BayesQuantify: an R package utilized to refine the ACMG/AMP criteria according to the Bayesian framework

Authors

Sihan Liu (刘思寒)[1, *], Xiaoshu Feng (冯小殊)[1], Fengxiao Bu (卜枫啸)[1, *]

Affiliations

[1] Institute of Rare Diseases, West China Hospital, Sichuan University, Chengdu 610041, China

[*] To whom correspondence should be addressed.

Abstract

Improving the precision and accuracy of variant classification in clinical genetic testing involves further specification and stratification of the ACMG/AMP criteria. The Bayesian framework proposed by ClinGen has provided a mathematical foundation for evidence refinement, successfully quantifying, and extending the evidence strengths of PS1, PS4, PM5, and PP3/BP4. However, existing software and tools designed for quantifying the evidence strength and establishing corresponding thresholds to refine the ACMG/AMP criteria are lacking. To address this gap, we have developed *BayesQuantify*, an R package that aims to provide users with a unified resource for quantifying the strength of evidence for ACMG/AMP criteria using a naive Bayes classifier. By analyzing publicly available data, we demonstrate *BayesQuantify*'s capability to offer objective and consistent refinement of the ACMG/AMP evidence. BayesQuantify is available from GitHub at https://github.com/liusihan/BayesQuantify.

Correspondence: bufengxiao@wchscu.cn; liusihan@wchscu.cn

Keywords: Bayesian framework; ACMG/AMP guidelines; R package; Evidence refinement

Introduction

Interpreting sequence variants in human genes is a pivotal aspect of diagnosing and managing genetic diseases[1,2]. However, this task presents considerable challenges due to the complexity and diversity of the genomic data and the lack of standardization and consistency across laboratories. To address these issues, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) have developed a set of guidelines for variant interpretation, comprising 28 criteria which support or refute the pathogenicity or benignity of a variant[3]. These criteria are grouped into five strength levels of evidence (stand-alone, very strong, strong, moderate, and supporting) and are combined according to specific rules to assign a variant into one of five categories: pathogenic (P), likely pathogenic (LP), uncertain significance (US), likely benign (LB), and benign (B). Since their introduction, the ACMG/AMP guidelines have facilitated the genetic diagnosis of suspected inherited disorders and have become widely adopted by clinical laboratories and researchers as the standard for variant interpretation and reporting[4].

In 2018, the ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI) proposed a Bayesian framework that transformed the ACMG/AMP guidelines into a probabilistic system[5]. This framework assumes four levels of evidence and exponentially scaled odds of pathogenicity (OP). It assigns a posterior probability of pathogenicity (Post_P) to a variant by multiplying the OP for each level of evidence (O_{PVSt}, O_{PSt}, O_{PM}, O_{PSu}) with a prior probability of pathogenicity (Prior_P). The resulting posterior probability is then compared with probability thresholds to categorize variants: (1) the pathogenic combinations would have Post_P > 0.99, (2) the Post_P of the likely pathogenic combinations would be 0.90 < Post_P ≤ 0.99, (3) the Post_P of the likely benign combinations would be 0.001 ≤ Post_P < 0.10, and (4) the one benign category would have Post_P < 0.001. By modelling the ACMG/AMP guidelines with a Prior_P of 0.10 and O_{PVSt} set to 350, 16 out of 18 existing ACMG/AMP evidence combinations were found to be mathematically consistent with the overall framework, validating the approach adopted by the ACMG/AMP.

The Bayesian framework presents opportunities for refining evidence and combining rules, as well as supporting efforts to automate components of variant pathogenicity assessments. Researchers have quantified and extended the strengths of evidence criteria such as PS1, PM5, and PP3/BP4 according to the Bayesian framework[6]. Importantly, thresholds for PS4, PS4_Moderate, and PS4_Supporting have been quantified by aligning positive likelihood ratio (LR^+) and local positive likelihood ratio (lr^+) values with theoretical OP for each evidence strength in our prior study[7]. The specification and stratification of evidence are essential to establish objective, unbiased, and consistent variant classification[8-10]. However, there is a lack of existing software and tools tailored to quantify the strength of evidence and establish appropriate thresholds to refine the ACMG/AMP criteria.

In this study, we develop *BayesQuantify*, an R package that implements a Bayesian framework for quantifying the strength of evidence for ACMG/AMP criteria using a naive Bayes classifier. *BayesQuantify* provides functions to estimate OP for each level of evidence strength and the posterior probability of pathogenicity for the ACMG/AMP combining rules. Furthermore, *BayesQuantify* facilitates the calculation of LR^+ and lr^+ for various tested cutoffs, aligning the 95% lower boundary of LR^+/lr^+ with OP to establish thresholds. Additionally, users can employ *BayesQuantify* to classify variants in accordance with ACMG/AMP combining rules, Bayesian framework, and scaled point system, as well as visualize both the data and statistic results[3,5,11]. The *BayesQuantify* is available from GitHub at https://github.com/liusihan/BayesQuantify.

BayesQuantify workflow overview

The functions available through *BayesQuantify* were designed to quantify the evidence strength and thresholds following the algorithm proposed by ClinGen SVI. The package requires one essential input: a *data.frame* containing variants classification results, including the variants, ACMG/AMP evidence, variant classification, and tested features. The functions included in *BayesQuantify* consists of six main steps (as shown in Figure 1):

1. **Determining theoretical OP**: The *op_postp()* function calculates the OP and posterior probability of pathogenicity and benignity for strong, moderate, and supporting evidence levels based on provided prior probability and O_{PVSt}. The *auto_select_postp()* function automatically computes reasonable OP for very strong, strong, moderate, and supporting evidence by comparing existing ACMG/AMP evidence combinations to reach probability thresholds with a given Prior_P.

2. **Variants of US (VUS) subclassification:** To construct a truth set of credibly classified variants to refine evidence, the *VUS_classify()* function categorizes VUS into six levels (hot, warm, tepid, cool, cold, and ice cold) according to the Association for Clinical Genomic Science (ACGS) Best Practice Guidelines: 1) hot: 1 strong + 1 supporting or 2 moderate + 1 supporting or 1 moderate + 3 supporting evidence; 2) warm: 1 strong or 2 moderate or 1 moderate + 2 supporting or 4 supporting evidence; 3) tepid: 1 moderate + 1 supporting or 3 supporting evidence; 4) cool: 1 moderate or 2 supporting evidence; 5) cold: 1 supporting evidence; 6) ice cold: no supporting evidence. Variants classified as cool, cold, or ice cold are considered benign-leaning VUS (BL-VUS), while variants classified as hot, warm, or tepid are considered pathogenic-leaning VUS (PL-VUS).

3. **Calculating LR^+:** For binary variables, the *LR()* function computes various metrics for each tested cutoff, including true positive (TP, the number of P/LP variants above a tested cutoff), false positive (FP, the number of BL-VUS/B/LB variants above a tested cutoff), true negative (TN, the number of BL-VUS/B/LB variants below a tested cutoff), and false negative (FN, the number of P/LP variants below a tested cutoff), overall accuracy, true positive rate (sensitivity), true negative rate

(specificity), positive predictive value (PPV), negative predictive value (NPV), F1 score, and LR^+. Estimates of the 95% confidence interval (CI) of LR^+ are generated using bootstrapping in the R package, bootLR[12]. The lower boundary of LR+ (LR^+_LB) is used to determine evidence strength by matching with the OP thresholds generated in step 1.

4. **Calculating lr^+:** For quantitative variables, initially, the *local_bootstrapped_lr()* function sorts all unique tested cutoffs and positions each cutoff at the center of a sliding window (0.01). The posterior probability of pathogenicity and benignity is then calculated for each tested cutoff within the interval, considering a minimum of 100 disease-causing and non-pathogenic variants. Additionally, the one-sided 95% confidence bound for each estimated lr^+ is determined through 10,000 bootstrapping iterations to assess evidence strength.

5. **Variant classification:** The *ACMG_Classification()*, *BCF()*, and *Point_Classification()* functions classify variants into five classes according to existing ACMG/AMP combinations rules, Bayesian framework, and scaled point system, respectively.

6. **Visualization:** The *multi_plot()*, *heatmap_LR()*, and *plot_lr()* functions facilitate the visualization of variant characteristics, statistical metrics, and local posterior probability curves for different tested cutoffs.

All these steps are implemented within the R programming environment to consolidate all intermediate data generated throughout the workflow, ensuring transparency and user-friendly operation. For a detailed explanation of each step with code and examples, see the website of the package (https://github.com/liusihan/BayesQuantify).

Example applications

Example datasets included in *BayesQuantify*

To evaluate its performance, we utilized *BayesQuantify* on two distinct datasets: the ClinGen Curated Variants dataset (referred to as dataset 1) and the ClinVar 2019 dataset (referred to as dataset 2). The ClinGen Curated Variants data, retrieved from the ClinGen Evidence Repository as of April 7th, 2024 (https://erepo.clinicalgenome.org/evrepo/), serves as an FDA-recognized human genetic variant database containing expert-curated assertions regarding variants' pathogenicity and supporting evidence summaries[10]. This dataset encompasses classification summaries for 6,768 curated variants across 74 diseases, including 1850 P, 1463 LP, 679 LB, 775 B, and 2001 US variants (Figure 2 and Table S1).

The ClinVar 2019 dataset, compiled by Pejaver et al., aims to estimate the thresholds of PP3/BP4 criteria for different computational tools and was obtained from https://zenodo.org/records/8347415 [6]. This dataset includes missense variants with an

allele frequency (AF) below 0.01 in the Genome Aggregation Database (gnomAD v.2.1) and from genes with at least one pathogenic variant of any type in ClinVar[13]. In total, 11,834 variants (2,787 P/LP and 6,327 B/LB variants) from 1,914 genes are included in this dataset (Table S2).

Defining the OP for each strength level in the datasets

To establish the OP for each strength level in both datasets, we initially generated truth datasets by excluding PL-VUS. In dataset 1, this resulted in 5,800 variants (referred to as truth dataset 1), comprising 1,850 P, 1,463 LP, 775 B, 679 LB, and 1,033 BL-VUS variants. Correspondingly, the truth dataset 2 encompassed 2,787 disease-causing (P/LP) variants and 6,327 non-pathogenic (B/LB) variants.

Subsequently, we calculated the OP for each evidence strength in the two datasets using the *auto_select_postp()* function. For dataset 1, we adopted the OP values proposed by ClinGen SVI: $O_{PSu} \geq 2.08$, $O_{PM} \geq 4.33$, $O_{PSt} \geq 18.7$, and $O_{PVSt} \geq 350.0$, assuming a Prior_P of 0.1. This choice was influenced by the biased proportion (0.58) of disease-causing variants in dataset 1. Additionally, theoretical OP values for each pathogenic evidence strength in the ClinVar 2019 set were determined as ≥2.41, ≥5.80, ≥33.65, and ≥1132 for supporting, moderate, strong, and very strong evidence, respectively (Table S3). These calculations were based on a set Prior_P of 0.0441, as estimated by Pejaver et al., reflecting the prevalence in the dataset.

Evidence strength assessment

The frequency of each ACMG-AMP code invoked in truth dataset 1 is shown in Figure 3. Notably, all lines of evidence were applied at least once, except for PP5 and BP6, which have been removed in accordance with SVI recommendations. PM2 emerged as the most frequently applied pathogenic criterion, being utilized in 3,072 P/LP variants, 580 BL-VUS, and 114 B/LB variants, respectively (Table S4).

In the 2015 ACMG/AMP guidelines, PM2 is defined as absent from controls or at an extremely low frequency if recessive in the general population database. In the ClinGen curated dataset, PM2 was designated as moderate evidence and supporting evidence in accordance with gene-/disease-specific guidelines. Consequently, a total of 3,766 variants were tagged with PM2 or PM2_Supporting, including 3,072 P/LP variants and 694 BL-VUS/B/LB variants. The accuracy was 0.839, PPV was 0.816, sensitivity was 0.927, specificity was 0.721, and LR^+ was 3.323 (95% CI, 3.121–3.549), reaching the supporting evidence level (Table S5). Furthermore, 1,102 variants were identified with a moderate evidence level of PM2, comprising 1,002 P/LP variants and 100 BL-VUS/B/LB variants. The accuracy was 0.584, PPV was 0.910, sensitivity was 0.302, specificity was 0.960, and LR^+ was 7.522 (95% CI, 6.206–9.276), attaining the moderate evidence level (LR^+_LB ≥ 4.33).

Additionally, utilizing *BayesQuantify*, we were able to reproduce published results in

the ClinVar 2019 dataset. Through 10,000 bootstrap iterations, we calculated local posterior probabilities for each unique PrimateAI score, confirming that PrimateAI scores >0.775 and >0.848 can attain supporting and moderate evidence levels, respectively, consistent with published findings (Figure 4 and Table S6).

Conclusions

The *BayesQuantify* R package is intended to improve the flexibility and convenience of quantifying the strength of evidence and associated thresholds for ACMG/AMP criteria through a Bayesian framework. The example datasets included in this publication illustrate how *BayesQuantify* can offer objective and consistent refinement of evidence, thereby contributing to the development and enhancement of the ACMG/AMP guidelines. *BayesQuantify* is available from GitHub at https://github.com/liusihan/BayesQuantify.

Funding

This work was supported by the Natural Science Foundation of Sichuan Province, China (Grant No. 2024NSFSC1767), the National Natural Science Foundation of China (Grant No. 82171836) and the 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (Grant No. ZYJC20002).

Acknowledgements

We would like to extend a special thanks to Information Center of West China Hospital for technical support. Figure 1 created with BioRender.com.

Conflict of interest

None declared.

Data availability

The ClinGen Curated Variants data was downloaded from https://erepo.clinicalgenome.org/evrepo/. The ClinVar 2019 set was sourced from https://zenodo.org/records/8347415. These datasets are provided in the supplementary tables and are also included in the *BayesQuantify* package.

References

1. Giles, H.H., Hegde, M.R., Lyon, E., Stanley, C.M., Kerr, I.D., Garlapow, M.E., and Eggington, J.M. (2021). The Science and Art of Clinical Genetic Variant Classification and Its Impact on Test Accuracy. Annu Rev Genomics Hum Genet *22*, 285-307. 10.1146/annurev-genom-121620-082709.
2. Marian, A.J. (2020). Clinical Interpretation and Management of Genetic Variants. JACC Basic Transl Sci *5*, 1029-1042. 10.1016/j.jacbts.2020.05.013.
3. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (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. Genet Med *17*, 405-424. 10.1038/gim.2015.30.
4. Niehaus, A., Azzariti, D.R., Harrison, S.M., DiStefano, M.T., Hemphill, S.E., Senol-Cosar, O., and Rehm, H.L. (2019). A survey assessing adoption of the ACMG-AMP guidelines for interpreting sequence variants and identification of areas for continued improvement. Genet Med *21*, 1699-1701. 10.1038/s41436-018-0432-7.
5. Tavtigian, S.V., Greenblatt, M.S., Harrison, S.M., Nussbaum, R.L., Prabhu, S.A., Boucher, K.M., Biesecker, L.G., and ClinGen Sequence Variant Interpretation Working, G. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med *20*, 1054-1060. 10.1038/gim.2017.210.
6. Pejaver, V., Byrne, A.B., Feng, B.J., Pagel, K.A., Mooney, S.D., Karchin, R., O'Donnell-Luria, A., Harrison, S.M., Tavtigian, S.V., Greenblatt, M.S., et al. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet *109*, 2163-2177. 10.1016/j.ajhg.2022.10.013.
7. Liu, S., Zhong, M., Huang, Y., Zhang, Q., Chen, T., Xu, X., Peng, W., Wang, X., Feng, X., Kang, L., et al. (2023). Quantitative thresholds for variant enrichment in 13,845 cases: improving pathogenicity classification in genetic hearing loss. Genome Med *15*, 116. 10.1186/s13073-023-01271-7.
8. Amendola, L.M., Muenzen, K., Biesecker, L.G., Bowling, K.M., Cooper, G.M., Dorschner, M.O., Driscoll, C., Foreman, A.K.M., Golden-Grant, K., Greally, J.M., et al. (2020). Variant Classification Concordance using the ACMG-AMP Variant Interpretation Guidelines across Nine Genomic Implementation Research Studies. Am J Hum Genet *107*, 932-941. 10.1016/j.ajhg.2020.09.011.
9. Patel, M.J., DiStefano, M.T., Oza, A.M., Hughes, M.Y., Wilcox, E.H., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K., Shen, J., et al. (2021). Disease-specific ACMG/AMP guidelines improve sequence variant interpretation for hearing loss. Genet Med *23*, 2208-2212. 10.1038/s41436-021-01254-2.
10. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen--the Clinical Genome Resource. N Engl J Med *372*, 2235-2242. 10.1056/NEJMsr1406261.
11. Tavtigian, S.V., Harrison, S.M., Boucher, K.M., and Biesecker, L.G. (2020). Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. Hum Mutat *41*, 1734-1737. 10.1002/humu.24088.

12. Marill, K.A., Chang, Y., Wong, K.F., and Friedman, A.B. (2017). Estimating negative likelihood ratio confidence when test sensitivity is 100%: A bootstrapping approach. Stat Methods Med Res *26*, 1936-1948. 10.1177/0962280215592907.

13. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434-443. 10.1038/s41586-020-2308-7.

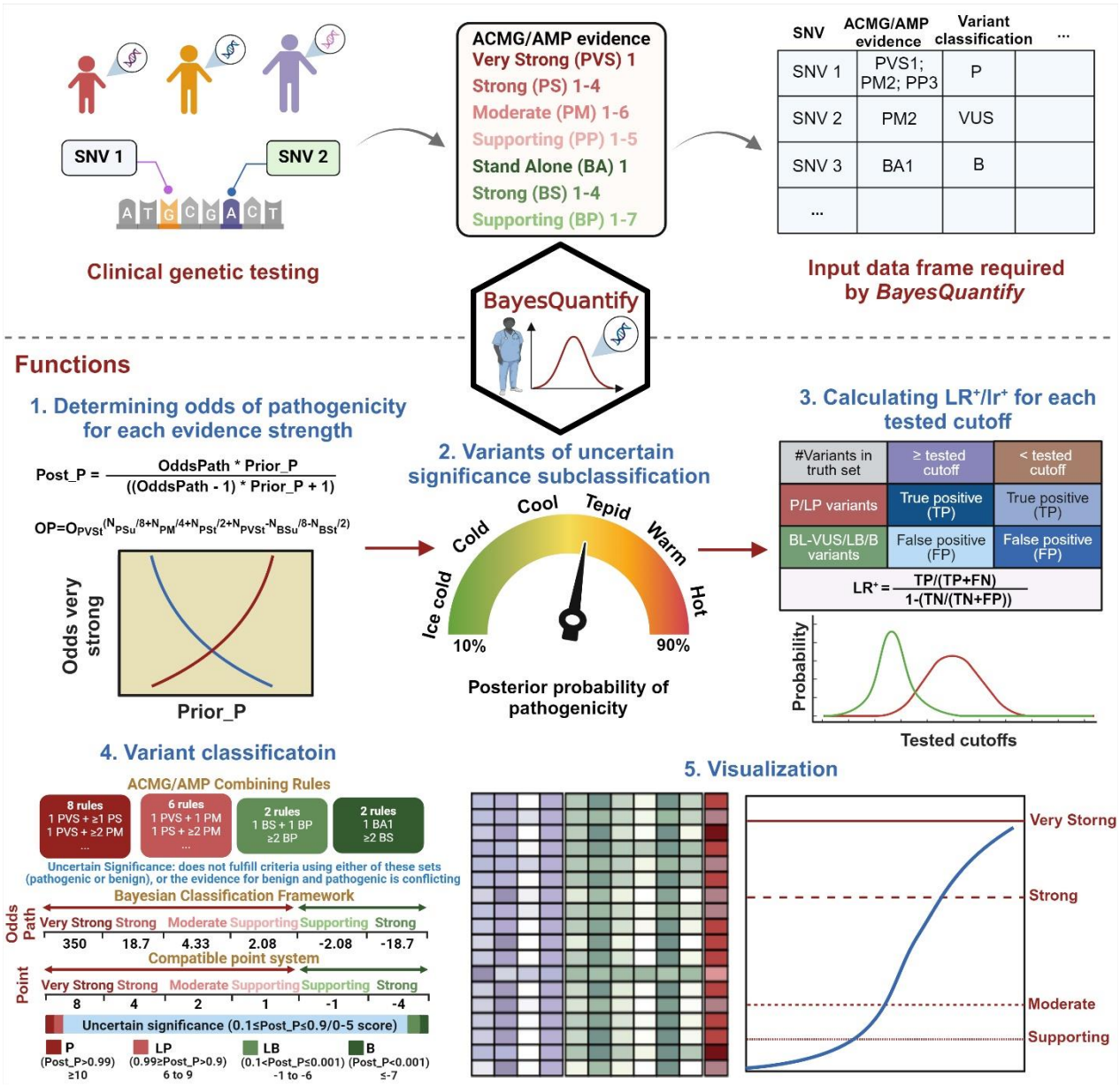
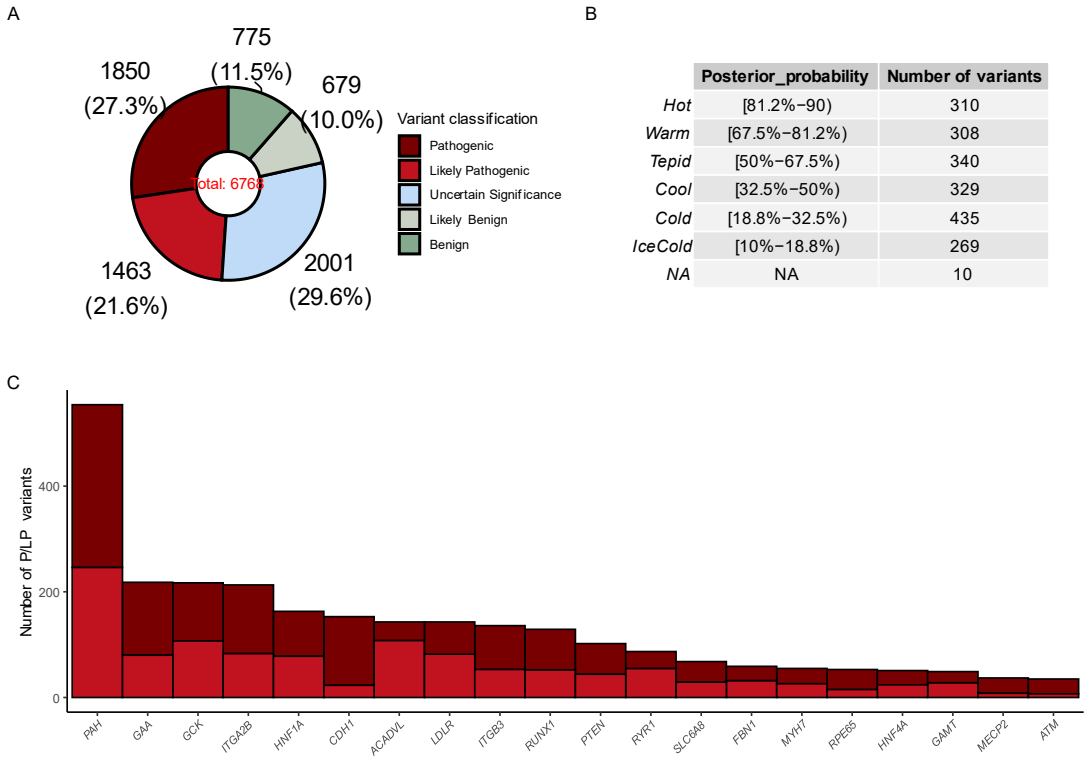


Figure 1: Schematic overview of *BayesQuantify*. *BayesQuantify* begins with the input of variant classification results dataset. Subsequently, it computes the positive likelihood ratio (LR^+) and local positive likelihood ratio (lr^+), which are then aligned with predefined odds of pathogenicity (OP) for each evidence strength to refine evidence. Prior_P: prior probability of pathogenicity; Post_P: posterior probability of pathogenicity.



	Posterior_probability	Number of variants
Hot	[81.2%−90)	310
Warm	[67.5%−81.2%)	308
Tepid	[50%−67.5%)	340
Cool	[32.5%−50%)	329
Cold	[18.8%−32.5%)	435
IceCold	[10%−18.8%)	269
NA	NA	10

Figure 2: Characteristics of the 6,768 variants in the ClinGen Curated Variants dataset. A. Summary of classification for the 6,768 variants. B. Further subclassification of variants of uncertain significance (VUS) to hot, warm, tepid, cool, cold, and ice-cold groups, according to the ACGS Best Practice Guidelines. Variants categorized as "cool/cold/ice cold" were defined as benign-leaning VUS (BL-VUS), while those classified as "hot/warm/tepid" were defined as pathogenic-leaning VUS (PL-VUS). C. Distribution of pathogenic (P) and likely pathogenic (LP) variants across the top twenty genes.

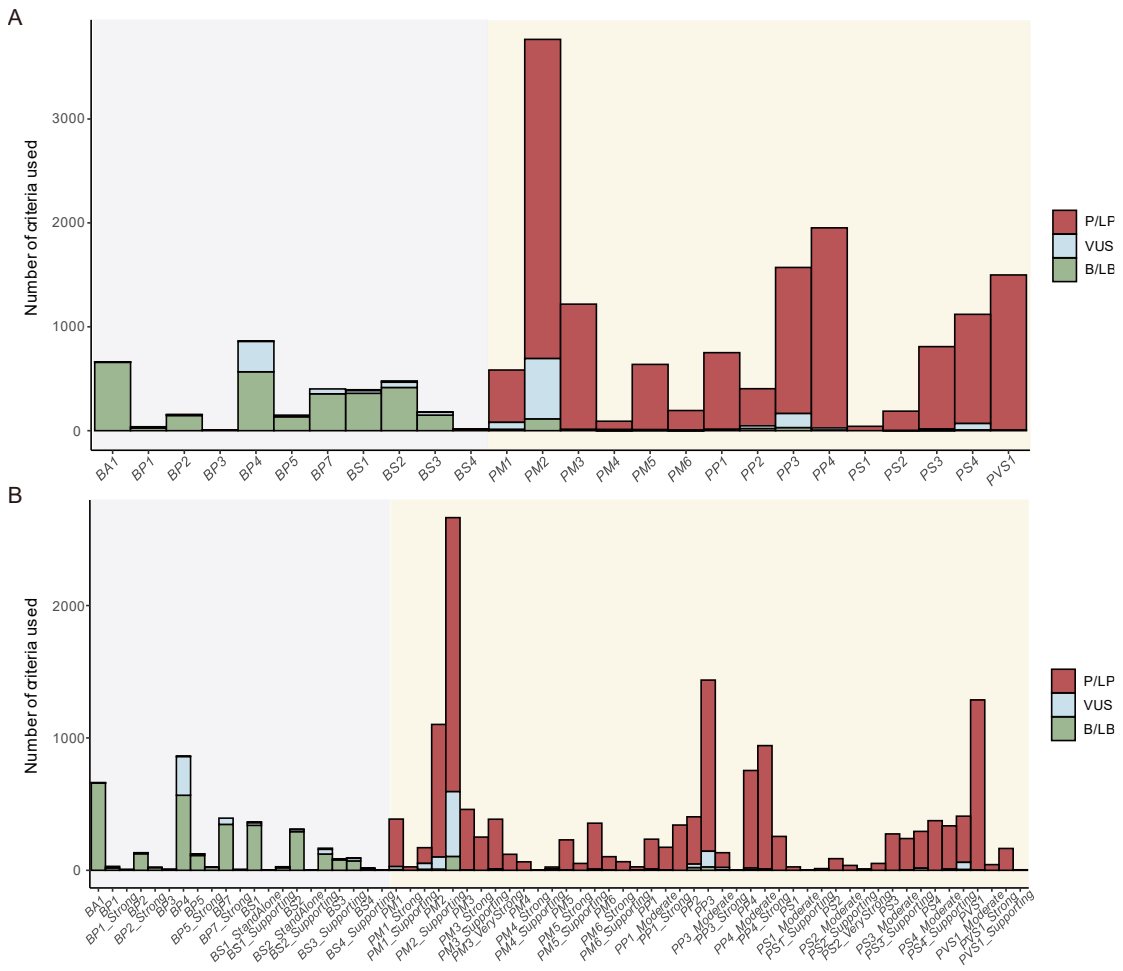


Figure 3: Frequency of the ACMG/AMP criteria applied in the truth dataset 1.

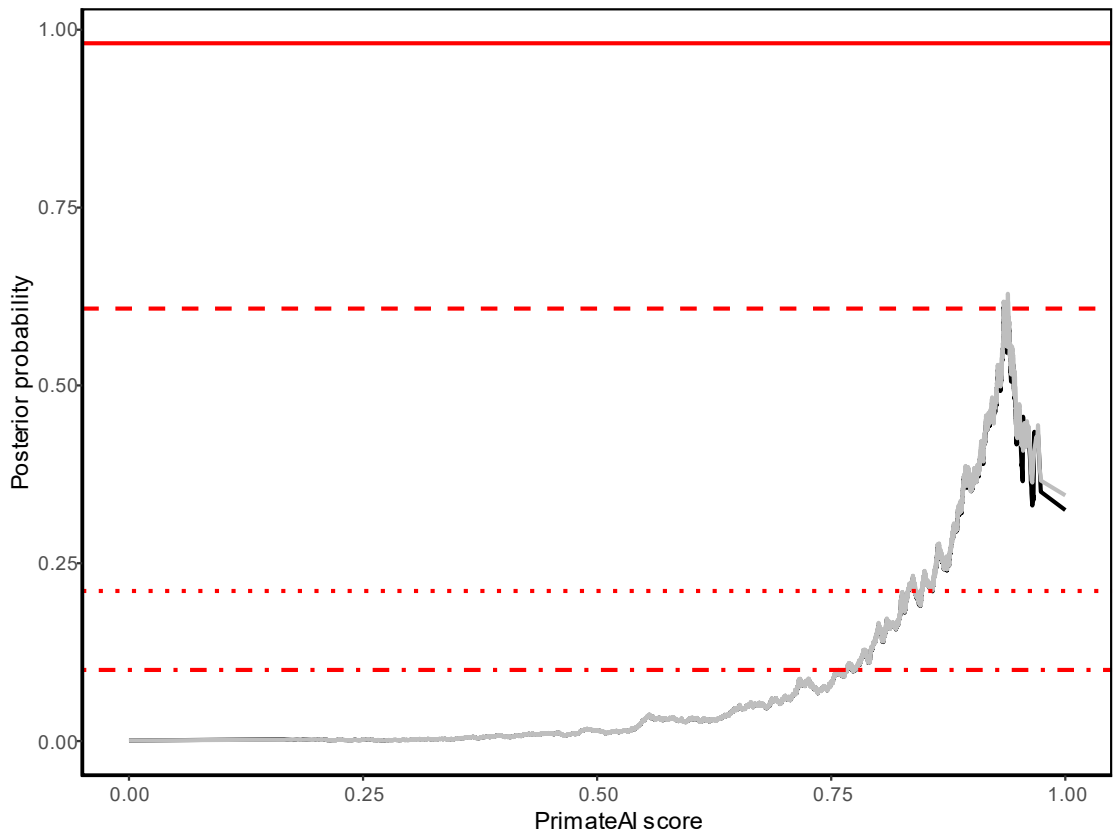


Figure 4: Local posterior probability curves of various PrimateAI scores in the ClinVar 2019 dataset. The horizontal lines depict the posterior probability thresholds for supporting, moderate, strong, and very strong evidence. The black curves represent the posterior probability estimated from ClinVar 2019 dataset, while the grey curves represent one-sided 95% confidence intervals calculated from 10,000 bootstrap samples of this dataset.