

Research Article

SchizoCircuit: An Auditable, Claim-Governed Nextflow Workflow for Synthetic GWAS-to-Gene Evidence Integration

Regio Marcos Pinto Abreu Filho¹

1. Psychiatry Department, Military Police of the State of Rio de Janeiro (PMERJ), Brazil

Background. Locus-to-gene prioritization can combine association statistics, fine mapping, molecular QTL evidence, regulatory annotation, and literature-derived context, but heterogeneous evidence layers are vulnerable to double counting, hidden thresholds, non-reproducible execution, and claims that exceed what the data support. We developed SchizoCircuit as a claim-governed, auditable workflow motivated by schizophrenia genomics but evaluated only on frozen synthetic engineering controls.

Methods. The Nextflow DSL2 workflow comprises ten modules for GWAS ingest, quality control, supplied-LD validation, Wakefield approximate-Bayes-factor fine mapping with an explicit null state, QTL harmonization, single-causal colocalization with prior sensitivity analysis, evidence-graph construction, policy-based grading, candidate-card generation, and run metadata. A versioned policy defines G0-G4 evidence grades and forbids biological or clinical promotion from synthetic inputs. The frozen benchmark contains positive, ambiguous, and null loci. Reproducibility was tested across Windows and Linux, Python and Nextflow engines, Docker clean runs, resume/cache behavior, unit and workflow contracts, manifests, software bills of materials, and hash-locked package payloads.

Results. All supplied LD matrices passed structural validation. The positive control produced a single-variant 95% credible set, posterior locus causality 0.99999, robust QTL colocalization, and a G2 single-leading-gene status. The ambiguous control retained two genes at G2 without forced winner selection; one QTL signal was explicitly prior-sensitive. The null control retained a diffuse conditional PIP distribution, posterior H_0 0.97968, and G0 insufficient-evidence status. Twenty-eight scientific artifacts were equivalent between Python and containerized Nextflow execution; two clean Docker runs were reproducible, and all ten tasks were cached on resume. The final baseline passed 102 Python

tests and 11 of 11 nf-test contracts. CycloneDX and SPDX inventories, container identity, and Debian package bytes and contents were independently verified.

Conclusions. SchizoCircuit demonstrates that a locus-to-gene workflow can make thresholds, evidence dependence, uncertainty, provenance, and claim ceilings executable and testable. The study establishes internal correctness and reproducibility for a synthetic benchmark only. It does not identify schizophrenia genes, biomarkers, therapeutic targets, mechanisms, or clinically useful predictors.

Corresponding author: Regio Marcos Pinto Abreu Filho, regiomarcosabreu@gmail.com

1. Introduction

Large schizophrenia genome-wide association studies have identified hundreds of associated loci and have used fine mapping, functional genomic annotation, and transcriptomic resources to prioritize candidate genes ^[1]. The transition from an associated locus to a defensible effector-gene hypothesis nevertheless remains difficult because variants can be linked to several nearby genes, molecular traits can share linkage disequilibrium without sharing a causal variant, and multiple evidence channels can derive from the same underlying association signal. General locus-to-gene resources have therefore combined genetics and functional data while emphasizing calibrated prediction rather than direct biological proof ^[2].

The methodological problem is not only statistical. Computational pipelines often leave important decisions implicit: which evidence channels count, what constitutes independence, how quality-control failures propagate, whether ambiguity is preserved or collapsed, which priors affect colocalization, and which outputs are permitted to support a biological claim. When these rules are not executable, a negative result can reflect post-hoc thresholding and a positive result can reflect unrecognized evidence duplication.

The initial version of this manuscript proposed a conservative claim-governance concept but presented a narrative candidate attrition result that was not reconstructable from inspectable matrices, operational thresholds, or an executable release. Independent audit therefore recommended a complete reframing as a methods, software, and reproducibility paper. The current study implements that recommendation.

The former narrative candidate set and gene-specific interpretations are not treated as evidence in this revision.

SchizoCircuit was designed around four principles. First, statistical transformations and policy rules must be explicit and versioned. Second, evidence that shares upstream data must not be counted as independent merely because it appears in different files. Third, the system must preserve positive, ambiguous, and null control behavior rather than force a ranked winner at every locus. Fourth, every candidate output must inherit a claim ceiling that is no stronger than the verified evidence and dataset class.

The objective was to determine whether these principles could be implemented as a deterministic, modular workflow whose outputs remain stable across operating systems, execution engines, clean reruns, and cache-resume paths. The study does not estimate predictive accuracy, biological validity, or clinical utility. Instead, it asks whether a frozen synthetic benchmark can test internal logic, provenance, uncertainty propagation, and reproducibility with sufficient precision for independent audit.

2. Study design and scope

This was a computational methods and software-verification study. No human participants, primary biological samples, patient-level records, animal experiments, or newly collected clinical data were used. All inputs were explicitly labeled synthetic engineering fixtures, and the pipeline carried `external_validity=false` through its context, graphs, candidate cards, and final manifests.

The benchmark was deliberately small enough for complete inspection. It contained three loci, nine GWAS-like rows, nine QTL-like rows, three supplied linkage-disequilibrium matrices, and seven structured annotation rows. The positive locus was configured to support one gene through two independent channels. The ambiguous locus was configured to support two genes through distinct channel pairs, including one prior-sensitive colocalization. The null locus contained weak association statistics and context-only annotation. These controls tested success, ambiguity preservation, and abstention/insufficient-evidence behavior within the same code path.

The project name reflects the intended future domain of schizophrenia locus-to-gene analysis, but the synthetic symbols `GENE_POS`, `GENE_AMB_A`, `GENE_AMB_B`, and `GENE_NULL` are placeholders. No output in this paper should be interpreted as evidence about an actual gene or locus.

3. Methods

3.1. Frozen baseline and identity

The manuscript was frozen against workflow version 4.3.0 and policy version 4.2.0. Scientific source identity was represented by a SHA-256 digest over the files that define the analytical core. The same scientific-core hash was observed in Windows and WSL/Linux. A stable analysis identifier was derived from workflow version, policy identity, and canonicalized scientific-input hashes; invocation-specific run identifiers and timestamps were excluded from scientific equivalence comparisons.

Identifier	Frozen value
Workflow version	4.3.0
Policy version	4.2.0
Scientific-core SHA-256	ce0ffc01e54e261913816db4140cd8c0d412097d8a6c74692f77c111a4efc252
Stable analysis ID	f14c896233b7db15f97ec98962ee6fd37fbbd5e1e6cfb7aa8837742a277c9636
Container tag	schizocircuit-v4:stage5a-ce0ffc01e54e
Container image ID	sha256:7907562b00b5183ec8af26ca68718af18e51b0bf212067fbee89745c31f1dca3
Base image	python:3.13.13-slim-bookworm@sha256:355bfa66770995d7e9a0da4b3473b44d0cb451f6b56f5615ad9c39e3c4eca03f
Verified release bundle SHA-256	ef07df26e4b094334a99f2874094b08f5ecb856695cb84d66158727f7eb5548d

Table 1. Frozen analytical and release identifiers.

Note. The long identifiers are printed in full so the manuscript can be matched to the verified release independently of filenames.

3.2. Synthetic benchmark fixtures

The GWAS-like table used GRCh38 coordinates, EUR ancestry labels, sample size 100,000, and deterministic beta, standard-error, p-value, allele-frequency, and imputation-information fields. The QTL-like table used a DLPFC tissue label and deterministic eQTL or sQTL effects. These labels reproduce a familiar schema but do not represent observed schizophrenia or brain-molecular results. The provenance file states: “Association and QTL effects are deterministic engineering fixtures, not observed schizophrenia results.”

The positive control (POS1) included a dominant association at rsPOS1 and annotations linking that variant to GENE_POS through a coding-consequence channel and a separate eQTL-colocalization channel. The ambiguous control (AMB1) included two correlated associated variants and two candidate genes, each supported by a QTL-colocalization channel and a distinct enhancer-link channel. The null control (NULL1) contained small z scores, diffuse variant evidence, and a context-only annotation that was not permitted to count toward G1 or G2.

3.3. Workflow architecture

The workflow was implemented in Nextflow DSL2 as ten named processes with explicit file contracts and fail-fast error propagation. Scientific outputs were emitted as TSV, JSON, GraphML, and Markdown; execution observers included trace, report, timeline, and DAG files. The Python reference runner and Nextflow processes invoked the same package functions, which reduced the risk of implementing two divergent scientific algorithms.

No.	Process	Primary contract	Principal outputs
01	INGEST_GWAS	Validate schema; preserve rows; create analysis context	01_gwas_ingested.tsv; 00_analysis_context.json
02	QC_HARMONIZE	Numerical, allele, build, z-p, duplicate, and palindromic checks	pass/fail TSVs; QC report
03	VALIDATE_LD	Validate supplied matrices; do not compute LD	LD report
04	FINE_MAP	Wakefield ABF with explicit H0 and 95% credible sets	variant results; credible sets; report
05	INGEST_QTL	Validate QTL schema and harmonize alleles	harmonized/failure TSVs; report
06	COLOCALIZE	Single-causal ABF colocalization and p12 sensitivity	central results; sensitivity table; report
07	BUILD_EVIDENCE_GRAPH	Persist evidence, provenance, dependence, and contradictions	JSON and GraphML graphs
08	APPLY_POLICY	Evaluate versioned G0–G4 rules and locus status	policy decisions and report
09	GENERATE_CANDIDATE_CARDS	Render claim-limited JSON/Markdown records	four cards and index
10	RUN_METADATA	Finalize execution identity and manifest	run metadata; SHA-256 manifest

Table 2. Ten-module Nextflow DSL2 architecture.

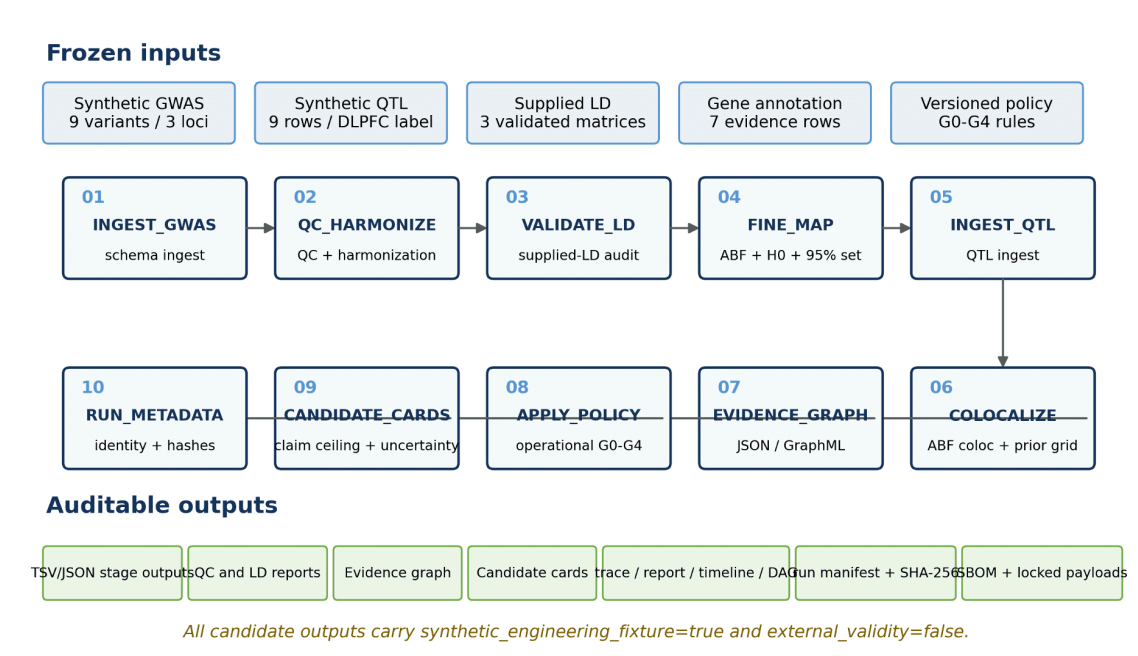


Figure 1. Workflow architecture. Frozen synthetic inputs enter ten modular processes. Scientific outputs, execution observers, evidence graphs, candidate cards, and manifests are emitted for audit. The Python reference runner and Nextflow processes use the same scientific package.

3.4. GWAS quality control and deterministic p-value reconstruction

The ingest module required an exact column contract and preserved accepted input rows without silent transformation. The QC module validated numeric ranges, positive standard errors, allele membership, effect/other allele consistency, GRCh38 build, imputation information, unique variant identifiers, and conflicting duplicate positions. For each valid row it recomputed $z = \beta / SE$ and the two-sided normal p value. To eliminate cross-platform differences observed at approximately 10^{-16} , the final implementation used a deterministic high-precision evaluation and serialized p values at a fixed significant-digit policy.

The ratio between the reported and recomputed p value was used as a diagnostic. Ratios greater than 2 triggered REVIEW_REQUIRED and ratios greater than 10 triggered FAIL. Palindromic A/T or C/G variants with effect-allele frequency between 0.40 and 0.60 also triggered review. A failed locus was represented as QC_FAILED, not silently converted into low evidence.

3.5. Linkage-disequilibrium validation

LD matrices were supplied as benchmark inputs and were never described as computed by the workflow. The validator checked squareness, finite values, range [1][1], unit diagonal, symmetry, positive semidefiniteness, and exact variant-label order. Minimum eigenvalues and condition numbers were recorded. A missing or invalid matrix blocked downstream interpretation through the NO_VALID_LD_REFERENCE state.

3.6. Approximate-Bayes-factor fine mapping

The fine-mapping component was intentionally a transparent baseline rather than a multi-causal production model. For variant j , $V_j = SE_j^2$, $z_j = \beta_j / SE_j$, and the Wakefield approximate Bayes factor was calculated as:

$$ABF_j = \sqrt{V_j / (V_j + W)} * \exp[z_j^2 * W / 2(V_j + W)]$$

The prior effect variance W was 0.04 (prior standard deviation 0.2) [3]. A locus-level prior probability of causality of 0.05 introduced an explicit H_0 state. Conditional PIPs were normalized among variants given a causal locus; unconditional PIPs multiplied those conditional values by the posterior probability that the locus was causal. Variants were sorted deterministically and accumulated until the 95% credible-set threshold was reached. The method label in every output was ABF_SINGLE_CAUSAL_WITH_H0_LD_UNAWARE_BASELINE. It was not presented as SuSiE or FINEMAP [4] [5].

3.7. QTL harmonization and colocalization

QTL rows were matched by variant identifier, build, ancestry label, and allele orientation. Reversed effect/other alleles caused a sign flip and were recorded; incompatible alleles caused an explicit failure/abstention. QTL type and dataset identifier remained part of the evidence identity.

Colocalization implemented a summary-statistic ABF baseline with one causal variant per trait [6]. Using $p_1 = p_2 = 10^{-4}$, central $p_{12} = 10^{-5}$, and p_{12} sensitivity grid $\{10^{-6}, 10^{-5}, 10^{-4}\}$, the unnormalized hypothesis weights were:

$$\begin{aligned}
H_0 &= 1 \\
H_1 &= p_1 * \text{sum}(\text{ABF_GWAS}) \\
H_2 &= p_2 * \text{sum}(\text{ABF_QTL}) \\
H_3 &= p_1 * p_2 * [\text{sum}(\text{ABF_GWAS}) * \text{sum}(\text{ABF_QTL}) - \text{sum}(\text{ABF_GWAS} * \text{ABF_QTL})] \\
H_4 &= p_{12} * \text{sum}(\text{ABF_GWAS} * \text{ABF_QTL})
\end{aligned}$$

The weights were normalized to PP0-PP4. Variant sets had to match exactly and contain at least two variants. A PP4 threshold of 0.50 was required for a QTL-colocalization channel to count. The entire p12 grid was retained because prior misspecification can materially alter posterior inference, and the single-causal assumption can be unrealistic [7].

3.8. Evidence graph, dependence model, and operational grades

Evidence was persisted as JSON and GraphML with nodes for loci, variants, genes, datasets, methods, policy, and run identity. Each evidence record included the evidence channel, track, source dataset, cohort identifier, publication placeholder, support direction, contradiction status, independence group, and provenance hash. Fine mapping and colocalization were not counted as independent evidence channels because both derived from the same synthetic GWAS association statistics. Independence for G2 was counted only from distinct annotation-defined groups.

Grade	Name	Operational rule	Permitted interpretation
G0	CONTEXT	No counted support channel or locus posterior below policy floor	Context only; not a candidate
G1	LOCUS SUPPORTED	At least one eligible counted channel	Hypothesis-generating locus support
G2	CONVERGENT EFFECTOR CANDIDATE	At least two eligible channels in at least two independence groups; QC and LD PASS; complete provenance; no major unresolved contradiction	Synthetic convergent-candidate status only
G3	EXPERIMENTALLY SUPPORTED MEDIATOR	G2 plus functional perturbation, directional coherence, and orthogonal confirmation	Unreachable in this synthetic release
G4	HUMAN TRANSLATIONAL HYPOTHESIS	G3 plus therapeutic-tractability evidence, independent human replication, and safety assessment	Unreachable in this synthetic release

Table 3. Versioned claim-governance grades.

Note. G2 requires both channel and independence-group multiplicity. Duplicate evidence and evidence from the same independence group cannot promote a gene. Synthetic fixtures are prohibited from producing G3 or G4.

Locus status was evaluated separately from gene grade. A locus could be QC_FAILED, NO_VALID_LD_REFERENCE, ABSTAINED, CONFLICTING_EVIDENCE, MULTIPLE_PLAUSIBLE_GENES, SINGLE_LEADING_GENE, or INSUFFICIENT_EVIDENCE. The policy contained no code path that selected a winner when more than one gene reached the leading-grade threshold.

3.9. Candidate cards, claim firewall, and manifest

For each locus-gene pair, the workflow generated JSON and Markdown candidate cards containing grade, locus status, counted channels, independence groups, uncertainty, blockers, source hashes, analysis identity, policy version, and a reproducibility hash. Every card included the mandatory sentence “This is an engineering control and does not constitute biological validation.” A prohibited-claim scan blocked

phrases that would imply a proven causal gene, validated biomarker, therapeutic target, treatment efficacy, clinical utility, or individual forensic prediction.

The final manifest recorded SHA-256 hashes for all scientific outputs and execution metadata, excluding the manifest itself to avoid circularity. Verification recomputed the hashes and failed on drift. The stable analysis identifier excluded run-specific timestamps and UUIDs, whereas invocation identifiers remained available for traceability.

3.10. Reproducibility and supply-chain verification

The verification program was staged. Cross-platform hardening normalized text encoding, line endings, serialization order, floating-point output policy, and Python hash-seed sensitivity. Python runs were repeated in Windows and Linux. The modular Nextflow pipeline was executed in clean runs and with resume. Containerized runs used a non-root image built from a base pinned by digest, with the package imported from site-packages rather than a mounted source tree.

Scientific equivalence compared 28 designated artifacts byte-for-byte for non-JSON files and semantically after removal of invocation-only fields for JSON. Trace verification required all ten processes, expected completion or cache status, and the expected container. nf-test 0.9.5 executed ten process contracts and one end-to-end contract under Nextflow 25.10.2 ^{[8][9]}. The final Python regression suite contained 102 tests spanning unit, integration, negative, determinism, cross-platform, container, provenance, supply-chain, and evidence-contract checks.

Supply-chain verification included the base-image digest, final image identity, installed Python distributions, CycloneDX and SPDX software bills of materials, and independent Syft 1.51.0 reconciliation. Hash-locked Debian payloads for procp and libproc2-0 were checked by byte size, SHA-256, package metadata, architecture, archive contents, and live image installation. The final release archive was hashed and matched to the identifiers in Table 1. These practices follow established recommendations for reproducible computational research and FAIR research objects ^{[10][11][12][13][14]}.

4. Results

4.1. Quality-control and LD behavior

All nine GWAS-like rows passed the configured numerical and allele checks; no palindromic variant had effect-allele frequency in the 0.40–0.60 review interval. The three supplied LD matrices passed label-

order, symmetry, range, diagonal, and positive-semidefinite checks. Minimum eigenvalues were 0.1983 for POS1, 0.0786 for AMB1, and 0.8882 for NULL1; condition numbers were 9.64, 25.22, and 1.25, respectively. These values are diagnostics of the supplied engineering matrices, not estimates from population genotypes.

4.2. Fine-mapping control behavior

POS1 produced posterior locus causality 0.9999897 and posterior H_0 1.03×10^{-5} . rsPOS1 had conditional PIP 0.9999967 and alone formed the 95% credible set. AMB1 produced posterior locus causality 0.9719614; rsAMB1 and rsAMB2 had conditional PIPs 0.9371250 and 0.0626935 and jointly formed the credible set. NULL1 produced posterior locus causality 0.0203180 and posterior H_0 0.9796820. Its conditional PIPs were diffuse (0.3338, 0.3499, and 0.3163), and all variants were required for the nominal credible set. Including H_0 therefore prevented the null locus from being described as confidently causal merely because conditional PIPs must sum to one.

4.3. Colocalization and prior sensitivity

The POS1-GENE_POS eQTL pair had central PP4 0.999984 and remained between 0.999836 and 0.999998 across the p12 grid. AMB1-GENE_AMB_A had central PP4 0.987554 and range 0.888076–0.998741. AMB1-GENE_AMB_B had central PP4 0.949744 but a broader range of 0.653956–0.994736. The workflow preserved this range in the candidate card as an uncertainty field rather than reporting only the central posterior.

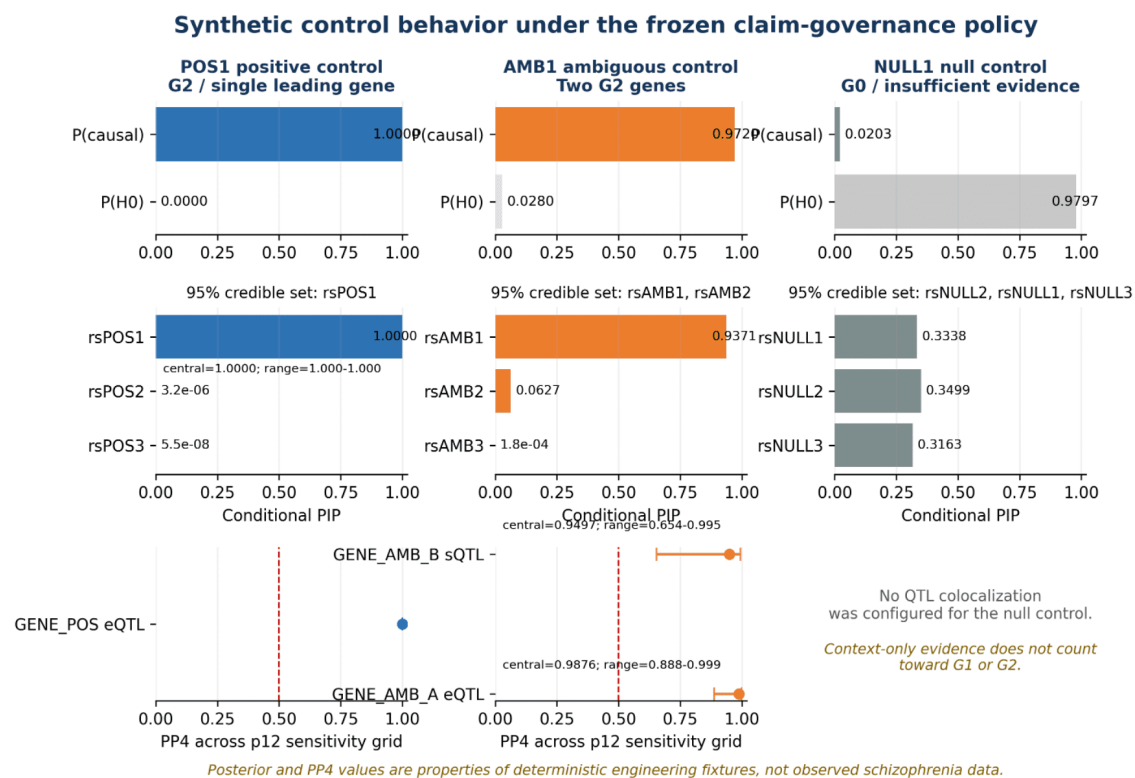


Figure 2. Prespecified synthetic control behavior. Top panels show posterior locus causality and H0; middle panels show conditional variant PIPs and 95% credible sets; bottom panels show PP4 central values and the p12 sensitivity range. These are deterministic engineering outputs, not disease estimates.

4.4. Policy outcomes

The policy reproduced all three prespecified behaviors. GENE_POS reached G2 through a strong coding-consequence channel and an eQTL-colocalization channel in two independence groups; POS1 was SINGLE_LEADING_GENE. GENE_AMB_A and GENE_AMB_B each reached G2 through separate QTL and enhancer-link groups, so AMB1 was MULTIPLE_PLAUSIBLE_GENES. No tie-breaker or rank score collapsed the locus to one gene. GENE_NULL retained only context evidence, was blocked by low locus posterior causality, and remained G0 with INSUFFICIENT_EVIDENCE.

Locus	Gene	P(causal)	95% credible set	PP4 central (grid range)	Grade	Locus status
POS1	GENE_POS	0.99999	rsPOS1	0.99998 (0.99984–1.00000)	G2	SINGLE_LEADING_GENE
AMB1	GENE_AMB_A	0.97196	rsAMB1, rsAMB2	0.98755 (0.88808–0.99874)	G2	MULTIPLE_PLAUSIBLE_GENES
AMB1	GENE_AMB_B	0.97196	rsAMB1, rsAMB2	0.94974 (0.65396–0.99474)	G2	MULTIPLE_PLAUSIBLE_GENES
NULL1	GENE_NULL	0.02032	all 3 variants	not configured	G0	INSUFFICIENT_EVIDENCE

Table 4. Synthetic benchmark outcomes under the frozen policy.

Note. The locus-level posterior is repeated for both AMB1 genes. PP4 was not configured for the null context-only control.

4.5. Cross-platform, engine, and resume equivalence

The scientific-core hash was identical in Windows and WSL/Linux. Repeated Python executions produced the same analysis identifier and equivalent scientific outputs after exclusion of invocation-only metadata. The ten-process Nextflow graph completed in two independent clean Docker runs. The trace recorded the expected image for all processes. Python reference output and Docker/Nextflow output were equivalent for 28 designated scientific artifacts with no recorded differences. A resume run recovered all ten processes from cache while preserving the analysis identifier and scientific outputs.

4.6. Testing and supply-chain closure

The final source tree passed 102 Python tests in both Windows and Linux. nf-test executed 11 contracts: one for each of the ten processes and one full-workflow contract; all passed. The non-root container used Python 3.13.13, imported SchizoCircuit from site-packages, and contained the expected scientific-core hash. Syft-generated CycloneDX and SPDX inventories contained the expected Python and Debian packages and reconciled without problems. The procps and libproc2-0 packages matched the locked

versions, architecture, sizes, hashes, and expected archive contents. The Stage 5B finalizer reported status VERIFIED with an empty problems list.

Verification layer	Comparison/contract	Observed result	Status
Windows/Linux source identity	scientific-core SHA-256	identical	VERIFIED
Python reproducibility	independent clean runs	stable analysis ID	VERIFIED
Nextflow modular execution	10 processes	10/10 completed	VERIFIED
Python vs Docker/Nextflow	28 scientific artifacts	no differences	VERIFIED
Docker clean run A vs B	scientific outputs	reproducible	VERIFIED
Resume/cache	same work directory	10/10 cached	VERIFIED
Python regression	Windows and Linux	102/102 passed	VERIFIED
nf-test contracts	10 processes + workflow	11/11 passed	VERIFIED
SBOM cross-check	CycloneDX + SPDX via Syft	expected inventory present	VERIFIED
Debian payloads	hash, size, metadata, contents	procps and libproc2-0 matched	VERIFIED

Table 5. Verification matrix for the frozen Stage 5A-5B release.

Verification stack for the frozen Stage 5A-5B baseline

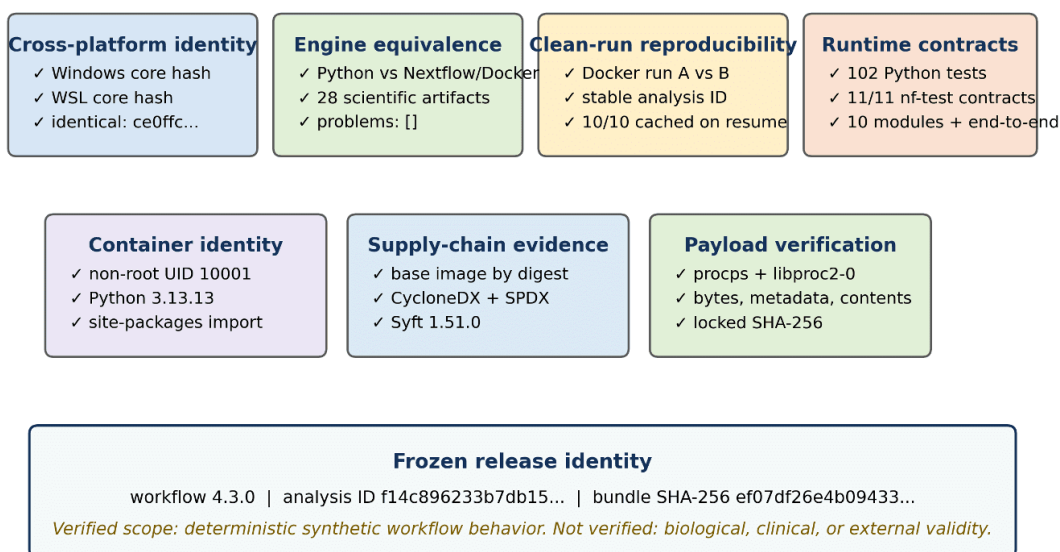


Figure 3. Verification stack. The release was checked across operating systems, execution engines, clean reruns, cache-resume behavior, unit and workflow contracts, container identity, software bills of materials, and package payloads. The verified scope remains internal synthetic workflow behavior.

5. Discussion

The principal contribution of SchizoCircuit is not a new schizophrenia candidate. It is an executable separation between evidence accumulation and claim authorization. Statistical outputs, annotations, policy rules, and language constraints are represented as inspectable artifacts rather than prose-only judgments. The synthetic controls show why this separation matters: a positive control can be promoted when independent criteria are met, an ambiguous locus can retain multiple candidates, and a null locus can remain non-causal despite normalized within-locus PIPs.

The explicit H0 state is particularly important for auditability. A conditional PIP distribution always sums to one among variants; without a locus-level null state, a weak locus can appear to contain a leading variant simply because one variant must be largest. In the null fixture, conditional PIPs were approximately one third each, but posterior H0 was 0.97968. The policy correctly used the locus posterior to block promotion.

The ambiguous control illustrates a second failure mode: conflating gene prioritization with winner selection. Both GENE_AMB_A and GENE_AMB_B satisfied the same operational G2 requirements through distinct evidence groups. The appropriate output was therefore MULTIPLE_PLAUSIBLE_GENES. This differs from predictive resources that are designed to rank or score genes across large empirical loci ^[2]; SchizoCircuit in its current form is a governance and audit framework, not a trained predictor or comparative benchmark.

The colocalization sensitivity grid also demonstrates the value of carrying uncertainty into downstream records. The central PP4 for GENE_AMB_B exceeded the policy threshold, but the lower-grid value was 0.654 rather than near one. A single central number would conceal this dependence on p12. The card therefore records prior sensitivity while preserving the prespecified G2 result. In empirical analyses, multi-signal methods and aligned LD should be considered when the single-causal assumption is doubtful ^{[7][4][5]}.

The reproducibility program addressed defects that cannot be resolved by a Dockerfile alone. Cross-platform floating-point differences, line-ending conversion, source-tree import shadowing, stale WSL integration mounts, missing runtime utilities, cache identity, and supply-chain provenance were all treated as testable failure modes. The final evidence chain links scientific input hashes, code identity, analysis identity, container identity, process traces, candidate-card hashes, SBOMs, and a release-bundle checksum. This aligns with reproducible-computation and FAIR research-object principles ^{[10][11][12][13][14]}.

A further correction is the removal of the unstructured literature resolver from the verified grading core. Biomedical literature supports the rationale and future extension of the system, but no title-level search, abstract parser, or citation count contributes to G0–G4 in this release. This avoids converting incomplete literature retrieval into a molecular evidence route and directly resolves a major weakness of the earlier manuscript.

The verified release is therefore suitable as a methods and software artifact for further scrutiny. The next scientific phase would require a separate empirical protocol: preregistered observed schizophrenia summary statistics, ancestry-matched LD, frozen QTL resources, multi-causal fine mapping, calibration controls drawn from real loci, blinded assessment of policy outputs, and independent replication. Such a phase would create a new baseline and must not inherit biological validity from the synthetic benchmark.

6. Limitations

- All association, QTL, annotation, cohort, publication, and gene identifiers are engineering fixtures. The benchmark has no external, biological, or clinical validity.
- The fine-mapping method is a single-causal, LD-unaware Wakefield-ABF baseline. The LD matrices are validated and reported but are not used in the posterior model.
- The colocalization model assumes one causal variant per trait and is evaluated only on matched synthetic variant sets. Prior sensitivity is assessed, but empirical calibration is absent.
- The benchmark contains three loci and four gene cards. It tests specified behaviors but cannot estimate predictive accuracy, calibration, recall, or generalization.
- Ancestry is labeled EUR in the fixture, but the study does not evaluate ancestry transfer, LD-reference mismatch, or population heterogeneity.
- The completed verification does not include a formal mutation-testing campaign or an independent third-party clean-room reproduction. The paper therefore does not describe the software as mutation-tested, production-grade, or high-assurance.
- The release archive is hash-identified but has not been assigned a persistent repository DOI within this manuscript. Long-term deposition should preserve the verified bytes and checksum.

7. Conclusion

SchizoCircuit converts claim governance from a narrative caution into an executable workflow. The frozen synthetic benchmark demonstrates deterministic positive, ambiguous, and null behavior; operational G0-G4 grades; explicit evidence dependence; propagated uncertainty; reproducible Python, Nextflow, and Docker execution; tested resume behavior; and supply-chain traceability. These results support a narrow methods claim: the workflow behaves internally as specified under predefined fixtures. They do not support a claim about schizophrenia genes, mechanisms, biomarkers, targets, or patient-level prediction.

Abbreviations

ABF, approximate Bayes factor; CI, continuous integration; DLPFC, dorsolateral prefrontal cortex; GWAS, genome-wide association study; LD, linkage disequilibrium; OCI, Open Container Initiative; PIP, posterior

inclusion probability; PP4, posterior probability of a shared causal variant under the colocalization model; QTL, quantitative trait locus; SBOM, software bill of materials; WSL, Windows Subsystem for Linux.

Notes

Scope

The disease framing motivates the future application domain. Every association, QTL effect, annotation identifier, gene symbol, and control outcome analyzed here is a deterministic engineering fixture. The present paper evaluates workflow behavior, not schizophrenia biology.

Statements and Declarations

Funding

No external funding was received for this work.

Potential Competing Interests

The author declares no competing interests.

Ethics

Not applicable. This study used deterministic synthetic engineering fixtures and no human participants, patient-level data, animal experiments, identifiable records, or new biological samples.

Data Availability

The executable source, synthetic fixtures, verification reports, nf-test contracts, software bills of materials, hash-locked payload records, and audit materials are supplied as a verified supplementary release bundle. The bundle SHA-256 is ef07df26e4b094334a99f2874094b08f5ecb856695cb84d66158727f7eb5548d. The bundle must be verified before use; a filename alone is not an identity guarantee.

The frozen code identity is workflow version 4.3.0 with scientific-core SHA-256 ce0ffc01e54e261913816db4140cd8c0d412097d8a6c74692f77c111a4efc252. The container image used for

Stage 5A verification is schizocircuit-v4:stage5a-ce0ffc01e54e with image ID sha256:7907562b00b5183ec8af26ca68718af18e51b0bf212067fbee89745c31fdca3.

Author Contributions

Regio Marcos Pinto Abreu Filho conceived the study, defined the claim-governance framework, supervised pipeline development, executed the workstation verification program, interpreted the results, and wrote and approved the final manuscript.

Use of Generative AI

Generative AI tools were used during iterative drafting, code generation, troubleshooting, audit design, and editorial revision. All computational results reported here were executed in controlled local environments. The author reviewed the source, tests, logs, outputs, references, and final text and accepts full responsibility for the integrity and interpretation of the work.

References

1. ^{a, b, c}Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. (2022). "Mapping Genomic Loci Implicates Genes and Synaptic Biology in Schizophrenia." *Nature*. **604**:502–508. doi:[10.1038/s41586-022-04434-5](https://doi.org/10.1038/s41586-022-04434-5).
2. ^{a, b}Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. (2021). "An Open Approach to Systematically Prioritize Causal Variants and Genes at All Published Human GWAS Trait-Associated Loci." *Nat Genet*. **53**:1527–1533. doi:[10.1038/s41588-021-00945-5](https://doi.org/10.1038/s41588-021-00945-5).
3. ^ΔWakefield J (2009). "Bayes Factors for Genome-Wide Association Studies: Comparison with P-Values." *Genet Epidemiol*. **33**:79–86. doi:[10.1002/gepi.20359](https://doi.org/10.1002/gepi.20359).
4. ^{a, b}Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016). "FINEMAP: Efficient Variable Selection Using Summary Data from Genome-Wide Association Studies." *Bioinformatics*. **32**:1493–1501. doi:[10.1093/bioinformatics/btw018](https://doi.org/10.1093/bioinformatics/btw018).
5. ^{a, b}Wang G, Sarkar A, Carbonetto P, Stephens M (2020). "A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping." *J R Stat Soc Series B Stat Methodol*. **82**:1273–1300. doi:[10.1111/rssb.12388](https://doi.org/10.1111/rssb.12388).
6. ^ΔGiambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. (2014). "Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics." *PLoS Genet*. **10**:

e1004383. doi:[10.1371/journal.pgen.1004383](https://doi.org/10.1371/journal.pgen.1004383).

7. ^aWallace C (2020). "Eliciting Priors and Relaxing the Single Causal Variant Assumption in Colocalisation Analyses." *PLoS Genet.* **16**:e1008720. doi:[10.1371/journal.pgen.1008720](https://doi.org/10.1371/journal.pgen.1008720).
8. ^ΔDi Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C (2017). "Nextflow Enables Reproducible Computational Workflows." *Nat Biotechnol.* **35**:316–319. doi:[10.1038/nbt.3820](https://doi.org/10.1038/nbt.3820).
9. ^ΔForer L, Schönherr S (2025). "Improving the Reliability, Quality, and Maintainability of Bioinformatics Pipelines with nf-test." *GigaScience.* **14**:giaf130. doi:[10.1093/gigascience/giaf130](https://doi.org/10.1093/gigascience/giaf130).
10. ^aSandve GK, Nekrutenko A, Taylor J, Hovig E (2013). "Ten Simple Rules for Reproducible Computational Research." *PLoS Comput Biol.* **9**:e1003285. doi:[10.1371/journal.pcbi.1003285](https://doi.org/10.1371/journal.pcbi.1003285).
11. ^aWilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. (2016). "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Sci Data.* **3**:160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
12. ^aSoiland-Reyes S, Sefton P, Crosas M, Castro LJ, Coppens F, Fernández JM, et al. (2022). "Packaging Research Artefacts with RO-Crate." *Data Sci.* **5**:97–138. doi:[10.3233/DS-210053](https://doi.org/10.3233/DS-210053).
13. ^aLebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, Corsar D, et al. (2013). "PROV-O: The PROV Ontology." *W3C Recommendation*.
14. ^aMunafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, et al. (2017). "A Manifesto for Reproducible Science." *Nat Hum Behav.* **1**:0021. doi:[10.1038/s41562-016-0021](https://doi.org/10.1038/s41562-016-0021).

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.