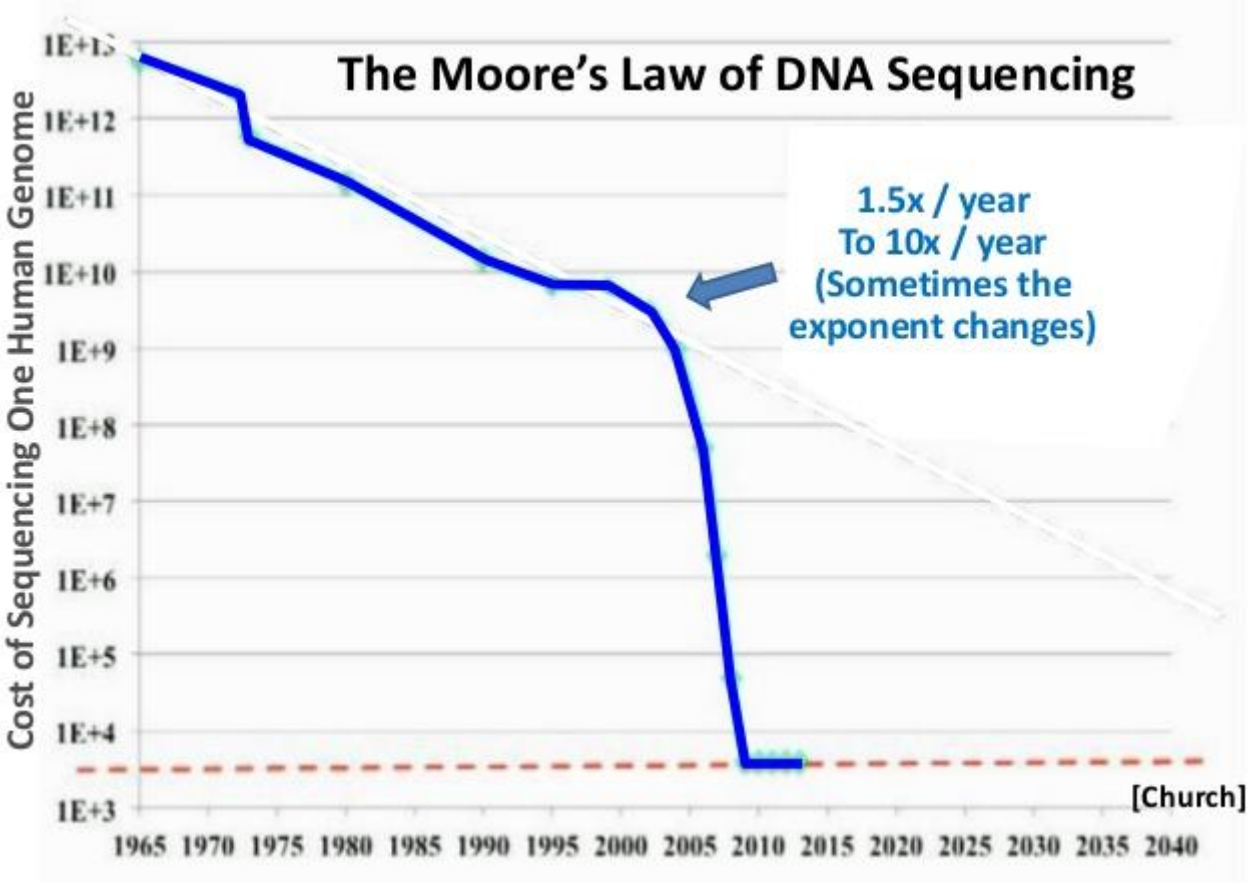

DTU

DTU Health Technology
Bioinformatics

Mutations and Disease

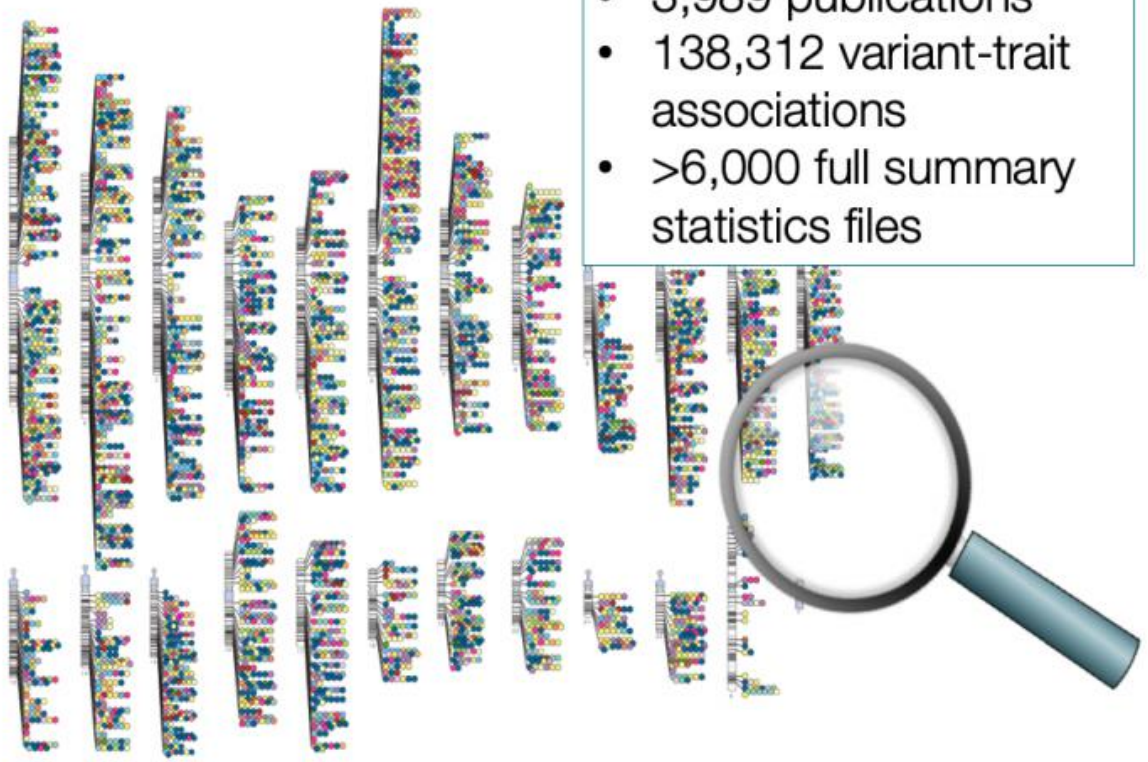
Shyam Gopalakrishnan
Associate Professor
Section of Evolutionary Genomics, KU
Section of Bioinformatics, DTU
shyam.g@gmail.com

Biological relevance of mutations



Biological relevance of mutations

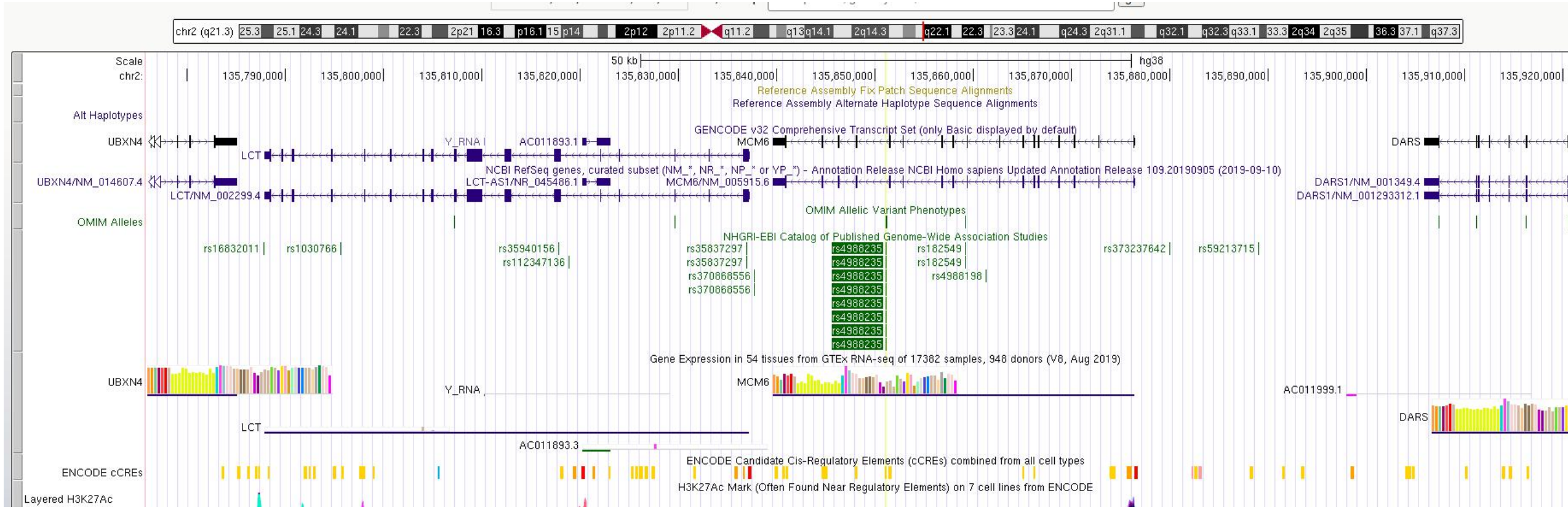
GWAS Catalog



As of May 2019
- 3,989 publications
- 138,312 variant-trait associations
- >6,000 full summary statistics files

Example of mutations affecting phenotype/disease

Example of mutations affecting phenotype/disease



Example of mutations affecting phenotype/disease

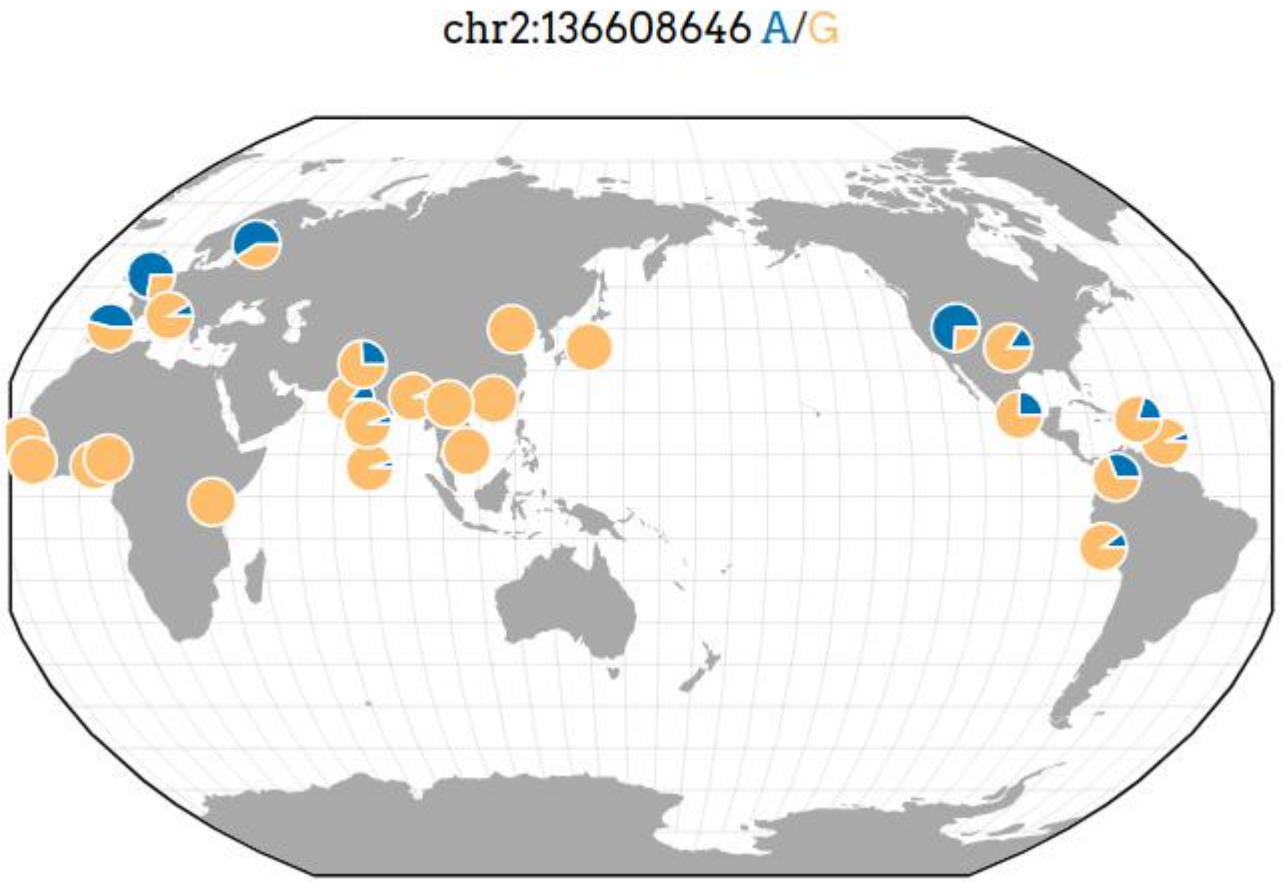
1000genomes (hg19) rs4988235 random

Orientation minus

Stabilized minus

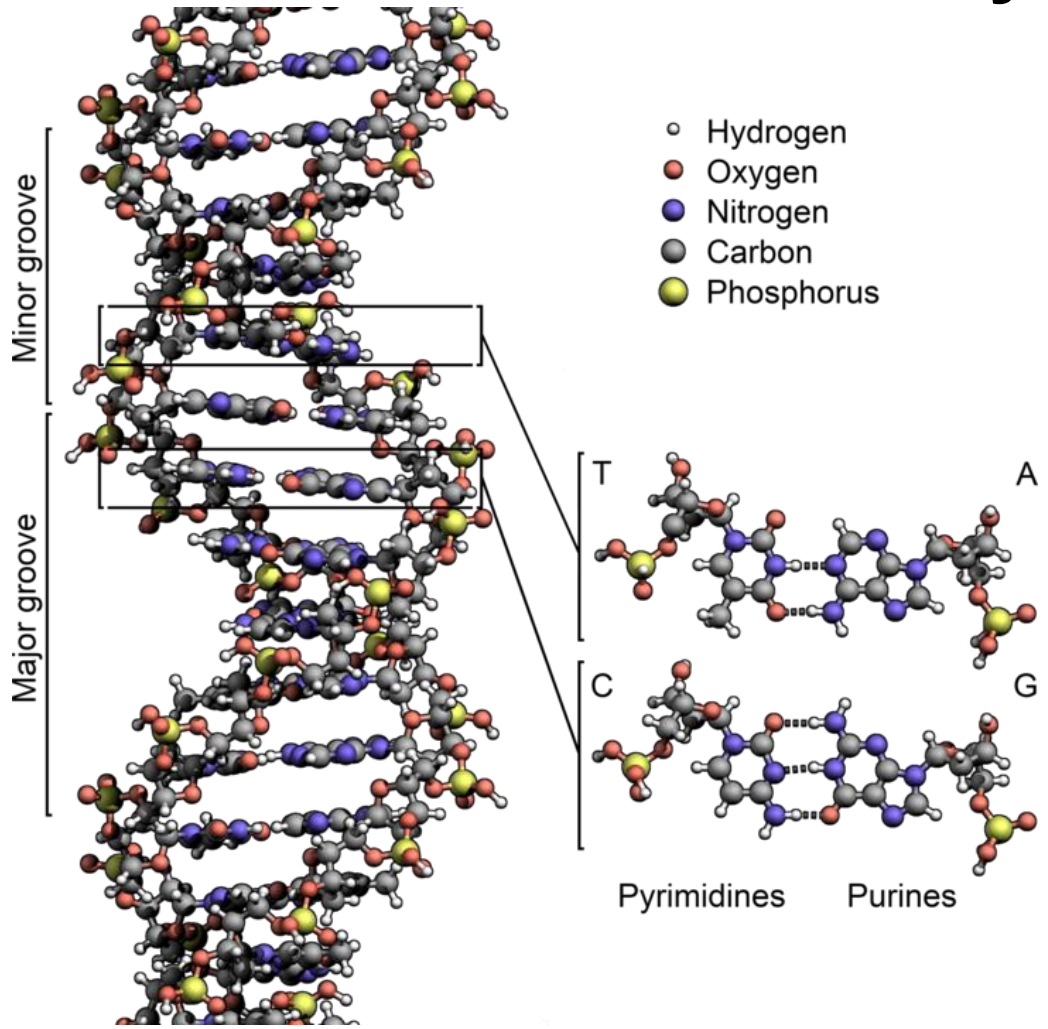
Geno	Mag	Summary
(C;C)	2.5	likely to be lactose intolerant as an adult
(C;T)	1.1	likely to be able to digest milk as an adult
(T;T)	1.1	can digest milk

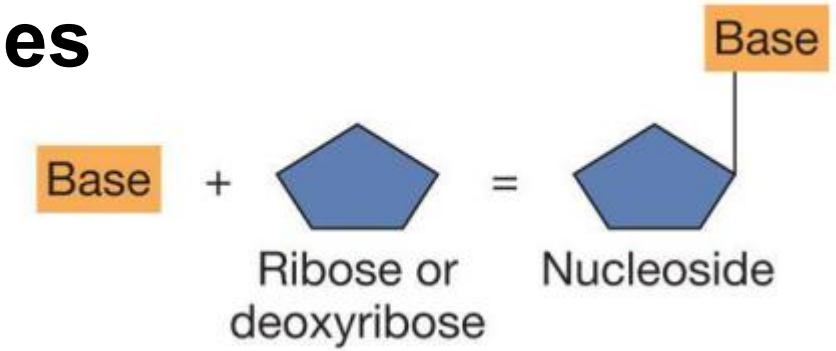
Reference GRCh38 38.1/141

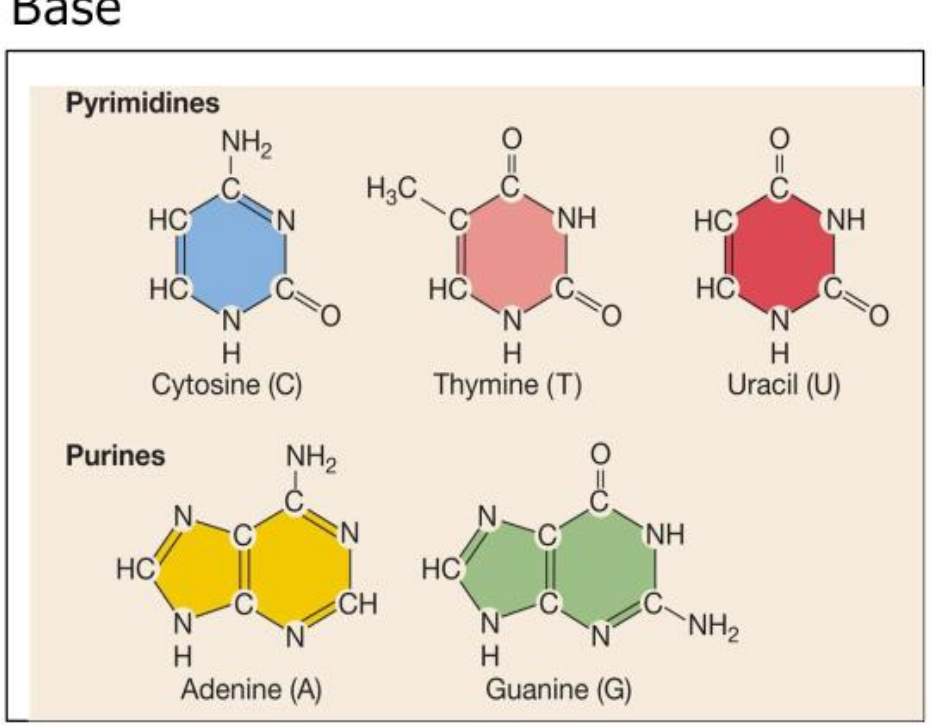


Part 1: What are mutations, and their types?

DNA: Purines and Pyrimidines







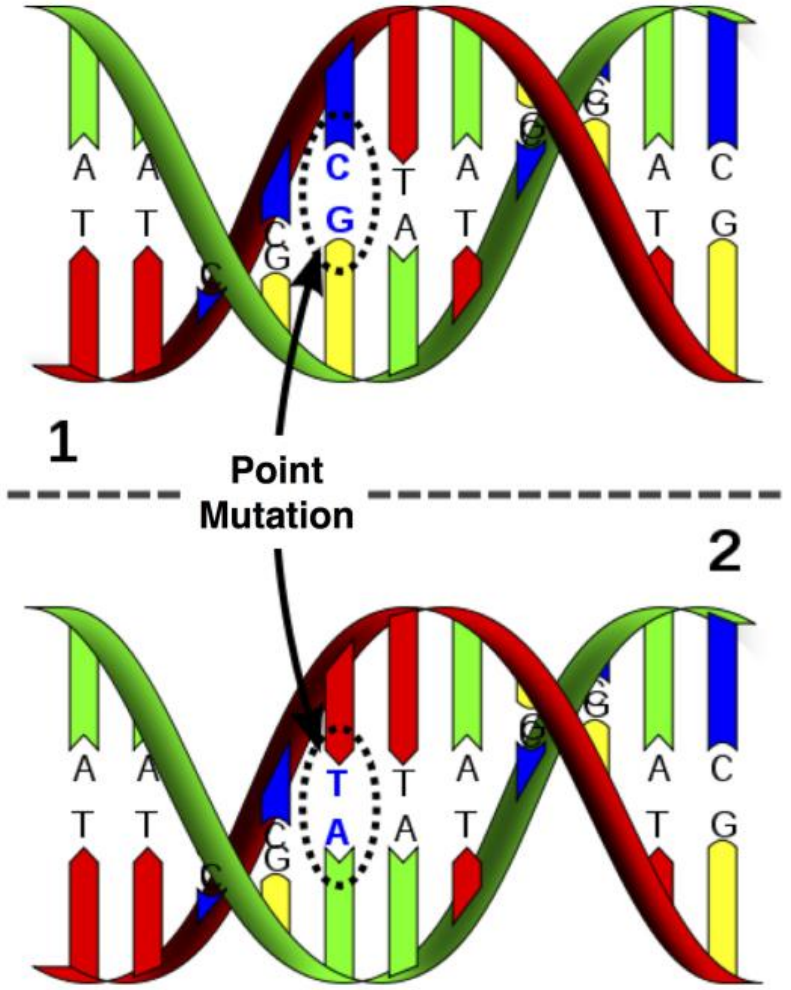
Variations: Mutations and Polymorphisms

- Permanent changes in DNA sequence

Single nucleotide variation (SNV)
- Change in the sequence at a single nucleotide
  - two alleles (states) in the population
- UV, replication error, chemicals ...

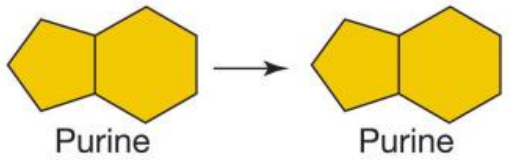
Polymorphism (SNP)
- SNV with a minimum frequency in the population (1% minor allele frequency)

Mutation
- Usually implies association to disease or particular phenotype

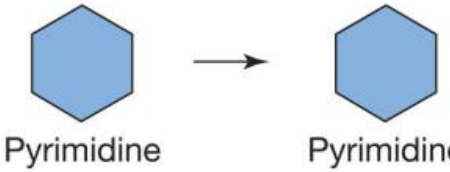


Transitions and Transversions

Transition: Same class of nucleotide base

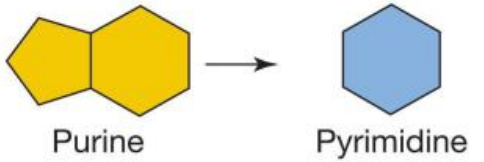


A → G
G → A

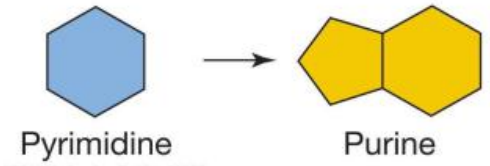


T → C
C → T

Transversion: Different class of nucleotide base

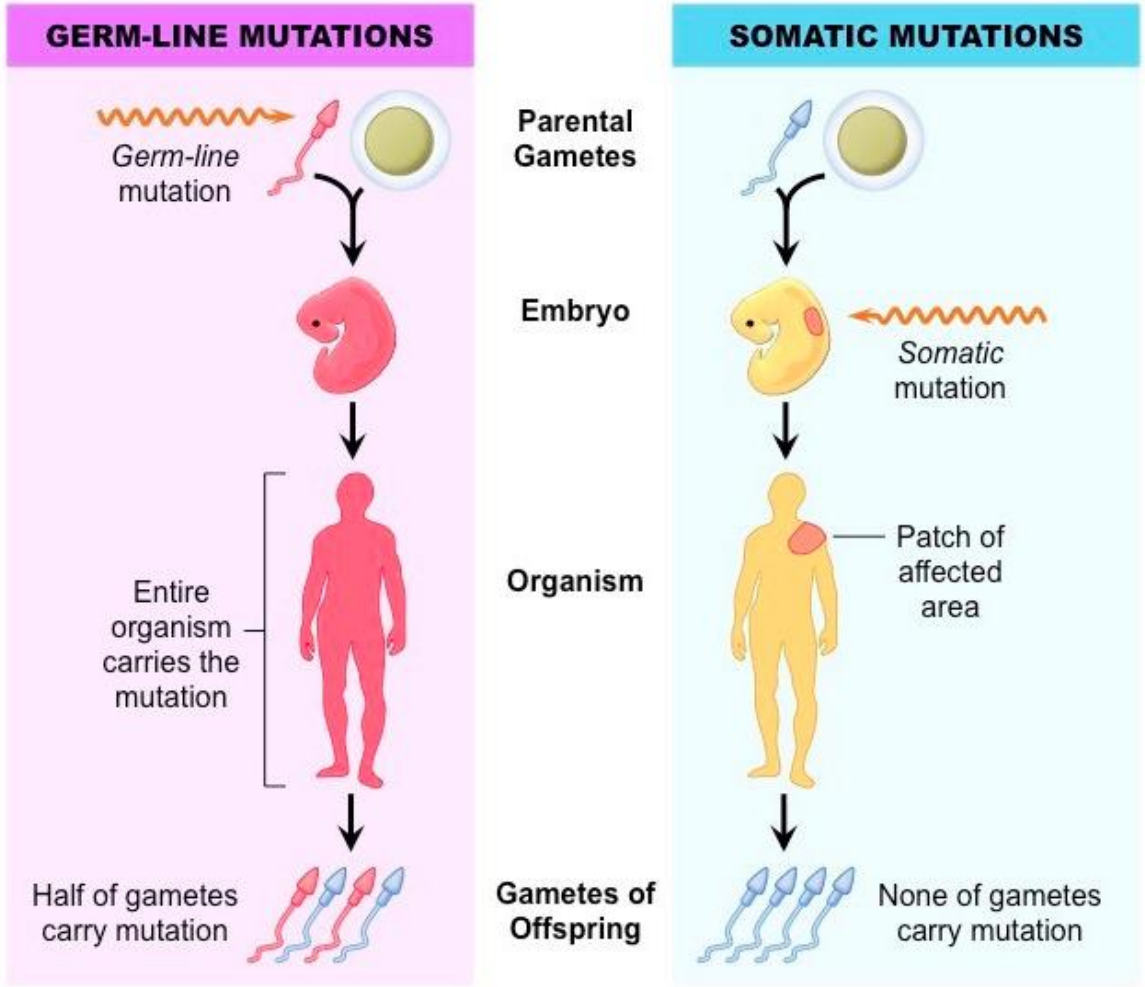


A → C
A → T
G → C
G → T



C → A
C → G
T → A
T → G

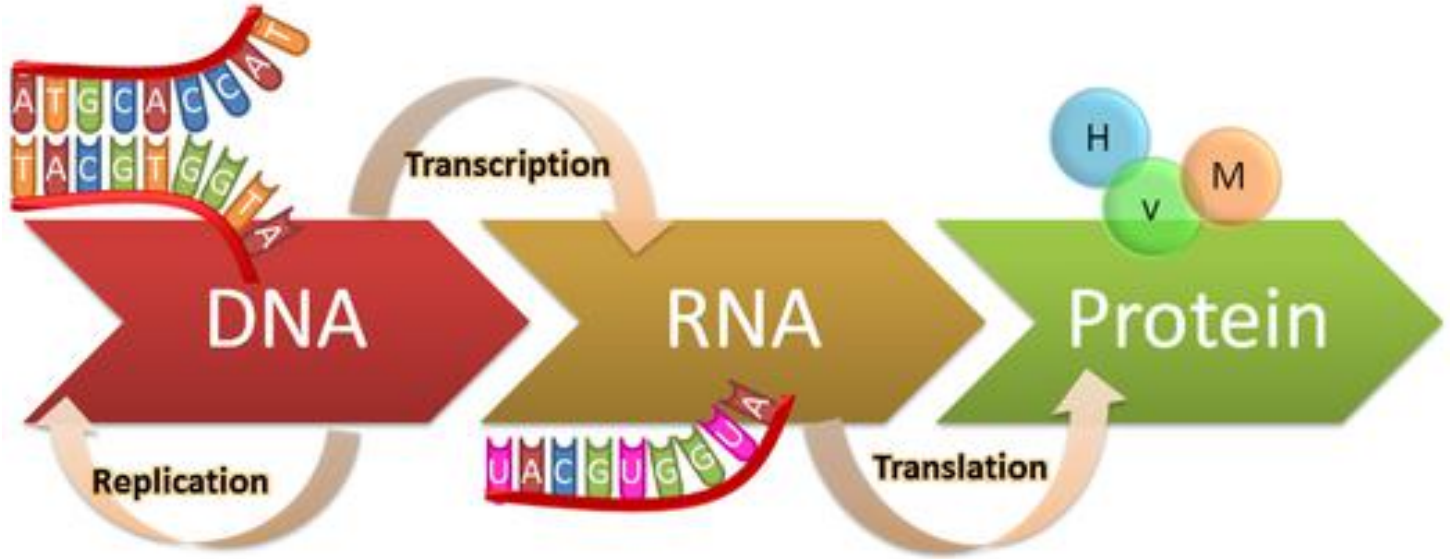
Somatic vs Germline mutations



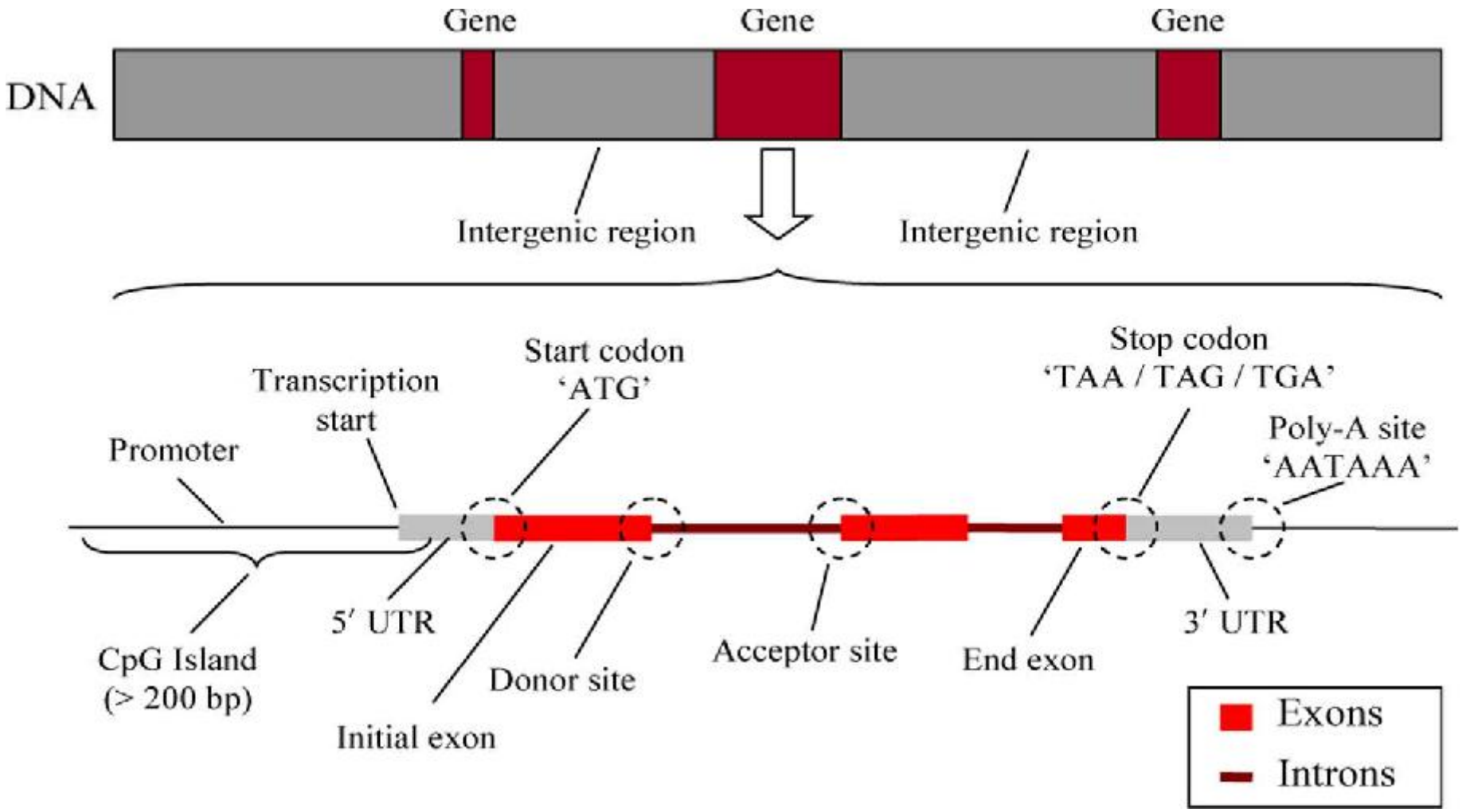
Germline Mutations
Present in **all** cells
Transmitted to offspring
Fixate in population (SNP)

Somatic Mutations
-Present only in **some** cells
-Not transmitted to offspring
-Do **not fixate** in population

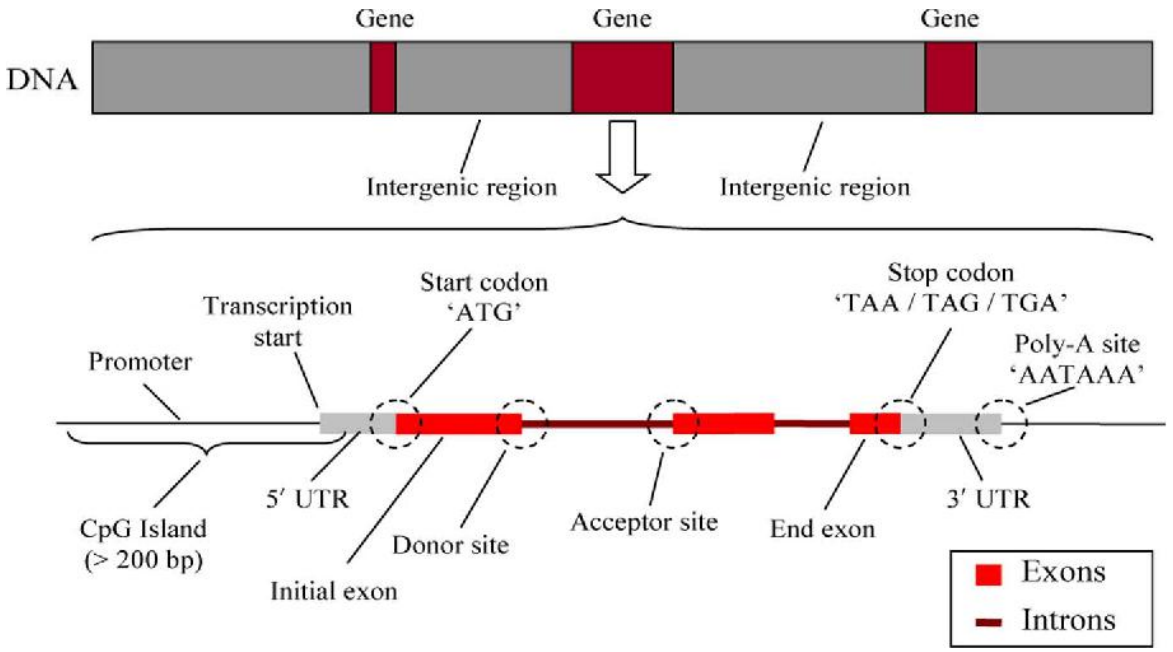
Central dogma of molecular biology



Parts of the genome



Coding vs Non-Coding mutation



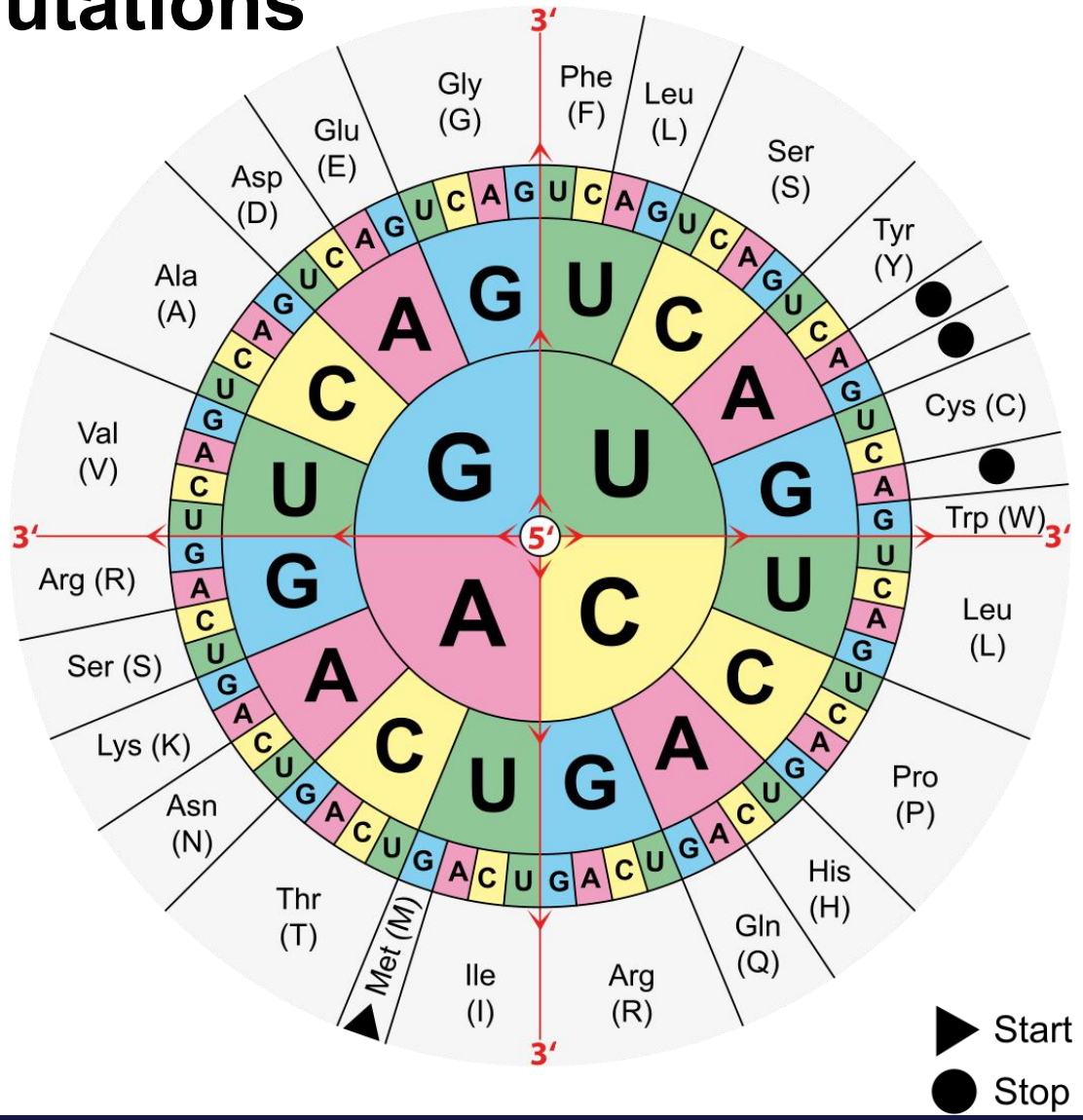
Coding mutation: Located in regions of the genome that code for proteins [Exons]

Non-coding mutation: Located in regions of the genome that do not result in proteins [Everything but the exons]

Consequences of coding mutations

1. Synonymous (slient) mutation

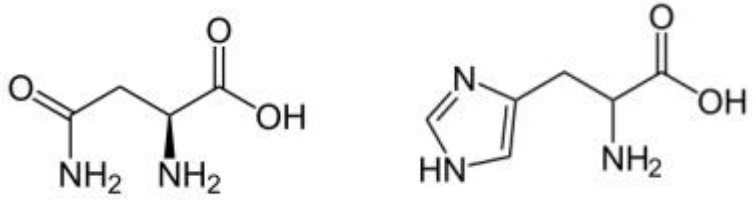
   - no change in the amino acid sequence
     ... CCA ... → ... CCC ...
     Proline → Proline

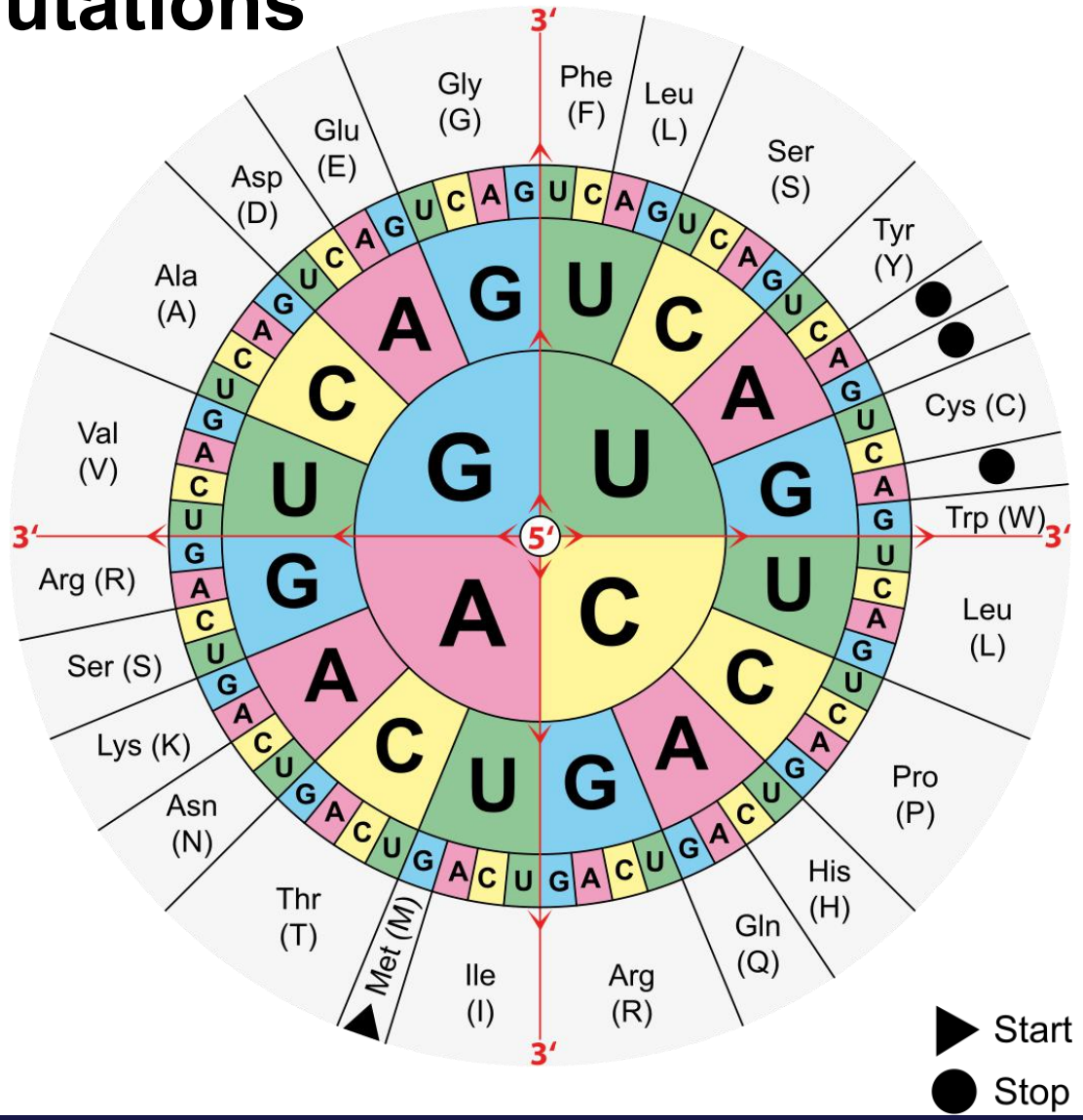


Consequences of coding mutations

2. Non-synonymous (mis-sense) mutation

- **change** in the amino acid sequence
  ... AAC ... → ... CAC ...
  Aspargine → Histidine





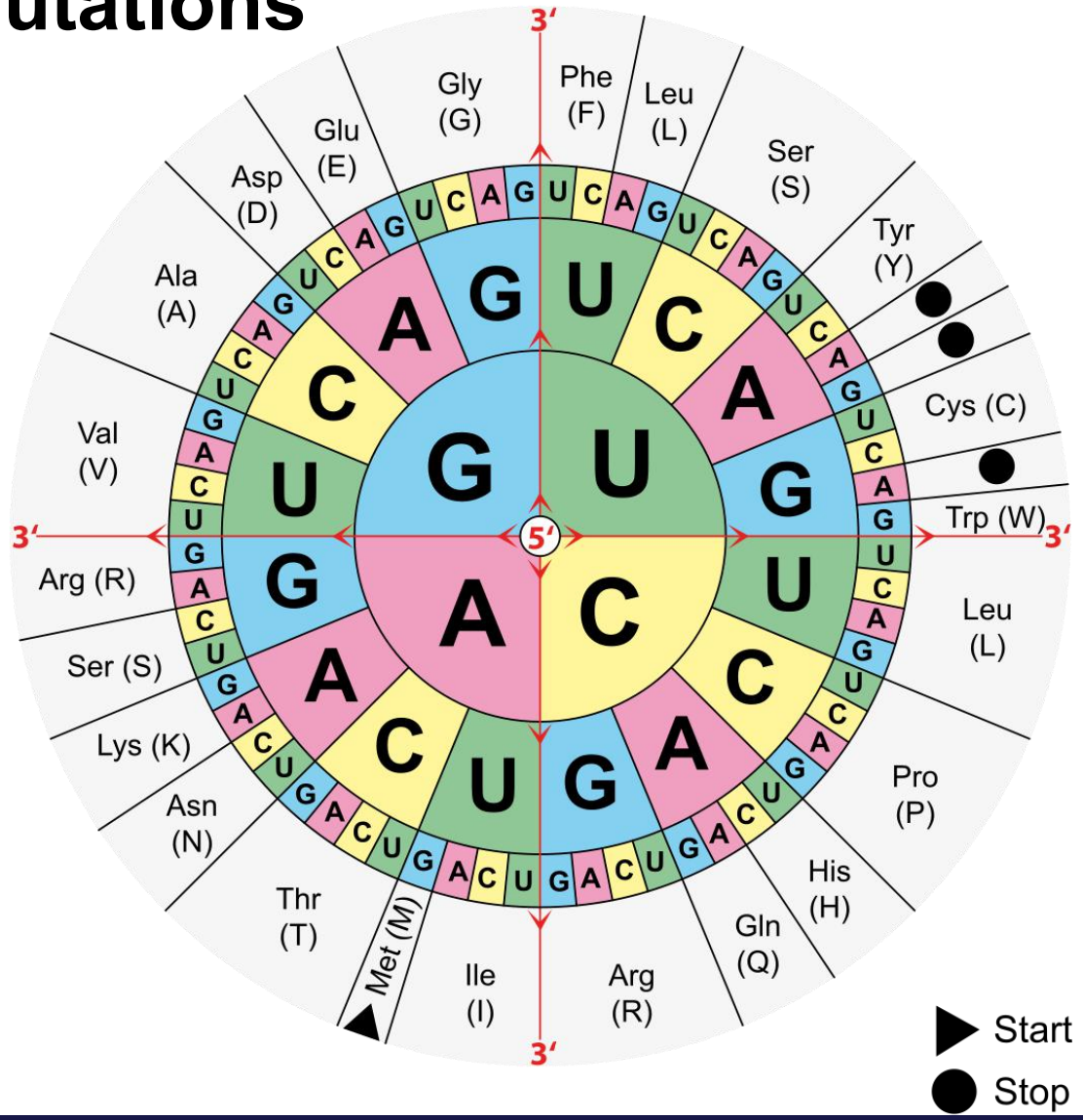
Consequences of coding mutations

3. Non-sense mutation (early stop codon):

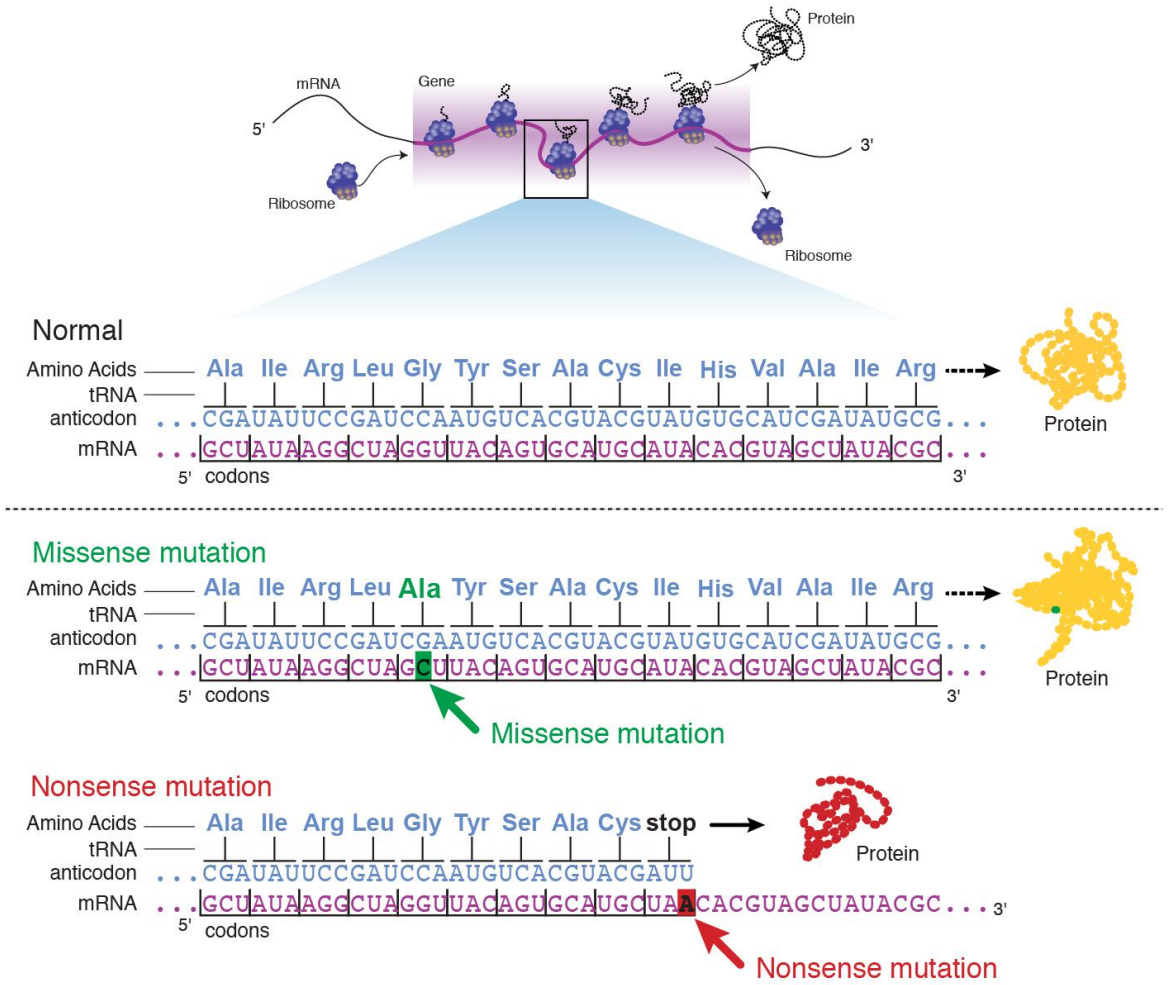
- Early termination of a protein: replace an amino acid with a stop codon

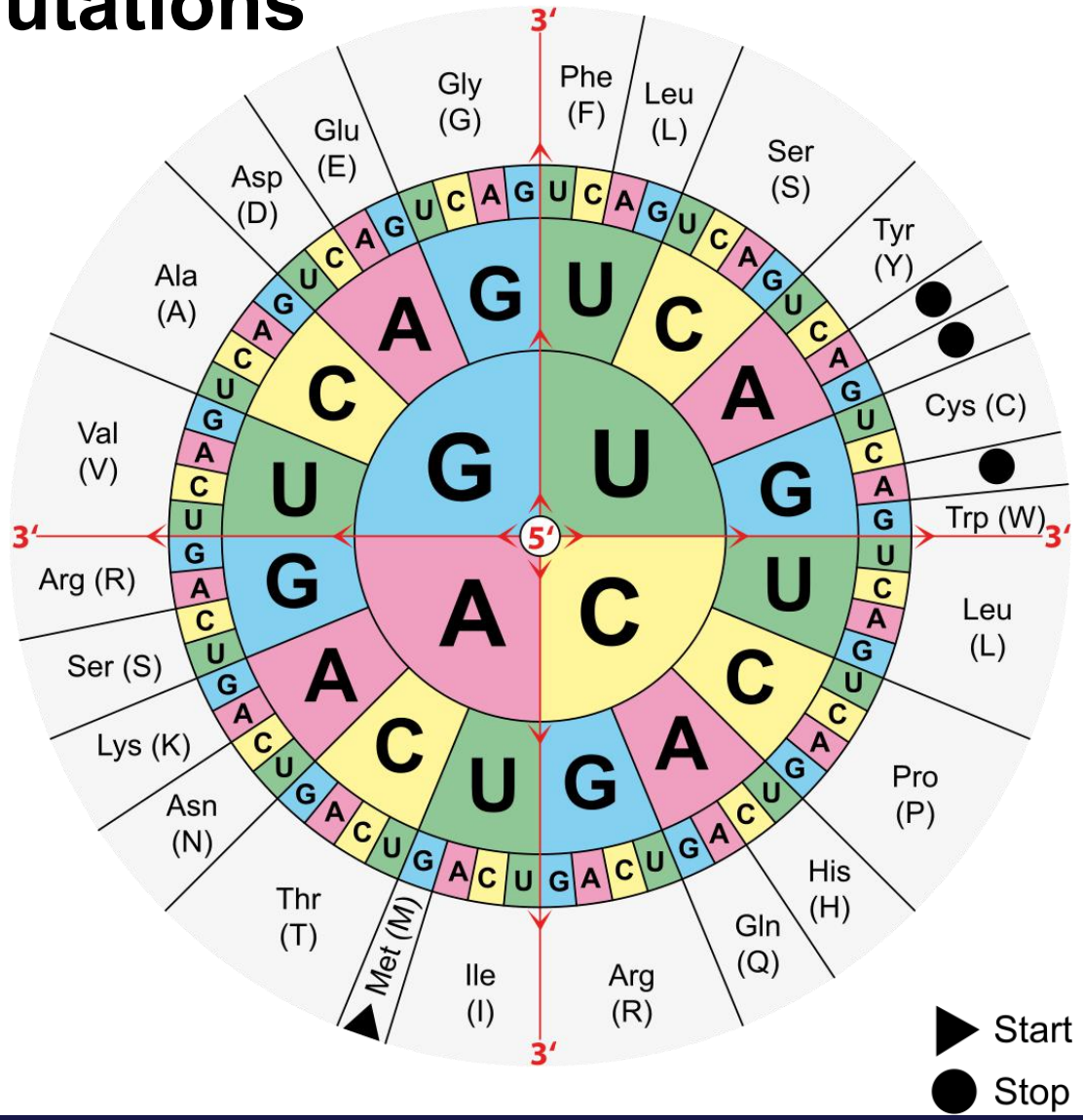
... UGU ... → ... UGA ...

Cystine → opal stop codon



Consequences of coding mutations

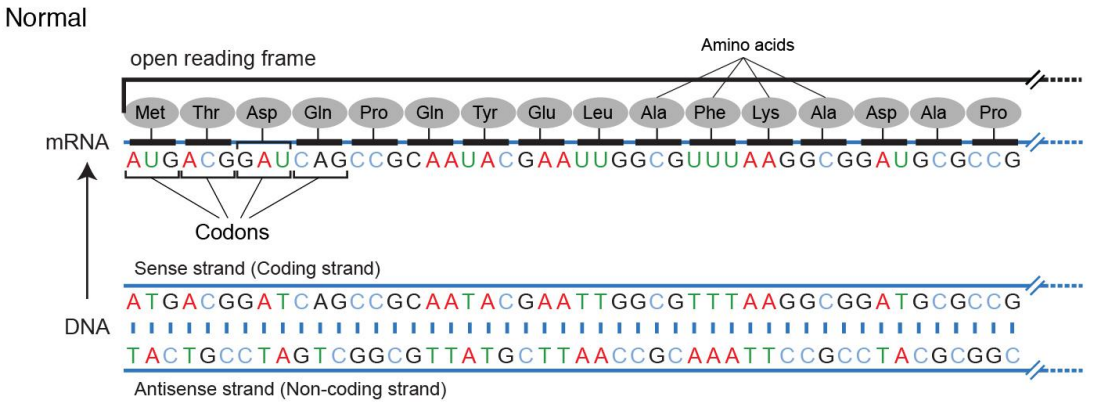


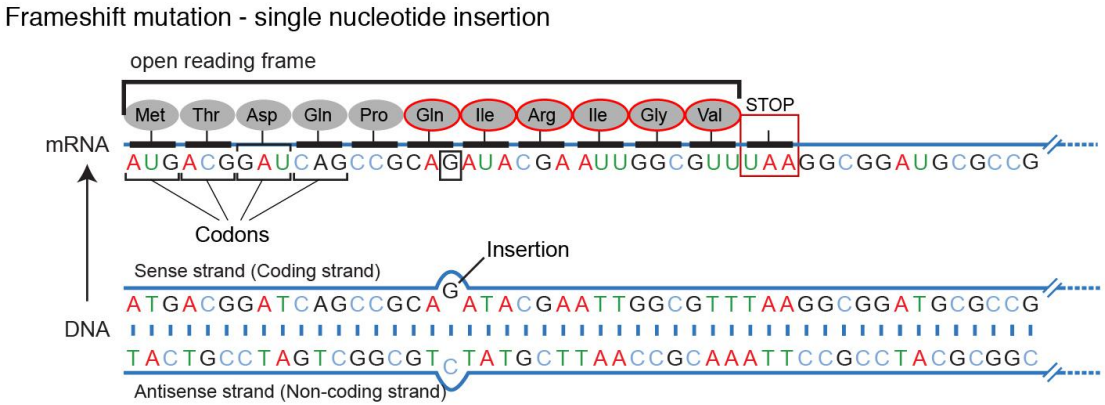


Consequences of coding mutations

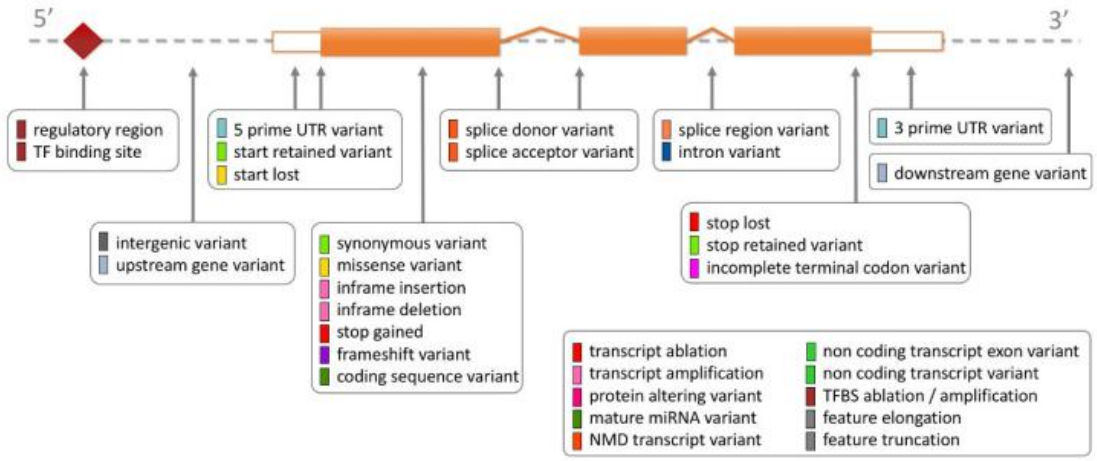
4. Frame shift mutation:

- Alter the open reading frame - change the grouping of the 3 bases that comprise a codon
- Cause by an small insertion/deletion
- Results in a different protein





What about the non-coding mutations?



* SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice region variant	LOW
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Incomplete terminal codon variant	LOW
start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains	SO:0002019	Start retained variant	LOW

More complex variations

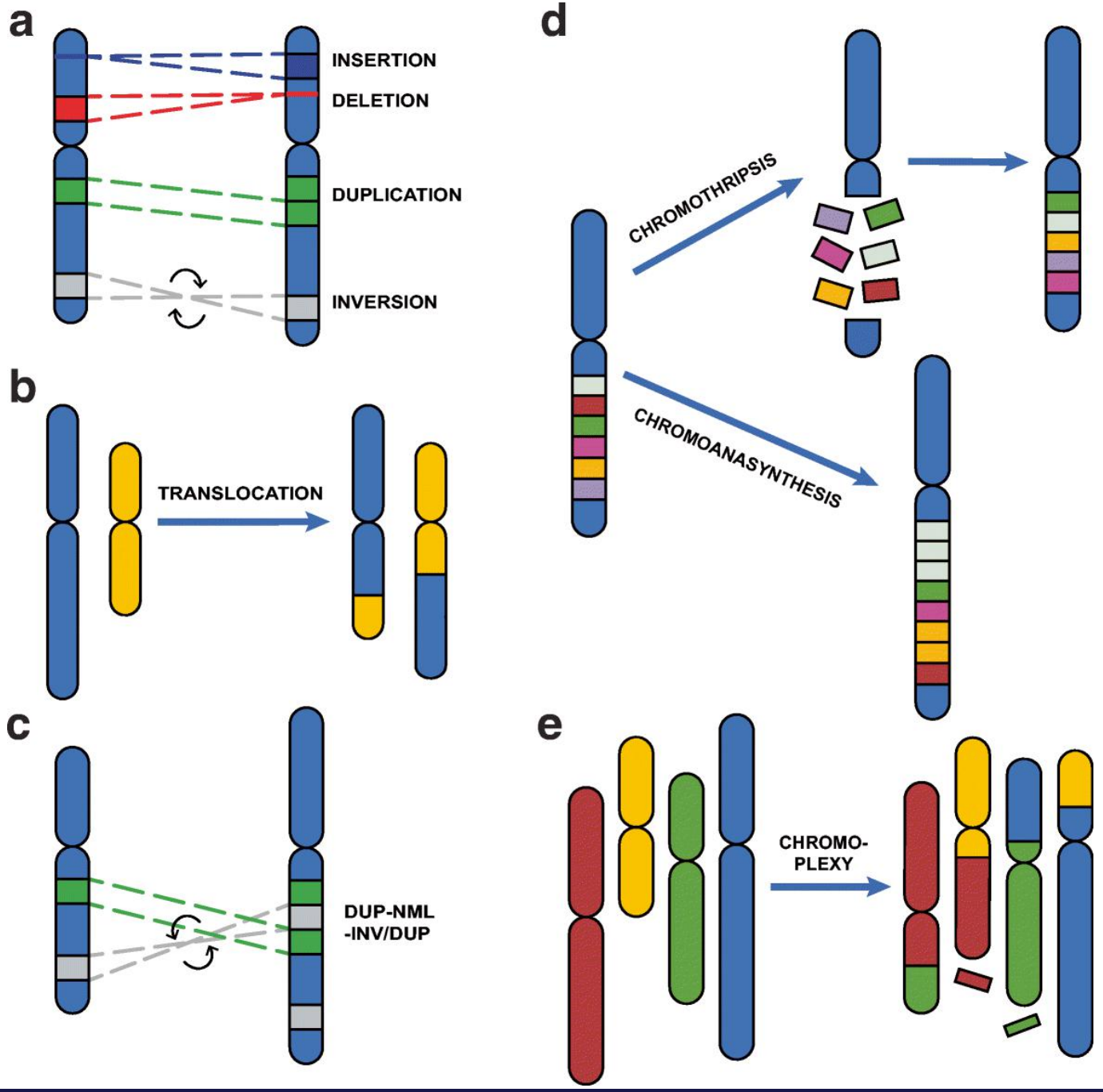
Insertion

Deletion

Inversion

Duplication

Translocation

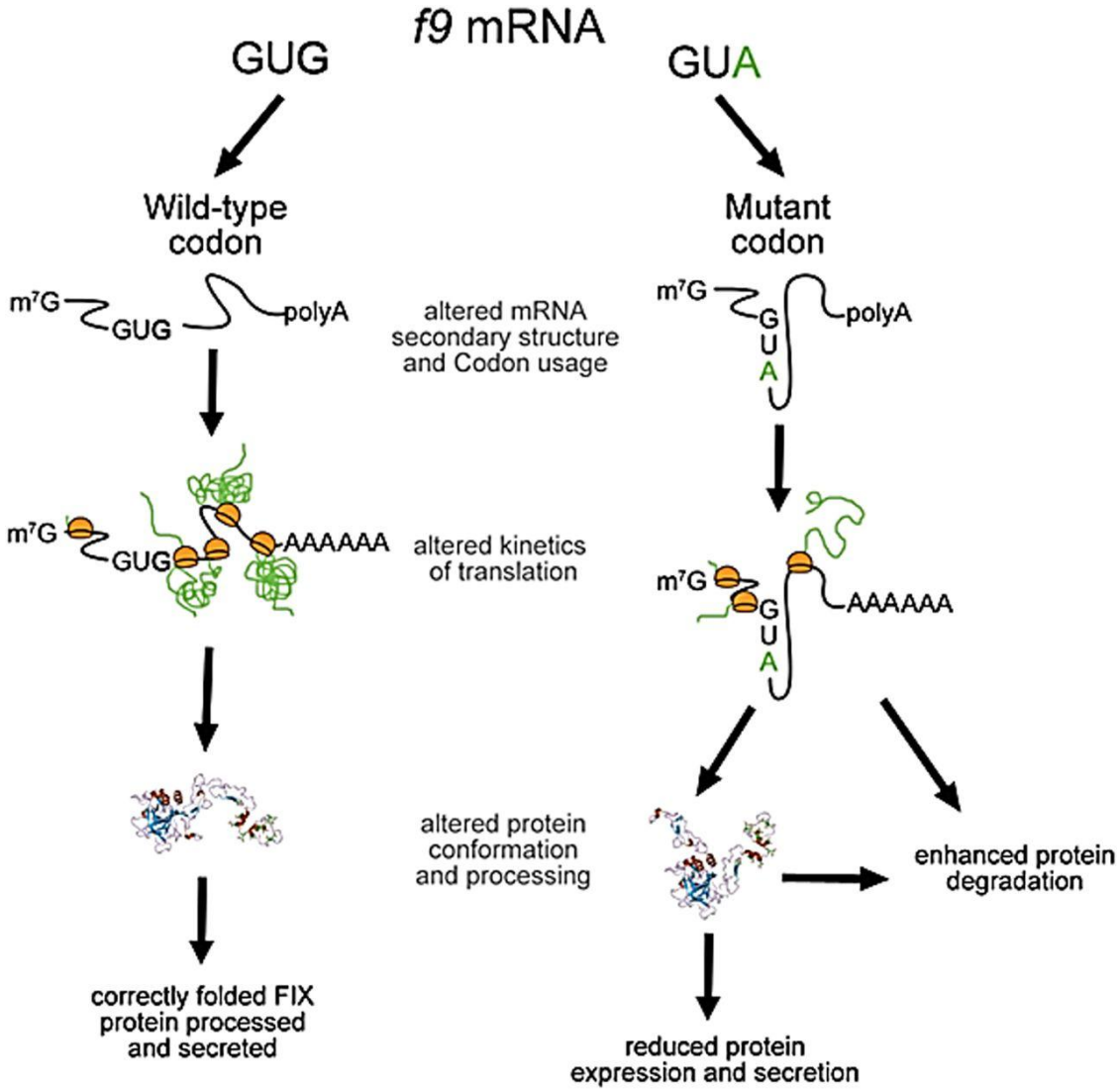


Part 2: Mutations and diseases

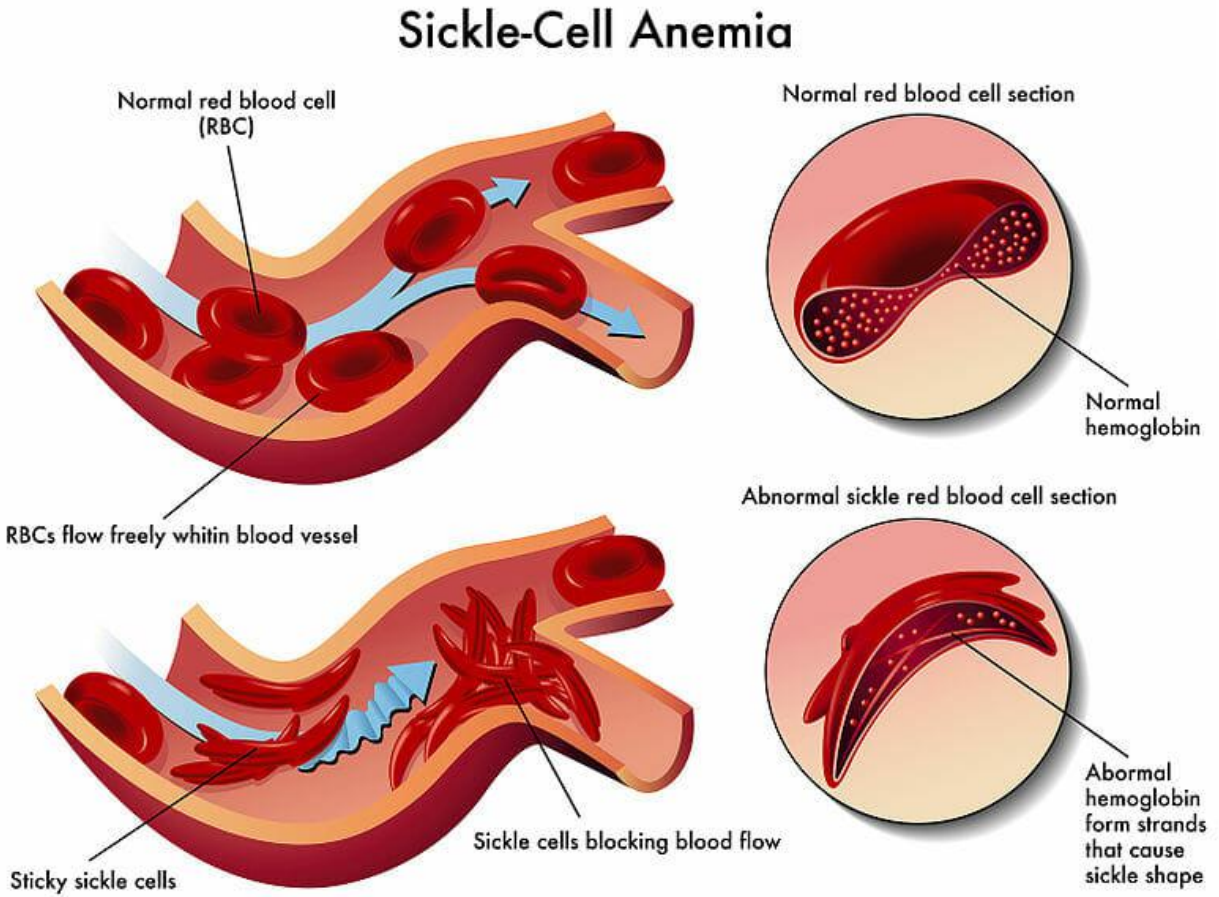
Mutations affect protein structure

Non-synonymous mutations

Non-sense mutations



Mutations affect protein structure



RESEARCH Open Access

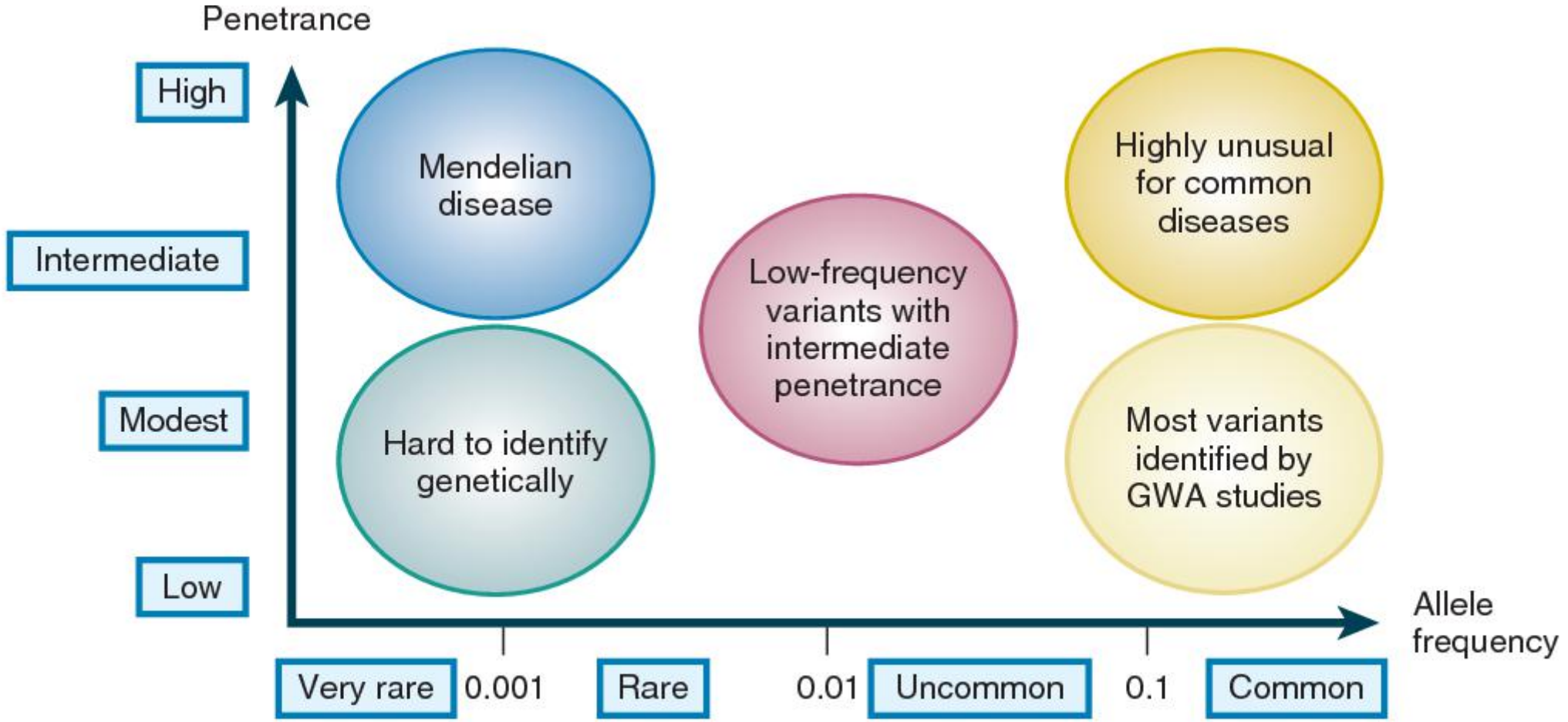
A novel lipoprotein lipase gene missense mutation in Chinese patients with severe hypertriglyceridemia and pancreatitis

Tan-Zhou Chen[1†], Sai-Li Xie[1†], Rong Jin[2] and Zhi-Ming Huang[1*]

Novel missense mutations in exon 15 of desmoglein-2: Role of the intracellular cadherin segment in arrhythmogenic right ventricular cardiomyopathy?

Katja Gehmlich, PhD,* Angeliki Asimaki, PhD,† Thomas J. Cahill, MA, MRCP,* Elisabeth Ehler, PhD,‡ Petros Syrris, PhD,* Elisabetta Zachara, MD,§ Federica Re, MD,§ Andrea Avella, MD,‖ Lorenzo Monserrat, MD,¶ Jeffrey E. Saffitz, MD, PhD,‖ William J. McKenna, MD, FRCP*

Mutations and disease



etic architecture of common and mendelian diseases. At one end of the spectrum are mendelian diseases cause by few variants in few

Understanding mutations using genomics

Traditionally

1 Mutation
=
1 Disease

Phenotype
Function
Mechanism

Lots of hard work

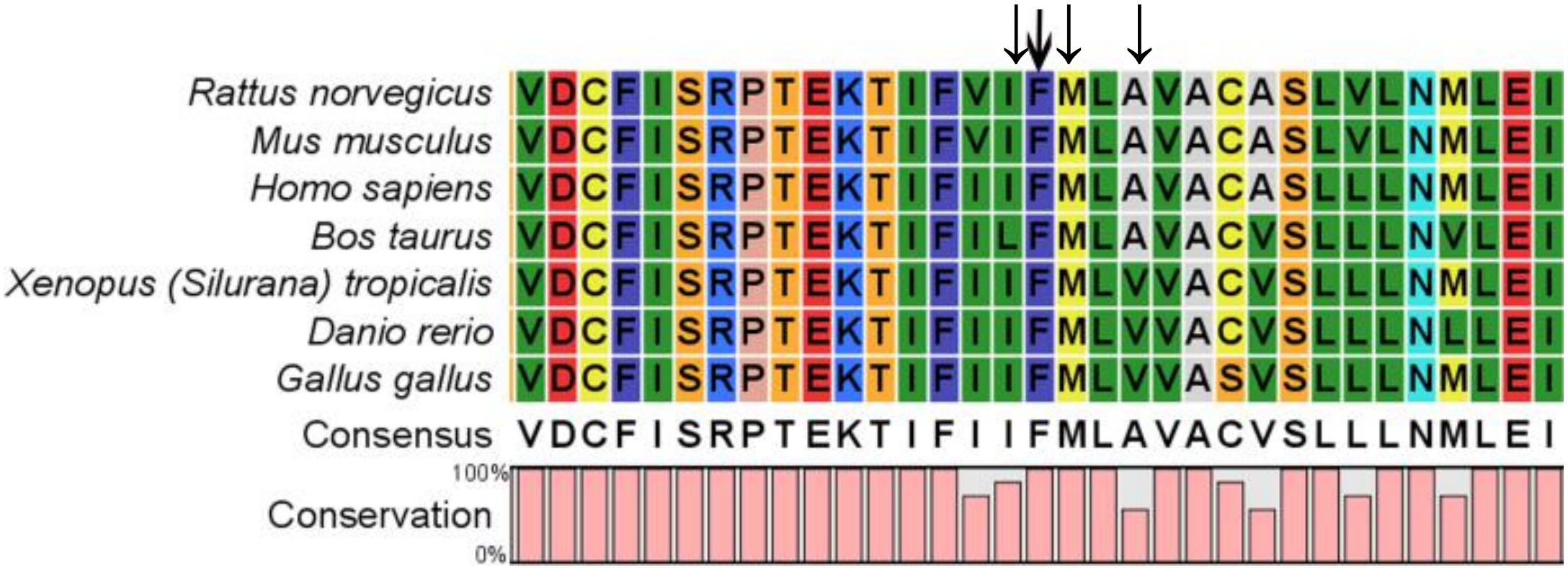
Now (Bioinformatics)

X Mutations
In
Y Patients
And
Z Conditions

Prediction of
Pathogenicity /
Prioritization

Unfeasible

Pathogenicity of mutations: Conservation



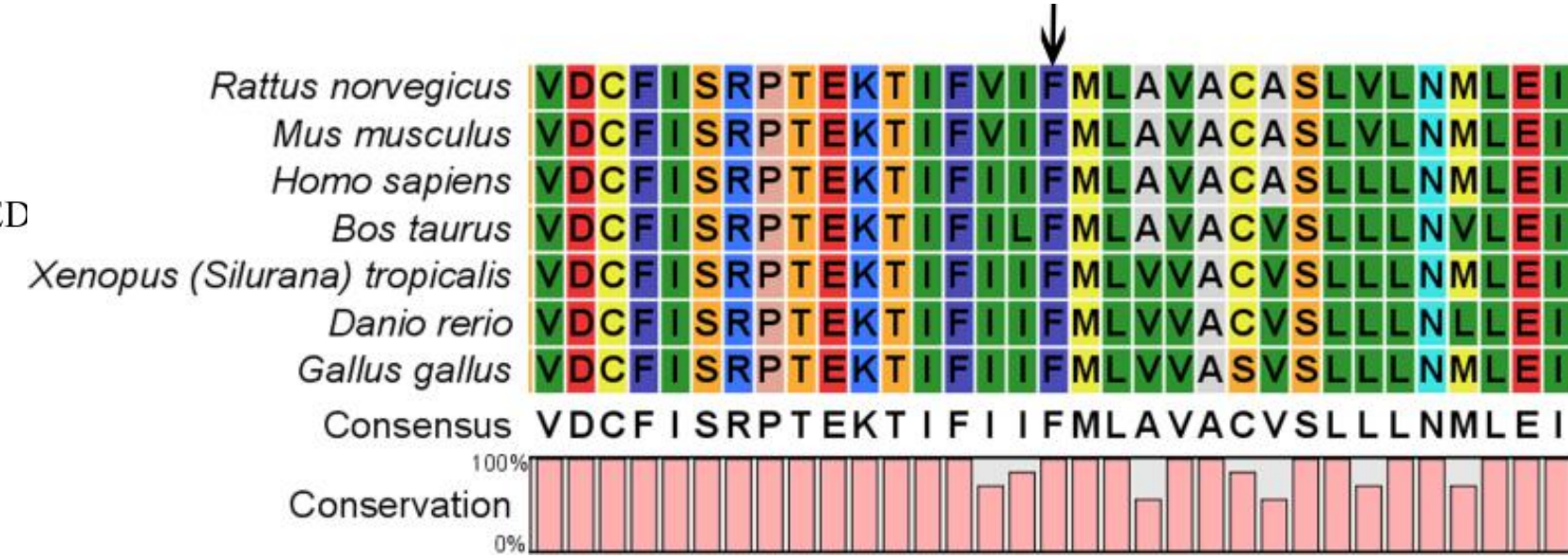
Pathogenicity of mutations: Beyond conservation

2001 – SIFT, SNP3D-stability
2002 – Polyphen
2003 – Panther
2004 – PMUT, PFAM LogRE
2005 – LS-SNP, SNP3D-seq

2007 – SNAP, CanPredict, SAPRED
Torkamani (Kinases)

2009 – SNPs&GO
2010 – Polyphen-2, MuD

2012 – KinMut (Kinases)
2013 – NetDiseaseSNP



Where do we go from here?