TABLE OF CONTENTS

1 Introduction

The American College of Medical Genetics and Genomics (ACMG) developed guidance for the interpretation of sequence variants. This guideline describes the standards and guidelines for classification of sequence variants using criteria informed by expert opinion and empirical data from the ACMG, the Association for Molecular Pathology (AMP) and the College of American Pathologists (CAP). In 2015, ACMG and AMP published the updated standards and guidelines for the clinical interpretation of sequence variants with respect to human diseases, based on 28 criteria. (doi: 10.1038/gim.2015.30)

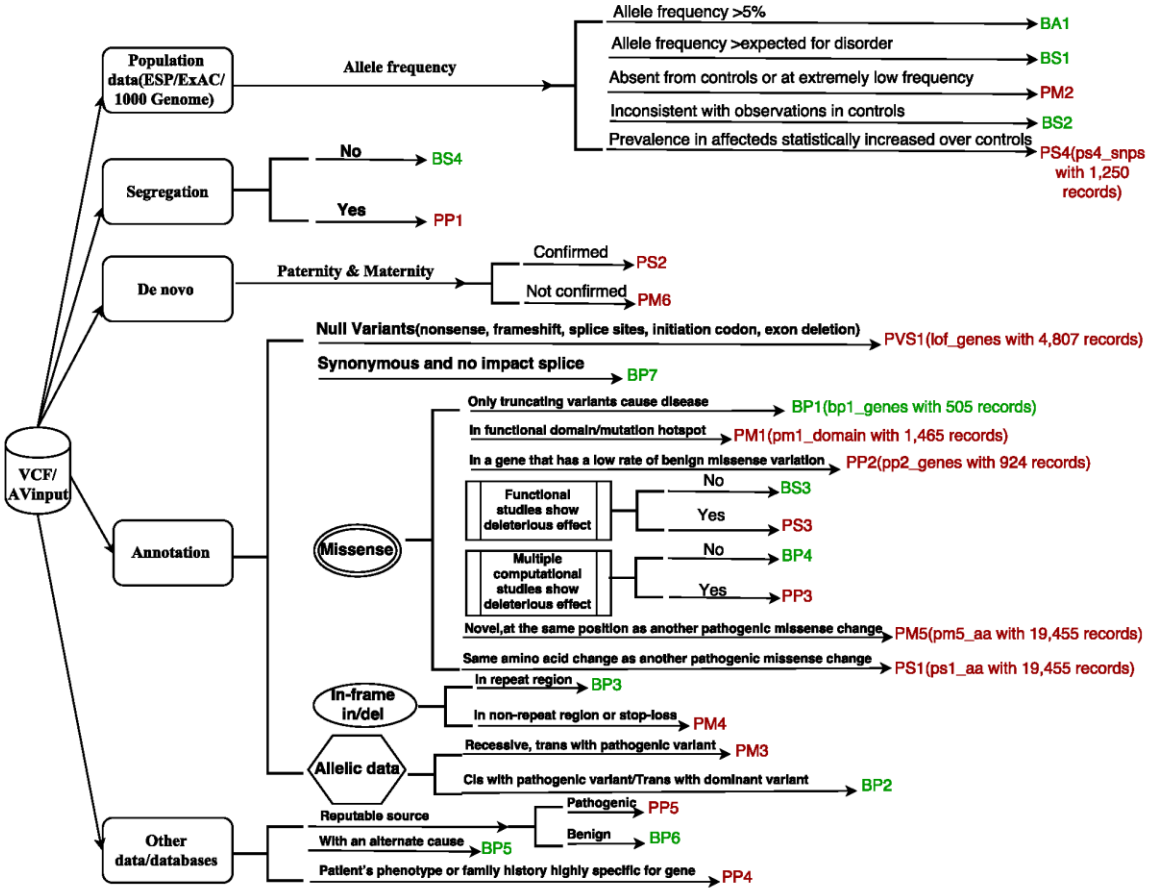


Figure 1. An illustration of the 28 criteria from the ACMG/AMP2015 guidelines.

InterVar is a bioinformatics software tool for clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline. The input to InterVar is an annotated file generated from ANNOVAR, while the output of InterVar is the classification of variants into 'Benign', 'Likely benign', 'Uncertain significance', 'Likely pathogenic' and 'Pathogenic', together with detailed evidence code. InterVar can take a pre-annotated file or a VCF file as input and generate automated interpretation on 18 criteria. Furthermore, the group developed a companion web server wInterVar to enable user-friendly variants interpretation with an automated interpretation step and a manual adjustment step.

2 How to install

InterVar is a python script for variant interpretation of clinical significance. To download, please visit https://github.com/WGLab/InterVar/releases or

Usage:

```
wget https://github.com/WGLab/InterVar/archive/master.zip -O InterVar.zip
unzip InterVar.zip
cd InterVar-master/
ls -alrt .
python Intervar.py
```

Prior to using InterVar, you need to make sure you have installed Python >=2.6.6, ANNOVAR version >= 2016-02-01, download other files such as mim2gene.txt from OMIM (should be generated: >= 2016-09). Several third-party researchers have provided additional annotation program and datasets that can be used by InterVar directly. However, users need to agree to specific license terms set forth by the third parties:

- ANNOVAR main package: The latest version of ANNOVAR (2016Feb01) can be downloaded from ANNOVAR website (openbioinformatics.org) (registration required). ANNOVAR is written in Perl and can be run as a standalone application on diverse hardware systems where standard Perl modules are installed.
- OMIM dataset: Please download the mim2gene.txt from OMIM at OMIM Download Request. Please use the updated files from OMIM, outdated files will bring problems of InterVar.

After installing Annovar and OMIM dataset, copy three ANNOVAR perl files: annotate_variation.pl table_annovar.pl convert2annovar.pl to InterVar's folder of InterVar-master. Also please copy the mim2gene.txt to intervardb folder of InterVar-master/intervardb.

Usage:

```
cp -f annotate_variation.pl table_annovar.pl convert2annovar.pl ~/InterVar-master/
cp -f mim2gene.txt ~/InterVar-master/intervardb/
```

Lastly, navigate to folder of InterVar-master, check and edit the config.ini. The names and locations in config.ini should match with your downloaded files:

- mim2gene = %(database_intervar)s/mim2gene.txt (This line is for location of the OMIM's mim2gene file)
- convert2annovar = ./convert2annovar.pl (This line is for location of ANNOVAR's convert2annovar.pl file)
- table_annovar = ./table_annovar.pl (This line is for location of ANNOVAR's table_annovar.pl file)
- annotate_variation = ./annotate_variation.pl (This line is for location of ANNOVAR's table_annovar.pl file)

For detail explanation, please refer to https://github.com/WGLab/InterVar/blob/master/docs/user-guide/manual.md .

3 How to use InterVar

3.1 Prepare input data

The required input for InterVar is a simple tab-delimited file including a list of variants that are already annotated with a set of required information, such as amino acid changes and allele frequency. Users can generate this input file themselves by using an in-house variant analysis workflow. Alternatively, InterVar can take pre-annotated VCF file from ANNOVAR software

Usage:

```
perl table_annovar.pl input.vcf humandb/ -buildver hg19 -remove -out output -protocol refGene,esp6500siv2_all,1000g2015aug_all,avsnp144,dbnsfp30a,clinvar_20160302,exac03,dbscsnv11,dbnsfp31a_interpro,rmsk,ensGene,knownGene -operation g,f,f,f,f,f,f,f,r,g,g -nastring . -vcfinput
```

Abbreviations after option “Protocol” are representing the databases you wish to annotate to your VCF. It can be included/excluded based on your needs.

- “esp6500siv2_all” is a database for allele frequency in the NHLBI Exome Sequencing Project (ESP6500)
- “refGene” is a database for gene annotation from RefSeq
- “1000 g2015aug_all” is a database for alternative allele frequency (AAF) in the 1000 Genomes Project27 (version August 2015)
- “exac03” is a database for AAF in the Exome Aggregation Consortium (ExAC) Browser28 (version 0.3)
- “dbnsfp30a” is a database for various functional deleteriousness prediction scores from dbNSFP29, 30 (version 3.0a)
- “clinvar_20160302” is for the variants reported in ClinVar20 (version 20160302)
- “avsnp144” is for the ANNOVAR-compiled dbSNP (version 144)
- “ensGene” is for gene annotation from Ensembl
- “knownGene” is for gene annotation from UCSC Known Genes

- “dbnsfp31a_interpro” is a database of the domain information from dbNSFP29, 30 and InterPro31 (which integrates information about protein families, domains, and functional sites)
- “dbscsnv11” is a database for predicting the splicing impact by Ada Boost and Random Forest,32
- “rmsk” is a database on the repeat masking track from the UCSC Genome Browser.

For more details, please refer to Annotation Tools Recommendation-Annovar or https://annovar.openbioinformatics.org/en/latest/

3.2 Analysis

Please be advice that for the first running, the InterVar will use the perl ./annotate_variation.pl to download the necessary ANNOVAR datasests, it will take some time. From second running, InterVar will not download the same ANNOVAR's datasets again.

Usage：

```
python Intervar.py -i example/ex1.avinput -o example/myanno
```

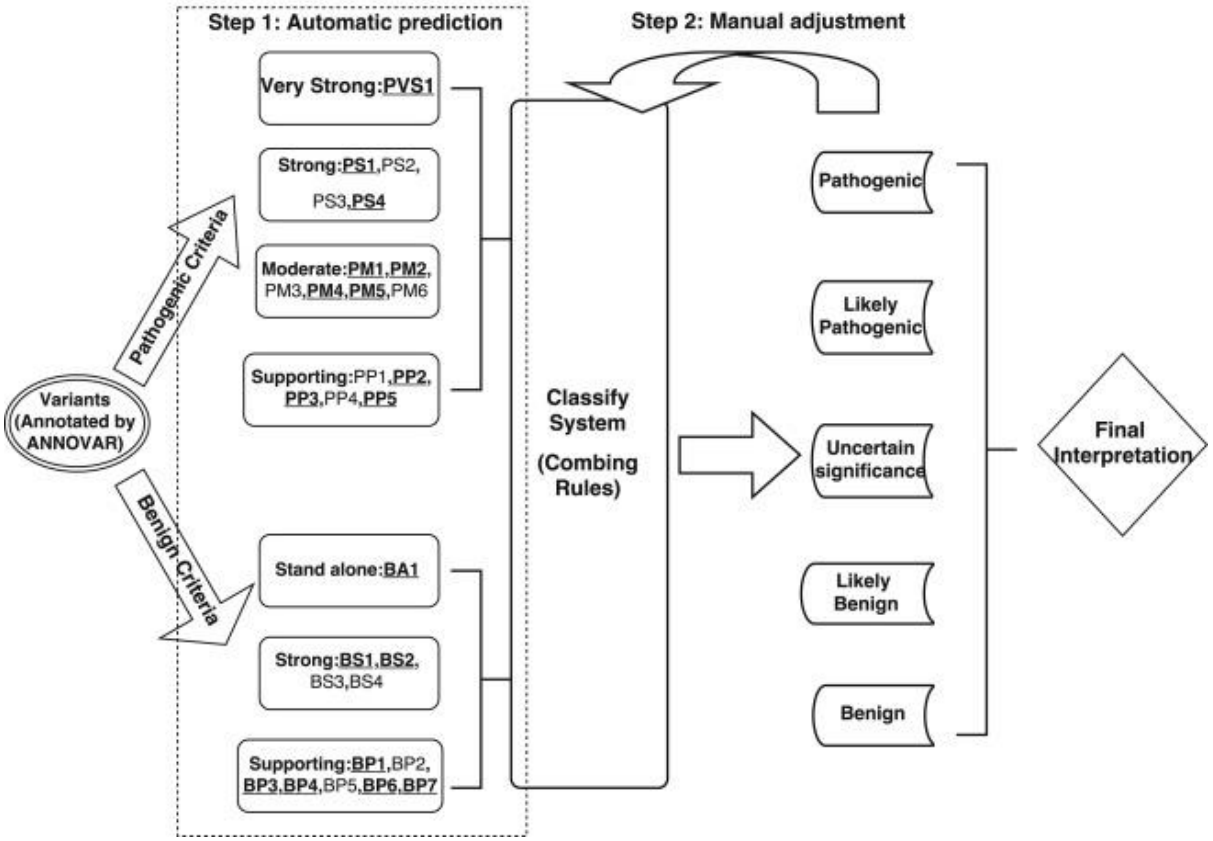
InterVar mainly consists of two major steps: (1) automated scoring on each of the 18 pieces of criteria and (2) manual review and adjustment on specific criteria to arrive at a final interpretation. By running the command, InterVar calls an annotation software, such as ANNOVAR, to obtain necessary annotation information on variants and then uses its own internal annotation database to supplement additional annotations. Using these annotations on variants and genes, InterVar performs a preliminary interpretation of the variant and presents all relevant evidence for manual review.

3.3 Output file interpretation (First step)

The output of InterVar is the classification of variants into 'Benign', 'Likely benign', 'Uncertain significance', 'Likely pathogenic' and 'Pathogenic', together with detailed evidence code. The result file is named as example/myanno.hg19_multianno.txt.intervar. It is tab-delimited, you can import this file into Excel. The column of "InterVar: InterVar and Evidence" give the InterVar interpretation result with all the criteria.

Based on the 2015 ACMG-AMP guidelines, the criteria fall into two sets: pathogenic or likely pathogenic (P/LP) and benign or likely benign (B/LB), whereas “uncertain significance” is assigned to variants for which the criteria for P/LP and B/LB are contradictory or not met. There are a total of 28 criteria:

- 16 criteria for the P/LP criterion are very strong (PVS1), strong (PS1–PS4), moderate (PM1–PM6), or supporting (PP1–PP5)
- 12 criteria for the B/LB criterion are stand-alone (BA1), strong (BS1–BS4), or supporting (BP1–BP7).



If a criterion is positive, InterVar will assign 1; otherwise, InterVar will assign 0. For these 28 criteria, InterVar can automatically generate predictions on 18 (PVS1, PS1, PS4, PM1, PM2, PM4, PM5, PP2, PP3, PP5, BA1, BS1, BS2, BP1, BP3, BP4, BP6, and BP7) according to the current annotation datasets (annotation from Annovar). For more information on the automated scoring system, please refer to http://dx.doi.org/10.1016/j.ajhg.2017.01.004.

While the rest (PS2, PS3, PM3, PM6, PP1, PP4, BS3, BS4, BP2, and BP5) require user input in the manual adjustment step.

3.4 Manual review and adjustment (Second step)

Taking below variant as example. This variant is annotated by InterVar as Uncertain significance with criteria of BS1=1, all other critera is 0;

```
1       67705958        67705958        G       A       IL23R   exonic  nonsynonymous SNV
InterVar: Uncertain significance PVS1=0 PS=[0, 0, 0, 0, 0] PM=[0, 0, 0, 0, 0, 0, 0] PP=[0,
0, 0, 0, 0, 0] BA1=0 BS=[1, 0, 0, 0, 0] BP=[0, 0, 0, 0, 0, 0, 0, 0]
```

After manual review, you want to add two criteria as PS3=1 for "Well-established in vitro or in vivo functional studies supportive of a damaging effect " and PM6=1 for "Assumed de novo, but without confirmation of paternity and maternity".

To do this, please vim a tab-delimited text file with name as "evdience.txt" and rerun InterVar annotation. Format of the evidence.txt should be as follow. If you want to add more than one criteria, it should be semicolon-delimited in the evidence_list column.

Chr Position Ref_allele Alt_allele evidence_list

Usage:

```
~/InterVar-master> cat evdience.txt
1       67705958        G       A       PS3=1;PM6=1


python Intervar.py -i example/ex1.avinput -o example/myanno --
evidence_file=evdience.txt
```

The annotation results change to Likely pathogenic, and you can find that PS3 change to 1 and PM6 change to 1. One strong pathogenic criterion (PS3) and one moderate pathogenic criteria (PM6) will make the clinical interpretation as Likely pathogenic.

Output:

```
1       67705958        67705958        G       A       IL23R   exonic  nonsynonymous SNV
InterVar: Likely pathogenic PVS1=0 PS=[0, 0, 1, 0, 0] PM=[0, 0, 0, 0, 0, 1, 0] PP=[0, 0, 0,
0, 0, 0] BA1=0 BS=[1, 0, 0, 0, 0] BP=[0, 0, 0, 0, 0, 0, 0, 0]
```

In the evidence list, you can also state the strength of the variant. For example, you want to increase the PM6 from moderate to strong.

```
grade_PXX=2;grade_BXX=3 1 for Strong; 2 for Moderate; 3 for Supporting
```

Usage:

```
~/InterVar-master> cat evdience.txt
1       67705958        G       A       PS3=1;PM6=1;grade_PM6=1;


python Intervar.py -i example/ex1.avinput -o example/myanno --
evidence_file=evdience.txt
```

Output:

```
1       67705958        67705958        G       A       IL23R   exonic  nonsynonymous SNV
InterVar: Pathogenic PVS1=0 PS=[0, 0, 1, 0, 0] PM=[0, 0, 0, 0, 0, 1, 0] PP=[0, 0, 0, 0, 0,
0] BA1=0 BS=[1, 0, 0, 0, 0] BP=[0, 0, 0, 0, 0, 0, 0, 0]
```

You can find that now the interpretation changed as Pathogenic. Two strong pathogenic criteria will make clinical significance to pathogenic. For criteria description, please refer to Figure 1 at section 1.0 and doi: 10.1038/gim.2015.30 .

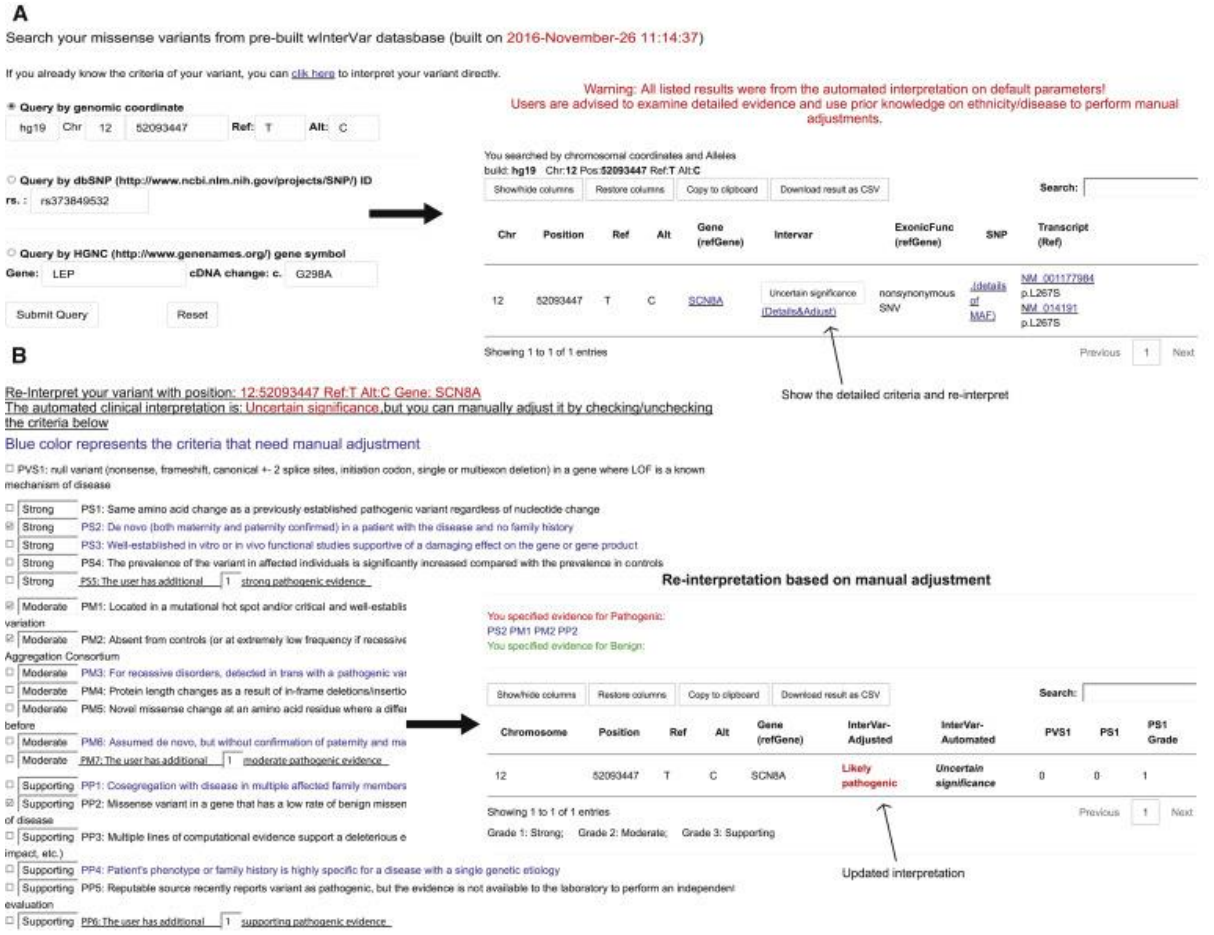
4 wInterVAR

wInterVar (https://wintervar.wglab.org/) is a web implementation of InterVar so that users can use an online web server to perform interpretation on individual variants without using command-line

tools. Unfortunately, the wInterVar server only takes variants one by one. You will need to download InterVar and run the command line tools to analyze many variants, such as an entire exome.

The wInterVar server has two steps for assessing and adjusting the clinical significance of variants:

1. Users first input a variant to obtain pre-computed, automated interpretation.
2. After reviewing the results of automated interpretation, users can then click the “adjust” button to perform the manual adjustment step by selecting and de-selecting appropriate criteria according to additional information and domain knowledge. The wInterVar server will then perform the final interpretation with the two-step procedure.



-End-