

ClinGen General Sequence Variant Curation Process

Standard Operating Procedure

Version 1.0

April 2019

The Clinical Genome Resource
ClinGen Variant Curation SOP Committee

Introduction	1
General Guidance	3
How to do a Literature Search	4
Curated Literature Evidence	7
Criteria Evaluation Labels	8
Modified Strength of ACMG/AMP Evidence Codes	9
Evaluation Summary	9
Basic Information Tab	11
Transcript Information	12
Population Tab	14
Variant Type Tab	18
Missense sub-tab	18
Loss of Function (LOF) sub-tab	25
Silent & Intron sub-tab	27
In-frame Indel sub-tab	28
Experimental Tab	29
Case/Segregation Tab	32
Gene-centric Tab	40
Evidence Summary	40
List of Resources	42
ClinGen Resources	44
Appendix I - Computational Tools	45
In silico predictors linked in VCI	45
Conservation analysis tools linked in VCI	45
Splice site predictors linked in VCI	45
Appendix II - PVS1 flowchart, SVI recommendations	47
Appendix III - Combining Criteria to Classify Sequence Variants	48

Introduction

The ClinGen General Sequence Variant Curation standard operating procedure (SOP) is designed to provide guidance on variant classification using ClinGen approved processes and tools. Standardized assertion criteria to classify clinical sequence variants associated with Mendelian disorders into a five-tier nomenclature system were developed in “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”¹, and include pathogenic (P), likely pathogenic (LP), uncertain significance (VUS), likely benign (LB), and benign (B). This SOP is intended to guide variant curators through the process of curating and scoring evidence for the ACMG/AMP assertion criteria (population data, computational and predictive analysis, functional criteria, and allelic and co-segregation data) using the ClinGen Variant Curation Interface (VCI).

Information from publicly available resources and internal laboratory data is curated and scored with respect to a variant-disease relationship. A classification for each variant is assigned based on assigning ACMG/AMP evidence codes and the appropriate strength of the evidence in the assertion categories. This SOP provides a general overview of the best practices to follow for curating a variant-phenotype relationship and subsequently assigning a classification including general recommendations from the ClinGen Sequence Variant Interpretation Working Group (SVI), which can be found on the SVI webpage of the ClinGen website (<https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation/>). This SOP serves as a general guide to the process of variant classification and is intended to complement disease-specific Variant Curation Expert Panel (VCEP) specifications when available. **Please note, if applicable disease or gene-specific specifications exist, those should be used in conjunction with this SOP.** Please check the ClinGen SVI webpage for a listing of all available VCEP specifications, as well as with the VCEP with which you may be working, for the most up-to-date disease-specific guideline specifications that may be in progress.

This protocol is designed to be used in conjunction with the ClinGen Variant Curation Interface (VCI, <https://curation.clinicalgenome.org/>) and other ClinGen tools such as the allele registry (http://reg.clinicalgenome.org/redmine/projects/registry/genboree_registry/landing), which are publicly available. To register to use the VCI, please contact ClinGen at clingen-helpdesk@lists.stanford.edu. To create a login for the allele registry, please use the following site: http://reg.clinicalgenome.org/redmine/projects/clingen_users/register_clingen_user/cg_users/new. Instructions for VCI registration and login are available in the VCI Help document, which can be found here: <https://github.com/ClinGen/clincoded/wiki/VCI-Curation-Help>.

General Guidance

There are several general factors to consider upon beginning variant curation in the VCI. While the VCI supports the evaluation and interpretation of variants that range in size from 1-20 bases in length, it currently returns evidence only for single nucleotide variants. Please keep this in mind while reviewing the self-populated tabs within the VCI.

The variant curation process begins with entering a variant into the VCI (Figure 1). Instructions for how to enter a variant into the VCI are available in the previously mentioned VCI help document (<https://github.com/ClinGen/clin coded/wiki/VCI-Curation-Help>). Either a ClinVar ID or Allele Registry ID is needed. Once a variant has been entered, a curator will enter the VCI in the 'evidence' view. To begin interpretation, click on the 'Interpretation +' button located on the top right of the screen (Figure 2). This will take the curator to the 'Variant Interpretation Record' and the ACMG criteria will be listed across the top of the page. Here, the curator is given the opportunity to associate the variant with a specific disease and mode of inheritance (Figure 3). The curator should discuss with the Expert Panel (if appropriate) which disease term(s) are appropriate for the gene. Instructions on adding this term are available in the VCI help document. While it is possible to interpret a variant without entering a disease and/or inheritance pattern, as sometimes the best disease term is not known until the evidence is gathered, this data is required for ClinGen VCEPs per the FDA approval documentation. Therefore, a disease upon which the assertion is based must be added before any disease-specific criteria are added and before approving the classification. As criteria are selected, they will be marked and color-coded (Figure 4A), thus recording the interpretation progress and tracking the type and strength of criteria met as well as calculated pathogenicity (Figure 4B). By clicking the 'Interpretation +' button, you are creating an assessment of that variant that is editable by all members of the curation interface affiliation under which you are working. Any criteria applied will not be viewable to other affiliations, but the evidence you enter (e.g. publications) will be saved for all curators to view. Of note, if your affiliation already has an existing interpretation, you may edit that interpretation by clicking on the blue and white edit icon under 'My Interpretation'.

Figure 1. Variant curation process in the VCI.

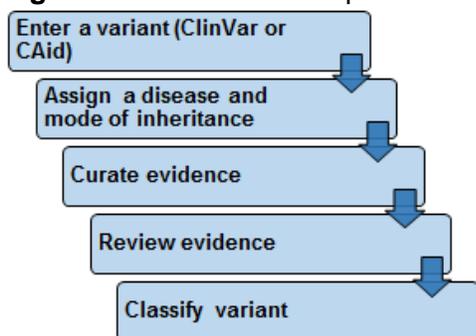


Figure 2. Variant Curation Interface (VCI) 'Evidence View'. To begin a variant interpretation, click on the 'Interpretation +' button.

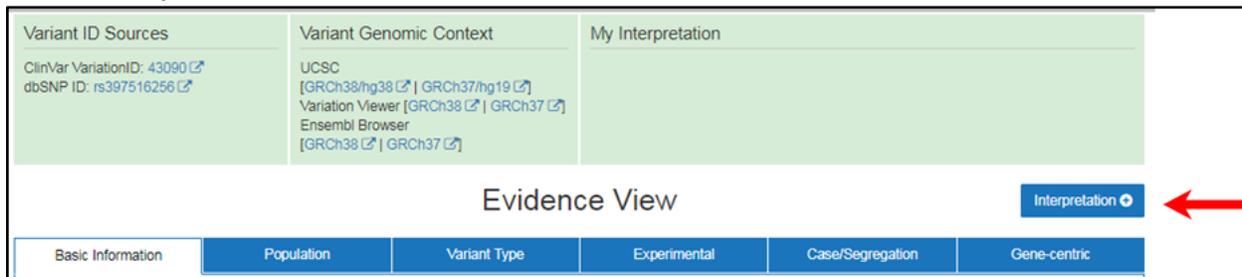
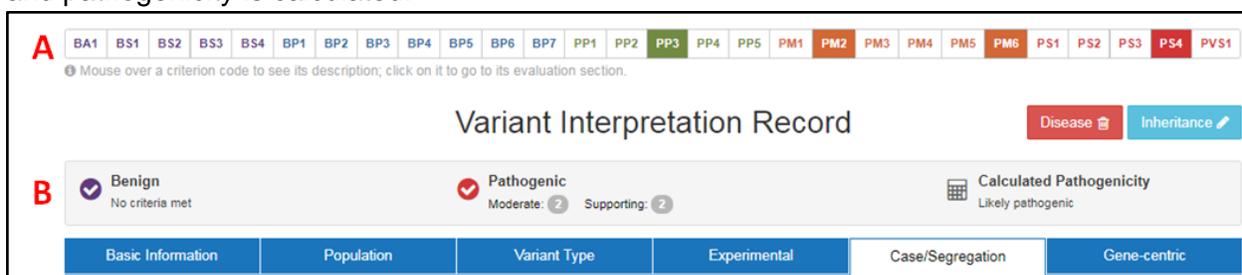


Figure 3. VCI 'Variant Interpretation Record' view. Here, the curator enters the disease and inheritance pattern.



Figure 4. A, Chosen criteria are marked and color-coded; B, Interpretation progress is recorded and pathogenicity is calculated.



How to do a Literature Search

An important first step for any curation is a comprehensive and effective literature search. In order to perform this, the biocurator must understand variant nomenclature, and know which resources are most useful in finding appropriate information. A comprehensive literature search will collect articles that allow the curator to apply ACMG criteria during the curation, including data on functional studies, case reports (*de novo*, segregation, phenotype, co-occurrence), molecular

characterization and case-control data (if available). Please keep in mind that the evidence collected may be both for and against pathogenicity.

Comprehensive literature searches will use a combination of search tools and databases. Examples of such that are publicly available or provide a version that is free of charge can be found in Table 1. Each has different strengths. For example, PubMed, while useful for scientific literature, will only identify a paper where the variant is contained in the title or abstract. Google has the capability of identifying variants found within the text of an article, as well as in supplemental tables; however, it will also find non-genetics related links. Google Scholar is limited to published academic literature, and is often better than Google at finding variants within a paper and within tables.

Table 1. Examples of publicly available databases and search tools.

Example Databases		
Human Gene Mutation Database* (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php)		
Gene-specific databases		
ClinVar		
ClinGen	Allele	Registry
http://reg.clinicalgenome.org/redmine/projects/registry/genboree_registry/landing)		
Example Search tools		
PubMed		
Google		
Google Scholar		
LitVar		
Mastermind** (https://mastermind.genomenon.com/)		

*HGMD is a commercial database that offers a less up-to-date version freely available to registered users from academic institutions or non-profit organizations.

** Mastermind is a commercial search engine that offers a basic plan free-of-charge.

To perform a comprehensive search, it is important to build an appropriate search term. Begin with the gene name and several versions of the variant nomenclature. A variant may have multiple identifiers for several reasons – strict Human Genome Variation Society (HGVS) format is not always followed, the Genome Reference Consortium human (GRCh) build has been updated (this is done periodically), alternate or historical gene names are used, an alternate transcript or non-standard nomenclature is used, etc. Some alternate transcripts and nomenclature can be found under the ‘basic information’ tab in the VCI and in the Allele Registry entry for the variant; this is discussed in more detail under this section in the SOP. Use of quotation marks, parentheses, and the AND/OR Boolean operators assist with collection of a broad, but not overwhelming, output. Examples of search strings that can be used for specific types of variants can be found in Figure 5.

Figure 5. Examples of search strings for different types of variants.

Variant Type	Lit search string 'HGNC gene symbol' AND ("variant codon nomenclature" OR "variant protein nomenclature" OR "variant GRCh location")
Nonsense	PTEN AND ("1003C>T" OR "1003 C>T" OR "Arg335Ter" OR "R335X" OR "Arg335STOP" OR "89720852")
Missense	NF1 AND ("277T>C" OR "277 T>C" OR "Cys93Arg" OR "C93R" OR "29486100")
Frameshift	MYBPC3 AND ("1028delC" OR "1028del" OR "Thr343MetfsX7" OR "T343MfsX7" OR "Thr343Metfs*" OR "T343Mfs*" OR "47367820")
Intronic	FBN1 AND ("1148-2A>C" OR "1148-2 A>C" OR "IVS10-2A>C" OR "IVS10-2 A>C" OR "48808561")
In-frame indel	MSH6 AND ("2157_2159delTAC" OR "2157delTAC" OR "2157del3" OR "Thr720del" OR "T720del" OR "48027279")
Different amino acid change at same residue	Variant of interest: TP53 His179Asn TP53 AND ("His179*" OR "H179*") – wouldn't include "NOT His179Asn" because one article might discuss both variants.

Search output can range from few (or zero) to hundreds of articles. The list will most likely include multiple links from the same source, including those that may have already been viewed, such as ClinVar or gnomAD, so the original output list can be reduced. While all links should be explored, not all will contain relevant data. For example, an article may mention the gene of interest, and

also the variant of interest but found in a different gene, or may describe the variant of interest but the paper may discuss downstream care of patients with the disease.

When reviewing variant specific literature, it is important to avoid double counting of probands. The same patient or family may be presented in multiple publications, and while this is not always readily apparent, there are certain clues to suggest it. Things to look for include author overlap between articles, use of the same study cohort, highly aligned specific clinical details, etc. If unsure, it is best to contact the authors and/or discuss inclusion of this data with the Expert Panel (if applicable).

Once it has been verified that the specific variant of interest is included in the publications, this literature can be collected and reviewed, and the data can be entered into the appropriate text boxes, with PMID, in the VCI as described below.

Curated Literature Evidence

Each of the tabs that collect ACMG/AMP criteria evidence contains a subsection entitled 'Curated Literature Evidence'. In this section, there is opportunity to add PMIDs to document publication data that support those criteria. Clicking on the "Add PMID" button provides a pop up box (Figure 6) where a PMID may be entered. The VCI will automatically retrieve the abstract from PubMed. Double check that the correct article has been retrieved and add the article. Once the article is added, a free text 'Evidence' box will appear where the curator can document relevant data from the article.

Figure 6. The "Add PMID" pop-up found in all tabs in the VCI. PMIDs are entered and automatically retrieved, and a free text evidence box appears once the article is added.

The image shows two screenshots of a web interface. The top screenshot is a pop-up window titled "Add new PubMed Article". It contains a text input field labeled "Enter a PMID *", a "Retrieve PubMed Article" button, and "Cancel" and "Add Article" buttons at the bottom. The bottom screenshot shows the "Curated Literature Evidence (Population)" section. It displays a retrieved article snippet: "Richards S et al. 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. 2015 May;17(5):405-24. PMID: 25741868". Below this is an "Evidence:" text area and "Edit PMID", "Save", and "Cancel" buttons.

For each reference, there are several pieces of data that should be captured. These include, but are not limited to, the following:

1. Details on proband phenotype
2. Proband race/ethnicity
3. Whether each proband described in the paper is heterozygous, homozygous, hemizygous, a compound heterozygote or a double heterozygote. Include other known or possible clinically relevant variants and their classification if possible.
4. Family history relevant to the disease being curated
5. *De novo* occurrence of any variants
6. Co-segregation data, including number of times the variant segregates with the disease, as well as the number of times the variant does not segregate with disease. Please keep in mind age of onset as well as penetrance. It is also relevant to include the number of affected and the number of unaffected individuals.
7. Control cohort information - number of observations in controls and number of controls tested, ethnicity, phenotype evaluation, etc.
8. Functional data

Limitations:

- At the current time, the VCI only allows for use of published data with a PMID. Pre-prints, abstracts, or other sources without a PMID cannot be cited in a linkable manner. To account for data not in PMIDs, a note describing the evidence and its source can be added to the text box for a given ACMG/AMP code application.
- A new page to allow input of variant data from these other data sources, including clinical laboratories is under development and will be included in the next release of this SOP.
- Some cohorts are published in more than one publication. If this is suspected, make a note of this to avoid over-counting.

Criteria Evaluation Labels

For each criterion, there are at least three evaluation label choices: Met, Not Met and Not Evaluated. It is important to understand how these labels are used, and how information considered when applying each label is captured and published in the final classification. To that end, the following definitions will be used to define each label:

Met: This label will be used if evidence meets specified rules for a given criterion. Curators are encouraged to add an explanation for usage that begins with the criterion code in the free text explanation box next to each criterion (e.g., PM2 is met because this variant is not found in gnomAD...). The curator should include all relevant PMIDs and data (e.g., PM2 Met; In PMID xxxx, ...). All explanation notes and PMIDs (with corresponding PMID notes) will be captured and published to the ClinGen Evidence Repository (<https://erepo.clinicalgenome.org/evrepo/>) in association with MET codes.

Not Met: This label is to be used if evidence is evaluated and determined **not** to meet the criterion. Again, curators are encouraged to add an explanation for not applying the code beginning with the criterion code and including relevant PMIDs; for example, “PM2 is not met as there are 100 instances of the variant in gnomAD”. All explanations will be captured and published to the ClinGen Evidence Repository.

Not Evaluated: This label is the default label and is currently intended to be used if there is no evidence to evaluate OR if the criterion is not applicable for the variant. It is helpful if the curator documents whether the label was actively chosen as opposed to the code not yet being considered; however, Not Evaluated codes are currently not captured and therefore any added notes are not published to the ClinGen Evidence Repository, nor are any relevant captured PMIDs.

Modified Strength of ACMG/AMP Evidence Codes

The ACMG/AMP guidelines for variant interpretation include the option of using professional judgement to move criteria listed as one weight to another weight. VCEPs are free to utilize these strength modifications in their specifications. These weight modifications are especially important for quantitative evidence types such as co-segregation with disease in affected family members (PP1), which can increase in strength with increasing segregation data. Alternatively, strength of given criteria can be downgraded in certain instances. For example, a well-established functional study supportive of a damaging role has a strong level of strength (PS3); however, it may be lessened to moderate or supporting with weaker or less definitive types of functional studies. The SVI has recommended standardized nomenclature to address these changes (https://www.clinicalgenome.org/site/assets/files/8490/svi_criteria_nomenclature_recommendation_v1_0.pdf). For strength-modified evidence, the original criteria code is followed by an underscore and the new level of strength (e.g. PP1_Strong, PP1_Moderate, PP1).

Evaluation Summary

Once all evidence has been evaluated, a classification can be made. The VCI will aggregate the selected criteria using the criteria combining rules outlined by the ACMG/AMP guidelines. This calculated classification, saved provisional and approved interpretation(s), criteria meeting an evaluation strength, criteria evaluated as “Not met” and criteria ‘Not yet evaluated” can all be found by accessing the evaluation summary via the ‘View Summary’ button shown in Figure 7. Please note that currently, any VCEP modifications to the original ACMG/AMP combining rules must be manually applied by overriding the classification with an explanation as described below.

Figure 7. The 'View Summary' button in the VCI accesses the evaluation summary page.

The screenshot displays a variant record for NM_000257.3(MYH7):c.5740G>A (p.Glu1914Lys). The variant is associated with dilated cardiomyopathy and has an autosomal dominant inheritance pattern. A red arrow points to the 'View Summary' button in the top right corner. The main content area is divided into three columns: Variant ID Sources (ClinVar VariationID: 43088, dbSNP ID: rs397516254), Variant Genomic Context (UCSC, GRCh38/hg38, GRCh37/hg19, Variation Viewer, Ensembl Browser), and My Interpretation (Disease: dilated cardiomyopathy, Calculated Pathogenicity: Likely pathogenic, Modified Pathogenicity: Not provided, Provisional/Approved Status: APPROVED, Interpretation Last Saved: 2018 Nov 05, 1:11 pm). Below the columns is a row of criterion codes (BA1, BS1, BS2, BS3, BS4, BP1, BP2, BP3, BP4, BP5, BP6, BP7, PP1, PP2, PP3, PP4, PP5, PM1, PM2, PM3, PM4, PM5, PM6, PS1, PS2, PS3, PS4, PVS1) with a tooltip that reads 'Mouse over a criterion code to see its description; click on it to go to its evaluation section.' The 'Variant Interpretation Record' section shows 'Benign' (No criteria met) and 'Pathogenic' (Moderate: 2, Supporting: 2) with a 'Calculated Pathogenicity' of 'Likely pathogenic'. There are also buttons for 'Disease' and 'Inheritance'.

On the evaluation summary page, there is also the opportunity to modify the calculated classification. This can be done based on expert opinion/clinical judgment, must have the approval of the Expert Panel or approving body, and the rationale must be clearly documented in the 'explain reason(s) for change' free text box located below the modified pathogenicity classification selection box. This explanation is required, and must also be included in the evidence summary for public display.

Basic Information Tab

This tab provides high-level information about nomenclature and ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) assertions of the variant chosen for interpretation. Each section is detailed below.

Genomic subsection: This section lists the proper HGVS notation for the genomic coordinates on the two most recent genome builds, GRCh37 and GRCh38. Genome builds can differ across resources such as population databases, so it is critical to double check the genome build when curating a variant.

ClinVar interpretation: The review status of the overall variant record is provided for the curator, along with the aggregated clinical significance, total number of submissions and the date last evaluated. This helps to alert the curator if the variant is expert-reviewed and to the concordance of the interpretation (i.e. if there are any conflicts). As published literature and information about a variant can change often, it is important to keep the date last evaluated in mind. In addition, this section may change over time as you classify the variant, as new data may be submitted to ClinVar.

The ClinVar entry can be accessed directly by clicking the “See data in ClinVar” arrow (A) in Figure 8 below. The ClinVar entry provides access to a lab’s summary evidence for the variant, which provides a short description of their rationale for classification. This can be found on the Summary Evidence Tab of ClinVar, in the Description column. Many ClinVar entries have links to publications that contain relevant information to the interpretation of the variant. This information can be found in the Citations column.

Figure 8. Basic tab in the Variant Curation Interface.

Overall ClinVar Interpretation					See data in ClinVar ↗
Review status:	Criteria provided, multiple submitters, no conflicts			Last evaluated:	Feb 14, 2017
Clinical significance:	Uncertain significance			Number of submission(s):	3
Interpretations Submitted to ClinVar (Germline SCVs only)					A See data in ClinVar ↗
Clinical significance (Last evaluated) B	Review Status (Assertion method) C	D Condition(s) (Mode of inheritance)	E Submitter - Study name	F Submission accession	
Uncertain significance (Apr 28, 2015)	criteria provided, single submitter EGL Classification Definitions	not specified [MedGen]	EGL Genetic Diagnostics, Eurofins Clinical Diagnostics ↗	SCV000231954.4	
Uncertain significance (Mar 25, 2015)	criteria provided, single submitter LMM Criteria	not specified [MedGen]	Laboratory for Molecular Medicine, Laboratory for Molecular Medicine (Partners HealthCare Personalized Medicine) ↗	SCV000272892.2	
Uncertain significance (Feb 14, 2017)	criteria provided, single submitter Counsyl Autosomal Recessive and X-Linked Classification Criteria (2018)	Usher syndrome, type 2A [MedGen] OMIM Retinitis pigmentosa 39 [MedGen] OMIM	Counsyl ↗	SCV000789566.1	

Information is also provided about each individual ClinVar submission, including the clinical significance (B), the date it was last evaluated (B), the review status (C), the assertion method (C), condition (D), mode of inheritance (D), the submitter (E), if a study name is associated with the submission (E), and the submission accession (SCV - submission to ClinVar) number (F). For more information about ClinVar terminology, please see Harrison et al 2016².

Please keep in mind that variants can have multiple interpretations based on the condition, mode of inheritance or whether reported as a somatic or germline variant. Therefore, it is critical to crosscheck information in column D with the condition and inheritance for the variant-disease association of interest. In general, ClinVar is best used to identify sources of published or unpublished data, and review other laboratory's rationales for classification. If an assertion in a ClinVar submission is discordant with the classification arrived at in the VCI or by the VCEP, it may be useful to reach out to the submitting laboratory to discuss the basis for classification.

Transcript Information

Below the ClinVar submission section is information regarding transcripts. A sequence variant can fall on multiple transcripts and may have differing molecular consequences depending on the transcript evaluated. Generally, although a primary transcript is chosen for evaluation, it is critical to evaluate the effect of the variant on all transcripts when proposing a variant classification, especially if the biologically and disease-relevant transcript is uncertain. This information is easily viewable under the Protein change and Molecular Consequence columns of this section. Most often, the longest transcript is chosen as the primary transcript, unless the variant is predicted to have a more severe impact on another transcript (e.g., if a variant is predicted to result in an intronic change on the longest transcript and a nonsense change on a shorter transcript) or a different transcript has been reported as primary or most biologically relevant. It is important to keep in mind the disease of interest when performing transcript evaluation. For more information on choosing transcripts for variant interpretation, please see DiStefano et al 2018³.

ClinVar primary transcript: The ClinVar primary transcript is a RefSeq transcript designated by ClinVar. The choice of the transcript is largely driven by which transcript(s) have previously been submitted to ClinVar for that particular variant.

RefSeq Transcripts: RefSeq is a transcript annotation and curation effort headed by the National Center for Biotechnology Information (NCBI). These transcripts are most commonly used by clinical labs, are independent of the genome build, and many are supported by manual literature curation. Transcripts can be preceded by multiple prefixes:

- NM, NR, NP: mRNA, non-coding RNA, and protein transcripts, respectively that have been curated by the RefSeq team and are supported by some evidence, whether it is published literature or GenBank cDNA or EST data.

- XM, XR, XP: mRNA, non-coding RNA, and protein transcripts, respectively that are predicted, but not confirmed with curation or evidence.

Ensembl Transcripts: Ensembl is a human genome annotation effort headed by the European Bioinformatics Institute (EBI) that includes a transcript annotation and curation effort. These transcripts are used by gnomAD/ExAC, are dependent upon the genome build, and are supported by computational and/or manual curation.

Although transcript mapping between these two resources can be difficult at times, a current effort is underway to harmonize RefSeq and Ensembl transcripts called The Matched Annotation from the NCBI and EBI (MANE) project. For more information on the MANE project, please see the Ensembl blog post: <http://www.ensembl.info/2018/10/12/our-new-joint-transcript-initiative-the-matched-annotation-from-the-ncbi-and-ebi-mane-project/>.

Within the VCI, the canonical transcripts for both RefSeq and Ensembl (defined here as the transcript with the longest coding sequence if the gene has translated transcripts, or the longest cDNA if it does not) are indicated by an asterisk.

External Resources: Links are provided (for each genome build) to navigate to UCSC (a genome browser developed by the University of California Santa Cruz), Variation Viewer (NCBI's genome browser), and Ensembl Browser (EBI's genome browser).

Population Tab

ACMG/AMP criteria codes: BA1, BS1, PM2

Background: This tab displays population frequency data, if available, from Genome Aggregation Database (gnomAD), Exome Aggregation Consortium (ExAC), Population Architecture using Genomics and Epidemiology (PAGE), 1000 Genomes (1000G) and the Exome Sequencing Project (ESP), including minor allele frequency (MAF) and genotype data (Figure 9A). The subpopulation and database pair with the highest MAF for the variant of interest is displayed in the "Highest Minor Allele Frequency" section (Figure 9B). For example, for the variant displayed below, the highest MAF (0.08194) is found in the East Asian subpopulation in gnomAD (Figure 9D, red box). These data are then displayed in the "Highest Minor Allele Frequency" section along with subpopulation, source, total number of variant alleles, and total number of alleles tested in that subpopulation. The VCI defaults to showing the upper and lower bounds of the 95% confidence interval to help users determine the level of confidence in the computed MAF.

Limitations:

- **Data is currently only auto-populated for nucleotide substitution variants.** If assessing a deletion or insertion variant, please manually search the population frequency resources using either the dbSNP ID (provided in the *Variant ID Sources* section at the top of the page) or the genomic positions (provided in the *Genomic* section on the *Basic Information* tab). The ClinGen Allele Registry also provides direct links to the population frequencies and is faster than some of the other searches. Curators can view variants identified in ExAC and gnomAD in the region (+/- 30 bp) of the deletion/insertion by clicking the links provided in the Population tab (Fig 9E).

Population Tab Criteria Evaluations:

General instructions:

- Please confirm variant information is correct by checking the genomic coordinates and/or dbSNP IDs.
- As criteria codes BA1, BS1, and PM2 are mutually exclusive, only one of the three criteria are allowed to be "Met."

BA1: Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium

- The default BA1 threshold (5%) is displayed in the MAF cutoff field.
- This threshold is often modified in a gene/disease-specific manner by the appropriate VCEPs, to account for variability of disease prevalence and genetic heterogeneity. Some VCEPs have also defined a set of variants (exclusion variants) whose frequency surpasses the designated BA1 threshold in one or more population data sets, but have a plausible argument for pathogenicity and should be exempted from this rule. For example, *BTD* NM_000060.4:c.1330G>C [p.Asp444His] has a MAF >5% in gnomAD (0.05558 in the Finnish population) but is classified as Pathogenic (ACMG/AMP criteria also applied:

PS3; PM3_Strong; PP3; PP4); therefore, the BA1 criteria is not applicable. Please check with the VCEP with which you are working for any specifications of their BA1 rule.

- The SVI has provided an updated definition of this criterion which states that “Allele frequency is >0.05 in any general continental population dataset of at least 2000 observed alleles and found in a gene without a gene- or variant-specific BA1 modification”. They have also provided an exclusion set of nine variants (exclusion variants defined above) with an allele frequency over 5% where BA1 does not apply⁴.
- The SVI has provided a mechanism for identifying further alleles that should be excluded from this BA1 criteria to be shared publicly through a submission form located on the ClinGen webpage (<https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation/>)

BS1: Allele frequency is greater than expected for disorder

- This is disease-specific. Please consult with the specific Expert Panel (if applicable) for details.
- The incidence of the disorder being analyzed can be found using the resources located in the resource section of this SOP (such as the Genetic Testing Registry). If the variant is found more frequently in the population (or at least one subpopulation in ExAC or gnomAD) than the disease occurs, this is considered strong evidence of a benign impact.

PM2: Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium

- Although the original code states “absent”, many VCEPs have modified this code in a disease-specific manner. Please consult with the specific Expert Panel (if applicable) for details.
- As with BS1, incidence of the disorder being analyzed can be found using available resources. If the variant is found at a frequency below the expected carrier frequency of a given autosomal recessive condition, this is considered moderate evidence of a pathogenic impact.
- Since coverage information is not currently displayed (for variants present or absent in population databases), please use the link-out to navigate to the population database to determine if the region has adequate coverage (typically >20X coverage).
If the variant is present in a population database, a ‘filter’ button will appear above the pre-populated data. If coverage of the region is adequate, this button will be a green ‘Pass’ (Figure 9D).
- Since the VCI is only able to display population frequency data for substitutions, indel variants may appear as “absent” when frequency data is available (Fig 9E). When assessing an indel variant, please click the “View the coverage of this region (+/- 30 bp)...” hyperlink to view the region and determine if the variant of interest is present. Additionally, the ClinGen Allele Registry links indels to gnomAD and can be used to determine if population frequency data is available.

Evidence Explanation Notes:

- In the Explanation field (9C), please list the frequency, source of data used, and the allele counts. For example, “Variant identified in 8.19% (1546/18868) of East Asian chromosomes in gnomAD” or, if application is based on the 95% CI, then “The filtering allele frequency is 7.81% for East Asian chromosomes by gnomAD (1546/18868 with 95% CI)”, for the variant below. Although the date you are working on the variant is maintained in the system, it is also helpful to specify in your comment the date on which this information is captured.

Figure 9. Population Criteria Evaluation. Variant with allele frequency (AF) returned. Only gnomAD and ExAC are displayed due to sizing.

A
Population Criteria Evaluation

BA1: Allele frequency is > 5% in ExAC, 1000 Genomes, or ESP

BS1: Allele frequency greater than expected due to disorder Disease-specific

PM2: Absent from controls (or at extremely low frequency if recessive) in ExAC, 1000 Genomes, or ESP

BA1:

- or -

BS1:

- or -

PM2:

MAF cutoff: %

Explanation:

Variant identified in 8.19% (1546/18868) of East Asian chromosomes in **gnomAD**

Evaluations for BA1, PM2, BS1 saved successfully!
Update

B

Subpopulation with Highest Minor Allele Frequency

Note: this calculation does not currently include PAGE study minor allele data

This reflects the highest MAF observed, as calculated by the interface, across all subpopulations in the versions of gnomAD, ExAC, 1000 Genomes, and ESP shown below.

<p>Subpopulation: East Asian</p> <p># Variant Alleles: 1546</p> <p>Total # Alleles Tested: 18868</p>	<p>Source: gnomAD</p> <p>Allele Frequency: 0.08194</p> <p>Desired CI: <input type="text" value="95"/></p> <p>CI - lower: 0.07811</p> <p>CI - upper: 0.08594</p>
---	--

Note: ExAC Constraint Scores displayed on the Gene-centric tab

C

D **gnomAD 13:20763612 C/T (GRCh37)**  [See data in gnomAD](#)

Exomes Filter: Pass
 Genomes Filter: Pass

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	1546	18868	88	0.08194
Ashkenazi Jewish	78	10126	0	0.00770
Other	30	6452	0	0.00465
Latino	92	34420	1	0.00267
European (Finnish)	44	25776	0	0.00171
European (Non-Finnish)	182	126030	1	0.00144
African	25	24028	1	0.00104
South Asian	14	30750	0	0.00046
Total	2011	276450	91	0.00727

ExAC 13:20763612 C/T (GRCh37)  [See data in ExAC](#)

Filter: Pass

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	626	8644	38	0.07242
Other	2	908	0	0.00220
Latino	25	11578	0	0.00216
European (Non-Finnish)	129	66682	1	0.00193
European (Finnish)	6	6614	0	0.00091
African	9	10406	0	0.00086
South Asian	2	16468	0	0.00012
Total	799	121300	39	0.00659

Variant with no AF returned

E **gnomAD 17:41215361 T/- (GRCh37)** 

Data is currently only returned for single nucleotide variants. [View the coverage of this region \(+/- 30 bp\) in gnomAD](#) for this variant.

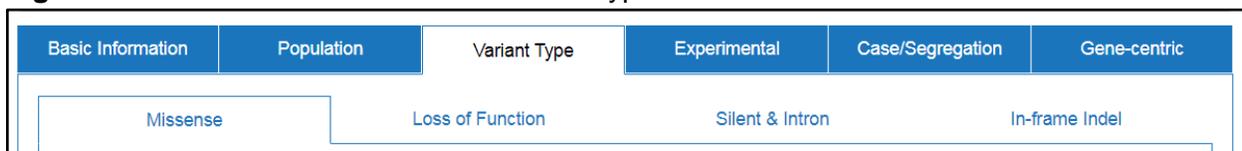
ExAC 17:41215361 T/- (GRCh37) 

Data is currently only returned for single nucleotide variants. [View the coverage of this region \(+/- 30 bp\) in ExAC](#) for this variant.

Variant Type Tab

Under the parent Variant Type tab are four sub-tabs separated out by variant type (Missense, Loss of Function, Silent & Intron, and In-frame Indel (Figure 10) with each sub-tab containing the ACMG/AMP criteria specific for that variant type. For example, criteria PS1 and PM5 (missense change in a codon with other pathogenic missense changes) are only found on the Missense tab, while PVS1 (LOF variant) is only found on the Loss of Function tab. Thus, **curators only need to complete the Variant Type sub-tab specific for their variant.**

Figure 10. Sub-tabs found under the Variant Type tab in the VCI.



Missense sub-tab

Functional, Conservation, and Splicing Predictors Section

ACMG/AMP criteria codes: PP3, BP4, PP2, BP1

Background: For missense variants, multiple prediction scores are displayed and separated out by type. The first type, “ClinGen Predictors” (Figure 11A), displays results from the meta-predictor REVEL⁵, which predicts the pathogenicity of missense variants based on a combination of scores from 13 individual tools: MutPred, FATHMM v2.3, VEST 3.0, Polyphen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP, and phastCons (please see Appendix I, Computation Tools, for links to each individual tool). The REVEL score can range from 0 to 1, with 1 most suggestive of pathogenicity. As many of the individual tools incorporated by REVEL are also displayed in the VCI (Other Predictors), the score from REVEL is most useful in assessing whether, in aggregate, the tools are suggestive of a pathogenic or benign impact. In general, a score > 0.75 is considered evidence of pathogenicity, though some VCEPs have specified different cutoffs for PP3 and BP4 or recommended the combination of REVEL and other predictors.

Below “ClinGen Predictors”, “Other Predictors” (Figure 11B) displays results from 13 *individual* prediction tools (SIFT, PolyPhen2-HDIV, PolyPhen2-HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, PROVEAN, MetSVM, MetLR, CADD, FATHMM-MKL, fitCons) and includes both the quantitative score and prediction classification from each tool. While there is some overlap between these 13 prediction tools and the 13 tools aggregated in REVEL, they are not identical. Informational links for these 13 individual tools can also be found in Appendix I, Computational Tools. As some prediction tools assess the impact of the missense change on all transcripts, these tools may display multiple quantitative scores and prediction classifications for

a single variant. To determine the prediction classification text from each single letter abbreviation displayed, hover over the information button  (Table 2).

The next section, Conservation Analysis (Figure 11C), contains links to, and displays results from, six conservation analysis tools. The conservation of a nucleotide at a particular position in the genome can give an indication of its importance. Selective constraint can be used to assess functional significance of a variant. There are multiple tools available for conservation analysis. The VCI utilizes several programs from the Phylogenetic Analysis with Space/Time Models (PHAST) software package, including phyloP100way, phyloP20way, phastCons100way and phastCons20way. The score for each is automatically calculated and provided in the VCI. Again, many of the predictors previously describe incorporate these conservation scores.

- PhyloP is a suite of programs that measure the calculated p-value of conservation versus acceleration (non-conserved) at each nucleotide across an aligned area, as compared to a model of neutral conservation. The absolute value score range is -14 to 3. A negative value represents faster than expected evolution, while a positive score indicates slower than expected evolution (areas predicted to be conserved). Information and tutorials can be found at the following link: <http://compugen.cshl.edu/phast/index.php>.
- PhastCons calculates conservation using a Phylogenetic Hidden Markov Model. The score ranges from 0 to 1, with a lower score indicating faster than expected evolution and a higher score indicating conservation among 17 vertebrate species. Additional information and tutorials for phastCons100way and phastCons20 way can be found here: <http://compugen.cshl.edu/phast/index.php>.
- The Genomic Evolutionary Rate Profiling (GERP++) (<http://mendel.stanford.edu/SidowLab/downloads/gerp/>) program estimates position-specific evolutionary constraint by measuring the number of substitutions expected minus those observed under neutral drift at each position independently. The calculated score varies with the level of conservation; positive scores indicate evolutionary constraint. Thresholds are chosen based on desired sensitivity and specificity, and may be dictated by VCEP; discuss with VCEP if applicable.
- SiPhy, available through the Broad Institute (http://portals.broadinstitute.org/genome_bio/siphy/index.html), uses deeply sequenced phylogeny data and multiple statistical tests to identify conservation patterns.

The last section under ‘Functional, Conservation, and Splicing Predictors’ is Splice Site Predictors (Figure 11D). Splice site predictors utilize bioinformatics tools and splice site consensus sequences to predict variant effects. While the VCI offers dynamic links to three of these tools - MaxEntScan, NNSPLICE and HumanSplicingFinder, it does not automatically import data from these tools. The curator must access and utilize these tools via the links provided if they wish to use this data. Many VCEPs have specified the splice predictor tools that are being used in variant classification as described in Limitations below.

- MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) is based on the 'Maximum Entropy Principle', and utilizes probabilistic models of short sequence motifs to account for non-adjacent and adjacent dependencies between nucleotide positions. The score generated is the difference between a reference allele and a variant, and a higher score implies a higher probability of a true splice site.
- NNSPLICE (http://www.fruitfly.org/seq_tools/splice.html) is part of the Berkeley Drosophila Genome Project and utilizes a neural network method trained to recognize 5' and 3' eukaryotic splice sites using a representative data set from *D. melanogaster*. The score ranges from 0-1, with anything above 0.5 indicative of a possible splice site gain.
- Human Splicing Finder (<http://www.umd.be/HSF3/HSF.shtml>) is a tool created to study pre-mRNA splicing by calculating consensus values of potential splice sites and branch points. Twelve algorithms are combined to identify various sequences that affect splicing in different capacities. Each algorithm calculates the potential for various types of splicing signals and although they have the same consensus values (0-100), the threshold is defined differently for each. In general, the higher the score, the more likely a potential splicing motif.

Figure 11. Predictors section of Missense tab in the VCI.

A ClinGen Predictors				
Source	Score Range	Score	Impact Threshold	Prediction
REVEL 🔗 (meta-predictor)	0 to 1	0.658	>0.75	higher score = higher pathogenicity
B Other Predictors 👉				
Source	Score Range	Score	Impact Threshold	Prediction 🔗
SIFT 🔗	--	0.001	<0.049	D
PolyPhen2-HDIV 🔗	0 to 1	0.606	--	P
PolyPhen2-HVAR 🔗	0 to 1	0.596	>0.447	P
LRT 🔗	0 to 1	0.01421	--	N
MutationTaster 🔗	0 to 1	0.999997, 0.999988	>0.5	N, N
MutationAssessor 🔗	-0.5135 to 6.49	2.7	>1.935	M
FATHMM 🔗	-16.13 to 10.64	-6.22, -6.22	<-1.51	D, D
PROVEAN 🔗	-14 to +14	-2.74	<-2.49	D
MetaSVM 🔗 (meta-predictor)	-2 to +3	1.0557	>0	D
MetaLR 🔗 (meta-predictor)	0 to 1	0.9745	>0.5	D
CADD 🔗 (meta-predictor)	-7.535 to 35.789	6.124358	>19 (inferred)	higher score = higher pathogenicity
FATHMM-MKL 🔗	--	0.22862	--	N
fitCons 🔗	0 to 1	0.638212	--	higher score = higher pathogenicity
C Conservation Analysis 👉				
Source	Score			
phyloP100way 🔗	2.434			
phyloP20way 🔗	0.935			
phastCons100way 🔗	0.381			
phastCons20way 🔗	0.997			
GERP++ 🔗	4.92			
SiPhy 🔗	12.3933			
D Splice Site Predictors				
Analyze using MaxEntScan 🔗 Analyze using NNSPLICE 🔗 Analyze using HumanSplicingFinder 🔗				

Table 2. Prediction classification text letter code key.

Predictor	Letter Code	Prediction
SIFT	D	Damaging
	T	Tolerated
PolyPhen2-HDIV PolyPhen2-HVAR	D	probably Damaging
	P	Possibly damaging
	B	Benign
LRT	D	Deleterious
	N	Neutral
	U	Unknown
MutationTaster	A	disease causing Automatic
	D	Disease causing
	P	Polymorphism automatic
	N	polymorphism
MutationAssessor	H	High (predicted functional)
	M	Medium (predicted functional)
	L	Low (predicted non-functional)
	N	Neutral
FATHMM	D	Damaging
	T	Tolerated
PROVEAN	D	Damaging
	N	Neutral
METASVM	D	Damaging
	T	Tolerated
METALR	D	Damaging
	T	Tolerated

Limitations:

- Splicing *in silico* tools can be tricky to utilize and to interpret, and more training is necessary than can be provided in this general SOP. Although the VCI provides links to all of these tools, each VCEP will determine the tools most useful for their group. Each Expert Panel should validate which of these tools are able to predict the native sites for each splice donor/acceptor for their gene, specify which tools will be used by their EP, and provide training to their biocurators. Before evaluating a variant using the provided links, determine the tools to be used and receive instruction in usage, thresholds, etc., with the appropriate VCEP.

Functional, Conservation, and Splicing Predictors Section Criteria Evaluations:

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.)

- To complete assessment of PP3 and BP4, please compare *in silico* predictors, conservation analysis predictors, and/or splice site predictors (not currently available within the VCI) to determine if **overall the predictors suggest a deleterious effect (PP3), no impact (BP4), or if predictions are mixed/unclear (no ACMG criteria selected)**.
 - **It is not necessary for all prediction tools to be concordant to apply PP3 or BP4, but expert judgment or VCEP specifications must be applied and a consistent threshold should be used for all the variants in that gene.**
- In application of PP3/BP4, a meta-predictor such as REVEL may be used in place of multiple predictors in the *in silico* analysis of missense variants.
- As the positive predictive value for a tool may vary by the gene, VCEPs may define thresholds based on a specific prediction tool or set of prediction tools for their gene of interest.
- Again, *in silico* tools to be used should be specified by the VCEP. In particular, splicing results often benefit from discussion with the full EP (or splicing experts within the panel) for proper application, and any abnormal results or questions should result in screenshots of the output and presentation to the EP.

PP2: Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease

As a general recommendation based on ExAC and gnomAD data, the SVI WG suggests a gene with a missense z-score (found on ExAC and gnomAD) $z > 3.09$ is more likely to be intolerant of missense changes. The z-score is calculated by comparing the expected versus observed number of missense variants. The higher the z-score, the higher the constraint or intolerance to variation.

- As this will be gene-specific, each VCEP will decide if PP2 applies for their gene. Please check with your VCEP if appropriate.

BP1: Missense variant in a gene for which primarily truncating variants are known to cause disease

- To complete assessment of PP2 and BP1, additional gene-centric knowledge is required regarding the variant spectrum of the gene/disease of interest. Currently, the only data displayed in the interface to support assessment of these criteria are the LOF ExAC Constraint Scores displayed on the Gene-centric tab (pLI). pLI is the probability of a gene being LOF-intolerant. Those genes with pLI scores of 0.9 and above have the highest likelihood of LOF intolerance in a heterozygous state. This score is most useful for haploinsufficient genes with pediatric onset conditions. It will not identify recessive disease genes that tolerate LOF carriers, or late onset disease genes where the variants are passed on to the next generation. ExAC also computes a missense constraint score (z), available on any gene page in ExAC, but this score is not currently displayed in the

interface. This z-score represents the deviation of the observed from expected variant count. A positive z-score indicates increased intolerance to variation (fewer variants than expected), and a negative z-score indicates variant tolerance (more variants than expected). GnomAD also computes missense and LOF gene constraint metrics; however, these are based on an observed versus expected score which is calculated slightly differently from the ExAC constraint metrics. This ratio is a continuous measure of gene tolerance to different classes of variant (i.e., missense, LOF) and a LOWER score indicates stronger selection (decreased tolerance) for that variant type.

- To thoroughly evaluate these criteria, it is recommended that curators assess the variant spectrum (missense/nonsense, splicing, indels, LOF, copy number variants, etc.) for the gene in ClinVar, The Human Gene Mutation Database (HGMD), ExAC, and gnomAD as described above, and receive guidance from the appropriate VCEP, if available. Please note that HGMD (<http://www.hgmd.cf.ac.uk/ac/index.php>) has two versions - a public version and a licensed version. The public version is less up-to-date, but freely available to users from academic institutions or non-profit organizations upon registration.
- As this will be gene-specific, each VCEP will decide if BP1 applies for their gene. Please check with your VCEP if appropriate.

Other Variants in Same Codon Section

ACMG/AMP criteria codes: PS1, PM5

Background: This section provides an alert to curators regarding additional ClinVar variants in the codon of interest. This indication of additional variants in the codon is **not** dependent on the classification of the additional variants or the variant type. To view the additional variants in ClinVar, click "Search ClinVar for variants in this codon". Please note, if the variant under assessment is already in ClinVar, the ClinVar search result will also return that variant. For example, if assessing variant NM_007294.3(BRCA1):c.5513T>A (p.Val1838Glu), the interface informs the curator that 3 additional variants are found in the Val1838 codon in ClinVar (Figure 12A). The search result screen in ClinVar will display 4 returned variants as the c.5513T>A (p.Val1838Glu) variant is already present in ClinVar (Figure 12B).

Other Variants in the Same Codon Criteria Evaluations:

PS1: Same amino acid change as a previously established pathogenic* variant regardless of nucleotide change

Example: Val to Leu caused by either G>C or G>T in the same codon. Caveat: Beware of changes that impact splicing rather than the predicted amino acid/protein change.

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

Example: Arg156Cys is pathogenic; now you observe Arg156His. Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.

- The alert to additional ClinVar variants in the codon is a helpful tool to identify these variants; however, assessment of the pathogenicity of the additional variants should not solely depend on pathogenic interpretations in ClinVar. Additionally, as ClinVar may not contain all known variants within a codon, curators should also search other variant databases. Curators must thoroughly evaluate the potential pathogenicity of other variants in the codon and not rely on the variant’s database designation, as these may not reflect current variant interpretation practice.
- These criteria are VCEP-specific; please consult the appropriate VCEP if applicable. Some examples of VCEP specifications are describe here:
 - Certain VCEPs do not require the missense change to be novel.
 - Certain VCEPs will consider PS1/PM5 if the previous missense change has a ‘pathogenic’ OR a ‘likely pathogenic’ classification.
 - Certain VCEPs require that an *in silico* tool predicts the missense change being interrogated to be more damaging than the known variant.
 - Certain VCEPs require complete curation for all variants at the relevant codon in order to completely assess whether PS1/PM5 should be applied.

*To meet either of these criteria, the other variant in the codon should reach a pathogenic (or likely pathogenic, if appropriate) classification without relying on PS1 or PM5, in order to avoid circular dependencies.

Figure 12. A, ‘Other Variant in Same Codon’ section of Missense tab, and B, ClinVar search result screen.

A Other Variants in Same Codon

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

PM5: Explanation:

PS1: Same amino acid change as a previously established pathogenic variant regardless of nucleotide change (has caveat) Disease-specific

PS1: Explanation:

ClinVar Variants

Additional ClinVar variants found in the same codon: 3 ([Search ClinVar for variants in this codon](#))

B Search results
Items: 4

	Variation Location	Gene(s)	Condition(s)	Clinical significance (Last reviewed)	Review status
1.	NM_007294.3(BRCA1):c.5513T>G (p.Val1838Gly) GRCh37: Chr17:41197774 GRCh38: Chr17:43045757	BRCA1	Breast-ovarian cancer, familial 1	Likely pathogenic (Jul 1, 2015)	criteria provided, single submitter
2.	NM_007294.3(BRCA1):c.5513T>A (p.Val1838Glu) GRCh37: Chr17:41197774 GRCh38: Chr17:43045757	BRCA1	Breast-ovarian cancer, familial 1, not provided	Pathogenic (Aug 10, 2015)	reviewed by expert panel
3.	NM_007294.3(BRCA1):c.5512G>T (p.Val1838Leu) GRCh37: Chr17:41197775 GRCh38: Chr17:43045758	BRCA1	Hereditary breast and ovarian cancer syndrome, Hereditary cancer-predisposing syndrome	Uncertain significance (Jun 13, 2015)	criteria provided, multiple submitters, no conflicts
4.	NM_007294.3(BRCA1):c.5512G>A (p.Val1838Met) GRCh37: Chr17:41197775 GRCh38: Chr17:43045758	BRCA1	Hereditary breast and ovarian cancer syndrome, not specified	Uncertain significance (Apr 15, 2015)	criteria provided, multiple submitters, no conflicts

Loss of Function (LOF) sub-tab

ACMG/AMP criteria code: PVS1

Background: Nonsense, frameshift, canonical splice site, initiation codon changes, and single or multi-exon deletions are all variant types that often (though not always) result in loss of protein function. The SVI has provided specific guidance on the use of this criterion⁶. A flowchart of these recommendations can be found in Appendix 2. To apply the PVS1 criteria to these types of variants, the variant should be expected to lead to a truncated mRNA that undergoes nonsense-mediated decay, resulting in complete loss of protein translation, and be in a gene associated with a condition where the disease mechanism should be consistent with LOF. The LOF tab in the VCI allows the user to capture any evidence that supports these criteria but does not specifically address how to evaluate them. The following information can be used to aid in determining whether PVS1, (or one of its strength modifications), can be appropriately applied to a suspected null/LOF variant. It is important to keep in mind that use of this criteria is also disease-specific and should be cross-referenced with the Expert Panel for which curation is being done, although the ClinGen reference is the primary guide.

LOF Sub-tab Criteria Evaluations:

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

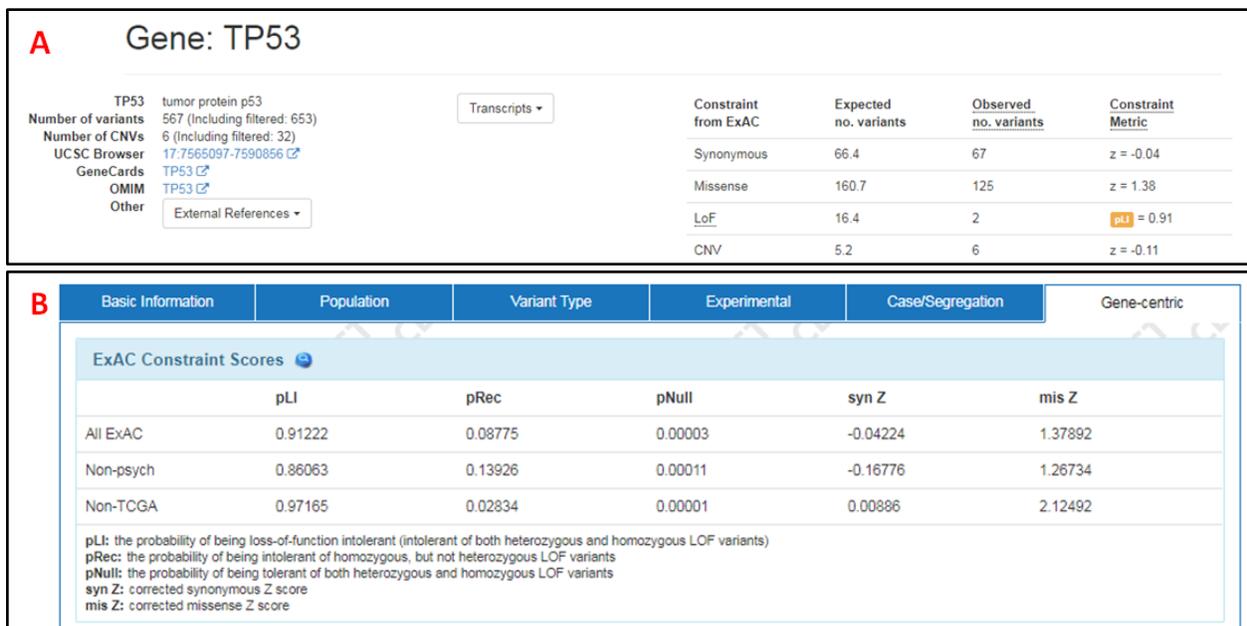
Disease Mechanism

This type of information is not always well established (particularly for recently identified and/or less-well studied disease genes), but in some instances is published in the literature. It is important to critically evaluate evidence supporting published assertions of LOF as the disease mechanism, since this can be assumed without sufficient supporting data in some publications. If strong genetic or functional evidence is not available, it is possible to estimate how well a gene tolerates LOF variants using the “constraint metrics” provided by ExAC (which are included on the gene-centric tab in the VCI) or gnomAD. These metrics provide a probability score (pLI) for a gene’s intolerance of LOF variants (more detail can be found on the ExAC website: <http://exac.broadinstitute.org/faq>)⁷. For dominant diseases caused by LOF variants the

expectation would be that the pLI score should be close to 1 (i.e. the gene is extremely intolerant to LOF variants) and there should be an excess of pathogenic LOF variants relative to other variant types. In gnomAD, an observed and expected variant score (o/e constraint metric) is also used to determine the probability that a given gene is intolerant to LOF variants (<https://gnomad.broadinstitute.org/>)⁸. The o/e metric differs from that of the pLI. It incorporates a 90% confidence interval, and has an opposite scale - the closer the o/e is to zero, the more likely the gene is LOF-constrained. At this time, these metric only includes nonsense and canonical splice site variants in this calculation; they do not include frameshift variants. Additionally, variant databases (internal or public such as ClinVar) can be used to assess the spectrum of reported pathogenic/likely pathogenic variants in a gene associated with the condition being curated. The Gene-centric tab automatically populates the ExAC constraint scores for a gene (Figure 13A and B), but only provides a link-out to view ClinVar variants in that gene. **This method for evaluating disease mechanism should be used with caution**, as it primarily reflects constraint due to reproductive fitness. It is most useful for pediatric disorders with a severe phenotype, and **should not be used for (most) adult onset Mendelian disorders** (e.g., *BRCA1* has a pLI of 0 and an o/e of 0.73). If curating with a VCEP, discuss disease mechanism and whether use of this metric is appropriate, with the Expert Panel.

An additional resource to consider is the ClinGen Dosage Sensitivity Map (<https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/>)⁹. Through this resource, genes and genomic regions are evaluated for evidence of haploinsufficiency (HI) and triplosensitivity (TS). While this resource was originally created to assist with the interpretation of copy number variants, those genes with sufficient evidence of HI (i.e., an HI score of “3” per this evaluation system) are thought to have LOF as a disease mechanism.

Figure 13. Example of ExAC constraint scores for TP53. A, Screenshot of data in ExAC; B, Screenshot of ExAC constraint score on Gene-centric tab.



Assessing Molecular Consequence

Variants that potentially result in LOF should be visualized using a genome browser to determine their location with respect to exon structure and the 3' (C-terminal) end of the gene. The Basic information tab provides a link out to multiple different genome browsers (Figure 14). Only those variants that result in premature termination codons > 50 bp upstream of the last exon-exon junction of the transcript are expected to result in nonsense-mediated decay (NMD) and consequently lead to LOF. Variants not predicted to undergo NMD may still impact a critical domain of the protein product or eliminate a large percentage of the protein (particularly if the protein is small and/or the last exon is very large). It is additionally important to consider which transcripts the variant impacts. In some scenarios, a nonsense variant may occur in only one transcript that is not biologically relevant; therefore, it is critical to determine if the LOF variant will impact a biologically relevant transcript.

For variants that may alter mRNA splicing (i.e. canonical splice site variants, internal exon deletions, initiation codon) it is important to consider what impact the variant will have on the predicted protein product. It is possible that such variants could result in in-frame insertions/deletions creating an intact protein of inappropriate size (shorter or larger), but still retain all necessary protein domains and thus have little impact on the overall protein function. If curating as part of a VCEP, that Expert Panel will provide guidance on how to modify the PVS1 evidence level. Anything that falls outside of the specified rules should be presented to the Expert Panel for discussion and to make a final decision on evidence level.

Figure 14. Link out to genome browsers on Basic Information tab.



Silent & Intron sub-tab

ACMG/AMP criteria code: BP7

Background: The major consideration for synonymous and/or intronic variants is whether they impact mRNA splicing. As previously mentioned, *in silico* splice site predictors can be useful in evaluating whether a variant might impact splicing, but these tools are not currently integrated into the VCI and instead are linked out to the appropriate sites directly within this tab (Figure 11D). This tab also provides the option of including any published functional information regarding the splicing impact of synonymous and/or intronic variants. The BP7 criteria also takes into consideration evolutionary conservation of the nucleotide across species. This information can be evaluated using the conservation track provided by most genome browsers (which are linked out on the Basic Information tab, see Figure 14 above).

Silent & Intron Subtab Criteria Evaluations:

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

- Some VCEPs do not require lack of a predicted splicing impact. Please check with the appropriate VCEP (if applicable).
- Nucleotide conservation comparison also varies by VCEP. Different VCEPs may specify alternate conservation tools, cutoffs to use, or species to evaluate for presence of other nucleotides.

In-frame Indel sub-tab

ACMG/AMP criteria codes: PM4, BP3

Background: An indel refers to a small insertion or deletion variant and is in-frame if the altered nucleotide number is divisible by three, keeping the same reading frame. Variants that extend or shorten a protein by deletion, insertion, or altering the stop codon to another amino acid can disrupt protein function. 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. It is important to understand whether the domain the variant alters is functionally important and conserved. There are examples of severe disorders that can result from the loss of a single amino acid, e.g. Coffin-Siris syndrome. 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 (linked directly within this tab). It can also help to assess population databases, such as ExAC and 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.

In-frame Indel sub-tab criteria evaluations:

PM4: Protein length changes as a result of in-frame deletions/insertions in a non-repeat region or stop-loss variant

- VCEPs may define cutoffs to apply PM4 for variants that cause premature truncation but not nonsense-mediated decay (similar to PVS1_moderate). Please discuss with specific VCEP if appropriate.

BP3: In-frame deletions/insertions in a repetitive region without a known function

- VCEPs may define if this rule applies, and if so, to what specific areas of their gene; please discuss with specific VCEP if appropriate.

Experimental Tab

Hotspot or functional domain section

ACMG/AMP criteria codes: PM1

Background: This tab provides a space to evaluate experimental evidence at the variant level (Figure 15). Mutational hotspots and functional domains are disease-specific and evaluated by the appropriate Expert Panel, which should be consulted for guidance.

Limitations:

- Currently, this tab does not automatically pull in data, and curators must manually add the PMID to curate evidence from published articles.

Hotspot or functional domain section Criteria Evaluations:

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

- Expert Panels determine gene/disease-specific hot spots and functional domains, including specific residues.

Figure 15. Space to evaluate experimental evidence.

The screenshot shows a web interface for evaluating experimental evidence. At the top, a blue header reads "Hotspot or functional domain". Below this, there is a light blue box containing the text: "PM1: Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of enzyme) without benign variation". To the right of this box is a dropdown menu labeled "PM1:" with "Not Evaluated" selected. Further right is a text area labeled "Explanation:". A blue "Save" button is positioned to the right of the explanation area. Below these elements is a section titled "Curated Literature Evidence (Hotspot or functional domain)". Inside this section, there is a blue "Add PMID" button followed by the instruction: "Select 'Add PMID' to curate and save a piece of evidence from a published article." At the bottom of this section, it says "No evidence added."

Experimental Studies
^

BS3: Well established *in vitro* or *in vivo* functional studies show no damaging effect on protein function or splicing
Disease-specific

PS3: Well established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product
Disease-specific

BS3:

- or -

PS3:

Explanation:

Curated Literature Evidence (Experimental Studies)

Select "Add PMID" to curate and save a piece of evidence from a published article.

No evidence added.

Experimental Studies Section

ACMG/AMP Criteria Codes: PS3, BS3

Background: Functional studies cover a broad range of experimentation. Only validated and approved assays should be evaluated for classification evidence. Appropriate assays are disease-specific and identified by the Expert Panel. The definition of a ‘validated’ assay is currently being addressed, and accepted functional assays must meet a high level of rigor and reproducibility. However, curators may still document all functional data in the free text ‘evidence’ box, and pull in the appropriate PMID of the article, regardless of whether the evidence code is applied, so that the data is captured for potential utilization at a later date. Curators are encouraged to present results from functional studies to Expert Panel members to decide whether PS3/BS3 applies, particularly if a study has not previously been reviewed by the Expert Panel or if results from functional studies appear discrepant.

Experimental studies section Criteria Evaluations:

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

PS3: Well established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product

- In the “Evidence” text box, curators should enter at least the assay and output under evaluation.
- Functional assays should be confirmed by expert review if they are curator identified.
- Functional assays that may not be explicitly specified by VCEPs may still meet criteria for decreased strength modifications for PS3 (i.e. PS3_Moderate or PS3_Supporting). These should be documented and shared with the VCEP before analysis.

General considerations:

- The existence of a functional assay is not necessarily sufficient for PS3 or BS3 criteria:
 - Exemplary functional assays should demonstrate reproducibility and experimental rigor, including benchmarking with variants of definitive clinical significance as determined by other evidence.
 - The assay should be relevant to disease mechanism and manifestation.
 - Curated evidence should demonstrate variant-level effect on the gene or gene product.
- Applying BS3 criteria should be done with caution, as negative results (equivalent to wild-type protein function) may support a benign classification but should be contingent on how thoroughly protein function is assessed. Multiple assays comprehensively probing gene function will more accurately represent clinical significance. VCEPs should have guidelines in place for biocurators to follow regarding the types of studies and results that may be used for BS3 to be applied; discuss with the appropriate VCEP as applicable.

Case/Segregation Tab

Assessing the de novo occurrence, frequency and inheritance of a variant is useful in identifying potential pathogenicity. This tab contains several sections broken down into population data, allelic evidence and co-segregation data criteria.

Observed in healthy adults section

ACMG/AMP criteria: BS2

Background: This criterion is disease-specific and most applicable if the condition in question is fully penetrant at an early age. Allele frequency can be found in publicly available databases such as gnomAD, or through a thorough literature search. In disease-specific instances, occurrence thresholds may be added; VCEPs may have customized this criteria in alternative ways so should be consulted (if applicable).

Observed in healthy adults criteria evaluations:

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

Case-control section*

ACMG/AMP criteria: PS4

Background: PS4 can be used for typical case-control studies when a relative risk or odds ratio (OR) > 5.0, and confidence interval not including 1.0, has been calculated to assess whether a variant is likely to be associated with a particular phenotype. In instances of very rare variants where case-control studies are not available, the prior observation of the variant in multiple unrelated patients with the same phenotype (and its absence in population databases) may be used as evidence. The number of observations in probands can be used to determine the application of PS4 or any of its modified strengths (see Table 3 for ClinGen VCEP examples). Number of probands is gathered through internal lab data, publicly available data and a thorough literature search.

*Although this section has been titled 'Case-control' in the VCI, due to specifications by many VCEPs this rule is more often used for proband counting.

Limitations:

- Access to internal data can impact rule application and/or strength.
- The VCI does not currently support the import of spreadsheets containing de-identified internal data such as proband count. Each entry must be logged independently into the "Explanation" free text box; proband count for strength level must be calculated by hand.
- Caution should be exercised to not duplicate counting of one proband reported multiple times in the literature or within both the literature and a clinical testing cohort.

Case control criteria evaluations:

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

Table 3. Examples of PS4 specification from different ClinGen VCEPs.

	PS4_Supporting	PS4_Moderate	PS4 (Strong)	PM2 threshold
MYH7 Hypertrophic cardiomyopathy	≥2 probands with consistent phenotypes	≥6 probands with consistent phenotypes	≥15 probands with consistent phenotypes	<0.004%
RASopathy	≥1 probands that meets RASopathy criteria*	≥3 probands that meet RASopathy criteria	≥5 probands that meet RASopathy criteria	Strictly absent
PTEN PTEN Hamartoma Tumor syndrome (PHTS)	>1 proband meets Cleveland Clinic criteria (30 score)**	>2 probands meets Cleveland Clinic criteria (30 score)	>4 probands meets Cleveland Clinic criteria (30 score)	<0.001%
CDH1 Hereditary diffuse gastric cancer (HDGC)	>1 HDGC family meets criteria***	>2 HDGC family meets criteria	>4 HDGC family meets criteria	<0.001%

For a more detailed explanation, see:

*https://www.clinicalgenome.org/site/assets/files/8852/clingen_rasopathy_acmg_specifications_v1-1.pdf

**<https://www.clinicalgenome.org/working-groups/clinical-domain/hereditary-cancer-clinical-domain-working-group/pten-variant-curation-expert-panel/announcements/summary-of-acmg-amp-classification-rules-specified-for-pten-variant-curation/>

***<https://www.clinicalgenome.org/working-groups/clinical-domain/hereditary-cancer-clinical-domain-working-group/cdh1-variant-curation-expert-panel/>

Segregation data section

ACMG/AMP criteria: BS4, PP1

Background: Co-segregation data can be used as evidence for or against variant pathogenicity. The more frequently a variant co-segregates with a disease phenotype, the more likely the variant and the disease locus are linked. This can be determined using a LOD (log of the odds) score, which is a statistical estimate of the proximity of two loci and the likelihood that they will be

inherited together. The higher the LOD score, the more likely the loci are linked. For dominant disorders, co-segregation of a variant in affected family members is counted as supportive pathogenic evidence. Conversely, when a variant is not found in an affected family member, this is considered good evidence that the variant is benign. In either case, several issues must be thoroughly investigated. Other variants in linkage disequilibrium with the linked variant must be considered when applying PP1, and thorough phenotypic evaluation of affected non-segregating family members must be done for BS4 to rule out phenocopies. Correct application of these criteria are hampered by issues of penetrance, expressivity, and age of onset of the disorder. Different conditions apply if diagnostic clarity or phenocopy are a concern.

Segregation data criteria evaluations:

PP1: Co-segregation with disease in multiple affected family members in a gene definitively known to cause disease. VCEPs may specify higher levels of evidence to be applied for higher numbers of meioses/families demonstrating segregation with disease (see Table 4 for VCEP examples).

- Please note that only affected individuals are counted as a part of this criterion. Unaffected individuals may complicate the picture due to incomplete or age/gender-related penetrance, variable expressivity, or phenocopies.

BS4: Lack of segregation in affected members of family

- Please note, BS4 non-segregation is meant for affected family members who do not carry the variant of interest (phenotype positive; genotype negative). Individuals who carry the variant but do not have the phenotype (genotype positive; phenotype negative) may be counted under BS2, if applicable for the gene/disease of interest with agreement from the Expert Panel.

Table 4. Segregation counts from approved VCEPs.

VCEP	PP1 segregation counts
MYH7	PP1_Strong: variant segregate with ≥ 7 meioses PP1_Moderate: variant segregates with ≥ 5 meioses PP1: variant segregates with ≥ 3 meioses
RASopathy	Segregation in more than one family is recommended PP1_Strong: ≥ 7 informative meioses PP1_Moderate: ≥ 5 informative meioses PP1: ≥ 3 informative meioses
PTEN	PP1_Strong: variant segregates with ≥ 7 meioses observed across ≥ 2 families PP1_Moderate: variant segregates with 5-6 meioses across ≥ 1 family PP1: variant segregates with 3-4 meioses across ≥ 1 family
CDH1	PP1_Strong: variant segregates with ≥ 7 meioses across ≥ 2 families PP1_Moderate: variant segregates with 5-6 meioses across ≥ 1 family PP1: variant segregates with 3-4 meioses across ≥ 1 family

Hearing Loss	<p>PP1_Strong: Segregation in 3 affected relatives for recessive and 5 affected relatives for dominant</p> <p>PP1_Moderate: Segregation in 2 affected relatives for recessive and 4 affected relatives for dominant</p> <p>PP1: Segregation in 1 affected relative for recessive and 2 affected relatives for dominant</p>
---------------------	--

De novo occurrence section

ACMG/AMP criteria: PS2, PM6

Background: There are different levels of strength given to *de novo* data. To be considered as strong evidence, parents must be tested and both maternity and paternity must be confirmed (using techniques such as identity panels or NGS trio analysis), or there must be multiple observations of *de novo* occurrence, the variant must be in a gene that is associated with a condition consistent with the phenotype of the patient, and there must be no family history of the disease. The evidence is considered moderate if the second two criteria are met but maternity and paternity have both not been confirmed via identity/parentage testing. Exome trio testing may be accepted as capable of confirming maternity and paternity.

Segregation data criteria evaluations:

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

- See Table 5.

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

- See Table 5.

Using parental confirmation, phenotypic consistency, and number of *de novo* observations, the SVI has proposed a point-based system for modified strength levels of these criteria (https://www.clinicalgenome.org/site/assets/files/8490/recommendation_PS2_and_PM6_ACMG_AMP_criteria_version_1_0.pdf), as shown below in Table 5.

The SVI also provides additional considerations for applying *de novo* criteria based on inheritance:

- X-linked conditions: if an X-linked variant occurs *de novo* in an unaffected carrier mother, and family history is consistent - i.e. she has no affected brothers/other male relatives apart from her affected son(s) – *de novo* criteria may be applied despite the fact that she is unaffected.
- Autosomal recessive conditions: for a *de novo* occurrence in a gene associated with an autosomal recessive condition without an additional pathogenic/likely pathogenic variant identified, the strength of evidence should be decreased by one level.

- Mosaicism: for cases with apparent germline mosaicism (multiple affected siblings with both parents negative for the variant), paternity/maternity must be confirmed in order for *de novo* criteria to apply.

Table 5. Points awarded per *de novo* occurrence

Phenotypic consistency	Points per Proband	
	Confirmed <i>de novo</i>	Assumed <i>de novo</i>
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

*Maximum allowable value of 1 may contribute to overall score

Recommendation for determining the appropriate ACMG/AMP evidence strength level for *de novo* occurrence(s)

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

Limitations:

- The *de novo* guidelines PM6 and PS2 are mutually exclusive; only one of the two criteria is allowed to be met, and the higher strength prevails.
- Phenotype must be consistent with the gene. Further phenotypic specifications may be applied by an Expert Panel for criteria application.

Allele data (cis/trans) section

ACMG/AMP criteria: BP2, PM3

Background: Phase of the variant is important in assessing its effect. When two variants are found on a single chromosome in the same gene, they are considered to be in *cis*; variants identified in a gene but on different chromosomes are in *trans*. Inheritance of a condition is key for applying these rules. For autosomal recessive conditions, PM3 is a method of counting probands (Table 6). Use of these rules for autosomal dominant disorders is dependent upon the

effect of homozygosity on the phenotype. For example, two pathogenic variants in *trans* for an AD disorder could result in embryonic lethality or a severe disease phenotype, in which case BP2 would be applicable (the variant would have to be benign by definition). This is a gene-dependent (and therefore EP-level) decision.

Allele data criteria evaluations:

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

- For recessive disorders, the strength of PM3 may be altered under certain conditions, as recommended by the SVI (see Table 6).

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

- In order to apply BP2, the pathogenic variant must have been observed independently of the variant being assessed. The criteria should not be applied if there is evidence of a synergistic effect.

Table 6. SVI Recommended Scoring, Allele Data.

Classification/zygosity of other variant	Points per proband	
	Known in trans	Phase unknown
Pathogenic/Likely pathogenic	1.0	0.5
Homozygous occurrence (Max points from homozygotes = 1.0)	0.5	N/A
Rare uncertain significance variant on other allele, OR Homozygous occurrence due to consanguinity, (Max point = 0.5)	0.25	N/A

PM3 Point Table			
PM3_Supporting	PM3	PM3_Strong	PM3_VeryStrong
0.5 points	1.0 points	2.0 points	4.0 points
Examples: - 1 observation w/ P/LP variant but phase unknown - 1 homozygous	Examples: - 1 observation w/ P/LP variant confirmed in <i>trans</i> - 2 homozygous		

occurrence - 2 observations of either homozygous occurrence due to consanguinity OR rare VUS	occurrences - 2 observations w/ P/LP variant but phase unknown		
--	---	--	--

Alternate mechanism for disease section

ACMG/AMP criteria: BP5

Background: If a variant is found in an affected individual with an obvious alternate cause of disease, this is supportive of a benign classification; however, there are exceptions.

Alternate mechanism for disease criteria evaluations:

BP5: Variant found in a case with an alternate molecular basis for disease

- Stronger for supporting benign classification in a gene for a dominant disorder as compared to a gene for a recessive disorder. Caution should be used when applying for a recessive disorder as the individual may be a carrier. For example, multiple genes can cause nonsyndromic hearing loss and a diagnosed individual may be a carrier for a pathogenic variant in one gene but harbor two pathogenic variants in *trans* in another gene causative for their disease.
- In some disorders, having multiple variants can contribute to more severe disease. If a novel variant is seen in a patient with a dominant disorder and a family history of a more mild presentation, this would not be considered supportive of a benign classification for the novel variant.
- In conditions where multigenic inheritance is known to occur, additional variants at a second locus may also be pathogenic. Some EPs may qualify the necessity of multiple observations or specify phenotypic considerations in order for this rule to apply.

Specificity of phenotype section

ACMG/AMP criteria: PP4

Background: This criterion is only appropriate for application when a phenotype is highly specific to a single gene. For example, all known cases of Pompe disease are caused by variants in the *GAA* gene, have a specific phenotype, and have a deficiency of activity of the *GAA* gene product, acid alpha glucosidase. Family history should be consistent with mode of inheritance, testing should be highly sensitive, and the gene should have little benign variation. The PP4 criterion may also be applied if non-genetic confirmatory assays are available, such as enzyme levels for biochemical disorders. For example, the Phenylalanine Hydroxylase Variant Curation Expert Panel (PAH VCEP) applies PP4 if a case has diagnostic plasma phenylalanine levels >120 umol/L. However, this criterion should not be modified in weight due to multiple observations of

the variant. This criterion may not be used for commonly encountered, non-specific phenotypes such as developmental delay or cardiomyopathy. These disorders have multiple genetic causes, and phenotype cannot be distinguished by the gene in which the pathogenic variant is located. Of note, several Expert Panels have wrapped PP4 into other evidence types (PS4, PS2/PM6, PP1) as opposed to having a separate evidence code.

Specificity of phenotype criteria evaluations:

PP4: Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

Reputable source section

ACMG/AMP criteria: BP6, PP5

ClinGen has determined that these rules should not be applied in any context. While they are part of the original 2015 ACMG/AMP Variant Classification guidelines, the SVI is of the opinion that evidence collected for variant interpretation should be limited to primary data, which is becoming more widely available through resources such as ClinVar, and that BP6 and PP5 are commonly misused ¹⁰. These criteria have been disabled in the VCI.

Gene-centric Tab

The gene-centric tab contains general information about the gene for use with the various guidelines. This includes the ExAC constraint scores as described in the LOF section, an 'other ClinVar variants in same gene' section with a link-out to ClinVar, and multiple link-outs to gene and protein resources (see Figure 16).

Figure 16. Gene-centric tab in VCI.

ExAC Constraint Scores					
	pLI	pRec	pNull	syn Z	mis Z
All ExAC	0.91222	0.08775	0.00003	-0.04224	1.37892
Non-psych	0.86063	0.13926	0.00011	-0.16776	1.26734
Non-TCGA	0.97165	0.02834	0.00001	0.00886	2.12492

pLI: the probability of being loss-of-function intolerant (intolerant of both heterozygous and homozygous LOF variants)
pRec: the probability of being intolerant of homozygous, but not heterozygous LOF variants
pNull: the probability of being tolerant of both heterozygous and homozygous LOF variants
syn Z: corrected synonymous Z score
mis Z: corrected missense Z score

Other ClinVar Variants in Same Gene

[Search ClinVar for variants in this gene](#)

Gene Resources

HGNC Symbol: [TP53](#)
Approved Name: tumor protein p53
Synonyms: p53, LFS1
Entrez Gene: [7157](#)
Ensembl: [ENSG00000141510](#)
GeneCards: [TP53](#)

Protein Resources

UniProtKB: [P04637](#)
Domains: [InterPro](#)
Structure: [PDBe](#)
Gene Ontology (Function/Process/Cellular Component): [AmiGO2](#) | [QuickGO](#)

Evidence Summary

Once all evidence has been curated into the VCI, a calculated classification and evaluation summary will be displayed. This calculated classification is derived from combining the applied ACMG/AMP guidelines (see key found in Appendix III), some of which may have been modified for some VCEPs during their specification processes. The curator has the option of saving this interpretation as provisional. The curator can also change the classification at this point, provided that he/she describes the reason for the change. The curator will be responsible for providing an evidence summary that contains all evidence used to arrive at the classification, including any PMIDs and de-identified internal data that has been used, as well as all criteria that met an evaluation strength. Of note, some Expert Panels have drafted template sentences to guide curators in standardizing evidence descriptions, as this summary will be publicly available. After review with the VCEP, the final approval for the classification can be documented by saving the

evaluation summary as 'Approved'. This summary can then be entered into ClinVar with a three-star review status via submission by the appropriate ClinGen VCEP.

List of Resources

References

1. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. *Genet Med*. 2015 May; 17(5):405-24.
2. Harrison SM, Riggs ER, Maglott DR, Lee JM, Azzariti DR, Niehaus A, Ramos EM, Martin CL, Landrum MJ, Rehm HL. Using ClinVar as a Resource to Support Variant Interpretation. *Curr Protoc Hum Genet*. 2016 Apr 1; 89:8.
3. DiStefano MT, Hemphill SE, Cushman BJ, Bowser MJ, Hynes E, Grant AR, Siegert RK, Oza AM, Gonzalez MA, Amr SS, Rehm HL, Abou Tayoun AN. Curating Clinically Relevant Transcripts for the Interpretation of Sequence Variants. *J Mol Diagn*. 2018 Nov; 20(6):789-801.
4. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG; ClinGen Sequence Variant Interpretation Working Group. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat*. 2018 Nov; 39(11):1525-1530.
5. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*. 2016 Oct 6; 99(4):877-885.
6. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, Harrison SM; ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018 Nov; 39(11):1517-1524.
7. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016 Aug 18; 536(7616):285-91.
8. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman

- E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, The Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. **doi:** <https://doi.org/10.1101/531210>.
9. Riggs ER, Church DM, Hanson K, Horner VL, Kaminsky EB, Kuhn RM, Wain KE, Williams ES, Aradhya S, Kearney HM, Ledbetter DH, South ST, Thorland EC, Martin CL. Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin Genet*. 2012 May; 81(5):403-12.
 10. Biesecker LG, Harrison SM. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med*. 2018 Mar 15.

ClinGen Resources

ClinGen website	https://www.clinicalgenome.org/
SVI webpage	https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation
CDWG webpage	https://www.clinicalgenome.org/working-groups/clinical-domain/
Variant Curation Interface	https://curation.clinicalgenome.org/
VCI Help document	https://github.com/ClinGen/clincoded/wiki/VCI-Curation-Help
ClinGen Allele Registry	http://reg.clinicalgenome.org/redmine/projects/registry/genboree_registry/landing
ClinGen Evidence Repository	https://erepo.clinicalgenome.org/evrepo/

Other useful resources

ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
OMIM	https://www.omim.org
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php
Decipher	https://decipher.sanger.ac.uk/
The Monarch Initiative	https://monarchinitiative.org/
Genetic Testing Registry	https://www.ncbi.nlm.nih.gov/gtr/

Sequence databases

RefSeq	https://www.ncbi.nlm.nih.gov/refseq/
Ensembl	http://uswest.ensembl.org/index.html
UCSC genome browser	https://genome.ucsc.edu/cgi-bin/hgGateway
Variation Viewer	https://www.ncbi.nlm.nih.gov/variation/view/overview/

Population databases

Genome Aggregation Database (gnomAD)	http://gnomad.broadinstitute.org/
Exome Aggregation Consortium (ExAC)	http://exac.broadinstitute.org/
Population Architecture using Genomics and Epidemiology (PAGE)	https://www.genome.gov/27541456/population-architecture-using-genomics-and-epidemiology/
1000 Genomes (1000G)	http://grch37.ensembl.org/index.html
Exome Sequencing Project (ESP)	http://evs.gs.washington.edu/EVS/
dbSNP	https://www.ncbi.nlm.nih.gov/projects/SNP/

Literature sources

PubMed	https://www.ncbi.nlm.nih.gov/pubmed/
Google Scholar	https://scholar.google.com/
bioRxiv	https://www.biorxiv.org/content/early/2017/06/12/148353
LitVar	https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/LitVar/

Appendix I - Computational Tools

In silico predictors linked in VCI

In silico tools are used to predict the effect of a variant on a particular gene. The scores for some of these tools are calculated directly in the VCI, and ranges, as well as prediction meaning are described. For further information, descriptions, and tutorials, please see each individual site.

REVEL	https://sites.google.com/site/revelgenomics/about
SIFT	http://sift.bii.a-star.edu.sg/
PolyPhen2-HDIV/HVAR	http://genetics.bwh.harvard.edu/pph2/
LRT	http://www.genetics.wustl.edu/jflab/lrt_query.html
MutationTaster	http://mutationtaster.org
MutationAssessor	http://mutationassessor.org/r3/
FATHMM	http://fathmm.biocompute.org.uk/
PROVEAN	http://provean.jcvi.org/index.php
MetaSVM MetaLR	https://sites.google.com/site/ipopgen/dbNSFP
CADD	https://cadd.gs.washington.edu/
FATHMM-MKL	http://fathmm.biocompute.org.uk/fathmmMKL.htm
fitCons	http://compgen.cshl.edu/fitCons/

Conservation analysis tools linked in VCI

phyloP100way phyloP20way	http://compgen.cshl.edu/phast/index.php
phastCons100way phastCons20way	http://compgen.cshl.edu/phast/index.php
GERP++	http://mendel.stanford.edu/SidowLab/downloads/gerp/
SiPhy	http://portals.broadinstitute.org/genome_bio/siphy/index.html

Splice site predictors linked in VCI

MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
------------	---

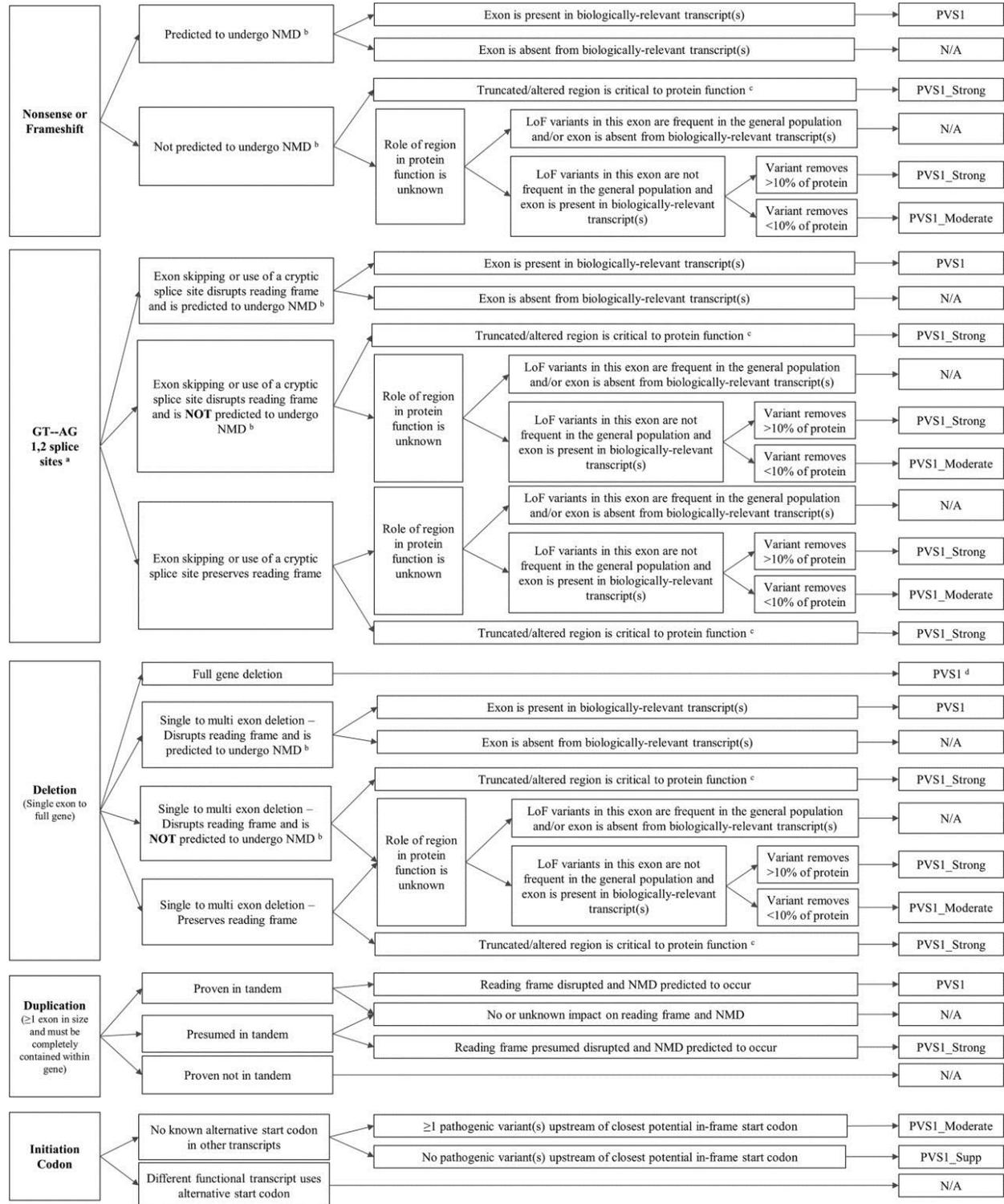
NNSPLICE

http://www.fruitfly.org/seq_tools/splice.html

HumanSplicingFinder

<http://www.umd.be/HSF3/HSF.shtml>

Appendix II - PVS1 flowchart, SVI recommendations



For full article, see PMID: 30192042

Appendix III - Combining Criteria to Classify Sequence Variants

Pathogenic

1. 1 Very Strong AND:
 - a. ≥ 1 strong, OR
 - b. ≥ 2 moderate, OR
 - c. 1 moderate and 1 supporting, OR
 - d. ≥ 2 supporting
2. ≥ 2 strong
3. 1 strong AND:
 - a. ≥ 3 moderate, OR
 - b. 2 moderate AND ≥ 2 supporting, OR
 - c. 1 moderate AND ≥ 4 supporting

Likely Pathogenic

1. 1 very strong AND 1 moderate, OR
2. 1 very strong AND 1-2 moderate, OR
3. 1 strong AND ≥ 2 supporting, OR
4. ≥ 3 moderate, OR
5. 2 moderate AND ≥ 2 supporting, OR
6. 1 moderate AND ≥ 4 supporting

Benign

1. 1 stand alone, OR
2. ≥ 2 strong

Likely Benign

1. 1 strong AND 1 supporting, OR
2. ≥ 2 supporting

Uncertain Significance

1. Other criteria shown above are not met, OR
2. The criteria for benign and pathogenic are contradictory

For full article, see PMID: 25741868