
DTU

DTU Health
Technology
Bioinformatics

Functional Human Variation

Gabriel Renaud
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gabriel.reno@gmail.com

Why is it important?

- Tay-Sachs is a genetic disorder
- Babies do not sit or crawl
- Followed by seizures, hearing loss
- Death occurs before 15
- The cause?

  4 bases out of 3,088,286,401 bases

http://www.curetay-sachs.org/

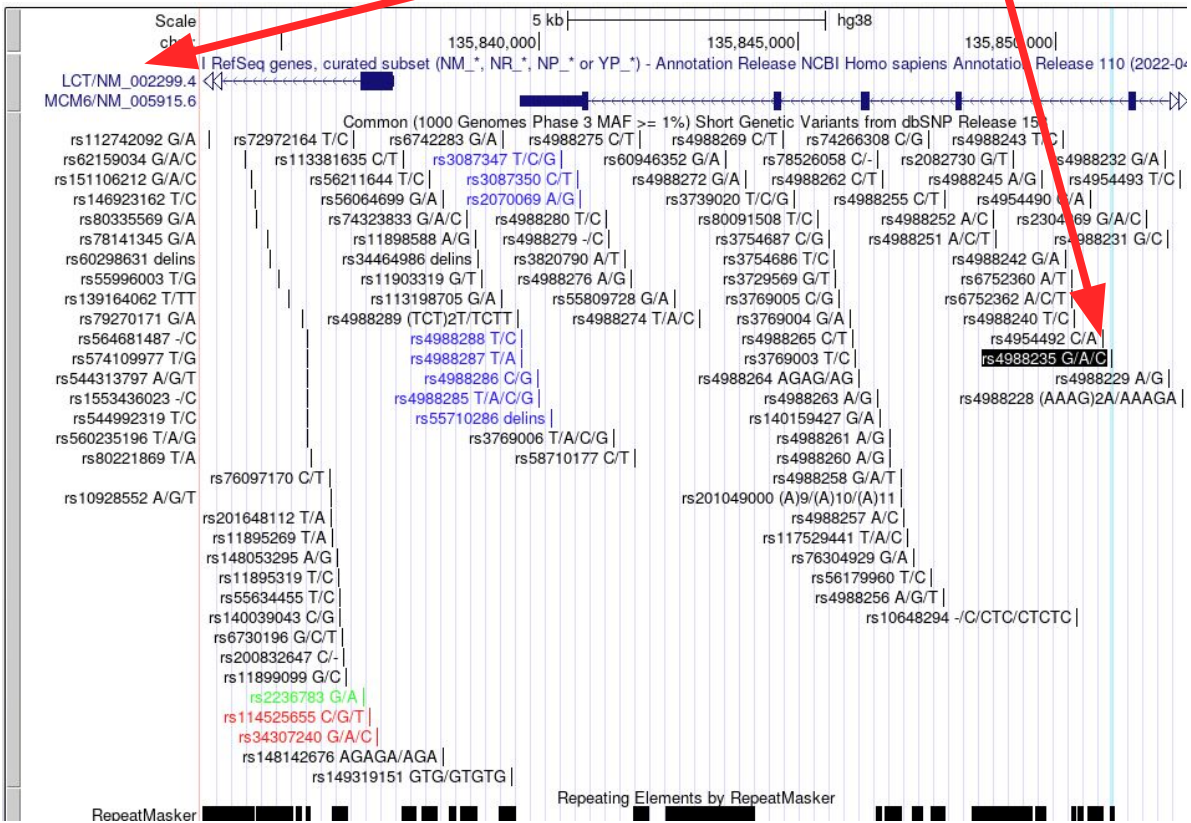
Got milk?


SORBET z MALIN
TRUSKAWKOWE
TIRAMISU
WANILIOWE
ŚLIWKA w CZEKOLADZIE

rs4988235 on chr2

genes:

SNPs:

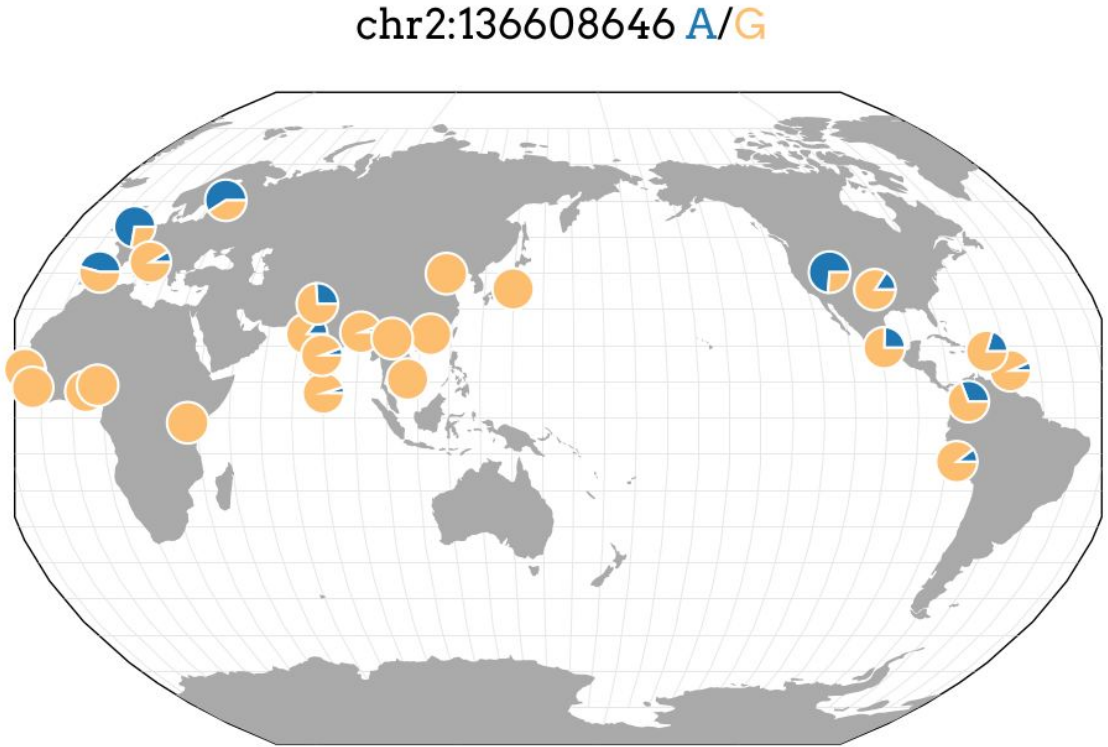


rs4988235

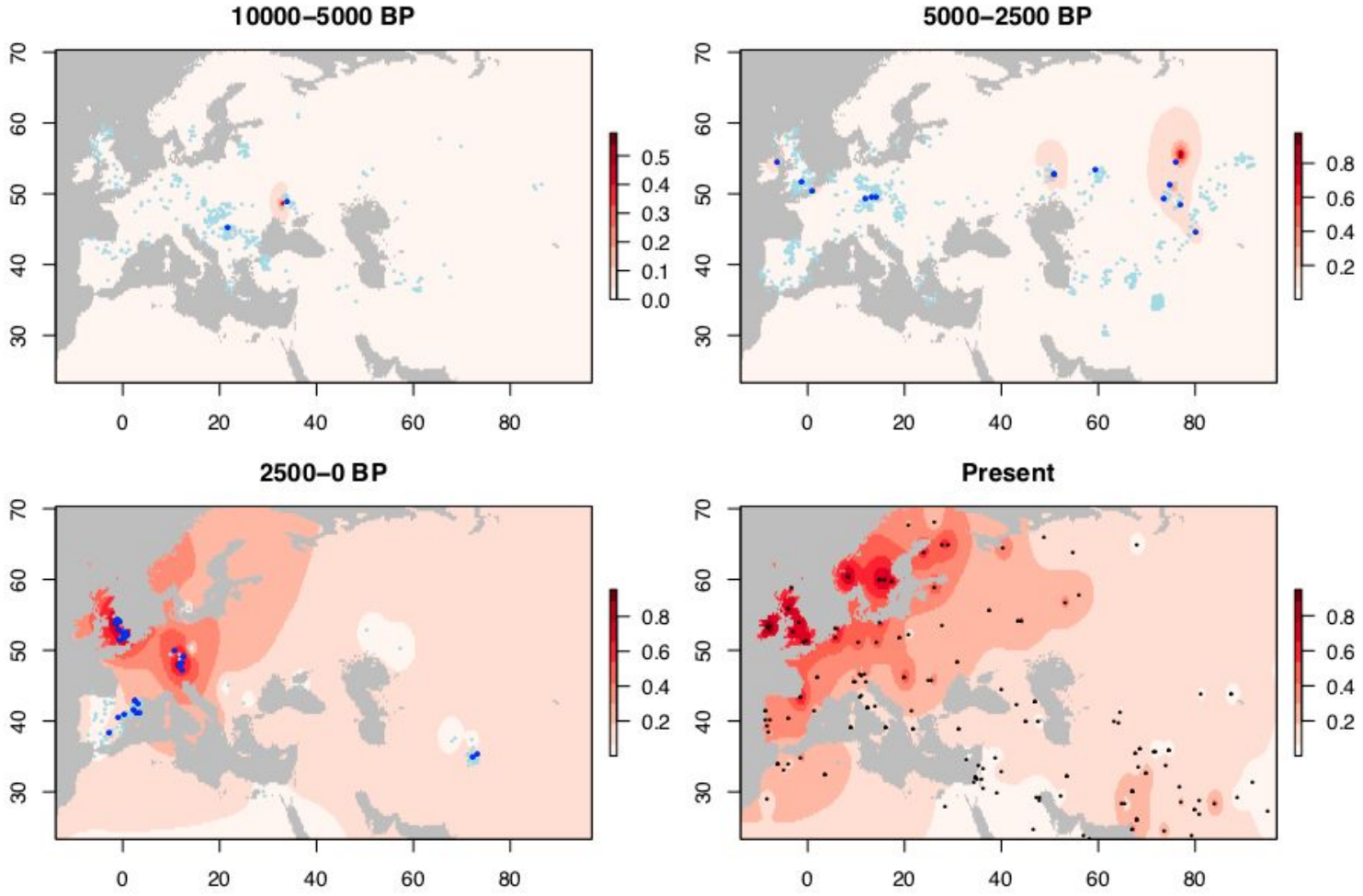
Orientation minus
Stabilized minus

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

source: snpedia

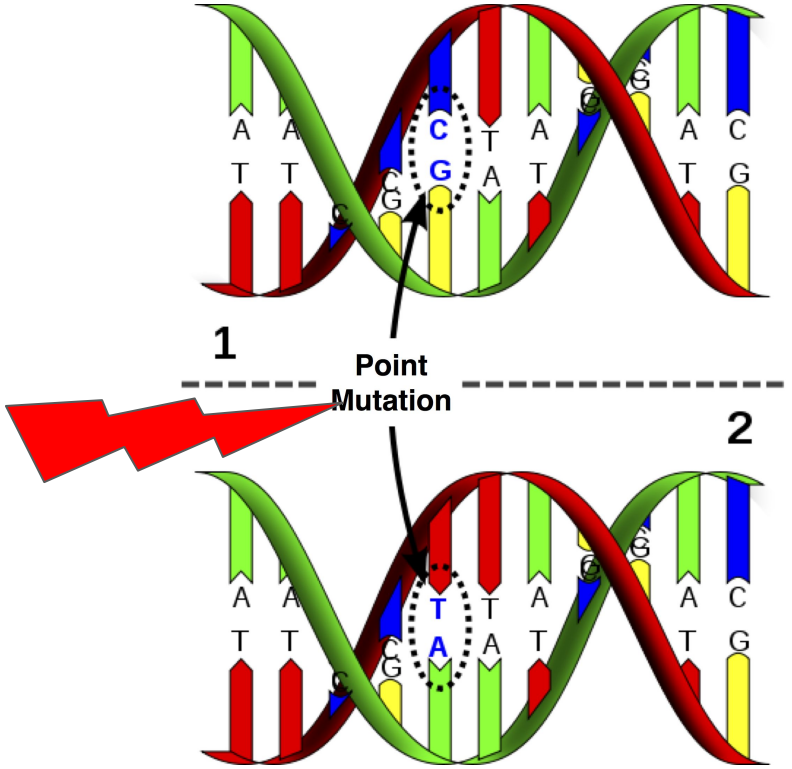


source: Geography of Genetic Variants Brows



source: http://mathii.github.io/2019/10/12/the-spread-of-the-european-lactase-persistence-allele

Variations: Mutations and Polymorphisms

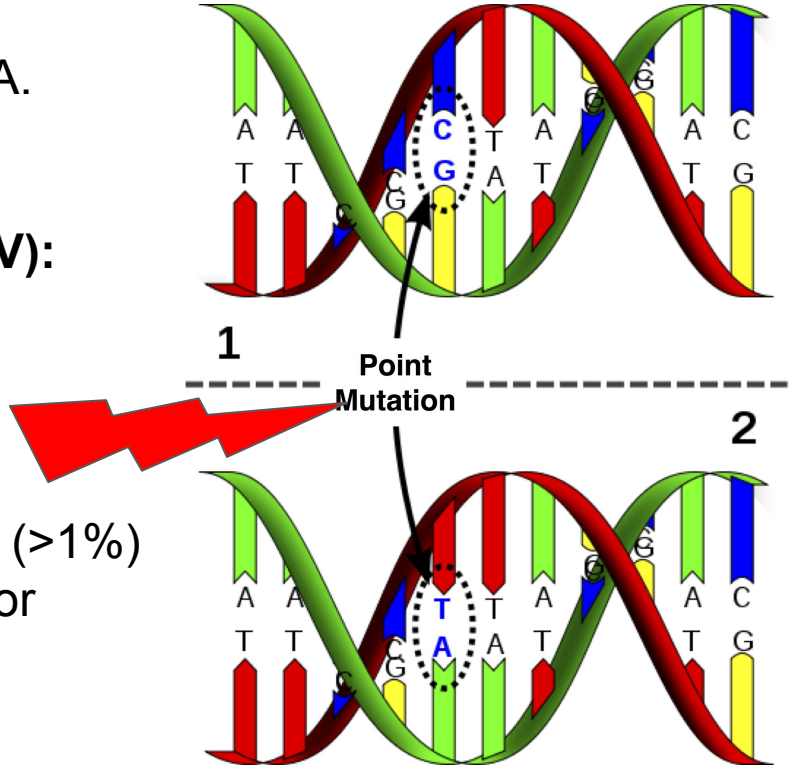


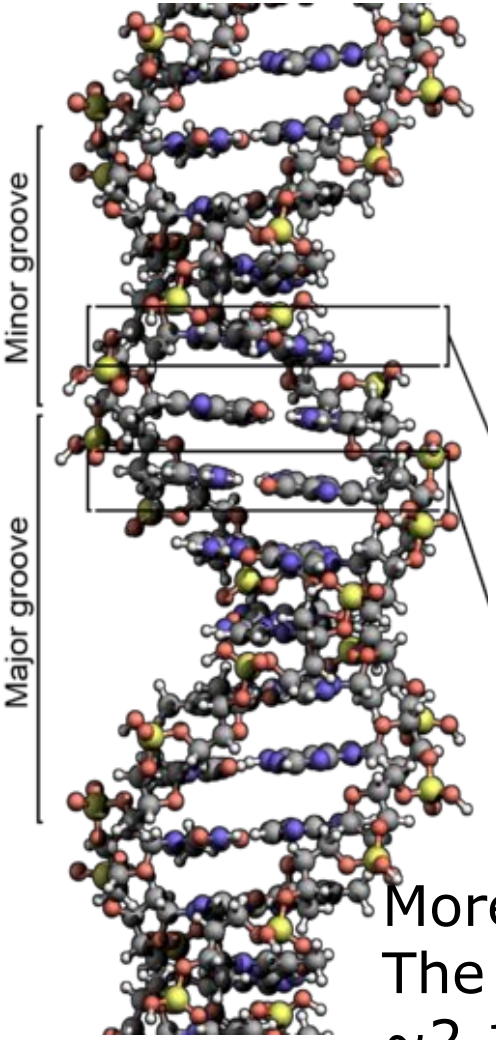
Variations: Mutations and Polymorphisms

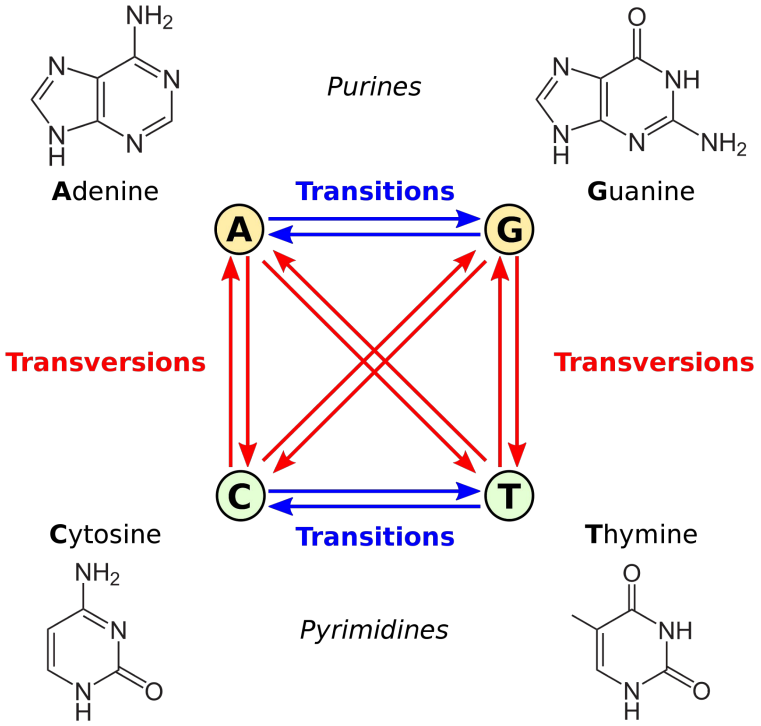
- Changes in the sequence of DNA.

Single Nucleotide Variation (SNV):
- single nucleotide is altered
- UV, chemicals, tobacco…

Polymorphism (SNP):
- Implies presence in population (>1%)
- Allele frequencies: Major / Minor





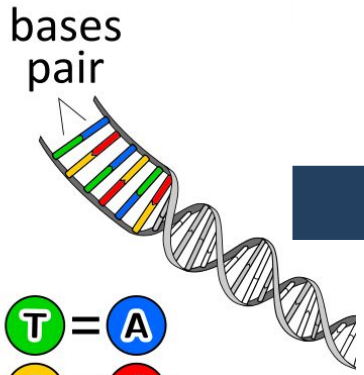


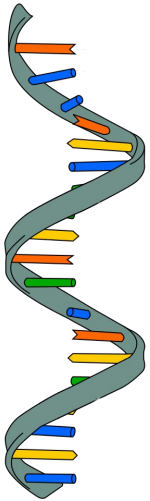
More transitions observed:

The transition/transversion ratio in human is ~2.1

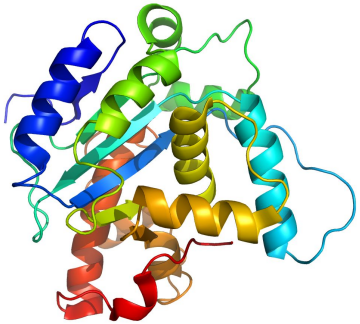
Why are mutations important?

DNA





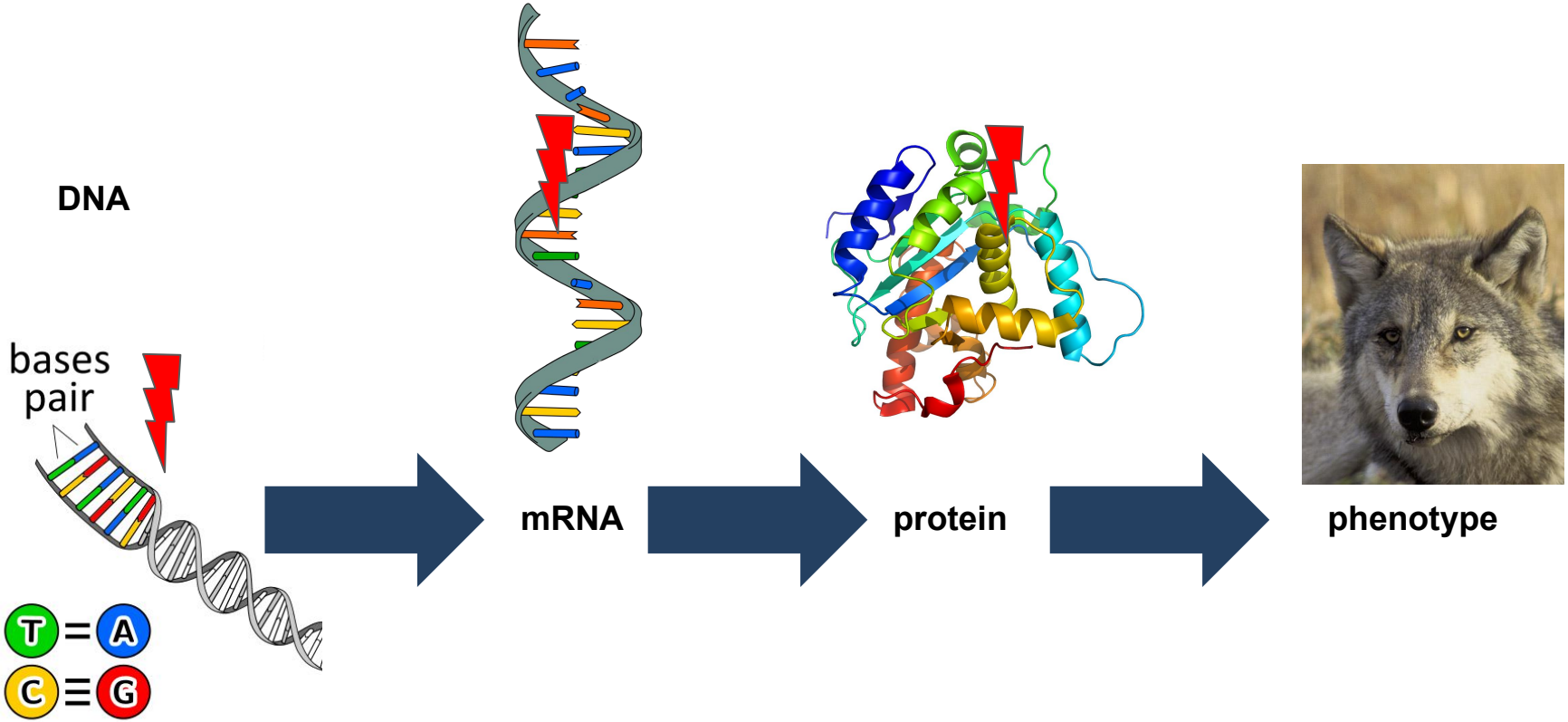
mRNA



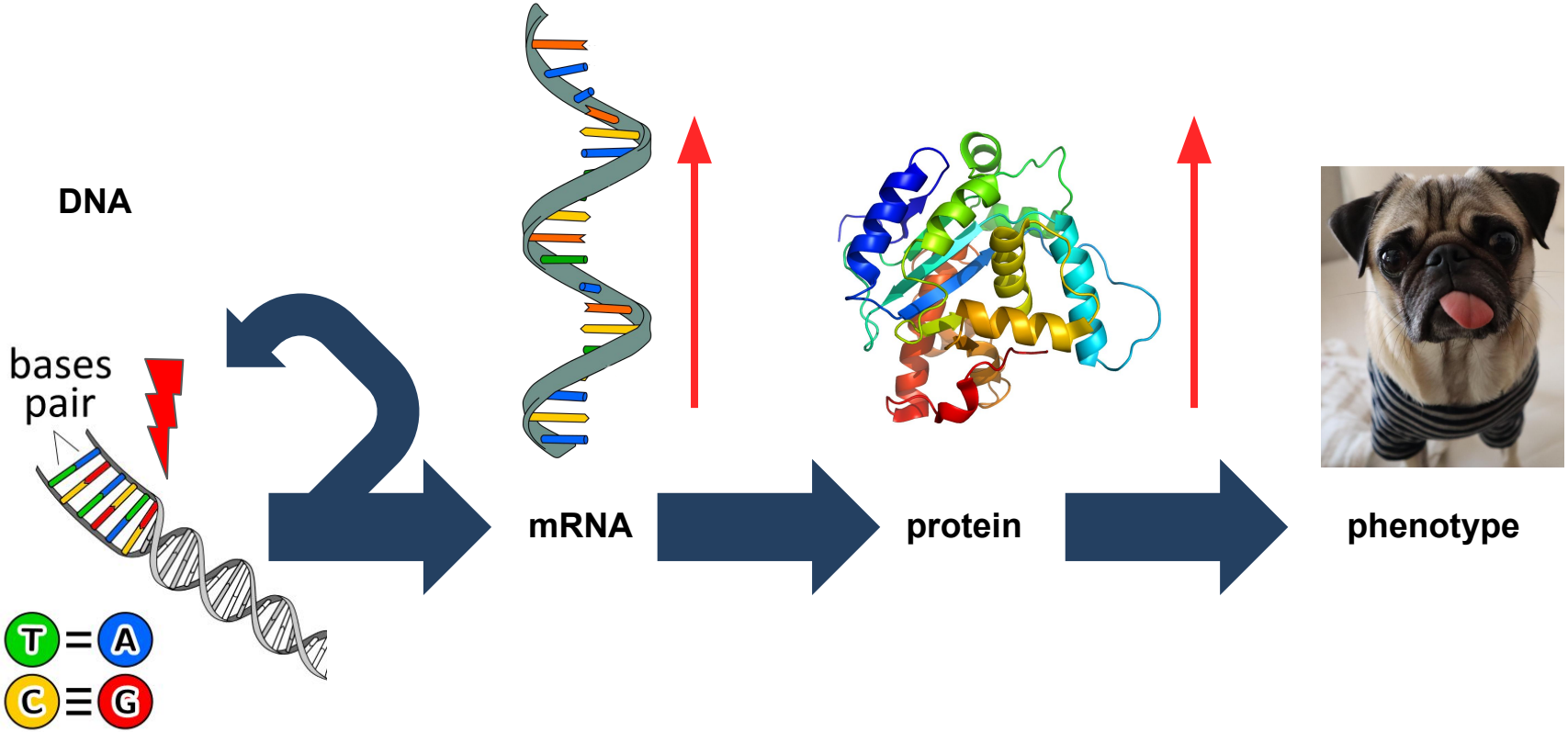
protein

phenotype

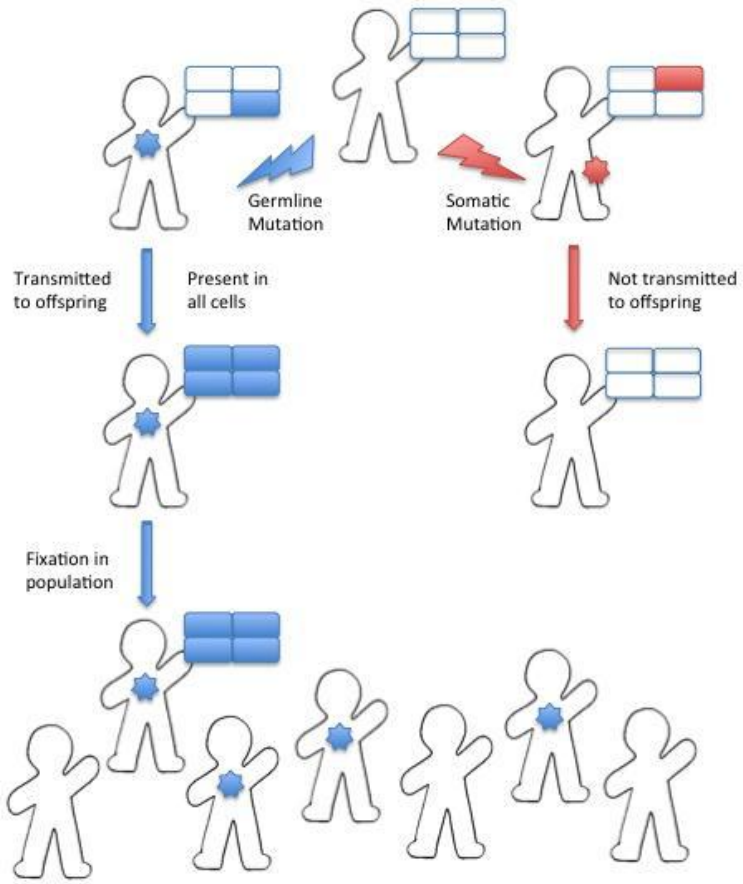
Change in the coding sequence



Change in the expression



Germline *vs* Somatic Mutations

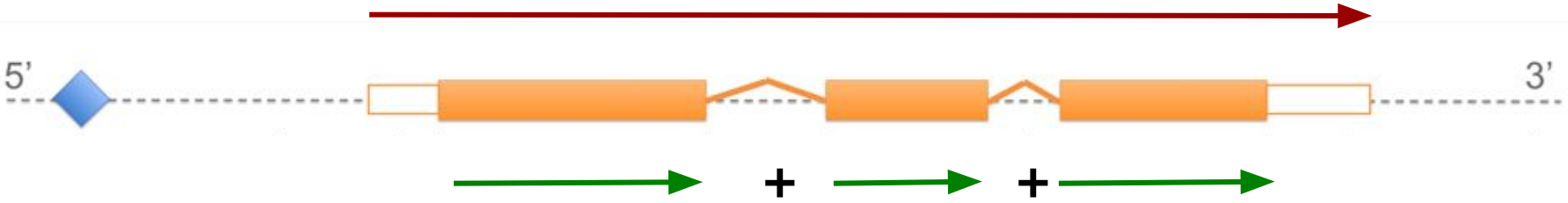


Germline Mutations
Present in **all** cells
Transmitted to offspring
Exists in population (SNP)

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

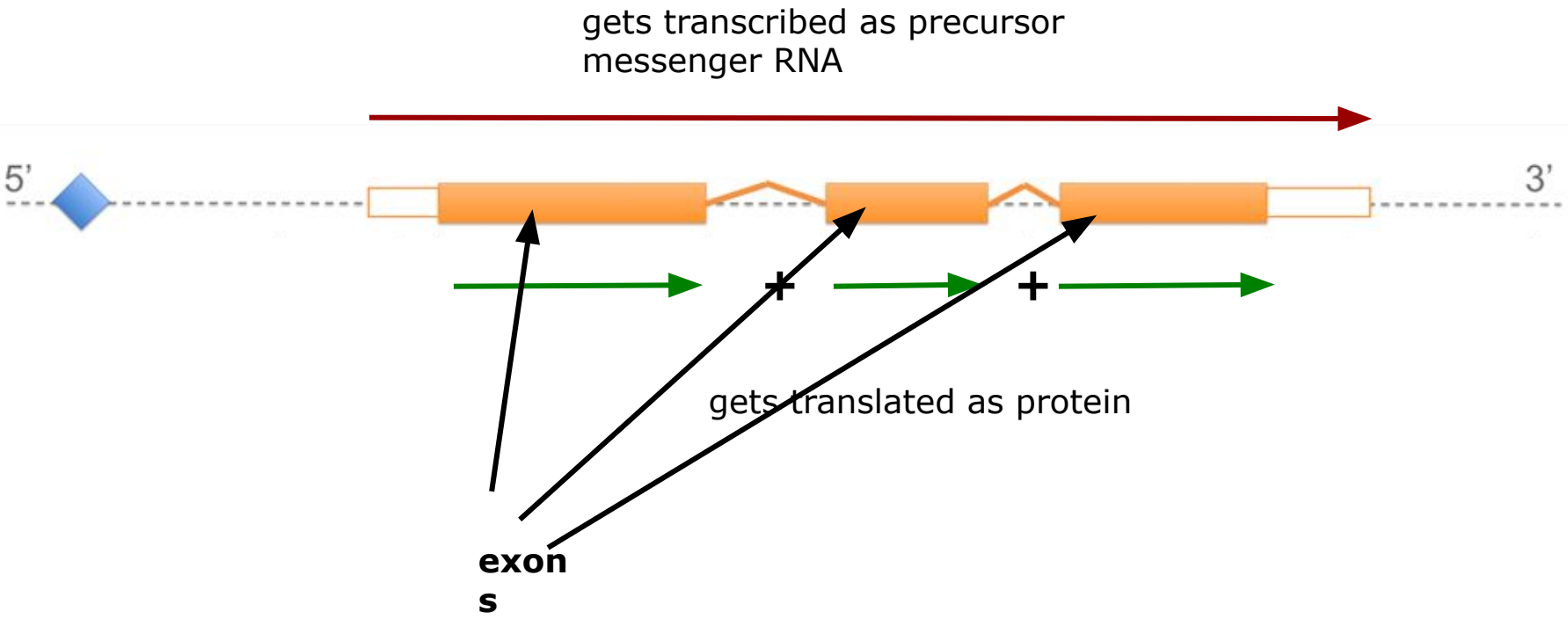
Some protein coding gene:

gets transcribed as precursor
messenger RNA

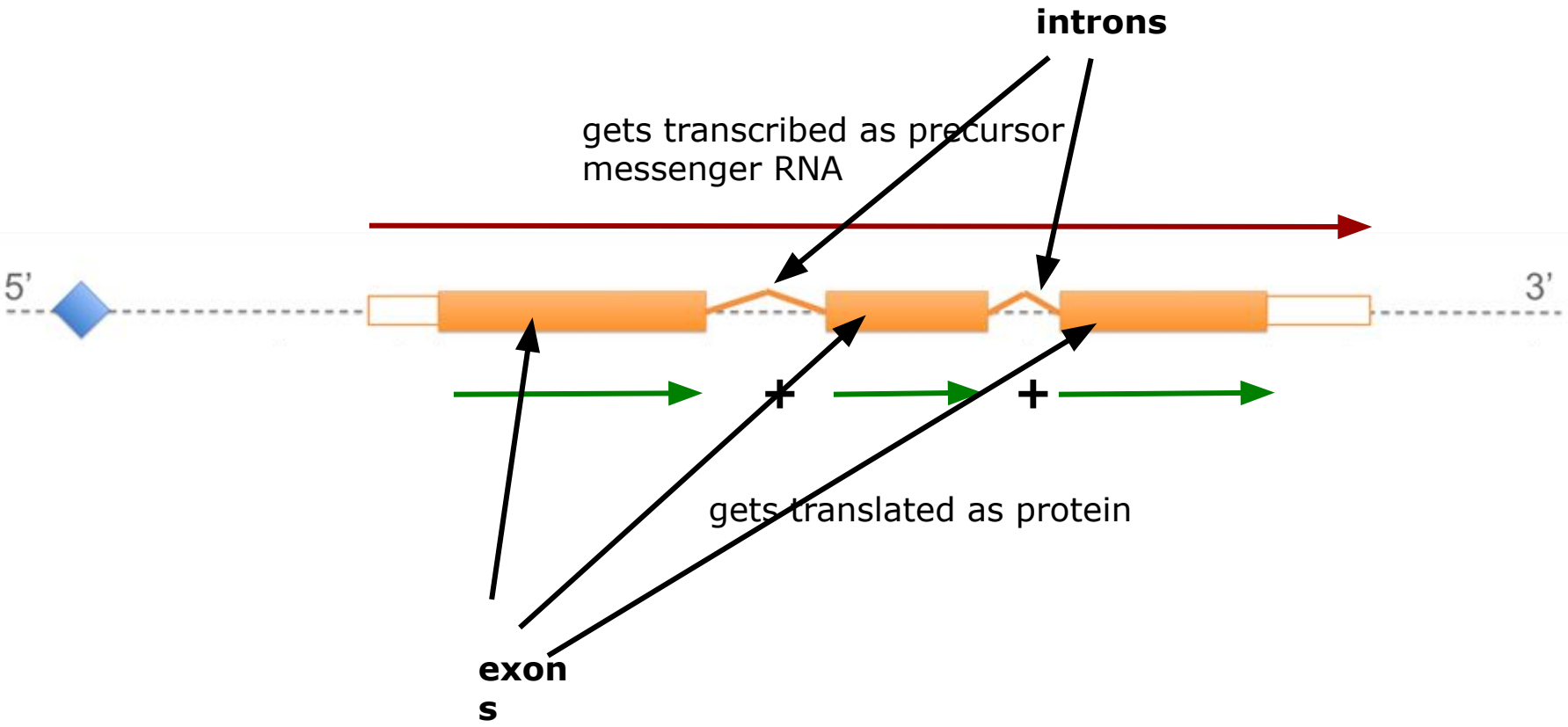


gets translated as protein

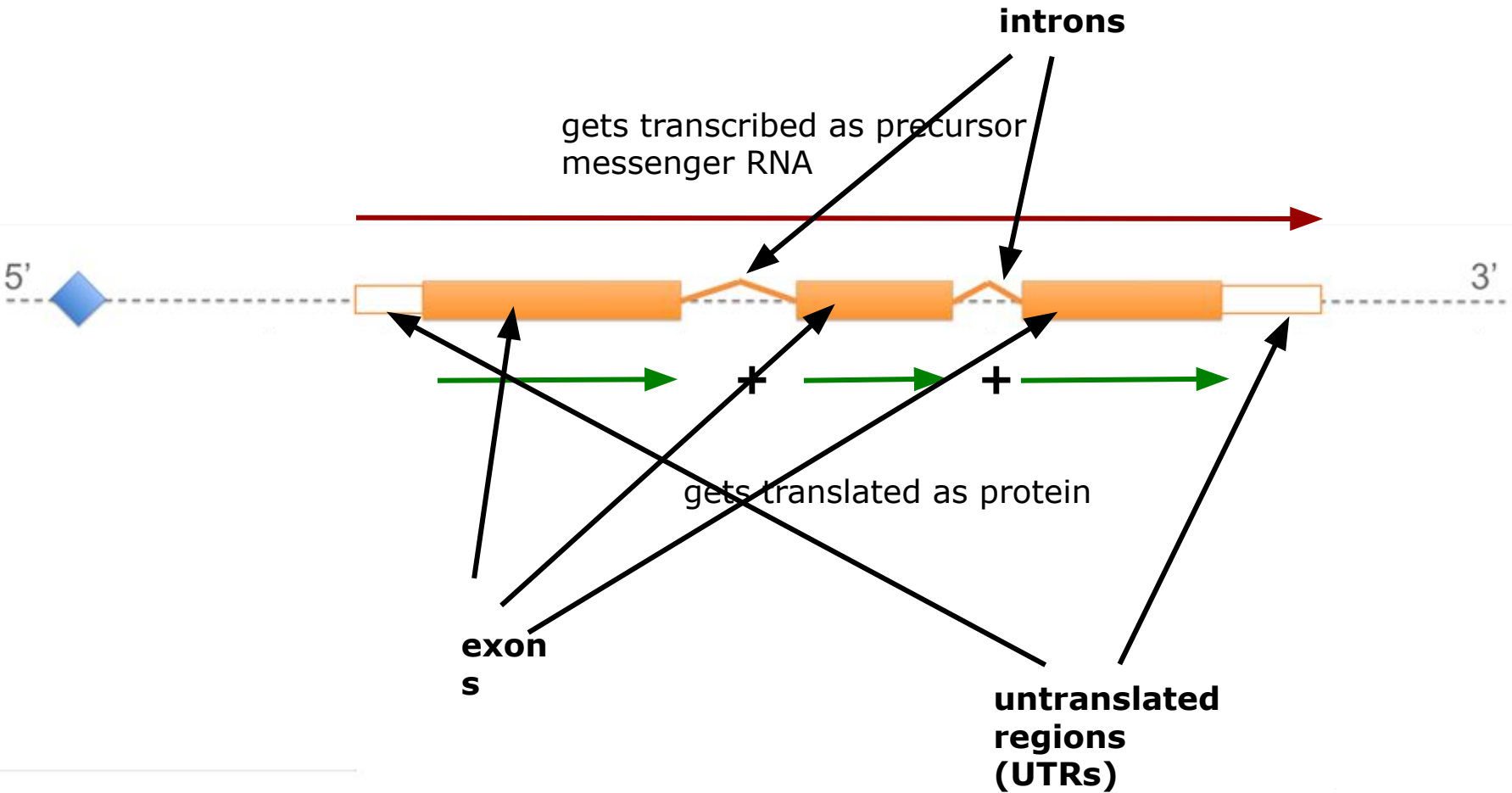
Some protein coding gene:



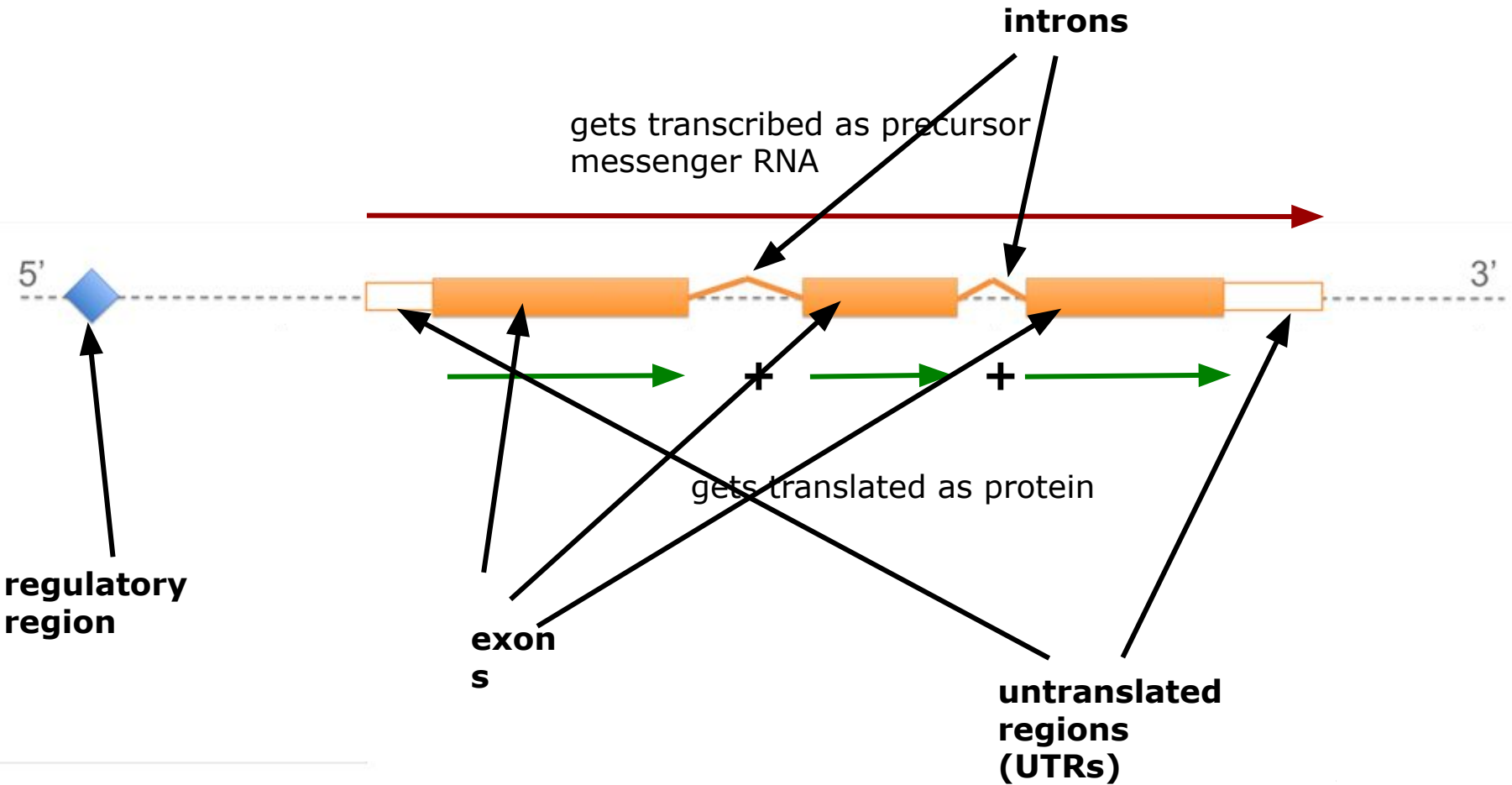
Some protein coding gene:



Some protein coding gene:

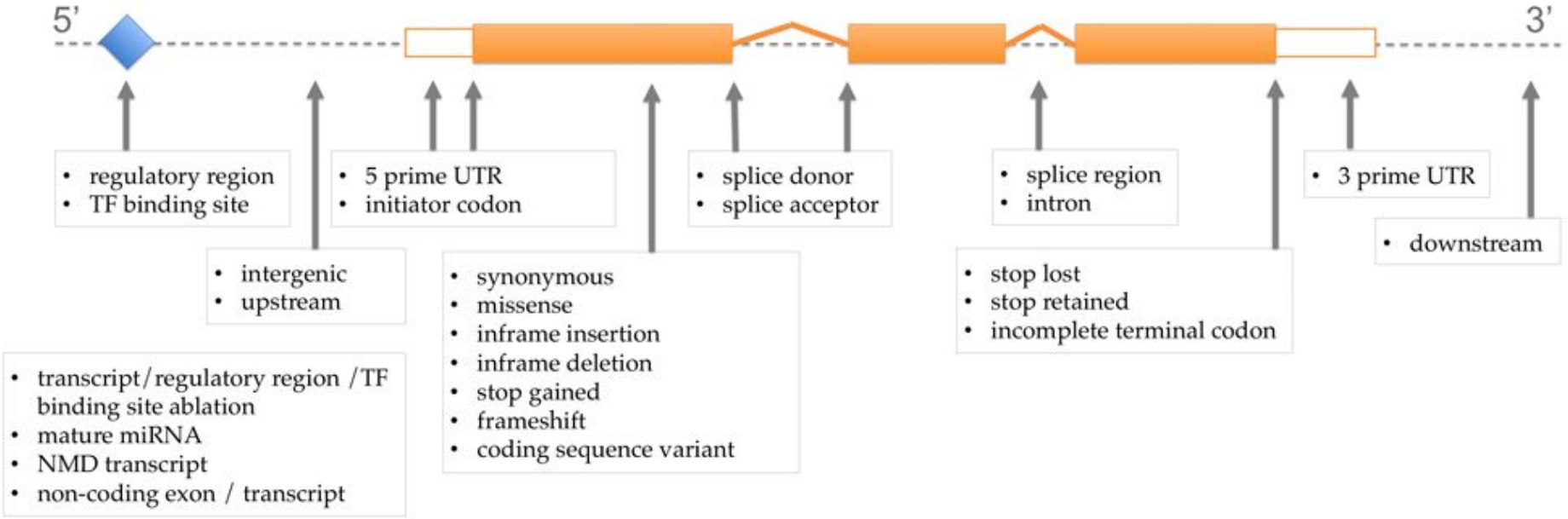


Some protein coding gene:

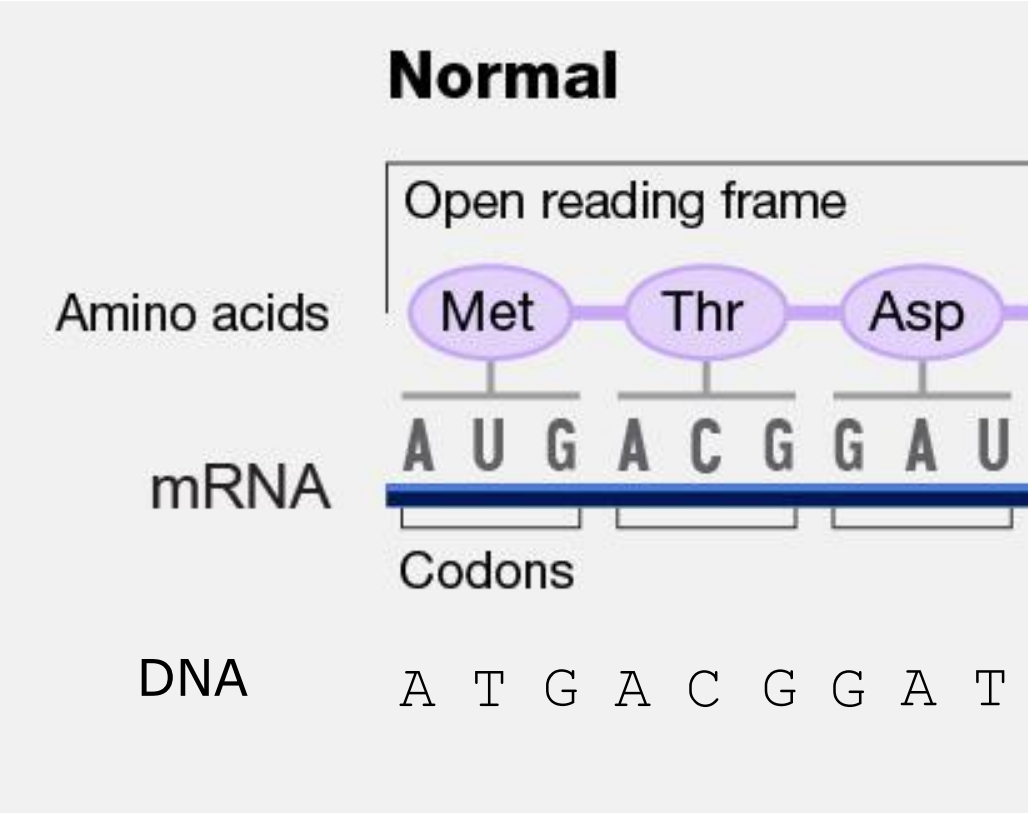


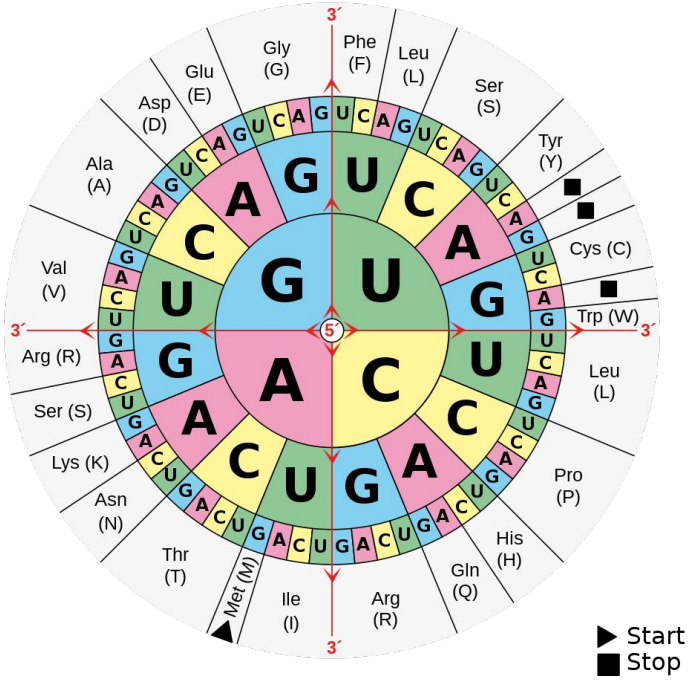
Coding *vs* Non-Coding variations

- There are different regions in the genome
- Mutations in different regions will have different consequences



RNA codons





source: genome.gov

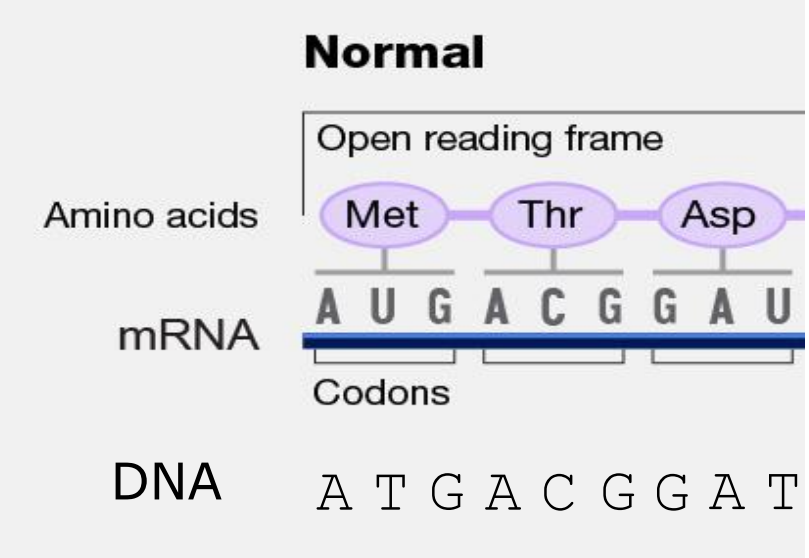
Consequences of coding mutations

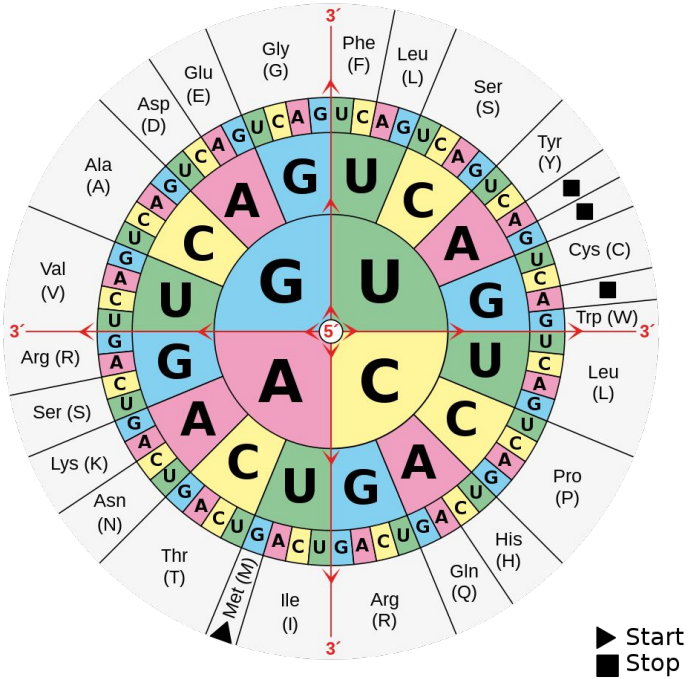
Silent

ACG -> ACA

Threonine

Threonine





source: genome.gov

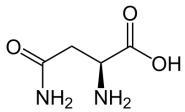
Consequences of coding mutations

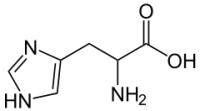
Degeneration of the genetic code

Non-Synonymous (missense)

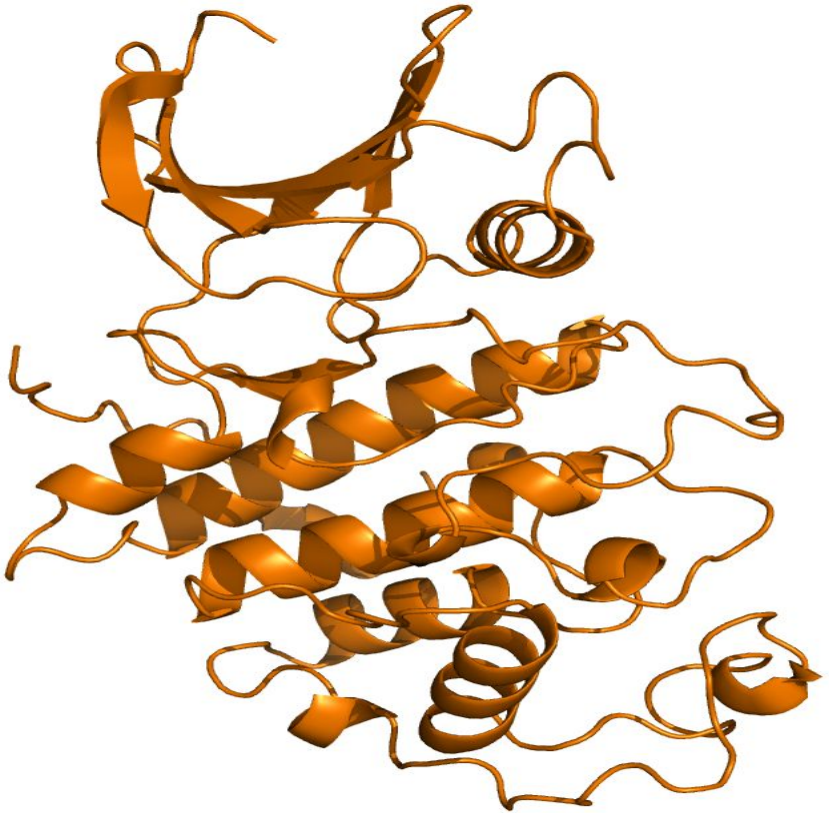
AAC -> CAC

Asparagine Histidine

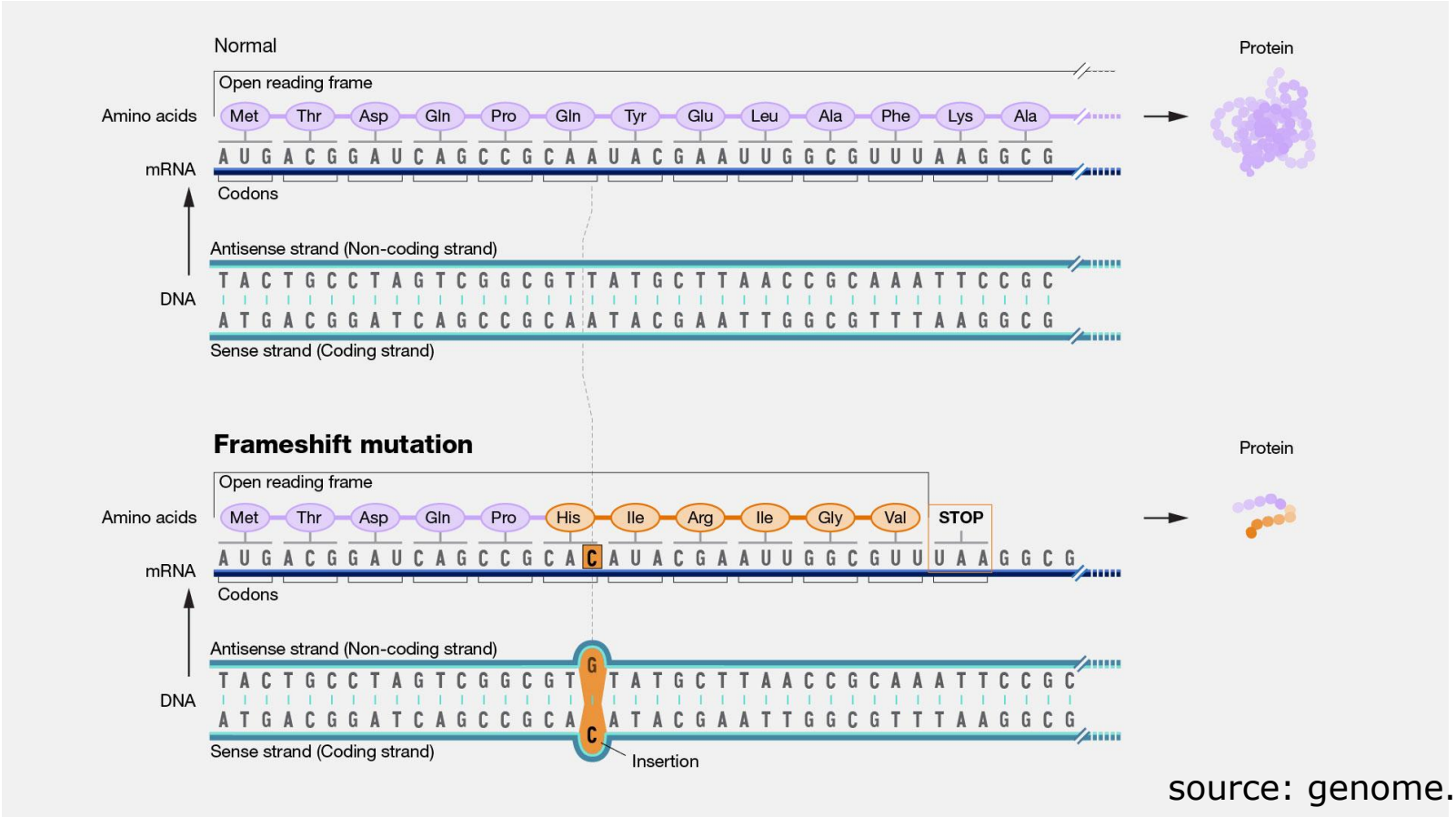




- Different chemical properties
- Altered protein structure

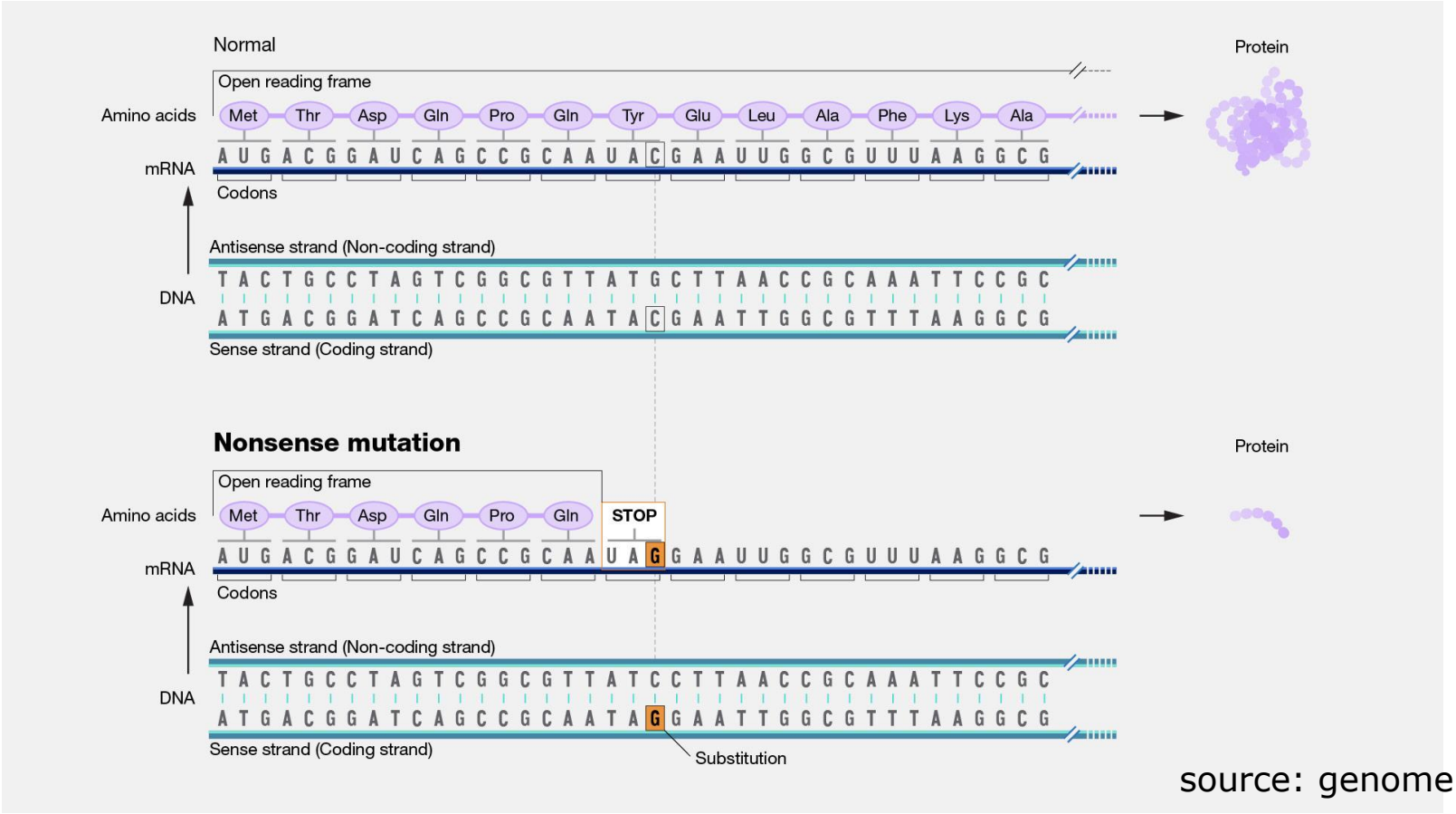


Consequences of coding mutations



source: genome.gov

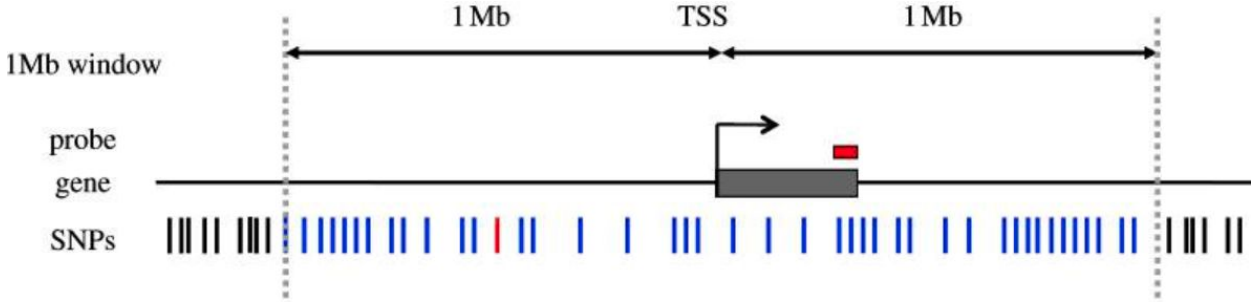
Consequences of coding mutations

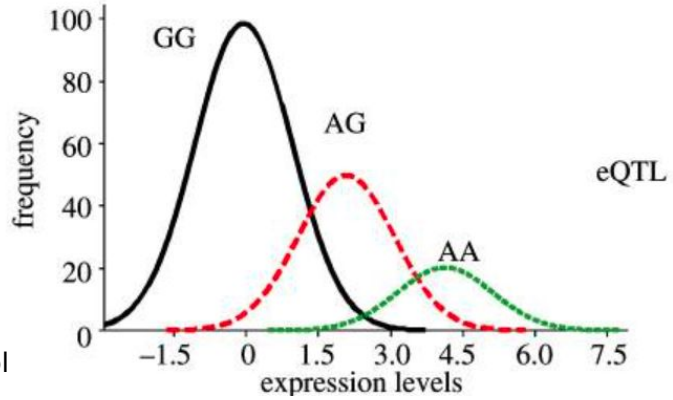


source: genome.gov

Consequences of non-coding mutations

1) change in expressions



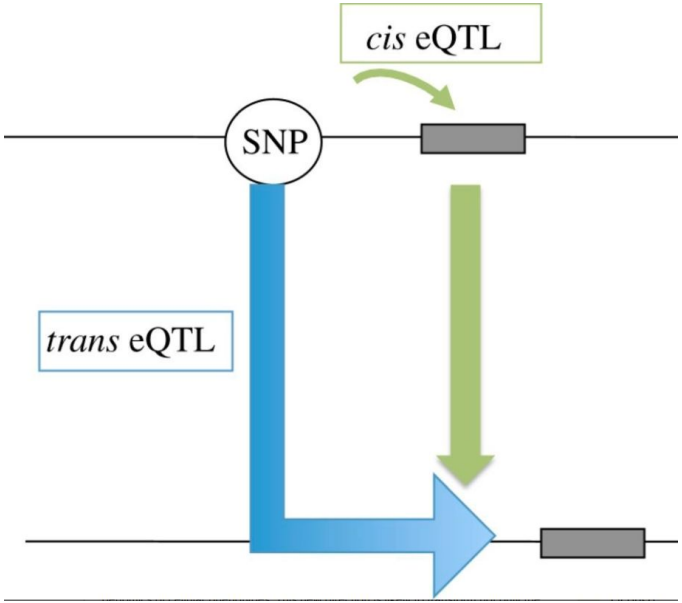


eQTL = **E**xpression **Q**uantitative **T**rait **L**oci = locus responsible for a fraction of the genetic variance of a gene expression phenotype

Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci. 2013 May 6;368(1620):20120362. doi: 10.1098/rstb.2012.0362. PMID:

Consequences of non-coding mutations

1) change in expressions

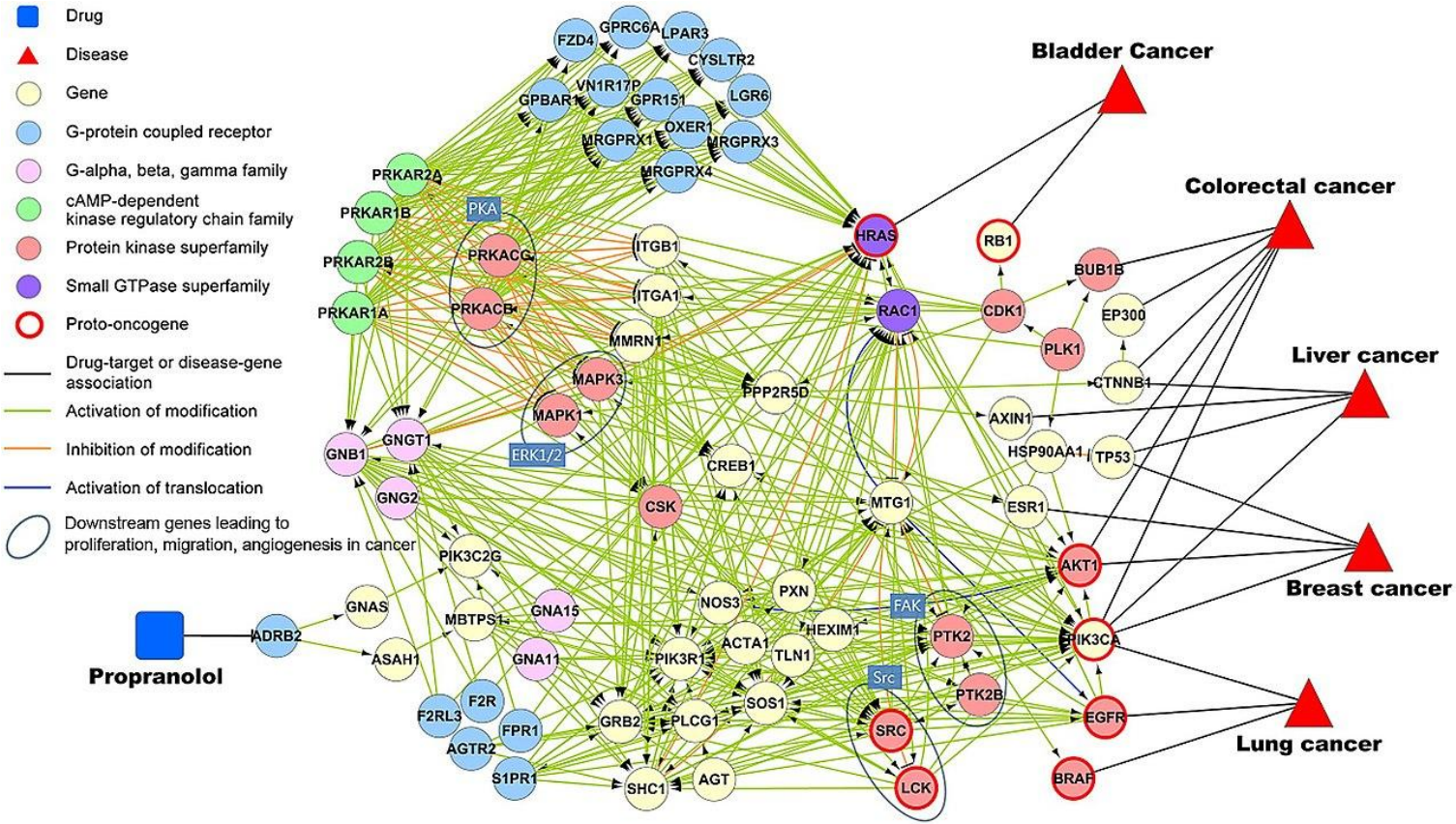


eQTL (expression quantitative trait loci)

locus responsible for a fraction of the genetic variance of a gene expression phenotype

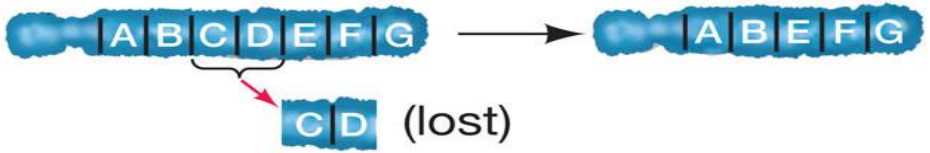
Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci. 2013 May 6;368(1620):20120362. doi: 10.1098/rstb.2012.0362. PMID:

Consequences of mutations…. it's not easy.

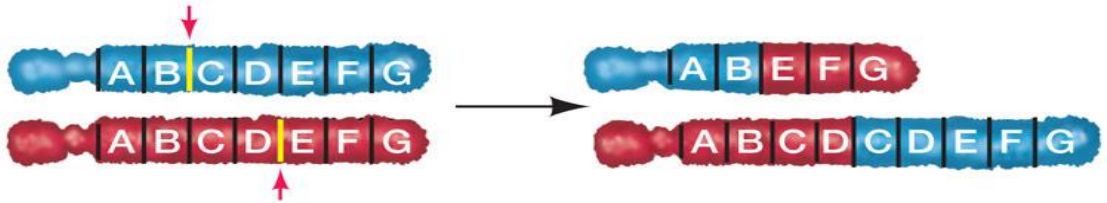


Chromosomal rearrangements

Deletion/Insertion



Deletion + Duplication:

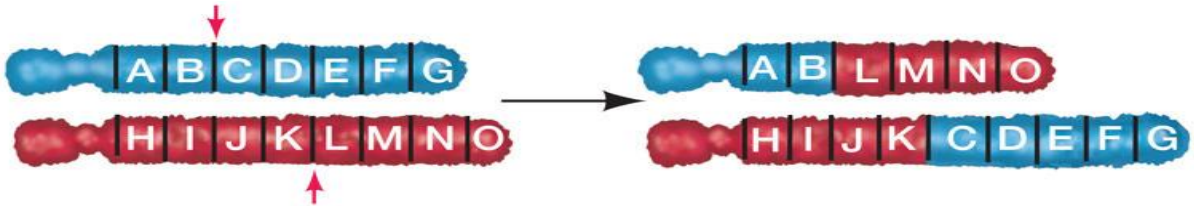


(LIFE 10e, Figure 15.4)

Chromosomal rearrangements

Inversion

Translocation:



(LIFE 10e, Figure 15.4)

Conservation and mutations

Human	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Chimpanzee	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Orangutan	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Macaque	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Mouse	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Rat	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Dog	E	S	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Horse	E	G	R	A	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Cow	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Opossum	E	T	K	T	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Zebrafish	V	G	K	L	P	K	S	I	D	Y	R	K	L	G	Y	V	T	S	V	K	N
Medaka	L	I	K	L	P	K	S	V	D	Y	R	K	K	G	M	V	T	S	V	K	N
Tetraodon	V	Q	R	L	P	R	N	L	D	Y	R	K	K	G	A	V	T	A	V	K	D
Tetraodon	G	F	E	T	P	P	S	V	D	W	R	K	A	G	L	V	S	P	V	Q	N
Fugu	G	A	D	L	P	Q	T	V	D	W	R	D	K	G	L	V	T	S	V	K	K

The mutation of which position looks more pathogenic?

Are there amino acids better tolerated than other in (red) position?

And what about the position in (green)?

(Ng, 2001)

SIFT: Pathogenicity of mutations

- SIFT is based on amino acid conservation across species
- Mutation of highly conserved ☐ Pathogenic
- http://blocks.fhcrc.org/sift/SIFT.html
- Pathogenic if **SIFT score < 0.05**

Human	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Chimpanzee	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Orangutan	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Macaque	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Mouse	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Rat	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Dog	E	S	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Horse	E	G	R	A	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Cow	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Opossum	E	T	K	T	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Zebrafish	V	G	K	L	P	K	S	I	D	Y	R	K	L	G	Y	V	T	S	V	K	N
Medaka	L	I	K	L	P	K	S	V	D	Y	R	K	K	G	M	V	T	S	V	K	N
Tetraodon	V	Q	R	L	P	R	N	L	D	Y	R	K	K	G	A	V	T	A	V	K	D
Tetraodon	G	F	E	T	P	P	S	V	D	W	R	K	A	G	L	V	S	P	V	Q	N
Fugu	G	A	D	L	P	Q	T	V	D	W	R	D	K	G	L	V	T	S	V	K	K

(Ng, 2001)

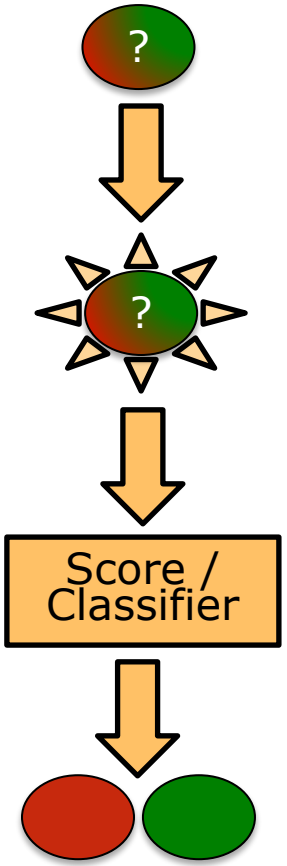
Methods to predict pathogenicity

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



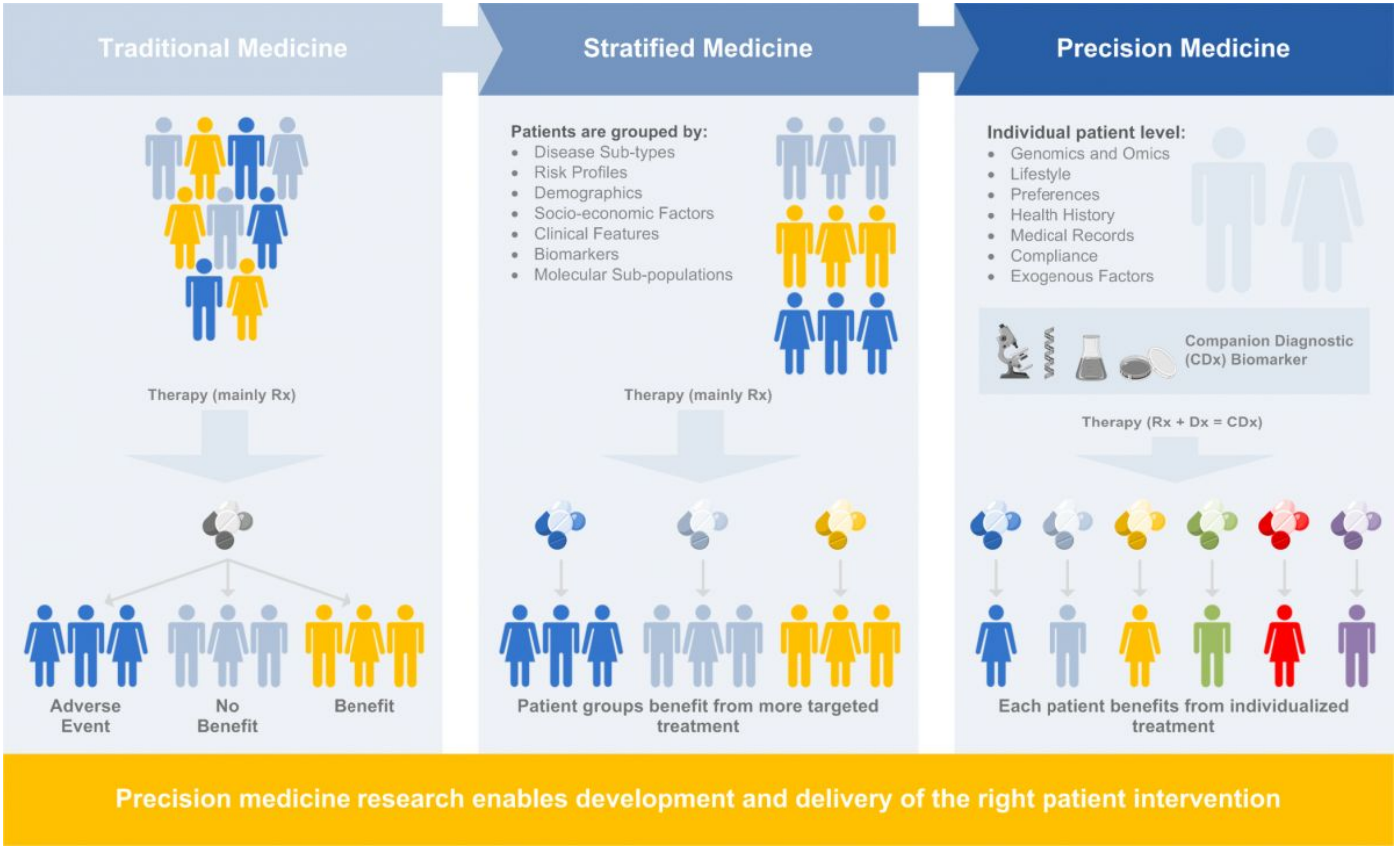
Novel Mutation

Features Added

Prediction of Pathogenicity

Pathogenic or Neutral Mutation

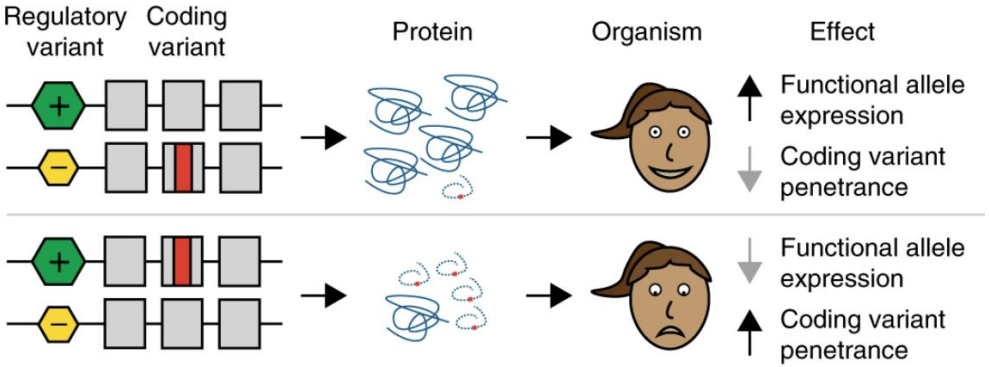
The future: Personalized/Stratified medicine



The future: Personalized/Stratified medicine

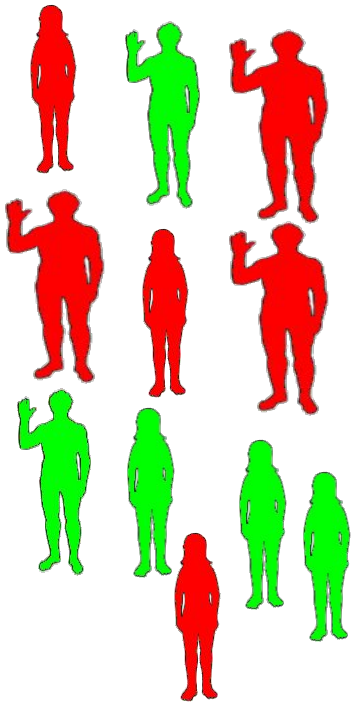
Hurdles to personalized medicine:

- no 1 mutation -> 1 gene
- rare variants
- variants on the same chr vs different: haplotypes



Castel, S.E., Cervera, A., Mohammadi, P. *et al.* Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet* 50, 1327–1334 (2018). https://doi.org/10.1038/s41588-018-0192-y

The future: Polygenic risk scores

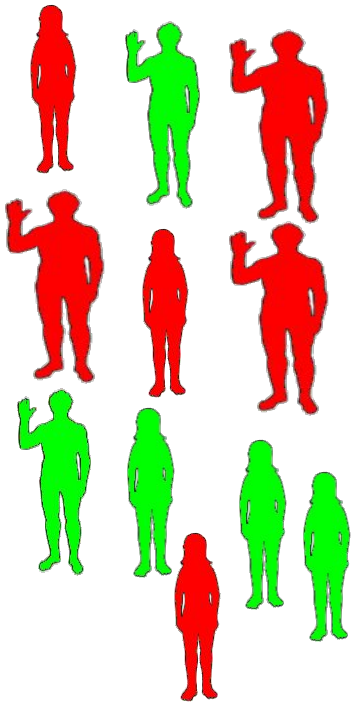


High risk for disease X

Low risk for disease X

The future: Polygenic risk scores

$$PRS = w_1 SNP_1 + w_2 SNP_2 + w_3 SNP_3 + \ldots$$



The future: Polygenic risk scores

$$PRS = w_1 SNP_1 + w_2 SNP_2 + w_3 SNP_3 + \ldots$$