

Thousands of pathogenic variants in ClinVar show signs of incomplete loss-of-function

Tatyana E. Lazareva

tatiana.ev.lazareva@gmail.com

D. O.Ott Research Institute of Obstetrics, Gynaecology, and Reproductology

Yury A. Barbitoff

D. O.Ott Research Institute of Obstetrics, Gynaecology, and Reproductology

Yulia A. Smirnova

D. O.Ott Research Institute of Obstetrics, Gynaecology, and Reproductology

Andrey S. Glotov

D. O.Ott Research Institute of Obstetrics, Gynaecology, and Reproductology

Short Report

Keywords: pLoF, PTV, ClinVar, genetic rescue

Posted Date: September 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10689306/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Thousands of pathogenic variants in ClinVar show signs of incomplete loss-of-function

Tatyana E. Lazareva^{1*†}, Yury A. Barbitoff^{1†}, Yulia A. Smirnova¹,
Andrey S. Glotov¹

^{1*}Dpt. of Genomics Medicine, D. O.Ott Research Institute of
Obstetrics, Gynaecology, and Reproductology, Mendeleev line, 3, City,
199034, St. Petersburg, Russian Federation.

*Corresponding author(s). E-mail(s): tatiana.ev.lazareva@gmail.com;
Contributing authors: barbitoff@bk.ru; yulnasa@gmail.com;
anglotov@mail.ru;

[†]These authors contributed equally to this work.

Abstract

Loss-of-function (LoF) is commonly considered the dominant cause of human rare disease. Genetic variants that lead to a putative LoF (pLoF) are viewed as the strongest candidates during variant interpretation in rare disease. However, numerous molecular mechanisms by which such variants can lead to incomplete LoF have been described, and such low-confidence LoF variants are frequently observed in studies of healthy individuals, including recent efforts involving carrier screening and reproductive genetic testing of our biobanking cohorts. This prompted us to test if the various measures of potential LoF evasion can be used to distinguish between pathogenic and benign variants from the commonly used ClinVar database. Our results confirm that some of the tools and metrics, such as Loss-of-Function Transcript Effect Estimator (LOFTEE) tool, proportion expressed across transcripts (pext) score, and the analysis of nonsense-mediated decay (NMD) are informative for determination of suspected variant pathogenicity, while others (such as the potential for translation reinitiation or computationally predicted splice rescue) are of lesser utility, at least when simple rules are used for batch filtering of variants. Of note, thousands of pathogenic variants from ClinVar showed signs of incomplete LoF according to at least one of the metrics, highlighting challenges with automated variant filtering on the basis of these criteria.

Keywords: pLoF, PTV, ClinVar, genetic rescue

1 Introduction

Protein-truncating variants (PTVs), including nonsense mutations, splice site variants, and frame shifting indels, are of ultimate significance in human rare disease diagnostics. Standards and guidelines for genetic variant interpretation, such as those provided by the American College of Medical Genetics and Genomics (ACMG), assign typically consider a potential loss-of-function (pLoF) effect of a variant as the strongest piece of evidence for variant pathogenicity [1]. However, it has been noted a long time ago that not every variant annotated as a pLoF actually leads to a loss of gene product function, while many genuine LoF variants still have little to no effect on the phenotype [2, 3].

Failure of a pLoF allele to result in a complete LoF is attributable to a variety of factors. These include discordant effects of a variant on different transcript isoforms encoded by the same gene, potential rescue by a second variant in the same haplotype, nonsense-mediated decay (NMD) evasion, and others, as noted in many studies of pLoF variants in healthy cohorts (e.g., [3, 4]). Recognition of potential LoF evasion has led to the development of specialized tools for identification and filtering of low-confidence pLoF variants, such as Loss-Of-Function Transcript Effect Estimator (LOFTEE) [5]. Multiple studies have demonstrated high rates of incomplete LoF among variants present in healthy individuals [6, 7].

In 2023, Singer-Berk et al. have proposed a framework for classifying pLoF variants into categories depending on the strength of evidence supporting potential incomplete LoF [8]. The set of features used for filtering includes information about transcript isoform expression, potential for NMD evasion, translation reinitiation, splice variant rescue, rescue by a second variant, and others. Application of this framework to a set of 1,113 variants in autosomal recessive (AR) disease genes present in healthy individuals from the Genome Aggregation Database (gnomAD) v.2.1 data showed that up to 27.3% of pLoF evade complete LoF. Of note, the LoF-evading variants were enriched with variants that were annotated as benign in the ClinVar database [8]. These results, as well as our own experience with interpreting pLoF variants observed in healthy individuals during genetic testing, prompted us to more systematically assess the proportions of variants leading to incomplete LoF in ClinVar. Our analysis confirmed that benign pLoF variants are enriched with low-confidence LoF alleles, but many such alleles are also present among variants reported to be pathogenic.

2 Results

According to Ensemble VEP annotation of ClinVar variants, a substantial subset (n=2,581,089) are annotated as PTVs, encompassing nonsense, frameshift, and essential splice site alterations and distributed across all ACMG classification categories.

We first tested an assumption that benign variants should be overrepresented in transcript isoforms with low expression levels. To this end, we used the pext metric which quantifies the proportional expression contribution of a specific genomic region across all annotated transcripts of a gene [6]. Contrary to our expectation, this analysis revealed no clear decrease in pext values among B/LB variants compared to other groups, with VUS exhibiting significantly higher average pext scores than other categories (Figure 1a). The same pattern was observed for maximum pext

values across all tissues (Figure 1b). However, the proportion of low- (< 0.1) value of both average and maximum pext for P/LP variants (2.1%/8.7%) was lower compared to B/LB (2.5%/10.2%), but not VUS (1.8%/7.7%, Figure 1c), with the highest numbers observed for variants with conflicting interpretations (4.3%/10.6%). These results confirm an enrichment of sites with low pext among benign variants, though the magnitude of this trend was subtle.

We next limited our analysis to a set of clinically relevant transcripts isoforms from the MANE Plus annotation, and assessed the concordance of LoF variant confidence assessed by LOFTEE with ClinVar pathogenicity classification. Analysis of LOFTEE confidence categories across ClinVar classifications revealed a consistent predominance of HC pLoF variants, comprising 84–98% of variants across all clinical significance groups. The highest proportion of HC variants was observed among P/LP ones (98.5%). Conversely, B/LB variants exhibited the highest proportion of LC predictions (15.8% vs. 1.5% for P/LP), supporting the utility of LOFTEE as an effective filter for identifying annotation artifacts or biologically irrelevant events. VUS and variants with COI showed intermediate distributions, with HC variants comprising 94.1% and 93.9%, respectively. Further stratification of P/LP variants by ClinVar review status revealed no significant enrichment of LC pLoF variants among variants with a single submitter (the lowest review status category), despite these variant groups being widely perceived as enriched with misclassified variants (Figure S1). Collectively, our findings confirm that while pLoF variants designated as HC by LOFTEE represent strong candidates for pathogenicity, LC predictions warrant additional scrutiny and functional validation before being considered clinically significant.

Still, many of the variants flagged by LOFTEE correspond to well-established pathogenic alleles. For example, one such variant that we observed in carrier screening of couples planning pregnancy is a frameshift variant in EYS (c.8648.8655del; p.Thr2883Lysfs*4). This variant is classified as LOFTEE LC and shows low pext scores (< 0.1), consistent with NMD escape, despite being responsible for 5% of AR retinitis pigmentosa (OMIM #602772) cases [9].

Pathogenic variants are also expected to have greater disruptive effects on splicing, as well as a lower proportion of splice rescue events. To test this, we analyzed predicted effects on splicing using SpliceAI across ClinVar classification categories (the analysis was limited to donor and acceptor site variants). In line with the expectation, P/LP variants exhibited substantially higher delta scores of acceptor/donor loss and acceptor gain, with a notable enrichment at values approaching 1.0, indicating strong evidence of splice donor or acceptor disruption. Conversely, B/LB variants displayed lower delta scores for acceptor/donor gain and broader range of delta scores for acceptor/donor loss. VUS demonstrated the broadest range of delta scores, reflecting the heterogeneous nature of this category. Variants with COI also showed intermediate delta score distributions, suggesting that discrepancies in classification may, in part, arise from differences in the interpretation of splice-altering evidence. Notably, substantial overlap in delta scores was observed between P/LP, VUS and COI variants, emphasizing that SpliceAI scores alone cannot definitively discriminate between benign and pathogenic variants. At the same time, we did not observe any evidence that benign variants are enriched with variants that are rescued by an in-frame cryptic

splice event, with even a contrary tendency observed (Figure 1d). These results suggest that the analysis of splice rescue requires a more nuanced approach and functional validation.

Yet another expected difference between pathogenic and benign variants is the lower proportion of NMD-escaping variants among the former. Analysis of NMD escape across ClinVar classes revealed a clear gradient: P/LP variants exhibited the lowest NMD escape rate (7.9%), whereas B/LB variants demonstrated the highest (26.2%). VUS and COI showed intermediate escape rates of 23.8% and 17.5%, respectively. Further stratification by ClinVar review status among P/LP variants revealed a slight but significant enrichment of NMD-escaping variants among sites with the lowest review status (no criteria provided) (11.7%).

Finally, another mechanism of pLoF allele rescue is translation reinitiation, which occurs when translation restarts at a downstream in-frame methionine codon following a premature termination event [8]. To assess the potential for rescue via translation reinitiation, we analyzed the presence and relative position of potential initiation codons downstream of pLoF variants. Across all categories, the majority of variants (94.6–95.3%) exhibited no signs of potential rescue by reinitiation (i.e., no methionine codon in the first 10% or 25% of the CDS), and no statistically significant differences were observed among ACMG classification groups.

3 Methods

3.1 Data acquisition

The analysis was performed on ClinVar [10] data v. 2026-04-15 based on the GRCh38 human reference genome assembly. Base-level proportion expressed across transcripts (pext) data was obtained from gnomAD v.4.1.1 [6]. Isoforms from the Matched Annotation by NCBI and Ensembl (MANE) were used unless indicated otherwise [11].

3.2 Variant annotation framework

We relied on the framework proposed by Singer-Berk et al. for variant annotation and classification [8]. Variants were annotated using the Ensembl Variant Effect Predictor v.110 (VEP) [12] with the LOFTEE [13] and SpliceAI plugins [14], as well as a custom file with pext scores.

For prediction of NMD, a computational method used by Torene et al. was applied [15]. For the analysis of potential translation re-initiation, protein sequences downstream of each variant were examined, and the coordinate of the next methionine residue was calculated. Variants were flagged as potentially rescued by re-initiation if the next methionine residue was located within the first 10% or 25% of the sequence. Splice rescue events were defined as a potential in-frame donor/acceptor site gain with high SpliceAI score (> 0.5) (analysis was restricted to variants with strong canonical splice site disruption).

All intermediate data files and code used for this analysis are openly available at: <https://github.com/tanya-lazareva/clinvplof>.

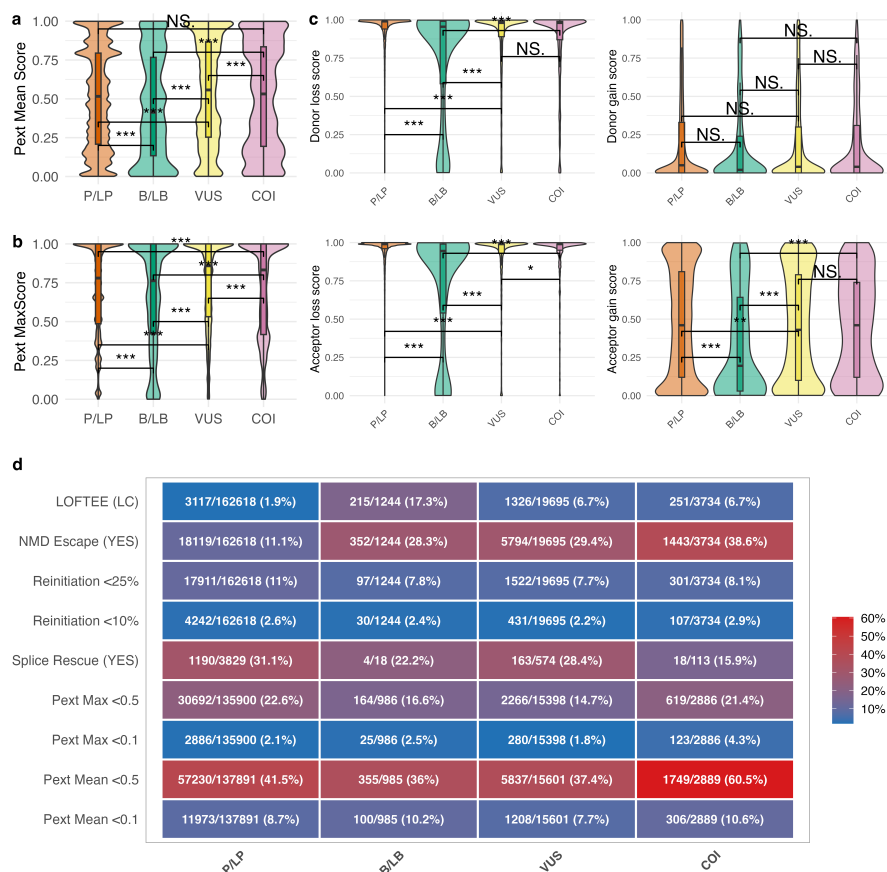


Fig. 1 Evidence of incomplete loss-of-function for protein-truncating variants across variant pathogenicity groups in ClinVar (P/LP - pathogenic/likely pathogenic; B/LB - benign/likely benign; VUS - variants of uncertain significance; COI - variants with conflicting interpretations of pathogenicity). (a) Maximum proportion expressed across transcripts (pext) score distributions across all tissues; (b) Average pext score distributions across tissues; (c) Distributions of SpliceAi scores for donor site loss, donor site gain, acceptor site loss, and acceptor site gain calculated for variants affecting donor and acceptor sites, respectively; (d) A heatmap showing proportions of variants filtered out by the indicated criteria.

4 Discussion

In this work, we have systematically evaluated the extent to which various measures predictive of incomplete LoF can be used to distinguish known pathogenic and benign variants in ClinVar. Overall, our results validate the utility of considering several widely used factors when interpreting the pathogenic potential of pLoF alleles. Thus, overrepresentation of variants annotated as low-confidence by LOFTEE among benign ones (15.8%) relative to pathogenic classifications (1.5%) demonstrates the utility of LOFTEE as a reliable filter for identifying genuine loss-of-function variants. Similarly, a clear gradient in the proportion of NMD-escaping variants across

ClinVar classifications (from pathogenic to benign) strongly supports the expectation that variants resulting in NMD-targeted transcripts are more likely to be pathogenic. The enrichment of high Splice AI delta scores in P/LP variants confirms the role of state-of-the-art splice predictors in determining the definitive impact of splice-affecting variants, even for sites that affect canonical splice sites. However, a substantial fraction of variants annotated as pathogenic still show signs of incomplete LoF according to these factors, emphasizing that: (i) existing pathogenicity classifications should be treated with caution, especially for variants with poor review status; and (ii) each of the metrics tested should not be considered as an ultimate filter that uniquely removes genuinely non-LoF variants.

Some of the metrics tested, however, turned out to be less informative, for discrimination between ACMG classification groups. Thus, we did not observe a clear enrichment of variants with higher pext scores among the P/LP variants. pLoF variants with lower pext scores that are classified as P/LP may exert their deleterious effects through disruption of critical low-expression transcripts that are essential in specific developmental or tissue contexts [6]. Additionally, the lower pext scores observed among P/LP variants may reflect the phenomenon wherein loss-of-function variants in genes with highly restricted expression patterns often lead to tissue-specific disorders. Besides pext, we also did not find significant evidence for the discriminative power of potential for translation reinitiation and cryptic 4 site-based rescue. For the former, this observation may in part be related to the fact that reinitiation was more commonly observed in individuals with milder phenotypes rather than healthy ones [16]. While our results may suggest that some mechanisms of LoF evasion may be of lower relevance on aggregate, careful consideration of possible mechanisms is required for accurate variant interpretation in rare disease, especially when evaluating genetic risks for healthy individuals or couples. Besides, our results caution against overreliance on single in silico metrics and emphasize the need for careful interpretation, particularly for variants with poor review status.

5 Limitations

While we believe that our findings emphasize the relatively higher importance of certain factors for the assessment of LoF evasion, our approach has several notable limitations that have to be considered. First, while LoF is frequently touted as the dominant disease mechanism, it is by far not the only one, with gain-of-function and dominant negative effects of variants playing a huge role (reviewed in [17]). Second, despite generally higher penetrance of pLoF variants [18], many pathogenic alleles that show signs of incomplete LoF may have lower penetrance or more subtle phenotypic effects, but not be completely silent. Last but not least, the approach we employed in this study was rather coarse-grained, as we did not perform manual review of the annotations taking into account specific aspects of a particular genomic region, as suggested by the authors of LoF curation frameworks [8].

Supplementary information. Supplementary Figure A1 provides data of incomplete loss-of-function for pathogenic and likely pathogenic protein-truncating variants across review status categories in ClinVar.

Acknowledgements. This work was supported by Multicenter Research Bioresource Collection "Family Reproductive Health" Project (No. 15.BRK.25.06) Agreement No. 075-15-2025-478 dated May 29, 2025

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding: Multicenter Research Bioresource Collection "Family Reproductive Health" Project (No. 15.BRK.25.06) Agreement No. 075-15-2025-478 dated May 29, 2025
- Conflict of interest/Competing interests: The authors declare no conflicts of interest.
- Ethics approval and consent to participate: Not applicable.
- Consent for publication: Not applicable.
- Data availability: All data and code pertinent to the analysis presented in this work are available at https://github.com/tanya-lazareva/clinv_plof (accessed on 5 of August 2026)
- Materials availability: Not applicable.
- Author contribution: Conceptualization Y.A.B. and A.S.G.; Data curation, formal analysis and writing, T.E.L. and Y.A.B.; Supervision Y.A.N. and A.S.G.; Visualization, T.E.L.. All authors have read and agreed to the published version of the manuscript.

References

- [1] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *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. *Genetics in medicine* **17**(5), 405–423 (2015)
- [2] MacArthur, D.G., Tyler-Smith, C.: Loss-of-function variants in the genomes of healthy humans. *Human molecular genetics* **19**(R2), 125–130 (2010)
- [3] MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., *et al.*: A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**(6070), 823–828 (2012)
- [4] Sulem, P., Helgason, H., Oddson, A., Stefansson, H., Gudjonsson, S.A., Zink, F., Hjartarson, E., Sigurdsson, G.T., Jonasdottir, A., Jonasdottir, A., *et al.*: Identification of a large set of rare complete human knockouts. *Nature genetics* **47**(5), 448–452 (2015)

- [5] Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.*: Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**(7616), 285–291 (2016)
- [6] Cummings, B.B., Karczewski, K.J., Kosmicki, J.A., Seaby, E.G., Watts, N.A., Singer-Berk, M., Mudge, J.M., Karjalainen, J., Satterstrom, F.K., O'Donnell-Luria, A.H., *et al.*: Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**(7809), 452–458 (2020)
- [7] Gudmundsson, S., Singer-Berk, M., Stenton, S.L., Goodrich, J.K., Wilson, M.W., Einson, J., Watts, N.A., Lappalainen, T., Rehm, H.L., *et al.*: Exploring penetrance of clinically relevant variants in over 800,000 humans from the genome aggregation database. *Nature communications* **16**(1), 9623 (2025)
- [8] Singer-Berk, M., Gudmundsson, S., Baxter, S., Seaby, E.G., England, E., Wood, J.C., Son, R.G., Watts, N.A., Karczewski, K.J., Harrison, S.M., *et al.*: Advanced variant classification framework reduces the false positive rate of predicted loss-of-function variants in population sequencing data. *The American Journal of Human Genetics* **110**(9), 1496–1508 (2023)
- [9] Placidi, G., Maltese, P.E., Savastano, M.C., D'Agostino, E., Cestroni, V., Bertelli, M., Chiurazzi, P., Maceroni, M., Minnella, A.M., Ziccardi, L., *et al.*: Retinitis pigmentosa associated with *eys* gene mutations: Disease severity staging and central retina atrophy. *Diagnostics* **13**(5), 850 (2023)
- [10] Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., Maglott, D.R.: Clinvar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research* **42**(D1), 980–985 (2014)
- [11] Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry, A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., *et al.*: A joint ncbi and embl-ebi transcript set for clinical genomics and research. *Nature* **604**(7905), 310–315 (2022)
- [12] McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., Cunningham, F.: The ensembl variant effect predictor. *Genome biology* **17**(1), 122 (2016)
- [13] Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., *et al.*: The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**(7809), 434–443 (2020)
- [14] Jaganathan, K., Panagiotopoulou, S.K., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., *et al.*: Predicting splicing from primary sequence with deep learning. *Cell* **176**(3), 535–548

(2019)

- [15] Torene, R.I., Sacoto, M.J.G., Millan, F., Zhang, Z., McGee, S., Oetjens, M., Heise, E., Chong, K., Sidlow, R., O’Grady, L., *et al.*: Systematic analysis of variants escaping nonsense-mediated decay uncovers candidate mendelian diseases. *The American Journal of Human Genetics* **111**(1), 70–81 (2024)
- [16] Gurvich, O.L., Maiti, B., Weiss, R.B., Aggarwal, G., Howard, M.T., Flanigan, K.M.: Dmd exon 1 truncating point mutations: amelioration of phenotype by alternative translation initiation in exon 6. *Human mutation* **30**(4), 633–640 (2009)
- [17] Backwell, L., Marsh, J.A.: Diverse molecular mechanisms underlying pathogenic protein mutations: beyond the loss-of-function paradigm. *Annual review of genomics and human genetics* **23**, 475–498 (2022)
- [18] Forrest, I.S., Chaudhary, K., Vy, H.M.T., Petrazzini, B.O., Bafna, S., Jordan, D.M., Rocheleau, G., Loos, R.J., Nadkarni, G.N., Cho, J.H., *et al.*: Population-based penetrance of deleterious clinical variants. *Jama* **327**(4), 350–359 (2022)

Appendix A

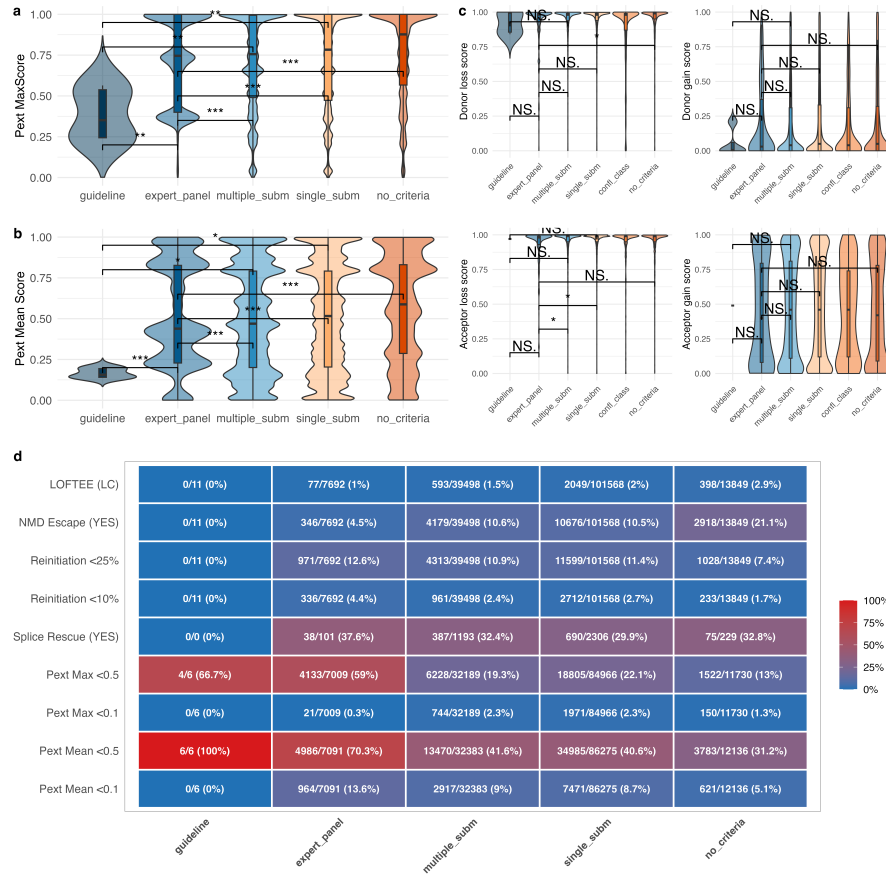


Fig. A1 Evidence of incomplete loss-of-function for pathogenic and likely pathogenic protein-truncating variants across review status categories in ClinVar. (a) Maximum proportion expressed across transcripts (pext) score distributions across all tissues; (b) Average pext score distributions across tissues; (c) Distributions of SpliceAi scores for donor site loss, donor site gain, acceptor site loss, and acceptor site gain calculated for variants affecting donor and acceptor sites, respectively; (d) A heatmap showing proportions of variants filtered out by the indicated criteria.