Today's lecture:

Types of mutations and their impact on protein function

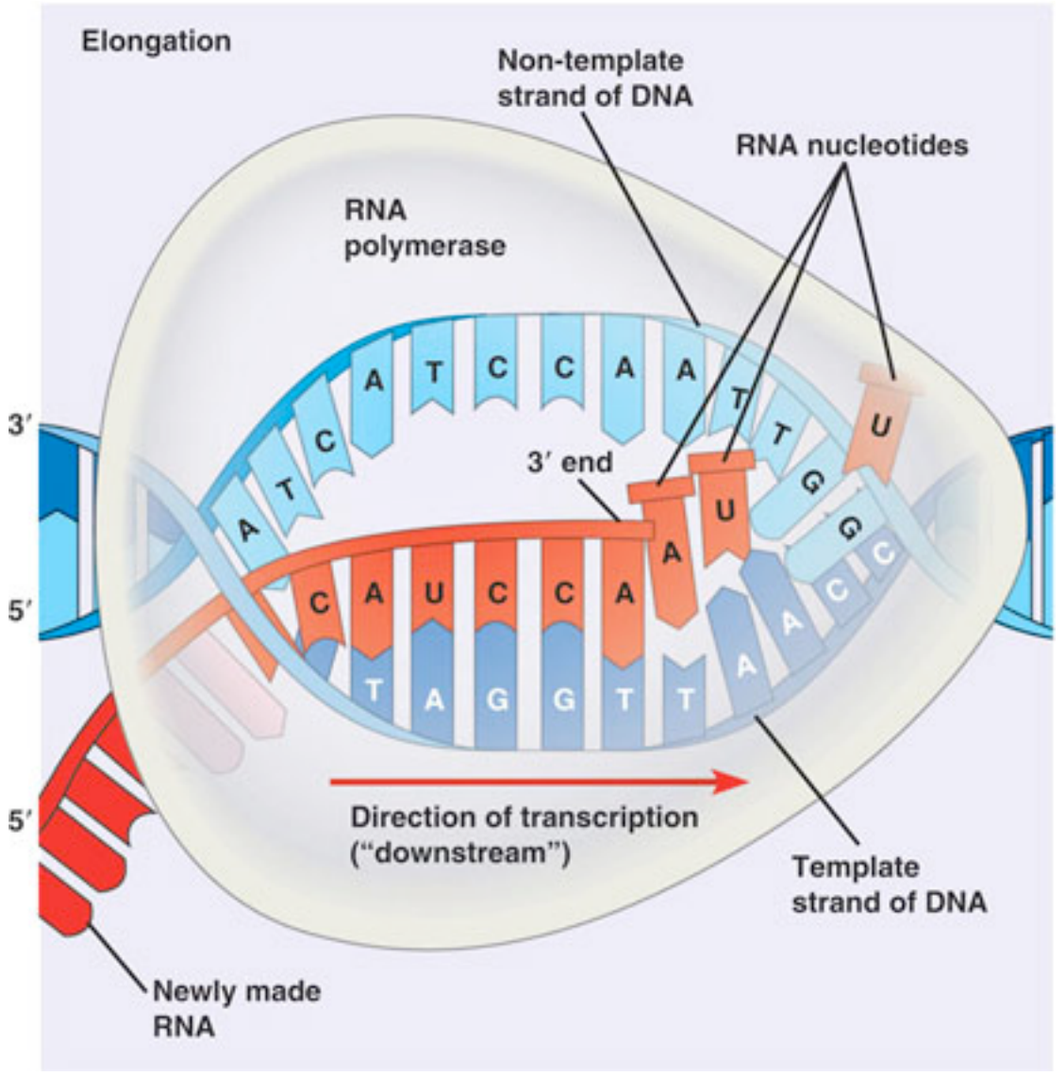
Mutations can be classified by their effect on the DNA sequence *OR* the encoded protein

From my Lecture 4 (10/1):

Classification of mutations by their effects on the DNA molecule

- **Substitution**: base is replaced by one of the other three bases
- **Deletion**: block of one or more DNA pairs is lost
- **Insertion**: block of one or more DNA pairs is added
- **Inversion**: 180˚ rotation of piece of DNA
- **Reciprocal translocation**: parts of nonhomologous chromosomes change places
- **Chromosomal rearrangements**: affect many genes at one time

The triplet nature of the genetic code means that base changes within coding sequence can have several different outcomes.



Second letter

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

Universal genetic code

I am not going to discuss the experiments that led to the deciphering of the genetic code. If you are interested, they are described in Chapter 8

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

	5'									3'
Wild-type mRNA		GCU	GGA	GCA	CCA	GGA	CAA	GAU	GGA	
Wild-type polypeptide	N	Ala	Gly	Ala	Pro	Gly	Gln	Asp	Gly	C
Silent mutation		GCU	GGA	GC**C**	CCA	GGA	CAA	GAU	GGA	
		Ala	Gly	Ala	Pro	Gly	Gln	Asp	Gly	
Missense mutation		GCU	GGA	GCA	CCA	**A**GA	CAA	GAU	GGA	
		Ala	Gly	Ala	Pro	Arg	Gln	Asp	Gly	
Nonsense mutation		GCU	GGA	GCA	CCA	GGA	**U**AA	GAU	GGA	
		Ala	Gly	Ala	Pro	Gly	Stop			
Frameshift mutation		GCU	GGA	GC**C**	ACC	AGG	ACA	AGA	UGG	A
		Ala	Gly	Ala	Thr	Arg	Thr	Arg	Trp	

Note:
these are all
substitutions

This one is an
insertion

Missense mutation: changes an amino acid to another amino acid. This may or may not affect protein function, depending on whether the change is "conservative" or "nonconservative," and what the amino acid actually does.

Nonsense mutation: changes an amino acid to a STOP codon, resulting in premature termination of translation.

"Silent" mutation: does not change an amino acid, but in some cases can still have a phenotypic effect, e.g., by speeding up or slowing down protein synthesis, or by affecting splicing.

Frameshift mutation: Deletion or insertion of a number of bases that is *not* a multiple of 3. Usually introduces premature STOP codons in addition to lots of amino acid changes.

Mutations outside the coding sequence can also impact gene expression

Promoter or enhancer* sequences

Termination signals

Splice donor and acceptor sites

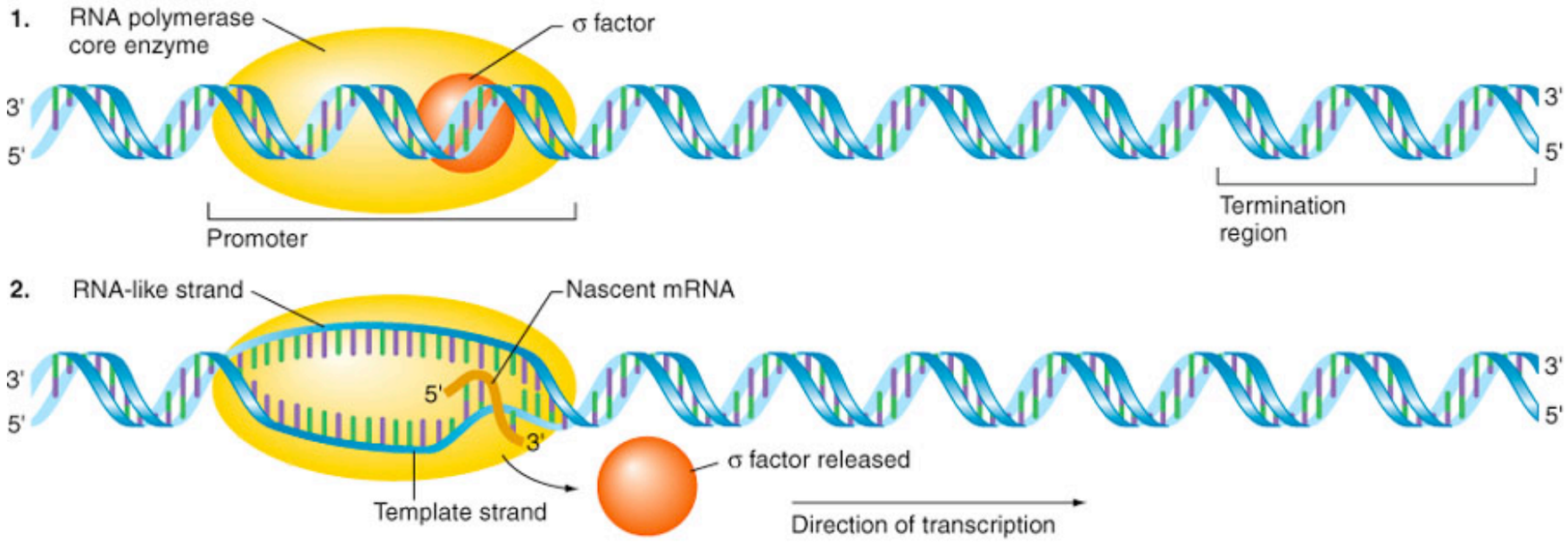
Ribosome binding sites

*Enhancers are regulatory elements that specify where and when particular genes are expressed

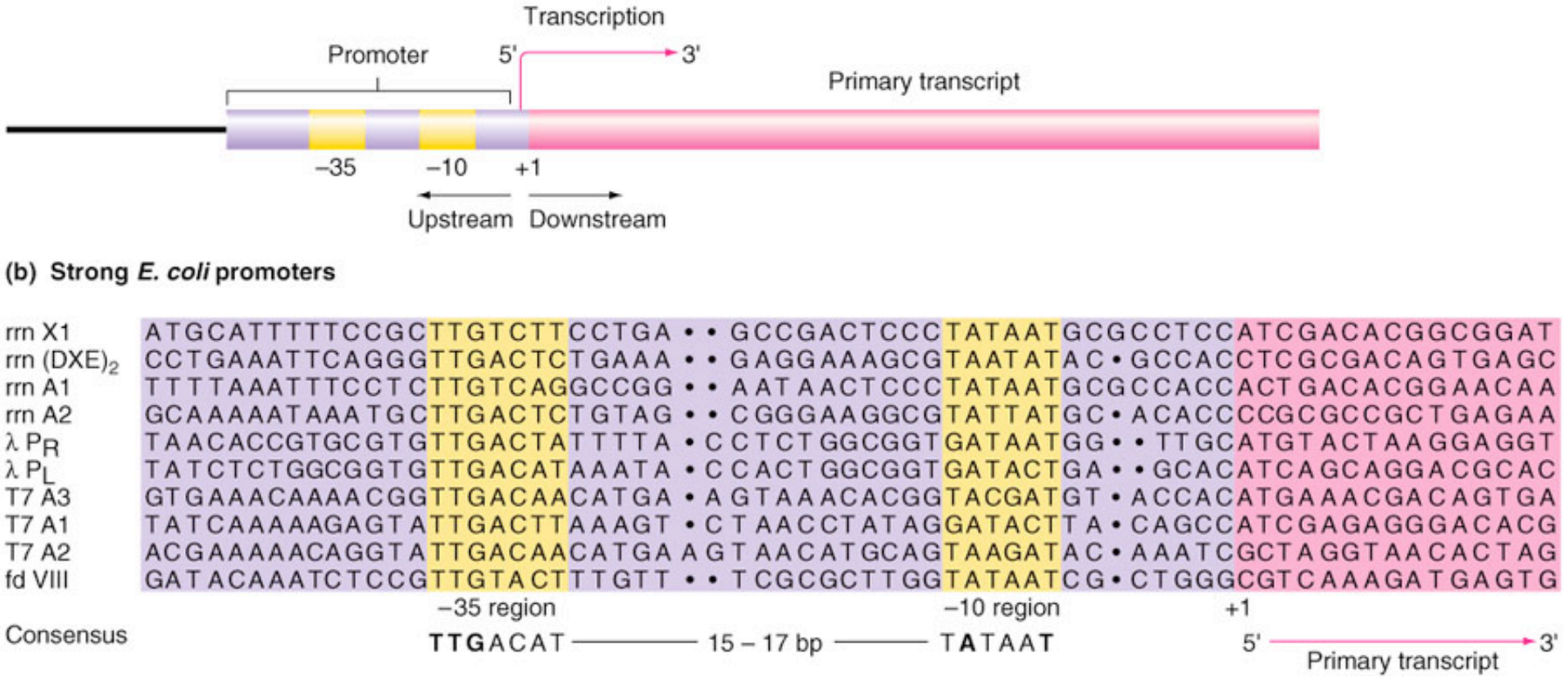
Not all of the mutable information in a gene is "coding."

A. Genes include information that tells the RNA polymerase where to start and stop (transcription initiation and termination signals).

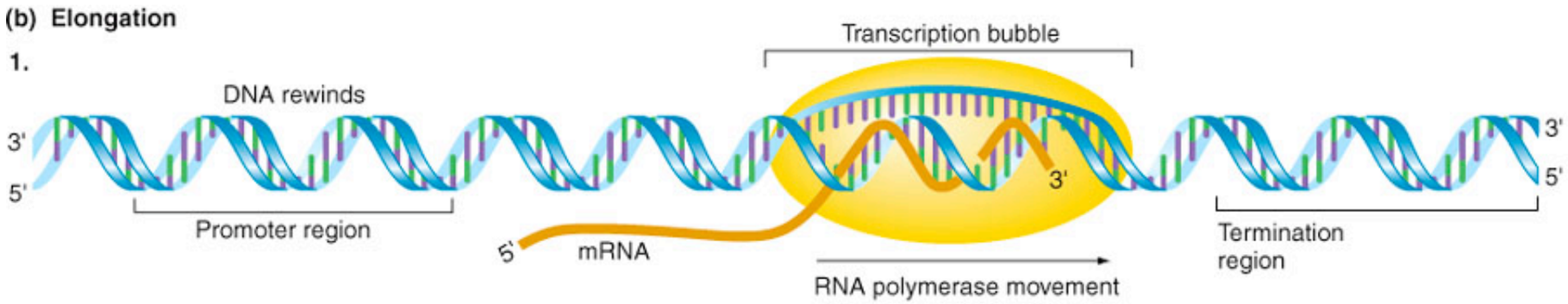
Initiation of transcription (bacteria)



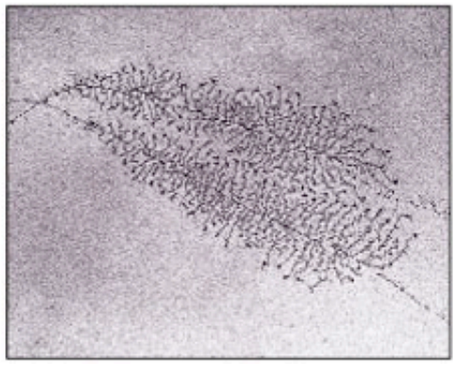
Initiation is controlled by short sequence elements called promoters, just upstream (5') of the gene

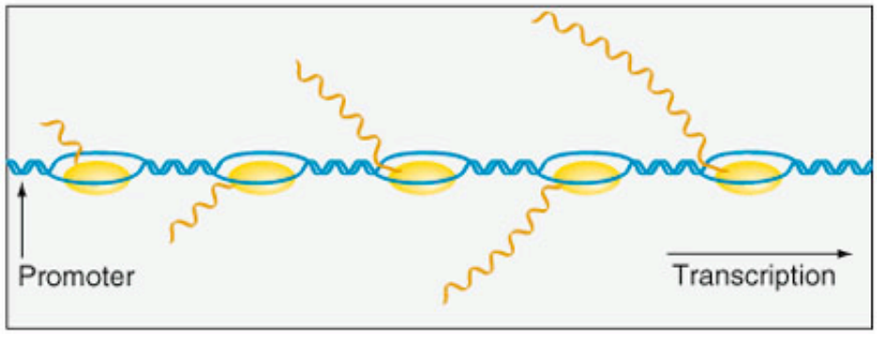


Elongation

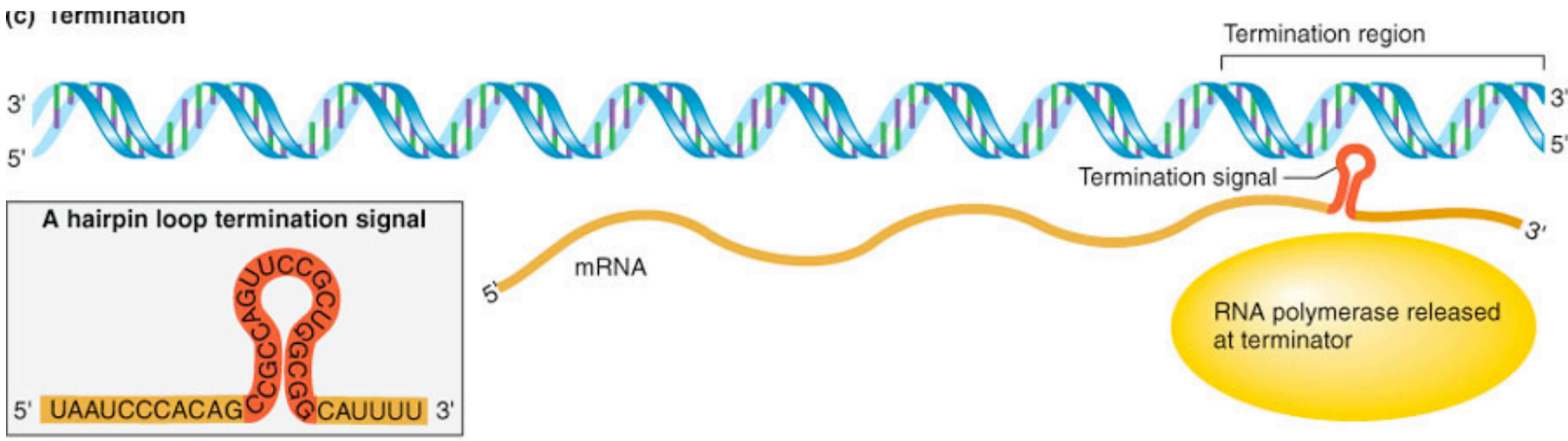


2.





Termination of transcription (bacteria)

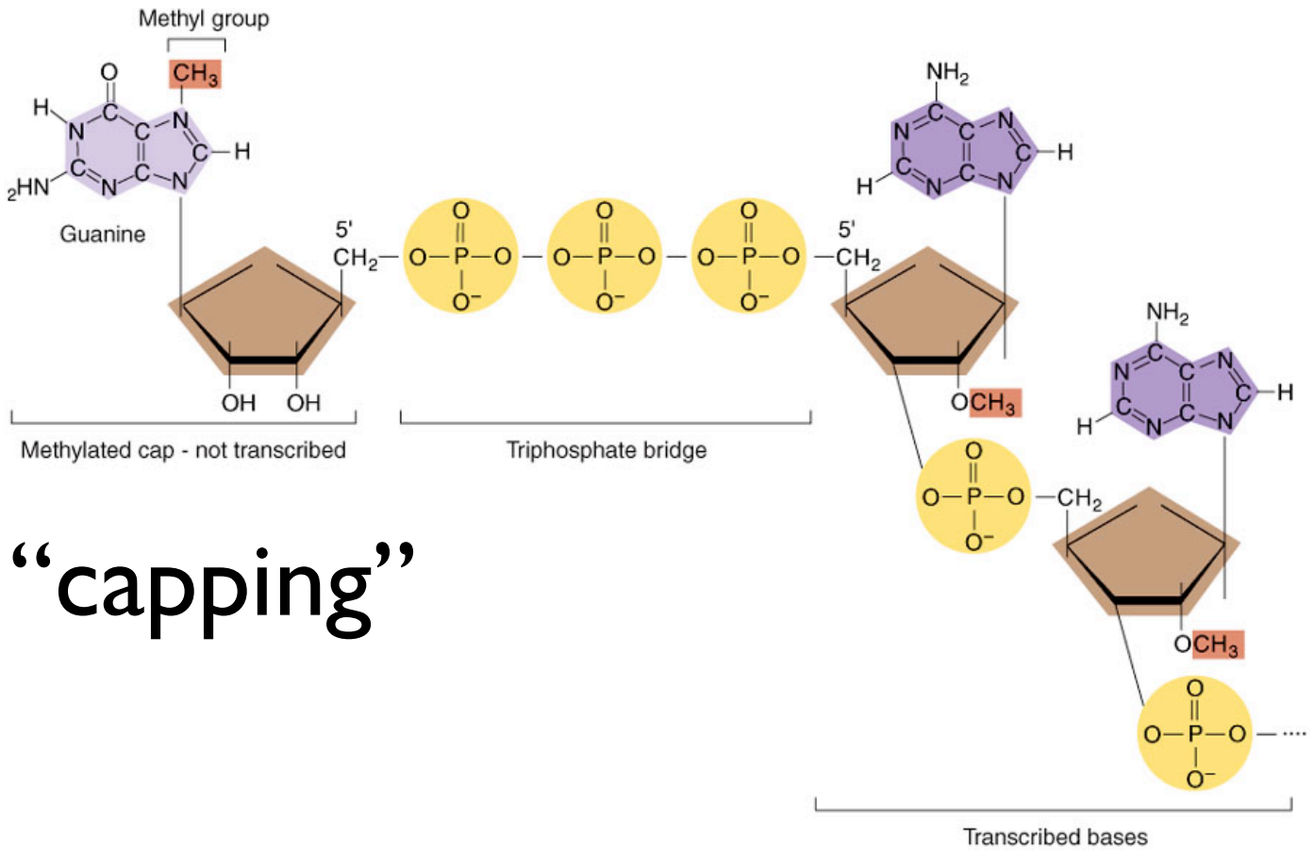


Not all of the mutable information in a gene is "coding."

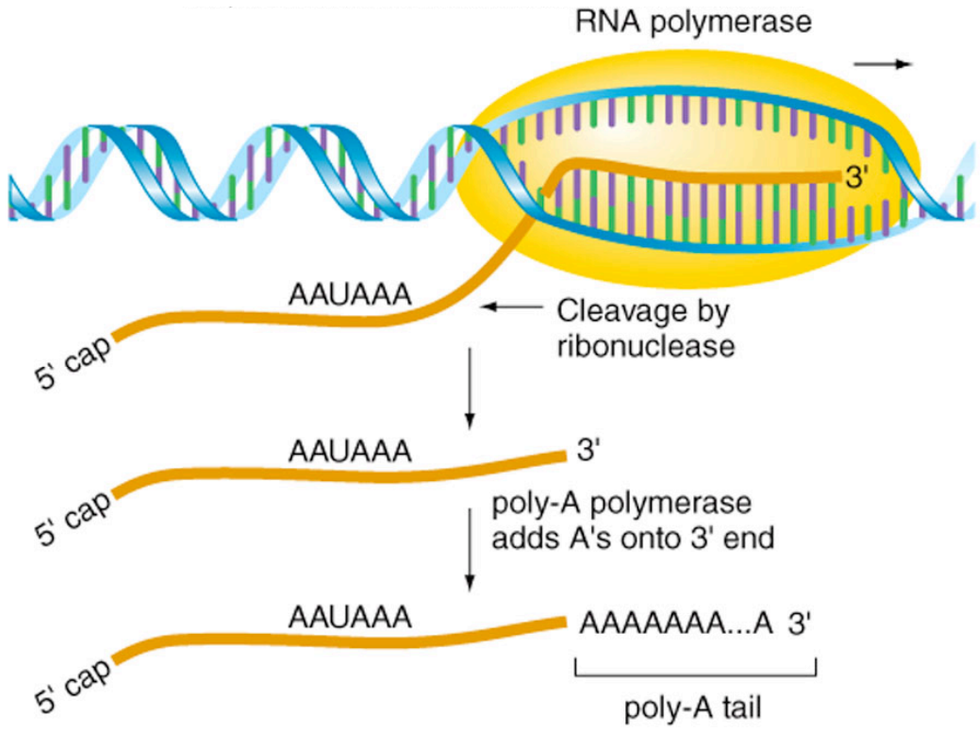
A. Genes include information that tells the *RNA polymerase* where to start and stop (transcription initiation and termination signals).

B. In eukaryotes, there is additional information that tells the splicing machinery where to cut and paste.

In eukaryotes, RNA synthesis and processing are more complex

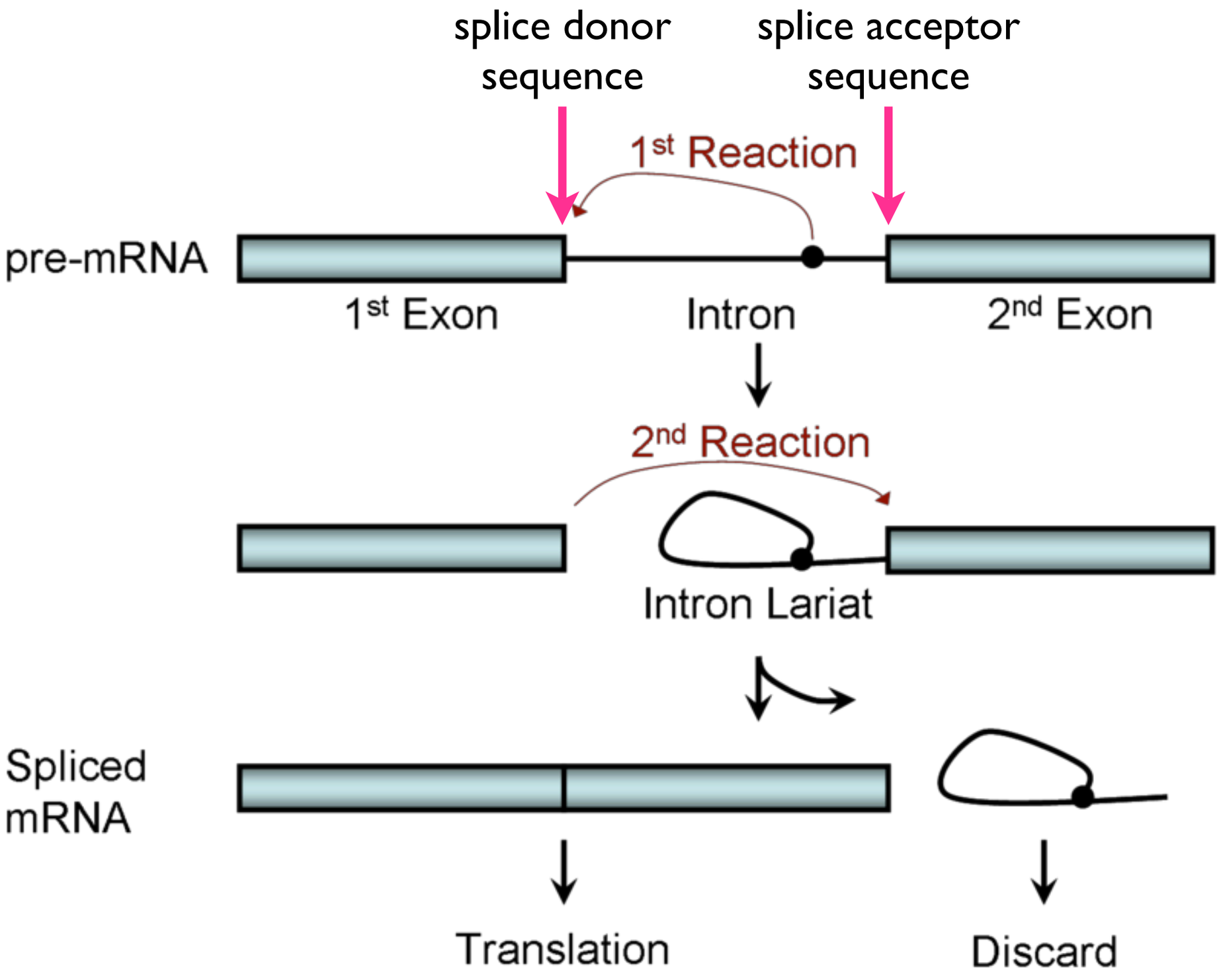


5' "capping"

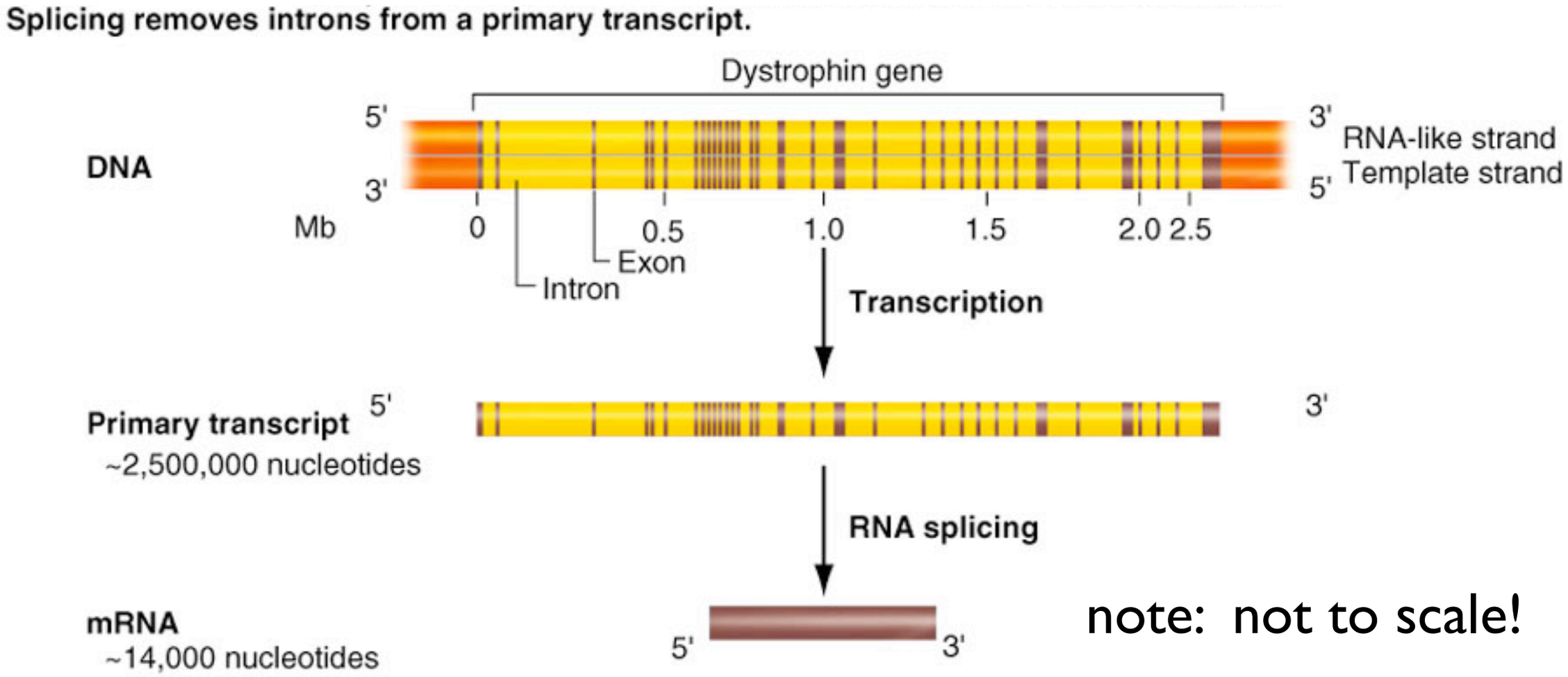


3' end polyadenylation (addition of poly-A tail)

Most eukaryotic genes contain *introns*, which are removed by a process called *splicing*



Most eukaryotic genes contain *introns*, which are removed by a process called *splicing*

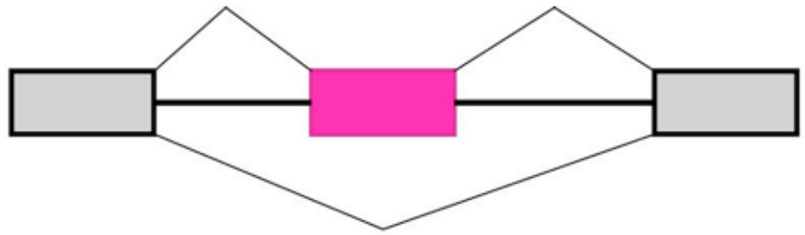


note: not to scale!

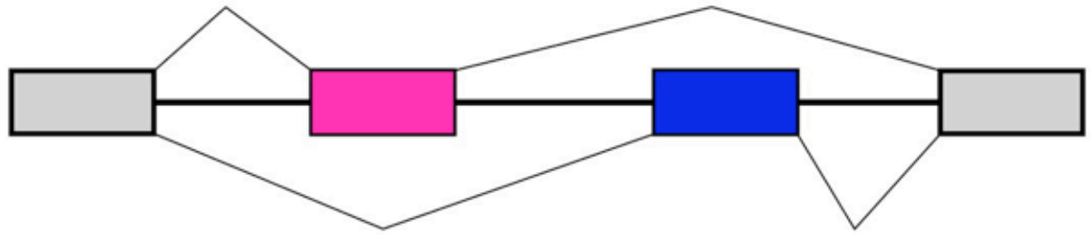
The "mature" mRNA has an added 5' cap and poly-A tail, and all of the introns removed. It can be MUCH smaller than the "primary transcript."

Sometimes there are multiple potential transcriptional start and/or splice sites

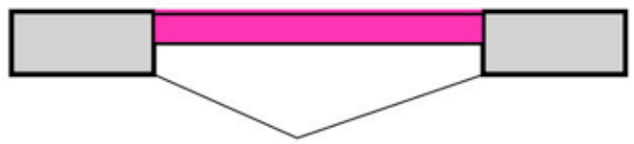
Cassette Exon



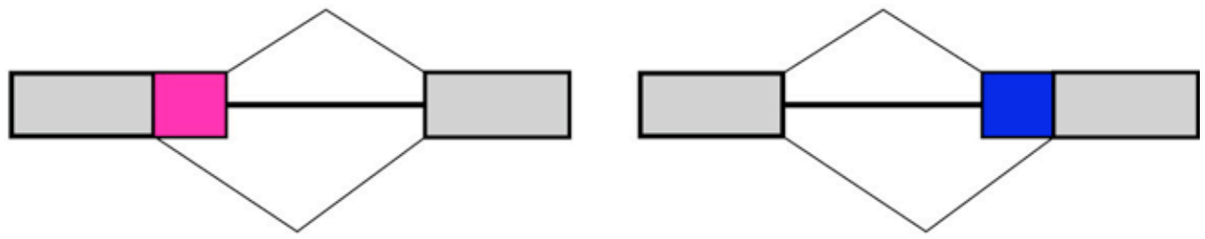
Mutually Exclusive Exons



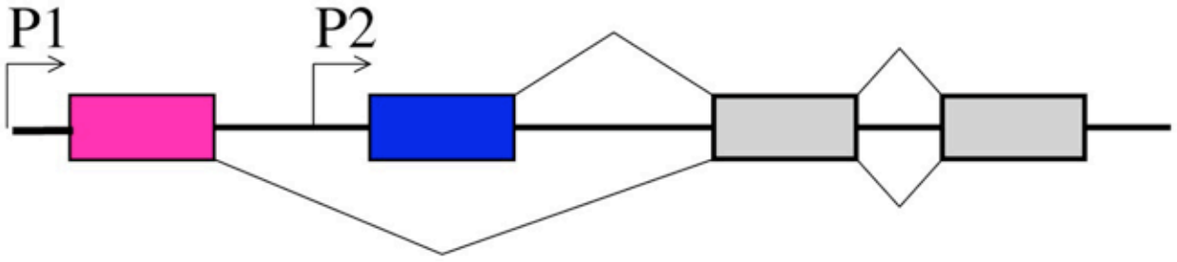
Intron Retention



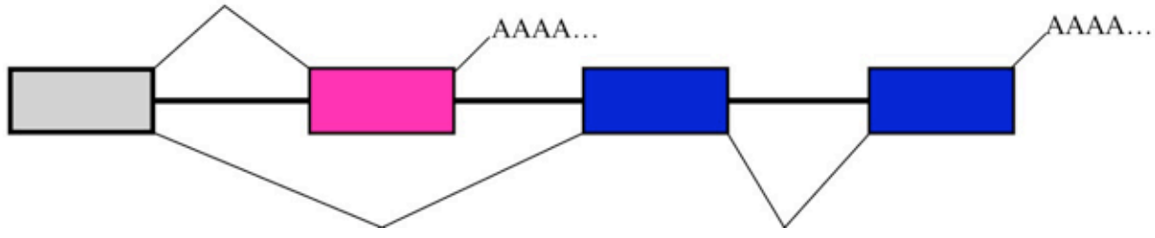
Alternative 5' or 3' Splice Sites



Alternative Promoters



Alternative Splicing and Polyadenlyation



Mutations outside the coding sequence can also impact gene expression

Promoter or enhancer* sequences

Termination signals

Splice donor and acceptor sites

Ribosome binding sites

*Enhancers are regulatory elements that specify where and when particular genes are expressed

Mutations are also classified by their impact on protein function:

Loss of function

Complete loss of the protein:
null, *loss-of-function*, *amorph*

Reduction of protein's ability to work:
hypomorph, *reduction-of-function*

Gain of function

Increase in the protein's function:
hypermorph, *gain-of-function*

A protein that interferes with the wild-type protein's function:
antimorph, *dominant negative*

Acquisition of a new function (or ectopic expression of the function):
neomorph, *dominant gain-of-function*

These terms are frequently misused, and also context-dependent

The distinction between loss-of-function and gain-of-function is not always super-clear.

Loss-of-function usually means that less of a protein is made or that some function of the protein has been compromised.

Loss-of-function mutations are usually recessive, since in most cases, a single "good" copy of the gene will suffice.

2 common types of exceptions:

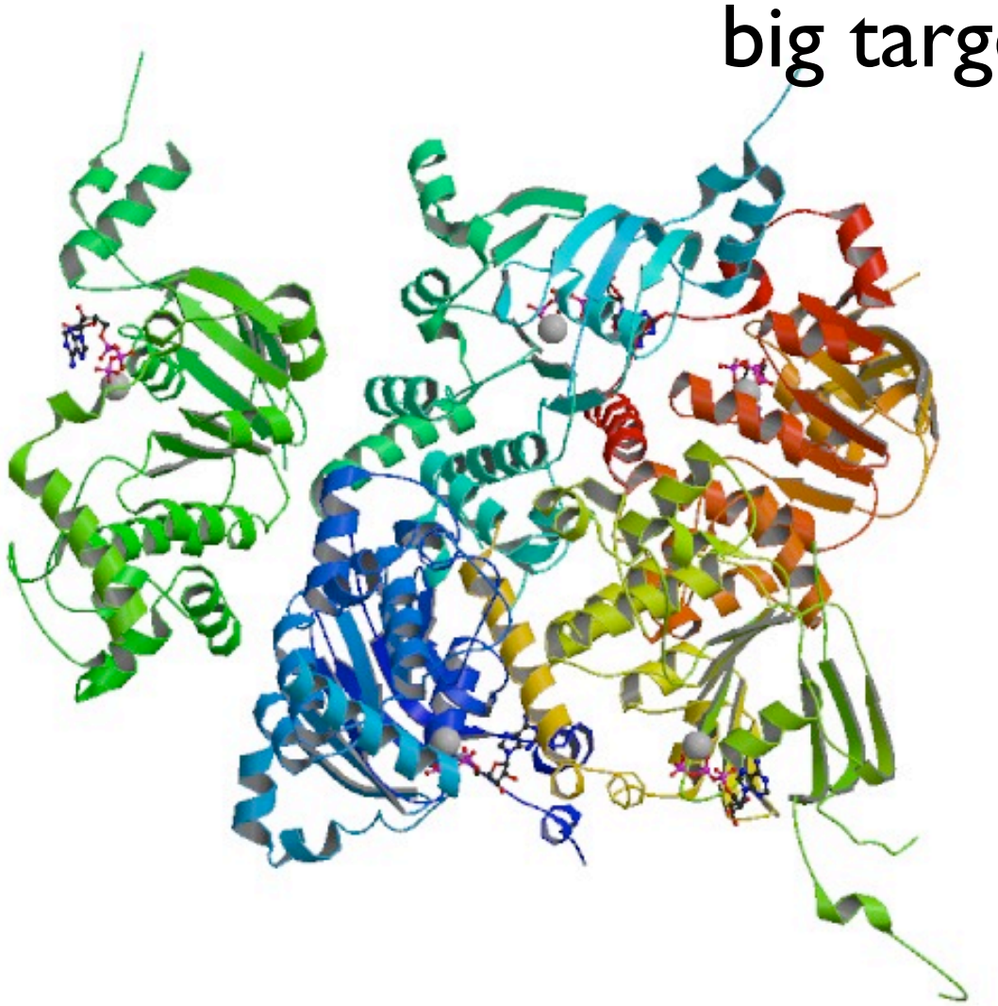
"**Haploinsufficiency**":
One copy is not enough

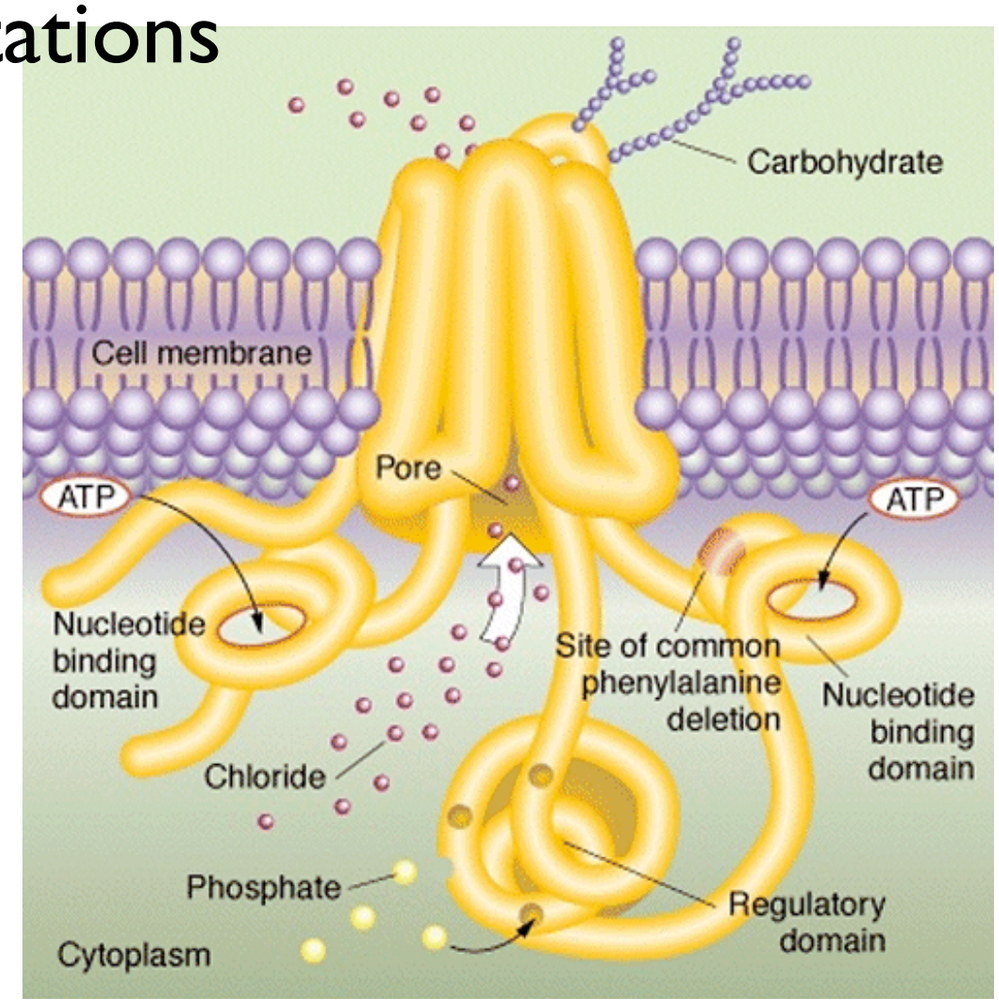
"**Dominant negative**" or "**antimorphic**" mutations:
The defective gene interferes with the function of the wild-type copy.
This is common with proteins that form polymeric structures, such as filaments.

Cystic Fibrosis shows the "expected" recessive pattern of inheritance for a loss-of-function allele of a gene

CFTR = cystic fibrosis transmembrane conductance regulator, a salt transporter required for normal function of the lungs, pancreas, and other tissues.

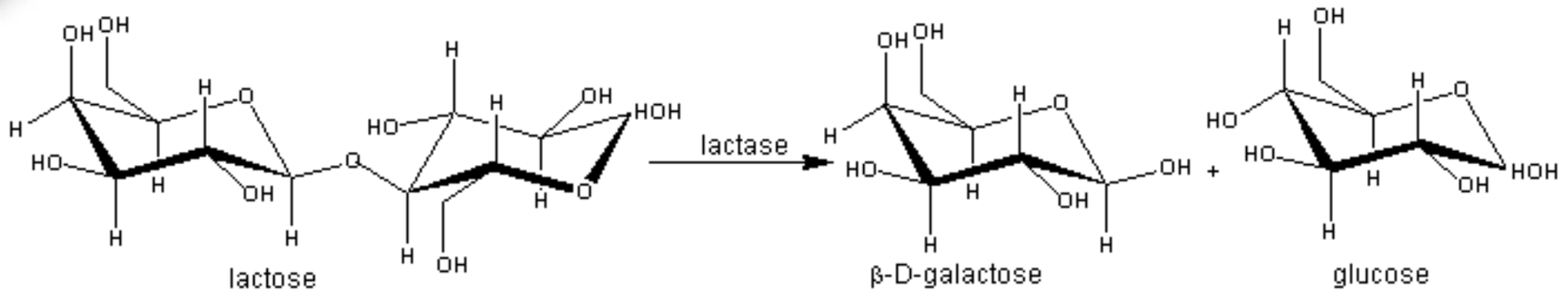
CFTR is a large gene that encodes a large protein, making it a big target for mutations





Another example of a recessive loss-of function allele:

Lactose intolerance is usually the result of "reduction-of-function" alleles that have low expression of the lactase enzyme in adults



Lactose *tolerance* (also known as *persistence*) is, historically speaking, the "mutant" form. Most mammals (including early humans) do not drink milk after infancy, and the lactose gene is usually inactivated (*i*.e., shut off). Many human populations, particularly in Europe, where dairy cows were domesticated, acquired the ability to metabolize lactose throughout adult life, most likely by mutation of regulatory elements in the lactase gene promoter region.

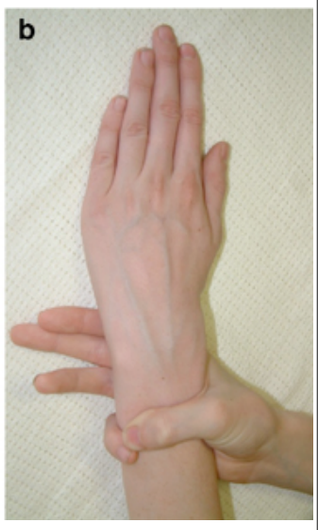
This has apparently happened independently among some east African populations.

Lactose intolerance is very prevalent among non-European populations.

Lactose tolerance is dominant over intolerance, for reasons that should be obvious. In other words, lactose intolerance shows recessive inheritance.

Vincent Schiavelli
1948-2005

Marfan syndrome is caused by "**dominant negative**" mutations in the *FBN1* gene

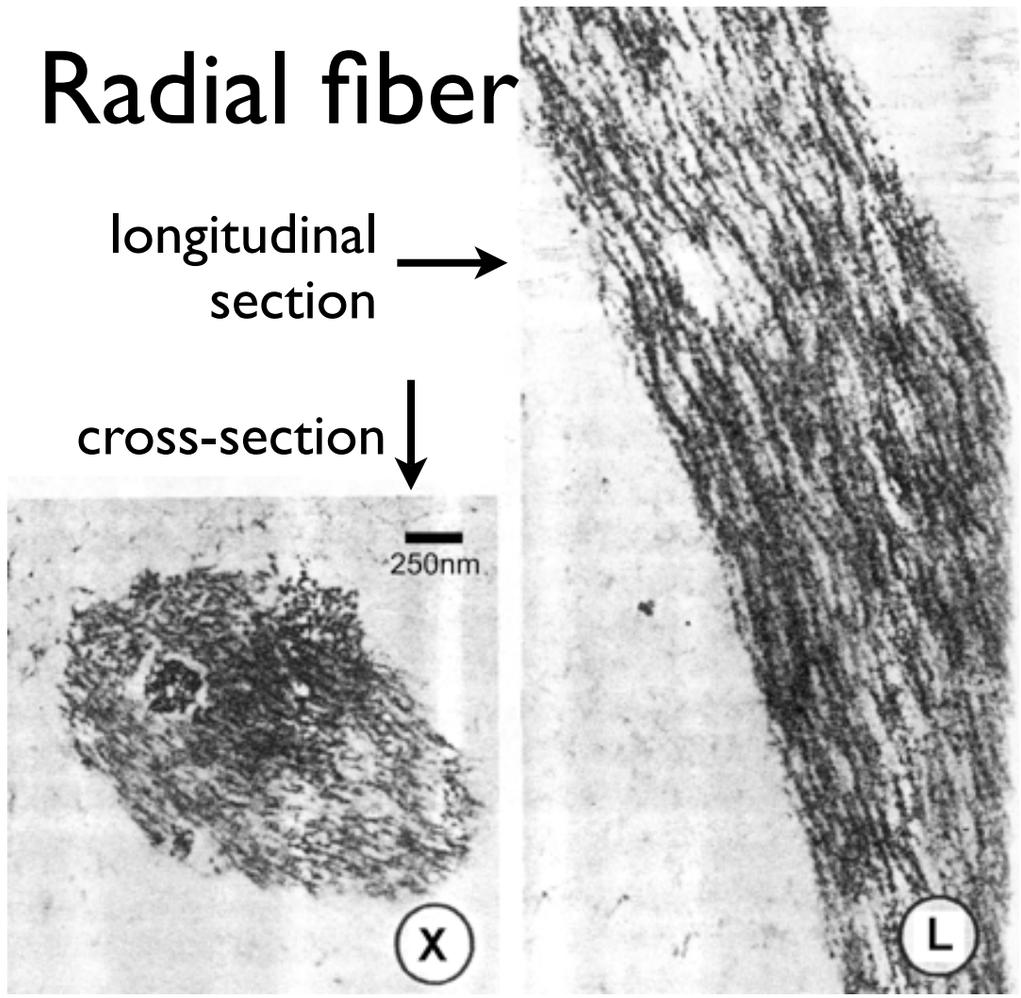


Marfan syndrome is caused by mutations that truncate the *FBN1* gene, which encodes Fibrillin-1, a protein that forms microfibrils in the extracellular matrix.

Radial fiber

longitudinal section →

cross-section ↓



Fibrillin-1

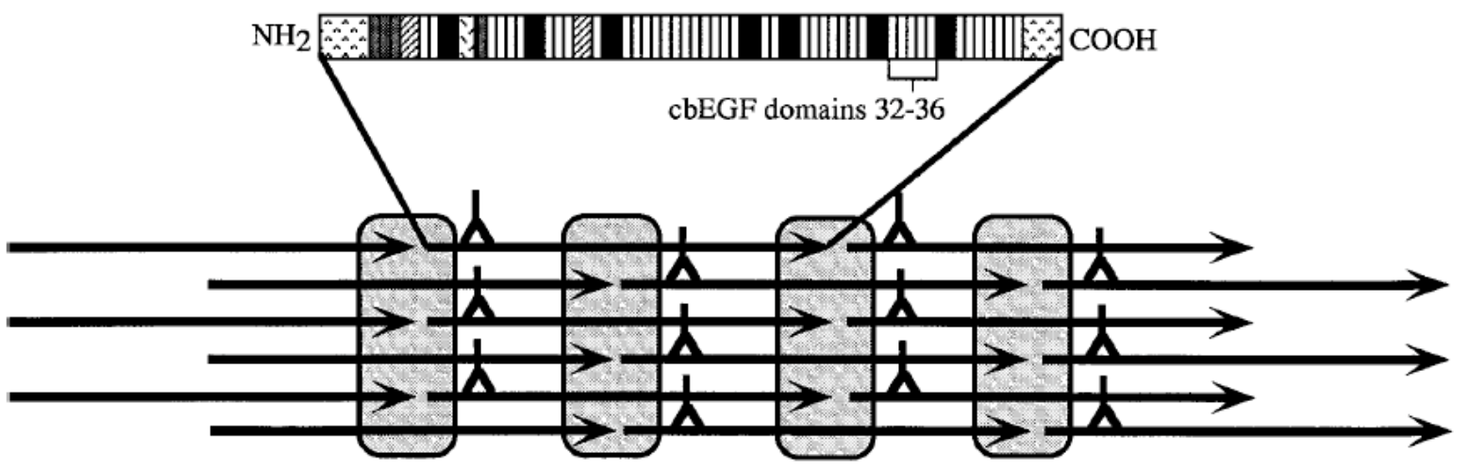
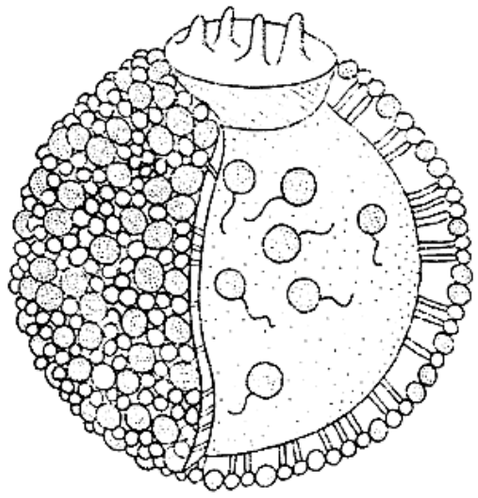


Figure 6. The Domain Structure of Human Fibrillin-1 and a Model for the Organization of Fibrillin Monomers within Connective Tissue Microfibrils

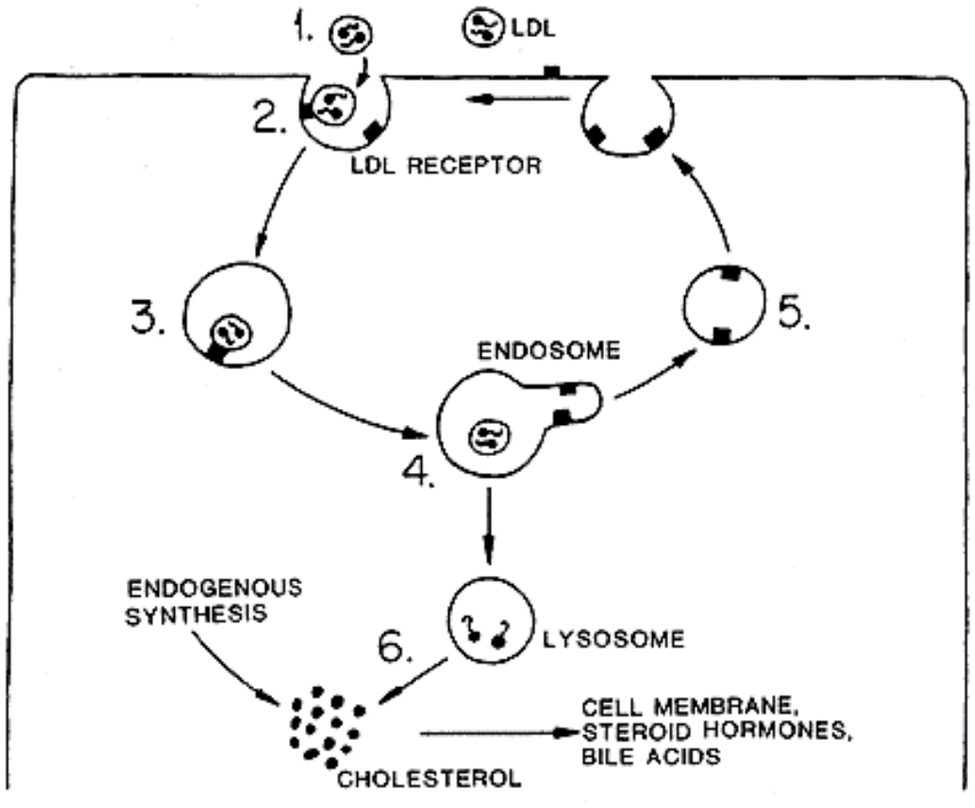
Fibrillin-1 assembles into long chains (microfibrils) that bundle together to form fibers
Defective Fibrillin-1 proteins disrupt the integrity of the chains.

Haploinsufficiency:

Familial Hypercholesterolemia (FH; high cholesterol) can result from having only one good copy of the LDL receptor gene



LDL (low density lipoprotein) particle



LDL uptake from blood

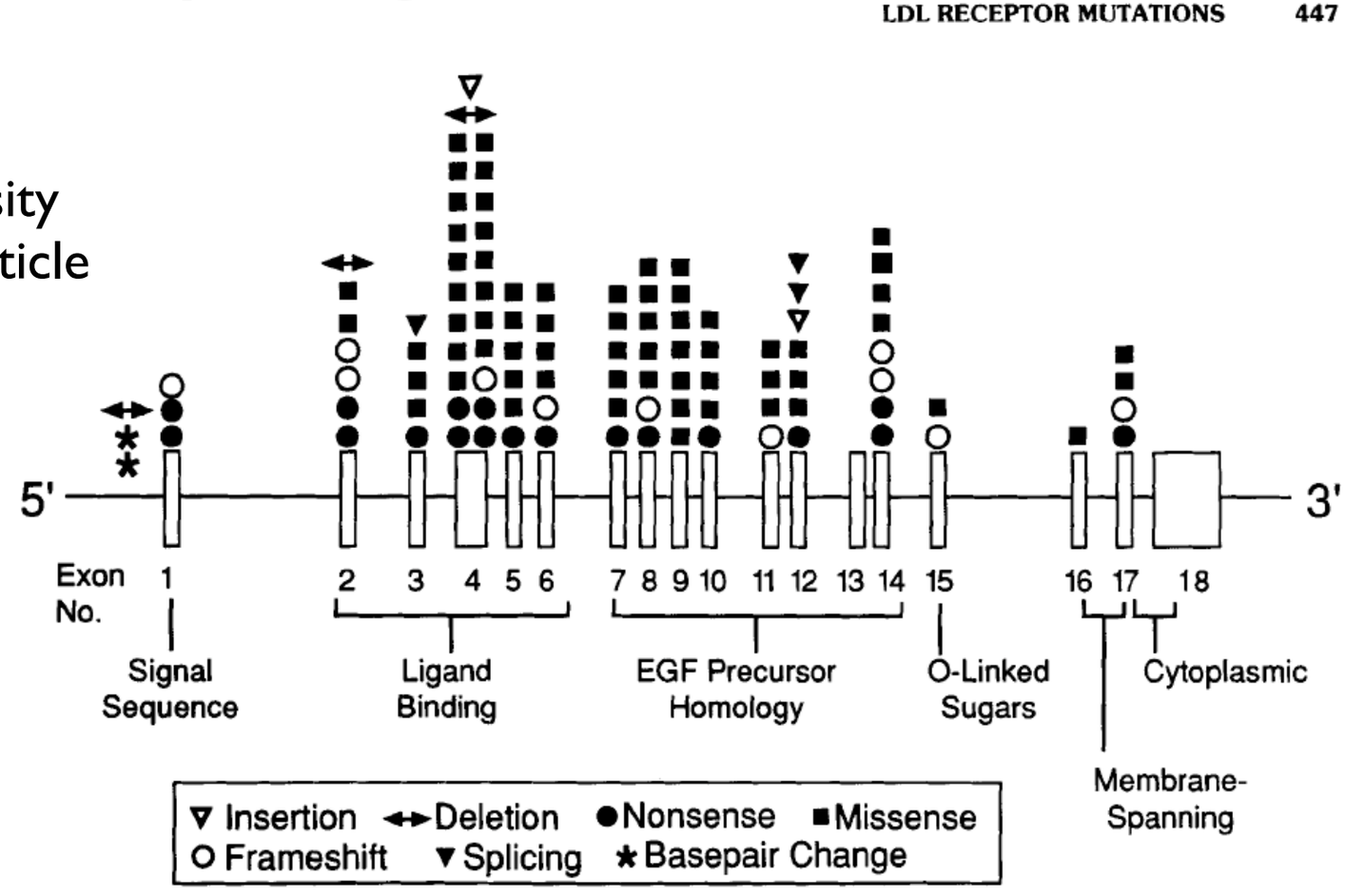
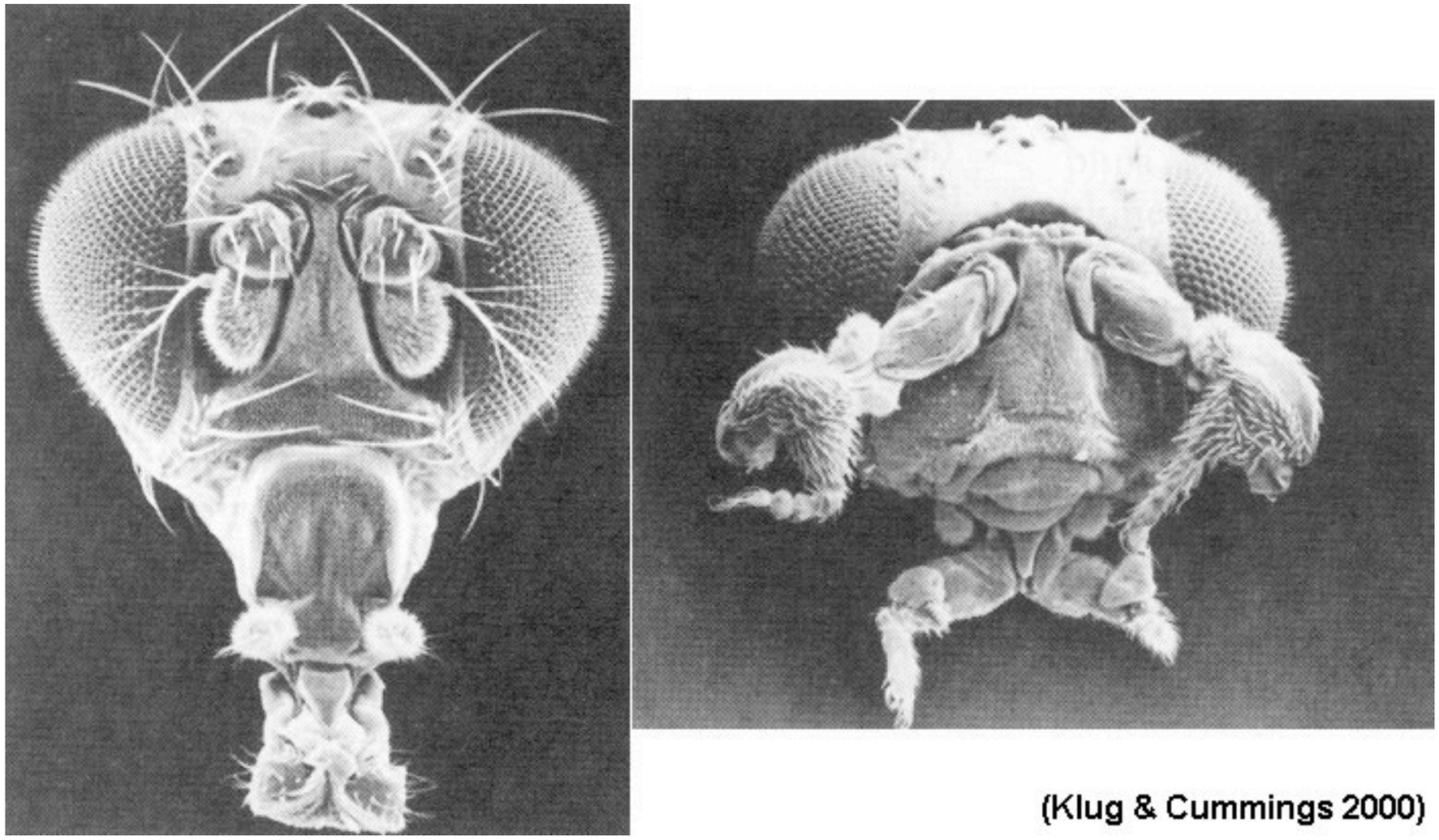


FIGURE 1. Point mutations and small in-frame deletions/insertions (<25 bp) in the LDL receptor gene in individuals with FH. Exons are shown as vertical boxes and introns as the lines connecting them. The map is drawn to approximate scale. Additional data for each mutation are given in Table 2.

Lots of different mutations cause dominant familial hypercholesterolemia (FH) by disrupting LDL receptor function

“Gain-of-function” mutations are almost always dominant



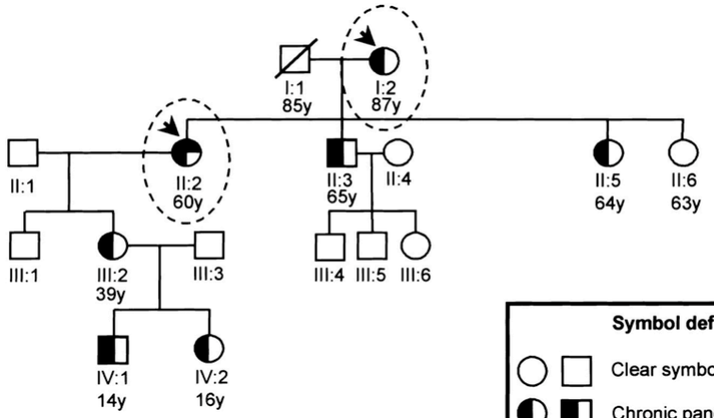
(Klug & Cummings 2000)

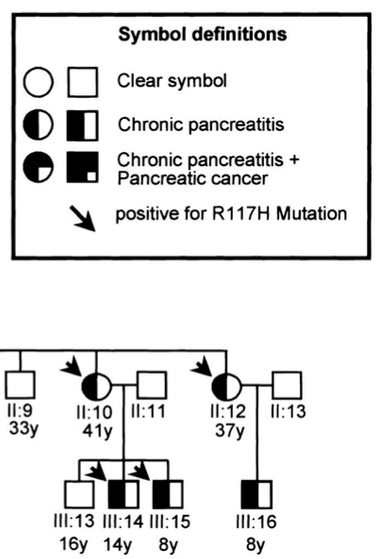
Antennapedia mutation in *Drosophila*

"Gain-of-function" mutations are almost always dominant

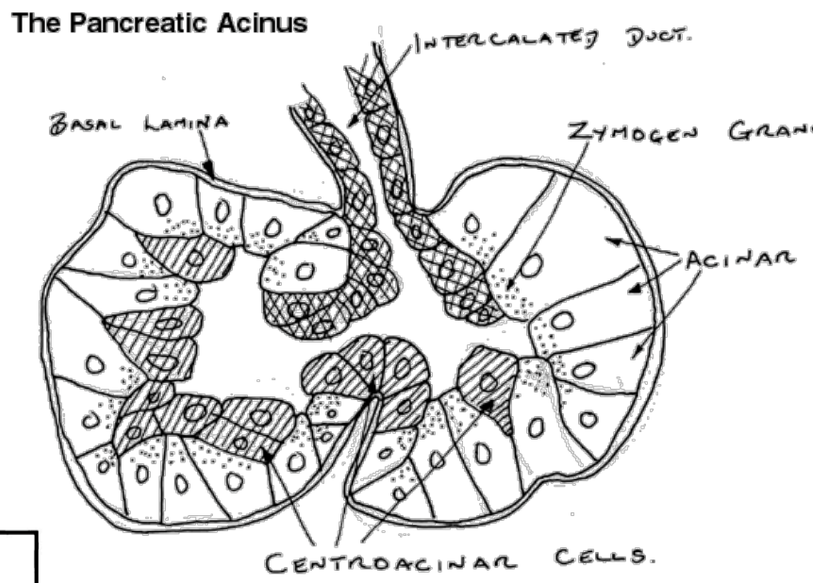
2 pedigrees showing dominant inheritance of pancreatitis

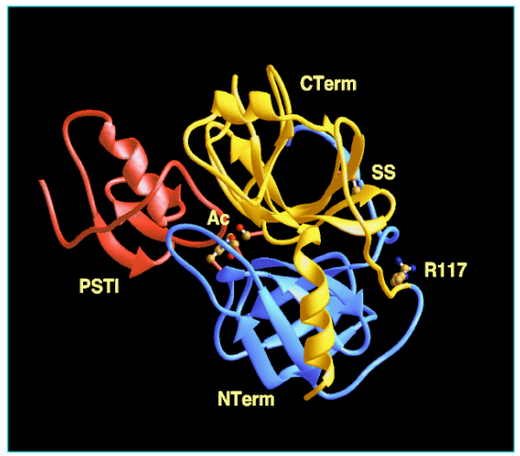
A



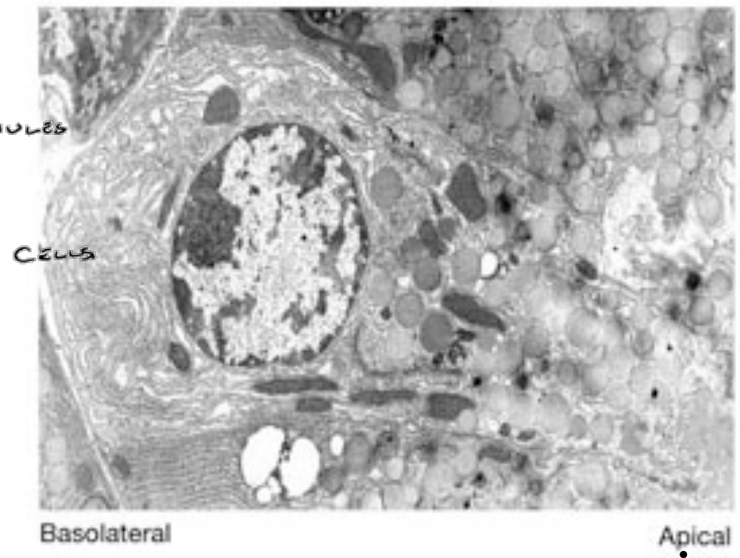


B

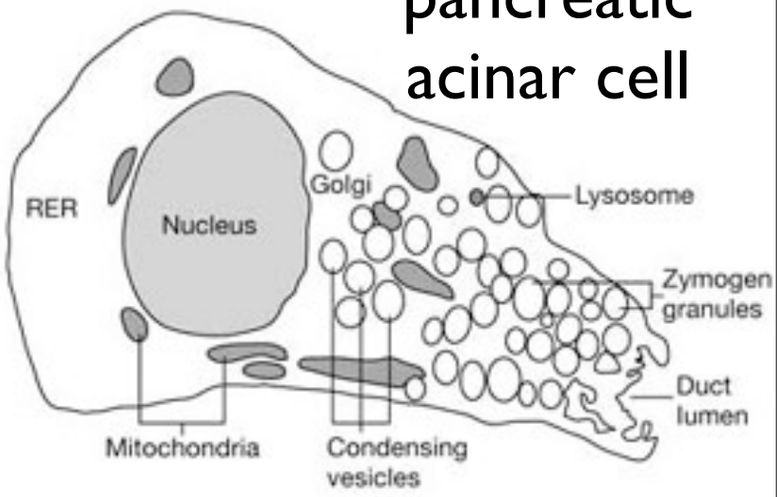




trypsin



pancreatic acinar cell

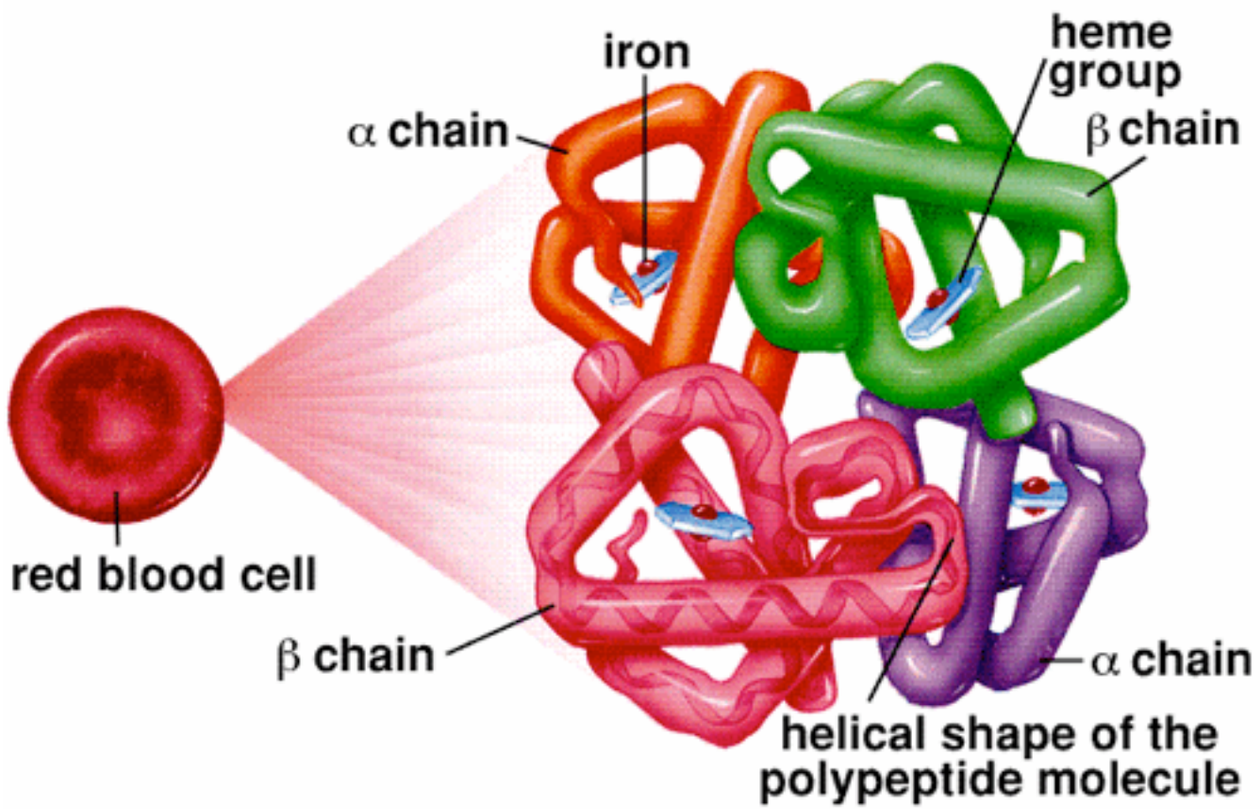


Hereditary pancreatitis is caused by a mutation that causes a digestive enzyme, trypsin, to become aberrantly active inside the pancreas.

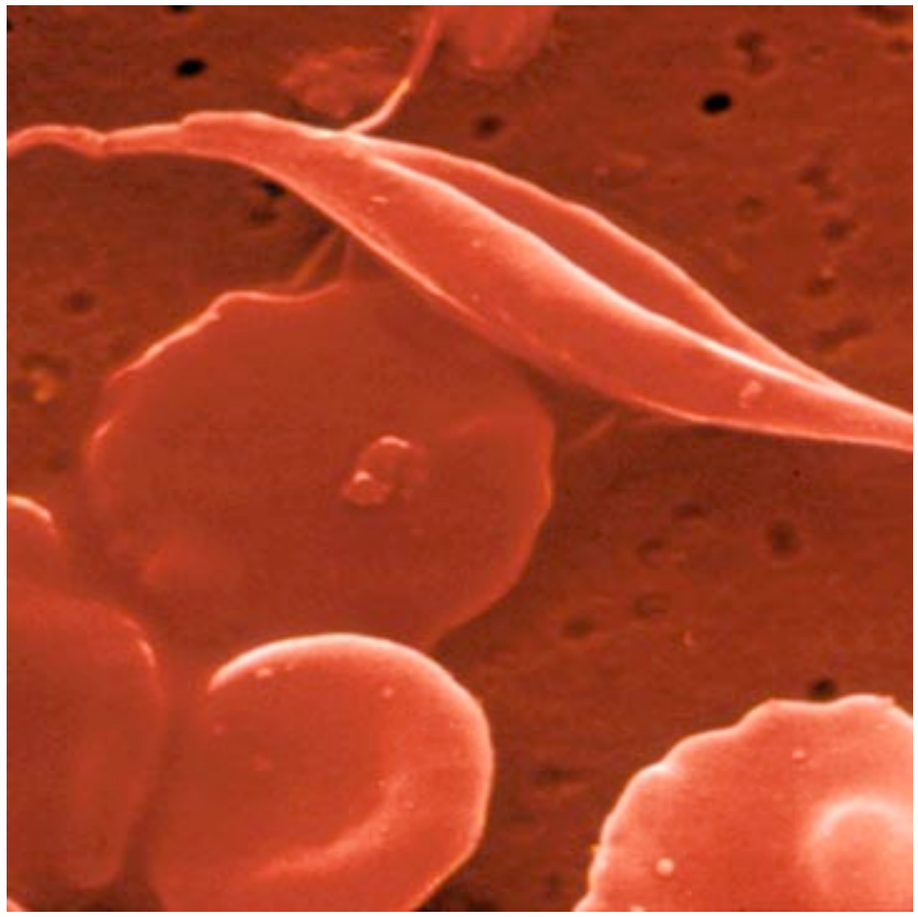
Normally, the pancreas is protected because active trypsin will destroy itself by cutting at R117. This will split the trypsin and inactivate it.

In HP, R117 is mutated to H117. This creates a "super-trypsin" that cannot be inactivated and leads to acute pancreatitis.

“Gain-of-function” is defined with respect to a specific function



Variations in the beta globin gene (*HbS* alleles) cause sickle cell anemia. The disease is inherited as a recessive trait, but the same mutations result in dominant inheritance of resistance to malaria.



sickle-shaped red blood cells tend to clump together, restricting oxygen delivery and causing more acute symptoms.

Thallasemia and G6PD are other recessive genetic diseases for which a single mutation confers malaria resistance.