

# VIP-HL: Semi-automated ACMG/AMP variant interpretation platform for genetic hearing loss

Jiguang Peng<sup>1</sup>, Jiale Xiang<sup>1</sup>, Xiangqian Jin<sup>1</sup>, Junhua Meng<sup>1</sup>, Nana Song<sup>1</sup>, Lisha Chen<sup>1</sup>, Ahmad Abou Tayoun<sup>2</sup>, and Zhiyu Peng<sup>1</sup>

<sup>1</sup>BGI Genomics

<sup>2</sup>Al Jalila Children’s Specialty Hospital

December 1, 2020

## Abstract

The American College of Medical Genetics and Genomics, and the Association for Molecular Pathology (ACMG/AMP) have proposed a set of evidence-based guidelines to support sequence variant interpretation. The ClinGen hearing loss expert panel (HL-EP) introduced further specifications into the ACMG/AMP framework for genetic hearing loss. This study developed a tool named VIP-HL, aiming to semi-automate the HL ACMG/AMP rules. VIP-HL aggregates information from external databases to automate 13 out of 24 ACMG/AMP rules specified by HL-EP, namely PVS1, PS1, PM1, PM2, PM4, PM5, PP3, BA1, BS1, BS2, BP3, BP4, and BP7. We benchmarked VIP-HL using 50 variants where 83 rules were activated by the ClinGen HL-EP. VIP-HL concordantly activated 96% (80/83) rules, significantly higher than that of by InterVar (47%; 39/83). Of 4948 ClinVar star 2+ variants from 142 deafness-related genes, VIP-HL achieved an overall variant interpretation concordance in 88.0% (4353/4948). VIP-HL is an integrated online tool for reliable automated variant classification in hearing loss genes. It assists curators in variant interpretation and provides a platform for users to share classifications with each other. VIP-HL is available with a user-friendly web interface at <http://hearing.genetics.bgi.com/>.

## Introduction

Hearing loss is a primary global health concern, affecting about 6.8% of the world’s population (Wilson, Tucci, Merson, & O’Donoghue, 2017). Genetic factors account for at least 50 to 60% of childhood hearing loss (Morton & Nance, 2006). To date, over 150 deafness-related genes were discovered (Azaiez et al., 2018). The advancement of next-generation sequencing technology and its continually decreasing cost have facilitated the genetic diagnosis of hearing loss (Abou Tayoun et al., 2016). One of the remaining significant challenges, however, is the accurate interpretation of a large number of identified variants.

In 2015, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) published a joint guideline to standardize variant interpretation (Richards et al., 2015). The guideline offered a common framework for curators to resolve variant classification differences, and to improve interpretation consistency across laboratories (Amendola et al., 2016). To evolve the guideline over time, National Institutes of Health-funded Clinical Genome Resource (ClinGen) developed groups of disease experts to refine recommendations for different disease areas, including genetic hearing loss (Oza et al., 2018), *MYH7* -associated inherited cardiomyopathies (Kelly et al., 2018), germline *CDH1* (Lee et al., 2018), and *PTEN* -associated hereditary cancer (Mester et al., 2018).

Using bioinformatics tools to facilitate and standardize variant interpretation is one of top priorities, and has been shown to be useful for curators (Kopanos et al., 2019). InterVar is one such tool semi-automating 18 criteria in the ACMG/AMP guidelines (Li & Wang, 2017). However, generic tools may not fulfill the need because many interpreting standards are disease-specific and vary dramatically amongst different diseases

(Kanavy et al., 2019). In 2018, two tools named CardioClassifier (Whiffin et al., 2018) and CardioVAI (Nicora et al., 2018) were explicitly developed for interpreting variants in cardiovascular diseases. In 2019, a semi-automated tool called “Variant Interpretation for Cancer” (VIC) was developed to accelerate the interpretation process in Cancer (He et al., 2019).

Here, we developed a **V**ariant **I**nterpretation **P**latform for genetic **H**earing **L**oss (VIP-HL). This new online tool utilizes the framework outlined by the ClinGen Hearing Loss Expert Panel (HL-EP) (Oza et al., 2018) to automatically annotate variants in 142 hearing loss related genes across 13 ACMG/AMP rules. VIP-HL is freely accessible for non-commercial users in a web server at <http://hearing.genetics.bgi.com/>.

## Methods

VIP-HL automates 13 out of 24 ACMG/AMP rules specified by ClinGen HL-EP, namely PVS1, PS1, PM1, PM2, PM4, PM5, PP3, BA1, BS1, BS2, BP3, BP4, and BP7 (Oza et al. 2018). The development and optimization of VIP-HL are described in three sections: (1) rule selection, optimization and implementation; (2) Benchmarking and comparative analysis; (3) web interface development.

### Rule selection, optimization and implementation

Twenty-four ACMG/AMP rules for genetic hearing loss were grouped into four sections: population (PM2, BA1, BS1, BS2), computational (PVS1, PS1, PM1, PM4, PM5, PP3, BP3, BP4, and BP7), case/segregation (PM3, PS2/PM6, PS4, PP1, PP4, BS4, BP2, and BP5) and functional (BS3 and PS3) data. VIP-HL automates all 13 population and computational rules, while case/segregation and functional criteria (n=11) still require manual curation.

For population rules (PM2, BA1, and BS1), our implementation of population frequency cutoffs relied on gnomAD v2.1 exomes and genomes combined dataset (Karczewski et al., 2020). The cutoffs followed the guideline for genetic hearing loss (Oza et al., 2018). The highest filtering allele frequency across all gnomAD populations (“popmax”) was applied as the allele frequency for each variant (Whiffin et al., 2017). To determine the BS2 criterion, we retrieved the homozygote number from the gnomAD control dataset. BS2 was applied when the homozygote number was greater than three for autosomal recessive disease, and one for autosomal dominant disease. It should be noted that ClinGen HL-EP did not determine the cutoffs of the homozygous number for BS2. If users disagree the default setting, it is adjustable in VIP-HL.

The parameterization of nine computational criteria (PVS1, PS1, PM1, PM4, PM5, PP3, BP3, BP4, and BP7) was described as follows. The specifications of PVS1 were clarified before (Abou Tayoun et al., 2018; Xiang, Peng, Baxter, & Peng, 2020). PM1 can be applied when a variant is in a mutational hotspot region or well-studied functional domain without benign variation (Oza et al., 2018; Richards et al., 2015). The hotspot region was determined based on the enrichment of pathogenic variants, as clarified in our recent work (Xiang, Peng, et al., 2020).

Pathogenic/likely pathogenic variants from ClinVar 20200629 release were retrieved to determine the existence of the same amino acid changes (PS1) and the different amino acid changes (PM5). For example, knowing that NM\_004004.6(*GJB2*):c.109G>A (p.Val37Ile) is a well-established pathogenic variant in ClinVar, we can now apply PM5 for NM\_004004.6(*GJB2*):c.109G>T (p.Val37Phe).

PM4 was applied when the in-frame deletion or insertion was not in the repetitive region which is evolutionary well conserved (Oza et al. 2018). The repetitive region was determined based on RepeatMasker (Chen, 2004) and Tandem repeats finder (Benson, 1999). The region with GERP scores higher than two was considered as evolutionarily conserved (Cooper et al., 2005). BP3 was applied while the in-frame indels were in repeat region without known function.

In silico tools selected for implementing PP3, BP4 and BP7 were REVEL (Ioannidis et al., 2016) and MaxEntScan (Yeo & Burge, 2004). PP3 was applied with a REVEL score of  $\geq 0.7$  and BP4 was applied with a REVEL score  $\geq 0.15$  (Oza et al., 2018). PP3 can also be applied when non-canonical splice variants were predicted to have an impact on splicing via MaxEntScan (Yeo & Burge, 2004). BP7 was employed

when a synonymous variant was predicted with no impact on splicing via MaxEntScan and the nucleotide is not highly conserved ( $\text{GERP} < 2$ ) (Davydov et al., 2010).

The final pathogenicity is reported in five tiers system proposed by the ACMG/AMP guideline, namely "Pathogenic (P)", "Likely Pathogenic (LP)", "Variant of Uncertain Significance (VUS)", "Likely Benign (LB)", and "Benign (B)" (Richards et al., 2015).

## Benchmarking and comparative analysis

In order to test VIP-HL performance, we constructed two benchmark datasets. The first dataset consisted of 50 out of 51 variants in the hearing loss gene that has been curated by the ClinGen HL-EP (Oza et al., 2018). We excluded NM\_206933.3:c.(?\_12295)\_(14133\_?)del in the *USH2A* gene because it is an exon-level deletion (Exons 63-64 deletion), which is currently not compatible with VIP-HL.

To assess the importance of disease-specific annotations, we compared activated rules by ClinGen HL-EP with those activated by VIP-HL and InterVar, and vice versa. Comparing rules that were not activated by ClinGen HL-EP, but activated by either VIP-HL or InterVar, we did not count the variants meriting the BA1 criterion because ClinGen HL-EP did not activate other criteria once a variant met the BA1 criterion. For example, ClinGen HL-EP assigned BA1 for NM\_005422.2:c.1111A>G in the *TECTA* gene and did not further activate other criteria. However, both VIP-HL and InterVar activated BS2 because 10518 homozygotes are reported in the gnomAD database for this variant (Karczewski et al., 2020). The InterVar code was downloaded from GitHub. All the settings were set as default.

The second dataset included 4948 variants in 142 deafness-related genes with ClinVar star 2+ (i.e., multiple submitters with assertion criteria, expert panel or practice guideline) (Landrum et al., 2018). These variants were selected because they had fewer misclassifications (Shah et al., 2018; Xiang, Yang, et al., 2020). The 142 deafness-related genes were curated by ClinGen HL-EP (DiStefano et al., 2019). The gene list and their gene-disease associations are listed in Supplementary Table 1.

## Web interface development

The front end of VIP-HL is written in HTML5 and JavaScript (VUE), with the back end implemented in JAVA and Python 3. Uploaded variants are annotated via Ensembl Variant Effect Predictor (McLaren et al., 2016). A total of 142 hearing loss genes as mentioned above were collected from the ClinGen website (DiStefano et al. 2019). We pre-annotated all variants in these genes from ClinVar 20200629 release.

## Results

### Benchmark analysis

Adopting ACMG/AMP guidelines for genetic hearing loss, 152 ACMG/AMP rules were activated for 50 pilot variants curated by the ClinGen HL-EP. Of these, 55% (83/152) rules could be automatically interpreted by VIP-HL, spanning PVS1, PM1, PM2, PM4, PM5, PP3, BA1, BS1, BP4, and BP7. Overall, VIP-HL achieved 96% (80/83) concordant interpretations compared with activations by ClinGen HL-EP (Figure 1a). Of 83 activated rules (Figure 1b), the three discordant activations between VIP-HL and ClinGen HL-EP were PM2\_Supporting (2 times) and BA1 (1 time). Those discrepancies were most likely due to using an updated version of gnomAD and the adoption of popmax filtering allele frequency (Whiffin et al. 2017) by VIP-HL (Table 1, variant #1-3). By comparison, InterVar only achieved 47% (39/83) concordant activations, significantly lower than that by VIP-HL (Figure 1b).

Both VIP-HL and InterVar activated certain rules that ClinGen HL-EP did not apply. Specifically, VIP-HL activated only four rules six times, including PM1 (2 times), PP3 (2 times), PM2 (1 time), PVS1 (1 time), BS2 (2 times) (Table 1, variant #4-8). InterVar, however, activated 11 rules 100 times (Supplementary Figure S1), spanning PVS1 (2 times), PM1 (15 times), PM2 (20 times), PP2 (2 times), PP3 (12 times), PP5 (25 times), BS1 (2 times), BS2 (8 times), BP1 (11 times), BP4 (1 times), BP6 (2 times). Notably, four rules (PP2, PP5, BP1, and BP6) that were considered not applicable for genetic hearing loss accounted for 40%

of the activations (40/100). These results demonstrated the reliable annotations by VIP-HL and reinforced the importance of disease-specific annotations in variant interpretation.

### Semi-automation of ClinVar variants

Of 4948 ClinVar variants in 142 deafness-related genes, VIP-HL achieved an overall variant interpretation concordance of 88.0% (4353/4948). The concordant interpretations were 57.1% (376/658) in P/LP variants, 93.6% (3083/3295) in B/LB variants, and 89.8% (894/995) in VUS variants, respectively (Table 2). The lower concordance in P/LP variants could be explained that VIP-HL is a semi-automated tool and only automated 13 out of 24 ACMG/AMP rules.

Of note, VIP-HL generated three P/LP classifications versus B/LB compared to ClinVar. The first one was NM\_153676.3(*USH1C*):c.2547-1G>T. It was submitted as likely benign in ClinVar, whereas VIP-HL assigned PVS1 and PM2, leading to a likely pathogenic classification. Manual curation revealed that this splicing variant affected an exon in transcript NM\_153676.3 with no detectable expression based on the Genotype-Tissue Expression (GTEx) database (GTEx Consortium, 2017). Thus, this variant should be classified as benign/likely benign. This example highlights the need for including information about the most biologically relevant transcripts in VIP-HL for accurate clinical variant interpretation (DiStefano et al., 2018).

The other two discrepant variants were synonymous (NM\_206933.3(*USH2A*):c.949C>A (p.Arg317=) and NM\_022124.6(*CDH23*):c.7362G>A (p.Thr2454=). Both were assigned likely benign classifications by VIP-HL due to the activation of BP4 and BP7. Specifically, NM\_206933.3:c.949C>A was submitted as a pathogenic variant in ClinVar. This variant led to abnormal splicing and a premature termination codon (Vaché et al., 2010), strongly supporting its pathogenic classification. As shown below, our VIP-HL user interface will enable curators to manually activate codes for functional studies to avoid such possible misclassifications. The second variant, NM\_022124.6(*CDH23*):c.7362G>A, was submitted as likely pathogenic. No supporting RNA/functional, case-level or segregation studies were provided to support this classification. Since the actual effect of this sequence change is still unknown, a likely benign classification may be more appropriate for this case.

Of 4948 ClinVar variants, the most utilized rules were population-related, spanning PM2, BA1, BS1, BS2 and their modified strength levels (supporting through very strong; Figure 2). BP4 and BP7 are also frequently used because 33.5% (1657/4948) deafness-related variants with ClinVar star 2+ were synonymous variants. As expected, the pathogenic criteria were enriched in pathogenic/likely pathogenic variants, whereas benign criteria were enriched in benign/likely benign variants (Figure 2).

### Features of VIP-HL web interface

Users can search the interface (<http://hearing.genetics.bgi.com/>) by gene name, chromosomal location, dbSNP identifier, and HGVS nomenclature (den Dunnen et al., 2016). If a gene is queried, the results will present the gene name, cytobands, protein name, external databases and the variant list. The variant list consists all the pre-annotated variants displayed with their pathogenicity and corresponding ACMG/AMP rules. For variants not pre-annotated, VIP-HL can execute the algorithm and provide the interpreting results promptly. Once variants that are not pre-annotated are queried, they will be automatically added to the variant list. If a specific variant is queried, VIP-HL displays the variant location (variant identifier), gene name, inheritance, cHGVS, pHGVS, ACMG/AMP criteria, and classifications (Figure 3).

The browser displays six sections, including necessary information, population, computation, case/segregation, function, and community sections (Figure 3). The automatically annotated ACMG/AMP rules are presented in the population and computation sections, providing detailed explanations of the ratings based on the guidelines of genetic hearing loss (Oza et al., 2018). Rules that require manual curation are presented in the case/segregation and function sections. The community section records the submissions by users who would like to share their classifications with others.

Users are free to adjust the evidence strength from “Supporting” to “VeryStrong” in 24 ACMG/AMP

rules. The classifications will update instantly with any rule(s) modifications. For rules grouped in the case/segregation and function sections, evidence to support the strength level can be manually added, enabling users to record and manage their curations.

## Discussion

Considering the substantial differences amongst diseases in terms of inheritance pattern, disease mechanism, phenotype, genetic and allelic heterogeneity, and prevalence, disease-specific guidelines are necessary for accurate and reliable interpretations (Rehm, Berg, & Plon, 2018). Following variant interpretation guidelines for genetic hearing loss (Oza et al., 2018), we developed a new computational tool, named VIP-HL, publicly available through a web interface (<http://hearing.genetics.bgi.com/>). To our knowledge, this is the first tool designed for automated variant interpretation in genetic hearing loss. Considering the high prevalence of hearing loss in the population, the availability of VIP-HL will significantly relieve the interpretation burdens for clinicians and curators.

Compared to rules activated by ClinGen HL-EP, VIP-HL showed a markedly high concordance (96%), indicating the reliability of interpreting hearing loss variants via VIP-HL. Of note, all the three discrepant activations (variants #1-3, Table 1) were attributable to population-based rules (BA1 and PM2), which depends on the adoption of popmax filtering allele frequency in extensive population studies (Whiffin et al., 2017). The ClinGen HL-EP used the ExAC database in the time of their research whereas we employed a larger dataset (gnomAD) as it was encouraged by the ClinGen HL-EP (Oza et al., 2018). Using these stringent allele frequencies empowers clinical genome interpretation without the removal of true pathogenic variants (Whiffin et al., 2019).

VIP-HL activated several rules that were not activated by ClinGen HL-EP, including PM1 and BS2. ClinGen HL-EP did not perform a systematic review of mutational hot spots or functional domains for all genes associated with hearing loss, and proposed that PM1 can be applied for *KCNQ4* pore-forming region (Oza et al., 2018). In this study, we used the enrichment of pathogenic/likely pathogenic variants to construct a set of important regions (Xiang, Peng, et al., 2020) which includes the *KCNQ4* pore-forming region. Additionally, although HL-EP did not elaborate on the cutoff for BS2, we used a conservative cutoff to automate this rule. It should be noted that the penetrance affects the application of BS2 but was not considered by VIP-HL. This led to activations of BS2 for NM\_004004.6:c.109G>A and NM\_004004.6:c.101T>C in the *GJB2* gene because 50 and 16 homozygotes were identified from the gnomAD control dataset, respectively. The two variants were well-known pathogenic variants with low penetrance (Shen et al., 2019). Nevertheless, VIP-HL is a semi-automatic tool and our user interface enables curators to manually adjust codes to avoid such possible misclassifications.

A further comparison between VIP-HL and ClinVar showed an overall interpretation concordance of 88.0%. In terms of pathogenic/likely pathogenic variants, the concordance was lower (57.1%). This could be explained that VIP-HL only automated 13 out of 24 ACMG/AMP rules. The nine case-level and segregation evidence and two functional evidence required manual curation from scientific literature. Of them, pathogenic rules are more frequently activated than benign rules (Oza et al., 2018). Prospectively, text-mining and machine learning techniques might serve as potential solutions. For example, Birgmeier and co-authors developed an end-to-end machine learning tool, named AVADA, for the automatic retrieval of variant evidence directly from full-text literature (Birgmeier et al., 2020). Suppose we can accumulate enormous datasets of evidence-related sentences or figures, in that case, it is possible to apply machine-learning approaches in the future for evidence retrieval and to automate the remaining ACMG/AMP rules in the next version of VIP-HL. In the meantime, our interface enables curators to manually activate the relevant codes after manual literature curation.

VIP-HL generated three P/LP classifications versus B/LB compared to ClinVar. All the three variants were related to the consideration of splicing impact. This discrepancy of NM\_153676.3:c.2547-1G>T was attributable to a lack of considerations of exon expression data, which ultimately led to inappropriate classifications. It is apparent that a splicing variant affecting a non-expressive exon should have less functional effects (DiS-

tefano et al., 2018). Recently, the transcript-level information from the GTEx project (Consortium, 2017) was utilized and proved that incorporating exon expression data can improve interpretations of putative loss-of-function variants (Cummings et al., 2020). The second and third variants (NM\_206933.3:c.949C>A and NM\_022124.6:c.7362G>A) were synonymous variants, and their splicing impact should be curated from public literature if available. Nevertheless, these results indicated the importance of expression data in variant interpretation.

To improve user experience and further facilitate variation interpretation via VIP-HL, we developed a user-friendly web interface, which we continue to grow and add useful features over time. For example, PM3, one of the most frequently activated rules in genetic hearing loss (Oza et al., 2018), relies on the variant's pathogenicity on the second allele. If this latter variant is introduced (in HGVS nomenclature) during the curation of PM3, VIP-HL can now provide the pathogenicity of this second variant as a reference for users. We expect such features and ongoing improvements would save curators the time and relieve the burden of variant interpretation.

VIP-HL has limitations. First, it is currently not applicable for exon-level copy number variations. Second, the allele frequency cutoffs were different for dominant and recessive hearing loss disorders. We first applied the cutoffs from the inheritance curated by ClinGen HL-EP for variants in a gene with both dominant and recessive inheritance. If both were available, we conservatively chose the cutoffs in recessive disorders. To avoid users falling into this pitfall, we highlighted the selected inheritance in the web interface of VIP-HL.

In conclusion, VIP-HL is an integrated online tool and search engine for variants in genetic hearing loss genes. It is also the first tool, to our knowledge, to consider the specifications proposed by ClinGen HL-EP for genetic hearing loss related variants. Providing reliable and reproducible annotations, VIP-HL not only facilitates variant interpretation but also provides a platform for users to share classifications with others.

**Data Availability Statement:** Data sharing is not applicable to this article as no new data were created or analyzed in this study.

**Conflict of Interest:** Jiguang Peng, Jiale Xiang, Xiangqian Jin, Junhua Meng Lisha Chen, Nana Song, and Zhiyu Peng were employed at BGI Genomics at the time of submission. No other conflicts relevant to this study should be reported.

## Figure legend

**Figure 1. Validation of VIP-HL.** (a) Comparing VIP-HL to a set of 50 hearing loss expert panel curated variant. (b) Counts of individual rules activated by ClinGen hearing loss expert panel, VIP-HL and InterVar for 50 pilot variants. Rules applied with a modified strength are denoted by the rule followed by \_P for Supporting.

**Figure 2. Frequency of rules applied for 4948 ClinVar variants .** Rules applied with a modified strength are denoted by the rule followed by \_P for Supporting, \_M for Moderate, \_S for Strong, and \_VS for Very Strong.

**Figure 3. VIP-HL web-interface workflow.** A variant could be searched by HGVS nomenclature, chromosomal location, or dbSNP identifier. Once a variant is queried, results are showed in three parts. The top-left corner displays variant information; the top-right corner displays ACMG/AMP criteria and final classification; the lower part displays six sections, including basic information, population, computation, case/segregation, function, and community. In each section, the strength of each criterion is adjustable based on users' evaluations.

## References

Abou Tayoun, A. N., Al Turki, S. H., Oza, A. M., Bowser, M. J., Hernandez, A. L., Funke, B. H., . . . Amr, S. S. (2016). Improving hearing loss gene testing: a systematic review of gene evidence toward more efficient next-generation sequencing-based diagnostic testing and interpretation. *Genetics in Medicine*, 18 (6), 545-553. doi:10.1038/gim.2015.141

- Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., . . . ClinGen Sequence Variant Interpretation Working, G. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation*, *39* (11), 1517-1524. doi:10.1002/humu.23626
- Amendola, L. M., Jarvik, G. P., Leo, M. C., McLaughlin, H. M., Akkari, Y., Amaral, M. D., . . . Rehm, H. L. (2016). Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *American Journal of Human Genetics*, *98* (6), 1067-1076. doi:10.1016/j.ajhg.2016.03.024
- Azaiez, H., Booth, K. T., Ephraim, S. S., Crone, B., Black-Ziegelbein, E. A., Marini, R. J., . . . Smith, R. J. H. (2018). Genomic Landscape and Mutational Signatures of Deafness-Associated Genes. *American Journal of Human Genetics*, *103* (4), 484-497. doi:10.1016/j.ajhg.2018.08.006
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27* (2), 573-580. doi:10.1093/nar/27.2.573
- Birgmeier, J., Deisseroth, C. A., Hayward, L. E., Galhardo, L. M. T., Tierno, A. P., Jagadeesh, K. A., . . . Bejerano, G. (2020). AVADA: toward automated pathogenic variant evidence retrieval directly from the full-text literature. *Genetics in Medicine*, *22* (2), 362-370. doi:10.1038/s41436-019-0643-6
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*, *Chapter 4* , Unit 4 10. doi:10.1002/0471250953.bi0410s05
- Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C. S., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15* (7), 901-913. doi:10.1101/gr.3577405
- Cummings, B. B., Karczewski, K. J., Kosmicki, J. A., Seaby, E. G., Watts, N. A., Singer-Berk, M., . . . MacArthur, D. G. (2020). Transcript expression-aware annotation improves rare variant interpretation. *Nature*, *581* (7809), 452-458. doi:10.1038/s41586-020-2329-2
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*, *6* (12), e1001025. doi:10.1371/journal.pcbi.1001025
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., . . . Taschner, P. E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, *37* (6), 564-569. doi:10.1002/humu.22981
- DiStefano, M. T., Hemphill, S. E., Cushman, B. J., Bowser, M. J., Hynes, E., Grant, A. R., . . . Abou Tayoun, A. N. (2018). Curating Clinically Relevant Transcripts for the Interpretation of Sequence Variants. *Journal of Molecular Diagnostics*, *20* (6), 789-801. doi:10.1016/j.jmoldx.2018.06.005
- DiStefano, M. T., Hemphill, S. E., Oza, A. M., Siegert, R. K., Grant, A. R., Hughes, M. Y., . . . ClinGen Hearing Loss Clinical Domain Working, G. (2019). ClinGen expert clinical validity curation of 164 hearing loss gene-disease pairs. *Genetics in Medicine*, *21* (10), 2239-2247. doi:10.1038/s41436-019-0487-0
- GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550* (7675), 204-213. doi:10.1038/nature24277
- He, M. M., Li, Q., Yan, M., Cao, H., Hu, Y., He, K. Y., . . . Wang, K. (2019). Variant Interpretation for Cancer (VIC): a computational tool for assessing clinical impacts of somatic variants. *Genome Medicine*, *11* (1), 53. doi:10.1186/s13073-019-0664-4
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*, *99* (4), 877-885. doi:10.1016/j.ajhg.2016.08.016

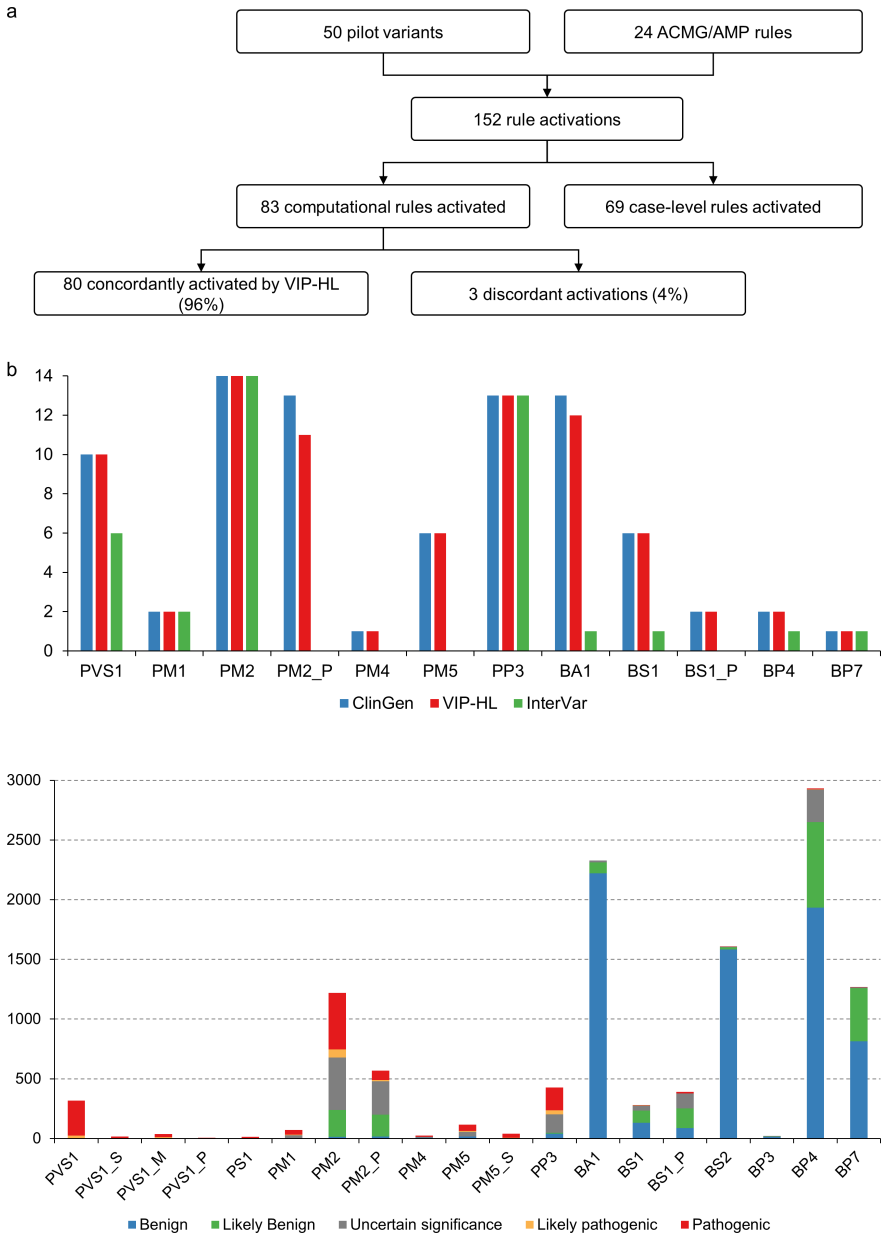
- Kanavy, D. M., McNulty, S. M., Jairath, M. K., Brnich, S. E., Bizon, C., Powell, B. C., & Berg, J. S. (2019). Comparative analysis of functional assay evidence use by ClinGen Variant Curation Expert Panels. *Genome Medicine*, 11 (1), 77. doi:10.1186/s13073-019-0683-1
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581 (7809), 434-443. doi:10.1038/s41586-020-2308-7
- Kelly, M. A., Caleshu, C., Morales, A., Buchan, J., Wolf, Z., Harrison, S. M., . . . Funke, B. (2018). Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genetics in Medicine*, 20 (3), 351-359. doi:10.1038/gim.2017.218
- Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics*, 35 (11), 1978-1980. doi:10.1093/bioinformatics/bty897
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46 (D1), D1062-D1067. doi:10.1093/nar/gkx1153
- Lee, K., Krempely, K., Roberts, M. E., Anderson, M. J., Carneiro, F., Chao, E., . . . Karam, R. (2018). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Human Mutation*, 39 (11), 1553-1568. doi:10.1002/humu.23650
- Li, Q., & Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *American Journal of Human Genetics*, 100 (2), 267-280. doi:10.1016/j.ajhg.2017.01.004
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., . . . Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17 (1), 122. doi:10.1186/s13059-016-0974-4
- Mester, J. L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K. S., . . . Eng, C. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Human Mutation*, 39 (11), 1581-1592. doi:10.1002/humu.23636
- Morton, C. C., & Nance, W. E. (2006). Newborn hearing screening—a silent revolution. *New England Journal of Medicine*, 354 (20), 2151-2164. doi:10.1056/NEJMra050700
- Nicora, G., Limongelli, I., Gambelli, P., Memmi, M., Malovini, A., Mazzanti, A., . . . Bellazzi, R. (2018). CardioVAI: An automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases. *Human Mutation*, 39 (12), 1835-1846. doi:10.1002/humu.23665
- Oza, A. M., DiStefano, M. T., Hemphill, S. E., Cushman, B. J., Grant, A. R., Siegert, R. K., . . . ClinGen Hearing Loss Clinical Domain Working, G. (2018). Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation*, 39 (11), 1593-1613. doi:10.1002/humu.23630
- Rehm, H. L., Berg, J. S., & Plon, S. E. (2018). ClinGen and ClinVar – Enabling Genomics in Precision Medicine. *Human Mutation*, 39 (11), 1473-1475. doi:10.1002/humu.23654
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). 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. *Genetics in Medicine*, 17 (5), 405-424. doi:10.1038/gim.2015.30
- Shah, N., Hou, Y. C., Yu, H. C., Sainger, R., Caskey, C. T., Venter, J. C., & Telenti, A. (2018). Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *American Journal of Human Genetics*, 102 (4), 609-619. doi:10.1016/j.ajhg.2018.02.019



- Shen, J., Oza, A. M., Del Castillo, I., Duzkale, H., Matsunaga, T., Pandya, A., . . . ClinGen Hearing Loss Working, G. (2019). Consensus interpretation of the p.Met34Thr and p.Val37Ile variants in GJB2 by the ClinGen Hearing Loss Expert Panel. *Genetics in Medicine*, *21* (11), 2442-2452. doi:10.1038/s41436-019-0535-9
- Vaché, C., Besnard, T., Blanchet, C., Baux, D., Larrieu, L., Faugère, V., . . . Roux, A.-F. (2010). Nasal epithelial cells are a reliable source to study splicing variants in Usher syndrome. *Human Mutation*, *31* (6), 734-741. doi:10.1002/humu.21255
- Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., . . . Ware, J. S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine*, *19* (10), 1151-1158. doi:10.1038/gim.2017.26
- Whiffin, N., Roberts, A. M., Minikel, E., Zappala, Z., Walsh, R., O'Donnell-Luria, A. H., . . . Ware, J. S. (2019). Using High-Resolution Variant Frequencies Empowers Clinical Genome Interpretation and Enables Investigation of Genetic Architecture. *American Journal of Human Genetics*, *104* (1), 187-190. doi:10.1016/j.ajhg.2018.11.012
- Whiffin, N., Walsh, R., Govind, R., Edwards, M., Ahmad, M., Zhang, X., . . . Cook, S. A. (2018). CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation. *Genetics in Medicine*, *20* (10), 1246-1254. doi:10.1038/gim.2017.258
- Wilson, B. S., Tucci, D. L., Merson, M. H., & O'Donoghue, G. M. (2017). Global hearing health care: new findings and perspectives. *Lancet*, *390* (10111), 2503-2515. doi:10.1016/S0140-6736(17)31073-5
- Xiang, J., Peng, J., Baxter, S., & Peng, Z. (2020). AutoPVS1: An automatic classification tool for PVS1 interpretation of null variants. *Human Mutation*, *41* (9), 1488-1498. doi:10.1002/humu.24051
- Xiang, J., Yang, J., Chen, L., Chen, Q., Yang, H., Sun, C., . . . Peng, Z. (2020). Reinterpretation of common pathogenic variants in ClinVar revealed a high proportion of downgrades. *Scientific Reports*, *10* (1), 331. doi:10.1038/s41598-019-57335-5
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*, *11* (2-3), 377-394. doi:10.1089/1066527041410418

## Hosted file

Tables.pdf available at <https://authorea.com/users/380351/articles/496373-vip-hl-semi-automated-acmg-amp-variant-interpretation-platform-for-genetic-hearing-loss>



# VIP-HL

## Variant Interpretation Platform for Genetic Hearing Loss

1 Search variant by HGVS nomenclature, chromosomal position or dbSNP identifier  
 2 Google link containing search strings  
 3 Variant information  
 4 ACMG/AMP criteria and classification  
 5 Reference related to this variant  
 6 Six sections  
 7 Adjustable rules  
 8 Detailed explanations of the ratings

1

2 Example: [GJB2\\_13-20763044-A-C](#), NM\_004004.5:c.109G>A, rs111033313, GJB2 c.250G>A

Variant ID : [J3-20763471-C-T](#) 2

Gene : [GJB2](#)

Inheritance : AD/AR

cHGVS : NM\_004004.5:c.250G>A

pHGVS : NP\_003995.2:p.Val84Met (V84M)

[Manual](#)

4 Classification: Likely Pathogenic  
 Criteria: PS3 Supporting, PM2, PM5 Strong, PP3

PSV2 PS3 PM2 PS2 PS4 PA1 PA2 PA3 PM4 PM5 PM6 PP1 PP2 PP3 PP4 PP5

BA1 BS1 BS2 BS3 BS4 BP1 BP2 BP3 BP4 BP5 BP6 BP7

5 PMID: [304029](#) [224799](#) [201143](#) [290391](#) [286339](#) [New](#)

Basic	Population	Composition	Coe/Segregation	Function	Community (0)
1 BA1	Unmet	BA1 should not be applied.			
2 BS1	Unmet	BS1 should not be applied.			
3 BS2	Unmet	0 homozygote(s) are reported in gnomAD, less than the threshold (≥3) for autosomal recessive variants.			
4 PM2	Moderate	7 Pexmax Filtering AF is 0.0000050896 in Ashkenazi Jewish in gnomAD, within the threshold (0zAF<0.00007) for autosomal recessive variants.           8			