Reclassification of variants of uncertain significance through integrated data aggregation and community curation in VarSome

P14.018.F

SAPHETOR

Christos Kitsoulis, Nefeli-Angeliki Paschou, Alma Mudrov, Antonis Koussounadis, Charles E. Chapple

varsome

Background

Variants of uncertain significance (VUS) are the most frequent outcome in germline genetic testing, with almost 30% of patients receiving only VUS results[1]. This delays clinical decisions and forces manual review, complicating the interpretation of rare alleles.

As genomic testing expands, rare variant discovery outpaces clinical curation. Because automated classifiers resolve only some VUS variants, scalable resolution requires systematic integration of diverse data sources. Here, we present an analysis of the performance of VarSome's automated germline classifier[2] (v13.15.2) which integrates evidence from 150+ genomic databases, literature, and community contributions in reclassifying ClinVar[3] VUS variants.

Methods

Of the 2,257,292 variants classified as VUS in ClinVar (data freeze: April 7, 2026), VarSome reclassifies 1,446,362 (64%) as non-VUS. For this study, variants in genes covered by ClinGen Variant Curation Expert Panels (VCEPs) were excluded to prevent confounding by gene-specific interpretation rules, resulting in a final dataset of **1,308,352 ClinVar VUS variants shifted out of VUS by VarSome.**

As missense variants are the most abundant and difficult to decipher, we decided to feature such cases. Representative case studies of missense variants upgraded toward a pathogenic verdict were selected, focusing on the contribution of external databases and expert in-house curation. Within these, VUS resolution was also assessed across two disease-specific genes.

Conclusion

Automated classification frameworks alone cannot fully resolve the current VUS burden. In this analysis, curated clinical evidence from in-house expert review, community contributions, and external databases such as LOVD was critical for reclassifying 1,308,352 ClinVar VUS variants at scale. Much of the resolving evidence originated outside ClinVar, indicating that actionable data for many unresolved variants already exist but remain fragmented across separate repositories. By systematically aggregating, harmonizing, and computationally synthesizing these disparate evidence streams, platforms like VarSome directly overcome the limitations of decentralized data. Consequently, VarSome provides the critical informatics infrastructure required to drive systemic reductions in VUS rates, optimize clinical diagnostic yield, and translate previously ambiguous genomic data into actionable patient diagnoses.

Results

Of the 1,308,352 variants in the final dataset (Table 1), approximately 96.5% were reclassified as benign (B) or likely benign (LB), and 3.5% (44,692) as pathogenic (P) or likely pathogenic (LP). Missense variants were the most common, accounting for 91.9% of the cohort (Figure 1) and 18,686 (41.8%) of these, spread across 5,351 genes, were reclassified as P or LP.

Foundational evidence categories constituted the majority of the total counts, as anticipated (Figure 2). Although the Curation / External Source category represented only 3.7%, expert curation, community contributions and external sources frequently supplied the necessary evidence for definitive P/LP variant resolution. A significant example is the incorporation of independent clinical data from the Leiden Open Variation Database[4] (LOVD).

Table 1. *Quantity of reclassified ClinVar VUS by VarSome*

Germline Classification	Number of variants	Number of variants (filtered)
Benign	141,640	134,918
Likely Benign	1,241,271	1,128,742
Likely Pathogenic	61,493	43,532
Pathogenic	1,958	1,160
Total	**1,446,362**	**1,308,352**

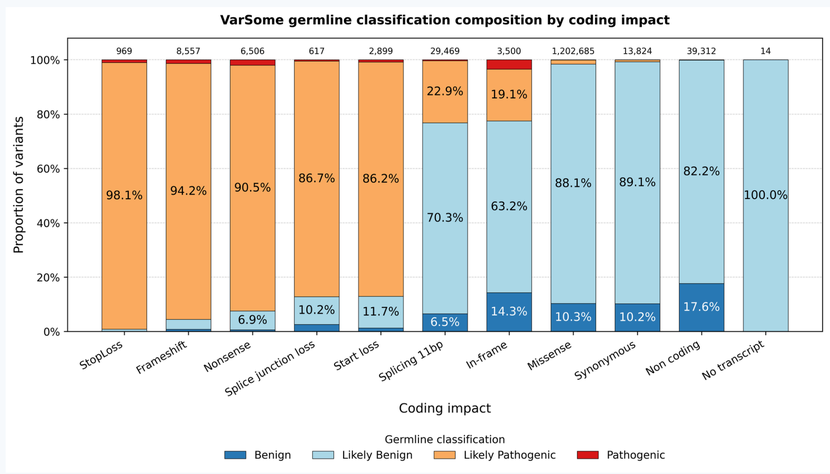


Figure 1. *Proportions of ACMG-aligned germline classifications across variant coding impact categories. Values at the top of each column indicate total sample size (N).*

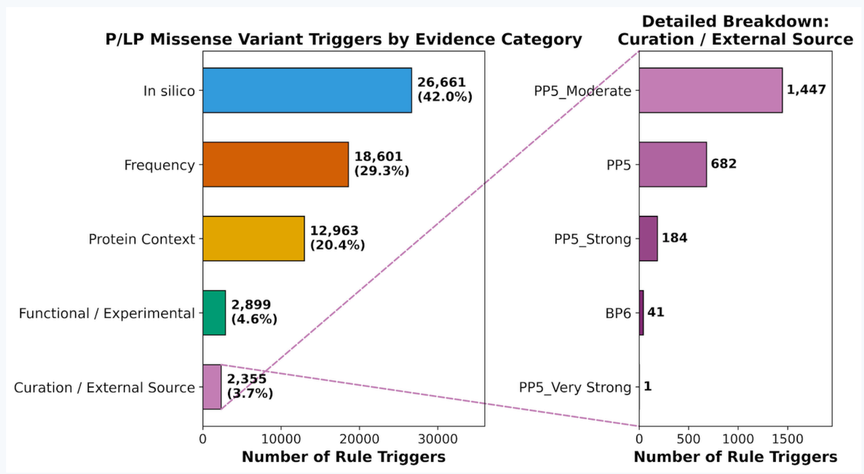
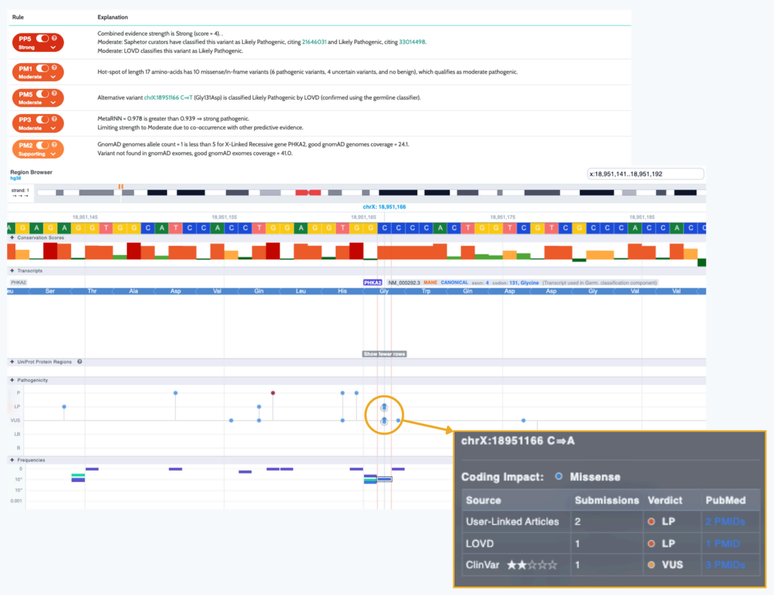


Figure 2. *Distribution of P/LP missense variant rule triggers. The left panel displays the total count and relative percentage of triggers across five major evidence categories. The right panel provides a detailed breakdown of the specific criteria rules applied within the Curation / External Source category (n=2,355)*

The PP5 criterion, supported primarily by LOVD submissions and expert manual curation, drove nearly 12.5% of upgraded classifications. LOVD contributed to approximately 80% of these PP5-driven resolutions, while manual curation and community contributions supported around 9% of cases.

Case 1 - Glycogen Storage Disease

Variant: PHKA2 (NM_000292):c.392G>T
Inheritance: X-linked
Final Verdict: Pathogenic
Reclassification Driver: Combined evidence of expert manual curation and LOVD contribution (Figure 3a).



Case 2 - Leber Congenital Amaurosis Type 9

Variant: NMNAT1 (NM_022787):c.451G>T
Inheritance: Autosomal Recessive
Final Verdict: Pathogenic
Reclassification Driver: Well-documented clinical data from LOVD to trigger specific clinical evidence rules (Figure 3b).

Figure 3. *Integrative variant interpretation for two missense mutations. The top panel displays the automated evaluation of ACMG criteria, highlighting triggered evidence rules and their calculated strengths. The central region browser provides genomic context, illustrating nucleotide conservation and mapped protein domains. Lower annotation tracks visualize the regional landscape of known variants by clinical significance and population frequency. The highlighted inset details the queried variant's coding impact and highlights conflicting external database submissions.*

References

1. Chen E, Facio FM, Aradhya KW, Rojahn S, Hatchell KE, Aguilar S, Ouyang K, Saitta S, Hanson-Kwan AK, Capurro NN, Takamine E, Jamuar SS, McKnight D, Johnson B, Aradhya S. Rates and Classification of Variants of Uncertain Significance in Hereditary Disease Genetic Testing. JAMA Netw Open. 2023 Oct 2;6(10):e2339571. doi: 10.1001/jamanetworkopen.2023.39571.
2. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. Bioinformatics. 2019 Jun 1;35(11):1978-1980. doi: 10.1093/bioinformatics/bty897.
3. Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., et al. (2011). LOVD v.2.0: the next generation in gene variant databases. Human Mutation, 32, 557-563. https://doi.org/10.1002/humu.21438
4. Landrum, M. J., Lee, J. M., Benson, M., et al. (2017). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research, 46, D1062-D1067. https://doi.org/10.1093/nar/gkx1153

Learn more:

Disclosures: C. Kitsoulis, N. A. Paschou, A. Mudrov, A. Koussounadis, C. E. Chapple are employees of Saphetor S.A., the creators of VarSome. The study was conducted using VarSome's internal datasets and classifier framework.