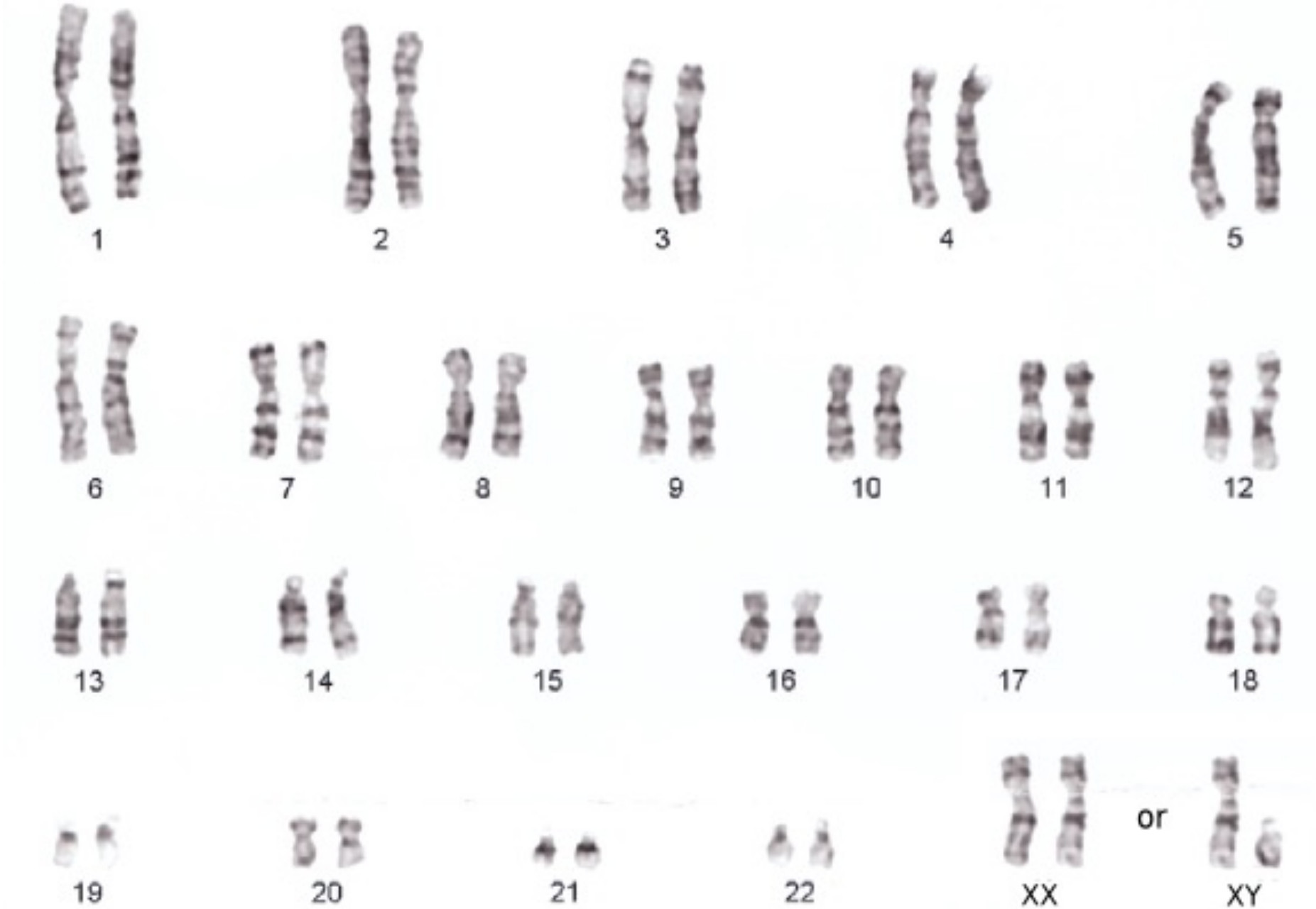
SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

Session 3:
Types of Human Genetic Variation

Learning Objectives

- Describe differences in types of genetic variation and how they affect phenotypes.
- Identify inheritance patterns of genotype-phenotype relationships.



Karyotype: Profile of an individual's chromosomes isolated from an individual cell. Autosomes: 1-22, sex chromosomes: XX/XY

Our Genome in Numbers

23 chromosome pairs

3.2 billion base-pairs (A,C,G,T)

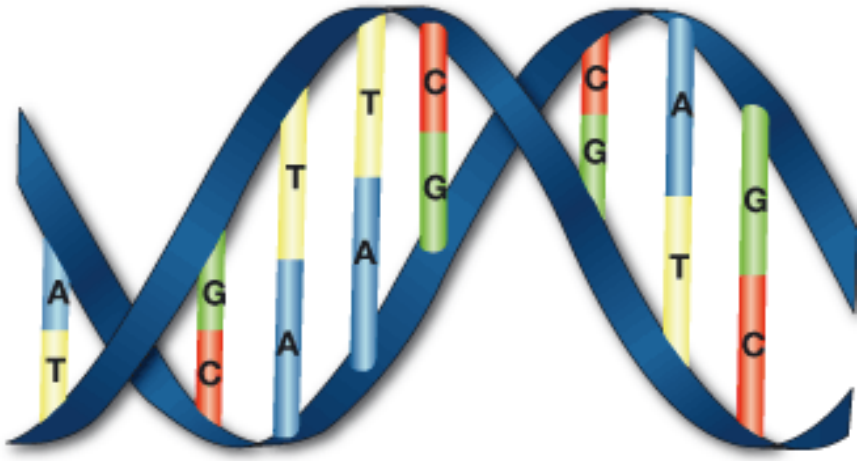
~20,000 genes

~1.5% of the genome is coding DNA

Genetically, two people chosen at random are likely to be ~99.9% identical

The human genome mutation rate is estimated to be ~1.1×10^{-8} per site per generation

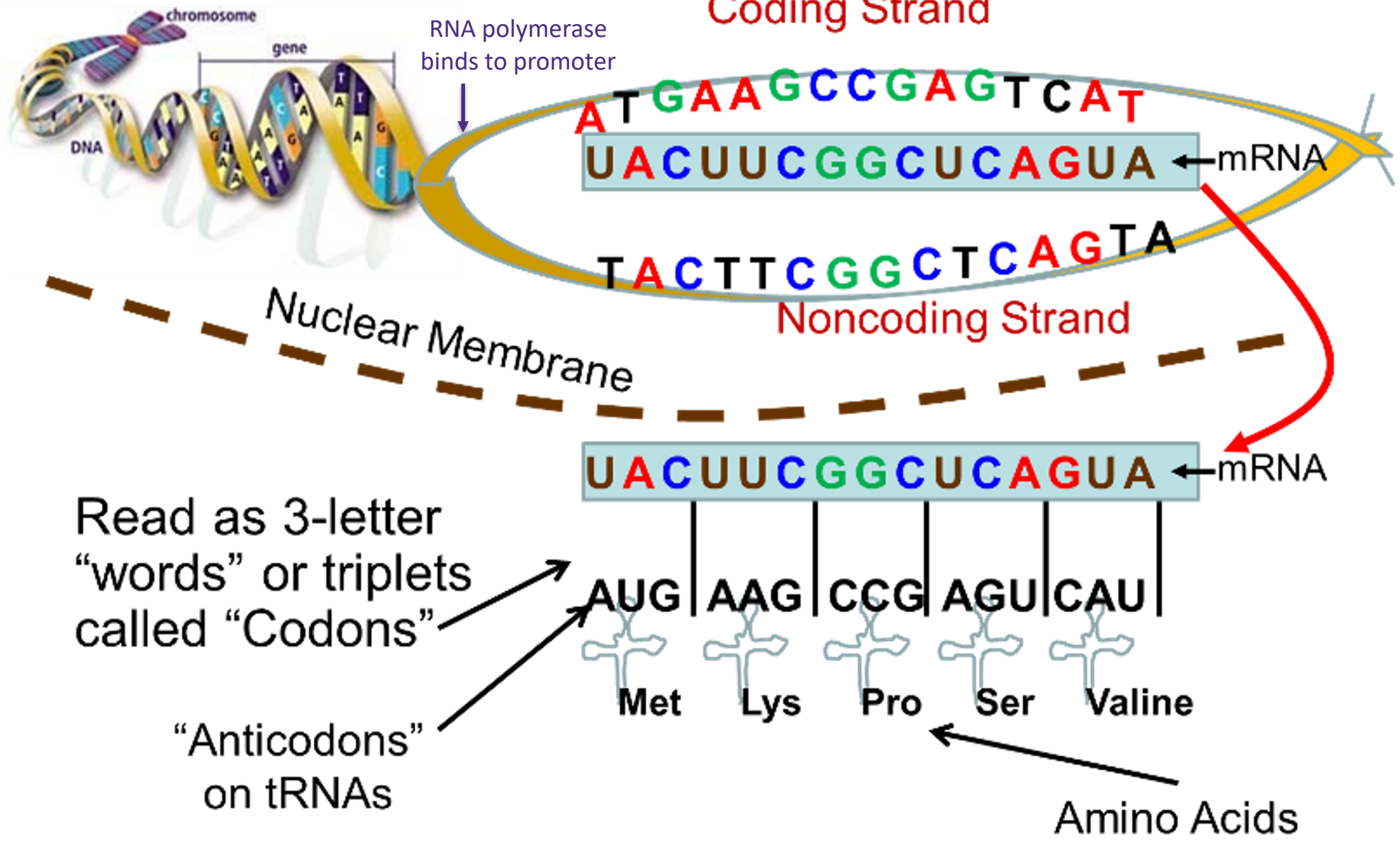
→ *A human genome accumulates ~ 64 new mutations per generation (6.4B bases total*1x10^{-8} mutation rate)*



Thymine (Yellow) = T Guanine (Green) = G

Adenine (Blue) = A Cytosine (Red) = C

Transcription and Translation



The Central Dogma

DNA

transcribed

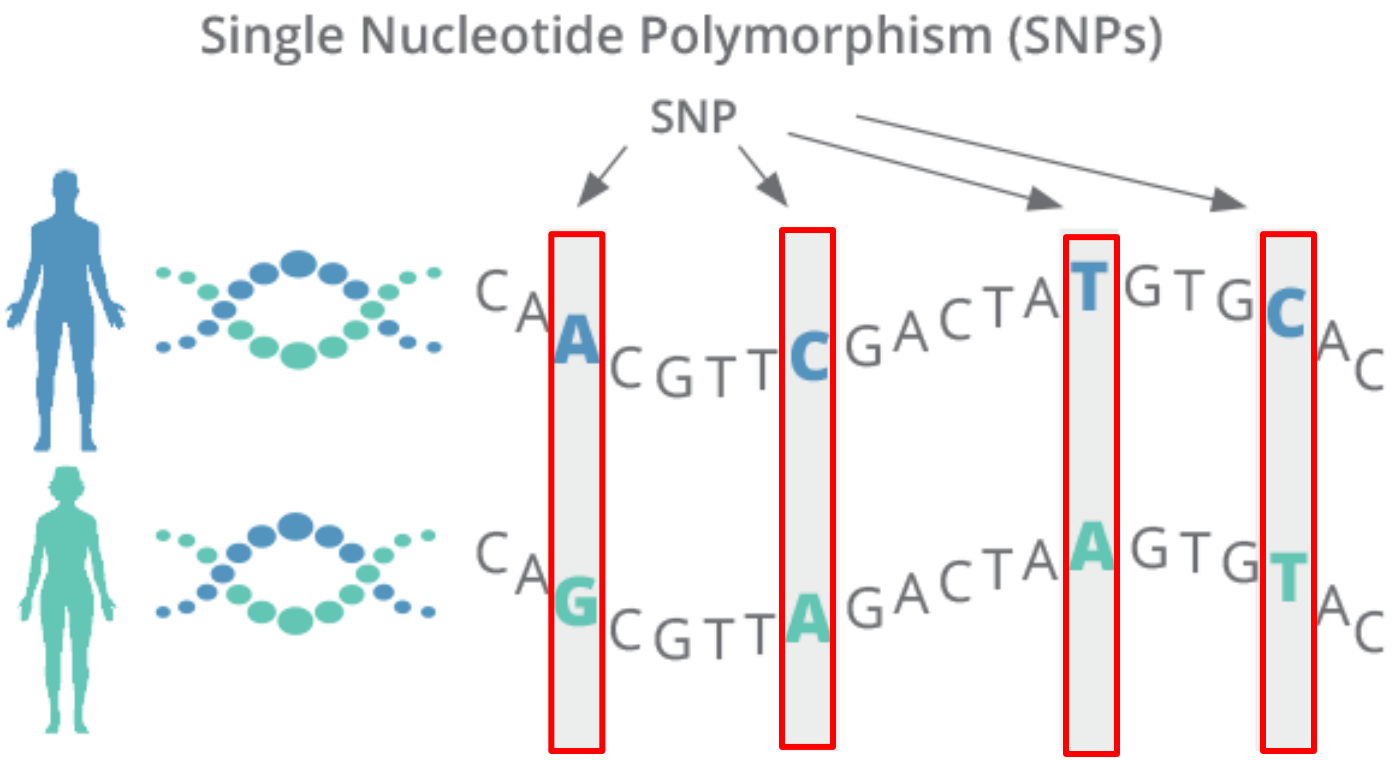
mRNA

translated

protein

4 possible bases (ACUG), 3 letters per codon → 4^3 = 64 possible codons that code for 20 amino acids

Genetic Variation: Single Nucleotide Polymorphisms (SNPs)



A recent study sequenced 40,722 individuals and identified 357 million SNPs

189 million SNPs (53%) only showed up in one individual (singletons)

Each person carried on average 3.6 million SNPs

Taliun et al, Nature 2021

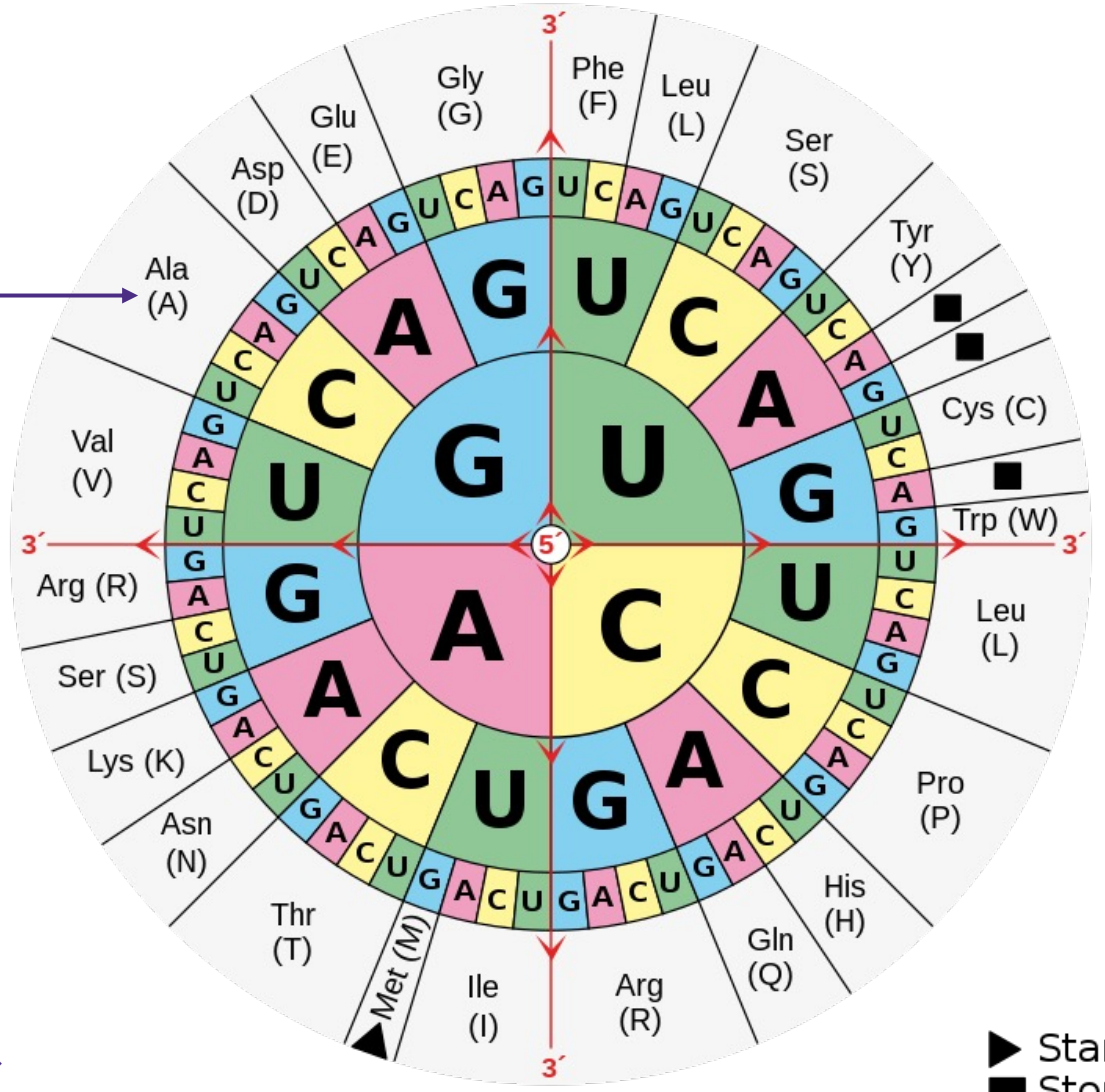
EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

Single base change = Single Nucleotide Polymorphism/Variant (SNP/SNV)

"Coding Variant" – affects protein translation

Synonymous Mutation

Mutated codon codes for the same amino acid (e.g., GCG [ala] → GCA [ala])

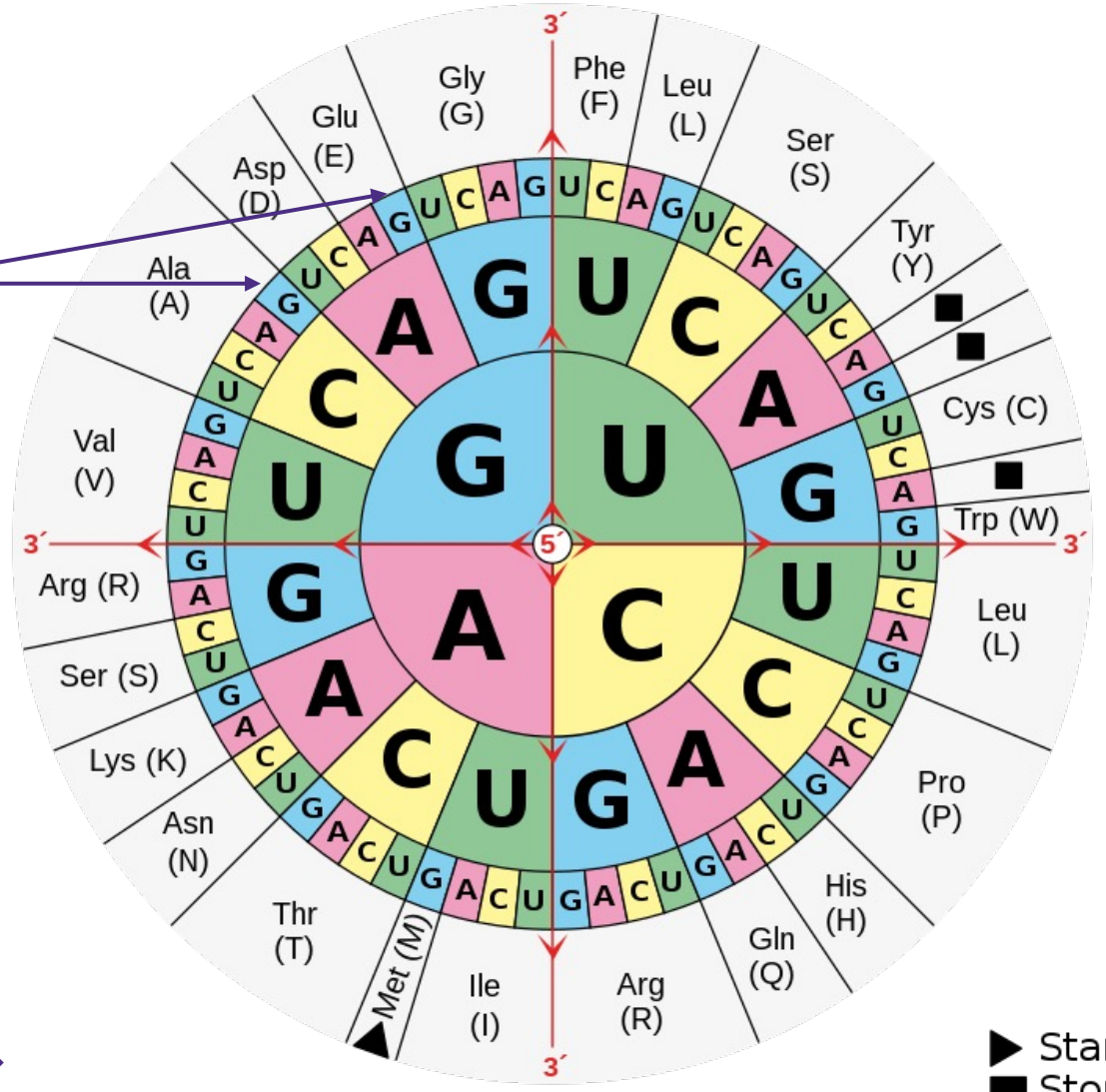


4 possible bases (ACUG), 3 letters per codon →
4^3 = 64 possible codons that code for 20 amino acids

Nonsynonymous mutations

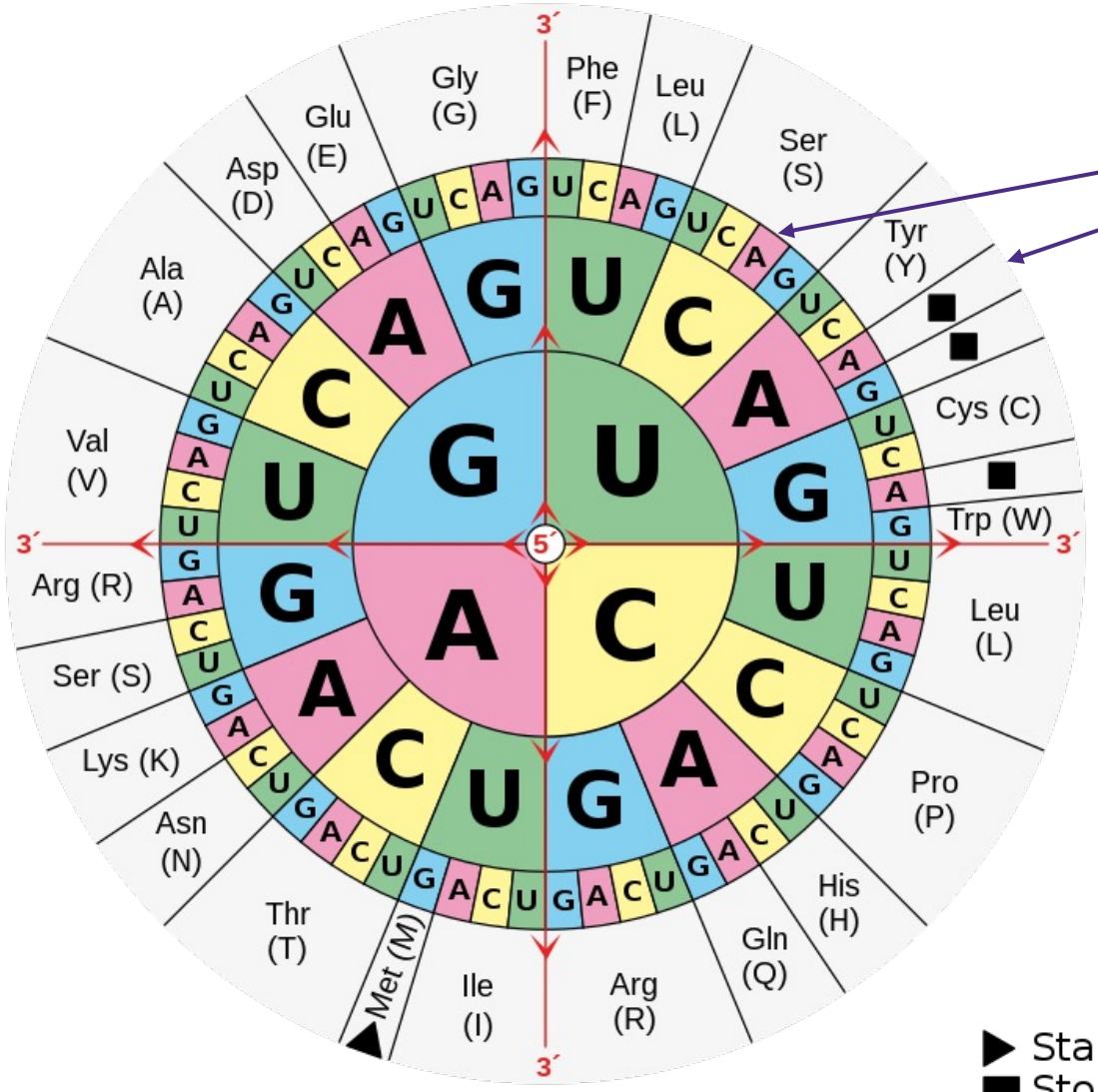
Missense Mutation

Mutated codon codes for a different amino acid (e.g., GCG [ala] → GAG [glu])



4 possible bases (ACUG), 3 letters per codon →
4^3 = 64 possible codons that code for 20 amino acids

Nonsynonymous mutations

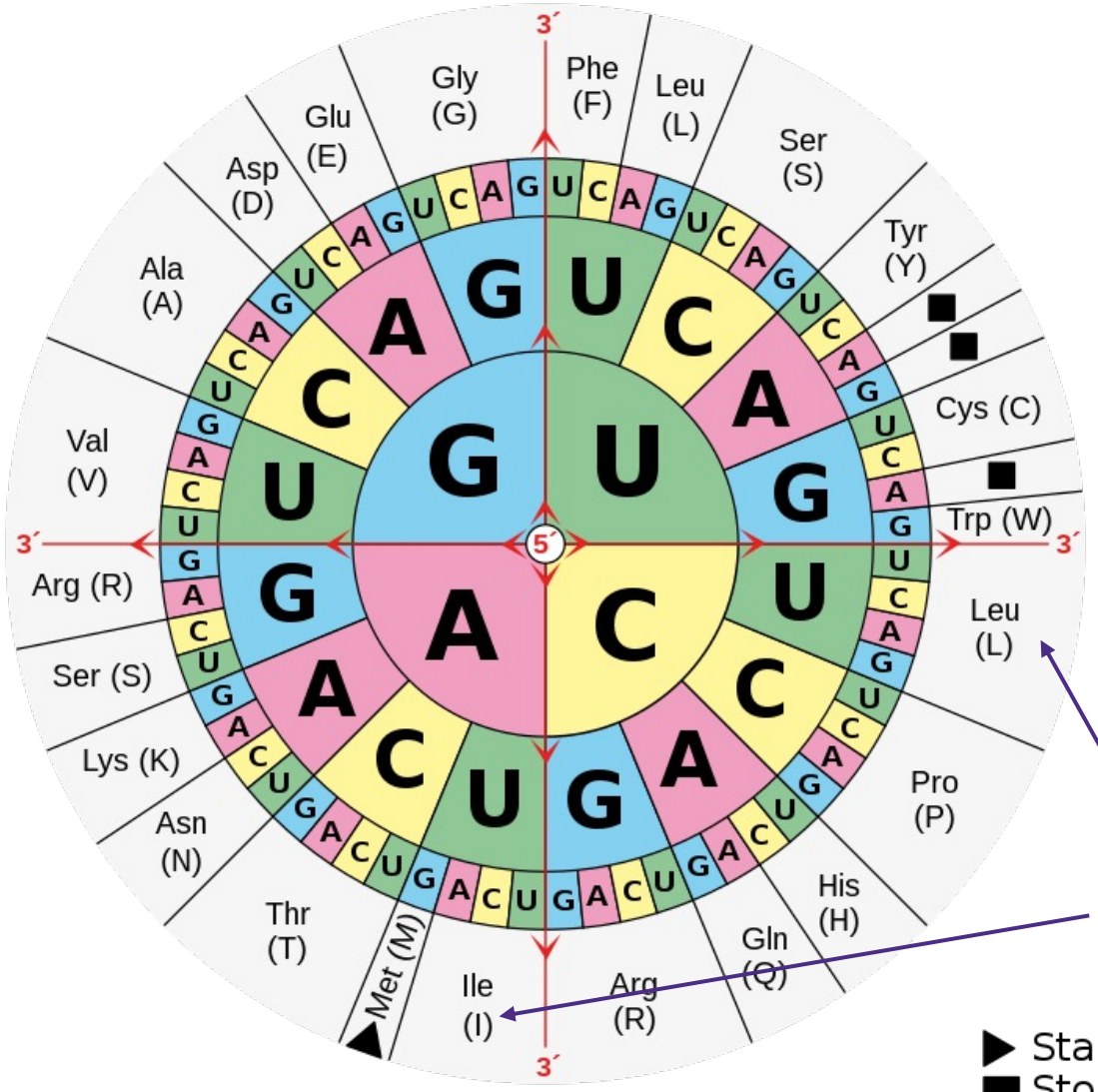


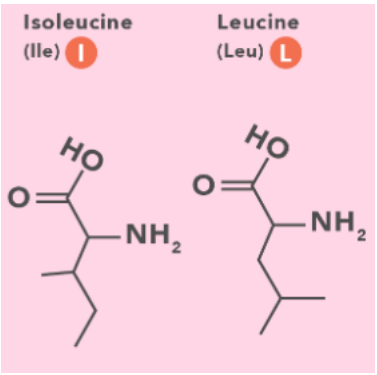
Nonsense mutation
Mutated codon is a premature stop codon (e.g., UCA [ser]→ UAA [stop])

4 possible bases (ACUG), 3 letters per codon →
4^3 = 64 possible codons that code for 20 amino acids

Some missense mutations are less severe

Bioinformatic tools to predict mutation severity
-Polyphen (3D protein structures)
-SIFT (uses probability of substitution across homologous sequences
-PhD-SNP
-PROVEAN



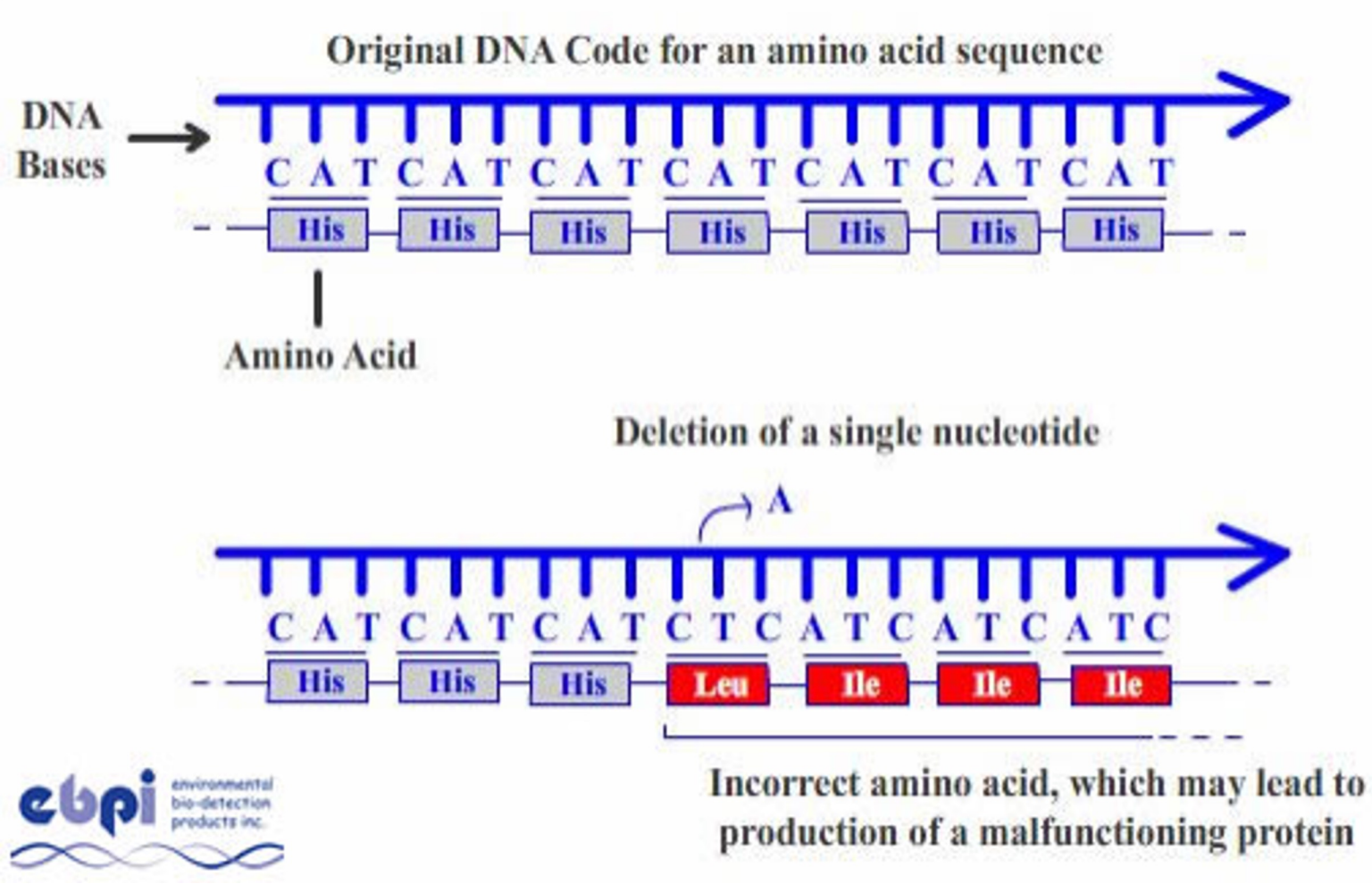


Both polar
Similarly have hydrophobic side chains

4 possible bases (ACUG), 3 letters per codon →
4^3 = 64 possible codons that code for 20 amino acids

Genetic Variation: Deletions/insertions

Frameshift mutations: caused by a deletion/insertion that shifts the way the sequence is read



Example of deletion: Cystic fibrosis F508del

Functioning CFTR sequence:

Nucleotide	ATC	ATC	TTT	GGT	GTT
Amino acid	Ile	Ile	Phe	Gly	Val

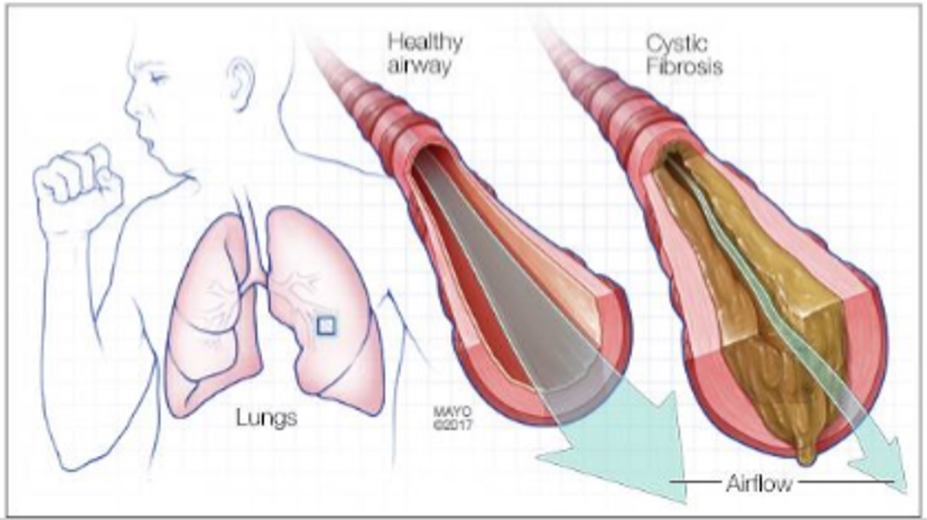
F508Del variant inactivating chloride channel:

Nucleotide	ATC	ATT	GGT	GTT
Amino acid	Ile	Ile	Gly	Val

Synonymous change
(ATC [Ile] → ATT [Ile])

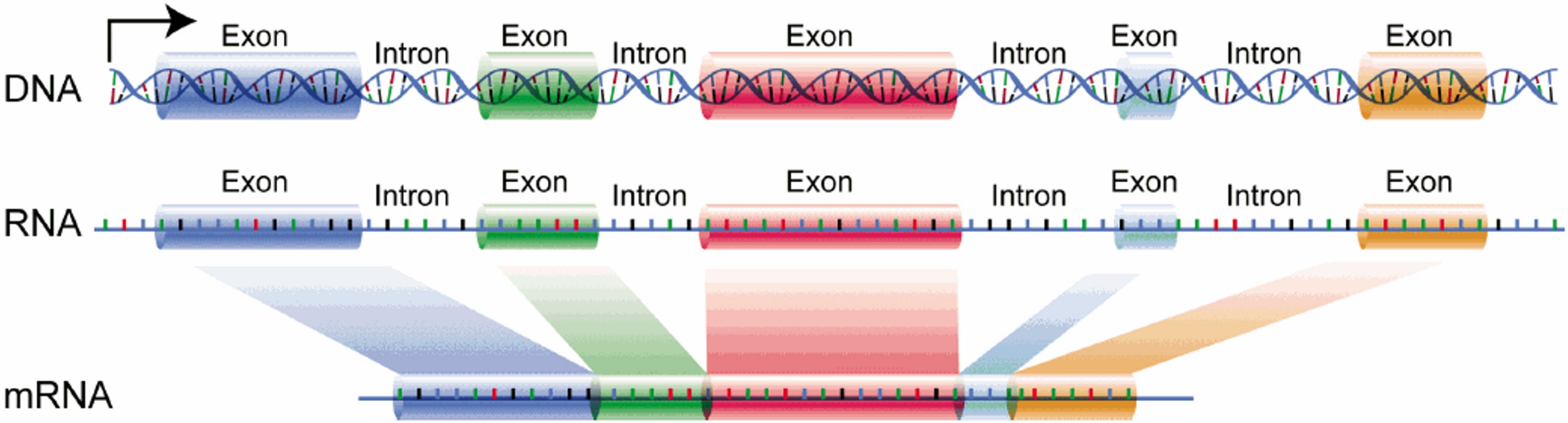
Phe is deleted, causes
CFTR protein to misfold

CFTR mutations cause mucus buildup in various organs

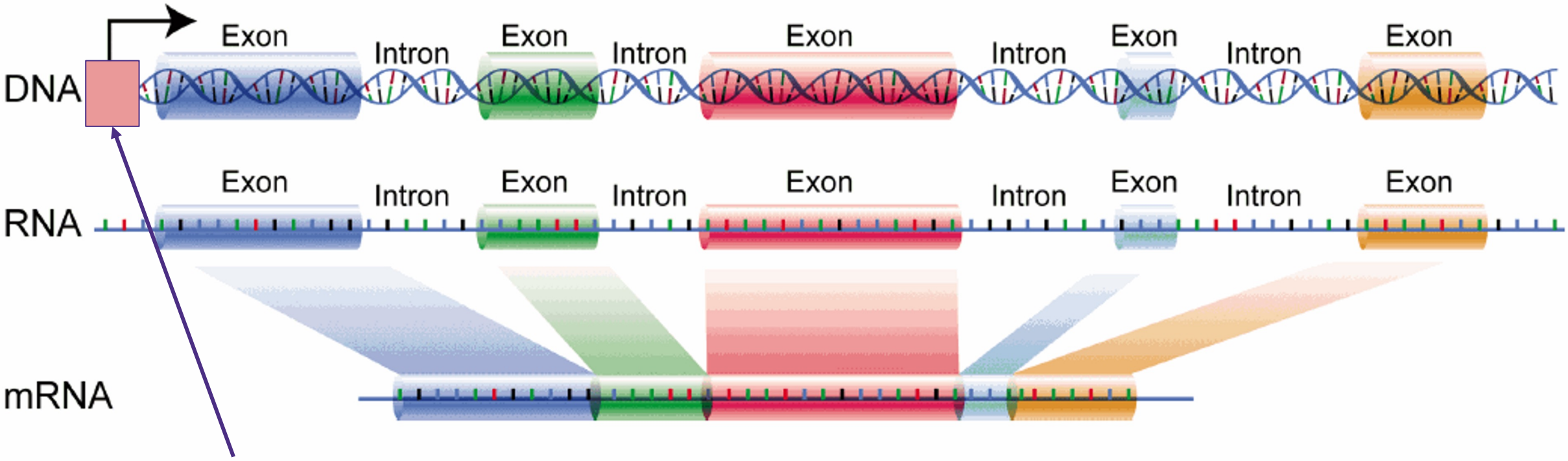


Sometimes mutations indirectly change the protein

A gene includes a lot of DNA that doesn't become protein

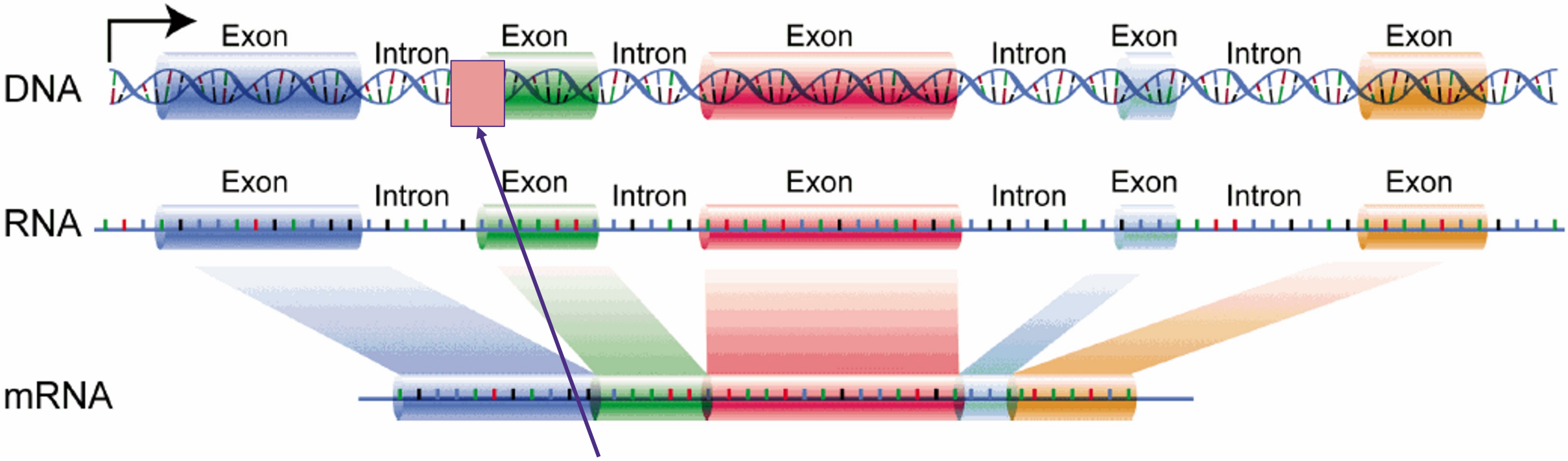


A gene includes a lot of DNA that doesn't become protein



A variant here can change gene "expression"

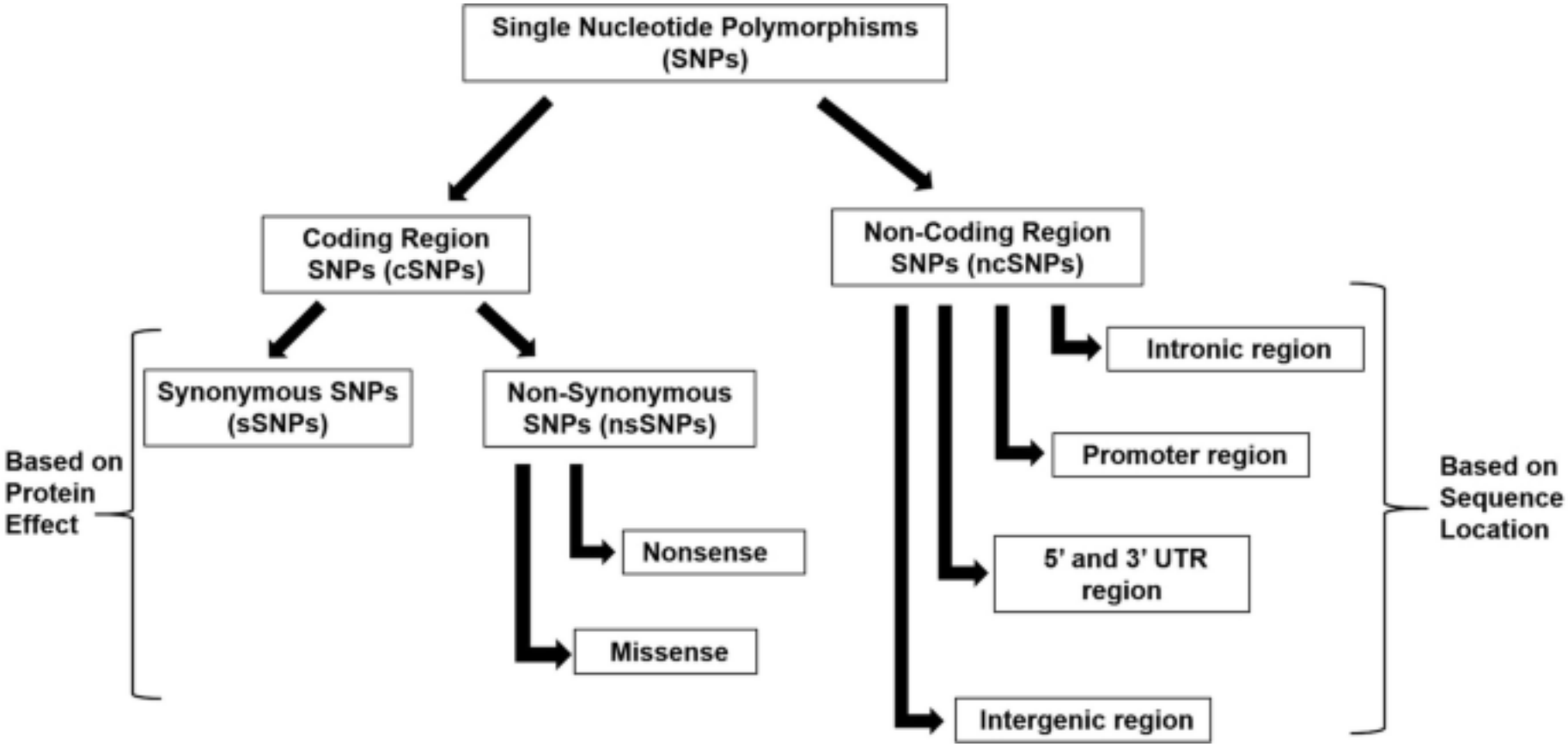
A gene includes a lot of DNA that doesn't become protein



Or here can change the "splice site" to make a different protein

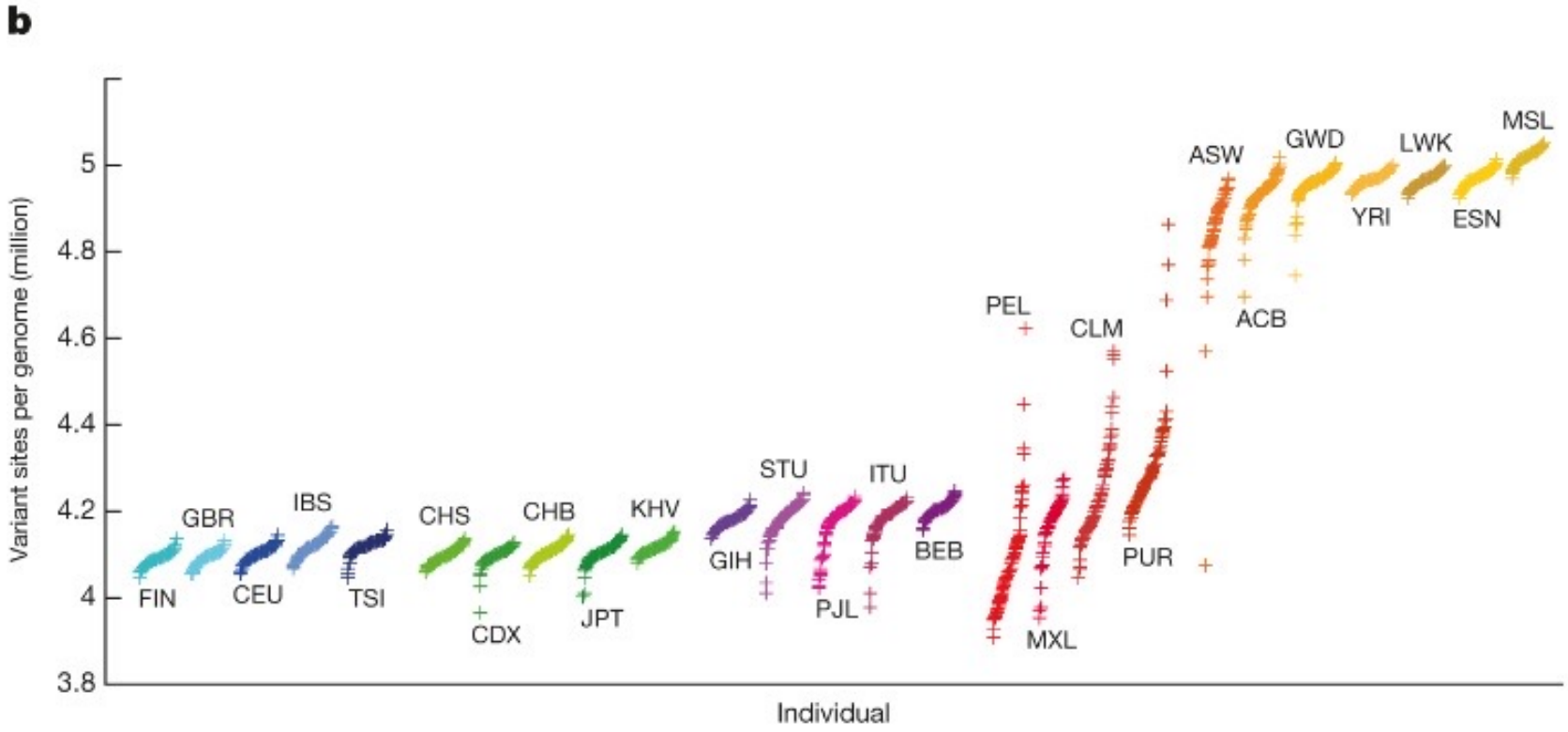
During transcription, introns are spliced out, removing non-coding regions before translation

Summary types of SNPs



Sukumar, S., Advances in Bioinformatics 2021

Genetic diversity is greatest in African Ancestry populations



	Population Color	Continental Group Color	Analysis Panel
Africa			AFR
			AFR
			AFR
			AFR
			AFR
			AFR/AMR
			AFR/AMR
Americas			AMR
			AMR
			AMR
			AMR
East Asian			EAS
			EAS
			EAS
			EAS
			EAS
European			EUR
			EUR
			EUR
			EUR
			EUR
South Asian			SAS
			SAS
			SAS
			SAS
			SAS

1000 Genomes project, Nature 2015

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

BREAKOUT ACTIVITY

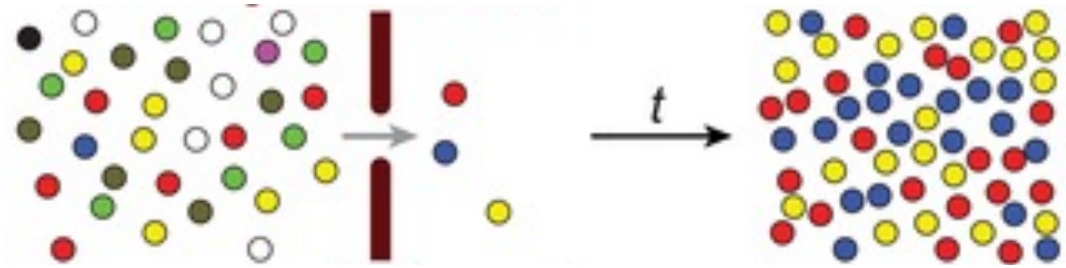
Q1. Why do we see the greatest genetic diversity in African ancestry populations?

Q2. Based on a recent sequencing project (N=40,722 individuals), each person carried on average 3.7 million genetic variants. Of those, 23,909 variants (0.6%) were located in coding regions, which constitute 1.5% of the genome. Why do you think this discrepancy (0.6% vs 1.5%) exists?

Q3. Look up “rs6025” in dbSNP (https://www.ncbi.nlm.nih.gov/snp/).

- 3a. What are the alleles identified at this location?
- 3b. What gene is this SNP located in?
- 3c. Click on the “frequency” tab to the left. What is the frequency of the minor allele (less common allele) overall? How do these frequencies differ by ancestral subgroups?

The “Out-of-Africa” migration is an example of a population bottleneck



deMenocal & Stringer, Nature 2016

Abel, PLoS Pathogenics 2015

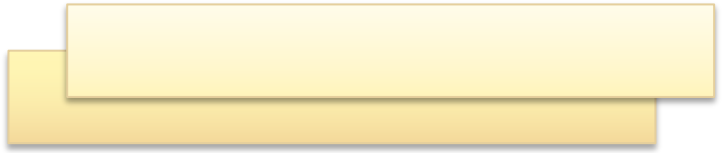
EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

DNA inheritance

We inherit “blocks” of the genome from our parents (and not independent base-pairs)

Fathers DNA

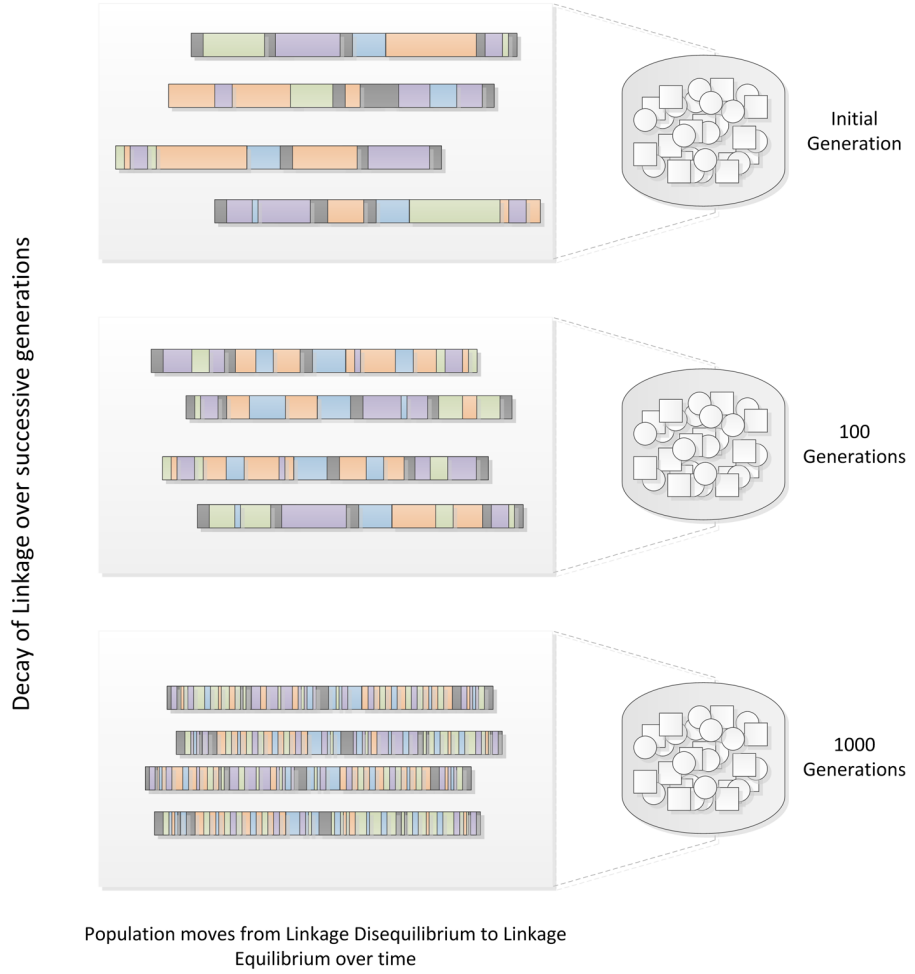
Mothers DNA



Childs DNA

The size of the blocks decreases with time

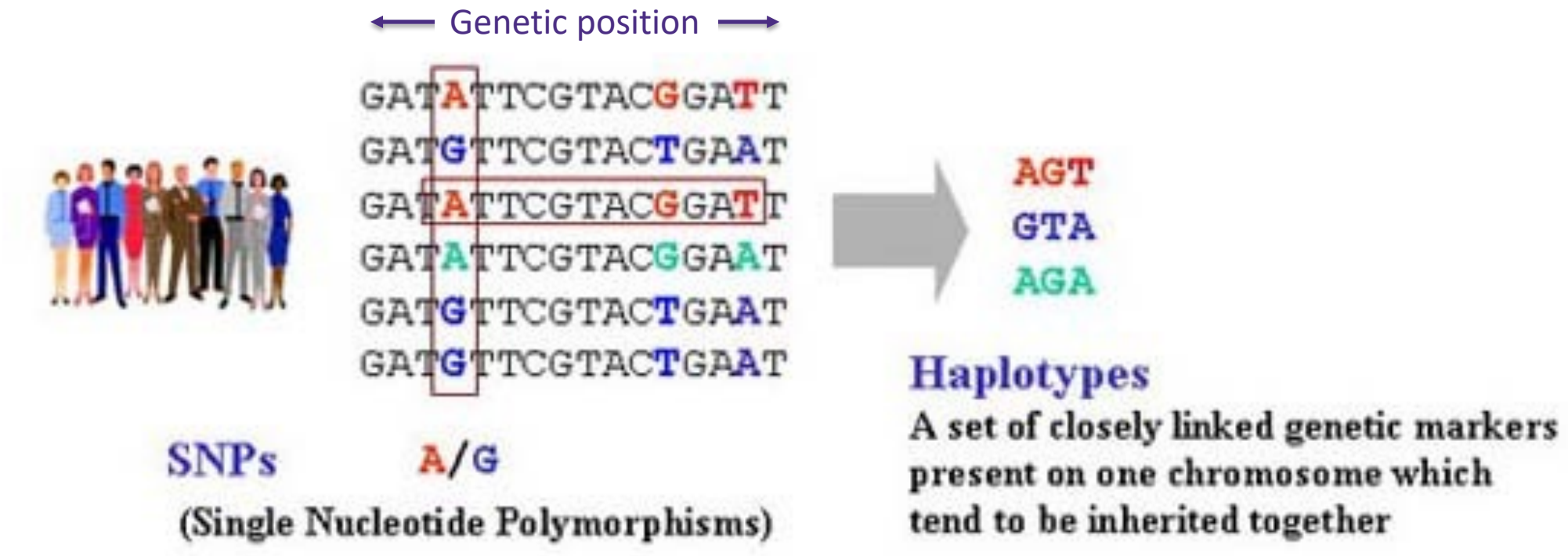
Descendants DNA



Bush & Moore, PLOS Comp Bio 2012

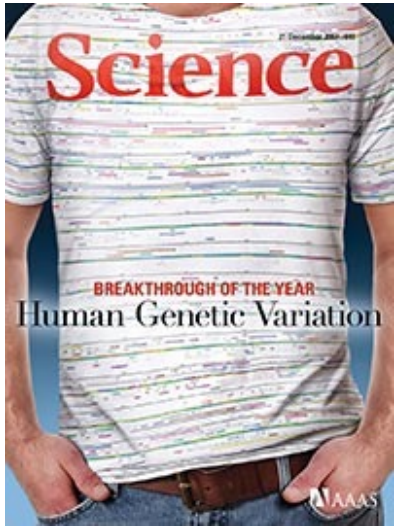
Haplotypes

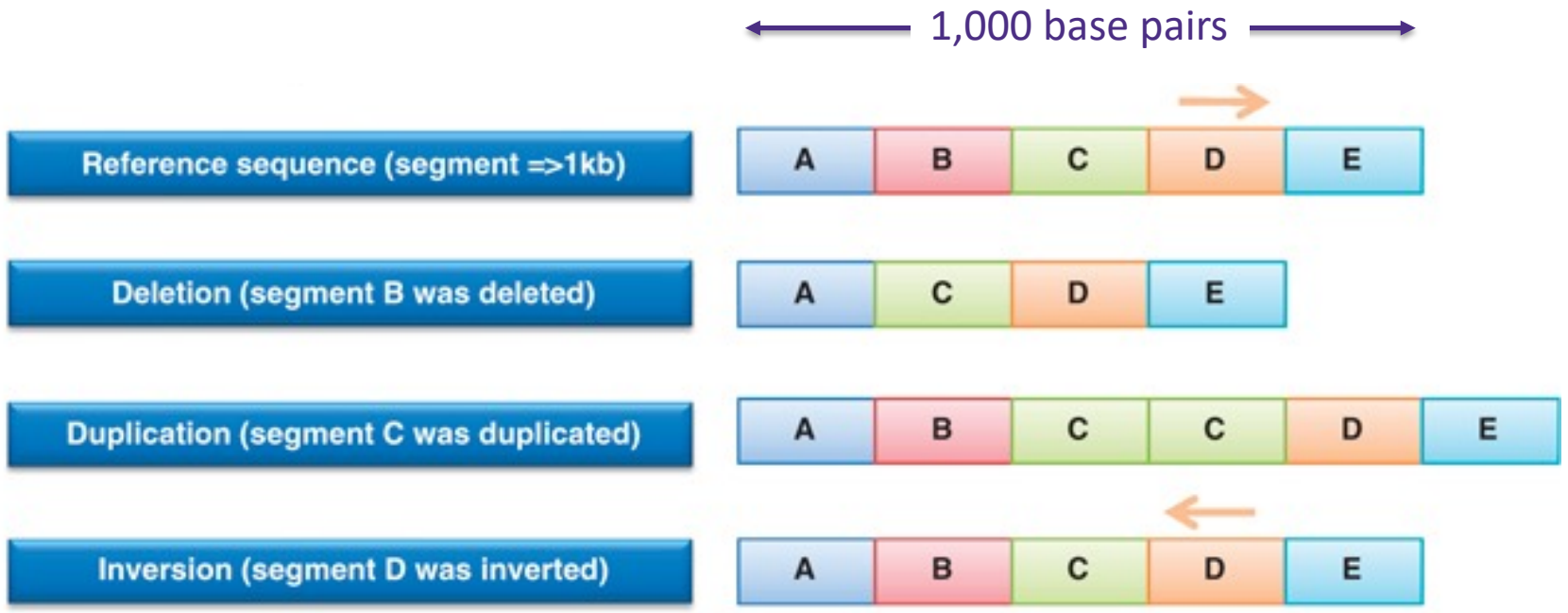
Specific combination of SNPs occurring on the same segment of chromosome that are inherited together.



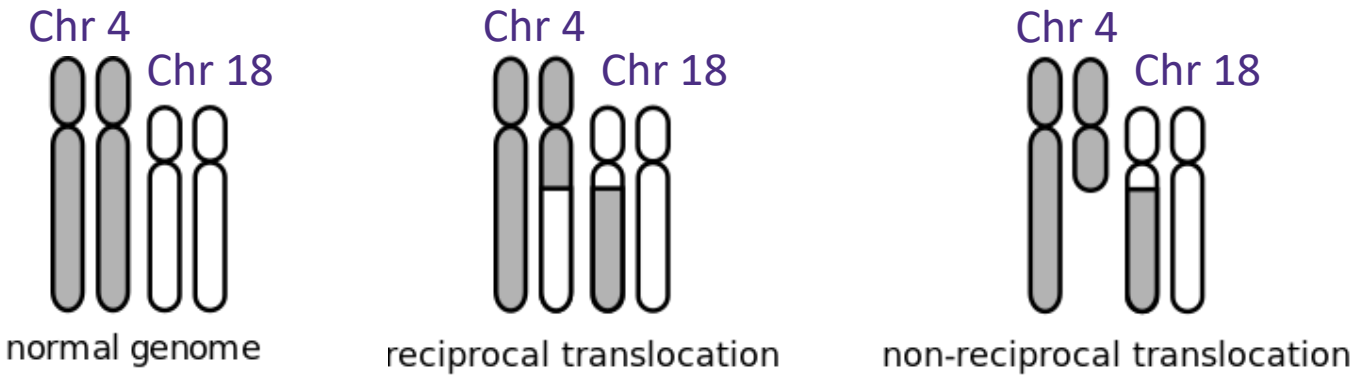
Besides single base changes, what are other types of changes?

Genetic Variation: Structural Variation

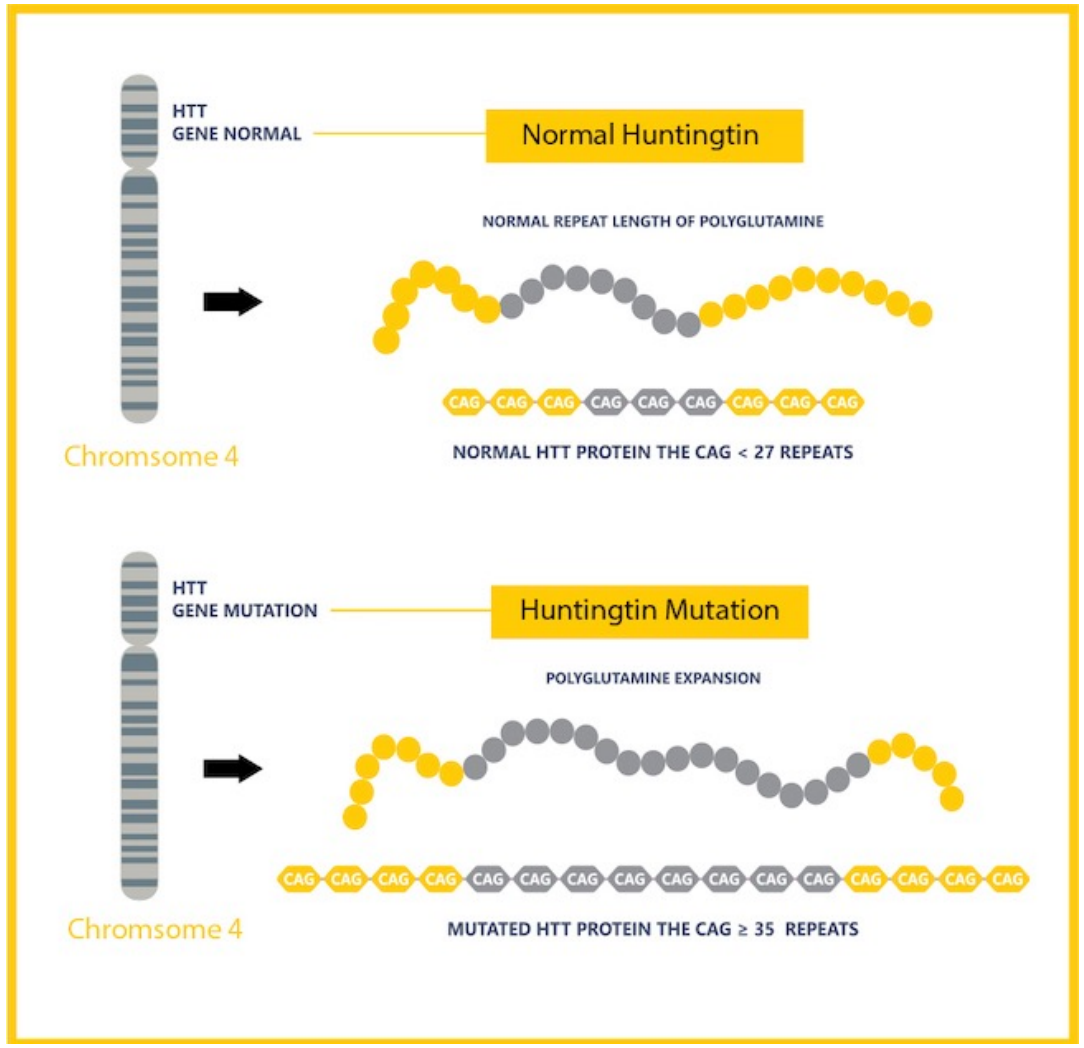




Translocation



Structural variation example: Tandem repeats in *HTT* (Huntington's disease)



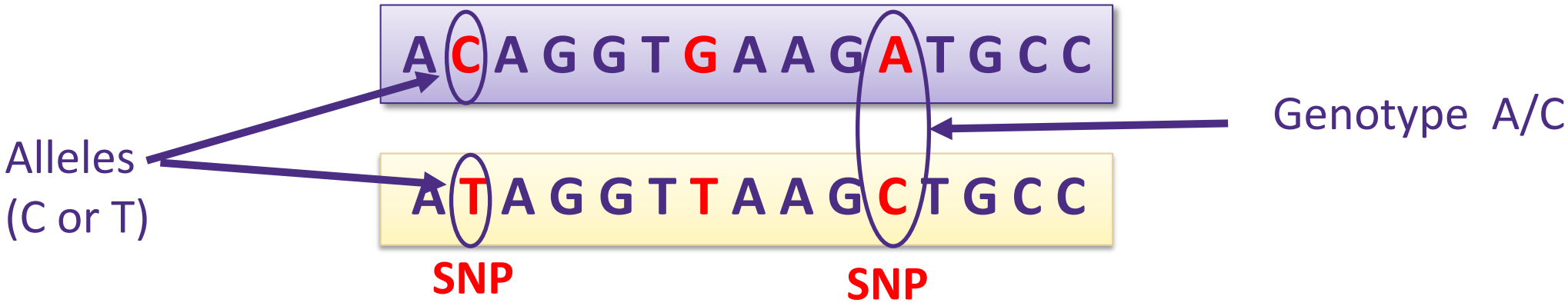
The normal function of huntingtin protein is unknown. The CAG repeats (polyglutamine) are in some way neurotoxic.

Huntington's status		CAG repeat length
Unaffected	Normal	10-26
	Intermediate allele	27-35
Affected	Reduced penetrance	36-39
	Full penetrance	40+

Alleles to genotypes and phenotypes

Allele vs. genotype

> We inherit two copies of each chromosome



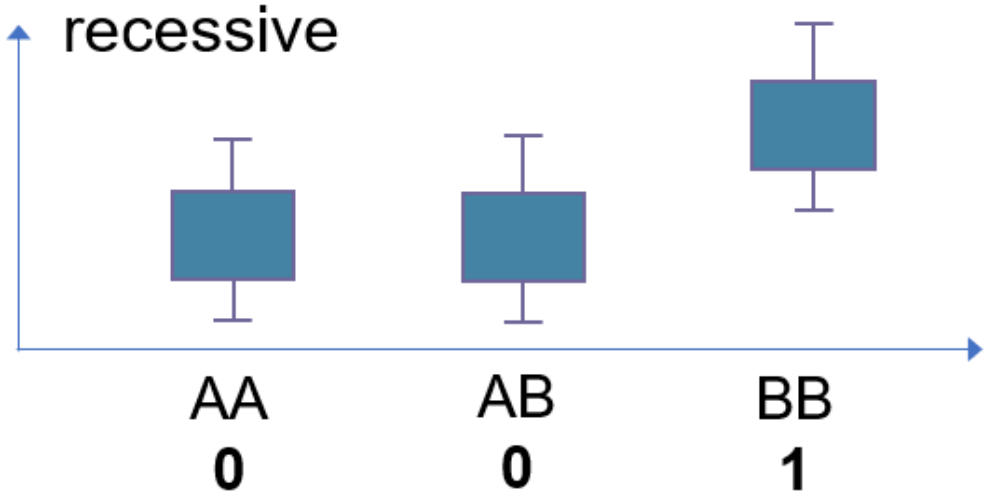
Genotypes
(A/A) – homozygous
(A/C) – heterozygous
(C/C) - homozygous

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

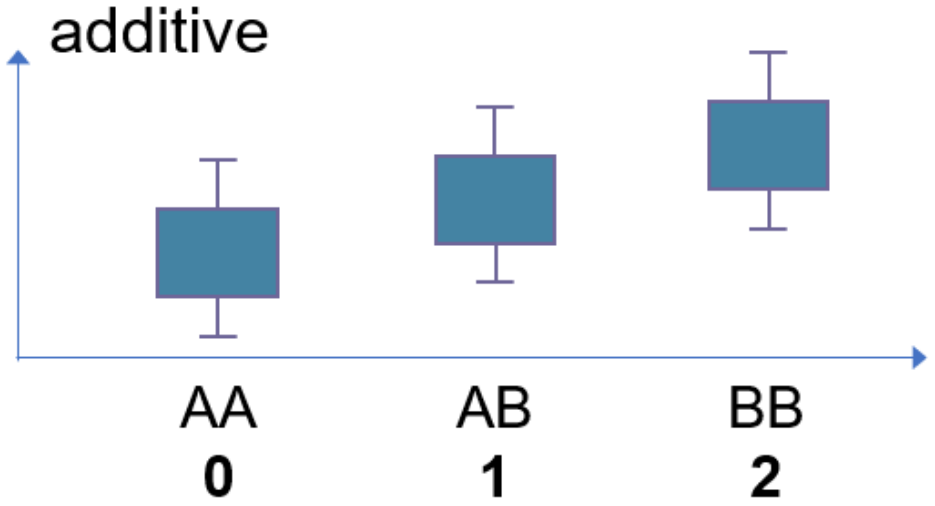
Inheritance Patterns: Genotype → Phenotype

- 2 copies of every gene/chromosome (most common)
  - Dominant (only need one copy of a variant to see the effect)
  - Recessive (need two copies of the variant to see the effect)
  - Additive (the effect of one variant is ½ that of two variants)

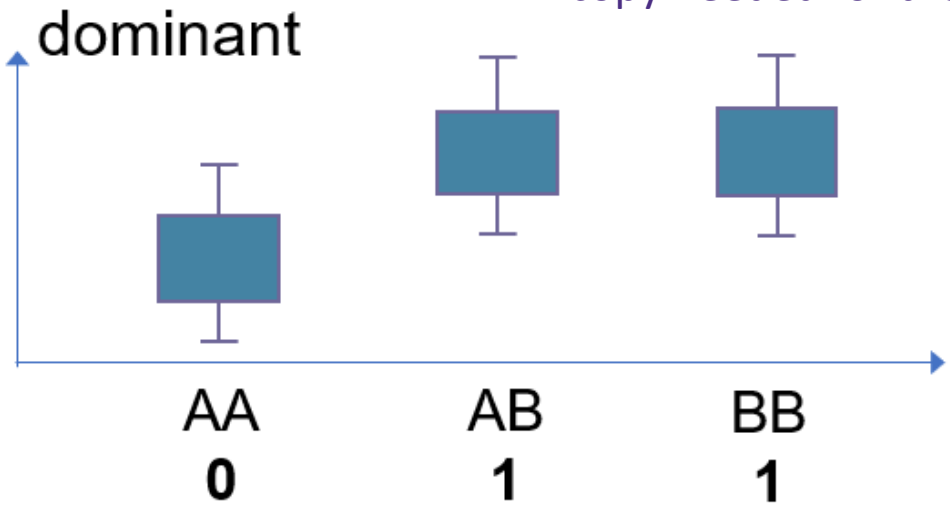
Inheritance Patterns: Genotype → Phenotype



Example: Cystic fibrosis
2 copies needed for the phenotype



Example: Huntington's disease
1 copy needed for the phenotype

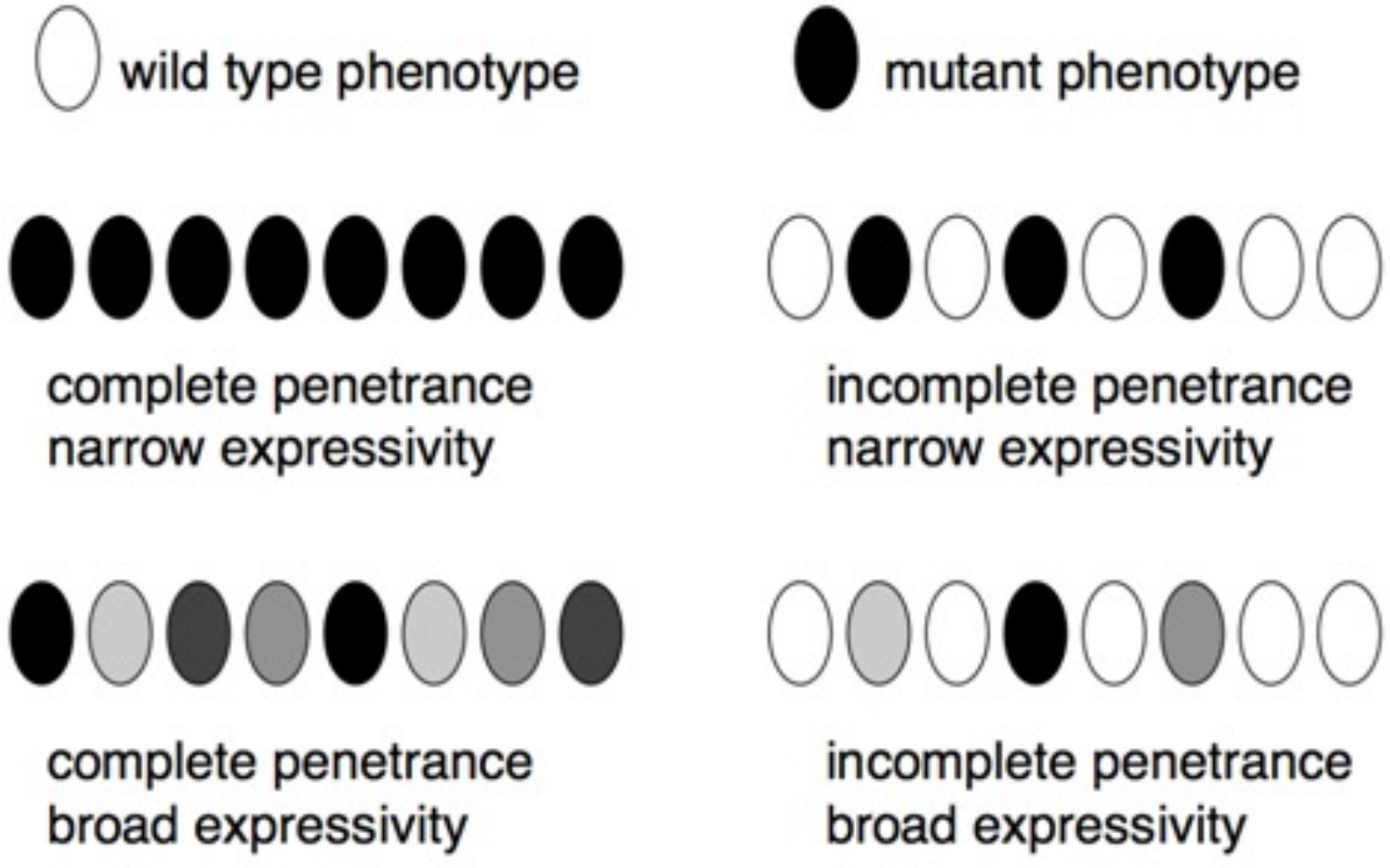


Genotype → Phenotypes

- **Mendelian phenotype** is one driven by variation at a single genetic locus (e.g., Huntington's disease, Cystic fibrosis).
- **Complex phenotype** does not show such simple patterns of inheritance (e.g., height).
  - oligogenic (a few genetic loci)
  - polygenic (many genetic loci)

Penetrance and Expressivity

Each oval represents an individual. All individuals have the same genotypes.



Penetrance: Proportion of individuals with the same genotype who have an expected phenotype

Expressivity: Extent to which a phenotype is expressed in individuals with a given genotype

Breakout Activity

1. Match the genetic term with the definition:

a. Nonsense
b. Heterozygous
c. Exon
d. Allele
e. Synonymous
f. Missense
g. Non-coding region
h. Haplotype
i. Autosomal
j. Phenotype
k. Genotype
l. Frameshift
m. Intron
n. Homozygous

____ Alternative forms of a gene or DNA base.
____ Genetic makeup of an individual at a particular DNA location based on both alleles.
____ Genotype consisting of two different alleles at a particular location.
____ DNA base change that does not change the translated amino acid.
____ Genotype consisting of two of the same alleles at a particular location.
____ Observable characteristics resulting from a genotype.
____ Concerning the 22 pairs of chromosomes that are not sex chromosomes.
____ Portion of gene that does not code for amino acids and appears in between exons.
____ Insertion or deletion mutation that changes the whole subsequent sequence of amino acids by changing the 3-codon groups for generating amino acids.
____ Portion of gene that encodes amino acids.
____ Section of DNA that does not become protein.
____ Substitution of a single DNA base that causes a stop in protein production.
____ DNA base change that changes the translated amino acid.
____ Set of DNA variations at several positions that are inherited together.

Breakout Activity

1. Match the genetic term with the definition:

a. Nonsense
b. Heterozygous
c. Exon
d. Allele
e. Synonymous
f. Missense
g. Non-coding region
h. Haplotype
i. Autosomal
j. Phenotype
k. Genotype
l. Frameshift
m. Intron
n. Homozygous

d___ Alternative forms of a gene or DNA base.
k___ Genetic makeup of an individual at a particular DNA location based on both alleles.
b___ Genotype consisting of two different alleles at a particular location.
e___ DNA base change that does not change the translated amino acid.
n___ Genotype consisting of two of the same alleles at a particular location.
j ___ Observable characteristics resulting from a genotype.
i ___ Concerning the 22 pairs of chromosomes that are not sex chromosomes.
m___ Portion of gene that does not code for amino acids and appears in between exons.
l ___ Insertion or deletion mutation that changes the whole subsequent sequence of amino acids by changing the 3-codon groups for generating amino acids.
c___ Portion of gene that encodes amino acids.
g___ Section of DNA that does not become protein.
a___ Substitution of a single DNA base that causes a stop in protein production.
f ___ DNA base change that changes the translated amino acid.
h___ Set of DNA variations at several positions that are inherited together.