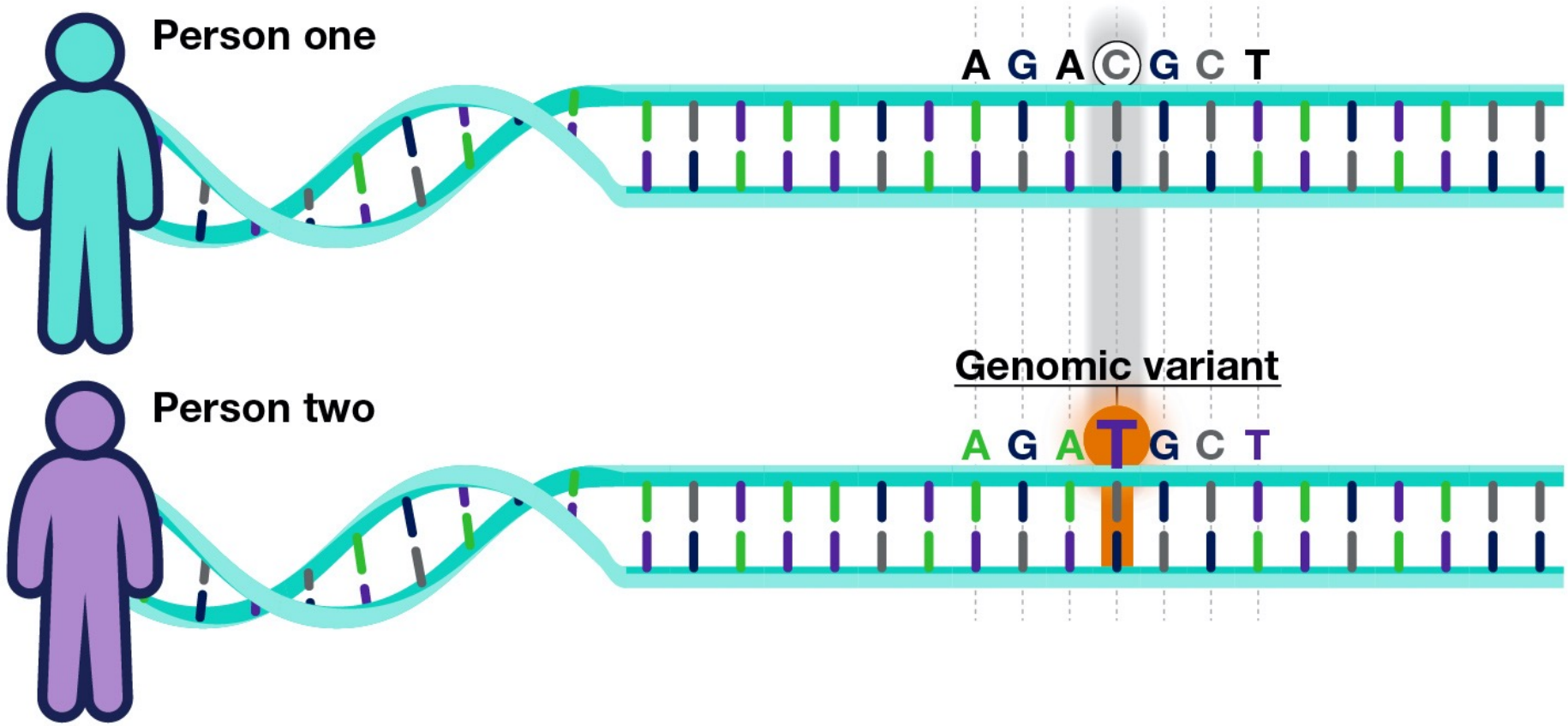

Person one
A G A C G C T
Genomic variant
Person two
A G A T G C T

1. HOW MANY VARIANTS DOES THE GENOME HAVE?

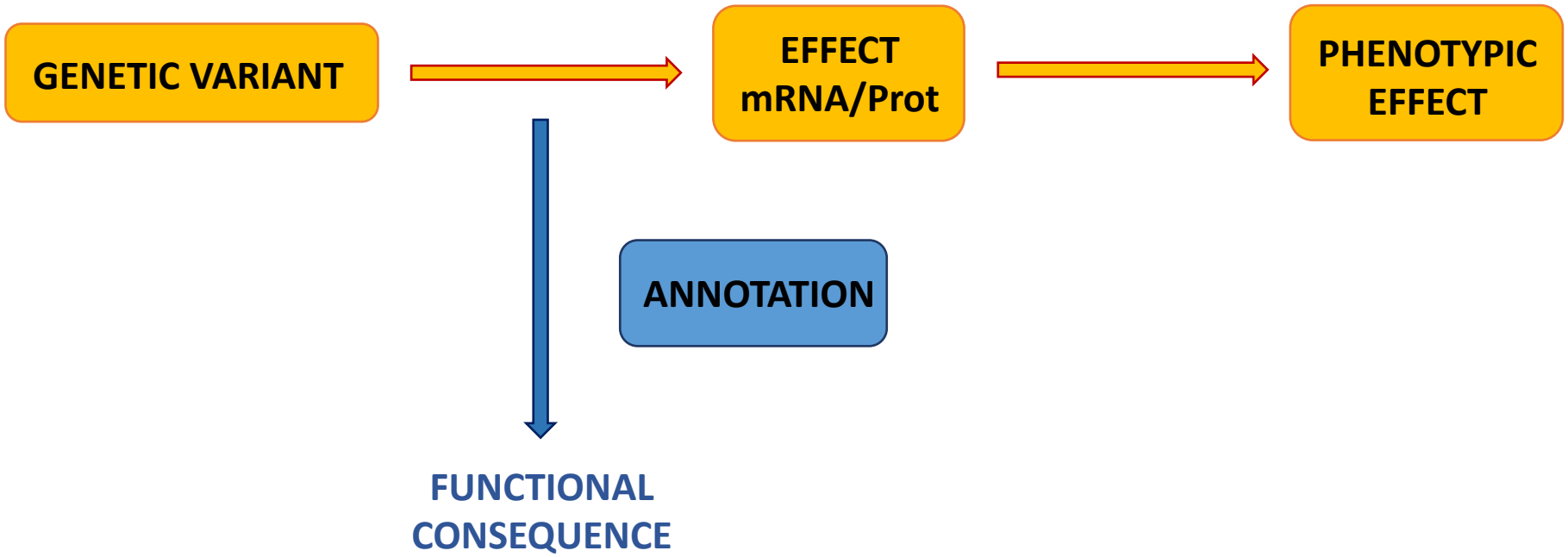
2. WHICH TYPE?

3. WHAT EFFECT DO THEY HAVE?

4. WHAT IS THE BIOLOGICAL SIGNIFICANCE?

3. WHAT EFFECT DO THEY HAVE?

Functional annotation



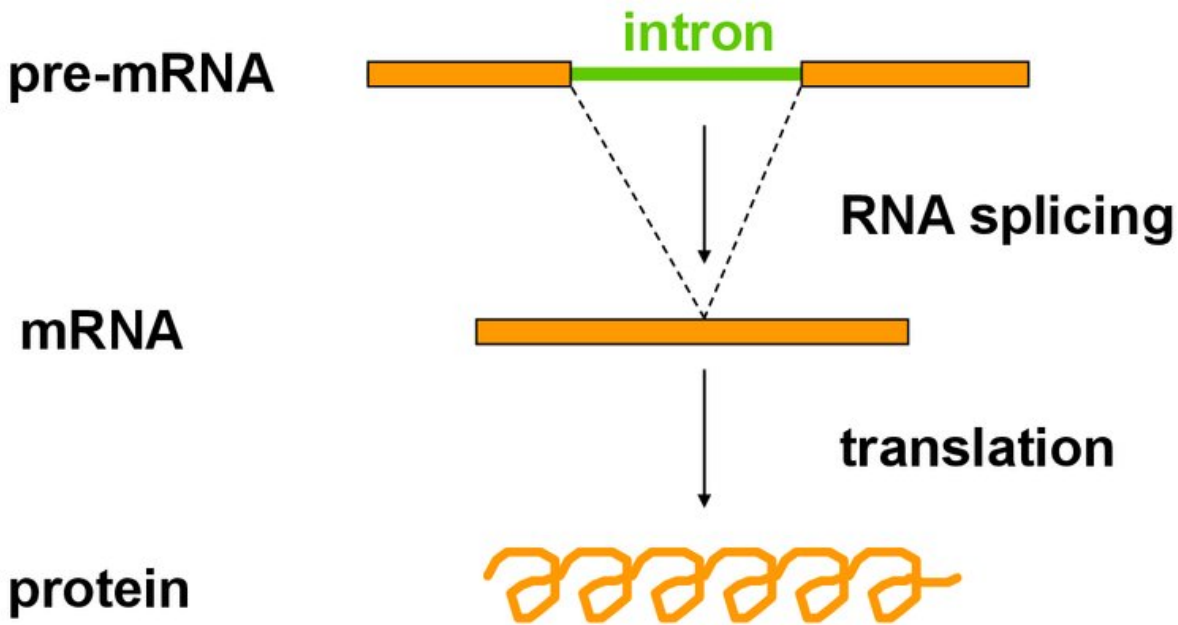
3. WHAT EFFECT DO THEY HAVE?

Loss of function (LoF) genetic variants

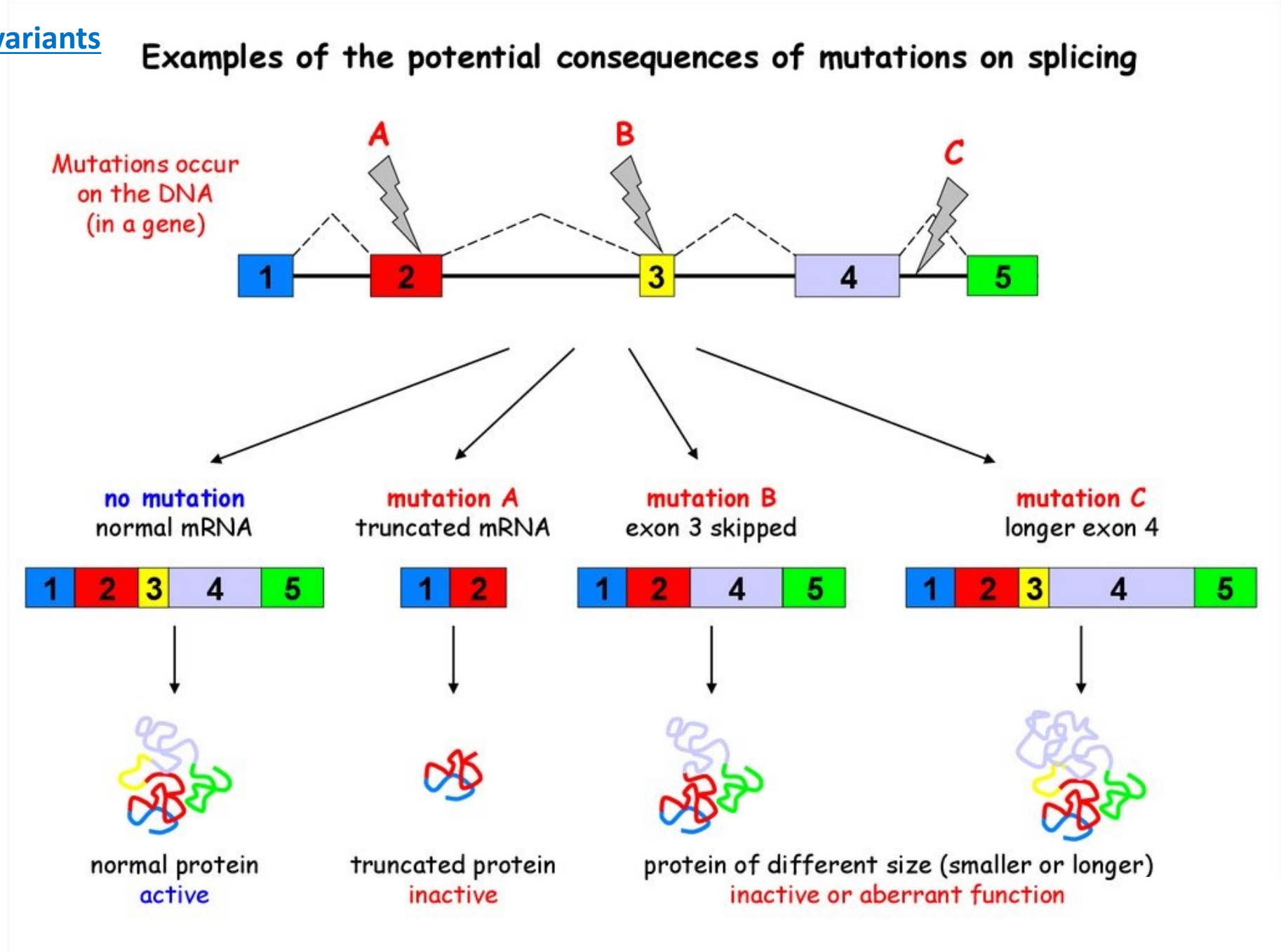
- Nonsense
- Splice site
- Frameshift indels
- Large deletions

Splice site genetic variants

The intron is also present in the RNA copy of the gene and must be removed by a process called "RNA splicing"

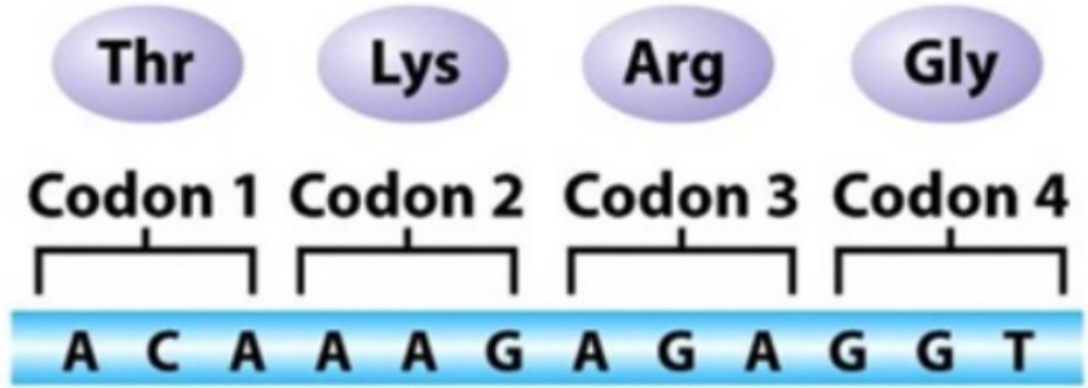


Splice site genetic variants



Indels in the coding region

Wild-type gene

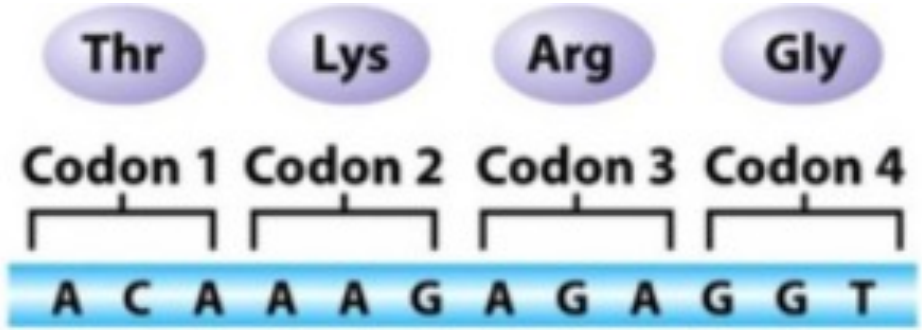


- Normal coding mRNA is called **in frame** when codons are arranged in specific sequence from start codon to stop codon.

- Insertion and deletion mutations introduce a disruption to the codon sequence called **frame-shift.**

Indels in the coding region

Wild-type gene



Gene with insertion

- Insertion point mutations introduce a base-pair or more to the sequence of the gene which causes a **frame-shift** downstream.

- The frame-shift may result in a premature stop codon down stream or changes in amino acid sequence.

Number of loss-of-function variants

Category	Filtered number/individual (CEU)	
	All	Homozygous
nonsense SNP	26.2	5.2
splice SNP	11.2	1.9
frameshift indel	38.2	9.2
large LoF deletion	28.3	6.2
total	103.9	22.3

2. TYPES OF GENETIC VARIANTS

Functional annotation

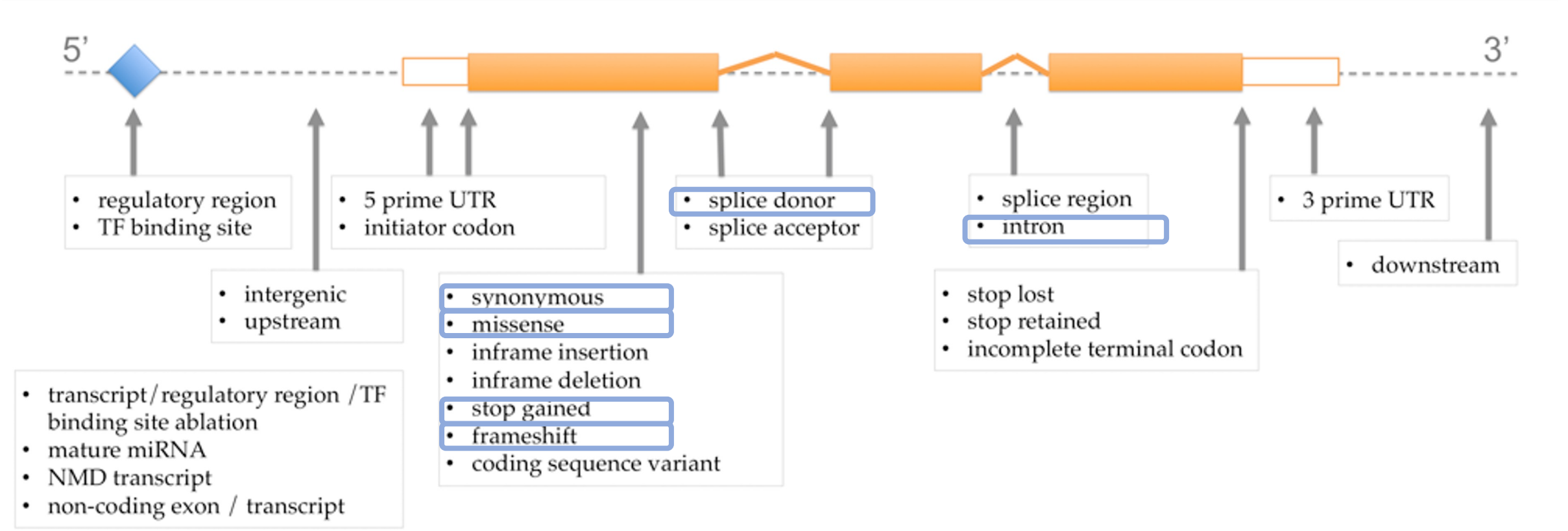
Table 1 | Median autosomal variant sites per genome

	AFR		AMR		EAS		EUR		SAS	
Samples	661		347		504		503		489	
Mean coverage	8.2		7.6		7.7		7.4		8.0	
	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons
SNPs	4.31M	14.5k	3.64M	12.0k	3.55M	14.8k	3.53M	11.4k	3.60M	14.4k
Indels	625k	-	557k	-	546k	-	546k	-	556k	-
Large deletions	1.1k	5	949	5	940	7	939	5	947	5
CNVs	170	1	153	1	158	1	157	1	165	1
MEI (Alu)	1.03k	0	845	0	899	1	919	0	889	0
MEI (L1)	138	0	118	0	130	0	123	0	123	0
MEI (SVA)	52	0	44	0	56	0	53	0	44	0
MEI (MT)	5	0	5	0	4	0	4	0	4	0
Inversions	12	0	9	0	10	0	9	0	11	0
Nonsynon	12.2k	139	10.4k	121	10.2k	144	10.2k	116	10.3k	144
Synon	13.8k	78	11.4k	67	11.2k	79	11.2k	59	11.4k	78
Intron	2.06M	7.33k	1.72M	6.12k	1.68M	7.39k	1.68M	5.68k	1.72M	7.20k
UTR	37.2k	168	30.8k	136	30.0k	169	30.0k	129	30.7k	168
Promoter	102k	430	84.3k	332	81.6k	425	82.2k	336	84.0k	430
Insulator	70.9k	248	59.0k	199	57.7k	252	57.7k	189	59.1k	243
Enhancer	354k	1.32k	295k	1.05k	289k	1.34k	288k	1.02k	295k	1.31k
TFBSs	927	4	759	3	748	4	749	3	765	3
Filtered LoF	182	4	152	3	153	4	149	3	151	3
HGMD-DM	20	0	18	0	16	1	18	2	16	0
GWAS	2.00k	0	2.07k	0	1.99k	0	2.08k	0	2.06k	0
ClinVar	28	0	30	1	24	0	29	1	27	1

See Supplementary Table 1 for continental population groupings. CNVs, copy-number variants; HGMD-DM, Human Gene Mutation Database disease mutations; k, thousand; LoF, loss-of-function; M, million; MEI, mobile element insertions.

3. EFFECT OF GENETIC VARIANTS

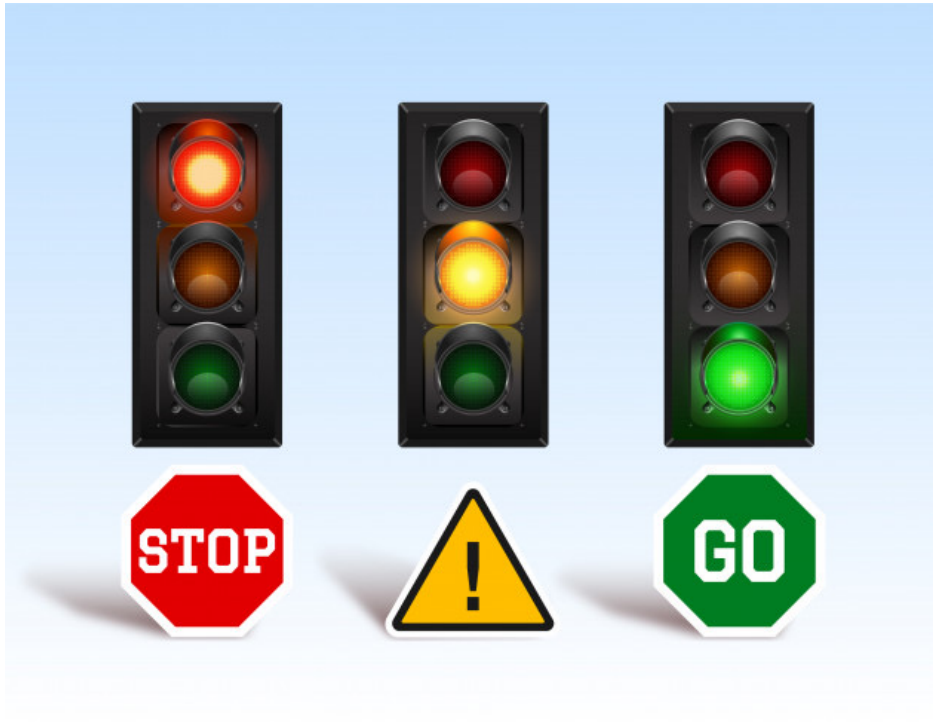
Functional annotation



3. EFFECT OF GENETIC VARIANTS

Functional annotation

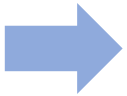
- synonymous
- splice donor
- missense
- frameshift
- intron
- stop gained



3. EFFECT OF GENETIC VARIANTS

Functional annotation

- synonymous
- splice donor
- missense
- frameshift
- intron
- stop gained

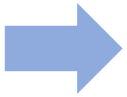


- splice donor
- frameshift
- stop gained
- missense
- synonymous
- intron

3. EFFECT OF GENETIC VARIANTS

Functional annotation

- synonymous
- splice donor
- missense
- frameshift
- intron
- stop gained



- splice donor
- frameshift
- stop gained
- missense
- synonymous
- intron

FUNCTIONAL ANNOTATION OF VARIANTS

➢ **What differentiates one pathogenic variant from a neutral one_**

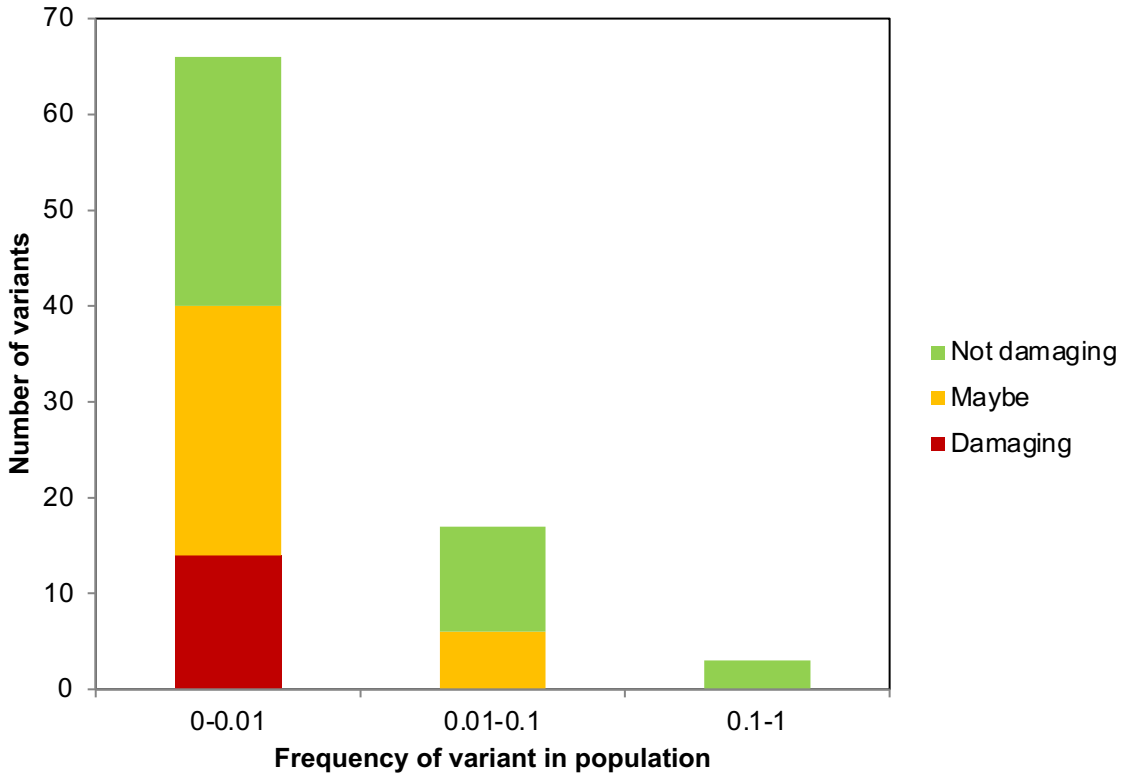
➢ **How can we use these properties to predict pathogeneicity?**

How would you predict the functional and biological consequences of a variant?

3. EFFECT OF GENETIC VARIANTS

Functional annotation

Genetic variant	Position	Change	Consequence	Gene	Freq
A	3:1403401-1403401	A/C	missense	CNTN6	0.003
B	12:49022301-49022301	G/A	missense	KMT2D	-
C	3:25634002-25634002	G/A	missense	TOP2B	4×10^{-6}
D	15:72353105-72353105	C/T	missense	HEXA	5×10^{-5}
E	4:38797314-38797314	C/A	synonymous	TLR1	0.38
F	7:117639961-117639961	C/T	intron	CFTR	6×10^{-5}
G	17:31230383-31230383	G/A	splice_donor	NF1	4×10^{-6}
H	11:63290453-63290453	G/A	stop_gained	SLC22A10	0.434



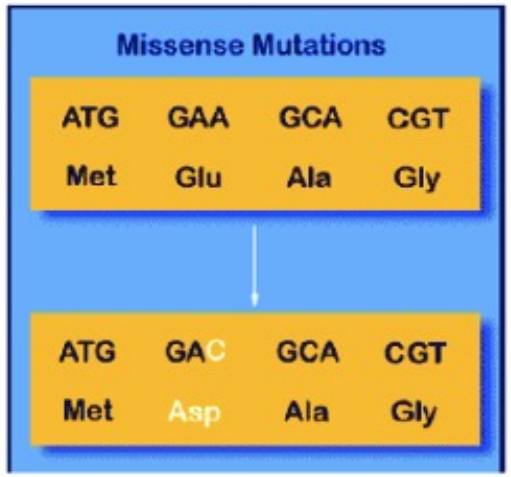
Marth *et al*. (2011) *Genome Biol*. **12**, R84

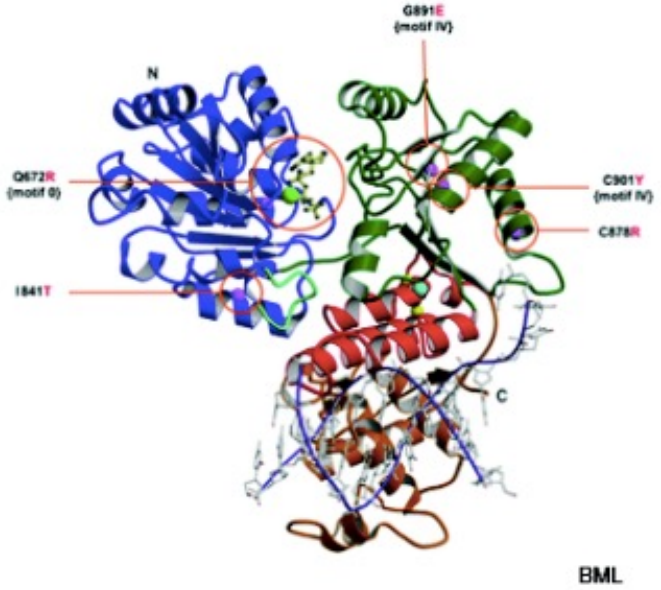
3. EFFECT OF GENETIC VARIANTS

Functional annotation

- splice donor
- frameshift
- stop gained
- missense
- synonymous
- intron

Functional effect prediction of missense changes





Functional annotation

Computational approaches to predict the impact of SNVs

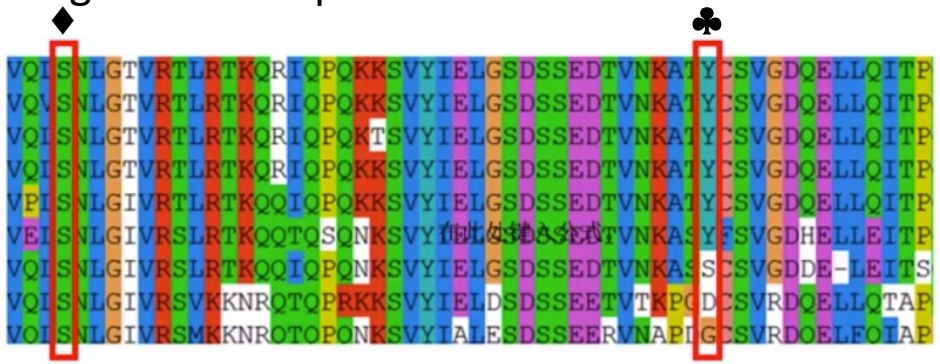
Sorting **I**ntolerant **F**rom **T**olerant substitutions (**SIFT**)

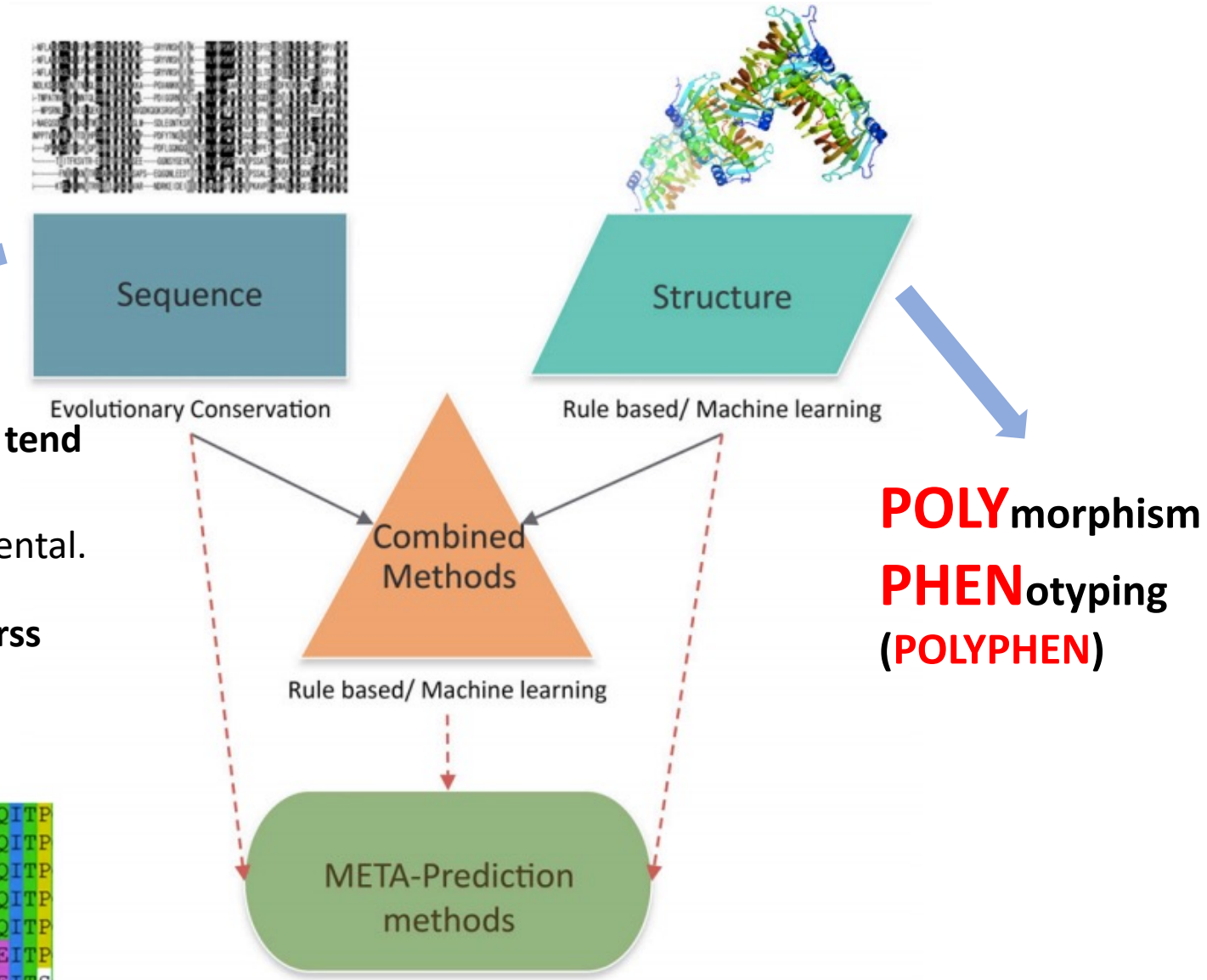
➢ **Funtionally important positions (e.g. active sites) tend to be conserved across species.**

- ♦ Changes at conserved sites tend to be detrimental.

➢ **Some positions show a high level of diversity acorss species.**

- ♣ Changes at these positions tend to be neutral.





POLYmorphism
PHENotyping
(**POLYPHEN**)

Functional annotation

POLYmorphism PHENotyping (POLYPHEN)

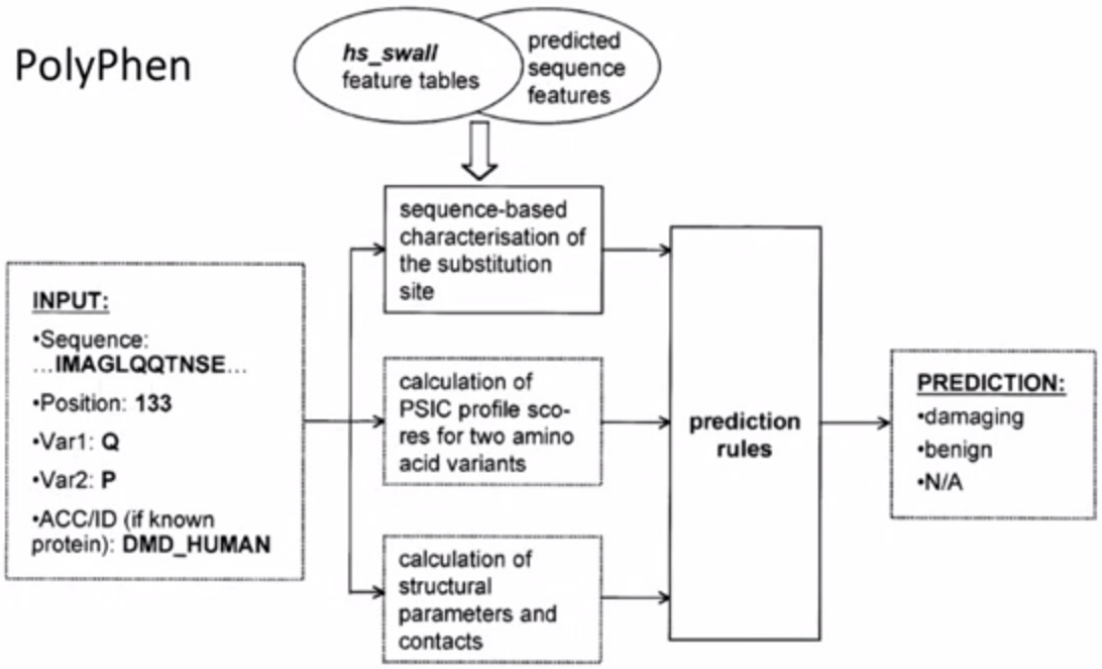
➢ **Predicts the impact of amino acid changes based on multiple sequence alignements and the protein structure.**

➢ **Assumes that:**

1. Amino acid changes at conserved positions have more probabilities of generating functional changes.

2. Changes affecting to active and interaction sites, protein solubility or stability will probably affect to its structure.

3. Changes in the structure of the protein will probably generate functional changes that is probable that originate phenotype changes.



Polyphen value	Qualitative prediction	Website display example
greater than 0.908	"Probably Damaging"	0.95
greater than 0.446 and less than or equal to 0.908	"Possibly Damaging"	0.5
less than or equal to 0.446	"Benign"	0.25
unknown	"Unknown"	unknown

3. EFFECT OF GENETIC VARIANTS

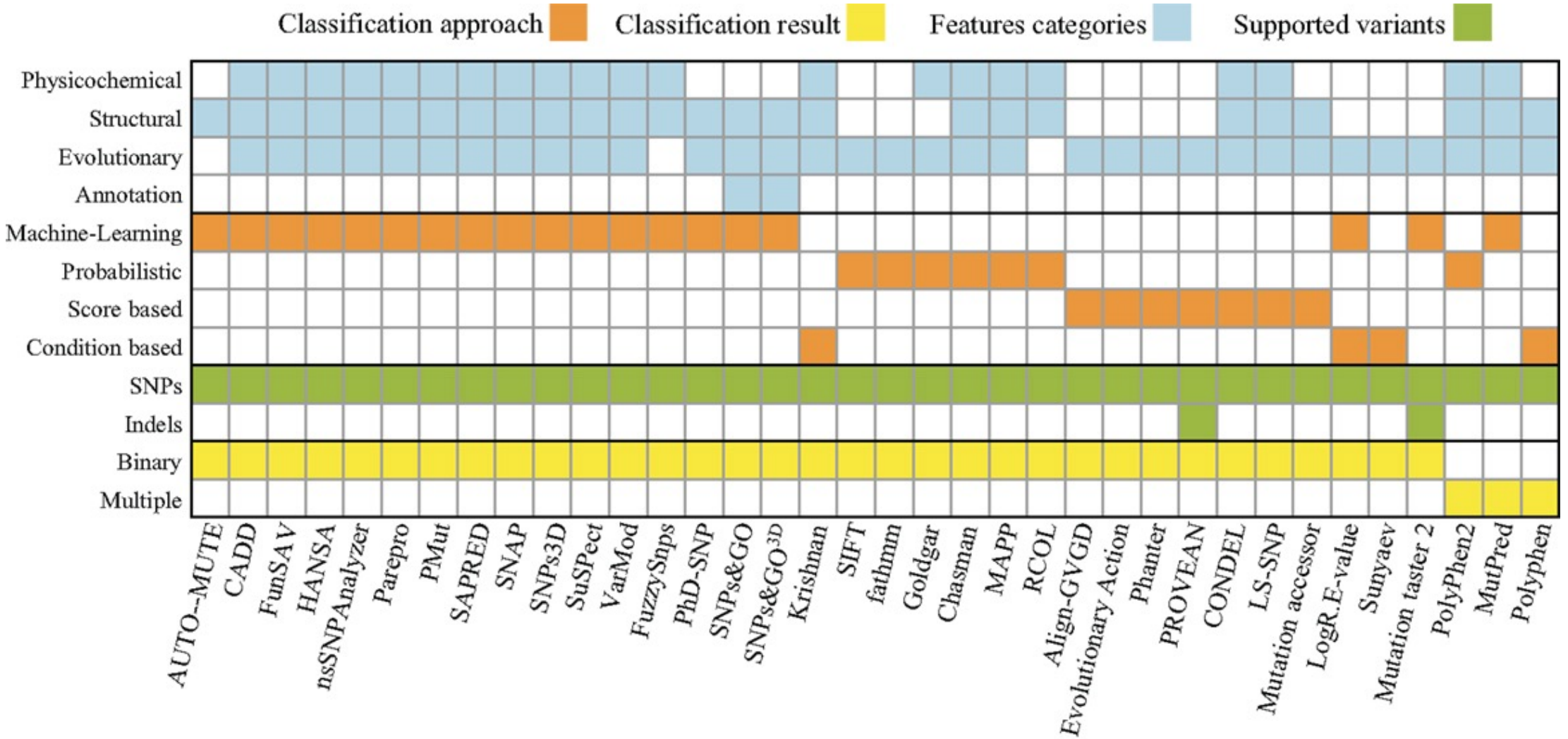
Functional annotation

Figure 1. Wordle of variant impact predictors.

De *VIPdb, a Genetic Variant Impact Predictor database*
Human Mutation 2019.

3. EFFECT OF GENETIC VARIANTS

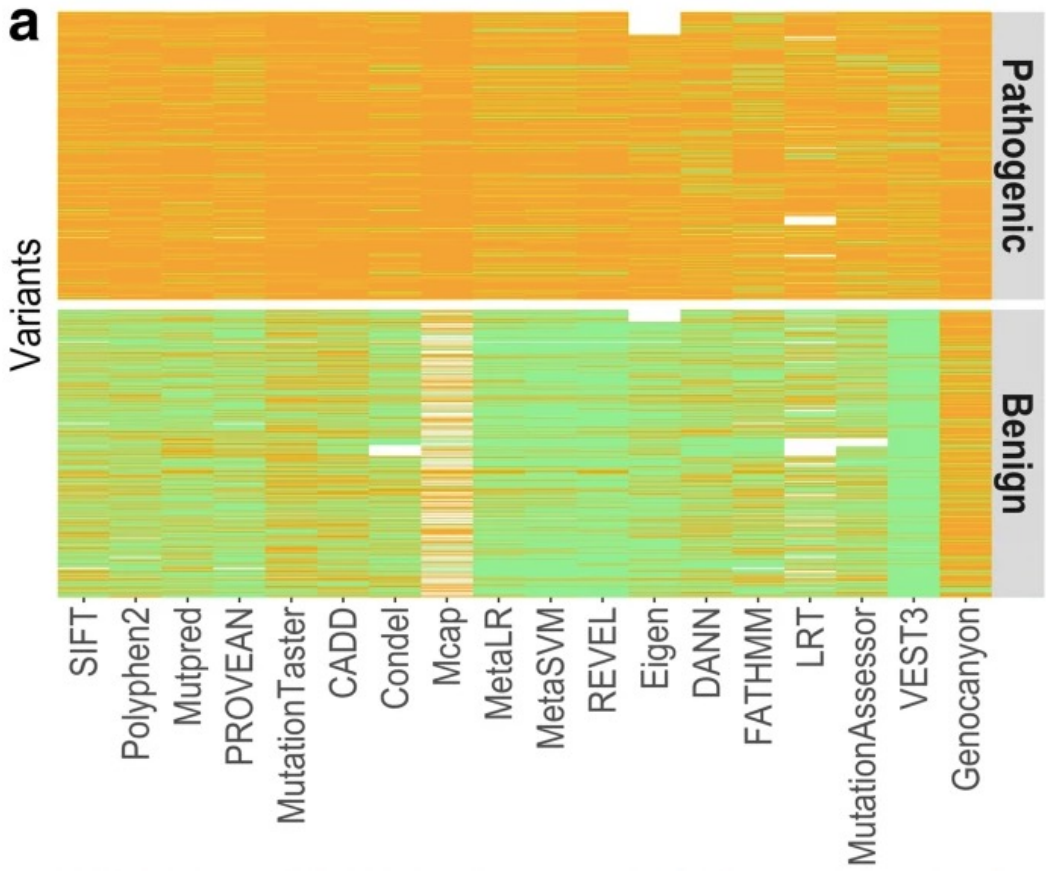
Functional annotation

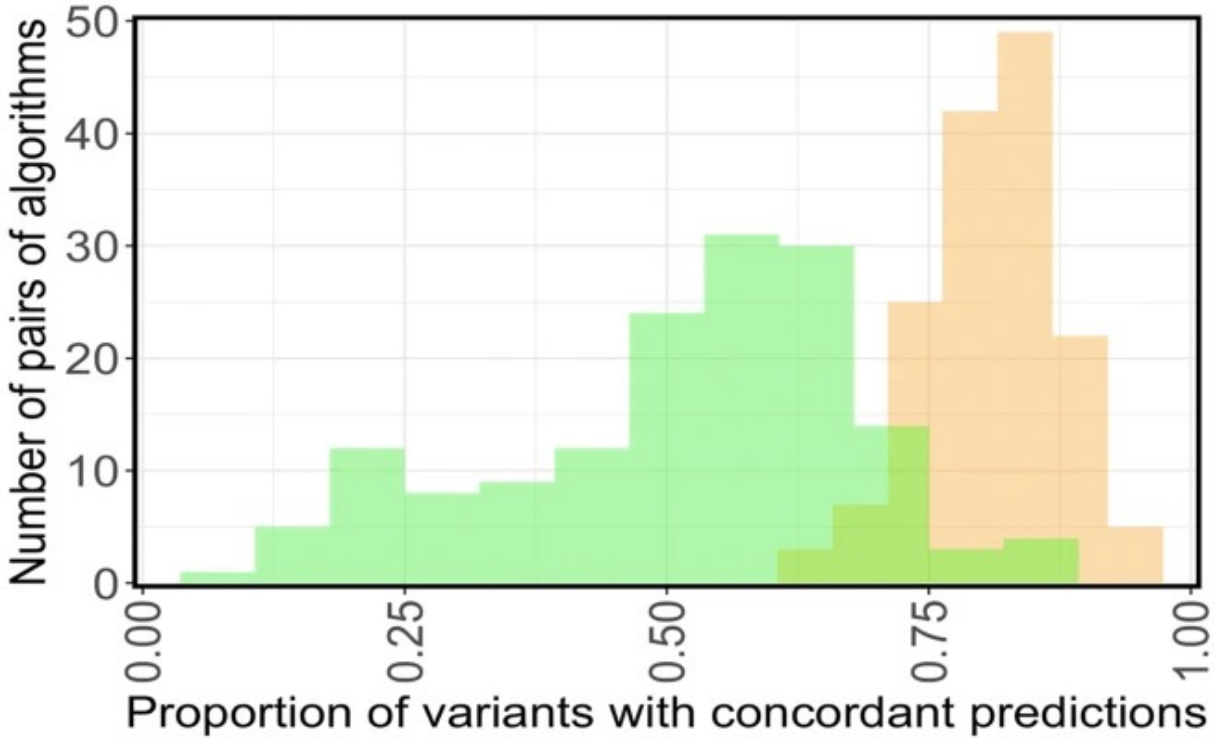


De *Analysis of genetic variation and potential Applications in genome-scale metabolic modeling*
Frontiers in Bioengineering and Biotechnology, 2015

3. EFFECT OF GENETIC VARIANTS

Functional annotation





De *Evalutation of in silico algorithms for use with ACMG/AMP clinical variant interpretion guidelines*
Genome Biology, 2017

3. EFFECT OF GENETIC VARIANTS

Functional annotation

Figure 1. Wordle of variant impact predictors.

3. EFFECT OF GENETIC VARIANTS

Functional annotation

Genetic variant	Position	Change	Consequence	Gene	Freq	SIFT	Polyphen	Mutation Assessor	CADD	Condel
A	3:1403401-1403401	A/C	missense	CNTN6	0.003	0.47	0	Neutral	9.52	Neutral
B	12:49022301-49022301	G/A	missense	KMT2D	-	0	1	Medium	32	0.945
C	3:25634002-25634002	G/A	missense	TOP2B	4×10^{-6}	0.01	0.152	Medium	29.3	0.778
D	15:72353105-72353105	C/T	missense	HEXA	5×10^{-5}	0	1	Damaging	32	0.945
E	4:38797314-38797314	C/A	synonymous	TLR1	0.38	-	-	-	0.29	-
F	7:117639961-117639961	C/T	intron	CFTR	6×10^{-5}	-	-	-	1.08	-
G	17:31230383-31230383	G/A	splice_donor	NF1	4×10^{-6}	-	-	-	34	-
H	11:63290453-63290453	G/A	stop_gained	SLC22A10	0.434	-	-	-	36	-

3. EFFECT OF GENETIC VARIANTS

Gene properties

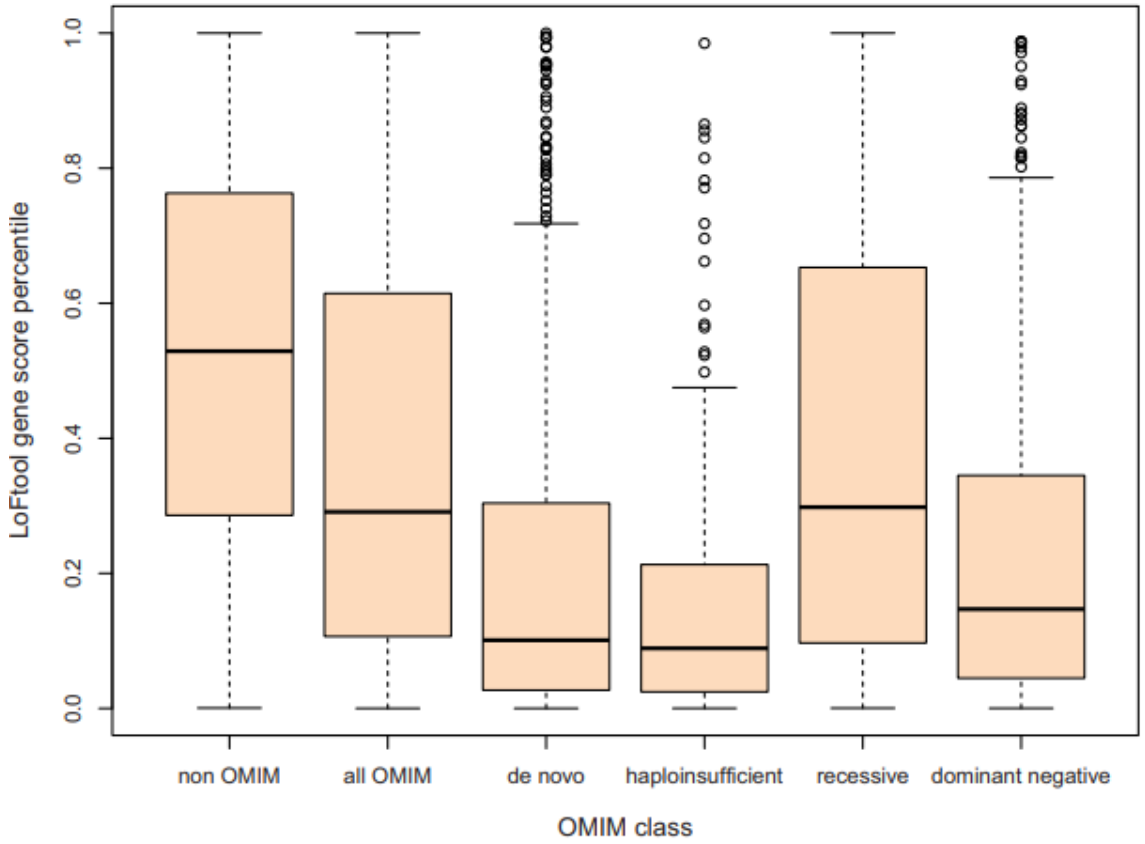
Number of loss-of-function variants per individual genome

Category	Filtered number/individual (CEU)	
	All	Homozygous
nonsense SNP	26.2	5.2
splice SNP	11.2	1.9
frameshift indel	38.2	9.2
large LoF deletion	28.3	6.2
total	103.9	22.3

Genome analysis

LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals

João Fadista[1,2,*], Nikolay Oskolkov[2], Ola Hansson[2] and Leif Groop[2,3]



3. EFFECT OF GENETIC VARIANTS

Gene properties

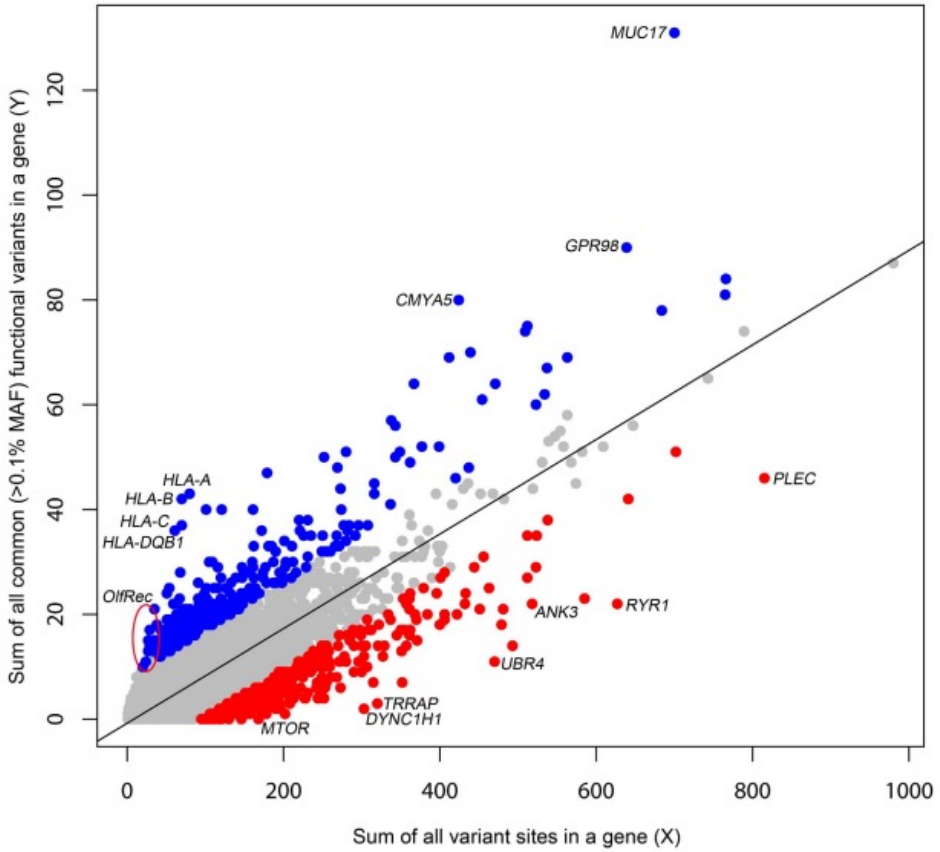
OPEN ACCESS Freely available online

PLOS | GENETICS

Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes

Slavé Petrovski[1,2]*, Quanli Wang[1], Erin L. Heinzen[1,3], Andrew S. Allen[1,4], David B. Goldstein[1]*

RVIS: residual variation intolerance score



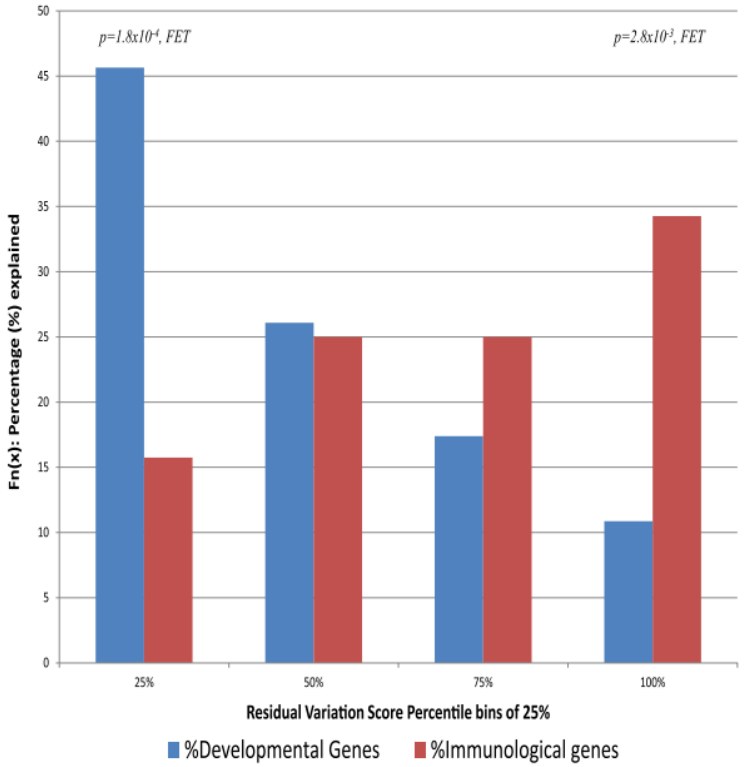


Figure 4. The proportion of genes explained by each of the 25-percentile bins (RVIS) for the human disease networks disorder class with the lowest "Developmental Disorders" and highest "Immunological Disorders" average residual variation intolerance score.
doi:10.1371/journal.pgen.1003709.g004

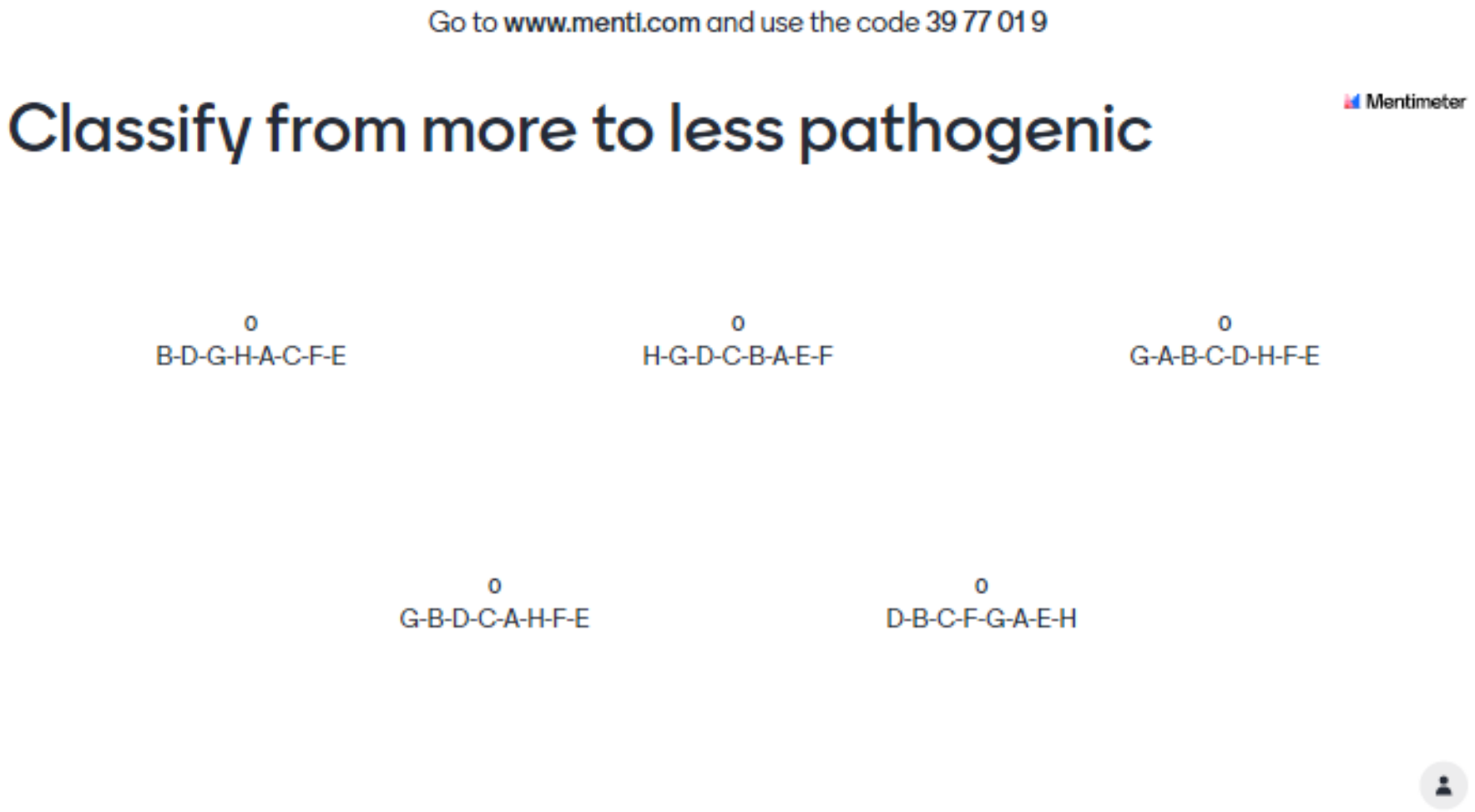
3. EFFECT OF GENETIC VARIANTS

Functional annotation

Genetic variant	Position	Change	Consequence	Gene	Freq	SIFT	Polyphen	Mutation Assessor	CADD	Condel	LoF tol	RVIS
A	3:1403401-1403401	A/C	missense	CNTN6	0.003	0.47	0	Neutral	9.52	Neutral	0.974	-0.92 (9.81%)
B	12:49022301-49022301	G/A	missense	KMT2D	-	0	1	Medium	32	0.945	-	-5.29 (0.06%)
C	3:25634002-25634002	G/A	missense	TOP2B	4×10^{-6}	0.01	0.152	Medium	29.3	0.778	0.82	-0.15 (42.28%)
D	15:72353105-72353105	C/T	missense	HEXA	5×10^{-5}	0	1	Damaging	32	0.945	0.17	-0.33 (30.7%)
E	4:38797314-38797314	C/A	synonymous	TLR1	0.38	-	-	-	0.29	-	0.963	1.32 (94.07%)
F	7:117639961-117639961	C/T	intron	CFTR	6×10^{-5}	-	-	-	1.08	-	0.0235	-0.51 (21.73%)
G	17:31230383-31230383	G/A	splice_donor	NF1	4×10^{-6}	-	-	-	34	-	0.116	-3.09 (0.47%)
H	11:63290453-63290453	G/A	stop_gained	SLC22A10	0.434	-	-	-	36	-	0.7	1.85 (97.12%)

3. EFFECT OF GENETIC VARIANTS

Functional annotation



3. EFFECT OF GENETIC VARIANTS

Functional annotation

Go to **www.menti.com** and use the code 39 77 01 9

Classify from more to less pathogenic

Mentimeter

0
B-D-G-H-A-C-F-E

0
H-G-D-C-B-A-E-F

0
G-A-B-C-D-H-F-E

0
G-B-D-C-A-H-F-E

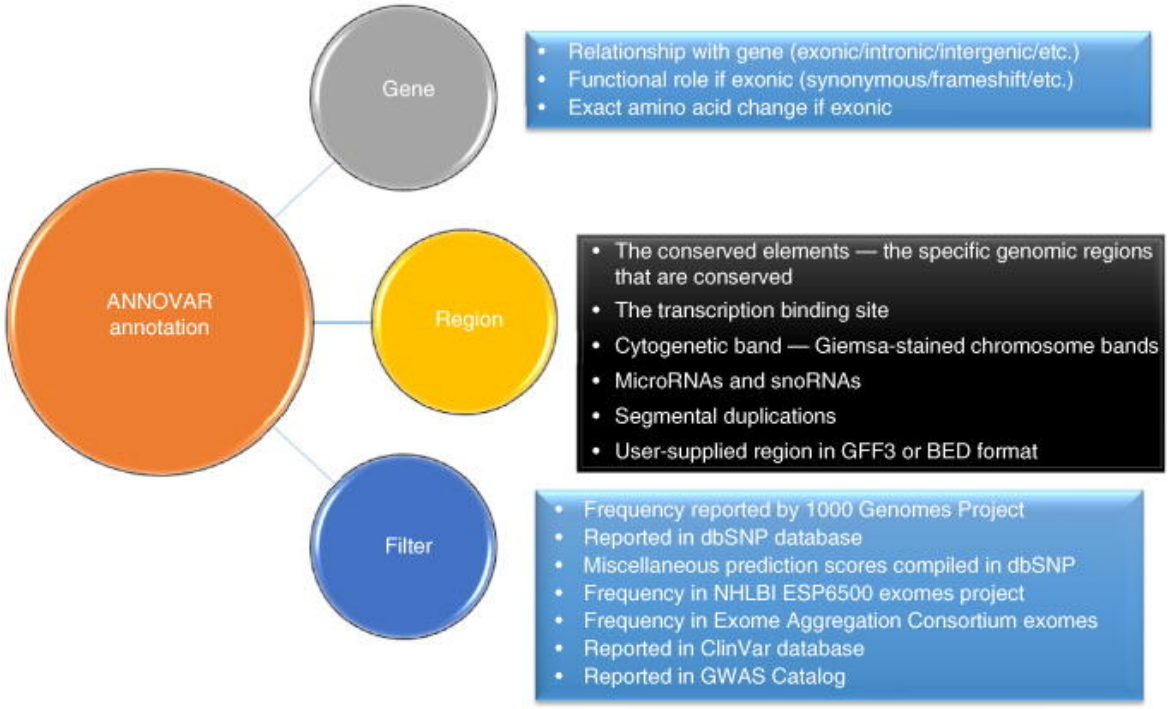
0
D-B-C-F-G-A-E-H

3. EFFECT OF GENETIC VARIANTS

Functional annotation

TOOLS FOR GENETIC VARIANTS FUNCTIONAL ANNOTATION

➢ ANNOVAR



De *Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR*
Nature Protocols, 2015

➢ ENSEMBL VARIANT EFFECT PREDICTOR (VEP)

Ensembl Variant Effect Predictor (VEP)

Simply input the coordinates of your variants and the nucleotide changes to find out the:

- **Genes** and **Transcripts** affected by the variants
- **Location** of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regulatory regions)
- **Consequence** of your variants on the protein sequence (e.g. stop gained, missense, stop lost, frameshift)
- **Known variants** that match yours, and associated minor allele frequencies from the **1000 Genomes Project**
- **SIFT** and **PolyPhen-2** scores for changes to protein sequence
- ... And more! See data types, versions.

Web interface

- Point-and-click interface
- Suits smaller volumes of data

Documentation

Launch VeP

Command line tool

- More options and flexibility
- For large volumes of data

Documentation

Clone from GitHub

Download (zip)

Pull Docker image from DockerHub

3. EFFECT OF GENETIC VARIANTS

Functional annotation

- G is the more pathogenic variant…….but is it really pathogenic?