A twelve-state notation for genotype-phenotype interpretation

Canonical semantics for evidence, inference, reporting, and audit

Dylan Lawless

Version 1.0 6 September 2026

Abstract

Genotype-phenotype interpretation often conflates biological truth, what an assay observed, how evidence was interpreted, and what was ultimately reported. We separate these quantities and define a finite audit state for a specified genotype-phenotype hypothesis. The state has three coordinates: latent explanatory truth, resolution of the genetic state required to evaluate the hypothesis, and interpretive status. Their Cartesian product contains twelve states,

$$\mathcal{S}_{\text{audit}} = \{0, 1\} \times \{\text{R}, \text{U}\} \times \{\text{C}, \text{D}, \text{N}\},$$

which are represented by compositional codes such as 1RC, 0UD, and 1UN. The code describes an audit configuration; it is not a posterior probability or a clinical decision. For each sufficiently specified causal hypothesis, the framework posits an observer-independent truth value; evidence and posterior probability quantify uncertainty about that truth rather than constituting it. Bayesian inference, measurement, reporting, and model versioning are defined separately. This document fixes the terminology and notation required to use the framework reproducibly.

1 Introduction

A genotype-phenotype result can fail for different reasons that are hidden by a standard true-positive, true-negative, false-positive, and false-negative table. A genetic explanation may be true but technically unresolved, resolved but misinterpreted, or unresolved because no interpretable evidence exists. Conversely, a non-explanatory hypothesis may be correctly rejected from resolved evidence or merely appear negative because relevant evidence is missing. These distinctions matter because they imply different probabilities, failure mechanisms, and routes to correction.

The framework therefore separates the path

hypothesis $\longrightarrow$ latent truth $\longrightarrow$ measurement and resolution $\longrightarrow$ interpretation $\longrightarrow$ report $\longrightarrow$ audit.

Only three coordinates define the twelve-state audit notation:

$$\boxed{\mathbf{s}_h = (T_h, Z_h, I_h)}$$

for a specified genotype-phenotype hypothesis H_h. Everything downstream, including a clinical report and the labels TP, TN, FP, FN, or no call, is kept separate.

Here, T_h denotes the *latent explanatory truth*: whether the sufficiently specified hypothesis H_h is actually true or false, independently of what is currently known or believed. It is not a database classification, expert consensus, or scientific conclusion. The assumption that such an observer-independent truth exists is the *axiomatic ground-truth postulate*. These concepts are defined below and the necessity for definition is developed further in **Section 6**.

Running example. Some examples are included so that readers can map to more familiar terms to see how we generalise. In most we use a hypothetical patient with cystic fibrosis and two candidate *CFTR* variants, V_1 and V_2. The case is illustrative: the variants, evidence states, and outcomes are chosen only to demonstrate the notation.

2 Conceptual foundations

Phenotype context. Let D denote the phenotype, disease, trait, response, or clinical context being explained.

Example. Here, D is the patient's cystic-fibrosis phenotype. In another study, D could be a measured LDL-cholesterol concentration. More generally, D specifies the phenotypic context being explained, independent of how it is represented or measured.

Reference model. Let $\mathcal{R}^{(v)}$ denote a versioned reference model. It is the specification required to make an interpretation reproducible and may contain phenotype definitions, transcripts, admissible event classes, inheritance assumptions, penetrance assumptions, population models, phase rules, assay scope, evidence sources, interpretation rules, thresholds, software, database snapshots, and standard operating procedures. The superscript v identifies a coherent version of that specification.

The reference model defines the interpretation problem; it does not define biological truth. Updating $\mathcal{R}^{(v)}$ may change an inferred probability or report without changing the underlying biological reality.

Example. For this case, $\mathcal{R}^{(v)}$ might specify *CFTR* transcript `NM_000492.4`, recessive inheritance, phase requirements, assay adequacy criteria, interpretation rules, and the database and software versions used.

Genotype-phenotype hypothesis. Let H_h be a precisely specified proposition about the explanatory role of a genetic event or genetic configuration for phenotype context D. The proposition should state the causal scope intended by the word "explanatory", for example a complete monogenic explanation, a necessary component of a recessive genotype, or a contributing component of a more complex causal configuration.

The unit of inference is therefore not necessarily a single sequence variant. A hypothesis may concern a dominant allele, a homozygous genotype, a compound heterozygous pair, a phased haplotype, a copy-number event, a structural rearrangement, a repeat expansion, mosaicism, or a multi-locus configuration.

Example. Suppose whole-genome sequencing identifies candidate variants V_1 and V_2 in *CFTR*. A hypothesis H_h is that V_1 and V_2, present in trans, form the recessive genetic configuration explaining the patient's phenotype.

Latent explanatory truth and the ground-truth postulate. The framework begins with a simple assumption: once a causal question is specified precisely, there is a fact about whether it is true, even if we cannot determine that fact with certainty.

For hypothesis H_h, define

$$T_h \in \{0, 1\},$$

where $T_h = 1$ if H_h is true and $T_h = 0$ if it is false.

The word *truth* is used deliberately. T_h is intended to represent the underlying biological reality itself: what is actually true in the world, independently of any observer, database, or model. T_h is not an expert consensus, database classification, accepted association, or current best judgement. Those are sources of evidence about the truth. They do not define the truth itself.

The framework therefore adopts an *axiomatic ground-truth postulate*: once H_h and its causal meaning are fixed, its truth value is a property of the underlying biological system and does not depend on what an observer knows about it.

T_h is *latent* because its value is generally not directly observable and may never be known with certainty. Scientific evidence changes our knowledge of T_h, not T_h itself.

$$\boxed{\text{causal truth} \neq \text{evidence about causal truth} \neq \text{belief about causal truth}}$$

Conceptually, an ideal observer with complete knowledge of the relevant causal system would know T_h. This does not require deterministic biology: H_h may itself describe a probabilistic causal effect.

The distinction between the latent explanatory truth T_h, the admissible events considered by the reference model, and the evidence used to infer T_h (probative evidence) is developed further in **Section 6**.

Example. Suppose a patient with cystic fibrosis has two candidate *CFTR* variants, V_1 and V_2. There is a real answer to whether this genetic configuration truly explains the disease: $T_h = 1$ if it does and $T_h = 0$ if it does not.
Even if many studies show a very strong association and a database classifies the genotype as "causal", that classification is still part of our evidence and model, not the ground truth itself. If the separate, unknown *axiomatic ground truth* is omitted, the model effectively assumes that current knowledge is complete and leaves no possibility that its conclusion is wrong. Keeping T_h separate preserves that possibility and allows the model to represent uncertainty about reality rather than equating the model with reality.

Genetic state, observation, and measurement. Let G_h denote the latent genetic configuration relevant to H_h, X_h the observed evidence, and M_h the measurement process. A generic measurement model is

$$P(X_h \mid G_h, M_h).$$

This layer accommodates coverage, mapping uncertainty, genotype quality, sensitivity, specificity, phasing, assay blind spots, and other properties that determine what can be observed.

Example. Suppose the patient's true *CFTR* sequence contains an $G \to A$ change relative to transcript `NM_000492.4`, producing a protein change such as `p.Gly551Asp`. That underlying DNA state exists whether or not we measure it correctly. A sequencing test might instead return only the reference base G, for example because the altered allele was not captured or confidently called. The true sequence, the sequence we observe, and the method used to observe it are therefore three different things.

Resolution. Let

$$Z_h \in \{\mathrm{R}, \mathrm{U}\},$$

where R means *resolved* and U means *unresolved*. Operationally,

$$Z_h = z(H_h, X_h, M_h, \mathcal{R}^{(v)}),$$

where $z(\cdot)$ is the versioned resolution rule. A hypothesis is resolved when the genetic state required to evaluate it has been determined to the pre-specified level of adequacy. It is unresolved when that state remains indeterminate because of missing data, insufficient quality, incomplete phase, assay limitation, incompatible representation, or another measurement deficit.

A resolved state is not necessarily alternate, pathogenic, or explanatory. A confidently observed reference genotype is resolved evidence. Thus,

$$\boxed{\text{resolved negative evidence} \neq \text{unresolved evidence.}}$$

Example. Suppose the hypothesis requires knowing whether the patient carries the *CFTR* deletion `NM_000492.4:c.1521_1523delCTT` (`p.Phe508del`). If high-quality sequence reads clearly show that the deletion is absent, that part of the genetic state is *resolved*, even though the result is negative. If the same region has insufficient coverage and we cannot tell whether the deletion is present or absent, it is *unresolved*. "Not present" and "not known" are therefore different results. A report such as "sequence missing", "result unavailable", or "could not be determined" should therefore be treated as unresolved, not as evidence that the variant is absent.

Interpretive state. Let

$$I_h \in \{\mathrm{C}, \mathrm{D}, \mathrm{N}\}.$$

Here C means that the interpretation of the available evidence is *concordant* with the specified audit adjudication under $\mathcal{R}^{(v)}$; D means that it is *discordant*; and N means that there is *no interpretable evidence* sufficient to support such an interpretation.

The third category is intentionally different in kind from the first two. Concordant and discordant describe the fidelity of an interpretation. N records that the evidential basis required for that interpretation is absent. Consequently, no interpretable evidence is not the same as a negative interpretation and does not necessarily imply that no report was issued.

Resolution and interpretability are distinct. $Z_h = \mathrm{U}$ does not imply $I_h = \mathrm{N}$. An unresolved genetic state may itself be handled concordantly or discordantly; N instead records that there is no sufficient evidential basis for an interpretation.

Example.
Suppose sequencing identifies two well-established disease-causing *CFTR* variants, `p.Phe508del` and `p.Gly551Asp`, on opposite copies of the gene. Under the specified review rules, the available evidence supports interpreting this pair as an explanation for cystic fibrosis.
If the analyst interprets the evidence in the way expected under those review rules, the interpretation is *concordant* ($I_h = \mathrm{C}$). In simple terms, the evidence has been understood in the intended way.
If the same evidence is interpreted differently from the specified adjudication, the interpretation is *discordant* ($I_h = \mathrm{D}$). In simple terms, the available evidence has been interpreted incorrectly relative to the stated review rules.
If instead the second variant were poorly understood and there were not enough evidence to decide what it means, the case would have *no interpretable evidence* ($I_h = \mathrm{N}$).
Thus, "interpreted as non-explanatory" and "not enough evidence to interpret" are different outcomes.

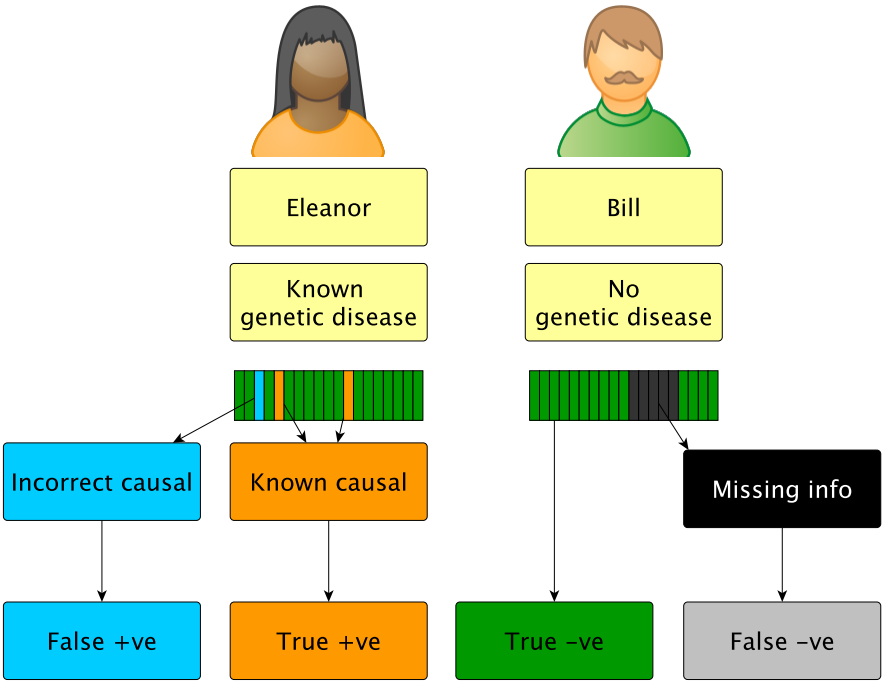

Figure 1: **Illustrative patient-level interpretation routes.** Representative examples showing explanatory, non-explanatory, and unresolved genetic findings, with compact audit-state codes shown as a secondary technical layer.

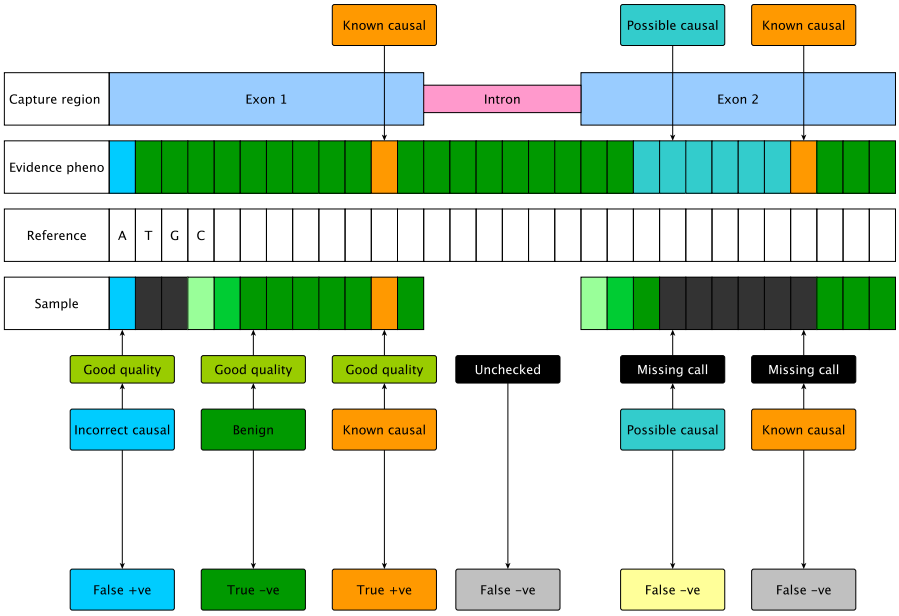

Figure 2: **Representative genomic evidence routes.** Examples separating underlying explanatory truth, observed sequence, interpretation, and downstream audit outcome. Missing sequence is represented separately from resolved negative evidence.

3 The twelve-state audit formalism

The twelve states comprise all combinations permitted by the three-coordinate audit formalism: explanatory truth, resolution of the required genetic state, and interpretive state. The describe whether the required genetic information has been resolved, and how the evidence can be interpreted.

One part of the state is the latent explanatory truth T_h: whether the genetic hypothesis really explains the phenotype. Because this true answer is usually unknown in a real patient, the complete audit state is usually also unknown at the time of analysis.

Figures 3 and **4** show the same twelve states in two complementary layouts: first as a branching process and then as a compact square arrangement. Both figures list the *final call* made from the available evidence and model. It is kept separate from the underlying ground truth downstream.

This distinction is important because the final call is what can be reported in practice, while the underlying truth may remain unknown.

For a specified hypothesis H_h, the canonical audit state is

$$\mathbf{s}_h = (T_h, Z_h, I_h),$$

where T_h is the latent explanatory truth defined above.

The state space is

$$\mathcal{S}_{\text{audit}} = \{0, 1\} \times \{\text{R}, \text{U}\} \times \{\text{C}, \text{D}, \text{N}\},$$

so

$$|\mathcal{S}_{\text{audit}}| = 2 \times 2 \times 3 = 12.$$

3.1 Compositional state codes

The compact state code is a short way to describe an audit state. It has three parts, read from left to right: latent explanatory truth, resolution, and interpretive state.

$$\underbrace{0/1}_{\text{explanatory truth}} \qquad \underbrace{\text{R/U}}_{\text{resolution}} \qquad \underbrace{\text{C/D/N}}_{\text{interpretation}}$$

The symbols have the following formal and plain-language meanings:

0	non-explanatory truth	the stated causal hypothesis is false
1	explanatory truth	the stated causal hypothesis is true
R	resolved	the required genetic state is known well enough
U	unresolved	the required genetic state is not known well enough
C	concordant	interpretation agrees with the specified audit adjudication
D	discordant	interpretation disagrees with the specified audit adjudication
N	no interpretable evidence	there is not enough evidence to interpret

The first character refers to the latent explanatory truth T_h, not to our current belief about it. The existence of this observer-independent truth value is the *axiomatic ground-truth postulate* defined above. This distinction is discussed further in Section 6.

For C and D, "correctly" and "incorrectly" refer to the interpretation of the available evidence, not to certainty about T_h. This distinction is necessary because T_h cannot be directly observed; we can only judge the available evidence and its interpretation.

Formally,

$$\text{code}(\mathbf{s}_h) \in \{\mathsf{0RC}, \mathsf{0RD}, \mathsf{0RN}, \mathsf{0UC}, \mathsf{0UD}, \mathsf{0UN}, \mathsf{1RC}, \mathsf{1RD}, \mathsf{1RN}, \mathsf{1UC}, \mathsf{1UD}, \mathsf{1UN}\}.$$

Each character has a fixed meaning, so the state can be reconstructed directly from the code.

In prospective analysis, the analyst does not usually know T_h. What can be reported is the conclusion supported by the available evidence and model, later defined formally as the *final call* $\widehat{T}_h$. The final call is therefore kept separate from the latent ground truth.

Cardinality is not probability. The twelve states enumerate distinct audit configurations; they are not assumed to occur with probability 1/12, and the framework does not impose a uniform prior over them.

In derivations, we use $\mathbf{s}_h = (T_h, Z_h, I_h)$ when the coordinates themselves are being manipulated. In figures, tables, simulations, and retrospective audit, the compact code provides a convenient shorthand. For example, $\mathsf{1RC}$ is shorthand for $\mathbf{s}_h = (1, \mathrm{R}, \mathrm{C})$.

Because the first character is the latent truth value T_h, a complete state code should not ordinarily be used as a prospective patient result. A prospective report instead records the evidence-based final call $\widehat{T}_h$, together with relevant quantities such as resolution and posterior support.

The codes are compositional by design. They should not be replaced by twelve independent names, arbitrary Greek letters, or sequential scenario numbers, because the three coordinates are scientifically meaningful and directly recoverable from the notation.

The representation preserves two distinctions that are particularly easy to lose in ordinary classification. First, resolved negative evidence is different from unresolved evidence. Second, absence of supporting evidence is different from evidence against a hypothesis. The same absence of evidence may be compatible with a non-explanatory truth yet conceal an explanatory truth. This asymmetry is one reason TP, TN, FP, and FN alone are insufficient to describe the process.

Table 1: **Canonical genotype–phenotype audit state codes.** The code is read from left to right as latent explanatory truth, resolution, and interpretive state. For example, 0RC corresponds formally to $\mathbf{s}_h = (0, \mathrm{R}, \mathrm{C})$.

Code	Plain-language meaning
0RC	Does not explain the phenotype; genetic state resolved; interpretation concordant.
0RD	Does not explain the phenotype; genetic state resolved; interpretation discordant.
0RN	Does not explain the phenotype; genetic state resolved; no interpretable evidence.
0UC	Does not explain the phenotype; genetic state unresolved; interpretation concordant.
0UD	Does not explain the phenotype; genetic state unresolved; interpretation discordant.
0UN	Does not explain the phenotype; genetic state unresolved; no interpretable evidence.
1RC	Explains the phenotype; genetic state resolved; interpretation concordant.
1RD	Explains the phenotype; genetic state resolved; interpretation discordant.
1RN	Explains the phenotype; genetic state resolved; no interpretable evidence.
1UC	Explains the phenotype; genetic state unresolved; interpretation concordant.
1UD	Explains the phenotype; genetic state unresolved; interpretation discordant.
1UN	Explains the phenotype; genetic state unresolved; no interpretable evidence.

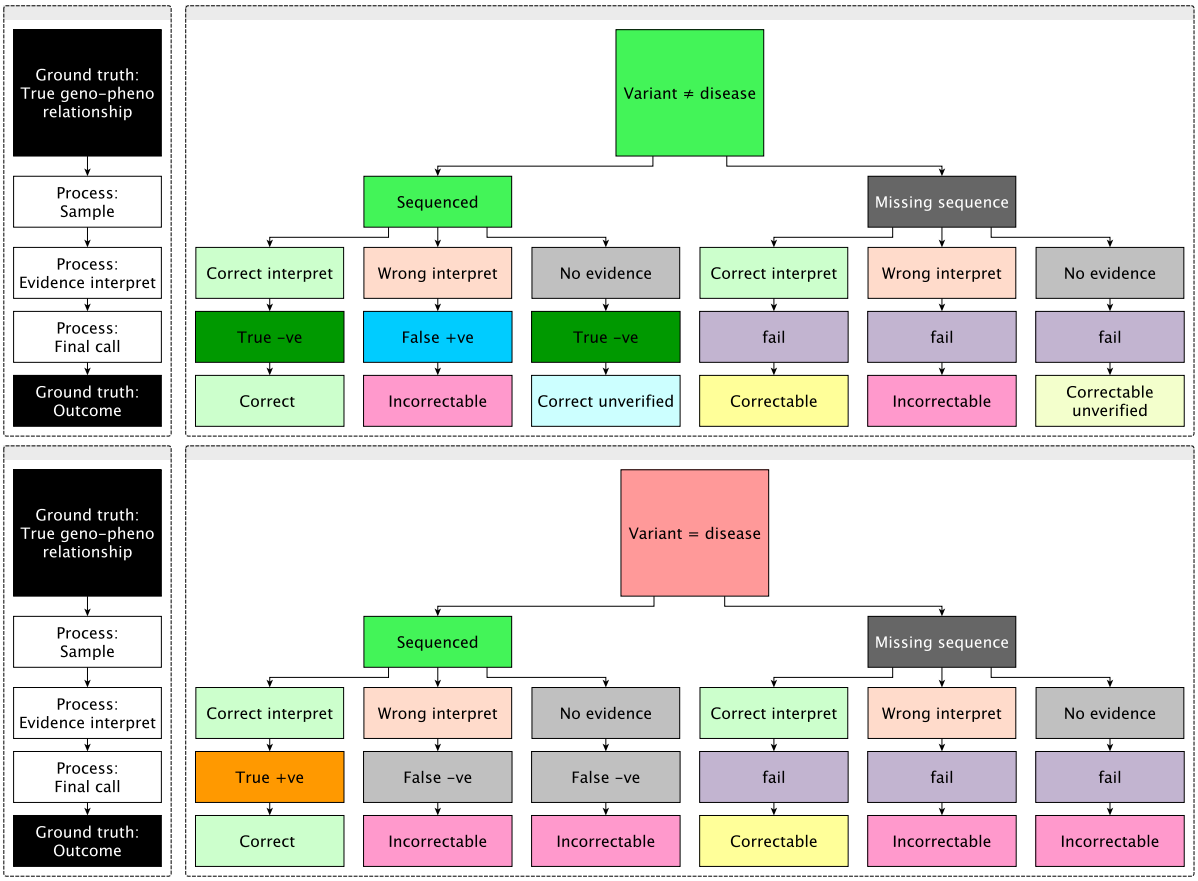

Figure 3: **Tree representation of the genotype–phenotype audit routes.** The figure separates the latent explanatory ground truth T_h, resolution Z_h, interpretation I_h, and the *final call* $\widehat{T}_h$. The final call is the conclusion produced from the available evidence and model; it is not the same as the underlying ground truth. Comparing the final call with T_h gives familiar audit outcomes such as true positive, false positive, true negative, and false negative.

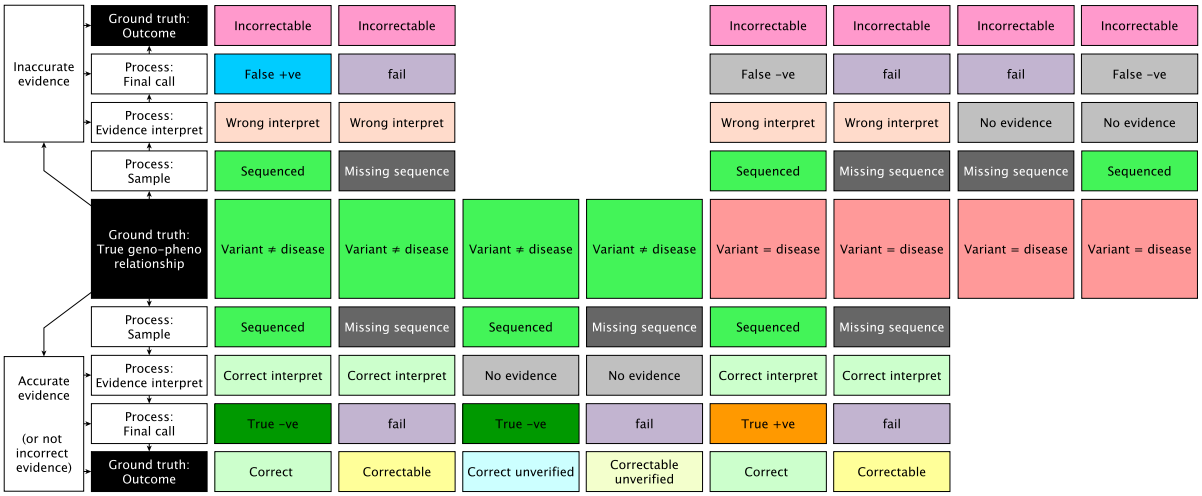

Figure 4: **Square representation of the twelve genotype–phenotype audit states.** The twelve states are the combinations of axiomatic explanatory truth T_h, resolution Z_h, and interpretive state I_h. The figure also shows the separate final call $\widehat{T}_h$, making clear that the scientist's reported conclusion and the latent ground truth are different quantities.

4 Bayesian inference over explanatory hypotheses

We use X for the complete evidence set available to an inference, X_i for an individual evidence component, and X_h for the observed evidence relevant to hypothesis H_h.

4.1 Prior plausibility, likelihood, and posterior attribution

Three probabilities answer different questions:

$$P(H_h \mid D, \mathcal{R}^{(v)})$$

is the prior plausibility of H_h;

$$P(X \mid H_h, D, \mathcal{R}^{(v)})$$

is the likelihood of observing evidence X if H_h were true; and

$$P(H_h \mid X, D, \mathcal{R}^{(v)})$$

is the posterior attribution after observing the evidence.

Bayes' rule connects them:

$$P(H_h \mid X, D, \mathcal{R}^{(v)}) = \frac{P(X \mid H_h, D, \mathcal{R}^{(v)})P(H_h \mid D, \mathcal{R}^{(v)})}{P(X \mid D, \mathcal{R}^{(v)})}.$$

In plain terms, a genetic explanation may be biologically possible but unlikely to occur; it may occur but be difficult to observe; and an observed finding may still fail to be the best explanation for the phenotype. Prior plausibility, probability of observation, and posterior attribution must therefore remain separate.

4.2 Hypothesis space and residual hypotheses

For mutually exclusive model selection, let

$$\mathcal{H} = \{H_\varnothing, H_{\text{other}}, H_1, \ldots, H_J\},$$

where $H_1, \ldots, H_J$ are enumerated genotype-phenotype hypotheses, $H_\varnothing$ denotes no genetic explanation within the defined causal domain, and H_{other} denotes a genetic explanation outside the enumerated set. These classes must form a partition if posterior probabilities are normalised by a simple sum:

$$\bigcup_{H \in \mathcal{H}} H = \Omega, \qquad H_g \cap H_h = \varnothing \quad \text{for } g \neq h.$$

Then

$$P(H_h \mid X, D, \mathcal{R}^{(v)}) = \frac{P(X \mid H_h, D, \mathcal{R}^{(v)})P(H_h \mid D, \mathcal{R}^{(v)})}{\sum_{H_g \in \mathcal{H}} P(X \mid H_g, D, \mathcal{R}^{(v)})P(H_g \mid D, \mathcal{R}^{(v)})}.$$

If causal explanations can coexist, the hypotheses must instead represent complete causal configurations or the model must use a multivariate latent state. Overlapping hypotheses should not be inserted into a mutually exclusive normalising sum.

5 Reporting and retrospective audit

A prospective report is generated from information available at the time of interpretation and therefore cannot depend on the latent truth T_h. Let

$$\widehat{T}_h \in \{0, 1, \varnothing\},$$

where 1 denotes an explanatory report, 0 a non-explanatory report, and $\varnothing$ no binary call. A general reporting policy is

$$\widehat{T}_h = d_\pi(X, D, \mathcal{R}^{(v)}),$$

where π contains the applicable thresholds and decision rules.

When T_h is known or stipulated, as in simulation or an oracle benchmark, the classical categories are

$$Q_h = \begin{cases} \text{TP}, & T_h = 1,\ \widehat{T}_h = 1, \\ \text{FN}, & T_h = 1,\ \widehat{T}_h = 0, \\ \text{FP}, & T_h = 0,\ \widehat{T}_h = 1, \\ \text{TN}, & T_h = 0,\ \widehat{T}_h = 0, \\ \text{NC}, & \widehat{T}_h = \varnothing. \end{cases}$$

TP, TN, FP, FN, and NC are therefore report-level audit projections, not primitive states of genetic causation.

In empirical studies, a reference adjudication may be used as a proxy for T_h, but it should not be identified with axiomatic ground truth unless that equivalence is an explicit assumption. Any substitution of an adjudicated proxy for T_h should therefore be stated explicitly.

A complete compact state code contains T_h. It should therefore be used as a latent-state or retrospective audit notation, not as a prospective assertion of truth when T_h is unknown. Prospective systems should report the posterior, the report decision, resolution status, or other quantities that are actually observed or inferred.

6 Ground-truth postulate, admissible events, and probative evidence

Three concepts must remain separate. The latent explanatory truth T_h concerns what is actually true of a sufficiently specified causal hypothesis; admissibility concerns what the reference model agrees to consider; and probative evidence concerns what changes our belief about competing hypotheses. In compact form,

latent explanatory truth	T_h	underlying causal truth
admissible events	$\mathcal{E}_{\text{adm}}$	modelled candidate universe
probative evidence	X_i	evidence that changes belief

Axiomatic ground-truth postulate. The axiomatic ground-truth postulate and latent explanatory truth T_h are defined above and are recalled here. For a fixed, sufficiently specified H_h, the framework assumes that there is an observer-independent fact about whether the proposition is true. This is the *axiomatic ground-truth postulate*, and the corresponding latent explanatory truth is T_h.

A database assertion, expert consensus, reference adjudication, or model output may provide evidence about T_h, but none is T_h itself. Even a conclusion that is widely accepted scientifically remains part of our evidence and knowledge, rather than the underlying truth it is intended to represent.

Admissible genetic events. The modelled candidate universe is a separate concept. Let

$$\mathcal{E}_{\text{adm}}(D, \mathcal{R}^{(v)}) = \{e_1, \ldots, e_K\}$$

denote the genetic events admitted for consideration under phenotype context D and reference model $\mathcal{R}^{(v)}$. Membership in $\mathcal{E}_{\text{adm}}$ states only that an event is eligible to participate in a modelled hypothesis; it does not assert that the event is present, pathogenic, causal, or explanatory in the case.

Complete genotype–phenotype hypotheses may require one or more admissible events. A hypothesis-construction rule may therefore be written

$$\mathcal{H}_{\text{enum}} = \Phi\Big(\mathcal{E}_{\text{adm}}(D, \mathcal{R}^{(v)}), D, \mathcal{R}^{(v)}\Big) = \{H_1, \ldots, H_J\}.$$

The number of admissible events need not equal the number of complete hypotheses.

We have previously demonstrated the handling of admissible genetic events, by the practical processing and implementation of qualifying variants (Lawless, et al. Bioinformatics 2026, doi:10.1093/bioinformatics/btaf676).

Probative genetic evidence. *Probative genetic evidence* is evidence that changes the relative support for competing hypotheses. For two specified hypotheses H_h and H_a, the evidential contribution of X_i can be written as a Bayes factor,

$$\mathrm{BF}_{ih,a} = \frac{P(X_i \mid H_h, D, \mathcal{R}^{(v)})}{P(X_i \mid H_a, D, \mathcal{R}^{(v)})}.$$

The posterior odds are

$$\frac{P(H_h \mid X_i, D, \mathcal{R}^{(v)})}{P(H_a \mid X_i, D, \mathcal{R}^{(v)})} = \mathrm{BF}_{ih,a} \frac{P(H_h \mid D, \mathcal{R}^{(v)})}{P(H_a \mid D, \mathcal{R}^{(v)})}.$$

Thus $\mathrm{BF}_{ih,a} = 1$ provides no pairwise discrimination. A value above one favours H_h, while a value below one favours H_a. Evidence against a hypothesis can therefore be probative. An implementation may define a practical threshold such as $|\log \mathrm{BF}_{ih,a}| > \epsilon$ when negligible changes should be treated as operationally non-probative.

Probative evidence changes our relative belief about causal hypotheses. It may become stronger or weaker as evidence accumulates, but it does not become the latent explanatory truth T_h. The ground-truth postulate concerns what is actually true; probative evidence concerns what the available evidence allows us to infer about it.

7 Versioning, semantic stability, and use

Every posterior probability, resolution assessment, interpretation, and report must be traceable to the reference model under which it was produced. Results from different $\mathcal{R}^{(v)}$ versions should not be pooled as though they came from one contemporaneous model unless the relevant evidence and assumptions have been harmonised or recalibrated. If a version change alters the proposition represented by H_h, the revised proposition should receive a new or explicitly revised hypothesis identifier.

The meanings of the twelve state codes are version-stable. Future extensions may add coordinates or richer evidence models, but they should not silently change the meanings of 0RC through 1UN. A semantic change to an established code requires a new version of this specification.

Authors should use "resolved" and "unresolved" rather than "sequenced" and "missing sequence" unless sequencing itself is the subject. "No interpretable evidence" should not be used as a synonym for "no report". TP, TN, FP, and FN should be reserved for retrospective comparison of a report with explanatory truth, not used as synonyms for pathogenic, benign, causal, non-causal, observed, or absent.

8 Notation reference

The framework begins from a simple assumption: reality exists independently of our measurements, models, beliefs, or conclusions. We call this the *axiomatic ground-truth postulate* and denote it, when needed, by A_{GT}.

For a sufficiently specified causal hypothesis H_h,

$$\mathsf{A}_{\mathrm{GT}} \quad \Longrightarrow \quad T_h \in \{0, 1\},$$

where T_h is the *latent explanatory truth*: what is actually true of that hypothesis. We refrain from operating on this A_{GT} notation formally but include it here as we expect to subsequently

use it as a mathematical quantity in future work. The remaining terms below describe what we can observe, model, infer, interpret, and report about that underlying reality.

Table 2: **Canonical notation.**

Symbol	Name	Meaning
$\mathcal{R}^{(v)}$	Reference model	Versioned specification of the interpretation problem.
D	Phenotype context	Disease, phenotype, trait, response, or other context being explained.
H_h	Genotype–phenotype hypothesis	A specified proposition about the explanatory role of a genetic event or configuration for D.
T_h	Latent explanatory truth	Observer-independent truth value of H_h, with $T_h \in \{0, 1\}$. It represents what is actually true, not an evidence-based judgement about it.
A_{GT}	Axiomatic ground-truth postulate	Framework postulate that, for a fixed and sufficiently specified H_h, a latent explanatory truth T_h exists independently of what is observed, inferred, believed, or currently accepted scientifically.
G_h	Latent genetic configuration	Underlying molecular or genotype state relevant to H_h.
X	Evidence set	Complete set of evidence available to an inference.
X_i	Evidence component	An individual component of X, used when describing the contribution of particular evidence.
X_h	Hypothesis-relevant observed evidence	Observed evidence relevant to hypothesis H_h.
M_h	Measurement process	Assay and measurement properties governing what can be observed.
Z_h	Resolution	R for resolved and U for unresolved, according to whether the genetic state required to evaluate H_h is known sufficiently well.
I_h	Interpretive state	C concordant, D discordant, and N no interpretable evidence, referring to the interpretation of the available evidence.
$\mathbf{s}_h$	Audit state	Canonical three-coordinate state (T_h, Z_h, I_h).
$\mathrm{code}(\mathbf{s}_h)$	State code	Compact compositional notation for $\mathbf{s}_h$, such as $\mathsf{1RC}$.
$\widehat{T}_h$	Final call	Evidence-based report decision: explanatory, non-explanatory, or no binary call. It is distinct from the latent explanatory truth T_h.
Q_h	Classical audit projection	TP, TN, FP, FN, or NC obtained by comparing the final call with T_h when T_h is known or stipulated.
$\mathcal{E}_{\mathrm{adm}}$	Admissible event set	Genetic events included in the modelled candidate universe. Membership does not imply that an event is present, pathogenic, causal, or explanatory.
$\mathcal{H}$	Hypothesis space	Enumerated and residual genotype–phenotype hypotheses used for inference.

Table 3: **Practical routes through the twelve canonical audit states.** The table is read from left to right, from latent explanatory truth through resolution and interpretation to the reported final call and its audit outcome. The final-call and audit columns show a simple practical reporting route; formally, the final call $\widehat{T}_h$ is defined separately by the reporting policy. We use the simple labels 'Variant $\neq$ phenotype' and 'Variant = phenotype' for readability. More precisely, these correspond to $T_h = 0$, meaning the specified hypothesis H_h is false, and $T_h = 1$, meaning H_h is true. This is because H_h may refer to a genetic configuration rather than a single variant, such as compound heterozygosity, a structural variant, a multi-locus configuration, or a probabilistic causal effect.

Code	Latent explanatory truth	Resolution (genetic state)	Interpretation (by the analyst)	Final call (reported conclusion)	Audit outcome (call vs truth)
0RC	Variant $\neq$ phenotype	Resolved	Correct interpretation	Non-explanatory	True negative
0RD	Variant $\neq$ phenotype	Resolved	Wrong interpretation	Explanatory	False positive
0RN	Variant $\neq$ phenotype	Resolved	Not enough evidence	No binary call	No call
0UC	Variant $\neq$ phenotype	Unresolved	Correctly recognised as unresolved	No binary call	No call
0UD	Variant $\neq$ phenotype	Unresolved	Wrong interpretation	Explanatory	False positive
0UN	Variant $\neq$ phenotype	Unresolved	Not enough evidence	No binary call	No call
1RC	Variant = phenotype	Resolved	Correct interpretation	Explanatory	True positive
1RD	Variant = phenotype	Resolved	Wrong interpretation	Non-explanatory	False negative
1RN	Variant = phenotype	Resolved	Not enough evidence	No binary call	No call
1UC	Variant = phenotype	Unresolved	Correctly recognised as unresolved	No binary call	No call
1UD	Variant = phenotype	Unresolved	Wrong interpretation	Non-explanatory	False negative
1UN	Variant = phenotype	Unresolved	Not enough evidence	No binary call	No call

9 Discussion

The principal contribution of this work is a reference formalism that keeps latent explanatory truth, measurement and resolution, interpretation, reporting, and retrospective audit distinct during genotype–phenotype interpretation. The twelve-state notation is not itself a causal estimator, variant classifier, or clinical decision rule. Rather, it provides a common representation in which different sources of uncertainty and error remain identifiable.

The latent explanatory truth T_h is the central inferential target of the framework. Evidence changes our belief about T_h, rather than defining it. The accuracy of that inference therefore depends on the adequacy of the available evidence, priors, likelihoods, measurement model, and reference model. In empirical studies, adjudicated reference standards may be useful for evaluation, but they remain evidence about T_h unless their equivalence to ground truth is explicitly assumed.

Several quantities are deliberately left implementation-specific. The framework does not prescribe how admissible events are selected, hypotheses are constructed, priors or likelihoods are estimated, sufficient resolution is defined, or posterior evidence is converted into a report. Likewise, the twelve states are complete with respect to the three coordinates defined here, not with respect to all biological or clinical complexity. Extensions may therefore introduce richer measurement, causal, or adjudication models while preserving the meanings of the canonical state codes.

Future work should test how accurately and reproducibly evidence moves posterior belief towards T_h. Simulation and oracle settings can evaluate inference when truth is known by construction, while empirical genomic studies can assess calibration, reproducibility, auditability, and error attribution under real reference models. Of particular interest is whether separating resolved negative evidence from unresolved evidence, and interpretive discordance from insufficient evid-

ence, reveals failure modes that are obscured by conventional TP, TN, FP, and FN summaries.