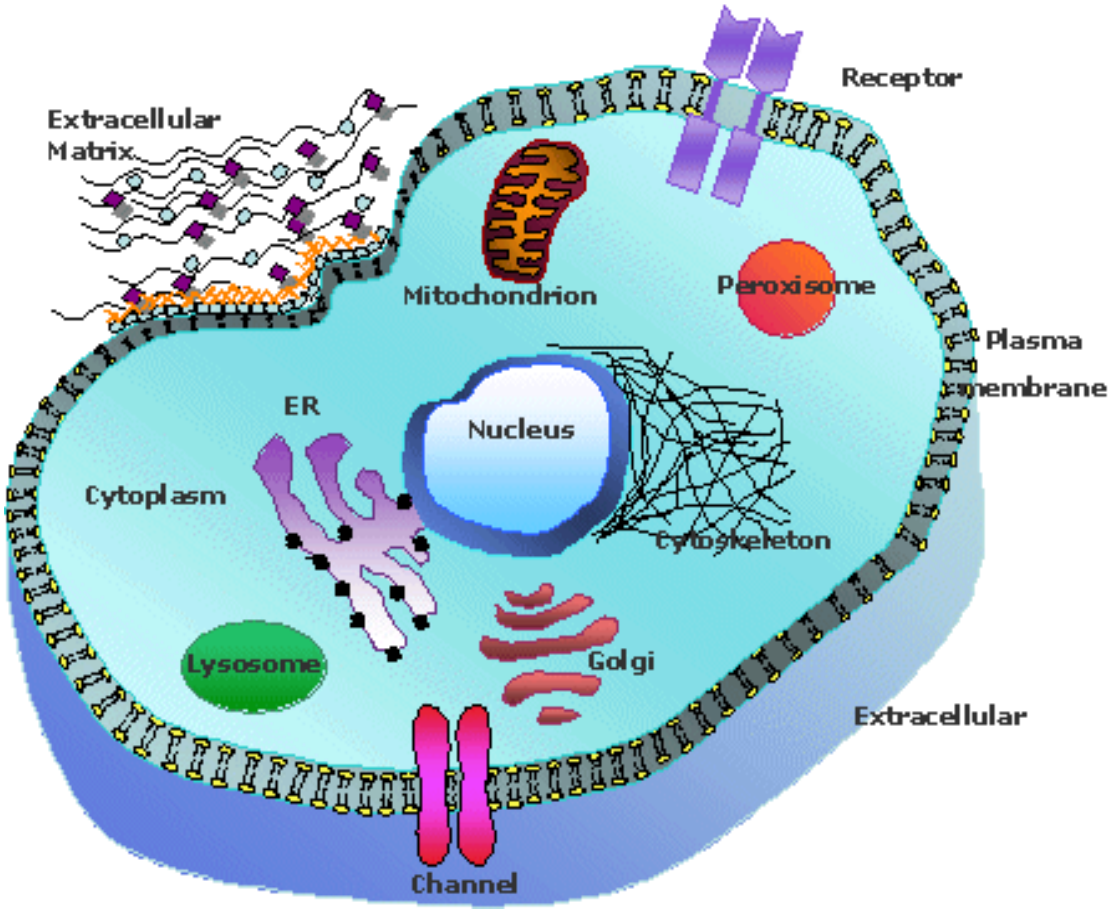
A Zero-Knowledge Based Introduction to Biology

Konstantinos (Gus) Katsiapis
25 Sep 2009

Thanks to Cory McLean and George Asimenos

Cells: Building Blocks of Life

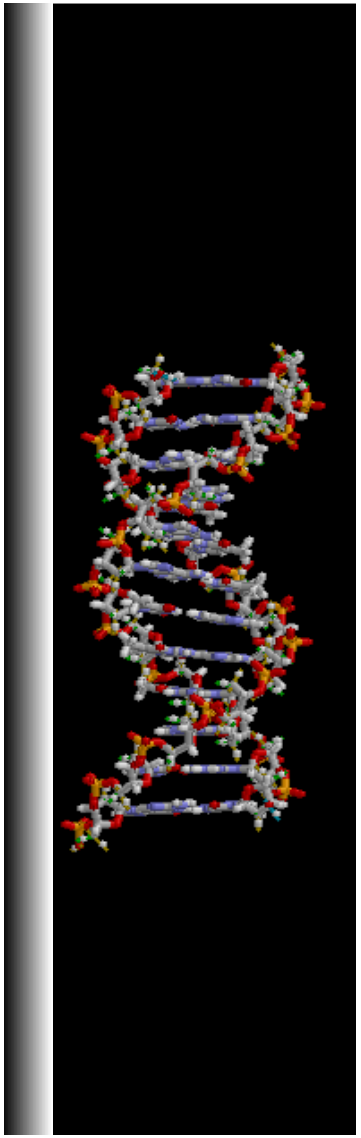
cell, membrane, cytoplasm, nucleus, mitochondrion



DNA: “Blueprints” for a cell

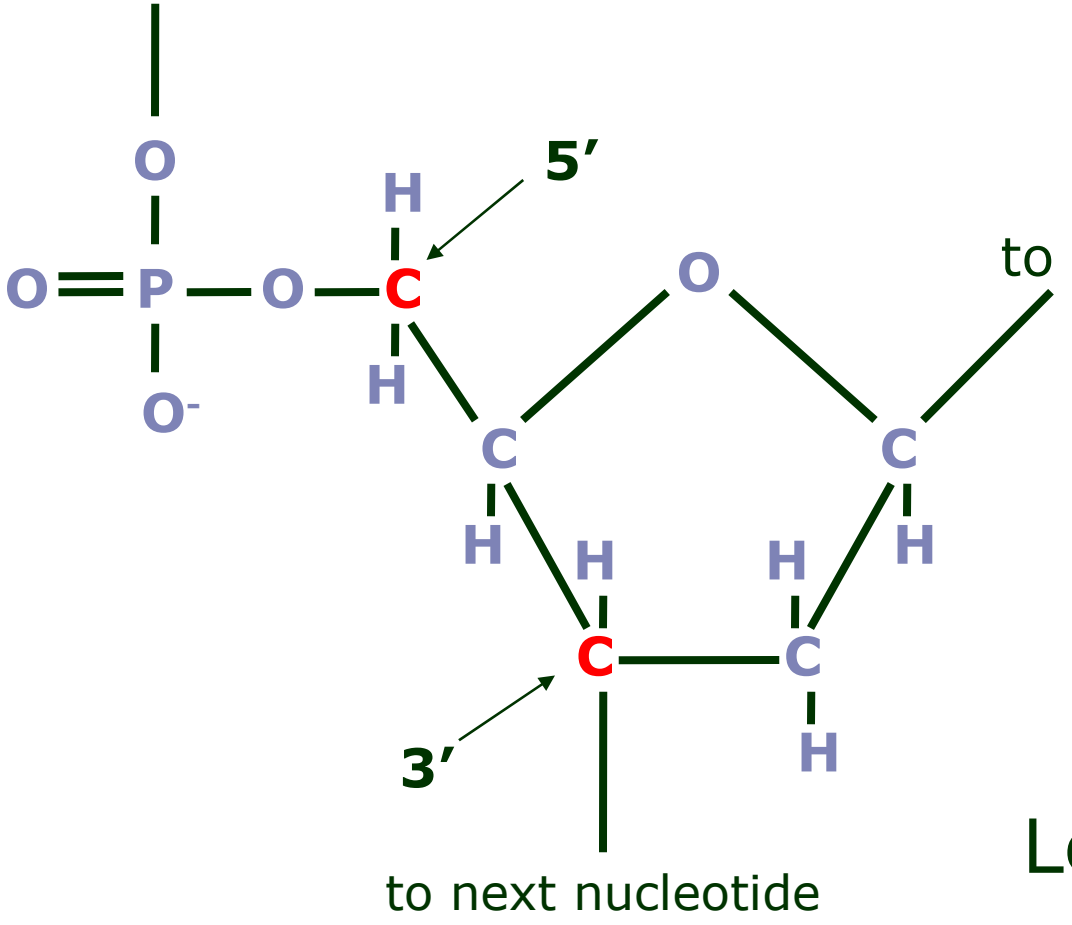
- Genetic information encoded in long strings of double-stranded DNA

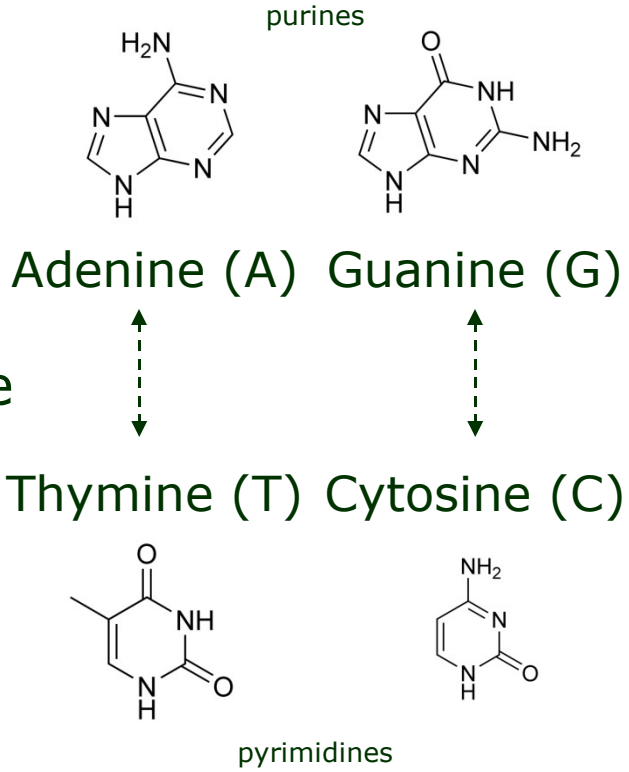
- DeoxyriboNucleic Acid comes in only four flavors: Adenine, Cytosine, Guanine, Thymine



Nucleotide

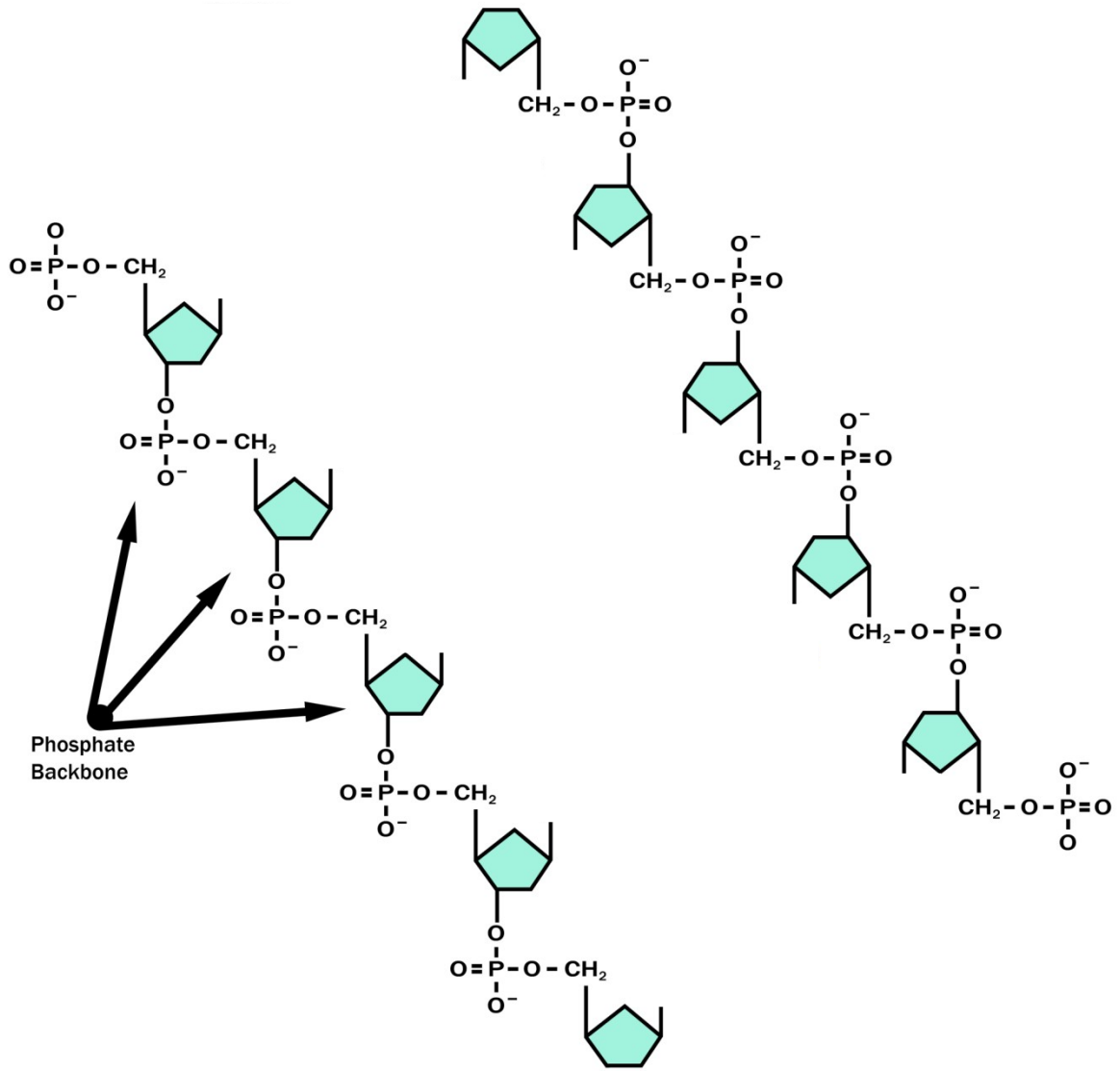
deoxyribose, nucleotide, base, purine (A,G), pyrimidine (T,C), 3’, 5’





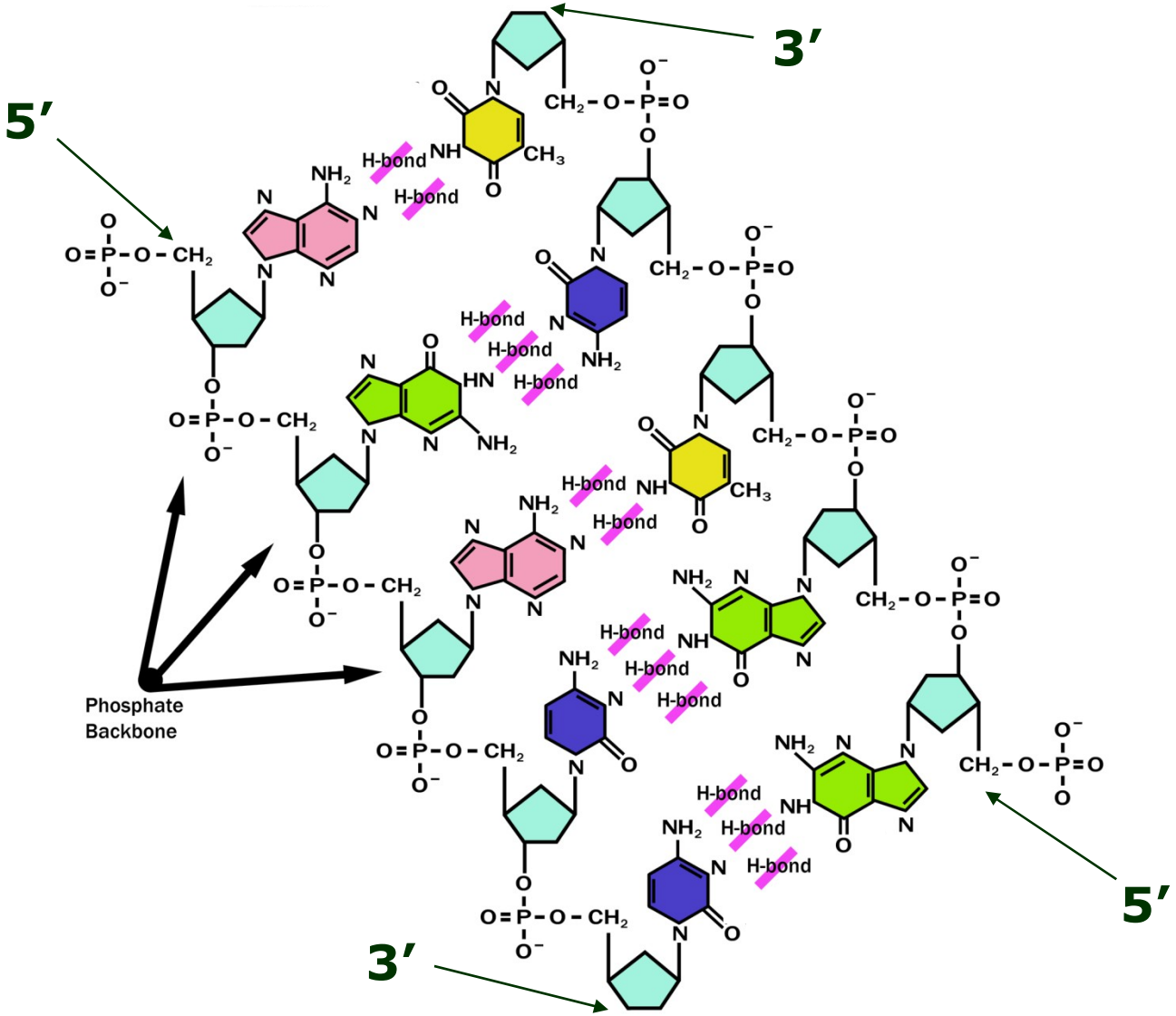
Let’s write “AGACC”!

“AGACC” (backbone)



“AGACC” (DNA)

deoxyribonucleic acid (DNA)



DNA is double stranded

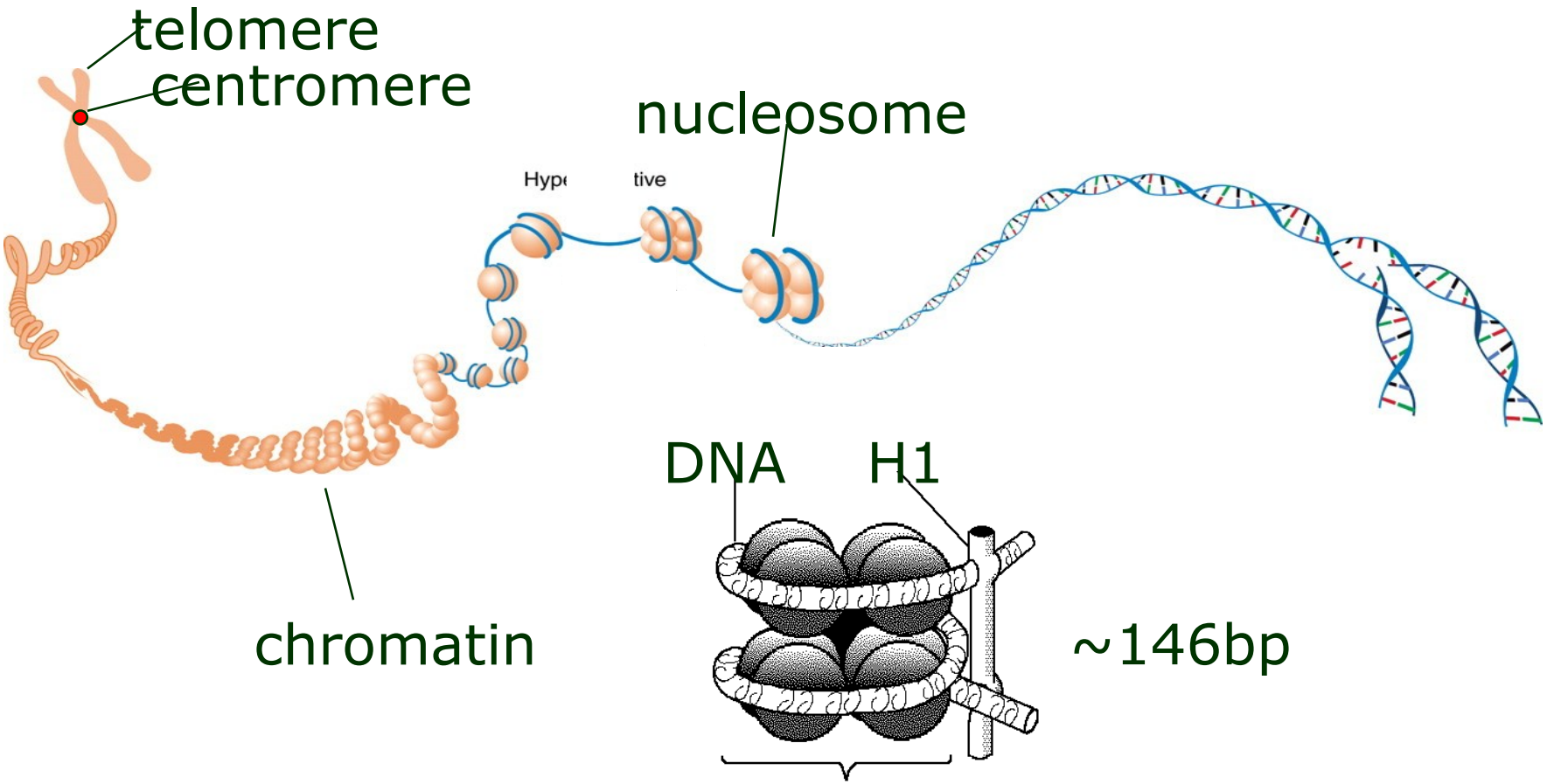
strand, reverse, complement, reverse-complement

DNA is always written 5’ to 3’

AGACC or GGTCT

DNA Packaging

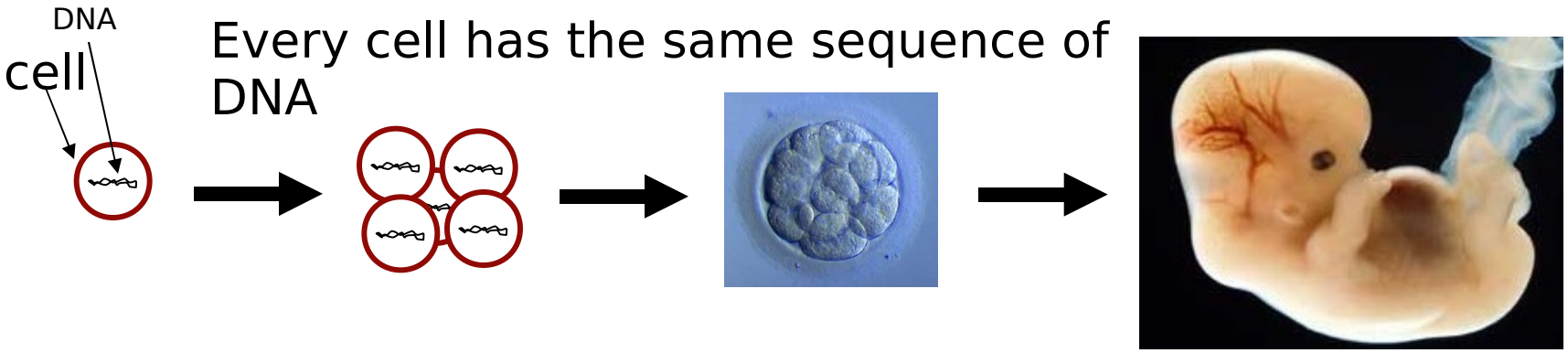
histone, nucleosome, chromatin, chromosome, centromere, telomere



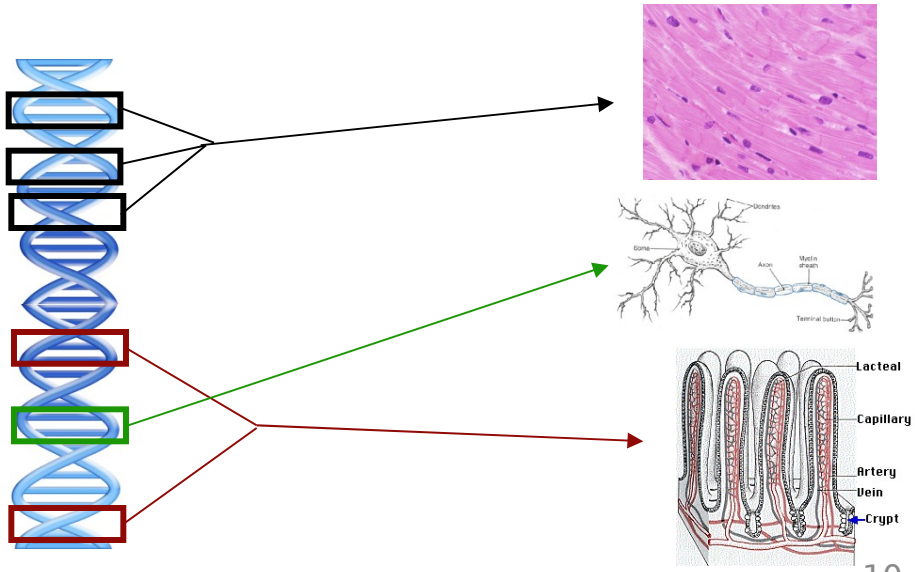
The Genome

- The genome is the full set of hereditary information for an organism
- Humans bundle two copies of the genome into 46 chromosomes in **every cell**
  - = 2 * (1-22 + X/Y)

Building an Organism

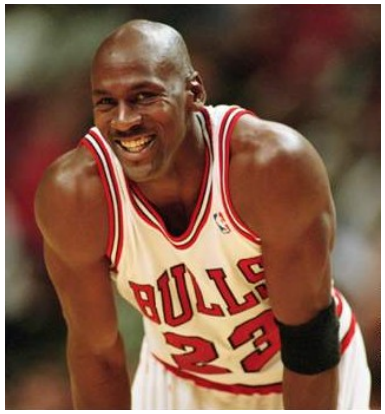


Subsets of the DNA sequence determine the identity and function of different cells

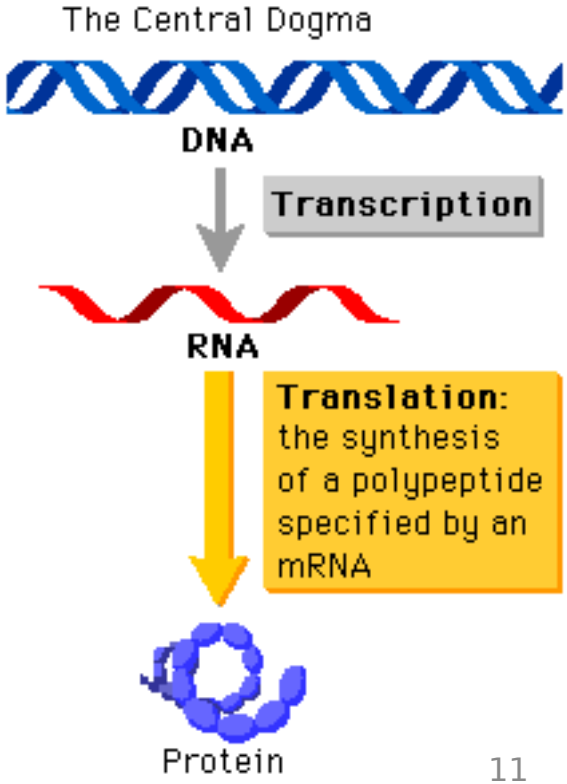


From DNA To Organism

?

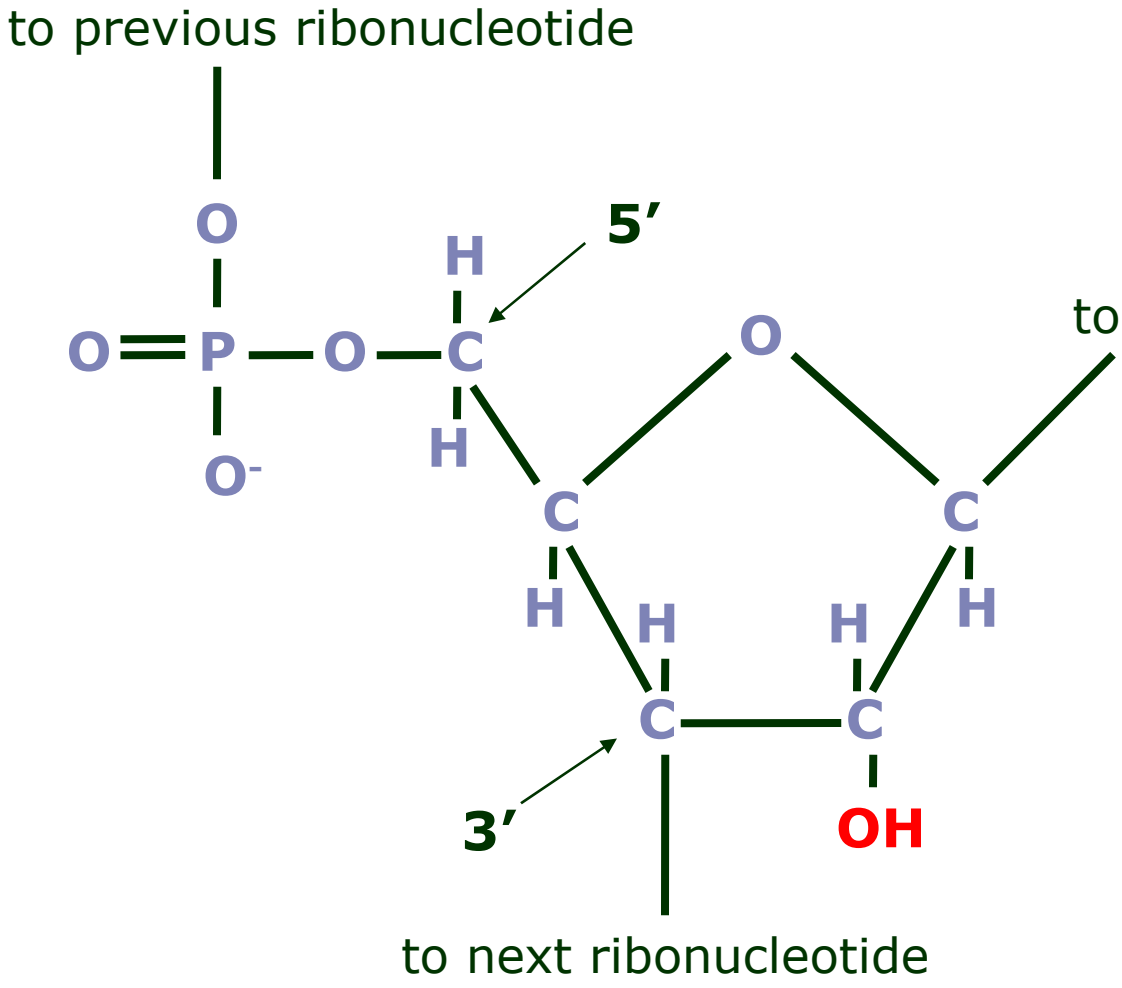


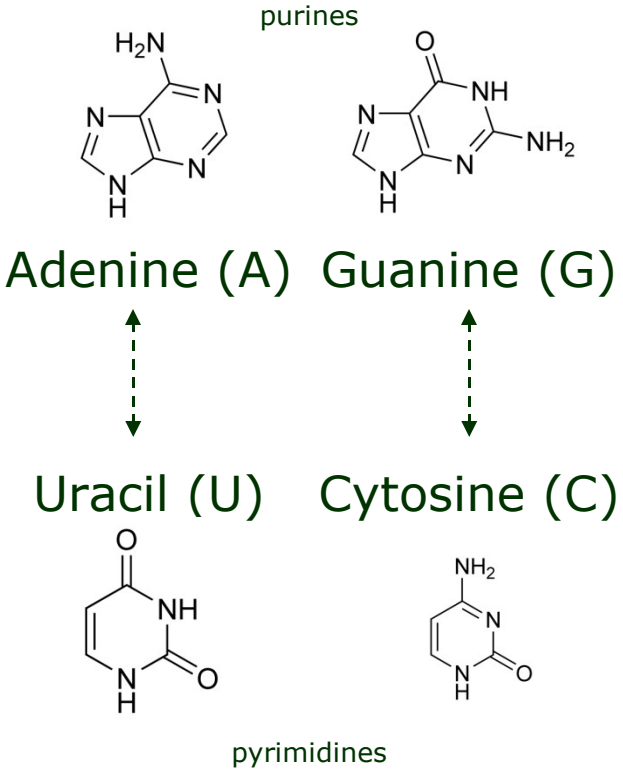
Proteins do most of the work in biology, and are encoded by subsequences of DNA, known as genes.



RNA

ribose, ribonucleotide, U





T → U

Genes & Proteins

gene, transcription, translation, protein

Double-stranded DNA

5′ TAGGATCGACTATATGGGATTACAAAGCATTTAGGGA...TCACCCTCTCTAGACTAGCATC 3′

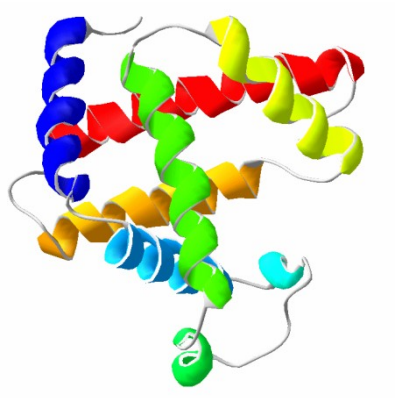
3′ ATCCTAGCTGATATACCCTAATGTTTCGTAAATCCCT...AGTGGGAGAGATCTGATCGTAG 5′

(transcription)

Single-stranded RNA AUGGGAUUACAAAGCAUUUAGGGA...UCACCCUCUCUAGACUAGCAUC

(translation)

protein



Gene Transcription

promoter

Promoter: A region of DNA facilitating transcription of a gene. Usually located one the same strand, upstream and nearby.

Gene Transcription

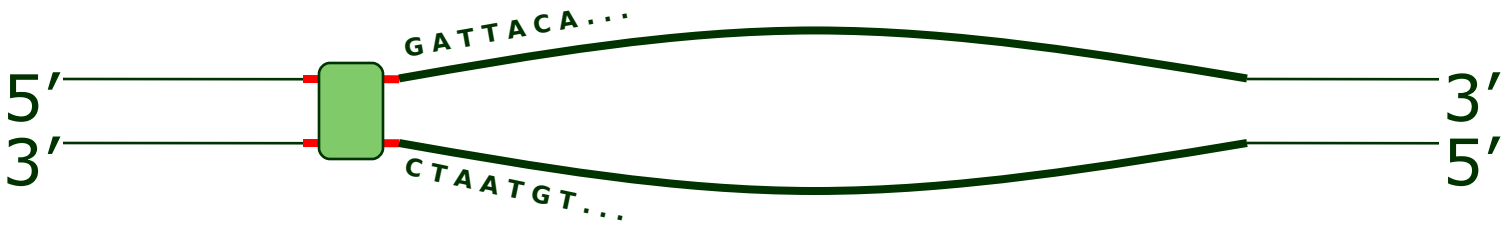
transcription factor, binding site, RNA polymerase

Transcription factors: a type of protein that binds to DNA and helps initiate gene transcription.

Transcription factor binding sites: Short sequences of DNA (6-20 bp) recognized and bound by TFs.

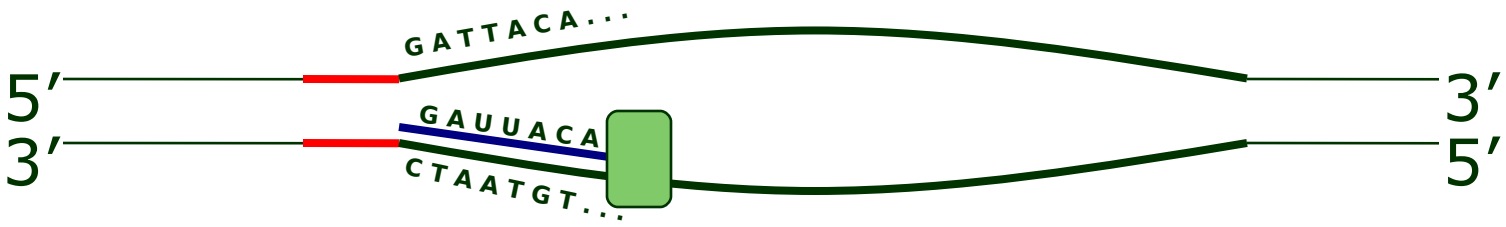
RNA polymerase binds a complex of TFs in the promoter.

Gene Transcription



The two strands are separated

Gene Transcription

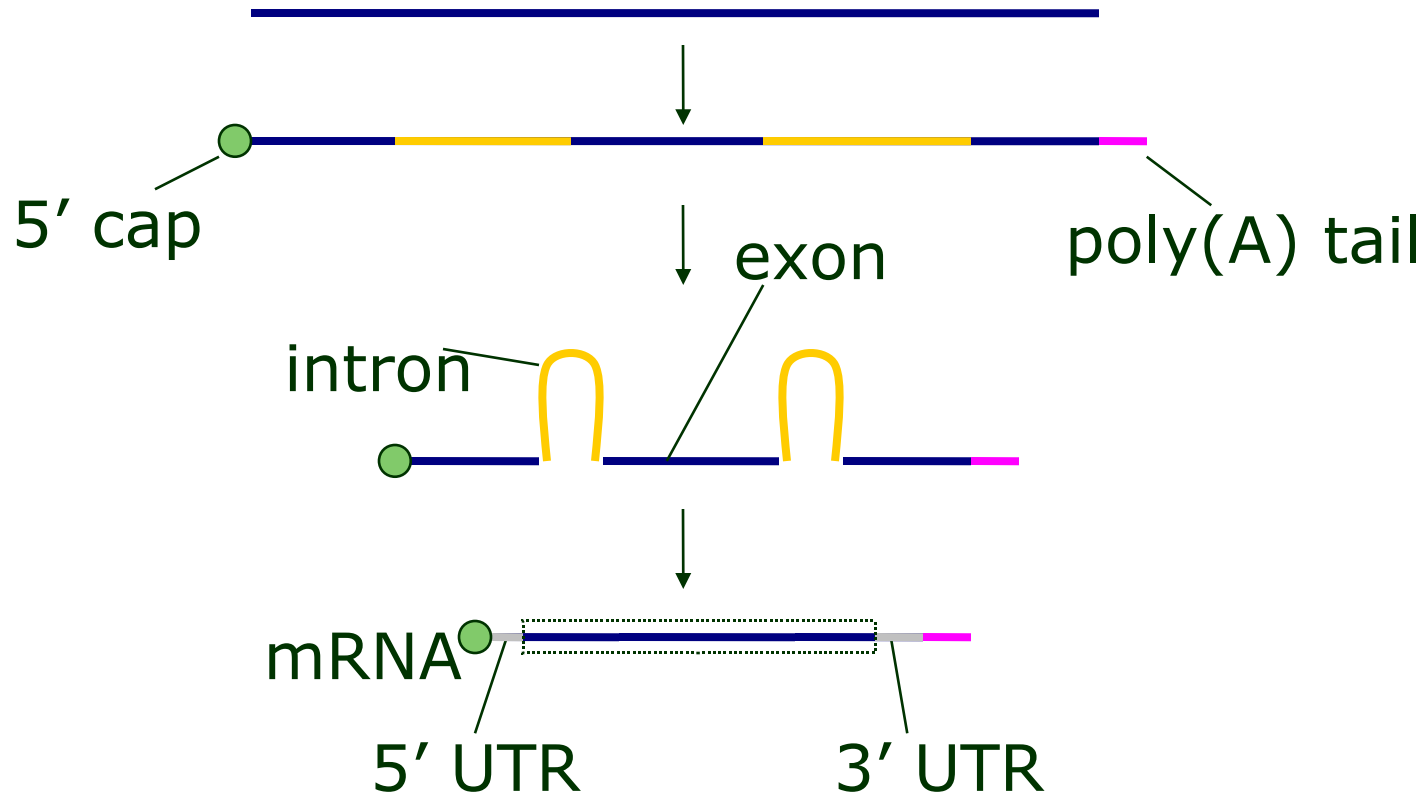


An RNA copy of the 5′→3′ sequence is created from the 3′→5′ template

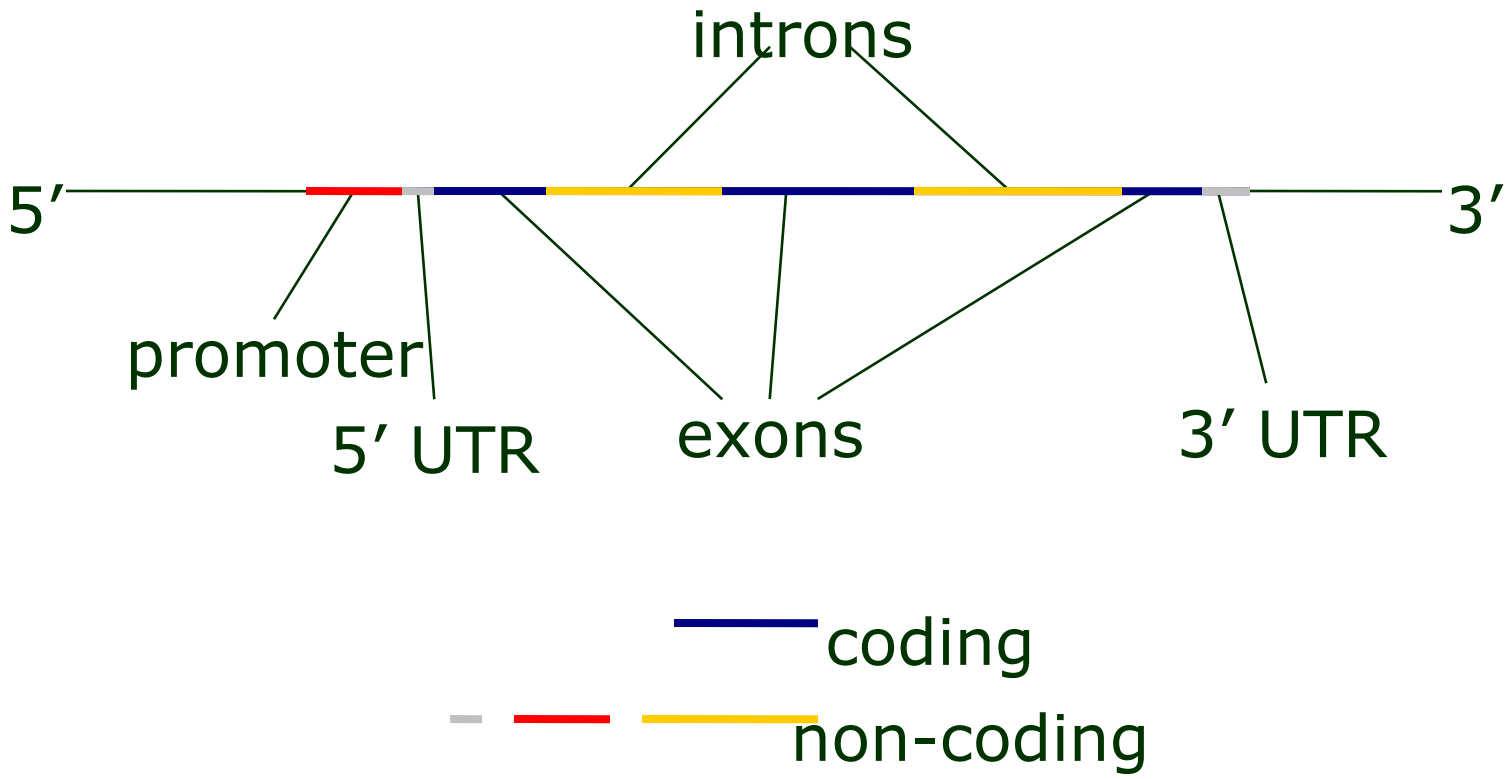
Gene Transcription

RNA Processing

5’ cap, polyadenylation, exon, intron, splicing, UTR, mRNA



Gene Structure



How many?
(Human Genome)

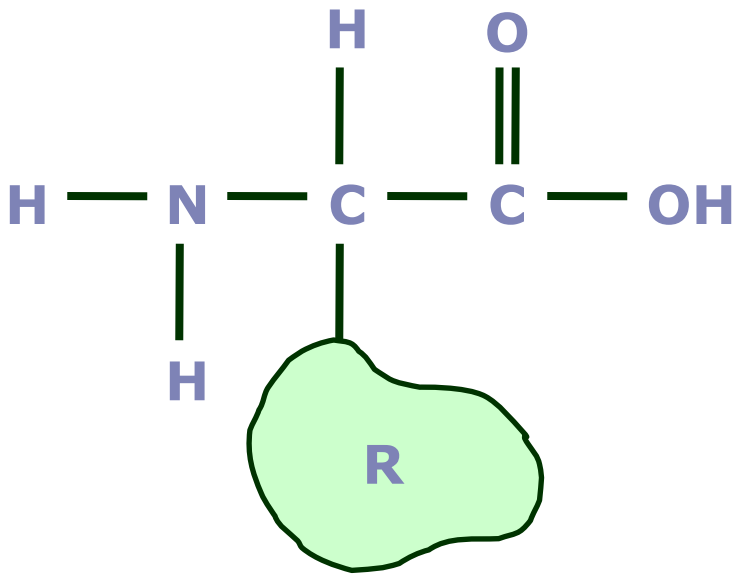
- Exons per gene:
  ~ 8 on average (max: 148)
- Nucleotides per exon:
  170 on average (max: 12k)
- Nucleotides per intron:
  5,500 on average (max: 500k)
- Nucleotides per gene:
  45k on average (max: 2,2M)

From RNA to Protein

- Proteins are long strings of amino acids joined by peptide bonds
- Translation from RNA sequence to amino acid sequence performed by ribosomes
- 20 amino acids → 3 RNA letters required to specify a single amino acid

Amino acid

amino acid

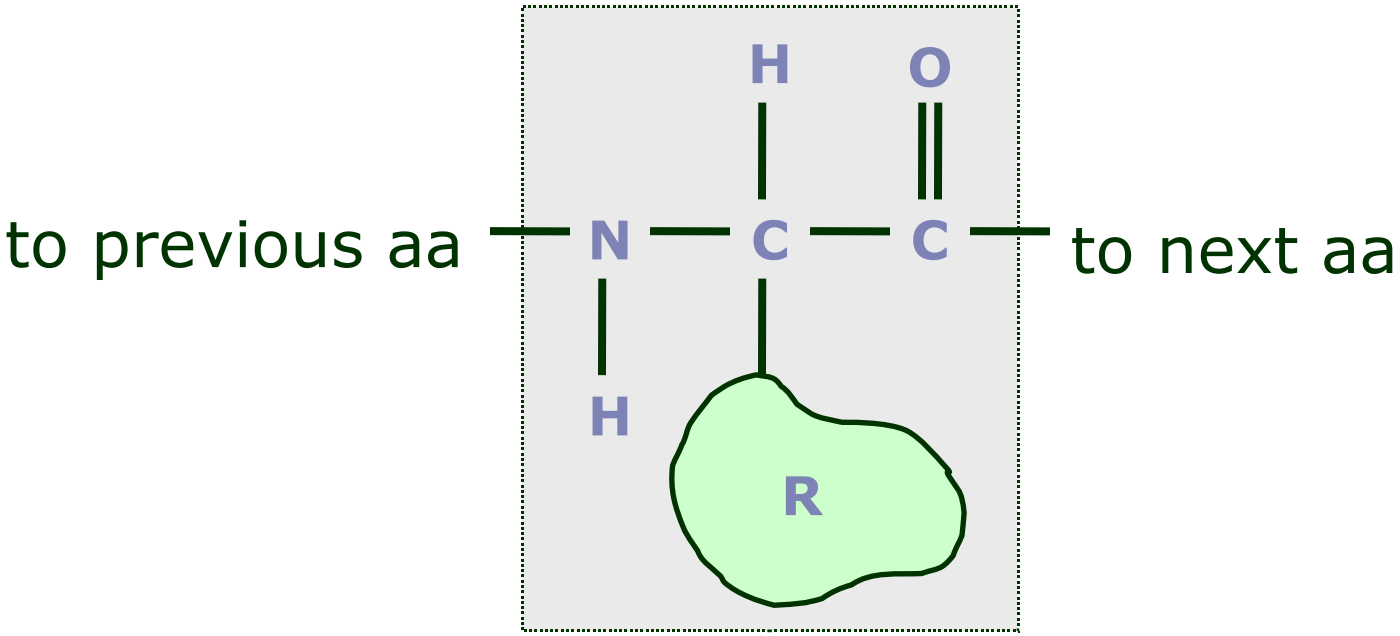


Alanine
Arginine
Asparagine
Aspartate
Cysteine
Glutamate
Glutamine
Glycine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Proline
Serine
Threonine
Tryptophan
Tyrosine
Valine

There are 20 standard amino acids

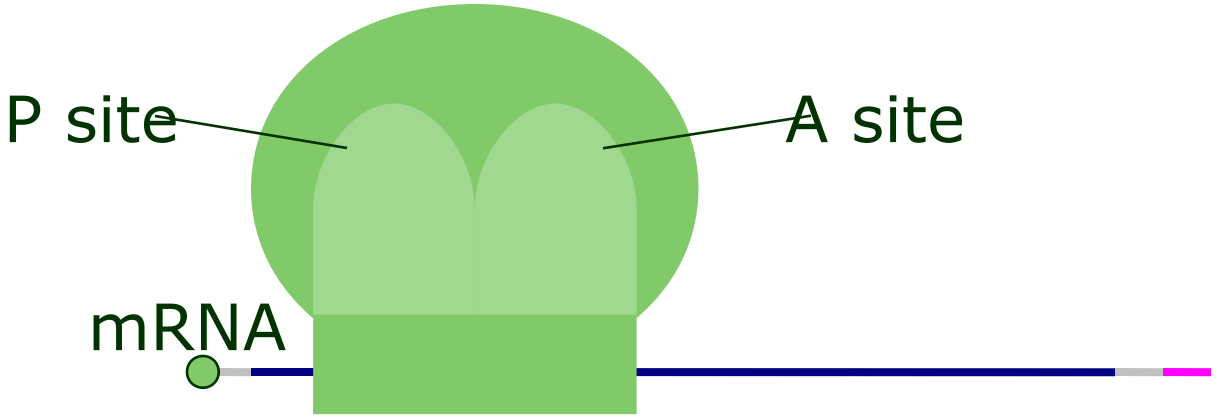
Proteins

N-terminus, C-terminus



Translation

ribosome, codon

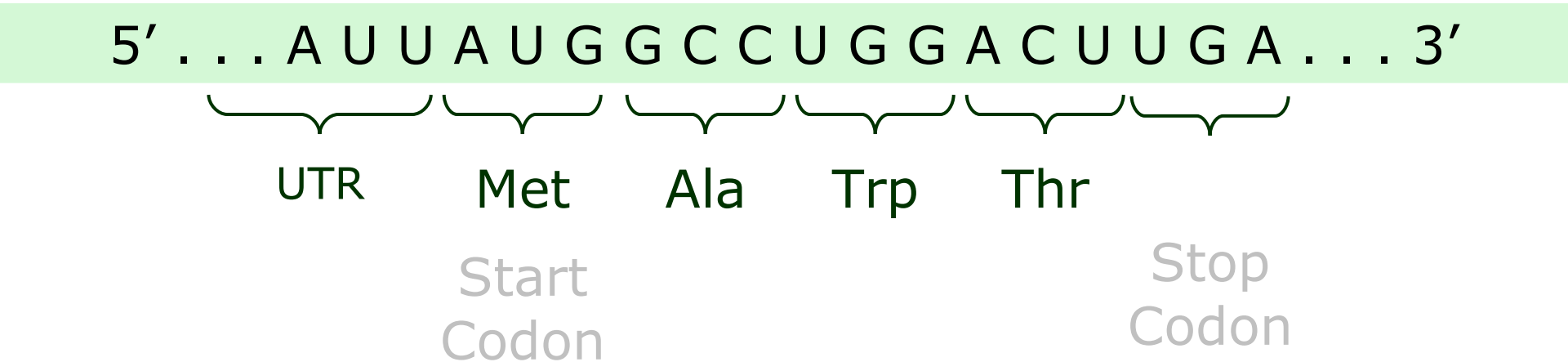


The ribosome (a complex of protein and RNA) synthesizes a protein by reading the mRNA in triplets (codons). Each codon is *translated* to an amino acid.

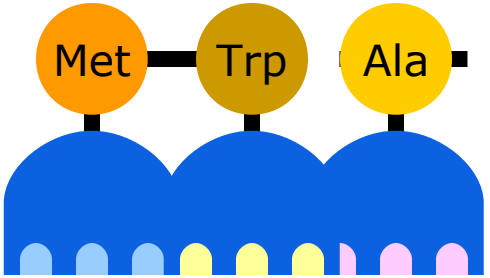
Translation

	U	C	A	G	
U	UUU **Phenylalanine** (Phe)	UCU **Serine** (Ser)	UAU **Tyrosine** (Tyr)	UGU **Cysteine** (Cys)	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA **Leucine** (Leu)	UCA Ser	UAA **STOP**	UGA **STOP**	**A**
	UUG Leu	UCG Ser	UAG **STOP**	UGG **Tryptophan** (Trp)	**G**
C	CUU **Leucine** (Leu)	CCU **Proline** (Pro)	CAU **Histidine** (His)	CGU **Arginine** (Arg)	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA **Glutamine** (Gln)	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU **Isoleucine** (Ile)	ACU **Threonine** (Thr)	AAU **Asparagine** (Asn)	AGU **Serine** (Ser)	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA **Lysine** (Lys)	AGA **Arginine** (Arg)	**A**
	AUG **Methionine** (Met) or **START**	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU **Valine** (Val)	GCU **Alanine** (Ala)	GAU **Aspartic acid** (Asp)	GGU **Glycine** (Gly)	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA **Glutamic acid** (Glu)	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

Translation



Translation



Transcription and Translation [3D Simulation]

Errors?

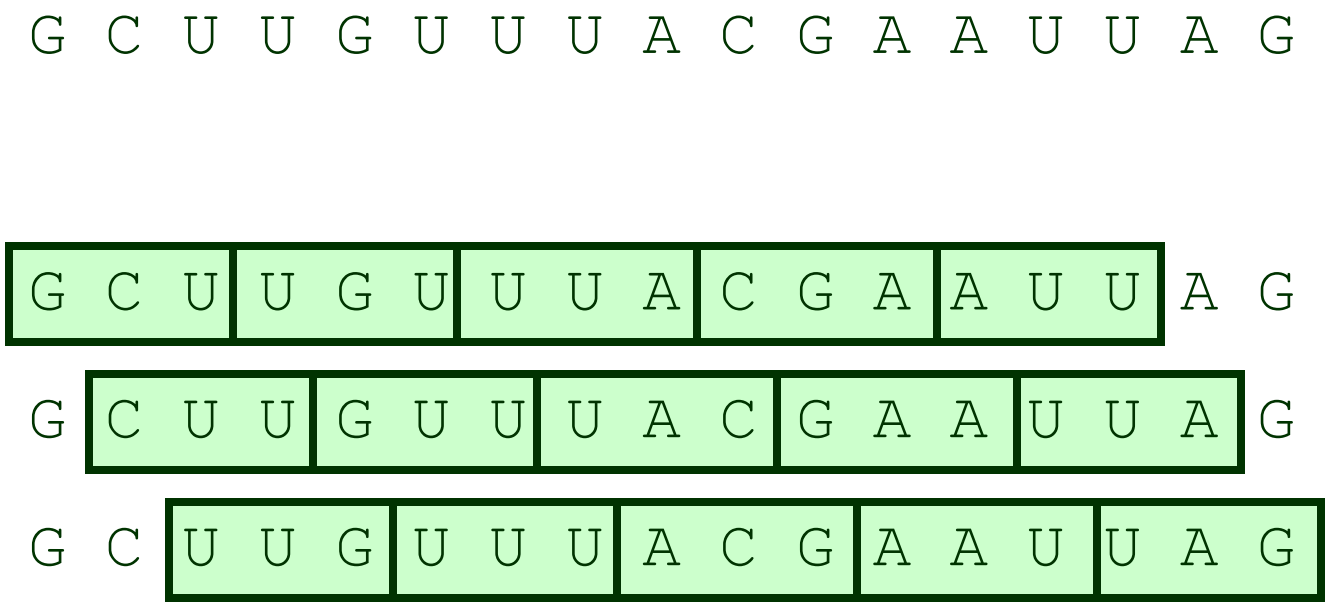
mutation

- What if the transcription / translation machinery makes mistakes?

- What is the effect of **mutations** in coding regions?

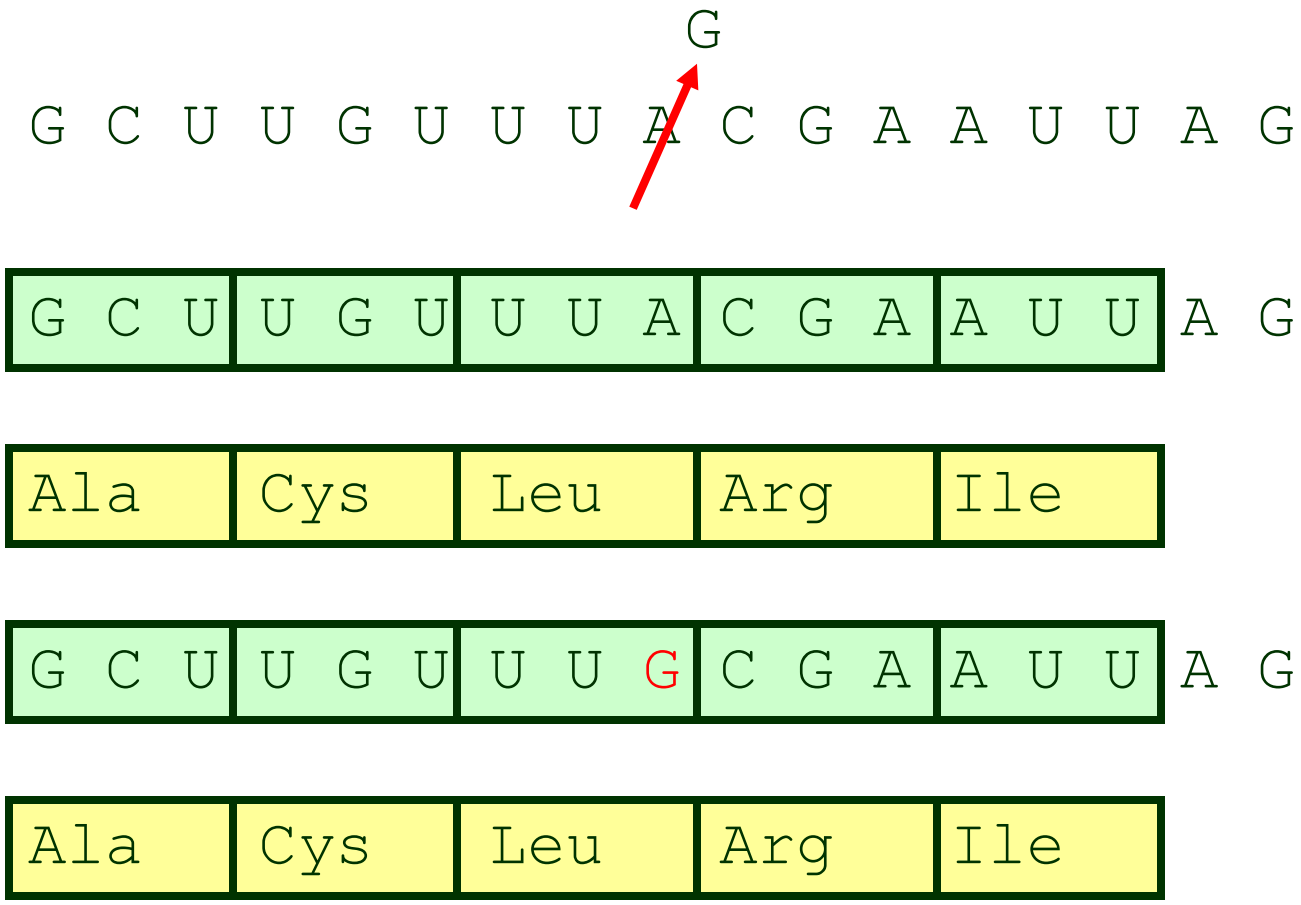
Reading Frames

reading frame



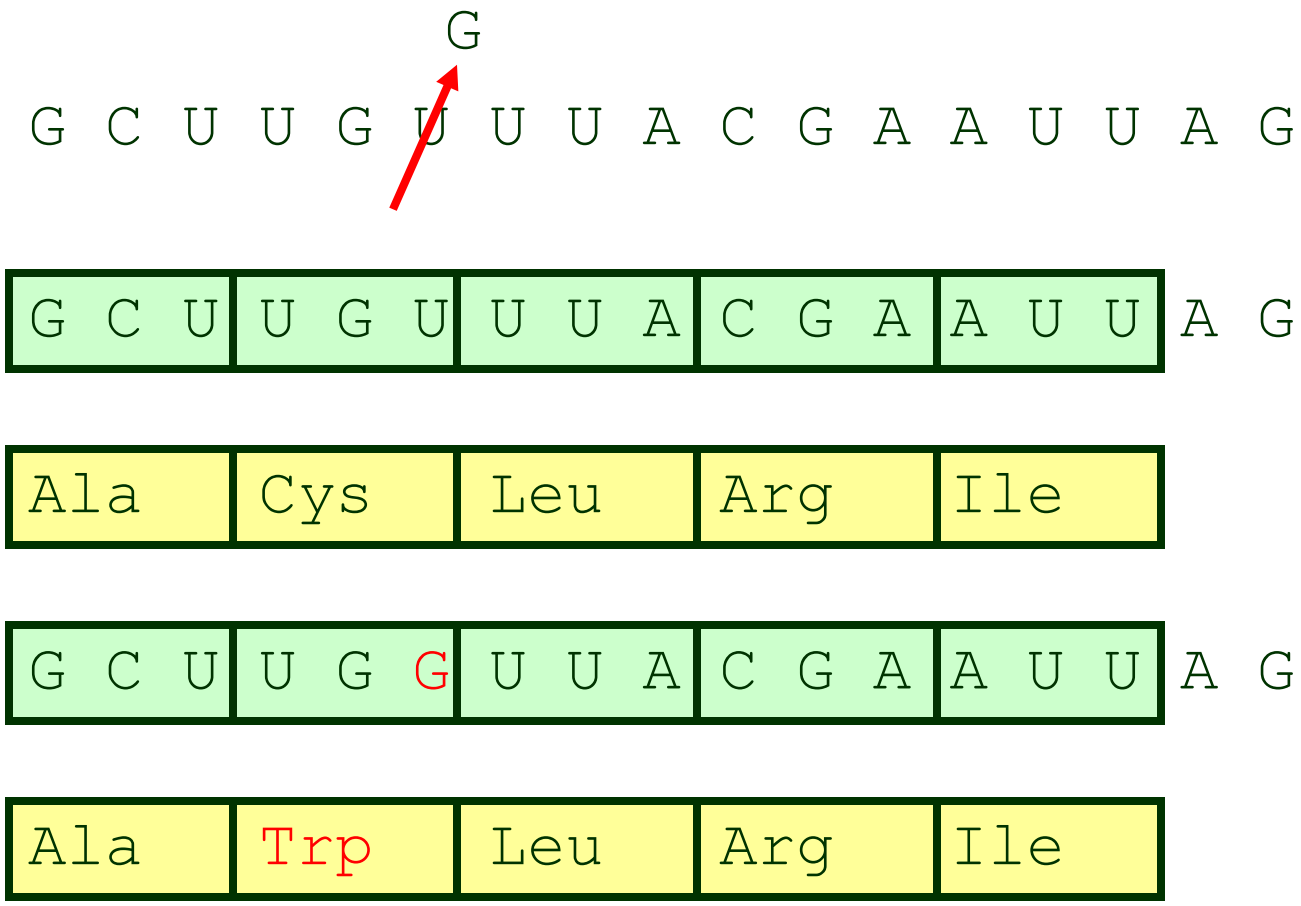
Synonymous Mutation

synonymous (silent) mutation, fourfold site



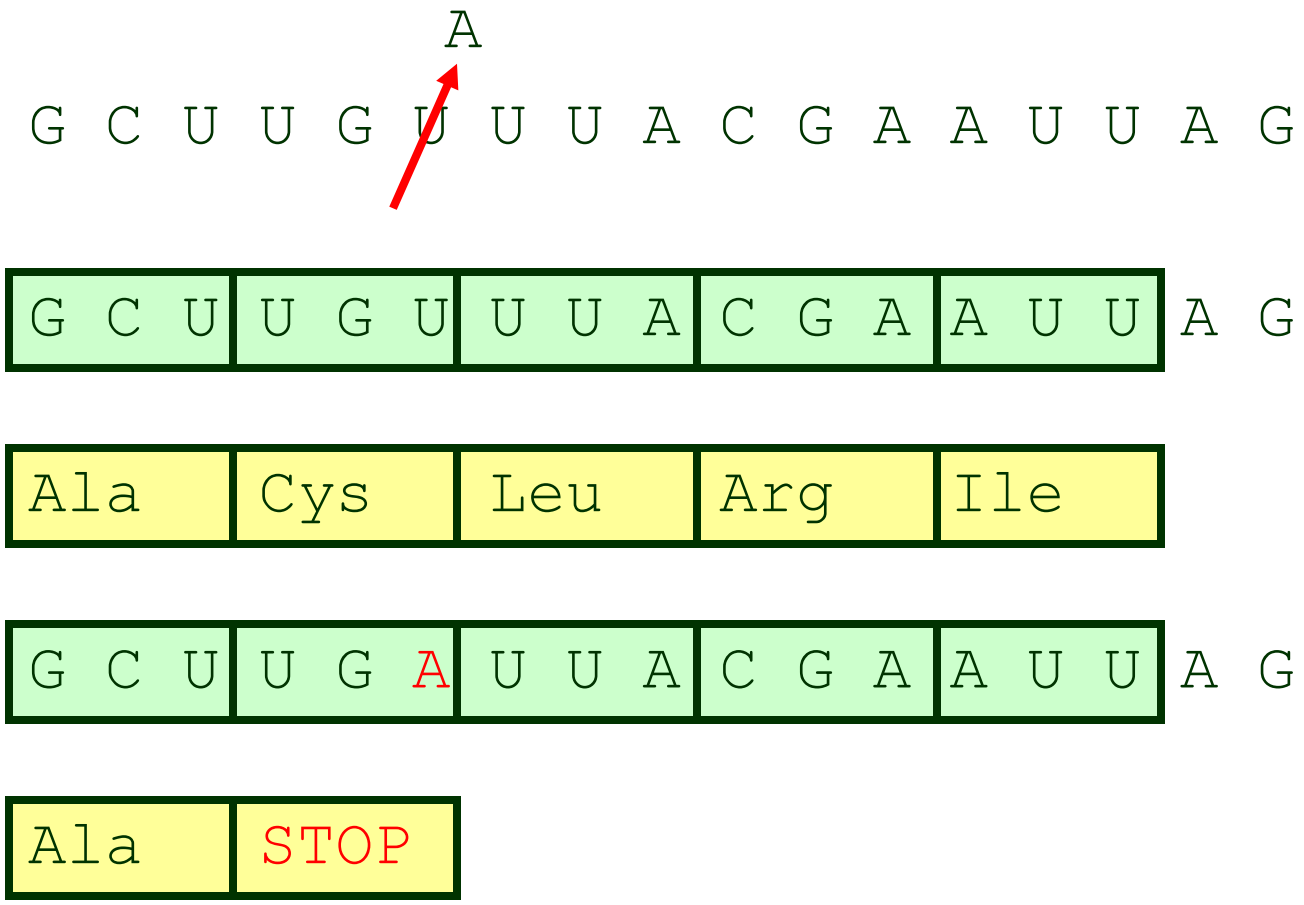
Missense Mutation

missense mutation



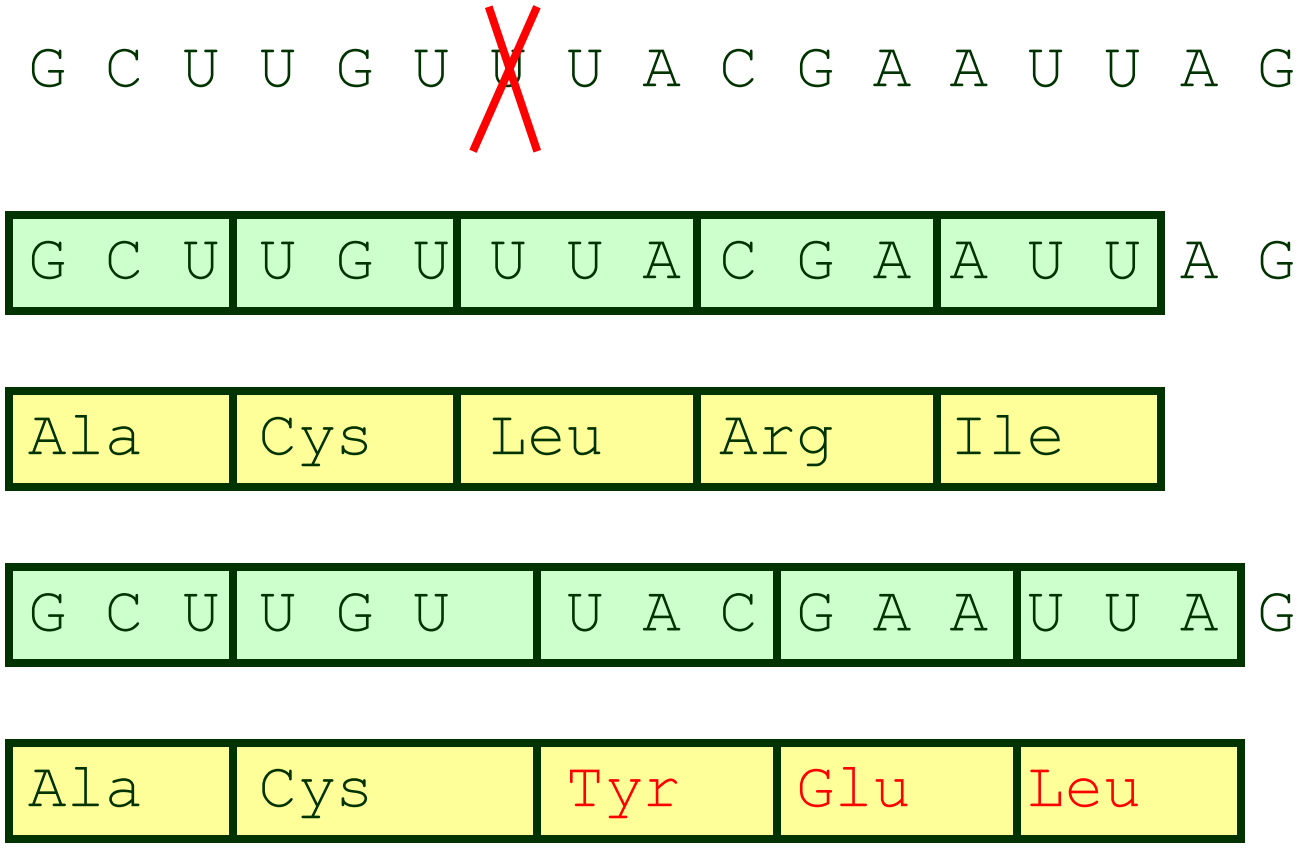
Nonsense Mutation

nonsense mutation



Frameshift

frameshift



Gene Expression Regulation

Regulation, signal transduction

- When should each gene be expressed?
- **Regulate** gene expression

  Examples:

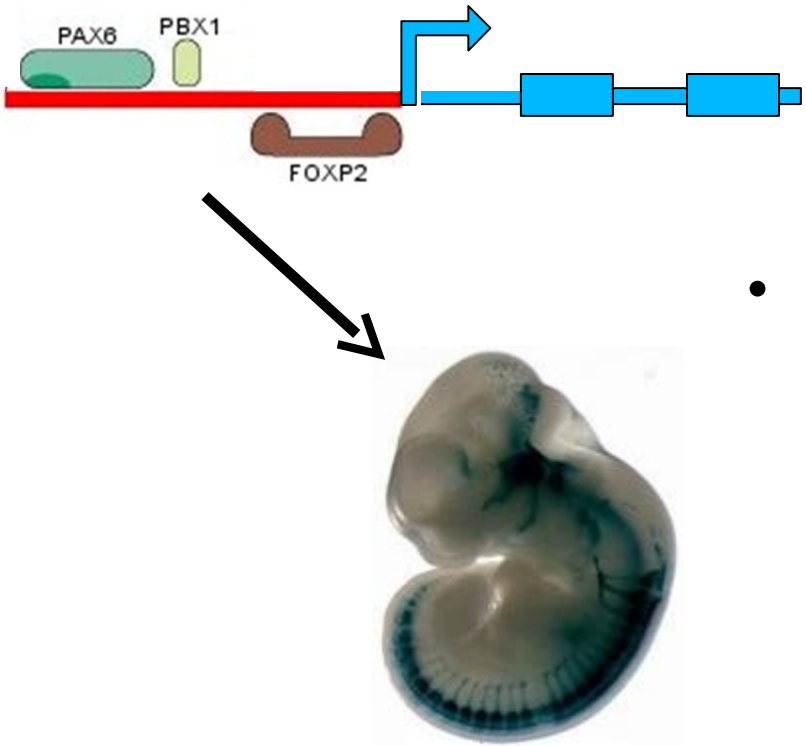
  - Make more of gene A when substance X is present
  - Stop making gene B once you have enough
  - Make genes C_1, C_2, C_3 simultaneously

- Why? Every cell has **same DNA** but each cell expresses **different proteins**.
- **Signal transduction**: One signal converted to another
  - Cascade has “master regulators” turning on many proteins, which in turn each turn on many proteins, …

Gene Regulation

- Gene expression is controlled at many levels:
  - DNA chromatin structure
  - Transcription
  - Post-transcriptional modification
  - RNA transport
  - Translation
  - mRNA degradation
  - Post-translational modification

Transcription Regulation



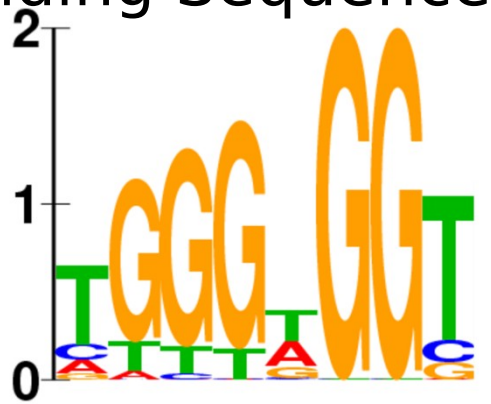
- Much gene regulation occurs at the level of transcription.
- Primary players:
  - Binding sites (BS) in *cis*-regulatory modules (CRMs)
  - Transcription factor (TF) proteins
  - RNA polymerase II
- Primary mechanism:
  - TFs link to BSs
  - Complex of TFs forms
  - Complex assists or inhibits formation of the RNA polymerase II machinery

Tx Factor Binding Sites

- Short, degenerate DNA sequences recognized by particular TFs

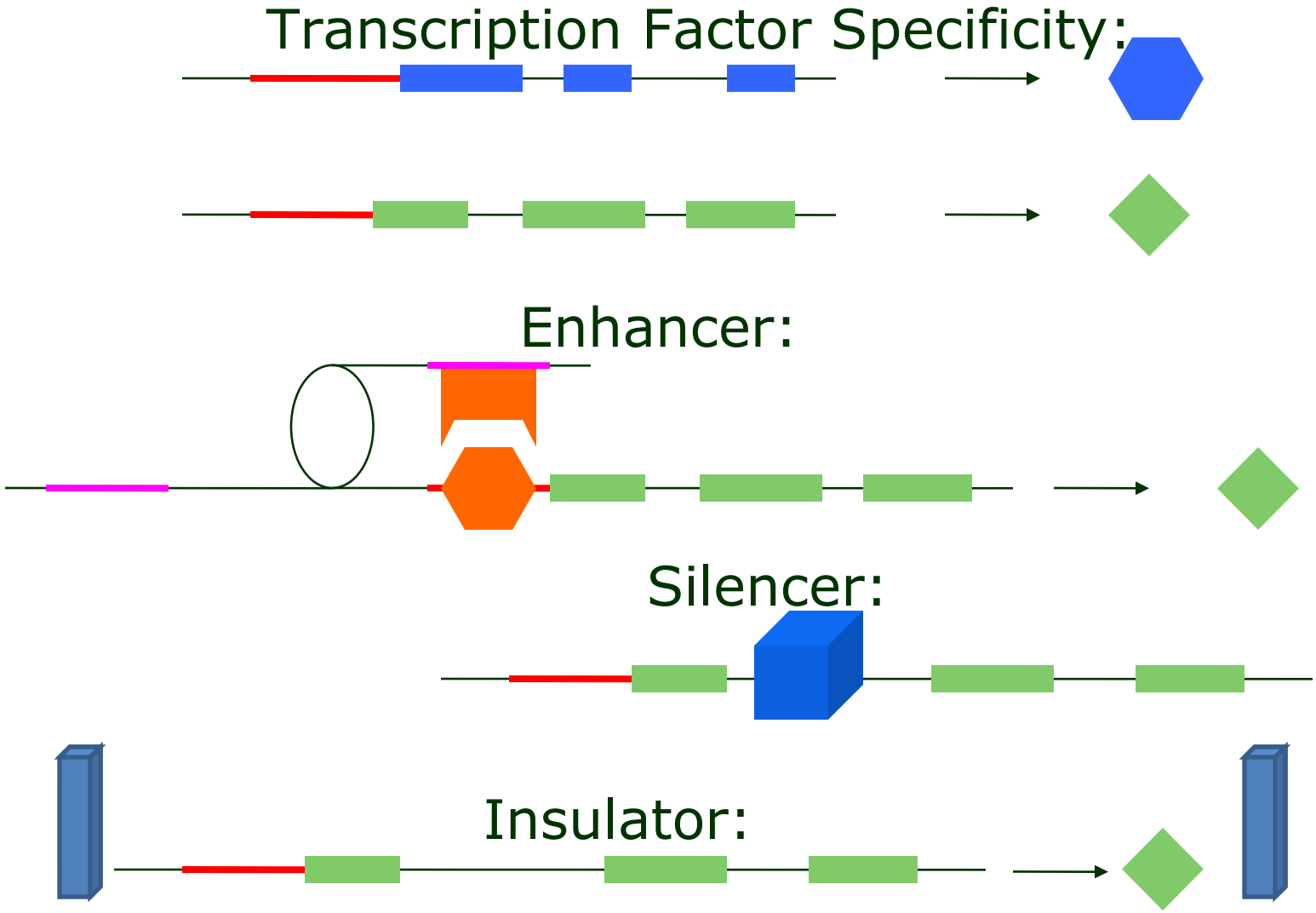
- For complex organisms, cooperative binding of multiple TFs required to initiate transcription

Binding Sequence Logo



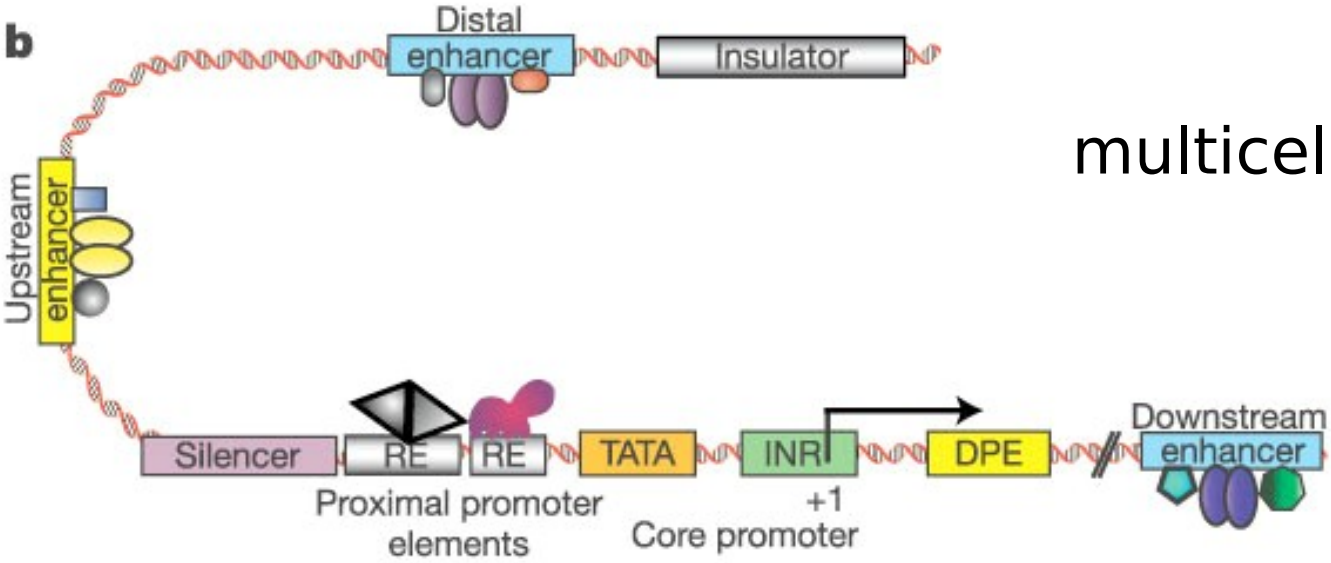
Transcription Regulation Mechanisms

enhancer, silencer, insulator



Unicellular vs. Multicellular

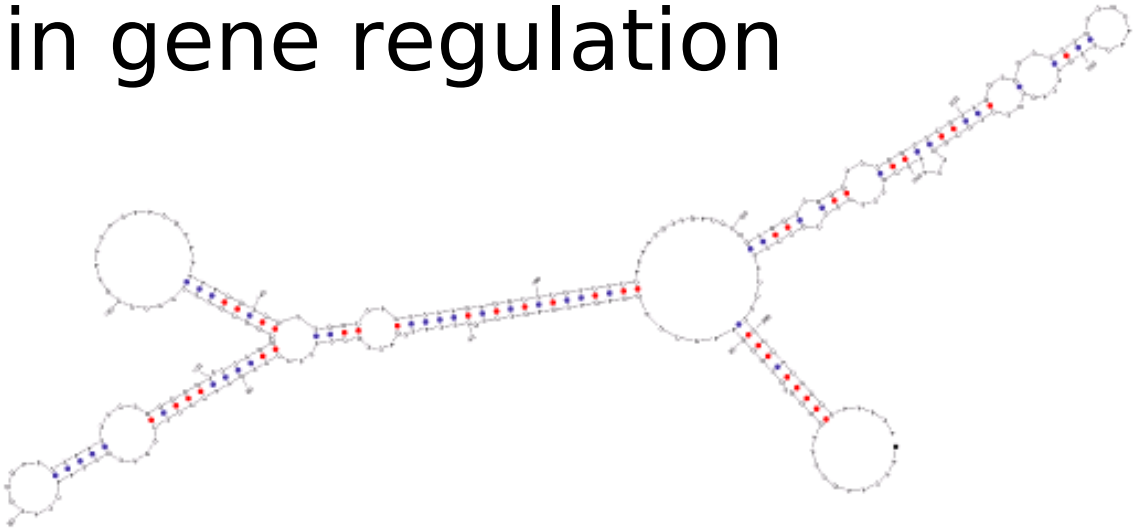
unicellular



multicellular

Non-coding RNAs

- RNAs transcribed from DNA but not translated into protein
- Structural ncRNAs: Conserved secondary structure (A-U, C-G, G-U)
- Involved in gene regulation



Summary

- All hereditary information encoded in double-stranded DNA
- Each cell in an organism has same DNA
- DNA → RNA → protein
- Proteins have many diverse roles in cell
- Gene regulation diversifies protein products within different cells

The end?


READING THE ENTIRE GENOME CODE...
Listen to this part: 'AGGCTAATC CGCATAACTG'
Wow!