A Bayesian model for quantifying genomic variant evidence sufficiency in Mendelian disease

Quant Group[1] and Dylan Lawless ** [2]

[1]The quantitative omic epidemiology group.
[2]Department of Intensive Care and Neonatology, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland.

December 2, 2025

*Addresses for correspondence: Dylan.Lawless@kispi.uzh.ch

Abstract

Motivation:

Genomic variant interpretation is carried out across hospitals, research groups, commercial providers, and national programmes, yet each uses different pipelines and internal rules. This creates outputs that are difficult to compare or verify, limiting reliable exchange of genomic findings. The Qualifying Variant Evidence Standard (QV ES) provides a national minimal evidence specification, but there is no statistical method for converting these binary rule outcomes into a quantitative measure of how complete the supporting scientific information is for a given variant.

Results:

We introduce a Bayesian framework, Quant ES, that transforms QV ES rule outcomes into an uncertainty calibrated measure of evidence sufficiency. QV ES uses a reverse logic in which each rule signals whether required biological or functional evidence is available. By modelling these indicators as observations of a latent sufficiency parameter, the method estimates the proportion of required evidence domains with verifiable information. The result reflects completeness of evidence, not pathogenicity, and is independent of provider-specific algorithms or intellectual property. Comparison with the genome wide sufficiency distribution quantifies each candidate's support relative to all alternatives. The approach is efficient, reproducible, and suitable for national-scale genomic workflows.

Availability:

All source code, figures, and simulation workflows are available at the Zenodo repository (pending). The framework is released under the MIT licence and will be supported for at least two years following publication.

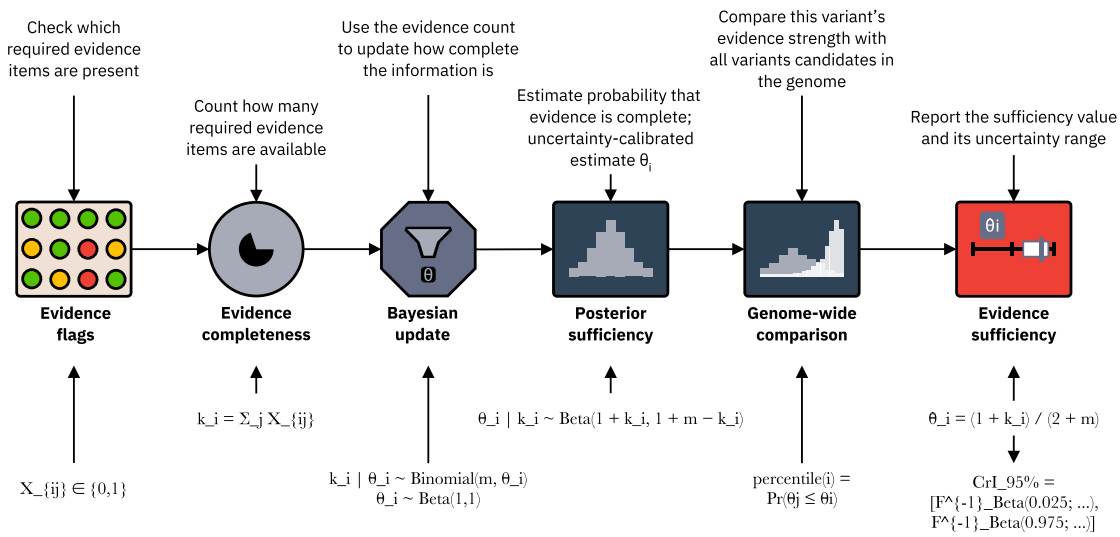


Figure 1: **A graphical overview of Quant Evidence Standard**. The method provides a clear, quantitative measure of how well each variant is supported by the scientific evidence required for Mendelian disease interpretation. Starting from the binary evidence matrix defined by the Qualifying Variant Evidence Standard, Quant ES uses a principled statistical model to measure the completeness of the available evidence. This separates WGS analysis outputs from downstream interpretation and delivers a universal, uncertainty calibrated assessment of evidential credibility.

Acronyms

GIAB Genome in a Bottle 8

HPO Human Phenotype Ontology 8

JSON JavaScript Object Notation 7

QV Qualifying Variant 8

QV ES Qualifying Variant Evidence Standard 5

QV ES interpret Qualifying Variant Evidence Standard interpretation rules 6

Quant ES Quantitative Evidence Sufficiency model 5

WGS Whole Genome Sequencing 8

YAML YAML Ain't Markup Language 7

1 Introduction

Genome interpretation in Mendelian disease depends critically on the availability of biological and functional evidence. Key elements include variant–gene validity, inheritance consistency, disease mechanism fit, population frequency, protein constraint, and phenotype matching. In practice, the completeness of these evidence categories varies widely across the genome. Some variants benefit from extensive prior characterisation, while many others fall into poorly annotated genomic regions where essential information is sparse or absent.

The QV Evidence Standard (QV-ES) defines a fixed national-scale registered rule set for capturing the minimal evidence required for clinical interpretation (1; 2). Genome analysis providers evaluate these rules during secondary analysis and return an evidence matrix in which each rule is recorded as TRUE, FALSE, or NA. Depending on the clinical context, these outcomes are then mapped to a binary present (1) or missing (0) indicator for downstream statistical analysis. For example, a homozygous loss-of-function variant in *CFTR* causing cystic fibrosis does not require confirmed parental genotypes to the same extent as a compound heterozygous configuration, and the corresponding inheritance flags are interpreted accordingly. Currently, however, the field lacks a principled method for converting this heterogeneous collection of rule outcomes into a quantitative and interpretable measure of evidence sufficiency. A statistical framework is therefore needed to provide a confidence interval that captures both the completeness of evidence for the candidate and the extent to which competing explanations are exist.

We introduce Quantitative Evidence Sufficiency model (Quant ES), a Bayesian framework (3; 4) that treats the m Qualifying Variant Evidence Standard (QV ES) rule outcomes as a vector of binary observations and infers, for each variant, a latent evidence sufficiency parameter θ_i that represents the expected proportion of required evidence domains for which reliable information is available. This parameter reflects the completeness of verifiable biological information rather than pathogenicity. By estimating the empirical distribution of θ_i across all observed variants, the framework places each candidate within the wider landscape of evidence sufficiency. This

situates each candidate within the observed evidence landscape without modelling causality (5).

By defining a shared evidence layer that is independent of any specific pipeline or institutional protocol, Quant ES provides a common interface through which genomic results can be compared and verified across clinical, research, and industry settings.

This approach is implemented in an open-source R package designed to operate directly on the evidence matrices returned by external analysis providers according to the QV ES standard. It requires no access to upstream pipeline code and no disclosure of proprietary algorithms. The method is computationally efficient, scalable to millions of variants, and suitable for national-level reproducibility mandates.

2 Methods

2.1 Approach

In general, current genomic workflows combine upstream variant calling, reference annotation, and downstream interpretation. Each institution typically maintains its own combination of tools and internal rules, which makes variant reports difficult to compare or verify across settings (6–14). The QV Evidence Standard defines a shared evidence layer based on a registered flag rule set evaluated from public reference data (examples in **Box 2-3**). These flag outcomes are then transformed into a standardised binary evidence matrix using a second registered interpretation rule set (example in **Table S1**). Our Bayesian framework operates on this binary matrix to support consistent interpretation across providers.

For each variant i, the registered Qualifying Variant Evidence Standard interpretation rules (QV ES interpret) rule set converts the raw flags into a binary vector of length m corresponding to evidence present (1) or missing (0):

$$X_{ij} = \begin{cases} 1 & \text{if the evidence requirement for rule } j \text{ is satisfied,} \\ 0 & \text{otherwise.} \end{cases}$$

The total evidence count for variant i is $k_i = \sum_{j=1}^{m} X_{ij}$, and the empirical proportion of satisfied rules is $\hat{p}_i = k_i/m$. This observable captures how much verifiable information supports any specific variant.

The sufficiency parameter θ_i represents the probability that a randomly selected evidence rule is satisfied for variant i. The model assumes $k_i \mid \theta_i \sim \text{Binomial}(m, \theta_i), \quad \theta_i \sim \text{Beta}(1, 1)$. By conjugacy, $\theta_i \mid k_i \sim \text{Beta}(1 + k_i,\ 1 + m - k_i)$. The posterior mean and credible interval (4; 15) provide an uncertainty calibrated description of evidence completeness for each variant (3; 16).

We analyse each variant's θ_i alongside the empirical mixture of all θ_i values (17) across the genome to characterise the background availability of biological and functional knowledge.

Monte Carlo draws from each posterior θ_i give a global set of samples from which we estimate the genome wide mean, median, and credible interval. These illustrate typical evidence availability in the case being analysed and form the reference set against which candidate variants are compared.

Given a list of candidate variants supplied by the provider, we compute the percentile of each θ_i within the global distribution as $\text{percentile}(i) = \Pr\left(\theta_j \leq \theta_i\right)$, with j indexing all variants. Percentiles quantify relative support across all evaluated variants.

High values indicate stronger support; low values indicate that many alternatives have comparable or better profiles.

2.2 Implementation

The method is implemented in an R package, `quantES`, which ingests the QV ES rule specification (YAML Ain't Markup Language (YAML) or JavaScript Object Notation (JSON)) (2), the provider-generated evidence matrix (variant-by-rule, encoded as 0/1), and optional lists of provider-prioritised variants. The package then produces per-variant posterior summaries (θ_i, CrI), global empirical evidence distributions, candidate percentiles, reproducible visual summaries, a structured inter-

pretation suitable for reporting such as for accredited clinical genetics. No pipeline internals, genotype likelihoods, or proprietary logic are required. All analyses depend solely on the evidence flags defined by the national QV ES standard.

2.3 Usage in a WGS validation study with GIAB and Exomiser

We evaluated the Qualifying Variant (QV) and Quant ES framework using whole genome sequencing from the Genome in a Bottle (GIAB) Trio (HG005–HG007, PRJNA200694, GRCh38 v4.2.1) (18). All processing parameters were defined in a publicly registered external QV YAML specification. Quality control and study filters were applied to illustrate a typical clinical Whole Genome Sequencing (WGS) pipeline. Site level thresholds on QUAL and INFO/DP were applied using BCFtools, complemented by per sample thresholds on FORMAT/DP and FORMAT/GQ with exclusion of missing genotypes. Composite filters required either all trio samples or at least one sample to pass the criteria. The filtered trio VCF was analysed with Exomiser (default configuration) using the family `.ped` file and without Human Phenotype Ontology (HPO) terms.

Exomiser outputs were then processed with the registered QV ES flag rule set to generate raw per variant evidence flags (TRUE, FALSE, or NA). A second registered QV ES interpret rule set converted these raw flags into the binary evidence matrix required by Quant ES, with values indicating evidence present (1) or missing (0). All QV ES rule sets used in this workflow are provided in YAML or JSON format (2). Quant ES was then run on the binary evidence matrix using default settings to generate per variant credible intervals and genome wide evidence summaries.

3 Results

3.1 Default Quant ES output and interpretation

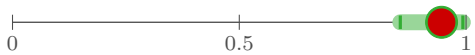
The report in **Box 1** illustrates the primary output generated by Quant ES. It summarises how completely a candidate variant is supported by established scientific evidence, independent of the analysis provider's internal ranking methods or proprietary algorithms. When paired with the recognised and database-registered QV ES flag set, this output provides a consistent evidential metric that can be used across clinical genetics, research workflows, and external analysis services.

Box 1: Example of QV ES evidence sufficiency output

QV-ES: `qv_variant_interpretation.yaml`
QV-ES registered MD5: 6343b82a254edc709f4dc766fe056253

Clinical scenario: autosomal recessive cystic fibrosis with a primary candidate variant in *CFTR*.
Candidate variant: `NM_000492.4:c.1521_1523delCTT (p.Phe508del)`.

Posterior evidence sufficiency: 0.946 (scale 0–1)
Credible interval: 0.855 to 0.993 (scale 0–1)
Percentile: 100 (scale 0–100)
Evidence scale:



Further information
Higher values indicate that most required evidence elements are available; lower values reflect sparse or incomplete support. For this candidate, almost all required domains are satisfied.
Total evaluated variants: 100
Global evidence sufficiency: CrI 0.475 (0.082 to 0.951).
This summarises the overall availability of scientific and clinical knowledge for this genome and is independent of the provider's internal methods. Across all variants, the mean θ is 0.475 and the median is 0.459, reflecting the typical level of interpretable evidence in this case.

CrI evidence summaries for the top ten candidates:
1. variant 94: 0.946 (0.855 to 0.993), percentile 100
2. variant 99: 0.946 (0.855 to 0.993), percentile 100
3. variant 35: 0.919 (0.813 to 0.982), percentile 98
4. ...

3.2 Validation in a WGS trio study with GIAB and Exomiser

To evaluate the practical use of the QV ES and Quant ES framework within an established whole genome analysis workflow, we applied it to the GIAB Trio (HG005–

HG007, GRCh38 v4.2.1) analysed with Exomiser (19). The Exomiser run produced its standard annotated variant list and candidate gene predictions, which we processed using a publicly registered QV YAML specification defining the required evidence rules. The first step was the generation of raw QV ES evidence flags for every variant. **Figure S4** shows these rule evaluations (TRUE, FALSE, or NA), computed directly from the registered flag rule set. The second step applied the registered QV ES interpretation rules to produce the binary matrix required for Quant ES. **Figure S5** shows the resulting interpreted matrix (1 = evidence present, 0 = evidence missing), which is structurally identical to the matrices used in the simulation study. Together, these two figures show that the Exomiser outputs can be converted into the standardised QV ES evidence representation required by Quant ES.

Quant ES then computed posterior evidence sufficiency for each variant. Across the trio, most variants exhibited intermediate θ values with wide credible intervals, reflecting variation in available annotation across the genome. A smaller subset showed high evidence sufficiency ($\theta > 0.8$), typically corresponding to well characterised disease genes. Percentile values quantified the relative standing of each variant among all evaluated alternatives. Variants prioritised by Exomiser also achieved high Quant ES percentiles. Variants with low percentiles were supported by less complete evidence. This behaviour is complementary to established variant effect predictors and interpretation tools (19–21).

Figure S2 summarises the interpreted evidence matrix, illustrating the heterogeneous availability of case specific and external reference evidence. **Figure S3** shows the observed evidence proportions and the corresponding posterior 95% credible intervals for each θ_i. **Figure 2** displays the genome wide mixture distribution of θ with overlaid density curves for the top Exomiser candidates: the variant with the strongest evidence sufficiency is marked in red, and the two Exomiser rank 1 phenotype driven candidates are shown in grey. The variant with the strongest evidence sufficiency was *RYR2*:ENST00000366574.7:c.6493dup:p.(Leu2165Profs*44) with AD MOI, with $\theta = 0.808$ (95% CrI 0.639 to 0.932) and a percentile of 100. This candidate emerged from 123 evaluated variants derived from an initial WGS call set of 5,188,814 variants.

Candidates that rise above the genome wide mixture are supported by substantially more interpretable evidence; Quant ES therefore provides a robust measure of how strongly the available evidence supports the causal hypotheses proposed by tools such as Exomiser.

Taken together, the results show that Quant ES provides a reproducible, transparent, and pipeline independent measure of evidence availability. The framework separates the evidential layer from the internal logic of analysis providers, enabling consistent interpretation, independent verification, and a shared evidential standard suitable for clinical diagnostics, research cohorts, and commercial genomic services.

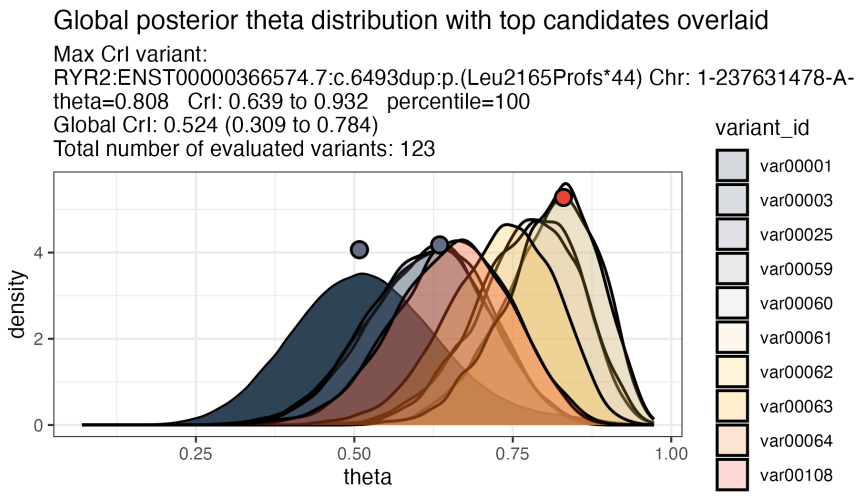


Figure 2: **Global evidence sufficiency distribution with candidate overlay (GIAB WGS trio).** The grey background shows the genome wide mixture distribution of θ across all evaluated variants. Overlaid lines represent the top Exomiser prioritised candidates. The red dot marks the variant with the strongest evidence sufficiency CrI, while the grey dots indicate the two Exomiser rank 1 predictions based on phenotype. Candidates that rise above the genome wide mixture display substantially more interpretative evidence, though this does not by itself indicate causality.

4 Conclusion

We presented a Bayesian framework that quantifies how completely each variant in a genome is supported by the scientific evidence required for Mendelian interpretation. By operating directly on the QV ES evidence matrix and separating biological knowledge availability from provider-specific analysis pipelines, the method provides a reproducible and transparent evidential layer suitable for clinical diagnostics, research, and large-scale genomic workflows. The posterior sufficiency parameter and its percentile within the genome-wide distribution offer a clear summary of both variant-level evidence completeness and the standing of each candidate among observable alternatives. This approach supports consistent interpretation, improves comparability across providers, and forms a practical foundation for national and international interoperability in genomic medicine.

Funding

This project was supported through the grant Swiss National Science Foundation 320030_201060, and NDS-2021-911 (SwissPedHealth) from the Swiss Personalized Health Network and the Strategic Focal Area 'Personalized Health and Related Technologies' of the ETH Domain (Swiss Federal Institutes of Technology).

Acknowledgements

Acknowledgements We would like to thank all the patients and families who have been providing advice on SwissPedHealth and its projects, as well as the clinical and research teams at the participating institutions.

Data availability

The data used in this manuscript is derived from open sources which are cited in methods. The data generated is available from the Zenodo repository (pending). The resulting data are also reproducible using the code repository (pending). WGS GIAB trio data (HG005–HG007, PRJNA200694, GRCh38 v4.2.1) from Wagner et al. (18) with the benchmark VCF files for small variants from GRCh38 at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/.

Contributions

DL designed the work and contributed to the manuscript.

Competing interests

The authors declare no competing interests.

Ethics statement

The projects were approved by the respective ethics committees of all participating centers (Cantonal Ethics Committee Bern, approval number KEK-029/11) and the study was conducted in accordance with the Declaration of Helsinki.

References

[1] Swiss Genomics Association. Swiss genomics association consensus guideline for evidence based genomic variant interpretation in mendelian disease. Draft guideline, Swiss Genomics Association, 2025. URL https://www.

swissgenomicsassociation.ch/releases/. Draft in progress. Final publication anticipated.

[2] Dylan Lawless, Ali Saadat, Mariam Ait Oumelloul, Simon Boutry, Veronika Stadler, Sabine Österle, Jan Armida, David Haerry, D. Sean Froese, Luregn J. Schlapbach, and Jacques Fellay. Application of qualifying variants for genomic analysis. *medrxiv*, May 2025. doi: 10.1101/2025.05.09.25324975.

[3] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics. Wiley, 1 edition, May 1994. ISBN 978-0-471-49464-5 978-0-470-31687-0. doi: 10.1002/9780470316870.

[4] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 0 edition, June 1995. ISBN 978-1-135-43941-5. doi: 10.1201/9780429258411.

[5] A. P. Dawid. Causal Inference without Counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, June 2000. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2000.10474210.

[6] Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W Grody, Madhuri Hegde, Elaine Lyon, Elaine Spector, 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.

[7] Marilyn M Li, Michael Datto, Eric J Duncavage, Shashikant Kulkarni, Neal I Lindeman, Somak Roy, Apostolia M Tsimberidou, Cindy L Vnencak-Jones, Daynna J Wolff, Anas Younes, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the association for molecular pathology, american society of clinical oncology, and college of american pathologists. *The Journal of molecular diagnostics*, 19(1):4–23, 2017.

[8] Quan Li and Kai Wang. Intervar: clinical interpretation of genetic variants by the 2015 acmg-amp guidelines. *The American Journal of Human Genetics*, 100 (2):267–280, 2017.

[9] Erin Rooney Riggs, Erica F Andersen, Athena M Cherry, Sibel Kantarci, Hutton Kearney, Ankita Patel, Gordana Raca, Deborah I Ritter, Sarah T South, Erik C Thorland, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the american college of medical genetics and genomics (acmge and the clinical genome resource (clingen). *Genetics in Medicine*, 22(2):245–257, 2020.

[10] Sean V Tavtigian, Steven M Harrison, Kenneth M Boucher, and Leslie G Biesecker. Fitting a naturally scaled point system to the acmg/amp variant classification guidelines. *Human mutation*, 41(10):1734–1737, 2020.

[11] Brent S Pedersen, Joe M Brown, Harriet Dashnow, Amelia D Wallace, Matt Velinder, Martin Tristani-Firouzi, Joshua D Schiffman, Tatiana Tvrdik, Rong Mao, D Hunter Best, et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease. *NPJ Genomic Medicine*, 6(1): 1–8, 2021.

[12] Carl A Anderson, Fredrik H Pettersson, Geraldine M Clarke, Lon R Cardon, Andrew P Morris, and Krina T Zondervan. Data quality control in genetic case-control association studies. *Nature protocols*, 5(9):1564–1573, 2010. URL https://doi.org/10.1038/nprot.2010.116.

[13] Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina De Vries, Yukinori Okada, Alicia R Martin, Hilary C Martin, Tuuli Lappalainen, and Danielle Posthuma. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1):59, 2021. URL https://doi.org/10.1038/s43586-021-00056-9.

[14] Brent S Pedersen and Aaron R Quinlan. Vcfexpress: flexible, rapid user-expressions to filter and format VCFs. *Bioinformatics*, 41(3): btaf097, March 2025. ISSN 1367-4811. doi: 10.1093/bioinformatics/

btaf097. URL https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btaf097/8051444.

[15] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 1 edition, April 2003. ISBN 978-0-521-59271-0 978-0-511-79042-3. doi: 10.1017/CBO9780511790423.

[16] Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley, 1 edition, August 2005. ISBN 978-0-471-27246-5 978-0-471-71581-8. doi: 10.1002/0471715816.

[17] Bradley Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, 1 edition, August 2010. ISBN 978-0-521-19249-1 978-0-511-76136-2 978-1-107-61967-8. doi: 10.1017/CBO9780511761362.

[18] Justin Wagner, Nathan D. Olson, Lindsay Harris, Ziad Khan, Jesse Farek, Medhat Mahmoud, Ana Stankovic, Vladimir Kovacevic, Byunggil Yoo, Neil Miller, Jeffrey A. Rosenfeld, Bohan Ni, Samantha Zarate, Melanie Kirsche, Sergey Aganezov, Michael C. Schatz, Giuseppe Narzisi, Marta Byrska-Bishop, Wayne Clarke, Uday S. Evani, Charles Markello, Kishwar Shafin, Xin Zhou, Arend Sidow, Vikas Bansal, Peter Ebert, Tobias Marschall, Peter Lansdorp, Vincent Hanlon, Carl-Adam Mattsson, Alvaro Martinez Barrio, Ian T. Fiddes, Chunlin Xiao, Arkarachai Fungtammasan, Chen-Shan Chin, Aaron M. Wenger, William J. Rowell, Fritz J. Sedlazeck, Andrew Carroll, Marc Salit, and Justin M. Zook. Benchmarking challenging small variants with linked and long reads. *Cell Genomics*, 2(5):100128, May 2022. ISSN 2666979X. doi: 10.1016/j.xgen.2022.100128.

[19] Valentina Cipriani, Nikolas Pontikos, Gavin Arno, Panagiotis I. Sergouniotis, Eva Lenassi, Penpitcha Thawong, Daniel Danis, Michel Michaelides, Andrew R. Webster, Anthony T. Moore, Peter N. Robinson, Julius O.B. Jacobsen, and Damian Smedley. An Improved Phenotype-Driven Tool for Rare Mendelian Vari-

ant Prioritization: Benchmarking Exomiser on Real Patient Whole-Exome Data. *Genes*, 11(4):460, April 2020. ISSN 2073-4425. doi: 10.3390/genes11040460.

[20] Cristian Riccio, Max L. Jansen, Linlin Guo, and Andreas Ziegler. Variant effect predictors: A systematic review and practical guide. *Human Genetics*, 143(5): 625–634, May 2024. ISSN 1432-1203. doi: 10.1007/s00439-024-02670-5.

[21] Manuel Holtgrewe, Oliver Stolpe, Mikko Nieminen, Stefan Mundlos, Alexej Knaus, Uwe Kornak, Dominik Seelow, Lara Segebrecht, Malte Spielmann, Björn Fischer-Zirnsak, Felix Boschann, Ute Scholl, Nadja Ehmke, and Dieter Beule. VarFish: Comprehensive DNA variant analysis for diagnostics and research. *Nucleic Acids Research*, 48(W1):W162–W169, July 2020. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkaa241.

[22] Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435):1343–1370, 1996.

[23] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, September 2009. ISBN 978-0-511-80316-1 978-0-521-89560-6 978-0-521-74919-0. doi: 10.1017/CBO9780511803161.

[24] Michael S. Hamada, Alyson G. Wilson, C. Shane Reese, and Harry F. Martz. *Bayesian Reliability*. Springer Series in Statistics. Springer New York, New York, NY, 2008. ISBN 978-0-387-77948-5 978-0-387-77950-8. doi: 10.1007/978-0-387-77950-8.

5 Supplemental

5.1 QV ES evidence figure

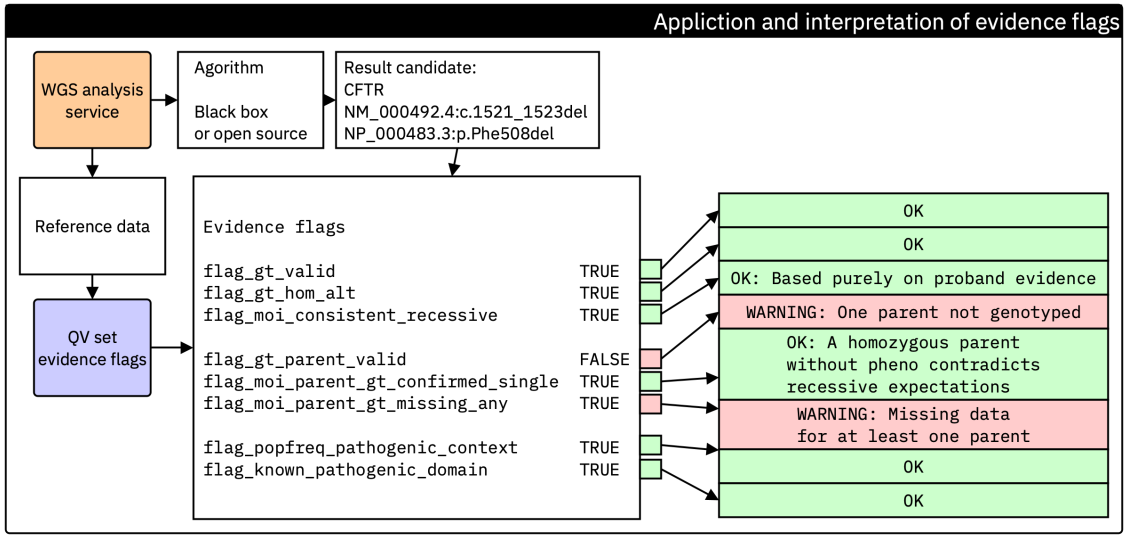


Figure S1: **Generation and interpretation of QV ES evidence flags.** After secondary analysis, each variant is evaluated against the registered QV ES flag rule set. Each rule checks whether a required piece of biological or functional evidence is present. The results form a set of raw evidence flags (TRUE, FALSE, or NA).

5.2 Validation figures

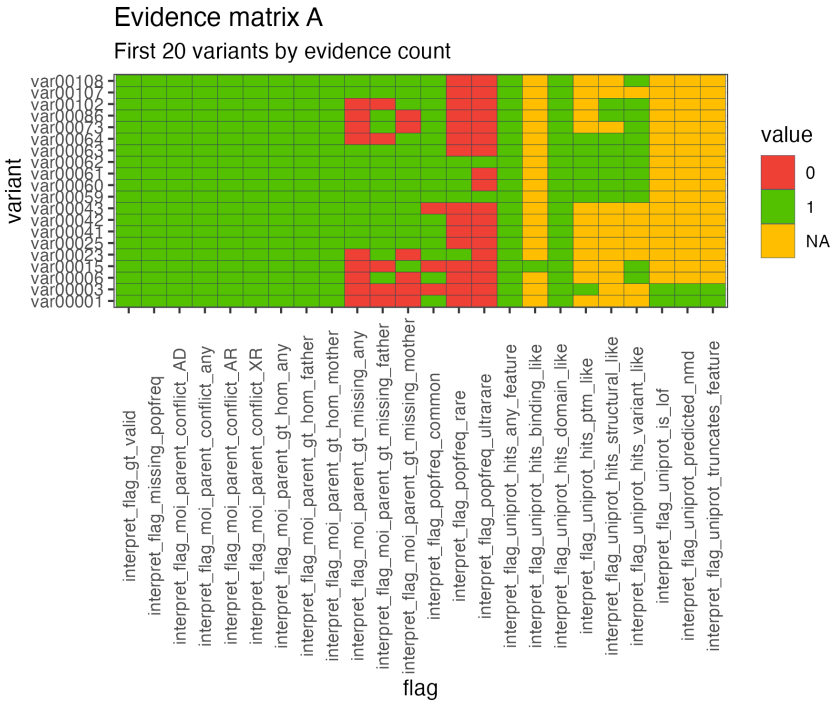


Figure S2: **QV-ES evidence interpretation matrix (GIAB WGS trio).** Each row corresponds to a variant and each column corresponds to a QV ES rule. Cells encode the derived binary evidence outcome (1 = evidence present, 0 = evidence missing) after converting the Exomiser outputs using the QV ES flag set. The heterogeneity across rules illustrates the variability in available scientific evidence in a real whole genome analysis. We illustrate NA evidence in yellow indicating that this was specifically checked and no current evidence exists.

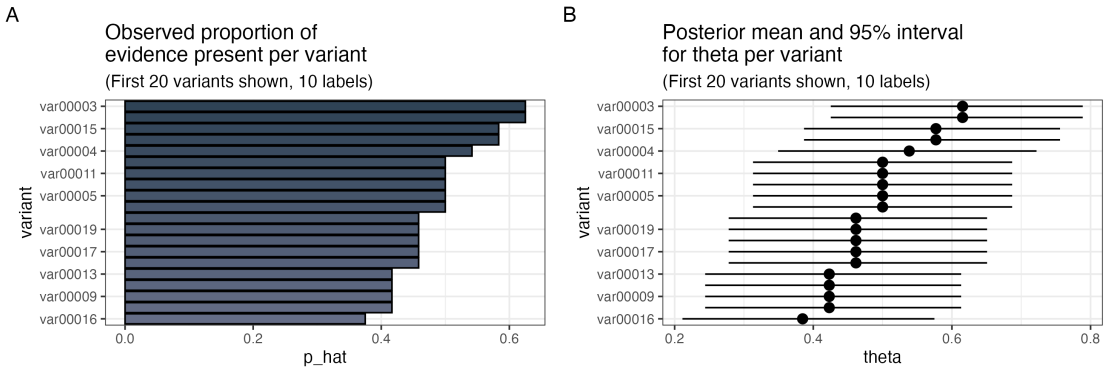


Figure S3: **Evidence sufficiency across variants (GIAB WGS trio). (A)** Observed proportion of satisfied evidence rules. Bars show $\hat{p}_i = k_i/m$ for the first 20 variants, illustrating the completeness of biological and functional evidence available for each locus. **(B)** Posterior means and 95% credible intervals for θ_i. The Bayesian model converts evidence counts into uncertainty calibrated estimates of evidence sufficiency under real clinical analysis conditions.

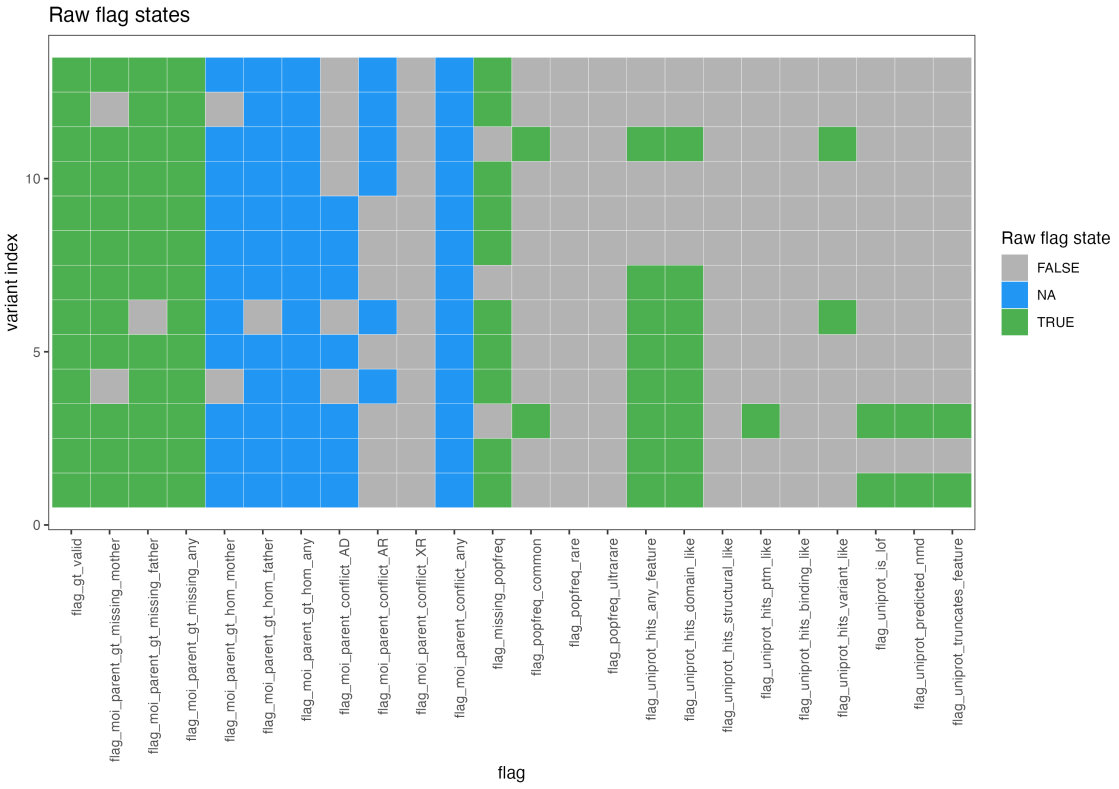


Figure S4: **Raw QV ES evidence flags from Exomiser outputs (GIAB WGS trio).** Rows represent variants and columns correspond to QV ES rules. Cells show the direct flag outcomes returned by the QV ES evaluation: TRUE (evidence present), FALSE (evidence absent), and NA (evidence checked but not currently available). These raw flags provide a standardised and provider independent summary of the evidence underlying each variant before statistical interpretation.

Figure S5: **Binary QV ES evidence interpretation matrix (GIAB WGS trio).** The registered QV ES interpret rules convert the raw flags into a binary matrix suitable for Quant ES analysis. Cells indicate evidence present (1) or evidence missing (0), following the clinical interpretation logic of the QV ES standard. The resulting matrix forms the input for Quant ES and enables consistent evidence comparison across all evaluated variants.

5.3 Example QV ES flags and interpretation rules

Box 2: Example QV ES YAML (parental genotype completeness)

```yaml
criteria:
  flag_moi_parent_gt_missing_mother:
    conditions:
      - field: mother_geno
        operator: '=='
        value: NA
    description: Mother genotype missing
  flag_moi_parent_gt_missing_any:
    conditions:
      - field: mother_geno
        operator: '=='
        value: NA
      - field: father_geno
        operator: '=='
        value: NA
    description: At least one parent genotype missing
    logic: or
notes:
  - flag_moi_parent_gt_missing_mother="means the mother's genotype
     is not available, so inheritance checks relying on her cannot
     be fully assessed."
  - flag_moi_parent_gt_missing_any="indicates that at least one parent
    genotype is missing, which reduces certainty for inheritance
    interpretation."
```

Box 3: Example QV ES YAML (gnomAD quality flags)

```yaml
criteria:
  gnomad_flags:
    conditions:
      - field: gnomad_flag
        operator: '=='
        value: NA
      - field: gnomad_flag
        operator: in
        value: AC0,AS-VQSR,InbreedingCoeff,RF,Not in exomes,
                    Not in genomes,  No data,Discrepant Frequencies,
                    CHIP,Monoallelic,Only heterozygous, MNV,LCR,
                    LC pLoF,pLoF Flag,NC Transcript,SEGDUP,
                    Common low heteroplasmy
    description: Variant has no gnomAD flags or will be reviewed
                    in report generation
    logic: or
notes:
  - gnomad: gnomAD flags are retained for review; presence of any
                    flag requires explicit mention or justification
                    in final causal report.
                    For information see:
                    https://gnomad.broadinstitute.org/help
```

Table S1: **QV ES interpret rules for parental homozygous genotype checks**. Raw flags of TRUE, FALSE, and NA outcomes are mapped to evidence present (1) or missing (0) according to whether the parental genotype supports or contradicts the proband's Mendelian configuration. Counterfactual reasoning is not part of this evidence layer.

Flag	**Raw value**	**Meaning**	**Interpretation**
flag moi parent gt hom mother	TRUE	Mother inconsistent homozygous reference	0
flag moi parent gt hom mother	FALSE	Mother genotype valid	1
flag moi parent gt hom mother	NA	Not recorded	0
flag moi parent gt hom father	TRUE	Father inconsistent homozygous reference	0
flag moi parent gt hom father	FALSE	Father genotype valid	1
flag moi parent gt hom father	NA	Not recorded	0
flag moi parent gt hom any	TRUE	Any inconsistent parental homozygous genotype	0
flag moi parent gt hom any	FALSE	No conflict	1
flag moi parent gt hom any	NA	Not recorded	0

6 Statistical basis

6.1 Introduction

Clinical variant interpretation depends on a diverse set of biological, functional, and population-level evidence sources. In many real cases, the completeness of this evidence varies substantially across the genome. Some gene–variant combinations have abundant supporting information from population datasets, disease databases, protein annotations, segregation information, and phenotypic fit. Others have limited or absent information in one or more of these domains. This variability affects the confidence with which any proposed causal variant can be evaluated.

The QV ES framework defines a finite set of evidence rules. Each rule corresponds to a required piece of knowledge that should be verifiable from established datasets or computational annotation systems. For a given variant, each rule is either satisfied or not satisfied. This structure lends itself naturally to a binomial evidence availability model. The analytical question is how to convert a large matrix of binary rule evaluations into a coherent probabilistic quantification of how completely each variant is supported by available scientific evidence.

We present a Bayesian approach based on conjugate Beta–Binomial inference to estimate, for each variant i, a posterior evidence sufficiency probability θ_i. This represents the probability that a randomly selected QV ES rule would be satisfied for that variant. In addition, we compare each candidate variant to the empirical distribution of θ_i across all variants in the genome. This comparison quantifies how well supported the candidate is relative to all alternative plausible explanations (22).

The method is simple, computationally efficient, and transparent. It can be applied to millions of variants in real genome analysis workflows. It does not depend on external variant scoring systems, and it separates biological knowledge availability from the analysis provider's heuristic ranking. The approach therefore produces an auditable and reproducible layer of statistical interpretation that complements clinical genetics practice.

6.2 Evidence Structure and Data Model

For each variant i we consider a set of m QV ES rules:

$$R_1, R_2, \ldots, R_m.$$

Each rule corresponds to a required evidence type, such as genotype validity, population frequency context, gene disease validity, functional domain support, phenotype consistency, or absence from controls.

For a variant i, each rule is recorded as:

$$X_{ij} = \begin{cases} 1 & \text{if rule } j \text{ is satisfied for variant } i, \\ 0 & \text{otherwise.} \end{cases}$$

This yields an evidence vector:

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{im}).$$

The total number of satisfied rules is

$$k_i = \sum_{j=1}^{m} X_{ij}.$$

This is the observed evidence count for variant i. It reflects neither pathogenicity nor penetrance. It reflects how much verifiable biological and functional knowledge is available from scientific reference sources.

6.3 Variant-Level Evidence Sufficiency Model

Likelihood

We model the observed evidence count k_i as a binomial quantity:

$$k_i \mid \theta_i \sim \text{Binomial}(m, \theta_i),$$

where $\theta_i \in [0, 1]$ denotes the probability that a randomly selected QV ES rule is satisfied for variant i. This is an evidence sufficiency parameter, not a disease probability.

Prior and Posterior

We use a Beta prior:

$$\theta_i \sim \text{Beta}(a, b),$$

with a default non-informative prior $a = 1$, $b = 1$ (22).

By conjugacy, the posterior distribution is:

$$\theta_i \mid k_i, m \sim \text{Beta}(a + k_i,\ b + m - k_i).$$

Using the uniform prior yields:

$$\theta_i \mid k_i, m \sim \text{Beta}(1 + k_i,\ 1 + m - k_i).$$

The posterior mean is:

$$\mathbb{E}[\theta_i \mid k_i] = \frac{1 + k_i}{2 + m}.$$

The central 95% credible interval is:

$$\theta_i^{\mathrm{CrI}} = \left[\mathrm{Beta}^{-1}(0.025; 1 + k_i,\ 1 + m - k_i),\ \mathrm{Beta}^{-1}(0.975; 1 + k_i,\ 1 + m - k_i)\right].$$

This gives a natural, interpretable measure of completeness of evidence.

6.4 Global Evidence Distribution

To understand the context of a candidate variant, we also study the distribution of θ_i across all variants in the genome. This global distribution describes the landscape of biological information present in the data and reflects how much of the genome is well supported by established scientific resources. For each variant, we draw Monte Carlo samples:

$$\theta_i^{(s)} \sim \mathrm{Beta}(1 + k_i,\ 1 + m - k_i), \qquad s = 1, \ldots, S.$$

Pooling across all variants gives a dataset reflecting the global availability of evidence. This forms the statistical backdrop for variant prioritisation. Important genome-level summaries include the global posterior mean, global median, and global 95% credible interval. These quantities summarise the completeness of scientific knowledge across the entire genome.

6.5 Candidate Comparison and Counterfactual Reasoning

Variant interpretation is inherently comparative. A variant is a plausible explanation for a phenotype only in the context of alternative explanations. The central question is not only how complete the evidence is for variant i but also how this compares with all other variants that could, in principle, explain the phenotype.

Our method addresses this by calculating the empirical percentile of each candidate as $\mathrm{percentile}(i) = \Pr\left(\theta_j \leq \theta_i \mid j = 1, \ldots, N\right)$, where N is the total number of evaluated variants.

This percentile gives a clear and intuitive quantity; a candidate in the 99th percentile has stronger evidence support than 99% of all evaluated variants in the genome while a candidate in the 10th percentile is weakly supported compared with most alternatives.

If the evidence sources are highly correlated with pathogenicity prediction, this percentile might be considered a direct diagnostic of counterfactual plausibility (5; 23). If many alternative variants have as much or more evidence support, the proposed candidate is less credible as an explanation.

6.6 Interpretation

Interpretation of variant level uncertainty and the role of alternative variants

A clear way to understand Quant ES is to compare two evaluated variants. Suppose variant A satisfies most QV ES rules, while variant B has many missing evidence indicators. In Quant ES each variant is analysed independently through its own evidence count

$$k_i = \sum_{j=1}^{m} X_{ij}, \qquad k_i \mid \theta_i \sim \text{Binomial}(m, \theta_i), \qquad \theta_i \mid k_i \sim \text{Beta}(1+k_i,\ 1+m-k_i).$$

The parameter θ_i is a marginal evidence sufficiency quantity that describes the expected proportion of QV ES domains supported for variant v_i. The credible interval for θ_A depends only on k_A and m. It does not broaden because variant B has sparse evidence. This is because Quant ES summarises the completeness of evidence for A under the current QV ES rules and reference data. It does not estimate the probability that A is the true causal variant among all alternatives.

Variant B affects A only through the comparative percentile. Quant ES computes

$$\text{percentile}(A) = \Pr_{v_j \in \mathcal{V}}\big(\theta_j \le \theta_A\big),$$

which measures how well supported A is relative to other variants that the workflow has exposed. If many alternatives have low θ_j values, A receives a high percentile. This reflects strong support for A within the evaluated set but does not rule out explanations outside that set.

One might ask whether missing evidence for B should contribute to global uncertainty that weakens confidence in A. In Bayesian terms this would alter the target of inference. Quant ES estimates a local evidential property for each variant, namely the completeness of the evidence available for that variant. It does not pool uncertainty across variants. Allowing missing evidence for B to widen the credible interval for θ_A would mix this local question with a different question about whether an understudied variant might become a stronger explanation if new data appear. That broader reasoning requires a model of causal structure across variants and lies outside the scope of the QV ES evidence layer.

Quant ES therefore distinguishes two kinds of uncertainty. The posterior for θ_i reflects uncertainty about the completeness of evidence for variant v_i. The percentile reflects how strongly supported the variant is relative to the alternatives visible in the evaluated set. Questions about unobserved variants or hypothetical future evidence remain outside the remit of Quant ES, which provides a transparent and standardised summary of evidence completeness rather than a full causal inference model.

Variant-Level Interpretation

A high value of θ_i means that most of the biological and functional evidence layers required by the QV ES standard are satisfied for variant i. The credible interval provides uncertainty quantification. A narrow interval near one indicates near-complete and consistent evidence. A wide interval near zero indicates sparse or conflicting evidence.

Relative Interpretation across Variants

The percentile expresses the support for the candidate relative to the support for all other variants. This addresses the practical clinical question of how well supported this explanation is compared with all other possible explanations currently observable in the genome. The combination of θ_i, its credible interval, and its percentile forms a comprehensive evidence-based profile for any candidate.

6.7 Methodological Advantages

The Bayesian evidence sufficiency framework has several strengths.

1. It separates biological knowledge availability from the analysis provider's variant filtering pipeline.
2. It quantifies uncertainty explicitly and with closed-form posterior distributions.
3. It ranks candidate variants not by arbitrary thresholds but by their posterior evidence sufficiency and their relative support in the full genome.
4. It is computationally trivial and suitable for millions of variants.
5. It is fully auditable and reproducible.
6. It does not encode pathogenicity directly, but rather the completeness and reliability of known information, which is a distinct and necessary layer of interpretation.

Figure S6 explains how the Quant ES system measures how much reliable scientific information is available for each genetic variant in a genome. The method starts by checking a set of evidence rules for every variant, for example how well the variant matches known disease mechanisms or whether it has been observed in relevant databases. These results form a simple table of green and red entries. The number of green entries tells us how much information is available. A mathematical model

then translates this count into a probability that the evidence is strong enough to support interpretation. Panel B shows how this probability grows as more evidence becomes available. Finally, we compare the evidence for the candidate variant with the evidence for all other variants across the genome. This gives two important answers. First, whether the evidence for the candidate is complete enough to support interpretation. Second, how it compares with all alternative variants that might explain the clinical scenario.

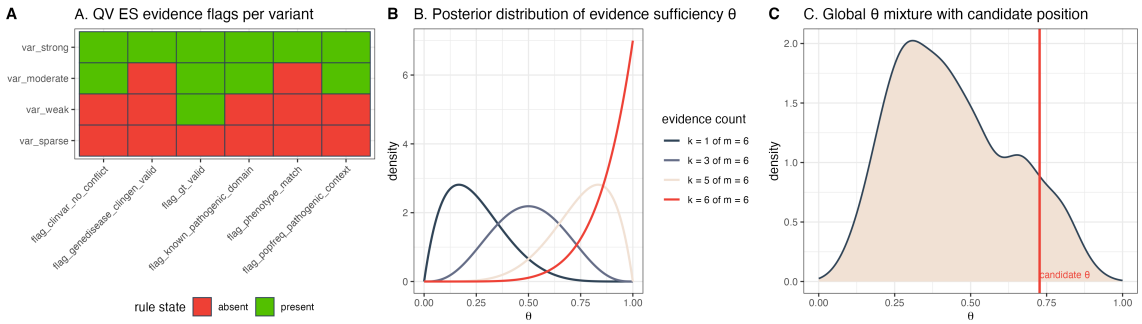


Figure S6: **Overview of the Quant ES Bayesian evidence model.** Panel A shows an illustrative evidence matrix in which QV ES rule outcomes are recorded for several hypothetical variants. Each row corresponds to a variant and each column corresponds to a required evidence element. Green cells indicate that the relevant evidence is available and red cells indicate that it is absent. Panel B displays the posterior distribution of the evidence sufficiency parameter θ for different counts of observed supporting evidence. Each curve corresponds to a value of k, the number of available evidence elements out of the total m rules. Larger values of k shift the posterior mass towards higher θ. Panel C shows the combined distribution of θ across all variants in a genome together with the position of a selected candidate. This global mixture describes the background level of biological knowledge available across the genome, while the candidate's location shows how its evidence compares with all other possible explanations.

6.8 Summary

We presented a Bayesian system for modelling QV ES rule satisfaction across variants. We model the observed rule outcomes as binary samples from a latent evidence suf-

ficiency parameter θ_i, which quantifies the completeness of biological and functional information available for variant i.

This quantity, θ_i, is a principled measure of evidence sufficiency (16; 24). We further compare each candidate with the empirical distribution of θ_i across the genome, obtaining a percentile that reflects relative support. The approach is simple, scalable, statistically coherent, and directly aligned with the realities of clinical genomics, where the limiting factor is often evidence availability rather than categorical statements about pathogenicity.

This framework can be included as a core methodological supplement in a statistical genomics application note or as a standalone statistical appendix in a clinical genomics study.

6.9 Simulation figures

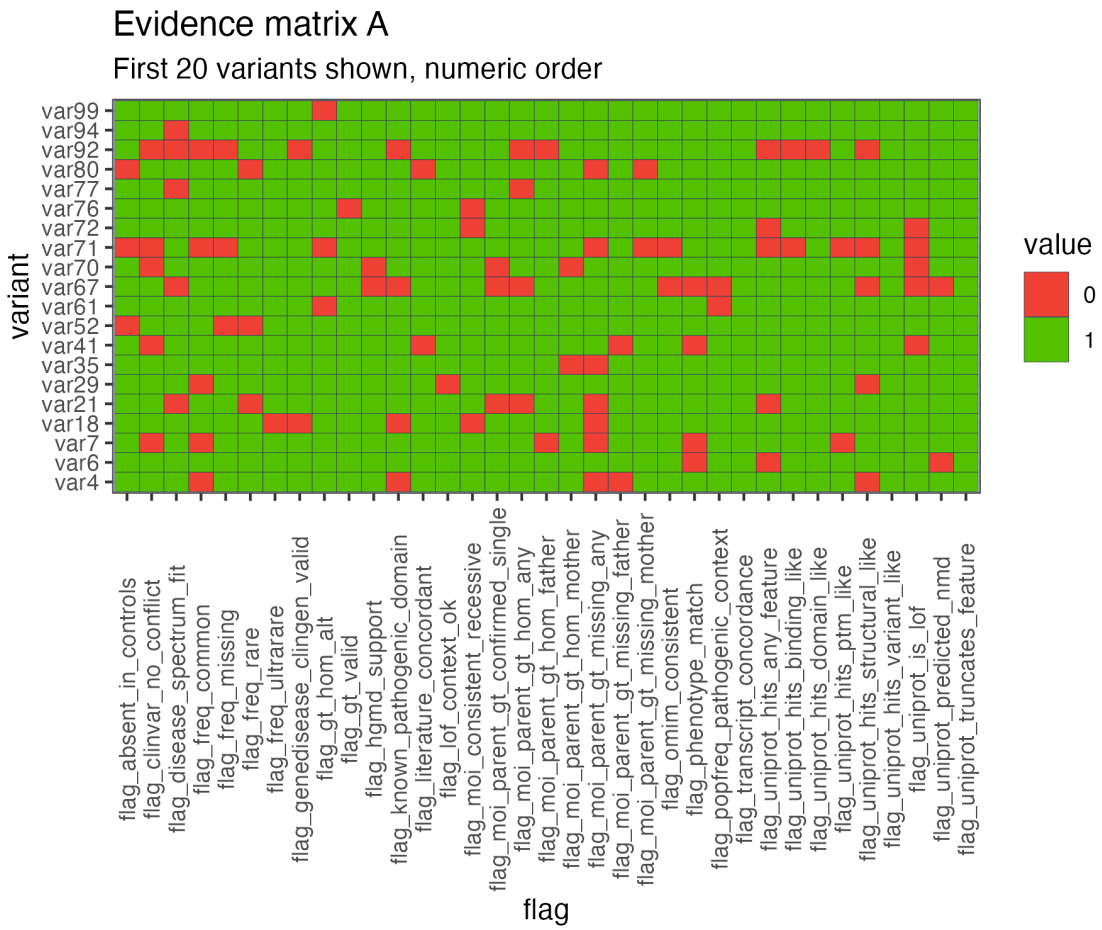


Figure S7: **QV ES evidence interpretation matrix (simulation study).** Each row is a simulated variant and each column is a QV ES rule. The simulated evidence matrix illustrates heterogeneity in rule satisfaction and motivates quantitative evidence modelling.

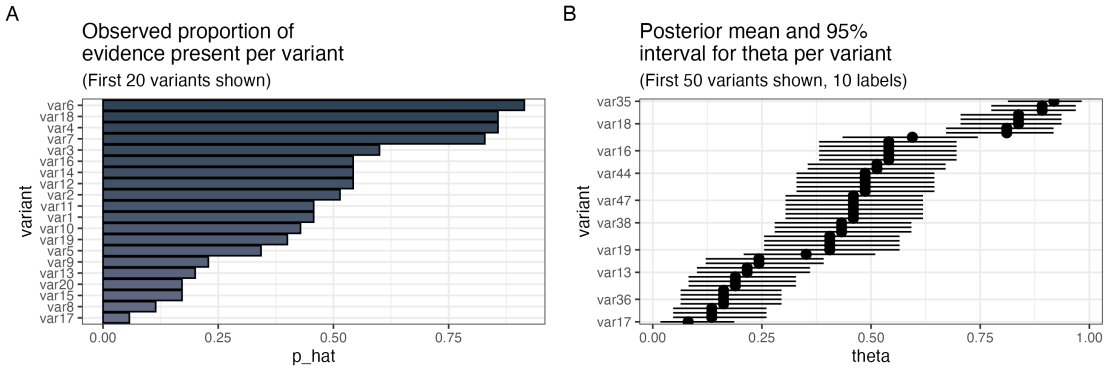


Figure S8: **Evidence sufficiency across variants (simulation study). (A)** Observed proportions $\hat{p}_i$. **(B)** Posterior means and 95% credible intervals for θ_i. These results demonstrate the theoretical behaviour of the Beta Binomial model under controlled conditions.

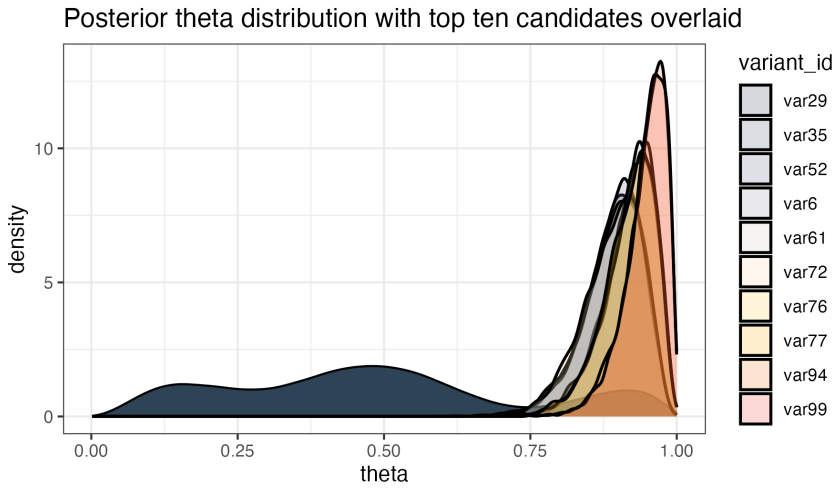


Figure S9: **Global evidence sufficiency distribution with candidate overlay (simulation study).** The grey background shows the simulated mixture distribution of θ. Candidate overlays illustrate how Quant ES separates high evidence and low evidence variants under theoretical assumptions.