HYPOTHESIS GENERATION AND SEMANTIC ATTRIBUTION IN INVERSE PERTURBATION MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding gene regulatory relationships at scale can help identify intervention points and anticipate unintended effects. However, connecting a black-box model's learned computation to biological knowledge remains difficult, particularly for unfamiliar cellular states. We investigate whether a sparse, typed gene network can generate inspectable hypotheses for steering malignant transcriptional programs toward expression patterns associated with favorable prognosis. Using LINCS L1000 CRISPR/Cas9 profiles paired with matched plate controls, we train a graph neural network to recover the perturbed gene from baseline and endpoint expression. Its nodes combine ProTrek-derived protein representations with expression features; its edges encode signed regulatory relationships and other gene interactions. Our focus is how representations learned from perturbation responses use these identities and relationships to support individual nominations. We then query the model in a specified cancer-cell context with endpoints that upregulate favorable prognostic markers and downregulate unfavorable ones. At this fixed request, integrating from mean projected protein representations to actual representations allocates a nomination score difference through node states, individual layer–edge messages, and residual connections. In detailed A549–PYGL and A549–CCDC86 cases, supporting and opposing contributions coexist. Connected supporting structures suggest hypotheses involving glycogen-associated regulatory/metabolic propagation and physical interactions associated with ribosome biogenesis. Preserving gene and relation identities makes these learned computational dependencies explicit, allowing comparison with biological knowledge and framing mechanistic hypotheses for experimental investigation.

1 INTRODUCTION

Perturbation experiments probe cellular function by measuring responses to interventions. Perturb-seq made such measurements scalable by combining pooled genetic screens with single-cell expression profiling (Adamson et al., 2016; Dixit et al., 2016), later extending to genome-scale surveys (Replogle et al., 2022). These data support models of how interventions change cell state, a central capability in the emerging AI virtual-cell vision (Bunne et al., 2024).

Within this setting, forward models predict expression responses to specified interventions (Klein et al., 2025; Roohani et al., 2024; Lotfollahi et al., 2023; Adduri et al., 2025). Inverse methods instead nominate interventions for a desired state (Gonzalez et al., 2026). We take baseline and desired expression profiles as inputs and rank single-gene perturbations, then inspect what supports those nominations.

Logistic regression offers a familiar starting point: its coefficients connect named inputs to a logit. In an interacting model, however, contributions depend on other inputs and on the comparison being made. We therefore ask whether the representation can preserve the scientific identities needed to inspect those dependencies.

To make these dependencies inspectable, we combine protein features, paired expression, and typed graph messages at gene/protein nodes (Figure 1). Under a fixed request, semantic attribution traces protein identity-dependent score differences while retaining gene and relationship identities. We

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

illustrate this account with PYGL and CCDC86 under one A549 prognosis request, using signed conductance, finite message suppression, and compound annotations to inform nomination-specific hypotheses. Predictive evaluation provides context for the learned representations we inspect.

2 Related work

Perturbation prediction and inverse design. PDGrapher couples perturbagen discovery with an intervention-conditioned response model (Gonzalez et al., 2026). Causal differential networks recover intervention targets from observational and perturbed distributions (Wu et al., 2025). Our direct inverse classifier learns on a fixed communication graph.

Biologically structured representations. DCell and DrugCell show how biological entities can organize neural computation (Ma et al., 2018; Kuenzi et al., 2020). We build on this principle using gene-indexed states and relation-specific message passing (Bronstein et al., 2021; Schlichtkrull et al., 2018) to inspect individual inverse nominations.

Attribution and graph explanations. IG allocates a scalar difference relative to a reference (Sundararajan et al., 2017), while conductance follows it through hidden representations (Dhamdhere et al., 2019). For graph explanations, GNNExplainer optimizes subgraph and feature masks (Ying et al., 2019), whereas GNN-LRP attributes predictions to higher-order relevant walks (Schnake et al., 2022). Our signed allocation through gene states, typed messages, and residuals precedes display selection. Routes then organize marginal transitions without higher-order route scores. Our evaluation concerns signed accounting and finite message effects; visual plausibility alone is not validation (Adebayo et al., 2018). We claim neither unique explanations nor comparative superiority.

3 Learning context-dependent gene representations

We learn these representations by identifying knockouts from paired expression, allowing graph neighbors to shape each gene's state.

3.1 Observed pairs and gene-indexed states

For training, each LINCS L1000 observation pairs a CRISPR/Cas9 knockout endpoint with matched plate controls and labels the targeted gene (Subramanian et al., 2017; Broad Institute, 2017). The Level-3 cohort contains 269,769 eligible pairs across 17 lines and 11 tissues. Using LINCS BING (Best Inferred Genes), we retain 978 measured and 9,095 inferred expression coordinates (10,073 total) to condition more nodes with protein features on cellular state. Training-only moments normalize both channels. We test new cell contexts by holding out whole lines in multi-line tissues; singleton tissues permit only physical-replicate holdouts. Whole-plate validation selects checkpoints (Appendix A).

The 17,308 nodes comprise expression genes, retained perturbation-target labels, and their one-hop neighbors in STRING and CollecTRI (Szklarczyk et al., 2025; Müller-Dott et al., 2023). Directed families have separate forward/reverse transport channels; computational reversal does not reverse biological causality. At each node,

$$h_i^{(0)} = m_i^p P_\theta(z_i) + m_i^x a_i \big[B_\theta(x_i^{(0)}) + A_\theta(x_i^{(1)})\big]. \tag{1}$$

The shared encoders combine frozen ProTrek-derived vectors (Su et al., 2026) and paired expression. Availability masks suppress missing inputs; a_i masks both expression channels together during training. The protein projector is learned.

Typed messages contextualize these states through repeated updates:

$$M_i^{(\ell)} = \sum_{r,j} \widetilde{A}_{ij}^r W_{\ell,r} h_j^{(\ell)}, \tag{2}$$

$$h_i^{(\ell+1)} = \mathrm{LN}_\ell\Big(h_i^{(\ell)} + \mathrm{SiLU}(M_i^{(\ell)})\Big), \quad \ell = 0, \ldots, L-1. \tag{3}$$

Adjacency fixes where messages travel; transforms and node-local LayerNorm are learned separately at each layer. The residual carries the current state, including earlier messages. $H^{(L)} \in \mathbb{R}^{N\times 128}$ retains one identifiable row per gene.

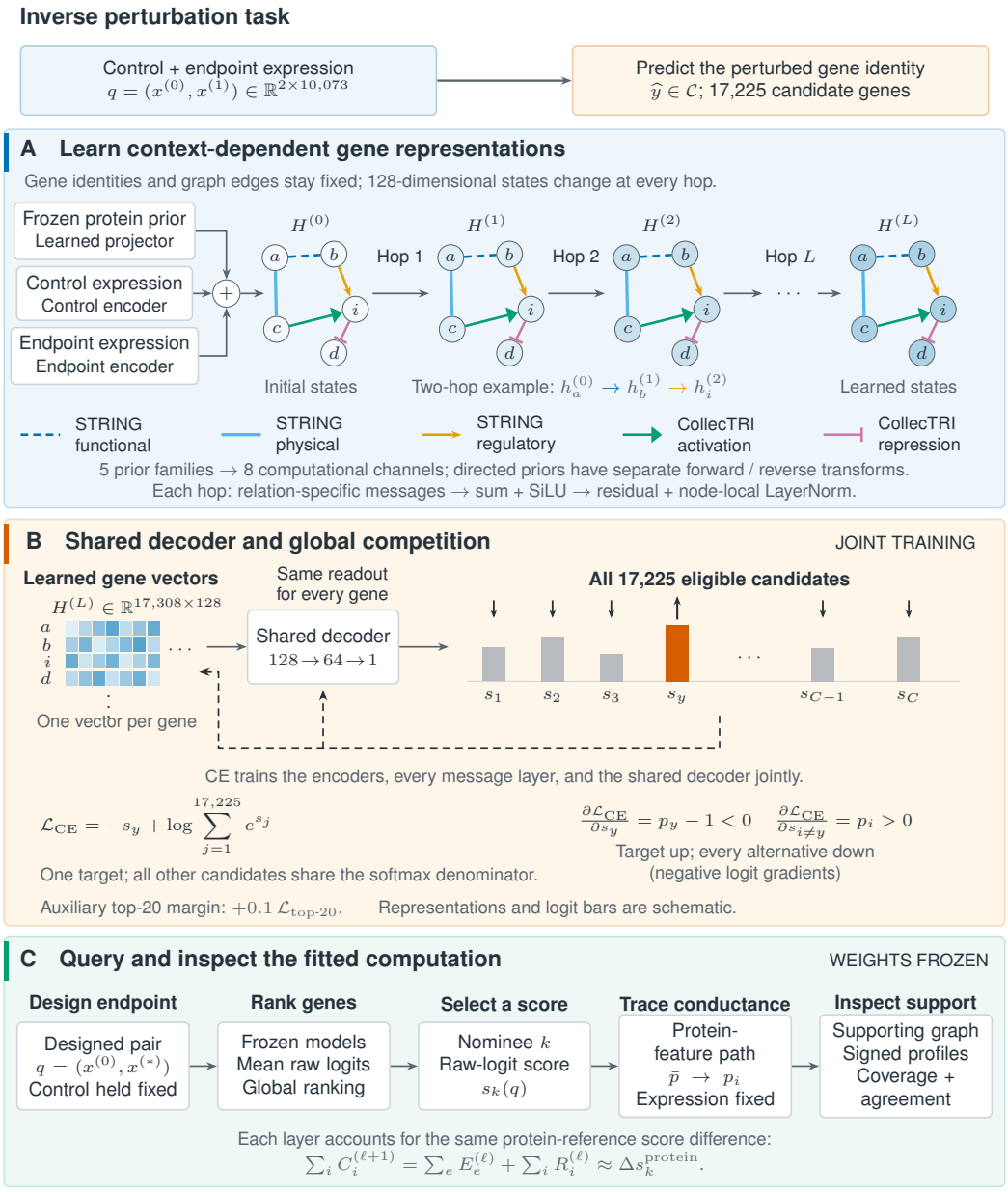

Figure 1: **Inverse perturbation prediction and computational inspection.** The task maps a control/endpoint expression pair ($2 \times 10{,}073$ inputs) to the perturbed gene's identity. **(A)** Shared encoders and typed message layers learn a 128-dimensional state per graph gene; five prior families supply eight transport channels. **(B)** A shared decoder scores 17,225 nodes with mapped ProTrek vectors; 83 others remain in the 17,308-node graph as context. Global CE jointly trains the representations and decoder. Arrows show negative logit gradients, not guaranteed changes under shared-parameter updates. **(C)** With weights frozen, a designed endpoint $x^{(*)}$ replaces the observed $x^{(1)}$; this expression request yields a mean-logit ranking across the held-out H1 and H3 fits. For each selected nominee, protein representations move from their common projected mean to their actual values while expression stays fixed. Conductance attributes the score difference through node states (C), messages (E), and carried-state residuals (R). Each layer accounts for the same difference; complete layers are not added. Corresponding signed scalar contributions are averaged before drawing a supporting graph. The accompanying profiles retain opposition and selection coverage. States and bars are schematic; edge glyphs denote biological priors, not attribution signs.

3.2 SHARED DECODING AND PREDICTIVE CONTEXT

The shared decoder ranks 17,225 nodes with mapped ProTrek vectors; 83 others remain in the computation but are ineligible candidates. Genes without positive training examples remain eligible; unseen-target accuracy is untested. Training uses global cross-entropy plus a margin against the hardest 20 incorrect candidates (Appendix B). Genetic labels keep the intended target explicit; multitarget chemical action is examined later through annotations (Keiser et al., 2009).

Uniform random ranking has expected MRR ≈ 0.00060 over these candidates. Across twelve fits, one-to-four-hop held-replicate MRR rises from 0.1213 to 0.1352, while unseen-line MRR is similar (0.1463 versus 0.1456) and CE worsens. Depth changes capacity and reach; no uniformly preferred depth is established.

4 FROM A DESIRED PROGRAM TO A NOMINATION

After training, we freeze the weights and replace the observed endpoint with a requested profile to nominate interventions. HPA survival associations supply expression directions, not causal effects (Uhlen et al., 2017; Yuan et al., 2025; Human Protein Atlas, n.d.). For A549, validated markers in *Lung Adenocarcinoma (validation)* specify 2 increases and 155 decreases, each of amplitude 2 H1-reference SDs. Other coordinates remain unchanged.

One native baseline/endpoint pair is transformed through each model's normalizer. Four-hop H1 and H3 both held A549 out of training and validation; their mean raw logit defines the ranking. PYGL and CCDC86 are unfavorable markers requested to decrease and remain eligible knockout candidates. Directly reducing an unfavorable marker agrees with its requested direction; a single-gene perturbation may also affect other coordinates. Candidates are ranked against the complete multigene request. Of the top 50 nominees, 20 have edited own inputs. The graph attribution conditions on this request; it does not separately quantify the nominee's own-expression contribution.

The notebook first follows the H1 query through encoded inputs and four message layers using one joint UMAP (McInnes et al., 2018). Points are genes, with final nomination colors held fixed (Appendix Figure 11). These are stages of one model, not expression trajectories. Geometry makes the representation visible; conductance supplies its quantitative account.

5 FOLLOWING A NOMINATION THROUGH NODES AND MESSAGES

To inspect a nomination, we trace supporting and opposing contributions through named genes and relationships relative to a protein reference.

5.1 FIX THE QUESTION AND PROTEIN REFERENCE

For nominee k and fixed expression request q, the reference replaces covered genes' projected proteins with their common mean. Write $p_i = P_\theta(z_i)$ and $r_i(q) = m_i^x[B_\theta(x_i^{(0)}) + A_\theta(x_i^{(1)})]$, the expression term in Equation 1 with inference mask $a_i = 1$. Then

$$\bar{p} = \frac{1}{|\mathcal{V}_p|} \sum_{i \in \mathcal{V}_p} p_i, \qquad h_i^{(0)}(\alpha) = r_i(q) + m_i^p \big[(1-\alpha)\bar{p} + \alpha p_i\big], \tag{4}$$

$$\Delta s_k^{\text{protein}} = s_k(1) - s_k(0), \qquad 0 \le \alpha \le 1. \tag{5}$$

The mean is taken after the learned projector over protein-covered nodes. All covered proteins move jointly, including those at unedited expression coordinates. Expression, weights, and graph remain fixed; representations and sensitivities change along the path. The reference removes between-gene protein differences, while expression and graph position can still distinguish genes: a shared decoder does not imply equal logits or a random guess. This reference isolates protein distinctions while preserving the requested cellular context. Expression attribution would answer the complementary question of which input edits change the nomination; the present account conditions on those edits.

5.2 MEASURE WHAT EACH STATE AND CONNECTION CARRIES

Conductance combines the score's sensitivity to a representation with that representation's movement along the path (Dhamdhere et al., 2019). The node value is

$$C_i^{(\ell)} = \int_0^1 \left\langle \frac{\partial s_k}{\partial h_i^{(\ell)}}, \frac{dh_i^{(\ell)}}{d\alpha} \right\rangle d\alpha. \tag{6}$$

At the initial state, $C_i^{(0)}$ is protein-feature IG summed within gene i. After message passing, $C_i^{(\ell)}$ follows the allocation carried by that gene's current representation, which can include earlier information from other genes. Every value concerns the same final nominated logit.

For an edge $e = (j \to i, r)$, let $m_e^{(\ell)} = \widetilde{A}_{ij}^r W_{\ell,r} h_j^{(\ell)}$. The receiver sums these messages into $M_i^{(\ell)}$, while $u_i^{(\ell)} = h_i^{(\ell)} + \mathrm{SiLU}(M_i^{(\ell)})$ is the pre-normalization state. Their two branches have separate conductances:

$$E_e^{(\ell)} = \int_0^1 \left\langle \frac{\partial s_k}{\partial M_i^{(\ell)}}, \frac{dm_e^{(\ell)}}{d\alpha} \right\rangle d\alpha, \tag{7}$$

$$R_i^{(\ell)} = \int_0^1 \left\langle \frac{\partial s_k}{\partial u_i^{(\ell)}}, \frac{dh_i^{(\ell)}}{d\alpha} \right\rangle d\alpha. \tag{8}$$

The residual carries the current representation, including earlier messages. The chain rule reconciles these contributions with the receiving node and the same total difference:

$$C_i^{(\ell+1)} = \sum_{e:e\to i} E_e^{(\ell)} + R_i^{(\ell)}, \qquad \sum_i C_i^{(\ell+1)} = \sum_e E_e^{(\ell)} + \sum_i R_i^{(\ell)} = \Delta s_k^{\mathrm{protein}}. \tag{9}$$

A node state provides one account; its incoming messages and residual provide the corresponding branch account. Each complete layer recovers the same difference, so layer totals are not added. Positive and negative values support or oppose that logit difference; relationship signs retain their separate biological-prior meaning.

5.3 COMPUTE THE FULL FIELD, THEN SELECT A READABLE VIEW

We evaluate the integrals in float64 using shared Gauss–Legendre points. Acceptance requires endpoint parity with ordinary scoring, full-array refinement, per-layer completeness, and incoming/outgoing reconciliation. Member calculations settled at 32 or 64 points; Appendix D reports the saved errors and implementation details.

H1 and H3 receive the same native request but have their own projectors and protein references. We average corresponding signed scalar node, message, and residual contributions, not hidden feature coordinates. Their mean actual logit is also the ranking score. Positive selection occurs after averaging so opposing member values can cancel.

For display, edited markers anchor widest positive routes through the layered computation, including residual transitions. The consensus uses complete member route witnesses, not paths assembled from unrelated fragments. The resulting network retains intermediates and relationship identities. It is a partial supporting view; signed profiles and per-layer coverage expose what it omits. Node size summarizes peak positive node conductance, fill shows the requested edit, and edge glyphs retain biological prior semantics (Appendix E).

6 TWO NOMINATIONS FROM THE SAME REQUEST

Applying this account to PYGL and CCDC86 reveals different attributed structures under the same request, motivating distinct hypotheses and computational checks.

6.1 PYGL: GLYCOGEN-ASSOCIATED REGULATORY AND METABOLIC PROPAGATION

PYGL has mean-logit rank 6, with member ranks 5 and 10. Its selected structure includes phosphorylase-kinase components, glycogen branching/storage machinery, and a CollecTRI

HIF1A→GYS1 activation route into glycogen synthesis (Figure 2). PYGL knockdown reduces proliferation and migration in lung cancer cells (He et al., 2023); subsequent NSCLC work connects PYGL-dependent glycogenolysis to glycolytic flux and tumor growth (Huo et al., 2026). Broader cancer studies connect hypoxia-induced glycogen metabolism to redox balance and avoidance of senescence (Favaro et al., 2012). Together these motivate a glycogenolysis-fueled glycolysis hypothesis in hypoxia-related signaling. The request does not specify oxygen tension, and the attribution measures neither metabolic flux nor the relative importance of metabolic and mitotic mechanisms.

The profile in Figure 3 shows supporting and opposing allocations recovering the +8.8719-logit protein-reference difference. At $3 \to 4$, signed messages contribute +4.58 and the residual +4.30 logit units. This motivates a propagation-oriented interpretation: glycogen-associated regulatory relationships connect the nominee to a broader metabolic program. Whether PYGL loss produces the requested expression program through these processes remains the biological prediction.

6.2 CCDC86: A PHYSICAL-INTERACTION CONTEXT FOR RIBOSOME BIOGENESIS

CCDC86 ranks 3 in the mean, with member ranks 4 and 2. Its selected network includes nucleolar/ribosome-associated genes, with STRING functional links to WDR18, NLE1, and large-subunit ribosomal proteins and physical links to EIF6, LAS1L, ZNF593, and GNL2 (Appendix Figure 13), motivating a hypothesis about partners involved in ribosome biogenesis and complex organization. CCDC86 has established chromosome-periphery and cell-division literature (Stamatiou et al., 2023) and functional evidence in glioma (Liu et al., 2025). The context-specific hypothesis is that CCDC86 loss can move A549 toward the requested program, with nucleolar/ribosome processes participating. An independently measured response aligned with the request would support the endpoint prediction; a confirmed perturbation without that response would count against it. A response without the proposed functional involvement would weaken the mechanistic lead.

After three message rounds, the smallest set carrying 90% of positive node conductance contains 8 nodes for PYGL but 1 for CCDC86. CCDC86's final residual carries +12.02 logit units, while final messages contribute -0.31 (Appendix Figure 12). Earlier information is already present in that carried state. Together, the cases illustrate propagation through regulatory/metabolic relationships and attribution involving physical partners; both use the same message-passing architecture. These are interpretations of distinct allocations, not measurements of metabolic flux, complex assembly, or biological timing. Many higher-ranked nominees are well-characterized cancer genes; we selected PYGL as a nominee with supporting lung-cancer literature and CCDC86 as one with limited lung-cancer characterization. Complete rankings and signed fields remain available.

6.3 CHECKING CONDUCTANCE WITH FINITE MESSAGE SUPPRESSION

The profiles allocate the protein-reference score difference; suppression checks whether sampled allocations track finite changes in that computation. Under the same request and held-out H1/H3 models, we suppress one typed, directed message at one transition. Other messages, residuals, and degree normalization remain fixed. For member f, let $s^{\setminus e}_{k,f}(\alpha)$ denote the resulting forward-pass score. We measure

$$D^{\text{actual}}_{e,f} = s_{k,f}(1) - s^{\setminus e}_{k,f}(1), \tag{10}$$

$$D^{\text{gap}}_{e,f} = [s_{k,f}(1) - s_{k,f}(0)] - [s^{\setminus e}_{k,f}(1) - s^{\setminus e}_{k,f}(0)]. \tag{11}$$

Gap drop tests the same actual-minus-reference contrast allocated by conductance; Actual drop measures the nomination-score change. We average member drops before correlating them with mean signed conductance. Neither finite drop is an attribution identity; no integral is recomputed.

Figure 4 summarizes 89 PYGL and 96 CCDC86 sampled messages. Gap correlations are positive at every transition for both nominees: CCDC86 ranges from $\rho = +0.63$ to +0.92 and PYGL from +0.09 to +0.54. Positive/negative selections have correspondingly signed average drops and control means lie closer to zero (Appendix Table 7). Individual PYGL gap signs agree less often (18/24 positive, 15/24 negative) than CCDC86's (23/24, 22/24); single-message deletion is a finite, nonlinear change that path-averaged conductance need not match edge by edge. These correlations describe the stratified sample, not all edges, total layer reliance, or necessity of the displayed network.

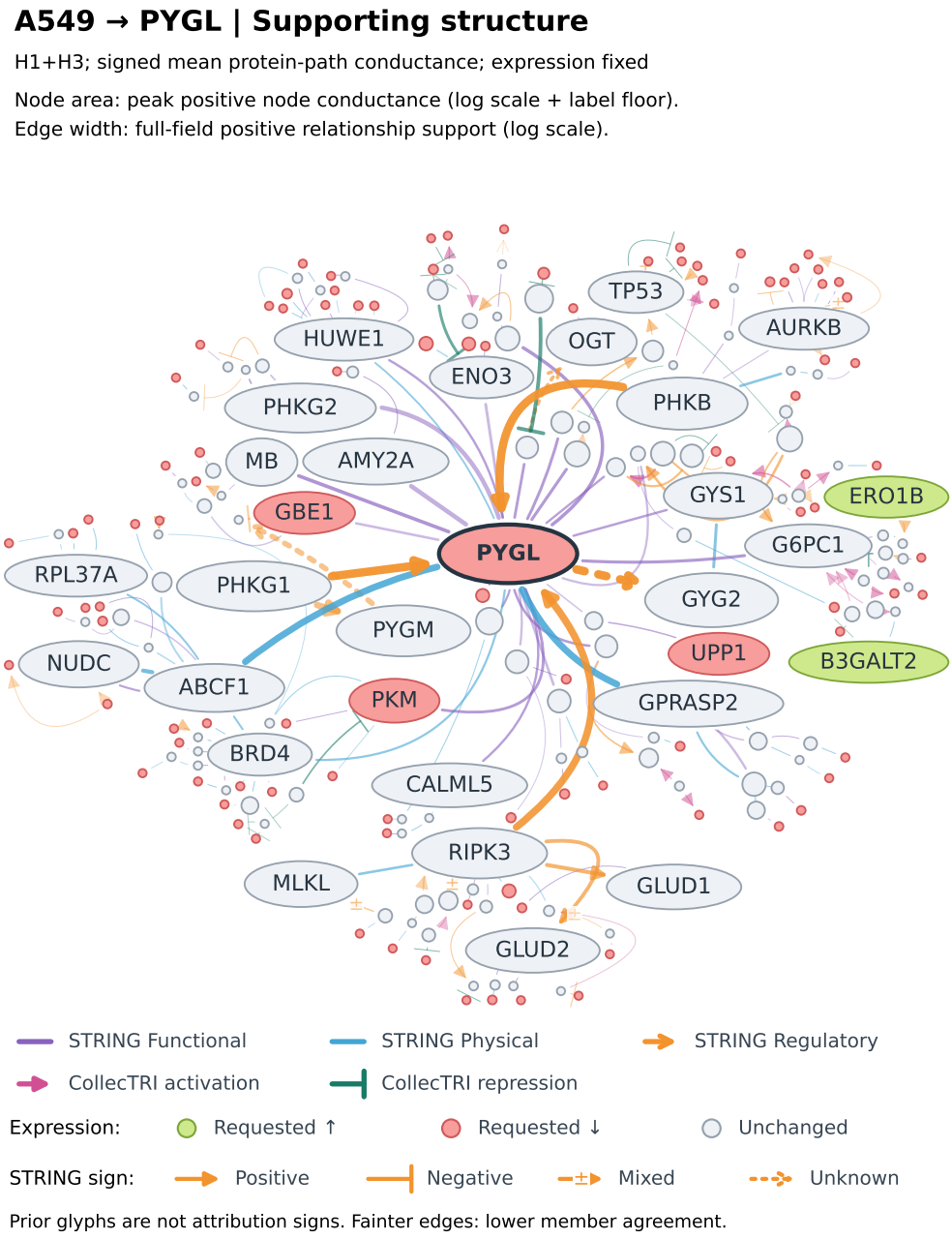

Figure 2: **PYGL: an inspected supporting structure.** Complete selected member-witnessed routes under the common A549 request, with signed scalar conductance averaged across H1/H3 before positive selection. Node fill denotes requested increase, decrease, or no edit. Size summarizes each gene's maximum positive node conductance; edge width summarizes full-field positive conductance for a displayed relationship. Colors and arrow/bar tips retain the supplied relationship family and biological prior sign, not the sign of a learned effect. Fainter links indicate lower member agreement. The profile in Figure 3 supplies signed accounting and the coverage of this selected view.

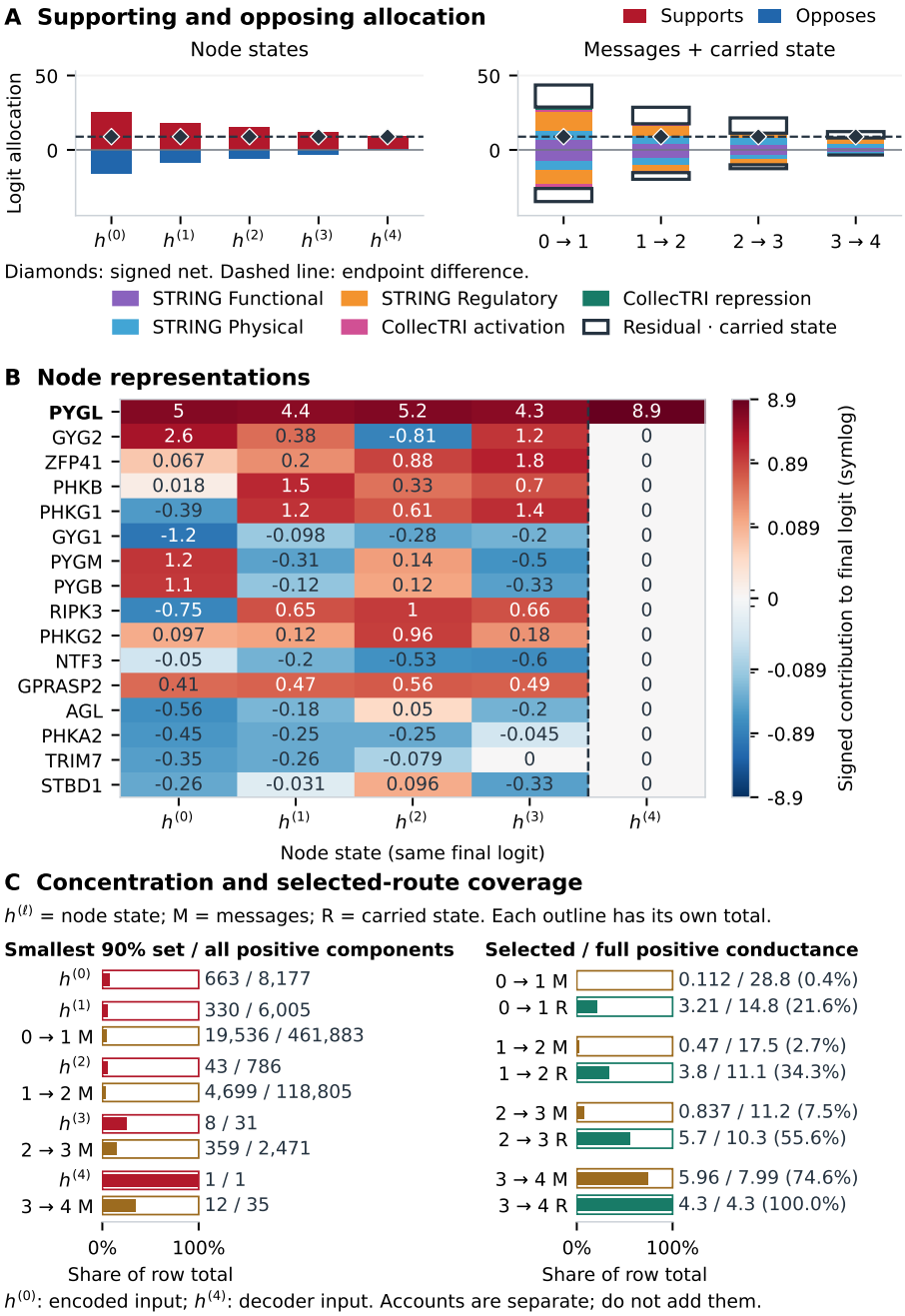

Figure 3: **PYGL: signed conductance behind the network.** A gives separate node and transition accounts of the same protein-reference difference. B attributes the same final logit at each node state, from encoded input $h^{(0)}$ to decoder input $h^{(4)}$; the nominee-only final column is architectural. C indexes node states by $h^{(\ell)}$ and message/residual transitions by $\ell \to \ell + 1$, separating component counts from selected positive mass with a different denominator for each row. Low early-layer coverage means the displayed routes retain a small share of that transition's positive conductance; it does not imply that early layers are unimportant or unable to distinguish the nominee. Residuals include earlier messages; complete layer totals are not added.

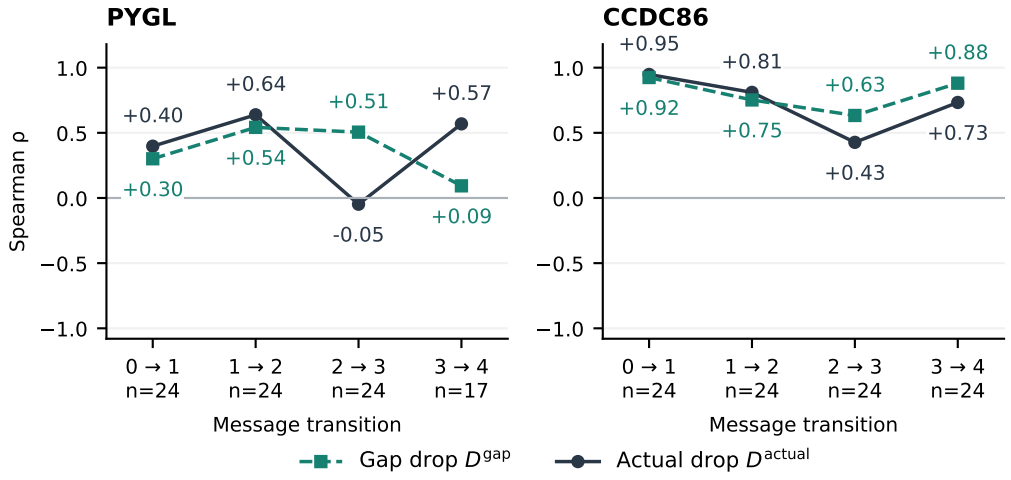

Figure 4: **Conductance and finite message suppression.** Correlation with Gap drop and Actual drop at each transition, using the case profiles' request and H1/H3 members. Each suppression is evaluated at the final nominee logit; n counts sampled messages. PYGL's final $n = 17$ comprises six positive, six negative, and the five available controls (Appendix F). Correlations describe these stratified samples.

7 From an inspectable hypothesis to compound evidence

The explorer is an interactive interface in the accompanying notebook for configuring expression requests and inspecting ranked nominees and their attribution networks. Its compound panel matches nominees by exact symbol to the pinned Broad Drug Repurposing Hub (Corsello et al., 2017; Broad Institute, 2025), returning source-reported mechanisms, all targets, stages, and indications without changing rankings or the selected network.

The source labels CP-316819 and CP-91149 as preclinical glycogen-phosphorylase inhibitors (Appendix Table 8); complete PYGL matches remain in the evidence bundle. CCDC86 has no exact-symbol match. These annotations identify possibilities to examine alongside the rationale, not established selectivity, action compatibility, or efficacy.

8 Limitations and conclusion

The observed-pair task measures recovery of a perturbation label, whereas the designed request is exploratory. Survival associations specify directions, not intervention effects or amplitudes, and a bulk cancer cohort is only a coarse match to a cell-line context. L1000 profiles aggregate expression across cell populations. Perturbation-associated differences may reflect both changes within cells and shifts in the proportions of cells occupying different cell-cycle or other cellular states. The present analysis does not distinguish these sources, so attributed structures should not be interpreted as phase-specific regulatory mechanisms. Training does not filter technical QC; the query baseline uses a QC-passing plate, and both nominees stay in the top 7 across all 44 such plates (Appendix H). Protein references and interpolation are computational comparisons. STRING/CollecTRI supply biological grounding for the graph; suppression checks message use within the fitted model. Neither establishes that a complete displayed route is a necessary biological mechanism.

The two examples establish case-specific computational accounts; stability across independent refits and population-wide generality remain untested. Preserving gene and relation identities connects nomination, signed allocation, and finite message effects in one inspectable account. The contrasting regulatory/metabolic and physical-interaction contexts, together with existing compound evidence, turn the nominations into explicit mechanistic hypotheses for experimental investigation.

AI USE STATEMENT

Generative AI tools assisted with conceptual framing and methodology design, implementation and debugging, data-cleaning and identifier-mapping code, numerical verification code, interpretation of saved results, literature discovery, figure preparation, and drafting and editing this manuscript. Expression requests follow an explicit rule applied to the stated source data. Reported numerical results are outputs of executed code, not numbers generated by a language model. Automated checks cover source and artifact identities, splits and coordinates, checkpoint restoration, selected derivative comparisons, attribution completeness and refinement, and manuscript exports. AI-assisted visual checks informed layout. These checks are not independent biological validation and do not replace the authors' responsibility for the final content.

REPRODUCIBILITY STATEMENT

The notebook presents the observations, model, learning procedure, request, and conductance calculations beside their shared implementation. The recorded load-mode execution has 39 code cells and identifies both primary cases, their member calculations, signed arrays, profiles, and finite suppression results. Frozen paper exports retain source hashes and numerical locators and build without raw assay matrices or model checkpoints. Appendix I distinguishes this current evidence from a separate historical full-training execution.

REFERENCES

Britt Adamson, Thomas M. Norman, Marco Jost, Min Y. Cho, James K. Nuñez, Yuwen Chen, Jacqueline E. Villalta, Luke A. Gilbert, Max A. Horlbeck, Marco Y. Hein, Ryan A. Pak, Andrew N. Gray, Carol A. Gross, Atray Dixit, Oren Parnas, Aviv Regev, and Jonathan S. Weissman. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*, 167(7):1867–1882.e21, 2016. doi: 10.1016/j.cell.2016.11.048.

Abhinav K. Adduri, Dhruv Gautam, Beatrice Bevilacqua, Alishba Imran, Rohan Shah, Mohsen Naghipourfar, Noam Teyssier, Rajesh Ilango, Sanjay Nagaraj, Mingze Dong, Chiara Ricci-Tam, Christopher Carpenter, Vishvak Subramanyam, Aidan Winters, Sravya Tirukkovular, Jeremy Sullivan, Brian S. Plosky, Basak Eraslan, Nicholas D. Youngblut, Jure Leskovec, Luke A. Gilbert, Silvana Konermann, Patrick D. Hsu, Alexander Dobin, Dave P. Burke, Hani Goodarzi, and Yusuf H. Roohani. Predicting cellular responses to perturbation across diverse contexts with State. *bioRxiv*, 2025. doi: 10.1101/2025.06.26.661135. URL https://www.biorxiv.org/content/10.1101/2025.06.26.661135v1. Preprint, version 1.

Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity Checks for Saliency Maps. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://papers.nips.cc/paper/8160-sanity-checks-for-saliency-maps.

Broad Institute. Perturbagen types and controls. https://clue.io/connectopedia/perturbagen_types_and_controls, 2017. Page last modified December 8, 2017; accessed September 25, 2026.

Broad Institute. CRISPR v1.0 Data release README. https://s3.amazonaws.com/assets.clue.io/clue/public/resources/Broad_Institute_LINCS_CRISPRv1.0_early_access_data_deposition_README.pdf, 2018a. July 2018 early-access release documentation; accessed September 25, 2026.

Broad Institute. Is the L1000 data "fully QC'd"? https://clue.io/connectopedia/l1000_qc, 2018b. Page last modified January 5, 2018; accessed September 25, 2026.

Broad Institute. Drug Repurposing Hub: drug information annotation snapshot. https://repo-hub.broadinstitute.org/public/data/repo-drug-annotation-20250818.txt, 2025. Snapshot dated August 18, 2025; checksum and queried rows retained with the code.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021. URL https://arxiv.org/abs/2104.13478v2. Version 2, revised May 2, 2021.

Charlotte Bunne, Yusuf Roohani, Yanay Rosen, Ankit Gupta, Xikun Zhang, Marcel Roed, Theo Alexandrov, Mohammed AlQuraishi, Patricia Brennan, Daniel B. Burkhardt, Andrea Califano, Jonah Cool, Abby F. Dernburg, Kirsty Ewing, Emily B. Fox, Matthias Haury, Amy E. Herr, Eric Horvitz, Patrick D. Hsu, Viren Jain, Gregory R. Johnson, Thomas Kalil, David R. Kelley, Shana O. Kelley, Anna Kreshuk, Tim Mitchison, Stephani Otte, Jay Shendure, Nicholas J. Sofroniew, Fabian Theis, Christina V. Theodoris, Srigokul Upadhyayula, Marc Valer, Bo Wang, Eric Xing, Serena Yeung-Levy, Marinka Zitnik, Theofanis Karaletsos, Aviv Regev, Emma Lundberg, Jure Leskovec, and Stephen R. Quake. How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell*, 187(25):7045–7063, 2024. doi: 10.1016/j.cell.2024.11.015.

Steven M Corsello, Joshua A Bittker, Zihan Liu, Joshua Gould, Patrick McCarren, Jodi E Hirschman, Stephen E Johnston, Anita Vrcic, Bang Wong, Mariya Khan, Jacob Asiedu, Rajiv Narayan, Christopher C Mader, Aravind Subramanian, and Todd R Golub. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nature Medicine*, 23(4):405–408, 2017. doi: 10.1038/nm.4306.

Kedar Dhamdhere, Mukund Sundararajan, and Qiqi Yan. How Important Is a Neuron? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=SylKoo0cKm.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P. Fulco, Livnat Jerby-Arnon, Nemanja D. Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M. Norman, Eric S. Lander, Jonathan S. Weissman, Nir Friedman, and Aviv Regev. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, 167(7):1853–1866.e17, 2016. doi: 10.1016/j.cell.2016.11.038.

Elena Favaro, Karim Bensaad, Mei G. Chong, Daniel A. Tennant, David J. P. Ferguson, Cameron Snell, Graham Steers, Helen Turley, Ji-Liang Li, Ulrich L. Günther, Francesca M. Buffa, Alan McIntyre, and Adrian L. Harris. Glucose utilization via glycogen phosphorylase sustains proliferation and prevents premature senescence in cancer cells. *Cell Metabolism*, 16(6):751–764, 2012. doi: 10.1016/j.cmet.2012.10.017.

Guadalupe Gonzalez, Xiang Lin, Isuru Herath, Kirill Veselkov, Michael Bronstein, and Marinka Zitnik. Combinatorial prediction of therapeutic perturbations using causally inspired neural networks. *Nature Biomedical Engineering*, 10(5):1008–1025, 2026. doi: 10.1038/s41551-025-01481-x. First published online September 9, 2025.

Xin-Ling He, Wen-Yu Lyu, Xin-Yuan Li, Hong Zhao, Lu Qi, and Jin-Jian Lu. Identification of glycogen phosphorylase L as a potential target for lung cancer. *Medical Oncology*, 40(7):211, 2023. doi: 10.1007/s12032-023-02069-8. URL https://doi.org/10.1007/s12032-023-02069-8.

Human Protein Atlas. The Cancer resource - methods summary. https://www.proteinatlas.org/humanproteome/cancer/method, n.d. Accessed September 25, 2026; the study's query uses pinned version 25.1 annotations.

Xingfa Huo, Helei Hou, Chuantao Zhang, Xueqin Duan, Hongwei Lan, Yufeng Li, Na Zhou, and Xiaochun Zhang. Chicoric acid targets PYGL to normalize glycogenolysis-driven glycolysis to suppress non-small cell lung cancer. *Cellular & Molecular Biology Letters*, 31:105, 2026. doi: 10.1186/s11658-026-00930-y.

Michael J. Keiser, Vincent Setola, John J. Irwin, Christian Laggner, Atheir I. Abbas, Sandra J. Hufeisen, Niels H. Jensen, Michael B. Kuijer, Roberto C. Matos, Thuy B. Tran, Ryan Whaley, Richard A. Glennon, Jérôme Hert, Kelan L. H. Thomas, Douglas D. Edwards, Brian K. Shoichet, and Bryan L. Roth. Predicting new molecular targets for known drugs. *Nature*, 462(7270):175–181, 2009. doi: 10.1038/nature08506.

Dominik Klein, Jonas Simon Fleck, Daniil Bobrovskiy, Lea Zimmermann, Sören Becker, Alessandro Palma, Leander Dony, Alejandro Tejada-Lapuerta, Guillaume Huguet, Hsiu-Chuan Lin, Nadezhda Azbukina, Fátima Sanchís-Calleja, Theo Uscidda, Artur Szalata, Manuel Gander, Aviv Regev, Barbara Treutlein, J. Gray Camp, and Fabian J. Theis. CellFlow enables generative single-cell phenotype modeling with flow matching. *bioRxiv*, 2025. doi: 10.1101/2025.04.11.648220. URL https://www.biorxiv.org/content/10.1101/2025.04.11.648220v1. Preprint, version 1.

Brent M Kuenzi, Jisoo Park, Samson H Fong, Kyle S Sanchez, John Lee, Jason F Kreisberg, Jianzhu Ma, and Trey Ideker. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell*, 38(5):672–684.e6, 2020. doi: 10.1016/j.ccell.2020.09.014. URL https://pubmed.ncbi.nlm.nih.gov/33096023/.

Jinping Liu, Dingyu Du, Yukai Huang, Jie Tian, Xuhui Wang, Longyi Chen, and Feng Wang. CCDC86-BHLHE40-ATF3 axis promotes aerobic glycolysis and tumor development in glioma. *Genes & Diseases*, 12(6):101643, 2025. doi: 10.1016/j.gendis.2025.101643. URL https://doi.org/10.1016/j.gendis.2025.101643.

Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, Leon Hetzel, Yuge Ji, Ignacio L Ibarra, Sanjay R Srivatsan, Mohsen Naghipourfar, Riza M Daza, Beth Martin, Jay Shendure, Jose L McFaline-Figueroa, Pierre Boyeau, F Alexander Wolf, Nafissa Yakubova, Stephan Günnemann, Cole Trapnell, David Lopez-Paz, and Fabian J Theis. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, 19(6):e11517, 2023. doi: 10.15252/msb.202211517. URL https://doi.org/10.15252/msb.202211517.

Jianzhu Ma, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan, and Trey Ideker. Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods*, 15(4):290–298, 2018. doi: 10.1038/nmeth.4627. URL https://pubmed.ncbi.nlm.nih.gov/29505029/.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018. URL https://arxiv.org/abs/1802.03426v3. Version 3, revised 2020.

Sophia Müller-Dott, Eirini Tsirvouli, Miguel Vazquez, Ricardo O Ramirez Flores, Pau Badia-I-Mompel, Robin Fallegger, Dénes Türei, Astrid Lægreid, and Julio Saez-Rodriguez. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Research*, 51(20):10934–10949, 2023. doi: 10.1093/nar/gkad841. URL https://doi.org/10.1093/nar/gkad841.

Joseph M. Replogle, Reuben A. Saunders, Angela N. Pogson, Jeffrey A. Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J. Wagner, Karen Adelman, Gila Lithwick-Yanai, Nika Iremadze, Florian Oberstrass, Doron Lipson, Jessica L. Bonnar, Marco Jost, Thomas M. Norman, and Jonathan S. Weissman. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14):2559–2575.e28, 2022. doi: 10.1016/j.cell.2022.05.013.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nature Biotechnology*, 42(6):927–935, 2024. doi: 10.1038/s41587-023-01905-6. First published online August 17, 2023.

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web: ESWC 2018*, pp. 593–607. Springer, 2018. doi: 10.1007/978-3-319-93417-4_38. URL https://doi.org/10.1007/978-3-319-93417-4_38.

Thomas Schnake, Oliver Eberle, Jonas Lederer, Shinichi Nakajima, Kristof T. Schütt, Klaus-Robert Müller, and Grégoire Montavon. Higher-Order Explanations of Graph Neural Networks via Relevant Walks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7581–7596, 2022. doi: 10.1109/TPAMI.2021.3115452. URL https://ieeexplore.ieee.org/document/9547794/.

Konstantinos Stamatiou, Aldona Chmielewska, Shinya Ohta, William C. Earnshaw, and Paola Vagnarelli. CCDC86 is a novel Ki-67-interacting protein important for cell division. *Journal of Cell Science*, 136(2):jcs260391, 2023. doi: 10.1242/jcs.260391. URL https://doi.org/10.1242/jcs.260391.

Jin Su, Yan He, Shiyang You, Shiyu Jiang, Xibin Zhou, Xuting Zhang, Yuxuan Wang, Xining Su, Igor Tolstoy, Xing Chang, Hongyuan Lu, and Fajie Yuan. A trimodal protein language model enables advanced protein searches. *Nature Biotechnology*, 44(8):1310–1316, 2026. doi: 10.1038/s41587-025-02836-0. First published online October 2, 2025.

Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, David L Lahr, Jodi E Hirschman, Zihan Liu, Melanie Donahue, Bina Julian, Mariya Khan, David Wadden, Ian C Smith, Daniel Lam, Arthur Liberzon, Courtney Toder, Mukta Bagul, Marek Orzechowski, Oana M Enache, Federica Piccioni, Sarah A Johnson, Nicholas J Lyons, Alice H Berger, Alykhan F Shamji, Angela N Brooks, Anita Vrcic, Corey Flynn, Jacqueline Rosains, David Y Takeda, Roger Hu, Desiree Davison, Justin Lamb, Kristin Ardlie, Larson Hogstrom, Peyton Greenside, Nathanael S Gray, Paul A Clemons, Serena Silver, Xiaoyun Wu, Wen-Ning Zhao, Willis Read-Button, Xiaohua Wu, Stephen J Haggarty, Lucienne V Ronco, Jesse S Boehm, Stuart L Schreiber, John G Doench, Joshua A Bittker, David E Root, Bang Wong, and Todd R Golub. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6):1437–1452.e17, 2017. doi: 10.1016/j.cell.2017.10.049. URL https://doi.org/10.1016/j.cell.2017.10.049.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 3319–3328. PMLR, 2017. URL https://proceedings.mlr.press/v70/sundararajan17a.html.

Damian Szklarczyk, Katerina Nastou, Mikaela Koutrouli, Rebecca Kirsch, Farrokh Mehryary, Radja Hachilif, Dewei Hu, Matteo E Peluso, Qingyao Huang, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, Peer Bork, Lars J Jensen, and Christian von Mering. The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Research*, 53(D1):D730–D737, 2025. doi: 10.1093/nar/gkae1113. URL https://doi.org/10.1093/nar/gkae1113.

Mathias Uhlen, Cheng Zhang, Sunjae Lee, Evelina Sjöstedt, Linn Fagerberg, Gholamreza Bidkhori, Rui Benfeitas, Muhammad Arif, Zhengtao Liu, Fredrik Edfors, Kemal Sanli, Kalle von Feilitzen, Per Oksvold, Emma Lundberg, Sophia Hober, Peter Nilsson, Johanna Mattsson, Jochen M Schwenk, Hans Brunnström, Bengt Glimelius, Tobias Sjöblom, Per-Henrik Edqvist, Dijana Djureinovic, Patrick Micke, Cecilia Lindskog, Adil Mardinoglu, and Fredrik Ponten. A pathology atlas of the human cancer transcriptome. *Science*, 357(6352):eaan2507, 2017. doi: 10.1126/science.aan2507. URL https://doi.org/10.1126/science.aan2507.

Virginia Vassilevska, R. Ryan Williams, and Raphael Yuster. All Pairs Bottleneck Paths and Max-Min Matrix Products in Truly Subcubic Time. *Theory of Computing*, 5(9):173–189, 2009. doi: 10.4086/toc.2009.v005a009. URL https://toc.cs.uchicago.edu/articles/v005a009/.

Menghua Wu, Umesh Padia, Sean H. Murphy, Regina Barzilay, and Tommi Jaakkola. Identifying biological perturbation targets through causal differential networks. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 67537–67561. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/wu25v.html.

Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GNNExplainer: Generating Explanations for Graph Neural Networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html.

Meng Yuan, Cheng Zhang, Kalle Von Feilitzen, Martin Zwahlen, Mengnan Shi, Xiangyu Li, Hong Yang, Xiya Song, Hasan Turkez, Mathias Uhlén, and Adil Mardinoglu. The Human Pathology Atlas for deciphering the prognostic features of human cancers. *eBioMedicine*, 111:105495, 2025. doi: 10.1016/j.ebiom.2024.105495. URL https://doi.org/10.1016/j.ebiom.2024.105495. Published online December 10, 2024.

A DATA, GRAPH, AND VARIABLE DEFINITIONS

A.1 ASSAY AND RETENTION

The pinned inputs are CLUE's LINCS2020 beta Level-3 CRISPR-treated and control GCTX matrices, sample metadata, and gene annotations. L1000 measures expression abundance rather than transcription rates. The selected `trt_xpr` modality denotes CRISPR/Cas9 knockout, not CRISPR interference. Level-3 values are processed expression, not raw counts or Level-5 differential signatures. All retained treated wells are measured at 96 hours; dose is missing and is not interpreted as zero.

Retention requires physical replicate X1, X2, or X3, a nonmissing target label, and matched controls on the same plate. The target vocabulary is the intersection of labels present in X1/X2 and X3 before hybrid splitting. This produces 270,331 paired wells on 915 plates, with 21,687 matched controls. Of 5,119 retained target identities, 5,108 map to eligible positive labels, yielding 269,769 supervised pairs. Protein-covered candidate nodes form a broader set; a never-positive candidate receives competition-based negative supervision but is not an established biological negative or a validated zero-shot prediction.

LINCS supplies expression estimates inferred from the 978 measured landmarks; we use these supplied values without fitting a new inference model. The BING (Best Inferred Genes) space combines landmarks with genes whose inferred expression passed the provider's validation against RNA-seq (Subramanian et al., 2017). In the pinned LINCS annotation, 10,174 genes belong to BING and 10,086 of these are annotated protein-coding. The historical identifier intersection retains 10,073 coordinates: all 978 landmarks and 9,095 inferred genes. This ordered feature table is preserved as a contract; no GTEx expression enters this study. Extending expression coverage lets more gene nodes combine cellular context with protein features and typed relationships. Inferred coordinates remain dependent estimates, and we make no claim of exact variance preservation. Native control values are averaged with float64 accumulation and stored in float32 before normalization.

A.2 TRAINING-ONLY COORDINATES AND SPLIT SEMANTICS

For rotation f, eligible inner-training treated wells T_f determine the per-gene moments for both input channels:

$$\mu_{fj} = \frac{1}{|T_f|} \sum_{n \in T_f} v_{nj}, \qquad \sigma^2_{fj} = \frac{1}{|T_f|} \sum_{n \in T_f} (v_{nj} - \mu_{fj})^2, \qquad x^{(c)}_j = \frac{v^{(c)}_j - \mu_{fj}}{s_{fj}}. \tag{12}$$

Here $s_{fj} = 1$ if $\sigma_{fj} < 10^{-6}$ and $s_{fj} = \sigma_{fj}$ otherwise. Moments use population variance, float64 accumulation, and float32 storage. Controls, validation, test, and constructed requests never fit these moments. There is no extra log transform, clipping, rank transform, or per-profile centering.

H1/H2/H3 are hybrid split rotations, not X1/X2/X3 physical replicates or independent random seeds. Multi-line tissues hold out an entire cell line, including every plate and replicate. Singleton tissues hold out one complete physical replicate while other replicates of that line remain in development. Within each development line, a seeded 10% of whole plates is assigned to validation, rounding up while retaining a training plate. Validation selects the minimum unmasked CE and does not include the outer test.

H1 and H3 each contain 30,921 A549 test rows and no A549 training or validation rows. H2 includes A549 in development and is not a primary consensus member. Repeated holdouts of the same line are not independent biological units. Neither evaluation setting holds out target genes.

A.3 BASELINE QUALITY

The query baseline is the native matched-control mean for the lexicographically first retained A549 sample whose plate's matched controls all carry `qc_pass=1` in the CLUE metadata: `XPR008_A549.311_96H_X3_B35`, with 20 controls. Failed and unlabeled controls both disqualify a plate, and the rule uses QC metadata alone, so it is fixed before any model is queried. It is a deterministic selection, not an estimate of a typical cell.

Training retention records but does not filter technical QC. Raw metadata retain separate failed, passing, and missing labels. CLUE cautions against assuming that all downloaded L1000 columns pass QC (Broad Institute, 2018b); its CRISPR documentation also distinguishes technical quality from functional perturbation checks (Broad Institute, 2018a). The baseline's conductance is an explanation of that model input, not evidence that it represents a typical A549 state. Appendix H reports robustness across all QC-passing A549 plates.

A.4 Graph construction and protein inputs

The ordered expression features and resolved retained perturbation targets span 11,791 distinct genes. The unresolved historical label RIEG2 is also retained for audit, giving 11,792 seed nodes. Adding 5,516 one-hop neighbors across STRING v12.5 and CollecTRI yields 17,308 nodes; we then retain the selected relations within that universe. Resolved nodes use canonical HGNC symbols; all nodes follow lexicographic symbol order. STRING pairs retain their maximum confidence after mapping; a threshold of 700 removes lower-confidence pairs, and STRING self-pairs are excluded. The evidence-filtered CollecTRI subset requires a PMID-decided sign and at least two distinct underlying resources after normalizing resource suffixes and excluding the umbrella label. Source signed self-edges are retained.

Five biological families become eight transport channels: STRING functional and physical expand bidirectionally within their channels; STRING regulatory, CollecTRI activation, and CollecTRI repression each have separate forward and reverse channels. Their learned transformations are not constrained to positive or negative latent effects. STRING confidence determines retention, not the runtime message coefficient. For sender j and receiver i,

$$\widetilde{A}^r_{ij} = \frac{\mathbf{1}[(j \to i) \in E_r]}{\sqrt{\max(1, d^r_{\text{in},i}) \max(1, d^r_{\text{out},j})}}. \tag{13}$$

All relation channels read the same preceding state and are summed before SiLU. No additional self-loops are introduced; the residual provides a direct carried-state path.

Protein vectors are reconstructed from the pinned published SwissProt ProTrek-650M FAISS index; setup does not run a new sequence encoder. HGNC UniProt mappings, with STRING accession fallback, provide vectors for 17,225 nodes. Candidate eligibility follows this availability mask, rather than an HGNC protein-coding filter. The 83 nodes without mapped vectors remain computational context but are excluded from ranking; five are annotated protein-coding in the pinned HGNC table. The shared projector maps 1,024 dimensions through 256 to 128 with SiLU between affine maps. Baseline and endpoint encoders are separate scalar-to-128 affine maps. Missing protein or expression suppresses that component, including expression-encoder biases when the input is unavailable. Of the 10,073 expression-bearing nodes, 10,071 also have protein vectors.

The official September 4, 2026 HGNC archive replaced an unavailable August 28 source during acquisition. The recorded reconstruction reproduced seven ordered graph tables and both protein arrays byte-for-byte; source metadata retain the two source identities and the equivalence record. This is acquisition provenance, not a change of graph or scientific interpretation.

B Training and predictive characterization

The shared decoder is affine $128 \to 64$, SiLU, affine $64 \to 1$. Its eligible outputs compete in one global softmax. For correct target y and the highest-scoring $K_{\text{eff}} = \min(20, |\mathcal{V}_p| - 1)$ incorrect candidates $\mathcal{H}_K$, training uses

$$\mathcal{L}(q, y) = -s_y + \log \sum_{i \in \mathcal{V}_p} e^{s_i} + \frac{\lambda}{K_{\text{eff}}} \sum_{j \in \mathcal{H}_K} [m + s_j - s_y]_+, \qquad \lambda = 0.1, \quad m = 1. \tag{14}$$

Only the auxiliary margin is restricted to hard negatives. Cross-entropy compares all candidates for every pair. The input graph and protein vectors are frozen; the projector, expression encoders, relation transforms, LayerNorm parameters, and decoder are learned. Expression encoders initialize at zero but remain trainable. A Bernoulli keep mask of probability 0.85 hides each available gene's

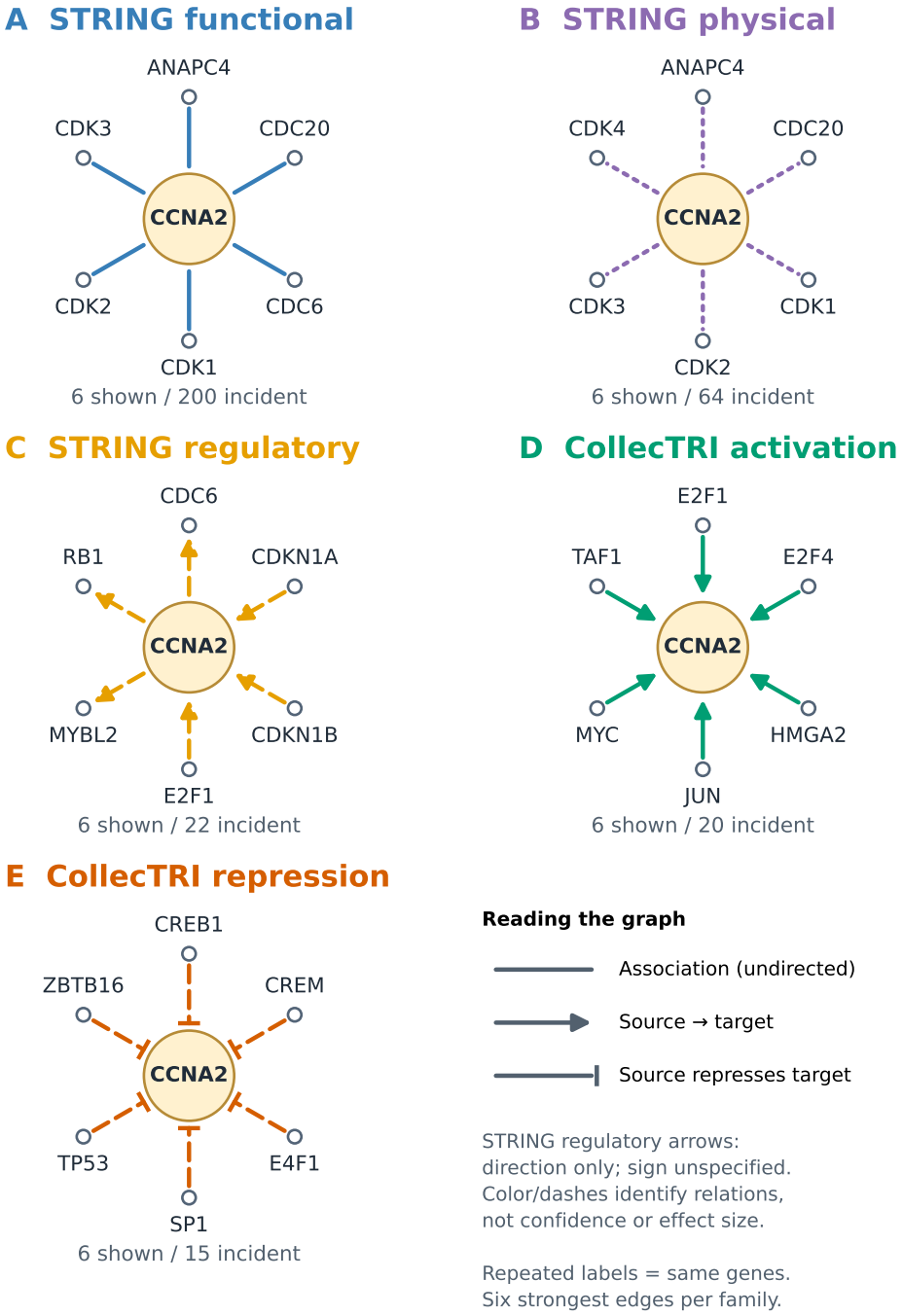

Figure 5: **The relation vocabulary before learning.** The focal gene is the most frequent retained target with incident edges in all five families, with alphabetical tie-breaking. Each panel shows six strongest incident edges by source confidence or reference count, then source/target names. Selection uses no model score. The asset reports omitted incident edges and excludes neighbor–neighbor edges. Repeated labels are the same genes; prior direction/sign are not attribution signs.

Table 1: Notation shared by the model and attribution. Node states are indexed from 0 through L; transition ℓ connects states ℓ and $\ell + 1$.

Symbol	Meaning
$N, F, \lvert\mathcal{V}_p\rvert, d$	17,308 graph nodes, 10,073 inputs, 17,225 protein-covered candidates, hidden width 128.
$q = (x^{(0)}, x^{(1)})$	Ordered baseline/endpoint pair, shape $[2, F]$.
m_i^p, m_i^x, a_i	Protein/expression availability and paired training keep mask.
$p_i, r_i(q), \bar{p}$	Projected protein, fixed expression encoding, and projected protein-population mean.
$h_i^{(\ell)}, H^{(\ell)}$	A node vector in $\mathbb{R}^d$ and the all-node matrix in $\mathbb{R}^{N\times d}$.
$m_e^{(\ell)}, M_i^{(\ell)}, u_i^{(\ell)}$	Individual typed message, receiver message sum, and pre-LayerNorm carried-state-plus-message sum.
$s_k, \Delta s_k^{\text{protein}}$	Nominated raw logit and its actual-minus-reference protein-path difference.
$C_i^{(\ell)}, E_e^{(\ell)}, R_i^{(\ell)}$	Signed node, message, and residual conductance, in nominated-logit units.

Table 2: Predictive context for the interpreted models. Equal-tissue means keep five unseen-line tissues and six held-replicate tissues separate. Differences are descriptive: these overlapping rotations are not independent seeds, and the study does not isolate depth from capacity or nonlinear computation.

Depth	Unseen MRR	CE	Replicate MRR	CE
1	0.1463	7.3530	0.1213	7.4106
2	0.1414	7.4310	0.1296	7.1058
3	0.1408	7.4816	0.1325	6.9100
4	0.1456	7.5455	0.1352	6.7383

two expression channels together, without inverted-dropout rescaling. Evaluation uses no random mask.

AdamW uses learning rate 1.00×10^{-3}, weight decay 1.00×10^{-4}, and gradient-norm clipping at 5, without a scheduler or mixed precision in the declared path. The budget is 30 epochs with patience 5 and minimum CE improvement zero; training and validation use seed 20260903. Four CUDA ranks each process a nominal 16 pairs. Balanced, disjoint, unpadded shards include every training row; local loss sums are scaled by world size over global example count before DDP gradient averaging, including uneven final shards. Validation CE selects the checkpoint, after which full outer tests are evaluated.

The twelve fits share the declared graph, ordered features, candidate policy, and non-depth training recipe; within each rotation their split and normalizer identities match. Increasing depth improves the held-replicate macro metrics but does not uniformly improve transfer to unseen lines. These observations characterize the trained objects rather than establish a universally preferred depth or competitive advantage over another architecture.

For target rank r_a and target softmax probability $p_a(y_a)$, MRR is $n^{-1}\sum_a r_a^{-1}$ and CE is $-n^{-1}\sum_a \log p_a(y_a)$. A uniformly random ordering of the $|\mathcal{V}_p| = 17{,}225$ candidates has expected MRR $|\mathcal{V}_p|^{-1}\sum_{r=1}^{|\mathcal{V}_p|} r^{-1} \approx 0.00060$. The held-out rank convention gives ties their average occupied rank: $r_a = 1 + g_a + (e_a - 1)/2$, with g_a strictly greater scores and e_a equal scores including the target. Hits uses the recorded `torch.topk` convention. The nomination table separately uses minimum tied rank; a fixed node order resolves displayed ties.

Group means are pooled by example count within a fold/evaluation unit. Repeated holdouts of a unit are averaged, then distinct units receive equal weight within tissue and tissues receive equal weight

Table 3: Saved checkpoint selection and eligible populations. Recorded epochs include training after the selected checkpoint. Counts repeat across depths because split identities match.

Depth	Fit	Selected	Recorded	Train	Validation	Test
1	H1	30	30	111,512	14,053	144,204
1	H2	30	30	167,325	21,291	81,153
1	H3	23	28	111,987	13,921	143,861
2	H1	17	22	111,512	14,053	144,204
2	H2	29	30	167,325	21,291	81,153
2	H3	23	28	111,987	13,921	143,861
3	H1	16	21	111,512	14,053	144,204
3	H2	21	26	167,325	21,291	81,153
3	H3	19	24	111,987	13,921	143,861
4	H1	11	16	111,512	14,053	144,204
4	H2	20	25	167,325	21,291	81,153
4	H3	17	22	111,987	13,921	143,861

Table 4: Complete macro metrics from the same saved test groups.

Depth	Setting	MRR	CE	Hits@1	Hits@10
1	Held replicate	0.1213	7.4106	0.1027	0.1485
1	Unseen line	0.1463	7.3530	0.1218	0.1848
2	Held replicate	0.1296	7.1058	0.1116	0.1566
2	Unseen line	0.1414	7.4310	0.1215	0.1711
3	Held replicate	0.1325	6.9100	0.1125	0.1602
3	Unseen line	0.1408	7.4816	0.1187	0.1745
4	Held replicate	0.1352	6.7383	0.1156	0.1639
4	Unseen line	0.1456	7.5455	0.1257	0.1751

within a setting:

$$m_{fu} = \frac{\sum_{g\in(f,u)} n_g m_g}{\sum_{g\in(f,u)} n_g}, \quad m_u = \frac{1}{|F_u|}\sum_f m_{fu}, \quad m_t = \frac{1}{|U_t|}\sum_u m_u, \quad m_{\text{setting}} = \frac{1}{|T|}\sum_t m_t. \tag{15}$$

For unseen-line tests, physical replicates are pooled within a line/fold before repeated line holdouts are averaged. For held-replicate tests, the unit is a line/physical-replicate pair. This weights tissues deliberately rather than estimating the average recorded well. Repeated rotations do not support independence-based confidence intervals. Group medians cannot recover a pooled median, and MRR cannot recover mean rank.

C THE REQUEST AND ITS TWO GEOMETRIES

C.1 PROGNOSTIC EVIDENCE AND COORDINATES

HPA separates high- and low-expression patient groups at an expression cutoff and assesses survival association. We use the validated tier in the pinned 25.1 table, which identifies prognostic associations replicated across cohorts (Yuan et al., 2025; Human Protein Atlas, n.d.). Its entries are P values, not expression amplitudes or causal effect sizes. We use only validated favorable/unfavorable entries for *Lung Adenocarcinoma (validation)*, deduplicate by Ensembl identity, and exclude conflicting directions or ambiguous mappings. Graph-only markers cannot be edited expression inputs. The resulting primary request has 157 mapped coordinates, not the whole raw HPA marker list.

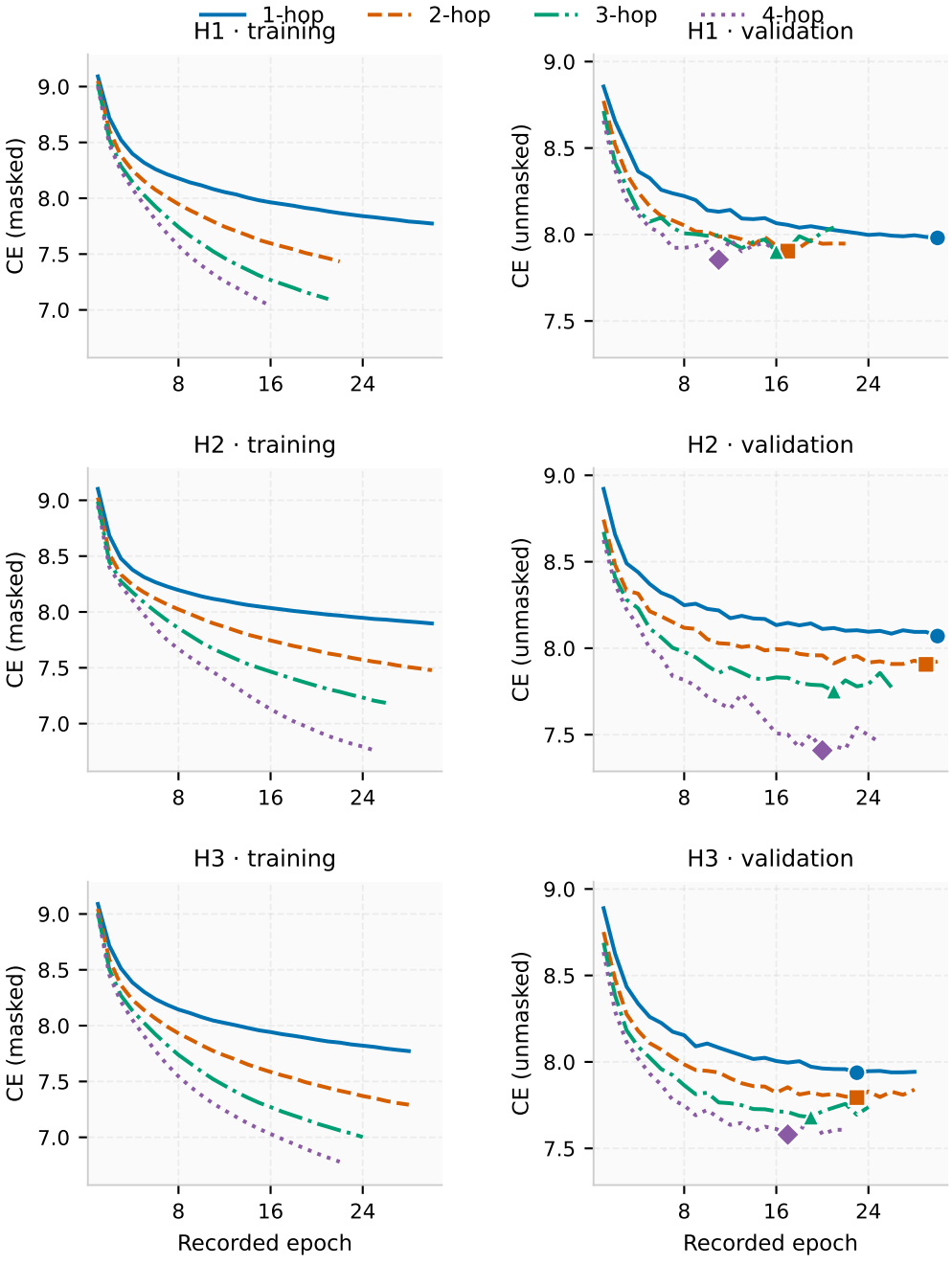

Figure 6: **Recorded training and validation histories.** Rows denote rotations and columns distinguish masked online training CE from unmasked epoch-end validation CE. Curves stop at the last recorded epoch; markers identify minimum-validation-CE checkpoints. The two CE measurements differ in masking and weight timing, so their gap is not a pure generalization gap.

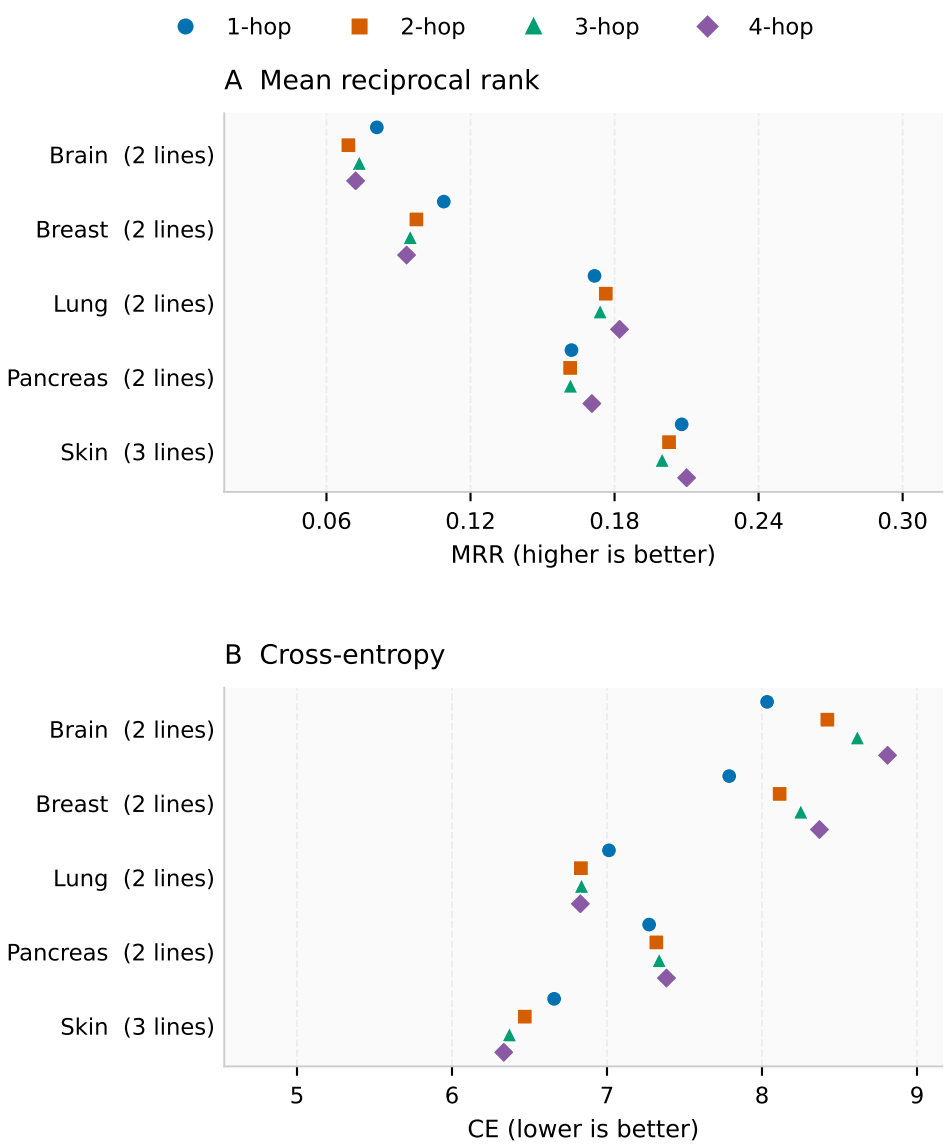

Figure 7: **Held-out-cell-line variation.** Repeated holdouts are averaged before distinct lines receive equal weight within tissue. Depth markers are offset for legibility, not to represent uncertainty. Values are unchanged saved evaluations.

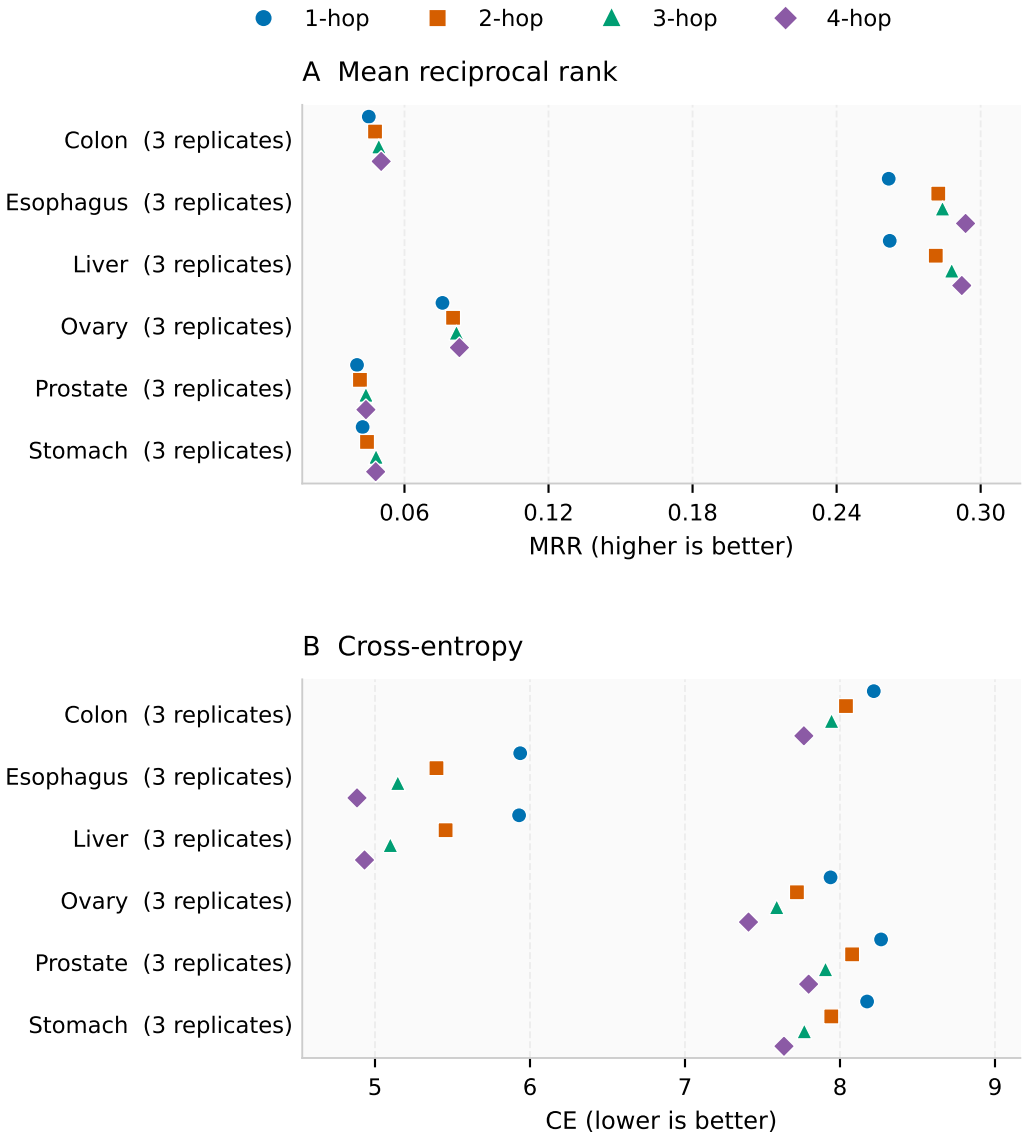

Figure 8: **Held-out-replicate variation.** Each line/physical-replicate unit receives equal weight within its singleton tissue. This describes another replicate of a known line, not unseen-line transfer.

Let c^{nat} be the exact matched native control mean and (μ^R, s^R) the H1 reference moments. The request is

$$c^R = (c^{\mathrm{nat}} - \mu^R) \oslash s^R, \qquad t^R = c^R + a\mathbf{1}_{S_+} - a\mathbf{1}_{S_-}, \qquad a = 2, \tag{16}$$

$$q^{\mathrm{nat}} = (q^R \odot s^R) + \mu^R, \qquad q_f = (q^{\mathrm{nat}} - \mu^f) \oslash s^f, \qquad f \in \{\mathrm{H1}, \mathrm{H3}\}. \tag{17}$$

The reference pair is formed in float32, then transferred through native units in float64. This retains the declared physical request and its rounding. All other coordinates remain unchanged; there is no clipping or profile-wide renormalization. H1 defines the edit units, each member's normalizer defines scoring coordinates, and the projected protein mean defines the separate attribution reference. Both primary nominees' own expression coordinates decrease by two reference SDs.

The primary fits are sibling rotations of the specified four-hop run, not a search over the newest or best available checkpoint. Model membership is selected from saved train/validation/test exposure. The explorer labels contexts without whole-line-held-out models as seen-cell exploration rather than presenting them as held out. A missing HPA cohort or evidence tier is not silently replaced by another policy.

C.2 Observed response geometry

One observed-displacement point is a retained knockout well. It uses the complete vector $(v_n^{\mathrm{endpoint}} - \bar{v}^{\mathrm{plate(n)}}) \oslash s^{\mathrm{H1}}$, retaining replicate and partition membership without unit normalization. The exploratory UMAP uses all retained wells and is not part of predictor fitting or checkpoint selection. A second view uses the same coordinates, colored by the within-tissue rank of full-feature cosine alignment with favorable-up/unfavorable-down directions. That descriptive overview uses explicitly matched TCGA cohorts and potential-plus-validated markers, separately from the primary validation-only A549 request. Unmatched cohorts or undefined cosines are unscored, not imputed.

C.3 Learned gene-state geometry

The representation view instead contains one row per gene at each stage of the same frozen H1 model and request. One CPU UMAP fit uses all 86,540 unscaled node-stage vectors, Euclidean distance, 30 neighbors, minimum distance 0.15, 500 epochs, random initialization, seed 20260903, and one job. Neither logits nor ranks enter the fit. No stage-specific centering or normalization is performed.

D Conductance calculations and numerical evidence

D.1 Initial features and complete computational layers

At each integration point, the same learned updates in Equation 3 act on the interpolated initial state. The projected protein endpoints and mean are detached, fixed values; all trained parameters and graph coefficients are unchanged. For a covered node, its initial tangent is $p_i - \bar{p}$; for a missing-protein node it is zero. Initial feature attribution is

$$I_{i,a} = (p_{i,a} - \bar{p}_a) \int_0^1 \left. \frac{\partial s_k}{\partial h_{i,a}^{(0)}} \right|_\alpha d\alpha, \qquad C_i^{(0)} = \sum_a I_{i,a}, \tag{18}$$

with $I_{i,a} = 0$ where protein is missing. A node can have zero own initial protein attribution yet carry nonzero later conductance through incoming messages.

The node-local decoder makes the nominee's final state the only directly read final node. Its conductance therefore equals the full difference at that layer, up to numerical integration error. Earlier layers can distribute the same difference across many nodes. This architectural endpoint is explicitly separated in the profile heatmaps; it is not evidence that earlier computation lacked network dependence.

D.2 Full forward path, reverse sensitivities, and forward tangents

For each quadrature position, `stages_from_initial` evaluates the actual full graph, recording node states, summed messages, and pre-LayerNorm states. One backward call obtains the nominated

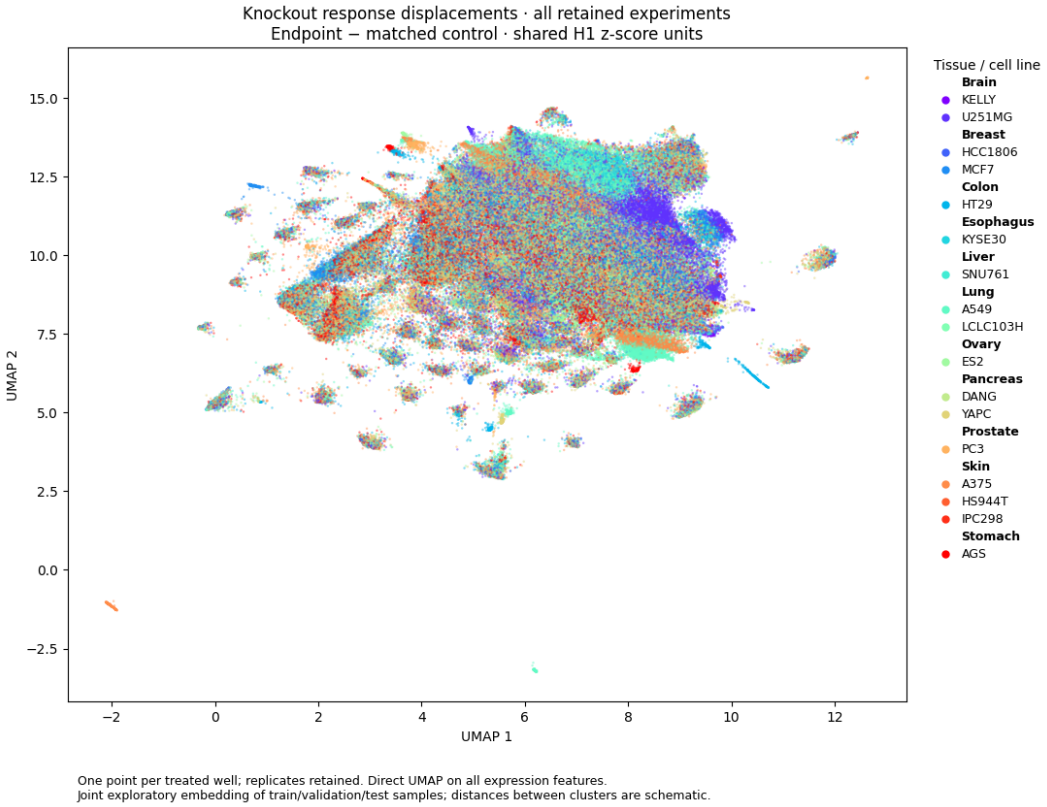

Figure 9: **Observed knockout displacements.** One point represents one retained well in shared H1-scale displacement coordinates. Colors identify cell lines, grouped by tissue. This is a joint exploratory projection of training, validation, and test observations, not a held-out performance estimate or a measured biological trajectory.

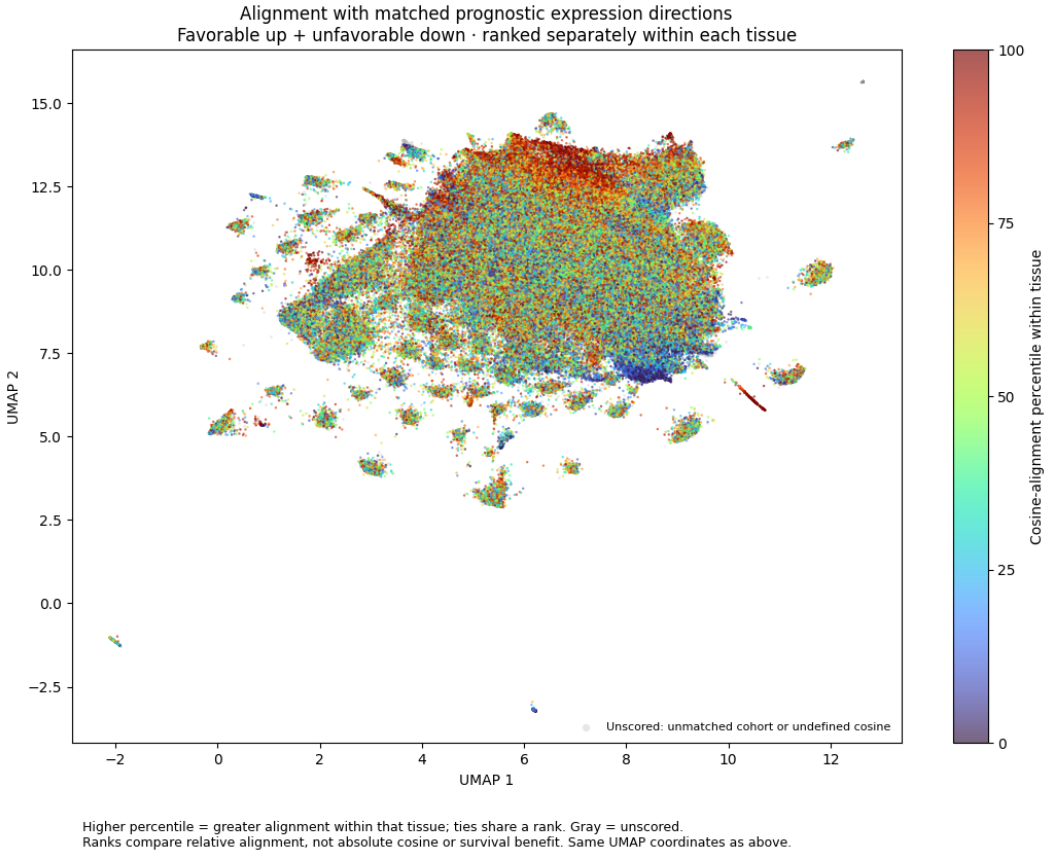

Figure 10: **The same observations, recolored by prognostic alignment.** Percentiles rank full-feature cosine alignment within each tissue; they are not absolute cosine, survival benefit, or a cross-tissue effect-size scale. Gray marks unscored observations. The embedding is unchanged from Figure 9, and this descriptive marker policy is distinct from the primary query.

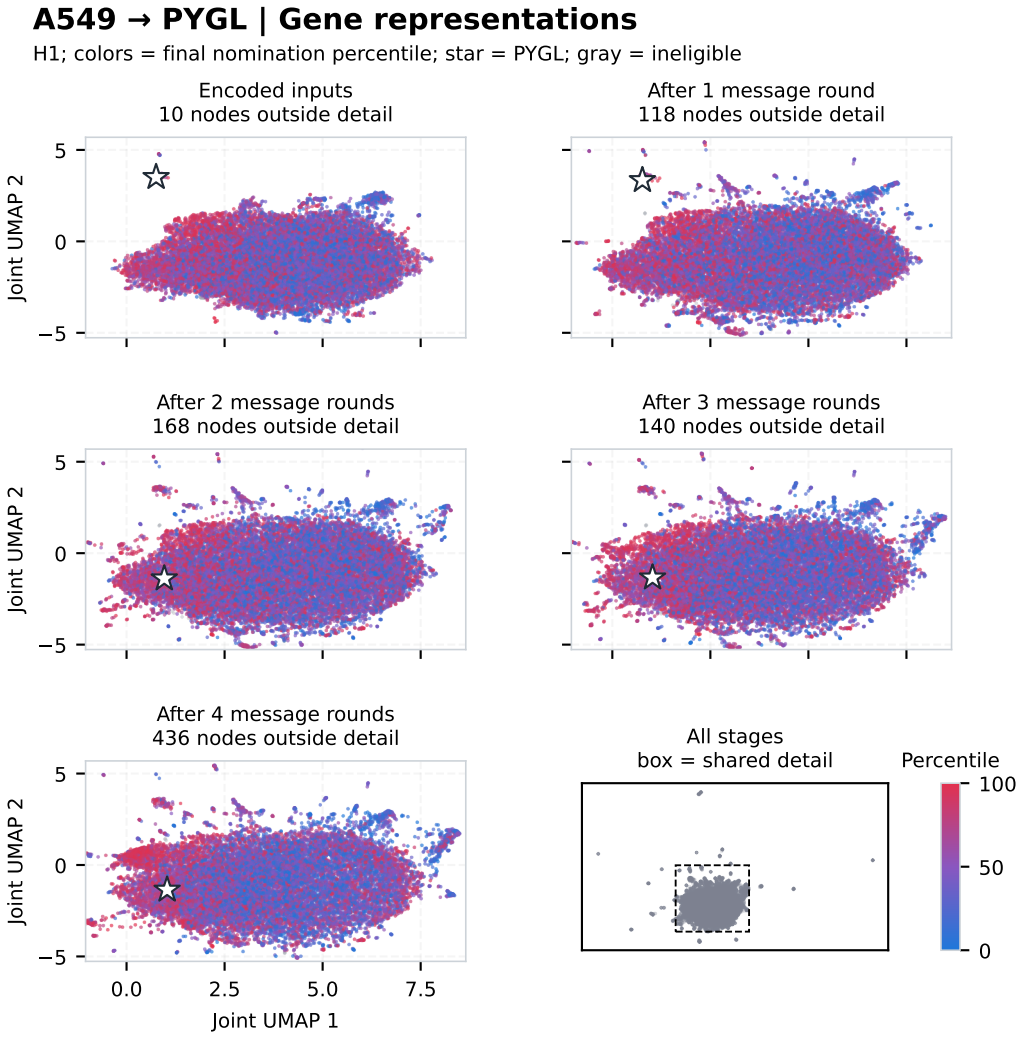

Figure 11: **Following one request through gene representations.** Encoded inputs and four successive message states belong to the same trained H1 model. Each point remains one gene. The white star marks PYGL; colors hold the H1 final nomination percentile fixed across panels, and gray marks ineligible candidates. The detail window and axes are shared, with omitted-point counts recorded per panel; the final overview supplies full-extent context. The projection describes representation geometry, not attribution or biological time.

Table 5: Saved numerical evidence for each primary case and model. Completeness checks compare full signed cuts to the endpoint difference; reconciliation checks both ends of each transition. These are recorded computations, not biological-validation measurements.

Nominee	Fit	Points	Completeness error	Reconciliation error
PYGL	H1	32	9.77e-06	1.78e-15
PYGL	H3	64	3.20e-08	7.11e-15
CCDC86	H1	32	1.23e-06	8.88e-15
CCDC86	H3	32	3.13e-06	8.88e-15

score's sensitivities to those tensors. Tangents propagate forward through the same transforms. For edge $e = (j \to i, r)$,

$$\dot{m}_e^{(\ell)} = \widetilde{A}^r_{ij} W_{\ell,r} \dot{h}_j^{(\ell)}, \qquad \dot{u}_i^{(\ell)} = \dot{h}_i^{(\ell)} + \mathrm{SiLU}'(M_i^{(\ell)}) \odot \dot{M}_i^{(\ell)}, \qquad \dot{h}_i^{(\ell+1)} = D\,\mathrm{LN}_\ell(u_i^{(\ell)})[\dot{u}_i^{(\ell)}]. \tag{19}$$

A dot denotes differentiation with respect to α. The LayerNorm directional derivative includes centering, variance, epsilon, and learned scale. It is not an interpolation between independently chosen hidden-layer endpoints.

Message conductance contracts each transformed sender tangent with the receiver's message-sum sensitivity and fixed normalized coefficient. Contractions are chunked without materializing the full edge-by-hidden Jacobian. The gradient includes SiLU, LayerNorm, later messages, and the decoder; residual sensitivities bypass SiLU but include downstream normalization and computation. Receptive-field bookkeeping retains every layer-edge instance that can reach the nominee, including residual reachability. Only structurally unrelated contractions are omitted; the forward path remains the full graph.

The implementation compares analytic tangents and edge contractions against explicit autograd calculations on small test models. It separately verifies that the actual-protein endpoint reproduces ordinary model scoring. These checks concern the chosen mathematical instrument, not the biological validity of its reference.

D.3 Quadrature and reconciliation

Gauss–Legendre points and weights are transformed from $[-1, 1]$ to $[0, 1]$. All feature, node, message, and residual integrals share the same positions and weights. We compare 16 versus 32 points, continuing to 64 or 128 if required. For each attribution array or specified layer slice V,

$$\|\widehat{V}_{2Q} - \widehat{V}_Q\|_1 \le 10^{-5} + 10^{-3}\|\widehat{V}_{2Q}\|_1. \tag{20}$$

Each complete signed cut must also recover the independently evaluated score difference within the absolute-plus-relative tolerance based on its own absolute mass. This avoids judging numerical accuracy solely by a cancellation-sensitive total. Node/residual slices and message slices are checked separately per layer. Endpoint parity uses absolute and relative tolerances of 10^{-10}.

In addition to receiving-node reconciliation, we check the allocation leaving each sender:

$$C_j^{(\ell)} = \sum_{e=(j\to i,r)} E_e^{(\ell)} + R_j^{(\ell)}. \tag{21}$$

Maximum incoming/outgoing discrepancies must be at most 10^{-5}. Failure to satisfy the criteria by 128 points stops the calculation. The saved accepted orders and errors for every case/member are reported below.

E Consensus, route selection, and display semantics

The two models receive byte-identical native expression pairs but retain their own normalizers, projectors, means, and decoders. Candidate identities, graph relations, and layer indices align. We

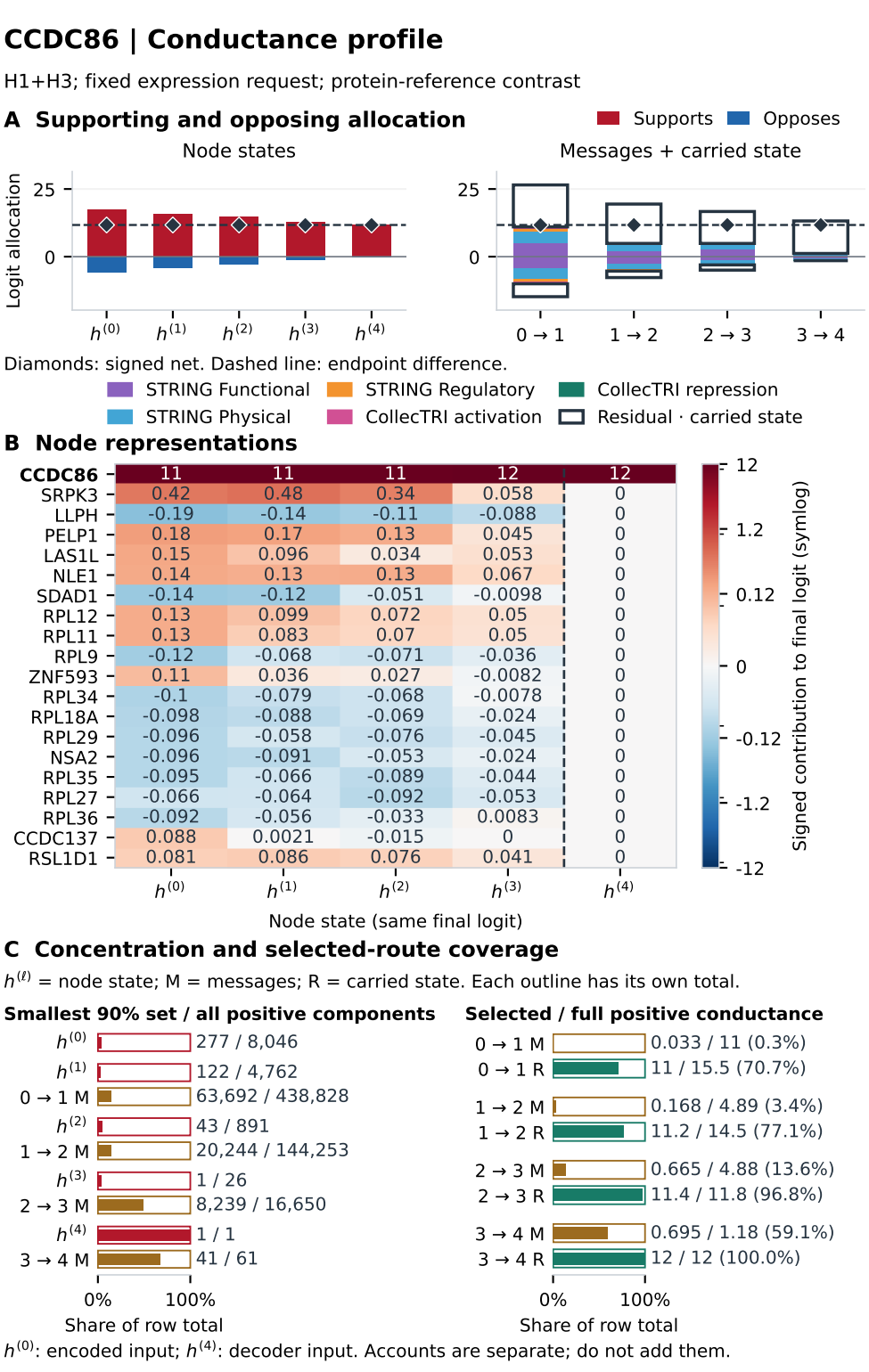

Figure 12: **CCDC86: a different intermediate allocation.** The quantities match Figure 3, accounting for a +11.7054-logit protein-reference difference. Heatmap colors share a within-case scale, not a common cross-case scale. Final-node concentration is architectural; message and residual coverage have separate full-positive denominators. The complete supporting network is in Appendix Figure 13.

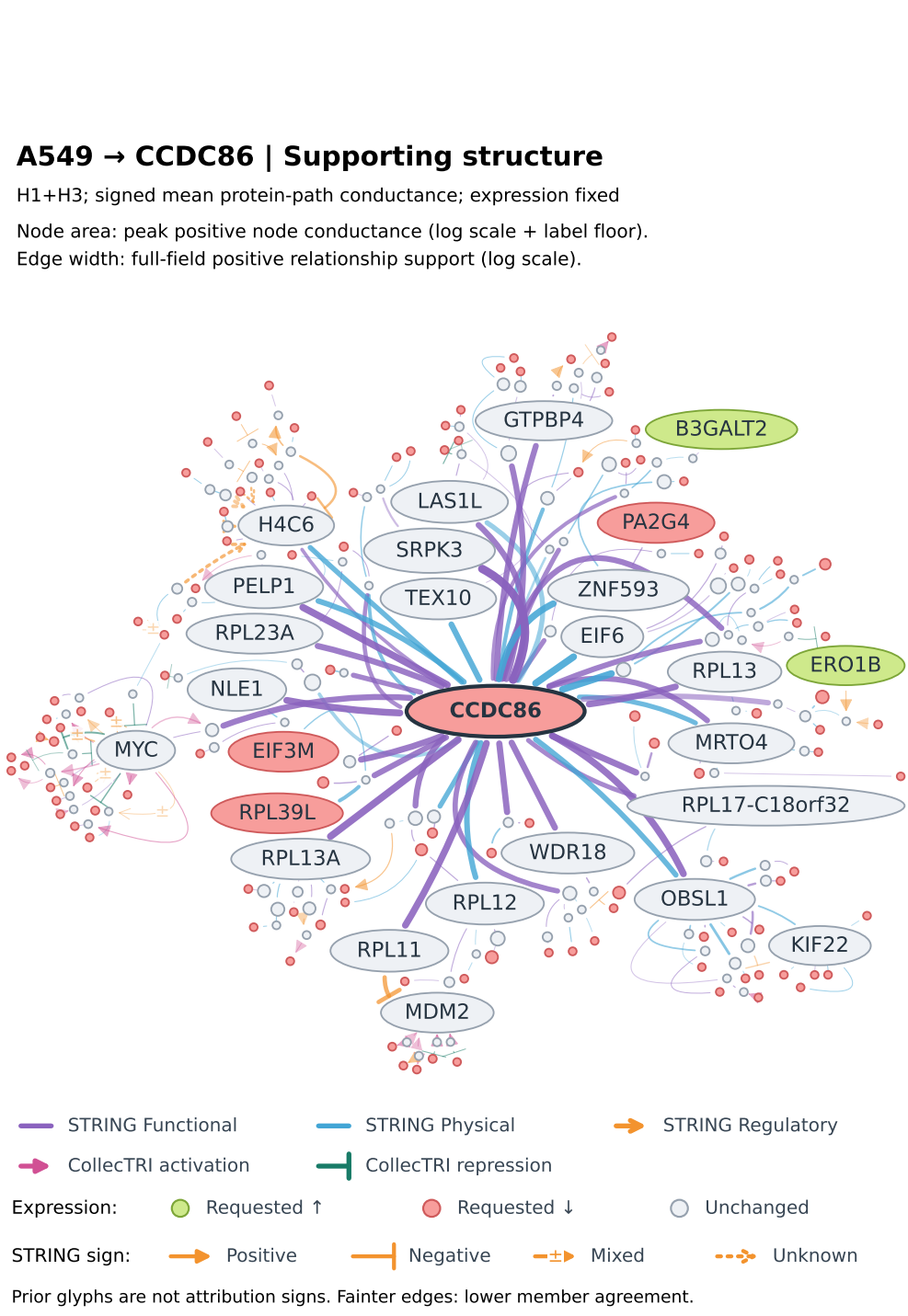

Figure 13: **CCDC86 under the same expression request.** The selection retains a different supporting structure, including nucleolar/ribosome-associated intermediates. The definitions of node size, edge width, edit colors, prior glyphs, and model agreement are the same as in Figure 2; display scales are local to each case, not a shared quantitative cross-case scale. These connections describe the fitted computation on the supplied graph, not an experimentally established causal chain in A549.

Table 6: The two fixed-request nominations. Mean-logit ranks use the same 17,225 candidates. Member ranks expose fit variation. N_{90} counts the smallest set carrying 90% of positive node conductance after three message rounds, before the final decoder state. It does not measure how many genes are biologically necessary.

Nominee	Consensus rank	H1 / H3 rank	$\Delta s^{\text{protein}}$	N_{90} after 3
PYGL	6	5 / 10	+8.8719	8
CCDC86	3	4 / 2	+11.7054	1

average signed scalar allocations after each member passes its numerical checks:

$$\bar{C}_i^{(\ell)} = \tfrac{1}{2} \sum_{f \in \{\text{H1},\text{H3}\}} C_{i,f}^{(\ell)}, \qquad \Delta \bar{s}_k^{\text{protein}} = \tfrac{1}{2} \sum_f [s_{k,f}(1) - s_{k,f}(0)]. \tag{22}$$

The same rule applies to messages and residuals. Initial feature channels are summed within each gene before cross-model averaging; hidden coordinate bases are not aligned or averaged. Mean actual scores agree with the ranking table. Opposing member contributions can cancel before any positive-part display rule.

E.1 Select a structure without changing the attribution

For one member, a route connects an edited marker's initial state to the nominee's final state with exactly one transition per layer, using positive messages or residuals. A widest route maximizes its minimum transition conductance (Vassilevska et al., 2009). Ties prefer residuals, then relation and original-edge order. The union retains every node and connection used by at least one selected route; shared instances are counted once. Unreachable anchors remain recorded rather than receiving invented connectors.

For consensus, candidate routes are complete member witnesses. A witness is eligible only when every corresponding mean message/residual conductance exceeds 10^{-5}; among eligible witnesses, the greatest bottleneck is selected for each anchor. This does not assemble a route from unrelated positive fragments of different fits. The path is a display of positive marginal transition allocations: positivity does not establish the sign or necessity of an isolated complete route. Exact ordered-route integration is a separate optional inspection, not a reported experiment on the primary displayed paths.

E.2 Read the overview and profile as complementary objects

Node size uses $\max_\ell [C_i^{(\ell)}]_+$, with a logarithmic mapping and a visibility/text floor. It is not an incident-edge sum or a conserved cross-layer total. Edge width sums positive conductance over all full-field layer instances of the displayed biological relationship; route selection determines which relationships appear, not the width's denominator. Reverse computational channels return to source-prior orientation for the overview. STRING physical/functional pairs remain undirected; regulatory sign annotations preserve positive, negative, mixed, or unassigned source evidence.

Names are the union of the 18 largest displayed node summaries, 12 highest displayed degrees, up to four strongest favorable-marker summaries, up to four strongest unfavorable-marker summaries, and the nominee. Overlap reduces the label count. Unnamed genes remain in the selected graph and exports. Node fill denotes the requested edit; neutral includes unchanged inputs and graph context without mapped expression. Neither a prior repression bar nor a requested decrease fixes the sign of a learned score contribution.

Full signed node arrays and typed message records underlie the profiles. Panel A separates positive mass from opposing magnitude for each complete node layer and message/residual transition. Panel B selects positive and opposing peaks, retains raw values with one within-case symmetric-log scale, and attributes the same final prediction in every column. Panel C's upper part counts components needed for 90% of positive conductance; its lower part compares selected positive conductance with full positive conductance, separately for messages and residuals. A row with zero positive mass has undefined coverage, not an invented zero or 100%.

Table 7: **Suppression summaries by selection group.** Each row pools layer-message instances across four transitions after equal-weight averaging across H1/H3. Means are signed logit units; they can conceal cancellation. Sign matches count $E_e D_e > 0$ for individual positive/negative selections, separately for Actual and Gap drop; zero effects do not match. Controls have no prescribed sign. These are descriptive summaries of the selected instances.

Nominee	Selection	n	$\overline{E}$	$\overline{D}^{\text{actual}}$	$\overline{D}^{\text{gap}}$	Sign matches Actual; Gap
PYGL	High positive	24	+0.665	+0.295	+0.215	18/24; 18/24
PYGL	High negative	24	-0.360	-0.122	-0.145	14/24; 15/24
PYGL	Low $\lvert E\rvert$	21	-0.001	-0.012	+0.104	–
PYGL	Matched random	20	+0.014	+0.006	+0.002	–
CCDC86	High positive	24	+0.062	+0.062	+0.060	22/24; 23/24
CCDC86	High negative	24	-0.050	-0.035	-0.046	21/24; 22/24
CCDC86	Low $\lvert E\rvert$	24	-0.000	-0.004	-0.004	–
CCDC86	Matched random	24	+0.000	+0.002	-0.005	–

The overview's log scales and profile color scales are case-specific display conventions. Quantitative cross-case comparisons use the recorded values and denominators. Layout positions and edge bends carry no biological distance or temporal meaning. Paper print reflow retains the scientific values, identities, selected nodes/relationships, and their source records.

F FINITE MESSAGE SUPPRESSION

The conductance values, native expression request, protein references, and checkpoint members match Sections 5–6. Both members held A549 out of training and validation. Sampling uses the full signed receptive field, rather than only the positive routes selected for the network overview.

At each of the four transitions, we select up to six largest positive and six most negative mean message conductances. A 10^{-5} sign cutoff excludes near-zero entries; it is not a significance or integration-error threshold. All transitions here contain six extremes of each sign, well above that cutoff. For each positive selection we sample a low-$|E|$ control from the lowest absolute-conductance quartile and a random control in the same transition and typed relation. Selections are disjoint, unavailable controls are omitted, and seed 20260924 fixes the selection before suppression outcomes are evaluated. Each transition contains 24 sampled messages, except PYGL's final transition, which contains 17 because only three low-$|E|$ and two random controls are available.

An intervention sets one original layer/relation/edge gate from one to zero. Reverse channels, other occurrences of the same relationship, residuals, and original degree factors remain unchanged. We recompute the full forward pass at both actual and reference protein endpoints with frozen weights. The binary gate intervention produces a continuous score change through the nonlinear model. Conductance follows those nonlinearities along the original protein path; neither finite drop is algebraically equal to the original message conductance. Equal-weight mean member drops are compared with signed mean conductance using Spearman correlation. Selecting extremes and controls changes the sampled distribution, so these correlations are not estimates for all messages or significance claims.

Table 7 shows directional agreement in the group means, with more consistent individual Gap-drop signs for CCDC86 than PYGL. Smaller mean control effects do not imply zero effects for every control. Pooling transitions summarizes the sampled messages without measuring the combined effect of suppressing a group or establishing independent biological replicates.

Some tested messages also lie on selected overview routes. For example, suppressing the final STRING regulatory PHKB $\rightarrow$ PYGL message lowers the actual score and protein-reference gap. The final GYG2 $\rightarrow$ PYGL reverse-channel message lowers the actual score but increases the gap, because the reference score falls further. This explains why the two contrasts must be retained. A pooled overview relationship can occur at several transitions; a layer-specific suppression tests only one instance. Residual paths and other messages remain available, so agreement does not rule out redundancy or nonlinear interactions.

Table 8: PYGL matches whose pinned source labels are *Preclinical* and *glycogen phosphorylase inhibitor*. This declared subset is illustrative; complete matches and their other targets are retained. Stages and targets are source annotations, not measurements performed here.

Compound	Reported stage	Annotated target
CP-316819	Preclinical	PYGL
CP-91149	Preclinical	PYGL

The frozen evidence contains the selection, consensus and member effects, correlations, and source study receipts. The notebook retains the complete signed scatterplots and matched-control diagnostics. The paper summary reproduces the correlation panel only.

G Compound evidence and scope

The explorer uses the pinned August 18, 2025 Broad Drug Repurposing Hub annotation snapshot, checked by SHA-256 before use. Exact gene symbols are matched against the full target-list field. No fuzzy alias expansion or clinical-status inference is performed, and no selected gene is sent to the source server during offline matching.

For PYGL, the full set includes records with broad target lists, metabolite-related annotations, and differing source-reported stages. The main table uses the explicit source-label filter *Preclinical* plus *glycogen phosphorylase inhibitor*; it does not claim those rows are exhaustive pharmacology or that a one-gene annotation proves selectivity. The full matched rows and source identity are preserved in the frozen paper evidence. CCDC86 has no exact-symbol match in this resource, which is not evidence that no compound exists or that the target is undruggable.

A source label such as *Launched* or *Phase 3* is not an independent check of present FDA approval, oncology indication, trial status, target engagement, or an effect compatible with knockout. The role of the lookup is to place existing evidence beside a model-generated hypothesis for a researcher to examine.

H Robustness across QC-passing baselines

The query baseline (Appendix A) is one of 44 A549 plates whose matched controls all pass QC (20–33 controls each). To test whether the reported cases depend on that choice, we repeat the query with each of these plates as a single-plate baseline—the same construction used for every training pair—and with a pooled mean of all 1,790 QC-passing A549 controls across 78 plates. The pooled baseline is smoother than any training baseline and is reported separately. Weights, graph, request rule, validated markers, and H1/H3 members are unchanged.

Table 9: **Rank robustness across QC-passing baselines.** Mean-logit ranks over H1/H3 among 17,225 candidates. The final row summarizes all 44 single-plate QC-passing baselines. Overlap and Spearman ρ compare the full ranking with the query baseline.

Baseline	PYGL (H1/H3)	CCDC86 (H1/H3)	Top-20 overlap	ρ
Query baseline (single plate)	6 (5/10)	3 (4/2)	20/20	1.00
Pooled QC-passing controls	5 (3/9)	2 (4/2)	18/20	0.98
44 QC-passing plates, median [range]	5 [2–7]	2 [2–6]	–	–

Both nominees rank within the top 7 on every QC-passing plate (Table 9). Member disagreement persists across plates (PYGL: H1 ranks 1–7, H3 ranks 6–15), so it is a property of the fits rather than of a particular baseline.

We also repeated protein-path conductance under the pooled baseline. For PYGL, 12 of the 15 largest positive node contributors are shared with the query baseline, including GYG2, PHKB, PHKG1, PHKG2, PYGM, and PYGB; the protein-reference difference is +5.44, N_{90} after three rounds is 7, and final-transition messages contribute +1.73 against a residual of +3.71, so the

glycogen-associated structure is stable while the magnitude of final-layer message support varies with the baseline. For CCDC86, all 15 largest contributors are shared, and the difference (+12.21), $N_{90} = 1$, and residual-dominated final transition (+11.96) are essentially unchanged. Outputs and scripts are included with the code.

I EXECUTION RECORDS AND REPRODUCIBILITY BOUNDARY

The current notebook's saved load-mode execution restores identified weights, verifies the common request and held-out membership, and exports both protein-conductance cases with their member checks, selected routes, full signed arrays, finite suppression results, and figure records. The paper freezes small identified result snapshots so numerical macros, tables, and figures are traceable without loading raw assays or weights during compilation. A paper-only build is distinct from rerunning the scientific workflow.

An earlier clean-room execution, recorded separately in `reports/reproduction_execution.json`, acquired the public inputs, prepared them, and completed twelve full CUDA depth/rotation fits and a 43-code-cell notebook with an earlier message-gate analysis. It produced a different fit. That record supports execution of that historical recipe; it is not presented as independent reproduction of the current 39-code-cell protein-conductance notebook, the two current nominations, or their stability across refits. The historical artifacts remain preserved rather than being relabeled as current results.