Blue text are notes from source logic material

Purple text are remaining questions/areas of uncertainty

Input: One variant, VCF format

Output: Precalculatable ACMG evidence scores with reason for score assignment

**Nomenclature**

Ensure input variant is left normalized VCF format. If variant is entered as HGVS c., convert to VCF with pyHGVS

Display the HGVS c., and p., for the primary and disease-relevant transcript from multiple annotators: cartagenia, snpeff, pyhgvs

Verify (cartagenia, snpeff, pyhgvs) produce HGVS format, consider, and test possible differences on MVL for: VEP, annovar, VAT

**Population Data**

**Data sources:**

Variant allele frequency database

Variant allele frequency database coverage data (possibly just gnomAD)

Gene information table

**Algorithm:**

Use variant’s highest sub-population/ethnicity MAF with AC > 2000

Funda: “No MAF information due to poor sequence quality (<20X)”: skip

If AF appears = 0% (no individuals with variants), the coverage at that position > ~20X/sample

SOP: “MAF is >=5%, this is considered very strong evidence that the variant is benign… If the variant is <5% but greater than >1%, then a likely benign classification should be considered”

Hard cutoffs:

>5%: BA1

1-5%: BS1

<1%: PM2

Can we assume a fair disease allele frequency cutoff is ~0.1% more than the most frequent MVL pathogenic variant? (assumes most common disease alleles are already discovered – will require curation of highest AF variants in pathogenic variant database)

Or 0.1% more than the assumed carrier AF/RR cutoff?

Need to compare known patho AF to whatever cutoff is decided based on carrier AF/RR

Disease-specific cutoffs:

> 10x higher than disease allele frequency: BA1

Higher than disease allele frequency: BS1

Disease allele frequency is consistent with a plausible pathogenic variant: PM2

Which phenotypes should be blacklisted from having BS2 applied from gnomAD?

Observed in gnomAD for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder AND high penetrance (cutoff?) AND not adult-onset: BS2

**Variant database with pathogenicity assertion**

**Data sources:**

Pathogenic variant database

**Algorithm:**

Non-MVL pathogenic variant reports from diagnostic cases (affected patients) without literature evidence

All reports must be consistent (B/LB, V, LP/P)

Pathogenic: PP5

Benign: BP6

ACMG: “[For PS4:] The prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls [or presence at a disease appropriate frequency], may be used as moderate level of evidence”

If PM2 and number of unrelated affected patients > 1: PS4-M

[Uncertain criteria but ACMG usage excel shows as option]: PS4-P

**Predictions and computational evidence for deleteriousness/splicing**

**Data sources:**

Gene information table

Pathogenic variant database

Refgene table for lengths and boundaries of exons (Opt: Uniprot or protein sequence db)

Missense predictors (Alamut and dbNSFP)

Splicing predictors (Alamut and dbscSNV)

MultiZ 100 vertebrate alignment (Opt: Conservation predictors)

RepeatMasker

Amino acid substitution matrix (grantham / blossom)

**Nonsense and Frameshift**

**Algorithm:**

*ACMG paper: “PVS1… complete absence of the gene product by lack of transcription or nonsense-mediated decay”*

Above being the ideal definition but later implied to potentially tolerate existing protein product but with *negligible functionality*

*“[CSS variants could lead to an…] in-frame deletion/insertion, which could retain the critical domains of the protein and lead to a minor length change (PM4)”*

Gray area being what is ‘minor’ and should there be outlet for PM4…

LOF disease mechanism

NMD (New stop codon NOT in last exon, or 55bp at 3’ end of penultimate exon)

IF any 1

>= 2 pathogenic variants with PTC same codon (encompasses PS1), 3’, or 2% of transcript length 5’

> 10% of transcript is missing/frameshifted

> 5% extension of protein length

: PVS1

ELSE:

PM4

**Canonical Splice Site (CSS)**

**Algorithm:** Assumes exon-skipping consequence

SOP: rather simple since they use single-exon deletion = PVS1 which is the minimal expected consequence

*“For canonical splice variants, do not use the 10% rule or 2% rule… these variants have protein effects that may be difficult to predict (ie, entire exon skipping)”*

*“PVS1 (…single or multi-exon deletion…)”*

Zach:

LOF disease mechanism

Not dispensable exon from gene information table

: PVS1

**Start-loss**

**Algorithm:**

LOF disease mechanism: PVS1

Other start-loss variants in the same codon reported as pathogenic: PS1

**Stop-loss**

**Algorithm:**

PM4

[Uncertain criteria but ACMG usage excel shows as option]: PM4-S

[Uncertain criteria but ACMG usage excel shows as option]: PM4-VS

Other stop-loss variants in the same codon reported as pathogenic: PS1

**In-frame InDel**

**Algorithm:**

Not in simple repeat by repeatmasker: PM4

[Uncertain criteria but ACMG usage excel shows as option]: PM4-S

[Uncertain criteria but ACMG usage excel shows as option]: PM4-VS

In simple repeat by repeatmasker: BP3

**Nonsynonymous**

**Algorithm:**

*ACMG: “The variant amino acid change being present in multiple nonhuman mammalian species in an otherwise well-conserved region, suggesting the amino acid change would not compromise function,* ***can be considered strong evidence for a benign interpretation****”* clear example where a manual override to benign is recommended which doesn’t necessarily fit into guideline scoring

PP3 BP4 *“Ex: missense/splicing/conservation preds must be 100% concordant”*

PM5 [missense where same AA affected as patho]

PS1 [same (ex: missense) AA sub as patho]

dbscSNV ada\_score > 0.97: PP3

Use SIFT, PolyPhen2, MutationTaster for v1 preds, recommend update after performance assessment of REVEL, metaSVM, CADD.

SIFT AND PolyPhen2 AND MutationTaster = deleterious: PP3

AA absolutely conserved in mammals: PP3

SIFT AND PolyPhen2 AND MutationTaster = tolerated: BP4

Variant amino acid is reference in > 1 mammal: BP4

Same amino acid substitution as pathogenic variant: PS1

Same amino acid affected as pathogenic variant and substitution is similar or more biochemically different: PM5

Present in manually curated hotspot location from gene information table (GlyXY motif, active site, .bed interval of critical residues): PM1

[Uncertain criteria but ACMG usage excel shows as option]: PM1-S

Gene’s valid disease variant type is “GOF(missense only)”: PP2

Gene’s valid disease variant type is “LOF-only”: BP1

Clarify satisfactory variants for BP1, LOF-only = (From usage probably stopgain, CSS, startloss, but require PVS1 for stopgain)

**Synonymous or intronic (non-canonical splicing consensus)**

**Algorithm:**

dbscSNV ada\_score > 0.97: PP3

dbscSNV ada\_score < 0.6: BP7

**Remaining purple:**

Verify (cartagenia, snpeff, pyhgvs) produce HGVS format, consider, and test possible differences on MVL for: VEP, annovar, VAT

Can we assume a fair disease allele frequency cutoff is ~0.1% more than the most frequent MVL pathogenic variant? (assumes most common disease alleles are already discovered – will require curation of highest AF variants in pathogenic variant database)

Or 0.1% more than the assumed carrier AF/RR cutoff?

Need to compare known patho AF to whatever cutoff is decided based on carrier AF/RR

Which phenotypes should be blacklisted from having BS2 applied from gnomAD?

Because of the inclusion of affected individuals in gnomAD

Because of inappropriate age of onset or penetrance

[Uncertain criteria but ACMG usage excel shows as option]: PS4-P

[Uncertain criteria but ACMG usage excel shows as option]: PM4-S

[Uncertain criteria but ACMG usage excel shows as option]: PM4-VS

Use SIFT, PolyPhen2, MutationTaster for v1 preds, recommend update after performance assessment of REVEL, metaSVM, CADD.

[Uncertain criteria but ACMG usage excel shows as option]: PM1-S

Clarify satisfactory variants for BP1, LOF-only = (From usage probably stopgain, CSS, startloss, but require PVS1 for stopgain)

**ClinVar submitter stats from Dianwei’s ClinVar**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Submitter** | **Variants** | **Submissions** | **Diagnostic?** | **db other than clinvar?** |
| Illumina Clinical Services Laboratory,Illumina | 137417 | 138260 | yes | No |
| GeneDx | 69106 | 69121 | yes | Cant find |
| Invitae | 44663 | 44663 | yes | Yes but not their variants |
| EGL Genetic Diagnostics,Eurofins Clinical Diagnostics | 30041 | 30063 | yes | Yes |
| OMIM | 23154 | 23154 | no |  |
| Ambry Genetics | 21169 | 21183 | yes | Only cancer? |
| Laboratory for Molecular Medicine,Partners HealthCare Personalized Medicine | 18847 | 18848 | yes |  |
| PreventionGenetics | 15716 | 15718 | yes | Cant find |
| Genetic Services Laboratory, University of Chicago | 15156 | 15161 |  |  |
| Database of Curated Mutations (DoCM) | 6947 | 6947 | no | Clinvar has more |
| Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) | 6101 | 6102 |  |  |
| Counsyl | 6101 | 6101 |  |  |
| GeneReviews | 5502 | 5509 | no |  |
| ARUP Laboratories, Molecular Genetics and Genomics | 3439 | 3439 |  |  |
| Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA), c/o University of Cambridge | 3057 | 3057 |  |  |
| ITMI | 2699 | 2699 |  |  |
| Quest Diagnostics Nichols Institute San Juan Capistrano | 2695 | 2695 |  |  |
| Center for Pediatric Genomic Medicine,Children's Mercy Hospital and Clinics | 2358 | 2358 |  |  |
| Sharing Clinical Reports Project (SCRP) | 2239 | 2239 |  |  |
| Biesecker Lab/Human Development Section,National Institutes of Health | 2218 | 2218 |  |  |
| InSiGHT | 2120 | 2120 |  |  |
| Breast Cancer Information Core (BIC) (BRCA2) | 1900 | 1900 |  |  |
| Department of Pathology and Laboratory Medicine,Sinai Health System | 1890 | 1890 |  |  |
| Division of Genomic Diagnostics,The Children's Hospital of Philadelphia | 1744 | 1746 |  |  |
| Breast Cancer Information Core (BIC) (BRCA1) | 1658 | 1658 |  |  |
| Tuberous sclerosis database (TSC2) | 1587 | 1608 |  |  |
| Retina International | 1528 | 1532 |  |  |
| Laboratory Corporation of America, | 1526 | 1527 |  |  |
| Cardiovascular Biomedical Research Unit,Royal Brompton & Harefield NHS Foundation Trust | 1511 | 1522 |  |  |
| Praxis fuer Humangenetik Tuebingen, | 1488 | 1489 |  |  |
| LDLR-LOVD, British Heart Foundation | 1401 | 1401 |  |  |
| ARUP Institute,ARUP Laboratories | 1234 | 1268 |  |  |
| Blueprint Genetics | 1214 | 1214 |  |  |
| Epithelial Biology, Institute of Medical Biology, Singapore | 1080 | 1086 |  |  |
| Systems Biology Platform Zhejiang California International NanoSystems Institute | 1024 | 1024 |  |  |
| Research Molecular Genetics Laboratory,Women's College Hospital, University of Toronto | 1008 | 1008 |  |  |