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 8, 2025

Abstract

This document aims to bridge clinical, research, and commercial domains by defining transparent, reproducible standards for genomic data interpretation and reporting. The framework integrates structured variant annotation, inheritance assessment, and functional evidence to support consistent and traceable decisions in clinical genomics. Emphasis is placed on automation, interoperability, and adherence to international standards to ensure that genomic interpretation remains accurate, reproducible, and aligned with both scientific and regulatory expectations. While more advanced analytical frameworks, including those using artificial intelligence or probabilistic inference, can build upon this foundation, this guideline defines the minimum technical and procedural standard required to ensure accuracy and transferability across domains. **Conclusions**: [1]

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 Background

3 Principles of evidence-based variant interpretation

4 Preference for established standards

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 (3).

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.

5 Structured variant annotation and automation of interpretation

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.

6   Variant annotation specification

We will examples of YAML/JSON and figures - see
https://va-spec.ga4gh.org/en/latest/examples/
acmg-variant-pathogenicity-statement-with-evidence.html.

GA4GH VA-Spec provides a structured, interoperable framework for representing variant interpretations, evidence lines, provenance, and clinical assertions using shared semantics. It allows algorithmic outputs, literature evidence, cohort data, functional studies, and clinical propositions to be encoded in a transparent, traceable format that different laboratories and software providers can exchange without loss of meaning. This makes it highly suitable for improving variant interpretation reporting, particularly when combining automated tool outputs with manual clinical review, while ensuring that each evidence source, method, and agent is explicitly recorded. It also aligns with established frameworks such as ACMG/AMP, enabling rule-based evidence to be represented in a consistent, computable form.

However, VA-Spec is not itself a method for deciding pathogenicity or clinical actionability. It does not score or weight evidence, validate sequencing quality, enforce inheritance logic, detect biologically implausible results, resolve contradictory evidence, or judge whether the total evidence meets a clinical threshold. It cannot determine whether allele frequency conflicts with disease prevalence, whether zygosity matches the disorder model, whether phenotype congruence is sufficient, or whether alternative diagnoses were adequately excluded. A VA-Spec record can be structurally valid yet clinically incomplete if it contains only algorithmic predictions without essential genetic and diagnostic checks.

For this reason, VA-Spec should be used as the evidence representation and exchange layer, while the clinical interpretation framework must sit above it. This additional

layer must provide explicit evidence weighting, quality thresholds, segregation validation, phenotype matching logic, mechanism-of-disease assessment, differential diagnosis tracking, and clear delineation between algorithmic inference and expert clinical judgement. When combined, VA-Spec enables standardised, auditable reporting, but the clinical decision model must supply the reasoning, safeguards, and quantitative evaluation required for patient-grade variant assessment.

Example content. The example omits full representations of these VRS and CatVRS Variation objects - as these are large structures that are the remit of other GKS Specifications.

Box 1: GA4GH variant annotation example

```
{
  "SCV000778434.1": {
    "id": "SCV000778434.1",
    "type": "Statement",
    // ClinVar assertion describing pathogenicity claim for a single variant-
disease pair
    "proposition": {
      "id": "ex:Proposition001",
      "type": "VariantPathogenicityProposition",
      "subjectVariant": "clinvar/208366",
      "predicate": "isCausalFor",
      "objectCondition": {
        "id": "clinvar.trait/939",
        "conceptType": "Disease",
        "name": "Autosomal dominant nonsyndromic hearing loss 2A",
        "primaryCoding": {
          "code": "C2677637",
          "system": "https://www.ncbi.nlm.nih.gov/medgen/",
          "iris": ["http://identifiers.org/medgen/C2677637"]
        }
      },
      // Qualifier for degree of penetrance (here stated as high)
      "penetranceQualifier": {
        "primaryCoding": { "code": "high", "system": "local:pathogenicity-
penetrance-qualifier" },
        "name": "high"
      }
    },
    "direction": "supports",
    // Overall assertion strength assigned by submitter/guideline
    "strength": {
      "primaryCoding": { "code": "established", "system": "ImplementerSystem1" }
    },
    // Final classification label applied under a guideline (e.g. ACMG 2015)
    "classification": {
      "primaryCoding": { "code": "disease-
causing", "system": "ACMG-2015" }
    },
    // Provenance: who evaluated/submitted and when
```

7 Data provenance and sample quality control

8 Variant representation and normalisation

8.1 Nomenclatures

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. (1) and Den Dunnen et al. (2), 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.

9 Functional evidence databases

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** 1).

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.

10 Automated QV (qualifying variant) framework

10.1 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.

```
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


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
```
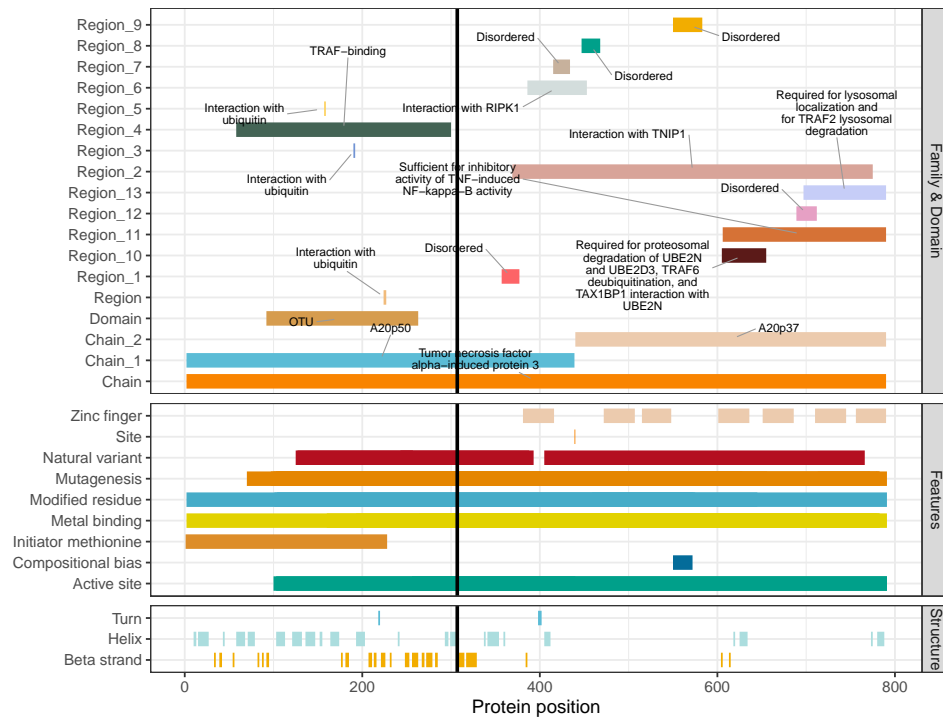
Figure 1: **Functional evidence tracks from UniProt annotations.** The example illustrates how protein-level features such as domains, motifs, and catalytic sites provide structured evidence supporting interpretation of coding variants. Overlaps between variant positions and curated functional regions 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.

```
criteria gnomad_flags field=gnomad_flag operator="in" value="ACO,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"
```

```
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."
```

10.2 Qualifying variant protocol design and standardised approaches

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.

10.3 QV set preparation

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.

10.4 QV sets for reporting

10.5 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 2: Example application of the `gnomad_flags` rule to patient variants. Entries automatically wrap within the column width for compact layout.

10.6 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 4: 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.

10.7   Pedigree and inheritance information

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.

10.8 Pedigree and inheritance information

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.

11 Population frequency and cohort context

12 Inheritance and segregation evidence

13 Phenotype matching and gene-disease validity

14 Counter-evidence and conflict resolution

15 Manual clinical genetics review checkpoints

16 Semi-quantitative scoring rubric

17 End-to-end WGS interpretation workflow

18 Patient and public representation

Patient and public representation (**ppie!** (**ppie!**)) applies across clinical, research, and commercial contexts. The framework should record whether, and how, the duty of representation was fulfilled in each setting.

For clinical or hospital-based cases, the record should indicate whether the patient or guardian had the opportunity to discuss findings with a qualified genetics expert and whether follow-up support was offered.

For commercial services where results are delivered online, the record should state whether the customer was informed about the availability of expert consultation, explanatory materials, or optional counselling.

For research or large cohort studies without individual feedback, documentation should confirm that participants were informed about the nature of data use and that no individual result return was planned.

Recording these details ensures consistent accountability and transparency in all contexts without storing identifiable or sensitive personal information.

19 Recommendations for bioinformatics in clinical practice

Clinical bioinformatics forms the analytical backbone of genomic medicine. It bridges raw sequencing data and clinical interpretation, and its reliability determines the accuracy of every downstream result. To maintain national consistency and international credibility, Switzerland should adopt a harmonised framework for sequencing-based clinical bioinformatics aligned with the recommendations of the Nordic Alliance for Clinical Genomics (NACG) (5). While the NACG framework was developed for the Nordic region, its principles provide an appropriate foundation for Swiss implementation. The key recommendations, summarised in their Table 1, outline the operational, technical, and quality criteria expected of any accredited clinical bioinformatics unit.

The Swiss Genomics Association endorses the core sentiment of these recommendations: that clinical bioinformatics must operate at the same professional and regulatory standards as accredited medical laboratories. Bioinformatics pipelines should be validated, reproducible, and transparent, with defined responsibilities for quality management, data integrity, and patient safety. Use of the GRCh38/hg38 reference genome, strict version control, and structured documentation of analytical steps are expected across all clinical operations. Pipelines must be tested systematically, covering unit, integration, and end-to-end validation, and benchmarked against established reference datasets such as Genome in a Bottle (GIAB) and SEQC2, supplemented by local recall testing of verified clinical samples.

From an infrastructure perspective, secure, isolated computing environments are essential, with preference for air-gapped, clinical-grade high-performance systems. Software should be encapsulated in containerised or environment-controlled frameworks to ensure reproducibility and auditable traceability. Version-controlled source code, accompanied by peer-reviewed updates and complete change logs, is mandatory for clinical deployment.

Data integrity and identity verification are further non-negotiable components. File hashing, sample fingerprinting, and checks for sex, ancestry, and relatedness provide safeguards against sample mix-ups or data corruption. In-house databases of recurrent or artefactual variant calls should be maintained to filter false positives, particularly for structural variants.

Finally, the Association emphasises that clinical bioinformatics is not solely a computational discipline but a multidisciplinary field requiring expertise in software engineering, data management, human genetics, and quality assurance. Sustainable national implementation depends on cross-site collaboration, shared reference datasets, and continual alignment with evolving ISO and GA4GH standards.

Together, these principles define the expected Swiss standard for clinical bioinformatics: evidence-based, reproducible, securely managed, and accredited under frameworks equivalent to ISO 15189. Adhering to these practices will ensure that genomic analysis in Switzerland remains accurate, auditable, and of enduring public value.

Table 5: Recommendations for sequencing-based clinical bioinformatics, based on Lavrichenko et al. (5).

1. Genome build **hg38** is the national reference for alignment.
2. Standard analyses: SNV, CNV, SV, STR, LOH, variant annotation, PRS (optional), and for cancer, TMB, HRD, MSI.
3. Use multiple tools for structural variant calling.
4. Apply in-house datasets to filter recurrent or false-positive SV calls.
5. Use air-gapped, clinical-grade HPC and IT systems.
6. Operate under ISO 15189 or equivalent accreditation.
7. Employ standardised file formats and terminologies.
8. Document and test pipelines for accuracy and reproducibility.
9. Subject production code to manual review and testing.
10. Manage all code and documentation under strict version control (e.g. Git).
11. Validate pipelines with GIAB (germline) and SEQC2 (somatic) truth sets, plus local recall tests.
12. Conduct unit, integration, system, and end-to-end tests.
13. Verify data integrity using file hashing (e.g. MD5, SHA1).
14. Confirm sample identity via fingerprinting or inferred traits.
15. Encapsulate software using containers or controlled environments.
16. Maintain multidisciplinary teams covering software, data, QA, and human genetics expertise.

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] 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.

[2] 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.

[3] 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.

[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

C Supplemental

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