Omics in human diseases

Index

- Omics data and Biological databases
- NGS methods
- **NGS data analysis**
- Prediction and interpretation of pathogenic variants
- Protein-protein interaction networks

Course organization 2022/2023

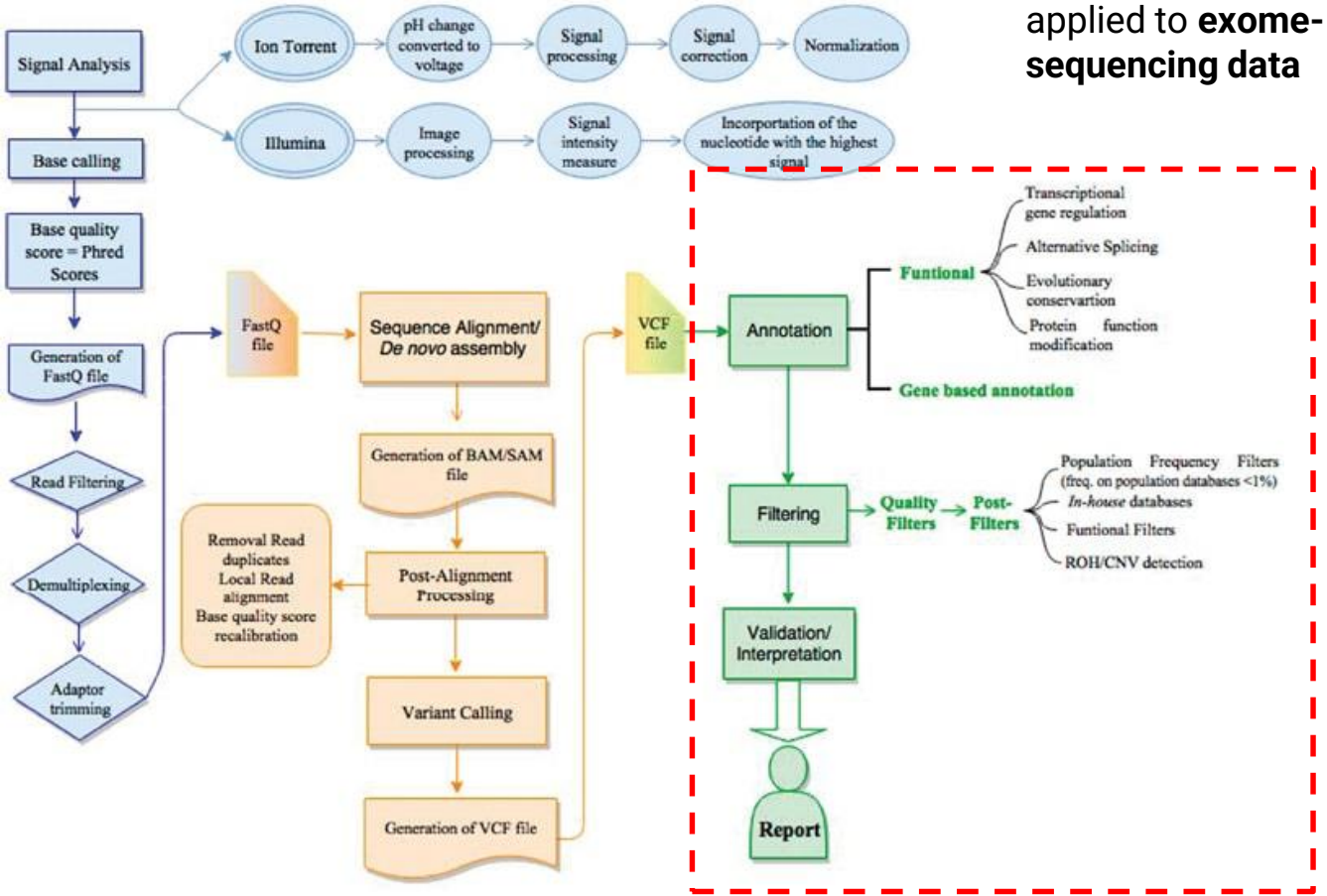
Frontal lecture/ guided practical activity

How to pass the exam: multiple choice quiz (50%) + results from practical activities (50%) + Bonus points, e.g. summary of previous lecture (up to 10%)

Mail: **emanuela.leonardi@unipd.it**

NGS analysis workflow

The same data analysis tools used for WGS can be applied to **exome-sequencing data**



BCF tools

- BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF

- All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed

- Several commands can thus be combined with Unix pipes.
    - **annotate** .. edit VCF files, add or remove annotations
    - **concat** .. concatenate VCF/BCF files from the same set of samples
    - **convert** .. convert VCF/BCF to other formats and back
    - **filter** .. filter VCF/BCF files using fixed thresholds
    - **gtcheck** .. check sample concordance, detect sample swaps and contamination
    - **head** .. view VCF/BCF file headers
    - **isec** .. intersections of VCF/BCF files
    - **merge** .. merge VCF/BCF files files from non-overlapping sample sets
    - **plugin** .. run user-defined plugin
    - **query** .. transform VCF/BCF into user-defined formats
    - **stats** .. produce VCF/BCF stats (former vcfcheck)
    - ……

Standard pipelines for NGS analysis

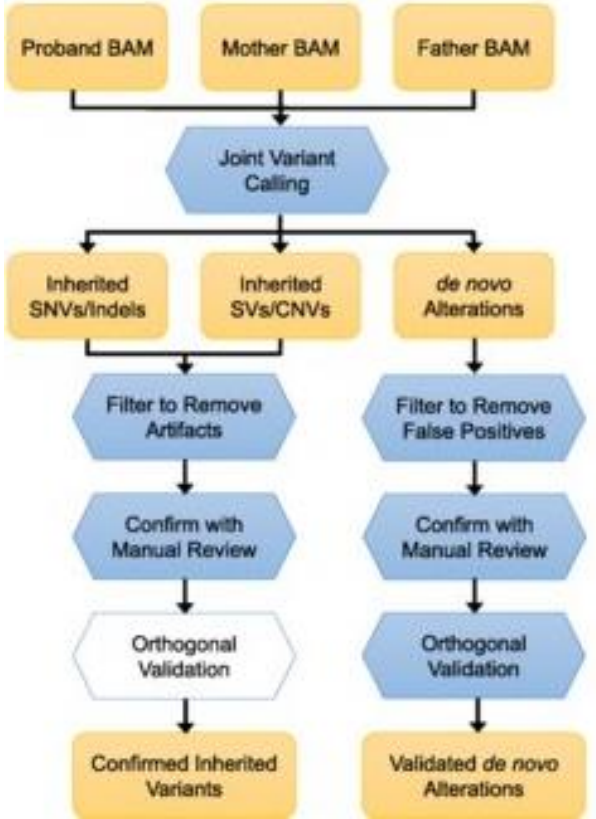
Individual versus joint variant calling

Individual

- Perform variant calling in each sample
- Merge VCF files with **BCFtools**
- The VCF contains entries only for positions that are variants in a particular sample

Joint

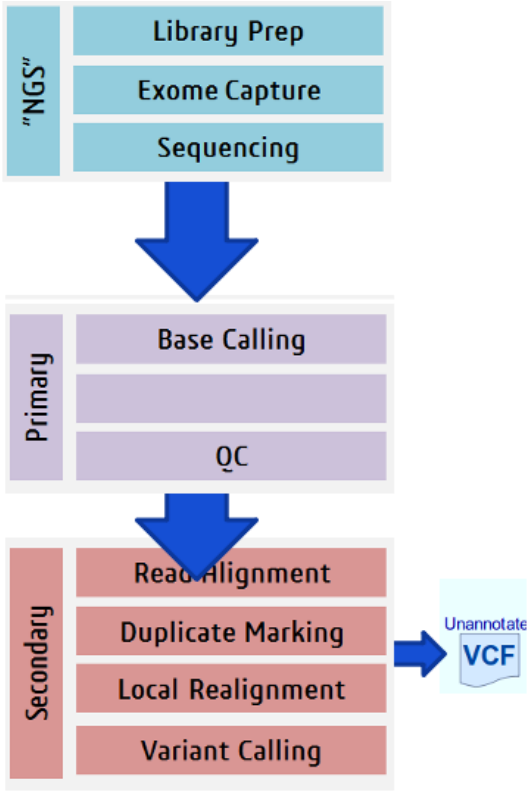
- Perform variant calling in all samples simultaneously
- Called genotypes for every sample at **all variant positions**
- Differentiates among match with reference sequence and low coverage positions
- For trios, direct inference of variant **phase** (cis or trans)
- Increase sensitivity of variant calling in low-coverage regions



Somatic variant calling

- 10% of cancer patients harbor **germline** predisposition variants
- Clinical tumor sequencing aims to search **somatic variants**
- Sequencing DNA from a tumor sample and a matched control sample from the same patient
- Specific **variant callers** for somatic variants (MuTect2, Strelka, VarScan2) use simultaneous alignment from tumor and normal samples
- **Challenges**: Tumor purity, low frequency of somatic variants, type of specifment (formalin fixed, paraffin embedded)
- **An Ensemble approach** that combines the results of two or more callers offer the best balance of sensitivity and specificity

is this a pathogenic variant?

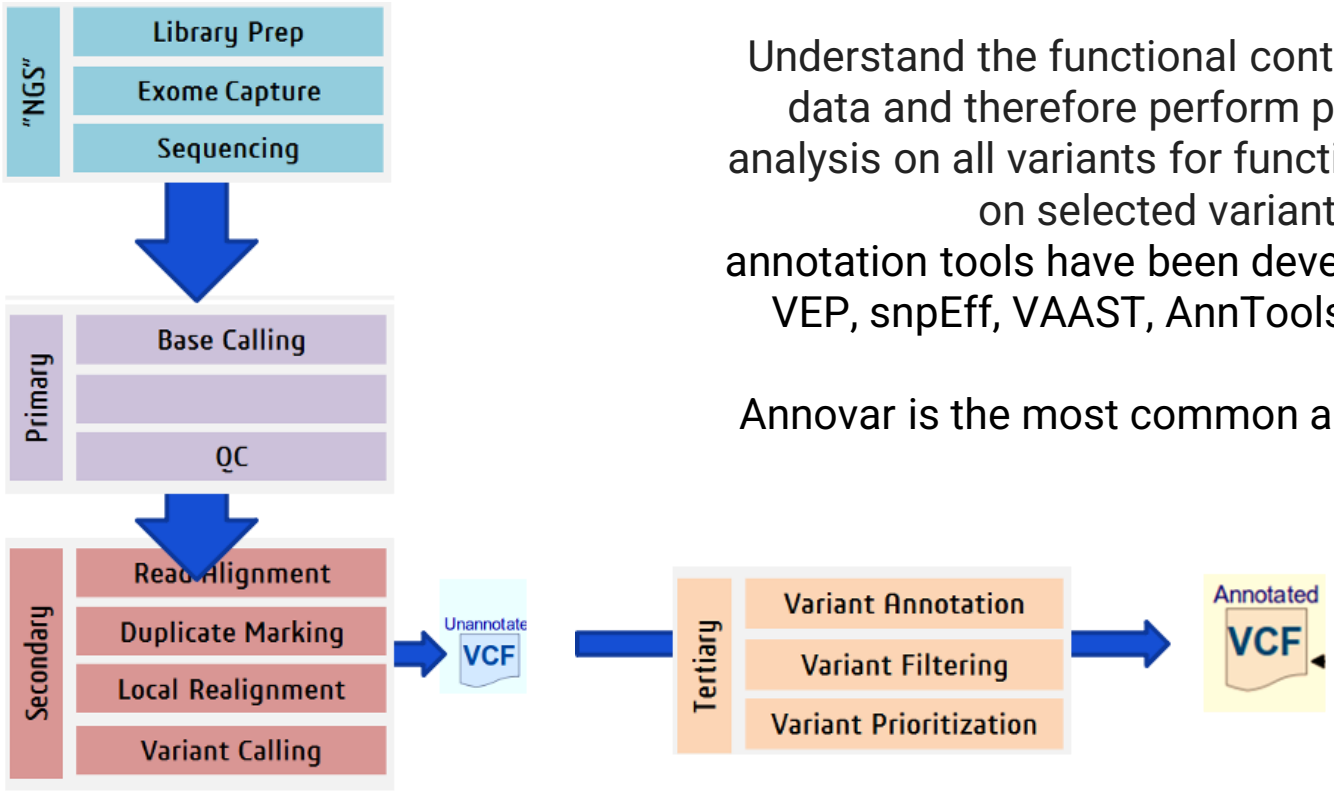



```
##FILTER=<ID=HETEROPLASMY,Description="Called heterozygous genotype on mitochondrial contig">
##FORMAT=<ID=GP,Number=G,Type=Float,Description="Estimated Genotype Probability">
##FILTER=<ID=IMP,Description="Set if true: IMP==1">
##FILTER=<ID=BOOSTED,Description="Set if true: BOOSTED==1">
##FILTER=<ID=LOWDP,Description="Set if GQ>20 and 10<=DP<=20">
##FILTER=<ID=LOWQ,Description="Set if GQ<=20 or DP<10">
##FILTER=<ID=NOTVALIDATED,Description="Set if variant falls outside of analytic range">
##FORMAT=<ID=GL,Number=G,Type=Float,Description="Genotype likelihoods">
##FORMAT=<ID=VAR_TYPE,Number=.,Type=String,Description="Variant type: SNV, INSERTION, DELETION, SUBSTITUTION, MNV, COMPLEX">
##FORMAT=<ID=VAR_CONTEXT,Number=.,Type=String,Description="Variant genomic context: STR-expansion, STR-contraction, STR-proximal">
##FORMAT=<ID=STR_MAX_LEN,Number=1,Type=Integer,Description="Maximum observed STR sequence length">
##FORMAT=<ID=STR_PERIOD,Number=1,Type=Integer,Description="Repetition period for STR variants">
##FORMAT=<ID=STR_TIMES,Number=1,Type=Float,Description="Number of repetition for STR variants">
##pipeline=helix-v2.6.1
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	PC-TA537BFRZC6B332GAGD'
chr1	55039878	.	A	ACTG	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47805173	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47793959	.	C	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr13	32319070	.	T	A,TA	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11113686	.	A	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	21011802	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	5977709	.	T	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43094795	.	A	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11102787	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	6003794	.	T	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr3	37028782	.	AG	CC	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47798826	.	A	AAC	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	5987451	.	CTT	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr13	32340378	.	AGCAAG	ATGCTG	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	21038086	.	C	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11123663	.	C	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr1	55039330	.	G	GGAGGA	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	21010226	.	CTCA	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	6003018	.	A	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43124094	.	G	GCCT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43124097	.	TT	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43045679	.	C	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr13	32398769	.	A	AT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	5973402	.	CTGA	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47806206	.	A	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11128084	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47806452	.	G	GGGG	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47801152	.	TTGG	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11120166	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11111506	.	T	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47805601	.	A	AT,ATT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47805601	.	AT	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11128142	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43053463	.	C	CACA	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47806751	.	CTT	C,CT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43125260	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43124135	.	C	CAT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43124745	.	GTTTTT	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43044346	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47806983	.	A	AGTTC	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11133511	.	TTA	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	21038086	.	C	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11113534	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	21012365	.	A	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr7	5997399	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11120188	.	T	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr19	11116388	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47805638	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr1	55057514	.	G	A	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47793092	.	T	C	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47793601	.	C	T	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr2	47806206	.	A	G	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV

BioComputing

is this a pathogenic variant?



Understand the functional content within the data and therefore perform prioritization analysis on all variants for functional follow-up on selected variants
annotation tools have been developed, such as VEP, snpEff, VAAST, AnnTools, ANNOVAR

Annovar is the most common annotation tool

ANNOVAR (ANNOtate VARiation)

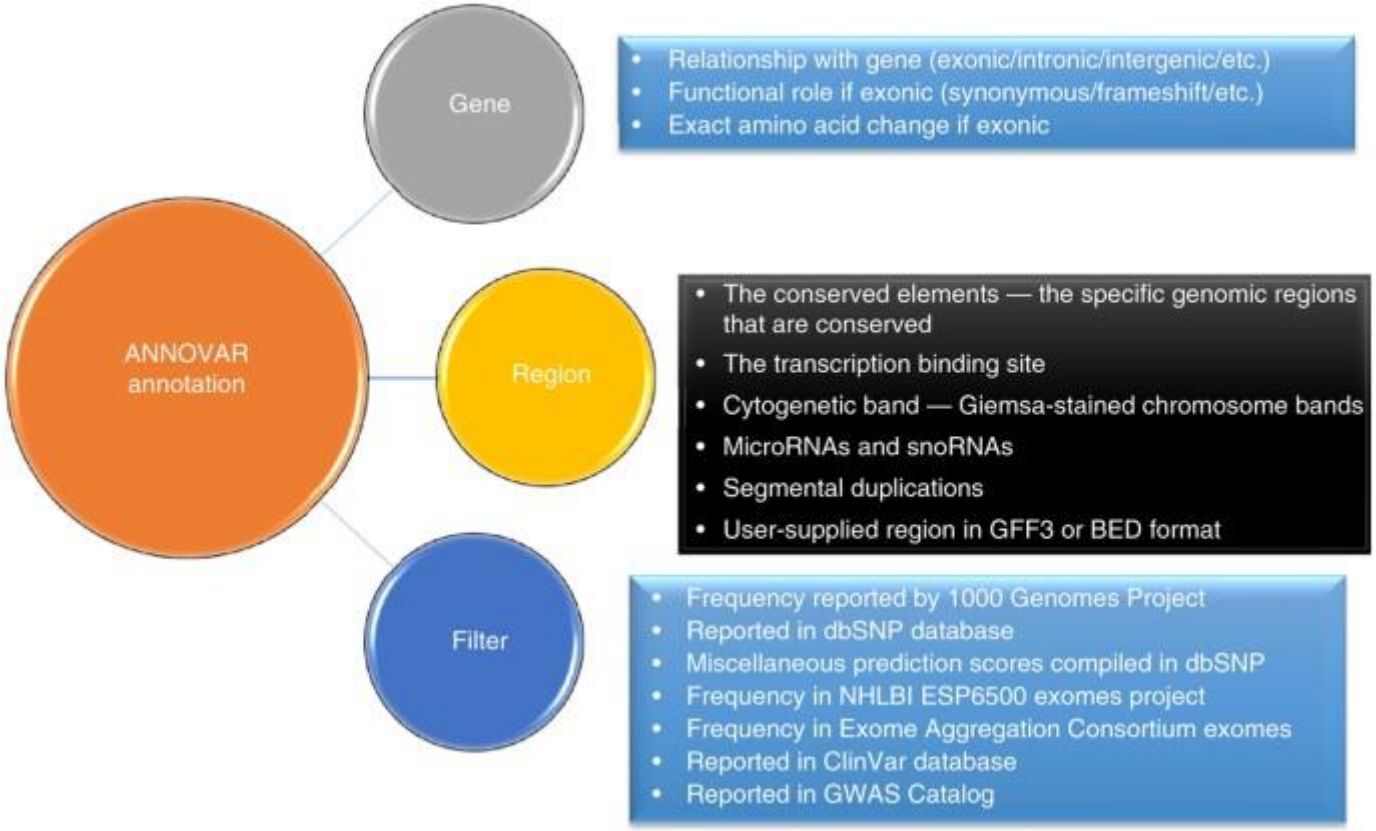
http://annovar.openbioinformatics.org/

- It is a command-line tool **written in the Perl** programming language, which can be executed on a variety of operating systems with a Perl interpreter installed.
- **VCF files** as input
- Outputs an annotated variant file in several different formats (such as annotated VCF file, **tab-delimited text file** or comma-delimited text file) with annotations for each variant in the input file

- Download ANNOVAR package with **databases** for gene- and region-based annotations and filtering
- **User-contributed datasets**
  - regSNP intron: machine learning algorithm to prioritize the disease-causing probability of intronic SNVs
  - LoFtool score: gene loss-of-function score percentiles
  - RVIS-ESV score: RVIS score measures genetic intolerance of genes to functional mutations
  - SPIDEX: Machine-learning prediction on how genetic variants affect RNA splicing.

ANNOVAR (ANNOtate VARiation)

http://annovar.openbioinformatics.org/



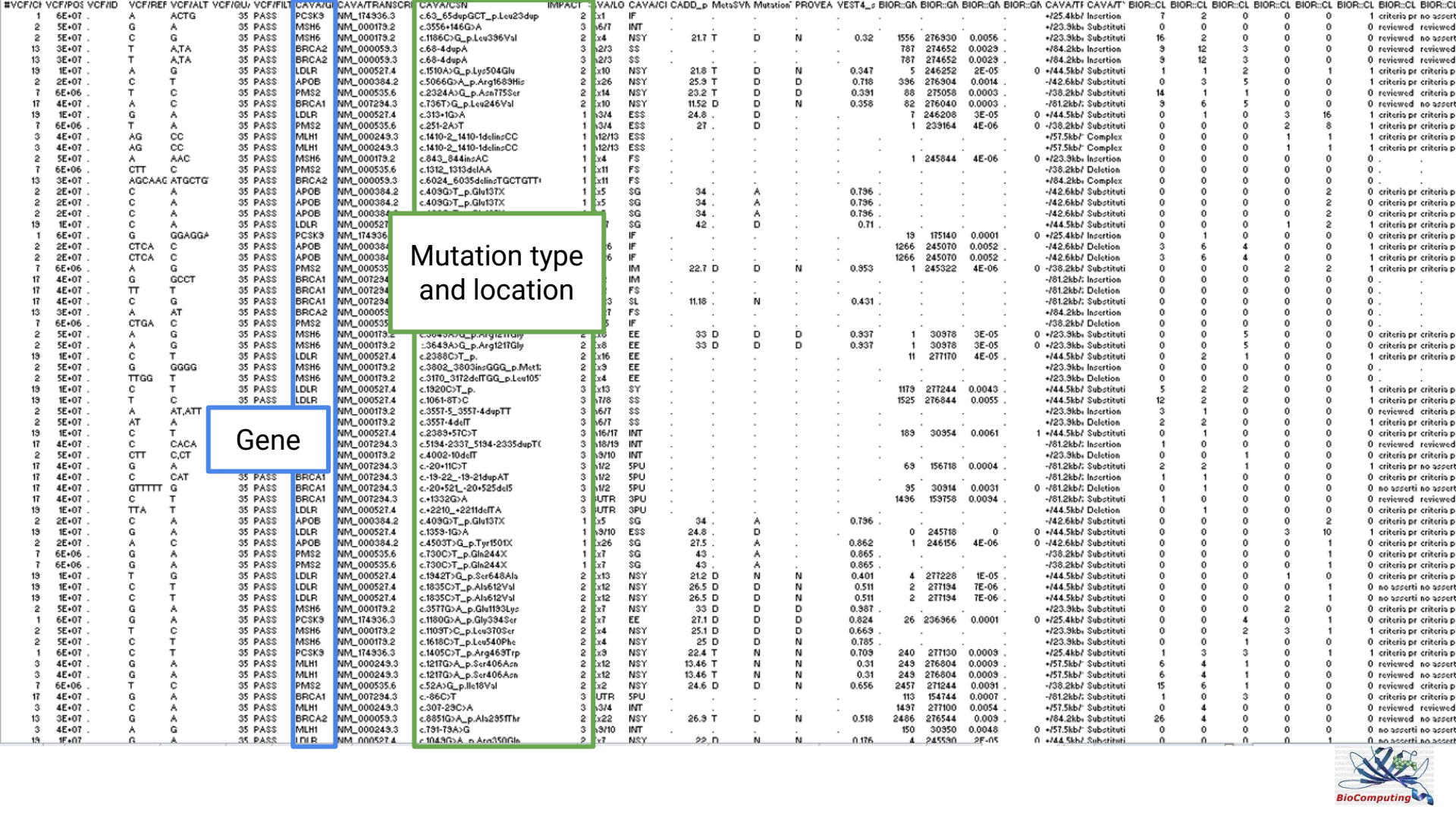
And various types of variant-deleteriousness prediction scores

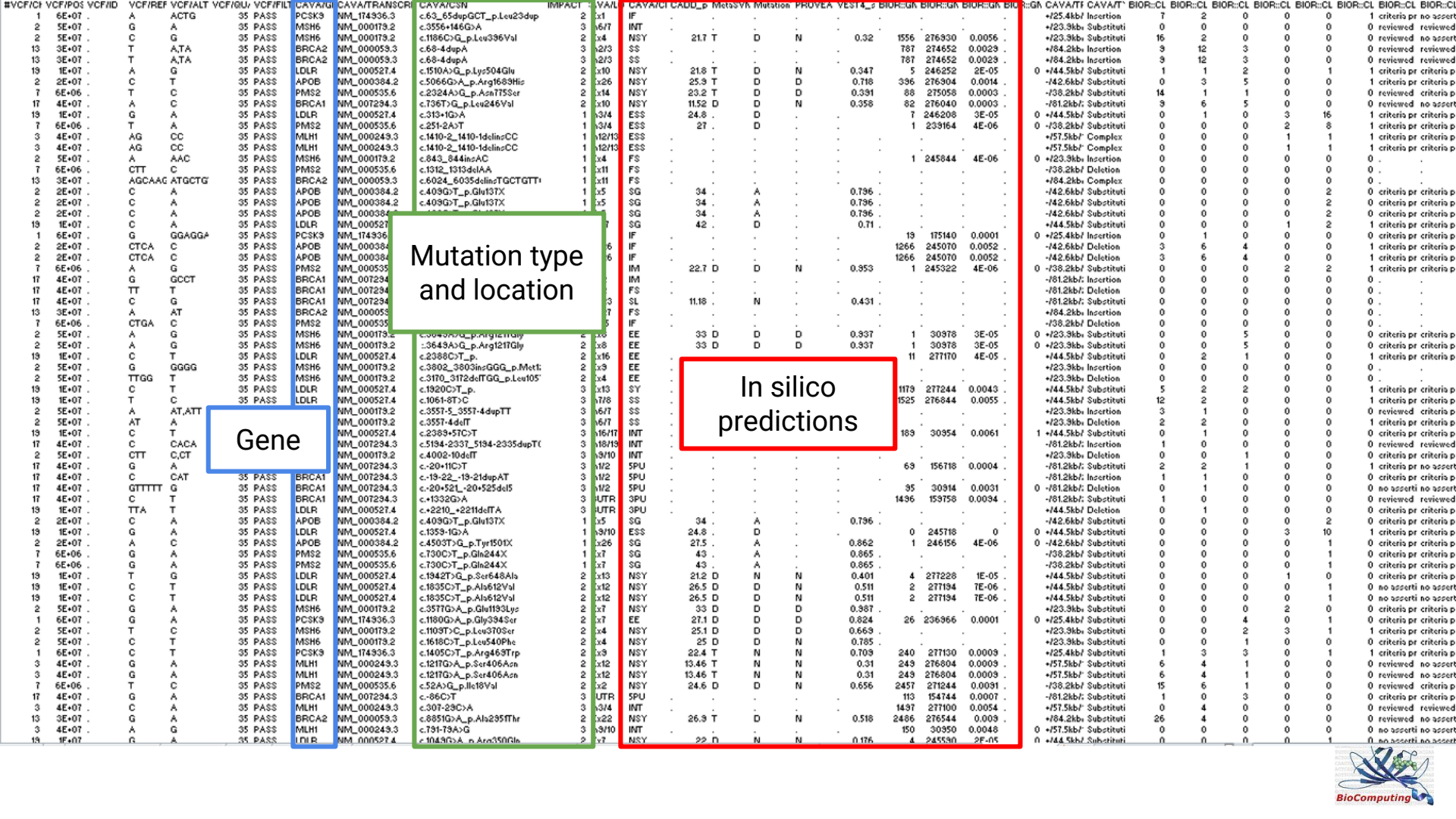
#VCF/CI	VCF/POS	VCF/ID	VCF/REF	VCF/ALT	VCF/QU/	VCF/FILT	CAVA/GI	CAVA/TRANSCRII	CAVA/CSN	IMPACT	:AVA/LO	CAVA/CI	CADD_p	MetaSVN	Mutation	PROVEA	VEST4_s	BIOR::GN	BIOR::GN	BIOR::GN	BIOR::GN	CAVA/TF	CAVA/T'	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL
1	6E+07	.	A	ACTG	35	PASS	PCSK9	NM_174936.3	c.63_65dupGCT_p.Leu23dup	2	Ex1	IF	.	.	.	.	.	.	.	.	.	+/25.4kb/	Insertion	7	2	0	0	0	1	criteria pr	no assert
2	5E+07	.	G	A	35	PASS	MSH6	NM_000179.2	c.3556+146G>A	3	In6/7	INT	.	.	.	.	.	.	.	.	.	+/23.9kb	Substituti	0	0	0	0	0	0	reviewed	reviewed
2	5E+07	.	C	G	35	PASS	MSH6	NM_000179.2	c.1186C>G_p.Leu396Val	2	Ex4	NSY	21.7	T	D	N	0.32	1556	276930	0.0056	.	+/23.9kb	Substituti	16	2	0	0	0	0	reviewed	no assert
13	3E+07	.	T	A,TA	35	PASS	BRCA2	NM_000059.3	c.68-4dupA	3	In2/3	SS	.	.	.	.	.	787	274652	0.0029	.	+/84.2kb	Insertion	9	12	3	0	0	0	reviewed	reviewed
13	3E+07	.	T	A,TA	35	PASS	BRCA2	NM_000059.3	c.68-4dupA	3	In2/3	SS	.	.	.	.	.	787	274652	0.0029	.	+/84.2kb	Insertion	9	12	3	0	0	0	reviewed	reviewed
19	1E+07	.	A	G	35	PASS	LDLR	NM_000527.4	c.1510A>G_p.Lys504Glu	2	Ex10	NSY	21.8	T	D	N	0.347	5	246252	2E-05	0	+/44.5kb/	Substituti	1	1	2	0	1	1	criteria pr	criteria p
2	2E+07	.	C	T	35	PASS	APOB	NM_000384.2	c.5066G>A_p.Arg1689His	2	Ex26	NSY	25.9	T	D	D	0.718	396	276904	0.0014	.	-/42.6kb/	Substituti	0	3	5	0	0	1	criteria pr	criteria p
7	6E+06	.	T	C	35	PASS	PMS2	NM_000535.6	c.2324A>G_p.Asn775Ser	2	Ex14	NSY	23.2	T	D	D	0.391	88	275058	0.0003	.	-/38.2kb/	Substituti	14	1	1	0	0	0	reviewed	criteria p
17	4E+07	.	A	C	35	PASS	BRCA1	NM_007294.3	c.736T>G_p.Leu246Val	2	Ex10	NSY	11.52	D	D	N	0.358	82	276040	0.0003	.	-/81.2kb/:	Substituti	9	6	5	0	0	0	reviewed	no assert
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.313+1G>A	1	In3/4	ESS	24.8	.	D	.	.	7	246208	3E-05	0	+/44.5kb/	Substituti	0	1	0	3	16	1	criteria pr	criteria p
7	6E+06	.	T	A	35	PASS	PMS2	NM_000535.6	c.251-2A>T	1	In3/4	ESS	27	.	D	.	.	1	239164	4E-06	0	-/38.2kb/	Substituti	0	0	0	2	8	1	criteria pr	criteria p
3	4E+07	.	AG	CC	35	PASS	MLH1	NM_000249.3	c.1410-2_1410-1delinsCC	1	In12/13	ESS	.	.	.	.	.	.	.	.	.	+/57.5kb/	Complex	0	0	0	1	1	1	criteria pr	criteria p
3	4E+07	.	AG	CC	35	PASS	MLH1	NM_000249.3	c.1410-2_1410-1delinsCC	1	In12/13	ESS	.	.	.	.	.	.	.	.	.	+/57.5kb/	Complex	0	0	0	1	1	1	criteria pr	criteria p
2	5E+07	.	A	AAC	35	PASS	MSH6	NM_000179.2	c.843_844insAC	1	Ex4	FS	.	.	.	.	.	1	245844	4E-06	0	+/23.9kb	Insertion	0	0	0	0	0	0	.	.
7	6E+06	.	CTT	C	35	PASS	PMS2	NM_000535.6	c.1312_1313delAA	1	Ex11	FS	.	.	.	.	.	.	.	.	.	-/38.2kb/	Deletion	0	0	0	0	0	0	.	.
13	3E+07	.	AGCAAG	ATGCTG	35	PASS	BRCA2	NM_000059.3	c.6024_6035delinsTGCTGTT	1	Ex11	FS	.	.	.	.	.	.	.	.	.	+/84.2kb	Complex	0	0	0	0	0	0	.	.
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
19	1E+07	.	C	A	35	PASS	LDLR	NM_000527.4	c.2546C>A_p.Ser849X	1	Ex17	SG	42	.	D	.	0.71	.	.	.	.	+/44.5kb/	Substituti	0	0	0	1	2	1	criteria pr	criteria p
1	6E+07	.	G	GGAGGA	35	PASS	PCSK9	NM_174936.3	c.101_106dup6_p.Glu34_Asp3	2	Ex1	IF	.	.	.	.	.	19	175140	0.0001	0	+/25.4kb/	Insertion	0	1	0	0	0	0	criteria pr	criteria p
2	2E+07	.	CTCA	C	35	PASS	APOB	NM_000384.2	c.6639_6641delTGA_p.Asp2	2	Ex26	IF	.	.	.	.	.	1266	245070	0.0052	.	-/42.6kb/	Deletion	3	6	4	0	0	1	criteria pr	criteria p
2	2E+07	.	CTCA	C	35	PASS	APOB	NM_000384.2	c.6639_6641delTGA_p.Asp2	2	Ex26	IF	.	.	.	.	.	1266	245070	0.0052	.	-/42.6kb/	Deletion	3	6	4	0	0	1	criteria pr	criteria p
7	6E+06	.	A	G	35	PASS	PMS2	NM_000535.6	c.2T>C_p.Met1?	2	Ex1	IM	22.7	D	D	N	0.953	1	245322	4E-06	0	-/38.2kb/	Substituti	0	0	0	2	2	1	criteria pr	criteria p
17	4E+07	.	G	GCCT	35	PASS	BRCA1	NM_007294.3	c.2_3insAGG_p.Met1?	2	Ex2	IM	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	0	0	0	0	0	0	.	.
17	4E+07	.	TT	T	35	PASS	BRCA1	NM_007294.3	c.1delA_p.Met1?	1	Ex2	FS	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Deletion	0	0	0	0	0	0	.	.
17	4E+07	.	C	G	35	PASS	BRCA1	NM_007294.3	c.5591G>C_p.X1864SerextX3	2	Ex23	SL	11.18	.	N	.	0.431	.	.	.	.	-/81.2kb/:	Substituti	0	0	0	0	0	0	.	.
13	3E+07	.	A	AT	35	PASS	BRCA2	NM_000059.3	c.10256_10257insT	1	Ex27	FS	.	.	.	.	.	.	.	.	.	+/84.2kb	Insertion	0	0	0	0	0	0	.	.
7	6E+06	.	CTGA	C	35	PASS	PMS2	NM_000535.6	c.2583_2585delGAA_p.Gln86	2	Ex15	IF	.	.	.	.	.	.	.	.	.	-/38.2kb/	Deletion	0	0	0	0	0	0	.	.
2	5E+07	.	A	G	35	PASS	MSH6	NM_000179.2	c.3649A>G_p.Arg1217Gly	2	Ex8	EE	33	D	D	D	0.937	1	30978	3E-05	0	+/23.9kb	Substituti	0	0	5	0	0	0	criteria pr	criteria p
2	5E+07	.	A	G	35	PASS	MSH6	NM_000179.2	:.3649A>G_p.Arg1217Gly	2	Ex8	EE	33	D	D	D	0.937	1	30978	3E-05	0	+/23.9kb	Substituti	0	0	5	0	0	0	criteria pr	criteria p
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.2388C>T_p.	2	Ex16	EE	.	.	.	.	.	11	277170	4E-05	.	+/44.5kb/	Substituti	0	2	1	0	0	1	criteria pr	criteria p
2	5E+07	.	G	GGGG	35	PASS	MSH6	NM_000179.2	c.3802_3803insGGG_p.Met1	2	Ex9	EE	.	.	.	.	.	.	.	.	.	+/23.9kb	Insertion	0	0	0	0	0	0	.	.
2	5E+07	.	TTGG	T	35	PASS	MSH6	NM_000179.2	c.3170_3172delTGG_p.Leu105	2	Ex4	EE	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	0	0	0	0	0	0	.	.
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1920C>T_p.	3	Ex13	SY	.	.	.	.	.	1179	277244	0.0043	.	+/44.5kb/	Substituti	5	2	2	0	0	1	criteria pr	criteria p
19	1E+07	.	T	C	35	PASS	LDLR	NM_000527.4	c.1061-8T>C	3	In7/8	SS	.	.	.	.	.	1525	276844	0.0055	.	+/44.5kb/	Substituti	12	2	0	0	0	1	criteria pr	criteria p
2	5E+07	.	A	AT,ATT	35	PASS	MSH6	NM_000179.2	c.3557-5_3557-4dupTT	3	In6/7	SS	.	.	.	.	.	.	.	.	.	+/23.9kb	Insertion	3	1	0	0	0	0	reviewed	criteria p
2	5E+07	.	AT	A	35	PASS	MSH6	NM_000179.2	c.3557-4delT	3	In6/7	SS	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	2	2	0	0	0	1	criteria pr	criteria p
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.2389+57C>T	3	In16/17	INT	.	.	.	.	.	189	30954	0.0061	1	+/44.5kb/	Substituti	0	1	0	0	0	0	criteria pr	criteria p
17	4E+07	.	C	CACA	35	PASS	BRCA1	NM_007294.3	c.5194-2337_5194-2335dupT	3	In18/19	INT	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	1	0	0	0	0	0	reviewed	reviewed
2	5E+07	.	CTT	C,CT	35	PASS	MSH6	NM_000179.2	c.4002-10delT	3	In9/10	INT	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	0	0	1	0	0	0	criteria pr	criteria p
17	4E+07	.	G	A	35	PASS	BRCA1	NM_007294.3	c.-20+11C>T	3	In1/2	5PU	.	.	.	.	.	69	156718	0.0004	.	-/81.2kb/:	Substituti	2	2	1	0	0	1	criteria pr	no assert
17	4E+07	.	C	CAT	35	PASS	BRCA1	NM_007294.3	c.-19-22_-19-21dupAT	3	In1/2	5PU	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	1	1	0	0	0	0	criteria pr	criteria p
17	4E+07	.	GTTTTT	G	35	PASS	BRCA1	NM_007294.3	c.-20+521_-20+525del5	3	In1/2	5PU	.	.	.	.	.	95	30914	0.0031	0	-/81.2kb/:	Deletion	0	1	0	0	0	0	no asserti	no assert
17	4E+07	.	C	T	35	PASS	BRCA1	NM_007294.3	c.+1332G>A	3	3UTR	3PU	.	.	.	.	.	1496	159758	0.0094	.	-/81.2kb/:	Substituti	1	0	0	0	0	0	reviewed	reviewed
19	1E+07	.	TTA	T	35	PASS	LDLR	NM_000527.4	c.+2210_+2211delTA	3	3UTR	3PU	.	.	.	.	.	.	.	.	.	+/44.5kb/	Deletion	0	1	0	0	0	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.1359-1G>A	1	In9/10	ESS	24.8	.	D	.	.	0	245718	0	0	+/44.5kb/	Substituti	0	0	0	3	10	1	criteria pr	criteria p
2	2E+07	.	A	C	35	PASS	APOB	NM_000384.2	c.4503T>G_p.Tyr1501X	1	Ex26	SG	27.5	.	A	.	0.862	1	246156	4E-06	0	-/42.6kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
7	6E+06	.	G	A	35	PASS	PMS2	NM_000535.6	c.730C>T_p.Gln244X	1	Ex7	SG	43	.	A	.	0.865	.	.	.	.	-/38.2kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
7	6E+06	.	G	A	35	PASS	PMS2	NM_000535.6	c.730C>T_p.Gln244X	1	Ex7	SG	43	.	A	.	0.865	.	.	.	.	-/38.2kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
19	1E+07	.	T	G	35	PASS	LDLR	NM_000527.4	c.1942T>G_p.Ser648Ala	2	Ex13	NSY	21.2	D	N	N	0.401	4	277228	1E-05	.	+/44.5kb/	Substituti	0	0	0	1	0	0	criteria pr	criteria p
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1835C>T_p.Ala612Val	2	Ex12	NSY	26.5	D	D	N	0.511	2	277194	7E-06	.	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1835C>T_p.Ala612Val	2	Ex12	NSY	26.5	D	D	N	0.511	2	277194	7E-06	.	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert
2	5E+07	.	G	A	35	PASS	MSH6	NM_000179.2	c.3577G>A_p.Glu1193Lys	2	Ex7	NSY	33	D	D	D	0.987	.	.	.	.	+/23.9kb	Substituti	0	0	0	2	0	0	criteria pr	criteria p
1	6E+07	.	G	A	35	PASS	PCSK9	NM_174936.3	c.1180G>A_p.Gly394Ser	2	Ex7	EE	27.1	D	D	D	0.824	26	236966	0.0001	0	+/25.4kb/	Substituti	0	0	4	0	1	0	criteria pr	criteria p
2	5E+07	.	T	C	35	PASS	MSH6	NM_000179.2	c.1109T>C_p.Leu370Ser	2	Ex4	NSY	25.1	D	D	D	0.669	.	.	.	.	+/23.9kb	Substituti	0	0	2	3	1	1	criteria pr	criteria p
2	5E+07	.	C	T	35	PASS	MSH6	NM_000179.2	c.1618C>T_p.Leu540Phe	2	Ex4	NSY	25	D	D	N	0.785	.	.	.	.	+/23.9kb	Substituti	0	0	1	0	0	0	criteria pr	criteria p
1	6E+07	.	C	T	35	PASS	PCSK9	NM_174936.3	c.1405C>T_p.Arg469Trp	2	Ex9	NSY	22.4	T	N	N	0.709	240	277130	0.0009	.	+/25.4kb/	Substituti	1	3	3	0	1	1	criteria pr	criteria p
3	4E+07	.	G	A	35	PASS	MLH1	NM_000249.3	c.1217G>A_p.Ser406Asn	2	Ex12	NSY	13.46	T	N	N	0.31	249	276804	0.0009	.	+/57.5kb/	Substituti	6	4	1	0	0	0	reviewed	no assert
3	4E+07	.	G	A	35	PASS	MLH1	NM_000249.3	c.1217G>A_p.Ser406Asn	2	Ex12	NSY	13.46	T	N	N	0.31	249	276804	0.0009	.	+/57.5kb/	Substituti	6	4	1	0	0	0	reviewed	no assert
7	6E+06	.	T	C	35	PASS	PMS2	NM_000535.6	c.52A>G_p.Ile18Val	2	Ex2	NSY	24.6	D	D	N	0.656	2457	271244	0.0091	.	-/38.2kb/	Substituti	15	6	1	0	0	0	reviewed	criteria p
17	4E+07	.	G	A	35	PASS	BRCA1	NM_007294.3	c.-86C>T	3	5UTR	5PU	.	.	.	.	.	113	154744	0.0007	.	-/81.2kb/:	Substituti	1	0	3	0	0	0	criteria pr	criteria p
3	4E+07	.	C	A	35	PASS	MLH1	NM_000249.3	c.307-29C>A	3	In3/4	INT	.	.	.	.	.	1497	277100	0.0054	.	+/57.5kb/	Substituti	0	4	0	0	0	0	reviewed	reviewed
13	3E+07	.	G	A	35	PASS	BRCA2	NM_000059.3	c.8851G>A_p.Ala2951Thr	2	Ex22	NSY	26.9	T	D	N	0.518	2486	276544	0.009	.	+/84.2kb	Substituti	26	4	0	0	0	0	reviewed	no assert
3	4E+07	.	A	G	35	PASS	MLH1	NM_000249.3	c.791-79A>G	3	In9/10	INT	.	.	.	.	.	150	30950	0.0048	0	+/57.5kb/	Substituti	0	0	0	0	0	0	no asserti	no assert
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.1049G>A_p.Arg350Gln	2	Ex7	NSY	22	D	N	N	0.176	4	245590	2E-05	0	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert

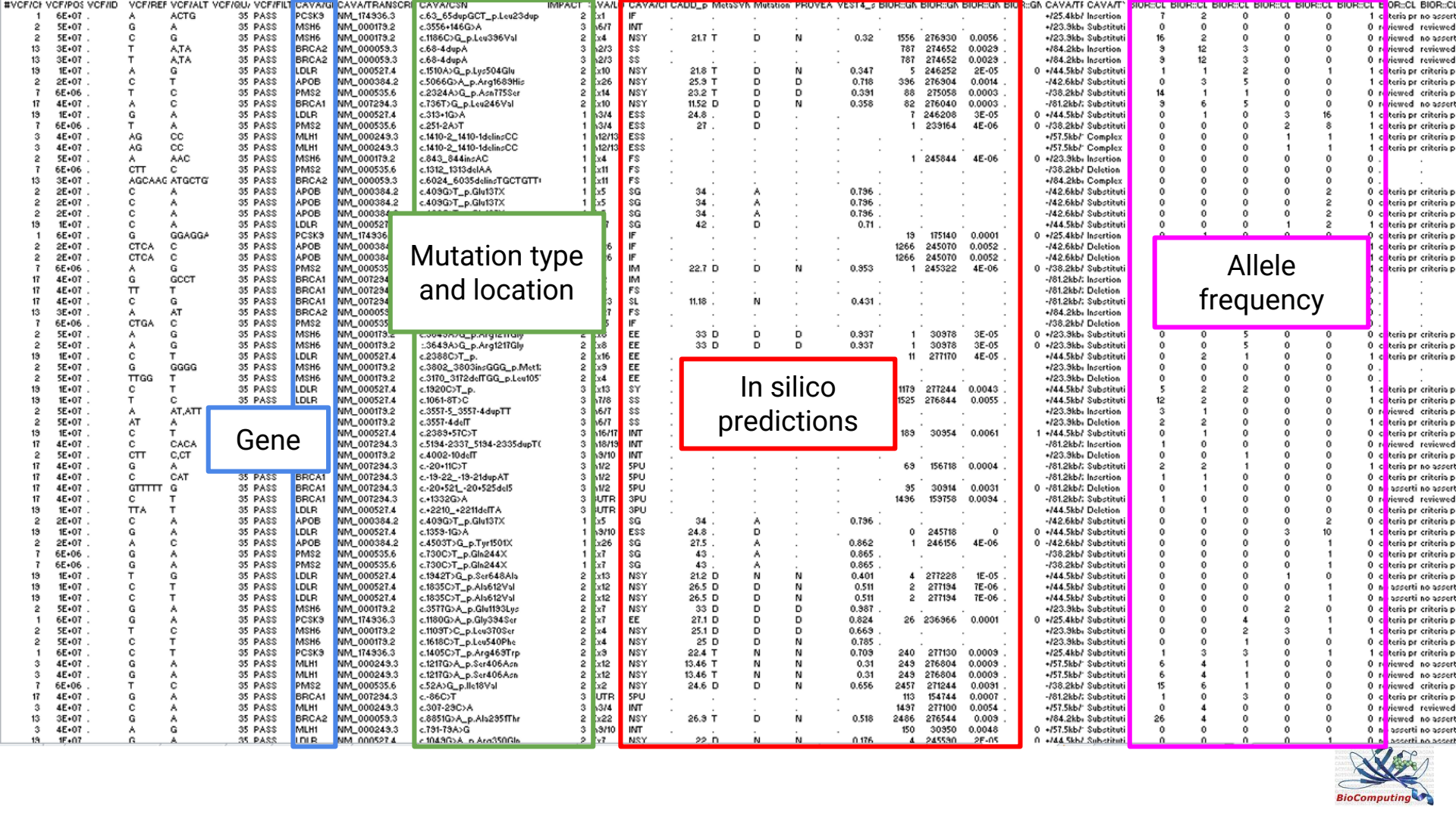
BioComputing

GENE

#VCF/CI	VCF/POS	VCF/ID	VCF/REF	VCF/ALT	VCF/QU/	VCF/FILT	CAVA/GI	CAVA/TRANSCRIF	CAVA/CSN	IMPACT	:AVA/LO	CAVA/CI	CADD_p	MetaSVN	Mutation	PROVEA	VEST4_s	BIOR::GN	BIOR::GN	BIOR::GN	BIOR::GN	CAVA/TF	CAVA/TY	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL	BIOR::CL
1	6E+07	.	A	ACTG	35	PASS	PCSK9	NM_174936.3	c.63_65dupGCT_p.Leu23dup	2	Ex1	IF	.	.	.	.	.	.	.	.	.	+/25.4kb/	Insertion	7	2	0	0	0	1	criteria pr	no assert
2	5E+07	.	G	A	35	PASS	MSH6	NM_000179.2	c.3556+146G>A	3	In6/7	INT	.	.	.	.	.	.	.	.	.	+/23.9kb	Substituti	0	0	0	0	0	0	reviewed	reviewed
2	5E+07	.	C	G	35	PASS	MSH6	NM_000179.2	c.1186C>G_p.Leu396Val	2	Ex4	NSY	21.7	T	D	N	0.32	1556	276930	0.0056	.	+/23.9kb	Substituti	16	2	0	0	0	0	reviewed	no assert
13	3E+07	.	T	A,TA	35	PASS	BRCA2	NM_000059.3	c.68-4dupA	3	In2/3	SS	.	.	.	.	.	787	274652	0.0029	.	+/84.2kb	Insertion	9	12	3	0	0	0	reviewed	reviewed
13	3E+07	.	T	A,TA	35	PASS	BRCA2	NM_000059.3	c.68-4dupA	3	In2/3	SS	.	.	.	.	.	787	274652	0.0029	.	+/84.2kb	Insertion	9	12	3	0	0	0	reviewed	reviewed
19	1E+07	.	A	G	35	PASS	LDLR	NM_000527.4	c.1510A>G_p.Lys504Glu	2	Ex10	NSY	21.8	T	D	N	0.347	5	246252	2E-05	0	+/44.5kb/	Substituti	1	1	2	0	1	1	criteria pr	criteria p
2	2E+07	.	C	T	35	PASS	APOB	NM_000384.2	c.5066G>A_p.Arg1689His	2	Ex26	NSY	25.9	T	D	D	0.718	396	276904	0.0014	.	-/42.6kb/	Substituti	0	3	5	0	0	1	criteria pr	criteria p
7	6E+06	.	T	C	35	PASS	PMS2	NM_000535.6	c.2324A>G_p.Asn775Ser	2	Ex14	NSY	23.2	T	D	D	0.391	88	275058	0.0003	.	-/38.2kb/	Substituti	14	1	1	0	0	0	reviewed	criteria p
17	4E+07	.	A	C	35	PASS	BRCA1	NM_007294.3	c.736T>G_p.Leu246Val	2	Ex10	NSY	11.52	D	D	N	0.358	82	276040	0.0003	.	-/81.2kb/:	Substituti	9	6	5	0	0	0	reviewed	no assert
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.313+1G>A	1	In3/4	ESS	24.8	.	D	.	.	7	246208	3E-05	0	+/44.5kb/	Substituti	0	1	0	3	16	1	criteria pr	criteria p
7	6E+06	.	T	A	35	PASS	PMS2	NM_000535.6	c.251-2A>T	1	In3/4	ESS	27	.	D	.	.	1	239164	4E-06	0	-/38.2kb/	Substituti	0	0	0	2	8	1	criteria pr	criteria p
3	4E+07	.	AG	CC	35	PASS	MLH1	NM_000249.3	c.1410-2_1410-1delinsCC	1	In12/13	ESS	.	.	.	.	.	.	.	.	.	+/57.5kb/	Complex	0	0	0	1	1	1	criteria pr	criteria p
3	4E+07	.	AG	CC	35	PASS	MLH1	NM_000249.3	c.1410-2_1410-1delinsCC	1	In12/13	ESS	.	.	.	.	.	.	.	.	.	+/57.5kb/	Complex	0	0	0	1	1	1	criteria pr	criteria p
2	5E+07	.	A	AAC	35	PASS	MSH6	NM_000179.2	c.843_844insAC	1	Ex4	FS	.	.	.	.	.	1	245844	4E-06	0	+/23.9kb	Insertion	0	0	0	0	0	0	.	.
7	6E+06	.	CTT	C	35	PASS	PMS2	NM_000535.6	c.1312_1313delAA	1	Ex11	FS	.	.	.	.	.	.	.	.	.	-/38.2kb/	Deletion	0	0	0	0	0	0	.	.
13	3E+07	.	AGCAAC	ATGCTG	35	PASS	BRCA2	NM_000059.3	c.6024_6035delinsTGCTGTT	1	Ex11	FS	.	.	.	.	.	.	.	.	.	+/84.2kb	Complex	0	0	0	0	0	0	.	.
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
19	1E+07	.	C	A	35	PASS	LDLR	NM_000527.4	c.2546C>A_p.Ser849X	1	Ex17	SG	42	.	D	.	0.71	.	.	.	.	+/44.5kb/	Substituti	0	0	0	1	2	1	criteria pr	criteria p
1	6E+07	.	G	GGAGGA	35	PASS	PCSK9	NM_174936.3	c.101_106dup6_p.Glu34_Asp3	2	Ex1	IF	.	.	.	.	.	19	175140	0.0001	0	+/25.4kb/	Insertion	0	1	0	0	0	0	criteria pr	criteria p
2	2E+07	.	CTCA	C	35	PASS	APOB	NM_000384.2	c.6639_6641delTGA_p.Asp2	2	Ex26	IF	.	.	.	.	.	1266	245070	0.0052	.	-/42.6kb/	Deletion	3	6	4	0	0	1	criteria pr	criteria p
2	2E+07	.	CTCA	C	35	PASS	APOB	NM_000384.2	c.6639_6641delTGA_p.Asp2	2	Ex26	IF	.	.	.	.	.	1266	245070	0.0052	.	-/42.6kb/	Deletion	3	6	4	0	0	1	criteria pr	criteria p
7	6E+06	.	A	G	35	PASS	PMS2	NM_000535.6	c.2T>C_p.Met1?	2	Ex1	IM	22.7	D	D	N	0.953	1	245322	4E-06	0	-/38.2kb/	Substituti	0	0	0	2	2	1	criteria pr	criteria p
17	4E+07	.	G	GCCT	35	PASS	BRCA1	NM_007294.3	c.2_3insAGG_p.Met1?	2	Ex2	IM	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	0	0	0	0	0	0	.	.
17	4E+07	.	TT	T	35	PASS	BRCA1	NM_007294.3	c.1delA_p.Met1?	1	Ex2	FS	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Deletion	0	0	0	0	0	0	.	.
17	4E+07	.	C	G	35	PASS	BRCA1	NM_007294.3	c.5591G>C_p.X1864SerextX3	2	Ex23	SL	11.18	.	N	.	0.431	.	.	.	.	-/81.2kb/:	Substituti	0	0	0	0	0	0	.	.
13	3E+07	.	A	AT	35	PASS	BRCA2	NM_000059.3	c.10256_10257insT	1	Ex27	FS	.	.	.	.	.	.	.	.	.	+/84.2kb	Insertion	0	0	0	0	0	0	.	.
7	6E+06	.	CTGA	C	35	PASS	PMS2	NM_000535.6	c.2583_2585delGAA_p.Gln86	2	Ex15	IF	.	.	.	.	.	.	.	.	.	-/38.2kb/	Deletion	0	0	0	0	0	0	.	.
2	5E+07	.	A	G	35	PASS	MSH6	NM_000179.2	c.3649A>G_p.Arg1217Gly	2	Ex8	EE	33	D	D	D	0.937	1	30978	3E-05	0	+/23.9kb	Substituti	0	0	5	0	0	0	criteria pr	criteria p
2	5E+07	.	A	G	35	PASS	MSH6	NM_000179.2	:.3649A>G_p.Arg1217Gly	2	Ex8	EE	33	D	D	D	0.937	1	30978	3E-05	0	+/23.9kb	Substituti	0	0	5	0	0	0	criteria pr	criteria p
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.2388C>T_p.	2	Ex16	EE	.	.	.	.	.	11	277170	4E-05	.	+/44.5kb/	Substituti	0	2	1	0	0	1	criteria pr	criteria p
2	5E+07	.	G	GGGG	35	PASS	MSH6	NM_000179.2	c.3802_3803insGGG_p.Met1	2	Ex9	EE	.	.	.	.	.	.	.	.	.	+/23.9kb	Insertion	0	0	0	0	0	0	.	.
2	5E+07	.	TTGG	T	35	PASS	MSH6	NM_000179.2	c.3170_3172delTGG_p.Leu105	2	Ex4	EE	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	0	0	0	0	0	0	.	.
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1920C>T_p.	3	Ex13	SY	.	.	.	.	.	1179	277244	0.0043	.	+/44.5kb/	Substituti	5	2	2	0	0	1	criteria pr	criteria p
19	1E+07	.	T	C	35	PASS	LDLR	NM_000527.4	c.1061-8T>C	3	In7/8	SS	.	.	.	.	.	1525	276844	0.0055	.	+/44.5kb/	Substituti	12	2	0	0	0	1	criteria pr	criteria p
2	5E+07	.	A	AT,ATT				NM_000179.2	c.3557-5_3557-4dupTT	3	In6/7	SS	.	.	.	.	.	.	.	.	.	+/23.9kb	Insertion	3	1	0	0	0	0	reviewed	criteria p
2	5E+07	.	AT	A				NM_000179.2	c.3557-4delT	3	In6/7	SS	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	2	2	0	0	0	1	criteria pr	criteria p
19	1E+07	.	C	T				NM_000527.4	c.2389+57C>T	3	In16/17	INT	.	.	.	.	.	189	30954	0.0061	1	+/44.5kb/	Substituti	0	1	0	0	0	0	criteria pr	criteria p
17	4E+07	.	C	CACA				NM_007294.3	c.5194-2337_5194-2335dupT(	3	In18/19	INT	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	1	0	0	0	0	0	reviewed	reviewed
2	5E+07	.	CTT	C,CT				NM_000179.2	c.4002-10delT	3	In9/10	INT	.	.	.	.	.	.	.	.	.	+/23.9kb	Deletion	0	0	1	0	0	0	criteria pr	criteria p
17	4E+07	.	G	A				NM_007294.3	c.-20+11C>T	3	In1/2	5PU	.	.	.	.	.	69	156718	0.0004	.	-/81.2kb/:	Substituti	2	2	1	0	0	1	criteria pr	no assert
17	4E+07	.	C	CAT	35	PASS	BRCA1	NM_007294.3	c.-19-22_-19-21dupAT	3	In1/2	5PU	.	.	.	.	.	.	.	.	.	-/81.2kb/:	Insertion	1	1	0	0	0	0	criteria pr	criteria p
17	4E+07	.	GTTTTT	G	35	PASS	BRCA1	NM_007294.3	c.-20+521_-20+525del5	3	In1/2	5PU	.	.	.	.	.	95	30914	0.0031	0	-/81.2kb/:	Deletion	0	1	0	0	0	0	no asserti	no assert
17	4E+07	.	C	T	35	PASS	BRCA1	NM_007294.3	c.+1332G>A	3	3UTR	3PU	.	.	.	.	.	1496	159758	0.0094	.	-/81.2kb/:	Substituti	1	0	0	0	0	0	reviewed	reviewed
19	1E+07	.	TTA	T	35	PASS	LDLR	NM_000527.4	c.+2210_+2211delTA	3	3UTR	3PU	.	.	.	.	.	.	.	.	.	+/44.5kb/	Deletion	0	1	0	0	0	0	criteria pr	criteria p
2	2E+07	.	C	A	35	PASS	APOB	NM_000384.2	c.409G>T_p.Glu137X	1	Ex5	SG	34	.	A	.	0.796	.	.	.	.	-/42.6kb/	Substituti	0	0	0	0	2	0	criteria pr	criteria p
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.1359-1G>A	1	In9/10	ESS	24.8	.	D	.	.	0	245718	0	0	+/44.5kb/	Substituti	0	0	0	3	10	1	criteria pr	criteria p
2	2E+07	.	A	C	35	PASS	APOB	NM_000384.2	c.4503T>G_p.Tyr1501X	1	Ex26	SG	27.5	.	A	.	0.862	1	246156	4E-06	0	-/42.6kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
7	6E+06	.	G	A	35	PASS	PMS2	NM_000535.6	c.730C>T_p.Gln244X	1	Ex7	SG	43	.	A	.	0.865	.	.	.	.	-/38.2kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
7	6E+06	.	G	A	35	PASS	PMS2	NM_000535.6	c.730C>T_p.Gln244X	1	Ex7	SG	43	.	A	.	0.865	.	.	.	.	-/38.2kb/	Substituti	0	0	0	0	1	0	criteria pr	criteria p
19	1E+07	.	T	G	35	PASS	LDLR	NM_000527.4	c.1942T>G_p.Ser648Ala	2	Ex13	NSY	21.2	D	N	N	0.401	4	277228	1E-05	.	+/44.5kb/	Substituti	0	0	0	1	0	0	criteria pr	criteria p
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1835C>T_p.Ala612Val	2	Ex12	NSY	26.5	D	D	N	0.511	2	277194	7E-06	.	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert
19	1E+07	.	C	T	35	PASS	LDLR	NM_000527.4	c.1835C>T_p.Ala612Val	2	Ex12	NSY	26.5	D	D	N	0.511	2	277194	7E-06	.	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert
2	5E+07	.	G	A	35	PASS	MSH6	NM_000179.2	c.3577G>A_p.Glu1193Lys	2	Ex7	NSY	33	D	D	D	0.987	.	.	.	.	+/23.9kb	Substituti	0	0	0	2	0	0	criteria pr	criteria p
1	6E+07	.	G	A	35	PASS	PCSK9	NM_174936.3	c.1180G>A_p.Gly394Ser	2	Ex7	EE	27.1	D	D	D	0.824	26	236966	0.0001	0	+/25.4kb/	Substituti	0	0	4	0	1	0	criteria pr	criteria p
2	5E+07	.	T	C	35	PASS	MSH6	NM_000179.2	c.1109T>C_p.Leu370Ser	2	Ex4	NSY	25.1	D	D	D	0.669	.	.	.	.	+/23.9kb	Substituti	0	0	2	3	1	1	criteria pr	criteria p
2	5E+07	.	C	T	35	PASS	MSH6	NM_000179.2	c.1618C>T_p.Leu540Phe	2	Ex4	NSY	25	D	D	N	0.785	.	.	.	.	+/23.9kb	Substituti	0	0	1	0	0	0	criteria pr	criteria p
1	6E+07	.	C	T	35	PASS	PCSK9	NM_174936.3	c.1405C>T_p.Arg469Trp	2	Ex9	NSY	22.4	T	N	N	0.709	240	277130	0.0009	.	+/25.4kb/	Substituti	1	3	3	0	1	1	criteria pr	criteria p
3	4E+07	.	G	A	35	PASS	MLH1	NM_000249.3	c.1217G>A_p.Ser406Asn	2	Ex12	NSY	13.46	T	N	N	0.31	249	276804	0.0009	.	+/57.5kb/	Substituti	6	4	1	0	0	0	reviewed	no assert
3	4E+07	.	G	A	35	PASS	MLH1	NM_000249.3	c.1217G>A_p.Ser406Asn	2	Ex12	NSY	13.46	T	N	N	0.31	249	276804	0.0009	.	+/57.5kb/	Substituti	6	4	1	0	0	0	reviewed	no assert
7	6E+06	.	T	C	35	PASS	PMS2	NM_000535.6	c.52A>G_p.Ile18Val	2	Ex2	NSY	24.6	D	D	N	0.656	2457	271244	0.0091	.	-/38.2kb/	Substituti	15	6	1	0	0	0	reviewed	criteria p
17	4E+07	.	G	A	35	PASS	BRCA1	NM_007294.3	c.-86C>T	3	5UTR	5PU	.	.	.	.	.	113	154744	0.0007	.	-/81.2kb/:	Substituti	1	0	3	0	0	0	criteria pr	criteria p
3	4E+07	.	C	A	35	PASS	MLH1	NM_000249.3	c.307-29C>A	3	In3/4	INT	.	.	.	.	.	1497	277100	0.0054	.	+/57.5kb/	Substituti	0	4	0	0	0	0	reviewed	reviewed
13	3E+07	.	G	A	35	PASS	BRCA2	NM_000059.3	c.8851G>A_p.Ala2951Thr	2	Ex22	NSY	26.9	T	D	N	0.518	2486	276544	0.009	.	+/84.2kb	Substituti	26	4	0	0	0	0	reviewed	no assert
3	4E+07	.	A	G	35	PASS	MLH1	NM_000249.3	c.791-79A>G	3	In9/10	INT	.	.	.	.	.	150	30950	0.0048	0	+/57.5kb/	Substituti	0	0	0	0	0	0	no asserti	no assert
19	1E+07	.	G	A	35	PASS	LDLR	NM_000527.4	c.1049G>A_p.Arg350Gln	2	Ex7	NSY	22	D	N	N	0.176	4	245590	2E-05	0	+/44.5kb/	Substituti	0	0	0	0	1	0	no asserti	no assert


Mutation type and location
Gene


Mutation type and location
Gene
In silico predictions

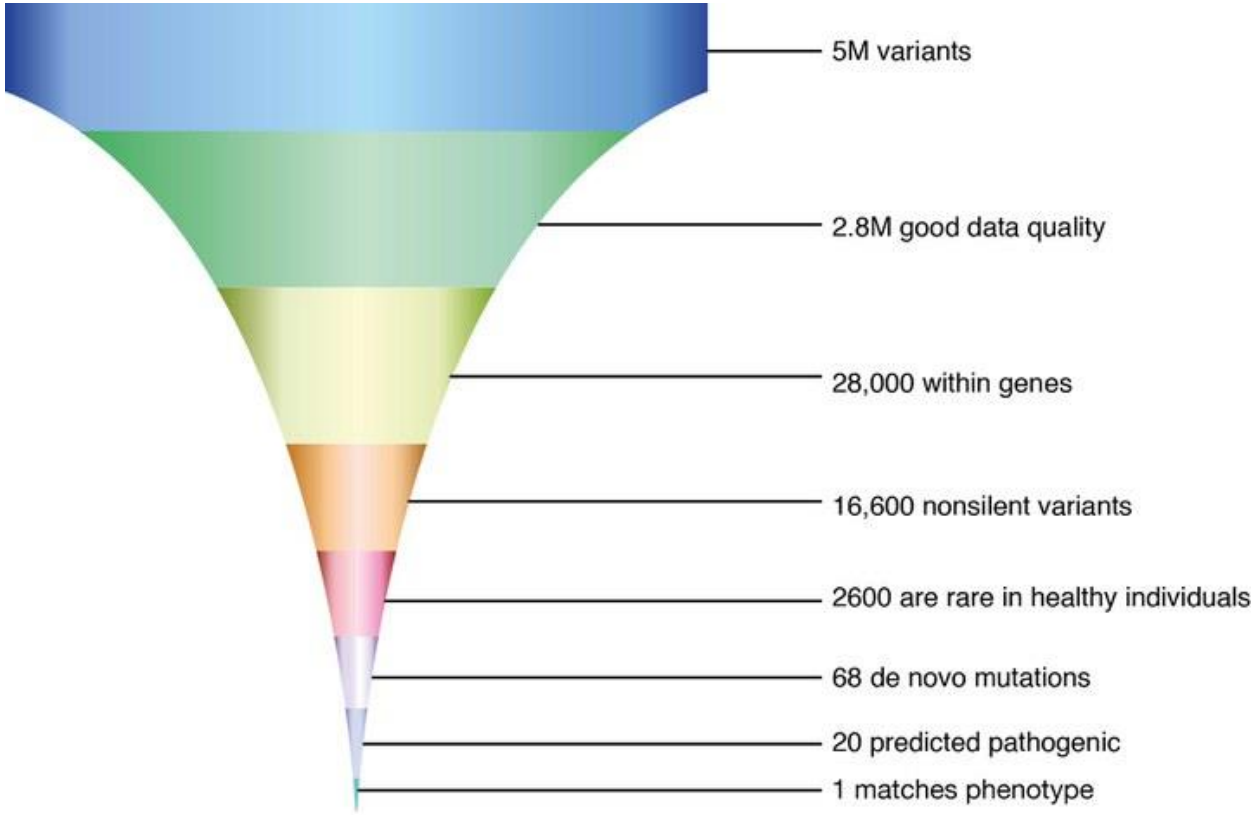

Mutation type and location
Gene
In silico predictions
Allele frequency
BioComputing

Variant filtering: criteria

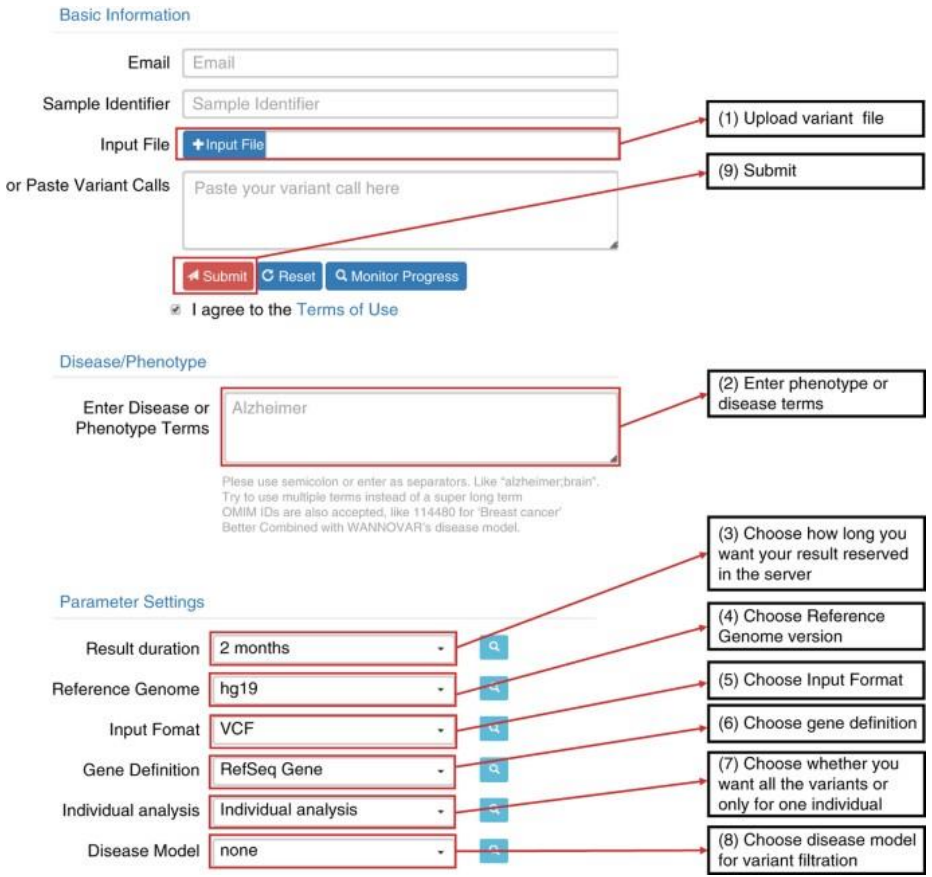
- Genes list
- Sequencing parameters (filter artifactuals)
- Variant types (exonic – intronic;
- Variant class (frameshift - synonymous)
- Population frequency
- Pathogenicity prediction score
- Conservation score
- Clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline
- Presence in mutation databases

(PMID: 28118812;
https://www.nature.com/articles/s41525-021-00227-3)

Variant filtering: criteria



wANNOVAR *(http://wannovar.usc.edu)*



wANNOVAR implemented a:

Variant-reduction pipeline based on commonly used **filters** and **disease models** such as selecting:

- only the nonsynonymous variants and splicing variants
- rare or novel variants in the 1000 Genomes Project database
- predicted deleterious variants.

Phenotype-based variant prioritization sample's specific phenotype or disease information is available and may help identify causal variants.

wANNOVAR limitations

- Default 'variant-reduction' schemes (disease models) may not be optimal for the **specific use case** (e.g. eliminate true causal variants in some scenarios during the filtering procedure)

- **complex diseases**: rule-based 'hard' filtering versus a probabilistic prioritization approach

- Originally developed as annotation tools for genetic variants from a **single genome**, with limited functionality on analyzing **multiple** genomes (not support case-control association analysis, or family-based association analysis)

- ANNOVAR provides several different types of **deleteriousness-prediction scores**, and it leaves the choice of selecting annotations to users

Variant filtering: criteria

- **Genes list**
- Sequencing parameters (filter artifactuals)
- Variant types (exonic – intronic;
- Variant class (frameshift - synonymous)
- Population frequency
- Pathogenicity prediction score
- Conservation score
- Clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline
- Presence in mutation databases

(PMID: 28118812;
https://www.nature.com/articles/s41525-021-00227-3)

Variant filtering: identify variants in specific genomic regions

- Gene panel disease-oriented
- Identity-by-descent (IBD), the detection of shared segments inherited from a common ancestor
- Linkage analysis
- Exome capture regions, capture array manufacturers will provide the regions in BED file

Browser Extensible Data (BED) format

- Provides a flexible way to define the data lines that are displayed in an annotation track. These are generally used for user defined sequence features as well as **graphical representations of features**.
- BED lines have **three** required fields and **nine** additional optional fields.
- The first three required **BED fields** are:
    1. **chrom** - The name of the chromosome
    2. **chromStart** - The starting position of the feature in the chromosome or scaffold. The first base in a chromosome is numbered 0.
    3. **chromEnd** - The ending position of the feature in the chromosome

```
chr1 213941196 213942363
chr1 213942363 213943530
chr1 213943530 213944697
chr2 158364697 158365864
chr2 158365864 158367031
chr3 127477031 127478198
chr3 127478198 127479365
chr3 127479365 127480532
chr3 127480532 127481699
```

Browser Extensible Data (BED) format

- **What software uses bed files?**
- Alignment viewers can use these data to graphically display certain features.
- bedtools uses this format to query for nearby features.
- Some annotation files are in this format.
- Feature detection packages use this as output.

- **How are these files generated?**
- Feature detection algorithms.
- Lots of databases that hold certain genomic features report their data in this format.
- Sometimes manually curated from alignments (via bedtools, bamtools, etc.).

Variant filtering: identify variants in specific genomic regions

- Users can supply your own region annotation databases in generic, BED or GFF formats.
- Region-based annotation looks for overlap of a query variant with a region (this region could be a single position) in a database, and it does not care about exact match of positions, and it does not care about nucleotide identity at all.
- Users can select annotation tracks that are already provided by the UCSC Genome Browser annotation databases
- Annotate variants against GFF3-formatted annotation databases, using the region-based annotation procedure

General Feature Format (GFF) / General Transfer Format (GTF)

- GFF or GTF are a tab-delimited text file that holds information any and every feature that can be applied to a nucleic acid or protein sequence.
- Everything from CDS, microRNAs, binding domains, ORFs, and more can be handled by this format.
- many variations of the original GFF format, latest accepted format (GFF3)

Sample GTF output from Ensembl data dump:

```
1 transcribed_unprocessed_pseudogene  gene        11869 14409 . + . gene_id "ENSG00000223972"; gene_name "DDX11L1"; gene_source "havana"; gene_biotype "transcrik
1 processed_transcript                transcript  11869 14409 . + . gene_id "ENSG00000223972"; transcript_id "ENST00000456328"; gene_name "DDX11L1"; gene_sourc e
```

Sample GFF output from Ensembl export:

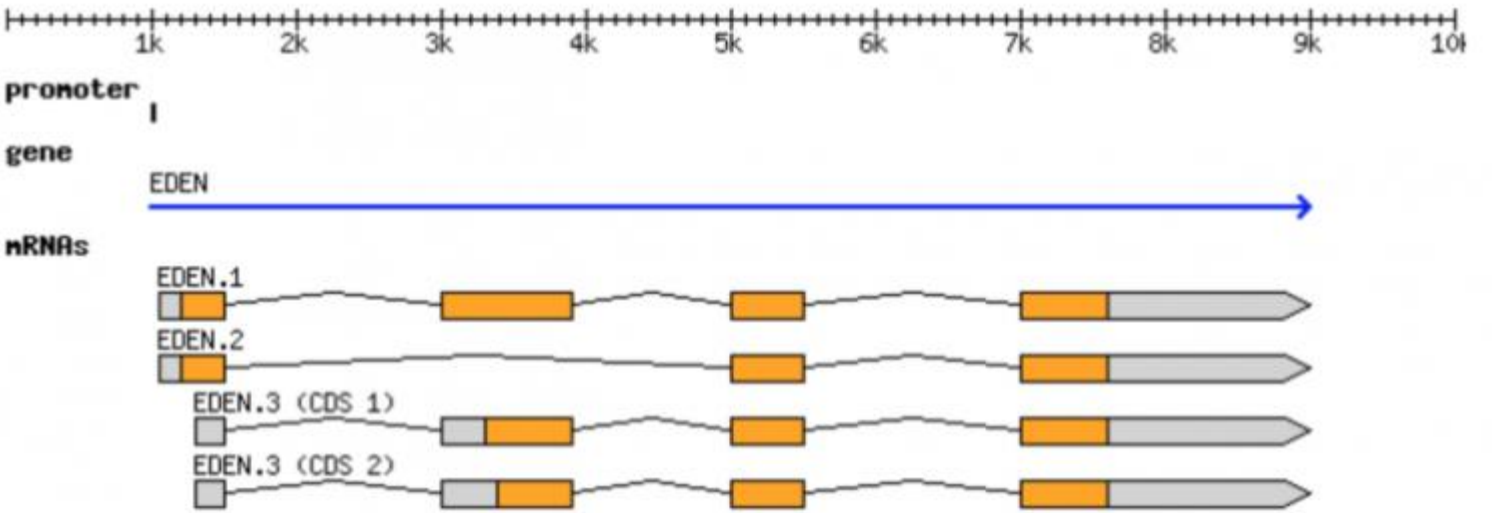
```
X       Ensembl Repeat  2419108 2419128 42      .       .       hid=trf; hstart=1; hend=21
X       Ensembl Repeat  2419108 2419410 2502    -       .       hid=AluSx; hstart=1; hend=303
X       Ensembl Repeat  2419108 2419128 0       .       .       hid=dust; hstart=2419108; hend=2419128
X       Ensembl Pred.trans.     2416676 2418760 450.19  -       2       genscan=GENSCAN00000019335
X       Ensembl Variation       2413425 2413425 .       +       .
X       Ensembl Variation       2413805 2413805 .       +       .
```

GFF3 has 9 required fields

1. Sequence ID
2. Source *(Describes the algorithm that generated this feature, e.g Genescane or Genebank).*
3. Feature Type *(Describes what the feature is: mRNA, domain, exon, etc.)*
4. Feature Start
5. Feature End
6. Score *(Typically E-values for sequence similarity and P-values for predictions).*
7. Strand
8. Phase *(Indicates where the feature begins with reference to the reading frame).*
9. Attributes *(A semicolon-separated list of tag-value pairs, providing additional information about each feature)*

GFF3 file

Graphical representation of a canonical gene



The same information can be represented in GFF3 format

GFF3 file

```
0  ##gff-version 3.2.1
1  ##sequence-region ctg123 1 1497228
2  ctg123 . gene            1000  9000  .  +  .  ID=gene00001;Name=EDEN
3  ctg123 . TF_binding_site 1000  1012  .  +  .  ID=tfbs00001;Parent=gene00001
4  ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00001;Parent=gene00001;Name=EDEN.1
5  ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00002;Parent=gene00001;Name=EDEN.2
6  ctg123 . mRNA            1300  9000  .  +  .  ID=mRNA00003;Parent=gene00001;Name=EDEN.3
7  ctg123 . exon            1300  1500  .  +  .  ID=exon00001;Parent=mRNA00003
8  ctg123 . exon            1050  1500  .  +  .  ID=exon00002;Parent=mRNA00001,mRNA00002
9  ctg123 . exon            3000  3902  .  +  .  ID=exon00003;Parent=mRNA00001,mRNA00003
10 ctg123 . exon            5000  5500  .  +  .  ID=exon00004;Parent=mRNA00001,mRNA00002,mRNA00003
11 ctg123 . exon            7000  9000  .  +  .  ID=exon00005;Parent=mRNA00001,mRNA00002,mRNA00003
12 ctg123 . CDS             1201  1500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
13 ctg123 . CDS             3000  3902  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
14 ctg123 . CDS             5000  5500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
15 ctg123 . CDS             7000  7600  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
16 ctg123 . CDS             1201  1500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
17 ctg123 . CDS             5000  5500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
18 ctg123 . CDS             7000  7600  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
19 ctg123 . CDS             3301  3902  .  +  0  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
20 ctg123 . CDS             5000  5500  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
21 ctg123 . CDS             7000  7600  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
22 ctg123 . CDS             3391  3902  .  +  0  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
23 ctg123 . CDS             5000  5500  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
24 ctg123 . CDS             7000  7600  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
```

GFF3 file

- **What Software uses GFF3?**
- Any tool that requires information about gene position for analysis such as:
    - Mapping RNA-seq such as Tophat, HTSeq, Genome Browsers like IGV

- **How is this file generated?**
- Feature identification software report motifs/features in this format.
- Almost all sequence annotation databases report in this format.

Variant filtering: criteria

- Genes list
- **Sequencing parameters (filter artifactuals)**
- Variant types (exonic – intronic;
- Variant class (frameshift - synonymous)
- Population frequency
- Pathogenicity prediction score
- Conservation score
- Clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline
- Presence in mutation databases

(PMID: 28118812;
https://www.nature.com/articles/s41525-021-00227-3)

Variant filtering: sequencing parameters

NGS data are prone to artifactual variant calls due to e.g. short-read alignment

- **Allele balance** (the ratio of reads aligned at a variant locus that support the alternate allele) 0.2 - 0.8. More stringent threshold of 0.3–0.7 has a very high transmission rate, and an estimated false negative rate of ~1.41%
- **Genotype quality** (GQ) threshold: >20
- Read **Depth of Coverage** (DP): >10
- Presence of **repetitive or polymeric** sequences

Variant filtering: criteria

- Genes list
- Sequencing parameters (filter artifactuals)
- **Variant types (exonic – intronic)**
- **Variant class (frameshift - synonymous)**
- Population frequency
- Pathogenicity prediction score
- Conservation score
- Clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline
- Presence in mutation databases

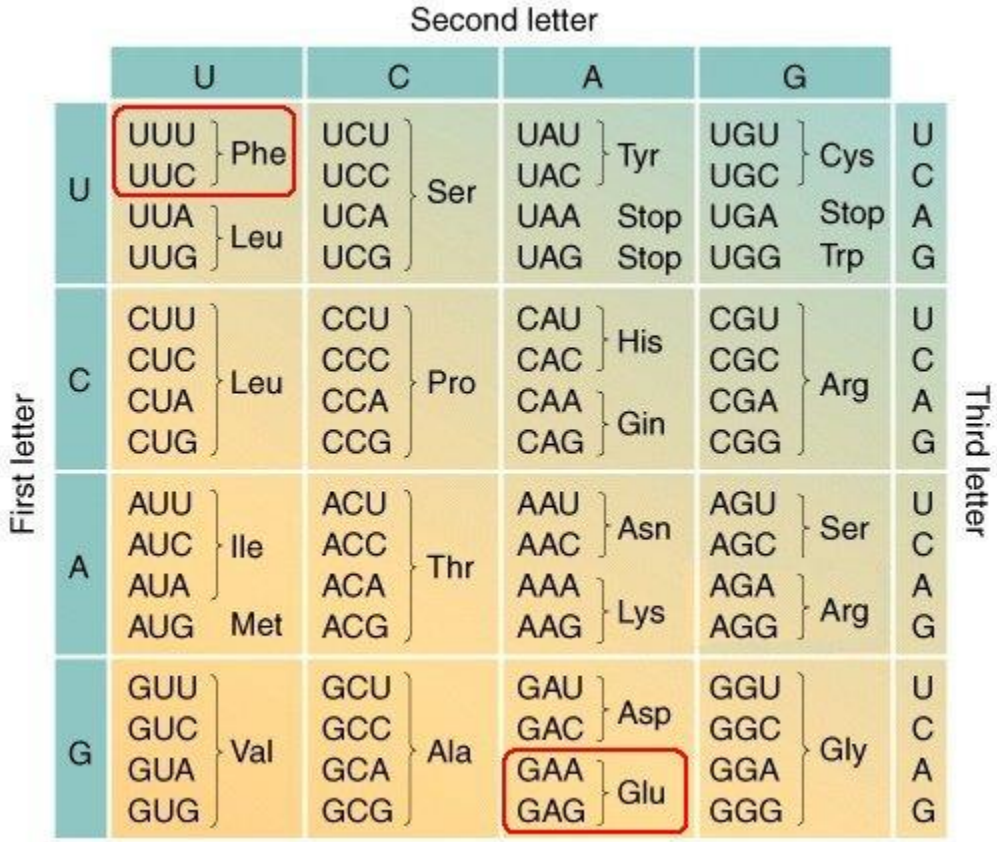
(PMID: 28118812;
https://www.nature.com/articles/s41525-021-00227-3)

Types of mutations

- Point (missense, same sense, stop gain, …)
- Frameshift
- Splicing
- Regulatory
- Insertion/deletion (small)
- Insertion/deletion (large)
- Repeats

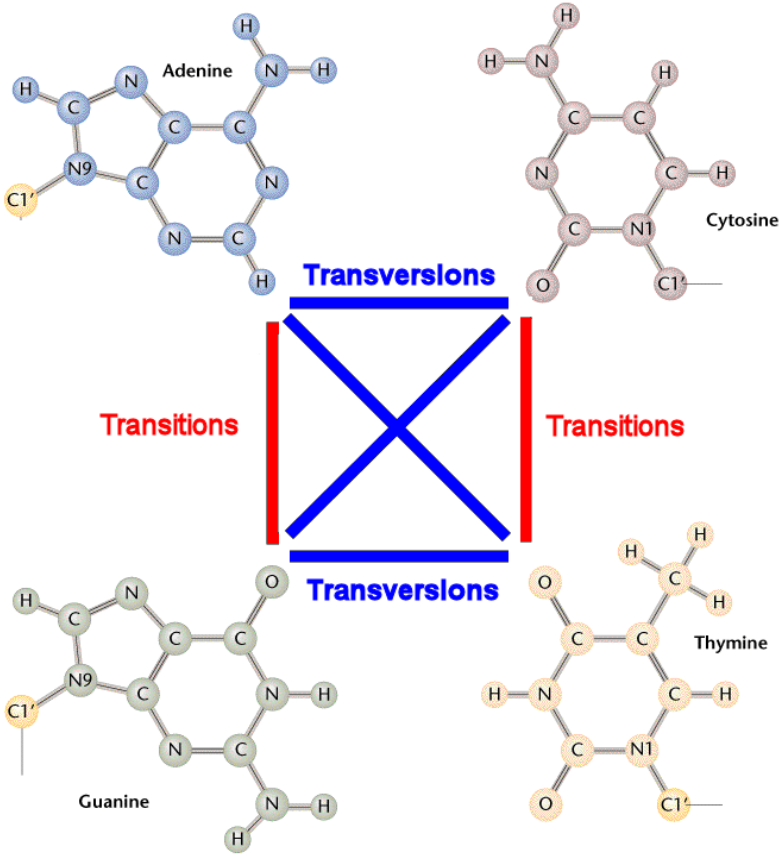
BioComputing

Point mutations



Point mutations

Transitions
involve bases of similar shape



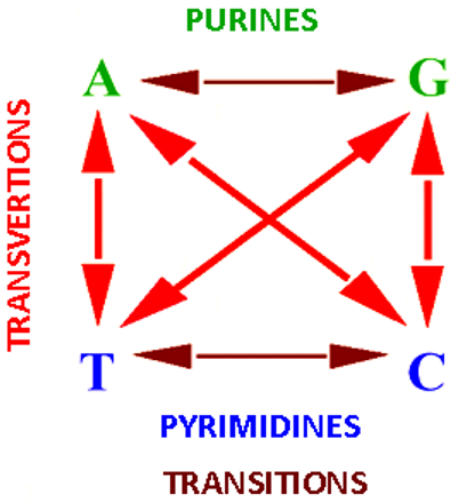
Transitions are generated at **higher frequency** than transversions

two-ring
purines (**A** - **G**)

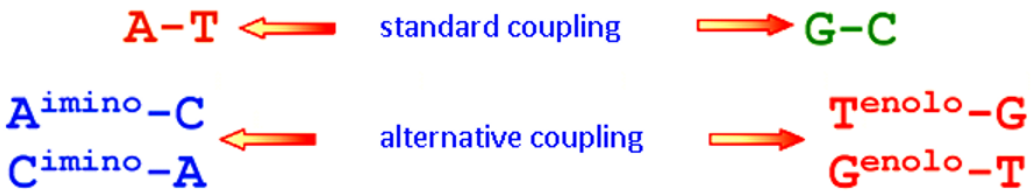
one- ring
pyrimidines (**C** - **T**)

Interchanges of **purine for pyrimidine bases**

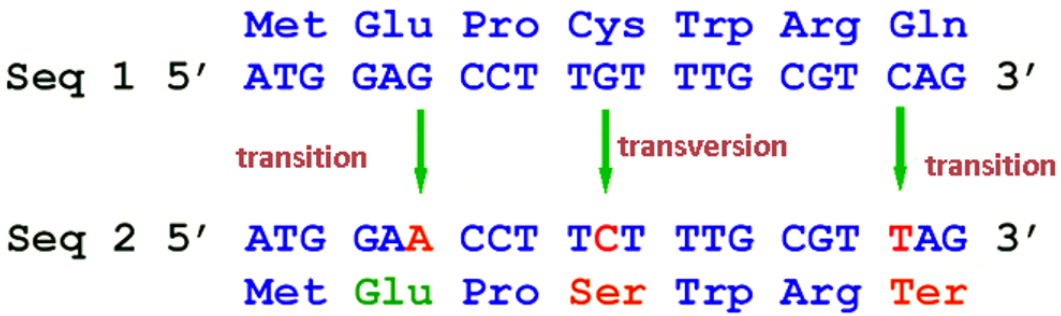
Point mutations



Point mutations are allowed due to the physicochemical properties of nucleobases. They can take the shape of several conformational states, facilitating alternative coupling to those discovered by Watson & Crick

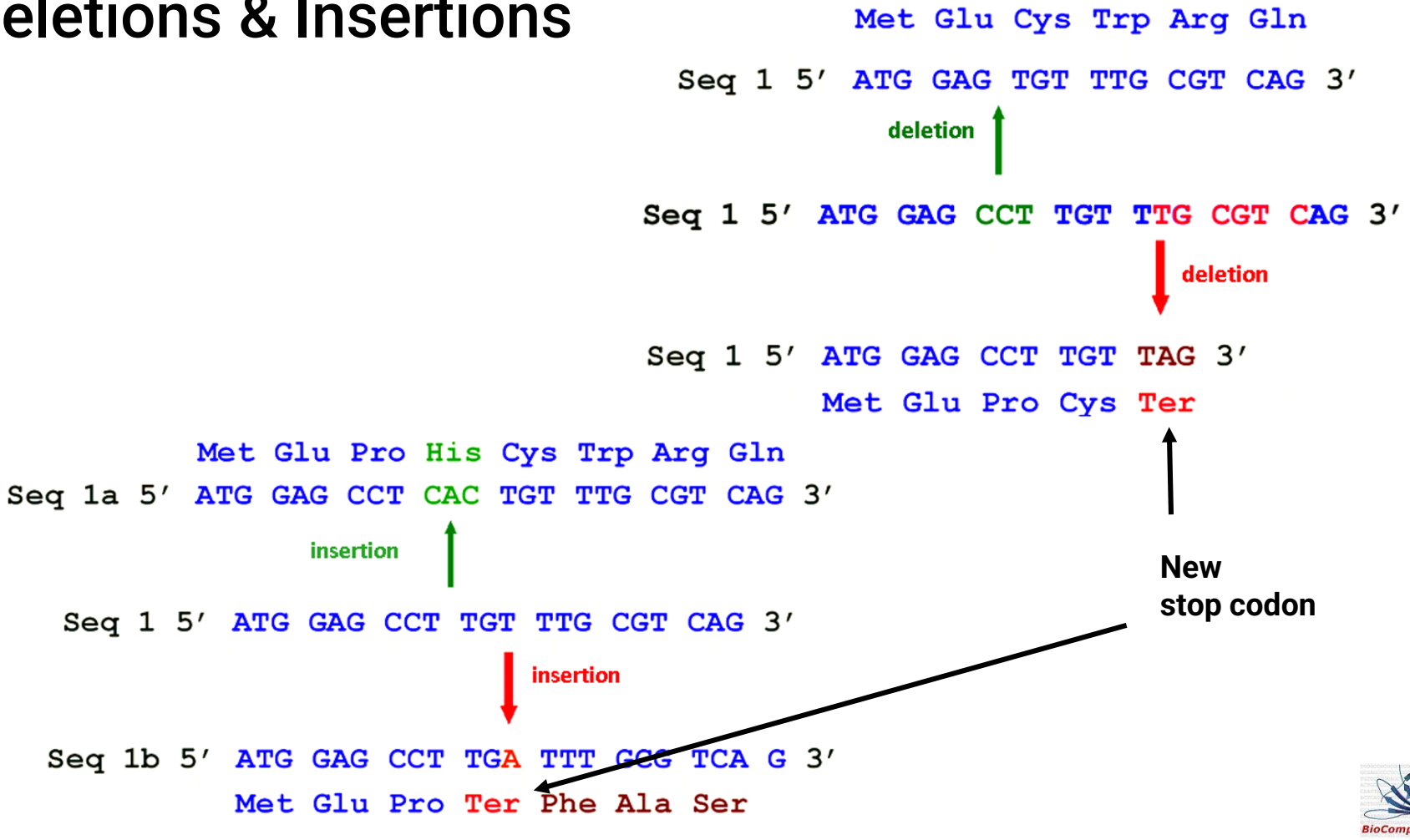


Dna nucleobases substitution events



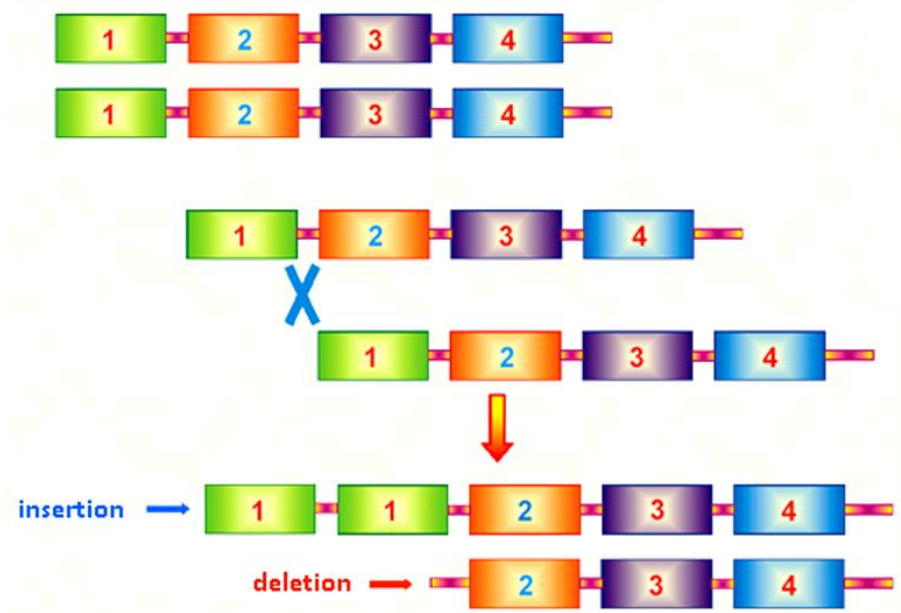
Transitions are less likely to result in amino acid substitutions, and are therefore more likely to persist as "***silent substitutions***" in populations as **single nucleotide polymorphisms** (**SNPs**).

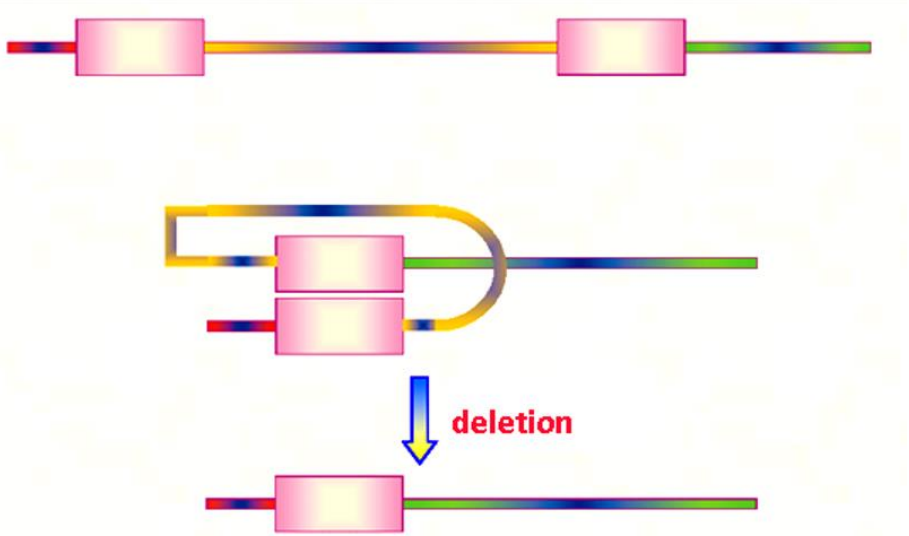
Deletions & Insertions



Unequal Crossing over

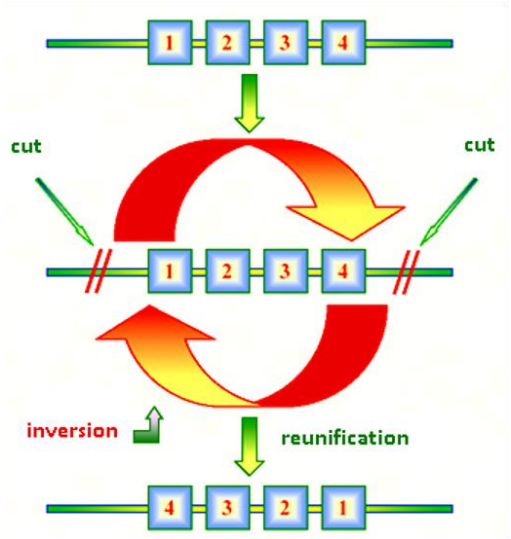
Inter-chromosomal pairing

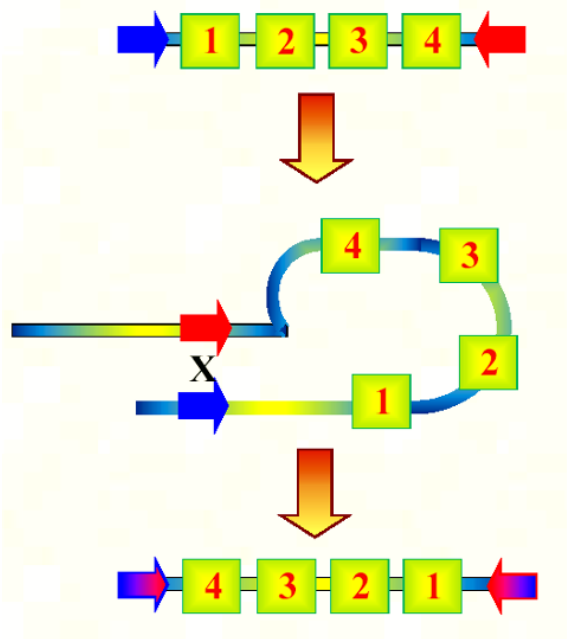




Intra-chromosomal pairing

Inversion





Bases from complementary DNA strand

Seq 1 5′ ATG GAG CCT TGT TTG CGT CAG 3′

inversion

Seq 1 5′ ATG GAG ACA AGG TTG CGT CAG 3′

Met Glu Thr Arg Trp Arg Gln

Types of mutations

1. MSH2 NM_000251.2 c.1590A>G_p.Glu530=
2. BRCA1 NM_007294.3 c.736T>G_p.Leu246Val
3. BRCA2 NM_000059.3 c.6474delT_p.Gln2159AsnfsTer9
4. PMS2 NM_000535.6 c.730C>T_p.Gln244X
5. PCSK9 NM_174936.3 c.524-1G>A
6. MLH1 NM_000249.3 c.307-29C>A

Sequence variant nomenclature

https://varnomen.hgvs.org/

SPECIAL ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

HGVS Recommendations for the Description of Sequence Variants: 2016 Update

Hum Mutat (2016) 37:564-569

Johan T. den Dunnen,[1]* Raymond Dalgleish,[2] Donna R. Maglott,[3] Reece K. Hart,[4] Marc S. Greenblatt,[5] Jean McGowan-Jordan,[6] Anne-Francoise Roux,[7] Timothy Smith,[8] Stylianos E. Antonarakis,[9] and Peter E.M. Taschner[10] on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO)

"mutation" > disease-associated variant
"polymorphism" > not disease-associated
"pathogenic" > disease-associated

http://www.hgvs.org/varnomen/HGVS-basics2017.pdf

BioComputing

Sequence variant nomenclature

- DNA A, G, C, T g.957A>T, c.63-3T>C
- RNA a, g, c, u r.957a>u, r.(?), r.spl?
- protein (mostly deduced) three / one letter amino acid code * = stop codon p.(His78Gln)
- use official HGNC gene symbols
- provide reference sequence covering complete sequence largest transcript preferably a LRG e.g. LRG-123 give accession.version number e.g. NM 012654.3
- indicate type of Reference Sequence:
    - coding DNA c.
    - genomic g.
    - mitochondrial m.
    - non-coding RNA n.
    - RNA r.
    - protein p.

Mutalyzer: https://mutalyzer.nl/

Bioinformatics, 37(18), 2021, 2811–2817
doi: 10.1093/bioinformatics/btab051
Advance Access Publication Date: 4 February 2021
Original Paper

Sequence analysis

Mutalyzer 2: next generation HGVS nomenclature checker

Mihai Lefter [1,*], Jonathan K. Vis[1,2], Martijn Vermaat[1], Johan T. den Dunnen[1,2], Peter E. M. Taschner[1,3] and Jeroen F. J. Laros [1,2,4]

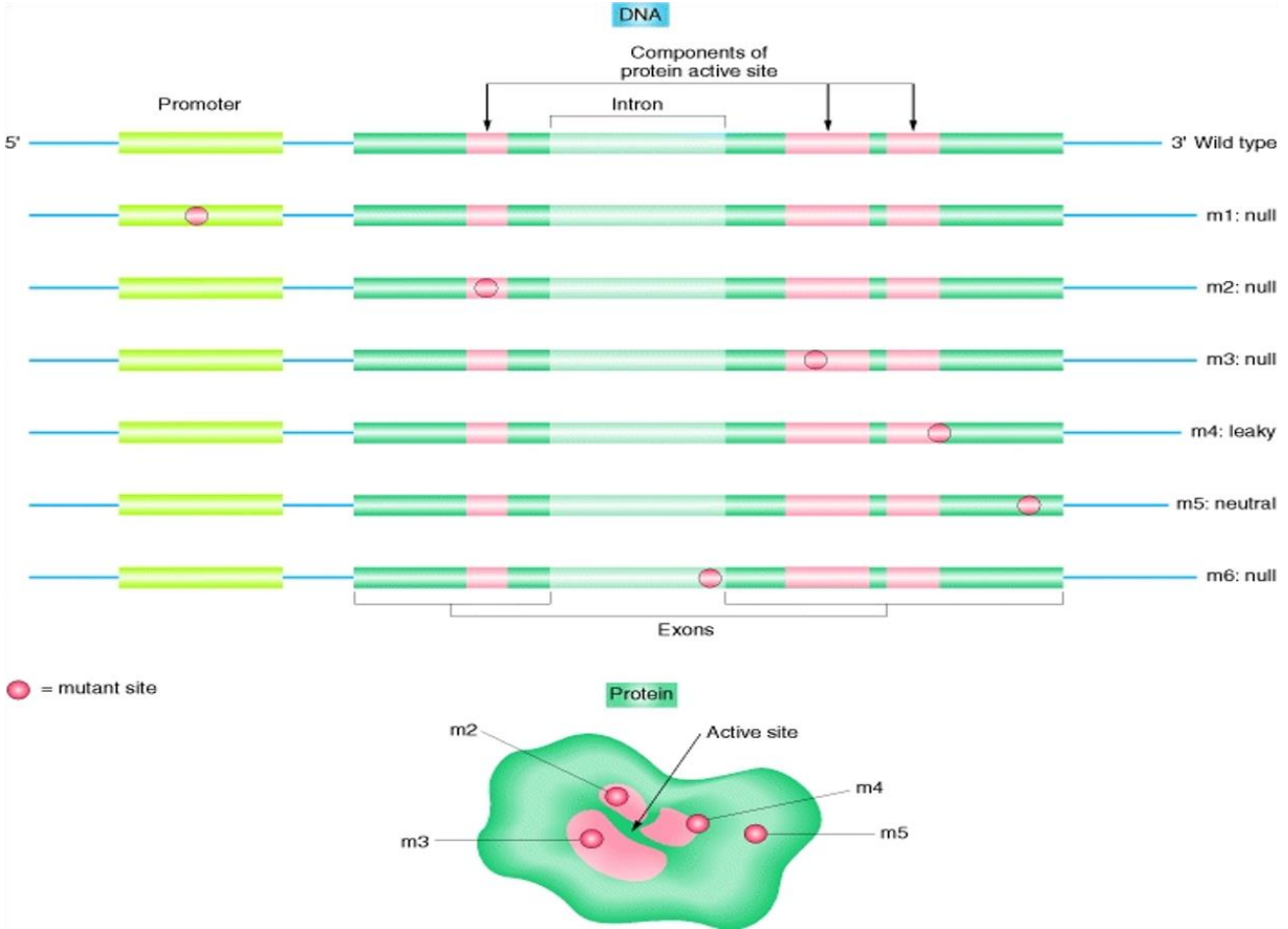
Mutalyzer 2 tool suite is designed to automatically apply the HGVS guidelines

The source code is available on GitHub (https://github.com/mutalyzer/mutalyzer)

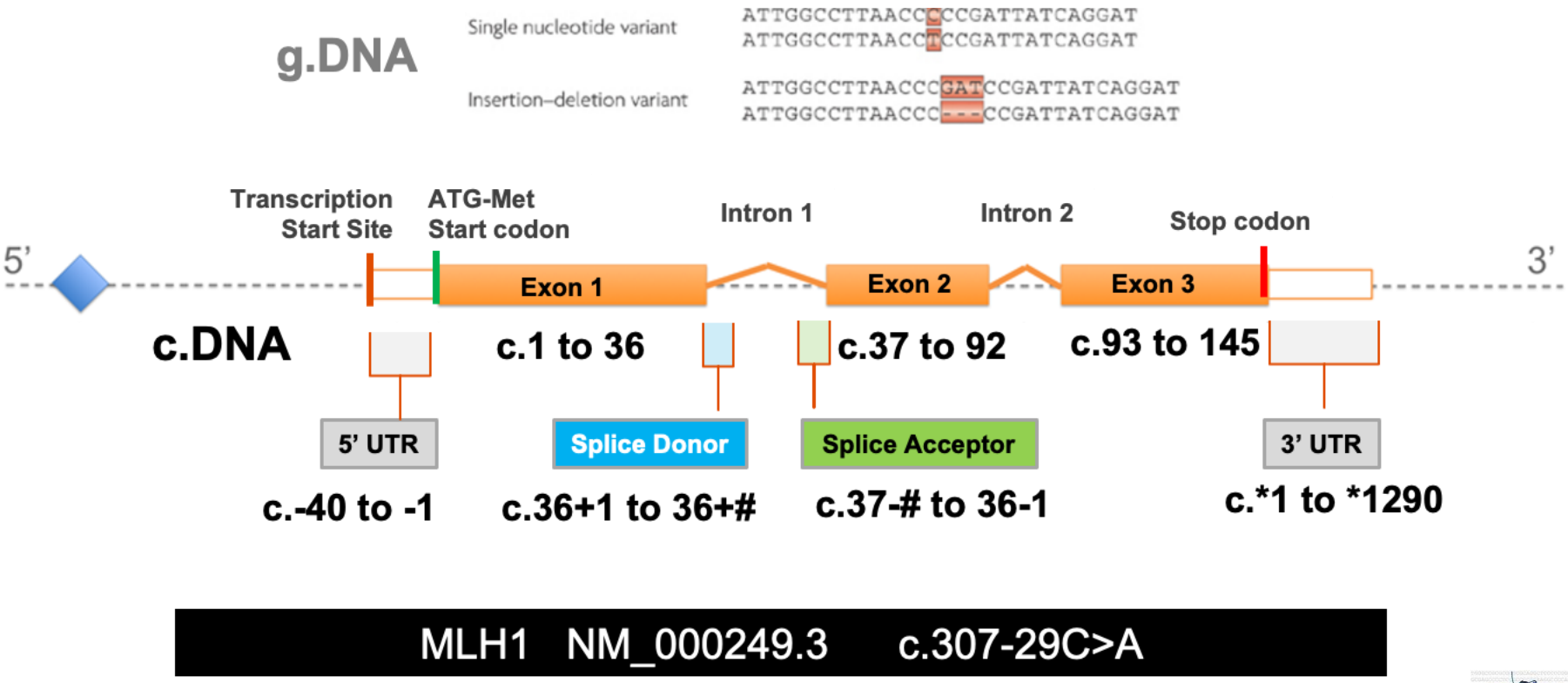
- *Name Checker / Normalizer,* provides checking and disambiguation of variant descriptions
- *Position Converter*, which converts descriptions between chromosomal and transcript references
- *Description Extractor*, generates HGVS variant descriptions given a reference sequence and an observed sequence
- *Mapper*, maps a description to another reference.
- *Batch Processor*, can be used to process up to 50 descriptions with the Normalizer

BioComputing

Missense mutations: defects of folding or function?

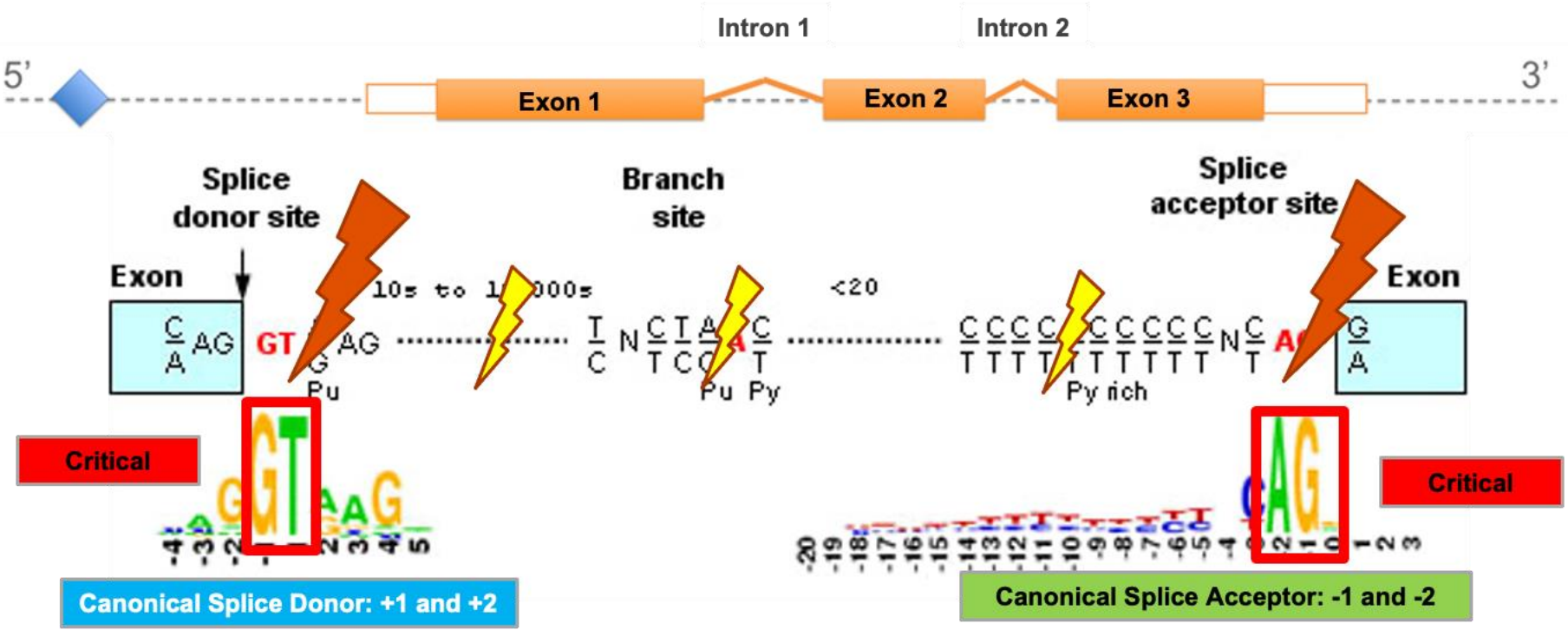


Missense mutations: defects of folding or function?

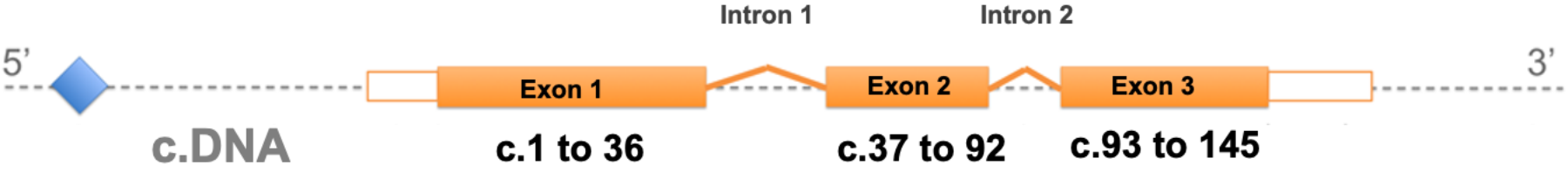


MLH1 NM_000249.3 c.307-29C>A

Missense mutations: defects of folding or function?



Missense mutations: defects of folding or function?



Protein Consequences:

	No mutation	Point mutations			
		Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr

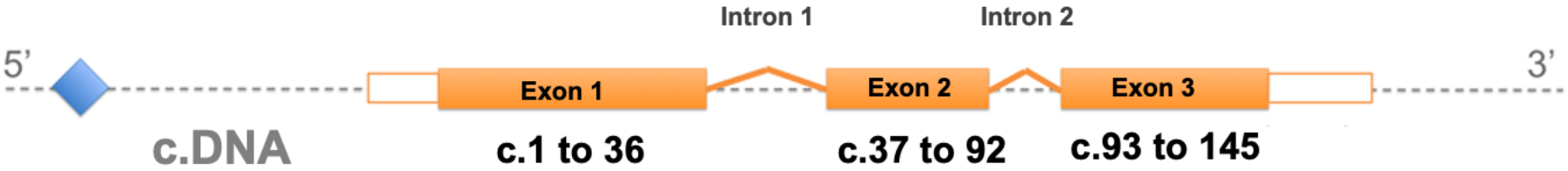
MSH2
c.1590A>G
p.Glu530=
"synonymous"

PMS2
c.730C>T
p.Gln244X
p.Gln244Ter

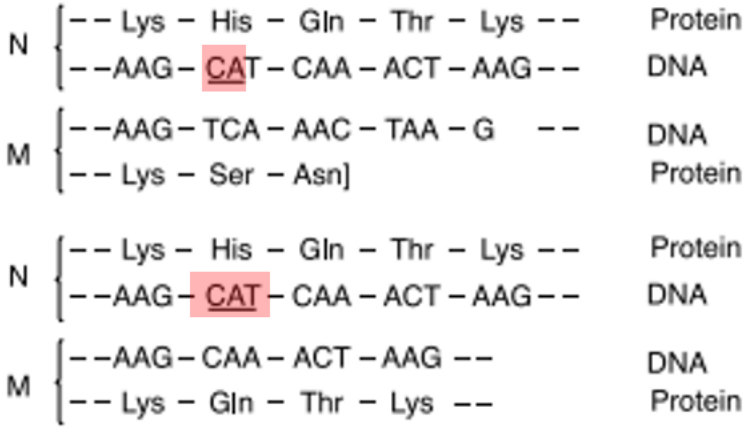
BRCA1
c.736T>G
p.Leu246Val

The functional impact of an SNP is depends on its **location** (i.e., coding or non-coding region) and **effects** to the related **protein sequence** (i.e., synonymous or nonsynonymous).

Missense mutations: defects of folding or function?



Protein Consequences

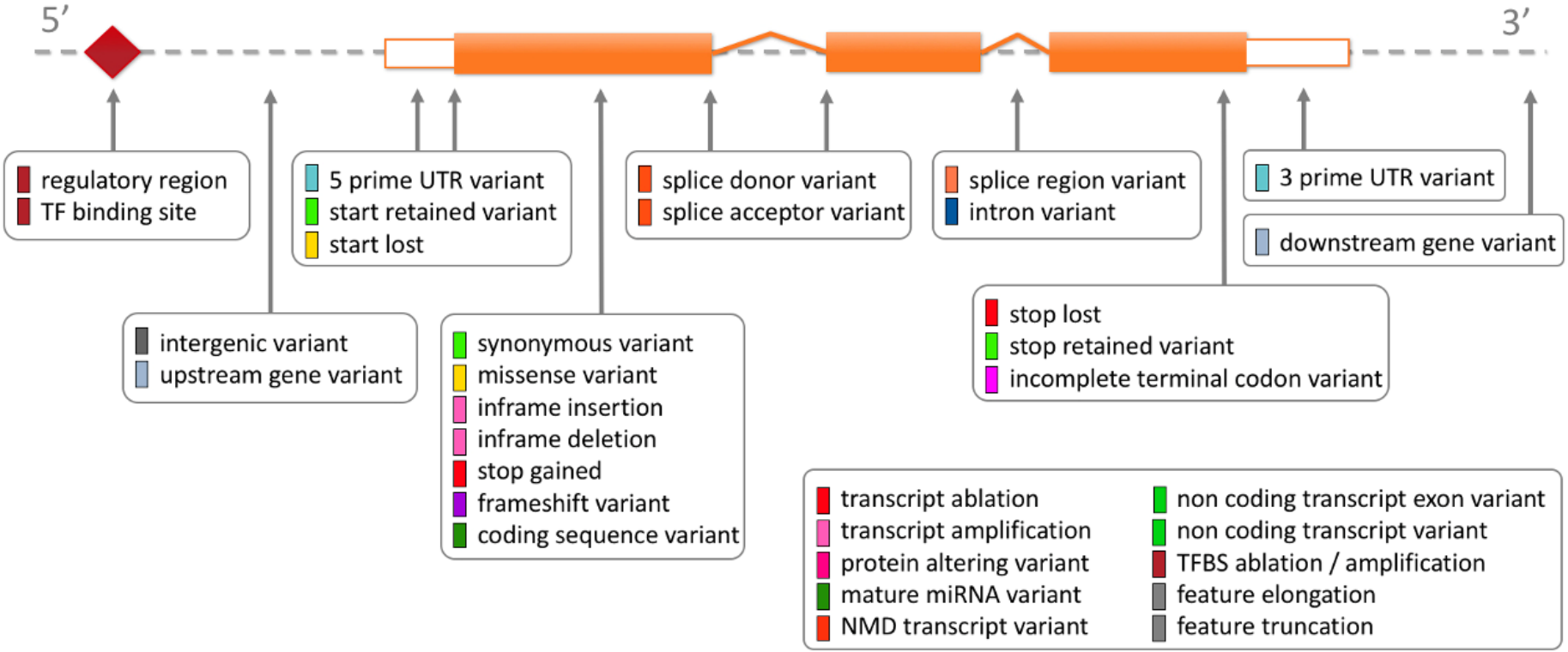


frameshift

In frame deletion

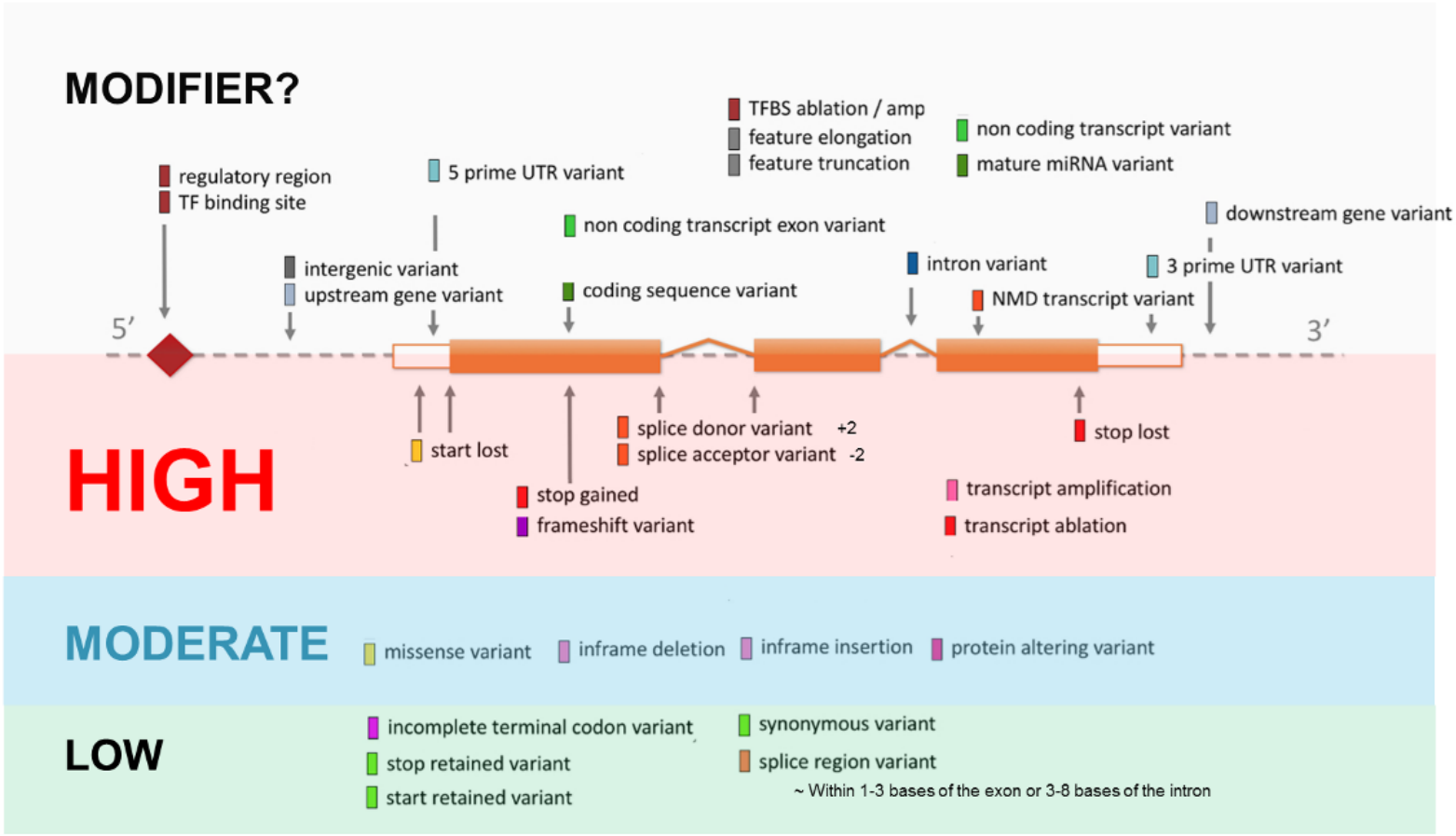
SNPs leading to the **truncation** of a protein sequence are mostly considered to **inhibit protein function**

Variant Severity: Variable definitions but helps prioritize



https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html

Variant Severity: Variable definitions but helps prioritize



Variant filtering: criteria

- Genes list
- Sequencing parameters (filter artifactuals)
- Variant types (exonic – intronic)
- Variant class (frameshift - synonymous)
- **Population frequency**
- Pathogenicity prediction score
- Conservation score
- Clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline
- Presence in mutation databases

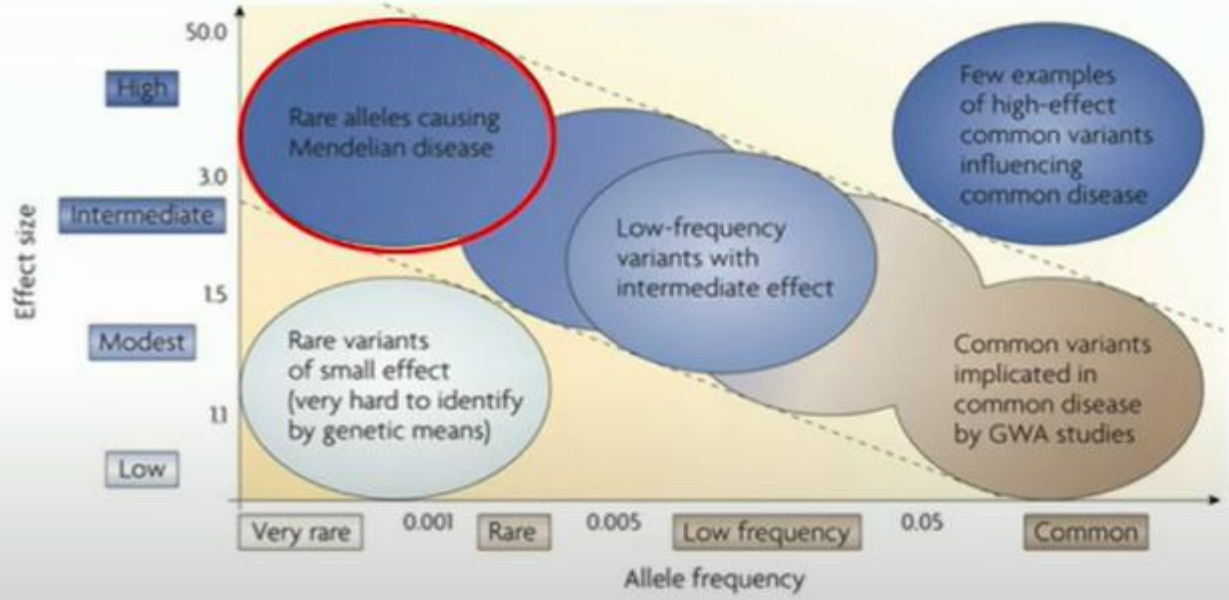
(PMID: 28118812;
https://www.nature.com/articles/s41525-021-00227-3)

Variant filtering: Variant frequency

What's in an exome

- **Every genome contains many rare, potentially functional variants**
  - ~500 rare missense variants (1/3 of which are predicted damaging by *in silico* predictors)
  - ~100 LoF variants: ~20 homozygous, ~20 rare
  - ~100 rare variants in known disease genes
  - ~50 reported disease-causing mutations (!)
  - 1-2 *de novo* coding mutations
  - Unknown number of sequencing errors

Mendelian disease: Mainly looking for rare variants with large effect size



Antonarakis, *et al.*, Nature Reviews Genetics (2010) 11, 380-384.

Nature Reviews | Genetics

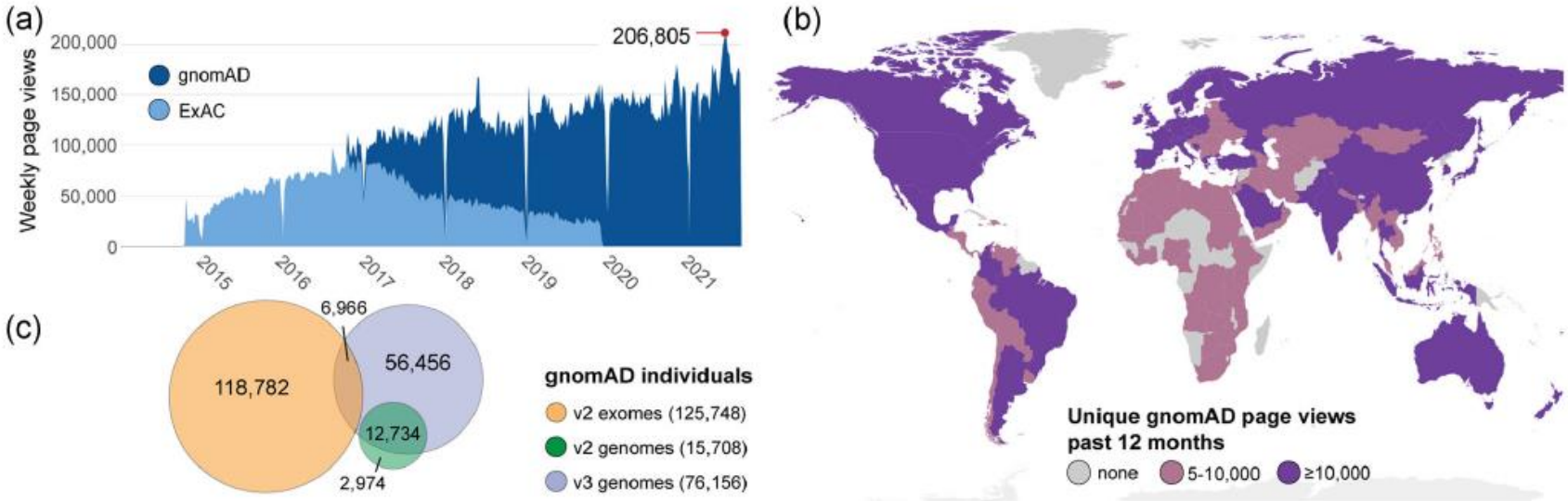
Variant filtering: Variant frequency

- **Population databases**
- GnomAD
- ExAC
- Trans-Omics for PrecisionMedicine (TOPMed)-BRAVO
- Geisinger Healthcare System DiscovEHR dataset

- **Internal laboratory panel of normal**

Variant filtering: GnomAD

https://doi.org/10.1002/humu.24309



Variant filtering: Genome Aggregation Database (GnomAD)

- **195,000 individuals**
- 85% of all possible synonymous CpG-to TpG transitions observed; across non-CpG trinucleotide contexts, less than 12% of possible synonymous variants; 4% of nonsense variants
- With the existing sample size of gnomAD, each individual has
  - **200 very rare coding variants** (gnomAD allele frequency <0.1%).
  - **tens of variants that are absent from gnomAD**
  - **27 ± 13 novel coding variants** that are absent in all other gnomAD individuals (variants unique to that individual
- Case–control from common adult-onset disease studies (type 2 diabetes, psychiatric disorders, and cardiovascular disease (60 studies)
- Database of Genotypes and Phenotypes (dbGaP)
- Report the population with highest Allele frequency for each variant (Popmax AF)

Variant filtering: GnomAD

- Uniform joint variant calling
- variant quality control (QC)
- individuals known to be affected with severe pediatric disease excluded
- Quality Score Recalibration (VQSR) have been applied to distinguish true genetic variants from artifacts
- depth ≥ 10, genotype quality ≥ 20, minor allele fraction ≥ 0.2 for nonreference heterozygous variants

- Using gnomAD - tips and tricks (video)
- gnomAD: Using large genomic data sets to interpret human genetic variation (video)

Variant filtering: Variant frequency

Disease model

- **De novo or Dominant** Absent in Public database or MAF < 0.001 in each of the eight gnomAD populations (e.g., African, Latino, East Asian, etc.)
- For **dominant,** the number of homozygous alternate alleles in gnomAD to be <10
- Recessive MAF < 0.01
- Complex MAF < 0.0001
- Somatic: high confidence: absent in Public database or MAF < 0.001

Identifying de novo variants

De novo definition: A variant that has arisen in an individual for the first time and is not inherited from a parent.

- can be caused by chance, impaired DNA repair, or increased mutation of the genome due to mutagens such as radiation and particular chemicals
- The human **de novo mutation rate 1.29 x 10^{-8}** per base pair per generation
- **70 de novo** against 4-5 million inherited variants, for each proband
- In protein coding exome we expect **1 de novo** mutation against 50.000 inherited variants
- For 99.9% variant calling precision, **50 false-positive** calls for each de novo mutations

Filtering for artifactual and population frequency

Variant filtering: de novo variants