Genetics in Medicine (2024) 26, 101213

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine

BRIEF REPORT

# Assessment of the evidence yield for the calibrated PP3/BP4 computational recommendations

Sarah L. Stenton[1,2], Vikas Pejaver[3,4], Timothy Bergquist[3], Leslie G. Biesecker[5], Alicia B. Byrne[1], Emily A.W. Nadeau[6], Marc S. Greenblatt[6], Steven M. Harrison[1,7], Sean V. Tavtigian[8], Predrag Radivojac[9], Steven E. Brenner[10], Anne O'Donnell-Luria[1,2,*], ClinGen Sequence Variant Interpretation Working Group

ARTICLE INFO

*Article history:*
Received 11 March 2024
Received in revised form
10 July 2024
Accepted 11 July 2024
Available online 25 July 2024

*Keywords:*
ACMG/AMP recommendations
Clinical classification
Computational predictors
PP3/BP4 criteria
Variant interpretation

ABSTRACT

**Purpose:** : To investigate the number of rare missense variants observed in human genome sequences by ACMG/AMP PP3/BP4 evidence strength, following the ClinGen-calibrated PP3/BP4 computational recommendations.

**Methods:** Missense variants from the genome sequences of 300 probands from the Rare Genomes Project with suspected rare disease were analyzed using computational prediction tools that were able to reach PP3_Strong and BP4_Moderate evidence strengths (BayesDel, MutPred2, REVEL, and VEST4). The numbers of variants at each evidence strength were analyzed across disease-associated genes and genome-wide.

**Results:** From a median of 75.5 rare (≤1% allele frequency) missense variants in disease-associated genes per proband, a median of one reached PP3_Strong, 3-5 PP3_Moderate, and 3-5 PP3_Supporting. Most were allocated BP4 evidence (median 41-49 per proband) or were indeterminate (median 17.5-19 per proband). Extending the analysis to all protein-coding genes genome-wide, the number of variants reaching PP3_Strong score thresholds increased approximately 2.6-fold compared with disease-associated genes, with a median per proband of 1-3 PP3_Strong, 8-16 PP3_Moderate, and 10-17 PP3_Supporting.

**Conclusion:** A small number of variants per proband reached PP3_Strong and PP3_Moderate in 3424 disease-associated genes. Although not the intended use of the recommendations, this was also observed genome-wide. Use of PP3/BP4 evidence as recommended from calibrated computational prediction tools in the clinical diagnostic laboratory is unlikely to inappropriately contribute to the classification of an excessive number of variants as pathogenic or likely pathogenic by ACMG/AMP rules.



## Introduction

Genetic testing identifies many variants of uncertain significance (VUS), of which the majority of coding variants are missense (nonsynonymous).[1] Limited availability of functional data means that it is often necessary to turn to computational in silico prediction tools for evidence of deleteriousness. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) provide a sequence variant classification

The names of the ClinGen Sequence Variant Interpretation Working Group members will appear at the end of the article.

*Correspondence and requests for materials should be addressed to Anne O'Donnell-Luria, Boston Children's Hospital, 300 Longwood Avenue, CLS15023 Boston, MA 02115. *Email address:* odonnell@broadinstitute.org

Affiliations are at the end of the document.

doi: https://doi.org/10.1016/j.gim.2024.101213
1098-3600/

framework to combine distinct lines of evidence of pathogenicity or benignity of varying strengths to reach a final variant classification (benign [B], likely benign [LB], VUS, likely pathogenic [LP], or pathogenic [P]). In the 2015 recommendations, in silico evidence (PP3 and BP4) was capped at "Supporting" for or against pathogenicity.[2] Furthermore, no explicit recommendations concerning the prediction tools or thresholds to be used were specified, enabling nonstandardized application of criteria and resulting in inconsistencies in variant classification between clinical diagnostic laboratories.[3]

Recently, Pejaver et al[4] refined the use of computational prediction tools to provide evidence of pathogenicity using the Bayesian adaptation of the ACMG/AMP framework.[4,5] For 13 computational prediction tools frequently used in clinical workflows, evidence-based calibrated thresholds were introduced corresponding to "Supporting," "Moderate," "Strong," and "Very Strong" PP3/BP4 evidence strengths, and also defined an indeterminate range. These thresholds demonstrated that the initial framework underweighted evidence from computational prediction tools because many had the ability to provide evidence beyond "Supporting" strength.

Because of the release of the PP3/BP4 recommendations, we have received questions from users regarding the key steps to implementation, calling for practical guidance on the intended use of the PP3/BP4 recommendations for variant curation in disease-associated genes (see Box 1). In particular, concerns have arisen because of the impression that an excessive number of variants are reaching PP3_Strong. Here, by demonstrating the level of PP3/BP4 evidence allocated to rare missense variants in the genome sequences of patients with rare disease, we specifically aimed to address these concerns.

## Materials and Methods

### Study participants and data

Genome sequencing (GS) data were obtained from the Rare Genomes Project at the Broad Institute of MIT and Harvard.[6] All participants signed informed consent including the use of data for research purposes (Mass General Brigham IRB protocol 2016P001422). Participant demographics are displayed in Supplemental Table 1. Sequencing was performed on DNA purified from blood by the Broad Institute Genomics Platform on an Illumina sequencer to 30× average depth. Raw sequence reads were aligned to the GRCh38 reference genome. Variants were called with Genome Analysis Toolkit version 4.1.8.0[7] in the form of single nucleotide variants (SNVs) and small insertions/deletions (indels). Variants were filtered at the site level with Genome Analysis Toolkit Variant Quality Score Recalibration.

### Missense variant extraction and annotation

Missense variants identified by the Ensembl Variant Effect Predictor (VEP)[8] using Matched Annotation from NCBI and EMBL-EBI (MANE) Select transcripts[9] were extracted from the GS data. Only variants with genotype quality ≥40, depth ≥10, and allele balance ≥0.2 were retained for analyses. Allele-frequency (AF) thresholds of ≤5% and ≤1% global and population-max "popmax" AF in gnomAD v3.1.2 genomes were applied (the highest allele frequency for nonbottlenecked populations).[10] Precomputed scores from 4 in silico (meta)predictors that were able to reach PP3_Strong and BP4_Moderate in the Pejaver et al[4] calibration were included in the analysis. BayesDel (without minor allele frequency),[11] REVEL,[12] and VEST4[13] were annotated using the dbNSFP4.4a database.[14] MutPred2 scores were also generated.[15] For transcript-specific predictors, MutPred2, REVEL, and VEST4, the MANE Select transcript was used for score annotation.

Using recommended thresholds,[4] PP3/BP4 evidence strength per variant was annotated for each prediction tool. The number of missense variants summed across probands and per proband by prediction tool and evidence strength was assessed in disease-associated genes, classified as "Definitive," "Strong," or "Moderate" in the Gene Curation Coalition Database (last accessed Jul 21, 2023) (3424 genes-1004 autosomal dominant only [AD-only], 1903 autosomal recessive only [AR-only], 517 other [includes gene that are both AD and AR])[16] and genome-wide. These methods were also repeated for missense variants according to the VEP "most severe consequence" across all transcripts (see Supplemental Methods).

Statistical analyses. Proportions between 2 groups were compared with 2-tailed binomial tests with Bonferroni correction for multiple testing. Bootstrap resampling with replacement (1000 iterations) was performed to provide a 95% confidence interval (CI) for the mean.

## Results

### Detection of missense variants in disease-associated genes

The GS data set included 300 probands with rare disease. Across protein-coding genes genome-wide, a median of 8781.5 missense variants per proband (range 8383-10,616) passing QC thresholds were detected in MANE Select transcripts. Applying a ≤1% AF threshold in the gnomAD v3 genomes data set, we found 75,384 unique missense variants across 15,566 genes (median 321 per proband, range 244-847). Within Gene Curation Coalition Moderate, Strong, and Definitive disease-associated genes, the number

of unique variants dropped to 17,789 across 2899 genes, and a median of 75.5 variants per proband (range 53-186). Variant counts following each step in QC and AF filtering are displayed in Supplemental Table 2.

## PP3/BP4 evidence strength of missense variants in disease-associated genes

In disease-associated genes, a median of 1 missense variant (mean 0.8-1) per proband reached PP3_Strong per analyzed prediction tool, 3-5 variants (mean 3.4-4.9) reached PP3_Moderate, and 3-5 (mean 3.6-5.2) reached PP3_Supporting (Table 1, Figure 1A). Summed across all probands, 227 to 313 PP3_Strong variants were found in a total of 153 to 196 disease-associated genes, accounting for 0.96% to 1.3% of all analyzed variants (BayesDel 1.3% [95% CI 1.321-1.330], MutPred2 1.0% [95% CI 1.037-1.044], REVEL 0.96% [95% CI 0.957-0.966], and VEST4 1.1% [95% CI 1.113-1.114]). PP3_Moderate-Strong variants were more frequent in AR genes (mean 2.6-4.1 per proband) than in AD genes (mean 0.7-1.0 per proband) ($P$ value $\leq$ .0001 for all prediction tools, 2-tailed binomial test with Bonferroni correction) (Figure 1B and Supplemental Figure 1). At a more stringent threshold of 0.1% AF, routinely used for dominant disease, a mean of 0.6 to 0.7 PP3_Moderate-Strong variants were detected in AD genes per proband. ClinVar provides classifications for 54% to 72% of the unique variants with PP3_Strong evidence per prediction tool, of which 12% to 29% are currently reported as P/LP, 63% to 79% as VUS or conflicting interpretations of pathogenicity, and 7% to 10% as B/LB (Supplemental Figure 2). The majority of analyzed variants were allocated BP4 evidence (median 41-49 per proband, 53%-64% of all analyzed variants) or were indeterminate (median 17.5-19 per proband, 23%-25% of all analyzed variants) (Supplemental Figure 3). Using a preliminary calibration, the newly released prediction tool AlphaMissense[17] was generally consistent with each of these figures (data not shown).

Using a less stringent AF threshold of $\leq$5% resulted in a subtle increase in variants with PP3 evidence in disease-associated genes (median = 1 PP3_Strong, median = 4-6 PP3_Moderate, median = 5-6 PP3_Supporting) (Supplemental Table 3). Using the VEP "most severe consequence" across all transcripts to detect variants, a rare disease analysis approach that is sometimes used to increase the detection of potentially deleterious missense variants in alternative transcripts versus using only MANE Select transcripts, we also did not see many more variants reaching PP3_Supporting-Strong in disease-associated genes (median = 1 PP3_Strong, median = 3-5 PP3_Moderate, median = 4-5 PP3_Supporting) (Supplemental Table 4).

Although the ACMG/AMP sequence variant classification framework is intended to be used in genes associated with Mendelian disease, we also analyzed variants genome-wide and found 1-3 PP3_Strong, 8-16 PP3_Moderate, and 10-17 PP3_Supporting variants per proband (Supplemental Table 5). Summed across all probands, a total of 447 to 847 variants reaching PP3_Strong score thresholds were found in 317 to 587 protein-coding genes genome-wide, accounting for 0.442% to 0.837% of all analyzed variants (BayesDel 0.611% [95% CI 0.610-0.613%], MutPred2 0.837% [95% CI 0.835-0.838%], REVEL 0.442% [95% CI 0.441-0.444%], VEST4 0.792% [95% CI 0.791-0.794%]). This equates to a 2-3.5-fold (mean 2.6-fold) higher number than found at the same AF threshold in disease-associated genes only.

## Discussion

The use of computational prediction tools to provide evidence of pathogenicity and benignity within the ACMG/AMP framework was recently refined by Pejaver et al,[4] and certain prediction tools were found capable of reaching "Strong" and "Very Strong" evidence for PP3 and BP4 codes, respectively. These changes were expected to have important implications for the final classification of missense variants in the clinical diagnostic setting, given that previously the codes were capped at "Supporting" and could only be applied if "Multiple" lines of computational evidence support a deleterious effect on the gene or gene product.[2]

Through various scientific meetings and interactions after the release of the recommendations, concerns were raised because of the impression that an excessive number of PP3_Strong variants are generated. To explore these concerns, we assessed the observed number of rare missense variants by PP3/BP4 evidence strength in the genome sequences of 300 research participants with rare disease.

In our analyses, at $\leq$1% AF, a standard threshold in rare disease analysis, we found a median of 1 PP3_Strong variant per individual (range 0-4) across ~950 of over 3400 analyzed disease-associated genes. Most variants had evidence of benignity or no evidence (indeterminate). The rate of PP3_Strong variants among all rare missense variants in disease-associated genes per individual ranged from 0.96% to 1.3%, which is similar to previous reports for variants sampled from gnomAD (1.4%-1.7%) in an analysis using a distinctly defined set of disease-associated genes.[4] PP3_Moderate and PP3_Strong variants were more often observed per proband in genes associated with disorders having AR inheritance (in which heterozygous deleterious variants can be expected that are nondiagnostic in the individual) compared with genes associated with disorders having AD inheritance, in which heterozygous variants that are nondiagnostic would represent false positives. Moreover, for PP3_Strong variants with ClinVar classifications available, $\leq$10% were classified as B/LB. These figures demonstrate the low frequency of PP3_Strong evidence contributing to a false-positive diagnosis and also confirm that gnomAD is an appropriate reference set for score calibration for application in variant classification.

To better understand why users reported an excess of PP3_Strong variants, we also extended our analyses to more

**Table 1** Number of rare (≤1% AF) missense variants in disease-associated genes per proband by ACMG PP3/BP4 evidence strength within MANE Select transcripts

| | Computational Prediction Tool | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BayesDel | | MutPred2 | | REVEL | | VEST4 | |
| ACMG/AMP Evidence Code | Median (range) | Mean (95% CI) | Median (range) | Mean (95% CI) | Median (range) | Mean (95% CI) | Median (range) | Mean (95% CI) |
| PP3_Strong | 1 (0-4) | 1 (0.9-1.2) | 1 (0-4) | 0.8 (0.7-0.9) | 1 (0-4) | 0.8 (0.7-0.9) | 1 (0-4) | 0.9 (0.8-1.0) |
| PP3_Moderate | 4 (0-10) | 4.1 (3.9-4.3) | 4 (0-12) | 4 (3.74-4.19) | 3 (0-9) | 3.4 (3.2-3.6) | 5 (0-11) | 4.9 (4.7-5.1) |
| PP3_Supporting | 4 (0-13) | 4.6 (4.3-4.8) | 4 (0-11) | 3.8 (3.58-4.04) | 3 (0-9) | 3.6 (3.4-3.8) | 5 (0-13) | 5.2 (4.9-5.4) |
| Indeterminate | 18 (7-35) | 18.6 (18.0-19.1) | 19 (8-44) | 19.7 (19.2-20.3) | 18 (8-36) | 18.4 (17.9-19.0) | 17.5 (4-31) | 17.9 (17.4-18.3) |
| BP4_Supporting | 17 (8-48) | 17.8 (17.2-18.4) | 18 (9-39) | 18.2 (17.6-18.8) | 11 (3-23) | 11 (10.6-11.4) | 11 (2-40) | 11.3 (10.9-11.8) |
| BP4_Moderate | 31 (15-91) | 32.3 (31.3-33.4) | 31 (14-96) | 32.2 (31.1-33.2) | 28 (14-84) | 29.3 (28.3-30.3) | 32 (15-91) | 32.8 (31.7-34.0) |
| BP4_Strong | - | - | 0 (0-2) | 0.1 (0-0.1) | 1 (0-11) | 1.6 (1.4-1.8) | - | - |
| BP4_Very Strong | - | - | - | - | 0 (0-1) | 0 (0-0.1) | - | - |
| No score | 0 (0-3) | 0.3 (0.28-0.40) | 0 (0-0) | 0 (0-0) | 10 (2-28) | 10.7 (10.2-11.1) | 5 (1-19) | 5.8 (5.5-6.1) |

"-" indicates that the given prediction tool is not able to provide BP4 evidence of this strength.
*AF*, allele frequency; *CI*, confidence interval.

frequent variants up to 5% AF, the threshold for stand-alone evidence of benignity in the ACMG/AMP guidance, and to variants that are missense on alternative transcripts (VEP "most severe consequence"). These analyses did not result in a considerable increase in the number of PP3_Strong variants. Furthermore, though Pejaver et al[4] made no recommendation about running computational prediction tools genome-wide, given that the thresholds are calibrated for disease-associated genes only, we applied the same thresholds to variants genome-wide. We found an approximately 2.6-fold increase in the number of variants reaching PP3_Strong score thresholds genome-wide compared with within disease-associated genes only, consistent with the genome having ~5-fold as many genes as covered by ACMG/AMP classification rules and the prior for pathogenicity genome-wide being ~5-fold lower (~1%)[15,18] than for disease-associated genes (~4.5%).[4]

Importantly, deleterious in silico prediction does not equate to pathogenicity and, in the absence of additional evidence, one line of "Strong" evidence from the PP3 code classifies a variant as a VUS in the ACMG/AMP framework. In the case that a variant does reach P or LP classification in combination with other codes, there is a 99% or 90% posterior probability of pathogenicity, respectively, which implies that 1% to 10% of variants may not actually be causative of disease. The PP3/BP4 codes should be used within the framework of the ACMG/AMP recommendations, including the updates that have been made by ClinGen to determine the pathogenicity of a variant. Code combination requires great care, and there are a number of important caveats. In particular, (meta)predictors may use data partially captured by other codes, notably key domains and critical residues and population AF, increasing the risk of double counting of evidence (see Supplemental Discussion for further recommendations on code combination). For 2 reasons, methods that incorporate allele frequency as an explicit feature, as well as those that had strong implicit use of allele frequency, were therefore not calibrated by Pejaver et al.[4] First, use of a variant impact predictor incorporating allele frequency will limit use of lines of evidence depending upon allele frequency, such as BA1, in variant classification. In practice, this means that such methods are impractical to use in most clinical classification pipelines. Second, methods using allele frequency typically need to be calibrated distinctly for different AF thresholds (Rastogi R, Chung R, Li S, et al. Critical assessment of missense variant effect predictors on disease-relevant variant data. Published online June 8, 2024:2024.06.06.597828. bioRxiv. 2024. https://doi.org/10.1101/2024.06.06.597828), and we lack sufficient data to conduct robust calibration in that manner.

The PP3/BP4 calibration by Pejaver et al[4] does have limitations. It was performed on variants classified in the past several years that were not used in the training sets of the analyzed prediction tools and may be nonrepresentative of novel variants to be classified. Moreover, computational prediction tools were assumed not to have played a major role in the classification of the variants used for the

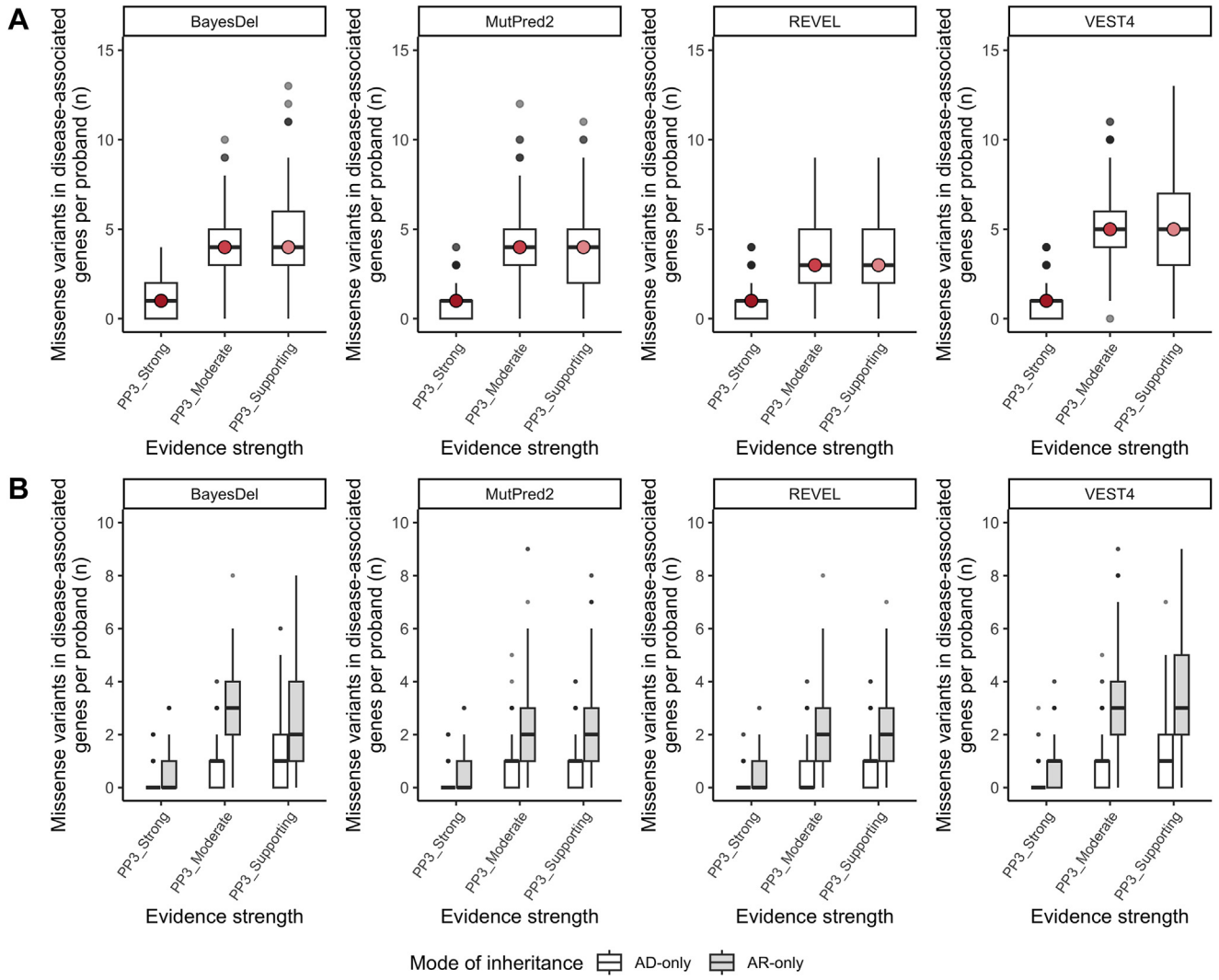


**Figure 1 PP3 evidence strength of missense variants in disease-associated genes.** A. Rare (≤1% AF) missense variants in disease-associated genes per proband by PP3 evidence strength for analyzed computational prediction tools. B. Rare (≤1% AF) missense variants in disease-associated genes with PP3 evidence per proband by evidence strength and reported mode of inheritance (AD-only and AR-only) for analyzed computational tools. Boxplots correspond to the first, second, and third quartile of data, with whiskers denoting 1.5 × IQR. Outliers are displayed as individual points.

calibration. Given these limitations, we appeal to diagnostic labs to report the computational prediction tool and version used for PP3/BP4 evidence, both to determine the impact the PP3/BP4 recommendations are having on the final classification of missense variants and to ensure that we are able to continue to evaluate the performance of computational predictors in the future.

## Conclusion

A small number of variants per rare disease proband reached PP3_Strong and PP3_Moderate in disease-associated genes following the calibrated PP3/BP4 computational recommendations. Computational methods are therefore unlikely to inappropriately classify variants as P/LP by ACMG/AMP rules in the clinical setting by the ACMG/AMP framework.

## Data Availability

Sequence Compressed Reference-orientated Alignment Map (CRAM) files and metadata for the Rare Genomes Project is available through the Broad Institute Data Use Oversight System (DUOS) at duos.broadinstitute.org under data set IDs DUOS-000008 (HMB) and DUOS-000143 (GRU) or via dbGaP accession numbers phs003047 (GREGoR data set). Access is managed by a data access

**Box 1.** Key steps in implementing the PP3/BP4 missense variant recommendations

**How should PP3/BP4 evidence be used for missense variants?**
- -Only apply PP3/BP4 evidence in genes where missense variants are known to cause disease.
- -Use a single computational prediction tool, preferably one able to reach PP3_Strong and BP4_Moderate.
- -Select the prediction tool in advance of seeing the scores, and preferably before knowing the results of any other line of evidence (ie, do not "cherry pick" a prediction tool).
- -Use ClinGen-recommended[4] established thresholds unless or until there is superseding gene or region-specific guidance.

**How can computational PP3 evidence be combined with other ACMG/AMP codes?**
- -Use the PP3/BP4 evidence within the ACMG/AMP rules for variant classification and implement all updated recommendations together.
- -Code combination must avoid double counting of evidence, for example:
  - -The combined evidence strength of PP3 and PM1 is capped at "Strong."
  - -Prediction tools that explicitly incorporate allele frequency should not be combined with independent allele-frequency evidence.

**How can the calibrations be customized by expert panels to specific genes or regions?**
- -Apply the single prediction tool that performs best for the gene(s) or region.
- -If there is evidence that the prediction tool yields inappropriate predictions for a specific gene or region, well-informed judgment may be used to adjust PP3/BP4 evidence.
- -For genes with sufficient number of benign and pathogenic missense variants, it may be possible to perform gene (or region) specific calibration.

**Can the calibration of these methods be trusted?**
- -PP3/BP4 are empirically calibrated evidence codes.
- -Confounders that could be addressed directly were eliminated.
- -As with any approach, it is expected that the evidence strength provided will be too high or too low for some variants when applying the calibrated PP3/BP4 codes.
- -The calibrated codes have been extensively validated, including in this study.

**What are some of the limitations to the calibration?**
- -Variants used for the calibration may not be representative of novel variants to be classified.
- -Computational prediction tools were assumed not to have had a major role in the classification of variants used in the calibration.
- -The calibration provides the evidence strength, on average, across the thousands of genes assessed; however, it is a probability that will vary across genes.

**Will more calibrations need to be performed in the future?**
- -New and revised methods will require independent validated calibration.

committee and is based on intended use of the requester and allowed use of the data submitter as defined by consent codes (some are health/medical/biomedical research [HMB] and some are general research use [GRU]).

## Funding

S.L.S. is supported by a fellowship from the Manton Center for Orphan Disease Research at Boston Children's Hospital. Data were provided by Broad Institute of MIT and Harvard Center for Mendelian Genomics with funding to A.O.'D.-L. by the National Human Genome Research Institute (NHGRI) grants UM1 HG008900 and U01 HG011755, and by the Chan Zuckerberg Initiative through an advised fund of the Silicon Valley Community Foundation grant 2020-224274, and by research funding from Illumina Inc. P.R. is supported by NHGRI grant U01 HG012022. V.P. is supported by National Library of Medicine grant R00LM012992. E.A.W.N., M.S.G., and S.V.T. are supported by the National Institutes of Health grant R01 CA264971. A.B.B. and S.M.H. are supported by National Institutes of Health grant U24 HG006834. L.G.B. is supported by Z01 HG200328-18. ClinGen is primarily funded by the NHGRI with co-funding from the National Cancer Institute, through the following grants: U24 HG009649 (to Baylor/Stanford), U24 HG006834 (to Broad/Geisinger), and U24 HG009650 (to UNC/Kaiser).

## Author Contributions

Conceptualization: V.P., L.G.B., S.V.T., P.R., S.E.B., A.O.'D.-L.; Data Curation: S.L.S., V.P., T.B.; Formal Analysis: S.L.S.; Funding Acquisition: V.P., L.G.B., M.S.G., S.V.T., P.R., A.O.'D.-L.; Methodology: S.L.S., V.P., T.B., L.G.B., A.B.B., E.A.W.N., M.S.G., S.M.H., S.V.T., P.R., S.E.B., A.O.'D.-L.; Supervision: V.P., P.R., A.O.'D.-L.; Visualization: S.L.S.; Writing-original draft: S.L.S., S.E.B, A.O.'D.-L.; Writing-review and editing: S.L.S., V.P., T.B., L.G.B., A.B.B., E.A.W.N., M.S.G., S.M.H., S.V.T., P.R., S.E.B., A.O.'D.-L.

## Ethics Declaration

All participants signed informed consent including the use of data for research purposes (Mass General Brigham IRB protocol 2016P001422).

## Members Of Multiauthor Work Group

**ClinGen Sequence Variant Interpretation Working Group:** Leslie G. Biesecker, Steven M. Harrison, Ahmad A. Tayoun, Jonathan S. Berg, Steven E. Brenner, Garry R. Cutting, Sian Ellard, Marc S. Greenblatt, Peter Kang, Izabela Karbassi, Rachel Karchin, Jessica Mester, Anne O'Donnell-Luria, Tina Pesaran, Sharon E. Plon, Heidi L. Rehm, Natasha T. Strande, Sean V. Tavtigian, and Scott Topper.

## Acknowledgments

The authors thank the clinicians and scientists providing feedback to members of the ClinGen Sequence Variant Interpretation Working Group on their experience applying the calibrated PP3/BP4 computational recommendations that prompted this study.

## Conflict of Interest

Leslie G. Biesecker receives royalties from Wolters-Kluwer for authorship of UpToDate, is a member of the Illumina Medical Ethics Committee, and receives research support from Merck, Inc. Anne O'Donnell-Luria oversees the Rare Genomes Project, which received research funding from Illumina Inc. Steven M. Harrison is an employee of Ambry Genetics. Vikas Pejaver and Predrag Radivojac participated in the development of some of the tools assessed in this study. Steven E. Brenner receives research support at UC Berkeley from Tata Consultancy Services. All other authors declare no conflicts of interest.

## Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2024.101213) contains supplemental material, which is available to authorized users.

## Affiliations

[1]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; [2]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA; [3]Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY; [4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; [5]Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; [6]Department of Medicine and University of Vermont Cancer Center, University of Vermont, Larner College of Medicine, Burlington, VT; [7]Ambry Genetics, Aliso Viejo, CA; [8]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT; [9]Khoury College of Computer Sciences, Northeastern University, Boston, MA; [10]Department of Plant and Microbial Biology and Center for Computational Biology, University of California, (redundant), Berkeley, CA

## References

1. McInnes G, Sharo AG, Koleske ML, et al. Opportunities and challenges for the computational interpretation of rare variation in clinically important genes. *Am J Hum Genet*. 2021;108(4):535-548. http://doi.org/10.1016/j.ajhg.2021.03.003
2. Richards S, Aziz N, Bale 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. *Genet Med*. 2015;17(5):405-424. http://doi.org/10.1038/gim.2015.30
3. Harrison SM, Dolinsky JS, Knight Johnson AE, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med*. 2017;19(10):1096-1104. http://doi.org/10.1038/gim.2017.14
4. Pejaver V, Byrne AB, Feng BJ, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet*. 2022;109(12):2163-2177. http://doi.org/10.1016/j.ajhg.2022.10.013
5. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. 2018;20(9):1054-1060. http://doi.org/10.1038/gim.2017.210
6. Serrano JG, O'Leary M, VanNoy GE, et al. Advancing understanding of inequities in rare disease genomics. *Clinical Therapeutics*. 2023;45(8):745-753. http://doi.org/10.1016/j.clinthera.2023.06.010
7. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. http://doi.org/10.1101/gr.107524.110
8. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122. http://doi.org/10.1186/s13059-016-0974-4
9. Morales J, Pujar S, Loveland JE, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*. 2022;604(7905):310-315. http://doi.org/10.1038/s41586-022-04558-8
10. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*. 2024;625:92-100.
11. Feng BJ. PERCH: a unified framework for disease gene prioritization. *Hum Mutat*. 2017;38(3):243-251. http://doi.org/10.1002/humu.23158
12. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J*

*Hum Genet*. 2016;99(4):877-885. http://doi.org/10.1016/j.ajhg.2016.08.016

13. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics*. 2013;14(suppl 3):S3. http://doi.org/10.1186/1471-2164-14-S3-S3
14. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med*. 2020;12(1):103. http://doi.org/10.1186/s13073-020-00803-9
15. Pejaver V, Urresti J, Lugo-Martinez J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*. 2020;11(1):5918. http://doi.org/10.1038/s41467-020-19669-x
16. DiStefano MT, Goehringer S, Babb L, et al. The Gene Curation Coalition: a global effort to harmonize gene-disease evidence resources. *Genet Med*. 2022;24(8):1732-1742. http://doi.org/10.1016/j.gim.2022.04.017
17. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381(6664):eadg7492. http://doi.org/10.1126/science.adg7492
18. Critical Assessment of Genome Interpretation Consortium. CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods. *Genome Biol*. 2024;25(1):53. http://doi.org/10.1186/s13059-023-03113-6