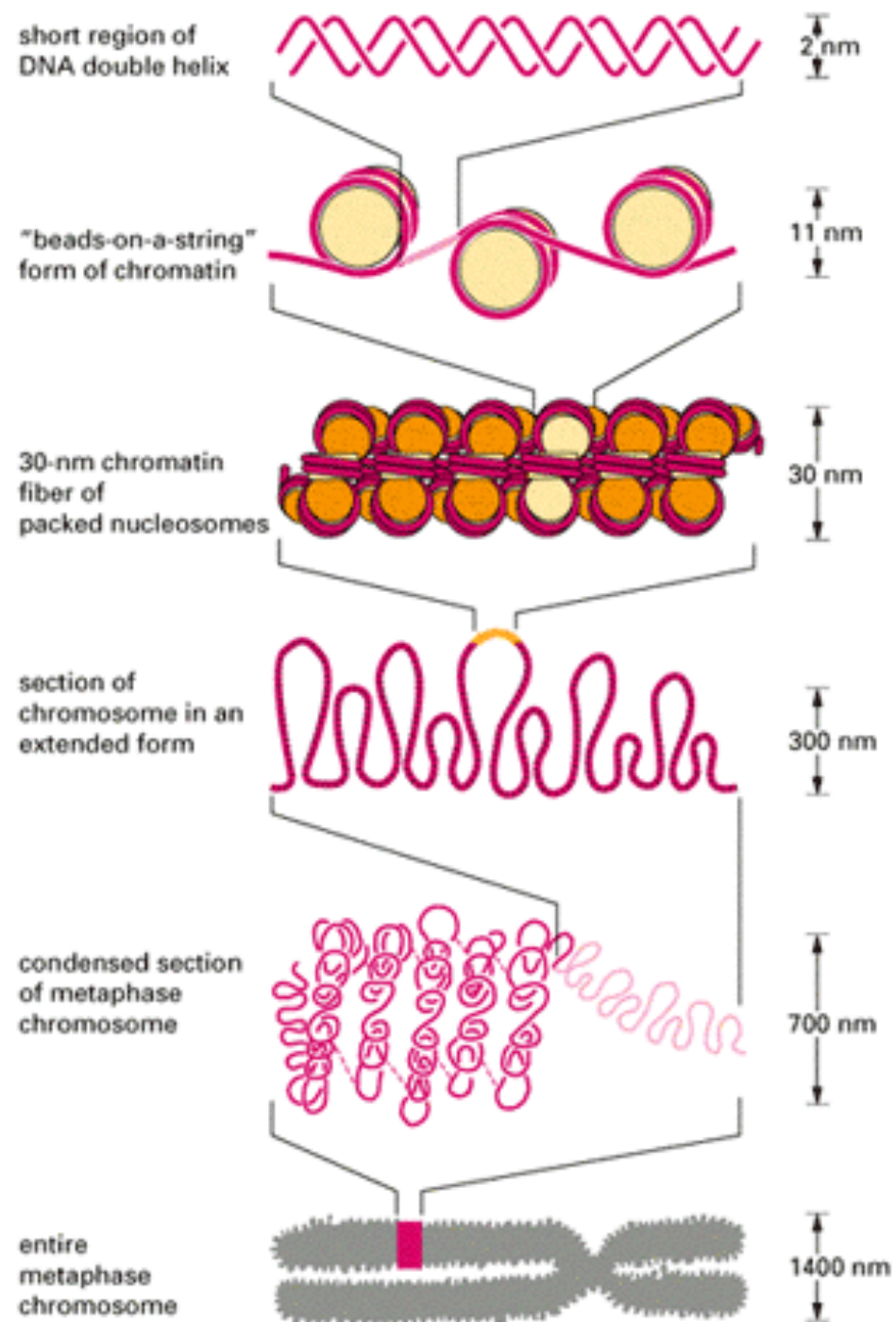
DNAs (Deoxyribonucleic acids) are molecules to store genetic information of a living organism.

DNA consists of two polymers made from four types of nucleotides: adenine (A) guanine (G), cytosine (C) and thymine (T).

Purines: A, G; Pyrimidines: C, T

Two polymers are complementary to each other and form a double-helix structure

$$\begin{array}{c}
 5' - \text{ACCGTTTCGACGGTAA} - 3' \\
 | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \quad | \\
 3' - \text{TGGCAAGCTGCCATT} - 5'
 \end{array}$$




chromatin

Measurement

- For a small enough piece, we can measure the sequence of bases, referred to as *sequencing*
- Human Genome Project



D. melanogaster, Science, 2000



H. sapiens, Nature, 2000
and Science, 2000



M. musculus, Nature, 2002

Genome

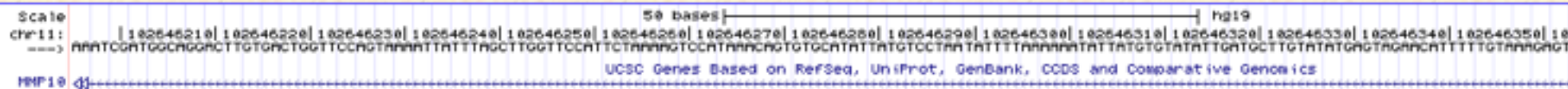
[illegible]

**Total amount of DNA in human genome:
3 * 10⁹ base pairs (bp)**

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr11:102,646,196-102,646,31 [gene](#) [jump](#) [clear](#) size 187 bp. [configure](#)



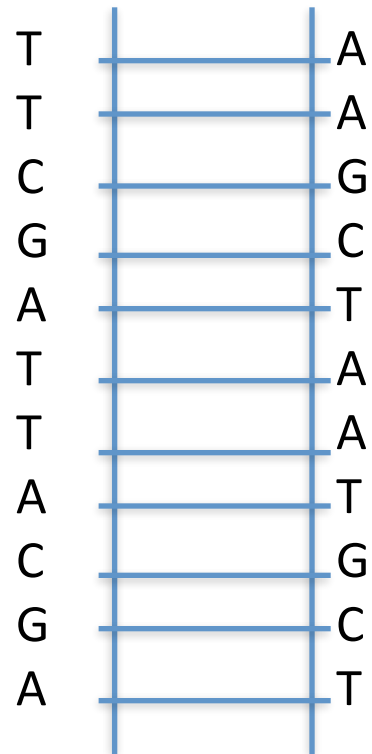
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

move start

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

< 2.0 >

Replication



T
T
C
G
A
T
T
A
C
G
A

A
A
G
C
T
A
A
T
G
C
T

T
T
C
G
A
T
T
A
C
G
A

C C C G T A A
G T
A T T T G

T T G G G T A A T
G C

A T G G G T C A A
T T A

T T T A G T A G

A A T G T C

A
A
G
C
T
A
A
T
G
C
T



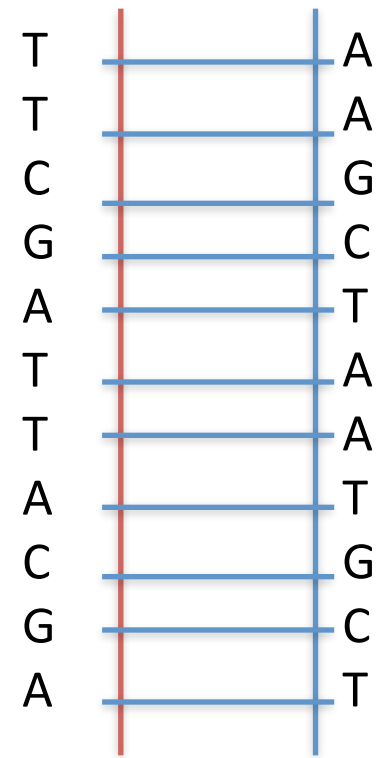
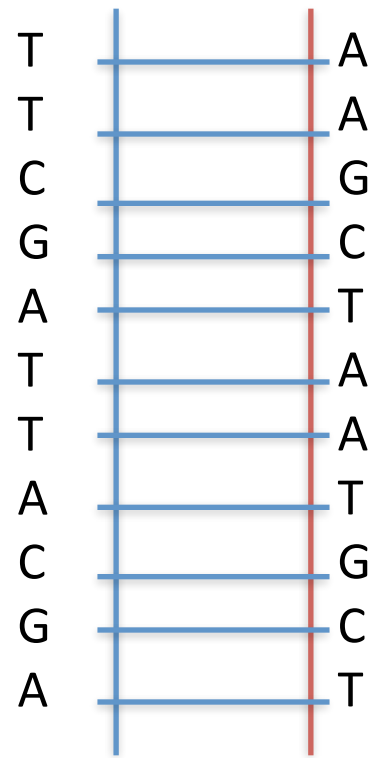
nucleotides available in cells

T
T
C
G
A
T
T
A
C
G
A

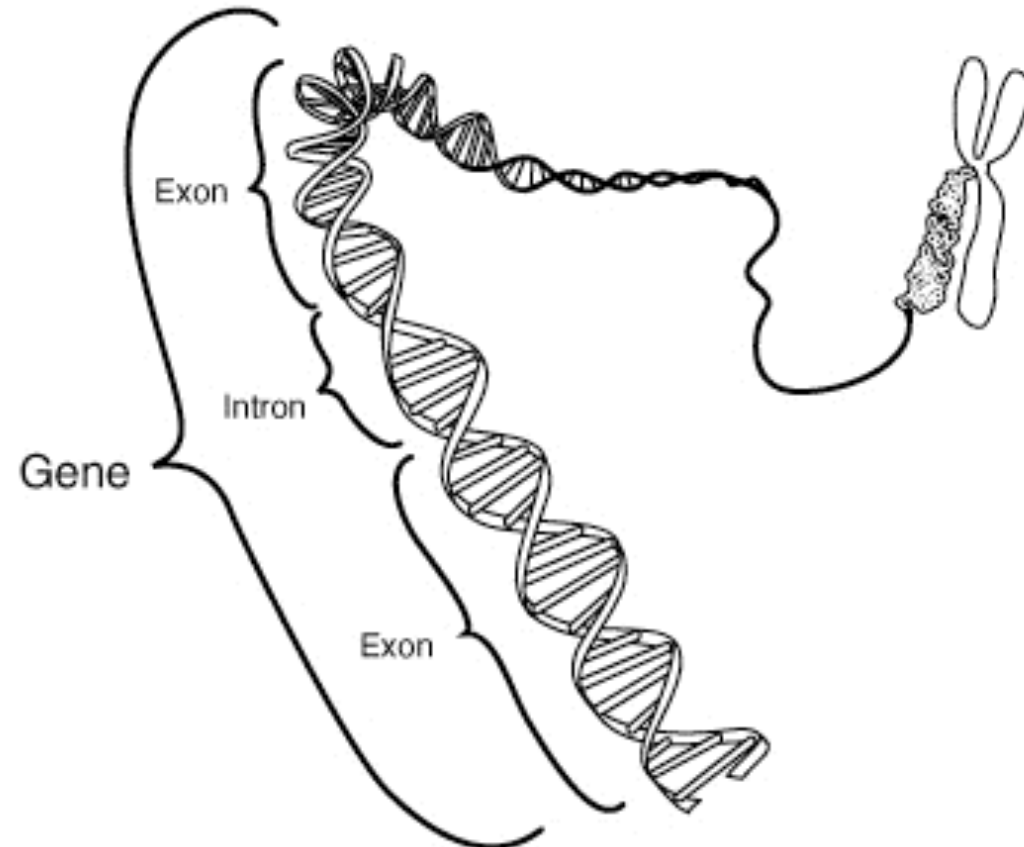
A
A
G
C
T
A
A
T
G
C
T

T
T
C
G
A
T
T
A
C
G
A

A
A
G
C
T
A
A
T
G
C
T



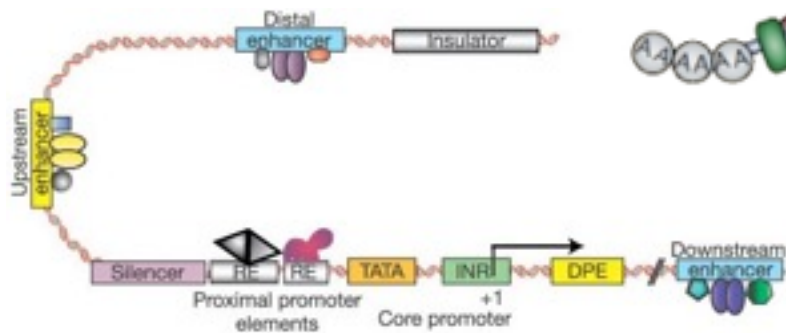
Genes



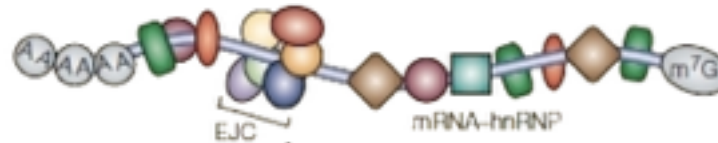
Central Dogma

Genes encode proteins which are transcribed into mRNA and translated into proteins.

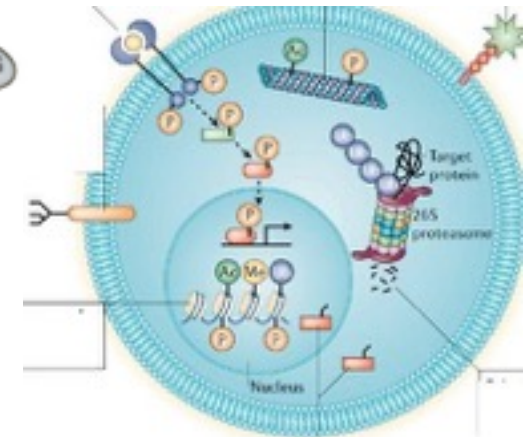
DNA



RNA



Proteins



Transcription

DNA



CT AGCGCT C
| | | | | | | |
GATCG CGAG



mRNA

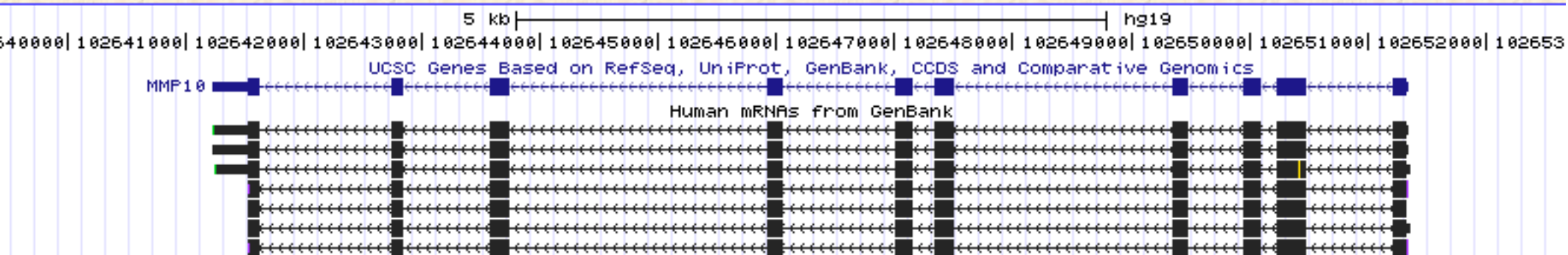
CUAGCG

Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

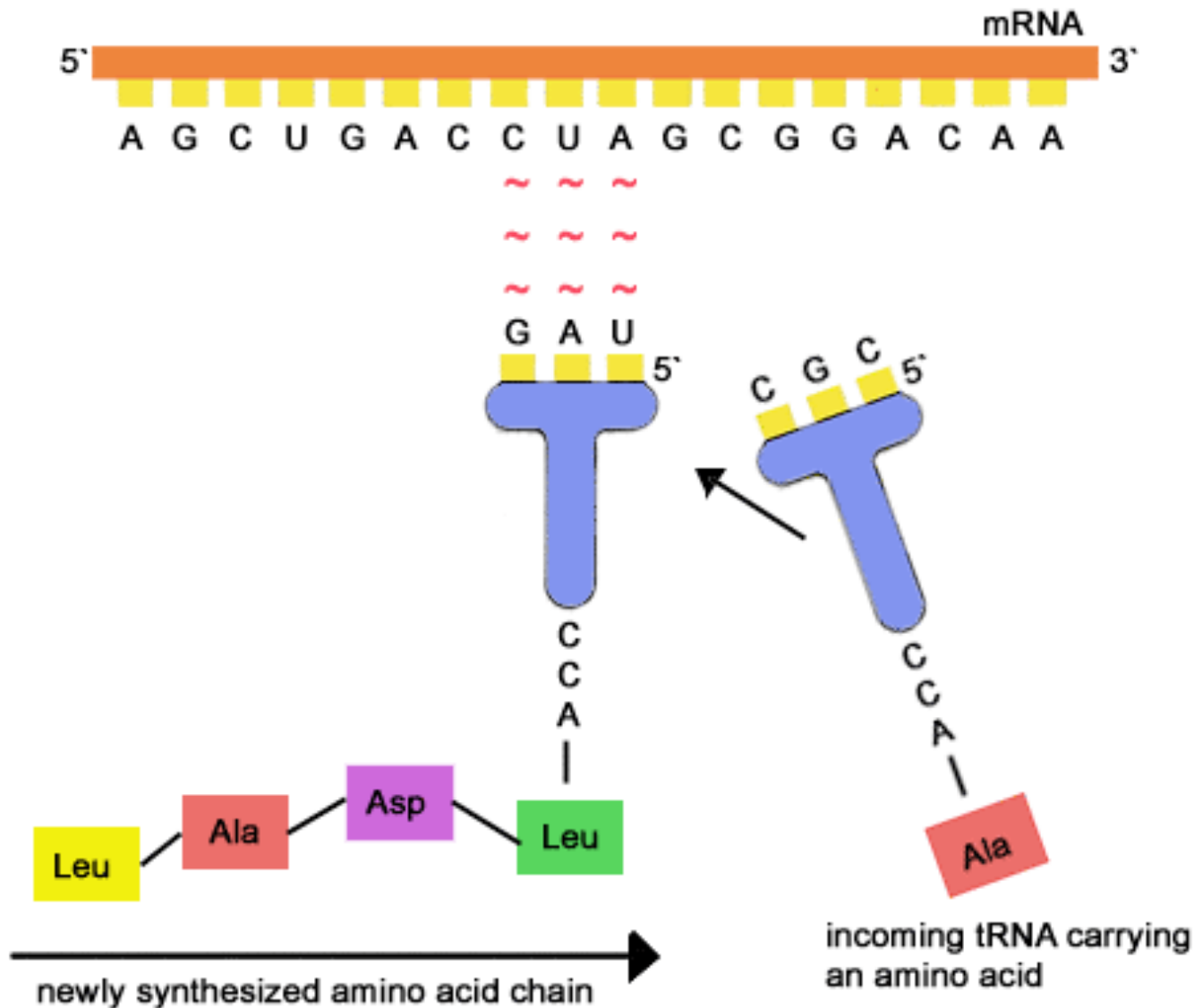
n/search chr11:102,636,939-102,655,6: gene jump clear size 18,700 bp. configure

5.1 p14.3 14.1 11p13 11p12 p11.2 12.1 q13.4 11q14.1 q14.3 11q21q22.1 11q22.3 11q23.3 24.2 q25



for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down. Drag tracks left or right to new position.

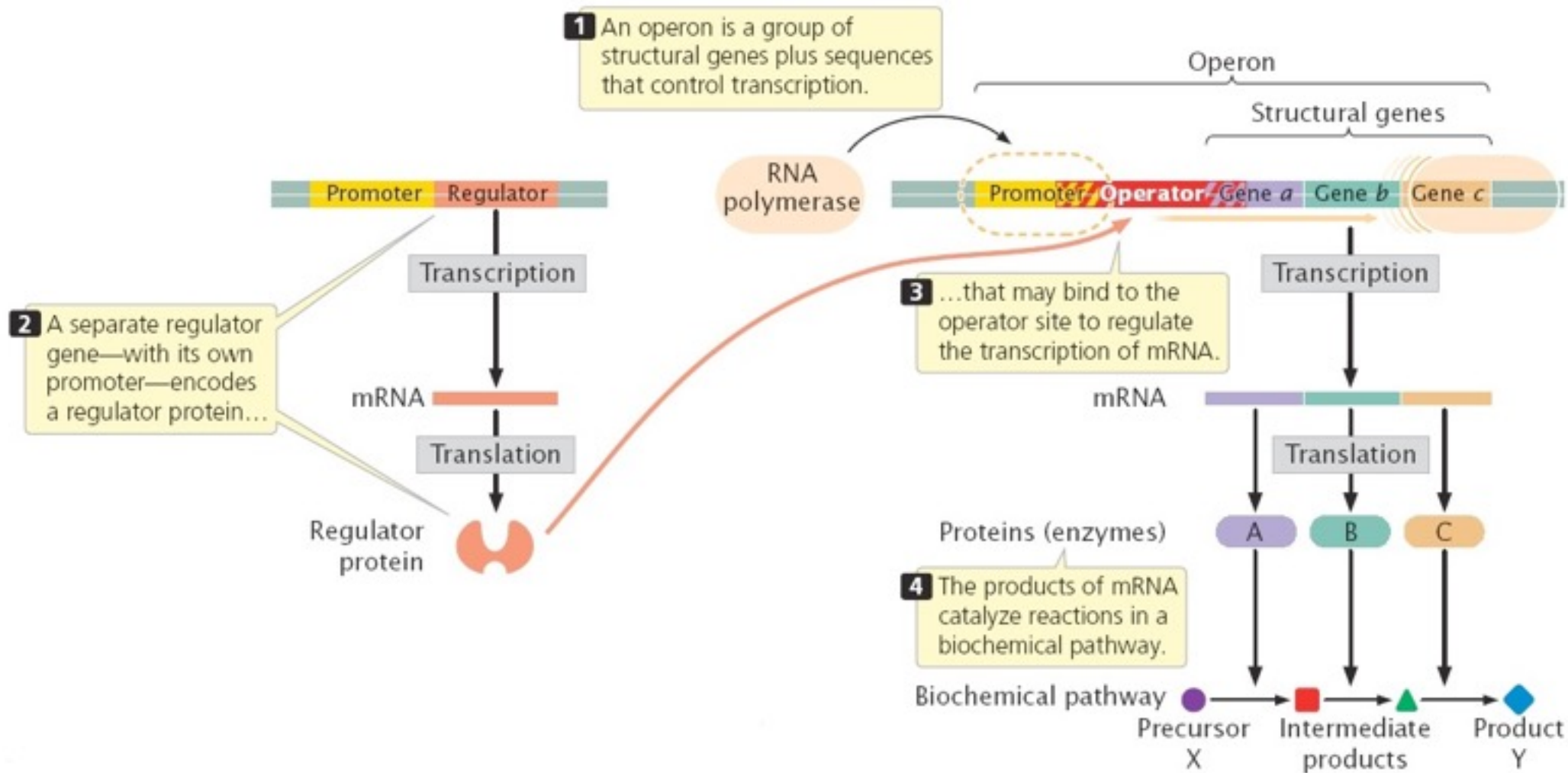
Translation



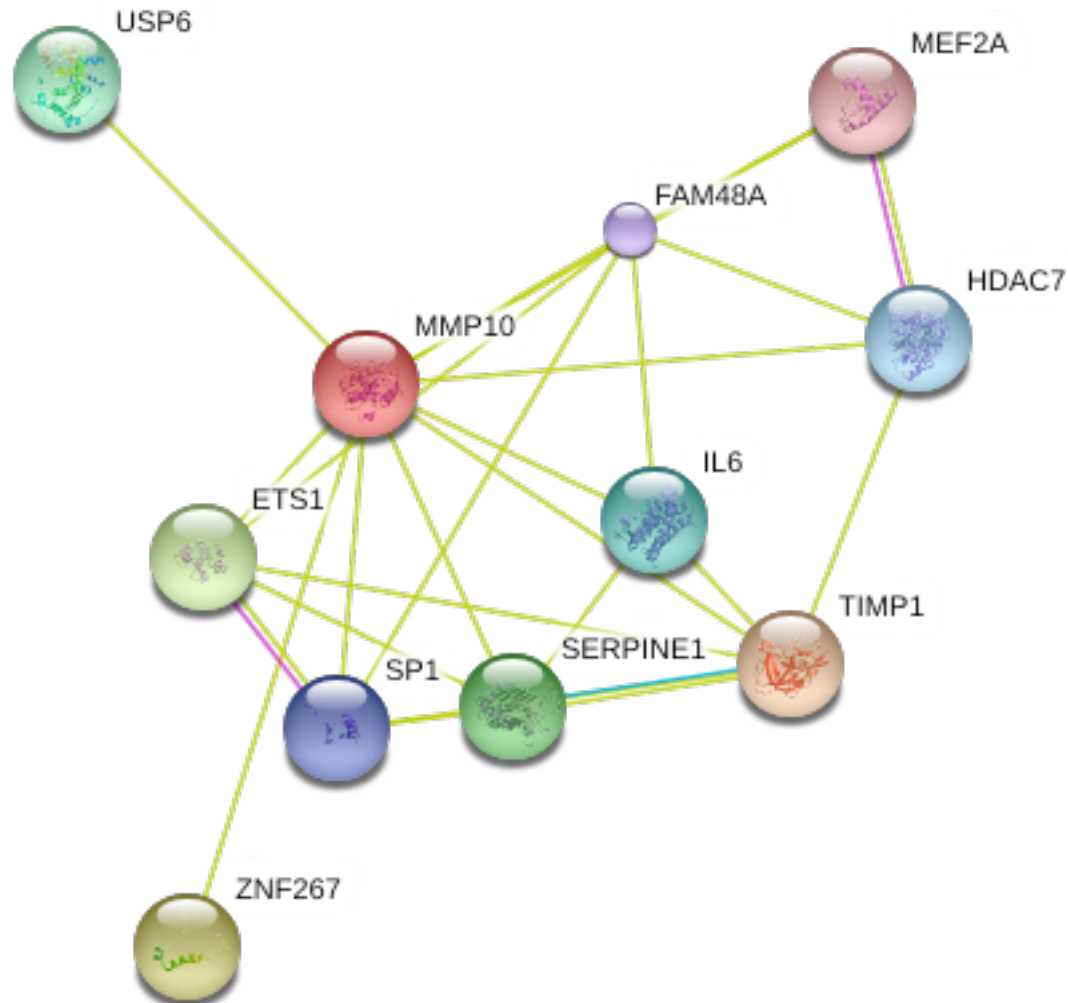
The genetic code

		Second Letter					
		T	C	A	G		
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA Stop TAG Stop	TGT } Cys TGC } TGA Stop TGG Trp	T	C
	C	CTT } CTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	CAT } His CAC } CAA Gln CAG }	CGT } CGC } Arg CGA } CGG }	T	C
	A	ATT } Ile ATC } ATA } ATG Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA Lys AAG }	AGT } Ser AGC } AGA Arg AGG }	T	C
	G	GTT } Val GTC } GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA Glu GAG }	GGT } GGC } Gly GGA } GGG }	T	C
						Third Letter	
						A	G

gene regulation



gene regulation



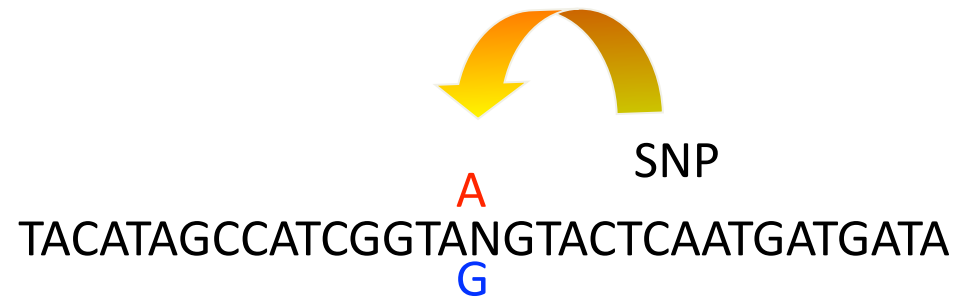
What makes them different?



Much human variation is due to difference in ~ 6 million base pairs (0.1 % of genome) referred to as SNPs

Single Nucleotide Polymorphism (SNP)

Genomic DNA:



Three genotypes

AA

Mother

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTTCATGAGTTACTACTAT

Father

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTTCATGAGTTACTACTAT

AG

Mother

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTTCATGAGTTACTACTAT

Father

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATCCATGAGTTACTACTAT

GG

Mother

TACATAGCCATCGGTA^GGTACTCAATGATGATA
ATGTATCGGTAGCCAT^CCATGAGTTACTACTAT

Father

TACATAGCCATCGGTA^GGTACTCAATGATGATA
ATGTATCGGTAGCCAT^CCATGAGTTACTACTAT

DNA Sequence Variation in a Gene Can Change the Protein Produced by the Genetic Code

Gene A from Person 1

GCA AGA GAT AAT TGT...
Ala Arg Asp Asn Cys ...
1 2 3 4 5

Protein Products



Gene A from Person 2

Codon change made no difference in amino acid sequence

GCG AGA GAT AAT TGT...
Ala Arg Asp Asn Cys ...
1 2 3 4 5

Gene A from Person 3

Codon change resulted in a different amino acid at position 2

GCA AAA GAT AAT TGT...
Ala Lys Asp Asn Cys ...
1 2 3 4 5

OR



Health or Disease?

