

Consensus guideline for verifiable genomic decision infrastructure

Quant Group¹, Dylan Lawless *², and the joint SGA committee³

¹The quantitative omic epidemiology group.

³The Swiss Genomics Association committee.

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

November 12, 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

Motivation: national genomic interoperability gap, reproducibility limits, manual interpretation burden.

Results: introduction of a unified three-pillar framework enabling end-to-end sample and data provenance, workflow-agnostic genomic evidence representation, and probabilistic variant evidence quantification with explicit accounting for absence of evidence.

Availability: non-commercial access, no mandatory registration, longevity greater than two years, links to schemas, APIs and validation suites.

For internal draft reference only:

Pillar 1 - [demo](#).

Pillar 2 demo - [demo](#).

Pillar 3 - [in-progress](#).

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

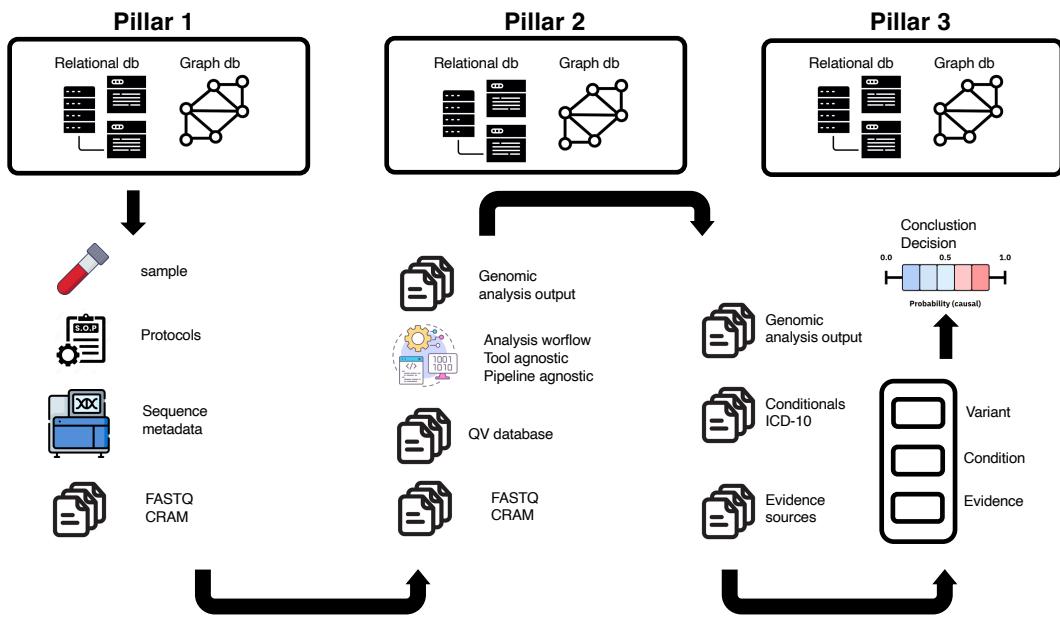


Figure 1:

Contents

1	Introduction	4
2	System design principles	5
3	System of record, system of meaning	5
4	Three pillar framework	6
4.1	pillar 1: semantic sample-to-sequence provenance	6
5	pillar 2: workflow-agnostic analysis representation	7
6	pillar 3: probabilistic variant evidence quantification	8
7	Unified three-pillar data contract	9
8	National integration and interoperability	10
9	Implementation and availability	10
10	Discussion	10
11	Conclusion	10
12	supplementary information	11
13	Supplemental	12

1 Introduction

National-scale genomics depends on shared evidence models, auditable provenance, and interoperable data contracts that remain stable across institutions, technologies, and time. Fragmentation across clinical, research, and commercial pipelines leads to non-equivalent evidence states, duplicated effort, and variable interpretation outcomes, limiting both clinical reliability and large-scale automation. A national framework must therefore serve multiple sectors simultaneously, while preserving traceability, reproducibility, and international compatibility.

The Swiss Genomics Association (SGA) brings together academic, clinical, healthcare,

industry, technical, regulatory, ethics, and public stakeholders to address this need through open, consensus-driven standards. SGA outputs are developed for voluntary adoption, informed by peer-reviewed evidence, and designed to support long-term national infrastructure, cross-border collaboration, and responsible innovation. The association’s remit spans the full lifecycle of genomic implementation, from sequence generation and data governance to variant interpretation, decision support, and clinical application.

To support a unified national approach, we propose a three-pillar architecture guided by SGA principles of persistence, transparency, sector inclusivity, and global alignment. This framework ensures that genomic evidence remains computable, auditable, and comparable across hospitals, universities, laboratories, industry partners, public institutions, and citizen-facing initiatives. By unifying how evidence is generated, structured, and interpreted, Switzerland can deliver a coordinated genomic strategy that supports precision medicine, research, clinical care, and innovation within a shared, future-ready foundation.

2 System design principles

Principles include workflow independence, machine-readable evidence, explicit uncertainty, full lineage, and no forced centralisation. Alignment includes SPHN, GA4GH, RDF semantics, and international schema harmonisation. Long-term design traits include versioned schemas, public reference implementations, auditability, graceful deprecation, decision traceability, and digital clinical readiness.

3 System of record, system of meaning

National genomic infrastructure must remain functional, auditable, and interpretable for decades, while biological meaning, file formats, knowledge graphs, and sequencing technologies evolve. A 20-year strategy must therefore separate the system of record from the system of meaning.

A national platform benefits from a hybrid model. Relational databases provide maximal stability for structured clinical and operational data where schemas are known, versioned, and query patterns are predictable. They offer mature tooling, strong governance controls, high transaction throughput, and wide talent availability.

They are a suitable choice for national patient records, audit logs, longitudinal timelines, test catalogues, quality metrics, consent records, and operational reporting.

In contrast, omics data and their interpretations form a dense, evolving knowledge network where schema flexibility, federated linking, identifier stability, and semantic reasoning are essential. Graph and ontology-based approaches are better suited for capturing biological relationships, evidence graphs, gene-disease networks, phenotypes, assay provenance, and cross-institution semantic alignment. These layers prioritise model extensibility rather than transaction performance and are optimally deployed as an interoperable knowledge tier rather than a replacement for core clinical databases.

This combination ensures long-term sustainability: structured records remain stable and performant, while biological meaning can evolve without schema migrations or system redesign. Clear synchronisation contracts between the two layers ensure reliability without conflating architectural responsibilities.

4 Three pillar framework

Pillar 1 provides standardised genomic provenance from wet sample to raw data. Pillar 2 enables normalised representation of analysis variables irrespective of tooling or pipeline. Pillar 3 delivers probabilistic variant evidence quantification for clinical-grade decision support. The data contract flows left to right and evidence trace flows right to left, and both are queryable, versioned, and reproducible.

4.1 pillar 1: semantic sample-to-sequence provenance

Pillar 1 guarantees an unbroken, technology-independent lineage from biosample acquisition to sequence generation, ensuring that any genome can be interpreted through a deterministic, auditable, and re-executable provenance record.

The model prioritises semantic interoperability rather than a single mandated serialisation, but uses explicit relationship graphs where appropriate to demonstrate portability across institutions. The reference implementation expresses provenance via RDF/Turtle to illustrate a fully machine-readable graph structure that aligns with national interoperability frameworks such as SPHN and demonstrates compatibility with clinical and research systems.

The provenance model captures core entities including sampling events, biospecimen handling, sequencing assays, instruments, runs, derived files, analysts, laboratory sites, quality metrics, and custody boundaries. Entities are represented using resolvable identifiers, enabling unambiguous tracking across institutions. The structural design builds on semantic approaches for clinical-genomic harmonisation (1) and aligns with FAIR practices for standardised vocabularies, persistent identifiers, and model transparency (? ? ?).

Relationships encode temporal order, processing dependencies, quality observations, and operational context using constrained predicates to avoid ambiguity and free-text interpretation. Shape constraints enforce topology, required fields, and referential integrity, ensuring that every genome has a provably complete lineage. Provenance metadata is stored alongside, but cryptographically separable from, primary sequence files, permitting governance workflows that inspect, validate, and exchange lineage without accessing genomic content.

Reference graphs, conformance profiles, and validation suites provide stable compliance targets for sequencing providers, hospitals, and analytic workflows. Automated validation includes identifier resolution, ontology mapping, temporal ordering, graph completeness, and schema contract testing, ensuring equivalent interpretation outcomes regardless of infrastructure, vendor, or analysis engine.

By treating provenance as an immutable semantic asset rather than a workflow artefact, pillar 1 ensures national interoperability, long-term reproducibility, and evidence traceability for clinical and research genomes across decades.

5 pillar 2: workflow-agnostic analysis representation

Pillar 2 operationalises the published qualifying variant (QV) model (2), which defines variant selection criteria as external, versioned, and pipeline-independent objects rather than embedded code heuristics. The core aim is to ensure that genomic evidence produced by different analytical toolchains can be expressed, exchanged, and evaluated as equivalent evidence statements, rather than non-comparable pipeline outputs.

The QV framework formalises filtering logic as declarative parameters, decoupled

from workflow execution and bound to immutable identifiers, enabling deterministic re-instantiation of variant selection across compute environments. Each QV record is an auditable analytic artefact, retaining rule provenance, parameter bounds, dataset context, callable region constraints, and tool-agnostic variable definitions. These records are content-hashed for integrity and referenced by persistent identifiers, permitting unambiguous evidence comparison across studies and institutions.

Pipeline portability is achieved by modelling QVs as normalised analytic constraints that apply equally to short-read, long-read, single-sample, and joint-calling workflows without altering underlying definitions. This permits equivalent interpretation of allele frequency thresholds, genotype quality, coverage bounds, inheritance filters, allelic balance constraints, and locus inclusion rules, regardless of whether computed by GATK, DeepVariant, DRAGEN, or alternative callers. Benchmarking in the original implementation demonstrated that QV-driven workflows reproduce conventional filtering results while improving reproducibility, inspection transparency, and cross-analysis comparability.

QV data contracts guarantee lossless serialisation of analytic intent, enabling independent validation, deterministic recomputation, and cross-tool evidence federation. The model supports selective disclosure of variant evidence sets without exposing full raw VCFs, aligning with federated data governance requirements. Compliance testing is supported through reference datasets, reproducibility benchmarks, portable schema definitions, and query validation suites.

By transforming QV logic into a portable representation layer, pillar 2 removes pipeline and software as hidden variables in genomic evidence generation. Downstream interpretation systems therefore receive a normalised evidence state, rather than pipeline-specific artefacts, enabling interoperability, reproducibility, and systematic evidence tracking at a national infrastructure scale.

6 pillar 3: probabilistic variant evidence quantification

Purpose is to replace heuristic interpretation with calibrated evidence models including unobserved or low-coverage states. Implementation includes likelihood-based evidence weights, explicit modelling of non-observation, evidence aggregation into clinical-grade posterior summaries, and GA4GH-aligned evidence

blocks. Validation includes calibrated truth sets, decision concordance benchmarks, and robustness evaluation across missing data regimes. Output guarantees auditable, machine-readable clinical decision evidence without manual reinterpretation.

7 Unified three-pillar data contract

This approach treats variant interpretation as structured evidence aggregation rather than categorical labelling. Evidence is encoded using GA4GH Variant Annotation (VA-Spec) and supporting Genomic Knowledge Standards such as the Variation Representation Specification (VRS), enabling machine-readable statements that link a variant, a condition, and evidence lines with full provenance. Each evidence item (e.g., population frequency, functional assays, segregation data, phenotype concordance, literature, or tool-derived scores) is modelled as a discrete, traceable object with defined method, direction of support, strength descriptor, and source identifier.

Rather than collapsing evidence into ACMG-style classes alone, VA-Spec records assertions as composable data structures, supporting quantitative or probabilistic frameworks layered above the annotation model. It allows heterogeneous evidence types, including cohort observations, MAVE assays, computational predictors, curated databases, and phenotype-driven prioritisation outputs from tools such as Exomiser or LIRICAL, to co-exist in a uniform schema. Assertions retain links to the originating agent and reference (laboratory, software, curator, publication), enabling auditability, reproducibility, and automated re-evaluation as new evidence is added.

This model separates evidence representation from clinical decision logic. VA-Spec captures *what* evidence exists and *how* it relates to a variant-disease claim, but does not determine *whether* the total evidence meets diagnostic thresholds. A clinical interpretation layer must therefore implement weighting, inheritance validation, phenotype concordance rules, disease mechanism checks, contradictory evidence handling, population prevalence alignment, and minimum reporting requirements before clinical conclusions are issued.

By storing variant evidence as structured, interoperable statements rather than static classifications, this framework enables national-scale knowledge aggregation, computational reassessment, cross-laboratory exchange, and long-term interpretability without loss of evidential detail or provenance.

Disease causality and variant pathogenicity are used for illustration, but the same

evidence framework applies to any variant classification, conditional on the causal claim being evaluated, including evidence supporting benign status or relevance to an alternative condition.

8 National integration and interoperability

The Swiss deployment model uses federated adoption without centralised control. Compatibility includes GA4GH, SPHN, HL7 FHIR Genomics, and international partner schemas. Governance is supported through public versioning, open test vectors, voluntary conformance, archival persistence, and transparent decision history under SGA stewardship.

9 Implementation and availability

Public endpoints provide schemas, APIs, and reference implementations. Access is non-commercial and free, with no required login. Browser-agnostic containerised reference deployments, validation toolkits, and example queries are provided. Maintenance is guaranteed for a minimum of two years, with versioned releases and deprecation timelines.

10 Discussion

Impacts include removal of manual reinterpretation, comparability across institutions, and clinical automation with traceable evidence. Limitations include initial integration cost, ecosystem alignment, and training requirements. Future scope includes national federation, longitudinal evidence accumulation, and automated regulatory evidence exports.

11 Conclusion

This work establishes a national framework for genomic evidence that prioritises provenance, interoperability, and quantified uncertainty. It defines a durable, tool-independent standard for variant interpretation that is applicable across clinical

care, research, healthcare systems, and industry. By aligning evidence representation, auditability, and decision logic, the framework enables consistent genomic interpretation across sectors and institutions in Switzerland, while maintaining compatibility with international standards for collaboration and data exchange.

12 supplementary information

S1 covers schema specifications for RDF, QV, and the evidence model. S2 includes validation reports and conformance tests. S3 provides example payloads, queries, and benchmarks. S4 describes the governance model and version history. S5 contains the reference implementation and deployment guide.

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] Eelke Van Der Horst, Deepak Unni, Femke Kopmels, Jan Armida, Vasundra Touré, Wouter Franke, Katrin Cramer, Elisa Cirillo, and Sabine Österle. Bridging Clinical and Genomic Knowledge: An Extension of the SPHN RDF Schema for Seamless Integration and FAIRification of Omics Data, December 2023.
- [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. May 2025. doi: 10.1101/2025.05.09.25324975.

Acronyms

SGA	Swiss Genomics Association	4
-----	----------------------------	---

13 Supplemental

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