Day	Time (PT)	Lead	Topics	Details
Wed	11:45-12.30	Alie/Sara	Class Intro	Intro to class/agenda and topics Technology/logistics Breakout room introductions
			BREAK	
	12.45-1.30	Sara	Overview of Genetic Epi	Intro to Epidemiology and Genetic Epidemiology
			BREAK	
	1:45-2:30	Alie	Human genetic variation	Types of genetic variation

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

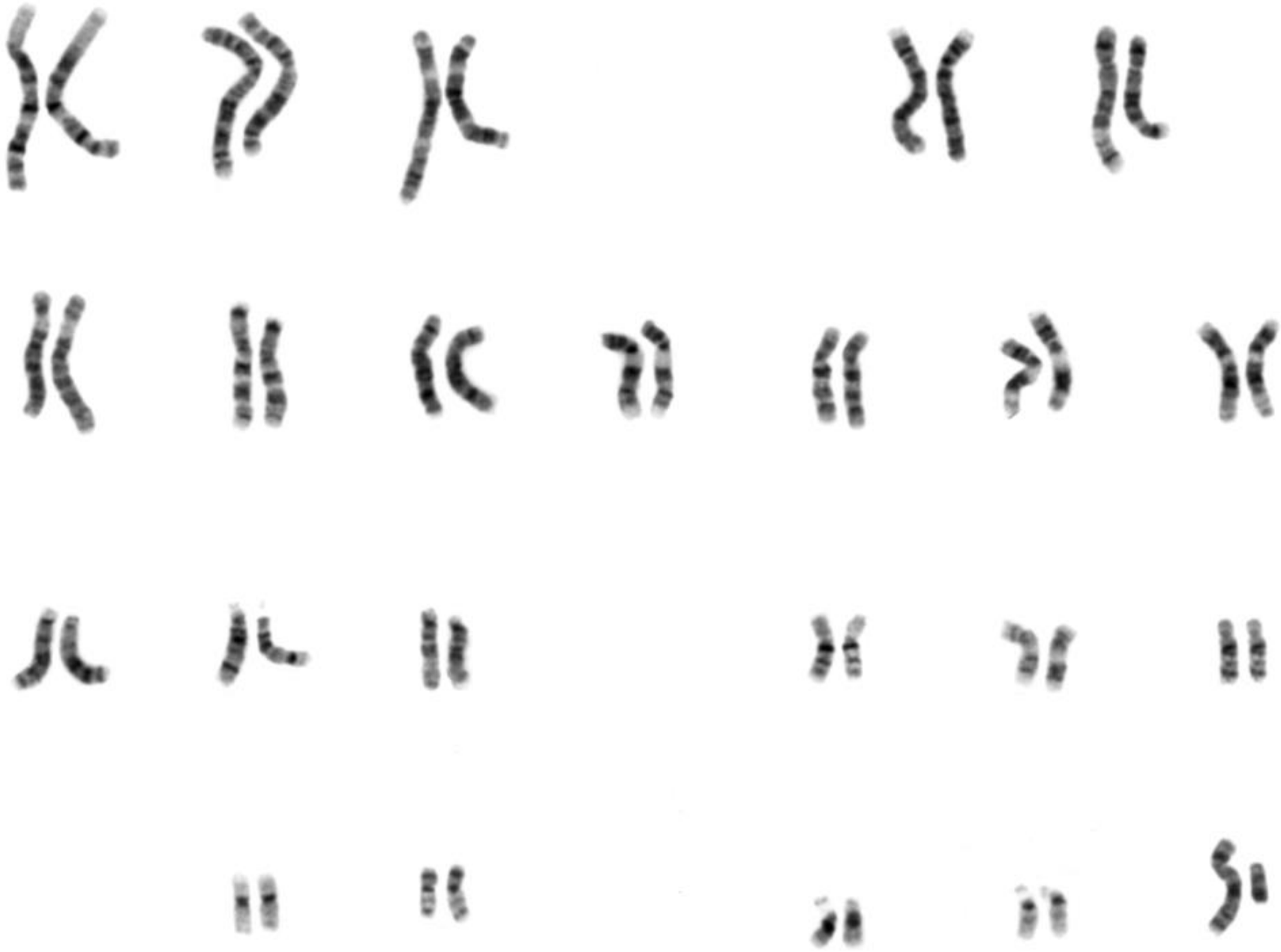
SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

SISG 2022:
Module 11
Genetic Epidemiology

Human Genetic Variation

Learning Objectives

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



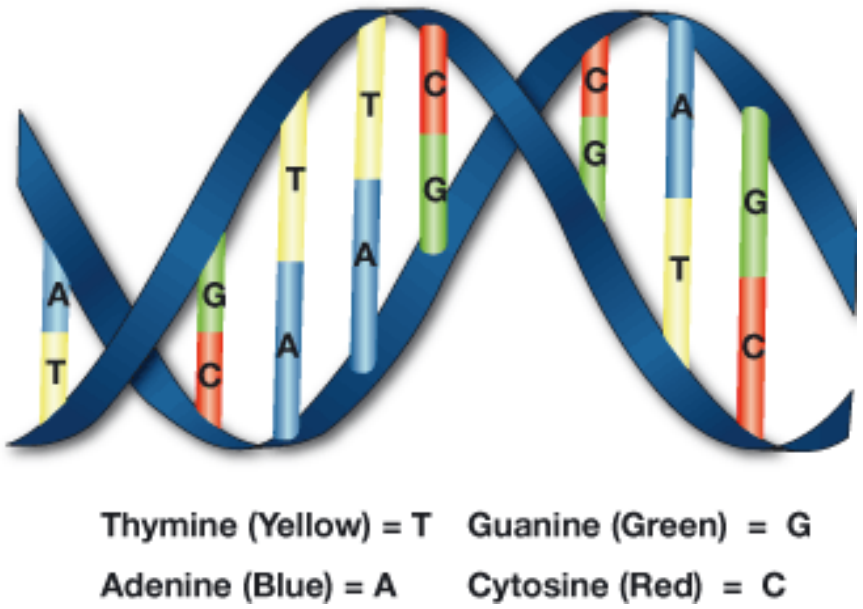
Our Genome in Numbers

23 chromosome pairs

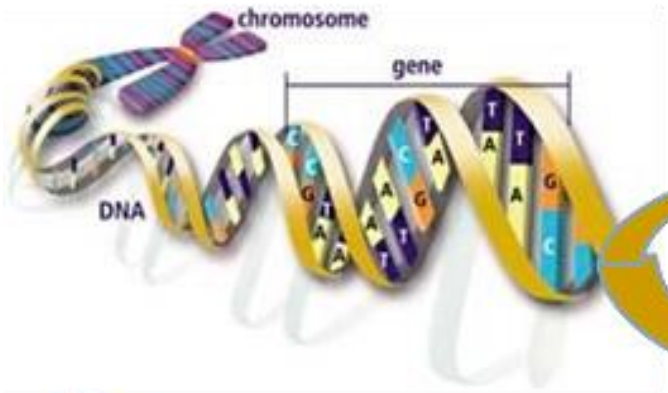
3.2 billion base-pairs (A,C,G,T)

~20,000 genes

~1.5% of the genome is coding DNA



Transcription and Translation



Coding Strand

ATGAAGCCGAGTCAT

UACUUCGGCUCAGUA ← mRNA

TACTTCGGCTCAGTA

Noncoding Strand

Nuclear Membrane

UACUUCGGCUCAGUA ← mRNA

Read as 3-letter "words" or triplets called "Codons"

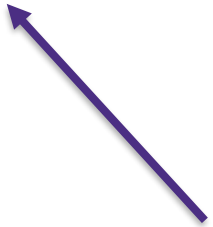
AUG | AAG | CCG | AGU | CAU

Met Lys Pro Ser Valine

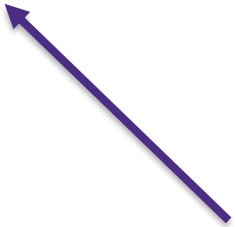
"Anticodons" on tRNAs

Amino Acids

Genetic variation to phenotype variation

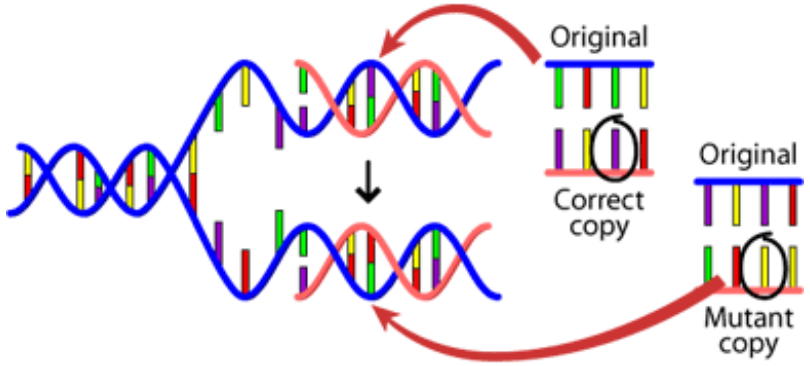


The changes in DNA



What we actually see (disease, trait)

Mutations happen all the time, with every replication



Human genome mutation rate is ~1.1×10^{-8} per site per generation.

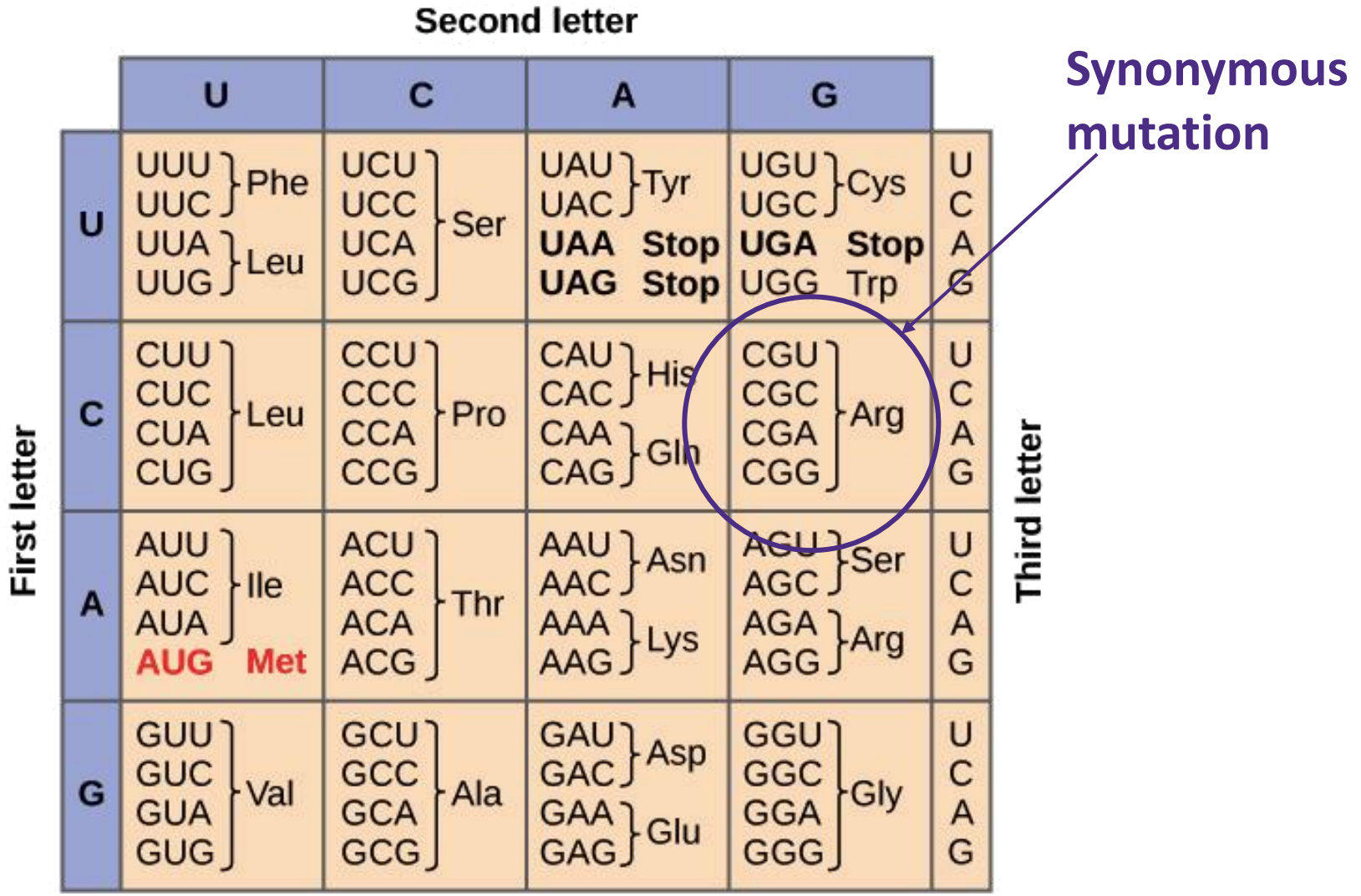
Human genome is over 3 billion base pairs.

Each genome: 3,000,000,000 sites
Mutation rate: 0.000000011 errors/site

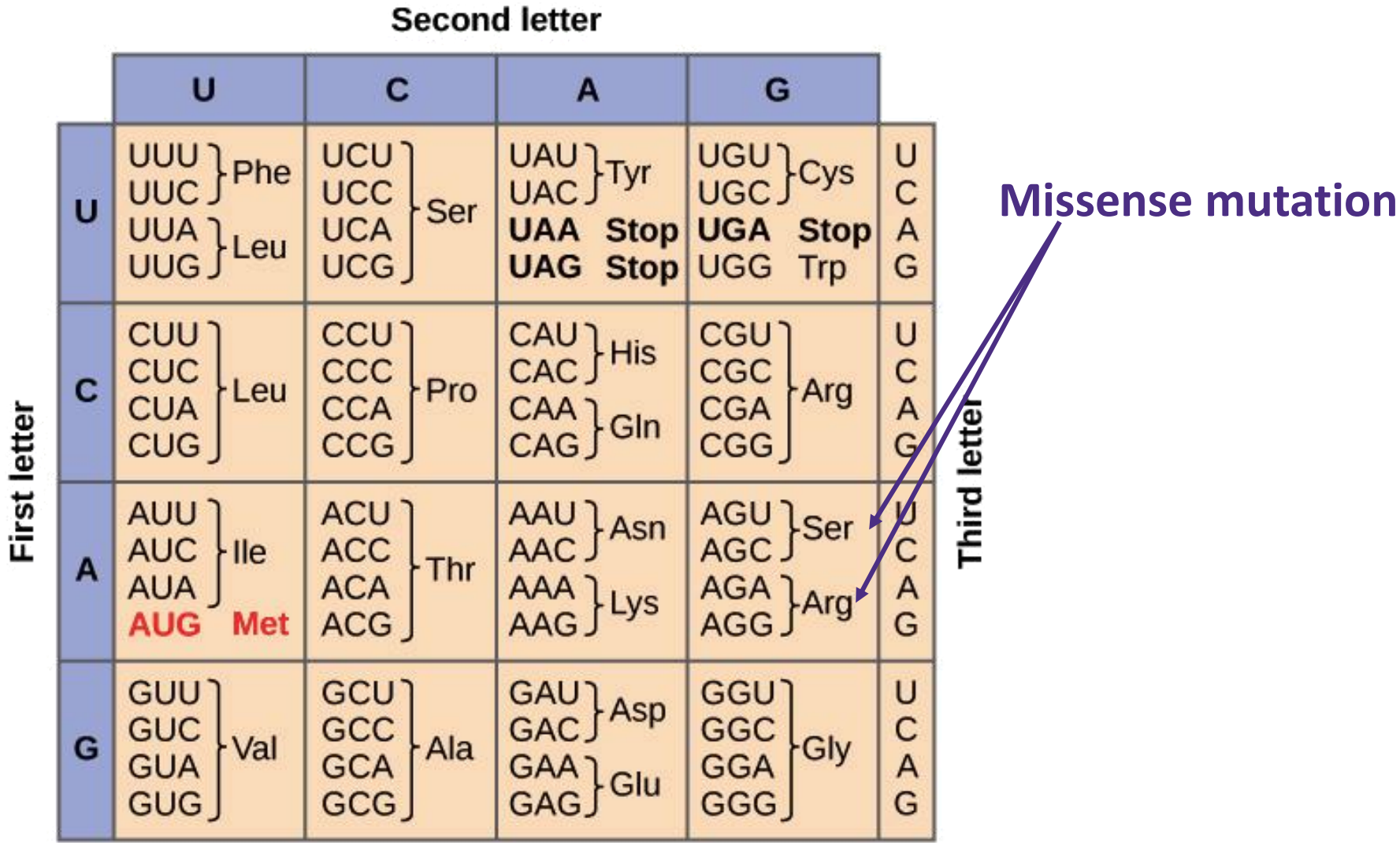
How many new mutations do you expect in each cell replication?

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

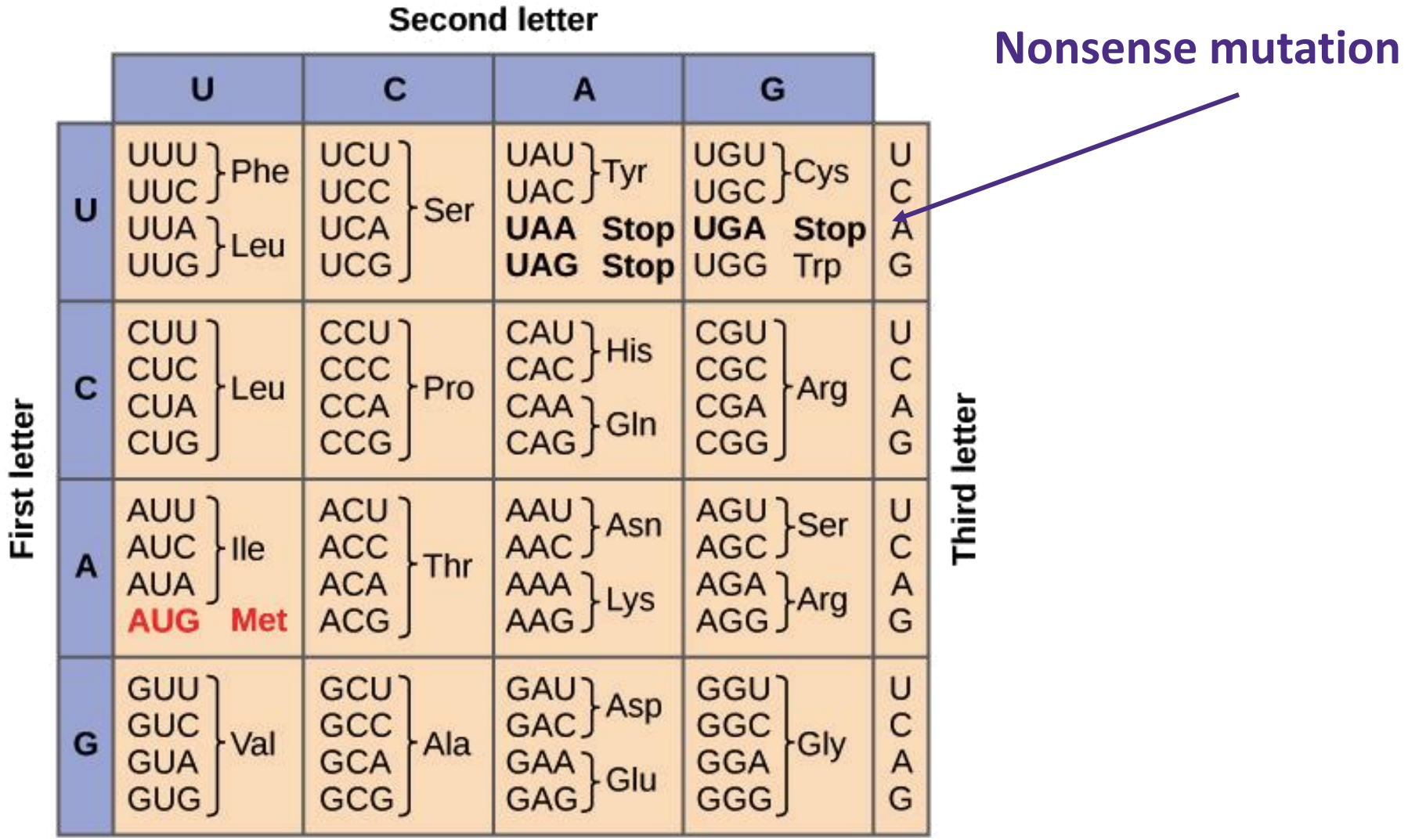
"Coding Variant" – affects protein translation



Nonsynonymous mutations



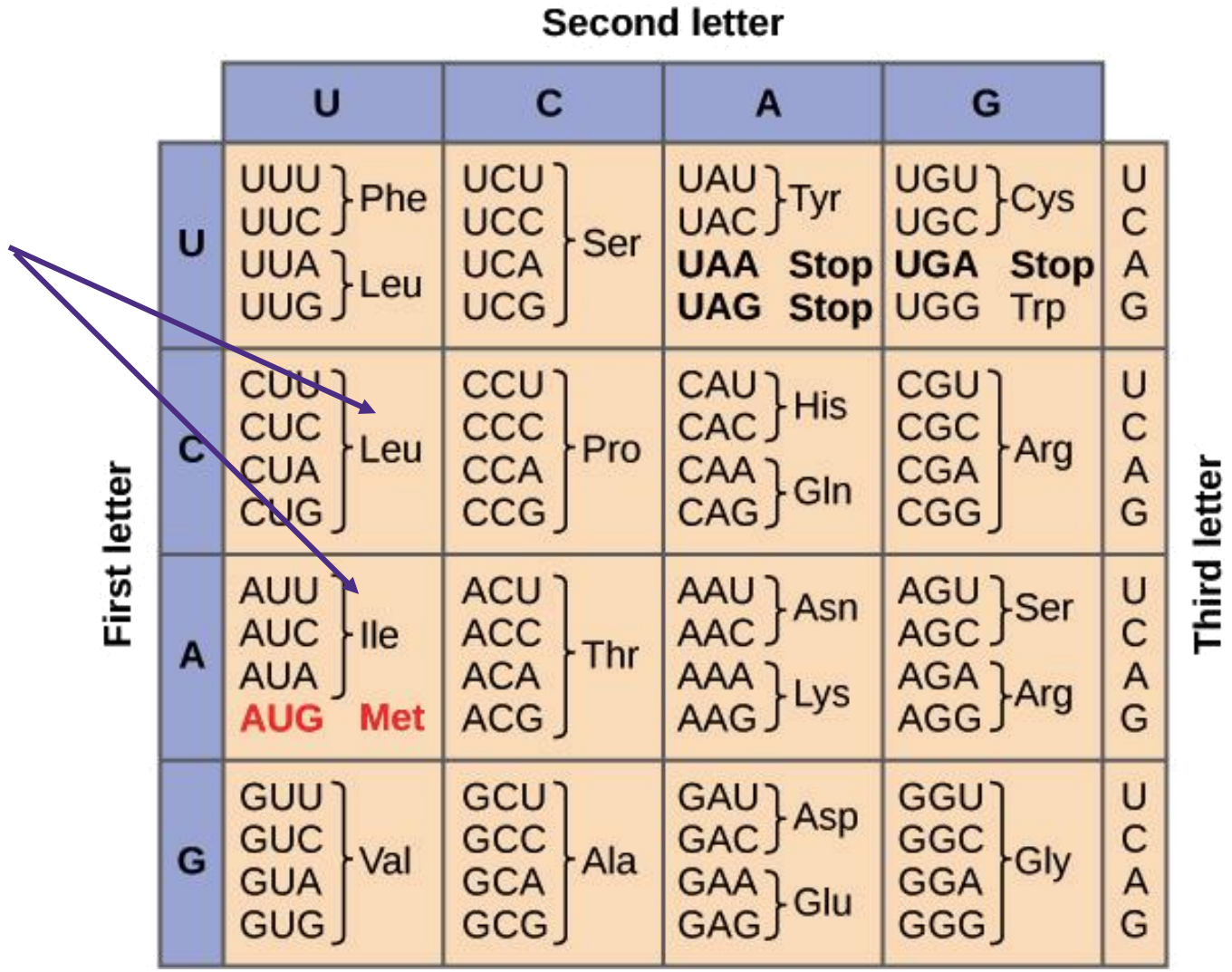
Nonsynonymous mutations



Nonsense mutation

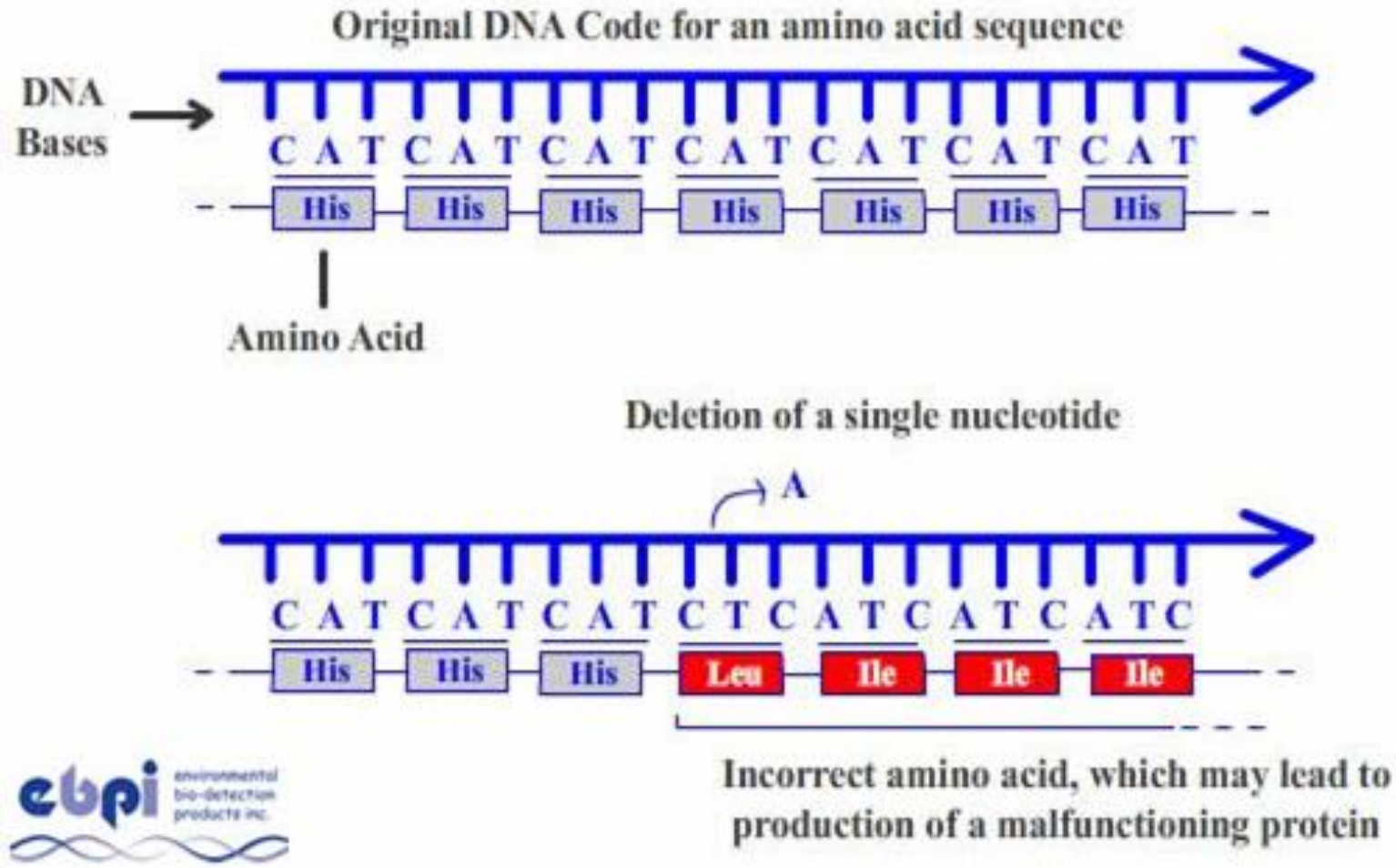
Some missense mutations can be less bad

Both polar



Deletions/insertions

"Frameshifts"



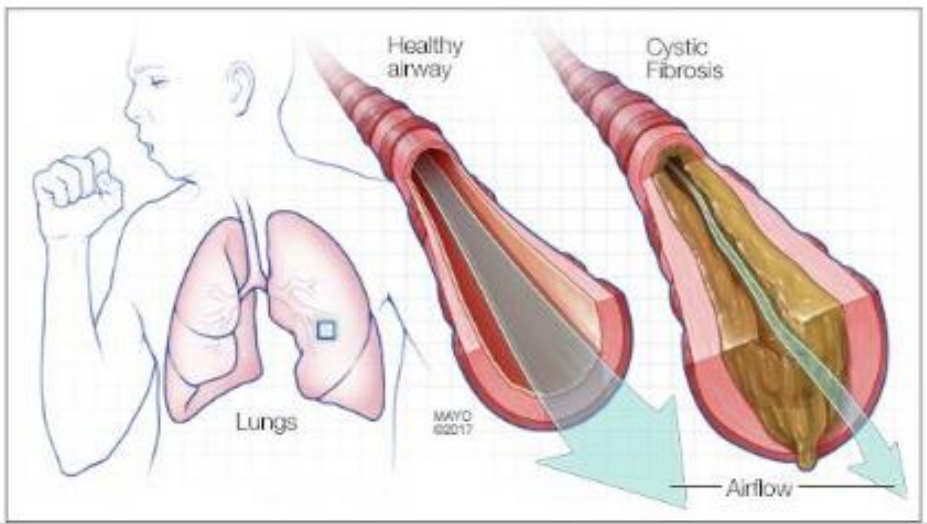
Deletion – cystic fibrosis F508del

Functioning CFTR sequence:

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

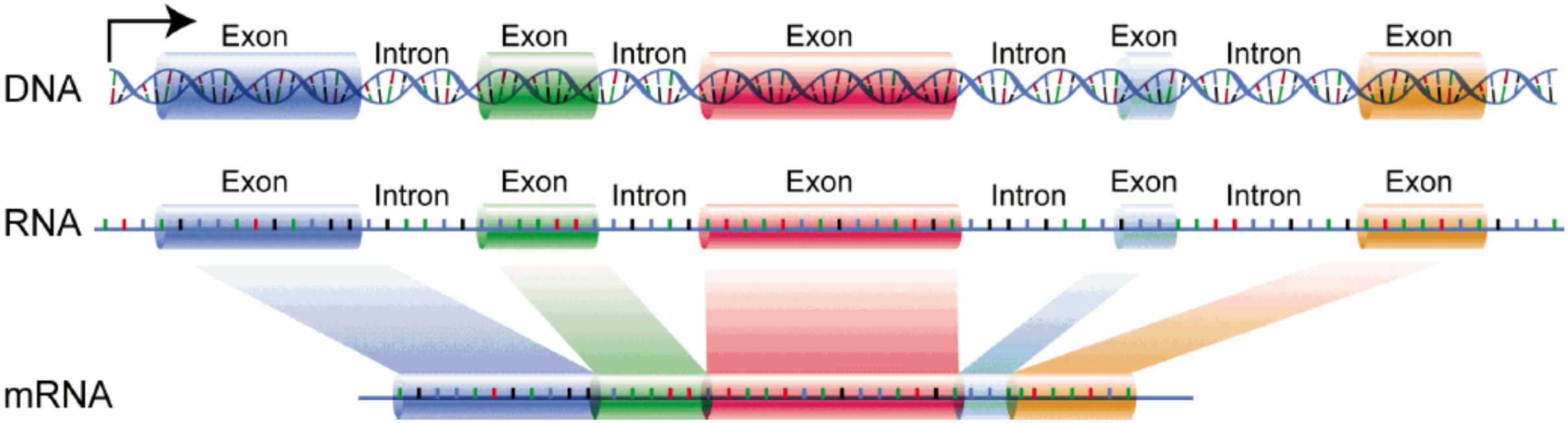
F508Del variant inactivating chloride channel:

Nucleotide	ATC	ATT	GGT	GTT
Amino acid	Ile	Ile	Gly	Val

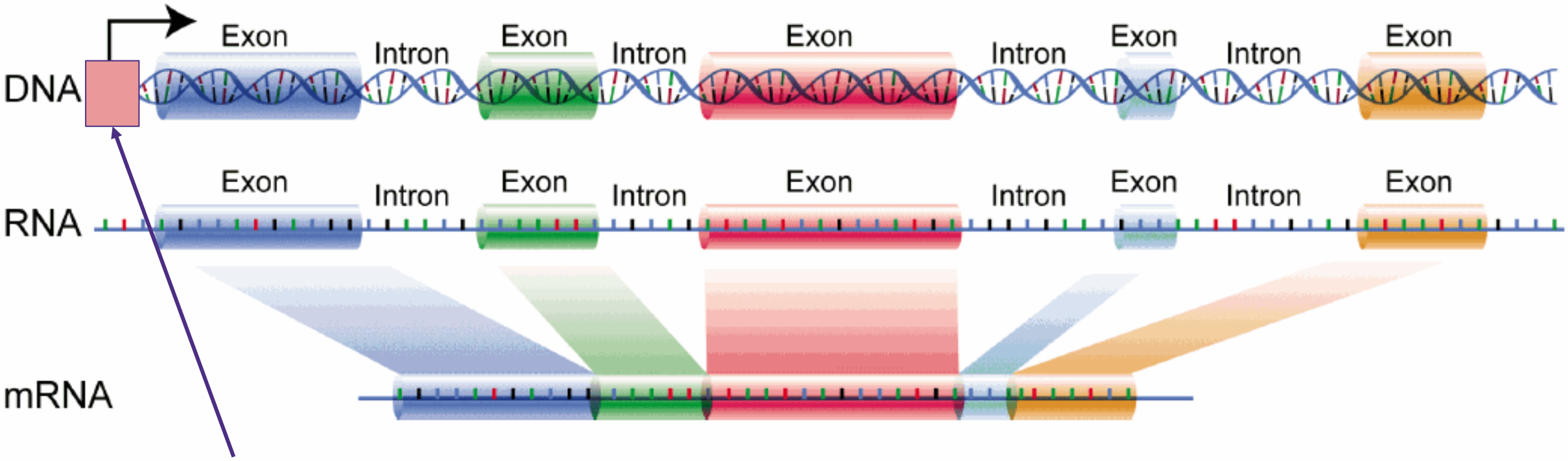


Sometimes we don’t change the protein itself…

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

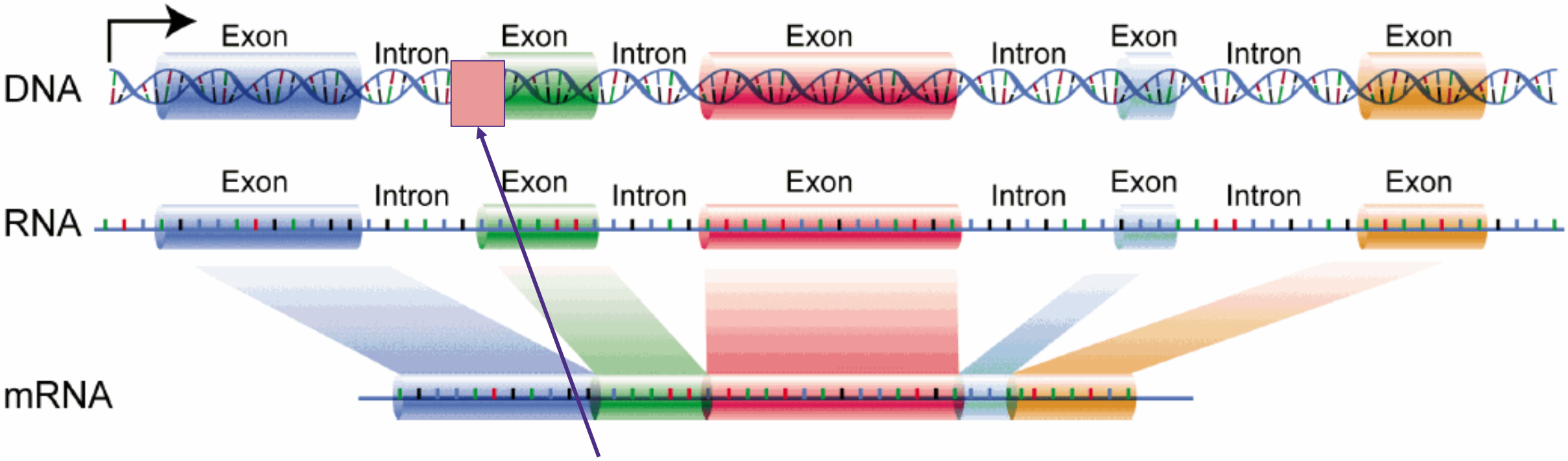


A gene includes a lot of DNA that doesn't become protein



A variant here can change gene "expression"

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



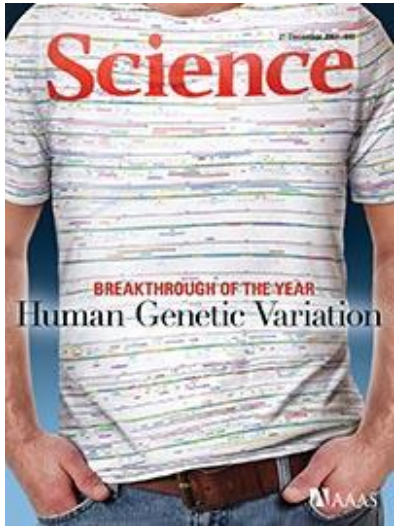
Or here can change the "splice site" to make a different protein

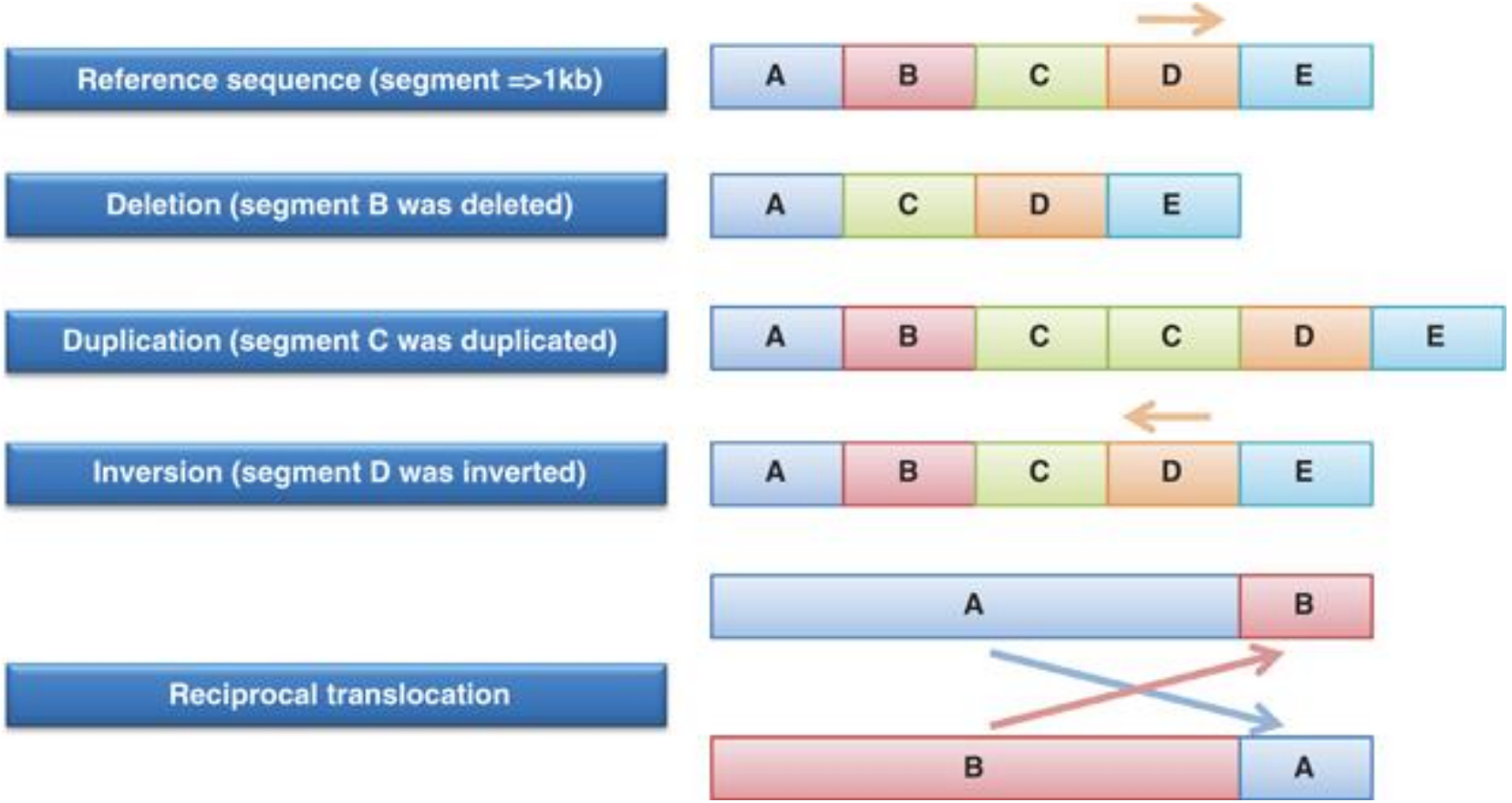
Zoom breakout – discuss Q1

A recent study sequenced the genome of 2,504 individuals and identified 84.7 million SNPs (single nucleotide polymorphisms) between the participants. On average, each individual carried 3.5-4.3 million SNPs each. About 0.5% of those SNPs were in coding regions of genes. Remember, 1.5% of the genome is in a coding region. Why might only 0.5% of variants be in coding regions compared to what would be expected if SNPs were randomly allocated throughout the genome?

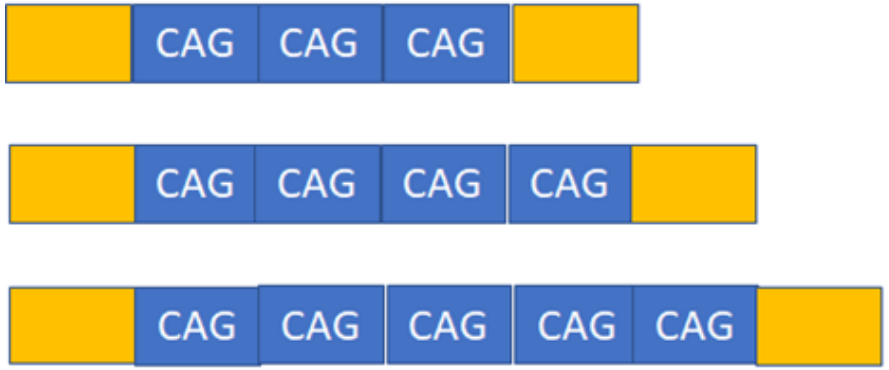
Besides single base changes, what types of changes can we have?

Genetic Variation – structural variation





Tandem Repeats in *HTT* (Huntington's disease)



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

Repeat count	Classification	Disease status
<28	Normal	Unaffected
28–35	Intermediate	Unaffected
36–40	Reduced-penetrance	May be affected
>40	Full-penetrance	Affected

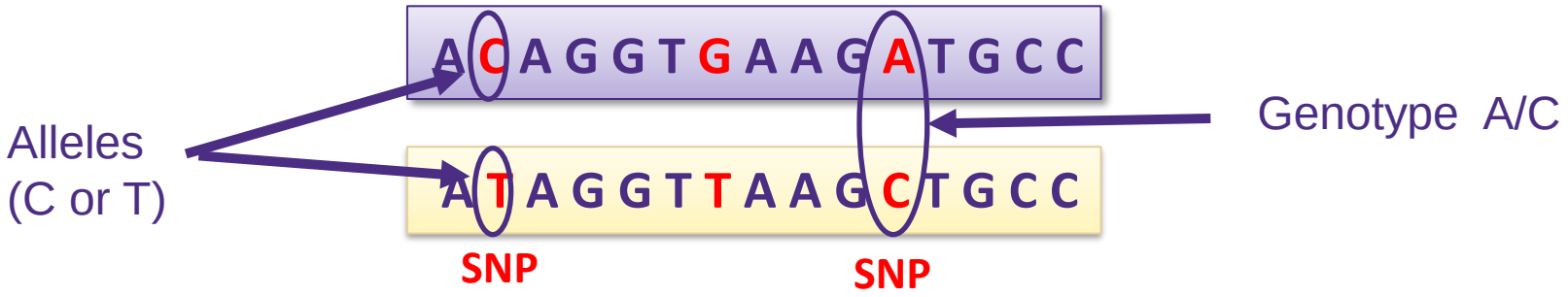
Italics = gene, normal font = protein: *HTT gene* encodes the Htt protein

Alleles to genotypes and phenotypes

Terminology

We inherit two copies of each chromosome

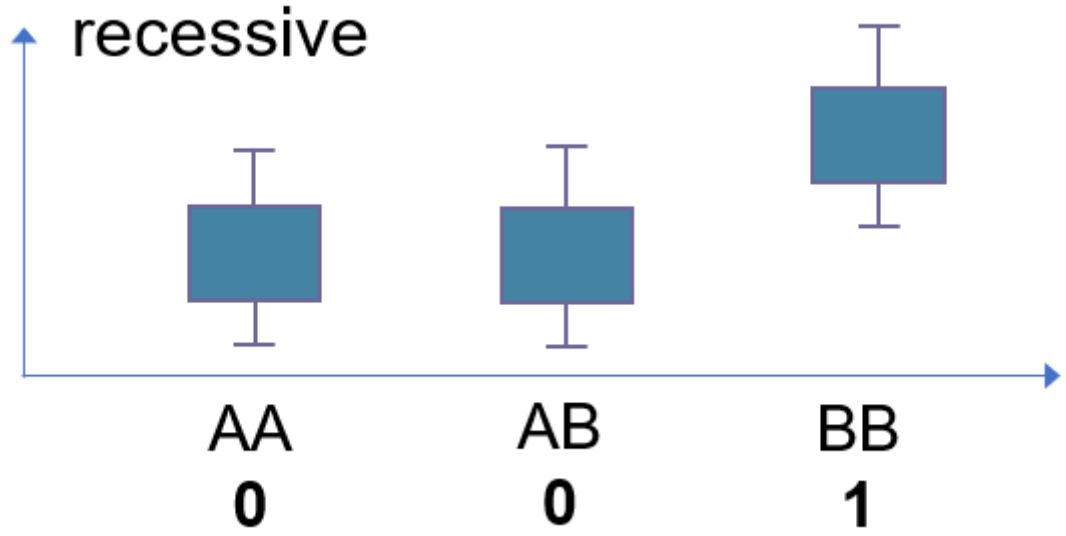
Allele vs. Genotype

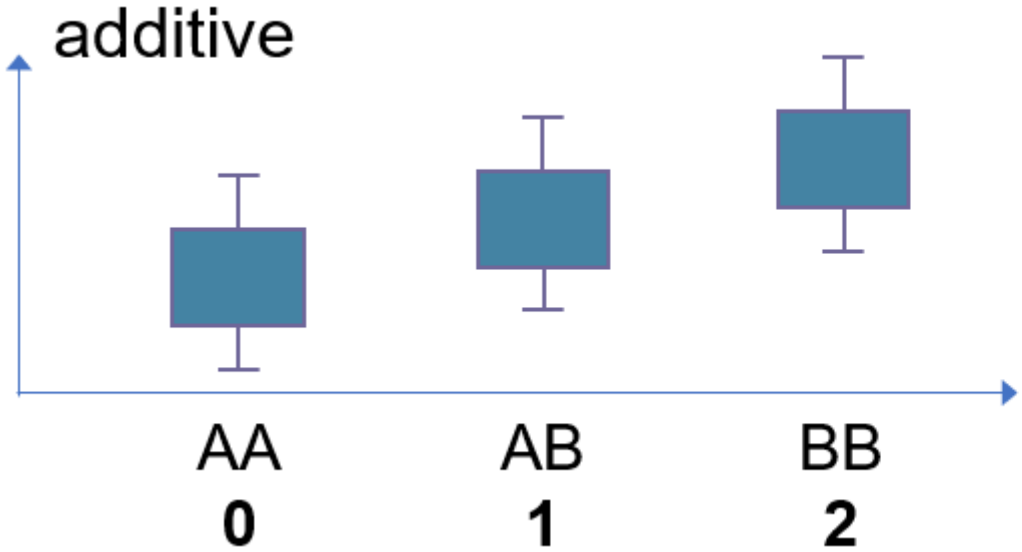


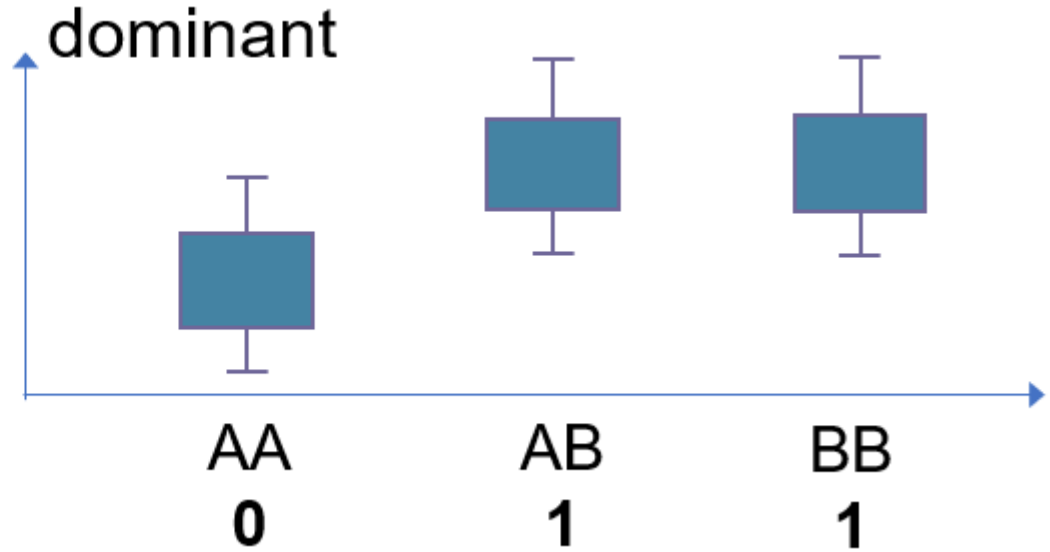
Genotypes
(A/A) – homozygous
(A/C) – heterozygous
(C/C) - homozygous

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)

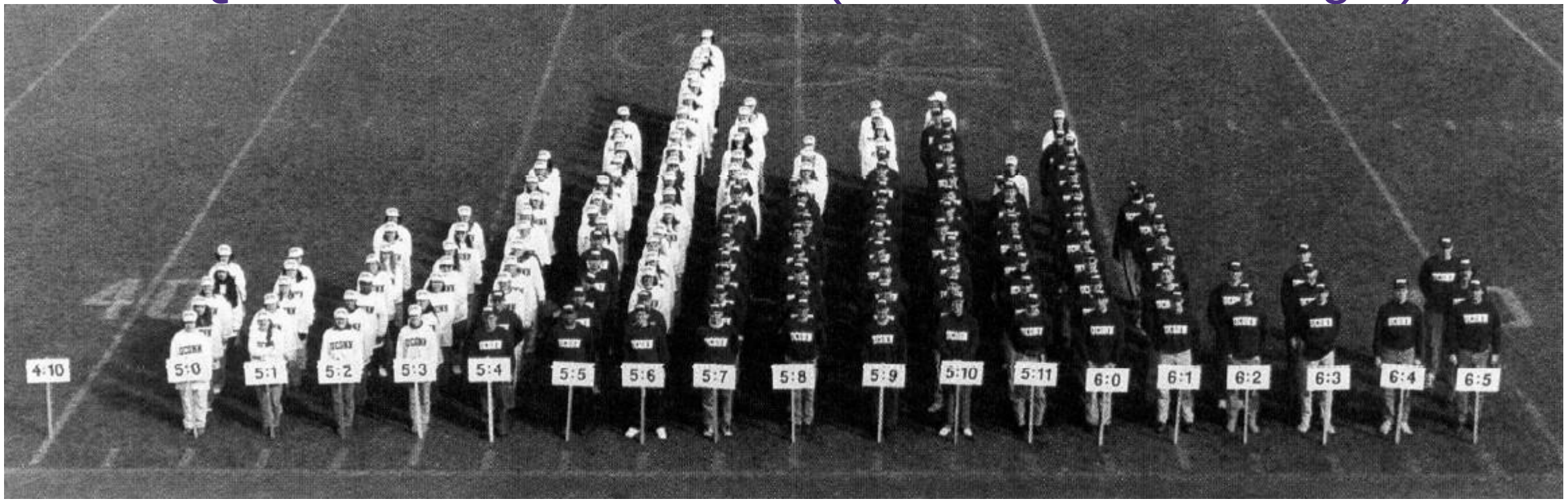

recessive
AA
0
AB
0
BB
1


additive
AA
0
AB
1
BB
2


dominant
AA
0
AB
1
BB
1

Genotypes and Phenotypes

- **Binary outcomes** (yes/no, i.e. disease status)
- **Quantitative outcomes** (continuous, i.e. height)

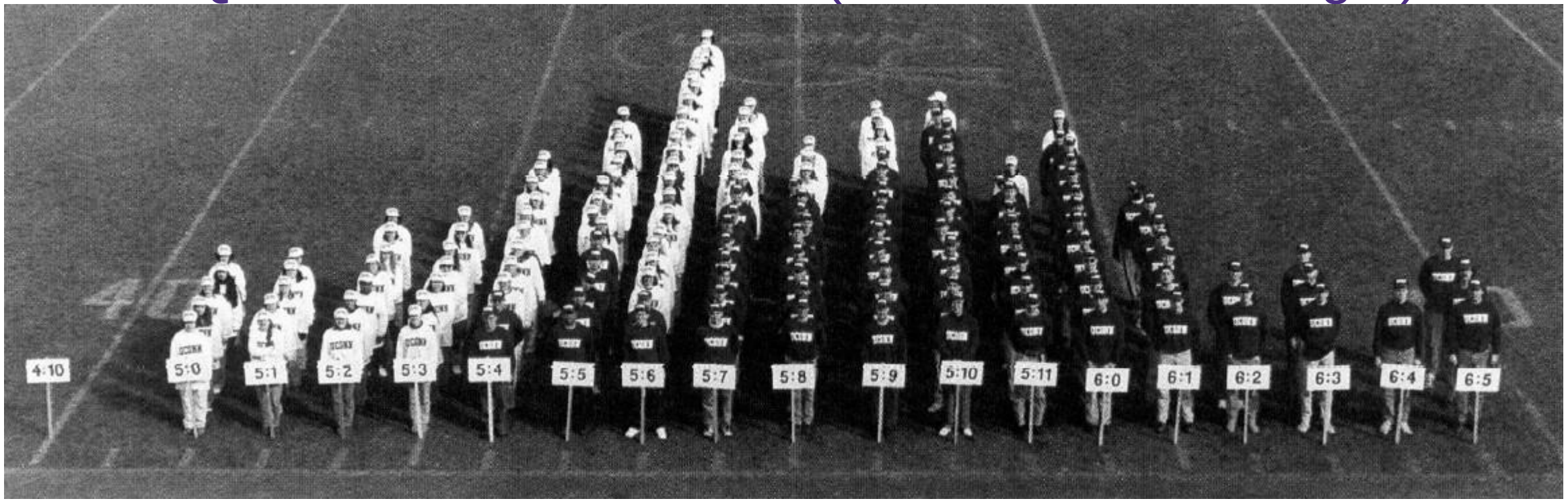


Genotypes and Phenotypes

Zoom chat: How might you turn “height” into a binary variable?

What would be your approach?

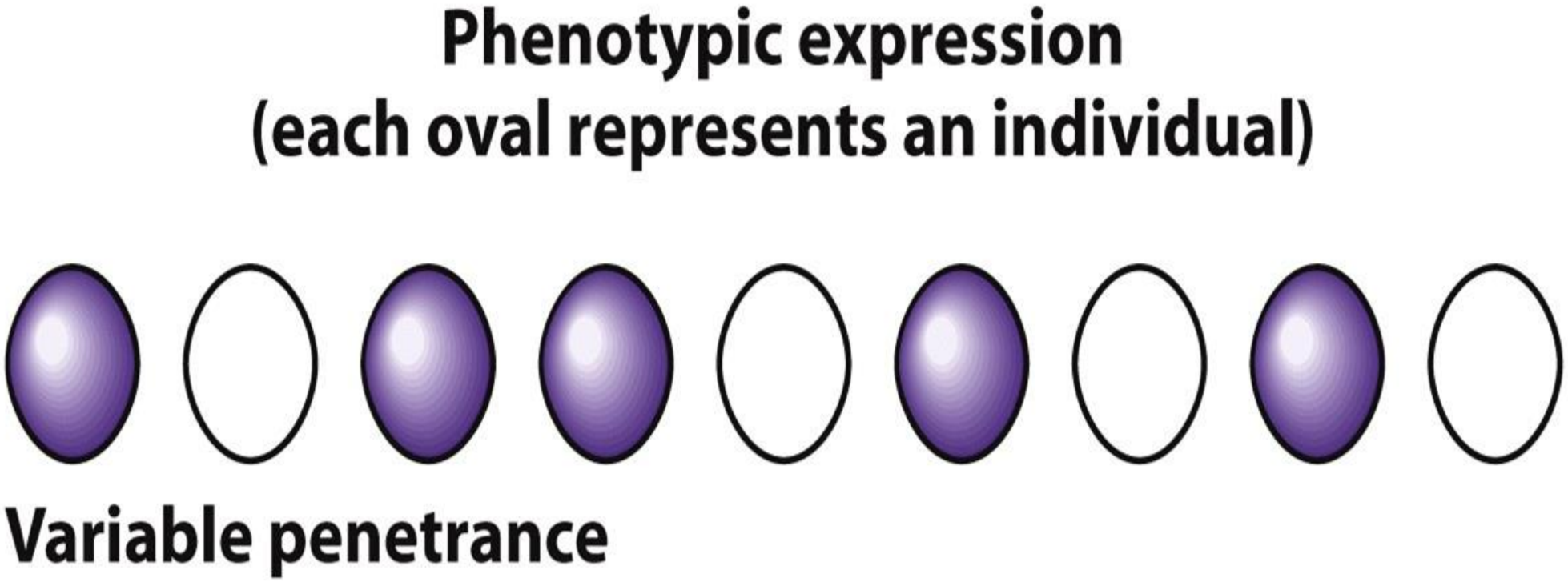
- **Binary outcomes** (yes/no, i.e. disease status)
- **Quantitative outcomes** (continuous, i.e. height)



Genotypes and Phenotypes

- **Mendelian phenotype** is one driven by variation at a single genetic locus.
- **Complex phenotype** does not show such simple patterns of inheritance.
  - oligogenic (a few genetic loci)
  - polygenic (many genetic loci)

Same genetic pattern, different phenotype



Haplotypes

Specific combination of SNPs occurring on the same segment of chromosome. This depends on Linkage Disequilibrium, which we will discuss later

GATATTCGTACGGATT
GATGTTCGTACTGAAT
GATATTCGTACGGATT
GATATTCGTACGGAAT
GATGTTCGTACTGAAT
GATGTTCGTACTGAAT

SNPs A/G

(Single Nucleotide Polymorphisms)

AGT
GTA
AGA

Haplotypes

A set of closely linked genetic markers present on one chromosome which tend to be inherited together

Reference SNP (rs) Report

Download

Switch to classic site

rs776746

Current Build 152
Released October 2, 2018

Organism	*Homo sapiens*
Position	chr7:99672916 (GRCh38.p12)
Alleles	T>C
Variation Type	SNV Single Nucleotide Variation
Frequency	T=0.28922 (36317/125568, TOPMED) T=0.2653 (8204/30920, GnomAD) T=0.379 (1896/5008, 1000G) (+ 3 more)
Clinical Significance	Reported in ClinVar
Gene : Consequence	CYP3A5 : Splice Acceptor Variant ZSCAN25 : Intron Variant
Publications	386 citations
Genomic View	See rs on genome

FEEDBACK

Variant Details

Clinical Significance

Frequency

Genomic Placements

Sequence name	Change
GRCh37.p13 chr 7	NC_000007.13:g.99270539C>T
GRCh38.p12 chr 7	NC_000007.14:g.99672916T>C

Navigate to dbSNP: https://www.ncbi.nlm.nih.gov/snp

- Search: rs4646438

Zoom Breakout

- Activities #2 and #3 (terminology and dbSNP navigation)

Conclusions

- Genetic variation can affect single nucleotides or larger sections of genetic sequence.
- Their effect depends on where they are in the genome.
- Changes occur in each generation and propagate through families.