

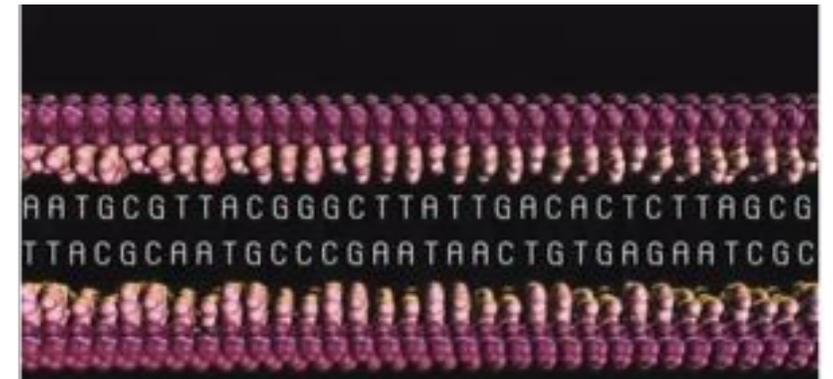
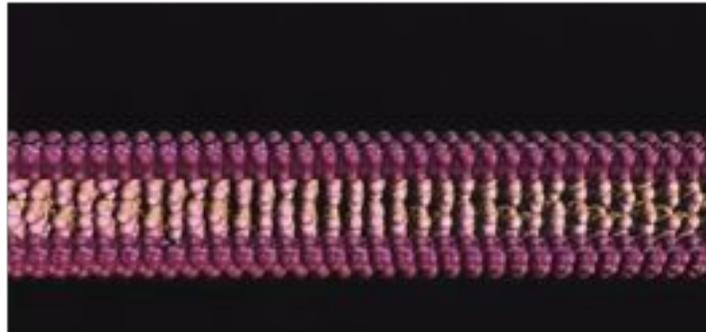
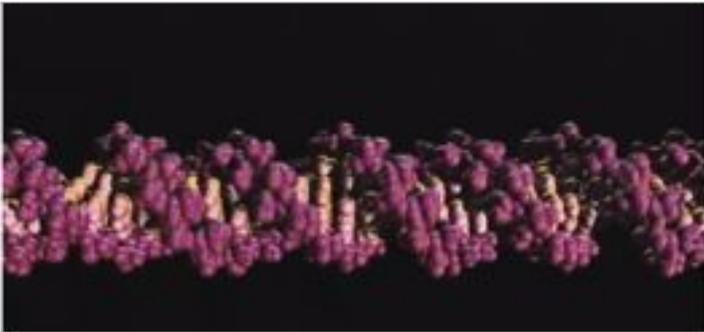
# Molecules, Central Dogma and Phenotypes

- Cells and organisms are busy, complex, 4D places — lots going on, short timelines, routine business AND response to changing situations-- need clear instructions, a lot of communication and coordination



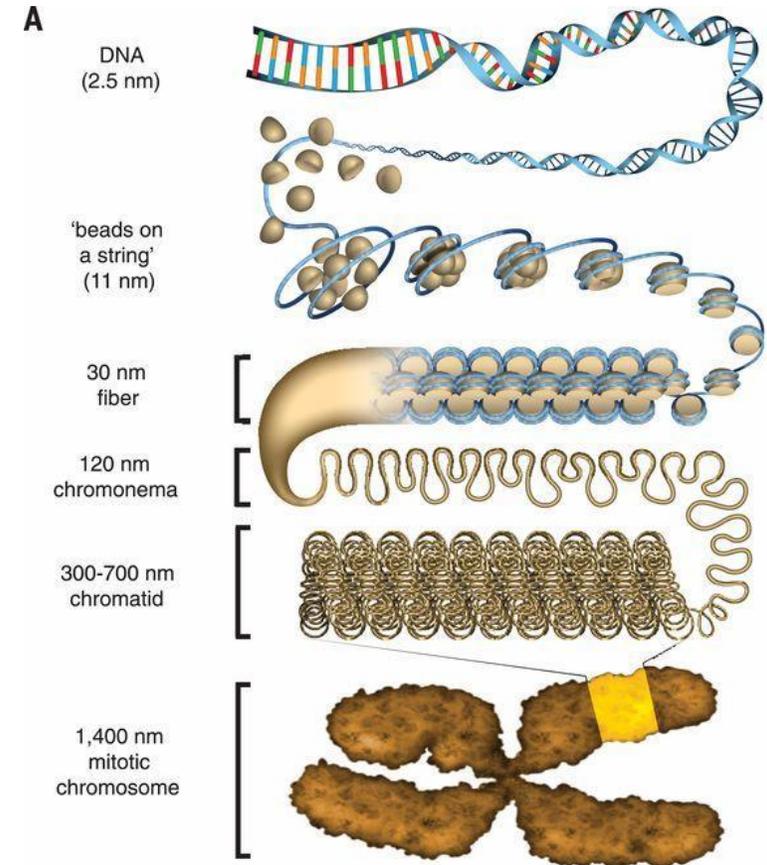
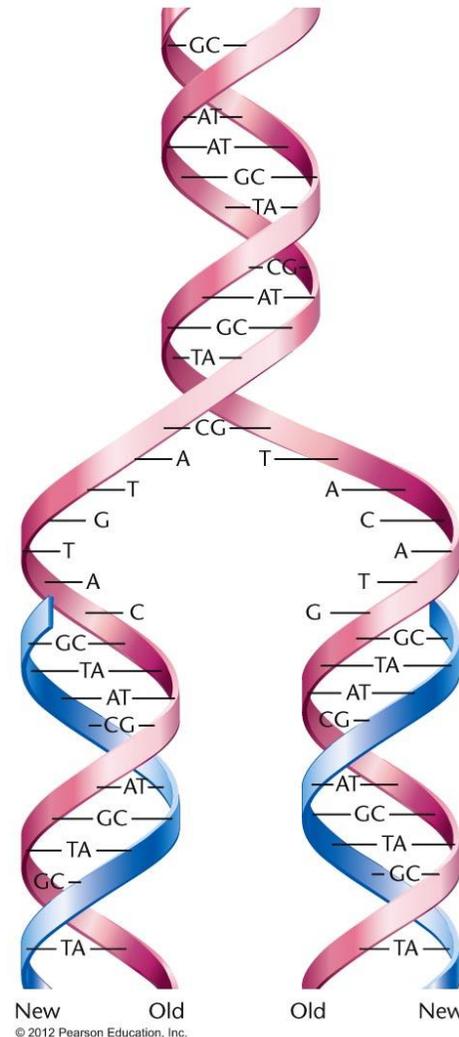
- “Genotype” is the complete set of genetic information within a cell or organism; “Phenotype” is any outward manifestation of that genotype

- **DNA** – informational molecule made of 4 building blocks (“**bases**” A, C, G, T)  
— most (not all!) information is in **1)** which building blocks and **2)** in what order
  - Double-stranded (two strings of bases that pair with one another), following VERY strict rules for pairing with each other— **A always with T** and **C always with G**; extensive machinery for monitoring and for fixing if violated
  - These rules allow faithful copying, but mistakes happen (“**mutations**”)
  - DNA generally thought of in “**base pairs**” (bp)—humans have ~3.2 billion per cell



- How strict pairing rules allow copying:
  - Separate the two strands of DNA and use each single strand to “template” an exact copy
  - Wherever A on the “old” template, put a T opposite it in the “new” strand being made; wherever there is a C in the old strand, put a G in the new strand, etc. [A—T, C—G, G—C, T—A]
- For spatial management and equal distribution into two new cells, all this DNA is tightly packaged using protein into “**chromatin**”, and discreet bundles of chromatin are called “**chromosomes**”

- DNA Learning Center  
[VIDEO: Chromosomes](#)

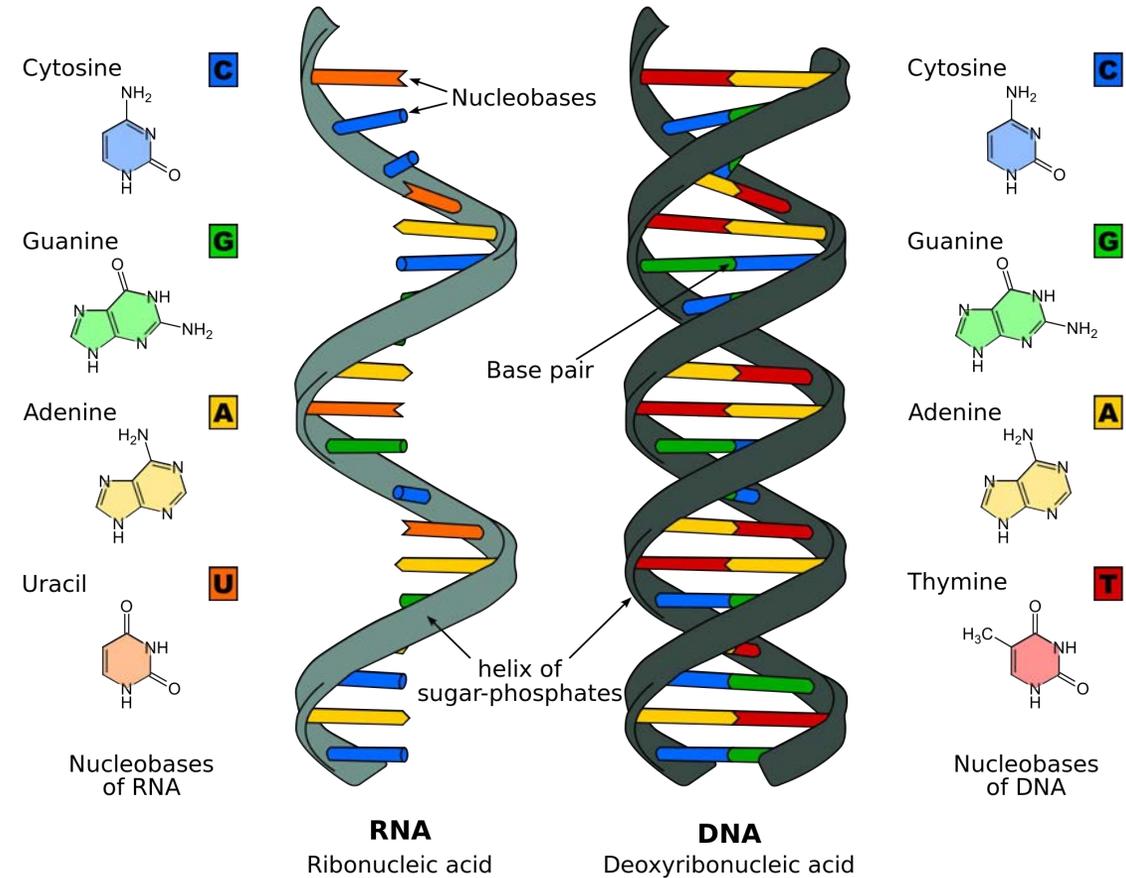


- **RNA**– another informational molecule

- Chemically similar to DNA (probably OLDER) but less stable, some different building blocks (bases), less strict base pairing rules and more flexible “backbone”
- 4 major bases (A, C, G, **U** (instead of T, that can pair with A or G) **plus >150 MORE weird variations on bases**, many that also pair with more than one other base

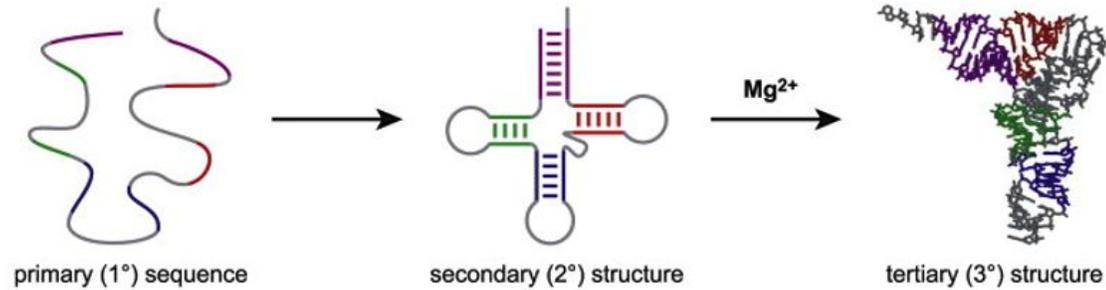
- **BOTTOM LINE?** Allows RNA to fold up into all sorts of shapes and even make double stranded parts of the same RNA molecule

- All this folding lets these RNAs not only encode things, it lets the molecules DO things (act like enzymes, carry cargoes, serve as a 3D scaffold, etc.)

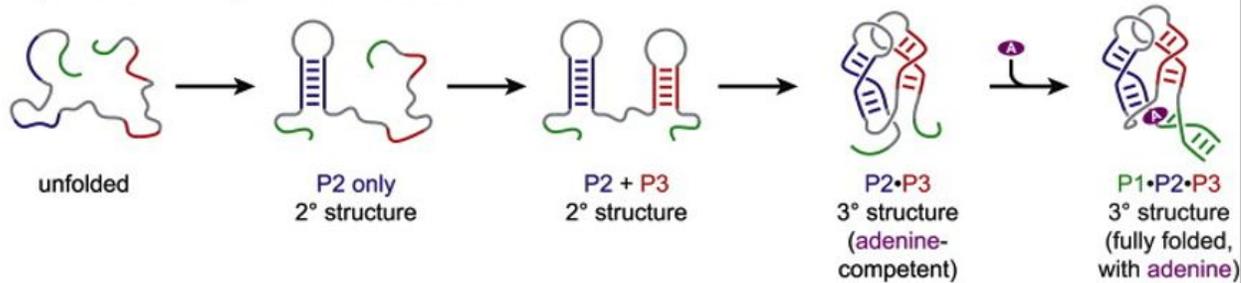


Folding into various shapes is what allows RNAs to have different functions

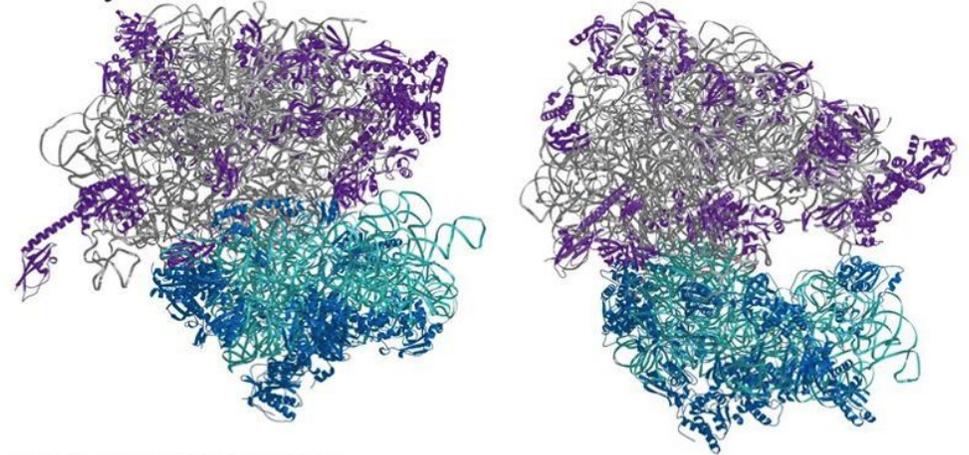
**A** Transfer RNA (yeast tRNA<sup>Phe</sup>)



**B** Aptamer domain of *pbuE* adenine riboswitch



The ribosome: Ribosomal RNAs play both a structural and catalytic role



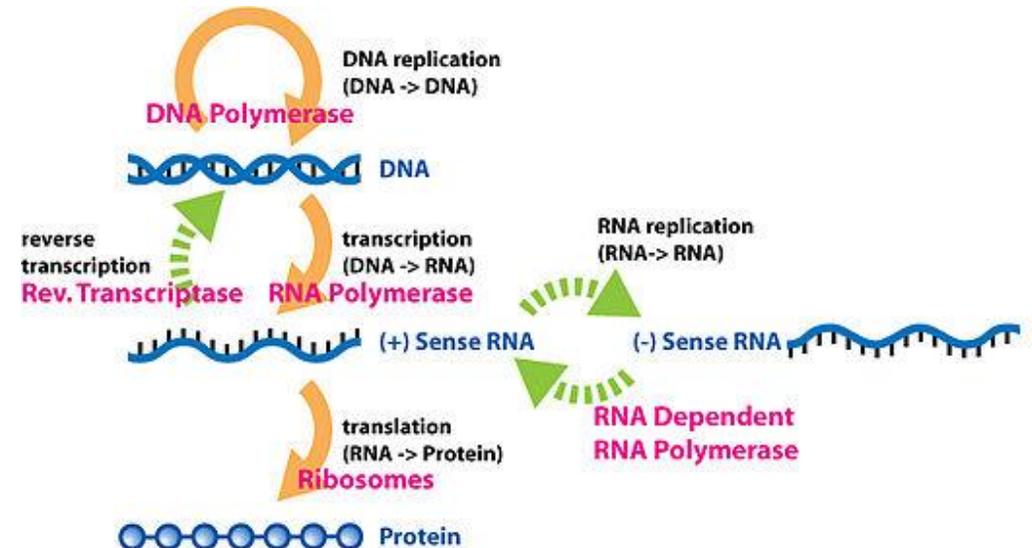
- **Protein**—also a chain of building blocks, but this time 20 of them and a different set of molecules, “amino acids”
  - Which amino acids and what order determine how a protein will fold up in 3D, how flexible it can be under different conditions
- Shape it folds into, details about surface charge, what in the cell it interacts with, when and where it is made all determine what function the protein has in a cell
  - Those functions and interactions among tens of thousands of proteins are important for determining what we see as traits and characteristics



- From information to traits: how does the information get used?
- DNA gets used as a template to make RNA, using base pairing rules
  - **TRANSCRIPTION** of DNA into an RNA copy (= “a transcript”)
  - Signals in the DNA sequence that define where, when and under what conditions TRANSCRIPTION should start
  - Notice the **green dotted arrows**? Things we have learned since the Central Dogma was first proposed!
- ~1% of RNA transcripts made are “**messenger RNA**” (mRNA)
  - mRNA gets **TRANSLATED**, three letters (bases) at a time, into a different language, a PROTEIN, following a GENETIC CODE
  - Each possible combination of three bases means “add one of the 20 different amino acids”

• **DNA** → **RNA** → **PROTEIN**

• DNA Learning Center  
[Video: Central Dogma](#)



- Triplet code (read three bases at a time = a “codon”)
  - 2 bases at a time could only specify  $4^2 = 16$  amino acids
  - Code is non-overlapping, there is a “reading frame”
    - THE BIG CAT ATE THE FAT RAT
    - T HEB IGC ATA TET HEF ATR AT
    - TH EBI GCA TAT ETH EFA TRA T
- Starts at a ***fixed point***, not randomly within each gene (establish ***correct*** reading frame)
  - AUG = “START” as well as MET; also notice 3 codons do not encode any amino acid = “STOP”
- This code is “degenerate” (more than one codon for most amino acids; 64 3-base combos, 20 amino acids)

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

- A translatable sequence begins with a start codon and ends with a stop codon
  - Typically numerous “out of frame” stop codons to prevent accidental translation of incorrect proteins
- Adding or deleting bases (“frameshift mutations”) can change reading frame;
  - THE BIG CAR TAT ETH EFA TRA T
  - Translation runs into a stop codon very quickly

Putative ORF protein sequence

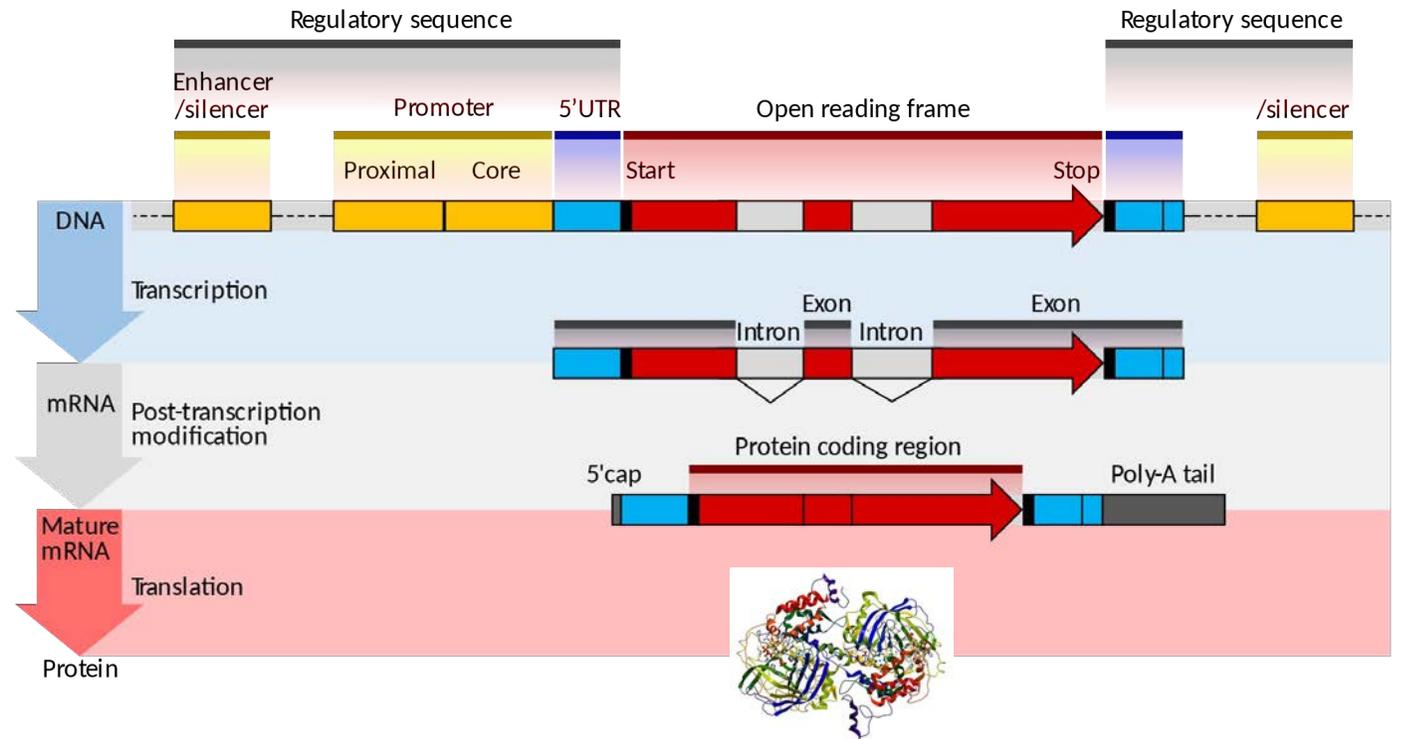
Cys Val Arg Met Val Ser Ala Ile Cys Ser Phe Phe Val Phe Tyr Asn Arg Glu Arg Arg His Phe Leu Arg Lys Gly Gly Thr Cys Arg Lys Met Glu Asn Ile Leu Asp Leu  
 Trp Asn Gln Ala Leu Ala Gln Ile Glu Lys Lys Leu Ser Lys Pro Ser Phe Glu Thr Trp Met Lys Ser Thr Lys Ala His Ser Leu Gln Gly Asp Thr Leu Thr Ile Thr Ala Pro

Putative protein is 477 amino acids long.

3 letter code Copy to clipboard

- **GENE**– any sequence of DNA or RNA that encodes a product

- Defined by specific features
- Signals that start and stop
  - Transcription
  - Translation
- Splicing of RNA
  - Parts that get removed = “introns”
  - Parts that are kept = “exons”



- **GENOME**– usually means all of the DNA in a cell (note: some genomes are RNA)

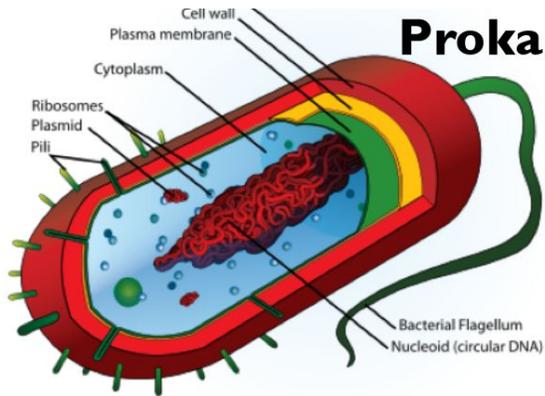
- “TRANSCRIPT-OME”, “PROTE-OME”, “METABOL-OME”, “THUNDERD-OME”
- **Also “PanGenome”**



- “Sequencing a genome” means determining the precise order of ALL of the bases in an organism—all of them. In you, about 3 billion. But that is just the beginning of the story
  - How they are packaged matters as well (remember “chromatin”?)
  - Imagine DNA sequences are letters that make words and sentences. Think of each gene as a sentence. Which sentences are **READ together (“co-expressed”)** frame different stories (for example, different cell types). Defining which sentences are **READABLE** can have the same impact, and THAT is Epigenetics and the Epigenome.
- 
- The sun rose gently in the sky
  - The horses grazed peacefully in the meadow
- 
- The smoke rose gently into the sky

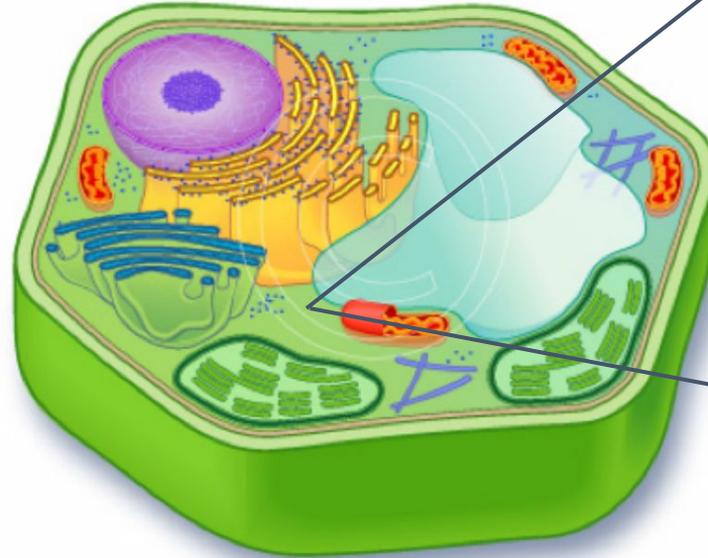
- “Sequencing a genome” means determining the precise order of ALL of the bases in an organism—all of them. In you, about 3 billion. But that is just the beginning of the story
  - How they are packaged matters as well (remember “chromatin”?)
  - Imagine DNA sequences are letters that make words and sentences. Think of each gene as a sentence. Which sentences are **READ together (“co-expressed”)** frame different stories (for example, different cell types). Defining which sentences are **READABLE** can have the same impact, and THAT is Epigenetics and the Epigenome.
- 
- The sun rose gently in the sky
- 
- The cow knocked over the lantern and the house burned to the ground.
  - The smoke rose gently into the sky

- **NUCLEUS** – a membrane-bound compartment in a EUKARYOTIC cell that contains the chromosomes and where most transcription takes place; mitochondria and chloroplasts have DNA, too

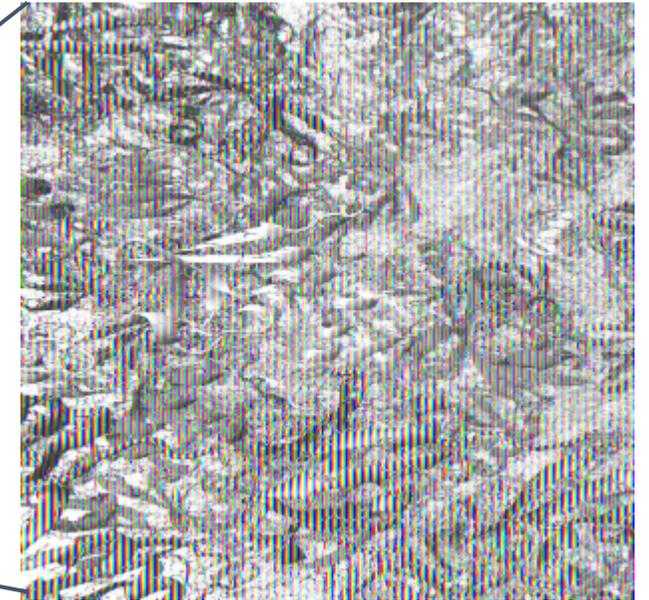


**Prokaryotic Cell (Bacteria)**

**Eukaryotic Cell (Plant)**

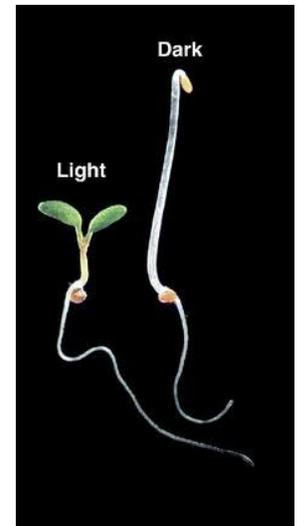


**Eukaryotic Cell (Animal)**



**“Cell Atlas”**

- **GENOTYPE** = The collective genetic information in an organism
- **PHENOTYPE** = Any outward manifestation of the genotype
  - Typically thought of as a physical trait, BUT...
  - ...Can also be alteration of a biochemical compound, presence of a behavior, a misfolded protein, absence of a peak in a chromatogram or a band in a gel, etc.
- **PHENOTYPE** results from the interaction of a **GENOTYPE** with the **ENVIRONMENT**
  - That means that the **SAME GENOTYPE** can result in **DIFFERENT PHENOTYPES**



# Molecules, Central Dogma and Phenotypes

- Cells and organisms are busy, complex, 4D places — lots going on, short timelines, routine business AND response to changing situations-- need clear instructions, a lot of communication and coordination
-  DNA → RNA → **PROTEIN**
- “Genotype” is the complete set of genetic information within a cell or organism; “Phenotype” is any outward manifestation of that genotype

- 1) Describe an example (that is NOT the one you just saw) in which two identical genotypes produce different phenotypes.
- 2) Which do you think might have a more severe effect on a gene, changing a single base from a G to an A or removing the G altogether? Explain your answer.
- 3) True story...Humans have recently synthesized four new, ADDITIONAL DNA bases (P, Z, B and S) that have similar, strict base pairing rules among themselves (P -- Z and B -- S). What do you think an eight base DNA code has the potential to allow?
- 4) Gene annotation software looks for particular DNA sequence characteristics occurring together to define genes *de novo*. What are potential pros and cons of this approach?