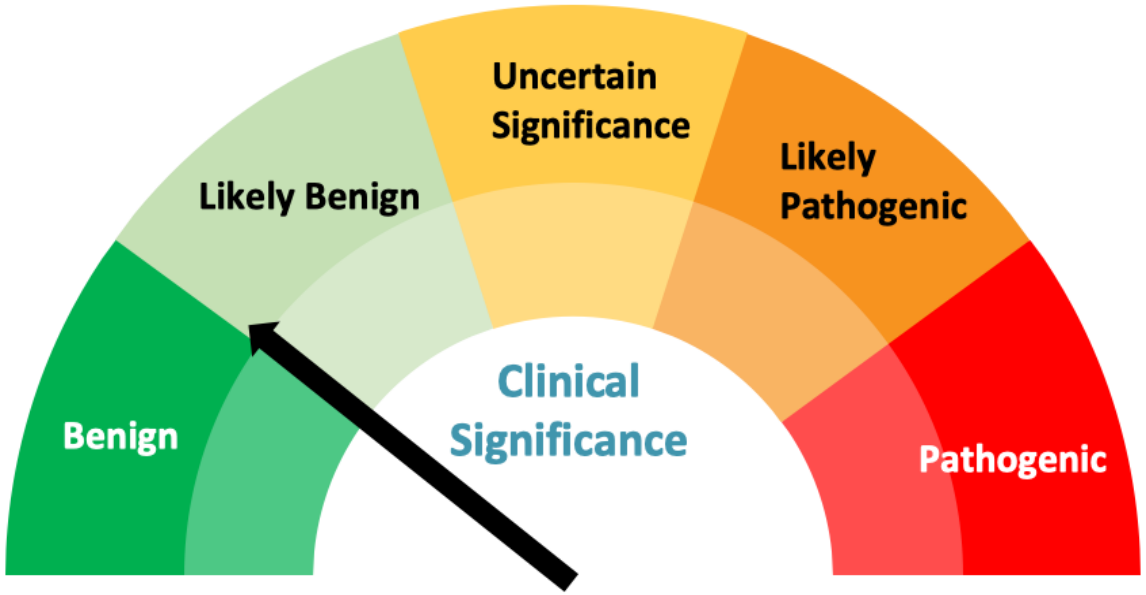
Variant Classification using ACMG/AMP Interpreting Sequence Guidelines

Mohamed Z. Alimohamed

Clinical Genomics

- Clinical genomics involve using the whole genome of a patient to diagnose diseases based on the variants found in their genome and other factors such as symptoms and family history. The variant information uncovered through whole genome sequencing can also help adjust management and treatment strategies for that particular patient and possibly his family members or relatives.

Categories for classifying the clinical significance of genetic variants

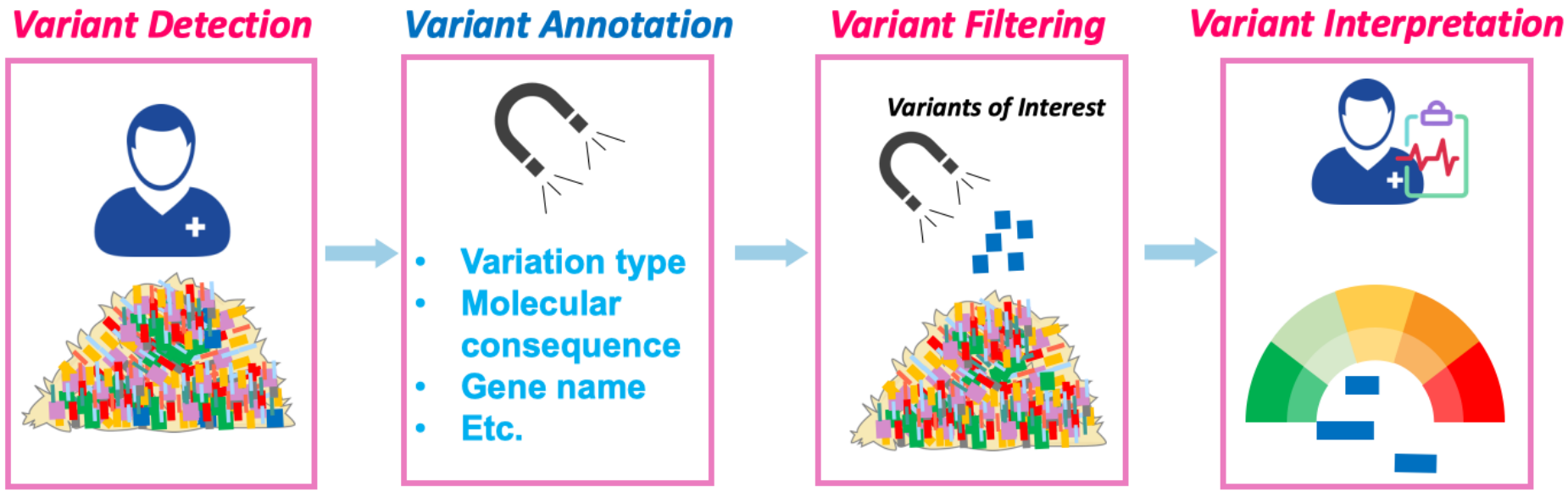


- 1. Benign: Variant for which there is sufficient evidence to support that they do not cause the suspected disease in a patient
- 2. Likely Benign: For variants having moderate, but not definitive evidence to show that it is not disease causing
- 3. Likely Pathogenic: Variants having moderate, but not definitive evidence to show that they are disease causing
- 4. Pathogenic: Variant for which there is sufficient well-documented evidence to support that they are disease causing
- 5. Uncertain Significance: Variants for which existing evidence is not sufficient to classify them as benign or pathogenic and additional evidence is needed to determine whether or not the variant is disease causing

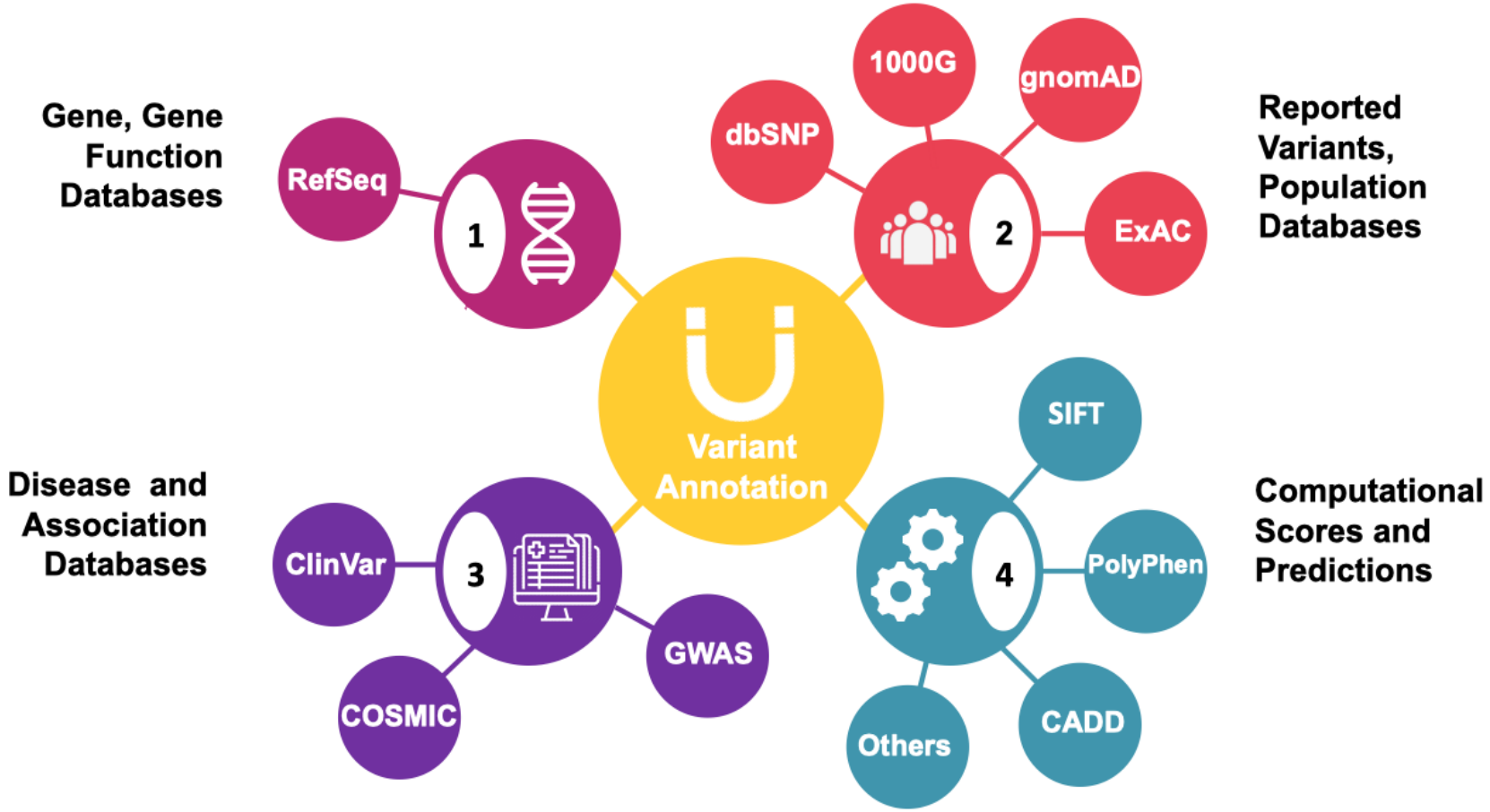
Variant Annotation

- To detect the possible impact of a variant in a clinical context, we need to know additional information about the variant. Assigning additional information against variants is known as variant annotation. Based on this additional information, a subset of variants that interest us can be picked out from a person's genome, thus helping us get closer to the task of detecting a variant of clinical importance. This additional information can include:
- 1. Variant type (Substitution, Deletion, Insertion, Duplication)
- 2. Molecular consequence of the variant (Missense, Nonsense, Frameshift, Non-Coding, UTR)
- 3. Gene in which the variant lies
- 4. Whether the variant lies in a coding or a non-coding region
- 5. Amino acid changes if the variant lies in a coding region
- 6. Whether the variant is already documented in databases and
- 7. The clinical significance previously reported for that variant (if any)

Steps involved in analyzing and interpreting genomic variants



Different types of information that can be annotated against variants and the respective databases that can be used to get that information



Databases for Annotating Variants

- 1. Gene Based Annotations:
- It is important to annotate variants with attributes like gene names, variant type, and amino acid changes. Gene based annotations also allow us to do a codon analysis and determine whether the variant is synonymous, nonsynonymous, frameshift, non-frameshift, nonsense (stop-gain) and so on.
- Gene based annotation can be done by mapping variants to the RefSeq database, which will provide the gene name in which the variant lies, the transcript details in HGVS notations, amino acid changes (if it is lying in a protein coding region) and other important information like variant type (intergenic, upstream, downstream, UTR, intronic, splicing or exonic) and molecular consequence

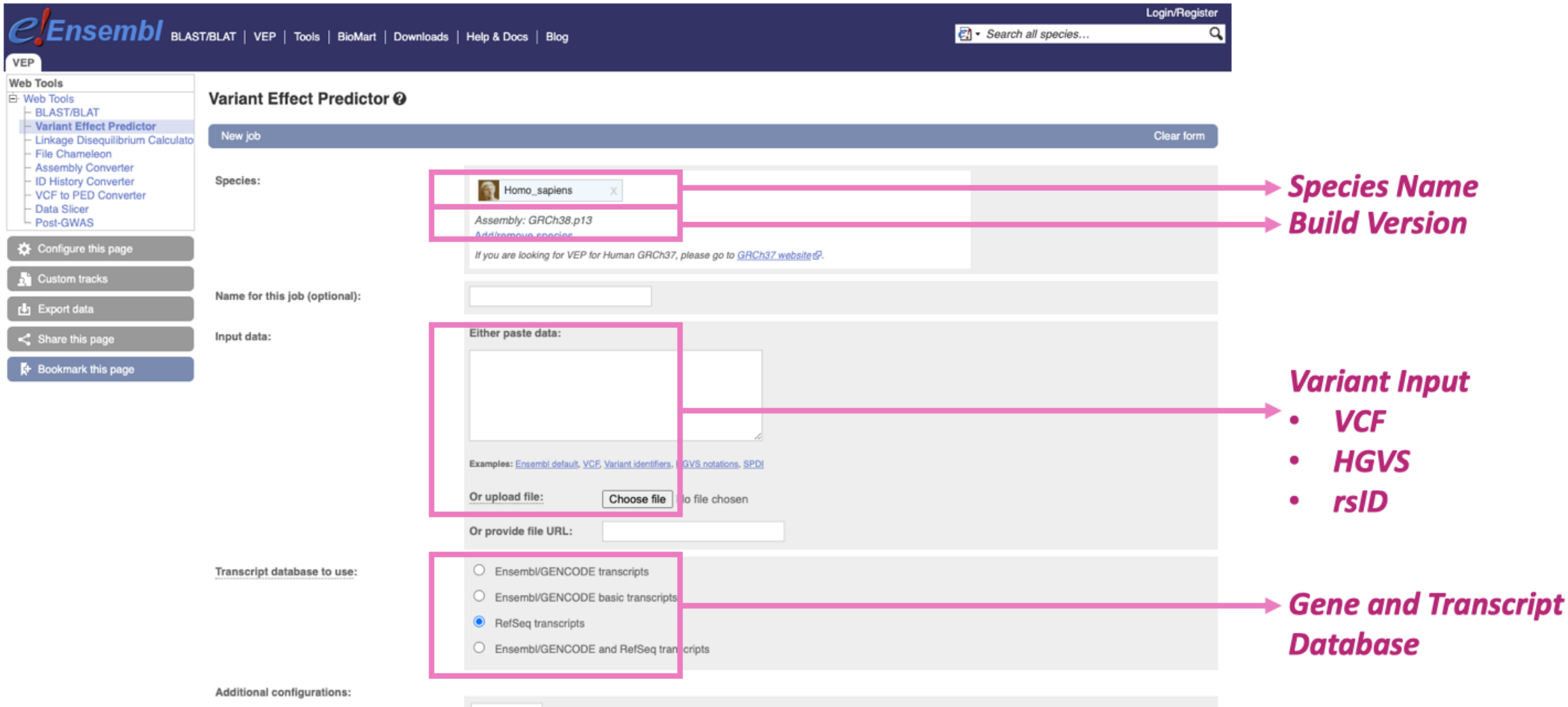
- 2. Annotating against Known Variant Databases:
- Checking if a variant is already reported in databases like dbSNP can help uncover additional information about the variant such as the publications that report the variant. A lot of variants and their allele frequencies in healthy individuals of different populations have been catalogued through projects like the 1000 Genomes Project and The Genome Aggregation Database (gnomAD). Mapping detected variants to such databases can help compare allele frequencies of the variant in different populations. A variant thus having a high allele frequency in a population is likely to be a benign variant while disease causing variants are expected to either have a low frequency of occurrence or not be reported in such population databases.

- 3. Annotating Known Clinical Phenotypes:
- Databases like ClinVar catalog genomic variants that have been reported in context of a disease along with their reported clinical significance and the evidence used to assign the significance. Searching and annotating detected variants against such databases will help uncover if the variant has been previously reported in context of any disease, including its clinical significance. For instance, a patient's variant previously reported as pathogenic in ClinVar for a particular disease provides a high amount of evidence to prioritize that variant as disease causing for the patient.

- 4. Annotating Computational Predicted Scores
- Many computational methods and tools have been developed that predict whether an amino acid change can affect normal protein function and thus result in disease. These tools are useful for gathering evidence against variants detected in a patient especially those that have not been reported previously. All such tools assign scores to the variants that can be used to analyze their predicted impact. These tools define a range of scores that are assigned, a cut-off value for the scores and the respective predictions based on the cut-off value

Tools for Variant Annotation

- Variant annotation tools and packages help ease the task of functionally annotating variants using various databases. Such tools provide a compact summary of the annotations to help analyze and interpret them in a clinical context.
- Variant Effect Predictor (VeP) is one such tool that helps in the analysis, annotation, and prioritization of genomic variants. It has various databases linked with it that help annotate important features such as gene name, variant type, codon information, transcript details, allele frequency in different databases, computational tool predictions and known disease phenotype along with the clinical significance. VeP is free to use and is also available as a web-browser based tool. It can be accessed at https://ensembl.org/Tools/VEP


Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search all species...
VEP
Web Tools
Web Tools
BLAST/BLAT
Variant Effect Predictor
Linkage Disequilibrium Calculato
File Chameleon
Assembly Converter
ID History Converter
VCF to PED Converter
Data Slicer
Post-GWAS
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Variant Effect Predictor
New job
Clear form
Species:
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:
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
Additional configurations:
Species Name
Build Version
Variant Input
VCF
HGVS
rsID
Gene and Transcript Database

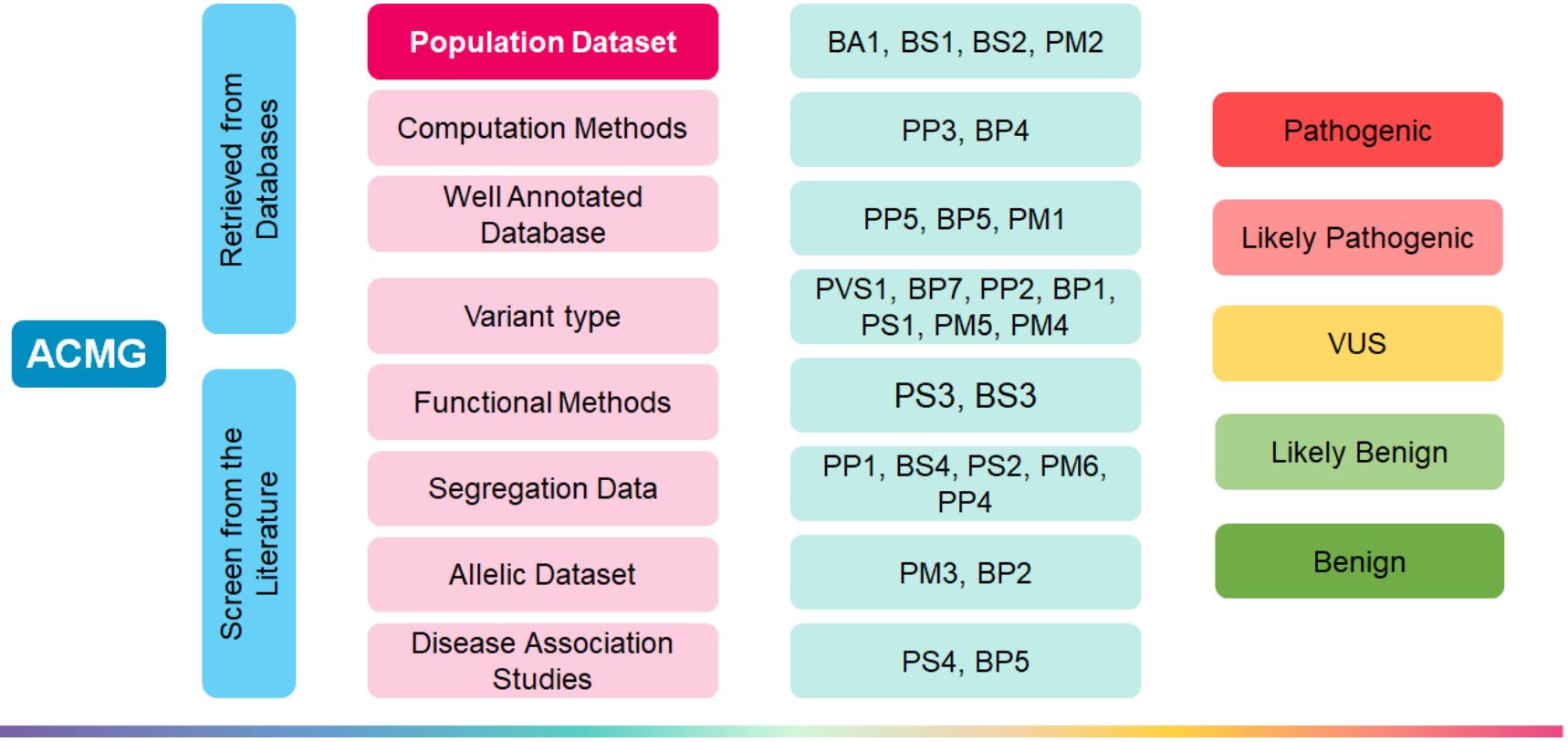
Features of VeP:

- VeP supports both GRCh38 (hg38) and CRCh37 (hg19) builds of the human reference genome
- Input File Types for VeP: VeP accepts different file formats or notations to input your variants for annotation. These include:
- 1. VCF format (Variant Call format file)
- 2. HGVS notations
- 3. Variant Identifiers in the form of rsIDs (dbSNP IDs)
- Transcript Database Options: For annotating gene names and transcript details, VeP provides 4 database options, including the RefGene database
- Supported Annotations:
- 1. Gene Based (Gene name, transcript details, codon information)
- 2. Variant Type and Molecular Consequence
- 3. Finding variants known variants databases like dbSNP
- 4. Computational scores and predictions
- 5. Allele frequency in 1000 Genomes Project, gnomAD and ESP6500 databases
- 6. Known disease phenotypes and clinical significance
- 7. Other annotations like conservation scores

ACMG/AMP guidelines based variant interpretation

- To address the above problem, the experts at the American College of Medical Genetics and Genomics (ACMG) and the Association of the Molecular Pathology (AMP) had put forth the 28 standard guidelines for the interpretation of the variants. These 28 guidelines have been represented and termed as the 28 ACMG attributes.
- The terminology of these attributes is according to the variant if it is disease-causing i.e. pathogenic then denoted by P and not disease-causing i.e. benign then denoted by B. Further pathogenic and benign guidelines were divided based on the guidelines' stringency. For the pathogenic variant, it is further classified into very strong denoted with VS, strong denoted with S, moderate with M, and supporting with P. Whereas, the benign variants are divided into further categories as stand-alone or very strong denoted with A, strong with S,
- and supporting with P. However the guidelines for benign variant do not have the moderate evidence. Further, based on the number of guidelines for the severity of the disorder, the attributes are further termed as: for very strong pathogenic and benign, there is a single guideline so it is named as PVS1 and BA1 respectively. While for the strong evidence within the pathogenic attribute, there are 4 guidelines, which are termed as PS1, PS2, PS3, and PS4, and so on for each guideline, the attributes have been termed.

ACMG/AMP guidelines based variant interpretation



Evidence Framework

	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*	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	Non-segregation with disease *BS4*		Co-segregation with disease in multiple affected family members *PP1*	Increased segregation data →		
De novo Data				*De novo* (without paternity & maternity confirmed) *PM6*	*De novo* (paternity & 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*			

Criteria for pathogenic classification

Very strong evidence of pathogenicity

PVS1 Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease

Caveats:

- Beware of genes where LOF is not a known disease mechanism (e.g. *GFAP, MTH7*)
- Use caution interpreting LOF variants at the extreme 3' end of a gene
- Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact
- Use caution in the presence of multiple transcripts

Strong evidence of pathogenicity

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

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

PS2 *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history

Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, *etc*. can contribute to non-maternity

PS3 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

PS4 The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls

Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR or OR does not include 1.0. See manuscript for detailed guidance.

Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Moderate evidence of pathogenicity

PM1 Located in a mutational hot spot and/or critical and well-established functional domain (*e.g.* active site of an enzyme) without benign variation

PM2 Absent from controls (or at extremely low frequency if recessive) (see Table 6) in Exome Sequencing Project, 1000 Genomes or ExAC

Caveat: Population data for indels may be poorly called by next generation sequencing

PM3 For recessive disorders, detected in *trans* with a pathogenic variant

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

PM4 Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants

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

Caveat: Beware of changes that impact splicing rather than at the amino

acid/protein level

PM6 Assumed *de novo*, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity

PP1	Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data
PP2	Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc) Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.
PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic etiology
PP5	Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for benign classification

Stand-Alone evidence of benign impact

BA1 Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC

Strong evidence of benign impact

BS1 Allele frequency is greater than expected for disorder (see table 6)

BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age

BS3 Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing

BS4 Lack of segregation in affected members of a family

Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.

Supporting evidence of benign impact

BP1	Missense variant in a gene for which primarily truncating variants are known to cause disease
BP2	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
BP3	In-frame deletions/insertions in a repetitive region without a known function
BP4	Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc) Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
BP5	Variant found in a case with an alternate molecular basis for disease
BP6	Reputable source recently reports variant as benign but the evidence is not available to the laboratory to perform an independent evaluation
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

Rules for combining criteria to classify variants

Pathogenic

1. 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)
2. ≥2 Strong (PS1–PS4) *OR*
3. 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

1. 1 Very Strong (PVS1) *AND* 1 Moderate (PM1–PM6) *OR*
2. 1 Strong (PS1–PS4) *AND* 1–2 Moderate (PM1–PM6) *OR*
3. 1 Strong (PS1–PS4) *AND* ≥2 Supporting (PP1–PP5) *OR*
4. ≥3 Moderate (PM1–PM6) *OR*
5. 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
6. 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Benign

1. 1 Stand-Alone (BA1) *OR*
2. ≥2 Strong (BS1–BS4)

Likely Benign

1. 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) *OR*
2. ≥2 Supporting (BP1–BP7)

PAH - NM_000277.2:c.1147C>T (p.Gln383Ter)

	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			

PAH established causal gene for phenylketonuria (PKU)

- Variant is loss-of-function and in a gene in which loss-of-function causes disease
  - **PVS1** applied
- Variant found in 1/16,254 (0.006%) in African chromosomes in gnomAD
  - **PM2** applied
- Variant observed in 1 proband with classic PKU who also carried variant p.Arg408Trp in trans (variant called Pathogenic by 18 labs in ClinVar)
  - **PM3** applied
- In total have PVS1, PM2, and PM3 applied = 1 Pathogenic VeryStrong and 2 Pathogenic Moderate

PAH - NM_000277.2:c.1147C>T (p.Gln383Ter)

	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			

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)

- In total have PVS1, PM2, and PM3 applied = 1 Pathogenic VeryStrong and 2 Pathogenic Moderate