

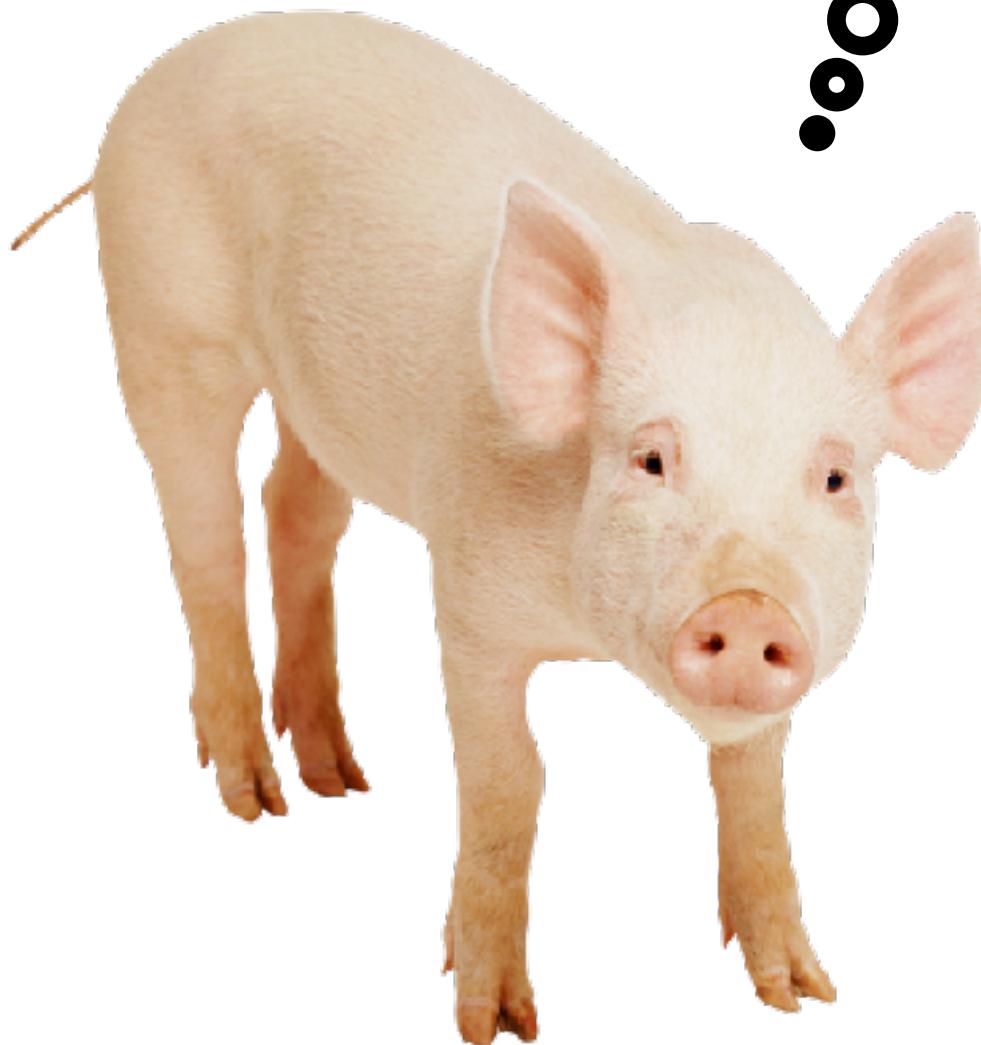
A super quick introduction to molecular biology and genomics

Héctor Corrada Bravo
Dept. of Computer Science
Center for Bioinformatics and Computational Biology
University of Maryland

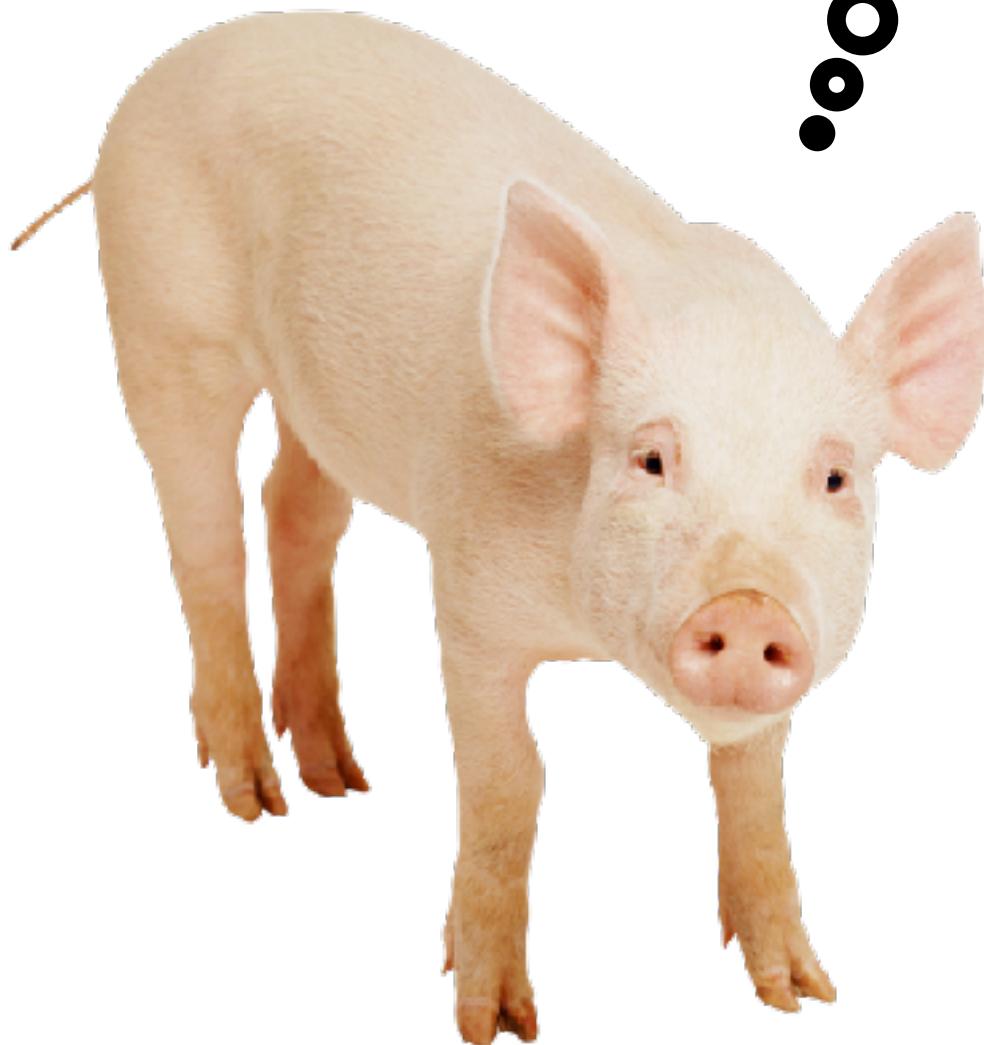
Key terms

- 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

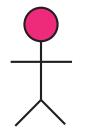
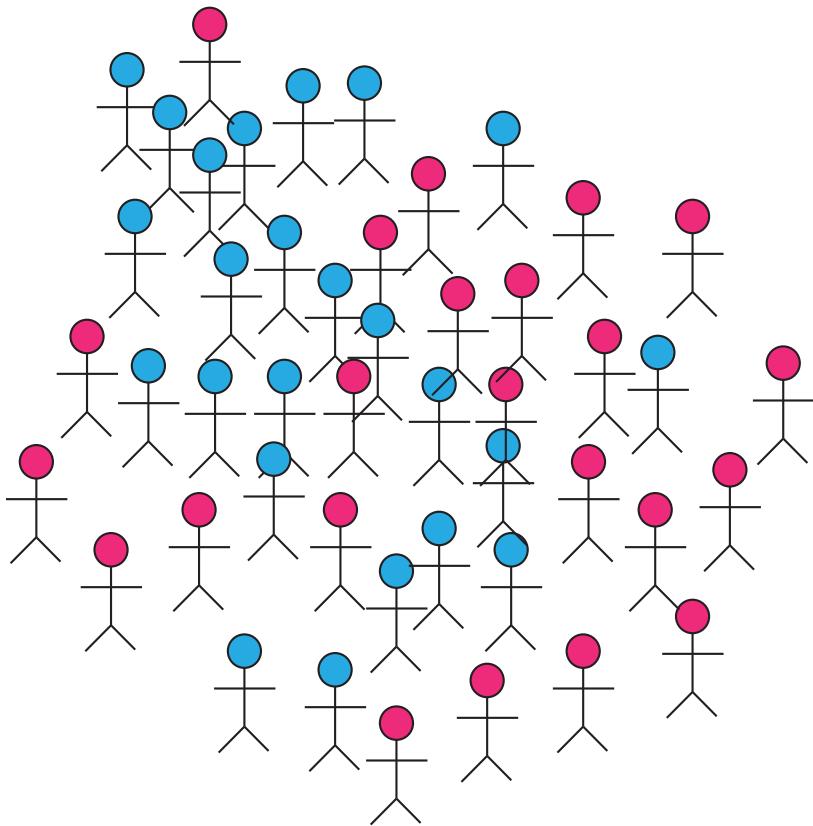
- *phenotype*: characteristics (traits) of an organism
- characteristics due to cellular structures and activities
 - mostly carried out by *proteins*
- Examples:

<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>

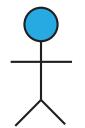
Genetics

- *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?

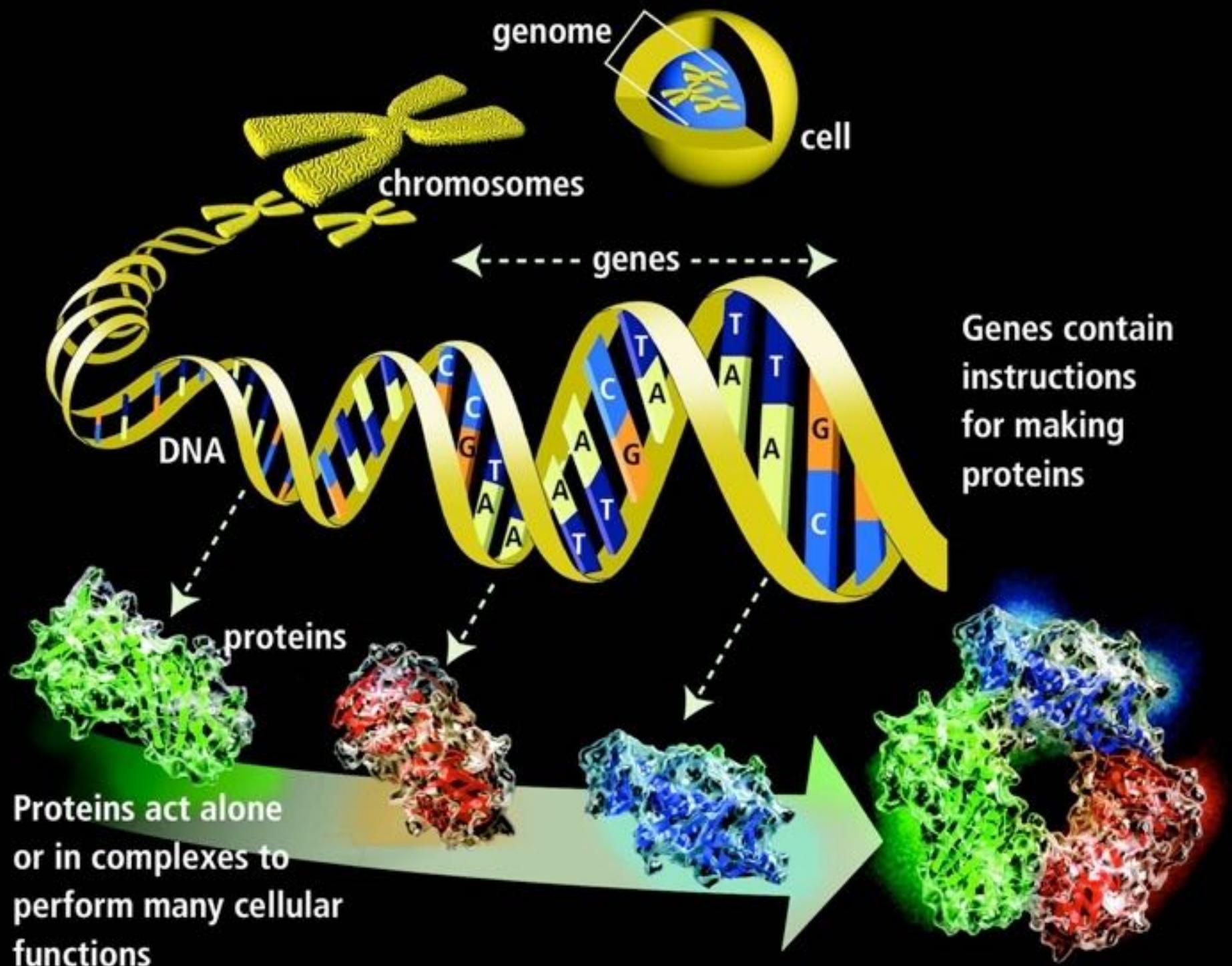


cancer



healthy

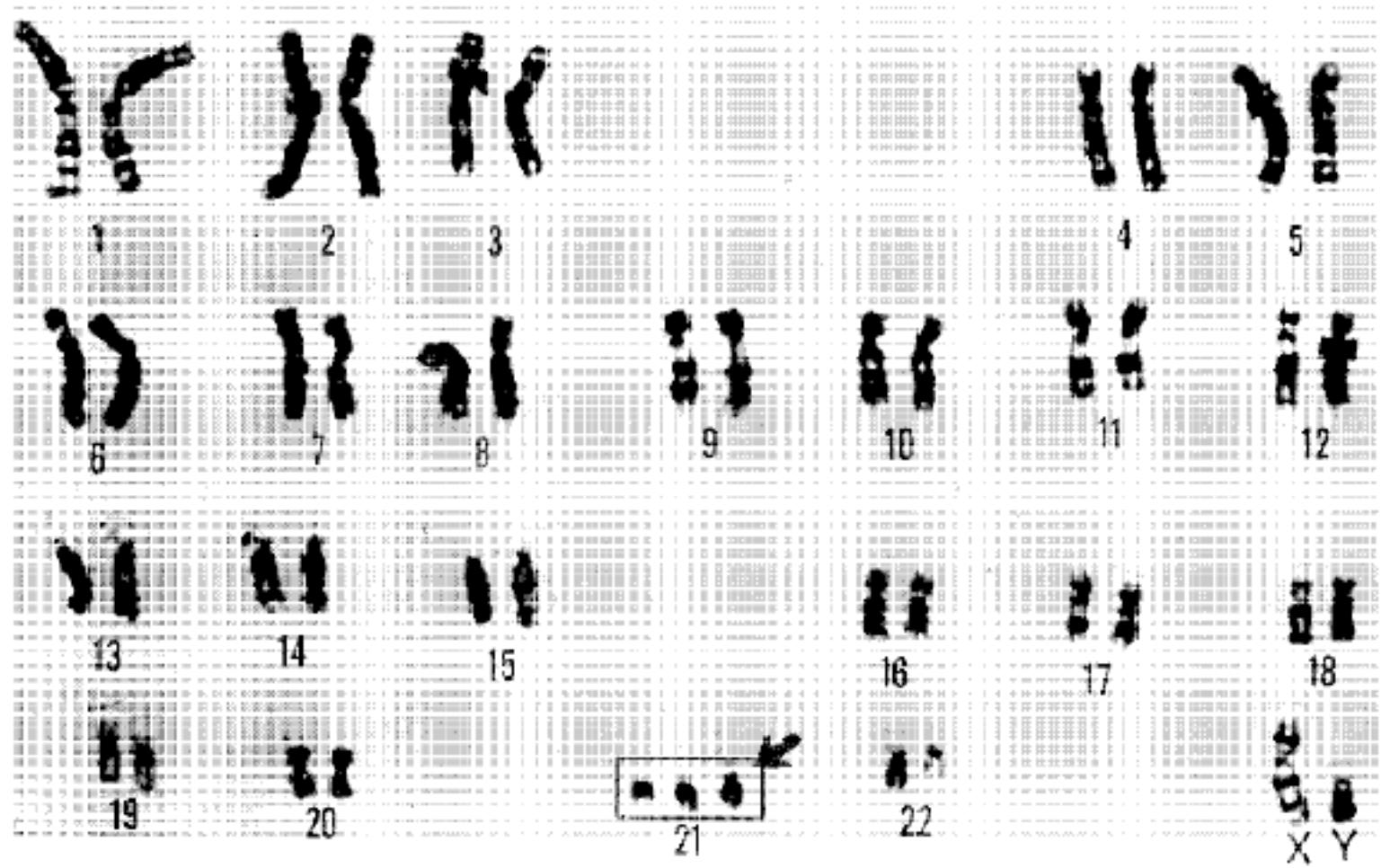
- Study the **molecular** basis of variation in development and disease
- Using **high-throughput** experimental methods
 - algorithms
 - ML
 - data management
 - modeling



What is Genomics?

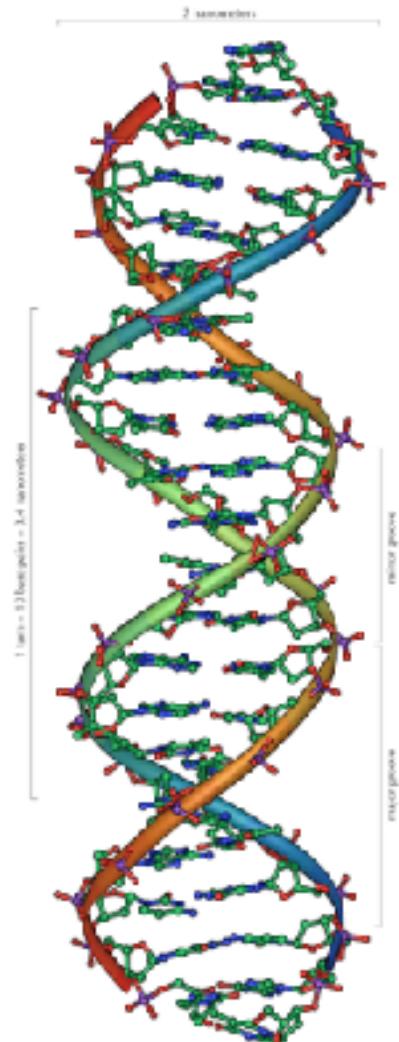
- Each cell contains a complete copy of an organism's **genome**, or blueprint for all cellular structures and activities.
- The genome is distributed along **chromosomes**, which are made of compressed and entwined DNA.
- Cells are of many different types (e.g. blood, skin, nerve cells), but all can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

Chromosomes



These are actually human. And for a down syndrome patient

DNA



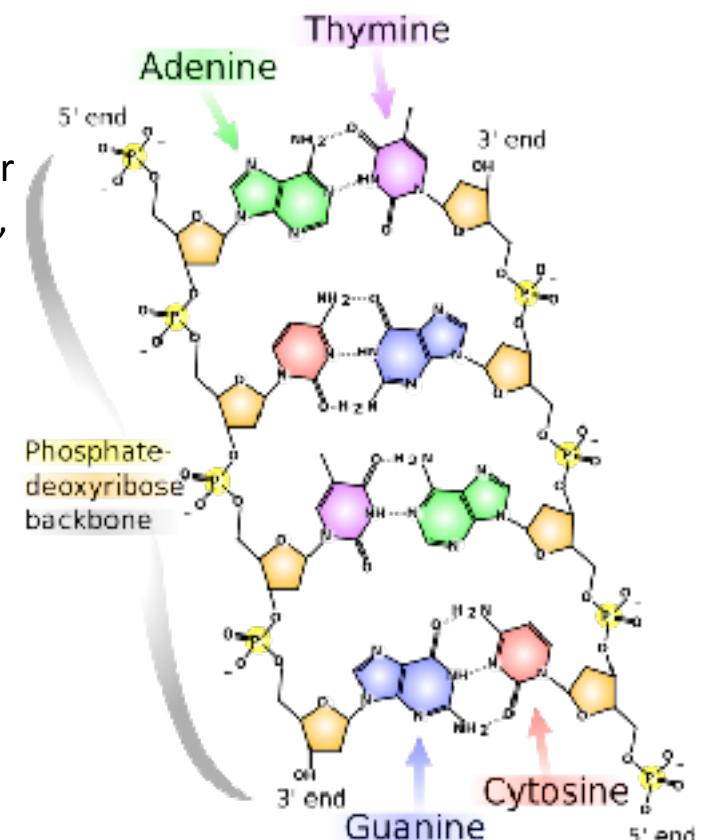
DNAs (Deoxyribonucleic acids) are molecules to store genetic information of a living organism.

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

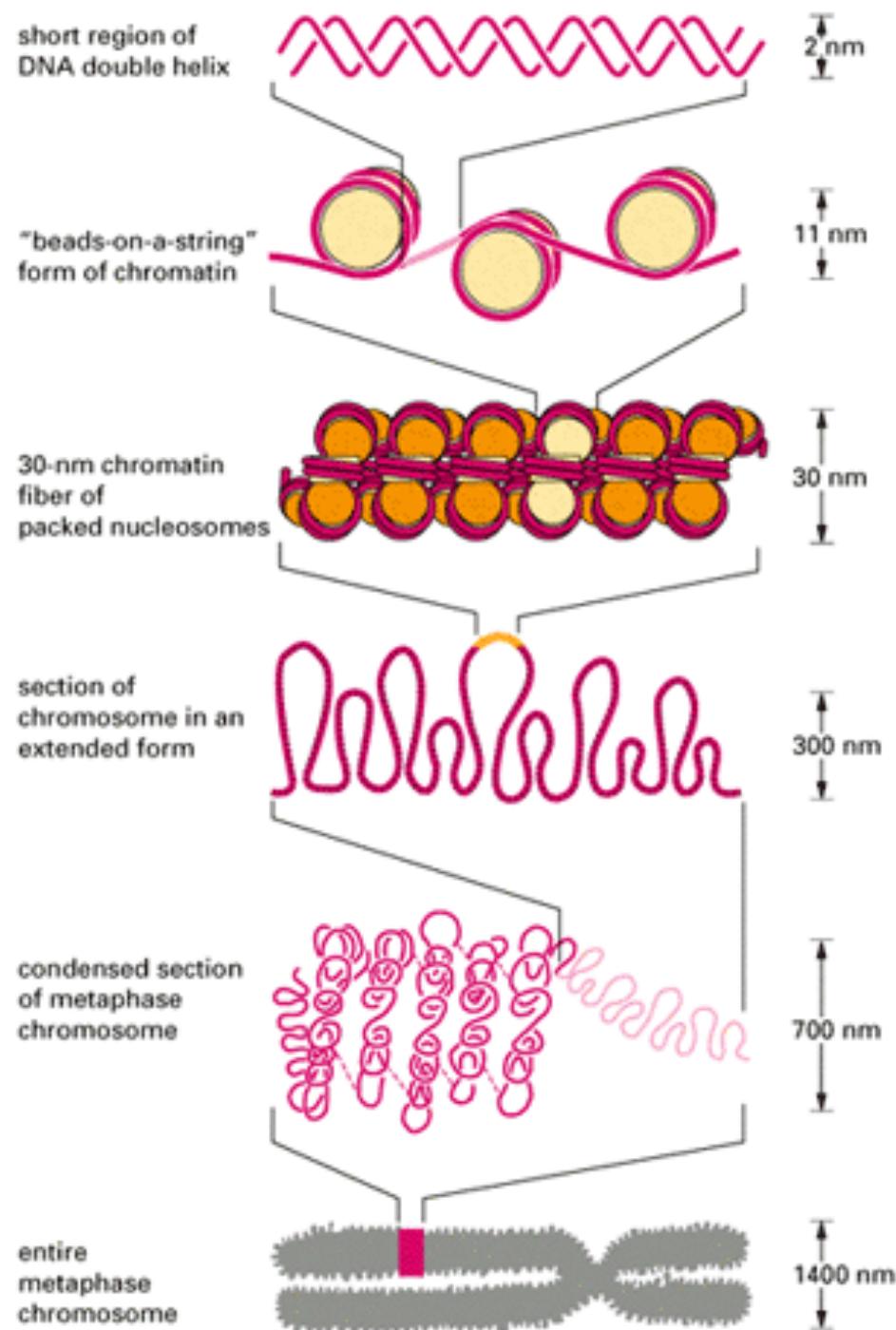
Purines: A, G; Pyrimidines: C, T

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

5' -ACCGTTCGACGGTAA-3'
 |||||||
3' -TGGCAAGCTGCCATT-5'



Watson and Crick 1953



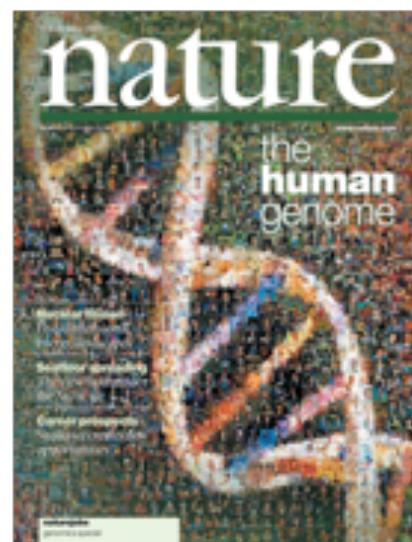
chromatin

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

TCAGTTGGAGCTGCTCCCCACGGCCTCTCCTCACATTCCACGTCTGTAGCTCTATGACCTCACCTTGAGTCCCTCCTCACACACTGAC
ATGAAAAGGCACATGAGGATCCTCAAATACCCCGTATCAGTCTCAGGGTAGCTCATAGCCTGGACAGGGCCCCCTGGGGTTGCGCCC
AGGTCCAGGCGGGGATGCACAGCAACAGTCACCGAAGCAGAAGCCGTACAGTGGTATGGCTGGCAGTAGCTGGCACAGAGCTGCCAT
GGCGGTGGACGTTGGGTTCCGAGGGTTGTGAGAACGGGCCACGGGCCCTGAGCGGTCCCTATTGCTAGGGCAGAATGCCCTCAGTAGA
AATTTCAAAAGCGTCTCGCGGGTCTGTAGGGGGTGGCCGAAGCCTCTAGGGGATCCCTCGAGGCTGCTGGCCTGCCGTCCAGG
GGACAAGGAGCCAGAGTCAGGTGGGCTGTTGCCAGGGGTAAGGGAGGCTGATGTCTGGAGTCCGGATGGACCACCTGCAGAGGAGAGAC
ATAGGTCAACACAGGGAGGTAGGATGGTGTGATGTTCCACCCACAAAAGAAAACCTATTCTTAGAACCTCCAGGATGTGAATCCTGCCT
GCACCTGCACAGCTGGCTGGAGGCATATAGCCACTGCCCATAGATCTCAACTTACCTCACAAACCAACTGCCAGGCCTAAGTTCTGCC
TCAAAAC TGCCAAGGCCTGGATAGCCAAGAGCCTGGGTCTGGAAATATGCAACCATAAATAGTAGCTTTAGAAGTATAAGGCTCTGTT
TCTGGGTCAATTAGTGTGTTTCACCTGTCCCCAGCCCTAACAGCCAGGTGTGGCCAGAACGAAATGTACTGTAAGAGCAGAGCAAAACTTC
CACACAGATAGTCTGTTAGGCAATACATCTGCCTGACTATTAGGAATCTGGTTCTGGCCTCTGTACAAAGCTGGAGCAACACAGTG
GCCACATCAATCAAAGGACCGTGACCAACTCAAAGTCGGTGAGCTGTACCTATTAGGCTCTGTAACAGAACAGATTCAACTA
CAGCTCAGCAGGGCATCGTCACGGGTGTGTGTGTGTGTGTGTGTGTGTGTGGGGGGGGGGGGTGGACAGAGGACGGGAC
ACAATTCACTGCCAGCCCTCTCCTCAAGGAAGGCTGCTTAGCCTGGACTGGAATACACATTCTGTAAACATGGTGGGGCCTCA
GGCAAGCCAGAGTTGGAGCCTCCTTAACTCTCAAGGTGAGCATCTTGACTTGGAGGGTGGGGTAAGGAAGGAACCTGTGGAC
TCCTCCCTACAAGACAGAAAAGGAATAAGCCACGAAGACAATAACGATTTGTATCAAGCGCCTCTCCCATTCACTGACAATGA
AATCAAATTGGACCCCTGCAAGCATTACACCCAGCAGAGTGGACACAGCACCGTCCAGAACGGGAGCAAACATGTGCTCAGAGCGAGCA
TAGCCCTGTGGTTCTGTCCCCAATGGCTGTCAGAAAGGCCTGAACAAAGGAGAAAATTGACACGGTCACATTCTGGGTGTGGTAAAGTGCTC
AGCTGTGTCTATACTTGGGTTTGAT...

**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 2x 10x base ZOOM OUT 1.5x 2x 10x

position/search chr1:102,646,196 102,646,3 genome jump clear size 187 bp, configuration

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.

Replication

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

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

TTA

TTT 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

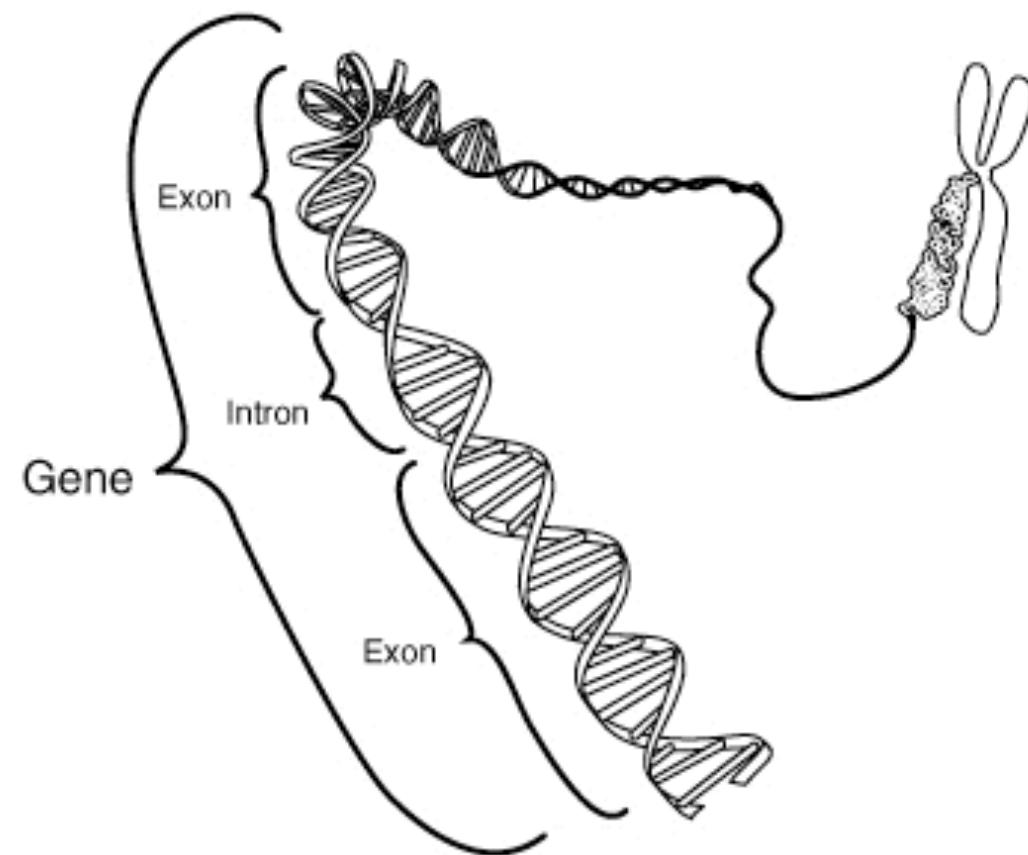
T
T
C
G
A
T
T
A
C
G
A

A
A
G
C
T
A
A
T
G
C
T

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

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

Genes

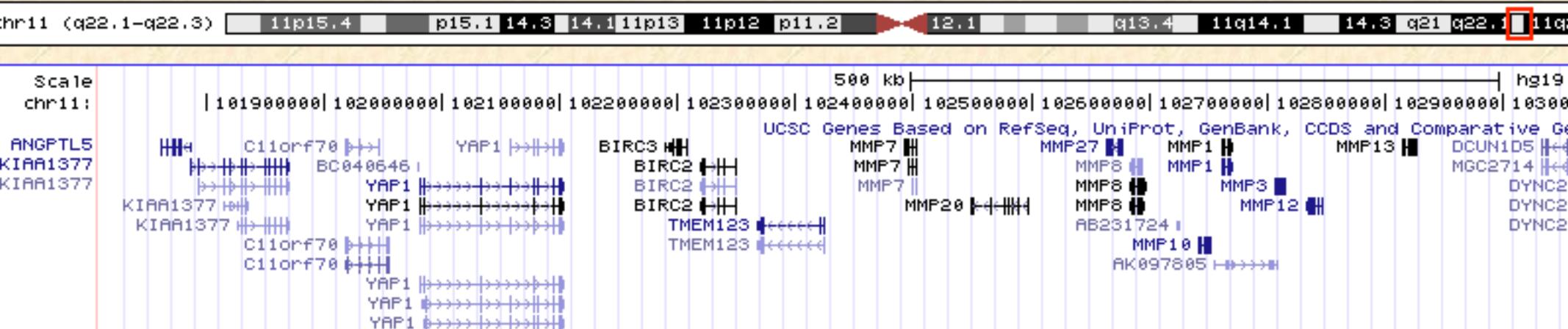


[Genomes](#)[Blat](#)[Tables](#)[Gene Sorter](#)[PCR](#)[DNA](#)[Convert](#)[PDF/PS](#)[Se](#)

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

move zoom in [1.5x](#) [3x](#) [10x](#) [base](#) zoom out [1.5x](#) [3x](#)

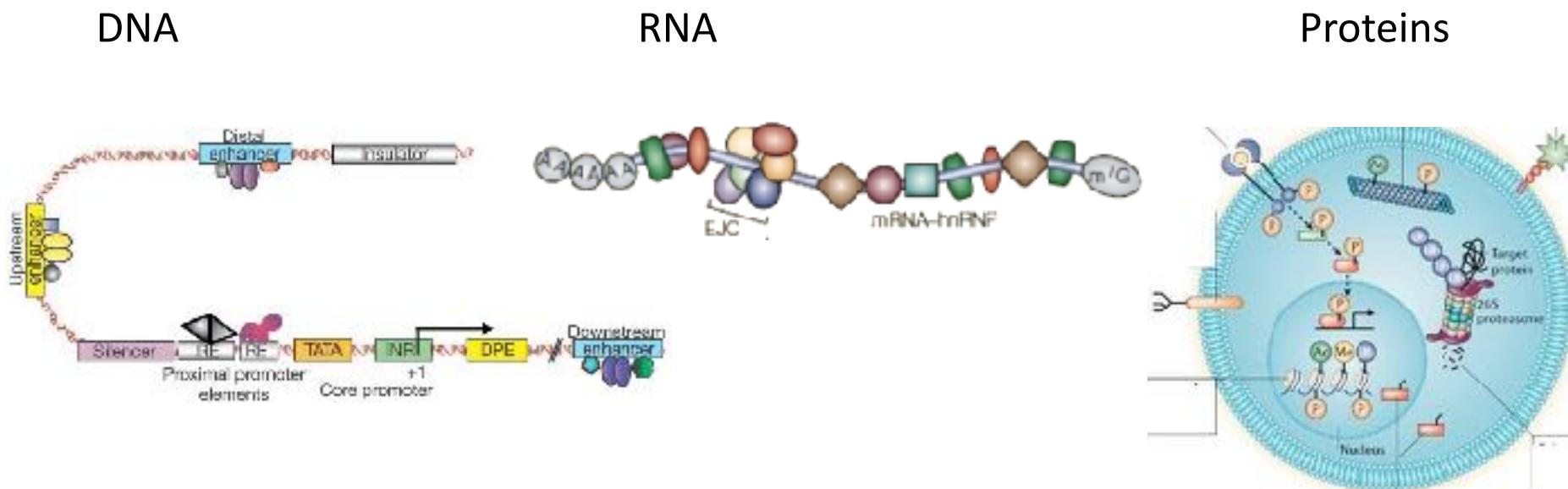
position/search [chr11:101,711,289-103,581,21](#) gene [jump](#) [clear](#) size 1,870,000 bp. [config](#)



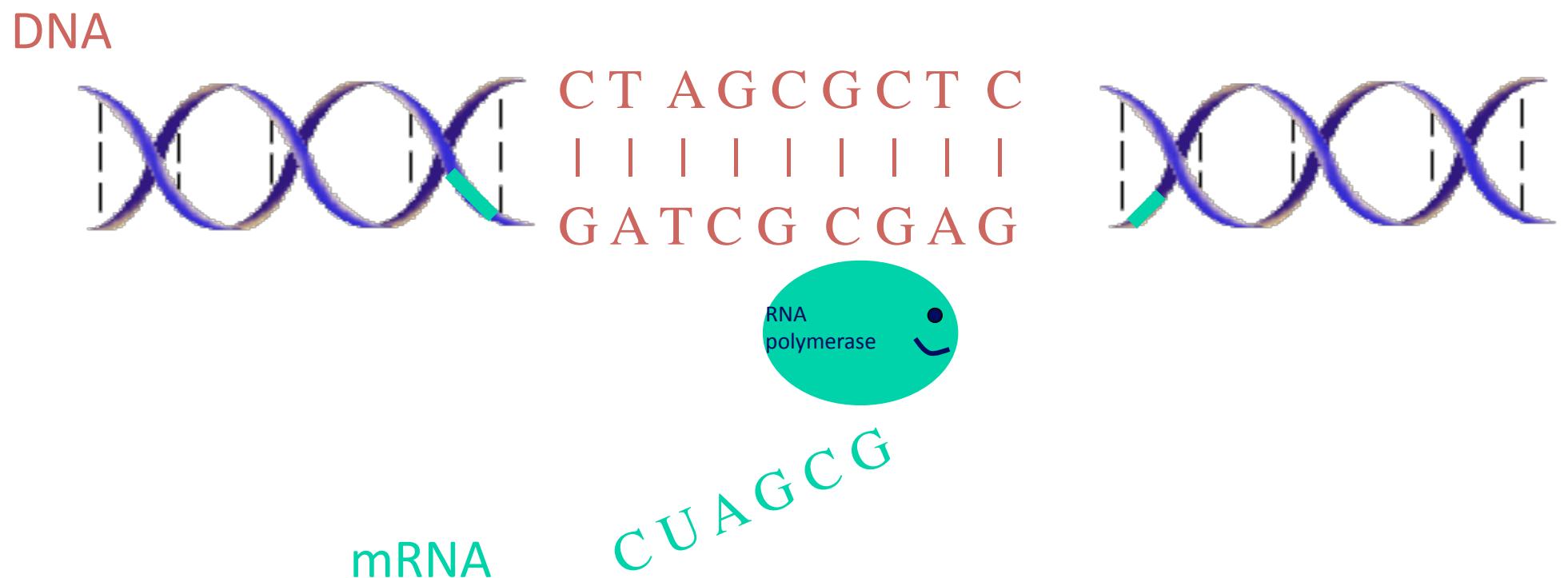
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options.

Central Dogma

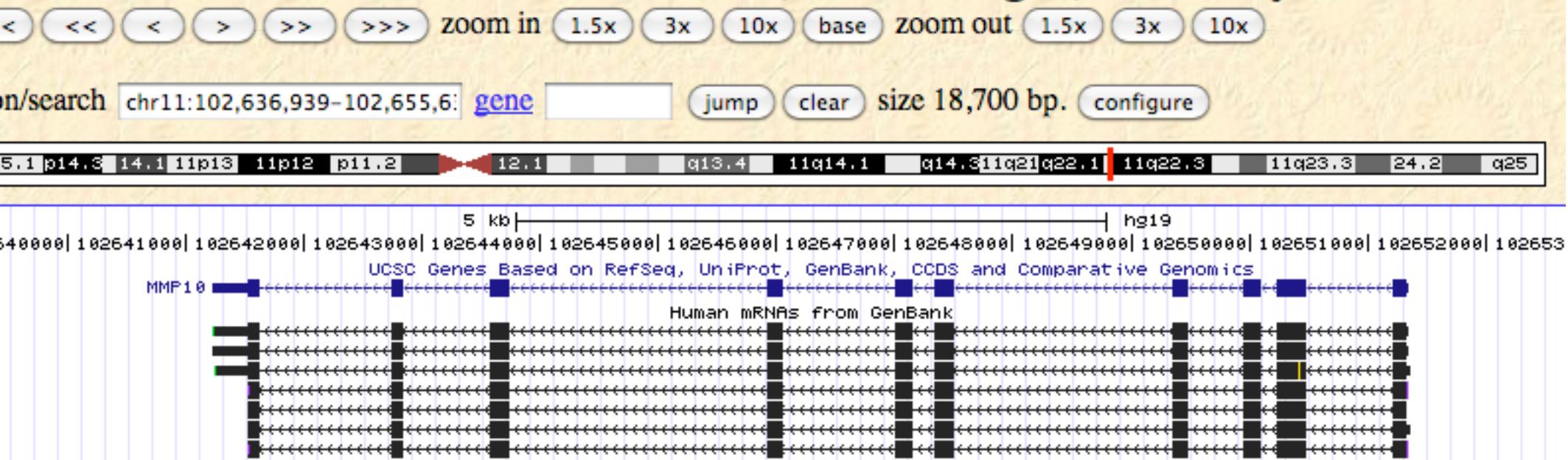
Genes encode proteins which are transcribed into mRNA and translated into proteins.



Transcription

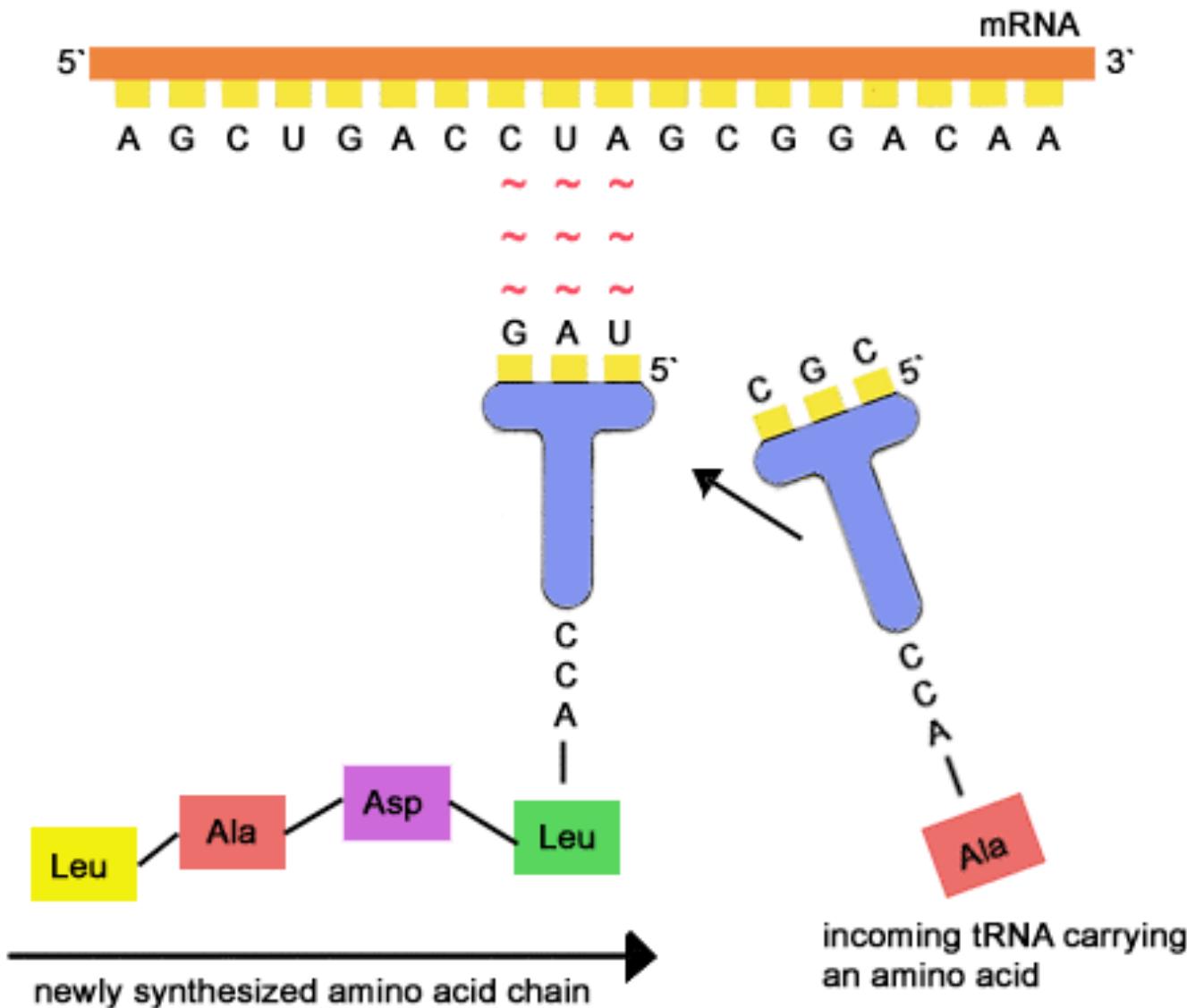


Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly



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.

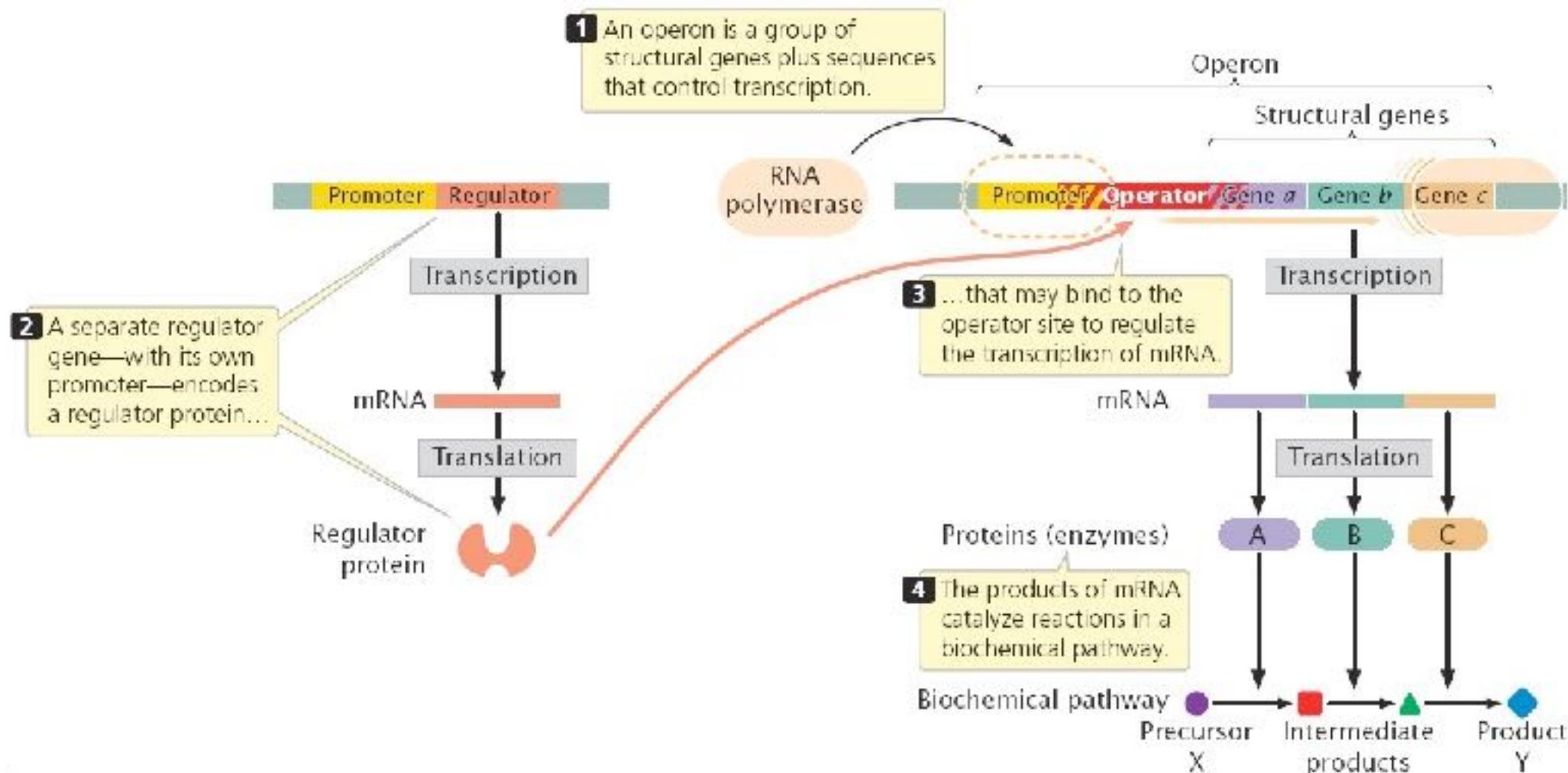
Translation



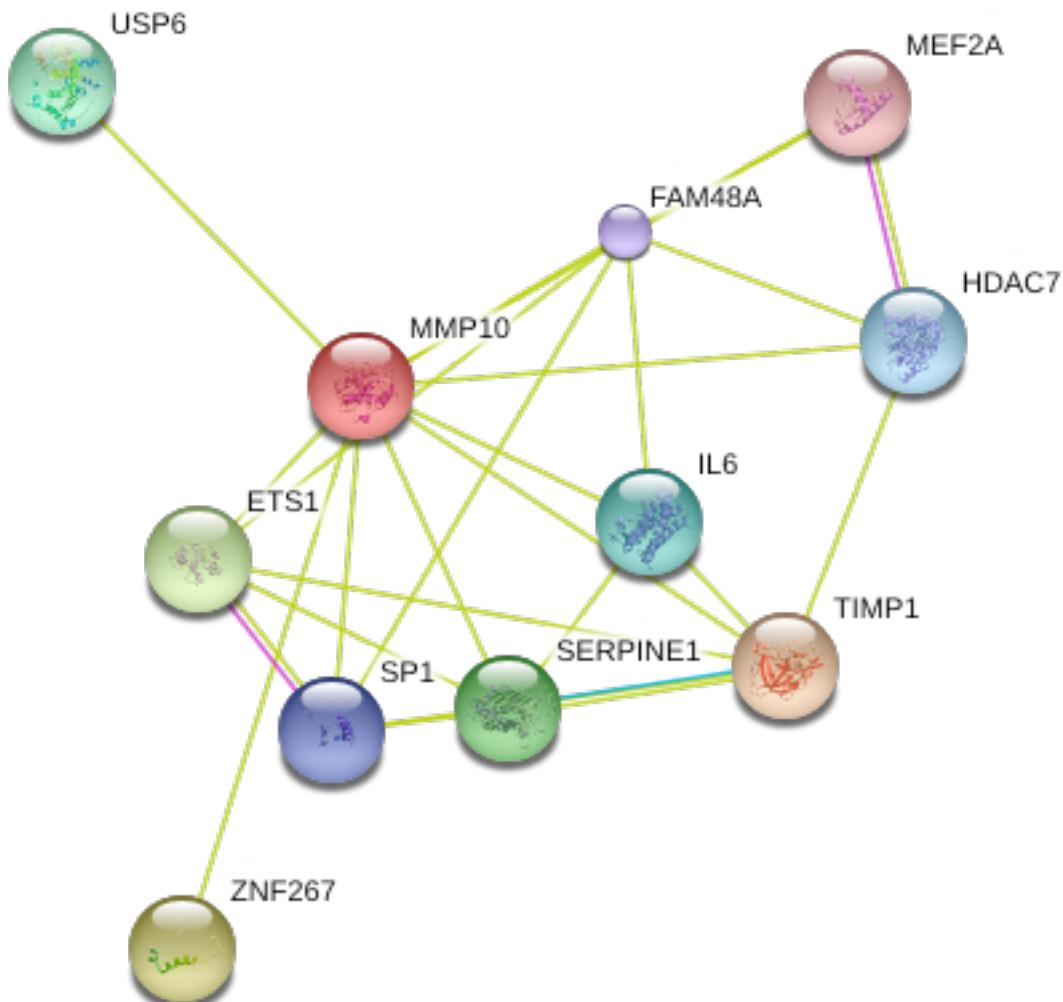
The genetic code

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

gene regulation



gene regulation



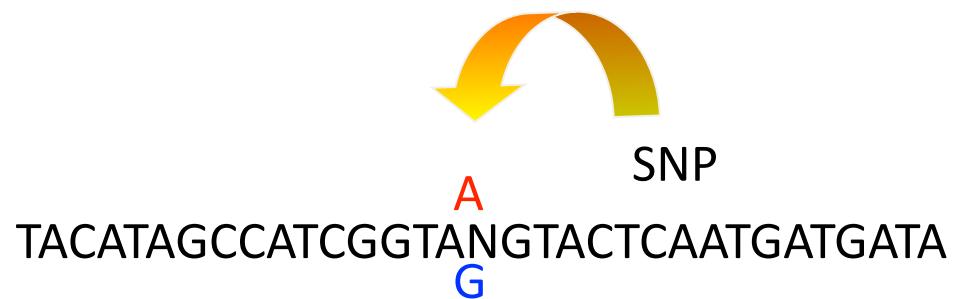
What makes them different?



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
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

Father

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

AG

Mother

TACATAGCCATCGGTAGTACTCAATGATGATA
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

Father

TACATAGCCATCGGTGGTACTCAATGATGATA
ATGTATCGGTAGCCATCCCATGAGTTACTACTAT

GG

Mother

TACATAGCCATCGGTAG**G**TACTCAATGATGATA
ATGTATCGGTAGCCAT**C**CATGAGTTACTACTAT

Father

TACATAGCCATCGGTAG**G**TACTCAATGATGATA
ATGTATCGGTAGCCAT**C**CATGAGTTACTACTAT

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

Gene A from Person 1



Gene A from Person 2

Codon change made no difference in amino acid sequence



Gene A from Person 3

Codon change resulted in a different amino acid at position 2



Health or Disease?

Person 1

DNA Sequence

A A A T T T



Normal protein



Some DNA variations have no negative effects

Person 2

A A T T T T



Low or nonfunctioning protein



Other variations lead to disease (e.g., sickle cell) or increased susceptibility to disease (e.g., lung cancer)

Person 3

A A C T T T

