Swiss Genomics Association consensus guideline for evidence-based genomic variant interpretation in Mendelian disease

Quant Group[1], Dylan Lawless *[2], and the joint SGA committee[3]

[1]The quantitative omic epidemiology group.

[3]The Swiss Genomics Association commitee.

[2]Department of Intensive Care and Neonatology, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland.

November 23, 2025

DRAFT – NOT FINAL

This document is a **live public working draft** and does not represent a completed or officially endorsed release. All statements, analyses, and opinions are those of the authors and have not yet undergone joint committee review or external peer review. Content is provided for transparency and collaborative development and may be revised substantially prior to publication.

Abstract

Genomic medicine programmes across the world are advancing rapidly, especially with AI, but variant interpretation remains implemented differently across national, institutional and commercial efforts, with no shared agreement on how supporting evidence should be structured, recorded or exchanged. As a result, equivalent data are repeatedly recomputed, manually curated or privately remodelled, while shared learning, verification and reuse remain limited. Key evidence including inheritance, provenance, population context, functional data and conflicting observations is inconsistently captured, reducing transparency and interoperability and slowing collective progress.

We define a harmonised, tool-agnostic specification that establishes a shared data architecture and scientific approach for variant interpretation in Mendelian disease. The guideline sets minimum requirements for representing, linking and auditing evidence, including sequence and sample provenance, variant normalisation, segregation logic, phenotype alignment, evidence grading, conflict handling and versioned synthesis statements. It emphasises structured, reviewable reasoning rather than fixed classification labels and is designed to complement, not replace, existing standards, tools and workflows.

By formalising a standardised approach to quantifying interpretation, this framework enables distributed groups to solve genomics problems in alignment rather than in parallel isolation. It reduces duplication, supports independent validation, improves automation and ensures that advances in one sector directly strengthen others. The model establishes a shared, exchangeable foundation for interpretation that remains transparent, interoperable and evolvable, strengthening cooperation between national initiatives, healthcare, research and industry. [1]

[1]* Addresses for correspondence: Dylan.Lawless@uzh.ch.

Availability: This data is integrated in https://iei-genetics.github.io.

Contents

1 Introduction

Genome sequencing is now routine in research, diagnostics, and commercial genomics, but the evidence supporting variant interpretations is recorded, structured, and justified in widely different ways. Algorithmic scores, proprietary rankings, or pathogenicity labels are often treated as conclusions rather than evidence, while key determinants of validity including sequence provenance, variant identity, inheritance logic, mechanism fit, population context, phenotype alignment, and contradictory observations are inconsistently captured or reported. This fragmentation limits reproducibility, weakens cumulative knowledge, and slows safe integration of AI-driven and cross-sector genomic innovation.

The Swiss Genomics Association is building a national, open infrastructure for shared standards at the intersection of healthcare, research and industry. Our aim is not to replace existing frameworks or tools, but to define the evidential foundation they depend on: transparent, computable, and auditable variant interpretation that preserves biological nuance, remains interoperable across platforms, and retains full reasoning rather than only final labels. By establishing common expectations for evidence traceability, uncertainty, mechanistic reasoning, inheritance verification and phenotype grounding, this guideline enables consistent interpretation that can be exchanged, validated, and built upon by laboratories, companies, hospitals and researchers alike.

This effort strengthens national coordination where common standards are still underdefined, ensuring genomic data remains clinically meaningful, scientifically reusable, and technically compatible with future care models, analytical innovation and AI-supported interpretation.

2 Purpose and scope of the guideline

3 Architectural context and national motivation

4 Design principles for evidence-based variant interpretation

5 Data provenance and sample quality requirements

6 Evidence rule framework

We set an important conceptual basis around evidence and counter evidence. The entire framework is based on reverse reasoning: For every variant proposed by an upstream tool, the national system tests all available evidence domains for signals that contradict, weaken, or fail to support the hypothesis that this variant is the cause of disease. By default we also need certain agreements on the structure for allowing an interoperable use of results.

6.1 Structure of evidence rules

6.2 Versioning and audit trace

6.3 Optional reference implementation (QV framework)

QV file guideline content

Listing 1: Excerpt from the QV interpretation prompt used with the QV Builder app. Line wrapping is shown for display only. For actual use, refer to the original source file or the corresponding official QV set release.

```
draft style of prompts used to make the variant interpretation flagging
    ↪  rules before converting to YAML/JSON:

```
meta qv_set_id="qv_rare_singlecase_interpretation_v1_20251105"
meta version="1.0.0"
meta title="Rare variant single case interpretation"
meta created="2025-11-05"
meta authors=DylanLawless

GnomAD flags
criteria gnomad_flags logic=or desc="Variant has no gnomAD flags or
    ↪ will be reviewed in report generation"
criteria gnomad_flags field=gnomad_flag operator==" value=NA
criteria gnomad_flags 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"
note "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"

ACMG flags
criteria acmg_criteria logic=or desc="Variant has no benign ACMG
    ↪ evidence or will be reviewed in report generation"
criteria acmg_criteria field=ACMG_criteria operator="not_contains"
    ↪ value="B"
criteria acmg_criteria field=ACMG_criteria operator="contains" value="
    ↪ BA1,BS1,BS2,BS3,BS4,BP1,BP2,BP3,BP4,BP5,BP6,BP7,BP8"
note "ACMG: ACMG benign evidence codes are retained for review;
    ↪ variants carrying these codes must be justified in the final
    ↪ causal interpretation report"
note "ACMG: For details on ACMG criteria see Richards et al. 2015, Li
    ↪ et al. 2017, Riggs et al. 2020, Tavtigian et al. 2020"
```
```

```
note "ACMG: Variants containing benign evidence codes (BA, BS, BP) are
    ↪ not excluded but must be explicitly reviewed in the final report"

criteria ppie_reporting logic=or desc="Record presence and fulfilment
    ↪ of patient or public representation (PPIE) requirements"
criteria ppie_reporting field=ppie_status operator="in" value="
    ↪ fulfilled,not_applicable,declined,pending"
note "PPIE: For each case, record whether patient or public
    ↪ representation requirements were met according to context (
    ↪ clinical, research, or commercial)."
note "PPIE: Minimum record includes: opportunity for expert
    ↪ consultation, whether discussion occurred, and whether non-
    ↪ actionable findings were addressed."
note "PPIE: In commercial contexts, the report itself may satisfy this
    ↪ requirement if it conveys the relevant information clearly and
    ↪ accessibly without requiring direct consultation."
note "PPIE: In research or cohort studies, record whether participants
    ↪ were informed about data use and whether individual feedback was
    ↪ planned."


# Inheritance pattern

criteria gt_valid field="proband_geno" operator="in" value="HET,HOM,
    ↪ HEMI"

criteria comp_het group_by="sample,SYMBOL" count=">=2" desc="Two
    ↪ distinct variants in one gene"

criteria not_hom_parent field="proband_geno" operator="==" value="HOM"
criteria not_hom_parent field="mother_geno" operator="!=" value="HOM"
criteria not_hom_parent field="father_geno" operator="!=" value="HOM"

criteria not_hemi_parent field="proband_geno" operator="==" value="HEMI
    ↪ "
criteria not_hemi_parent field="father_geno" operator="!=" value="HEMI"

criteria not_same_parent field="PARENT_GENOTYPE_MATCH" operator="=="
```

```
    ↪ value="false"

criteria phase_ok field="PHASE" operator="in" value="TRANS,UNKNOWN"

criteria allow_dn field="DE_NOVO" operator="==" value="true" logic="or"

note "inheritance priority ordering applied downstream: compound_het >
    ↪ biallelic_hom > de_novo > possible_compound_het >
    ↪ single_het_parent > proband_geno"
note "compound_het includes inherited, inherited+de novo, and dual de
    ↪ novo in trans (phasing required downstream)"
note "missing parent genotypes allow de_novo and possible_compound_het,
    ↪  and may still support compound_het depending on counts"
note "genotype mapping applied upstream: 00=REF, 01=HET, 10=HET, 11=HOM
    ↪ "


# RNA transcript support from controls

criteria rna_in_controls field="RNA_TPM_controls_median" operator=">"
    ↪ value=0.1 desc="Transcript detected in controls"
criteria rna_in_any_control field="RNA_TPM_controls_max" operator=">"
    ↪ value=0.0 desc="Transcript seen in at least one control"
criteria rna_patient_quantifiable field="RNA_TPM_patient" operator=">"
    ↪ value=0 desc="Patient transcript quantified"

criteria rna_pass logic="and" desc="RNA quant suitable for
    ↪ interpretation"
criteria rna_pass field="RNA_TPM_controls_max" operator=">" value=0
criteria rna_pass field="RNA_TPM_patient" operator=">" value=0

criteria rna_fail logic="and" desc="Transcript not seen in any control,
    ↪  quantification unreliable"
criteria rna_fail field="RNA_TPM_controls_max" operator="==" value=0

note "RNA filter: transcript must be observed in at least one control
    ↪ sample to consider patient quantification reliable"
note "RNA FAIL indicates no expression in all controls, patient value
```

```
    ↪ should not be interpreted quantitatively"
````

Variant interpretation depends on how whole genome data are prepared and analysed. Each step, from sequencing and variant calling to annotation and filtering, defines what can be detected or missed. A qualifying variant (QV) protocol makes these steps explicit by describing the rules that determine variant inclusion and interpretation (4).

A standard QV protocol should specify the sequencing method, genome build, and tools used; the quality thresholds for coverage and genotype confidence; the genomic regions or panels considered; and the annotation or classification systems applied. Defining these parameters in a structured QV file separates logic from execution, allowing reproducible, auditable workflows. Each file has a version and checksum, linking it directly to analysis outputs.

This standardisation ensures that variant findings are traceable and comparable across studies. It also supports automated pipelines that integrate multiple evidence sources, such as population, functional, and multiomic data, in line with frameworks like the GA4GH Variant Annotation model. The result is consistent, transparent, and quantifiable variant interpretation.

6.4 Implemented rule set

Qualifying Variant (QV) sets define structured, reproducible variant interpretation rules using a transparent, machine-readable YAML format. Each QV file specifies metadata, filters, and evidence-based criteria in a simple key–value syntax, enabling precise replication of variant selection and interpretation logic across studies. QV sets can be composed interactively using the online https://switzerlandomics.ch/pages/qv_builder/, which provides an intuitive interface for defining rules, previewing YAML output, and ensuring consistency with established standards. This framework promotes FAIR data principles and harmonised variant interpretation for both research and clinical applications.

7 Evidence flags

7.1 Evidence domains

All evidence domains test whether the observable data contradict the hypothesis that a variant is the cause of disease. Each domain represents a structured attempt at falsification. Population frequency, inheritance, molecular function, phenotype fit, and external assertions are all evaluated using the same principles.

7.2 Final flag set

Box 1: Flags

flag_gt_valid
flag_moi_parent_gt_missing_mother
flag_moi_parent_gt_missing_father
flag_moi_parent_gt_missing_any
flag_moi_parent_gt_hom_mother
flag_moi_parent_gt_hom_father
flag_moi_parent_gt_hom_any
flag_moi_parent_conflict_AD
flag_moi_parent_conflict_AR
flag_moi_parent_conflict_XR
flag_moi_parent_conflict_any
flag_popfreq_common
flag_popfreq_rare
flag_popfreq_ultrarare
flag_missing_popfreq
flag_uniprot_hits_any_feature
flag_uniprot_hits_domain_like
flag_uniprot_hits_structural_like
flag_uniprot_hits_ptm_like
flag_uniprot_hits_binding_like
flag_uniprot_hits_variant_like
flag_uniprot_is_lof
flag_uniprot_predicted_nmd
flag_uniprot_truncates_feature

7.3 Interpretation of flag combinations

8 Structured representation and interoperability

8.1 Alignment with GA4GH Variant Representation and Variant Annotation models

Automation, interoperability, and reproducibility in genomics rely on the use of internationally recognised standards rather than isolated or proprietary tools. Validated frameworks enable traceable, comparable, and scalable handling of genomic and phenotypic data across research and clinical contexts (1).

Core standards include the **Variant Call Format (VCF)**, **Binary Alignment Map (BAM/CRAM)**, and the **Phenopacket Schema** from the **Global Alliance for Genomics and Health (GA4GH)**. The **Human Phenotype Ontology (HPO)** and **Orphanet Rare Disease Ontology (ORDO)** support structured phenotyping and disease mapping, while the **Beacon protocol** enables federated search across distributed datasets without sharing raw data.

Using such standards ensures long-term compatibility, supports automation, and improves accuracy through consistent validation and version control. Tools such as *Exomiser*, *LIRICAL*, and *PhenIX* show how ontology-based approaches enable automated, phenotype-driven variant prioritisation. For cloud analysis, platforms like **Terra** and pipelines such as **GATK** provide transparent, reproducible environments aligned with GA4GH and international data-sharing frameworks.

8.2 GA4GH Variant Annotation model

See https://va-spec.ga4gh.org/en/latest/examples/acmg-variant-pathogenicity-statement-with-evidence.html for example.

Variant interpretation has traditionally relied on expert review and rule-based systems such as the ACMG/AMP criteria, where evidence is manually classified into discrete levels (e.g. “strong”, “moderate”, “supporting”). While this provides clinical transparency, it limits scalability and does not easily incorporate quantitative or experimental data generated outside the sequencing pipeline.

Structured annotation frameworks, such as the **GA4GH Variant Annotation (VA) specification**, provide a way to formalise how evidence supports or refutes pathogenicity, linking statements to their provenance and strength. Rather than a human-readable label alone, each assertion is represented as a computable object that connects a variant, a condition, and all supporting evidence lines. These may include cohort allele frequencies, functional assays, or other study results, each qualified by method, direction, and strength of support. This structure allows integration of diverse evidence types while retaining traceability to original data sources.

Such a model enables a shift from categorical to quantitative reasoning. A variant's pathogenicity statement can accumulate weighted evidence from multiple domains: ACMG-derived criteria, *in vitro* functional data (e.g. MAVE), RNA and protein evidence, population studies, or model organism data. Each evidence line can be standardised using controlled vocabularies and shared identifiers, allowing aggregation across studies and automated computation of posterior probabilities.

Tools such as *Exomiser* and *LIRICAL* illustrate early automation based on ACMG-compatible logic and phenotype-driven scoring. Extending these with the GA4GH VA model allows incorporation of continuous, probabilistic evidence rather than threshold-based categories. This approach transforms variant classification from interpretive judgment into quantifiable inference, reducing subjective variability and supporting automated re-evaluation as new evidence emerges.

By representing all evidence as structured, machine-interpretable data with defined provenance, the GA4GH framework provides a foundation for fully automatable, transparent, and continuously updatable variant interpretation.

The Variation Representation Specification (VRS) is development by the GA4GH in the Genomic Knowledge Standards (GKS) Work stream;
https://www.ga4gh.org/product/variation-representation/.

8.3 Exchange formats and machine readable outputs

9 Probabilistic interpretation in Pillar 3

9.1 Relationship to Pillar 1 and Pillar 2 inputs

9.2 Evidence aggregation logic

10 Clinical review checkpoints

10.1 Required manual assessments

10.2 Boundary of automated reasoning

11 Reporting and synthesis statements

12 Auditability, versioning, and reproducibility

13 Limitations and scope

14 Future extensions

15 Conclusion

Acknowledgements

Contributions

DL designed the analyses and wrote the manuscript.

Competing interest

The authors declare no competing interest.

Ethics statement

This study only used data which was previously published and publicly available, as cited in the manuscript.

Data availability

The data used in this manuscript is derived from open sources which are cited in methods. The data generated is available from ...

Funding

References

[1] Michael H. Duyzend, Pilar Cacheiro, Julius O. B. Jacobsen, Jessica Giordano, Harrison Brand, Ronald J. Wapner, Michael E. Talkowski, Peter N. Robinson, and Damian Smedley. Improving prenatal diagnosis through standards and aggregation. *Prenatal Diagnosis*, 44(4):454–464, April 2024. ISSN 0197-3851, 1097-0223. doi: 10.1002/pd.6522.

[2] Reece K. Hart, Ivo F. A. C. Fokkema, Marina DiStefano, Ros Hastings, Jeroen F. J. Laros, Rachel Taylor, Alex H. Wagner, and Johan T. Den Dunnen. HGVS Nomenclature 2024: Improvements to community engagement, usability, and computability. *Genome Medicine*, 16(1):149, December 2024. ISSN 1756-994X. doi: 10.1186/s13073-024-01421-5.

[3] Johan T. Den Dunnen, Raymond Dalgleish, Donna R. Maglott, Reece K. Hart, Marc S. Greenblatt, Jean McGowan-Jordan, Anne-Francoise Roux, Timothy Smith, Stylianos E. Antonarakis, Peter E.M. Taschner, and on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP),

and the Human Genome Organisation (HUGO). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, 37(6):564–569, June 2016. ISSN 10597794. doi: 10.1002/humu.22981.

[4] 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. May 2025. doi: 10.1101/2025.05.09.25324975.

[5] Ksenia Lavrichenko, Emilie Sofie Engdal, Rasmus L. Marvig, Anders Jemt, Jone Marius Vignes, Henrikki Almusa, Kristine Bilgrav Saether, Eiríkur Briem, Eva Caceres, Edda María Elvarsdóttir, Magnús Halldór Gíslason, Maria K. Haanpää, Viktor Henmyr, Ronja Hotakainen, Eevi Kaasinen, Roan Kanninga, Sofia Khan, Mary Gertrude Lie-Nielsen, Majbritt Busk Madsen, Niklas Mähler, Khurram Maqbool, Ramprasad Neethiraj, Karl Nyrén, Minna Paavola, Peter Pruisscher, Ying Sheng, Ashish Kumar Singh, Aashish Srivastava, Thomas K. Stautland, Daniel T. Andreasen, Esmee Ten Berk De Boer, Søren Vang, Valtteri Wirta, and Frederik Otzen Bagger. Recommendations for bioinformatics in clinical practice. *Genome Medicine*, 17(1):124, October 2025. ISSN 1756-994X. doi: 10.1186/s13073-025-01543-4.

Acronyms

16 Supplemental

Supplemental data are presented under the same headings that correspond to their relevant main text sections.

17 Appendix

17.1 Example rule files

17.2 Example YAML or JSON evidence objects

17.3 Worked examples of flagged variants

18 Evidence flags

18.1 Evidence domains

18.1.1 Counter-evidence and conflict handling

18.1.2 Normalised variant representation and nomenclature

Accurate variant description depends on consistent use of internationally recognised nomenclature systems. The Human Genome Variation Society (HGVS) nomenclature provides the authoritative standard for describing sequence variants at the DNA, RNA, and protein levels. It ensures that each variant is expressed unambiguously and reproducibly across clinical reports, publications, and databases.

The HGVS Nomenclature is maintained by the HGVS HGVS Variant Nomenclature Committee (HVNC) under the H Human Genome Organisation (HUGO) and is widely implemented across major genomic resources and clinical interpretation platforms. Current recommendations are detailed in Hart et al. (2) and Den Dunnen et al. (3), which formalise the syntax, reference sequence alignment, and conventions for variant expression.

Within this guideline, all variants should be reported according to the most recent HGVS Nomenclature release, aligned to an approved reference sequence (RefSeq or

Ensembl transcript). Both coding (`c.`) and protein (`p.`) level annotations should be provided where possible. Genomic coordinates should follow the Genome Reference Consortium Human Build 38 (GRCh38) reference assembly.

Using standardised nomenclature ensures interoperability between laboratories, software tools, and public databases, reducing ambiguity in variant exchange and supporting precise traceability in both research and clinical reporting.

18.1.3 ACMG criteria with counter-factual evidence

The `acmg_criteria` rule interprets the condensed ACMG criteria column for each variant and flags variants that carry benign evidence. When a variant has passed upstream prioritisation tools such as Exomiser, the ACMG criteria are typically recorded in a single column as a list of applied evidence codes (for example, `PVS1, PM2, PP3`, or `BA1`). The downstream rule inspects this condensed string to determine whether it contains any benign evidence codes, recognised by the presence of the letter "B" (for example, `BA1`, `BS1--BS4`, or `BP1--BP8`). Variants with such codes are retained but marked for review, similar to those carrying gnomAD quality flags. This ensures that variants with benign evidence are not automatically excluded but instead require explicit assessment or justification in the final clinical interpretation.

Rank based scoring such as that from Tavanian offers benefits but similarly can miss the presence of counter-factual evidence that a variant might be considered benign but remain prioritised by the presence of other pathogenic flags.

variant_id	ACMG_criteria	outcome	report_action
1	PVS1, PM2, PP3	pass	include normally
2	PVS1, PM2, BA1	pass (flagged)	include with note: "contains benign ACMG evidence (BA1); review interpretation."
3	BP4	pass (flagged)	include with note: "benign supporting evidence; verify consistency with phenotype."
4	PS2, PM5	pass	include normally

Table S2: Example application of the `acmg_criteria` rule to patient variants. Variants containing benign ACMG evidence codes are retained but flagged for explicit review in the final report.

18.1.4 Population frequency evidence

18.1.5 Conditional evidence rules and source-dependent quality checks

Reference databases and the case sample may share variant-calling or sequencing biases. Therefore, databases which carry flagged variants should be examined through conditional automation or manual review. For instance, GnomAD is a key reference for interpreting variants in single-case analyses. The `gnomad_flags` rule ensures that flagged variants are reviewed rather than excluded. Variants without a flag (`NA`) pass directly, while those with recognised gnomAD flags pass with review status.

Common flags include `AS-VQSR` (allele-specific quality recalibration), `RF` (random forest outlier), `LC pLoF` (low-confidence loss-of-function), and `SEGDUP` (segmental duplication). These indicate potential technical or annotation uncertainty rather than confirmed artefacts. We recommend that variants are retained in reporting but require justification or comment to ensure transparency and traceability.

variant_id	gnomad_flag	outcome	report_action
1	NA	pass	include normally
2	LC pLoF	pass (flagged)	include with note: "gnomAD LC pLoF; review interpretation."
3	UnknownFlag	fail	hold for manual review

Table S4: Example application of the `gnomad_flags` rule to patient variants. Entries automatically wrap within the column width for compact layout.

18.1.6 Inheritance and segregation evidence

Version 1

When pedigree data are available, Whole Genome Sequencing (WGS) enables direct evaluation of inheritance models for each variant. Genotypes are interpreted across proband and parents using standard representations such as `REF`, `HET`, and `HOM`, or equivalent encodings used in variant data formats (`0,1,2` in PLINK, `0/0`, `0/1`, `1/1`, or phased forms such as `0|1` in VCF). From these data, the inheritance pattern for each variant is determined, such as de novo, homozygous, heterozygous, or compound heterozygous, and this information is then evaluated in relation to the known gene–disease relationship. For example, the observed segregation pattern may provide

supporting or contradictory evidence for anAutosomal Dominant (AD), Autosomal Recessive (AR), or X-Linked (XL) disease mechanism.

Where genotype data are incomplete, inheritance may be inferred from clinical features, family history, or segregation information, but such cases are explicitly flagged as uncertain. Scenarios of incomplete penetrance, such as a heterozygous variant inherited from an unaffected parent, are also recorded because they influence the strength of causal interpretation. Each inheritance assessment includes both the inferred pattern and the type of supporting evidence, ensuring that interpretative conclusions in the genetic report transparently reflect the available data and its confidence level.

Version 2

Accurate interpretation of inheritance requires integrating two complementary sources of evidence: (1) the *Mode of Inheritance (MOI)* defined by curated reference datasets that describe known gene–disease mechanisms, and (2) the *observed inheritance pattern* derived from family genotype or clinical data.

The reference MOI defines the expected transmission mechanism for a gene–disease pair, typically AD, AR, or XL. Structured datasets such as PanelAppRex (**?**) harmonise these annotations across thousands of curated panels. Foundational sources including Genomics England's PanelApp and PanelApp Australia (**?**) provide continuously updated expert curation underpinning national infrastructures such as the NHS National Genomic Test Directory and the 100,000 Genomes Project. The MOI field thus serves as an evidence-based prior — a quantitative expectation of how pathogenic variants in a given gene are likely to segregate.

The observed inheritance pattern, by contrast, is determined from the case data. When trio or family WGS is available, inheritance can be assessed directly from genotype encodings (`REF`, `HET`, `HOM`, or `0/1`, `1/1`, `0|1`). This pattern may confirm or contradict the reference MOI. For example, a de novo heterozygous variant in an AD gene supports causality, whereas biallelic variants in an AR gene are expected.

Cases with incomplete genotype data require inference from clinical or segregation information and must be explicitly flagged as uncertain. Incomplete penetrance, mosaicism, or unaffected carriers (e.g. heterozygous variants in AR genes) should be documented, as they influence the posterior probability of pathogenicity (**?**).

To ensure consistency, each variant interpretation should record both the reference MOI (from curated databases) and the observed inheritance pattern (from patient

data). This dual recording enables probabilistic interpretation frameworks, such as Quant (**?**), to integrate population frequencies, genotype configurations, and inheritance priors under Hardy–Weinberg equilibrium. Together, these components quantify diagnostic confidence and prevent misclassification of variants arising from uncertain or incomplete pedigree information.

18.1.7 Functional and molecular evidence

Functional evidence from UniProt is integrated by detecting positional overlap between variant amino acid coordinates and annotated protein features recorded in UniProt GFF files. Each UniProt entry provides structured annotations describing biochemical, structural, and functional properties of the protein. For each variant, the affected residue position is compared against these annotated intervals, and any intersection is recorded as supporting evidence in the QV interpretation framework. This approach ensures that experimental and curated protein-level information contributes directly to variant interpretation and reporting (**Figure S1**).

The annotated features used as evidence sources include catalytic and binding sites, metal and nucleotide binding regions, and other experimentally defined functional motifs. Structural features such as helices, beta strands, and coiled coils provide spatial context for potential conformational disruption. Domain- and family-level annotations, including domains, motifs, and topological regions, capture conserved structural organisation and functional domains. Additional layers include post-translational modification sites, mutagenesis data, and known sequence variants curated in UniProt. Processing and localisation signals (such as signal peptides, transit peptides, and cleavage products) and cautionary sequence annotations (for example, frameshifts or sequence uncertainty) are also recorded.

By systematically linking these feature classes to variant coordinates, the framework records not only where functional or structural evidence exists, but also the type of information present—whether experimental, inferred, or computational. This enables each variant interpretation to transparently reflect the available molecular evidence supporting its classification.

18.1.8 Phenotype and gene–disease validity evidence

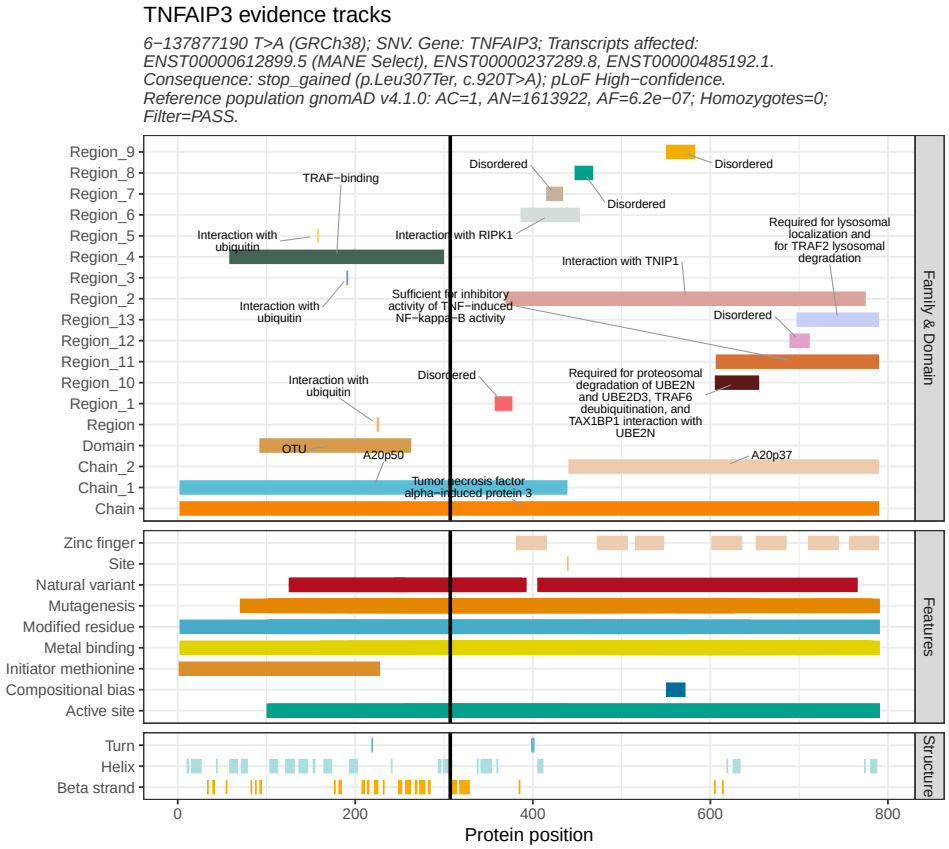

Figure S1: **Functional evidence tracks from UniProt annotations.** The example illustrates how protein-level features such as domains, motifs, and catalytic sites provide structured evidence supporting interpretation of coding variants. Overlaps between variant positions and curated functional regions indicate potential mechanistic relevance, while the absence of overlap suggests limited or indirect evidence. This evidence framework guides the strength of interpretation in clinical reporting, ensuring that well-supported variants are highlighted and uncertain findings are transparently qualified.