TARGEN: A STRUCTURAL SCORE BEYOND MOLECULAR EVIDENCE FOR CANCER TARGET GENE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cancer-associated gene prediction requires integrating molecular measurements with biological relations. We study whether learning the predictive contributions of a candidate's connection partners improves ranking beyond molecular and functional scores and their neighborhood propagation. We develop TARGEN, a regularized additive model that combines multi-omics data, optional pretrained GeneRAIN representations, and functional gene sets. Alongside direct scores and their propagation through protein–protein interaction and functional relations, TARGEN learns node coefficients from training labels and aggregates them into a structural score. For the linear special case with fixed relation operators and scale, we derive a stationary-point identity under L2 regularization: the structural score is a weighted sum of training residuals, with weights determined by shared-neighborhood similarity. This identity explains how training labels contribute to a candidate's structural score. In five repetitions of ten-fold cross-validation, TARGEN achieves pan-cancer average precision (AP) of 0.921 and cancer-specific macro-average AP of 0.841. It obtains the highest mean AP among the evaluated configurations for pan-cancer prediction and in 15 of 16 cancer types. Compared with a control that fixes node coefficients to zero, the full model improves pan-cancer AP by 0.278 percentage points and cancer-specific macro-average AP by 1.663 percentage points. Each model repeats selection and fitting independently; mean AP improves in all 16 cancer types. Public patient-derived organoid data further support growth dependency on the prescreened candidate GRB2 in organoids from 34 of 49 patients with esophageal adenocarcinoma. Together, the score decomposition, linear analysis, and paired ablation explain the structural score and support its complementary predictive value in the evaluated tasks.

1 INTRODUCTION

Cancer-associated gene prediction prioritizes candidates for follow-up experiments from a large gene pool, providing clues to tumorigenesis. Multi-omics data describe gene mutations, expression, DNA methylation, and copy-number alterations; protein–protein interaction (PPI) networks describe interactions between gene-encoded proteins; and functional gene sets provide pathway, functional-category, and perturbation-related annotations. These sources describe genes' molecular states and their relationships with other genes. GenePlexus learns network connection patterns characteristic of an input gene set to identify candidates with similar network neighborhoods, including genes without direct links to the input genes (Mancuso et al., 2022). Converting this information into useful predictive evidence underlies improved candidate ranking.

Existing methods use graph representation learning and multi-source integration to predict cancer genes from molecular data and biological relations. EMOGI integrates multi-omics features and PPI networks through graph convolution, using both a gene's own information and its neighborhood (Schulte-Sasse et al., 2021). DISFusion combines relations derived from PPI networks and functional gene sets to enhance their shared information (Deng et al., 2025). GRAFT further combines multi-network representations, gene-set representations, and network structural encodings, using a Transformer to model relationships between genes (Cho & Cho, 2026a). These methods learn and integrate gene representations to use information from different sources. Building on this work, 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.

ask a more specific question: can the identities of a candidate's connected genes yield a supplementary score that further improves ranking beyond existing molecular and functional information and its neighborhood propagation?

Learning the predictive role of network connections from training labels provides a basis for such a supplementary score. Liu et al. (2020) used gene connections as inputs to regularized supervised gene classification. Correct and Smooth (C&S) propagates differences between training labels and predictions over a network to correct and smooth classifier predictions (Huang et al., 2021). Drawing on these ideas, we apply regularized additive modeling to cancer-associated gene prediction. Here, evidence propagation means a weighted sum of connected genes' molecular and functional scores along PPI and gene-set relations. We learn each gene's contribution as a connection partner from training labels. Each candidate aggregates these coefficients along its connections to form the structural score, which contributes to prediction together with the existing scores. We first isolate the additive structural score in the linear model, then assess node coefficients within the complete predictor.

TARGEN investigates a focused hypothesis: after direct molecular and functional evidence has been scored and propagated, the connection partners of a candidate gene can still provide an additional label-derived structural signal for ranking cancer-associated genes.

- We introduce a trainable connection-partner structural score, $G\mathbf{s}$, that complements local evidence and standard graph propagation within a jointly optimized predictor.
- We isolate this score in controlled reduced-evidence experiments and degree-preserving graph rewiring tests, showing that its gains are positive across the evaluated settings and depend on the observed biological network structure.
- We derive a linear stationary-point attribution identity that makes the structural contribution traceable, and evaluate the complete model across cancer-specific benchmark tasks, public comparison configurations, and a prespecified organoid follow-up case study.

2 Related Work

2.1 Multi-omics and Biological Network Learning

EMOGI integrates mutation, methylation, expression, and copy-number features through graph convolution (Schulte-Sasse et al., 2021). MODIG applies attention across relation types (Zhao et al., 2022), whereas EMGNN connects gene representations across molecular networks (Chatzianastasis et al., 2023). TREE integrates molecular and network representations through a Transformer (Su et al., 2025a). MTGCN uses molecular and structural features to jointly optimize gene classification and link prediction (Peng et al., 2022).

Recent architectures adapt network representations at different levels. IMI-driver integrates multiple cancer-specific gene networks (Shi et al., 2024). ATTAG combines network-derived embeddings with topology-adaptive graph convolution and attention to predict cancer driver genes in bladder urothelial carcinoma (BLCA) (Li et al., 2026). DriverMONI uses condition-specific PPI subnetworks in multimodal learning (Park & Jo, 2026); ORBIT uses context-adaptive reweighting and hyperbolic prototype learning (Cho & Cho, 2026b). TARGEN fixes relation operators within each fit and distinguishes direct scoring, evidence propagation, and node-coefficient aggregation. This decomposition enables separate interventions on the scoring paths.

2.2 Functional Annotations and Pretrained Representations

Curated gene sets encode pathways, perturbation signatures, and ontology membership. DISHyper represents these memberships as hyperedges (Deng et al., 2024). DISFusion combines PPI and gene-set views through graph and hypergraph encoders (Deng et al., 2025). GRAFT integrates PPI, pathway co-occurrence, and Gene Ontology similarity through a Transformer (Cho & Cho, 2026a). TARGEN uses C2&C5 membership to construct gene-set scores and functional relations as its functional annotation inputs.

Geneformer and scGPT learn gene and cell representations from single-cell transcriptomes (Theodoris et al., 2023; Cui et al., 2024). GeneRAIN learns gene-expression relationships from

approximately 410,000 human bulk RNA-sequencing samples (Su et al., 2025b). TARGEN uses principal components of its frozen representations as molecular inputs.

2.3 Supervised Connection Learning and Residual Propagation

Supervised learning can directly use adjacency, diffusion, or embedding features to predict gene annotations (Liu et al., 2020). GenePlexus applies this approach to gene prioritization, including lung adenocarcinoma prediction (Yuan et al., 2025); LOTUS combines molecular and PPI diffusion kernels (Collier et al., 2019). UniMP unifies feature and label propagation through masked label prediction (Shi et al., 2021). C&S propagates training residuals and smooths classifier predictions (Huang et al., 2021).

TARGEN learns regularized node coefficients jointly with molecular and gene-set scores. Its linear identity links the fitted structural score to residuals of the joint model. We fix $\mathbf{s} = 0$ and refit the remaining parameters to evaluate the complementary value of node coefficients through the structural score and bounded correction.

3 Method

3.1 Problem Definition and Molecular Features

Let $\mathcal{V} = \{1, \ldots, N\}$ be the candidate gene set, $\mathcal{T} \subset \mathcal{V}$ the training gene set, and y_i the binary label of training gene $i \in \mathcal{T}$. TARGEN predicts a pan-cancer or cancer-specific association score for each gene. Learning is transductive: molecular features and biological relations are available over a fixed candidate set, while the supervised loss uses only training labels.

Four molecular feature types across 16 cancer contexts form 64 numerical inputs: mutation score (MF), DNA methylation (METH), gene expression (GE), and copy-number alteration (CNA). Each feature is scaled by its root mean square (RMS) over observed entries in training genes, including observed zeros. Observation masks identify entries used to estimate the scale. Both excluded entries and measured zeros produce zero inputs; the numerical molecular features constitute the omics input. Denote the resulting matrix by $O \in \mathbb{R}^{N\times 64}$.

We also use 32 principal components of frozen GeneRAIN representations (Su et al., 2025b). A label-free principal component analysis (PCA) transform is fitted on candidate genes with representations and then fixed. In each fit, the PCA output is scaled as a block using covered training genes to obtain $Z \in \mathbb{R}^{N\times 32}$. Internal training-set validation compares a configuration without GeneRAIN with four GeneRAIN coefficient regularization strengths. The feature matrix is $X = [O, Z]$ when GeneRAIN is included and $X = O$ otherwise. Selection within the training set therefore yields 96 or 64 features. Appendix A describes the transformation and selection procedures.

3.2 Biological Relations and Propagation

Protein–protein interaction networks. We use six PPI networks: ConsensusPathDB (CPDB), STRING, iRefIndex v9 and v15, MULTINET, and PCNet. Let W_r be the original adjacency matrix of network r and D_r its degree matrix. The fixed normalized operator is

$$P_r = D_r^{-1/2} W_r D_r^{-1/2}, \qquad r = 1, \ldots, 6. \tag{1}$$

Rows for zero-degree nodes remain zero. GeneRAIN enters through the feature matrix; PPI operators retain the original interactions and normalization.

Functional gene sets. We use the MSigDB C2&C5 membership matrices released with DISFusion (Deng et al., 2025). Each fit retains sets containing at least three positive training genes. For the retained membership matrix $H \in \{0,1\}^{N\times M}$, define

$$B = D_g^{-1/2} H D_e^{-1/2}, \qquad A = D_B^{-1/2} B, \qquad K = BB^\top, \tag{2}$$

where $(D_g)_{ii} = \sum_e H_{ie}$, $(D_e)_{ee} = \sum_i H_{ie}$, and $(D_B)_{ii} = \sum_e B_{ie}^2$. Inverse scaling factors for zero degrees are set to zero. Nonzero rows of A have unit Euclidean norm, and K retains its diagonal. Matrix A provides direct gene-set scores, while K supplies the seventh propagation relation.

Relation fusion. For each gene, available PPI and gene-set relations each receive half the total weight. If only one class is available, it receives all the weight; the PPI share is divided equally among available networks. Weights are zero when neither class is available. Denoting these node-specific weights by ω_{ir}, define

$$L_r = g\, diag(\omega_{:r}) P_r \ (r \le 6), \quad L_7 = g\, diag(\omega_{:7}) K, \quad \mathcal{G} = \sum_{r=1}^{7} L_r, \tag{3}$$

where $g = 0.500$. Node-specific fusion weights allow the operator to be asymmetric. Gene sets, fused relations, and scales are constructed from each training subset and held fixed during fitting and prediction.

3.3 Structural Score and Overall Prediction

Let β denote feature coefficients, θ gene-set coefficients, and $\mathbf{s} \in \mathbb{R}^N$ trainable node coefficients. Define

$$\begin{aligned} a = A\theta, \quad u = X\beta + a, \quad v = u + \alpha\mathbf{s}, \\ \alpha = \left(\frac{\|A_{\mathcal{T},:}\|_F^2}{\|\mathcal{G}_{\mathcal{T},:}\|_F^2} \right)^{1/2}, \qquad G = \alpha\mathcal{G}. \end{aligned} \tag{4}$$

The scale uses the combined operator, including cross-relation terms, and remains fixed during optimization. Relation aggregation decomposes as $\mathcal{G}v = \mathcal{G}u + G\mathbf{s}$: $\mathcal{G}u$ propagates molecular and functional scores, while $G\mathbf{s}$ aggregates node coefficients learned from training labels. We define $G\mathbf{s}$ as the structural score. For candidate gene i, it is $\alpha \sum_j \mathcal{G}_{ij} s_j$, where s_j represents the contribution of gene j, as a connection partner, to its neighbors' predictions.

The relation-specific correction inputs are

$$h_r = L_r v \quad (r = 1, \ldots, 6), \qquad h_7 = a + \sum_{r=1}^{6} L_r a + L_7 v. \tag{5}$$

Equivalently, $h_7 = (I + \mathcal{G})a + L_7(X\beta + \alpha\mathbf{s})$ combines gene-set contributions with feature and node scores propagated through functional relations. With global intercept b and seven shared correction parameters γ_r, predictions are

$$\begin{aligned} z = b1 + u + \mathcal{G}u + G\mathbf{s} + c(v, a), \\ c_i(v, a) = \textstyle\sum_{r=1}^{7} \tanh(\gamma_r) \tanh(h_{ir}/\tau), \\ p_i = sigmoid(z_i), \end{aligned} \tag{6}$$

where $\tau = 1$. Each relation contributes a correction with absolute value at most 1. All score components share the same classification objective.

3.4 Training Objective and Model Selection

The objective combines mean binary cross-entropy over training genes with quadratic penalties:

$$\begin{aligned} \mathcal{L} = & \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} [log(1 + expz_i) - y_i z_i] \\ & + \frac{\lambda}{2} \left(\| \beta_O \|_2^2 + \kappa \| \beta_Z \|_2^2 + \theta^\top M_A \theta \right) + \frac{\lambda_s}{2} \| \mathbf{s} \|_2^2 + \frac{\lambda_\gamma}{2} \| \gamma \|_2^2 . \end{aligned} \tag{7}$$

Here, β_O and β_Z correspond to the numerical omics and GeneRAIN blocks. Configurations without GeneRAIN omit the latter and its penalty. All four omics modalities use the same coefficient penalty. For retained gene set S_e, with $\mathcal{T}_+$ denoting positive training genes, define

$$\pi_e = \frac{|S_e \cap \mathcal{T}_+|}{|S_e \cap \mathcal{V}|}, \qquad \bar{\pi} = \frac{1}{M} \sum_e \pi_e, \qquad M_A = diag(\bar{\pi}/\pi_e). \tag{8}$$

We set $\lambda = \lambda_s = 10^{-4}$ and $\lambda_\gamma = 10^{-3}$. Within each outer training fold, a class-stratified 20% validation subset selects, by AP, either no GeneRAIN or $\kappa \in \{1, 10, 100, 1000\}$. Preprocessing is first estimated from inner training genes, then rebuilt using all outer training genes before fitting the selected model. All trainable parameters are initialized to zero and optimized jointly by L-BFGS-B; the intercept is unpenalized. Appendix A details optimization and sparse computation.

3.5 Source of Supervision for the Linear Structural Score

Fixing $\mathcal{G}$ and α and setting $\gamma = 0$ gives the linear logit

$$z = b\mathbf{1} + (I + \mathcal{G})(X\beta + A\theta) + G\mathbf{s}. \tag{9}$$

At a stationary point with $\lambda_s > 0$, let $\epsilon_j = y_j - p_j$ be the residual for training gene j. The structural score $f_{s,i} = (G\mathbf{s})_i$ of candidate gene i satisfies

$$\begin{aligned} \mathbf{s}^* &= \frac{1}{|\mathcal{T}|\lambda_s} G_{\mathcal{T},:}^\top \epsilon_{\mathcal{T}}, \\ f_{s,i} &= \frac{1}{|\mathcal{T}|\lambda_s} \sum_{j \in \mathcal{T}} (GG^\top)_{ij}(y_j - p_j). \end{aligned} \tag{10}$$

The kernel $GG^\top$ measures weighted shared-neighborhood similarity, and residuals come from the joint linear fit. With fixed operators and scale, this identity decomposes the fitted linear structural score by training gene. Appendix A provides the derivation. This decomposition applies to the linear special case defined here.

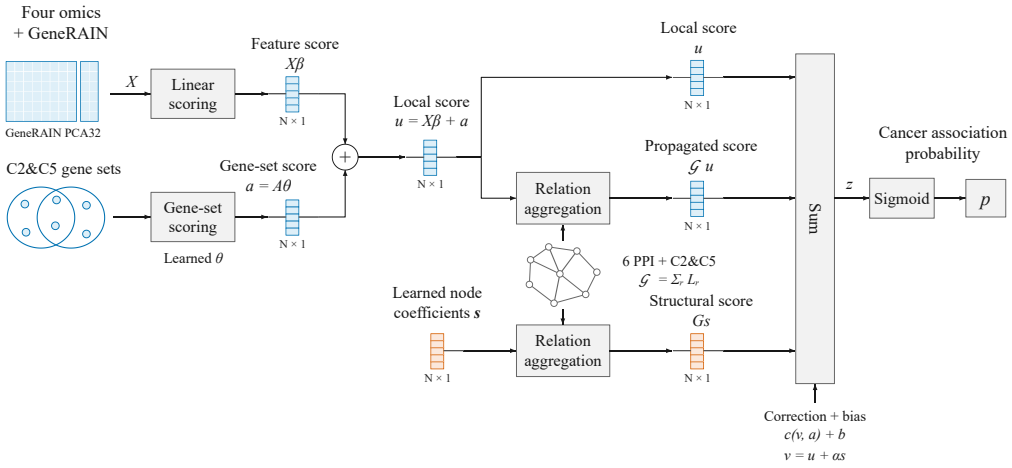


Figure 1: TARGEN framework. Four omics modalities and optional GeneRAIN principal components yield the feature score $X\beta$, which is added to the C2&C5 gene-set score $a = A\theta$ to obtain the gene's own score u. Six PPI networks and C2&C5 functional relations form the fused operator $\mathcal{G}$, which propagates u and aggregates trainable node coefficients $\mathbf{s}$ to yield the propagated score $\mathcal{G}u$ and structural score $G\mathbf{s}$. These scores, the bounded correction $c(v, a)$, and bias b are summed and passed through sigmoid to obtain the cancer-association probability p, where $G = \alpha\mathcal{G}$ and $v = u + \alpha\mathbf{s}$.

4 Experiments

4.1 Setup and Evaluation Protocol

The 16 cancer-specific tasks use DriverDB-derived positive labels and share 1,030 negative genes. The pan-cancer task uses 796 positive and 2,187 negative genes from the CPDB benchmark released with DISFusion. Labels, splits, and public model configurations follow Appendix C.

Each cancer-specific model uses the target cancer's four molecular columns; the pan-cancer model uses 64 columns across all 16 cancers. Inner training-set validation selects whether to include 32 GeneRAIN principal components. Six PPI networks and training-filtered C2&C5 sets provide relations. Appendices A and E describe implementation and data sources.

Five repetitions of ten-fold cross-validation give 50 test-fold values per metric. We report their equally weighted means and sample standard deviations; macro-averages weight the 16 cancers equally. Normalized AP is $(AP - \pi)/(1 - \pi)$, where π is test-fold positive prevalence. We compute

AP differences from unrounded metrics in percentage points (pp). Paired gene-cluster resampling holds fitted predictions fixed. We report two-sided intervals and one-sided tests for higher reference AP (Appendix D).

4.2 Isolated structural-score gains under reduced evidence

Structural gains across data-retention levels. We cross four numerical-omics retention levels with four training-label retention levels (Figure 2A). The evidence-propagation reference combines the gene's own score with a weighted sum of connected genes' scores, $z = b\mathbf{1} + u + \mathcal{G}u$; the structural model adds Gs. Both retain GeneRAIN selection and C2&C5, with bounded correction disabled and preprocessing rebuilt within each fit (Appendix I.2). The paired AP difference is positive in all 16 settings, with an equally weighted mean gain of +0.557 pp. Averaging over the other dimension, the gain increases by 0.503 pp from full to zero retention of numerical omics and by 0.389 pp from full to 25% training labels. Table I.4 retains the cell values, endpoint contrasts, and intervals.

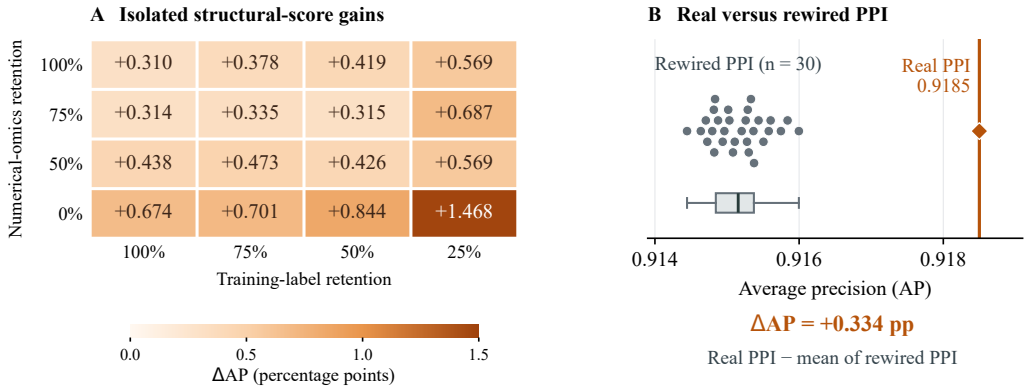


Figure 2: **Structural gains increase when direct evidence is reduced and depend on observed network topology.** (A) Pan-cancer ΔAP in percentage points: the linear structural model minus the evidence-propagation reference, with bounded correction disabled in both. Each cell averages the same 50 paired test folds; all 16 cells have one-sided $p < 0.05$. At 0% numerical-omics retention, GeneRAIN selection, gene sets, and relation information remain available. (B) Each gray point is the mean AP over 50 folds for one of 30 degree-preserving PPI rewiring instances; the box below the points shows the median and interquartile range, and whiskers span the observed range. The vertical line marks real-PPI AP. We rewire only the structural branch's PPI connections; evidence propagation and C2&C5 retain real relations.

Connection identities carry information beyond degree. The linear structural model using real PPI exceeds all 30 degree-preserving rewired instances, with a paired gain of 0.334 pp over their mean (Figure 2B). Rewiring changes only the PPI component of the structural-scoring branch; evidence propagation, C2&C5, and structural-scale calibration retain their real-relation settings. Every instance repeats selection and fitting on the same 50 folds. This contrast links predictive value to which genes are connected while preserving their unweighted degrees. Appendix F reports all instances and the separate GeneRAIN correspondence control.

4.3 Node-coefficient ablation in the complete model

We assess the contribution of trainable node coefficients by fixing $\mathbf{s} = 0$ throughout training and prediction. This removes Gs and the contribution of $\mathbf{s}$ to bounded correction. The control retains molecular and gene-set scores, evidence propagation, and bounded correction with trainable γ. Each model uses the same outer splits and independently repeats inner selection and outer fitting (Appendix D.8).

The full model improves cancer-specific macro-average AP from 0.825 to 0.841, a gain of 1.663 percentage points. Mean differences are positive in all 16 cancers, with the largest gains in THCA, BLCA, and LUAD: 4.760, 3.789, and 2.857 points, respectively. The macro-average gain has a

two-sided 95% interval of [1.291, 2.035] points. One-sided tests give $p < 0.05$ for the macro-average and 14 cancers (Table 1). The paired comparison supports the predictive contribution of node coefficients through the structural score and bounded correction.

Table 1: Node-coefficient ablation with $\mathbf{s} = 0$; both models retain bounded correction. AP is the mean over 50 test folds; the cancer-specific macro-average weights all 16 cancers equally. ΔAP is full model minus $\mathbf{s} = 0$, in percentage points (pp), using unrounded values. One-sided tests assess higher AP for the full model.

Task	**Full AP**	$\mathbf{s} = 0$ **AP**	Δ**AP (pp)**	p
Cancer-specific macro-average	0.841	0.825	+1.663	**<0.05**
BLCA	0.732	0.694	+3.789	**<0.05**
BRCA	0.846	0.833	+1.366	**<0.05**
CESC	0.938	0.930	+0.787	**<0.05**
COAD	0.875	0.861	+1.403	**<0.05**
ESCA	0.841	0.815	+2.588	**<0.05**
HNSC	0.821	0.808	+1.207	**<0.05**
KIRC	0.923	0.915	+0.770	**<0.05**
KIRP	0.884	0.880	+0.403	0.223
LIHC	0.941	0.934	+0.695	**<0.05**
LUAD	0.697	0.668	+2.857	**<0.05**
LUSC	0.710	0.691	+1.870	**<0.05**
PRAD	0.914	0.913	+0.147	0.252
READ	0.909	0.897	+1.231	**<0.05**
STAD	0.794	0.779	+1.497	**<0.05**
THCA	0.765	0.717	+4.760	**<0.05**
UCEC	0.871	0.858	+1.243	**<0.05**
Pan-cancer	0.921	0.918	+0.278	**<0.05**

4.4 Cancer-specific prediction results

TARGEN achieves the highest mean AP among the compared configurations in 15 of 16 cancer types (Table 2), with a cancer-specific macro-average AP of 0.841. The corresponding macro-averages are 0.755 for DISFusion, 0.751 for deepCDG, and 0.742 for EMOGI. In THCA, deepCDG achieves a mean AP of 0.841, exceeding TARGEN's 0.765.

Within each cancer type, the four methods share labeled genes and training/test splits and use only the target cancer's molecular feature columns. Each method retains its own molecular modalities, network resources, and training settings; Appendix D.4 details implementations. These results show that TARGEN ranks candidate genes well in most evaluated cancer types.

4.5 Comparison of pan-cancer prediction results

TARGEN obtains the highest mean pan-cancer AP, nAP, and AUROC among the public method configurations evaluated here (Table 3). Comparators include DISFusion (Deng et al., 2025), GRAFT (Cho & Cho, 2026a), EMGNN (Chatzianastasis et al., 2023), deepCDG (Wu et al., 2025), and EMOGI (Schulte-Sasse et al., 2021). Mean AP is 0.921 for TARGEN, compared with 0.908 for DISFusion and 0.891 for deepCDG. The difference from DISFusion is approximately 1.3 AP percentage points.

Methods use the same collection of molecular data sources while retaining their respective modalities, network resources, preprocessing, and gene coverage. Table 3 summarizes all 16 method configurations. This comparison evaluates the predictions from each complete configuration.

The $\mathbf{s} = 0$ control reaches AP 0.918, with a full-model gain of 0.278 pp (Table 1). This gain is smaller than the cancer-specific macro-average gain. The one-sided test gives $p < 0.05$ (Appendix D.8).

Table 2: Prediction results across 16 cancer types. AP is the mean ± sample standard deviation over five repetitions of ten-fold cross-validation; bold indicates the highest mean in each row. Macro-averages weight the 16 cancer types equally.

Cancer	TARGEN	DISFusion	deepCDG	EMOGI
BLCA	**0.732 ± 0.122**	0.605±0.127	0.616±0.118	0.596±0.123
BRCA	**0.846 ± 0.057**	0.740±0.087	0.750±0.080	0.759±0.077
CESC	**0.938 ± 0.019**	0.883±0.031	0.842±0.038	0.820±0.037
COAD	**0.875 ± 0.040**	0.782±0.059	0.765±0.056	0.742±0.061
ESCA	**0.841 ± 0.077**	0.646±0.114	0.622±0.120	0.637±0.127
HNSC	**0.821 ± 0.082**	0.817±0.082	0.813±0.087	0.808±0.104
KIRC	**0.923 ± 0.048**	0.770±0.099	0.757±0.091	0.786±0.099
KIRP	**0.884 ± 0.075**	0.869±0.079	0.857±0.088	0.877±0.085
LIHC	**0.941 ± 0.026**	0.866±0.050	0.846±0.051	0.827±0.051
LUAD	**0.697 ± 0.136**	0.602±0.124	0.638±0.119	0.606±0.123
LUSC	**0.710 ± 0.114**	0.586±0.132	0.536±0.129	0.534±0.132
PRAD	**0.914 ± 0.053**	0.874±0.056	0.859±0.065	0.889±0.061
READ	**0.909 ± 0.033**	0.800±0.049	0.737±0.057	0.696±0.063
STAD	**0.794 ± 0.072**	0.733±0.067	0.732±0.071	0.720±0.083
THCA	0.765±0.263	0.666±0.282	**0.841 ± 0.236**	0.776±0.246
UCEC	**0.871 ± 0.027**	0.836±0.034	0.807±0.040	0.795±0.045
Macro-average AP	**0.841**	0.755	0.751	0.742

Table 3: Pan-cancer prediction results. AP and AUROC are means ± sample standard deviations over 50 folds; nAP is the mean of fold-wise values.

Model / network	AP	nAP	AUROC	MCC	F1
TARGEN	**0.921±0.028**	**0.892**	**0.958±0.016**	**0.792**	**0.841±0.037**
DISFusion-CPDB	0.908±0.026	0.874	0.955±0.015	0.762	0.825±0.036
deepCDG-iRefIndex v15	0.891±0.034	0.812	0.905±0.029	0.641	0.795±0.042
deepCDG-MULTINET	0.888±0.034	0.818	0.914±0.023	0.654	0.792±0.037
deepCDG-PCNet	0.884±0.027	0.841	0.938±0.016	0.708	0.782±0.043
deepCDG-STRING	0.881±0.029	0.812	0.912±0.020	0.653	0.779±0.040
GRAFT-CPDB	0.880±0.030	0.833	0.935±0.019	0.723	0.800±0.034
EMGNN (six networks)	0.879±0.035	0.835	0.938±0.019	0.710	0.784±0.038
deepCDG-CPDB	0.869±0.034	0.821	0.931±0.018	0.689	0.772±0.037
deepCDG-iRefIndex v9	0.868±0.033	0.817	0.928±0.019	0.685	0.773±0.034
EMOGI-PCNet	0.831±0.028	0.768	0.908±0.018	0.315	0.519±0.021
EMOGI-MULTINET	0.821±0.046	0.709	0.868±0.028	0.247	0.595±0.018
EMOGI-iRefIndex v15	0.810±0.046	0.674	0.844±0.033	0.066	0.592±0.019
EMOGI-STRING	0.783±0.040	0.660	0.848±0.027	0.285	0.589±0.017
EMOGI-CPDB	0.778±0.034	0.698	0.885±0.018	0.270	0.490±0.013
EMOGI-iRefIndex v9	0.773±0.039	0.685	0.876±0.020	0.228	0.484±0.014

4.6 ORGANOID EVIDENCE FOR CANDIDATE CANCER GENES

We assess four prescreened TARGEN candidates in public patient-derived organoids (Herranz-Ors et al., 2026), using one observation per patient after excluding overlapping samples (Table 4). GRB2 in esophageal adenocarcinoma meets the prespecified support rule, providing external evidence of growth dependency. Appendix H details candidate selection, cohort processing, and the support rule.

Table 4: Dependency of four prescreened candidates in patient-derived organoids.

Gene	Rank	CRC		EAC	
		Dependent/tested	**Median LFC**	**Dependent/tested**	**Median LFC**
XRCC2	44	33/78	-0.465	14/49	-0.336
CRKL	63	16/78	-0.303	14/49	-0.297
PTK2	92	20/78	-0.408	29/49	-0.477
GRB2	98	27/78	-0.432	34/49	-0.612

CRC: colorectal cancer; EAC: esophageal adenocarcinoma. Counts report dependent/tested patients. LFC: scaled log2 fold change after knockout; more negative values indicate greater growth impairment.

5 DISCUSSION

TARGEN applies regularized additive modeling to cancer-associated gene prediction, using biological networks both to propagate existing gene scores and to aggregate node contributions learned from training labels. The structural score $G\mathbf{s}$ turns candidates' connection partners into supplementary predictive information. Comparisons with public models evaluate the complete method, whereas the $\mathbf{s} = 0$ ablation tests the contribution of trainable node coefficients. Pan-cancer and cancer-specific macro-average AP gains support their value for ranking. Public organoid data further support growth dependency on the prescreened candidate GRB2 in esophageal adenocarcinoma.

Experiments and theoretical analysis describe the predictive role and source of supervision of the structural score, respectively. The $\mathbf{s} = 0$ ablation measures the combined contribution of node coefficients through $G\mathbf{s}$ and bounded correction, while refitting the remaining parameters. In the linear special case, with fixed relation operators and scale, the stationary-point identity under L2 regularization decomposes the structural score into contributions from training residuals through shared neighborhoods. Together, these analyses show how training labels and network relations can be used for candidate gene ranking.

The scope of evaluation remains constrained by benchmarks and data sources. Comparisons with public models retain their respective modalities, graph resources, preprocessing, and training budgets, and therefore describe differences between the complete configurations evaluated. Functional annotations and labels may share evidence, and molecular assay coverage and cohort composition also affect inputs. Fixed biological relations do not describe cancer-specific interactions; the conditions for the linear stationary-point identity also differ from the full model's bounded correction. Organoid results support growth dependency on selected candidates; enrichment across the full ranking and clinical utility require their own evidence.

AI USE STATEMENT

We used generative AI tools to assist with manuscript drafting, translation and editing, figure and table preparation, and consistency checks between results and conclusions.

REFERENCES

Chatzianastasis, Michail; Vazirgiannis, Michalis; Zhang, Zijun. Explainable Multilayer Graph Neural Network for cancer gene prediction. Bioinformatics, 39(11): btad643, 2023.

Cho, Sang-Pil; Cho, Young-Rae. GRAFT: a graph-aware fusion transformer for cancer driver gene prediction. Briefings in Bioinformatics, 27(1): bbaf706, 2026a.

Cho, Sang-Pil; Cho, Young-Rae. ORBIT: Oncogenic Representation Learning via Bi-Prototype Contrastive Learning in Hyperbolic Space for cancer driver gene identification. Artificial Intelligence in Medicine, 181: 103486, 2026b.

Collier, Olivier; Stoven, Véronique; Vert, Jean-Philippe. LOTUS: A single- and multitask machine learning algorithm for the prediction of cancer driver genes. PLOS Computational Biology, 15(9): e1007381, 2019.

Cui, Haotian; Wang, Chloe; Maan, Hassaan; Pang, Kuan; Luo, Fengning; Duan, Nan; Wang, Bo. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods, 21: 1470–1480, 2024.

Deng, Chao; Li, Hong-Dong; Zhang, Li-Shen; Liu, Yiwei; Li, Yaohang; Wang, Jianxin. Identifying new cancer genes based on the integration of annotated gene sets via hypergraph neural networks. Bioinformatics, 40(Supplement 1): i511–i520, 2024.

Deng, Chao; Li, Hongdong; Wang, Jianxin. Improving Cancer Gene Prediction by Enhancing Common Information Between the PPI Network and Gene Functional Association. Proceedings of the AAAI Conference on Artificial Intelligence, 39(1): 137–145, 2025.

EMOGI official analysis notebook: `https://github.com/schulter/EMOGI/blob/master/analysis/newly_predicted_cancergenes.ipynb`

Herranz-Ors, C., et al. (2026). A tumour-derived organoid biobank maps cancer gene dependencies. Nature, 657, 765–774. `https://doi.org/10.1038/s41586-026-10830-y`

Himmelstein, Daniel S.; Baranzini, Sergio E. Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes. PLOS Computational Biology, 11(7): e1004259, 2015.

Huang, Qian; He, Horace; Singh, Abhay; Lim, Ser-Nam; Benson, Austin R. Combining Label Propagation and Simple Models Out-performs Graph Neural Networks. International Conference on Learning Representations, 2021.

Li, Sa; Shader, Jonah; Bhattacharya, Abhijeet; Ma, Tianle. Integration of Multi-Omics Data With Topology Adaptive Graph Convolutional Network for Cancer Driver Gene Identification, 2026.

Liu, Renming; Mancuso, Christopher A.; Yannakopoulos, Anna; Johnson, Kayla A.; Krishnan, Arjun. Supervised learning is an accurate method for network-based gene classification. Bioinformatics, 36(11): 3457–3465, 2020.

Liu, Chia-Hsin; Lai, Yo-Liang; Shen, Pei-Chun; Liu, Hsiu-Cheng; Tsai, Meng-Hsin; Wang, Yu-De; Lin, Wen-Jen; Chen, Fang-Hsin; Li, Chia-Yang; Wang, Shu-Chi; Hung, Mien-Chie; Cheng, Wei-Chung. DriverDBv4: a multi-omics integration database for cancer driver gene research. Nucleic Acids Research, 52(D1): D1246–D1252, 2024.

Mancuso, Christopher A.; Bills, Patrick S.; Krum, Douglas; Newsted, Jacob; Liu, Renming; Krishnan, Arjun. GenePlexus: a web-server for gene discovery using network-based machine learning. Nucleic Acids Research, 50(W1): W358–W366, 2022. `https://doi.org/10.1093/nar/gkac335`

Neiswender, J. V., Maffa, S., Brenan, L., et al. (2026). A dependency map enhanced with next-generation 3D cancer models. Nature, 657, 789–799. `https://doi.org/10.1038/s41586-026-10843-7`

Park, Jeongho; Jo, Kyuri. DriverMONI: Cancer Driver Gene Prediction With Multimodal Deep Learning Integrating Multiomics Data and Condition-Specific Network Information, 2026.

Peng, Wei; Tang, Qi; Dai, Wei; Chen, Tielin. Improving cancer driver gene identification using multi-task learning on graph convolutional network, 2022.

Schulte-Sasse, Roman; Budach, Stefan; Hnisz, Denes; Marsico, Annalisa. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. Nature Machine Intelligence, 3(6): 513–526, 2021.

Shi, Yunsheng; Huang, Zhengjie; Feng, Shikun; Zhong, Hui; Wang, Wenjing; Sun, Yu. Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification, 2021.

Shi, Peiting; Han, Junmin; Zhang, Yinghao; Li, Guanpu; Zhou, Xionghui. IMI-driver: Integrating multi-level gene networks and multi-omics for cancer driver gene identification, 2024.

Su, Xiaorui; Hu, Pengwei; Li, Dongxu; Zhao, Bowei; Niu, Zhaomeng; Herget, Thomas; Yu, Philip S.; Hu, Lun. Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning. Nature Biomedical Engineering, 9(3): 371–389, 2025a.

Su, Zheng; Fang, Mingyan; Smolnikov, Andrei; Dinger, Marcel E.; Oates, Emily C.; Vafaee, Fatemeh. GeneRAIN: multifaceted representation of genes via deep learning of gene expression networks. Genome Biology, 26: 288, 2025b.

Theodoris, Christina V.; Xiao, Ling; Chopra, Anant; Chaffin, Mark D.; Al Sayed, Zeina R.; Hill, Matthew C.; Mantineo, Helene; Brydon, Elizabeth M.; Zeng, Zexian; Liu, X. Shirley; Ellinor, Patrick T. Transfer learning enables predictions in network biology. Nature, 618: 616–624, 2023.

Wu, Yingzhuo; Xu, Jialuo; Li, Junming; Gu, Jia; Shang, Xuequn; Li, Xingyi. Deep graph convolutional network-based multi-omics integration for cancer driver gene identification. Briefings in Bioinformatics, 26(4): bbaf364, 2025.

Yuan, Fei; Zhang, Yu-Hang; Huang, FeiMing; Cao, Xiaoyu; Chen, Lei; Li, JiaBo; Shen, WenFeng; Feng, KaiYan; Bao, YuSheng; Huang, Tao; Cai, Yu-Dong. Prediction of Lung Adenocarcinoma Driver Genes Through Protein–Protein Interaction Networks Utilizing GenePlexus. Proteomics, 25(5--6): e202400296, 2025.

Zhao, Wenyi; Gu, Xun; Chen, Shuqing; Wu, Jian; Zhou, Zhan. MODIG: integrating multi-omics and multi-dimensional gene network for cancer driver gene identification based on graph attention network model. Bioinformatics, 38(21): 4901–4907, 2022.

A Implementation and Training

A.1 Numerical Omics Preprocessing

The 64 numerical columns comprise four modalities across 16 cancer contexts, ordered first by modality and then by cancer type. Mutation scores retain their supplied magnitudes. Expression and methylation use absolute tumor–normal differences; CNA uses the nonnegative fraction of altered records. These inputs describe alteration magnitudes or frequencies. Appendix G.2 compares the current GE/METH representation with a joint replacement that retains both signs. Let F_{ij} denote these values and m_{ij} the observation mask. A mask value of 1 indicates source-supported measurement evidence, including measured zeros. For the current training set $\mathcal{T}$, define

$$\rho_j = \left(\frac{\sum_{i \in \mathcal{T}} m_{ij} F_{ij}^2}{\sum_{i \in \mathcal{T}} m_{ij}} \right)^{1/2}, \qquad O_{ij} = m_{ij} F_{ij} / \rho_j. \tag{11}$$

If the denominator or RMS is zero, set $\rho_j = 1$. All genes within a fit use the same training-estimated scale. Observed zeros contribute to the denominator; unavailable or uncertain entries remain zero. Masks serve only as preprocessing metadata. Appendix E describes the construction of molecular summary features.

A.2 Frozen GeneRAIN Principal Components

GeneRAIN provides 200-dimensional representations. Exact gene-symbol matching yields representations for 18,404 of 34,842 candidate gene records. We subtract the representation mean over these genes and normalize each centered vector to unit Euclidean length. We then center again, project onto the first 32 principal components, and retain their unwhitened scales. This label-free transform uses represented genes from the full candidate set and remains fixed throughout cross-validation. Genes without representations receive all-zero PCA rows. The input block contains 32 numerical principal components.

Let $\widetilde{Z}$ be the fixed PCA output and $\mathcal{C}$ the set of genes with nonzero PCA rows. The block scale and input for each fit are

$$\eta = \left(\frac{1}{64 |\mathcal{T} \cap \mathcal{C}|} \sum_{i \in \mathcal{T} \cap \mathcal{C}} \| \widetilde{Z}_i \|_2^2 \right)^{1/2}, \qquad Z = \widetilde{Z} / \eta. \tag{12}$$

The mean squared norm of represented training rows is thus 64. Each fit requires a finite, positive block scale and uses the fixed label-free PCA transform.

A.3 Construction within Training Folds and Model Selection

Inner validation in each fit selects GeneRAIN inclusion and penalty settings as described.

Each fit reconstructs numerical RMS scales, C2&C5 filtering, gene-set penalty weights, relation-fusion weights, and the structural scale from the current training subset. Each retained set contains at least three positive training genes. The structural scale uses the squared Frobenius norm of the combined propagation operator, including all cross-relation terms. Both its numerator and denominator must be positive. Gene-set relations retain diagonal entries.

Within each outer training set, a class-stratified 20% validation subset compares five configurations: no GeneRAIN, or GeneRAIN coefficient penalty multipliers of 1, 10, 100, or 1000. The remaining 80% is used for all training-dependent preprocessing. The inner-split seed is $20260915 + f$, where the fold index is $f \in \{0, \ldots, 9\}$. Selection maximizes inner-validation AP. Exact ties favor no GeneRAIN, followed by stronger GeneRAIN penalties. The selected configuration is then rebuilt on the complete outer training set and fitted from zero. Preprocessing, selection, and fitting use labels from the corresponding training subset.

Of the full pan-cancer model's 50 outer folds, four select the 64-dimensional numerical omics input and 46 select numerical omics plus GeneRAIN principal components. Reported gains average over all 50 folds and evaluate the complete training-set selection procedure.

A.4 Optimization

All trainable feature, gene-set, node, correction, and intercept parameters start at zero. We use L-BFGS-B with analytical gradients, unbounded coefficients, and double-precision objective calculations. Optimization allows at most 1,500 iterations, a gradient tolerance of 10^{-7}, a relative objective tolerance of 10^{-12}, and 30 steps per line search; the quasi-Newton approximation retains 20 correction-vector pairs. The intercept is unpenalized. Base penalties are 10^{-4} for numerical omics and node coefficients and 10^{-3} for correction coefficients. GeneRAIN coefficients use the selected multiplier of the base feature penalty. Gene-set coefficients use the inverse relative-to-mean training-positive proportions defined in Section 3.4.

A.5 Sparse Computation and Complexity

PPI operators and gene-set membership matrices use compressed sparse row format. Functional-relation multiplication is implemented as $Kx = B(B^\top x)$, and the fused operator sums weighted sparse products. The objective and gradient are computed on supervised output rows while retaining state over the full candidate set.

Let E be the total number of PPI nonzero entries, J the number of membership-matrix nonzeros, and d the input width. With seven fixed channels, one objective/gradient evaluation requires $O(E + J + Nd + N + M)$ arithmetic operations. The model has $P = N + M + d + 8$ trainable parameters, including the intercept and correction coefficients. Sparse operators, features, and model state require $O(E + J + Nd + N + M)$ storage. L-BFGS-B retains 20 history-vector pairs, adding $O(20P)$ storage and vector operations. Total optimization work also depends on the number of objective evaluations.

Before optimization, exact structural-scale calibration partitions the fused operator $\mathcal{G}$ into blocks of at most 64 rows and accumulates squared norms, retaining cross-relation terms. This calculation expands shared-membership pairs within blocks, so its cost depends on membership overlap. A temporary block contains at most $64N$ entries; K and $\mathcal{G}$ remain implicit in the implementation.

A.6 Linear Structural-score Identity

For the linear model, let $\epsilon_{\mathcal{T}} = y_{\mathcal{T}} - p_{\mathcal{T}}$ and $n_T = |\mathcal{T}|$. With fixed propagation operators and scale,

$$\nabla_{\mathbf{s}} \mathcal{L} = -\frac{1}{n_T} G_{\mathcal{T},:}^\top \epsilon_{\mathcal{T}} + \lambda_s \mathbf{s}. \tag{13}$$

Setting the gradient to zero and substituting into $G\mathbf{s}$ yields Equation (10). Because the fused operator can be asymmetric, the kernel is $GG^\top$. This identity provides an additive decomposition of the structural score at the fitted solution.

B GeneRAIN Evaluation

B.1 Comparison with a Model without GeneRAIN

Both methods in the GeneRAIN comparison retain numerical omics, original PPI operators, and training-constructed C2&C5 relations. The reference model omits the representation block. TARGEN selects between this configuration and the four GeneRAIN penalty multipliers described in Appendix A. Both methods use identical outer splits and labeled genes.

Mean AP increases from 0.917 ± 0.029 to 0.921 ± 0.028. The paired difference is +0.418 AP percentage points, with a paired 95% interval of [+0.107, +0.732]. The mean difference is also positive in every repetition (Table B.1). Each row equally averages its ten fold-wise AP values.

Table B.1: GeneRAIN comparison by repetition. AP is the equally weighted mean over the corresponding ten folds; differences use unrounded scores.

Repetition	Without GeneRAIN	TARGEN	ΔAP (pp)
1	0.917	0.921	+0.364
2	0.917	0.921	+0.381
3	0.913	0.918	+0.497
4	0.917	0.921	+0.369
5	0.919	0.924	+0.479

B.2 Recall and F1

At a fixed probability threshold of 0.500, the GeneRAIN selection procedure also improves recall and F1 over the model without representations (Table B.2). Recall increases by +2.235 percentage points and F1 by +1.336 percentage points. Intervals use the same 2,000 paired gene-cluster resamples as the original GeneRAIN comparison, averaging fold-wise metrics over 50 folds. These auxiliary metrics use the interval construction for the GeneRAIN baseline comparison in Appendix D.3.

Table B.2: Auxiliary metrics for the GeneRAIN comparison. Values are means ± sample standard deviations over 50 folds. Differences are TARGEN minus the model without GeneRAIN, in percentage points.

Metric	Without GeneRAIN	TARGEN	Δ (pp)	95% interval (pp)
Recall	0.766±0.052	0.789±0.047	+2.235	[+1.142,+3.280]
F1	0.828±0.039	0.841±0.037	+1.336	[+0.540,+2.091]

B.3 Gene–Representation Correspondence

The current correspondence experiment uses 30 independent row permutations among represented genes while retaining the original inputs, relations, and inner-selection procedure. Appendix F.1 provides the protocol, all instance results, and paired inference computed together with PPI randomization.

C Evaluation Labels and Coverage of Public Methods

C.1 Pan-cancer Labels in the CPDB Benchmark

We use the fixed positive and negative lists accompanying the ConsensusPathDB (CPDB) cancer-gene benchmark released with DISFusion and based on the EMOGI benchmark (Deng et al., 2025; Schulte-Sasse et al., 2021). The released files `796true.txt` and `2187false.txt` contain 796 positive and 2,187 negative genes, respectively. Case-insensitive exact gene-symbol matching confirms agreement with all labeled genes used in the pan-cancer splits. CPDB denotes the network setting; the accompanying reference lists provide cancer-association labels.

Positive labels in this benchmark derive from sources including known cancer genes in the Network of Cancer Genes (NCG) and DigSEE literature evidence. Following EMOGI preprocessing, negative-label construction excludes genes with cancer or disease annotations, including NCG candidates, Online Mendelian Inheritance in Man (OMIM) entries, COSMIC annotations, and KEGG cancer-pathway members. The released CPDB lists remain fixed during evaluation. The remaining 31,859 unlabeled candidate genes do not enter the supervised loss.

Public-method results come from reproductions using the same collection of molecular data sources, with each method's modality selection, network resources, and preprocessing. TARGEN is implemented as in Section 3.

Outer splits are mapped to each method's graph by gene symbol, and labeled genes present in that graph are evaluated. Table C.1 records the resulting training and test coverage. TARGEN and

DISFusion-CPDB cover all 2,983 labeled genes. The ranking in Table 3 uses each configuration's evaluable gene set, with label coverage reported in Table C.1.

Table C.1: Label coverage of public method configurations. The omics count is the number of molecular modalities; TARGEN also uses GeneRAIN features selected within the training set.

Model / network	Omics count	Positive / negative
TARGEN	4	796/2,187
DISFusion-CPDB	3	796/2,187
deepCDG-iRefIndex v15	3	750/1,050
deepCDG-MULTINET	3	743/1,194
deepCDG-PCNet	3	796/2,109
deepCDG-STRING	3	745/1,304
GRAFT-CPDB	3	787/2,002
EMGNN (six networks)	4	796/2,187
deepCDG-CPDB	3	796/2,187
deepCDG-iRefIndex v9	3	783/2,023
EMOGI-PCNet	4	796/2,109
EMOGI-MULTINET	4	743/1,194
EMOGI-iRefIndex v15	4	750/1,050
EMOGI-STRING	4	745/1,304
EMOGI-CPDB	4	796/2,187
EMOGI-iRefIndex v9	4	783/2,023

For test-fold positive proportion π, normalized AP is $(AP - \pi)/(1 - \pi)$. Normalized AP and the Matthews correlation coefficient (MCC) are computed fold-wise and averaged over all 50 folds. Classification metrics use a probability threshold of 0.500.

C.2 Cancer-specific Labels and Splits

Positive labels. Cancer-specific positive labels are derived from computationally identified candidates in DriverDB, using data downloaded on September 10, 2026 and fixed during evaluation. For 16 cancer types, we use candidates supported by at least three detection tools in the corresponding The Cancer Genome Atlas (TCGA) projects. We cite the DriverDB resource description (Liu et al., 2024) and separately record the download date and data version; the website identified the release as version 5 at download.

Negative labels. The CPDB benchmark supplies 2,187 initial negative candidates in `2187false.txt`. Before defining cross-validation folds, we exclude genes appearing in a positive-label list for any of the 118 projects in the above DriverDB release, removing 1,157 genes from the negative pool. Restriction to genes jointly covered by TARGEN and the three CPDB-based comparator implementations leaves 1,030 negatives used consistently across all 16 cancer types. Positive labels are also restricted to this common set of 13,627 genes. Gene alignment uses case-insensitive exact symbol matching. The resulting positive and negative reference lists define the labeled scope of supervised learning; all other genes remain unlabeled. Table C.2 gives positive-label counts.

For each cancer type, all four methods share every labeled gene and the complete training/test splits. Five repetitions of stratified ten-fold splitting use seeds 1234–1238. Each gene appears in exactly one test fold per repetition. Each method completes 800 outer fits across 16 cancer types, for 3,200 fits evaluated in total. This cancer-specific set differs from the pan-cancer label set above.

Table C.2: Cancer-specific positive-label counts. Every cancer type uses the 1,030 negatives defined by the procedure. The four methods in Table 2 share these labels and training/test splits.

Cancer	Positive genes	Cancer	Positive genes
BLCA	116	LIHC	349
BRCA	136	LUAD	109
CESC	640	LUSC	144
COAD	307	PRAD	182
ESCA	112	READ	427
HNSC	124	STAD	283
KIRC	120	THCA	21
KIRP	109	UCEC	542

D Controls and Paired Statistical Analysis

D.1 Definitions of Biological-input Controls

We set coefficients for all 64 numerical omics inputs to zero while retaining candidate GeneRAIN configurations, functional relations, node coefficients, and other trainable terms. In candidate configurations without GeneRAIN, all feature coefficients are zero. The remaining parameters are then refitted. This intervention removes all contributions of numerical omics feature scores throughout the model.

C2&C5 functional knowledge. Full resource ablation removes the direct gene-set score, gene-set relations, and their correction term together. The remaining PPI relations define $\mathcal{G}$, and the structural scale is recalibrated as $\alpha = \sqrt{|\mathcal{T}|/ \parallel \mathcal{G}_{\mathcal{T},:} \parallel_F^2}$. The numerator assigns one unit of calibration mass to each training gene, and prediction uses the remaining PPI operators. The model is refitted with the recalibrated relations.

Direct gene-set score. We set $\theta = 0$, retaining the selected gene sets, relation operators, and calibrated structural scale. This removes $A\theta$ and its downstream score contributions while retaining gene-set relations, then refits the remaining parameters.

GeneRAIN. The reference without GeneRAIN uses the same original PPI operators and numerical omics but omits representation features. The correspondence control permutes representation rows among covered genes using 30 independent mappings and the same model-selection procedure. Both comparisons retain the original PPI operators.

D.2 Degree-preserving Connection Controls

Connection experiments use the linear model with bounded correction disabled. Direct evidence includes numerical omics, the selected GeneRAIN block, and learned gene-set scores, giving the logit

$$z = b\mathbf{1} + u + \mathcal{G}_P u + G_S \mathbf{s}, \qquad G_S = \alpha \mathcal{G}_S. \tag{14}$$

The real model uses original relations in both operators. Randomized controls retain the evidence-propagation operator $\mathcal{G}_P$ and change only the PPI part of $\mathcal{G}_S$. C2&C5 retains real relations in both parts, and the structural scale retains its calibration under real relations. Conditions share outer splits and inner selection, with all active coefficients refitted.

Each PPI source independently undergoes degree-preserving double-edge swaps, accepting five successful swaps per undirected edge. Its seed is the base seed plus 101 times the zero-based source index. The procedure rejects self-loops and duplicate edges and preserves each gene's unweighted degree and network availability. Normalized edge weights and weighted degrees are recomputed for randomized graphs before fitting. Appendix F.1 reports all 30 instances and paired comparisons against real relations.

D.3 Paired Estimation and Confidence Intervals

Let $AP_{rf}^{(a)}$ denote the AP of method a in repetition r, fold f. The paired estimate is

$$\Delta_{pp} = 100 \left[\frac{1}{50} \sum_{r=1}^{5} \sum_{f=1}^{10} \left(AP_{rf}^{(a)} - AP_{rf}^{(b)} \right) \right]. \tag{15}$$

Multiplication by 100 converts the difference to AP percentage points. Differences are computed from full-precision scores before rounding. Values following $\pm$ are sample standard deviations of the 50 fold scores.

Intervals use class-stratified paired resampling with genes as clusters. All occurrences of the same gene across repetitions and methods share the same bootstrap multiplicity. These shared weights are used to recompute fold-wise metrics and average them across folds, preserving method pairing and repeated predictions for each gene.

Score and biological-input controls use percentile intervals, as do AP, recall, and F1 in the original GeneRAIN baseline comparison. Additional score controls retain their original simultaneous intervals: a family of m comparisons uses quantiles $0.050/(2m)$ and $1 - 0.050/(2m)$. The three-comparison family comprises GeneRAIN correspondence, direct-only GeneRAIN scoring, and removal of node coefficients from the full model. The eleven-comparison family comprises propagation and structural supervision, two post-prediction corrections, and seven PPI randomization comparisons. Updated correspondence, randomization, and linear structural-supervision results use the expanded experiments in Appendix F; the remaining controls here retain their original family sizes.

Resampling changes gene counts in saved out-of-fold predictions while holding fitted parameters, hyperparameter selections, and graph instances fixed. Intervals describe variation under this resampling scheme. We standardized reporting to one-sided p-values after the existing analyses were completed, using saved resampling results without model retraining. The organoid analysis retains its prespecified candidates, contexts, and support rule; its p-value reporting is standardized here.

The pan-cancer EMOGI configurations in Table 3 use a PyTorch reproduction. Weighted binary cross-entropy directly receives raw logits for training nodes, with positive-class weight 30. Prediction applies sigmoid once. Across six networks, 300 fits use each network's native 64-dimensional omics, the original gene-mapped splits, and layer widths 64, 50, 100, and 1. Each fit uses 7,000 epochs, Adam with learning rate 0.001, dropout 0.500, and a penalty equal to 0.050 times half the squared norm of the first-layer weights. An independent weighted logistic-loss calculation verifies loss and gradients. Supplementary directory `emogi_corrected/` provides fold-wise metrics, out-of-fold probabilities, coverage, and metric-recomputation code.

D.4 Cancer-specific Inputs, Implementations, and Aggregation

For each target cancer, TARGEN uses four numerical columns: mutation score, methylation, expression, and copy-number alteration. Observation masks determine which observed training entries enter RMS preprocessing. Selection within training decides whether to add 32 GeneRAIN principal components, yielding four or 36 inputs. The two prediction tasks share the frozen GeneRAIN transform, six PPI operators, C2&C5 procedure, regularization strengths, and optimizer. The GeneRAIN block retains the normalization target of 64 in Equation (12). Inner 20% validation selects GeneRAIN inclusion and its penalty multiplier.

DISFusion and deepCDG retain mutation, methylation, and expression inputs and CPDB resources; EMOGI retains all four molecular modalities. Each method receives only columns for the corresponding cancer type. Fixed training budgets are 200 epochs for DISFusion, 1,200 for deepCDG, and 7,000 for EMOGI. Methods share labels and splits while retaining their respective biological inputs, graph coverage, preprocessing, and training budgets.

The cancer-specific EMOGI implementation uses the same weighted raw-logit loss and training settings, with layer widths 4, 50, 100, and 1. Table 2 reuses 800 completed fits, with verification of

source-code and checkpoint hashes, splits, feature columns, saved predictions, and metrics. Each fit uses the target cancer's four molecular columns and training labels.

We reuse completed DISFusion and deepCDG fits whose source configurations, seeds, training/test splits, and gene identifiers match, and complete the remaining fits with identical settings. Cancer-specific AP is the equally weighted mean of 50 separately computed test-fold AP values. Macro-average AP equally averages the 16 cancer-specific means. Bold indicates the highest mean AP for each cancer type.

D.5 Additional Scores and Resource Contributions

The score-restricted models in Table I.1 distinguish prediction from direct molecular features from prediction through learned node coefficients. Feature-only scoring uses $z = b\mathbf{1} + X\beta$, retaining numerical omics and the GeneRAIN block selected within training. Node-coefficient scoring uses $z = b\mathbf{1} + G\mathbf{s}$, with feature coefficients, gene-set coefficients, and bounded-correction parameters fixed to zero. The latter aggregates node coefficients through PPI and selected C2&C5 relations. Each model fits its active coefficients and intercept on training genes.

The direct molecular and gene-set score control in Table I.1 fits $z = b\mathbf{1} + X\beta + A\theta$. It retains numerical omics, the training-selected GeneRAIN block, and direct gene-set scores, while disabling evidence propagation, node coefficients, and bounded correction.

Table D.1: Additional score and resource comparisons. Full-model AP is 0.921 ± 0.028. AP is the mean $\pm$ sample standard deviation over 50 folds. Differences are full model minus control, in AP percentage points. See Appendix D.3 for interval construction.

Control	AP	ΔAP (pp)	95% interval (pp)
CPDB as the only PPI source	0.913±0.028	+0.779	[+0.410,+1.136]
Without CNA	0.916±0.028	+0.517	[+0.121,+0.901]
Without MULTINET	0.916±0.028	+0.464	[+0.224,+0.714]
Without PCNet	0.919±0.028	+0.186	[+0.110,+0.262]
Without STRING	0.919±0.029	+0.175	[+0.038,+0.312]
MF as the only omics modality	0.916±0.025	+0.450	[-0.257,+1.155]
CNA as the only omics modality	0.918±0.026	+0.317	[-0.452,+1.075]
GE as the only omics modality	0.903±0.026	+1.734	[+0.904,+2.552]
METH as the only omics modality	0.907±0.029	+1.424	[+0.551,+2.223]

PPI resources. Restricting PPI input to CPDB retains C2&C5 relations, numerical omics, and GeneRAIN selection. The six-network model gains +0.779 AP percentage points over this configuration. Removing MULTINET, PCNet, or STRING gives paired full-model gains of +0.464, +0.186, and +0.175 percentage points, respectively. Each removal excludes the network from evidence propagation and node-coefficient aggregation and sets its correction-channel input to zero. Relation-fusion weights and the structural scale are then recomputed from remaining resources, and the model is refitted. These comparisons support contributions from MULTINET, PCNet, and STRING.

Molecular modalities. Removing CNA excludes its 16 numerical columns while retaining the other three modalities, GeneRAIN selection, PPI networks, and C2&C5 knowledge. The full model gains +0.517 AP percentage points. Each single-omics configuration retains its corresponding 16 numerical columns together with GeneRAIN selection, PPI networks, and C2&C5 knowledge. Table D.1 records all four modalities: AP is 0.916 for MF and 0.918 for CNA. The GE and METH configurations give AP values of 0.903 ± 0.026 and 0.907 ± 0.029, respectively, supporting gains from the complete molecular input over these two configurations. Active parameters are refitted within outer training sets, with selection of GeneRAIN inclusion and regularization strength.

D.6 Supplementary Comparisons of Score Components

Table D.2 examines evidence propagation, node coefficients in the full model, and where GeneRAIN scores enter prediction. The propagation control removes $\mathcal{G}u$ from the linear model while retaining direct molecular and gene-set scores and structural supervision. The full-model control sets $\mathbf{s} = 0$

while retaining bounded correction. The direct GeneRAIN control adds representation scores to the gene's own prediction. Reparameterizing node coefficients can express the same family of prediction functions; the quadratic penalty acts on the control's coefficient coordinates. Each control repeats fitting and training-set selection.

Table D.2: Additional score comparisons. AP is the mean ± sample standard deviation across folds. The first row uses linear-model AP 0.919 ± 0.029 as reference; the other two use full-model AP 0.921 ± 0.028. Differences are reference minus control, in AP percentage points. Intervals retain the original comparison-family size m.

Control	**AP**	**ΔAP**	**Simultaneous 95% interval**	m
Without evidence propagation	0.915±0.030	+0.328	[-0.042,+0.661]	11
Full model with $\mathbf{s} = 0$	0.918±0.028	+0.278	–	–
Direct-only GeneRAIN score	0.920±0.028	+0.071	[-0.003,+0.147]	3

The reference configurations have positive mean differences in these comparisons. Appendix I.2 evaluates linear structural gains under 16 omics-data and training-label retention conditions.

D.7 Alternative Fitting and Regularization Settings

Table D.3 records two additional fitting procedures. The regularization-selection variant sets $\gamma = 0$ and selects a common multiplier of the base feature, gene-set, and node penalties from $\{0.100, 1, 10\}$ within training, jointly with GeneRAIN inclusion and its penalty. The staged variant first fits the linear model with $\mathbf{s} = 0$, fixes its predictions and coefficients, and then fits the structural score; inner validation selects the node-penalty multiplier from the same set. Both variants have mean AP close to the full model.

Table D.3: Score models under different fitting settings. Values are means ± sample standard deviations on the same 50 test folds.

Configuration	**AP**
Full model	0.921±0.028
Linear model, fixed base penalties	0.919±0.029
Linear model, base penalties selected within training	0.921±0.028
Structural score fitted after fixing the evidence-score model	0.920±0.029

D.8 Node-coefficient Ablation with Bounded Correction

We fix every node coefficient to zero during training and prediction, removing its contribution to the additive structural score and bounded correction. The control retains molecular inputs, functional relations, PPI, and bounded correction with trainable γ. It refits all remaining parameters with the original regularization and optimization settings. Pan-cancer and all 16 cancer-specific tasks use the same five repetitions of ten-fold splitting as the full model. Each control independently repeats inner selection and outer fitting. We combine 800 new cancer-specific fits with the 50 completed pan-cancer $\mathbf{s} = 0$ fits and reuse the corresponding full-model predictions.

We average AP equally over 50 folds and weight the 16 cancers equally for the macro-average. Both studies freeze all test predictions before evaluation and use 5,000 paired gene-cluster bootstrap draws. Cancer-specific draws assign exponential weights to genes, shared across methods, repetitions, and cancers. Pan-cancer draws retain the original class-stratified gene counts, shared across methods and repetitions. These procedures retain all repeated predictions for a gene within the same resampling unit.

For the observed difference $\Delta = AP_{\text{full}} - AP_{\mathbf{s}=0}$, let $e_b = \Delta_b^* - \Delta$. The one-sided test of higher full-model AP uses $p = (1 + \sum_{b=1}^{B} \mathbf{1}\{e_b \geq \Delta\})/(B + 1)$, with $B = 5000$. Table 1 reports these tests for each task and the cancer-specific macro-average. The cancer-specific macro-average

uses the two-sided 95% interval $\Delta \pm Q_{0.95}(|e_b|)$, giving [1.291, 2.035] percentage points, where Q denotes an empirical quantile.

E Data Sources and Molecular Feature Preprocessing

E.1 Public Sources and Sample Selection

Molecular inputs comprise 64 gene-level features from four modalities across 16 cancer types, aligned to 34,842 candidate gene records. This fixed-order collection adds H2BFS and SLC49A3 to the previously curated 34,840 records to retain all benchmark labels. The count refers to aligned records before further gene-identifier harmonization. We construct cohort-level summaries from 23 cBioPortal studies, 31 Gene Expression Omnibus (GEO) series, and one Open Science Framework (OSF) project. Among these 55 source identifiers, some studies contribute multiple cancer types or modalities. Gene-level aggregation integrates molecular summaries from the patient collections in these sources.

Cancer abbreviations are BLCA (bladder cancer), BRCA (breast cancer), CESC (cervical cancer), COAD (colon adenocarcinoma), ESCA (esophageal cancer), HNSC (head and neck squamous cell carcinoma), KIRC (kidney clear cell carcinoma), KIRP (kidney papillary cell carcinoma), LIHC (hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), PRAD (prostate adenocarcinoma), READ (rectal adenocarcinoma), STAD (stomach adenocarcinoma), THCA (thyroid cancer), and UCEC (endometrial cancer). Individual studies may represent more specific histological or clinical subgroups within these categories.

Mutation and copy-number data are obtained from cBioPortal using study-specific molecular profiles, sample lists, clinical annotations, and gene-panel definitions for targeted assays. Mixed-cancer studies are subset by cancer annotations. Expression and methylation inputs are publicly processed series or supplementary matrices with platform annotations and sample metadata. Tumor and normal specimens are selected within the same study and platform. KIRP methylation uses seven tumor–normal pairs from the OSF project. Tables E.1–E.4 list sources and selected sample counts for all 64 cancer–modality combinations. Supplementary file `data_sources/data_sources_64.tsv` additionally records molecular-profile identifiers and effective denominator ranges for targeted assays.

Analysis begins with selected molecular records and publicly processed matrices. Sample counts refer to tissue specimens; some studies contain repeated or spatially distinct specimens. Expression and methylation features use differences between tumor-group and normal-group summaries.

E.2 Gene Mapping and Molecular Summaries

Entrez identifiers in mutation and copy-number records are mapped to gene symbols and aligned to the candidate set. Expression and methylation probes are mapped to genes through platform annotations; the KIRP expression platform uses Entrez annotations associated with probe identifiers. Symbol matching is case-insensitive. Probes annotated to multiple mapped genes contribute separately to each gene through the aggregation procedures below.

Mutation score. Mutation scoring retains nonsilent mutation records in selected specimens and excludes specimens with more than 1,000 such records. Let k_{icj} be the retained mutation-record count for gene i in specimen j of cancer type c, t_{cj} the specimen's total nonsilent-record count, ℓ_i the reference gene-length parameter, ℓ_{tot} the fixed total annotation-record length defined below, and $\mathcal{S}_{ic}$ the retained specimens contributing to gene i. The score is

$$m_{ic} = \frac{1}{|\mathcal{S}_{ic}|} \sum_{j \in \mathcal{S}_{ic}} \frac{k_{icj}}{1 + t_{cj}\ell_i/\ell_{tot}}. \tag{16}$$

For the targeted mutation sources for CESC, KIRP, and READ, $\mathcal{S}_{ic}$ includes only specimens whose declared assay panels cover gene i, and k_{icj} counts only records within that panel. Burden t_{cj} counts nonsilent records downloaded for selected specimens before candidate mapping and panel restriction. Other sources use the retained specimen count as denominator. When no specimen is

eligible for a gene, a zero score and separate evidence metadata are stored. The resulting feature is a mutation-record score adjusted for mutation burden and length.

Reference lengths are calculated from GENCODE release 29 basic exon annotations by summing absolute differences between exon end and start coordinates across annotation records. This sum retains repeated exon records across transcripts. Before candidate mapping, all positive gene-length parameters in the annotation are summed to obtain $\ell_{tot} = 203,190,246$, which is shared across sources and specimens. Genes without positive mapped lengths receive the median positive reference length; these replacements do not change ℓ_{tot}. Scoring uses annotation-record lengths as a fixed reference, with $\mathcal{S}_{ic}$ defining the assayed specimen set. Because these inputs do not provide targeted capture or callable intervals, this reference does not calibrate assay-specific absolute mutation rates. Mutation filtering removes 65 specimens: 56 COAD, 1 LIHC, 2 LUAD, 1 PRAD, and 5 STAD. Copy-number sample selection follows each source's own criteria.

Gene expression. Exported expression matrices for BLCA, HNSC, and PRAD are treated as nonnegative intensity-scale matrices and transformed by $\log_2(max(x, 0) + 1)$; other sources retain the supplied transformed scale. Scale detection computes the 95th percentile from finite values for up to 250 mapped probes and up to 20 selected specimens per probe. If this value exceeds 100, the same $\log_2(max(x, 0) + 1)$ transform is applied. Each probe's feature is median tumor expression minus median normal expression; the gene-level effect is the median effect of available mapped probes. Both groups must contain finite values, and the resulting difference must be finite. Summaries use only finite values, and finite mapped expression effects form the input features.

DNA methylation. For each retained probe, mean normal signal is subtracted from mean tumor signal, then probe effects mapped to the same gene are averaged. Beta-value sources are used directly; M-values are converted by $1/(1 + 2^{-x})$. The KIRC methylated DNA immunoprecipitation microarray (MeDIP-chip) source transforms its log-ratio signals by $1/(1 + 10^{-x})$. This feature quantifies differences in methylation enrichment on the assay-specific scale.

Probe selection follows platform-specific annotations. Relevant 450K and EPIC arrays use the transcription-start-site-proximal groups TSS200/TSS1500; 27K arrays require an absolute distance of at most 1,000 bases from the transcription start site; other platforms use available distance or accession annotations. Some general annotation records with missing distance fields are retained. The OSF methylation matrix retains only probes with finite values in all 14 selected specimens. Construction preserves upstream processing of the supplied matrices and applies the annotation and finite-value criteria described here.

Copy-number alteration. The feature is the fraction of eligible gene-level records with an absolute discrete copy-number value of at least 1. For targeted profiles, records must have valid values and genes within each specimen's declared assay panel; duplicate gene–specimen records are checked during construction. Denominators are therefore gene-specific. For other profiles, all mapped nonmissing records in selected specimens are used. Neutral copy-number values contribute to the denominator. The score counts both gains and losses, including low-level gains and shallow deletions supplied by the source profile.

E.3 Numerical Representation and RMS Scaling within Training Folds

Mutation scores retain their original aggregated values; DNA methylation and expression use absolute tumor–normal differences. CNA retains the nonnegative fraction of altered records, combining copy-number gains and losses. Appendix G.2 reports a comparison retaining GE/METH signs jointly. Each cancer–modality column is scaled by the RMS over observed genes in the current training subset, as in Equation (11). Observed numerical zeros contribute to the denominator. Unavailable or uncertain entries are excluded from scale estimation and remain zero after transformation. Scaling uses the common candidate set and current training subset; across different PPI coverage patterns, molecular values retain the observation availability determined by their sources.

The resulting 64 numerical columns form the omics block. Observation masks determine entries included in preprocessing; the numerical values themselves form prediction features. Each inner fit estimates its own numerical scales, and the selected model re-estimates scales on the complete outer training set before fitting. Label-dependent C2&C5 filtering and penalty weights use the same training-subset boundaries.

E.4 Observation Masks and Assay Coverage

We retain four source-evidence categories: unavailable or excluded entries, observed nonzero values, observed numerical zeros, and mutation zeros from broad-coverage sequencing without gene-specific assay evidence. Observation masks include observed nonzero values and observed numerical zeros. Entries with uncertain coverage are excluded from scale estimation. Source-evidence metadata record these categories, distinguishing measured zeros from excluded entries that also receive zero inputs.

Sources differ in assay coverage, tissue composition, disease stage, and histological subtype. For example, KIRC methylation uses MeDIP-chip, the READ copy-number source supports only 16 mapped genes, and some thyroid and endometrial mutation sources represent particular histological subgroups. Genes are the evaluation units, and their molecular features summarize these source-specific assays and specimens.

E.5 Functional Gene-set Resources

We use C2&C5 membership matrices released with DISFusion and retain its repository's name-based filtering rules. C2 sets are excluded if their names contain underscore-delimited CANCER, TUMOR, or NEOPLASM tokens; for C5, this exclusion applies only to sets with the HP prefix. After mapping genes to the candidate set, label-free filtering removes empty sets and sets with fewer than two mapped members. This yields 20,623 fixed gene sets before fold-specific filtering. As in Section 3.2, each fit retains sets containing at least three positive members in the current training subset.

In the fixed matrix, 9,537 sets contain fewer than three benchmark-positive genes, including 4,272 with none; the reference here is all 796 benchmark positives. The positive-count threshold is applied separately to each training subset. The archived DISFusion matrices and preprocessing rules define the exact resource used in this study, which includes both cancer-related gene sets and general functional annotations.

E.6 Pretrained Gene Representations

The model uses the frozen 200-dimensional GeneRAIN representations described by Su et al. (2025b). Exact symbol matching aligns vectors to candidate genes. A fixed label-free PCA transform retains 32 components, followed by block scaling using represented training genes, as detailed in Appendix A. Genes without available vectors receive zero representation features. All six PPI operators retain their original edge definitions and normalization throughout the representation comparison.

E.7 Source Inventory

The following tables list source accessions for each cancer type and modality. T and N denote selected tumor and normal specimens. Mutation sample counts give selected specimens before and after nonsilent-record filtering. Restricting targeted records to each specimen's declared panel does not change these specimen counts, but effective denominators remain gene-specific.

Table E.1: Mutation-score sources.

Cancer	Study identifier	Assay scope	Specimens
BLCA	blca_bgi	Broad coverage	99 → 99
BRCA	brca_bccrc	Broad coverage	65 → 65
CESC	msk_met_2021	Targeted	39 → 39
COAD	coad_silu_2022	Broad coverage	281 → 225
ESCA	escc_icgc	Broad coverage	88 → 88
HNSC	hnsc_jhu	Broad coverage	32 → 32
KIRC	ccrcc_utokyo_2013	Broad coverage	106 → 106
KIRP	msk_impact_50k_2026	Targeted	69 → 69
LIHC	hcc_inserm_fr_2015	Broad coverage	243 → 242

Continued on next page

Table E.1 (continued)

Cancer	Study identifier	Assay scope	Specimens
LUAD	luad_oncosg_2020	Broad coverage	302 → 300
LUSC	nsclc_tracerx_2017	Broad coverage	155 → 155
PRAD	prostate_dkfz_2018	Broad coverage	313 → 312
READ	rectal_msk_2022	Targeted	612 → 612
STAD	stad_pfizer_uhongkong	Broad coverage	100 → 95
THCA	thyroid_gatci_2024	Broad coverage	190 → 190
UCEC	uccc_nih_2017	Broad coverage	16 → 16

Sample counts indicate selected specimens and, after the arrow, specimens retained after excluding those with more than 1,000 nonsilent records. Targeted scoring uses gene-specific panel denominators; zero-value coverage follows the source-evidence rules in Appendix E.

Table E.2: DNA methylation sources.

Cancer	Source accession	Platform	T / N
BLCA	GSE33510	GPL8490	149 / 6
BRCA	GSE31979	GPL8490	103 / 21
CESC	GSE211668	GPL13534	62 / 19
COAD	GSE29490	GPL8490	24 / 24
ESCA	GSE178212	GPL13534	24 / 16
HNSC	GSE25093	GPL8490	91 / 18
KIRC	GSE166734	GPL19930	11 / 11
KIRP	y8bh2	GPL13534	7 / 7
LIHC	GSE176273	GPL21145	16 / 16
LUAD	GSE137481	GPL8490	77 / 77
LUSC	GSE121849	GPL13534	20 / 13
PRAD	GSE39603	GPL9183	87 / 24
READ	GSE132668	GPL23976	32 / 5
STAD	GSE127857	GPL13534	24 / 34
THCA	GSE97466	GPL13534	74 / 50
UCEC	GSE155760	GPL23976	33 / 13

All entries except KIRP are GEO accessions. The OSF project provides seven tumor–normal pairs, with probes required to be finite across all 14 specimens. GPL19930 is MeDIP-chip; other methylation platforms use their respective array-signal definitions. For GSE155760, uterine cancer and normal endometrial specimens are selected.

Table E.3: Gene-expression sources.

Cancer	Source accession	Platform	T / N
BLCA	GSE3167	GPL96	41 / 9
BRCA	GSE70947	GPL13607	148 / 148
CESC	GSE7803	GPL96	21 / 10
COAD	GSE39582	GPL570	566 / 19
ESCA	GSE20347	GPL571	17 / 17
HNSC	GSE6631	GPL8300	22 / 22
KIRC	GSE40435	GPL10558	101 / 101
KIRP	GSE48352	GPL16311	24 / 8
LIHC	GSE25097	GPL10687	268 / 243
LUAD	GSE10072	GPL96	58 / 49
LUSC	GSE19188	GPL570	27 / 65
PRAD	GSE6919	GPL8300	65 / 81
READ	GSE20842	GPL4133	65 / 65
STAD	GSE29272	GPL96	134 / 134
THCA	GSE60542	GPL570	28 / 24

Continued on next page

Table E.3 (continued)

Cancer	Source accession	Platform	T / N
UCEC	GSE63678	GPL571	7 / 5

Counts are selected tumor and normal specimens within the same study and platform. Aggregation uses the separate medians of the tumor and normal groups.

Table E.4: Copy-number alteration sources.

Cancer	Study identifier	Assay scope	Specimens
BLCA	blca_msk_2024	Targeted	184
BRCA	brca_igr_2015	Broad coverage	216
CESC	msk_met_2021	Targeted	39
COAD	coad_silu_2022	Broad coverage	281
ESCA	egc_msk_2017	Targeted	43
HNSC	hnc_mskcc_2016	Targeted	151
KIRC	msk_met_2021	Targeted	421
KIRP	msk_impact_50k_2026	Targeted	69
LIHC	hcc_jcopo_msk_2023	Targeted	37
LUAD	luad_oncosg_2020	Broad coverage	302
LUSC	msk_met_2021	Targeted	537
PRAD	prad_fhcrc	Broad coverage	149
READ	rectal_msk_2022	Targeted	612
STAD	egc_msk_2017	Targeted	147
THCA	thyroid_gatci_2024	Broad coverage	137
UCEC	ucec_msk_2018	Targeted	197

Selected specimen counts differ from gene-specific denominators. Targeted summaries use only valid records within each specimen's declared panel; broad-coverage profiles use their available mapped numerical records.

F Supplementary Controlled Evaluations

Experimental-record index. Availability records contain per-fold metrics, held-out predictions, condition summaries, and resampling arrays for all 16 retention conditions in Table I.4. Original randomization, architecture, and adaptation summaries retain their respective comparison families.

F.1 Randomization Experiments with Thirty Instances

Base seeds 2026091701–2026091730 generate 30 randomized PPI graphs and 30 GeneRAIN permutations. Each instance uses the same five repetitions of ten-fold cross-validation and repeats training-set selection. Graph experiments use the linear model with $\gamma = 0$; representation experiments use the full model.

The PPI intervention follows Appendix D.2: independent double-edge swaps within each source preserve unweighted degrees, candidate records and their order, and network availability. Five successful swaps are performed per undirected edge. Original symmetric degree normalization is recomputed after rewiring, while the structural scale retains its real-relation calibration. Evidence propagation and C2&C5 retain real connections. Normalized weights change with rewired graphs, and each instance refits active coefficients and performs GeneRAIN selection.

The GeneRAIN intervention permutes 32-dimensional principal-component rows among represented genes, leaving uncovered rows zero. It preserves dimensionality and coverage, recomputes block scaling over covered training genes, and repeats inner selection between no GeneRAIN and four coefficient penalties. The model retains original PPI connections and the same training-dependent C2&C5 procedure.

Section 4.2 reports the PPI comparison, and Table F.2 reports the GeneRAIN comparison. Each real reference is compared with its average over 30 instances, with a one-sided p-value for each contrast. Table F.1 lists all instances in seed order. Each real mean AP exceeds all 30 corresponding instance means; paired inference concerns the difference between the real condition and the instance average.

Table F.1: All 30 random instances in each control family. Each entry is mean AP $\pm$ sample standard deviation over 50 test folds. Real references are 0.919 ± 0.029 for the linear PPI model and 0.921 ± 0.028 for the full GeneRAIN model.

Base seed	PPI rewiring	GeneRAIN permutation
2026091701	0.915±0.029	0.916±0.029
2026091702	0.915±0.029	0.916±0.029
2026091703	0.915±0.029	0.916±0.029
2026091704	0.915±0.029	0.916±0.029
2026091705	0.916±0.029	0.916±0.029
2026091706	0.915±0.029	0.915±0.030
2026091707	0.915±0.029	0.917±0.029
2026091708	0.915±0.029	0.916±0.029
2026091709	0.916±0.029	0.916±0.029
2026091710	0.916±0.029	0.916±0.030
2026091711	0.915±0.029	0.917±0.029
2026091712	0.915±0.029	0.917±0.029
2026091713	0.915±0.029	0.916±0.030
2026091714	0.915±0.029	0.916±0.029
2026091715	0.915±0.029	0.916±0.030
2026091716	0.914±0.029	0.916±0.029
2026091717	0.915±0.029	0.916±0.029
2026091718	0.916±0.029	0.916±0.029
2026091719	0.915±0.029	0.916±0.029
2026091720	0.915±0.029	0.917±0.029
2026091721	0.915±0.029	0.916±0.029
2026091722	0.915±0.029	0.916±0.029
2026091723	0.915±0.029	0.916±0.029
2026091724	0.915±0.029	0.916±0.029
2026091725	0.916±0.028	0.916±0.030
2026091726	0.915±0.029	0.916±0.030
2026091727	0.915±0.029	0.916±0.029
2026091728	0.916±0.028	0.916±0.029
2026091729	0.915±0.029	0.916±0.029
2026091730	0.915±0.029	0.916±0.029

Table F.2: GeneRAIN correspondence in the full model. The real condition is compared with average metrics across 30 row permutations. Differences and two-sided percentile 95% intervals are in AP percentage points; one-sided p-values are reported.

Intervention	Real condition	Randomized mean	ΔAP	95% interval	One-sided p
GeneRAIN permutation (full model)	0.921	0.916	+0.458	[+0.141,+0.777]	$<$0.05

G Supplementary Input and Score Controls

Accompanying records for the ten input and score controls below retain all comparisons, fold-wise AP/AUROC/nAP, ordinary and simultaneous intervals, one-sided p-values, and 500 completion records. Fitted models rank by continuous logits; fixed predictors use aggregated scores. Stratified tables use separately identified AP from concatenated out-of-fold predictions. Independent recomputation from saved predictions reproduces reported metrics, and recomputation from saved

bootstrap draws reproduces original paired statistics; one-sided p-values in this manuscript are recalculated from saved draws. Input and prediction checks also confirm the same full-model reference across the three comparison families.

G.1 Predictive Information beyond Mutation and Count Summaries

TARGEN outperforms four predictors based on mutation scores or counts of biological measurements and relations (Table I.3). These controls use the same 2,983 labeled genes, comprising 796 positives and 2,187 negatives, and the same five repetitions of ten-fold cross-validation. The reference reuses saved full-model predictions for 50 folds.

Two fixed predictors first scale mutation scores for the 16 cancer types by the RMS over observed training entries, then take their mean or maximum. Another predictor fits logistic regression on the same 16 columns. These predictors use only the burden- and length-adjusted mutation-record scores defined in Appendix E.

Count-feature logistic regression uses 11 variables: unweighted node degree in each of six PPI networks, gene-set membership count, and observed measurement counts for each of four modalities. Degree and annotation counts are transformed by $log(1 + x)$, then all 11 variables are scaled using training means and standard deviations. The static annotation pool is filtered by mapped set size, independently of positive-training-gene filtering. These diagnostic variables belong to the count-feature control; TARGEN retains numerical molecular inputs and masks used for preprocessing.

Both logistic regressions use the same inner 20% validation split and select inverse regularization strength from $C \in \{0.010, 0.100, 1, 10, 100\}$ by AP. Models use unweighted logistic loss, an L_2 penalty, and the L-BFGS solver, with at most 5,000 iterations and tolerance 10^{-8}. After selection, scales are re-estimated and models refitted using outer training genes.

The full model's advantage over these controls supports predictive information beyond fixed mutation aggregation and the fitted summaries tested here. In particular, its gain over the 11-dimensional count model indicates information beyond this combination of connectivity, annotation counts, and measurement counts.

G.2 Joint Gene Sets and Molecular Magnitudes

Using C2 and C5 jointly yields higher mean AP than either resource alone (Table I.2). Each resource intervention retains six PPI networks and repeats training-positive set filtering, relation fusion, structural-scale calibration, and inner GeneRAIN selection. These comparisons evaluate complete resource-selection and fitting procedures.

The signed-input control jointly replaces absolute GE and METH differences with their source-derived signed aggregate values. Numerical dimensionality, observation masks, MF, and CNA remain fixed. Squaring removes signs, so both representations use the same RMS scales over observed training entries. Table I.2 reports gains over this joint replacement, supporting the current magnitude representation within the evaluated model and joint comparison of these two inputs.

G.3 Sensitivity of the Seventh Correction Input

Equation (5) combines the direct gene-set score, its PPI propagation, and the seventh-relation term. We separately fit three simplifications: $L_7 v$, $a + L_7 v$, and $\sum_{r=1}^{6} L_r a + L_7 v$. Each variant retains node coefficients $\mathbf{s}$, the other six correction channels, numerical inputs, and gene-set resources, and repeats inner GeneRAIN selection and fitting.

Table G.1: Variants of the seventh correction input. Equation (5) gives the full input, with AP 0.921 ± 0.028. Rows show replacements; the PPI gene-set term is the sum of gene-set scores propagated through six PPI relations. Differences are in AP percentage points. Two-sided simultaneous 95% intervals for three comparisons are retained, with one-sided p-values reported.

Control	**AP**	**ΔAP (pp)**	**Simultaneous 95% interval (pp)**	**One-sided p**
$L_7 v$	0.919±0.028	+0.197	[-0.067,+0.470]	<0.05
$a + L_7 v$	0.920±0.028	+0.045	[-0.009,+0.099]	<0.05
PPI gene-set term + $L_7 v$	0.919±0.028	+0.199	[-0.065,+0.468]	<0.05

AP for the three variants ranges from 0.919 to 0.920, close to the full model. Full-model mean gains are 0.197, 0.045, and 0.199 percentage points, with one-sided $p < 0.05$ for all three variants; all two-sided simultaneous intervals cross zero. These comparisons characterize sensitivity to the seventh correction input. Section 4.2 tests the linear structural score; Section 4.3 tests node coefficients in the full model.

G.4 Prediction within Count Strata

Evaluation genes are independently divided into low, middle, and high groups using label-free tertiles of annotation count, total PPI degree, or observed numerical measurement count. Within each group, both predictors use exactly the same evaluation genes. For each repetition, out-of-fold predictions are concatenated and AP is computed within the group; the five AP values are then averaged.

Table G.2 retains all nine groups. TARGEN exceeds the count-feature predictor in every group and repetition. These descriptive comparisons show observed gains at different count levels; variables are stratified separately, and positive counts are also reported.

Table G.2: Descriptive AP within count strata. Values average AP computed from concatenated out-of-fold predictions separately in each of five repetitions. ΔAP is in AP percentage points. The count-feature control is the 11-dimensional logistic regression in Table I.3.

Stratification variable	**Group**	**Genes**	**Positives**	**TARGEN**	**Count model**	**ΔAP**
Annotation count	Low	995	27	0.475	0.055	+41.991
Annotation count	Middle	1008	141	0.708	0.536	+17.124
Annotation count	High	980	628	0.963	0.952	+1.130
PPI degree	Low	996	34	0.563	0.201	+36.217
PPI degree	Middle	996	249	0.862	0.784	+7.831
PPI degree	High	991	513	0.960	0.952	+0.763
Measurement count	Low	1070	34	0.435	0.327	+10.777
Measurement count	Middle	1048	185	0.733	0.573	+16.058
Measurement count	High	865	577	0.972	0.963	+0.976

G.5 Relation Weights and Correction Parameters

Table G.3: Sensitivity to relation weights and correction parameters. AP is the mean ± sample standard deviation over five repetitions of ten-fold cross-validation. ΔAP is consistently default model minus the current row, in AP percentage points, computed from unrounded values. Two-sided simultaneous 95% intervals retain the original Bonferroni quantiles. One-sided tests consistently assess higher AP for the default model.

Condition	**AP**	**ΔAP (pp)**	**Simultaneous 95% interval (pp)**	**One-sided p**
Default TARGEN	0.921±0.028	Reference	Reference	Reference
All corrections disabled ($\gamma = 0$)	0.919±0.029	Reference	Reference	Reference
PPI class weight 0.250	0.919±0.028	+0.211	[-0.026,+0.466]	<0.05
PPI class weight 0.750	0.921±0.028	+0.008	[-0.173,+0.179]	0.446
Propagation coefficient $g = 0.250$	0.919±0.029	+0.185	[+0.025,+0.349]	<0.05
Propagation coefficient $g = 1.000$	0.921±0.027	-0.035	[-0.248,+0.176]	0.666
Learned weights across PPI networks	0.920±0.028	+0.076	[-0.031,+0.192]	<0.05
CPDB correction disabled	0.920±0.028	+0.037	[-0.033,+0.113]	0.093
STRING correction disabled	0.920±0.028	+0.045	[-0.055,+0.144]	0.105
iRefIndex v9 correction disabled	0.921±0.028	+0.003	[-0.076,+0.095]	0.455
iRefIndex v15 correction disabled	0.921±0.028	-0.019	[-0.127,+0.088]	0.699
MULTINET correction disabled	0.920±0.028	+0.040	[-0.059,+0.133]	0.135
PCNet correction disabled	0.920±0.028	+0.089	[+0.021,+0.163]	<0.05
Gene-set correction disabled	0.918±0.029	+0.310	[-0.004,+0.646]	<0.05

The default propagation coefficient is 0.500, with PPI and gene-set relations each receiving half the class weight. Changes to class weights affect only genes with both PPI and gene-set relations; other genes retain default allocations. The learned-weight variant retains class allocations, learns positive weights for the six PPI networks, and renormalizes them over each gene's available networks. It uses five logit coordinates, initializes uniform weights, and applies an L2 penalty coefficient of 0.001. Weight variants recompute scale α by Equation (4); when learning weights, the objective differentiates through scale changes. For changes in the positive scalar g, this rescaling leaves $\alpha\mathcal{G}$ unchanged. Each individual correction ablation fixes one $\gamma_r = 0$, retains all relations in evidence propagation and structural scoring, and refits remaining parameters. All 12 variants repeat inner GeneRAIN selection and outer fitting, totaling 600 outer and 3,000 inner fits. The default and all-corrections-disabled references each reuse 50 fits. Intervals use 5,000 class-stratified gene-cluster bootstrap draws, sharing weights for each gene across repetitions and conditions while holding fitted predictions fixed. The two reference rows are excluded from the two comparison families above. Supplementary records are in `relation_sensitivity/`.

H External Replication in Patient-derived Organoids

H.1 Prescreened Candidates and Primary Results

We evaluate prescreened TARGEN candidates in public patient-derived tumor organoid data (Herranz-Ors et al., 2026). Prior dependency data identified four candidates among the top 100 unlabeled genes that were absent from the DepMap 25Q3 common-essential gene list. These four genes and two cancer contexts were fixed before inspecting candidate outcomes. After excluding documented model and patient overlap and requiring malignant histology, we include 146 organoids from 143 patients, including 78 patients with colorectal cancer and 49 with esophageal adenocarcinoma. We aggregate observations by patient. The following subsections describe screening and processing.

Organoids from 34 of 49 patients with esophageal adenocarcinoma show dependency on GRB2 (rank 98), with a median scaled log2 fold change (LFC) of -0.612 (Table 4). The support rule requires at least five patients, dependency in at least half, and median LFC below -0.500. GRB2 in esophageal adenocarcinoma meets the full support rule, providing external support for follow-up experiments on this candidate.

H.2 Defining the Replication Cohort by Patient Identity

We reanalyze official Supplementary Tables C.4, 4, 5, and 6 from Herranz-Ors et al. (2026), downloaded on September 19, 2026. Source data contain 162 screened organoids from 157 patients. Public sample identifiers, aliases, and Broad overlap flags identify 13 models overlapping the earlier NextGen resource. Excluding these models and all organoids sharing their patient identifiers removes 14 organoids, leaving 148 organoids from 144 patients.

We require the malignant morphology behavior code /3 from the International Classification of Diseases for Oncology (ICD-O). Removing one adenoma and one sample with unspecified malignant morphology leaves 146 organoids from 143 patients. Samples with missing cancer codes are retained if malignant morphology and primary tissue are known. Colorectal cancer and esophageal adenocarcinoma are the largest patient groups and define the primary analysis contexts. Identity checks cover documented overlap; undocumented cross-resource relationships remain uncertain.

H.3 One Effect and Dependency Call per Patient

We use author-provided BAGEL2 binary dependency calls (false discovery rate below 0.050) and published scaled log2 fold changes (LFCs). Nonessential controls are centered at zero and essential controls at -1; more negative values indicate greater growth impairment after knockout. These differ from the earlier Chronos scores and are analyzed separately.

For multiple libraries of the same quality-controlled model, we take median LFC and require a strict majority of libraries for a binary dependency call. Within each cancer type, the same aggregation is then applied to multiple organoids from one patient; ties yield no positive dependency call. Each gene's denominator includes patients with both binary and continuous measurements.

H.4 All Candidates Retained in Main and Sensitivity Analyses

Table 4 reports all eight primary comparisons; Table H.1 supplements their intervals. For GRB2 in esophageal adenocarcinoma, the Wilson 95% interval for the dependency proportion is [55.472%, 80.485%], and the patient-bootstrap interval for median LFC is $[-0.796, -0.532]$. Figures H.1 and H.2 retain all primary comparisons and patient observations.

Table H.2 retains three top-200 candidates beyond those in Table 4. KLF5 meets the support rule in both primary cancer types, and SRF meets it in colorectal cancer. These are descriptive top-200 results outside the four-gene testing family. Table H.3 retains coverage and support counts at each candidate-set size, including the top 50.

The binomial tests evaluate dependency in a majority of patients for selected candidates. Ranking enrichment, comparative prediction results, tumor selectivity, pharmacological response, and normal-tissue safety each require their own evidence.

Anonymous supplementary directory `organoid_replication/` contains all fixed top-200 results, model- and patient-level effects, cohort exclusions, essentiality flags, and portable records of the frozen protocol. Original protocol hashes are retained; portable copies remove local filesystem identity information.

Table H.1: Two-sided Wilson 95% confidence intervals for dependency proportions in all eight candidate comparisons.

Gene	Cancer	Wilson interval
XRCC2	CRC	[0.320, 0.534]
XRCC2	EAC	[0.178, 0.424]
CRKL	CRC	[0.130, 0.308]
CRKL	EAC	[0.178, 0.424]
PTK2	CRC	[0.173, 0.363]
PTK2	EAC	[0.452, 0.718]
GRB2	CRC	[0.250, 0.457]
GRB2	EAC	[0.555, 0.805]

Table H.2: Descriptive top-200 results for three prior candidates beyond Table 4. SRF, KLF5, and TRAF2 are outside the primary testing family.

Gene	TARGEN rank	Cancer	Dependent/tested	Median LFC
SRF	118	CRC	45/78	-0.655
SRF	118	EAC	27/49	-0.468
KLF5	130	CRC	56/78	-1.025
KLF5	130	EAC	32/49	-0.756
TRAF2	200	CRC	23/78	-0.410
TRAF2	200	EAC	19/49	-0.413

Table H.3: Candidate-set coverage and support. A gene is covered if measured in at least five patients in either primary cancer type, and supported if it meets the support rule in either type; support rates here are not filtered by p-values. Common-essential exclusions use DepMap 25Q3. Support rate is the fraction of measured genes meeting the support rule.

Set size K	Gene set	Fixed count	Measured	Coverage	Supported	Support rate
50	All	50	47	94.000%	0	0.000%
50	Excluding common-essential genes	49	46	93.878%	0	0.000%
100	All	100	96	96.000%	3	3.125%
100	Excluding common-essential genes	96	92	95.833%	1	1.087%
200	All	200	192	96.000%	9	4.688%
200	Excluding common-essential genes	189	183	96.825%	3	1.639%

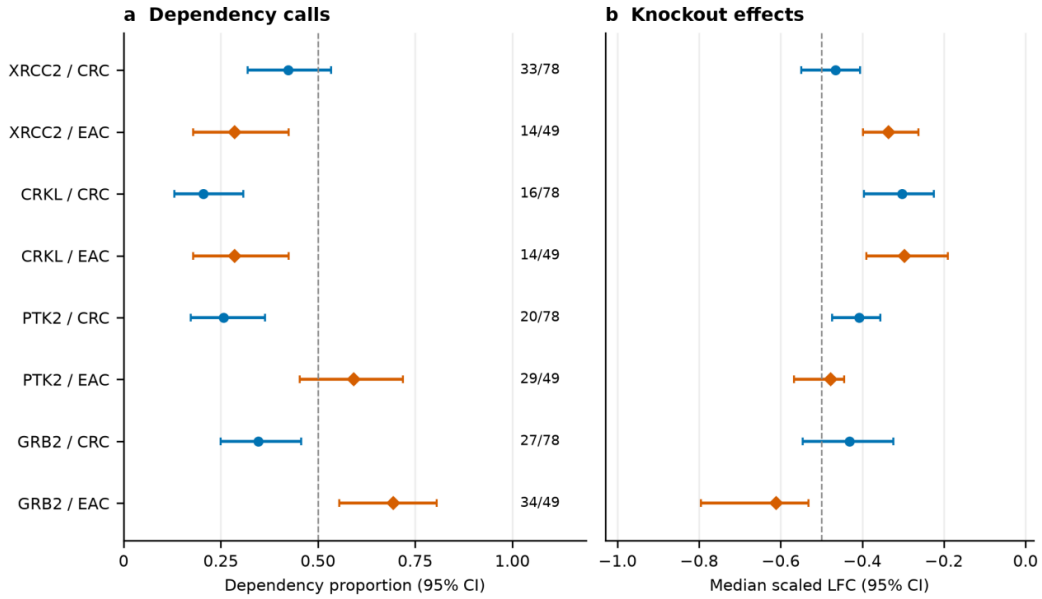


Figure H.1: Results for four prespecified top-100 candidates in colorectal cancer (78 patients, 80 organoids) and esophageal adenocarcinoma (49 patients, 49 organoids). Left: dependency proportions with two-sided 95% Wilson intervals, labeled as dependent/tested patients. Right: median scaled LFC with two-sided pointwise 95% intervals from 5,000 patient-bootstrap draws. Libraries and organoids are first aggregated within patients. Dashed lines mark analysis thresholds of proportion 0.500 and median LFC -0.500. All eight comparisons use one-sided exact binomial tests; only GRB2/EAC meets both the full support rule and one-sided $p < 0.05$. Tests assess within-cancer dependency proportions. Source: Herranz-Ors et al. (2026), analyzed as described above.

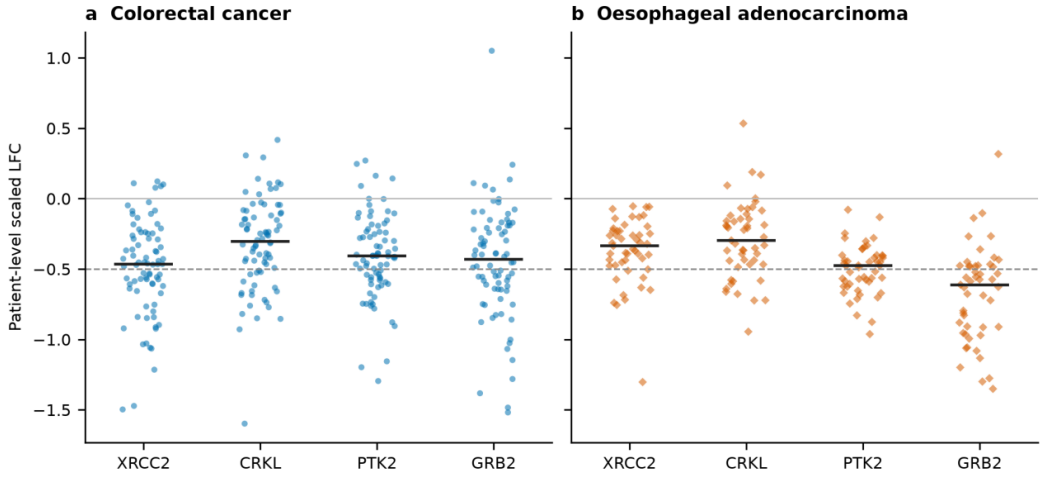


Figure H.2: Each point is a patient's aggregated scaled LFC. Solid lines indicate medians; the dashed line is -0.500. Each gene includes all 78 colorectal cancer and 49 esophageal adenocarcinoma patients. More negative values indicate greater growth impairment after knockout. Horizontal jitter only displays overlapping points and does not enter the analysis. All patients are retained, including effects inconsistent with dependency calls. Source: Herranz-Ors et al. (2026), analyzed as described above.

I Supplementary Model Comparisons and Score Controls

I.1 Biological-input and Score Controls

The full model outperforms direct molecular and gene-set scoring by 1.481 AP percentage points (Table I.1). Feature-only and node-coefficient-only scores provide references with narrower inputs. Removing numerical omics, C2&C5 knowledge, or GeneRAIN reduces mean AP, supporting their contributions within the evaluated configurations. GeneRAIN improves mean AP in every repetition; Appendix B also reports recall and F1.

Removing trainable gene-set scores while retaining functional relations lowers AP by 1.283 percentage points. This intervention removes gene-set scores from direct and downstream paths while retaining shared-membership relations, supporting learned gene-set scoring in addition to the use of relations. Joint C2 and C5 also outperform either alone; the current GE/METH magnitude representation outperforms the joint signed-input replacement (Table I.2).

TARGEN outperforms mutation-only predictors and the 11-variable predictor using PPI degrees, gene-set membership counts, and observed measurement counts (Table I.3). The gain of 3.174 percentage points over the count-feature model supports predictive information beyond these summaries. Appendix G gives fitting procedures and count-stratified results. These comparisons connect the joint model's prediction results to its molecular inputs and its two uses of functional annotations.

Table I.1: Joint scoring and biological-input controls. Full-model AP is 0.921 ± 0.028. Differences are full model minus control, in AP percentage points; interval construction follows Appendix D.3.

Control	AP	ΔAP (pp)	95% interval (pp)
Direct molecular and gene-set scores	0.906±0.032	+1.481	[+0.967,+1.979]
Feature-only score	0.869±0.033	+5.201	[+4.199,+6.157]
Node-coefficient-only score	0.879±0.030	+4.155	[+2.780,+5.464]
Without numerical omics	0.907±0.025	+1.363	[+0.435,+2.265]
Without C2&C5 knowledge	0.897±0.030	+2.371	[+1.777,+2.969]
Without gene-set scores	0.908±0.029	+1.283	[+0.855,+1.721]
Without GeneRAIN	0.917±0.029	+0.418	[+0.107,+0.732]

Table I.2: Functional gene sets and molecular representations. Differences are full model minus control, in AP percentage points. Two-sided simultaneous 95% intervals for three comparisons are retained, with one-sided p-values reported.

Control	AP	ΔAP (pp)	Simultaneous 95% interval (pp)	p
C2 only	0.915±0.028	+0.580	[+0.068,+1.088]	<0.05
C5 only	0.910±0.029	+1.052	[+0.616,+1.525]	<0.05
Signed GE/METH	0.913±0.028	+0.826	[+0.003,+1.654]	<0.05

Table I.3: Mutation-score and count-feature predictors. Differences are full model minus control, in AP percentage points. Two-sided simultaneous 95% intervals for four comparisons are retained, with one-sided p-values reported.

Control	AP	ΔAP (pp)	Simultaneous 95% interval (pp)	p
Mean mutation score	0.579±0.048	+34.212	[+30.146,+37.500]	<0.05
Maximum mutation score	0.540±0.044	+38.129	[+34.075,+41.486]	<0.05
Mutation-only logistic regression	0.698±0.041	+22.298	[+19.138,+25.252]	<0.05
Count-feature logistic regression	0.889±0.026	+3.174	[+1.625,+4.845]	<0.05

I.2 Availability Protocol

The 16 pan-cancer conditions in Section 4.2 cross numerical omics data retention of 100%, 75%, 50%, and 0% with training-label retention of 25%, 50%, 75%, and 100%. A fixed random sequence (seed 2026091801) generates nested masks over observed numerical omics entries: entries retained at 50% also remain at 75%. Retained entries keep their original values, while excluded entries contribute zero numerical input. Class-stratified training-label subsets use a fixed random ordering within each outer training fold and remain nested across retention levels. The outer test genes remain identical across all conditions. GeneRAIN and C2&C5 remain available; each fit rebuilds preprocessing and filters gene sets using only its retained training subset. Each new condition repeats inner model selection and outer fitting on the original five repetitions of ten-fold cross-validation.

Table I.4 reports the structural-score gain separately for all 16 conditions. The overall gain weights these conditions equally, and endpoint contrasts average over the four levels of the other dimension. Inference uses 5,000 paired gene-cluster bootstrap draws (seed 2026091802), sharing the same gene weights across conditions and repetitions. The intervals in Table I.4 are two-sided pointwise percentile intervals; the tests are one-sided. Saved records contain per-fold metrics, held-out predictions, bootstrap draws, pointwise and simultaneous intervals, and Holm-adjusted results for the 16 condition-specific comparisons. The PPI randomization protocol and results remain as described in Appendix F.

I.3 Numerical Structural-score Gains

Table I.4: Pan-cancer structural-score ΔAP (percentage points). Panel A averages 50 paired test folds. In B, the mean weights 16 conditions equally; endpoint contrasts average over the other dimension; the PPI contrast compares real relations with the mean of 30 rewired instances. Intervals are two-sided pointwise 95%; tests are one-sided.

A. Structural-score ΔAP across label and omics retention combinations

Numerical omics retained	**100% labels**	**75% labels**	**50% labels**	**25% labels**
100%	+0.310*	+0.378*	+0.419*	+0.569*
75%	+0.314*	+0.335*	+0.315*	+0.687*
50%	+0.438*	+0.473*	+0.426*	+0.569*
0%	+0.674*	+0.701*	+0.844*	+1.468*

TARGEN ΔAP = AP with Gs $-$ AP without it; *: one-sided $p < 0.05$ for higher AP with Gs.

B. Mean ΔAP and control contrasts

Contrast	**ΔAP**	**95% interval**	p
Mean ΔAP over the 16 conditions in A	+0.557	[+0.352,+0.756]	<0.05
Structural ΔAP: 0% $-$ 100% omics	+0.503	[+0.200,+0.798]	<0.05
Structural ΔAP: 25% $-$ 100% labels	+0.389	[+0.142,+0.629]	<0.05
AP: real PPI minus degree-preserving rewired PPI	+0.334	[+0.118,+0.564]	<0.05