QUANTUM-INSPIRED DUAL-PATH GRAPH LEARNING FOR MULTIMODAL BIOMEDICAL TRIADIC PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Biomedical triadic prediction identifies context-dependent associations among molecular entities (e.g., RNAs, genes, and proteins), drugs, and diseases to support drug repurposing and candidate screening. However, molecular sequences, chemical structures, and textual descriptions differ in dimension and statistical properties, making direct fusion difficult. Moreover, missing modality features limit the information available for prediction, while noisy features can interfere with useful signals from other modalities. We propose Quantum-Inspired Dual-Path Graph Learning (QDGL) to integrate heterogeneous modalities and improve triadic prediction with missing or noisy modality features. First, Hilbert-sphere encoding maps modality features onto a shared complex unit sphere through learnable amplitudes and phases, providing a common space for fusion. Second, reliability-weighted fusion learns entity-specific weights to adjust the contributions of available modalities and constructs both vector representations and mixed density states. The vector path aggregates neighborhood features, while the density path propagates mixed states through state matching. Together, these paths combine available modality features and graph relations to mitigate the impact of missing features. Vector and measurement-based tensor interactions are combined for triadic association classification and disease ranking given a molecule–drug pair. Experiments on three biomedical benchmarks (ncRNADrug, CTD, and PrimeKG) show that QDGL achieves the highest mean Hit@1, MRR, and AUC among the compared methods, with maximum relative improvements of 12.04%, 8.05%, and 4.72%, respectively, over the strongest baseline for each dataset and metric. Further experiments demonstrate generalization to unseen entity pairs and robustness to missing features and structural feature perturbations, while ablations support the effectiveness of key designs.

1 INTRODUCTION

Biomedical triadic prediction identifies context-dependent associations among molecular entities, drugs, and diseases to support drug repurposing and candidate prioritization. Molecular entities, including noncoding RNAs, genes, and proteins, provide context for drug–disease associations. Biomedical knowledge graphs organize known relations for association prediction (Huang et al., 2024), and recent studies combine heterogeneous graphs with multimodal attributes to predict drug–target–disease triplets (Zheng et al., 2025). Molecular sequences, drug structures, textual descriptions, and ontology information complement graph relations in predicting unobserved associations.

However, these sources differ in representation, completeness, and quality. Feature dimensions and statistics vary across modalities, and some entities lack certain modality features (Tang et al., 2024). Noisy modalities can also suppress useful signals from others (Chaudhuri et al., 2025). Recent studies use relation-guided modality representations and expert combinations to coordinate modality-specific predictions (Zhang et al., 2025b). Triadic prediction further requires combining these features with relations among all three entity types to assess their joint association. The challenge is to integrate heterogeneous modalities while limiting noise effects and using the remaining attributes and graph relations when some modalities are missing.

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

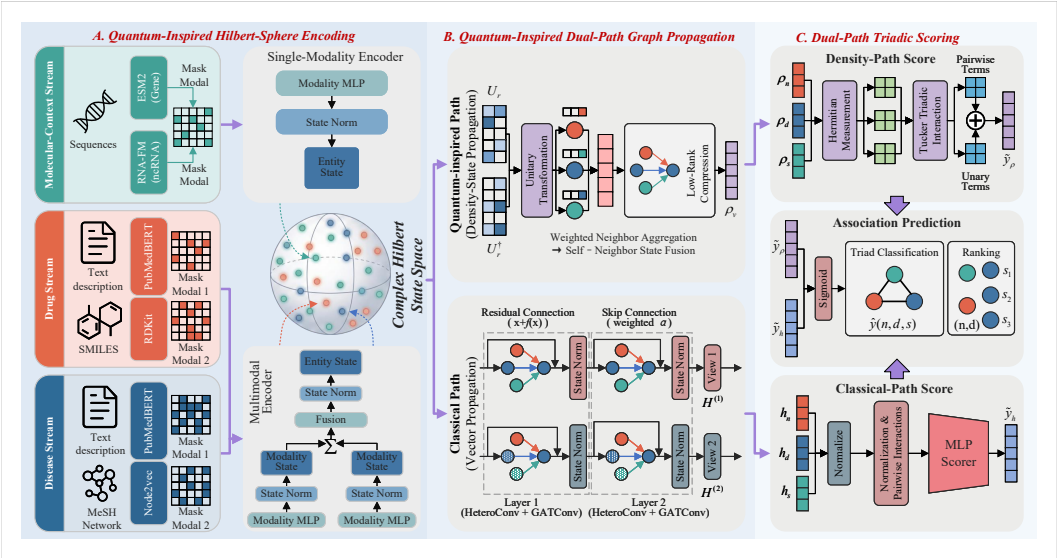


Figure 1: Overview of QDGL. Multimodal encoding and dual-path graph propagation produce vector and density representations, whose scores are combined for triad classification and candidate ranking.

We propose Quantum-Inspired Dual-Path Graph Learning (QDGL) to integrate heterogeneous modalities for triadic prediction with missing or noisy features. Drawing on density representations and measurements (Zhang et al., 2025a), we first design Hilbert-sphere encoding with learnable amplitudes and phases to map features onto a shared complex unit sphere for fusion and state matching. Second, reliability-weighted fusion adjusts available modalities' contributions using entity-specific weights. Weighted superposition constructs vector representations, while mixing modality density operators forms mixed density states for state matching. The vector path aggregates neighborhood features through heterogeneous graph attention, while the density path propagates mixed states through relation-conditioned transformations and state matching. Combining vector-path feature interactions with density-path measurement-based tensor interactions enables triadic association classification and disease ranking for a molecule–drug pair.

We compare QDGL with eleven baselines from four categories: knowledge graph embedding, quantum-inspired knowledge graph embedding, graph neural networks, and triadic completion and multimodal models. Across three biomedical benchmarks (ncRNADrug, CTD, and PrimeKG), QDGL achieves the highest mean Hit@1, MRR, and AUC among these methods. Further experiments demonstrate unseen-pair generalization and robustness to missing features and graph relations. Structural feature perturbations examine modality responses, while state-representation controls, ablations, and dual-path analyses support key designs.

The main contributions are as follows:

- We propose QDGL with learnable amplitude–phase Hilbert-sphere encoding, providing normalized complex states for heterogeneous modality fusion and relational modeling.
- We design reliability-weighted fusion and dual-path graph propagation to learn vector representations and mixed density states, combining vector and measurement-based tensor interactions for triadic association prediction.
- QDGL achieves the highest mean Hit@1, MRR, and AUC among the compared methods on three biomedical benchmarks, with further evidence of unseen-pair generalization, robustness under missing information and structural feature perturbations, and the effectiveness of key designs.

2 METHODOLOGY

As shown in Figure 1, QDGL constructs vector representations and mixed density states through quantum-inspired Hilbert-sphere encoding and reliability-weighted fusion, propagates them along two graph paths, and combines their scores for triadic association classification and candidate disease ranking. The training procedure and theoretical proofs are provided in Algorithm S1 and Appendix S5.3, respectively.

2.1 TASK DEFINITION AND FEATURE CONSTRUCTION

We study molecular–drug–disease triadic association prediction. The three entity sets are denoted by $\mathcal{N}$, $\mathcal{D}$, and $\mathcal{S}$, respectively, with observed associations forming $\mathcal{T}^+ \subseteq \mathcal{N} \times \mathcal{D} \times \mathcal{S}$. The model learns an association score $\widehat{y}(n, d, s) \in [0, 1]$ for classifying candidate triads and ranking candidate diseases for a given molecular–drug pair. Pairwise projections of positive training triads form a heterogeneous graph containing molecular–drug, drug–disease, and molecular–disease relations for subsequent relational propagation.

We construct modality-specific features for the three entity types. ncRNA sequences are encoded by RNA-FM, while protein sequences, including those mapped from genes, are encoded by ESM2. Drugs use PubMedBERT representations of textual descriptions and ECFP4 molecular fingerprints generated by RDKit. Diseases use PubMedBERT representations of textual definitions and Node2Vec embeddings derived from MeSH or MONDO ontologies. Preprocessing details and feature dimensions are provided in Appendix S3.

2.2 QUANTUM-INSPIRED HILBERT-SPHERE ENCODING

To support unified state fusion and interaction computations across modalities with different feature dimensions and statistics, QDGL uses modality-specific amplitude–phase encoders to map input features onto the complex Hilbert unit sphere:

$$\boldsymbol{\psi}_v^{(m)} = \sqrt{\mathrm{softmax}\left(\mathbf{a}_v^{(m)}\right)} \odot \exp\left(i\pi \tanh\left(\boldsymbol{\varphi}_v^{(m)}\right)\right), \qquad \left[\mathbf{a}_v^{(m)}, \boldsymbol{\varphi}_v^{(m)}\right] = f_{\mathrm{state}}^{(m)}\left(\mathbf{u}_v^{(m)}\right), \quad (1)$$

where $\mathbf{u}_v^{(m)} \in \mathbb{R}^{d_m}$ is the input feature of node v under modality m, and $f_{\mathrm{state}}^{(m)}$ outputs unnormalized weight parameters $\mathbf{a}_v^{(m)} \in \mathbb{R}^D$ and phase parameters $\boldsymbol{\varphi}_v^{(m)} \in \mathbb{R}^D$. Softmax normalizes the squared amplitudes, while the complex exponential introduces phase without changing amplitude. Consequently, $\|\boldsymbol{\psi}_v^{(m)}\|_2 = 1$, so all modality states lie on the unit sphere of the same complex Hilbert space $\mathbb{C}^D$.

For weighted modality fusion, we construct a density operator from the outer product of each state vector:

$$\rho_v^{(m)} = |\psi_v^{(m)}\rangle\langle\psi_v^{(m)}|, \mathrm{Tr}(\rho_v^{(m)}) = 1, \quad (2)$$

where $\rho_v^{(m)}$ is a rank-one pure-state density operator. Its diagonal entries represent normalized latent component weights, while its off-diagonal entries contain amplitude products and phase differences, allowing subsequent state matching to use both component weights and relative phases.

2.3 QUANTUM-INSPIRED DUAL-PATH GRAPH PROPAGATION

Based on these modality states, QDGL learns node-specific fusion weights to construct the initial representations of the density and classical paths. For a node v with at least one available modality, modality-specific MLPs compute reliability scores $e_v^{(m)} = f_{\mathrm{rel}}^{(m)}(\mathbf{u}_v^{(m)})$, which are normalized over its available modality set $\mathcal{M}_v$:

$$\pi_v^{(m)} = \frac{\exp(e_v^{(m)})}{\sum_{k \in \mathcal{M}_v} \exp(e_v^{(k)})}, \qquad m \in \mathcal{M}_v, \quad (3)$$

where $\pi_v^{(m)}$ is a modality weight learned through joint training, representing the relative contribution of modality m to the node's fused representation. The weighted sum of modality density operators gives the initial mixed state of the density path:

$$\rho_v^{(0)} = \sum_{m \in \mathcal{M}_v} \pi_v^{(m)} \rho_v^{(m)} = \sum_{m \in \mathcal{M}_v} \pi_v^{(m)} |\psi_v^{(m)}\rangle\langle\psi_v^{(m)}|. \tag{4}$$

When only one modality is available, its weight is one and the mixed state reduces to the corresponding pure state. When all modalities are missing, we encode learnable missing-modality representations and mix the resulting density operators with equal weights to construct the initial node representation.

The classical path uses the same modality states and weights to form a complex vector through weighted superposition and unit normalization, then concatenates its real and imaginary parts:

$$\widetilde{\psi}_v^{(0)} = \mathrm{Norm}\left(\sum_{m \in \mathcal{M}_v} \sqrt{\pi_v^{(m)}}\, \psi_v^{(m)} \right), \qquad \mathbf{h}_v^{(0)} = \mathrm{Re}(\widetilde{\psi}_v^{(0)}) \oplus \mathrm{Im}(\widetilde{\psi}_v^{(0)}), \tag{5}$$

where $\mathrm{Norm}(\cdot)$ denotes unit normalization, $\oplus$ denotes vector concatenation, and $\mathbf{h}_v^{(0)}$ is the initial real-valued representation of the classical path. The two paths thus organize the same node's multimodal information as a vector and a density state, respectively.

The classical path aggregates neighbor features through relation-aware attention and incorporates the initial representation as a residual to update the target node:

$$\mathbf{h}_v^{(l+1)} = \mathrm{Norm}\left(\mathrm{ReLU}\left(\sum_{r \in \mathcal{R}_{\tau(v)}} \sum_{u \in \mathcal{N}_r(v)} \alpha_{uv,r}^{(l)} \mathbf{W}_r^{(l)} \mathbf{h}_u^{(l)} \right) + \mathbf{R}_{\tau(v)}^{(l)} \mathbf{h}_v^{(0)} \right), \tag{6}$$

where $\tau(v)$ denotes the node type, $\mathcal{R}_{\tau(v)}$ is the corresponding set of incoming relation types, and $\mathcal{N}_r(v)$ is the neighborhood under relation r. The attention weight $\alpha_{uv,r}^{(l)}$ is normalized within this neighborhood, and $\mathbf{W}_r^{(l)}$ is a relation-specific linear transformation. The classical path uses $L_v = 2$ propagation layers. The residual map $\mathbf{R}_{\tau(v)}^{(l)}$ is a type-specific linear projection in the first layer and learnable scalar scaling in the second. The final vector representation is $\mathbf{h}_v = \mathbf{h}_v^{(L_v)}$.

The quantum-inspired path propagates mixed density states and uses state overlap for neighborhood matching. For a neighbor u connected to target node v by relation r, we first apply a relation-specific unitary transformation:

$$\overline{\rho}_{u,r}^{(l)} = \mathbf{U}_r \rho_u^{(l)} \mathbf{U}_r^{\dagger}, \tag{7}$$

where $\mathbf{U}_r$ is a unitary transformation parameterized by a learnable relation phase, $\dagger$ denotes the conjugate transpose, and $\overline{\rho}_{u,r}^{(l)}$ is the transformed neighbor state. Attention scores are then computed from the Hilbert–Schmidt overlap between the target state and the transformed neighbor state:

$$\xi_{uv,r}^{(l)} = \frac{\sqrt{D}}{T_{\mathrm{att}}} \mathrm{Tr}\left(\rho_v^{(l)} \overline{\rho}_{u,r}^{(l)} \right) + b_r, \tag{8}$$

where $T_{\mathrm{att}} > 0$ is the attention temperature and b_r is a relation-specific bias. Neighborhood softmax converts the scores into weights $\gamma_{uv,r}^{(l)}$, which are used to fuse neighbor information with the node's current state:

$$\rho_v^{(l+1)} = \mathcal{C}_{R_c}\left(\eta_{\tau(v)} \rho_v^{(l)} + \left(1 - \eta_{\tau(v)}\right) \sum_{(u,r) \in \mathcal{N}(v)} \gamma_{uv,r}^{(l)} \overline{\rho}_{u,r}^{(l)} \right), \tag{9}$$

where $\mathcal{N}(v)$ contains the target node's neighbors together with their relation types, and $\eta_{\tau(v)} \in [0,1]$ is a type-specific mixing coefficient. The operator $\mathcal{C}_{R_c}$ compresses the fused state to rank at most R_c, preserving positive semidefiniteness and unit trace.

After L_ρ layers of these state updates, we further fuse the propagated state with the initial mixed state through a density residual:

$$\rho_v = \beta \rho_v^{(L_\rho)} + (1-\beta)\rho_v^{(0)}, \tag{10}$$

where $\beta \in (0,1)$ is a learnable mixing coefficient balancing propagated neighborhood information and initial multimodal information. The final density state ρ_v and classical-path representation $\mathbf{h}_v$ are jointly used for triadic association scoring.

2.4 Dual-Path Triadic Scoring

The triadic scorer computes interaction scores from vector representations and density states separately, then combines them for association prediction. The classical branch constructs pairwise features $\mathbf{c}_{ab}^{-} = |\mathbf{h}_a - \mathbf{h}_b|$ and $\mathbf{c}_{ab}^{\odot} = \mathbf{h}_a \odot \mathbf{h}_b$ from the normalized node representations. The node features, absolute differences, and products are concatenated and passed to an MLP:

$$\widetilde{y}_h = \mathrm{MLP}\Big(\mathbf{h}_n \oplus \mathbf{h}_d \oplus \mathbf{h}_s \oplus \mathbf{c}_{nd}^{-} \oplus \mathbf{c}_{ds}^{-} \oplus \mathbf{c}_{ns}^{-} \oplus \mathbf{c}_{nd}^{\odot} \oplus \mathbf{c}_{ds}^{\odot} \oplus \mathbf{c}_{ns}^{\odot}\Big). \tag{11}$$

The quantum-inspired branch uses entity-type-specific Hermitian measurements to convert density states into real-valued interaction features. In column-vector notation, the measurement readout is

$$\mathbf{q}_v = \mathbf{P}_{\tau(v)} \left[\mathrm{Tr}(\rho_v \mathbf{M}_{\tau(v),j}) - D^{-1}\right]_{j=1}^{J}, \tag{12}$$

where $\mathbf{M}_{\tau(v),j} = |\mathbf{b}_{\tau(v),j}\rangle\langle \mathbf{b}_{\tau(v),j}|$ is a rank-one Hermitian measurement operator constructed from a learnable unit vector, J is the number of measurements, and $\mathbf{P}_{\tau(v)}$ is a type-specific learnable linear map. After subtracting the maximally mixed-state measurement baseline D^{-1}, the readout features form the density-path score through a Tucker-parameterized triadic interaction together with pairwise and unary terms:

$$\widetilde{y}_\rho = b + \kappa \left[\langle \mathcal{G}, \mathbf{q}_n \otimes \mathbf{q}_d \otimes \mathbf{q}_s \rangle + \sum_{(a,b)\in\mathcal{P}} \mathbf{q}_a^\top \mathbf{A}_{\tau(a),\tau(b)} \mathbf{q}_b + \sum_{a\in\{n,d,s\}} \mathbf{u}_{\tau(a)}^\top \mathbf{q}_a \right], \tag{13}$$

where $\mathcal{P} = \{(n,d),(d,s),(n,s)\}$ is the set of entity pairs in a triad, and $\mathcal{G}$ is the Tucker core modeling triadic interactions. The type-specific parameters $\mathbf{A}_{\tau(a),\tau(b)}$ and $\mathbf{u}_{\tau(a)}$ are the pairwise interaction matrix and unary weight vector, respectively, while b is a bias and $\kappa > 0$ is a scaling coefficient.

The two path scores are fused as $\widetilde{y}(n,d,s) = \widetilde{y}_h + g\widetilde{y}_\rho$ and passed through a sigmoid function to obtain the association prediction:

$$\widehat{y}(n,d,s) = \sigma(\widetilde{y}(n,d,s)) = \sigma(\widetilde{y}_h + g\widetilde{y}_\rho), \tag{14}$$

where $g \in (0,1)$ controls the contribution of the density-path score, and $\sigma(\cdot)$ is the sigmoid function. The prediction is used for triadic association classification, while the fused score ranks candidate diseases for a fixed molecular–drug query.

The model jointly optimizes association classification and disease ranking, with cross-modal and cross-view constraints coordinating representation learning. During training, a base view retains the original training graph, while a perturbed view applies relation-wise edge dropout and message dropout. Both views execute the dual-path encoding described above and share model parameters. The overall loss is

$$\mathcal{L} = \mathcal{L}_C + \lambda_R(t)\mathcal{L}_R + \lambda_S\mathcal{L}_S + \lambda_N\mathcal{L}_N + \lambda_P\mathcal{L}_P + \lambda_\rho\mathcal{L}_\rho, \tag{15}$$

Table 1: Dataset sizes and sparsity. A/B/C denote molecules/drugs/diseases. One-shot denotes one occurrence in positive triads.

Statistic	ncRNADrug	CTD	PrimeKG
#Triads	24,211	397,598	5,496
#Entities (A/B/C)	12,374 / 181 / 156	6,800 / 4,230 / 1,464	344 / 650 / 439
Median ent. freq (A/B/C)	1 / 3 / 4	11 / 6 / 11	4 / 3 / 4
One-shot ent. (% A/B/C)	81.26 / 34.81 / 28.85	7.81 / 18.75 / 16.60	21.22 / 21.54 / 28.70
Gini (A/B/C)	0.454 / 0.913 / 0.886	0.816 / 0.892 / 0.895	0.711 / 0.656 / 0.723
#Pairs (AB/BC/AC)	21,056 / 683 / 19,096	146,427 / 44,575 / 25,588	2,079 / 2,266 / 1,828
Median pair freq. (AB/BC/AC)	1 / 2 / 1	1 / 3 / 8	2 / 1 / 2
One-shot pairs (% AB/BC/AC)	92.40 / 46.56 / 82.82	53.03 / 30.21 / 10.50	44.40 / 57.37 / 48.74

where $\mathcal{L}_C$ uses binary cross-entropy to distinguish positive and negative triads, and $\mathcal{L}_R$ uses pairwise Softplus to rank the observed disease above sampled negative diseases for a fixed (n, d) query. $\mathcal{L}_S$ coordinates different modality representations of the same entity through mutual prediction with stop-gradient targets, $\mathcal{L}_N$ contrasts node representations across the two views, $\mathcal{L}_P$ aligns class prototypes of positive and negative triads across the two views, and $\mathcal{L}_\rho$ aligns the propagated density states. With t denoting the training epoch, the ranking weight $\lambda_R(t)$ is zero for the first E_w epochs and equals λ_R thereafter, while the remaining λ coefficients are fixed auxiliary-loss weights. The complete training objectives are defined in Appendix S5.1.1.

3 Experiments

3.1 Datasets and Evaluation Metrics

We conduct experiments on three triadic association datasets: ncRNADrug (Cao et al., 2024), CTD (Davis et al., 2025), and PrimeKG (Chandak et al., 2023). Triads in ncRNADrug are derived from curated drug-response records, while those in CTD and PrimeKG are constructed from pairwise-relation closures, as defined in Appendix S3. Table 1 summarizes dataset sizes and frequency distributions. ncRNADrug contains many low-frequency molecular entities and entity pairs, CTD is the largest dataset with a pronounced long-tailed entity frequency distribution, and PrimeKG contains many low-frequency pairs across all three pair types.

We use AUC to evaluate triadic association classification and Hit@1 and MRR to evaluate disease ranking. The main results table reports sampled ranking, where the target disease is ranked against 50 randomly sampled negative diseases for each fixed molecular–drug query (Teru et al., 2020). Metric definitions are provided in Appendix S4.4.

3.2 Implementation Details

We perform random five-fold cross-validation on deduplicated positive triads, holding out one fold for testing and reserving validation data from the remainder. Classification negatives are generated by replacing entities in the corresponding positive triads. Training negatives for classification and ranking are filtered against training positives only, while validation candidates are filtered against the union of training and validation positives and fixed within each fold. Test ranking retains the target disease and filters out other known positive diseases for the same query using all known positives. We optimize the model with AdamW. Hyperparameter selection, early stopping, and model selection are guided by validation performance. Detailed settings are provided in Appendix S4.

Random triad splitting allows training and test sets to share constituent entity pairs and therefore primarily evaluates triadic completion within an existing relational context. Appendix S6.6 further examines the effects of pair reuse and generalization to unseen pairs through pair-overlap-stratified and pair-disjoint evaluations.

Table 2: Comparison with SOTA methods. **Bold** and underline mark the best and second-best results. $\Delta(\%) = (\text{Best} - \text{Second})/\text{Second} \times 100$.

Method	ncRNADrug			CTD			PrimeKG		
	Hit@1	MRR	AUC	Hit@1	MRR	AUC	Hit@1	MRR	AUC
Knowledge Graph Embedding (KGE)									
TransE	28.95 ± 7.98	43.33 ± 7.32	61.83 ± 2.12	44.35 ± 2.22	60.54 ± 1.83	89.69 ± 0.08	83.24 ± 2.66	86.40 ± 1.94	86.33 ± 0.74
RotatE	40.92 ± 0.82	50.08 ± 0.60	56.90 ± 0.59	72.46 ± 0.26	81.92 ± 0.21	93.92 ± 0.08	86.72 ± 0.61	89.61 ± 0.35	91.20 ± 0.34
Quantum-inspired Knowledge Graph Embedding									
QIQE-KGC	40.17 ± 0.69	58.34 ± 0.44	70.40 ± 0.87	58.09 ± 0.21	70.58 ± 0.16	81.61 ± 0.17	65.81 ± 1.64	73.97 ± 1.32	79.86 ± 0.55
QLogicE	50.96 ± 0.47	64.75 ± 0.42	61.31 ± 0.40	64.33 ± 0.16	76.23 ± 0.13	93.64 ± 0.09	84.53 ± 1.49	87.87 ± 0.99	87.67 ± 0.52
Graph Neural Networks (GNN and HeteroGNN)									
SAGMM	32.97 ± 1.01	51.75 ± 1.32	82.09 ± 0.86	46.22 ± 1.06	62.56 ± 0.92	91.87 ± 0.51	74.63 ± 9.84	83.28 ± 6.85	95.33 ± 1.08
MTGNN	41.37 ± 4.13	59.39 ± 2.68	76.48 ± 2.44	43.71 ± 1.34	59.90 ± 1.53	90.00 ± 0.70	60.93 ± 6.53	73.87 ± 4.35	92.49 ± 1.36
SpikingHAN	44.31 ± 2.20	61.00 ± 1.54	84.53 ± 0.93	35.67 ± 2.04	54.14 ± 1.10	85.59 ± 1.79	64.94 ± 3.62	76.98 ± 2.30	94.18 ± 0.47
Triadic Completion and Multimodal Models									
ConvNTC	52.72 ± 6.01	67.86 ± 4.76	83.48 ± 0.28	39.65 ± 0.70	58.39 ± 0.59	90.51 ± 0.15	3.55 ± 0.52	11.99 ± 0.81	55.10 ± 1.34
SAFKGC	61.19 ± 4.36	73.04 ± 3.39	87.38 ± 0.98	78.64 ± 3.92	87.05 ± 2.59	97.68 ± 0.36	$\underline{86.99 \pm 1.62}$	$\underline{91.46 \pm 0.98}$	$\underline{96.67 \pm 0.34}$
TriMoGCL	49.38 ± 2.68	65.56 ± 3.38	$\underline{88.86 \pm 0.52}$	79.14 ± 1.00	87.62 ± 0.57	$\underline{97.94 \pm 0.04}$	78.18 ± 0.92	86.28 ± 0.62	96.58 ± 0.31
HySAE	$\underline{71.58 \pm 1.46}$	$\underline{80.58 \pm 1.22}$	75.18 ± 1.15	$\underline{90.68 \pm 0.30}$	$\underline{94.21 \pm 0.21}$	82.90 ± 0.35	82.37 ± 2.07	87.29 ± 1.36	86.87 ± 1.36
QDGL (Ours)	$\mathbf{80.20 \pm 0.45}$	$\mathbf{87.07 \pm 0.34}$	$\mathbf{93.05 \pm 0.54}$	$\mathbf{96.98 \pm 0.08}$	$\mathbf{98.03 \pm 0.04}$	$\mathbf{99.08 \pm 0.02}$	$\mathbf{94.34 \pm 0.35}$	$\mathbf{96.18 \pm 0.34}$	$\mathbf{98.07 \pm 0.20}$
Δ (vs 2nd)	+12.04%	+8.05%	+4.72%	+6.95%	+4.05%	+1.16%	+8.45%	+5.16%	+1.45%

Table 3: Ablation results for QDGL. **Bold** and underline mark the best and second-best results.

Variant	ncRNADrug			CTD			PrimeKG		
	Hit@1	MRR	AUC	Hit@1	MRR	AUC	Hit@1	MRR	AUC
w/o phase	75.64 ± 0.82	83.56 ± 0.51	89.92 ± 1.18	94.68 ± 0.14	96.77 ± 0.07	98.86 ± 0.04	85.06 ± 0.59	89.98 ± 0.41	96.86 ± 0.28
w/o mixed-state fusion	77.04 ± 1.05	84.59 ± 0.62	92.23 ± 0.72	95.04 ± 0.06	96.96 ± 0.03	98.93 ± 0.02	89.36 ± 0.51	92.65 ± 0.42	97.32 ± 0.47
w/o reliability weighting	78.26 ± 0.58	85.46 ± 0.39	91.71 ± 0.87	95.27 ± 0.14	97.12 ± 0.06	98.94 ± 0.04	90.27 ± 0.40	93.21 ± 0.29	97.37 ± 0.25
w/o state normalization	54.35 ± 1.61	68.21 ± 1.26	84.73 ± 0.99	91.99 ± 0.42	95.35 ± 0.21	98.71 ± 0.04	68.28 ± 6.32	78.67 ± 4.55	94.12 ± 0.74
w/o dual path	$\underline{79.46 \pm 0.66}$	$\underline{86.47 \pm 0.42}$	91.06 ± 0.56	$\underline{95.56 \pm 0.13}$	$\underline{97.32 \pm 0.08}$	$\underline{99.05 \pm 0.03}$	$\underline{93.65 \pm 0.45}$	$\underline{95.72 \pm 0.40}$	$\underline{97.69 \pm 0.20}$
fixed random phase	75.61 ± 0.85	83.56 ± 0.60	90.56 ± 1.24	94.98 ± 0.07	96.91 ± 0.03	98.93 ± 0.04	85.24 ± 1.77	89.96 ± 1.00	96.77 ± 0.55
vMF density block	78.73 ± 0.54	85.98 ± 0.42	$\underline{92.31 \pm 0.52}$	92.86 ± 0.13	95.89 ± 0.07	98.71 ± 0.02	93.07 ± 0.69	95.24 ± 0.49	97.64 ± 0.15
spherical cosine	74.16 ± 2.15	82.15 ± 1.50	79.18 ± 0.97	82.89 ± 0.31	89.74 ± 0.19	97.07 ± 0.06	79.46 ± 2.03	85.96 ± 1.43	86.09 ± 0.96
QDGL (Ours)	$\mathbf{80.20 \pm 0.45}$	$\mathbf{87.07 \pm 0.34}$	$\mathbf{93.05 \pm 0.54}$	$\mathbf{96.98 \pm 0.08}$	$\mathbf{98.03 \pm 0.04}$	$\mathbf{99.08 \pm 0.02}$	$\mathbf{94.34 \pm 0.35}$	$\mathbf{96.18 \pm 0.34}$	$\mathbf{98.07 \pm 0.20}$

3.3 Comparison with State-of-the-Art Methods

We compare QDGL with eleven representative baselines spanning knowledge graph embedding methods TransE (Bordes et al., 2013) and RotatE (Sun et al., 2019), quantum-inspired knowledge graph embedding methods QIQE-KGC (Li et al., 2023b) and QLogicE (Chen et al., 2022a), graph neural network methods SAGMM (Meena et al., 2026), MTGNN (Zheng et al., 2025), and SpikingHAN (Cao et al., 2026), and triadic completion and multimodal methods ConvNTC (Liu et al., 2025), SAFKGC (Wang et al., 2026), TriMoGCL (Kang et al., 2025), and HySAE (Li et al., 2025). All methods use the same data splits and evaluation protocol. Further analyses use each category's best baseline by mean performance in Table 2. Baseline settings are provided in Appendix S4.3.

Table 2 shows that QDGL achieves the best classification and ranking results on all three datasets. Baseline strengths differ between these two tasks. HySAE achieves the strongest baseline ranking performance on ncRNADrug and CTD, but its AUC is lower than TriMoGCL's, indicating that prioritizing the true disease within a fixed query and distinguishing positive from negative triads place different demands on a model. QDGL improves both aspects. On ncRNADrug, its Hit@1 improves by 12.04% relative to HySAE, placing the true disease first for more queries, while its AUC exceeds that of the strongest classification baseline, TriMoGCL, by 4.19 percentage points. The accompanying MRR improvement further indicates better placement of true diseases toward the top of the ranking overall. QDGL also leads on all three metrics on CTD and PrimeKG.

3.4 Ablation Study

Table 3 examines the contributions of phase modeling, mixed-state fusion, reliability weighting, state normalization, and dual-path propagation, with two real-valued spherical controls comparing predictive performance across state representations. Variant definitions and loss ablations appear in Appendices S6.2.1 and S6.2.3, respectively.

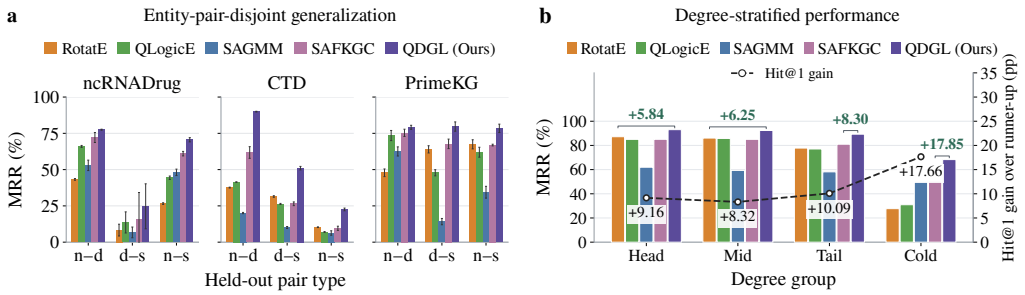


Figure 2: (a) Full-candidate MRR under entity-pair isolation (mean ± SD). n, d, and s denote molecule, drug, and disease, respectively. (b) MRR across weakest-link degree groups. The dashed curve and green labels indicate QDGL's Hit@1 and MRR gains over the strongest baseline, respectively, in percentage points.

State normalization and learnable phase both affect discrimination among candidate diseases. Removing normalization lowers Hit@1 by more than 25 percentage points on both ncRNADrug and PrimeKG. Normalization constrains state scales during fusion and propagation, providing a consistent basis for matching states across nodes. Among the two real-valued spherical controls, the vMF density block, which combines distribution-based fusion with learnable triadic scoring, outperforms spherical cosine, which directly computes pairwise cosine similarities, but still falls below QDGL in ranking performance. The phase ablations further reveal the role of learnable relationships between states. On PrimeKG, setting phase to zero or fixing it to random values lowers Hit@1 by more than 9 percentage points. Both variants retain learnable probability amplitudes, and their shared degradation indicates that adjusting amplitudes alone does not achieve the full model's candidate discrimination. Learnable phase allows relative relationships between states to adapt to the association task, providing additional discriminative information for disease ranking.

Mixed-state fusion preserves the composition of multimodal states, while reliability weighting adjusts their contributions. On PrimeKG, replacing mixed states with a single pure state lowers Hit@1 by 4.98 percentage points, illustrating the value of retaining distinct modality-state information for candidate ranking. Uniform fusion weights also reduce classification and ranking performance on all three datasets, indicating that adapting modality contributions to node attributes improves joint prediction. The dual-path design further combines vector interactions and density-state matching in triadic scoring. Compared with retaining only the vector path, the full model improves every metric on all three datasets, including a 1.99-percentage-point AUC gain on ncRNADrug. The density path thus complements the discriminative information in vector representations, improving both positive–negative association discrimination and prioritization of true diseases.

3.5 Entity-Pair Generalization and Relational Sparsity

We examine QDGL's disease-ranking performance when constituent entity pairs are unseen or individual entities have limited relational information.

Entity-Pair Isolation. Random triad splitting allows constituent pairs to recur across splits. We therefore separately hold out molecular–drug, drug–disease, and molecular–disease pairs, preventing the selected pair type from recurring across training, validation, and test sets. Figure 2(a) shows that QDGL achieves the highest mean full-candidate MRR in all three settings on each dataset. On CTD, its MRR under drug–disease pair isolation exceeds the strongest baseline by 19.53 percentage points. QDGL can thus prioritize diseases without a training-positive record for the target drug–disease pair. The three settings examine generalization to new query combinations and to new disease-related pairings. Detailed results appear in Appendix S6.6.2.

Relational Sparsity Analysis. We group triads by the minimum training-graph degree of their constituent entities to examine prediction when one entity lacks relational information. Figure 2(b) shows that QDGL achieves the highest macro-averaged MRR in all four groups, with larger advantages toward lower-degree groups. MRR is averaged equally across the three datasets within each degree group. In the Cold group, where at least one entity has no training-graph neighbors, its

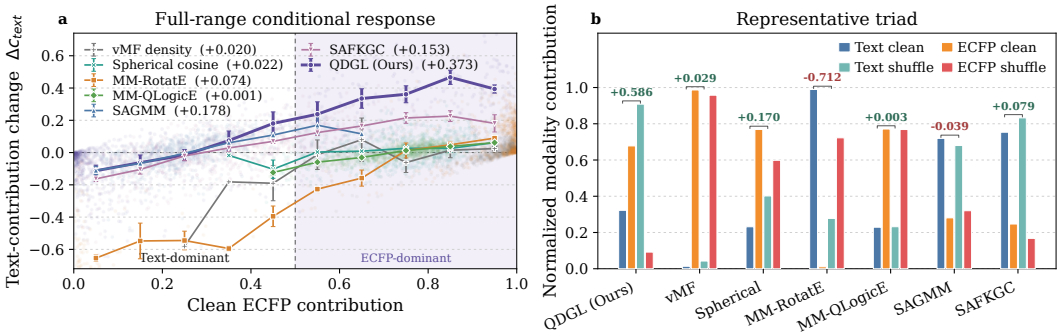


Figure 3: Modality responses to ECFP shuffling on ncRNADrug. (a) Text-contribution changes versus initial ECFP contributions. (b) Contributions before and after shuffling for a representative triad, with text-contribution changes annotated.

MRR exceeds the strongest baseline by 17.85 percentage points. Here, the model cannot directly aggregate relational information from that entity's neighbors and must use entity attributes and the context of the other entities in the triad for prediction. QDGL's leading performance in this group demonstrates its ability to distinguish candidate diseases under incomplete relational information.

3.6 Modality Contribution Responses to Perturbation

To examine how models adjust their reliance on different modalities, we randomly permute observed ECFP features across drugs on ncRNADrug, preserving observation masks and the marginal feature distribution while disrupting the correspondence between structural features and drugs. Figure 3 compares population-level responses and a representative case. MM-RotatE and MM-QLogicE denote adaptations with multimodal encoding and reliability-weighted fusion.

After ECFP perturbation, QDGL increases text contributions in cases that initially rely primarily on structural information. In Figure 3(a), its mean text contribution increases by 0.3726, with positive changes in 95.3% of these observations. Both the mean increase and the positive-response fraction are the highest among the compared methods. The adjustment is thus substantial in magnitude and widespread among initially ECFP-dominant cases. The representative triad in Figure 3(b) further illustrates this response: QDGL's text contribution increases by 0.5859, shifting the prediction from structural to textual dominance. When the correspondence between structural features and drugs is disrupted, the model reduces its relative reliance on that modality and increases the contribution of unperturbed text information.

Responses under different initial modality reliance are analyzed in Appendix S6.4.1, and association case studies across the three datasets are presented in Appendix S6.7.

4 Conclusion

This work introduces QDGL to address difficulties in fusing heterogeneous modalities and predicting biomedical triadic associations with missing or noisy features. Hilbert-sphere encoding unifies modality representations, while reliability-weighted fusion adjusts available modalities' contributions to construct both vectors and mixed density states. Vector-based neighborhood aggregation and density-state matching jointly exploit entity features and graph relations for triadic classification and disease ranking. QDGL achieves the best mean classification and ranking results among the compared methods across three biomedical benchmarks, with strong performance under unseen-pair splits, sparse relations, missing modalities, and structural feature perturbations. Ablations identify component contributions, while dual-path analyses reveal complementary roles of the two representations in candidate ranking. QDGL offers an effective approach to prioritizing molecular–drug–disease associations when modality features are missing or unreliable.

REFERENCES

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

Buqing Cao, Qian Peng, Xiang Xie, Liang Chen, Min Shi, and Jianxun Liu. Spiking heterogeneous graph attention networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 19853–19861, 2026.

Xinyu Cao, Xu Zhou, Fei Hou, Yu-e Huang, Mengqin Yuan, Min Long, Sina Chen, Wanyue Lei, Jicun Zhu, Jiahao Chen, et al. ncrnadrug: a database for validated and predicted ncrnas associated with drug resistance and targeted by drugs. *Nucleic Acids Research*, 52(D1):D1393–D1399, 2024.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific data*, 10(1):67, 2023.

Abhra Chaudhuri, Anjan Dutta, Tu Bui, and Serban Georgescu. A closer look at multimodal representation collapse. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 7555–7577. PMLR, 13–19 Jul 2025.

Panfeng Chen, Yisong Wang, Xiaomin Yu, and Renyan Feng. Qlogice: Quantum logic empowered embedding for knowledge graph completion. *Knowledge-Based Systems*, 239:107963, 2022a.

Xiang Chen, Ningyu Zhang, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, and Huajun Chen. Hybrid transformer with multi-level fusion for multimodal knowledge graph completion. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pp. 904–915, 2022b.

Allan Peter Davis, Thomas C Wiegers, Daniela Sciaky, Fern Barkalow, Melissa Strong, Brent Wyatt, Jolene Wiegers, Roy McMorran, Sakib Abrar, and Carolyn J Mattingly. Comparative toxicogenomics database's 20th anniversary: update 2025. *Nucleic acids research*, 53(D1):D1328–D1334, 2025.

Jeffrey A Engelman, Kreshnik Zejnullahu, Tetsuya Mitsudomi, Youngchul Song, Courtney Hyland, Joon Oh Park, Neal Lindeman, Christopher-Michael Gale, Xiaojun Zhao, James Christensen, et al. Met amplification leads to gefitinib resistance in lung cancer by activating erbb3 signaling. *science*, 316(5827):1039–1043, 2007.

Riddhi S Gupta, Carolyn E Wood, Teyl Engstrom, Jason D Pole, and Sally Shrapnel. A systematic review of quantum machine learning for digital health. *NPJ digital medicine*, 8(1):237, 2025.

Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N Nadkarni, Benjamin S Glicksberg, Nils Gehlenborg, and Marinka Zitnik. A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30(12):3601–3613, 2024.

Minoru Kanehisa, Miho Furumichi, Yoko Sato, Masayuki Kawashima, and Mari Ishiguro-Watanabe. Kegg for taxonomy-based analysis of pathways and genomes. *Nucleic acids research*, 51(D1): D587–D592, 2023.

Chuanze Kang, Zonghuan Liu, and Han Zhang. A comprehensive graph neural network method for predicting triplet motifs in disease–drug–gene interactions. *Bioinformatics*, 41(2):btaf023, 2025.

Amin Li, Weiya Cao, Xueke Liu, Yinci Zhang, Yongfang Ma, Ruyue Xu, Rongbo Zhang, Xinkuang Liu, Shuping Zhou, Ruikai Wang, et al. Gefitinib sensitization of cisplatin-resistant wild-type egfr non-small cell lung cancer cells. *Journal of Cancer Research and Clinical Oncology*, 146 (7):1737–1749, 2020.

Jipeng Li, Shanfeng Li, Zhe Chen, Jianhua Wang, Ying Chen, Zhengyang Xu, Mingwei Jin, and Wanjun Yu. mir-326 reverses chemoresistance in human lung adenocarcinoma cells by targeting specificity protein 1. *Tumor Biology*, 37(10):13287–13294, 2016.

Juanhui Li, Harry Shomer, Haitao Mao, Shenglai Zeng, Yao Ma, Neil Shah, Jiliang Tang, and Dawei Yin. Evaluating graph neural networks for link prediction: Current pitfalls and new benchmarking. *Advances in neural information processing systems*, 36:3853–3866, 2023a.

Linyu Li, Xuan Zhang, Zhi Jin, Chen Gao, Rui Zhu, Yuqin Liang, and Yubing Ma. Knowledge graph completion method based on quantum embedding and quaternion interaction enhancement. *Information Sciences*, 648:119548, 2023b.

Zhao Li, Xin Wang, Jun Zhao, Feng Feng, Zirui Chen, and Jianxin Li. Hysae: an efficient semantic-enhanced representation learning model for knowledge hypergraph link prediction. In *Proceedings of the ACM on Web Conference 2025*, pp. 86–97, 2025.

Pei Liu, Xiao Liang, Yue Li, and Jiawei Luo. Convntc: convolutional neural tensor completion for detecting "a–a–b" type biological triplets. *Briefings in Bioinformatics*, 26(4):bbaf372, 2025.

Yu Liu, Fei-Ya Du, Wei Chen, Pei-Fen Fu, Min-Ya Yao, and Shu-Sen Zheng. G15 sensitizes epithelial breast cancer cells to doxorubicin by preventing epithelial-mesenchymal transition through inhibition of gpr30. *American Journal of Translational Research*, 7(5):967–975, 2015.

Brian Madajewski, Michael A Boatman, Ivan Martinez, Julia H Carter, and Erik A Bey. Nad (p) h quinone oxidoreductase-1 expression promotes self-renewal and therapeutic resistance in non-small cell lung cancer. *Genes*, 14(3):607, 2023.

Makoto Maemondo, Akira Inoue, Kunihiko Kobayashi, Shunichi Sugawara, Satoshi Oizumi, Hiroshi Isobe, Akihiko Gemma, Masao Harada, Hirohisa Yoshizawa, Ichiro Kinoshita, et al. Gefitinib or chemotherapy for non–small-cell lung cancer with mutated egfr. *New England journal of medicine*, 362(25):2380–2388, 2010.

Mohit Meena, Yash Punjabi, Vishal Sharma, Mahesh Chandran, et al. Self-adaptive graph mixture of models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 24344–24352, 2026.

Keichin NG. Ensemble learning methods and heterogeneous graph network fusion: building drug-gene-disease triple association prediction models. *Briefings in Bioinformatics*, 26(4):bbaf369, 2025.

Pablo Perdomo-Quinteiro and Alberto Belmonte-Hernández. Knowledge graphs for drug repurposing: a review of databases and methods. *Briefings in bioinformatics*, 25(6):bbae461, 2024.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.

Chao-Yue Sun, Ying Zhu, Xiao-Feng Li, Xie-Qi Wang, Li-Peng Tang, Zu-Qing Su, Cai-Yun Li, Guang-Juan Zheng, and Bing Feng. Scutellarin increases cisplatin-induced apoptosis and autophagy to overcome cisplatin resistance in non-small cell lung cancer via erk/p53 and c-met/akt signaling pathways. *Frontiers in pharmacology*, 9:92, 2018.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2019.

Zhenchao Tang, Guanxing Chen, Shouzhi Chen, Jianhua Yao, Linlin You, and Calvin Yu-Chian Chen. Modal-nexus auto-encoder for multi-modality cellular data integration and imputation. *Nature Communications*, 15(1):9021, 2024.

Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In *International conference on machine learning*, pp. 9448–9457. PMLR, 2020.

Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. In *International Conference on Learning Representations*, 2018.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, volume 48, pp. 2071–2080. PMLR, 2016.

Lei Wang, Richong Zhang, Junfan Chen, Jaein Kim, Wei Xiao, and Hui Liu. Revealing multimodal information trade-off in multimodal knowledge graph completion. *Expert Systems with Applications*, 322:132003, 2026.

Zhankun Xiong, Shichao Liu, Feng Huang, Ziyan Wang, Xuan Liu, Zhongfei Zhang, and Wen Zhang. Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 5339–5347, 2023.

Chenchen Zhang, Qiuchi Li, Dawei Song, and Prayag Tiwari. Quantum-inspired semantic matching based on neural networks with the duality of density matrices. *Engineering Applications of Artificial Intelligence*, 140:109667, 2025a.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32, 2019.

Xuelin Zhang, Jie Zhu, Ruiyun Xing, Yi Tie, Hanjiang Fu, Xiaofei Zheng, and Bingxiang Yu. mir-513a-3p sensitizes human lung adenocarcinoma cells to chemotherapy by targeting gstp1. *Lung cancer*, 77(3):488–494, 2012.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu, Wen Zhang, and Huajun Chen. Native: Multi-modal knowledge graph completion in the wild. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pp. 91–101, 2024.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Binbin Hu, Ziqi Liu, Wen Zhang, Huajun Chen, et al. Multiple heads are better than one: Mixture of modality knowledge experts for entity representation learning. In *International Conference on Learning Representations*, volume 2025, pp. 54811–54828, 2025b.

Lidan Zheng, Simeng Zhang, Yihao Li, Yang Liu, Qian Ge, Lingxi Gu, Yu Xie, Xiao Wang, Yunfei Ma, Junfei Liu, et al. Mtgnn: A drug–target–disease triplet association prediction model based on multimodal heterogeneous graph neural networks and direction-aware metapaths. *Journal of Chemical Information and Modeling*, 65(12):5921–5933, 2025.

APPENDIX

APPENDIX CONTENTS

S1 Statements **14**
S1.1 AI Use Statement 14
S1.2 Ethics Statement 14
S1.3 Reproducibility Statement 14

S2 Related Work **14**
S2.1 Multimodal Biomedical Association Prediction and Higher-Order Relational Learning 14
S2.2 Complex-Valued and Quantum-Inspired Representation Learning 14

S3 Dataset Description and Preprocessing **15**

S4 Reproducibility Details **16**
S4.1 Computing Environment 16
S4.2 Training Hyperparameters 16
S4.3 Baseline Settings 17
S4.4 Evaluation Metrics 17

S5 Algorithmic Procedure, Computational Complexity, and Theoretical Properties **18**
S5.1 Training Procedure 18
S5.2 Computational Complexity Analysis 20
S5.3 Theoretical Properties and Proofs 21

S6 Supplementary Experiments and Analyses **23**
S6.1 Supplementary Analysis of Ranking and Discrimination 23
S6.2 Component Ablations and Dual-Path Analysis 25
S6.3 Dual-Path Robustness under Missing Information 29
S6.4 Modality Reliability and Perturbation Responses 31
S6.5 Assessing Sensitivity to False Negatives during Training 34
S6.6 Analyzing Entity-Pair Overlap and Generalization 34
S6.7 Triadic Association Case Studies 41

S7 Limitations **45**

S1 STATEMENTS

S1.1 AI USE STATEMENT

Generative AI tools were used solely for language polishing and improving the clarity and readability of the manuscript. All AI-assisted edits were reviewed by the authors. The authors take full responsibility for the final content of the manuscript.

S1.2 ETHICS STATEMENT

This study uses publicly available biomedical association data and does not involve direct intervention with human participants. The model outputs are intended for computational hypothesis prioritization and should not be interpreted as clinical recommendations without independent experimental and clinical validation.

S1.3 REPRODUCIBILITY STATEMENT

The main paper specifies the model, optimization objectives, datasets, evaluation metrics, and comparison protocol. The appendix provides feature preprocessing, computing environment, training hyperparameters, baseline adaptation, metric definitions, ablation details, complexity analysis, and additional case studies. Anonymous source code and configuration files will be supplied as supplementary material for verification.

S2 RELATED WORK

S2.1 MULTIMODAL BIOMEDICAL ASSOCIATION PREDICTION AND HIGHER-ORDER RELATIONAL LEARNING

Multimodal biomedical association prediction combines entity attributes and relational structure to identify potential associations in heterogeneous data. Knowledge graphs organize drug, molecular, and disease information from different sources into relational networks, providing a basis for data integration and association inference in drug repurposing (Perdomo-Quinteiro & Belmonte-Hernández, 2024). In multimodal knowledge graph completion, transformer-based fusion combines entity side information with structural information to predict missing links (Chen et al., 2022b), while research on incomplete-modality settings examines how uneven information availability affects prediction (Zhang et al., 2024). Biomedical graph models also combine molecular attributes with multi-relational neighborhoods, for example by using drug molecular-graph features and drug interaction structures to predict interaction events (Xiong et al., 2023). These studies use feature fusion and graph propagation to convert multi-source information into entity representations for association prediction.

Triadic association prediction concerns joint associations among molecules, drugs, and diseases. Prior work extracts node embeddings through relational graph convolution and combines them with ensemble learning to predict drug–gene–disease associations (NG, 2025). This task raises two connected questions: how multimodal information forms node representations and how the three entity types participate in joint scoring. QDGL introduces density representations alongside vector representations, mixing modality-specific states with learned weights to describe each node's modality composition. Using the propagated representations, the model constructs triadic interactions conditioned on the distinct roles of molecules, drugs, and diseases, connecting modality fusion with joint association modeling within one framework.

S2.2 COMPLEX-VALUED AND QUANTUM-INSPIRED REPRESENTATION LEARNING

Complex-valued and hypercomplex representations offer approaches to structured relational modeling by extending vector spaces and their operations. ComplEx models asymmetric relations using complex embeddings and Hermitian products (Trouillon et al., 2016). Complex neural networks apply complex-valued operations to deep feature extraction through corresponding network layers and learning mechanisms (Trabelsi et al., 2018). Quaternion embeddings use hypercomplex operations

to describe interactions between entities and relations (Zhang et al., 2019). These methods primarily concern the coordinate forms of vector representations and their interaction operations. Density representations provide a way to combine multiple states. QDGL encodes individual modalities as states, forms node density operators through weighted mixing, and updates these representations during relational propagation.

Quantum machine learning studies predictive models through feature mappings and state operations in Hilbert spaces, and its applications to digital health have been systematically reviewed (Gupta et al., 2025). Hilbert-space feature methods encode inputs as states and connect state inner products with kernel methods (Schuld & Killoran, 2019). Quantum-enhanced feature maps further explore this representation approach for supervised classification (Havlíček et al., 2019). QDGL applies related representation ideas to multimodal heterogeneous graphs, propagating node density states through relation-conditioned transformations and neighborhood aggregation, then obtaining triadic scoring features through measurement readout. The quantum-inspired and classical vector paths jointly contribute to association prediction, producing scores through density-state interactions and vector-feature interactions, respectively.

S3 Dataset Description and Preprocessing

We evaluate QDGL on three biomedical triadic association benchmarks with the following positive-triad definitions.

ncRNADrug. We extract ncRNAs, drugs, and the corresponding diseases from manually curated drug-response records, standardize names and identifiers, and remove duplicate triads. A positive triad indicates an association between an ncRNA and drug response in the corresponding disease context, covering both sensitivity and resistance records. The prediction task concerns the presence of an association without distinguishing the direction of drug response.

CTD. We construct triads by combining human chemical–gene relations with chemical–disease and gene–disease relations annotated with direct evidence. A triad is included in the positive set when all three constituent relations are present. A positive thus represents a chemical–gene–disease combination supported by its constituent pairwise relations. The construction does not require the three relations to be jointly validated in the same study.

PrimeKG. We extract closed triads from drug–protein, protein–disease, and drug–disease relations. Drug–protein relations include target, enzyme, transporter, and carrier relations. Protein–disease relations use association records, and drug–disease relations include indications and off-label use, excluding contraindications. Combinations with all three relations present are labeled positive, with the selected relation types collapsed into a single association label for triad prediction.

For each dataset, entity identifiers are canonicalized within each entity type, duplicated triads are removed, and entity vocabularies are constructed before cross-validation. Only triads whose three entities can be mapped to the corresponding vocabularies are retained. The graph encoder uses only the training portion of each fold to construct relational neighborhoods.

Feature preprocessing is performed by entity type. For sequence entities, raw sequences are first canonicalized and then converted into entity-level embeddings with pretrained sequence models. ncRNA sequences are split into fixed-length windows when they exceed the RNA-FM length budget. Each window is passed through RNA-FM, token representations are averaged into a 640-dimensional window vector, and window vectors belonging to the same ncRNA are averaged to obtain one sequence embedding. Protein sequences and genes that can be mapped to proteins are processed with ESM2. Long sequences are divided into overlapping chunks, final-layer residue representations are averaged after removing special tokens, and chunk embeddings are averaged into a 1280-dimensional entity representation.

For text modalities, drug descriptions and disease definitions are normalized and encoded by PubMedBERT into 768-dimensional sentence embeddings, which serve as textual modality vectors. Empty descriptions are assigned a zero vector and a missing mask. For chemical structure, canonical SMILES strings are parsed by RDKit and converted into 2048-bit Morgan fingerprints with radius 2, corresponding to ECFP4. Molecules without valid SMILES receive a zero fingerprint and

a missing structural mask. For disease structure, ontology identifiers are connected through MeSH or MONDO hierarchy edges, Node2Vec is trained on the resulting disease-related ontology graph, and ontology-code vectors linked to the same disease are averaged into a 128-dimensional structural feature.

Table S1 summarizes the encoder and output dimensionality for each modality.

Table S1: Feature encoders and output dimensions used in the three benchmarks.

Feature type	Raw input	Encoder	Dimension
ncRNA sequence	Canonicalized RNA sequence	RNA-FM	640
Protein or gene sequence	Protein sequence or mapped gene product	ESM2	1280
Drug or disease text	Drug descriptions and disease definitions	PubMedBERT	768
Chemical structure	Drug or chemical SMILES	RDKit ECFP4	2048
Disease ontology	MeSH or MONDO hierarchy codes	Node2Vec	128

Entity features are represented by dense vectors with binary observation masks. A mask value of one indicates an available feature, while zero indicates an unavailable feature. Modality-specific projection layers map observed features into the common state dimension, and reliability-weighted mixed-state fusion normalizes their contributions over available modalities.

S4 Reproducibility Details

S4.1 Computing Environment

Table S2 lists the hardware and software environment, and Table S3 summarizes the computational resources used by QDGL.

Table S2: Experimental computing environment.

Component	Configuration
GPU	NVIDIA A40 GPUs
CPU	Intel Xeon Gold 6334, 3.60GHz, 32 logical cores
Host memory	256 GB
GPU memory	48 GB per A40
Operating system	Linux
Deep learning framework	PyTorch and PyTorch Geometric
Feature tools	RDKit, RNA-FM, ESM2, PubMedBERT, Node2Vec
Evaluation tools	scikit-learn and NumPy

Table S3: Computational resources of QDGL. Training GPU hours sum training time across five folds, excluding setup and evaluation.

Dataset	Peak GPU memory (GB)	Parameters (M)	Forward GFLOPs	Training GPU hours
ncRNADrug	9.8	3.7	18.9	2.0
CTD	31.2	3.8	40.3	20.9
PrimeKG	3.2	3.8	6.0	0.5

S4.2 Training Hyperparameters

Table S4 lists the training hyperparameters of QDGL on the three datasets. Models are selected using the mean validation score over AUC, Hit@1, and MRR.

Table S4: Training hyperparameters of QDGL. Gray rows indicate identical settings across the three datasets.

Hyperparameter	ncRNADrug	CTD	PrimeKG
Cross-validation (folds)	5	5	5
Validation fraction	10%	10%	10%
Optimizer	AdamW	AdamW	AdamW
Learning rate	1.78×10^{-4}	2.33×10^{-4}	1.95×10^{-4}
Weight decay	8.30×10^{-5}	1.54×10^{-4}	7.38×10^{-5}
Maximum epochs	500	200	500
Validation interval (epochs)	5	10	5
Early-stopping patience (checks)	12	20	20
Complex-state dimension	160	160	160
Real-valued dimension	320	320	320
Graph propagation layers	2	2	2
Attention heads per relation	1	1	1
Hidden dimension	96	96	96
Measurements	48	48	48
Tucker core rank	16	16	16
Model dropout	0.04	0.05	0.07
Edge dropout	0.21	0.20	0.14
Message dropout	0.07	0.10	0.06
Triadic batch size	8192	16384	4096
Triadic negatives per positive	1	1	1
Evaluation ranking negatives per query	50	50	50
Evaluation ranking batch size	4096	8192	4096
Training ranking negatives per positive	1	1	1
Base-view classification weight ω_1	1.00	1.00	1.00
Perturbed-view classification weight ω_2	1.17	1.10	0.91
Ranking loss weight	0.14	0.10	0.14
Ranking warmup (epochs)	15	30	20
Siamese consistency weight	0.09	0.15	0.10
NCE loss weight	0.03	0.03	0.03
Prototype loss weight	0.09	0.10	0.07
Density-view loss weight	1.21×10^{-3}	5.00×10^{-3}	2.18×10^{-3}
NCE temperature τ_{nce}	0.10	0.10	0.10
Prototype temperature τ_{proto}	0.20	0.20	0.20

S4.3 Baseline Settings

All baselines share the data splits, positive-filtering rules, and ranking candidates with QDGL, with 50 negative disease candidates per positive query. Each method scores candidate diseases given a molecular–drug context.

ID-based knowledge graph embedding methods use entity identifiers. Graph models use the graph constructed from the training fold, and methods supporting multimodal inputs use the same pre-trained features. Hyperparameters are selected on the validation set, with method-specific settings guided by the original implementations where available.

For extended analyses, we equally average the nine mean scores in Table 2, covering Hit@1, MRR, and AUC across three datasets, and select the highest-scoring baseline within each category. This yields RotatE, QLogicE, SAGMM, and SAFKGC, with aggregate means of 73.75%, 74.59%, 68.97%, and 84.46%, respectively.

S4.4 Evaluation Metrics

We use AUC to evaluate binary discrimination and Hit@1 and MRR to evaluate disease ranking.

At a prediction threshold τ, let $\mathrm{TP}(\tau)$, $\mathrm{FP}(\tau)$, $\mathrm{TN}(\tau)$, and $\mathrm{FN}(\tau)$ denote the numbers of true positives, false positives, true negatives, and false negatives, respectively. The true-positive and false-positive rates are

$$\mathrm{TPR}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FN}(\tau)}, \qquad \mathrm{FPR}(\tau) = \frac{\mathrm{FP}(\tau)}{\mathrm{FP}(\tau) + \mathrm{TN}(\tau)}. \tag{S1}$$

Varying τ yields the ROC curve, whose area is

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(u)\,\mathrm{d}u, \qquad u = \mathrm{FPR}. \tag{S2}$$

For sampled ranking, each query contains one held-out positive disease and 50 sampled negative disease candidates. Let $\mathcal{Q}$ be the test query set and r_q the rank of the positive disease for query q, starting from one. Hit@1 and MRR are defined as

$$\mathrm{Hit@1} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \mathbb{I}(r_q = 1), \qquad \mathrm{MRR} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \frac{1}{r_q}, \tag{S3}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Metrics are computed separately on each test fold and reported as five-fold means and standard deviations.

S5 Algorithmic Procedure, Computational Complexity, and Theoretical Properties

S5.1 Training Procedure

Algorithm S1 summarizes QDGL training. The two graph views share all model parameters. The superscript $[a]$ identifies the view, while (m) and (l) denote modality and propagation layer, respectively.

S5.1.1 Training Objectives

QDGL jointly optimizes association classification, disease ranking, and consistency at the modality, node, class-prototype, and density-state levels. The base and perturbed graph views are denoted by G_1 and G_2, with shared model parameters. For a labeled batch $\mathcal{B}$, triad $\chi = (n, d, s)$ has label $y_\chi \in \{0, 1\}$ and fused score $\widetilde{y}_\chi^{[a]} = \widetilde{y}_{h,\chi}^{[a]} + g\widetilde{y}_{\rho,\chi}^{[a]}$ in view $a \in \{1, 2\}$. Its predicted probability is $\widehat{y}_\chi^{[a]} = \sigma(\widetilde{y}_\chi^{[a]})$.

Triad classification. Binary cross-entropy supervises association prediction in both views:

$$\mathcal{L}_C = -\frac{1}{|\mathcal{B}|} \sum_{\chi \in \mathcal{B}} \sum_{a=1}^{2} \omega_a \left[y_\chi \log \widehat{y}_\chi^{[a]} + (1 - y_\chi) \log(1 - \widehat{y}_\chi^{[a]}) \right], \tag{S4}$$

where ω_1 and ω_2 weight the base- and perturbed-view classification losses, respectively.

Disease ranking. For each positive triad $\chi = (n, d, s) \in \mathcal{B}^+$, one negative disease s_χ^- is sampled with (n, d) fixed. The same negative is used in both views. The ranking objective is

$$\mathcal{L}_R = \frac{1}{2|\mathcal{B}^+|} \sum_{\chi=(n,d,s) \in \mathcal{B}^+} \sum_{a=1}^{2} \log\Big(1 + \exp\Big[\widetilde{y}^{[a]}(n, d, s_\chi^-) - \widetilde{y}^{[a]}(n, d, s)\Big]\Big), \tag{S5}$$

where $\mathcal{B}^+$ is the positive subset of $\mathcal{B}$. This objective encourages the observed disease to score above the negative disease, with $\mathcal{L}_R = 0$ when $\mathcal{B}^+$ is empty.

Shared similarity for auxiliary objectives. Modality consistency, node contrast, and prototype alignment use normalized Born similarity. For a real-valued representation $\mathbf{x} = [\mathbf{x}_\mathrm{R}; \mathbf{x}_\mathrm{I}] \in \mathbb{R}^{2d_\mathrm{aux}}$, define $\widehat{\mathbf{x}} = (\mathbf{x}_\mathrm{R} + i\mathbf{x}_\mathrm{I})/\sqrt{\|\mathbf{x}\|_2^2 + \epsilon}$. The similarity is

$$s_\mathrm{B}(\mathbf{x}, \mathbf{y}) = \left|\widehat{\mathbf{x}}^\dagger \widehat{\mathbf{y}}\right|^2, \tag{S6}$$

where $\mathbf{x}_\mathrm{R}$ and $\mathbf{x}_\mathrm{I}$ define the d_aux-dimensional real and imaginary components, $\dagger$ denotes conjugate transpose, and ϵ is a small positive constant that stabilizes normalization.

Algorithm S1 QDGL Training Procedure.

Require: Training triads $\mathcal{T}_{\mathrm{tr}}^{+}$, validation data, modality features $\mathbf{u}_v^{(m)}$, available modalities $\mathcal{M}_v$, epoch budget T, ranking warm-up E_w
Ensure: Parameters $\theta^{\star}$ selected by validation performance

Construct graph G from $\mathcal{T}_{\mathrm{tr}}^{+}$ and initialize θ
Initialize the best validation score to $-\infty$
for $t = 1, \ldots, T$ **do**
  Corrupt training triads, exclude only $\mathcal{T}_{\mathrm{tr}}^{+}$, and form labeled batches
  Set $\lambda_R(t) = 0$ for $t \le E_w$, and $\lambda_R(t) = \lambda_R$ otherwise
  for each labeled batch $\mathcal{B}$ **do**
    Set $G_1 = G$ and obtain G_2 by relation-wise edge dropout
    for each view $a \in \{1, 2\}$ **do**
      Compute $\boldsymbol{\psi}_v^{(m)}$ and $\pi_v^{(m)}$ using shared encoders ▷ Eqs. (1)–(3)
      Initialize density states $\rho_v^{(0)}$ and vectors $\mathbf{h}_v^{(0)}$ ▷ Eqs. (4)–(5)
      Enable message dropout in both propagation paths only when $a = 2$
      Propagate vectors on G_a for L_v layers to obtain $\mathbf{h}_v^{[a]}$ ▷ Eq. (6)
      Propagate density states for L_ρ layers with unitary transport and overlap attention
      Apply $\mathcal{C}_{R_c}$ at each layer to retain rank at most R_c ▷ Eqs. (7)–(9)
      Fuse propagated and initial density states to obtain $\rho_v^{[a]}$ ▷ Eq. (10)
      Compute measurement features $\mathbf{q}_v^{[a]}$ from $\rho_v^{[a]}$ ▷ Eq. (12)
      Compute $\widetilde{y}_{h,\chi}^{[a]}$ and $\widetilde{y}_{\rho,\chi}^{[a]}$ for each $\chi = (n, d, s) \in \mathcal{B}$ ▷ Eqs. (11), (13)
      Fuse scores $\widetilde{y}_{\chi}^{[a]} = \widetilde{y}_{h,\chi}^{[a]} + g\widetilde{y}_{\rho,\chi}^{[a]}$
      Obtain probabilities $\widehat{y}_{\chi}^{[a]} = \sigma(\widetilde{y}_{\chi}^{[a]})$ ▷ Eq. (14)
    end for
    Compute $\mathcal{L}_C$ from $\widehat{y}_{\chi}^{[1]}$, $\widehat{y}_{\chi}^{[2]}$, and labels y_χ ▷ Eq. (S4)
    Set $\mathcal{L}_R = 0$
    if $\lambda_R(t) > 0$ and the positive subset $\mathcal{B}^{+}$ is nonempty **then**
      For each $\chi = (n, d, s) \in \mathcal{B}^{+}$, sample s_χ^{-} with $(n, d, s_\chi^{-}) \notin \mathcal{T}_{\mathrm{tr}}^{+}$
      Use the same s_χ^{-} in both views to compute $\mathcal{L}_R$ ▷ Eq. (S5)
    end if
    Compute auxiliary losses $\mathcal{L}_S$, $\mathcal{L}_N$, $\mathcal{L}_P$, and $\mathcal{L}_\rho$ ▷ Appendix S5.1.1
    Combine losses using Eq. (15) and update θ by backpropagation
  end for
  if validation is scheduled **then**
    Evaluate on G without perturbations using $(\mathrm{AUC} + \mathrm{Hit@1} + \mathrm{MRR})/3$
    Save $\theta^{\star} \leftarrow \theta$ if the validation score improves
    Stop if the early-stopping criterion is met
  end if
end for
return $\theta^{\star}$

Modality-branch consistency. In the base view, each modality state of a drug or disease is represented as $\mathbf{x}_v^{(m)} = \mathrm{Re}(\psi_v^{(m)}) \oplus \mathrm{Im}(\psi_v^{(m)})$. Type-specific projection and prediction heads produce $\mathbf{z}_{\mathrm{proj},v}^{(m)} = f_{\mathrm{proj},\tau(v)}(\mathbf{x}_v^{(m)})$ and $\mathbf{p}_v^{(m)} = f_{\mathrm{pred},\tau(v)}(\mathbf{z}_{\mathrm{proj},v}^{(m)})$. The two modalities share these heads within each entity type. Their bidirectional prediction loss is

$$\mathcal{L}_S = -\sum_{\tau \in \{\mathrm{drug},\mathrm{disease}\}} \frac{1}{2|\mathcal{V}_\tau|} \sum_{v \in \mathcal{V}_\tau} \Big[s_{\mathrm{B}}\Big(\mathbf{p}_v^{(1)}, \mathrm{sg}(\mathbf{z}_{\mathrm{proj},v}^{(2)})\Big) + s_{\mathrm{B}}\Big(\mathbf{p}_v^{(2)}, \mathrm{sg}(\mathbf{z}_{\mathrm{proj},v}^{(1)})\Big) \Big], \tag{S7}$$

where $\mathcal{V}_\tau$ is the node set of type τ and $\mathrm{sg}(\cdot)$ stops gradients through the target representation. This objective coordinates modality branches through mutual prediction.

Cross-view node contrast. The representations of the same node across graph views form a positive pair, while other nodes of the same type serve as contrastive alternatives:

$$\mathcal{L}_N = -\frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}}\frac{1}{|\mathcal{U}_\tau|}\sum_{v\in\mathcal{U}_\tau}\log\frac{\exp\Big(s_{\mathrm{B}}(\mathbf{h}_v^{[1]},\mathbf{h}_v^{[2]})/T_N\Big)}{\sum_{u\in\mathcal{U}_\tau}\exp\Big(s_{\mathrm{B}}(\mathbf{h}_v^{[1]},\mathbf{h}_u^{[2]})/T_N\Big)}, \tag{S8}$$

where $\mathcal{T} = \{\mathrm{molecular}, \mathrm{drug}, \mathrm{disease}\}$, $\mathbf{h}_v^{[a]}$ is the propagated vector representation in view a, $\mathcal{U}_\tau \subseteq \mathcal{V}_\tau$ is the node set used for cross-view contrast, and T_N is a positive temperature parameter. The base view anchors the contrast against the perturbed view.

Cross-view prototype consistency. Class-level alignment uses triad interaction features constructed from the vector representations:

$$\begin{aligned}\mathbf{f}_\chi^{[a]} = {}& \mathbf{h}_n^{[a]} \oplus \mathbf{h}_d^{[a]} \oplus \mathbf{h}_s^{[a]} \\ & \oplus |\mathbf{h}_n^{[a]} - \mathbf{h}_d^{[a]}| \oplus |\mathbf{h}_d^{[a]} - \mathbf{h}_s^{[a]}| \oplus |\mathbf{h}_n^{[a]} - \mathbf{h}_s^{[a]}| \\ & \oplus (\mathbf{h}_n^{[a]} \odot \mathbf{h}_d^{[a]}) \oplus (\mathbf{h}_d^{[a]} \odot \mathbf{h}_s^{[a]}) \oplus (\mathbf{h}_n^{[a]} \odot \mathbf{h}_s^{[a]}),\end{aligned} \tag{S9}$$

where $\oplus$ and $\odot$ denote concatenation and element-wise multiplication. For class $c \in \{0, 1\}$, the within-batch prototype is

$$\boldsymbol{\mu}_c^{[a]} = \frac{1}{|\mathcal{B}_c|}\sum_{\chi\in\mathcal{B}_c}\mathbf{f}_\chi^{[a]}, \qquad \mathcal{B}_c = \{\chi \in \mathcal{B} : y_\chi = c\}, \tag{S10}$$

with a zero prototype when $\mathcal{B}_c$ is empty. Prototype consistency is defined by cross-view class matching:

$$\mathcal{L}_P = -\frac{1}{2}\sum_{c\in\{0,1\}}\log\frac{\exp\Big(s_{\mathrm{B}}(\boldsymbol{\mu}_c^{[1]},\boldsymbol{\mu}_c^{[2]})/T_P\Big)}{\sum_{c'\in\{0,1\}}\exp\Big(s_{\mathrm{B}}(\boldsymbol{\mu}_c^{[1]},\boldsymbol{\mu}_{c'}^{[2]})/T_P\Big)}, \tag{S11}$$

where T_P is a positive temperature parameter. This objective matches same-class prototypes across views while distinguishing positive from negative classes. Both node contrast and prototype consistency propagate gradients through the representations of both views.

Density-state consistency. The final density states used by the quantum-inspired scoring path are aligned across graph views:

$$\mathcal{L}_\rho = \frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}}\frac{1}{|\mathcal{W}_\tau|}\sum_{v\in\mathcal{W}_\tau}\left[1 - \frac{\mathrm{Tr}(\rho_v^{[1]}\rho_v^{[2]})}{\|\rho_v^{[1]}\|_F\|\rho_v^{[2]}\|_F}\right], \tag{S12}$$

where $\mathcal{W}_\tau \subseteq \mathcal{V}_\tau$ is the node set used for cross-view density alignment and $\|\cdot\|_F$ is the Frobenius norm. The loss measures disagreement between the two views through normalized density-state overlap. Its nonnegativity and zero-loss condition are established in Proposition S4 in Appendix S5.3.

S5.2 Computational Complexity Analysis

We analyze the computation after pretrained modality features have been extracted. Let N and E be the numbers of entities and training-graph edges, M the maximum number of modalities per entity, f the maximum input feature dimension, and D the complex-state dimension. Let L_v and L_ρ denote the vector- and density-path depths, h the maximum vector-network width, and R an upper bound on the number of density-state components, including initial and residual components. The numbers of relation types, attention heads, and Householder transformations per relation are fixed. When analyzing scaling with data size, we also hold feature dimensions and model configuration fixed.

State encoding and dual-path propagation. Modality projection, amplitude–phase encoding, and reliability-weighted fusion require $O(NM(fD + D^2))$ operations. Node transformations and message aggregation in the vector path cost $O(L_v(Nh^2 + Eh))$. In each density-path layer, edge-wise state matching and aggregation cost $O(EDR^2)$, while nodewise low-rank compression costs $O(NDR^2 + NR^3)$. One complete graph encoding therefore has complexity

$$C_{\text{enc}} = O\Big(NM(fD + D^2) + L_v(Nh^2 + Eh) + L_\rho\big((N + E)DR^2 + NR^3\big)\Big), \tag{S13}$$

where the R^3 term arises from small-matrix factorization in low-rank compression. One graph encoding costs $O(N + E)$ with respect to graph size, and low-rank node density states require $O(NDR)$ storage.

Triadic scoring. Given the encoded node states, the model scores triads through vector interactions and density-state measurements. Let J be the number of Hermitian measurements, p the dimension of each Tucker-core mode, and H_{MLP} the hidden width of the classical MLP. For each triad, vector scoring costs $O(DH_{\text{MLP}})$, and density-state measurement and projection cost $O(JDR+Jp)$. Including the Tucker interaction, the per-triad scoring complexity is

$$C_{\text{score}} = O\big(DH_{\text{MLP}} + JDR + Jp + p^3\big)\,. \tag{S14}$$

The p^3 term comes from Tucker-core contraction and dominates the pairwise $O(p^2)$ and unary $O(p)$ terms. The per-triad scoring cost is independent of graph size.

Training. Training encodes the base and perturbed graphs for each batch. Let B be the number of batches per epoch and Q the total number of triad evaluations for the supervised and ranking losses, counting repeated evaluations.

Auxiliary losses add to these encoding and scoring costs. Let S_c and S_ρ bound the numbers of nodes sampled per batch for contrastive learning and density consistency. These losses cost $O(S_c^2 D)$ and $O(S_\rho DR^2)$, respectively. Fixed-width Siamese projections and prototype alignment are absorbed into the encoding and scoring bounds. Including backpropagation, the per-epoch complexity is

$$C_{\text{epoch}} = O\big(BC_{\text{enc}} + QC_{\text{score}} + B\big(S_c^2 D + S_\rho DR^2\big)\big)\,. \tag{S15}$$

Repeated graph encoding contributes $O(B(N + E))$ per epoch.

Inference. Unlike training, which repeats graph encoding for each batch, inference encodes the graph once and reuses the node states to score candidate triads. For Q_{test} candidates, the total cost is

$$C_{\text{infer}} = O(C_{\text{enc}} + Q_{\text{test}} C_{\text{score}})\,. \tag{S16}$$

This reduces to $O(N + E + Q_{\text{test}})$, scaling linearly with graph size and the number of candidate triads.

S5.3 Theoretical Properties and Proofs

We analyze the density-state operations underlying QDGL's representation, propagation, reliability weighting, and cross-view objective. Let $\mathcal{D}_D$ denote the Hermitian positive semidefinite matrices in $\mathbb{C}^{D\times D}$ with unit trace. We write $\|\cdot\|_1$ for the matrix trace norm or the vector ℓ_1 norm, as appropriate, and $\|\cdot\|_F$ for the Frobenius norm. The statements concern the exactly normalized mathematical operators, with nonnegative mixture weights summing to one.

Proposition S1 (Mixed-state representation). Let $\|\psi_m\|_2 = 1$, $\pi_m \geq 0$, and $\sum_m \pi_m = 1$. Then $\rho = \sum_m \pi_m \psi_m \psi_m^\dagger$ belongs to $\mathcal{D}_D$, and

$$1 - \operatorname{Tr}(\rho^2) = \sum_{i,j} \pi_i \pi_j \big(1 - |\langle \psi_i, \psi_j \rangle|^2\big).$$

The state is pure if and only if all positive-weight components differ only by a global phase.

Proof. Each outer product is Hermitian. For any x, $x^\dagger \rho x = \sum_m \pi_m |\psi_m^\dagger x|^2 \geq 0$, and $\operatorname{Tr}\rho = \sum_m \pi_m \|\psi_m\|_2^2 = 1$. Expanding the square gives

$$\operatorname{Tr}(\rho^2) = \sum_{i,j} \pi_i \pi_j \operatorname{Tr}(\psi_i \psi_i^\dagger \psi_j \psi_j^\dagger) = \sum_{i,j} \pi_i \pi_j |\langle \psi_i, \psi_j \rangle|^2.$$

Subtracting this expression from $\sum_{i,j} \pi_i \pi_j = 1$ proves the identity. Every summand in the resulting expression is nonnegative by Cauchy–Schwarz. It vanishes for every positive-weight pair exactly when $\psi_j = e^{i\theta_{ij}} \psi_i$, in which case all corresponding outer products coincide and ρ has rank one. Conversely, a unit-trace positive semidefinite matrix of rank one has purity one, giving the same equality condition. □

This identity connects mixed-state purity to weighted differences between encoded modalities. Unlike the outer product of a single normalized vector, whose purity is always one, a mixture can represent nonzero disagreement between its components. This is a property of the encoded states, not a calibration statement about label uncertainty.

Proposition S2 (Validity under propagation and compression). Suppose the input density states belong to $\mathcal{D}_D$, each relation transform U_k is unitary, $\eta \in [0, 1]$, and $\gamma_k \geq 0$. Define

$$A = \eta \rho_v + (1 - \eta) \sum_k \gamma_k U_k \rho_k U_k^\dagger.$$

For a basis $Q \in \mathbb{C}^{D \times R_c}$ satisfying $Q^\dagger Q = I_{R_c}$, with $1 \leq R_c \leq D$, and $\varepsilon > 0$, let

$$H = Q^\dagger A Q + \varepsilon I_{R_c} = L L^\dagger, \qquad B = QL, \qquad \mathcal{C}_{R_c,\varepsilon}(A) = \frac{BB^\dagger}{\operatorname{Tr}(BB^\dagger)}.$$

Then $\mathcal{C}_{R_c,\varepsilon}(A) \in \mathcal{D}_D$ and its rank is at most R_c.

Proof. Unitary conjugation preserves positive semidefiniteness and trace, since $x^\dagger U_k \rho_k U_k^\dagger x \geq 0$ and $\operatorname{Tr}(U_k \rho_k U_k^\dagger) = \operatorname{Tr}\rho_k = 1$. Thus A is positive semidefinite under nonnegative aggregation, including when the neighbor set is empty. For every nonzero z, $z^\dagger H z = (Qz)^\dagger A (Qz) + \varepsilon \|z\|_2^2 > 0$, so the Cholesky factor L exists and is invertible. Consequently, $BB^\dagger$ is positive semidefinite and

$$\operatorname{Tr}(BB^\dagger) = \operatorname{Tr} H = \operatorname{Tr}(Q^\dagger A Q) + \varepsilon R_c > 0.$$

Normalization gives unit trace. Since $Q^\dagger Q = I_{R_c}$, Q has rank R_c. Multiplication by the invertible matrix L preserves this rank, so $\operatorname{rank}(B) = \operatorname{rank}(QL) = R_c$. Hence $BB^\dagger$ and its normalized state also have rank R_c, and every column of B is nonzero. Writing $B = [b_1, \ldots, b_{R_c}]$ gives the factorized representation

$$\mathcal{C}_{R_c,\varepsilon}(A) = \sum_{j=1}^{R_c} w_j \widehat{\psi}_j \widehat{\psi}_j^\dagger, \qquad w_j = \frac{\|b_j\|_2^2}{\sum_k \|b_k\|_2^2}, \qquad \widehat{\psi}_j = \frac{b_j}{\|b_j\|_2}.$$

The positive column norms make both the mixture weights and the normalized components well defined. □

The relation transforms used by QDGL satisfy the unitary assumption under exact normalization. A diagonal phase matrix $P = \operatorname{diag}(e^{i\theta_j})$ satisfies $P^\dagger P = I$. For a unit vector v, the Householder factor $F_v = I - 2vv^\dagger$ satisfies

$$F_v^\dagger F_v = (I - 2vv^\dagger)^2 = I - 4vv^\dagger + 4v(v^\dagger v)v^\dagger = I.$$

Products of these factors are unitary. Their adjoints, obtained by reversing the factor order and taking each adjoint, are also unitary and define the reverse-relation transforms.

Applying the proposition layer by layer shows that density propagation preserves valid states without materializing dense density matrices. The initial-state residual $\beta \rho_v^{(L_\rho)} + (1 - \beta)\rho_v^{(0)}$, for $\beta \in [0, 1]$, is also a density state by convexity. The rank inequality for a sum gives its bound $\min(D, R_c + r_0)$, where $r_0 = \operatorname{rank}\rho_v^{(0)}$ and R_c is the maximum rank retained by each compression layer. State validity holds for any orthonormal compression basis, independently of its approximation quality.

Proposition S3 (Local perturbation under reliability weighting). For density states $\rho_m, \rho'_m \in \mathcal{D}_D$ and simplex weights π, π', let $\rho = \sum_m \pi_m \rho_m$ and $\rho' = \sum_m \pi'_m \rho'_m$. Then

$$\|\rho - \rho'\|_1 \le \sum_m \pi_m \|\rho_m - \rho'_m\|_1 + \|\pi - \pi'\|_1.$$

If only modality k changes and the weights remain fixed, this reduces to $\|\rho - \rho'\|_1 \le \pi_k \|\rho_k - \rho'_k\|_1 \le 2\pi_k$.

Proof. Add and subtract $\sum_m \pi_m \rho'_m$ to obtain

$$\rho - \rho' = \sum_m \pi_m (\rho_m - \rho'_m) + \sum_m (\pi_m - \pi'_m) \rho'_m.$$

The triangle inequality and $\|\rho'_m\|_1 = \operatorname{Tr} \rho'_m = 1$ give the first bound. For a single perturbed modality, the weight-change term is zero and $\|\rho_k - \rho'_k\|_1 \le \|\rho_k\|_1 + \|\rho'_k\|_1 = 2$. □

For modality removal, let $\delta < 1$ be the original total weight of the removed components. If retained states are unchanged and their original weights are renormalized, the retained mixture ρ_K satisfies $\|\rho - \rho_K\|_1 \le 2\delta$. For $0 < \delta < 1$, write $\rho = (1 - \delta)\rho_K + \delta \rho_J$, where ρ_J is the normalized removed mixture, and apply $\|\rho_J - \rho_K\|_1 \le 2$. The case $\delta = 0$ is immediate. These bounds explain how reliability weights control local state sensitivity, where δ is removed weight mass for an entity, not the dataset-wide missing rate. The bounds apply before subsequent graph propagation and scoring.

Proposition S4 (Density-view alignment). For $\rho, \sigma \in \mathcal{D}_D$, define the per-node alignment loss

$$\ell_\rho(\rho, \sigma) = 1 - \frac{\operatorname{Tr}(\rho\sigma)}{\|\rho\|_F \|\sigma\|_F}.$$

Then $0 \le \ell_\rho \le 1$, and $\ell_\rho = 0$ if and only if $\rho = \sigma$. Moreover,

$$\left\| \frac{\rho}{\|\rho\|_F} - \frac{\sigma}{\|\sigma\|_F} \right\|_F^2 = 2\ell_\rho(\rho, \sigma).$$

Proof. Positive semidefiniteness gives $\operatorname{Tr}(\rho\sigma) = \operatorname{Tr}(\rho^{1/2}\sigma\rho^{1/2}) \ge 0$. The Hilbert–Schmidt Cauchy–Schwarz inequality bounds this quantity above by $\|\rho\|_F \|\sigma\|_F$, which is positive for unit-trace states. Equality holds exactly when $\rho = a\sigma$ for some $a > 0$. Taking traces gives $a = 1$, proving the zero-loss condition. Expanding the squared Frobenius distance between the normalized matrices proves the final identity. □

Thus density-view alignment directly constrains differences between the density representations of the same node across views. It complements vector-level consistency by supervising the state used by the density scoring path. This pairwise alignment property does not by itself enforce separation between different nodes, which is learned jointly with the prediction and other training objectives.

S6 Supplementary Experiments and Analyses

This section further analyzes QDGL's predictive performance, dual-path complementarity, and the roles of key components. We examine its performance under missing features and graph structure, modality perturbations, false-negative training labels, and disjoint splits. Finally, triadic association case studies combine candidate ranking, local relation analysis, and literature evidence to illustrate the biomedical context of model predictions.

S6.1 Supplementary Analysis of Ranking and Discrimination

We further examine top-ranked retrieval, discrimination between positive and negative triads, and sensitivity to candidate-pool size.

S6.1.1 Top-Ranked Retrieval Performance

Figure S1 presents Hit@K curves under sampled and full candidate ranking. The sampled setting includes one true disease and 50 sampled negative diseases per query. The full candidate setting

scores all diseases and filters out other known positive diseases for the same query. Hit@K measures the proportion of queries whose target disease appears among the top K predictions.

Figure S1 shows that QDGL achieves higher Hit@1, Hit@5, and Hit@10 than the plotted baselines on all three datasets under sampled candidate ranking. As K increases, the methods progressively retrieve true diseases at lower ranks, narrowing the gaps in hit rates.

When the candidate set expands to all diseases, QDGL maintains leading hit rates at the top of the ranking on all three datasets. On CTD, its Hit@1 reaches 81.32%, compared with 34.28% for the strongest plotted baseline, and its Hit@5 reaches 95.71%. Even within the full disease set, QDGL ranks the true disease first for most queries and covers a larger proportion of queries within the top five predictions. The high coverage among top-ranked predictions on ncRNADrug and PrimeKG likewise demonstrates its ability to prioritize true diseases within larger candidate sets.

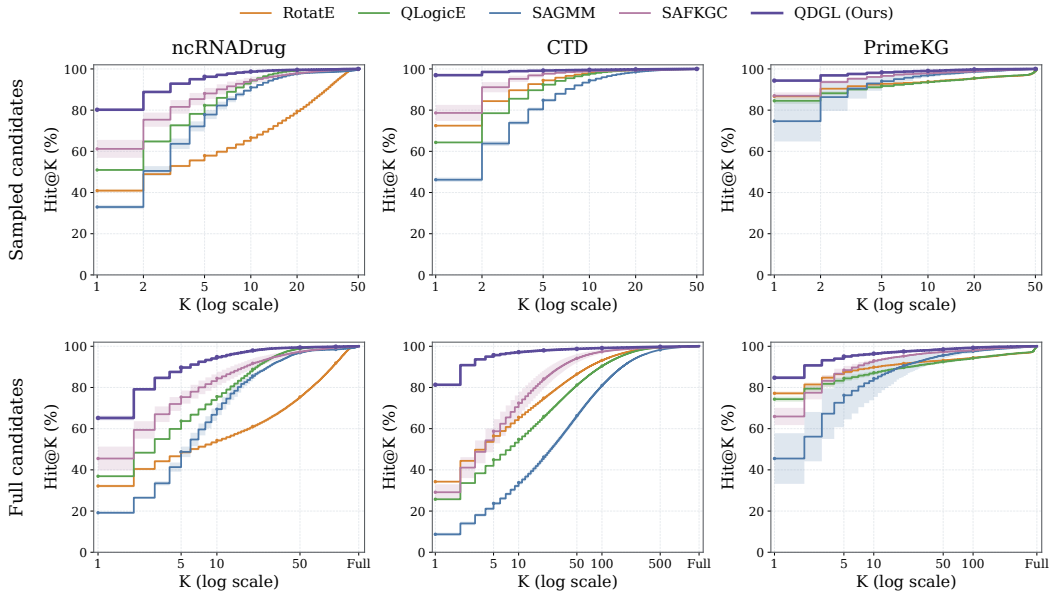


Figure S1: Hit@K under sampled (top) and full (bottom) candidate ranking. Lines and shading denote the five-fold mean and standard deviation.

S6.1.2 Discrimination Between Positive and Negative Associations

Figure S2 presents the ROC curves corresponding to Table 2. QDGL achieves the highest mean AUC among the plotted methods on all three datasets. On CTD and PrimeKG, its ROC curves reach high true-positive rates at low false-positive rates. On ncRNADrug, its advantage extends over a broad range of false-positive rates. These results show that QDGL combines effective prioritization of true diseases with discrimination between positive and negative triads across decision thresholds.

S6.1.3 Candidate-Pool Size Sensitivity

Candidate-pool size affects link prediction evaluation (Li et al., 2023a). Figure S3 compares ranking performance as the candidate pool expands to all diseases, with results averaged equally across the three datasets. MRR decreases for all methods as the pool grows, but QDGL exhibits a smaller decline, widening its margin over the strongest baseline shown. Its MRR advantage increases from 4.07 percentage points under the smallest candidate setting to 25.21 percentage points under full-candidate ranking, where its Hit@1 advantage reaches 29.22 percentage points. These results indicate that QDGL retains its ability to prioritize true diseases as more candidates compete with the target.

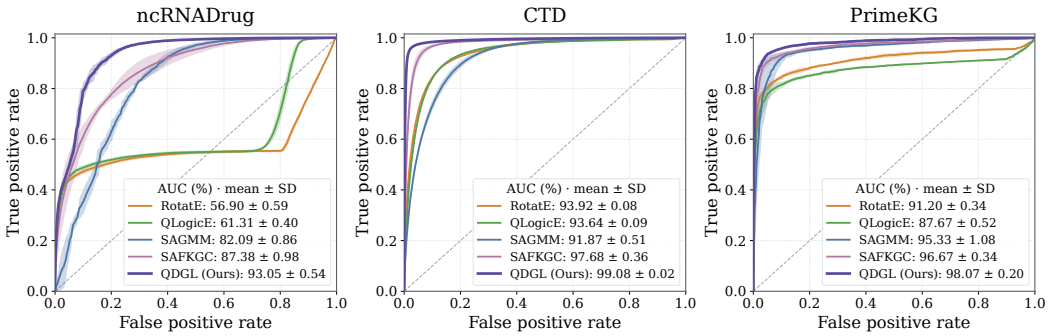


Figure S2: ROC curves for triadic association classification (five-fold mean ± standard deviation).

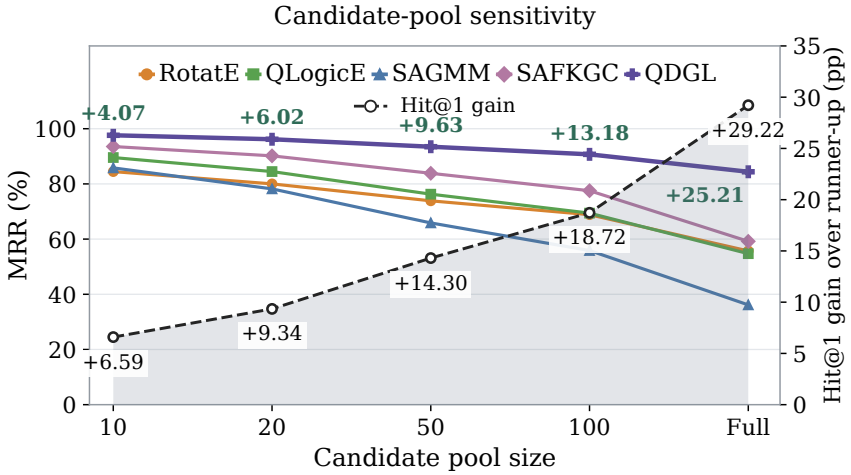


Figure S3: Ranking performance across candidate-pool sizes, averaged equally over the three datasets. Full denotes all disease candidates. Solid curves show MRR. The dashed curve and green labels show QDGL's Hit@1 and MRR gains over the strongest baseline shown, respectively, in percentage points.

S6.2 Component Ablations and Dual-Path Analysis

This section further examines the roles of QDGL's components and training objectives. We first define the ablation variants in Table 3, report supplementary controls for state representations and triadic scoring, and compare predictive performance under different loss configurations. We then analyze how the two paths complement each other in joint prediction through their score contributions and changes in candidate ranks, followed by a visualization of triad representations before and after training.

S6.2.1 Ablation Variant Settings

The component ablation variants in Table 3 are defined as follows:

- **w/o phase**: Sets the phase to zero while retaining learnable probability amplitudes.
- **w/o mixed-state fusion**: Replaces each entity's mixed-state representation with a single graph-propagated pure state for scoring.
- **w/o reliability weighting**: Assigns uniform fusion weights to each entity's available modalities.
- **w/o state normalization**: Disables state normalization in modality fusion and graph propagation.

- **w/o dual path**: Retains graph propagation along the vector path and the MLP triadic scorer, while removing density-path propagation and scoring.
- **fixed random phase**: Randomly initializes the phase and keeps it fixed during training, while retaining learnable probability amplitudes.

Table 3 also includes two real-valued spherical controls:

- **vMF density block**: Replaces quantum mixed-state fusion with von Mises–Fisher (vMF) distribution fusion and uses an MLP triadic scorer.
- **spherical cosine**: Uses zero-phase, normalized real-valued states and scores each triad by the mean cosine similarity over its three entity pairs.

S6.2.2 State-Representation and Triadic-Interaction Controls

We compare QDGL with simplified alternatives along two dimensions: multimodal state representation and triadic association scoring. Figure S4 includes four controls:

- **Normalized complex vector**: Uses normalized complex vectors in place of density states.
- **Matched real vector**: Uses a parameter-matched real-vector alternative to density states.
- **w/o density triadic term**: Removes the density path's explicit three-way interaction while retaining joint triadic scoring in the vector path.
- **Pairwise additive scoring**: Retains density states but restricts both paths to additive pairwise scoring.

For state representation, QDGL achieves higher mean classification and ranking scores than both vector alternatives across all three datasets. On PrimeKG, the normalized complex-vector and parameter-matched real-vector controls achieve Hit@1 scores of 92.70% and 92.61%, respectively, compared with 94.34% for QDGL, which also attains higher MRR. The two vector alternatives perform similarly, while the full model further improves the ranking of true diseases. QDGL represents modalities as weighted state components, allowing subsequent propagation and measurement readout to use differences in modality composition.

For triadic scoring, full QDGL also achieves higher mean classification and ranking scores than both the control without the density triadic term and the additive pairwise control. On CTD, additive pairwise scoring reaches an AUC of 98.84%, close to QDGL's 99.08%, but its Hit@1 is 93.39%, compared with 96.98% for QDGL. The gap is particularly apparent in placing the true disease first among competing candidates. Additive pairwise scoring accumulates evidence from individual entity pairs, whereas the triadic term explicitly models the joint compatibility of a molecule, a drug, and a disease.

S6.2.3 Loss Ablation

We examine the effects of loss terms and their combinations on prediction performance across the three datasets through individual removal and progressive combination:

- **Individual removal**: Remove the ranking, NCE contrastive, Siamese consistency, prototype, or density-view consistency loss separately, corresponding to the w/o variants in the table.
- **Progressive combination**: Train three configurations separately: supervised only, supervised + ranking, and supervised + ranking + NCE.

Table S5 reports five-fold means and standard deviations for Hit@1, MRR, and AUC.

In Table S5, full QDGL achieves the highest mean Hit@1, MRR, and AUC on all three datasets. The ranking loss directly constrains the relative scores of the true disease and competing candidates within each query, supplementing positive–negative triad discrimination with supervision on candidate ordering. On CTD, removing the ranking loss leaves AUC high but reduces Hit@1 by 1.76 percentage points, indicating that strong overall discrimination does not guarantee that the true disease ranks first.

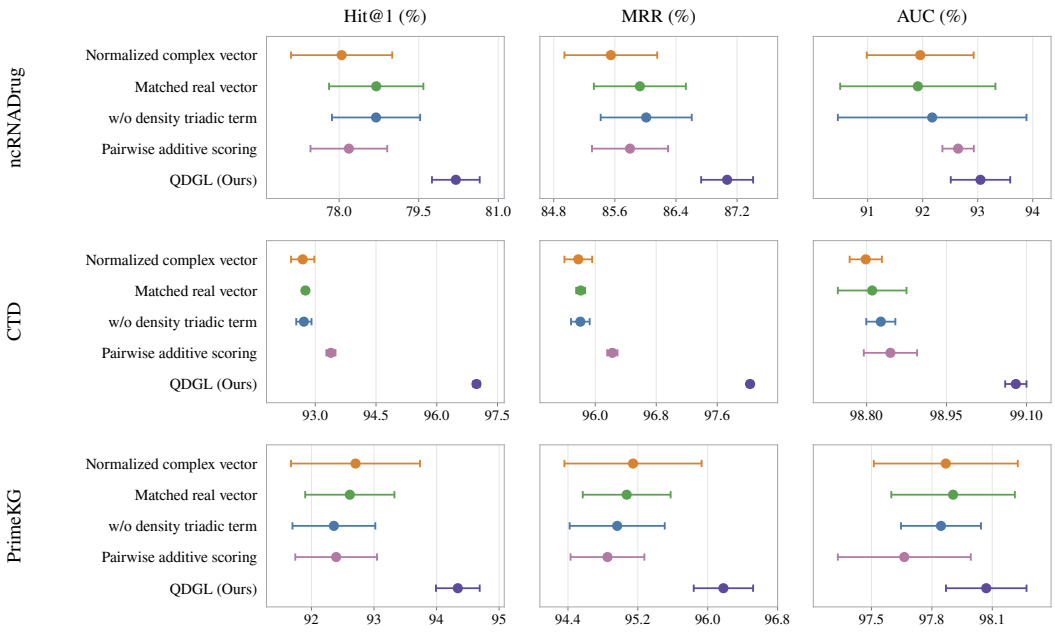


Figure S4: State-representation and triadic-interaction controls. Points and error bars denote means and standard deviations.

Representation constraints supplement association supervision with cross-view and cross-modal training signals. NCE aligns the same node across graph views while distinguishing different nodes, encouraging discriminative entity representations under neighborhood perturbations. Removing NCE on ncRNADrug reduces AUC by 1.64 percentage points and also degrades ranking performance, demonstrating the contribution of node representation learning to both discrimination and ranking. Siamese consistency further coordinates modality branches, while prototype loss aligns class-level representations of positive and negative triads across views, extending the constraints from individual nodes to modality branches and triad classes. Removing either objective degrades performance, indicating that these constraints complement node-level contrastive learning. On PrimeKG, the combination of supervised, ranking, and NCE losses remains 3.15 percentage points below the full model in Hit@1. Together with the individual-removal results, this gap indicates that constraints on modality branches, class representations, and density states also contribute to the final ranking performance beyond learning discriminative node representations.

Density states directly contribute to candidate scoring in the quantum-inspired path. We therefore introduce density-view consistency to align the propagated density states of the same entity across the original and perturbed graphs, aiming to reduce the influence of neighborhood perturbations on this path's representations. On CTD, removing this loss reduces Hit@1 by 1.77 percentage points while leaving AUC nearly unchanged. Thus, even with similar overall positive–negative discrimination, this constraint affects whether the true disease ranks first.

S6.2.4 Dual-Path Contributions and Complementarity

To examine how the two paths jointly improve candidate ranking, we compare the vector, quantum-inspired, and combined readouts of the same QDGL model. A path margin is the true-disease score minus the competing-candidate score, with a positive value indicating a preference for the true disease. The quantum-inspired margin includes its learned mixing weight, and the two path margins sum to the combined margin.

Figure S5(a) plots the two path margins against the highest-scoring negative under the combined readout. When both margins are positive, both paths place the true disease above that candidate. Opposite signs indicate different judgments about the same candidate. Among these queries, a positive combined margin illustrates complementary correction: one path favors the negative candidate,

Table S5: Loss ablations of QDGL. **Bold** and underline mark the best and second-best results.

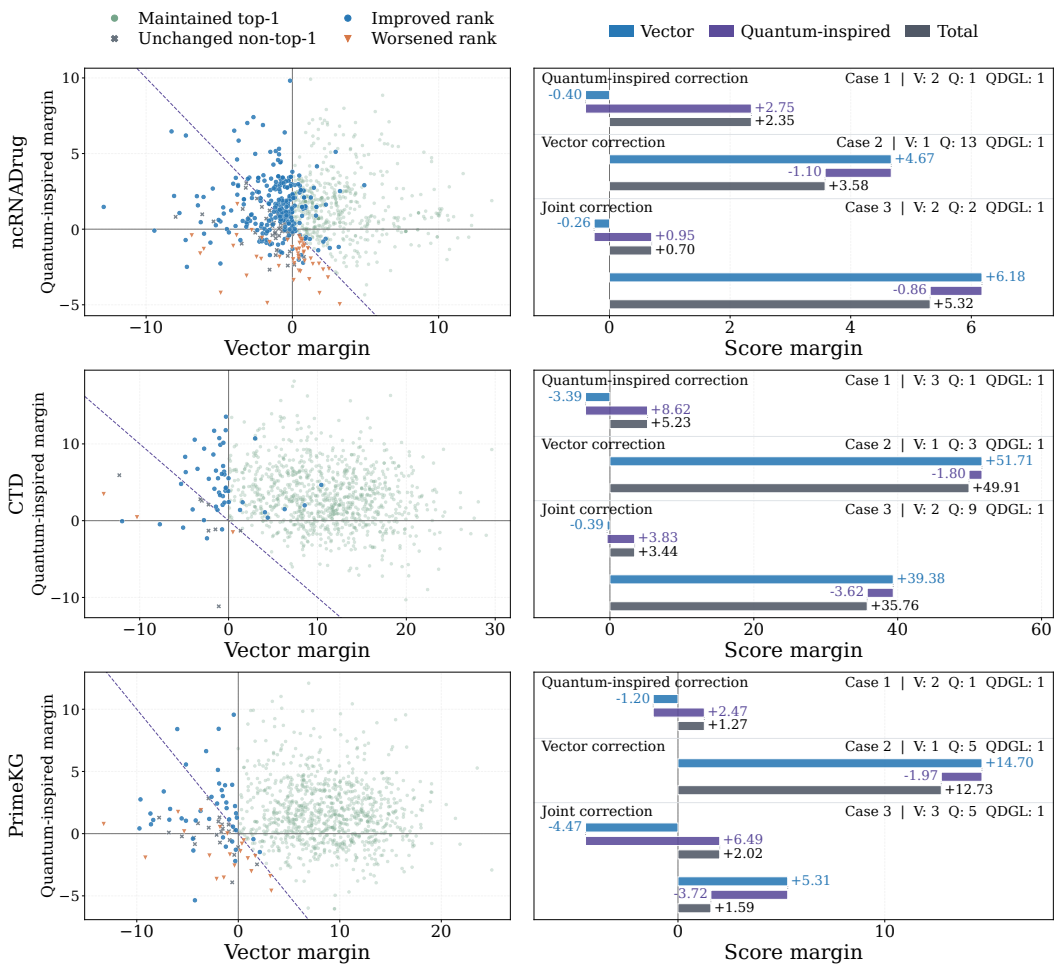
ncRNADrug	Hit@1	MRR	AUC
w/o ranking loss	78.13 ± 0.72	85.37 ± 0.37	92.31 ± 0.31
w/o NCE loss	77.71 ± 1.09	84.87 ± 0.70	91.41 ± 0.75
w/o Siamese consistency	78.25 ± 1.00	85.40 ± 0.58	92.03 ± 0.43
w/o prototype loss	77.77 ± 1.05	85.04 ± 0.64	92.10 ± 0.50
w/o density-view loss	$\underline{80.03 \pm 0.29}$	$\underline{86.92 \pm 0.30}$	$\underline{92.74 \pm 0.43}$
supervised only	77.97 ± 0.58	85.13 ± 0.34	91.74 ± 0.56
supervised + ranking	77.67 ± 1.49	84.93 ± 0.87	91.19 ± 0.77
supervised + ranking + NCE	77.78 ± 0.97	85.14 ± 0.64	92.31 ± 0.75
QDGL	$\mathbf{80.20 \pm 0.45}$	$\mathbf{87.07 \pm 0.34}$	$\mathbf{93.05 \pm 0.54}$
CTD	Hit@1	MRR	AUC
w/o ranking loss	95.22 ± 0.13	97.09 ± 0.08	98.94 ± 0.03
w/o NCE loss	95.47 ± 0.14	97.20 ± 0.10	99.01 ± 0.03
w/o Siamese consistency	95.26 ± 0.11	97.11 ± 0.07	98.97 ± 0.02
w/o prototype loss	95.58 ± 0.10	97.28 ± 0.07	98.98 ± 0.02
w/o density-view loss	95.21 ± 0.15	97.18 ± 0.08	$\underline{99.05 \pm 0.04}$
supervised only	95.61 ± 0.11	97.34 ± 0.07	99.01 ± 0.02
supervised + ranking	$\underline{95.69 \pm 0.05}$	$\underline{97.37 \pm 0.03}$	98.99 ± 0.03
supervised + ranking + NCE	95.48 ± 0.09	97.23 ± 0.05	98.98 ± 0.01
QDGL	$\mathbf{96.98 \pm 0.08}$	$\mathbf{98.03 \pm 0.04}$	$\mathbf{99.08 \pm 0.02}$
PrimeKG	Hit@1	MRR	AUC
w/o ranking loss	91.23 ± 0.41	93.88 ± 0.40	97.44 ± 0.21
w/o NCE loss	90.85 ± 0.72	93.67 ± 0.54	97.46 ± 0.36
w/o Siamese consistency	90.98 ± 0.58	93.69 ± 0.51	97.21 ± 0.27
w/o prototype loss	91.54 ± 0.31	94.10 ± 0.37	97.50 ± 0.26
w/o density-view loss	$\underline{93.94 \pm 0.53}$	$\underline{95.92 \pm 0.44}$	$\underline{98.03 \pm 0.18}$
supervised only	90.34 ± 0.62	93.21 ± 0.47	97.19 ± 0.47
supervised + ranking	90.72 ± 0.60	93.54 ± 0.56	97.20 ± 0.43
supervised + ranking + NCE	91.19 ± 0.56	93.75 ± 0.48	97.32 ± 0.30
QDGL	$\mathbf{94.34 \pm 0.35}$	$\mathbf{96.18 \pm 0.34}$	$\mathbf{98.07 \pm 0.20}$

while the positive margin from the other enables the combined scorer to place the true disease ahead. The paths can thus reinforce a correct ordering or contribute different judgments about competing candidates. Blue points identify rank improvements under the combined readout relative to the vector readout, connecting path margins with changes in the final candidate ordering.

Figure S5(b) illustrates how the paths handle competing candidates through quantum-inspired correction, vector correction, and joint correction. Cases are selected near the median path margins within each ranking pattern. The first two use the highest-scoring negative under the vector and quantum-inspired readouts, respectively, to show how the other path corrects the ordering. Joint correction compares the highest-scoring negatives under the two individual readouts in the upper and lower bar groups, respectively. For PrimeKG Case 3, the true disease ranks third under the vector readout and fifth under the quantum-inspired readout, but first under the combined readout. Neither individual readout places the true disease first, yet they favor different incorrect candidates, whose score advantages over the true disease are reduced by the other path. The joint-correction cases shown for ncRNADrug and CTD likewise recover the correct top-ranked prediction. These cases demonstrate complementary discrimination among competing candidates, enabling the joint model to achieve a correct top-ranked prediction that neither individual readout provides.

S6.2.5 Representation Distributions Before and After Training

To examine how training affects triad representations, we compare the distributions of the same positive and negative triads before and after training. We combine vector-path hidden features with density-path interaction features and project them into two dimensions using a shared PCA and t-SNE mapping, with consistent coordinate ranges to show changes in the class distributions.



(a) Joint distribution of path margins

(b) Examples of path complementarity

Figure S5: Dual-path margin distributions and complementary cases. Scatter colors indicate rank changes relative to the vector readout. The dashed line marks zero combined margin. V, Q, and QDGL denote ranks under the vector, quantum-inspired, and combined readouts.

Figure S6 shows that positive and negative triads are intermingled across multiple regions before training on all three datasets, with local clusters exhibiting no clear class separation. After training, CTD exhibits two more distinct class regions, accompanied by separated peaks in the marginal distributions, indicating reduced class overlap in the projected space. ncRNADrug and PrimeKG retain multiple local clusters while showing clearer differences between the class distributions. For example, ncRNADrug retains substantial overlap along the horizontal projection but exhibits clearer separation along the vertical projection, revealing a more distinct class structure in the joint two-dimensional distribution.

S6.3 Dual-Path Robustness under Missing Information

To examine the role of joint dual-path modeling under incomplete information, we reduce both entity features and graph structure and compare the resulting performance of full QDGL with its two single-path variants. Specifically, we randomly mask entity-level modality features and remove edges from the training message-passing graph at the same missing rate, evaluating at 10%, 30%, 50%, and 80%. QDGL (vector-only) and QDGL (quantum-only) use the vector path and the quantum-inspired density path, respectively. The two single-path variants provide references for

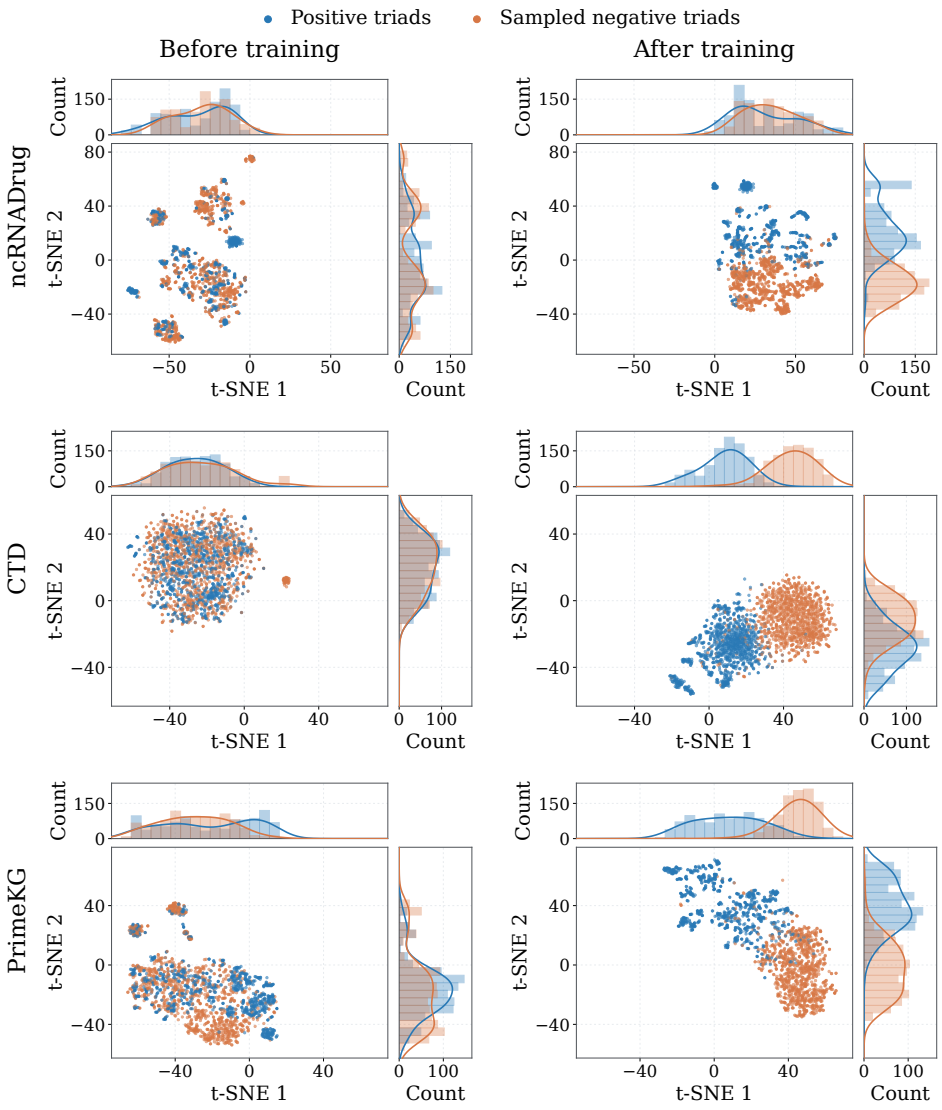


Figure S6: QDGL triad representations before and after training. Blue and orange denote positive and sampled negative triads. Marginal histograms and density curves show distributions along the two projection axes.

prediction using each representation alone, while increasing missing rates probes how well their joint use retains performance as available information decreases.

Figure S7 shows that full QDGL achieves the highest mean performance across all three datasets and missing rates. On ncRNADrug, both single-path variants exhibit declining ranking performance as missingness increases, while the full model remains relatively stable at low to moderate missing rates. At 80% missingness, Hit@1 falls to approximately 64% for both single-path variants, whereas QDGL retains 77.89%. The joint model also maintains higher MRR, extending its gains to both top-ranked disease identification and candidate ranking.

The widening gap between the full model and the stronger single-path variant at higher missing rates further illustrates the value of joint dual-path modeling. On PrimeKG, QDGL's Hit@1 advantage over the vector-only variant increases from 3.91 percentage points at 10% missingness to 23.86 percentage points at 80%. Although severe information loss affects all three models, the joint model exhibits a smaller performance decline, with more pronounced gains as available evidence decreases. On CTD, the density-only variant performs below the vector-only variant, yet full QDGL

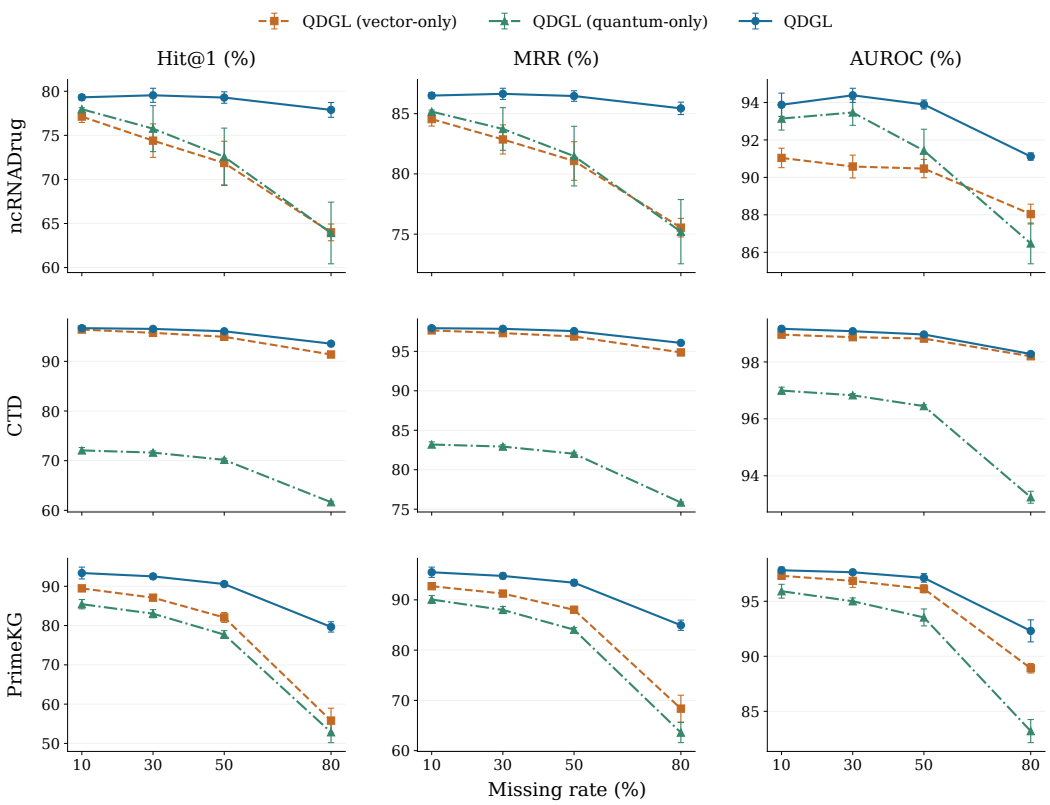


Figure S7: Performance of QDGL and its single-path variants under missing information. Error bars indicate standard deviations.

achieves higher mean performance at every missing rate when both paths are used jointly. At 80% missingness, its Hit@1 exceeds the vector-only variant by 2.18 percentage points. Despite the density path's weaker standalone predictions, the joint model still improves top-ranked identification of the true disease.

The vector path learns interaction representations from entity features and relational context, while the density path preserves modality composition and relative weights and contributes to candidate scoring through state matching. The two paths use the remaining information in different ways. When one path has difficulty distinguishing the true disease from a competing candidate, scores from the other path can provide additional discriminative evidence and improve the joint ranking.

S6.4 Modality Reliability and Perturbation Responses

We compare modality-weight responses across drugs, examine predictive performance under ECFP perturbation, and analyze how different inputs contribute to candidate discrimination through query-level modality masking.

S6.4.1 Case Responses under Different Initial Modality Reliance

To compare modality allocation across drugs under structural perturbation, we randomly permute drug ECFP features on ncRNADrug while keeping text and other inputs unchanged, and compare QDGL's text and structural weights before and after perturbation. The analysis covers 165 drugs with both modalities, with perturbed weights averaged over five random permutations. An ECFP weight of at least 0.5 defines ECFP dominance, with the remaining cases classified as text-dominant. The dominant modalities before and after perturbation define four response types. Table S6 summarizes their proportions. Table S7 presents representative drugs whose clean and perturbed weights are close to the respective medians of each response type.

Table S6: Modality-response distribution before and after ECFP perturbation, pooled across five folds.

Clean dominance	Shuffled dominance	Proportion (%)
ECFP	Text	12.61
ECFP	ECFP	8.12
Text	Text	75.52
Text	ECFP	3.76

Table S7: Representative drugs from each response type. Arrows denote clean-to-shuffled weights, and Δw_{text} is the text-weight change.

Drug	Dominance	Text weight	ECFP weight	Δw_{text}
Everolimus	ECFP → Text	$0.2664 \rightarrow 0.5920$	$0.7336 \rightarrow 0.4080$	$+0.3257$
Lapatinib	ECFP → ECFP	$0.0726 \rightarrow 0.3513$	$0.9274 \rightarrow 0.6487$	$+0.2787$
Nutlin-3	Text → Text	$0.9593 \rightarrow 0.8939$	$0.0407 \rightarrow 0.1061$	-0.0654
Regorafenib	Text → ECFP	$0.6932 \rightarrow 0.4685$	$0.3068 \rightarrow 0.5315$	-0.2247

After ECFP permutation, drugs initially dominated by structural inputs show an overall tendency to shift weight toward text. Across the five folds, approximately 61% of initially ECFP-dominant results become text-dominant. Everolimus increases its text weight from 0.2664 to 0.5920, making text its primary information source. Lapatinib also increases its text weight while reducing its structural weight. These two cases show that, when structural inputs change, QDGL can reduce their share in modality fusion and increase the relative weight of text, which is not directly perturbed.

Nutlin-3, which initially relies primarily on text, retains text dominance. Although its text weight decreases, it remains approximately 89% after perturbation, with structural information receiving a much smaller weight. Whereas Everolimus shifts to text after perturbation, Nutlin-3 uses text as its primary information source both before and after perturbation. The change in structural inputs does not alter which modality receives greater weight.

A small proportion of cases exhibit adjustment in the opposite direction. Regorafenib increases its ECFP weight after structural perturbation, changing from text dominance to ECFP dominance. This response type accounts for 3.76% of the pooled results across five folds. This case illustrates that weight changes induced by structural permutation differ across drugs, including both shifts toward text that is not directly perturbed and increased weighting of the permuted structural input.

S6.4.2 Prediction Performance under Structural Feature Perturbation

To examine the effect of incorrect structural information on prediction, we randomly permute ECFP features among ncRNADrug drugs with observed fingerprints. This preserves feature availability and the overall fingerprint distribution while disrupting the correspondence between drugs and their structures. Only ECFP features are perturbed, with the trained model and remaining evaluation conditions fixed. Ranking uses the same 50 negative diseases per test triad, and five-fold performance is summarized after averaging five permutations within each fold.

Table S8: Reference and ECFP-perturbed performance on ncRNADrug (mean ± SD, %). Differences are reference minus perturbed performance, in percentage points.

	Hit@1			MRR			AUC		
Method	Reference	Perturbed	Difference	Reference	Perturbed	Difference	Reference	Perturbed	Difference
QDGL	80.20 ± 0.45	75.53 ± 2.56	4.67	87.07 ± 0.34	84.19 ± 1.72	2.88	93.05 ± 0.54	90.12 ± 1.51	2.93
vMF density	78.73 ± 0.54	26.07 ± 4.48	52.66	85.98 ± 0.42	41.15 ± 3.96	44.83	92.31 ± 0.52	67.57 ± 1.11	24.74
Spherical cosine	74.16 ± 2.15	30.47 ± 4.12	43.69	82.15 ± 1.50	41.09 ± 3.41	41.06	79.18 ± 0.97	56.18 ± 1.72	23.00
MM-RotatE	72.70 ± 1.57	38.41 ± 4.45	34.29	81.71 ± 0.98	51.42 ± 4.22	30.29	77.54 ± 1.84	64.03 ± 2.85	13.51
MM-QLogicE	75.56 ± 0.53	32.77 ± 3.80	42.79	83.74 ± 0.49	51.80 ± 3.71	31.94	86.65 ± 1.11	66.32 ± 1.57	20.33
SAGMM	32.97 ± 1.01	32.97 ± 1.05	0.00	51.75 ± 1.32	51.73 ± 1.34	0.02	82.09 ± 0.86	82.03 ± 0.83	0.06
SAFKGC	61.19 ± 4.36	50.65 ± 4.58	10.54	73.04 ± 3.39	64.60 ± 4.01	8.44	87.38 ± 0.98	80.92 ± 4.22	6.46

Among the reported perturbed results, QDGL achieves the highest mean performance on all three metrics, with a Hit@1 of 75.53%. Thus, even after disrupting the drug–structure correspondence, the model ranks the true disease first for approximately three quarters of the test triads. In the model design, structural information and text jointly contribute to entity representations. Reducing the structural weight can decrease the share of incorrect structural inputs in fusion, while attributes and graph relations that are not directly perturbed provide other information for prediction. The shifts toward text observed for some drugs in the preceding subsection offer a clue to this explanation. The present results describe the prediction performance of the complete model under structural perturbation.

The two real-valued spherical controls further illustrate performance differences under structural perturbation. The vMF density control uses distribution-based fusion and a learnable triadic scorer. Its reference Hit@1 is less than two percentage points below QDGL, but its reported perturbed performance is substantially lower. Spherical cosine, which uses normalized real-valued states and pairwise cosine similarities, also exhibits low ranking performance under perturbation. In particular, the close reference performance of vMF density and QDGL does not carry over to the perturbed results, showing that similar prediction performance with intact inputs can accompany different performance under structural perturbation.

The SOTA methods also differ in their perturbed performance. MM-RotatE and MM-QLogicE both achieve lower perturbed ranking metrics than SAFKGC, indicating differences in their ability to distinguish candidate diseases under incorrect structural inputs. SAGMM has nearly identical results before and after perturbation, with a Hit@1 of 32.97% in both conditions, illustrating that little change can coexist with low prediction performance. In comparison, QDGL maintains a high top-ranked hit rate under structural perturbation and also achieves the highest reported mean MRR and AUC, with advantages in both candidate ranking and positive–negative discrimination.

S6.4.3 Role of Modality Information in Candidate Discrimination

To examine each modality's contribution to candidate discrimination, we mask modality inputs in turn across the relevant entities and compare the score margin between the true disease and the highest-scoring negative before masking. The model, graph, and candidate set remain fixed. Figure S8 shows the unmasked margin minus the masked margin. Positive values indicate that removing the modality reduces the true disease's score advantage over this fixed competitor.

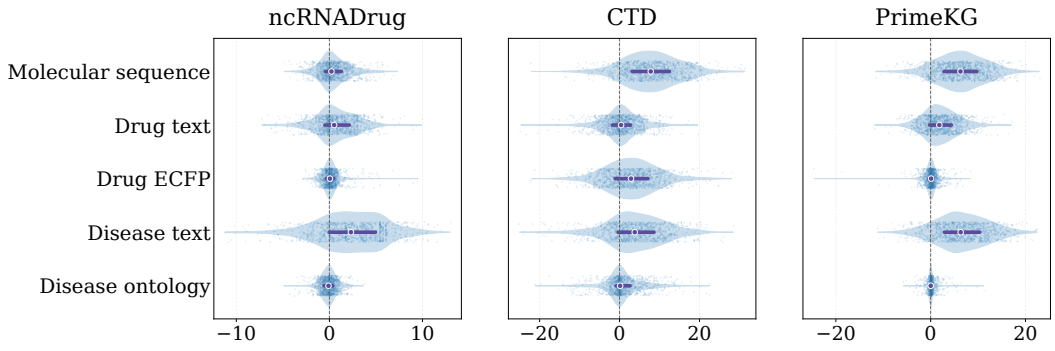


Figure S8: Query-level margin changes under modality masking. Positive values indicate a margin decrease after masking. Purple markers show medians and interquartile ranges.

On ncRNADrug, candidate discrimination is most sensitive to masking disease text. The median margin decrease is 2.31, the largest among the modalities on this dataset. This indicates that disease descriptions contribute to assessing the compatibility between a molecular–drug pair and a candidate disease. Removing this information reduces the true disease's relative score advantage over a high-scoring negative. In comparison, margin changes under drug ECFP masking are more concentrated near zero, indicating that the trained model relies less on this structural input when other information remains available.

On CTD, molecular sequence has the most pronounced role in candidate discrimination. Masking sequence produces a median margin decrease of 7.81, and even the lower quartile is positive,

indicating that approximately three quarters or more of the queries exhibit a margin decrease. The effect therefore extends across most queries and is not driven by a small number of extreme changes. Disease text and drug ECFP also yield positive median margin decreases, indicating that the model additionally uses disease semantics and drug structure to distinguish candidates. However, their distributions span zero, showing that the direction and magnitude of their effects vary across queries.

On PrimeKG, candidate discrimination relies on both molecular sequence and disease text. Masking either modality produces a similar median margin decrease, with both interquartile ranges entirely positive, indicating effects on most queries. Even with the remaining modalities and graph relations available, removing either molecular sequence or disease text weakens discrimination between the true disease and its competing candidate. Across the three datasets, masking reveals distinct modality dependencies: disease text has the most prominent role on ncRNADrug, molecular sequence has the strongest effect on CTD, and PrimeKG uses both molecular sequence and disease text to distinguish true diseases from high-scoring competitors.

S6.5 Assessing Sensitivity to False Negatives during Training

Incomplete association annotations can cause genuine triads to be treated as negatives during training. Incorrect negative labels drive the model to lower the scores of these genuine associations, conflicting with the learning objective of association prediction. To examine the effect of this erroneous supervision, we withhold a fixed 10% of the original training positives from positive supervision and graph construction, then replace 0%, 5%, or 10% of the training negatives with triads sampled from this hidden pool and labeled negative. The contamination rate is the fraction of negative training examples replaced. All rates use the same retained training set and hidden pool, with positive and negative sample counts unchanged. The 0% condition also uses the reduced training set to isolate the effect of incorrect negative labels.

The intervention targets the binary cross-entropy negative pool. Auxiliary ranking samplers retain their original rules and exclude hidden training positives. Validation and test labels and candidate sets remain fixed, with known positives filtered from negative candidates, so that performance changes across contamination levels reflect the effect of erroneous supervision during training.

As shown in Figure S9, QDGL achieves the highest mean Hit@1, MRR, and AUC at every tested contamination level on all three datasets. On PrimeKG, QDGL combines a stronger uncontaminated reference with a smaller decline in ranking performance as contamination increases. From 0% to 10% contamination, its MRR decreases by 4.17 percentage points, compared with 6.60 points for SAFKGC, widening the ranking gap between the two methods. Thus, QDGL better preserves the positions of true diseases relative to negative candidates when some genuine associations receive incorrect negative supervision. Its advantage under contamination reflects both a stronger starting point and a smaller performance loss.

On CTD, QDGL still places nearly 99% of true diseases in the top five under 10% contamination, retaining high candidate coverage. In comparison, its Hit@1 decreases by 10.70 percentage points from the uncontaminated reference, indicating a more pronounced effect on candidate prioritization. Meanwhile, AUC decreases by only 1.14 points, suggesting relatively stable overall positive–negative discrimination. Together, these results indicate that QDGL retains strong candidate-screening ability under contamination, while fine-grained discrimination among high-ranking candidates is more sensitive to erroneous supervision.

On ncRNADrug, erroneous negative supervision also affects positive–negative discrimination, with QDGL's AUC decreasing from 92.46% to 85.56%. This decrease indicates that a genuine association is less likely to receive a higher score than a randomly selected negative, extending the effect of contamination beyond the choice of the first-ranked candidate. Nevertheless, QDGL maintains higher AUC and ranking metrics than the compared methods at every tested contamination level, indicating that it remains more effective at distinguishing genuine associations and prioritizing them among candidates when supervision contains errors.

S6.6 Analyzing Entity-Pair Overlap and Generalization

Even when a test triad is absent from training, its constituent entity pairs may occur in other training triads, allowing the model to draw on these previously observed combinations for prediction. To

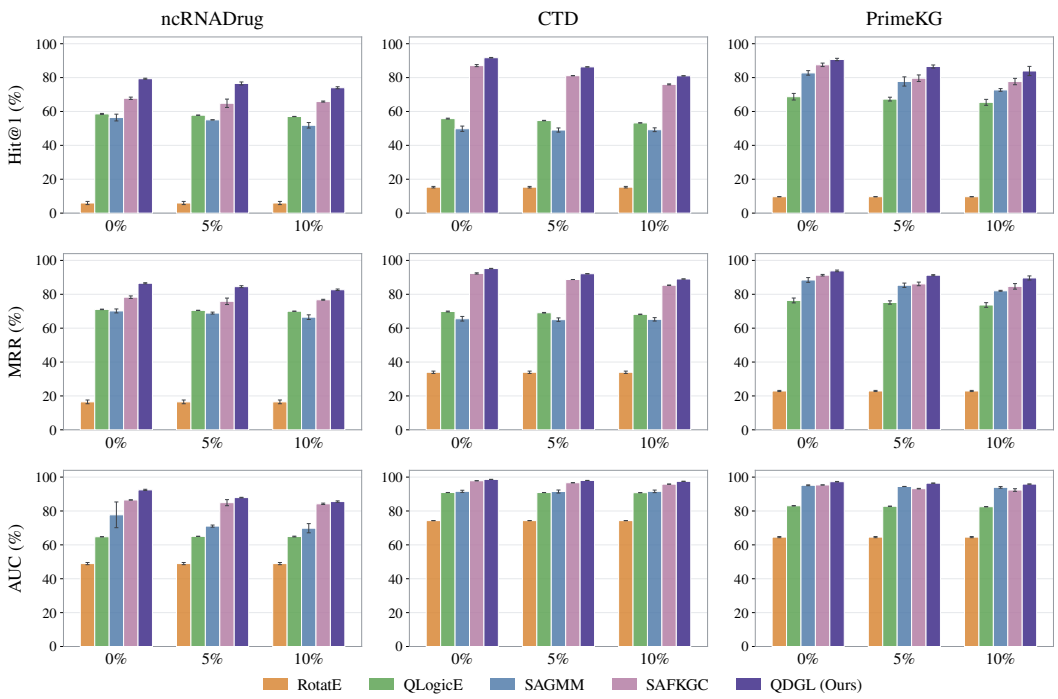


Figure S9: Effects of false-negative training labels on ranking and classification performance. The horizontal axis indicates the proportion of true triads mislabeled as negative among the training negatives. Bars and error bars denote means and standard deviations.

examine QDGL's generalization under different conditions of information reuse, we first analyze entity-pair overlap and the corresponding ranking performance in the original split. We then construct entity-pair-disjoint and unseen-entity splits to evaluate predictions for new entity combinations and triadic associations involving unseen entities.

S6.6.1 Entity-Pair Overlap and Stratified Ranking

To examine QDGL's ranking performance across different levels of entity-pair overlap, we stratify test triads by the occurrence of their constituent pairs in the positive training triads of the same fold. For a triad (n, d, s), we check the molecular–drug pair (n, d), drug–disease pair (d, s), and molecular–disease pair (n, s). PairSeen$k$ denotes the subset with exactly k seen pairs. An unseen query is a molecular–drug pair absent from training positives. An unseen pair does not require its individual entities to be unseen.

Table S9: Entity-pair overlap (five-fold mean $\pm$ sample SD, %) and pooled test-subset counts (proportions in parentheses).

Statistic	ncRNADrug	CTD	PrimeKG
Test triads	24,211	397,598	5,496
Molecular–drug overlap (%)	16.93 ± 0.28	75.66 ± 0.23	76.75 ± 0.66
Drug–disease overlap (%)	98.40 ± 0.11	95.30 ± 0.12	71.42 ± 1.05
Molecular–disease overlap (%)	28.52 ± 0.52	98.87 ± 0.03	79.82 ± 0.78
PairSeen0	121 (0.50%)	176 (0.04%)	82 (1.49%)
PairSeen1	15,306 (63.22%)	5,960 (1.50%)	780 (14.19%)
PairSeen2	6,832 (28.22%)	107,513 (27.04%)	2,152 (39.16%)
PairSeen3	1,952 (8.06%)	283,949 (71.42%)	2,482 (45.16%)

Table S9 shows distinct entity-pair overlap patterns across the three datasets. On ncRNADrug, drug–disease overlap reaches 98.40%, while molecular–drug overlap is much lower, so many test triads involve new molecular–drug combinations within known drug–disease associations. CTD and PrimeKG are dominated by PairSeen2 and PairSeen3, with more test triads composed of pairwise associations observed during training. These differences cover varying degrees of relational reuse and combinatorial novelty. Stratified evaluation further examines how QDGL's ranking gains are distributed across these prediction settings, particularly whether they persist for samples with limited entity-pair overlap and unseen query combinations.

We compare methods on identical test triads and disease candidates within each stratum. Each query contains one true disease and 50 distinct negative diseases filtered against the global positive set. Figure S10 reports the five-fold mean of paired Hit@1 and MRR differences between QDGL and each baseline for all test triads, PairSeen0, the subset with at most one seen pair, and unseen queries.

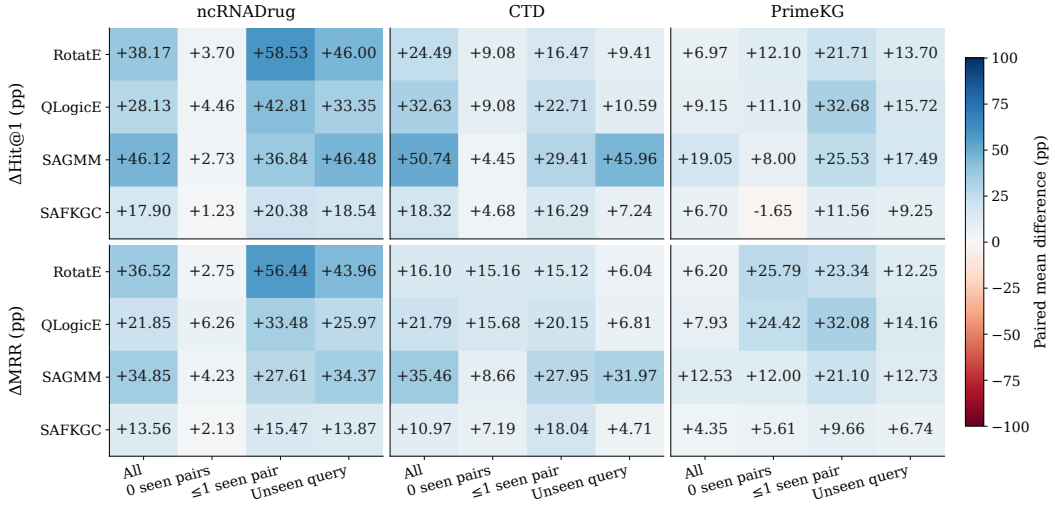


Figure S10: Ranking gains of QDGL over each baseline across entity-pair overlap strata (five-fold mean paired differences, percentage points). Blue/red favors QDGL/the baseline.

In the subset with at most one seen pair, QDGL achieves higher Hit@1 and MRR than every displayed baseline on all three datasets. Each test triad in this subset contains at least two entity pairs absent from training positives, allowing evaluation of predictions for low-overlap combinations. Improvements in both metrics indicate that QDGL places more true diseases first and achieves better overall ranking quality.

In the unseen-query subset, QDGL also outperforms every baseline on Hit@1 and MRR. This stratum specifically examines molecular–drug combinations absent from training positives, complementing the low-overlap subset defined by the number of seen constituent pairs.

In PairSeen0, where all three constituent pairs are unseen, QDGL retains positive MRR gains against every baseline. On PrimeKG, its Hit@1 is 1.65 percentage points below SAFKGC while its MRR is higher. Hit@1 measures only whether the true disease ranks first, whereas MRR also accounts for its positions beyond the first rank. QDGL's greater reciprocal-rank contributions from non-first positions more than offset the gap in first-rank accuracy, yielding higher MRR. Across the strata, QDGL's ranking gains cover low-overlap samples and unseen queries, demonstrating its predictive capability under different degrees of relational reuse and for new query combinations.

S6.6.2 Entity-Pair-Disjoint Evaluation

This section considers association screening when specific pairings among previously recorded entities lack association records. For example, a molecule and a drug may each have existing records while their combination remains unrecorded, motivating a search for relevant diseases. Alternatively, the drug or molecule may lack a recorded pairing with the target disease, requiring triadic

prediction in the context of the third entity. We use three entity-pair-disjoint settings to simulate missing association records for molecular–drug, drug–disease, and molecular–disease pairs, respectively, and examine whether the model can prioritize diseases for subsequent validation of these unseen combinations.

Specifically, we separately partition positive triads by molecular–drug pairs (n, d), drug–disease pairs (d, s), and molecular–disease pairs (n, s). All positives sharing the selected pair are assigned to the same split, making that pair type disjoint across training, validation, and test sets while allowing individual entities to overlap. Each method is retrained on the corresponding training split. All three settings retain disease ranking for a molecular–drug query. Because this section focuses on disease prioritization, we use Hit@1 and MRR to assess first-rank accuracy and candidate ranking quality, with full-candidate evaluation examining screening across the complete disease candidate set. Figures S11–S13 compare QDGL with RotatE, QLogicE, SAGMM, and SAFKGC in sampled and full-candidate ranking on the three datasets.



Figure S11: Molecular–drug-pair-disjoint ranking performance across three datasets (mean ± sample SD, %). Annotations show gains over the strongest baseline in percentage points.

Under query-pair isolation, QDGL's predictive advantages for unseen molecular–drug combinations extend to the complete disease candidate set. In Figure S11, QDGL achieves the highest mean Hit@1 and MRR in both sampled and full-candidate ranking on all three datasets. On CTD in particular, its Hit@1 advantage over SAFKGC increases from 7.45 percentage points with sampled candidates to 36.86 percentage points with all candidates. Both methods achieve high first-rank accuracy among limited candidates, but the complete candidate set introduces more competitors that may outrank the true disease, making the difference in their discrimination ability more apparent. QDGL retains more correct first-rank predictions as the candidate set expands, demonstrating its ability to prioritize targets for new query combinations among a large set of diseases.

Figures S12 and S13 reveal a clear directional difference between the two disease-related pair-isolation settings. On ncRNADrug, QDGL achieves a full-candidate MRR of 70.57% under molecular–disease pair isolation, exceeding its result under drug–disease pair isolation. The model ranks candidates more effectively when the molecule–target disease pairing is unseen and drug–disease associations are allowed to remain in training. CTD exhibits the opposite relative pattern: full-candidate MRR under drug–disease pair isolation is 51.01%, higher than under molecular–

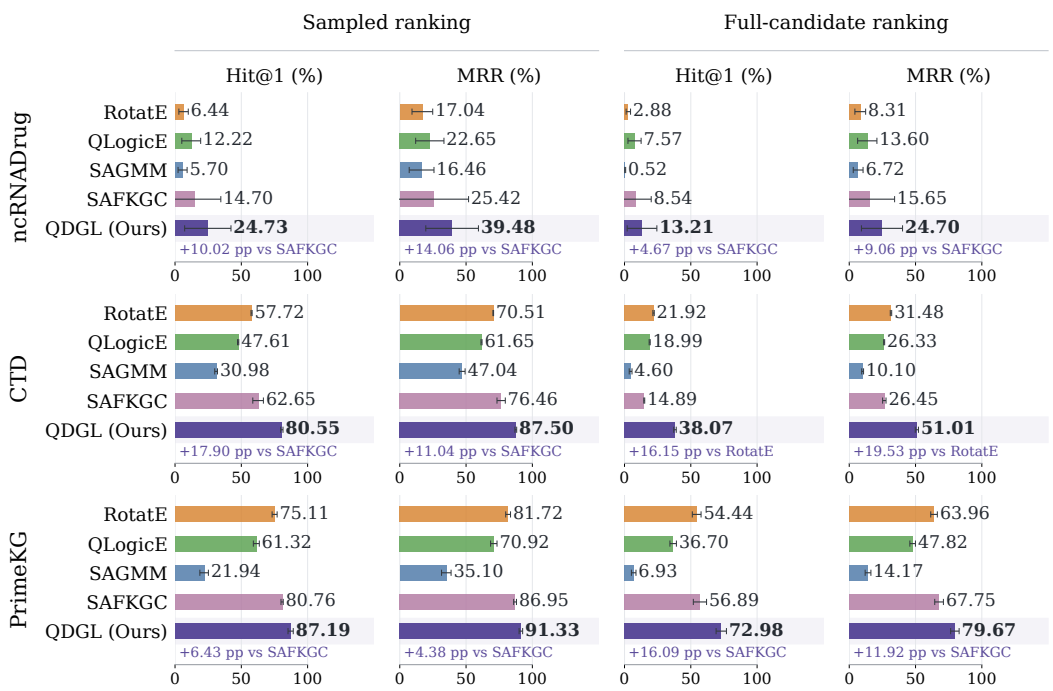


Figure S12: Drug–disease-pair-disjoint ranking performance across three datasets (mean ± sample SD, %). Annotations show gains over the strongest baseline in percentage points.



Figure S13: Molecular–disease-pair-disjoint ranking performance across three datasets (mean ± sample SD, %). Annotations show gains over the strongest baseline in percentage points.

disease pair isolation. Here, QDGL ranks candidate diseases more effectively when molecular–disease associations are allowed to remain in training and the target drug–disease pairing is unseen.

On PrimeKG, QDGL maintains a full-candidate MRR of approximately 79% under molecular–drug, drug–disease, and molecular–disease pair isolation, showing relatively consistent ranking performance across pair types.

S6.6.3 Unseen-Entity Generalization

Association prediction in practice must accommodate drugs, molecules, or diseases absent from the training data. We construct three unseen-entity splits with no training–test entity overlap for the held-out type and consistently evaluate disease ranking for a molecular–drug query. Unseen-drug and unseen-molecule settings assess whether the model can recommend relevant diseases for queries containing new entities, while the unseen-disease setting assesses whether predictions extend to target diseases without training associations.

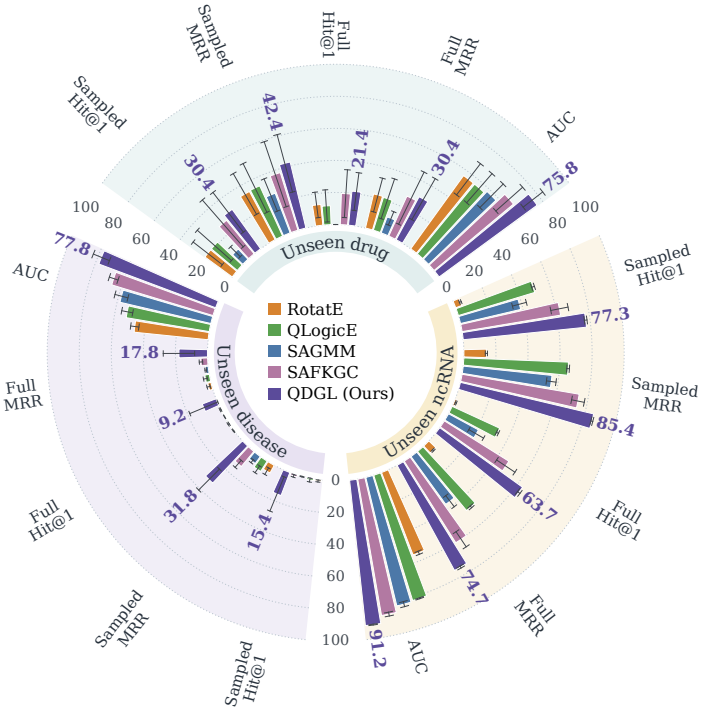


Figure S14: Unseen-entity generalization on ncRNADrug. Bars and error bars show means and sample standard deviations (%).

In Figure S14, QDGL achieves a full-candidate MRR of 74.75% for unseen ncRNAs on ncRNADrug, exceeding the strongest baseline, SAFKGC, by 14.50 percentage points, alongside higher Hit@1. Despite the absence of training associations for these ncRNAs, the model retains better ranking performance across the full disease candidate set.

On CTD, Figure S15 shows distinct ranking gains for new drug queries and new disease targets. For unseen drugs, QDGL's full-candidate MRR is close to SAFKGC, but its higher Hit@1 means that more queries directly identify the true disease as the first choice, which is particularly relevant when only a few candidates can be investigated. For unseen diseases, QDGL achieves a sampled MRR of 27.28%, more than twice that of the strongest baseline, together with higher Hit@1. The improvement covers both first-choice accuracy and candidate ranking quality measured by reciprocal rank. QDGL retains the highest Hit@1 and MRR when ranking all candidate diseases, extending this relative advantage to a larger disease-retrieval space.

Broader applicability across unseen entity types is evident in the PrimeKG results. Figure S16 shows that QDGL outperforms the comparison methods in mean ranking performance for unseen drugs, genes, and diseases. For unseen diseases in particular, full-candidate MRR reaches 46.77%, exceeding the strongest baseline by 12.49 percentage points. This setting requires distinguishing

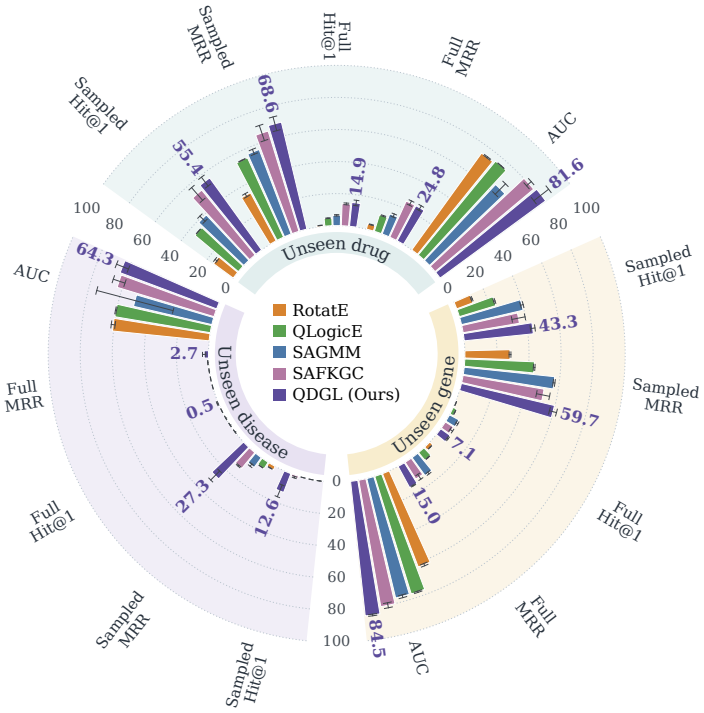


Figure S15: Unseen-entity generalization on CTD. Bars and error bars show means and sample standard deviations (%).

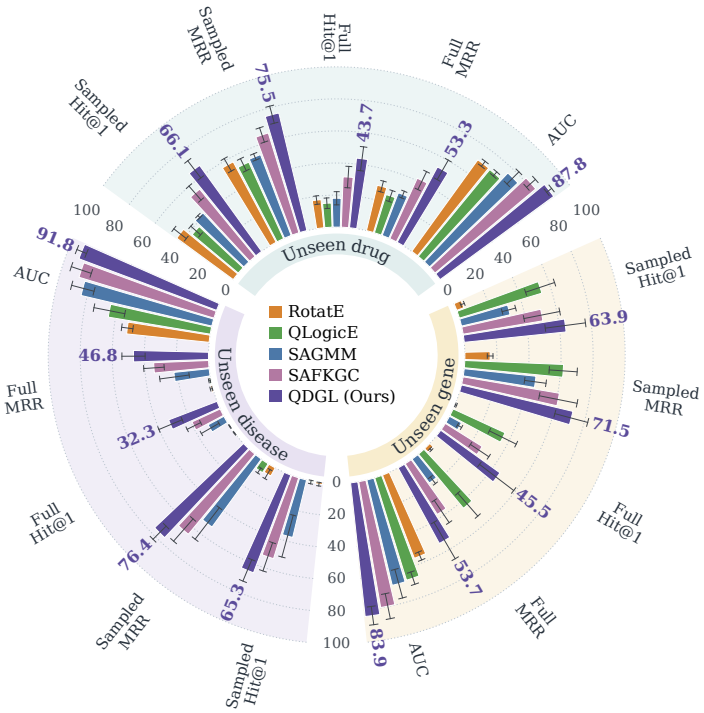


Figure S16: Unseen-entity generalization on PrimeKG. Bars and error bars show means and sample standard deviations (%).

target diseases without their training associations, and the full-candidate ranking performance supports extending association prediction to new disease contexts. Together with the unseen-drug and unseen-gene results, these findings support both identifying relevant diseases for new drugs or genes and screening new disease associations for existing molecular–drug combinations.

S6.7 Triadic Association Case Studies

This section presents triadic association case studies of drug response in non-small-cell lung cancer and breast neoplasms, focusing on ncRNAs and genes associated with drug sensitivity and resistance. The primary task ranks candidate diseases for a fixed molecular–drug pair. The case studies instead examine molecules that may participate in drug response within a specific drug–disease context, fixing the drug and disease and ranking candidate molecules using the same triadic association scoring function. We present the complete Top-10 candidates on ncRNADrug, CTD, and PrimeKG and discuss their biological roles in light of the literature. We further use local relation masking to examine how selected predictions depend on graph relations, a multi-query network to display shared molecular associations across drug contexts, and pathway enrichment to characterize the functional processes involving candidate genes.

S6.7.1 Local Relation Effects on Selected Triads

miR-326 participates in post-transcriptional gene regulation and can modulate cisplatin sensitivity in lung adenocarcinoma cells by targeting the transcription factor SP1. Previous research found reduced miR-326 expression in cisplatin-resistant cells and showed that restoring its expression reversed resistance in experimental models. In tumor tissues from lung adenocarcinoma patients receiving cisplatin-based chemotherapy, low miR-326 expression was also associated with high SP1 expression and reduced cisplatin sensitivity (Li et al., 2016). This regulatory relationship connects changes in molecular expression to chemotherapy response. We therefore examine the miR-326–Cisplatin–non-small-cell lung cancer (NSCLC) triad, focusing on its candidate rank and dependence on local graph relations.

For the Cisplatin–NSCLC query on ncRNADrug, QDGL ranks miR-326 seventh among 12,192 candidate ncRNAs. Table S10 lists the complete Top-10, six of which have direct triadic literature evidence. This miR-326 triad is absent from the dataset's positive-triad set, but published lung adenocarcinoma research has documented its association with cisplatin response. The case thus illustrates the complementarity between dataset annotations and existing biological knowledge, with a highly ranked candidate providing a lead for supplementing an unrecorded association.

GPER1 (GPR30) participates in estrogen signaling and is associated with breast cancer cell states and drug responses. Cell-based research found that doxorubicin treatment could induce epithelial–mesenchymal transition and that downregulating GPR30 suppressed this process. The GPR30 antagonist G15 also increased doxorubicin sensitivity in breast cancer cells with epithelial phenotypes (Liu et al., 2015). These findings make GPER1 a relevant subject for studying the connections between receptor regulation, cell-state changes, and chemotherapy response, motivating our analysis of the GPER1–Doxorubicin–breast-neoplasms triad.

For the Doxorubicin–breast neoplasms query on CTD, GPER1 ranks third among 6,542 candidate genes. Table S11 lists the complete Top-10, eight of which have direct triadic literature evidence.

To examine how these two predictions depend on local graph relations, we assess score and rank changes under individual and joint relation masking. Figure S17 shows the complete Top-10 on the left and the six screened neighborhood relations whose individual removal causes the largest decreases in the target score, together with background connections, on the right. The displayed relations are pairwise projections of training triads, and edge widths reflect their effects on the target score.

The local network of miR-326 connects cisplatin, NSCLC, and other tumor-related entities. Jointly masking the six selected relations changes its rank from 7 to 322, showing that these neighborhood connections substantially influence the candidate's ranking.

The local network of GPER1 includes breast neoplasms and related chemical neighborhoods. Jointly masking the six selected relations changes its rank from 3 to 5, leaving it among the leading candi-

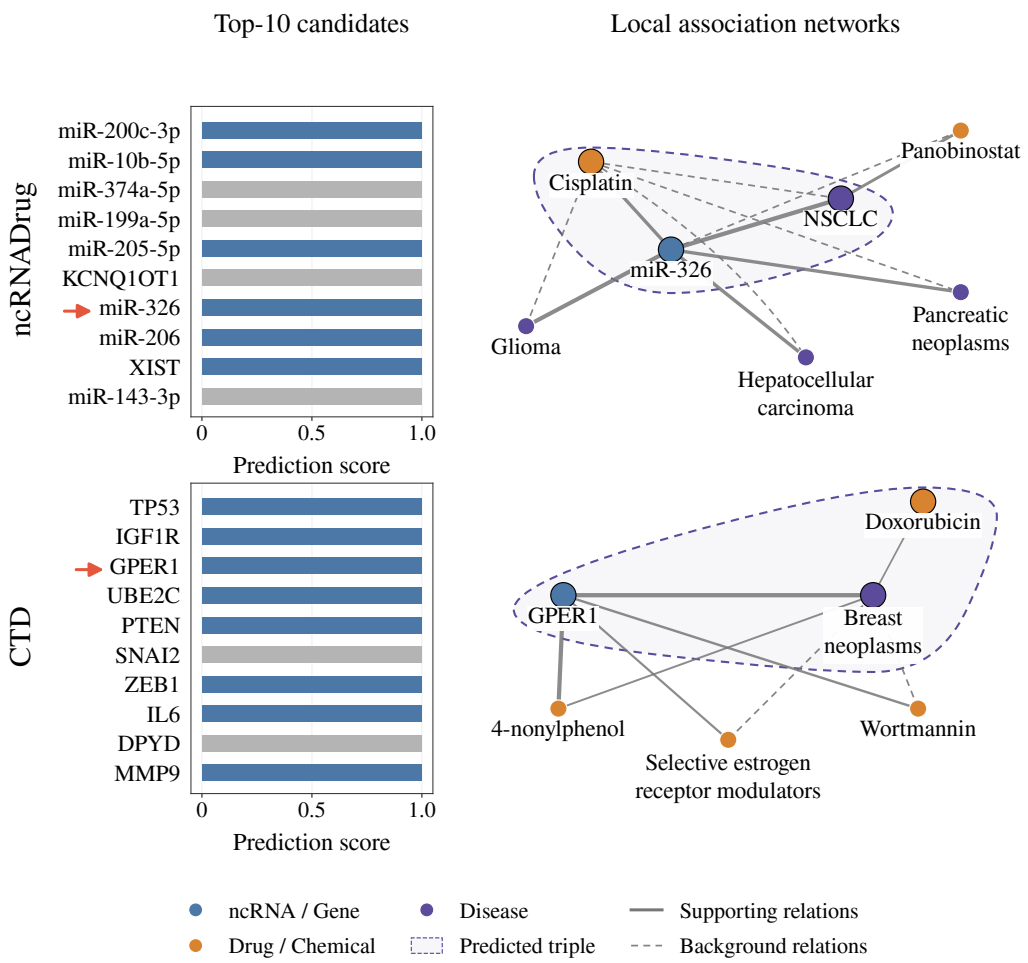


Figure S17: Top-10 candidates and local relation effects on ncRNADrug and CTD. Blue and gray bars denote candidates with and without verified direct literature support, respectively. Bar lengths show prediction scores. Arrows mark candidates expanded on the right. Thicker solid edges indicate larger decreases in the target score when individually masked. Edge widths are comparable only within each panel.

dates. The two cases exhibit different degrees of local dependence. The ranking of miR-326 is more sensitive to the selected relations, whereas that of GPER1 changes less after their removal.

S6.7.2 Triadic Associations across Drug Contexts on PrimeKG

Drug response in non-small-cell lung cancer is closely related to the molecular state of tumor cells. Sensitizing EGFR mutations provide an important molecular basis for gefitinib treatment, while cisplatin resistance also involves tumor-cell self-renewal and survival regulation (Maemondo et al., 2010; Madajewski et al., 2023). Comparing genes associated with the two drugs in the same disease context allows examination of the molecular factors involved in their respective responses and their overlap. We select the Gefitinib–NSCLC and Cisplatin–NSCLC queries on PrimeKG and organize their complete Top-10 candidates into the triadic network in Figure S18. Tables S12 and S13 list the candidate ranks and literature evidence, with direct literature support for 14 of the 20 candidate triads.

For the Gefitinib–NSCLC query, QDGL ranks EGFR first and also places MET among the leading candidates. These genes concern the drug target and bypass resistance, respectively. Gefitinib acts by inhibiting EGFR signaling, whereas MET amplification can activate ERBB3-dependent PI3K

Table S10: ncRNADrug Top-10 triads for the Cisplatin–NSCLC query.

Rank	ncRNA	Drug	Disease	PMID
1	hsa-miR-200c-3p	Cisplatin	NSCLC	20696752
2	hsa-miR-10b-5p	Cisplatin	NSCLC	34165168
3	hsa-miR-374a-5p	Cisplatin	NSCLC	—
4	hsa-miR-199a-5p	Cisplatin	NSCLC	—
5	hsa-miR-205-5p	Cisplatin	NSCLC	24084898
6	KCNQ1OT1	Cisplatin	NSCLC	—
7	hsa-miR-326	Cisplatin	NSCLC	27460077
8	hsa-miR-206	Cisplatin	NSCLC	27014910
9	XIST	Cisplatin	NSCLC	32209729
10	hsa-miR-143-3p	Cisplatin	NSCLC	—

PMIDs cite direct triadic evidence. A dash denotes a triad without verified direct literature support.

Table S11: CTD Top-10 triads for the Doxorubicin–breast-neoplasms query.

Rank	Gene	Drug	Disease	PMID
1	TP53	Doxorubicin	Breast neoplasms	22698404
2	IGF1R	Doxorubicin	Breast neoplasms	22493363
3	GPER1	Doxorubicin	Breast neoplasms	26175858
4	UBE2C	Doxorubicin	Breast neoplasms	24072565
5	PTEN	Doxorubicin	Breast neoplasms	36276272
6	SNAI2	Doxorubicin	Breast neoplasms	—
7	ZEB1	Doxorubicin	Breast neoplasms	37356459
8	IL6	Doxorubicin	Breast neoplasms	11751408
9	DPYD	Doxorubicin	Breast neoplasms	—
10	MMP9	Doxorubicin	Breast neoplasms	22516258

PMIDs cite direct triadic evidence. A dash denotes a triad without verified direct literature support.

signaling, allowing tumor cells to maintain survival signaling despite EGFR inhibition (Engelman et al., 2007).

For the Cisplatin–NSCLC query, the top-ranked candidate NQO1 is associated with resistant cell states. Previous research found that reducing NQO1 expression impaired NSCLC cell self-renewal and increased the susceptibility of cisplatin-refractory tumor spheroids to drug treatment (Madajewski et al., 2023). Another highly ranked candidate, GSTP1, was upregulated in cisplatin-resistant lung adenocarcinoma cells, and miR-513a-3p-mediated regulation of GSTP1 enhanced cisplatin-induced apoptosis (Zhang et al., 2012).

Shared genes such as EGFR and MET connect the two candidate sets, but the meanings of their associations vary with the drug context. EGFR ranks first for gefitinib and sixth for cisplatin. The former concerns a direct drug target, while research has linked abnormal EGFR activation to cisplatin resistance in wild-type EGFR NSCLC cells (Li et al., 2020). MET is implicated in bypass resistance in the gefitinib context and in AKT signaling and autophagy regulation in cisplatin-resistant cell models (Sun et al., 2018). Figure S18 displays each shared gene's associations separately across drug contexts, retaining their respective drug-response backgrounds. The cisplatin association for RRM1 concerns a gemcitabine–cisplatin combination regimen and is marked separately in Table S13.

S6.7.3 Functional Pathway Enrichment of Candidate Genes

To analyze the biological processes shared by candidate genes, we extend the three CTD and PrimeKG drug–disease queries above to their Top-50 candidates and perform human KEGG pathway enrichment using g:Profiler (Kanehisa et al., 2023). Each query uses its eligible candidate genes as the background, and enrichment tests are corrected for FDR. Figure S19 displays representative

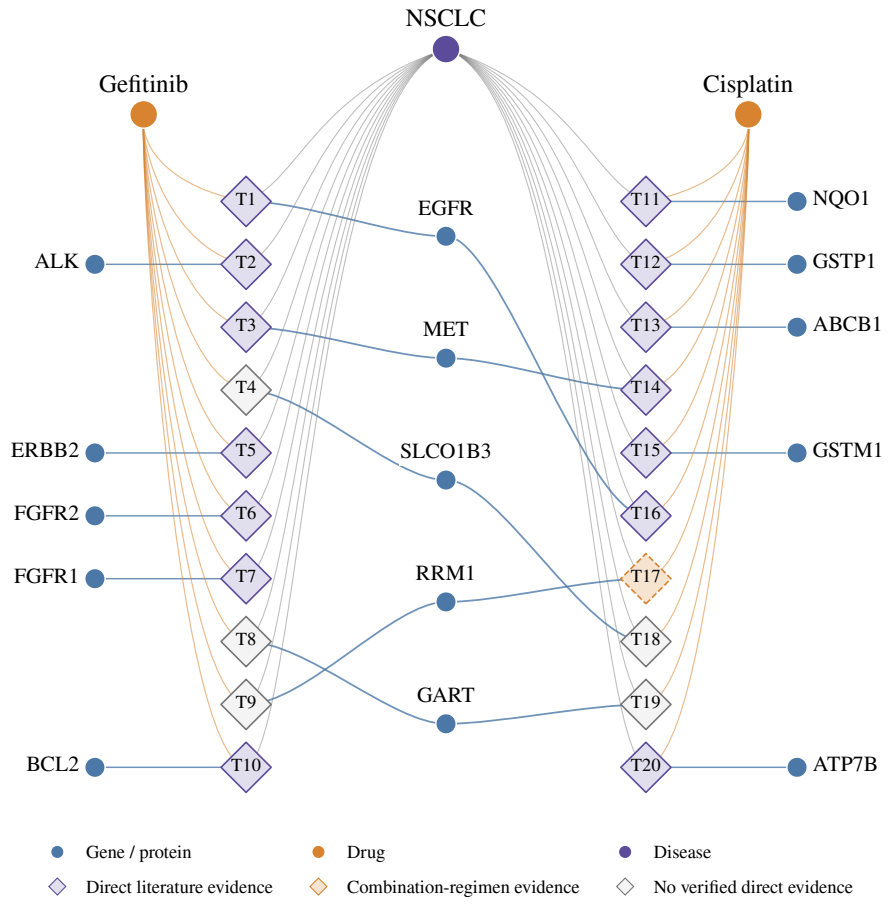


Figure S18: PrimeKG triadic network for two Top-10 candidate lists. Each diamond and its links represent a gene–drug–disease triad. T1–T10 and T11–T20 follow the ranks in Tables S12 and S13, respectively.

Table S12: PrimeKG Top-10 triads for the Gefitinib–NSCLC query (T1–T10).

Rank	Gene	Drug	Disease	PMID
1	EGFR	Gefitinib	NSCLC	20573926
2	ALK	Gefitinib	NSCLC	36387181
3	MET	Gefitinib	NSCLC	17463250
4	SLCO1B3	Gefitinib	NSCLC	—
5	ERBB2	Gefitinib	NSCLC	28105244
6	FGFR2	Gefitinib	NSCLC	21152424
7	FGFR1	Gefitinib	NSCLC	23552882
8	GART	Gefitinib	NSCLC	—
9	RRM1	Gefitinib	NSCLC	—
10	BCL2	Gefitinib	NSCLC	29422965

PMIDs cite direct triadic evidence. A dash denotes a triad without verified direct literature support.

functional and antineoplastic drug-resistance pathways with FDR below 0.05 and at least three gene hits, retaining up to six terms per query after redundancy filtering.

For the Doxorubicin–breast neoplasms query, candidate genes span damage responses, DNA repair, and cell-survival regulation. Genes including TP53 and ATM map to p53 signaling, while ATM, RAD51C, and RAD54L map to homologous recombination, linking the candidates to damage-

Table S13: PrimeKG Top-10 triads for the Cisplatin–NSCLC query (T11–T20).

Rank	Gene	Drug	Disease	PMID
1	NQO1	Cisplatin	NSCLC	36980879
2	GSTP1	Cisplatin	NSCLC	22749944
3	ABCB1	Cisplatin	NSCLC	30718500
4	MET	Cisplatin	NSCLC	29487530
5	GSTM1	Cisplatin	NSCLC	22085405
6	EGFR	Cisplatin	NSCLC	32342201
7	RRM1	Cisplatin	NSCLC	21496433†
8	SLCO1B3	Cisplatin	NSCLC	—
9	GART	Cisplatin	NSCLC	—
10	ATP7B	Cisplatin	NSCLC	22966294

PMIDs cite direct triadic evidence, except † (combination-regimen evidence). A dash denotes a triad without verified direct literature support.

induced cell-cycle arrest and apoptosis regulation and to DNA double-strand break repair, respectively. Enrichment of PI3K–Akt signaling further connects candidates such as IGF1R and PTEN to cell growth and survival regulation.

For the Gefitinib–NSCLC query, the EGFR tyrosine kinase inhibitor resistance pathway includes candidates such as EGFR, MET, and ERBB2, together with signaling regulators including RAF1, BRAF, and MTOR. Together with MAPK pathway enrichment, these results place the candidate set across receptor and downstream signaling regulation.

For the Cisplatin–NSCLC query, platinum drug resistance is the most significant displayed pathway, covering 9 of the 49 mapped input genes. Candidates including GSTP1, GSTM1, ATP7A, ATP7B, and ABCC2 are involved in glutathione-related detoxification and drug transport, processes that may influence platinum-drug responses by altering effective intracellular drug exposure. Glutathione metabolism enrichment further links GSTP1 with other candidates involved in detoxification and metabolism.

S7 LIMITATIONS

QDGL is trained and evaluated using existing triadic association annotations. Unannotated candidates may include genuine associations absent from current annotations, making performance assessment dependent on annotation completeness. Our false-negative sensitivity experiments examine changes in predictions across contamination levels. Their results characterize model performance under the specified perturbations and do not establish whether unannotated associations in real data are positive or negative.

Local relation masking, literature verification, and pathway enrichment provide model-level and biological interpretations of candidate associations, but this study does not include new wet-lab validation of the predicted candidates. Relation masking reveals how changes in model inputs affect predictions and does not directly establish the biological causal roles of candidate molecules. Candidate triads without direct experimental support require molecular functional perturbation and drug-response experiments in the corresponding drug–disease context to test the associations and determine the direction of their effects.

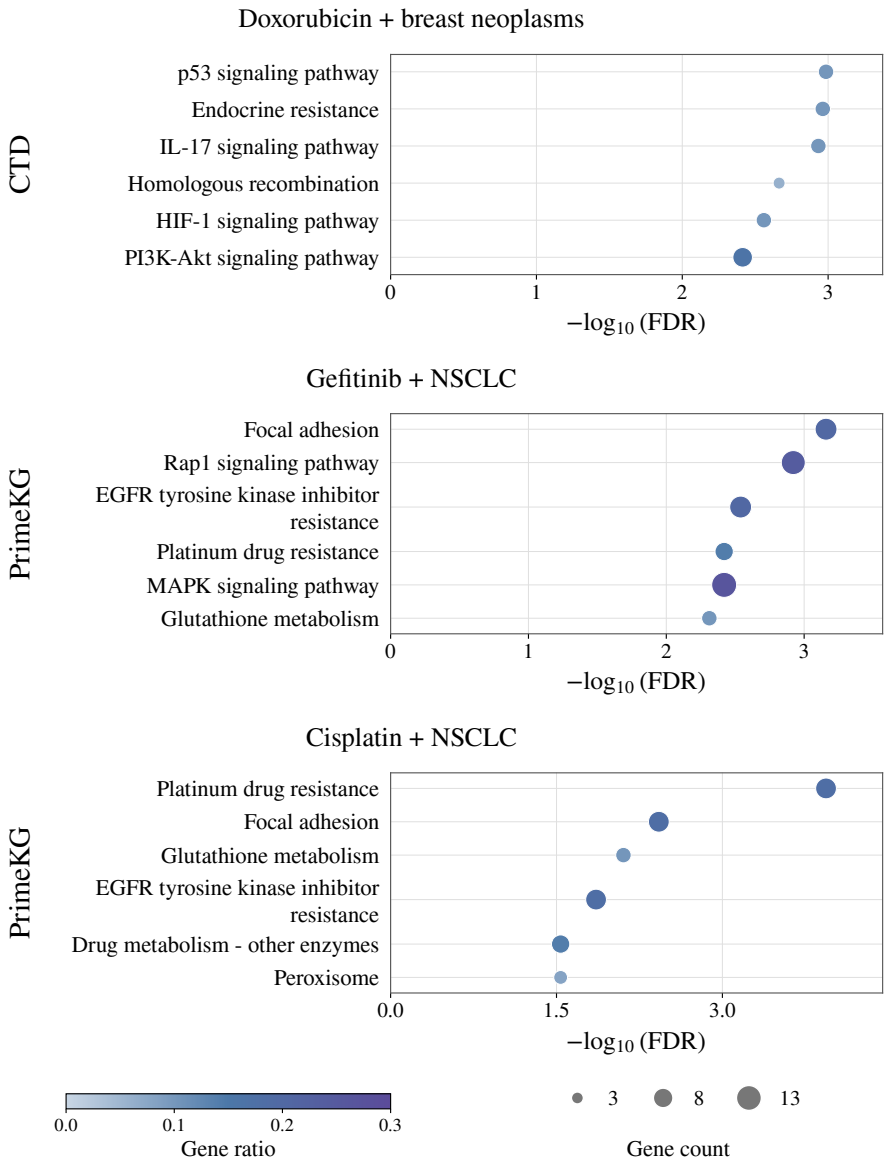


Figure S19: KEGG pathway enrichment of Top-50 candidate genes for three drug–disease queries. Bubble size and color indicate the number and proportion of mapped input genes in each pathway. The horizontal axis shows $-\log_{10}(\mathrm{FDR})$.