MAYO CLINIC

CENTER FOR
INDIVIDUALIZED MEDICINE

Clinical Variant Interpretation
June 9th, 2022

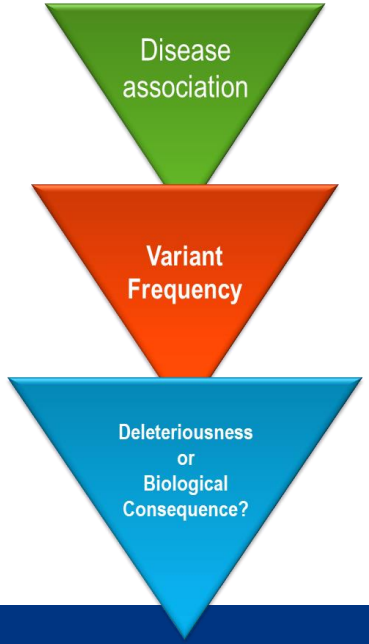
Erica Macke, PhD
Postdoctoral Research Fellow
Translational Omics Program
Quantitative Health Sciences Department
macke.erica@mayo.edu

Objectives:

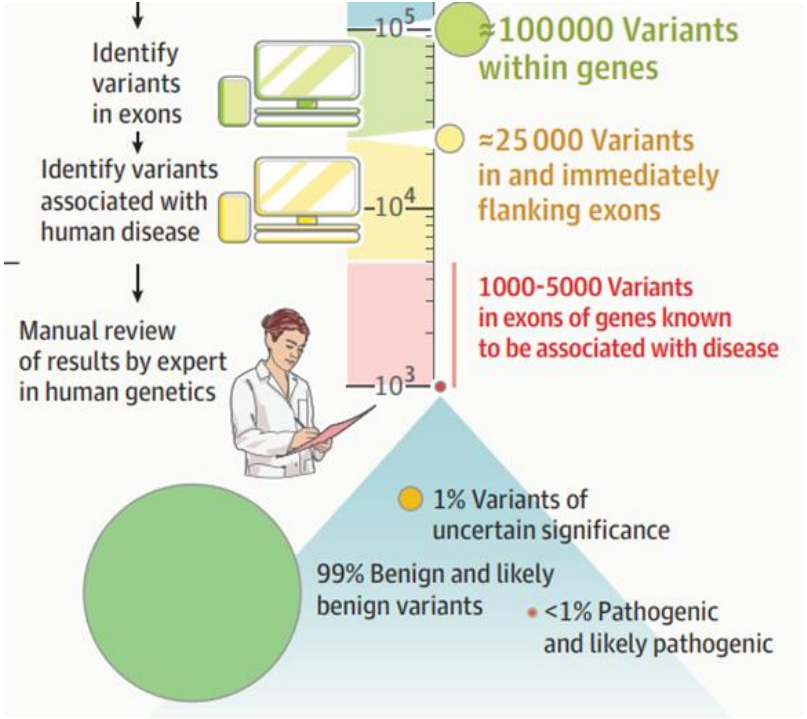
- Understand basics of variant prioritization and nomenclature
- Overview of the current framework for analysis and interpretation of sequence variants for monogenic disorders
- Overview of key available resources and their utility with variant interpretation.

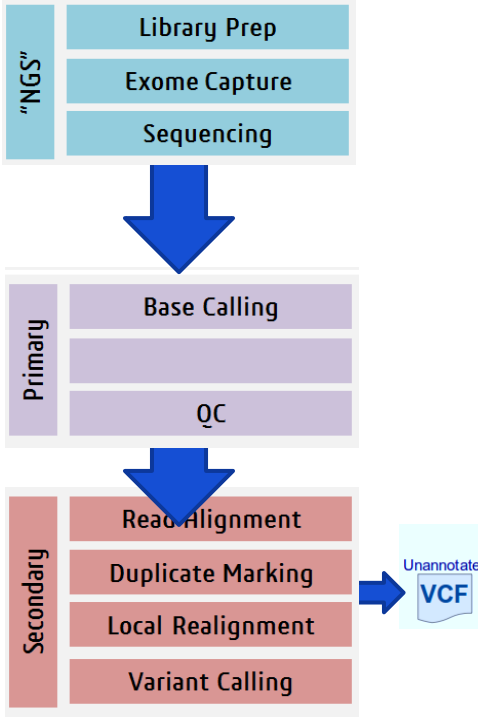
Variant prioritization: why does it matter?

I have detected variants in my patient sample. What information can I use to interpret them?



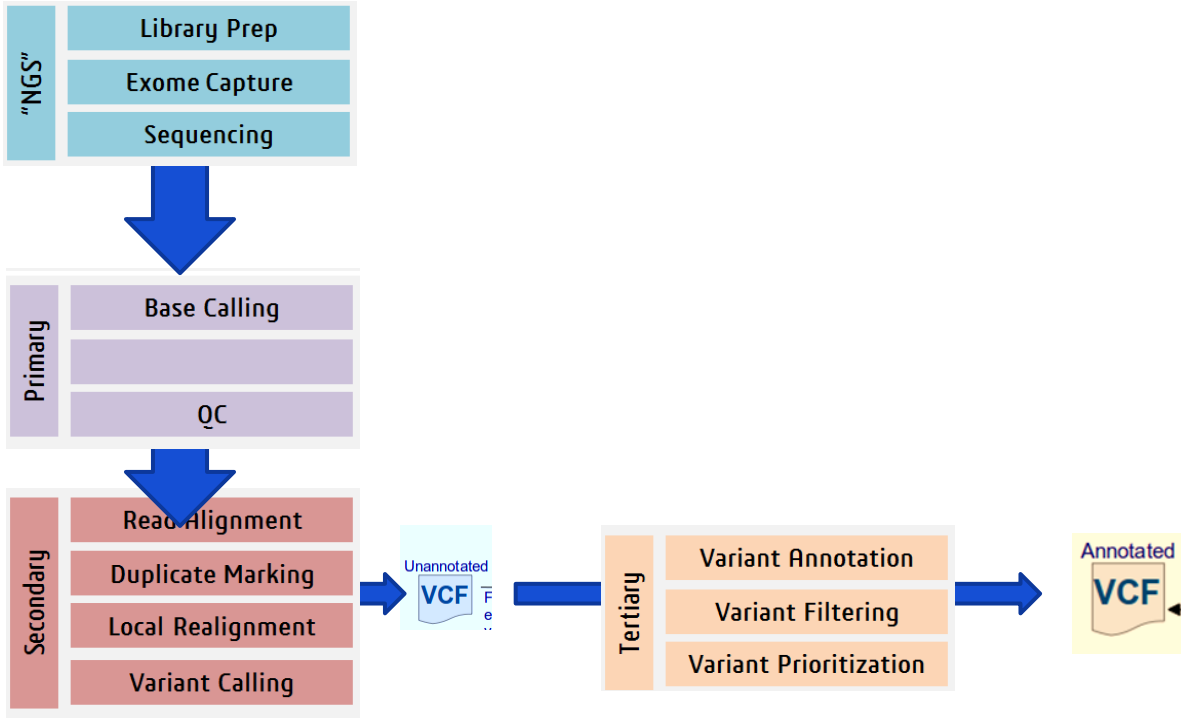
- Gene disease association
- Allele frequency
- In silico predictions

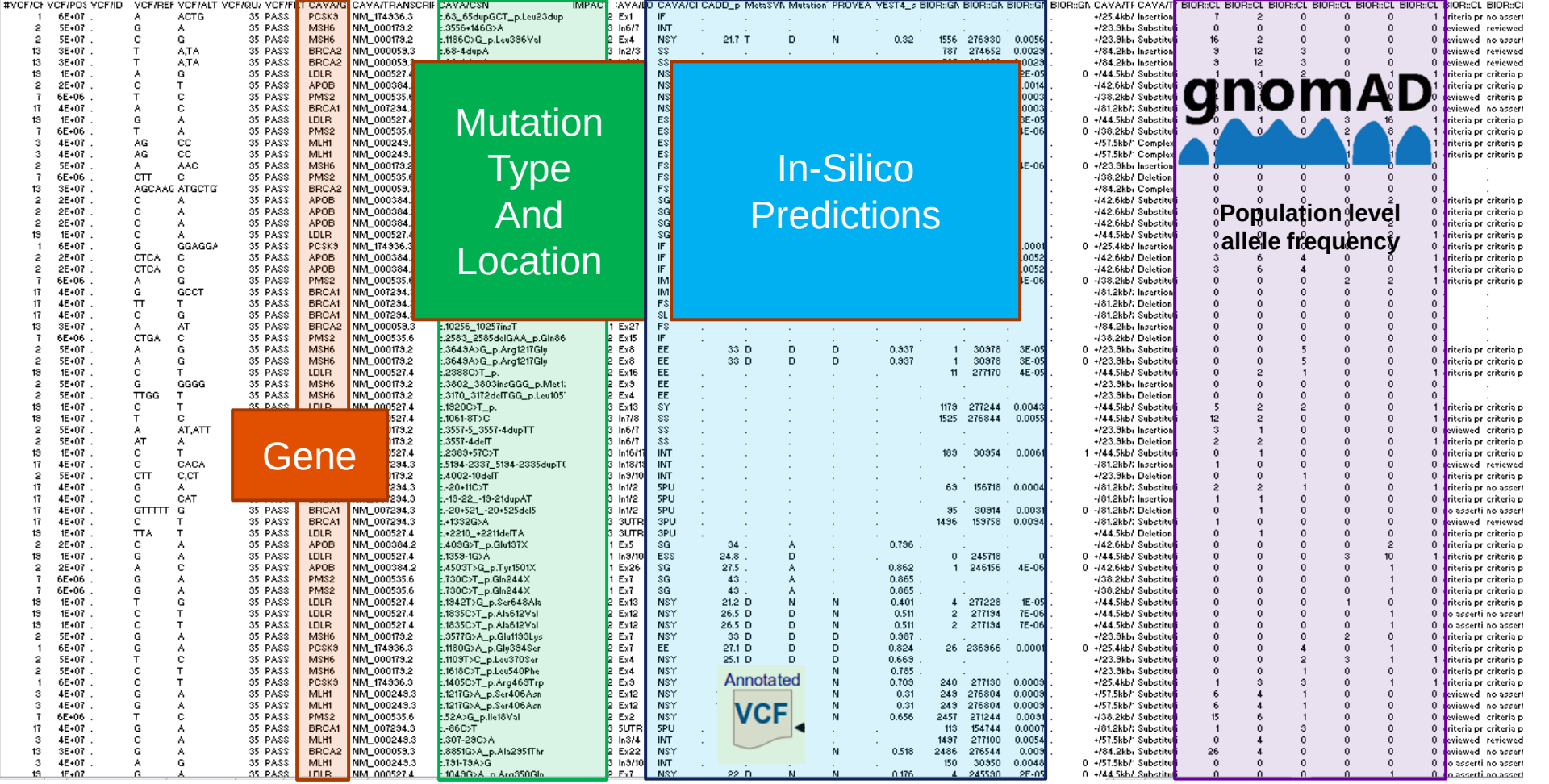




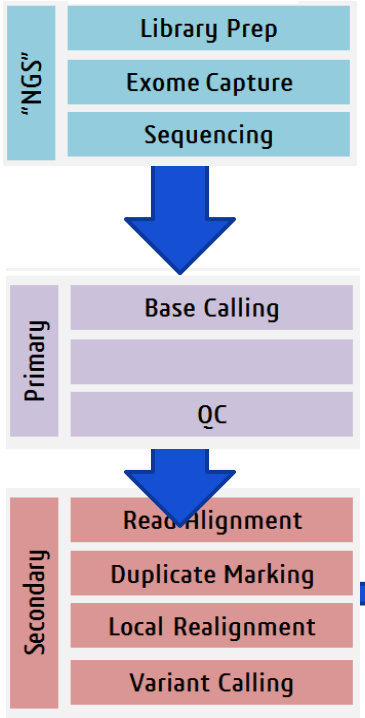

```
##FILTER=<ID=HETEROPLASMY,Description="Called heterozygous genotype on mitochondrial contig">
##FORMAT=<ID=GP,Number=G,Type=Float,Description="Estimated Genotype Probability">
##FILTER=<ID=IMP,Description="Set if true: IMP==1">
##FILTER=<ID=BOOSTED,Description="Set if true: BOOSTED==1">
##FILTER=<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-TA537BFRZC6B332GA6D'
chr1	55033879	.	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	47793163	.	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	5377709	.	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	47788826	.	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	.	AGCAAC	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	55033930	.	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	43124034	.	G	GCCT	35	PASS	.	GT:AD:DP:GQ:VAR_TYPE	0/1:500,500:1000:99:SNV
chr17	43124037	.	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	32338769	.	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	47806383	.	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	5937399	.	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	47793032	.	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

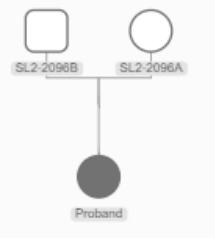

"NGS"
Library Prep
Exome Capture
Sequencing
Primary
Base Calling
QC
Secondary
Read Alignment
Duplicate Marking
Local Realignment
Variant Calling
Unannotated
VCF
Tertiary
Variant Annotation
Variant Filtering
Variant Prioritization
Annotated
VCF


Mutation Type And Location
In-Silico Predictions
gnomAD
Population level allele frequency
Gene
Annotated VCF

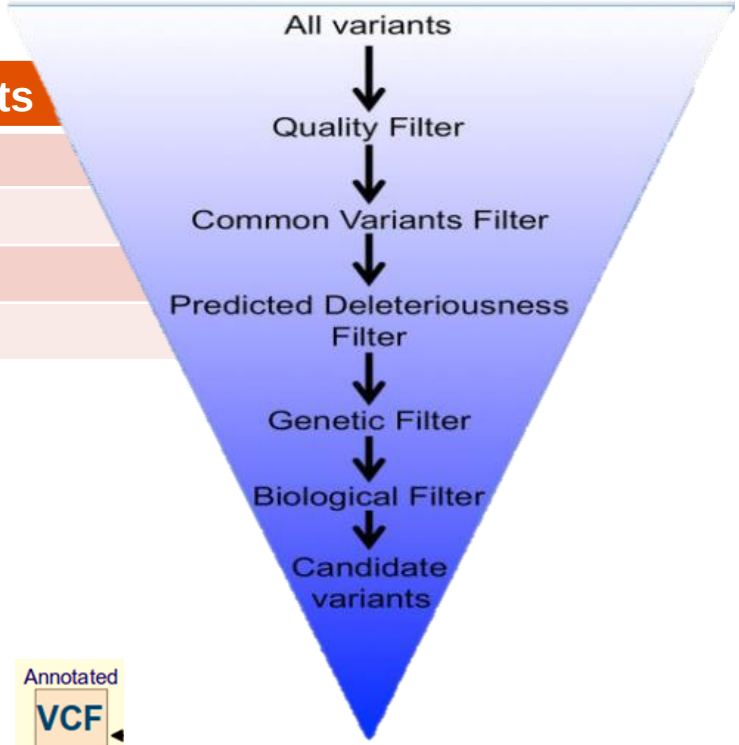
Class 2



WES Trio



Filter	Variants
None	125,746
DP>10, QUAL>20	25,541
MAF<0.01	1,263
Coding regions	551



HGVS nomenclature

Sequence Variant Nomenclature | Recommendations | Background Materials | Recent Additions | Contact Us | Version 20.05

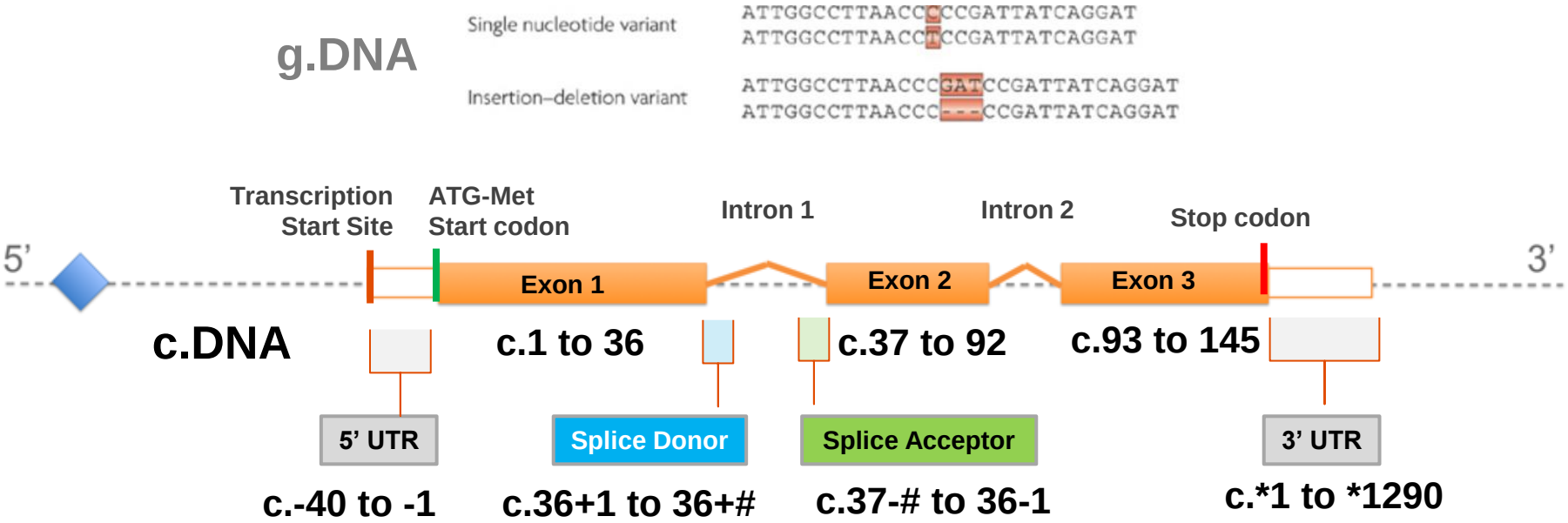
- a **letter prefix** is mandatory to indicate the type of reference sequence used. Accepted prefixes are;
  - "**c.**" for a coding DNA reference sequence
  - "**g.**" for a linear genomic reference sequence
  - "**m.**" for a mitochondrial DNA reference sequence
  - "**n.**" for a non-coding DNA reference sequence
  - "**o.**" for a circular genomic reference sequence
  - "**p.**" for a protein reference sequence
  - "**r.**" for an RNA reference sequence (transcript)
- numbering of the residues (nucleotide or amino acid) in relation to the reference sequence used should **follow the approved scheme** (*see Numbering*)

gDNA: Chr12(GRCh37): g.53703386C>G
cDNA: NM_015665.5(AAAS): c.809G>C
Protein: p.(Arg270Pro)

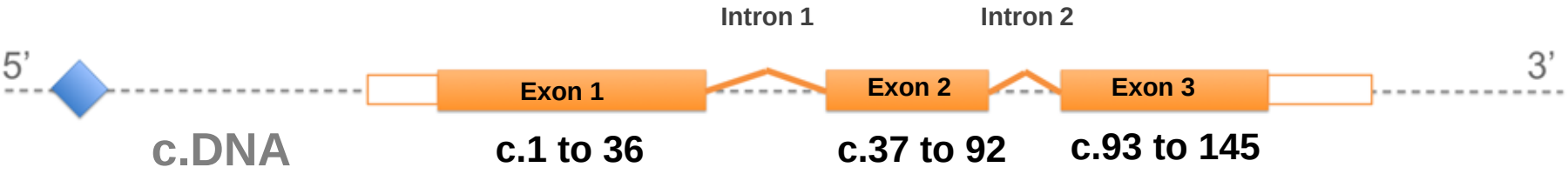
Other important considerations when looking at nomenclature

- Which reference genome is being used
- Transcript ID (one genomic variant can be described differently in the coding sequence depending on transcript)

cDNA examples



Protein level examples



Protein Consequences:

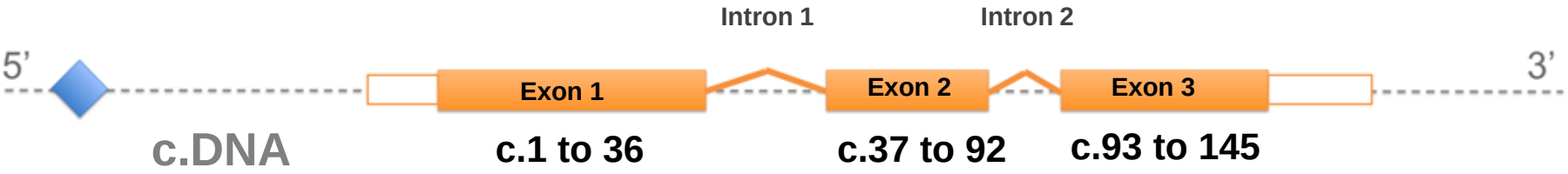
	No mutation	Point mutations			
		Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr

MSH2
c.1590A>G
p.(Glu530=)
“synonymous”

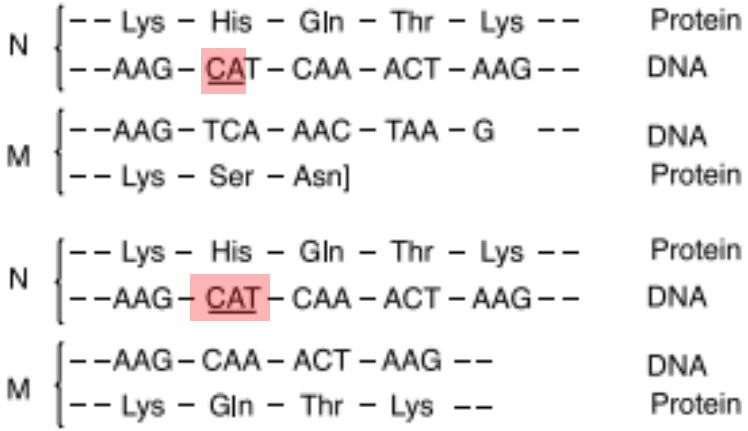
PMS2
c.730C>T
p.(Gln244*)
p.(Gln244Ter)

BRCA1
c.736T>G
p.(Leu246Val)

Protein level examples



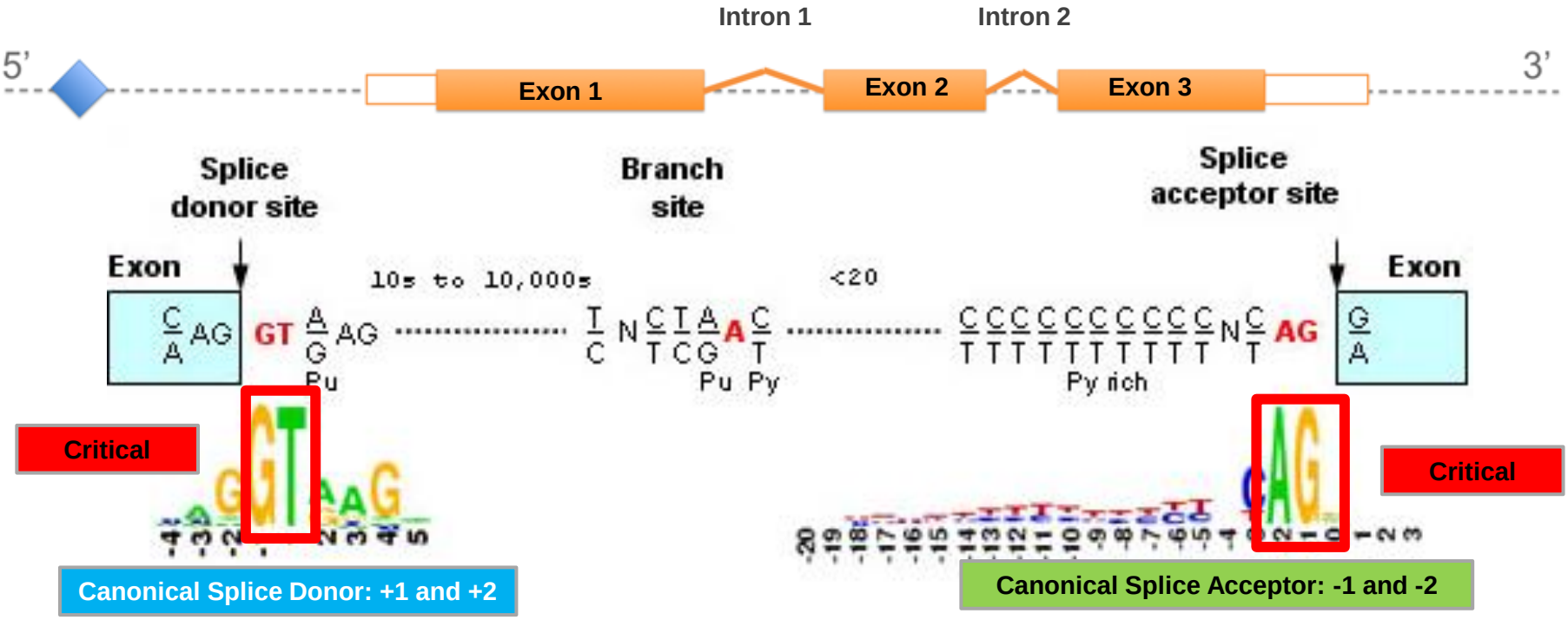
Protein Consequences



Frameshift
c.6474delCA
p.(Ser2159Asnfs*9)

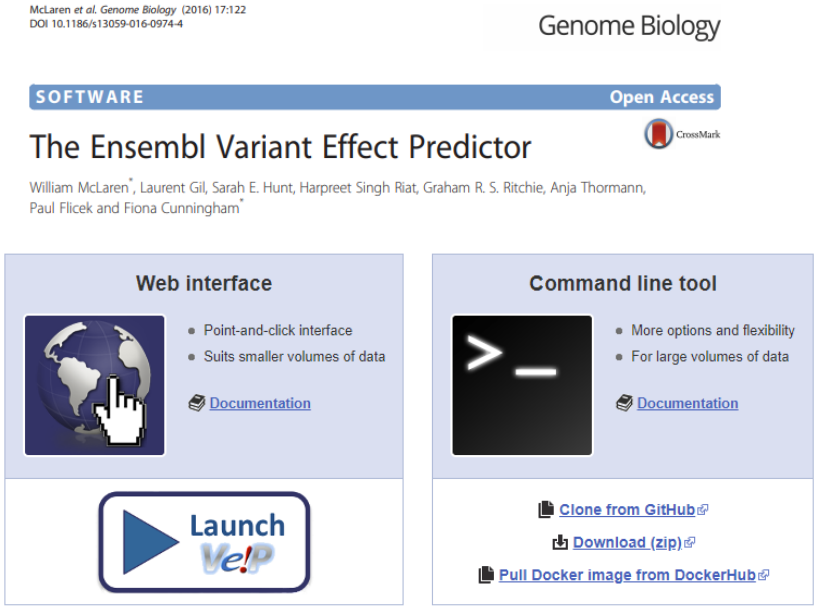
In-frame deletion
c.6639_6641delTGA
p.(Asp2213del)

Splicing disruptions

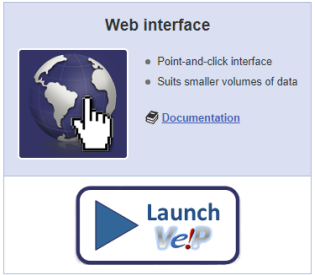


VEP: Variant Effect Predictor

- VEP determines:
  - Variant effect (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.
  - 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



VEP: Variant Effect Predictor



Variant Effect Predictor

New job

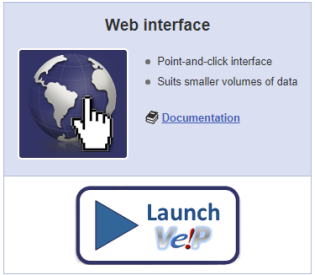
Recent jobs

Refresh

Show/hide columns (1 hidden)

Analysis	Jobs	Submitted at
Variant Effect Predictor	VEP analysis of pasted data in Homo_sapiens Done [View results]	06/02/2020, 18:59 (GMT)
Variant Effect Predictor	VEP analysis of pasted data in Homo_sapiens Done [View results]	24/01/2020, 21:53 (GMT)
Variant Effect Predictor	VEP analysis of pasted data in Homo_sapiens Done [View results]	24/01/2020, 21:43 (GMT)

VEP: Variant Effect Predictor



Species: Human (Homo sapiens)

Assembly: GRCh38.p13 *(If you are looking for VEP for Human GRCh37, please go to GRCh37 website.)*

Name for this job (optional):

Input data:

Either paste data:

```
17 43047665 . C T . . .
```

Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, SPDI

Or upload file: Choose File No file chosen

Or provide file URL:

Transcript database to use:

- Ensembl/GENCODE transcripts
- Ensembl/GENCODE basic transcripts
- RefSeq transcripts
- Ensembl/GENCODE and RefSeq transcripts

Identifiers ⊞ *Additional identifiers for genes, transcripts and variants*

Variants and frequency data ⊞ *Co-located variants and frequency data*

Additional annotations ⊞ *Addtional transcript, protein and regulatory annotations*

Predictions ⊞ *Variant predictions, e.g. SIFT, PolyPhen*

Filtering options ⊞ *Pre-filter results by frequency or consequence type*

Advanced options ⊞ *Settings to optimise VEP*

https://useast.ensembl.org/info/docs/tools/vep/index.html

VEP: Variant Effect Predictor

Species: Human (Homo sapiens)

Assembly: GRCh38.p13 *(If you are looking for VEP for Human GRCh37, please go to GRCh37 website.)*

Name for this job (optional):

Input data:

Either paste data:

```
17 43047665 . C T . . .
```

Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, SPDI

Or upload file: Choose File No file chosen

Or provide file URL:

Transcript database to use:

- Ensembl/GENCODE transcripts
- Ensembl/GENCODE basic transcripts
- RefSeq transcripts
- Ensembl/GENCODE and RefSeq transcripts

Identifiers *Additional identifiers for genes, transcripts and variants*

Variants and frequency data *Co-located variants and frequency data*

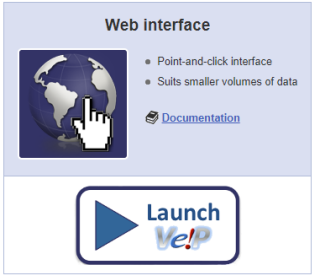
Analysis	Jobs	Submitted at
Variant Effect Predictor	VEP analysis of pasted data in Homo_sapiens Queued	26/02/2020, 04:29 (GMT)

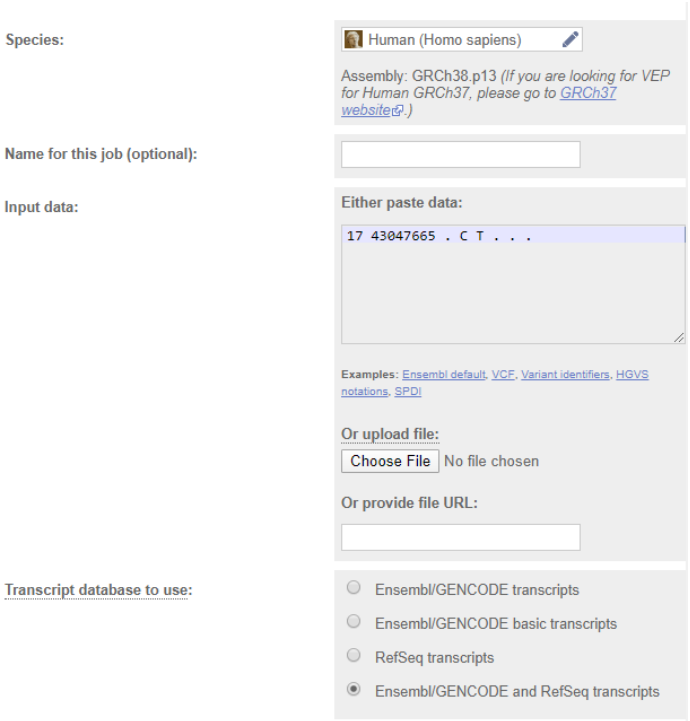
Filtering options *Pre-filter results by frequency or consequence type*

Advanced options *Settings to optimise VEP*

https://useast.ensembl.org/info/docs/tools/vep/index.html

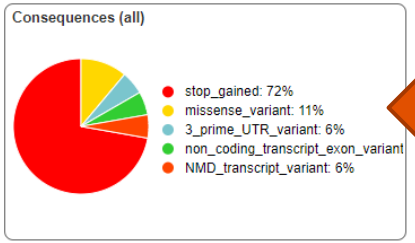
VEP: Variant Effect Predictor

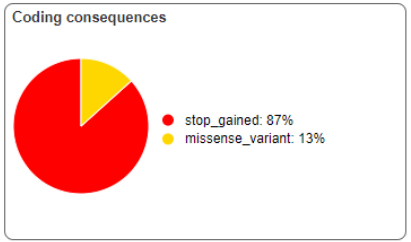




BRCA1
c.5445G>A
p.Trp1815X

Category	Count
Variants processed	1
Variants filtered out	0
Novel / existing variants	0 (0.0) / 1 (100.0)
Overlapped genes	2
Overlapped transcripts	17
Overlapped regulatory features	0





https://useast.ensembl.org/info/docs/tools/vep/index.html

VEP: Variant Effect Predictor

Uploaded variant	Location	Allele	Consequence	Symbol	Exon	HGVSc	HGVSp	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variant	Feature strand	MANE
.	17:43047665-43047665	T	stop_gained	BRCA1	21/22	ENST00000352993.7:c.2019G>A	ENSP00000312236.5:p.Trp673Ter	2138	2019	673	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	22/23	ENST00000357654.9:c.5445G>A	ENSP00000350283.3:p.Trp1815Ter	5558	5445	1815	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	NM_007294.4
.	17:43047665-43047665	T	3_prime_UTR_variant, NMD_transcript_variant	BRCA1	22/23	ENST00000461221.5:c.*5228G>A	-	5546	-	-	-	-	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	missense_variant	BRCA1	21/22	ENST00000468300.5:c.2059G>A	ENSP00000417148.1:p.Asp687Asn	2253	2059	687	D/N	GAC/AAC	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	23/24	ENST00000471181.7:c.5508G>A	ENSP00000418960.2:p.Trp1836Ter	5740	5508	1836	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	22/23	ENST00000491747.6:c.2133G>A	ENSP00000420705.2:p.Trp711Ter	2232	2133	711	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	21/22	ENST00000493795.5:c.5304G>A	ENSP00000418775.1:p.Trp1768Ter	5536	5304	1768	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	7/8	ENST00000586385.5:c.375G>A	ENSP00000465818.1:p.Trp125Ter	519	375	125	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	10/11	ENST00000591534.5:c.918G>A	ENSP00000467329.1:p.Trp306Ter	1020	918	306	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	4/5	ENST00000591849.5:c.144G>A	ENSP00000465347.1:p.Trp48Ter	301	144	48	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	14/15	ENST00000644379.1:c.1832G>A	ENSP00000496570.1:p.Trp611Ter	1832	1833	611	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-
.	17:43047665-43047665	T	stop_gained	BRCA1	22/23	NM_007294.4:c.5445G>A	NP_009225.1:p.Trp1815Ter	5558	5445	1815	W/*	TGG/TGA	rs397509284, CM042679, COSV58785802	-1	-

Gene: BRCA1 ENSG00000012048

MAYO CLINIC

Description	BRCA1 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1100]
Gene Synonyms	BRCC1, FANCS, PPP1R53, RNF53
Location	Chromosome 17: 43,044,295-43,170,245 reverse strand.

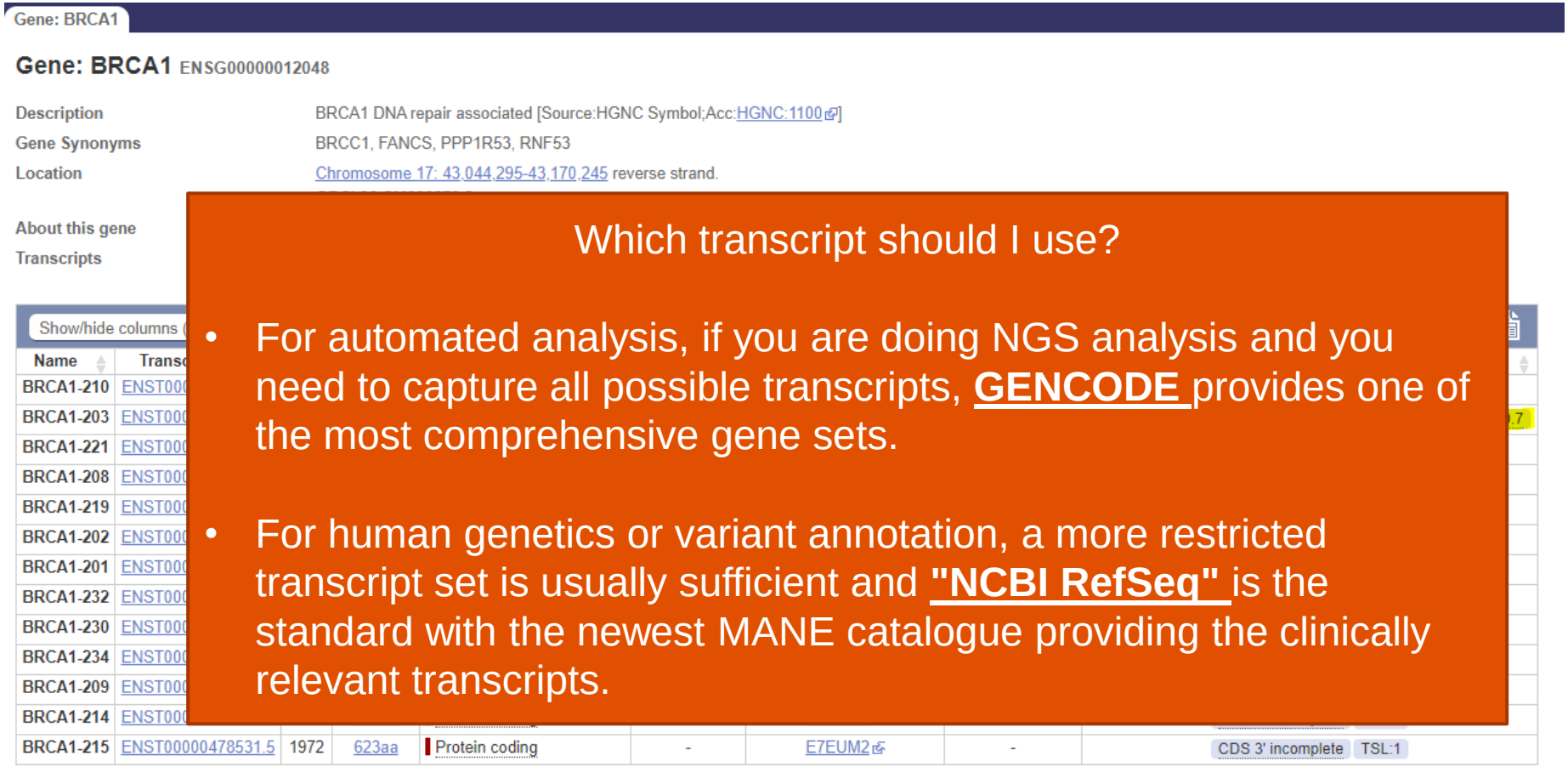
Gene: BRCA1

Gene: BRCA1 ENSG00000012048

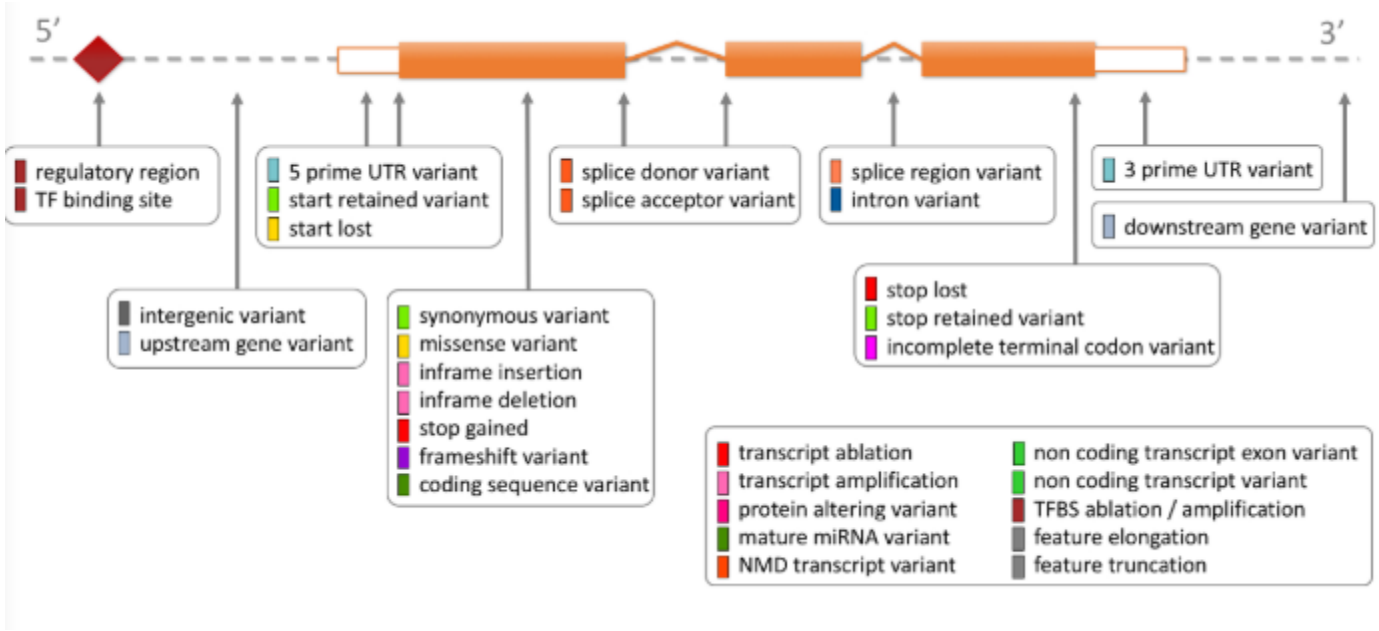
Description	BRCA1 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1100]
Gene Synonyms	BRCC1, FANCS, PPP1R53, RNF53
Location	Chromosome 17: 43,044,295-43,170,245 reverse strand. GRCh38:CM000679.2
About this gene	This gene has 34 transcripts (splice variants), 233 orthologues, is a member of 1 Ensembl protein family and is associated with 80 phenotypes.
Transcripts	Hide transcript table

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq Match	Flags
BRCA1-210	ENST00000471181.7	7270	1884aa	Protein coding	CCDS11456	P38398	-	TSL:1 GENCODE basic APPRIS P4
BRCA1-203	ENST00000357654.9	7088	1863aa	Protein coding	CCDS11453	P38398	NM_007294.4	TSL:1 GENCODE basic APPRIS ALT2 MANE Select v0.7
BRCA1-221	ENST00000493795.5	5732	1816aa	Protein coding	CCDS11459	P38398	-	TSL:5 GENCODE basic
BRCA1-208	ENST00000468300.5	3273	699aa	Protein coding	CCDS11455	P38398	-	TSL:1 GENCODE basic
BRCA1-219	ENST00000491747.6	2379	759aa	Protein coding	CCDS11454	A0A024R1V0 P38398	-	TSL:5 GENCODE basic
BRCA1-202	ENST00000354071.7	4497	1399aa	Protein coding	-	Q5YLB2	-	TSL:1 GENCODE basic
BRCA1-201	ENST00000352993.7	3668	721aa	Protein coding	-	A0A024R1Z8 P38398	-	TSL:5 GENCODE basic
BRCA1-232	ENST00000644379.1	2571	659aa	Protein coding	-	A0A2R8Y7V5	-	CDS 5' incomplete
BRCA1-230	ENST00000634433.1	2534	798aa	Protein coding	-	A0A0U1RRA9	-	CDS 3' incomplete TSL:5
BRCA1-234	ENST00000652672.1	2291	601aa	Protein coding	-	A0A494C182	-	CDS 3' incomplete
BRCA1-209	ENST00000470026.5	2108	649aa	Protein coding	-	E7EWN5	-	CDS 3' incomplete TSL:1
BRCA1-214	ENST00000477152.5	1980	622aa	Protein coding	-	E9PH68	-	CDS 3' incomplete TSL:1
BRCA1-215	ENST00000478531.5	1972	623aa	Protein coding	-	E7EUM2	-	CDS 3' incomplete TSL:1

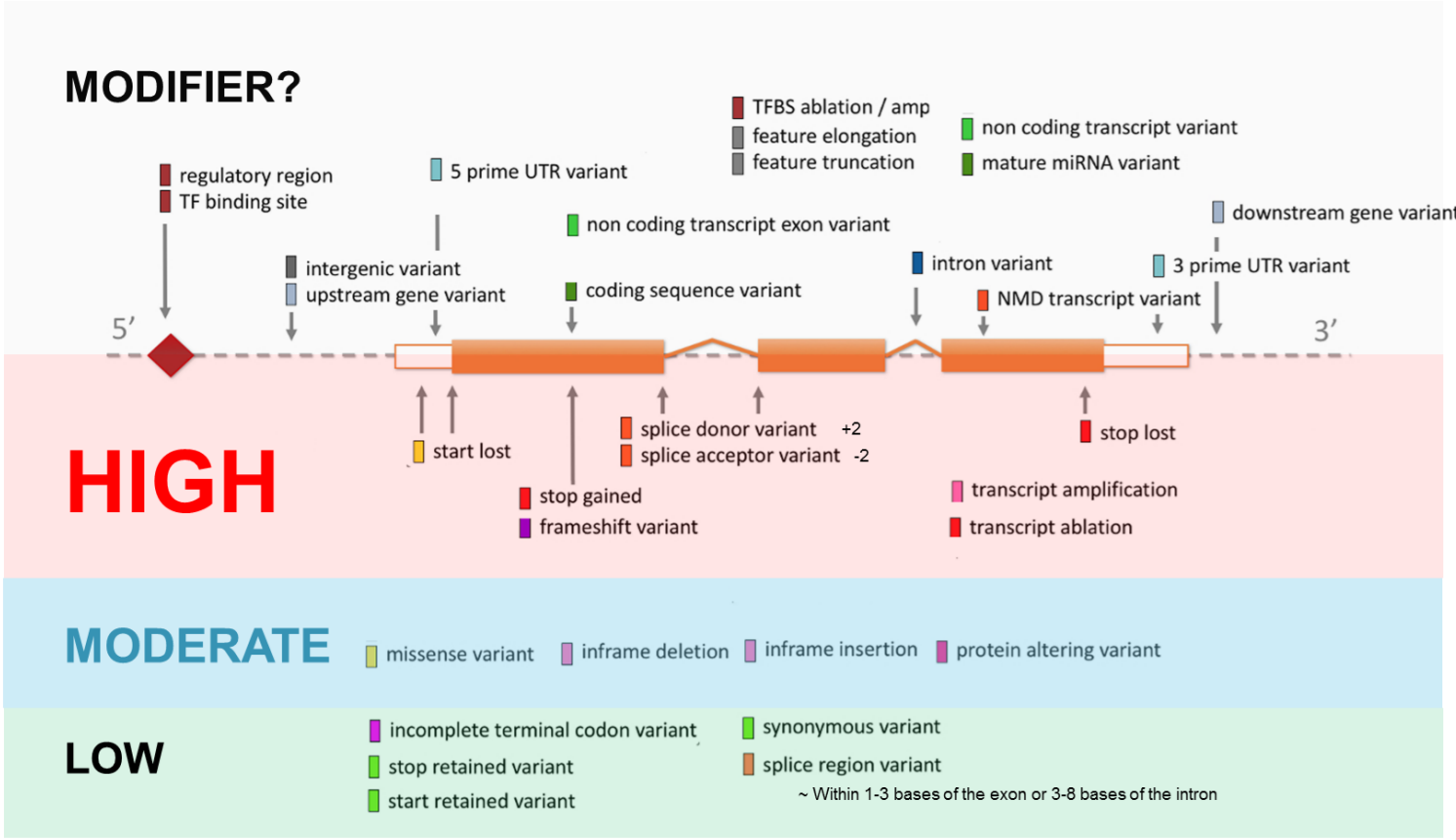

Gene: BRCA1
Gene: BRCA1 ENSG00000012048
Description
BRCA1 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1100]
Gene Synonyms
BRCC1, FANCS, PPP1R53, RNF53
Location
Chromosome 17: 43,044,295-43,170,245 reverse strand.
About this gene
Transcripts
Show/hide columns
Name
BRCA1-210
BRCA1-203
BRCA1-221
BRCA1-208
BRCA1-219
BRCA1-202
BRCA1-201
BRCA1-232
BRCA1-230
BRCA1-234
BRCA1-209
BRCA1-214
BRCA1-215
ENST00000478531.5
1972
623aa
Protein coding
E7EUM2
CDS 3' incomplete
TSL:1
Which transcript should I use?
For automated analysis, if you are doing NGS analysis and you need to capture all possible transcripts, GENCODE provides one of the most comprehensive gene sets.
For human genetics or variant annotation, a more restricted transcript set is usually sufficient and "NCBI RefSeq" is the standard with the newest MANE catalogue providing the clinically relevant transcripts.

Variant Severity: Variable definitions but helps prioritize

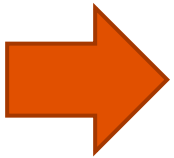


https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html

Variant Severity: Variable definitions but helps prioritize



Where is it? Which exon?
Regulatory elements nearby?
Visualization is key!



#CHROM	POS	ID	REF	ALT	QUAL	FILTER
chr1	55038873	.	A	ACTG	35	PASS
chr2	47805173	.	G	A	35	PASS
chr2	47733163	.	C	G	35	PASS
chr13	32319070	.	T	A,TA	35	PASS
chr19	11113686	.	A	G	35	PASS
chr2	21011802	.	C	T	35	PASS
chr7	5977709	.	T	C	35	PASS
chr17	43094795	.	A	C	35	PASS
chr19	11102787	.	G	A	35	PASS
chr7	6003794	.	T	A	35	PASS
chr3	37028782	.	AG	CC	35	PASS
chr2	47798826	.	A	AAC	35	PASS
chr7	5987451	.	CTT	C	35	PASS
chr13	32340378	.	AGCAAC	ATGCTG	35	PASS
chr2	21038086	.	C	A	35	PASS

Experimentally Defined Genomic Features:
UCSC Genome Browser - Visualization

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

UCSC

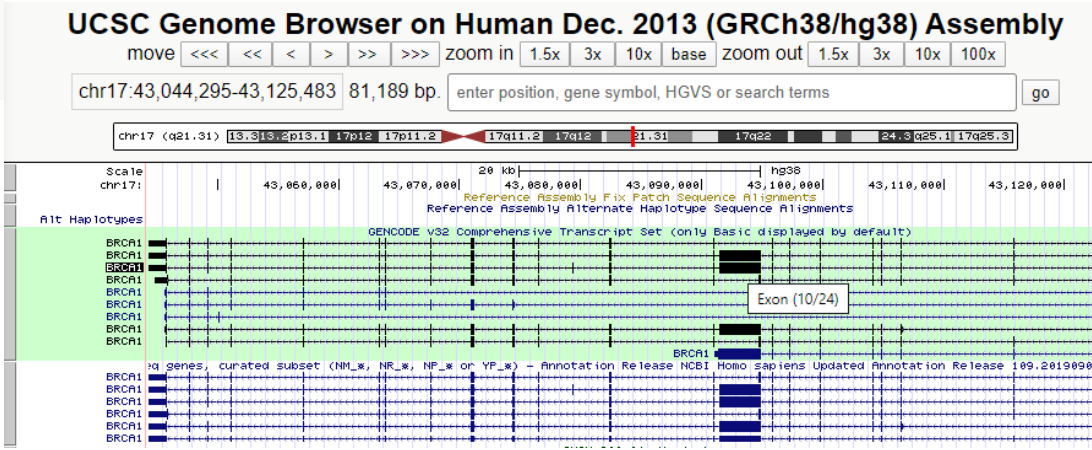
D756–D761 *Nucleic Acids Research, 2020, Vol. 48, Database issue* *Published online 6 November 2019*
doi: 10.1093/nar/gkz1012

UCSC Genome Browser enters 20th year

Christopher M. Lee[1,*], Galt P. Barber[1], Jonathan Casper[1], Hiram Clawson[1], Mark Diekhans[1], Jairo Navarro Gonzalez[1], Angie S. Hinrichs[1], Brian T. Lee[1], Luis R. Nassar[1], Conner C. Powell[1], Brian J. Raney[1], Kate R. Rosenbloom[1], Daniel Schmelter[1], Matthew L. Speir[1], Ann S. Zweig[1], David Haussler[1,2], Maximilian Haeussler[1], Robert M. Kuhn[1] and W. James Kent[1]

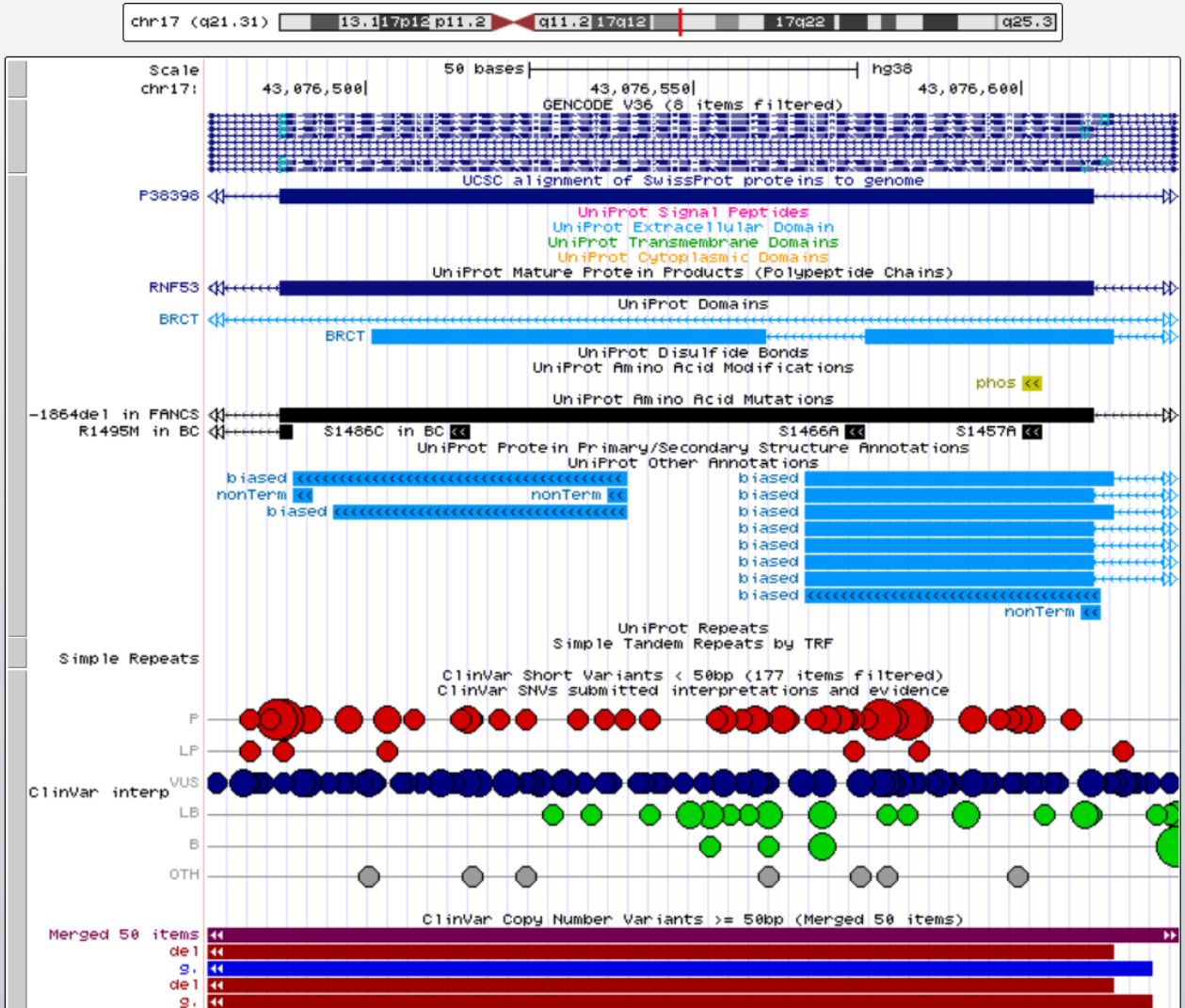
[1]Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA and [2]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA

Received September 11, 2019; Revised October 16, 2019; Editorial Decision October 17, 2019; Accepted October 25, 2019



- Web-based viewer for genome sequence data and annotations.
- Steadily added data and software features to the website since first coming online in July 2000, and currently hosts 206 assemblies from 105 species

https://genome.ucsc.edu/index.html


chr17 (q21.31)
13.1 17p12 p11.2
q11.2 17q12
17q22
q25.3
Scale
chr17:
50 bases
hg38
43,076,500
43,076,550
43,076,600
GENCODE V36 (8 items filtered)
UCSC alignment of SwissProt proteins to genome
P38398
UniProt Signal Peptides
UniProt Extracellular Domain
UniProt Transmembrane Domains
UniProt Cytoplasmic Domains
UniProt Mature Protein Products (Polypeptide Chains)
RNF53
UniProt Domains
BRCT
BRCT
UniProt Disulfide Bonds
UniProt Amino Acid Modifications
phos
UniProt Amino Acid Mutations
-1864del in FANCS
R1495M in BC
S1486C in BC
S1466A
S1457A
UniProt Protein Primary/Secondary Structure Annotations
UniProt Other Annotations
biased
nonTerm
UniProt Repeats
Simple Tandem Repeats by TRF
Simple Repeats
ClinVar Short Variants < 50bp (177 items filtered)
ClinVar SNVs submitted interpretations and evidence
ClinVar interp
P
LP
VUS
LB
B
OTH
ClinVar Copy Number Variants >= 50bp (Merged 50 items)
Merged 50 items
del
dup
del
dup

Variant Interpretation: Rationale

- Is a previously published variant associated with a disease phenotype pathogenic?
- Are all variants observed in a control population benign?
- What evidence to we use to ultimately classify a variant?
- How do we ensure consistency among clinicians, clinical laboratories, and researchers?

15-20 years ago…

- If you thought a gene may be implicated in a specific disease:
  - You could screen a cohort of patients and look for variants in said gene
  - If you identified a variant in a patient and did not find it in 50 control samples (100 alleles!!!) you could deem this as a pathogenic variant
- Does this make sense statistically?

Common framework and criteria for germline variant classification

ACMG

American College of Medical Genetics and Genomics

Translating Genes Into Health®

AMP ASSOCIATION FOR MOLECULAR PATHOLOGY

cap

Common framework and criteria for germline variant classification

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

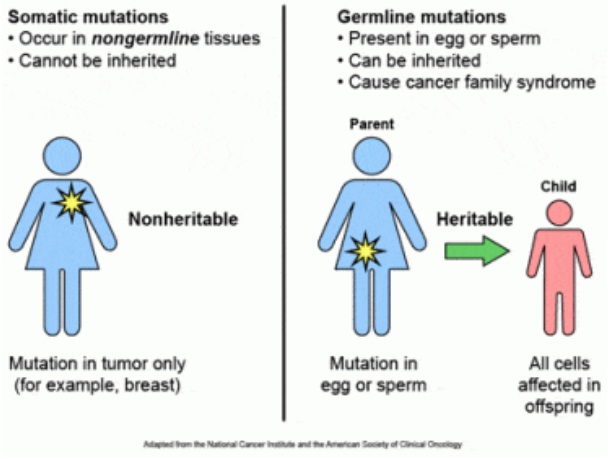
Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

2015 ACMG guidelines

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

- These recommendations primarily apply to genetic tests used in clinical laboratories including **genotyping, single genes, panels, exomes and genomes.**
- **It is not intended** for the interpretation of **somatic variation, pharmacogenomic variants, or variants in genes associated with multigenic non-Mendelian complex disorders.**



Class
Pathogenic
Likely Pathogenic
Variant of Uncertain Significance
Likely Benign
Benign

2015 ACMG guidelines

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

- These recommendations primarily apply to genetic tests used in clinical laboratories including **genotyping, single genes, panels, exomes and genomes.**
- **It is not intended** for the interpretation of **somatic variation, pharmacogenomic variants, or variants in genes associated with multigenic non-Mendelian complex disorders.**
- Care must be taken when applying these rules to candidate genes ("genes of uncertain significance", **GUS**)
- This report recommends the use of specific standard terminology: 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' to describe variants identified in **Mendelian disorders.**

Class
Pathogenic
Likely Pathogenic
Variant of Uncertain Significance
Likely Benign
Benign

We interpret by sorting variants into categories

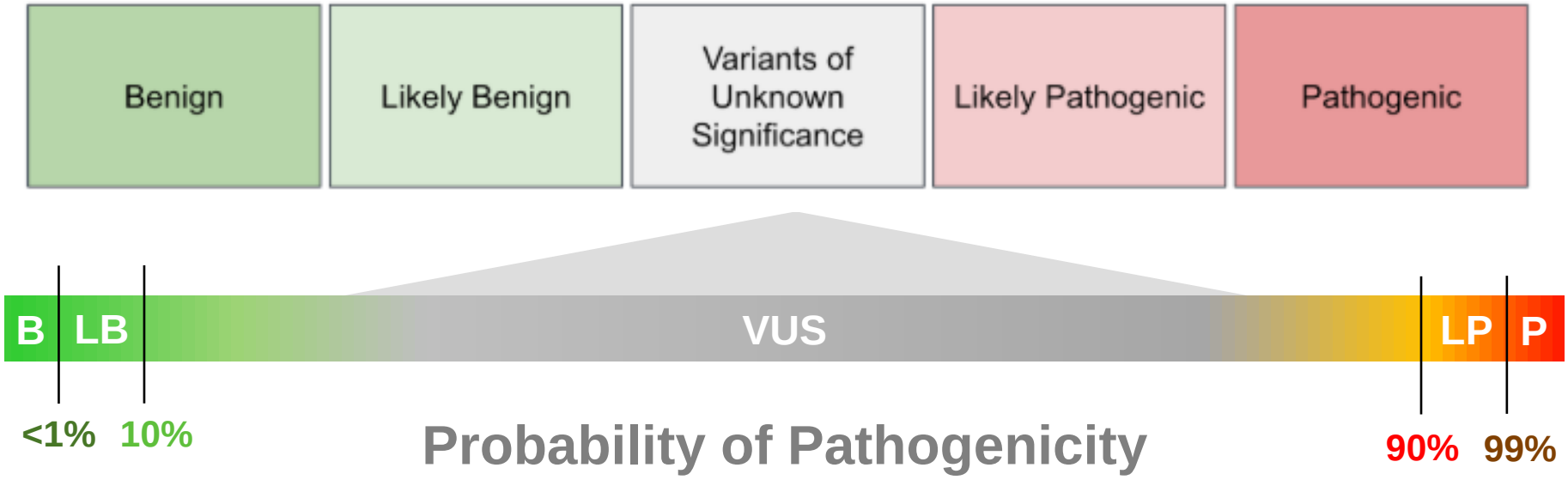
***Variant:** An alteration in the normal sequence of a gene:

Benign	Variants of Unknown Significance	Pathogenic
A variant that does not appear to have a *deleterious* effect often associated with a "normal" or no human phenotype.	A variant whose association with disease risk is unknown.	A variant which is proven to be deleterious to protein or gene function and is associated with a particular human phenotype or disease

Caution: **A deleterious variant is not always pathogenic or disease causing.**

What is `Likely`?

The rules proposed to classify sequence variants follows is a heuristic system for variant classification that is compatible with a formal, quantitative, Bayesian classifier.



2015 ACMG Guidelines

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

- The guideline defined 28 criteria, with codes that addressed types of variant evidence. Each evidence type or criterion code was assigned a direction, benign (B) or pathogenic (P), and a level of strength: stand-alone (A), very strong (VS), strong (S), moderate (M), or supporting (P).

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

ACMG 2015 guidelines discrete criteria have a strong quantitative correlation with the odds of pathogenicity of a variant.

		BENIGN CRITERIA		PATHOGENIC CRITERIA			
Strength of Evidence		Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Odds of Pathogenicity*		-18.7	-2.08	2.08	4.33	18.7	350.0
Evidence Category and Corresponding ACMG/AMP Codes	Population Data	*BA1*+ BS1 BS2			PM2	PS4	
	Allelic Evidence & Co-Segregation Data	BS4	BP2 BP5	PP1 →	PM3 PM6	PS2	
	Computation & Predictive Data		BP1 BP3 BP4 BP7	PP2 PP3	PM1 PM4 PM5	PS1	PVS1
	Functional Data	BS3				PS3	
	Other		BP6	PP4 PP5			

Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND*
	(a) ≥1 Strong (PS1–PS4) *OR*
	(b) ≥2 Moderate (PM1–PM6) *OR*
	(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR*
	(d) ≥2 Supporting (PP1–PP5)
	(ii) ≥2 Strong (PS1–PS4) *OR*
	(iii) 1 Strong (PS1–PS4) *AND*
	(a)≥3 Moderate (PM1–PM6) *OR*
	(b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
	(c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR*
	(ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR*
	(iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR*
	(iv) ≥3 Moderate (PM1–PM6) *OR*
	(v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR*
	(vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR*
	(ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR*
	(ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR
	(ii) the criteria for benign and pathogenic are contradictory

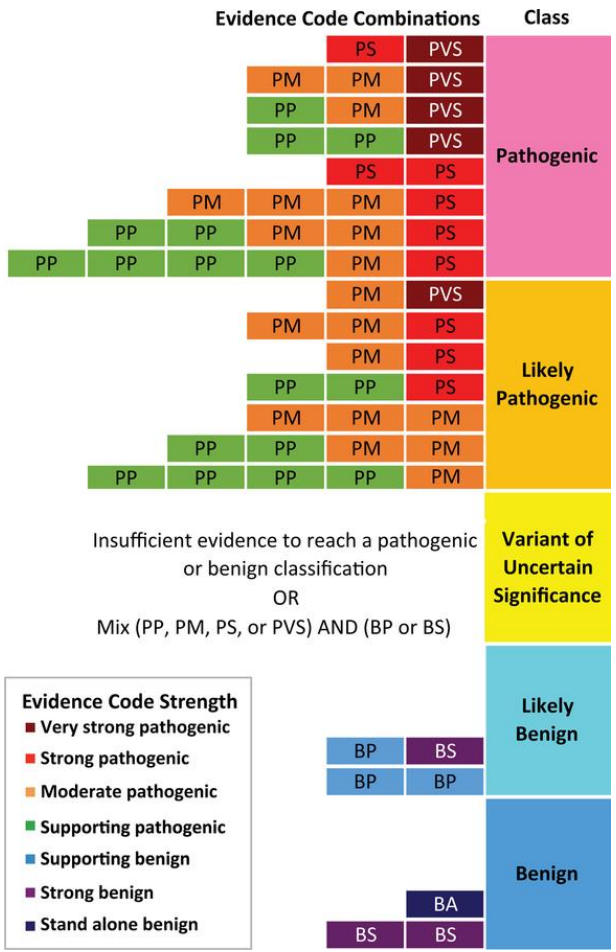


Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

PS4 + PM2 + PP1

Pathogenic

Level of Evidence: *Strong*

= Likely Pathogenic

Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met *OR* (ii) the criteria for benign and pathogenic are contradictory

PS4 + PM2 + PP1

Pathogenic

Level of Evidence: *Strong*

= Likely Pathogenic

BP4

Benign

Level of Evidence: *Supporting*

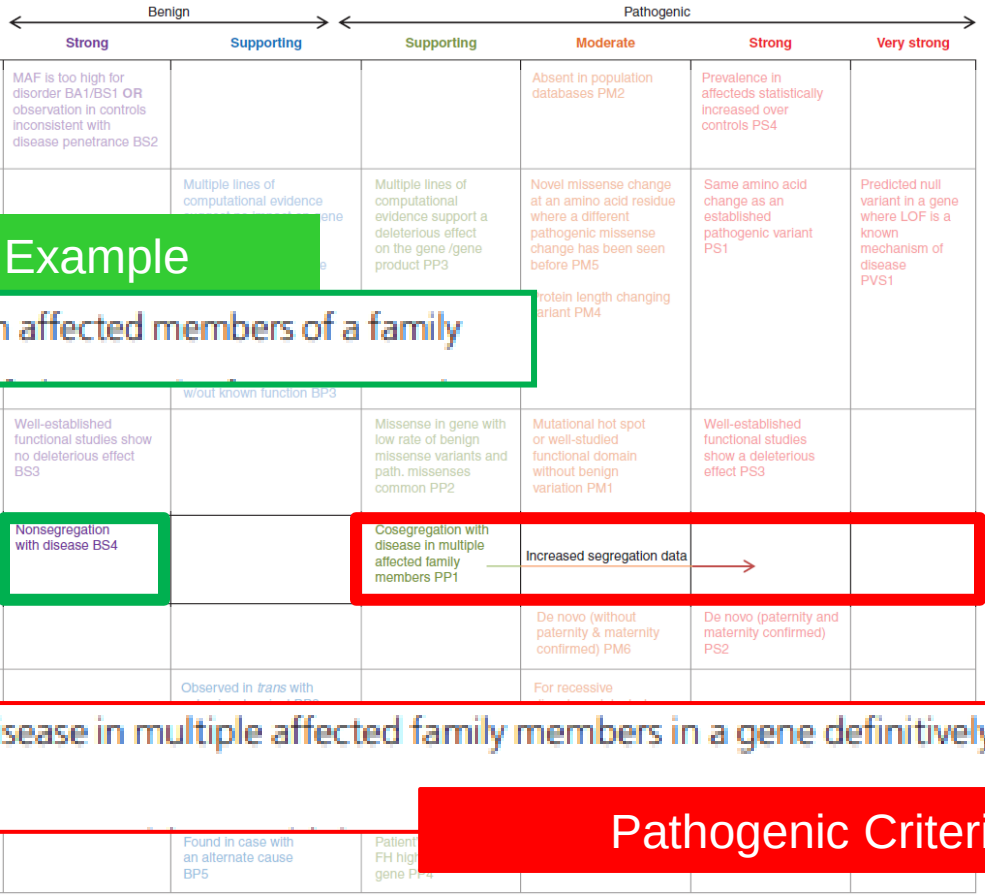
Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Conflicting evidence example:

PS4 + PM2 + BP2 + BP4

= Variant of Uncertain Significance (VUS or VOUS)



Benign Criteria Example

BS4 Lack of segregation in affected members of a family

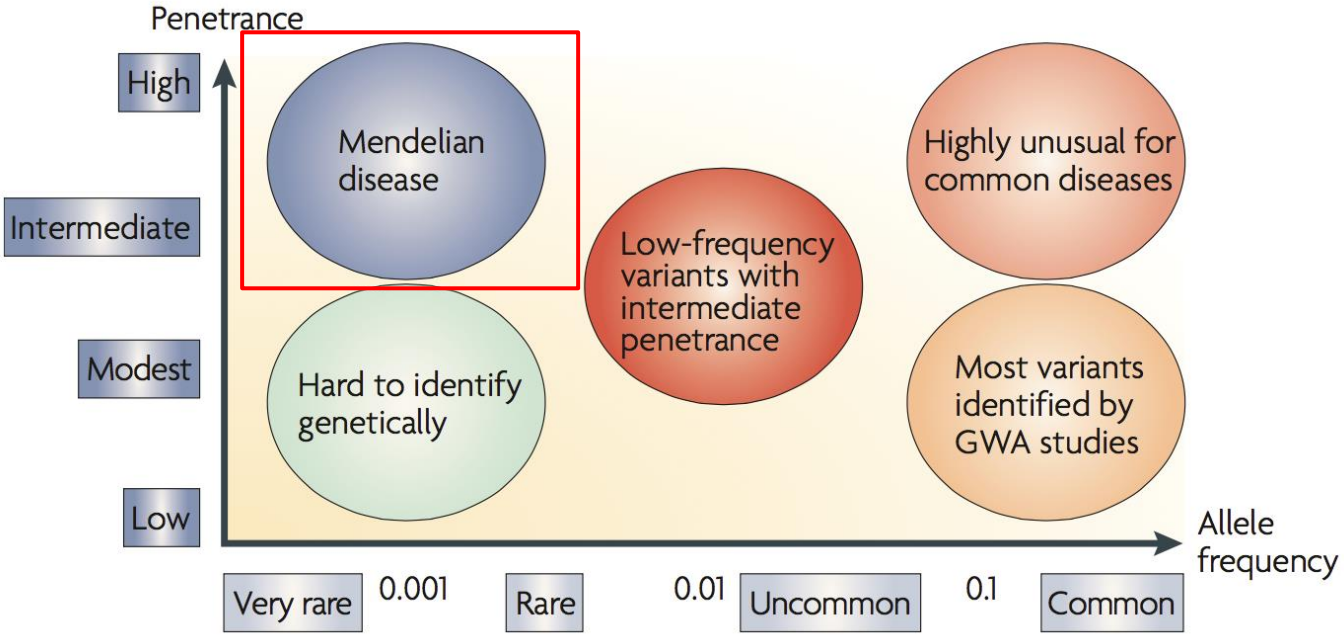
Segregation Data Criterion

PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease

Pathogenic Criteria Example

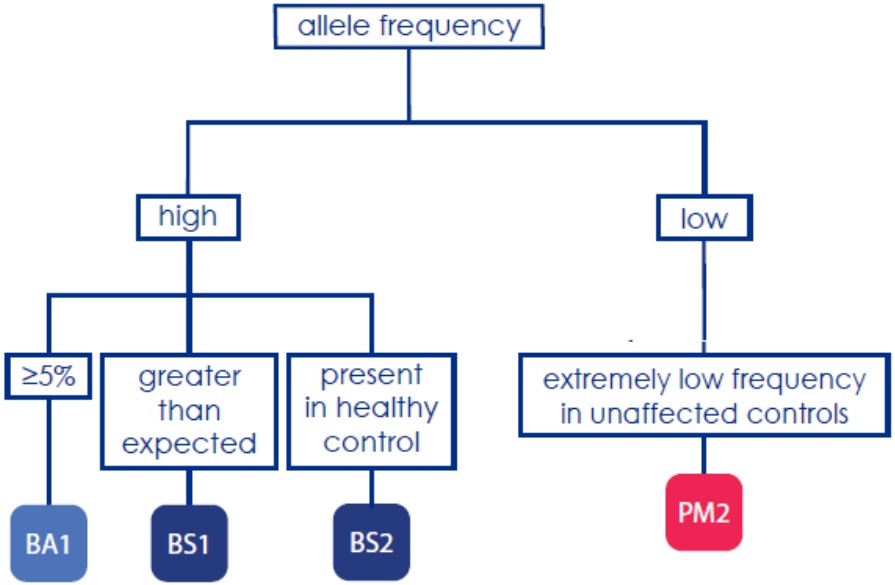
	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

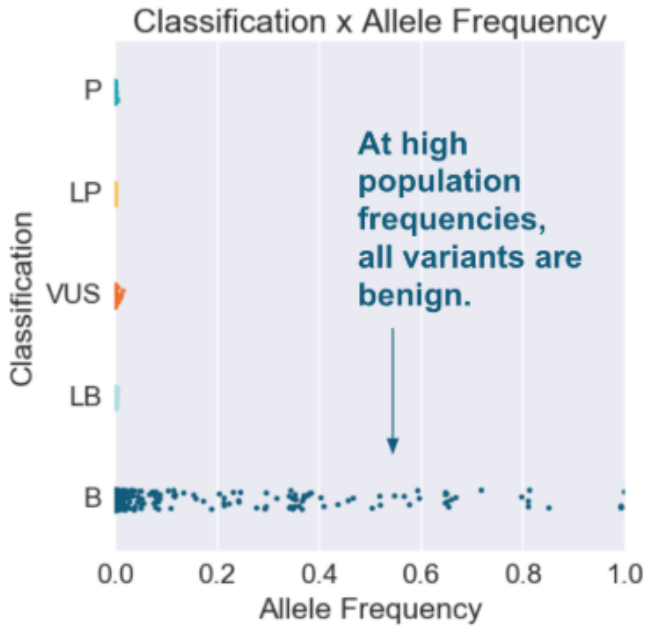
Allele Frequency (BA1, BS1, PM2)
How common is this variant?



https://gnomad.broadinstitute.org/

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	





The cutoffs of each of these criteria depends on many factors such as: Prevalence of disease, age of onset, and penetrance

Allele Frequency (BA1, BS1, PM2)
How common is this variant?

BA1:

>5% allele frequency in any general continental population of at least 2,000 alleles for a gene without a gene or variant specific recommendation.

BS1:

Allele frequency is greater than expected for disorder (lower than BA1)

PM2:

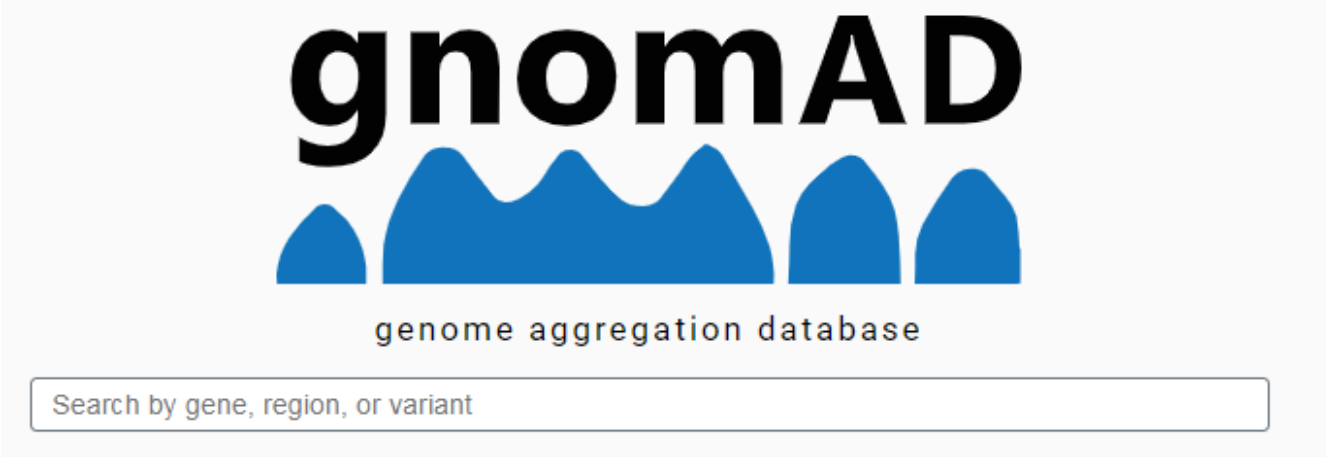
Absent from controls (or at an extremely low frequency if recessive).

https://gnomad.broadinstitute.org/

Exome Aggregation Consortium (ExAC)

gnomAD browser

125,748 exome sequences
15,708 whole-genome sequences
141,456 individuals



http://gnomad.broadinstitute.org/

gnomAD

Single nucleotide variant: 6-51944718-G-A (GRCh37)

	Exomes	Genomes	Total
Filter	Pass	Pass	
Allele Count	2	1	3
Allele Number	251434	31396	282830
Allele Frequency	0.000007954	0.00003185	0.00001061
Popmax Filtering AF (95% confidence)	–	–	
Number of homozygotes	0	0	0

References

- dbSNP (rs727504096)
- UCSC
- ClinVar (177240)

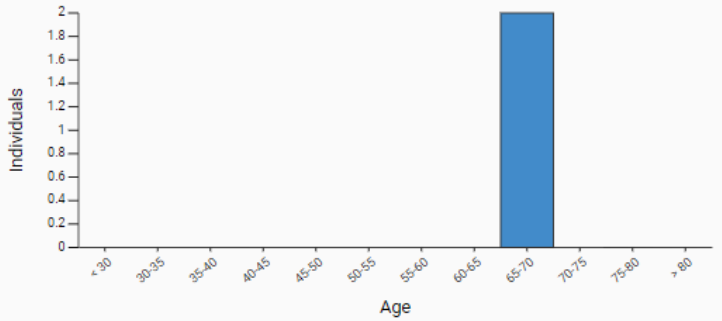
Report

- Report this variant
- Request additional information

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	2	24962	0	0.00008012
East Asian	1	19950	0	0.00005013
Latino	0	35438	0	0.000
Ashkenazi Jewish	0	10370	0	0.000
European (Finnish)	0	25122	0	0.000
European (non-Finnish)	0	129146	0	0.000
Other	0	7226	0	0.000
South Asian	0	30616	0	0.000
Male	2	153358	0	0.00001304
Female	1	129472	0	0.000007724
Total	**3**	**282830**	**0**	**0.00001061**

Include: Exomes Genomes

Age Distribution



Heterozygous Variant Carriers

Exomes Genomes

List of nine variants for which there was some evidence of pathogenicity even though the MAF was high for these variants!

Gene	Variant	Classification	applied (not including BA1 or BS1)	ClinVar ID	ClinGen Allele Registry ID	Chr	Position	Ref	Alt	ExAC Source Pop	ExAC Source Pop MAF	ClinVar disease entry
ACAD9	NM_014049.4: c.-44_-41dupTAAG	VUS	PS3_Supporting; BS2	1018	CA114709	3	128,598,490	C	CTAAG	AFR	0.1261	Deficiency of Acyl-CoA dehydrogenase family, member 9
GJB2	NM_004004.5: c.109G>A (p.Val37Ile)	Pathogenic	PS4; PP1_Strong; PM3_VeryStrong; PS3_Moderate	17023	CA172210	13	20,763,612	C	T	EAS	0.07242	Deafness, autosomal recessive
HFE	NM_000410.3: c.187C>G (p.His63Asp)	Pathogenic*	PS4	10	CA113797	6	26,091,179	C	G	NFE	0.1368	Hereditary hemochromatosis
HFE	NM_000410.3: c.845G>A (p.Cys282Tyr)	Pathogenic*	PS4; PP3	9	CA113795	6	26,093,141	G	A	NFE	0.05135	Hereditary hemochromatosis
MEFV	NM_000243.2: c.1105C>T (p.Pro369Ser)	VUS	PM3; PM5	2551	CA280114	16	3,299,586	G	A	EAS	0.07156	Familial Mediterranean fever
MEFV	NM_000243.2: c.1223G>A (p.Arg408Gln)	VUS	PM3; PM5	2552	CA280116	16	3,299,468	C	T	EAS	0.05407	Familial Mediterranean fever
PIBF1	NM_006346.2: c.1214G>A (p.Arg405Gln)	VUS	PM3; BS2	217689	CA210261	13	73,409,497	G	A	AMR	0.09858	Joubert syndrome
ACADS	NM_000017.3: c.511C>T (p.Arg171Trp)	VUS	PS3_Moderate; PM3; PP3	3830	CA312214	12	121,175,678	C	T	FIN #	0.06589	Deficiency of butyryl-CoA dehydrogenase
BTD	NM_000060.4: c.1330G>C (p.Asp444His)	Pathogenic	PS3; PM3_Strong; PP3; PP4	1900	CA090886	3	15,686,693	G	C	FIN #	0.05398	Biotidinase deficiency

*ACMG/AMP criteria selected does not match the classification as these variants are common low-penetrant variants and the ACMG/AMP guidelines are not designed for this variant type

Detected at >5% MAF only in Finnish population (see text).

Genomic coordinates on build GRCh37

AFR: African/African American, EAS: East Asian, NFE: Non-Finnish European, AMR: Latino, FIN=Finnish

Allele Frequency (BA1, BS1, PM2)
Other considerations…

What is a control population? Unselected?

For dominant disorders (AD):
Adult-onset disorders could be represented in the gnomAD database in still unaffected probands. Instead of controls the database could be refer to better as "general population".

PM2:
Absent from controls (or at an extremely low frequency if recessive).

ClinGen Sequence Variant Interpretation Recommendation for PM2 - Version 1.0
Working Group Page: https://clinicalgenome.org/working-groups/sequence-variant-interpretation/
Date Approved: September 4, 2020

SVI Recommendation for Absence/Rarity (PM2) - Version 1.0

The ClinGen Sequence Variant Interpretation (SVI) Working Group proposes decreasing the weight of criterion PM2 ("Absent from controls, or at extremely low frequency if recessive, in Exome Sequencing Project, 1000Genomes Project, or Exome Aggregation Consortium") from a Moderate strength level to a Supporting strength level (PM2_Supporting).

PM2_supporting

Demotion of this category to supporting is currently recommended.

https://clinicalgenome.org/site/assets/files/5182/pm2_-_svi_recommendation_-_approved_sept2020.pdf

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

- Focus on the "Pathogenic criteria"

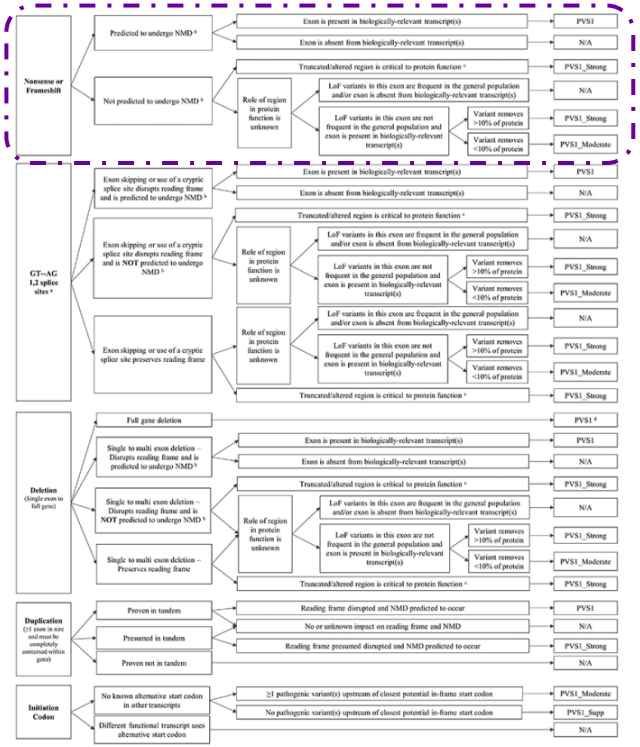
Richards CS et al. *Gen Med.* 2015;17:405-423

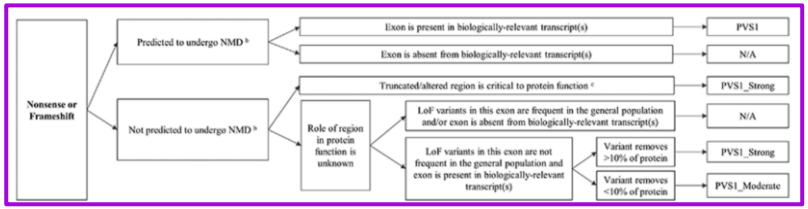
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

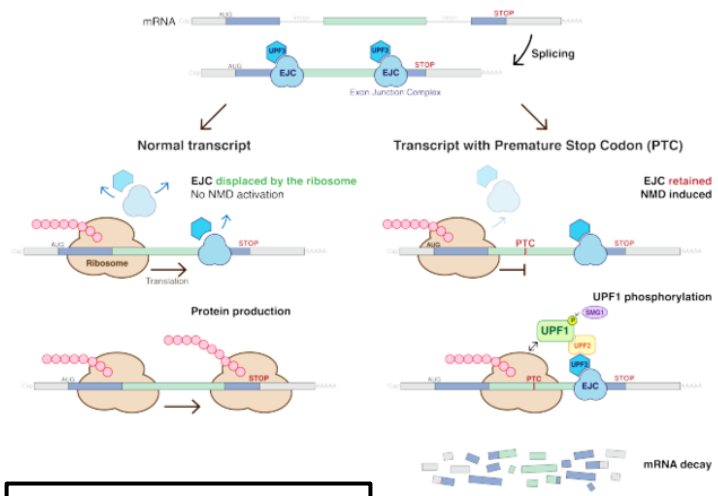
Loss of Function Criteria (PVS1) (only "very strong" level of evidence)

- Null variant in a gene where loss of function (LoF) is a known mechanism of disease.

PVS1



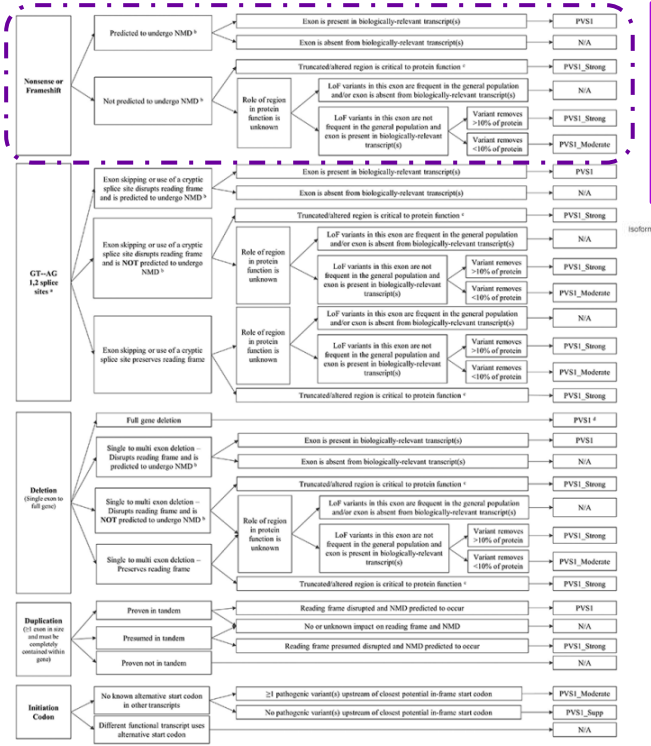


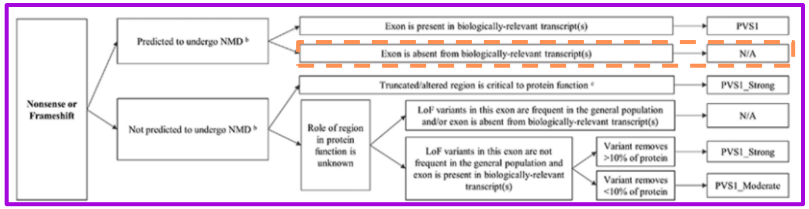


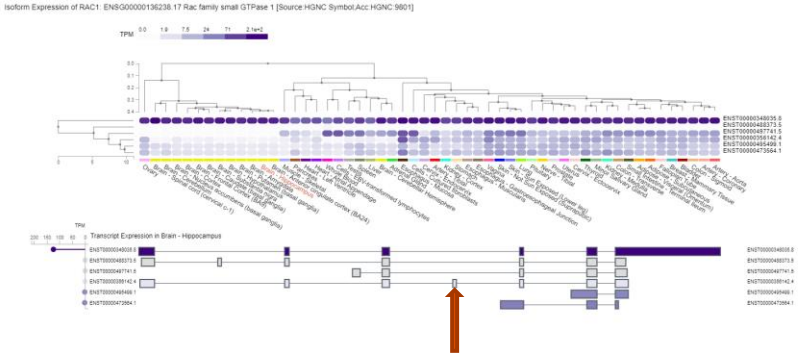
If the termination codon is downstream of or within about 50 nucleotides of the final exon-junction complex then the transcript is translated normally.

https://onlinelibrary.wiley.com/doi/full/10.1002/humu.236

PVS1

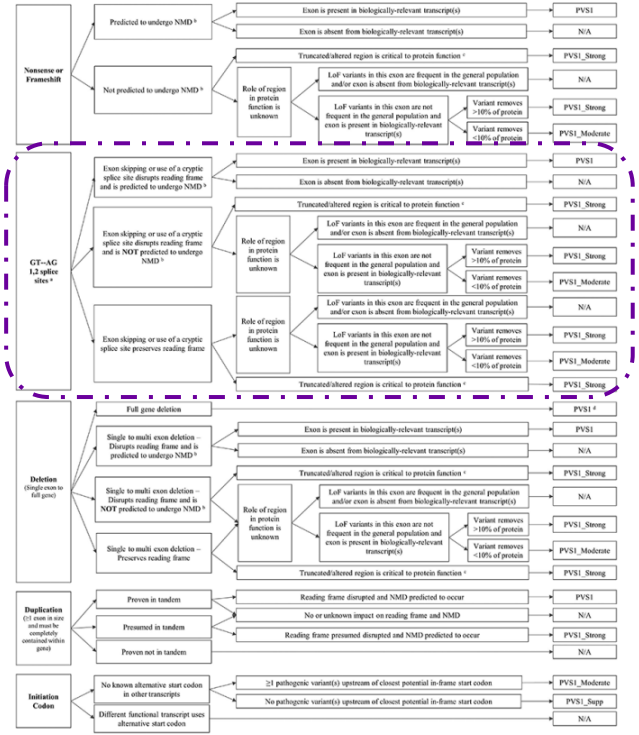


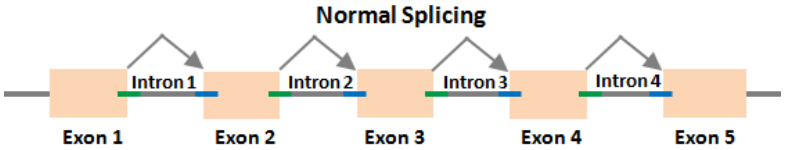




What if I had a variant here?

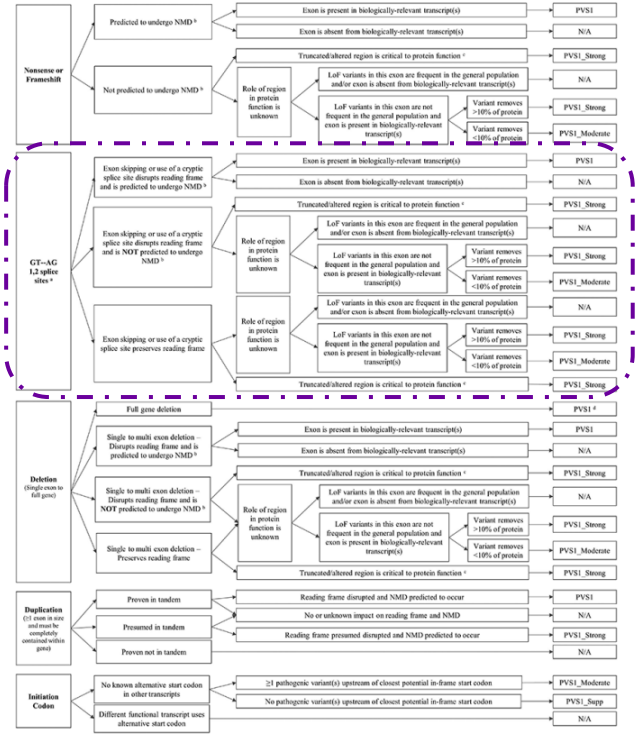
PVS1

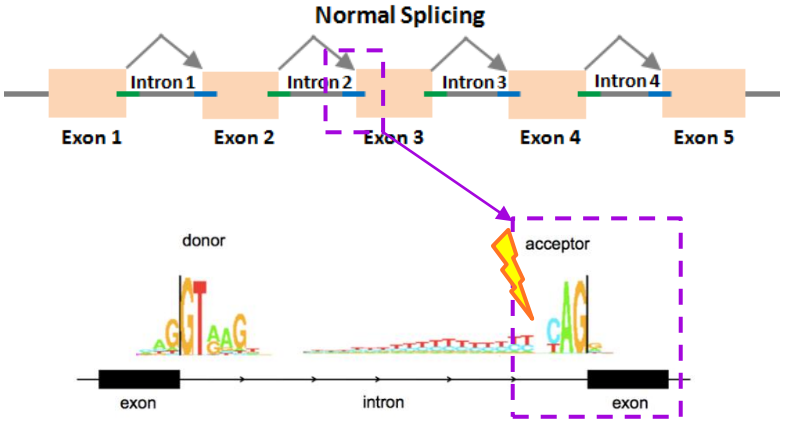




https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23626

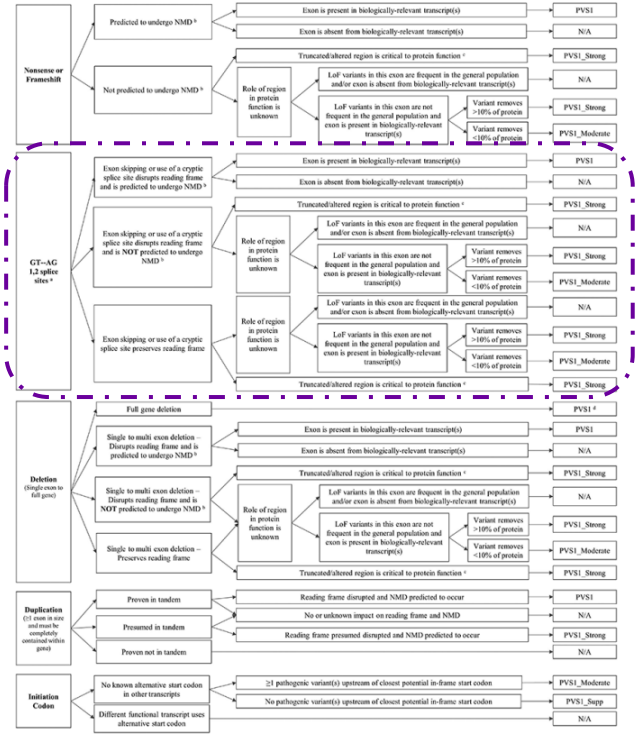
PVS1

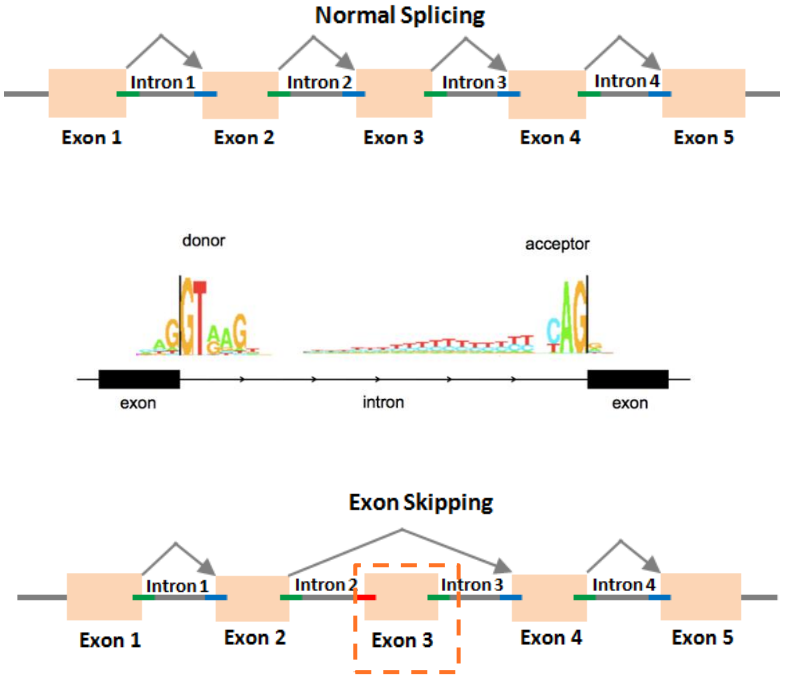




https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23626

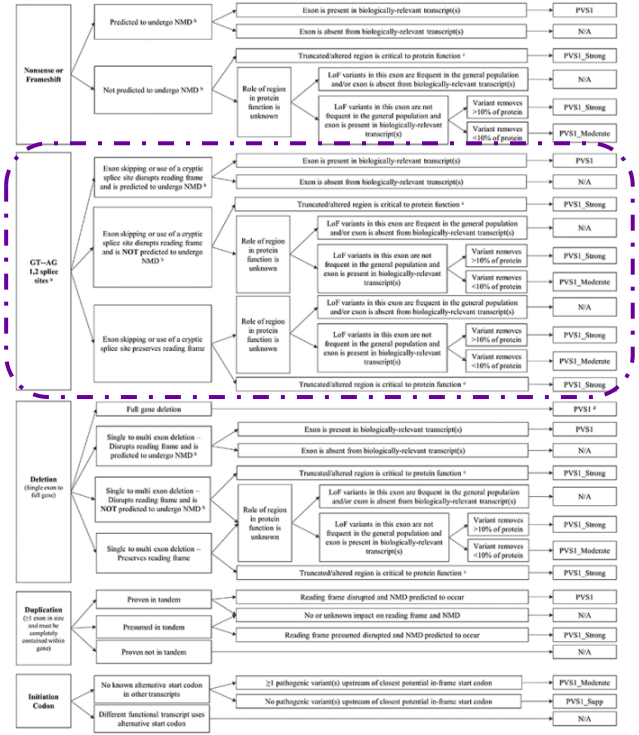
PVS1

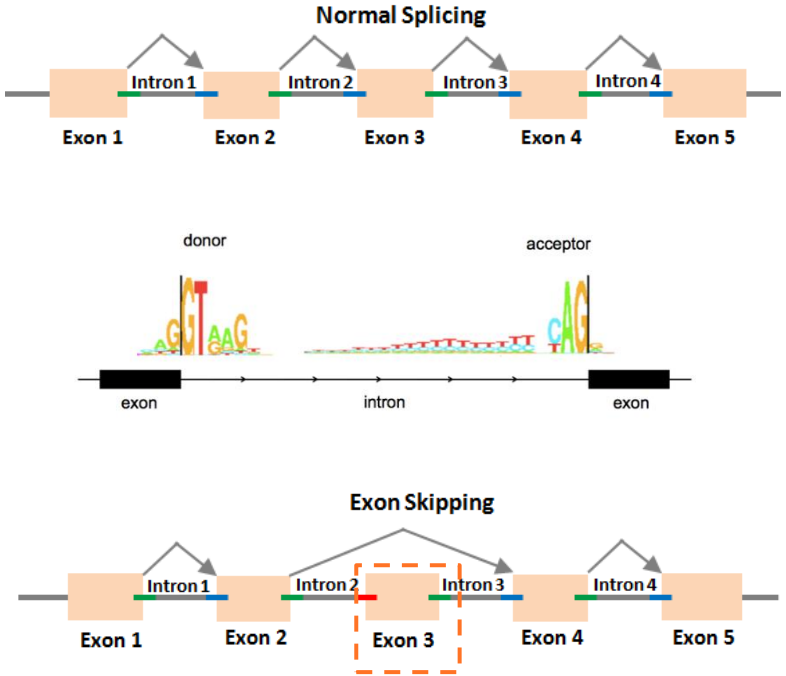




https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23626

PVS1- How to investigate if LOF is a `known mechanism of disease`



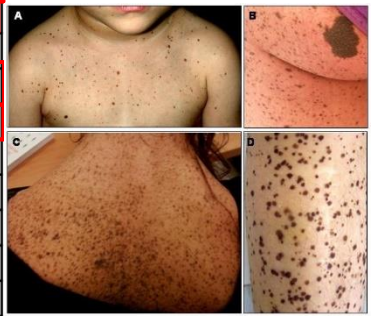


https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23626

Table 3.

Missense and LoF annotations and curations per gene from ClinGen Variant Curation Expert Panels

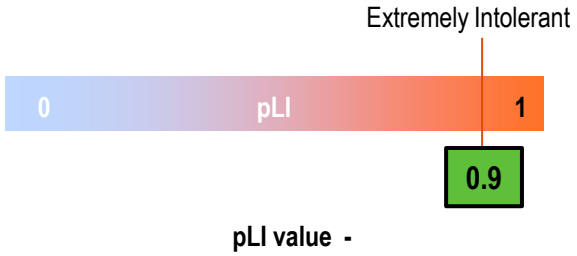
Gene	Disease Area (MOI)	HI Score	gnomAD LoF *oe* metric (90% CI)	PVS1?	Missense Z score (ExAC / gnomAD)	PP2?
MYH7	Cardio (AD)	0	0.45 (0.35–0.57)	Yes (Mod)	6.54 / 3.93	No
BRAF		1	0.1 (0.05–0.21)	No	3.99 / 3.72	Yes
HRAS		0	0.36 (0.16–0.93)	No	2.69 / 1.51	Yes
KRAS		0	0.63 (0.34–1.24)	No	1.36 / 2.32	Yes
MAP2K1		0	0.15 (0.07–0.38)	No	3.43 / 3.11	Yes
MAP2K2	RAS (AD)	1	0.1 (0.04–0.33)	No	1.48 / 1.87	Yes
PTPN11		3	0.03 (0.01–0.14)	No	3.43 / 3.13	Yes
RAF1		0	0.19 (0.11–0.35)	No	2.82 / 2.46	Yes
SHOC2		-	0 (0.00–0.14)	No	2.57 / 2.97	Yes
SOS1		0	0.07 (0.03–0.14)	No	2.18 / 3.05	Yes
PTEN	PHTS (AD)	3	0.24 (0.13–0.51)	Yes	3.71 / 3.49	Yes
CDH1	HDGC (AD)	3	0.25 (0.15–0.43)	Yes	0.81 / 0.71	No
PAH	PKU (AR)	30	1.12 (0.84–1.50)	Yes	−1.54 / −0.65	No
CDH23	HL (AR)	30	0.38 (0.26–0.57)	Yes	−0.24 / 0.71	No
GJB2		-	2.62 (1.39–1.98)	Yes	−1.07 / 1.17	No
MYO7A		-	0.7 (0.58–0.85)	Yes	−1.44 / 1.07	No
SLC26A4		-	0.89 (0.68–1.18)	Yes	−3.23 / −2.01	No
TECTA		30	0.45 (0.35–0.58)	Yes	2.3 / 1.61	No
USH2A		30	0.76 (0.67–0.86)	Yes	−5.12 / −2.47	No
COCH	HL (AD)	-	0.59 (0.40–0.91)	No	0.34 / 0.68	No
KCNQ4		-	0.22 (0.12–0.41)	Yes	2.73 / 1.83	No
MYO6		-	0.3 (0.22–0.42)	Yes	1.02 / 1.39	No
TECTA		30	0.45 (0.35–0.58)	No	2.3 / 1.61	No



https://doi.org/10.1186/1750-1172-3-13

How do I know if loss of function variants cause disease?

- pLI score - Probability a gene is haploinsufficient - where heterozygous LoFs are not tolerated. >0.9 is a common threshold. Particularly good for autosomal dominant disease.



PTPN11 protein tyrosine phosphatase non-receptor... Dataset gnomAD v2.1.1 gnomAD SVs v2.1

Genome build GRCh37 / hg19
Ensembl gene ID ENSG00000179295.11
Ensembl canonical transcript ENST00000351677.2
Other transcripts
ENST00000392597.1, ENST00000530818.1, ENST00000531326.1
Region 12:112856155-112947717
External resources Ensembl, UCSC Browser, and more

Constraint

Category	Expected SNVs	Observed SNVs	Constraint metrics
Synonymous	123.7	112	Z = 0.82 o/e = 0.91 (0.78 - 1.06)
Missense	331.3	171	Z = 3.13 o/e = 0.52 (0.46 - 0.59)
pLoF	35.2	1	pLI = 1 o/e = 0.03 (0.01 - 0.14)

Constraint metrics based on Ensembl canonical transcript (ENST00000351677.2).

https://gnomad.broadinstitute.org/gene/ENSG00000179295?dataset=gnomad_r2_1

PVS1- How to investigate if LOF is a `known mechanism of disease`

*176876
Table of Contents

ICD+

* 176876

PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR-TYPE, 11; PTPN11

Alternative titles; symbols

PROTEIN-TYROSINE PHOSPHATASE 2C; PTP2C
TYROSINE PHOSPHATASE SHP2; SHP2

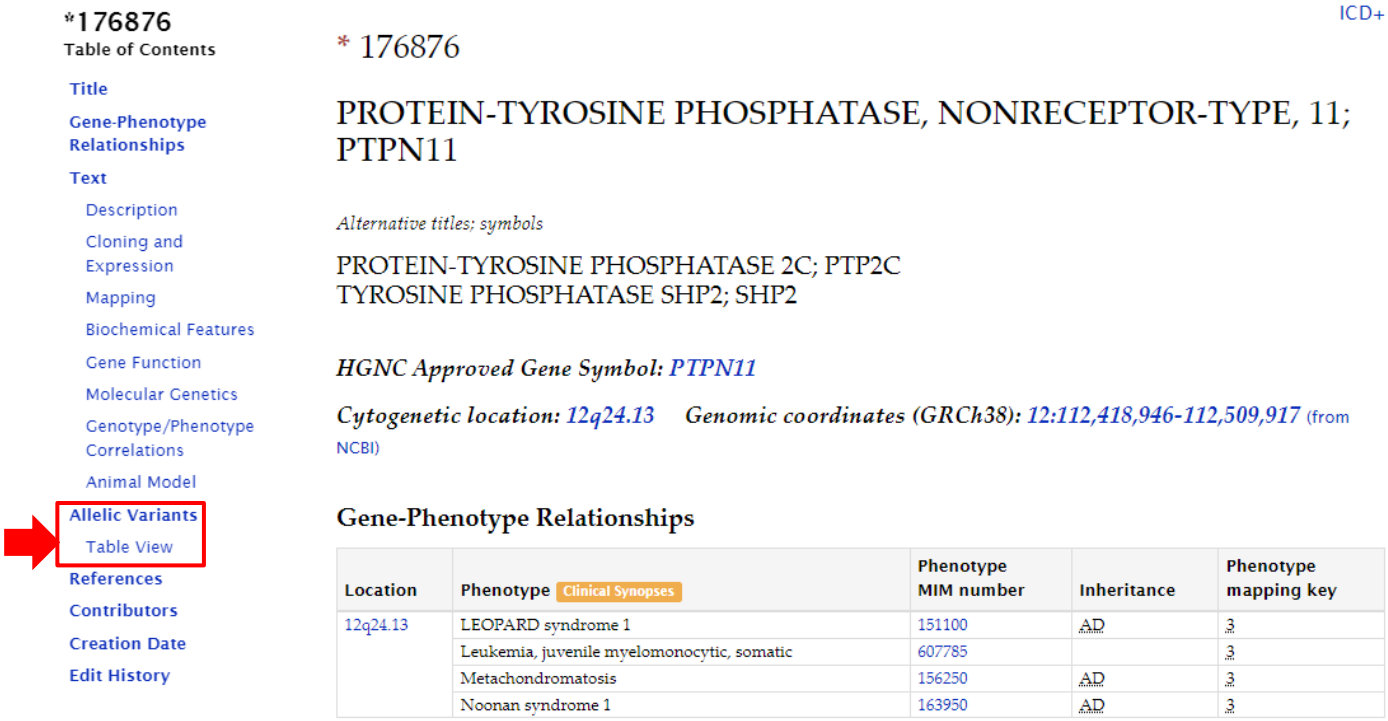
HGNC Approved Gene Symbol: *PTPN11*

Cytogenetic location: *12q24.13* ***Genomic coordinates (GRCh38):*** *12:112,418,946-112,509,917* (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype Clinical Synopses	Phenotype MIM number	Inheritance	Phenotype mapping key
12q24.13	LEOPARD syndrome 1	151100	AD	3
	Leukemia, juvenile myelomonocytic, somatic	607785		3
	Metachondromatosis	156250	AD	3
	Noonan syndrome 1	163950	AD	3

OMIM - Online Mendelian Inheritance in Man®



*176876
Table of Contents

- Title
- Gene-Phenotype Relationships
- Text
  - Description
  - Cloning and Expression
  - Mapping
  - Biochemical Features
  - Gene Function
  - Molecular Genetics
  - Genotype/Phenotype Correlations
  - Animal Model
- Allelic Variants
  - Table View
- References
- Contributors
- Creation Date
- Edit History

ICD+

* 176876

PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR-TYPE, 11; PTPN11

Alternative titles; symbols

PROTEIN-TYROSINE PHOSPHATASE 2C; PTP2C
TYROSINE PHOSPHATASE SHP2; SHP2

HGNC Approved Gene Symbol: PTPN11

Cytogenetic location: 12q24.13 *Genomic coordinates (GRCh38): 12:112,418,946-112,509,917* (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype Clinical Synopses	Phenotype MIM number	Inheritance	Phenotype mapping key
12q24.13	LEOPARD syndrome 1	151100	AD	3
	Leukemia, juvenile myelomonocytic, somatic	607785		3
	Metachondromatosis	156250	AD	3
	Noonan syndrome 1	163950	AD	3

OMIM - Online Mendelian Inheritance in Man®

PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR-TYPE, 11; PTPN11

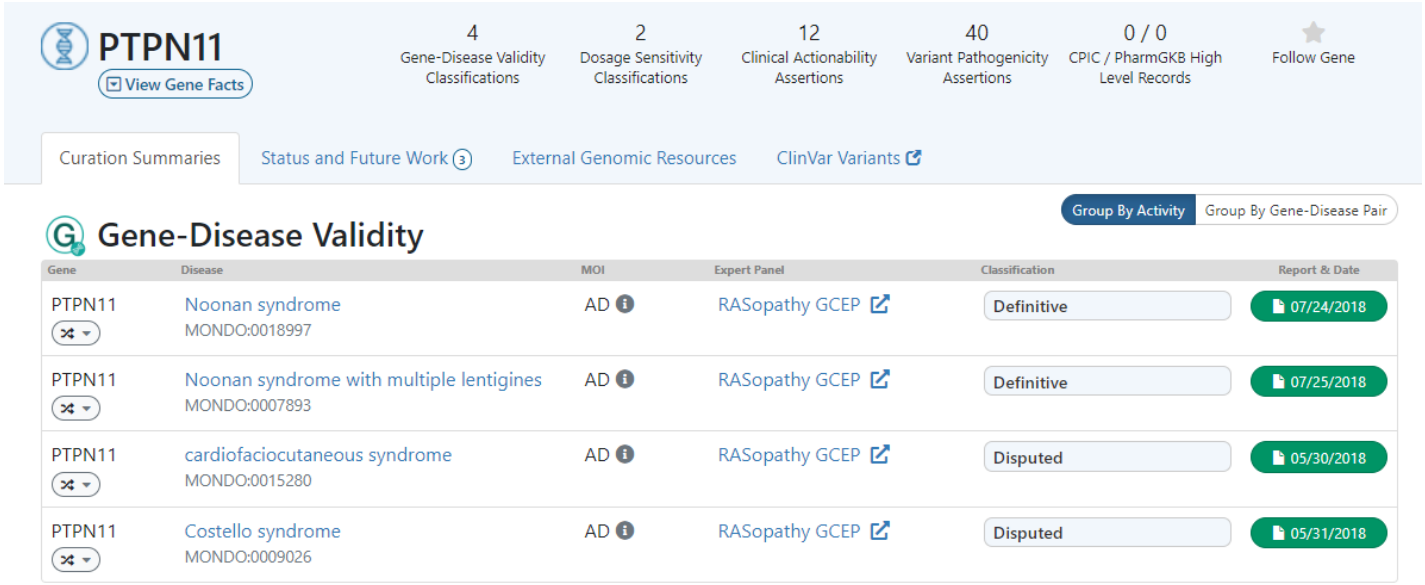
Allelic Variants (36 Selected Examples) :

All ClinVar Variants

Number	Phenotype	Mutation	SNP	gnomAD	ClinVar
.0001	NOONAN SYNDROME 1	PTPN11, ALA72SER	rs121918453	-	RCV000014252...
.0002	NOONAN SYNDROME 1	PTPN11, ALA72GLY	rs121918454	-	RCV000014253...
.0003	NOONAN SYNDROME 1	PTPN11, ASN308ASP	rs28933386	rs28933386	RCV000014254...
.0004	NOONAN SYNDROME 1	PTPN11, ASN308SER	rs121918455	-	RCV000014255...
.0005	LEOPARD SYNDROME 1	PTPN11, TYR279CYS	rs121918456	-	RCV000030620...
.0006	LEOPARD SYNDROME 1	PTPN11, THR468MET	rs121918457	rs121918457	RCV000033533...
.0007	NOONAN SYNDROME 1	PTPN11, SER502THR	rs121918458	-	RCV000014260...
.0008	NOONAN SYNDROME 1	PTPN11, TYR63CYS	rs121918459	rs121918459	RCV000014261...
.0009	NOONAN SYNDROME 1	PTPN11, TYR62ASP	rs121918460	rs121918460	RCV000014257...
.0010	NOONAN SYNDROME 1	PTPN11, ASP61GLY	rs121918461	-	RCV000014258...
.0011	NOONAN SYNDROME 1	PTPN11, THR73ILE	rs121918462	-	RCV000014262...
.0012	NOONAN SYNDROME 1	PTPN11, PHE285SER	rs121918463	-	RCV000014263...
.0013	MOVED TO 176876.0011	-	-	-	-
.0014	LEUKEMIA, JUVENILE MYELOMONOCYTIC, SOMATIC	PTPN11, GLU76LYS	rs121918464	-	RCV000014264...
.0015	LEUKEMIA, JUVENILE MYELOMONOCYTIC, SOMATIC	PTPN11, GLU76VAL	rs121918465	-	RCV000014265...
.0016	LEUKEMIA, JUVENILE MYELOMONOCYTIC, SOMATIC	PTPN11, GLU76GLY			
.0017	LEUKEMIA, JUVENILE MYELOMONOCYTIC, SOMATIC	PTPN11, GLU76ALA			
.0018	NOONAN SYNDROME	PTPN11, GLN79ARG			
.0019	NOONAN SYNDROME	PTPN11, THR411MET			
.0020	LEOPARD SYNDROME 1	PTPN11, ALA461THR			
.0021	LEOPARD SYNDROME 1	PTPN11, GLY464ALA			
.0022	LEOPARD SYNDROME 1	PTPN11, GLN510PRO			
.0023	NOONAN SYNDROME	PTPN11, GLN510ARG	rs121918470	rs121918470	RCV000014273...

- Most of the variants associated with the phenotype are Missense

ClinGen



- Noonan syndrome is believed to be caused **by gain-of-function** defects in *PTPN11* (PMID:11992261), and LEOPARD syndrome is believed to be caused by **dominant-negative mechanisms** (PMID: 16358218). Evidence gathered for the haploinsufficiency rating for this gene is related to the metachondromatosis phenotype.

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

PS1

PS1	Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val->Leu caused by either G>C or G>T in the same codon

(b)

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

PS1

PS1	Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val->Leu caused by either G>C or G>T in the same codon

(b)

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G

CUU → CUC

Both are Leucine

Caveat:	Beware of changes that impact splicing rather than at the amino acid/protein level

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

PM5

PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

- PM5

PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

Example: Arg156His is pathogenic; now you observe Arg156Cys

(b)

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G

Leu257Pro - Pathogenic

CUU → CCU

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

- PM5

PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

Example: Arg156His is pathogenic; now you observe Arg156Cys

(b)

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G

Leu257Pro - Pathogenic

CUU → CCU

Leu257His - ???

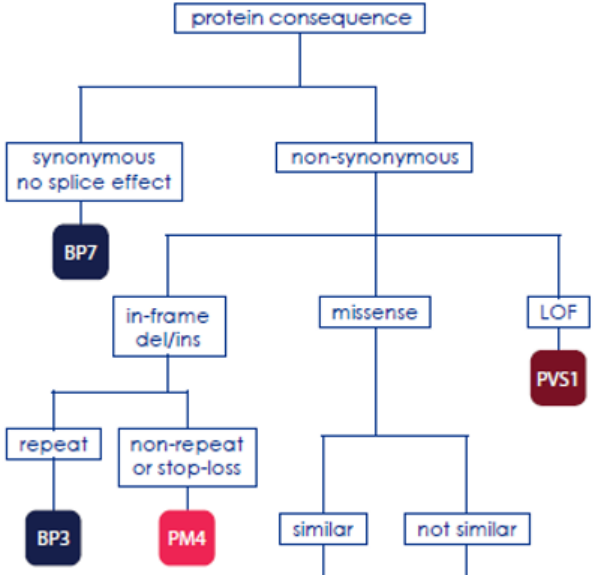
CUU → CAU

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

PM4

- Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants.



In-frame removal or insertion of amino acids

PM4

Protein length changes as a result of in-frame deletions/insertions

NM_000179.3(MSH6):
c.535_546del
p.(Ala179_Ala182del)

- Insertions/deletions that occur in repetitive regions are more likely to be of little functional impact; therefore, it is important to assess the surrounding sequence for repetitiveness using a genome browser.
- It can also help to assess population databases, such gnomAD, for high confidence variant calls that indicate the site is multi-allelic, which could indicate that the region is prone to indels that are generally tolerated, depending on the overall allele frequency.
- It is important to verify the functional impact the deletion or insertion might have. Does it affect the zinc-fingers of a transcription factor? Does it remove important aminoacids in the catalytic site?
- To prevent double-counting of this evidence type, we recommend that PM4 should not be applied for any variant in which PVS1, at any strength level, is also applied.

Stop loss: Protein extending variants

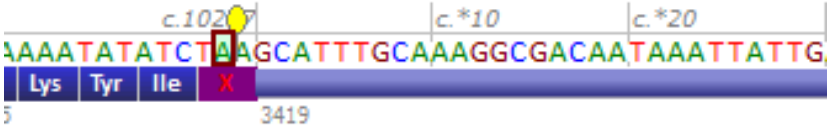
PM4

Protein length changes as a result of in-frame deletions/insertions

PM4

Stop-loss variant

- When a variant results in loss of the termination codon (stop-loss variant), the protein is extended; if a variant creates a premature termination codon (nonsense variant), the protein is shortened.



NM_000059.3(BRCA2):
c.10256_10257insT
p.(*3419Tyrext*18)

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

- PP3

PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)

Computational Impact Prediction

"In silico scores"

Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1

PP3

Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)

BP4

Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)

Pathogenicity Scores

20 | 1

BayesDel addAF *dbNSFP version 4.1*	addAF prediction Damaging		Mutation assessor *dbNSFP version 4.1*	prediction Medium	score 3.28	rankscore 0.9021
BayesDel noAF *dbNSFP version 4.1*	noAF prediction Damaging		MutationTaster *dbNSFP version 4.1*	Prediction Disease causing	Accuracy 1	converted rankscore 0.81
DANN *version 2014*	Score 0.9987		PROVEAN *dbNSFP version 4.1*	prediction Damaging	score -6.74, -6.44, -6.82, -6.71, -6.76	converted rankscore 0.9298
DEOGEN2 *dbNSFP version 4.1*	prediction Damaging, Tolerated		REVEL *dbNSFP version 4.1*	prediction Pathogenic	score 0.9599	rankscore 0.9939
			SIFT *dbNSFP version 4.1*	prediction Damaging	score 0	converted rankscore 0.9125
EIGEN *dbNSFP version 4.1*	prediction Pathogenic		SIFT4G *dbNSFP version 4.1*	prediction Damaging	score 0.002, 0.001, 0.003, 0	converted rankscore 0.9282
EIGEN PC *dbNSFP version 4.1*	prediction Pathogenic		PrimateAI *dbNSFP version 4.1*	prediction Tolerated	score 0.6591	rankscore 0.6128

Computational Impact Prediction

Considerations

PP3

BP4

- PP3 or BP4 can be used only once in a variant. Many algorithms used the same or very similar training data for their predictions, each algorithm cannot be counted as an independent criterion.
- Consistent threshold for the tool(s) should be used for all the variants in that gene.
- Currently, a **meta-predictor such as REVEL** may be used in place of multiple predictors in the in silico analysis of missense variants.
- Splicing in silico tools can be difficult to utilize and the interpretation is often not standardized.

Computational Impact Prediction

Splicing Scores

PP3

BP4

BP7

A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved.

Richards CS et al. *Gen Med.* 2015;17:405-423

Computational Impact Prediction

A Commonly Used Powerful Splicing Tool

PP3

BP4

BP7

SpliceAI

Deep neural network based on pre-mRNA transcript sequences that predicts splice sites using long-range primary genomic sequence flanking each position as input
(+/-50 bp as default; +/-10,000 bp maximum).

https://spliceailookup.broadinstitute.org/

Δ type	Δ score	pre-mRNA position
Acceptor Loss	0.00	
Donor Loss	0.72	0 bp
Acceptor Gain	0.00	
Donor Gain	0.01	-47 bp

SpliceAI provides a table with delta scores (0-1) for acceptor loss, donor loss, acceptor gain, and donor gain within the designated flanking sequence. The delta score indicates the probability that the variant will alter splicing at the pre-mRNA position indicated

Scores are not deterministic of biological effect/deleteriousness, they are used as “supporting evidence”

gDNA: Chr6(GRCh37):g.51720765A>G
cDNA: NM_138694.3(PKHD1):c.7837T>C
Protein: p.Trp2613Arg

Polyphen-2: Probably damaging
CADD: 29
M-CAP: Probably
PredictSNP2: Deleterious

Scores agree towards SNV being deleterious

Likelihood of pathogenicity is affected, not determined.

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	with disease BS4		disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Functional Evidence:

PS3

Functional Consequence

Criteria for classifying pathogenic variants

Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product.

Note:

Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.

BS3

No Functional Consequence

Criteria for classifying benign variants

Well-established *in vitro* or *in vivo* functional studies show no damaging effect on protein function or splicing.

What defines a "well established" functional study or assay? How reliable? This is not simple.

Functional Evidence:

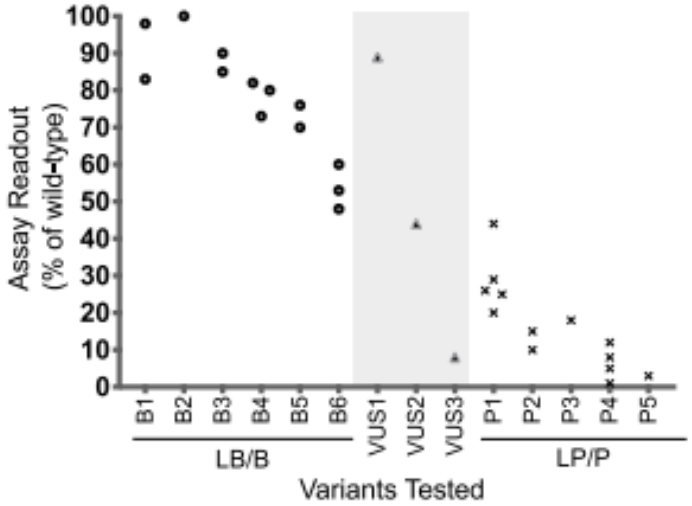
GUIDELINE Open Access

Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework

Sarah E. Brnich[1], Ahmad N. Abou Tayoun[2], Fergus J. Couch[3], Garry R. Cutting[4], Marc S. Greenblatt[5], Christopher D. Heinen[6], Dona M. Kanavy[1], Xi Luo[7], Shannon M. McNulty[1], Lea M. Starita[8,9], Sean V. Tavtigian[10], Matt W. Wright[11], Steven M. Harrison[12], Leslie G. Biesecker[13], Jonathan S. Berg[1*] and On behalf of the Clinical Genome Resource Sequence Variant Interpretation Working Group

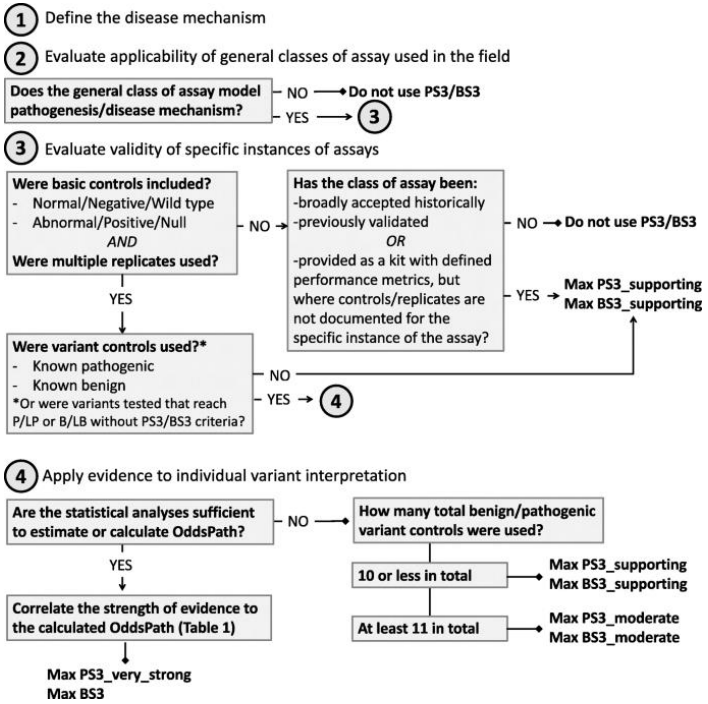
Table 3 Evidence strength equivalent of odds of pathogenicity

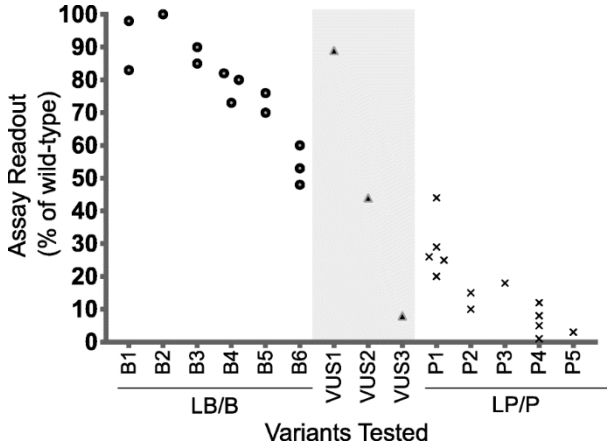
Odds of pathogenicity (OddsPath)	Evidence strength equivalent
< 0.053	BS3
< 0.23	BS3_moderate*
< 0.48	BS3_supporting
0.48–2.1	Indeterminate
> 2.1	PS3_supporting
> 4.3	PS3_moderate
> 18.7	PS3
> 350	PS3_very_strong



- Most functional evidence under these recommendations is demoted to PS3_supporting and in order to increase to moderate or strong, need to consider appropriate level of controls.
- Always consider if a test or assay is measuring the protein function or one of many.

Decision Tree to guide PS3/BS3 criterion





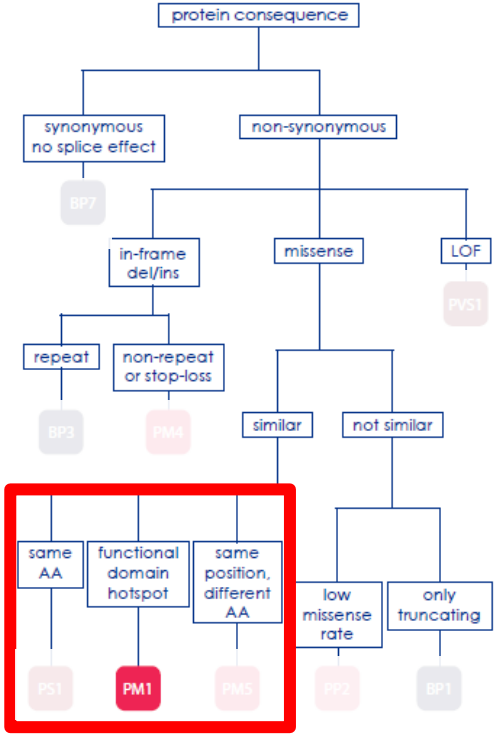
“Functional” Impact Prediction: Computational or Knowledge-based

PM1

Hotspot

Criteria for classifying pathogenic variants

Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation.



“Functional” Impact Prediction: Computational or Knowledge-based

PM1

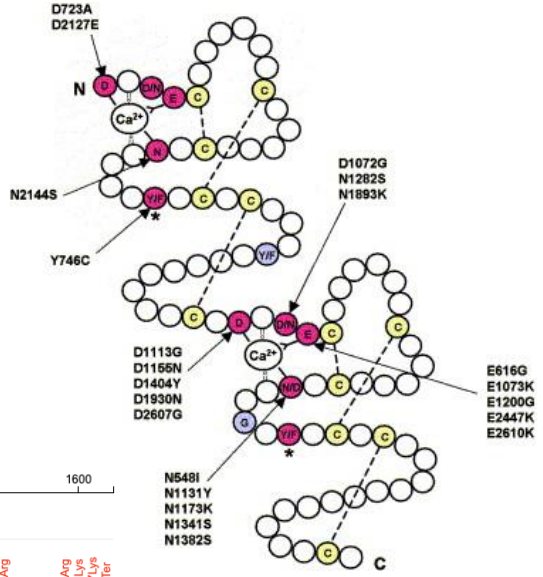
Hotspot

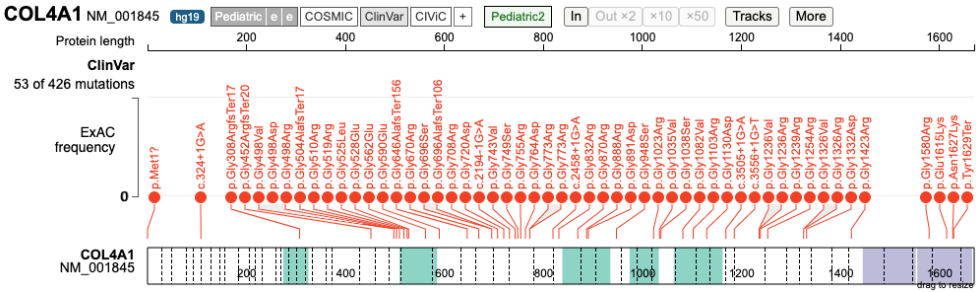
Circulation: Genomic and Precision Medicine
Volume 11, Issue 6, June 2018
https://doi.org/10.1161/CIRCGEN.117.002039

ORIGINAL ARTICLE

Tailoring the American College of Medical Genetics and Genomics and the Association for Molecular Pathology Guidelines for the Interpretation of Sequenced Variants in the *FBN1* Gene for Marfan Syndrome

Proposal for a Disease- and Gene-Specific Guideline





Richards CS et al. *Gen Med.* 2015;17:405-423

“Functional” Impact Prediction: Computational or Knowledge-based

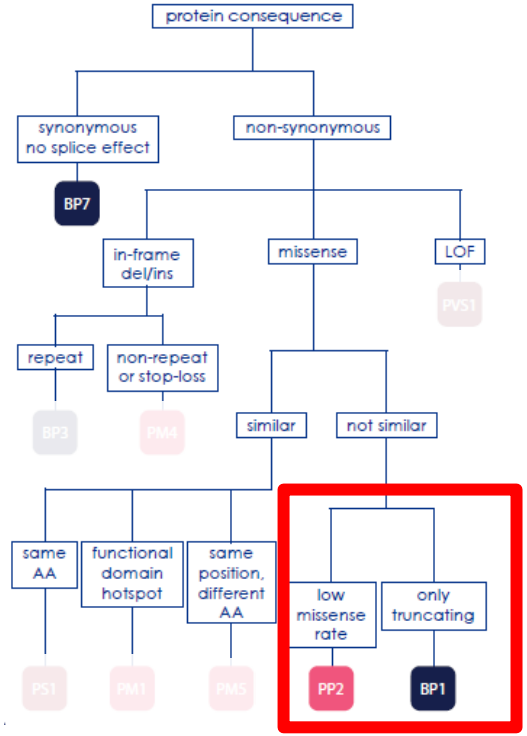
Reference laboratories are very conservative in the use of this criteria because of its subjectivity

Automated criteria

Rule	Explanation
PM1 Moderate	UniProt protein CLN6_HUMAN trans-membrane region 'Helical' has 6 non-VUS missense/in-frame/non-synonymous, variants (6 pathogenic and 0 benign), pathogenicity = 100.0% which is more than threshold 50.0%.

Richards CS et al. *Gen Med.* 2015;17:405-423

Impact Prediction: Computational or Knowledge-based

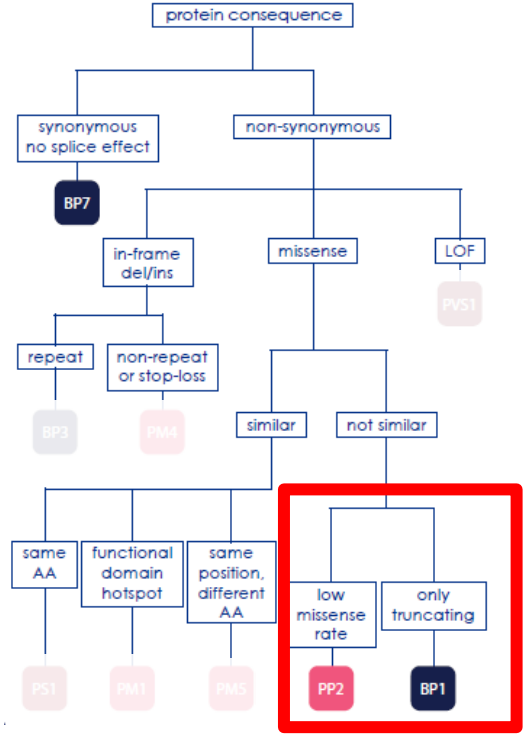


Impact Prediction: Computational or Knowledge-based

Constraint

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	163.3	147	Z = [illegible] o/e = [illegible] (0.79 - 1.03)
Missense	498.1	262	Z = 3.76 o/e = 0.53 (0.47 - 0.58)
pLoF	45.6	6	pLI = 1 o/e = 0.13 (0.07 - 0.26)

Rule of thumb Z-score “bigger” than 3



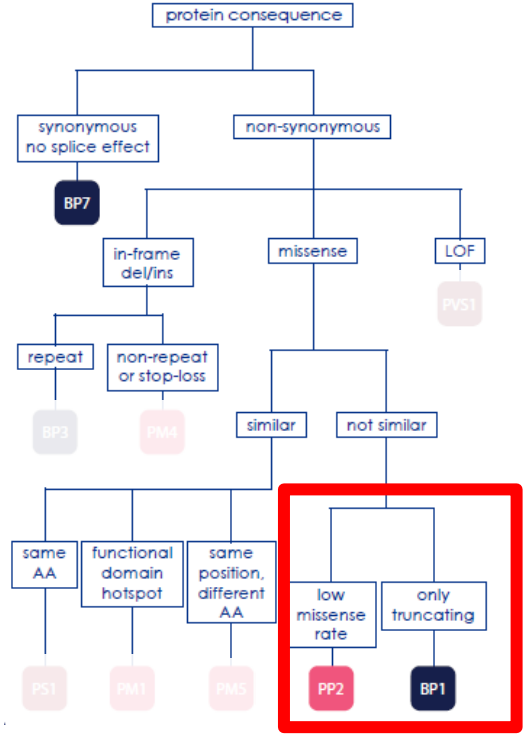
Richards CS et al. *Gen Med.* 2015;17:405-423

Impact Prediction: Computational or Knowledge-based

"Genic" Tolerance

Constraint

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	163.3	147	Z = [illegible] o/e = [illegible] (0.79 - 1.03)
Missense	498.1	262	Z = 3.76 o/e = 0.53 (0.47 - 0.58)
pLoF	45.6	6	pLI = 1 o/e = 0.13 (0.07 - 0.26)



Impact Prediction: Computational or Knowledge-based

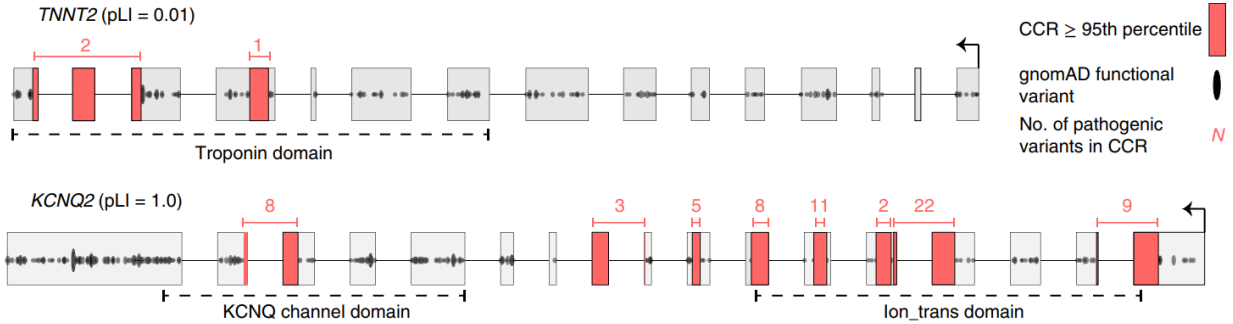
- Gene-wide summary measures of constraint are prone to overstating and understating constraint within specific regions of protein-coding genes

A map of constrained coding regions in the human genome

James M. Havrilla[1,2], Brent S. Pedersen[1,2], Ryan M. Layer[3,4] and Aaron R. Quinlan[1,2,5*]

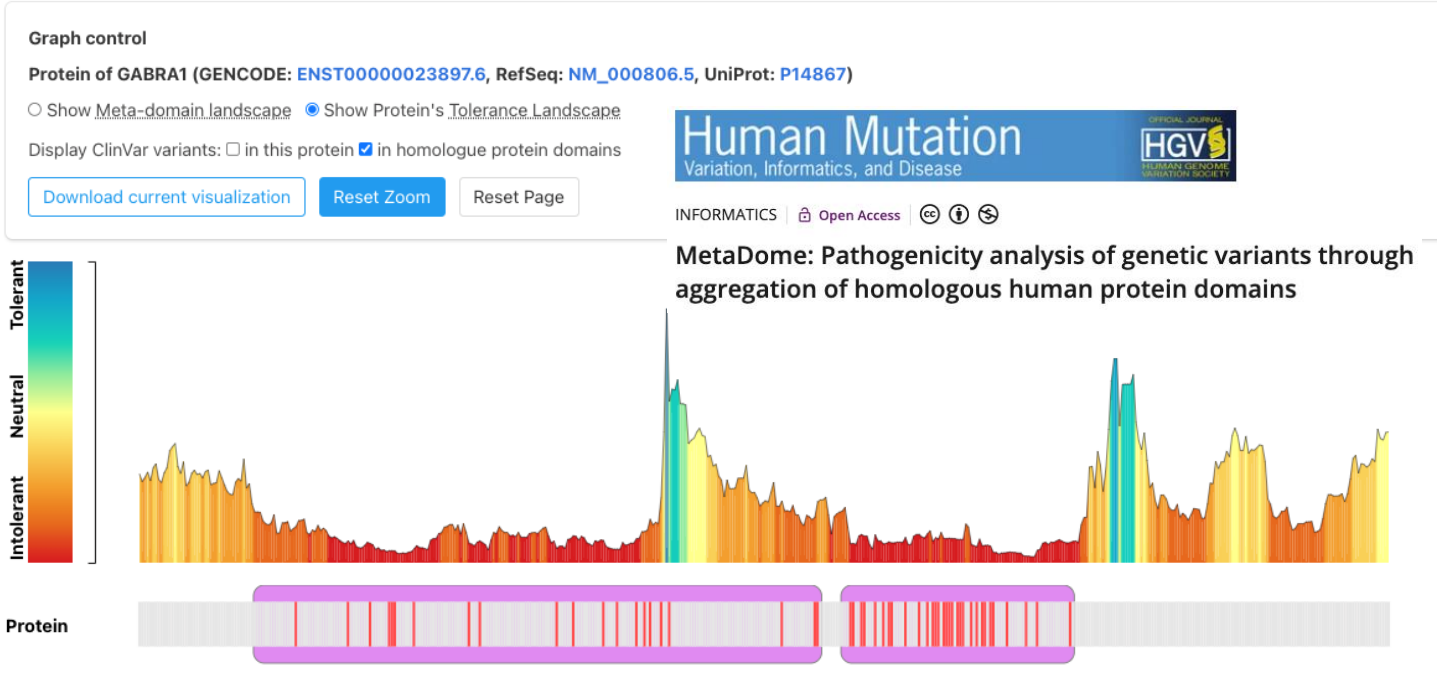
nature genetics

Regional intolerance correlates with important functional domains

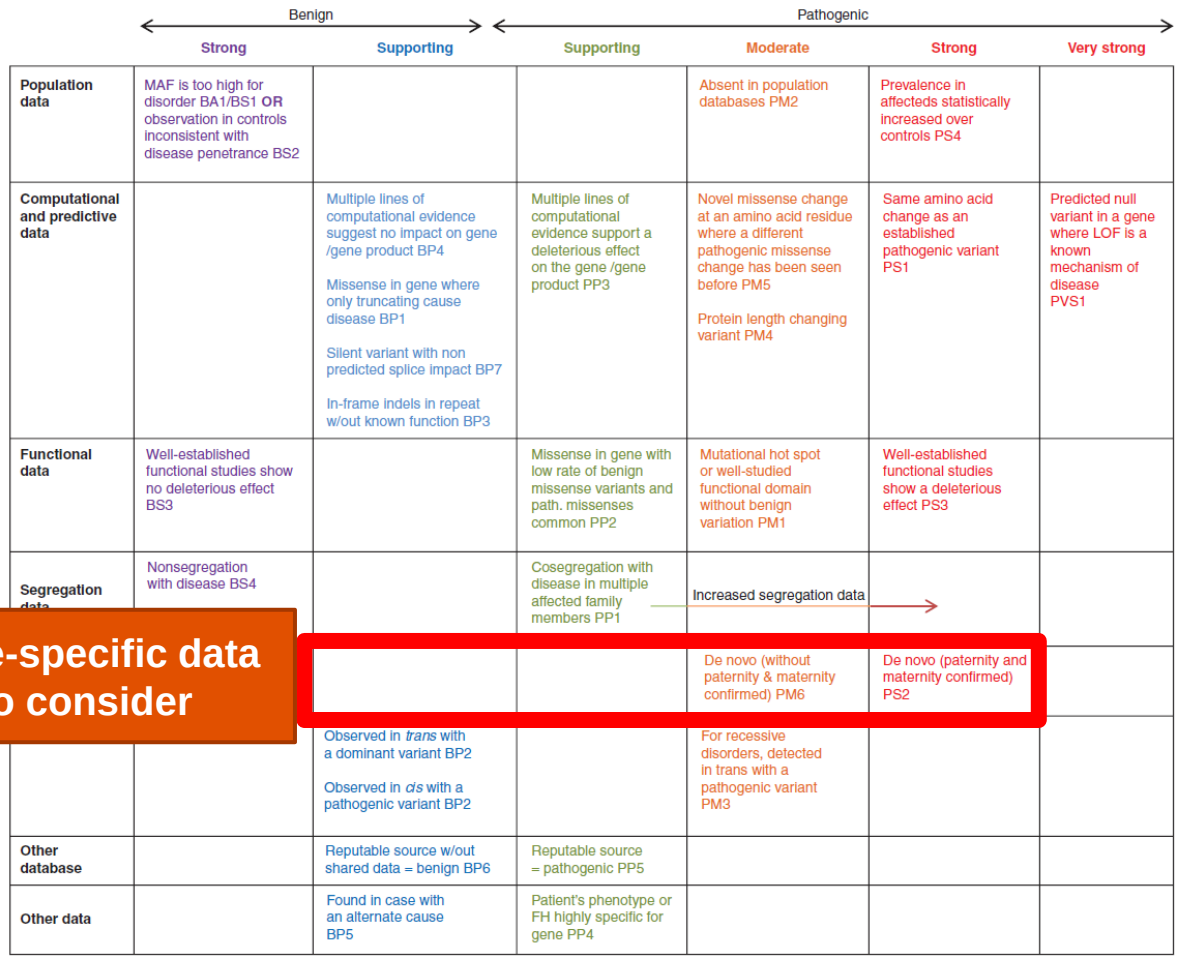


Impact Prediction: Computational or Knowledge-based

- Gene-wide summary measures of constraint are prone to overstating and understating constraint within specific regions of protein-coding genes

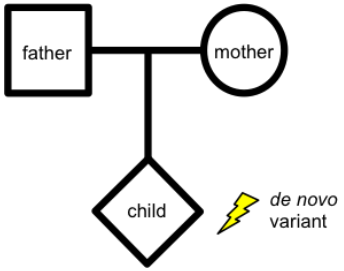


https://stuart.radboudumc.nl/metadome/dashboard



Case-specific data to consider

Case-Specific Evidence - Segregation Data



PS2
De Novo - Confirmed

PM6
De Novo - Not Confirmed

De novo (both maternity and paternity confirmed) in a patient with the disease and no family history.

PS2/PM6

- parental confirmed
- phenotype consistency
- number of *de novo* observations

Phenotypic consistency	Points per Proband	
	***de novo* with confirmed parental relationships**	***de novo* with unconfirmed parental relationships**
Phenotype highly specific for gene	2	1
Phenotype consistent with gene but not highly specific	1	0.5
Phenotype consistent with gene but not highly specific and high genetic heterogeneity**	0.5	0.25
Phenotype not consistent with gene	0	0

Supporting (PS2_Supporting or PM6_Supporting)	**Moderate** (PS2_Moderate or PM6)	**Strong** (PS2 or PM6_Strong)	**Very Strong** (PS2_VeryStrong or PM6_VeryStrong)
0.5	1	2	4

https://clinicalgenome.org/site/assets/files/3461/svi_proposal_for_de_novo_criteria_v1_1.pdf

PS2/PM6

- parental confirmed
- phenotype consistency
- number of *de novo* observations

Phenotypic consistency	Points per Proband	
	***de novo* with confirmed parental relationships**	***de novo* with unconfirmed parental relationships**
Phenotype highly specific for gene	2	1
Phenotype consistent with gene but not highly specific	1	0.5
Phenotype consistent with gene but not highly specific and high genetic heterogeneity**	0.5	0.25
Phenotype not consistent with gene	0	0

Supporting (PS2_Supporting or PM6_Supporting)	**Moderate** (PS2_Moderate or PM6)	**Strong** (PS2 or PM6_Strong)	**Very Strong** (PS2_VeryStrong or PM6_VeryStrong)
0.5	1	2	4

If a NIPBL variant was de novo in one patient with Cornelia de Lange syndrome, with confirmed parental relationships and de novo in two additional unrelated patients with Cornelia de Lange syndrome with unconfirmed parental relationships, then …

PS2/PM6

- parental confirmed
- phenotype consistency
- number of *de novo* observations

Phenotypic consistency	Points per Proband	
	***de novo* with confirmed parental relationships**	***de novo* with unconfirmed parental relationships**
Phenotype highly specific for gene	2	1
Phenotype consistent with gene but not highly specific	1	0.5
Phenotype consistent with gene but not highly specific and high genetic heterogeneity**	0.5	0.25
Phenotype not consistent with gene	0	0

Supporting (PS2_Supporting or PM6_Supporting)	**Moderate** (PS2_Moderate or PM6)	**Strong** (PS2 or PM6_Strong)	**Very Strong** (PS2_VeryStrong or PM6_VeryStrong)
0.5	1	2	4

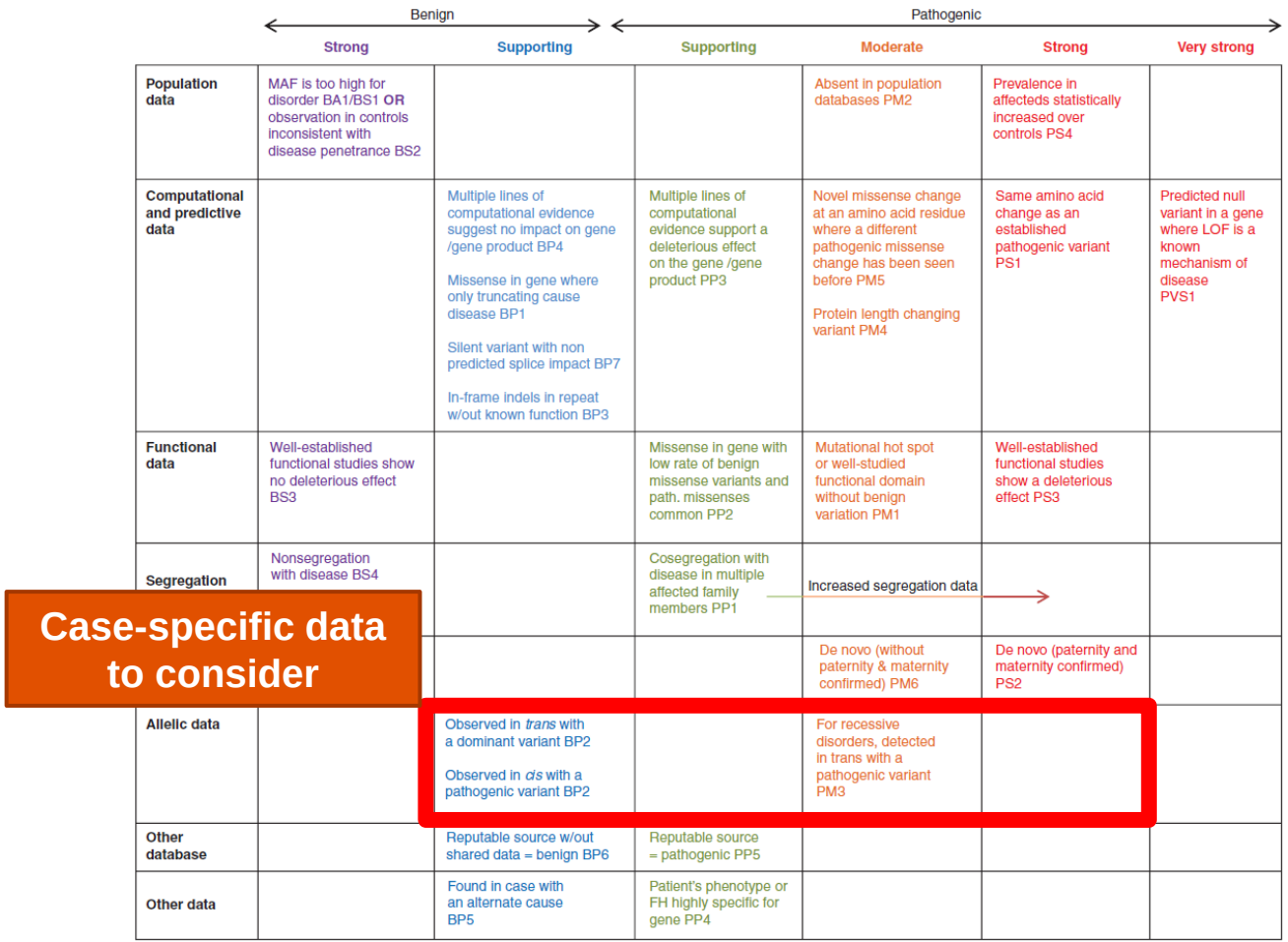
If a NIPBL variant was de novo in one patient with Cornelia de Lange syndrome, with confirmed parental relationships and de novo in two additional unrelated patients with Cornelia de Lange syndrome with unconfirmed parental relationships, then VeryStrong evidence level is applied (PS2_VeryStrong) based on combined point value of 4 (Table 2).

PS2/PM6 – Additional considerations

- A patient with early infantile epileptic encephalopathy and a de novo *SIK1* variant with confirmed parental relationships is awarded 1 point (as the patient's phenotype is consistent with the gene but not highly specific and the variant is de novo with confirmed parental relationships). If this patient is the only de novo occurrence for the variant, then a Moderate strength level (PS2_Moderate) is applied.

- A patient with nonsyndromic intellectual disability and a de novo *ASH1L* variant is awarded 0.5 points (as the variant is de novo with confirmed parental relationships and patient's phenotype is consistent with the gene but not highly specific and there is significant evidence of genetic heterogeneity). If this patient is the only de novo occurrence for the variant, then a Supporting strength level (PS2_Supporting) is applied.

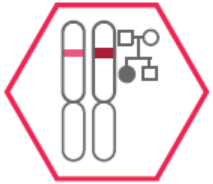
- A patient with developmental delay but no other features of Cornelia de Lange syndrome and a de novo *NIPBL* variant with unconfirmed parental relationships is awarded zero points as this phenotype is not consistent with the gene/disease association. If this patient was the only de novo occurrence for the variant, then no de novo criteria are applied.



	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Case-specific data to consider

Case-Specific Evidence – Allelic Data

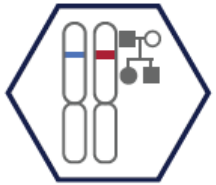


PM3

Trans

For recessive disorders, detected in trans with a pathogenic variant.

Note:
This requires testing of parents (or offspring) to determine phase.



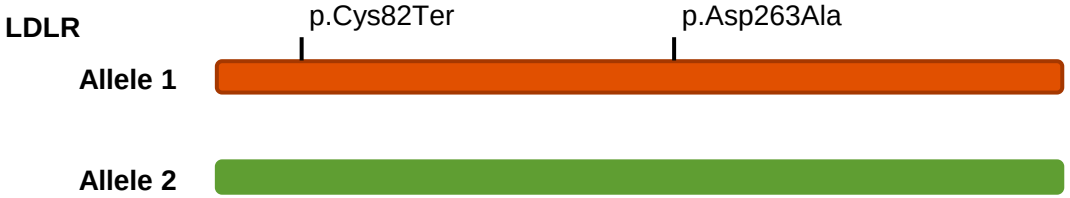
BP2

With *cis* Pathogenic

Observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern.

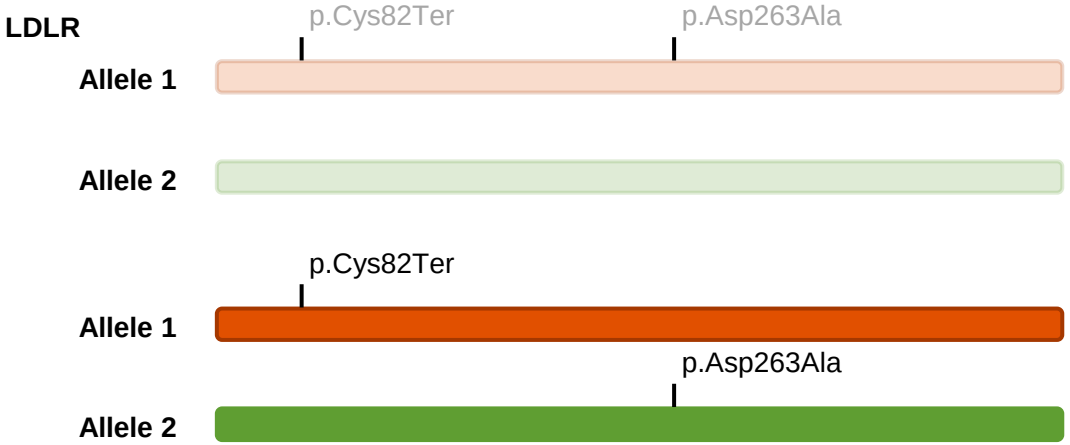
PM3/BP2

- Patient presents with Familial Hypercholesterolemia



PM3/BP2

- Patient presents with Familial Hypercholesterolemia



	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Case-specific data to consider

Case-Specific Evidence – Phenotype Specificity

PP4

Patient's phenotype or family history is highly specific for a disease with a single genetic etiology.

Richards CS et al. *Gen Med.* 2015;17:405-423

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Reputable sources

PP5/BP6

NM_000249.4(MLH1):c.931A>G (p.Lys311Glu)

Interpretation:	**Likely pathogenic**
Review status:	★★★☆ reviewed by expert panel
Submissions:	5 (Most recent: Sep 24, 2021)
Last evaluated:	Mar 9, 2018
Accession:	VCV000230595.10
Variation ID:	230595
Description:	single nucleotide variant

Who is reputable?

Submitted interpretations and evidence

Interpretation (Last evaluated)	Review status (Assertion criteria)	Condition (Inheritance)	Submitter	More information
Likely pathogenic (Mar 09, 2018)	reviewed by expert panel (Guidelines v2.3) Method: curation	Lynch syndrome I Affected status: yes Allele origin: germline	International Society for Gastrointestinal Hereditary Tumours (InSiGHT) Accession: SCV000740673.1 Submitted: (Mar 23, 2018)	Other databases http://www.insight-database.org/... Comment: Multifactorial probability: 0.999 but with conflicting data. Reduced classification to class 4 pending somatic information.
Uncertain significance (Dec 06, 2019)	criteria provided, single submitter (Ambry Autosomal Dominant and X-Linked criteria (3/2017)) Method: clinical testing	Hereditary cancer-predisposing syndrome Affected status: unknown Allele origin: germline	Ambry Genetics Accession: SCV000274194.5 Submitted: (Nov 30, 2020)	Comment: The p.K311E variant (also known as c.931A>G), located in coding exon 11 of the MLH1 gene, results from an A to G substitution at nucleotide ... (more)
Uncertain significance (Apr 24, 2019)	criteria provided, single submitter (LabCorp Variant Classification Summary - May 2015) Method: clinical testing	not specified Affected status: unknown Allele origin: germline	Women's Health and Genetics/Laboratory Corporation of America, LabCorp Accession: SCV000696173.3 Submitted: (Sep 24, 2019)	Comment: Variant summary: MLH1 c.931A>G (p.Lys311Glu) results in a conservative amino acid change located in the N-terminal domain (IPR002099) of the encoded protein sequence. Four of ... (more)
Uncertain significance (Jun 21, 2020)	criteria provided, single submitter (Invitae Variant Classification Sherloc (09022015)) Method: clinical testing	Hereditary nonpolyposis colorectal neoplasms Affected status: unknown Allele origin: germline	Invitae Accession: SCV000543638.6 Submitted: (Jan 07, 2021)	Publications: PubMed (4) Comment: This sequence change replaces lysine with glutamic acid at codon 311 of the MLH1 protein (p.Lys311Glu). The lysine residue is highly conserved and there is ... (more)
Uncertain significance (Jun 11, 2020)	criteria provided, single submitter (GeneDx Variant Classification Process June 2021) Method: clinical testing	Not Provided Affected status: yes Allele origin: germline	GeneDx Accession: SCV000565923.3 Submitted: (Sep 24, 2021)	Comment: Not observed at a significant frequency in large population cohorts (Lek et al., 2016); In silico analysis supports that this missense variant has a deleterious ... (more)

Knowledge Databases \ Previous Interpretations – ClinVar, HGMD

Variant of Interest: NM_000249.3(MLH1):c.1038G>T (p.**Gln346**His) (p.**Q346**H)

https://www.ncbi.nlm.nih.gov/clinvar/?term=Q346+%5Bvariant+name%5D+and+MLH1

	Variation *Location*	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Review status
	NM_000249.3(**MLH1**):c.1037A>G (p.Gln346Arg) *GRCh37:* Chr3:37061953 *GRCh38:* Chr3:37020462	MLH1	Q346R, Q248R, Q313R, Q105R, Q5R	Lynch syndrome	Pathogenic (Oct 18, 2018)	reviewed by expert panel
2.	NM_000249.3(**MLH1**):c.1038G>T (p.Gln346His) *GRCh37:* Chr3:37061954 *GRCh38:* Chr3:37020463	MLH1	Q346H, Q1 Q248H, Q3			expert
	NM_000249.3(**MLH1**):c.1038G>C (p.Gln346His) *GRCh37:* Chr3:37061954 *GRCh38:* Chr3:37020463	MLH1	Q346H, Q1 Q248H, Q3			expert

PM5

PS1

PS1 and PM5 is typically used in clinical laboratories if the ClinVar submission has a “review status” of 2 stars or multiple submitters of P and LP interpretations without any conflicts or 3 stars expert panel.

Knowledge Databases \ Previous Interpretations – ClinVar, HGMD

Variant of Interest: NM_000249.3(MLH1):c.1038G>T (p.**Gln346**His) (p.**Q346**H)

PS1

HGMD® Professional 2020.4

CM1812352	CAG-CAC	Gln346His	c.1038G>C	p.Q346H	DM	Colorectal cancer, non-polyposis	Shirts (2018) Am J Hum Genet **103**, 19
CM092210	CAG-CAT	Gln346His	c.1038G>T	p.Q346H	DM	Colorectal cancer, non-polyposis	Tang (2009) Clin Genet **75**, 334 Pagenstecher (2006) *Hum Genet* **119**: 9 [Functional characterisation] Zhu (2013) *Oncol Lett* **5**: 1710 [Additional report] 2 more reference(s)...

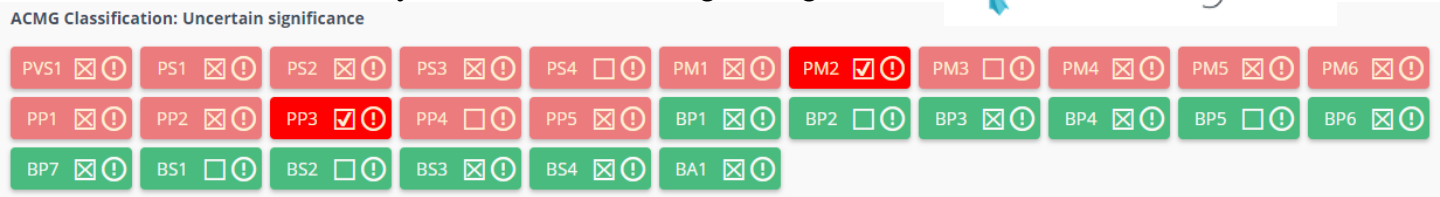
- Potential PS1 or PM5 – if there is literature available for the same missense variant or a similar substitution without a ClinVar assertion, carefully review the data for a potential application of PM5 or PS1.
- The variant has to stand on its own merits as P/LP for use of PS1 or PM5.

Publicly Available Calculators and Workflows

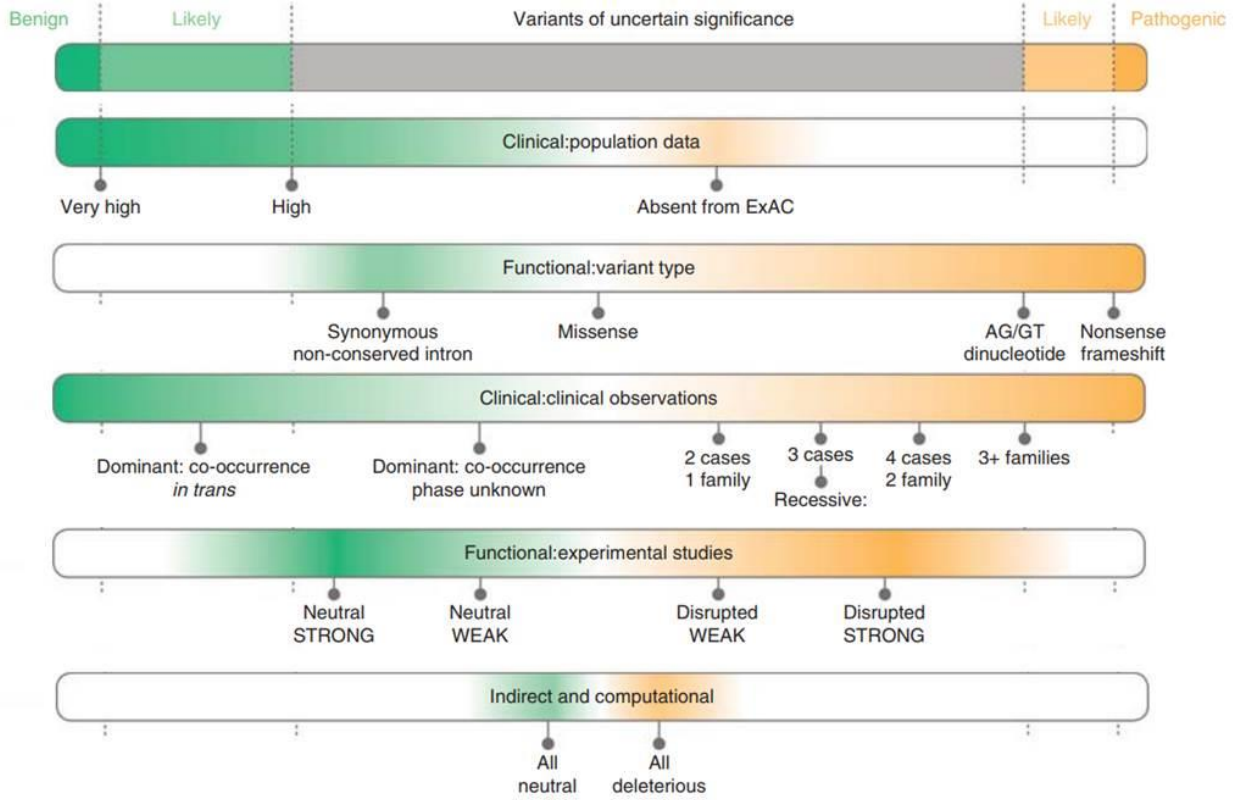
- Publically available tools that will help tally up your “points”
  - https://varsome.com/
  - http://wintervar.wglab.org/
  - http://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/
  - https://mobidetails.iurc.montp.inserm.fr/MD/

- Several analysis software integrate guideline

emedgene



The ACMG guidelines are not mandatory, or the only ones used



Framework Summary for Variant Interpretation – 6 key questions

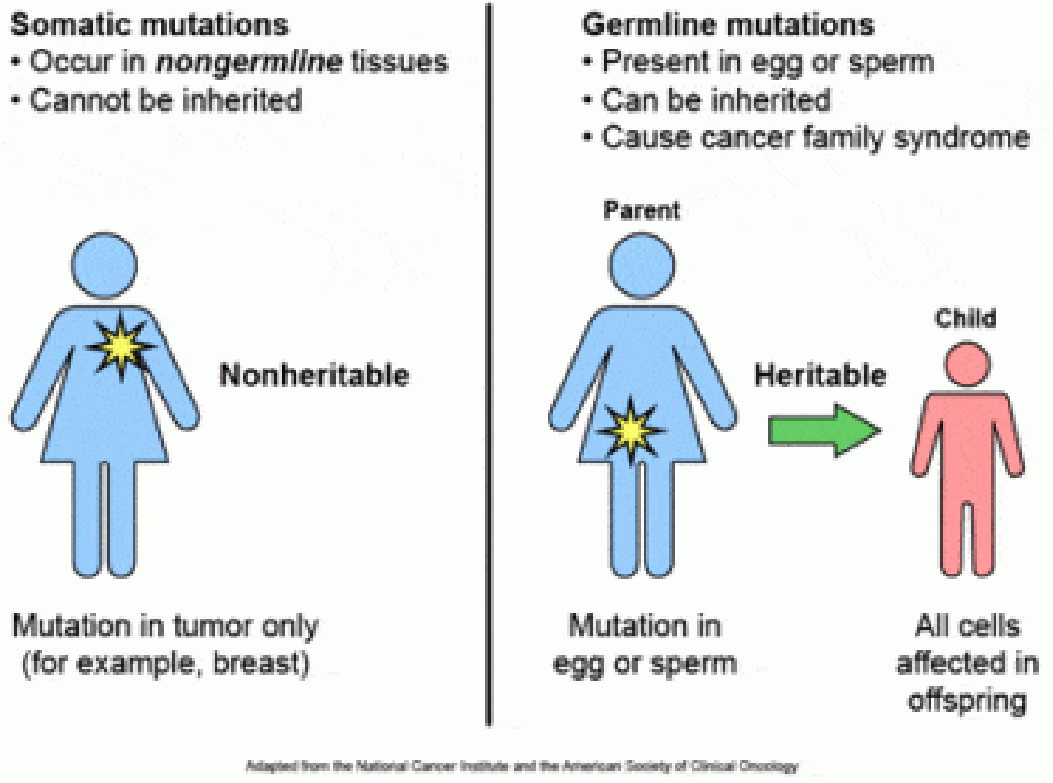
- Allele Frequency?
- What is the mechanism of disease?
- Known or predicted impact?
- Do we have functional evidence? How reliable?
- Phenotype overlaps with gene-disease association described?
- Does it segregate with disease?

Warning!

Germline and Somatic Classification and Catalogue Differences

OncoKB

COSMIC
Catalogue Of Somatic Mutations In Cancer



OMIM®

human phenotype ontology

ClinVar

Warning!

Germline and Somatic Classification and Catalogue Differences

Categories:

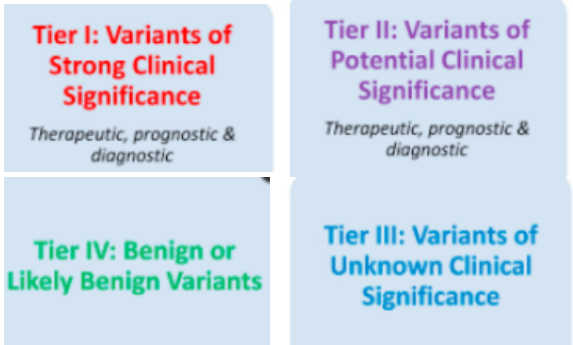
Diagnostic

Prognostic

Therapeutic

Somatic mutations
- Occur in ***nongermline*** tissues
- Cannot be inherited

Germline mutations
- Present in egg or sperm
- Can be inherited
- Cause cancer family syndrome



Categories:

Pathogenic

Likely Pathogenic

VUS – Variant of Uncertain Significance

Likely Benign

Benign

Questions?

Variant Interpretation Summary Example:
BRCA1 (NM_007294.3) c. 212G>C, p.(Arg71Thr)

SUMMARY

The heterozygous c.212G>C (p.R71T) variant was detected in the BRCA1 gene (NM_007294.3) and involves the last residue of exon 4 of 23.

This variant has been reported in a single affected individual with the associated disease (Harter et al., 2017; PMID: 29053726).

Functional testing has been performed for this variant and supports decreased protein function with a reduced expression of mRNA in transfected HAP1 cells. (Findlay et al, 2018, PMID 30209399).

Another amino acid substitution occurring in the same residue (p.Arg71Gly, p.Arg71Lys) has been determined to contribute to the disease associated with this gene.

The variant detected is absent in a large control population database without reported homozygotes (Karczewski et al., 2020, PMID: 32461654).

Multiple computational predictors suggest a damaging effect on gene or protein function.

Therefore, c.212G>C (p.R71T) in the BRCA1 gene is classified as ***Pathogenic.*** *Clinical correlation is recommended.*

CONCEPT	ACMG CRITERIAS CODE
Introduction	
PREVIOUSLY REPORTED CASES	PS4_Supporting
FUNCTIONAL TESTING	PS3
RESIDUE LEVEL ANNOTATION	PM5
ALLELE FREQUENCY	PM2
INSILICO PREDICTORS	PP3
Conclusion	

Variant Interpretation Summary Example:

BRCA1 (NM_007294.3) c. 212G>C, p.(Arg71Thr)

Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

ACMG CRITERIAS CODE

PS4_Supporting
PS3
PM5
PM2
PP3