strand
New Generation Healthcare

Variant classification assertion Criteria

Strand Life Sciences classifies variants based on data from several sources, including public variant databases (ClinVar, UniProt), curated databases (HGMD, in-house curation), population frequency databases (dbSNP, ExAC, Exome Variant Server, 1000 genomes), internal variant-phenotype databases, literature (Pubmed), bioinformatics predictors, conservation scores and segregation rules. Individual variant labels are based on the 5 ACMG classification levels:

1. Pathogenic
2. Likely Pathogenic
3. Variant of Uncertain Significance (VUS)
4. Likely Benign
5. Benign.

The details of how individual variant labels are determined is described below:

a) Pathogenic

A Variant is labeled *Pathogenic* if any one of the following holds:

- The **Override List**, that is maintained internally, explicitly states that the Variant is Pathogenic. Users can explicitly add variants to the Override List, with the criteria that every variant in the list has references and citations as evidence for the label.
- A **Credible Curated Source** indicates that the Variant is Pathogenic, and the variant's Minor Allele frequency is below a certain threshold in population frequency databases. Variants with conflicting clinical significance or non-mendelian phenotypes are dropped from this set.
- The Variant is reported in at least a certain number of affected individuals in the **StrandCurated-AffectedIndividuals** knowledgebase (this knowledgebase is curated from literature and records information about the various affected individuals reported in literature). In addition, its **Minor Allele Frequency** in population frequency databases is below a certain threshold. And finally, the Variant satisfies one of the following properties:
    - It causes **Protein Truncation** (on account of a premature stop, a frameshift or a start lost) to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold.
    - It causes **Mis-splicing** (on account of a splice donor or acceptor site change) to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold.
    - Some **Functional Validation** for the variant has been reported in literature.

- The Variant causes **Protein Truncation** to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold; further, another Variant which satisfies one of the above conditions lies in the truncated region, and its Minor Allele Frequency in population frequency databases is below a certain threshold.
- The Variant causes **Mis-splicing** to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold; further another Variant which satisfies one of the above conditions causes the same mis-splicing effect, and its Minor Allele Frequency in population frequency databases is below a certain threshold.

b) Likely pathogenic

A variant is called *Likely Pathogenic* if one of the following holds:

- A **Credible Curated Source** indicates that the Variant is Likely Pathogenic and the variant's Minor Allele frequency is below a certain threshold in population frequency databases. Variants with conflicting clinical significance or non-Mendelian phenotypes are dropped from this set.
- The Variant is reported in at least a certain number of affected individuals in the **StrandCurated-AffectedIndividuals** knowledgebase (this knowledgebase is curated from literature and records information about the various affected individuals reported in literature) – this number is smaller than the requirement for the variant to be called **Pathogenic**. In addition, its **Minor Allele Frequency** in population frequency databases is below a certain threshold. And finally, the Variant satisfies one of the following properties:
    - It causes **Protein Truncation** (on account of a premature stop, a frameshift or a start lost) to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold.
    - It causes **Mis-splicing** (on account of a splice donor or acceptor site change) to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold.
- The Variant causes **Protein Truncation** to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold and its Minor Allele Frequency in population frequency databases is below a certain threshold.
- The Variant causes **Mis-splicing** to a substantial extent, i.e., the amount of protein beyond the variant exceeds a certain threshold, and its Minor Allele Frequency in population frequency databases is below a certain threshold.

c) Variant of Uncertain Significance (VUS)

A variant is labeled *VUS* in various sub-categories as follows:

- The variant has been reported in literature, either with conflicting evidence, or weak evidence; further, its Minor Allele Frequency in population frequency databases is below a certain threshold.

- A missense variant is classified as *VUS* when its Minor Allele Frequency in population frequency databases is below a certain threshold **AND** one of the following conditions are met:
    - At least one well-known bioinformatics predictor from dbNSFP (https://sites.google.com/site/jpopgen/dbNSFP) predicts that the nucleotide change causes damage to the protein function. dbNSFP v2.8 uses seven prediction algorithms – SIFT, Polyphen2 HDIV, Polyphen2 HVAR, LRT, MutationTaster, MutationAssessor and FATHMM.
    - The variant location is conserved across multiple species (this is an indication of how important the location is to the function of the protein).
- The variant's Minor Allele Frequency in population frequency databases is below a certain threshold **AND** the variant causes an insertion or deletion of one or more amino acids.
- The variant's Minor Allele Frequency in population frequency databases is below a certain threshold **AND** the variant causes non-substantial protein shortening.
- The variant's Minor Allele Frequency in population frequency databases is below a certain threshold **AND** the variant causes possible missplicing, by virtue of being close to a splice site.
- The variant's Minor Allele Frequency in population frequency databases is below a certain threshold **AND** the variant is present in the promoter or untranslated region (UTR), causing a possible effect on gene expression.

d) Benign - A variant is labeled **Benign** if it has a Minor Allele Frequency higher than a certain threshold in population frequency databases (this indicates that the variant is commonly present in the general population).

e) Likely Benign - A variant is labeled **Likely Benign** if it doesn't fall into any of the other classes.