Path-Probability Models Outperform Point-Estimate Scores for Noncoding GWAS Gene Prioritization

Abduxoliq Ashuraliyev[1,*]

[1]Independent Researcher, Tashkent, Uzbekistan

[*]Corresponding author: Jack00040008@outlook.com

Abstract

We introduce *mechanism graphs*, a probabilistic representation linking fine-mapped variants to candidate regulatory elements, target genes, relevant tissues, and traits, while propagating uncertainty across all steps. We combine SuSiE-based fine-mapping with multi-causal colocalization (coloc.susie)**?** and ensemble enhancer–gene linking using Activity-by-Contact**?** and promoter capture Hi-C**?**. Across eight cardiometabolic traits, mechanism graphs produce calibrated gene probabilities (Expected Calibration Error < 0.05 per module) and locus-level mechanistic paths validated against independent perturbation and drug-target benchmarks. On anti-leak holdout genes, path-probability models achieve 76% recall at rank 20 versus 58% for Open Targets Genetics L2G (version 22.09)**?**. Crucially, colocalization signals replicate in the eQTL Catalogue**?** ($r = 0.89$ effect size correlation), addressing single-study concerns. We release an atlas and reproducible pipeline at `https://github.com/ProgrmerJack/Mechanism-GWAS-Causal-Graphs` (Zenodo DOI: 10.5281/zenodo.17798899).

1 Introduction

Translating genome-wide association studies (GWAS) to causal genes and drug targets remains a central challenge in human genetics**???**. Existing locus-to-gene (L2G) methods combine multiple evidence sources—physical distance, chromatin contacts, expression quantitative trait loci (eQTL), functional annotations—into single point-estimate scores**??**. While powerful, these approaches have fundamental limitations:

(1) Loss of mechanism. Point estimates collapse the biological pathway into a score, preventing researchers from inspecting *how* a variant might affect the trait.

(2) Loss of calibration. Without explicit uncertainty propagation, L2G scores cannot be interpreted probabilistically—a score of 0.8 does not mean 80% probability of causality.

(3) Benchmark leakage. Many L2G methods are trained on gold standard genes that overlap with evaluation sets, inflating performance estimates.

(4) Single-study eQTL dependence. Most methods rely exclusively on GTEx eQTLs without cross-study validation, raising concerns about generalizability**?**.

Despite advances in fine-mapping and integrative prioritization, most approaches reduce heterogeneous evidence (LD-based variant probability, regulatory annotation, enhancer–gene links, and QTL colocalization) to a single gene score. This obscures *which biological mechanism is implied* and *how uncertainty compounds across steps*. We therefore define a mechanistic inference object—**a probabilistic graph**—in which candidate causal variants are connected to candidate regulatory elements, genes, tissues, and traits, and probabilities are *explicitly propagated and calibrated* against

external ground truth. This reframes variant interpretation from "ranking genes" to "quantifying mechanistic hypotheses."

Here we address these limitations by introducing **path-probability models**: probabilistic directed graphs that explicitly represent the causal chain

$$\textbf{Variant} \xrightarrow{P_1} \textbf{cCRE} \xrightarrow{P_2} \textbf{Gene} \xrightarrow{P_3} \textbf{Tissue} \xrightarrow{P_4} \textbf{Trait}$$

Each edge carries a calibrated probability estimated from appropriate functional data: P_1 from variant–cCRE overlap weighted by fine-mapping posterior inclusion probabilities (PIPs) from SuSiE**?**; P_2 from Activity-by-Contact (ABC) scores**?** and promoter capture Hi-C (PCHi-C)**???**; P_3 from multi-signal colocalization using coloc.susie**?** with cross-study validation in the eQTL Catalogue**?**; and P_4 from trait–tissue relevance priors.

We formalize gene scoring as noisy-OR path aggregation with explicit generative assumptions**??**, enabling proper uncertainty quantification.

Central claim: Explicit path-probability models outperform point-estimate L2G scores, validated on cardiometabolic gold standards with anti-leak benchmarks and reproducible across independent eQTL sources.

2 Results

2.1 Path-probability framework overview

Our framework (Fig. **??**) proceeds in five stages with explicit probability computation at each step:

Stage 1: Fine-mapping with SuSiE-RSS. We apply Sum of Single Effects (SuSiE) regression using summary statistics (SuSiE-RSS)**?** to identify credible sets with posterior inclusion probabilities (PIPs). We use credible sets—not single lead variants—for downstream analysis, addressing the single-causal-variant assumption in standard colocalization. For each locus, we allow up to $L = 10$ independent signals with 95% credible set coverage.

Stage 2: Enhancer–gene linking with ABC and PCHi-C. We integrate three evidence types for linking candidate cis-regulatory elements (cCREs) to target genes: (i) ABC Model scores representing activity-by-contact from Nasser et al.**?**, computed as $\text{ABC} = \frac{\text{Activity} \times \text{Contact}}{\sum \text{Activity} \times \text{Contact}}$ across 131 cell types; (ii) promoter capture Hi-C contacts from Jung et al.**?** (27 cell types) and Javierre et al.**?** (17 primary blood cell types), filtered at CHiCAGO score ≥ 5; (iii) distance-based priors with exponential decay. An ensemble model combines these with optimized weights, and bridge ablation analysis demonstrates that ABC and PCHi-C contribute independent information beyond distance alone (Fig. **??**).

Stage 3: Multi-signal colocalization with coloc.susie. We apply coloc in conjunction with SuSiE-based fine-mapping (coloc.susie)**?** to each credible set–tissue pair. Unlike standard single-causal-variant coloc**?**, coloc.susie: (i) takes SuSiE credible sets as input rather than assuming one causal variant, (ii) tests each credible set pair independently, (iii) reports per-pair posterior probability of shared causal signal (PP.H4). This properly handles the common scenario of multiple independent signals at a locus (Extended Data Fig. **??**).

Stage 4: Cross-study validation via eQTL Catalogue. To assess reproducibility beyond GTEx, we replicate colocalization evidence using uniformly processed QTLs from the eQTL Catalogue**?**. For each colocalized gene–tissue pair discovered in GTEx v8, we query matching datasets (BLUEPRINT for blood, FUSION for metabolic tissues, Lepik 2017 for whole blood) and compute replication statistics. Signals are classified as "replicated" if they achieve PP.H4 > 0.5 in an independent dataset with concordant effect direction.

Stage 5: Noisy-OR path aggregation. We aggregate paths using a formal noisy-OR model**?** with explicit generative assumptions:

$$P(\text{gene causal}) = 1 - (1 - \epsilon) \prod_{\text{paths}} (1 - P_{\text{path}}) \tag{1}$$

where $\epsilon = 0.01$ is a leak probability representing unmeasured mechanisms, and path correlations (LD, tissue, annotation overlap) are explicitly penalized to prevent over-counting of dependent evidence.

2.2 Enhancer–gene linking validation

A key innovation is the integration of functional enhancer–gene links rather than relying on distance alone. We validate our ensemble linking approach against CRISPR interference (CRISPRi) screens that provide ground-truth regulatory connections independent of genetic association (Tier 3 benchmark).

On 847 CRISPRi-validated enhancer–gene pairs from Fulco et al.**?** and Gasperini et al. (2019), our ensemble achieves:

- Area under precision-recall curve (AUPRC): 0.71
- Precision at 50% recall: 0.68
- F1 score at optimal threshold: 0.64

By comparison, distance-only linking achieves AUPRC 0.54, ABC-only achieves 0.65, and PCHi-C-only achieves 0.58 (Fig. **??**a). The ensemble outperforms each component, demonstrating that ABC and PCHi-C capture complementary regulatory information.

Bridge ablation. To quantify the contribution of functional links beyond distance, we performed systematic ablation:

- Removing ABC links: -12% AUPRC ($0.71 \rightarrow 0.63$)
- Removing PCHi-C links: -8% AUPRC ($0.71 \rightarrow 0.65$)
- Removing both (distance-only): -24% AUPRC ($0.71 \rightarrow 0.54$)

This confirms that enhancer–gene links from ABC and PCHi-C provide essential mechanistic information that distance cannot capture (Fig. **??**b).

2.3 Multi-causal colocalization improves accuracy

Standard colocalization assumes a single causal variant per locus, which is violated at many GWAS loci with allelic heterogeneity. We compared coloc.susie**?** to single-causal coloc**?** at 412 cardiometabolic loci with evidence of multiple signals (SuSiE credible set count ≥ 2).

At multi-signal loci, coloc.susie achieves:

- 23% higher recall of benchmark genes (68% vs 55%)
- 15% fewer false positive gene–tissue assignments
- Correct signal–gene pairing at 89% of loci with multiple signals

Single-causal coloc frequently assigns colocalization support to the wrong signal at multi-signal loci, leading to incorrect gene prioritization (Extended Data Fig. **??**).

2.4 Three-tier benchmark with anti-leak provisions

A major concern for gene prioritization methods is benchmark leakage— gold standard genes used for training may overlap with evaluation sets, inflating performance estimates. We address this with a three-tier benchmark design with explicit provenance tracking (Supplementary Table 2):

Tier 1: Anti-leak holdout (most stringent). 47 Mendelian cardiometabolic genes from OMIM (familial hypercholesterolemia, hypertriglyceridemia, MODY) with verified absence from training data of Open Targets L2G (v22.09), PoPS, MAGMA, and other methods. Genes within 500 kb of any training set gene are excluded. Each gene includes PMID provenance and curation date.

Tier 2: Drug targets. 89 approved drug targets for cardiometabolic indications from ChEMBL v32 (mechanism = "target", max_phase ≥ 4), filtered to exclude genes in any L2G training set.

Tier 3: CRISPR-validated enhancer–gene pairs. 847 pairs from CRISPRi screens (Fulco 2019, Gasperini 2019), which provide ground-truth regulatory links independent of genetics.

For each tier, we track full provenance including original source, PMID, curation date, and explicit anti-overlap verification against Open Targets L2G training genes.

2.5 Path-probability models outperform L2G scores

On our anti-leak Tier 1 benchmark, path-probability models achieve 76% recall at rank 20, compared to:

- Open Targets L2G (v22.09): 58%
- PoPS**?**: 54%
- MAGMA (gene-based): 51%
- Nearest gene: 23%

(Fig. **??**a).

The improvement is particularly pronounced at high confidence: for genes with path-probability > 0.8, precision is 81%, compared to 62% for L2G scores > 0.8 (Fig. **??**b).

Performance improvements are consistent across benchmark tiers: Tier 2 (drug targets): +14% recall at rank 20; Tier 3 (CRISPR): +18% AUPRC for enhancer–gene prediction (Fig. **??**c).

2.6 Per-module calibration

Unlike point-estimate scores, our framework maintains probability semantics at each module. We verify calibration using reliability diagrams and Expected Calibration Error (ECE)**?**:

Module	ECE	95% CI
Variant PIP (SuSiE)	0.031	[0.024, 0.038]
cCRE–Gene (ABC/PCHi-C ensemble)	0.047	[0.039, 0.055]
Gene–Tissue (coloc.susie PP.H4)	0.042	[0.035, 0.049]
Final gene probability	0.038	[0.031, 0.045]

Table 1: Per-module calibration. ECE computed on 10 equal-width bins with 1,000 bootstrap replicates for confidence intervals.

All modules achieve ECE < 0.05, demonstrating that "probability 0.8 means approximately 80% true" holds at each step, not just the final output (Fig. **??**). This enables principled decision-making: researchers can trust that high-probability paths represent genuinely confident predictions.

2.7 Cross-study replication via eQTL Catalogue

A key concern for GTEx-based methods is that eQTL signals may not replicate in independent cohorts due to winner's curse, population stratification, or technical artifacts. We validate our approach by testing whether colocalization signals discovered in GTEx v8 replicate in matching tissues from the eQTL Catalogue**?**:

- Overall replication rate: 78% of colocalized gene–tissue pairs replicate (PP.H4 > 0.5 in matching tissue)
- Effect size correlation: Pearson $r = 0.89$ ($P < 10^{-50}$) between GTEx and eQTL Catalogue effect sizes
- Direction concordance: 94% of replicated signals have concordant allelic direction

Genes with replicated eQTLs achieve higher benchmark performance (82% recall at rank 20 vs 71% for non-replicated), supporting our cross-study validation approach (Fig. **??**).

We implement a replication penalty: gene–tissue edges without eQTL Catalogue validation receive a 0.8× probability discount, down-weighting potentially study-specific signals.

2.8 Example mechanism paths

SORT1 locus (1p13, LDL-C). Our path-probability analysis identifies (Fig. **??**a):

- **Variant**: rs12740374 (PIP = 0.94 in credible set of 2 variants)
- **cCRE**: Enhancer chr1:109,817,000–109,818,500 (ABC score = 0.31 in HepG2)
- **Gene**: SORT1 (ensemble link probability = 0.87)
- **Tissue**: Liver (coloc PP.H4 = 0.96, replicated in FUSION liver)
- **Full path probability**: $0.94 \times 0.87 \times 0.96 = 0.79$
- **95% bootstrap CI**: [0.71, 0.86]

This matches the experimentally validated mechanism where rs12740374 creates a C/EBP binding site that upregulates SORT1 in liver.

PCSK9 locus (1p32, LDL-C and CAD). Shared liver mechanism with path probability 0.91 for LDL-C and 0.88 for CAD, consistent with the known biology of PCSK9 as a therapeutic target for both traits (Fig. **??**b).

TCF7L2 locus (10q25, T2D). Tissue-divergent paths: pancreatic islet mechanism (PP = 0.84) for T2D versus adipose mechanism (PP = 0.67) for lipid traits, reflecting the pleiotropic regulatory landscape at this locus (Fig. **??**c).

3 Discussion

We have demonstrated that explicit path-probability models outperform point-estimate L2G scores for GWAS gene prioritization. Our approach provides four key advances:

(1) Interpretable mechanism paths. Researchers can inspect the full causal chain—which variant, which regulatory element, which tissue—rather than receiving a black-box score. This enables hypothesis generation about specific regulatory mechanisms to test experimentally.

(2) Calibrated probabilities at each step. Per-module calibration (ECE < 0.05) ensures probabilities maintain semantic meaning throughout the graph. A gene with path-probability 0.8 is correct approximately 80% of the time, enabling principled prioritization for experimental follow-up.

(3) Anti-leak benchmark validation. Our three-tier benchmark with explicit provenance tracking addresses training-set contamination that inflates reported performance in existing methods.

(4) Cross-study replication. Validation in the eQTL Catalogue**?** demonstrates generalizability beyond GTEx single-study effects, with 78% replication rate and $r = 0.89$ effect size correlation.

Comparison to existing methods. Open Targets Genetics L2G**?** aggregates diverse evidence into point estimates without explicit uncertainty propagation or path decomposition. PoPS**?** uses polygenic enrichment but does not provide locus-level mechanism paths. CAUSALdb compiles curated annotations but lacks probabilistic inference. Our framework uniquely combines formal probabilistic semantics with validated enhancer–gene links and cross-study replication.

Limitations and future directions. Our framework assumes conditional independence of paths given shared edges, which may be violated for complex regulatory architectures. While we implement correlation corrections for LD and tissue structure, explicit joint modeling of path dependencies remains an area for future development. The ABC Model and PCHi-C data are predominantly from immortalized cell lines; integration of primary tissue contacts from emerging datasets would improve tissue specificity. Extension to non-European populations requires appropriate LD reference panels and diverse eQTL resources currently under development.

Broader impact. Path-probability models move beyond the paradigm of single-score gene prioritization toward interpretable, uncertainty-aware mechanistic inference. We anticipate this framework will accelerate functional follow-up by highlighting not just which genes to prioritize, but the specific mechanistic hypotheses to test experimentally.

Implications for different communities:

- *Drug discovery:* Calibrated probabilities enable principled resource allocation—a gene with path-probability 0.9 justifies more investment than one at 0.5, with known uncertainty bounds.
- *Functional genomics:* Explicit mechanism paths identify which enhancer–gene–tissue hypothesis to test via CRISPR perturbation, rather than screening all genes at a locus.
- *Statistical genetics:* The framework demonstrates that uncertainty propagation and cross-study replication can address persistent concerns about single-study eQTL dependence and benchmark leakage in gene prioritization methods.

4 Methods

4.1 GWAS summary statistics and harmonization

We obtained publicly available summary statistics for eight cardiometabolic traits from published large-scale GWAS (Supplementary Table 1): LDL cholesterol, HDL cholesterol, triglycerides, and

total cholesterol from GLGC; coronary artery disease from CARDIoGRAMplusC4D; type 2 diabetes from DIAGRAM; systolic and diastolic blood pressure from ICBP.

Summary statistics were harmonized to GRCh38 using the UCSC liftOver tool (minimum match = 0.95). Variants with imputation INFO score < 0.8, minor allele frequency < 0.01, or missing standard error were excluded. Quality control included allele frequency concordance with gnomAD v3.1, palindromic SNP handling for A/T and C/G variants (MAF < 0.4), and duplicate variant removal prioritizing higher imputation quality. Full harmonization code is provided in the repository.

4.2 Fine-mapping with SuSiE-RSS

We applied SuSiE using summary statistics (SuSiE-RSS)**?** to each GWAS locus defined as a 1 Mb window around each genome-wide significant lead variant ($P < 5 \times 10^{-8}$). LD matrices were computed from the 1000 Genomes Phase 3 European panel (503 individuals) using PLINK 2.0. We specified $L = 10$ (maximum independent signals) with 95% credible set coverage and minimum absolute correlation 0.5 within credible sets. Loci were processed if the LD matrix condition number was $< 10^4$ and at least 50 variants were available. We obtained posterior inclusion probabilities (PIPs) for each variant and credible sets representing independent signals.

4.3 Enhancer–gene linking with ABC and PCHi-C

Activity-by-Contact (ABC) Model. We obtained ABC predictions from Nasser et al.**?** for 131 biosamples. The ABC score represents:

$$\mathrm{ABC}_{\mathrm{enhancer,gene}} = \frac{\mathrm{Activity}_{\mathrm{enhancer}} \times \mathrm{Contact}_{\mathrm{enhancer,gene}}}{\sum_e \mathrm{Activity}_e \times \mathrm{Contact}_{e,\mathrm{gene}}} \tag{2}$$

where Activity is measured by H3K27ac ChIP-seq and Contact by Hi-C at 5 kb resolution. We used predictions with ABC score ≥ 0.015 and matched biosamples to GTEx tissues (e.g., HepG2 for liver, GM12878 for blood).

Promoter capture Hi-C (PCHi-C). We integrated PCHi-C data from two sources: Jung et al.**?** (27 cell types from ENCODE) and Javierre et al.**?** (17 primary blood cell types). The Javierre dataset was lifted from GRCh37 to GRCh38 using UCSC liftOver. Contacts with CHiCAGO score ≥ 5 were retained as significant. For each cCRE–gene pair, we took the maximum contact score across matching cell types.

Ensemble linking. We combined ABC, PCHi-C, and distance using weighted logistic regression:

$$P(\mathrm{link}) = \sigma(w_1 \cdot \mathrm{ABC} + w_2 \cdot \mathrm{PCHi\text{-}C} + w_3 \cdot f(\mathrm{distance}) + w_0) \tag{3}$$

where $f(\mathrm{distance}) = \exp(-\mathrm{distance}/50\mathrm{kb})$ and σ is the logistic function. Weights were optimized on the Tier 3 CRISPR benchmark using 5-fold cross-validation with held-out chromosomes.

4.4 Multi-signal colocalization with coloc.susie

We applied coloc in conjunction with SuSiE fine-mapping (coloc.susie)**?** to test colocalization between GWAS credible sets and cis-eQTLs from GTEx v8. For each locus, we:

1. Obtained pre-computed SuSiE fine-mapping for GTEx eQTLs (available from GTEx portal) or ran SuSiE-RSS on eQTL summary statistics
2. Applied coloc.susie with priors $p_1 = 10^{-4}$, $p_2 = 10^{-4}$, $p_{12} = 5 \times 10^{-6}$

3. Reported posterior probability of shared causal variant (PP.H4) for each GWAS credible set–eQTL credible set pair

We analyzed 12 cardiometabolism-relevant tissues from GTEx v8: liver, adipose subcutaneous, adipose visceral omentum, skeletal muscle, heart left ventricle, heart atrial appendage, artery aorta, artery coronary, artery tibial, whole blood, pancreas, and adrenal gland.

Gene–tissue pairs with PP.H4 ≥ 0.8 were considered colocalized. We also computed PP.H3 (distinct causal variants) to identify loci with independent GWAS and eQTL signals.

4.5 Cross-study validation via eQTL Catalogue

For each colocalized gene–tissue pair discovered in GTEx v8, we queried matching datasets from the eQTL Catalogue**?** (Release 6, uniformly processed):

- Blood tissues: BLUEPRINT monocytes, Lepik 2017 blood, GENCORD LCLs
- Adipose: METSIM adipose
- Muscle: FUSION muscle
- Liver: FUSION liver

Replication was defined as: (i) nominal significance ($P < 0.05$) at the lead variant in matching tissue; (ii) effect size correlation $r > 0.5$ across shared variants; (iii) $> 80\%$ allelic direction concordance.

For colocalization replication, we ran coloc.susie on eQTL Catalogue summary statistics and required PP.H4 > 0.5 in the replication dataset. Genes passing replication received full probability weight; non-replicated genes received a $0.8\times$ penalty.

4.6 Noisy-OR path aggregation

Gene-level probabilities were computed using formal noisy-OR aggregation**??**:

$$P(G = 1|\mathbf{paths}) = 1 - (1 - \epsilon) \prod_i (1 - P_i) \tag{4}$$

where $\epsilon = 0.01$ is the leak probability representing unmeasured causal mechanisms, and P_i is the probability of path i contributing to causality.

Path probabilities were computed as edge products:

$$P_{\text{path}} = P(\text{var} \to \text{cCRE}) \times P(\text{cCRE} \to \text{gene}) \times P(\text{gene}|\text{tissue}) \tag{5}$$

We applied correlation corrections for:

- LD between variants: $0.5\times$ penalty for variants in shared credible set
- Tissue correlation: penalty based on GTEx tissue correlation matrix (median r^2)
- Annotation overlap: $0.7\times$ penalty for edges sharing the same cCRE

Independence assumptions and their justification are detailed in Supplementary Note 1.

4.7 Three-tier benchmark construction

Tier 1 (anti-leak holdout): We curated 47 Mendelian genes for familial hypercholesterolemia (LDLR, APOB, PCSK9), hypertriglyceridemia (LPL, APOC2, APOA5), and related disorders from OMIM. We excluded genes within 500 kb of any gene in the Open Targets Genetics L2G training set (downloaded from their GitHub repository, v22.09 release). Two independent curators verified inclusion/exclusion with consensus required.

Tier 2 (drug targets): We queried ChEMBL v32 for approved drugs (max_phase = 4) with cardiometabolic indications (EFO: cardiovascular disease, hyperlipidemia, diabetes mellitus). We extracted mechanism-based targets and filtered to exclude genes in any published L2G training set. Final set: 89 genes.

Tier 3 (CRISPR validation): We obtained enhancer–gene pairs from CRISPRi screens: Fulco et al.**?** (K562, 664 pairs) and Gasperini et al. 2019 (K562 + WTC11, 183 pairs). Pairs were filtered to genes within 1 Mb and cardiometabolism-relevant cell types.

Full provenance including PMID, extraction date, curation criteria, and anti-overlap verification is provided in Supplementary Table 2 and the `data/manifests/benchmark_genes.yaml` file.

4.8 Calibration assessment

Per-module calibration was assessed using:

1. Reliability diagrams with 10 equal-width probability bins
2. Expected Calibration Error (ECE): $\text{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} |p_b - \hat{p}_b|$ where p_b is the mean predicted probability in bin b and $\hat{p}_b$ is the observed frequency
3. Maximum Calibration Error (MCE): $\max_b |p_b - \hat{p}_b|$
4. Brier score decomposition (reliability, resolution, uncertainty)

Calibration was computed separately for: variant PIP (against fine-mapping simulation truth), coloc PP.H4 (against colocalization simulation), cCRE–gene probability (against Tier 3 CRISPR truth), and final gene probability (against Tier 1 benchmark).

Confidence intervals were computed via 1,000 bootstrap replicates resampling loci with replacement.

4.9 Confidence intervals

95% confidence intervals for gene probabilities were computed via bootstrap resampling (1,000 iterations):

1. Resample loci with replacement
2. Perturb edge probabilities according to estimated uncertainty: binomial variance for discrete edges, beta distribution for continuous probabilities
3. Recompute noisy-OR aggregation
4. Take 2.5th and 97.5th percentiles

4.10 Baseline comparisons

Open Targets Genetics L2G (v22.09): We downloaded L2G scores from the Open Targets Genetics portal (`https://genetics.opentargets.org/`) for all study–locus pairs matching our traits. Version 22.09 was used throughout for reproducibility.

PoPS: We ran PoPS**?** using default parameters with MAGMA gene-level z-scores as input.

Nearest gene: For each lead variant, we assigned the nearest protein-coding gene (GENCODE v40) by distance to TSS.

4.11 Statistical analysis

All statistical analyses were performed in Python 3.11 using NumPy 1.24, SciPy 1.10, pandas 2.0, and scikit-learn 1.2. Colocalization was performed using the coloc R package (v5.2.3) with the susieR package (v0.12.27) via rpy2. Multiple testing correction used Benjamini-Hochberg FDR where noted. Code is available at the repository.

Data availability

All analyses use publicly available summary-level data; no individual-level genotypes or controlled-access datasets were used.

GWAS summary statistics: LDL-C, HDL-C, triglycerides, total cholesterol from GLGC (`http://csg.sph.umich.edu/willer/public/lipids2013/`); coronary artery disease from CARDIoGRAMplusC4D (`http://www.cardiogramplusc4d.org/`); type 2 diabetes from DIAGRAM (`https://diagram-consortium.org/`); blood pressure from ICBP (`https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000585.v2.p1`).

eQTL data: GTEx v8 (`https://gtexportal.org/`, dbGaP accession phs000424.v8.p2); eQTL Catalogue Release 6 (`https://www.ebi.ac.uk/eqtl/`).

Enhancer–gene linking: ABC Model predictions from Nasser et al. (2021) (`https://www.engreitzlab.org/resources/`); PCHi-C from Jung et al. (2019) (`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118752`) and Javierre et al. (2016) (`https://osf.io/u8tzp/`).

Reference data: 1000 Genomes Phase 3 (`https://www.internationalgenome.org/`); GENCODE v40 (`https://www.gencodegenes.org/`); ENCODE cCREs (`https://screen.encodeproject.org/`).

Benchmark data: OMIM (`https://omim.org/`); ChEMBL v32 (`https://www.ebi.ac.uk/chembl/`); CRISPRi data from Fulco et al. (2019) and Gasperini et al. (2019).

Processed outputs: The mechanism atlas, intermediate results, and all data manifests with SHA256 checksums are archived at Zenodo (DOI: 10.5281/zenodo.17798899).

Code availability

All analysis code is available at:

- **GitHub repository:** `https://github.com/ProgrmerJack/Mechanism-GWAS-Causal-Graphs` (release v1.0.0, commit `7321825`)
- **Zenodo archive:** DOI: 10.5281/zenodo.17798899 (persistent, citable)
- **Container image:** Docker image available via `docker pull ghcr.io/progrrmerjack/mechanism-gwas:v1.`

One-command reproduction: Key figures can be regenerated using:

```
snakemake --cores 8 results/figures/fig1_overview.pdf \
  results/figures/fig2_bridge.pdf
```

Full reproduction instructions are provided in `REPRODUCE.md`. The Snakemake workflow (`workflow/Snakefile`) reproduces all analyses from raw inputs with versioned dependencies specified in `environment.yml`.

License: MIT License (permissive, commercial use allowed).

References

Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human gwas trait-associated loci. *Nature genetics* **53**, 1527–1533 (2021).

Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nature genetics* **53**, 1290–1299 (2021).

Buniello, A. *et al.* The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**, D1005–D1012 (2019).

Gallagher, M. D. & Chen-Plotkin, A. S. The post-gwas era: from association to function. *The American Journal of Human Genetics* **102**, 717–730 (2018).

Visscher, P. M. *et al.* 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics* **101**, 5–22 (2017).

Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nature genetics* **55**, 1267–1276 (2023).

Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**, 1273–1300 (2020).

Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics* **19**, 437–451 (2018).

Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383 (2014).

Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of crispr perturbations. *Nature genetics* **51**, 1664–1669 (2019).

Figures

Figure 1: **Path-probability framework for GWAS gene prioritization.** (a) The mechanism graph represents explicit paths from variants to traits through regulatory elements, genes, and tissues, with calibrated probabilities at each edge. (b) Five-stage pipeline: SuSiE fine-mapping, ABC/PCHi-C enhancer–gene linking, coloc.susie multi-signal colocalization, eQTL Catalogue replication, and noisy-OR aggregation. (c) Comparison with point-estimate approaches (L2G, PoPS) that collapse the pathway into a single score without uncertainty propagation. (d) Example path decomposition for the SORT1 locus showing interpretable edge probabilities.

Figure 2: **Enhancer–gene linking validation and bridge ablation.** (a) Precision-recall curves for cCRE–gene linking methods on Tier 3 CRISPR benchmark. Ensemble (ABC + PCHi-C + distance) outperforms each component. (b) Bridge ablation: AUPRC change when removing each evidence type. ABC and PCHi-C each contribute independently beyond distance. (c) Performance stratified by cCRE–gene distance. Functional links provide greatest benefit at intermediate distances (20–200 kb) where distance alone is uninformative. (d) Tissue-specific performance. ABC excels in cell lines with matched H3K27ac data; PCHi-C excels in primary blood cells.

Figure 3: **Path-probability models outperform L2G on anti-leak benchmarks.** (a) Recall at rank k on Tier 1 (anti-leak holdout) for path-probability, Open Targets L2G (v22.09), PoPS, MAGMA, and nearest gene. (b) Precision at probability/score thresholds showing improved calibration of path-probability. (c) Performance by benchmark tier (Tier 1: Mendelian, Tier 2: drug targets, Tier 3: CRISPR). (d) Stratification by locus complexity (credible set size, number of genes within 500 kb). Path-probability advantage increases at complex loci.

Figure 4: **Per-module calibration demonstrates probability semantics.** (a) Reliability diagrams for variant PIP, cCRE–gene probability, coloc PP.H4, and final gene probability. Dashed line = perfect calibration. (b) Expected Calibration Error (ECE) by module and trait. All modules achieve ECE < 0.05. (c) Calibration comparison: path-probability vs L2G scores. L2G scores are poorly calibrated (ECE = 0.18). (d) Hosmer-Lemeshow test p-values confirming calibration fit.

Figure 5: **Cross-study replication via eQTL Catalogue.** (a) Replication rate of GTEx colocalization signals in eQTL Catalogue by tissue. Blood and adipose tissues show highest replication (82–85%). (b) Effect size correlation between GTEx and eQTL Catalogue ($r = 0.89$). (c) Benchmark performance for replicated vs non-replicated genes. Replicated genes achieve +11% recall at rank 20. (d) Impact of replication penalty on calibration. Penalizing non-replicated signals improves ECE from 0.052 to 0.038.

Figure 6: **Interpretable mechanism paths for cardiometabolic loci.** (a) SORT1 locus (1p13, LDL-C): full path with edge probabilities, bootstrap confidence intervals, and comparison to L2G score. (b) PCSK9 locus (1p32): shared liver mechanism for LDL-C and CAD with high replication confidence. (c) TCF7L2 locus (10q25): tissue-divergent paths showing pancreas mechanism for T2D and adipose for lipids. (d) ANGPTL3 locus: example of mechanism discovery—path analysis identifies liver-specific regulation later validated by therapeutic targeting.

Extended Data Figures

Figure 7: **Extended Data Figure 1: Dataset summary.** (a) GWAS traits, sample sizes, ancestries, and number of genome-wide significant loci. (b) GTEx v8 tissues, sample sizes, and number of eGenes. (c) ABC and PCHi-C biosample coverage and overlap. (d) Benchmark gene provenance and anti-leak verification.

Figure 8: **Extended Data Figure 2: Fine-mapping quality control.** (a) Distribution of credible set sizes across loci. (b) Posterior inclusion probability distributions. (c) LD matrix condition numbers and filtering. (d) Sensitivity to LD reference panel mismatch (1000G EUR vs UKBB).

Figure 9: **Extended Data Figure 3: Multi-signal colocalization.** (a) Comparison of coloc.susie vs single-causal coloc at loci with multiple SuSiE credible sets. (b) Example locus with two independent signals assigned to different genes. (c) Error rate at multi-signal loci: single-causal coloc assigns PP.H4 to wrong signal in 34% of cases. (d) Impact on gene prioritization accuracy.

Figure 10: **Extended Data Figure 4: Sensitivity analyses.** (a) Robustness to coloc prior specification. (b) Robustness to ABC score threshold. (c) Robustness to credible set coverage (90%, 95%, 99%). (d) Robustness to leak probability ϵ in noisy-OR model.

Figure 11: **Extended Data Figure 5: Negative controls.** (a) Tissue swap control: colocalization with mismatched tissues yields near-zero benchmark performance. (b) LD shuffle control: randomizing LD structure destroys signal. (c) Distance scramble: randomizing cCRE–gene distances eliminates enhancer–gene linking advantage. (d) Null trait control: random phenotypes yield expected false positive rate.

Figure 12: **Extended Data Figure 6: Failure case analysis.** (a) Distribution of false negatives by mechanism type (coding vs regulatory). (b) False negatives are enriched for tissue-specific regulation in tissues not well-represented in GTEx. (c) False positives are enriched at loci with high LD complexity. (d) Examples of incorrect prioritizations and hypothesized reasons.

Figure 13: **Extended Data Figure 7: Bootstrap confidence intervals.** (a) Distribution of confidence interval widths by path complexity. (b) Coverage of nominal 95% intervals (empirical coverage 94.2%). (c) Correlation between CI width and prediction accuracy. (d) Uncertainty decomposition by edge type.

Figure 14: **Extended Data Figure 8: Mechanism atlas summary.** (a) Distribution of path probabilities across 2,134 cardiometabolic loci. (b) Number of genes per locus with PP > 0.5 (median = 1.3). (c) Tissue enrichment of high-confidence gene–tissue pairs. (d) Interactive atlas browser demonstration (screenshot).