Report

Quant Group[1] and Dylan Lawless *[6]

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

November 3, 2025

Abstract

Background:

Methods:

Results:

Conclusions: [1]

1 Introduction

2 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 (**?**).

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.

[1]* Addresses for correspondence: Dylan.Lawless@uzh.ch.
Availability: This data is integrated in `https://iei-genetics.github.io`.

3 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.

4 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 (**?**).

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.

5 Compliance and accreditation framework

Genomic data processing and analysis in Switzerland are governed by national and international legal standards designed to ensure data protection, traceability, and clinical reliability. The primary legal framework is the **Federal Act on Data Protection (FADP, SR 235.1)**, which regulates the processing of personal and sensitive data, including genetic information. Key obligations under the FADP include lawful processing (Arts. 6, 31), the establishment of Data Processing Agreements between controllers and processors (Art. 9), the implementation of technical and organisational safeguards such as encryption and access logging (Art. 8), maintenance of records of processing activities (Art. 12), breach notification procedures (Art. 24), and regulated cross-border data transfers requiring adequacy decisions or safeguards (Arts. 16–18).

Under Swiss law, platforms that store or share genomic data are typically considered

data processors, not *medical devices*, unless marketed for diagnostic, therapeutic, or monitoring purposes. In such cases, the regulatory scope shifts from the FADP to the **Medical Devices Ordinance (MedDO, SR 812.213)** or the **In Vitro Diagnostic Medical Devices Ordinance (IvDO, SR 812.219)**, corresponding to the EU's **MDR 2017/745** and **IVDR 2017/746**. Medical device classification is determined by intended use and requires conformity assessment, a Technical File, risk management under **ISO 14971**, and registration in **swissdamed**.

For non-medical genomic infrastructure, compliance with FADP and, where applicable, the **General Data Protection Regulation (GDPR)** is sufficient. GDPR may apply when data from EU subjects are processed (Art. 3). Voluntary alignment with international standards such as **ISO 27001** for information security and **ISO 15189** for quality management in medical laboratories is recommended to enhance transparency and interoperability across research, clinical, and industrial settings.

5.1 Accreditation standards

Accreditation provides verifiable assurance that laboratory and computational processes meet international quality benchmarks. The **ISO 15189** standard defines competence and quality management requirements for medical laboratories, encompassing both wet-lab and analytical bioinformatics activities. Laboratories accredited under ISO 15189 operate with validated procedures, documented workflows, and reproducible performance metrics suitable for clinical-grade diagnostics.

Within Switzerland, the *Health 2030 Genome Center* represents a reference implementation. Its ISO 15189-accredited Diagnostic Sequencing Platform (DSP) delivers whole-genome, whole-exome, and RNA-seq services under a certified quality system, including validated fast-turnaround workflows capable of data delivery within seven days of sample receipt. This demonstrates that high-quality accreditation and rapid clinical response can be achieved concurrently.

Accreditation should extend beyond laboratory sequencing to cover computational analysis, data management, and quality assurance frameworks. Validation, reproducibility testing, version control, and full documentation of pipelines must form part of the accredited quality management system. Laboratories and institutions that align with ISO 15189 or equivalent frameworks ensure that genomic data in Switzerland remain verifiable, comparable, and suitable for both clinical and research integration at national and international levels.

6 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) (1). 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 1: Recommendations for sequencing-based clinical bioinformatics. Reproduced based on Lavrichenko et al. (1) .

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.

7 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] 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

8 Supplemental

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