MORGaN: self-supervised multi-relational graph learning for druggable gene discovery

Anonymous Author(s)
Affiliation
Address
email

Abstract

Accurate identification of druggable targets remains a critical challenge in drug discovery due to the inherent complexity of biology and the scarcity of labeled data. We present **MORGaN**, the first *masked auto-encoder* that natively operates on *heterogeneous* **m**ulti-**o**mic **g**ene **n**etworks with diverse biological relation types. MORGaN learns structure-aware node embeddings without supervision, leveraging multi-relation topology through a cross-relation message-passing architecture. We deploy MORGaN for **druggable gene discovery**, using its representations to identify candidate therapeutic targets. Despite using no additional labels, MORGaN outperforms state-of-the-art models across all metrics (AUPR: $0.815 \rightarrow 0.888$; $+9\%$). Ablation studies highlight the importance of both relation diversity and architectural design in achieving these gains. Post-hoc analyses uncover pathway-coherent subgraphs that help explain predictions, supporting biological interpretability. MORGaN enables label-efficient, interpretable, and *fast* graph learning for drug discovery and other data-scarce biomedical tasks. Code and documentation are available at this link.

1 Introduction

Drug discovery is complex and time-consuming, marked by costly financial investments and a high risk of failure [32]. Bringing a new drug to market can take over a decade and cost upwards of 2.6 billion USD, with failure rates remaining high across all stages of development [16, 32, 39]. A major contributor to this attrition is the inherent complexity in correctly identifying molecular targets whose modulation translates into clinical benefit [10]. Improving target selection is therefore critical to accelerating therapeutic development and reducing the associated costs [32, 39].

Cancer exemplifies this challenge, as tumorigenesis results from intricate, context-dependent interactions across genetic, epigenetic, and proteomic layers [4, 5, 7, 15, 37, 47]. Yet, conventional methods for identifying druggable genes – genes whose products can be therapeutically targeted – often rely on single data modalities or predefined pathways, overlooking crucial cross-layer interactions and tumor heterogeneity [7, 15, 37, 47]. Even network-based methods often use PPIs alone, overlooking diverse gene–gene relationships; models that jointly capture multiple biological relation types to provide a systems-level view are required [8, 36].

Graph neural networks (GNNs) naturally model such structure via message passing, merging network topology and features into context-aware embeddings [14, 22, 36, 41, 45]. Particularly, Relational Graph Convolutional Networks (RGCNs) extend GNNs to handle heterogeneous graphs with multiple edge types, making them ideally suited for multi-relational biological networks [34, 42]. However, applying these methods to druggable gene prediction remains challenging due to the scarcity and bias of labeled datasets [8].

Submitted to the AI for Science workshop (NeurIPS 2025). Do not distribute.

To address these challenges, we propose MORGaN, a self-supervised multi-relational graph learning framework that integrates multi-omic features with diverse biological relationships for robust druggable gene discovery (Figure 1). By leveraging masked feature reconstruction [8, 17], MORGaN captures underlying biological structures, exploiting both labeled and unlabeled information to produce generalizable gene embeddings. Consequently, MORGaN significantly enhances predictive accuracy, scalability, and biological interpretability compared to existing methods.

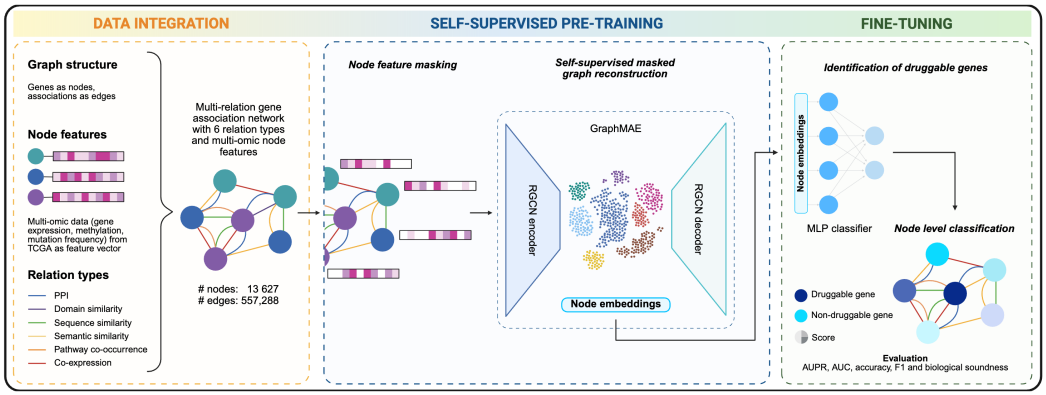


Figure 1: **MORGaN overview.** *Graph construction and data integration*: build a six-relation, multi-omic gene network. *Self-supervised pre-training*: a RGCN-based masked autoencoder (GraphMAE) reconstructs the missing features and generates node embeddings. *Fine-tuning*: an MLP uses these embeddings to classify druggable genes, evaluated with AUPR, AUROC, accuracy and F1.

1.1 Contributions

1. **Masked auto-encoding for heterogeneous graphs.** To our knowledge, MORGaN introduces the first MAE that operates on *multi-relation, multi-omics* graphs, unifying six biological edge types in a single self-supervised objective.
2. **State-of-the-art performance.** MORGaN surpasses current SOTA across key metrics (AUPR $+9\%$, AUROC $+3\%$, Accuracy $+15\%$ and F1 $+13\%$).
3. **Ultra-lightweight kernel.** A vertically-stacked sparse matrix [42] and basis decomposition [34] collapse R separate message-passing steps into one sparse-dense matrix multiplication, cutting per-epoch time by $\sim 80\%$.
4. **Plug-and-play extensibility.** The model is disease-agnostic; re-targeting only requires new node features and labels. We validate disease- and task-level transfer in out-of-distribution studies; see Appendix I.
5. **Out-of-the-box interpretability.** GNNExplainer sub-graphs translate each prediction into a minimal set of genes – providing falsifiable hypotheses for experimental validation.

2 Related work

Multi-omic GNNs for cancer biology. Early work showed that fusing several omics layers with network structure helps uncover oncogenic mechanisms. Schulte-Sasse et al. [35] integrated mutation, expression and copy-number profiles with a GCN to pinpoint novel cancer genes and their pathways. More recently, MOGAT used graph attention over concatenated omics views to refine cancer-subtype labels, confirming that attention can weight modalities adaptively [41]. SMG added self-supervised masking to a PPI graph, improving essential- and cancer-gene recall under scarce labels [8]. IMI-driver extended this idea by stacking distinct functional networks (PPI, co-expression, co-methylation, etc.) and training a multi-graph GCN to prioritize tumor drivers across 33 TCGA cohorts, outperforming feature-only baselines by 15–20% AUROC [37]. MODIG [49] and the follow-up MDMNI-DGD [23] generalized from single PPIs to five- and six-edge-type multiplex graphs; per-layer attention and view-level fusion lifted driver- and druggable-gene AUPR by 6–10% over PPI-only models. Yet, all of these models train each edge type in isolation or require heavy message-passing loops, limiting scalability.

Self-supervised and masked graph learning. Generic graph SSL replaces expensive node labels with pretext tasks. Masked autoencoders such as GraphMAE reconstruct hidden node attributes and beat contrastive objectives on 21 benchmarks [17]; SMG applies the same principle to cancer PPIs [8]. Orthogonal work explores contrastive signals: HeCo co-contrasts schema- and meta-path views on heterogeneous graphs to lift performance without labels [25], while SpaMask couples masking and contrastive heads for spatial-omics clustering [27].

3 MORGaN

MORGaN combines heterogeneous biological knowledge with graph representation learning to predict druggable genes in three stages (Fig. 1):

1. **Graph construction and data integration** – we construct a heterogeneous, multi-relational gene graph that fuses six biological interaction types with multi-omic node features;
2. **Self-supervised representation learning** – a relational graph convolution masked autoencoder distills structure-aware embeddings from the unlabeled graph;
3. **Task-specific fine-tuning and interpretation** – the embeddings are fine-tuned to identify druggable genes, and post-hoc explainers expose the components that drive each prediction.

This design unites broad biological priors with graph deep learning to deliver accurate, transparent, and computationally efficient predictions.

3.1 Graph construction and data integration

We represent the gene interaction landscape as a heterogeneous, multi-relational graph $G = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where each node $v_i \in \mathcal{V}$ represents a gene. For each biological relationship type $r \in \mathcal{R}$, we define a relation-specific edge set $\mathcal{E}_r \subseteq \mathcal{V} \times \mathcal{V}$ and we define the full graph as the union over all relations $\mathcal{E} = \bigcup_{r \in \mathcal{R}} \mathcal{E}_r$. The pan-cancer graph contains 13 627 genes and 557 288 edges across six relation types. For further details, see Appendix A.

Relations. Following MDMNI-DGD [23], we incorporate six biologically grounded relation types, based on protein-protein interaction networks (PPI), gene co-expression, pathway co-occurrence, gene ontology semantic similarity, and sequence similarity. Self-loops are added to preserve each gene's own features during message passing. For further details, see Appendix A.2.1.

Node features. Each gene node v_i is associated with a multi-omic feature vector x_i, obtained by concatenating log10-transformed somatic mutation frequencies, copy number alteration (CNA) z-scores, DNA methylation β-values, and log-normalized gene expression values. All features are extracted from The Cancer Genome Atlas (TCGA) pan-cancer dataset [8, 43], spanning 29,446 tumor samples across 16 cancer types, as in SMG [8]. For further details, see Appendix A.2.2.

Labels. Positive druggability labels correspond to Tier 1 targets defined by Finan et al. [10], i.e. proteins with approved drugs or clinical candidates; an equal number of negatives is randomly sampled from the remaining non-target genes to balance class distributions. See Appendix A.2.3.

3.2 Model architecture

We adopt the relational graph convolutional network (RGCN) of Schlichtkrull et al. [34] and convert it into a *graph masked autoencoder* (GraphMAE) [17].

Message passing formulation. For layer l the hidden state of node v_i is updated via:

$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right), \tag{1}$$

where $h_i^{(l)}$ represents the hidden state of node v_i at layer l, $\mathcal{N}_i^r$ is the set of i's neighbors under relation r, $W_r^{(l)}$ and $W_0^{(l)}$ are trainable relation-specific and self-loop weight matrices, respectively, and $c_{i,r}$ is a normalization constant to ensure numerical stability.

Vertical stacking for sparse message passing. To exploit the fast sparse–dense multiplication (`spmm`) available in PyTorch while still updating all relation types at once, we concatenate the R relation-specific adjacency matrices $\{A_r\}_{r=1}^{R}$ *vertically* into a single sparse block matrix $A_v \in \mathbb{R}^{(RN)\times N}$, as introduced by Thanapalasingam et al. [42]. During each RGCN layer, we first mix topological and feature information with one call to $\texttt{spmm}(A_v, X)$, producing a relation-expanded feature matrix of shape $(RN) \times d_{in}$. This matrix is then reshaped back to $N \times (R\, d_{in})$ and multiplied by a stacked weight matrix to yield the next-layer embeddings. Because the projection to higher dimensions happens *after* the sparse multiplication, vertical stacking keeps memory usage low and scales well to large graphs with modest input dimensionality.

Weight decomposition. To manage parameter complexity with multiple relation types, we implement basis decomposition [34]. Each relation-specific matrix is expressed as a linear combination of a shared set of B basis matrices $\{V_b\}_{b=1}^{B}$:

$$W_r = \sum_{b=1}^{B} a_{rb} V_b, \tag{2}$$

where $V_b \in \mathbb{R}^{d_{\text{in}} \times d_{\text{out}}}$ are global basis matrices shared across all relations, and $a_{rb} \in \mathbb{R}$ are relation-specific learnable coefficients. This formulation significantly reduces parameter count compared to using unique weights per relation, while preserving expressiveness through learned compositions. In our implementation, we set $B = 2$ to strike a balance between model flexibility and generalization capacity.

Normalization and dropout. Each layer applies layer normalization to the concatenated relation outputs, adds a residual connection, and then dropout ($p = 0.2$).

Implementation details. All models are implemented in PyTorch 2.6.0 [31] and PyTorch-Geometric 2.6.1 [9]. Relation weights W_r, bases V_b, and coefficients $a_{r,b}$ use Xavier uniform (gain $= \sqrt{2}$ for PReLU) initialization; the self-loop matrix W_0 uses Kaiming initialization. We fix random seeds (Python, NumPy, PyTorch, PyG) to 0 and 1 and report mean $\pm$ std over 3 runs per seed.

3.3 Training

We adopt a two-phase training strategy with Adam [21].

Self-supervised pre-training. Following GraphMAE [17], we randomly mask 50% of node features and reconstruct them using a scaled cosine loss:

$$\mathcal{L}_{\text{rec}} = \frac{1}{|\widetilde{\mathcal{V}}|} \sum_{v_i \in \widetilde{\mathcal{V}}} \left(1 - \frac{x_i^T \hat{x}_i}{|x_i||\hat{x}_i|}\right)^{\gamma}, \quad \gamma \geq 1, \tag{3}$$

where x_i and $\hat{x}_i$ denote the original and reconstructed feature vectors, respectively, and γ controls the loss sharpness. Pre-training runs for 100 epochs with an initial learning rate 10^{-2}, weight decay 10^{-3}, cosine decay 10^{-6}, $\gamma = 3$, and early stopping (patience 10). Hyper-parameters were selected via a grid sweep (see Appendix C); the best configuration is used throughout the paper. Through this pre-training stage, the model learns compressed embeddings that encode both multi-omic profiles and relational context, serving as a robust foundation for downstream classification.

Fine-tuning (supervised). The frozen embeddings feed an MLP classifier optimized with weighted binary cross-entropy:

$$\mathcal{L}_{\text{class}} = -w \left[y \log \sigma(p) + (1 - y) \log(1 - \sigma(p))\right], \tag{4}$$

with label-dependent weights w to handle class imbalance. We train for up to 200 epochs (learning rate 5×10^{-3}, weight decay 10^{-4}, gradient-clip 1.0) with early stopping (patience 20) on validation AUPR. Hyper-parameters were selected in the same sweep used for pre-training (see Appendix C).

3.4 Experimental setup and evaluation

Repeated shuffle-split validation. We generate two independent, stratified train/validation/test splits (80% / 10% / 10% of nodes) using different random seeds. Each split is trained three times with different weight initializations, giving six runs in total. We report mean $\pm$ s.d. of AUPR, AUROC, Accuracy, and F1 across these runs.

Baseline models. We benchmark MORGaN against eight alternatives that span feature-only, homogeneous-graph and heterogeneous-graph approaches:

1. *Logistic Regression* – feature-only
2. *Multilayer Perceptron* (MLP) – feature-only)
3. *GCN* – vanilla graph convolution on a 1-dimensional PPI graph
4. *GAT* – graph attention network on the PPI graph
5. *SMG-GCN* [8] – GCN with self-supervised pre-training on the PPI graph
6. *SMG-GAT* [8] – GAT with self-supervised pre-training on the PPI graph
7. *MODIG* [49] – heterogeneous graph model without pre-training
8. *MDMNI-DGD* [23] – heterogeneous graph model without pre-training

The two *feature-only* models use the concatenated multi-omic vectors. The four *homogeneous* baselines (3-6) operate on a *single-relation* PPI graph and therefore lack the multi-relational context exploited by MORGaN. The two *heterogeneous* baselines (7-8) share the full multi-relational topology with MORGaN but do not include its self-supervised pre-training stage. All models receive identical node features and use the same train/validation/test splits; hyper-parameters are selected by grid search on the validation fold. See Appendix B for further details.

This design cleanly isolates MORGaN's architectural and training contributions while ensuring a fair, rigorously repeated comparison to both feature-based and graph-based alternatives.

Interpretability. For high-confidence predictions ($p > 0.9$), we use GNNExplainer [48] to highlight the network edges and gene features that most drive each call. We then use Enrichr to run *pathway enrichment analysis* – a simple check of whether those highlighted genes occur together in well-known biological pathways more often than expected by chance – reporting pathways that pass a false discovery rate (FDR) threshold of < 0.05.

4 Results

4.1 Overall predictive performance

Table 1 summarizes mean performance $\pm$ s.d. over six stratified shuffle–split runs (see §3.4) of MORGaN against eight alternative approaches. MORGaN tops every metric, pushing the state of the art into the 0.9-range on almost every curve–based measure:

- **Precision–recall.** MORGaN achieves 0.888 ± 0.004, an absolute gain of $+0.073$ (+9.0 %) over the strongest competitor (MDMNI-DGD, 0.815 ± 0.019). The gap exceeds the largest baseline standard deviation (0.052 for MLP) by more than 40 %, underscoring significance.
- **Discrimination ability (AUROC).** Our framework reaches 0.907 ± 0.005, outperforming the next best model by $+0.030$ (+3.4 %). Even methods that exploit the *same* multi-relational graph but skip self-supervised pre-training (MODIG, MDMNI-DGD) lag behind.
- **Balanced classification (F1 and accuracy).** MORGaN's F1$= 0.917$ and Acc$= 0.915$ translate to absolute gains of $10 - 12$ percentage points (pp) over the closest challengers, reflecting superior recall without sacrificing precision.

Table 1: **Test-set performance of MORGaN versus eight baselines** on the druggable-gene prediction task (mean ± s.d.). **Bold** numbers indicate the best score per column; *italic* numbers mark the second best. All models receive the same multi-omic node features; heterogeneous methods (bottom block) also share the identical six-relation graph.

Model	AUPR	AUROC	Accuracy	F1 Score
Logistic regression	0.749 ± 0.045	0.682 ± 0.055	0.620 ± 0.944	0.577 ± 0.096
MLP	0.675 ± 0.052	0.722 ± 0.045	0.722 ± 0.045	0.701 ± 0.035
GCN	0.721 ± 0.020	0.766 ± 0.005	0.715 ± 0.025	0.722 ± 0.037
GAT	0.699 ± 0.005	0.764 ± 0.009	0.724 ± 0.015	0.742 ± 0.005
SMG-GCN [8]	0.714 ± 0.009	0.763 ± 0.011	0.729 ± 0.029	0.724 ± 0.014
SMG-GAT [8]	0.708 ± 0.005	0.776 ± 0.005	0.732 ± 0.027	0.751 ± 0.014
MODIG [49]	0.764 ± 0.017	0.837 ± 0.015	0.794 ± 0.009	0.810 ± 0.007
MDMNI-DGD [23]	0.815 ± 0.019	0.877 ± 0.003	0.664 ± 0.038	0.741 ± 0.022
MORGaN (no pre-training)	$\mathit{0.879 \pm 0.006}$	$\mathit{0.900 \pm 0.006}$	$\mathit{0.898 \pm 0.006}$	$\mathit{0.902 \pm 0.006}$
MORGaN (with pre-training)	$\mathbf{0.888 \pm 0.004}$	$\mathbf{0.907 \pm 0.005}$	$\mathbf{0.915 \pm 0.005}$	$\mathbf{0.917 \pm 0.004}$

Why does MORGaN win? A head-to-head inside each baseline family points to two factors:

1. *Heterogeneous edges matter.* Substituting the single-edge view of GAT with the full six-relation interactome already yields a strong lift (AUPR $+0.065$ from GAT to MODIG). This gain confirms that druggability signals are not confined to one molecular relationship but are dispersed across many.
2. *Architecture matters more.* MORGaN does not just stack relation-specific layers; its encoder processes every edge type in one coherent pass, allowing information to flow *between* relations inside each layer. This cross-relation coupling unlocks another sizable margin over the best heterogeneous competitor (AUPR $+0.073$ versus MDMNI-DGD) and leaves even the ablated MORGaN (no pre-training) far ahead.

Together, these observations show that (**i**) embracing the full diversity of biological interactions and (**ii**) employing an architecture specifically designed to fuse those interactions in-layer are both necessary – and mutually reinforcing – for state-of-the-art druggable-gene discovery.

We further test robustness under both distribution and task shift: (i) disease shift by re-training MORGaN on an Alzheimer's disease (AD) graph built analogously to the cancer setting; and (ii) task shift, by applying the framework to essential-gene ranking. We observe qualitatively consistent trends under both types of shift; see Appendix I for full protocols and results.

4.2 Ablation on biological relations and omics features

We next asked which inputs matter most to MORGaN's performance on the cancer dataset. Three complementary ablations were performed (full tables in Appendix E).

Edge-type "drop-one" study. Starting from the six-relation graph, we removed one edge type at a time and re-trained the full pipeline. Figure 2 (blue boxes) plots the resulting AUPR distributions; the line at $n = 6$ is the unablated model. Removing *GO semantic similarity* produces the largest drop ($0.888 \rightarrow 0.878$, -1.0 pp). PPI and pathway co-occurrence are nearly as important (-0.9 pp each), whereas sequence similarity and co-expression are slightly noisy: cutting them *improves* AUPR to 0.900 and 0.899. Notably, every five-relation variant still beats the best single-relation model (GO only, 0.883), confirming the benefit of diverse biological context. Robustness checks across five popular PPI databases (Appendix Table 4) show that STRING-db gives the strongest lift (AUPR $= 0.971$), but MORGaN consistently outperforms baselines regardless of the underlying interactome.

Randomized edge controls. To verify that gains arise from genuine biology rather than extra parameters, we degree-preserved-shuffled each edge type and repeated the experiment (grey boxes in Fig. 2). AUPR collapsed to ~ 0.5 – close to random guessing – for every configuration, demonstrating that MORGaN relies on meaningful topology, not mere edge density.

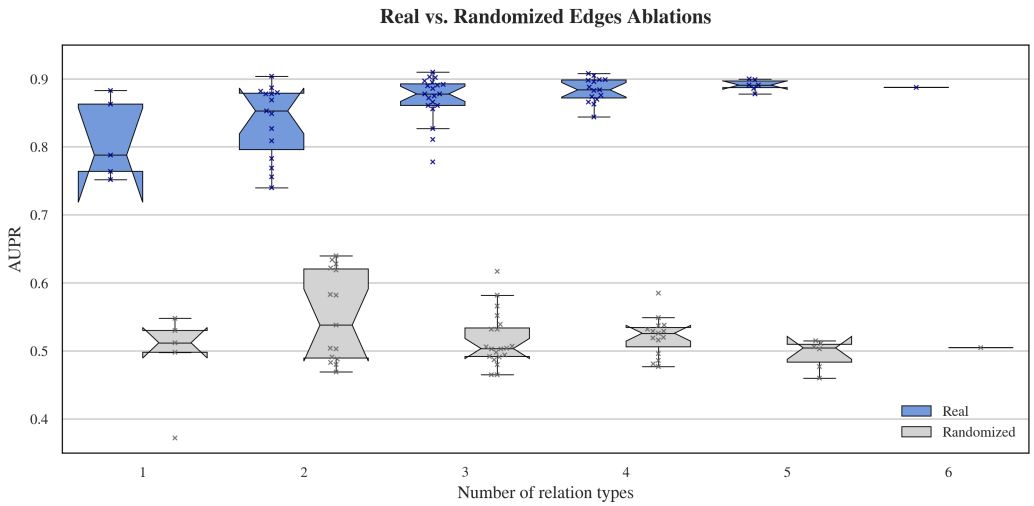


Figure 2: For every subset size n we plot AUPR over all combinations of real edges (blue) and their randomized counterparts (grey). The "box" at $n = 6$ reduces to a thin line because only one configuration – the full graph – exists.

Omics feature ablation. A similar "leave-one-modality-out" analysis suggests that *copy-number alterations (CNA)* are the most informative modality: achieving the highest performance when used in isolation. Gene expression, by contrast, was the noisiest – its removal *increased* AUPR by +0.009, consistent with evidence that bulk transcriptomics yields limited signal strength, given the many measured features but relatively few samples [1]. Nevertheless, the full four-omics model offered the best overall trade-off, achieving the highest AUROC (0.907) and second-highest AUPR (see Appendix E.2).

Takeaways. (1) Semantic proximity in Gene Ontology terms supplies critical long-range cues that cannot be captured by PPI edges alone. (2) Multi-relation structure is genuinely informative – performance falls to chance once biology is scrambled. (3) Results are stable across alternative PPI layers, with STRING-db performing best but all sources sustaining MORGaN's edge. (4) While CNA dominates in isolation, combining complementary omics restores false negatives that CNA alone misses, yielding the best F1. These findings validate the design choices behind MORGaN.

4.3 Biological interpretability and case studies

MORGaN rediscovers established targets and proposes novel candidates. MORGaN correctly retrieves hallmark cancer-druggable genes such as *EGFR*, *HER2*, *BRAF*, *ALK*, *MET*, and *RET*. Beyond these, the model flags *954 additional genes* with posterior probability $p > 0.9$. Many prioritized genes substantially overlap with independent resources (see Appendix H), supporting the reliability of MORGaN's predictions. Furthermore, pathway enrichment, literature mining, and local explanations (below) indicate that many of these are, in fact, promising putative targets.

Latent structure mirrors biology. In the learned embedding space, positives cluster distinctly from non-targets (Fig. 3). This separation indicates that the self-supervised pre-training has distilled a biologically meaningful representation in which druggable genes occupy a distinct region of latent space.

Pathway-level evidence and mechanistic explanations support predictions. Predicted positives are enriched for receptor-mediated signaling and canonical cancer hallmarks (FDR< 0.05; see Appendix F). GNNExplainer recovers pathway-coherent subgraphs that support these predictions (see Appendix G).

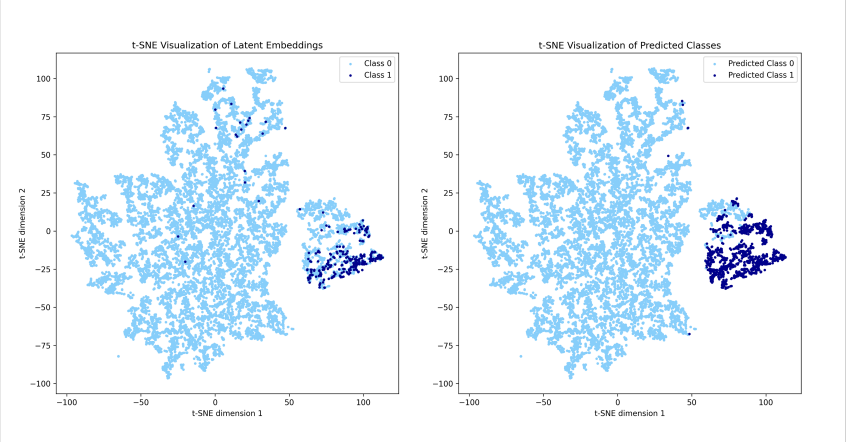


Figure 3: t-SNE visualization of MORGaN embeddings. Each point is a gene, colored by ground-truth class (left) or model prediction (right). The learned latent space clearly separates druggable (dark blue) from non-druggable (light blue) genes.

Combining global evidence (hallmark-pathway enrichment) with local mechanistic explanations shows that MORGaN both rediscovers established targets and delivers a set of 954 high-confidence, biologically coherent putative targets for experimental validation.

4.4 Computational efficiency

Thanks to vertically stacked sparse message passing and basis-matrix weight decomposition (§3.2), MORGaN trains in 24.3 ± 2.9 s end-to-end – about $65\times$ and $23\times$ faster than MODIG and MDMNI-DGD, respectively; full hardware and per-stage breakdown are in Appendix D.

5 Limitations and future work

This study relies on a *static* interaction graph that merges data across tissues and time; it therefore misses the dynamic rewiring that accompanies disease progression. Moreover, our data is derived from public resources that may over-represent well-studied genes and common cancers, which can introduce dataset bias and obscure disease-driven network rewiring. However, we choose TCGA because it provides breadth, standardization, and reproducibility. In addition, we treat druggability as a binary label, whereas in practice it lies on a spectrum shaped by modality, structure, and clinical tractability; future versions could adopt multi-label or continuous endpoints. Nonetheless, binary labels mirror industrial practice – at a given point in time, for a specific therapeutic class (here, small-molecule drugs), targets are operationally treated as either druggable or not.

6 Conclusion

This work introduces MORGAN, a self-supervised multi-relational graph framework that integrates six biological edge types with four omics modalities to prioritize druggable genes in cancer. By combining relational structure with masked feature reconstruction, MORGaN bridges two gaps in target discovery: (i) the siloing of heterogeneous interaction data and (ii) the scarcity of high-quality labels. On a pan-cancer benchmark, MORGaN beats eight strong baselines and re-discovers hallmark pathways, without sacrificing efficiency. Masked pre-training cuts label dependence, efficiency enables large graph sweeps, and explainer subgraphs provide falsifiable mechanisms. Being disease-agnostic, MORGAN transfers to other pipelines with minimal changes (Appendix I). Code and configs are available on GitHub.

References

[1] P. Angerer, L. Simon, S. Tritschler, F. A. Wolf, D. Fischer, and F. J. Theis. Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology*, 4:85–91, Aug. 2017. ISSN 2452-3100. doi: 10.1016/j.coisb.2017.07.004. URL https://www.sciencedirect.com/science/article/pii/S245231001730077X.

[2] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, May 2000. ISSN 1546-1718. doi: 10.1038/75556. URL https://www.nature.com/articles/ng0500_25. Publisher: Nature Publishing Group.

[3] S. Branciamore, G. Gogoshin, A. S. Rodin, and A. J. Myers. Changes in expression of VGF, SPECC1L, HLA-DRA and RANBP3L act with APOE E4 to alter risk for late onset Alzheimer's disease. *Scientific Reports*, 14(1):14954, June 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-65010-7. URL https://www.nature.com/articles/s41598-024-65010-7. Publisher: Nature Publishing Group.

[4] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, 127(16):3029–3030, 2021. ISSN 1097-0142. doi: 10.1002/cncr.33587. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.33587. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncr.33587.

[5] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024. ISSN 1542-4863. doi: 10.3322/caac.21834.

[6] M. Cannon, J. Stevenson, K. Stahl, R. Basu, A. Coffman, S. Kiwala, J. McMichael, K. Kuzma, D. Morrissey, K. Cotto, E. Mardis, O. Griffith, M. Griffith, and A. Wagner. DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Research*, 52(D1):D1227–D1235, Jan. 2024. ISSN 0305-1048. doi: 10.1093/nar/gkad1040. URL https://doi.org/10.1093/nar/gkad1040.

[7] L. Chin, J. N. Andersen, and P. A. Futreal. Cancer genomics: from discovery science to personalized medicine. *Nature Medicine*, 17(3):297–303, Mar. 2011. ISSN 1546-170X. doi: 10.1038/nm.2323. URL https://www.nature.com/articles/nm.2323. Publisher: Nature Publishing Group.

[8] Y. Cui, Z. Wang, X. Wang, Y. Zhang, Y. Zhang, T. Pan, Z. Zhang, S. Li, Y. Guo, T. Akutsu, and J. Song. SMG: self-supervised masked graph learning for cancer gene identification. *Briefings in Bioinformatics*, 24(6):bbad406, Nov. 2023. ISSN 1477-4054. doi: 10.1093/bib/bbad406. URL https://doi.org/10.1093/bib/bbad406.

[9] M. Fey and J. E. Lenssen. Fast Graph Representation Learning with PyTorch Geometric, Apr. 2019. URL http://arxiv.org/abs/1903.02428. arXiv:1903.02428 [cs].

[10] C. Finan, A. Gaulton, F. A. Kruger, R. T. Lumbers, T. Shah, J. Engmann, L. Galver, R. Kelley, A. Karlsson, R. Santos, J. P. Overington, A. D. Hingorani, and J. P. Casas. The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine*, 9(383):eaag1166, Mar. 2017. doi: 10.1126/scitranslmed.aag1166. URL https://www.science.org/doi/10.1126/scitranslmed.aag1166. Publisher: American Association for the Advancement of Science.

[11] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta. Pfam: the protein families database. *Nucleic Acids Research*, 42(Database issue):D222–D230, Jan. 2014. ISSN 0305-1048. doi: 10.1093/nar/gkt1223. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965110/.

[12] M. I. Gabitto, K. J. Travaglini, V. M. Rachleff, E. S. Kaplan, B. Long, J. Ariza, Y. Ding, J. T. Mahoney, N. Dee, J. Goldy, E. J. Melief, A. Agrawal, O. Kana, X. Zhen, S. T. Barlow, K. Brouner, J. Campos, J. Campos, A. J. Carr, T. Casper, R. Chakrabarty, M. Clark, J. Cool, R. Dalley, M. Darvas, S.-L. Ding, T. Dolbeare, T. Egdorf, L. Esposito, R. Ferrer, L. E. Fleckenstein, R. Gala, A. Gary, E. Gelfand, J. Gloe, N. Guilford, J. Guzman, D. Hirschstein, W. Ho, M. Hupp, T. Jarsky, N. Johansen, B. E. Kalmbach, L. M. Keene, S. Khawand, M. D. Kilgore, A. Kirkland, M. Kunst, B. R. Lee, M. Leytze, C. L. Mac Donald, J. Malone, Z. Maltzer, N. Martin, R. McCue, D. McMillen, G. Mena, E. Meyerdierks, K. P. Meyers, T. Mollenkopf, M. Montine, A. L. Nolan, J. K. Nyhus, P. A. Olsen, M. Pacleb, C. M. Pagan, N. Peña, T. Pham, C. A. Pom, N. Postupna, C. Rimorin, A. Ruiz, G. A. Saldi, A. M. Schantz, N. V. Shapovalova, S. A. Sorensen, B. Staats, M. Sullivan, S. M. Sunkin, C. Thompson, M. Tieu, J. T. Ting, A. Torkelson, T. Tran, N. J. Valera Cuevas, S. Walling-Bell, M.-Q. Wang, J. Waters, A. M. Wilson, M. Xiao, D. Haynor, N. M. Gatto, S. Jayadev, S. Mufti, L. Ng, S. Mukherjee, P. K. Crane, C. S. Latimer, B. P. Levi, K. A. Smith, J. L. Close, J. A. Miller, R. D. Hodge, E. B. Larson, T. J. Grabowski, M. Hawrylycz, C. D. Keene, and E. S. Lein. Integrated multimodal cell atlas of Alzheimer's disease. *Nature Neuroscience*, 27(12):2366–2383, Dec. 2024. ISSN 1546-1726. doi: 10.1038/s41593-024-01774-5. URL https://www.nature.com/articles/s41593-024-01774-5. Publisher: Nature Publishing Group.

[13] I. Granata, L. Maddalena, M. Manzo, M. R. Guarracino, and M. Giordano. HELP: A computational framework for labelling and predicting human common and context-specific essential genes. *PLOS Computational Biology*, 20(9):e1012076, Sept. 2024. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012076. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012076. Publisher: Public Library of Science.

[14] W. L. Hamilton, R. Ying, and J. Leskovec. Inductive Representation Learning on Large Graphs, Sept. 2018. URL http://arxiv.org/abs/1706.02216. arXiv:1706.02216 [cs].

[15] Y. J. Heo, C. Hwa, G.-H. Lee, J.-M. Park, and J.-Y. An. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes. *Molecules and Cells*, 44 (7):433–443, July 2021. ISSN 1016-8478. doi: 10.14348/molcells.2021.0042. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8334347/.

[16] I. V. Hinkson, B. Madej, and E. A. Stahlberg. Accelerating Therapeutics for Opportunities in Medicine: A Paradigm Shift in Drug Discovery. *Frontiers in Pharmacology*, 11:770, June 2020. ISSN 1663-9812. doi: 10.3389/fphar.2020.00770. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7339658/.

[17] Z. Hou, X. Liu, Y. Cen, Y. Dong, H. Yang, C. Wang, and J. Tang. GraphMAE: Self-Supervised Masked Graph Autoencoders, July 2022. URL http://arxiv.org/abs/2205.10803. arXiv:2205.10803 [cs].

[18] J. K. Huang, D. E. Carlin, M. K. Yu, W. Zhang, J. F. Kreisberg, P. Tamayo, and T. Ideker. Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Systems*, 6(4):484–495.e5, Apr. 2018. ISSN 2405-4712, 2405-4720. doi: 10.1016/j.cels.2018.03.001. URL https://www.cell.com/cell-systems/abstract/S2405-4712(18)30095-4. Publisher: Elsevier.

[19] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, Jan. 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.27. URL https://doi.org/10.1093/nar/28.1.27.

[20] M. Kanehisa, M. Furumichi, Y. Sato, Y. Matsuura, and M. Ishiguro-Watanabe. KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*, 53(D1):D672–D677, Jan. 2025. ISSN 1362-4962. doi: 10.1093/nar/gkae909. URL https://doi.org/10.1093/nar/gkae909.

[21] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization, Jan. 2017. URL http://arxiv.org/abs/1412.6980. arXiv:1412.6980 [cs].

[22] T. N. Kipf and M. Welling. Semi-Supervised Classification with Graph Convolutional Networks, Feb. 2017. URL http://arxiv.org/abs/1609.02907. arXiv:1609.02907 [cs].

[23] J. Li, B. Li, X. Zhang, X. Ma, and Z. Li. MDMNI-DGD: A novel graph neural network approach for druggable gene discovery based on the integration of multi-omics data and the multi-view network. *Computers in Biology and Medicine*, 185:109511, Feb. 2025. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2024.109511. URL https://www.sciencedirect.com/science/article/pii/S0010482524015968.

[24] K.-H. Lim, S. Yang, S.-H. Kim, and J.-Y. Joo. Identifying New COVID-19 Receptor Neuropilin-1 in Severe Alzheimer's Disease Patients Group Brain Using Genome-Wide Association Study Approach. *Frontiers in Genetics*, 12, Oct. 2021. ISSN 1664-8021. doi: 10.3389/fgene.2021.741175. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.741175/full. Publisher: Frontiers.

[25] N. Liu, X. Wang, H. Han, and C. Shi. Hierarchical Contrastive Learning Enhanced Heterogeneous Graph Neural Network, Mar. 2024. URL http://arxiv.org/abs/2304.12228. arXiv:2304.12228 [cs].

[26] W.-C. Lo, C.-C. Lee, C.-Y. Lee, and P.-C. Lyu. CPDB: a database of circular permutation in proteins. *Nucleic Acids Research*, 37(suppl_1):D328–D332, Jan. 2009. ISSN 0305-1048. doi: 10.1093/nar/gkn679. URL https://doi.org/10.1093/nar/gkn679.

[27] W. Min, D. Fang, J. Chen, and S. Zhang. SpaMask: Dual masking graph autoencoder with contrastive learning for spatial transcriptomics. *PLOS Computational Biology*, 21(4):e1012881, Apr. 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012881. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012881. Publisher: Public Library of Science.

[28] D. Paes, R. Lardenoije, R. M. Carollo, J. A. Y. Roubroeks, M. Schepers, P. Coleman, D. Mastroeni, E. Delvaux, E. Pishva, K. Lunnon, T. Vanmierlo, D. van den Hove, and J. Prickaerts. Increased isoform-specific phosphodiesterase 4D expression is associated with pathology and cognitive impairment in Alzheimer's disease. *Neurobiology of Aging*, 97:56–64, Jan. 2021. ISSN 1558-1497. doi: 10.1016/j.neurobiolaging.2020.10.004.

[29] P.-C. Pao, D. Patnaik, L. A. Watson, F. Gao, L. Pan, J. Wang, C. Adaikkan, J. Penney, H. P. Cam, W.-C. Huang, L. Pantano, A. Lee, A. Nott, T. X. Phan, E. Gjoneska, S. Elmsaouri, S. J. Haggarty, and L.-H. Tsai. HDAC1 modulates OGG1-initiated oxidative DNA damage repair in the aging brain and Alzheimer's disease. *Nature Communications*, 11(1):2484, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16361-y. URL https://www.nature.com/articles/s41467-020-16361-y. Publisher: Nature Publishing Group.

[30] T. A. Pascoal, M. Chamoun, E. Lax, H.-Y. Wey, M. Shin, K. P. Ng, M. S. Kang, S. Mathotaarachchi, A. L. Benedet, J. Therriault, F. Z. Lussier, F. A. Schroeder, J. M. DuBois, B. G. Hightower, T. M. Gilbert, N. R. Zürcher, C. Wang, R. Hopewell, M. Chakravarty, M. Savard, E. Thomas, S. Mohaddes, S. Farzin, A. Salaciak, S. Tullo, A. C. Cuello, J.-P. Soucy, G. Massarweh, H. Hwang, E. Kobayashi, B. T. Hyman, B. C. Dickerson, M.-C. Guiot, M. Szyf, S. Gauthier, J. M. Hooker, and P. Rosa-Neto. [11C]Martinostat PET analysis reveals reduced HDAC I availability in Alzheimer's disease. *Nature Communications*, 13(1):4171, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-30653-5. URL https://www.nature.com/articles/s41467-022-30653-5. Publisher: Nature Publishing Group.

[31] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library, Dec. 2019. URL http://arxiv.org/abs/1912.01703. arXiv:1912.01703 [cs].

[32] F. W. Pun, I. V. Ozerov, and A. Zhavoronkov. AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*, 44(9):561–572, Sept. 2023. ISSN 0165-6147. doi: 10.1016/j.tips.2023.06.010. URL https://www.cell.com/trends/pharmacological-sciences/abstract/S0165-6147(23)00137-2. Publisher: Elsevier.

[33] S. Razick, G. Magklaras, and I. M. Donaldson. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics*, 9(1):405, Sept. 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-405. URL https://doi.org/10.1186/1471-2105-9-405.

[34] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling. Modeling Relational Data with Graph Convolutional Networks, Oct. 2017. URL http://arxiv.org/abs/1703.06103. arXiv:1703.06103 [stat].

[35] R. Schulte-Sasse, S. Budach, D. Hnisz, and A. Marsico. Graph Convolutional Networks Improve the Prediction of Cancer Driver Genes. In I. V. Tetko, V. Kůrková, P. Karpov, and F. Theis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions*, pages 658–668, Cham, 2019. Springer International Publishing. ISBN 978-3-030-30493-5. doi: 10.1007/978-3-030-30493-5_60.

[36] R. Schulte-Sasse, S. Budach, D. Hnisz, and A. Marsico. 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, June 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00325-y. URL https://www.nature.com/articles/s42256-021-00325-y. Publisher: Nature Publishing Group.

[37] P. Shi, J. Han, Y. Zhang, G. Li, and X. Zhou. IMI-driver: Integrating multi-level gene networks and multi-omics for cancer driver gene identification. *PLOS Computational Biology*, 20 (8):e1012389, Aug. 2024. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1012389. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11379397/.

[38] A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, and J. B. Hogenesch. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences*, 101(16):6062–6067, Apr. 2004. doi: 10.1073/pnas.0400782101. URL https://www.pnas.org/doi/10.1073/pnas.0400782101. Publisher: Proceedings of the National Academy of Sciences.

[39] D. Sun, W. Gao, H. Hu, and S. Zhou. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7):3049–3062, July 2022. ISSN 2211-3835. doi: 10.1016/j.apsb.2022.02.002. URL https://www.sciencedirect.com/science/article/pii/S2211383522000521.

[40] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. Doncheva, S. Pyysalo, P. Bork, L. Jensen, and C. von Mering. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, Jan. 2023. ISSN 0305-1048. doi: 10.1093/nar/gkac1000. URL https://doi.org/10.1093/nar/gkac1000.

[41] R. B. Tanvir, M. M. Islam, M. Sobhan, D. Luo, and A. M. Mondal. MOGAT: A Multi-Omics Integration Framework Using Graph Attention Networks for Cancer Subtype Prediction. *International Journal of Molecular Sciences*, 25(5):2788, Feb. 2024. ISSN 1422-0067. doi: 10.3390/ijms25052788. URL https://www.mdpi.com/1422-0067/25/5/2788.

[42] T. Thanapalasingam, L. v. Berkel, P. Bloem, and P. Groth. Relational Graph Convolutional Networks: A Closer Look. *PeerJ Computer Science*, 8:e1073, Nov. 2022. ISSN 2376-5992. doi: 10.7717/peerj-cs.1073. URL http://arxiv.org/abs/2107.10015. arXiv:2107.10015 [cs].

[43] The Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, 45(10):1113–1120, Oct. 2013. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2764. URL https://www.nature.com/articles/ng.2764.

[44] The Gene Ontology Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, J. Ramsey, D. A. Siegele, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V. Nugnes, F. Quaglia,

S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. dos Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P. Asanitthong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhum, K. Y. C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L. De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, M. A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A. Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K. Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M.-C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, H. Bye-A-Jee, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta, N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka, and M. Westerfield. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, May 2023. ISSN 1943-2631. doi: 10.1093/genetics/iyad031. URL https://doi.org/10.1093/genetics/iyad031.

[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph Attention Networks, Feb. 2018. URL http://arxiv.org/abs/1710.10903. arXiv:1710.10903 [stat].

[46] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen. A new method to measure the semantic similarity of GO terms. *Bioinformatics*, 23(10):1274–1281, May 2007. ISSN 1367-4803. doi: 10.1093/bioinformatics/btm087. URL https://doi.org/10.1093/bioinformatics/btm087.

[47] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(suppl_1):D668–D672, Jan. 2006. ISSN 0305-1048. doi: 10.1093/nar/gkj067. URL https://doi.org/10.1093/nar/gkj067.

[48] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec. GNNExplainer: Generating Explanations for Graph Neural Networks, Nov. 2019. URL http://arxiv.org/abs/1903.03894. arXiv:1903.03894 [cs].

[49] W. Zhao, X. Gu, S. Chen, J. Wu, and Z. Zhou. 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, Nov. 2022. ISSN 1367-4803. doi: 10.1093/bioinformatics/btac622. URL https://doi.org/10.1093/bioinformatics/btac622.

Appendix

A Graph construction and data integration

A.1 Graph summary and descriptive statistics

Table 2 lists the edge count and filtering threshold used for each of the six relation types that form the heterogeneous gene network. The graph is moderately sparse (overall density <0.004), with a heavy-tailed degree distribution typical of biological interaction maps (details in the supplied Jupyter notebook). All subsequent experiments use this exact graph unless stated otherwise.

Table 2: Edge statistics for the heterogeneous gene graph.

Relation type	**Threshold**	**#Edges**
CPDB PPI	score ≥ 0.50	504 378
Co-expression	$\lvert r\rvert \geq 0.80$	34 982
Pathway co-occurrence	Jaccard ≥ 0.60	8 964
GO semantic similarity	Wang ≥ 0.80	8 606
Sequence similarity	top 5 % bitscore	150
Domain similarity	Jaccard ≥ 0.30	208

A.2 Components

A.2.1 Relations

The heterogeneous MORGaN graph contains six *complementary* edge types. Each captures a different notion of functional similarity; combining them lets the model reconcile noisy, partially overlapping evidence rather than over-focusing on any single assay. We consider the following relation types:

- **Protein–protein interaction (PPI).** Proteins are large biomolecules composed of amino-acid chains encoded by genes. A PPI edge is added when two proteins form a physical complex – e.g. an enzyme binds its substrate or two receptors dimerize – detected by assays such as yeast-two-hybrid or affinity purification. We connect the *genes* that encode the interacting proteins with an undirected edge. Because small molecule drugs also act at this physical level, PPI edges supply high-resolution mechanistic context. High-confidence protein-protein interactions are obtained from one of STRING-db [40] (score ≥ 0.8), CPDB [26] (score ≥ 0.5), IRefIndex v.1 and v.4 [33] (score ≥ 0.8), and PCNet [18] (default threshold). CPDB is used as a default.
- **Co-expression.** RNA-seq quantifies how often each gene is transcribed across thousands of samples; higher counts mean the gene is more active. If two genes' expression profiles are consistently correlated, we add an edge, reflecting shared regulation by common transcription factors or signaling programs – even when their proteins never touch. Co-expression therefore contributes *regulatory* information that PPI alone cannot provide. An edge is added between genes with an absolute Pearson correlation ≥ 0.7 across 79 healthy human tissues, based on GSE1133 [38].
- **Pathway co-occurrence.** KEGG [19, 20] curate step-by-step biochemical pathways (e.g. "MAPK signaling"). Genes that appear in the *same* pathways are linked because they participate in a shared biological process. This injects human knowledge and adds a loose sense of up-/down-stream directionality without exploding the number of edge types. We compute the Jaccard similarity of KEGG [19, 20] pathway memberships:

$$J_p(G_1, G_2) = \frac{|P_{G_1} \cap P_{G_2}|}{|P_{G_1} \cup P_{G_2}|} \tag{5}$$

  and include an edge where similarity ≥ 0.60.
- **GO semantic similarity.** The Gene Ontology (GO) is a controlled vocabulary with three name-spaces: *Biological Process* (what the gene does), *Molecular Function* (how), and

Cellular Component (where) [2, 44]. Terms are assigned by curators and automated pipelines. GO edges generalize "same pathway" and cover genes that lack rich KEGG annotations. We compute the geometric mean of best-match-average (BMA) Wang scores [46] across the GO Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) ontologies [2, 44]:

$$R_{GO}(G_1, G_2) = (\mathrm{Sim}^{\mathrm{BP}}_{\mathrm{BMA}}(G_1, G_2) \times \mathrm{Sim}^{\mathrm{MF}}_{\mathrm{BMA}}(G_1, G_2) \times \mathrm{Sim}^{\mathrm{CC}}_{\mathrm{BMA}}(G_1, G_2))^{\frac{1}{3}} \quad (6)$$

and add an edge where $R_{GO}(G_1, G_2) > 0.80$.

- **Sequence similarity.** A sequence similarity edge joins proteins whose sequences align with high statistical confidence. Such homology implies a common ancestor and often a shared 3-D fold or catalytic pocket, allowing MORGaN to transfer knowledge from well-studied family members to poorly characterized relatives. We add an edge to the top 5% BLAST bit-scores (normalized for sequence length) between non-identical gene pairs.
- **Domain similarity.** Pfam domains are recurrent, modular sequence blocks that fold into functional units (e.g. SH2, zinc-finger). We connect two proteins if the Jaccard similarity between their Pfam domain sets exceeds 0.30 [11]. Whereas full-length sequence similarity is global, domain similarity edges focus on the local pockets – pinpointing druggable pockets that recur across otherwise dissimilar proteins, which has proved useful for scaffold hopping in medicinal chemistry.

Why multiple relations? Biology is inherently multi-scale: genes can be co-expressed yet never touch, or interact directly yet be regulated in opposite ways. Integrating multiple edge types allows the model to draw from these multiple relation types.

A.2.2 Node features

Each gene is associated with a four-view multi-omic vector that aggregates evidence about how the gene is *altered* or *active* in sixteen different cancer types: KIRC (kidney renal clear cell carcinoma), BRCA (breast invasive carcinoma), READ (rectum adenocarcinoma), PRAD (prostate adenocarcinoma), STAD (stomach adenocarcinoma), HNSC (head and neck squamous cell carcinoma), LUAD (lung adenocarcinoma), THCA (thyroid carcinoma), BLCA (bladder urothelial carcinoma), ESCA (esophageal carcinoma), LIHC (liver hepatocellular carcinoma), UCEC (uterine corpus endometrial carcinoma), COAD (colon adenocarcinoma), LUSC (lung squamous cell carcinoma), CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma), and KIRP (kidney renal papillary cell carcinoma). This representation allows the model to exploit both pan-cancer regularities and tissue-specific idiosyncrasies in a unified space. The following omics types are included:

- **Copy-number alteration (CNA).** Chromosomal instability can duplicate or delete large DNA segments. We encode the resulting $\log_2$ copy-ratio for each gene. Amplifications drive oncogenes via dosage; deletions can inactivate tumour suppressors; either type of alteration increases the gene's potential therapeutic relevance by changing pathway dynamics and dependencies.
- **Gene expression.** TPM-normalised RNA-seq counts serve as a proxy for transcriptional activity along the canoncial DNA $\rightarrow$ mRNA $\rightarrow$ protein axis. High expression marks pathway engagement and potential vulnerability; zero or strongly tissue-specific expression highlights candidates for of potential on-target toxicity.
- **Mutation frequency.** A *non-synonymous* variant changes an amino-acid and can alter protein function. We supply the fraction of tumours (TCGA) carrying at least one non-synonymous hit in each gene. Recurrent hits point to cancer *drivers*; high frequency therefore raises the prior that a gene is causally important—and druggable.
- **Methylation.** CpG methylation at a promoter recruits proteins that compact chromatin and block transcription – known as *epigenetic silencing*. The β value (where 0 = unmethylated, 1 = fully methylated) distinguishes permanently "switched-off" genes from merely low-copy ones, helping the model avoid nominating silent targets.

Rationale and complementarity. Taken together, these four views cover structural (CNA), regulatory (expression, methylation), and genetic (mutation) evidence. This complementarity provides orthogonal signals that no single modality alone can provide, and enables the encoder to disambiguate

mechanisms (e.g., high expression due to amplification versus loss of expression due to promoter hypermethylation).

Data source and reproducibility. We derive these features from TCGA, a widely used and rigorously curated resource for cancer genomics [43]. Its breadth, depth, and transparent processing pipelines enable reproducible comparisons across studies and typically provide stronger statistical power than smaller proprietary cohorts. While under-representation of rare histologies and understudied genes remains a limitation of any centralized resource, TCGA's standardization and multi-omic scope make it an appropriate foundation for building generalizable target representations at scale.

Extensibility to additional modalities. MORGaN is feature-agnostic: any per-gene descriptor can be appended to the node feature vector without architectural changes. In particular, structural and chemoinformatics descriptors – such as binding-site fingerprints, pocket hydrophobicity, or docking-derived scores – are natural complements to biological priors. Embedding these signals would involve augmenting the node features with quantities derived from 3D structures or *in silico* screening. Because the present work focuses on *upstream* target prioritization from multi-omic and network context, a full end-to-end fusion with chemoinformatics is left for future work; we view this as an exciting extension toward unifying biological and chemical modalities in a single graph-learning pipeline.

A.2.3 Labels

Positive–unlabeled formulation. We frame the task as positive–unlabeled (PU) learning. High-confidence positives – FDA-approved or clinically validated drug targets – are known. However, *true* negatives do not exist: a gene without clinical evidence is not necessarily undruggable. To reflect this epistemic asymmetry, we treat the remaining genes as unlabeled and, for each train/validation/test split, sample negatives uniformly at random from this pool. This approach (i) avoids penalizing understudied genes, (ii) allows estimation of class-conditional risk without inventing a questionable "non-druggable" set, and (iii) yields conservative evaluations because improvements must persist across independent negative samplings.

Moreover, druggability labels are intrinsically skewed (on the order of $\sim$150 Tier-1 positives versus $\sim$16,000 unlabeled genes). There is no authoritative set of genes that are *provably* undruggable, and previously intractable targets continue to become amenable with new modalities (e.g., PROTACs, molecular glues, mRNA therapeutics). We therefore create negatives by resampling a subset of unlabeled genes for every split:

- **Bias dilution.** Because the negative pool changes with each split, the classifier cannot overfit to idiosyncrasies of any single hand-curated list. Despite resampling, metric standard deviations remain low, indicating stable performance.
- **Graph neutrality.** Resampled negatives retain their full connectivity and multi-omic features, preserving the structural context established during pre-training. The model continues to learn from each gene's neighbourhood and attributes even when a given gene is temporarily treated as negative, thereby avoiding topological artefacts that would arise from pruning or rewiring nodes.
- **Forward compatibility.** If a gene is later reclassified as druggable (e.g., due to a new modality), past experiments remain valid because that gene was never canonically fixed as negative. Benchmarks can be rerun with an updated label file without invalidating prior protocols.

These design choices mitigate pathway memorization, manage extreme class imbalance, and keep the evaluation protocol adaptable to methodological and pharmacological advances.

Why binary labels in practice. In principle, "druggability" spans a continuum of chemical tractability that evolves with technology. In practice, however, industrial target pipelines employ *discrete gates* (e.g., evidence of a small-molecule binder, a clinical candidate, or regulatory approval). We therefore label Tier-1 targets (approved or clinical candidates) as positives and sample negatives uniformly from unlabeled genes, mirroring how pipelines prioritize targets. This operational definition enables fair, reproducible benchmarking and aligns with prior work [10], while remaining compatible with future re-labeling as the field progresses.

B Baselines

We keep the node features (§3.1) and the stratified 80/10/10 train–validation–test splits described in §3.4 the same across all baselines. Hyper-parameters are selected by grid search on the validation fold and seeds are fixed to `0, 1` for full reproducibility. Mean ± s.d. over six runs are reported in Table 1 of the main paper. The eight baselines fall into three tiers: feature-only, homogeneous (single-relation) graph and heterogeneous (multi-relation) graph.

B.1 Feature-only baselines

Logistic regression. A linear classifier with L_2 regularization trained on the node features, with no graph structure We use `scikit-learn`'s `LogisticRegression(max_iter = 1000, penalty = "l2", solver = "lbfgs")` and optimize the inverse regularization strength C over $\{0.01, 0.1, 1, 10\}$. Class weights are set inverse-frequency to counter the 1:1 positive/unlabeled sampling. This baseline tests whether a *strictly linear* decision boundary in feature space can already separate druggable from non-druggable genes.

Multilayer perceptron (MLP). Identical input as above, but with two hidden layers to capture non-linear feature interactions. Architecture: `[in` $\rightarrow$ `64` $\rightarrow$ `32` $\rightarrow$ `2]` with ReLU, dropout 0.2 after each hidden layer, and softmax output. Optimiser: Adam (lr $= 1 \times 10^{-3}$, weight-decay 5×10^{-4}), batch size 256, 100 epochs, early stopping (patience 20). Validation tuning sweeps hidden size $\{32, 64, 128\}$ and learning rate $\{1 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}\}$. Serves as a capacity-matched non-graph baseline.

B.2 Homogeneous-graph baselines

Graph convolutional network (GCN). The vanilla spectral GCN operating on the *single* PPI edge set. Best configuration from the grid: two layers, hidden 128, PReLU activation, dropout 0.2, weight-decay 1×10^{-4}. Input is a graph where nodes represent genes, node features are the same as above, and edges are derived from PPIs.

Graph attention network (GAT). Multi-head attention on the same PPI graph. We use three layers with hidden 64 per head, LeakyReLU(0.2), feat-drop 0.2 and attn-drop 0.2. Heads are concatenated inside the network and averaged in the output layer. Edge-specific attention weights let the model down-weight noisy PPI links, providing a stronger yet still homogeneous comparator.

SMG-based (self-supervised masking) Following Cui et al. [8] we add a masked-feature reconstruction pre-text stage to the GCN and GAT backbones. Mask ratio 0.5, 100 pre-training epochs (lr 1×10^{-2}, weight-decay 1×10^{-3}, cosine decay), then fine-tune as above for at most 200 epochs (lr 5×10^{-3}). This pair isolates the effect of *self-supervision* while holding the single-relation topology constant.

B.3 Heterogeneous-graph baselines

MODIG. The multi-omics, multi-relation GAT of Zhao et al. [49] trained on our six-edge-type graph. Each relation is processed by its own two-layer GAT; relation-specific embeddings are fused with learned view-level attention before a final MLP classifier. We keep the authors' recommended settings (hidden 128, 8 heads, dropout 0.3) and tune only the learning rate. MODIG gauges the benefit of heterogeneous edges without any self-supervised pre-training.

MDMNI-DGD. The meta-path DNN of Li et al. [23] – a six-view extension of MODIG that stacks dense layers on hand-crafted meta-path incidence vectors. We train the model on our dataset, following the original paper – we use three hidden layers (256–128–64, dropout 0.3) and Adam (lr 1×10^{-3}). This baseline retains heterogeneous information but replaces GNN message passing with fully-connected fusion, testing whether explicit relational reasoning is needed.

Together, these baselines allus us to disentangle the contributions of (i) multi-omic feature depth, (ii) homogeneous versus heterogeneous topology, and (iii) self-supervised pre-training, ultimately demonstrating the incremental value added by each MORGaN component.

C Self-supervised masked pre-training

Pre-training dynamics. Figure 4 shows that the scaled-cosine reconstruction loss drops sharply during the first ten epochs, then converges smoothly, indicating that the model quickly captures first-order correlations and subsequently refines higher-order structure. The frozen embeddings obtained after 100 epochs serve as initialization for the downstream druggability classifier.

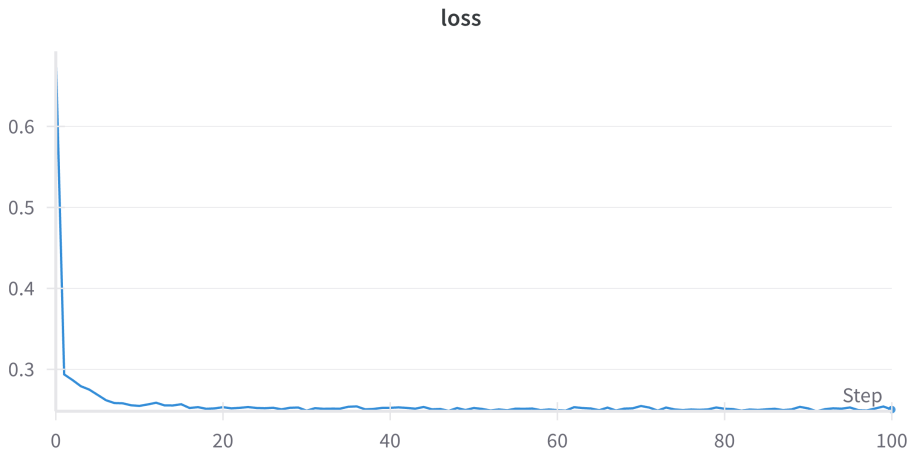


Figure 4: Scaled-cosine reconstruction loss during masked pre-training (mean ± s.d. over six splits).

Hyper-parameter search in brief. A grid of 192 runs varied mask ratio (0.1–0.8), depth (1–4 RGCN layers), learning rate, weight decay and activation. The best AUPR clustered around a mask ratio of 0.5, two or three layers, PReLU activation, learning rate 10^{-2} for pre-training and 5×10^{-3} for fine-tuning, and weight decay 10^{-3} / 10^{-4} respectively (Fig. 5). These values constitute the default configuration shipped in the supplementary `config.yaml`; all reported results use that setting.

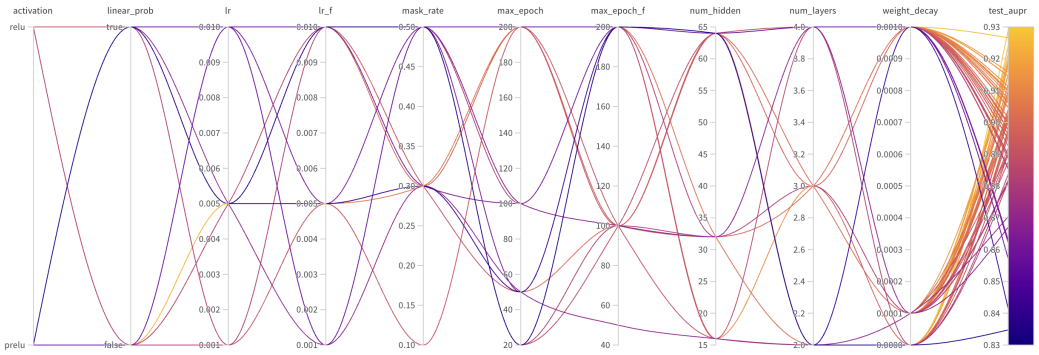


Figure 5: Parallel-coordinates view of the 192-run hyper-parameter sweep; colour encodes test AUPR. Orange lines highlight the high-performing region described in the text.

Sensitivity to masking ratio. Masking ratio (feature corruption). Raising the fraction of masked features from 5% to 50% consistently improves downstream metrics, with AUPR rising by ~ 4 pp and AUROC by ~ 3 pp. A higher mask rate forces the encoder to rely more heavily on relational context instead of relying on raw features, leading to richer, more transferable embeddings. Beyond that, performance eventually degrades.

Sensitivity to loss exponent. Increasing the loss exponent γ (the error curvature) in the SCE reconstruction loss steepens the penalty on large reconstruction errors. This gradually lifts AUPR from 0.907 ($\gamma = 0.5$) to 0.926 ($\gamma = 5.0$), but the gains are modest (< 2 pp AUPR over a ten-fold

change) and all standard deviations overlap, indicating that the model remains broadly insensitive to the precise curvature of the loss.

Hence, performance improves with stronger feature corruption and a steeper loss, but the increments are small. Why such robustness? We believe that it can be traced back to two things:

1. **Aggregated objective.** The MAE sums residuals over six relation types and multidimensional features, so changing the weight on any individual error, via masking or γ, has a diluted global effect.
2. **Masking as a regulariser.** Even relatively moderate corruption ($\geq 30\%$) regularises the model; once in this regime, additional changes are unlikely to reshape the learned space.

Practically, this means MORGaN can be deployed with default settings (e.g. 50% masking, $\gamma = 3$), still achieving within $1 - 2$ pp of the best scores - greatly simplifying hyper-parameter tuning while underscoring the model's inherent robustness.

C.1 Defaults

- **Default hyperparameters:** mask ratio $= 0.5$, γ=3, 2–3 relation layers (PReLU), LR 1×10^{-2} (pre-train), 5×10^{-3} (fine-tune).
- **Early stopping:** monitor validation AUPR with patience 20 epochs.
- **Splits:** report mean $\pm$ s.d. over k seeds; use consistent positive fractions across splits.
- **Compute:** single spmm per layer via basis decomposition for efficiency; training-time wall clock improvements observed vs. per-relation updates.

D Computational requirements and efficiency

All timings were obtained on a **MacBook Pro (Apple M3, 8-core CPU, 16 GB RAM, macOS 15.4.1)** with no GPU acceleration. Table 3 compares MORGaN to the two strongest heterogeneous baselines.

Table 3: Runtime on the six-relation graph (mean ± s.d. over six runs).

Model	**CPU time / epoch (s)**	**End-to-end time (s)**
MODIG	18.60 ± 1.18	$1\,582 \pm 116$
MDMNI-DGD	5.69 ± 0.82	566 ± 16
MORGaN	$\mathbf{0.23 \pm 0.07}$	$\mathbf{24.3 \pm 2.9}$

Key numbers. MORGaN trains $\approx 80\times$ **faster per epoch** than MODIG and completes the full pre-train + fine-tune pipeline $\approx 65\times$ **faster**. Put differently, a hyper-parameter sweep that takes one day with MODIG finishes in under 30 minutes with MORGaN on a traditional laptop.

E Ablation studies

The main paper shows that MORGaN outperforms eight strong baselines; the natural follow-up question is *why*. We therefore conduct six systematic ablation experiments, which all run on the same train–validation–test splits and are evaluated with the same metrics as the main results (AUPR is the headline score).

1. **PPI-source comparisons** (Table 4) swap the base PPI layer among five popular databases (STRING-db, PCNet, CPDB, IRefIndex 2015, IRefIndex 2009) while holding all other relations and features constant. CPDB is used in all main-paper experiments.
2. **Feature ablations** (Table 5) isolate the importance of the four node-feature modalities (CNA, gene expression, methylation, mutation frequency) by training MORGaN on every single, pairwise, triple, and full combination.
3. **Edge-type ablations** (Tables 6-7) repeat the experiment for the six biological relation types.
4. **Randomized-edge control ablations** (Tables 8-9) replace each real edge set with a degree-preserved shuffle keeping node features unchanged. Performance dropping to chance under this perturbation demonstrates that the improvements arise from genuine biology rather than increased edge density or model capacity.
5. **Domain-restricted (organ-system) training** (Table 10) tests whether pan-cancer gains arise from cross-tumour transfer or from a few dominant entities. We retrain MORGaN on *organ-specific feature and label subsets* while holding graph topology fixed.
6. **Model ablations.** We swap the basis-decomposed RGCN encoder for a relational GIN (RGIN) with matched depth/width/parameters and identical pre-training task, decoder, and schedule to probe whether gains are operator-specific or persist across encoder families (Table 11). In addition, we ablate the efficiency components – vertical stacking and weight decomposition – showing that stacking provides the dominant speedup while decomposition preserves this throughput, reduces parameters via sharing, and acts as a mild regularizer (Table 12).

All ablation results are averaged over the six stratified shuffle–split runs described in §3.4; one standard deviation is shown for completeness. The next subsections present the detailed numbers and summarize the key observations.

E.1 Comparison between PPI datasets

Table 4: **PPI-source comparison.** Performance of MORGaN when the PPI layer is sourced from five popular interaction databases. All other edge types and node features are kept identical. **Bold** numbers highlight the best score *within each column*. STRING-db provides the most informative PPI set, pushing AUPR to 0.971, whereas the older IRefIndex releases yield lower accuracy despite comparable AUPR/AUROC figures.

Features	AUPR	AUROC	Accuracy	F1
CPDB	0.888 ± 0.004	0.906 ± 0.004	0.917 ± 0.004	0.919 ± 0.004
IRefIndex 2015	0.949 ± 0.008	0.944 ± 0.004	0.866 ± 0.011	0.869 ± 0.010
IRefIndex	0.949 ± 0.008	0.944 ± 0.004	0.866 ± 0.011	0.869 ± 0.010
PCNet	0.950 ± 0.008	0.941 ± 0.007	0.893 ± 0.004	0.888 ± 0.004
STRINGdb	**0.971 ± 0.002**	**0.970 ± 0.001**	**0.927 ± 0.007**	**0.927 ± 0.007**

E.2 Feature ablations

Table 5: **Ablation of the four input omics modalities**. Blocks separated by lines correspond to (top to bottom) single-, pair-, triple- and four-modality configurations. **Bold** numbers highlight the best score within each column, and *italics* highlight the second-best. Copy-number alterations (CNA) are the most informative modality on their own, whereas combining CNA with gene expression (GE) or mutation frequency (MF) restores accuracy and F1 to the highest levels. Using all four modalities yields a balanced performance but does not surpass the best CNA–based subsets on AUPR.

Features	AUPR	AUROC	Accuracy	F1
Copy Number Alterations (CNA)	**0.908 ± 0.002**	*0.927 ± 0.001*	0.907 ± 0.005	0.909 ± 0.005
Gene Expression (GE)	0.859 ± 0.007	0.919 ± 0.002	0.917 ± 0.004	0.919 ± 0.004
Methylation (METH)	0.884 ± 0.003	0.907 ± 0.002	0.913 ± 0.004	0.914 ± 0.004
Mutation Frequency (MF)	0.866 ± 0.002	0.909 ± 0.002	0.900 ± 0.008	0.902 ± 0.007
CNA + GE	0.874 ± 0.011	0.920 ± 0.007	**0.919 ± 0.000**	**0.921 ± 0.000**
CNA + METH	0.893 ± 0.005	0.910 ± 0.007	0.909 ± 0.004	0.911 ± 0.004
CNA + MF	**0.908 ± 0.004**	**0.929 ± 0.003**	0.911 ± 0.000	0.913 ± 0.000
GE + METH	0.891 ± 0.018	0.907 ± 0.002	0.909 ± 0.008	0.911 ± 0.007
GE + MF	0.881 ± 0.005	0.920 ± 0.002	0.917 ± 0.004	0.918 ± 0.004
METH + MF	0.890 ± 0.004	0.909 ± 0.001	0.915 ± 0.005	0.917 ± 0.005
CNA + GE + METH	0.891 ± 0.002	0.912 ± 0.002	0.917 ± 0.004	0.919 ± 0.004
CNA + GE + MF	0.886 ± 0.006	0.918 ± 0.004	**0.919 ± 0.000**	**0.921 ± 0.000**
CNA + METH + MF	0.897 ± 0.005	0.916 ± 0.006	0.911 ± 0.007	0.913 ± 0.006
GE + METH + MF	**0.908 ± 0.006**	0.913 ± 0.007	0.896 ± 0.008	0.901 ± 0.007
CNA + GE + METH + MF	0.888 ± 0.005	0.906 ± 0.004	0.917 ± 0.004	0.919 ± 0.004

E.3 Edge ablations

Table 6: **Edge–type ablation**, part I (up to four relation types). Each row shows test performance when the heterogeneous graph is restricted to the specified subset of biological relations. Values are mean ± s.d. over the six splits described in §3.4. The full six-relation result (AUPR = 0.888, cf. Table 1) is given for reference in Table 7. The horizontal rules separate 1-, 2-, 3- and 4-relation configurations. **Bold** numbers mark the best score within each block.

Relations	AUPR	AUROC	Accuracy	F1
Co-expression (Coexpr.)	0.788 ± 0.006	0.788 ± 0.003	0.805 ± 0.000	0.782 ± 0.000
Domain Similarity (DomSim)	0.764 ± 0.000	0.533 ± 0.000	0.537 ± 0.000	0.123 ± 0.000
GO Semantic Similarity (GO)	**0.883 ± 0.002**	0.819 ± 0.001	0.805 ± 0.000	0.778 ± 0.000
Pathway Co-occurrence (Path)	0.863 ± 0.008	**0.847 ± 0.004**	**0.843 ± 0.004**	**0.832 ± 0.005**
Sequence Similarity (SeqSim)	0.752 ± 0.000	0.508 ± 0.000	0.512 ± 0.000	0.032 ± 0.000
Coexpr. + DomSim	0.809 ± 0.007	0.798 ± 0.002	0.813 ± 0.000	0.793 ± 0.000
Coexpr. + GO	0.882 ± 0.015	0.892 ± 0.003	0.878 ± 0.000	0.878 ± 0.001
Coexpr. + PPI	0.827 ± 0.021	0.881 ± 0.001	0.835 ± 0.015	0.825 ± 0.013
Coexpr. + Path.	0.878 ± 0.007	**0.904 ± 0.003**	**0.894 ± 0.011**	**0.894 ± 0.013**
Coexpr. + SeqSim	0.783 ± 0.001	0.790 ± 0.001	0.805 ± 0.000	0.782 ± 0.000
DomSim + GO	0.887 ± 0.008	0.825 ± 0.015	0.805 ± 0.000	0.778 ± 0.000
DomSim + PPI	0.756 ± 0.007	0.804 ± 0.004	0.738 ± 0.041	0.740 ± 0.068
DomSim + Path.	0.878 ± 0.003	0.855 ± 0.007	0.841 ± 0.008	0.830 ± 0.010
DomSim + SeqSim	0.769 ± 0.000	0.541 ± 0.000	0.545 ± 0.000	0.152 ± 0.000
GO + PPI	0.849 ± 0.030	0.789 ± 0.074	0.813 ± 0.016	0.792 ± 0.028
GO + Path.	**0.904 ± 0.003**	0.901 ± 0.001	**0.894 ± 0.000**	**0.894 ± 0.001**
GO + SeqSim	0.880 ± 0.004	0.812 ± 0.002	0.805 ± 0.000	0.778 ± 0.000
PPI + Path.	0.869 ± 0.007	**0.904 ± 0.002**	0.843 ± 0.004	0.835 ± 0.006
PPI + SeqSim	0.740 ± 0.014	0.789 ± 0.014	0.726 ± 0.043	0.729 ± 0.059
Path. + SeqSim	0.853 ± 0.004	0.839 ± 0.008	0.835 ± 0.004	0.822 ± 0.005
Coexpr. + DomSim + GO	0.897 ± 0.010	0.895 ± 0.002	0.878 ± 0.007	0.877 ± 0.008
Coexpr. + DomSim + PPI	0.856 ± 0.015	0.895 ± 0.002	0.833 ± 0.019	0.826 ± 0.018
Coexpr. + DomSim + Path.	0.861 ± 0.023	0.902 ± 0.004	0.909 ± 0.004	0.910 ± 0.004
Coexpr. + DomSim + SeqSim	0.811 ± 0.004	0.802 ± 0.003	0.813 ± 0.000	0.793 ± 0.000
Coexpr. + GO + PPI	0.903 ± 0.018	0.909 ± 0.033	0.876 ± 0.004	0.876 ± 0.005
Coexpr. + GO + Path.	0.891 ± 0.001	**0.918 ± 0.001**	**0.917 ± 0.004**	**0.918 ± 0.004**
Coexpr. + GO + SeqSim	0.890 ± 0.025	0.881 ± 0.009	0.872 ± 0.004	0.870 ± 0.004
Coexpr. + PPI + Path.	0.878 ± 0.005	0.917 ± 0.003	0.909 ± 0.004	0.910 ± 0.004
Coexpr. + PPI + SeqSim	0.827 ± 0.015	0.874 ± 0.004	0.819 ± 0.010	0.810 ± 0.021
Coexpr. + Path. + SeqSim	0.866 ± 0.032	0.900 ± 0.003	0.909 ± 0.004	0.910 ± 0.004
DomSim + GO + PPI	0.861 ± 0.031	0.800 ± 0.068	0.807 ± 0.004	0.782 ± 0.008
DomSim + GO + Path.	**0.910 ± 0.001**	0.910 ± 0.000	0.902 ± 0.000	0.903 ± 0.000
DomSim + GO + SeqSim	0.892 ± 0.004	0.835 ± 0.009	0.811 ± 0.004	0.786 ± 0.006
DomSim + PPI + Path.	0.886 ± 0.007	0.914 ± 0.008	0.860 ± 0.012	0.852 ± 0.015
DomSim + PPI + SeqSim	0.778 ± 0.006	0.810 ± 0.006	0.754 ± 0.008	0.766 ± 0.009
DomSim + Path. + SeqSim	0.875 ± 0.005	0.856 ± 0.007	0.848 ± 0.008	0.837 ± 0.010
GO + PPI + Path.	0.895 ± 0.005	0.889 ± 0.014	0.894 ± 0.000	0.894 ± 0.000
GO + PPI + SeqSim	0.872 ± 0.029	0.850 ± 0.066	0.797 ± 0.022	0.779 ± 0.007
GO + Path. + SeqSim	0.902 ± 0.002	0.901 ± 0.001	0.894 ± 0.000	0.894 ± 0.000
PPI + Path. + SeqSim	0.878 ± 0.009	0.905 ± 0.007	0.850 ± 0.008	0.840 ± 0.010
Coexpr. + DomSim + GO + PPI	0.905 ± 0.018	0.907 ± 0.031	0.882 ± 0.008	0.882 ± 0.007
Coexpr. + DomSim + GO + Path	0.899 ± 0.001	**0.920 ± 0.001**	**0.919 ± 0.000**	**0.921 ± 0.000**
Coexpr. + DomSim + GO + SeqSim	0.898 ± 0.010	0.878 ± 0.013	0.870 ± 0.015	0.868 ± 0.016
Coexpr. + DomSim + PPI + Path.	0.870 ± 0.030	0.905 ± 0.016	0.909 ± 0.004	0.910 ± 0.004
Coexpr. + DomSim + PPI + SeqSim	0.844 ± 0.026	0.885 ± 0.010	0.833 ± 0.010	0.837 ± 0.007
Coexpr. + DomSim + Path. + SeqSim	0.863 ± 0.022	0.902 ± 0.005	0.909 ± 0.004	0.910 ± 0.004
Coexpr. + GO + PPI + Path.	0.896 ± 0.003	0.915 ± 0.003	0.915 ± 0.008	0.917 ± 0.008
Coexpr. + GO + PPI + SeqSim	0.876 ± 0.002	0.861 ± 0.001	0.880 ± 0.010	0.879 ± 0.010
Coexpr. + GO + Path. + SeqSim	0.884 ± 0.003	0.915 ± 0.000	**0.919 ± 0.000**	**0.921 ± 0.000**
Coexpr. + PPI + Path. + SeqSim	0.866 ± 0.022	0.900 ± 0.004	0.904 ± 0.012	0.906 ± 0.012

Table 7: **Edge–type ablation**, part II (four to six relation types, continued). This table completes the sweep by listing the remaining four- and five-relation subsets followed by the full six-relation graph (bottom row). Metrics are reported as mean ± s.d. over six runs.

Relations	**AUPR**	**AUROC**	**Accuracy**	**F1**
DomSim + GO + PPI + Path.	0.899 ± 0.008	0.893 ± 0.018	0.902 ± 0.000	0.903 ± 0.000
DomSim + GO + PPI + SeqSim	0.888 ± 0.010	0.889 ± 0.006	0.801 ± 0.024	0.786 ± 0.005
DomSim + GO + Path. + SeqSim	**0.908 ± 0.001**	0.909 ± 0.001	0.900 ± 0.004	0.901 ± 0.004
DomSim + PPI + Path. + SeqSim	0.874 ± 0.005	0.899 ± 0.009	0.837 ± 0.018	0.829 ± 0.015
GO + PPI + Path. + SeqSim	0.883 ± 0.004	0.868 ± 0.006	0.890 ± 0.005	0.890 ± 0.005
Coexpr. + DomSim + GO + PPI + Path.	**0.900 ± 0.005**	**0.919 ± 0.002**	**0.917 ± 0.004**	**0.919 ± 0.004**
Coexpr. + DomSim + GO + PPI + SeqSim	0.891 ± 0.014	0.880 ± 0.031	0.884 ± 0.010	0.885 ± 0.007
Coexpr. + DomSim + GO + Path. + SeqSim	0.891 ± 0.002	0.918 ± 0.001	**0.917 ± 0.004**	0.918 ± 0.004
Coexpr. + DomSim + PPI + Path. + SeqSim	0.878 ± 0.010	0.906 ± 0.011	0.894 ± 0.020	0.897 ± 0.018
Coexpr. + GO + PPI + Path. + SeqSim	0.886 ± 0.007	0.910 ± 0.006	0.913 ± 0.004	0.915 ± 0.004
DomSim + GO + PPI + Path. + SeqSim	0.899 ± 0.010	0.896 ± 0.021	0.896 ± 0.004	0.897 ± 0.004
Coexpr. + DomSim + GO + PPI + Path. + SeqSim	0.888 ± 0.004	0.907 ± 0.005	0.915 ± 0.005	0.917 ± 0.004

E.4 Edge ablations (randomized)

Table 8: **Randomized–edge ablation**, part I (up to three relation types). For each subset of biological relations we replace every edge with a degree-preserved shuffle, keeping node features unchanged. Performance collapses to chance level (AUPR ≈ 0.5, AUROC ≈ 0.5), demonstrating that MORGaN's gains in Table 6 come from *biologically meaningful* topology rather than edge density or parameter count. Horizontal rules separate 1-, 2- and 3-relation configurations; values are mean ± s.d. over six runs.

Relations	AUPR	AUROC	Accuracy	F1
Coexpr.	0.512 ± 0.050	0.499 ± 0.060	0.503 ± 0.046	0.494 ± 0.072
DomSim	0.372 ± 0.141	0.496 ± 0.005	0.504 ± 0.000	0.349 ± 0.367
GO	0.548 ± 0.036	0.531 ± 0.039	0.520 ± 0.045	0.482 ± 0.112
Path.	0.530 ± 0.038	0.500 ± 0.018	0.502 ± 0.015	0.479 ± 0.157
SeqSim	0.498 ± 0.289	0.496 ± 0.005	0.498 ± 0.004	0.166 ± 0.332
Coexpr. + DomSim	0.480 ± 0.040	0.494 ± 0.053	0.488 ± 0.018	0.491 ± 0.059
Coexpr. + GO	0.504 ± 0.026	0.514 ± 0.022	0.528 ± 0.018	0.504 ± 0.078
Coexpr. + PPI	0.622 ± 0.148	0.494 ± 0.025	0.504 ± 0.007	0.141 ± 0.164
Coexpr. + Path.	0.583 ± 0.030	0.579 ± 0.019	0.553 ± 0.015	0.575 ± 0.027
Coexpr. + SeqSim	0.469 ± 0.044	0.460 ± 0.071	0.480 ± 0.043	0.482 ± 0.060
DomSim + GO	0.538 ± 0.054	0.515 ± 0.040	0.520 ± 0.015	0.494 ± 0.151
DomSim + PPI	0.582 ± 0.119	0.520 ± 0.047	0.512 ± 0.033	0.396 ± 0.287
DomSim + Path.	0.491 ± 0.013	0.473 ± 0.018	0.492 ± 0.005	0.341 ± 0.168
DomSim + SeqSim	0.628 ± 0.145	0.512 ± 0.014	0.508 ± 0.014	0.669 ± 0.006
GO + PPI	0.619 ± 0.149	0.495 ± 0.014	0.508 ± 0.008	0.166 ± 0.261
GO + Path.	0.503 ± 0.042	0.474 ± 0.052	0.484 ± 0.024	0.478 ± 0.094
GO + SeqSim	0.483 ± 0.042	0.476 ± 0.027	0.496 ± 0.034	0.358 ± 0.159
PPI + Path.	0.640 ± 0.134	0.497 ± 0.047	0.502 ± 0.040	0.359 ± 0.311
PPI + SeqSim	0.634 ± 0.132	0.525 ± 0.040	0.512 ± 0.022	0.337 ± 0.371
Path. + SeqSim	0.488 ± 0.025	0.485 ± 0.037	0.494 ± 0.023	0.473 ± 0.172
Coexpr. + DomSim + GO	0.539 ± 0.023	0.524 ± 0.032	0.514 ± 0.012	0.496 ± 0.026
Coexpr. + DomSim + PPI	0.492 ± 0.012	0.487 ± 0.009	0.498 ± 0.004	0.479 ± 0.071
Coexpr. + DomSim + Path.	0.507 ± 0.018	0.501 ± 0.028	0.520 ± 0.015	0.519 ± 0.023
Coexpr. + DomSim + SeqSim	0.494 ± 0.015	0.478 ± 0.034	0.482 ± 0.032	0.463 ± 0.078
Coexpr. + GO + PPI	0.617 ± 0.066	0.590 ± 0.067	0.549 ± 0.062	0.561 ± 0.053
Coexpr. + GO + Path.	0.465 ± 0.045	0.463 ± 0.082	0.480 ± 0.072	0.438 ± 0.081
Coexpr. + GO + SeqSim	0.566 ± 0.048	0.549 ± 0.055	0.551 ± 0.033	0.564 ± 0.030
Coexpr. + PPI + Path.	0.480 ± 0.022	0.485 ± 0.031	0.472 ± 0.027	0.468 ± 0.177
Coexpr. + PPI + SeqSim	0.465 ± 0.030	0.447 ± 0.051	0.467 ± 0.060	0.395 ± 0.113
Coexpr. + Path. + SeqSim	0.487 ± 0.036	0.493 ± 0.037	0.504 ± 0.016	0.541 ± 0.087

Table 9: **Randomized–edge ablation**, part II (three to six relation types). Continuation of Table 8, covering the remaining three-, four-, five- and full six-relation shuffles. Even with all six relation layers present but randomized, MORGaN remains close to random guessing, reinforcing that the real multi-relation structure (Table 7) is essential for predictive power.

Relations	AUPR	AUROC	Accuracy	F1
DomSim + GO + PPI	0.503 ± 0.051	0.522 ± 0.061	0.490 ± 0.029	0.467 ± 0.152
DomSim + GO + Path.	0.503 ± 0.048	0.531 ± 0.057	0.537 ± 0.040	0.546 ± 0.079
DomSim + GO + SeqSim	0.532 ± 0.026	0.509 ± 0.020	0.498 ± 0.017	0.552 ± 0.115
DomSim + PPI + Path.	0.582 ± 0.131	0.518 ± 0.063	0.510 ± 0.028	0.420 ± 0.316
DomSim + PPI + SeqSim	0.532 ± 0.008	0.537 ± 0.022	0.533 ± 0.041	0.432 ± 0.224
DomSim + Path. + SeqSim	0.498 ± 0.050	0.523 ± 0.047	0.524 ± 0.037	0.489 ± 0.132
GO + PPI + Path.	0.491 ± 0.063	0.470 ± 0.075	0.480 ± 0.060	0.471 ± 0.176
GO + PPI + SeqSim	0.506 ± 0.049	0.481 ± 0.034	0.486 ± 0.026	0.472 ± 0.114
GO + Path. + SeqSim	0.504 ± 0.028	0.524 ± 0.052	0.533 ± 0.030	0.535 ± 0.074
PPI + Path. + SeqSim	0.552 ± 0.040	0.555 ± 0.023	0.539 ± 0.046	0.504 ± 0.127
Coexpr. + DomSim + GO + PPI	0.529 ± 0.070	0.494 ± 0.064	0.504 ± 0.083	0.522 ± 0.079
Coexpr. + DomSim + GO + Path.	0.532 ± 0.079	0.549 ± 0.102	0.541 ± 0.060	0.550 ± 0.054
Coexpr. + DomSim + GO + SeqSim	0.486 ± 0.045	0.489 ± 0.046	0.502 ± 0.031	0.465 ± 0.081
Coexpr. + DomSim + PPI + Path.	0.549 ± 0.053	0.551 ± 0.019	0.549 ± 0.037	0.460 ± 0.288
Coexpr. + DomSim + PPI + SeqSim	0.477 ± 0.018	0.443 ± 0.023	0.470 ± 0.022	0.488 ± 0.163
Coexpr. + DomSim + Path. + SeqSim	0.520 ± 0.053	0.461 ± 0.029	0.490 ± 0.014	0.449 ± 0.041
Coexpr. + GO + PPI + Path.	0.526 ± 0.021	0.504 ± 0.047	0.496 ± 0.026	0.271 ± 0.076
Coexpr. + GO + PPI + SeqSim	0.496 ± 0.081	0.475 ± 0.087	0.470 ± 0.027	0.509 ± 0.227
Coexpr. + GO + Path. + SeqSim	0.516 ± 0.040	0.507 ± 0.048	0.520 ± 0.030	0.486 ± 0.049
Coexpr. + PPI + Path. + SeqSim	0.538 ± 0.036	0.554 ± 0.026	0.541 ± 0.017	0.467 ± 0.172
DomSim + GO + PPI + Path.	0.519 ± 0.050	0.507 ± 0.040	0.512 ± 0.037	0.539 ± 0.094
DomSim + GO + PPI + SeqSim	0.529 ± 0.039	0.531 ± 0.026	0.533 ± 0.014	0.524 ± 0.093
DomSim + GO + Path. + SeqSim	0.585 ± 0.067	0.582 ± 0.061	0.561 ± 0.049	0.547 ± 0.037
DomSim + PPI + Path. + SeqSim	0.537 ± 0.066	0.536 ± 0.057	0.518 ± 0.017	0.524 ± 0.086
GO + PPI + Path. + SeqSim	0.481 ± 0.049	0.468 ± 0.055	0.492 ± 0.045	0.462 ± 0.151
Coexpr. + DomSim + GO + PPI + Path.	0.506 ± 0.019	0.536 ± 0.014	0.520 ± 0.040	0.493 ± 0.236
Coexpr. + DomSim + GO + PPI + SeqSim	0.511 ± 0.048	0.507 ± 0.038	0.496 ± 0.000	0.463 ± 0.095
Coexpr. + DomSim + GO + Path. + SeqSim	0.503 ± 0.011	0.491 ± 0.011	0.504 ± 0.018	0.525 ± 0.087
Coexpr. + DomSim + PPI + Path. + SeqSim	0.460 ± 0.031	0.449 ± 0.039	0.472 ± 0.040	0.405 ± 0.092
Coexpr. + GO + PPI + Path. + SeqSim	0.515 ± 0.051	0.507 ± 0.052	0.506 ± 0.054	0.341 ± 0.262
DomSim + GO + PPI + Path. + SeqSim	0.477 ± 0.012	0.469 ± 0.016	0.484 ± 0.019	0.419 ± 0.221
Coexpr. + DomSim + GO + PPI + Path. + SeqSim	0.505 ± 0.022	0.479 ± 0.036	0.480 ± 0.038	0.472 ± 0.122

E.5 Domain-restricted (organ-system) training

To determine whether MORGaN's accuracy is driven by a handful of tumor entities or is truly pan-cancer, we trained six separate models, each restricted to one "organ-system" (omics features retained only for the cancer types listed in brackets), based on those already included in the pan-cancer feature set used to train our original model:

- Head and Neck [HNSC]
- Gastro-intestinal [ESCA, STAD, LIHC, COAD, READ]
- Respiratory [LUAD, LUSC]
- Genitourinary [KIRC, KIRP, BLCA, PRAD]
- Reproductive [UCEC, CESC, BRCA]
- Endocrine [THCA]

The table below reports mean $\pm$ s.d. over three random splits ($70/15/15\%$).

Across the profiled organ systems, performance is uniformly strong (AUPR $= 0.892 - 0.919$; AUROC $= 0.891 - 0.927$; Acc $= 0.874 - 0.905$; F1 $= 0.877 - 0.908$), indicating that MORGaN's accuracy is not driven by a single tissue context. Variation is modest (absolute AUPR spread ≤ 0.027

Table 10: Performance by tissue group (mean ± s.d.).

Tissue group	AUPR	AUROC	Accuracy	F1
Gastrointestinal	0.898 ± 0.003	0.913 ± 0.005	0.892 ± 0.004	0.896 ± 0.004
Respiratory	0.919 ± 0.003	0.927 ± 0.002	0.890 ± 0.004	0.896 ± 0.005
Head and neck	0.893 ± 0.008	0.901 ± 0.003	0.874 ± 0.012	0.877 ± 0.016
Genitourinary	0.910 ± 0.009	0.922 ± 0.010	0.905 ± 0.004	0.908 ± 0.005
Reproductive	0.892 ± 0.007	0.891 ± 0.007	0.874 ± 0.009	0.877 ± 0.010

with s.d. ≤ 0.016), and tracks data availability: the Respiratory group achieves the highest scores (AUPR 0.919 ± 0.003, AUROC 0.927 ± 0.002), while Head & Neck and Reproductive, which have fewer established positives, are slightly lower but remain well within the high-performing regime (AUPR $\approx 0.892 - 0.893$, AUROC $\approx 0.891 - 0.901$). Gastrointestinal and Genitourinary are consistently competitive (e.g., AUPR 0.898 and 0.910; AUROC 0.913 and 0.922, respectively). In short, MORGaN generalizes across cancer types; although joint pan-cancer training yields the single best overall model, the per-tissue experiments show that it retains high fidelity even when feature sets are restricted to smaller, system-specific vectors.

E.6 Model ablations

E.6.1 Encoder-family

To test whether MORGaN's gains depend on the specific relational operator, we replace the basis-decomposed **RGCN** encoder/decoder with a **Relational GIN (RGIN)** backbone while keeping the pre-training objective, decoder head, data splits, optimization schedule, and regularization unchanged. We match depth/width to keep parameter count and per-epoch compute comparable.

Table 11 reports mean ± s.d. across the same splits used elsewhere. RGIN performance is comparable with that achieved by our RGCN configuration, indicating that MORGaN's gains primarily arise from the multi-relation masking objective and the information in the heterogeneous graph rather than from a particular choice of message-passing operator.

Table 11: Encoder-family ablation: replacing RGCN with RGIN inside MORGaN (mean ± s.d. across identical splits).

	AUPR	AUROC	Accuracy	F1
RGCN	0.888 ± 0.004	0.907 ± 0.005	0.915 ± 0.005	0.917 ± 0.004
RGIN	0.908 ± 0.005	0.913 ± 0.011	0.898 ± 0.007	0.902 ± 0.007

Takeaway. Comparable results with RGIN suggest the framework is robust to encoder choice; the core driver is the self-supervised multi-relation formulation combined with rich graph context.

E.6.2 Weight decomposition and vertical stacking

We assessed the effect of *weight decomposition* (basis sharing across relations) and *vertical stacking* (single `spmm` over a stacked relation matrix) on both efficiency and accuracy. Runtime was measured on the same data and training schedule.

Efficiency. Vertical stacking accounts for the dominant speedup versus a naive per-relation pass. Adding weight decomposition maintains this fast regime while reducing parameter count via sharing. Without employing vertical stacking and weight decomposition, MORGaN training exhibits a substantially higher runtime ($\sim$ seconds per iteration compared to 0.23 seconds per iteration). With vertical stacking but without weight decomposition, the runtime was approximately 4.26iterations per second.

Accuracy. With vertical stacking but *without* weight decomposition, we observed slightly higher metrics; however, given the large efficiency/parameter benefits of decomposition and its regularizing effect, we retain it as the default. Reported means ± s.d. over the same splits:

Table 12: Performance with vertical stacking but without weight decomposition.

	AUPR	**AUROC**	**Accuracy**	**F1**
Vertical stacking and decomposition	0.888 ± 0.004	0.907 ± 0.005	0.915 ± 0.005	0.917 ± 0.004
Vertical stacking and no decomposition	0.912 ± 0.010	0.913 ± 0.004	0.894 ± 0.010	0.897 ± 0.012

Takeaway. Vertical stacking delivers the primary runtime gain, while weight decomposition preserves that efficiency, reduces parameters through sharing, and serves as an implicit regulariser; we therefore keep decomposition in MORGaN's default encoder.

F Pathway enrichment analysis

Rationale. Given a ranked list of genes from MORGaN (high score = predicted druggable), *pathway enrichment* asks: "Do the top-ranked genes cluster in curated biological pathways more than we would expect by chance?" If so, that provides *external validity*: the model is concentrating probability mass on coherent processes (e.g., cell cycle, receptor signaling) rather than on idiosyncratic single genes.

Pipeline in brief. We use GSEA (Gene Set Enrichment Analysis) in the "pre-ranked" mode:

1. **Rank genes.** Sort all genes by MORGaN's prediction score.
2. **Choose gene sets.** Use curated pathway collections (e.g., KEGG, GO). Each set is simply a list of genes that participate in a process.
3. **Enrichment statistic.** For each pathway, GSEA computes a running-sum statistic that increases when a pathway gene is encountered high in the ranking and decreases otherwise. The maximum deviation of this walk is the raw enrichment score.
4. **Normalization and significance.** Scores are **normalized** by gene-set size, yielding the **NES** (Normalized Enrichment Score), which lets large and small pathways be compared. Significance is assessed by permutation to form a null distribution; we report nominal p-values (NOM p) and multiple-testing–corrected **FDR q-values**.

Analyzed gene sets. We run GSEA on two sets of predictions: **(A)** all genes predicted as positive by MORGaN, and **(N)** the subset of *novel* positives with no prior druggability annotation. Tables 13–14 and Fig. 6 summarize the most significant results (FDR < 0.05).

NES measures *how strongly* a pathway is enriched at the top of the ranking after accounting for set size. FDR q controls for testing many pathways at once (analogous to a false discovery rate in multiple-hypothesis testing). The bar plots in Fig. 6 compare NES across pathway categories; darker bars refer to results on set (A) and lighter bars refer to set (N).

Cancer hallmarks. Both sets recover core oncogenic programs – *cell cycle*, *p53*, *PI3K–Akt*, and composite *pathways in cancer* – indicating that high-scoring genes cluster in well-established cancer biology (Table 13).

Therapeutically actionable signalling. The strongest signals are *receptor-mediated* pathways, led by *neuroactive ligand–receptor interaction* and several **GPCR** cascades (Table 14). GPCRs and related receptors are classic drug targets because they are membrane-exposed, ligandable, and already richly represented in approved medicines. Enrichment here suggests MORGaN's scores align with historically "druggable" target classes rather than random gene families.

Immune and metabolism niches. In the novel set (N), we observe *cytokine–cytokine receptor interaction*, *hematopoietic cell lineage*, and xenobiotic/retinol metabolism. These point to *immunomodulatory* mechanisms (e.g., tuning tumor–immune interactions) and to metabolic processes associated with drug processing and resistance – fertile ground for new targets.

Pathways overlap and are correlated; FDR addresses multiple testing, and NES mitigates gene-set size effects, but some redundancy is expected. Because MORGaN is trained with multi-omic and network context, we consider pathway-level enrichment a complementary sanity check that the model's global ranking is biologically coherent.

Taken together, the enrichment profile shows that MORGaN both *rediscovers* canonical drug classes (external validity) and *highlights* plausible novel targets for follow-up (novel set N).

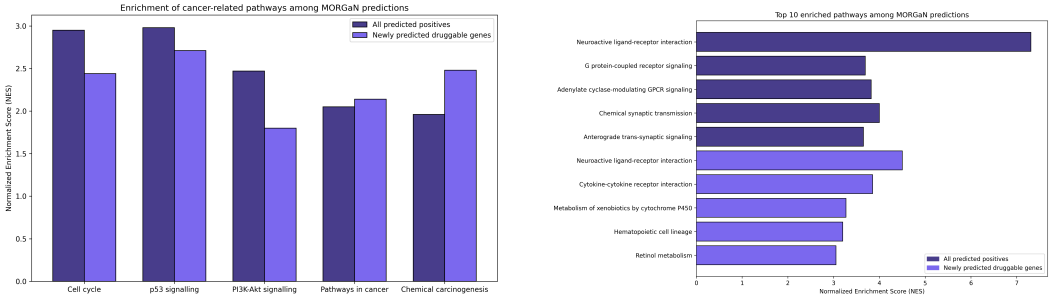


Figure 6: **Visual summary of pathway enrichment analysis results.** *Left:* Normalized enrichment score (NES) for five hallmark cancer pathways. *Right:* Ten most significant pathways overall. Dark bars = all predicted positives; light bars = novel predictions only.

Table 13: Enrichment of hallmark cancer pathways among MORGaN predictions. Normalized enrichment score (NES), FDR q-value and nominal p-value (NOM p) are shown for both established druggable genes (A) and newly predicted candidates (N). All listed pathways pass FDR ≤ 0.05 and NOM $p \leq 0.01$.

Pathway	**Group**	**NES**	**FDR** q	**NOM** p
Cell cycle (KEGG)	A	2.95	0.00049	0.000
Cell cycle (KEGG)	N	2.44	0.00610	0.0023
p53 signaling pathway (KEGG)	A	2.98	0.00098	0.000
p53 signaling pathway (KEGG)	N	2.71	0.00043	0.000
PI3K–Akt signaling pathway (KEGG)	A	2.47	0.00112	0.000
PI3K–Akt signaling pathway (KEGG)	N	1.80	0.04700	0.0077
Pathways in cancer (KEGG)	A	2.05	0.01580	0.0031
Pathways in cancer (KEGG)	N	2.14	0.01050	0.0014
Chemical carcinogenesis (KEGG)	A	1.96	0.02410	0.0050
Chemical carcinogenesis (KEGG)	N	2.48	0.00057	0.000

Table 14: Top five pathways enriched among all (A) and novel (N) MORGaN-predicted druggable genes. Metrics as in Table 13.

Pathway	**Group**	**NES**	**FDR** q	**NOM** p
Neuroactive ligand–receptor interaction (KEGG)	A	7.31	0.000	0.000
G protein-coupled receptor signaling (GO)	A	3.69	0.000	0.000
Adenylate cyclase-modulating GPCR signaling (GO)	A	3.82	0.000	0.000
Chemical synaptic transmission (GO)	A	4.00	0.000	0.000
Anterograde trans-synaptic signaling (GO)	A	3.65	0.000	0.000
Neuroactive ligand–receptor interaction (KEGG)	N	4.50	0.000	0.000
Cytokine–cytokine receptor interaction (KEGG)	N	3.85	0.000	0.000
Xenobiotic metabolism by cytochrome P450 (KEGG)	N	3.27	0.000	0.000
Hematopoietic cell lineage (KEGG)	N	3.20	0.000	0.000
Retinol metabolism (KEGG)	N	3.05	0.000	0.000

G Local interpretability: case studies

Deep graph models often deliver accurate predictions while leaving the mechanistic "*why*" opaque. We ask: "*Which subgraph structure and which feature dimensions were most influential for MORGaN's decision on a specific gene?*" Local explanations help users assess faithfulness, spot failure modes, and form testable hypotheses.

To examine MORGaN's decision process we apply **GNNExplainer** [48], which learns *soft masks* over (i) edges ($M_E \in [0,1]^{|E|}$) and (ii) feature dimensions ($M_F \in [0,1]^d$). The explainer optimizes these masks to *maximize the mutual information* between the masked inputs and the model's output for the target node:

$$\max_{M_E,\, M_F} \; I(Y;\, f(G \odot M_E,\; X \odot M_F)) \tag{7}$$

where f is the frozen trained model, G is the graph (adjacency), X are node features, and $\odot$ denotes element-wise masking. In practice, this is implemented with a differentiable surrogate objective (e.g., cross-entropy on the target logit), plus sparsity and entropy regularizers that encourage compact, human-readable explanations. Thresholding M_E yields an *explanation subgraph*; the thicker the edge, the higher its attribution weight.

Fig. 8 displays subgraphs with the top-20 edges by mask weight for four case genes (two established: *EGFR*, *NOTCH1*; two high-confidence novel: *LAMA3*, *IL4R*). The focal node is enlarged; edge width encodes importance. Fig. 7 aggregates the feature mask into a *cancer-type* $\times$ *omic-layer* heat-map, so we can see whether *structure vs. features*, and *which modality*, drove the call.

G.1 Case studies

a) EGFR – validating known biology. The highest-weight edges connect *EGFR* to *TP53*, *CDK2*, and *CTNNB1*. These neighbors sit on well-studied axes that link receptor tyrosine-kinase signaling to proliferation control: *CDK2* is a core cell-cycle kinase (G1/S transition), *TP53* constrains damaged cells from cycling, and *CTNNB1* (β-catenin) mediates Wnt pathway transcriptional programs that reinforce growth signals. The feature mask assigns large weights to copy-number and expression channels in lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC), indicating that MORGaN's *per-gene* score for *EGFR* is supported by both (i) a *structural motif* tying EGFR to cell-cycle checkpoints and (ii) *omics evidence* of amplification/over-expression in the histologies where EGFR inhibitors are first-line therapy.

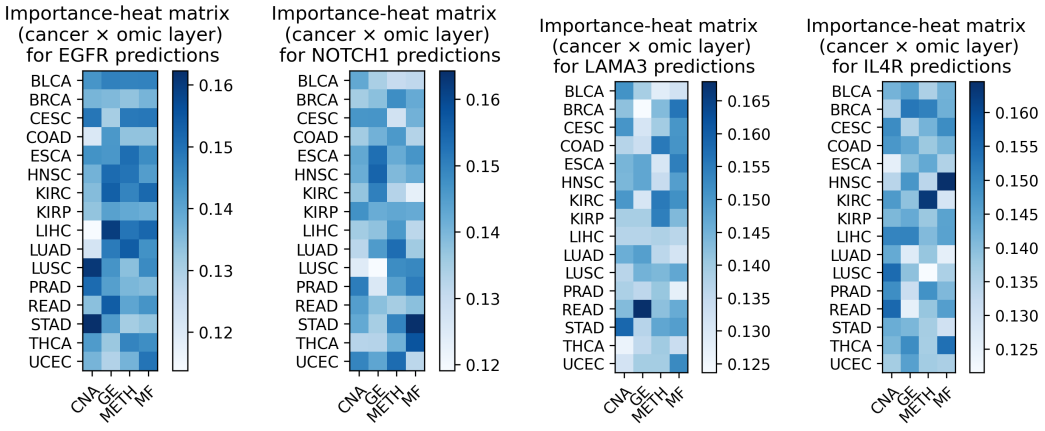
b) NOTCH1 – pathway-centered evidence. Instead of a star around *NOTCH1*, the mask emphasizes two tightly connected patterns: (i) a receptor–kinase crosstalk motif involving *ERBB4* and *MAPK9* (JNK), and (ii) a transcriptional decision module with *RBPJ*, the canonical DNA-binding partner for Notch intracellular domain. This says the model is using *multi-hop pathway context* – how Notch signalling routes into MAPK and transcription – rather than just counting direct interactors. Feature-wise, the importance is spread across expression and methylation channels, which is consistent with NOTCH pathway activity being regulated by both ligand/receptor levels and downstream transcriptional state. The selection of small, interconnected motifs implies the predictor relies on *substructures with function*, not just local density or centrality.

c) LAMA3 – extracellular-matrix lead. For the unlabeled candidate *LAMA3* (a laminin subunit in basement membrane), salient neighbors include *ITGA4* (integrin receptor) and *SMAD1/2* (TGF-β effectors). Together these mark *ECM–integrin–TGF* crosstalk: integrins sense matrix composition and stiffness, transmit signals that modulate SMAD activity, and jointly regulate adhesion, migration, and invasion. The feature mask concentrates in bladder and thyroid contexts, with expression and methylation dimensions carrying the largest weights, suggesting tumor settings where ECM remodeling is particularly informative for the model's decision. For a novel prediction, a coherent *mechanistic neighborhood* plus aligned feature evidence is stronger than either alone. The model is not "hallucinating" from topology.

d) IL4R – immune-evasion angle. The subgraph highlights edges to *AKT2* (PI3K/AKT survival signaling), *TP53BP1* (DNA-damage signaling), and *RAC1* (actin cytoskeleton and motility). This

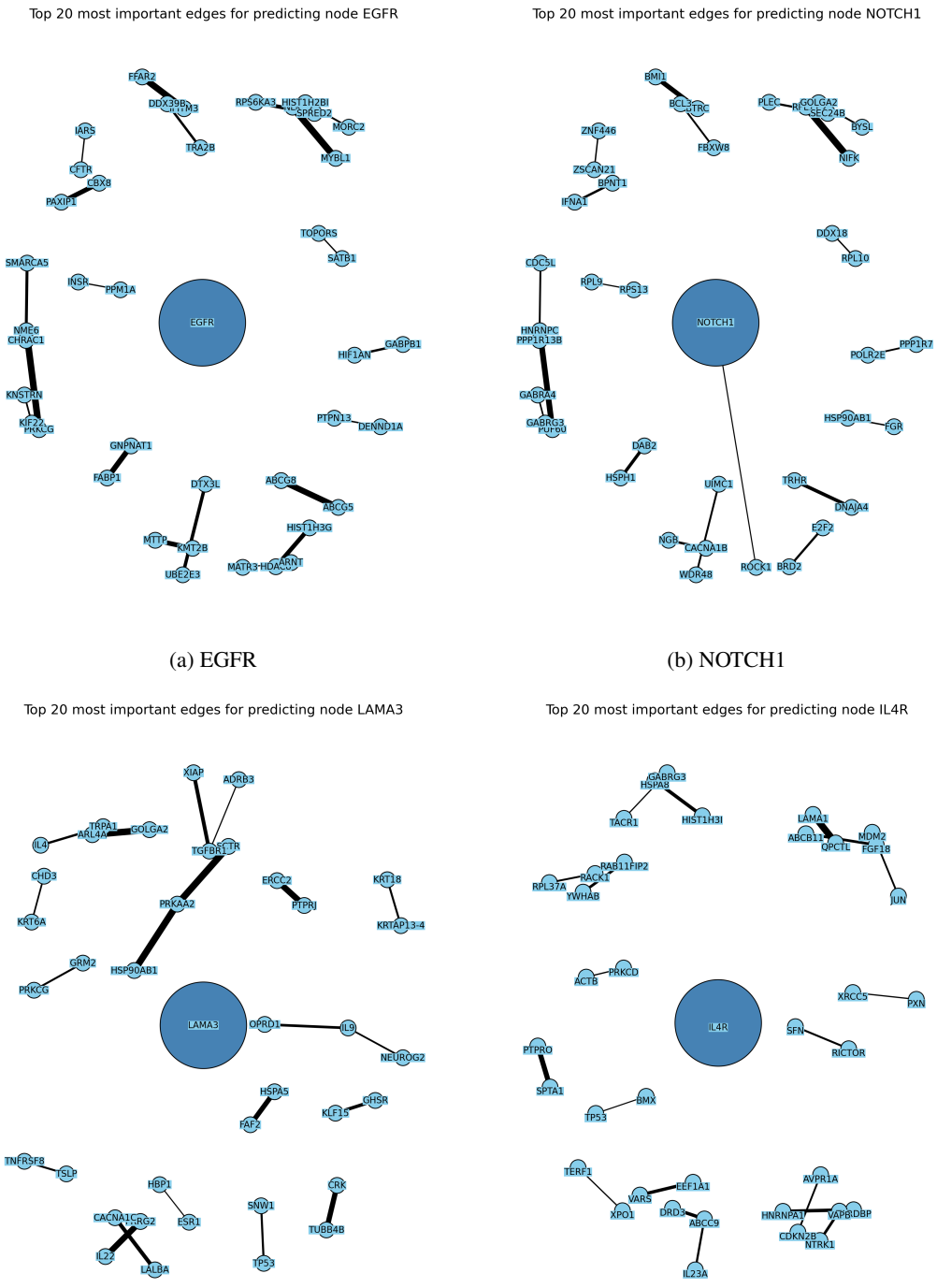
context is expected for *IL4R*, a cytokine receptor that modulates immune and survival pathways: IL-4/IL-13 signaling can activate PI3K/AKT, reshape cytoskeletal dynamics via Rho GTPases, and influence DNA-damage responses indirectly through cell-state changes. The feature mask is strongest in colorectal and lung cancers, with expression and CNA dimensions dominating, again matching settings in which cytokine-driven immune escape and microenvironmental interactions are prominent. Receptor localization (membrane), a signal-integration neighborhood, and high-weight omic channels together form a druggability-consistent explanation. Indeed, the explanation aligns with literature linking IL-4/IL-13 signaling to macrophage polarization and immune escape, supporting IL4R as a promising immuno-oncology target.

Overall, the explanations are *compact*, *stable*, and *mechanistically plausible*, letting us trace MORGaN's "YES" decisions back to specific relational motifs and 'omic signals – useful both as a faithfulness check and as a hypothesis generator for downstream experiments.



(a) EGFR (b) NOTCH1 (c) LAMA3 (d) IL4R

Figure 7: Heat-map visualization of node-feature importance for the same four driver genes. Each panel shows a cancer-type × omic-layer matrix; color intensity is proportional to the contribution weight assigned by GNNExplainer (darker = higher importance).



(a) EGFR

(b) NOTCH1

(c) LAMA3

(d) IL4R

Figure 8: Sub-graphs with the 20 most influential edges (edge width $\propto$ contribution) for four driver genes. The central node is enlarged and darkened.

H External concordance

We compared MORGaN's high-confidence positives ($p > 0.9$) with two external resources: DGIdb [6] and the Finan et al. [10] druggable-genome atlas. Table 15 reports overlaps and proportions. The substantial concordance – particularly the three-way intersection – supports MORGaN's ability to recover genes independently recognized as druggable.

Table 15: Overlap between MORGaN positives and external resources.

Resource	Overlap (genes)	% of MORGaN positives
DGIdb [6]	50	69.1%
Finan [10]	106	74.9%
DGIdb ∩ Finan ∩ MORGaN	609	63.8%

Overall, **80.2%** (765/954) of MORGaN's high-confidence predictions are supported by at least one external resource (DGIdb or Finan), with 63.8% (609/954) shared by both.

Reproducibility note (MDMNI-DGD). We attempted to include MDMNI-DGD predictions for a broader comparison; however, the supplementary gene list referenced in their paper was not accessible (the downloadable file appeared corrupted across multiple attempts). We will add this comparison if/when an updated file becomes available.

I Out-of-distribution experiments

I.1 Alzheimer's disease

Setup. To test disease-agnostic generalization, we built an Alzheimer's disease (AD) network using Alzheimer-specific multi-omic profiles ($\log_2$ fold-change gene expression and chromatin accessibility) and the same six biological relation types used in the pan-cancer graph (derived from [12]). We re-trained MORGaN end-to-end with the identical pre-training and fine-tuning protocol and evaluated on the same split strategy as in the cancer experiments.

Results. Performance remains strong under this domain shift, with a small drop relative to oncology (Table 16). This suggests that the self-supervised, multi-relation objective captures disease-general structure that transfers beyond cancer.

Table 16: Alzheimer's disease: mean $\pm$ s.d. over splits.

	AUPR	AUROC	Accuracy	F1
MORGaN (AD)	0.892 ± 0.022	0.908 ± 0.009	0.840 ± 0.009	0.847 ± 0.008

Qualitative sanity checks. Among high-scoring predictions without prior druggability labels ("false positives" under our operational binary label), MORGaN prioritizes genes with AD-relevant evidence, including *PDE4D* (amyloid/tau pathology; cognitive decline) [28], *HLA-DRA* (upregulated; neuroinflammation) [3], members of the *HDAC* family (pharmacological modulation ameliorates cognitive deficits in AD models) [29, 30], as well as *NTRK1* (nervous system development) and *NRP1* (neuronal migration, angiogenesis; upregulated in AD models) [24]. These examples support that out-of-distribution predictions remain biologically plausible.

I.2 Essential genes

Setup. To illustrate the task-agnostic utility of MORGaN embeddings, we evaluated a distinct prediction task: gene essentiality. We used proxy labels derived from prior predictions [13] (subset to **E** (essential)) and applied the same training/evaluation protocol (architecture and schedule unchanged), treating this as a separate downstream classification problem.

Results. Despite the weaker, prediction-derived labels, performance is competitive (Table 17), indicating that MORGaN learns task-general representations that transfer to essentiality beyond the original druggability objective.

Table 17: Essential gene prediction: mean $\pm$ s.d. over splits.

	AUPR	AUROC	Accuracy	F1
MORGaN (essential)	0.765 ± 0.015	0.835 ± 0.008	0.772 ± 0.009	0.797 ± 0.008

Takeaway. Across both experiments, MORGaN's multi-relation self-supervision yields embeddings that generalize across *diseases* (AD) and *tasks* (essentiality), with only modest degradation under distribution shift and competitive performance under weaker labels.